ELSEVIER

Statistics & Probability Letters 42 (1999) 33–37

STATISTICS & PROBABILITY LETTERS

On the posterior distribution of a location parameter from a strongly unimodal distribution

Chunsheng Ma

Department of Statistics, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada

Received July 1997; received in revised form December 1997

Abstract

For a single observation $X=x$ from a distribution having unknown location parameter Θ, we investigate the behaviour of the posterior distribution as x varies, when the likelihood or the prior distribution is strongly unimodal. Extensions and generalizations of several results in the literature are obtained.

Keywords: Location parameter; Logconcavity; Normal distribution; Posterior distribution; Prior distribution; Strong unimodality; TP_2

1. Introduction

A distribution function $F(x)$ is unimodal if there exists a value $x=a$ such that $F(x)$ is convex for $x<a$ and concave for $x>a$. It is said to be strongly unimodal if the convolution of F with any unimodal distribution is unimodal. According to Ibragimov (1956), a non-degenerate distribution F is strongly unimodal if and only if it has a density f that is log-concave within some open interval. Such a density is also called a PF_2 density; see Barlow and Proschan (1981). Typical examples of strongly unimodal distributions include: normal, truncated normal, exponential, double exponential (or Laplace), logistic, uniform, extreme value, Pareto, triangular and some of the Gamma, the beta and Weibull distributions. For descriptions of the multivariate unimodality, we refer to Barndorff–Nielsen (1978).

In this paper we consider inference from a single observation X about a location parameter Θ. Specifically, given $\Theta=\theta$, the density function of X has the form $f(x-\theta)$ for a known function f that is assumed to be log-concave within some open interval, while the prior Θ has a proper density. Our purpose is to investigate the behaviour of the posterior distribution of Θ, as $X=x$ varies. It is shown that the posterior density $\pi(\theta\,|\,x)$ is TP_2 in (θ,x) for any prior, provided that $f(\cdot)$ is log-concave. This important property results in many consequences, which include those derived in the literature under stronger conditions or in the special case where the likelihood or the prior is a normal distribution. Employing the dual argument of Dawid (1973), we also consider the behaviour of the posterior distribution of $X-\Theta$, as $X=x$ varies.

0167-7152/99/$ – see front matter
PII: S0167-7152(98)00183-7

2. Main results

Throughout this paper assume that given $\Theta=\theta$, X has a density of the form $f(x-\theta)$, and that Θ possesses a density $\pi(\theta)$. Denote by $\pi(\theta\,|\,x)$ the posterior density. All results are given in the univariate case, except for Theorem 4. The term "increasing" means non-decreasing and "decreasing" means "non-increasing".

Theorem 1. *If $f(\cdot)$ is log-concave, then for any prior $\pi(\cdot)$,*
(i) *$\pi(\theta\,|\,x)$ is TP_2 in (θ,x), namely,*

$$\begin{vmatrix} \pi(\theta_1\,|\,x_1) & \pi(\theta_2\,|\,x_1) \\ \pi(\theta_1\,|\,x_2) & \pi(\theta_2\,|\,x_2) \end{vmatrix} \geqslant 0, \quad \textit{for every } \theta_1\leqslant\theta_2 \textit{ and } x_1\leqslant x_2; \tag{2.1}$$

(ii) *Both*

$$\frac{P(\Theta\leqslant\theta\,|\,X=x_2)}{P(\Theta\leqslant\theta\,|\,X=x_1)} \quad \textit{and} \quad \frac{P(\Theta\geqslant\theta\,|\,X=x_2)}{P(\Theta\geqslant\theta\,|\,X=x_1)}$$

are increasing in θ for every $x_1<x_2$;
(iii) *Both*

$$\frac{P(\Theta\leqslant\theta_2\,|\,X=x)}{P(\Theta\leqslant\theta_1\,|\,X=x)} \quad \textit{and} \quad \frac{P(\Theta\geqslant\theta_2\,|\,X=x)}{P(\Theta\geqslant\theta_1\,|\,X=x)}$$

are increasing in x for every $\theta_1<\theta_2$;
(iv)

$$\frac{E\{\alpha(\Theta)\,|\,X=x\}}{E\{\beta(\Theta)\,|\,X=x\}}$$

is increasing in x for all functions α and β, for which the integrals are well-defined, β is non-negative, and α/β and β are increasing.

Proof. (i) Define $p(x)=\int_{-\infty}^{\infty} f(x-\theta)\pi(\theta)\,\mathrm{d}\theta$. Then

$$\begin{vmatrix} \pi(\theta_1\,|\,x_1) & \pi(\theta_2\,|\,x_1) \\ \pi(\theta_1\,|\,x_2) & \pi(\theta_2\,|\,x_2) \end{vmatrix} = \frac{\pi(\theta_1)\pi(\theta_2)}{p(x_1)p(x_2)} \begin{vmatrix} f(x_1-\theta_1) & f(x_1-\theta_2) \\ f(x_2-\theta_1) & f(x_2-\theta_2) \end{vmatrix}.$$

As indicated by Barlow and Proschan (1981, p. 76), the logconcavity of $f(\cdot)$ is equivalent to the TP_2 property of $f(x-\theta)$ in (θ,x), from which Eq. (2.1) follows immediately.

(ii) Since $\pi(\theta\,|\,x)$ is TP_2 in (θ,x), $P(\Theta\leqslant\theta\,|\,X=x)$ is also TP_2 in (θ,x). In fact, for $\theta_1<\theta_2$ and $x_1<x_2$ one has

$$\begin{vmatrix} P(\Theta\leqslant\theta_1\,|\,X=x_1) & P(\Theta\leqslant\theta_2\,|\,X=x_1) \\ P(\Theta\leqslant\theta_1\,|\,X=x_2) & P(\Theta\leqslant\theta_2\,|\,X=x_2) \end{vmatrix} = \begin{vmatrix} P(\Theta\leqslant\theta_1\,|\,X=x_1) & P(\theta_1\leqslant\Theta\leqslant\theta_2\,|\,X=x_1) \\ P(\Theta\leqslant\theta_1\,|\,X=x_2) & P(\theta_1\leqslant\Theta\leqslant\theta_2\,|\,X=x_2) \end{vmatrix}$$

$$= \int\!\!\int_{t_1\leqslant\theta_1\leqslant t_2\leqslant\theta_2} \begin{vmatrix} \pi(t_1\,|\,x_1) & \pi(t_2\,|\,x_1) \\ \pi(t_1\,|\,x_2) & \pi(t_2\,|\,x_2) \end{vmatrix} \mathrm{d}t_1\,\mathrm{d}t_2 \geqslant 0.$$

Similarly, $P(\Theta\geqslant\theta\,|\,X=x)$ is TP_2 in (θ,x).

(iii) It is an alternative version of Part (ii).

(iv) This follows from Lemma 2 of Bickel and Lehmann (1975) or Corollary of Capéraà (1988). This completes the proof of Theorem 1. □

An immediate consequence of Theorem 1(iii) is that for every $\theta_1 < \theta_2$, the limits of

$$\frac{P(\Theta \leqslant \theta_2 \mid X = x)}{P(\Theta \leqslant \theta_1 \mid X = x)} \quad \text{and} \quad \frac{P(\Theta \geqslant \theta_2 \mid X = x)}{P(\Theta \geqslant \theta_1 \mid X = x)}$$

exist as $x \to \infty$. Letting $\theta_2 \to \infty$ in Theorem 1(iii) yields

Corollary 1.1. *$P(\Theta \leqslant \theta \mid X = x)$ is decreasing in x for all θ and for any prior, provided that $f(\cdot)$ is log-concave.*

Corollary 1.2 below was proved by O'Hagan (1979) under a further condition that $f(\cdot)$ has a bounded derivative. It states that a strongly unimodal distribution is outlier-resistant in the sense that it yields posterior distributions which continue to respond to an extreme observation as it goes to infinity, regardless of prior information. His precise definition of the outlier-resistant distribution is repeated here for convenience.

Definition (*outlier-resistance*). Let $X_1, X_2, \ldots, X_{n+1}$ be i.i.d. given $\Theta = \theta$ with densities $f(x_i - \theta)$. The distribution having density $f(\cdot)$ is said to be *outlier-resistant* if

$$P(\Theta \leqslant c \mid X_1 = x_1, \ldots, X_{n+1} = x_{n+1})$$

is a decreasing function of x_{n+1} for all $c, x_1, \ldots, x_n$, for all $n = 0, 1, 2, \ldots$, and for any prior distribution for Θ.

Corollary 1.2. *A strongly unimodal distribution is outlier-resistant.*

Proof. As shown in the proof of O'Hagan (1979, Theorem 2), it suffices to consider a single observation $X = x$ with density $f(x - \theta)$. The desired result follows now from Corollary 1.1. □

Using Corollary 1.2 it is easy to see that Theorem of Goldstein (1983) holds even if $f(\cdot)$ has no bounded derivative.

Taking $\alpha(t) = \psi(t)$ and $\beta(t) \equiv 1$ in Theorem 1(iv), one obtains Corollary 1.3 below, which is shown by Andrews et al. (1972, Lemma 1) in the particular case where f is a normal density and $\psi(t) = t$. See also Mitchell (1994).

Corollary 1.3. *$E\{\psi(\Theta) \mid X = x\}$ is increasing in x for all increasing function $\psi(\cdot)$ for which the integral is well-defined, if $f(\cdot)$ is log-concave.*

Let $\Phi = X - \Theta$. Then

$$X = \Theta + \Phi, \tag{2.2}$$

where Θ and Φ are independent with densities $\pi(\cdot)$ and $f(\cdot)$ respectively. An important fact is that Θ and Φ enter symmetrically into the relation (2.2), so that from observation of X one is able to make inferences about Θ, or equivalently about Φ. Consequently, there is an interesting symmetry in the problem between the roles of $f(\cdot)$ and $\pi(\cdot)$. More precisely, given $X = x$, the posterior density of Θ is proportional to $f(x - \theta)\pi(\theta)$ when the prior Θ has density $\pi(\theta)$; the posterior density of Φ is proportional to $\pi(x - \phi) f(\phi)$, in which the roles of $f(\cdot)$ and $\pi(\cdot)$, as likelihood and prior respectively, are reversed. From Theorem 1, this dual argument yields

Theorem 2. *If $\pi(\cdot)$ is log-concave, then for any likelihood,*

(i) *Both*

$$\frac{P(X - \Theta \leqslant \theta \mid X = x_2)}{P(X - \Theta \leqslant \theta \mid X = x_1)} \quad \text{and} \quad \frac{P(X - \Theta \geqslant \theta \mid X = x_2)}{P(X - \Theta \geqslant \theta \mid X = x_1)}$$

are increasing in θ for every $x_1 < x_2$;

(ii) *Both*

$$\frac{P(X-\Theta\leqslant\theta_2\,|\,X=x)}{P(X-\Theta\leqslant\theta_1\,|\,X=x)} \quad \textit{and} \quad \frac{P(X-\Theta\geqslant\theta_2\,|\,X=x)}{P(X-\Theta\geqslant\theta_1\,|\,X=x)}$$

are increasing in x for every $\theta_1<\theta_2$;
(iii)

$$\frac{E\{\alpha(X-\Theta)\,|\,X=x\}}{E\{\beta(X-\Theta)\,|\,X=x\}}$$

is increasing in x for all functions α *and* β, *for which the integrals are well-defined,* β *is non-negative and* α/β *and* β *are increasing.*

Corollary 2.1. $P(\Theta-x\leqslant\theta\,|\,X=x)$ *is decreasing in x for all* θ.

Corollary 2.2. $E\{\psi(X-\Theta)\,|\,X=x\}$ *is increasing in x for all increasing function* $\psi(\cdot)$ *for which the integral is well-defined.*

Corollary 2.3. $E(\Theta\,|\,X=x)-x$ *is decreasing in x.*

Corollary 2.3 is derived by Mitchell (1994) in the particular case where $f(\cdot)$ is a normal and $\pi(\cdot)$ is a double-exponential density.

Theorem 3. *If both* $f(\cdot)$ *and* $\pi(\cdot)$ *are log-concave, then* $E\{\psi(\Theta,X-\Theta)\,|\,X=x\}$ *is increasing in x for any function* $\psi(\cdot,\cdot)$ *which is increasing in each of its arguments.*

Proof. Consider the relationship (2.2). Under the assumption of the theorem, Θ and $\Phi=X-\Theta$ are independent with PF_2 densities $\pi(\cdot)$ and $f(\cdot)$ respectively. The assertion of Theorem 3 is readily obtained by using a theorem of Efron (1965). □

Now suppose that both X and Θ are $n\times 1$ vectors, given $\Theta=\theta$, X has a density of the form $f(x-\theta)$, and that Θ possesses a density $\pi(\theta)$. For two symmetric matrices Σ_1 and Σ_2, by $\Sigma_1\leqslant\Sigma_2$ we mean that $\Sigma_2-\Sigma_1$ is non-negative definite.

Theorem 4. *If* $f(\cdot)$ *is a multivariate normal density with known variance matrix* Σ *and* $\pi(\cdot)$ *is log-concave in the n-dimensional Euclidean space* $\mathbb{R}^n$, *then for all* $x\in\mathbb{R}^n$,

$$\mathrm{Var}(\Theta\,|\,X=x)\leqslant\Sigma. \tag{2.3}$$

Proof. Let

$$Q(x)=\log\int f(x-\theta)\pi(\theta)\,\mathrm{d}\theta,$$

and

$$\nabla Q(x)=\left(\frac{\partial}{\partial x_1}Q(x),\ldots,\frac{\partial}{\partial x_n}Q(x)\right)^{\mathrm{T}},\quad \nabla^2 Q(x)=\left(\frac{\partial^2}{\partial x_i\partial x_j}Q(x)\right)_{n\times n},$$

where T stands for the transpose operation. It can be shown, in a similar way to that of Polson (1991) and Pericchi and Smith (1992), that

$$E(\Theta \mid X = x) - x = \Sigma \cdot \bigtriangledown Q(x),$$

and

$$\mathrm{Var}(\Theta \mid X = x) = \Sigma + \Sigma \cdot \bigtriangledown^2 Q(x) \cdot \Sigma.$$

According to Barndorff–Nielsen (1978, Corollary 6.1, p. 96), $Q(x)$ is concave, so that the matrix $-\bigtriangledown^2 Q(x)$ is non-negative definite, which in turn implies that $-\Sigma \bigtriangledown^2 Q(x)\Sigma$ is non-negative definite. So is $\Sigma - \mathrm{Var}(\Theta \mid X = x)$. The proof of Theorem 4 is established. □

In the univariate case, Eq. (2.3) is illustrated by Mitchell (1994) when $\pi(\cdot)$ is a double-exponential density and the likelihood is normal. It is different from that given by Theorem of O'Hagan (1981), where Θ is assumed to have a symmetric unimodal prior density and to satisfy his condition (1). An important example is a Student-t prior, for which Pericchi and Sansó (1995) show that $\lim_{x\to\infty}\{E(\Theta \mid X = x) - x\} = 0$, and O'Hagan (1981) and Table 2 of Pericchi and Smith (1992) illustrate that there exists a $x_0 > 0$ such that $\mathrm{Var}(\Theta \mid X = x_0) > \sigma^2$. Using a Student-$t$ prior to provide a robust analysis for a univariate normal location parameter, rather than using a log-concave prior such as the double-exponential, has been suggested by O'Hagan (1979), and Pericchi and Smith (1992), among others. For a multivariate normal location parameter, Eq. (2.3) says that the posterior variance in the case of a log-concave prior is never greater than the sample variance Σ, which implies that an alternative prior should be employed in Bayesian robustness studies.

Acknowledgements

This work is supported in part by a NSERC Operating grant. I am very grateful to Professors Anthony O'Hagan and Harry Joe for their helpful comments on an earlier draft, which was written at The University of Sydney, Australia.

References

Andrews, R.L., Arnold, J.C., Krutchkoff, R.G., 1972. Shrinkage of the posterior measure in the normal case. Biometrika 59, 693–695.
Barlow, R.E., Proschan, F., 1981. Statistical theory of reliability and life testing: probability models. To Begin With, Silver Spring, MD.
Barndorff–Nielsen, O.E., 1978, Information and Exponential Families in Statistical Theory. Wiley, New York.
Bickel, P.J., Lehmann, E.L., 1975. Descriptive statistics for nonparametric models, II, location. Ann. Statist. 3, 1045–1069.
Capéraà, P., 1988. Tail ordering and asymptotic efficiency of rank tests. Ann. Statist. 16, 470–478.
Dawid, A.P., 1973. Posterior expectations for large observations. Biometrika 60, 664–667.
Efron, B., 1965. Increasing properties of Pólya frequency functions. Ann. Math. Statist. 36, 272–279.
Goldstein, M., 1983. Outlier resistant distributions: where does the probability go? J. Roy. Statist. Soc. B 45, 355–357.
Ibragimov, I.A., 1956. On the composition of unimodal distributions. Theoret. Probab. Appl. 1, 255–260.
Mitchell, A.F.S., 1994. A note on posterior moments for a normal mean with double-exponential prior. J. Roy. Statist. Soc. B 56, 605–610.
O'Hagan, A., 1979. On outlier rejection phenomena in Bayes inference. J. Roy. Statist. Soc. B 41, 358–367.
O'Hagan, A., 1981. A moment of indecision. Biometrika 68, 329–330.
Pericchi, L.R., Sansó, B., 1995. A note on bounded influence in Bayesian analysis. Biometrika 82, 223–225.
Pericchi, L.R., Smith, A.F.M., 1992. Exact and approximate posterior moments for a normal location parameter. J. Roy. Statist. Soc. B 54, 793–804.
Polson, N.G., 1991. A representation of the posterior mean for a location model. Biometrika 78, 426–430.