

Bayesian Joint Estimation of Binomial Proportions

Marlos A. G. Viana *

University of Illinois at Chicago

November 29, 2004

Abstract

Testing the hypothesis H that $k > 1$ binomial parameters are equal and jointly estimating these parameters are related problems. A Bayesian argument can simultaneously answer these inference questions: to test the hypothesis H the posterior probability $\lambda = \lambda(H | x)$ of H given the experimental data x can be used; to estimate each binomial parameter, their Bayesian estimates under H and its complement \bar{H} are combined, with weights λ and $1 - \lambda$, respectively.

Key words: Joint estimation, Shrinkage, Bayesian hypothesis testing, Bayes factor

1 Introduction

Joint estimation is a frequently occurring problem, as pointed out by Novick, Lewis & Jackson (1973). For example, collaborative pilot studies are applications in which preliminary, small sample data from several participating centers are reviewed based on likely comparable experimental criteria. In this case, collateral information contained in observations from the other groups may effectively contribute to estimating the parameters of interest in each group. Table 1 shows the observed proportions of cases with eyes free of corneal scarring among patients participating in a five center ophthalmological clinical trial. These rates are thought to compromise the use of the cornea as a reflective surface which is used in keratometry and photokeratoscopy to measure the curvature of the cornea. This information is of value for detecting and measuring astigmatism in pre- and post-eye surgery patients, as well as an aid for fitting contact lenses. The observed proportions are (maximum likelihood) joint estimates of the corresponding binomial probabilities indicated by $\omega = (\omega_1, \dots, \omega_5)$. Of interest here is how to estimate ω taking into account the likelihood of the hypothesis of homogeneity $H : \omega_1 = \dots = \omega_5$. More specifically, the objective of this study is to suggest a Bayesian solution to the problem of jointly estimating $k > 1$ binomial parameters *in conjunction with* the adjacent problem of testing, also in the Bayesian sense, the hypothesis H of parametric homogeneity. Consequently, the proposed methodology unifies the solutions to both problems.

In fact, if $\lambda = \lambda(H | x)$ is a posterior probability of the hypothesis H , the proposed estimate of the i -th binomial parameter ω_i has the form

$$\hat{\omega}_i = \lambda \hat{\omega}_{i,H} + (1 - \lambda) \hat{\omega}_{i,\bar{H}} ,$$

*This in an electronic version of the original publication that appeared in the Journal of Educational Statistics (1991) Vol. 16 No.4, pp. 331-343, with minor editorial changes. The author is indebted to I.Olkin and C.A. de B. Pereira for critical suggestions regarding earlier drafts of this manuscript, to an associate editor and two referees for their helpful comments and to T. McMahon for providing the keratometry data. The research for this paper was supported in part by grant EY1729 to the UIC Eye Center from the National Institutes of Health, Bethesda, MD and a grant from the Illinois Eye Fund.

Table 1: Corneal scarring data.

center	scars absent	cases	joint MLE
1	7	17	0.412
2	6	7	0.857
3	11	15	0.733
4	17	32	0.531
5	24	37	0.649

where $\hat{\omega}_{i,H}$ and $\hat{\omega}_{i,\bar{H}}$ are Bayes estimates of ω_i under the hypothesis H and its complement \bar{H} , respectively, $i = 1, \dots, k$. The dependence between the likelihood of the homogeneity hypothesis H , the estimation problem, and the form of the suggested joint estimates is considered in the next Section, while the Bayesian evaluation of H and the actual expression for $\hat{\omega}_i$ are derived in Section 3.

Particular solutions to either one of the problems are, of course, well known. The classical test for the homogeneity hypothesis is typically a chi-squared test or Fisher's exact test e.g. D'Agostino, Chase & Belanger (1988), although even such common basic procedures are subjects of controversy, as pointed out recently by Little (1989). The homogeneity hypothesis and its effect in pooling event rates from clinical trials have been considered by, among others, Demets (1987) & Berlin, Laird, Sacks & Chalmers (1989). Large sample joint estimators of the binomial parameters have been discussed by several authors as well: Leonard (1976, 1977) has shown that the exchangeability argument leads to joint estimates of the logistic transformation of these parameters. Novick, Lewis & Jackson (1973) obtained large sample posterior joint modal estimates of the arcsin square root transform of each proportion based on Lindley's concept of collateral information (e.g. Lindley & Smith, 1972).

In general, these estimated values are obtained numerically, and correspond to the vector of posterior means, or modes, of the parameter of interest (e.g. Viana, 1980). In the empirical Bayes context, Efron & Morris (1973, 1975) have derived the corresponding joint estimates, a summary of which can be seen in Casella (1985) who clearly pointed out the relationship between the resulting estimates and the underlying hypothesis of parametric homogeneity to jointly estimating a vector of means.

Joint posterior means and variances of the population means for the one-parameter exponential families, based on equal-sized samples, have been obtained by Stroud (1988), with applications to the joint estimation of the logit transformation of the success probabilities.

The *form* of the estimates suggested in the paper is analogous to other existing joint estimates which, indirectly, also take into account the likelihood of the hypothesis of parametric homogeneity. The classical examples are briefly reviewed in Section 2, pointing to the analogy and to the natural dependence between the estimation problem and the likelihood of the homogeneity hypothesis. In Section 3 the estimation and testing procedures are presented in detail while real data applications are discussed in Section 4.

2 Joint Estimation and Parametric Homogeneity

The proposed joint estimates have the form $\hat{\omega}_i = \lambda\hat{\omega}_{i,H} + (1 - \lambda)\hat{\omega}_{i,\bar{H}}$, where $\hat{\omega}_{i,H}$ and $\hat{\omega}_{i,\bar{H}}$ are Bayes estimates of ω_i under the hypothesis H and its complement \bar{H} , respectively, and λ is the posterior probability of the hypothesis H . The purpose of the Section is to justify, by analogy, the form of the proposed estimates.

First, note that if it is known that H is true than the (maximum likelihood) estimate of ω_i is obtained by

pooling the total number of positive results $\sum x_i$, the total number of trials, $\sum n_i$ and determining the corresponding proportion of positive results. For example, from Table 1, $\hat{\omega}_i = \hat{\omega}_{i,H} = 65/108 = 0.601$ $i = 1, \dots, 5$. Alternatively, if it is known that H is false (without specifying its *form* e.g. $\omega_1 \leq \dots \leq \omega_k$), then the MLE of ω_i is simply the corresponding proportion of positive results, so that $\hat{\omega}_i = \hat{\omega}_{i,\bar{H}} = x_i/n_i$ $i = 1, \dots, 5$, following the example, $\hat{\omega}_1 = 7/17 = 0.412$. Otherwise, if there is uncertainty about H and λ reflects its likelihood, it is reasonable to weight these two estimates by λ and $1 - \lambda$, respectively, that is, $\hat{\omega}_i = \lambda \hat{\omega}_{i,H} + (1 - \lambda) \hat{\omega}_{i,\bar{H}}$. The argument can be extended to $m \geq 3$ alternative hypotheses H_1, \dots, H_m , with corresponding probabilities $\lambda_1, \dots, \lambda_m$, $\sum_{j=1}^m \lambda_j = 1$, in which case $\hat{\omega}_i = \sum_{j=1}^m \lambda_j \hat{\omega}_{i,H_j}$, $i = 1, \dots, k$. Note that if $m = 2$, then $H = H_1$ and $\bar{H} = H_2$.

Second, note that the following joint estimates shrink each exchangeable component to a center value. The amount of shrinkage being proportional to the likelihood of the homogeneity hypothesis. They justify, by analogy, the form of the suggested joint estimates.

Consider first the problem of estimating several means, as described in Lindley (1972). In the simplest case, suppose that the data is based on $X_i \sim N(\omega_i, 1)$, $i = 1, \dots, k$, with the additional assumption of $\omega_1, \dots, \omega_k$ being exchangeable. Then, De Finetti Theorem (e.g. Heath & Sudderth, 1976) leads to the conditional joint distribution $\omega_1, \dots, \omega_k \mid \xi \sim N(\xi, \sigma^2)$. When ξ is integrated out and the posterior expectation of ω_i is obtained, it follows that

$$\hat{\omega}_i = w\bar{x} + (1 - w)x_i$$

with $w = \frac{\tau}{1+\tau}$, and $\tau = \sigma^{-2}$. In this case, $\frac{w}{1-w} = \tau$ is the ratio of the precisions so that w is proportional to the degree of homogeneity of $\omega_1, \dots, \omega_k$. The homogeneity corresponds to the limit case $\tau = \infty$, in which case $w = 1$ and $\hat{\omega}_i = \bar{x}$. In the ANOVA extension of the above problem, $X_{ij} \sim N(\omega_i, \sigma_w^2)$, $i = 1, \dots, k$, $j = 1, \dots, n_i$, and the exchangeability assumption leads to the posterior mean of ω_i :

$$\hat{\omega}_i = w_i\bar{x} + (1 - w_i)\bar{x}_i$$

where $\frac{w_i}{1-w_i} = \frac{\sigma_w^2}{n_i\sigma_b^2}$ and, as usual, $\bar{x} = \frac{\sum w_i\bar{x}_i}{\sum w_i}$, with σ_w^2 and σ_b^2 being the *within* and *between* groups variances, respectively. Again, w_i reflects the (relative) strength of the homogeneity hypothesis, under which $\sigma_b = 0$, $w_i = 1$, and consequently, $\hat{\omega}_i = \bar{x}_i$.

In the empirical Bayes approach to this same problem, (Efron & Morris, 1973, 1975, Casella, 1985), suppose, in the simplest case, that $X_i \sim N(\omega_i, \sigma^2)$, $i = 1, \dots, k$ where σ is known. The Bayesian assumption $\omega_i \sim N(\mu, \tau^2)$ leads to the posterior (mean) estimate of the i -th component,

$$\hat{\omega}_i = \frac{\sigma^2}{\sigma^2 + \tau^2} \times \mu + \frac{\tau^2}{\sigma^2 + \tau^2} \times x_i, \quad i = 1, \dots, k. \quad (1)$$

The empirical Bayes method looks at the marginal distribution of X_i to infer about μ and τ . In fact, $X_i \sim N(\mu, \sigma^2 + \tau^2)$, so that $E[\bar{X}] = \mu$ and

$$E\left[\frac{(k-3)\sigma^2}{\sum (X_i - \bar{X})^2}\right] = \frac{\sigma^2}{\sigma^2 + \tau^2}.$$

Substituting the resulting estimates into (1) leads to

$$\bar{\omega}_{i,EB} = \left[\frac{(k-3)\sigma^2}{\sum (x_i - \bar{x})^2}\right] \times \bar{x} + \left[1 - \frac{(k-3)\sigma^2}{\sum (x_i - \bar{x})^2}\right] \times x_i,$$

which is equivalent (Casella, 1985) to

$$\hat{\omega}_{i,EB} = \frac{k-3}{k-1} T^{-1} \bar{x} + \left[1 - \frac{k-3}{k-1} T^{-1}\right] x_i.$$

Under the homogeneity assumption, however,

$$T = \frac{\sum (x_i - \bar{x})^2 / (k-1)}{\sigma^2} \sim \chi_{k-1}^2 / (k-1),$$

so that large values of T would lead to the rejection of $H : \omega_1 = \dots = \omega_k$. As pointed out by Casella, as T becomes large (and the data support \bar{H}), $\hat{\omega}_{i,E}$ reflects more of x_i and less of \bar{x} . Consequently, $\hat{\omega}_{i,E}$ reflects the estimates x_i , \bar{x} and the likelihood of the homogeneity hypothesis as well.

3 Joint Estimation of Binomial Parameters

The assessment of the homogeneity hypothesis is considered first. The likelihood $v(\omega)$ on the data $\{x_1, \dots, x_k\}$ is assumed to be

$$v(\omega) \propto \prod_{i=1}^k \omega_i^{x_i} (1 - \omega_i)^{y_i},$$

defined for $\omega \in \Omega = (0, 1)^k$, where x_i is the number of events of probability ω_i , among $n_i = x_i + y_i$ trials, $i = 1, \dots, k$. The posterior evaluation of H can be obtained following the Jeffreys method in which the posterior odds ratio O in favor of the hypothesis H is the product of the prior odds ratio $\frac{q}{1-q}$ and the ratio K of averaged likelihoods under the hypothesis H and its complement \bar{H} . The ratio K is referred to as Bayes factor (Jeffreys, 1961, Press, 1989, Lee, 1989). Under certain assumptions on v , the Bayes factor can be expressed as

$$K = \frac{\int_{\Omega} (v \circ \phi) du_1 \dots du_k}{\int_{\Omega} (v \circ \bar{\phi}) du_1 \dots du_k}, \quad (2)$$

where ϕ projects ω into the equiangular line along with one of its coordinates (restricting v to the subset corresponding to H) and $\bar{\phi}$ is the identity mapping (restricting v to \bar{H}). The posterior odds in favor of H is $O = Kq/(1-q)$ and the posterior probability of H is $\lambda(H | x) = O/(O+1)$.

The resulting posterior odds (from K) and conditional Bayes estimates (relative to uniform priors) have the form

$$\begin{aligned} O &= \left[\prod_{i=1}^k \frac{(n_i + 1)!}{x_i! y_i!} \right] \frac{x_i! y_i!}{(n_i + 1)!} \frac{q}{1-q}, \\ \hat{\omega}_{i,H} &= \frac{x_i + 1}{n_i + 2}, \\ \hat{\omega}_{i,\bar{H}} &= \frac{x_i + 1}{n_i + 2}, \quad i = 1, \dots, k, \end{aligned} \quad (3)$$

where the dot subscript indicates summation of the corresponding variable over k groups. Consequently, the i -th binomial proportion is estimated by

$$\hat{\omega}_i = \frac{O}{O+1} \frac{x_i + 1}{n_i + 2} + \frac{1}{O+1} \frac{x_i + 1}{n_i + 2}, \quad i = 1, \dots, k. \quad (4)$$

In summary, note that the form of the joint estimates described above was suggested by analogy, the expression for the posterior odds on H followed from Jeffreys method, and the conditional estimates are simply the usual posterior binomial estimates under uniform prior specification.

Alternatively, if the hypothesis H of homogeneity is represented by the set event $H : \omega \in H$ and a posterior probability given the experimental data $x = \{x_1, \dots, x_k\}$ can be derived and denoted by $\lambda(\cdot | x)$, then the

¹Except for a subset of P -probability 0 for all P absolutely continuous with respect to $du_1 \dots du_k$.

Table 2: Joint Estimates of Corneal Scarring Proportions.

center	x	n	joint mle	joint Bayes
1	7	17	0.412	0.562
2	6	7	0.857	0.638
3	11	15	0.733	0.622
4	17	32	0.531	0.585
5	24	37	0.649	0.609

hypothesis H can be evaluated by its posterior probability $\lambda(H | x)$. In fact, the formal dependence between the resulting estimates and the posterior probability of H has been studied by Viana (1990) who showed that each posterior marginal probability of $\lambda(H | x)$ leads to a Bayesian estimate (with respect to an appropriate prior probability measure λ) of the corresponding component ω_i which is a weighted average of a posterior estimate of ω_i under the hypothesis H and a posterior estimate under its complement \bar{H} , with corresponding weights $\lambda(H | x)$ and $\lambda(\bar{H} | x)$;

$$\hat{\omega}_i = \lambda(H | x)\hat{\omega}_{i,H} + \lambda(\bar{H} | x)\hat{\omega}_{i,\bar{H}}.$$

Moreover, it turns out that $\lambda(H | x) = O/(O + 1)$, $\hat{\omega}_{i,H}$ and $\hat{\omega}_{i,\bar{H}}$ are exactly as described by equation (3).

Example: Table 2 shows the resulting joint MLE and Bayesian estimates of proportions of cases not affected by corneal scars, based on data from 5 participating centers. The prior odds in favor of the homogeneity hypothesis was taken to be 1:1. The posterior odds, determined by (3) is

$$O = \left[\prod_{i=1}^5 \frac{(n_i + 1)!}{x_i!y_i!} \right] \frac{65!43!}{109!} = 3.7342,$$

so that, consequently, $\lambda(H | x) = 0.789$. The estimated proportion of cases not affected by scars at site number 1, under H and its complement, is, respectively:

$$\begin{aligned} \hat{\omega}_{1,H} &= \frac{66}{110} = 0.6, \\ \hat{\omega}_{1,\bar{H}} &= \frac{8}{19} = 0.421, \end{aligned}$$

so that the estimated proportion of cases free of corneal scars at site 1 is, from (4):

$$\hat{\omega}_1 = 0.789 \times 0.6 + 0.211 \times 0.421 = 0.562.$$

Performance of the Bayesian Joint Estimates The suggested estimates can be shown to perform comparably well with respect to the joint MLE and the EB joint estimates of Efron and Morris (1975). Other large sample joint estimates have been proposed by Leonard (1977) and Novick, Lewis & Jackson (1973). Tables 3 and 4 show the results of a Monte Carlo study with $k=3$ groups in which 3000 triples $\omega = (\omega_1, \omega_2, \omega_3)$ were uniformly generated in the unit cube $(0, 1)^3$, and assigned to satisfy the null hypothesis $H : \omega_1 = \omega_2 = \omega_3$ with probability 1/2. Corresponding to each ω generated, a triplet (x_1, x_2, x_3) of binomial counts was simulated, based on $n_1 = n_2 = n_3 = 20$ trials and respective Bernoulli probabilities $(\omega_1, \omega_2, \omega_3)$. The posterior probability $\lambda = \lambda(H | x)$ of hypothesis H , the corresponding joint estimates $\hat{\omega}_i$ based on (3), and the joint MLE $(x_1/n_1, x_2/n_2, x_3/n_3)$ were obtained. The observed Bayes quadratic loss $L_{Bayes} = \sum_{i=1}^3 (\hat{\omega}_i - \omega_i)^2$, the corresponding expected MLE quadratic loss $L_{mle} = \sum_{i=1}^3 \omega_i(1 - \omega_i)/n_i$ and their difference

$$\delta = L_{mle} - L_{Bayes}$$

were then obtained. Table 3 shows the estimated confidence limits for the expected value of $\lambda(H | x)$ as a function of D , the distance between ω and the equiangular line $\omega_1 = \omega_2 = \omega_3$. Table 4 shows the estimated

Table 3: Simulated $\lambda = \lambda(H | x)$, $\bar{\lambda}$, confidence limits (U,L) and $P(\lambda < .5)$ based on 3000 joint binomial estimates of ω ($k = 3, n_1 = n_2 = n_3 = 20$) as a function of the distance D between the simulated ω to the equiangular line.

D	count	λ	L 95% CI	U 95% CI	$P(\lambda < .5)$
D=0	1517	.8150519	.8058283	.8242755	.059
$0 < D \leq .1$	81	.7518091	.7118928	.7917255	.148
$.1 < D \leq .2$	190	.5826667	.5566041	.6087292	.321
$.2 < D \leq .3$	315	.3085924	.2883511	.3288337	.701
$.3 < D \leq .4$	328	.1241955	.1043594	.1440316	.899
$.4 < D \leq .5$	239	.0188578	-.0043800	.0420956	1.000
$.5 < D \leq .6$	186	.0014471	-.0248942	.0277884	1.000
$.6 < D \leq .7$	125	.0000644	-.0320676	.0321965	1.000
$.7 < D \leq .8$	19	.0000000	-.0824170	.0824170	1.000

Table 4: Simulated $\delta = L_{mle} - L_{Bayes}$, $\bar{\delta}$ and confidence limits based on 3000 joint binomial estimates of ω ($k = 3, n_1 = n_2 = n_3 = 20$) as a function of the distance D between the simulated ω to the equiangular line.

D	count	δ	lower 95% CI	upper 95% CI
D=0	1517	.0144331	.0133498	.0155164
$.0 < D \leq .1$	81	.0107410	.0060530	.0154291
$.1 < D \leq .2$	190	.0032317	.0001707	.0062927
$.2 < D \leq .3$	315	-.0007649	-.0031422	.0016124
$.3 < D \leq .4$	328	.0029249	.0005952	.0052546
$.4 < D \leq .5$	239	.0064893	.0037600	.0092185
$.5 < D \leq .6$	186	.0072698	.0041761	.0103635
$.6 < D \leq .7$	125	.0086408	.0048669	.0124146
$.7 < D \leq .8$	19	.0201875	.0105078	.0298671

confidence limits for the expected value of δ as a function of D . Note, from Table 3, that except for values of D close to H, the posterior probability correctly rejects the homogeneity hypothesis, as expected. In fact, if H is to be rejected when $\lambda(H | x) < 0.5$, the (average) power $P(\lambda < 0.5 | D)$ seems to be quite reasonable. Table 4 shows that the suggested joint estimate is at least as good as the joint MLE, in the quadratic loss sense. In fact, the data suggest that $\delta \geq 0$ for all values of D considered.

4 Examples of Applications and Discussion

The following examples illustrate the application of the suggested estimates, including the evaluation of the homogeneity hypothesis. A computer code in BASIC generates all numerical results. Note, however, that λ and the proposed joint estimates can be easily computed with the help of only a desk calculator and a table of log factorials (e.g. Novick & Jackson, 1974).

Iowa Test of Educational Development. Table 5 shows the Bayesian analysis of the data described by Novick, Lewis and Jackson (1973). It includes the proportions of students attaining the 70th percentile in the ITED among 11 Midwestern high schools, who would qualify for an honors course (joint MLE), and their joint Bayesian estimates: (1) under prior odds of 1:1 in favor of the homogeneity hypothesis H; (2) under prior odds of 1:10 on H; and (NLJ) Novick, Lewis & Jackson (1973) joint large sample modal estimates. The

corresponding posterior probabilities $\lambda(H | x)$ obtained from (3) were 0.991 (even odds) and 0.523 (.01:1). The later seems to closely correspond to the prior distribution indicating a wide variation of the proportions across high schools, adopted by Novick, Lewis and Jackson. The resulting estimates (2) and (NLJ) are fairly comparable in their relative magnitude. The fact that H remains more likely than \bar{H} suggests that the investigator's original beliefs concerning the variability of the proportions might have been incorrect, as pointed out by Novick.

CPP Reading Scale Table 6 shows the estimated proportions of students attaining the 50th percentile on the reading scale of the Career Planning Profile in 10 vocational-technical colleges, and the corresponding Bayesian joint estimates : (1) under even prior odds on H ; (2) under prior odds of 10:1 on H ; (NLJ) Novick, Lewis and Jackson joint large sample modal estimates. The corresponding posterior probabilities $\lambda(H | x)$ obtained from (3) were 0.033 (even odds) and 0.256 (10:1), pointing to the wider variation among the proportions. Note that the NLJ joint estimates corresponding to smaller sample sizes are shrunk closely to the overall mean. The suggested Bayes estimates show a constant shrinkage factor, namely $\lambda(H | x)$. Note, however, that $\lambda(H | x)$ reflects each component sample size. In general, the resulting estimates (2) and (NLJ) are fairly comparable in their relative magnitude.

Estimation of item difficulties Table 7 shows the estimated percentage of persons in a specific population that can be expected to respond correctly to the item. The items were proposed for inclusion in an examination for a university social studies course and expected to have difficulties as near 0.6 as possible. The posterior probability $\lambda(H | x)$ remains less than 0.1 even against prior odds of 1000:1 in favor of H , suggesting, as pointed out by NLJ's results as well, that the items did not reach the expected difficulties. Again, the resulting joint estimates are fairly comparable in their relative magnitude.

Casella's Baseball data Table 8 shows the batting averages (BA) of 7 major league baseball players after their first 45 at bats, reported by Casella (1985). The problem is to estimate their final batting averages (final BA). The Table includes Efron and Morris empirical Bayes estimates (e Bayes), the suggested Bayesian joint estimates, the final BA, and the resulting reduction in MSE when the corresponding player is removed from the set. The prior odds were taken to be 1:1. The resulting $\lambda(H | x)$ was 0.997. Consequently, the suggested estimates are virtually shrunk to the Bayes estimate of the BA under H , 0.287. The overall (7 players) MLE prediction error $\gamma_{MLE} = \sum(\hat{\omega}_i - \omega_i)^2$ with respect to final BA (ω) is 0.03296, while $\gamma_{EB} = 0.010784$ and $\gamma_{Bayes} = 0.005765$. As pointed out by Casella, the EB shows an improvement in prediction error of $0.010784 \div 0.03296 = 0.32$, meaning a 67% reduction in prediction error. Nevertheless, the suggested Bayes estimate shows an improvement in prediction error of $0.005765 \div 0.03296 = 0.17$, meaning an overall 82% reduction in prediction error, relative to the joint MLE. Note that when player 1 (quite similar early and final BAs) is removed from the set, the reduction in prediction error is as high as 91%, while when player 7 is removed (quite unstable early and final BAs), the reduction falls to 45%.

San Francisco Giants vs NY Mets, 1986 Table 9 shows the batting averages of SF Giants when playing against the New York Mets, cumulative to July of that year (early BA), the career Giants vs Mets BA, the corresponding Bayesian joint estimates and reduction in prediction error (MSE %) when that player is removed from the set. The prior odds on H were taken to be 1:1. The posterior odds were 15.9:1 and $\lambda(H | x) = 0.941$. The suggested estimates are quite close to the Bayes estimate under H . The overall (9 players) MLE prediction error γ_{MLE} with respect to career BA (ω) is 0.24745, while $\gamma_B = 0.01509$. Consequently, the suggested joint Bayesian estimates shows an improvement in prediction error of $0.01509 \div 0.24745 = 0.06$, meaning a 94% reduction in prediction error with respect to the joint MLE.

In summary, the Bayesian joint estimate suggested in the paper has the following distinct properties : (a) each component of the joint estimate is a weighted average of its Bayesian estimate under H and its complement \bar{H} , with weights equal to the corresponding posterior probabilities of H and its complement. Consequently, each binomial estimate has a constant shrinkage factor, namely the likelihood of the homogeneity hypothesis;

Table 5: Joint estimates of proportions of students attaining the 70th percentile in the ITED among 11 Midwestern high schools.

school	x	n	joint mle	Bayes (1)	Bayes (2)	Bayes (NLJ)
1	10	15	0.667	0.526	0.583	0.591
2	13	21	0.619	0.525	0.565	0.576
3	5	16	0.313	0.523	0.433	0.424
4	10	17	0.588	0.525	0.550	0.556
5	6	15	0.400	0.523	0.471	0.469
6	11	17	0.647	0.525	0.576	0.585
7	9	18	0.500	0.524	0.513	0.513
8	15	20	0.750	0.526	0.621	0.647
9	12	30	0.400	0.523	0.468	0.447
10	8	19	0.421	0.524	0.479	0.472
11	7	14	0.500	0.524	0.513	0.515

Table 6: Joint Estimates of Proportions Attaining the 50 Percentile on the Reading Scale of the CPP in 10 Vocational- Technical Colleges.

school	x	n	joint mle	Bayes (1)	Bayes (2)	Bayes (NLJ)
1	13	20	0.650	0.635	0.626	0.621
2	2	6	0.333	0.382	0.432	0.502
3	17	26	0.654	0.641	0.631	0.628
4	6	20	0.300	0.327	0.390	0.408
5	32	40	0.800	0.779	0.738	0.748
6	3	14	0.214	0.262	0.339	0.384
7	21	34	0.618	0.611	0.608	0.606
8	4	4	1.000	0.825	0.773	0.688
9	9	13	0.692	0.664	0.649	0.634
10	24	42	0.571	0.569	0.576	0.573

(b) the suggested method can be applied without restriction to the number k of proportions, and without restriction to the corresponding sample sizes n_1, \dots, n_k . The estimates do not require numerical methods and, in fact, can be obtained with the help of only a simple calculator and a table of log factorials; (c) the posterior odds derived from the Bayes factor (2) can be obtained under *more informative* prior specification for ω under H , e.g. a given univariate Beta density $Beta(a, a')$, and under its complement, e.g. a product $\prod_{i=1}^k Beta(a_i, a'_i)$ of univariate Beta densities; (d) the alternative hypothesis may be decomposed into component hypotheses such as $H_1 : \omega_1 > \omega_2 = \dots = \omega_k, \dots, H_k : \omega_k > \omega_1 = \dots = \omega_{k-1}$, in which case each binomial component is estimated as a weighted average of its estimates under H_1, H_2, \dots, H_k with weights $\lambda(H_1 | x), \dots, \lambda(H_k | x)$. Finally, note that the argument suggested in the paper may be extended to multinomial proportions (e.g. 2x2 tables) and other univariate natural exponential distributions.

Table 7: Joint Estimates of Item Difficulties for Six Social Studies Items.

item	x	n	joint mle	Bayes	Bayes (NLJ)
1	54	57	0.947	0.932	0.913
2	22	57	0.386	0.390	0.438
3	30	57	0.526	0.525	0.555
4	48	57	0.842	0.831	0.817
5	44	57	0.772	0.763	0.757
6	35	57	0.614	0.610	0.627

Table 8: Batting averages of 7 major league baseball players after their first 45 at bats; empirical Bayes (e Bayes), Bayes and MSE reduction relative to final BA when corresponding player is removed from the set.

player	H	AB	BA	e Bayes BA	Bayes	final BA	MSE %
1	18	45	0.400	0.341	0.287	0.346	91
2	16	45	0.356	0.321	0.287	0.279	75
3	14	45	0.311	0.300	0.287	0.276	82
4	13	45	0.289	0.289	0.287	0.266	84
5	11	45	0.244	0.266	0.287	0.271	84
6	10	45	0.222	0.255	0.287	0.266	84
7	8	45	0.178	0.230	0.287	0.318	45

Table 9: 1986 Batting averages of San Francisco Giants vs NY Mets.

player	H	AB	early BA	Bayes BA	career AB	MSE %
Brenly	4	30	0.133	0.290	0.267	83
Brown	6	21	0.286	0.299	0.237	92
Davis	14	30	0.467	0.309	0.250	98
Gladden	10	30	0.333	0.302	0.249	93
Leonard	11	28	0.393	0.305	0.280	96
Maldonado	1	11	0.091	0.290	0.292	88
Spilman	1	2	0.500	0.311	0.267	93
Uribe	7	28	0.250	0.297	0.291	89
Youngblood	0	2	0.000	0.296	0.253	91

References

- Berlin, J. A., Laird, N. M., Sacks, H. S. & Chalmers, T. C. (1989), 'A comparison of statistical methods for combining event rates from clinical trials', *Statistics in Medicine* **8**, 141–151.
- Casella, G. (1985), 'An introduction to empirical Bayes data analysis', *The American Statistician* **39**(2), 38–97.
- D'Agostino, R., Chase, W. & Belanger, A. (1988), 'The appropriateness of some common procedures for testing the equality of two independent binomial populations', *The American Statistician* **42**(3), 198–202.
- Demets, D. L. (1987), 'Methods for combining randomized clinical trials: Strengths and limitations', *Statistics in Medicine* **6**, 341–348.
- Efron, B. & Morris, C. (1973), 'Stein's estimation rule and its competitors', *Journal of the American Statistical Association* **68**, 117–130.
- Efron, B. & Morris, C. (1975), 'Data analysis using Stein's estimators and its generalizations', *Journal of American Statistical Association* **70**, 311–319.
- Heath, D. & Sudderth, W. (1976), 'De Finetti's theorem on exchangeable variables', *The American Statistician* **30**(4), 188–191.
- Jeffreys, H. (1961), *Theory of Probability*, 3rd edn, Clarendon Press, Oxford.
- Lee, P. M. (1989), *Bayesian Statistics: An Introduction*, Oxford University Press, New York.
- Leonard, T. (1977), 'Bayesian simultaneous estimation for several multinomial distributions', *Comm. Statistics A* **6**, 619–630.
- Lindley, D. V. (1972), *Bayesian Statistics, a Review*, SIAM, Philadelphia.
- Novick, M., Lewis, M. & Jackson, P. (1973), 'The estimation of proportions in m groups', *Psychometrika* **38**(1), 19–46.
- Novick, M. R. & Jackson, P. (1974), *Statistical Methods for Educational and Psychological Research*, McGraw-Hill, New York.
- Press, S. J. (1989), *Bayesian Statistics*, John Wiley, New York.
- Stroud, T. (1991), 'Hierarchical Bayes predictive means and variances with applications to sample survey inference', *Communications in Statistics* (20), 13–36.
- Viana, M. A. G. (1980), 'Statistical methods for summarizing independent correlational results', *Journal of Educational Statistics* **5**(1), 83–104.
- Viana, M. A. G. (1994), 'Bayesian joint estimates and the hypothesis of parametric homogeneity', *Technical Report*.

Marlos A.G. Viana
College of Medicine
University of Illinois at Chicago m/c 648
1855 West Taylor Street
Chicago, Illinois 60612