Unbiasedness and Bayes Estimators

Siamak Noorbaloochi
Center for Chronic Disease Outcomes Research
Minneapolis VA Medical Center
and
Glen Meeden[1]
School of Statistics
University of Minnesota[2]

A simple geometric representation of Bayes and unbiased rules for squared error loss is provided. Some orthogonality relationships between them and the functions they are estimating are proved. Bayes estimators are shown to be behave asymptotically like unbiased estimators.

Key Words: Unbiasedness, Bayes estimators, squared error loss and consistency

---

[1]Research supported in part by NSF Grant DMS 9971331

[2]Glen Meeden
School of Statistics
313 Ford Hall
224 Church ST S.E.
University of Minnesota
Minneapolis, MN 55455-0460

1 Introduction

Let X be a random variable, possible vector valued, with a family of possible probability distributions indexed by the parameter $\theta \in \Theta$. Suppose γ, some real-valued function defined on Θ, is to be estimated using X. An estimator δ is said to be unbiased for γ if $E_\theta \delta(X) = \gamma(\theta)$ for all $\theta \in \Theta$. Lehmann (1951) proposed a generalization of this notion of unbiasedness which takes into account the loss function for the problem. Noorbaloochi and Meeden (1983) proposed a generalization of Lehmann's definition but which depends on a prior distribution π for θ. Assuming squared error loss let the Bayes risk of an estimator δ for estimating γ be denoted by $r(\delta, \gamma; \pi)$. Then under their definition δ is unbiased for estimating γ for the prior π if

$$r(\delta, \gamma; \pi) = \inf_{\gamma'} r(\delta, \gamma'; \pi)$$

Under very weak assumptions it is easy to see that this definition reduces to the usual one. See their paper for further details.

In this note we observe that the prior π can be used to define a norm over all square integrable real-valued functions of X and θ. Under this norm some orthogonality relationships between Bayes and unbiased estimators and the functions they are estimating are proved. It is then demonstrated that Bayes estimators behave asymptotically like unbiased estimators.

2 Notation

We begin by introducing some additional notation. Let μ be some measure on the sample space of X and ν a measure on Θ. Let $p_\theta(\cdot)$ be the density of X with respect to μ for the parameter θ. Let π be the prior density for θ with respect to ν. Let

$$\mathcal{H}_\pi = \{h(x, \theta) : \int \int h^2(x, \theta) p_\theta(x) \pi(\theta) d\mu \, d\nu < \infty\}$$

the space of square-integrable $(\mu \times \nu)$ real-valued functions defined on $\mathcal{X} \times \Theta$. Note $\mathcal{H}_\pi$ becomes a Hilbert space when it is equipped with the inner product

$$(h_1, h_2) = \int \int h_1(x, \theta) h_2(x, \theta) p_\theta(x) \pi(\theta) d\mu \, d\nu$$

Let $\|h\|_\pi = \sqrt{(h,h)}$ denote the norm of h. We include the subscript π to remind us that $\mathcal{H}_\pi$ does depend on π.

There are two linear subspaces of $\mathcal{H}_\pi$ which are important for us. The first is

$$\Gamma_\pi = \{\gamma(\theta) : \int \gamma^2(\theta)\pi(\theta)d\nu < \infty\}$$

which is the set of all square integrable functions of θ. The second is

$$\Delta_\pi = \{\delta(x) : \int \delta^2(x)m(x)d\mu < \infty\}$$

where $m(x) = \int p_\theta(x)\pi(\theta)d\nu$. This is the linear subspace of all square integrable functions of X.

Note with this notation we have that

$$\|\delta - \gamma\|_\pi^2 = r(\delta, \gamma; \pi)$$

Let γ be a fixed member of Γ_π and suppose that U some member of Δ_π is an unbiased estimator of γ. Then by Noorbalhoochi and Meeden (1983) it follows that

$$\|U - \gamma\|_\pi^2 = \inf_{\gamma' \in \Gamma_\pi} \|U - \gamma'\|_\pi^2$$

That is γ is the projection of U unto Γ_π. While on the other hand if δ_π is the Bayes estimator of γ then

$$\|\delta_\pi - \gamma\|_\pi^2 = \inf_{\delta \in \Delta_\pi} \|\delta - \gamma\|_\pi^2$$

and so δ_π is the projection of γ unto Δ_π and its Bayes risk is finite. Finally δ_π is the unbiased estimator of some function of θ, say $\alpha(\theta)$. Then we also must have that

$$\|\delta_\pi - \alpha\|_\pi^2 = \inf_{\gamma' \in \Gamma_\pi} \|\delta_\pi - \gamma'\|_\pi^2$$

since α is the projection of δ_π unto Γ_π. These relations are represented in the diagram.

put the diagram about here

3 The Results

We begin with some relationships which are suggested by the diagram.

Theorem 1. *Let $U \in \Delta_\pi$ be an unbiased estimator of $\gamma \in \Gamma_\pi$. Let δ_π be the Bayes estimator of γ and suppose δ_π is an unbiased estimator of α. Then the Pythagorean equality holds for the Bayes risk of an unbiased estimator and of the Bayes estimator, that is*

$$\|U\|_\pi^2 = \|U - \gamma\|_\pi^2 + \|\gamma\|_\pi^2 \tag{1}$$
$$\|\delta_\pi\|_\pi^2 = \|\delta_\pi - \alpha\|_\pi^2 + \|\alpha\|_\pi^2 \tag{2}$$

Also there is an orthogonal partition of the Bayes risk of an unbiased estimator and of the Bayes estimator, that is

$$\|U - \gamma\|_\pi^2 = \|\delta_\pi - \gamma\|_\pi^2 + \|U - \delta_\pi\|_\pi^2 \tag{3}$$
$$\|\delta_\pi - \gamma\|_\pi^2 = \|\delta_\pi - \alpha\|_\pi^2 + \|\gamma - \alpha\|_\pi^2 \tag{4}$$

The proof of the theorem is immediate by inspection of the diagram. However these results are also easy to prove more formally. For example equation 4 follows by adding and subtracting α in the left hand side and then squaring out the term. The fact that the cross product term is zero follows by conditioning on θ and using the fact that δ_π is unbiased for α. This result is just a Bayesian version of the usual partitioning of mean squared error into variance and bias squared.

Two more general orthogonality relationships are also evident from the diagram. That is

$$(U - \gamma, \gamma') = 0 \text{ for all } \gamma' \in \Gamma_\pi$$
$$(\delta_\pi - \gamma, \delta) = 0 \text{ for all } \delta \in \Delta_\pi$$

It is well known that except in very special circumstances an estimator cannot be both unbiased and Bayes for estimating γ. See for example Blackwell and Girshick (1954) and Bickel and Blackwell (1967). The result is clear in the present setup.

Theorem 2. *If δ_π, the Bayes estimator of γ, is also an unbiased estimator of γ then its Bayes risk is zero.*

Proof. Recall that

$$\|\gamma\|_\pi^2 = \|\delta_\pi - \gamma\|_\pi^2 + \|\delta_\pi\|_\pi^2$$

Now since by assumption $\alpha = \gamma$ we can substitute equation 2 of the previous theorem for $\|\delta_\pi\|_\pi^2$ in the above equation and the result follows. □

Intuitively a Bayesian with a Bayes risk equal to zero believes that they have prefect information about the parameter. In the diagram this means that there is zero distance between the two spaces Γ_π and Δ_π. This does not happen except in very special circumstances. However asymptotically this happens in many statistical problems and suggests that if there is a consistent unbiased estimator then Bayes estimators should behave asymptotically as the unbiased estimator.

Theorem 3. *Suppose that given θ $X_1, \ldots, X_n$ are independent and identically distributed as the random variable X. For a sample of size one let $U \in \Delta_\pi$ be an unbiased estimator of $\gamma \in \Gamma_\pi$. Let $U_n = \sum_{i=1}^n U(X_i)/n$ and $\delta_{\pi,n}$ be the Bayes estimator of γ based on the sample of size n. Then*

i. $\|\delta_{\pi,n} - \gamma\|_{\pi,n}^2 \to 0$ *as* $n \to \infty$

ii. $\|U_n - \delta_{\pi,n}\|_{\pi,n}^2 \to 0$ *as* $n \to \infty$

Proof. Let τ be the Bayes risk of U for estimating γ when $n = 1$. Then the Bayes risk of U_n is just τ/n. But the Bayes risk for $\delta_{\pi,n}$ for estimating γ is no greater than the Bayes risk of U_n so part i follows. Now part ii follows from part i and equation 3 of Theorem 1 for the sample size n problem because both of the terms involving γ go to zero as $n \to \infty$. □

The second part of the theorem implies for large n that a Bayesian whose prior is π believes with high probability that their estimator will be close to the unbiased estimator. Note another Bayesian with a different prior believes the same thing. So they both expect agreement as the sample size increases. This result is somewhat in the spirit of one in Blackwell and Dubins (1962).

Most of the standard arguments for the consistency of Bayes estimators start with considering the asymptotic distribution of the posterior distribution, see for example Johnson (1970). These arguments are closely related to the asymptotic behavior of the maximum likelihood estimator and need to make various assumptions about the probability model. For a discussion of this point and more references see O'Hagan (1994). Here however we tie the asymptotic behavior of the Bayes estimator to that of an unbiased estimator.

Our lack of assumptions and the relative simplicity of the argument comes at the cost of being unable to say anything directly about the asymptotic behavior of the posterior distribution. We close by showing that in some cases it is possible to say something more about the posterior.

Suppose Θ is a set of real numbers. Given $\theta \in \Theta$, $X_1, \ldots, X_n$ are independent and identically distributed copies of X. Suppose $\gamma \in \Gamma_\pi$ is a strictly increasing function with an unbiased estimator $U \in \Delta_\pi$. Let a be a fixed real number and $I_a(z) = 1$ when $z \leq a$ and 0 otherwise. Let

$$U_n = \sum_{i=1}^{n} U(X_i)/n, \quad W_n = I_a(U_n),$$
$$\phi(\theta) = I_a(\gamma(\theta)) \text{ and } \phi_n(\theta) = E_\theta(W_N) = P_\theta(U_n \leq a)$$

Let π be a prior whose support is Θ and $\delta_{\pi,n}$ and $\delta^*_{\pi,n}$ be the Bayes estimators of ϕ_n and ϕ respectively. We have

$$\|\phi_n - \phi\|^2_{\pi,n} \to 0 \quad \text{and} \quad \|W_n - \phi\|^2_{\pi,n} \to 0 \tag{5}$$

as $n \to \infty$. The first equation follows from the fact that $\phi_n(\theta) \to \phi(\theta)$ for $\theta \neq \gamma^{-1}(a)$. To see the second first note that

$$\begin{aligned} E_\theta(W_n - \phi(\theta))^2 &= E_\theta(W_n) - 2\phi(\theta)E_\theta(W_n) + \phi(\theta) \\ &= P_\theta(U_n \leq a) - 2I_a(\gamma(\theta))E_\theta(W_n) + I_a(\gamma(\theta)) \end{aligned}$$

and so

$$\begin{aligned} E_\theta(W_n - \phi(\theta))^2 &= 1 - P_\theta(U_n \leq a) \quad &\text{if} \quad \gamma(\theta) \leq a \\ &= P_\theta(U_n \leq a) \quad &\text{if} \quad \gamma(\theta) > a \end{aligned}$$

From this we have that $E_\theta(W_n - \phi(\theta))^2 \to 0$ if $\gamma(\theta) \neq a$ and the result follows. The next theorem extends the results of Theorem 3.

Theorem 4. *Assuming the above setup we have that*

i. $\|\delta^*_{\pi,n} - \phi\|^2_{\pi,n} \to 0$ *as* $n \to \infty$

ii. $\|W_n - \delta^*_{\pi,n}\|^2_{\pi,n} \to 0$ *as* $n \to \infty$

Proof. Note since

$$\|W_n - \delta^*_{\pi,n}\|^2_{\pi,n} \leq \|W_n - \phi\|^2_{\pi,n} + \|\phi - \delta^*_{\pi,n}\|^2_{\pi,n}$$

part *ii* follows from part *i* and the second part of equation 5.

To prove part *i* we note that

$$\begin{aligned}
\|\delta^*_{\pi,n} - \phi\|^2_{\pi,n} &\le \|\delta_{\pi,n} - \phi\|^2_{\pi,n} \\
&\le \|\delta_{\pi,n} - \phi_n\|^2_{\pi,n} + \|\phi_n - \phi\|^2_{\pi,n} \\
&\le \|W_n - \phi_n\|^2_{\pi,n} + \|\phi_n - \phi\|^2_{\pi,n} \\
&\le \|W_n - \phi\|^2_{\pi,n} + 2\|\phi_n - \phi\|^2_{\pi,n} \qquad (6)
\end{aligned}$$

where the first and third inequalities follow because $\delta^*_{\pi,n}$ is Bayes for ϕ and $\delta_{\pi,n}$ is Bayes for ϕ_n. But equation 6 goes to 0 by equation 5 and the proof is complete. □

Now $\delta^*_{\pi,n}$ is just the posterior probability of the event that $\gamma(\theta) \le a$. Hence part *ii* of Theorem 4 implies that a Bayesian whose prior is π believes with high probability their posterior probability of this event will be close to W_n which does not depend on their prior. So once again two different Bayesian would expect agreement as the sample size increases.

References

Bickel, P. J. and Blackwell, D. (1967). A note on Bayes estimates. *Annals of Mathematical Statistics*, 38:1907–1911.

Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. *Annals of Mathematical Statistics*, 33:882–886.

Blackwell, D. and Girshick, M. A. (1954). *Theory of games and statistical decisions*. John Wiley, New York.

Jonhson, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Annals of Mathematical Statistics*, 41:851–854.

Lehmann, E. L. (1951). A general concept of unbiasedness. *Annals of Mathematical Statistics*, 22:587–592.

Noorbaloochi, S. and Meeden, G. (1983). Unbiasedness as the dual of being Bayes. *Journal of the American Statistical Association*, 78:619–623.

O'Hagen, A. (1994). *Kendal's advanced theory of statistics Volume 2b Bayesian inference*. Halsted Press, New York.

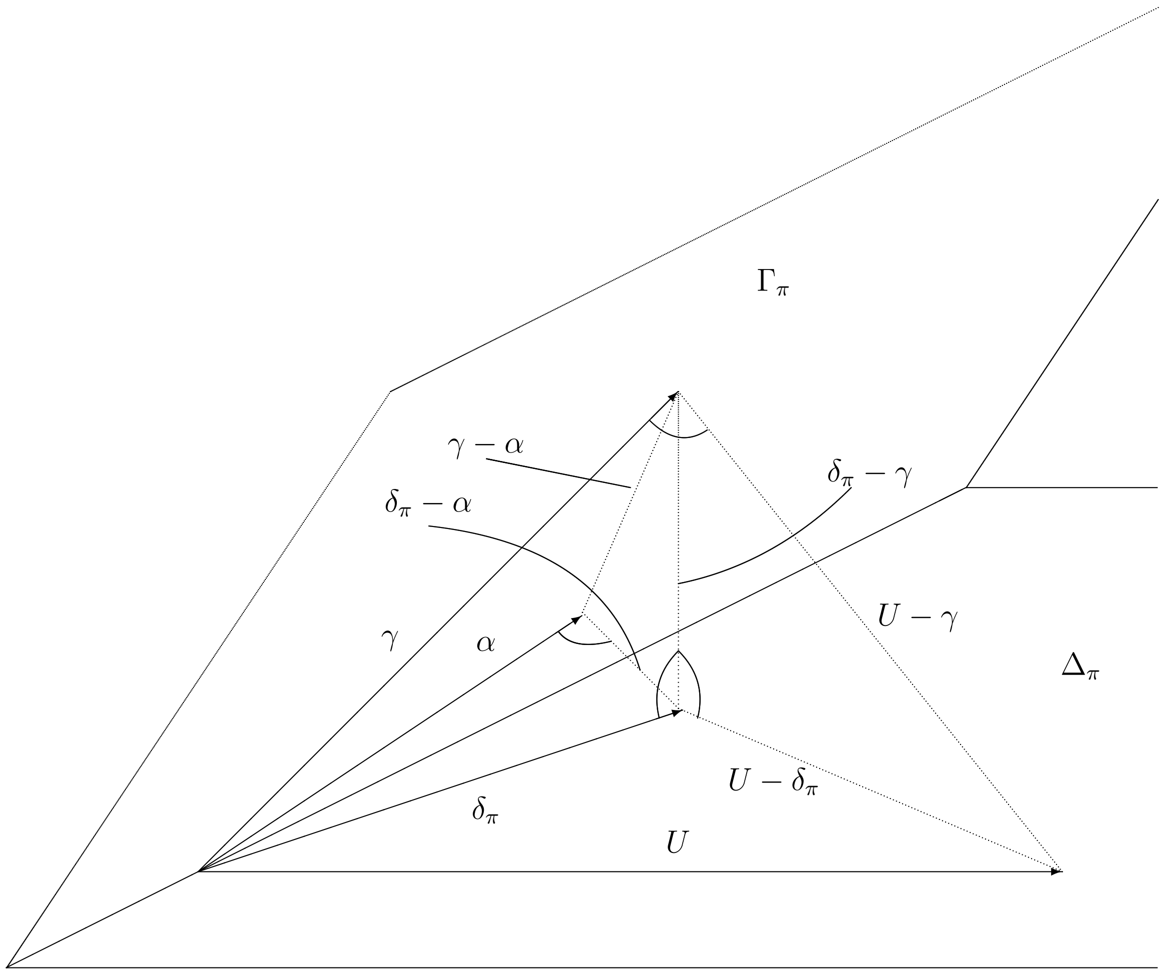

Figure 1: The linear subspaces Γ_π and Δ_π