On the Bayesian Sequential Estimation Problem in k-Parameter Exponential Family

Byoung Chang Yoon* & Jae Joo Kim**

abstract

The Bayesian sequential estmiation problem for k parameters exponential families is considered using loss related to the Fisher information. Tractable expressions for the Bayes estimator and the posterior expected loss are found, and the myopic or one-step-ahead stopping rule is defined. Sufficient conditions are given for optimality of the myopic procedure, and the myopic procedure is shown to be asymptotically optimal in all cases considered.

1. INTRODUCTION

The Bayesian sequential estimation problem in one parameter exponential families is considered by many authors using a family of loss functions related to the Fisher information.

In 1967, Whittle and Lane [12] showed that with loss proportional to the information multiplied by squared error for parameters of exponential families, the Bayes sequential procedure (BSP) is a fixed sample size procedure.

In 1977, Cabilio [2] showed that using loss proportional to a second power of the information multiplied by squared error for a binomial success probability, he yielded a cost sequence in the monotone case thereby deriving the optimality of the myopic stopping rule.

*Department of Mathematics, Korea Military Academy.
**Department of Computer Science & Statistics, Seoul National University.

Also, in 1977, Shapiro and Wardrop [9] showed that using loss proportional to a power (between one and three) of the information multiplied by squared error for a Poisson process rate, they yielded costs and optimality of the myopic stopping rule in the monotone case.

In 1978, Shapiro and Wardrop [11] considered the Bayes sequential estimation problem for one parameter exponential families with loss functions generalizing those mentioned above.

In this paper, we consider the Bayes sequential estimation problem for k parameters exponential families with loss functions which are introduced by Shapiro and Wardrop [11].

In Section 2, some distribution theory is introduced in order to derive the form of the Bayes estimator and the posterior expected loss.

In Section 3, a myopic stopping rule is defined and sufficient conditions for the optimality of the myopic stopping rule are given.

In Section, 4 the myopic stopping rule is shown to be asymptotically optimal in the case that cost c approaches to zero. Also, the limiting ratio of the expected cost using the myopic stopping rule and the expected cost using the Bayes fixed sample size procedure (BFSSP) is derived in order to compare between them.

In Section 5, the results of Section 2 through 4 are applied in three examples.

In section 6, the concluding remarks are given.

2. DISTRIBUTION THEORY

Let the distribution function of a random variable X be $F_\theta(x)$, where $\theta'=(\theta_1,\theta_2,\cdots\theta_k)$ is a k-vector of parameters. If this is of the exponential family, then we can write in the form

(2.1) $dF_\theta(x)=\exp\ [\alpha(\theta')\beta(x)-\gamma(\theta)]d\mu(x)$,

where $\alpha(\theta)$ and $\beta(x)$ are k-vectors and $\mu\ (\cdot)$ is a σ-finite measure on the real line and $\beta(\cdot)$ is a measurable function on the real line. Suppose that the parameters take values in a parameter space, all identifiable and that $\alpha(\theta)$ and $\gamma(\theta)$ are differentiable. Then we can choose the parameters so that

(2.2) $E[\beta[(X)]=\theta$.

Since the total integral of (2.1) is unity,

$$0=\frac{\partial}{\partial\theta_i}\int dF_\theta(x)=\sum_{j=1}^{k}\frac{\partial\alpha_j}{\partial\theta_i}E[\beta_j(x)]-\frac{\partial\gamma}{\partial\theta_i}$$

Therefore (2.2) implies that

$$(2.3)\quad \frac{\partial\gamma}{\partial\theta_i}=\sum_{j=1}^{k}\theta_j\frac{\partial\alpha_j}{\partial\theta_i},\quad i=1,2,\cdots k$$

Conversely, if the matrix $\{\partial\alpha_j/\partial\theta_i\}$ is non-singular, then (2.3) implies (2.2). Let $f_\theta(x)$ be the probability density function of X. If (2.2) holds and the matrix $\{\partial\alpha_j/\partial\theta_i\}$ is non-singular, then we may define the information matrix $J=\{J_{ij}\}$, where

$$J_{ij}=E\left[-\frac{\partial^2\log f_\theta(x)}{\partial\theta_i\,\partial\theta_j}\right]$$

From (2.2) and (2.3) we can deduce that

$$\begin{aligned}J_{ij}&=\frac{\partial^2\gamma}{\partial\theta_i\partial\theta_j}-\sum_{s=1}^{k}\theta_s\frac{\partial^2\alpha_s}{\partial\theta_i\partial\theta_j}\\&=\frac{\partial}{\partial\theta_i}\left(\frac{\partial\gamma}{\partial\theta_j}-\sum_{s=1}^{k}\theta_s\frac{\partial\alpha_s}{\partial\theta_j}\right)+\frac{\partial\alpha_i}{\partial\theta_j}\\&=\frac{\partial\alpha_i}{\partial\theta_j}\end{aligned}$$

Similarly, we obtain $J_{ij}=\frac{\partial\alpha_j}{\partial\theta_i}$. So that

$$(2.4)\quad J_{ij}=\frac{\partial\alpha_i}{\partial\theta_j}=\frac{\partial\alpha_j}{\partial\theta_i}$$

For u,v real, define

$$(2.5)\quad K(u,v)=\int\exp\,[\alpha(\theta)'u-\gamma(\theta)v]d\theta,$$

where $u'=(u_1,u_2\cdots u_k)$ is a k-vector, $d\theta=d\theta_1\,d\theta_2\cdots d\theta_k$, and let $A_0=\{(u,v)$ in R^{k+1} : $K(u,v)<\infty\}$. Then A_0 is the "natural parameter space" for the parameters of the prior distribution. The natural conjugate prior density of θ is

$$(2.6)\quad \lambda_0(\theta)=K(T_0,\,N_0)^{-1}\exp[\alpha(\theta)'T_0-\gamma(\theta)N_0],$$

where $T_0'=(T_{10},T_{20},\cdots T_{k0})$ is a k-vector and (T_0,N_0) in A_0 is required.

Suppose that given θ, random variables $X_1,X_2,\cdots X_n$ are independent and identically distributed f_θ. Then the joint density of $X_1,X_2,\cdots X_n$ is

$$f(X_n|\theta)=\exp[\alpha(\theta)'\sum_{i=1}^{n}\beta(X_i)-n_\gamma(\theta)],$$

where $X_n=(X_1,X_2,\cdots X_n)$.

Thus, the posterior density of θ given X_n is

$$(2.7)\quad \lambda_n(\theta)=K(T_n,N_n)^{-1}\exp[\alpha(\theta)'T_n-\gamma(\theta)N_n],$$

where $T_n'=(T_{1n},\,T_{2n}\cdots\cdots T_{kn})=T_0'+\sum_{i=1}^{n}\beta(X_i)'$ and $N_n=N_0+n$.

Also, A_0 should contain (T_n, N_n) with probability one.

The loss function to be considered is related to the Fisher information. For real numbers a and b, define

(2.8) $B_{a,b}(\theta)=\exp[-\alpha(\theta)'a+\gamma(\theta)b]$,

where $a'=(a_1,a_2,\cdots a_k)$ is a k-vector and loss

(2.8) $L_{a,b}(\theta,\hat{\theta})=B_{a,b}(\theta)(\hat{\theta}-\theta)'J(\hat{\theta}-\theta)$,

where $\hat{\theta}$ is the estimator of θ.

Before proceeding further, several restrictions are needed to insure finite expected loss. Define

set $A_1=\{(u,v) \text{ in } R^{k+1} : \int J \exp[\alpha(\theta)'u-\gamma(\theta)v]d\theta<\infty\}$,

$A_2=\{(u,v) \text{ in } R^{k+1} : \int \theta' J\theta \exp[\alpha(\theta)'u-\gamma(\theta)v]d\theta<\infty\}$, and

$A=A_0\cap A_1\cap A_2$. Then for all n and fixed a and b,

(2.10) $(T_n-a,\ N_n-b)$ in A and (T_n, N_n) in A_0 with probability one

must be assumed. This condition simply restricts the choice of prior parameters (T_0, N_0) in terms of a and b and implies that $E\ B_{a,b}(\theta)<\infty$.

To obtain the Bayes estimator, the following conditions must be satisfied for (u,v) in A:

(2.11) For all $i=1,2,\cdots,k$,

$$\int\frac{\partial}{\partial\theta_i}\exp[\alpha(\theta)'u-\gamma(\theta)v]d\theta=0$$

$$\text{and }\int\frac{\partial}{\partial\theta_i}\theta_i\exp[\alpha(\theta)'u-r(\theta)v]d\theta=0.$$

The expected loss given X_n is

$$E[L_{a,b}(\theta,\hat{\theta})\,|\,X_n]=K(T_n,N_n)^{-1}\int\exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)]\ (\hat{\theta}-\theta)'J(\hat{\theta}-\theta)d\theta.$$

To minimize the expected loss,

Let $\dfrac{\partial E(L_{a,b}(\theta,\hat{\theta})\,|\,X_n]}{\partial\hat{\theta}}=0.$

Then $\int\exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)]J\hat{\theta}d\theta$

$$=\int\exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)]J\ \theta\ d\theta$$

and $\int\exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)]J\left(\hat{\theta}-\dfrac{T_n-a}{N_n-b}\right)d\theta$

$$=\int\exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)]\left[\frac{\partial\gamma_j}{\partial\theta_i}-\sum_j\frac{T_{in}-a_j}{N_n-b}\ \frac{\partial\gamma_j}{\partial\theta_i}\right]d\theta$$

$$=\int \exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)\,(N_n-b)]\Big[\frac{\partial\gamma}{\partial\theta_i}-\sum_j \frac{T_{jn}-a_j}{N_n-b}\ \frac{\partial\alpha_j}{\partial\theta_i}\Big]\, d\theta$$

$$=-\frac{1}{N_n-b}\Big\{\int\frac{\partial}{\partial\theta_i}\ \exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)]\, d\theta\Big\}=0.$$

Thus we obtain the Bayes estimator of θ given X_n

(2. 12) $\hat{\theta}_n(a,b)=\hat{\theta}_n=\dfrac{T_n-a}{N_n-b}$

and the posterior expected loss is

$$E[L_{a,b}(\theta,\hat{\theta})\,|\,X_n]$$

$$=K(T_n,N_n)^{-1}\int \exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)]$$

$$\Big(\frac{T_n-a}{N_n-b}-\theta\Big)'J\Big(\frac{T_n-a}{N_n-b}-\theta\Big)\, d\theta$$

$$=K(T_n,N_n)^{-1}\int \exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)]$$

$$\Big[\sum_j \frac{T_{jn}-a_j}{N_n-b}\ \frac{\partial\alpha_j}{\partial\theta_i}-\frac{\partial\gamma}{\partial\theta_i}\Big]\Big(\frac{T_n-a}{N_n-b}-\theta\Big)\, d\theta$$

$$=K(T_n,N_n)^{-1}\frac{1}{N_n-b}\Big\{\sum_i\int\frac{\partial}{\partial\theta_i}\ \exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)]$$

$$\Big(\frac{T_{in}-a_i}{N_n-b}-\theta_i\Big)+k\int \exp[\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)\Big]\, d\theta$$

$$=K(T_n,N_n)^{-1}\frac{k}{N_n-b}\ \int \exp\ [\alpha(\theta)'(T_n-a)-\gamma(\theta)(N_n-b)]d\theta.$$

Thus we obtain the posterior expected loss

(2. 13) $E_nL_{a,b}=k(N_n-b)^{-1}K(T_n-a,N_n-b)K(T_n,N_n)^{-1}$

This section is ended with a theorem which will be of major use in the next sectio-ns. Define

(2. 14) $Y_n(a,b)=K(T_n-a,N_n-b)K(T_n,N_n)^{-1}$

Theorem 2. 1. For fixed a and b, $\{Y_n(a,b)\}$ is a uniformly integrable martingale sequence.

Remark. This theorem is an extended form of Theorem 1. 1 in Shapiro and Wardrop [11].

3. MYOPIC STOPPING RULES

Of interest are sequential estimation procedures $(t,\hat{\theta}_t)$ where t is a stopping time with respect to F_n, σ-algebra generated by $(X_1,X_2,\cdots,X_n)$ and $\hat{\theta}_t$ is an F_t-measurable function. The object is to find the Bayes sequential procedure, the proceduer which minimizes the expected total cost. Due to the Bayesian nature of the problem the estimator $\hat{\theta}_t$ can be taken to be the Bayes estimator of θ given F_t, and henceforth, the procedure$(t, \hat{\theta}_t)$ is identified with stopping time t. Note that $\hat{\theta}_t$ is given by (2.12) with n replaced by t.

In the present problem, the total cost of observing n random variables from f_θ is

(3.1) $\mathcal{T}_n(a, b)=E_nL_{a,b}+cn,$

where $E_nL_{a,b}$ is given in (2.13), and cost $c>0$.

First, a definition is needed.

Definition. Myopic rule t is defined to be the first $n\geq 0$ such that $E[\mathcal{T}_{n+1}(a, b)\,|\,F_n]\geq\mathcal{T}_n(a, b)$ (*cf.* [11]).

Define stopping time $\tau_{a,b}$ be equal to the first $n\geq 0$ such that

(3.2) $(N_n-b+1)c\geq E_nL_{a,b}$.

Theorem 3.1.

i) The stopping rule $\tau_{a,b}$ is the myopic rule.

ii) $E\tau_{a,b}<\infty$.

iii) If $(N_n-b+1)^{-1}E_nL_{a,b}$ is nonincreasing in n, then $\tau_{a,b}$ is optimal (the Bayes sequential procedure).

Remark. This theorem is an extended form of Theorem 2.1 in Shapiro and Wardrop [11].

Although the condition in (iii) of Theorem 3.1 can be stated in terms of the $K(u, v)$ functions, the condition must be checked in each application. In many instances, the rule $\tau_{a,b}$ will be optimal for restricted values of a and b and possibly restricted N_0. For example, if f_θ is normal with unknown mean μ and variance σ^2, then $a_1=a_2=b=0$ gives a fixed sample size procedure and the condition in (iii).

4. ASYMPTOTIC RESULTS

In this section, two practical questions are considered. First, in those cases where the myopic rule may not be optimal, how much is lost by using the myopic rule instead of the unknown optimal rule? Theorem 4.2 below shows that as cost $c\to 0$, the myopic rule performs as well as any rule; i.e. it is asymptotically optimal. Second, since there is generally some inconvenience in implementing a sequential procedure, how much is gained when the myopic rule is used instead of the Bayes fixed sample size procedure

(BFSSP)? Theorem 4.1 below gives the limiting ratio of the expected cost using the myopic rule and the expected cost using the BFSSP. The asymptotic properties given in this section are derived by allowing cost c to tend to zero.

First, the BFSSP is derived. For $n \geq 0$,

$$(4.1)\quad E\mathcal{T}_n = EY_n k(N_0-b+n)^{-1}+cn$$
$$= EB_{a,b}(\theta)k(N_0-b+n)^{-1}+cn$$

since $\{Y_n(a,b)\}$ is a martingale sequence. The integer $n'=n'_{a,b}(c)$ which minimizes (4.1) is the BFSSP.

Let $\dfrac{dE\mathcal{T}_n}{dn}=0$

Then $n'=c^{-1/2}[kEB_{a,b}(\theta)]^{1/2}-N_0+b$ and

$$(4.2)\quad \lim_{c\to 0} c^{-1/2}E(\mathcal{T}_{n'}) = \lim_{c\to 0} c^{-1/2}[EB_{a,b}(\theta)k(N_0-b+n')^{-1}+cn']$$
$$=2[kEB_{a,b}(\theta)]^{1/2}=2k^{1/2}[EB_{a,b}(\theta)]^{1/2}.$$

For comparison with (4.2), the limit of $c^{-1/2}E\mathcal{T}\tau_{a,b}$ will be computed.

Theorem 4.1. For the myopic rule $\tau_{a,b}$ and the BFSSP, n',

$$(4.3)\quad \lim_{c\to 0} c^{-1/2}E(\mathcal{T}\tau_{a,b})=2k^{1/2}E[B_{a,b}(\theta)^{1/2}], \text{ and}$$

$$(4.4)\quad \lim_{c\to 0} (\mathcal{T}\tau_{a,b})/E(\mathcal{T}_{n'})=\frac{E[B_{a,b}(\theta)^{1/2}]}{[EB_{a,b}(\theta)]^{1/2}}\leq 1.$$

Remark. This theorem is an extended form of Theorem 3.1 in Shapiro and Wardrop [11].

Theorem 4.2. Let $\{t_c\}$, $(c\to 0)$, be any sequence of stopping times and let $\tau_{a,b}$ be the myopic rule, then

$$(4.5)\quad \lim_{c\to 0}\inf c^{-1/2}E(\mathcal{T}t_c)\geq \lim_{c\to 0} c^{-1/2}E(\mathcal{T}\tau_{a,b})$$

Remark. This theorem is an extended form of Theorem 3.2 in Shapiro and Wardrop [11].

5. EXAMPLES

In this section, the results of Sections 2 through 4 are applied in three cases with particular emphasis on the asymptotic savings, $h_{a,b}(T_0, N_0)$, realized when the myopic rule is used rather than the BFSSP. Because of Theorem 4.2, no other sequential procedure gives more savings (asymptotically) than the myopic rule. In general, $1-h_{a,b}$ is the limiting ratio of the expected cost using the myopic rule and the expected cost using the BFSSP. Theorem 4.1. yields

$$(5.1)\ 1-h_{a,b}(T_0, N_0)=\frac{E[B_{a,b}(\theta)^{1/2}]}{[EB_{a,b}(\theta)]^{1/2}}$$

$$=K(T_0-a/2, N_0-b/2)K(T_0-a, N_0-b)^{-1/2}K(T_0, N_0)^{-1/2}$$

Example 1. Normal variable with unknown mean and variance. Here

$$f_\theta(x)=\exp\Big[\frac{\mu x}{\sigma^2}-\frac{x^2}{2\sigma^2}-\frac{\mu^2}{2\sigma^2}-\log\{\sqrt{2\pi}\sigma\}\Big]$$

A possible choice of functions is $\beta_1(x)=x, \beta_2(x)=x^2$. The parameters must then be chosen as

$$\theta_1=E[\beta_1(x)]=(\mu),$$
$$\theta_2=E[\beta_2(x)]=\mu^2+\sigma^2.$$

Then

$$\alpha_1(\theta)=\frac{\mu}{\sigma^2}=\frac{\theta_1}{\theta_2-\theta_1^2},$$

$$\alpha_2(\theta)=-\frac{1}{2\sigma^2}=-\frac{1}{2(\theta_2-\theta_1^2)},$$

$$\gamma(\theta)=\frac{\theta_1^2}{2(\theta_2-\theta_1^2)}+\frac{1}{2}\log(\theta_2-\theta_1^2)+\frac{1}{2}\log\ (2\pi).$$

This gives information matrix

$$J=\{J_{ij}\}=\begin{bmatrix}\frac{\theta_2+\theta_1^2}{(\theta_2-\theta_1^2)^2} & -\frac{\theta_1}{(\theta_2-\theta_1^2)^2}\\ -\frac{\theta_1}{(\theta_2-\theta_1^2)^2} & \frac{1}{2(\theta_2-\theta_1^2)^2}\end{bmatrix}$$

With $B_{a,b}(\theta)=\exp\Big[-\frac{\mu}{\sigma^2}a_1+\frac{1}{2\sigma^2}a_2+\frac{1}{2}\Big\{\frac{\mu^2}{\sigma^2}+\log(2\pi\sigma^2)\Big\}b\Big]$, resulting in

$$L_{a,b}(\theta,\hat\theta)=B_{a,b}(\theta)\Big[\frac{(\hat\theta_1-\mu)^2}{\sigma^2}+\frac{1}{2}\Big\{\frac{(\hat\theta_1-\mu)^2}{\sigma^2}+\frac{(\hat\theta_2-\hat\theta_1^2)-\sigma^2}{\sigma^2}\Big\}^2\Big]$$

The prior parameters are restricted to $-\infty<T_{10}<\infty$ and $T_{20}>\max\ (0, a_2)$ and $N_0>\max\ (3, b+3)$ to give (T_n-a, N_n-b) in A as defined in (2.10).

Also, $K(u,v)=(2\pi)^{-(v-1)/2}v^{-1/2}\Big(\frac{vu_2-u_1^2}{2v}\Big)^{-(v+3)/2}\Gamma\Big(\frac{v-3}{2}\Big)$

From the results of Section 2 it follows that the optimal estimates for $E(x)$ and $E(x^2)$ are

$$\hat\theta_1=\frac{1}{N_0+n-b}\Big(T_{10}+\sum_{i=1}^{n}x_i-a_1\Big)$$

$$\hat\theta_2=\frac{1}{N_0+n-b}\Big(T_{20}+\sum_{i=1}^{n}x_i^2-a_2\Big),\ \text{and}$$

the posterior expected loss, given any sample $(x_1, x_2 \dots, x_n)$, is

$$E_n L_{a,b} = \left(\frac{2}{N_0+n-b}\right) K(T_n - a, N_0+n-b) K(T_n, N_0+n)^{-1}$$

where $\left(T_n' = T_{10} + \sum_{i=1}^{n} x_i,\ T_{20} + \sum_{i=1}^{n} x_i^2\right)$, $a' = (a_1, a_2)$.

Next, from the condition in (iii) of Theorem 3.1 it follows that the costs are in the nonincreasing case (yielding optimality of the myopic rule $\tau_{a,b}$) if $a_1 = a_2 = b = 0$ which yields a fixed sample size procedure, i.e. $(N_n - b + 1)^{-1}$ $E_n L_{a,b} = 2/(N_0+n+1)(N_0+n)$. In particular, if $a_1 = a_2 = 0$ and $b > 0$, then

$$1 - h_{0,b}\ (T_0, N_0) = K(T_0, N_0 - b/2) K(T_0, N_0 - b)^{-1/2} K(T_0, N_0)^{-1/2}$$

$$= [N_0(N_0 - b)(N_0 - b/2)^{-2}]^{1/4}$$

$$\left[\left\{\frac{N_0 T_{20} - T_{10}^2}{N_0}\right\}^{N_0-3} \left\{\frac{(N_0-b) T_{20} - T_{10}^2}{N_0 - b}\right\}^{N_0-b-3} \left\{\frac{(N_0 - b/2) T_{20} - T_{10}^2}{N_0 - b/2}\right\}^{-2N_0+b+6}\right]^{1/4}$$

$$\left[\Gamma\left(\frac{N_0-3}{2}\right)\Gamma\left(\frac{N_0-b-3}{2}\right)\Gamma\left(\frac{N_0-b/2-3}{2}\right)^{-2}\right]^{-1/2}$$

If T_{10} and T_{20} are fixed, and $N_0 \to \infty$, then $1 - h_{0,b} \to 1$. If $N_0 \to b$, then $1 - h_{0,b} \to 0$ for $b > 0$ which indicates an asymptotic savings approaching 100% with the myopic procedure.

Example 2. A gamma distribution with unknown parameters α and β ($\alpha > 0$, $\beta > 0$). Here

$$f_\theta(x) = \exp[-\beta x + (\alpha - 1) \log x + \alpha \log \beta - \log \Gamma(\alpha)].$$

A possible choice of functions is $\beta_1(x) = x, \beta_2(x) = \log x$. The parameters must then be chosen as

$$\theta_1 = E[\beta_1(x)] = \frac{\alpha}{\beta},$$

$$\theta_2 = E[\beta_2(x)] = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \log \beta,$$

that is, $\beta = \dfrac{\alpha}{\theta_2}$,

$$\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \log \alpha = \theta_2 - \log \theta_1,$$

which is to be solved for α. In this case, the equation is not easily solvable and it is necessary to resort to numerical methods, using tables for $\Gamma'(\alpha)/\Gamma(\alpha)$.

Example 3. The trivariate normal distribution with unknown means $\theta_1, \theta_2, \theta_3$. Here

$$f_\theta(x_1, x_2, x_3) = (2\pi)^{-3/2} (\sigma_1 \sigma_2 \sigma_3)^{-1} \Delta^{-1/2}$$

$$\exp\left[-\frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} A_{ij} \frac{(x_i - \theta_i)}{\sigma_i} \frac{(x_j - \theta_j)}{\sigma_j}\right]$$

where $\Delta=1-\rho_{23}^2-\rho_{13}^2-\rho_{12}^2+2\rho_{23}\rho_{13}\rho_{12}$,

$$A_{11}=(1-\rho_{23}^2)\Delta^{-1},\quad A_{22}=(1-\rho_{13}^2)\Delta^{-1},\quad A_{33}=(1-\rho_{12}^2)\Delta^{-1},$$

$$A_{12}=A_{21}=(\rho_{13}\rho_{23}-\rho_{12})\Delta^{-1},\quad A_{13}=A_{31}=(\rho_{12}\rho_{23}-\rho_{13})\Delta^{-1},$$

$$A_{23}=A_{32}=(\rho_{12}\rho_{13}-\rho_{23})\Delta^{-1}.$$

A possible choice of functions is

$$\beta_1(x)=x_1,\quad \beta_2(x)=x_2,\quad \beta_3(x)=x_3.$$

The parameters must then be chosen as

$$\theta_1=E[\beta_1(x)],\quad \theta_2=E[\beta_2(x)],\quad \theta_3=E[\beta_3(x)].$$

Then $\alpha_1(\theta)=\dfrac{A_{11}\theta_1}{\sigma_1^2}+\dfrac{A_{12}\theta_2}{\sigma_1\sigma_2}+\dfrac{A_{13}\theta_3}{\sigma_1\sigma_3}$,

$$\alpha_2(\theta)=\frac{A_{12}\theta_1}{\sigma_1\sigma_2}+\frac{A_{22}\theta_2}{\sigma_2^2}+\frac{A_{23}\theta_3}{\sigma_2\sigma_3},$$

$$\alpha_3(\theta)=\frac{A_{13}\theta_1}{\sigma_1\sigma_3}+\frac{A_{23}\theta_2}{\sigma_2\sigma_3}+\frac{A_{33}\theta_3}{\sigma_3^2},$$

$$\gamma(\theta)=\frac{A_{11}\theta_1^2}{2\sigma_1^2}+\frac{A_{12}\theta_1\theta_2}{\sigma_1\sigma_2}+\frac{A_{22}\theta_2^2}{2\sigma_2^2}+\frac{A_{23}\theta_2\theta_3}{\sigma_2\sigma_3}+\frac{A_{33}\theta_3^2}{2\sigma_3^2}+\frac{A_{13}\theta_1\theta_3}{\sigma_1\sigma_3}$$

This gives information matrix

$$J=\{J_{ij}\}=\begin{bmatrix}\dfrac{A_{11}}{\sigma_1^2} & \dfrac{A_{12}}{\sigma_1\sigma_2} & \dfrac{A_{13}}{\sigma_1\sigma_3}\\[2ex] \dfrac{A_{12}}{\sigma_1\sigma_2} & \dfrac{A_{22}}{\sigma_2^2} & \dfrac{A_{23}}{\sigma_2\sigma_3}\\[2ex] \dfrac{A_{13}}{\sigma_1\sigma_3} & \dfrac{A_{23}}{\sigma_2\sigma_3} & \dfrac{A_{33}}{\sigma_3^2}\end{bmatrix}$$

with $B_{a,b}\ (\theta)=\exp\ [-\alpha_1(\theta)a_1-\alpha_2(\theta)a_2-\alpha_3(\theta)a_3+\gamma(\theta)b]$, resulting in

$$\begin{aligned}L_{a,b}(\theta,\hat\theta)&=B_{a,b}(\theta)(\hat\theta-\theta)'J(\hat\theta-\theta)\\&=B_{a,b}(\theta)\Big[\frac{A_{11}}{\sigma_1^2}(\hat\theta_1-\theta_1)^2+\frac{2A_{12}}{\sigma_1\sigma_2}(\hat\theta_1-\theta_1)(\hat\theta_2-\theta_2)+\frac{A_{22}}{\sigma_2^2}(\hat\theta_2-\theta_2)^2\\&\quad+\frac{2A_{23}}{\sigma_2\sigma_3}(\hat\theta_2-\theta_2)(\hat\theta_3-\theta_3)+\frac{A_{33}}{\sigma_3^2}(\hat\theta_3-\theta_3)^2+\frac{2A_{13}}{\sigma_1\sigma_3}(\hat\theta_1-\theta_1)(\hat\theta_3-\theta_3)\Big]\end{aligned}$$

The prior parameters are restricted to $-\infty<T_{10}<\infty$, $-\infty<T_{20}<\infty$, $-\infty<T_{30}<\infty$ and $N_0>\max\ (0,b)$ to give (T_n-a, N_n-b) in A as defined in (2.10). Also,

$$\begin{aligned}K(u,v)&=(2\pi)^{3/2}v^{-3/2}\ (\sigma_1\sigma_2\sigma_3)[A_{11}(A_{22}A_{33}-A_{23}^2)\\&\quad+A_{12}(A_{13}A_{23}-A_{12}A_{33})+A_{13}(A_{12}A_{23}-A_{13}A_{22})]^{-1/2}\\&\quad\exp\Big[\frac{A_{11}u_1^2}{2v\sigma_1^2}+\frac{A_{12}u_1u_2}{v\sigma_1\sigma_2}+\frac{A_{22}u_2^2}{2v\sigma_2^2}\\&\quad+\frac{A_{23}u_2u_3}{v\sigma_2\sigma_3}+\frac{A_{33}u_3^2}{2v\sigma_3^2}+\frac{A_{13}u_1u_3}{v\sigma_1\sigma_3}\Big]\end{aligned}$$

From the results of Section 2 it follows that the optimal estimates for θ_1, θ_2 and θ_3 are

$$\hat{\theta}_1=\frac{1}{N_0+n-b}(T_{10}+\sum_{j=1}^{n}x_{1j}-a_1),$$

$$\hat{\theta}_2=\frac{1}{N_0+n-b}(T_{20}+\sum_{j=1}^{n}x_{2j}-a_2),$$

$$\hat{\theta}_3=\frac{1}{N_0+n-b}(T_{30}+\sum_{j=1}^{n}x_{3j}-a_3), \text{ and}$$

the posterior expected loss, given any sample $(x_1, x_2, \dots, x_n)$, is

$$E_nL_{a,b}=\frac{3}{N_0+n-b}\ K(T_n-a, N_0+n-b)K(T_n, N_0+n)^{-1}$$

where $T_n'=(T_{10}+\sum_{j=1}^{n}x_{1j},\ T_{20}+\sum_{j=1}^{n}x_{2j},\ T_{30}+\sum_{j=1}^{n}x_{3j}),\ a'=(a_1, a_2, a_3)$.

Next, from the condition in (iii) of Theorem 3.1 it follows that the costs are in the nonincreasing case (implying optimality of the myopic rule $\tau_{a,b}$) if $a_1=a_2=a_3=b=0$ which yields a fixed sample size procedure,. i.e. $(N_n-b+1)^{-1}E_nL_{a,b}=3/(N_0+n+1)\ (N_0+n)$. If $a_1=a_2=a_3=0$ and $b>0$, then

$$1-h_{0,b}(T_0, N_0)=K(T_0, N_0-b/2)K(T_0, N_0-b)^{-1/2}K(T_0, N_0)^{-1/2}$$

$$=[N_0(N_0-b)\ (N_0-b/2)^{-2}]^{3/4}\exp\Big[-\Big\{\Big(\frac{A_{11}T_{10}^2}{\sigma_1^2}+\frac{A_{12}T_{10}T_{20}}{\sigma_1\sigma_2}+\frac{A_{22}T_{20}^2}{\sigma_2^2}$$

$$+\frac{A_{23}T_{20}T_{30}}{\sigma_2\sigma_3}+\frac{A_{33}T_{30}^2}{\sigma_3^2}+\frac{A_{13}T_{10}T_{30}}{\sigma_1\sigma_3}\Big)N_0^{-2}\Big\}b^2N_0(N_0-b)^{-1}(2N_0-b)^{-1}(1/2)\Big]$$

If T_{10}/N_0, T_{20}/N_0 and T_{30}/N_0 are held fixed while $N_0\to\infty$ (sharper prior knowledge), then $1-h_{0,b}\to 1$. If $N_0\to b$, then $1-h_{0,b}\to 0$ for $b>0$. Also, if T_{10}, T_{20}, T_{30} and N_0 are fixed, and $\sigma_1^2\to 0$, $\sigma_2^2\to 0$ and $\sigma_3^2\to 0$, then $1-h_{0,b}\to 0$ indicating that greater precision in the observations yields more asymptotic savings (up to 100%) with the myopic procedure. If $\sigma_1^2\to\infty$, $\sigma_2^2\to\infty$ and $\sigma_3^2\to\infty$, then $1-h_{0,b}\to[N_0(N_0-b)(N_0-b/2)^{-2}]^{3/4}$.

6. CONCLUDING REMARKS

The considering of exponential families of form (2.1) along with loss $L_{a,b}$ leads to tractable expressions for both the Bayes estimator of θ and the posterior expected loss. The class of loss functions $L_{a,b}$ is a rich class in general. Since the posterior expected loss is a multiple of $Y_n\ (a, b)$, a martingale sequence, the myopic stopping rule can also be expressed easily and form of the rule is particularly intuitive: stop sampling when sampling cost (approximately $(N_n-b+1)\ c$) exceeds estimation cost $(E_nL_{a,b})$. Also, sufficient conditions for nonincreasing case costs (and hence optimality of the myopic rule) result from this general expression of the posterior loss. Finally, the role played by Y_n (a, b) leads to the asymptotic results of Section 4 giving the asymptotic optimality of the myopic stopping rule in all cases.

In (iii) of Theorem 3. 1, we have not derived necessary conditions yielding optimality of the myopic rule. Also, we have not considered the Bayes sequential estimation problem in the case that the conjugate prior distribution does not belong to exponential family. These problems deserve further consideration and will be studied in the future.

REFERENCES

(1) Breiman, L. (1968). *Probability*. Addison-Wesley: Reading.

(2) Cabilio, P. (1977). Sequential estimation in Bernoulli trials. *Ann. Statist.* 2 342~356.

(3) Chow, Y.S., Robbins, H., and Seigmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin: Boston.

(4) Degroot, M.H. (1970). *Optimal Statistical Decisions*. McGraw-Hill: New York.

(5) El-Sayyad, G.M. and Freeman, P.R. (1973). Bayesian sequential estimation of a Poisson process rate. *Biometrika* 60 289~296

(6) Ferguson, T.S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic: New York.

(7) Lehmann, E.L. (1959). *Testing Statistical Hypotheses*. Wiley: New York.

(8) Nevic, B. (1977). Bayes sequential estimation of a Poisson rate. Technical Report #134, Department of Statistics, Carnegie-Mellon University.

(9) Shapiro, C.P. and Wardrop, R.L. (1977). The continuous time Bayes sequential procedure for estimating the arrival rate of a poisson process and large sample properties. Technical Report #1753, Mathematics Research Center, University of Wisconsin, Madison.

(10) Shapiro, C.P. and Wardrop, R.L. (1978). Simultaneous Bayesian sequential estimation for success probabilities. Technical Report#380, Department of Statistics and probability, Michigan State University.

(11) Shapiro, C.P. and Wardrop, R.L. (1978). Bayesian sequential estimation for one parameter exponential families. Technical Report #537, Department of Statistics, University of Wisconsin, Madison.

(12) Whittle, P. and Lane, R.O.D. (1967). A class of situations in which a sequential procedure is non-sequential. *Biometrika* 54 229~234.