Technical Report No. 209

036040-5-T

THE FREE RUNNING DETECTION PROBLEM:

THE DETECTION OF RANDOMLY OCCURRING PULSES AS THEY OCCUR

by

R. M. Heitmeyer

Approved by: Theodore G. Birdsall

for

COOLEY ELECTRONICS LABORATORY

Department of Electrical and Computer Engineering
The University of Michigan
Ann Arbor, Michigan

Contract No. N00014-67-A-0181-0032
Office of Naval Research
Department of the Navy
Arlington, Va. 22217

September 1971

ABSTRACT

This study provides an extension of classical detection theory by defining and analyzing two classes of decision devices. The common feature of these devices is that they both operate in real time to detect the presence of each pulse in a sequence of randomly occurring pulses, while that pulse is still present. These decision devices are distinguished by the fact that one, the respond-once device, seeks to detect each pulse once and only once, whereas the other device, the respond-and-hold device, seeks to respond at each instant of time at which a pulse is present.

In this study, the problem of obtaining the decision devices that are optimum in the above setting is referred to as the free running detection problem. The basic assumptions are as follows. It is assumed that all of the pulses are identically shaped and completely known up to their arrival times. Moreover, we assume that it is possible to observe these pulses only through an observation that contains additive noise. Finally, it is assumed that the decision devices are allowed to make decisions on the presence or absence of pulses only at discrete, equally-spaced points in time.

The mathematical model for the free running detection problem is constructed in terms of the general decision theory model. This provides the theoretical basis for describing the performance of both

classes of decision devices and for obtaining decision devices that are optimum with respect to the Bayes criteria.

The performance of the respond-once decision device is described in terms of the average number of pulses that are detected (detection rate), the average number of false alarms (false alarm rate), and the average number of times a pulse is detected more than once (extra detection rate). It is shown that the detection rate can be increased for the same false alarm rate by increasing the extra detection rate.

The performance of the respond-and-hold decision devices is described in terms of the detection rate, a detection duty, and a false alarm duty. These last two quantities are a measure of the amount of time that the decision device responds when a pulse is present. For the respond-and-hold devices, it is seen that the detection duty can be increased for a fixed false alarm duty only by decreasing the detection rate.

FOREWORD

In the past two decades, classical detection theory has been profitably applied to the design and evaluation of active surveillance systems. The applicability of detection theory is a result of the natural temporal structure inherent in these systems. That is, in the active surveillance system it is sufficient to make a decision at the end of an interval of time on the presence or absence of signal in the received waveform. In certain passive surveillance situations, however, the lack of temporal structure precludes the use of a "forced-response" detection theory as a realistic model for the systems detector. One such situation arises when targets are continually entering and departing from the field of view and it is desired to detect these targets while they are still present. In a situation such as this a "free-response" detection theory must be used to model the system detector. This report establishes the foundations of such a theory. A recursive relationship describing the optimum detector is developed and a scheme for evaluating the performance of any free-response detector is obtained in terms of an "ROC surface." Finally, some numerical results are presented as applications of the theory.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (CONT.)

LIST OF ILLUSTRATIONS (CONT.)

LIST OF ILLUSTRATIONS (CONT.)

LIST OF ILLUSTRATIONS (CONT.)

LIST OF ILLUSTRATIONS (CONT.)

LIST OF ILLUSTRATIONS (CONT.)

LIST OF SYMBOLS

Symbol	Definition
$A_k(.,.)$	response set at time t_k
$\mathcal{A}$	action space
a	parameter of the discrete prior probability law
a_k	decision at time t_k
$\vec{a}$	total decision
C	correct rejection outcome
D	detected pulse outcome
$\mathcal{D}(m)$	the set of decision devices for a fixed m
D_p, D_D, D_F	pulse duty, detection duty, false alarm duty
F	false alarm outcome
$F(\cdot)$	commutative distribution function
$I_B(\cdot)$	indicator function for the set B
$K(\cdot)$	threshold function for the $m=1$ Bayes decision device
$K^0(\cdot), \hat{K}^0(\cdot)$	threshold function for the $m=2$ Bayes decision device at time t_k given that $a_{k-1}=0$, the limit of the sequence $K_k^0(\cdot)$

LIST OF SYMBOLS (Cont.)

Symbol	Definition
$K_k^1(\cdot)$, $\hat{K}^1(\cdot)$	threshold function for the $m=2$ Bayes decision device at time t_k given that $a_{k-1} = 1$, the limit of the sequence $K_k(\cdot)$
$L(\cdot, \cdot)$	total loss function
$L_k(\cdot, \cdot, \cdot)$	loss at time t_k
L_D, L_F, L_M, L_X	the loss per detect pulse, false alarm, missed pulse and extra detection, respectively
L_C^T, L_D^T, L_F^T, L_M^T	the loss per unit correct rejection time, per unit detection time, per unit false alarm time and per unit miss time respectively
$\ell_k(x_k \mid \vec{x}_{k-1}, \theta_k, \theta_{k-1})$	the Raydon-Nikodym derivative of $P(dx \mid \theta_k, \theta_{k-1})$ with respect to $P(dx \mid \vec{x}_{k-1})$
$\ell_k(x_k \mid \theta_{k-1}, S_{co})$, $\ell_k(x_k \mid \theta_k, S_c)$, $\ell_k(x_k \mid \theta_k, \theta_{k-1}, S_{co}, S_o)$	the carry-over pulse likelihood ratio, the current pulse likelihood ratio, the combined pulse likelihood ratio
$\ell_k(x_k)$	the likelihood ratio for white noise and translation invariant pulses
M	missed pulse outcome
MPD	maximum pulse detection device
MDT	maximum detection time device
m	the number of decision opportunities per pulse

LIST OF SYMBOLS (Cont.)

Symbol	Definition
n	the stochastic process representing noise
N	the total number of decision times
$N_C, N_D, N_F, N_M, N_R, N_X$	the total number of correct rejection outcomes, detected pulse outcomes, false alarm outcomes, missed pulse outcomes, rest outcomes, and extra outcomes, respectively
N_N, N_S	the total number of decision times where noise alone is present and where signal is present, respectively
N_{R_N}, N_{R_S}	the total number of decision times at which responses are made and noise is present, and signal is present, respectively
N_o	noise power per unit bandwidth
N_p	the total number of pulses
$O_k(\cdot, \cdot, \cdot)$	odds ratio density at time t_k
P_k	carry over odds ratio
$p(t)$	the basic pulse shape
$p^k(t)$	the basic pulse shape with arrival time $t_k - T_p$
$P(\cdot \mid \cdot)$	conditional probability measures on the observation space
$p(\theta_k \mid \theta_{k-1})$	prior conditional density
$p(\theta_k)$	prior unconditional density
P_D, P_F	probability of detection and the probability of false alarm

Symbol	Definition
$Pr[B]$	probability of the event B
q	number of allowable pulse arrival times per T_p seconds
R	rest outcome
$R_k(\vec{x}_k, a_k, \tau_k)$	posterior risk at time t_k
$R_C, R_D, R_F, R_M, R_R, R_X$	the normalized rates for a correct rejection, a detection, a false alarm, a miss, a rest, and an extra detection, respectively
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^{[0,T]}$	the set of real valued functions defined on $[0, T]$
R-O	respond-once decision device
R-H	respond-and-hold decision device
$r_C, r_D, r_F, r_M, r_R, r_X$	the unnormalized rates for the six basic outcomes listed above
r_N, r_S	the expected value per unit time of N_N and N_S
r_p	the pulse rate
$s(t, \cdot)$	total signal
s	the number of allowable pulse arrival times per Δ seconds
$\mathscr{S}, \tilde{\mathscr{S}}$	respond-once performance set and respond-and-hold performance set

LIST OF SYMBOLS (Cont.)

Symbol	Definition
T	length of the total observation interval
T_C, T_D, T_F, T_M	total amount of correct rejection time, detection time, false alarm time, and miss time, respectively
T_N, T_S	total amount of noise alone time and signal time, respectively
T_p	minimum time between pulse arrivals
$T_{t'}(\cdot, \cdot)$	test function at time t'
$\mathscr{T}$	set of allowable decision times
$U(\cdot, \cdot)$	odds ratio density updating function
W_F, W_X	normalized false alarm loss, normalized extra detection loss
$W_X^*(W_F)$	the function defining the inhibit rule
$\vec{x}, \vec{X}$	total observation, total observation space
x_k, X_k	current observation, current observation space
$\vec{x}_k, \vec{X}_k$	total past observation, total past observation space
X	extra detection outcome
α	parameter of prior probability law
$\delta(\cdot)$	decision rule
$\delta_k(\cdot)$	component decision rule

Symbol	Definition
Δ	decision time separation
$\epsilon_{\{c\}}(\cdot)$	the measure that assigns unit mass to the point c
η_i	inter-arrival time between i-1 st and ith pulse
$\vec{\theta}, \Theta$	total parameter, total parameter space
θ_k, Θ_k	parameter for the interval $[t_k - T_p, t_k)$, $[t_k - T_p, t_k]$
$\lambda_{\mathscr{P}}, \lambda_{\tilde{\mathscr{P}}}$	respond-once ROC surface, respond-and-hold ROC surface
ν	separation between the allowable pulse arrival time
$\Pi(\cdot)$	prior and posterior probability distribution functions on the space Θ
σ_i	arrival time of the ith pulse
τ_t	most recent response function

LIST OF APPENDICES

CHAPTER I

THE FREE RUNNING DETECTION PROBLEM

The free running detection problem (FRD) is a decision problem that is concerned with the detection of randomly occurring pulses as they occur. In this study we will develop the basic decision theoretic model for this problem and use this model to obtain the optimum (Bayes) decision rules. This is done in Chapters II through V. In this chapter, we introduce the FRD problem as the decision problem associated with two different pulse-monitoring systems. We then examine certain properties of the decision rules for the FRD problem from an intuitive point of view.

1.1 The Continuous Action and the Discrete Action Pulse Monitoring Systems

By a pulse monitoring system, we mean a system which takes as its input randomly occurring pulses (unknown arrival times) and produces at its output an action for each pulse that occurs. Such a system is shown in Fig. 1.1.

In this section two different pulse monitoring systems are examined. The first system of interest will be referred to as the continuous action pulse monitoring system. This system is characterized by the fact that it takes an action at each instant of time that

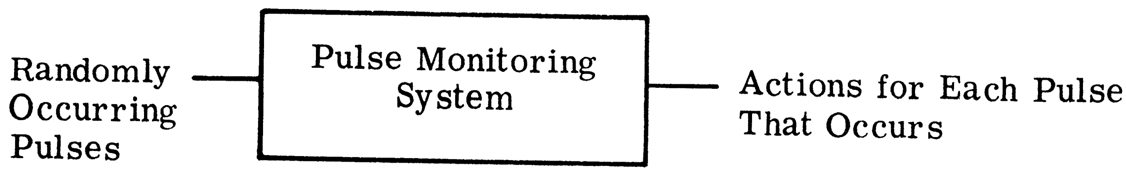


Fig. 1.1. General pulse monitoring system

a pulse is present. A simple example of such a system is an alarm system. Here the appropriate action is to sound an alarm precisely at those times at which a pulse is present.

In contrast to the continuous action system, we also consider the discrete action pulse monitoring system. This system is characterized by the fact that it takes a single action each time a pulse occurs. Perhaps the simplest example of a discrete action system is a system that provides a running count of the number of pulses that have occurred up to the current time. In this system, the appropriate action is to increment a counter each time a pulse occurs.

Additional insight into the nature of these systems is obtained by decomposing the general system model into two components as shown in Fig. 1.2.

The components in this system configuration have the following interpretation. The decision device takes as its input the randomly occurring pulses and extracts from these pulses the information that is needed by the action device in order to take the appropriate

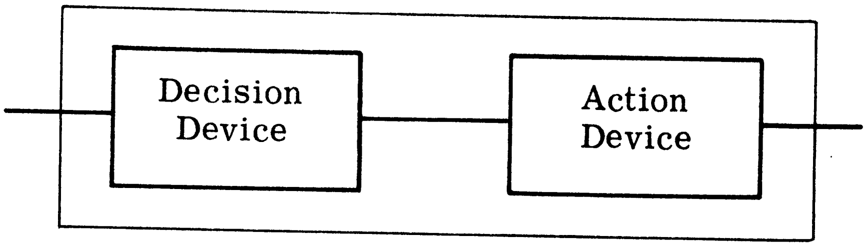


Fig. 1.2. The general pulse monitoring system as a decision device and an action device

action. The action device then takes the necessary action.

The continuous action system can be interpreted in terms of this model by noting that the information required by the action device is simply the knowledge of those times when signal pulses are present. Thus, we may take the decision device to be a device which produces the function

$$a(t) = \begin{cases} 1 & \text{if some pulse is present at time } t \\ 0 & \text{otherwise} \end{cases} \qquad (1.1)$$

We may then take the action device to be the device which produces the appropriate action at those times when $a(t) = 1$.

Now consider the discrete action system. In this system, the information needed by the action device is the knowledge of a single time when each pulse is present. This information is provided by a decision device that produces the function

$$a(t) = \begin{cases} 1 & \text{if some pulse arrives at time } t \\ 0 & \text{otherwise} \end{cases} \tag{1.2}$$

As above we may take the action device to be a device that is on for $a(t) = 1$ and off otherwise.

In the work to follow the output functions of the decision devices, $a(t)$, are referred to as response functions and the values $a(t) = 1$ as responses. A typical input sequence of pulses and the corresponding response functions for both decision devices are shown in Fig. 3 a, b, c.

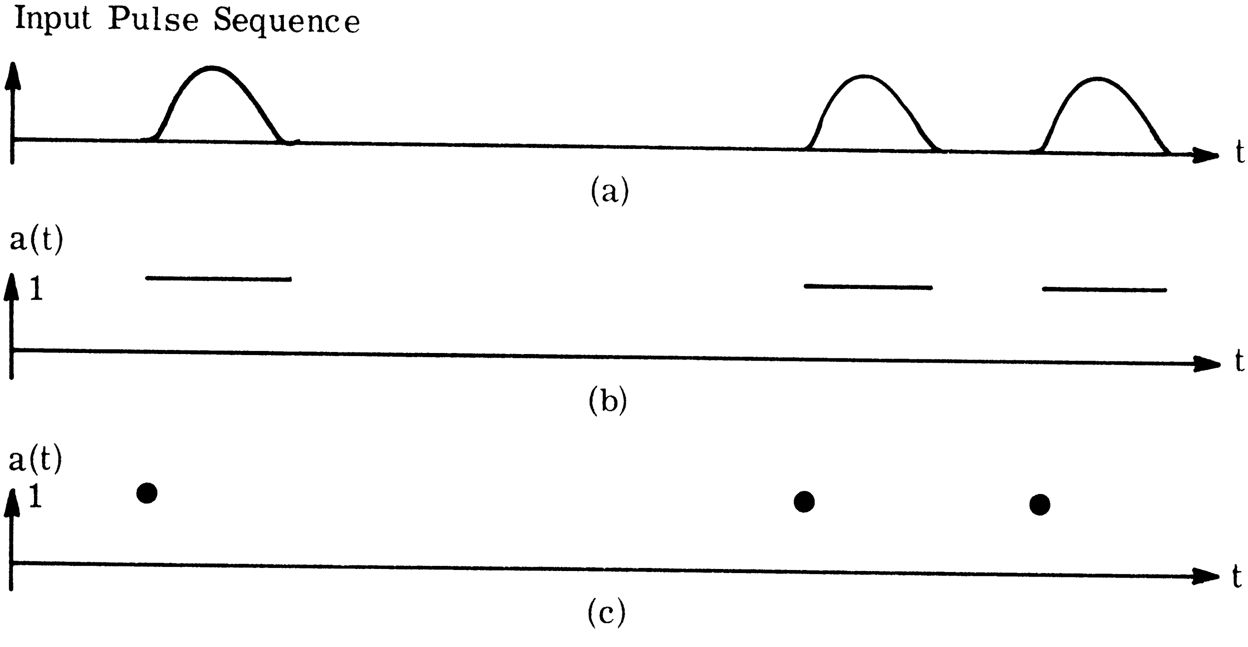


Fig. 1.3. The relation between the input pulse sequence and the response functions for the continuous action and the discrete action decision devices.
(a) input pulse sequence,
(b) continuous action response function,
(c) discrete action response function

The point to be noted in connection with the above system model is that both the continuous action and the discrete action systems can be characterized by a decision device that produces a binary valued response function $a(t)$. In both of these systems the responses $(a(t) = 1)$ occur at times when the presence of a pulse calls for an action to be taken. Thus, in both of these systems, we may view the decision device as a device that detects randomly occurring pulses as they occur. The decision devices for the two systems differ, however, in the fact that the continuous action decision device seeks to respond as long as a pulse is present, whereas the discrete action decision device seeks to respond once and only once to each pulse that occurs. To emphasize this distinction we will hereafter refer to the continuous action decision device as a respond-and-hold (R-H) decision device and we will refer to the discrete action system decision device as a respond-once (R-O) decision device.

1.2 The Performance Degradation Due to Noisy Inputs

When the input to the decision devices of the preceding section consists only of the randomly occurring pulses, the ideal performance illustrated in Fig. 1.3 can be realized. But, if noise is also present at the input to the decision device, this performance will be degraded. In this section the nature of this performance degradation is examined.

We consider first the effect of noise on the performance of the R-H decision device. Figure 1.4 shows a typical pulse sequence, a corresponding noisy input and a typical respond-and-hold response function.

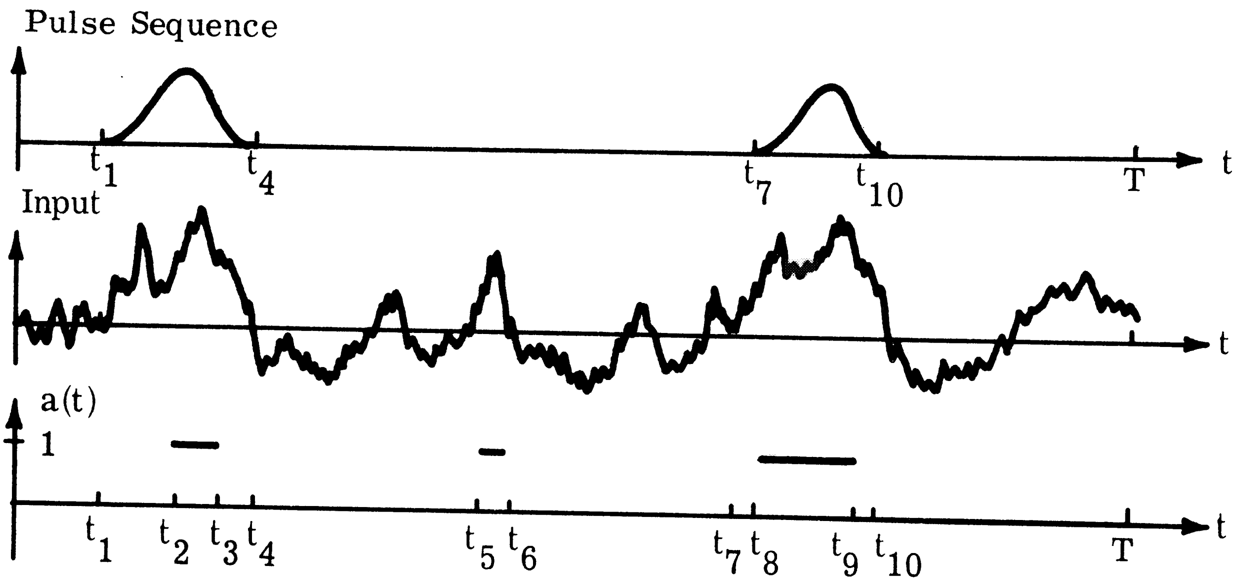


Fig. 1.4. Typical inputs and outputs for a respond-and-hold decision device

By comparing the pulse sequence and the response function in the above figure it is clear that the time axis can be divided into four sets, detection times [$a(t) = 1$ and pulse present], false alarm times [$a(t) = 1$ and no pulse present], miss times [$a(t) = 0$ and pulse present] and correct rejection times [a(t) = 0 and no pulse present]. For example, if we let T_D, T_F, T_M and T_C denote the total amount of detection time, false alarm time, miss time and correct rejection time in the interval [0, T] respectively, then from Fig. 1.4 above,

$$T_D = (t_3 - t_2) + (t_9 - t_8)$$

$$T_F = (t_6 - t_5)$$

$$T_M = (t_2 - t_1) + (t_4 - t_3) + (t_8 - t_7) + (t_{10} - t_9)$$

$$T_C = t_1 + (t_5 - t_4) + (t_7 - t_6) + (T - t_{10})$$

A numerical expression for the performance degradation of the R-H decision device can be obtained in terms of a loss function. A natural definition for the loss function for this system is

$$L = L_D{}^T\, T_D + L_M{}^T\, T_M + L_F{}^T\, T_F + L_C{}^T\, T_C \qquad (1.3)$$

Here $L_D{}^T$, $L_M{}^T$, $L_F{}^T$ and $L_C{}^T$ represent the loss per unit detection time, per unit miss time, per unit false alarm time and per unit correct rejection time, respectively. It should be noted that, in all physical situations, detection time is preferred to miss time and correct rejection time is preferred to false alarm time. Thus we may assume that $L_D{}^T \le L_M{}^T$ and $L_C{}^T \le L_F{}^T$.

A more general definition of the respond-and-hold loss function is needed for systems that must show special concern for the possibility of missing a pulse altogether. For example, in Fig. 1.5 we have shown a typical pulse sequence and two different response functions. Both of these response functions have been constructed to have the same values of T_D, T_M, T_F, T_C, so that by Eq. 1.3, both response functions have the same loss. But, the response function $a_1(t)$ detects all three pulses (responds at least once to each pulse), whereas the response function $a_2(t)$ detects only one of the pulses. This fact may be incorporated into the loss function by adding in the term

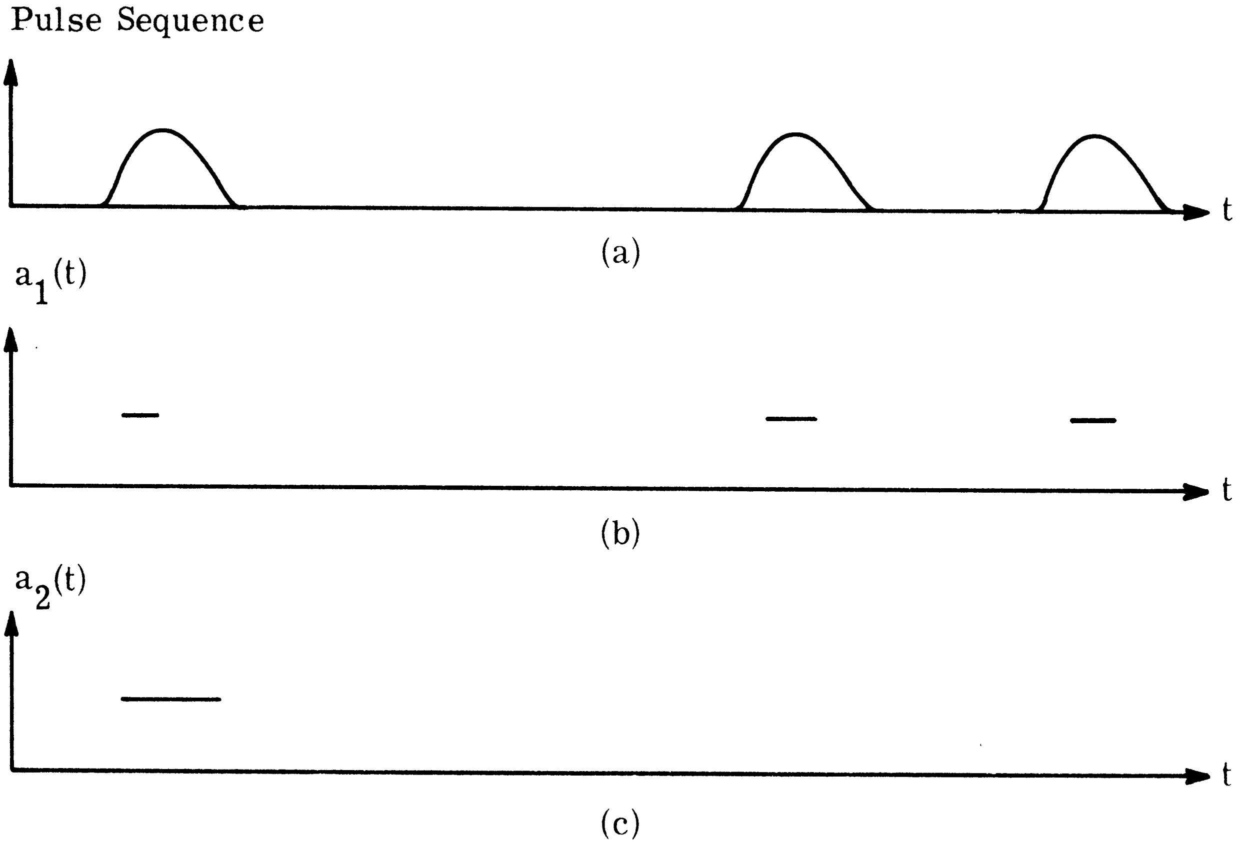


Fig. 1.5. A pulse sequence and two different response functions

$$L_D N_D + L_M N_M$$

where N_D and N_M represent the number of pulses that are detected and the number of pulses that are missed, respectively, and L_D and L_M represent the loss for a detected pulse and the loss for a missed pulse, respectively. The complete expression for the respond-and-hold loss function is then

$$\begin{aligned} L &= L_D N_D + L_M N_M \\ &\quad + L_D^T T_D + L_M^T T_M + L_F^T T_F + L_C^T T_C \end{aligned} \tag{1.4}$$

As above, physical considerations allow us to assume that $L_D \leq L_M$.

Before turning to the analysis of the respond-once decision device, it should be noted that the six quantities involved in the respond-and-hold loss function are not independent. This fact is a consequence of the following observations. First, note that each pulse that occurs must either be detected or missed. Thus, if we denote the number of pulses to occur by N_p, then we may write

$$N_p = N_D + N_M \tag{1.5a}$$

Next, note that, if a pulse is present at time t', then t' must either be a detection time or a miss time. On the other hand, if no pulse is present at t', then t' must be either a false alarm time or a correct rejection time. Thus, if we denote the total amount of time occupied by pulses by T_S and the total amount of time occupied by noise alone by T_N, then we may write

$$T_S = T_D + T_M \tag{1.5b}$$

and

$$T_N = T_F + T_C \tag{1.5c}$$

We may then substitute for N_M, T_M, and T_C from Eq. 1.5 into Eq. 1.4 to obtain the equivalent expression for the respond-and-hold loss function.

$$L = [(L_D - L_M) N_D + (L_D^T - L_M^T) T_D + (L_F^T - L_C^T) T_F] + [L_M N_p + L_M^T T_S + L_C^T T_N] \tag{1.6}$$

Now the point to be noted here is that the second bracketed term on the right hand side of Eq. 1.6 depends only on the pulse content of the interval of concern, $[0, T]$, and not on the particular decision device. Thus, for the purposes of comparing the performance of different decision devices, it is sufficient to consider the loss function

$$\tilde{L} = (L_D - L_M) N_D + (L_D^T - L_M^T) T_D + (L_F^T - L_C^T) T_F \tag{1.7}$$

Next we consider the performance degradation of the respond-once decision device. First, recall from the preceding section that the objective of the discrete action system is to take a single action each time a pulse occurs. This is accomplished in the noiseless case by requiring that the decision device respond at the instant that each pulse arrives. (See Fig. 1.3c.) But, when noise is present at the input to the system, this is no longer a reasonable requirement. Instead, it is usually sufficient to require only that the decision device respond within some prespecified time of the arrival of a pulse in order to detect that pulse and fulfill the objective of the system. Thus, we shall say that a pulse is detected if some response occurs within T_p seconds of its arrival. A detected pulse is shown in Fig. 1.6.

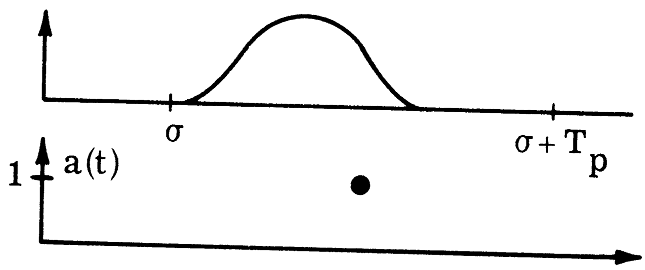


Fig. 1.6. A detected pulse

In order to insure that one response cannot detect more than one pulse, we will also assume that the arrival times of the pulses are separated by at least T_p seconds.

Now consider a typical pulse sequence, a corresponding noisy input and a typical respond-once response function as shown in Fig. 1.7.

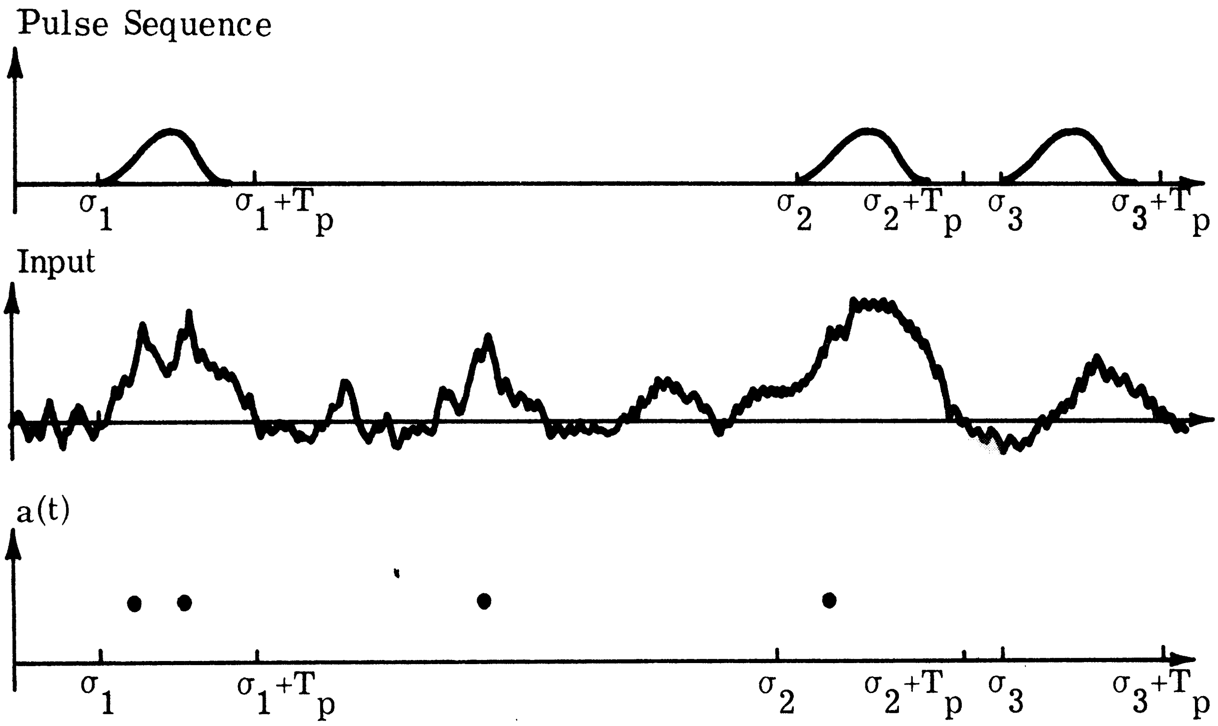


Fig. 1.7. Typical inputs and outputs for a respond once decision device

Here, it is seen that the effect of noise on the decision device performance is to cause some pulses to be missed, some pulses to be detected more than once, and some false alarms to be made. Let N_F be the number of false alarms that occur and let N_X be the number of extra detections that occur. Then the performance degradation can be expressed numerically by defining the loss function for the respond-once decision device to be

$$L = L_D N_D + L_M N_M + L_X N_X + L_F N_F \qquad (1.8)$$

where L_X and L_F are the loss per extra detection and per false alarm, respectively.

As above the respond-once loss function may be separated into a part that depends only on the pulse content of the interval of concern. This is done by substituting Eq. 1.5a into Eq. 1.8. The result is

$$L = \tilde{L} + L_M N_p \qquad (1.9)$$

where

$$\tilde{L} = (L_D - L_M) N_D + L_X N_X + L_F N_F \qquad (1.10)$$

In the above paragraphs, loss functions have been introduced as a means of measuring the performance degradation of the respond-and-hold and the respond-once decision devices. The basic assumption here is that, in the noisy case, the overall objectives of the system can be characterized by assigning specific values to the losses that appear in these functions. Thus, this section concludes with several important examples that illustrate the role these losses play in characterizing the system objective.

First, consider the respond-and-hold loss function of Eq. 1.7. The most important special case here occurs when $L_D = L_M$. In this case, the loss function depends only on the total amount of detection time T_D and the total amount of false alarm time T_F but not on the number of detected pulses N_D. Thus, this case corresponds to a decision device that seeks only to respond as long as possible to each pulse that occurs with no extra penalty for missing a pulse. It might be noted here that if we also set $L_D^T - L_M^T = L_F^T - L_C^T$ then this loss function provides a measure of the mean square error between the ideal response function for the noiseless case and the actual response function. In the following, this loss function is referred to as the maximum detection time (MTD) loss function.

As a contrast to the MTD loss function suppose that

$$L_D - L_M << L_D^T - L_M^T < 0$$

In this case, the system incurs a large gain, $-(L_D - L_M)$, for each detected pulse, but only a small gain, $-(L_D^T - L_M^T)$, for the total amount of detection time. Thus, this decision device seeks to respond at least for a short period of time to each pulse that occurs in order to detect that pulse, but the actual amount of detection time is not that important.

Next consider the respond-once loss function of Eq. 1.10. Here, L_X is the loss for an extra detection and L_F is the loss for a false alarm. Since both of these outcomes are undesirable, we may assume that $L_X, \ L_F \geq 0$. The case $L_X > L_F$ corresponds to a system that prefers false alarms to extra detections and, if $L_F < L_X$, the converse is true. The case $L_X = L_F$ corresponds to a system that views extra detections and false alarms as equivalent as is the case in the example of the running counter of Section 1.1.

An important special case of the respond-once loss function is defined by the condition $L_X = 0$. In this case, the loss function depends only on the number of pulses that are detected and the number of false alarms. Thus, this decision device simply seeks to respond "at least" once to each pulse that occurs rather than "once and only once" as is the case when $L_X > 0$. In the following, we will refer to this case as the maximum pulse detection (MPD) case.

1.3 The Free Running Detection Problem

In the preceding section we considered the effect of noise on the performance of two decision devices that arise in certain pulse monitoring systems. As a measure of the resulting performance degradation, a loss function for each device was introduced. In this study we shall be concerned with the problem of obtaining the decision devices that are optimum in the sense that they minimize the average loss. This problem shall be referred to as the free running detection problem (FRD). In this section, a formal statement of the FRD problem is presented.

We begin by specifying the decision device input. As discussed in the preceding sections, this input consists of a sequence of randomly occurring pulses together with noise. Consider first the pulse sequence itself. In this study it is assumed that each pulse is identical in shape and completely known except for its arrival time. These pulses may then be characterized in terms of a signal $s(t, \sigma_1, \sigma_2, \ldots)$ defined by

$$s(t, \sigma_1, \sigma_2, \ldots) = \sum_j p(t - \sigma_j) \tag{1.11}$$

where σ_j is the arrival time of the jth pulse and where $p(t)$ is the shape of the pulse with arrival time $t = 0$. In addition, it is assumed that the pulses are non-overlapping in the sense that there exists a number T_p with the property that

$$\sigma_{j+1} \geq \sigma_j + T_p \qquad \text{for all } j \tag{1.12}$$

and such that $p(t)$ is non-zero only over the interval $[0, T_p)$. The number T_p is referred to as the pulse duration. A typical pulse $p(t)$ and signal $s(t, \sigma_1, \sigma_2, \ldots)$ are shown in Fig. 1.8. It should be mentioned in connection with the definition of the signal $s(t, \sigma_1, \sigma_2)$ that we allow $p(t)$ to be zero over part of the interval $[0, T_p)$. For example, the waveform shown in Fig. 1.9 is an acceptable pulse waveform.

To complete the definition of the decision device input it is assumed that the noise is additive. That is, if the input is denoted

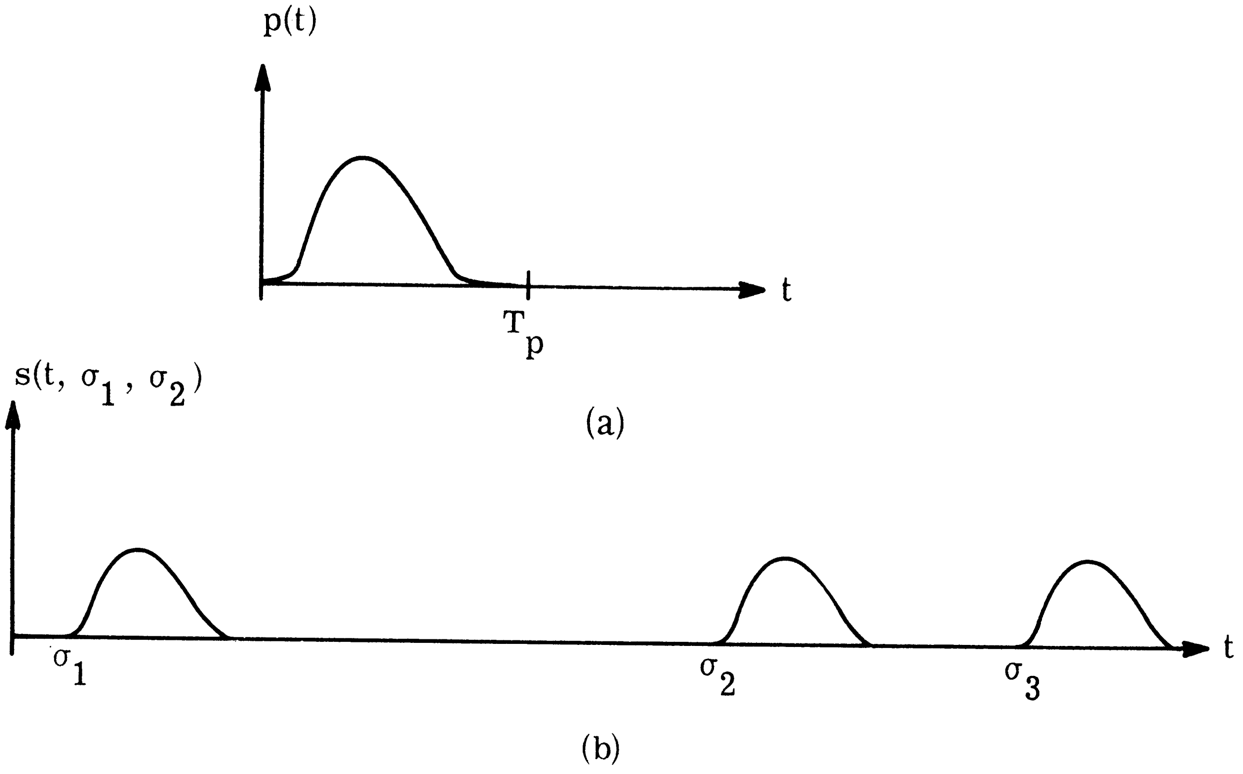


Fig. 1.8. A typical pulse p(t) and signal $s(t, \sigma_1, \sigma_2, \ldots)$

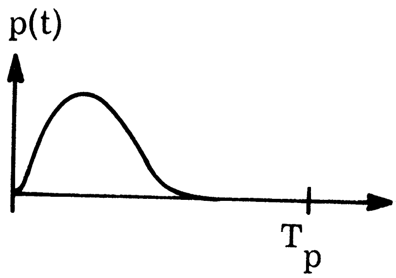


Fig. 1.9. A pulse p(t) that is zero over part of the interval $[0, T_p)$

by $x(t)$, then

$$x(t) = n(t) + s(t, \sigma_1, \sigma_2, \ldots) \qquad t \in [0, T]$$

where $n(t)$ is a realization of some known stochastic process which represents the noise in the system. In the following we will also refer to $x(t)$ as the observation and we will refer to the interval $[0, T]$ as the total observation interval.

Next, consider the output of the decision device. As discussed in the previous sections, this output is a binary valued function $a(t)$, $a(t) \in \{0, 1\}$, called the response function. For the respond-once decision device, we will impose the condition that the number of responses in the interval $[0, T]$ is finite. That is, it is assumed that the set of response times,

$$\{t\,; a(t) = 1\}$$

consists of a finite number of isolated points. For the respond-and-hold decision device, we will stipulate that responses may occur only over half-open intervals. More specifically, we will assume that the set of response times can be written as a finite union of disjoint intervals of the form $[a, b)$. In the following, the value of the response function at a particular time t', $a(t')$, will often be referred to as the decision at time t'.

Let us now turn to the description of the decision device itself.

This device operates in real time to produce the response function $a(t)$ from the observation $x(t)$. Thus, we may formally consider the decision device as a function δ which maps observations $x(t)$ into response functions $a(t)$. We may incorporate the fact that the decision device must operate in real time by requiring that δ have the property that $a(t')$ can depend only on $x(t'')$ for $t'' < t'$. In the following we will also refer to δ as a decision rule.

The basic definition of the FRD problem is completed by defining the respond-and-hold and the respond-once loss functions. From Eqs. 1.4 and 1.8, it is seen that the respond-and-hold loss function is a weighted sum of the six quantities T_D, T_M, T_C, T_F, N_D and N_M, and the respond-once loss function is a weighted sum of the four quantities, N_F, N_X, N_D and N_M. Thus, in order to define these loss functions, it is only necessary to specify how these quantities are calculated from each response function $a(t)$ and signal $s(t, \sigma_1, \sigma_2, \ldots)$. To this end, we will classify each possible decision $a(t')$ into one of six different outcomes, a false alarm (F), a current rejection (C), a detected pulse (D), an extra detection (X), a missed pulse (M) and a rest (R). It will then be shown that the quantities that appear in the loss functions can be defined in terms of these outcomes.

First, define a false alarm and a correct rejection. Now both of these outcomes occur when no pulse is present. Thus, if we define the condition that no pulse is present at time t' by the condition that $t' \notin [\sigma_j, \sigma_j + T_p)$ for any σ_j, then:

> A false alarm (F) occurs at time t' iff $a(t') = 1$ and no pulse is present at time t'

and

> A correct rejection (C) occurs at time t' iff $a(t') = 0$ and no pulse is present at time t'

Next, we define the detected pulse outcome and the extra detection outcome. Both of these outcomes occur when a pulse is present at time t'; that is, when $t' \in [\sigma_j, \sigma_j + T_p)$ for some σ_j. Now, the idea here is to consider the first response to a pulse as a detected pulse outcome and any other response to that pulse as an extra detection. Formally,

> A detected pulse (D) occurs at time t' iff
> $a(t') = 1$ and $t' \in [\sigma_j, \sigma_j + T_p)$ for some σ_j
> and $a(t'') = 0$ for $t'' \in (\sigma_j, t')$

and

An extra detection (X) occurs at time t' iff

$$a(t') = 1 \quad \text{and} \quad t' \in [\sigma_j, \sigma_j + T_p) \text{ for some } \sigma_j$$
$$\text{and} \quad a(t'') = 1 \text{ for some } t'' \in (\sigma_j, t')$$

Finally, we define the rest outcome and the missed pulse outcome. These outcomes also occur when a pulse is present. Here the idea is to define a miss outcome as the decision not to respond to an undetected pulse at the last remaining opportunity to detect that pulse. Any other decision not to respond when a pulse is present will be a rest outcome. Formally,

A missed pulse (M) occurs at time t' iff

$$a(t') = 0 \quad \text{and} \quad t' = \sigma_j + T_p \text{ for some } \sigma_j$$
$$\text{and} \quad a(t'') = 0 \text{ for all } t'' \in (\sigma_j, \sigma_j + T_p)$$

and

A rest (R) occurs at time t' iff

$$a(t') = 0 \quad \text{and} \quad t' \in (\sigma_j, \sigma_j + T_p) \text{ for some } j$$
or
$$a(t') = 0 \qquad t' = \sigma_j + T_p \text{ and } a(t'') = 1 \text{ for}$$
$$\text{some } t'' \in (\sigma_j, \sigma_j + T_p)$$

The above definitions are illustrated in Fig. 1.10. The outcomes associated with the decision times $t_1, \ldots, t_7$ are shown beneath these times.

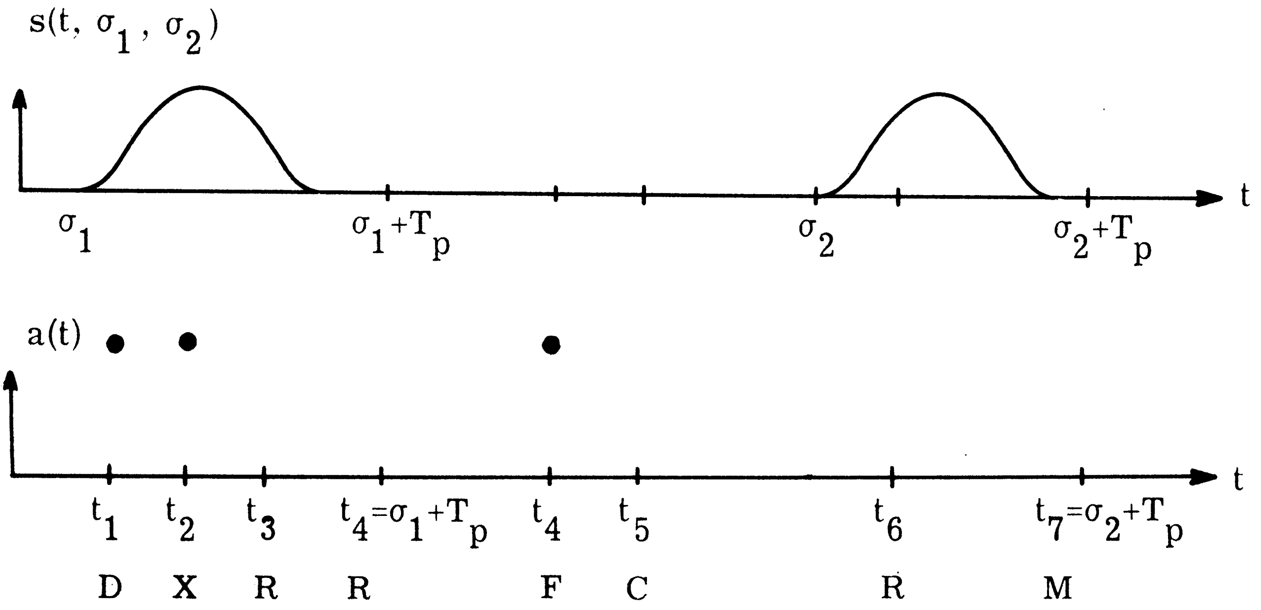


Fig. 1.10. The different possible decision outcomes

It remains to show that the quantities N_D, N_M, N_X and N_F for the respond-once loss function and the quantities N_D, N_M, T_D, T_F, T_C and T_M for the respond-and-hold loss functions can be expressed in terms of the above six outcomes. This is done as follows. First, set the quantities N_D and N_M equal to the number of detected pulse outcomes and the number of missed pulse outcomes respectively. These quantities are always well-defined, since there can only be one detected pulse outcome or one missed pulse outcome for each pulse that occurs, and there can only be a finite number of pulses that occur in the interval $[0, T]$. Next, set N_F and N_X, for the respond-once device, equal to the total number of false alarm outcomes and extra detection outcomes, respectively. These quantities are well-defined, since a false alarm or an extra detection can occur only when $a(t) = 1$ and, by definition, the respond-once response function has only a finite number of responses. Finally, define T_D, T_M, T_C and T_F for the respond-and-hold decision device by:

T_D = the total extra detection time

T_M = the total rest time

T_C = the total correct rejection time

T_F = the total false alarm time

It should be emphasized that T_D, T_M, T_C and T_F as defined here are numerically the same as the T_D, T_M, T_C and T_F of the preceding section.

In the above paragraphs, we have formally defined the basic model for the FRD problem. In Chapter II, we will interpret this model in terms of the general decision theory model and in the remaining chapters we will use this theory to obtain optimum (Bayes) decision rules and to analyze their performance.

1.4 An Intuitive Look at the FRD Decision Device

Since, for the most part, the analysis in this study is of a formal nature, we will pause at this point to analyze the structure of the decision device from an intuitive point of view.

We begin by considering the quantities that determine a particular decision $a(t')$. First of all, note that in general this decision depends on the total past observation $x(t'')$, $t'' < t'$. Thus, it can be expected that the decision device contains a memory for storing this past observation. In Chapter IV, however, we will show that, if the noise satisfies a certain independence condition, then it is sufficient to store a certain density function which can be updated as time progresses.

Next, consider the dependence of the decision $a(t')$ on the past decisions $a(t'')$, $t'' < t'$. This dependence is due to the fact that, if a pulse is present at time t', then the outcome of the decision $a(t')$ depends on whether or not that pulse has been previously detected. For example, a response at time t' to the pulse shown in Fig. 1.11 results in a detected pulse, if $a(t'') = 0$ for $t'' \epsilon (\sigma, t')$, but it results in an extra detection if $a(t'') = 1$ for some $t'' \epsilon (\sigma, t')$. On the other hand, a decision not to respond at time $\sigma + T_p$ results in a missed pulse if $a(t'') = 0$ for $t'' \epsilon (\sigma, \sigma + T_p)$, but it results in a rest if $a(t'') = 1$ for some $t'' \epsilon (\sigma, \sigma + T_p)$.

Since the outcome of the decision $a(t')$ may depend on the previous decision $a(t'')$, $t'' < t'$, it also seems necessary to expect that the general decision device must contain a memory for storing these past decisions. But note from the above arguments that this dependence occurs only if a pulse is present at time t' and then only through the condition that the current pulse has been previously detected. Now we may also express this condition if we know only the time

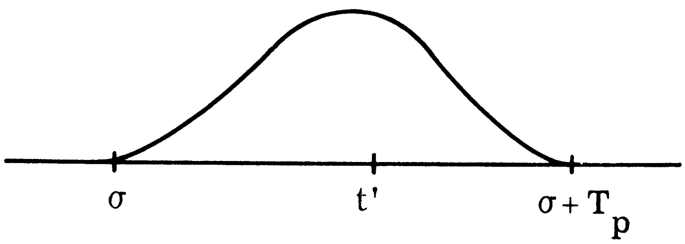


Fig. 1.11. A pulse present at time t'

at which the last response was made. To see that this is so, denote by $\tau^r_{t'}$ the time of the most recent response prior to time t'. Then, from Fig. 1.12, it is seen that, if $\tau^r_{t'} > \sigma$, then the pulse present at time t' has already been detected, but, if $\tau^r_{t'} \leq \sigma$, then that pulse has not been detected. Thus, it seems reasonable to conclude that the general decision device need remember only the time of the most recent response and not the total past decision history. In the following, this conjecture is referred to as the most recent response rule. In Chapter III it will be shown that essentially all decision devices obey the most recent response rule.

We now consider the form of the decision device. In the above paragraphs, it has been argued that the decision device must contain a memory for storing the total past observation and a memory for storing the value of the most recent response function. Now the

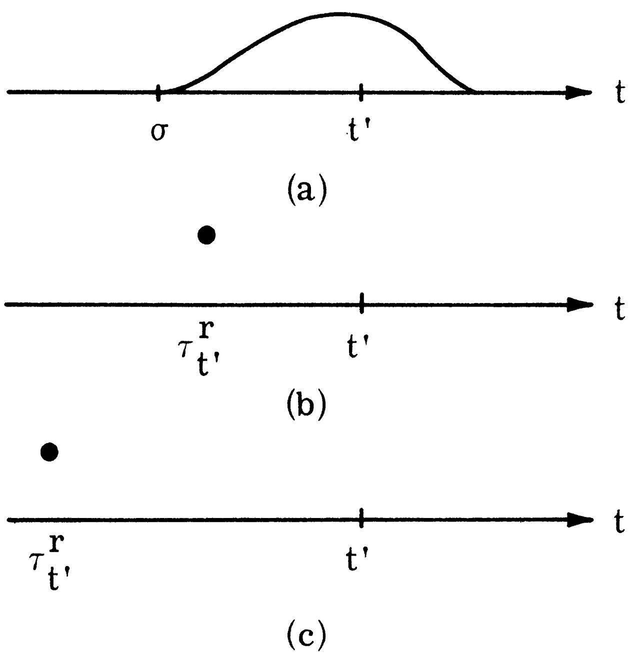


Fig. 1.12. An illustration of the most recent response time
(a) a pulse present at time t'
(b) the case $\tau^r_{t'} > \sigma$
(c) the case $\tau^r_{t'} \leq \sigma$

decision at time t', $a(t')$, is based on some function of these two quantities. Let us denote this function by $T_{t'}(x(t''), t'' < t'; \tau^r_{t'})$. Then without loss of generality, we may assume that the decision $a(t')$ is determined according to

$$a(t') = \begin{cases} 1 & \text{if } T_{t'}(x(t''), t'' < t; \ \tau^r_{t'}) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

It is then possible to characterize the form of the general decision device as shown in Fig. 1.13.

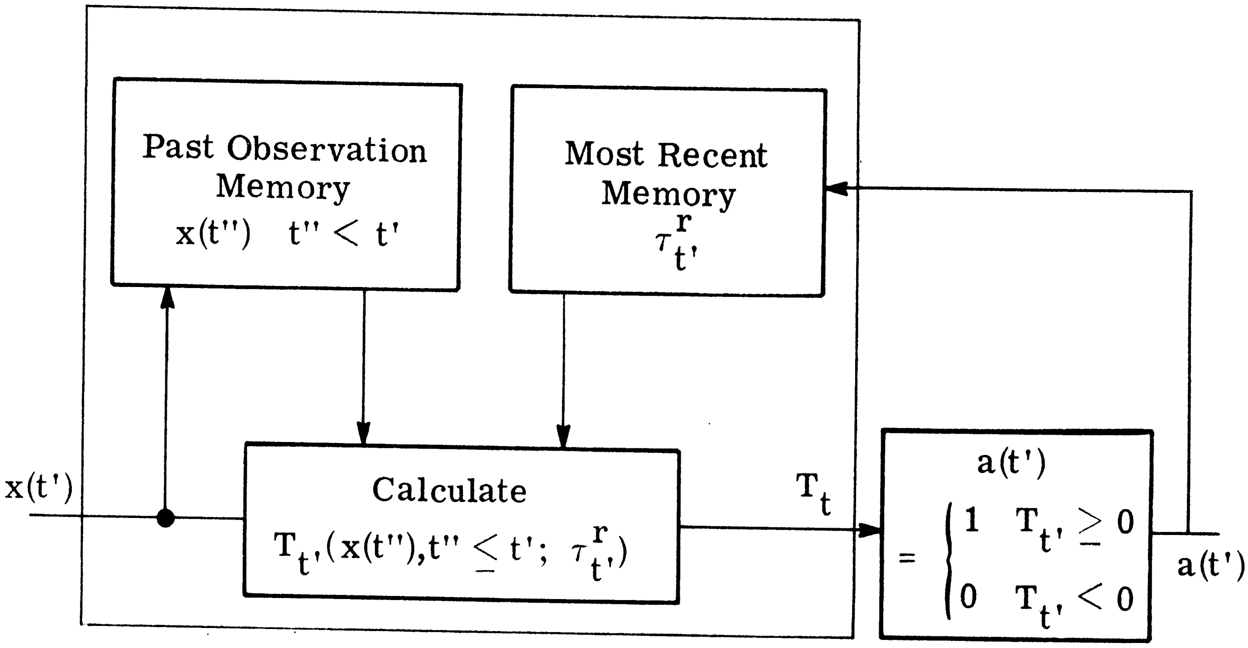


Fig. 1.13. The form of the decision device

The intuitive idea behind Fig. 1.13 is that the decision device first filters or smooths the observation by a time varying filter which also depends on the time of the most recent response. The filter output, $T_{t'}$, is then compared with a threshold of zero to determine the current value of the response function. The operation of this configuration is illustrated in Fig. 1.14 in terms of a typical signal $s(t, \sigma_1, \sigma_2, \ldots)$, a corresponding input $x(t)$, the filter output $T_{t'}$ and the resulting response function $a(t)$ for a respond-and-hold decision device.

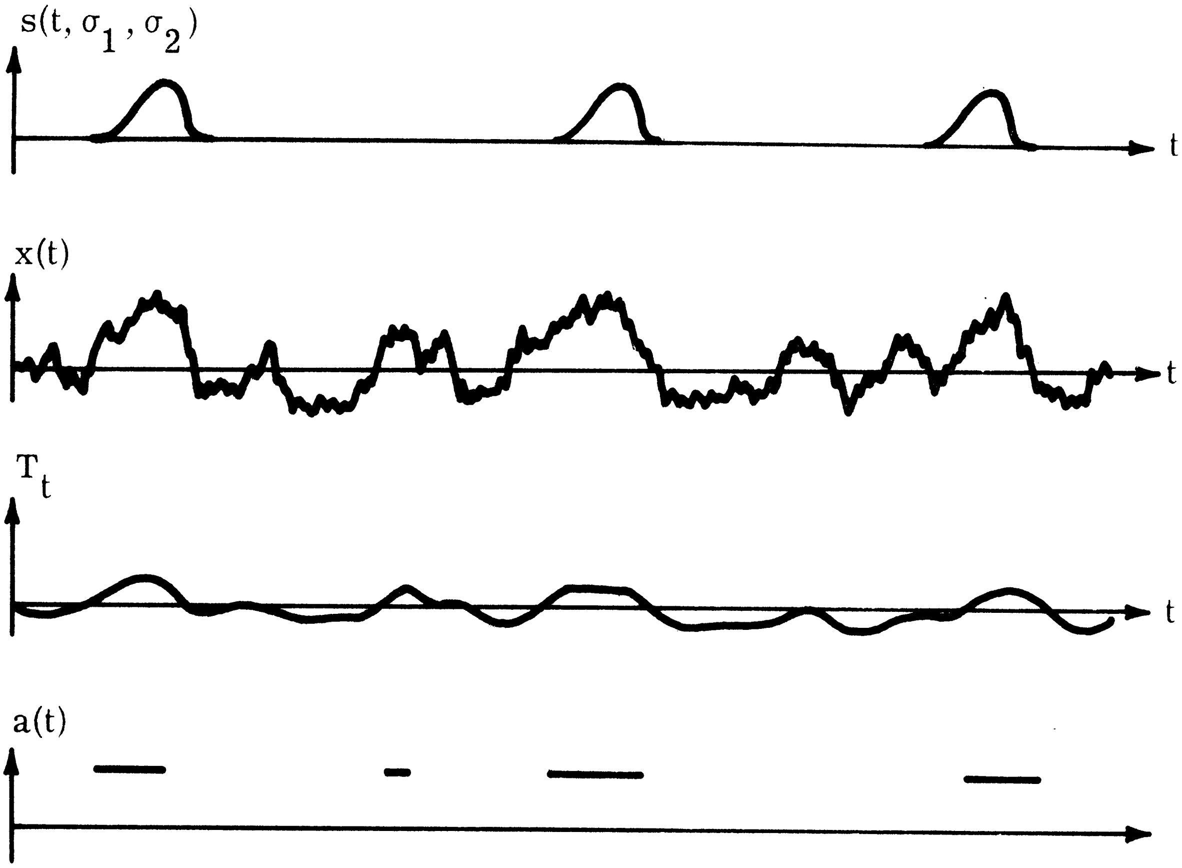


Fig. 1.14. The signals present in a respond-and-hold decision device with the configuration of Fig. 1.13

In the above paragraphs it has been argued that the generic form of the optimum decision device is as illustrated in Fig. 1.13. This argument is based solely on a consideration of the quantities that determine a particular decision $a(t')$. In the next paragraphs we will draw several conjectures on the character of specific decision devices based on a consideration of the loss functions.

Consider first the respond-once device. This device seeks to respond to each pulse that occurs while avoiding false alarms and extra detections. One way to preclude extra detections and yet still have the opportunity to detect each pulse is to avoid making more than one response in any interval of length T_p seconds. (Recall that pulses are separated by at least T_p seconds.) Stated another way, the decision device can refuse to respond at time t' regardless of the observation if the time of the most recent response $\tau_{t'}^r$ is greater than $t' - T_p$. In the following, this principle is referred to as the inhibit rule, since any decision device that obeys the inhibit rule is inhibited from responding for T_p seconds after each response. The reasoning behind the inhibit rule and the fact that large values of the loss L_X reflect a strong desire to avoid extra detections suggest the following conjecture.

<u>Conjecture I.</u> If the loss L_X is sufficiently large in relation to L_F, then the optimum respond-once device satisfies the inhibit rule.

Next, note that if, in fact, a decision device satisfies the inhibit rule, then the number of extraneous detections N_X is equal to zero, so that in this case the loss does not depend on L_X. (See Eq. 1.10.) Thus, it is also reasonable to conjecture that the structure of those decision devices that satisfy the inhibit rule does not depend on L_X. Stated formally,

<u>Conjecture II.</u> For each L_F, there exists a $L_X^0(L_F)$ such that for $L_X \geq L_X^0(L_F)$ the optimum decision device satisfies the inhibit rule and does not otherwise depend on the particular value of L_X.

Finally, consider the respond-once decision devices associated with small values of L_X. These values reflect only a mild desire to avoid extra detections. Thus, it is reasonable to suppose that the optimum decision device for small L_X will not satisfy the inhibit rule if in so doing it can increase the number of pulse detections without increasing the number of false alarms. This reasoning suggests

<u>Conjecture III.</u> For the optimum respond-once decision devices that do not satisfy the inhibit rule there is a trade-off between the average number of detected pulses N_D and the average number of extra detections N_X. Moreover, the maximum average N_D occurs for the MPD decision device $(W_X = 0)$.

We now turn to the respond-and-hold decision devices. First, recall that the purpose of the decision feedback loop in the system diagram of Fig. 1.13 is to help the device distinguish between an initial response to a pulse and a later response to the same pulse. This was seen to be necessary, since the initial response resulted in a detection, whereas the later response results in an extra detection. But, note that the loss function for the MDT decision device does not distinguish between initial detections and extra detections, since this function does not depend on the number of detected pulses N_D. Thus, it can be asserted that, in this case, the decision feedback loop is not necessary. Hence,

Conjecture IV. The optimum MDT decision device does not contain a decision feedback loop.

Finally, consider the respond-and-hold loss function for which $(L_D - L_M) << (L_D^T - L_M^T) < 0$. As mentioned above, this function places a high premium on the number of detected pulses but a low premium on the total amount of detection time. A reasonable procedure for the optimum decision device in this case is to respond only for short periods of time during the center of each pulse as illustrated by the response function $a_1(t)$ in Fig. 1.5. This would increase the likelihood that each pulse is detected but it would reduce the amount of false alarm time that might otherwise result from responding for a longer period of time to each pulse. On the other hand, this would also reduce the total amount of detection but this is of little consequence since the gain for detection time $-(L_D^T - L_M^T)$ is small. Thus, it is reasonable to assume that the effect of increasing the number of detected pulses is to decrease the amount of detection time. That is,

Conjective V. For the respond-and-hold decision device there is a trade-off between the average number of detected pulses and the average detection time with the maximum average detection time occurring for the MDT decision device.

We conclude this section with an example of an R-H decision device and an R-O decision device. It will be demonstrated later that both of these decision devices are sub-optimal by comparing their performances with the performance of the optimum decision device. Nevertheless, these decision devices are simple and practical and, in certain situations, they perform almost as well as the optimum decision devices.

To define these decision devices, it is sufficient to specify their test functions T_t. For the R-H decision device, we define T_t by

$$T_{t'}(x(t''),\ t'' < t';\ \tau_{t'}^r) = C_{t'}(x(t''),\ t'' < t') - \beta$$

where β is some prespecified threshold and

$$C_{t'}(x(t''),\ t'' < t') = \int_{t'-T_p}^{t'} x(t)\, p(t - (t' - T_p))\, dt$$

For the R-O decision device we have essentially the same definition except now we enforce the inhibit rule. This is done by setting $T_{t'}$ equal to $-|\epsilon|$ (ϵ an arbitrary non-zero number) whenever $\tau^r_{t'} > t' - T_p$. The resulting definition is

$$T_{t'}(x(t''),\ t'' < t';\ \tau^r_{t'})$$

$$= \begin{cases} C_{t'}(x(t''),\ t'' < t') - \beta & \tau^r_{t'} \leq t' - T_p \\ -|\epsilon| & \tau^r_{t'} > t' - T_p \end{cases}$$

The engineering significance of these decision devices lies in the fact that the function $C_{t'}$ can be realized as the output of a filter that is matched to the basic pulse shape $p(t)$. That is, $C_{t'}$ can be obtained by filtering the observation $x(t)$ with a linear time-invariant filter with impulse response

$$h(t) = p(T_p - t)$$

Thus, we may realize these systems as shown in Fig. 1.15. In the following discussion, these decision devices are referred to as matched filter (MF) decision devices.

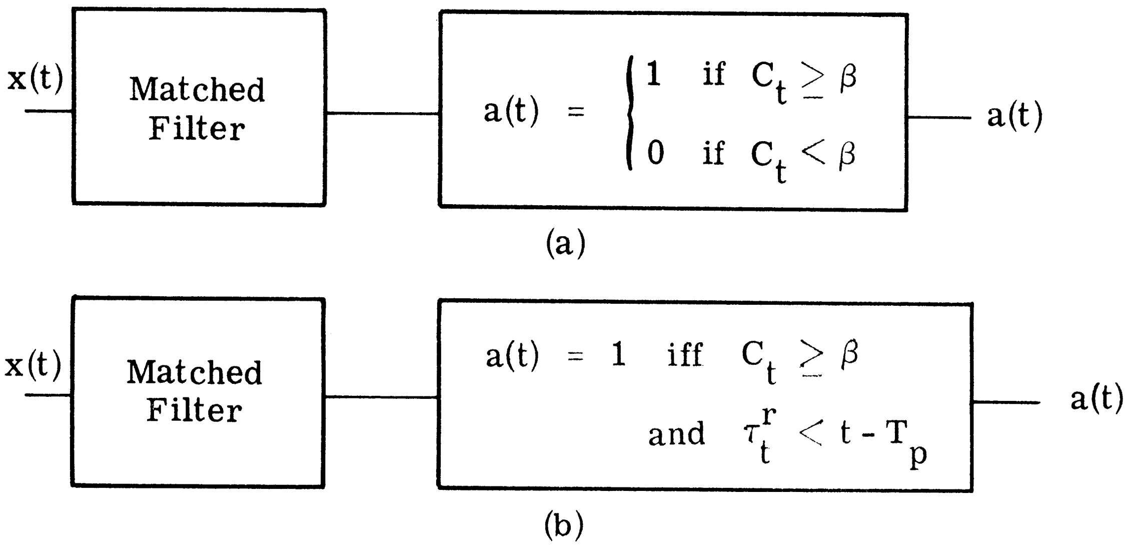


Fig. 1.15. The matched filter decision device
(a) respond-and-hold case, (b) respond-once case

1.5 Historical Remark

To this author's knowledge there is no existing theory for the free running detection problem. This lack of theoretical results is probably due to certain mathematical and conceptual difficulties that arise in carrying out a theoretical analysis. For example, the fact that different decisions are interdependent makes it impossible to determine the optimum decision device by considering a single decision independently of the other decisions.

Nevertheless, in the psycho-physics literature there has been some attention given to the problem that we have characterized here by the respond-once decision device. We shall mention only three of these references. A more complete bibliography is found in Ref. 3.

The first reference is a paper by Egan, Greenberg and Schulman (Ref. 1). This paper serves mainly to define the basic respond-once paradigm and to present techniques for experimentally evaluating the performance of the human decision-maker. The second reference is a paper by Lucas (Ref. 2). In this paper, Lucas introduces essentially the same decision device as the respond-once MF of the preceding section and he calculates certain bounds on the performance of this device. The third reference is a paper by Luce (Ref. 3). This paper assumes that responses occur according to some renewal process and on the basis of this assumption certain probabilities are calculated.

1.6 Organization of This Study

In Chapter I we introduced the free running detection problem and formulated several conjectures on the character of the respond-once and the respond-and-hold decision devices. In this study, we will develop the theory for the FRD problem in three parts: (1) the definition of the basic decision model, (2) the description of the performance for the two classes of decision devices, and (3) the derivation of the optimum (Bayes) decision devices. As a result, we will be able to verify the conjectures of this chapter.

The organization of this study is as follows. Chapter I provides the basic definition of the FRD problem. In Chapter II we proceed formally to obtain the basic decision model for this problem as a special case of the general decision problem. In Chapter III, we develop a procedure for evaluating the performance of different decision devices in terms of the average rates of detections, detections and false alarms. This performance characterization is then used to verify some of the conjectures in Section 1. 4. In Chapter IV we analyze the model of Chapter II to obtain a general recursive description of the optimum decision rules. Then, some sharper results are derived under the assumption that the observations are conditionally independent. We also verify Conjecture IV of Section 1. 4 in this chapter. In Chapter V, we turn to the computer to obtain a description of certain optimum decision rules and to evaluate numerically the performance of these decision rules as well as the performance of the matched filter decision devices. Chapter VI summarizes the results of this study.

CHAPTER II

THE MATHEMATICAL MODEL FOR THE FREE RUNNING DETECTION PROBLEM

The analysis of any problem involving the making of decisions requires precise definitions of many things. First of all, the nature of those entities for which decisions are desired must be known. For example, in the detection of randomly occurring pulses, it must be possible to characterize all of the pulse sequences that may occur. Also, it is necessary to be able to express any prior knowledge regarding these pulses such as the rate at which they are occurring. Secondly, one must know precisely what kind of decisions are to be made. That is, is a typical decision a statement of the form "a pulse is present at the current time" or is it a statement of the form "the pulse present at the current time arrived at time t_0"? Thirdly, one must know precisely what information is available on which to base these decisions and how this information is statistically related to the unknown quantity. Lastly, one must have a criteria by which to judge the "quality" of different decision procedures.

In most decision problems the above quantities can be specified in a fairly straightforward manner without resorting to a formal mathematical structure. But, in contrast to most decision problems, the FRD problem is quite complex. Moreover, it involves the use of certain concepts not usually found in most of the classical decision problems. For this reason, we shall call on the formal theory of decision-making to provide a theoretical basis for the mathematical model. The first section of this chapter reviews the basic elements of decision theory. The remainder of the chapter is devoted to interpreting the FRD problem in terms of this theory.

2.1 Basic Elements of Decision Theory

In this section we introduce the basic elements of decision theory. The definitions introduced here closely parallel the development found in Chapters I and II of Ref. 4.

To begin, consider the triple $(\mathcal{A}, \Theta, L(\cdot, \cdot))$. The first element of this triple, $\mathcal{A}$, is known as the action space. It is the set of all possible actions or decisions available to the decision-maker. The second element, Θ, is the state space. It is the set of all possible states of nature that are of interest to the decision-maker. The third element, $L(\cdot, \cdot)$ is the loss function. This is a real-valued function defined on the product $\mathcal{A} \times \Theta$. These elements have the following interpretation. A state, $\theta \epsilon \Theta$, unknown to the decision-maker, is in effect. The decision-maker chooses an action $a \epsilon \mathcal{A}$ and then incurs a loss given by $L(a, \theta)$.

The objective of decision theory is to provide the decision-maker with some rational rule for choosing an action, $a \epsilon \mathcal{A}$, so as to incur a small loss. To assist in this choice, the decision-maker is allowed to observe some quantity that is statistically related to θ. Let X denote the space of all of these possible observations. The decision-maker then observes an observation $x \in X$ and on the basis of that observation he chooses an $a \in \mathscr{A}$. The rule governing the choice of $a \in \mathscr{A}$ for each $x \in X$ is known as the decision rule employed by the decision-maker. Formally, the decision rule, δ, is a mapping from X to $\mathscr{A}$. The statistical dependence of $x \in X$ on θ is provided by the family of probability measures $\{P(dx| \theta); \theta \in \Theta\}$ which are assumed to be known.

We are now in the position to present the formal definition of a decision model.

> A decision model is a triple $(\mathscr{A}, \Theta, L(\cdot, \cdot))$ coupled with a random experiment which consists of an observation space X and observation probabilities $\{P(dx | \theta); \theta \in \Theta\}$.
>
> A solution to a decision problem is a decision rule $\delta \in \mathscr{D}$ where $\mathscr{D}$ is the set of all functions from X to $\mathscr{A}$ that are measurable with respect to $P(dx | \theta)$ for each $\theta \in \Theta$.

It should be mentioned that, in the above definition of a decision model, the action space $\mathscr{A}$ must be a measurable space so that the definition of δ as a measurable function makes sense.

In a given decision problem, it is desired to obtain a decision rule which satisfies some specific criterion of optimality. One such criterion of optimality is the Bayes criterion. The procedure used to obtain a decision rule that is optimum in the Bayes sense is known as the Bayes procedure and the resulting decision rule is known as the Bayes decision rule. In the bulk of the work which follows, we shall be concerned with Bayes decision rules.

To investigate further the notion of Bayes optimality, it is necessary to introduce two additional entities. The first of these is the prior probability law. The prior probability law, $\Pi(\cdot)$, is a probability measure defined on the state space Θ (Θ must now be considered as a measurable space), with the interpretation that $\Pi(d\theta)$ represents the decision-maker's "subjective feelings", prior to the observation, as to which state of nature θ is in effect.

The second concept needed for the Bayes procedure is the Bayes risk. This is defined as follows. First, recall from above that the loss function $L(\cdot, \cdot)$ is a real-valued mapping on $\mathscr{A} \times \Theta$. The space $\mathscr{A} \times \Theta$ can be interpreted as a measurable space since both $\mathscr{A}$ and Θ are measurable spaces. If we now require that L be a measurable mapping into the real line, then, for each $\delta \in \mathscr{D}$, the mapping $L(\delta(x), \theta)$ is a measurable mapping defined on $X \times \Theta$. With this provision we define the Bayes risk $r(\delta)$ of the decision rule $\delta \in \mathscr{D}$ with respect to the prior probability law $\Pi(\cdot)$ as

$$r(\delta) = E\{L(\delta(x), \theta)\}$$

where the expected value is taken with respect to the joint measure defined by

$$P(dx, d\theta) = P(dx| \theta) \Pi(d\theta)$$

We are now in a position to define the Bayes decision rule δ^0. Specifically, a decision rule δ^0 is said to be Bayes with respect to $\Pi(\cdot)$ iff

$$r(\delta^0) = \inf_{\delta \in \mathscr{D}} r(\delta)$$

The number $r(\delta^0)$ is referred to as the minimum Bayes risk. It should be mentioned in passing that, although, in a given decision problem, a Bayes decision rule may not exist, the minimum Bayes risk always exists.

Before proceeding, it is desirable to introduce some additional notation. In applying the Bayes procedure it is often necessary to take the expected value with respect to different probability laws. Which law is intended can be conveniently denoted by writing underneath the expected value operator the random variables involved. As examples, $\underset{x, \theta}{E}$ denotes the expected value with respect to $P(dx, d\theta)$, $\underset{x|\theta}{E}$ denotes the expected value with respect to $P(dx | \theta)$ and $\underset{\theta}{E}$ denotes the expected value with respect to $\Pi(d\theta)$. We then have the well known identities

$$\underset{x,\theta}{E}\{\cdot\} = \underset{\theta}{E}\left\{\underset{x|\theta}{E}\{\cdot\}\right\}$$

and

$$\underset{x,\theta}{E}\{\cdot\} = \underset{x}{E}\left\{\underset{\theta|x}{E}\{\cdot\}\right\}$$

where of course

$$P(dx) = \int_{\Theta} P(dx, d\theta)$$

and

$$P(d\theta\,|x) = \frac{P(dx\,|\,\theta)\,\Pi(d\theta)}{P(dx)}$$

To provide an example of the above definition and to introduce some techniques and results which are useful later, we have cast the well known classical detection problem in the notation of the general decision model. This is done in Appendix A. The remainder of this chapter is devoted to interpreting the FRD problem in terms of the general decision model.

2.2 The Action Space

The action space is the set of all possible actions available to the decision-maker. In Chapter I it was seen that we can identify these actions with response functions $a(t)$. Thus, in general, we should take the action space for the FRD problem to be the set of all binary valued functions satisfying the definitions of Section 1. 3.

To deal directly with such an action space, however, introduces certain mathematical difficulties in determining the Bayes decision rule. To circumvent these difficulties we will follow a strategy that is common in problems that deal with the continuum. Specifically, the action space will be further restricted to response functions that allow decisions to be made at only a finite number of equally spaced decision times. It is then reasonable to expect that, as the separation between the allowable decision times becomes small, the resulting Bayes solution will closely approximate the solution obtained from an unrestricted action space.

We proceed with the definition of the restricted action space. To begin, define the set of allowable decision times $\mathscr{T}$ by

$$\mathscr{T} = \{t_i \,;\, t_i = i\Delta,\ i=1,\ \ldots,\ N\} \tag{2.1}$$

where the decision time separation Δ and the number of decision times N are related by

$$\Delta = T/N \tag{2.2}$$

Now denote the value of the response function at time t_i (the decision at time t_i) by a_i. That is

$$a(t_i) = a_i$$

Then we define the action space for the respond-once decision device to be those functions a(t) such that

$$a(t) = \begin{cases} 0 & t \notin \mathscr{T} \\ a_i & t \in \mathscr{T} \end{cases} \tag{2.3}$$

and we define the action space for the respond-and-hold decision device to be those functions a(t) such that

$$a(t) = a_i \qquad t \in [t_{i-1}, t_i) \tag{2.4}$$

An example of a typical respond-once response function and a typical respond-and-hold response function appears in Fig. 2.1.

The technical advantage of the above definitions lies in the fact that a response function for either the R-O case or the R-H case may be represented by the binary N-tuple

$$\vec{a} = (a_1, a_2, \ldots, a_N), \qquad a_i \in \{0, 1\}$$

Thus, we may characterize the action space for either case as the set of all such N-tuples,

$$\mathscr{A} = \left\{ \vec{a} = (a_1, \ldots, a_N);\ a_i \in \{0, 1\} \right\}$$

Henceforth, an element $\vec{a} \in \mathscr{A}$ will be referred to as a response vector. The response function a(t) associated with the response vector

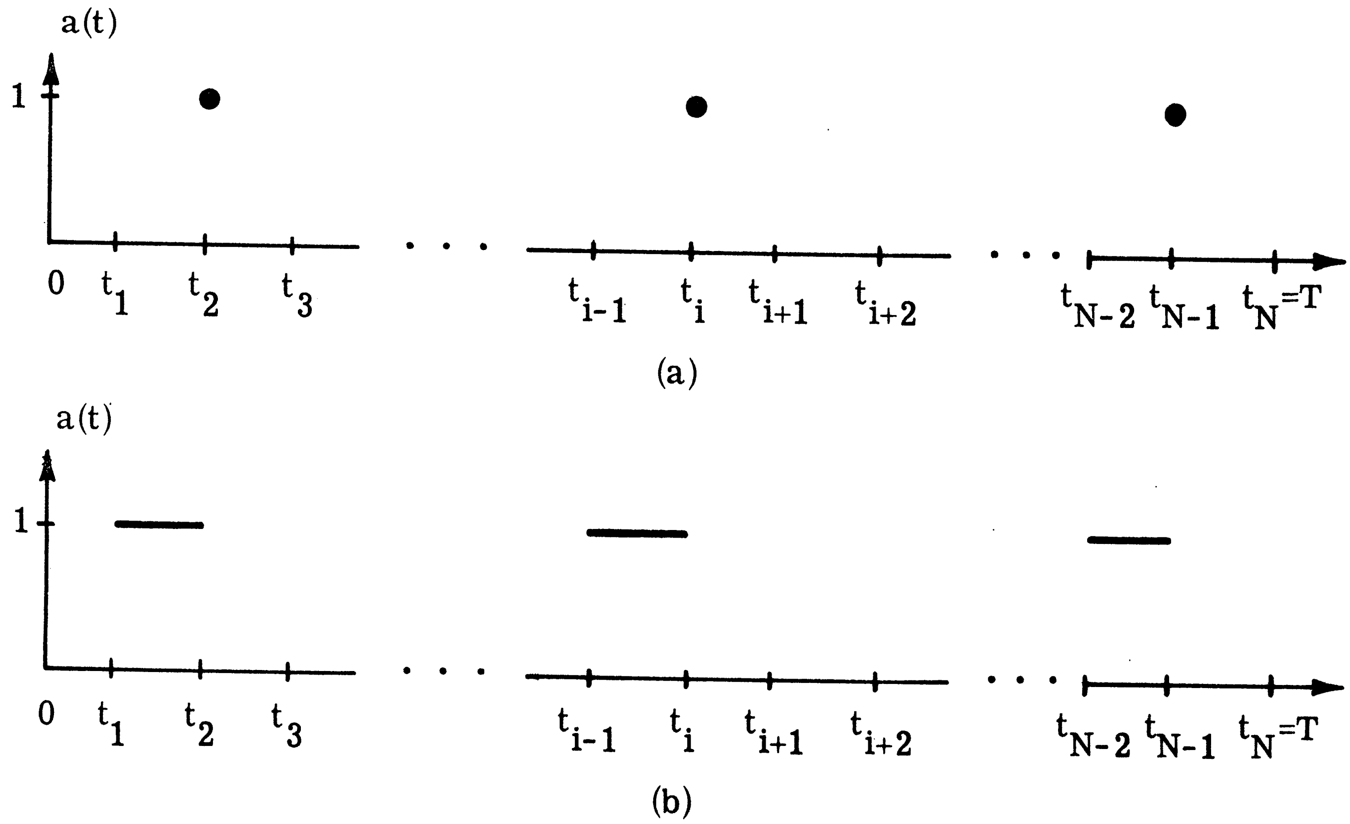
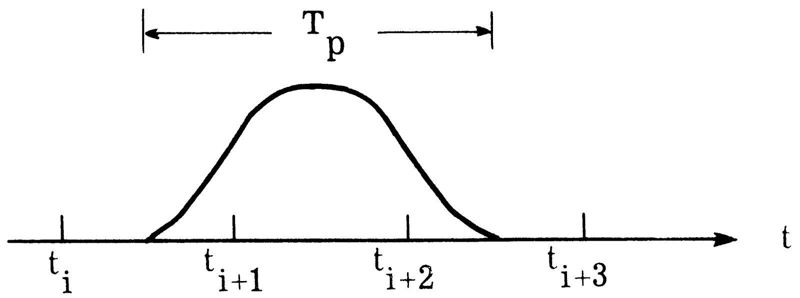


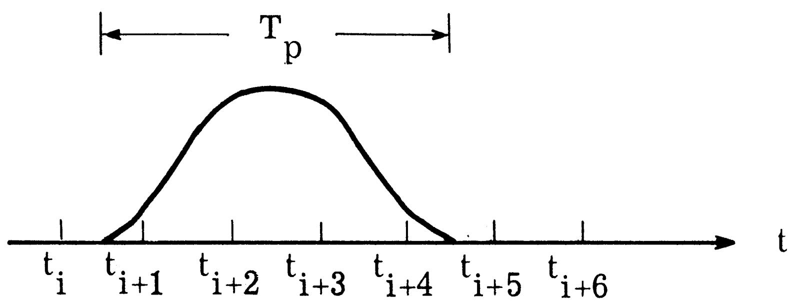
Fig. 2.1. Typical response functions
(a) respond-once response function
(b) respond-and-hold loss function

$\vec{a}$ is obtained from either Eq. 2.3 or Eq. 2.4 depending on the case at hand.

We conclude this section with a discussion of the significance of the decision time separation Δ. First, it should be noted that for a fixed pulse duration, T_p, Δ determines the number of decision opportunities available to detect each pulse. For example, suppose T and T_p are fixed and $\Delta = T_p/2$. Then, any pulse that occurs covers two allowable decision times, so that the decision-device has two opportunities to detect each pulse. On the other hand, if Δ is decreased to $T_p/4$, then each pulse covers four decision times so that now there are four opportunities to detect each pulse. These cases are illustrated in Fig. 2.2.



(a)



(b)

Fig. 2.2. The number of decision opportunities per pulse;
(a) $\Delta = T_p/2$ (b) $\Delta = T_p/4$

In the light of the above comments, it is reasonable to expect that the performance of the decision-device should improve as the number of decision opportunities per pulse increases. Indeed, this is shown to be the case in Chapter IV. Thus, it is not the absolute durations of Δ and T_p that are important but rather the ratio of these quantities since this ratio gives the number of decision opportunities per pulse. To simplify the following analysis we shall assume that Δ is only allowed to take on values such that the number of decision opportunities T_p/Δ is integer valued. We will denote this number by $m = T_p/\Delta$.

2.3 Parameter Space and Prior Probabilities

As discussed in Chapter I, the signal present at the input to the decision-device consists of a train of non-overlapping, randomly occurring pulses. In this section we shall make this notion precise. This is done by means of a probabilistic description of the arrival times of the individual pulses. This description is then used as a basis for defining the parameter set Θ and the a priori probability law on $\Pi(d\theta)$.

To begin we recall from Section 1.3 that a signal is represented as a function

$$s(t, \sigma_1, \ldots, \sigma_{N_p}) = \sum_{i=1}^{N_p} p(t - \sigma_i)$$

This representation can be interpreted as a mapping which identifies signals with sequences of arrival times $\sigma_1, \sigma_2, \ldots, \sigma_{N_p}$. Therefore, it is possible to answer probabilistic questions concerning the presence of pulses, if we have a probabilistic description of the arrival times. In the discussion which follows we will obtain such a description.

The theoretical basis for the probabilistic description of the pulse arrival times is found in the theory of renewal processes. We shall pause briefly to review those aspects of renewal theory that are of interest here. Formally, any sequence of positive, independent, identically distributed random variables, $\{\eta_i\}$, can be used to define a renewal process. The random variables η_i are known as inter-arrival times. Physically, η_i represents the elapsed time between the (i-1)st event and the ith event. The arrival times, σ_j, are related to the inter-arrival times, η_i, by

$$\sigma_j = \sum_{i=1}^{j} \eta_i$$

An important concept in renewal theory is the concept of the arrival rate. Let $N_p(\sigma_1, \ldots)$ be the number of arrivals in the interval $[0, T]$. Then the arrival rate is defined as

$$r_p(T) = \frac{E[N_p]}{T} \tag{2.5}$$

In general $r_p(T)$ depends on T. In fact, the fundamental result of renewal theory, the renewal theorem, states that under suitable conditions, $r_p(T) \to (E[T_i])^{-1}$ as $T \to \infty$. However, in some situations, it is desirable to deal with processes for which $r_p(T)$ does not depend on T. A renewal process with this property can be obtained by allowing the arrival time to the first event, η_1, to have a distribution different than η_i, $i > 1$. The resulting process is then known as a "delayed" renewal process. Specifically, if the cumulative distribution function for η_1, $F_{\eta_1}(\eta)$, is redefined by

$$F_{\eta_1}(\eta) = \frac{1}{E[\eta_2]} \int_0^{\eta} [1 - F_{\eta_2}(u)] \; du \tag{2.6}$$

then the resulting "delayed" renewal process has the rate

$$r_p(T) = \frac{1}{E[\eta_2]} \tag{2.7}$$

A more comprehensive discussion of renewal processes along with precise statements and proofs of the above facts can be found in Chapter XI of Ref. 5. We shall now apply these concepts to obtain a probabilistic model for the pulse arrivals.

The simplest renewal process is the Poisson process or the "No Memory Waiting Time" process. This process is characterized by the fact that the inter-arrival times are exponentially distributed.

It is the Poisson process which is usually used to describe the notion of "randomly occurring events".

The Poisson process shall be modified in two ways to obtain the model for the pulse arrival times. The first and simplest modification will be to incorporate the provision that the pulses be non-overlapping in the basic Poisson process. This is accomplished simply by translating the density of the arrival time T_p units in time. Thus, the density of the inter-arrival times $f_{\eta_i}(\eta)$ shall be defined as

$$f_{\eta_i}(\eta) = \begin{cases} 0 & \eta < T_p \\ \alpha e^{-\alpha(\eta - T_p)} & \eta \geq T_p \end{cases} \tag{2.8}$$

This density is illustrated in Fig. 2.3.

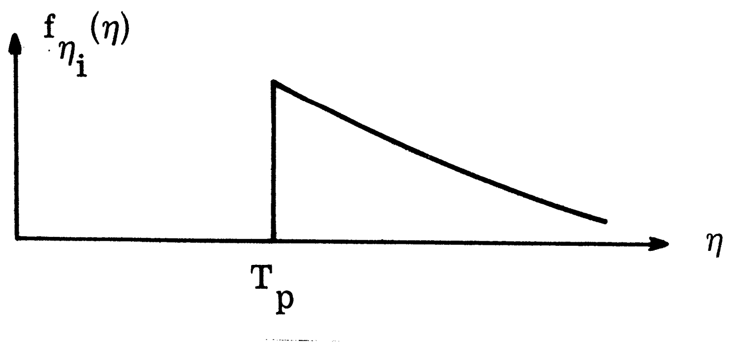


Fig. 2.3. The density of the ith inter-arrival time

The second modification to the Poisson process will be to redefine the density of the first arrival time, $f_{\eta_1}(\eta)$, so as to obtain a delayed process with a constant rate.

This modification is made for two reasons. First, it greatly simplifies the mathematical model to deal with signals whose pulses are arriving at a rate independent of T. The second reason is more subtle. In all ordinary renewal processes, with the single exception of the (unmodified) Poisson process, the time origin takes on a special significance. Physically, this significance can be interpreted in the following way. If the pulses can be considered as being generated by some device, then the ordinary renewal process describes the pulse arrivals as if the generator were turned on at time $t = 0$. When this is the case, the process appears to "evolve" beginning from the time $t = 0$.

The evolution of the ordinary renewal process can be seen in terms of the behavior of the random variable v_t, known as the residual waiting time. Specifically, v_t is the time elapsed from the current time, t, to the next pulse arrival. Now, for an ordinary renewal process, the distribution of v_t depends on the specific time t. It is well known, however (Ref. 5), that as $t \to \infty$ the distribution of v_t approaches a limiting distribution, thus indicating that the process has reached a "steady state". Moreover, it can be shown that the limiting distribution of v_t is precisely the distribution of the random variable η_1 in the delayed renewal process. This fact allows one to interpret the delayed renewal process as a renewal process which began at time $t = -\infty$, or, stated in other terms, as a renewal process which has reached its "steady state". Thus, the inclusion of the second modification assumes that the decision-maker knows only that the pulse generator was turned on in the far distant past and that the pulses are occurring at some constant rate r_p.

In summary, the pulse arrival times will be assumed to be defined by the delayed renewal process with inter-arrival times $\{\eta_i\}$. The density of η_i, $i > 1$, is given by Eq. 2.8. The density of η_1 can readily be determined from Eq. 2.6. The result is

$$f_{\eta_1}(\eta) = \begin{cases} 0 & \eta < 0 \\ (T_p + (\alpha^{-1})) & 0 \leq \eta < T_p \\ (T_p + (\alpha^{-1}))^{-1} \exp\left[-\alpha(\eta - T_p)\right] & T_p \leq \eta \end{cases} \tag{2.9}$$

The pulse rate r_p is determined from Eq. 2.7. Specifically,

$$r_p = (T_p + \alpha^{-1})^{-1} \tag{2.10}$$

Before proceeding to the description of the state space, we shall present a "discretized" version of the above process. Physically, this amounts to restricting the pulse arrivals to occur only at regularly spaced points on the time axis. To be more specific, we divide the time axis into equal intervals of length $\nu = T_p/q$ for some fixed integer q. Then it is assumed that pulses can arrive only at the times $t_i = i\nu$.

To obtain the probability law in the discrete setting, we appeal to the fact that randomly occurring events are described by geometrically distributed inter-arrival times. Thus, in analogy to the continuous case, the probabilities of the inter-arrival times η_i, $i > 1$ will be defined as

$$\Pr[\eta_i = j\nu] = \begin{cases} 0 & j < q \\ a^{j-q}(1-a) & j \geq q \end{cases} \tag{2.11}$$

This probability law is illustrated in Fig. 2.4.

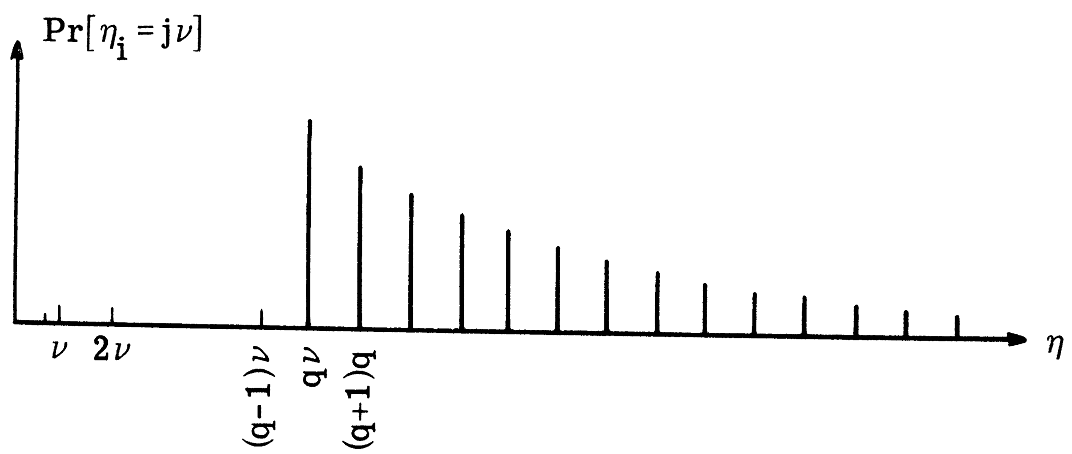


Fig. 2.4. Discrete inter-arrival time probability law

The probability law for the first arrival η_1 will be defined as

$$\Pr[\eta_1 = j\nu] = \begin{cases} 0 & j < 0 \\ [q + a/(1-a)]^{-1} & 1 \leq j < q \\ [q + a/(1-a)]^{-1} a^{j-q+1} & q \leq j \end{cases} \quad (2.12)$$

It is shown in Appendix E that with $\{\eta_i\}$ as defined above the rate r_p is constant and equal to

$$r_p = \Big[[a/(1-a)]\,\nu + T_p\Big]^{-1} \quad (2.13)$$

We turn now to the definition of the state space Θ and the prior probability law $\Pi(d\theta)$.

In the beginning of this section it was shown that the signals s(t) can be identified with the pulse arrival times $\sigma_1, \sigma_2, \ldots$. Therefore, we might take the set of all possible sequences of arrival times as the state space Θ . Then, the prior probability law, $\Pi(d\theta)$, would be defined directly in terms of the delay renewal process. However, this would introduce the following difficulty. At each allowable decision time $t_i \in \mathscr{T}$, the decision-maker is concerned with whether or not an undetected pulse is present at time t_i . The difficulty lies in the fact that the events of this kind are difficult to express directly in terms of the arrival times or inter-arrival times. It would be far more advantageous to have a description of Θ in terms of N-tuples, $\vec{\theta} = (\theta_1, \ldots, \theta_N)$, where θ_i is defined so as to express the prior information relevant to the decision-maker at the time $t_i \epsilon \mathscr{T}$. In the next paragraphs, the state space Θ is constructed in terms of random variables θ_i that have this property.

The random variables θ_i, $i = 1, \ldots, N$ are defined in terms of the pulse arrival times $\{\sigma_j\}$. Specifically, for $i = 1, \ldots, N$, we define

$$\theta_i = \begin{cases} \sigma_j - (t_i - T_p) & \text{if some} \quad \sigma_j \epsilon [t_i - T_p, t_i) \\ T_p & \text{otherwise} \end{cases} \tag{2.14}$$

These random variables are well-defined since we have assumed that $\sigma_{j+1} \geq \sigma_j + T_p$ so that it is not possible for more than one pulse to arrive in the interval $[t_i - T_p, t_i)$. We might also point out that this definition assumes that the signal $s(t; \sigma_1, \ldots, \sigma_{N_p})$ is defined over the interval $[-(T_p - \Delta), T]$.

The above definition of θ_i is interpreted with the help of Fig. 2.5. Here it is seen that, if a pulse arrives in the interval $[t_i - T_p, t_i)$,

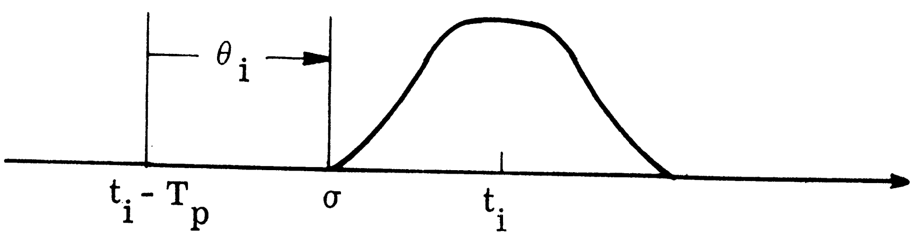


Fig. 2.5. The relation of θ_i to a pulse arriving in the interval $[t_i - T_p, t_i)$

then θ_i gives the time at which that pulse arrives as measured from time $t_i - T_p$. Note that when a pulse is present at time t_i, $\theta_i \epsilon [0, T_p)$. On the other hand, if no pulse is present at time t_i, then $\theta_i = T_p$. Thus, the range of the random variable θ_i is $\Theta_i = [0, T_p]$.

With the above definition at hand we may define the state space Θ as the set,

$$\Theta = \{\vec{\theta} = (\theta_1, \ldots, \theta_N) : \theta_i \epsilon \Theta_i = [0, T_p]\}$$

These definitions are illustrated in Fig. 2.6, which shows a typical pulse sequence and the resulting values of $\vec{\theta} = (\theta_1, \ldots, \theta_N)$.

For the above definition of Θ to be valid, we must be able to represent the signal $s(t, \sigma_1, \sigma_2, \ldots, \sigma_{N_p})$ in terms of $\vec{\theta}$. To show that this is possible and to establish a result for later use we will prove that

$$s(t, \vec{\theta}) = \sum_{i=1}^{N} I_{\{\theta_i \epsilon [0, \Delta)\}}(\theta_i) \quad p\left(t - (\theta_i + t_i - T_p)\right) \qquad (2.15)$$

We proceed as follows. Suppose that $\sigma_1, \sigma_2, \ldots, \sigma_{N_p}$ denote the arrival times of the pulses in $s(t, \sigma_1, \sigma_2, \ldots, \sigma_{N_p})$. Consider the jth pulse, arriving at time σ_j. This pulse is contained in the interval $[\sigma_j, \sigma_j + T_p)$ and no other pulse may also arrive in this interval. Let i^* be the largest integer such that the decision time $t_{i^*} - T_p$ is less

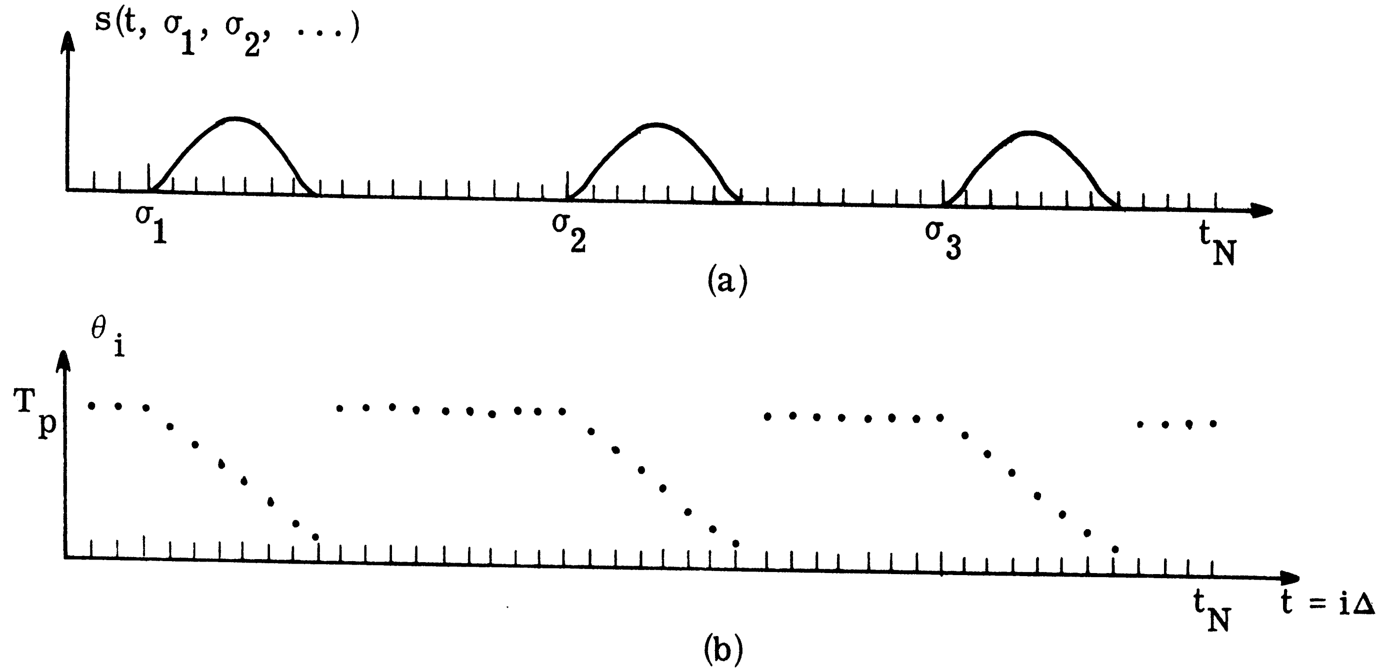


Fig. 2.6. An illustration of $\vec{\theta}$
(a) a typical pulse sequence
(b) the corresponding value of $\vec{\theta} = (\theta_1, \ldots, \theta_N)$

than or equal to σ_j as shown in Fig. 2.7. From this figure it is clear that the interval $[\sigma_j, \sigma_j + T_p)$ contains only the m decision times

$$t_{i^*+m-1}, \ldots, t_{i^*}$$

The random variables associated with these decision times are

$$\theta_{i^*+m-1}, \ldots, \theta_{i^*}$$

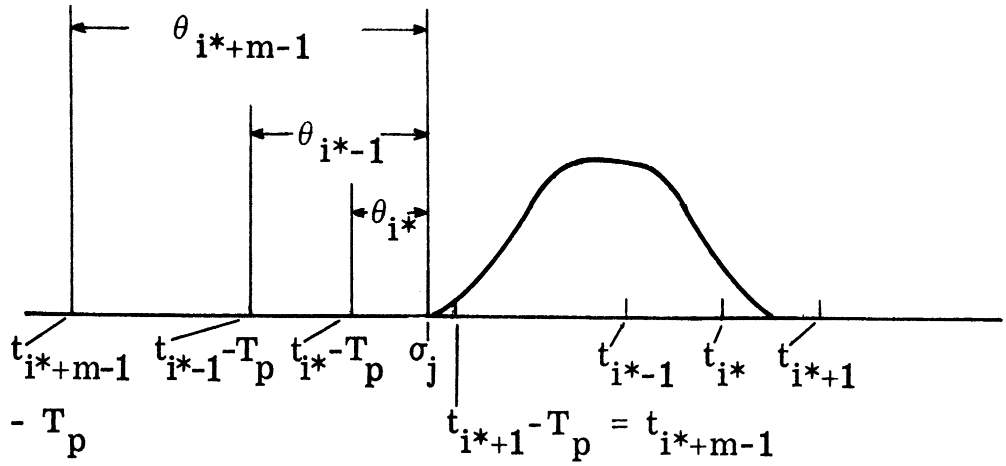


Fig. 2.7. The definition of i*

Also from Fig. 2.7 it is clear that the only one of these random variables that takes on a value in the interval $[0, \Delta)$ is θ_{i^*}. Thus, the only contribution from the jth pulse to the sum in Eq. 2.15 is the i*th term,

$$I_{\{\theta_{i^*} \epsilon [0, \Delta)\}} (\theta_i^*) \quad p\left(t - (\theta_{i^*} + t_{i^*} - T_p)\right)$$

But this term is equal to $p(t - \sigma_j)$ since by definition $\sigma_j = \theta_{i^*} + t_{i^*} - T_p$. Since this reasoning applies for all of the pulse arrivals $\sigma_1, \sigma_2, \ldots, \sigma_{N_p}$, Eq. 2.15 is verified.

The remaining task in this section is to specify the prior probability law $\Pi(d\theta)$. This law is obtained as that induced from the delayed renewal process by the transformation of Eq. 2.14. The computations involved in obtaining $\Pi(d\theta)$ are for the most part tedious and uninteresting and for that reason they are left to Appendices B and D. The main result of these computations is that for both the discrete and continuous pulse arrival time cases, the induced probability law, $\Pi(d\theta)$, is a stationary Markov probability law on the random variables $\theta_1, \ldots, \theta_N$. In Appendix B, $\Pi(d\theta)$ is shown to be Markov. In the continuous case, the conditional cumulative distribution function $F(c_k | c_{k-1}) = Pr[\theta_k < c | \theta_{k-1} = c_{k-1}]$ is given by

$$F(c|c_{k-1}) = \begin{cases} \left\{\begin{array}{ll} 0 & c < c_{k-1} + (T_p - \Delta) \\ 1 - e^{-\alpha(c - (c_{k-1} + T_p - \Delta))} & c_{k-1} + T_p - \Delta \leq c < T_p \\ 1 & T_p \leq c_k \end{array}\right\} & c_{k-1} \epsilon [0, \Delta) \\ \left\{\begin{array}{ll} 0 & c < c_{k-1} - \Delta \\ 1 & c_{k-1} - \Delta \leq \end{array}\right\} & c_{k-1} \epsilon [\Delta, T_p) \\ \left\{\begin{array}{ll} 0 & c < T_p - \Delta \\ 1 - e^{-\alpha(c - (T_p - \Delta))} & T_p - \Delta \leq c < T_p \\ 1 & T_p \leq c \end{array}\right\} & c_{k-1} = T_p \end{cases} \tag{2.16}$$

This function is illustrated in Fig. 2.8.

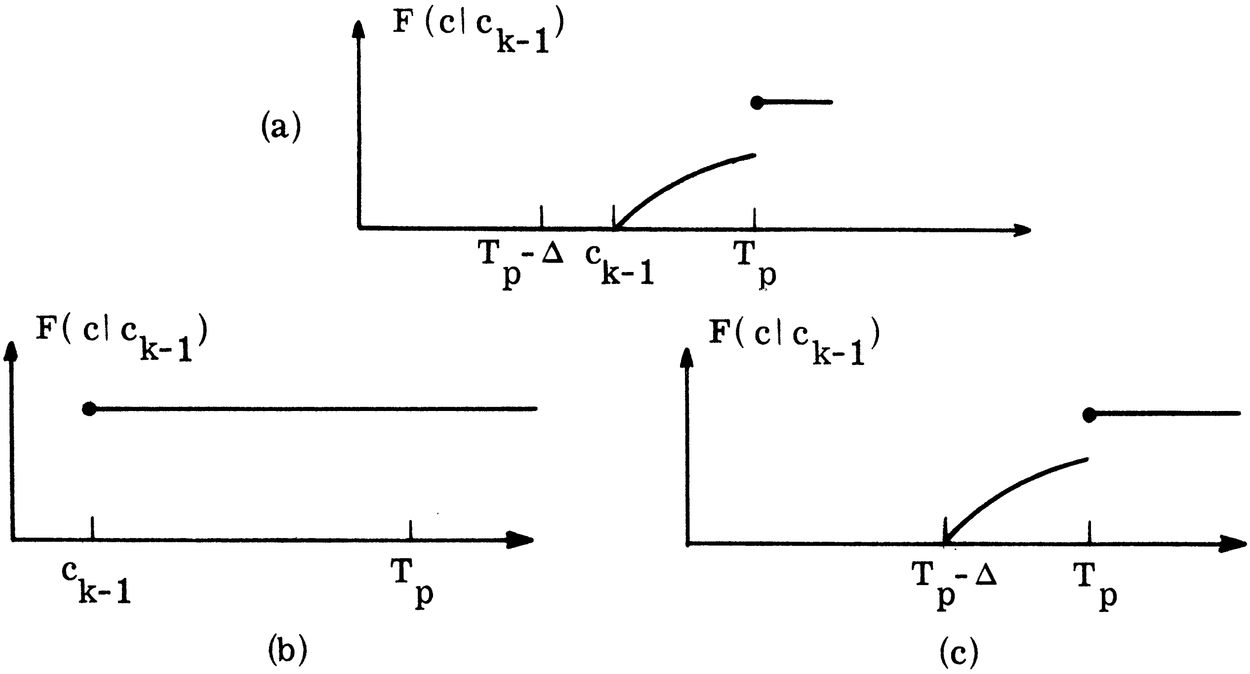


Fig. 2.8. Conditional cumulative distribution function $F(c \mid c_{k-1})$
(a) $c_{k-1} \in [0, \Delta)$, (b) $c_{k-1} \in [\Delta, T_p)$, (c) $c_{k-1} = T_p$

For the discrete case it must be assumed that the decision separation time, Δ, can be divided into an integral number of intervals of length ν. If this number is denoted by $s = \Delta/\nu$, then in the discrete case the conditional probability law is given by

$$\Pr[\theta_k = \ell\nu \mid \theta_{k-1} = j\nu] =$$

$$\begin{cases} \left.\begin{cases} 0 & \ell < j+q-s \\ (1-a)\, a^{\ell-(j+q-s)} & j+q-s \leq \ell < q \\ a^{s-j} & \ell = q \end{cases}\right\} & j \in \{0, \ldots, s-1\} \\ \left.\begin{cases} 0 & \ell \neq j-s \\ 1 & \ell = j-s \end{cases}\right\} & j \in \{s, \ldots, q-1\} \\ \left.\begin{cases} 0 & \ell < q-s \\ (1-a)\, a^{\ell-(q-s)} & q-s \leq \ell < q \\ a^{s} & \ell = q \end{cases}\right\} & j = q \end{cases} \tag{2.17}$$

This function is illustrated in Fig. 2.9.

In Appendix D the Markov law for $\Pi(d\theta)$ is shown to be stationary. In the continuous case, the unconditional cumulative distribution for θ_k, $k \geq 1$ is given by

$$\Pr[\theta_k \leq c] = \begin{cases} c(T_p + (\alpha)^{-1})^{-1} & c \in [0, T_p) \\ 1 & T_p \leq c \end{cases} \tag{2.18}$$

and in the discrete case

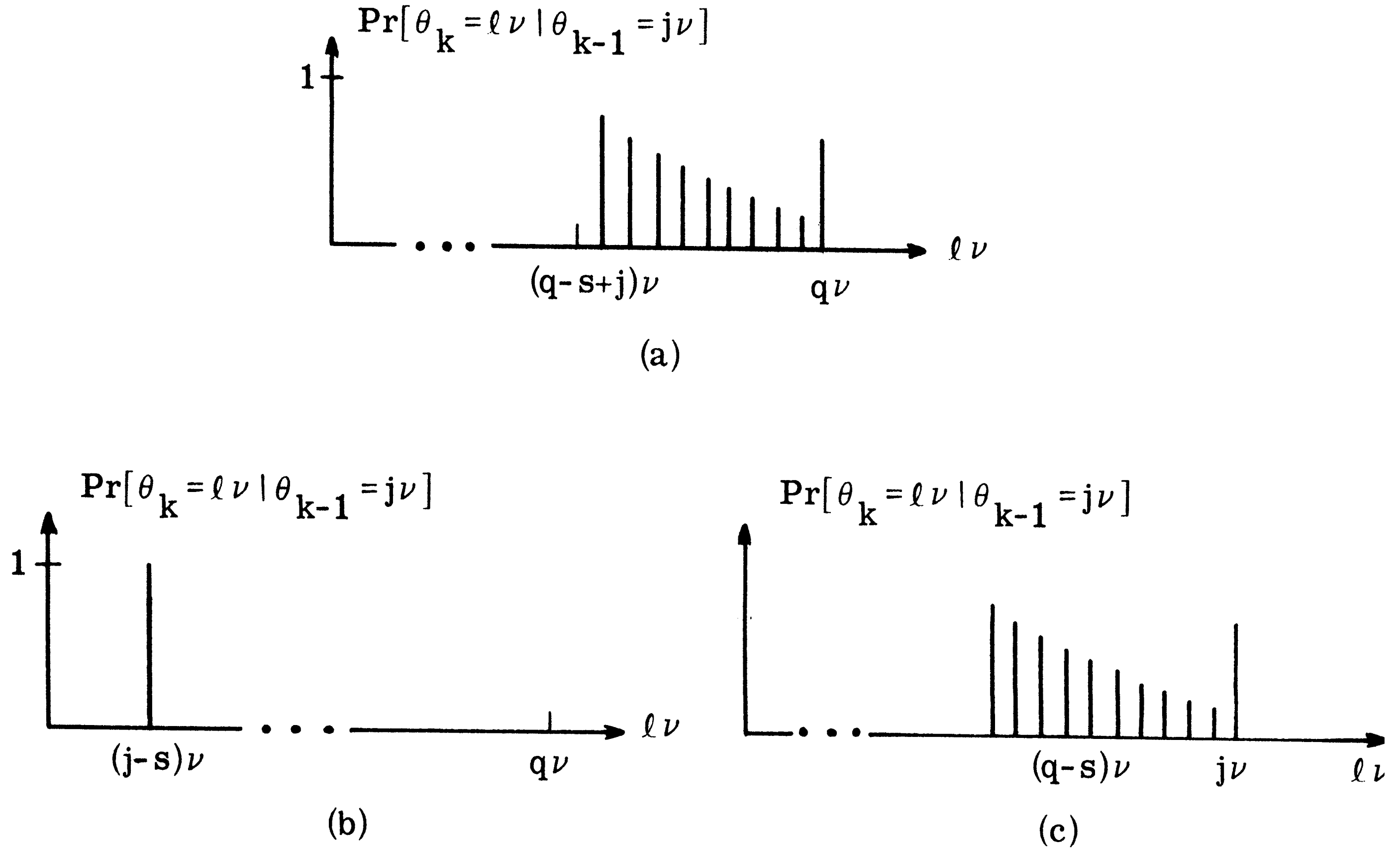


Fig. 2.9. Conditional probability law $P(\theta_k = \ell\nu | \theta_{k-1} = j\nu)$
(a) $j \in \{0, \ldots, s-1\}$, (b) $j \in \{s, \ldots, q-1\}$, (c) $j = q$

$$\Pr[\theta_k = \ell\nu] = \begin{cases} [q + a/(1-a)]^{-1} & \ell \in \{0, \ldots, q-1\} \\ (a/(1-a))\,[q + a/(1-a)]^{-1} & \ell = q \end{cases} \tag{2.19}$$

Both of these probability laws are seen to be uniform for $\theta_k \in [0, T_p)$ with a jump at $\theta_k = T_p$.

It has been shown above that the prior probability law $\Pi(d\theta)$ is expressed in terms of discrete probability laws and cumulative distribution functions. In the later chapters it will be of great value to be able to express both of these probability laws in terms of densities. For this reason we will introduce a concept which is common in mathematical statistics, specifically, the generalized density.

The theoretical basis for the generalized density is the Raydon-Nikodym theorem.

Theorem (Radon-Nikodym). If ω, μ are two measures defined on the measurable space $(Y, \mathscr{B})$ such that ω is absolutely continuous with respect to μ (written $\omega << \mu$), and μ is σ-finite, then there exists a function $p: Y \rightarrow \mathbb{R}$ such that for any measurable set $A \epsilon \mathscr{B}$,

$$\omega(A) = \int_A p(y)\, \mu(dy)$$

The application of this theorem that is of interest here is due to the fact that if ω is a probability measure, then

$$\Pr[A] = \omega(A) = \int_A p(y)\, \mu(dy)$$

so that the function $p(\cdot)$ has the usual property of a probability density. In this setting, the function $p(\cdot)$ is commonly referred to as the generalized density of the probability law ω with respect to the measure μ and is often written as $p(y) = \omega(dy)/\mu(dy)$. A more comprehensive discussion of the generalized density can be found in Ref. 6.

In Appendix C we apply the above ideas to obtain generalized densities for the prior probability law $\Pi(d\theta)$. Definitions of absolute continuity and σ-finite measures are also found there.

2.4 The Observation Space and Real Time Decision Rules

In this section we specify the observation space and the structure of the decision rules for the FRD problem.

The observation space, **X**, is considered first. This space is the set of all possible observations that may occur together with a family of probability measures $\{P(dx|\theta);\ \theta \in \Theta\}$. Now from Chapter I it is recalled that the FRD decision device observes a waveform that consists of a train of pulses immersed in additive noise. Stated more formally, it has been assumed that some noise process, $n = \{n(t);\ t \in [0, T]\}$, defined on some appropriate probability space $(\Omega, \mathscr{F}, P)$, is specified and that the observation x consists of a realization of that process plus some signal $s(t;\theta)$, $\theta \in \Theta$. That is,

$$x(t) = n(t) + s(t, \theta) \qquad \theta \in \Theta \qquad t \in [0, T] \tag{2.20}$$

This equation defines for each $\theta \in \Theta$ a stochastic process $x = \{x(t);\ t \in [0, T]\}$, on $(\Omega, \mathscr{F}, P)$. Now the viewpoint to be adopted here is as follows. For each $\omega \in \Omega$, $x(t, \omega)$ is a real-valued function defined on $[0, T]$, so that the stochastic process x can be viewed as a mapping from the space $(\Omega, \mathscr{F}, P)$ to the set of real-valued functions defined on $[0, T]$, $\mathbb{R}^{[0, T]}$. Moreover, it is possible to define a σ-algebra, $\mathscr{B}$, of subsets of $\mathbb{R}^{[0,T]}$ so that the mapping x can be considered as a random variable taking values in $\mathbb{R}^{[0,T]}$. Thus, for each $\theta \in \Theta$, we may speak of the probability measure $P_\theta(dx)$ as that measure induced by the transformation of Eq. 2.20. We may then take as the observation space X, the space $(\mathbb{R}^{[0,T]}, \mathscr{B})$ and we may define the family of conditional probability measures $\{P(dx|\theta); \theta \in \Theta\}$ by $P(dx|\theta) = P_\theta(dx)$. Further details on this point of view can be found in Chapter 12 of Ref. 6.

The observation space X defined in the above paragraph is the set of all the observations $x(t)$, $t \in [0,T]$. This space shall be referred to as the total observation space and an $x \in X$ shall be referred to as a total observation. Now in the FRD model, decisions are made at each of the decision times $t_i \in \mathscr{T}$. Since each of these decisions is based only on the observation up to time t_i, we must also be able to speak of the observation $x(t)$, $t \in [0, t_i)$. This can be done in the following straightforward way. For a given set of allowable decision times $\mathscr{T} = \{t_i;\ t_i = i\Delta,\ i = 1, \ldots, N)$, consider the intervals $[t_{i-1}, t_i)$. These intervals partition the interval $[0,T)$. For each $i = 1, \ldots, N$, we can consider the stochastic process $x_i = \{x(t);\ t \in [t_{i-1}, t_i)\}$. An observation space X_i can be defined for x_i in precisely the same manner as above. This observation space shall be referred to as the space of current observations and an $x_i \in X$, as a current observation. In addition, we can define the observation space

$$\vec{X}_i = X_1 \times \ldots \times X_i = \prod_{j=1}^{i} X_i$$

This observation space will be referred to as the space of observations up to time t_i and an observation

$$\vec{x}_i = (x_1, \ldots, x_i) \in \vec{X}_i$$

will be referred to as the observation up to time t_i.

The probability distributions for x_i and $\vec{x}_i$ have the obvious definitions in terms of the noise process n and the definitions of $x_i = \{x(t)\, ;\, t \in [t_{i-1}, t_i)\}$ and $\vec{x}_i = \{x(t)\, ;\, t \in [0, t_i)\}$. A few properties of these distributions which will be of use later might be mentioned at this point. First, it is easily verified from Eq. 2.15 that for $t \in [0, t_i)$, $s(t, \theta)$ can be written as

$$s(t, \theta) = \sum_{j=1}^{i-1} I_{\{\theta_j \in [0, \Delta)\}}(\theta_j) \; p\Big(t - (\theta_j + t_j - T_p)\Big)$$

$$+ \; I_{\{\theta_i \in [0, T_p)\}}(\theta_i) \; p\Big(t - (\theta_i + t_i - T_p)\Big)$$

$$t \in [0, t_i)$$

so that for $t \in [0, t_i)$, $s(t, \theta)$ and thus $x(t)$ depend only on $\vec{\theta}_i = (\theta_1, \ldots, \theta_i)$. Therefore, $P(d\vec{x}_i | \theta)$ depends only on $\vec{\theta}_i$. By the same reasoning it can be concluded that for $t \epsilon [t_{i-1}, t_i)$, $s(t, \theta)$ can be written as

$$s(t, \theta) = I_{\{\theta_{i-1} \epsilon [0, T_p)\}}(\theta_i) \; p\Big(t - (\theta_{i-1} + t_{i-1} - T_p)\Big)$$
$$+ I_{\{\theta_i \epsilon [0, \Delta)\}}(\theta_i) \; p\Big(t - (\theta_i + t_i - T_p)\Big)$$
$$t_i \epsilon [t_{i-1}, t_i) \qquad (2.21)$$

Thus, $P(dx_i | \theta) = P(dx_i | \theta_i, \theta_{i-1})$.

The remainder of this section is concerned with the decision rule δ. The decision rule δ is a mapping from the observation space X to the action space $\mathscr{A}$. In the general decision problem, nothing more need be said about the structure of the decision rule. In the FRD model, however, it is assumed that the decisions are made in real time. This assumption imposes a special structure on the decision rule δ.

To investigate this structure, consider a decision time $t_i \epsilon \mathscr{T}$. At this time the decision-maker either responds, $a_i = 1$, or he doesn't respond, $a_i = 0$. But since the decisions are to be made in real time, the decision-maker can base this decision only on the information which has become available up to time t_i. This information is of two kinds, the observation up to time t_i, $\vec{x}_i$, and the knowledge of what responses have already been made, $\vec{a}_i = (a_1, \ldots, a_i)$.

Therefore, the decision at time t_i can be a function only of $\vec{x}_i$ and $\vec{a}_i$. This function will be denoted by δ_i and referred to as the component of the decision function δ at time t_i. We then have $a_i = \delta_i(\vec{x}_i, \vec{a}_i)$. The total decision rule δ is defined as the N-tuple $\delta = (\delta_1, \ldots, \delta_N)$. It should be emphasized in connection with this definition that in order to specify a_i at some time t_i by $a_i = \delta(\vec{x}_i, \vec{a}_i)$, the past decisions $\vec{a}_{i-1}$ must have already been specified.

2.5 The Loss Function

It is recalled from Section 2.1 that the loss function is a mapping of pairs $(a, \theta) \epsilon \mathscr{A} x \Theta$ into the real numbers. The interpretation here is that if the decision-maker chooses the response vector $\vec{a} \epsilon \mathscr{A}$ and if the parameter vector $\vec{\theta} \epsilon \Theta$ is in effect, then a loss $L(\vec{a}, \vec{\theta})$ is incurred. Now, in Section 1.3, two loss functions, a respond-and-hold loss function and a respond-once loss function, were defined. In this section, these loss functions will be expressed in the notation of this chapter. It will be seen that both of these loss functions can be described in terms of a single canonical form.

We begin by expressing the six basic outcomes defined in Section 1.3 in terms of $\vec{a}$ and $\vec{\theta}$. First it is recalled that these outcomes depend on the following conditions: whether or not a response is made at time t_i, whether or not a pulse is present at time t_i, and whether or not a pulse, present at time t_i, is detected prior to time t_i. Now, by definition, a response is made at time t_i if $a_i = 1$, and no response is made at t_i if $a_i = 0$. Furthermore, a pulse is present at time t_i if $\theta_i \in [0, T_p)$, and no pulse is present at time t_i if $\theta_i = T_p$. Thus it remains to find an expression for the condition that a pulse present at time t_i is detected prior to time t_i.

To accomplish this, define the most recent response function τ_i; $i = 1, \ldots, N$ as follows. Suppose that at time $t_i \in \mathscr{T}$ responses have been made at the times $t_{r_1} < t_{r_2} < \ldots < t_{r_\ell} < t_i$. Then define the value of the most recent response function at time t_i by

$$\tau_i = \begin{cases} t_{r_\ell} - (t_i - T_p) & \text{if} \quad t_{r_\ell} \in (t_i - T_p, t_i) \\ 0 & \text{if} \quad t_{r_\ell} \notin (t_i - T_p, t_i) \end{cases}$$

The interpretation of this function is as follows. If no responses occur in the interval $(t_i - T_p, t_i)$ then $\tau_i = 0$. On the other hand, if at least one response occurs in this interval, then τ_i specifies the time of the most recent response as measured from $t_i - T_p$. This case is shown in Fig. 2.10. The past responses are denoted by crosses.

To apply the above definition, note that if a pulse present at time t_i is detected prior to time t_i, then a response must have occurred after that pulse has arrived (see Fig. 1.11). Thus, we must have

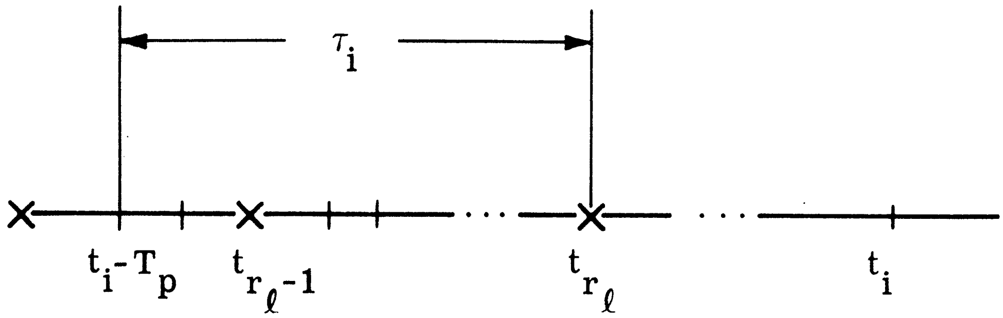


Fig. 2.10. The definition of the most recent response function

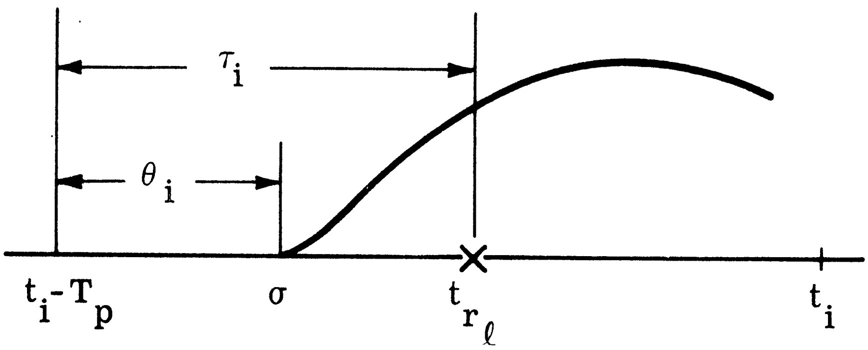


Fig. 2.11. A previously detected pulse

$$t_{r_\ell} > \sigma$$

or

$$t_{r_\ell} - (t_i - T_p) > \sigma - (t_i - T_p)$$

or finally

$$\tau_i > \theta_i$$

Conversely, the condition that a pulse present at time t_i is not detected prior to time t_i is equivalent to the condition,

$$\tau_i \leq \theta_i$$

With these observations at hand we may immediately write down the definitions of a false alarm, a correct rejection, a detected pulse and an extra detection in terms of $\vec{a}$ and $\vec{\theta}$ from the corresponding definitions in Section 1.3.

A false alarm (F) occurs at time t_i iff

$$a_i = 1 \quad \text{and} \quad \theta_i = T_p$$

A correct rejection (C) occurs at time t_i iff

$$a_i = 0 \quad \text{and} \quad \theta_i = T_p$$

A detected pulse (D) occurs at time t_i iff

$$a_i = 1, \quad \theta_i \epsilon [0, T_p) \quad \text{and} \quad \tau_i \leq \theta_i$$

An extra detection (X) occurs at time t_i iff

$$a_i = 1 \quad \theta_i \epsilon [0, T_p) \quad \text{and} \quad \tau_i > \theta_i$$

To carry over the definitions of a missed pulse and a rest, note that the condition that time t_i is the last decision opportunity to detect a pulse present at time t_i is equivalent to the condition $\theta_i \epsilon [0, \Delta)$. (See Fig. 2.12.) Thus we must replace the statement $t_i = \sigma + T_p$ in the definitions of Section 1.3 by the statement $\theta_i \epsilon [0, \Delta)$. The resulting definitions for $t_i < t_N$ are as follows.

A missed pulse (M) occurs at time t_i iff

$$a_i = 0, \quad \tau_i = 0 \quad \text{and} \quad \theta_i \epsilon [0, \Delta)$$

A rest (R) occurs at time t_i iff

$$a_i = 0 \quad \text{and} \quad \theta_i \epsilon [\Delta, T_p)$$

or

$$a_i = 0 \quad \text{and} \quad \tau_i > 0 \quad \text{and} \quad \theta_i \epsilon [0, \Delta)$$

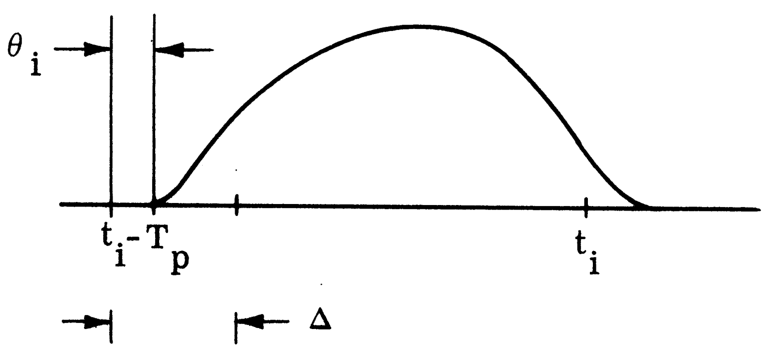


Fig. 2.12. The condition that time t_i is the last decision opportunity to detect a pulse present at time t_i

The terminal time $t_N = T$ requires special consideration since if $a_N = 0$, any pulse present at time t_N that has not been previously detected will be missed.

A missed pulse occurs at time t_N iff

$$a_N = 0 \quad \tau_N < \theta_N \quad \theta_N \epsilon [0, T_p)$$

A correct rejection occurs at time t_N iff

$$a_N = 0 \quad \tau_N \geq \theta_N \quad \theta_N \epsilon [0, T_p)$$

The above definitions of the six basic outcomes are illustrated in Fig. 2.13 below. The outcome associated with a particular decision time t_i is shown beneath that time.

Next it is shown that both the respond-once and the respond-and-hold loss functions defined in Chapter I can be written in terms of a single canonical form. First, consider the respond-once loss function. According to Eq. 1.10, this function is given by

$$\widetilde{L} = (L_D - L_M) N_D + L_X N_X + L_F N_F$$

where it is assumed that

$$L_D < L_M, \quad L_X \geq 0 \quad \text{and} \quad L_F > 0$$

In the remainder of this report it will be convenient to normalize

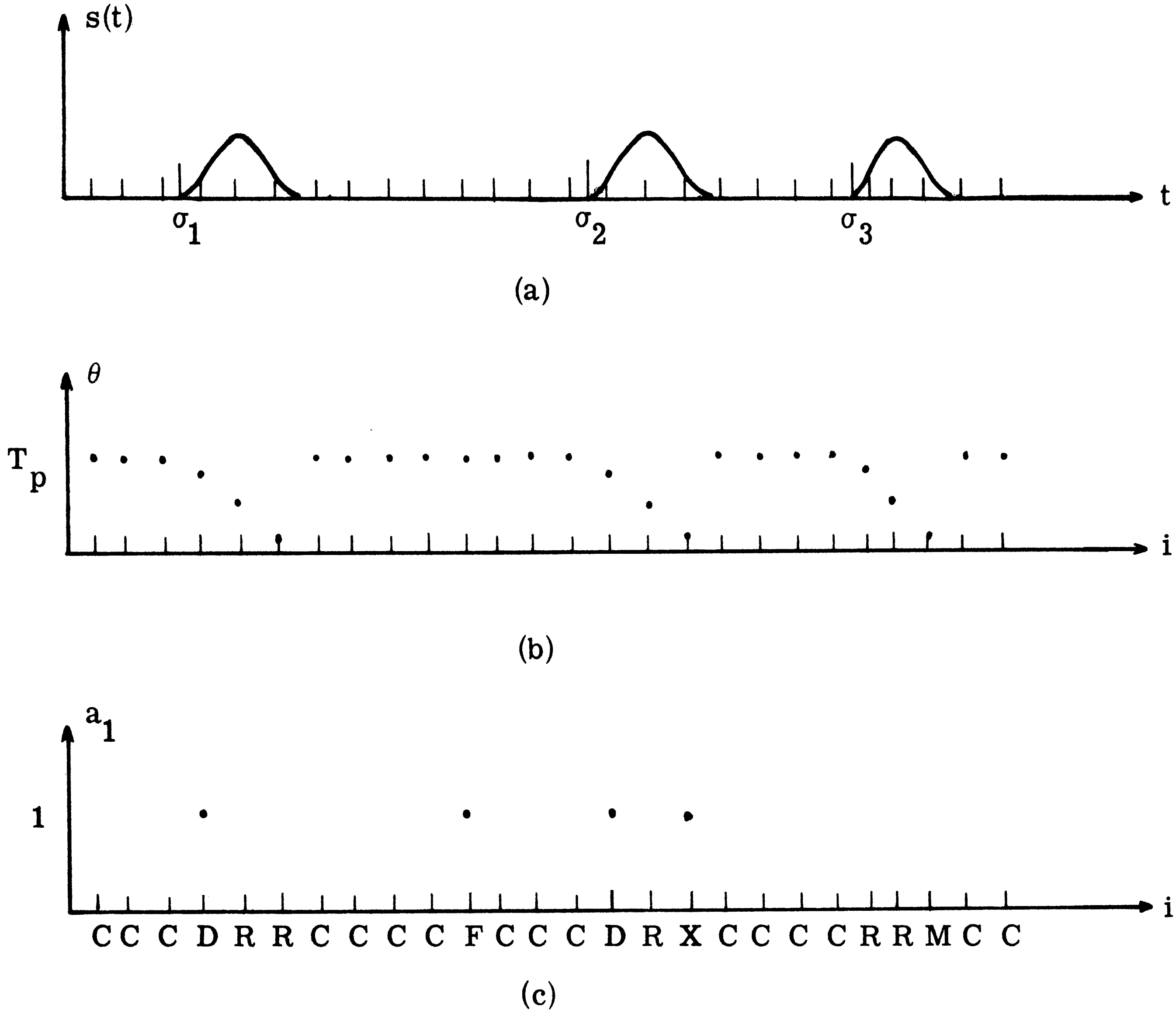


Fig. 2.13. An illustration of the six basic events
(a) a typical pulse sequence,
(b) the corresponding values of θ_i, and
(c) a typical response vector and the corresponding decision outcomes

this loss function by dividing through by $L_M - L_D$. The resulting loss function can then be written as

$$L = W_X N_X + W_F N_F - N_D \tag{2.22a}$$

where

$$W_X = L_X/(L_M - L_D) \geq 0 \tag{2.22b}$$

and

$$W_F = L_F/(L_M - L_D) > 0 \tag{2.22c}$$

Next, consider the respond-and-hold loss function. From Eq. 1.7,

$$\tilde{L} = (L_D - L_M)\, N_D + (L_D^T - L_M^T)\, T_D + (L_F^T - L_C^T)\, T_F$$

where

$$L_D \leq L_M, \quad L_D^T < L_M^T \quad \text{and} \quad L_F^T < L_C^T$$

First, it is necessary to relate T_D and T_F to N_D, N_X and N_F. To do this, note from Fig. 2.14 that, if it is assumed that the pulses can arrive only at discrete times and that these times are also decision times, then

$$T_F = \Delta N_F \tag{2.23a}$$

and

$$T_D = \Delta(N_X + N_D) \qquad (2.23b)$$

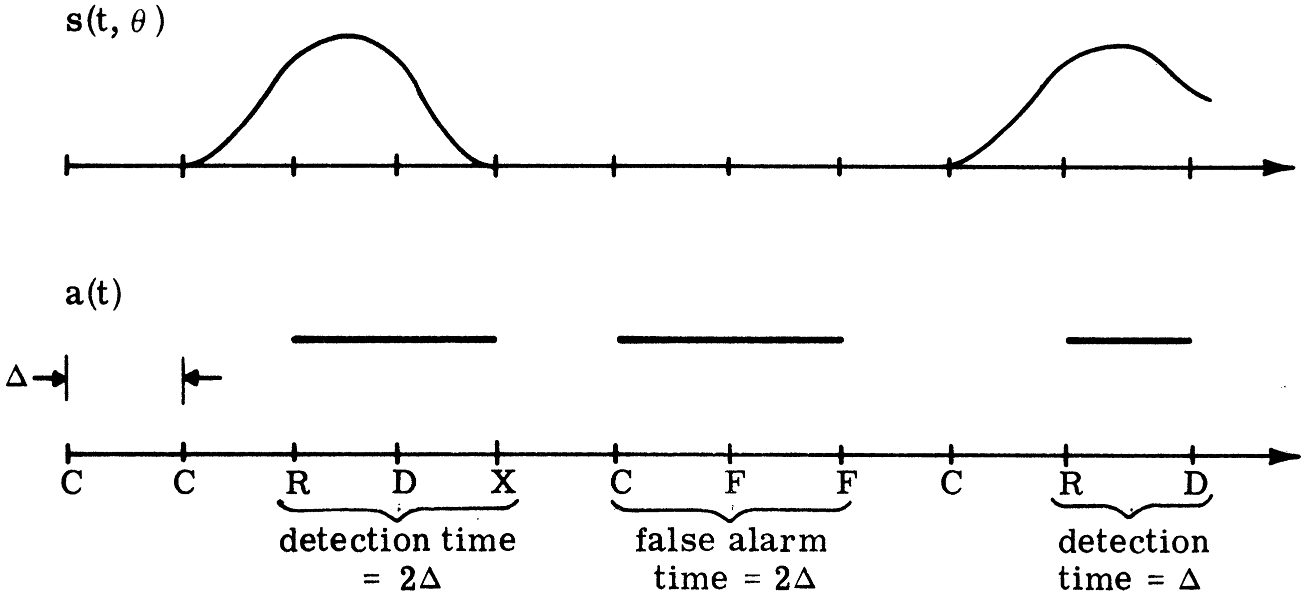


Fig. 2.14. The relation between T_D, T_F and N_X, N_D and N_F

It might also be pointed out that if the pulses are allowed to arrive continuously, then Eq. 2.23 still holds approximately, providing Δ is small. In either case, we will substitute Eq. 2.23 into Eq. 2.22a to write the loss function as

$$\tilde{L} = [(L_D - L_M) + \Delta(L_D^T - L_M^T)]\, N_D + \Delta(L_D^T - L_M^T)\, N_X$$
$$+ \Delta(L_F^T - L_C^T)\, N_F$$

Finally, this function is normalized by dividing through by $L_M - L_D + \Delta(L_M^T - L_D^T)$. The resulting respond-and-hold loss function is

$$L = W_X'' N_X + W_F'' N_F - N_D \tag{2.24a}$$

where

$$W_X'' = (L_D^T - L_M^T) \Big/ [(L_M - L_D) + \Delta(L_M^T - L_D^T)] \tag{2.24b}$$

and

$$W_F'' = (L_F^T - L_C^T) \Big/ [(L_M - L_D) + \Delta(L_M^T - L_D^T)] \tag{2.24c}$$

In the above paragraphs we have shown that both the respond-once and the respond-and-hold loss functions can be written in terms of the single canonical form

$$L = W_X N_X + W_F N_F - N_D \tag{2.25}$$

Thus, for the purpose of analysis we can work with the loss function of Eq. 2.25. When it is necessary to interpret this function in terms of a specific decision device, set W_X and W_F equal to W_X' and W_F' or to W_X'' and W_F'' depending on whether the decision device is a respond-once or a respond-and-hold decision device.

It might also be pointed out here that we may distinguish between the two cases by whether or not $W_X \geq 0$. That is, it is easily concluded from Eqs. 2.22 and 2.24 that the respond-once decision devices are associated with values of $W_X \geq 0$ and the respond-and-hold decision devices are associated with values of W_X satisfying

$$-1 \leq W_X < 0$$

The special case $W_X = 0$ is the maximum pulse detection (MPD) device and the special case $W_X = -1$ is the maximum detection time (MDT) device.

To complete the definition of the loss function it is necessary to find expressions for N_X, N_D and N_F in terms of $\vec{a}$ and $\vec{\theta}$. To this end, we appeal to the definitions of the extra detection outcome, the detected pulse outcome and the false alarm outcome. From these definitions it is clear that

$$N_X(\vec{a}, \vec{\theta}) = \sum_{i=1}^{N} I\,(\tau_i, a_i, \theta_i) \quad \{a_i=1,\ \theta_i < \tau_i,\ \theta_i \in [0, T_p)\} \tag{2.26a}$$

$$N_D(\vec{a}, \vec{\theta}) = \sum_{i=1}^{N} I\,(\tau_i, a_i, \theta_i) \quad \{a_i=1,\ \theta_i \geq \tau_i,\ \theta_i \in [0, T_p)\} \tag{2.26b}$$

and

$$N_F(\vec{a}, \vec{\theta}) = \sum_{i=1}^{N} I_{\{a_i=1,\ \theta_i=T_p\}}(a_i, \theta_i) \qquad (2.26c)$$

where $I_B(\theta)$ denotes the indicator function for the set B. We may then substitute these expressions into Eq. 2.25 to obtain

$$L(\vec{a}, \vec{\theta}) = \sum_{i=1}^{N} L_i(\tau_i, a_i, \theta) \qquad (2.27a)$$

where

$$L_i(\tau_i, a_i, \theta) = W_F\, I_{\{a_i=1,\ \theta_i=T_p\}}(a_i, \theta_i) + W_X\, I_{\{a_i=1,\ \theta_i<\tau_i,\ \theta_i\in[0,T_p)\}}(\tau_i, a_i, \theta_i)$$
$$- I_{\{a_i=1,\ \theta_i\geq\tau_i,\ \theta_i\in[0,T_p)\}}(\tau_i, a_i, \theta_i) \qquad (2.27b)$$

This completes the derivation of the loss function for the FRD problem.

CHAPTER III

THE PERFORMANCE OF FRD DECISION DEVICES

In Chapter II we introduced the mathematical model for the FRD problem. In Section 2.1 of that chapter the Bayes procedure was introduced as the means of obtaining optimum decision rules. This procedure is useful primarily because it has the advantage of being applicable in any decision problem where it is possible to characterize the losses and the prior knowledge.

The Bayes risk $r(\delta)$ can also be used to compare the quality of suboptimum rules. Specifically, if δ' and δ'' are two decision rules, then δ' is said to be "better than" δ'' if the Bayes risk $r(\delta')$ is less than the Bayes risk $r(\delta'')$. This means of comparing performance, however, has one major drawback. Namely, it loses meaning if the losses and the prior probabilities are not held constant. For example, in classical detection theory it is possible to find two decision rules δ' and δ'', where δ' is the optimum Bayes rule for a particular set of losses and prior probabilities, δ'' is suboptimum for a different set of losses and prior probabilities, but yet $r(\delta'') < r(\delta')$.

What is needed is a means of comparing the performance of different decision rules which is independent of the specific losses and prior knowledge. In classical detection theory, this is done in terms of performance sets and ROC curves. In this chapter we will develop an analogous characterization for the FRD problem.

3.1 A Review of Performance in Classical Detection Theory

In classical detection theory, two numbers P_D and P_F, the probability of detection and the probability of false alarm, are calculated for each decision rule. These numbers are plotted as a point (P_F, P_D) in the unit square. The locus of all points corresponding to all possible decision rules is known as the performance set. Through the notion of the randomized decision rule it can be shown that the performance set is convex. Moreover, it contains the points $(0,0)$ and $(1,1)$, the former corresponding to the receiver that never responds and the latter corresponding to the receiver that always responds. Since this set is convex, it also contains the line segment connecting these points. This line segment is usually referred to as the chance diagonal. An additional property of the performance set is that it is symmetric about the point $(1/2, 1/2)$. That is, if (P_F, P_D) is a point in the set corresponding to a particular decision rule, then there is a decision rule, known as the antipodal decision rule, whose performance plots as the point $(1 - P_F, 1 - P_D)$. A typical performance sst is shown in Fig. 3.1.

A criterion of performance can be defined in terms of the performance set as follows. Specifically, a decision rule corresponding

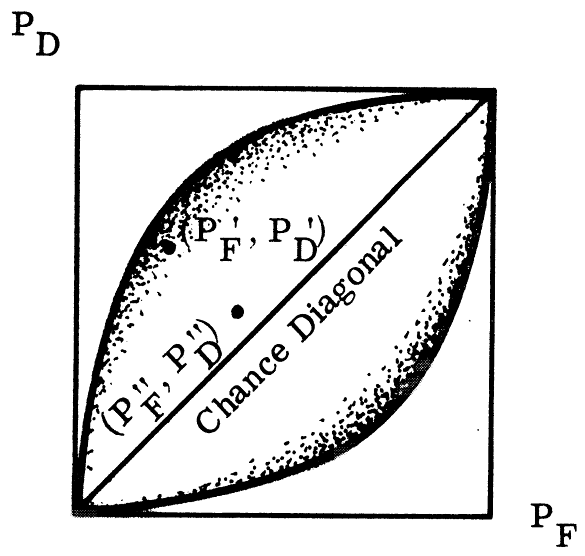


Fig. 3.1. A typical performance set for the classical detection problem

to the point (P'_F, P'_D) is said to be "better than" a decision rule corresponding to the point (P''_F, P''_D) if the point (P'_F, P'_D) lies to the left and above the point (P''_F, P''_D). That is, if $P'_F \leq P''_F$ and $P'_D > P''_D$. It is quickly recognized that the "best" decision rules are those corresponding to points on the upper boundary of the performance set. This upper boundary is referred to as the optimum receiver operating characteristic (ROC). One of the main results of classical detection theory is that the ROC is actually contained in the performance set, so that, in fact, the best decision rules do exist. A more comprehensive discussion of these concepts can be found in Ref. 8 .

In classical detection theory, the Bayes procedure can be interpreted in terms of the performance set in the following manner.

Recall from Appendix A that the Bayes risk of a decision rule δ can be written as

$$r(\delta) = Pr[SN](L_D - L_M) P_D + Pr[N](L_F - L_C) P_F$$

$$+ (Pr[SN] L_M + Pr[N] L_C)$$

or

$$P_D = \frac{W_F}{L_0} P_F - [r(\delta) - (W_M + W_C/L_0)] \qquad (3.1)$$

where

$$W_F = (L_F - L_C)/(L_M - L_D)$$

$$W_M = L_M/(L_M - L_D), \quad W_C = L_C/(L_M - L_D)$$

and

$$L_0 = Pr[SN]/Pr[N]$$

In the unit square this equation plots as a line with slope W_F/L_0 and intercept

$$c(\delta) = -\left[\frac{r(\delta) - (L_M + L_C/L_0)}{(L_M - L_C)}\right]$$

The Bayes decision rule δ_0 is then characterized by the point

(P_F^0, P_D^0) that supports the Bayes risk line of Eq. 3.1. This is illustrated in Fig. 3.2.

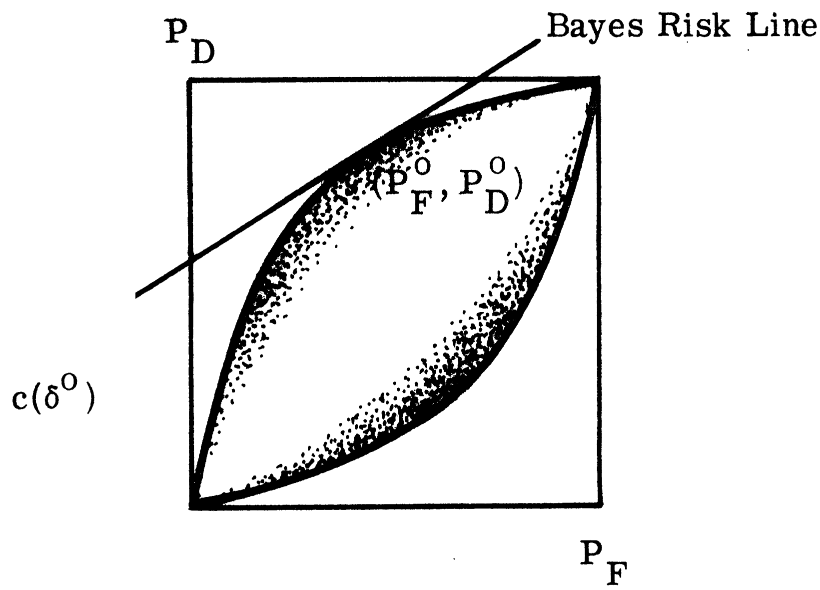


Fig. 3.2. The geometrical relation between the Bayes rule and the performance set

3.2 Performance Sets for the FRD Decision Rules

In the preceding section, it was seen that, in classical detection theory, the performance set is defined in terms of two probabilities, P_D and P_F. Since FRD decision devices also make detections, false alarms and in addition extra detections, it might seem reasonable to define the performance set in terms of three probabilities, P_D, P_F and P_X. The difficulty here, however, is that these probabilities now depend on the decision times t_i. This is easily seen to be the case by considering the decision device that does not respond at the decision times t_i, for i even. For this device, P_D, P_F and P_X must be zero at these times, since a detection, a false alarm or an extra detection can occur only if the device responds. Thus, in general, we must associate sequences of probabilities to the FRD decision rules.

The difficulty in dealing directly with sequences of probabilities can be avoided by dealing instead with the rates at which the pertinent events occur. This approach will be followed here. In this section, we first define the rates of the various events. Next, we investigate the relations that exist between these rates. Finally, we define two performance sets, one for R-O decision devices and one for R-H decision devices, in terms of certain "normalized rates."

To begin, we define the rate of the event β, where β denotes one of the six events: false alarm, detection, extra detection, rest, correct rest, or miss. Recall from Section 2.5 that $N_\beta(\delta(x), \theta)$ denotes the number of times the event β occurs when the observation is x, the decision rule δ is used, and the pulses are described by θ. Now, the rate of the event β, denoted by $r_\beta(\delta)$, is the average number of times the event β occurs per unit time. Thus we have the definition

$$r_\beta(\delta) = \frac{\underset{x,\theta}{E}\,[N_\beta(\delta(x), \theta)]}{T} \qquad (3.2)$$

An alternate form for this definition can be obtained by substituting the expressions for $N_\beta(\delta(x), \theta)$ given in Eq. 2.26 in Section 2.5 into Eq. 3.2. Then,

$$r_F(\delta) = \frac{\underset{x,\theta}{E}\left[\sum_{j=1}^{N} I_{\{a_j=1,\,\theta_j=T_p\}}(a_j, \theta_j)\right]}{T}$$

$$= \frac{\sum_{j=1}^{N} \underset{x,\theta}{E}\left[I_{\{a_j=1,\,\theta_j=T_p\}}(a_j, \theta_j)\right]}{T}$$

or

$$r_F(\delta) = \sum_{j=1}^{N} \frac{\Pr[a_j=1, \theta_j=T_p]}{T} \tag{3.3a}$$

Similarly,

$$r_D(\delta) = \sum_{j=1}^{N} \frac{\Pr[a_j=1, \tau_j \leq \theta_j;\ \theta_j \epsilon [0, T_p)]}{T} \tag{3.3b}$$

and

$$r_X(\delta) = \sum_{j=1}^{N} \frac{\Pr[a_j=1, \tau_j > \theta_j;\ \theta_j \epsilon [0, T_p)]}{T} \tag{3.3c}$$

Thus, in general, $r_\beta(\delta)$ can be expressed as

$$r_\beta(\delta) = \frac{\sum_{j=1}^{N} P_j^\beta}{T} \tag{3.4}$$

where

$$P_j^\beta = \Pr[\text{event } \beta \text{ occurs at time } t_j] \; .$$

It should be emphasized that P_j^β is not the same as the corresponding P_β in classical detection theory. For example, P_j^D is the probability that a detection occurs at time t_j whereas P_D is the probability that a detection occurs given that signal is present.

Next, we investigate the relation that exists between the different rates. To this end, note from Fig. 2.13 that at any decision time that is not covered by a pulse either a false alarm or a correct rejection occurs. Thus, if $N_N(\theta)$ is the number of decision times in the interval $[0,T]$ that are not covered by a pulse, then

$$N_F(\delta(x),\theta) + N_C(\delta(x),\theta) = N_N(\theta) \tag{3.5}$$

Furthermore, at any decision time that is covered by a pulse, either a detection, an extraneous detection, a rest or a miss occurs. Thus, if $N_S(\theta)$ is the number of decision times in the interval $[0,T]$ that are covered by a pulse, then

$$N_D(\delta(x),\theta) + N_X(\delta(x),\theta) + N_R(\delta(x),\theta) + N_M(\delta(x),\theta) = N_S(\theta) \tag{3.6}$$

Now, recall from Eq. 1.5 that

$$N_D(\delta(x), \theta) + N_M(\delta(x), \theta) = N_p(\delta) \qquad (3.7)$$

If Eq. 3.7 is substituted into Eq. 3.6 the result is

$$N_X(\delta(x), \theta) + N_R(\delta(x), \theta) = N_S(\theta) - N_p(\theta) \qquad (3.8)$$

Finally, divide Eqs. 3.5 and 3.6, and 3.8 by T and take expected values to obtain

$$r_F + r_C = r_N \qquad (3.9a)$$

$$r_X + r_R = r_S - r_p \qquad (3.9b)$$

$$r_D + r_M = r_p \qquad (3.9c)$$

The significance of Eq. 3.9 is as follows. Each decision device δ is completely characterized by the six rates $r_F(\delta)$, $r_C(\delta)$, $r_D(\delta)$, $r_M(\delta)$, $r_X(\delta)$, $r_R(\delta)$. Now if r_p, r_N and r_S are fixed, then the right hand sides of the above equations are constant. Thus, the performance of each receiver can be completely specified by the three rates $r_D(\delta)$, $r_F(\delta)$ and $r_X(\delta)$.

The dependency of Eq. 3.9 on the numbers r_p, r_S, and r_N can be removed in the following way. First, note from Appendix E that

$$r_p = \frac{Pr[\theta_i \epsilon [0, T_p)]}{T_p} \tag{3.10a}$$

$$r_N = m \frac{Pr[\theta_i = T_p]}{T_p} \tag{3.10b}$$

$$r_S = m \frac{Pr[\theta_i \epsilon [0, T_p)]}{T_p} = m\, r_p \tag{3.10c}$$

Here $Pr[\theta_i, \epsilon [0, T_p)]$ and $Pr[\theta_i = T_p]$ are the prior probabilities of the events "a pulse arrives in an interval of length T_p" and "no pulse arrives in an interval of length T_p", respectively. Now make the following definitions:

$$R_D = r_D/r_p \qquad R_X = r_X/r_p$$

$$R_M = r_M/r_p \qquad R_R = r_R/r_p$$

$$R_F = r_F/(r_N/m) = r_F/[Pr[\theta = T_p]/T_p]$$

$$R_C = r_C/(r_N/m) = r_C/[Pr[\theta = T_p]/T_p] \tag{3.11}$$

Finally, substitute Eq. 3.11 into Eq. 3.9. The result is

$$R_D + R_M = 1 \tag{3 12a}$$

$$R_X + R_R = m-1 \tag{3.12b}$$

$$R_F + R_C = m \tag{3.12c}$$

These equations depend only on the number of decision opportunities per pulse, m. The quantities R_β in Eq. 3.11 will be referred to hereafter as the normalized rates of the events β or, if it is clear from the context, as simply the rate of the event β. These normalized rates can be interpreted as follows. The rates r_D, r_X, r_M and r_R are rates of events that occur when a signal pulse is present. Thus, R_D, R_X, R_M and R_R are the results of normalizing these rates by dividing through by $r_p = Pr[\theta \in [0, T_p)]/T_p$. On the other hand, the rates r_F and r_C are rates of events that occur when no pulse is present. Thus, R_F and R_C are the results of normalizing r_F and r_C by dividing through by $r_N/m = Pr[\theta_i = T_p]/T_p$.

$$\lim_{N\to\infty} r_\beta = \lim_{N\to\infty} \left[\sum_{i=1}^{N} \frac{P_i^\beta}{N\Delta} \right]$$

We now define the performance set for the respond-once decision devices. As in classical detection theory, we should like the performance set to characterize the performance of all of the possible R - O decision devices. From Eq. 3.12 it is seen that this can be done by specifying for each device δ the three rates, $R_X(\delta)$, $R_F(\delta)$, and $R_D(\delta)$. Thus, we shall make the following formal definition.

The performance set $\mathscr{S}(m)$ is the subset of $\mathbb{R}^3$ defined by

$$\mathscr{S}(m) = \{(y_1, y_2, y_3) \in \mathbb{R}^3 \text{ ; there exists a decision rule } \delta \in \mathscr{D}(m) \text{ with } R_X(\delta) = y_1,\ R_F(\delta) = y_2 \text{ and } R_D(\delta) = y_3\}$$

We shall sometimes refer to $\mathscr{S}(m)$ as the image of the set of decision rules $\mathscr{D}(m)$ and to a point

$$\mathbb{R}(\delta) = (R_X(\delta),\ R_F(\delta),\ R_D(\delta))$$

as the image of the decision rule δ .

The performance set $\mathscr{S}(m)$ is most aptly suited to the respond-once decision devices, since, for these devices, the quantities of interest are R_X, R_F and R_D. In dealing with the respond-and-hold decision devices, however, we should like to relate these quantities to the average detection time and the average false alarm time. This can be done as follows. Define the normalized detection duty D_D and the normalized false alarm duty D_F by

$$D_D = \frac{E\{T_D / T\}}{E\{T_S / T\}}$$

and

$$D_F = \frac{E\{T_F/T\}}{E\{T_N/T\}}$$

Now recall from Eq. 2.23 of Section 2.5 that

$$T_F = \Delta N_F$$

and

$$T_D = \Delta(N_D + N_X)$$

Also, note that by employing the same reasoning used to obtain Eq. 2.23, we may write

$$T_S = \Delta N_S$$

and

$$T_N = \Delta N_N$$

Now, substitute for T_D and T_S into the equation for D_D to obtain

$$D_D = \frac{E\{\Delta(N_D + N_X)/T\}}{E\{\Delta N_S/T\}}$$

$$= \frac{r_D + r_X}{r_S}$$

$$= \frac{r_D + r_X}{mr_p}$$

or finally

$$D_D = \frac{1}{m}(R_D + R_X)$$

with the last equality following from Eq. 3.9. Next, substitute for T_F and T_N into the equation for D_F to obtain

$$D_F = \frac{E\{\Delta N_F/T\}}{E\{\Delta N_N/T\}}$$

$$= \frac{r_F}{r_N}$$

$$= \left[\frac{r_F}{(r_N/m)}\right]\frac{1}{m}$$

or finally

$$D_F = \frac{R_F}{m}$$

Now the point to be made here is that if we have the performance point,

$$R(\delta) = (R_X(\delta),\ R_F(\delta),\ R_D(\delta))$$

then we can associate with this point another performance point

$$D(\delta) = (D_D(\delta),\ D_F(\delta),\ R_D(\delta))$$

by the transformation

$$D_D(\delta) = (R_X(\delta) + R_D(\delta))/m \tag{3.13}$$

$$D_F(\delta) = R_F(\delta)/m$$

$$R_D(\delta) = R_D(\delta)$$

Thus we can also define the performance set for the respond-and-hold decision devices, $\tilde{\mathscr{P}}(m)$, either by the definition

$$\tilde{\mathscr{P}}(m) = \{y \in \mathbb{R}^3 \text{ ; there exists a } \delta \in \mathscr{D}(m) \text{ with}$$

$$D_D(\delta) = y_1,\ D_F(\delta) = y_2 \text{ and } R_D(\delta) = y_3\}$$

or as the image of $\mathscr{P}(m)$ under the transformation of Eq. 3.13. In either case, we will hereafter refer to the performance sets $\mathscr{P}(m)$ and $\tilde{\mathscr{P}}_T(m)$ as the respond-once performance set and the respond-and-hold performance set, respectively.

3.3 Some General Properties of the Performance Sets

In this section, some general properties of the performance sets $\mathscr{S}(m)$ and $\tilde{\mathscr{S}}(m)$ are developed. First some simple bounds on these sets are derived. Next, a notion of randomized decision rules is introduced to insure that $\mathscr{S}(m)$ and $\tilde{\mathscr{S}}(m)$ are convex. This convexity is then used to extend the notion of a chance diagonal to the FRD problem. Next the relationship between the performance of a given decision device δ and the performance of the "inverse" antipodal decision device is determined. Finally, we verify the "most recent response" rule introduced in Section 1.3.

We first show that the set $\mathscr{S}(m)$ is contained in a triangular shaped wedge in the positive quadrant of $\mathbb{R}^3$. That $\mathscr{S}(m)$ lies in the positive quadrant is immediately evident from the fact that, by definition, $r_\beta(\delta)$ is non-negative and, hence, $R_\beta(\delta)$ is also non-negative. Furthermore, from Eq. 3.12 it follows that $\mathscr{S}(m)$ must lie in the rectangular volume.

$$0 \leq R_X \leq m-1$$

$$0 \leq R_F \leq m$$

$$0 \leq R_D \leq 1$$

The final result appears in the following theorem and its corollary.

Theorem 3.1 The performance set $\mathcal{S}(m)$ is contained in the region defined by

$$0 \leq R_F \leq m \tag{3.14}$$

$$0 \leq R_X \leq (m-1)$$

and $$R_X/(m-1) \leq R_D \leq 1$$

Proof. From the above remarks, it is only necessary to show that

$$\frac{R_X}{m-1} \leq R_D$$

We reason as follows. An extra detection cannot be covered by a pulse that is not detected. Furthermore, any pulse that is detected cannot cover more than m-1 extra detections. Thus, $N_X \leq (m-1) N_D$. Dividing by T, taking the expected value, and then normalizing yields the desired result.

Corollary 3.2 The performance set $\tilde{\mathcal{S}}(m)$ is contained in the region.

$$0 \leq D_D \leq 1 \tag{3.15}$$

$$0 \leq D_F \leq 1$$

$$D_D \leq R_D \leq \min[1, D_D/m]$$

Proof. The result follows immediately by applying the linear transformation of Eq. 3.13 to the triangular wedge of Theorem 3.1.

The regions of Theorem 3.1 and its corollary are illustrated in Fig. 3.3.

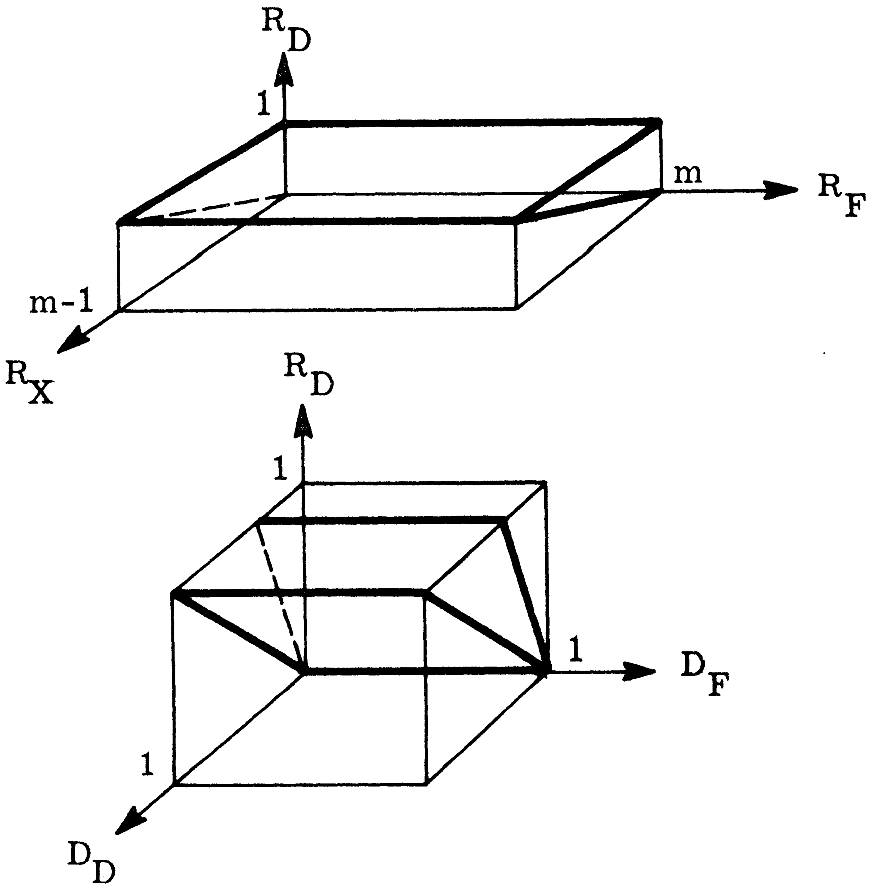


Fig. 3.3. The region containing $\mathcal{S}(m)$ and $\tilde{\mathcal{S}}(m)$

Next we show that the performance sets are convex. In classical detection theory, this is accomplished by introducing the notion of the randomized decision rule. The same idea can be extended to the FRD problem. The physical interpretation is as follows. Suppose that δ' and δ'' are two decision devices. Further, suppose that, before a decision device is to be put into operation, we perform a random experiment with two outcomes, say "A" and "B", with probabilities p and 1-p respectively. If the outcome "A" is observed, we use δ', and, if the outcome "B" is observed, we use δ''. Then, the β-rate for the randomized receiver δ is

$$r_\beta(\delta) = \frac{E[N_\beta(\delta(x),\theta)\,|\delta=\delta']\ p + E[N_\beta(\delta(x),\theta)\,|\delta=\delta'']\ (1-p)}{T}$$

$$= r_\beta(\delta')\ p + r_\beta(\delta'')\ (1-p)$$

Dividing through by the proper normalizing factor, yields

$$R_\beta(\delta) = R_\beta(\delta')\ p + R_\beta(\delta'')\ (1-p) \tag{3.16}$$

From Eq. 3.16 we obtain the desired theorem and its corollary.

Theorem 3.3 The performance set $\mathcal{S}(m)$ is convex.

Proof. It is recalled that a set S is convex, if, given any two points $a, b \in S$, the line segment

$$ap + b(1-p),\ p \in [0,1]$$

is in S. Now, we have shown that, given any two decision devices δ' and δ'' with images

$$R(\delta') = (R_x(\delta'),\ R_F(\delta'),\ R_D(\delta'))$$

and

$$R(\delta'') = (R_X(\delta''),\ R_F(\delta''),\ R_D(\delta''))$$

in $\mathscr{S}(m)$, there exists a decision device δ with image

$$R(\delta')\, p + R(\delta'')\, (1-p)$$

in $\mathscr{S}(m)$. Since this holds for all $p \in [0,1]$, it follows that $\mathscr{S}(m)$ is convex.

Corollary 3.4 The performance set $\mathscr{S}(m)$ is convex.

Proof. The set $\tilde{\mathscr{S}}(m)$ is obtained as the image of $\mathscr{S}(m)$ under the linear transformation of Eq. 3.13. Thus, the conclusion follows, since the image of a convex set under a linear transformation is convex.

The convexity properly can be used to extend the notion of a chance diagonal to the FRD problem. The procedure is basically the same as that followed in classical detection theory. First, consider the decision device that never responds. Since a detection, a false alarm or an extraneous detection can occur only if the decision devices respond, we have $r_D = r_F = r_X = 0$, and thus $R_D = R_F = R_X = 0$. Hence, the point $(0,\ 0,\ 0) \in \mathscr{S}(m)$.

Next consider the decision device that always responds. For this device, $r_C = r_R = r_M = 0$, since correct rejections, rests and misses occur only when the device does not respond. Thus,

$R_C = R_R = R_M = 0$ and by Eq. 3.12 $R_X = m-1$, $R_F = m$ and $R_D = 1$. Thus, the point $(m-1, m, 1) \epsilon \mathscr{S}(m)$.

Finally, consider the decision device that responds every mth time. Its performance is determined as follows. First, note that since each pulse covers m successive decision times, each pulse must cover one and only one decision time at which a response is made. Thus, $N_D = N_p$ and $N_X = 0$. It follows that $R_D = 1$ and $R_X = 0$. Now, r_F is given by

$$r_F = \sum_{i=1}^{N} \frac{Pr[a_i = 1, \theta_i = T_p]}{T}$$

But, if $i \neq m$, $a_i = 0$ so that $Pr[a_i = 1, \theta_i = T_p] = 0$. On the other hand, if $i = m$, then $a_i = 1$ so that $Pr[a_i = 1, \theta_i = T_p] = Pr[\theta_i = T_p]$. It then follows that,

$$r_F = \frac{(N/m)\ Pr[\theta_i = T_p]}{N\Delta}$$

$$= \frac{Pr[\theta_i = T_p]}{T_p}$$

$$= r_N/m$$

so that by Eq. 3.11

$$R_F = 1$$

Thus, the receiver that responds every mth time has the performance point (0,1,1).

The convexity theorem can now be applied to the performance points (0,0,0), (0,1,1) and (m-1,m,1) to obtain additional points in $\mathscr{S}$(m). By Theorem 3.3 every point on the line segments connecting these points must be in $\mathscr{S}$(m). Moreover, by convexity, every point in the segment of the plane bounded by these lines and segments is in $\mathscr{S}_T$(m). (See Fig. 3.4.)

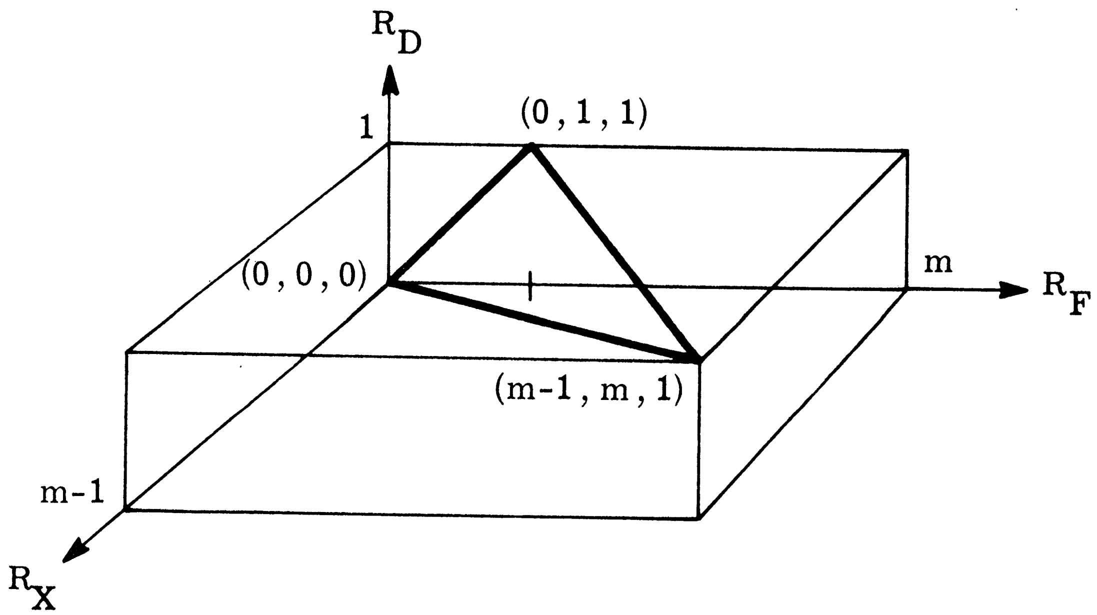


Fig. 3.4 The randomization between the performance points (0,0,0), (0,1,1) and (m-1,m,1)

The segment of the plane shown in Fig. 3.4 can be interpreted as follows. First, note that every point on the line connecting the points (0,0,0) and (m-1,m,1) can be obtained by randomizing between a decision device that always responds and a decision device that never responds. Thus, this line has the usual property of a chance diagonal. Next from the remarks preceeding the statement of Theorem 3.8 in Section 3.4, it is clear that, among all decision devices that have images in the $R_X = 0$ plane, the decision device with image $(0,1,1)$ responds as often as possible. Thus, the line connecting the points $(0,0,0)$ and $(0,1,1)$ has the usual property of a chance diagonal. Finally, it is noted that any other point in the segment of the plane illustrated in Fig. 3.4 can be obtained by randomizing between points on the lines.

$$(0,0,0)\ p + (0,1,1)\ (1-p) \qquad p \in [0,1]$$

and

$$(0,0,0)\ p + (m-1,m,1)\ (1-p) \qquad p \in [0,1]$$

Thus, the segment of the plane shown in Fig. 3.4 can be interpreted as a "chance plane" for the performance set $\mathcal{S}(m)$. The equation of this plane is easily seen to be

$$R_X + R_D - R_F = 0 \tag{3.17}$$

A chance plane for the performance set $\tilde{\mathcal{S}}(m)$ can be obtained by applying the transformation of Eq. 3.13 to Eq. 3.17. The resulting chance plane is then given by

$$D_D - D_F = 0 \tag{3.18}$$

The segment of this plane that is contained in $\tilde{\mathscr{S}}(m)$ is illustrated in Fig. 3.5.

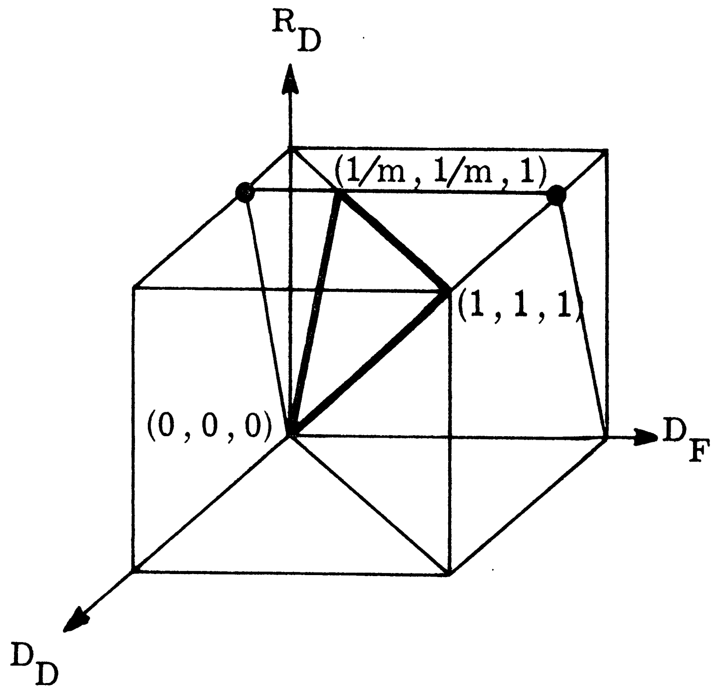


Fig. 3.5 The chance plane for $\tilde{\mathscr{S}}(m)$

Next we investigate a certain relationship that exists between the points in $\mathscr{S}(m)$ and $\tilde{\mathscr{S}}(m)$. To begin, it is recalled that, in classical detection theory, the antipodal decision rule δ^a of a decision rule δ is the decision rule that responds only when δ does not respond, and conversely. This idea can be extended to the FRD problem as follows. For a given decision rule

$$\delta = (\delta_1, \ldots, \delta_N)$$

define the antipodal decision rule, δ^a, by

$$\delta^a = (\delta_1^a, \ldots, \delta_N^a)$$

where

$$\delta_i^a(\vec{x}_i, \vec{a}_{i-1}) = \begin{cases} 1 & \text{if} \quad \delta_i(\vec{x}_i, \vec{a}_{i-1}) = 0 \\ 0 & \text{if} \quad \delta_i(\vec{x}_i, \vec{a}_{i-1}) = 1 \end{cases}$$

The interpretation here is that δ^a responds at those times that δ does not respond and fails to respond at those times that δ responds.

The performance of the antipodal decision rule δ^a is related to the performance of δ by the following theorem and its corollary.

Theorem 3.5 If δ^a is the antipodal decision rule for δ, then

$$R_F(\delta^a) = m - R_F(\delta) \tag{3.19a}$$

$$R_D(\delta^a) + R_X(\delta^a) = m - [R_D(\delta) + R_X(\delta)] \tag{3.19b}$$

Proof. Fix the decision device δ, the observation X and the pulse sequence θ, and let $N_{r_N}(\delta(x), \theta)$ be the number of responses that occur at the decision times that are not covered by a pulse. Similarly, let $N_{r_S}(\delta(x), \theta)$ be the number of responses occurring at decision times that are covered by a pulse. Then it follows immediately that,

$$N_{r_N}(\delta(x), \theta) = N_F(\delta(x), \theta)$$

and

$$N_{r_S}(\delta(x), \theta) = N_D(\delta(x), \theta) + N_X(\delta(x), \theta)$$

Now the antipodal device δ^a responds at those times that δ does not respond at and vice versa. Thus,

$$N_{r_N}(\delta^a(x), \theta) = N_N(\theta) - N_{r_N}(\delta(x), \theta)$$

and

$$N_{r_S}(\delta^a(x), \theta) = N_S(\theta) - N_{r_S}(\delta(x), \theta)$$

Combining the above equations results in

$$N_F(\delta^a(x), \theta) = N_N(\theta) - N_F(\delta(x), \theta)$$

and

$$N_D(\delta^a(x), \theta) + N_X(\delta^a(x), \theta) = N_S(\theta) - \left(N_D(\delta(x), \theta) + N_X(\delta(x), \theta)\right)$$

Dividing by T, taking expected values and normalizing yields Eq. 3.19. This completes the proof.

<u>Corollary 3.6</u> If δ^a is the antipodal decision rule for δ, then

$$D_F(\delta^a) = 1 - D_F(\delta)$$

$$D_D(\delta^a) = 1 - D_D(\delta) \qquad (3.20b)$$

<u>Proof.</u> Eq. 3.20 follows immediately from Eq. 3.19 by the transformation of Eq. 3.13.

The above theorem and its corollary have an interesting geometrical interpretation. To see this, we first combine Eqs. 3.19a and b to obtain

$$R_X(\delta^a) + R_D(\delta^a) - R_F(\delta^a)$$

$$= - \left[R_X(\delta) + R_D(\delta) - R_F(\delta)\right] \qquad (3.21)$$

Next, it is noted that the plane defined by

$$R_X + R_D - R_F = c$$

is a plane parallel to the chance plane and located at a normal distance of c away from the chance plane. Thus it is evident from Eq. 3.21 that the two points $R(\delta^a)$ and $R(\delta)$ are located on opposite sides of the chance plane and at equal distances from the chance plane. Precisely the same reasoning can be used to conclude that the two points $D(\delta^a)$ and $D(\delta)$ are on opposite sides of the chance plane for $\tilde{\mathscr{P}}(m)$ and at equal distances from that plane.

This section is concluded by strengthening the most recent response rule. Recall that it was conjectured in Chapter I that it is sufficient to consider only decision devices that depend upon the previous decisions only through the time of the most recent response in the interval $(t_i - T_p, t_i)$. This conjecture is now shown to be true in the following sense. Given any point in the performance set, $\mathscr{S}(m)$, there is a decision device that obeys the most recent response rule and that has that point as its image. Formally, we have the following theorem.

<u>Theorem 3.7</u> If $(y_1, y_2, y_3) \in \mathscr{S}_T(m)$, then there exists a $\delta = (\delta_1, \ldots, \delta N) \in \mathscr{D}(m)$ such that $y_1 = R_x(\delta)$, $y_2 = R_x(\delta)$ and $y_3 = R_D(\delta)$ and

$$\delta_i(\vec{x}_i, \vec{a}_{i-1}) = \delta_i(\vec{x}_i, \tau_i)$$

where τ_i is the most recent response function of Section 2.3.

The proof of Theorem 3.7 is found in Appendix F.

3.4 The Inhibit Rule and the $R_X = 0$ Plane

In Section 1.4 we introduced the ROMF decision device. This device responds if the output of a matched filter exceeds a fixed threshold β and if no responses have been made at any of the previous $m-1$ decision opportunities. This last condition is imposed to insure that no extra detections result. We have called such a condition the "inhibit" rule. In this section, we investigate decision devices which obey the inhibit rule.

Before proceeding, the notion of the inhibit rule must be made more precise. In Appendix F it is shown that the component δ_i of any decision rule δ can be expressed in terms of a response set $A_i(\vec{x}_{i-1}, \tau_i)$. Specifically,

$$\delta_i(\vec{x}_i, \tau_i) = \begin{cases} 1 & \text{if} \quad x_i \in A_i(\vec{x}_{i-1}, \tau_i) \\ 0 & \text{if} \quad x_i \notin A_i(\vec{x}_{i-1}, \tau_i) \end{cases}$$

The interpretation of the response set $A_i(\vec{x}_{i-1}, \tau_i)$ is that for $\vec{x}_{i-1}$ and τ_i fixed, $A_i(\vec{x}_{i-1}, \tau_i)$ is the set of current observations, $x_i \in X_i$, for which the decision device responds at the decision time t_i. Now, the condition $\tau_i > 0$ is equivalent to the condition that at least one response has been made within the last $m-1$ decision times. Thus, if the probability of the set $A_i(\vec{x}_{i-1}, \tau_i)$ for $\vec{x}_{i-1}$ fixed and any pulse configuration θ_i is greater than zero for $\tau_i > 0$, then the decision device is allowed to violate the inhibit rule. Therefore, for a decision device to satisfy the inhibit rule,we must have

$$\Pr[A_i(\vec{x}_{i-1}, \tau_i) | \vec{x}_{i-1}, \theta_i] = 0 \qquad \text{for } \tau_i > 0$$

This probability depends on all the past observations $\vec{x}_{i-1}$ and all the possible current pulse configurations θ_i. Mathematically speaking, it is only reasonable to require that this probability be zero except perhaps on subsets of $\vec{X}_{i-1} \times \Theta_i$ of probability zero with respect to $P(d\vec{x}_{i-1}, d\theta_i)$. It is common practice in measure theory to indicate that a certain proposition holds, except perhaps on subsets of measure zero with respect to a measure μ, by appending to the statement of that proposition the symbol $[\mu]$, which is read "almost everywhere with respect to μ." If we follow this convention here, then the formal definition of the inhibit rule is as follows.

The decision rule δ satisfies the inhibit rule iff for $i = 1, 2, \ldots, N$

$$\Pr[A_i(\vec{x}_{i-1}, \tau_i) | \vec{x}_{i-1}, \theta_i] = 0 \text{ for } \tau_i > 0 \qquad [P(d\vec{x}_{i-1}, d\theta_i)] \tag{3.22}$$

The significance of this definition is easily seen from Eq. F1.b in Appendix F. Specifically, if δ satisfies the above definition then the probability of making an extra detection at time t_i, P_i^X, is

$$P_i^X =$$

$$\int_{\theta_i \in [0, T_p)} \left[\sum_{j \,\ni\, j\Delta > 0}^{m-1} \left[\int_{\vec{X}_{i-1}^j} \left[\int_{A_i(\vec{x}_{i-1}, \tau_i = j\Delta)} P(dx_i | \vec{x}_{i-1}, \theta_i) \right] P(d\vec{x}_{i-1} | \theta_i) \right] P(d\theta_i) \right]$$

$$= 0$$

since

$$\int_{A_i(\vec{x}_{i-1},\ \tau_i=j\Delta)} P(dx_i|\vec{x}_{i-1},\ \theta_i) = 0 \qquad [P(d\vec{x}_{i-1},\ d\theta_i)]$$

Furthermore, from Eq. 3.4 it follows that $R_X = 0$. Thus, the performance point for a decision device that satisfies the inhibit rule lies completely in the $R_X = 0$ plane.

The main result of this section is that, under relatively mild conditions on the family $\{P(dx|\theta);\ \theta \in \Theta\}$, the converse statement holds. Namely, the only decision devices with images in the $R_X = 0$ plane are those that obey the inhibit rule.

Theorem 3.8 Suppose for all i

$$P(dx_i|\vec{x}_{i-1},\theta_i) >> P(dx_i|\vec{x}_{i-1},\theta_i=T_p) \qquad [P(d\vec{x}_{i-1},d\theta_i]$$

Then, given any performance point in the $R_X = 0$ plane of $\mathscr{S}(m)$, there exists a decision device that satisfies the inhibit rule and that has that point as its image.

Proof. It is desired to show that given any performance point in $\mathscr{S}(m)$ with $R_X(\delta) = 0$, then the response sets for the decision device δ satisfy

$$\int_{A_i(\vec{x}_{i-1},\tau_i=j\Delta)} P(dx_i|\vec{x}_{i-1},\theta_i) = 0 \qquad \text{for} \qquad j = 1,\ldots, m-1 \tag{3.23}$$

except perhaps on a set $A \subset \vec{X}_{i-1} \times \Theta_i$ with zero probability. Now suppose it can be shown that Eq. 3.23 holds except perhaps on a set $A^1 \subset \vec{X}_{i-1} \times [0, T_p)$ with zero probability. Then, since, by hypothesis, $P(dx_i|\vec{x}_{i-1},\theta_i) >> P(dx_i|\vec{x}_{i-1},\theta_i = T_p)$ $[P(d\vec{x}_{i-1}, d\theta_i)]$, we have

$$\int_{A_i(\vec{x}_{i-1},\ \tau_i=j\Delta)} P(dx_i|\vec{x}_{i-1},\theta_i = T_p) = 0 \qquad \text{for} \qquad j = 1,\ldots, m-1$$

Thus, it is sufficient to verify Eq. 3.23 up to subsets of probability zero contained in $\vec{X}_{i-1} x[0, T_p)$. The proof is by contradiction. Suppose there exists some $j \in \{1,\ldots, m-1\}$ and some set $B \subset \vec{X}_{i-1} x[0, T_p)$ with

$$\int_B P(dx_i, d\theta_i) > 0$$

such that

$$\int_{A(\vec{x}_{i-1},\tau_i=j\Delta)} P(dx_i|\vec{x}_{i-1},\theta_i) > 0 \qquad \text{for} \qquad (\vec{x}_{i-1},\theta_i) \in B$$

Then it follows from Eq. F1.b in Appendix F that

$$R_X(\delta) > P_i^X > \int_{[0,T_p)} \left[\int_{\vec{X}_{i-1}^j} \left[\int_{A_i(\vec{x}_{i-1},\tau_i=j\Delta)} P(dx_i | \vec{x}_{i-1}, \theta_i) \right] P(d\vec{x}_{i-1} | \theta_i) \right] \Pi(d\theta_i)$$

Moreover, by the definition of $A_i(\vec{x}_{i-1}, \tau_i = j\Delta)$ in Appendix F, $B \subset \vec{X}_{i-1}^j \times [0, T_p)$ so that

$$\int_{[0,T_p)} \left[\int_{\vec{X}_{i-1}^j} \left[\int_{A_i(\vec{x}_{i-1},\tau_i=j\Delta)} P(dx_i | \vec{x}_{i-1}, \theta_i) \right] P(d\vec{x}_{i-1} | \theta_i) \right] \Pi(d\theta_i)$$

$$= \int_{\vec{X}_{i-1}^j \times [0,T_p)} \left[\int_{A_i(\vec{x}_{i-1},\tau_i=j\Delta)} P(dx_i | \vec{x}_{i-1}, \theta_i) \right] P(d\vec{x}_{i-1}, d\theta_i)$$

$$> \int_{B} \left[\int_{A_i(\vec{x}_{i-1},\tau_i=j\Delta)} P(dx_i | \vec{x}_{i-1}, \theta_i) \right] P(d\vec{x}_{i-1}, d\theta_i)$$

$$> 0$$

Thus, $R_X > 0$ which contradicts the assumption that δ has its image in the $R_X = 0$ plane. This completes the proof.

As a corollary to the above theorem, we may strengthen Theorem 3.5 for those decision devices that satisfy the inhibit rule.

Theorem 3.9 Under the hypothesis of Theorem 3.8, if δ satisfies the inhibit rule and if $m>1$, then the antipodal decision device, δ^a, has the image

$$R_F(\delta^a) = m - R_F(\delta)$$

$$R_X(\delta^a) = m-1 - R_D(\delta)$$

$$R_D(\delta^a) = 1$$

Proof. Since δ satisfies the inhibit rule, the responses must occur no more frequently than every mth decision time. Therefore, the antipodal decision device cannot make no-response decisions any more frequently than every mth decision time. Hence, every pulse that occurs must be detected. Thus

$$N_D(\delta^a(x), \theta) = N_p(\theta)$$

so that

$$R_D(\delta^a) = 1$$

Substituting $R_D(\delta^a) = 1$ and $R_X(\delta) = 0$ into Eq. 3.19 gives the desired result.

3.5 Optimality and ROC Surfaces

In this section we borrow the idea of the ROC curve from classical detection theory and apply this idea to describe the optimum performance of the FRD decision devices. The result is two ROC surfaces, one for respond-once decision devices and one for respond-and-hold decision devices.

The procedure used to define the ROC surfaces is in complete analogy with classical detection theory. We first define a concept of "better than" in terms of the performance sets and then we use this concept to define the ROC surface as representing the "best" possible performance.

Consider first the respond-once case. The basic idea here is to say that one decision device is "better than" another if on the average it detects more pulses without making more false alarms or extra detections. More precisely, we will say that δ' is better than δ'' iff

$$R_X(\delta') \leq R_X(\delta'')$$

$$R_F(\delta') \leq R_F(\delta'')$$

$$R_D(\delta') > R_D(\delta'')$$

To obtain a geometrical interpretation of this definition we define the "inner quadrant at the point z" as the subset $Q_z^i \subset \mathbb{R}^3$ by

$$Q_z^i = \{y \in \mathbb{R}^3 ;\ y_1 \leq z_1,\ y_2 \leq z_2,\ y_3 > z_3\}$$

Then, it follows from the above definition that the set of performance points that are better than the point

$$R(\delta'') = (R_X(\delta''),\ R_F(\delta''),\ R_D(\delta''))$$

is the set

$$Q^i_{R(\delta'')} \cap \mathscr{S}$$

Such a set is illustrated in Fig. 3.6.

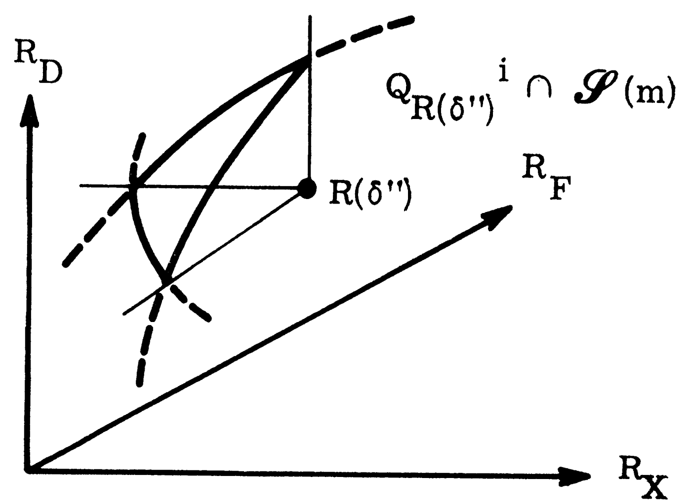


Fig. 3.6 The set of performance points better than $R(\delta'')$

We will now define the respond-once ROC surface to consist of those points that represent the best possible performance. From the above figure it is clear that $R(\delta'')$ represents the best possible performance if $Q^i_{R(\delta'')}$ is just touching that portion of the upper surface that is closest to the $R_X = 0$ plane. But, this occurs when the point $R(\delta'')$ is the only point in $Q^i_{R(\delta'')} \cap \mathscr{S}$. Thus we have the definition:

The respond-once ROC surface $\lambda_{\mathscr{S}(m)}$ is the set

$$\lambda_{\mathscr{S}} = \{y \in \mathbb{R}^3 \ ; \ Q^i_y \cap \mathscr{S}(m) = \{y\} \}$$

A typical respond-once ROC surface is shown in Fig. 3.7.

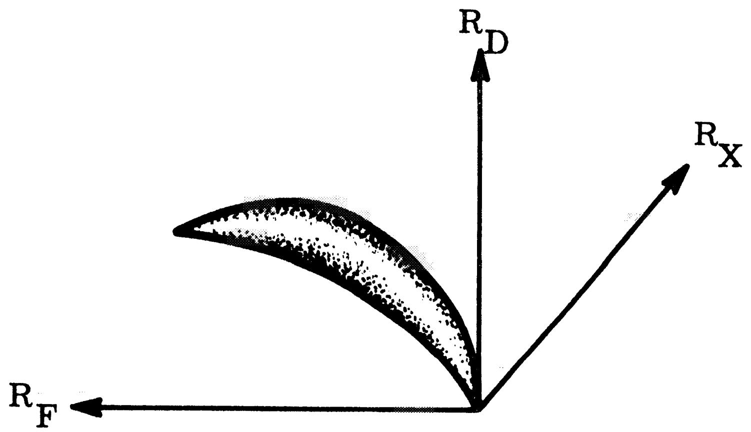


Fig. 3.7. A typical respond-once ROC surface

We treat the respond-and-hold case in the same way except now we deal with the performance set $\tilde{\mathscr{S}}(m)$. The underlying idea here is to say that one decision device is "better than" another if it detects more pulses and has more detection time but less false alarm time. That is, we will say that δ' is better than δ'' if

$$D_D(\delta') > D_D(\delta'')$$

$$D_F(\delta') \leq D_F(\delta'')$$

$$R_D(\delta') > R_D(\delta'')$$

The geometrical interpretation of this definition is now expressed in terms of the "outer quadrant at the point z". This is the set Q^o defined by

$$Q^o = \{y \in \mathbb{R}^3 \; ; \; y_1 > z_1, \; y_2 \leq z_2, \; y_3 > z_3\}$$

It then follows that the set of performance points that are better than the point

$$D(\delta'') = (D_D(\delta''), \; D_F(\delta''), \; R_D(\delta''))$$

is the set

$$Q^o_{D(\delta'')} \cap \tilde{\mathscr{P}}$$

An illustration appears in Fig. 3.8.

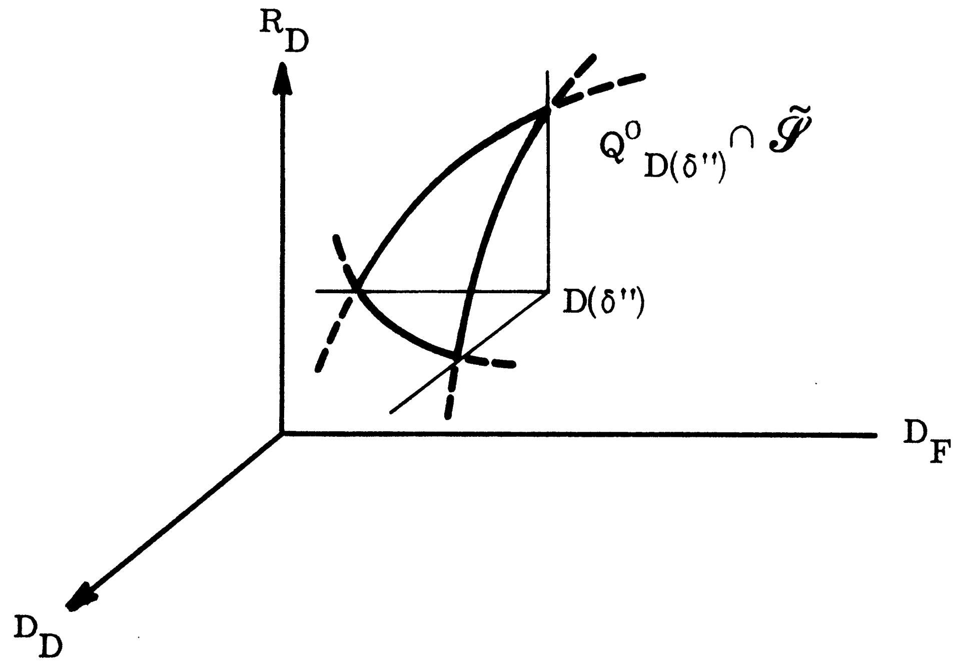


Fig. 3.8. The set of performance points better than D(δ")

The respond-and-hold ROC surface is the set of the best performance points in $\tilde{\mathcal{S}}(m)$. In the same manner as before, this set is seen to consist of that portion of the upper surface of $\tilde{\mathcal{S}}$ that is farthest from the $D_D = 0$ plane. The formal definition of this surface is as follows.

The respond-and-hold ROC surface $\lambda_{\tilde{\mathcal{S}}}$ is the set

$$\lambda_{\tilde{\mathcal{S}}} = \{y \in \mathbb{R}^3 \; ; \; Q_y^0 \cap \tilde{\mathcal{S}} = \{y\}\}$$

A typical respond-and-hold ROC surface is shown in Fig. 3.9.

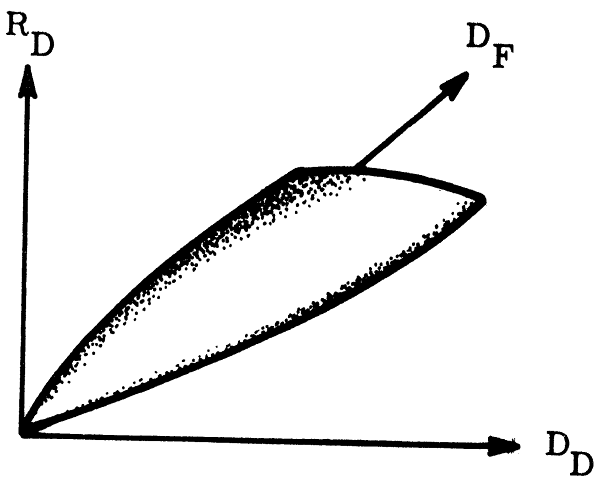


Fig. 3.9 A typical respond-and-hold ROC surface

As an application of the ideas of this section we will obtain a result on the effect of increasing the number of decision opportunities per pulse, m, on the performance. Specifically, it is shown that the performance gets better, or at least gets no worse, if m is doubled. This result follows from the fact that the ROC surfaces are contained in the upper boundaries of the performance sets and from the following theorem.

Theorem 3.10

$$\mathscr{P}(m) \subset \mathscr{P}(2m) \qquad (3.23a)$$

$$\tilde{\mathscr{P}}(m) \subset \tilde{\mathscr{P}}(2m) \qquad (3.23b)$$

Proof. Since the performance set $\tilde{\mathscr{P}}(m)$ can be obtained from the performance set $\mathscr{S}(m)$ by the transformation of Eq. 3.13 it is sufficient to prove only Eq. 3.23a. This is done as follows. It is desired to show that given any decision rule $\delta \in \mathscr{D}(m)$ with performance point $R(\delta) \in \mathscr{S}(m)$, there exists a rule $\delta' \in \mathscr{D}(2m)$ with the same performance point. To show this, define $\delta' = (\delta_1', \delta_2', \ldots, \delta_N')$ by

$$\delta'(\cdot,\cdot) = \begin{cases} \delta_i(\cdot,\cdot) & i \text{ even} \\ 0 & i \text{ odd} \end{cases}$$

Now, note that if $\mathscr{T}$ is the set of decision times for decision rule in $\mathscr{D}(2m)$, then $\mathscr{T}$ can be written as

$$\mathscr{T} = \{t_i = i\Delta;\ i \text{ even}\} \cup \{t_i = i\Delta;\ i \text{ odd}\}$$

But, the set $\{t_i = i\Delta;\ i \text{ even}\}$ is precisely the set of decision times for decision rules in $\mathscr{D}(m)$. (See Fig. 3.10.) Since δ' responds iff δ does for $t_i \in \{t_i = i\Delta;\ i \text{ even}\}$ and never responds for $t_i \in \{t_i = i\Delta;\ i \text{ odd}\}$, it follows that the rates R_X, R_F, R_D must be the same decision rules. This completes the proof.

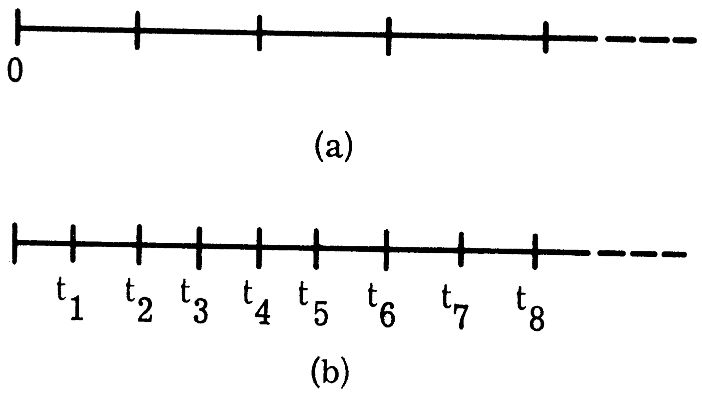


Fig. 3.10. The relation between the decision time for decision devices in $\mathscr{D}(2m)$ and $\mathscr{D}(m)$
(a) the decision times associated with $\mathscr{D}(m)$
(b) the decision times associated with $\mathscr{D}(2m)$

3.6 Optimality and the Bayes Decision Rules

In classical detection theory the performance of the Bayes decision rules is geometrically related to the ROC curve by the fact that the Bayes risk line intersects the ROC curve at the point of optimum performance. (See Fig. 3.2.) This relation can be used both to show that the concepts of Bayes optimality and ROC optimality are equivalent and to relate the losses in the Bayes risk to the probability of detection and probability of false alarm. In this section, these ideas are extended to the FRD problem.

We begin by establishing a relationship between the Bayes risk and the respond-once performance set $\mathscr{S}(m)$. To this end, it is first necessary to obtain an expression for the Bayes risk of a decision rule δ in terms of the $R_X(\delta)$, $R_F(\delta)$ and $R_D(\delta)$. This is done as follows. In Section 2.1 we defined the Bayes risk as the average loss. In this section only, however, it will be more convenient to normalize this risk by dividing by the length of the total observation interval T . We then have

$$r(\delta) = \frac{E[L(\delta(x), \theta)]}{T}$$

Substituting for the loss function $L(\cdot, \cdot)$ from Eq. 2.25 yields

$$r(\delta) = \frac{E\{W_X N_X(\delta(x), \theta) + W_F N_F(\delta(x), \theta) - N_D(\delta(x), \theta)\}}{T}$$

$$= W_X \frac{E\{N_X(\delta(x), \theta)\}}{T} + W_F \frac{E\{N_F(\delta(x), \theta)\}}{T} - \frac{E\{N_D(\delta(x), \theta)\}}{T}$$

$$= W_X r_X(\delta) + W_F r_F(\delta) - r_D(\delta)$$

Next, substitute for $r_X(\delta)$, $r_F(\delta)$ and $r_D(\delta)$ from Eq. 3.11. The result is

$$\frac{r(\delta)}{r_p} = W_X R_X(\delta) + W_F L_0^{-1} R_F(\delta) - R_D(\delta) \qquad (3.24)$$

where

$$L_0 = r_p/(r_N/m) = Pr[\theta_i \in [0, T_p)]/ Pr[\theta_i = T_p]$$

Now, note that Eq. 3.24 defines the plane in (R_X, R_F, R_D) space that is normal to the vector

$$(W_X, \ W_F/L_0, \ -1)$$

and which has intercepts

$$R_X = r(\delta)/r_p W_X$$

$$R_F = r(\delta)/r_p W_F L_0^{-1}$$

$$R_D = -r(\delta)/r_p$$

The significance of this plane lies in the fact that it contains all performance points $R(\delta)$ that result in the same Bayes risk. Moreover, if we denote the Bayes rule by δ^o, then as $r(\delta)$ is allowed to approach its minimum value of $r(\delta^o)$, (the minimum Bayes risk), then this plane approaches the plane that intersects the upper surface of $\mathscr{S}(m)$ at the point $R(\delta^o)$. Thus, in complete analogy to classical detection theory, the performance points for the Bayes decision rules may be characterized as those points for which the Bayes risk plane intersects the upper surface of $\mathscr{S}(m)$. The geometrical relation here is shown in Fig. 3.11.

We are now in a position to relate the respond-once Bayes decision rules (the case $W_X \geq 0$, $W_F > 0$) to the respond-once ROC

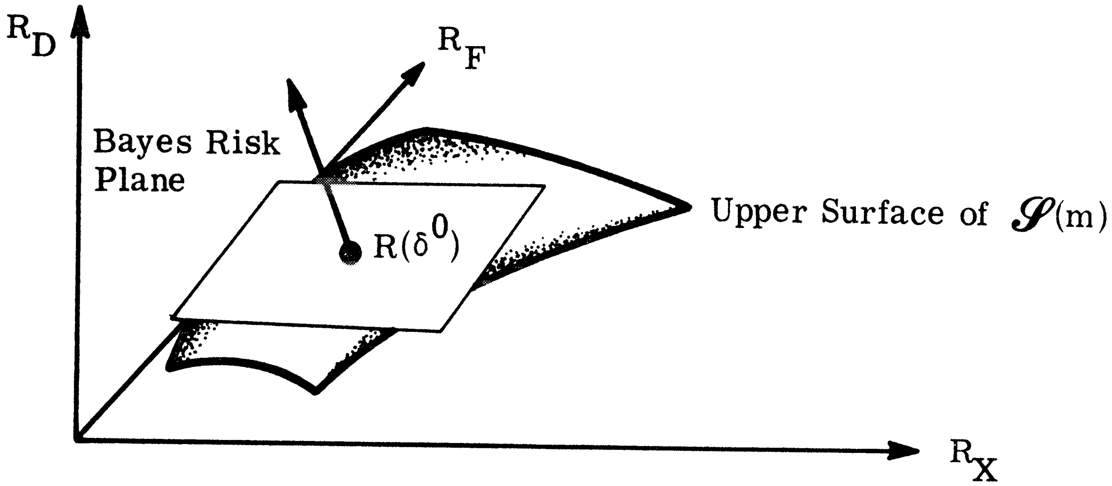


Fig. 3.11. The Bayes risk plane for the respond-once case

surface $\lambda_{\mathscr{S}}$. To this end, fix $R_F = R_F^*$ and consider a section of the upper surface of $\mathscr{S}(m)$. Since $\mathscr{S}(m)$ is convex, this section appears as a continuous convex function $f_{R_F}(R_X) = R_D$ in the $R_F = R_F^*$ plane as shown in Fig. 3.12.

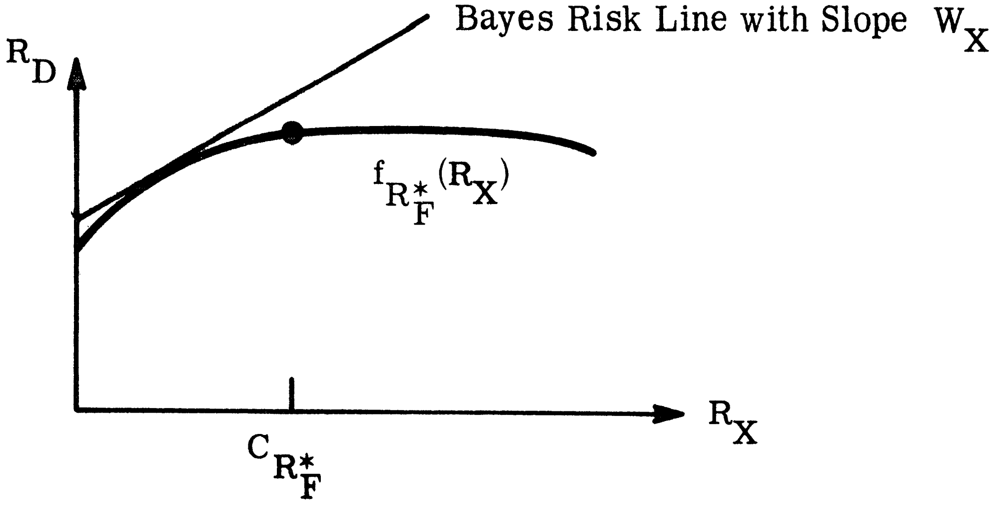


Fig. 3.12. A section of the upper surface of $\mathscr{S}(m)$ at $R_F = R_F^*$

Now, let $C_{R_F^*}$ be the smallest value of R_X for which $f_{R_F^*}(R_X)$ takes on its maximum value. Specifically,

$$C_{R_F^*} = \inf \left\{ R_X' \; ; \; f_{R_F^*}(R_X') = \sup_{R_X} f_{R_F^*}(R_X) \right\}$$

Then the section of the respond-once ROC surface at $R_F = R_F^*$ is given by

$$R_D = f_{R_F^*}(R_X) \quad \text{for} \quad R_X \in [0, C_{R_F^*}]$$

Next, note that the Bayes risk plane for the Bayes decision rule δ^o with $R_F(\delta^o) = R_F^*$ must intersect the section $f_{R_F^*}(R_X) = R_D$ as a line

$$R_D = W_X R_X + [W_F L_0^{-1}) R_F^* - r(\delta^o)/r_p]$$

as shown in Fig. 3.12. Since the slope of this line is W_X, it is immediately clear that this line intersects the respond-once ROC curve iff $W_X \geq 0$. Thus, all Bayes rules associated with $W_X \geq 0$ have performance points lying on the respond-once ROC surface. Moreover, for any point on the respond-once ROC surface, there is a Bayes rule for some $W_X \geq 0$ with that performance point.

Before turning to the respond-and-hold case, conjectures I, II and III of Section 1.4 are verified. This is done in the following two theorems. The first of these theorems says that if the loss W_X is increased past a certain value then there is only one Bayes rule and that rule is an inhibit rule. This is the content of conjectures I and II.

Theorem 3.11. For each $W_F > 0$ there exists a $W_F^*(W_F) \geq 0$ such that, if δ^* is a Bayes rule for the losses W_X^* and W_F, then

(1) δ^* is an inhibit rule

and

(2) if $\tilde{W}_X > W_X^*(W_F)$, then δ^* is also a Bayes rule for the losses $\tilde{W}_X$ and W_F.

Proof. Fix $W_F = W_F^*$ and define $W_X^*(W_F^*)$ to be that value of W_X such that the Bayes risk plane is tangent to the upper surface of $\mathscr{S}$ at some point in the $R_X = 0$ plane. Denote the Bayes rule associated with the losses $W_X^*(W_F^*)$ and W_F by δ^*. This construction is shown in Fig. 3.13 in which the Bayes risk plane appears as the line tangent to the section $f_{R_F(\delta^*)}(R_X)$ at the point $R_X = 0$. Now, by construction, $R_X(\delta^*) = 0$ so that, from Theorem 3.8, δ^* must be an inhibit rule.

It remains to show that δ^* is also a Bayes rule for the losses $\tilde{W}_X > W_X^*(W_F^*)$ and W_F^*. To accomplish this, denote the Bayes risk of an arbitrary rule δ for the losses $\tilde{W}_X$ and W_F^* by $\tilde{r}(\delta)$ and the Bayes risk of δ for the losses $W_X^*(W_F^*)$ and W_F^* by $\hat{r}(\delta)$. Then,

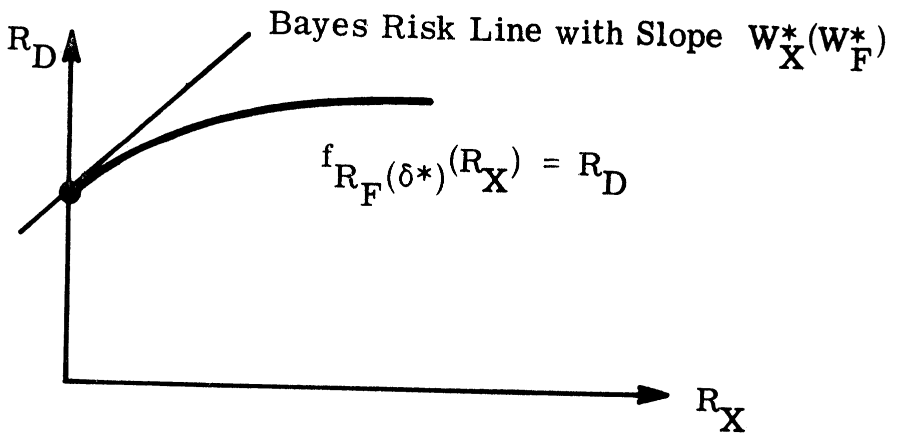


Fig. 3.13. The Bayes risk line tangent to $f_{R_F(\delta^*)}(W_X)$ at the point $R_X = 0$

$$\tilde{r}(\delta) = \tilde{W}_X R_X(\delta) + W_F^* L_0^{-1} R_F(\delta) - R_D(\delta)$$

$$\geq W_X^*(W_F^*) R_X(\delta) + W_F^* L_0^{-1} R_F(\delta) - R_D(\delta) = \hat{r}(\delta)$$

But, by the definition of a Bayes rule,

$$\hat{r}(\delta^*) \leq \hat{r}(\delta)$$

Thus

$$\hat{r}(\delta^*) \leq \tilde{r}(\delta)$$

The proof is then completed by noting that, since $R_X(\delta^*) = 0$,

$$\hat{r}(\delta^*) = \tilde{r}(\delta^*)$$

so that δ^* satisfies

$$\tilde{r}(\delta^*) = \inf_{\delta} \tilde{r}(\delta)$$

The next theorem says that the smaller the value of W_X the larger the value of R_D for a fixed value of R_F. This is the content of conjecture III.

Theorem 3.12. If δ' and δ'' are Bayes rules for the losses W_X' and W_X'' respectively and if $R_F(\delta') = R_F(\delta'') = R_F^*$, then

$$0 \leq W_X' < W_X''$$

implies

$$R_D(\delta') \geq R_D(\delta'')$$

In lieu of a formal proof of this theorem we will refer to Fig. 3.14. Here it is seen that, since the slope W_X' of the Bayes risk line for δ' is less than the slope W_X'' of the Bayes risk line for δ'', $R_D(\delta'')$ must be greater than or equal to $R_D(\delta'')$. Moreover, the maximum value of R_D for $R_F = R_F^*$ will occur for the Bayes rule with $W_X = 0$.

The only comment that might be made about the above geometrical argument is that it is possible that the section $f_{R_F^*}(R_X)$ is a

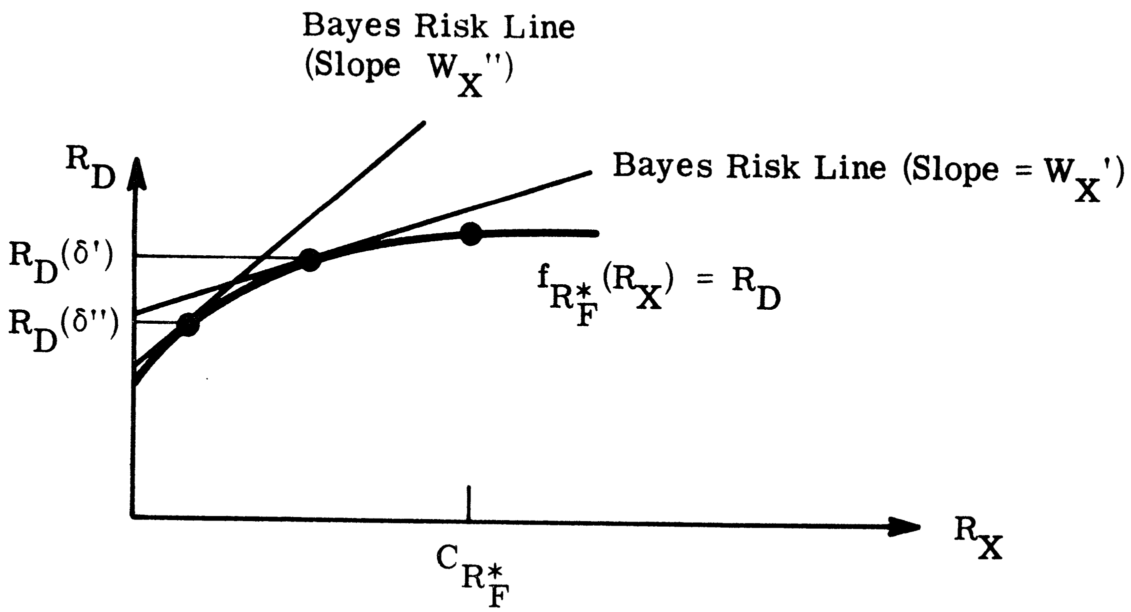


Fig. 3.14. The intersection of the two Bayes risk lines with the section $f_{R_F^*}(R_X)$

decreasing function of R_X for $R_X > 0$. In this case, Theorems 3.11 and 3.12 together say that there is only one Bayes rule with $R_F = R_F^*$ and that this rule is an inhibit rule corresponding to the loss $W_X = W_X^*(W_F) = 0$. We might point out here, however, that at least in the actual computations described in Chapter 5 this was not the case.

We now turn to the respond-and-hold case. First, we establish the geometrical relation between the Bayes risk and the respond-and-hold performance set $\tilde{\mathscr{P}}(m)$. To do this, substitute for $R_X(\delta)$ and $R_F(\delta)$ from Eq. 3.13 into Eq. 3.24 to get an alternate expression for the Bayes risk $r(\delta)$. Namely,

$$\frac{r(\delta)}{r_p} = (W_X m)\, D_D(\delta) + (W_F L_0^{-1} m)\, D_F(\delta) - (1 + W_X)\, R_D(\delta) \quad (3.25)$$

As above, this equation also defines a plane except now in terms of the variables D_D, D_F and R_F. The normal vector is

$$(W_X m,\ W_F L_0^{-1} m,\ -(1 + W_X))$$

and the intercepts are

$$D_D = r(\delta)/r_p\, W_X\, m$$

$$D_F = r(\delta)/r_p\, W_F\, L_0^{-1}\, m$$

$$R_D = -\, r(\delta)/r_p\, (1 + W_X)$$

The geometrical interpretation of the Bayes risk plane and the Bayes decision rule δ^0 is the same as in the respond-once case. This relation is illustrated in Fig. 3.15.

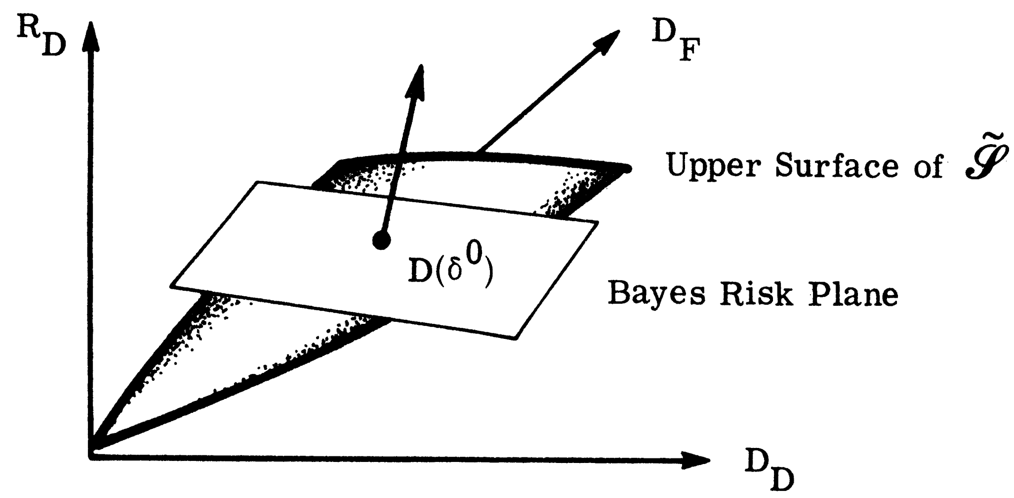


Fig. 3.15. The Bayes risk plane for the respond-and-hold case

To relate the respond-and-hold Bayes decision rules (the case $-1 \leq W_X < 0$, $W_F > 0$) to the respond-and-hold ROC surface $\lambda_{\tilde{\mathscr{S}}}$, we proceed as before. First, denote the section of the upper surface of $\tilde{\mathscr{S}}$ at $D_F = D_F^*$ by $g_{D_F^*}(D_D) = R_D$. Then, define $C_{D_F^*}$ to be the largest value of D_D for which $g_{D_F^*}(D_D) = R_D$ takes on its maximum value. Specifically

$$C_{D_F^*} = \sup \left\{ D_D' \, ; \, g_{D_F^*}(D_D') = \sup_{D_D} g_{D_F^*}(D_D) \right\}$$

Finally, note that the Bayes risk plane intersects $g_{D^*}(D_D)$ as the line

$$R_D = \left[\frac{W_X m}{1 + W_X} \right] D_D + \left[\left(\frac{W_F L_0^{-1} m}{1 + W_X} \right) D_F^* - \frac{r(\delta)}{r_p(1 + W_X)} \right]$$

These ideas are illustrated in Fig. 3.16. Then proceeding as before, note that the portion of the section $g_{D_F^*}(D_D)$ that is contained in the respond-and-hold ROC surface is given by

$$R_D = g_{D_F^*}(D_D) \qquad \text{for} \qquad D_D > C_{D_F^*}$$

Moreover, as W_X ranges from -1 to 0, the slope of the Bayes risk line ranges from $-\infty$ to 0. Thus, as before, it follows from Fig. 3.16 that all respond-and-hold Bayes rules have performance

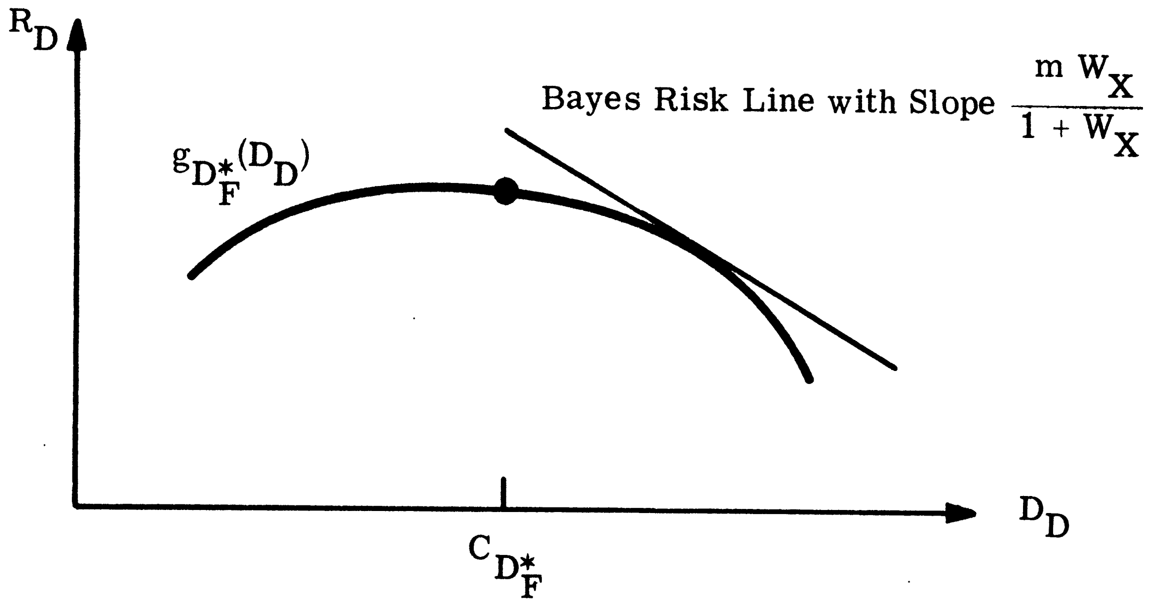


Fig. 3.16. The relation of the Bayes risk line with slope $W_X m/(1+W_X)$ to the section $g_{D_F^*}(D_D)$

points on the respond-and-hold ROC surface and, for any point on $\lambda_{\tilde{\mathscr{P}}}$, there is a respond-and-hold Bayes rule with that performance point.

Next, we verify conjecture V of Section 1.4. This conjecture says that as W_X increases, the number of detections increases but the detection time decreases. This is stated formally by the following theorem.

Theorem 3.13. If δ' and δ'' are two Bayes rules corresponding to the losses W_X' and W_X'' respectively and such that $D_F(\delta') = D_F(\delta'') = D_F^*$, then

$$W_X' > W_X''$$

implies

$$R_D(\delta') \geq R_D(\delta'')$$

and

$$D_D(\delta') \leq D_D(\delta'')$$

As with Theorem 3.12, we will reason from the geometrical relation between the section $g_{D_F^*}(D_D)$ and the Bayes risk line. This relation is shown in Fig. 3.17 below.

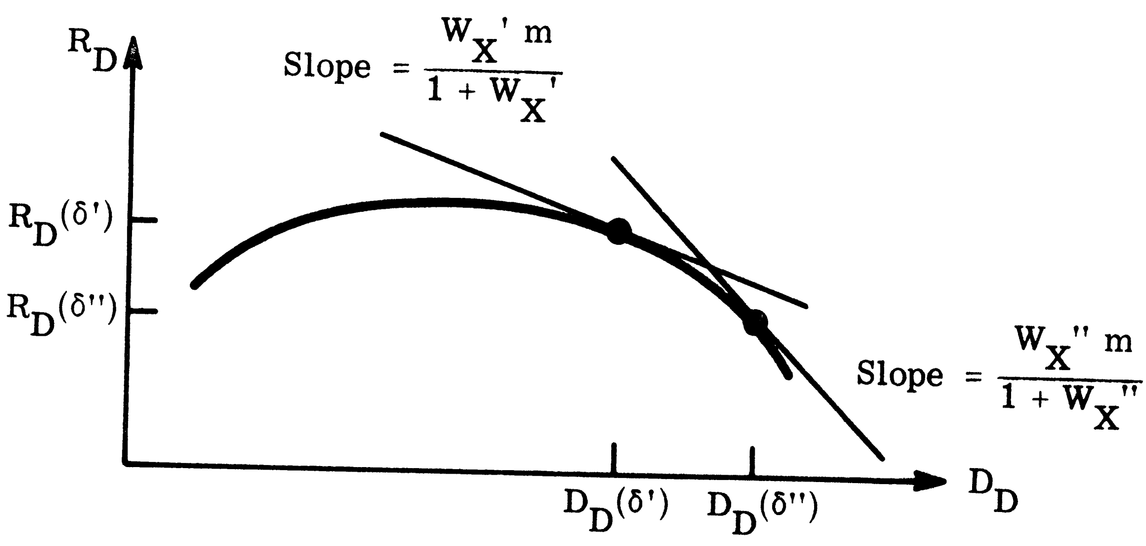


Fig. 3.17. The intersection of the Bayes risk lines with the section $g_{D_F^*}(D_D)$

Here, it is seen that, since $0 > W_X' > W_X'' \geq -1$ implies that the slopes of the Bayes risk lines satisfy

$$0 > \frac{W_X' m}{1 + W_X'} > \frac{W_X'' m}{1 + W_X''} \geq -\infty$$

then the intersection point corresponding to W_X'' must be to the right of the intersection point corresponding to W_X'. Thus, it is clear from Fig. 3.17 that

$$R_D(\delta') \geq R_D(\delta'')$$

and

$$D_D(\delta') \leq D_D(\delta'')$$

Moreover, it is to be noted that the maximum value of D_D and the minimum value of R_D for $D_F = D_F^*$ will occur for $W_X = -1$ since then the slope is $-\infty$.

This section is concluded by summarizing the results of the above theorems. The main ideas of these theorems are contained in Fig. 3.18.

In the diagram of Fig. 3.18 the region

$$W_X \geq -1 \qquad W_F > 0$$

has been divided into five special cases. The first three cases

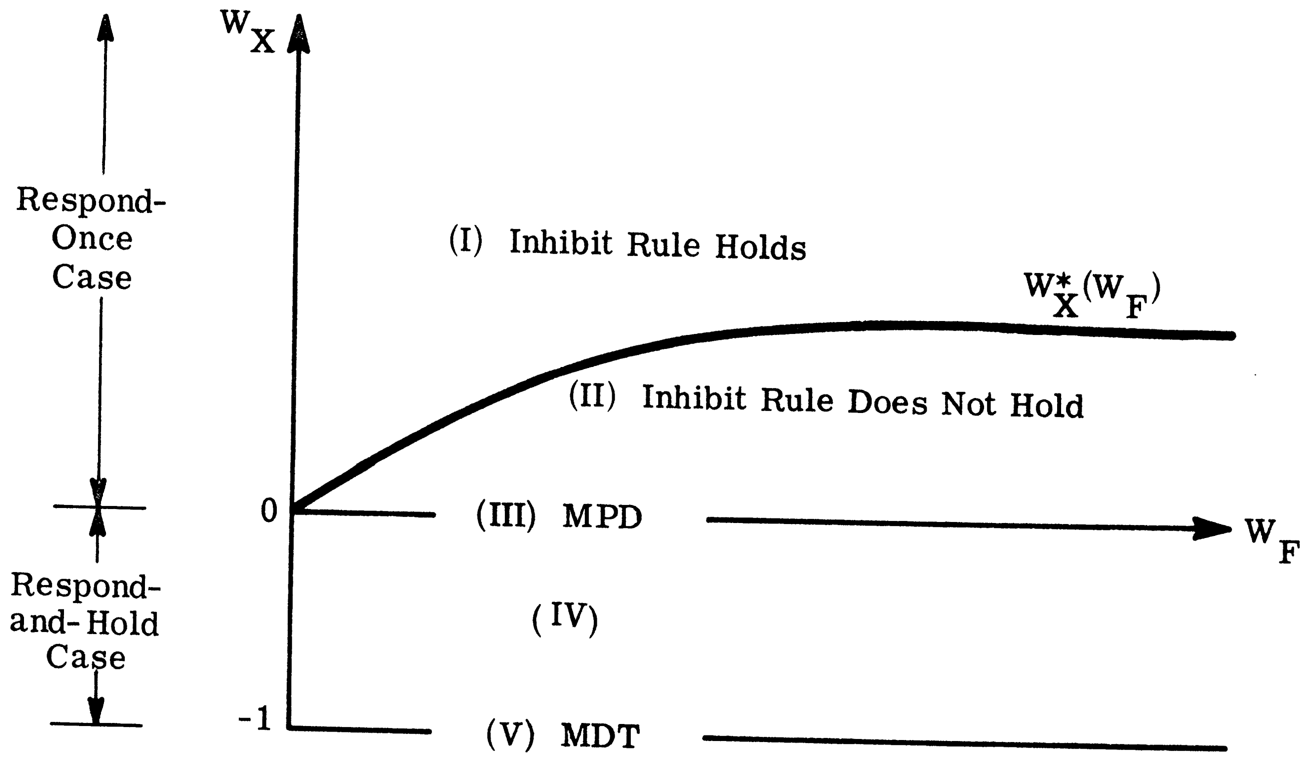


Fig. 3.18. The relation of the losses W_X, W_F to the behavior of the Bayes decision rules

correspond to the general respond-once decision device and the last two correspond to the general respond-and-hold decision devices. The important characteristics of these special cases are as follows.

$$\underline{\text{Case I}} \qquad W_X \geq W_X^*(W_F), \qquad W_F > 0$$

The Bayes decision devices associated with losses in this region satisfy the inhibit rule and, as a consequence, they depend only on the value of W_F. Their performance is characterized by the fact that they attain the lowest values of R_D and R_X for a fixed value of R_F .

$$\underline{\text{Case II}} \quad 0 < W_X < W_X^*(W_F) \qquad W_F > 0$$

The Bayes decision devices associated with these losses do not satisfy the inhibit rule and therefore they depend on both the losses W_X and W_F . Their performance is characterized by the fact that an increase in R_D can be obtained at the expense of an increase in R_X for constant R_F . This takes place for decreasing values of W_X .

$$\underline{\text{Case III}} \quad W_X = 0 \qquad W_F > 0$$

This set of losses defines the MPD decision devices. These decision devices also do not satisfy the inhibit rule and their structure will depend on both W_X and W_F . The singular importance of these devices lies in the fact that they achieve the highest possible R_D for a fixed R_F . It might also be noted that on the basis of the numerical calculations of Chapter 5, the structure of these decision devices is unique among the different cases.

$$\underline{\text{Case IV}} \quad -1 < W_X < 0 \qquad W_F > 0$$

The Bayes decision devices associated with these losses are respond-and-hold devices. Their performance is characterized by the fact that increases in R_D are obtained at the expense of decreases in D_D for fixed values of R_F. These increases occur with increasing values of W_X.

<u>Case V</u> $W_X = -1$ $W_F > 0$

The Bayes decision devices defined by these losses are the MDT devices. Their special significance lies in the fact that these devices achieve the maximum D_D for a fixed D_F. In the next chapter, it will be shown that these devices also have a unique structure.

CHAPTER IV

THE BAYES DECISION RULES FOR THE FRD PROBLEM

In Chapter II the FRD problem is cast in terms of the general decision model. In this chapter, we use this model to develop a recursive procedure for obtaining the Bayes decision rules. Then, the resulting equations are used to establish a general model for the Bayes decision device. Next, we investigate two special cases for which the recursive procedure is easily implemented. Finally, the additional assumption of conditionally independent observations is applied in order to simplify the general decision device model.

4.1 The Bayes Decision Rule

In Section 2.1, a Bayes decision rule was defined as any decision rule δ^o satisfying

$$r(\delta^o) = \inf_{\delta \in \mathcal{D}} r(\delta)$$

where

$$r(\delta) = E_{x,\theta} [L(\delta(x), \theta)]$$

The philosophy for minimizing $r(\delta)$ to obtain $\delta^o = (\delta_1^{\ o}, \delta_2^{\ o}, \ldots, \delta_N^{\ o})$ is well known. Specifically, $r(\delta)$ is minimized over the components δ_k one at a time starting with δ_N and proceeding through decreasing k. The result is a recursive scheme for obtaining $\delta_k^{\,o}$ in terms of δ_{k+1}. The formal derivation of this procedure, however, is mechanical in nature and as such it sheds little light on the character of the Bayes rule itself. For this reason, we will content ourselves with a heuristic argument here, leaving the formal derivation to Appendix E.

As a preliminary, we will investigate in further detail the most recent response function τ_k of Section 2.5. It will be shown that τ_k can be written as a function of τ_{k-1} and the preceding decision a_{k-1}. Specifically,

$$\tau_k = \begin{cases} T_p - \Delta & a_{k-1} = 1 \\ \max[0, \tau_{k-1} - \Delta] & a_{k-1} = 0 \end{cases} \tag{4.1}$$

We proceed as follows. Recall from Section 2.5 that the most recent responses function τ_k is given by

$$\tau_k = \begin{cases} t_{r_\ell} - (t_k - T_p) & \text{if the most recent response occurs at time } t_{r_\ell} \in (t_k - T_p,\ t_k) \\ 0 & \text{if no response occurs in } (t_k - T_p,\ t_k) \end{cases}$$

Suppose first that $a_{k-1} = 1$, so that a response occurred at time t_{k-1}. Then, $\tau_k = t_{k-1} - (t_k - T_p) = T_p - \Delta$. This verifies Eq. 4.1

when $a_{k-1} = 1$. Next, suppose that $a_{k-1} = 0$. There are two cases, $\tau_{k-1} \leq \Delta$ and $\tau_{k-1} > \Delta$. If $\tau_{k-1} \leq \Delta$, then it follows that no responses occurred in the interval $(t_k - T_p, t_{k-1})$. Thus, in this case, $\tau_k = 0$. On the other hand, if $\tau_{k-1} > \Delta$, then from Fig. 4.1 it is seen that the most recent response occurred at time

$$t_{r_\ell} \in (t_k - T_p, t_{k-1})$$

and that

$$\tau_k = \tau_{k-1} - \Delta$$

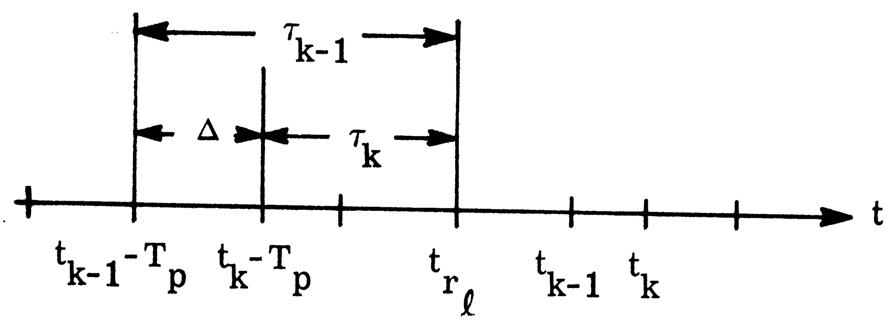


Fig. 4.1. The relation of τ_{i-1} to τ_i when $a_i = 0$ and $\tau_i > \Delta$

Thus, for either of the two cases associated with the condition $a_{k-1} = 0$, τ_k can be written as

$$\tau_k = \max[0, \tau_{k-1} - \Delta]$$

This completes the verification of Eq. 4.1.

The computational procedure for determining the Bayes decision rule will now be developed. To begin, we fix the decision time $t_k \epsilon \mathscr{T}$ and consider the dependency of the total loss $L(\vec{a}, \vec{\theta})$ on the decision a_k. First, recall from Eq. 2.27 that the total loss is given by

$$L(\vec{a}, \vec{\theta}) = \sum_{k=1}^{N} L_k(\tau_k, a_k, \theta_k)$$

where $L_k(\tau_k, a_k, \theta_k)$ represents the loss arising from the current decision a_k when the most recent response is given by τ_k and the pulse configuration in the interval $[t_k - T_p, t_k)$ is given by θ_k. Clearly, the total loss $L(\vec{a}, \vec{\theta})$ depends on the decision a_k through the "immediate" loss $L_k(\tau_k, a_k, \theta_k)$. But now consider the situation at time t_{k+1}. At this time, a decision a_{k+1} is made and the loss $L_{k+1}(\tau_{k+1}, a_{k+1}, \theta_{k+1})$ is incurred. But, at time t_{k+1}, the value of the most recent response function τ_{k+1} depends on whether or not a response was made at time t_k according to

$$\tau_{k+1} = \tau(a_k, \tau_k) = \begin{cases} T_p - \Delta & \text{if} \quad a_k = 1 \\ \max[0, \tau_k - \Delta] & \text{if} \quad a_k = 0 \end{cases}$$

Thus, the loss at time t_{k+1}, $L_{k+1}(\tau_{k+1}, a_{k+1}, \theta_{k+1})$, depends on a_k.

The above reasoning can be extended to any future decision time t_{k+j}, j=1, ..., N-k, by noting that τ_{k+j} depends on all of the preceding decisions and hence on the decision a_k. Thus, we may conclude that the decision a_k affects the total loss $L(a, \theta)$ through each of the functions

$$L_{k+j}(\tau_{k+j}, a_{k+j}, \theta_{k+j}) \qquad j = 0, 1, \ldots, N-k \qquad (4.2)$$

As a concrete example of this dependency, consider the situation shown in Fig. 4.2.

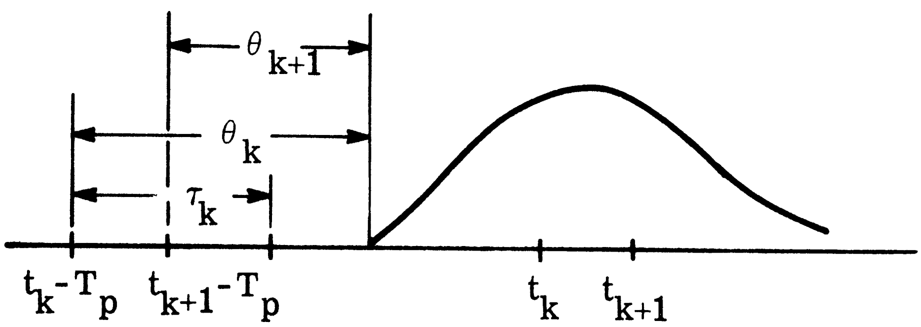


Fig. 4.2. An example of the effect of the decision at time t_k on the decision at time t_{k+1}

Here we have a pulse that is present at both the times t_k and t_{k+1}. As shown, $\tau_k = 2\Delta$ which is less than θ_k so that at time t_k the pulse has not been detected.

First, suppose that the decision device responds at t_k. Then a detection results, so that

$$L_k(\tau_k = 2\Delta,\ a_k = 1,\ \theta_k > \tau_k) = L_D$$

Then at time t_{k+1}, $\tau_{k+1} = T_p - \Delta$ which is greater than θ_{k+1} so that the loss at time t_{k+1} is

$$L_{k+1}(\tau_{k+1} = T_p - \Delta,\ a_{k+1},\ \theta_{k+1} < \tau_{k+1}) = \begin{cases} L_X & a_{k+1} = 1 \\ 0 & a_{k+1} = 0 \end{cases}$$

On the other hand, if the decision device does not respond at time t_k, then $a_k = 0$, and

$$L_k(\tau_k = 2\Delta,\ a_k = 0,\ \theta_k > \tau_k) = 0$$

But, $\tau_{k+1} = \max[0,\ \tau_k - \Delta] = \Delta$ which is less than θ_{k+1} so that the loss at time t_{k+1} is given by

$$L_{k+1}(\tau_{k+1} = \Delta,\ a_{k+1},\ \theta_{k+1} > \tau_{k+1}) = \begin{cases} L_D & \text{if } a_{k+1} = 1 \\ 0 & \text{if } a_{k+1} = 0 \end{cases}$$

The next step in the derivation of the Bayes rule δ^o is to determine an expression for the posterior risk associated with the decision a_k. This risk, to be denoted hereafter by $R_k(\vec{x}_k, \tau_k, a_k)$, is the posterior average of the loss due to the decision a_k. We can immediately write down the contribution to $R_k(\vec{x}_k, \tau_k, a_k)$ due to the immediate loss $L_k(\cdot, \cdot, \cdot)$. Specifically,

$$\underset{\theta_k | \vec{x}_k}{E} L_k(\tau_k, a_k, \theta_k)$$

Now, at time t_{k+1}, the posterior risk associated with the loss $L_{k+1}(\cdot, \cdot, \cdot)$ is

$$\underset{\theta_{k+1} | \vec{x}_{k+1}}{E} L_{k+1}(\tau_{k+1}, a_{k+1}, \theta_{k+1})$$

But, at time t_k only the observation up to time t_k is known. Thus, we must average this term with respect to the "unobserved" observation x_{k+1}, conditioned on the observed value $\vec{x}_k$. The resulting contribution to $R_k(\tau_k, a_k)$ is

$$\underset{x_{k+1} | \vec{x}_k}{E} \left\{ \underset{\theta_{k+1} | \vec{x}_{k+1}}{E} \{L_{k+1}(\tau_{k+1}, a_{k+1}, \theta_{k+1})\} \right\}$$

Similarly, the contribution to the posterior risk of the decision a_k due to the loss at time t_{k+j}, $j = 1, \ldots, N-k$, is

$$\underset{x_{k+1}, \ldots, x_{k+j} | \vec{x}_k}{E} \left\{ \underset{\theta_{k+j} | \vec{x}_{k+j}}{E} \{L_{k+i}(\tau_{k+j}, a_{k+j}, \theta_{k+j})\} \right\}$$

Adding these contributions gives, for the total posterior risk of the decision a_k,

$$R_k(\vec{x}_k, \tau_k, a_k) \tag{4.3}$$

$$= \begin{cases} \underset{\theta_k | \vec{x}_k}{E} \{L_k(\tau_k, a_k, \theta_k)\} \\ + \sum_{j=1}^{N-k} \underset{x_{k+1}, \ldots, x | \vec{x}_k}{E} \left\{ \underset{\theta_{k+j} | \vec{x}_{k+j}}{E} \{L_{k+j}(\tau_{k+j}, a_{k+j}, \theta_{k+j})\} \right\} & k < N \\ \underset{\theta_N | \vec{x}_N}{E} \{L_N(\tau_N, a_N, \theta_N)\} & k = N \end{cases}$$

To complete the derivation of the Bayes decision rule

$$\delta^o = (\delta_1^o, \ldots, \delta_N^o)$$

we apply the well known principle that δ^o is the function which minimizes the posterior risk of the decision a_k. In the problem at hand, a_k takes on one of two values, $a_k = 1$, denoting a response at time t_k, and $a_k = 0$, denoting a no-response decision at time t_k. Thus, the posterior risk of the decision a_k is minimized by the decision device that responds at time t_k iff the posterior risk of responding, $R_k(\vec{x}_k, \tau_k, a_k{=}1)$, is less than or equal to the posterior risk of not responding, $R_k(\vec{x}_k, \tau_k, a_k{=}0)$. Stated formally,

$$\delta_k^o(\tau_k, \vec{x}_k) = \begin{cases} 1 & \text{if } R_k(\vec{x}_k, \tau_k, a_k{=}1) \le R_k(\vec{x}_k, \tau_k, a_k{=}0) \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

It is to be noted in connection with Eqs. 4.3 and 4.4 that since $R_k(\vec{x}_k, \tau_k, a_k)$ depends on the decisions

$$a_{k+j} = \delta^o_{k+j}(\tau_{k+j}, \vec{x}_{k+j}) \qquad j = 1, \ldots, N-k$$

the complete Bayes rule $\delta^o = (\delta_1{}^o, \ldots, \delta_N{}^o)$ must be obtained recursively by solving Eq. 4.4 for $\delta_N{}^o$, then δ^o_{N-1}, and so on.

An alternative expression for the Bayes decision rule can be obtained in the following way. For each $\tau_k \epsilon \{0, \ldots, m-1\}$ and each $\vec{x}_{k-1} \epsilon \vec{X}_{k-1}$ define the set of current observation $A_k(\vec{x}_{k-1}, \tau_k) \subset X_k$ by

$$A_k(\vec{x}_{k-1}, \tau_k) = \{x_k ; T_k(\vec{x}_k, \tau_k) \geq 0\} \tag{4.5}$$

where

$$T_k(\vec{x}_k, \tau_k) = R_k(\vec{x}_k, \tau_k, a_k{=}0) - R_k(\vec{x}_k, \tau_k, a_k{=}1) \tag{4.6}$$

The set $A_k(\vec{x}_{k-1}, \tau_k)$ is the response set for $\delta^o_k(\vec{x}_k, \tau_k)$ (see Section 3.3), and the function $T_k(\vec{x}_k, \tau_k)$ is the net posterior risk of not responding at time t_k. In terms of $A_k(\cdot, \cdot)$ and $T_k(\cdot, \cdot)$, we can write δ^o_k either in the form

$$\delta^o_k(\vec{x}_k, \tau_k) = \begin{cases} 1 & x_k \epsilon A_k(\vec{x}_{k-1}, \tau_k) \\ 0 & x_k \notin A_k(\vec{x}_{k-1}, \tau_k) \end{cases} \tag{4.7a}$$

or, in the form

$$\delta_k^o(\vec{x}_k, \tau_k) = \begin{cases} 1 & T_k(\vec{x}_k, \tau_k) \geq 0 \\ 0 & T_k(\vec{x}_k, \tau_k) < 0 \end{cases} \tag{4.7b}$$

It is to be noted from Eq. 4.7b that the function $T_k(\vec{x}_k, \tau_k)$ plays the same role as the test function of Section 1.4.

Our next task is to obtain a block diagram for the Bayes decision rule. To this end, we appeal to Eqs. 4.7b, 4.6, 4.3 and 4.1 and conclude that the configuration of the general Bayes decision device is as illustrated in Fig. 4.3. The operation of this device is as follows. At time t_{k-1}, the decision device has observed the observation $\vec{x}_{k-1}$ and made some decision a_{k-1}. The observation $\vec{x}_{k-1}$ is then stored in memory so that when the current observation x_k is observed, the decision device will have available the total past observation $\vec{x}_k = (\vec{x}_{k-1}, x_k)$. Simultaneously, the decision a_{k-1} is used to calculate τ_k according to

$$\tau_k = \begin{cases} T_p - \Delta & a_{k-1} = 1 \\ \max[0, \tau_{k-1} - \Delta] & a_{k-1} = 0 \end{cases}$$

The value of τ_k is then stored in a memory so that it too is available at time t_k. Finally, at time t_k, the decision device calculates $T_k(\vec{x}_k, \tau_k)$ according to Eqs. 4.6 and 4.3 and responds iff

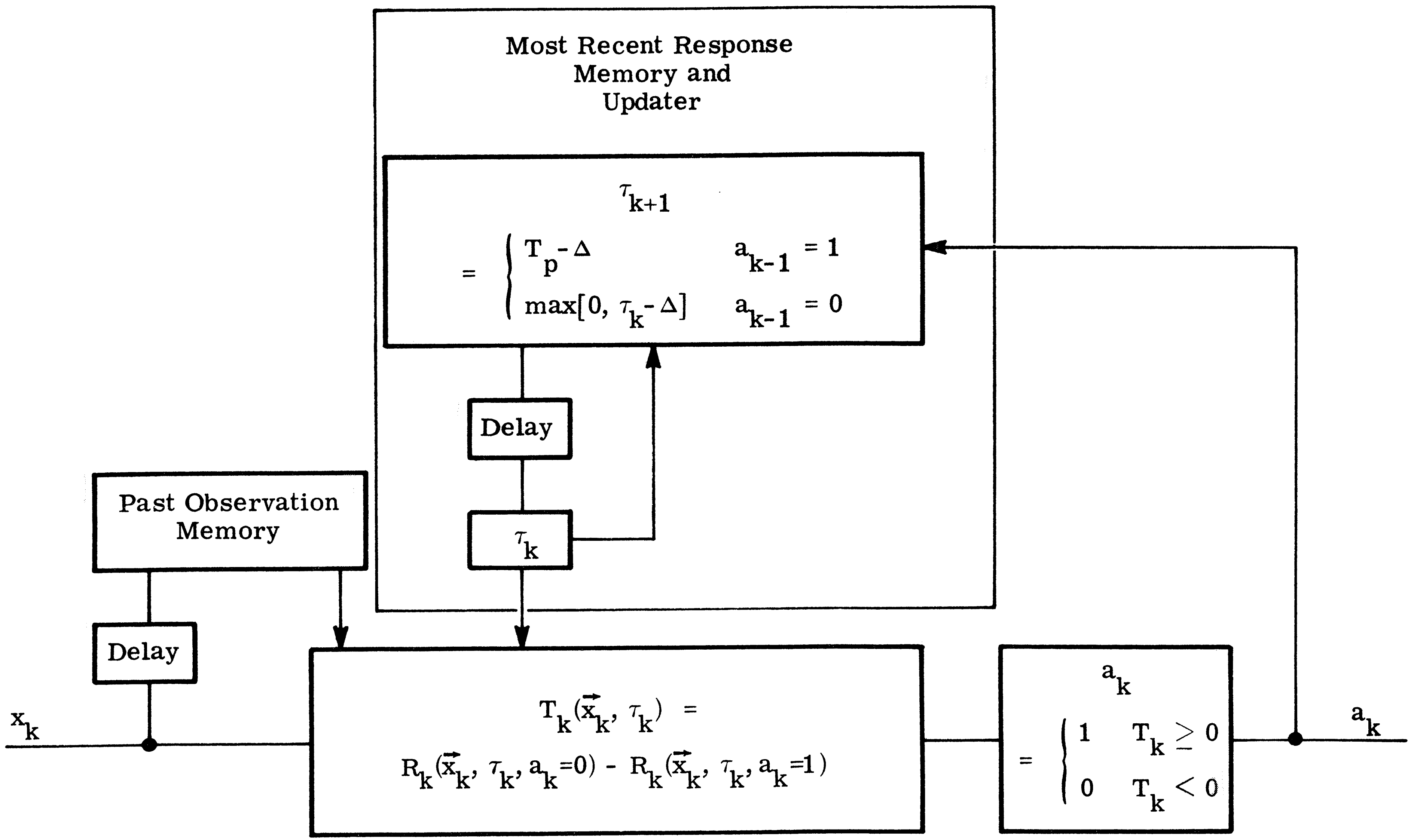


Fig. 4.3. The general Bayes decision device

$$T_k(\vec{x}_k, \tau_k) \geq 0$$

Theoretically, the general Bayes decision device is completely determined by the block diagram of Fig. 4.3. In practice, however, this device is difficult to implement for at least two reasons. First, since the past observation memory must store the total past observation from time $t = 0$ up to the current time $t = t_k$, its capacity must grow linearly in time. Thus, it is not practical to think of allowing this decision device to operate indefinitely.

The second difficulty in implementing the general Bayes decision device arises in the calculation of $T_k(\vec{x}_k, \tau_k)$. To see where this problem lies, replace the expected value operators in Eq. 4.3 by integrals with respect to the appropriate probability measures and note that

$$R_{k+1}(\vec{x}_{k+1}, a_{k+1}, \tau_{k+1}(\tau_k, a_k))$$

$$= \begin{cases} R_{k+1}(\vec{x}_{k+1}, 1, \tau_{k+1}(\tau_k, a_k)) & \text{for } x_{k+1} \in {}^1A_{k+1}(\vec{x}_k, \tau_{k+1}) \\ R_{k+1}(\vec{x}_{k+1}, 0, \tau_{k+1}(\tau_k, a_k)) & \text{for } x_{k+1} \in {}^0A_{k+1}(\vec{x}_k, \tau_k) \end{cases}$$

where

$${}^tA_{k+1}(\vec{x}_k, \tau_{k+1}) = \begin{cases} X_{k+1} \setminus A_{k+1}(\vec{x}_k, \tau_{k+1}) & t = 1 \\ A_{k+1}(\vec{x}_k, \tau_{k+1}) & t = 0 \end{cases}$$

The resulting expression for $R_{k+1}(\cdot, \cdot, \cdot)$ is

$$R_k(\vec{x}_k, \tau_k, a_k) \tag{4.8}$$

$$= \begin{cases} \int_{\Theta_k} L_k(\tau_k, a_k, \theta_k)\, \Pi(d\theta_k | \vec{x}_k) \\ \quad + \sum_{t=0}^{1} \int_{{}^t A_{k+1}(\vec{x}_k, \tau_{k+1})} R_{k+1}(\vec{x}_{k+1}, a_{k+1}=t, \tau_{k+1}(a_k, \tau_k))\, P(dx_{k+1} | \vec{x}_k) & k < N \\ \int_{\Theta_N} L(\tau_N, a_N, \theta_N)\, \Pi(d\theta_N | \vec{x}_N) & k = N \end{cases}$$

Now, from Eq. 4.8 it is seen that in order to calculate $R_k(\vec{x}_k, \tau_k, a_k)$ and hence $T_k(\vec{x}_k, \tau_k, a_k)$, it is necessary to have available $\Pi(d\theta_k | \vec{x}_k)$ and $P(dx_{k+1} | \vec{x}_k)$. But, in the general case, these probability measures can be obtained from the known probability measures $\{P(dx_k | \theta);\ \theta \in \Theta\}$ and $\Pi(d\theta)$ only by integrating over the whole space $\vec{X} \times \Theta$. It is certainly not practical to carry out these integrations at each decision time t_k.

Fortunately, both of the above difficulties can be eliminated if we place certain restrictions on the observation probability laws. The result is that the probability measures relevant to the calculation of $T_k(\vec{x}_k, \tau_k)$ can be calculated from the probability measures relevant to the calculation of $T_{k-1}(\vec{x}_{k-1}, \tau_{k-1})$ along with the current observation x_k. When this is the case, the computational procedure is often referred to as an updating procedure. In Sections 4.3 and 4.4 we will investigate such a procedure. In the next section, two special cases of the general Bayes decision device are examined.

4.2 Two Special Cases: The Maximum Detection Time Decision Device and the $m = 1$ Decision Device

In Section 4.1 it was seen that at time t_k the general Bayes decision device must calculate $T_k(\vec{x}_k, \tau_k)$ according to Eqs. 4.6 and 4.8. But, these equations involve the optimum component decision rules at time t_{k+1} through the response set $A_{k+1}(\cdot, \cdot)$. Thus, to get an explicit expression for $T_k(\vec{x}_k, \tau_k)$, one must first solve for $T_N(\vec{x}_N, \tau_N)$ and then proceed recursively to solve for $T_{N-1}(\vec{x}_{N-1}, \tau_{N-1})$, $\ldots$, $T_{k+1}(\vec{x}_{k+1}, \tau_{k+1})$, $T_k(\vec{x}_k, \tau_k)$. There are two special cases, however, where the solution of this recursive procedure can be obtained easily. In this section, we treat these cases.

Case 1. The Maximum Detection Time (MDT) decision device.

The MDT decision device is a respond-and-hold decision device that is characterized by a loss function that depends only on the amount of detection time and false alarm time. In Chapter III it was seen that among all R-H decision devices, this device maximizes the detection duty D_D for a fixed false alarm duty D_F. Thus, the structure of the MDT decision device is of particular interest. This structure is given explicitly in the following theorem.

Theorem 4.1. The test function for the MDT decision device is given by

$$T_k(\vec{x}_k, \tau_k) = T_k(\vec{x}_k) \tag{4.9}$$

$$= \int_{\theta_k \epsilon [0, T_p)} \Pi(d\theta | \vec{x}_k) - W_F \int_{\theta_k = T_p} \Pi(d\theta_k | \vec{x}_k)$$

Proof. We will show that Eq. 4.9 holds by also showing that

$$R_k(\vec{x}_k, \tau_k, a_k) = \mathop{E}_{\theta_k | \vec{x}_k} L_k(\tau_k, a_k, \theta_k) + f_k(\vec{x}_k) \tag{4.10}$$

where $f_k(\vec{x}_k)$ does not depend on τ_k. The proof is by induction on the index k.

To begin, substitute for the loss function from Eq. 2.27 of Section 2.5 to conclude that

$$\mathop{E}_{\theta_k | \vec{x}_k} L_k(\tau_k, a_k, \theta_k)$$

$$= \begin{cases} 0 & a_k = 0 \\ W_X \int_{\{\theta_k < \tau_k, \, \theta_k \epsilon [0, T_p)\}} \Pi(d\theta_k | \vec{x}_k) - \int_{\{\theta_k \geq \tau_k, \, \theta_k \epsilon [0, T_p)\}} \Pi(d\theta_k | \vec{x}_k) & \\ \qquad + W_F \int_{\{\theta_k = T_p\}} \Pi(d\theta_k | \vec{x}_k) & a_k = 1 \end{cases}$$

Next, combine the first two integrals on the right hand side of the above expression into a single integral by using the facts that $W_X = -1$ for the MDT loss function and that

$$\{\theta_k \geq \tau_k,\ \theta_k \in [0, T_p)\} \cup \{\theta_k < \tau_k,\ \theta_k \in [0, T_p)\}$$
$$= \{\theta_k \in [0, T_p)\}$$

The result is

$$\underset{\theta_k | \vec{x}_k}{E} L_k(\tau_k, a_k, \theta_k) \tag{4.11}$$

$$= \begin{cases} 0 & a_k = 0 \\ -\left[\int_{\theta_k \in [0, T_p)} \Pi(d\theta_k | \vec{x}_k) - W_F \int_{\{\theta_k = T_p\}} \Pi(d\theta_k | \vec{x}_k) \right] & a_k = 1 \end{cases}$$

Now, for $k = N$ we note from Eq. 4.3 that Eq. 4.10 is satisfied with $f_N(\vec{x}_N) = 0$. Moreover,

$$T_N(\vec{x}_N, \tau_N) = R_N(\vec{x}_N, \tau_N, a_N{=}0) - R_N(\vec{x}_N, \tau_N, a_N{=}1)$$
$$= \int_{\{\theta_N \in [0, T_p)\}} \Pi(d\theta_N | \vec{x}_N) - W_F \int_{\theta_N = T_p} \Pi(d\theta_N | \vec{x}_N)$$

so that Eq. 4.9 holds for $k = N$.

Now, suppose that Eq. 4.9 and Eq. 4.10 are satisfied for indices $k+1, \ldots, N$. Then, $T_{k+1}(\vec{x}_{k+1}, \tau_{k+1})$ does not depend on

τ_{k+1}. Thus, from Eq. 4.5,

$$A_{k+1}(\vec{x}_k, \tau_{k+1}) = \{x_{k+1} ; T_{k+1}(\vec{x}_k, \tau_{k+1}) \geq 0\}$$

$$= \{x_{k+1} ; T_{k+1}(\vec{x}_k) \geq 0\}$$

or

$$A_{k+1}(\vec{x}_k, \tau_{k+1}) = A_{k+1}(\vec{x}_k) \tag{4.12}$$

Therefore, the response set A_{k+1} does not depend on τ_{k+1}. Next, substitute Eq. 4.11 into Eq. 4.10 to obtain

$$R_{k+1}(\vec{x}_{k+1}, \tau_{k+1}, a_{k+1})$$

$$= \begin{cases} f_{k+1}(\vec{x}_{k+1}) & a_{k+1} = 0 \\ - \left[\int\limits_{\theta_{k+1} \epsilon [0, T_p)} \Pi(d\theta_{k+1} | \vec{x}_{k+1}) - W_F \int\limits_{\theta_{k+1} = T_p} \Pi(d\theta_k | x_k) \right] + f_{k+1}(\vec{x}_{k+1}) & a_{k+1} = 1 \end{cases} \tag{4.13}$$

Then, substitute for R_{k+1} from Eq. 4.13 and for A_{k+1} from Eq. 4.12 into Eq. 4.8. The result is

$$R_k(\vec{x}_k, \tau_k, a_k)$$

$$= \int_{\Theta_k} L_k(\tau_k, a_k, \theta_k)\, \Pi(d\theta_k | \vec{x}_k)$$

$$+ \left\{ \int_{{}_0A_{k+1}(\vec{x}_k)} f_{k+1}(\vec{x}_{k+1})\, P(dx_{k+1} | \vec{x}_k) \right.$$

$$+ \int_{{}_1A_{k+1}(\vec{x}_k)} \left[- \left[\int_{\theta_{k+1} \epsilon [0, T_p)} \Pi(d\theta_{k+1} | \vec{x}_{k+1}) - W_F \int_{\theta_{k+1} = T_p} \Pi(d\theta_{k+1} | \vec{x}_{k+1}) \right] + f_{k+1}(\vec{x}_{k+1}) \right]$$

$$\left. \times\ P(dx_{k+1} | \vec{x}_k) \right\}$$

or

$$R_k(\vec{x}_k, \tau_k, a_k) = \underset{\theta_k | \vec{x}_k}{E}\ L_k(\tau_k, a_k, \theta_k) + f_k(\vec{x}_k)$$

where we have set the bracketed term equal to $f_k(\vec{x}_k)$, since it depends only on $\vec{x}_k$. Thus we have established Eq. 4.10 for the index k. The proof is completed by substituting for

$$\underset{\theta_k | \vec{x}_k}{E}\ L_k(\tau_k, a_k, \theta_k)$$

from Eq. 4.11 into Eq. 4.10 and then substituting into Eq. 4.6. The result is Eq. 4.9 as was to be shown.

On the basis of the above theorem we can simplify the block diagram for the MDT decision device to that shown in Fig. 4.4.

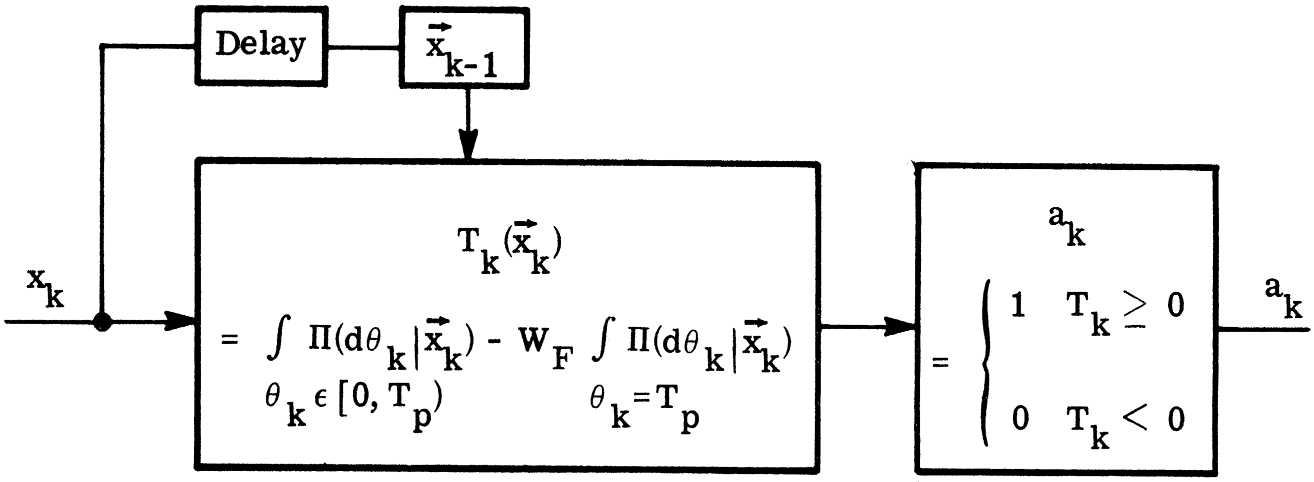


Fig. 4.4. The MDT decision device

Several properties of the MDT decision device might be pointed out here. First, it is no longer necessary to provide a decision feedback loop to calculate τ_k, since τ_k does not appear in the test function T_k. Thus, we have verified Conjecture IV of Section 1.4. Secondly, it is noted that the test function T_k no longer depends on the distribution $P(dx_{k+1} | \vec{x}_k)$. Thus the only distribution that is involved in the decision at time t_k is the posterior distribution $\Pi(d\theta_k | \vec{x}_k)$.

The single most important property of the MDT decision device, however, is contained in the following corollary to Theorem 4.1.

Corollary 4.2. The MDT decision device may be considered as the optimum detector, in the sense of classical detection theory, for detecting the arrival of a pulse within the last T_p seconds.

Proof. In Appendix A it is seen that, in general, the optimum detector responds iff

$$\int_{\Theta_{SN}} \Pi(d\theta \mid x) \geq W_F \int_{\Theta_N} \Pi(d\theta \mid x)$$

where Θ_{SN} are those parameters to be associated with a response decision and Θ_N are those parameters to be associated with a no-response decision. Now, from Section 2.3 it is clear that if we wish the optimum detector to respond at time t_k iff a pulse has arrived in the last T_p seconds, then we must take

$$\begin{aligned}\Theta_{SN} &= \{\vec{\theta} = (\theta_1, \ldots, \theta_k) ; \theta_k \epsilon [0, T_p)\} \\ &= \Theta_1 \times \ldots \times \Theta_{k-1} \times [0, T_p)\end{aligned}$$

and

$$\begin{aligned}\Theta_N &= \{\vec{\theta} = (\theta_1, \ldots, \theta_k) ; \theta_k = T_p\} \\ &= \Theta_1 \times \ldots \times \Theta_k \times \{T_p\}\end{aligned}$$

Also, in the notation of Section 2.4,

$$x = \vec{x}_k$$

Thus, we can write the response condition for the optimum detector as

$$\int_{\Theta_1 \times \ldots \times \Theta_{k-1} \times [0, T_p)} \Pi(d\theta_k, \ldots, d\theta_1 | \vec{x}_k) \geq W_F \int_{\Theta_1 \times \ldots \times \Theta_{k-1} \times \{T_p\}} \Pi(d\theta_k, \ldots, d\theta_1 | \vec{x}_k)$$

or, equivalently, as

$$\int_{[0, T_p)} \Pi(d\theta_k | \vec{x}_k) \geq W_F \int_{\{T_p\}} \Pi(d\theta_k | \vec{x}_k)$$

But, this is precisely the same response condition as the response condition

$$T_k(\vec{x}_k) \geq 0$$

where $T_k(\vec{x}_k)$ is given in Eq. 4.2. Thus, the optimum detector from classical detection theory is equivalent to the MDT decision device. This completes the proof.

In the above paragraphs we have seen that, because of the special nature of the MDT loss function, the structure of the decision device is considerably simplified. This also turns out to be the case for the Bayes decision device that makes decisions at intervals of T_p seconds.

m = 1 Case

The quantity m is the number of decision opportunities per pulse. In Section 3.5 it was seen that, in general, the performance of FRD decision devices increases with increasing m . Thus, the case m = 1 represents a lower bound on the performance of the Bayes decision devices.

The structure of the m = 1 Bayes decision device is given in the following theorem.

Theorem 4.3. The test function for the m = 1 Bayes decision device is given by Eq. 4.9.

Proof. For m = 1, $\Delta = T_p$, so that the value of the most recent response function τ_k is given by

$$\tau_k = \begin{cases} T_p - \Delta = T_p - T_p = 0 & a_{k-1} = 1 \\ \max[0, \tau_{k-1} - \Delta] = \max[0, \tau_{k-1} - T_p] = 0 & a_{k-1} = 0 \end{cases}$$

Now, substitute $\tau_k = 0$ into Eq. 2.27 for the loss function to conclude that

$$\underset{\theta_k|\vec{x}_k}{E} L_k(\tau_k, a_k, \theta_k)$$

$$= \begin{cases} 0 & a_k = 0 \\ -\left[\int\limits_{\theta_k \in [0, T_p)} \Pi(d\theta_k|\vec{x}_k) - W_F \int\limits_{\theta_k = T_p} \Pi(d\theta_k|\vec{x}_k) \right] & a_k = 1 \end{cases}$$

But, this is precisely the same expression as Eq. 4.11 in the proof to Theorem 4.1. Thus the conclusion to Theorem 4.3 follows by the same inductive argument.

The remarks following Theorem 4.1 also apply to the $m = 1$ case. The block diagram of Fig. 4.4 also describes the structure of the $m = 1$ decision device and the interpretation of this device in terms of classical detection theory is also appropriate. It is emphasized, however, that Theorem 4.1 applies to a specific loss function (the MDT loss function) and for any $m \geq 1$, whereas Theorem 4.3 applies to any FRD loss function so long as $m = 1$.

4.3 The Updating Equation for Conditionally Independent Observations

In Section 4.1 we discussed the desirability of obtaining an updating procedure for the general Bayes decision device. In this section the basic updating equation is derived. In the next section we will use this equation to simplify the general Bayes decision device.

Generally speaking, updating procedures do not exist without certain restrictions on the prior probability measure $\Pi(d\theta)$ and the observation measures $\{P(dx|\theta); \theta \epsilon \Theta\}$. One set of restrictions that yields tractable results requires that $\Pi(d\theta)$ be Markov and that $\{P(dx|\theta); \theta \epsilon \Theta\}$ be conditionally independent. Now, in Chapter II it was seen that $\Pi(d\theta)$ satisfies the Markov property,

$$\Pi(d\theta_k|\theta_{k-1}, \dots, \theta_1) = \Pi(d\theta_k|\theta_{k-1})$$

Thus, we only need the assumption of conditionally independent observations to get a tractable updating procedure.

In the signal plus noise setting of Section 2.4, the assumption of conditionally independent observations is really an assumption on the independence of the noise in different observation intervals. This is seen as follows. Recall from Section 2.4 that the observation x_k depends only on the noise in the interval $[t_{k-1}, t_k)$ and on the signal parameters θ_k and θ_{k-1}. Now, suppose that the noise in the interval $[t_{k-1}, t_k)$ can be considered as being independent of the noise in the interval $[t_{k-2}, t_{k-1})$. Then, it follows that for a fixed signal parameter θ, the observations x_k and x_{k-1} are independent. That is,

$$\begin{aligned} P(dx_{k-1}, dx_k|\theta) &= P(dx_{k-1}|\theta)\, P(dx_k|\theta) \\ &= P(dx_{k-1}|\theta_{k-1}, \theta_{k-2})\, P(dx_k|\theta_k, \theta_{k-1}) \end{aligned}$$

If this assumption holds for all of the intervals $[t_{k-1}, t_k)$ then the observations are said to be conditionally independent. Formally, we have the following definition.

The observations are conditionally independent iff for each $\theta \in \Theta$, and $k = 1, \ldots, N$,

$$P(dx_1, \ldots, dx_k | \theta) = \prod_{j=1}^{k} P(dx_j | \theta_j, \theta_{j-1}) \tag{4.14}$$

We now state three theorems which together provide the theoretical basis for the updating equation.

<u>Theorem 4.4.</u> If $\Pi(d\theta)$ is Markov and the observations are conditionally independent, then

$$\Pi(d\theta_k | \vec{x}_k) = \int_{\Theta_{k-1}} \frac{P(dx_k | \theta_k, \theta_{k-1})}{P(dx_k | \vec{x}_{k-1})} \Pi(d\theta_k | \theta_{k-1}) \Pi(d\theta_{k-1} | \vec{x}_{k-1}) \tag{4.15}$$

The proof of this theorem is interesting in its own right. The philosophy of the proof, however, is already established in the existing literature, most notably Ref. 9, and although there are some differences here, it is felt that this proof, along with the proofs of the remaining theorems in this section, is best left to Appendix H.

In the next theorem we replace the ratio

$$P(dx_k | \theta_k, \theta_{k-1}) / P(dx_k | \vec{x}_{k-1})$$

in Eq. 4.15 by a Raydon-Nikodym derivative.

<u>Theorem 4.5.</u> If the hypothesis of Theorem 4.4 holds, then the Raydon-Nikodym derivative

$$\ell(x_k | \vec{x}_{k-1}, \theta_k, \theta_{k-1}) = P(dx_k | \theta_k, \theta_{k-1}) / P(dx_k | \vec{x}_{k-1}) \qquad (4.16)$$

exists and

$$\Pi(d\theta_k | \vec{x}_k) = \int_{\Theta_k} \ell(x_k | \vec{x}_{k-1}, \theta_k, \theta_{k-1}) \, \Pi(d\theta_k | \theta_{k-1}) \, \Pi(d\theta_{k-1} | \vec{x}_{k-1}) \qquad (4.17)$$

Next we determine a generalized density for $\Pi(d\theta_k | \vec{x}_k)$. To do this, recall from Section 2.3 that we must first find a σ-finite measure μ that dominates $\Pi(d\theta_k | \vec{x}_k)$. (A measure μ dominates $\Pi(d\theta_k | \vec{x}_k)$ if $\Pi(d\theta_k | \vec{x}_k) << \mu(d\theta_k)$.) Now, for practical reasons, it is desired that this measure μ work for all $\vec{x}_k$. Thus, we seek a measure μ that dominates each member of the family $\{\Pi(d\theta_k | \vec{x}_k); \vec{x}_k \epsilon \vec{X}_k\}$. Such a measure is not always possible to find. As an example, it can be shown that there is no σ-finite measure μ that dominates the family $\{\Pi(d\theta_k | \theta_{k-1}) \ \theta_{k-1} \epsilon [0, T_p]\}$, when $\Pi(d\theta_k | \theta_{k-1})$ is the continuous prior distribution of Section 2.3.

Nevertheless, consider the physical situation at time t_k. The prior probability of θ_k is given by the density

$$p(\theta_k) = \begin{cases} (T_p + \alpha^{-1})^{-1} & \theta_k \in [0, T_p) \\ \alpha^{-1}(T_p + \alpha^{-1})^{-1} & \theta_k = T_p \end{cases}$$

with respect to the measure $\mu(d\theta_k) = d\theta_k + \epsilon_{\{T_p\}}(d\theta_k)$. Heuristically speaking, it seems reasonable to expect that the effect of any observation $\vec{x}_k$ on the distribution of θ_k should be to no more than "continuously" redistribute the probabilities. That is, it might be expected that $\{\Pi(d\theta_k|\vec{x}_k); \vec{x}_k \in \vec{X}_k\}$ is dominated by $\mu(d\theta_k) = d\theta_k + \epsilon_{\{T_p\}}(d\theta_k)$. A statement to this effect is provided by Theorem 4.6.

Theorem 4.6. If the hypothesis of Theorem 4.4 holds, then the family $\{\Pi(d\theta_k|\vec{x}_k); \vec{x}_k \in \vec{X}_k\}$ is dominated by the measure $\mu(d\theta_k) = \mu_{[0,T_p)}(d\theta_k) + \epsilon_{\{T_p\}}(d\theta_k)$ where

$$\mu_{[0,T_p)}(d\theta_k) = \begin{cases} d\theta_k & \text{for continuously arriving pulses} \\ \sum_{j=1}^{q-1} \epsilon_{\{j\nu\}}(d\theta_k) & \text{for pulses with discrete arrival time} \end{cases} \tag{4.18}$$

and there exists a generalized density $p(\theta_k|\vec{x}_k)$ with respect to $\mu(d\theta_k)$ such that

$$\Pi(d\theta_k|\vec{x}_k) = p(\theta_k|\vec{x}_k)\,\mu(d\theta_k) \tag{4.19}$$

An expression for $p(\theta_k|\vec{x}_k)$ in terms of $p(\theta_{k-1}|\vec{x}_{k-1})$ and $\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})$ is found in Eq. H.8 of Appendix H. This expression provides the means for updating the posterior densities $p(\theta_k|\vec{x}_k)$. The difficulty, however, is that this expression involves the function $\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})$ which depends on $\vec{x}_{k-1}$. Thus, it is still necessary to provide a memory for storing the past observation $\vec{x}_{k-1}$. In the next section, however, it will be shown that it is sufficient for the decision device to have available the function

$$O_k(\theta_k|\vec{x}_k) = \begin{cases} \begin{cases} \dfrac{p(\theta_k|\vec{x}_k)}{p(\theta_k=T_p|\vec{x}_k)} & \theta_k \in [0, T_p) \\ 1 & \theta_k = T_p \end{cases} & p(\theta_k=T_p|\vec{x}_k) \neq 0 \\ \begin{cases} p(\theta_k|\vec{x}_k) & \theta_k \in [0, T_p) \\ 0 & \theta_k = T_p \end{cases} & p(\theta_k=T_p|\vec{x}_k) = 0 \end{cases} \tag{4.20}$$

The advantage of this fact is that the function $O_k(\theta_k|\vec{x}_k)$ can be updated solely in terms of $O_k(\theta_k|\vec{x}_k)$ and x_k. Thus, the decision device need only provide a memory for $O_k(\theta_k|\vec{x}_k)$. This result is a consequence of the following theorem.

Theorem 4.7. If the hypothesis of Theorem 4.4 holds and if in addition

$$P(dx_k|\theta_k, \theta_{k-1}) << P(dx_k|\theta_k = T_p, \theta_{k-1} = T_p) \quad \text{all} \quad \theta_{k-1}, \theta_k$$

then

$$O_k(\theta_k|\vec{x}_k) =$$

$$\frac{\left[I_{\{[T_p-\Delta, T_p)}(\theta_k) \int_{\theta_{k-1} \in [0, \theta_k - (T_p - \Delta)] \cup \{T_p\}} \ell(x_k|\theta_k, \theta_{k-1})\, p(\theta_k|\theta_{k-1})\, O_{k-1}(\theta_{k-1}|\vec{x}_{k-1})\, \mu(d\theta_{k-1}) + I_{\{[0, T_p-\Delta)\}}(\theta_k)\, \ell(x_k|\theta_k, \theta_{k-1} = \theta_k + \Delta)\, O_{k-1}(\theta_{k-1} = \theta_k + \Delta|\vec{x}_{k-1}) \right]}{\int_{\theta_{k-1} \in [0, \Delta) \cup \{T_p\}} \ell(x_k|\theta_k = T_p, \theta_{k-1})\, p(\theta_k|\theta_{k-1})\, O_{k-1}(\theta_{k-1}|\vec{x}_{k-1})\, \mu(d\theta_{k-1})} \tag{4.21}$$

where

$$\ell(x_k|\theta_k, \theta_{k-1}) = \frac{P(dx_k|\theta_k, \theta_{k-1})}{P(dx_k|\theta_k = T_p, \theta_{k-1} = T_p)} \tag{4.22}$$

In the proof to Theorem 4.7 above we use only the hypothesis that the observations are conditionally independent, that θ is Markov with conditional density given in Appendix C, and that the family $\{P(dx_k|\theta_k, \theta_{k-1})\}$ is dominated by $P(dx_k|\theta_k = \theta_{k-1} = T_p)$. But, we have also assumed in this study that the observation x is statistically related to the state θ by the "signal plus noise" relationships of Section 2.4. This added assumption makes it possible to obtain more explicit expressions for the updating equation. These expressions and their derivations are found in Appendix I.

We now apply Theorem 4.7 to deduce the updating procedure for the function $O_k(\theta_k|\vec{x}_k)$. To this end, note from Eq. 4.21 that $O_k(\theta_k|\vec{x}_k)$ depends on the past observation $\vec{x}_{k-1}$ only through the function $O_{k-1}(\theta_{k-1}|\vec{x}_{k-1})$. More precisely, suppose $\vec{x}_{k-1}'$ and $\vec{x}_{k-1}''$ are any two past observations such that

$$O_{k-1}(\cdot\,|\vec{x}_{k-1}') = O_{k-1}(\cdot\,|\vec{x}_{k-1}'')$$

Then, for any x_k, if $\vec{x}_k' = (\vec{x}_{k-1}', x_k)$ and $\vec{x}_k'' = (\vec{x}_{k-1}', x_k)$, then

$$O_k(\cdot\,|\vec{x}_k') = O_k(\cdot\,|\vec{x}_k'')$$

Thus, without loss of generality we can write

$$O_k(\theta_k|\vec{x}_k) = O_k(\theta_k, x_k, O_{k-1})$$

We can then consider Eq. 4.21 as defining a function U that maps the current observation x_k and the function $O_{k-1}(\cdot, x_{k-1}, O_{k-2})$ into the function $O_k(\cdot, x_k, O_{k-1})$. Specifically,

$$O_k(\cdot, x_k, O_{k-1}) = U(O_{k-1}(\cdot, x_{k-1}, O_{k-2}), x_k) \tag{4.23}$$

Then, in terms of this function, the updating procedure can be illustrated as in Fig. 4.5.

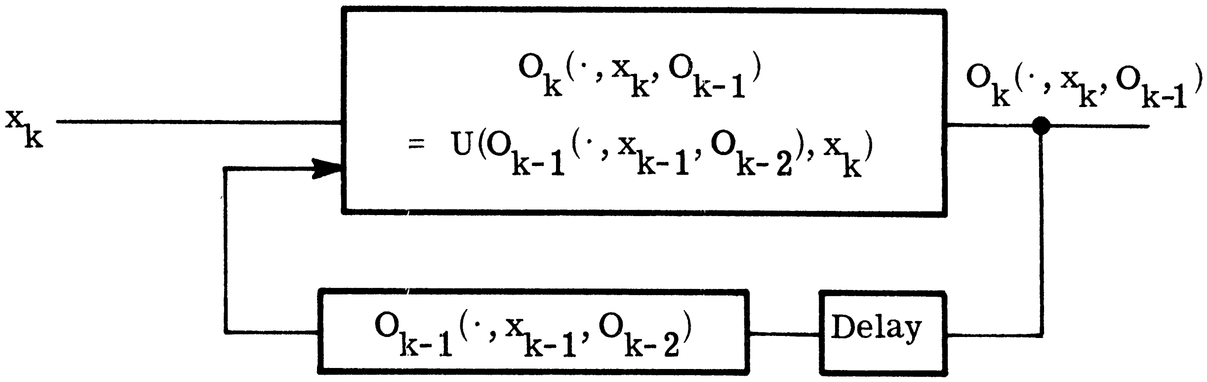


Fig. 4.5. The updating procedure for $O_k(\cdot, x_k, O_{k-1})$

We conclude this section with an interpretation of the function $O_k(\theta_k, x_k, O_{k-1})$. To do this, consider the integral

$$\int_{[0, T_p)} O_k(\theta_k, x_k, O_{k-1})\, \mu(d\theta_k) = \left(\int_{[0, T_p)} p_k(\theta_k | \vec{x}_k)\, \mu(d\theta_k) \right) \Big/ p_k(\theta_k = T_p | \vec{x}_k)$$

In the denominator of this expression we have

$$p_k(\theta_k|\vec{x}_k) = Pr[\theta_k = T_p|\vec{x}_k]$$

$$= Pr[\text{no pulse arrives in } [t_k - T_p, t_k)|\vec{x}_k]$$

In the numerator, we have

$$\int_{\Theta'} p_k(\theta_k|\vec{x}_k)\,\mu(d\theta_k) = Pr[\theta_k \epsilon [0, T_p)|\vec{x}_k]$$

$$= Pr[\text{a pulse arrives in } [0, T_p)|\vec{x}_k]$$

Thus, the integral of $O_k(\theta_k, x_k, O_{k-1})$ over the set $\{\theta_k \epsilon [0, T_p)\}$ gives the ratio of the posterior probability that signal is present to the posterior probability that no signal is present. Now, in the literature it is customary to refer to this ratio as the odds ratio. Thus, in this context $O_k(\theta_k, x_k, O_{k-1})$ can be viewed as an "odds ratio density," since its integral gives the corresponding odds ratio.

4.4 The Bayes Decision Device for Conditionally Independent Observations

In Section 4.1 it was pointed out that there are two basic difficulties in implementing the general Bayes decision device: the fact that the past observation memory must grow linearly in time and the fact that the distributions

$$\Pi(d\theta_k|\vec{x}_k) \quad \text{and} \quad P(dx_{k+1}|\vec{x}_k)$$

must be calculated at each decision time t_k. In this section we will apply the updating equation of the preceding section to replace the observation memory by a memory of fixed size and to replace the distribution $P(dx_{k+1} | \vec{x}_k)$ by the known distributions

$$P(dx_{k+1} | \theta_{k+1}, \theta_k) \quad \text{and} \quad \Pi(d\theta_{k+1} | \theta_k)$$

This is done in the following three theorems.

Theorem 4.8. Under the assumptions of Theorem 4.4,

$$T_k(\vec{x}_k, \theta_k) = \int_{\Theta_k} T_k(\vec{x}_k, \tau_k, \theta_k)\, p(\theta_k | \vec{x}_k)\, \mu(d\theta_k) \tag{4.24}$$

where

$$T_k(\vec{x}_k, \tau_k, \theta_k) = R_k(\vec{x}_k, \tau_k, a_k{=}0, \theta_k) - R_k(\vec{x}_k, \tau_k, a_k{=}1, \theta_k) \tag{4.25}$$

and

$$R_k(\vec{x}_k, \tau_k, a_k, \theta_k) \tag{4.26}$$

$$\begin{cases} L_k(\tau_k, a_k, \theta_k) \\ \quad + \displaystyle\sum_{t=0}^{1} \left\{ \int_{\Theta_k} \int_{{}^tA_{k+1}(\vec{x}_k, \tau_{k+1})} R_{k+1}(\vec{x}_{k+1}, \tau_{k+1}, a_{k+1}{=}t, \theta_{k+1})\, P(dx_{k+1} | \theta_{k+1}, \theta_k) \right. \\ \qquad\qquad \left. \times\, \Pi(d\theta_{k+1} | \theta_k) \right\} & k < N \\ L_N(\tau_N, a_N, \theta_N) & k = N \end{cases}$$

Proof. We begin by showing that $R_k(\vec{x}_k, \tau_k, a_k)$ can be written as

$$R_k(\vec{x}_k, \tau_k, a_k) = \int_{\Theta_k} R_k(\vec{x}_k, \tau_k, a_k, \theta_k)\, \Pi(d\theta_k | \vec{x}_k) \tag{4.27}$$

where $R_k(\vec{x}_k, \tau_k, a_k, \theta_k)$ satisfies Eq. 4.26. The proof is by induction on the index k. For $k = N$, it follows from Eq. 4.8 that

$$R_N(\vec{x}_N, \tau_N, a_N) = \int_{\Theta_N} L_N(\tau_N, a_N, \theta_N)\, \Pi(d\theta_N | \vec{x}_N)$$

$$= \int_{\Theta_N} R_N(\vec{x}_N, \tau_N, a_N, \theta_N)\, \Pi(d\theta_N | \vec{x}_N)$$

where

$$R_N(\vec{x}_N, \tau_N, a_N, \theta_N) = L_N(\tau_N, a_N, \theta_N)$$

Thus, the result holds for $k = N$. Now, assume Eq. 4.26 and Eq. 4.27 are satisfied for indices $k+1, \ldots, N$. From Eq. 4.8 we have

$$R_k(\vec{x}_k, \tau_k, a_k) \tag{4.28}$$

$$= \int_{\Theta_k} L_k(\tau_k, a_k, \theta_k)\, \Pi(d\theta_k | \vec{x}_k)$$

$$+ \sum_{t=0}^{1} \left\{ \int_{{}^tA_{k+1}(\vec{x}_k, \tau_{k+1})} R_{k+1}(\vec{x}_{k+1}, \tau_{k+1}, a_{k+1}=t)\, P(dx_{k+1} | \vec{x}_k) \right\}$$

Denote the second term in Eq. 4.28 by $B_k(\vec{x}_k, \tau_{k+1})$ and apply the induction hypothesis of Eq. 4.27 with k replaced by k+1 to conclude that

$$B_k(\vec{x}_k, \tau_{k+1}) \tag{4.29}$$

$$= \sum_{t=0}^{1} \left\{ \int_{{}^tA_{k+1}(\vec{x}_k, \tau_{k+1})} \left[\int_{\Theta_{k+1}} R_{k+1}(\vec{x}_{k+1}, \tau_{k+1}, a_{k+1}=t, \theta_{k+1}) \times \Pi(d\theta_{k+1} | \vec{x}_{k+1}) \right] P(dx_{k+1} | \vec{x}_k) \right\}$$

Next, substitute for $\Pi(d\theta_{k+1} | \vec{x}_{k+1})\, P(dx_{k+1} | \vec{x}_k)$ from Theorem 4.4 into Eq. 4.29 to obtain

$$B_k(\vec{x}_k, \tau_{k+1}) =$$

$$\int_{\Theta_k} \left[\sum_{t=0}^{1} \int_{\Theta_{k+1}} \int_{{}^tA_{k+1}(\vec{x}_k, \tau_{k+1})} R_{k+1}(\vec{x}_{k+1}, \tau_{k+1}, a_{k+1}=t, \theta_{k+1})\, P(dx_{k+1} | \theta_{k+1}, \theta_k) \times \Pi(d\theta_{k+1} | \theta_k) \right] \Pi(d\theta_k | \vec{x}_k) \tag{4.30}$$

Finally, substitute Eq. 4.30 for the second term of Eq. 4.28 and combine the two integrals into a single integral. The result is

$$R_k(\vec{x}_k, \tau_k, a_k) =$$

$$\int_{\Theta_k} \Bigg\{ L(\tau_k, a_k, \theta_k)$$

$$+ \sum_{t=0}^{1} \int_{\Theta_{k+1}} \int_{{}^tA_{k+1}(\vec{x}_k, \tau_{k+1})} R_{k+1}(\vec{x}_{k+1}, \tau_{k+1}, a_{k+1}=t, \theta_{k+1})\, P(dx_{k+1} | \theta_{k+1}, \theta_k)$$

$$\times\, \Pi(d\theta_{k+1} | \theta_k) \Bigg\}\, \Pi(d\theta_k | \vec{x}_k)$$

$$= \int_{\Theta_k} R_k(\vec{x}_k, \tau_k, a_k, \theta_k)\, \Pi(d\theta_k | \vec{x}_k)$$

as was to be shown.

It remains to verify Eqs. 4.24 and 4.25. To do this, substitute for $R_k(\vec{x}_k, \tau_k, a_k)$ from Eq. 4.27 into Eq. 4.6 and combine the two integrals into a single integral. The result is

$$T_k(\vec{x}_k, \tau_k) = \int_{\Theta_k} R_k(\vec{x}_k, \tau_k, \theta_k)\, \Pi(d\theta_k | \vec{x}_k)$$

where $T_k(\vec{x}_k, \tau_k, \theta_k)$ is given by Eq. 4.25. The proof is then completed by using Theorem 4.6 to write $\Pi(d\theta_k | \vec{x}_k)$ as $p(\theta_k | \vec{x}_k)\, \mu(d\theta_k)$.

In the next theorem the test function $T_k(\vec{x}_k, \tau_k)$ is written in terms of the odds ratio density $O_k(\theta_k, x_k, O_{k-1})$.

<u>Theorem 4.9.</u> Under the hypothesis of Theorem 4.6, the component decision rule $\delta_k(\vec{x}_k, \tau_k)$ is given by

$$\delta_k(\vec{x}_k, \tau_k) = \begin{cases} 1 & \text{if } \tilde{T}_k(\vec{x}_k, \tau_k) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.30a)$$

where

$$\tilde{T}_k(\vec{x}_k, \tau_k) = \int_{\Theta_k} T_k(\vec{x}_k, \tau_k, \theta_k)\, O_k(\theta_k, x_k, O_{k-1})\, \mu(d\theta_k) \quad (4.30b)$$

Proof. Expand the integral in Eq. 4.24 to obtain

$$T_k(\vec{x}_k, \tau_k) = \int_{\theta_k \in [0, T_p)} T_k(\vec{x}_k, \tau_k, \theta_k)\, p(\theta_k|\vec{x}_k)\, \mu_{[0, T_p)}(d\theta_k)$$

$$+ T_k(\vec{x}_k, \tau_k, \theta_k = T_p)\, p(\theta_k = T_p|\vec{x}_k)$$

First, suppose $p(\theta_k = T_p|\vec{x}_k) \neq 0$. Then, we may write

$$T_k(\vec{x}_k, \tau_k)/p(\theta_k = T_p|\vec{x}_k)$$

$$= \int_{[0, T_p)} T_k(\vec{x}_k, \tau_k, \theta_k)\, [p(\theta_k|\vec{x}_k)/p(\theta_k = T_p|\vec{x}_k)]\, \mu_{[0, T_p)}(d\theta_k)$$

$$+ T_k(\vec{x}_k, \tau_k, \theta_k = T_p)$$

$$= \int_{\Theta_k} T_k(\vec{x}_k, \tau_k, \theta_k)\, O_k(\theta_k, x_k, O_{k-1})\, \mu(d\theta_k)$$

where the last equality follows by first substituting in the odds ratio density from Eq. 4.20 and then applying the definition of $\mu(d\theta_k)$ from

Theorem 4.6 to combine the two integrals into a single integral.

On the other hand, if $p(\theta_k = T_p | \vec{x}_k) = 0$, then

$$T_k(\vec{x}_k, \tau_k) = \int_{[0, T_p)} T_k(\vec{x}_k, \tau_k, \theta_k)\, p(\theta_k | \vec{x}_k)\, \mu_{[0, T_p)}(d\theta_k)$$

$$= \int_{\Theta_k} T_k(\vec{x}_k, \tau_k, \theta_k)\, O_k(\theta_k, x_k, O_{k-1})\, \mu(d\theta_k)$$

In either case, the condition

$$T_k(\vec{x}_k, \tau_k) \geq 0$$

is equivalent to the condition

$$\widetilde{T}_k(\vec{x}_k, \tau_k) = \int_{\Theta_k} T_k(\vec{x}_k, \tau_k, \theta_k)\, O_k(\theta_k, x_k, O_{k-1})\, \mu(d\theta_k) \geq 0$$

as was to be shown.

As the final step, we show that the decision rule at time t_k depends only on x_k, O_{k-1} and τ_k.

<u>Theorem 4.10.</u> The functions $\widetilde{T}_k(\vec{x}_k, \tau_k)$ and $R_k(\vec{x}_k, \tau_k, a_k, \theta_k)$ and the response set $A_k(\vec{x}_{k-1}, \tau_k)$ depend on the observation $\vec{x}_k$ only through x_k and O_k. That is,

$$\widetilde{T}_k(\vec{x}_k, \tau_k) = \widetilde{T}_k(O_{k-1}, x_k, \tau_{k-1}) \qquad (4.32a)$$

$$R_k(\vec{x}_k, \tau_k, a_k, \theta_k) = R_k(O_{k-1}, x_k, \tau_k, a_k, \theta_k) \qquad (4.32b)$$

and

$$A_k(\vec{x}_{k-1}, \tau_k) = A_k(O_{k-1}, \tau_k) \tag{4.32c}$$

The proof to this theorem is a straightforward proof by induction. The details are found in Appendix J.

On the basis of the above theorems, we may replace the past observation memory in the block diagram of Fig. 4.3 by a memory for the previous odds ratio density O_{k-1} and an updating device for obtaining O_k from x_k and O_{k-1}. The resulting block diagram is shown in Fig. 4.6.

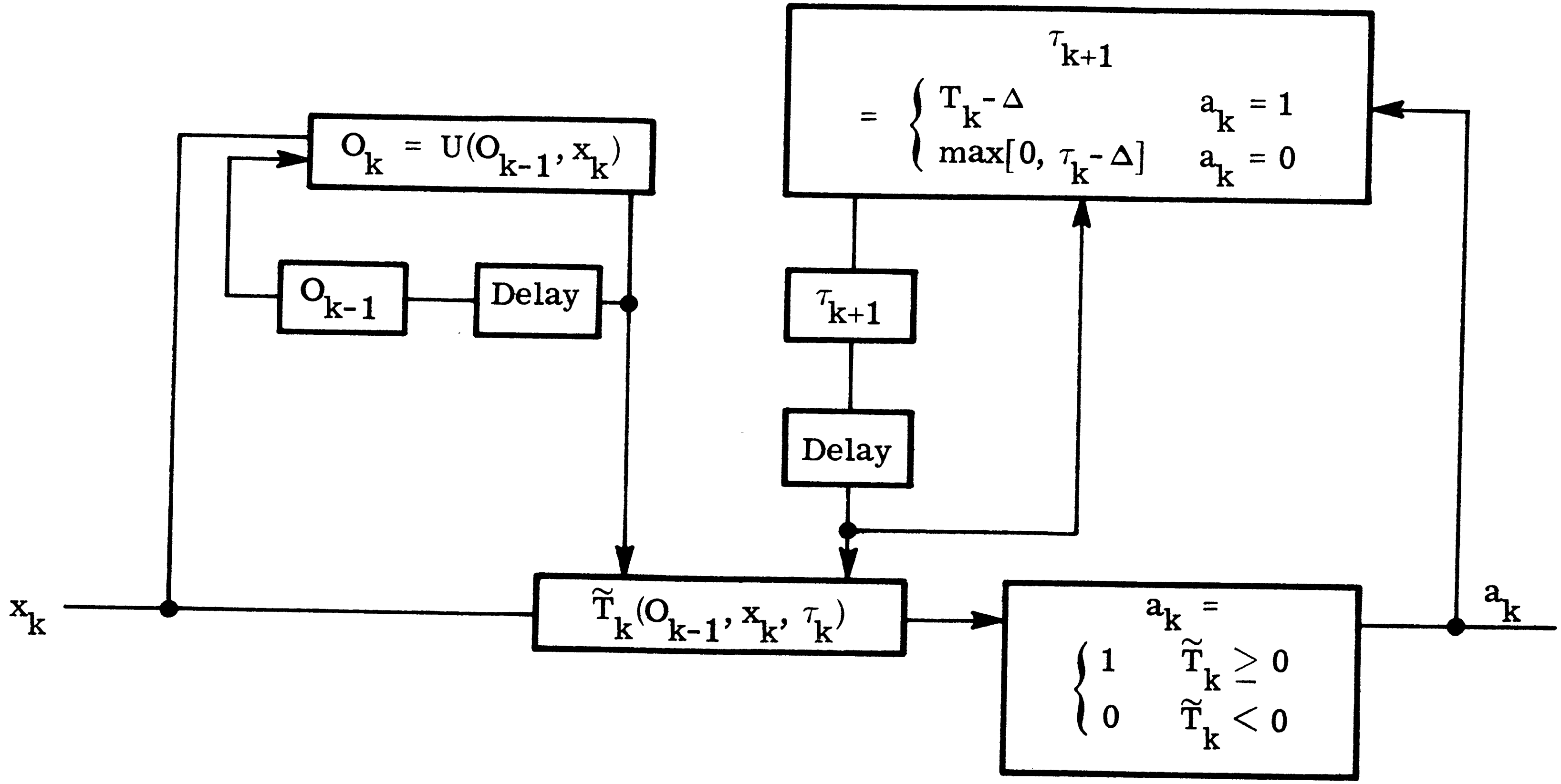


Fig. 4.6. The Bayes decision device for conditionally independent observations

CHAPTER V

A NUMERICAL SOLUTION FOR THE OPTIMUM BAYES DECISION DEVICE AND ITS ROC SURFACE

In the preceding chapter we developed the structure of the optimum Bayes decision device up to the solution of the recursive relation for the test function $T_k(\vec{x}_k, \tau_k)$. In this chapter we will carry out this recursive computation numerically for the $m = 2$ Bayes decision device. For the purposes of comparison we will also discuss the $m = 1$ Bayes decision device and the matched filter decision devices of Chapter I. Finally, we will obtain a numerical description of the ROC surfaces for these decision devices.

5.1 The Basic Setting

In order to simplify the computations in this chapter, certain additional restrictions will be made on the probability laws involved in the general model. Specifically, it is assumed that the pulse arrivals are discrete, that the noise is white Gaussian noise and that the pulse waveforms satisfy a certain translation property. In this section these assumptions are stated precisely and their effect on the basic mathematical model is determined.

To begin, we review the basic temporal structure for the $m = 2$ case. According to the definitions of Chapter II, the decision time separation Δ is

$$\Delta = T_p/m = T_p/2$$

the set of allowable decision times $\mathcal{T}$ is

$$\mathcal{T} = \{t_k \,;\, t_k = (T_p/2)k \,;\, k = 1, 2, \ldots, N\}$$

and the current observation x_k is given by

$$x_k = x(t) \qquad\qquad t \in [t_k - T_p/2, t_k)$$

The temporal relation between these quantities is shown in Fig. 5.1.

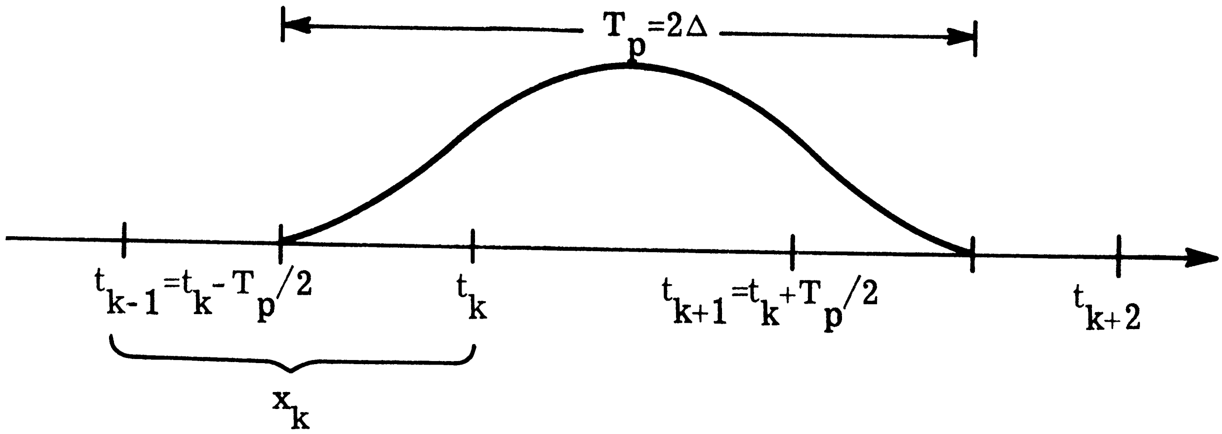


Fig. 5.1. The temporal relation between Δ, T_p, and x_k

We now state the assumptions for this chapter.

Our first assumption is that the pulse arrival times are discrete and that there are two allowable pulse arrival times per T_p seconds. Thus, in terms of the notation of Section 2.3, we assume that q is equal to 2 and that the pulse arrival time separation $\nu = T_p/q = T_p/2$. It is to be noted that the pulse separation time ν is equal to the decision separation time Δ .

As a consequence of this assumption, the random variable θ_k takes on only the three values. $\theta_k = 0,\ T_p/2$, are associated with pulse arrivals in the interval $[t_k - T_p, t_k]$ and the value $\theta_k = T_p$ denotes "noise alone" in the interval $[t_k - T_p, t_k]$. The pulse configurations associated with each of the possible values of θ_k are illustrated in Fig. 5.2 along with the temporal relation between ν and T_p .

To obtain the specific form of the prior probability law, we appeal to the relevant equations of Section 2.3. First, from Eq. 2.19 with $q = 2$ we have

$$\Pr[\theta_k = \ell\ ] = \begin{cases} \dfrac{1}{[2 + a/(1-a)]} & \ell = 0, 1 \\[2ex] \dfrac{[a/(1-a)]}{[2 + a/(1-a)]} & \ell = 2 \end{cases} \tag{5.1}$$

Next, the conditional prior probabilities $\Pr[\theta_k = \ell\delta \mid \theta_k = j\delta]$ are obtained from Eq. 2.17 by setting $q = 2$ and $s = \Delta/\nu = 1$. The results are most conveniently expressed in terms of the transition matrix $\{p_{\ell,j}\}$, where $p_{\ell,j} = \Pr[\theta_k = \ell\nu \mid \theta_{k-1} = j\nu]$. This matrix is

$$\{p_{\ell,j}\} = \begin{bmatrix} 0 & (1-a) & a \\ 1 & 0 & 0 \\ 0 & (1-a) & a \end{bmatrix} \tag{5.2}$$

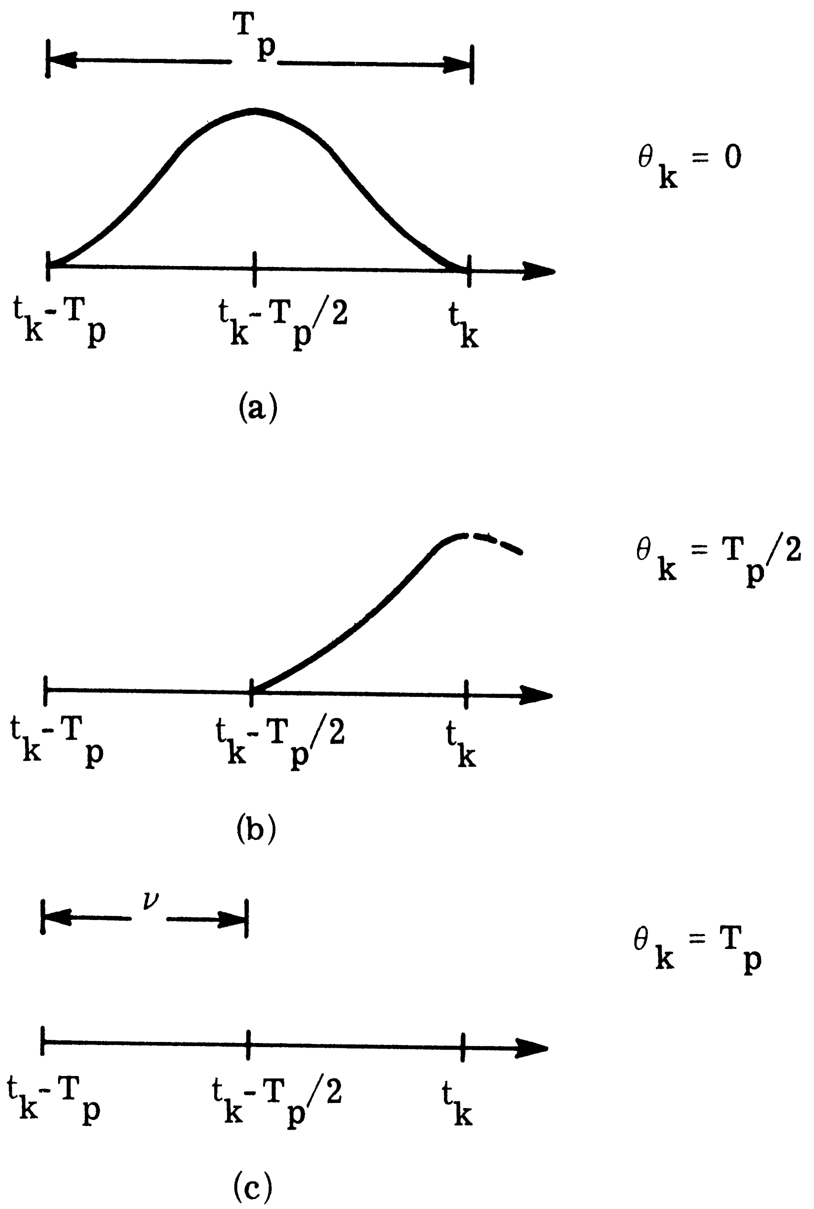


Fig. 5.2. The pulse configurations for the different values of θ_k. (a) $\theta_k = 0$, (b) $\theta_k = T_p/2$, (c) $\theta_k = T_p$

Finally, we determine the pulse rate r_p from Eq. 2.13 with $q = 2$ and $\nu = T_p/2$. The result is

$$r_p = \frac{1}{T_p} \left[\frac{2}{[2 + a/(1-a)]} \right]$$

or

$$r_p = \frac{1}{T_p} D_p \tag{5.3}$$

where

$$D_p = \frac{2}{[2 + a/(1-a)]} \tag{5.4}$$

From Eq. 5.3 and the definition of the pulse rate r_p, we have

$$D_p = \frac{E\{T_p N_p\}}{T}$$

But, since the pulses are non-overlapping, the quantity $T_p N_p$ is equal to the total amount of time occupied by signal T_S, so that

$$D_p = \frac{E\{T_S\}}{T}$$

Thus, we may interpret D_p as the average pulse duty.

Our next assumption is that the noise is white Gaussian noise. From this assumption we may obtain the update equation for the odds ratio density by substituting Eq. I.21 into Eq. I.15 in App. I, and then setting $q = 2$, $s = 1$ and $\nu = T_p/2$. The result is

$$O_k(\theta_k, x_k, O_{k-1}) \tag{5.5a}$$

$$= \begin{cases} \ell_k(x_k \mid \theta_k = 0, S_c)\, a^{-1} \left[\dfrac{O_{k-1}(\theta_{k-1} = T_p/2, x_{k-1}, O_{k-2})}{1 + O_{k-1}(\theta_{k-1} = 0, x_{k-1}, O_{k-2})} \right] & \theta_k = 0 \\ \ell_k(x_k \mid \theta_k = T_p/2, S_c)\, a^{-1}\, (1 - a) & \theta_k = T_p/ \\ 1 & \theta_k = T_p \end{cases}$$

where, from Eq. I.19,

$$\ell_k(x_k \mid \theta_k, S_c) = \exp \left\{ \frac{1}{N_o} \int_{t_k - T_p/2}^{t_k} x(t)\, p^k(t - \theta_k)\, dt - \right.$$

$$\left. \frac{1}{2 N_o} \int_{t_k - T_p/2}^{t_k} [p^k(t - \theta_k)]^2\, dt \right\} \tag{5.5b}$$

The quantity N_o is the noise power per unit bandwidth.

The above equation for $O_k(\theta_k, x_k, O_{k-1})$ can be expressed in a more compact form by denoting the quantity that appears in brackets in Eq. 5.5 by P_{k-1}. Thus we shall define P_{k-1} by

$$P_{k-1} = \frac{O_{k-1}(\theta_k = T_p/2, x_k, O_{k-2})}{1 + O_{k-1}(\theta_k = 0, x_k, O_{k-2})}$$

First, we show that P_{k-1} also satisfies an update equation. To do this, replace k-1 by k in the above equation and substitute for $O_k(\cdot, \cdot, \cdot)$ from Eq. 5.5. The result is

$$P_k = \frac{\ell_k(x_k \mid \theta_k = T_p/2, S_c)\, a^{-1}(1-a)}{1 + \ell_k(x_k \mid \theta_k = 0, S_c)\, a^{-1} \left[\frac{O_{k-1}(\theta_{k-1} = T_p/2, x_{k-1}, O_{k-2})}{1 + O_{k-1}(\theta_{k-1} = 0, x_{k-1}, O_{k-2})}\right]}$$

Next, multiply the numerator and denominator by "a" and note that the bracketed quantity in the denominator is P_{k-1}. The resulting up-dating equation for P_k is

$$P_k = \frac{\ell_k(x_k \mid \theta_k = T_p/2, S_c)\,(1-a)}{a + \ell_k(x_k \mid \theta_k = 0, S_c)\, P_{k-1}} \qquad (5.7)$$

To obtain an interpretation for P_k, substitute

$$O_{k-1}(\theta_{k-1}, x_{k-1}, O_{k-2}) = \frac{\Pr[\theta_{k-1} \mid \vec{x}_{k-1}]}{\Pr[\theta_{k-1} = T_p \mid \vec{x}_{k-1}]}$$

into Eq. 5.6 and multiply numerator and denominator by $\Pr[\theta_{k-1} = T_p \mid \vec{x}_{k-1}]$. The result is

$$P_{k-1} = \frac{\Pr[\theta_{k-1} = T_p/2 \mid \vec{x}_{k-1}]}{\Pr[\theta_{k-1} = T_p \mid \vec{x}_{k-1}] + \Pr[\theta_{k-1} = 0 \mid \vec{x}_{k-1}]}$$

or, since

$$\Pr[\theta_{k-1} = T_p \mid \vec{x}_{k-1}] + \Pr[\theta_{k-1} = 0 \mid \vec{x}_{k-1}] = 1 - \Pr[\theta_{k-1} = T_p/2 \mid \vec{x}_{k-1}]$$

we may also write

$$P_{k-1} = \frac{\Pr[\theta_{k-1} = T_p/2 \mid \vec{x}_{k-1}]}{1 - \Pr[\theta_{k-1} = T_p/2 \mid \vec{x}_{k-1}]}$$

From the above equation it is evident that the quantity P_{k-1} can be interpreted as the posterior odds based on the observation $\vec{x}_{k-1}$ that $\theta_{k-1} = T_p/2$. The event $\theta_{k-1} = T_p/2$ is shown in Fig. 5.3 .

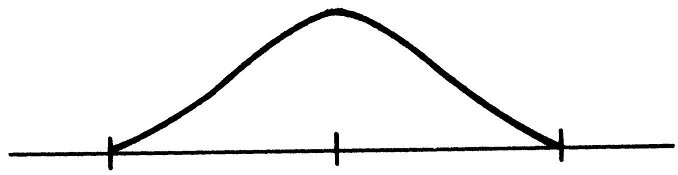


Fig. 5.3. The event $\theta_{k-1} = T_p/2$

Now, the significance of this event at the decision time t_k is that $\theta_{k-1} = T_p/2$ represents a pulse that is present at time t_{k-1} and that also has energy in the interval $[t_{k-1}, t_k)$. In Appendix I we referred to such a pulse as a carry-over pulse. Thus, we will refer to the odds ratio P_{k-1} as the carry-over odds ratio.

Finally, we assume that the first half of the pulse $p(t)$ "looks the same" as the last half. More precisely, it is assumed that

$$p(t) = p(t - T_p/2) \qquad t \in [T_p/2, T_p)$$

It follows that if $p^k(t)$ is the basic pulse with arrival time $t_k - T_p$, $p^k(t) = p[t - (t_k - T_p)]$, then $p^k(t)$ satisfies

$$p^k(t) = p^k(t - T_p/2) \qquad t \in [t_k - T_p/2, t_k)$$

An example of a pulse p(t) which satisfies this translation property is shown in Fig. 5.4.

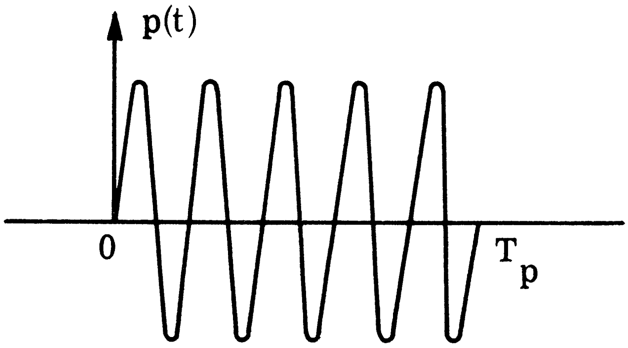


Fig. 5.4. A pulse which satisfies the translation property

As a consequence of this assumption it is seen from Eq. 5.5a that

$$\ell_k(x_k \mid \theta_k = 0, S_c) = \ell_k(x_k \mid \theta_k = T_p/2, S_c)$$

Thus at time t_k the odds ratio density $O_k(\theta_k, x_{k-1}, O_{k-1})$ depends on the observation only through the past carry-over odds P_{k-1} and a single likelihood ratio. Denote this likelihood ratio by

$$\ell_k(x_k) = \exp\left\{\frac{1}{N_o}\int_{t_k - T_p/2}^{t_k} x(t)\, p^k(t)\, dt - \frac{1}{2N_o}\int_{t_k - T_p/2}^{t_k} [p^k(t)]^2\, dt\right\} \tag{5.9}$$

Then, it follows from Eq. 5.6 and 5.9 that the update equation for the odds ratio density may be written as

$$O_k(\theta_k, x_k, O_{k-1}) = \begin{cases} \ell_k(x_k)\, a^{-1}\, P_{k-1} & \theta_k = 0 \\ \ell_k(x_k)\, a^{-1}\, (1-a) & \theta_k = T_p/2 \\ 1 & \theta_k = T_p \end{cases} \tag{5.10}$$

and the update equation for the carry-over odds ratio becomes

$$P_k = \frac{(1-a)\, \ell_k(x_k)}{a + \ell_k(x_k)\, P_{k-1}} \tag{5.11}$$

One last result will be necessary in order to describe the $m = 2$ Bayes decision rule. Specifically, we must have a description of the probability measures of the likelihood ratio $\ell_k(x_k)$ that are induced from the probability measures of x_k by Eq. 5.9. In Appendix K it is shown that there are two such distributions on ℓ_k that must be considered, one arising when x_k consists of noise alone and one arising when x_k consists of noise plus a signal pulse. Here we need only the fact that both of these probability measures can be described in terms of density functions that depend only on the signal-to-noise ratio of the basic pulse. In the following we will denote this signal-to-noise ratio by d_T^2 where

$$d_T^2 = \frac{1}{N_o} \int_0^{T_p} p(t)^2\, dt \tag{5.12}$$

5.2 The m = 1 Bayes Decision Device

In Section 4.2 we obtained the general solution of the recursive equations for the m = 1 Bayes decision device. There it was seen that the resulting decision device is the same as that obtained from classical detection theory (classical pulse detector). In this section, we will interpret this device in terms of the basic setting of the preceding section. We will then be able to compare both the structure and the performance of this device with the m = 2 Bayes decision device.

To begin, recall from Theorem 4.3 that the response condition for the m = 1 Bayes decision device is given by

$$\int_{\theta_k \in [0, T_p)} \Pi(d\theta_k \mid \vec{x}_k) \geq W_F \int_{\{\theta_k = T_p\}} \Pi(d\theta_k \mid \vec{x}_k)$$

Now, to interpret this condition in terms of the assumptions of Section 5.1, use Theorem 4.6 to replace $\Pi(d\theta_k \mid \vec{x}_k)$ by $p(\theta_k \mid \vec{x}_k)\ \mu(d\theta_k)$. Then, use Eq. 4.20 to substitute for $p(\theta_k \mid \vec{x}_k)$ in terms of $O_k(x_k, \theta_k, O_{k-1})$. The resulting response condition is

$$\int_{\theta_k \in [0, T_p)} O_k(\theta_k, x_k, O_{k-1})\ \mu(d\theta_k) \geq W_F \tag{5.13}$$

The next step is to note from the discrete case of Eq. 4.18 with q = 2 that

$$\mu(d\theta_k) = \epsilon_{\{0\}}(d\theta_k) + \epsilon_{\{T_p/2\}}(d\theta_k) + \epsilon_{\{T_p\}}(d\theta_k)$$

Thus we may evaluate the integral in the inequality (5.13) to express the response condition as

$$O_k(\theta_k = 0,\ x_k,\ O_{k-1}) + O_k(\theta_k = T_p/2,\ x_k,\ O_{k-1}) \geq W_F \tag{5.14}$$

Next, note that the set of decision times for the $m = 1$ decision device can be considered as those decision times for the $m = 2$ decision device that have even indices. (See Fig. 3.10.) Thus, we can substitute for $O_k(\theta_k, x_k, O_{k-1})$ from Eq. 5.10 if we allow decisions only for k even and if we realize that x_k and thus $\ell_k(x_k)$ and $P_k(x_k, P_{k-1})$ pertain to the observation over the interval $[t_k - T_p/2, t_k)$. Carrying out this substitution results in the response condition

$$\ell_k(x_k)\, a^{-1} P_{k-1} + \ell_k(x_k)\, a^{-1}(1-a) \geq W_F$$

or equivalently

$$T_k(P_{k-1}, \ell_k) \geq 0 \tag{5.15a}$$

where

$$T_k(P_{k-1}, \ell_k) = \ell_k - K(P_{k-1}) \tag{5.15b}$$

with

$$K(P_{k-1}) = \frac{W_F\, a}{1 - a + P_{k-1}} \tag{5.15c}$$

The function $K(P_{k-1})$ is illustrated in Fig. 5.5.

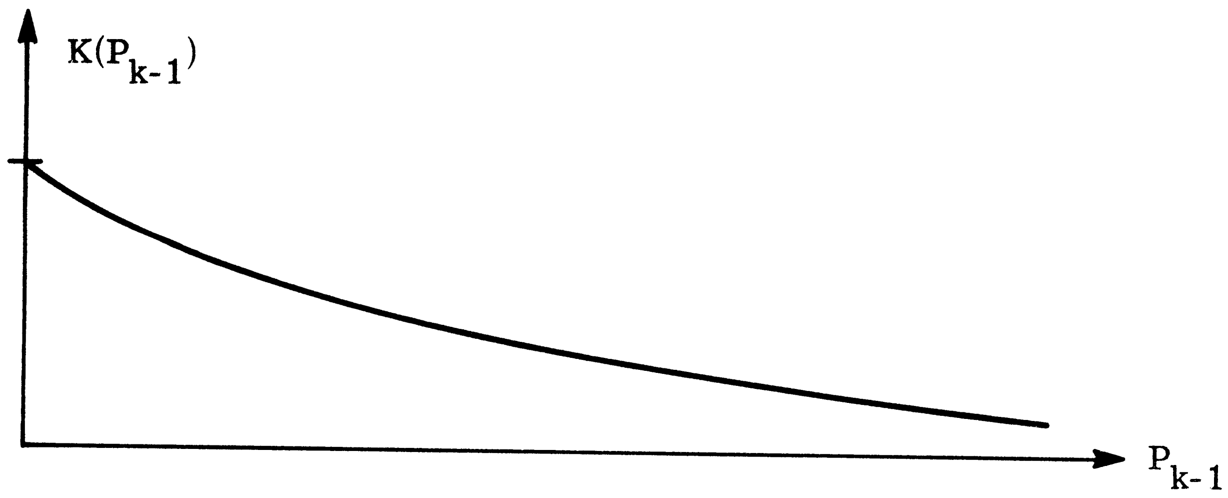


Fig. 5.5. The threshold function for the $m = 1$ Bayes decision device

The response condition of Eq. 5.15 and the updating procedure for P_k of Eq. 5.11 completely determine the block diagram for the $m = 1$ Bayes decision device.

This block diagram is illustrated in Fig. 5.6.

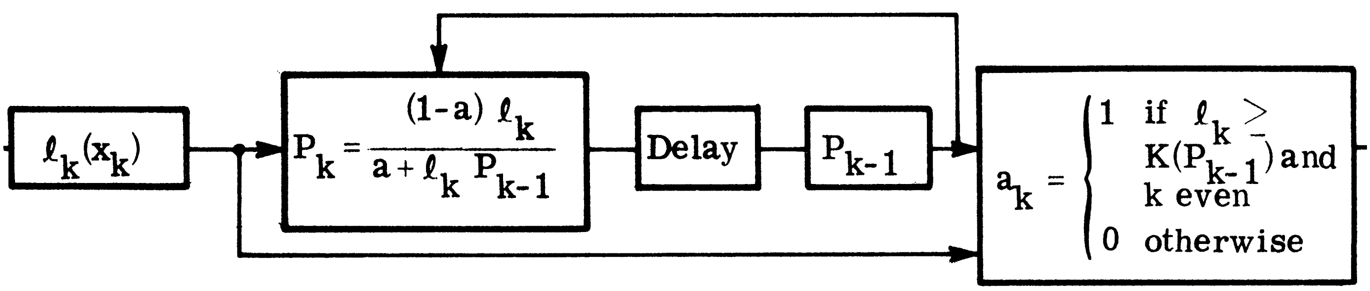


Fig. 5.6. The $m = 1$ Bayes decision device

The operation of this device is as follows. At time t_k the value of the carry-over odds ratio P_{k-1}, which has been stored in memory from time t_{k-1}, is used to determine a value of the threshold function $K(P_{k-1})$. The decision device then compares the current value of the likelihood ratio $\ell_k(x_k)$ with the threshold $K(P_{k-1})$ and responds iff

$$\ell_k(x_k) \geq K(P_{k-1}) \qquad \text{and } k \text{ is even}$$

Finally, the values of $\ell_k(x_k)$ and P_{k-1} are used to determine P_k for use at the next decision time.

We conclude this section with a few interpretive remarks on the $m = 1$ Bayes decision device. First, note from the response condition

$$\ell(x_k) \geq K(P_{k-1}) = \frac{W_F a}{(1-a) + P_{k-1}}$$

that the decision at time t_k can be considered as a likelihood ratio test against a threshold that depends on the information prior to time t_{k-1} through the statistic $P_{k-1}(\vec{x}_{k-1})$. Here, the likelihood ratio $\ell_k(x_k)$ is a measure of the evidence in the current observation x_k that signal energy is present if the interval $[t_k - T_p/2, t_k)$, whereas $P_{k-1}(\vec{x}_{k-1})$ is a measure of the evidence based on the past observation $\vec{x}_{k-1}$ that the carry-over pulse is present. These quantities and the carry-over pulse are illustrated in Fig. 5.7.

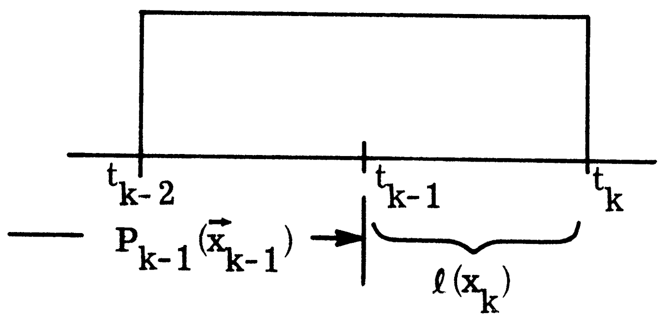


Fig. 5.7. The temporal relation of the carry-over pulse, the previous carry-over odds $P_{k-1}(\vec{x}_{k-1})$ and the current likelihood ratio $\ell(x_k)$

The form of the threshold function $K(P_{k-1})$ can be interpreted as follows. The large values of P_{k-1} indicate strong evidence that the carry-over pulse is present. For these values, $K(P_{k-1})$ is small as seen from Fig. 5. 5, and thus little evidence is needed from $\ell(x_k)$ in order to respond. On the other hand, small values of P_{k-1} indicate strong evidence that the carry-over pulse is not present. For these values, $K(P_{k-1})$ is larger, so that more evidence from $\ell(x_k)$ is needed in order to respond.

The extreme values of P_{k-1} are interesting. Suppose first that $P_{k-1} = +\infty$. Then, the decision device knows with absolute certainty on the basis of $\vec{x}_{k-1}$ that the carry-over pulse is present. Now, since the carry-over pulse must be detected at time t_k , the decision device should respond regardless of the information present in the likelihood ratio $\ell(x_k)$. But, this is compatible with the response condition, since

$$K(P_{k-1} = +\infty) = 0$$

On the other hand, suppose $P_{k-1} = 0$. Then, the decision-maker knows with absolute certainty on the basis of $\vec{x}_{k-1}$ that the carry-over pulse is not present. Now, from Fig. 5.2 it is seen that the only other possibilities are that $\theta_k = T_p$ so that no pulse is present in the interval $[t_k - T_p, t_k)$, or that $\theta_k = T_p/2$ so that a pulse arrives at time t_{k-1} . But, at time t_k these two possibilities can be distinguished only on the basis of whether or not there is signal energy in the interval $[t_k - T_p/2, t_k)$. Thus, from classical detection theory, we would expect that the optimum decision device would respond iff $\ell(x_k)$ exceeds the threshold

$$W_F \frac{\Pr[\theta_k = T_p \mid \theta_k \neq 0]}{\Pr[\theta = T_p/2 \mid \theta_k \neq 0]}$$

But, this is also compatible with the response condition for the $m = 1$ Bayes decision device, since, by substituting for the conditional probabilities from Eq. 5.2, we conclude that

$$W_F \frac{\Pr[\theta_k = T_p \mid \theta_k \neq 0]}{\Pr[\theta_k = T_p/2 \mid \theta_k = 0]}$$

$$= W_F \frac{a}{1 - a}$$

$$= K(P_{k-1} = 0)$$

5.3 The m = 2 Bayes Decision Device

In this section we present the results of a numerical solution for the m = 2 Bayes decision device. To obtain these results, we have interpreted the recursive equations of Section 4.4 in terms of the assumptions of Section 5.1. The resulting equations were then solved on a digital computer. The details of this procedure are found in Appendices K and L.

To begin, we first draw on Theorem K.3 of Appendix K to conclude that, as in the m = 1 case, the test function T_k depends on the total past observation $\vec{x}_k$ only through P_{k-1} and ℓ_k. Thus, in the m = 2 case we may also write

$$T_k(\vec{x}_k, \tau_k) = T_k(P_{k-1}, \ell_k, \tau_k)$$

Next, we state two basic properties of the numerical solution for $T_k(P_{k-1}, \ell_k, \tau_k)$. These properties have not been established theoretically but have only been verified through the computations.

The first of these properties states that, as in the m = 1 case, the m = 2 Bayes decision device can be considered as a likelihood ratio test against a threshold function that depends on the previous value of P_{k-1}. In the m = 2 case, however, the threshold function also depends on the specific decision time t_k and on the past decision a_{k-1} as well.

<u>Property I: Likelihood Ratio Property.</u> For each set of values of the losses W_X and W_F , there exist two sequences of functions

$$\{K_k^{\,0}(P_{k-1})\} \quad \text{and} \quad \{K_k^{\,1}(P_{k-1})\}$$

such that

$$T_k(P_{k-1}, \ell_k, \tau_k = 0) = \ell_k - K_k^{\,0}(P_{k-1})$$

and

$$T_k(P_{k-1}, \ell_k, \tau_k = T_p/2) = \ell_k - K_k^{\,1}(P_{k-1})$$

The above property specifies the form of the response condition for the $m = 2$ case. These conditions together with the updating equation for P_k allow us to conclude that the general form of the $m = 2$ Bayes decision device is as shown in Fig. 5.8.

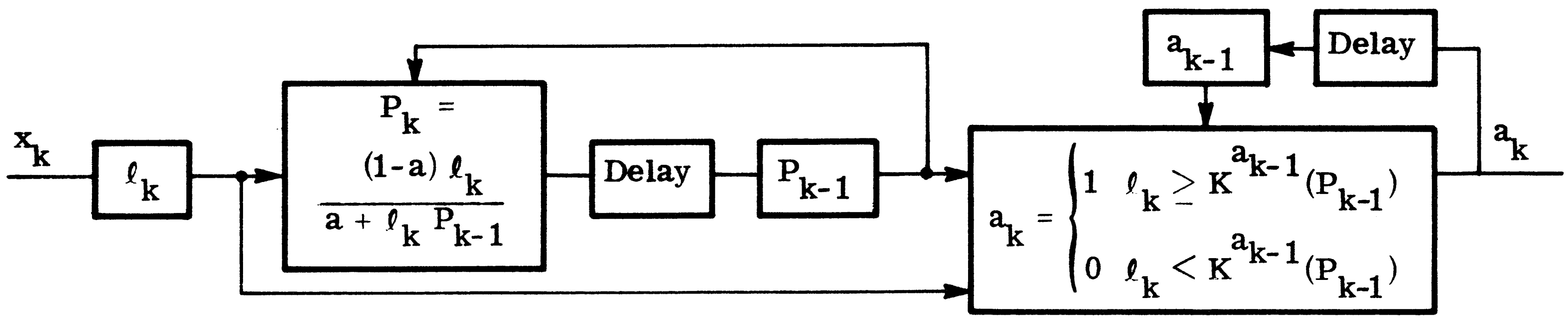


Fig. 5.8. The general form of the $m = 2$ Bayes decision device

The operation of this device is much the same as the $m = 1$ decision device. At time t_k the value of the previous carry-over odds ratio P_{k-1} and the previous decision a_{k-1} are both stored in memory. The previous decision a_{k-1} is used to determine which of the two threshold functions, $K_k^0(\cdot)$ or $K_k^1(\cdot)$, is to be used at time t_k, and the value of P_{k-1} determines the specific value of the threshold function. Then, at time t_k the device responds iff it did not respond at time t_{k-1} and

$$\ell_k \geq K_k^0(P_{k-1})$$

or iff it did respond at time t_{k-1} and

$$\ell_k \geq K_k^1(P_{k-1})$$

Finally, the value of P_{k-1} is used along with the value of ℓ_k to determine P_k for use at the next decision time. The value of a_k replaces a_{k-1} in memory.

The second of the two basic properties states that, except for decision times t_k close to the terminal decision time t_N, the structure of the $m = 2$ Bayes decision device does not depend on time. More precisely,

Property II: Asymptotic Stationarity Property. For each set of losses W_X and W_F, the two sequences of threshold functions

$$\{K_k^0(P_{k-1})\} \quad \text{and} \quad \{K_k^1(P_{k-1})\}$$

converge through decreasing indices K to limiting threshold functions

$$\hat{K}^0(P_{k-1}) \quad \text{and} \quad \hat{K}^1(P_{k-1})$$

respectively.

The interpretation of this property in terms of the system model of Fig. 5.8 is that, except for t_k near t_N, we may replace the threshold functions $K_k^0(P_{k-1})$ and $K_k^1(P_{k-1})$ by $\hat{K}^0(P_{k-1})$ and $\hat{K}^1(P_{k-1})$. Thus, for these decision times, none of the components in the decision device depend on the particular decision time t_k. We might point out that it is certainly reasonable to expect that Property II holds in the setting of this chapter, since we have assumed that both the noise process and the prior probabilities are stationary.

We now turn to the task of describing the $m = 2$ Bayes decision device for different values of the losses W_X and W_F. To this end, we note from Property I that this device is completely specified by the sequences of threshold functions

$$\{K_k^0(P_{k-1})\} \quad \text{and} \quad \{K_k^1(P_{k-1})\}$$

Moreover, by Property II these functions converge to limiting threshold functions

$$\hat{K}^0(P_{k-1}) \quad \text{and} \quad \hat{K}^1(P_{k-1})$$

Thus, we may describe the stationary behavior in the $m = 2$ case completely in terms of these two functions. We will defer the discussion of the convergence of the threshold functions to Appendix L.

In order to provide a basis for the description of the limiting threshold functions, we recall from Section 3.6 that the general Bayes decision device can be divided into five special cases depending on the values of the losses W_X and W_F. For the $m = 2$ Bayes decision device, each of these special cases results in a distinct form for the threshold functions. Thus we may describe the numerical results by presenting specific examples for each of the five cases.

Case I: $W_X^*(F) < W_X, W_F > 0$

The Bayes decision devices for these values of W_X and W_F are respond-once decision devices which satisfy the inhibit rule and which are independent of the loss W_X. Typical threshold functions for this case are shown in Fig. 5.9 and 5.10.

There are several features of these functions that we shall take special note of here. First of all, it is noted that $\hat{K}^1(P_{k-1}) = +\infty$ for all P_{k-1}. This fact is a consequence of the inhibit rule property. To see this, we note that in the $m = 2$ case the inhibit rule says that the decision device cannot respond at time t_k if it has just responded at time t_{k-1}. Thus we must have

$$T_k(P_{k-1}, \ell_k, \tau_k = T_p/2)$$
$$= \ell_k - \hat{K}^1(P_{k-1}) < 0$$

for all ℓ_k and P_{k-1}. But this is true for all ℓ_k iff we take

$$\hat{K}^1(P_{k-1}) = +\infty \quad \text{for all } P_{k-1}$$

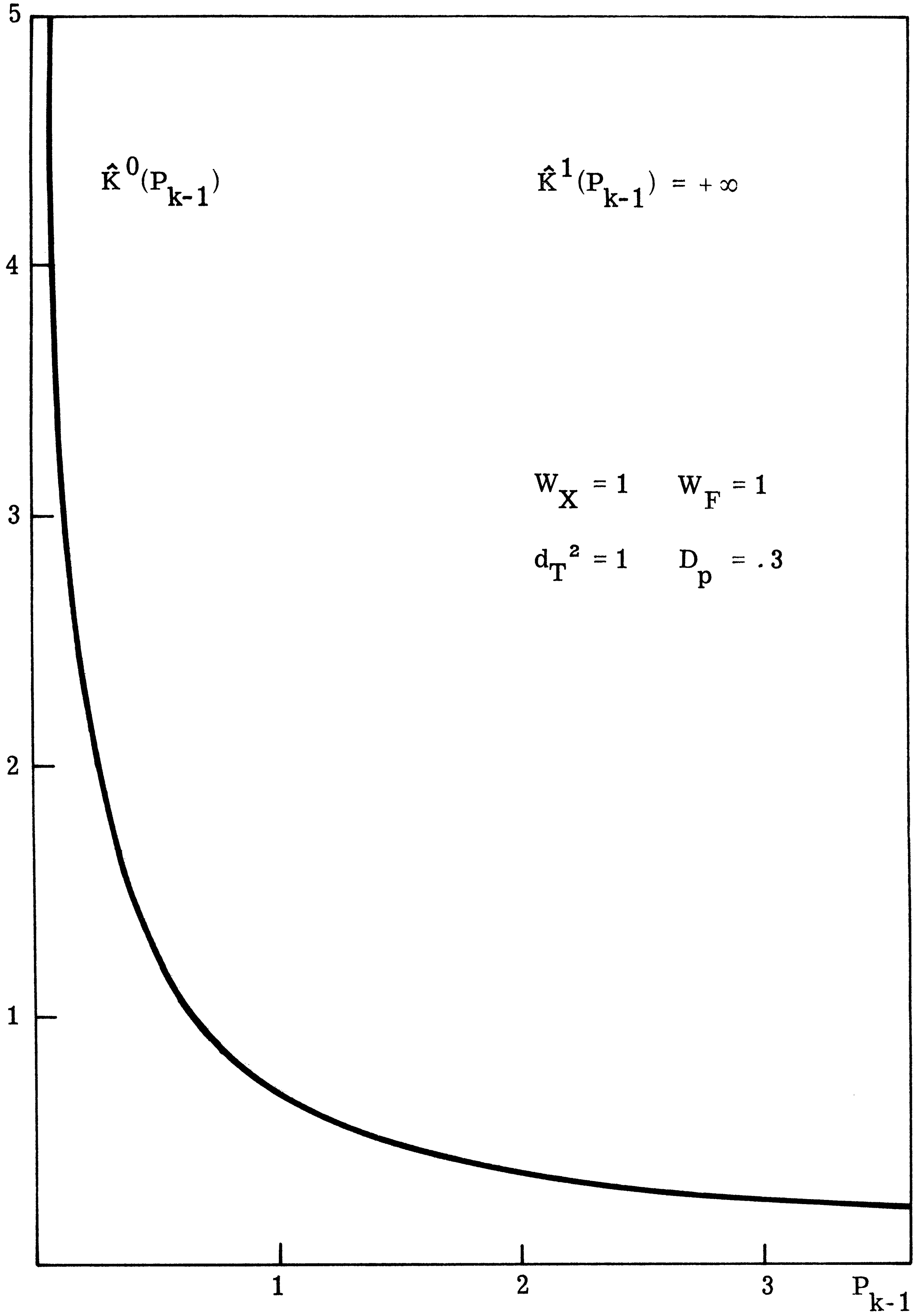


Fig. 5.9. Threshold functions for Case I;
$d_T^2 = 1$, $D_p = .3$, $W_X = 1$, $W_F = 1$

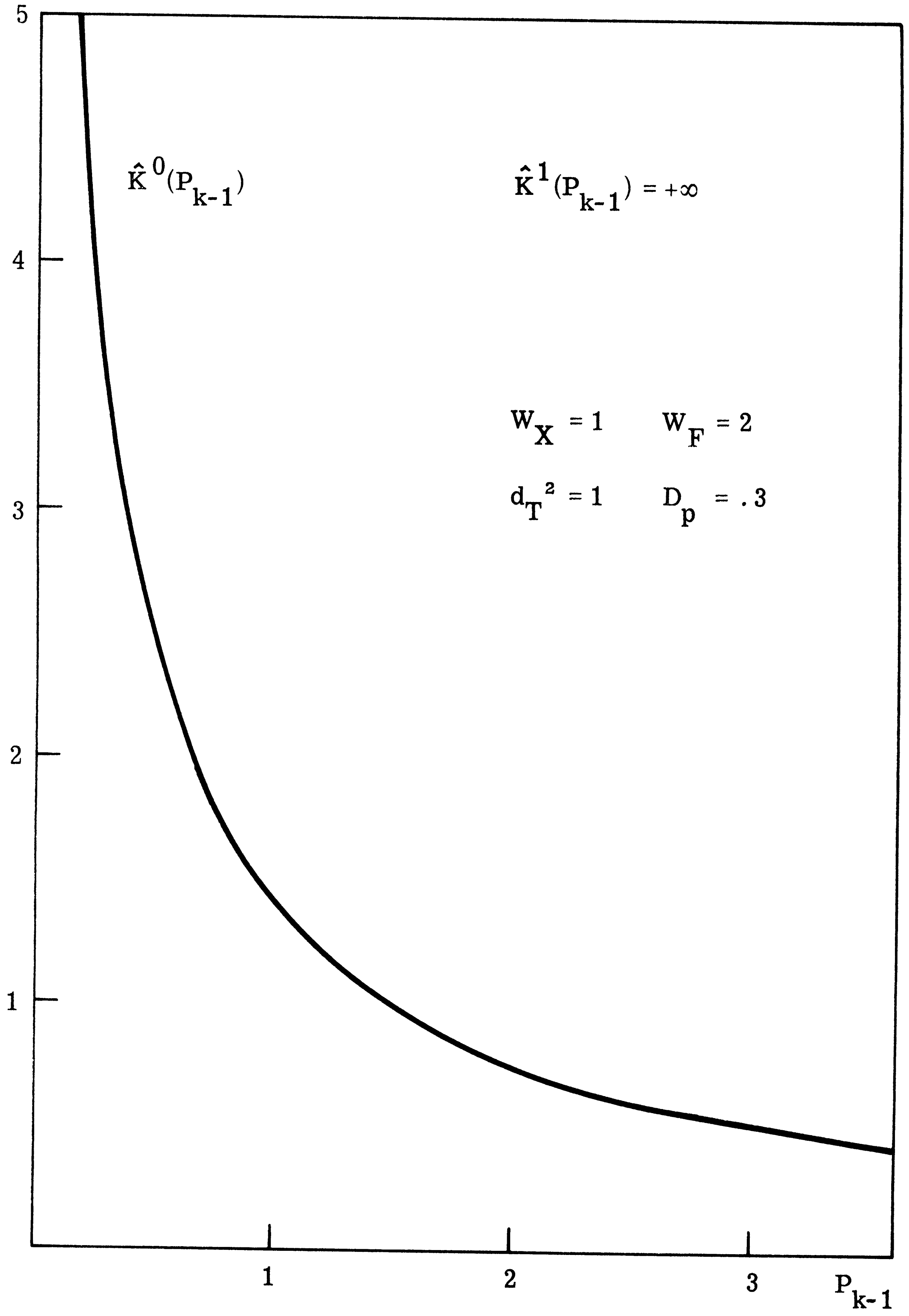


Fig. 5.10. Threshold functions for Case I;
$d_T^2 = 1$, $D_p = .3$, $W_X = 1$, $W_F = 2$

Next, it is noted that the threshold function $\hat{K}^0(P_{k-1})$ approaches $+\infty$ as P_{k-1} approaches 0. In the following paragraphs it will be seen that this property holds for all respond-once decision devices. To gain some insight into this property, consider the situation when $P_{k-1} = 0$. Now, from the preceding section we recall that when $P_{k-1} = 0$, the decision device is assured that either no pulse is present at time t_k, or if a pulse is present at time t_k, then t_k is the first opportunity to detect that pulse. Thus the response condition

$$\ell_k \geq \hat{K}^0(P_{k-1} = 0) = +\infty$$

says that, in this situation, the decision device will always defer a response decision to the next decision time.

A few additional comments on the results for Case I are appropriate here. First of all, it has been verified computationally that the threshold functions are completely independent of the loss W_X. This is to be expected from Theorem 3.11 which says that any Bayes decision device that satisfies the inhibit rule depends only on W_F and not W_X. Secondly, we point out that we have not been able to obtain an analytical expression for the function $W_X^*(W_F)$. Nevertheless, by trying different values of the losses W_X and W_F and noting for which of these values the inhibit rule holds, we have obtained a numerical description of this function. This description appears in Fig. 5.11 for $d_T^2 = 1$ and $D_p = .3$.

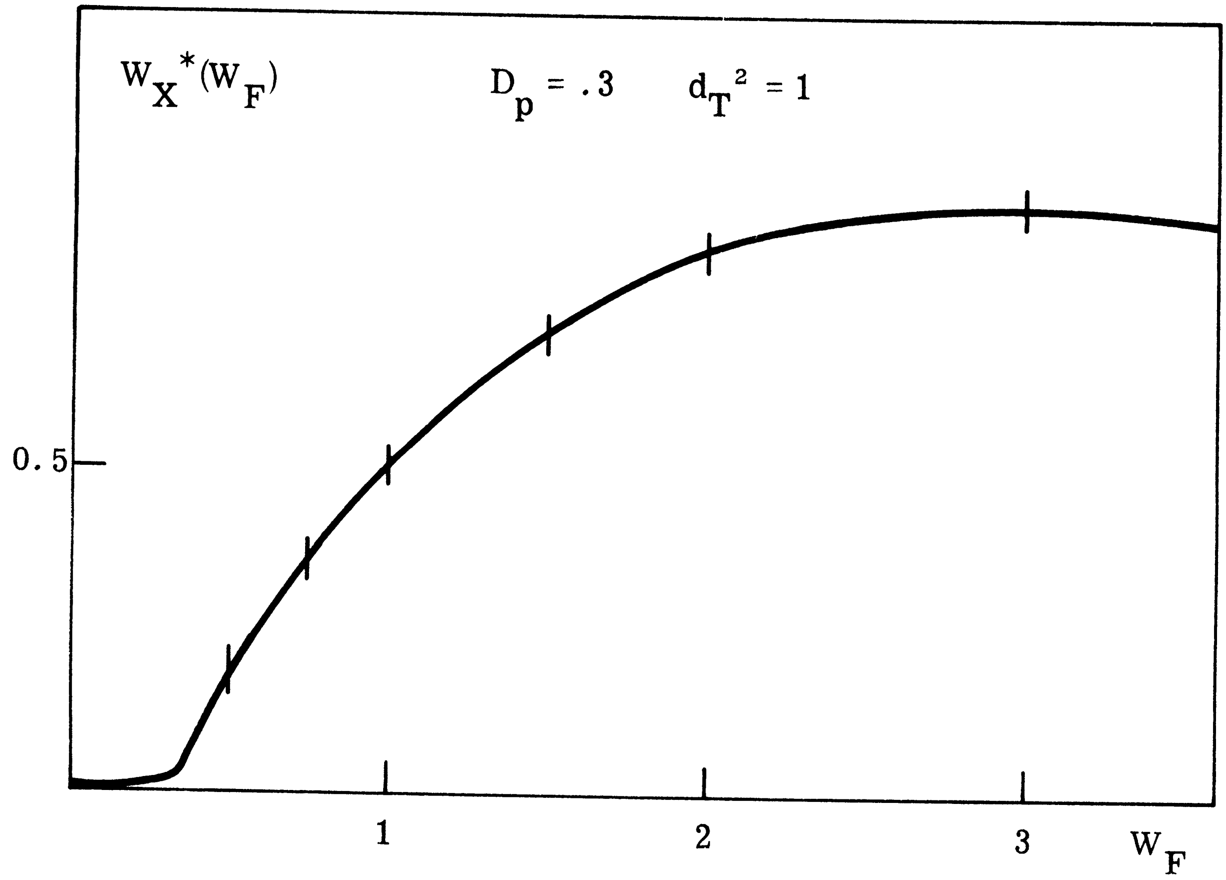


Fig. 5.11. The function $W_X^*(W_F)$

Case II: $0 < W_X \leq W_X^*(W_F),\ W_F > 0$

The Bayes decision devices for these losses are respond-once decision devices which do not satisfy the inhibit rule property. Typical threshold functions are shown in Figs. 5.12 through 5.15.

We make the following observations. First of all, it is noted that for constant W_F, $\hat{K}^1(P_{k-1})$ decreases with decreasing W_X. This says that the decision device with the smaller value of W_X will break the inhibit rule more often than the decision device with the

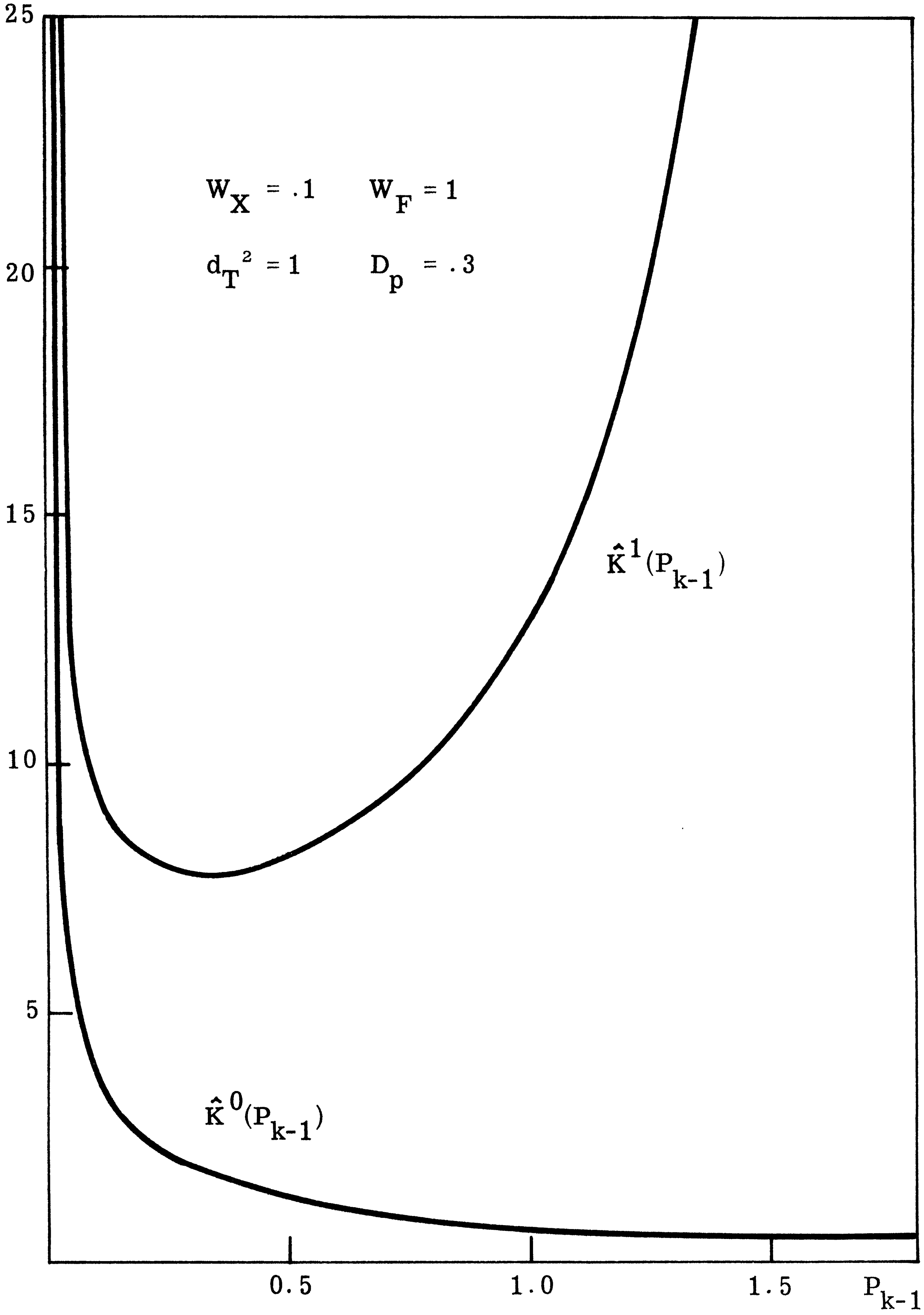


Fig. 5.12. Threshold functions for Case II; $d_T{}^2 = 1$, $D_p = .3$, $W_X = .1$, $W_F = 1$

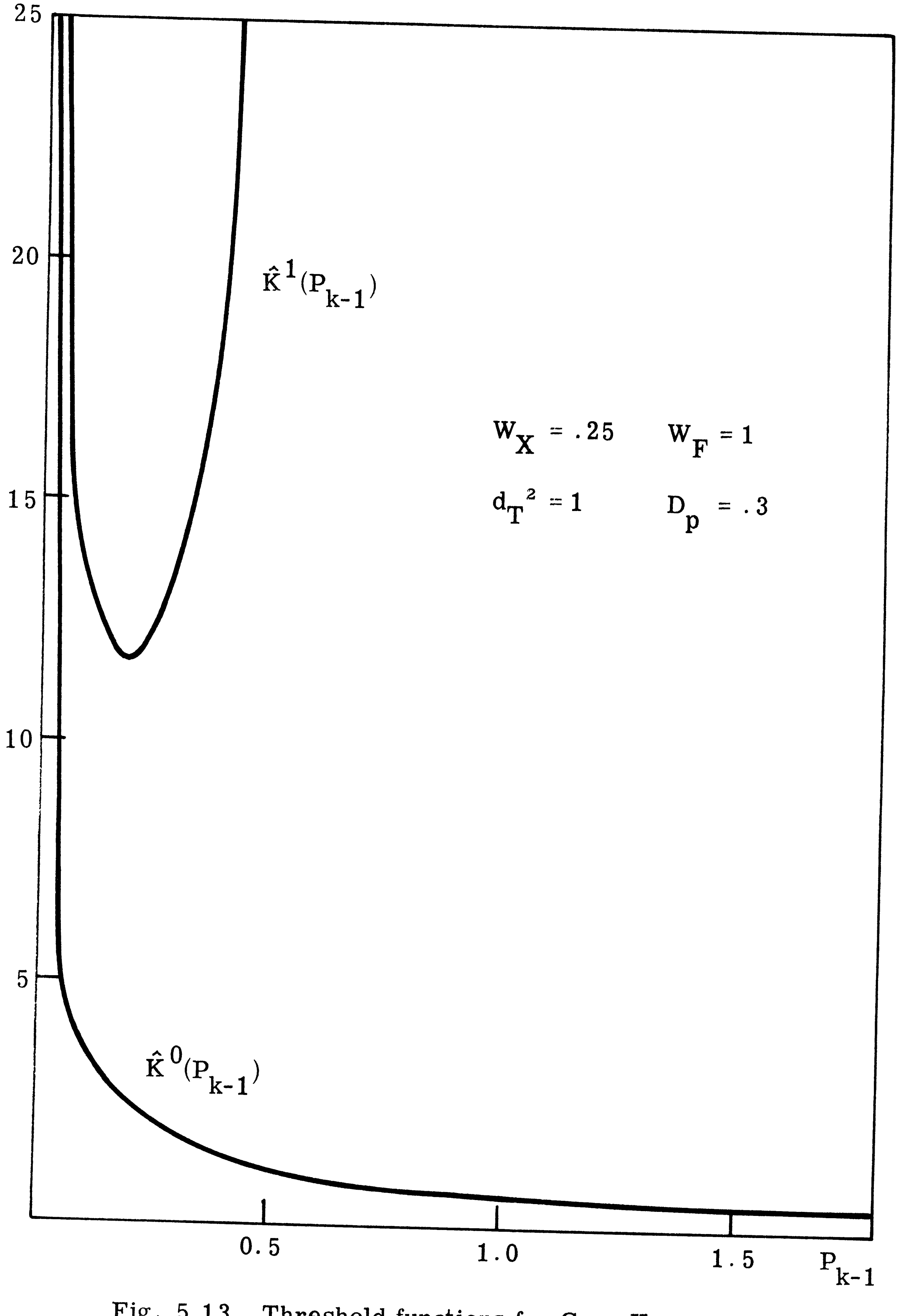


Fig. 5.13. Threshold functions for Case II; $d_T{}^2 = 1$, $D_p = 3$, $W_X = .25$, $W_F = 1$

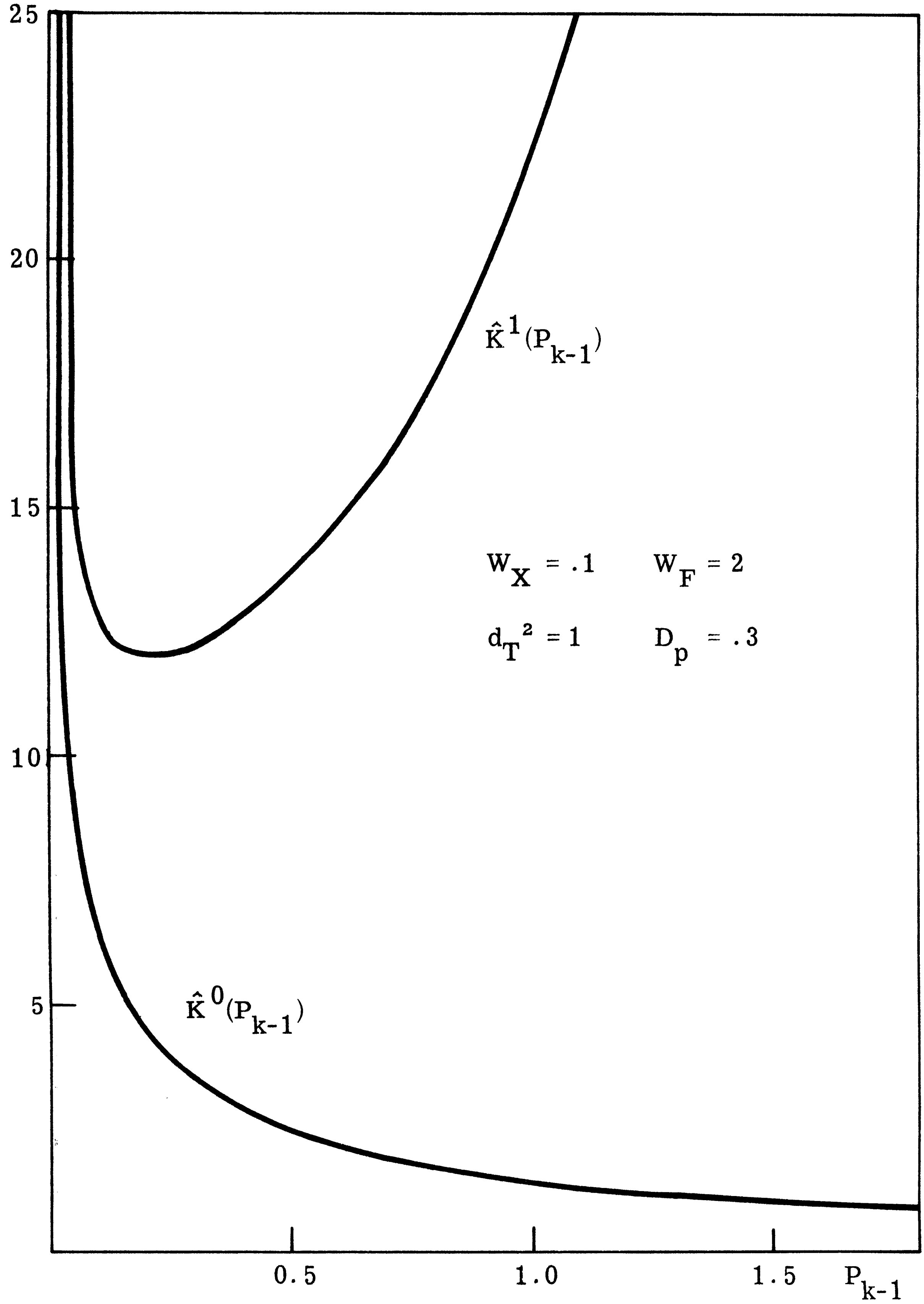


Fig. 5.14. Threshold functions for Case II;
$d_T^2 = 1$, $D_p = .3$, $W_X = .1$, $W_F = 2$

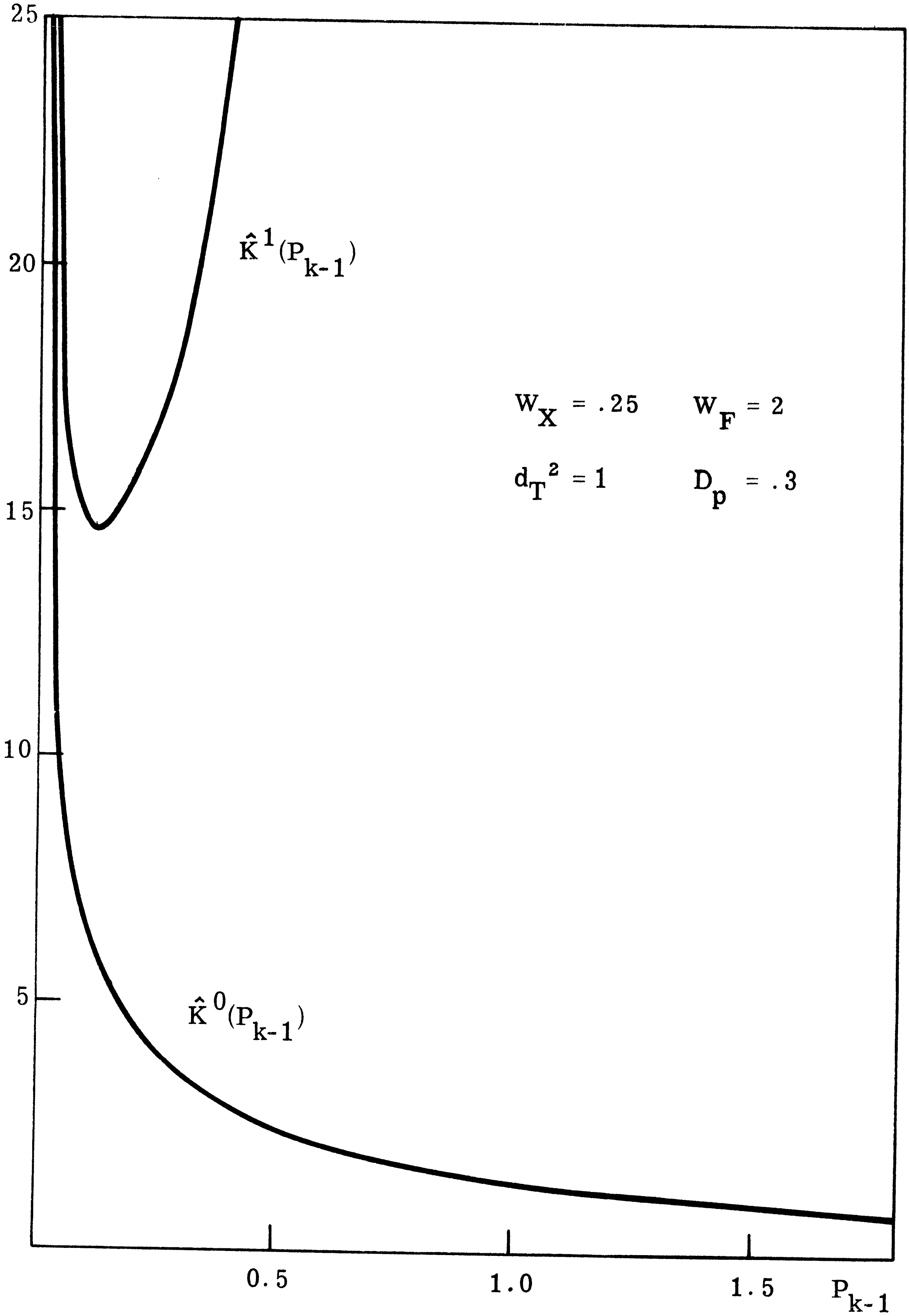


Fig. 5.15. Threshold functions for Case II; $d_T^2 = 1$, $D_p = .3$, $W_X = .25$, $W_F = 2$

larger value of W_X . This is certainly compatible with the physical interpretation of the loss W_X, since large values of W_X reflect a strong desire to avoid extra detections and extra detections can occur only when the inhibit rule is violated.

Next, it is noted that $\hat{K}^1(P_{k-1})$ is always larger than $\hat{K}^0(P_{k-1})$, and that, although $\hat{K}^1(P_{k-1})$ is finite for some values of P_{k-1}, it approaches $+\infty$ for both small and large values of P_{k-1}. The interpretation here is that the decision device can break the inhibit rule only when the past information on the carry-over pulse is inconclusive and then only if ℓ_k is large.

Finally, by comparing Fig. 5.12 with Fig. 5.13, and 5.14 with Fig. 5.15, it is seen that $\hat{K}^0(P_{k-1})$ is completely independent of the value of W_X . Moreover, for the same value of W_F the function $\hat{K}^0(P_{k-1})$ in Case II is identical to the function $\hat{K}^0(P_{k-1})$ in Case I. We shall return to this point in the discussion of Case III below.

<u>Case III: $W_X = 0$, $W_F > 0$</u>

The Bayes decision devices for these losses are the maximum pulse detection devices. Typical threshold functions for these devices are shown in Figs. 5.16 and 5.17.

There are two observations to be made here. First of all, the function $\hat{K}^1(P_{k-1})$ remains finite for large values of P_{k-1} . Thus, even if the decision device knows that the carry-over pulse is present, ($P_{k-1} = +\infty$), and even if it has detected this pulse at the

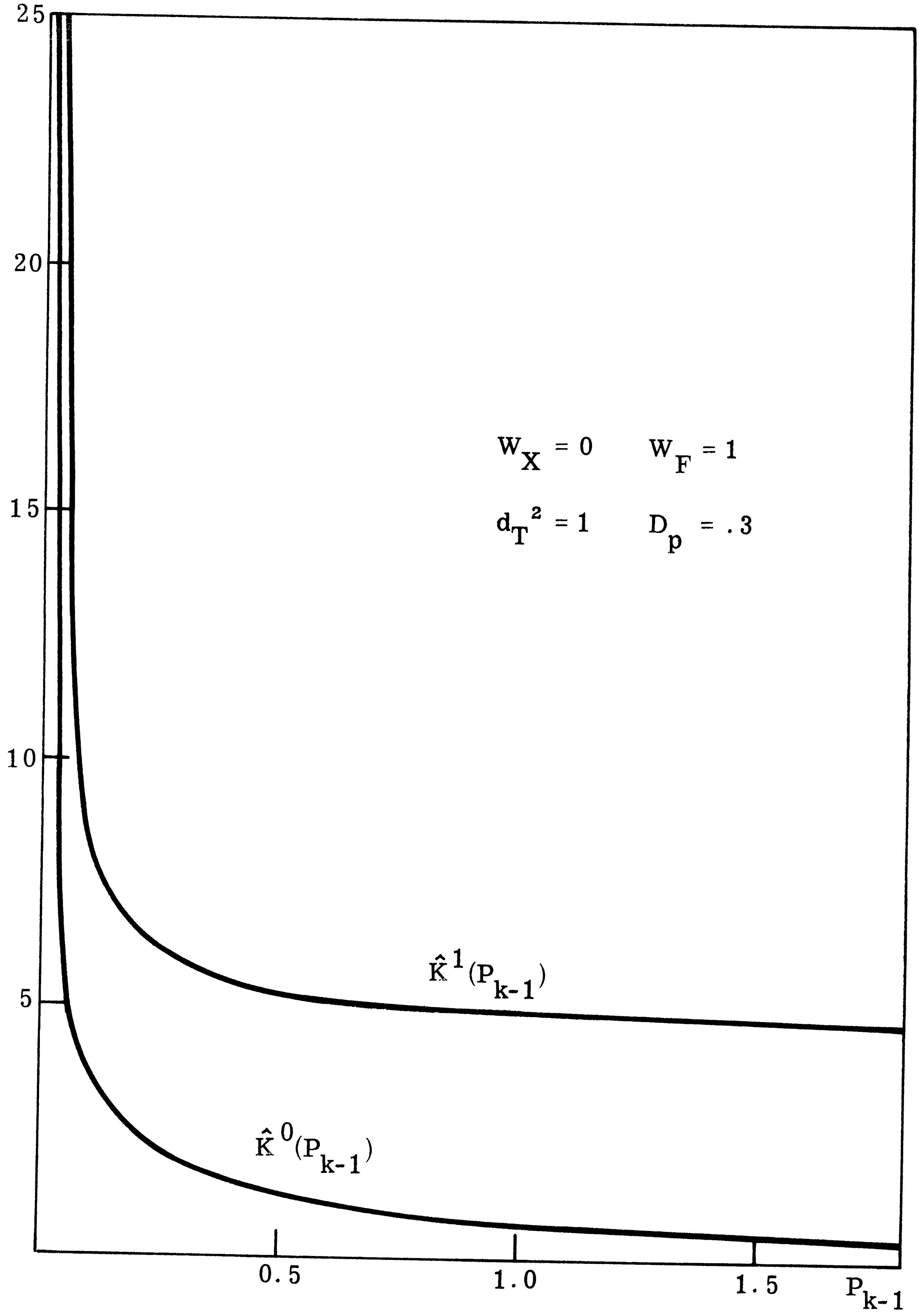


Fig. 5.16. Threshold functions for Case III; $d_T^2 = 1$, $D_p = .3$, $W_X = 0$, $W_F = 1$

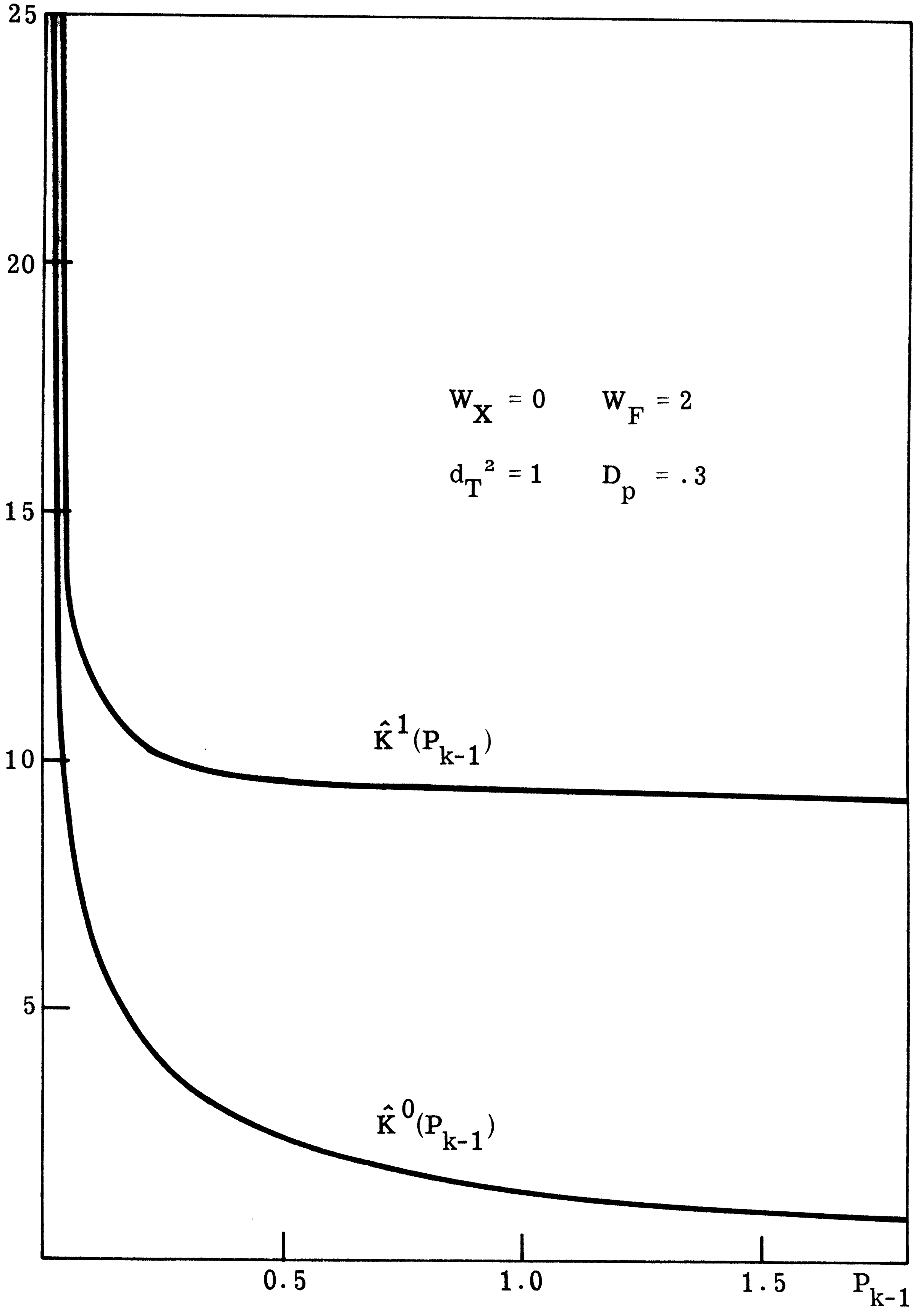


Fig. 5.17. Threshold functions for Case III; $d_T^2 = 1$, $D_p = .3$, $W_X = 0$, $W_F = 2$

preceding decision time t_{k-1}, it may still respond again at time t_k. It is to be noted, however, that the extra detection that results incurs no extra loss since $W_X = 0$ for the MPD decision device.

The second and most important observation is that for the same value of W_F the threshold function $K^0(P_{k-1})$ in Case III is identically equal to the threshold function $K^0(P_{k-1})$ in Cases II and I. Thus, for any respond-once decision device (Case I, II or III), the threshold function $K^0(P_{k-1})$ does not depend on W_X. Hence, the only effect of the loss W_X on the respond-once decision devices is to determine the function $\hat{K}^1(P_{k-1})$.

The remaining two cases pertain to the respond-and-hold decision devices.

Case IV: $-1 < W_X < 0$, $W_F > 0$

The Bayes decision devices for these losses are those decision devices which balance the detection rate R_D against the detection duty D_D. Typical threshold functions for this case are shown in Figs. 5.18 through 5.21.

There are several observations to be made here. First of all, it is noted that both the threshold function $\hat{K}^1(P_{k-1})$ and $\hat{K}^0(P_{k-1})$ depend on both the losses W_X and W_F. Moreover, these functions decrease with either decreasing W_X or decreasing W_F.

Secondly, both the threshold functions are finite at the origin. Thus, even if the decision device knows with absolute certainty that

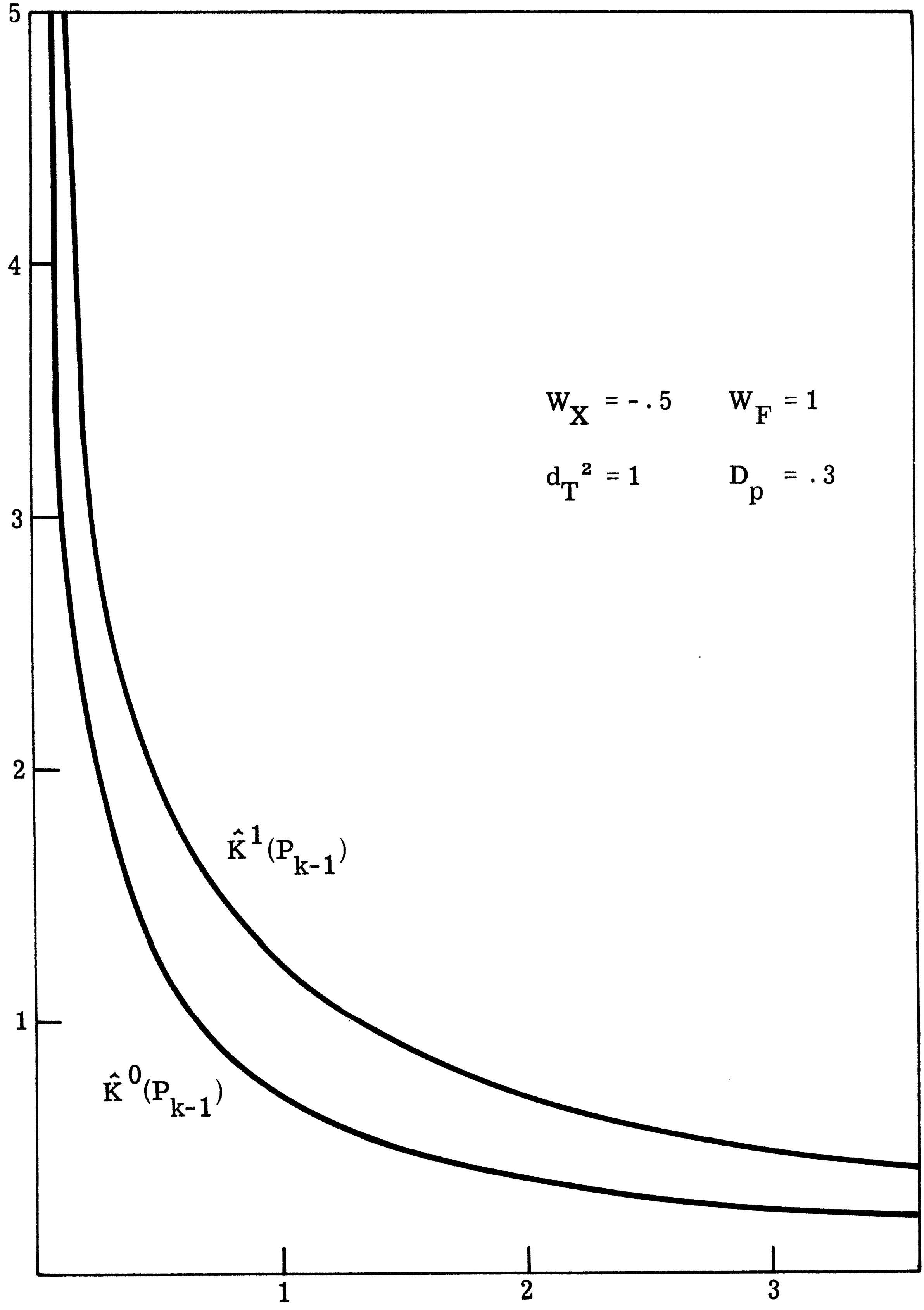


Fig. 5.18. Threshold functions for Case IV; $d_T^2 = 1$, $D_p = .3$, $W_X = -.5$, $W_F = 1$

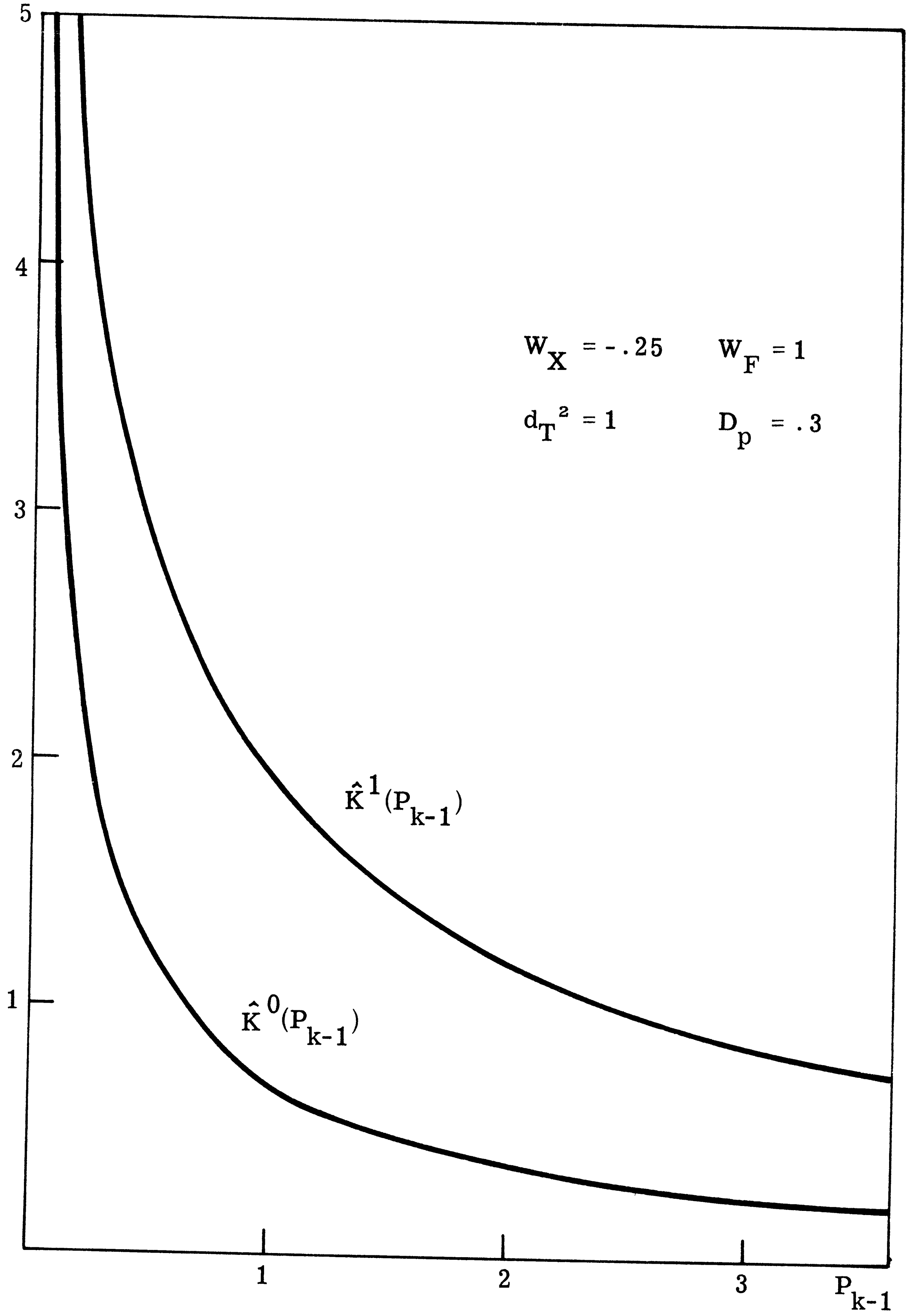


Fig. 5.19. Threshold functions for Case IV; $d_T^2 = 1$, $D_p = .3$, $W_X = -.25$, $W_F = 1$

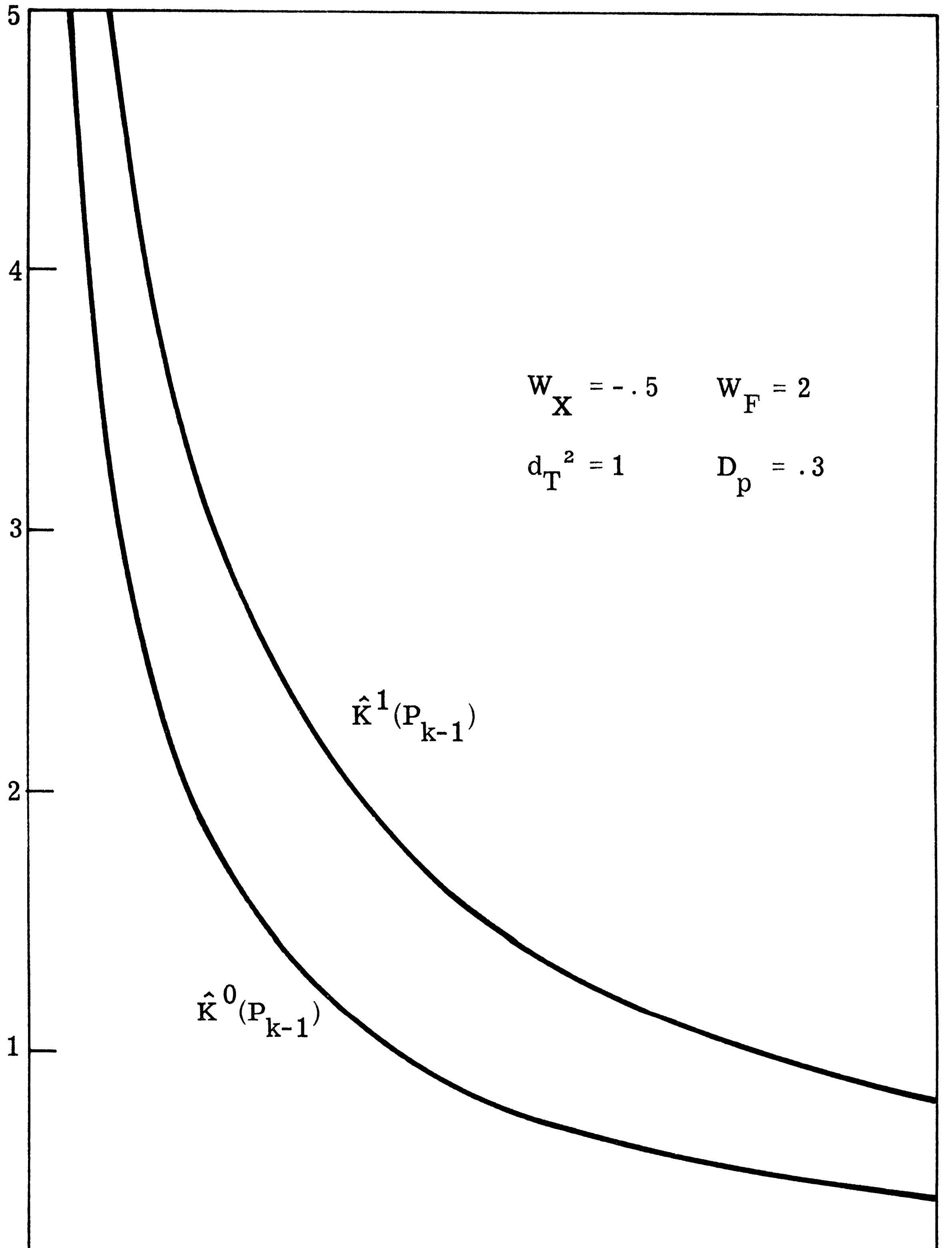


Fig. 5.20. Threshold functions for Case IV; $d_T^2 = 1$, $D_p = .3$, $W_X = -.5$, $W_F = 2$

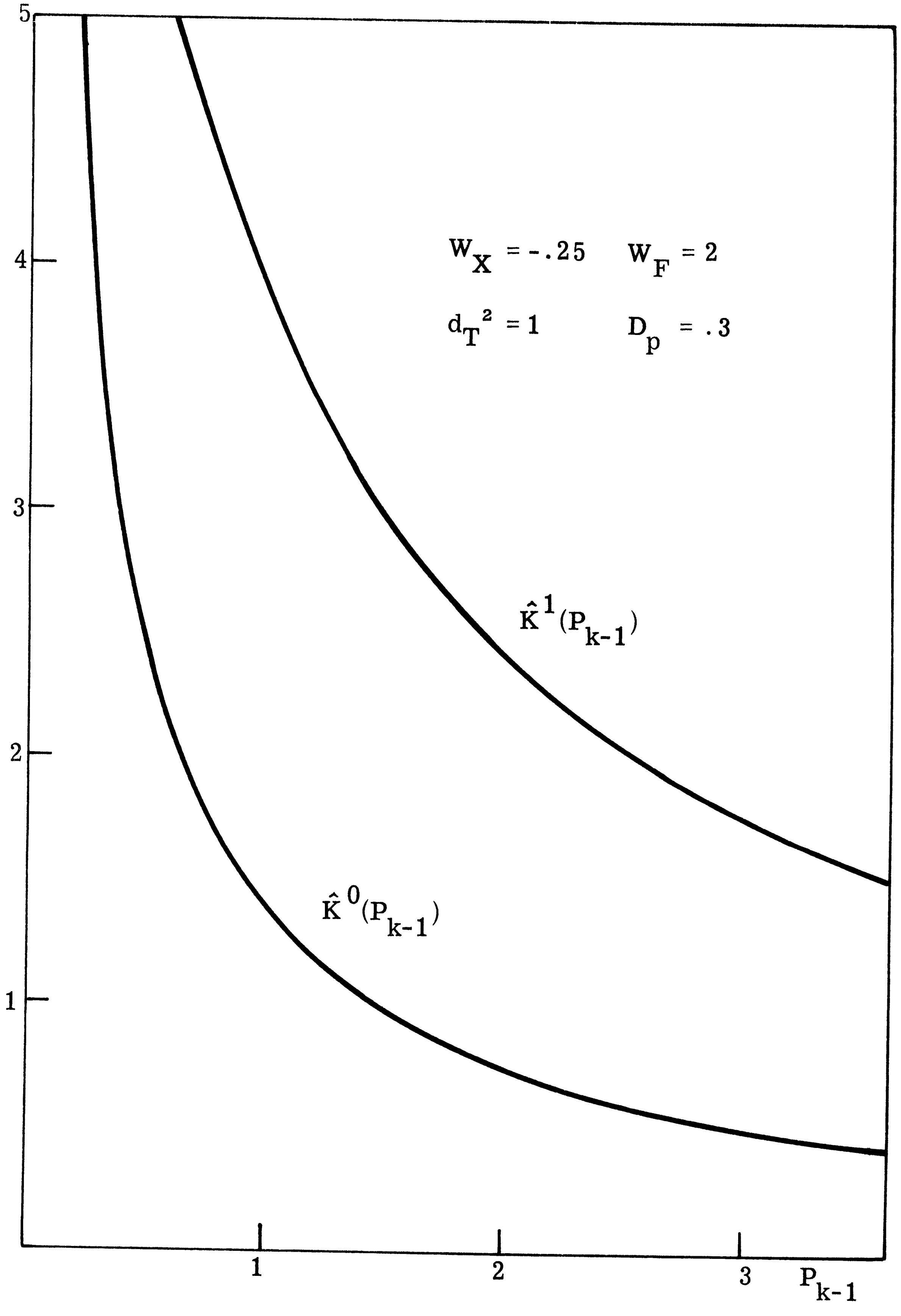


Fig. 5.21. Threshold functions for Case IV; $d_T^2 = 1$, $D_p = .3$, $W_X = -.25$, $W_F = 2$

time t_k is the first opportunity to detect a pulse present at that time, ($P_{k-1} = 0$), then it will still respond at time t_k even if it responded at time t_{k-1}. This behavior is typical of all respond-and-hold decision devices.

Finally, we note that all of the threshold functions satisfy

$$\hat{K}^0(P_{k-1}) < \hat{K}^1(P_{k-1}) \qquad \text{for all } P_{k-1}$$

This condition says that the decision device is more likely to respond at time t_k if it did not respond at the preceding time t_{k-1} then if it did respond at time t_{k-1}. That this is reasonable can be seen by noting that the response condition

$$\ell_k \geq \hat{K}^0(P_{k-1})$$

can result only in a detected pulse outcome or a false alarm, whereas the response condition

$$\ell_k \geq \hat{K}^1(P_{k-1})$$

can also result in an extra detection outcome. But, in the respond-and-hold case the extra detection outcome only results in a gain of Δ units of detection time, whereas the detected pulse outcome results in Δ units of detection time plus the gain for a detected pulse. Thus, it is reasonable to expect that the decision device is less likely to respond if it has just responded at the preceding decision time.

<u>Case V: $W_X = -1$, $W_F > 0$</u>

The Bayes decision devices for these losses are the maximum detection devices. In this case the response condition can be obtained analytically. The procedure is to note from Section 4.2 that the response condition is the same as the $m = 1$ Bayes decision device, except that decisions are made at each decision time t_k. Thus, we may conclude from Eq. 5.15 that

$$\hat{K}^0(P_{k-1}) = \hat{K}^1(P_{k-1}) = K(P_{k-1})$$

$$= \frac{W_F a}{(1-a) + P_{k-1}}$$

An example of these threshold functions is shown in Fig. 5.22.

We might point out that since both of the threshold functions are equal, the MDT decision device does not need to store the preceding decision. Thus, the decision feedback loop can be removed from the general block diagram of Fig. 5.8. The resulting block diagram is shown in Fig. 5.23.

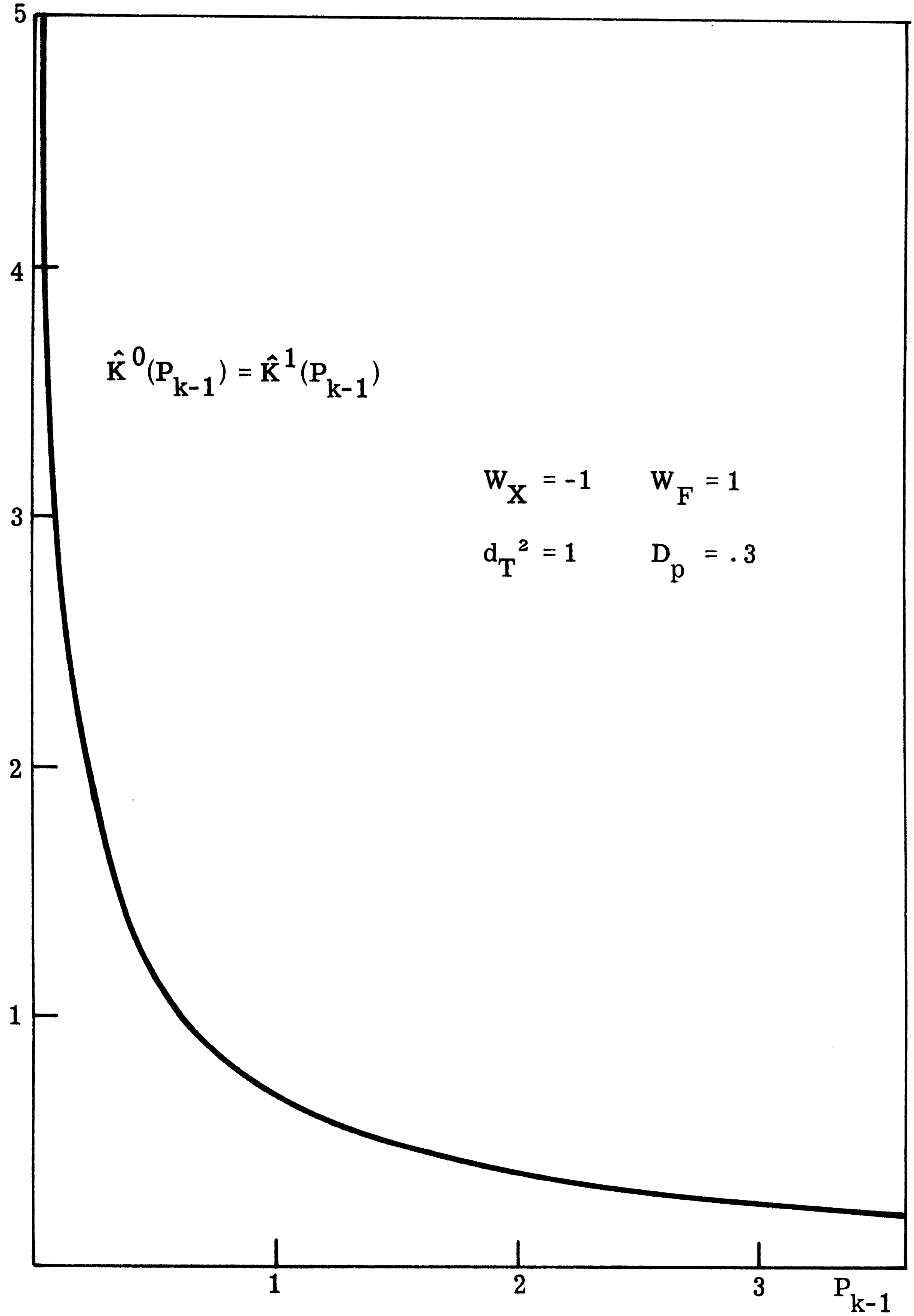


Fig. 5.22. Threshold functions for Case V;
$d_T{}^2 = 1$, $D_p = .3$, $W_X = -1$, $W_F = 1$

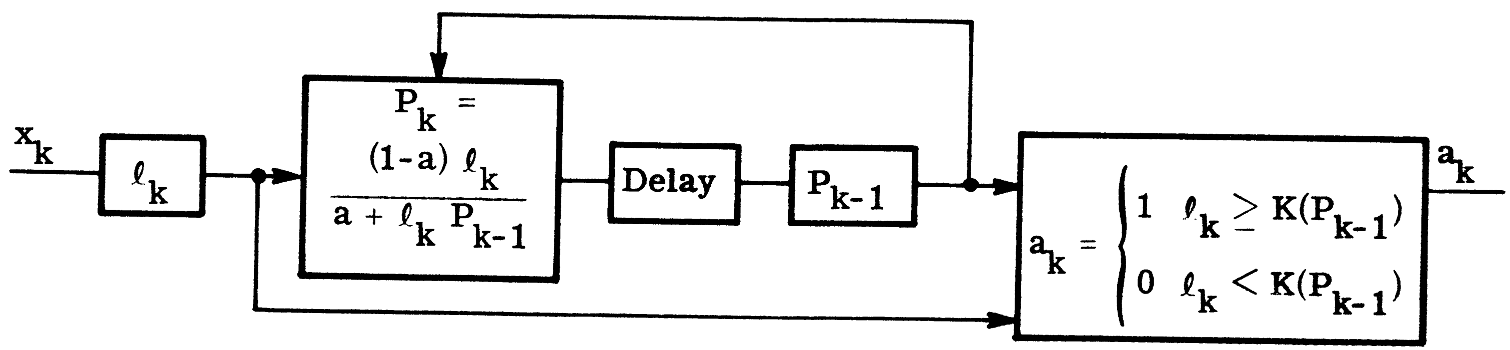


Fig. 5.23. Block diagram for $m = 2$ maximum detection time decision device

In concluding this section, we note that all of the above examples for the five cases were calculated for the signal-to-noise ratio $d_T^2 = 1$ and the duty $D_p = .3$. In general, the basic form of these functions does not depend on d_T^2 and D_p, but the functions themselves decrease with increasing d_T^2 or increasing D_p. The one exception is the MDT threshold function which depends only on D_p.

Examples of this behavior appear in Figs. 5.24 through 5.27. The decrease in the threshold functions for an increase in D_p can be seen by comparing Figs. 5.24 and 5.25 with Figs. 5.12 and 5.18. The decrease in the threshold functions for an increase in d_T can be seen by comparing Figs. 5.26 and 5.27 with Figs. 5.12 and 5.18.

5.4 The Matched Filter Decision Devices

In Section 1.4 we introduced a matched filter (MF) decision device for both the respond-once and the respond-and-hold case. There we mentioned that, although these devices are suboptimal, they are quite practical from an engineering point of view. In this section, we will adapt these decision devices to the basic setting of this chapter. This will provide a common basis for comparing the structure and the performance of these devices with the optimum $m = 1$ and $m = 2$ Bayes decision devices.

We begin by recalling from Section 1.4 that at the decision time t_k both the MF devices calculate the quantity

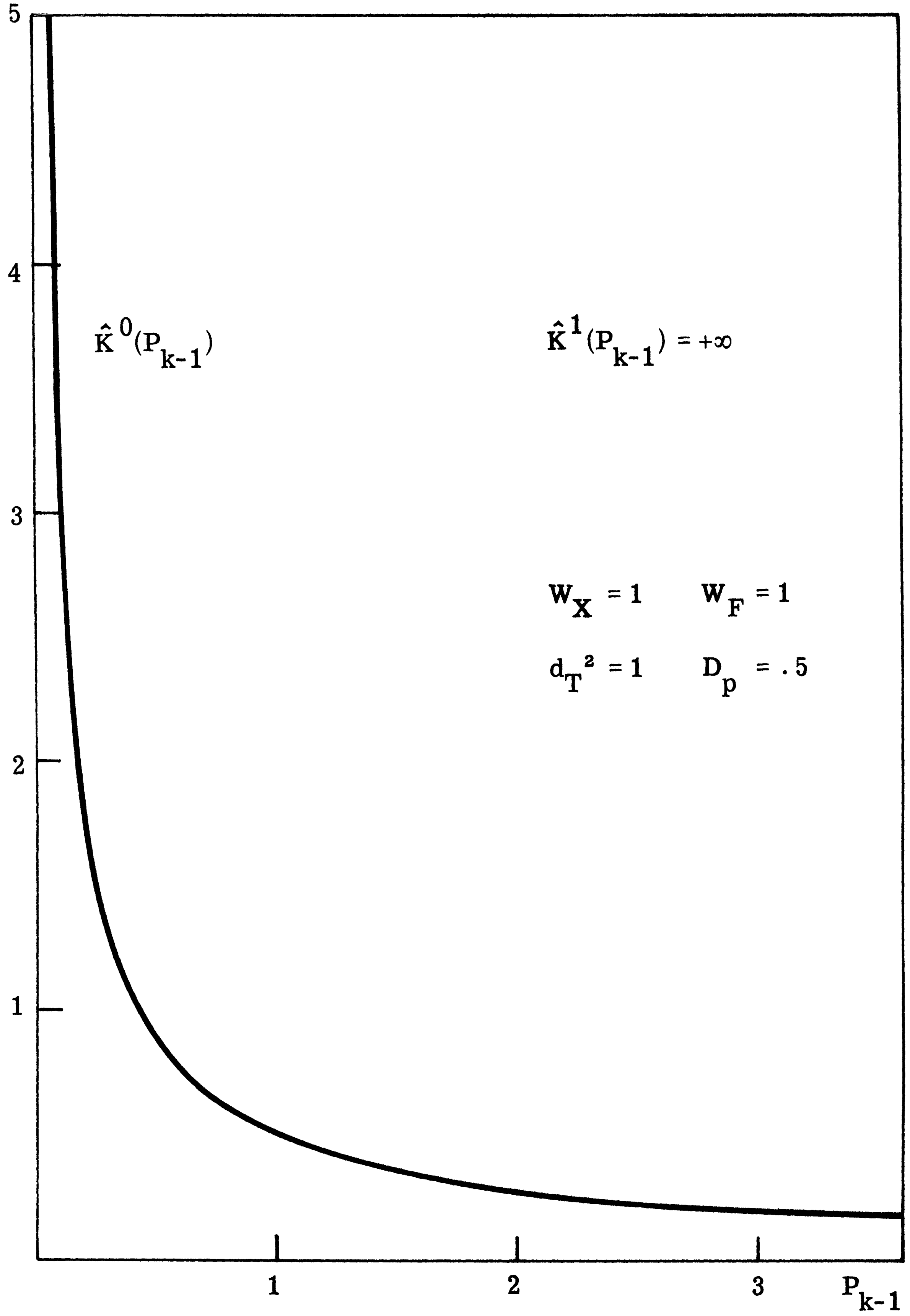


Fig. 5.24. Threshold functions for Case I;
$d_T^2 = 1$, $D_p = .5$, $W_X = 1$, $W_F = 1$

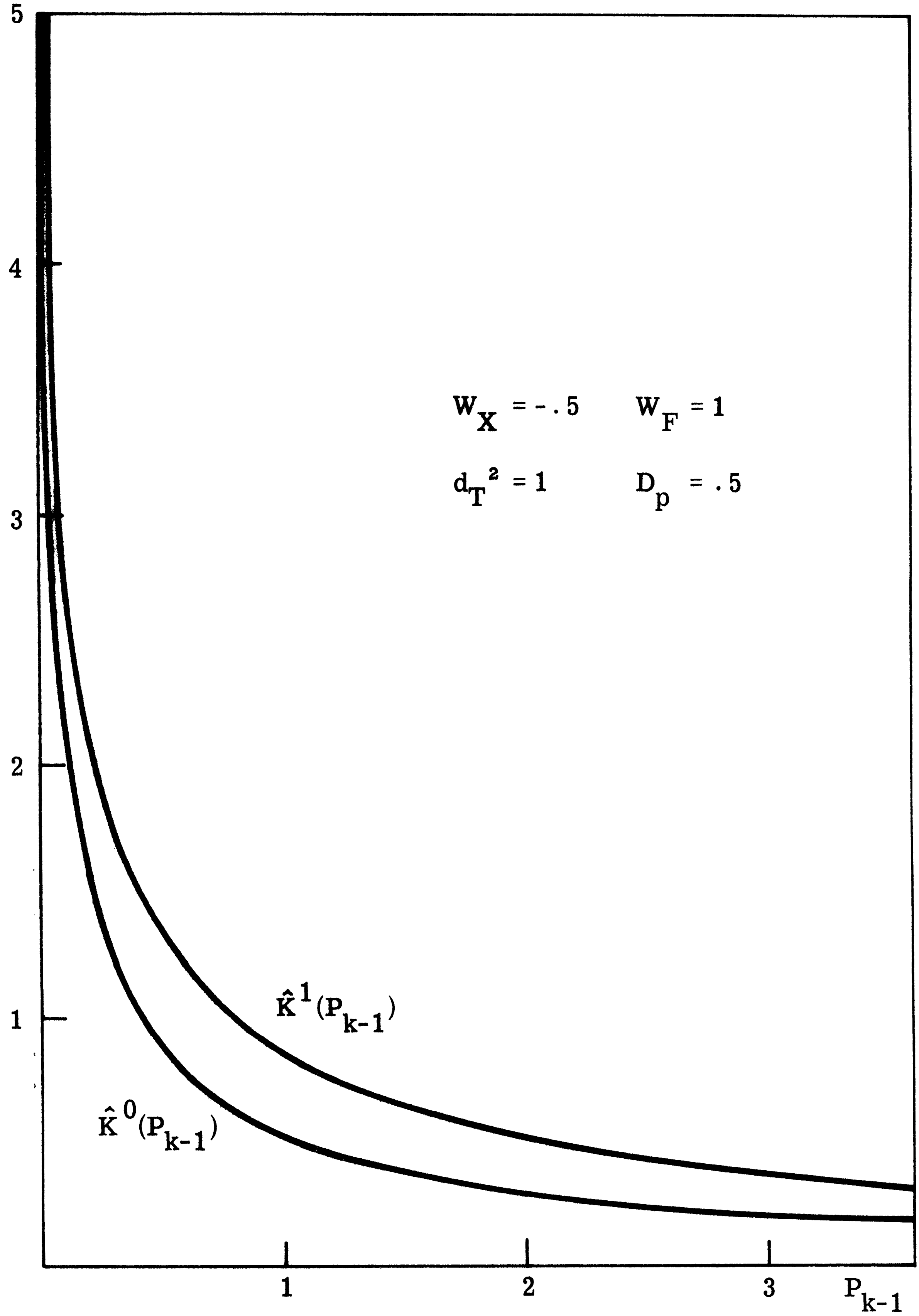


Fig. 5.25. Threshold functions for Case IV; $d_T^2 = 1$, $D_p = .5$, $W_X = -.5$, $W_F = 1$

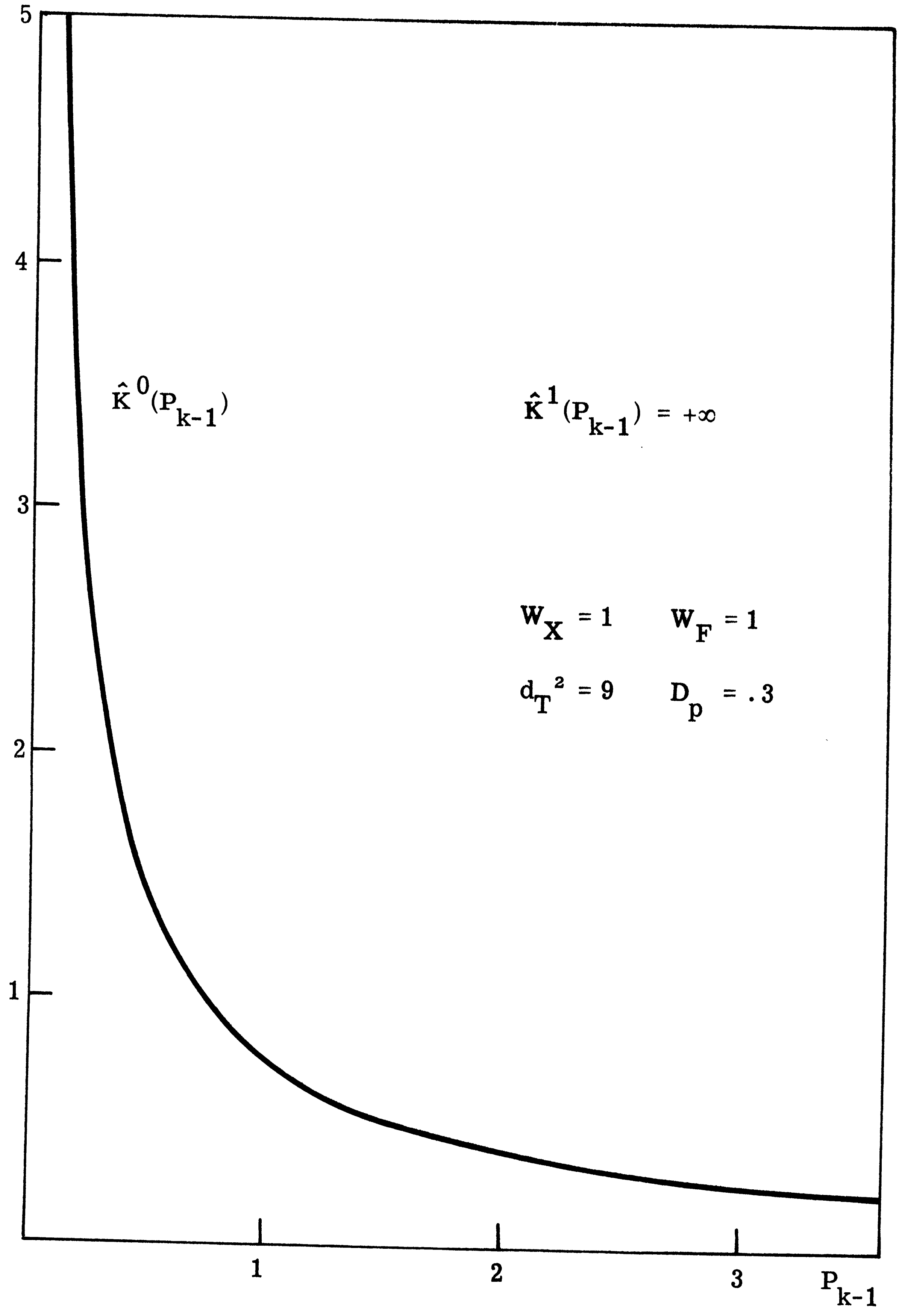


Fig. 5.26. Threshold functions for Case I; $d_T^2 = 9$, $D_p = .3$, $W_X = 1$, $W_F = 1$

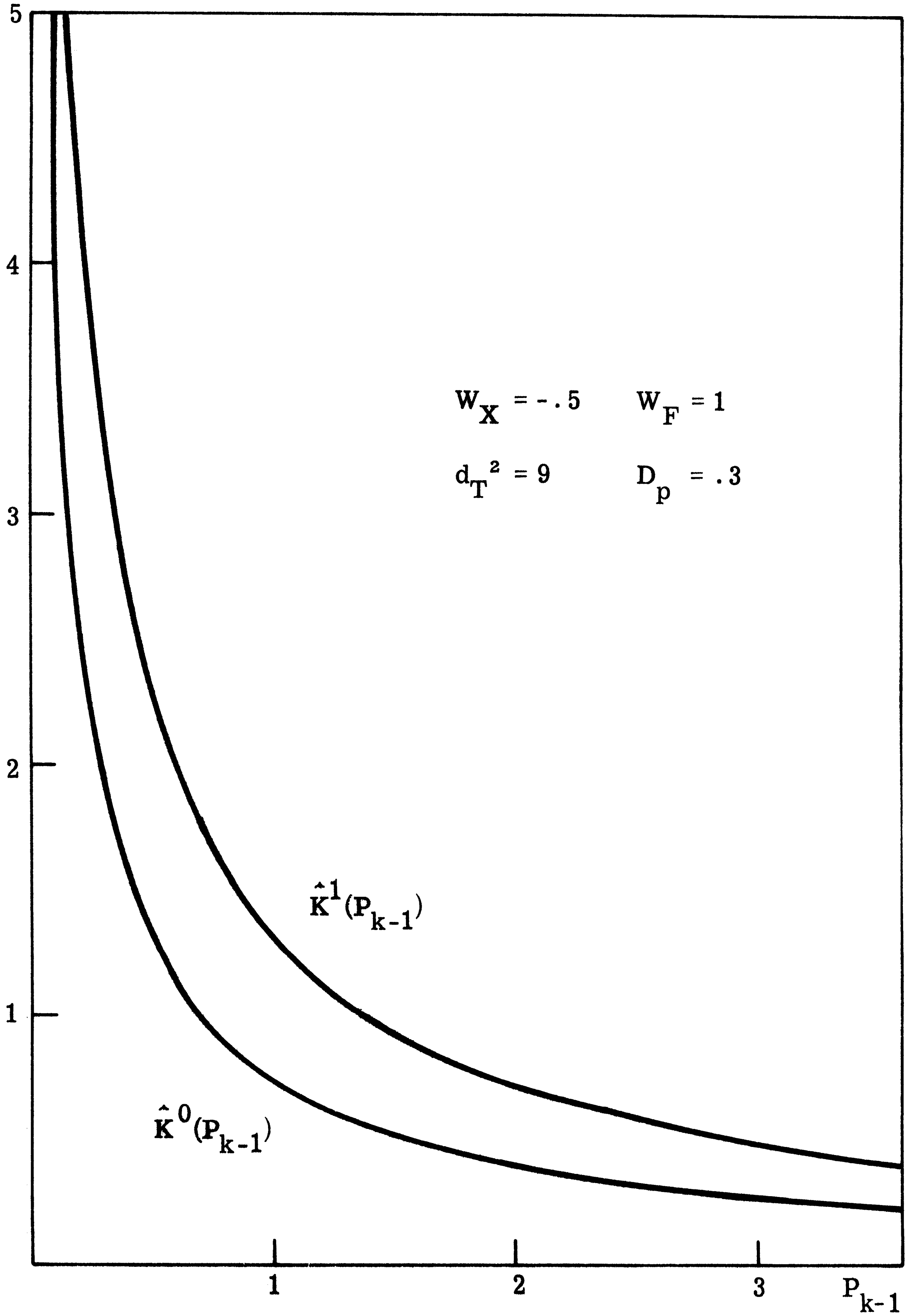


Fig. 5.27. Threshold functions for Case IV; $d_T^2 = 9$, $D_p = .3$, $W_X = -.5$, $W_F = 1$

$$C_{t_k}(x(t''); t'' < t_k) = \int_{t_k - T_p}^{t_k} x(t)\, p(t - (t_k - T_p))\, dt$$

To interpret this quantity in terms of the assumptions of this chapter, we first break the integral into two parts, and then use the pulse translation property and the fact that $t_k = t_{k-1} + t_p/2$ to write

$$C_{t_k} = \int_{t_k - T_p}^{t_k - T_p/2} x(t)\, p(t - (t_k - T_p))\, dt + \int_{t_k - T_p/2}^{t_k} x(t)\, p(t - (t_k - T_p))\, dt$$

$$= \int_{t_k - T_p}^{t_{k-1}} x(t)\, p(t - (t_{k-1} - T_p))\, dt + \int_{t_k - T_p/2}^{t_k} x(t)\, p(t - (t_k - T_p))\, dt$$

or

$$\frac{C_k}{N_0} = \frac{1}{N_0} \int_{t_{k-1} - T_p/2}^{t_{k-1}} x(t)\, p^{k-1}(t)\, dt + \frac{1}{N_0} \int_{t_k - T_p/2}^{t_k} x(t)\, p^{k}(t)\, dt \qquad (5.16)$$

Next, add and subtract

$$\frac{d_T^{\,2}}{2} = \frac{1}{2N_0} \int_{t_k - T_p}^{t_k} [p^k(t)]^2\, dt$$

$$= \frac{1}{2N_0} \int_{t_{k-1} - T_p/2}^{t_{k-1}} [p^{k-1}(t)]^2\, dt + \frac{1}{2N_0} \int_{t_k - T_p/2}^{t_k} [p^k(t)]^2\, dt$$

to Eq. 5.16 to obtain

$$\frac{C_{t_k}}{N_0} = \left(\frac{1}{N_0} \int_{t_{k-1} - T_p/2}^{t_{k-1}} x(t)\, p^{k-1}(t)\, dt - \frac{1}{2N_0} \int_{t_{k-1} - T_p/2}^{t_{k-1}} [p^{k-1}(t)]^2\, dt \right)$$

$$+ \left(\frac{1}{N_0} \int_{t_k - T_p/2}^{t_k} x(t)\, p^k(t)\, dt - \frac{1}{2N_0} \int_{t_k - T_p/2}^{t_k} [p^k(t)]^2\, dt \right)$$

$$+ \frac{d_T^2}{2}$$

or

$$\frac{C_{t_k}}{N_0} - \frac{d_T^2}{2} = \ln \ell_{k-1}(x_{k-1}) + \ln \ell_k(x_k) \qquad (5.17)$$

$$= \ln [\ell_{k-1}(x_{k-1}) \cdot \ell_k(x_k)]$$

The next step is to replace the response condition

$$C_{t_k} \geq \beta$$

by the equivalent response condition

$$\exp \left\{ \frac{C_{t_k}}{N_0} - \frac{d_T^2}{2} \right\} \geq \exp \left\{ \frac{B}{N_0} - \frac{d_T^2}{2} \right\} = \beta'$$

Now, substitute Eq. 5.17 into the above expression to write the response condition as

$$\ell_{k-1}(x_{k-1})\, \ell_k(x_k) \geq \beta'$$

or finally, as

$$\ell_k(x_k) \geq \beta' / \ell_k(x_{k-1}) \qquad (5.18)$$

We are now in a position to describe the MF decision devices. To this end, note from Section 1.4 that the RHMF responds whenever inequality (5.18) is satisfied. On the other hand, the ROMF device must also satisfy the inhibit rule. Thus, for this device the response condition is

$$\ell_k(x_k) \geq \begin{cases} \beta' / \ell_k(x_k) & \text{if} \quad a_{k-1} = 0 \\ +\infty & \text{if} \quad a_{k-1} = 1 \end{cases}$$

The block diagrams for these devices are shown in Fig. 5.28.

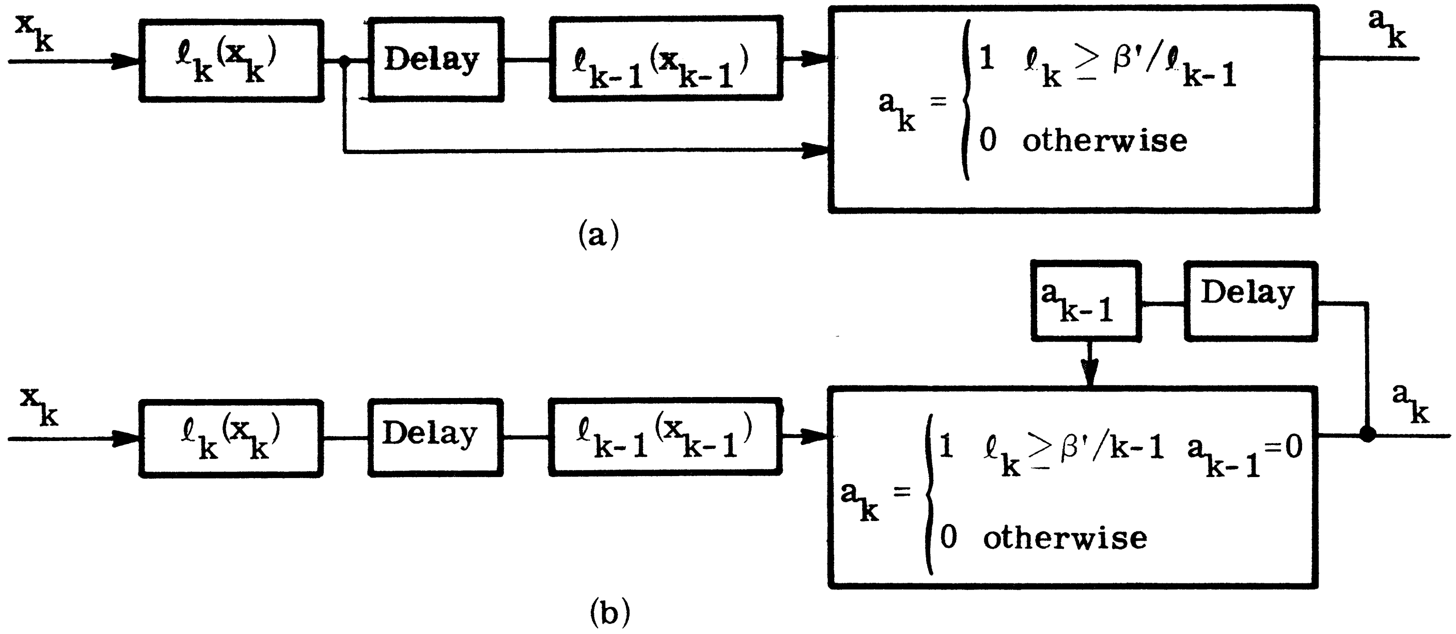


Fig. 5.28. The MF decision device
(a) respond-and-hold case, (b) respond-once case

We might point out here that these block diagrams are essentially the same as the block diagrams for the $m = 2$ Bayes decision devices; the main difference here is that the threshold function β'/ℓ_k does not depend on P_{k-1} .

5.5 Performance

In Chapter III it was seen that the performance of the FRD decision devices can be described in terms of either the respond-once performance set $\mathscr{P}(m)$ or the respond-and-hold performance set $\tilde{\mathscr{P}}(m)$. Moreover, the optimum performance for these devices is described in terms of either of the two ROC surfaces $\lambda_{\mathscr{P}(m)}$ or $\lambda_{\tilde{\mathscr{P}}(m)}$. In this section, we present a numerical description of these surfaces for the $m = 2$ and the $m = 1$ case. Also, we will present a numerical description of the performance for the MF decision devices of the preceding section.

To obtain the description of the ROC surfaces, it is sufficient to determine the performance points

$$R(\delta) = (R_X(\delta),\ R_F(\delta),\ R_D(\delta))$$

and

$$D(\delta) = (D_D(\delta),\ D_F(\delta),\ R_D(\delta))$$

for the Bayes decision devices. This has been done by simulating the operation of these devices over a long period of time and measuring the resulting values of $R(\delta)$ and $D(\delta)$. The details of this procedure are found in Appendix M. The performance description for the MF decision devices has also been obtained by this procedure.

Before presenting the numerical results, we will introduce certain bounds on the ROC surfaces. First, we consider the bounds for the respond-once ROC surface $\lambda_{\mathscr{S}(2)}$. To this end, we assert that the performance of a respond-once decision device cannot exceed the performance of the classical signal-known-exactly (SKE) decision device with the same signal-to-noise ratio d_T^2 . This follows from the fact that in the SKE paradigm the decision device always bases its decision on an observation that is influenced by the whole pulse, whereas in the FRD paradigm the decision device must sometimes make decisions on an observation that is influenced by only part of the pulse. The basic SKE paradigm is shown in Fig. 5.29.

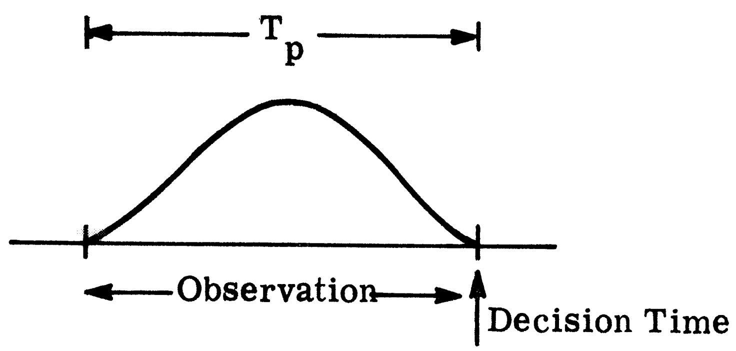


Fig. 5.29 . The SKE paradigm for signal-to-noise ratio d_T^2

Next, we assert that in the $m = 2$ case the performance of the optimum respond-once decision device will always exceed the performance of the optimum SKE decision device that bases its decision on only half of the signal energy. The basic paradigm here is shown in Fig. 5.30.

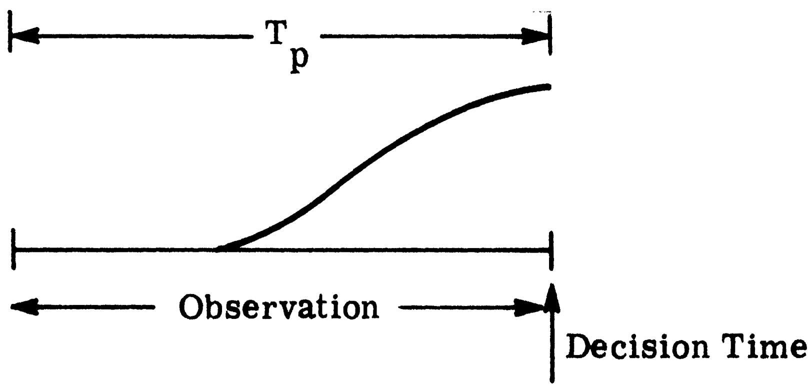


Fig. 5.30. The SKE paradigm for signal-to-noise ratio $d_T^2/2$

The validity of this assertion follows from the fact that the $m = 2$ respond-once decision device always bases its decisions on at least half of the signal energy.

To interpret the above statements in terms of the ROC surface $\lambda_{\mathscr{S}(2)}$, it is necessary to relate the probability of detection P_D and the probability of false alarm P_F for the SKE decision device to the rates R_X, R_F and R_D. This is done in Appendix N. There it is seen that, for the SKE decision device,

$$R_X = 0$$

$$R_F = P_F$$

$$R_D = P_D$$

It then follows from the above statements that the projection of $\lambda_{\mathscr{S}(2)}$ onto the $R_X = 0$ plane must not exceed the ROC curve for the SKE case with signal-to-noise ratio d_T^2 and must lie on or above the ROC curve for the SKE case with signal-to-noise ratio $d_T^2/2$. Thus this projection must lie completely in the region shown in Fig. 5.31 below. It is to be noted here that these curves, as well as the remaining curves in this section, are plotted on normal-normal paper.

We may use the same argument as above to obtain an upper bound for the ROC surface $\lambda_{\tilde{\mathscr{S}}(2)}$. This is done by associating T_p seconds of detection time to each detection decision and T_p seconds of false alarm time to each false alarm decision. It then follows that the projection of the ROC surface, $\lambda_{\tilde{\mathscr{S}}(2)}$, onto the $R_D = 0$ plane cannot exceed the ROC curve for the SKE decision device with the same signal-to-noise ratio. This latter curve is obtained by noting from Appendix N that for the SKE decision device

$$D_D = P_D$$

$$D_F = P_F$$

$$R_D = P_D$$

The region containing the projection of $\lambda_{\tilde{\mathscr{S}}(2)}$ is shown in Fig. 5.32.

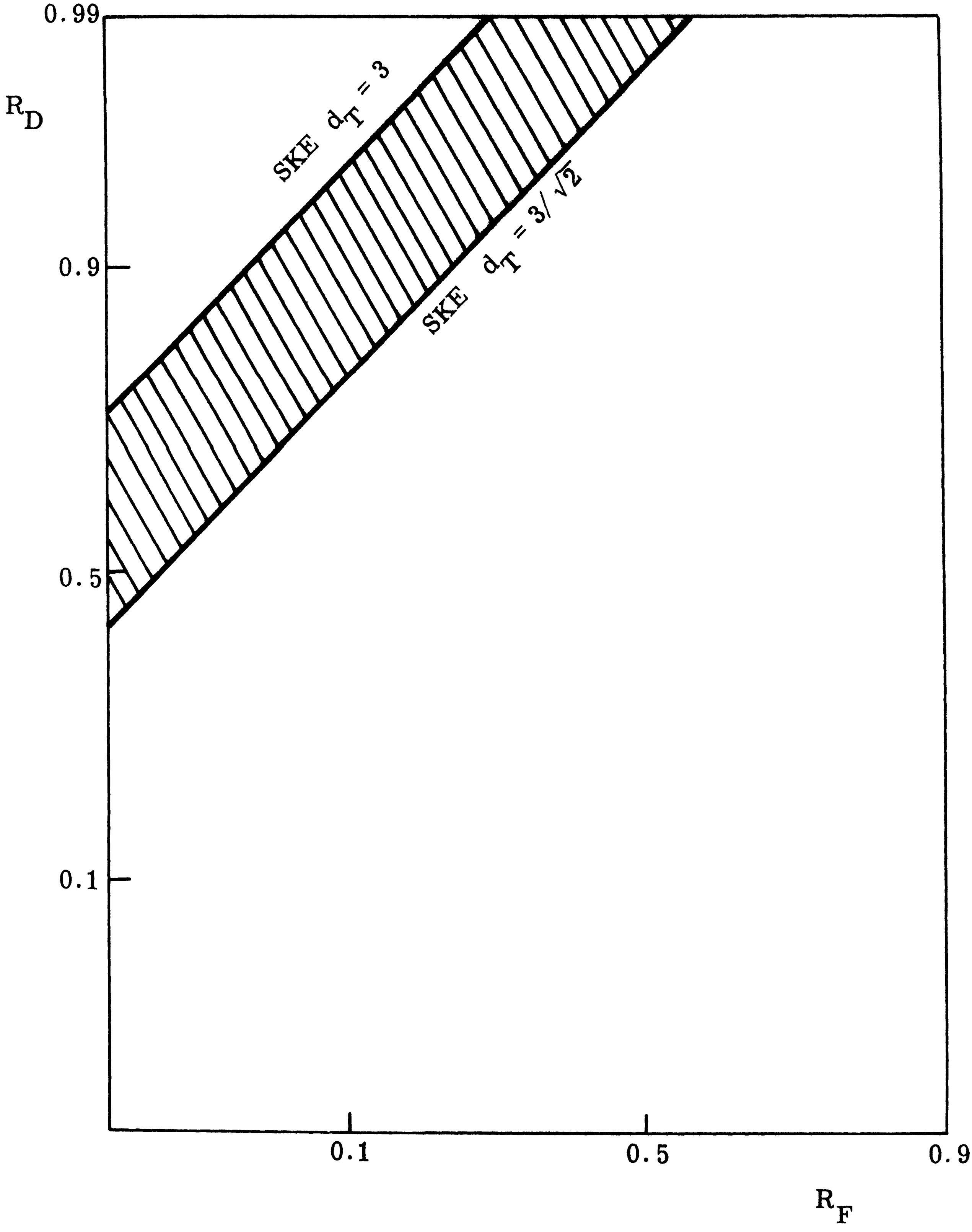


Fig. 5.31. The region containing the projection of $\lambda_{\mathscr{S}(2)}$

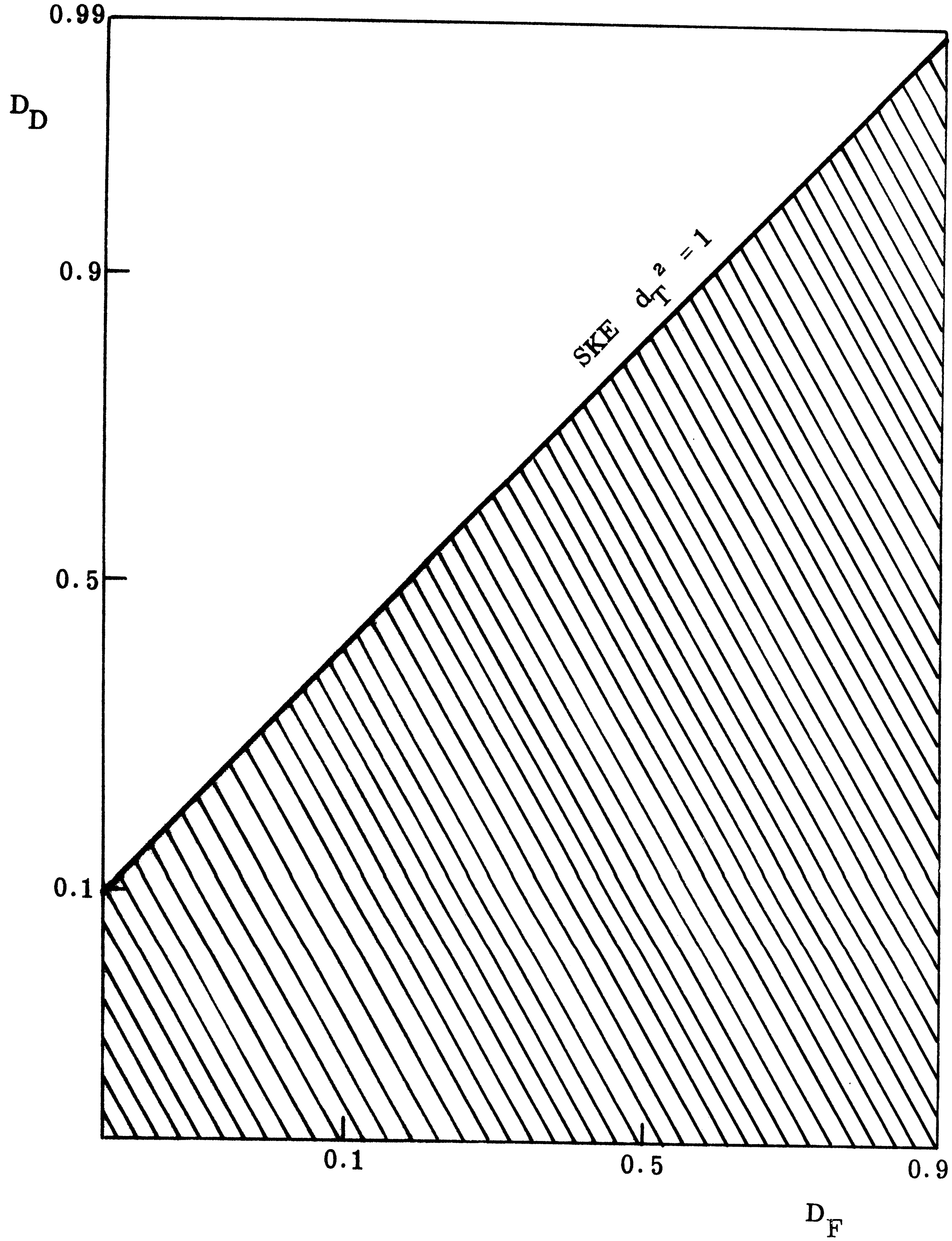


Fig. 5.32. The region containing the projection of $\lambda_{\tilde{\mathscr{S}}(2)}$.

We now turn to the description of the performance of the respond-once decision devices. First we consider the $m = 2$ decision device. Recall from Section 3.6 that the $m = 2$ Bayes decision devices can be classified into three special cases. The Case I decision devices satisfy the inhibit rule and thus their performance plots as a curve in the $R_X = 0$ plane. This curve has the property that it represents the smallest values of the detection rate R_D. The Case III decision devices are the MPD devices. These devices do not satisfy the inhibit rule and thus their performance plots as a curve in the region $R_X > 0$. This curve has the property that it represents the largest values of R_D. Finally, the Case II devices are those devices that achieve a trade-off between R_D and R_X. These devices also do not satisfy the inhibit rule. Their performance constitutes the rest of the ROC surface $\lambda_{\mathscr{S}(2)}$ in the region $R_X > 0$.

To illustrate the performance in the $m = 2$ case, we have plotted the inhibit rule curve and the projection of the MPD curve into the $R_X = 0$ plane. The area bounded by these curves can then be considered as the projection of $\lambda_{\mathscr{S}(2)}$ into the $R_X = 0$ plane. These curves are shown in Fig. 5.33 for $d_T = \sqrt{d_T^2} = 3$ along with the upper and lower bounds discussed above.

In Fig. 5.33, it is seen that for the low signal-to-noise ratio, ($d_T{}^2 = 1$), the surface $\lambda_{\mathscr{S}(2)}$ is essentially independent of R_X, since the projection of the MPD performance curve and the inhibit rule

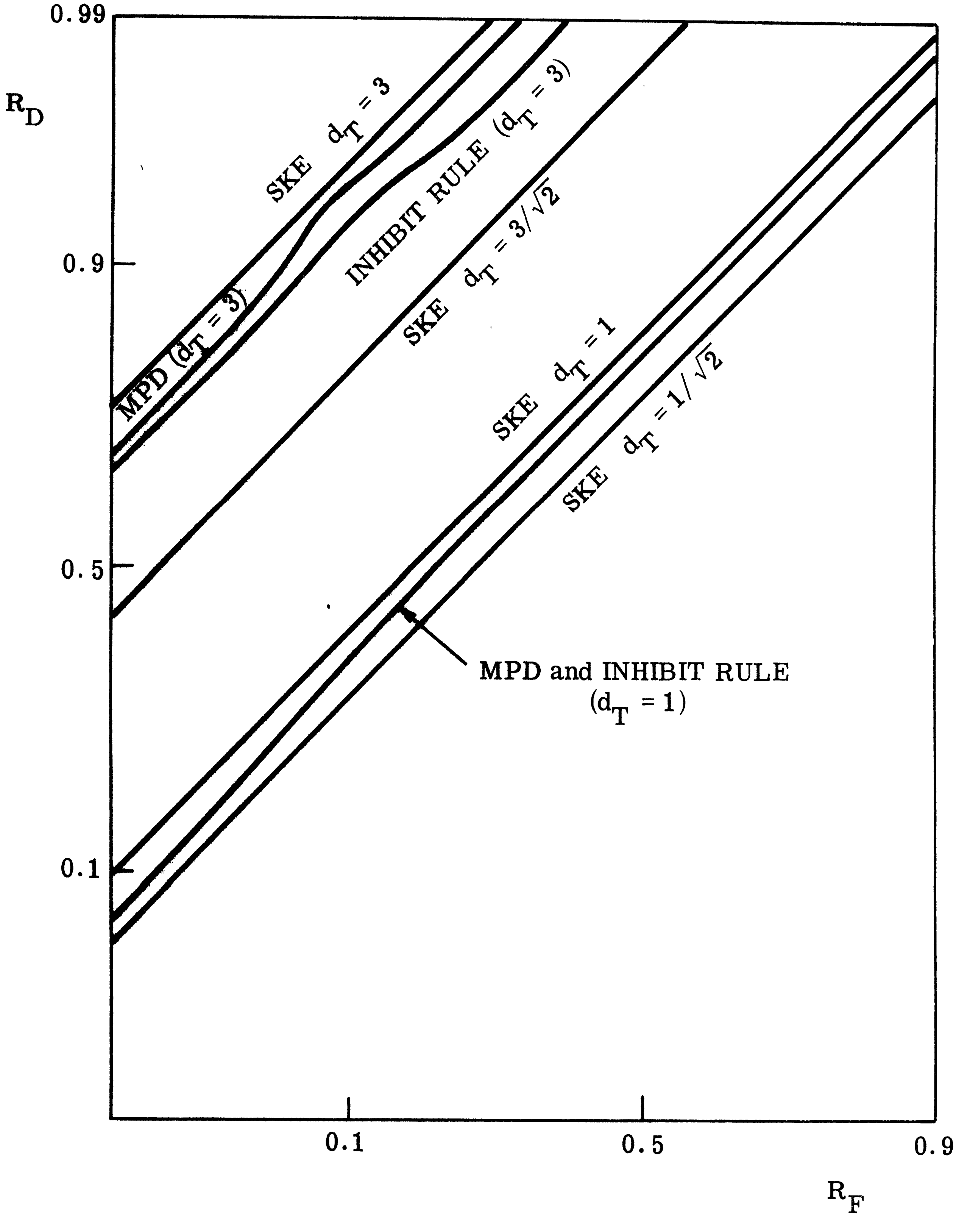


Fig. 5.33. The projection of $\lambda\mathscr{S}_{(2)}$ into the $R_X = 0$ plane for $d_T = 1, 3$.

performance curve coincide. Thus, for low signal-to-noise ratios, one can expect little or no improvement in the detection rate R_D for increases in the extra detection rate R_X. On the other hand, if the signal-to-noise ratio is large, $(d_T^2 = 9)$, there is a noticeable difference in the two curves, so that in this case one can expect a substantial increase in R_D for an increase in R_X.

Next, we consider the performance of the $m = 1$ R-O decision devices. In this case the ROC surface $\lambda_{\mathscr{P}(1)}$ degenerates to a curve in the $R_X = 0$ plane, since these devices are incapable of making extra detections. Figure 5.34 illustrates $\lambda_{\mathscr{P}(1)}$ for $d_T = 1, 3$ along with the upper and lower bounds.

The main point to be made here is that these curves lie only slightly above the lower bound and, as can be seen by comparing Fig. 5.34 with Fig. 5.33, they lie substantially below the projection of $\lambda_{\mathscr{P}(2)}$. Thus, as is to be expected from Theorem 3.9, there is a substantial increase in the optimum performance for an increase in the number of detection opportunities m.

Finally, we consider the performance of the ROMF decision devices. As discussed in the preceding section, these devices also have two decision opportunities per pulse, $(m = 2)$, and they also satisfy the inhibit rule so that their performance plots as a curve in the $R_X = 0$ plane. Furthermore, it should be noted that since these devices are suboptimal, their performance can be expected to depend on the duty D_p.

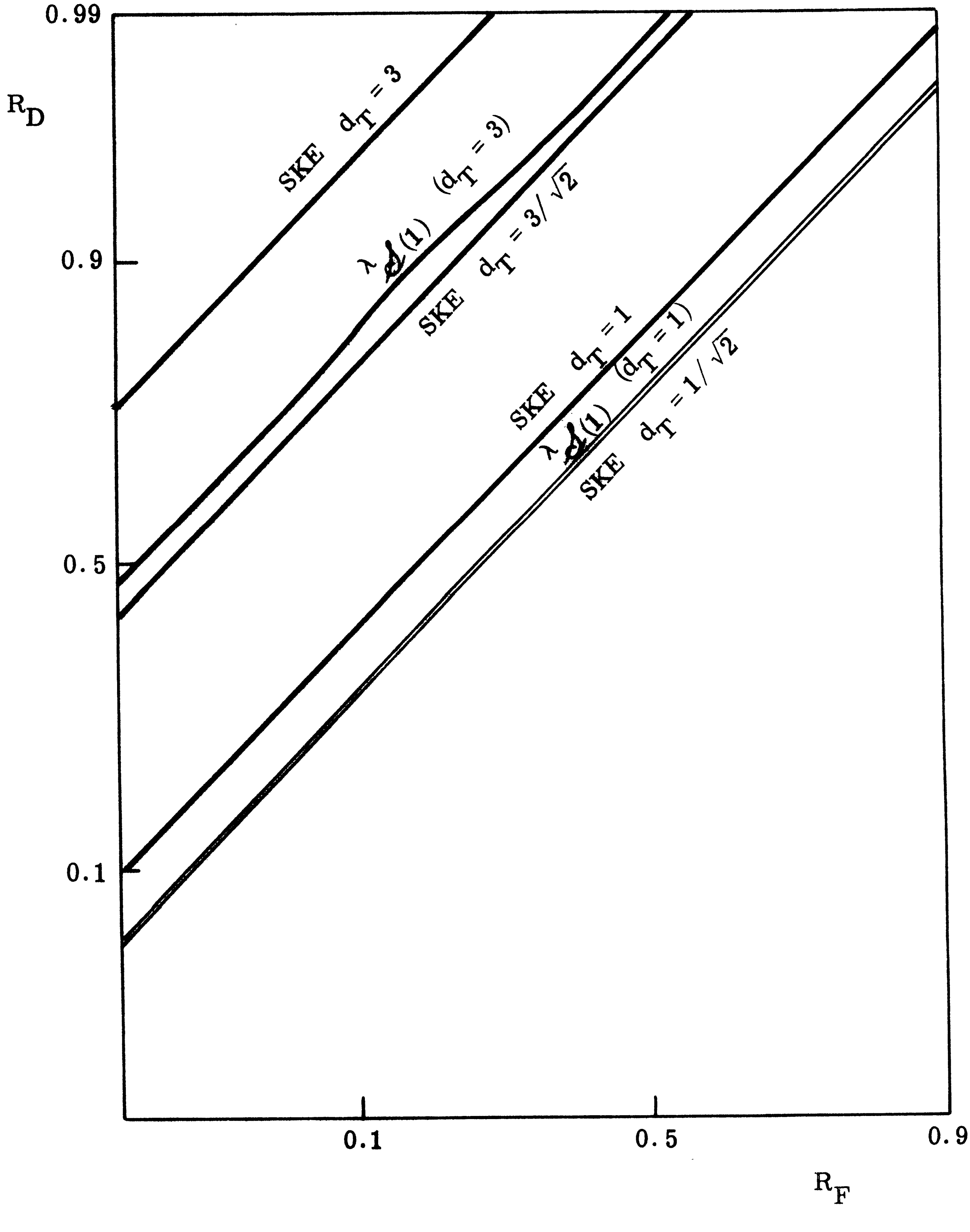


Fig. 5.34. The ROC curve $\lambda \mathscr{S}_{(1)}$

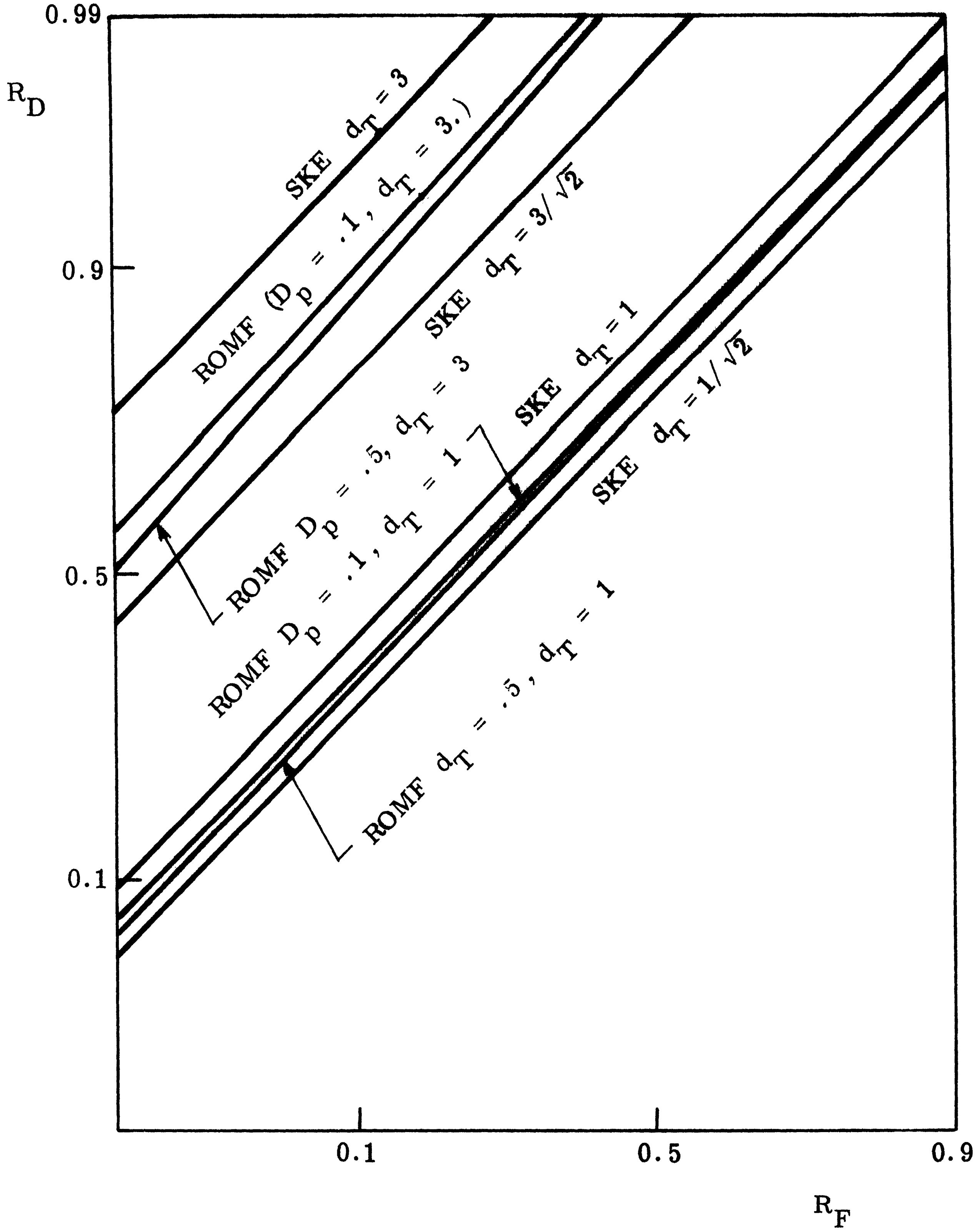


Fig. 5.35. Performance curves for the ROMF decision devices.

This is seen to be the case in Fig. 5.35, where we have plotted the ROMF performance curves for $d_T = 1, 3$, and $D_p = .1, .5$, along with the upper and lower bounds.

It should be noted from Fig. 5.35 that the performance of the ROMF device for $D_p = .1$ shows a substantial improvement over the performance for $D_p = .5$ for the higher signal-to-noise ratio.

We conclude the respond-once case by comparing the performance of the three different R-O decision devices for $d_T = 3$. This is done in Fig. 5.36. In this figure it is seen, as expected, that the ROMF performance curves lie below the projection of $\lambda_{\mathscr{S}(2)}$. On the other hand, these curves lie above $\lambda_{\mathscr{S}(1)}$. Thus we have an example of a suboptimum $m = 2$ device whose performance exceeds the performance of the optimum $m = 1$ devices.

We turn now to the performance of the respond-and-hold decision devices. First, we treat the $m = 2$ decision devices. Recall from Section 3.6 that the R-H Bayes decision devices have been divided into two cases, Case IV and Case V. The Case V decision devices are the MDT devices, $(W_X = -1)$. These devices achieve the maximum values of D_D for fixed D_F. The Case IV decision devices are those devices that sacrifice detection time for an increased R_D. Their performance constitutes the rest of the ROC surface $\lambda_{\tilde{\mathscr{S}}(2)}$. To illustrate $\lambda_{\tilde{\mathscr{S}}(2)}$ we have projected the MDT performance curve and the curves for the Bayes decision devices for $W_X = -.5$, $W_X = -1.$,

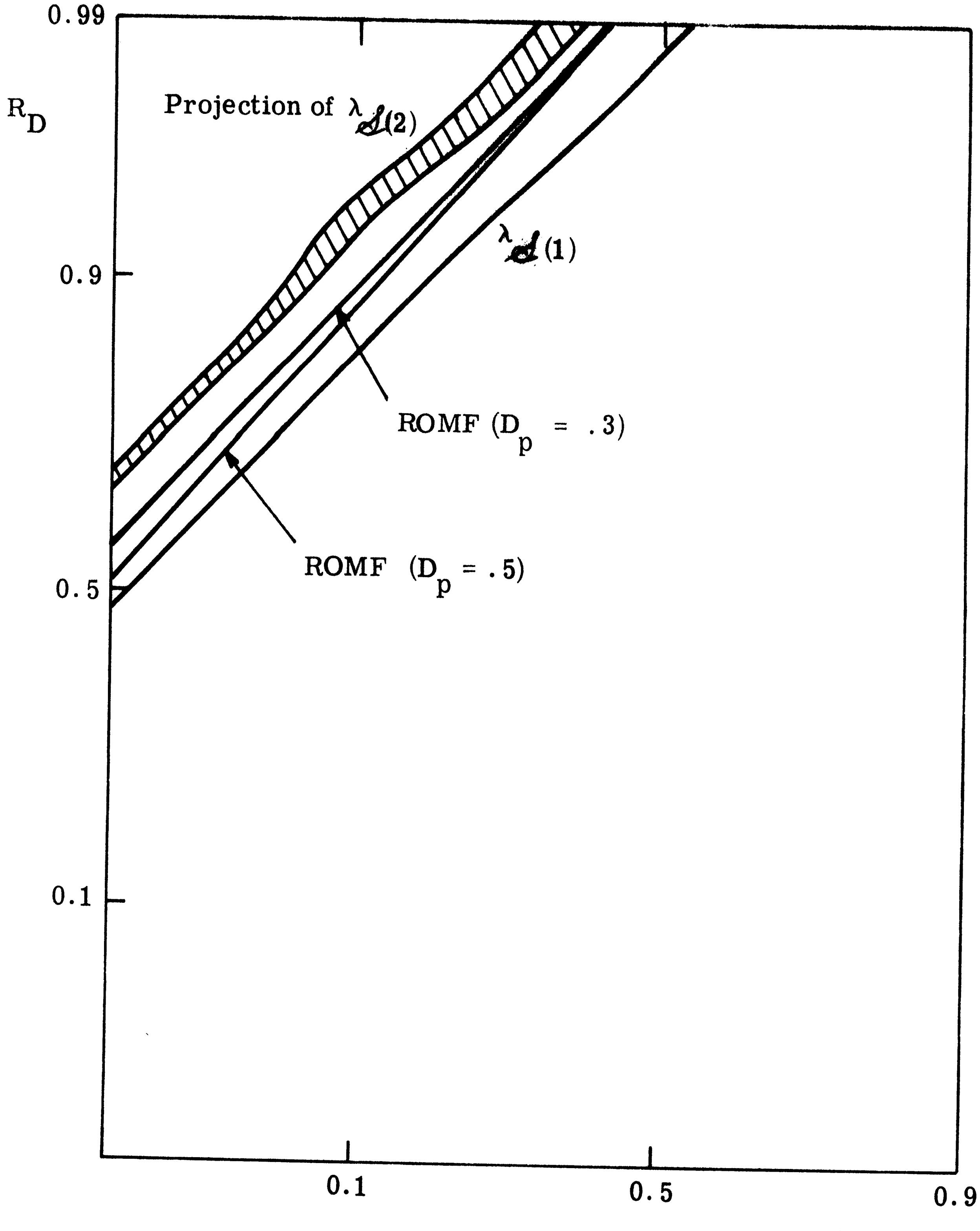


Fig. 5.36. The performance of the R-O decision devices.

and $W_X = 0$ into the $R_D = 0$ plane and the $D_D = 0$ plane. These curves are shown in Figs. 5.37a, b and 5.38a, b for $d_T = 1, 3$.

Here it is seen that the MDT has the largest values of D_D, but the smallest values of R_D, for fixed D_F. On the other hand, the curve associated with $W_X = 0$ has the smallest values of D_D and the largest values of R_D for fixed D_F. This is to be expected from the results of Chapter III. We might also point out here that the projection of the $W_X = 0$ curve into the $R_D = 0$ plane passes through the point

$$R_D = 1$$

$$D_F = .5$$

$$D_D = .5$$

This can be verified analytically by noting that this point is the performance point of the decision device that responds every other time. (See Section 3.4.)

Next we consider the performance of the RHMF. As in the respond-once case, this device is a suboptimal $m = 2$ device whose performance depends on the duty D_p. The projections of the performance curves are shown in Figs. 5.39 and 5.40 for $d_T = 1, 3$, and $D = .1, .3$.

By comparing these curves with the curves in Figs. 5.37 and 5.38, it is seen that there is a substantial degradation in the performance

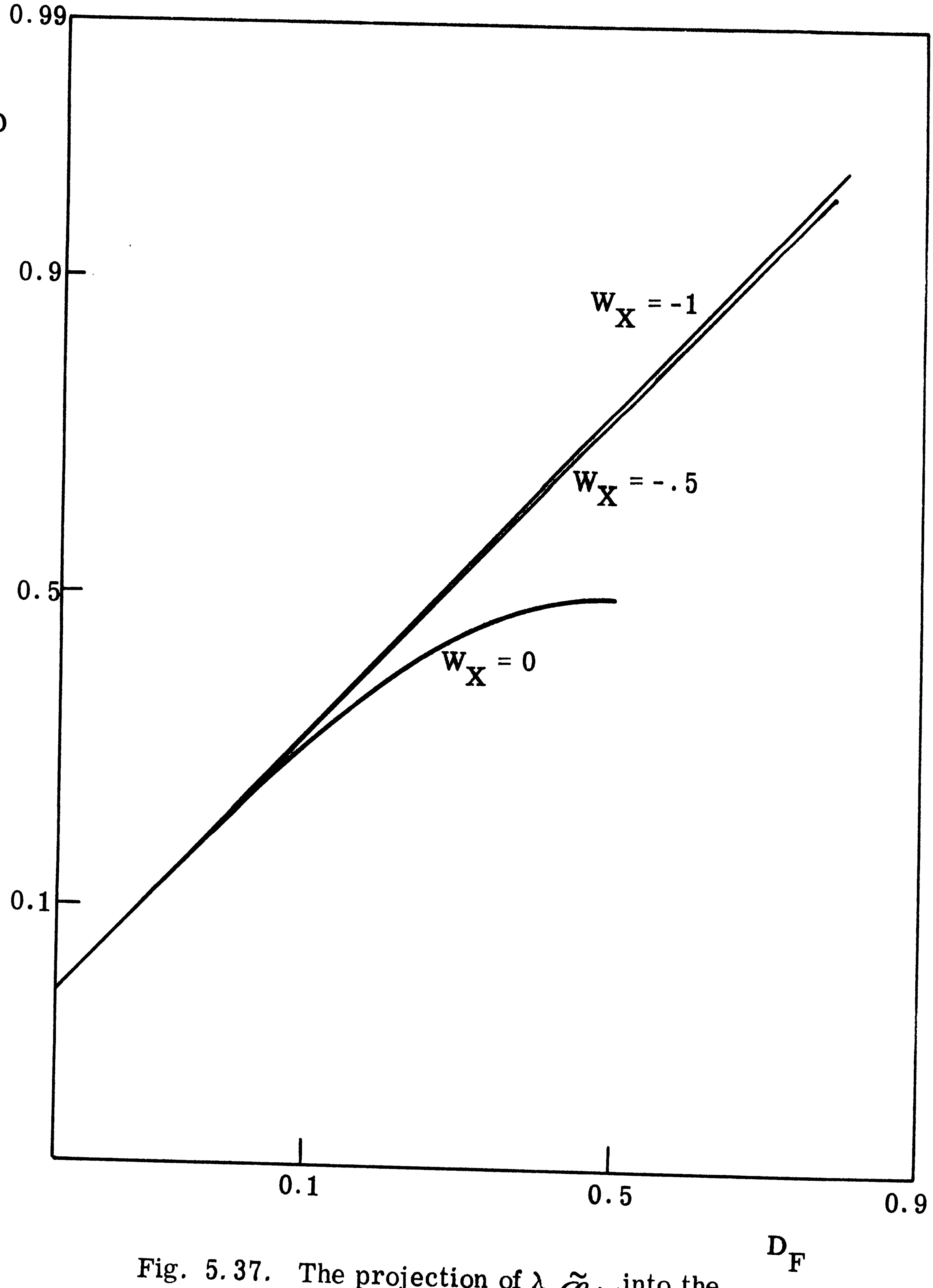


Fig. 5.37. The projection of $\lambda\, \tilde{\mathscr{P}}_{(2)}$ into the $R_D = 0$ plane

(a) $d_T = 1$.

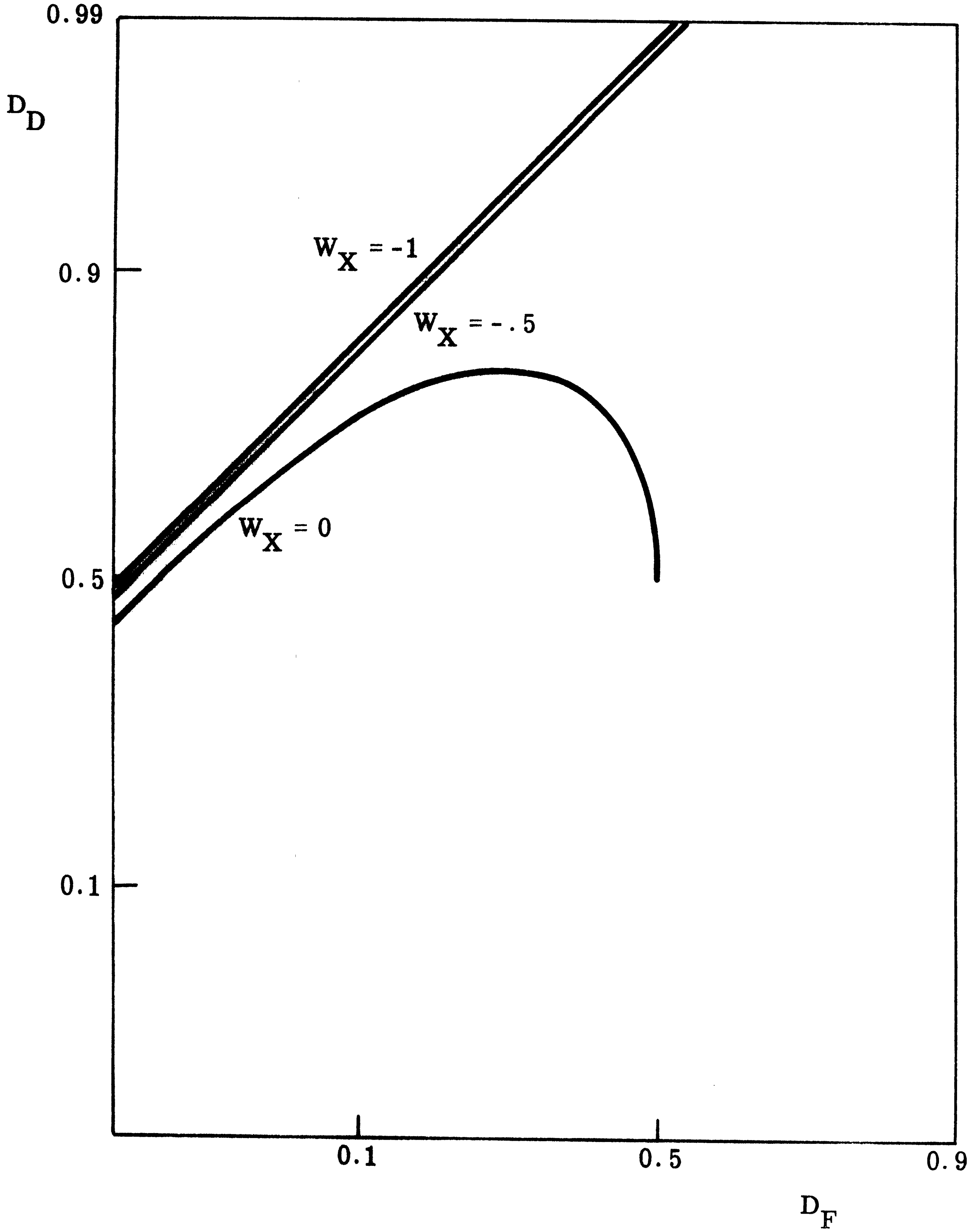


Fig. 5.37. The projection of $\lambda\, \tilde{\mathscr{P}}_{(2)}$ into the $R_D = 0$ plane
(b) $d_T = 3$.

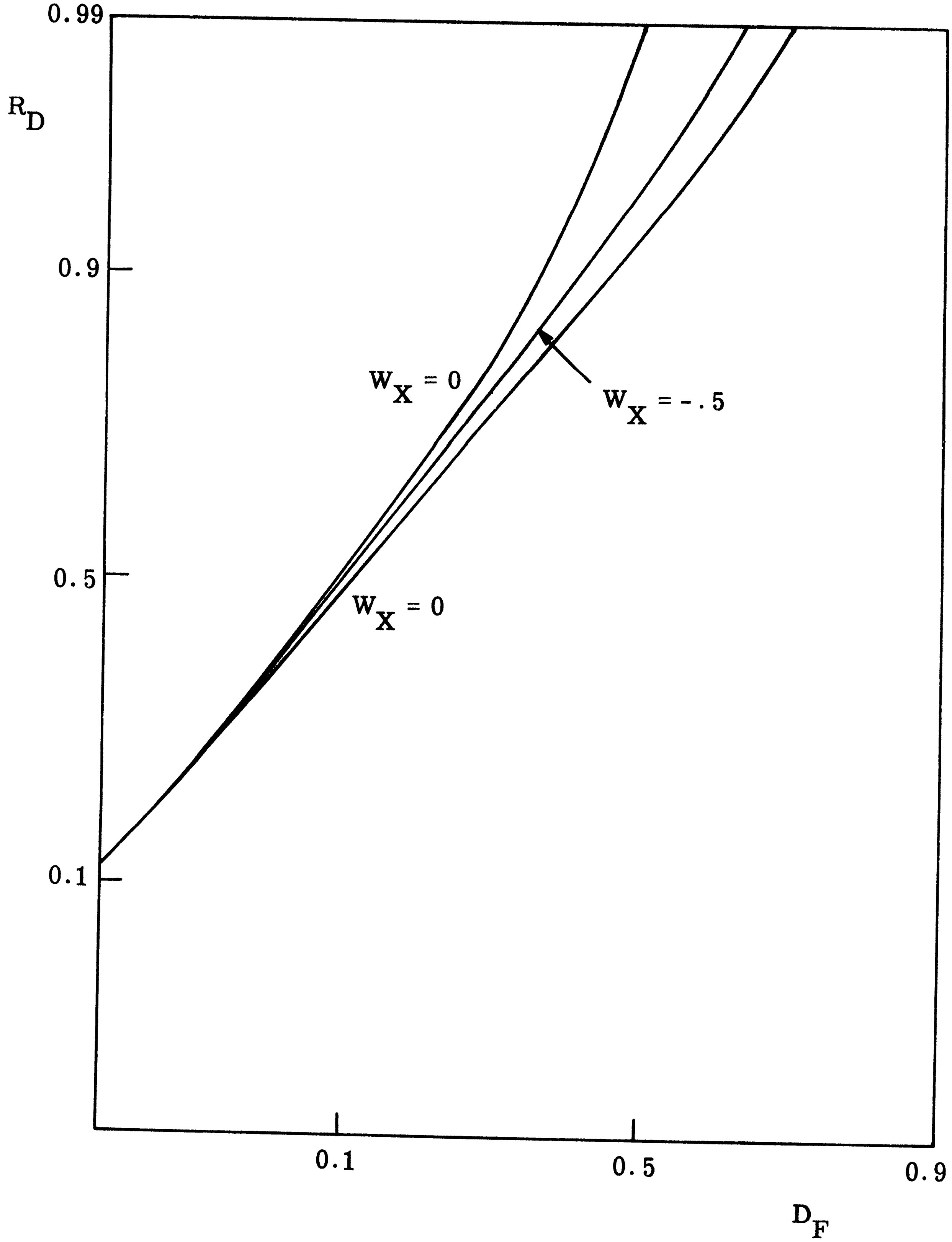


Fig. 5.38. The projection of $\lambda \tilde{\mathscr{P}}(2)$ into the $D_D = 0$ plane
(a) $d_T = 1.$

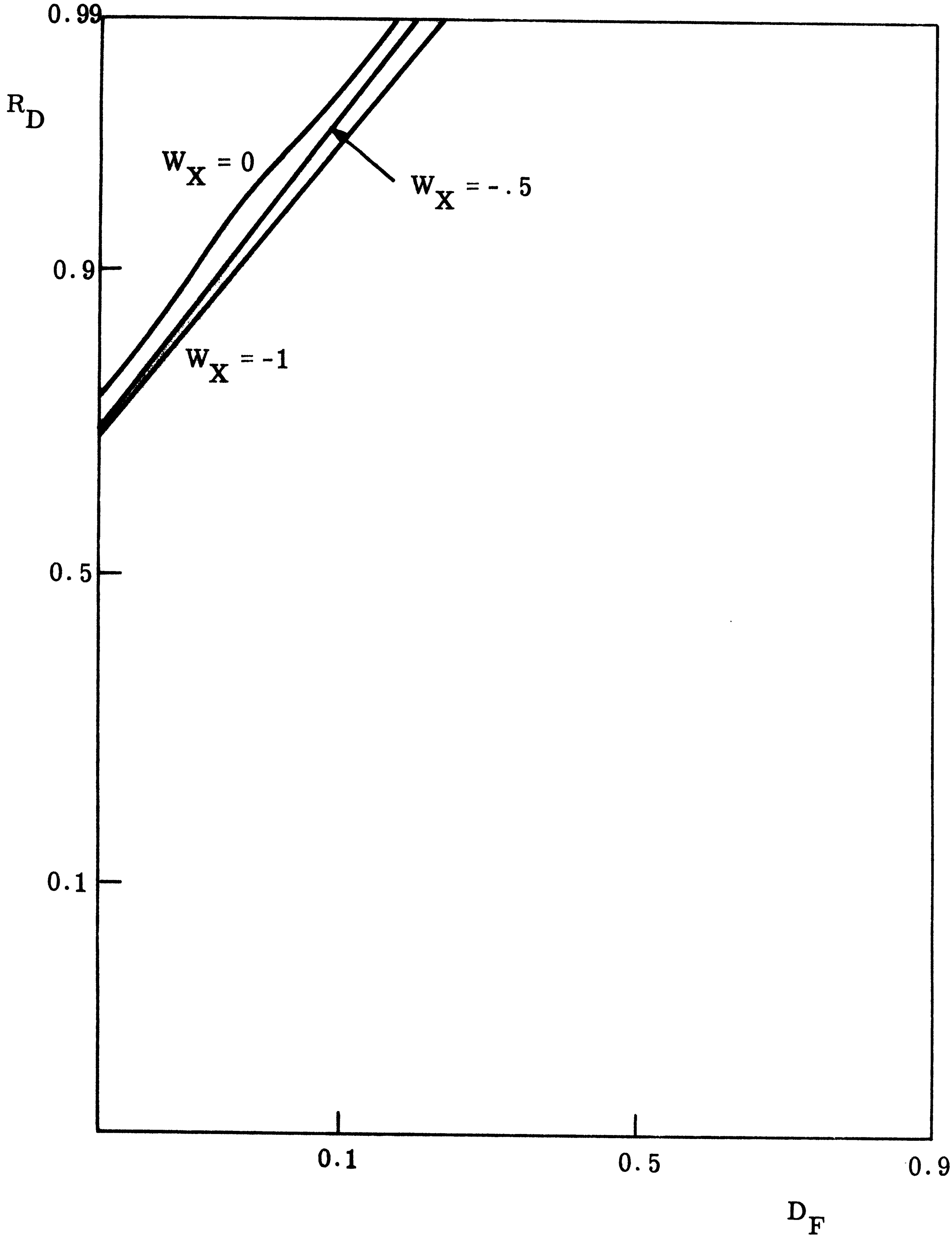


Fig. 5.38. The projection of λ $\tilde{\mathscr{S}}_{(2)}$ into the $D_D = 0$ plane
(b) $d_T = 3$.

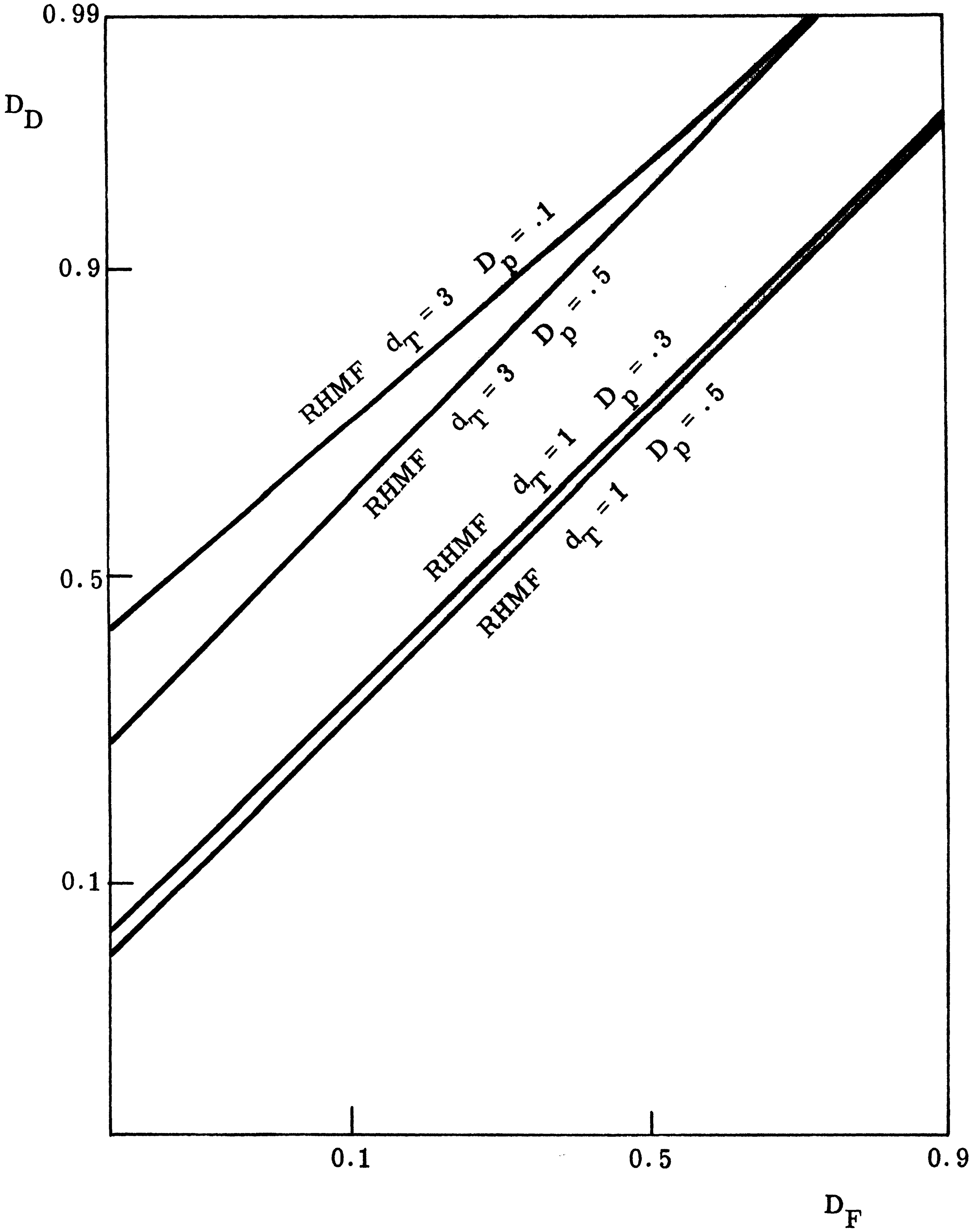


Fig. 5.39. The projections of the performance curves for the RHMF into the $R_D = 0$ plane.

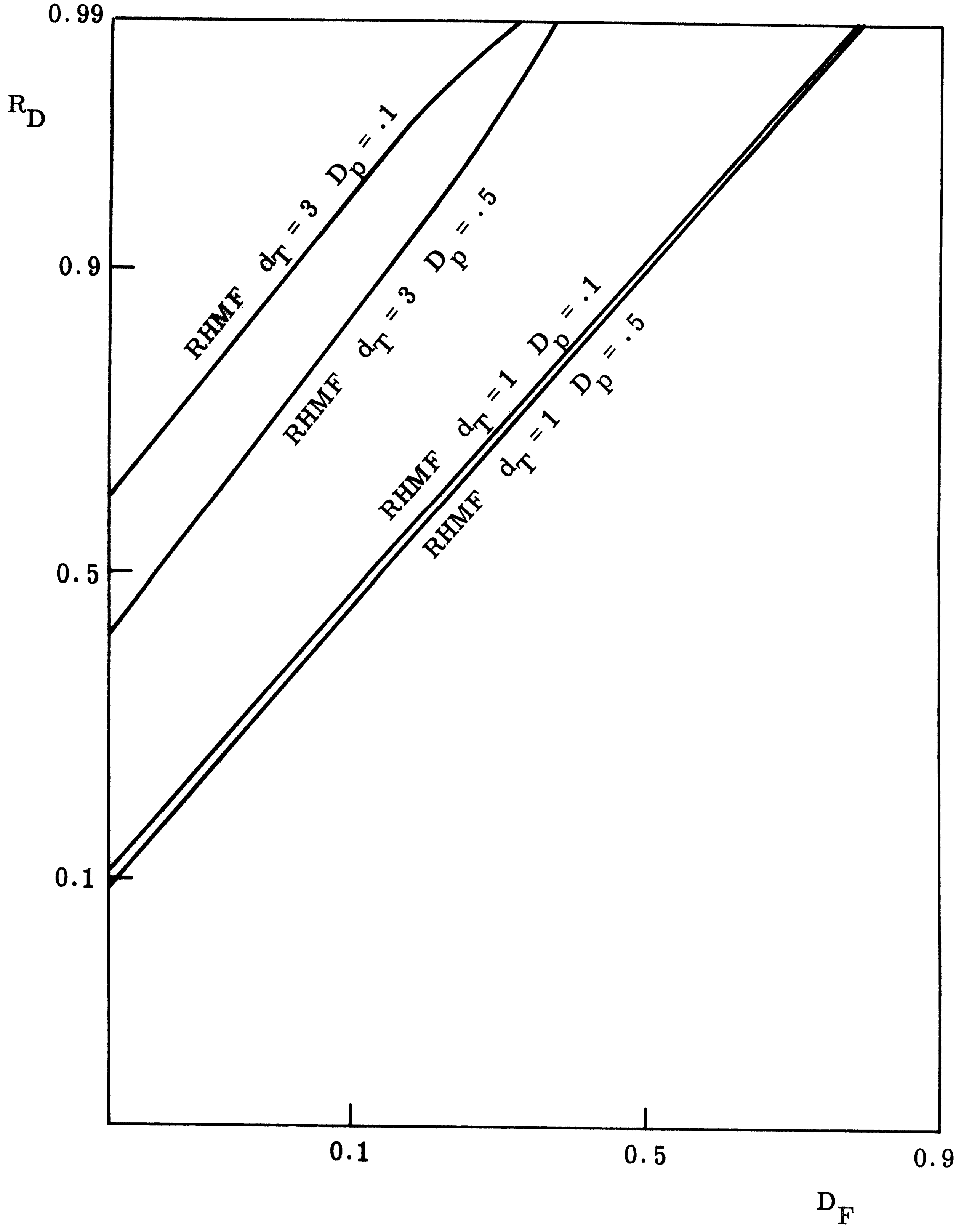


Fig. 5.40. The projections of the performance curves for the RHMF into the $D_D = 0$ plane.

of the RHMF, although, as in the respond-once case, the least amount of degradation occurs for the small value of the duty D_p.

CHAPTER VI

SUMMARY AND CONCLUSIONS

In many physical situations it is desired to detect the presence of randomly occurring pulses as they occur. In this study we have defined and analyzed two classes of decision devices that function in this setting. The respond-once decision devices seek to respond at a single instant of time within a prespecified time after the arrival of each pulse. These devices are applicable in a situation where it is desired to take a single action for each pulse that occurs. The respond-and-hold decision devices seek to respond at each instant of time at which a pulse is present. These devices are applicable in a situation where it is desired to take a continuous action during the time when each pulse is present.

To describe the performance degradation due to noise at the input of the R-O and the R-H decision devices, we have defined two loss functions. The loss function for the R-O decision devices measures the performance degradation in terms of the number of detected pulses, the number of false alarms and the number of extra detections. The special R-O decision device for which the loss for an extra detection is zero has been referred to as the maximum pulse detection device. The loss function for the respond-and-hold decision devices measures the performance degradation in terms of the number of pulses that are detected, the amount of detection time, and the amount of false alarm time. We have referred to the special R-H decision device whose performance depends only on the amount of detection time and false alarm time, and not on the number of pulses that are detected, as the maximum detection time device.

To the author's knowledge, no formal theory exists for the R-O and the R-H decision devices described above. In psychophysics, some work has been done to characterize the human decision-maker in the respond-once setting, while in systems engineering, it is generally agreed that the matched filter decision device described in Chapter I will function reasonably well in either setting. Nevertheless, no effort has been made to obtain optimum decision devices and to characterize their performance. This problem, which we have referred to as the free running detection problem, is the subject of this study.

To analyze the free running detection problem, it has been necessary to do three things: define the basic decision model, characterize the performance of both classes of decision devices, and obtain the optimum decision devices. The methods used in carrying out these tasks and the results are summarized in the following paragraphs.

In order to provide a solid theoretical foundation for the basic model, the definition of the FRD problem formulated in

Chapter I has been interpreted in terms of the general decision theory model ($\mathscr{A}$, Θ, L). This was done in Chapter II. There it was seen that the action space, $\mathscr{A}$, could be defined as the set of binary N-tuples, if we added the restriction that the decision device be allowed to make decisions (change the value of the response function) only at equally spaced times on the time axis. The separation between these decision times is an important parameter, since it determines the number of decision opportunities available to detect each pulse.

The state space for the FRD problem was defined in terms of the pulse arrival times. The procedure was to assume that either the pulses can arrive at any time as long as they do not overlap, or else to assume that they can arrive only at discrete, equally spaced times. In the continuous arrival time case, we assumed that the time between the completion of one pulse and the arrival of the next pulse can be described as an exponentially distributed random variable. In the discrete case, we assumed that this inter-arrival time can be described as a geometrically distributed random variable. The next step was to define the distribution of the arrival time of the first pulse, so that the pulse rate (the expected number of pulses per unit time) was independent of time. Finally, we defined the parameter space in terms of a sequence of random variables, $\theta = (\theta_1, \ldots, \theta_N)$, so that the ith random variable θ_i describes the pulse arrivals relevant to the decision device at the ith decision time. As a consequence of these definitions, it was seen that the sequence θ could be described in terms of a stationary, first order Markov probability law.

To complete the description of the decision model, we applied the formal definitions of Chapter I to interpret the loss functions in terms of the action space and the parameter space. It was seen that the loss functions for the R-O and the R-H decision devices could be expressed in terms of a single canonical form.

The means of describing the performance of the FRD decision devices was developed in Chapter III. The procedure followed there parallels that of classical detection theory. Specifically, a detection rate R_D, a false alarm rate R_F, and an extra detection rate R_X were defined for each R-O decision device. It was then possible to characterize the performance of the whole class of R-O decision devices in terms of the performance set $\mathscr{S}$. Similarly, it was seen that it was possible to describe the performance of an R-H decision device by the detection rate R_D, a detection duty D_D, and a false alarm duty D_F. This gave rise to the performance set $\tilde{\mathscr{S}}$ as the means of describing the performance of the class of R-H decision devices.

The next step was to establish some basic properties of the performance sets $\mathscr{S}$ and $\tilde{\mathscr{S}}$. First, it was shown that both

$\mathscr{P}$ and $\tilde{\mathscr{P}}$ lie in triangular regions in $\mathbb{R}^3$. Next, we introduced the concept of a randomized decision rule to show that both performance sets are convex. The convexity property was then used to define the concept of a chance plane for both performance sets. Finally, we extended the notion of the antipodal decision device from the classical detection problem to the FRD problem in order to obtain a reflection property.

With the above properties at hand, we were able to verify two conjectures introduced in Chapter I. First, it was seen that any FRD decision device had the property that the decision at a particular time t' depends on the previous decisions only through the time of the last response. This principle was introduced in Chapter I as the most recent response rule. Next, we proved that the only decision devices that never make extra detections are those devices that never make more than one response within an interval of duration equal to the pulse duration. This principle was introduced in Chapter I as the inhibit rule.

In the final sections of Chapter III, we used the performance sets $\mathscr{P}$ and $\tilde{\mathscr{P}}$ to define a notion of optimality that is independent of the loss function and the prior probability laws. There it was seen that this optimum performance could be described in terms of two ROC surfaces, one for the R-O decision devices and one for the R-H decision devices. Moreover, we were also able to show that this notion of optimality and the notion of Bayes optimality are equivalent.

This last fact together with the geometrical properties of the ROC surface enabled us to verify certain conjectures introduced in Chapter I. First, it was shown that, if the loss for an extra detection is sufficiently large with respect to the loss for a false alarm, then the corresponding R-O decision device satisfies the inhibit rule. Moreover, the structure of the decision device is then independent of the particular value of the extra detection loss. Secondly, we were able to show that, for those R-O decision devices that do not satisfy the inhibit rule, an increase in the detection rate can be obtained at the expense of an increase in the extra detection rate; the maximum values of the detection rate occur for the MPD decision device. As a final result, it was seen that the performance of the R-H decision devices also reflects a trade-off. Specifically, an increase in the detection duty comes at the expense of a decrease in the detection rate with the maximum values of the detection duty occurring for the MDT decision device.

In Chapter IV, we completed the theoretical portion of this study by obtaining a description of the optimum Bayes decision devices. The first step was the derivation of a general expression for the Bayes decision rule at the decision time t_k in terms of the Bayes decision rule at time t_{k+1}. From this expression it was possible to verify the configuration of the Bayes decision device that was introduced in Chapter I. Next, we considered the structure of the MDT decision device and the structure of the decision device that is allowed one decision opportunity per pulse. It was seen that both of these decision devices are equivalent to the classical detection theory device which responds if a pulse is present at the current time.

In the remaining sections of Chapter IV, we added the assumption of conditionally independent observations to obtain an updating procedure for the Bayes decision device in order to replace the past observation memory with a memory of fixed size. This is certainly a necessary requirement of any decision device that is to be operated over long periods of time. It should be mentioned, however, that the resulting structure of the Bayes decision device is still considerably more complicated than the structure of the matched filter decision devices.

In Chapter V, the theory of the preceding chapters was applied to obtain specific examples of FRD decision devices. In particular, we considered the m=1 Bayes decision devices (m is the number of decision opportunities per pulse) and the m=2 Bayes decision devices. The structure of the m=1 devices could be obtained analytically from the results of Chapter IV, but the structure of the m=2 devices had to be obtained numerically. For both of these devices it was seen that the decision at the decision time t_k is made by comparing a likelihood ratio with a threshold that depends upon a certain statistic of the past observations. The threshold for the m=2 devices also depends upon the preceding decision, but the threshold for the m=1 devices is independent of the preceding decisions.

The performance of the Bayes decision devices considered in Chapter V was calculated by simulating their operation over long periods of time. It was seen that the ROC surfaces for both the m=2 and the m=1 devices exhibited the properties developed in Chapter III. These properties were most apparent, however, for the higher signal-to-noise ratio. For example, the performance of the m=2 MDT devices showed a substantial improvement over the performance of the m=2 inhibit rule devices for a high signal-to-noise ratio but essentially no improvement for the low signal-to-noise ratio.

As a final example, we calculated the performance of the matched filter decision devices in the m=2 setting and noted that their performance was inferior to that of the Bayes decision devices. It might be emphasized, however, that since these calculations were made only in the m=2 setting, the question of the performance degradation of the matched filter devices for large values of m remains open.

APPENDIX A

THE CLASSICAL DETECTION PROBLEM AS A GENERAL DECISION PROBLEM

In this appendix the classical detection problem is interpreted in terms of the general decision model.

To begin, we identify the elements $\mathscr{A}$, Θ, L. Consider first the action space $\mathscr{A}$. There are two possible decisions in the classical detection situation; decide that a pulse is present or decide that no pulse is present. Denote the first of these alternatives by "1" and the second by "0". Then the action space $\mathscr{A}$ is given by

$$\mathscr{A} = \left\{ a \, ; a \epsilon \, \{0, 1\} \right\}$$

Next, consider the parameter space Θ. Two possible situations can occur; either signal is present or no signal is present in the observation interval. Denote the set of parameters which influence the observation when signal is present by Θ_{SN} and denote the set of parameters which influence the observation when no signal is present by Θ_N. (In many situations, Θ_N will consist of a single point indicating the event "no signal present".) Then, the state space Θ is given by

$$\Theta = \Theta_{SN} \cup \Theta_N$$

Now consider the prior probability law $\Pi(\cdot)$ on Θ.

In general, this probability law can be written in the form

$$\Pi(d\theta) = \Pi(d\theta \mid SN)\ Pr[SN]\ I_{\{\Theta_{SN}\}}(\theta) + \Pi(d\theta \mid N)\ Pr[N]\ I_{\{\Theta_N\}}(\theta)$$

where

$\Pi(d\theta \mid SN)$ is the prior probability law on Θ_{SN} given that signal is present,

$\Pi(d\theta \mid N)$ is the prior probability law on Θ_N given that no signal is present,

$Pr[SN]$ and $P[N]$ are the prior probabilities of the events "signal present" and "no signal present", respectively,

and $I_{\{\Theta'\}}(\theta)$ is the usual indicator function defined by

$$I_{\{\Theta'\}}(\theta) = \begin{cases} 0 & \theta \notin \Theta' \\ 1 & \theta \in \Theta' \end{cases}$$

Finally, consider the loss function $L(\cdot\ ,\ \cdot)$. In the classical detection problem, this function has the form illustrated in Fig. A.1.

	N_N	N_{SN}
$a = 0$	L_C	L_M
$a = 1$	L_F	L_D

Fig. A.1. The loss function for the classical detection problem

where L_C, L_M, L_F and L_D are the losses for a "correct rejection", a "miss", a "false alarm" and a "detection", respectively.

Next we calculate the Bayes risk of an arbitrary decision rule $\delta \epsilon \mathscr{D}$. By definition,

$$r(\delta) = \underset{x, \theta}{E} [L(\delta(x), \theta)]$$

or

$$r(\delta) = \underset{\theta}{E} \{\underset{x|\theta}{E} [L(\delta(x), \theta)]\}$$

For the decision rule δ , define the set, $A \subset X$, by

$$A = \{x ; \delta(x) = 1\}$$

Then the Bayes risk can be written as

$$r(\delta) = \Pr[N]\left[L_F \int_{\Theta_N}\int_A P(dx\,|\,\theta)\,\Pi(d\theta\,|\,N) + L_C \int_{\Theta_N}\int_{A^c} P(dx\,|\,\theta)\,\Pi(d\theta\,|\,N)\right]$$

$$+ \Pr[SN]\left[L_D \int_{\Theta_{SN}}\int_A P(dx\,|\,\theta)\,\Pi(d\theta\,|\,SN) + L_M \int_{\Theta_{SN}}\int_{A^c} P(dx\,|\,\theta)\,\Pi(d\theta\,|\,SN)\right]$$

The usual probabilities of detection and false alarm, P_D and P_F, are defined by

$$P_D = \int_{\Theta_{SN}}\int_A P(dx\,|\,\theta)\,\Pi(d\theta\,|\,SN) \qquad \text{(A. 1a)}$$

$$P_F = \int_{\Theta_N}\int_A P(dx\,|\,\theta)\,\Pi(d\theta\,|\,N) \qquad \text{(A. 1b)}$$

In terms of these probabilities, the Bayes risk can be written as

$$r(\delta) = \Pr[SN]\,(L_D - L_M)\,P_D + \Pr[N]\,(L_F - L_C)\,P_F$$

$$+ \Pr[SN]\,L_M + \Pr[N]\,L_C \qquad \text{(A. 2)}$$

Finally, we shall derive an expression which defines the optimum Bayes decision rule. We have

$$r(\delta) = \mathop{E}_{x,\theta}\,[L(\delta(x), \theta)]$$

$$= \mathop{E}_{x}\left[\mathop{E}_{\theta|x}\,[L(\delta(x), \theta)]\right]$$

By a well known argument in classical detection theory, the Bayes rule δ^0 is that rule which associates with $x \in X$ the $a^0(x) \in \mathscr{A}$ satisfying

$$\underset{\theta|x}{E}[L(a^0(x), \theta)] = \min \left[\underset{\theta|x}{E}[L(a=0, \theta)],\ \underset{\theta|x}{E}[L(a=1, \theta)] \right]$$

or

$$\delta(x) = \begin{cases} 1 & \text{if } \underset{\theta|x}{E}[L(a=1, \theta)] \leq \underset{\theta|x}{E}[L(a=0, \theta)] \\ 0 & \text{otherwise} \end{cases}$$

The quantity $\underset{\theta|x}{E}[L(\delta(x) = a, \theta)]$ is known as the posterior risk of the decision rule δ. A direct substitution into the condition

$$\underset{\theta|x}{E}[L(a=1, \theta)] \leq \underset{\theta|x}{E}[L(a=0, \theta)],$$

results in

$$L_F \int_{\Theta_N} \Pi(d\theta|x) + L_D \int_{\Theta_{SN}} \Pi(d\theta|x)$$

$$\leq L_C \int_{\Theta_N} \Pi(d\theta|x) + L_M \int_{\Theta_{SN}} \Pi(d\theta|x)$$

or

$$\int_{\Theta_{SN}} \Pi(d\theta|x) \geq \frac{(L_F - L_C)}{(L_M - L_D)} \int_{\Theta_N} \Pi(d\theta|x) \qquad \text{(A. 3)}$$

where $\Pi(d\theta|x)$ is the posterior distribution of θ given x. An alternate expression for the response condition can be obtained from the inequality (A. 3) by expressing $\Pi(d\theta|x)$ in terms of the observation distributions and the prior probability distribution. Namely,

$$\frac{\int_{\Theta_{SN}} P(dx|\theta)\,\Pi(d\theta|SN)}{\int_{\Theta_N} P(dx|\theta)\,\Pi(d\theta|N)} \geq \frac{(L_F - L_C)}{(L_M - L_C)} \quad \frac{Pr[N]}{Pr[SN]}$$

To obtain a more usable form for the above response condition it is necessary to place additional restrictions on the probability laws. One restriction that yields a familiar result is to assume that there is some parameter $\theta_0 \epsilon \Theta$ such that $P(dx|\theta) << P(dx|\theta_0)$ for all $\theta \epsilon \Theta$. ($<<$ denotes absolute continuity; see Appendix C.) We may then write the response condition as

$$\frac{\int_{\Theta_{SN}} \ell(x|\theta)\,P(d\theta|SN)}{\int_{\Theta_N} \ell(x|\theta)\,P(d\theta|N)} \geq \left[\frac{L_F - L_C}{L_M - L_D}\right] \quad \frac{Pr[N]}{Pr[SN]}$$

where $\ell(x|\theta)$ is the likelihood ratio

$$\ell(x|\theta) = \frac{P(dx|\theta)}{P(dx|\theta_0)}$$

APPENDIX B

DERIVATION OF THE PRIOR PROBABILITY LAWS -- PART I

In this appendix, the conditional probability law on θ_k, given $\theta_{k-1}, \dots, \theta_1$, is calculated. As a consequence of this calcultation, it will be seen that $\Pi(d\theta)$ satisfies the Markov property

$$\Pi(d\theta_k \mid \theta_{k-1}, \dots, \theta_1) = \Pi(d\theta_k \mid \theta_{k-1})$$

We proceed as follows. From the definition of $\vec{\theta} = (\theta_1, \dots, \theta_N)$ in Section 2.3, it is noted that the random variable θ_k depends on whether or not a pulse arrives in the interval $[t_k - T_p, t_k)$. The conditioning event $\theta_{k-1} = c_{k-1}, \dots, \theta_1 = c_1$ depends on pulse arrivals in the interval $[-(T_p - \Delta), t_{k-1})$. This basic temporal relation between θ_k, θ_{k-1}, $\dots$, θ_1 is shown in Fig. B. 1.

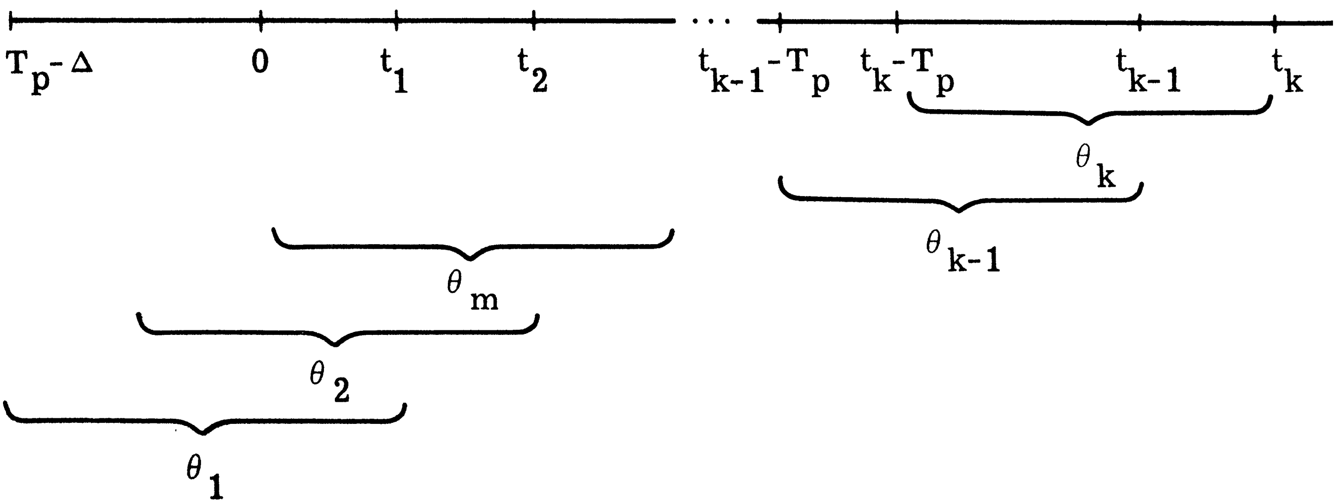


Fig. B. 1. The temporal relation between θ_k, θ_{k-1}, $\dots$, θ_1

Suppose first of all that no pulses arrive in the interval $[-(T_p - \Delta), t_{k-1})$. Then, $\theta_{k-1} = \dots = \theta_1 = T_p$. Now, if a pulse arrives in the interval $[t_k - T_p, t_k)$, then its arrival time must be equal to η_1 and, by the definition of θ_k,

$$\begin{aligned}\theta_k &= \eta_1 - (t_k - T_p) \\ &= \eta_1 - (t_{k-1} + \Delta - T_p) \\ &= \eta_1 - t_{k-1} + (T_p - \Delta)\end{aligned}$$

Thus the condition $\theta_k \leq c < T_p$ is equivalent to the condition

$$\eta_1 \leq c + t_{k-1} - (T_p - \Delta)$$

On the other hand, since it is assumed that no pulses arrive in the interval $[-(T_p - \Delta), t_{k-1})$, the arrival time η_1 must also satisfy

$$\eta_1 \geq t_{k-1}$$

Thus, for the continuous case we may write

$$\begin{aligned}&\Pr[\theta_k \leq c < T_p \mid \theta_{k-1} = \dots = \theta_1 = T_p] \\ &\quad = \Pr[\eta_1 \leq c - (T_p - \Delta) + t_{k-1} \mid t_{k-1} \leq \eta_1]\end{aligned} \tag{B. 1a}$$

For the discrete case

$$\Pr[\theta_k = c < T_p | \theta_{k-1} = \dots = \theta_1 = T_p]$$
$$= \Pr[\eta_1 = c - (T_p - \Delta) + t_{k-1} | t_{k-1} \leq \eta_1] \qquad \text{(B. 1b)}$$

Next, suppose that at least one pulse arrives in the interval $[-(T_p - \Delta), t_{k-1})$. Denote the arrival time of the last pulse to arrive in this interval by σ_ℓ .

We consider first the case $\sigma_\ell \in [t_k - T_p, t_{k-1})$ as shown in Fig. B. 2.

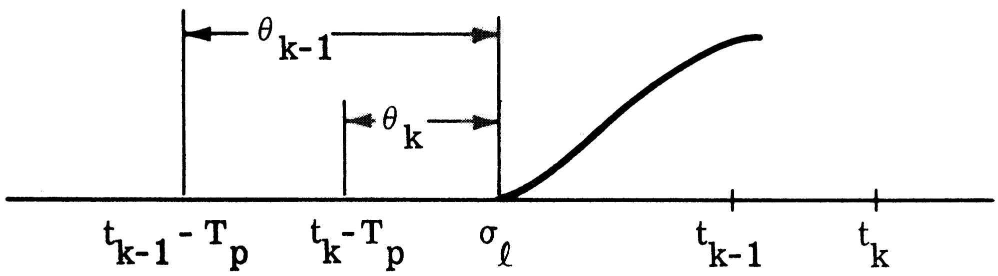


Fig. B. 2. The case $\sigma_\ell \in [t_k - T_p, t_{k-1})$

Here, it is seen that $\theta_{k-1} \in [\Delta, T_p)$ and that θ_k is related to θ_{k-1} by

$$\theta_k = \theta_{k-1} - \Delta$$

Thus, for the continuous case,

$$\Pr[\theta_k \leq c < T_p \mid \theta_{k-1} = c_{k-1}, \ldots, \theta_1 = c_1] = \begin{cases} 0 & 0 < c_{k-1} - \Delta \\ 1 & 0 \geq c_{k-1} - \Delta \end{cases} \qquad \text{(B.2a)}$$

and, for the discrete case,

$$\Pr[\theta_k = c < T_p \mid \theta_{k-1} = c_{k-1}, \ldots, \theta_1 = c_1] = \begin{cases} 0 & c \neq c_{k-1} - \Delta \\ 1 & c = c_{k-1} - \Delta \end{cases} \qquad \text{(B.2b)}$$

Next, suppose $\sigma_\ell \epsilon [t_{k-1} - T_p, t_k - T_p)$ as shown in Fig. B.3.

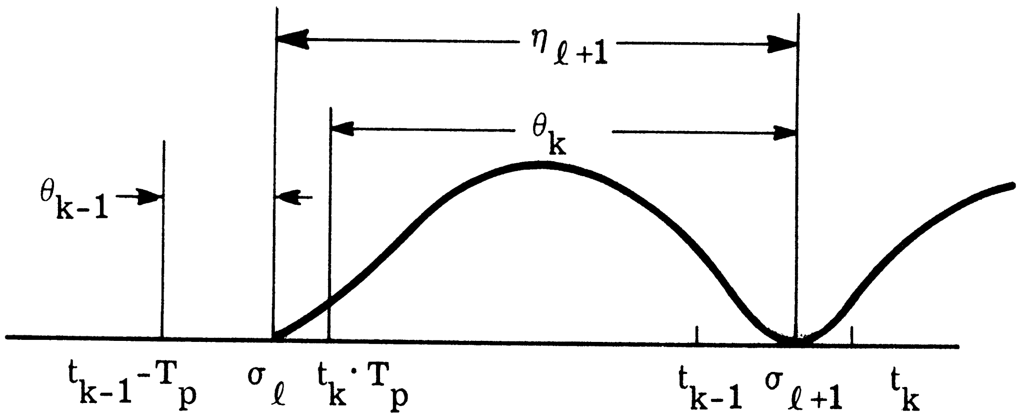


Fig. B.3. The case $\sigma_\ell \epsilon [t_{k-1} - T_p, t_k - T_p)$

In this case, $\theta_{k-1} \epsilon [0, \Delta)$. Now, if a pulse arrives in $[t_k - T_p, t_k)$, then its arrival time is given by

$$\sigma_{\ell+1} = \sigma_\ell + \eta_{\ell+1}$$

Substituting this relation into the definition of θ_k results in

$$\theta_k = \sigma_\ell + \eta_{\ell+1} - (t_k - T_p)$$

$$= \eta_{\ell+1} + \sigma_\ell - t_{k-1} + [T_p - \Delta]$$

Thus the condition $\theta_k \leq c$ is equivalent to the condition

$$\eta_{\ell+1} \leq c - [T_p - \Delta] + t_{k-1} - \sigma_\ell$$

On the other hand, the condition that σ_ℓ is the last pulse to occur in the interval $[-(T_p - \Delta), t_{k-1})$ is equivalent to

$$\eta_{\ell+1} \geq t_{k-1} - \sigma_\ell$$

Thus, for the continuous case, we have

$$\Pr[\theta_k \leq c < T_p \mid \theta_{k-1} = c_{k-1}, \ldots, \theta_1 = c_1]$$
$$= \Pr[\eta_{\ell+1} \leq c - [T_p - \Delta] + t_{k-1} - \sigma_\ell \mid \eta_{\ell+1} \geq t_{k-1} - \sigma_\ell] \quad \text{(B. 3a)}$$

and for the discrete case

$$\Pr[\theta_k = c < T_p \mid \theta_{k-1} = c_{k-1}, \ldots, \theta_1 = c_1]$$
$$= \Pr[\eta_{\ell+1} = c - [T_p - \Delta] + t_{k-1} - \sigma_\ell \mid \eta_{\ell+1} \geq t_{k-1} - \sigma_\ell] \quad \text{(B. 3b)}$$

The last possibility to consider is the case $\sigma_\ell \epsilon [-(T_p - \Delta), t_{k-1} - T_p)$. This case is shown in Fig. B. 4.

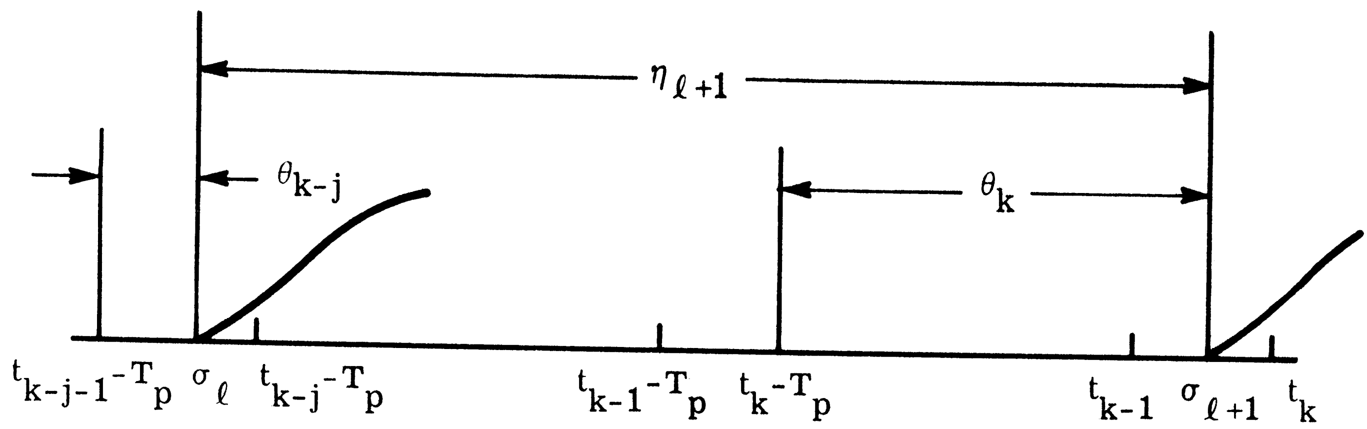


Fig. B.4. The case $\sigma_\ell \epsilon [-(T_p - \Delta), t_{k-1} - T_p)$

As shown here, $\sigma_\ell \epsilon [t_{k-j-1} - T_p, t_{k-j} - T_p)$, for some $j > 1$, so that $\theta_{k-j} \epsilon [0, \Delta)$ and $\theta_{k-j+1} = \dots = \theta_{k-1} = T_p$. Now as before the condition $\theta_k \leq c$ is equivalent to the condition

$$\eta_{\ell+1} \leq c - [T_p - \Delta] + t_{k-1} - \sigma_\ell$$

and the condition that σ_ℓ is the last pulse to arrive in $[-(T_1 - \Delta), t_{k-1})$ is equivalent to the condition

$$\eta_{\ell+1} \geq t_{k-1} - \sigma_\ell$$

Thus, in the continuous case,

$$\Pr[\theta_k \leq c < T_p \mid \theta_{k-1} = \dots = \theta_{k-j+1} = T_p, \theta_{k-j} = c_{k-j}, \dots, \theta_1 = c_1]$$

$$= \Pr[\eta_{\ell+1} \leq c - [T_p - \Delta] + t_{k-1} - \sigma_\ell \mid \eta_{\ell+1} \geq t_{k-1} - \sigma_\ell] \qquad \text{(B.4a)}$$

and, in the discrete case,

$$\Pr[\theta_k = c < T_p \mid \theta_{k-1} = \cdots = \theta_{k-j+1} = T_p,\ \theta_{k-j} = c_{k-j},\ \ldots,\ \theta_1 = c_1]$$

$$= \Pr[\eta_{\ell+1} = c - (T_p - \Delta) + t_{k-1} - \sigma_\ell \mid \eta_{\ell+1} \geq t_{k-1} - \sigma_\ell] \tag{B.4b}$$

We are now in a position to calculate the conditional probability laws. The continuously arriving pulse case is considered first. First suppose that $\theta_{k-1} \epsilon [\Delta, T_p)$. Then, from Eq. B.2a

$$\Pr[\theta_k \leq c < T_p \mid \theta_{k-1} = c_{k-1},\ \ldots,\ \theta_1 = c_1]$$

$$= \begin{cases} 0 & c < c_{k-1} - \Delta \\ 1 & c \geq c_{k-1} - \Delta \end{cases} \tag{B.5}$$

Next, suppose that $\theta_{k-1} \epsilon [0, \Delta)$. The equation of interest here is Eq. B.3a. To begin, note from Fig. B.3 that $t_{k-1} - \sigma_\ell$ is related to θ_{k-1} by

$$t_{k-1} - \sigma_\ell = T_p - \theta_{k-1}$$

Now substitute for $t_{k-1} - \sigma_\ell$ from above into Eq. B.3a and set $\theta_{k-1} = c_{k-1}$. The result is

$$\Pr[\theta_k \leq c < T_p | \theta_{k-1} = c_{k-1}, \ldots, \theta_1 = c_1] =$$

$$\Pr[\eta_{\ell+1} \leq c - [T_p - \Delta] + T_p - c_{k-1} | \eta_{\ell+1} \geq T_p - c_{k-1}]$$

Next apply Bayes rule to write the right hand side of the above equation as

$$\frac{\Pr[T_p - c_{k-1} \leq \eta_{\ell+1} \leq c - c_{k-1} + \Delta]}{\Pr[\eta_{\ell+1} \geq T_p - c_{k-1}]}$$

Finally, from the density in Eq. 2.8 of the text, it follows that for $c_{k-1} \epsilon [0, \Delta)$

$$\Pr[\eta_{\ell+1} \leq T_p - c_{k-1}] = 1$$

and

$$\Pr[T_p - c_{k-1} \leq \eta_{\ell+1} \leq c - c_{k-1} + \Delta]$$

$$= \begin{cases} 1 - e^{-\alpha(c - (c_{k-1} + T_p - \Delta))} & \text{if} \quad c \geq c_{k-1} + T_p - \Delta \\ 0 & \text{otherwise} \end{cases}$$

Thus, for $\theta_{k-1} = c_{k-1} \epsilon [0, \Delta)$,

$$\Pr[\theta_k \leq c \,|\, \theta_{k-1} = c_{k-1}, \ldots, \theta_1 = c_1] \tag{B.6}$$

$$= \begin{cases} 0 & c \leq c_{k-1} + [T_p - \Delta] \\ 1 - e^{-\alpha(c - (c_{k-1} + T_p - \Delta))} & c_{k-1} + T_p - \Delta \leq c < T_p \\ 1 & T_p \leq c \end{cases}$$

Finally suppose that $c_{k-1} = T_p$. There are two possibilities. Either $\theta_{k-2} = \ldots = \theta_1 = T_p$, or there is some largest index $j > 1$ such that $\theta_{k-j} \epsilon [0, \Delta)$ and $\theta_{k-j+1} = \ldots = \theta_{k-1} = T_p$. Equation B.1a applies in the first case and Eq. B.4a applies in the second. In the first case, the probability on the right hand side of Eq. B.1a is determined from the density in Eq. 2.9 of the text, and, in the second case, the probability on the right hand side of Eq. B.4a is determined from the density in Eq. 2.8 of the text. In either case we have

$$\Pr[\theta_k \leq c \,|\, \theta_{k-1} = T_p, \theta_{k-2} = c_{k-2}, \ldots, \theta_1 = c_1] \tag{B.7}$$

$$= \begin{cases} 0 & c \leq T_p - \Delta \\ 1 - e^{-\alpha(c - (T_p - \Delta))} & T_p - \Delta \leq c < T_p \\ 1 & T_p \leq c \end{cases}$$

This completes the derivation for the continuous case. The conditional cumulative distributions are given by Eqs. B.5, B.6 and B.7.

Furthermore, since the right hand sides of these equations do not depend on the values of $\theta_1, \ldots, \theta_{k-2}$, we have shown that the Markov property holds.

We turn now to the discrete case. In this case we have assumed that the separation between the allowable pulse arrival times, ν , is related to T_p by

$$q\nu = T_p$$

for same integer q . Moreover, we also have

$$m\Delta = T_p$$

Now we will also assume that Δ can be divided into an integral number, s , of intervals of length ν . Thus, we may write

$$\Delta = s\nu$$

As a last bit of notation we shall write c and c_{k-1} as

$$c = \ell\nu \quad \text{and} \quad c_{k-1} = j\nu$$

First of all we consider the case $c_{k-1} \epsilon [\Delta, T_p)$, or equivalently the case $j \epsilon \{s, \ldots, q-1\}$. Equation B.2b applies here. The result is that

$$\Pr[\theta_k = \ell\nu | \theta_{k-1} = j\nu, \ldots, \theta_1 = c_1]$$

$$= \begin{cases} 0 & \ell \neq j - s \\ 1 & \ell = j - s \end{cases} \qquad \text{(B. 8)}$$

Next consider the case $\theta_{k-1} \epsilon [0, \Delta)$. In this case $j \epsilon \{0, \ldots, s-1\}$. If we reason in precisely the same way as in the continuously arriving pulse case for $\theta_{k-1} \epsilon [0, \Delta)$, then we conclude

$$\Pr[\theta_k = \ell \nu | \theta_{k-1} = j\nu, \ldots, \theta_1] \tag{B.9}$$

$$= \begin{cases} 0 & \ell < q-s-j \\ (1-a)\, a^{\ell-(j+q-s)} & j+q-s \leq \ell < q \\ a^{s-j} & \ell = q \end{cases}$$

Finally consider the case $c_{k-1} = T_p$. In this case, $j = q$. Again the reasoning is the same as in the corresponding continuous arrival time case. The result is

$$\Pr[\theta_k = \ell \nu | \theta_{k-1} = j\nu, \ldots, \theta_1] \tag{B.10}$$

$$= \begin{cases} 0 & \ell < q-s \\ (1-a)\, a^{\ell-(q-s)} & q-s \leq \ell < q \\ a^{s} & \ell = q \end{cases}$$

The conditional probability law in the discrete case is given by Eqs. B.8, B.9 and B.10. Again this law satisfies the Markov property since the right hand sides of these equations depend only on the value of θ_{k-1}

APPENDIX C

DERIVATION OF THE GENERALIZED DENSITIES FOR THE PRIOR PROBABILITY LAWS

In this appendix we obtain generalized densities for the prior probability laws of Section 2.3. The procedure will be to find σ-finite measures $\mu(d\theta_k)$ such that $\Pi(d\theta_k) << \mu(d\theta_k)$.

As a preliminary we review the definitions of absolute continuity and σ-finite measures. Let ω and μ be two measures defined on the same measurable space $(Y, \mathscr{B})$. Then ω is absolutely continuous with respect to μ if given any measurable set $A \in \mathscr{B}$ such that $\mu(A) = 0$, then $\omega(A) = 0$. The measure μ is said to be σ-finite if there exists a countable partition of measurable sets $A_i \in \mathscr{B}$, $i = 1, 2, \ldots,$ such that $\mu(A_i) < \infty$ for all i.

We return to the problem at hand. Consider the discrete arrival time case. In this case the relevant probability laws are given by Eqs. 2.17 and 2.19. These laws assign positive mass only to the points $\theta_k = \ell \nu;\ \ell = 0, \ldots, q$. Thus, if we define $\mu(d\theta_k)$ to be the measure which assigns unit mass to each of the points $\theta_k = \ell \nu;\ \ell = 0, \ldots, q$, then clearly these laws are absolutely continuous with respect to $\mu(d\theta_k)$. ($\mu(d\theta_k)$ is often referred to as the counting measure.) Moreover, if A is any measurable set then

$$\Pr[\theta_k \epsilon A \mid \theta_{k-1} = j\nu] = \sum_{\ell \ni \ell \nu \epsilon A} \Pr[\theta_k = \ell \nu \mid \theta_{k-1} = j\nu]$$

$$= \int_A \Pr[\theta_k = \ell\nu \mid \theta_{k-1} = j\nu]\, \mu(d\theta_k)$$

and similarly

$$\Pr[\theta_k \epsilon A] = \int_A \Pr[\theta_k = \ell\nu \mid \theta_{k-1} = j\nu]\, \mu(d\theta_k)$$

Thus, by the Radon-Nikodym theorem, we can take the densities $p(\theta_k = \ell\nu \mid \theta_{k-1} = j\nu)$, $p(\theta_k = \ell\nu)$ to be the functions

$$p(\theta_k = \ell\nu) = \Pr[\theta_k = \ell\nu] \tag{C. 1a}$$

and

$$p(\theta_k = \ell\nu \mid \theta_{k-1} = j\nu) = \Pr[\theta_k = \ell\nu \mid \theta_{k-1} = j\nu] \tag{C. 1b}$$

Next we treat the continuous case. In this case the relevant probability laws are given by Eqs. 2. 16 and 2. 18. Consider first the unconditional distribution function of Eq. 2. 18. This function defines a probability law that consists of a mixture of a probability law that is absolutely continuous with respect to Lebesque measure on $[0, T_p)$ and a probability law which assigns unit mass to the point $\theta_k = T_p$. Thus, we may define $\mu(d\theta_k)$ by

$$\mu(d\theta_k) = d\theta_k + \epsilon_{\{T_p\}}(d\theta_k)$$

where $d\theta_k$ is the Lebesque measure and $\epsilon_{\{T_p\}}(d\theta_k)$ is the measure which assigns unit mass to the point $\theta_k = T_p$. It is then easy to conclude from the Radon-Nikodym theorem that the generalized density $p(\theta_k)$ with respect to $\mu(d\theta_k)$ is given by

$$p(\theta_k) = \begin{cases} [T_p + (\alpha^{-1})]^{-1} & \theta_k \epsilon [0, T_p) \\ (\alpha^{-1})[T_p + (\alpha^{-1})]^{-1} & \theta_k = T_p \end{cases} \tag{C.2a}$$

Finally, consider the conditional probability law defined by Eq. 2.16. Suppose first that $\theta_k \epsilon [0, \Delta)$ or $\theta_k \epsilon \{T_p\}$. In this case the conditional probability law is again absolutely continuous with respect to $\mu(d\theta_k) = d\theta_k + \epsilon_{\{T_p\}}(d\theta_k)$ and it is easily seen that we may take $p(\theta_k | \theta_{k-1})$ to be the function

$$
p(\theta_k | \theta_{k-1}) = \tag{C.2b}
$$

$$
\begin{cases}
\begin{cases}
0 & \theta_k \notin [\theta_{k-1} + T_p - \Delta,\ T_p] \\
\alpha e^{-\alpha(\theta_k - [\theta_{k-1} + T_p - \Delta])} & \theta_k \in [\theta_{k-1} + T_p - \Delta,\ T_p) \\
e^{-\alpha(\Delta - \theta_k)} & \theta_k = T_p
\end{cases} & \theta_{k-1} \in [0, \Delta) \\
\begin{cases}
0 & \theta_k \notin [T_p - \Delta,\ T_p] \\
\alpha e^{-\alpha(\theta_k - [T_p - \Delta])} & \theta_k \in [T_p - \Delta,\ T_p) \\
e^{-\alpha \Delta} & \theta_k = T_p
\end{cases} & \theta_{k-1} = T_p
\end{cases}
$$

This function is illustrated in Fig. C.1 below.

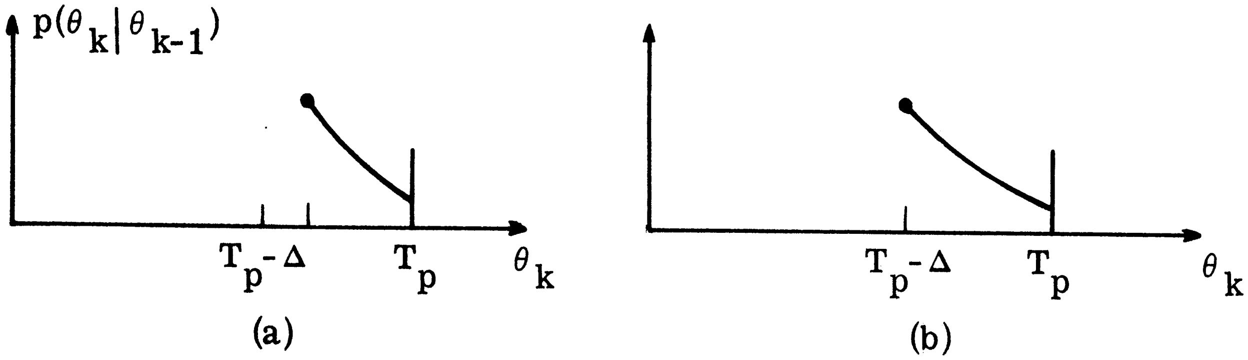


Fig. C.1. The density $p(\theta_k | \theta_{k-1})$ for $\theta_{k-1} \in [0, \Delta) \cup \{T_p\}$ (a) $\theta_{k-1} \in [0, \Delta)$ (b) $\theta_{k-1} = T_p$

Next suppose that $\theta_{k-1} \epsilon [\Delta, T_p)$. In this case, the probability law defined by $F(c_k | c_{k-1})$ assigns unit mass to the point $\theta_k = c_{k-1} - \Delta$. That is, if we define the measure $\epsilon_{\{c\}}(d\theta_k)$ to be the measure which assigns unit mass to the point c, then for $\theta_{k-1} \epsilon [\Delta, T_p)$

$$\Pi(d\theta_k | \theta_{k-1}) = \epsilon_{\{\theta_{k-1} - \Delta\}}(d\theta_k)$$

Now the difficulty here is that it is not possible to find a σ-finite measure $\mu(d\theta_k)$ which does not depend on θ_{k-1} and such that

$$\epsilon_{\{\theta_k - \Delta\}}(d\theta_k) << \mu(d\theta_k) \quad \text{for all } \theta_{k-1} \epsilon [\Delta, T_p)$$

Thus, in this case, it is not possible to find a generalized density

$$p(\theta_k | \theta_{k-1}) = \Pi(d\theta_k | \theta_{k-1}) / \mu(d\theta_k)$$

for which $\mu(d\theta_k)$ is independent of θ_{k-1}.

The results of this appendix can be summarized as follows. Define the measure $\mu(d\theta_k)$ by

$$\mu(d\theta_k) = \mu_{[0, T_p)}(d\theta_k) + \epsilon_{\{T_p\}}(d\theta_k) \qquad \text{(C. 3a)}$$

where

$$\mu_{[0,\, T_p)}(d\theta_k) = \begin{cases} d\theta_k & \text{in the continuous case} \\ \sum_{j=0}^{q-1} \epsilon_{\{j\nu\}}(d\theta_k) & \text{in the discrete case} \end{cases} \tag{C.3b}$$

Then, in both the continuous case and the discrete case there exists an unconditional density $p(\theta_k)$ such that

$$\Pi(d\theta_k) = p(\theta_k)\, \mu(d\theta_k) \tag{C.4}$$

The density $p(\theta_k)$ is given in Eqs. C.1a and C.2a. In addition, for $\theta_k \epsilon [0, \Delta) \cup \{T_p\}$, there exists a conditional density $p(\theta_k | \theta_{k-1})$ such that

$$\Pi(d\theta_k | \theta_{k-1}) = \begin{cases} p(\theta_k | \theta_{k-1})\, \mu(d\theta_k) & \theta_k \epsilon [0, \Delta) \cup \{T_p\} \\ \epsilon_{\{\theta_{k-1} - \Delta\}}(d\theta_k) & \theta_k \epsilon [\Delta, T_p) \end{cases} \tag{C.5}$$

The density $p(\theta_k | \theta_{k-1})$ is given in Eqs. C.1b and C.2b.

APPENDIX D

DERIVATION OF THE PRIOR PROBABILITY LAWS -- PART II

In this appendix we show that for both the continuous arrival time case and the discrete arrival time case the unconditional density $p_k(\theta_k)$ is independent of k . Now the most direct approach would be to simply perform the necessary computations separately in each case. However, an alternate procedure is to first derive an expression for the unconditional density that is valid in both cases and then evaluate that expression to obtain the desired result. This procedure has the added advantage that it automatically provides a result that is fundamental to the proof of Theorem 4.6 in Chapter IV. Thus, this alternate procedure will be followed here.

To begin, it is noted that since $\Pi(d\theta)$ is Markov with a stationary transition density, it is only necessary to show that $\Pi(d\theta_2)$ can be described by an unconditional density $p_2(\cdot)$ which is equal to the unconditional density $p_1(\cdot)$.

We proceed as follows. The unconditional distribution $\Pi(d\theta_2)$ is related to $\Pi(d\theta_2|\theta_1)$ and $\Pi(d\theta_2)$ by

$$\Pi(d\theta_2) = \int_{\Theta_1} \Pi(d\theta_2|\theta_1)\,\Pi(d\theta_1) \qquad \text{(D. 1)}$$

Now from Appendix C we have

$$\Pi(d\theta_1) = p(\theta_1)\,\mu(d\theta_1) \tag{D.2}$$

where

$$\mu(d\theta_1) = \mu_{[0,\,T_p)}(d\theta_1) + \epsilon_{\{T_p\}}(d\theta_1)$$

and

$$\mu_{[0,\,T_p)}(d\theta_1) = \begin{cases} d\theta_1 & \text{in the continuous case} \\ \sum_{j=1}^{q-1} \epsilon_{\{j\nu\}}(d\theta_1) & \text{in the discrete case} \end{cases}$$

Now substitute Eq. D. 2 into Eq. D. 1 to obtain

$$\Pi(d\theta_2) = \int_{\Theta_1} p(\theta_1)\,\Pi(d\theta_2 | \theta_1)\mu(d\theta_1) \tag{D.3}$$

The basic procedure to be followed here is to derive a relation of the form

$$\int_E \Pi(d\theta_2) = \int_A [\qquad]\,\mu(d\theta_2)$$

for all measurable sets A . Then by the Radon-Nikodym theorem we can identify the bracketed quantity on the right hand side as the density $p_2(\theta_2)$ with respect to the measure $\mu(d\theta_2)$. This procedure will be used to obtain desired result as a special case of a more general theorem.

Theorem D.1. Let $f(\theta_1, \theta_2)$ be any non-negative function that is $\mu(\cdot)$ measurable in θ_1 for each θ_2. Define the measure $\omega(d\theta_2)$ by

$$\omega(d\theta_2) = \int_{\Theta_1} f(\theta_2, \theta_1)\, \Pi(d\theta_2 | \theta_1)\mu(d\theta_1) \tag{D.4}$$

Then the Radon-Nikodym derivative of $\omega(d\theta_2)$ with respect to $\mu(d\theta_2)$ exists and is given by

$$\frac{\omega(d\theta_2)}{\mu(d\theta_2)} = I_{[T_p-\Delta, T_p]}(\theta_2) \int_{\tilde{\Theta}_1(\theta_2)\cup\{T_p\}} f(\theta_2, \theta_1)p(\theta_2|\theta_1)\mu(d\theta_1) + I_{[0, T_p-\Delta)}(\theta_2)\; f(\theta_2, \theta_1 = \theta_2 + \Delta) \tag{D.5}$$

where

$$\tilde{\Theta}_1(\theta_2) = \begin{cases} [0, \theta_2 - (T_p - \Delta)] & \theta_2 \in [T_p - \Delta, T_p) \\ [0, \Delta) & \theta_2 = T_p \end{cases}$$

Proof. We begin by breaking the integral in Eq. D.4 into two parts to obtain

$$\omega(d\theta_2) = \int_{[0,\Delta)\cup\{T_p\}} f(\theta_2, \theta_1)\, \Pi(d\theta_2|\theta_1) + \int_{[\Delta, T_p)} f(\theta_2, \theta_1)\, \Pi(d\theta_2|\theta_1)\mu(d\theta_1) \tag{D. 6}$$

Now note from Eq. C. 5 of Appendix C that for $\theta_2 \epsilon [0, \Delta) \cup \{T_p\}$, $\Pi(d\theta_2|\theta_1)$ is described by a density $p(\theta_2|\theta_1)$ with respect to $\mu(d\theta_2)$. Thus we may write

$$\Pi(d\theta_2|\theta_1) = p(\theta_2|\theta_1)\mu(d\theta_2) \qquad \theta_1 \epsilon [0, \Delta) \cup \{T_p\}$$

Also from Eq. C. 5

$$\Pi(d\theta_2|\theta_1) = \epsilon_{\{\theta_1 - \Delta\}}(d\theta_2) \qquad \theta_1 \epsilon [\Delta, T_p)$$

Substitute these expressions for $\Pi(d\theta_2|\theta_1)$ into Eq. D. 6. The result is

$$\omega(d\theta_2) = \left[\int_{[0,\Delta)\cup\{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2|\theta_1)\, \mu(d\theta_1)\right] \mu(d\theta_2) + \int_{[\Delta, T_p)} f(\theta_2, \theta_1)\, \epsilon_{\{\theta_1 - \Delta\}}(d\theta_2)\mu(d\theta_1)$$

Now let $A \subset \Theta_2$ be any measurable set. Then

$$\int_A \omega(d\theta_2) = \int_A \left[\int_{[0,\Delta) \cup \{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2|\theta_1) \mu(d\theta_1) \right] \mu(d\theta_1)$$
$$+ \int_A \int_{[\Delta, T_p)} f(\theta_2, \theta_1)\, \epsilon_{\{\theta_1 - \Delta\}}(d\theta_2)\, \mu(d\theta_1) \qquad \text{(D. 7)}$$

In Lemma D. 3, which is proven at the end of this appendix, it is shown that the second integral in Eq. D. 7 can be written as

$$\int_A I_{[0,\, T_p - \Delta)}(\theta_2)\;\; f(\theta_2, \theta_1 = \theta_2 + \Delta)\, \mu(d\theta_2)$$

Thus we can rewrite Eq. D. 7 as

$$\int_A \omega(d\theta_2) \qquad \text{(D. 8)}$$

$$= \int_A \left[\int_{[0,\Delta) \cup \{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2|\theta_1)\, \mu(d\theta_1) + I_{[0,\, T_p - \Delta)}(\theta_2)\;\; f(\theta_2, \theta_1 = \theta_2 + \Delta) \right] \mu(d\theta_2)$$

Then since Eq. D. 8 holds for any measurable set A, we may take the Radon-Nikodym derivative $\omega(d\theta_2)/\mu(d\theta_2)$ to be the function

$$\frac{\omega(d\theta_2)}{\mu(d\theta_2)} = \int_{[0,\Delta) \cup \{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2|\theta_1) \mu(d\theta_1) + I_{[0,\, T_p - \Delta)}(\theta_2)\;\; f(\theta_2, \theta_1 = \theta_2 + \Delta) \qquad \text{(D. 9)}$$

To complete the proof, we examine in greater detail the first integral on the right hand side of Eq. D.9. This integral can be written as

$$\int_{[0,\Delta)} f(\theta_2, \theta_1)\, p(\theta_2|\theta_1)\mu(d\theta_2) =$$

$$\int_{[0,\Delta)} f(\theta_2, \theta_1) p(\theta_2|\theta_1)\mu(d\theta_1) + \int_{\{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2|\theta_1)\mu(d\theta_1) \tag{D. 10}$$

The second integral in Eq. D. 10 can be evaluated immediately as follows:

$$\int_{\{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2|\theta_1)\mu(d\theta_1) = \int_{\{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2|\theta_1)\, \epsilon_{\{T_p\}}(d\theta_1)$$

$$= f(\theta_2, \theta_1 = T_p)\, p(\theta_2|\theta_1 = T_p)$$

Now note from Eqs. C. 1b and C. 2b of Appendix C and Eq. 2. 17 of Section 2.3 that for $\theta_1 = T_p$ the density $p(\theta_2|\theta_1 = T_p)$ is non-zero only for θ_2 satisfying $T_p - \Delta \leq \theta_2 \leq T_p$. Thus we may write this second integral as

$$\int_{\{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2|\theta_1)\mu(d\theta_1) = I_{[T_p - \Delta, T_p]}(\theta_2)\; f(\theta_2, \theta_1 = T_p)\, p(\theta_2|\theta_1 = T_p) \tag{D. 11}$$

Next consider the first integral in Eq. D. 10. Here we note from Eqs. C. 1b, C. 2b and Eq. 2. 17 of Section 2.3 that for $\theta_1 \epsilon [0, \Delta)$, $p(\theta_2|\theta_1)$ is non-zero only for θ_2 satisfying

$$\theta_1 \leq \theta_2 \leq T_p$$

This region is shown as the cross-hatched area in Fig. D.1.

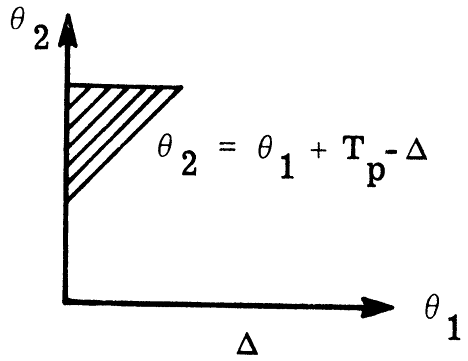


Fig. D.1. The region in which $p(\theta_2 | \theta_1)$, $\theta_1 \epsilon [0, \Delta)$, is positive

From this figure it is clear that the integration with respect to $\mu(d\theta_1)$ need be performed only for θ_1 satisfying

$$0 \leq \theta_1 \leq \theta_2 - (T_p - \Delta) \qquad \theta_2 \epsilon [T_p - \Delta, T_p)$$

and $$0 \leq \theta_1 < \Delta \qquad \theta_2 = T_p$$

Thus, if we define the subset $\tilde{\Theta}_1(\theta_2) \subset \Theta_1$ by

$$\tilde{\Theta}_1(\theta_2) = \begin{cases} [0, \theta_2 - (T_p - \Delta)] & \theta_2 \epsilon [T_p - \Delta, T_p) \\ [0, \Delta) & \theta_2 = T_p \end{cases}$$

then we may write the first integral as

$$I_{[T_p - \Delta, T_p]}(\theta_2) \int_{\tilde{\Theta}_1(\theta_2)} f(\theta_2, \theta_1)\, p(\theta_2 | \theta_1)\, \mu(d\theta_1) \qquad \text{(D. 12)}$$

Now substitute Eqs. D. 11 and D. 12 into Eq. D. 10 to obtain

$$\int_{[0, \Delta) \cup \{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2 | \theta_1)\, \mu(d\theta_1)$$

$$= I_{[T_p - \Delta, T_p]}(\theta_2) \left[\int_{\tilde{\Theta}_1(\theta_2)} f(\theta_2, \theta_1)\, p(\theta_2 | \theta_1)\, \mu(d\theta_1) + f(\theta_2, \theta_1 = T_p)\, p(\theta_2 | \theta_1 = T_p) \right]$$

or

$$\int_{[0, \Delta) \cup \{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2 | \theta_1)\, \mu(d\theta_1) \qquad \text{(D. 13)}$$

$$= I_{[T_p - \Delta, T_p]}(\theta_2) \left[\int_{\tilde{\Theta}_1(\theta_2)} f(\theta_2, \theta_1)\, p(\theta_2 | \theta_1)\, \mu(d\theta_1) + \int_{\{T_p\}} f(\theta_2, \theta_1)\, p(\theta_2 | \theta_1)\, \mu(d\theta_1) \right]$$

Substitution of Eq. D. 13 into Eq. D. 9 and combining the integrals over $\tilde{\Theta}_1(\theta_2)$ and $\{T_p\}$ into a single integral completes the proof.

The result of interest in this appendix is provided by the following corollary.

Corollary D. 2.

$$p_2(\theta_2) = I_{[T_p - \Delta,\, T_p]}(\theta_2) \left[\int_{\tilde{\Theta}_1(\theta_2)} p(\theta_2 | \theta_1) p_1(\theta_1) \mu(d\theta_1) + p(\theta_2 | \theta_1 = T_p)\, p_1(\theta_1 = T_p) \right] + I_{[0,\, T_p - \Delta)}(\theta_2)\; p_1(\theta_2 + \Delta) \qquad \text{(D. 14)}$$

Proof. This result follows immediately from the above theorem by taking $f(\theta_2, \theta_1) = p_1(\theta_1)$ and then rating that $\Pi(d\theta_2) = \omega(d\theta_2)$.

To compute the density $p_2(\theta_2)$ from Eq. D. 14 we first need the density $p_1(\theta_1)$. In the continuous case it is quickly concluded from Eq. 2.9 in the text and the definition of θ_1 that $p_1(\theta_1)$ is given by

$$p_1(\theta_1) = \begin{cases} (T_p + \alpha^{-1})^{-1} & 0 \le \theta_1 < T_p \\ \alpha^{-1}(T_p + \alpha^{-1})^{-1} & \theta_1 = T_p \end{cases} \qquad \text{(D. 15)}$$

In the discrete case we appeal to Eq. 2.12 in the text to conclude that

$$p_1(\theta_1) = \begin{cases} [q + a/(1-a)]^{-1} & 0 \le \theta_1 < T_p \\ [a/(1-a)]\,[q + a/(1-a)]^{-1} & \theta_1 = T_p \end{cases} \qquad \text{(D. 16)}$$

In both cases we have

$$p_2(\theta_1) = \begin{cases} K_1 & 0 \leq \theta_1 < T_p \\ K_2 K_1 & \theta_1 = T_p \end{cases} \tag{D.17}$$

where the definitions of K_1 and K_2 are provided by Eqs. D. 15 and D. 16.

Now substitute for $p_1(\theta_1)$ from Eq. D. 17 into Eq. D. 14 to obtain

$$p_2(\theta_2) = \begin{cases} K_1 & \theta_2 \epsilon [0, T_p - \Delta) \\ K_1 \left[\int_{\tilde{\Theta}_1(\theta_2)} p(\theta_2 | \theta_1) \mu(d\theta_1) + p(\theta_2 | \theta_1 = T_p) K_2 \right] & \theta_2 \epsilon [T_p - \Delta, T_p] \end{cases} \tag{D.18}$$

Now for $\theta_2 \epsilon [0, T_p - \Delta)$, $p_2(\theta_2) = K_1$ which is the desired result. The next step in the derivation is to show that for $\theta_2 \epsilon [T_p - \Delta, T_p)$, the bracketed quantity is equal to unity, for then it follows that for all $\theta_2 \epsilon [0, T_p)$, $p_2(\theta_2) = K_1$. Finally the derivation is completed by showing that for $\theta_2 = T_p$, the bracketed quantity is equal to K_2, for then we have $p_2(\theta_2 = T_p) = K_1 K_2$ as desired.

We shall treat the continuous case first. For notational convenience we shall denote the bracketed quantity in Eq. D. 18 as $\mathscr{B}(\theta_2)$. Suppose first that $\theta_2 \epsilon [T_p - \Delta, T_p)$. Then substitute for $p(\theta_2 | \theta_1)$ from Eq. C. 2b of Appendix C into the expression for $\mathscr{B}(\theta_2)$, and note that for the range of interest $\mu(d\theta_2) = d\theta_2$. The result is

$$
\begin{aligned}
\mathscr{B}(\theta_2) &= \int_0^{\theta_2 - (T_p - \Delta)} \alpha \; e^{-\alpha(\theta_2 - (\theta_1 + T_p - \Delta))} d\theta_1 + K_2 K_1^{-1} e^{-\alpha(\theta_2 - (T_p - \Delta))} \\
&= 1 - e^{-\alpha(\theta_2 - (T_p - \Delta))} + e^{-\alpha(\theta_2 - (T_p - \Delta))} \\
&= 1
\end{aligned}
$$

Next suppose $\theta_2 = T_p$. Then, if the expression for $p(\theta_2 = T_p | \theta_1)$ from Eq. C. 2b is substituted into the expression for $\mathscr{B}(\theta_2 = T_p)$, the result is

$$
\begin{aligned}
\mathscr{B}(\theta_2 = T_p) &= \int_0^{\Delta} e^{-\alpha(\Delta - \theta_1)} d\theta_1 + K_2 \; e^{-\alpha \Delta} \\
&= \alpha^{-1}(1 - e^{-\alpha\Delta}) + K_2 \; e^{-\alpha\Delta} \\
&= K_2
\end{aligned}
$$

since

$$
\alpha_1^{-1} = K_2
$$

Now consider the discrete case. Suppose first that $\theta_2 \epsilon [T_p - \Delta, T_p)$. Then by substituting for $p(\theta_2|\theta_1) = \Pr[\theta_2|\theta_1]$ from Eq. 2.17 in the text and noting from Eq. C.3b of Appendix C that

$$\mu(d\theta_2) = \sum_{j=1}^{q-1} \epsilon_{\{j\nu\}}(d\theta_2)$$

we obtain

$$\mathscr{B}(\theta_2) = \sum_{j=0}^{j=\ell-(q-s)} (1-a)\, a^{\ell-(j+q-s)} + K_2(1-a)\, a^{\ell-(q-s)}$$

$$= (1-a)\, a^{\ell-(q-s)} \sum_{j=0}^{\ell-(q-s)} (a^{-1})^j + a^{\ell-(q-s)+1}$$

$$= (1-a)\, a^{\ell-(q-s)} \frac{1-(a^{-1})^{\ell-(q-s)+1}}{(1-a^{-1})} + a^{\ell-(q-s)+1}$$

$$= -a^{\ell-(q-s)+1} + 1 + a^{\ell-(q-s)+1}$$

$$= 1$$

Finally suppose $\theta_2 = T_p$. Then the substitution for $p(\theta_2 = T_p|\theta_1)$ from Eq. 2.17 into the expression for $\mathscr{B}(\theta_2 = T_p)$ yields

$$\mathcal{B}(\theta_2) = \sum_{j=0}^{j=s-1} a^{s-j} + K_2 a^s$$

$$= a^s \sum_{j=0}^{s-1} (a^{-1})^j + \frac{a^{s+1}}{1-a}$$

$$= a^s \frac{(1-a^{-s})}{(1-a^{-1})} + \frac{a^{s+1}}{1-a}$$

$$= \frac{a}{1-a} = K_2$$

This completes the derivation of $p(\theta_2)$.

To complete the logic of this appendix we need to prove the lemma used in the proof of Theorem D. 1.

Lemma D. 3. Let $f(\theta_2, \theta_1)$ be any function which is measurable with respect to $\mu(d\theta_1)$ for each θ_2 and let A be any measurable subset of $[0, T_p)$. Then

$$\int_A \left[\int_{[\Delta, T_p)} \epsilon_{\{\theta_1 - \Delta\}} (d\theta_2)\, f(\theta_2, \theta_1)\mu(d\theta_1) \right]$$

$$= \int_A I_{[0, T_p - \Delta)}(\theta_2) \quad f(\theta_2, \theta_1 = \theta_2 + \Delta)\mu(d\theta_2) \qquad \text{(D. 19)}$$

Proof. Write the left hand side of Eq. D. 19 as

$$\int_{[\Delta, T_p)} \left[\int_A f(\theta_2, \theta_1)\ \epsilon_{\{\theta_1 - \Delta\}}(d\theta_2) \right] \mu(d\theta_1)$$

Now consider the inner integral. We have

$$\int_A f(\theta_2, \theta_1)\ \epsilon_{\{\theta_1 - \Delta\}}(d\theta_2) = \begin{cases} f(\theta_2 = \theta_1 - \Delta, \theta_1) & \theta_1 - \Delta \in A \\ 0 & \text{otherwise} \end{cases}$$

or equivalently

$$\int_A f(\theta_2, \theta_1)\ \epsilon_{\{\theta_1 - \Delta\}}(d\theta_2) = I_{\{A + \Delta\}}(\theta_1)\, f(\theta_2 = \theta_1 - \Delta, \theta_1)$$

where

$$A + \Delta = \{y' \mid y' = y + \Delta \text{ for } y \in A\}$$

Thus, the total integral can be written as

$$\int_{[\Delta, T_p)} I_{\{A + \Delta\}}(\theta_1)\, f(\theta_2 = \theta_1 - \Delta, \theta_1) \mu(d\theta_1)$$

$$= \int_{[\Delta, T_p) \cap (A + \Delta)} f(\theta_2 = \theta_1 - \Delta, \theta_1)\ \mu_{[0, T_p)}(d\theta_1) \qquad \text{(D.20)}$$

where we used the fact that on $[0, T_p)$, $\mu(d\theta_1) = \mu_{[0, T_p)}(d\theta_1)$.
Now make a change of variable by letting

$$\tilde{\theta}_2 = \theta_1 - \Delta$$

Then we have

$$[\Delta, T_p) \longrightarrow [0, T_p - \Delta)$$

$$A + \Delta \longrightarrow A$$

so that

$$[\Delta, T_p) \cap (A + \Delta) \longrightarrow [0, T_p - \Delta) \cap A$$

Furthermore since $\mu_{[0, T_p)}$ is either the Lebesque measure or the counting measure it follows that for any set $A' \subset [0, T_p - \Delta)$

$$\mu_{[0, T_p)}(A' + \Delta) = \mu_{[0, T_p)}(A')$$

Thus we may write the integral in Eq. D.20 as

$$\int_{[0, T_p - \Delta) \cap A} f(\tilde{\theta}_2, \theta_1 = \tilde{\theta}_2 + \Delta)\mu_{[0, T_p)}(d\tilde{\theta}_2)$$

$$= \int_A I_{[0, T_p - \Delta)}(\tilde{\theta}_2) \quad f(\tilde{\theta}_2, \theta_1 = \tilde{\theta}_2 + \Delta) \quad (d\tilde{\theta}_2)$$

which is the desired result.

APPENDIX E

EXPRESSIONS FOR THE RATES r_p, r_N AND r_S

In this appendix we calculate the rates r_p, r_N and r_S.

The temporal relation of the random variables θ_i is shown in Fig. B.1 of Appendix B. It is evident from this figure that $N_p(\theta)$ can be written as

$$N_p(\theta) = \sum_{i=m}^{N-1} I_{\{\theta_i \epsilon [0, \Delta)\}}(\theta_i) + I_{\{\theta_N \epsilon [0, T_p)\}}(\theta_N)$$

Taking expected values and dividing through by T results in

$$r_p = \frac{E\{N_p(\theta)\}}{T}$$

$$= \frac{E\left\{\sum_{i=m}^{N-1} I_{\{\theta_i \epsilon [0, \Delta)\}}(\theta_i) + I_{\{\theta_N \epsilon [0, T_p)\}}(\theta_N)\right\}}{T}$$

$$= \frac{\sum_{i=m}^{N-1} E\left\{I_{\{\theta_i \epsilon [0, \Delta)\}}(\theta_i)\right\} + E\left\{I_{\{\theta_N \epsilon [0, T_p)\}}(\theta_N)\right\}}{T}$$

or

$$r_p = \frac{(N-m)\,Pr[\theta_i \epsilon [0, \Delta)] + Pr[\theta_N \epsilon [0, T_p)]}{T} \qquad (E.1)$$

Next, from Eqs. 2.18 and 2.19 of Section 2.3 we note that in both the continuous case and the discrete case we have

$$Pr[\theta_N \epsilon [0, T_p)] = m\,Pr[\theta_i \epsilon [0, \Delta)] \qquad (E.2)$$

Now substitute Eq. E.2 into Eq. E.1 and write T as $N\Delta$. The result is

$$r_p = \frac{(N-m)\,Pr[\theta_i \epsilon [0, \Delta)] + m\,Pr[\theta_i \epsilon [0, \Delta)]}{N\Delta}$$

or

$$r_p = \frac{Pr[\theta_i \epsilon [0, \Delta)]}{\Delta} \qquad (E.3)$$

Finally, divide numerator and denominator by m and note that $T_p = m\Delta$. Then the resulting equation for r_p is

$$r_p = \frac{Pr[\theta_i \epsilon [0, T_p)]}{T_p} \qquad (E.4)$$

Next expressions for r_N and r_S are derived. From the definitions of N_N and N_S in Section 3.2, it is seen that $N_N(\theta)$ and $N_S(\theta)$ can be written as

$$N_N(\theta) = \sum_{i=1}^{N} I_{\{\theta_i = T_p\}}(\theta_i)$$

$$N_S(\theta) = \sum_{i=1}^{N} I_{\{\theta_i \epsilon [0, T_p)\}}(\theta_i)$$

Dividing by T and taking expected values results in

$$r_N = \frac{\sum_{i=1}^{N} Pr[\theta_i = T_p]}{T}$$

$$r_S = \frac{\sum_{i=1}^{N} Pr[\theta_i \epsilon [0, T_p)]}{T}$$

or

$$r_N = \frac{N\, Pr[\theta = T_p]}{N\Delta} = \frac{Pr[\theta = T_p]}{\Delta} \quad (E.5)$$

and

$$r_S = \frac{N\, Pr[\theta \epsilon [0, T_p)]}{N\Delta} = \frac{Pr[\theta \epsilon [0, T_p)]}{\Delta} \quad (E.6)$$

APPENDIX F

THE PROOF OF THE MOST RECENT RESPONSE RULE

In this appendix we prove that each performance point in the set $\mathscr{S}$ can be obtained by a decision device that satisfies the most recent response rule.

Theorem F. 1. If $(y_1, y_2, y_3) \in \mathscr{S}_T(m)$, then there exists a $\delta \in \mathscr{D}(m)$ such that $y_1 = R_X(\delta)$, $y_2 = R_X(\delta)$ and $y_3 = R_D(\delta)$ and

$$\delta_i(\vec{x}_i, \vec{a}_{i-1}) = \delta_i(\vec{x}_i, \tau_i)$$

Proof. Since $(y_1, y_2, y_3) \in \mathscr{S}_T(m)$ implies that there is at least one receiver δ which has $y_1 = R_X(\delta)$, $y_2 = R_F(\delta)$ and $y_3 = R_D(\delta)$ it is sufficient to show that r_X, r_F and r_D can be written in terms of a decision rule whose component rules can be expressed in the form

$$a_i = \delta_i(\vec{x}_i, \tau_i)$$

To this end consider first the expression for $r_D(\delta)$:

$$r_D(\delta) = \sum_{i=1}^{N} \frac{\underset{x,\theta}{E}\left\{ I_{\{a_i = 1,\ \tau_i \leq \theta_i,\ \theta_i \in [0, T_p)\}}(\tau_i, a_i, \theta_i) \right\}}{T}$$

The ith term in this sum can be written as

$$P_i^D = \underset{x,\,\theta}{E}\left\{ I_{\{a_i=1,\ \tau_i \le \theta_i,\ \theta_i \in [0,\,T_p)\}}(\tau_i\,,\ a_i\,,\ \theta_i) \right\}$$

But $$a_i = \delta_i(\vec{x}_i\,,\ \vec{a}_{i-1})$$

and $$\tau_i = \hat{\tau}_i(\vec{a}_{i-1}) = \hat{\tau}_i(\vec{x}_{i-1})$$

Thus, the indicator function in the above expression depends only on $\vec{x}_i$ and θ_i . We thus have

$$P_i^D = \underset{\vec{x}_i,\,\theta_i}{E}\left\{ I_{\{a_i=1,\ \tau_i \le \theta_i,\ \theta_i \in [0,\,T_p)\}}(\tau_i\,,\ a_i\,,\ \theta_i) \right\}$$

$$= \underset{\theta_i}{E}\left\{ \underset{\vec{x}_{i-1}|\theta_i}{E}\left\{ \underset{x_i|\vec{x}_{i-1},\,\theta_i}{E}\left\{ I_{\{a_i=1,\ \tau_i \le \theta_i,\ \theta_i \in [0,\,T_p)\}}(\tau_i\,,\ a_i\,,\ \theta_i) \right\}\right\}\right\}$$

Now, since τ_i takes on one of the m values, 0, 1, . . . , m-1, the space of observations up to time τ_{i-1}, $\vec{X}_{i-1}$, can be partitioned into the sets

$$\vec{X}_{i-1}^j = \{\vec{x}_{i-1} | \tau_i = j\Delta\} \qquad j = 0, \ldots, m-1$$

Now, for each $\vec{x}_{i-1}$ define the response set

$$A_i(\vec{x}_{i-1},\ \tau_i = j\Delta) = \{x_i | a_i = 1,\ \vec{x}_{i-1} \in \vec{X}_{i-1}^j,\ x_i \in X_i\}$$

Note that in terms of this set, the decision rule can be written as

$$\delta_i(\vec{x}_i, a_{i-1}, \ldots, a_1) = \begin{cases} 1 & x_i \in A_i(\vec{x}_{i-1}, \tau_i) \\ 0 & x_i \notin A_i(\vec{x}_{i-1}, \tau_i) \end{cases}$$

We are now in the position to evaluate the indicator function. Fix $\theta_i \in [0, T_p)$. Then, for $\vec{x}_{i-1} \in \vec{X}^j_{i-1}$ such that $j\Delta > \theta_i$, $\tau_i > \theta_i$, so that

$$I_{\{a_i = 1,\ \tau_i \leq \theta_i,\ \theta_i \in [0, T_p)\}}(\tau_i, a_i, \theta_i) = 0$$

On the other hand, for $\vec{x}_{i-1} \in \vec{X}^j_{i-1}$ such that $j\Delta \leq \theta_i$, $\tau_i \leq \theta_i$, so that

$$I_{\{a_i = 1,\ \tau_i \leq \theta_i,\ \theta_i \in [0, T_p)\}} = \begin{cases} 1 & x_i \in A_i(\vec{x}_{i-1}, \tau_i = j\Delta) \\ 0 & x_i \notin A_i(\vec{x}_{i-1}, \tau_i = j\Delta) \end{cases}$$

Thus, we can write

$$P_i^D = \int_{\theta_i \epsilon [0,\, T_p)} \left[\int_{\vec{X}_{i-1}} \left[\int_{X_i} I_{\{a_i=1,\, \tau_i \leq \theta_i,\, \theta_i \epsilon [0,\, T_p)\}} (\tau_i, a_i, \theta_i) \; P(dx_i | \vec{x}_{i-1}, \theta_i) \right] \times P(d\vec{x}_{i-1} | \theta_i) \right] \Pi(d\theta_i)$$

$$= \int_{\theta_i \epsilon [0,\, T_p)} \left[\sum_{j \ni j\Delta \leq \theta_i} \left[\int_{\vec{X}^j_{i-1}} \left[\int_{A_i(\vec{x}_{i-1},\, \tau_i = j\Delta)} P(dx_i | \vec{x}_{i-1}, \theta_i) \right] P(d\vec{x}_{i-1} | \theta_i) \right] \Pi(d\theta_i) \right]$$

(F. 1a)

The ith term in the expression for $r_X(\delta)$ can be determined in the same way, the only difference being that instead of summing over $j \ni j\Delta \leq \theta_i$ we must sum over $j \ni j\Delta > \theta_i$. The result is

$$P_i^X = \int_{\theta_i \epsilon [0, T_p)} \left[\sum_{j \ni j\Delta > \theta_i}^{m-1} \left[\int_{\vec{X}^j_{i-1}} \left[\int_{A_i(\vec{x}_{i-1},\, \tau_i = j\Delta)} P(dx_i | \vec{x}_{i-1}, \theta_i) \right] P(d\vec{x}_{i-1} | \theta_i) \right] \Pi(d\theta_i) \right]$$

(F. 1b)

Finally, the ith term in the expression for $r_F(\delta)$ can be written as

$$P_i^F = \int_{\theta_i = T_p} \left[\sum_{j=0}^{m-1} \int_{\vec{X}^j_{i-1}} \left[\int_{A_i(\vec{x}_{i-1},\, \tau_i = j\Delta)} P(dx_i | \vec{x}_{i-1}, \theta_i) \right] P(d\vec{x}_{i-1} | \theta_i) \right] \Pi(d\theta_i)$$

(F. 1c)

This completes the proof since

$$\delta_i(\vec{x}_i, a_{i-1}, \ldots, a_i) = \begin{cases} 1 & x_i \in A_i(\vec{x}_{i-1}, \tau_i) \\ 0 & x_i \notin A_i(\vec{x}_{i-1}, \tau_i) \end{cases}$$

implies that

$$\delta_i(\vec{x}_i, \vec{a}_{i-1}) = \delta_i(\vec{x}_i, \tau_i)$$

and thus δ_i depends only on the time of the most recent response τ_i.

APPENDIX G

THE DERIVATION OF THE RECURSION RELATION FOR THE BAYES DECISION RULE

In this appendix the recursive procedure for determining the Bayes decision rule is derived.

The Bayes decision rule is obtained by finding a δ^o that satisfies

$$r(\delta^o) = \inf_{\delta \in \mathscr{D}} r(\delta) \tag{G.1}$$

where

$$r(\delta) = \frac{\underset{x,\theta}{E} \; L(\delta(x), \theta)}{T} \tag{G.2}$$

We begin by developing an expression for $r(\delta)$. From Section 2.5 we have

$$L(\vec{a}, \vec{\theta}) = \sum_{i=1}^{N} L_i(\tau_i, a_i, \theta_i)$$

where from Eq. 4.1

$$\tau_i = \tau(a_{i-1}, \tau_{i-1})$$

A decision rule δ is an n-tuple $\vec{\delta} = (\delta_1, \ldots, \delta_N)$ where δ_i is a function of the form

$$a_i = \delta_i(\vec{x}_i, \vec{a}_{i-1})$$

where $\vec{a}_i = (a_1, \ldots, a_i)$.

Now consider the function $L(\delta(x), \theta)$. From the above definitions it is clear that

$$L(\delta(x), \theta) = \sum_{i=1}^{N} L_i(\tau_i, \delta_i(\vec{x}_i, \vec{a}_{i-1}), \theta_i) \qquad \text{(G. 3)}$$

where

$$\tau_i = \tau(\delta_{i-1}(\vec{x}_{i-1}, \vec{a}_{i-2}), \tau_{i-1}) \qquad \text{(G. 4)}$$

Substituting Eq. G. 3 into Eq. G. 2 and distributing the expected value operator over the sum results in

$$r(\delta) = \sum_{i=1}^{N} \underset{x,\theta}{E} \{L_i(\tau_i, \delta_i(\vec{x}_i, \vec{a}_{i-1}), \theta_i\}/T$$

Now consider the ith term in the above sum. From Eq. G. 4 it is seen that τ_i depends on the past decisions only through $\vec{\delta}_{i-1}$. Therefore $L_i(\cdot, \cdot, \cdot)$ depends only on $\vec{\delta}_i$. Moreover, since $\vec{\delta}_i$ depends only on the observation up to time τ_i, $\vec{x}_i$, the random variable $L_i(\cdot, \cdot, \cdot)$ depends only on $\vec{x}_i$ and θ_i. Therefore, the expected value need be taken only over $\vec{X}_i \times \Theta_i$. Thus, we may write

$$r(\delta) = \sum_{i=1}^{N} r_i(\vec{\delta}_i) \qquad \text{(G. 5)}$$

where

$$r_i(\vec{\delta}_i) = \underset{\vec{x}_i,\, \theta_i}{E} \{L_i(\tau_i, \delta_i(\vec{x}_i, \vec{a}_{i-1}), \theta_i)\} \tag{G.6}$$

With Eq. G.5 for $r(\delta)$ at hand, we may proceed with the derivation. We seek $\delta^o = (\delta_1^{\,o}, \ldots, \delta_N^{\,o})$ such that

$$r(\delta_1^{\,o}, \ldots, \delta_N^{\,o}) = \inf_{\delta_1, \ldots, \delta_N} r(\delta_1, \ldots, \delta_N)$$

First note that by a well known property of the infinium the above minimization can be carried out over individual components one at a time. That is,

$$\inf_{\delta} r(\delta) = \inf_{\delta_1, \ldots, \delta_{k-1}} \left[\inf_{\delta_k, \ldots, \delta_N} r(\delta_1, \ldots, \delta_N) \right]$$

Now substitute for $r(\delta_1, \ldots, \delta_N)$ from Eq. G.5 into the right hand side above. The result is

$$\inf_{\delta} r(\delta) = \inf_{\delta_1, \ldots, \delta_{k-1}} \left[\inf_{\delta_k, \ldots, \delta_N} \sum_{i=1}^{N} r_i(\vec{\delta}_i) \right]$$

or

$$\inf_{\delta} r(\delta) = \inf_{\vec{\delta}_{k-1}} \left[\sum_{i=1}^{k-1} r_i(\vec{\delta}_i) + \inf_{\delta_k, \ldots, \delta_N} \left[\sum_{i=k}^{N} r_i(\vec{\delta}_i) \right] \right] \tag{G. 7}$$

where $\vec{\delta}_{k-1} = (\delta_1, \ldots, \delta_{k-1})$.

This last equality follows from the fact that $\sum_{i=1}^{k-1} r(\vec{\delta}_i)$ does not depend on $\delta_k, \ldots, \delta_N$. Now define the function $G_k(\vec{\delta}_{k-1})$ to be the second term in Eq. G. 7. Specifically,

$$G_k(\vec{\delta}_{k-1}) = \inf_{\delta_k, \ldots, \delta_N} \left[\sum_{i=k}^{N} r_i(\vec{\delta}_i) \right] \tag{G. 8}$$

Then, in terms of $G_k(\cdot)$, Eq. D. 7 becomes

$$\inf_{\delta} r(\delta) = \inf_{\delta_1, \ldots, \delta_{k-1}} \left[\sum_{i=1}^{k-1} r_i(\vec{\delta}_i) + G_k(\vec{\delta}_{k-1}) \right]$$

$$= G_1$$

Next we obtain a recursive relation satisfied for the functions $G_k(\cdot)$. From the defining Eq. G. 8 we have

$$G_k(\vec{\delta}_{k-1}) = \inf_{\delta_k} \left[\inf_{\delta_{k+1}, \ldots, \delta_N} \left[\sum_{i=k}^{N} r_i(\vec{\delta}_i) \right] \right]$$

$$= \inf_{\delta_k} \left[r_k(\vec{\delta}_k) + \inf_{\delta_{k+1}, \ldots, \delta_N} \left[\sum_{i=k+1}^{N} r_i(\vec{\delta}_i) \right] \right]$$

But

$$\inf_{\delta_{k+1}, \ldots, \delta_N} \left[\sum_{i=k+1}^{N} r_i(\vec{\delta}_i) \right] = G_{k+1}(\vec{\delta}_k)$$

Thus,

$$G_k(\vec{\delta}_{k-1}) = \inf_{\delta_k} \left[r_k(\vec{\delta}_k) + G_{k+1}(\vec{\delta}_k) \right]$$

Equation G. 9 provides the basis for determining the Bayes rule δ^o and the minimum Bayes risk $r(\delta_0)$. First $\delta_N{}^o$ is determined from

$$\begin{aligned} G_N(\vec{\delta}_{N-1}) &= \inf_{\delta_N} r_N(\vec{\delta}_N) \\ &= \inf_{\delta_N} \left[\underset{\vec{x}_N, \theta_N}{E} \{L_N(\tau_N, \delta_N(\vec{x}_N, \vec{a}_{N-1}), \theta_N)\} \right] \end{aligned}$$

or

$$G_N(\vec{\delta}_{N-1}) = \inf_{\delta_N} \left[\underset{\vec{x}_N}{E} \left\{ \underset{\theta_N | x_N}{E} \{L_N(\tau_N, \delta_N(\vec{x}_N, \vec{a}_{N-1}), \theta_N)\} \right\} \right] \tag{G. 10}$$

The procedure is as follows. Recall from Eq. G. 4 that

$$\tau_N = \tau(\delta_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2}), \tau_{N-1})$$

so that for $\vec{\delta}_{N-1}$ fixed, τ_N is a fixed function of the observation $\vec{x}_{N-1}$. Moreover, $\vec{a}_{N-1}$ depends only on $\vec{x}_{N-1}$ through $\vec{\delta}_{N-1}(\vec{x}_{N-1})=\vec{a}_{N-1}$. Thus the inner expectation in Eq. G. 10 depends only on $\vec{x}_N$ and δ_N. It then follows by the standard argument of Appendix A that the δ_N that minimizes Eq. E. 10 is given by

$$\delta_N^o(\vec{x}_N, \vec{a}_{N-1}) = \begin{cases} 1 & \text{if } \underset{\theta_N|\vec{x}_N}{E}\{L_N(\tau_N, 1, \theta_N)\} \leq \underset{\theta_N|\vec{x}_N}{E}\{L_N(\tau_N, 0, \theta_N)\} \\ 0 & \text{otherwise} \end{cases}$$

The function δ_N^o can be written in a more convenient form in terms of indicator functions as follows. For each $\vec{x}_{N-1} \epsilon \vec{X}_{N-1}$ and $\tau_N \epsilon \{0, \ldots, m-1\}$ define the set of current observations $A_N(\vec{x}_{N-1}, \tau_N) \subset X_N$ by

$$\begin{aligned} A_N(\vec{x}_{N-1}, \tau_N) &= \Big\{ x_N \Big| \underset{\theta_N|\vec{x}_N}{E}\{L_N(\tau_N, 0, \theta_N)\} \\ &\quad - \underset{\theta_N|\vec{x}_N}{E}\{L_N(\tau_N, 1, \theta_N)\} \geq 0 \Big\} \\ &= \{x_N | \delta_N^o(\vec{x}_N, \vec{a}_{N-1}) = 1\} \end{aligned} \tag{G. 11}$$

Then $\delta_N^o(\cdot, \cdot)$ can be written as

$$\delta_N^o(\vec{x}_N, \vec{a}_N) = I_{A_N(\vec{x}_{N-1}, \tau_N)}(x_N)$$

or

$$\delta_N^{\,o}(\vec{x}_N, \tau_N) = I_{A_N(x_{N-1}, \tau_N)}(x_N) \qquad (G.12)$$

with the last equality following from the fact that $\delta_N^{\,o}(\cdot\,,\,\cdot)$ depends on the past decisions only through the function τ_N.

Next we obtain an expression for $G_N(\delta_{N-1})$. First it is noted that since $\delta_N^{\,o}(\vec{x}_N, \tau_N)$ minimizes $G_N(\delta_{N-1})$ we have

$$G_{N-1}(\delta_{N-1}) = \underset{\vec{x}_N}{E} \; \underset{\theta_N | \vec{x}_N}{E} \; L_N(\tau_N, \delta_N^{\,o}(\vec{x}_N, \tau_N), \theta_N)$$

$$= \int_{\vec{X}_N} \underset{\theta_N | \vec{x}_N}{E} \; L_N(\tau_N, \delta_N^{\,o}(\vec{x}_N, \tau_N), \theta_N) \; P(d\vec{x}_N)$$

$$= \int_{\vec{X}_{N-1}} \left[\int_{X_N} \underset{\theta_N | x_N}{E} \; L_N(\tau_N, \delta_N^{\,o}(\vec{x}_N, \tau_N), \theta_N) \; P(dx_N | \vec{x}_{N-1}) \right] \times P(dx_{N-1})$$

$$= \underset{\vec{X}_{N-1}}{E} \int_{X_N} \underset{\theta_N | x_N}{E} \; L_N(\tau_N, \delta_N^{\,o}(\vec{x}_N, \tau_N), \theta_N) \; P(dx_N | \vec{x}_{N-1})$$

Now use the fact that

$$\underset{\theta_N|\vec{x}_N}{E} \{L_N(\tau_N, \delta_N^o(\vec{x}_N, \tau_N), \theta_N\}$$

$$= \begin{cases} \underset{\theta_N|\vec{x}_N}{E} \{L_N(\tau_N, 1, \theta_N)\} & \text{on} \quad \vec{X}_{N-1} \, x\, {}^1A_N(\vec{x}_{N-1}, \tau_N) \\ \\ \underset{\theta_N|\vec{x}_N}{E} \{L_N(\tau_N, 0, \theta_N)\} & \text{on} \quad \vec{X}_{N-1} \, x\, {}^0A_N(\vec{x}_{N-1}, \tau_N) \end{cases}$$

where

$${}^0A_N(\vec{x}_{N-1}, \tau_N) = A_N(\vec{x}_{N-1}, \tau_N)$$

to write $G_N(\vec{\delta}_{N-1})$ as

$$G_N(\vec{\delta}_{N-1}) = \underset{\vec{x}_{N-1}}{E} \left\{ \sum_{\ell=0}^{1} \int_{{}^\ell A_N(\vec{x}_{N-1}, \tau_N)} R_N(\vec{x}_N, \tau_N, \ell) \, P(d\vec{x}_N | \vec{x}_{N-1}) \right\} \qquad \text{(G. 13)}$$

where

$$R_N(\vec{x}_N, \tau_N, j) = \underset{\theta_N|\vec{x}_N}{E} \{L_N(\tau_N, j, \theta_N)\}$$

The function δ_{N-1}^o is determined next. This function is defined by

$$G_{N-1}(\vec{\delta}_{N-2}) = \inf_{\delta_{N-1}} [r_{N-1}(\vec{\delta}_{N-1}) + G_N(\vec{\delta}_{N-1})] \qquad \text{(G. 14)}$$

Substitute for $r_{N-1}(\vec{\delta}_{N-1})$ from Eq. G.6 and for $G_N(\vec{\delta}_{N-1})$ from Eq. G.13 into Eq. G.14 to obtain

$$G_{N-1}(\vec{\delta}_{N-2}) =$$

$$\inf_{\delta_{N-1}} \left[\underset{\vec{x}_{N-1}}{E} \left\{ \underset{\theta_{N-1}|\vec{x}_{N-1}}{E} \{L_{N-1}(\tau_{N-1}, \delta_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2}), \theta_{N-1})\} \right\} \right.$$
$$\left. + \underset{\vec{x}_{N-1}}{E} \left\{ \sum_{\ell=0}^{1} \int R_N(\vec{x}_N, \tau_N[\delta_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2}), \tau_{N-1}], \ell)\, P(dx_N|\vec{x}_{N-1})\ {}^{\ell}A_N(\vec{x}_{N-1}, \tau_N[\delta_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2}), \tau_{N-1}]) \right\} \right]$$

or

$$G_{N-1}(\vec{\delta}_{N-2}) = \inf_{\delta_{N-1}} \left[\underset{\vec{x}_{N-1}}{E} \{R_{N-1}(\vec{x}_{N-1}, \tau_{N-1}, \delta_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2}))\} \right] \tag{G.15}$$

where

$$R_{N-1}(\vec{x}_{N-1}, \tau_{N-1}, \delta_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2}))$$

$$= \underset{\theta_{N-1}|\vec{x}_{N-1}}{E} \left\{ L_{N-1}(\tau_{N-1}, \delta_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2}), \theta_{N-1}) \right\}$$

$$+ \sum_{\ell=0}^{1} \int R_N(\vec{x}_N, \tau_N[\delta_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2}), \tau_{N-1}], \ell)\, P(dx_N|\vec{x}_{N-1})\ {}^{\ell}A_N(\vec{x}_{N-1}, \tau_N[\delta_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2}), \tau_{N-1}])$$

Now apply the standard argument to Eq. G. 15 to conclude that $\delta^o_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2})$ is given by

$$\delta^o_{N-1}(\vec{x}_{N-1}, \vec{a}_{N-2}) = \delta^o_{N-1}(\vec{x}_{N-1}, \tau_{N-1}) \tag{G. 16}$$

$$= \begin{cases} 1 & \text{if} \quad R_{N-1}(\vec{x}_{N-1}, \tau_{N-1}, 0) \geq R_{N-1}(\vec{x}_{N-1}, \tau_{N-1}, 1) \\ 0 & \text{otherwise} \end{cases}$$

As before we may express δ^o_{N-1} in terms of indicator functions by defining $A_{N-1}(\vec{x}_{N-2}, \tau_{N-1})$ by

$$^{\ell}A_{N-1}(\vec{x}_{N-2}, \tau_{N-1}) = \{x_{N-1} | \delta^o_{N-1}(\vec{x}_{N-1}, \tau_{N-1}) = \ell\}$$

and we may use the sets $^{\ell}A_{N-1}(\vec{x}_{N-1}, \tau_{N-1})$ to express $G_{N-1}(\vec{\delta}_{N-2})$ as

$$G_{N-1}(\vec{\delta}_{N-2}) =$$

$$\underset{\vec{x}_{N-2}}{E} \left[\sum_{\ell=0}^{1} \int_{^{\ell}A_{N-2}(\vec{x}_{N-2}, \tau_{N-1})} R_{N-1}(\vec{x}_{N-1}, \tau_{N-1}, \ell)\, P(dx_{N-1} | \vec{x}_{N-1}) \right]$$

With the basic character of the recursive procedure established we may conclude that the kth decision rule $\delta_k(\vec{x}_k, \vec{a}_{k-1})$ is given by

$$\delta_k^{\,o}(\vec{x}_k, \vec{a}_{k-1}) = \delta_k^{\,o}(\vec{x}_k, \tau_k) \tag{G. 17a}$$

$$= \begin{cases} 1 & R_k(\vec{x}_k, \tau_k, 0) \leq R_k(\vec{x}_k, \tau_{k-1}) \\ 0 & \text{otherwise} \end{cases}$$

where

$$R_k(\vec{x}_k, \tau_k, j_k) \tag{G. 17b}$$

$$= \underset{\theta_k | \vec{x}_k}{E} \{L_k(\tau_k, j_k, \theta_k)\}$$

$$+ \sum_{\ell=0}^{1} \int_{{}^{\ell}A_{k+1}(\vec{x}_k,\, \tau_{k+1}[j_k,\, \tau_{k-1}])} R_{k+1}(\vec{x}_{k+1}, \tau_{k+1}[j_k, \tau_{k-1}], \ell)\, P(dx_k | \vec{x}_{k-1})$$

and where

$${}^{\ell}A_{k+1}(\vec{x}_k, \tau_{k+1}) = \{x_k ; \delta_{k+1}^{\,o}(\vec{x}_{k+1}, \tau_{k+1}) = \ell\} \tag{G. 17c}$$

A formal proof to this effect can be obtained by the method of induction. The procedure is to assume the result holds for $\delta_{k+1}^{\,o}$, $\ldots$, $\delta_N^{\,o}$ and then to show that the result holds for $\delta_k^{\,o}$ by using precisely the same argument as used to obtain $\delta_{N-1}^{\,o}$. We shall not go through the steps of this argument here.

APPENDIX H

DERIVATION OF THE UPDATING EQUATION FOR THE ODDS RATIO DENSITY

In this appendix, the theorems of Section 4. 3 are proven. We begin by first proving three lemmas. The first lemma establishes a formula which is useful in proving the second and third lemmas. These last two lemmas are then used to prove Theorem 4. 4.

Lemma H. 1. If $\vec{x} = (x_1, \ldots, x_N)$ is conditionally independent for each $\theta \in \Theta$, then, for $j \leq i \leq k$,

$$P(dx_1, \ldots, dx_i \mid \theta_k, \ldots, \theta_j)$$

$$\int_{\Theta_{j-1} \times \cdots \times \Theta_1} P(dx_1, \ldots, dx_i \mid \theta_1, \ldots, \theta_i)\, \Pi(d\theta_{j-1}, \ldots, d\theta_1 \mid \theta_k, \ldots, \theta_j)$$

Proof. By Bayes law,

$$P(dx_1, \ldots, dx_i \mid \theta_k, \ldots, \theta_j) = \frac{\Pi(d\theta_k, \ldots, d\theta_j \mid \vec{x}_i)\, P(dx_1, \ldots, dx_i)}{\Pi(d\theta_k, \ldots, d\theta_j)} \tag{H. 1}$$

Write

$$\Pi(d\theta_k, \ldots, d\theta_j \mid \vec{x}_i) = \int_{\Theta_{j-1} \times \cdots \times \Theta_1} \Pi(d\theta_k, \ldots, d\theta_1 \mid \vec{x}_i)$$

or

$$\Pi(d\theta_k, \ldots, d\theta_j | \vec{x}_i) \qquad \text{(H. 2)}$$

$$= \int_{\Theta_{j-1} x \ldots x \Theta_1} \frac{P(dx_1, \ldots, dx_i | \theta_k, \ldots, \theta_1) \, \Pi(d\theta_k, \ldots, d\theta_1)}{P(dx_1, \ldots, dx_i)}$$

But

$$P(dx_1, \ldots, dx_i | \theta_k, \ldots, \theta_1) = \int_{X_{i+1} x \ldots x X_k} P(dx_1, \ldots, dx_k | \theta_k, \ldots, \theta_1)$$

$$= \int_{X_{i+1} x \ldots x X_k} \prod_{\ell=1}^{k} P(dx_\ell | \theta_\ell, \theta_{\ell-1})$$

$$= \prod_{\ell=1}^{i} P(dx_\ell | \theta_\ell, \ell_{\ell-1}) \int_{X_{i+1} x \ldots x X_k} P(dx_{i+1}, \ldots, dx_k | \theta_k, \ldots, \theta_i)$$

$$= \prod_{\ell=1}^{i} P(dx_\ell | \theta_\ell, \theta_{\ell-1})$$

or

$$P(dx_1, \ldots, dx_i | \theta_k, \ldots, \theta_1) = P(dx_1, \ldots, dx_i | \theta_i, \ldots, \theta_1) \qquad \text{(H. 3)}$$

Substitution of H. 3 into H. 2 gives

$$\Pi(d\theta_k,\ldots,d\theta_j|\vec{x}_i)$$

$$\int_{\Theta_{j-1}\times\ldots\times\Theta_1} \frac{P(dx_1,\ldots,dx_i|\theta_i,\ldots,\theta_1)\,\Pi(d\theta_k,\ldots,d\theta_1)}{P(dx_1,\ldots,dx_i)}$$

or

$$\Pi(d\theta_k,\ldots,d\theta_j|\vec{x}_i) \tag{H. 4}$$

$$\int_{\Theta_{j-1}\times\ldots\times\Theta_1} \frac{P(dx_1,\ldots,dx_i|\theta_i,\ldots,\theta_1)\,\Pi(d\theta_k,\ldots,d\theta_j)\,\Pi(d\theta_{j-1},\ldots,d\theta_1|\theta_k,\ldots,\theta_1)}{P(dx_1,\ldots,dx_i)}$$

Substitution of H. 4 into H. 1 and cancelling the appropriate terms yields the desired result.

<u>Lemma H. 2.</u> If $\vec{x} = (x_1,\ldots,x_N)$ is conditionally independent for each θ, then

$$\Pi(d\theta_k, d\theta_{k-1}|\vec{x}_k) = \frac{P(dx_k|\theta_k,\theta_{k-1})\,\Pi(d\theta_k, d\theta_{k-1}|\vec{x}_{k-1})}{P(dx_k|\vec{x}_{k-1})}$$

<u>Proof.</u>

We begin with Bayes law:

$$\Pi(d\theta_k, d\theta_{k-1} | \vec{x}_k) = \frac{P(dx_1, \dots, dx_k | \theta_k, \theta_{k-1}) \, \Pi(d\theta_k, d\theta_{k-1})}{P(dx_1, \dots, dx_k)}$$

$$= \frac{P(dx_k | \vec{x}_{k-1}, \theta_k, \theta_{k-1}) \, P(dx_1, \dots, dx_{k-1} | \theta_k, \theta_{k-1}) \, \Pi(d\theta_k, d\theta_{k-1})}{P(dx_k | \vec{x}_{k-1}) \, P(dx_1, \dots, dx_{k-1})}$$

or

$$\Pi(d\theta_k, d\theta_{k-1} | x_1, \dots, x_k) = \frac{P(dx_k | \vec{x}_{k-1}, \theta_k, \theta_{k-1}) \, \Pi(d\theta_k, d\theta_{k-1} | \vec{x}_{k-1})}{P(dx_k | \vec{x}_{k-1})} \tag{H. 5}$$

Now, consider $P(dx_k | \vec{x}_{k-1}, \theta_k, \theta_{k-1})$

$$= P(dx_1, \dots, dx_k | \theta_k, \theta_{k-1}) / P(dx_1, \dots, dx_{k-1} | \theta_k, \theta_{k-1})$$

Apply Lemma H. 1 to the numerator and denominator of the above. The result is

$$P(dx_k | \vec{x}_{k-1}, \theta_k, \theta_{k-1}) \tag{H. 6}$$

$$= \frac{\int_{\Theta_{k-2}} P(dx_1, \dots, dx_k | \theta_k, \dots, \theta_1) \, \Pi(d\theta_{k-2}, \dots, d\theta_1 | \theta_k, \theta_{k-1})}{\int_{\Theta_{k-2}} P(dx_1, \dots, dx_{k-1} | \theta_{k-1}, \dots, \theta_1) \, \Pi(d\theta_{k-2}, \dots, d\theta_1 | \theta_k, \theta_{k-1})}$$

But, by using the conditional independence of $x_1, \dots, x_N$, the numerator of (H. 6) can be written as

$$P(x_k | \theta_k, \theta_{k-1}) \int_{\Theta_{k-2}} P(dx_1, \ldots, dx_{k-1} | \theta_{k-1}, \ldots, \theta_1)$$

$$\cdot \Pi(d\theta_{k-2}, \ldots, d\theta_1 | \theta_k, \theta_{k-1})$$

Substituting into (H. 6) and cancelling gives

$$P(dx_k | \vec{x}_{k-1}, \theta_k, \theta_{k-1}) = P(dx_k | \theta_k, \theta_{k-1}) \tag{H. 7}$$

Finally, substitution of (H. 7) into (H. 5) completes the proof.

The above two lemmas have used only the assumption of conditional independence. The third lemma also requires the Markov property.

<u>Lemma H. 3.</u> If $\vec{x} = (x_1, \ldots, x_N)$ is conditionally independent and θ is Markov, then

$$\Pi(d\theta_k | \theta_{k-1}, \vec{x}_{k-1}) = \Pi(d\theta_k | \theta_{k-1})$$

<u>Proof.</u> We have

$$\Pi(d\theta_k | \theta_{k-1}, \vec{x}_{k-1}) = \frac{\Pi(d\theta_k, d\theta_{k-1} | \vec{x}_{k-1})}{\Pi(d\theta_{k-1} | \vec{x}_{k-1})}$$

$$= \frac{P(dx_1, \ldots, dx_{k-1} | \theta_k, \theta_{k-1}) \; \Pi(d\theta_k, d\theta_{k-1})}{P(dx_1, \ldots, dx_{k-1} | \theta_{k-1}) \quad \Pi(d\theta_{k-1})}$$

$$= \left[\frac{P(dx_1, \ldots, dx_{k-1} | \theta_k, \theta_{k-1})}{P(dx_1, \ldots, dx_{k-1} | \theta_{k-1})} \right] \Pi(d\theta_k | \theta_{k-1})$$

The proof is completed if it can be shown that the bracketed quantity is equal to unity. Applying Lemma H. 1 to both the denominator and the numerator of the bracketed quantity results in

$$\frac{\int_{\Theta_{j-2} \times \ldots \times \Theta_1} P(dx_1, \ldots, dx_{k-1} | \theta_{k-1}, \ldots, \theta_1) \, \Pi(d\theta_{k-2}, \ldots, d\theta_1 | \theta_k, \theta_{k-1})}{\int_{\Theta_{j-2} \times \ldots \times \Theta_1} P(dx_1, \ldots, dx_{k-1} | \theta_{k-1}, \ldots, \theta_1) \, \Pi(d\theta_{k-2}, \ldots, d\theta_1 | \theta_k)}$$

This quantity is equal to unity if

$$\Pi(d\theta_{k-2}, \ldots, d\theta_1 | \theta_k, \theta_{k-1}) = \Pi(d\theta_{k-2}, \ldots, d\theta_1 | \theta_k)$$

But this fact follows easily from the Markov assumption on θ . This completes the proof. We are now in a position to prove Theorem 4. 4.

<u>Theorem 4. 4.</u> If $x = (x_1, \ldots, x_N)$ is conditionally independent and if θ is Markov, then

$$\Pi(d\theta_k | \vec{x}_k) = \int_{\Theta_{k-1}} \frac{P(dx_k | \theta_k, \theta_{k-1})}{P(dx_k | \vec{x}_{k-1})} \Pi(d\theta_k | \theta_{k-1}) \, \Pi(d\theta_{k-1} | \vec{x}_{k-1})$$

Proof. From Lemma H.2 we have

$$\Pi(d\theta_k|\vec{x}_k) = \int_{\Theta_{k-1}} \Pi(d\theta_k, d\theta_{k-1}|\vec{x}_k)$$

$$= \int_{\Theta_{k-1}} \frac{P(dx_k|\theta_k, \theta_{k-1})}{P(dx_k|x_1, \ldots, x_{k-1})} \Pi(d\theta_k, d\theta_{k-1}|\vec{x}_{k-1})$$

Now,

$$\Pi(d\theta_k, d\theta_{k-1}|\vec{x}_{k-1}) = \Pi(d\theta_k|\theta_{k-1}, \vec{x}_{k-1}) \Pi(d\theta_{k-1}|\vec{x}_{k-1})$$

$$= \Pi(d\theta_k|\theta_{k-1}) \Pi(d\theta_{k-1}|\vec{x}_{k-1})$$

Substitution completes the proof.

Next we prove Theorem 4. 5. To this end it is sufficient to prove that the function

$$\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1}) = \frac{P(dx_k|\theta_k, \theta_{k-1})}{P(dx_k|x_{k-1})}$$

exists. The theorem then follows as a corollary to Theorem 4. 4.

Proof. Consider first the distribution $P(dx_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})$. Clearly, this distribution is absolutely continuous with respect to $P(dx_k|\vec{x}_{k-1})$, since the conditioning of $P(dx_k|\vec{x}_{k-1})$ by θ_k, θ_{k-1}

cannot take away any sets of measures zero. Thus, the Radon-Nykodym theorem can be applied to $P(dx_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})$ to obtain the function $\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})$ satisfying

$$\int_E P(dx_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1}) = \int_a \ell(x_k|x_{k-1}, \theta_k, \theta_{k-1}) P(dx_k|\vec{x}_{k-1})$$

for any measurable set A . But, in Lemma H.2, it is shown that

$$P(dx_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1}) = P(dx_k|\theta_k, \theta_{k-1})$$

so that

$$\int_A P(dx_k|\theta_k, \theta_{k-1}) = \int_A \ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1}) P(dx_k|\vec{x}_{k-1})$$

for all measurable sets A . Thus the function $\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})$ is also a Radon-Nykodym derivative of $P(dx_k|\theta_k, \theta_{k-1})$ with respect to $P(dx_k|\vec{x}_{k-1})$. This completes the proof.

Next we prove Theorem 4.6. Here, it is desired to show that $\Pi(d\theta_k|\vec{x}_k) << \mu(d\theta_k)$ for all $\vec{x}_k \epsilon \vec{X}_k$ and to determine the density $p_k(\theta_k|\vec{x}_k)$ with respect to $\mu(d\theta_k)$. The proof is by induction on k . First, for k=1 we have from Bayes law

$$\Pi(d\theta_1|x_1) = \frac{P(dx_1|\theta_1)}{P(dx_1)} \Pi(d\theta_1)$$

Now, using the fact that $\Pi(d\theta_1) = p(\theta_1)\mu(d\theta_1)$,

$$\Pi(d\theta_1|x_1) = \left[\frac{P(x_1|\theta_2)}{P(dx_2)} \; p(\theta_1)\right] \mu\,(d\theta_1)$$

Then, if A is any measurable set of $\mu(\cdot)$ measure zero, then A also has measure zero with respect to $\Pi(d\theta_1|x_1)$. Thus, $\Pi(d\theta_1|x_1) << \mu(d\theta_1)$ for all $x_1 \epsilon X_1$, and moreover the density $p_1(\theta_1|x_1)$ is given by

$$p_1(\theta_1|x_1) = \frac{P(x_1|\theta_1)}{P(dx_1)} \; p(\theta_1)$$

Next, we establish the induction step. Specifically, it is assumed that $\Pi(d\theta_{k-1}|\vec{x}_{k-1}) << \mu(d\theta_{k-1})$ for all $\vec{x}_{k-1} \epsilon \vec{X}_{k-1}$, and it will be shown that $\Pi(d\theta_k|\vec{x}_k) << \mu(d\theta_k)$ for all $\vec{x}_k \epsilon X_k$.

We begin by noting from Theorem 4. 5 that

$$\Pi(d\theta_k|\vec{x}_k) = \int_{\Theta_{k-1}} \ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})\, \Pi(d\theta_k|\theta_{k-1}) \times \Pi(d\theta_{k-1}|x_{k-1})$$

Now, by the induction hypothesis, we may replace $\Pi(d\theta_{k-1}|\vec{x}_{k-1})$ by $p(\theta_{k-1}|\vec{x}_{k-1})\mu(d\theta_{k-1})$ in the equation above to obtain

$$\Pi(d\theta_k|\vec{x}_k) = \int_{\Theta_{k-1}} [\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})\, p(\theta_{k-1}|\vec{x}_{k-1})] \times$$

$$\Pi(d\theta_k|\theta_{k-1})\,\mu(d\theta_{k-1})$$

Now apply Theorem D. 1 in Appendix D for the general index k and with the function $f(\theta_k, \theta_{k-1})$ replaced by $[\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})$ $p_{k-1}(\theta_{k-1}|\vec{x}_{k-1})]$ to conclude that the density $p_k(\theta_k|\vec{x}_k)$ exists and is given by

$$p(\theta_k|x_k) \qquad \text{(H. 8)}$$

$$= I_{\{[0,T_p-\Delta)\}}(\theta_k) \left[\ell(x|\vec{x}_{k-1}, \theta_k, \theta_{k-1} = \theta_k+\Delta)\, p(\theta_{k-1} = \theta_k+\Delta|\vec{x}_{k-1})\right]$$

$$+ I_{\{[T_p-\Delta, T_p]\}}(\theta_k) \left[\int_{k-1\,\epsilon\,\tilde{\Theta}_{k-1}(\theta_k)\,\cup\,\{T_p\}} \ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})\, p(\theta_k|\theta_{k-1})\, p(\theta_{k-1}|\vec{x}_{k-1})\, \mu(d\theta_{k-1})\right]$$

We conclude this appendix with the proof of Theorem 4. 7.

<u>Proof.</u> First substitute for $p(\theta_k|\vec{x}_k)$ and $p(\theta_k = T_p|\vec{x}_k)$ from Eq. H. 8 into Eq. 4. 30 in the text to obtain

$$O_k(\theta_k \; x_k) =$$

$$\frac{\left[I_{\{[T_p-\Delta, T_p)\}}(\theta_k) \int_{\theta_{k-1} \in [0, \theta_k-(T_p-\Delta)] \cup \{T_p\}} \ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1}) p(\theta_k|\theta_{k-1}) p(\theta_{k-1}|\vec{x}_{k-1}) \mu(d\theta_{k-1}) + I_{\{[0, T_p-\Delta)\}}(\theta_k) \, \ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1} = \theta_k+\Delta) p(\theta_{k-1} = \theta_k+\Delta|\vec{x}_{k-1}) \right]}{\left[\int_{\theta_{k-1} \in [0,\Delta) \cup \{T_p\}} \ell(x_k|\vec{x}_{k-1}, \theta_k = T_p, \theta_{k-1}) p(\theta_k = T_p|\theta_{k-1}) p(\theta_{k-1}|\vec{x}_{k-1}) \mu(d\theta_{k-1}) \right]}$$

Now divide numerator and denominator by $p_{k-1}(\theta_{k-1} = T_p|\vec{x}_{k-1})$. The result is

$$O_k(\theta_k|\vec{x}_k) = \tag{H.9}$$

$$\frac{\left[I_{\{[T_p-\Delta, T_p)\}}(\theta_k) \int_{\theta_{k-1} \in [0, \theta_k-(T_p-\Delta)] \cup \{T_p\}} \ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1}) p(\theta_k|\theta_{k-1}) O_{k-1}(\theta_{k-1}|\vec{x}_{k-1}) \mu(d\theta_{k-1}) + I_{\{[0, T_p-\Delta)\}}(\theta_k) \, \ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1} = \theta_k+\Delta) O_{k-1}(\theta_{k-1} = \theta_k+\Delta|\vec{x}_{k-1}) \right]}{\left[\int_{\theta_{k-1} \in [0,\Delta) \cup \{T_p\}} \ell(x_k|\vec{x}_{k-1}, \theta_k = T_p, \theta_{k-1}) p(\theta_k = T_p|\theta_{k-1}) O_{k-1}(\theta_{k-1}|\vec{x}_{k-1}) \mu(d\theta_{k-1}) \right]}$$

Equation H.9 is in the desired form except for the term $\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})$. Now, the idea is to break out the dependency on $\vec{x}_{k-1}$. This can be done as follows. By hypothesis, $P(dx_k|\theta_k, \theta_{k-1}) << P(dx_k|\theta_k = T_1, \theta_{k-1} = T_p)$.

Therefore, by the Radon-Nykodym theorem, there exists a function $\ell(x_k|\theta_k, \theta_{k-1})$ such that

$$\ell(x_k|\theta_k, \theta_{k-1}) = \frac{P(dx_k|\theta_k, \theta_{k-1})}{P(dx_k|\theta_k = T_p, \theta_{k-1} = T_p)}$$

On the other hand, $P(dx_k|\theta_k = T_p, \theta_{k-1} = T_p) << P(dx_k|\vec{x}_{k-1})$ by Theorem 4.5. Therefore, there also exists a function $\ell(x_k|\vec{x}_{k-1}, \theta_k = T_p, \theta_{k-1} = T_p)$ such that

$$\ell(x_k|\vec{x}_{k-1}, \theta_k = T_p, \theta_{k-1} = T_p) = \frac{P(dx_k|\theta_k = T_p, \theta_{k-1} = T_p)}{P(dx_k|\vec{x}_{k-1})}$$

Then, by the well known chain rule for Radon-Nykodym derivatives, we have

$$\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1}) = \frac{P(dx_k|\theta_k, \theta_{k-1})}{P(dx_k|\vec{x}_{k-1})}$$

$$= \frac{P(dx_k|\theta_k, \theta_{k-1})}{P(dx_k|\theta_k = T_p, \theta_{k-1} = T_p)} \cdot \frac{P(dx_k|\theta_k = T_p, \theta_{k-1} = T_p)}{P(dx_k|\vec{x}_{k-1})}$$

or

$$\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1}) = \ell(x_k|\theta_k, \theta_{k-1}) \cdot \ell(x_k|\vec{x}_{k-1}, \theta_k = T_p, \theta_{k-1} = T_p) \tag{H. 10}$$

Now substitute for $\ell(x_k|\vec{x}_{k-1}, \theta_k, \theta_{k-1})$ from Eq. H. 10 into Eq. H. 9. Next, note that, since the factor $\ell(x_k|\vec{x}_{k-1}, \theta_k = T_p, \theta_{k-1} = T_p)$ does not depend on θ_k and θ_{k-1}, it cancels out. The result is the desired expression.

APPENDIX I

EXPRESSIONS FOR THE ODDS RATIO DENSITY

In this appendix, the assumption of additive noise is applied to the updating equation of Theorem 4.7 to get specific expressions for the odds ratio density.

We begin by considering the function

$$\ell(x_k \mid \theta_k, \theta_{k-1}) = \frac{P(dx_k \mid \theta_k, \theta_{k-1})}{P(dx_k \mid \theta_k = \theta_{k-1} = T_p)} \tag{I. 1}$$

This function is normally viewed as a function on the current observation space X_k with the pulse parameters θ_k, θ_{k-1} held fixed. But, from Eq. 4.21, it is seen that in order to calculate $O_k(\theta_k, x_k, O_{k-1})$, the current observation x_k is held fixed and the parameters θ_{k-1}, θ_k are allowed to vary over certain subsets of $\Theta_{k-1} x\, \Theta_k$. Thus, to implement the update procedure, it is necessary to determine the behavior of $\ell(x_k \mid \theta_k, \theta_{k-1})$ for the values of θ_{k-1}, θ_k that appear in the integrations in Eq. 4.21. Specifically,

$$\theta_k \in [T_p - \Delta, T_p) \quad , \quad \theta_{k-1} \in [0, \theta_k - (T_p - \Delta)] \cup \{T_p\} \tag{I. 2a}$$

$$\theta_k = T_p \quad , \quad \theta_{k-1} \in [0, \Delta) \cup \{T_p\} \tag{I. 2b}$$

and

$$\theta_k \epsilon [0, T_p - \Delta) \quad , \quad \theta_{k-1} = \theta_k + \Delta \tag{I. 3}$$

To evaluate $\ell(x_k | \theta_{k-1})$ on the sets defined above, it is necessary to specify in more detail the probability measures on the right hand side of Eq. I. 1. It may be recalled from Section 2. 4 that, for additive noise, the observation process is defined by

$$x(t) = n(t) + s(t, \theta_k, \theta_{k-1}) \qquad t \epsilon [t_{k-1}, t_k) \tag{I. 4}$$

The function $s(t, \theta_k, \theta_{k-1})$ is given by Eq. 2. 21. This equation can be simplified by denoting the pulse arriving at time $t_k - T_p$ by

$$p^k(t) = p(t - (t_k - T_p)) \tag{I. 5}$$

Then, we may write $s(t, \theta_k, \theta_{k-1})$ as

$$\begin{aligned} &s(t, \theta_k, \theta_{k-1}) \\ &\quad = I_{\{\theta_{k-1} \epsilon [0, \Delta)\}}(\theta_{k-1})\, p^k(t - (\theta_{k-1} - \Delta)) + I_{\{\theta_k \epsilon [0, T_p)\}}(\theta_k)\, p^k(t - \theta_k) \end{aligned} \tag{I. 6}$$

Our first task is to restrict $s(t, \theta_k, \theta_{k-1})$ to those values of θ_k and θ_{k-1} satisfying Eq. I. 2. If these values are substituted into Eq. I. 6, then $s(t, \theta_{k-1}, \theta_k)$ can be written as

$$s(t,\theta_k,\theta_{k-1}) = \begin{cases} 0 & \theta_{k-1}\in\{0\}\cup\{T_p\} & \theta_k = T_p \\ p^k(t-(\theta_{k-1}-\Delta)) & \theta_{k-1}\in(0,\Delta) & \theta_k = T_p \\ p^k(t-\theta_k) & \theta_{k-1}\in\{0\}\cup\{T_p\} & \theta_k\in[T_p-\Delta, T_p) \\ p^k(t-(\theta_{k-1}-\Delta)) & & \\ \quad +p^k(t-\theta_k) & \theta_{k-1}\in(0,\theta_k-(T_p-\Delta)] & \theta_k\in[T_p-\Delta,T_p) \end{cases} \tag{I. 7}$$

where we have excluded the parameter value $\theta_{k-1} = 0$, since this value does not result in a pulse contribution to the interval $[t_{k-1}, t_k)$. The function $s(t, \theta_k, \theta_{k-1})$ is illustrated in Fig. I. 1 for the values of θ_k and θ_{k-1} of Eq. I. 2.

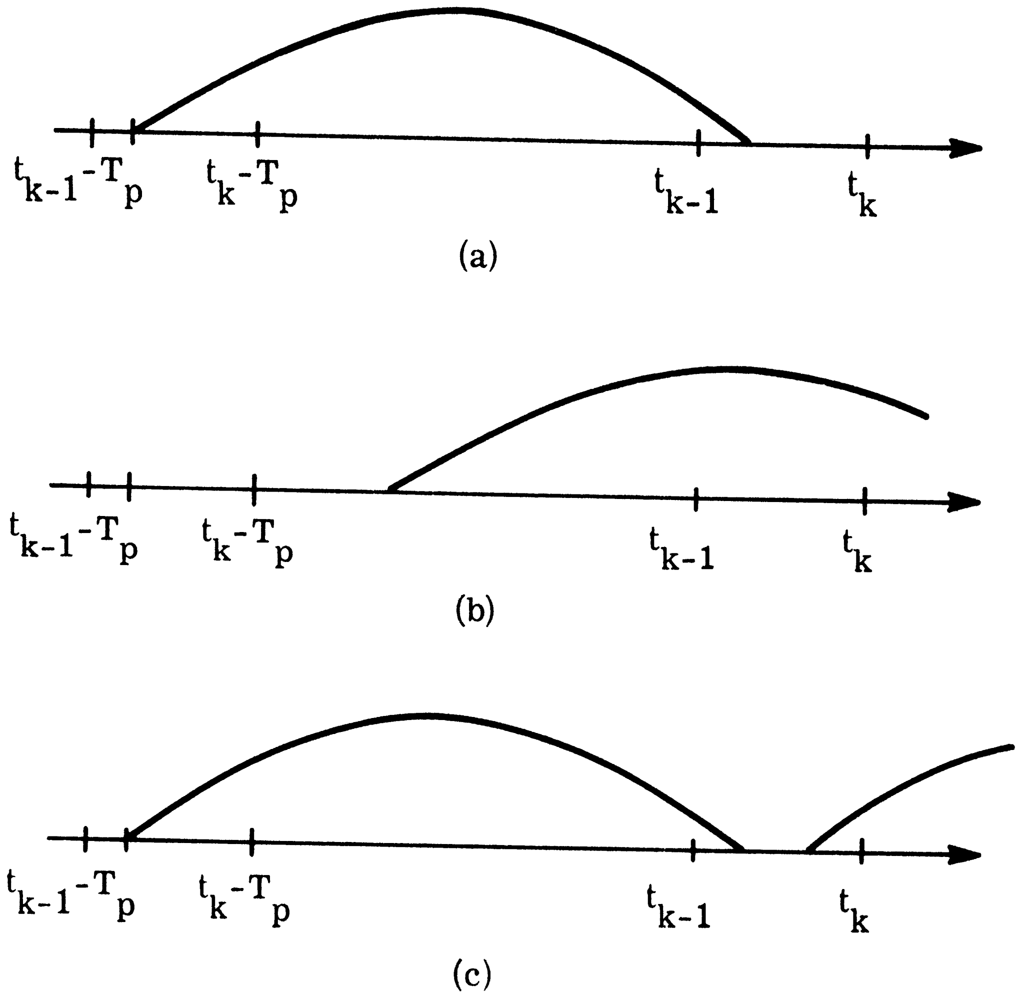


Fig. I. 1. The function $s(t, \theta_k, \theta_{k-1})$ (a) $\theta_{k-1} \epsilon (0, \Delta)$, $\theta_k = T_p$, (b) $\theta_{k-1} = T_p$, $\theta_k \epsilon [T_p - \Delta, T_p)$, (c) $\theta_{k-1} \epsilon [0, \theta_k - (T_p - \Delta)]$, $\theta_k \epsilon [T_p - \Delta, T_p)$

In the above figure, it is seen that the observation x_k in the current observation interval $[t_{k-1}, t_k)$ is influenced either by a pulse that is present at the decision time t_k, $p^k(t-\theta_k)$, or by a pulse that is not present at the decision time t_k, $p^k(t-(\theta_{k-1} - \Delta))$, or both. In the following, we shall refer to the pulse $p^k(t - \theta_k)$ as the "current pulse", and we will refer to the pulse $p^k(t - (\theta_{k-1} - \Delta))$ as the carry-over pulse.

Next, we consider the parameter values of Eq. I. 3. If these values are substituted into Eq. I. 6, the result is

$$s(t, \theta_k, \theta_{k-1}) = p^k(t - \theta_k) \qquad \theta_{k-1} = \theta_k \;\Delta, \;\; \theta_k \in [0, T_p - \Delta) \tag{I. 8}$$

This case is illustrated in Fig. I. 2.

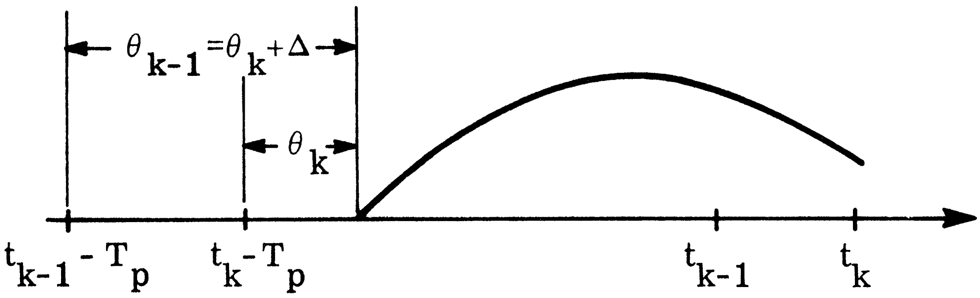


Fig. I. 2. $s(t, \theta_k, \theta_{k-1})$; $\theta_{k-1} \in [0, T_p - \Delta)$, $\theta_k = \theta_{k-1} + \Delta$

We are now in the position to describe the function $\ell(x_k | \theta_k, \theta_{k-1})$. First, substitute for $s(t, \theta_k, \theta_{k-1})$ from Eqs. I. 8 and I. 9 into Eq. I. 4. The result is

$$x(t) = \begin{cases} n(t) & \theta_{k-1} \epsilon \{0\} \cup \{T_p\} \quad , \quad \theta_k = T_p \\ n(t) + p^k(t-(\theta_{k-1}-\Delta)) & \theta_{k-1} \epsilon (0,\Delta) \quad , \quad \theta_k = T_p \\ n(t) + p^k(t-\theta_k) & \theta_{k-1} \epsilon \{0\} \cup \{T_p\} \quad , \quad \theta_k \epsilon [T_p - \Delta, T_p) \\ & \text{or } \theta_{k-1} = \theta_k + \Delta \quad , \quad \theta_k \epsilon [0, T_p - \Delta) \\ n(t) + p^k(t-(\theta_{k-1}-\Delta)) & \\ \quad + p^k(t-\theta_k) & \theta_{k-1} \epsilon [0, \theta_k - (T_p - \Delta)] \, , \quad \theta_k \epsilon [T_p - \Delta, T_p) \end{cases}$$

(I. 9)

Now make the following definitions.

Define $P(dx_k | N)$ to be the distribution of the process

$$x(t) = n(t) \qquad \text{(I. 10a)}$$

Define $P(dx_k | \theta_{k-1}, S_{co})$ to be the distribution of the process

$$x(t) = n(t) + p^k(t-(\theta_{k-1} - \Delta)) \qquad \theta_{k-1} \epsilon (0, \Delta), \ \theta_k = T_p$$

(I. 10b)

Define $P(dx_k | \theta_k, S_c)$ to be the distribution of the process

$$x(t) = n(t) + p^k(t-\theta_k) \qquad \theta_{k-1} \epsilon \{0\} \cup \{T_p\} \quad \theta_k \epsilon [T_p - \Delta, T_p)$$

$$\text{or } \theta_{k-1} = \theta_k + \Delta \qquad \theta_k \epsilon [0, T_p - \Delta)$$

(I. 10c)

Define $P(dx_k | \theta_{k-1}, \theta_k, S_{co}, S_c)$ to be the distribution of the process

$$x(t) = n(t) + p^k(t-(\theta_{k-1} - \Delta)) + p^k(t-\theta_k) \tag{I.10d}$$

$$\theta_{k-1} \in (0, \theta_k - (T_p - \Delta)] \quad \theta_k \in [T_p - \Delta, T_p)$$

Then, we may write

$$\ell(x_k | \theta_k, \theta_{k-1}) \tag{I.11}$$

$$= \begin{cases} 1 & \theta_{k-1} \in \{0\} \cup \{T_p\} & \theta_k = T_p \\ \ell(x_k | \theta_{k-1}, S_{co}) & \theta_{k-1} \in (0, \Delta) & \theta_k = T_p \\ \ell(x_k | \theta_k, S_c) & \theta_{k-1} \in \{0\} \cup \{T_p\} & \theta_k \in [T_p - \Delta, T_p) \\ & \text{or } \theta_{k-1} = \theta_k + \Delta & \theta_k \in [0, T_p - \Delta) \\ \ell(x_k | \theta_k, \theta_{k-1}, S_c, S_{co}) & \theta_{k-1} \in (0, \theta_k - (T_p - \Delta)] & \theta_k \in [T_p - \Delta, T_p) \end{cases}$$

where

$$\ell(x_k | \theta_{k-1}, S_{co}) = P(dx_k | \theta_{k-1}, S_{co}) / P(dx_k | N) \tag{I.12a}$$

$$\ell(x_k | \theta_k, S_c) = P(dx_k | \theta_k, S_c) / P(dx | N) \tag{I.12b}$$

and

$$\ell(x_k | \theta_{k-1}, \theta_k, S_{co}, S_c) = P(dx_k | \theta_{k-1}, \theta_k, S_{co}, S_c) / P(dx_k | N) \tag{I.12c}$$

The functions, $\ell(x_k | \theta_{k-1}, S_{co})$, $\ell(x_k | \theta_k, S_c)$ and $\ell(x_k | \theta_{k-1}, \theta_k, S_{co}, S_c)$ are "ratios" of distributions describing the observation x_k when signal is present to the distribution describing the observation x_k when noise alone is present. In classical detection theory, these functions are usually referred to as conditional likelihood ratios. In this context, $\ell(x_k|\theta_{k-1}, S_{co})$ is the likelihood ratio of the carry-over pulse, $\ell(x_k|\theta_k, S_c)$, is the likelihood ratio of the current pulse, and $\ell(x_k|\theta_k, \theta_{k-1}, S_{co}, S_c)$ is the likelihood ratio of both pulses.

Using the above results, we may now obtain the desired expressions for the odds ratio density. First, we write the update equation for the odds ratio density $O_k(\theta_k, x_k, O_{k-1})$ in terms of the likelihood ratios. This is done by substituting Eq. I. 11 into the update equation, Eq. 4.21, and evaluating the integrals over the set $\{\theta_k = T_p\}$. The result is

$$O_k(\theta_k, x_k, O_{k-1}) = \tag{I. 13}$$

$$\begin{cases} \dfrac{\ell(x_k|\theta_k, S_c)\, O_{k-1}(\theta_{k-1} = \theta_k + \Delta, \vec{x}_{k-1}, O_{k-2})}{D} & \theta_k \in [0, T_p - \Delta) \\[2ex] \dfrac{\begin{aligned} &\int_{\{\theta_{k-1} = 0\}} \ell(x_k|\theta_k, S_c)\, p(\theta_k|\theta_{k-1})\, O_{k-1}(\theta_{k-1}, x_{k-1}, O_{k-2}), \mu(d\theta_{k-1}) \\ &+ \int_{\theta_{k-1} \in (0, \theta_k - T_p - \Delta)]} \ell(x_k|\theta_k, \theta_{k-1}, S_{co}, S_c)\, p(\theta_k|\theta_{k-1})\, O_{k-1}(\theta_{k-1}, x_{k-1}, O_{k-2})\, \mu(d\theta_{k-1}) \\ &+ \ell(x_k|\theta_k, S_c)\, p(\theta_k|\theta_{k-1} = T_p) \end{aligned}}{D} & \theta_k \in [0, T_p - \Delta) \end{cases}$$

where

$$D = \int_{\{\theta_{k-1} = 0\}} p(\theta_k = T_p | \theta_{k-1})\, O_{k-1}(\theta_{k-1}, x_{k-1}, O_{k-2}) \mu(d\theta_{k-1})$$

$$+ \int_{\theta_{k-1} \epsilon (0,\Delta)} \ell(x_k | \theta_{k-1}, S_c)\, p(\theta_k = T_p | \theta_{k-1})\, O_{k-1}(\theta_{k-1}, \vec{x}_{k-1}, O_{k-2})\, \mu(d\theta_{k-1})$$

$$+ \; p(\theta_k = T_p | \theta_{k-1} = T_p)$$

The next step is to substitute for $p(\theta_k | \theta_{k-1})$ from Appendix C. This is accomplished in the continuous case by substituting $p(\theta_k | \theta_{k-1})$ from Eq. C.2b and integrating with respect to the measure

$$\mu(d\theta_k) = d\theta_k + \epsilon_{\{T_p\}}(d\theta_k)$$

and in the discrete case by substituting $p(\theta_k = \ell\nu | \theta_{k-1} = j\nu)$ from Eq. C.1 b and integrating the measure with respect to

$$\mu(d\theta_k) = \sum_{j=1}^{q} \epsilon_{\{j\nu\}}(d\theta_k)$$

In the continuous case, the results are

$$O_k(\theta_k, x_k, O_{k-1}) = \tag{I. 14}$$

$$\begin{cases} \ell(x_k|\theta_k, S_c) \times \left[\dfrac{e^{\alpha\Delta} O_{k-1}(\theta_{k-1} = \theta_k + \Delta, x_{k-1}, O_{k-2})}{\int_0^{\Delta} \ell(x_k|\theta_{k-1}, S_{co})\, e^{\alpha\theta_{k-1}} O_{k-1}(\theta_{k-1}, x_{k-1}, O_{k-2})\, d\theta_{k-1} + 1} \right] & \theta_k \epsilon [0, T_p - \Delta) \\ \alpha\, e^{\alpha(T_p - \theta_k)} \times \left[\dfrac{\int_0^{\theta_k - (T_p - \Delta)} \ell(x_k|\theta_k, \theta_{k-1}, S_{co}, S_c)\, e^{\alpha\theta_{k-1}} O_{k-1}(\theta_{k-1}, x_{k-1}, O_{k-2})\, d\theta_{k-1} + \ell(x_k\ \theta_k, S_c)}{\int_0^{\Delta} \ell(x_k|\theta_{k-1}, S_{co})\, e^{\alpha\theta_{k-1}} O_{k-1}(\theta_{k-1}, x_{k-1}, O_{k-2})\, d\theta_{k-1} + 1} \right] & \theta_k \epsilon [T_p - \Delta, T_p) \end{cases}$$

For the discrete case

$$O_k(\theta_k = \ell\nu, x_k) = \tag{I.15}$$

$$\begin{cases} \ell(x_k \mid \theta_k = \ell\nu, S_c) \\ \quad \left[\dfrac{a^{-s}\, O_{k-1}(\theta_{k-1} = \theta_k + s\nu, x_{k-1}, O_{k-2})}{D}\right] & \ell \in \{0, \ldots, q-s-1\} \\ (1-a)\, a^{-[q-\ell]} \\ \quad \left[\dfrac{\begin{array}{l}\ell(x_k \mid \theta_k = \ell\nu, S_c)\, O_{k-1}(\theta_{k-1} = 0, x_{k-1}, O_{k-2}) \\ \quad + \displaystyle\sum_{j=1}^{j=\ell-(q-s)} \ell(x_k \mid \theta_k = \ell\nu, \theta_{k-1} = j\nu, S_{co}, S_c) \\ \quad a^{-j}\, O_{k-1}(\theta_{k-1} = j\nu, x_{k-1}, O_{k-2}) + \ell(x_k \mid \theta_k = \ell\nu, S_c)\end{array}}{D}\right] & \ell \in \{q-s, \ldots, q-1\} \end{cases}$$

where

$$D = O_{k-1}(\theta_{k-1} = 0, x_{k-1}, O_{k-2}) + \sum_{j=1}^{s-1}\Big[\ell(x_k \mid \theta_{k-1} = j\nu, S_{co})$$

$$a^{-j}\, O_{k-1}(\theta_{k-1} = j\nu, x_{k-1}, O_{k-2})\Big] + 1$$

In concluding this appendix we will consider the special case of white Gaussian noise. In this case, it is well known (Ref. 10) that the likelihood ratio of Eq. I. 12c, $\ell(x_k | \theta_k, \theta_{k-1}, S_{co}, S_c)$, can be written as

$$\exp\left\{ \frac{1}{N_o} \int_{t_{k-1}}^{t_k} x(t)\, s(t, \theta_k, \theta_{k-1})\, dt - \frac{1}{2N_o} \int_{t_{k-1}}^{t_k} s(t, \theta_k, \theta_{k-1})^2\, dt \right\} \tag{I. 16}$$

where N_o is the noise power per unit bandwidth and

$$s(t, \theta_k, \theta_{k-1}) = p^k(t - (\theta_{k-1} - \Delta)) + p^k(t - \theta_k) \tag{I. 17}$$

Now, by definition, the pulses $p^k(t - (\theta_{k-1} - \Delta))$ and $p^k(t - \theta_k)$ do not overlap, so that

$$p^k(t - (\theta_{k-1} - \Delta))\, p^k(t - \theta_k) = 0 \quad \text{for} \quad t \in [t_{k-1}, t_k)$$

It then follows that

$$s^2(t, \theta_k, \theta_{k-1}) = p^k(t - (\theta_{k-1} - \Delta))^2 + p^k(t - \theta_k)^2 \tag{I. 18}$$

Thus, we may substitute Eq. I. 17 and Eq. I. 18 into Eq. I. 16 and separate the terms involving $p^k(t - (\theta_{k-1} - \Delta))$ and $p^k(t - \theta_k)$ to obtain

$$\exp\left\{\frac{1}{N_o}\int_{t_{k-1}}^{t_k} x(t)\, p^k(t-(\theta_{k-1}-\Delta))\, dt - \frac{1}{2N_o}\int_{t_{k-1}}^{t_k} p^k(t-(\theta_{k-1}-\Delta))^2\, dt\right\}$$

$$x\ \exp\left\{\frac{1}{N_o}\int_{t_{k-1}}^{t_k} x(t)\, p^k(t-\theta_k)\, dt - \frac{1}{2N_o}\int_{t_{k-1}}^{t_k} p^k(t-\theta_k)^2\, dt\right\}$$

But, in the white noise case, we also have

$$\ell(x_k|\theta_{k-1}, S_{co}) =$$

$$\exp\left\{\frac{1}{N_o}\int_{t_{k-1}}^{t_k} x(t)\, p^k(t-(\theta_{k-1}-\Delta))\, dt - \frac{1}{2N_o}\int_{t_{k-1}}^{t_k} p^k(t-(\theta_{k-1}-\Delta))^2\, dt\right\}$$

(I. 19)

and

$$\ell(x_k|\theta_k, S_c) =$$

$$\exp\left\{\frac{1}{N_o}\int_{t_{k-1}}^{t_k} x(t)\, p^k(t-(\theta_k)\, dt - \frac{1}{2N_o}\int_{t_{k-1}}^{t_k} p^k(t-\theta_k)^2\, dt\right\} \qquad \text{(I. 20)}$$

Thus, we may conclude that

$$\ell(x_k|\theta_k, \theta_{k-1}, S_{co}, S_c) = \ell(x_k|\theta_{k-1}, S_{co})\,\ell(x_k|\theta_k, S_c) \qquad \text{(I. 21)}$$

The update equation in the white noise case can be obtained by substituting Eq. I. 21 into Eqs. I. 14 and I. 15.

APPENDIX J

THE PROOF OF THEOREM 4.10

In this appendix, we prove Theorem 4.10. To do this, we must show that

$$\tilde{T}_k(\vec{x}_k, \tau_k) = \tilde{T}_k(O_{k-1}, x_k, \tau_k) \tag{J.1}$$

$$R_k(\vec{x}_k, \tau_k, a_k, \theta_k) = R_k(O_k, \tau_k, a_k, \theta_k)$$

$$A_k(\vec{x}_{k-1}, \tau_k) = A_k(O_{k-1}, \tau_k)$$

<u>Proof.</u> The proof is by induction on the index k. For $k = N$ we have

$$\begin{aligned} R_N(\vec{x}_N, \tau_N, a_N, \theta_N) &= L_N(\tau_N, a_N, \theta_N) \\ &= R_N(O_N, \tau_N, a_N, \theta_N) \end{aligned}$$

Thus

$$\begin{aligned} T_N(\vec{x}_N, \tau_N, \theta_N) &= R_N(\vec{x}_N, \tau_N, a_N = 0, \theta_N) - R_N(\vec{x}_N, \tau_N, a_N = 1, \theta_N) \\ &= R_N(O_N, \tau_N, a_N = 0, \theta_N) - R_N(O_N, \tau_N, a_N = 1, \theta_N) \\ &= T_N(O_N, \tau_N, a_N, \theta_N) \end{aligned}$$

and

$$\tilde{T}_N(\vec{x}_N, \tau_N) = \int_{\Theta_N} T_N(\vec{x}_N, \tau_N, \theta_N)\, O_N(\theta_N, x_N, O_{N-1})\, \mu(d\theta_N)$$

$$= \int_{\Theta_N} T_N(O_N, \tau_N, \theta_N)\, O_N(\theta_N, x_N, O_{N-1})\, \mu(d\theta_N)$$

$$= \int_{\Theta_N} T_N(U(O_{N-1}, x_N), \tau_N, \theta_N)\, U(O_{N-1}, x_N)\, \mu(d\theta_N)$$

$$= T_N(O_{N-1}, x_N, \tau_N)$$

and finally

$$A_N(\vec{x}_N, \tau_N) = \{x_N | \tilde{T}_N(\vec{x}_N, \tau_N) \geq 0\}$$

$$= \{x_N | \tilde{T}_N(O_{N-1}, x_N, \tau_N) \geq 0\}$$

$$= A_N(O_{N-1}, \tau_N)$$

Now assume Eq. J. 1 holds for indices $k+1, \ldots, N$. Then,

$$A_{k+1}(\vec{x}_k, \tau_k) = A_{k+1}(O_k, \tau_k)$$

and

$$R_{k+1}(\vec{x}_{k+1}, \tau_{k+1}, a_{k+1}, \theta_{k+1})$$

$$= R_{k+1}(O_{k+1}, \tau_{k+1}, a_{k+1}, \theta_{k+1})$$

$$= R_{k+1}(U(O_k, x_{k+1}), \tau_{k+1}, a_{k+1}, \theta_{k+1})$$

Substitute these quantities into Eq. 4.8 for $R_k(\vec{x}_k, \tau_k, a_k, \theta_k)$ and note that the right hand side that results depends on $\vec{x}_k$ only through O_k . Thus,

$$R_k(\vec{x}_k, \tau_k, a_k, \theta_k) = R_k(O_k, \tau_k, a_k, \theta_k)$$

We may then conclude that

$$\tilde{T}_k(\vec{x}_k, \tau_k) = \tilde{T}_k(O_{k-1}, x_k, \tau_k)$$

and

$$A_k(\vec{x}_{k-1}, \tau_k) = A_k(O_{k-1}, \tau_k)$$

by using precisely the same reasoning as was used for $k = N$. This completes the proof.

APPENDIX K

THE RECURSIVE RELATIONS FOR THE m = 2 BAYES DECISION DEVICE

In this appendix we adapt the recursive equations of Chapter IV to the basic setting of Section 5.1. We begin by obtaining an expression for $\tilde{T}_k(O_{k-1}, x_k, \tau_k)$ by applying the assumptions of Section 5.1 to the formula for $T_k(O_{k-1}, x_k, \tau_k)$ of Section 4.4. This is done in the following theorem.

Theorem K.1. Under the assumptions of Section 5.1

$$a\,\tilde{T}_k(O_{k-1}, x_k, \tau_k) \tag{K.1}$$

$$= \begin{cases} \ell_N \left[(1-a) + [1-(1+W_X)\,a_{N-1}]\,P_{N-1}\right] & k = N \\ \ell_k \left[(1-a) + [1-(1+W_X)\,a_{k-1}]\,P_{k-1}\right] - W_F\, a & \\ + (\ell_k P_{k-1} + a)\,[(F_k^0(O_k, a_k{=}0) - F_k^0(O_k, a_k{=}1)) & \\ \quad + P_k(F_k^1(O_k, a_k{=}0) - F_k^1(O_k, a_k{=}1))] & k < N \end{cases}$$

where

$$F_k^0(O_k, a_k) \tag{K.2a}$$

$$= \sum_{t=0}^{1} \int_{{}^tA_{k+1}(O_k,\,(T_p/2)\,a_k)} R_{k+1}(O_{k+1}, \tau_{k+1}, t, 0)\ P(dx_{k+1} \mid \theta_{k+1} = 0, S_c)$$

and

$$F_k^1(O_k, a_k) \tag{K.2b}$$

$$= \sum_{t=0}^{1} \int_{{}^t A_{k+1}(O_k, T_p/2\ a_k)} \{(1-a)\, R_{k+1}(O_{k+1}, \tau_{k+1}, t, T_p/2)\, P(dx_{k+1} | \theta_{k+1} = 0, S_c)$$

$$+ a\, R_{k+1}(O_{k+1}, \tau_{k+1}, t, T_p)\, P(dx_{k+1} | N)\}$$

and

$$P(dx_{k+1} | S) = P(dx_{k+1} | \theta_{k+1} = 0, S_c) \tag{K.2c}$$

and

$$R_{k+1}(O_{k+1}, \tau_{k+1}, a_{k+1}, \theta_{k+1}) \tag{K.3}$$

$$= \begin{cases} L_{k+1}(\tau_{k+1}, a_{k+1}, \theta_{k+1}) + F_{k+1}^0(O_{k+1}, a_{k+1}) & \theta_{k+1} = 0,\ T_p \\ L_{k+1}(\tau_{k+1}, a_{k+1}, \theta_{k+1}) + F_{k+1}^1(O_{k+1}, a_{k+1}) & \theta_{k+1} = T_p/2 \end{cases}$$

<u>Proof.</u> From Eqs. 4.32a and 4.31b we have

$$\widetilde{T}_k(O_{k-1}, x_k, \tau_k)$$

$$= \int_{\Theta_k} T_k(O_{k-1}, x_k, \tau_k, \theta_k)\, O_k(O_{k-1}, x_k, \theta_k)\, \mu(d\theta_k)$$

Now, since θ_k takes on only the values, 0, $T_p/2$, and T_p, we may evaluate the above integral with respect to the measure

$$\mu(d\theta_k) = \epsilon_{\{0\}}(d\theta_k) + \epsilon_{\{T_p/2\}}(d\theta_k) + \epsilon_{\{T_p\}}(d\theta_k)$$

to obtain

$$\tilde{T}_k(O_{k-1}, x_k, \tau_k) \tag{K. 4}$$

$$= T_k(O_{k-1}, x_k, \tau_k, \theta_k = 0)\, O_k(\theta_k = 0, x_k, O_{k-1})$$

$$+ T_k(O_{k-1}, x_k, \tau_k, \theta_k = T_p/2)\, O_k(\theta_k = T_p/2, x_k, O_{k-1})$$

$$+ T_k(O_{k-1}, x_k, \tau_k, \theta_k = T_p) \times 1 .$$

Next, substitute for $O_k(\theta_k, x_k, O_{k-1})$ from Eq. 5.10 to obtain

$$\tilde{T}_k(O_{k-1}, x_k, \tau_k) \tag{K. 5}$$

$$= \ell_k\, a^{-1}[P_{k-1}\, T_k(O_{k-1}, x_k, \tau_k, \theta_k = 0)$$

$$+ (1-a)\, T_k(O_{k-1}, x_k, \tau_k, \theta_k = T_p/2)]$$

$$+ T_k(O_{k-1}, x_k, \tau_k, \theta_k = T_p)$$

From Eqs. 4.32b and 4.26, we have

$$T_k(O_{k-1}, x_k, \tau_k, \theta_k) \tag{K.6}$$

$$= R_k(O_k, \tau_k, a_k = 0, \theta_k) - R_k(O_k, \tau_k, a_k = 1, \theta_k)$$

where

$$R_k(O_k, \tau_k, a_k, \theta_k) = L(\tau_k, a_k, \theta_k) + F_k(O_k, \tau_{k+1}, \theta_k) \tag{K.7}$$

and where

$$F_k(O_k, \tau_{k+1}, \theta_k) \tag{K.8}$$

$$= \int_{k+1} \left(\sum_{t=0}^{1} \int_{{}^{t}A_{k+1}(O_k, \tau_{k+1})} R_{k+1}(O_{k+1}, \tau_{k+1}, a_{k+1} = t, \theta_{k+1}) \, P(dx_{k+1} | \theta_{k+1}, \theta_k) \right) \Pi(d\theta_{k+1} | \theta_k)$$

Denote the integrand in Eq. K. 8 by $B_{k+1}(O_k, \tau_{k+1}, a_{k+1}, \theta_{k+1}, \theta_k)$. Then,

$$F_k(O_k, \tau_{k+1}, \theta_k) = \int_{\Theta_{k+1}} B_{k+1}(O_k, \tau_{k+1}, a_{k+1}, \theta_{k+1}, \theta_k) \, \Pi(d\theta_{k+1} | \theta_k) \tag{K.9}$$

Next, use the discrete conditional probability distribution with q=2 to evaluate the above integral. The result is

$$F_k(O_k, \tau_{k+1}, \theta_k) \tag{K.10}$$

$$= B_{k+1}(O_k, \tau_{k+1}, a_{k+1}, \theta_{k+1} = 0, \theta_k) \Pr[\theta_{k+1} = 0 | \theta_k]$$

$$+ B_{k+1}(O_k, \tau_{k+1}, a_{k+1}, \theta_k = T_p/2, \theta_k) \Pr[\theta_{k+1} = T_p/2 | \theta_k]$$

$$+ B_{k+1}(O_k, \tau_{k+1}, a_{k+1}, \theta_k = T_p, \theta_k) \Pr[\theta_{k+1} = T_p | \theta_k]$$

Now, substitute for $\Pr[\theta_k = \ell T_p/2 \; \theta_k]$ from Eq. 5.2 into Eq. K.10 to obtain

$$F_k(O_k, \tau_{k+1}, \theta_k) \tag{K.11}$$

$$= \begin{cases} B_{k+1}(O_k, \tau_{k+1}, a_{k+1}, \theta_{k+1} = T_p/2, \theta_k)(1 - a) & \\ \quad + B_{k+1}(O_k, \tau_{k+1}, a_{k+1}, \theta_{k+1} = T_p, \theta_k)\, a & \theta_k = 0, T_p \\ B_{k+1}(O_k, \tau_{k+1}, a_{k+1}, \theta_{k+1} = 0, \; \theta_{k+1} = T_p/2) & \theta_k = T_p/2 \end{cases}$$

The next step is to evaluate $B_{k+1}(\cdot, \cdot, \cdot, \cdot,)$. To this end, consider the distribution $P(dx_{k+1} | \theta_{k+1}, \theta_k)$. This is the distribution of the process

$$x(t) = n(t) + s(t, \theta_{k+1}, \theta_k) \qquad t \in [t_k, t_{k+1})$$

where from Eqs. I.7 and I.8 in Appendix I

$$s(t, \theta_{k+1}, \theta_k) = \begin{cases} p^{k+1}(t) & \theta_{k+1} = 0 \qquad \theta_k = T_p/2 \\ 0 & \theta_{k+1} = T_p \qquad \theta_k = 0,\ T_p \\ p^{k+1}(t - T_p/2) & \theta_{k+1} = T_p/2 \qquad \theta_k = 0,\ T_p \end{cases}$$

But, by the pulse translation assumption of Section 5.1, $p^{k+1}(t) = p^{k+1}(t - T_p/2)$ for $t \in [t_k, t_{k+1})$. Thus, the process $x(t)$ can be written as

$$x(t) = n(t) + p^{k+1}(t) \qquad \theta_{k+1} = 0, \quad \theta_k = T_p/2$$
$$\text{and } \theta_{k+1} = T_p/2, \quad \theta_k = 0,\ T_p \tag{K.12a}$$

and
$$x(t) = n(t) \qquad \theta_{k+1} = T_p, \quad \theta_k = 0,\ T_p \tag{K.12b}$$

Now, from Eq. I.10c, it is seen that the distribution of the process of Eq. K.12a is $P(dx_{k+1} | \theta_{k+1} = 0, S_c)$ and from Eq. I.10c, it is seen that the distribution of the process of Eq. K.12b is $P(dx_{k+1} | N)$. Thus, if we write

$$P(dx_{k+1} | S) = P(dx_{k+1} | \theta_{k+1} = 0,\ S_c)$$

then we may substitute for $P(dx_{k+1} | \theta_{k+1}, \theta_k)$ into the definition of

$B_{k+1}(O_k, \tau_{k+1}, a_{k+1}, \theta_{k+1}, \theta_k)$ and then substitute into **Eq.** K. 11 to obtain

$$F_k(O_k, \tau_{k+1}, \theta_k)$$

$$= \sum_{t=0}^{1} \left[(1-a) \int_{{}^tA_{k+1}(O_k, \tau_{k+1})} R_{k+1}(O_{k+1}, \tau_{k+1}, a_{k+1} = t, O_{k+1} = T_p/2)\ P(dx_{k+1}|S_c) \right.$$

$$\left. + \ a \int_{{}^tA_{k+1}(O_k, \tau_{k+1})} R_{k+1}(O_{k+1}, \tau_{k+1}, a_{k+1} = t, \theta_{k+1} = T_p)\ P(dx_{k+1}|N) \right]$$

for $\theta_{k+1} = 0,\ T_p$, and

$$F_k(O_k, \tau_{k+1}, \theta_k = T_p/2)$$

$$= \sum_{t=0}^{1} \left[\int_{{}^tA_{k+1}(O_k, \tau_{k+1})} R_{k+1}(O_{k+1}, \tau_{k+1}, a_{k+1} = t, \theta_{k+1} = T_p)\ P(dx_{k+1}|S_c) \right]$$

Finally, define $F_k^0(O_k, a_k)$ and $F_k^1(O_k, a_k)$ by

$$F_k^0(O_k, a_k) = F_k(O_k, \tau_{k+1}, \theta_{k+1} = 0) \qquad \text{(K. 13a)}$$

$$F_k^1(O_k, a_k) = F_k(O_k, \tau_{k+1}, \theta_{k+1} = T_p/2) \qquad \text{(K. 13b)}$$

and thus establishes Eqs. K. 1a and K. 1b.

Next, note that Eq. K. 3 then follows immediately from Eq. K. 7 and the definitions of Eq. K. 13.

It remains to verify Eq. K. 2. To do this, substitute Eq. K. 7 into Eq. K. 6 to obtain

$$T_k(O_{k-1}, x_k, \tau_k, a_k)$$

$$= \begin{cases} [L_k(\tau_k, 0, 0) - L_k(\tau_k, 1, 0)] + [F_k^0(O_k, a_k = 0) - F_k^0(O_k, a_k = 1)] \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \theta_k = 0 \\ [L_k(\tau_k, 0, T_p/2) - L_k(\tau_k, 1, T_p/2)] + [F_k^1(O_k, a_k = 0) - F_k^1(O_k, a_k = 1)] \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \theta_k = T_p/2 \\ [L_k(\tau_k, 0, T_p) - L_k(\tau_k, 1, T_p)] + [F_k^0(O_k, a_k = 0) \cdot F_k^0(O_k, a_k = 1)] \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \theta_k = T_p \end{cases}$$

Now, from Eq. 2.27 with $m = 2$

$$L_k(\tau_k, a_k = 0, \theta_k) = 0$$

and

$$L_k(\tau_k, a_k = 1, \theta_k)$$

$$= \begin{cases} \begin{cases} -1 & \tau_k = 0 \\ W_X & \tau_k = T_p \;\; 2 \end{cases} & \theta_k = 0 \\ -1 & \theta_k = T_p/2 \\ W_F & \theta_k = T_p \end{cases}$$

or, equivalently,

$$L_k(\tau_k, a_k = 1, \theta_k)$$

$$= \begin{cases} -[1 - (1 + W_X)\, a_{k-1}] & \theta_k = 0 \\ -1 & \theta_k = T_p/2 \\ W_F & \theta_k = T_p \end{cases}$$

Thus we may write

$$T_k(O_{k-1}, x_k, \tau_k, \theta_k) \quad \text{(K. 14)}$$

$$= \begin{cases} [1 - (1+W_X) a_{k-1}] + [F_k^0(O_k, a_k = 0) - F_k^0(O_k, a_k = 1)] & \theta_k = 0 \\ 1 + [F_k^1(O_k, a_k = 0) - F_k^1(O_k, a_k = 1)] & \theta_k = T_p/2 \\ - W_F + [F_k^0(O_k, a_k = 0) - F_k^1(O_k, a_k = 1)] & \theta_k = T_p \end{cases}$$

Now, substitute Eq. K. 14 into Eq. K. 5 and rearrange to obtain

$$\widetilde{T}_k(O_{k-1}, x_k, \tau_k)$$

$$= \ell_k a^{-1}[(1-a) + [1 - (1 + W_X) a_{k-1}] P_{k-1}] - W_F$$

$$+ \Big\{ \ell_k a^{-1}(1-a) (F_k^1(O_k, a_k = 0) - F_k^1(O_k, a_k = 1))$$

$$+ \ell_k a^{-1} P_{k-1} (F_k^0(O_k, a_k = 0) - F_k^0(O_k, a_k = 1))$$

$$+ (F^0(O_k, a_k = 0) - F^0(O_k, a_k = 1)) \Big\}$$

or

$$a\,\tilde{T}_k(O_{k-1}, x_k, \tau_k) \tag{K.15}$$

$$= \ell_k \left[(1-a) + [1-(1+W_X)a_{k-1}]\,P_{k-1}\right] - W_F\, a$$

$$+ \left\{(\ell_k P_k + a)\left[(F_k^0(O_k, a_k=0) - F_k^1(O_k, a_k=1))\right.\right.$$

$$\left.\left. + \frac{(1-a)\,\ell_k}{(\ell_k P_k + a)}\,(F_k^1(O_k, a_k=0) - F_k^1(O_k, a_k=1))\right]\right\}$$

Finally, note from Eq. 5.11 in Section 5.1 that

$$P_k = \frac{(1-a)\,\ell_k}{a + \ell_k\,P_k}$$

Substitution into Eq. 5.15 results in Eq. K.1. This completes the proof.

Next, we examine the dependence of the functions $\tilde{T}_k(O_{k-1}, x_k, \tau_k)$ and $F_k^1(O_k, a_k)$ and the set $A_k(O_{k-1}, \tau_k)$ on the odds ratio density O_k. In Section 5.1, O_k is written in terms of P_{k-1} and ℓ_k. Thus it seems reasonable to expect that $T_k(\cdot, \cdot, \cdot)$, $F_k^1(\cdot, \cdot)$ and $A_k(\cdot, \cdot)$ can also be written in terms of P_{k-1}, ℓ_k. A statement to this effect constitutes the following theorem.

<u>Theorem K.2.</u> We may write

$$\tilde{T}_k(O_{k-1}, x_k, \tau_k) = \tilde{T}_k'(P_{k-1}, \ell_k, \tau_k) \qquad \text{(K. 16a)}$$

$$F_k^1(O_k, a_k) = F_k^1(P_k, a_k) \qquad \text{(K. 16b)}$$

$$R_{k+1}(O_{k+1}, \tau_{k+1}, a_{k+1}, \theta_{k+1}) = \qquad \text{(K. 16c)}$$

$$R_{k+1}(P_k, \ell_{k+1}, \tau_{k+1}, a_{k+1}, \theta_{k+1})$$

A proof to this theorem can be obtained by induction on the index k . The steps of this proof parallel almost exactly the steps in the proof to Theorem 4. 10 in Appendix J and shall not be repeated here.

Next, note that rather than take the response set $A_k(P_{k-1}, a_{k-1})$ as a subset of the observation space X_k we may instead take the response set as

$$\tilde{A}_k(P_{k-1}, \tau_k) = \left\{ \ell_k \middle| T_k(P_{k-1}, \ell_k, a_{k-1}) \geq 0 \right\} \qquad \text{(K. 17)}$$

This follows from the fact that the response condition

$$T_k(P_{k-1}, \ell_k, a_{k-1}) \geq 0$$

depends on the current observation x_k , only through the likelihood ratio

$$\ell_k(x_k) = \exp\left\{ \frac{1}{N_o} \int_{t_k - T_p/2}^{t_k} x(t)\, p^k(t)\, dt - \frac{1}{2N_o} \int_{t_k - T_p/2}^{t_k} [p^k(t)]^2\, dt \right\}$$

Thus the condition

$$x_k \in A_k(P_{k-1}, \tau_k) = \left\{x_k \middle| \tilde{T}_k(P_{k-1}, \ell_k(x_k), a_{k-1}) \geq 0\right\}$$

is equivalent to the condition

$$\ell_k \in \tilde{A}_k(P_{k-1}, \tau_k) = \left\{\ell_k \middle| \tilde{T}_k(P_{k-1}, \ell_k, a_{k-1}) \geq 0\right\}$$

This fact results in the following theorem.

Theorem K. 3. We may write

$$\int_{A_k(O_{k-1}, \tau_k)} (\quad) P(dx_k | \tfrac{S}{N}) = \int_{\tilde{A}_k(P_{k-1}, \tau_k)} (\quad) f(\ell_k | \tfrac{S}{N}) \, d\ell_k \qquad \text{(K. 18)}$$

where the density $f(\ell_k | \frac{S}{N})$ is obtained from the relation

$$\ell_k = \exp\{z_k\} \qquad \text{(K. 19)}$$

where z_k is a real valued normal random variable with

$$E[z_k | \tfrac{S}{N}] = \pm d_T^2/4 \qquad \text{(K. 20a)}$$

and

$$\mathrm{Var}[z_k | \tfrac{S}{N}] = d_T^2/2 \qquad \text{(K. 20b)}$$

and d_T^2 is the signal to noise of the basic pulse

$$d_T^{\,2} = \frac{1}{N_o} \int_0^{T_p} [p(t)]^2\, dt \qquad \text{(K. 21)}$$

Proof. It follows immediately from the statements preceding the theorem that

$$\int_{A_k(O_{k-1}, \tau_k)} (\quad) P(dx_k|^S_N) = \int_{A_k(P_{k-1}, \tau_k)} (\quad) P(d\ell_k|^S_N) \qquad \text{(K. 22)}$$

where $P(d\ell_k|^S_N)$ is the distribution induced from $P(dx_k|^S_N)$ by the mapping in Eq. 5.9. Now, write the argument of the exponent in Eq. 5.9 as

$$z_k(x_k) = \frac{1}{N_o} \int_{t_k - T_p/2}^{t_k} x(t)\, p^k(t)\, dt - \frac{1}{2N_o} \int_{t_k - T_p/2}^{t_k} [p^k(t)]^2\, dt$$

In the case of additive white Gaussian noise, it is well known that z_k is distributed as a normal random variable with parameters

$$E[z_k|^S_N] = \pm d^2/2 \qquad \text{(K. 23a)}$$

and

$$\mathrm{Var}[z_k|^S_N] = d^2 \qquad \text{(K. 23b)}$$

where

$$d^2 = \frac{1}{N_o} \int_{t_k - T_p/2}^{t_k} [p^k(t)]^2 \, dt$$

The next step is to relate d^2 to the signal-to-noise ratio $d_T^{\,2}$ of Eq. K.21. First, it is noted from the fact that

$$p^k(t) = p(t - (t_k - T_p))$$

that $d_T^{\,2}$ can be written as

$$d_T^{\,2} = \frac{1}{N_o} \int_{t_k - T_p}^{t_k} [p^k(t)]^2 \, dt$$

or

$$d_T^{\,2} = \frac{1}{N_o} \int_{t_k - T_p}^{t_k - T_p/2} [p^k(t)]^2 \, dt + \frac{1}{N_o} \int_{t_k - T_p/2}^{t_k} [p^k(t)]^2 \, dt$$

Now, apply the pulse translation property of Eq. 5.8 to the first integral in the above expression to conclude that

$$d_T^{\,2} = \frac{1}{N_o} \int_{t_k - T_p/2}^{t_k} [p^k(t)]^2 \, dt + \frac{1}{N_o} \int_{t_k - T_p/2}^{t_k} [p^k(t)]^2 \, dt$$

$$= 2 \, \frac{1}{N_o} \int_{t_k - T_p/2}^{t_k} [p^k(t)]^2 \, dt$$

or

$$d_T^{\,2} = 2 \, d^2 \qquad \text{(K. 24)}$$

Finally, substitute for d^2 from Eq. K.24 into Eq. K.23a, b to get Eq. K.20.

The proof is then completed by noting that, since z_k has a random variable, ℓ_k must have a density function. Thus, we may replace $P(d\ell_k | \frac{S}{N})$ in Eq. K.22 by $f(\ell_k | \frac{S}{N}) \, d\ell_k$ to get Eq. K.18. This completes the proof.

In the next appendix, we describe the computer algorithm used to solve the recursive equations developed here.

APPENDIX L

THE COMPUTER ALGORITHM FOR OBTAINING THE m = 2 BAYES DECISION DEVICE

In this appendix the computer algorithm used to obtain the description of the $m = 2$ Bayes decision device is discussed.

From Section 5.3 and Appendix K it is seen that it is necessary to solve for the response sets

$$\tilde{A}_k(P_{k-1}, \tau_k) = \{\ell_k ; \tilde{T}_k(P_{k-1}, \ell_k, \tau_k) \geq 0\}$$

where

$$a\,\tilde{T}_k(P_{k-1}, x_k, \tau_k) \tag{L.1}$$

$$= \begin{cases} \ell_k \Big[(1-a) + [1 - (1+W_X)\, a_{k-1}]\, P_{k-1}\Big] - W_F a \\ \quad + (\ell_k P_{k-1} + a) \Big[\Big(F_k^{\,0}(P_k, a_k{=}0) - F_k^{\,0}(P_k, a_k{=}1)\Big) \\ \quad + P_k \Big(F_k^{\,1}(P_k, a_k{=}0) - F_k^{\,1}(P_k, a_k{=}1)\Big)\Big] & k < N \\ \ell_N \Big[(1-a) + [1 - (1+W_X)\, a_{N-1}]\, P_{N-1}\Big] & k = N \end{cases}$$

and where

$$F_k^{\,0}(P_k, a_k) \tag{L. 2a}$$

$$= \sum_{t=0}^{1} \int_{t\tilde{A}_{k+1}(P_k, (T_p/2)\, a_k)} R_{k+1}(P_{k+1}, (T_p/2)\, a_k, t, 0)\, f(\ell_{k+1} \mid S)\, d\ell_{k+1}$$

$$F_k^{\,1}(P_k, a_k) \tag{L. 2b}$$

$$= \sum_{t=0}^{1} \int_{t\tilde{A}_{k+1}(P_k, T_p/2\, a_k)} \{(1-a)\, R_{k+1}(P_{k+1}, (T_p/2)\, a_k, t, T_p/2)\, f(\ell_{k+1} \mid S)$$

$$+ \; a\, R_{k+1}(P_{k+1}, (T_p/2)\, a_k, t, T_p)\, f(\ell_{k+1} \mid N)\}\, d\ell_{k+1}$$

and

$$R_{k+1}(P_{k+1}, \tau_{k+1}, a_{k+1}, \theta_{k+1}) \tag{L. 3}$$

$$= \begin{cases} L(\tau_{k+1}, a_{k+1}, \theta_{k+1}) + F_{k+1}^{\,1}(P_{k+1}, a_{k+1}) & \theta_{k+1} = 0, T_p \\[2ex] L(\tau_{k+1}, a_{k+1}, \theta_{k+1}) + F_{k+1}^{\,1}(P_{k+1}, a_{k+1}) & \theta_{k+1} = T_p/2 \end{cases}$$

These equations are to be solved recursively, beginning with the index $k = n$ and proceeding through decreasing values of k . This procedure is indicated in Fig. L.1.

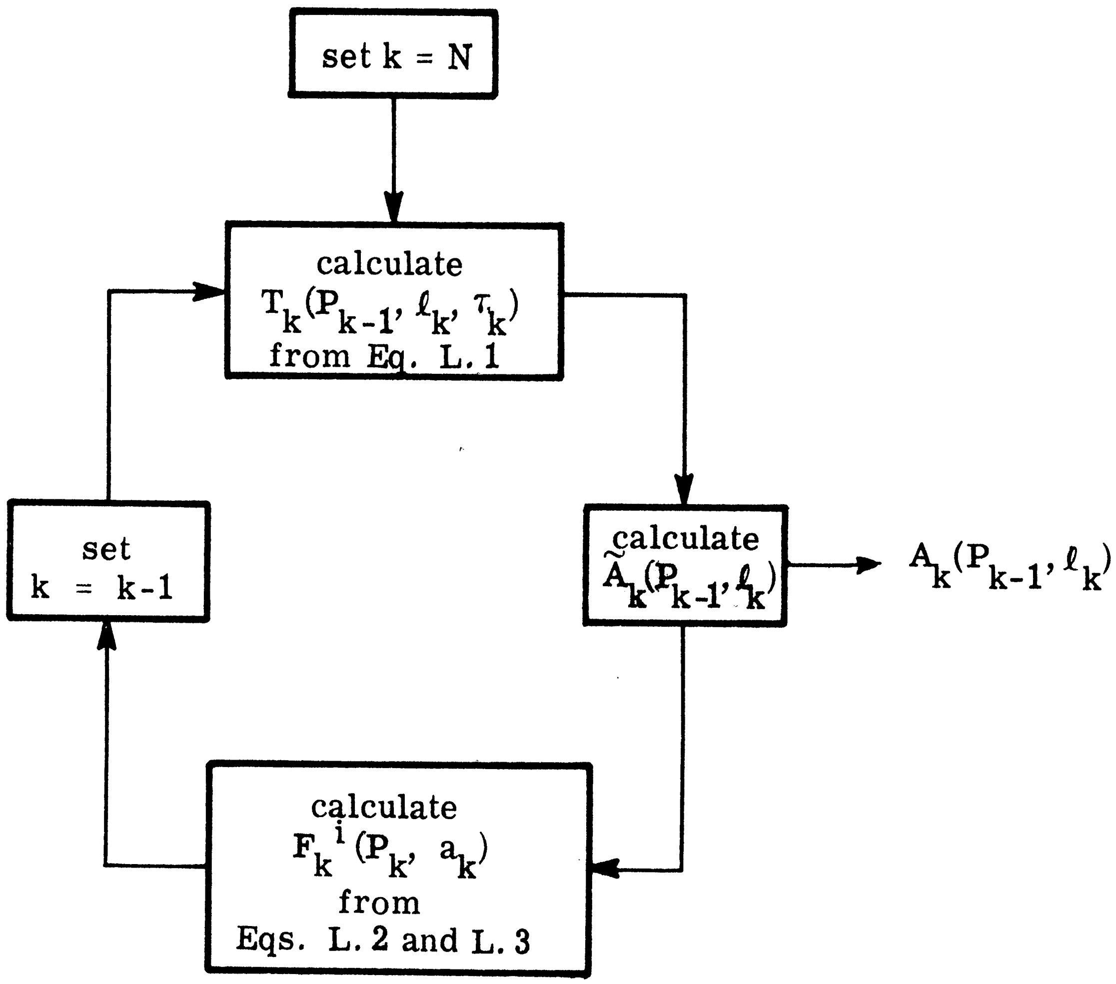


Fig. L.1. The recursive procedure

To carry out the computational procedure indicated above, it was first established that, for each value of $P_{k-1} > 0$ and $\tau_k = (T_p/2)\, a_{k-1}$, there exists at most one root of the function $\tilde{T}_k(P_{k-1}, \ell_k, \tau_k)$. Moreover, this root, to be denoted by $K_k^{a_{k-1}}(P_{k-1})$, has the property that

$$\tilde{T}_k(P_{k-1}, \ell_k, \tau_k) \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} 0$$

for

$$\ell_k \begin{Bmatrix} \geq \\ \leq \end{Bmatrix} K_k^{a_{k-1}}(P_{k-1})$$

This fact was established by carrying out the computational procedure a stage at a time and checking the function $\tilde{T}_k(P_{k-1}, \ell_k, \tau_k)$ at each stage.

As a consequence of the above fact, the response set $A_k(P_{k-1}, \tau_k)$ can be written in the form

$$A_k(P_{k-1}, \tau_k = \{ \ell_k \,;\, \ell_k \geq K_k^{a_{k-1}}(P_{k-1}) \} \qquad \text{(L. 4)}$$

Thus the likelihood ratio property of Section 4. 3 is established.

The function $K_k^{a_{k-1}}(P_{k-1})$ was obtained at each stage of the computation by a subroutine that calculates the root of the function $\tilde{T}_k(P_{k-1}, \ell_k, \tau_k)$. This subroutine operates by first determining if a root lies in the interval $[0, K_u]$, where K_u is equal to the five-standard deviation point for the random variable z_k under the condition "signal present". If no root is present in the interval $[0, K_u]$, $K_k^{a_{k-1}}(P_{k-1})$ is set equal to K_u. If a root is present, then the subroutine proceeds by a standard interval-bisection algorithm to locate that root to within 10^{-6} of its true value.

The procedure for calculating $F_k^i(P_k, a_k)$ is as follows. First, the values of $F_{k+1}^i(P_k, a_k)$, which have been stored from the previous stage, are used to determine $R_{k+1}(\cdot, \cdot, \cdot)$ according to Eq. L.3. These values are calculated for $\tau_{k+1} \epsilon \{0, T_p/2\}$, $a_{k+1} \epsilon \{0, 1\}$, $\theta_{k+1} \epsilon \{0, T_p\}$ and for $P_{k-1} = (.05)\, j$; $j=1, \ldots, 100$. Then the values of $R_{k+1}(\cdot, \cdot, \cdot, \cdot)$ and the threshold function $K^{a_{k-1}} P_{k-1})$ from the preceding stage are used to calculate $F_k^i(P_k, a_k)$ according to Eq. L.2a and L.2b. The integrals in these equations are evaluated using a seven-point Gaussian quadrature routine and a seven-point Gaussian-Hermite routine. The values $R_k(\cdot, \cdot, \cdot, \cdot)$ in the integrand of these integrals are obtained by a three-point Lagrange interpolation procedure from the values of $R_{k+1}(\cdot, \cdot, \cdot, \cdot)$ at the points $P_{k+1} = (.05)\, i$, $i = 1, \ldots, 100$. Finally, the values of $F_k^i(P_k, a_k)$ are stored for use at the next stage.

To complete the computations in one stage of the recursive procedure of Fig. L.1, it is necessary to calculate $\tilde{T}_k(P_{k-1}, \ell_k, \tau_k)$ according to Eq. L.1. Actually, this computation is carried in the root-finding routine using the stored values of $F_k^i(P_k, a_k)$.

In conclusion, we point out that it was only necessary to carry out the computations for six stages in order to obtain a reasonable approximation for the limiting function $\hat{K}^{a_{k-1}}(P_{k-1})$. This fact was established by calculating the difference

$$\hat{K}_{k+1}{}^i(u) - \hat{K}_k{}^i(u) \quad ; \quad i=0,1$$

and terminating the calculations when this difference became less than 10^{-2}.

APPENDIX M

THE SIMULATION PROCEDURE

In this appendix we describe the procedure used to obtain the ROC surfaces discussed in Section 5.5.

We begin by recalling the basic definitions of the normalized rates R_X , R_F and R_D . Specifically

$$R_X = r_X / r_p$$

$$R_F = r_F / (r_N / m)$$

$$R_D = r_D / r_p$$

where

$$r_\beta = \frac{E\{N_\beta(\delta(x), \theta)\}}{T}$$

Now, use the equalities

$$r_p = \frac{\Pr[\theta_i \in [0, T_p)]}{T_p} = \frac{D_p}{T_p}$$

$$r_N / m = \frac{1 - \Pr[\theta \in [0, T_p)]}{T_p} = \frac{1 - D_p}{T_p}$$

$$T_p = m\Delta \; ; \; T = N\Delta$$

to write

$$R_X = \left[\frac{E\{N_X(\delta(x), \theta)\}}{N} \right] \frac{m}{D_p}$$

$$R_F = \left[\frac{E\{N_F(\delta(x), \theta)\}}{N}\right] \frac{m}{1-D_p}$$

$$R_D = \left[\frac{E\{N_D(\delta(x), \theta)\}}{N}\right] \frac{m}{D_p}$$

Now the point here is that, for a fixed number of decision opportunities per pulse, m, and a fixed pulse duty, D_p,

$$R_\beta(\delta) \ \alpha \ E\left\{\frac{N_\beta(\delta(x), \theta)}{N}\right\}$$

Thus to determine the rate $R_\beta(\delta)$ for a specific decision device δ, it is only necessary to determine

$$E\left\{\frac{N_\beta(\delta(x), \theta)}{N}\right\}$$

To do this, we recall that $N_\beta(\delta(x), \theta)$ represents the total number of occurrences of the event β after N decision times when the decision device δ is used and the observation is $x = (x_1, \ldots, x_N)$ and the signal is described by $\theta = (\theta_1, \ldots, \theta_N)$. Thus, we may interpret the quantity

$$\frac{N_\beta(\delta(x), \theta)}{N}$$

as the relative frequency of the event β in N decision times. Now, if N is large and x^0 and θ^0 are typical realizations of the observation sequence x and state sequence θ, then it is reasonable to expect that this relative frequency approaches some limiting value

$$\frac{\hat{N}_\beta(\delta(x^0),\ \theta^0)}{N}$$

and that

$$E\left\{\frac{N_\beta(\delta(x),\ \theta)}{N}\right\} = \frac{\hat{N}_\beta(\delta(x^0),\ \theta^0)}{N}$$

Indeed, this is the underlying philosophy of the law of large numbers. Unfortunately, it appears quite difficult to establish this result theoretically due to the complex nature of the random variable $N_\beta(\delta(x), \theta)$. We have, however, tested the validity of this result experimentally by calculating $N_\beta(\delta(x^0), \theta^0)$ for different values of x^0 and θ^0 and noting that for $N \geq 8000$ the quantity

$$\frac{N_\beta(\delta(x^0),\ \theta^0)}{N}$$

remains constant out to the fourth decimal place.

We turn now to the question of how $N_\beta(\delta(x^0),\ \theta^0)/N$ is calculated. First, note that all the decision devices of interest in this chapter can be represented as shown in Fig. M.1. In this figure,

$$\ell_k(x_k) = \ell\ [z_k(x_k)]$$

where z_k ; k=1, 2, . . . , are Gaussian random variables with a

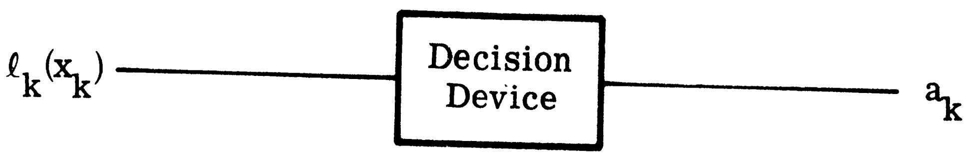


Fig. M.1. The generic form for the decision devices of Chapter 5 mean and variance that depend on the signal content of the interval $[t_k - T_p/2, t_k)$. More precisely, if we note that signal is present in the current interval iff $\theta_k = 0, T_p/2$ and signal is not present in the current interval iff $\theta_k = T_p$, then we may apply Theorem I-3 of Appendix I to conclude that

$$E[z_k \mid S] = E[z_k \mid \theta_k = 0, T_p/2] = d_T^2/4$$

$$E[z_k \mid N] = E[z_k \mid \theta_k = T_p] = -d_T^2/4$$

$$\mathrm{Var}[z_k \mid {S \atop N}] = d_T^2/2$$

Now the point here is that if we have a realization θ^0 of the sequence $\theta = (\theta_1, \ldots, \theta_N)$ then we can calculate a realization ℓ^0 of the sequence $\ell = (\ell_1, \ell_2, \ldots, \ell_N)$ by using a Gaussian random number generator to calculate z_k with mean and variance determined by θ_k . We can then use a particular decision device to determine the sequence of decisions $(a_1, a_2, \ldots, a_N)$. Finally, the number of occurrences of the event β can be determined according to the table in Fig. M. 2.

a_{k-1} a_k	$\theta_k = 0$	$\theta_k = T_p/2$	$\theta_k = T_p$
0 0	miss	---	---
0 1	detection	detection	false alarm
1 0	---	---	---
1 1	extra detection	detection	false alarm

Fig. M. 2. The relation of the events β to the decisions a_{k-1} and a_k and the state θ_k

It remains to describe how the realization of the sequence $\theta = (\theta_1, \ldots, \theta_N)$ is calculated. To this end, we note from the transition matrix of Eq. 5.2 that θ is described by a state diagram as shown in Fig. M.3.

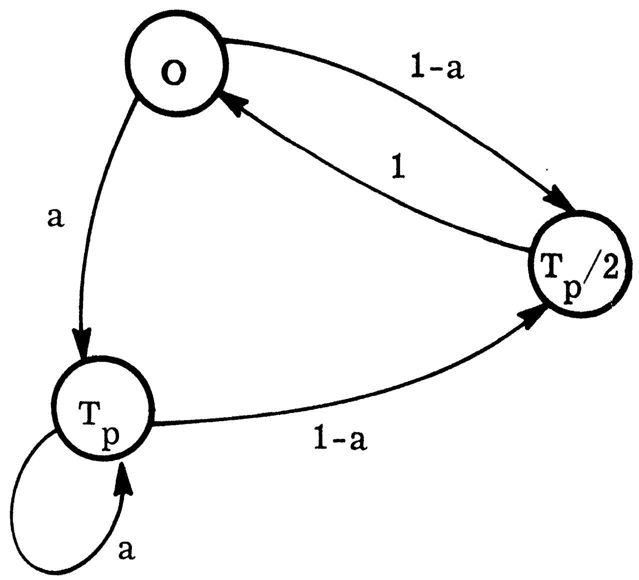


Fig. M.3. State transition diagram for the random variables θ_k

The interpretation of this diagram is as follows. If $\theta_k = T_p/2j$, $j = 0, 1, 2,$ then θ_{k+1} takes on the value indicated by the arrow emanating from the appropriate node with the probability shown next to that arrow. For example, if

$$\theta_k = 0$$

then

$$\theta_{k+1} = \begin{cases} T_p/2 & \text{with probability } 1 - a \\ T_p & \text{with probability } a \end{cases}$$

Thus, if we have a sequence of random variables, $b_1, b_2, \ldots, b_N$, such that

$$Pr[b_k = j] = \begin{cases} 1 - a & j = 0 \\ a & j = 1 \end{cases}$$

then we can determine $\theta = \theta_1, \ldots, \theta_N$ recursively by

$$\theta_{k+1} = \begin{cases} \begin{cases} T_p/2 & b_k = 0 \\ T_p & b_k = 1 \end{cases} & \theta_k = 0 \\ \begin{cases} 0 \end{cases} & \theta_k = T_p/2 \\ \begin{cases} T_p/2 & b_k = 0 \\ T_p & b_k = 1 \end{cases} & \theta_k = T_p \end{cases}$$

The sequence, $b_1, \ldots, b_m$, can be obtained from a uniform random number generator.

APPENDIX N

THE RATES R_X, R_F AND R_D FOR THE SKE DECISION DEVICE

In this appendix we relate the probabilities of detection and false alarm, P_D and P_F, for the SKE classical detection decision device to R_D, R_F and R_X and D_D and D_F.

The basic idea here is to view the SKE device as making successive independent decisions over contiguous intervals of time. More precisely, we divide the time axis into N equal intervals of length T_p and assume that the decision device makes decisions at the end of each interval. We may then consider the total decision of this device as the vector $\vec{a} = (a_1, \ldots, a_N)$, $a_i \epsilon \{0, 1\}$, where a_i represents the decision at the end of the ith interval. In addition, one may describe the total signal by the vector $\theta = (\theta_1, \ldots, \theta_N)$, $\theta_i \epsilon\, \Theta_S \cup \Theta_N$, where Θ_S, Θ_N represent the signal parameters and the noise parameters which affect the observation over the ith interval. (See Appendix A.) It is to be noted here that we have assumed that the noise is independent from interval to interval and that any signal associated with the parameter θ_i is completely contained within that interval.

We now appeal to the basic definitions in Section 2.5 for an extra detection, a false alarm, and a detected pulse to write

$$N_X = 0$$

$$N_D = \sum_{i=1}^{N} I(a_i, \theta_i) \quad \{a_i = 1, \Theta_S\}$$

and

$$N_F = \sum_{i=1}^{N} I(a_i, \theta_i) \quad \{a_i = 1, \Theta_N\}$$

Then, from the definitions of r_X, r_D and r_F in Section 3.2,

$$r_X = 0$$

$$r_D = \sum_{i=1}^{N} \frac{Pr[a_i = 1, \Theta_S]}{N\,T_p}$$

and

$$r_F = \sum_{i=1}^{N} \frac{Pr[a_i = 1, \Theta_N]}{N\,T_p}$$

Next, since the probabilities in the above sums are identical,

$$r_X = 0$$

$$r_D = \frac{Pr[a_i = 1, \Theta_S]}{T_p} = Pr[a_i = 1 \mid \Theta_S] \frac{Pr[\Theta_S]}{T_p}$$

and

$$r_F = \frac{Pr[a_i = 1, \Theta_N]}{T_p} = Pr[a_i = 1 \mid \Theta_N] \frac{Pr[\Theta_N]}{T_p}$$

or finally

$$R_X = 0$$

$$R_D = r_D / \Pr[\Theta_S] / T_p = \Pr[a_i = 1 \mid \Theta_S] = P_D$$

and

$$R_F = r_F / \Pr[\Theta_N] / T_p = \Pr[a_i = 1 \mid \Theta_N] = P_F$$

To relate P_D and P_F to D_D and D_F we simply appeal to Eq. 3.13 with $m = 1$. The result is

$$D_D = P_D$$

and

$$D_F = P_F$$

REFERENCES

1. Egan, J. P., Greenberg, G. Z., and Schulman, A. I. "Operating Characteristics, Signal Detectability, and the Method of Free Response." JASA, 1961, Vol. 33, pp. 993-1007.

2. Lucas, P. A. "Human Performance in Low-Signal-Probability Tasks." JASA, 1967, Vol. 42, No. 1, pp. 158-178.

3. Luce, R. D. "A Model for Detection in Temporally Unstructured Experiments with a Poisson Distribution of Signal Presentations." Journal of Mathematical Psychology, 1966, Vol. 3, pp. 48-64.

4. Ferguson, T. S. Mathematical Statistics: A Decision Theoretic Approach. New York: Academic Press, 1967.

5. Feller, W. An Introduction to Probability Theory and Its Applications. Volume 2. New York: John Wiley and Sons, Inc., 1966.

6. Lehmann, E. L. Testing Statistical Hypotheses. New York: John Wiley and Sons, Inc., 1959.

7. Breiman, L. Probability. Reading, Massachusetts: Addison-Wesley Publishing Company, 1968.

8. Birdsall, T. G. "The Theory of Signal Detectability: ROC Curves and Their Character." Doctoral dissertation. Ann Arbor: The University of Michigan, 1966.

9. Hilborn, Jr., C. G. and Lainiotis, D. G. "Recursive Computations for the Optimal Tracking of Time-Varying Parameters." IEEE Transactions on Information Theory, Vol. IT-14, No. 3, pp. 514-515.

10. Van Trees, H. L. Detection, Estimation, and Modulation Theory, Part I. New York: John Wiley and Sons, Inc. 1968.

DISTRIBUTION LIST

	No. of Copies
Office of Naval Research (Code 468)	1
(Code 102-OS)	1
(Code 480)	1
Navy Department Washington, D. C. 20360	
Director, Naval Research Laboratory Technical Information Division Washington, D. C. 20390	6
Director Office of Naval Research Branch Office 1030 East Green Street Pasadena, California 91101	1
Office of Naval Research San Francisco Area Office 1076 Mission Street San Francisco, California 94103	1
Director Office of Naval Research Bramch Office 495 Summer Street Boston, Massachusetts 02210	1
Office of Naval Research New York Area Office 207 West 24th Street New York, New York 10011	1
Director Office of Naval Research Branch Office 219 South Dearborn Street Chicago, Illinois 60604	1
Commanding Officer Office of Naval Research Branch Office Box 38 FPO New York 09510	8

DISTRIBUTION LIST (Cont.)

	No. of Copies
Commander Naval Ordnance Laboratory Acoustics Division White Oak, Silver Spring, Maryland 20907 Attn: Dr. Zaka Slawsky	1
Commanding Officer Naval Ship Research & Development Center Annapolis, Maryland 21401	1
Commander Naval Undersea Research & Development Center San Diego, California 92132 Attn: Dr. Dan Andrews Mr. Henry Aurand	2
Chief Scientist Navy Underwater Sound Reference Division P. O. Box 8337 Orlando, Florida 32800	1
Commanding Officer and Director Navy Underwater Systems Center Fort Trumbull New London, Connecticut 06321	1
Commander Naval Air Development Center Johnsville, Warminster, Pennsylvania 18974	1
Commanding Officer and Director Naval Ship Research and Development Center Washington, D. C. 20007	1
Superintendent Naval Postgraduate School Monterey, California 93940	1
Commanding Officer & Director Naval Ship Research & Development Center* Panama City, Florida 32402	1

*Formerly Mine Defense Lab.

DISTRIBUTION LIST (Cont.)

	No. of Copies
Naval Underwater Weapons Research & Engineering Station Newport, Rhode Island 02840	1
Superintendent Naval Academy Annapolis, Maryland 21401	1
Scientific and Technical Information Center 4301 Suitland Road Washington, D. C. 20390 Attn: Dr. T. Williams Mr. E. Bissett	2
Commander Naval Ordnance Systems Command Code ORD-03C Navy Department Washington, D. C. 20360	1
Commander Naval Ship Systems Command Code SHIPS 037 Navy Department Washington, D. C. 20360	1
Commander Naval Ship Systems Command Code SHIPS 00V1 Washington, D. C. 20360 Attn: CDR Bruce Gilchrist Mr. Carey D. Smith	2
Commander Naval Undersea Research & Development Center 3202 E. Foothill Boulevard Pasadena, California 91107	1
Commanding Officer Fleet Numerical Weather Facility Monterey, California 93940	1

DISTRIBUTION LIST (Cont.)

	No. of Copies
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
Dr. James Probus Office of the Assistant Secretary of the Navy (R&D) Room 4E741, The Pentagon Washington, D. C. 20350	1
Mr. Allan D. Simon Office of the Secretary of Defense DDR&E Room 3E1040, The Pentagon Washington, D. C. 20301	1
CAPT J. Kelly Naval Electronics Systems Command Code EPO-3 Washington, D. C. 20360	1
Chief of Naval Operations Room 5B718, The Pentagon Washington, D. C. 20350 Attn: Mr. Benjamin Rosenberg	1
Chief of Naval Operations Rm 4C559, The Pentagon Washington, D. C. 20350 Attn: DCR J. M. Van Metre	1
Chief of Naval Operations 801 No. Randolph St. Arlington, Virginia 22203	1
Dr. Melvin J. Jacobson Rensselaer Polytechnic Institute Troy, New York 12181	1
Dr. Charles Stutt General Electric Co. P. O. Box 1088 Schenectady, New York 12301	1

DISTRIBUTION LIST (Cont.)

	No. of Copies
Dr. Alan Winder EDO Corporation College Point, New York 11356	1
Dr. T. G. Birdsall Cooley Electronics Lab. University of Michigan Ann Arbor, Michigan 48105	1
Dr. John Steinberg University of Miami Institute of Marine & Atmospheric Sciences Miami, Florida 33149	1
Mr. Robert Cunningham Bendix Corporation 11600 Sherman Way North Hollywood, California 91606	1
Dr. H. S. Hayre University of Houston Cullen Boulevard Houston, Texas 77004	1
Dr. Robert R. Brockhurst Woods Hole Oceanographic Institute Woods Hole, Massachusetts 02543	1
Dr. Stephen Wolff Johns Hopkins University Baltimore, Maryland 21218	1
Dr. M. A. Basin Litton Industries 8000 Woodley Avenue Van Nuys, California 91409	1
Dr. Albert Nuttall Navy Underwater Systems Center Fort Trumbull New London, Connecticut 06320	1

DISTRIBUTION LIST (Cont.)

	No. of Copies
Dr. Philip Stocklin Raytheon Company P. O. Box 360 Newport, Rhode Island 02841	1
Dr. H. W. Marsh Navy Underwater Systems Center Fort Trumbull New London, Connecticut 06320	1
Dr. David Middleton 35 Concord Ave., Apt. #1 Cambridge, Massachusetts 02138	1
Mr. Richard Vesper Perkin-Elmer Corporation Electro-Optical Division Norwald, Connecticut 06852	1
Dr. Donald W. Tufts University of Rhode Island Kingston, Rhode Island 02881	1
Dr. Loren W. Nolte Dept. of Electrical Engineering Duke University Durham, North Carolina 27706	1
Mr. S. W. Autrey Hughes Aircraft Co. P. O. Box 3310 Fullerton, California 92634	1
Dr. Thomas W. Ellis Texas Instruments, Inc. 13500 North Central Expressway Dallas, Texas 75231	1
Mr. Robert Swarts Honeywell, Inc. Marine Systems Center 5303 Shilshole Ave., N.W. Seattle, Washington 98107	1

DISTRIBUTION LIST (Cont.)

	No. of Copies
Mr. Gordon R. Hamilton Palisades Geophysical Sofar Station APO New York 09856	1
Mr. Charles Loda Institute for Defense Analyses 400 Army-Navy Drive Arlington, Virginia 22202	1
Mr. Beaumont Buck General Motors Corporation Defense Research Division 6767 Holister Ave. Goleta, California 93017	1
Dr. M. Weinstein Underwater Systems, Inc. 8121 Georgia Avenue Silver Spring, Maryland 20910	1
Dr. Harold Saxton 1601 Research Blvd. TRACOR, Inc. Rockville, Maryland 20850	1
Dr. Thomas G. Kincaid General Electric Company P. O. Box 1088 Schenectady, New York	1
Applied Research Laboratories The University of Texas at Austin Austin, Texas 78712 Attn: Dr. Loyd Hampton Dr. Charles Wood	3
Dr. Paul McElroy Woods Hole Oceanographic Institution Woods Hole, Massachusetts 02543	1

DISTRIBUTION LIST (Cont.)

	No. of Copies
Dr. John Bouyoucos General Dynamics/ Electronics 1400 N. Goodman Street P. O. Box 226 Rochester, New York 14603	1
Hydrospace Research Corporation 5541 Nicholson Lane Rockville, Maryland 20852 Attn: CDR Craig Olson	1
Cooley Electronics Laboratory University of Michigan Ann Arbor, Michigan 48105	25

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Cooley Electronics Laboratory University of Michigan Ann Arbor, Michigan 48105	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

THE FREE RUNNING DETECTION PROBLEM: THE DETECTION OF RANDOMLY OCCURRING PULSES AS THEY OCCUR

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report No. 209 - 036040-5-T

5. AUTHOR(S) (First name, middle initial, last name)

Richard M. Heitmeyer

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1971	410	10

8a. CONTRACT OR GRANT NO. N00014-67-A-0181-0032	9a. ORIGINATOR'S REPORT NUMBER(S) 036040-5-T
b. PROJECT NO. c. d.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) TR 209

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research Department of the Navy Arlington, Va. 22217

13. ABSTRACT

This study provides an extension of classical detection theory by defining and analyzing two classes of decision devices that both operate in real time to detect the presence of each pulse in a sequence of randomly occurring pulses while that pulse is still present. The respond-once device seeks to detect each pulse once and only once whereas the other device, the respond-and-hold device, seeks to respond at each instant of time at which a pulse is present. The problem of obtaining the decision devices that are optimum in the above setting is the free running detection problem. It is assumed that all pulses are identically shaped and known up to their arrival times and that it is possible to observe these pulses only through an observation that contains additive noise. It is also assumed that the decision devices are allowed to make decisions on the presence or absence of pulses only at discrete, equally spaced points in time. The mathematical model is constructed in terms of the general decision theory model providing the theoretical basis for describing the performance of both classes of decision devices and for obtaining decision devices that are optimum with respect to the Bayes criteria. The performance of the respond-once decision device is described in terms of the average number of pulses detected, the average number of false alarms, and the average number of times a pulse is detected more than once. It is shown that the detection rate can be increased for the same false alarm rate by increasing the extra detection rate. The performance of the respond-and-hold decision devices is described in terms of the detection rate, duty, and false alarm duty.

DD FORM 1473 1 NOV 65 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

A-31408

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Random pulses						
Optimum detection						
Free response						