On a proper Bayes, but inadmissible estimator [1]

PANKAJ BHAGWAT[a,b], ÉRIC MARCHAND[a]
a Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1

b IISER Pune, Pune, India 411008 (e-mails: pankaj.bhagwat@students.iiserpune.ac.in; eric.marchand@usherbrooke.ca)

SUMMARY

We present an example of a proper Bayes point estimator which is inadmissible. It occurs for a negative binomial model with shape parameter a, probability parameter p, prior densities of the form $\pi(a,p) = \beta\, g(a)\,(1-p)^{\beta-1}$, and for estimating the population mean $\mu = a(1-p)/p$ under squared error loss. Other intriguing features are exhibited such as the constancy of the Bayes estimator with respect to the choice of g, including degenerate cases or known a cases.

AMS 2010 subject classifications: 62C10l 62C15, 62F10; 62F15

Keywords and phrases: Bayes estimator; Inadmissibility; Negative Binomial

1. Introduction

Bayesian methods are intimately linked to statistical decision theory, they possess desirable theoretical properties, such as coherence and, in general, good frequentist risk properties. However, proper Bayes estimators need not be admissible. So much and more is known (see for instance Berger, 1985; Lehmann, 1983; and the discussion following Theorem 2.1 below). Nonetheless, such examples remain surprising and instructive, especially when they occur in simple situations that are also relevant in practice. We report and comment on such a situation that occurred recently in studying Bayesian posterior analysis for a negative binomial model. Moreover, the example which we present exhibits other intriguing features, such as the constancy of a Bayes point estimator with respect to a large and dispersed class of priors.

As implied by the following well-known result (e.g., Ferguson, 1968; Lehmann, 1983), the inadmissibility of a proper Bayes estimator can only occur when the Bayes risk is infinite.

Lemma 1.1. *Consider model $X \sim p_\theta$, $\theta \in \Theta$, and the problem of estimating $\tau(\theta)$ under loss $L(\theta,\delta)$. Let $\delta_\pi(X)$ be a unique* [2] *Bayes estimator for a proper prior density π defined with respect to a $\sigma-$finite measure τ. Then, the estimator $\delta_\pi(X)$ is admissible when $r_\pi(\delta_\pi) < \infty$, with $r_\pi(\delta)$ the Bayes risk of δ given by*

$$r_\pi(\delta) = \int_\Theta \mathbb{E}\, L\left(\theta, \delta(x)\right)\, \pi(\theta)\, d\tau(\theta)\,.$$

[1] November 6, 2018
[2] Up to an equivalence

2. The example

Let $X_1, \ldots, X_n$, $n \geq 2$, be independently distributed NBI(a, p), $a > 0, p \in (0, 1)$, with common marginal probability mass function

$$\mathbb{P}(X_i = t) \;=\; \frac{(a)_t}{t!}\, p^a\, (1-p)^t\, \mathbb{I}_{\mathbb{N}}(t)\,, i = 1, \ldots, n, \tag{2.1}$$

with ascending factorial $(a)_t = a(a+1)\ldots(a+t-1)$ for $t > 1$, $(a)_0 = 1$. We take both a and p to be both unknown. The negative binomial model is one of the better known and appealing models for count data, in particular for over-dispersed data, with mean lower the variance, as

$$\mathbb{E}(X_1) \;=\mu\; = a\,\frac{1-p}{p} < \sigma^2 \;=\; \mathbb{V}(X_1) \;=\; a\,\frac{1-p}{p^2}\,,$$

for all $a > 0, p \in (0, 1)$. Now, consider estimating μ under squared error loss $(\delta - \mu)^2$ with Bayesian estimators given by $\mathbb{E}(\mu|x_1, \ldots, x_n)$ as long as $\mathbb{E}(\mu^2|x_1, \ldots, x_n) < \infty$ for all $x_1, \ldots, x_n$. Our main example is the following and relates to a joint prior for (a, p) which factorizes into independent components $a \sim g$ and $p \sim Beta(1, \beta)$.

Theorem 2.1. *Consider $X_1, \ldots, X_n \sim$ NBI(a,p) as in (2.1) with prior density*

$$\pi_g(a, p) \;=\; \beta\, g(a)(1-p)^{\beta-1}\,,$$

with $\beta > 0$ and g being a density relative to $\sigma-$finite measure ν such that $\int_1^\infty g(a)d\nu(a) \;=\; 1$, and such that all positive integer moments exist, i.e., $\int_1^\infty a^m\, g(a)\, d\nu(a) \;<\; \infty$, for $m = 1, 2, \ldots$. Then, the Bayes estimator of $\mu\; = a\,(1-p)/p$ for loss $(\delta - \mu)^2$, and with respect to prior $\pi_g(a, p)$ is given by $\delta_{\pi_g}(x_1, \ldots, x_n) \;=\; \frac{\sum_{i=1}^n x_i + \beta}{n}$, irrespective of g. Furthermore, δ_{π_g} is inadmissible and dominated by the unbiased estimator $\delta_0(x_1, \ldots, x_n) \;=\; \frac{\sum_{i=1}^n x_i}{n}$.

Proof. The inadmissibility of δ_{π_g} is immediate as $\mathbb{V}(\delta_{\pi_g}) \;=\; \mathbb{V}(\delta_0) \;=\; \sigma^2/n$, and δ_{π_g} is positively biased for μ as opposed to δ_0 which is unbiased for μ.

Now, for evaluating δ_{π_g}, we have the likelihood

$$p(x_1, \ldots, x_n|a, p) \;=\; \left\{\prod_i \frac{(a)_{x_i}}{x_i!}\right\}\, p^{na}\,(1-p)^{\sum_i x_i}\, \prod_i \mathbb{I}_{\mathbb{N}}(x_i)$$

yielding the posterior density

$$(a, p)|x_1, \ldots, x_n \;\sim\; g(a)\, p^{na}(1-p)^{\sum_i x_i + \beta - 1}\,; a \geq 1, 0 < p < 1\,.$$

From this, we obtain

$$
\begin{aligned}
\mathbb{E}(\mu|x_1,\ldots,x_n) &= \mathbb{E}(a\,\frac{1-p}{p}|x_1,\ldots,x_n) \\
&= \frac{\int_1^\infty a\,\prod_i (a)_{x_i}\, g(a)\int_0^1 p^{na-1}(1-p)^{\sum_i x_i+\beta} dp\, da}{\int_1^\infty \prod_i (a)_{x_i}\, g(a)\int_0^1 p^{na}(1-p)^{\sum_i x_i+\beta-1} dp\, da} \\
&= \frac{\int_1^\infty a\,\prod_i (a)_{x_i}\, g(a)\,\frac{\Gamma(na)\,\Gamma(x_i+\beta+1)}{\Gamma(na+\sum_i x_i+\beta+1)}\, da}{\int_1^\infty \prod_i (a)_{x_i}\, g(a)\,\frac{\Gamma(na+1)\,\Gamma(\sum_i x_i+\beta)}{\Gamma(na+\sum_i x_i+\beta+1)}\, da} \\
&= \left(\frac{\sum_i x_i\,+\,\beta}{n}\right)\frac{\int_1^\infty \prod_i (a)_{x_i}\, g(a)\,\{\Gamma(na+\sum_i x_i+\beta+1)\}^{-1}\, da}{\int_1^\infty \prod_i (a)_{x_i}\, g(a)\,\{\Gamma(na+\sum_i x_i+\beta+1)\}^{-1}\, da} \qquad (2.2)\\
&= \frac{\sum_i x_i\,+\,\beta}{n},
\end{aligned}
$$

as stated. In the above, the integrals in (2.2) do exist since $\prod_i (a)_{x_i}$ is a polynomial of degree $t = \sum_i x_i$, while $\Gamma(na+\sum_i x_i+\beta+1) > (na+\sum_i x_i)!$ with $(na+\sum_i x_i)!$ also a polynomial of degree t.

To conclude the proof, we require the posterior variance $\mathbb{V}(\mu|x_1,\ldots,x_n)$ to exist and we show that this to be the case under the given conditions on g. Indeed, proceeding with a decomposition as above, we have

$$
\begin{aligned}
\mathbb{E}(\mu^2|x_1,\ldots,x_n) &= \mathbb{E}(a^2\,\frac{(1-p)^2}{p^2}|x_1,\ldots,x_n) \\
&= \frac{\int_1^\infty a^2\,\prod_i (a)_{x_i}\, g(a)\int_0^1 p^{na-2}(1-p)^{\sum_i x_i+\beta+1} dp\, da}{\int_1^\infty \prod_i (a)_{x_i}\, g(a)\int_0^1 p^{na}(1-p)^{\sum_i x_i+\beta-1} dp\, da} \\
&= (\sum_i x_i+\beta)\,(\sum_i x_i+\beta+1)\,\frac{\int_1^\infty a^2\,\prod_i (a)_{x_i}\, g(a)\,\frac{\Gamma(na-1)}{\Gamma(na+\sum_i x_i+\beta+1)}\, da}{\int_1^\infty a^2\,\prod_i (a)_{x_i}\, g(a)\,\frac{\Gamma(na+1)}{\Gamma(na+\sum_i x_i+\beta+1)}\, da},
\end{aligned}
$$

which is finite since (i) both $\frac{\Gamma(na\pm1)}{\Gamma(na+\sum_i x_i+\beta+1)}$ are bounded by 1, and (ii) $\int_1^\infty a^2 (a)_{x_i}\, g(a)\, da \leq \int_1^\infty a^2 (a+\max_i x_i)^n\, g(a)\, da < \infty$, with the finiteness moment condition on the density g. □

Theorem 2.1 exhibits the motivating purpose and the main feature of this note. It adds to a small collection of known examples where a proper Bayes estimator is inadmissible. Earlier examples appear in Lehmann (1983, page 270) with a Gamma model and inverse-Gamma prior, as well as in Berger (1985) (see Robert, 2001, Section 8.2, who reports on both of these examples). In the latter case, one takes $X|\theta \sim N(\theta,1)$, prior $\theta \sim N(0,1)$ and weighted squared-error loss $e^{3\theta^2/4}\,(\delta-\theta)^2$ for estimating θ. Calculations yield the Bayes estimator $\delta_\pi(X) = \mathbb{E}(e^{3\theta^2/4}\,\theta|X)/\mathbb{E}(e^{3\theta^2/4}\,|X) = 2X$. Clearly, with larger bias in absolute value and larger variance than the unbiased estimator $\delta_0(X) = X$, the frequentist risk of $\delta_\pi(X)$ is quite poor, as represented by the ratio of risks $\frac{R(\theta,\delta_\pi)}{R(\theta,\delta_0)} = (4+\theta^2)$.

Theorem 2.1 does exhibit other surprises though. Although both the posterior distribution and variance of $\mu = a(1-p)/p$ do depend on the choice of g for the prior $(a,p) \sim \beta\, g(a)(1-p)^{\beta-1}$, the posterior expectation obtained is independent of g, and whether or not the prior is discrete or continuous for instance. Moreover, it has to be the case for degenerate a, in other words cases where a is known and $p \sim Beta(1,\beta)$. Indeed, this can verified directly by deriving the posterior

$p|x_1,\ldots,x_n \sim \text{Beta}(na+1, \sum_i x_i + \beta)$, from which one obtains $\mathbb{E}(a(1-p)/p \mid x_1,\ldots,x_n) = \frac{\sum_i x_i+\beta}{n}$. Despite its greater simplicity, to our knowledge, this known a case result has not been reported on before. The Bayes estimator here is still inadmissible and dominated by $\bar{X}$ as an estimate of $\mu = a(1-p)/p$. Furthermore, the unbiased estimator $\bar{X}$ is itself inadmissible under squared error loss and dominated by $\frac{na}{na+1}\bar{X}$ (Ferguson, 1968, problem 12, page 86).

Remark 2.1. *If one happens to derive the result of Theorem 2.1 for the known $a > 1$ case, one then can see that Theorem 2.1's expression for the Bayes estimator will hold since*

$$\mathbb{E}(\mu|X_1,\ldots,X_n) = \mathbb{E}^{a|X_1,\ldots,X_n}\{\mathbb{E}\,(a(1-p)/p|X_1,\ldots,X_n,a) = \mathbb{E}^{a|X_1,\ldots,X_n}\{\frac{\sum_i X_i+\beta}{n}\} = \frac{\sum_i X_i+\beta}{n}.$$

3. Concluding Remarks

We have provided an original example of a proper Bayes estimator which is inadmissible. It actually arose in a perfectly natural setting, part of an ongoing study of Bayesian inference for a negative binomial model. In such a case, as well as for the earlier known examples in the literature, the Bayes estimators are obtained in a coherent manner from a proper prior and by making trustworthy inferences from the posterior distribution. As described by Berger (1985, Section 4.8.1), such an unsettling or paradoxical situation, which can only possibly happen when the Bayes risk is infinite (Lemma 1.1), is alleviated with the use of a bounded loss function for which the Bayes risk cannot be infinite.

Still, the surprise persists given the deep connections in statistical decision theory between complete classes of estimators and Bayes estimators. In the other direction, it has long been known that the collection of proper Bayes estimators is not large enough to generally contain all admissible estimators and, typically, one requires the inclusion of some generalized Bayes estimators. A well known example arises for the multivariate normal model with X for $X \sim N_p(\theta, I_p)$ and squared error loss $\|d-\theta\|^2$ for estimating θ. Here, X is generalized Bayes for the improper prior density $\pi(\theta) = 1$, admissible for $p = 1, 2$, but inadmissible for $p \geq 3$. There exist many deep findings related to the admissibility of generalized Bayes estimators (e.g., Rukhin, 1975, as well as all the references below, among others).

Acknowledgements

Both authors are thankful to the Mitacs Global Link program which supported a visit of Pankaj Bhagwat to the Université de Sherbrooke, and made possible this collaborative effort. Author Marchand gratefully acknowledges research support of the Natural Sciences and Engineering Research Council of Canada.

References

[1] Berger, J. (1985). *Statistical Decision Theory and Related Topics.* Springer Texts in Statistics. Springer-Verlag, New York. Second edition.

[2] Ferguson, (1968). *Mathematical Statistics. A decision-theoretic approach.* Academic Press. San Diego.

[3] Lehmann, E.L. (1983). *Theory of point estimation.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

[4] Robert, C. (2001). *The Bayesian choice. From decision-theoretic foundations to computational implementation.* Translated and revised from the French original by the author. Second edition. Springer Texts in Statistics. Springer-Verlag, New York.

[5] Rukhin, A. (1995). Admissibility: Survey of a concept in progress. *International Statistical Review*, **63**, pp. 95-115