BAYES RISKS OF ESTIMATORS OF ESTIMABLE PARAMETERS

Yamato, Hajime
Department of Mathematics, Faculty of Science, Kagoshima University

https://doi.org/10.5109/13136

出版情報：統計数理研究. 19 (1/2), pp.13-21, 1980-03. Research Association of Statistical Sciences
バージョン：
権利関係：

BAYES RISKS OF ESTIMATORS OF ESTIMABLE PARAMETERS

By

Hajime YAMATO*

(Received July 20, 1979)

Summary

Bayes risks are evaluated for estimators of an estimable parameter of degree 1 or 2, which are U-statistics, differentiable statistical functions, Bayes estimates and limits of Bayes estimates, using squared error loss and a Ferguson's Dirichlet process prior. The relations among the Bayes risks of the above estimators are obtained under certain conditions.

1. Introduction

Let (X, A) be a measurable space and Θ be the set of all distributions on (X, A). We denote the k-fold product of the measurable space (X, A) by (X^k, A^k). Let θ_1 and θ_2 be an estimable parameter of degree 1 and 2, respectively, of a distribution $P \in \Theta$. Then there exists statistics $h_1(x)$ and $h_2(x, y)$ such that

$$\theta_1 = \int_X h_1(x) dP(x) , \tag{1.1}$$

$$\theta_2 = \int_{X^2} h_2(x, y) dP(x) dP(y) . \tag{1.2}$$

For the estimable parameter of degree 2, throughout this paper, we consider θ_2 with h_2 such that $h_2(x, y) = h_2(y, x)$ and $h_2(x, x) = 0$ for any $x, y \in X$.

As estimators of estimable parameters, U-statistics and differentiable statistical functions are well known. (See, for example, Hoeffding (1948) and von Mises (1947).) For an estimable parameter of degree 1, the U-statistic is identical with the differentiable statistical function, which is given by

$$U_1 = \sum_{i=1}^{n} h_1(X_i)/n , \tag{1.3}$$

where $X_1, \cdots, X_n$ denotes a sample of size n obtained from the distribution $P \in \Theta$. As a prior distribution for a distribution in a nonparametric Bayesian problem,

* Department of Mathematics, Faculty of Science, Kagoshima University, Kagoshima 890, Japan

Ferguson (1973) introduced the Dirichlet process and applied to many problems. Against squared error loss and a Dirichlet process prior $P\in D(\alpha)$ the Bayes estimate of θ_1 is

$$\hat{\theta}_1=q_n\phi_1+(1-q_n)U_1, \tag{1.4}$$

where α is a σ-additive nonnull finite measure on (X, A), $\phi_1=\int_X h_1(x)dQ(x)$ with $Q(\cdot)=\alpha(\cdot)/\alpha(X)$ and $q_n=\alpha(X)/[\alpha(X)+n]$. The limit of Bayes estimate, which is obtained from $\hat{\theta}_1$ by letting $\alpha(X)$ tend to zero with fixed Q, is identical with U_1 (see Ferguson (1973)).

For an estimable parameter of degree 2, the U-statistic is

$$U_2=\sum_{i\neq j} h_2(X_i, X_j)/[n(n-1)] \tag{1.5}$$

and the differentiable statistical function is

$$\tilde{\theta}_2=\sum_{i\neq j} h_2(X_i, X_j)/n^2. \tag{1.6}$$

Against squared error loss and a Dirichlet process prior $P\in D(\alpha)$ the Bayes estimate of θ_2 is

$$\begin{aligned}\hat{\theta}_2=\alpha(X)[\alpha(X)+n+1]^{-1}[q_n^2\phi_2+2q_n(1-q_n)\sum_{i=1}^{n} f(X_i)/n \\ +(1-q_n)^2\sum_{i\neq j} h_2(X_i, X_j)/n^2],\end{aligned} \tag{1.7}$$

where $\phi_2=\int_{X^2} h_2(x, y)dQ(x)dQ(y)$ and $f(x)=\int_X h_2(x, y)dQ(y)$. The limit of Bayes estimate, which is obtained from $\hat{\theta}_2$ by letting $\alpha(X)$ tend to zero with fixed Q, is given by

$$\theta_2^*=\sum_{i\neq j} h_2(X_i, X_j)/[n(n+1)] \tag{1.8}$$

(see Yamato (1977a)).

The purpose of this paper is to evaluate the Bayes risks of the above estimators using squared error loss and a Dirichlet process prior and to see the relations among the Bayes risks of the estimators. For the mean, in which case $h_1(x)=x$, Korwar and Hollander (1976) evaluates the Bayes risks and obtaines the exact relation in conjunction with the empirical Bayes estimation of the mean.

In the section 2, preliminarily we shall state about the expectations of integrals with respect to a distribution P which is a Dirichlet process on (X, A) with parameter α, i. e., $P\in D(\alpha)$.

In the section 3, against squared error loss and a Dirichlet process prior the Bayes risks of the above estimators are evaluated. For the estimable parameter of degree 1, exact relation between the Bayes risks is obtained, which is analogous to the case of the mean given by Korwar and Hollander (1976), (4.4). For the estimable parameter of degree 2, the relations among the Bayes risks are obtained for a small $\alpha(X)$.

2. Preliminaries

We shall prepare the evaluations of expectations of integrals with respect to a distribution $P\in D(\alpha)$ for the section 3.

LEMMA 1 (Ferguson). *Let $P\in D(\alpha)$ and g be a measurable real-valued function defined on (X, A). If $\int_X |g(x)|\,d\alpha(x)$ is finite, $\int_X |g(x)|\,dP(x)$ is finite with probability one and*

$$E_D\int_X g(x)dP(x)=\int_X g(x)dQ(x)$$

where E_D denotes the expectation with respect to the Dirichlet process $D(\alpha)$.

The following lemmas 2 and 3 are essentially identical with the lemmas 4 and 3, respectively, of Yamato (1977a), except for the symmetry of integrand.

LEMMA 2. *Let $P\in D(\alpha)$ and g be a measurable real-valued function defined on (X^2, A^2). If $\int_{X^2}|g(x, y)|\,d\alpha(x)d\alpha(y)$ and $\int_X |g(x, x)|\,d\alpha(x)$ are finite then*

$$\int_{X^2}|g(x, y)|\,dP(x)dP(y)$$

is finite with probability one and

$$E_D\int_{X^2}g(x, y)dP(x)dP(y)$$

$$=\left[\alpha(X)\int_{X^2}g(x, y)dQ(x)dQ(y)+\int_X g(x, x)dQ(x)\right]/[\alpha(X)+1].$$

LEMMA 3. *Let $P\in D(\alpha)$ and g be a measurable real-valued function defined on (X^3, A^3). If $\int_{X^3}|g(x, y, z)|\,d\alpha(x)d\alpha(y)d\alpha(z)$, $\int_{X^2}|g(x, x, y)|\,d\alpha(x)d\alpha(y)$,*

$$\int_{X^2}|g(x, y, x)|\,d\alpha(x)d\alpha(y), \qquad \int_{X^2}g(y, x, x)|\,d\alpha(x)d\alpha(y)$$

and $\int_X |g(x, x, x)d\alpha(x)$ are finite, then $\int_{X^3}|g(x, y, z)|\,dP(x)dP(y)dP(z)$ is finite with probability one and we have

$$E_D\int_{X^3}g(x, y, z)dP(x)dP(y)dP(z)$$

$$=[(\alpha(X)+1)(\alpha(X)+2)]^{-1}\left[\alpha^2(X)\int_{X^3}g(x, y, z)dQ(x)dQ(y)dQ(y)\right.$$

$$+\alpha(X)\int_{X^2}[g(x, x, y)+g(x, y, x)+g(y, x, x)]dQ(x)dQ(y)$$

$$\left.+2\int_X g(x, x, x)dQ(x)\right].$$

The following lemma 4 needs the lemma 4 with $k=4$ of Yamato (1977b). The method of proofs of the above lemmas 2 and 3 and the following lemma 4 is essentially similar to that of the proof of the lemma 1, which is Ferguson's (1973) Theorem 3. So, the proofs of the lemma 2, 3 and 4 are omitted.

LEMMA 4. *Let* $P\in D(\alpha)$. *If* $\int_{X^2} h^2(x, y)d\alpha(x)d\alpha(y)$ *is finite then*

$$E_D\left[\int_{X^2} h_2(x, y)dP(x)dP(y)\right]^2$$

$$=\alpha(X)[(\alpha(X)+1)(\alpha(X)+2)(\alpha(X)+3)]^{-1}\left[2\int_{X^2} h_2^2(x, y)dQ(x)dQ(y)\right.$$

$$+4\alpha(X)\int_{X^3} h_2(x, y)h_2(y, z)dQ(x)dQ(y)dQ(z)$$

$$\left.+\alpha^2(X)\left[\int_{X^2} h_2(x, y)dQ(x)dQ(y)\right]^2\right].$$

In what follows we suppose the existences of the integrals; $\int_X h_1^2(x)d\alpha(x)$ and $\int_{X^2} h_2^2(x, y)d\alpha(x)d\alpha(y)$. For abbreviation, we use the following notations;

$$\phi_1=\int_X h_1(x)dQ(x), \quad \zeta_1=\int_X h_1^2(x)dQ(x), \quad \phi^2=\int_{X^2} h_2(x, y)dQ(x)dQ(y),$$

$$\zeta_2=\int_{X^2} h_2^2(x, y)dQ(x)dQ(y), \quad f(x)=\int_X h_2(x, y)dQ(y), \quad \psi=\int_X f^2(z)dQ(z).$$

Then by the lemma 1 we have for $P\in D(\alpha)$

$$E_D\int_X f(z)dP(z)=\phi_2, \tag{2.1}$$

$$E_D\int_X f^2(z)dP(z)=\psi. \tag{2.2}$$

By the lemma 2 we have, for $P\in D(\alpha)$,

$$E_L\theta_2=\alpha(X)\phi_2/[\alpha(X)+1], \tag{2.3}$$

$$E_D\int_{X^2} h_2^2(x, y)dP(x)dP(y)=\alpha(X)\zeta_2/[\alpha(X)+1], \tag{2.4}$$

$$E_D\left[\int_X f(z)dP(z)\right]^2=[\alpha(X)\phi_2^2+\psi]/[\alpha(X)+1], \tag{2.5}$$

$$E_D\int_{X^2} f(x)h_2(x, y)dP(x)dP(y)=\alpha(X)\psi/[\alpha(X)+1]. \tag{2.6}$$

The lemma 3 yields, for $P\in D(\alpha)$,

$$\begin{aligned} &E_D\int_{X^3} h_2(x, y)h_2(y, z)dP(x)dP(y)dP(z) \\ &\quad =\alpha(X)[\zeta_2+\alpha(X)\psi]/[(\alpha(X)+1)(\alpha(X)+2)], \end{aligned} \tag{2.7}$$

(2.8) $$E_D\left[\theta_2\left(\int_X f(z)dP(z)\right)\right]=\alpha(X)[2\psi+\alpha(X)\phi_2^2]/[(\alpha(X)+1)(\alpha(X)+2)]\,.$$

For $P\in D(\alpha)$ the lemma 4 is rewritten as

(2.9) $$E_D\theta_2^2=\alpha(X)[2\zeta_2+4\alpha(X)\psi+\alpha^2(X)\phi_2^2]/[(\alpha(X)+1)(\alpha(X)+2)(\alpha(X)+3)]\,.$$

3. Bayes risks

3.1 Estimable parameter of degree 1

Against squared error loss and a Dirichlet process prior $P\in D(\alpha)$, we shall denote the Bayes risks of the estimator U_1 and $\hat{\theta}_1$ by $r(\alpha, U_1)$ and $r(\alpha, \hat{\theta}_1)$, respectively. Then we have

$$r(\alpha, U_1)=E_D E_{X|P}(U_1-\theta_1)^2\,,$$

$$r(\alpha, \hat{\theta}_1)=E_D E_{X|P}(\hat{\theta}_1-\theta_1)^2\,,$$

where $E_{X|P}$ denotes the conditional expectation with respect to $X_1, \cdots, X_n$ given P.

THEOREM 1. *Against squared error loss and a Dirichlet process prior $P\in D(\alpha)$, the Bayes risk of the U-statistic U_1 is*

$$r(\alpha, U_1)=\alpha(X)(\zeta_1-\phi_1^2)/[n(\alpha(X)+1)]$$

and the Bayes risk of the Bayes estimate $\hat{\theta}_1$ is

$$r(\alpha, \hat{\theta}_1)=\alpha(X)(\zeta_1-\phi_1^2)/[(\alpha(X)+1)(\alpha(X)+n)]\,.$$

PROOF. By applying the lemmas 1 and 2 to the equation;

$$\begin{aligned}E_{X|P}(U_1-\theta_1)^2&=E_{X|P}U_1^2-\theta_1^2\\&=n^{-1}\left[\int_X h_1^2(x)dP(x)-\int_{X^2}h_1(x)h_1(y)dP(x)dP(y)\right]\end{aligned}$$

we have

$$\begin{aligned}r(\alpha, U_1)&=n^{-1}\left[\int_X h_1^2(x)dQ(x)\right.\\&\quad\left.-\left\{\alpha(X)\left[\int_X h_1(x)dQ(x)\right]^2+\int_X h_1^2(x)dQ(x)\right\}/\{\alpha(X)+1\}\right]\\&=\alpha(X)(\zeta_1-\phi_1^2)/[n(\alpha(X)+1)]\,.\end{aligned}$$

For the Bayes estimate, we have

$$\begin{aligned}E_{X|P}(\hat{\theta}_1-\theta_1)^2&=E_{X|P}[q_n(\phi_1-\theta_1)+(1-q_n)(U_1-\theta_1)]^2\\&=q_n^2(\theta_1-\phi_1)^2+(1-q_n)^2E_{X|P}(U_1-\theta_1)^2\,,\end{aligned}$$

because U_1 is an unbiased estimator of θ_1. Thus

(3.1) $$r(\alpha, \hat{\theta}_1)=q_n^2E_D(\theta_1-\phi_1)^2+(1-q_n)^2r(\alpha, U_1)\,.$$

By the lemmas 1 and 2, we have

$$E_D(\theta_1-\phi_1)^2=E_D\theta_1^2-\phi_1^2=(\zeta_1-\phi_1^2)/[\alpha(X)+1]\,. \tag{3.2}$$

The combination of (3.1) and (3.2) yields

$$r(\alpha,\,\hat\theta_1)=\alpha(X)(\zeta_1-\phi_1^2)/[(\alpha(X)+1)(\alpha(X)+n)]\,.$$

From the above theorem we have

COROLLARY 1.

$$r(\alpha,\,U_1)=(1+\alpha(X)/n)r(\alpha,\,\hat\theta_1)\,.$$

The above results with $h_1(x)=x$ are obtained by Korwar and Hollander (1976). From the corollary 1 we have the following

COROLLARY 2.

$$\lim_{\alpha(X)\to 0} r(\alpha,\,U_1)/r(\alpha,\,\hat\theta_1)=1\,,$$

except for the case that h_1 is constant almost everywhere with respect to the measure α.

From the proposition 1 of Ferguson (1973), the exceptional case of the corollary 2 leads to that h_1 is constant almost everywhere with respect to a probability distribution $P(\in D(\alpha))$ with probability one and therefore that θ_1 is constant with probability one. Thus the exceptional case of the corollary 2 is negligible.

3.2 Estimable parameter of degree 2

Against squared error loss and a Dirichlet process prior $P\in D(\alpha)$ we shall denote the Bayes risks of the estimators U_2, $\tilde\theta_2$, $\hat\theta_2$ and θ_2^* by $r(\alpha,\,U_2)$, $r(\alpha,\,\tilde\theta_2)$, $r(\alpha,\,\hat\theta_2)$ and $r(\alpha,\,\theta_2^*)$, respectively. For example, $r(\alpha,\,U_2)=E_D E_{X|P}(U_2-\theta_2)^2$. Then for the U-statistic U_2 we have the following

THOREM 2. *Against squared error loss and a Dirichlet process prior $P\in D(\alpha)$, the Bayes risk of the U-statistic U_2, which is given by* (1.5), *is*

$$\begin{aligned} r(\alpha,\,U_2)=2\alpha(X)[&(\alpha(X)+1)(\alpha(X)+2n)\zeta_2\\ &+2\alpha(X)\{(n-2)\alpha(X)-n\}\phi-(2n-3)\alpha(X)^2\phi_2^2]\\ &/[n(n-1)(\alpha(X)+1)(\alpha(X)+2)(\alpha(X)+3)]\,. \end{aligned} \tag{3.3}$$

PROOF. We have easily

$$\begin{aligned} &E_{X|P}(U_2-\theta_2)^2\\ &\quad=E_{X|P}\Big[\sum_{i\neq j}\{h_2(X_i,\,X_j)-\theta_2\}\Big]^2/[n(n-1)]^2\\ &\quad=2\Big[\int_{X^2}h_2^2(x,\,y)dP(x)dP(y)+2(n-2)\int_{X^3}h_2(x,\,y)h_2(y,\,z)dP(x)dP(x)dP(z)\\ &\qquad-(2n-3)\theta_2^2\Big]/[n(n-1)]\,. \end{aligned}$$

By taking the expectation of the above equation with respect to the Dirichlet process $D(\alpha)$ and applying (2.4), (2.7), (2.9), we have the desired form (3.3).

For the differentiable statistical function and the limit of Bayes estimate, we have

$$E_{X|P}(\tilde{\theta}_2-\theta_2)^2=[(n-1)^2E_{X|P}(U_2-\theta_2)^2+\theta_2^2]/n^2$$

$$E_{X|P}(\theta_2^*-\theta_2)^2=[(n-1)^2E_{X|P}(U_2-\theta_2)^2+4\theta_2^2]/(n+1)^2$$

By applying (2.9) and (3.3) to the expectations of the above equations with respect to the Dirichlet process $D(\alpha)$, we have

(3.4) $$r(\alpha,\tilde{\theta}_2)=\alpha(X)[2\{(n-1)(\alpha(X)+1)(\alpha(X)+2n)+n\}\zeta_2 +4\alpha(X)(n-2)\{(n-1)\alpha(X)-n\}\phi-(4n-3)(n-2)\alpha^2(X)\phi_2^2] /[n^3(\alpha(X)+1)(\alpha(X)+2)(\alpha(X)+3)]$$

(3.5) $$r(\alpha,\theta_2^*)=2\alpha(X)[\{(n-1)(\alpha(X)+1)(\alpha(X)+2n)+4n\}\zeta_2 +2\alpha(X)\{(n-1)(n-2)\alpha(X)-n(n-5)\}\phi-(2n-1)(n-3)\alpha^2(X)\phi_2^2] /[n(n+1)^2(\alpha(X)+1)(\alpha(X)+2)(\alpha(X)+3)].$$

For the Bayes estimate, we have the following

THOREM 3. *Against squared error loss and a Dirichlet process prior $P\in D(\alpha)$, the Bayes risk of the Bayes estimate $\hat{\theta}_2$, which is given by (1.7), is*

(3.6) $$r(\alpha,\hat{\theta}_2)=2\alpha(X)[(\alpha(X)+1)(\alpha(X)+2n)\zeta_2 +2\alpha(X)\{\alpha^2(X)+(n+1)\alpha(X)-n)\phi-\{2(\alpha(X)+n)+3\}\alpha^2(X)\phi_2^2] /[(\alpha(X)+n)(\alpha(X)+n+1)(\alpha(X)+1)(\alpha(X)+2)(\alpha(X)+3)].$$

PROOF. Taking the conditional expectation, with respect to $X_1, \cdots X_n$ given P, of the square of

$$\hat{\theta}_2-\theta_2=[\alpha(X)+n][\alpha(X)+n+1]^{-1}[q_n^2\{\phi_2-[\alpha(X)+1]\theta_2/\alpha(X)\} +2q_n(1-q_n)\sum_i\{f(X_i)-[\alpha(X)+1]\theta_2/\alpha(X)\}/n +(1-q_n)^2\sum_{i\neq j}\{h_2(X_i,X_j)-\theta_2\}/n^2],$$

we have the following relation;

$$E_{X|P}(\hat{\theta}_2-\theta_2)^2=[(\alpha(X)+n)(\alpha(X)+n+1)]^{-2} \times[\{([\alpha((X)+1][\alpha(X)+2n])^2-2n(n-1)(2n-3)\}\theta_2^2 -2\alpha^2(X)[\alpha(X)+1][\alpha(X)+2n]\phi_2\theta_2+\alpha^4(X)\phi_2^2 +4n\alpha^2(X)\int_X f^2(x)dP(x)+4n(n-1)\{\alpha(X)\int_X f(x)dP(x)]^2$$

$$-4n\alpha(X)[\alpha(X)+2][\alpha(X)+2n-1]\theta_2\int_{\mathcal{X}} f(x)dP(x)$$

$$+2n(n-1)\int_{\mathcal{X}^2} h_2^2(x, y)dP(x)dP(y)$$

$$+4n(n-1)(n-2)\int_{\mathcal{X}^2} h_2(x, y)h_2(y, z)dP(x)dP(y)dP(z)$$

$$+4n\alpha^3(X)\phi_2\int_{\mathcal{X}} f(x)dP(x)$$

$$+8n(n-1)\alpha(X)\int_{\mathcal{X}^2} f(x)h_2(x, y)dP(x)dP(y)].$$

By applying (2.1), (2.2), $\cdots$, (2.9) to the expectation of the above equation with respect to the Dirichlet process $D(\alpha)$, we have the desired form (3.6).

From (3.3), (3.4), (3.5) and (3.6) we have

$$\lim_{\alpha(X)\to 0} r(\alpha, U_2)/\alpha(X)=2\zeta_2/3(n-1),$$

$$\lim_{\alpha(X)\to 0} r(\alpha, \theta_2)/\alpha(X)=(2n-1)\zeta_2/3n^2,$$

$$\lim_{\alpha(X)\to 0} r(\alpha, \theta_2^*)/\alpha(X)=2\zeta_2/3(n+1),$$

$$\lim_{\alpha(X)\to 0} r(\alpha, \hat{\theta}_2)/\alpha(X)=2\zeta_2/3(n+1),$$

where in the operation of limits the probability measure Q is fixed. Thus for a sufficiently small $\alpha(X)$ and $n\geqq 2$, we have

$$r(\alpha, \hat{\theta}_2)\leqq r(\alpha, \theta_2^*)<r(\alpha, \tilde{\theta}_2)<r(\alpha, U_2),$$

(the first inequality is due to the optimality of $\hat{\theta}_2$) and

$$\lim_{\alpha(X)\to 0} r(\alpha, \theta_2^*)/r(\alpha, \hat{\theta}_2)=1, \tag{3.7}$$

where in the operation of limit the probability measure Q is fixed. In the above (3.7), the exceptional case is the one that $h_2=0$ almost everywhere with respect to the measure α. Because from the inequalities $0\leqq\phi\leqq\zeta$, $\phi_2^2\leqq\zeta_2$ and the inequality $\alpha^2(X)+(n+1)\alpha(X)-n<0$ for a sufficiently small $\alpha(X)$ with any fixed positive integer n, we have

$$r(\alpha, \hat{\theta}_2)\geqq 2\alpha(X)\alpha(X)+2n)\zeta_2$$
$$/[(\alpha(X)+n)(\alpha(X)+n+1)(\alpha(X)+1)(\alpha(X)+2)(\alpha(X)+3)].$$

The right hand side equals to 0 if and only if $\zeta_2=0$, that is $h_2=0$ almost everwhere with respect to the measure α, in which case $r(\alpha, \hat{\theta}_2)$ also equals to 0. By the same reason as the one following the corollary of the theorem 1, the exceptional case in (3.7) leads to $\theta_2=0$ with probability one and therefore it is negligible.

Acknowlegements

The author is grateful to Professor A. Kubo for his encouragement and advice.

References

[1] FERGUSON, T. S., *A Bayesian analysis of some nonparametric problems.* *Ann. Statist.* **1** (1973) 209-230.

[2] HOEFFDING, W., *A class of statistics with asymptotically normal distribution.* *Ann. Math. Statist.* **19** (1948), 293-325.

[3] KORWAR, R.M. and HOLLANDER, M., *Empirical Bayes estimation of a distribution function.* *Ann. Statist.* **4** (1976), 581-588.

[4] VON MISES, R., *On the asymptotic distribution of differentiable statistical functions.* *Ann. Math. Statist.* **18** (1947), 309-348.

[5] YAMATO, H., *Relations between limiting Bayes estimates and the U-statistics for estimable parameter of degrees 2 and 3.* *Commun. Statist.-Theory Methods.* **6** (1977a), 55-66.

[6] YAMATO, H., *Relations between limiting Bayes estimates and U-statistics for estimable parameters.* *J. Japan Statist. Soc.* **7** (1977b), 57-66.