Ann. Inst. Statist. Math.
36 (1984), Part A, 1-9

A PRELIMINARY TEST PROCEDURE FOR THE SCALE PARAMETER OF EXPONENTIAL DISTRIBUTION WHEN THE SELECTION PARAMETER IS UNKNOWN*

KATUOMI HIRANO

(Received June 29, 1983; revised Nov. 28, 1983)

Summary

A preliminary test estimator is considered for the scale parameter of the two-parameter exponential distribution with unknown selection parameter, where the distribution does not satisfy the regularity condition of Wilks' theorem—the density is not differentiable. A method of specifying the level of significance of the preliminary test based on is proposed AIC.

1. Introduction

Let $X_1, X_2, \cdots, X_r$ be the first r ordered observations out of a sample of size n $(3 \leqq r \leqq n)$ from the two-parameter exponential distribution with the probability density function

$$f(x;\eta,\theta)=\frac{1}{\theta}\exp\left\{-\frac{x-\eta}{\theta}\right\}, \quad x\geqq\eta, \quad -\infty<\eta<\infty, \quad \theta>0$$

(abbr. as EP(η,θ)). We are interested in estimating the scale parameter θ. If we assume that η is known, the unbiased estimator is $T_r(\eta)/r$ where

$$T_r(\eta)=\sum_{i=1}^{r}(X_i-\eta)+(n-r)(X_r-\eta),$$

and the minimum mean squared error estimator, among the class of estimators of the form $a(r)T_r(\eta)$, is $T_r(\eta)/(r+1)$. Similarly, if η is unknown, the unbiased estimator is $S_r/(r-1)$ where

* This work was partly supported by Scientific Research Fund No. 58450058 from the Ministry of Education of Japan.
AMS 1980 subject classifications: Primary 62F10 62F03; Secondary 62C99.
Key words: Two-parameter exponential distribution, testimator, significance level, AIC.

$$S_r=\sum_{i=1}^{r}(X_i-X_1)+(n-r)(X_r-X_1)\,,$$

and the minimum mean squared error estimator, among the class of estimators of the form $b(r)S_r$, is S_r/r. We consider only two classes of estimators $\{a(r)T_r(\eta)\}$ and $\{b(r)S_r\}$.

The problem considered here is similar to the one considered by Davis and Arnold [4] in that the scale parameter θ is to be estimated in the presence of the nuisance parameter η. We assume that a point estimate, η_0 $(\eta\geqq\eta_0)$, of the selection parameter η is available in advance and, the preliminary test will be conducted on this point estimate. The preliminary test estimator (abbr. as testimator) is then defined as follows:

$$\hat{\theta}_{PT}=\begin{cases} a(r)T_r(\eta_0) & \text{if } \eta=\eta_0 \text{ is accepted}\,, \\ b(r)S_r & \text{if } \eta>\eta_0 \text{ is accepted}\,. \end{cases}$$

The testimator $\hat{\theta}_{PT}$ always depends on the significance level α for testing the null hypothesis $H\colon \eta=\eta_0$ against the alternative hypothesis $K\colon \eta>\eta_0$. Thus one of the main problems in this paper is to specify the necessary α.

In theory the choice of the level of significance α is arbitrary. It has become customary to choose for α one of the standard values such as 0.01, 0.05 or 0.1. However when we consider the procedure of preliminary test estimation as an estimation procedure, the choice of α can not be arbitrary, because the testimator always depends on α and we can not uniquely determine it. Thus we should find the optimal α in terms of a criterion which is reasonable for the problem. Here we shall adopt the method of minimizing $\mathrm{AIC}=-2\log L(\hat{\boldsymbol{\theta}})+2k$, where $L(\boldsymbol{\theta})$ denotes the likelihood, and the unknown parameter $\boldsymbol{\theta}$ is k-dimensional. For the detail of the method, see Akaike [1]. In Section 3, the risks of $\hat{\theta}_{PT}$ with some standard values of α are compared with the ones of $\hat{\theta}_{PT}$ with α decided by the minimum AIC procedure.

Next, we discuss the loss function used in this paper. The squared error loss function has been traditionally used in the estimation theory. But it is not always adequate one. In particular, for the scale problem it is out of balance in the sense that the maximum loss for low values of θ is finite, while the maximum loss for large values of θ is infinite. We think that the following loss functions, which are all scale invariant, are intuitively reasonable for the scale parameter $\theta>0$.

$$\frac{\hat{\theta}}{\theta}+\frac{\theta}{\hat{\theta}}-2 \quad \text{(Wasan [12])}\,, \qquad \max\left\{\frac{\hat{\theta}}{\theta}-1,\ \frac{\theta}{\hat{\theta}}-1\right\} \quad \text{(Hirano [5])}\,,$$

$$\left(\log\frac{\hat{\theta}}{\theta}\right)^2 \quad \text{and} \quad \frac{\hat{\theta}}{\theta}-1-\log\frac{\hat{\theta}}{\theta} \quad \text{(Brown [3])}.$$

In this paper we are going to use the first one.

2. $\hat{\theta}_{PT}$ and AIC

Suppose we are given an estimate, η_0 ($\eta \geqq \eta_0$), of the selection parameter η. For testing $H: \eta=\eta_0$ against $K: \eta>\eta_0$, the likelihood ratio test statistic L is

$$L=2r\log\left\{1+\frac{F(2, 2r-2)}{r-1}\right\}$$

where $F(2, 2r-2)=(r-1)n(X_1-\eta_0)/S_r$, and under H, $F(2, 2r-2)$ follows the F-distribution with $(2, 2r-2)$ degrees of freedom. Then we obtain a critical region for testing H against K:

$$L\geqq c_\alpha \quad \text{where} \quad c_\alpha=\frac{2r}{r-1}\log\frac{1}{\alpha},$$

and α ($0<\alpha\leqq 1$) denotes the level of significance of the test. Then the testimator $\hat{\theta}_{PT}$ can be written as

$$\hat{\theta}_{PT}=\begin{cases} a(r)T_r(\eta_0) & \text{if } 0<F(2, 2r-2)<(r-1)\left\{\left(\dfrac{1}{\alpha}\right)^{1/(r-1)}-1\right\} \\ b(r)S_r & \text{otherwise}. \end{cases}$$

On the other hand, we can view $\hat{\theta}_{PT}$ as a statistical procedure in which one of the two models (distributions) EP(η_0, θ) and EP(η, θ) is selected first and then the parameter θ is estimated assuming the sample is drawn from the selected distribution. To select the model we use AIC.

The derivation of AIC is based on the fact that the log-likelihood ratio statistic will asymptotically be distributed as a chi-square variable with degrees of freedom k (Wilks' theorem). See Akaike [1]. In our problem the two-parameter exponential distribution does not satisfy the regularity condition of the theorem (see, for example, Wilks [14]). However we have

$$\Pr\{L\leqq x\}=1-\exp\{-(r-1)x/(2r)\} \quad \text{under } H,$$

that is the statistic L follows EP$(0, 2r/(r-1))$, or $(r-1)L/r$ exactly follows the chi-square distribution with 2 degrees of freedom. Thus we can use AIC also in this case and AIC statistic is given by AIC$=-2\log$(the maximum likelihood)$+2kr/(r-1)$ where k is the number of parameters estimated in the model. Note that these considerations are not asymptotic, but exact. We specify the value of α for testing H against K on the basis of AIC. The AIC's under H and K are given by

$$\text{AIC}\,(H)=c(n,\,r)+2r\log T_r(\eta_0)/r+2r/(r-1)$$

and

$$\text{AIC}\,(K)=c(n,\,r)+2r\log S_r/r+4r/(r-1)$$

respectively, where $c(n,\,r)=-2\log n!/(n-r)!+2r$. Thus the following relation holds;

$$\text{AIC}\,(H)-\text{AIC}\,(K)<0 \iff L<2r/(r-1)\,.$$

Under H we have

$$1-\alpha=\Pr\{\text{AIC}\,(H)-\text{AIC}\,(K)<0\}=1-e^{-1}\,,$$

that is, $\alpha=e^{-1}=0.3678\cdots$.

3. Risk of $\hat{\theta}_{PT}$

To evaluate the risk $R(\hat{\theta}_{PT})$ of the testimator $\hat{\theta}_{PT}$, based on the loss function $l(\theta,\,\hat{\theta})=\hat{\theta}/\theta+\theta/\hat{\theta}-2$, we use the well-known results that $2S_r/\theta$ and $2n(X_1-\eta)/\theta$ are independently distributed as $\chi^2(2r-2)$ and $\chi^2(2)$, respectively.

Now, we obtain the risk of $\hat{\theta}_{PT}$ as

$$\begin{aligned}
R(\hat{\theta}_{PT})&=\mathrm{E}\,(l(\theta,\,\hat{\theta}_{PT}))\\
&=\iint_{0<y<c}\left\{\frac{a(r)}{2}x+\frac{2}{a(r)x}\right\}f(x,\,y)dxdy\\
&\quad+\iint_{y\geqq c}\left\{\frac{b(r)}{2}\,\frac{x}{1+y}+\frac{2}{b(r)}\,\frac{1+y}{x}\right\}f(x,\,y)dxdy\\
&=ra(r)\left\{1-\Gamma\left(r+1,\,\frac{\xi}{2c}\right)\right\}+\frac{\xi a(r)}{2}\left\{1-\Gamma\left(r,\,\frac{\xi}{2c}\right)\right\}\\
&\quad-ra(r)(1+c)^{1-r}e^{\xi/2}\left\{1-\Gamma\left(r+1,\,\frac{\xi(1+c)}{2c}\right)\right\}\\
&\quad+\frac{e^{\xi/2}}{a(r)\Gamma(r)}\left(-\frac{\xi}{2}\right)^{r-1}\int_{\xi(1+c)/(2c)}^{\infty}t^{-1}e^{-t}dt\\
&\quad+\frac{e^{\xi/2}}{a(r)\Gamma(r)}\sum_{k=1}^{r-1}{}_{r-1}C_k\left(-\frac{\xi}{2}\right)^{r-1-k}\Gamma(k)\left\{1-\Gamma\left(k,\,\frac{\xi(1+c)}{2c}\right)\right\}\\
&\quad-\frac{e^{\xi/2}(1+c)^{1-r}}{(r-1)a(r)}\left\{1-\Gamma\left(r-1,\,\frac{\xi(1+c)}{2c}\right)\right\}+(r-1)b(r)\Gamma\left(r+1,\,\frac{\xi}{2c}\right)
\end{aligned}$$

$$+(r-1)b(r)e^{\xi/2}(1+c)^{-r}\left\{1-\Gamma\left(r+1,\ \frac{\xi(1+c)}{2c}\right)\right\}$$
$$+\frac{1}{(r-2)b(r)}\Gamma\left(r-1,\ \frac{\xi}{2c}\right)+\frac{e^{\xi/2}(1+c)^{2-r}}{(r-2)b(r)}\left\{1-\Gamma\left(r-1,\ \frac{\xi(1+c)}{2c}\right)\right\}-2$$

where

$$f(x,y)=\frac{1}{\Gamma(r-1)}\left(\frac{1}{2}\right)^r\frac{x^{r-1}}{(y+1)^r}\exp\left\{-\frac{x}{2}+\frac{\xi}{2}\right\},\quad x>\xi\,,\quad y>\frac{\xi}{x-\xi}>0\,,$$

$$\Gamma(m,x)=\frac{1}{\Gamma(m)}\int_0^x y^{m-1}e^{-y}dy\,,\qquad c=(1/\alpha)^{1/(r-1)}-1$$

and

$$\xi=2n(\eta-\eta_0)/\theta>0\,.$$

The derivations are omitted for brevity.

Next, the risks of $a(r)T_r(\eta)$ and $b(r)S_r$ are

$$R(a(r)T_r(\eta))=ra(r)+1/\{(r-1)a(r)\}-2$$
$$\geqq 2(\sqrt{r/(r-1)}-1)\,,$$

and

$$R(b(r)S_r)=(r-1)b(r)+1/\{(r-2)b(r)\}-2$$
$$\geqq 2(\sqrt{(r-1)/(r-2)}-1)$$

respectively, and the equalities hold in the two inequalities if and only if $a^*(r)=1/\sqrt{r(r-1)}$ and $b^*(r)=1/\sqrt{(r-1)(r-2)}$, respectively. We are interested in $\hat{\theta}^*_{PT}$ defined by replacing $a(r)$ and $b(r)$ in the definition of $\hat{\theta}_{PT}$ by $a^*(r)$ and $b^*(r)$, respectively.

Figs. 1–4 give efficiencies, Eff (α), of $\hat{\theta}^*_{PT}$ relative to $b^*(r)S_r$ where the efficiency is defined as

$$\text{Eff}\,(\alpha)=\frac{R(b^*(r)S_r)}{R(\hat{\theta}^*_{PT})}\,,$$

for $n=10$, $r=5$, for $n=20$, $r=5, 10, 15$, for selected values of $(\eta-\eta_0)/\theta=\delta$ and for α $(0\leqq\alpha\leqq 1)$.

Our specification based on the minimum AIC procedure for the selection of models is $\alpha=e^{-1}$, given in Section 2. From Figures this choice seems to be with a minimax type optimality.

Remark. Note that we can extend the definition of $\hat{\theta}_{PT}$ for $\alpha=0$

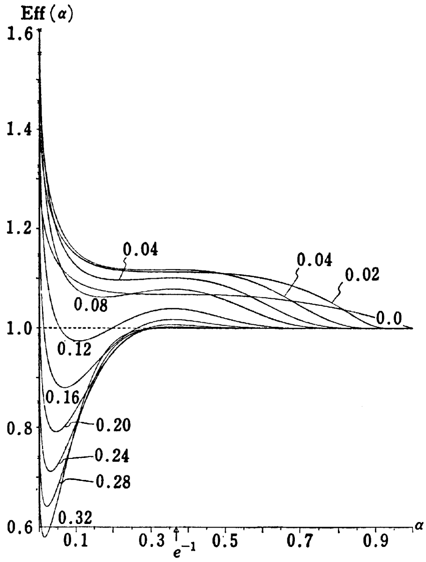

Fig. 1. Efficiency of $\hat{\theta}^*_{PT}$; $n=10$, $r=5$
$\delta=(\eta-\eta_0)/\theta=0.0$, 0.02(0.02)
0.08(0.04)0.32

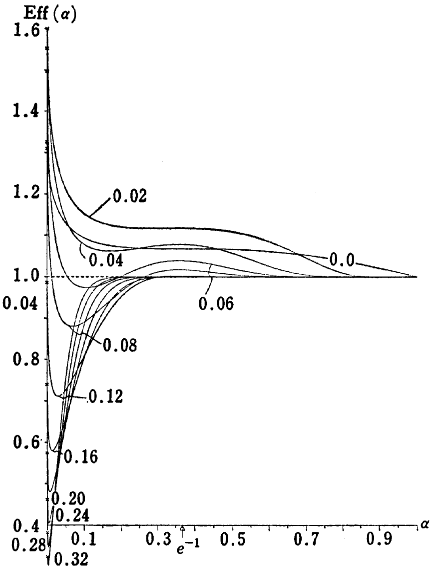

Fig. 2. Efficiency of $\hat{\theta}^*_{PT}$; $n=20$, $r=5$
$\delta=(\eta-\eta_0)/\theta=0.0$, 0.02(0.02)
0.08(0.04)0.32

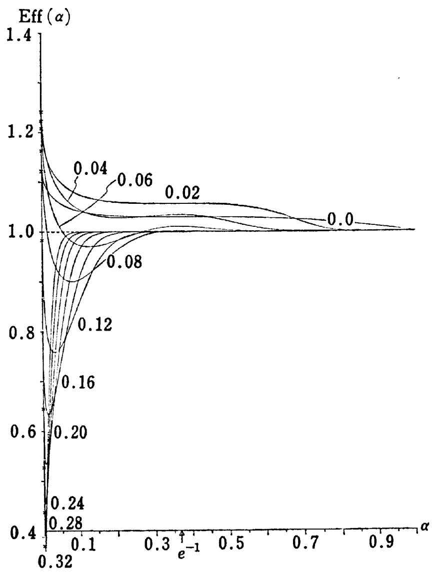

Fig. 3. Efficiency of $\hat{\theta}^*_{PT}$; $n=20$, $r=10$
$\delta=(\eta-\eta_0)/\theta=0.0$, 0.02(0.02)
0.08(0.04)0.32

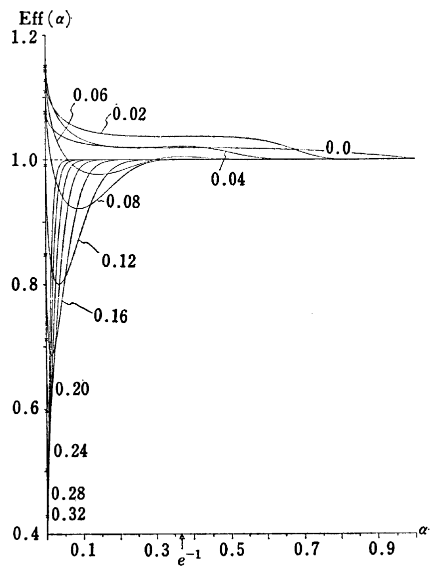

Fig. 4. Efficiency of $\hat{\theta}^*_{PT}$; $n=20$, $r=15$
$\delta=(\eta-\eta_0)/\theta=0.0$, 0.02(0.02)
0.08(0.04)0.32

by putting $\hat{\theta}_{PT}=a(r)T_r(\eta_0)$. In this case

$$R(a(r)T_r(\eta_0))=\lim_{\alpha\to 0} R(\hat{\theta}_{PT})$$
$$=a(r)(r+\xi/2)+\frac{e^{\xi/2}}{a(r)\Gamma(r)}\left(-\frac{\xi}{2}\right)^{r-1}\int_{\xi/2}^{\infty} t^{-1}e^{-t}dt$$
$$+\frac{e^{\xi/2}}{a(r)(r-1)}\left\{1-\Gamma\left(r-1, \frac{\xi}{2}\right)\right\}$$
$$+\frac{e^{\xi/2}}{a(r)}\sum_{i=1}^{r-2}\left(-\frac{\xi}{2}\right)^{i}\frac{1}{i!(r-1-i)}\left\{1-\Gamma\left(r-1-i, \frac{\xi}{2}\right)\right\}-2$$
$$=R(\hat{\theta}_{PT}, 0)\,, \quad \text{say}\,.$$

We also obtain

$$R(\hat{\theta}_{PT}, 0)=R(a(r)T_r(\eta)) \qquad \text{for } \xi=0\,,$$

and

$$R(\hat{\theta}_{PT})=R(b(r)S_r) \qquad \text{for } \alpha=1\,.$$

4. Conclusions

First, look at the Figs., and try to specify the optimal choice of α. We will never wish to choose for α one of the standard values such as 0.01, 0.05 and 0.1, because these values can give extremely low efficiency for some values of δ. On the other hand, as we saw in Section 2, the minimum AIC chose $\alpha=e^{-1}$. It can be seen from Figs. that this choice will never give extremely low efficiencies for all δ. This result is warning us that when we consider the procedure of preliminary test estimation as an estimation procedure, we should pay much attention to the choice of the level of significance of the preliminary test.

What is the optimal α? It may be difficult to find α that minimizes the risk of the maximum regret. For example, the method of Ohtani and Toyoda [11] cannot be applied to the present problem. We should try to use some other criterion. The method of minimizing AIC is one possibility and the numerical results reported in the Figs. show that using AIC for the selection of α provides an effective method for the problem.

5. Discussion

For the model EP(η, θ) there is the problem of determining the estimator of θ. For example, for the model EP(η_0, θ) with $\eta=\eta_0$ known, the estimator minimizing the risk is $a^*(r)T_r(\eta)$ and for EP(η, θ), it is $b^*(r)S_r$. In view of the fact that the m.l.e. is $T_r(\eta_0)/r$ for EP(η_0, θ), and is S_r/r for $\mathrm{EP}(\eta, \theta)$, we think the problem of choosing between the models $\mathrm{EP}(\eta_0, \theta)$ and $\mathrm{EP}(\eta, \theta)$ based on AIC, that is determining α, and the problem of determing the estimator that minimizes the risk associated with the loss function should be quite distinct from each other.

Many statisticians considered the use of a preliminary test for the situation in which one has two samples, each providing an estimate of a common unknown parameter. On the basis of the preliminary test, the decision is made whether or not to pool the two samples. Since the decision always depends on the level of significance of the preliminary test, the problem of specifying it has been discussed. Toyoda and Wallace [12] and Ohtani and Toyoda [11] discussed two-sample preliminary test estimation for the variance of the normal distribution. Since all of them satisfy the regularity condition of Wilks' theorem, we can use the minimum AIC procedure to determine the necessary level of significance. See Hirano [8] [9] [10]. Similarly, for the mean of the normal distribution, see Hirano [7].

Davis and Arnold [4], Hirano [6] and Bhattacharya and Srivastava [2] discussed the use of the preliminary tests for the situation in which only one sample is taken, but more than one estimator is considered. These are also regular cases in the above sense. The method stated in this paper is applicable to these and many other situations. For example, for the problems by Davis and Arnold, and Bhattacharya and Srivastava we can easily conclude that reasonable choices of the necessary significance levels are all about $0.15\cdots$ for two-sided preliminary tests. For the problem discussed by Davis and Arnold this result is close to their specification. Bhattacharya and Srivastava do not deal with the problem of the specification of α.

Acknowledgement

The author wishes to thank the referee for careful reading of an earlier draft of this paper.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory*, (eds. B. N. Petrov and F. Csáki), Akademiai Kiado, Budapest, 267-281.

[2] Bhattacharya, S. K. and Srivastava, V. K. (1974). A preliminary test procedure in life testing, *J. Amer. Statist. Ass.*, **69**, 726-729.

[3] Brown, L. D. (1968). Inadmissibility of the usual estimators of scale parameters, *Ann. Math. Statist.*, **39**, 29-48.

[4] Davis, R. L. and Arnold, J. C. (1970). An efficient preliminary test estimator for the variance of a normal population when the mean is unknown, *Biometrika*, **57**, 647-677.

[5] Hirano, K. (1973). Some properties of an estimator for the variance of a normal distribution, *Ann. Inst. Statist. Math.*, **25**, 479-492.

[6] Hirano, K. (1977). Estimation procedures based on preliminary test, shrinkage technique and information criterion, *Ann. Inst. Statist. Math.*, **29** A, 21-34. (Correction, 29 A, 317)

[7] Hirano, K. (1978a). On level of significance of the preliminary test in pooling means, *Ann. Inst. Statist. Math.*, **30** A, 1-8.

[8] Hirano, K. (1978b). A note on the level of significance of the preliminary test in pooling variances, *J. Japan Statist. Soc.*, **8**, 71-75.

[9] Hirano, K. (1980a). On the test estimators for estimating variance of a normal distribution, *Recent Developments in Statistical Inference and Data Analysis*, (ed. Matusita, K.), North-Holland Publishing Company, 105-115.

[10] Hirano, K. (1980b). The significance points in preliminary test in estimating variance based on two samples (in Japanese), *Proc. Inst. Statist. Math.*, **27**, 23-34.

[11] Ohtani, K. and Toyoda, T. (1978). Minimax regret critical value for a preliminary test in pooling variances, *J. Japan Statist. Soc.*, **8**, 15-20.

[12] Toyoda, Y. and Wallace, T. D. (1975). Estimation variance after a preliminary test of homogeneity and optimal levels of significance for the pre-test, *J. of Econometrics*, **3**, 395-404.

[13] Wasan, M. T. (1970). *Parametric Estimation*, McGraw-Hill Book Company, New York.

[14] Wilks, S. S. (1962). *Mathematical Statistics*, John Wiley & Sons Inc., New York.