BAYES AND EMPIRICAL BAYES PROCEDURES
FOR COMPARING PARAMETERS

BY

MARC J. SOBEL

TECHNICAL REPORT NO. 382
SEPTEMBER 1991

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS89–05874

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Bayes and Empirical Bayes Procedures
For Comparing Parameters

By

Marc J. Sobel

Technical Report No. 382
September 1991

Prepared Under the Auspices
Of
National Science Foundation Grant DMS89–05874

Department of Statistics
Stanford University
Stanford, California
BAYES AND EMPIRICAL BAYES PROCEDURES
FOR COMPARING PARAMETERS

MARC J. SOBEL*

ABSTRACT

The hitting averages of a set of N=33 baseball players during world series play are modeled in two different ways according to whether all of the hitting averages are assumed to have independent binomial or independent normal distributions. It is desired to construct ranking procedures for comparing these hitting averages (corresponding to each of the models considered) under circumstances where the number of times at bat of each of the players may differ by a large amount from each other. A loss function is constructed which represents an appropriate measure of distance between a given ranking of the players and the true ranking of the players. Bayes and Empirical Bayes procedures based on this loss function are proposed and compared with "adapted ranking procedures" which arise through Bayes, Empirical Bayes, or Maximum Likelihood estimation of the hitting averages of the baseball players. Bayes and Empirical Bayes procedures constructed under both the normal and binomial models are shown to be transitive. Empirical Bayes ranking procedures constructed here arise from maximizing a function of the prior parameters which measures the amount of confidence we have in a Bayes procedure. It is shown that the ranking procedures proposed here are more robust than adapted ranking procedures.

*Marc Sobel is Assistant Professor, Department of Statistics, Temple University, Phil., PA 19122
1. INTRODUCTION

A great deal of work has been done toward solving the problem of simultaneously estimating the location parameters from $k$ independent (but not necessarily identical) populations on the basis of observing independent samples from each of these $k$ populations under various assumptions about the underlying model (see e.g., Efron and Morris (1973) and (1975); Maritz and Lwin (1989); and Berger (1985). Simultaneous Bayes, Empirical Bayes, and Hierarchical Bayes estimators of location parameters have been shown to have many desirable properties. Consider for example the problem of comparing the world series hitting abilities (i.e., ideal number of hits out of total "at bats") of baseball players in world series play. We model this in two different ways: I) A transformed version of the number of hits of each given player is constructed: each such version has a normal distribution independent of all other players with mean and variance depending on the number of "at bats" of the players (see e.g., Efron and Morris (1975)); and II) It is assumed that the number of hits of each of the players has a binomial distribution independent of the other players; Model I will henceforth be called the normal model and Model II the binomial model. Under the normal model the authors cited above have demonstrated many desirable properties for Empirical Bayes simultaneous estimators of the transformed proportion of hits for each of the players. To our knowledge little work has been done using such techniques to simultaneously estimate the ideal proportions of hits of each of the players in the binomial model. Adopting a different viewpoint, we are interested in the players primarily in terms of the relative ranks of their hitting abilities.

Of course, players could be ranked by first estimating their relative abilities using Bayes, Empirical Bayes, or maximum likelihood simultaneous estimators and then ranking the players on the basis of these estimates but, as we shall see, there are built-in limitations to this strategy. Let us adopt the terminology, "adapted ranking procedure" for one which first simultaneously estimates the hitting abilities of the players using Bayes, Empirical Bayes, or maximum likelihood estimators and then ranks (the hitting abilities of) the players on the basis of these estimates. In this paper we employ a ranking-type loss function measuring how different the proposed ranking is from the true one. This differs from the conventional estimation-type loss functions used to evaluate simultane-
ous estimation procedures. Bayes and Empirical Bayes ranking procedures based on this ranking type loss function are constructed under the assumption that the parameters are independently distributed according to various conjugate prior distributions. Empirical Bayes ranking procedures arise through maximizing a function of the prior parameters (hyperparameters) which measures the amount of confidence we have in a ranking. The Bayes and Empirical Bayes procedures proposed here are all admissible relative to the loss function proposed below. A substantial body of work has been published concerning the question of constructing procedures for ranking items (see e.g., Gupta and Hsiao (1983), Laird and Louis (1989), Berger and Deeley (1988), and Portnoy (1982)).

Adapted ranking procedures arise from and are evaluated according to a Euclidean-type distance between the (vector) estimator and the (vector) parameter to be estimated while Bayes and Empirical Bayes ranking procedures arise from and are evaluated according to a notion of the ranking distance between the procedure and the true (parametric) ranking. This distinction between Bayes and Empirical Bayes ranking and adapted ranking procedures serves to indicate differences of efficiency, robustness, and etc.. between the two types of procedures: 1) We show below that, for given observed data, Bayes ranking and adapted ranking procedures frequently differ substantially from one another (depending on the assumed joint prior distribution of the parameters); 2) Relative comparisons (or orderings) between parameters are harder to make accurately, the closer they are to one another, whereas, it is "easier" to simultaneously estimate parameters in this case; and 3) It is also shown below that Bayes adapted ranking procedures are non-robust in the sense of being sensitive to the presence of (past) short hitting slumps or streaks while Bayes (non-adaptive) ranking procedures are robust in this sense; (past) short hitting slumps or streaks are modeled as small changes in the assumed joint prior distribution of the parameters. The differences between the robustness properties of adapted and non-adapted ranking procedures arise primarily from the fact that Bayes estimators are usually more sensitive than Bayes tests and confidence intervals to small changes in the prior distribution. Essentially, Bayes ranking procedures, as defined below, can be regarded as a combination of many correlated tests each one of which determines which of two
parameters are larger.

In a 1973 paper Efron and Morris (i.e., Efron and Morris (1973)) discuss the notion of "(posterior) relative savings loss". This quantity measures the posterior regret associated with using a given simultaneous estimation procedure \( \delta \) rather than the relevant Bayes (estimation) procedure. We adopt an analogous notion of regret substituting a ranking distance for the standard Euclidean distance. This allows us to measure a) the posterior regret associated with using a given adapted ranking procedure rather than the analogous Bayes ranking procedure and b) the sensitivity of a Bayes ranking procedure to changes in the prior distribution.

2. CONSTRUCTION OF THE GENERAL MODEL

In what follows we employ the terminology \( I(\bullet) \) for the indicator function of \( \bullet \), \( a_1, \ldots, a_N \) for some known real constants, and \( \mathcal{S}, \Omega \subset \mathbb{R} \) for the respective sample and parameter spaces of the random variables \( W_1, \ldots, W_N \), defined below. A certain number \( N \) of independent variables \( W_1, \ldots, W_N \) with respective densities (or probability functions)

\[
f_i(w|\psi_i) = \exp \{ \psi_i T(w) + a_i d(\psi_i) + S_i(w) \} I_S(w), \quad \psi_i \in \Omega_i (i = 1, \ldots, N)
\]  

(2.1)

all of which belong to a particular one-parameter exponential family, are observed (see e.g., Lehmann (1986)); these are the (possibly transformed) number of hits of the \( N \) players in world series competition. The parameters \( \psi_1, \ldots, \psi_N \) correspond to (possibly transformed) measures of the hitting ability of the \( N \) players considered; the goal is to rank these \( N \) parameters in **ascending order** (e.g., higher ranks indicate greater hitting ability) on the basis of the observed number of hits of each of the \( N \) players. We assume the parameters \( \psi_1, \ldots, \psi_N \) are independently distributed according to continuous conjugate prior densities \( \pi_1(\psi_1), \ldots, \pi_N(\psi_N) \) having the form

\[
\pi_i(\psi_i|t_i, s_i) \propto \exp \{ t_i \psi_i + s_i d(\psi_i) \} I(\psi_i \in \Omega) \quad (i = 1, \ldots, N)
\]  

(2.2)

The terminology \( \mathcal{F}(d, \Omega) \) is used to denote the particular exponential family specified by (2.2). The quantities \( t_1, \ldots, t_N \) and \( s_1, \ldots, s_N \) are hyperparameters which are either known in advance (i.e., a Bayes formulation), estimated on the basis of the data (i.e., an Empirical Bayes formulation), or have a known distribution (i.e., a hierarchical Bayes formulation).
The formulation introduced above has three important advantages: 1) It covers the cases of widest interest, 2) It seems to be the most general formulation under which very general conditions for the transitivity of the ranking procedures constructed here hold, and 3) It provides examples of Bayes and Empirical Bayes ranking procedures which differ markedly from the corresponding Bayes and Empirical Bayes adapted ranking procedures.

2.1 Construction of the Loss Function

In this subsection we define the loss $L(\psi, \delta)$ incurred by a procedure $\delta$ designed to rank the $N$ parameters $\psi_1, \ldots, \psi_N$. In the sequel the abbreviated notation $P_\pi(\bullet|W)$ for the posterior probability of the event $\bullet$ given the observations $W = (W_1, \ldots, W_N)$ will be used. Define a "tournament" procedure $\delta$ as one which specifies a binary relation between any pair $(\psi_\alpha, \psi_\beta)$ ($1 \leq \alpha, \beta \leq N; \alpha \neq \beta$) of parameters; we write $\delta_\alpha < \delta_\beta$ in case the parameter $\psi_\beta$ is judged to be larger than $\psi_\alpha$ and $\delta_\beta < \delta_\alpha$ if the converse is true; we write $\delta_\alpha \equiv \delta_\beta$ in case the parameter $\psi_\alpha$ is judged to be neither larger nor smaller than the parameter $\psi_\beta$. In case of a tie (e.g., $\delta_\alpha \equiv \delta_\beta$) we randomize in the usual way; we refer to this as rank randomization in the presence of ties. Note that by assumption of the continuity of the prior densities considered we can and will assume that there are no a-priori ties among the parameters $\psi_1, \ldots, \psi_N$, even though a given ranking procedure may specify some number of ties. Tournament procedures need not correspond to ranking procedures (i.e., procedures which assign a unique well-defined rank to each of the parameters); they may lack the property of being transitive (i.e., $\delta_k < \delta_j$, and $\delta_j < \delta_i \Rightarrow \delta_k < \delta_i$; ($1 \leq i, j, k \leq N$)). We use the terminology "ranking procedure" to denote each transitive tournament procedure referred to in this paper; in all other cases we refer to intransitive tournament procedures. Below, in section 2.5 general conditions (on the assumed model and joint prior distribution), under which a tournament procedure is a ranking procedure, are demonstrated; all of the procedures constructed in the remainder of this paper satisfy these conditions. This justifies our calling all the procedures discussed in the sequel ranking procedures.

We employ the loss function $L(\psi, \delta)$ which penalizes the practitioner by the amount

$$L(\psi, \delta) = \sum_{1 \leq \alpha \neq \beta \leq N} I(\psi_\alpha < \psi_\beta)I(\delta_\beta < \delta_\alpha) + (1/2) \sum_{1 \leq \alpha \neq \beta \leq N} I(\delta_\alpha \equiv \delta_\beta)$$ (2.3)
when the parameters are given by \( \psi = (\psi_1, ..., \psi_N) \) and the ranking procedure \( \delta \) is used to rank the parameters. This loss function originally arose in the context of rank tests (see e.g., Kendall(1970)) but has since found other statistical applications as a loss function (see e.g., Sobel(1989a), Sobel(1989b), and Govindarajulu (1974)).

It is easy to show that the Bayes ranking procedure, denoted here by \( \delta^* = \delta^*(W) \), which minimizes the Bayes risk associated with the loss function \( L \), defined above, and given independent prior densities \( \pi_1, ..., \pi_N \) as above has the form

\[
\begin{align*}
\delta^*_\alpha &< \delta^*_\beta \quad \text{if} \quad P_\pi(\psi_\alpha > \psi_\beta | W) < (1/2) \\
\delta^*_\beta &< \delta^*_\alpha \quad \text{if} \quad P_\pi(\psi_\alpha > \psi_\beta | W) > (1/2) \\
\delta^*_\alpha & = \delta^*_\beta \quad \text{otherwise}
\end{align*}
\]

for \( 1 \leq \alpha \neq \beta \leq N \). In the sequel, without loss of generality, we omit discussion of ties for all Bayes and Empirical Bayes (non-adaptive) ranking procedures proposed here and assume that the above defined rank-randomization technique is always used when appropriate.

2.2 Construction of Bayes Ranking Procedure

We consider two standard ways of modeling the number of hits out of number of ATBATS in world series play. It will be seen below that each model gives rise to a Bayes ranking procedure different from the maximum likelihood (ML) adapted ranking procedure (which ranks these \( N=33 \) players according to the values of the sample proportion of hits \( \hat{p}_1, ..., \hat{p}_N \) given in Table 1A of APPENDIX I) and different (under the Binomial but not the normal model) from the adapted ranking procedure (PM) which ranks players on the basis of their posterior mean hitting averages.

2.3 Bayes Ranking Procedures: Binomial Model

Under the Binomial model it is assumed that the observed number of hits \( X_i \) of player \( i \) has a binomial distribution with probability function \( f_i(x_i|p_i) = \binom{n_i}{x_i} p_i^{x_i}(1 - p_i)^{n_i-x_i} \) (independent of the number of hits (HITS) of the other 32 players) with known number of ATBATS \( n_i \) and with unknown hitting probability \( p_i \) (i=1,...,33). The natural conjugate prior for this exponential family is the beta density having the form \( \pi_\beta(p|\eta, \zeta) \propto p^{\eta-1}(1-p)^{\zeta-1} \) with positive real parameters \( \eta \) and \( \zeta \); we use the notation \( \mathcal{B}(\eta, \zeta) \) for this common
prior density. In accordance with the general model given at the beginning of this section, we assume the parameters \( p_1, \ldots, p_N \) are independently distributed with respective prior densities \( B(n_1, \zeta_1), \ldots, B(n_N, \zeta_N) \). The resulting Bayes ranking procedure for this model can be easily shown to rank \( p_\alpha \) above \( p_\beta \) iff

\[
\int_{p_\alpha > p_\beta} f_\alpha(x_\alpha | p_\alpha) \pi_B(p_\alpha | n_\alpha, \zeta_\alpha) f_\beta(x_\beta | p_\beta) \pi_B(p_\beta | n_\beta, \zeta_\beta) \; dp_\alpha \; dp_\beta > .5, \tag{2.4}
\]

(1 \leq \alpha \neq \beta \leq N). Table 1, given below, contains the number of ATBATS for 7 players in world series play between 1948 and 1955 and the hitting proportions of these players (expressed as a number out of 1000 in accordance with standard usage). Each player is ranked in ascending order in three different ways; we let \( \text{Rank}_{\text{MLE}} \) denote the players ranks (out of 33) induced by ranking players according to the maximum likelihood estimators of their respective hitting proportions, \( \text{Rank}_{\text{BR}} \) denote the players ranks (out of 33) given by the Bayes ranking procedure when the prior distribution of the hitting proportions is independently and identically distributed as \( B(.05, 1) \), and \( \text{Rank}_{\text{PME}} \) denote the players ranks (out of 33) given by ranking players according their respective posterior mean estimates when the prior distribution of the hitting proportions is independently and identically distributed as \( B(.05, 1) \). These 7 players are included in Table 1A of APPENDIX I as the 1st, 4th, 10th, 11th, 17th, 19th, and 22nd players respectively.

Table 1: BINOMIAL RANKING MODEL

<table>
<thead>
<tr>
<th>Known Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta = .05 )</td>
</tr>
<tr>
<td>( \zeta = 1 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>ATBATS</th>
<th>( \hat{p} )</th>
<th>\text{Rank}_{\text{MLE}}</th>
<th>\text{Rank}_{\text{BR}}</th>
<th>\text{Rank}_{\text{PME}}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bauer</td>
<td>94</td>
<td>202</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Byrne</td>
<td>7</td>
<td>285</td>
<td>24</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Edwards</td>
<td>2</td>
<td>500</td>
<td>33</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td>Furillo</td>
<td>82</td>
<td>256</td>
<td>17</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>Johnson</td>
<td>7</td>
<td>143</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Lopat</td>
<td>19</td>
<td>210</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>McDougal</td>
<td>99</td>
<td>222</td>
<td>11</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

In what follows some conclusions are drawn from Table 1. The Bayes ranking proce-
due is closer to the maximum likelihood procedure for hitters who tend to have smaller hitting averages and farther away otherwise; this is a consequence of the prior assumption that the hitting proportions will tend to be small which arises from choosing $\eta = .05$ and $\zeta = 1$. The Bayes ranking procedure is similar to both the adapted maximum likelihood and adapted Bayes procedures when the number of ATBATS is large; this follows, in view of the fact that the prior distributions of the parameters are identical, from the law of large numbers. The baseball players Edwards and Byrne provide examples of the divergence between the different procedures for moderately small number of ATBATS. The procedure which ranks players according to their posterior mean proportion of hits tends, in general, to be closer to the Bayes ranking procedure than the maximum likelihood-based ranking procedure even though, as in the case of Byrne, it can diverge substantially from it.

2.4 Bayes Ranking Procedures: Normal Model

Define approximately normal independent variables, $Y_i = \sqrt{n_i} \text{ARCSIN} \left( 2 \left( X_i/n_i \right) - 1 \right)$ and parameters $\theta_i = \text{ARCSIN} \left( 2p_i - 1 \right)$ (where ARCSIN refers to the arcsine function) $(i = 1, ..., N)$. Notice that ranking the p’s is equivalent to ranking the $\theta$’s. The parameters $\theta_1, ..., \theta_N$ correspond respectively in this model to the parameters $\psi_1, ..., \psi_N$, as defined above. The natural conjugate prior for this exponential family is the normal density having the form,

$$
\pi_N(\theta | \mu, \tau) \propto \exp \left\{ - (\theta - \mu)^2 / \left(2\tau^2 \right) \right\}
$$

with real parameter $\mu$ and positive parameter $\tau^2$; we use the notation $\mathcal{N}(\mu, \tau^2)$ for this density. It is assumed that the parameters $\theta_1, ..., \theta_N$ are independently distributed according to the respective densities $\mathcal{N}(\mu_1, \tau_1^2), ..., \mathcal{N}(\mu_N, \tau_N^2)$. In order to simplify notation we define the additional variables

$$
Y_i^* = \sqrt{n_i} Y_i + \frac{\mu_i}{\tau_i^2}, \quad (i = 1, ..., N)
$$

the Bayes estimators (posterior means)

$$
\hat{\theta}_i^{(B)} = \frac{\sqrt{n_i} Y_i + \frac{\mu_i}{\tau_i^2}}{n_i + \frac{1}{\tau_i^2}} \quad (2.5)
$$

of $\theta_i$, $(i = 1, ..., N)$ and the posterior variances

$$
\sigma_i^2(Y) = n_i + \frac{1}{\tau_i^2} \quad (i = 1, ..., N).
$$
The resulting Bayes ranking procedure for this (normal) model can be easily shown to rank $\theta_\alpha$ above $\theta_\beta$ iff $\theta_\alpha^{(B)} > \theta_\beta^{(B)}$ (1 $\leq \alpha \neq \beta \leq N$) since in this case the posterior mean and median of each of the $\theta$'s is the same; the transitivity of Bayes tournament procedures under this model is an immediate consequence of this fact. Algebra shows that $\theta_\alpha^{(B)} > \theta_\beta^{(B)}$ iff

$$
\left( n_\beta + \frac{1}{\tau_\beta^2} \right) Y_\alpha^* > \left( n_\alpha + \frac{1}{\tau_\alpha^2} \right) Y_\beta^*
$$

(2.6)

An analysis of equation (2.5) yields that if $\tau_1, ..., \tau_N$ all tend to infinity, this reduces to ranking $\theta_\alpha$ above $\theta_\beta$ iff

$$
\frac{Y_\alpha}{\sqrt{n_\alpha}} > \frac{Y_\beta}{\sqrt{n_\beta}}. \quad (1 \leq \alpha \neq \beta \leq N)
$$

We note that the quantity $\frac{Y_i}{\sqrt{n_i}}$ is equivalent, up to a constant of proportionality, to the quantity which maximizes the classical likelihood (regarded as a function of $\theta_1, ..., \theta_N$) and hence in this case the Bayes ranking procedure reduces to the adapted maximum likelihood ranking procedure. If, on the other hand, $\tau_1, ..., \tau_N$ all tend to 0, this reduces to ranking $\theta_\alpha$ above $\theta_\beta$ iff $\mu_\alpha > \mu_\beta$ (1 $\leq \alpha \neq \beta \leq N$); thus the Bayes ranking procedure reduces in this case to ranking on the basis of the mean hyperparameters $\mu_1, ..., \mu_N$. The Bayes ranking procedure corresponding to using positive finite values of $\tau_1, ..., \tau_N$ represents a compromise between these two extremes. Note finally that if the numbers $n_1, ..., n_N$ and prior variances $\tau_1^2, ..., \tau_N^2$ are such that the posterior variances $\sigma_1^2(Y), ..., \sigma_N^2(Y)$ are all equal, then the parameters $\theta_1, ..., \theta_N$ are stochastically ordered; Laird and Louis (1989) have shown that in this case the Bayes ranking, constructed above, is equivalent to a number of other Bayes ranking formulations; Laird and Louis note that in the contrary case where the posterior variances are not equal, the discrepancy these different ranking procedures depends on how much variation there is between the posterior variances. In the binomial model the posterior distributions of the parameters $p_1, ..., p_N$ are not usually stochastically ordered; hence the above result will rarely apply in this case.

In Table 2, given below, we give the number of HITS out of total ATBATS for 8 players listed in APPENDIX I, Table 1A as the 2nd, 6th, 7th, 9th, 20th, 21st, 23rd, and 30th players.
30th players. Each player is ranked in ascending order (of hitting ability) in two different ways; we let $\text{RANK}_{\text{MLE}}$ denote the ranking induced by a comparison of maximum likelihood estimates of players abilities and $\text{Rank}_B$ denote the players ranks given by the Bayes ranking procedure corresponding to the prior density $\mathcal{N}(-.52, 8.3)$. We note that the value $\mu = -.52$ has been selected for the location hyperparameter $\mu$ because it maximizes the "marginal likelihood" (see e.g., section 3 of this paper) while the value $\tau^2 = 8.3$ for the scale hyperparameter has been selected arbitrarily to reflect the prior belief that the $\theta$'s vary to this degree.

Table 2: NORMAL RANKING MODEL
KNOWN HYPERPARAMETERS

$\mu = - .52; \tau^2 = 8.3$

Ranking out of 33 Players Listed in Table 1A of APPENDIX I

<table>
<thead>
<tr>
<th>Name</th>
<th>ATBATS</th>
<th>HITS</th>
<th>$\hat{p}$</th>
<th>Rank$_{\text{MLE}}$</th>
<th>Rank$_B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berra</td>
<td>127</td>
<td>35</td>
<td>276</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td>Mantle</td>
<td>68</td>
<td>18</td>
<td>265</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>Martin</td>
<td>72</td>
<td>25</td>
<td>347</td>
<td>31</td>
<td>33</td>
</tr>
<tr>
<td>Rizzuto</td>
<td>118</td>
<td>27</td>
<td>228</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Dimaggio</td>
<td>54</td>
<td>12</td>
<td>222</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Collins</td>
<td>66</td>
<td>10</td>
<td>152</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Coleman</td>
<td>45</td>
<td>11</td>
<td>244</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>Mize</td>
<td>42</td>
<td>12</td>
<td>286</td>
<td>24</td>
<td>25</td>
</tr>
</tbody>
</table>

We note that the Bayes ranking procedure under the normal model seems to diverge from the adapted ranking $\text{Rank}_{\text{MLE}}$ more than under the binomial model; this arises in consequence of the assumed hyperparameters as well as from errors due to the use of the ARCSINE transformation, defined above.

2.5 Transitivity Properties of Bayes Tournament Procedures

In this section we demonstrate transitivity properties for the ranking procedures discussed in this paper. In the sequel, for a given parameter space $\Omega$ and real valued function $d$, the exponential family $\mathcal{F}(d, \Omega)$, defined by (2.2), will be said to have the $h$-property
iff for any independent random variables \( X, Y \) having respective densities \( \pi(\psi|t_x, s_x) \) and \( \pi(\psi|t_y, s_y) \) both of which belong to \( \mathcal{F}(d, \Omega) \) and for some function \( h \) and non-negative constant \( c \),

\[
P(X > Y|t_x, t_y, s_x, s_y) \geq 0.5 \quad \text{iff} \quad h(t_x, s_x) - h(t_y, s_y) \geq c.
\]

The exponential family \( \mathcal{F}(d, \Omega) \) will be said to have the transitivity property iff for any independent random variables \( X, Y, \) and \( Z \) each of which is distributed according to some distribution in \( \mathcal{F}(d, \Omega) \),

\[
P(X > Y) \geq (1/2); \quad \text{and} \quad P(Y > Z) \geq (1/2) \quad \rightarrow \quad P(X > Z) \geq (1/2)
\]

We can then show

**Lemma 2.1**

Under the general setup introduced at the beginning of this section, if the exponential family \( \mathcal{F}(d, \Omega) \) has (for some function \( h \)) the \( h \)-property then it has the transitivity property.

Proof: See APPENDIX II

In view of Lemma 2.1, a Bayes ranking procedure will be transitive if, for a suitable function \( h \), it has the \( h \)-property. This is used to demonstrate

**Theorem 2.1**

Under both the normal or binomial models illustrated above Bayes ranking procedures are transitive.

Proof: See APPENDIX II

Theorem 2.1 also leads to conditions under which, for any given hitting data and appropriate conjugate prior distributions, the adapted ranking by posterior means is different from the Bayes ranking under the Binomial model. We restrict ourself to the case involving only two hitters; analogous, although more technical conditions, can be given in
Corollary 2.1

For any independent binomial random variables $X_1, X_2$, there exists beta priors for which $P(p_1 > p_2 | X_1, X_2) \geq .5$ but $E(p_1 | X_1) < E(p_2 | X_2)$.

Proof: See APPENDIX II

3. POSTERIOR ROBUSTNESS OF RANKING PROCEDURES

In this section it is shown that Bayes ranking procedures, as constructed in section 2, are robust in the sense of being relatively insensitive to short hitting slumps or streaks. As was stated earlier, we model this by looking at the effect of 'prior' contamination (i.e., the assumption that the actual prior density for each parameter is slightly different than was claimed) on the ranking of the given set of baseball players. These robustness results are in sharp contrast to well-known results (see e.g., Berger (1985) and Hartigan (1984)) concerned with the sensitivity of simultaneous Bayes estimators to changes in the prior distribution of the parameters.

In the sequel, for Bayes ranking procedures $\delta^\nu$ and $\delta^\lambda$ corresponding respectively to the priors $\nu$ and $\lambda$, we adopt the measure of posterior regret (PR):

$$\text{PR}(\delta^\nu, \delta^\lambda | W) = \frac{L(\delta^\nu, \delta^\lambda)}{\binom{N}{2}} ,$$

with $L(\cdot, \cdot)$, as defined in (2.1), the ranking distance between the two procedures; the quantity $\binom{N}{2} = \binom{33}{2} = 528$ is the maximum possible ranking distance between procedures of this type. Thus, $\text{PR}(\cdot, \cdot)$ is the ranking distance between the two given Bayes ranking procedures with the distance scaled so as to be between 0 and 1.

In view of the above definition of posterior regret, an appropriate measure of posterior robustness of a Bayes ranking procedure is given as follows: An $\epsilon$-contamination family for $\pi_1, \ldots, \pi_N$ corresponds to the set $\Pi_\epsilon$ of joint prior densities $(\pi_1^{(q_1)}, \ldots, \pi_N^{(q_N)})$ with

$$\pi_i^{(q_i)} = (1 - \epsilon) + \epsilon q_i \quad (1 \leq i \leq N)$$

where $q$ is an arbitrary density. We use the notation $\delta^\pi$ (respectively $\delta^{\pi(q)}$) for the Bayes ranking procedure corresponding to the joint prior $\pi =: \prod_{i=1}^N \pi_i(\theta_i)$ (respectively for the
joint prior $\pi^{(q)} = \prod_{i=1}^{N} \pi_i^{(q_i)}(\theta_i)$. The maximal posterior regret (MPR) associated with the prior $\pi$ is

$$\text{MPR}(\pi, \epsilon) = \max \left\{ \text{PR}(\delta^\pi, \delta^{\pi^{(q)}}) : \pi^{(q)} \in \Pi_\epsilon \right\}$$

(3.2)

The function MPR defines an appropriate measure of robustness constrained to lie between 0 and 1 (inclusive). It is readily apparent that adapted ranking procedures are never robust in the above sense since, by choosing some (but not all) of the priors $q$ to put full mass on the value plus infinity, the corresponding posterior means will be ranked above the rest. Our next result shows, by contrast with this, that the Bayes ranking procedure constructed above is robust. In order to simplify notation and reduce technical complexity we show this result in case all the parameters $\theta_1, \ldots, \theta_N$ are independent and identically distributed according to a single prior $\pi = N(\mu, \tau^2)$ with known hyperparameters $\mu$ and $\tau^2$; the extension to the general case for both the normal and binomial models is analogous.

Below, we calculate maximal posterior regrets under the normal model when the prior distribution of the $\theta$’s are independent and identically distributed for different values of $\epsilon$. Preparatory to our main result we need some additional notation. We use the following simplified notation: for the posterior probability that $\theta_\alpha > \theta_\beta$ we use

$$P(\alpha, \beta) = P_{N(\mu, \tau^2)}(\theta_\alpha > \theta_\beta);$$

(3.3)

for the supremum of the joint density of $(Y_\alpha, Y_\beta)$ over the set $\theta_\alpha > \theta_\beta$ we use

$$\text{SUP}(\alpha, \beta) = \begin{cases} \frac{1}{\sqrt{2\pi}} \phi \left\{ Y_\beta - \sqrt{\frac{n_\beta}{n_\alpha}} Y_\alpha \right\} & \text{if } (Y_\alpha/\sqrt{n_\alpha}) \geq (Y_\beta/\sqrt{n_\beta}) \\ \frac{1}{\sqrt{2\pi}} \phi \left\{ Y_\alpha - \sqrt{\frac{n_\alpha}{n_\beta}} Y_\beta \right\} & \text{if } (Y_\alpha/\sqrt{n_\alpha}) < (Y_\beta/\sqrt{n_\beta}), n_\alpha > n_\beta \\ \frac{1}{\sqrt{2\pi}} \phi \left\{ Y_\alpha - \sqrt{\frac{n_\alpha}{n_\beta}} Y_\beta \right\} & \text{otherwise}; \end{cases}$$

(3.4)

for the marginal density of $(Y_\alpha, Y_\beta)$ we use

$$m(\alpha, \beta) = \phi \{(Y_\alpha - \mu)/\sqrt{1 + \tau^2}\} \phi \{(Y_\beta - \mu)/\sqrt{1 + \tau^2}\};$$

(3.5)

and put

$$Q(\alpha, \beta) = \frac{P(\alpha, \beta)\text{SUP}(\beta, \alpha)}{(1 - \epsilon)m(\alpha, \beta) + \epsilon \text{SUP}(\beta, \alpha)},$$

(3.6)
over the set $1 \leq \alpha \neq \beta \leq N$. We can then show, in the notation introduced by equations (3.3), (3.4), (3.5), and (3.6), that

**Theorem 3.1**

The maximal posterior regret $\text{MPR}\{\mathcal{N}(\mu, \tau^2), \epsilon\}$ associated with Bayes ranking procedures under the normal model satisfies the inequality

$$\text{MPR}\{\mathcal{N}(\mu, \tau^2), \epsilon\} \leq \sum_{I\{ - \epsilon Q(\alpha, \beta) \leq 0.5 - P(\alpha, \beta) \leq \epsilon Q(\beta, \alpha) \}} \left( \begin{array}{c} N \\alpha \neq \beta \end{array} \right),$$

(3.7)

the sum being over the set $(1 \leq \alpha, \beta \leq N; \alpha \neq \beta)$.

Proof: See APPENDIX II

Theorem 3.1 shows that for asymptotically small values of $\epsilon$ (i.e., in asymptotically small $\epsilon$-contamination neighborhoods) the maximal posterior regret is also asymptotically small. This fact is confirmed empirically by Table 3, given below, which reports the value of $\text{MPR}\{\mathcal{N}(\mu, \tau^2), \epsilon\}$ when $\mu = -0.52$, $\tau^2 = 8.3$, and $\epsilon$ takes on various values between .01 and 1.

**Table 4: MPR FOR THE NORMAL MODEL**

<table>
<thead>
<tr>
<th>$\theta_1, \ldots, \theta_N$</th>
<th>IID $\mathcal{N}(-0.52, 8.3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\epsilon$</td>
<td>MPR</td>
</tr>
<tr>
<td>.01</td>
<td>.031</td>
</tr>
<tr>
<td>.02</td>
<td>.031</td>
</tr>
<tr>
<td>.04</td>
<td>.156</td>
</tr>
<tr>
<td>.05</td>
<td>.188</td>
</tr>
<tr>
<td>.06</td>
<td>.188</td>
</tr>
<tr>
<td>.07</td>
<td>.188</td>
</tr>
<tr>
<td>.08</td>
<td>.188</td>
</tr>
<tr>
<td>.09</td>
<td>.250</td>
</tr>
<tr>
<td>.10</td>
<td>.281</td>
</tr>
</tbody>
</table>

Note that we can interpret the results of the table as saying that a contamination factor of $\epsilon = .09$ gives rise to a worst or least favorable ranking which is one quarter of the
ranking distance from the Bayes ranking in the normal model given above.

4. EMPIRICAL BAYES RANKING PROCEDURES

An Empirical Bayes approach to the problem of constructing ranking procedures involves using Bayes ranking procedures, corresponding to either the normal or binomial model studied in Section 2, when the hyperparameters have been estimated from the data. We first prove some general results. This is followed by constructing Empirical Bayes ranking procedures for the binomial model; numerical results are given showing how Empirical Bayes procedures rank some of the baseball players listed in TABLE 1A, APPENDIX I in this case. Discussion of Empirical Bayes ranking procedures under the normal model is omitted; the interested reader may consult Laird and Louis (1989) and Berger and Deeley (1988).

We return, for the moment, to the general model described at the beginning of Section 2. Let us adopt the notation \( \psi = (\psi_1, ..., \psi_N) \) for the vector of \( \psi \) parameters, \( \gamma = (\gamma_1, ..., \gamma_N) \) for the corresponding vector of hyperparameters, \( \pi(\psi|\gamma) \) for the joint prior distribution of the vector \( \psi \), and

\[
\text{MIN}(\alpha, \beta|\gamma_\alpha, \gamma_\beta) = (1/2) \min \{ P_{\pi(\psi|\gamma)}(\psi_\alpha > \psi_\beta|W), 1 - P_{\pi(\psi|\gamma)}(\psi_\alpha > \psi_\beta|W) \};
\]

(1 \leq \alpha \neq \beta \leq N). The hyperparametric vector \( \gamma \) is estimated so as to minimize the posterior Bayes risk given by

\[
\mathcal{R}(\gamma|W) = E[L(\psi, \delta^{\pi(\psi|\gamma)})]; \tag{4.1}
\]

the expectation on the right hand side of (4.1) denotes posterior expectation. The posterior Bayes risk, defined by (4.1) above, is

\[
\mathcal{R}(\gamma|W) = \sum_{1 \leq \alpha \neq \beta \leq N} \text{MIN}(\alpha, \beta|\gamma_\alpha, \gamma_\beta). \tag{4.2}
\]

The empirical bayes (vector) estimate \( \gamma^{EB} =: (\gamma_1^{EB}, ..., \gamma_N^{EB}) \) of \( \gamma = (\gamma_1, ..., \gamma_N) \) is chosen to have the property that

\[
\mathcal{R}(\gamma^{EB}|W) = \min \{ \mathcal{R}(\gamma|W); \quad \gamma \in \Gamma \}, \tag{4.3}
\]

14
where \( \Gamma \) refers to the space of permissible values of \( \gamma \). We can then show the following result.

**Theorem 4.1**

The Empirical Bayes estimate \( \gamma_{EB} \), defined above, has the property that it maximizes the quantity

\[
RCON(\gamma|W) = \sum (1/2) - P_{\pi(\psi|\gamma)}(\psi_\alpha > \psi_\beta|W); \tag{4.4}
\]

the sum on the right hand side of (4.4) is over the set \( 1 \leq \alpha \neq \beta \leq N \).

**Proof:** See APPENDIX II

Note that we use the notation "RCON" (standing for ranking confidence) for the quantity defined on the right hand side of (4.4) because this quantity constitutes a measure of how much confidence we have in the given Bayes ranking procedure; the size of "RCON" is a reflection of the amount of evidence present for believing that the given ranking is an accurate reflection of the world series hitting abilities of the players. Thus this theorem states that an Empirical Bayes ranking procedure should use those values of the hyperparameters which maximizes ones confidence in the Bayes ranking procedure.

**Proof:** See APPENDIX II

The marginal log-likelihood of the data corresponding to using the prior density \( \pi(\psi|\gamma) \) is given by,

\[
\mathcal{L}(\gamma|W) = \sum_{i=1}^{N} \log \left\{ \int f_i(W_i|\psi_i)\pi(\psi_i|\gamma_i) \, d\psi_i \right\}, \tag{4.5}
\]

(Empirical) Bayes estimation procedures frequently involve estimating the hyperparametric vector \( \gamma \) so as to maximize the marginal likelihood, given by the right hand side of (4.5), (see e.g., Efron and Morris (1975)); we use the notation \( \hat{\gamma} \) for these respective maximizing values, \( \delta^\gamma \) for the Bayes ranking procedure employing the estimate \( \hat{\gamma} \) and \( \text{Rank}_{\gamma} \) for the ranks of (baseball) players arising from using the procedure \( \delta^\gamma \).

**4.1 Empirical Bayes Ranking Procedures: Simplified Binomial Model**

It is assumed in this subsection, for the sake of simplicity, that the parameters \( p_1, \ldots, p_N \)
are independent and identically distributed according to a single beta distribution $B(\eta, \zeta)$. Analogous, although technically more difficult, results hold in the more general binomial model. Our aim is to show how the Empirical Bayes estimates of the parameters might be constructed in a particular case. It follows from Theorem 4.1 that under the binomial model the Empirical Bayes estimates of $\eta$ and $\zeta$ are those which maximize the confidence function.

$$RCON(\eta, \zeta|X) = \sum |(1/2) - \int \int_{p_\alpha > p_\beta} f_\alpha(X_\alpha|p_\alpha) f_\beta(X_\beta|p_\beta) \pi_B(p_\alpha|\eta, \zeta) \pi_B(p_\beta|\eta, \zeta) \, dp_\alpha \, dp_\beta|.$$  \hspace{1cm} (4.6)

**Theorem 4.2**

We have that

$$\lim_{\eta, \zeta \to \infty} \quad RCON(\eta, \zeta|X) = 0,$$  \hspace{1cm} (4.7)

the limit taken as $\eta$ and/or $\zeta$ tend, at whatever rate, to positive infinity.

**Proof:** See APPENDIX II.

It follows from Theorem 4.2 that the empirical Bayes estimates $\eta_{EB}$ and $\zeta_{EB}$, as defined above, are finite numbers. Our next result presents a theoretical characterization of these values.

We adopt the simplified notation

$$P(\alpha, \beta|\eta, \zeta) = P_{\pi_\beta}(p_\alpha > p_\beta|X) \quad (1 \leq \alpha \neq \beta \leq N)$$  \hspace{1cm} (4.8)

for the posterior probability that $p_\alpha$ is larger than $p_\beta$, $D_\eta$ (respectively $D_\zeta$, $D_{\eta, \zeta}$, etc.) for the partial derivative operator with respect to $\eta$ (respective $\zeta$, $\eta$ and $\zeta$, etc.), $D$ for the differential operator $D_{\eta, \eta}D_{\zeta, \zeta} - D_{\eta, \zeta}D_{\eta, \zeta}$, and $S(.5)$ for the set of $\alpha, \beta$'s $(1 \leq \alpha, \beta \leq N)$ for which $P(\alpha, \beta|\eta, \zeta) > .5$. We can then show that

**Theorem 4.3**

The Empirical Bayes estimators $\eta_{EB}$ and $\zeta_{EB}$ have the property that

$$\frac{\sum_{S(.5)} D_\eta P(\alpha, \beta|\eta, \zeta)}{\sum D_\eta P(\alpha, \beta|\eta, \zeta)} = .5$$  \hspace{1cm} (4.9)
\[
\frac{\sum_{s(5)} D_{\eta}P(\alpha, \beta | \eta, \zeta)}{\sum D_{\eta}P(\alpha, \beta | \eta, \zeta)} = .5, \tag{4.10}
\]
\[
\frac{\sum_{s(5)} D_p(\alpha, \beta | \eta, \zeta)}{\sum D_p(\alpha, \beta | \eta, \zeta)} < .5. \tag{4.11}
\]

when \( \eta = \eta_{EB} \) and \( \zeta = \zeta_{EB} \); all of the above sums unless otherwise specified are over the set, \( 1 \leq \alpha \neq \beta \leq N \).

Proof: See APPENDIX II

We note the quantities \( \eta \) and \( \zeta \) which satisfy equations (4.9) and (4.10) correspond to critical values of the posterior Bayes risk while equation (4.11) insures that the given critical values also constitute a minimum of the posterior Bayes risk function.

Table 4, below, gives the value of this posterior Bayes risk for the data of TABLE 1A, APPENDIX I when some different values of \( \eta, \zeta \) are used; the last column corresponds to the posterior regret \( PR(\delta_{\eta, \zeta}, \delta_{MLE}) \) associated with using the Bayes ranking procedure \( \delta_{\eta, \zeta} \) (with given values of \( \eta, \zeta \)) rather than the adapted maximum likelihood procedure \( \delta_{MLE} \) (arising from comparing batting averages of the different players) for the \( N = 33 \) players listed in TABLE 1A, APPENDIX I. Table 4 reveals the dependence of the Bayes ranking procedure on the values of the prior parameters \( \eta, \zeta \). It follows from Table 4 that the posterior Bayes risk is minimized when \( \eta = .1 \) and \( \zeta = 2.2 \); thus we obtain that \( \eta_{EB} = .1 \) and \( \zeta_{EB} = 2.2 \). The marginal log-likelihood of this data is maximized for the prior values \( \eta = 6 \) and \( \zeta = 15 \). Table 5, presented below, compares the ranks of baseball players for a portion of the data given in TABLE 1A, APPENDIX 1 when the Empirical Bayes procedure \( \delta_{EB} \), the adapted maximum likelihood ranking procedure \( \delta_{MLE} \), and the Bayes ranking procedure \( \delta_{\eta, \zeta} \) (called the marginal maximum likelihood ranking procedure) are used. The posterior regrets are listed at the bottom of the table. We use the notation \( \text{RANK}_{\text{EB}} \) for the ranks of players using the ranking procedure \( \delta_{\text{MLE}} \), and \( \text{RANK}_{\eta, \zeta} \) for the ranks of players using the ranking procedure \( \delta_{\eta, \zeta} \).
Table 4: Posterior Bayes Risk: Binomial Model

| $\eta$ | $\zeta$ | $\mathcal{R}(\eta, \zeta | X)$ | PR($\delta^{\eta, \zeta}, \delta^{\text{MLE}}$) |
|-------|--------|-------------------------------|---------------------------------|
| 0.1   | 0.60   | 69.602                        | .022                            |
| 0.1   | 1.40   | 66.718                        | .053                            |
| 0.1*  | 2.20** | 62.217                        | .098                            |
| 0.1   | 3.00   | 63.815                        | .121                            |
| 0.2   | 0.60   | 69.121                        | .009                            |
| 0.2   | 1.40   | 68.441                        | .040                            |
| 0.2   | 2.20   | 64.901                        | .080                            |
| 0.2   | 3.00   | 63.288                        | .115                            |
| 0.3   | 0.60   | 69.158                        | .008                            |
| 0.3   | 1.40   | 69.607                        | .032                            |
| 0.3   | 2.20   | 66.420                        | .064                            |
| 0.3   | 3.00   | 65.653                        | .104                            |

*: This is $\eta_{EB}$

**: This is $\zeta_{EB}$.

Table 5: Comparison of Empirical Bayes, Adapted MLE, and Marginal MLE Ranking Procedures

$\eta_{EB} = .1$; $\zeta_{EB} = 2.2$; $\bar{\eta} = 6$; $\bar{\zeta} = 15$.

<table>
<thead>
<tr>
<th>Name</th>
<th>ATBATS</th>
<th>HITS</th>
<th>$\hat{p}$</th>
<th>Rank$_{\bar{\theta}, \bar{\zeta}}$</th>
<th>Rank$_{EB}$</th>
<th>Rank$_{MLE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campanella</td>
<td>92</td>
<td>23</td>
<td>250</td>
<td>15</td>
<td>18</td>
<td>15</td>
</tr>
<tr>
<td>Hodges</td>
<td>84</td>
<td>19</td>
<td>226</td>
<td>9</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Jorgansen</td>
<td>11</td>
<td>2</td>
<td>182</td>
<td>12</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Olmo</td>
<td>11</td>
<td>3</td>
<td>273</td>
<td>20</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>Pafko</td>
<td>21</td>
<td>4</td>
<td>190</td>
<td>7</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Raschi</td>
<td>18</td>
<td>3</td>
<td>167</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Reese</td>
<td>97</td>
<td>29</td>
<td>298</td>
<td>26</td>
<td>28</td>
<td>25</td>
</tr>
<tr>
<td>Robinson</td>
<td>86</td>
<td>19</td>
<td>220</td>
<td>6</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Woodling</td>
<td>85</td>
<td>27</td>
<td>318</td>
<td>29</td>
<td>29</td>
<td>28</td>
</tr>
</tbody>
</table>

We have that

$$\text{PR}(\delta^{\text{EB}}, \delta^{\bar{\theta}, \bar{\zeta}}) = .160; \quad \text{PR}(\delta^{\bar{\theta}, \bar{\zeta}}, \delta^{\text{MLE}}) = .081.$$
References


### APPENDIX I

**TABLE 1A: PLAYERS HITS, ATBATS, AND HITTING AVERAGES IN WORLD SERIES PLAY 1948-1955**

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>TAB</th>
<th>HITS</th>
<th>( \hat{p} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bauer</td>
<td>94</td>
<td>19</td>
<td>202</td>
</tr>
<tr>
<td>2</td>
<td>Berra</td>
<td>127</td>
<td>35</td>
<td>276</td>
</tr>
<tr>
<td>3</td>
<td>Brown</td>
<td>38</td>
<td>15</td>
<td>395</td>
</tr>
<tr>
<td>4</td>
<td>Byrne</td>
<td>7</td>
<td>2</td>
<td>285</td>
</tr>
<tr>
<td>5</td>
<td>Campanella</td>
<td>92</td>
<td>23</td>
<td>250</td>
</tr>
<tr>
<td>6</td>
<td>Coleman</td>
<td>45</td>
<td>11</td>
<td>244</td>
</tr>
<tr>
<td>7</td>
<td>Collins</td>
<td>66</td>
<td>10</td>
<td>152</td>
</tr>
<tr>
<td>8</td>
<td>Cox</td>
<td>53</td>
<td>16</td>
<td>302</td>
</tr>
<tr>
<td>9</td>
<td>Dimaggio</td>
<td>54</td>
<td>12</td>
<td>222</td>
</tr>
<tr>
<td>10</td>
<td>Edwards</td>
<td>2</td>
<td>1</td>
<td>500</td>
</tr>
<tr>
<td>11</td>
<td>Furillo</td>
<td>82</td>
<td>21</td>
<td>256</td>
</tr>
<tr>
<td>12</td>
<td>Gilliam</td>
<td>51</td>
<td>15</td>
<td>294</td>
</tr>
<tr>
<td>13</td>
<td>Henrich</td>
<td>19</td>
<td>5</td>
<td>263</td>
</tr>
<tr>
<td>14</td>
<td>Hermanski</td>
<td>13</td>
<td>4</td>
<td>307</td>
</tr>
<tr>
<td>15</td>
<td>Hodges</td>
<td>84</td>
<td>19</td>
<td>226</td>
</tr>
<tr>
<td>16</td>
<td>Howard</td>
<td>26</td>
<td>5</td>
<td>192</td>
</tr>
<tr>
<td>17</td>
<td>Johnson</td>
<td>7</td>
<td>1</td>
<td>143</td>
</tr>
<tr>
<td>18</td>
<td>Jorgensen</td>
<td>11</td>
<td>2</td>
<td>182</td>
</tr>
<tr>
<td>19</td>
<td>Lopat</td>
<td>19</td>
<td>4</td>
<td>210</td>
</tr>
<tr>
<td>20</td>
<td>Mantle</td>
<td>68</td>
<td>18</td>
<td>265</td>
</tr>
<tr>
<td>21</td>
<td>Martin</td>
<td>72</td>
<td>25</td>
<td>347</td>
</tr>
<tr>
<td>22</td>
<td>McDougal</td>
<td>99</td>
<td>22</td>
<td>222</td>
</tr>
<tr>
<td>23</td>
<td>Mize</td>
<td>42</td>
<td>12</td>
<td>286</td>
</tr>
<tr>
<td>24</td>
<td>Noren</td>
<td>28</td>
<td>4</td>
<td>143</td>
</tr>
<tr>
<td>25</td>
<td>Olmo</td>
<td>11</td>
<td>3</td>
<td>273</td>
</tr>
<tr>
<td>26</td>
<td>Pafko</td>
<td>21</td>
<td>4</td>
<td>190</td>
</tr>
<tr>
<td>27</td>
<td>Raschi</td>
<td>18</td>
<td>3</td>
<td>167</td>
</tr>
<tr>
<td>28</td>
<td>Reese</td>
<td>97</td>
<td>29</td>
<td>298</td>
</tr>
<tr>
<td>29</td>
<td>Reynolds</td>
<td>22</td>
<td>6</td>
<td>273</td>
</tr>
<tr>
<td>30</td>
<td>Rizzuto</td>
<td>118</td>
<td>27</td>
<td>228</td>
</tr>
<tr>
<td>31</td>
<td>Robinson</td>
<td>86</td>
<td>19</td>
<td>220</td>
</tr>
<tr>
<td>32</td>
<td>Snyder</td>
<td>100</td>
<td>29</td>
<td>290</td>
</tr>
<tr>
<td>33</td>
<td>Woodling</td>
<td>85</td>
<td>27</td>
<td>318</td>
</tr>
</tbody>
</table>

*: The hitting averages \( \hat{p} \) are given here out of 1000 as is standard in baseball literature.
APPENDIX II

Proof of Lemma 2.1

Let $X,Y,Z$ be three independent random variables with distributions from the exponential family $\mathcal{F}(d,\Omega)$ with respective sets of parameters $\{s_x,t_x\},\{s_y,t_y\}$, and $\{s_z,t_z\}$. It follows by definition that if the exponential family $\mathcal{F}(d,\Omega)$ satisfies the h-condition and if

$$P(X > Y) \geq (1/2) \quad \text{and} \quad P(Y > Z) \geq (1/2),$$

then

$$h(t_x,s_x) - h(t_y,s_y) > c \quad \text{(A.1)}$$

$$h(t_y,s_y) - h(t_z,s_z) > c \quad \text{(A.2)}$$

It follows by adding the right and left hand sides of inequalities (A.1) and (A.2) together that

$$h(t_x,s_x) - h(t_z,s_z) > 2c \geq c \quad \text{(A.3)}$$

It follows by definition from (A.3) that

$$P(X > Z) \geq (1/2). \quad \text{(A.4)}$$

This completes the proof of the lemma.

Proof of Theorem 2.1

It follows from Lemma 2.1 that the theorem will follow once we have shown that the normal and binomial models both have the h-property (for different functions $h$). We show this first for the normal model: Let $X,Y$ be independently distributed according to the normal exponential family

$$\pi(x|t,s) \propto \exp\{tx + sd(x)\}$$

with $d(x) = -\frac{(x^2)}{2}$. Letting $h(t,s) = \frac{t}{s}$, it follows that

$$P(X > Y) = \Phi\left\{ \frac{h(t_x,s_x) - h(t_y,s_y)}{\sqrt{(1/s_x) + (1/s_y)}} \right\}; \quad \text{(A.5)}$$
with $\Phi$ denoting the standard normal distribution function. Clearly, the expression on the left hand side of the inequality sign in (A.5) is greater than or equal to (1/2) iff

$$h(t_z, s_z) - h(t_y, s_y) \geq 0$$

which demonstrates the main result in this case.

We now show the result for the binomial model. Let $F(d_B, \Omega_B)$ denote the exponential family with $d_B(u) = -\ln(1 + \exp(u))$ and $\Omega_B = \mathbb{R}$ and put $h_B(t, s) = t - s$. Let $X, Y$ be independent random variables each distributed according to a distribution in the family $F(d_B, \Omega_B)$ with respective sets of parameters $\{t_x, s_x\}$ and $\{t_y, s_y\}$. Define the functions $\Delta h_B = \Delta h_B(t_x, s_x, t_y, s_y)$ by $\Delta h_B = h_B(t_x, s_x) - h_B(t_y, s_y)$ and $g = g(t_x, s_x, t_y, s_y)$ by $g = P(X > Y)$. It follows from calculus that

$$\frac{\partial g}{\partial \Delta h_B} = \left( \frac{\partial g}{\partial t_x} - \frac{\partial g}{\partial t_y} \right) - \left( \frac{\partial g}{\partial s_x} - \frac{\partial g}{\partial s_y} \right)$$

(A.6)

We have by definition that

$$g = \frac{\int \int_{u>v} \exp \{t_x u + s_x d_B(u) + t_y v + s_y d_B(v)\} \, du \, dv}{\int \exp \{t_x u + s_x d_B(u) + t_y v + s_y d_B(v)\}}$$

(A.7)

In the sequel we use the notation COV to denote covariance. It follows from (A.7) and algebraic manipulation that

$$\frac{\partial g}{\partial t_x} - \frac{\partial g}{\partial t_y} \propto \text{COV}\{X - Y, I(X > Y)\}$$

(A.8)

and

$$\frac{\partial g}{\partial s_x} - \frac{\partial g}{\partial s_y} \propto \text{COV}\{d_B(X) - d_B(Y), I(X > Y)\}$$

(A.9)

up to the same positive constant of proportionality. Thus, by (A.6) and (A.9) we obtain that

$$\frac{\partial g}{\partial \Delta h_B} \propto \text{COV}\{X - Y, I(X > Y)\} - \text{COV}\{d(X) - d(Y), I(X > Y)\}$$

(A.10)

(up to a positive constant of proportionality). Now it is clear by definition of the function $d_B$ that $d_B(X) - d_B(Y)$ is a decreasing function of $X - Y$; hence

$$\text{COV}\{d_B(X) - d_B(Y), I(X > Y)\} < 0,$$

(A.11)
while $X - Y$ is an increasing function of $X - Y$; hence

$$\text{COV}\{X - Y; I(X > Y)\} > 0.$$ (A.12)

It follows from (A.10), (A.11), and (A.12) that

$$\frac{\partial g}{\partial \Delta h_B} > 0$$ (A.13)

The result (A.13) establishes that $g$ is an increasing function of $h_B$. On the other hand clearly $h_B$ is 0 when the random variables $X$ and $Y$ have identical distributions; it follows from these facts that

$$g \geq (1/2) \iff \Delta h_B \geq c$$ (A.14)

for some positive constant $c$. The result follows.

**Proof of Corollary 2.1**

Let $\pi_1, \pi_2$ be two prior densities with sets of parameters $\{t_1, s_1\}$ and $\{t_2, s_2\}$ and assume the number of hits $X_1, X_2$ are independent with binomial distributions having respective sample sizes $n_1$ and $n_2$. It follows that the posterior densities

$$\pi_1(\psi_1|X) \propto \exp \left( (t_1 + X_1)\psi_1 + (s_1 + n_1 - X_1) \right)$$ (A.15)

and

$$\pi_2(\psi_2|X) \propto \exp \left( (t_2 + X_2)\psi_2 + (s_2 + n_2 - X_2) \right)$$ (A.16)

Using the transformation $\psi = \ln \left\{ \frac{e^p}{1 - p} \right\}$, we can easily show that

$$E[\psi_i|X] = \frac{n_i - X_i + t_i}{n_i + s_i + t_i} \quad (i = 1, 2)$$ (A.17)

It follows from algebra that Lemma 2 states that, for the value of $c$ mentioned there,

$$P(\psi_1 > \psi_2|X) > (1/2) \iff 2(X_1 - X_2) - (n_1 - n_2) + (e_1 - e_2) + (f_1 - f_2) > c$$ (A.18)

Put $e_2(M) = e_1 + M$ and $f_2(M) = f_1 + M - (n_1 - n_2 - 2(X_1 - X_2)) - c$ with $M$ chosen large enough so that both $e_2$ and $f_2$ are positive; notice that for any such value of $M$ the inequality on the right hand side of the expression (A.18) holds. Define

$$D(\psi_1, \psi_2; e_1, e_2, f_1, f_2) = [E(\psi_1|X) - E(\psi_2|X)]$$
to be the difference of the posterior means of $\psi_1$ and $\psi_2$. It follows from (A.17) that

$$\lim_{M \to \infty} E(\psi_2|X) = 1.$$  

Hence,

$$\lim_{M \to \infty} D(\psi_1, \psi_2; e_1, e_2(M), f_1, f_2(M)) < 0. \quad (A.19)$$

It follows from the above that we can choose an $M$ large enough so that

$$D(\psi_1, \psi_2; e_1, e_2(M), f_1, f_2(M)) < 0 \quad (A.20)$$

while retaining the property that $P(\psi_1 > \psi_2|X) \geq (1/2)$. This completes the proof of the corollary.

**Proof of Theorem 3.1**

It follows from Theorem 4, section 4.7.4 of Berger (1985) that

$$\inf_{\pi \in \Pi_\epsilon} \{ (.5) - P_\pi(\theta_\alpha > \theta_\beta|Y) \} \geq -cQ(\beta, \alpha) \quad (1 \leq \alpha \neq \beta \leq N) \quad (A.21)$$

and

$$\sup_{\pi \in \Pi_\epsilon} \{ (.5) - P_\pi(\theta_\alpha > \theta_\beta|Y) \}, \quad (1 \leq \beta \leq N) \quad (A.22)$$

in the notation introduced prior to Theorem 3.1. It follows by definition of the Bayes ranking procedures introduced in section 2 that for any prior $\pi \in \Pi_\epsilon$, the Bayes ranking procedures $\delta^\pi$ and $\delta^N(\mu, \tau^2)$ ranking $\theta_\alpha$ and $\theta_\beta$ in the opposite order only if equations (A.21) and (A.22) both hold. The result (3.7) follows from this by definition of the loss function $L$. This completes the proof of the theorem.

**Proof of Theorem 4.1**

For any real numbers $a$ and $b$, we have the mathematical identity,

$$\min\{a, b\} = \frac{a + b - |a - b|}{2} \quad (A.23)$$

Setting $a = P_\pi(\psi_\alpha > \psi_\beta|W)$ and $b = 1 - a$, it follows from (A.23) that,

$$\text{MIN}(\alpha, \beta|\gamma_\alpha, \beta) = .5 - |1 - 2a| \quad (1 \leq \alpha \neq \beta \leq N) \quad (A.24)$$
Thus the values of $\gamma = (\gamma_1, \ldots, \gamma_N)$ which minimize the right hand side of (4.2) are the same ones which minimize the sum of the quantities on the right hand side of (A.24) or what is the same thing the same ones which maximize the right hand side of (4.4). This completes the proof of the theorem.

**Proof of Theorem 4.2**

Equation (4.9) will follow when we have shown that

$$
\lim_{\eta, \zeta \to \infty} \frac{\int \int_{\alpha > \beta} \frac{p_\alpha^{X_\alpha + \eta - 1} q_\alpha^{n_\alpha - X_\alpha + \zeta - 1}}{p_\beta^{X_\beta + \eta - 1} q_\beta^{n_\beta - X_\beta + \zeta - 1}} \frac{d\alpha}{d\beta}}{\int \int \frac{p_\alpha^{X_\alpha + \eta - 1} q_\alpha^{n_\alpha - X_\alpha + \zeta - 1}}{p_\beta^{X_\beta + \eta - 1} q_\beta^{n_\beta - X_\beta + \zeta - 1}} \frac{d\alpha}{d\beta}} = .5
$$

(A.25)

$(1 \leq \alpha \neq \beta \leq N)$ with $q_\alpha = 1 - p_\alpha$. Essentially, equation (A.25) can be written in the form

$$
\lim_{\eta, \zeta \to \infty} \frac{\int \int_{\alpha > \beta} \exp \left\{ \eta \left[ \ln(p_\alpha p_\beta) \right] + \zeta \left[ \ln(q_\alpha q_\beta) \right] \right\} g(p_\alpha, p_\beta) \frac{d\alpha}{d\beta}}{\int \int \exp \left\{ \eta \left[ \ln(p_\alpha p_\beta) \right] + \zeta \left[ \ln(q_\alpha q_\beta) \right] \right\} g(p_\alpha, p_\beta) \frac{d\alpha}{d\beta}} = .5,
$$

(A.26)

with $g(p_\alpha, p_\beta)$ some function of the $p$'s which does not involve $\eta$ and $\zeta$. The result follows from Laplace's Asymptotic expansion theorem (see e.g., De Brijn (1975)) applied to the left hand side of equation (A.26). This completes the proof of the theorem.

**Proof of Theorem 4.3**

Theorem 4.3 follows from Theorems 4.1 and 4.2 and standard results from analysis by differentiation of equation (4.8) with respect to both $\eta$ and $\zeta$. This completes the proof of the theorem.