Ann. Inst. Statist. Math.
39 (1987), Part A, 211-218

ON THE NORMALITY A POSTERIORI FOR EXPONENTIAL DISTRIBUTIONS, USING THE BAYESIAN ESTIMATION

MONICA E. BAD DUMITRESCU

(Received June 10, 1985)

Summary

The Bayesian estimation problem for the parameter θ of an exponential probability distribution is considered, when it is assumed that θ has a natural conjugate prior density and a loss-function depending on the squared error is used. It is shown that, with probability one, the posterior density of the Bayesian—centered and scaled parameter converges pointwise to the normal probability density. The weak convergence of the posterior distributions to the normal distribution follows directly. Both correct and incorrect models are studied and the asymptotic normality is stated respectively.

1. Introduction

The asymptotic behaviour of the posterior distribution for exponential models has been discussed widely in the literature. Le Cam [5], Berk [4] and other authors have studied the consistency and the asymptotic normality of the maximum likelihood estimate when the model being used actually governs the data. Berk [2], [3] has established the conditions under which a sequence of posterior distributions converges weakly to a degenerate distribution, even when the model is incorrect. Crain and Morgan [6] have shown that the posterior probability density of the centered and scaled parameter, using the maximum likelihood estimation, tends pointwise to the normal density with probability one, when the model is correct.

In this paper we study the limiting behaviour of the scaled and centered parameter with respect to the Bayesian estimation, for the exponential-type distributions.

In [1] it is shown that for a natural conjugate prior distribution, the loss equal to the squared error, multiplied by an appropriate fac-

Key words and phrases: Bayesian estimate, posterior distribution, asymptotic normality.

tor, leads to tractible expressions for the Bayesian estimator and the posterior expected loss.

In Section 2 we list some basic properties of the considered exponential models and we briefly review the construction of the Bayesian estimation $\hat{\theta}_n$. In Section 3 we suppose that the exponential model actually governs the data (the model is correct) and we show that, with probability one, the random variable $\sqrt{n}(\theta-\hat{\theta}_n)$ converges weakly to the normal distribution with the mean zero and the variance related to the Fisher information. In Section 4 we consider the case of an incorrect model. Using Berk's result concerning the weak convergence of a sequence of posterior distributions to a degenerate distribution, we establish the asymptotic normality of the variable $\sqrt{n}(\theta-\hat{\theta}_n)$ in this case too.

2. Exponential distributions . Construction of the Bayesian estimation

Let us consider a measurable space $(\Omega, \mathcal{K})$ and a random variable f, which takes values in $(\mathcal{X}, \mathcal{B})$, where $\mathcal{X}\subset R$ and $\mathcal{B}$ is the σ-field of the borelian subsets of $\mathcal{X}$. We suppose that the distribution function of f, F_θ, depends on the parameter θ. We shall also denote by F_θ the joint distribution of a sequence of independent, identically distributed (F_θ) random variables $f_1, f_2, \cdots$.

We suppose that F_θ is an exponential distribution, whose probability density with respect to a σ-finite measure μ on $\mathcal{X}$ belongs to the family

$$\mathcal{P}=\{p(x|\theta)=\exp[a(\theta)b(x)-c(a(\theta))],\ x\in\mathcal{X},\ \theta\in\Theta\},$$

where

$$\Theta=\left\{\theta\in R\,\middle|\,\int_{\mathcal{X}}\exp[a(\theta)b(x)-c(a(\theta))]d\mu(x)<\infty\right\}.$$

We notice that

$$\Theta=\left\{\theta\in R\,\middle|\,c(a(\theta))=\ln\int_{\mathcal{X}}\exp[a(\theta)b(x)]d\mu(x)<\infty\right\}.$$

Assumptions on the exponential model:

A.1. $\operatorname{Int}\Theta\neq\Phi$ (Θ has nonempty interior).

A.2. $\theta\to a(\theta)$ is a diffeomorfism between Θ and $a(\Theta)$.

A.3. μ is not supported on a flat.

Properties of the exponential model:

We list here some basic properties of the exponential model, which are to be found, in their general form, in Berk [4].

P.1. Θ is a convex subset of R.
$\Theta_1 = \{\theta \in \Theta | E_\theta[|b(f)|] < \infty\}$ is a convex set too, and $\text{Int}\,\Theta \subset \Theta_1$ (here E_θ denotes the expectation with respect to F_θ)

P.2. $c(a(\theta))$ is convex and lower semi-continuous on R, continuous and infinitely differentiable on $\text{Int}\,\Theta$.

P.3. $\theta \to c'(a(\theta))$ is $1-1$ on Θ_1 and $c''(a(\theta)) > 0$ for $\theta \in \text{Int}\,\Theta$.

Construction of the Bayesian estimation:

Following Shapiro's and Wardrop's method ([6]) for the construction of the Bayesian estimation of the mean parameter θ of an one-parameter exponential family, the estimation of $E_\theta[b(f)]$ for a general exponential family has been constructed in [1].

For the here considered exponential model, we assume that θ has a natural conjugate prior density

$$\lambda_0(\theta) = \frac{\exp[a(\theta)b(\alpha) - \beta c(a(\theta))]}{\int_\Theta \exp[a(\theta)b(\alpha) - \beta c(a(\theta))]d\theta},$$

where

$$(\alpha, \beta) \in \left\{(x, y) \in R^2 \middle| \int_\Theta \exp[a(\theta)b(x) - yc(a(\theta))]d\theta < \infty\right\}.$$

Let us consider $f_1, \cdots, f_n$ independent, identically distributed random variables, with the probability density $p(x|\theta) \in \mathcal{P}$.

Then the posterior density of θ, given $(x_1, \cdots, x_n)$ is

$$\lambda_n(\theta) = \frac{\exp\left\{a(\theta)\left[b(\alpha) + \sum_{i=1}^n b(x_i)\right] - (\beta + n)c(a(\theta))\right\}}{\int_\Theta \exp\left\{a(\theta)\left[b(\alpha) + \sum_{i=1}^n b(x_i)\right] - (\beta + n)c(a(\theta))\right\}d\theta}.$$

We are interested in the estimation of $\varphi(\theta) = E_\theta[b(f)] = c'(a(\theta))$. Let us consider the loss-function

$$L(\theta, \tilde{\theta}) = L(\varphi, \tilde{\varphi}) = a'(\theta)[\varphi(\theta) - \tilde{\varphi}]^2.$$

The minimum posterior expected loss

$$\inf_{\tilde{\varphi}} \int_\Theta a'(\theta)[\varphi(\theta) - \tilde{\varphi}]^2 \lambda_n(\theta)d\theta$$

is attained for

$$\hat{\varphi}_n = \frac{\int_\Theta a'(\theta)\varphi(\theta)\lambda_n(\theta)d\theta}{\int_\Theta a'(\theta)\lambda_n(\theta)d\theta}.$$

Hence, $\hat{\varphi}_n$ is the solution of the equation

$$\int_\Theta a'(\theta)(\hat{\varphi}_n - c'(a(\theta)))\lambda_n(\theta)d\theta = 0 .$$

From the condition

$$\int_\Theta \left\{\frac{d}{d\theta} \exp\left[a(\theta)\left(b(\alpha)+\sum_{i=1}^n b(x_i)\right)-(\beta+n)c(a(\theta))\right]\right\} d\theta = 0 ,$$

we get

$$\int_\Theta a'(\theta)\left[\left(b(\alpha)+\sum_{i=1}^n b(x_i)\right)-(\beta+n)c'(a(\theta))\right]\lambda_n(\theta)d\theta = 0 .$$

Then the Bayesian estimation of $\varphi(\theta)$ is

$$\hat{\varphi}_n(x_1, \cdots, x_n) = \frac{b(\alpha)+\sum_{i=1}^n b(x_i)}{\beta+n} .$$

3. Asymptotic normality when the model is correct

Let us consider that the exponential model actually governs the data. It means that the true, but unknown value of the parameter, θ_0, lies in the interior of Θ.

By the strong law of large numbers, $\hat{\varphi}_n \to \varphi(\theta_0)$ F_{θ_0}—almost sure for $n \to \infty$.

Let $\hat{\theta}_n$ be the solution of the equation $c'(a(\hat{\theta}_n)) = \hat{\varphi}_n$. Then $\hat{\theta}_n$ is the Bayesian estimation of the parameter θ and $\hat{\theta}_n \to \theta_0$ F_{θ_0}—a.s. for $n \to \infty$.

It is well established ([5]) that the posterior distribution Λ_n becomes degenerate at θ_0 for $n \to \infty$, F_{θ_0}—a.s. Then, for any neighborhood V of θ_0, $\lim_{n\to\infty} \Lambda_n(V) = 1$ F_{θ_0}—a.s.

On the basis of these two facts, we establish the following result.

THEOREM 1. *We assume that the considered exponential model is correct,* A.1.-A.3. *are verified and* $\hat{\theta}_n$ *is the Bayesian estimation of* θ. *Then, the random variable* $\sqrt{n}(\theta - \hat{\theta}_n)$ *converges in distribution for* $n \to \infty$ *to the normal distribution* $N\left(0, \frac{1}{(a'(\theta_0))^2 c''(a(\theta_0))}\right)$ F_{θ_0}*—almost sure.*

PROOF. We consider the variable $Z_n = \sqrt{n}(\theta - \hat{\theta}_n)$. The posterior probability density of Z_n is

$$q_n(z) = \frac{\exp\left\{a\left(\frac{z}{\sqrt{n}}+\hat{\theta}_n\right)\left[b(\alpha)+\sum_{i=1}^n b(x_i)\right]-(\beta+n)c\left(a\left(\frac{z}{\sqrt{n}}+\hat{\theta}_n\right)\right)\right\}}{\sqrt{n}\int_\Theta \exp\left\{a(\theta)\left[b(\alpha)+\sum_{i=1}^n b(x_i)\right]-(\beta+n)c(a(\theta))\right\} d\theta} . \tag{3.1}$$

Let us denote

$$u=\frac{z}{\sqrt{n}}, \qquad \bar{b}_n=\frac{1}{n}\sum_{i=1}^{n} b(x_i),$$

$$\xi_n(u)=a(u+\hat{\theta}_n)\Big(\frac{b(\alpha)}{n}+\bar{b}_n\Big)-\Big(1+\frac{\beta}{n}\Big)c(a(u+\hat{\theta}_n)).$$

Then $q_n(z)$ can be written as

$$\frac{\exp\{n[\xi_n(u)-\xi_n(0)]\}}{\sqrt{n}\int_\Theta \exp\Big\{n\Big[a(\theta)\Big(\frac{b(\alpha)}{n}+\bar{b}_n\Big)-\Big(1+\frac{\beta}{n}\Big)c(a(\theta))-\xi_n(0)\Big]\Big\}d\theta}. \tag{3.2}$$

By the Taylor series expansion of $\xi_n(u)$ about $u=0$ we get $\xi_n(u)-\xi_n(0) \simeq \frac{1}{2}u^2\cdot\xi_n''(v)$, where v lies on the line segment between u and the origin.

Then, the numerator of (3.2) becomes

$$\exp\Big\{\frac{z^2}{2}a''(v+\hat{\theta}_n)\Big[\Big(\frac{b(\alpha)}{n}+\bar{b}_n\Big)-\Big(1+\frac{\beta}{n}\Big)c'(a(v+\hat{\theta}_n))\Big]$$
$$-\frac{z^2}{2}(a'(v+\hat{\theta}_n))^2\Big(1+\frac{\beta}{n}\Big)c''(a(v+\hat{\theta}_n))\Big\}.$$

Since $\hat{\theta}_n \to \theta_0$ for $n\to\infty$ F_{θ_0}-a.s., with probability one, the numerator of (3.2) tends pointwise to

$$\exp\Big[-\frac{z^2}{2}(a'(\theta_0))^2c''(a(\theta_0))\Big].$$

Now, by the consistency a posteriori, for any neighborhood V of θ_0, $\lim_{n\to\infty}\int_V \lambda_n(\theta)d\theta=1$ F_{θ_0}-a.s.

Hence

$$\lim_{n\to\infty}\frac{\int_V \exp\Big\{n\Big[a(\theta)\Big(\frac{b(\alpha)}{n}+\bar{b}_n\Big)-\Big(1+\frac{\beta}{n}\Big)c(a(\theta))\Big]\Big\}d\theta}{\int_\Theta \exp\Big\{n\Big[a(\theta)\Big(\frac{b(\alpha)}{n}+\bar{b}_n\Big)-\Big(1+\frac{\beta}{n}\Big)c(a(\theta))\Big]\Big\}d\theta}=1.$$

Since $\hat{\theta}_n \to \theta_0$ for $n\to\infty$ F_{θ_0}-a.s., for any positive δ, there exists a neighborhood V of θ_0 such that

$$V\subset\{\theta\,|\,|\theta-\hat{\theta}_n|<\delta\} \quad \text{or} \quad V\subset\{|u|<\delta\}.$$

Then

$$\lim_{n\to\infty} \frac{\int_{|u|<\delta} \exp\{n[\xi_n(u)-\xi_n(0)]\}du}{\sqrt{n}\int_\Theta \exp\left\{n\left[a(\theta)\left(\frac{b(\alpha)}{n}+\bar{b}_n\right)-\left(1+\frac{\beta}{n}\right)c(a(\theta))-\xi_n(0)\right]\right\}d\theta}=1\,.$$

Then, the denominator of (3.2) has the same limiting behaviour as the expression

$$\int_{|z|<\delta\sqrt{n}} \exp\left\{-\frac{1}{2}z^2(a'(v+\hat{\theta}_n))^2\left(1+\frac{\beta}{n}\right)c''(a(v+\hat{\theta}_n))\right\}dz\,. \tag{3.3}$$

Since the integrant in (3.3) tends pointwise to $\exp\left[-\frac{1}{2}z^2(a'(\theta_0))^2\times c''(a(\theta_0))\right]$ F_{θ_0}-a.s. and is bounded above by the integrable function $\exp\left(-\frac{1}{2}z^2d\right)$, where

$$d=\inf_{|\theta-\theta_0|<2\delta}[(a'(\theta))^2c''(a(\theta))]\,,$$

it follows from the Lebesgue dominated convergence theorem that the denominator of (3.2) tends to

$$\int \exp\left[-\frac{1}{2}z^2(a'(\theta_0))^2c''(a(\theta_0))\right]dz\,.$$

Thus, the pesterior density $q_n(z)$ of the variable $Z_n=\sqrt{n}(\theta-\hat{\theta}_n)$ tends pointwise to the probability density $N(0;\ 1/(a'(\theta_0))^2c''(a(\theta_0)))$ F_{θ_0}-a.s.. The convergence in distribution follows immediately from Scheffè's theorem.

In this proof we have used the technique introduced by Crain and Morgan [6].

4. Asymptotic normality when the model is incorrect

Let us consider the same exponential model

$$\mathcal{P}=\{p(x|\theta)=\exp[a(\theta)b(x)-c(a(\theta))],\ x\in\mathcal{X},\ \theta\in\Theta\}\,,$$

where

$$\Theta=\left\{\theta\in R\,\middle|\,c(a(\theta))=\ln\int_{\mathcal{X}}\exp[a(\theta)b(x)]d\mu(x)<\infty\right\}.$$

We assume that the real, unknown distribution of the random variable f is F, not belonging to the family $\mathcal{P}$. In the following, F will refer also to the joint distribution of the sequence of independent, identically distributed random variables $f_1, f_2, \cdots$.

We denote $m_F = E[b(f)]$, where the expectation is taken with respect to the distribution F.

By the strong law of large numbers, the constructed Bayesian estimation of $\varphi(\theta)$, $\hat{\varphi}_n = \frac{1}{\beta+n}\left(b(\alpha)+\sum_{i=1}^{n} b(f_i)\right)$ converges to m_F for $n \to \infty$, F-a.s.

Let

$$\gamma(\theta) = E[\ln p(f|\theta)] = a(\theta)m_F - c(a(\theta)) \quad \text{and} \quad \gamma^* = \sup_{\theta \in \Theta} \gamma(\theta),$$

$$D = \{\xi \in R \mid \sup_{\theta \in \Theta} [a(\theta)\xi - c(a(\theta))] < \infty\},$$

$$H_n = \left\{\xi \in R \,\middle|\, \int_\Theta \exp\{a(\theta)[b(\alpha)+n\xi] - (\beta+n)c(a(\theta))\} d\theta < \infty\right\}.$$

Berk [3] has established the conditions under which a sequence of posterior distributions converges weakly to a degenerate distribution. For the considered exponential model, consistency a posteriori is given in the next proposition, which follows from Berk's result.

PROPOSITION. *Let us assume that the exponential model satisfies* A.1.–A.3. *and that the real probability distribution F verifies the following conditions:*

B.1. *$b(x)$ belongs to the linear span of $\bigcup_n (H_n \cap D)$ with probability one.*

B.2. $E[|b(f)|] < \infty$.

B.3. *m_F belongs to the interior of $\bigcup_n (H_n \cap D)$.*

Let θ^* be the value at which $\gamma(\theta)$ attains its supremum: $\gamma(\theta^*) = \gamma^*$.

Then, the posterior distribution Λ_n converges weakly to the distribution degenerate at θ^*.

For the considered model and the Bayesian estimation $\hat{\varphi}_n$ let $\gamma_n = \sup_{\theta \in \Theta} [a(\theta)\hat{\varphi}_n - c(a(\theta))]$. Since $\hat{\varphi}_n \to m_F$ for $n \to \infty$, F-a.s., we get $\gamma_n \to \gamma^*$ for $n \to \infty$, F-a.s.

If $\hat{\varphi}_n$ belongs to $\operatorname{Int} D$, let $\hat{\theta}_n$ be defined by the relation

$$\gamma_n = a(\hat{\theta}_n)\hat{\varphi}_n - c(a(\hat{\theta}_n)).$$

Then $\hat{\theta}_n \to \theta^*$ for $n \to \infty$, F-a.s.

Using this fact and the consistency a posteriori, we establish the asymptotic normality in the case of an incorrect model.

THEOREM 2. *If the exponential model satisfies the assumptions* A.1.–A.3. *and the real distribution F satisfies the conditions* B.1.–B.3., *the random variable $\sqrt{n}(\theta - \hat{\theta}_n)$ converges in distribution to the normal distribution $N(0;\ 1/(a'(\theta^*))^2 c''(a(\theta^*)))$ F-almost sure.*

We notice that the variance of the limiting normal distributions depends on Fisher's information in both cases.

UNIVERSITY OF BUCHAREST

REFERENCES

[1] Bad Dumitrescu, M. (1985). On two sequential estimations for one-parameter exponential families, *Annals of the University of Bucharest*, 3-11.

[2] Berk, R. (1966). Limiting behaviour of posterior distributions when the model is incorrect, *Ann. Math. Statist.*, **37**, 51-59.

[3] Berk, R. (1970). Consistency a posteriori, *Ann. Math. Statist.*, **41**, 894-907.

[4] Berk. R. (1972). Consistency and asymptotic normality of MLE's for exponential models, *Ann. Math. Statist.*, **43**, 193-205.

[5] Le Cam L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates, *Univ. Calif. Publ. Statist.*, **1**, 277-300.

[6] Crain, B. and Morgan, R. (1975). Asymptotic normality of the posterior distribution for exponential models, *Ann. Statist.*, **3**, 223-227.

[7] Shapiro, C. P. and Wardrop, R. L. (1980). Bayesian sequential estimation for one-parameter exponential families, *J. Amer. Statist. Ass.*, **75**, No. 372, 984-989.