The Korean Communications
in Statistics Vol. 8, No. 1, 2001
pp. 87-96

Empirical Bayes Estimation of the Binomial and Normal Parameters

Jee Chang Hong[1] and Inha Jung[2]

Abstract

We consider the empirical Bayes estimation problems with the binomial and normal components when the prior distributions are unknown but are assumed to be in certain families. There may be the families of all distributions on the parameter space or subfamilies such as the parametric families of conjugate priors. We treat both cases and establish the asymptotic optimality for the corresponding decision procedures.

Keywords : empirical Bayes estimation, asymptotically optimal, conjugate prior

1. Introduction

1.1. Bayes decision problem

Consider a statistical decision problem(called a component problem hereafter) with the following elements :

(1) A parameter space Θ, with a generic element θ. θ is the "state of nature" which is unknown to us.

(2) An action space A, with generic element a.

(3) A loss function L on $A \times \Theta$ to $[0, \infty)$ with $L(a, \theta)$ representing the loss of taking action a when θ is the true state of nature.

(4) A prior distribution G on Θ.

1) Fulltime Lecturer, Department of Liberal Arts, Tong Hae University, Tonghae, Kangwon, 240-713, Korea,
E-mail : jchong@mail.tonghae.ac.kr
2) Associate Professor, Division of Information and Computer Engineering, Ajou University, Suwon, Kyonggi, 442-749, Korea,
E-mail : ihjung@madang.ajou.ac.kr

(5) A random variable X taking values in a space $\aleph$, and for a given realization θ of a random variable having distribution G, X has a specified probability density $f_\theta(\cdot)$ with respect to a σ-finite measure μ on a σ-field in $\aleph$.

The statistical decision problem is to choose a decision function $t:\aleph\to A$. Assuming that $L(t(\cdot),\theta)$ is a measurable function on $\aleph\times\Theta$, the average(or expected) loss function of t is given by

$$R(t,\theta)=\int_\aleph L(t(x),\theta)f_\theta(x)d\mu(x). \tag{1.1}$$

The overall expected loss with $\theta\sim G$ is

$$R(t,G)=\int_\Theta R(t,\theta)dG(\theta) \tag{1.2}$$

called the Bayes risk of t relative to G. Thus,

$$R(t,G)=\int_\aleph \phi_G(t(x),x)d\mu(x), \tag{1.3}$$

where

$$\phi_G(a,x)=\int_\Theta L(a,\theta)f_\theta(x)dG(\theta). \tag{1.4}$$

Let $t_G:\ \aleph\to A$ be defined by

$$\phi_G(t(x),x)=\min_{a\in A}\phi_G(a,x). \tag{1.5}$$

Then, for any decision function t,

$$R(t_G,G)=\int_\aleph \min_{a\in A}\phi_G(a,x)d\mu(x)\le R(t,G) \tag{1.6}$$

so that, defining

$$R(G)=R(t_G,G)=\int_\aleph \phi_G(t_G(x),x)d\mu(x), \tag{1.7}$$

we have

$$R(G)=\min_t R(t,G). \tag{1.8}$$

Any decision function t_G satisfying (1.5) minimizes the Bayes risk relative to G, and called a Bayes decision function relative to G. The functional $R(G)$ defined by (1.7) is called the Bayes envelope functional of G(Berger, 1980). We are mainly concerned with the estimation problem under squared error loss defined by $L(\theta,a)=(\theta-a)^2$. A Bayes estimate $t_G(x)$ is a minimizer of $\phi_G(a,x)$ given by (1.4) over all $a\in A$, for each $x\in\aleph$, that is,

$$\int_\Theta(\theta-t_G(x))^2f_\theta(x)dG(\theta)=\min_{a\in A}\int_\Theta(\theta-a)^2f_\theta(x)dG(\theta).$$

It is well-known that $t_G(x)$ is given by the posterior mean of θ, i.e.,

$$t_G(x) = \mathrm{E}(\theta|x). \tag{1.9}$$

1.2. Empirical Bayes decision problem

When G is known, the Bayes decision function t_G defined by (1.5) is the most idealistic one. Unfortunately, G is not completely known in most cases. Robbins(1955, 1964) introduced the empirical Bayes approach to statistical decision problem which is applicable when we are faced with an independent sequence of Bayes decision problems having similar structure. Suppose now that the component decision problem described in Section 1.1 occurs repeatedly and independently with the same unknown prior distribution G in each repetition of the decision problem. That is, let $(\theta_1, X_1), \cdots, (\theta_N, X_N), \cdots$ be a sequence of independent, identically distribution(i.i.d.) random vectors where θ_N are i.i.d. G and where X_N has density $f_{\theta_N}(\cdot)$ with respect to μ given θ_N. $X_1, \cdots, X_N, \cdots$ are observable whereas $\theta_1, \cdots, \theta_N, \cdots$ are not. Viewing this setup at stage $(N+1)$ with G unknown, we have already accumulated the $X_1, \cdots, X_N$ and X_{N+1}, and we want to make a decision about θ_{N+1} under the loss L. Since G is assumed to be unknown, and since $X_1, \cdots, X_N$ is a random sample from the population with density

$$f_G(x) = \int_\Theta f_\theta(x)dG(\theta) \tag{1.10}$$

with respect to μ, it is reasonable to expect that $X_1, \cdots, X_N$ do contain some information about G. Eliciting this information about G from $X_1, \cdots, X_N$ and then using it to define

$$t_N(\cdot) = t_N(X_1, \cdots, X_N ; \cdot) \tag{1.11}$$

a decision rule for use in the $(N+1)$th decision problem to decide about θ_{N+1}, we incur an expected loss at stage $(N+1)$ given by

$$\begin{aligned} R_N(T, G) &= \mathrm{E}[R(t_N(\cdot), G)] \\ &= \int_{\aleph} \mathrm{E}\phi_G(t_N(x), x)d\mu(x) \\ &= \int_{\aleph}\int_\Theta \mathrm{E}[L(t_N(x), \theta)]f_\theta(x)dG(\theta)d\mu(x) \end{aligned} \tag{1.12}$$

with $T=\{t_N\}$. The search for rules $\{t_N\}$ which are asymptotical optimal relative to G for every distribution G within a certain class has taken basically two tracks. The first track is to use values $X_1, \cdots, X_N$ to form an estimate of G, call it $\hat{G}_N$, and then let t_N be a Bayes decision rule with respect to $\hat{G}_N$, i.e., let $t_N = t_{\hat{G}_N}$. The second track is to estimate the form of the Bayes decision rule directly without estimating G first.

Let $\mathscr{G}$ be a class of priors on Θ to which G is expected to belong. The class $\mathscr{G}$ may be the class of all priors on Θ or a subclass of priors on Θ. If $\mathscr{G}$ is the class of all priors, we do not assume any functional form of the unknown prior G, and the corresponding empirical Bayes problem is called nonparametric empirical Bayes. When G is assumed to be in the conjugate family, then $\mathscr{G}$ is a parametric subclass of all priors on Θ. Empirical Bayes problem corresponding to this parametric subclass is called parametric empirical Bayes. It is obvious from the equality (1.12) and the definition of $R(G)$ that

$$R_N(T, G) \geq R(G) \tag{1.13}$$

for all $T=\{t_N\}$ for all $G \in \mathscr{G}$. The sequence of decision functions $T=\{t_N\}$, where t_N is defined by (1.11) is referred to as an empirical Bayes decision procedure.

Definition 1.1. If $\lim_{N \to \infty} R_N(T, G) = R(G)$ for all $G \in \mathscr{G}$, we say that empirical Bayes decision procedure T is asymptotically optimal.

The following lemma is well-known and useful to prove asymptotic optimality of an empirical Bayes estimation procedure $T=\{t_N\}$ when squared error loss is used.

Lemma 1.2.

$$\begin{aligned} 0 &\leq R(t_N(X_1, \cdots, X_N ; X_{N+1}), G) - R(G) \\ &= \mathrm{E}[(t_N(X_1, \cdots, X_N ; X_{N+1}) - t_G(X_{N+1}))^2], \end{aligned}$$

where E denotes the expectation with respect to X_{N+1}.

Asymptotically optimal empirical Bayes estimation procedures with binomial and normal components are considered in Section 2 and 3 for both nonparametric and parametric situations. For the parametric empirical Bayes we follow the first tract and for the nonparametric empirical Bayes we follow the second track to obtain asymptotical optimal empirical Bayes estimates.

2. Esimation of the binomial parameter

Let $(\theta_1, X_1), \cdots, (\theta_N, X_N), \cdots$ be a sequence of independent and identically distributed random vectors, in which the conditional distribution of X_i given θ_i is binomial (m, θ_i) for each $i=1,2,\cdots$, and $\theta_1, \theta_2, \cdots$ are i.i.d. $G \in \mathscr{G}$. Here m is given positive integer. We consider the asymptotical optimal empirical Bayes estimates of θ_{N+1} based on $X_1, \cdots, X_{N+1}$.

2.1. Nonparametric empirical Bayes estimate

Let $\mathscr{G} = \left\{ G \mid \int_0^1 |\theta| dG(\theta) < \infty \right\}$. A Bayes estimate $t_G(x)$ of θ_{N+1} based on $X_{N+1} = x$ under squared error loss is given by posterior mean of θ, i.e., for $x = 0, 1, \cdots, m$,

$$\begin{aligned} t_G(x) &= \mathrm{E}(\theta_{N+1} | x) \\ &= \frac{\binom{m}{x} \int_0^1 \theta^{x+1}(1-\theta)^{m-x} dG(\theta)}{\binom{m}{x} \int_0^1 \theta^{x}(1-\theta)^{m-x} dG(\theta)} \\ &= \frac{x+1}{m+1} \frac{f_{G,m+1}(x+1)}{f_{G,m}(x)}, \end{aligned} \tag{2.1}$$

where $f_{G,m}(x) = \int_0^1 f_\theta(x) dG(\theta)$ is the marginal density function for X_i's.

Let $f_{N,m}(x)$ be a consistent estimate of $f_{G,m}(x)$, i.e., $f_{N,m}(x) \to f_{G,m}(x)$ a.s. as $N \to \infty$ for all x. Viewing (2.1), we obtain an empirical Bayes estimate

$$t_N(x) = \frac{x+1}{m+1} \frac{f_{N,m+1}(x+1)}{f_{N,m}(x)} \qquad x = 0, 1, \cdots, m. \tag{2.2}$$

It follows that for each fixed x $t_N(x) \to t_G(x)$ a.s. for any $G \in \mathscr{G}$ as $N \to \infty$.

Theorem 2.1. The empirical Bayes estimation procedure $T = \{t_N\}$ defined by (2.2) is asymptotically optimal.

Proof. It is easily seen that, for all sufficiently large N, $|t_N(x) - t_G(x)| < 1$. Since $t_N(x) \to t_G(x)$ a.s. as $N \to \infty$, it follows from dominated convergence theorem and Lemma 1.2 that $T = \{t_N\}$ is asymptotically optimal. □

Example 2.2. (Consistent estimate of $f_{G,m}(x)$) Consider the sequence of random variables $X'_1, \cdots, X'_N, \cdots$, where X'_N denotes the number of successes in the first $(m-1)$ out of the m trials which produced X_N successes, and let

$$f_{N,m-1}(x) = \frac{\text{number of terms } X'_1, \cdots, X'_N \text{ which are equal } x}{N}.$$

Then $f_{N,m-1}(x) \to f_{G,m-1}(x)$ a.s. as $N \to \infty$, by the strong law of large numbers.

2.2. Parametric empirical Bayes estimate

Suppose that G is in the class $\mathscr{G}$ of conjugate priors. Conjugate prior for binomial is beta distribution. Thus $\mathscr{G} =\{ \text{beta}(\alpha,\beta) \mid \alpha>0,\beta>0\}$. With the $G=\text{beta}(\alpha,\beta)$ in $\mathscr{G}$, posterior distribution of θ is $\text{beta}(\alpha+X_{N+1}, m+\beta-X_{N+1})$. A Bayes estimate is given by the posterior mean of θ_{N+1}, i.e.,

$$t_G(X_{N+1}) = \mathrm{E}(\theta_{N+1} \mid X_{N+1}) = \frac{\alpha+X_{N+1}}{\alpha+\beta+m}. \tag{2.3}$$

Let $\hat{\alpha}_N$, $\hat{\beta}_N$ be estimates of α, β using data $X_1,\cdots,X_N$. An empirical Bayes estimate $t_N(X_1,\cdots,X_N\ ; X_{N+1})$ corresponding to (2.3) is obtained by substituting α, β by $\hat{\alpha}_N$, $\hat{\beta}_N$, i.e.,

$$t_N(X_1,\cdots,X_N\ ; X_{N+1}) = \frac{\hat{\alpha}_N+X_{N+1}}{\hat{\alpha}_N+\hat{\beta}_N+m}. \tag{2.4}$$

Using the method of moments we obtain consistent estimates of α, β. Let $\xi=\mathrm{E}\,\theta$, $\eta=\mathrm{E}\,\theta^2$ be the first two moments of θ.

Remark 2.3. Let $\overline{X}_N$ and S_N^2 denote the sample mean and the sample variance of $X_1,\cdots,X_N$, respectively. It is easily seen that

$$\mathrm{E}\,\overline{X}_N = m\xi, \qquad \mathrm{E}\,S_N^2 = m(\xi-\eta). \tag{2.5}$$

Since $G=\text{beta}(\alpha,\beta)$, we have

$$\alpha=\frac{\xi(\xi-\eta)}{\eta-\xi^2}, \qquad \beta=\frac{(1-\xi)(\xi-\eta)}{\xi-\eta^2}. \tag{2.6}$$

Utilizing Remark 2.3 (2.6), the method of moments leads to the estimates of α, β through the estimates of ξ, η ;

$$\hat{\alpha}_N=\left(\frac{\hat{\xi}_N(\hat{\xi}_N-\hat{\eta}_N)}{\hat{\eta}_N-\hat{\xi}_N^{\,2}}\right)^+, \qquad \hat{\beta}_N=\left(\frac{(1-\hat{\xi}_N)(\hat{\xi}_N-\hat{\eta}_N)}{\hat{\eta}_N-\hat{\xi}_N^{\,2}}\right)^+, \tag{2.7}$$

where

$$\hat{\xi}_N=\frac{1}{m}\overline{X}_N, \qquad \hat{\eta}_N=\frac{1}{m}(\overline{X}_N-S_N^2). \tag{2.8}$$

Remark 2.4. $\hat{\alpha}_N$, $\hat{\beta}_N$ are consistent estimates of α, β.

Proof. By the strong law of large numbers, $\hat{\xi}_N\to\xi$ a.s. and $\hat{\eta}_N\to\eta$ a.s. as $N\to\infty$. Consequently, $\hat{\alpha}_N\to\alpha$ a.s. and $\hat{\beta}_N\to\beta$ a.s. as $N\to\infty$. □

We now prove the asymptotic optimality of the empirical Bayes decision procedure.

Theorem 2.5.(Maritz and Lwin(1989)). The empirical Bayes estimation procedure $T=\{t_N\}$ defined by (2.4), (2.7) and (2.8) is asymptotically optimal.

3. Estimation of the normal mean

Let $(X_1,\theta_1),\cdots,(X_N,\theta_N),\cdots$ be a sequence of independent and identically distributed random vectors, where the conditional distribution of X_i given θ_i is $\mathrm{N}(\theta_i,\sigma^2)$ for each $i=1,2,\cdots$, and $\theta_1,\theta_2,\cdots$ are i.i.d. $G\in\mathscr{C}$ with $0<\sigma^2<\infty$ known.

3.1. Nonparametric empirical Bayes estimate

Let $\mathscr{C}=\left\{G \mid \int|\theta|dG(\theta)<\infty \text{ and } G \text{ has support } [K_1,K_2]\right\}$, where K_1,K_2 are given numbers. A Bayes estimate of θ_{N+1} at $G\in\mathscr{C}$ based on the observation $X_{N+1}=x$ is given by posterior mean of θ_{N+1}, i.e.,

$$t_G(x)=\mathrm{E}[\theta_{N+1}|X_{N+1}=x]=\frac{\int\theta f_\theta(x)dG(\theta)}{\int f_\theta(x)dG(\theta)},$$

where $f_\theta(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{(x-\mu)^2}{2\sigma^2}]$, $-\infty<x<\infty$. Integration by parts applied to the numerator of the last equality gives

$$t_G(x)=x-\frac{f'_G(x)}{f_G(x)},\qquad -\infty<x<\infty, \tag{3.1}$$

where $f_G(x)=\int f_\theta(x)dG(\theta)$ is the marginal density of the X_i's and $f'_G(x)=\frac{d}{dx}f_G(x)$.

Now, $X_1,\cdots,X_N$ are N independent observations from f_G. Let $f_N(x)=f_N(X_1,\cdots,X_N;x)$, $f'_N(x)=f'_N(X_1,\cdots,X_N;x)$ denote the consistent estimates of $f_G(x)$, $f'_G(x)$ based on $X_1,\cdots,X_N$, respectively. An empirical Bayes estimates can be defined by

$$t_N(x)=t_N(X_1,\cdots,X_N;x)=x-\frac{f'_N(x)}{f_N(x)}. \tag{3.2}$$

Let $F_G(x)=\int_{-\infty}^{x}f_G(t)dt$, the marginal cumulative distribution function of X_N. Define

$$F_N(X_1,\cdots,X_N;x)=\frac{\text{number of terms } X_1,\cdots,X_N\le x}{N}, \tag{3.3}$$

the empirical distribution function determined by $X_1,\cdots,X_N$. Choose a sequence of positive numbers $\{c_N\}$ and define, for each $N=1,2,\cdots$,

$$f_N(X_1,\cdots,X_N\,;x)=\frac{F_N(X_1,\cdots,X_N\,;x+c_N)-F_N(X_1,\cdots,X_N\,;x-c_N)}{2c_N}, \tag{3.4}$$

and

$$f'_N(X_1,\cdots,X_N\,;x)=\frac{f_N(X_1,\cdots,X_N\,;x+c_N)-f_N(X_1,\cdots,X_N\,;x-c_N)}{2c_N}. \tag{3.5}$$

Remark 3.1. (i) If $c_N\to 0$, $N\cdot c_N\to\infty$ as $N\to\infty$ then $f_N(X_1,\cdots,X_N\,;x)\to f_G(x)$ in probability for all $x\in\aleph$.
(ii) If $c_N\to 0$, $N\cdot c_N^3\to\infty$ as $N\to\infty$ then $f'_N(X_1,\cdots,X_N\,;x)\to f'_G(x)$ in probability for all $x\in\aleph$.
Proof. Since $X_1,\cdots,X_N$ are independent with the distribution function F_G, $2N\cdot c_N f_N(x)\sim$ binomial $(N, F_G(x+c_N)-F_G(x-c_N))$ and

$$\mathrm{E}f_N(x)=\frac{1}{2c_N}[F_G(x+c_N)-F_G(x-c_N)],$$

$$\mathrm{var}f_N(x)=\frac{1}{4N\cdot c_N^2}[F_G(x+c_N)-F_G(x-c_N)][1-F_G(x+c_N)+F_G(x-c_N)]$$

for all x, $N=1,2,\cdots$. By the Chebyshev's inequality, for all ε

$$\begin{aligned}&\mathrm{P}\{|f_N(x)-\mathrm{E}f_N(x)|>\varepsilon\}\\ &\le\frac{\mathrm{var}f_N(x)}{\varepsilon^2}\\ &=\frac{1}{4N\cdot c_N^2\varepsilon^2}[F_G(x+c_N)-F_G(x-c_N)][1-F_G(x+c_N)+F_G(x-c_N)]\\ &=\frac{1}{2N\cdot c_N^2\varepsilon^2}f_G(x+c_N^*)[1-F_G(x+c_N)+F_G(x-c_N)]\\ &\to 0\quad\text{as } N\to\infty.\end{aligned}$$

where $|c_N^*|\le c_N$ for each $N=1,2,\cdots$. Therefore, $f_N(x)-\mathrm{E}f_N(x)\to 0$ in probability, provided that $c_N\to 0$, $N\cdot c_N\to\infty$ as $N\to\infty$. It is clear that $\mathrm{E}f_N(x)\to f_G(x)$ as $c_N\to 0$. Therefore, $f_N(x)\to f_G(x)$ in probability as $N\to\infty$, for all $x\in\aleph$. This proves (i). Proof of (ii) is similar. □

Theorem 3.2. Let $T=\{t_N\}$, where t_N is defined by (3.2), (3.3), (3.4) and (3.5). Then T is asymptotically optimal for all $G\in\mathscr{G}$.
Proof. By Remark 3.1, $t_N(x)\to t_G(x)$ in probability for all x. Since G is concentrated on the compact interval $[K_1,K_2]$, $|t_N(X_{N+1})-t_G(X_{N+1})|<|K_1|+|K_2|<\infty$. By dominated convergence theorem, $\lim_N \mathrm{E}[(t_N(X_{N+1})-t_G(X_{N+1}))^2]=0$ and by Lemma 1.2, $T=\{t_N\}$ is

asymptotically optimal at G. □

3.2. Parametric empirical Bayes estimate

Suppose that the unknown prior G is in the parametric family of normal priors $\mathscr{G} = \{N(\mu, \tau^2) \mid -\infty < \mu < \infty, 0 < \tau < \infty\}$. For $G = N(\mu, \tau^2) \in \mathscr{G}$, posterior distribution of θ_{N+1} given X_{N+1} is $N(\frac{\sigma^2\mu + \tau^2 X_{N+1}}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2})$. Therefore, a Bayes estimate of θ_{N+1} is

$$t_G(X_{N+1}) = \frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2} X_{N+1}. \tag{3.6}$$

Since σ^2 is known, we let $\sigma^2 = 1$ without loss of generality, and (3.6) can be written as

$$t_G(X_{N+1}) = \frac{1}{1+\tau^2}\mu + \frac{\tau^2}{1+\tau^2} X_{N+1}. \tag{3.7}$$

An empirical Bayes estimate corresponding to (3.7) is obtained by substituting τ^2 and μ by their appropriate estimates $\hat{\tau}_N^2$ and $\hat{\mu}_N$ based on $X_1, \cdots, X_N$. From the distributional assumptions and the properties of the conditional expectation we have the following remark.

Remark 3.3. For each $N = 1, 2, \cdots$,

$$\begin{aligned} \mathrm{E}[\frac{1}{N}\sum_{i=1}^{N} X_i] &= \mu, \\ \mathrm{E}[\frac{1}{N}\sum_{i=1}^{N} X_i^2] &= \mu^2 + \tau^2 + 1. \end{aligned} \tag{3.8}$$

Define

$$\begin{aligned} \hat{\mu}_N &= \overline{X}_N, \\ \hat{\tau}_N^2 &= \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \overline{X}_N^2 - 1 = \left((1-\frac{1}{N})S_N^2 - 1\right)^+. \end{aligned} \tag{3.9}$$

By the strong low of large numbers and Remark 3.3 we have,

Remark 3.4. $\hat{\mu}_N \to \mu$ a.s. and $\hat{\tau}_N^2 \to \tau^2$ a.s. as $N \to \infty$.

Define $T = \{t_N\}$ as

$$t_N(X_{N+1}) = \frac{1}{1+\hat{\tau}_N^2}\overline{X}_N + \frac{\hat{\tau}_N^2}{1+\hat{\tau}_N^2} X_{N+1} \tag{3.10}$$

for each $N = 1, 2, \cdots$, where $\hat{\tau}_N^2$ is given by (3.9). By Remark 3.4, the empirical Bayes estimate $t_N(X_{N+1})$ given by (3.10) converges to $t_G(X_{N+1})$ as $N \to \infty$.

Theorem 3.5.(Maritz and Lwin(1989)). $T=\{t_N\}$ in (3.10) is asymptotically optimal.

4. Concluding remarks

We have considered the nonparametric and parametric empirical Bayes estimation procedures with the binomial and normal components and proved their asymptotic optimality. Particularly in the parametric empirical Bayes estimation for the normal mean asymptotic optimality was proved without the assumption of compact support for the prior.

Estimating densities and their derivatives are essential in case of nonparametric empirical Bayes estimation. There should be more efficient ways than those used in this paper based upon the empirical distributions. In case of the parametric estimation, other procedures including the method of maximum likelihood can be compared to the method of moments employed here.

References

[1] Berger, J. O. (1980). *Statistical Decision Theory,* Springer-Verlag, New York.

[2] Billingsley, P. (1986). *Probability and Measure,* John Wiley & Sons, New York.

[3] Maritz, J. S. and Lwin, T. (1989). *Empirical Bayes Methods,* 2nd edition, Chapman and Hall, London.

[4] Robbins, H. (1955). An empirical Bayes approach to statistics, *Proceedings of the Third Berkeley Symposium on Statistics and Probability,* Vol. 1, 157-164.

[5] Robbins, H. (1964). The empirical Bayes approach to statistical decision problem, *Annals of Mathematical Statistics,* Vol. 35, 1-20.