IMS Collections
Contemporary Developments in Bayesian Analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman
Vol. 8 (2012) 131–153

DOI: 10.1214/11-IMSCOLL810

Multiagent estimators of an exponential mean

Constance van Eeden and James V. Zidek*

University of British Columbia

Abstract: Some Bayesian agents must produce a joint estimator of the mean of an exponentially distributed random variable S from a sample of realizations S. Their priors may differ but they have the same utility function. For the case of two agents, the Pareto efficient boundary of the utility set generated by the class of all non-randomized linear estimation rules is explored in this paper. Conditions are given that make those rules G-complete within the class of non-randomized linear estimators, meaning that optimum non-random estimators can be found on the Pareto boundary thereby providing a basis for a meaningful consensus.

Contents

1 Introduction . . . 131
2 A single agent . . . 132
3 The multiagent prediction problem . . . 136
4 Consensual choice . . . 141
5 Multiagent estimation . . . 144
6 Discussion . . . 146
A Appendix . . . 147
References . . . 152

1. Introduction

This paper extends the work of van Eeden and Zidek (1994) on the multiagent decision problem of estimating the mean λ of an exponentially distributed random variable S by a group of Bayesian decision makers. Here unlike the earlier work, the size of that group G is finite. However, the setup is otherwise identical to that in the earlier work whose salient results will be included in Section 2. However the nature of the results here differs markedly from the earlier ones.

The need to estimate λ arises in diverse areas of application, notably in environmental risk management. There λ can be the intensity of the homogeneous Poisson process with a random inter–event time S, the events being the exceedance of a threshold by a harzardous substance. Such an application is made by Tobías and

Department of Statistics, University of British Columbia, 6356 Agriculture Rd, Vancouver, BC, Canada, V6T 1Z2, e-mail: vaneeden@stat.ubc.ca, jim@stat.ubc.ca

*The research reported in this article was partially supported by the Natural Science and Engineering Research Council of Canada.

AMS 2000 subject classifications: Primary 62C10, 62C15; secondary 62F10

Keywords and phrases: conjugate utility, exponential–mean estimation, forecasting, G–admissibility, multiagent decision theory, Pareto optimality

Scotto (2005) where the event is the exceedance of a regulatory standard by the ozone field over Barcelona. They estimate the return period for such an exceedance for a randomly selected year, a quantile which they find to be a linear function of λ. Alternatively S could be the size of such an exceedance whose distribution can be approximated by an exponential distribution with mean λ when that threshold is sufficiently high.

Although Tobías and Scotto (2005) consider only the case of a single (non–Bayesian) agent, panels are commonly used in assessing regulatory standards. The method described in this paper is viewed as a normative rather than a descriptive approach by which such a panel of G Bayesians might ideally be used to select such an estimator and an imperative to make rational joint decision seems a reasonable requirement to impose on such a panel.

Facilitated by modern technology, decisions like the prediction of S are increasingly been made by groups of agents and this paper develops a normative formulation of the decision problem for such a group. To quote Parsons and Wooldridge (2002): "In the last few years, there has been increasing interest in the use of techniques from decision theory and game theory for analyzing and implementing agent systems." However it goes back a long way (see for example Radner, 1962).

Section 2 incorporates basic results of van Eeden and Zidek (1994) and sets up a decision theoretical framework for a single agent. In particular, a conjugate utility function for predicting S, is developed. That leads to an equivalent problem for the estimation of λ, the unknown scale parameter of the distribution of S. With a judicious approximation, the latter also has a conjugate utility function, providing the mathematical tractability needed to enable analytical progress. The Bayes estimator of λ is found and that in turn can be turned into a predictor of S. While recognizing that in practice the costs of incorrect prediction would in many situations be context dependent, we believe subsequent analysis will provide useful guidance to decision makers in situations where our utilities may be deemed unsuitable.

Section 3 considers the case of $G > 1$ agents and a number of general results are given. However solutions of that problem proves much more challenging than that for a single agent and hence our principle result on the so–called "group admissibility" of a predictor concerns the case of just $G = 2$.

Amongst other challenges, optimal solutions for the multi– unlike the single – agent problem can be randomized in situations where agent prior opinions are sufficiently divergent. However as the number of observations increases, their opinions will converge as reflected by their posterior distributions and eventually a non–randomized predictor will prove jointly acceptable as a consensual choice. Section 4 addresses the question of how large n must be to achieve that consensus. The pre-posterior probability of their attainment is found and shown to be 1 in certain situations.

Section 5 then explores the implications of adopting three paradigms for the multi – agent problem, comparing and contrasting the results obtained. Finally, Section 6 discusses various aspects of the problem addressed in this paper and returns to general issues presented in this introduction.

The Appendix contains the proofs of the lemmas as well as of one of the theorems. Further details of the proofs can be found in van Eeden and Zidek (2005).

2. A single agent

This section reviews results presented by van Eeden and Zidek (1994) for the case of a single Bayesian. A conjugate prior and utility function are assumed to gain mathematical tractability at the expense of completeness, to better enable us to address conceptual issues in later sections.

Prior to posterior distribution

Since the decision–maker observes $S_1, \ldots, S_n \overset{i.i.d}{\sim} \exp(\lambda)$, the sufficient statistic $T = \sum_{i=1}^{n} S_i$ contains all relevant sample information and has, conditionally on λ, a gamma distribution with density

$$f_T(t|\lambda) = \frac{t^{n-1}}{\lambda^n \Gamma(n)} \exp\left(-\frac{t}{\lambda}\right),\ t > 0.$$

Further assume the Agent has a (conjugate) inverted gamma prior density for λ given by

$$\pi(\lambda|\theta) = \frac{\beta^{\alpha-1}}{\lambda^{\alpha}\Gamma(\alpha-1)} \exp\left(-\frac{\beta}{\lambda}\right),\ \lambda > 0, \tag{2.1}$$

where $\theta = (\alpha, \beta)$ denotes the vector of hyperparameters. The mean of this prior is, for $\alpha > 2$, given by $E(\lambda|\theta) \equiv \mu(\lambda|\theta) = \beta/(\alpha - 2)$ and this decision maker's marginal density function for T is

$$f_T(t|\theta) = \frac{\Gamma(\alpha+n-1)}{\Gamma(n)\Gamma(\alpha-1)} \frac{(t/\beta)^{n-1}}{\beta(1+t/\beta)^{(\alpha+n-1)}},\ t > 0. \tag{2.2}$$

His posterior density function, conditional on the data, that is on the value of the sufficient statistics $T = t$, is given by

$$\pi(\lambda|t,\theta) = \frac{(\beta+t)^{\alpha+n-1}}{\Gamma(\alpha+n-1)\lambda^{\alpha+n}} \exp\left(-\frac{\beta+t}{\lambda}\right). \tag{2.3}$$

Finally, observe that this Agent's posterior mean, conditional on the data, $E[\lambda|t,\theta] = (\beta+t)/(\alpha+n-2)$ has the familiar form

$$E[\lambda|t,\theta] = \frac{\alpha-2}{\alpha+n-2}\mu(\lambda|\theta) + \frac{n}{\alpha+n-2}\hat{\lambda}_{MLE},$$

a weighted average of the prior mean and the maximum likelihood estimator of λ.

Selecting a utility function

This subsection uses the prediction problem to suggest reasonable choices for utility functions in the estimation problem. Prediction is often the ultimate goal of inference even when it is recast as estimation as in Tobias and Scotto (2005). In fact, Akaike (1981) citing his earlier work makes prediction the goal:

> "Akaike (1977) introduced a principle of statistical model building, the entropy maximization principle, which regards any statistical activity as an effort to maximize the expected entropy of the resulting estimate of the distribution of a future observation. The principle is characterized by the introduction of the entropy criterion and the predictive point of view."

The specific problem of predicting observable exponentials was considered in a Bayesian framework by Geisser (1985) although his goal was predictive intervals. Moreover he did not consider the associated estimation problem. Beginning with his paper, the prediction of such exponentials, their close cousins, Pareto random variables that commonly arise in applications in economics and finance, and others has generated a large research literature (Al–Hussaini and Ahmad, 2003).

Following Akaike we begin by seeking an estimator $\hat{\lambda}$ that yields a good probabilistic forecasting (or predictive) distribution $f_S(s|\hat{\lambda})$ for $S \sim \mathit{Exponential}(\lambda)$ that in turn would yield prediction intervals as well as a point predictor. Conditional on λ, an Akaike criterion functional is given by

$$\left.\begin{aligned} I &= \int f_S(s|\lambda)\log\left(\frac{f(s|\lambda)}{f(s|\hat{\lambda})}\right)ds \\ &= \int f_S(s|\lambda)\log f(s|\lambda)ds - \int f_S(s|\lambda)\log f(s|\hat{\lambda})ds = \frac{\lambda}{\hat{\lambda}} - \log\frac{\lambda}{\hat{\lambda}} - 1, \end{aligned}\right\} \tag{2.4}$$

the Kullback-Leibler measure of the discrepancy between the true distribution and the predictive distribution for S. Ideally $\hat{\lambda}$ should be chosen as $\hat{\lambda} = \lambda$ to minimize (2.4), the so-called *entropy loss* function. However it and its associated utility function $-I$, being unbounded, are unrealistic. Moreover, its mathematical intractability makes analysis of proposed estimators difficult.

Alternatively we could move from this celebrated approach to the more general one in decision analysis, that optimum decisions maximize the expected utility. Proceeding in this fashion we postulate a utility $U(S,\hat{S})$ for the prediction problem. Its expectation conditional on λ

$$E[U(\hat{S},S)|t,\lambda] \tag{2.5}$$

would provide a profile of its performance as a function of λ.

Following the approach of Lindley (1976) we might select the utility for computational convenience as

$$U(\hat{S},S) \equiv \frac{\gamma^{\frac{1}{2}}}{\sqrt{2\pi}}e\left[\frac{S}{\hat{S}}\exp\left(1-\frac{S}{\hat{S}}\right)\right]^{\gamma}. \tag{2.6}$$

Its utility profile function of λ can now be computed as

$$\left.\begin{aligned} E[U(\hat{S},S)|t,\lambda] &= \frac{1}{\sqrt{2\pi}}\gamma^{\frac{1}{2}}e\int f_S(s|\lambda)\left[\frac{s}{\hat{S}}\exp\left(1-\frac{s}{\hat{S}}\right)\right]^{\gamma}ds \\ &= \frac{1}{\sqrt{2\pi}}\gamma^{\frac{1}{2}}e^{1+\gamma}\left[\frac{1}{\hat{S}}\right]^{\gamma}\frac{1}{\lambda}\int s^{\gamma}\exp\left\{-s\left(\frac{1}{\lambda}+\frac{\gamma}{\hat{S}}\right)\right\}ds \\ &= \frac{1}{\sqrt{2\pi}}\Gamma(1+\gamma)e^{1+\gamma}\gamma^{-\gamma-\frac{1}{2}}\frac{\hat{S}}{\lambda}\left(1+\frac{\hat{S}}{\gamma\lambda}\right)^{-(1+\gamma)}. \end{aligned}\right\} \tag{2.7}$$

Formally substituting $\hat{\lambda}$ for $\hat{S}$ in (2.7) yields a utility function for the estimation of λ. In particular it is maximized by the choice $\hat{\lambda} = \lambda$ as would be required of any reasonable such utility.

Although that utility is not very tractable, we can find one that is by letting $\gamma \to \infty$.

Arriving at the approximation requires Stirling's approximation, which says that (when γ is large) $\Gamma(1+\gamma) \sim \sqrt{2\pi}\exp(-\gamma)\gamma^{\gamma+\frac{1}{2}}$. The result after the substitution

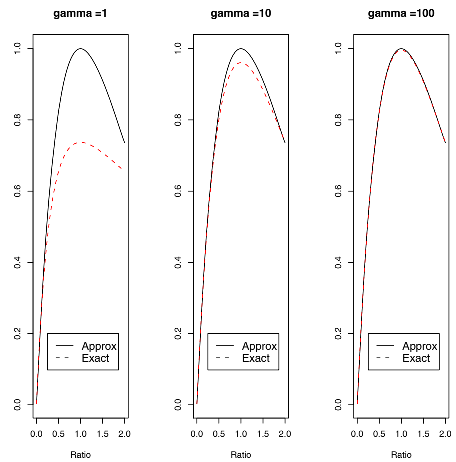

FIG 1. *Approximate and exact utility functions for estimating* λ *for varying* γ*, plotted against the* λ*-Ratio,* $\hat{\lambda}/\lambda$*.*

is a conjugate utility (Lindley, 1976)

$$u(\hat{\lambda}, \lambda) = \frac{\hat{\lambda}}{\lambda} \exp\left(1 - \frac{\hat{\lambda}}{\lambda}\right). \tag{2.8}$$

Figure 1 shows that the utility in (2.8) provides a good approximation to that in (2.7) even for fairly small values of γ.

However it has a disturbing feature seen on examining the utility for prediction in (2.6). As $\gamma \to \infty$ the latter increasingly concentrates on a shrinking neighborhood of $\hat{S} - S = 0$ while predictions outside that narrow region go unrewarded, not a seemingly realistic utility for prediction even in the pre–asymptotic case where γ is large. This finding calls into question the merit of that in (2.8) for estimating λ.

However, (2.8) derives from another prediction utility that is applicable in the situation described below. To see how it arises, we seek a solution $U_0(\hat{S}, S, \lambda)$ of

$$\int_0^\infty \frac{1}{\lambda} e^{-s/\lambda} U^*(\hat{S}, s, \lambda) ds = \frac{\hat{S}}{\lambda} e^{1-\hat{S}/\lambda}$$

and find

$$U_0(\hat{S}, S, \lambda) = \frac{e\hat{S}}{\lambda} I(\frac{\hat{S}}{S} < 1).$$

To verify this claim note that

$$\int_0^\infty \frac{1}{\lambda} e^{-s/\lambda} \frac{e\hat{S}}{\lambda} I(\frac{\hat{S}}{s} < 1) ds = \frac{\hat{S}}{\lambda} e \int_{\hat{S}}^\infty \frac{1}{\lambda} e^{-s/\lambda} ds = \frac{\hat{S}}{\lambda} e^{1-\hat{S}/\lambda}.$$

This utility (which is also scale invariant like its predecessor) covers the decision problem where the analyst is asked to provide an interval estimator $[\hat{S}, \infty)$ for S, $\hat{S}$ being as close as possible but below S. A less stringent family of utilities for the same situation would be

$$U_\delta(\hat{S}, S, \lambda) = \frac{e \min\{\hat{S}, S\}}{\lambda} I(\frac{\hat{S}}{S} < 1 + \delta)$$

for any $\delta > 0$. In any case, by formal substitution of $\hat{\lambda}$ for $\hat{S}$ in (2.8) we obtain a utility function for estimation that is the subject of further analysis in this paper, knowing that selecting a $\hat{\lambda}$ immediately yields a predictive lower bound for S as described above.

Estimating $\boldsymbol{\lambda}$

With the conjugate utility for λ derived in the last subsection, we can now find this Bayesian's predictor. However, before doing so, we present a lemma that will be useful in computing its conjugate utility and in the next section as well.

Lemma 2.1. *For the utility function (2.8), posterior (2.3) and an estimator of the form $\hat{\lambda}(t) = c_1 t + c_2$, $c_1 \geq 0$, $c_2 \geq 0$, the expected marginal utility is given by*

$$U(\hat{\lambda}, \theta) = \mathcal{E}u(\hat{\lambda}, \lambda) = \frac{e\beta^{\alpha-1}\{c_2(c_1+1)(\alpha-1) + c_1 n(c_2+\beta)\}}{(c_1+1)^{n+1}(c_2+\beta)^{\alpha}}. \tag{2.9}$$

The next theorem gives the Bayes rule for the conjugate utility (2.8).

Theorem 2.1. *For the utility function (2.8) and the prior (2.1), the Bayes estimator $\hat{\lambda}_B$ of λ is given by*

$$\hat{\lambda}_B(t) = \frac{t+\beta}{\alpha+n-1}. \tag{2.10}$$

The following corollary follows immediately from Lemma 2.1 with $c_1 = (\alpha+n-1)^{-1}$ and $c_2 = \beta(\alpha+n-1)^{-1}$.

Corollary 2.1. *The maximum value of the expected posterior utility is given by*

$$\frac{e\Gamma(\alpha+n)}{\Gamma(\alpha+n-1)} \, \frac{(\alpha+n-1)^{\alpha+n-1}}{(\alpha+n)^{\alpha+n}}. \tag{2.11}$$

Note that

$$\hat{\lambda}_B(t) = \frac{t+\beta}{\alpha+n-1} = \frac{\alpha+n}{\alpha+n-1}\,\mathcal{E}(\lambda \mid t),$$

where $\mathcal{E}(\lambda \mid t)$ is the Bayes estimator of λ for squared-error loss. Further, from (2.11) it is seen that the maximum expected posterior utility does not depend upon the data and depends on the prior only through α. This implies that the preposterior Bayes expected utility is also given by (2.11). In the next section each agent of a group of agents using the same (α, n) obtain, with their possibly different Bayes estimators, the same expected posterior as well as expected marginal utility.

3. The multiagent prediction problem

In this section the conjugate utility function (2.8) is used. The Bayes estimator is obtained and a G-complete-class result is presented.

Now consider a group consisting of G (Bayesian) agents and look at the problem of finding a G-complete class of estimators of λ using the expected marginal utility as the basis for comparisons between estimators. More specifically, the problem is to find a class $\mathcal{C}$ of estimators of λ such that, for each estimator $\hat{\lambda} \in \bar{\mathcal{C}}$, there exists a $\hat{\lambda}_1 \in \mathcal{C}$ with

$$\begin{cases} U(\hat{\lambda}_1, \theta) \geq U(\hat{\lambda}, \theta) & \text{for all } \theta \in \Theta \\ U(\hat{\lambda}_1, \theta) > U(\hat{\lambda}, \theta) & \text{for some } \theta \in \Theta. \end{cases}$$

We have succeeded in finding such a G-complete-class result for the special case where the group G consists of two agents and the estimators under consideration are of the form $\hat{\lambda}(t) = c_1 t + c_2$. However, for the proof of our G-complete-class result the following lemmas for the general case of an arbitrary number of agents are needed.

In Lemma 3.1 and Lemma 3.2, the expected utility $U(c_1 T + c_2, \theta)$ is studied as a function of c_1 and c_2 for $c_i \geq 0$, $i = 1, 2$.

Lemma 3.1. *Under the conditions of Lemma 2.1*

$$\frac{d}{dc_1} U(c_1 T + c_2, \theta) \left\{ \begin{array}{c} > \\ = \\ < \end{array} \right\} 0 \Longleftrightarrow c_2 \left\{ \begin{array}{c} < \\ = \\ > \end{array} \right\} \beta \frac{1 - c_1 n}{c_1(\alpha + n - 1) + \alpha - 2} .$$

Lemma 3.2. *Under the conditions of Lemma 2.1*

$$\frac{d}{dc_2} U(c_1 T + c_2, \theta) \left\{ \begin{array}{c} > \\ = \\ < \end{array} \right\} 0 \Longleftrightarrow c_2 \left\{ \begin{array}{c} < \\ = \\ > \end{array} \right\} \beta \frac{1 - (n-1) c_1}{c_1(\alpha + n - 1) + \alpha - 1} .$$

Now let (see Lemma 3.1 and Lemma 3.2)

$$h_1(c) = \frac{1 - cn}{c(\alpha + n - 1) + \alpha - 2} \qquad 0 \leq c \leq 1/n$$

$$h_2(c) = \frac{1 - c(n-1)}{c(\alpha + n - 1) + \alpha - 1} \qquad 0 \leq c \leq 1/(n-1)$$

and let $n > 1$ while $\alpha > 2$. Then h_1 and h_2 are each continuous and strictly decreasing in c with

$$\begin{cases} h_1(c) > 0 & \Longleftrightarrow \quad c < 1/n \\ h_2(c)) > 0 & \Longleftrightarrow \quad c < 1/(n-1) \\ h_1(0) = 1/(\alpha - 2) & \qquad h_1(1/n) = 0 \\ h_2(0) = 1/(\alpha - 1) & \qquad h_2(1/(n-1)) = 0. \end{cases}$$

Further, for $0 \leq c \leq 1/n$,

$$h_1(c) \left\{ \begin{array}{c} > \\ = \\ < \end{array} \right\} h_2(c) \Longleftrightarrow c \left\{ \begin{array}{c} < \\ = \\ > \end{array} \right\} \frac{1}{\alpha + n - 1} \tag{3.1}$$

and the pair (c_1, c_2) with

$$c_1 = \frac{1}{\alpha + n - 1} \text{ and } c_2 = \frac{\beta}{\alpha + n - 1} = \beta h_1 \left(\frac{1}{\alpha + n - 1} \right) = \beta h_2 \left(\frac{1}{\alpha + n - 1} \right)$$

gives the Bayes estimator $\hat{\lambda}_B$.

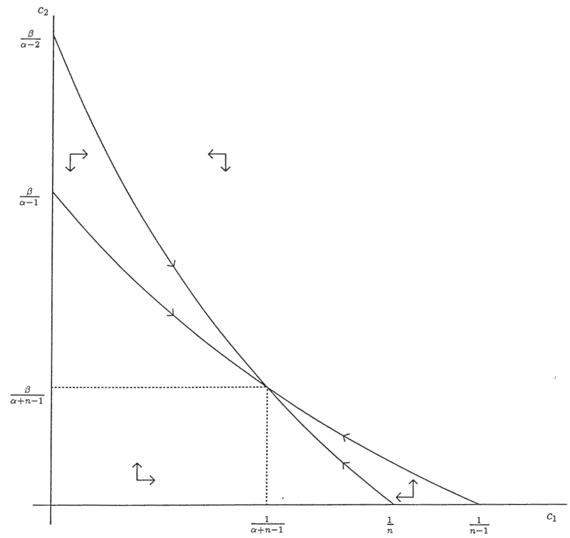

FIG 2. *Behavior of $U(c_1T + c_2, \theta)$ as a function of $c_1 > 0$ and $c_2 > 0$.*

Now let

$$
\left.\begin{aligned}
S_1(\beta) &= \{(c_1, c_2) \mid 0 \le c_1 \le 1/n, 0 \le c_2 \le \beta \min\{h_1(c_1), h_2(c_1\}\} \\
S_2(\beta) &= \{(c_1, c_2) \mid 0 \le c_1, c_2 > \beta \max\{0, h_1(c_1), h_2(c_1\}\} \\
S_3(\beta) &= \{(c_1, c_2) \mid 0 < c_1 < 1/(\alpha + n - 1), \beta h_2(c_1) < c_2 \le \beta h_1(c_1)\} \\
S_4(\beta) &= \{(c_1, c_2) \mid 1/(\alpha + n - 1) < c_1 \le 1/(n - 1), \beta h_1(c_1) < c_2 \le \beta h_2(c_1\}.
\end{aligned}\right\} \tag{3.2}
$$

Then it follows from Lemma 3.1, Lemma 3.2 and (3.1) that $U(c_1T + c_2, \theta)$ is, for each fixed β,

$$
\left.\begin{array}{ll}
(i) & \text{increasing in } c_1 \text{ and in } c_2 \text{ on } S_1(\beta) \\
(ii) & \text{decreasing in } c_1 \text{ and in } c_2 \text{ on } S_2(\beta) \\
(iii) & \text{increasing in } c_1 \text{ and decreasing in } c_2 \text{ on } S_3(\beta) \\
(iv) & \text{decreasing in } c_1 \text{ and increasing in } c_2 \text{ on } S_4(\beta).
\end{array}\right\} \tag{3.3}
$$

Finally, the next lemma gives the behaviour of $U(cT + c_2, \theta)$ as a function of c for $c_2 = \beta h_1(c)$ as well as for $c_2 = h_2(c)$.

Lemma 3.3. *For*

$$0 \le c \le \frac{1}{n}, \;\; c_2 = \beta h_1(c), \tag{3.4}$$

as well as for

$$0 \le c \le \frac{1}{n-1}, \;\; c_2 = \beta h_2(c), \tag{3.5}$$

$U(cT + c_2, \theta)$ is, for $c \ge 0$ and $c_2 \ge 0$, increasing in c for $c < 1/(\alpha + n - 1)$ and decreasing in c for $1/(\alpha + n - 1) < c$.

The above given properties of $U(c_1T + c_2, \theta)$ as a function of c_1 and c_2 are summarized in Figure 2, where the arrows indicate the direction in which $U(c_1T + c_2, \theta)$ increases.

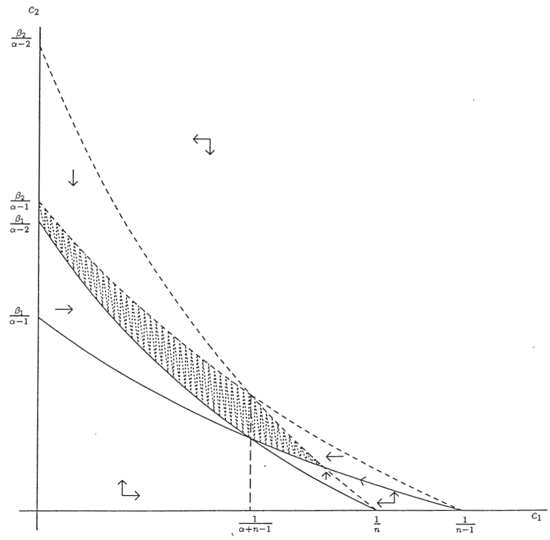

FIG 3. *Behavior of* $U(c_1T+c_2,\theta_j)$, $j=1,2$ *as a function of* c_1 *and* c_2 *and* β_j *when* $\beta_1(\alpha-2)^{-1} \leq \beta_2(\alpha-1)^{-1}$.

The following theorem gives our complete class result.

Theorem 3.1. *Consider a group consisting of two Bayesians, each using the utility function (2.8) with the posterior (2.3) with the same* α $(\alpha > 2)$ *but with different* β*'s,* β_1 *and* β_2 *with* $\beta_1 < \beta_2$*. Let* $n > 1$*. Then the class*

$$\mathcal{C} = \{\hat{\lambda}(T) = c_1T + c_2 \mid (c_1, c_2) \in S^*\}$$

where S^* *is the closure of the set* $\{(c_1, c_2) \ |(c_1, c_2) \in S_1(\beta) \cap S_2(\beta)\}$*, is a G-complete class of estimators of* λ *within the class of linear estimators.*

Proof. First note that S^* is not empty and contains points (c_1, c_2) with $c_1 < 1/(\alpha+n-1)$, points (c_1, c_2) with $c_1 > 1/(\alpha+n-1)$, as well as all points (c_1, c_2) with $c_1 = 1/(\alpha+n-1)$, $\beta_1/(\alpha+n-1) \leq c_2 \leq \beta_2/(\alpha+n-1)$. This can be seen as follows. First note that, by the definitions of $S_1(\beta)$ and $S_2(\beta)$

$$S^* = \{(c_1, c_2) \mid 0 \leq c_1 \leq 1/n, \beta_1 M(c_1) \leq c_2 \leq \beta_2 m(c_1)\}$$

where $m(c) = \min\{h_1(c), h_2(c)\}$ and $M(c) = \max\{h_1(c), h_2(c)\}$.

Further (see (3.1))

$$\beta_2 m\left(\frac{1}{\alpha+n-1}\right) - \beta_1 M\left(\frac{1}{\alpha+n-1}\right) = (\beta_2-\beta_1)h_1\left(\frac{1}{\alpha+n-1}\right) > 0.$$

That S^* is not empty then follows from the fact that $\beta_2 m(c) - \beta_1 M(c)$ is continuous in c on $0 \leq c \leq 1/(n-1)$.

Figures 3 and 4 summarize the properties of $U(c_1T+c_2,\theta_j)$, $j=1,2$ as functions of c_1 and c_2, as well as the relationships between the sets $S_i(\beta_j)$, $i=1,\ldots,4$, $j=1,2$.

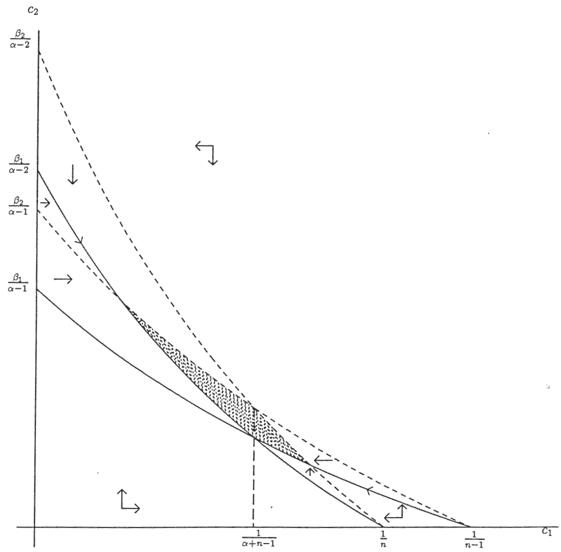

FIG 4. *Behavior of* $U(c_1T+c_2,\theta_j)$, $j=1,2$ *as a function of* c_1 *and* c_2 *and* β_j *when* $\beta_1(\alpha-2)^{-1} > \beta_2(\alpha-1)^{-1}$.

The two cases considered are

$$\left.\begin{array}{ll} (i)\ \dfrac{\beta_1}{\alpha-2} \le \dfrac{\beta_2}{\alpha-1} & \text{in Figure 3} \\[2ex] (ii)\ \dfrac{\beta_1}{\alpha-2} > \dfrac{\beta_2}{\alpha-1} & \text{in Figure 4.} \end{array}\right\} \tag{3.6}$$

To study the shape of S^*, let $H(c)=\beta_2 m(c)-\beta_1 M(c)$. First, consider the case where $0\le c\le 1/(\alpha+n-1)$. Then

$$\left.\begin{aligned} H(c) &= \beta_2 h_2(c)-\beta_1 h_1(c) \\ &= \beta_2\frac{1-c(n-1)}{c(\alpha+n-1+\alpha-1)}-\beta_1\frac{1-cn}{c(\alpha+n-1)+\alpha-2}. \end{aligned}\right\} \tag{3.7}$$

When $\beta_2/(\alpha-1)\ge\beta_1/(\alpha-2)$, it follows from (3.7) that $H(c)>0$ for $0\le c\le 1/(\alpha+n-1)$ because $1-c(n-1)\ge 1-cn>0$ and $\alpha-2<\alpha-1$. When $\beta_2/(\alpha-1)<\beta_1/(\alpha-2)$, $H(0)<0, H(1/(\alpha+n-1))>0$ and $H(c)=0$ has exactly one root, c_o, say, in the interval $[0,1/(\alpha+n-1)]$ because $H(c)\ge 0$ if and only if

$$(\alpha+n-1)(n\beta_1-(n-1)\beta_2)c^2+$$

$$[(\alpha+n-1)(\beta_2-\beta_1)+\beta_1 n(\alpha-1)-\beta_2(n-1)(\alpha-2)]c+\beta_2(\alpha-2)-\beta_1(\alpha-1)\ge 0.$$

Moreover, $H(c)<0$ for $0\le c\le c_o$ and $H(c)>0$ for $c_o<c\le 1/(\alpha+n-1)$.

Now consider the case where $1/(\alpha+n-1)\le c\le 1/n$. Then

$$\begin{aligned} H(c) &= \beta_2 h_1(c)-\beta_1 h_2(c) \\ &= \beta_2\frac{1-cn}{c(\alpha+n-1)+\alpha-1}-\beta_1\frac{1-c(n-1)}{c(\alpha+n-1)+\alpha-2} \end{aligned}$$

with $H(1/(\alpha+n-1)) > 0$ and $H(1/n) < 0$. That $H(c) = 0$ has exactly one root, c_o^* say, in the interval $(1/(\alpha+n-1), 1/n)$ follows from the fact that $H(c) \geq 0$ if and only if

$$(\alpha+n-1)(\beta_1(n-1)-\beta_2 n)c^2 +$$

$$[(\beta_2-\beta_1)(\alpha+n-1)-(\beta_2 n(\alpha-1)-\beta_1(n-1)(\alpha-2)]c+\beta_2(\alpha-1)-\beta_1(\alpha-2) \geq 0.$$

Further, of course, $H(c) > 0$ for $1/(\alpha+n-1) < c < c_o^*$ and $H(c) < 0$ for $c_o^* < c \leq 1/n$.

It now needs to be shown that, for every (c_1, c_2) not in S^*, there exists $(c_1', c_2') \in S^*$ such that

$$\left.\begin{aligned} &U(c_1'T + c_2', \theta_j) \geq U(c_1 T + c_2, \theta_j), && j = 1,2 \\ &U(c_1'T + c_2', \theta_j) > U(c_1 T + c_2, \theta_j) && \text{for some } j \in \{1,2\}. \end{aligned}\right\} \tag{3.8}$$

Such (c_1', c_2') can be obtained as follows (see also Figures 3.1–3.3). Start, e.g., with $(c_1, c_2) \in S_1(\beta_1)$. Then because $S_1(\beta_1) \subset S_1(\beta_2)$, $(c_1, c_2) \in S_1(\beta_2)$. Thus, one can, keeping c_1 fixed, increase each of the expected utilities by increasing c_2 until (c_1, c_2) satisfies

$$c_2 = \beta_1 m(c_1) = \beta_1 \min\{h_1(c_1), h_2(c_1)\}.$$

Then

(i) if $c_1 \leq 1/(\alpha+n-1)$, one can increase c_1 while keeping c_2 fixed. Each of the expected utilities then increases until (c_1, c_2) satisfies $c_2 = \beta_1 h_2(c_1)$. One then has reached S^* or, if not (as might be the case when $\beta_1/(\alpha-2) > \beta_2/(\alpha-1)$ one can "slide down" the curve $c_2 = \beta_1 h_1(c_1)$ and thus increase each of the expected utilities, until S^* is reached;

(ii) if $c_1 > 1/(\alpha+n-1)$, one can further increase c_2 until (c_1, c_2) satisfies $c_2 = \beta_1 h_2(c_1)$. Then one either has reached S^*, or one can "slide up" the curve $c_2 = \beta_1 h_2(c_1)$, increasing each of the expected utilities, until S^* is reached.

Similar reasoning works for the other cases. □

Remarks.

(i) We do not know whether C contains a proper subset which is G-comp1ete within the class of linear estimators.

(ii) In the above only nonrandomized estimators were considered. We do not have a similar result for the class of all estimators.

4. Consensual choice

In this section we consider the case where the group consists of two Bayesians, B_i, $i = 1, 2$, with the same conjugate utility function, while their priors have the same $\alpha > 2$ but different β's. They have the same data, t, available. Then (see (2.10)) B_i's preferred decision is $\hat{\lambda}_i = (t + \beta_i)/(\alpha+n-1)$, $i = 1, 2$ and the question we are looking at in this section is what decision $\hat{\lambda}$ these two Bayesians could agree upon as a compromise between their $\hat{\lambda}_i$.

To find an answer to this question we study (see (A.2)) the joint behavior of the expected posterior utilities $U_i(x,\theta_i \mid t)$, $i=1,2$ as a function of x. To simplify the notation we put, for $i=1,2$, $\delta_i=\beta_i+t$ and $U_i(x)=U_i(x,\theta_i \mid t)$. Then

$$U_i(x)=K_i\frac{x}{(\delta_i+x)^{\alpha+n}}, \qquad i=1,2,$$

where the K_i, $i=1,2$, are positive constants independent of x. Further we suppose, without loss of generality, that $\hat{\lambda}_1<\hat{\lambda}_2$.

The theorems below give the needed properties of $U_2(x)$ as a function of $U_1(x)$ for $x>0$.

Theorem 4.1. *For $x>0$*

$$\begin{aligned}\frac{dU_2}{dU_1}(x)\propto\frac{\hat{\lambda}_2-x}{\hat{\lambda}_1-x}\left(\frac{\delta_1+x}{\delta_2+x}\right)^{\alpha+n+1} \quad & > \quad 0 \text{ when } x<\hat{\lambda}_1\\ & = \quad \infty \text{ when } x=\hat{\lambda}_1\\ & < \quad 0 \text{ when } \hat{\lambda}_1<x<\hat{\lambda}_2\\ & = \quad 0 \text{ when } x=\hat{\lambda}_2\\ & > \quad 0 \text{ when } x>\hat{\lambda}_2.\end{aligned}$$

This theorem follows directly from Lemma A.1.

The following theorem, which follows directly from Lemma A.2, gives the convexity-concavity properties of U_2 as a function of U_1 for $x<\hat{\lambda}_1$ as well as for $x>\hat{\lambda}_2$.

Theorem 4.2. *For $x<\hat{\lambda}_1$ $U_2(x)$ is a convex function of $U_1(x)$, while for $x>\hat{\lambda}_2$ $U_2(x)$ is a concave function of $U_1(x)$.*

The next theorem gives the convexity-concavity properties of $U_2(x)$ as a function of $U_1(x)$ for $\hat{\lambda}_1<x<\hat{\lambda}_2$.

Let A, B and C be given by (A.7), let (see Lemma A.5) $s(\alpha,n)=8(\alpha+n)/(\alpha+n-1)$ and let $r(\alpha,n)$ be the unique solution >1 to $r^2+(2-s(\alpha,n))r+1=0$. Then

$$r(\alpha,n)=3+\frac{4}{\alpha+n-1}+\sqrt{\left(3+\frac{4}{\alpha+n-1}\right)^2-1} \tag{4.1}$$

and the following theorem follows from Lemma A.7.

Theorem 4.3. *On $(\hat{\lambda}_1,\hat{\lambda}_2)$*

1) *when $\hat{\lambda}_2/\hat{\lambda}_1\le r(\alpha,n)$, $U_2(x)$ is a concave function of $U_1(x)$;*
2) *when $\hat{\lambda}_2/\hat{\lambda}_1>r(\alpha,n)$, $U_2(x)$ is a concave, convex, concave function of $U_1(x)$ on, respectively, $(\hat{\lambda}_1,x_1],(x_1,x_2),[x_2,\hat{\lambda}_2)$, where $x_1<x_2$ are the roots to $Ax^2+Bx+C=0$.*

Remark. By Theorem 4.3 above we have concavity on $(\hat{\lambda}_1,\hat{\lambda}_2)$ if and only if

$$\frac{\hat{\lambda}_2}{\hat{\lambda}_1}\le r(\alpha,n)$$

or (see Lemma A.5) if and only if

$$\frac{\left(\hat{\lambda}_1+\hat{\lambda}_2\right)^2}{\hat{\lambda}_1\hat{\lambda}_2}\le s(\alpha,n).$$

This result is not in agreement with Theorem 4.3 of van Eeden and Zidek (1994) who have concavity on $(\hat{\lambda}_1, \hat{\lambda}_2)$ if and only if

$$\frac{\left(\hat{\lambda}_1 + \hat{\lambda}_2\right)^2}{\hat{\lambda}_1 \hat{\lambda}_2} \leq 4C_0^2$$

where

$$4C_0^2 = \frac{(\alpha+n)^2 - (\alpha+n) + 2}{\alpha(\alpha+n)^2}\alpha(\alpha+n)^2 \neq s(\alpha, n).$$

The 1994 result is incorrect.

From the theorems 4.2 and 4.3 it follows that, when $\hat{\lambda}_2/\hat{\lambda}_1 \leq r_1(\alpha, n)$,

$$\mathcal{C} = \{\hat{\lambda} \mid \hat{\lambda}_1 \leq \hat{\lambda} \leq \hat{\lambda}_2\}$$

is a complete class of decision rules within the class of all rules. But when $\hat{\lambda}_2/\hat{\lambda}_1 > r_1(\alpha, n)$, some of the rules in $\mathcal{C}$ can be improved upon by randomized rules. So in the latter case, optimality would force the two Bayesians into the practically objectionable position of having to resort to randomized rules to arrive at a consensual choice.

We now turn to the following question (which could be asked before the data are collected): "Conditional on λ, what is the probability that for two Bayesians using the same data, $\hat{\lambda}_2/\hat{\lambda}_1 \leq r_1(\alpha, n)$?". In other words: "Conditional on λ, what is the probability that the optimal consensual choice of two Bayesians can be reached with a nonrandomized rule?". This probability is given by

$$P_\lambda\left(\frac{T}{\lambda} \geq \frac{\beta_2 - r(\alpha,n)\beta_1}{\lambda(r(\alpha,n)-1)}\right) = P_\lambda\left(\frac{T}{\lambda} \geq \frac{\beta_2 - \beta_1}{\lambda(r(\alpha,n)-1)} - \frac{\beta_1}{\lambda}\right), \tag{4.2}$$

where, by the assumption made above, $\hat{\lambda}_1 < \hat{\lambda}_2$, $\beta_1 < \beta_2$.

Clearly, if the priors of the two Bayesians are not too far apart in the sense that

$$\beta_2 \leq \beta_1 r(\alpha, n), \tag{4.3}$$

they are sure to be able to reach consensus.

More properties of the probability (4.2) are given in the theorems 4.4 and 4.5 for which the following result is needed.

Lemma 4.1. *For $n = 1, 2, \ldots$,*

$$\left.\begin{aligned} 3 + \sqrt{8} &< r(\alpha, n+1) < r(\alpha, n) \\ &< r(\alpha, 0) = 3 + \frac{4}{\alpha-1} + \sqrt{\left(3 + \frac{4}{\alpha-1}\right)^2 - 1}. \end{aligned}\right\}$$

The proof of this lemma follows immediately from (4.1).

Theorem 4.4. *If, for some $N_o \geq 0$,*

$$r(\alpha, N_o + 1) < \frac{\beta_2}{\beta_1} \leq r(\alpha, N_o), \tag{4.4}$$

then the probability of consensus equals 1 *for every* $n \geq N_o$*. In particular, if*

$$\beta_2/\beta_1 \leq 3 + \sqrt{8},$$

the probability of consensus equals 1 *for all* $n \geq 0$*.*

Finally, if

$$\beta_2/\beta_1 > 3 + \frac{4}{\alpha - 1} + \sqrt{\left(3 + \frac{4}{\alpha - 1}\right)^2 - 1}$$

the probability of consensus is less than 1 *for all* $n \geq 0$*.*

Proof. The results follow immediately from (4.2) and (4.4). □

Theorem 4.5. *If the prior* β*'s do not satisfy (4.3), then (4.2) is less than* 1 *and the following hold:*

i) for fixed α*,* n *and* λ*, (4.2) increases as* $\beta_2 - r(\alpha, n)\beta_1$ *decreases, i.e. as the priors get closer together, the probability (4.2) increases;*
ii) for fixed α*,* n*,* β_1 *and* β_2*, the probability (4.2) increases as* λ *increases;*
iii) for fixed α*,* λ*,* β_1 *and* β_2*, the probability (4.2) converges to* 1 *as* $n \to \infty$*.*

Proof. The first result follows from (4.2). To see the second result, note that the distribution of T/λ does not depend on λ. For the third result, let

$$A_n = \frac{r(\alpha, n) - 1}{\beta_2 - \beta_1}.$$

Then it follows from (4.1) that $\sqrt{n}A_n \to \infty$. The result then follows from the fact that (4.2) can be written as

$$P_\lambda\left(\frac{T - n\lambda}{\lambda\sqrt{n}} \geq \frac{1}{\lambda A_n \sqrt{n}} - \frac{\beta_1}{\lambda\sqrt{n}} - \sqrt{n}\right)$$

and the asymptotic normality of $(T - n\lambda)/\lambda\sqrt{n}$. □

5. Multiagent estimation

The previous two sections explored the Pareto boundary from which the agents would select their joint estimator of λ with the knowledge, based on our analysis of Section 2, that the result would also solve a prediction problem. However, we have side-stepped the question of how the joint estimator might be chosen. That depends on the normative decision paradigm. Here are some possibilities:

1. The Organization is a third intelligent agent, *i.e.* a "supra Bayesian" (see, for example, Genest and Zidek (1986)), capable of combining the data and prior opinions (as data!) with his, her or its own prior and thereafter developing a conventional Bayes estimator.
2. The Organization can "pool" the separate posterior distributions to create a single posterior, again for use in a conventional analysis as above.
3. The problem can be treated as a multiagent decision problem where the agents would have an individually preferred (Bayesian) estimation strategy but would be forced to seek a compromise in a group decision problem.

However, selection of the best paradigm for a particular application would depend on the context. This section reviews the implications of choosing each of these paradigms.

First consider Paradigm #1. Suppose the G agents share with the supra - Bayesian, the conjugate prior (2.1) only with varying hyperparameters, so that Agent i has hyperparameters, $(\alpha_i, \beta_i),\ i = 1, \ldots, G$ while the supra-Bayesian has (α_0, β_0).

One common interpretation of conjugate priors leads us to an estimator for the supra-Bayesian. Thus, the supra-Bayesian (Agent 0) might well assume that the hyperparameters actually represent prior knowledge gained from the equivalent of repeated observations of the exponential random variable itself. Consequently the $\{\alpha_i\}$ represent the number of prior observations Agent i has made (that is, the amount of prior information i has) while the $\{\beta_i\}$ represent the values of their respective sufficient statistics, their prior counterparts of T, in other words.

Assuming independence of the agents' prior data leads to a likelihood for λ based on the prior data that can readily be combined with that based on the data (T). Then the results of Section 2 apply directly to yield the following estimator for the supra-Bayesian (Agent 0):

$$\hat{\lambda}_{Supra-Bayesian} = \frac{t + \beta_.}{\alpha_. + n}, \tag{5.1}$$

the "." subscript standing for summation over that subscript $i = 0, 1, \ldots, G$.

Although Paradigm #1 and the approach taken above lead directly to an estimator $\hat{\lambda}$, they have some objectionable features discussed in Section 6. In any case, the second paradigm enjoys appeal. Here instead of trying to "accumulate" the prior knowledge in the various priors, a single prior is adopted to "represent" or "typify" them. In particular, Genest and Zidek (1986) along with references therein, suggest the use of the geometric average of the priors to do so:

$$\begin{aligned} \pi_{Multagent}(\lambda) &\equiv \Pi_{i=1}^{G} \pi^{\omega_i}(\lambda|\alpha_i, \beta_i), \\ &\propto \left(\frac{1}{\lambda}\right)^{\sum_{i=1}^{G} \omega_i \alpha_i} \exp\left(-\frac{\sum_{i=1}^{G} \omega_i \beta_i}{\lambda}\right),\ \lambda > 0 \end{aligned}$$

where the weights $\{\omega_i\}, \omega_i \geq 0, \sum \omega_i = 1$, reflect the importance to be attached to each agent. Thus a conjugate prior is obtained. In the simplest case $\omega_i \equiv G^{-1}$ and then

$$\pi_{Multagent}(\lambda) \propto \left(\frac{1}{\lambda}\right)^{\bar{\alpha}} \exp\left(-\frac{\bar{\beta}}{\lambda}\right),\ \lambda > 0,$$

where $\bar{\alpha} \equiv G^{-1}\alpha_.$ and $\bar{\beta} \equiv G^{-1}\beta_.$.

Note that generally the weighted geometric average of the prior densities does not integrate to 1. However that is a non-issue. Afterall, both the utility function and likelihood functions are only defined up to a positive multiplicative scaling factor. Moreover, the prior gives the same Bayes rule no matter how it is scaled. In fact, in this case Section 2 again leads directly to a predictor

$$\hat{S} \equiv \hat{\lambda}_{Multiagent} = \frac{t + \bar{\beta}}{\bar{\alpha} + n}.$$

Paradigm #2 also has shortcomings in some situations and these are discussed in Section 6. In fact, neither #1 nor #2 will be suitable in situations where the groups of agents are required to act in their individual selfinterest and yet choose a compromise that recognizes their individual positions. Paradigm #3 is most appropriate in that case.

The linearity in t of (2.10) suggests restricting the search for a compromise to the class of linear estimators $\hat{S} \equiv \hat{\lambda} = c_1 t + c_2, t > 0$. Each agent's expected gain in utility for members of that class appears in (2.9). Selecting a compromise entails finding a solution concept on which the choice could be made. We adopt the one advocated in Weerahandi and Zidek (1983) that is based on maximizing the celebrated Nash-Kalai product of their utilities, that is their geometric average:

$$\begin{aligned} U_{Multiagent}(\hat{\lambda}) &\equiv \Pi_{i=1}^{G} U^{\omega_i}(\hat{\lambda}, \theta_i) \\ &\propto \frac{1}{(c_1+1)^{n+1}} \Pi_{i=1}^{G} \left(\frac{c_2(c_1+1)(\alpha_i - 1) + nc_1(c_2+\beta_i)}{(c_2+\beta_i)^{\alpha_i}} \right)^{\omega_i} \end{aligned}$$

where again the $\{\omega_i\}, \omega_i > 0, \sum_{i=1}^{G} \omega_i = 1$, represent the weights to be attached to each agent when seeking the compromise.

In general, the compromise predictor cannot be found in an explicit form even in the simplest case where $\omega_i \equiv G^{-1}$. Instead numerical methods would need to be used in specific cases. Moreover, as the results of Section 4 show, the Nash-Kalai solution may not be optimum in the class that includes randomized rules, even in the case $G = 2$ unless the conditions for consensus in that section are met. The results of Haines (2003) may be of value here.

6. Discussion

Observe that the logarithm of the conjugate utility (2.8) is

$$-\left[+\frac{\hat{\lambda}}{\lambda} - \log\left[\frac{\hat{\lambda}}{\lambda}\right] - 1 \right]. \tag{6.1}$$

Multiplying the result by -1 to convert it from a log utility to a loss function leads, curiously, to the entropy loss with the roles of λ and $\hat{\lambda}$ interchanged.

Generally parameters for sampling distributions are abstract quantities without physical meaning. That makes appealing the development of utility functions for estimating them from consideration of a dual prediction problem. The value of that approach is further enhanced by knowledge that usually prediction is the inferential goal even when cast in terms of parameter estimation. In fact, Geisser (1993) asserts that predictive inference is the primary purpose of statistical endeavor.

Surprisingly the duality between prediction and estimation within the decision theoretical framework has not received much attention perhaps because as Geisser argues, emphasis in statistics moved away from its original purpose of prediction to characterizing the "true state of nature". The results in Section 2 provide a curious example where a utility function for estimation actually corresponds to two very different prediction problems, only one of which seems like a realistic possibility.

This paper assumes the agents share their data. However, sharing may not be feasible in some situation so that Agent i has only T_i, the sufficient statistic from n_i observations on which to base an estimator of λ. Some of our results extend to this case in a straightforward manner. However, generally it proves much more challenging, corresponding to the case where not only the β's but also the α's vary in the prior distributions of the agents in (2.1). Here little analytic progress can be made.

In Section 5 three paradigms were invoked to find a predictor, #1 and #2 leading to an explicit result, while #3 leads to a criterion function that would need to be maximized numerically. Which of these is most appropriate will depend largely on the context. The first, #1, requires a supra-Bayesian (Agent 0) to supervise the other agents. The gain in utility function is Agent 0's. Even if having such an agent is feasible the derivation of Agent 0's predictor in Section 5 is too simplistic, supposing as it does the independence of the agents' prior data. In fact, their prior opinions will be shaped to a considerable extent by common knowledge. In the extreme case $\beta_i \equiv \beta$ and $\alpha_i \equiv \alpha$ when the agents have identical prior information. In general the supra-Bayesian would need to construct a likelihood function that reflects the correlation among these parameters. The result will be far less accumulated information than that reflected in the very optimistic (5.1). In other words, implementation of Paradigm #1 will require some sophisticated modeling by the supra-Bayesian. That agent's predictor will be much harder to find than our analysis suggests.

If the agents' opinions can be combined say by the organization they serve and a supra-Bayesian approach is not feasible, then Paradigm #2 obtains. The result is formally similar to that obtained above for #1. However, instead of trying to accumulate prior information as #1 does, it merely tries to deal with the competing posteriors by finding one that represents them. This shows that the two approaches differ in a very fundamental respect.

The last paradigm (#3) is the one to be used by autonomous agents required to find a compromise predictor. This one leads to difficult computational issues. Indeed, it is difficult to determine in general when grounds for consensus exist (that is when randomized predictors are unnecessary.) However, Section 5 does provide an explicit criterion for finding an optimum Nash-Kalai estimator.

Another solution criterion, a variation of a supra-Bayesian approach is also feasible. Suppose one Bayesian, i, is to be selected at random from among the G agents with probability ρ_i. The value of a predictor or estimator $(\hat{\lambda})$ will then be assessed using that agent's expected gain in utility function. However, the predictor must be selected in advance, without knowing which agent will be selected. Then to maximize the expected gain, the predictor should be chosen to maximize

$$\begin{aligned} U_{Supra} &\equiv \sum_{i=1}^{G} \rho_i U(\hat{\lambda}, \theta_i) \\ &\propto \frac{1}{(c_1+1)^{n+1}} \sum_{i=1}^{G} \rho_i \left(\frac{c_2(c_1+1)(\alpha_i - 1) + nc_1(c_2+\beta_i)}{(c_2+\beta_i)^{\alpha_i}} \right). \end{aligned}$$

This and other solution criteria remain to be explored in future work.

Appendix A: Appendix

Proof of Lemma 2.1. First note that conditionally on λ,

$$\left.\begin{aligned} &\mathcal{E}[u(c_1T + c_2, \lambda, \theta) \mid \lambda] \\ &= \frac{e}{\lambda} \int_0^\infty (c_1 t + c_2)\, e^{-(c_1 t + c_2)/\lambda}\, \frac{t^{n-1}}{\lambda^n \Gamma(n)}\, e^{-t/\lambda} dt \\ &= \frac{e^{(1-(c_2/\lambda))}}{\lambda(c_1+1)^{n+1}} \{c_1 n \lambda + c_2(c_1+1)\}. \end{aligned}\right\} \tag{A.1}$$

From (A.1) one obtains

$$\begin{aligned} U(c_1T+c_2,\theta) &= \mathcal{E}u(c_1T+c_2,\lambda,\theta) \\ &= \frac{e}{(c_1+1)^{n+1}}\frac{\beta^{\alpha-1}}{\Gamma(\alpha-1)}\left\{c_1 n\frac{\Gamma(\alpha-1)}{(c_2+\beta)^{\alpha-1}}+c_2(c_1+1)\frac{\Gamma(\alpha)}{(c_2+\beta)^{\alpha}}\right\}, \end{aligned}$$

from which the result follows immediately. □

Proof of Theorem 2.1. By (2.1) and the fact that for given $\lambda>0$, T has density

$$\frac{t^{n-1}e^{-t/\lambda}}{\lambda^n\Gamma(n)}I(t>0),$$

the joint density of T and λ is given by

$$\frac{t^{n-1}e^{-t/\lambda}}{\lambda^n\Gamma(n)}\frac{\beta^{\alpha-1}e^{-\beta/\lambda}}{\lambda^\alpha\Gamma(\alpha-1)}I(t>0,\lambda>0).$$

From (2.2) it then follows that the posterior density of λ is, for $t>0$, given by

$$\pi(\lambda\mid t)=\frac{(t+\beta)^{\alpha+n-1}}{\Gamma(\alpha+n-1)\lambda^{\alpha+n}}e^{-(t+\beta)/\lambda}I(\lambda>0).$$

For an estimator $\hat\lambda=\hat\lambda(T)$, the expected posterior utility becomes

$$\text{(A.2)}\quad \left.\begin{aligned} U(\hat\lambda,\theta\mid t) &= \mathcal{E}\left\{u(\hat\lambda,\lambda,\theta)\mid t\right\}=\mathcal{E}\left\{\frac{\hat\lambda}{\lambda}\exp\left\{1-\frac{\hat\lambda}{\lambda}\right\}\mid t\right\} \\ &= e\,\hat\lambda\,\frac{(t+\beta)^{\alpha+n-1}}{\Gamma(\alpha+n-1)}\,\frac{\Gamma(\alpha+n)}{(t+\beta+\hat\lambda)^{\alpha+n}}\,. \end{aligned}\right\}$$

The Bayes estimator $\hat\lambda_B$ maximizes $U(\hat\lambda,\theta\mid t)$ and it is easily seen that this maximum is attained for the $\hat\lambda$ satisfying $dU(\hat\lambda,\theta\mid t)/d\hat\lambda=0$ where

$$\begin{aligned} \frac{d}{d\hat\lambda}\log U(\hat\lambda,\theta\mid t) &= \frac{1}{\hat\lambda}-\frac{(\alpha+n)}{t+\beta+\hat\lambda}=\frac{t+\beta+\hat\lambda-\hat\lambda(\alpha+n)}{\hat\lambda(t+\beta+\hat\lambda)} \\ &= (\alpha+n-1)\frac{(t+\beta)/(\alpha+n-1)-\hat\lambda}{\hat\lambda(t+\beta+\hat\lambda)}\,. \end{aligned}$$

This proves (2.10). □

Proof of Lemma 3.1.

$$\frac{d}{dc_1}U(c_1T+c_2,\theta)\left\{\begin{matrix}>\\=\\<\end{matrix}\right\}0 \iff \frac{d}{dc_1}\frac{c_1n(c_2+\beta)+c_2(c_1+1)(\alpha-1)}{(c_1+1)^{n+1}}\left\{\begin{matrix}>\\=\\<\end{matrix}\right\}0$$

$$\iff c_2\quad\left\{\begin{matrix}<\\=\\>\end{matrix}\right\}\quad\beta\frac{1-c_1n}{c_1(\alpha+n-1)+\alpha-2}\,.$$

□

Proof of Lemma 3.2.

$$\frac{d}{dc_2}U(c_1T+c_2,\theta)\left\{\begin{array}{c}>\\=\\<\end{array}\right\}0\,\frac{d}{dc_2}\frac{c_1n(c_2+\beta)+c_2(c_1+1)(\alpha-1)}{(c_2+\beta)^\alpha}\left\{\begin{array}{c}>\\=\\<\end{array}\right\}0$$

$$\iff c_2\quad\left\{\begin{array}{c}<\\=\\>\end{array}\right\}\quad\beta\frac{1-(n-1)c_1}{c_1(\alpha+n-1)+\alpha-1}\,.$$

□

Proof of Lemma 3.3. To see this result for (3.4) note that

$$\left.\begin{array}{rcl}\frac{d}{dc}U(cT+\beta h_1(c),\theta) & = & \frac{d}{dc}U(cT+c_2,\theta))|_{c_2=\beta h_1(c)}+\\ & & \frac{d}{dc_2}U(cT+c_2,\theta))|_{c_2=\beta h_1(c)}\frac{d}{dc}\beta h_1(c).\end{array}\right\}\tag{A.3}$$

The first term on the right hand side of (A.3) is zero by Lemma 3.1. Furthermore,

$$\frac{d}{dc}h_1(c)<0 \text{ for } 0\le c\le\frac{1}{n},$$

so it is sufficient to show that

$$\frac{d}{dc_2}U(cT+c_2,\theta)|_{c_2=\beta h_1(c)}\left\{\begin{array}{c}<\\=\\>\end{array}\right\}0\iff c\left\{\begin{array}{c}<\\=\\>\end{array}\right\}\frac{1}{\alpha+n-1}.$$

But by Lemma 3.2

$$\frac{d}{dc_2}U(cT+c_2,\theta)\left\{\begin{array}{c}>\\=\\<\end{array}\right\}0\iff c_2\left\{\begin{array}{c}<\\=\\>\end{array}\right\}\beta h_2(c).$$

The result then follows from (3.1).

For a proof of the result when (3.5) holds, note that

$$\left.\begin{array}{rcl}\frac{d}{dc}U(cT+\beta h_2(c),\theta) & = & \frac{d}{dc}U(cT+c_2,\theta))|_{c_2=\beta h_2(c)}+\\ & & \frac{d}{dc_2}U(cT+c_2,\theta))|_{c_2=\beta h_2(c)}\frac{d}{dc}\beta h_2(c)\end{array}\right\}\tag{A.4}$$

with, by Lemma 3.2,

$$\frac{d}{dc_2}U(cT+c_2,\theta)|_{c_2=\beta h_2(c)}=0.$$

So, it is sufficient to show that

$$\frac{d}{dc}U(cT+c_2,\theta))|_{c_2=\beta h_2(c)}\left\{\begin{array}{c}>\\=\\<\end{array}\right\}0\iff c\left\{\begin{array}{c}<\\=\\>\end{array}\right\}\frac{1}{\alpha+n-1}\,.$$

But, by Lemma 3.1,

$$\frac{d}{dc}U(cT+c_2,\theta)\left\{\begin{array}{c}>\\=\\<\end{array}\right\}0 \iff c_2\left\{\begin{array}{c}<\\=\\>\end{array}\right\}\beta h_1(c)$$

and the result then follows from (3.1). □

Lemma A.1. *For $i=1,2$ and $x>0$,*

$$\frac{dU_i(x)}{dx}\propto(\alpha+n-1)\frac{1}{(\delta_i+x)^{\alpha+n+1}}(\hat{\lambda}_i-x).$$

Proof.

$$\begin{aligned}\frac{dU_i(x)}{dx} &\propto \frac{1}{(\delta_i+x)^{\alpha+n}}-(\alpha+n)\frac{x}{(\delta_i+x)^{\alpha+n+1}}\\ &= (\alpha+n-1)\frac{1}{(\delta_i+x)^{\alpha+n+1}}(\hat{\lambda}_i-x).\end{aligned}$$

□

Lemma A.2. *For $x>0$, $x\neq\hat{\lambda}_1$,*

$$\frac{d^2U_2}{dU_1^2}(x)=\frac{K(x)}{x-\hat{\lambda}_1}\left\{-(\hat{\lambda}_2-\hat{\lambda}_1)(\delta_1+x)(\delta_2+x)+(\alpha+n+1)(\delta_2-\delta_1)(\hat{\lambda}_2-x)(x-\hat{\lambda}_1)\right\},$$

where $K(x)>0$ for $x>0$.

Proof. Let $g(x)=dU_2(x)/dU_1(x)$. Then

$$\frac{d^2U_2}{dU_1^2}(x)=\frac{d\left(\dfrac{dU_2}{dU_1}\right)}{dU_1}(x)=\frac{dg(x)/dx}{dU_1(x)/dx} \tag{A.5}$$

with (see Theorem 4.1)

$$\left.\begin{aligned}\frac{dg(x)}{dx} &\propto \frac{d}{dx}\left(\frac{\hat{\lambda}_2-x}{\hat{\lambda}_1-x}\left(\frac{\delta_1+x}{\delta_2+x}\right)^{\alpha+n+1}\right)\\ &= \left(\frac{\delta_1+x}{\delta_2+x}\right)^{\alpha+n}\frac{1}{\delta_2+x}\frac{1}{x-\hat{\lambda}_1}\left(\frac{\hat{\lambda}_2-\hat{\lambda}_1}{x-\hat{\lambda}_1}(\delta_1+x)\right.\\ &+ \left.(\alpha+n+1)\frac{\delta_2-\delta_1}{\delta_2+x}(x-\hat{\lambda}_2)\right).\end{aligned}\right\} \tag{A.6}$$

The result then follows from (A.5) and Lemma A.1. □

Let

$$\left.\begin{aligned}A &= -(\alpha+n)^2\\ B &= (\alpha+n-1)(\hat{\lambda}_1+\hat{\lambda}_2)(\alpha+n)\\ C &= -2\hat{\lambda}_1\hat{\lambda}_2(\alpha+n-1)(\alpha+n)\end{aligned}\right\} \tag{A.7}$$

and let $H(x) = Ax^2 + Bx + C$. Then (see Lemma A.2) for $x \neq \hat{\lambda}_1$,

$$\frac{d^2U_2}{dU_1^2}(x) = \frac{H(x)}{K(x)(x-\hat{\lambda}_1)}. \tag{A.8}$$

The needed properties of $H(x)$ are given in the following lemmas.

Lemma A.3. *For $i = 1, 2$, $H(\hat{\lambda}_i) < 0$.*

Proof. For $i = 1, 2$

$$\begin{aligned} H(\hat{\lambda}_i) &= A\hat{\lambda}_i^2 + B\hat{\lambda}_i + C = \\ &= -\hat{\lambda}_i^2(\alpha+n)^2 - \hat{\lambda}_1\hat{\lambda}_2(\alpha+n-1)(\alpha+n) < 0. \end{aligned}$$

□

The following lemma follows directly from the definition of $H(x)$.

Lemma A.4. *For the derivative of $H(x)$ with respect to x we have*

$$\frac{d}{dx}H(x)\mid_{x = \hat{\lambda}_2}$$

$$= -(\hat{\lambda}_2 - \hat{\lambda}_1)(\alpha+n-1)(\alpha+n+2) - 2\hat{\lambda}_1 - 2\hat{\lambda}_2(\alpha+n-1) < 0.$$

The next lemma gives conditions under which $Ax^2 + Bx + C = 0$ has two, one or zero solutions.

Lemma A.5. *Let*

$$s(\alpha, n) = 8\frac{\alpha+n}{\alpha+n-1}. \tag{A.9}$$

Then

$$B^2 - 4AC \left\{ \begin{array}{c} < \\ = \\ > \end{array} \right\} 0 \Longleftrightarrow \frac{\hat{\lambda}_2}{\hat{\lambda}_1} \left\{ \begin{array}{c} < \\ = \\ > \end{array} \right\} r(\alpha, n)$$

where $r(\alpha, n)$ is the unique root > 1 of $r^2 + (2 - s(\alpha, n))r + 1 = 0$.

Proof. First note that

$$B^2 - 4AC = (\alpha+n-1)^2(\hat{\lambda}_1 + \hat{\lambda}_2)^2(\alpha+n)^2 - 8(\alpha+n)^3(\alpha+n-1)\hat{\lambda}_1\hat{\lambda}_2$$

$$\left\{ \begin{array}{c} < \\ = \\ > \end{array} \right\} 0 \Longleftrightarrow \frac{(\hat{\lambda}_1 + \hat{\lambda}_2)^2}{\hat{\lambda}_1\hat{\lambda}_2} \left\{ \begin{array}{c} < \\ = \\ > \end{array} \right\} s(\alpha, n).$$

Further, because $s(\alpha, n) > 8$, $r^2 + (2 - s(\alpha, n))r + 1 = 0$ has exactly two roots, $r_0 < r_1$, say, with $r_0 < 1 < r_1$. □

In the next lemma, assuming $B^2 - 4AC > 0$, the location of the roots of $Ax^2 + Bx + C = 0$ is investigated.

Lemma A.6.
$$B^2 - 4AC > 0 \Longrightarrow \hat{\lambda}_1 < x_1 < x_0 < x_2 < \hat{\lambda}_2,$$
where x_0 *maximizes* $Ax^2 + Bx + C$ *and* $x_1 < x_2$ *are the roots of* $Ax^2 + Bx + C = 0$.

Proof. First note that the lemmas A.3 and A.4 imply that $x_2 < \hat{\lambda}_2$. Further, $B > -2A\hat{\lambda}_1$ is equivalent to $B^2 > -2AB\hat{\lambda}_1$. So, in order to show that $x_0 > \hat{\lambda}_1$, it is sufficient to show that $4AC > -2AB\hat{\lambda}_1$. But

$$4AC > -2AB\hat{\lambda}_1 \Longleftrightarrow -2C > B\hat{\lambda}_1 \Longleftrightarrow$$

$$(n+\alpha-1)(n+\alpha)(\hat{\lambda}_1 + \hat{\lambda}_2) < 4(n+\alpha-1)(n+\alpha)\hat{\lambda}_2.$$

Then using the fact that $(n+\alpha) < 2(n+\alpha)$ and $\hat{\lambda}_1 + \hat{\lambda}_2 < 2\hat{\lambda}_2$ shows that $x_0 > \hat{\lambda}_1$. Finally, given that $x_0 > \hat{\lambda}_1$, it follows from Lemma A.3 with $i = 1$ that $x_1 > \hat{\lambda}_1$. □

From the above lemmas we get

Lemma A.7. *On* $(\hat{\lambda}_1, \hat{\lambda}_2)$

1) when $\dfrac{\hat{\lambda}_2}{\hat{\lambda}_1} \le r(\alpha, n)$
$$\frac{d^2U_2}{dU_1^2}(x) \le 0;$$

2) when $\dfrac{\hat{\lambda}_2}{\hat{\lambda}_1} > r(\alpha, n)$
$$\frac{d^2U_2}{dU_1^2}(x) \begin{cases} < 0 & \text{when } \hat{\lambda}_1 < x < x_1 \\ = 0 & \text{when } \hat{\lambda}_1 = x_1 \\ > 0 & \text{when } x_1 < x < x_2 \\ = 0 & \text{when } x = x_2 \\ < 0 & \text{when } x_2 < x < \hat{\lambda}_2. \end{cases}$$

References

Akaike, H. (1977). On entropy maximization principle. In: *Applications of statistics*, Proceedings of the symposium held at Wright State University, Dayton, Ohio, 14–18 June 1976. (P.R. Krishnaiah, Editor), Elsevier, North-Holland, 27–41.

Akaike, H. (1981). Likelihood of a model and information criteria. *J. Econometrics*, 16, 3–14.

Al-Hussaini, E.K. and Ahmed, A.E.A. (2003). On Bayesian interval prediction of future records. *Test*, 12, 79–99.

Geisser, S. (1985). Interval prediction for Pareto and exponential observables. *J. Econometrics*, 29, 173–185.

Geisser, S. (1993). *Predictive inference: An introduction.* London: Chapman and Hall.

Genest, C. and Zidek, J.V. (1986). Combining probability distributions: a critique and an annotated bibliography. *Statist. Sci.*, 1, 114–148.

Haines, L.M. (2003). An approach to simple bargaining games and related problems. *J. Statist. Plann. Inference*, 116, 353–366.

Lindley, D.V. (1976). A class of utility functions. *Ann. Statist.*, 4, 1–10.

Parsons, S. and Wooldridge, M. (2002). Game theory and decision theory in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 5, 243–254.

Radner, R. (1962). Team decision problems. *Ann. Math. Statist.*, 33, 857–881.

Tobías, A. and Scotto, M.G. (2005). Prediction of extreme ozone levels in Barcelona, Spain. *Environmental Monitoring and Assessment*, 100, 23–32.

van Eeden, C. and Zidek, J.V. (1994). Group Bayes estimation of the exponential mean: A retrospective view of the Wald theory. In: *Statistical Decision Theory and Related Topics V* (S.S. Gupta and J.O. Berger, eds.), Springer–Verlag, 35–49.

van Eeden, C. and Zidek, J.V. (2005). Multi-agent predictors of an exponential interevent time. Tech. Report Nr. 215, Department of Statistics, University of British Columbia, Vancouver, Canada.

Weerhandi, S. and Zidek, J.V. (1983). Elements of multi-Bayesian decision theory, *Ann. Statist.*, 11, 1032–1046.