Bayes Estimation Subject to Uncertainty About Parameter Constraints
Author(s): A. O'Hagan and Tom Leonard
Source: *Biometrika*, Vol. 63, No. 1 (Apr., 1976), pp. 201-203
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2335105

here only for determining the higher order determinants which are the most difficult to calculate by conventional methods.

We can prove a multivariate generalization of the theorem in §2 which states, under similar restrictions, that if $\{\partial l(\theta, \phi)/\partial\theta\}_g = 0$ for all values of the vector ϕ, where the subscript g means evaluation at $\theta = g(\phi)$, then the matrix $(\partial^2 l/\partial\gamma^2)_g$ is positive, or negative, definite if the two matrices $(\partial^2 l/\partial\theta^2)_g$ and $d\{(\partial l/\partial\phi)_g\}/d\phi$ are both positive, or both negative, definite. This theorem will be useful if, having solved $\partial l/\partial\theta = 0$ for θ in terms of ϕ, and having found that the corresponding pivots are negative, we find that the remaining equations, $(\partial l/\partial\phi)_g = 0$, require numerical solution; then a numerical study of the matrix $d\{(\partial l/\partial\phi)_g\}/d\phi$ is all that is required to establish the nature of the solutions.

The method presented here fails if one of the pivots is zero. In this case we must either investigate higher order derivatives or reparameterize. It would be interesting to know if the pivots could be interpreted as second derivatives in particular directions in the parameter space for then we might be able to take further derivatives just at the stage that gave the zero pivot. This matter needs further investigation.

This work was done while I was on leave from the New Zealand Ministry of Agriculture and Fisheries and the holder of a New Zealand National Research Advisory Council Fellowship. I would like to thank Professor D. R. Cox for his help.

References

Copas, J. B. (1972). The likelihood surface in the linear functional relationship problem. *J. R. Statist. Soc.* B **34**, 274–8.

Fraser, D. A. S. (1958). *Statistics, an Introduction*. New York: Wiley.

Jowett, G. H. (1963). Applications of Jordan's procedure for matrix inversion in multiple regression and multivariate distance analysis. *J. R. Statist. Soc.* B **25**, 352–7.

Richards, F. S. G. (1961). A method of maximum-likelihood estimation. *J. R. Statist. Soc.* B **23**, 469–75.

Solari, M. E. (1969). The 'maximum likelihood solution' of the problem of estimating a linear functional relationship. *J. R. Statist. Soc.* B **31**, 372–5.

[*Received April* 1975. *Revised September* 1975]

Bayes estimation subject to uncertainty about parameter constraints

By A. O'HAGAN and TOM LEONARD

Department of Statistics, University of Warwick, Coventry

Summary

When estimating a normal location parameter, an inequality constraint is sometimes indicated by theoretical considerations. If, however, this constraint is assumed *a priori* definitely to hold, then the data will never be permitted to formally contradict the theory. We propose a family of skew prior distributions which reflect uncertainty about the constraint.

Some key words: Hierarchical prior; Inequality constraints; Posterior mean; Skew prior; Truncated normal distribution.

Several estimation methods have been proposed for situations where unknown parameters are subject to inequality constraints. For example, O'Hagan (1973) considers this type of problem in a quadratic regression context, employing truncated prior distributions which lead to Bayesian alternatives to the constrained maximum likelihood estimates. The data may, however, provide strong evidence that the constraints are untrue, in which case it might sometimes be desirable for the estimates to contradict the constraints. In this note we construct a family of prior distributions with smooth positively skew densities, which enables us to express a degree of belief in the constraint, and permits the data to play a fuller role in the estimation procedure.

We confine attention to the simple situation where an observation x is normally distributed with known variance τ^2, conditionally upon its unknown mean θ. Suppose that prior considerations suggest the constraint $\theta \geqslant 0$, but that we are not completely convinced about this. Instead of taking the prior distribution of θ to be truncated to the interval $[0, \infty)$, we specify its structure via the following two stages.

Stage I. Given μ, θ has the untruncated normal distribution with mean μ and known variance σ^2.

Stage II. The first stage mean μ has the distribution obtained by truncating the normal distribution with known mean μ_0 and known variance σ_0^2 to the interval $[0, \infty)$.

The first stage variance σ^2 measures the degree of belief in the constraint, which is actually applied at the second stage. If $\sigma^2 = 0$ the constraint applies with probability one, whilst as $\sigma^2 \to \infty$ the distribution of θ becomes uniform over the whole of the interval $(-\infty, \infty)$. For $0 < \sigma^2 < \infty$ it is possible to show that $\mathrm{pr}(\theta \geqslant 0)$ is a strictly decreasing function of σ^2.

The prior density of θ over $(-\infty, \infty)$ is

$$\pi(\theta) \propto \phi\{(\sigma_0^2+\sigma^2)^{\frac{1}{2}}(\theta-\mu_0)\}\,\Phi\{(\sigma^{-2}+\sigma_0^{-2})^{-\frac{1}{2}}(\sigma^{-2}\theta+\sigma_0^{-2}\mu_0)\}, \tag{1}$$

where $\phi(.)$ and $\Phi(.)$ respectively denote the standard normal density and distribution function. The first contribution to the right-hand side of (1) takes the form of the more usual conjugate normal prior for θ, with mean μ_0 and variance $\sigma_0^2+\sigma^2$. This is modified by the second term which is an increasing function of θ, causing the density in (1) to become positively skew. Our new density possesses a thinner negative tail, and thicker positive tail than the normal prior, and is always a smooth function for $\theta \in (-\infty, \infty)$, unless $\sigma^2 = 0$ when θ has a truncated normal distribution. In the important case where $\sigma_0^2 = \infty$, we obtain the interesting improper density

$$\pi(\theta) \propto \Phi(\sigma^{-1}\theta). \tag{2}$$

When σ^2 lies between 0 and ∞, the distribution in (2) compromises between its two extremes, i.e. the uniform distributions over the ranges $[0, \infty)$ and $(-\infty, \infty)$ respectively.

Since the joint posterior distribution of θ and μ is truncated bivariate normal it is possible to use the results of Tallis (1961) to show that the posterior mean and variance of θ are

$$\begin{aligned} \theta^* &= E(\theta|x) = (\tau^{-2}+\sigma^{-2})^{-1}(\tau^{-2}x+\sigma^{-2}\mu^*), \\ \mathrm{var}(\theta|x) &= (\tau^{-2}+\sigma^{-2})^{-1}+\sigma_1^2\{1-\kappa(\sigma_1^{-1}\mu_1)\}, \end{aligned} \tag{3}$$

where

$$\mu^* = E(\mu|x) = \mu_1+\sigma_1 M(\sigma_1^{-1}\mu_1), \quad \kappa(z) = zM(z)+\{M(z)\}^2$$

with

$$M(z) = \phi(z)/\Phi(z), \quad \mu_1 = (\zeta^{-2}+\sigma_0^{-2})^{-1}(\zeta^{-2}x+\sigma_0^{-2}\mu_0), \quad \sigma_1^{-2} = \sigma_0^{-2}+\zeta^{-2}, \quad \zeta^2 = \sigma^2+\tau^2.$$

For $0 < \sigma^2 < \infty$, the posterior mean in (3) will be negative whenever $-x$ is large enough, although it takes account of the prior information about the possibility that $\theta \geqslant 0$. Consider, for example, the special case $\sigma_0^2 = \infty$, when we have

$$\theta^* = x+\zeta^{-1}\tau^2 M(\zeta^{-1}x). \tag{4}$$

The function in (4) asymptotically approximates its lower bound x as $x \to \infty$, so that it virtually ignores the constraint when x is large. As $x \to -\infty$ it approaches the asymptote $\theta^* = \rho x$, where $\rho = \zeta^{-2}\sigma^2 < 1$. Therefore, when $\sigma^2 > 0$ and x is large and negative, θ^* takes substantial account of the constraint, but may still be negative. For general x, the function in (4) increases x by an amount which is strictly decreasing in x.

For general σ_0^2 the quantity in (3) is always greater than

$$\hat{\theta} = (\tau^{-2}+\sigma_2^{-2})^{-1}(\tau^{-2}x+\sigma_2^{-2}\mu_0), \tag{5}$$

where $\sigma_2^2 = \sigma^2+\sigma_0^2$. The expression in (5) is the posterior mean obtained upon neglecting the second contribution to the right-hand side of (1). Our modification in (3) is similar in spirit to that in (4) but is slightly more complex since it allows for prior information about the location of θ.

We have demonstrated the feasibility of relaxing parameter constraints in such a way that meaningful estimates may be obtained, in particular when the data appear to contradict the constraints. Our approach combines a fairly realistic prior distribution with a straightforward posterior analysis. Although we have only examined a simple model, the general idea of deferring truncation to the second stage of a prior hierarchy can be usefully employed in more complex models without greatly increasing the technical problems. For example, in the one-way analysis of variance, or the simultaneous estimation of several regression lines, it is possible to construct exchangeable prior distributions, while still allowing for the possibility of constraints.

Even when there is no rationale of constrained parameters, distributions like (1) can be used to provide smooth skew prior distributions over $(-\infty, \infty)$ removing the assumption of symmetry implicit in most prior distributions in current use.

References

O'Hagan, A. (1973). Bayes estimation of a convex quadratic. *Biometrika* **60**, 565–72.

Tallis, G. M. (1961). The moment generating function of a truncated multi-normal distribution. *J. R. Statist. Soc.* B **23**, 223–9.

[*Received September* 1975]

Bayesian inference for the von Mises–Fisher distribution

By K. V. MARDIA and S. A. M. EL-ATOUM

Department of Statistics, University of Leeds

Summary

The main aim of this note is to give a theoretical discussion of Bayesian inference for the von Mises–Fisher distribution. The choice of particular priors is considered and the admissibility of certain Bayesian estimators studied. For the multisample case estimators are given. Some problems of testing hypotheses are summarized in the form of posterior odds against the null hypothesis.

Some key words: Bayesian directional data analysis; Loss function; One-way classification; Prior distribution; von Mises–Fisher distribution.

1. Introduction

The von Mises–Fisher distribution is the most important distribution in the statistics of directional data. Let $l' = (l_1, \ldots, l_p)$ be a unit random vector taking values on the surface of a p-dimensional hypersphere S_p of unit radius and having its centre at the origin. The unit random vector l is said to have p-variate von Mises–Fisher distribution, $M_p(\mu, \kappa)$, if its probability density function is given by

$$c_p(\kappa)\, e^{\kappa\mu' l} \quad (\kappa > 0, \mu'\mu = 1, l \in S_p), \tag{1·1}$$

where κ is the concentration parameter, μ is the mean direction vector and $c_p(\kappa)$ the constant factor (Watson & Williams, 1956; Mardia, 1972, 1975). For the von Mises case, $p = 2$, we say that a circular variable θ is distributed as $M(\mu_0, \kappa)$.

2. Prior distributions

2·1. *Conjugate and uniform prior distributions*

Let $L = (l_1, \ldots, l_n)$ be a random sample of unit vectors from $M_p(\mu, \kappa)$ with known κ. Assuming that the conjugate prior of μ is of the form $M_p(\mu^*, \kappa^*)$, the posterior density of μ is

$$h(\mu|L) = c_p(\kappa R) \exp(\kappa R \mu' \bar{l}_0), \tag{2·1}$$

where R is the resultant length and $\bar{l}_0$ is the vector of mean direction cosines obtained by combining the conjugate prior and current data. If κ^* tends to zero, the prior distribution of μ is uniform and the posterior density is completely determined by the likelihood. Since (1·1) is of the exponential family structure, the form given by (2·1) is expected.