An application of Bayesian statistical methods to adaptive radiotherapy

Kwok L Lam[1]**, Randall K Ten Haken**[1]**, Dale Litzenberg**[1]**,
James M Balter**[1] **and Stephen M Pollock**[2]

[1] Department of Radiation Oncology, The University of Michigan, Ann Arbor, MI 48109, USA
[2] Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor,
MI 48109, USA

Abstract
In adaptive radiotherapy, measured patient-specific setup variations are used to
modify the patient setup and treatment plan, potentially many times during the
treatment course. To estimate the setup adjustments and re-plan the treatment,
the measured data are usually processed using Kalman filtering or by computing
running averages. We propose, as an alternative, the use of Bayesian statistical
methods, which combine a population (prior) distribution of systematic and
random setup errors with the measurements to determine a patient-specific
(posterior) probability distribution. The posterior distribution can either be
used directly in the re-planning of the treatment or in the generation of statistics
needed for adjustments. Based on the assumption that day-to-day setup
variations are independent and identically distributed Normal distributions, we
can efficiently compute parameters of the posterior distribution from parameters
of the prior distribution and statistics of the measurements. We illustrate a
simple procedure to apply the method in practice to adaptive radiotherapy,
allowing for multiple adjustments of treatment parameters during the course of
treatment.

1. Introduction

Bayesian methods have been widely used in statistical inference. Applications range from
simple parameter estimation to the regularization of ill-posed inverse problems. Recently, the
use of Bayesian methods has also received attention in other settings such as the estimation of
uncertainties in dose distribution in inverse treatment planning (Unkelbach and Oelfke 2005).
Here we apply established Bayesian methods to sequential refinement of knowledge on a
specific patient as repeated measurements of the states of the patient are performed during the
course of treatment, often known as adaptive radiotherapy (Yan *et al* 1997). Since multiple
refinements may be performed during the course of a treatment, it is important that there is

a convenient procedure to repeatedly apply the method, in contrast to some applications that need only a one-time estimation.

During radiotherapy treatment planning, the distribution of error over a general patient population is often used to determine setup and treatment margins for an individual patient. In *adaptive* radiotherapy, however, individual patient characteristics are measured during treatment and used to refine the therapy by correcting for systematic setup error and creating individualized margins of treatment fields. In particular, fraction-to-fraction variations in setup for an individual patient can generate decisions for future treatments—for example, by using Kalman filtered data (Yan *et al* 1997) or sample statistics (Birkner *et al* 2003). Our group has been exploring the possibility of using the Bayesian method in estimating the patient-specific setup parameters (Ten Haken *et al* 2000, 2001, Lam *et al* 2004). We present here an approach to adaptive radiotherapy planning that efficiently and effectively combines a *sequence* of observed individual patient setup variations with data from the general population.

For illustrative purposes, we limit an individual patient's fraction-to-fraction setup variation to be a scalar X, representing the inferior–superior direction. (Extension to the realistic case of three-dimensional variation $[X, Y, Z]$ can be found in the appendix.) We first make the usual assumption that X has a Gaussian (Normal) distribution $f_X(x)$ with 'systematic error' given by the mean m and 'random' setup error given by the standard deviation s. That is, X has the probability density function (p.d.f.)

$$f_X(x; m, s) = \left(\frac{1}{2\pi s^2}\right)^{\frac{1}{2}} e^{-\frac{1}{2s^2}(x-m)^2}. \tag{1}$$

(Here, and in what follows, a capital letter denotes a random variable and the equivalent lower case letter its associated realization.)

A sequence of k independent and identically distributed setup variations over k fractions is the random k-vector $[X_1 \ X_2 \ \dots \ X_k]$, where X_i is the variation on the ith fraction. $[X_1 \ X_2 \ \dots \ X_k]$ therefore will have the joint p.d.f.

$$f_{X_1 \ X_2 \ \dots \ X_k}(x_1 \ x_2 \ \dots \ x_k; m, s) = \left(\frac{1}{2\pi s^2}\right)^{\frac{k}{2}} \exp\left(-\frac{1}{2s^2}\sum_{i=1}^{k}(x_i - m)^2\right), \qquad k = 1, 2, \dots, \tag{2}$$

where $x_i =$ the observed variation on the ith fraction.

At the core of our Bayesian analysis are three assumptions:

(a) there is a population of potential patients over which there is a distribution of values of the mean m and standard deviation s. That is, the mean and standard deviation are the *random variable* pair (M, S);

(b) a joint p.d.f. $f_{M,S}^{(0)}(m, s)$ (referred to as the 'prior' distribution or the 'population' distribution) on the pair (M, S) can be determined from existing patient population statistics;

(c) without information to the contrary, each patient can be considered to be randomly selected from this population.

2. Theory

2.1. Using Bayes theorem

Since an individual patient is randomly selected from the population, before taking any setup measurements this patient's variation has a Normal distribution with parameters (M, S) having p.d.f. $f_{M,S}^{(0)}(m, s)$.

Bayes theorem (the fundamental rule of inference) can then be used to combine imaging data from *individual patient* measurements governed by equation (1) with data taken from the *population* statistics as incorporated in $f_{M,S}^{(0)}(m, s)$. The result is an 'updated' (or 'posterior') joint p.d.f. for the pair (M, S) that can then be used to make subsequent treatment decisions.

In particular, after observing k daily setup-variation values $x_1\ x_2\ \ldots\ x_k$, the probability distribution of (M, S), which can be written as $f_{M,S}^{(k)}(m, s|x_1\ x_2\ \ldots\ x_k)$, will be updated according to Bayes theorem:

$$f_{M,S}^{(k)}(m, s|x_1\ x_2\ \ldots\ x_k) = \frac{f_{X_1\ X_2\ \ldots\ X_k}(x_1\ x_2\ \ldots\ x_k; m, s) f_{M,S}^{(0)}(m, s)}{\int_s \int_m f_{X_1\ X_2\ \ldots\ X_k}(x_1\ x_2\ \ldots\ x_k; m, s) f_{M,S}^{(0)}(m, s)\, \mathrm{d}m\, \mathrm{d}s}. \tag{3}$$

It is also possible to iteratively compute the updated distribution using

$$f_{M,S}^{(k)}(m, s|x_1\ x_2\ \ldots\ x_k) = \frac{f_{X_k}(x_k; m, s) f_{M,S}^{(k-1)}(m, s)}{\int_s \int_m f_{X_k}(x_k; m, s) f_{M,S}^{(k-1)}(m, s)\, \mathrm{d}m\, \mathrm{d}s}, \qquad k = 1, 2, 3, \ldots. \tag{4}$$

As a result, as data $x_1\ x_2\ \ldots\ x_k$ are collected, a sequence of posterior distributions $f_{M,S}^{(k)}(m, s|x_1\ x_2\ \ldots\ x_k)$ can be generated from the prior distribution $f_{M,S}^{(0)}(m, s)$. The prior distribution only has information about the population of patients. However, as an individual patient's data are collected, when the number of fractions k becomes large the posterior distribution $f_{M,S}^{(k)}(m, s|x_1\ x_2\ \ldots\ x_k)$ will tend to become peaked at $m = \theta_k$ (the sample mean) and $s = \bar{s}_k$ (the sample standard deviation), where

$$\theta_k \equiv \frac{1}{k} \sum_{i=1}^{k} x_i, \qquad \bar{s}_k \equiv \frac{1}{k} \sqrt{\sum_{i=1}^{k} (x_i - \theta_k)^2}.$$

Bayes theorem is the only formal way to gradually migrate from the population distribution of (M, S) to a distribution based on specific patient data. The generation of a posterior distribution from a prior distribution and observed data via equation (3) has been extensively studied and is widely used in statistical decision making.

2.2. Using a conjugate prior

Determining the population distribution $f_{M,S}^{(0)}(m, s)$ is a matter of collecting appropriate population data. However, there is both a computational and interpretive advantage in fitting $f_{M,S}^{(0)}(m, s)$ to a member of a family of distributions that is 'closed under sampling'. Also known as a 'conjugate family' of distributions, such a family has the property that the posterior distribution $f_{M,S}^{(k)}(m, s|x_1\ x_2\ \ldots\ x_k)$ will *also* be a member of the family.

To take advantage of this property and to make subsequent calculations tractable, we use an alternative form of the Normal distribution in which, instead of using the variance s^2, variability is expressed by a parameter called the 'precision' r which is defined to be $r \equiv 1/s^2$. If the variance is a random variable S^2, then the precision is also a random variable, given by $R \equiv 1/S^2$. Equation (1) can then be rewritten as

$$f_X(x; m, r) = N(x; m, r) \equiv \left(\frac{r}{2\pi}\right)^{\frac{1}{2}} \mathrm{e}^{-\frac{r}{2}(x-m)^2} \tag{5}$$

and equation (2) can be rewritten as

$$f_{X_1\ X_2\ \ldots\ X_k}(x_1\ x_2\ \ldots\ x_k; m, r) = \left(\frac{r}{2\pi}\right)^{\frac{k}{2}} \exp\left(-\frac{r}{2} \sum_{i=1}^{k} (x_i - m)^2\right), \qquad k = 1, 2, \ldots. \tag{6}$$

The conjugate family for (M, R) can be shown to be (see, for example, DeGroot 1970) the 'Normal–Gamma' distribution with parameters α, β, μ and τ:

$$f_{M,R}^{(0)}(m, r) = N\text{-}G(m, r; \alpha, \beta, \mu, \tau) \equiv N(m; \mu, \tau r)\gamma(r; \alpha, \beta), \tag{7}$$

where $N(\cdot; \cdot, \cdot)$ is the Normal probability distribution defined in equation (5) and

$$\gamma(t; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} \tag{8}$$

is the Gamma probability distribution, where $\Gamma(\alpha)$ is the standard gamma function (e.g., $\Gamma(i) = (i-1)!$ when i is a positive integer).

2.3. Interpreting the prior distribution

Since we intend to use the Normal–Gamma distribution of equation (7) to capture *a priori* knowledge of the population distribution of setup variability, it is helpful to examine and interpret its parameters in this context.

We first note that with this prior distribution it can be readily shown that the marginal distribution for M is a Student's t distribution (with 2α degrees of freedom, location parameter μ and shape parameter $\frac{\alpha\tau}{\beta}$) and the marginal distribution for R is $\gamma(r; \alpha, \beta)$. However, since the joint distribution given by equation (7) is not the product of these two marginals, M and R are not probabilistically independent. This is not a deficiency, however, in that (as noted by DeGroot 1970, p 170) 'even if the prior distribution of M and R specified that these variables were independent, their posterior distribution after the value of a single (individual patient's) observation had been noted would specify that they were dependent'.

The shapes of these marginal distributions (and of the joint distribution) are reasonable representations for patient populations, and the availability of four parameters allows a good deal of freedom obtaining a fit to the data. Moreover:

- The parameter μ, being the expected value of the systematic error M over all patients in the population, can reasonably be anticipated to equal 0 since there is no *a priori* reason to believe that this systematic error has a bias (in any dimension) high or low. In what follows, however, we allow μ to be general, since this does not create any analytical or computation difficulties.

- Given that the *random* error has some value for the standard deviation $S = s$ (and therefore a precision of $R = r = 1/s^2$), the parameter τ represents the precision with which the *systematic error* M is known, expressed as a value proportional to r. For example, suppose the population data support an assumption that the random error has a standard deviation of $s = 5$ mm, which implies $r = 1/5^2 = 0.04$ mm^{-2}. Further, suppose the data support using an expected value of systematic error M of $\mu = 0$ mm with an associated standard deviation of 8 mm, then the precision of M is $1/8^2 = 0.0156 = \tau r$ and so $\tau = 0.0156/0.04 = 0.39$.

In particular, it is straightforward to show that the prior (e.g., population) moments of M, S and R are

$$E[M] = \mu, \qquad \text{Var}[M] = \frac{\beta}{\tau(\alpha - 1)}, \tag{9}$$

$$E[R] = \frac{\alpha}{\beta}, \qquad \text{Var}[R] = \frac{\alpha^2}{\beta}, \tag{10}$$

$$E[S] = \frac{\Gamma\left(\alpha - \frac{1}{2}\right)\sqrt{\beta}}{\Gamma(\alpha)}, \qquad \text{Var}[S] = \frac{\beta}{\alpha - 1} - (E[S])^2, \tag{11}$$

where $\Gamma(y)$ is the standard gamma function.

2.4. Parameters of the posterior distribution

Since the prior distribution equation (7) is conjugate, the distribution after observing measurements **x** is also a Normal–Gamma distribution, with parameters α_k, β_k, μ_k and τ_k, where

$$\alpha_k = \alpha + k/2, \tag{12}$$

$$\beta_k = \beta + \frac{1}{2}\sum_{i=1}^{k}(x_i - \theta_k)^2 + \frac{k\tau(\theta_k - \mu)^2}{2(\tau + k)}, \tag{13}$$

$$\mu_k = \frac{\tau\mu + k\theta_k}{\tau + k}, \tag{14}$$

$$\tau_k = \tau + k. \tag{15}$$

Note that setting $k = 0$ (i.e., before any patient data are observed), these equations produce the prior parameters $\alpha = \alpha_0$, $\beta = \beta_0$, $\mu = \mu_0$ and $\tau = \tau_0$.

Defining the posterior moments of M, S and R, after observing **x**, to be $E[M_k]$, $E[S_k]$ and $E[R_k]$, equations (9)–(11) become

$$E[M_k] = \mu_k, \qquad \text{Var}[M_k] = \frac{\beta_k}{\tau_k(\alpha_k - 1)}, \tag{16}$$

$$E[R_k] = \frac{\alpha_k}{\beta_k}, \qquad \text{Var}[R_k] = \frac{\alpha_k^2}{\beta_k}, \tag{17}$$

$$E[S_k] = \frac{\Gamma\left(\alpha_k - \frac{1}{2}\right)\sqrt{\beta_k}}{\Gamma(\alpha_k)}, \qquad \text{Var}[S_k] = \frac{\beta_k}{\alpha_k - 1} - (E[S_k])^2. \tag{18}$$

We can see that the posterior distributions evolve smoothly from the (prior) population distribution to a patient-specific distribution. As $k \to \infty$, $E[M] \to \theta_k$ and $E[S] \to \bar{s}_k$, the sample mean and sample standard deviation, respectively. The effect of the population distribution is diminished as more data are accumulated.

It is also useful to compute the *unconditional* (sometimes called the 'predictive') distribution $f_{X_{k+1}}(x_{k+1})$ for patient variation X_{k+1} on the $(k + 1)$th fraction, given the joint distribution $f_{M,R}^{(k)}(m, r)$ computed on the basis of the previous k observations:

$$f_{X_{k+1}}(x_{k+1}) = \int_m \int_s f_{X_{k+1}}(x_{k+1}; m, r) f_{M,R}^{(k)}(m, r)\, \mathrm{d}m\, \mathrm{d}r$$

$$= \int_m \int_s f_{X_{k+1}}(x_{k+1}; m, r) N(m; \mu_k, \tau_k r)\gamma(r; \alpha_k, \beta_k)\, \mathrm{d}m\, \mathrm{d}r. \tag{19}$$

Even though $f_{M,R}^0(m, r)$ has a conjugate form, the predictive distribution $f_{X_{k+1}}(x_{k+1})$ does not have a convenient analytical representation, and must be computed numerically. However, it is possible to compute its moments, in particular,

$$E[X_{k+1}] = \mu_k, \tag{20}$$

$$\text{Var}[X_{k+1}] = \frac{\beta_k}{(\alpha_k - 1)}\frac{1 + \tau_k}{\tau_k}, \tag{21}$$

and, for the first fraction (when no patient observations have yet been made), $k = 0$ and therefore $E[X_1] = \mu$, $\text{Var}[X_1] = \frac{\beta}{(\alpha-1)}\frac{1+\tau}{\tau}$. These are particularly important for the practitioner, since they are the mean and variance of the patient's *total* setup error for the $(k + 1)$th fraction, taking into account population data as well as the individual patient's

data on the previous k fractions. Indeed, these are precisely the values that should be used in computation of treatment margins. Note, in particular, that in the Bayesian context the concatenation of uncertainty in the systematic error M and in the random error R is given by the multiplicative form of equation (21). However, this is equivalent to the weighted sum often used (e.g. van Herk *et al* 2000), since in general (suppressing the fraction index k)

$$
\begin{aligned}
\mathrm{Var}[X] &= E[X^2] - E^2[X] \\
&= \mathrm{Var}[M] + E\left[\frac{1}{R}\right] \\
&= \frac{\beta}{\tau(\alpha-1)} + \frac{\beta}{\alpha-1} \\
&= \frac{\beta}{(\alpha-1)}\frac{1+\tau}{\tau}.
\end{aligned}
$$

2.5. Estimating prior parameters

A critical step in the implementation of a Bayesian method is the accumulation of population data related to systematic setup error M and random setup error given by the variance S^2 (or, equivalently, by the precision R) in order to create the prior distribution $f_{M,R}^{(0)}(m,r)$. The most straightforward way to estimate the prior parameters α, β, μ and τ uses the method of moments. This can be done in one of two ways:

(a) Simultaneously solving the four equations (9) and (11) for α, β, μ and τ, where the expected values $E[\cdot]$ are replaced by the population sample averages and the variances $\mathrm{Var}[\cdot]$ are replaced by the population sample variances. Since solving for α using the two equations in (11) cannot be done analytically (it involves inverting the Gamma function), we instead can define

$$
\rho \equiv \frac{\mathrm{Var}[S]}{(E[S])^2} = \frac{(\Gamma(\alpha)/\Gamma(\alpha-0.5))^2}{\alpha-1} - 1. \tag{22}
$$

An approximation (providing an error of less than 0.2% for realistic values of α) is given by

$$
\alpha = \frac{0.2505}{\rho} + 1.0954 - 0.0849\,\mathrm{e}^{-\frac{0.7329}{\rho}}. \tag{23}
$$

(b) Simultaneously solving the four equations (9) and (10) for α, β, μ and τ. This is algebraically simpler than (a), but requires estimates (and associated uncertainties) of the population precision, which is not a conventional way to report population data.

In general, since $E[R] \neq E\left[\frac{1}{S^2}\right]$, these two methods will not produce exactly the same prior parameter values; the difference, however, can be shown to be small for reasonable parameter values. In what follows, since most population data sets contain measurements and estimates of within-patient variance rather than precision, we will use method (a).

2.6. Adaptive treatment planning

Before the beginning of therapy, and therefore before any measurements on an individual patient are taken, treatment planning is based on $f_{M,S}^{(0)}(m,s)$. The method can be as simple as using $E[M]$, the expected value of the mean M as given in equation (9), to determine setup corrections, and using $E[S]$, the expected value of the standard deviation S as given in equation (11) together with $\sqrt{\mathrm{Var}[M]}$, to determine the margin according to some 'recipe' (e.g. van Herk *et al* 2000). Since the density function $f_{M,S}^{(0)}(m,s)$ is available besides population

Table 1. Population (prior) moments and parameters.

$E[M]$ (mm)	0.0
$E[S]$ (mm)	5.4
$\mathrm{Var}[M]$ (mm^2)	10.1
$\mathrm{Var}[S]$ (mm^2)	3.6
α	3.1
β (mm^2)	67.8
μ (mm)	0.0
τ	3.21

sample averages and variances, it is preferable to use $\sqrt{\mathrm{Var}[X]}$ directly in the estimation of the margin, where $\mathrm{Var}[X]$ is given by equation (21) with $k = 0$. Or, a more sophisticated method of optimizing the treatment plan (e.g., Löf *et al* 1998) might be used, which requires the entire distribution $f_{M,S}^{(0)}(m, s)$.

As treatment progresses and data from the specific patient become available, the posterior parameter values α_k, β_k, μ_k and τ_k can be computed from equations (12)–(15) and used to compute $E[M_k]$, $E[S_k]$ and $\sqrt{\mathrm{Var}[M]}$ (or $\sqrt{\mathrm{Var}[X_k]}$ if it is used instead), using equations (16) and (18) or (21), which in turn will generate updated setup corrections and margins for the re-planning of treatment. If the treatment plan is one that involves optimization, the entire posterior distribution can be used in the associated computation.

Once the posterior distribution is determined, decisions can be made about the adaptive treatment of the patient. Various policies for adjusting the setup and margin of the treatment, based on the posterior distribution, can be formulated. For example, one may take a 'zero-action-level' policy and correct the systematic setup error and re-plan with an updated margin based on the moments of the predictive distribution computed in equations (20) and (21) from the posterior distribution parameters. Or, one might make a one-time adjustment to the setup and margin after the nth fraction using $E[X_n]$ and $\mathrm{Var}[X_n]$, where for example, $n = 5$.

3. Example

To illustrate the procedure, setup errors were obtained from a population of eight patients. The population distribution statistics and parameters, determined using method (a) of section 2.5, are shown in table 1. Figure 1 shows, for a patient selected from this population distribution, the measured left–right position (solid diamonds in the figure) on the kth fraction. The open squares and their associated error bars are the predictive moments $E[X_k]$ and $\sqrt{\mathrm{Var}[X_k]}$, respectively, computed from $\alpha_{k-1}, \beta_{k-1}, \mu_{k-1}$ and τ_{k-1} using equations (20) and (21) for $k = 2, 3, \ldots$. For $k = 1$, the shaded square and the (heavy) error bar show the predictive moments $E[X_1]$ and $\sqrt{\mathrm{Var}[X_1]}$ based completely on the (prior) population parameters.

At every fraction k, the margin for treatment planning is typically a multiple of $\sqrt{\mathrm{Var}[X_k]}$, and $E[X_k]$ can be used to correct for the systematic setup error. Note that for this particular patient, the expected systematic setup errors (open squares) are close to the population systematic setup error $E[M] = 0$, which implies that relatively small bias corrections are needed. The predictive random setup error (error bars on the open squares) gradually decreases from over 12 mm at $k = 1$ to about 6 mm as treatment progresses. Thus, the margin would be reduced over time if a Bayesian adaptive procedure were to be used for this patient.

4. Discussion and conclusions

Two common approaches to adaptive radiation therapy either use a form of Kalman filtering or use running statistics for a specific patient. The argument for the latter is that the decision
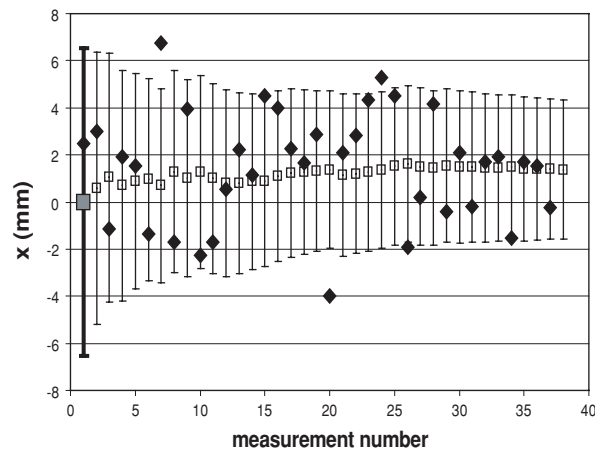
Figure 1. Typical patient-specific setup variances (black diamonds) during the course of treatment. The moments of the predictive distribution, $E[X_1]$ and $\sqrt{\text{Var}[X_1]}$, computed from the population distribution, are shown as the grey square data points and the (heavy) error bar at $k = 1$. The open squares show the updated predictive moments $E[X_k]$ and the associated error bars show $\sqrt{\text{Var}[X_k]}$, both computed by combining information from the population distribution with measured setup variations.

will not be affected by the population distribution which may not be representative of the individual patient, and that in the long run it will provide the best estimate for the parameters of the patient's distribution. However, the disadvantage is that there is only a finite number of fractions, and in order to benefit from adaptive radiation therapy, setup corrections need to be made as early as possible. The fluctuation in the running statistics is large during the initial phase of the treatment, and so the trade off is between making a confident decision and an early decision. Moreover, if population data *are* available and there is reason to believe that the particular patient has been 'selected' from this distribution, then a Bayes approach is the only consistent way to combine both sources of data.

It can be shown that the estimate of the systematic setup error using Kalman filtering as proposed by Yan *et al* (1997) is equal to $E[M_k]$ from the Normal–Gamma distribution for zero $E[M]$. However, the Kalman filter does not provide an estimate of the random setup error and the running standard deviation is usually used instead. This asymmetry is resolved by the Bayesian method which simultaneously provides information about the random setup error and the systematic setup error.

Bayesian methods provide a formal statistical inferential method to obtain the probability distribution of the systematic setup error and random setup error after a set of measurements is available. This probability distribution can be used directly in optimizations or to make decisions based on parameters such as the expected value, median or mode of the distribution. In this report, we have used a specific family of distribution functions for the prior (population) distribution and posterior (after observing errors over k fractions) distribution on (M, S), namely, the Normal–Gamma distribution on $(M, 1/S^2)$. This allows the parameters of the posterior distribution function to be determined analytically. Moreover, the family of Normal–Gamma distribution is a conjugate family of distributions under sampling. The procedure of computing the posterior distribution can be repeated without difficulty as more measurements are accumulated. We have not tested to see if the distribution on (M, S) for the patient population can be reasonably fitted to the Normal–Gamma distribution. However,

if a different distribution is more appropriate, the method can be implemented by using a numerical evaluation of equation (3) or (4).

Acknowledgment

Appendix. Extension to three dimensions

When the individual patient's fraction-to-fraction setup variation is the (random variable) 3-vector $\mathbf{X} \equiv [X, Y, Z]^t$ representing the three spatial dimensions, extension of the analysis is straightforward. We assume that $\mathbf{X}$ has a three-dimensional Normal distribution $f_{\mathbf{X}}(\mathbf{x})$ with three-dimensional mean vector $\mathbf{m}$ and 3×3 variance–covariance matrix $\mathbf{v}$. That is, $\mathbf{X}$ has the probability density function

$$f_{\mathbf{X}}(\mathbf{x}; \mathbf{m}, \mathbf{v}) = (2\pi)^{-3/2} |\mathbf{v}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^t \mathbf{v}^{-1}(\mathbf{x}-\mathbf{m})}. \tag{A.1}$$

A sequence of k independent and identically distributed setup variations over k fractions is the random $3 \times k$ matrix $[\mathbf{X}_1 \ \mathbf{X}_2 \ \ldots \ \mathbf{X}_k]$, where $\mathbf{X}_i$ is the variation on the ith fraction. $[\mathbf{X}_1 \ \mathbf{X}_2 \ \ldots \ \mathbf{X}_k]$ therefore will have the joint p.d.f.

$$f_{\mathbf{X}_1 \ \mathbf{X}_2 \ \ldots \ \mathbf{X}_k}(\mathbf{x}_1 \ \mathbf{x}_2 \ \ldots \ \mathbf{x}_k; \mathbf{m}, \mathbf{v}) = (2\pi)^{-3k/2} |\mathbf{v}|^{-\frac{k}{2}} e^{-[\frac{1}{2}\sum_{i=1}^{k}(\mathbf{x}_i-\mathbf{m})^t \mathbf{v}^{-1}(\mathbf{x}_i-\mathbf{m})]}. \tag{A.2}$$

When the values of the parameters $(\mathbf{m}, \mathbf{v})$ are unknown, we treat them as random variables $(\mathbf{M}, \mathbf{V})$ with prior (population) distribution $f_{\mathbf{M},\mathbf{V}}^{(0)}(\mathbf{m}, \mathbf{v})$, which can be obtained from patient population statistics.

In order to take advantage of the convenience of using a conjugate distribution, we define the precision matrix $\mathbf{R} \equiv \mathbf{V}^{-1}$ and let $f_{\mathbf{M},\mathbf{R}}^{(0)}(\mathbf{m}, \mathbf{r})$, the prior distribution on $(\mathbf{M}, \mathbf{R})$, be the Normal–Wishart distribution:

$$f_{\mathbf{M},\mathbf{R}}^{(0)}(\mathbf{m}, \mathbf{r}) = N_3(\mathbf{m}; \boldsymbol{\mu}, \tau \mathbf{r}) W(\mathbf{r}; \alpha, \beta). \tag{A.3}$$

In equation (A.3), $N_3(\mathbf{m}; \boldsymbol{\mu}, \tau \mathbf{r})$ is the three-dimensional Normal distribution

$$N_3(\mathbf{m}; \boldsymbol{\mu}, \tau \mathbf{r}) \equiv (2\pi)^{-3/2} \tau^{-\frac{1}{2}} |\mathbf{r}|^{-\frac{1}{2}} e^{-[\frac{1}{2\tau}(\mathbf{m}-\boldsymbol{\mu})^t \mathbf{r}^{-1}(\mathbf{m}-\boldsymbol{\mu})]} \tag{A.4}$$

and $W(\mathbf{r}; \alpha\beta)$ is the three-dimensional Wishart distribution

$$W(\mathbf{r}; \alpha, \beta) = C(\alpha) |\beta|^{-\alpha/2} |\mathbf{r}^{-(\alpha-4)/2}| e^{-\frac{1}{2}\text{trace}(\beta^{-1}\mathbf{r})}, \tag{A.5}$$

where $C(\alpha) = \left[2^{3\alpha/2} \pi^{3/2} \Gamma\left(\frac{\alpha}{2}\right) \Gamma\left(\frac{\alpha-1}{2}\right) \Gamma\left(\frac{\alpha-2}{2}\right)\right]^{-1}$. Note that the parameters α and τ are scalars equivalent to those used in the one-dimensional analysis of section 2.2, while $\boldsymbol{\mu}$ is a three-dimensional column vector and β is a symmetric positive-definite 3×3 matrix. The values of α and τ and the components of $\boldsymbol{\mu}$ and β can be obtained from population data by using the method of moments in an extension of the procedure outlined in section 2.3. After observing the three-dimensional errors $\mathbf{x}_1 \ \mathbf{x}_2 \ \ldots \ \mathbf{x}_k$, and defining the sample mean vector

$$\bar{\mathbf{x}} = \frac{1}{k} \sum_{i=1}^{k} \mathbf{x}_i$$

and

$$\phi = \frac{1}{k} \sum_{i=1}^{k} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t,$$

after k fractions the posterior parameters for the distribution on (M, R) are

$$\alpha_k = \alpha + k, \qquad \tau_k = \tau + k, \qquad \boldsymbol{\mu}_k = \frac{\tau\boldsymbol{\mu} + k\overline{\mathbf{x}}}{\tau + k},$$

$$\boldsymbol{\beta}_k = \boldsymbol{\beta} + \boldsymbol{\phi} + \frac{\tau\alpha}{\tau + \alpha}(\boldsymbol{\mu} - \overline{\mathbf{x}})(\boldsymbol{\mu} - \overline{\mathbf{x}})^t.$$

The posterior moments most useful to the practitioner, since they are the values needed in order to adapt the treatment plan for the $(k + 1)$th fraction, are given by the mean vector and variance–covariance matrix

$$E[\mathbf{X}_{k+1}] = \boldsymbol{\mu}_k, \qquad \text{Var}[\mathbf{X}_{k+1}] = \frac{(1 + \tau_k)(\alpha_k - 1)}{\tau_k}\boldsymbol{\beta}_k^{-1}. \qquad (\text{A.6})$$

References

Birkner M, Yan D, Alber M, Liang J and Nsslin F 2003 Adapting inverse planning to patient and organ geometrical variation: algorithm and implementation *Med. Phys.* **30** 2822–31

DeGroot M H 1970 *Optimal Statistical Decisions* (New York: McGraw-Hill)

Lam K L, Ten Haken R K and Pollock S 2004 Monte Carlo simulation of Bayes risk and Bayesian statistical decision theory in adaptive radiotherapy *Med. Phys.* **31** 1903–4 (abstract)

Löf J, Lind B K and Brahme A 1998 An adaptive control algorithm for optimization of intensity modulated radiotherapy considering uncertainties in beam profiles, patient set-up and internal organ motion *Phys. Med. Biol.* **43** 1605–28

Ten Haken R K, Balter J M, Brock K K, Dawson L A, Lam K L and Pollock S 2001 Combining patient specific and population based setup information for adaptive therapy *Radiother. Oncol.* **61** (Suppl 1) S36 (abstract)

Ten Haken R K, Balter J M and Lam K L 2000 Statistical control applications for dynamic refinement *Proc. 13th Int. Conf. on the Use of Computers in Radiation Therapy (ICCR)* ed W Schlegel and T Bortfeld (Heidelberg: Springer) pp 524–5

Unkelbach J and Oelfke U 2005 Incorporating organ movements in inverse planning: assessing dose uncertainties by Bayesian inference *Phys. Med. Biol.* **50** 121–39

van Herk M, Remeijer P, Rasch C and Lebesque J V 2000 The probability of correct target dosage: dose–population histograms for deriving treatment margins in radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **47** 1121–35

Yan D, Wong J, Vicini F, Michalski J, Pan C, Frazier A, Horwitz E and Martinez A 1997 Adaptive modification of treatment planning to minimize the deleterious effects of treatment setup errors *Int. J. Radiat. Oncol. Biol. Phys.* **38** 197–206