Multi-Agent Predictors of an Exponential Interevent Time*

Constance van Eeden & James V Zidek
University of British Columbia

November 29, 2005

Abstract

This paper presents a solution to a problem in multiagent statistical decision analysis. More precisely, a group of Bayesians are jointly required to predict an exponentially distributed random inter-event time, S. They are to base that prediction on shared, previously observed, independent and identically distributed realizations of S. Although their priors may be different, they have the same conjugate utility function for prediction. That utility is shown to yield a conjugate utility for estimating the population-mean inter-event time. Thus prediction gives way to estimation. For the case of just two agents, the Pareto efficient boundary of the set of utilities generated by the class of all non-randomized linear prediction rules is explored. Conditions are given under which those rules are G-complete within the class of non-randomized linear predictors, meaning that optimum non-random predictors can be found on the Pareto boundary thereby providing a basis for a meaningful consensus. It is shown that the pre-posterior probability of such consensus is 1 under reasonable, explicit conditions. Finally, several paradigms are considered for selecting the compromise predictor and their implied solutions for a general G characterized.

*The work reported in this paper was supported by a grant from the Natural Science and Engineering Research Council of Canada

1 Introduction

This paper concerns the prediction of a future exponentially distributed random variable S by a group of G Bayesian agents. These agents could in fact be robots, each monitoring a single node of a process, for example. The S in question could be an inter-event time or a survival time. In any case, the agents share n independently observed realizations of S. Moreover, although they may have different prior knowledge, they do have a common conjugate utility function. The latter could be viewed as that of the organization employing them.

This paper is more generally concerned with paradigms that can be invoked in multiagent decision problems. Although their implications are worked out in just the special context of the prediction problem addressed in this paper, the nature of those results are suggestive, indicating what might obtain in other contexts as well.

The paradigms explored in this paper are:

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 predictor.

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 multi-agent decision problem where the agents would have an individually preferred (Bayesian) prediction strategy but would be forced to seek a compromise in a group decision problem.

Interest focuses on the similarities as well as differences in the predictors they produce. However, selection of the best paradigm for a particular application would depend on the context.

With respect to Paradigm #3, the optimal joint predictor of S may well be randomized, since members of the class of conjugate utility functions adopted in this paper are not concave. In other words, the agents may have to select at random amongst a set of non-randomized predictors. However, that sort of procedure would not generally be considered practical. Therefore conditions are derived, at least in the case of $G = 2$ agents, that ensure the optimal procedure is non-randomized. Since those conditions depend on the data, their attainment is generally subject to sampling uncertainty. Yet surprisingly at least in some situations presented below, their attainment is certain *pre-posteriori.* Thus, in some situations the two agents are assured of a realistic consensual choice of a compromise predictor. The case of $G > 2$ remains an open problem.

The paper begins with Section 2 which addresses the case of a single agent. There a conjugate utility function for predicting the random exponential variable, S, is adopted. That leads to an equivalent problem for the estimation of λ. 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 viewed as a predictor of S for that Bayesian. Section 3 turns to the case of $G > 1$ agents and a number of general results are given. However, the principal result there, a group admissibility result, is for just $G = 2$. Is there a basis for a consensual choice of a nonrandomized estimator by these two agents? Section 4 addresses that question and develops conditions where the answer to this question is affirmative. 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 the three paradigms described above, 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.

2 A single agent

This section develops the problem of finding a predictor of an exponential random variable S for 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.
Suppose the decision maker observes $S_1, \ldots, S_n \overset{i.i.d}{\sim} \exp(\lambda)$ so that

$$\begin{aligned} f_{S_i}(t|\lambda) &= \frac{1}{\lambda}\exp\left(-\frac{s}{\lambda}\right),\ t > 0 \\ E(S_i|\lambda) &= \lambda \\ V(S_i|\lambda) &= \lambda^2 \end{aligned}$$

for all $i = i, \ldots, n$. Then if $T = \sum_{i=1}^{n} S_i$ denotes the sufficient statistic it has conditional density function,

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

Moreover,

$$\begin{aligned} E(T|\lambda) &\equiv \mu(T|\lambda) \\ &= n\lambda \text{ and} \\ V(T|\lambda) &\equiv \sigma^2(T|\lambda) \\ &= n\lambda^2. \end{aligned}$$

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

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

where $\theta = (\alpha, \beta, \gamma)$ denotes the vector of hyperparameters. Thus,

$$E(\lambda|\theta) \equiv \mu(\lambda|\theta)$$

$$
\begin{aligned}
&= \frac{\beta}{\alpha - 2}, \ \alpha > 2 \quad \text{and} \\
V(\lambda|\theta) &= \frac{\mu(\lambda|\theta)^2}{\alpha - 3}, \ \alpha > 3.
\end{aligned}
$$

Then this decision maker's marginal density function for T is

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

where

$$
C(n, \alpha) = \frac{\Gamma(\alpha + n - 1)}{\Gamma(n)\Gamma(\alpha - 1)}.
$$

It readily follows that

$$
\begin{aligned}
E[T|\theta] &\equiv \mu(T|\theta) \\
&= \frac{n}{\alpha} \\
V[T|\theta] &\equiv \sigma(T|\theta), \quad \text{and} \\
&= E[\sigma(T|\lambda)|\theta] + V[E(T|\lambda)|\theta] \\
&= \frac{n(\alpha + n - 2)\mu^2(T|\theta)}{(\alpha - 3)}.
\end{aligned}
$$

Finally we may compute the Agent's posterior density function conditional on the data, that is on the value of the sufficient statistics $T = t$,

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

It readily follows that

$$
\begin{aligned}
E[\lambda|t, \theta] &\equiv \mu(T|t, \theta) \\
&= \frac{\beta + t}{\alpha + n - 2}, \ \alpha + n > 2, \quad \text{and} \\
V[\lambda|t, \theta] &\equiv \sigma(T|t, \theta) \\
&= \frac{\mu^2(\lambda|t, \theta)}{(\alpha + n - 3)}, \ \alpha + n > 3.
\end{aligned}
$$

Observe that the just computed mean 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 the maximum likelihood estimator of λ.

One solution to the prediction problem is a predictive distribution, $f_S(s|\hat{\lambda})$ for S, $\hat{\lambda}$ being an estimator of λ. That solution yields not only a point predictor, $\hat{S} = \int s f_S(s|\hat{\lambda})ds$ but also prediction intervals.

Following a common practice, the performance of that predictive distribution conditional on λ, is quantified by

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

the Kullback-Leibler measure of the discrepancy between the true distribution and the predictive distribution for S. Ideally $\hat{\lambda}$ should be chosen to minimize the quantity in Equation (2.4). Thus the problem of finding a good predictor for S has formally turned into an estimation problem, that of estimating λ with the loss function given in Equation (2.4). The latter is often referred to as *entropy loss*.

However, that commonly used loss function turns out to be unrealistic since it is unbounded. Moreover, it is mathematically untractable. Thus, this formulation of the prediction problem has proven unsatisfactory. So we consider an alternative approach that looks for a point predictor of S instead of a predictive distribution. Moreover, for convenience we measure performance by means of a utility function $U(\hat{S}, S, \lambda)$ instead of a loss function. Its conditional expectation given λ is then given by

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

and we seek to maximize the expected gain in utility in Equation (2.5). Following Lindley (1976), we adopt a conjugate utility function

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

We now compute the conditional expectation of the utility in Equation (2.6) as

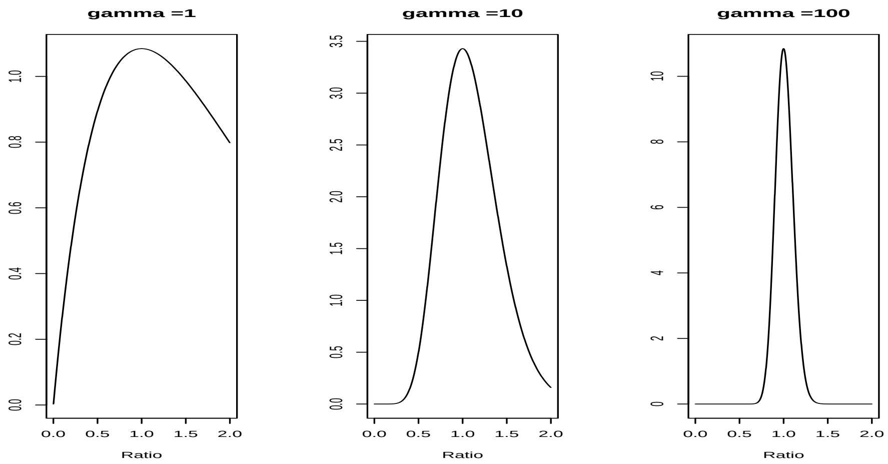

Figure 1: Conjugate utility functions for the prediction of S for varying γ, plotted against the S-Ratio, $\hat{S}/S$.

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

If we substitute $\hat{\lambda}$ for $\hat{S}$ in Equation (2.7) we obtain what may formally be viewed as a gain in 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 measure of utility.
However, that utility is not very tractable, in particular, it is not conjugate for the λ - estimation problem. But we can obtain a conjugate utility as an approximation by letting $\gamma \to \infty$. In fact as Figure 1 shows, the conjugate utility in Equation (2.6) for predicting S, which effectively tends to the 0-1 utility as γ increases, does not change very rapidly as λ increases. Moreover, the derived utility for estimating λ in Equation 2.8 is well approximated for λ's as small as 10 as Figure 2 demonstrates.

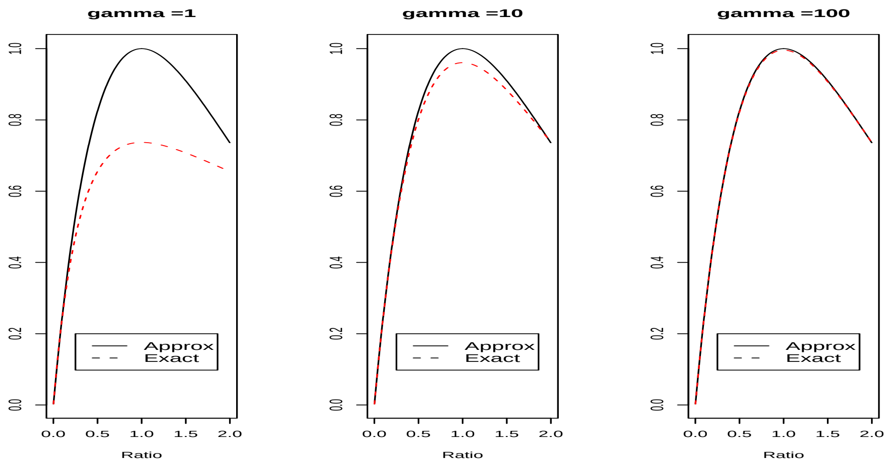

Figure 2: Approximate and exact utility functions for estimating λ for varying γ, plotted against the λ-Ratio, $\hat{\lambda}/\lambda$.

Arriving at the approximation requires Stirling's approximation,

$$\Gamma(1+\gamma) \sim \sqrt{(2\pi)}\exp(-\gamma)\gamma^{\gamma+\frac{1}{2}}.$$

The result after the substitution is a conjugate utility

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

In fact this has the same form as that adopted above for the prediction of S except that $\gamma = 1$ in this case. This then is the utility we adopt in this paper. With this conjugate utility for λ, 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*

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

Proof. First note that conditionally on λ,

$$\left.\begin{aligned}
&\mathcal{E}[u(c_1T+c_2,\lambda,\theta)\mid\lambda]\\
&=\frac{e}{\lambda}\int_0^\infty (c_1t+c_2)\ e^{-(c_1t+c_2)/\lambda}\ \frac{t^{n-1}}{\lambda^n\Gamma(n)}\ e^{-t/\lambda}dt\\
&=\frac{e^{(1-(c_2/\lambda))}}{\lambda^{n+1}\Gamma(n)}\int_0^\infty \left(\lambda\frac{c_1}{c_1+1}s+c_2\right)e^{-s}s^{n-1}\left(\frac{\lambda}{c_1+1}\right)^n ds\\
&=\frac{e^{(1-(c_2/\lambda))}}{\lambda^{n+1}\Gamma(n)}\left(\frac{\lambda}{c_1+1}\right)^n\left\{\lambda\frac{c_1}{c_1+1}\Gamma(n+1)+c_2\Gamma(n)\right\}\\
&=\frac{e^{(1-(c_2/\lambda))}}{\lambda(c_1+1)^{n+1}}\{c_1n\lambda+c_2(c_1+1)\}.
\end{aligned}\right\}\qquad(2.10)$$

From (2.10) one obtains

$$\begin{aligned}
U(c_1T+c_2,\theta) &= \mathcal{E}u(c_1T+c_2,\lambda,\theta)\\
&= \frac{e}{(c_1+1)^{n+1}}\left\{c_1n\int_0^\infty e^{-c_2/\lambda}\ \frac{\beta^{\alpha-1}}{\lambda^\alpha\Gamma(\alpha-1)}e^{-\beta/\lambda}d\lambda\right.\\
&+ \left. c_2(c_1+1)\int_0^\infty \frac{e^{-c_2/\lambda}}{\lambda}\frac{\beta^{\alpha-1}}{\lambda^\alpha\Gamma(\alpha-1)}e^{-\beta/\lambda}d\lambda\right\}\\
&= \frac{e}{(c_1+1)^{n+1}}\frac{\beta^{\alpha-1}}{\Gamma(\alpha-1)}\left\{c_1n\frac{\Gamma(\alpha-1)}{(c_2+\beta)^{\alpha-1}}+c_2(c_1+1)\frac{\Gamma(\alpha)}{(c_2+\beta)^\alpha}\right\},
\end{aligned}$$

from which the result follows immediately. ♡

We now determine the Bayes rule for the conjugate utility.

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*

$$\hat{\lambda}_B(t)=\frac{t+\beta}{\alpha+n-1}. \qquad (2.11)$$

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

$$\begin{cases}\dfrac{t^{n-1}e^{-t/\lambda}}{\lambda^n\Gamma(n)} & \text{if } t>0\\ 0 & \text{if } t\le 0,\end{cases}$$

the joint density of T and λ is given by

$$\begin{cases} \dfrac{t^{n-1}e^{-t/\lambda}}{\lambda^n\Gamma(n)}\dfrac{\beta^{\alpha-1}e^{-\beta/\lambda}}{\lambda^\alpha\Gamma(\alpha-1)} & \text{if } t>0, \lambda>0 \\ 0 & \text{if not.} \end{cases}$$

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

$$\pi(\lambda \mid t) = \begin{cases} \dfrac{(t+\beta)^{\alpha+n-1}}{\Gamma(\alpha+n-1)\lambda^{\alpha+n}} e^{-(t+\beta)/\lambda} & \text{if } \lambda > 0 \\ 0 & \text{if } \lambda \leq 0. \end{cases}$$

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

$$\left.\begin{aligned} U(\hat{\lambda}, \theta \mid t) &= \mathcal{E}\left\{u(\hat{\lambda}, \lambda, \theta) \mid t\right\} = \mathcal{E}\left\{\frac{\hat{\lambda}}{\lambda}\exp\left\{1-\frac{\hat{\lambda}}{\lambda}\right\} \mid t\right\} \\ &= e\,\hat{\lambda}\,\frac{(t+\beta)^{\alpha+n-1}}{\Gamma(\alpha+n-1)}\int_0^\infty \frac{1}{\lambda}\exp\left\{-\frac{t+\beta+\hat{\lambda}}{\lambda}\right\}\frac{1}{\lambda^{\alpha+n}}\,d\lambda \\ &= e\,\hat{\lambda}\,\frac{(t+\beta)^{\alpha+n-1}}{\Gamma(\alpha+n-1)}\,\frac{\Gamma(\alpha+n)}{(t+\beta+\hat{\lambda})^{\alpha+n}}\,. \end{aligned}\right\} \quad (2.12)$$

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

$$\frac{d}{d\hat{\lambda}}U(\hat{\lambda}, \theta \mid t) = 0$$

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 (2.11). ♡

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

$$\frac{e\Gamma(\alpha+n)}{\Gamma(\alpha+n-1)}\,\frac{(\alpha+n-1)^{\alpha+n-1}}{(\alpha+n)^{\alpha+n}}\,. \quad (2.13)$$

Proof. Apply Lemma 2.1 with $c_1 = (\alpha + n - 1)^{-1}$ and $c_2 = \beta(\alpha + n - 1)^{-1}$ to obtain (2.13). $\heartsuit$

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 for squared error loss. Further, from (2.13) 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 (2.13). In the next section each of the 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 in equation (2.8) is used. The Bayes estimator is obtained and a G-complete-class result is presented.

Now consider group G of (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*

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

Proof.

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

$$(c_1+1)[n(c_2+\beta)+c_2(\alpha-1)] \quad \left\{\begin{array}{c}>\\=\\<\end{array}\right\} \quad (n+1)[c_1n(c_2+\beta)+c_2(c_1+1)(\alpha-1)]$$

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

♡

Lemma 3.2 *Under the conditions of Lemma 2.1*

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

Proof.

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

$$(c_2+\beta)[c_1n+c_2(c_1+1)(\alpha-1)] \quad \left\{\begin{array}{c}>\\=\\<\end{array}\right\} \quad \alpha[c_1n(c_2+\beta)+c_2(c_1+1)(\alpha-1)]$$

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

♡

Now let (see Lemma 3.1 and Lemma 3.2)

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

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

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

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

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

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

and the pair (c_1, c_2) with

$$\begin{cases} c_1 = \dfrac{1}{\alpha + n - 1} \\ c_2 = \dfrac{\beta}{\alpha + n - 1} = \beta h_1\left(\dfrac{1}{\alpha + n - 1}\right) = \beta h_2\left(\dfrac{1}{\alpha + n - 1}\right) \end{cases}$$

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

Now let

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

Then it follows from Lemma 3.1, Lemma 3.2 and (3.1) that $U(c_1 T + c_2, \theta)$ 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 $1/(\alpha + n - 1) < c$.

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

$$\left.\begin{aligned} \frac{d}{dc}U(cT + \beta h_1(c), \theta) &= \frac{d}{dc}U(cT + c_2, \theta))|_{c_2=\beta h_1(c)} + \\ &\quad \frac{d}{dc_2}U(cT + c_2, \theta))|_{c_2=\beta h_1(c)} \frac{d}{dc}\beta h_1(c). \end{aligned}\right\} \tag{3.6}$$

The first term on the right hand side of (3.6) 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

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

But by Lemma 3.2

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

The result then follows from (3.1).

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

$$\left.\begin{aligned} \frac{d}{dc}U(cT + \beta h_2(c), \theta) &= \frac{d}{dc}U(cT + c_2, \theta))|_{c_2=\beta h_2(c)} + \\ &\quad \frac{d}{dc_2}U(cT + c_2, \theta))|_{c_2=\beta h_2(c)} \frac{d}{dc}\beta h_2(c) \end{aligned}\right\} \tag{3.7}$$

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

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

But, by Lemma 3.1,

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

and the result then follows from (3.1). ♡

The above given properties of $U(c_1T+c_2,\theta)$ as a function of c_1 and c_2 are summarized in Figure 3, where the arrows indicate the direction in which $U(c_1T+c_2,\theta)$ increases.

The following theorem gives our complete class result.

Theorem 3.1 *Let G be 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 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 $c_1=1/(\alpha+n-1)$, $\beta_1/(\alpha+n-1)\le c_2\le\beta_2/(\alpha+n-1)$. This can be seen as follows. First note that, by the definitions of $S_1(\beta)$ and $S_2(\beta)$

$$S^*=\{(c_1,c_2)\mid 0\le c_1\le 1/n, \beta_1M(c_1)\le c_2\le\beta_2 m(c_1)\}$$

where

$$\begin{aligned} m(c) &= \min\{h_1(c),h_2(c)\}\\ M(c) &= \max\{h_1(c),h_2(c)\}. \end{aligned}$$

Further (see (3.1))

$$\beta_2 m\left(\frac{1}{\alpha+n-1}\right)-\beta_1M(\frac{1}{\alpha+n-1})=(\beta_2-\beta_1)h_1\left(\frac{1}{\alpha+n-1}\right)>0.$$

That S^* is not empty then follows from the fact that $\beta_2 m(c) - \beta_1 M(c)$ is continuous in c on $0 \le c \le 1/(n-1)$.
Figures 4 and 5 summarize the properties of $U(c_1 T + c_2, \theta_j)$, $j = 1, 2$ as functions of c_1 and c_2, as well as the relationships between the sets $S_i(\beta_j)$, $i = 1, \ldots, 4$, $j = 1, 2$. The two cases considered are

$$\left.\begin{array}{ll} (i)\ \dfrac{\beta_1}{\alpha-2} \le \dfrac{\beta_2}{\alpha-1} & \text{in Figure 3.2} \\ (ii)\ \dfrac{\beta_1}{\alpha-2} > \dfrac{\beta_2}{\alpha-1} & \text{in Figure 3.3.} \end{array}\right\} \tag{3.8}$$

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

$$\left.\begin{aligned} H(c) &= \beta_2 h_2(c) - \beta_1 h_1(c) \\ &= \beta_2 \frac{1 - c(n-1)}{c(\alpha + n - 1 + \alpha - 1)} - \beta_1 \frac{1 - cn}{c(\alpha + n - 1) + \alpha - 2}. \end{aligned}\right\} \tag{3.9}$$

When $\beta_2/(\alpha-1) \ge \beta_1/(\alpha-2)$, it follows from (3.9) 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

$$\begin{aligned} &(\alpha + n - 1)(n\beta_1 - (n-1)\beta_2)c^2 + \\ &[(\alpha + n - 1)(\beta_2 - \beta_1) + \beta_1 n(\alpha - 1) - \beta_2(n-1)(\alpha - 2)]c + \beta_2(\alpha - 2) - \beta_1(\alpha - 1) \ge 0. \end{aligned}$$

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 $H(c) \ge 0$ if and only if

$$\begin{aligned} &(\alpha + n - 1)(\beta_1(n-1) - \beta_2 n)c^2 + \\ &[(\beta_2 - \beta_1)(\alpha + n - 1) - (\beta_2 n(\alpha - 1) - \beta_1(n-1)(\alpha - 2)]c + \beta_2(\alpha - 1) - \beta_1(\alpha - 2) \ge 0. \end{aligned}$$

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

$$
\left.
\begin{array}{ll}
U(c_1'T + c_2', \theta_j) \geq U(c_1 T + c_2, \theta_j), & j = 1, 2 \\
U(c_1'T + c_2', \theta_j) > U(c_1 T + c_2, \theta_j) & \text{for some } j \in \{1, 2\}.
\end{array}
\right\} \tag{3.10}
$$

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.

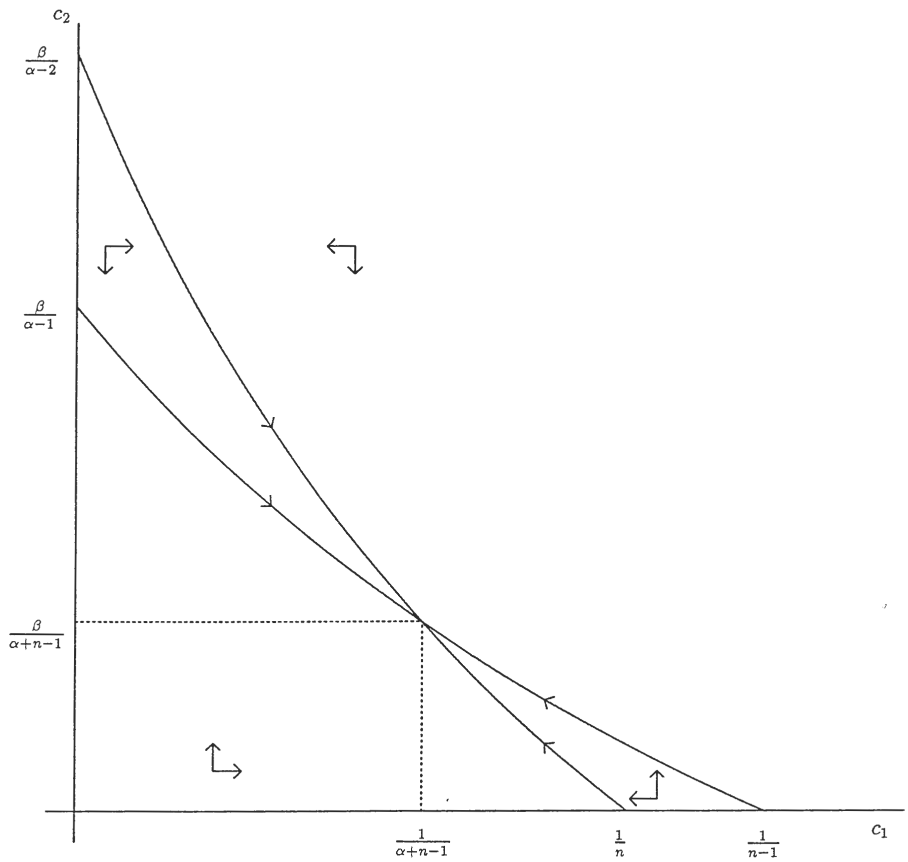

Figure 3: Behavior of $U(c_1T + c_2, \theta)$ as a function of $c_1 > 0$ and $c_2 > 0$

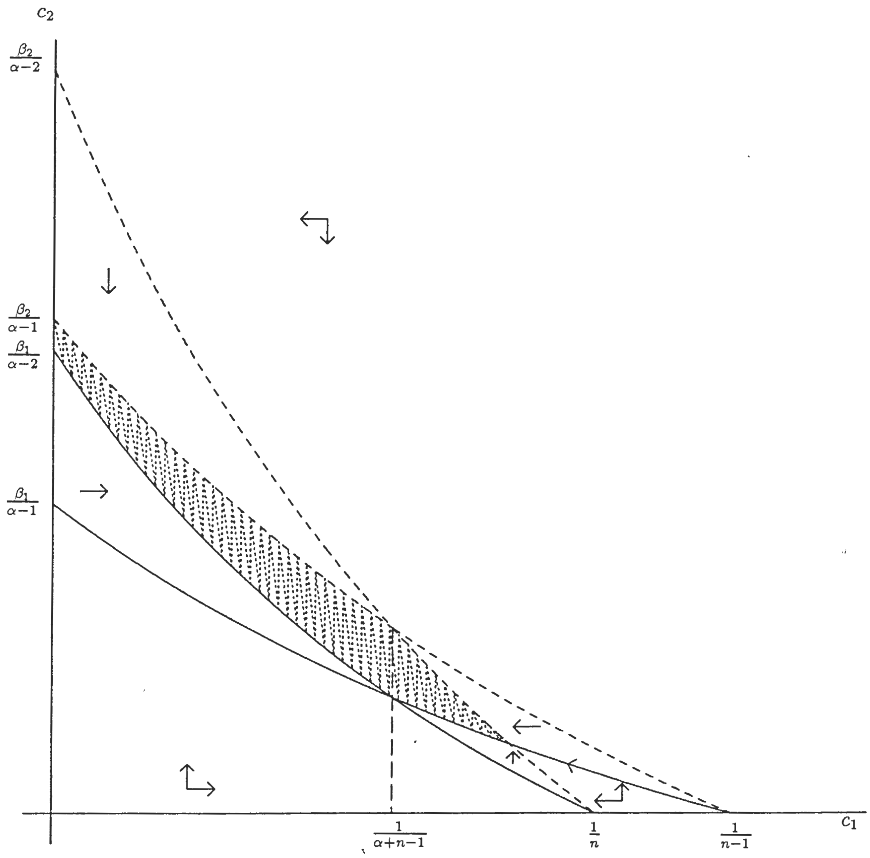

Figure 4: Behavior of $U_2(c_1T + c_2, \theta_j)$, $j = 1, 2$ as a function of c_1 and c_2 and β_j when $\beta_1(\alpha - 2)^{-1} \leq \beta_2(\alpha - 1)^{-1}$

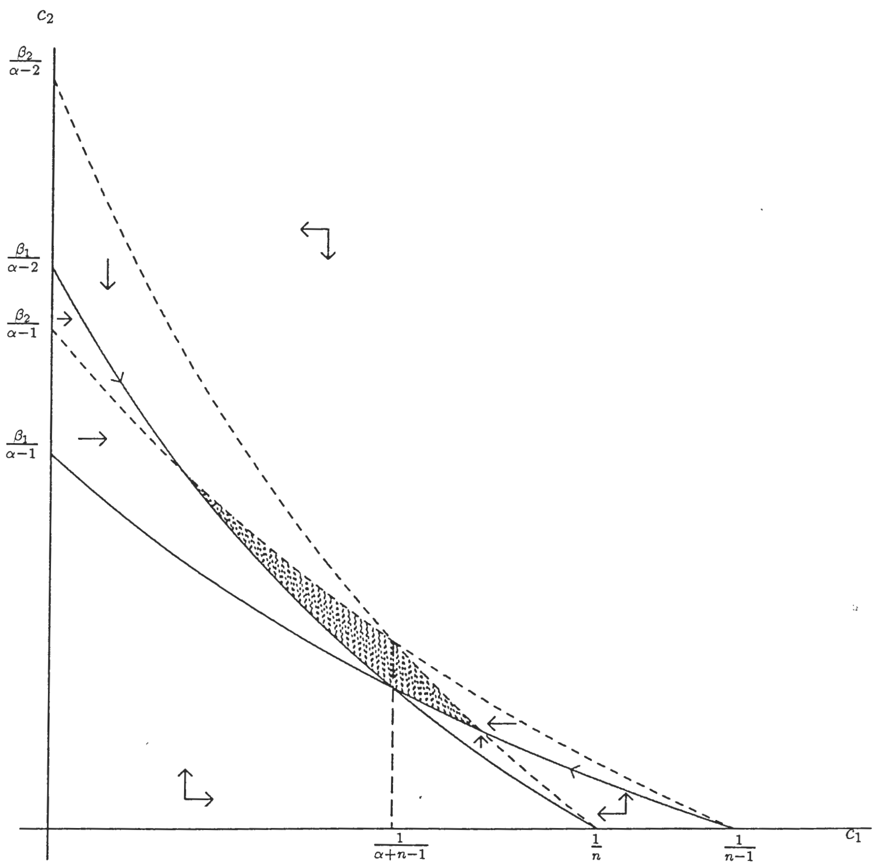

Figure 5: Behavior of $U_2(c_1T + c_2, \theta_j)$, $j = 1, 2$ as a function of c_1 and c_2 and β_j when $\beta_1(\alpha-2)^{-1} > \beta_2(\alpha-1)^{-1}$

4 Consensual Choice

In this section we consider the case where the group G 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 (2.11)) 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 (2.12)) 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 K_i is a positive constant independent of x. Further we suppose, without loss of generality, that $\hat{\lambda}_1 < \hat{\lambda}_2$.

The theorems below, whose proofs are given in the Appendix, give the needed properties of $U_2(x)$ as a function of $U_1(x)$ for $x > 0$.

Theorem 4.1 *For $x > 0$*

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

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.5), 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:
The result in Theorem 4.3 is not in agreement with Theorem 4.3 of van Eeden and Zidek (1994). 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

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

But, by Theorem 4.3 of van Eeden and Zidek (1994) we have concavity on $(\hat{\lambda}_1, \hat{\lambda}_2)$ if and only if

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

where

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

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

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

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, $\hat{\lambda}_2/\hat{\lambda}_1 \le r_1(\alpha, n)$?". 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

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

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

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

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 $n = 1, 2, \ldots,$*

$$\left.\begin{array}{rcl} 3+\sqrt{8} & < & r(\alpha, n+1) < r(\alpha, n) \\ & < & r(\alpha, 0) = 3 + \dfrac{4}{\alpha - 1} + \sqrt{\left(3 + \dfrac{4}{\alpha-1}\right)^2 - 1}. \end{array}\right\} \tag{4.4}$$

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

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

$$r(\alpha, N_o + 1) < \frac{\beta_2}{\beta_1} \leq r(\alpha, N_o), \tag{4.5}$$

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 (4.5). ♡

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 results 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 Selecting a predictor

The last two sections explored the Pareto boundary from which the agents would select their joint predictor of S. However, we have side-stepped the question of how the joint predictor might be chosen.

The answer depends on the specific context in which the predictor is being selected, in particular which of the three paradigms describe in the Introduction is to be used. This section investigates the implications of choosing each of these paradigms.

First consider Paradigm #1. Suppose the G agents share with the supra - Bayesian, the conjugate prior in Equation (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 a predictor 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 predictor for the supra-Bayesian (Agent 0):

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

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

Although Paradigm #1 and the approach taken above lead directly to a predictor for $\hat{S}$, 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:

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

the weights $\{\omega_i\}, \omega_i \geq 0, \sum \omega_i = 1$ reflecting 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. After all 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 self interest and yet choose a compromise that recognizes their individual positions. Paradigm #3 is most appropriate in that case.

The linearity in t of (2.11) suggests restricting the search for a compromise to the class of linear predictors $\hat{S} \equiv \hat{\lambda} = c_1 t + c_2, t > 0$. Each agent's expected gain in utility for members of that class appears in Equation (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 $\{\omega_i\}, \omega_i > 0, \sum_{i=1}^{G} \omega_1 = 1$ 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 of that includes randomized rules, even in the case $G = 2$ unless the conditions for consensus in that section are met.

6 Discussion

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

$$-\left[+ \frac{\hat{\lambda}}{\lambda} - \log\left[\frac{\hat{\lambda}}{\lambda}\right] - 1 \right]. \tag{6.2}$$

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.

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 Equation 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 fact 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 Equation 5.1. In other words, implementation of the Paradigm #1 will require some sophisticated modelling 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, it differs in a very fundamental respect that instead of trying to accumulate prior information as #1 does, it merely tries to deal with the competing priors 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 predictor.

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.

A Appendix

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

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

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

$$\frac{d^2U_2}{dU_1^2}(x) = \frac{d\left(\frac{dU_2}{dU_1}\right)}{dU_1}(x) = \frac{dg(x)/dx}{dU_1(x)/dx} \tag{A.3}$$

with (see Theorem 4.1)

$$\left.\begin{aligned}
\frac{dg(x)}{dx} &\propto \frac{d}{dx}\left(\frac{\hat{\lambda}_2 - x}{\hat{\lambda}_1 - x}\left(\frac{\delta_1 + x}{\delta_2 + x}\right)^{\alpha+n+1}\right) \\
&= \frac{d}{dx}\left(\left(\frac{\hat{\lambda}_2 - \hat{\lambda}_1}{\hat{\lambda}_1 - x} + 1\right)\left(1 - \frac{\delta_2 - \delta_1}{\delta_2 + x}\right)^{\alpha+n+1}\right) \\
&= \frac{\hat{\lambda}_2 - \hat{\lambda}_1}{(x - \hat{\lambda}_1)^2}\left(1 - \frac{\delta_2 - \delta_1}{\delta_2 + x}\right)^{\alpha+n+1} \\
&+ (\alpha + n + 1)\left(\frac{\hat{\lambda}_2 - \hat{\lambda}_1}{\hat{\lambda}_1 - x} + 1\right)\left(1 - \frac{\delta_2 - \delta_1}{\delta_2 + x}\right)^{\alpha+n}\frac{\delta_2 - \delta_1}{(\delta_2 + x)^2} \\
&= \left(\frac{\delta_1 + x}{\delta_2 + x}\right)^{\alpha+n}\frac{1}{\delta_2 + x}\frac{1}{x - \hat{\lambda}_1}\left(\frac{\hat{\lambda}_2 - \hat{\lambda}_1}{x - \hat{\lambda}_1}(\delta_1 + x)\right. \\
&+ \left.(\alpha + n + 1)\frac{\delta_2 - \delta_1}{\delta_2 + x}(x - \hat{\lambda}_2)\right).
\end{aligned}\right\} \tag{A.4}$$

The result then follows from (A.3) and Lemma A.1. ♡

Let

$$\begin{aligned}
A &= -(\alpha + n)^2 \\
B &= (\alpha + n - 1)(\hat{\lambda}_1 + \hat{\lambda}_2)(\alpha + n) \\
C &= -2\hat{\lambda}_1\hat{\lambda}_2(\alpha + n - 1)(\alpha + n)
\end{aligned} \tag{A.5}$$

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

$$\frac{d^2U_2}{dU_1^2}(x) = \frac{H(x)}{K(x)(x - \hat{\lambda}_1)}. \tag{A.6}$$

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 $i = 1, 2$

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

$$- \quad 2\hat{\lambda}_1\hat{\lambda}_2(\alpha+n-1)(\alpha+n)$$

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

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*

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

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

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

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

Lemma A.5 *Let*

$$s(\alpha,n) = 8\frac{\alpha+n}{\alpha+n-1}. \tag{A.7}$$

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 $r_0<1<r_1$. $\heartsuit$

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

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

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

But

$$(n + \alpha) < 2(n + \alpha)$$

and $\hat{\lambda}_1 + \hat{\lambda}_2 < 2\hat{\lambda}_2$ which proves 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.

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

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

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.

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