International Journal of Statistics and Probability; Vol. 5, No. 3; May 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Bayesian Sequential Estimation of the Inverse of the Pareto Shape Parameter

Mohamed Tahir[1]

[1] Department of Mathematics, University of Sharjah, Sharjah, United Arab Emirates

Correspondence: Mohamed Tahir, Department of Mathematics, University of Sharjah, Sharjah, United Arab Emirates. Tel: 971-50-9525614. E-mail: tahir_stat@yahoo.com

Received: February 11, 2016 Accepted: February 29, 2016 Online Published: April 14, 2016

doi:10.5539/ijsp.v5n3p49 URL: http://dx.doi.org/10.5539/ijsp.v5n3p49

Abstract

The problem addressed is that of developing a sequential procedure for estimating the inverse of the shape parameter of the Pareto distribution under the squared loss, assuming that the shape parameter is the value of a random variable having a density function with compact support and that the cost per observation is one unit. A stopping time is proposed and a second-order asymptotic expansion is obtained for the Bayes regret incurred by the proposed procedure.

Keywords: Bayes estimator, Bayes regret, Fatou's Lemma, martingales, posterior distribution, Pareto distribution, stopping time, uniform integrability.

1. Introduction

Let $X_1, \ldots, X_n$ denote independent observations to be taken sequentially from the Pareto distribution with p.d.f.

$$f_\theta(x) = \begin{cases} \dfrac{\theta}{x^{\theta+1}} & \text{if } x \geq 1 \\ 0 & \text{if not,} \end{cases} \tag{1}$$

where θ is an unknown positive number. The Pareto distribution was first developed in the late 1800s by the Italian economist Vilfredo Pareto, who used this distribution to describe the allocation of wealth among individuals. The Pareto distribution can also be used as the distribution for insurance loss, oil reserve in a oil field, standardized price return on an individual stock, area burnt in a forest fire and for many other real-life measurements.

A sequential method will be used to estimate the parameter $\delta(\theta) = 1/\theta$, subject to the loss function

$$L_a(\theta) = a^2\left[\hat{\delta}_n - \delta(\theta)\right]^2 + n, \tag{2}$$

where a is a known positive number, $\hat{\delta}_n$ is the Bayes estimator of $\delta(\theta)$ and assuming that the cost per observation is one unit. In a sequential investigation, the sample size n is not chosen in advance; instead, data are analyzed as they become available and whether to stop taking observations is decided according a stopping time t, say. t is a stopping time means that t takes on the values 1, 2, … and has the properties that $P\{t < \infty\} = 1$ and that $\{t = n\} \in \mathfrak{D}_n$ for each integer $n \geq 1$, where $\mathfrak{D}_n$ is the sigma-algebra generated by $X_1, , \ldots, X_n$. The advantage of using sequential methods in estimation or hypothesis testing problems is that procedures can be constructed with a substantially smaller number of observations compared to equally reliable procedures based on a predetermined sample size.

Throughout this paper, it is assumed that θ (the shape parameter) is a value of a random variable Θ having (prior) density function ξ with compact support in $(0, \infty)$ and the objective is to determine a stopping time t for which the Bayes regret (see (5) below) of the procedure $(t, \hat{\delta}_t)$ is as small as possible for large a. In order to anticipate the nature of the stopping time t, it is necessary to find the best fixed sample size. So, let E^ξ denote expectation with respect to a probability measure P^ξ, under which $X_1, X_2, \ldots$ are independent random variables with conditional p.d.f. as in (1), given $\Theta = \theta$, and Θ is a random variable having a (prior) density function ξ with compact support in $(0, \infty)$. Lemma 1 below states that if ξ is continuously differentiable on its compact support, then

$$\hat{\delta}_n \equiv E^\xi\left\{ \delta(\Theta) \,\middle|\, X_1, \ldots, X_n \right\} = \overline{Y}_n + \frac{1}{n} E^\xi\left\{ \frac{\xi'(\Theta)}{\xi(\Theta)} \mid X_1, \cdots, X_n \right\},$$

where (3)

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{with} \quad Y_i = \ln X_i,\ i = 1, \ldots, n.$$

Let E_θ denote conditional expectation, given $\Theta = \theta$. Since $Y_1, \ldots, Y_n$ are conditionally independent with common distribution the Exponential distribution with parameter θ, the risk incurred by estimating δ by (3) under the loss (2) is

$$R_a(n,\theta) = E_\theta\left[L_a(\theta)\right] = a^2 E_\theta\left[\left(\overline{Y}_n - \delta(\theta)\right)^2\right] + n + o\left(\frac{1}{n}\right) = \frac{a^2[\delta(\theta)]^2}{n} + n + o\left(\frac{1}{n}\right)$$

as $n \to \infty$. The approximate risk is minimized by choosing n adjacent to $n_a = a\delta(\theta)$. The minimum risk is approximately

$$R_a^* = R_a(n_a, \theta) = 2a\delta(\theta)$$

for large $a > 0$. Since n_a depends on the unknown value of θ, there is no fixed-sample-size procedure that attains the minimum risk R_a^* in practice. Therefore, we propose to use the sequential procedure $(t, \hat{\delta}_t)$ which stops the sampling process after observing $Y_1, \ldots, Y_t$ and estimates $\delta(\theta)$ by $\hat{\delta}_t$, where

$$t = \inf\left\{ n \geq m : n \geq a\hat{\delta}_n \right\} \tag{4}$$

and $m \geq 2$ being an initial sample size. The performance of the procedure $(t, \hat{\delta}_t)$ is measured by the Bayes regret, which is defined as

$$\bar{r}(a) = \int_0^\infty r_a(\theta)\xi(\theta)d\theta,$$

for $a > 0$, where

$$r_a(\theta) = E_\theta[L_a(t,\theta)] - R_a^* = E_\theta[a^2\left[\hat{\delta}_t - \delta(\theta)\right]^2 + t] - 2a\delta(\theta)$$

is the regret incurred by the procedure $(t, \hat{\delta}_t)$. The Bayes regret can be rewritten as

$$\bar{r}(a) = E^\xi\left[a^2\left(\hat{\delta}_t - \delta(\Theta)\right)^2 + t - 2a\delta(\Theta)\right] \tag{5}$$

for $a > 0$.

Bayesian sequential estimation problems were studied by Bickel and Yahav (1969), Alvo (1977), Rasmussen (1980), Shapiro and Wardrop (1980), Woodroofe (1981, 1985), Tahir (1989), Woodroofe and Hardwick (1990), among others.

In Section 2, preliminary results for the analysis of the Bayes regret are obtained. The main result is presented in Section 3. It provides an asymptotic expansion for the Bayes regret. The proposed procedure can be used to estimate the mean loss for insurance policyholders or the mean insured value of homes in an optimal fashion. It specifies how many insurance policyholders or homes should be selected and provides an estimate of the population mean, based on this number. Note that the estimate of the population mean is

$$\hat{\mu}_t = \frac{1}{1 - \hat{\delta}_t}$$

where t is given by (4).

2. Preliminary Results

Lemma 1

Let Ω denote the support of ξ and let $\overline{Y}_n$ be as in (3). If ξ is continuously differentiable on Ω, then

$$\hat{\delta}_n = \overline{Y}_n + \frac{1}{n} E^\xi\left\{ \frac{\xi'(\Theta)}{\xi(\Theta)} \mid X_1, \cdots, X_n \right\}.$$

Proof:

Let $x_1 > 0, \ldots, x_n > 0$ denote the observed values of $X_1, , \ldots, X_n$, respectively. Then, the likelihood function is

$$g_n(\theta) = \theta^n e^{-n(\theta+1)\bar{y}_n} = \exp\{n\ln\theta - n\theta\bar{y}_n - n\bar{y}_n\} \text{ for } \theta > 0,$$

where $\bar{y}_n = n^{-1}\sum_{i=1}^{n} \ln x_i$. Next, let $c_n = \int_\Omega g_n(\theta)\xi(\theta)d\theta$. Then,

$$E^\xi\left\{ \delta(\Theta) \mid X_1 = x_1, \cdots, X_n = x_n \right\} = \frac{1}{c_n}\int_\Omega \frac{1}{\theta} g_n(\theta)\xi(\theta)d\theta = \frac{1}{c_n}\int_\Omega \frac{1}{n} g_n'(\theta)\xi(\theta)d\theta + \bar{y}_n$$

$$= -\frac{1}{nc_n}\int_\Omega g_n(\theta)\xi'(\theta)d\theta + \bar{y}_n = \frac{1}{nc_n}\int_\Omega g_n(\theta)\frac{\xi'(\theta)}{\xi(\theta)}\xi(\theta)d\theta + \bar{y}_n$$

$$= \frac{1}{n}E^\xi\left\{\frac{\xi'(\Theta)}{\xi(\Theta)} \mid X_1 = x_1, \cdots, X_n = x_n\right\} + \bar{y}_n$$

using integration by parts. The lemma follows.

Lemma 2

Let Ω denote the support of ξ and let $\overline{Y}_n$ be as in (3). If ξ is twice continuously differentiable on Ω, then

$$E^\xi\left\{ \left(\delta(\Theta) - \overline{Y}_n\right)^2 \middle| X_1, \ldots, X_n\right\} = \frac{1}{n}E^\xi\left\{ \left[\delta(\Theta)\right]^2 \mid X_1, \ldots, X_n\right\} + \frac{1}{n^2}E^\xi\left\{\frac{\xi''(\theta)}{\xi(\theta)} \mid X_1, \ldots, X_n\right\}$$

for $n \geq 1$.

Proof:

Let $x_1 > 0, \ldots, x_n > 0$ denote the observed values of $X_1, , \ldots, X_n$, respectively and let

$$g_n(\theta) = \exp\{n\ln\theta - n\theta\bar{y}_n - n\bar{y}_n\}$$

denote the likelihood function. Also, let $c_n = \int_\Omega g_n(\theta)\xi(\theta)d\theta$. Then,

$$E^\xi\left\{ \left[\delta(\Theta) - \overline{Y}_n\right]^2 \middle| X_1 = x_1, \cdots, X_n = x_n\right\} = \frac{1}{c_n}\int_\Omega \left(\frac{1}{\theta} - \bar{y}_n\right)^2 g_n(\theta)\xi(\theta)d\theta$$

$$= \frac{1}{nc_n}\int_\Omega \left(\frac{1}{\theta} - \bar{y}_n\right) g_n'(\theta)\xi(\theta)d\theta = -\frac{1}{nc_n}\int_\Omega g_n(\theta)\frac{d}{d\theta}\left[\left(\frac{1}{\theta} - \bar{y}_n\right)\xi(\theta)\right]d\theta$$

$$= \frac{1}{nc_n}\int_\Omega \frac{1}{\theta^2} g_n(\theta)\xi(\theta)d\theta - \frac{1}{nc_n}\int_\Omega \left(\frac{1}{\theta} - \bar{y}_n\right) g_n(\theta)\xi'(\theta)d\theta$$

$$= \frac{1}{nc_n}\int_\Omega \frac{1}{\theta^2} g_n(\theta)\xi(\theta)d\theta - \frac{1}{n^2c_n}\int_\Omega g_n'(\theta)\xi'(\theta)d\theta$$

$$= \frac{1}{nc_n}\int_\Omega \frac{1}{\theta^2} g_n(\theta)\xi(\theta)d\theta + \frac{1}{n^2c_n}\int_\Omega g_n(\theta)\xi''(\theta)d\theta$$

$$= \frac{1}{n}E^\xi\left\{ \left[\delta(\Theta)\right]^2 \mid X_1 = x_1, \ldots, X_n = x_n\right\} + \frac{1}{n^2}E^\xi\left\{\frac{\xi''(\Theta)}{\xi(\Theta)} \mid X_1 = x_1, \cdots, X_n = x_n\right\}$$

using integration by parts. The lemma follows.

Let $\mathfrak{D}_t$ denote the sigma-algebra generated by $X_1, \ldots, X_t$. Lemma 1 and Lemma 2 imply that the Bayes regret in (5) becomes

$$\bar{r}(a) = E^\xi\left[\frac{a^2}{t}\left(U_t - \hat{\delta}_t^2\right) + \frac{a^2}{t}\left(\frac{t}{a} - \hat{\delta}_t\right)^2 + \frac{a^2}{t^2}V_t\right] \tag{6}$$

for any $a > 0$, where $U_t = E^\xi\{[\delta(\Theta)]^2 | \mathfrak{D}_t\}$ and $V_t = E^\xi\left\{\frac{\xi''(\Theta)}{\xi(\Theta)}\middle| \mathfrak{D}_t\right\}$

Lemma 3

Let t be defined by (4) with $m \geq 2$. Then,

(i) there exists $\delta_0 > 0$ such that $t \geq a\delta_0$ w.p.1 (P^ξ) for any $a > 0$ and

(ii) $\frac{t}{a} \to \delta(\Theta)$ w.p.1 (P^{ξ}) as $a \to \infty$.

Proof:

Let $[\theta_0, \theta_1]$ denote the support of ζ. Then,

$$t \geq a\hat{\delta}_t \geq a\frac{1}{\theta_1} = a\delta_0$$

by definition of t and $\hat{\delta}_t$. To establish Assertion (ii), observe that $a\hat{\delta}_t \leq t \leq a\hat{\delta}_t + m - 1$for $a > 0$, by definition of t. It follows that

$$\delta(\Theta) \leq \liminf_{a\to\infty} \frac{t}{a} \leq \limsup_{a\to\infty} \frac{t}{a} \leq \delta(\Theta) \qquad \text{w.p.1 } (P^{\xi})$$

since the sequence $\hat{\delta}_n$, $n \geq 1$, is a uniformly integrable martingale such that $\hat{\delta}_n \to \delta(\Theta)$ w.p.1 (P^{ξ}) as $n \to \infty$ and $t \to \infty$ w.p.1 as $a \to \infty$.

3. The Main Result

The Bayes regret in (6) can be rewritten as

$$\bar{r}(a) = E^{\xi}\left[\frac{a^2}{t}W_t + \frac{a^2}{t}\left(\frac{t}{a} - \hat{\delta}_t\right)^2 + \frac{a^2}{t^2}V_t\right] \tag{7}$$

for $a > 0$, where $W_t = Var\{\delta(\Theta)|\mathfrak{D}_t\}$.

Theorem

Let t be defined by (4) with $m \geq 2$ and let $\bar{r}(a)$ be as in (7). If ξ is continuously differentiable on its compact support, then

$$\bar{r}(a) = 1 + E^{\xi}\left[\frac{\xi''(\Theta)}{\delta^2(\Theta)\xi(\Theta)}\right] + o(1)$$

as $a \to \infty$.

The proof of the theorem requires Lemmas 4-6 below.

Lemma 4

Let t be defined by (4) with $m \geq 2$. Then

$$E^{\xi}\left[\frac{a^2}{t}\left(\frac{t}{a} - \hat{\delta}_t\right)^2\right] = o(1)$$

as $a \to \infty$.

Proof:

By definition of t, $a\hat{\delta}_t \leq t \leq a\hat{\delta}_t + m - 1$; so that

$$0 \leq \frac{t}{a} - \hat{\delta}_t \leq \frac{m-1}{a}$$

which implies that

$$\boldsymbol{\frac{a^2}{t}\left(\frac{t}{a} - \hat{\delta}_t\right)^2 \leq \frac{2(m-1)^2}{t} \leq \frac{2(m-1)^2}{a\delta_0}}$$

for some number $\boldsymbol{\delta_0 > 0}$, by the first assertion of Lemma 3. Thus,

$$E^{\xi}\left[\frac{a^2}{t}\left(\frac{t}{a} - \hat{\delta}_t\right)^2\right] \leq \frac{2(m-1)^2}{a\delta_0} \to 0$$

as $a \to \infty$.

Lemma 5

Let t be defined by (4) with $m \geq 2$. If ξ is continuously differentiable on its support, then $aW_t \to \delta(\Theta)$ w.p.1 (P^{ξ}) as a

$\to\infty$.

Proof:

Let $x_1, \ldots, x_n$ denote the observed values of $X_1, \ldots, X_n$, respectively and let $w_n = Var\{\delta(\Theta)|X_1 = x_1, \ldots, X_n = x_n\}$. Also, let $\hat{\theta}_n$ denote the maximum likelihood estimate of θ, based on $x_1, \ldots, x_n$. Then, $\delta(\hat{\theta}_n) = \bar{y}_n$ and

$$w_n \le E^{\xi}\left\{ \left[\delta(\Theta) - \bar{y}_n\right]^2 \middle| X_1 = x_1, \cdots, X_n = x_n \right\} = \frac{1}{n} E^{\xi}\left\{ \left[\delta(\Theta)\right]^2 \mid X_1 = x_1, \ldots, X_n = x_n \right\} + \frac{1}{n^2} E^{\xi}\left\{ \frac{\xi''(\Theta)}{\xi(\Theta)} \mid X_1 = x_1, \cdots, X_n = x_n \right\}$$

by Lemma 2. This implies that

$$aW_t \le \frac{a}{t}U_t + \frac{a}{t^2}V_t = \frac{a}{t}U_t + \left(\frac{1}{a}\right)\frac{a^2}{t^2}V_t$$

for any $a > 0$. Thus,

$$\limsup_{a\to\infty} aW_t \le \limsup_{a\to\infty} \frac{a}{t}U_t = \frac{1}{\delta(\Theta)}\left[\delta(\Theta)\right]^2 = \delta(\Theta) \text{ w.p.1 } (P^{\xi}) \text{ as } a\to\infty,$$

by the second assertion of Lemma 3 and the facts that U_n and V_n are martingales such that $U_n \to [\delta(\Theta)]^2$ and $V_n \to \xi''(\Theta)/\xi(\Theta)$ w.p.1 (P^{ξ}) as $n\to\infty$. Next, a Taylor's expansion for $\delta(\Theta)$ about $\Theta = \hat{\theta}_t$ yields

$$\delta(\Theta) = \delta(\hat{\theta}_t) + \delta'(\theta_t^*)(\Theta - \hat{\theta}_t) = \bar{Y}_t + \delta'(\theta_t^*)(\Theta - \hat{\theta}_t),$$

where θ_t^* is a random variable between Θ and $\hat{\theta}_t$. Thus,

$$\begin{aligned}\liminf_{a\to\infty} aW_t &= \liminf_{a\to\infty} aVar\left\{\delta'(\theta_t^*)(\Theta - \hat{\theta}_t)\,|X_1,\cdots,X_t\right\} \\ &= \liminf_{a\to\infty} \frac{a}{t}\frac{1}{[\delta(\hat{\theta}_t)]^2} Var\left\{\delta'(\theta_t^*)\sqrt{t}\delta(\hat{\theta}_t)(\Theta - \hat{\theta}_t)\,|X_1,\cdots,X_t\right\} \\ &\ge \frac{[\delta'(\Theta)]^2}{[\delta(\Theta)]^3} = \delta(\Theta)\end{aligned}$$

w.p.1 (P^{ξ}) as $a\to\infty$, by Fatou's Lemma, the second assertion of Lemma 3, the facts that $\delta(\hat{\theta}_t) \to \delta(\Theta)$ and $\delta'(\theta_t^*) \to \delta(\Theta)$ w.p.1 (P^{ξ}) as $a\to\infty$ and the fact that the posterior distribution of $\sqrt{n}\delta(\hat{\theta}_n)(\Theta - \hat{\theta}_n)$ converges to the Standard Normal distribution as $n\to\infty$ (see Bickel and Yahav (1969)).

Lemma 6

Let t be defined by (4) with $m \ge 2$ and let W_t be as in (7). If ξ is continuously differentiable on its compact support, then

$$\lim_{a\to\infty} E^{\xi}\left[\frac{a^2}{t}W_t\right] = 1.$$

Proof:

$\frac{t}{a} \to \delta(\Theta)$ w.p.1 (P^{ξ}) as $a\to\infty$, by the second assertion of Lemma 3 and $aW_t \to \delta(\Theta)$ w.p.1 (P^{ξ}) as $a\to\infty$, by Lemma 5. Thus, $\frac{a^2}{t}W_t \to 1$ w.p.1 (P^{ξ}) as $a\to\infty$. Moreover,

$$aW_t \le \frac{a}{t}U_t + \frac{a}{t^2}V_t = \frac{a}{t}\left(U_t + V_t\right)$$

as in the proof of Lemma 5 and this implies that

$$\frac{a^2}{t}W_t \le_t = \left(\frac{a}{t}\right)^2\left(U_t + V_t\right) \le \frac{1}{\delta_0^2}\left(U_t + V_t\right)$$

For any $a > 0$, by the first assertion of Lemma 3. It follows that, $\frac{a^2}{t} W_t, a > 0,$ are uniformly integrable since U_t and V_t are uniformly integrable martingales. The lemma follows.

4. Proof of the Theorem

The theorem follows by taking the limit as $a \to \infty$ in

$$\bar{r}(a) = E^{\xi}\left[\frac{a^2}{t} W_t\right] + E\left[\frac{a^2}{t}\left(\frac{t}{a} - \hat{\delta}_t\right)^2\right] + E\left[\frac{a^2}{t^2} V_t\right]$$

and using Lemma 4, Lemma 6 and the fact that

$$E\left[\frac{a^2}{t^2} V_t\right] \to E\left[\frac{1}{\delta^2(\Theta)} \frac{\xi''(\Theta)}{\xi(\Theta)}\right]$$

as $a \to \infty$ since $\frac{t}{a} \to \delta(\Theta)$ w.p.1 (P^{ξ}) as $a \to \infty$, $V_t \to \frac{\xi''(\Theta)}{\xi(\Theta)}$ w.p.1 (P^{ξ}) as $a \to \infty$ and $\frac{a^2}{t^2} V_t, a > 0,$ are uniformly integrable ($\frac{a^2}{t^2} \leq \frac{1}{\delta_0^2}$ and V_t, $a > 0$, is a uniformly integrable martingale).

5. Conclusions

We have proposed a Bayesian sequential procedure for estimating the inverse of the shape parameter of the type I Pareto distribution and provided a second-order asymptotic expansion for the regret incurred under the square error loss. The proposed procedure can be used to estimate the mean loss for insurance policyholders or the mean insured value of homes in an optimal fashion.

References

Alvo, M. (1977). Bayesian sequential estimation. *Ann. Statist., 5*, 955-968. http://dx.doi.org/10.1214/aos/1176343951

Bickel, P., & Yahav, J. (1969). Some Contributions to the Asymptotic Theory of Bayes Solutions. *Z. Wahrsch, 11*, 257-276. http://dx.doi.org/10.1007/BF00531650

Breiman, L. (1968). *Probability*. Addison-Wesley, New York.

Rasmussen, S. (1980). Bayesian sequential estimation. *Ann. Statist.*, 8, 1229-1234. http://dx.doi.org/10.1214/aos/1176345196

Rekab, K., & Tahir, M. (2004). An asymptotic second-order lower bound for the Bayes risk of a sequential procedure. *Sequential Analysis, 23*, 451-464. http://dx.doi.org/10.1081/SQA-200027055

Shapiro, C., & Wardrop, R. (1980). Bayesian sequential estimation for one-parameter exponential families. *J. Amer. Assoc., 75*, 984-988. http://dx.doi.org/10.1080/01621459.1980.10477583

Tahir, M. (1989). An Asymptotic Lower Bound for the Local Minimax Regret in Sequential Point Estimation. *Ann. Statist., 17*, 1335-1346. http://dx.doi.org/10.1214/aos/1176347273

Woodroofe, M. (1981). A.P.O. rules are asymptotically non-deficient for estimation with squared error loss. *Z Wahrsch, 58*, 331–341. http://dx.doi.org/10.1007/BF00542639

Woodroofe, M. (1985). Asymptotic Local Minimaxity In Sequential Point Estimation. *Ann. Statist., 13*, 676-688. http://dx.doi.org/10.1214/aos/1176349547

Woodroofe, M., & Hardwick, J. (1990). Sequential Allocation for Estimation problem with ethical costs. *Ann. Statist., 18*, 1358-1377. http://dx.doi.org/10.1214/aos/1176347754