RESEARCH SUPPORT
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

DECEMBER 1996

A FULL BAYESIAN ANALYSIS OF CIRCULAR DATA USING VON MISES DISTRIBUTION

WORKING PAPER #9612-37

PAUL DAMIEN
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

AND

STEPHEN WALKER
IMPERIAL COLLEGE, LONDON, UK

A Full Bayesian Analysis of Circular Data using von Mises Distribution

Paul Damien[1]
and
Stephen Walker[2]

[1] Department of Statistics and Management Science, School of Business, University of Michigan, Ann Arbor, 48109-1234, USA.

[2] Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2BZ

SUMMARY

This paper presents a full Bayesian analysis of circular data paying special attention to the von Mises distribution. We obtain samples from the posterior distribution, given an independent sample from the von Mises distribution, using the Gibbs sampler after the introduction of some strategic latent variables which ensures all the full conditional distributions are of known type.

Keywords: Circular data, latent variables, Gibbs sampler, von Mises distribution.

1 Introduction

The aim of this paper is to provide the basis for a full Bayesian analysis for circular data. In particular we will be concentrating on the von Mises distribution (von Mises, 1918) though our method of analysis should also be applicable to spherical and cylindrical data.

Circular data arise naturally in a number of areas and for a recent survey from a frequentist perspective the reader is referred to the book of Fisher et al. (1987). Frequentist approaches have also been Arnold (1941), Fisher (1953), Gumbel (1954), Mardia (1972, 1975) and Bingham (1964). Some theoretical results in this context also appears in Lévy (1939). The Bayesian literature is far less extensive and to some extent has not been very successful. This has been due to the difficulties in working with the parametric distributions commonly associated with circular data, in particular, the von Mises distribution. The von Mises distribution is the most important distribution in the statistics of circular data and is the 'natural' analogue on the circle of the normal distribution on the real line.

The earliest attempt at Bayesian inference for the von Mises distribution was given by Mardia and El-Atoum (1976). However these authors assume the concentration parameter to be known and only provide point estimates for the directional parameter. Guttorp and Lockhart (1988) consider the case when both parameters are unknown but are forced to use the posterior maximum likelihood (mode) estimate for the concentration parameter (Lenth, 1981).

The Bayesian analysis of Bagchi (1987) and Bagchi and Guttman (1988) also restrict their attention to the von Mises distribution. More recently Bagchi and Kadane (1991) developed Laplace approximations to posterior distributions involving the von Mises distribution. They only provide Bayes estimates (posterior means) for the cosine of the directional parameter after taking the concentration parameter to be known.

In this paper we implement a full Bayesian analysis involving the von Mises distribution in which both parameters are assumed unknown and assigned the conjugate prior distribution introduced by Guttorp and Lockhart (1988). We derive the posterior distribution and via the introduction of strategic latent variables (Damien and Walker, 1996) are able to use a Gibbs sampler (Smith and Roberts, 1993) to simulate from this posterior, providing the basis for a full Bayesian analysis.

2 The von Mises distribution

The von Mises distribution is a symmetric unimodal distribution which is the most common model for unimodal samples of circular data. The probability density function is given by

$$f(\theta|\mu,\kappa) = [2\pi I_0(\kappa)]^{-1}\exp[\kappa\cos(\theta-\mu)] \quad 0 \le \theta < 2\pi, \quad \kappa \ge 0,$$

where

$$I_0(\kappa) = (2\pi)^{-1}\int_0^{2\pi} \exp[\kappa\cos\phi]d\phi$$

is the modified Bessel function of order zero. The mean direction is μ and κ is the concentration parameter.

Prior distribution

We use the conjugate prior suggested by Guttorp and Lockhart (1988). This is given, up to a constant of proportionality, by

$$f(\mu,\kappa) \propto I_0^{-c}(\kappa)\exp[\kappa R_0\cos(\mu-\mu_0)].$$

We will only be considering the case when c is a non-negative integer which is in the spirit of the interpretation of the prior parameters. Essentially the prior parameters can be thought of as representing c observations in the direction μ_0 and R_0 can be thought of as the component on the x axis (i.e., in the known direction) of the resultant of c observations. The uniform distribution on the circle is a limiting case of this prior distribution.

Posterior distribution

Let $\theta = (\theta_1, ..., \theta_n)$ be a sample of size n. The posterior distribution of (μ,κ) is given, up to a constant of proportionality, by

$$f(\mu,\kappa|\theta) \propto I_0^{-m}(\kappa)\exp[\kappa R_n\cos(\mu-\mu_n)],$$

where $m = c + n$ and μ_n and R_n are obtained from

$$R_n\cos\mu_n = R_0\cos\mu_0 + \sum_i \cos\theta_i$$

and

$$R_n\sin\mu_n = R_0\sin\mu_0 + \sum_i \sin\theta_i.$$

In the next section we develop a Gibbs sampler for generating random variates from the posterior.

3 Inference via the Gibbs sampler

Our approach will depend on the introduction of latent variables to define a joint distribution with (μ,κ). This joint distribution will be constructed to ensure that all full conditional distributions (required for the Gibbs sampler) are of known type and can be sampled directly. Damien and Walker (1996) exemplify this concept in a variety of applications after developing suitable theory, and Walker and Damien (1996) extend the ideas to the analysis of data in the context of neutral to the right processes. Consider first the latent variables w, with w defined on $(0,\infty)$, and v, also defined on $(0,\infty)$, and define their joint distribution with (μ,κ) by

$$f(\mu,\kappa,w,v) \propto e^{-R_n\kappa} I\left(v < e^{R_n\kappa[1+\cos(\mu-\mu_n)]}\right)\left\{w^{m-1}e^{-wI_0(\kappa)}\right\}.$$

Clearly the marginal distribution for (μ,κ) is as required. Our next step involves writing $I_0(\kappa)$ in the form

$$I_0(\kappa) = \sum_{k=0}^{\infty} \lambda_k \kappa^{2k},$$

where $\lambda_k = (k!)^{-2}0.5^{2k}$. Therefore the joint distribution becomes

$$f(\mu,\kappa,w,v) \propto e^{-R_n\kappa} I\left(v < e^{R_n\kappa[1+\cos(\mu-\mu_n)]}\right)\left\{w^{m-1}\prod_{k=0}^{\infty} e^{-w\lambda_k\kappa^{2k}}\right\}.$$

Next we introduce the latent variable $u=(u_1,u_2,\ldots)$ and x and define the joint distribution with (μ,κ,w,v) by

$$f(\mu,\kappa,w,v,u,x) \propto$$

$$e^{-R_n\kappa} I\left(v < e^{R_n\kappa[1+\cos(\mu-\mu_n)]}, x < w^{m-1}\right)\left\{e^{-w}\prod_{k=1}^{\infty} I\left(u_k < e^{-w\lambda_k\kappa^{2k}}\right)\right\}.$$

Again it is clear that the marginal for distribution for (μ,κ) is as required.

To implement the Gibbs sampler we need the full conditional distributions (denoted by a $*$). Also the symbol U denotes the uniform distribution. The full conditional for x is given by

$$f^*(x) = U(0, w^{m-1})$$

and for v is given by

$$f^*(v) = U\Big(0, e^{R_n\kappa[1+\cos(\mu-\mu_n)]}\Big).$$

The full conditional for μ is given by

$$f^*(\mu) = U(A),$$

where A is given by the set $\mu_n + \cos^{-1}[(R_n\kappa)^{-1}\log v - 1]$. The full conditional for w is given by

$$f^*(w) \propto e^{-w} I\Big(x^{1/(m-1)} < w < \min_k\{-(\lambda_k\kappa^{2k})^{-1}\log u_k\}\Big).$$

and the full conditional for κ is given by

$$f^*(\kappa) \propto e^{-R_n\kappa} I\Big(\max\{0, v_n\} < \kappa < \min_k\{[-(w\lambda_k)^{-1}\log u_k]^{1/(2k)}\}\Big),$$

where $v_n = \log v/[R_n + R_n\cos(\mu - \mu_n)]$. Lastly, the full conditional for each of the u_k is given by

$$f^*(u_k) = U\Big(0, e^{-w\lambda_k\kappa^{2k}}\Big).$$

It is clear that we cannot sample all the u_ks since there are an infinite number of them. However we do not need to sample all of them in order to implement the Gibbs sampler (recall it is the (μ, κ) samples we are interested in). In fact it it easy to see from the full conditionals of w and κ that we only need to obtain $\min_k\{-(\lambda_k\kappa^{2k})^{-1}\log u_k\}$ and $\min_k\{[-(w\lambda_k)^{-1}\log u_k]^{1/(2k)}\}$. We will only concern ourselves with the former of these as the latter will follow by essentially the same method.

Since the full conditional for u_k is the uniform distribution on $(0, e^{-\tilde{w}\lambda_k\kappa^{2k}})$ we can take $u_k = \tau_k e^{-\tilde{w}\lambda_k\kappa^{2k}}$ where τ_k are iid $U(0,1)$ and $\tilde{w}$ is the 'current' w. Therefore the upper bound for the (exponential) full conditional distribution of w is given by

$$\min_k\{\tilde{w} - \alpha_k(\kappa)\log\tau_k\},$$

where $\alpha_k(\kappa) = (\lambda_k\kappa^{2k})^{-1}$, and note that $\alpha_k \to \infty$. Essentially then we only need to consider $\{\beta_k(\kappa) = -\alpha_k(\kappa)\log\tau_k\}$. In figure 1 we illustrate $\beta_k(\kappa)$ (on log scale) for $\kappa = 1, 5, 10, 20, 40$ and 80 and for each of these it is clear that obtaining the minimum value is very tractable.

From figure 1 we observe that the minimum value seems to occur when k is approximately $\kappa/2$ (certainly for the larger values of κ.) This can be explained theoretically. Our interest in the following discussion is in understanding $\bar{k} = k : M_\kappa = -\alpha_k(\kappa)\log\tau_k$ where $M_\kappa = \min_k\{-\alpha_k(\kappa)\log\tau_k\}$. Recall that the τ_ks are iid $U(0,1)$ and so clearly the 'expected' value of $\bar{k}$ will be that k minimising $(k!)^2(\kappa/2)^{-2k}$. Taking logarithms we see that this reduces to finding that k which minimises $\log k! - k\log(\kappa/2)$. If we allow k to take non-integral values and replace $k!$ by $\Gamma(k+1)$, where Γ represents the gamma function, then we want that k such that $\psi(k+1) = \log(\kappa/2)$ (here ψ is the digamma function). According to Abramowitz and Segun (1968) the asymptotic formula for ψ is given by $\psi(k) = \log k - 1/(2k) - 1/(12k)^2 \cdots$ and so a good approximation for the required k (in particular for large κ) is provided by $\kappa/2$.

To investigate this point further and to see what the variance of $\bar{k}$ looks like we simulated 1000 M_{80} random variables and the resulting histogram representation of the sample is given in figure 2. The mean value of the sample is 39.8 and the standard deviation is 4.6.

4 Numerical examples

Example 1. Here we consider the von Mises distribution and the roulette data with sample size 9 (see Mardia 1972, p.2). We simulated from the joint posterior distribution for (μ, κ) using the algorithm described in Section 3. We take $c = 0$, $\mu_0 = 0$ and $R_0 = 0$ and the resulting marginal posterior distributions for $\cos\mu$ and κ are presented in figure 3.

Example 2. We also consider the data set presented in Guttorp and Lockhart (1988, Table 1.) Seven signals from an unknown location (the target) are picked up at observed compass readings by receiver stations at known locations. If μ_i is the true direction from station i to the target then

$$f(\theta_i|\mu_i,\kappa) = [2\pi I_0(\kappa)]^{-1}\exp[\kappa\cos(\theta_i - \mu_i)].$$

Therefore only minor modifications are required to the algorithm outlined in the previous section. We take a flat prior for the μ_i and take

$$f(\kappa) \propto I_0^{-c}(\kappa)\exp(R_0\kappa)$$

as the prior for κ with $(R_0, c) = (5, 5)$ (see Guttorp and Lockhart for this prior and alternatives). The posterior plot of the location of the target is given in figure 4.

5 Discussion

In this paper we have provided a full Bayesian analysis of circular data modelled via the von Mises distribution. It is quite straightforward to extend the idea of latent variable substitution resulting in full conditional distributions of known type in the Gibbs sampler to observations on the surface of a p-dimensional hypersphere S_p of unit radius and centered on the origin. This would make use of the von Mises-Fisher distribution (see, for example, Mardia and El-Atoum, 1976) with density function given by

$$f(\theta|\mu,\kappa) = c_p(\kappa)\exp[\kappa\mu'\theta],$$

where as before μ is the mean direction, κ the concentration parameter and $c_p(\kappa) = \kappa^{(p-1)/2}/[(2\pi)^{p/2}I_{(p-1)/2}]$ is the constant factor where $I_r(\kappa)$ is the modified Bessel function of the first kind given (which is important to us) as the series

$$I_r(\kappa) = \sum_{k=0}^{\infty}\frac{(\kappa/2)^{2k+r}}{\Gamma(r+k+1)\Gamma(k+1)},$$

where Γ is the gamma function. Therefore we can proceed in the same way as for the circular data using the conjugate or noninformative priors proposed by Kardia and El-Atoum (1976).

6 References

Abramowitz,M. and Segun,I.A. (1968). *Handbook of Mathematical Functions.* Dover Publications Inc., New York.

Arnold,K.J. (1941). *On spherical probability distributions.* Ph.D. Thesis, Massachusetts Institute of Technology.

Bagchi,P. (1987). *Bayesian analysis of directional data.* Ph.D. Thesis, University of Toronto.

Bagchi,P. and Guttman,I. (1988). Theoretical considerations of the von Mises-Fisher distribution. *Journal of Applied Statistics* **15**, 149-169.

Bagchi,P. and Kadane,J.B. (1991). Laplace approximations to posterior moments and marginal distributions on circles, spheres, and cylinders. *The Canadian Journal of Statistics* **19**, 67-77.

Bingham,C. (1964). *Distributions on the sphere and on the projective plane.* Ph.D. Thesis, Yale University.

Damien,P. and Walker,S.G. (1996). Sampling probability densities via uniform random variables and a Gibbs sampler. *Submitted for publication.*

Fisher,N.I.,Lewis,T. and Embleton,B.J.J. *Statistical analysis of spherical data.* Cambridge University Press.

Fisher,R.A. (1953). Dispersion on a sphere. *Proceedings of the Royal Society* ***A217***, 295-305.

Gumbel,E.J. (1954). Applications of the circular normal distribution. *Journal of the American Statistical Association* **48**, 233-246.

Guttorp,P. and Lockhart,R.A. (1988). Finding the location of a signal: A Bayesian analysis. *Journal of the American Statistical Association* **83**, 322-329.

Lenth,R.V. (1981). On finding the source of a signal. *Technometrics* **23**, 149-154.

Lévy,P. (1939). L'addition des variables aléatoires défines sur une circonférences. *Bull. Soc. Math. France* **67**, 1-41.

Mardia,K.V. (1972). *Statistics of directional data.* London: Academic Press.

Mardia,K.V. (1975). Statistics of directional data. *Journal of the Royal Statistical Society Series B* **37**, 97-133.

Mardia,K.V. and El-Atoum,S.A.M. (1976). Bayesian inference for the von Mises-Fisher distribution. *Biometrika* **63**, 203-205.

Von Mises,R. (1918). Uber die "Ganzzahligkeit" der Atomgewicht und Verwandte Fragen. *Physikal. Z.* **19**, 490-500.

Smith,A.F.M. and Roberts,G.O. (1993). Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B* **55**, 3-23.

Walker,S.G. and Damien,P. (1996). A full Bayesian analysis involving a neutral to the right process. *Submitted for publication.*

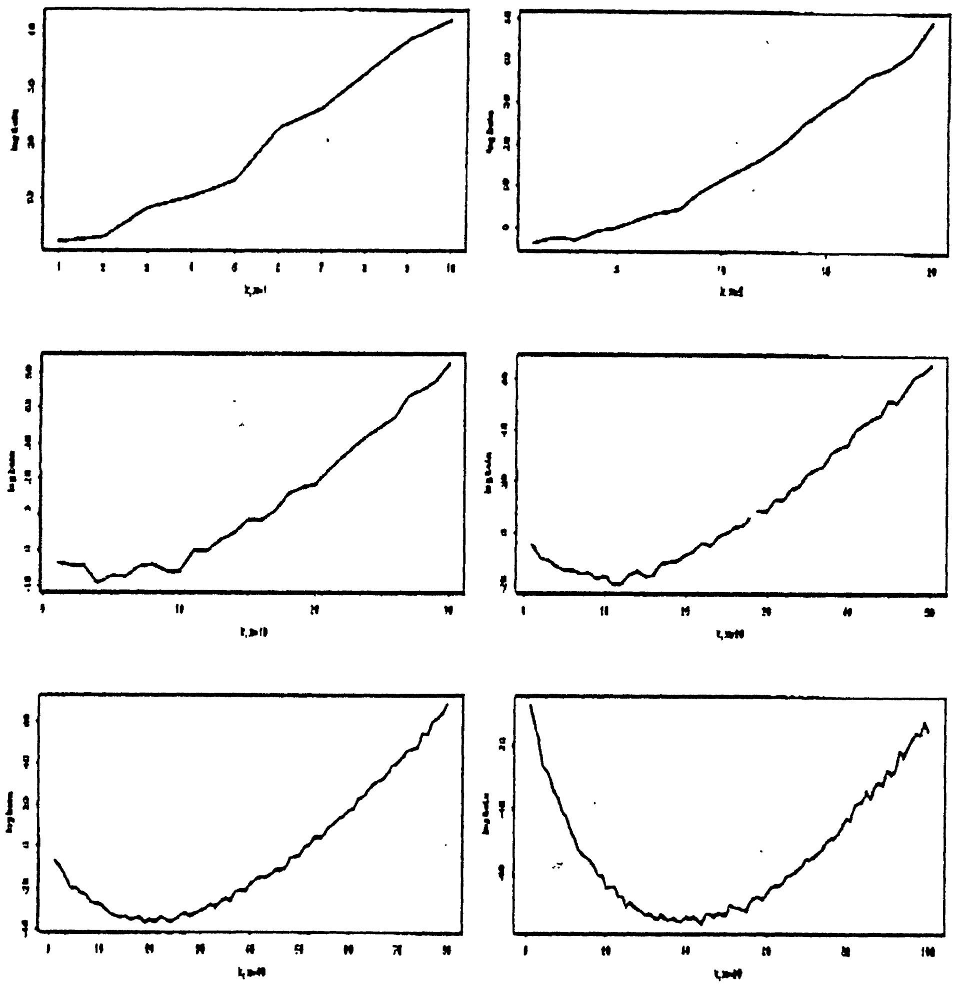

Figure 1: Plots of log beta (k,x) versus k for x=1,5,10,20,40 and 80

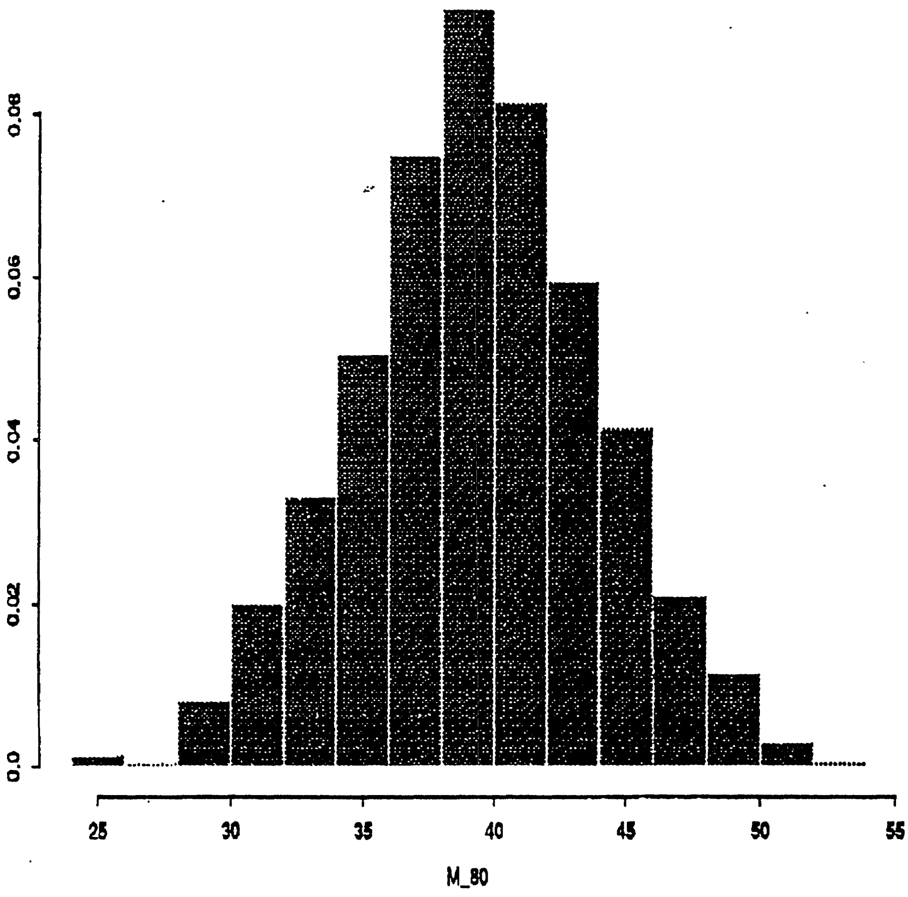

Figure 2: Histogram approximation of the density for M_{80}

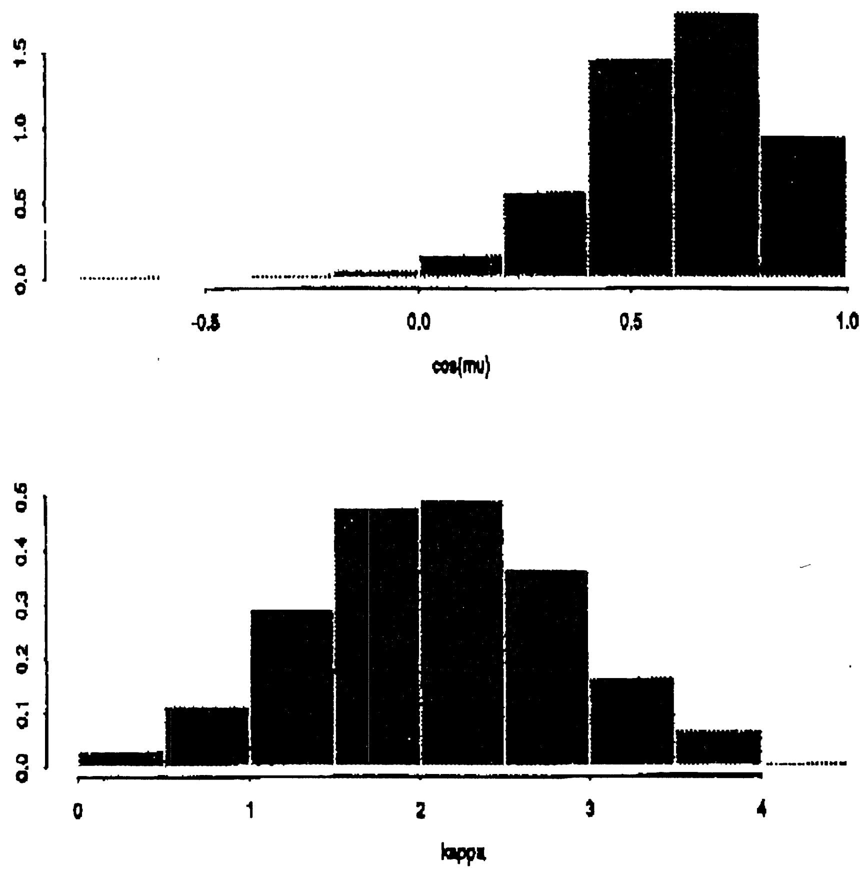

Figure 3: Histogram approximation of the marginal posterior densities for $\cos\mu$ (top) and κ (bottom)

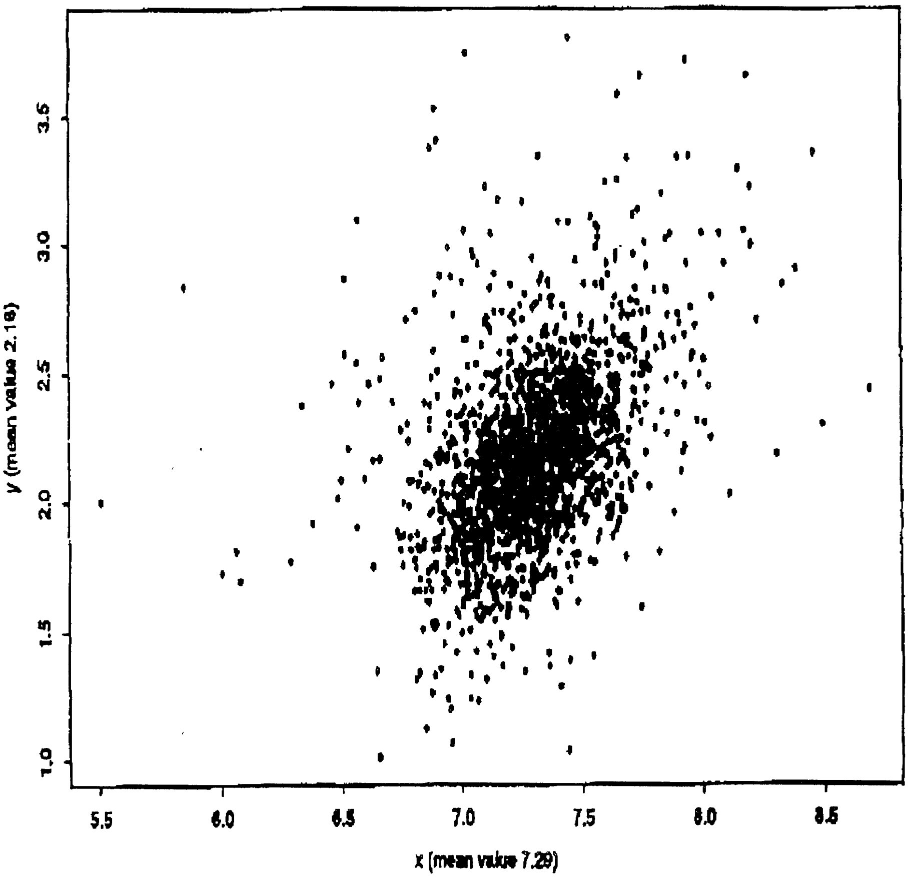

Figure 4: Posterior plot of location of the target