International Journal of Statistics and Probability; Vol. 6, No. 5; September 2017
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Insurance Mortality Rates, Performance Indicators, and Possibly Monotonic Population Proportions

John S. J. Hsu[1] & Thomas Leonard[2]

[1] Department of Statistics and Applied Probability, University of California, Santa Barbara, California, USA

[2] Hopetown Crescent, Edinburgh, Scotland, UK

Correspondence: John Hsu, Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA. E-mail: hsu@pstat.ucsb.edu

Received: June 8, 2017 Accepted: July 8, 2017 Online Published: July 25, 2017

doi:10.5539/ijsp.v6n5p29 URL: https://doi.org/10.5539/ijsp.v6n5p29

Abstract

Two applications are described of a probability model that can express uncertainty regarding a pre-specified monotonicity hypothesis for binomial proportions. The model also yields a random effects overdispersion formulation where the population proportions definitely satisfy a monotonicity specification. One application concerns an insurance data set recording mortalities of clients from ages 35 to 64. Two new actuarial graduation procedures are developed. The other application derives from a Veterans' administration hospital quality monitor and concerns the failure to return rates for psychiatric patients attending substance abuse clinics. While smoothed performance indicators are proposed, measures of their extra-binomial variation highlight problems experienced by evidence-based approaches when the data are uncontrolled.

Keywords: binomial, Pascal distribution, random effects, Bayesian probability model, over-dispersion, actuarial graduation, force of mortality, substance abuse, quality monitoring, performance indicators

1. Introduction

When investigating m population proportions $\theta_1, \theta_2, \ldots, \theta_m$ it is sometimes appropriate to consider monotonicity constraints

$$\theta_1 \leq \theta_2 \leq \ldots \leq \theta_m \quad , \tag{1}$$

based upon prior reasoning. The θ_i might for example represent the true success rates for a treatment, at consecutive time periods, where the success rates are thought to be non-decreasing in time. They may alternatively denote the success rates for m multi-centered trials, where the centers have been ordered according to preliminary performance indicators. More generally, we may have an isotonic regression situation where the θ_i are thought to possess the same ordering as an increasing covariate, for example, dose level.

A major theme of this paper lies in the argument that, while there may be some prior justification for the monotonicity constraints (1), the previous information may be insufficient to assume that (1) definitely holds. In section 3, hierarchical assumptions are introduced which, under binomial sampling assumptions, relax the investigator's prior belief in (1), thus permitting the observed data to refute (1). The probability model described may alternatively be interpreted as a random effects model for beta-binomial observations. The embedded extra-binomial variation then yields potentially quite different conclusions regarding the proposed monotonicity of the population proportions.

2. Two Data Sets

2.1 The Veterans' Administration Hospital Quality Monitor Data

We analyze part of a data set modeled by West and Aguilar (1997), Aguilar and West (1998), West et al. (1998), and Burgess et al. (2000), using Bayesian multiple time series. The subsample considered here provides information from the years 1992 and 1993 for $m = 159$ hospitals in the Veterans' Administration (VA) system. The 1993 data provide our dependent variables, and the 1992 data are used to calculate a set of explanatory variables.

Let y_i denote the number of individuals who failed to return for an outpatient visit within 30 days of discharge during 1993 out of the total number of annual discharges at the ith hospital, for $i = 1, 2, \ldots, m$. Then $p_i = y_i/n_i$ can be regarded as a performance indicator or measure of (lack of) quality for the ith hospital. The sample sizes range from 5 to 1142 with an average of $\bar{n} = 324.7$. Let x_i denote the corresponding proportion for the year 1992. For our first analysis, we attach our indices after reordering the hospitals according to increasing values $x_1 < x_2 < \cdots < x_m$. The rank ordering of the performance indicators for 1992 is thus taken into account when considering the rank ordering for 1993. Assumptions of

monotonic increasing population proportions for 1993, under binomial or beta-binomial assumptions for the y_i, will be investigated in section 6.

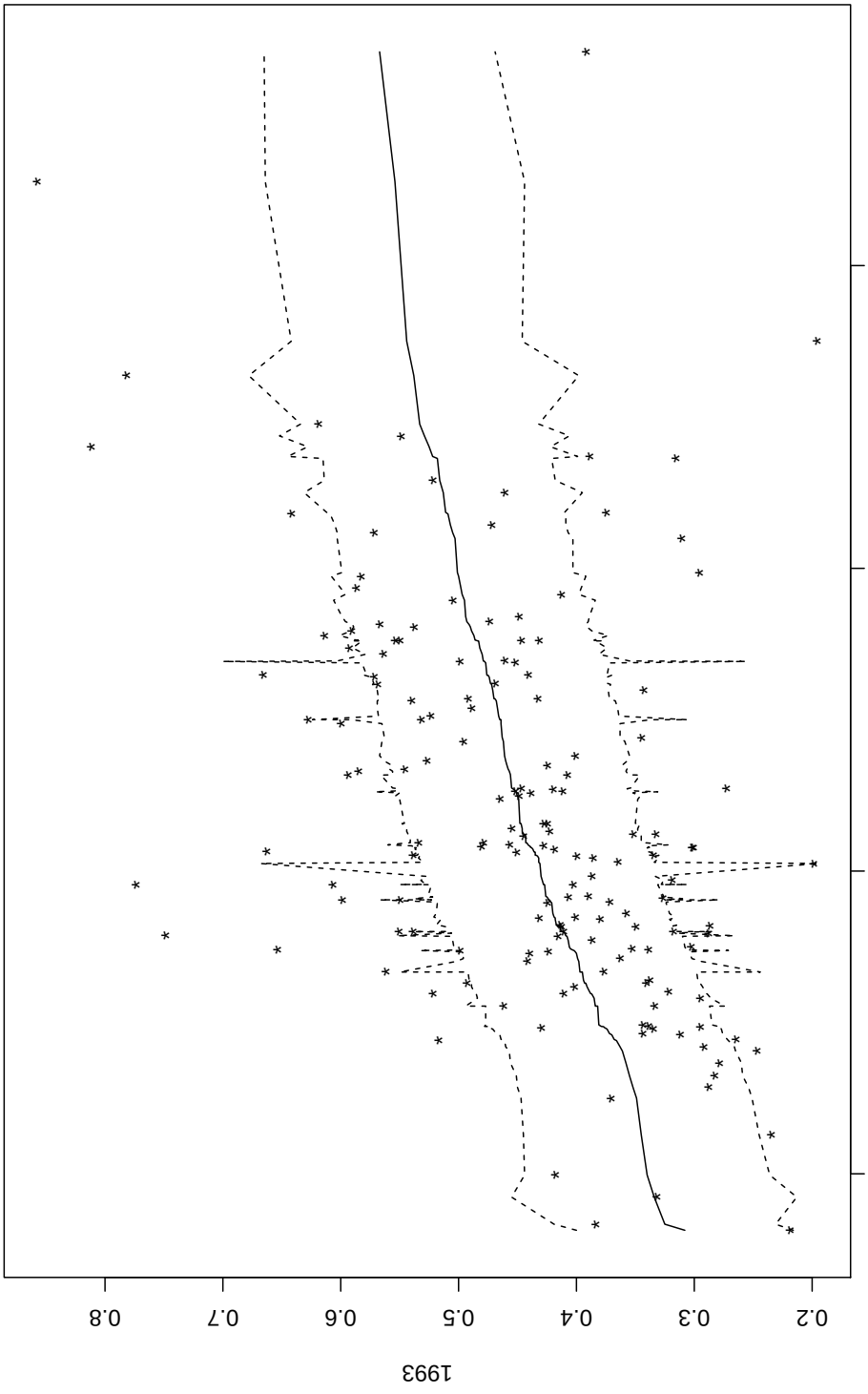

Figure 1: Raw Performance Indicators for 1992 and 1993

The association between the raw performance indicators is described by the entries to the scatterplot in Figure 1, which plot the p_i against the x_i. There is some overall increasing trend, but with considerable random scatter. The hospitals' raw performance indicators for 1992 do not provide good predictions of the performances for 1993. The solid plot describes a piecewise linear isotonic regression, as defined in sections 4.3 and 6, and justified under beta-binomial sampling assumptions. The abscissa of this plot provide smoothed performance indicators for 1993 which are consistent with

the rank ordering for 1992. The dotted plots add or subtract estimated standard errors of the 1993 sample proportions, which account for substantial extra-binomial variation. The magnitudes of the estimated standard errors provide guidance regarding the usefulness of the fitted performance indicators, for predictive rather than descriptive purposes. Further discussion is provided in section 6.

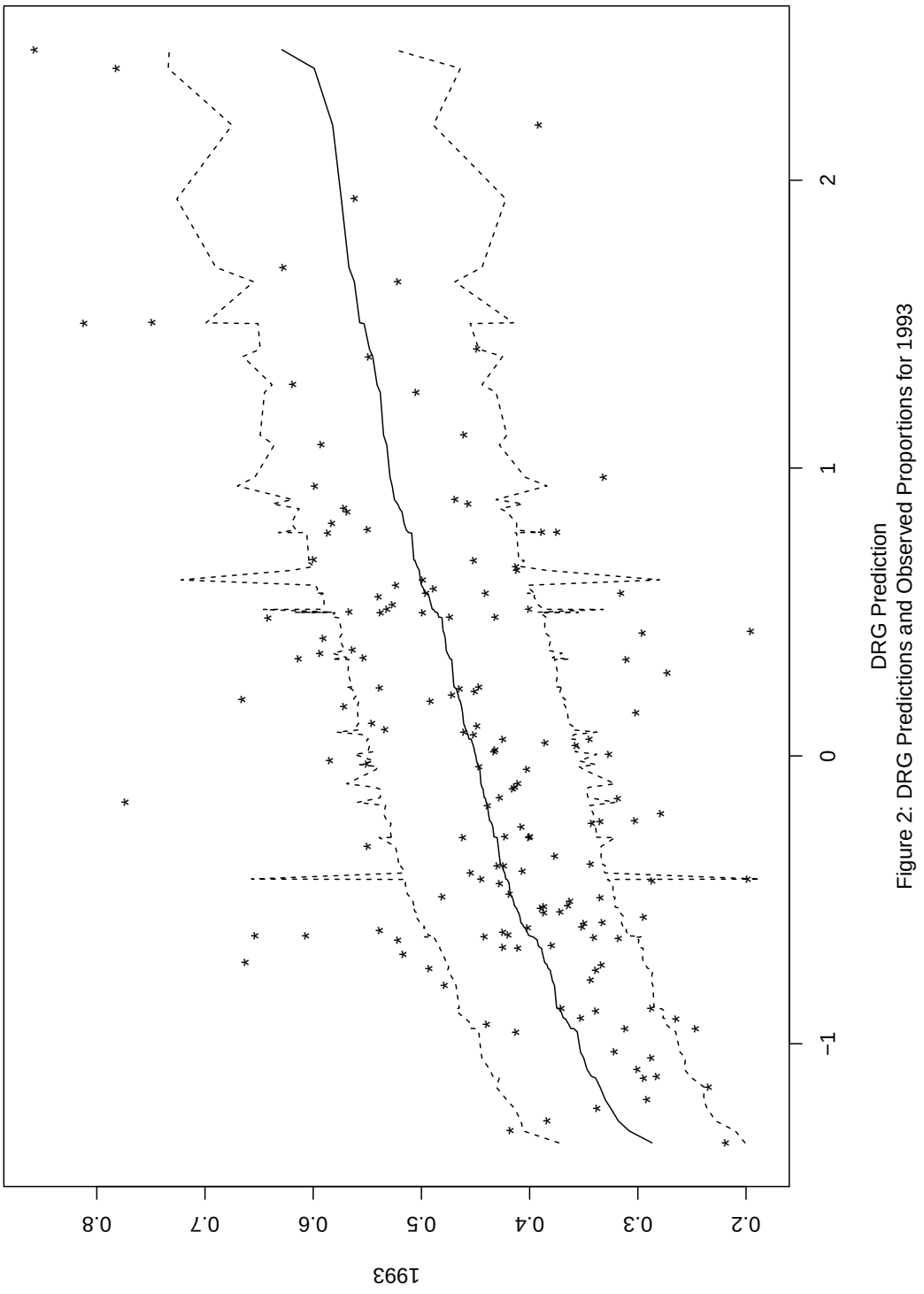

Figure 2: DRG Predictions and Observed Proportions for 1993

The preceding explanatory variables may be replaced by the VA's diagnostic related group (DRG) predictions that, for each hospital in each year, are supposed to provide predictions of the corresponding p_i. The DRG predictions for 1993 do not depend upon the sample proportions for years prior to 1993. In Figure 2, the p_i are plotted against the DRG predictions.

The performances of these predictions and the previous raw performance indicators are comparable. As the labeling of the x-axis of Figure 2 is quite compressed, when compared with Figure 1, the fitted isotonic graph, while similar in shape, represents a much steeper regression. The estimated standard errors of the corresponding sample proportions are however comparable.

In other analyses, not reported here, the explanatory variables were replaced by equally weighted or unequally weighted combinations of appropriately normalized proportions for 1992 and DRG predictions for 1993. Quite surprisingly, none of these combinations yielded substantive modifications to the shape of the isotonic regression graph, and the estimated standard errors were at best only marginally reduced. The inclusion of multiplicative interaction terms failed to improve the predictive performance.

2.2 Actuarial Graduation

The data in the second, third and fourth columns of Table 1 were collected and analyzed by Broffitt (1988) and reconsidered by Carlin (1992), Liu (2000), and Yang and Schwarz (2005). For $i = 1, 2, \ldots, 30$, the count y_i denotes the number of deaths out of n_i male clients of age a_i, for premium paying policies issued by an insurance company, with face amounts between \$10,000 and \$24,900. The previous authors assume in some cases the truth of the monotonicity hypothesis

$$H_0 : \theta_1 \leq \theta_2 \leq \ldots \leq \theta_m$$

for the corresponding underlying mortality rates, where $m = 30$. We furthermore take the y_i to be realizations of random variables Y_i satisfying the first stage assumptions specified in section 3, that is, the Y_i are independent given the θ_i, and possess binomial distributions with respective cell probabilities θ_i and sample sizes n_i.

Table 1. Mortality Rates Analysis

i	a_i	y_i	n_i	p_i	θ_i^*	sd(θ_i)	ξ_i^*	sd(ξ_i)
1	35	3	1172	0.256	0.115	0.068	0.086	0.201
2	36	1	2127	0.047	0.075	0.041	0.113	0.224
3	37	3	2744	0.109	0.112	0.046	0.138	0.239
4	38	2	2766	0.072	0.110	0.046	0.162	0.251
5	39	2	2463	0.081	0.130	0.051	0.187	0.261
6	40	4	2368	0.169	0.182	0.060	0.216	0.269
7	41	4	2310	0.173	0.201	0.063	0.245	0.277
8	42	7	2307	0.303	0.269	0.075	0.276	0.287
9	43	5	2060	0.243	0.264	0.073	0.303	0.295
10	44	2	1917	0.104	0.237	0.076	0.332	0.303
11	45	8	1931	0.414	0.366	0.090	0.370	0.311
12	46	13	1747	0.744	0.493	0.127	0.404	0.318
13	47	8	1580	0.506	0.433	0.103	0.430	0.325
14	48	2	1580	0.127	0.336	0.098	0.455	0.332
15	49	7	1468	0.477	0.463	0.106	0.488	0.340
16	50	4	1516	0.264	0.424	0.104	0.521	0.347
17	51	7	1372	0.510	0.525	0.115	0.562	0.354
18	52	4	1343	0.298	0.499	0.119	0.605	0.360
19	53	4	1304	0.307	0.546	0.129	0.664	0.366
20	54	11	1233	0.892	0.769	0.152	0.756	0.371
21	55	11	1205	0.913	0.840	0.158	0.843	0.377
22	56	13	1114	1.167	0.970	0.176	0.934	0.382
23	57	12	1048	1.145	1.028	0.181	1.020	0.390
24	58	12	1155	1.039	1.073	0.183	1.110	0.398
25	59	19	1019	1.865	1.345	0.230	1.219	0.406
26	60	12	945	1.270	1.279	0.211	1.307	0.418
27	61	16	853	1.876	1.485	0.244	1.414	0.433
28	62	12	750	1.600	1.511	0.250	1.519	0.454
29	63	6	693	1.866	1.485	0.283	1.645	0.483
30	64	10	594	1.684	1.890	0.369	1.959	0.572

Note, All entries to the last five columns have been multiplied by 100 and are therefore expressed in terms of percentages rather than proportions

In section 3, a prior distribution will be described for the θ_i, when they are not constrained, that expresses uncertainty in the monotonicity hypothesis (1) and permits the data to assist in the measurement of the posterior uncertainty in H_0. The corresponding posterior means and standard deviations of the θ_i are described in the sixth and seventh columns of Table 1. While substantially smoothing the raw mortality rates in the fifth column, some of the posterior means deviate quite noticeably from the monotonicity hypothesis. The magnitudes of the posterior standard deviations of the θ_i nevertheless suggest that the data are reasonably consistent with H_0 under the binomial sampling model.

As the data are uncontrolled there is no particular reason, apart from simplicity, to assume the preceding product binomial sampling model. The probabilistic assumptions of section 3 can alternatively be taken to represent a sampling model incorporating overdispersion, where the mortality rates or population proportions possess a random effects distribution and definitely satisfy a monotonicity specification paralleling (1). The posterior means and standard deviations of the mortality rates under this overdispersion model are given in the eighth and ninth columns of Table 1. The posterior means are more disperse than those under the product binomial sampling model and the posterior standard deviations are noticeably larger. As mortality rates are generally thought to increase with age, the random effects overdispersion interpretation is perhaps more appealing.

3. A Hierarchical Model

A four stage probability model with the following first two stages is employed:

Stage 1: Observations $Y_1, Y_2, \ldots, Y_m$ are independent and binomially distributed, given $\theta_1, \theta_2, \ldots, \theta_m$, with $Y_i|\theta_i \sim \text{BIN}(\theta_i, n_i)$, for $i = 1, 2, \ldots, m$.

Stage 2: The θ_i are independent and beta distributed, given an unknown parameter γ and respective conditional means ξ_i where, with the standard parameterization, $\theta_i|\gamma, \xi_i \sim \text{Beta}\{\gamma\xi_i, \gamma(1-\xi_i)\}$, for $i = 1, 2, \ldots, m$.

With the further assumption that the unknown ξ_i satisfy the monotonicity specification

$$\xi_1 \leq \xi_2 \leq \cdots \leq \xi_m \quad , \tag{2}$$

the preceding two stages can be interpreted in either of the following two ways:

(A) Let Stage 1 represent the sampling distribution of the Y_i and Stage 2 describe the first stage of a hierarchical prior distribution for the population proportions θ_i (further stages for γ and the conditional means ξ_i will be added below). In this case Stage 2 represents uncertainty in the belief that the monotonicity hypothesis (1) holds for the θ_i, thus extending an idea introduced by O'Hagan and Leonard (1976) in a single parameter normal situation. For given ξ_i and γ, the parameter θ_i can be said to possess a beta distribution with mean ξ_i and sample size γ, where this (prior) sample size measures the degree of belief in (1). As $\gamma \to \infty$ the monotonicity constraints are completely specified for the θ_i. A small value of γ represents substantial uncertainty in this hypothesis. Our formulation does not however require the specification of a definite value for γ, since the current data will typically provide considerable information regarding γ.

(B) The two stages may alternatively be combined. Unconditionally on θ_i, Y_i possesses a beta-binomial distribution, labeled by its parameters ξ_i and γ, and sample size n_i. The probability mass function of Y_i, given ξ_i and γ, is

$$p(Y_i = y_i|\xi_i, \gamma) = {}^{n_i}C_{y_i} l_i^*(\xi_i, \gamma) \quad ,$$

for $y_i = 0, 1, \ldots, n_i$, with ${}^{n_i}C_{y_i} = n_i!/y_i!(n_i - y_i)!$ and

$$l_i^*(\xi_i, \gamma) = \frac{B\{\gamma\xi_i + y_i, \gamma(1-\xi_i) + n_i - y_i\}}{B\{\gamma\xi_i, \gamma(1-\xi_i)\}} \quad , \tag{3}$$

where $B(a_0, a_1) = \Gamma(a_0 + a_1)/\Gamma(a_0)\Gamma(a_1)$ is the complete beta function with arguments a_0 and a_1. With the ξ_i now denoting our population proportions, we have a conditionally independent beta-binomial sampling model, within which the monotonicity specification in (2) is definitely satisfied as a modeling assumption. The plausibility of this specification may of course be further investigated.

In either case, the conditional distributions of the θ_i, given γ, the ξ_i and the observed values y_i of the Y_i are, for $i = 1, 2, \ldots, m$, independently beta with respective (posterior) sample sizes $n_i + \gamma$ and means

$$\theta_i^* = \rho_i p_i + (1 - \rho_i)\xi_i \quad , \tag{4}$$

where $p_i = y_i/n_i$ and

$$\rho_i = \rho_i(\gamma) = \frac{n_i}{n_i + \gamma} \quad . \tag{5}$$

In case (A), equation (4) describes the conditional posterior mean of θ_i. The θ_i^* compromise between the ξ_i satisfying the monotonicity specification (2), and the p_i, which can be taken to represent a general alternative hypothesis. Any data-based estimate of the average shrinkage proportion

$$\bar{\rho} = m^{-1} \sum_{i=1}^{m} \rho_i(\gamma) \tag{6}$$

can be interpreted as an overall measure, on a unit scale, of the evidence against the monotonicity hypothesis (1), and in favor of a general alternative hypothesis. The weighted modifications $\tilde{\rho} = \sum n_i \rho_i / N$ and $\hat{\rho} = \sum (n_i + \gamma)\rho_i / \sum (n_i + \gamma) = N/(N + \gamma)$, with $N = \sum n_i$ are more sensitive to values of the larger n_i.

Under the beta-binomial interpretation (B), the $P_i = Y_i/n_i$ are unbiased estimators of the ξ_i with respective variances $n_i^{-1} D_i \xi_i (1 - \xi_i)$, where $D_i = (n_i + \gamma)/(1 + \gamma)$ is the ith over-dispersion factor. These estimators do not however take account of (2). Moreover, not all of the $m + 1$ parameters γ and $\xi_1, \xi_2, \ldots, \xi_m$ are identifiable from the data, as there are just m observations. We consequently extend our conditionally independent beta-binomial model, by introducing the following random effects assumption:

Stage 3: Given $b_0 = \lambda\eta$ and $b_1 = \lambda(1 - \eta)$, the ξ_i possess the probability structure of the increasing order statistics based upon a random sample of size m from a Beta(b_0, b_1) distribution, that is, a beta distribution with mean η and sample size λ.

Our random effects beta-binomial sampling model for case (B) possesses just three parameters γ, λ, and η. When m is moderate to large, it is therefore possible to draw sensible proper Bayes inferences regarding these three identifiable parameters, and also for the ξ_i. Posterior estimates for γ and the ξ_i can thereby be imputed for the parameters of the preceding conditionally independent beta-binomial model. For computational convenience, we initially take the distribution of the parameters γ and λ in the prior assessment to be discrete. The prior distribution for the three parameters of our random effects model is selected as follows:

Stage 4: γ, λ, and η are independent, and $\eta \sim$ Beta(d_0, d_1). The distribution of γ assigns probabilities $\pi_1, \pi_2, \ldots, \pi_k$ to the points $g_1, g_2, \ldots, g_k$, and the distribution of λ assigns probabilities $\delta_1, \delta_2, \ldots, \delta_l$ to the points $h_1, h_2, \ldots, h_l$.

The assumption of prior independence of γ and λ can be relaxed by taking these parameters to possess a general discrete joint distribution on a $k \times l$ dimensional grid and practical choices of the prior parameters will be discussed in section 5. In a special case it will just be necessary to choose prior estimates n_0 and λ_0 for γ and λ, and, with $d_0 = d_1 = 1$, to then consider the sensitivity of the posterior inferences to the choices of n_0, λ_0, k, and l. Baseline values for n_0 and λ_0 will be recommended. Large values for k and l will yield close approximations to inferences under an interesting thick-tailed continuous prior distribution, which is effectively assumed.

In case (A), Stages 2, 3, and 4 provide a hierarchical prior distribution for the θ_i. Stage 3 permits input from the data regarding the values of the Stage 2 parameters ξ_i. Stage 4 facilitates input from the data regarding the value of γ, and the Stage 3 parameters $b_0 = \lambda\eta$ and $b_1 = \lambda(1 - \eta)$. Related hierarchical models for binomial probabilities, without the constraints in (2), provide alternatives to the binomial logit/normal prior or normal random effects developments by Leonard (1972, 1976), Warn et al. (2002), and many others.

4. Posterior Considerations

4.1 Posterior Inferences

In case (A) of section 3 the marginal posterior distribution of θ_i averages a beta distribution with sample size $n_i + \gamma$ and mean θ_i^* satisfying (4), with respect to the unconditional posterior distribution of γ and the ξ_i. All posterior quantities of interest for both cases (A) and (B) may be calculated, subject to a minor approximation, via standard Metropolis algorithm/MCMC procedures. Please see Appendix 2 for details. Unconditional posterior densities can be computed along with the means and standard deviations reported in the current paper.

For illustrative purposes only, note that the posterior distribution of the ξ_i, given γ, λ, and η, may be roughly approximated by taking the ξ_i to possess independent beta distributions, with respective sample sizes $D_i^{-1} n_i + \lambda$ and means

$$\xi_i^* = \frac{D_i^{-1} n_i p_i + \lambda\eta}{D_i^{-1} n_i + \lambda} \quad , \tag{7}$$

where $D_i = (n_i + \gamma)/(1 + \gamma)$, but then constraining these distributions to the region defined by (2). The expressions in (7) constrain the p_i towards a common unknown value η. The posterior means of the ξ_i are furthermore substantially influenced by the constraints in (2). As well as taking (2) into account, the unconditional posterior inferences create a partial pooling process which roughly speaking has the effect of flattening the ξ_i towards a pooled estimate for η.

When $d_0 = d_1 = 1$, η is estimated by a slightly adjusted center of location of the p_i. For example, the first posterior analysis of section 6, leading to the isotonic regression graph in Figure 1, yielded a posterior mean of 0.439 for η. This compares with the overall sample proportion $p^* = 0.425$, and the average sample proportion $\bar{p} = 0.444$, and accounts, via the shrinkages of the ξ_i, for a flattening of the isotonic regression graph. Pooled information from across the hospitals is thus incorporated. When judging the plausibility of a monotonic relationship, via the residual analysis of sections 6, it is important to realize that our regression graph meaningfully flattens steeper monotonic graphs which may better fit the data.

4.2 Two Useful Approximations and a Parameter of Interest

In Appendix 1, an approximation to the conditional distribution of the ξ_i, given the θ_i, γ, η, and λ, under Stages 2 and 3 of our probability model is justified unless γ, $b_0 = \lambda\eta$, or $b_1 = \lambda(1-\eta)$ is small. The approximation constrains m independent beta distributions to the region (2). These distributions may, for $i = 1, 2, \ldots, m$, be described as follows:

$$\xi_i|\theta_i, \gamma, \eta, \lambda \sim \text{Beta}\{\tilde{\lambda}\tilde{\xi}_i, \tilde{\lambda}(1-\tilde{\xi}_i)\} \quad , \tag{8}$$

where

$$\tilde{\xi}_i = \zeta\theta_i + (1-\zeta)\eta \quad , \tag{9}$$

and

$$\tilde{\lambda} = \gamma + \lambda + 1 \quad ,$$

with

$$\zeta = \frac{\gamma + 1}{\gamma + \lambda + 1} \quad . \tag{10}$$

This development highlights ζ in (10) as an interesting bounded function of γ and λ. As ζ approaches zero, the $\tilde{\xi}_i$ in (9) approach the common unknown value η. While the shrinkage proportions ρ_i in (5) relate to shrinkages of the θ_i towards the ordered ξ_i, the proportion ζ controls the shrinkages of the $\tilde{\xi}_i$ towards a common value η. Our preceding approximate conditional distribution for the ξ_i provides a key ingredient of the posterior computational procedures described in Appendix 2, and will be made more exact by acceptance sampling. The exact joint distribution of the ξ_i, given the θ_i, γ, η, and λ, initially takes the ξ_i to be independent, with respective densities

$$\tilde{\pi}(\xi_i) \propto \frac{\xi_i^{\eta\lambda-1}(1-\xi_i)^{\eta(1-\lambda)-1}\theta_i^{\gamma\xi_i}(1-\theta_i)^{\gamma(1-\xi_i)}}{B\{\gamma\xi_i, \gamma(1-\xi_i)\}} \quad , \tag{11}$$

for $0 < \xi_i < 1$ and $i = 1, 2, \ldots, m$, but then constrains the joint distribution of the ξ_i to the region (2). The acceptance sampling methodology refers to (11) without simulating from the corresponding exact distribution. In Appendix 2, the approximation

$$\eta|\boldsymbol{\xi}, \lambda, \mathbf{y} \sim \text{Beta}\{m(\lambda+1)\bar{\xi} + d_0, m(\lambda+1)(1-\bar{\xi}) + d_1\} \tag{12}$$

to the conditional posterior (or prior) distribution of η, given the ξ_i and λ, is also motivated, with $\bar{\xi}$ denoting the average ξ_i. The beta distribution in (12) possesses sample size $m(\lambda+1) + d_0 + d_1$, and mean

$$\tilde{\eta} = \frac{m(\lambda+1)\bar{\xi}}{m(\lambda+1) + d_0 + d_1} \quad , \tag{13}$$

which is close to $\bar{\xi}$ whenever $m(\lambda+1)$ is large compared with $d_0 + d_1$. The approximation in (12) may be contrasted with the exact conditional density

$$\pi(\eta|\boldsymbol{\xi}, \lambda, \mathbf{y}) \propto \pi(\eta)\tilde{l}(\eta, \lambda|\boldsymbol{\xi}) \quad , \tag{14}$$

for $0 < \eta < 1$, where $\pi(\eta)$ is a beta density with parameters d_0 and d_1, and

$$\tilde{l}(\eta, \lambda|\boldsymbol{\xi}) = \frac{\prod_{i=1}^{m}\xi_i^{\lambda\eta}(1-\xi_i)^{\lambda(1-\eta)}}{[B\{\lambda\eta, \lambda(1-\eta)\}]^m} \quad . \tag{15}$$

When justifying (12) and (15), it is important to note that the information provided about η and λ by fixed ordered values of the ξ_i is the same as when regarding the ξ_i as an unordered random sample from a beta distribution with mean η and sample size λ. This information is unaffected by knowledge of the data.

4.3 Regression Situations

The methodology underlying the isotonic regression examples of section 2.1 is now discussed. Consider case (B) of section 3, where each Y_i is taken to possess a beta-binomial distribution, conditional on parameters ξ_i and γ. Suppose that each Y_i and corresponding population proportion ξ_i is associated with a pre-specified value x_i of a covariate, where

$$x_1 \leq x_2 \leq \cdots \leq x_m \quad . \tag{16}$$

Assume that the ordering in (2) of the ξ_i is consistent with the ordering (16) of the x_i. A monotonic increasing regression of the ξ_i upon the x_i is therefore assumed. In situations where two or more of the x_i are equal, the ordering of the corresponding ξ_i should be based upon prior specification. Modifications to our procedure, which set two or more of the ξ_i equal, would alternatively be available. The posterior means of the ξ_i under our general analysis may be plotted against the x_i and connected by straight lines. If two or more of the x_i are equal, then the corresponding posterior means may be weighted according to the corresponding sample sizes. The recommended graph provides our estimated isotonic regression of the ξ_i upon the x_i. This semi-parametric approach provides an alternative to parametric procedures, see for example, Leonard and Novick (1986) and Lee and Nelder (1996), which replace stages 3 and 4 of our probability model, and the monotonicity assumption (2) by the specification of a functional form for the regression of the ξ_i upon the x_i. The precise modeling of this specification might sometimes present practical difficulties.

Our semi-parametric approach is also relevant to case (A) of section 3. If the posterior deviations of the θ_i from the ξ_i, are small, then the preceding estimated isotonic regression of the ξ_i upon the x_i can be used to meaningfully describe a fitted regression of the θ_i upon the x_i. Otherwise it is more important to report posterior inferences for the unconstrained θ_i. This contrasts with previous isotonic regression procedures for binomial data, for example, Barlow et al. (1972).

While our approach takes into account the ordering of the x_i, the specific values of the x_i are largely ignored in the posterior analysis, though they are re-introduced when plotting the regression of the ξ_i upon the x_i. Many isotonic regression procedures (e.g., Barlow et al. pp. 38 - 40) similarly trade information regarding the x_i for simplicity in the modeling procedure. Numerous possible adjustments to our method could however be considered. For example, when the regression of the ξ_i upon the x_i is thought to follow a segment of a concave function, (2) can be replaced by a decreasing slope specification. Information regarding the x_i can also be incorporated by generalizing Stage 2 of our probability model, by an assumption that $\theta_i|\gamma, \xi_i \sim \text{Beta}\{a_i\gamma\xi_i, a_i\gamma(1-\xi_i)\}$. The a_i adjust the sample size γ and may be specified subjectively as functions of several adjacent x_i.

5. Practical Prior Choices

The broad prior assumptions at Stage 4 of our probability model permit a wide spectrum of representation of prior beliefs, depending upon the information or views possessed by the statistician analyzing the data. However, in some practical situations, information external to the current data set may be sparse. In these circumstances, pragmatic choices should be made. For example, the values $d_0 = d_1 = 1$ lead to a uniform distribution for η on the unit interval. We will also assume that, for some specified n_0, the parameter

$$\rho_0 = n_0/(n_0 + \gamma)$$

is a priori uniformly distributed over the equally spaced grid of points $i/(k+1)$ for $i = 1, 2, \ldots, k$. Then the Stage 4 distribution for γ assigns equal prior probabilities $\pi_i = 1/k$ to the unequally spaced points

$$g_i = n_0(k - i + 1)/i \qquad (i = 1, 2, \ldots, k) \quad .$$

Since $E(\rho_0) = 1/2$, n_0 provides a prior estimate for γ, which is more sensible than the prior mean of γ. As k gets large, the distribution of ρ_0 approaches a continuous uniform distribution on the unit interval. In this limiting case γ possesses a Cauchy-tail prior density $\pi(\gamma) = n_0/(n_0 + \gamma)^2$, for $0 < \gamma < \infty$. No prior mean for γ exists in the limiting case owing to the extremely thick right tail of the prior distribution. The Cauchy-tail density contrasts with the log-Cauchy prior density assumed by Crook and Good (1982) for a multinomial smoothing parameter. In the current situation, the limiting conditional posterior density of γ given the ξ_i is

$$\pi(\gamma|\mathbf{y}, \boldsymbol{\xi}) \propto \pi(\gamma) \prod_{i=1}^{m} l_i^*(\xi_i, \gamma) \quad , \tag{17}$$

for $0 < \gamma < \infty$, where the contributions l_i^* to the product on the right hand side are defined in (3). Each ξ_i^* converges to unity as $\gamma \to \infty$, for any fixed ξ_i and y_i. Therefore the upper right tail of (17) invariably behaves like the upper right tail of $\pi(\gamma)$, for large values of γ.

Quite interestingly, if an improperly unfinitely uniform distribution with density $\pi(\gamma) \propto 1$, for $0 < \gamma < \infty$, is instead assumed for γ, then the density in (17) will never represent a proper distribution, thus invalidating the entire analysis. The Cauchy-tail prior density more appropriately controls the right tail of (17). This specification nevertheless represents quite sparse prior information regarding γ.

The parameter ρ_0 plays a somewhat similar role to the ρ_i satisfying (5) and (6), and can be interpreted as a shrinkage proportion relating to a hypothetical binomial experiment with sample size n_0. Under a beta prior distribution for θ_i with sample size γ and mean ξ_i, the posterior mean of θ_i, given only the hypothetical sample proportion p_0, is the weighted average compromise $\theta_i^o = \rho_0 p_0 + (1 - \rho_0)\xi_i$. A uniform distribution for γ rather than ρ_0, on an equally spaced grid, is much less appealing. This will become infinitely uniform as the width of the entire grid becomes large.

The choice of k should be based partly on considerations of computational simplicity. In practice, our prior assumptions for γ will however typically be justifiable only if the posterior inferences are insensitive to the choices of k and the prior estimate n_0. Reference will be made to a baseline value n^* for n_0, equal to the value of γ for which the average shrinkage proportion $\bar{\rho}$ in (6) is equal to 1/2. In pragmatic terms, n^* can be regarded as the value of γ for which, given the observed sample sizes, we judge the monotonicity hypothesis and a general alternative hypothesis to possess equal weight. When all the n_i are equal, n^* is equal to their common value. More generally n^* describes a robust center of location for the n_i.

With η, γ, and λ a priori independent, it is similarly assumed that, for some specified λ_0, the parameter

$$\zeta_0 = \lambda_0/(\lambda_0 + \lambda)$$

is uniformly distributed over the equally spaced grid of points $i/(l + 1)$, for $i = 1, 2, \ldots, l$. The corresponding distribution for λ assigns equal prior probabilities $\delta_i = i/(l + 1)$ to the unequally spaced points

$$h_i = \lambda_0(l - i + 1)/i \qquad (i = 1, 2, \ldots, l) \quad ,$$

yielding the Cauchy-tail prior density $\pi(\lambda) = \lambda_0/(\lambda_0 + \lambda)^2$, for $0 < \lambda < \infty$, in the limiting case, or l gets large. A sensitivity analysis with respective to the choices of l and the prior estimate λ_0 of λ should also be performed. As an alternative specification, the shrinkage proportion ζ in (10) could be taken to be uniformly distributed over the same grid. In this case γ and λ would not be independent.

When γ and λ are independent it may be reasonable to replace γ in (10) by its prior estimate n_0 before taking ζ to be uniformly distributed. This is the same as taking ζ_0 in (5.4) to be uniformly distributed, with the choice $\lambda_0 = n_0 + 1$ for the prior estimate of λ. Our prior estimate for the shrinkage proportion ζ, which controls the weighted average compromise (9), is then equal to the neutral value of 1/2. The specification $\lambda_0 = n_0+1$ should not therefore unduly bias our investigation of the monotonicity hypothesis, and is consequently recommended as a baseline choice. The initial baseline selections $n_0 = n^*$ and $\lambda_0 = n^* + 1$, when followed by a careful sensitivity analysis, promise a reasonably fair evaluation of the information regarding possible monotonicity contained in the current data.

Let $\tilde{\rho}^*$ and $\tilde{\zeta}^*$ denote the posterior means of the bounded parameters $\tilde{\rho} = n^*/(n^* + \gamma)$ and $\tilde{\zeta} = (n^* + 1)/(n^* + \lambda + 1)$ under the preceding prior assumptions, where the prior parameters n_0 and λ_0 may differ from the values n^* and $n^* + 1$. The posterior means of the unbounded parameters γ and λ invariably become arbitrarily large as k and l get large. We therefore recommend estimating γ and λ in the posterior assessment by the inverse transformations

$$\gamma^* = n^*(1 - \tilde{\rho}^*)/\tilde{\rho}^* \quad ,$$

and

$$\lambda^* = (n^* + 1)(1 - \tilde{\zeta}^*)/\tilde{\zeta}^* \quad .$$

Unconditional posterior inferences for the θ_i and ξ_i promise to be reasonably insensitive to the choices of k and l, since their posterior distributions, given γ and λ, depend only upon bounded functions of γ and λ.

6. Performance Indicators for Quality Monitoring

The conclusions described in section 2.1 for the data introduced there are now discussed further. The solid plot in Figure 1 describes the piecewise linear isotonic regression, defined in section 4.3, of the ξ_i, upon the 1992 raw proportions x_i. Smoothed performance indicators $\tilde{\xi}_1^*, \tilde{\xi}_2^*, \ldots, \tilde{\xi}_{159}^*$ for 1993, under conditionally independent beta-binomial assumptions, are thereby available. This ordering is consistent with the rank ordering of raw proportions for 1992. The posterior standard deviations of the ξ_i decrease from std(ξ_1) = 0.032 (with $n_1 = 350$ and $s_1 = 0.022$) to std(ξ_{69}) = 0.10 (with $n_{69} = 786$ and $s_{69} = 0.017$). They then increase from std($\tilde{\xi}_{109}^*$) = 0.010 (with $n_{109} = 301$ and $s_{109} = 0.027$) to std($\tilde{\xi}_{159}^*$) = 0.031 (with $n_{159} = 481$ and $s_{159} = 0.022$). They are however generally much smaller than the corresponding s_i.

After an initial sensitivity analysis, it was assumed that $k = 99$ and $l = 24$. The baseline values $n^* = 242.77$ and $n^* + 1 = 243.77$ are employed for γ_0 and η_0, and the posterior conclusions can again be shown to be reasonably insensitive to these assumptions. The posterior estimates for γ and λ are $\gamma^* = 25.99$ and $\lambda^* = 106.46$. As $\bar{\rho}$ has posterior mean 0.862 and standard deviation 0.013, with τ virtually equal to zero, there is negligible evidence to substantiate (1) under binomial sampling assumptions. As the shrinkage proportion ζ has posterior mean 0.214 and standard deviation 0.071, the $\tilde{\xi}_i^*$ are substantially smoothed towards a common value. The location parameter η possesses posterior mean 0.439 and standard deviation 0.010.

An intuitive overall evaluation of our monotonicity specification may be made by reference to the average squared normalized residual

$$W = \sum_{i=1}^{m} r_i^2/m \quad .$$

In the current example, $W = 1.006$. A full residual analysis, though not reported here, can be roughly inferred from Figure 2. This indicates that the data are largely consistent with (2). In other words, the performance indicators for 1993 are largely consistent with the rank ordering for 1992 when sensible extra-binomial variation is permitted. The most discrepant r_i, for hospitals 39, 44, 66, 77, 153, 157, and 158, were respectively 2.51, 2.38, 2.84, 2.36, 2.79, -3.56, and 2.76, corresponding to the sample sizes 220, 20, 40, 1630, 176, 702, and 78. However, when the four hospitals 39, 77, 153, 157 were dropped from the analysis a larger value of $W = 1.037$ was obtained. Moreover, several further discrepant residuals appeared. It was therefore decided to include all original 159 hospitals in the analysis.

The two dotted plots in Figure 1 graph the $p_i - s_i^*$ and the $p_i + s_i^*$ where

$$s_i^* = (D_i^*)^{\frac{1}{2}} \{\tilde{\xi}_i^*(1 - \tilde{\xi}_i^*)/n_i\}^{\frac{1}{2}}$$

for $i = 1, 2, \ldots, m$, with $D_i^* = (n_i + \gamma^*)/(1 + \gamma^*)$, is the estimated standard error of p_i under independent beta-binomial sampling assumptions. These estimated standard errors are quite large, ranging in magnitude from 0.092 to 0.237, though mainly in the region of 0.10. For a typical sample size of 250 our extra-binomial assumptions inflate the estimated standard errors by a factor of 3.20. The predictions of sample proportions for future years, with comparable sample sizes, are likely to be subject to greater random variability.

The solid graph in Figure 2 indicates that the performances for 1993 are also largely consistent with the rank ordering of the DRG predictions. The analysis assumed the same prior parameters as for Figure 1 and yielded $W = 1.005$, $\gamma^* = 24.38$, and $\lambda^* = 101.17$. The posterior means of $\bar{\rho}$, ζ, and η were 0.868, 0.209, and 0.443, with respective posterior standard deviations 0.013, 0.059, and 0.010. There is a remarkable similarity with the corresponding posterior quantities underlying the analysis for Figure 1. This further emphasizes the close comparability of the predictive performances of the quite different rank orderings, based upon the 1992 raw indicators, and the DRG predictions for 1993.

The accuracy of prediction from this noisy data set is open to some improvement by reference to the binomial logit/normal random effects time series formulation employed by West et al. (1998). See also Aguilar et al. (1999). This general paradigm offers considerable scope for incorporating information from years previous to 1992, and combining information across the hospitals. For, say 1993, West et al. assume a simple linear regression for the binomial logits, upon the logits of the DRG predictions. Separate fixed effects regression parameters are estimated for each year. Random error terms, expressing assumed autoregressive time dependence and the representing the substantial residual variation in the data, are added to the regression functions. Any estimated standard errors of the sample proportions should refer to appropriate marginal distributions under random effects assumptions, since these can express the extra-binomial variability inherent in the data. West et al. demonstrate that the total lower level random effects variability is very large, thus again highlighting possible difficulties with prediction. They obtain very useful descriptive conclusions regarding the regression coefficients. More generally, the usefulness of performance indicators and quality monitoring, for predictive rather than descriptive purposes, is open to further discussion when the data are not objectively generated by random sampling schemes.

Acknowledgements

The authors gratefully acknowledge Kam Wah Tsui, Chin-Shung Chung and Bob Mau, for their generous earlier contributions.

References

Aguilar, O., & West, M. (1998). Analysis of Hospital Quality Monitors Using Hierarchical Time Series Models. *B*ayesian Statistics in Science and Technology: Case Studies IV, (eds: C. Gatsonis, R.E. Kass, B. Carlin, A. Carriquiry, A. Gelman, I. Verdinelli, and M. West). New York, NY: Springer-Verlag.

Aguilar, O., Huerta, G., Prado, R., & West, M. (1999). Bayesian Inference on Latent Structure in Time Series. In *B*ayesian

Statistics 6 (eds: J. M. Bernardo, J. O. Berger, A. P. Dawid, & A.F.M. Smith). Oxford: Oxford University Press.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1972). *S*tatistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression. New York: John Wiley & Sons.

Broffitt, J. D. (1988). Increasing and Increasing Convex Bayesian Graduation (with discussion). *T*ransactions of Society of Actuaries, 40, 115-418.

Burgess, J., Lourdes, V., & West, M. (2000). Profiling Mental Health Provider Trends in Health Care Delivery Systems. *H*ealth Services and Outcomes Research Methodology, 1, 253-276. https://doi.org/10.1023/A:1011495210314

Carlin, B. P. (1992). A Simple Monte Carlo Approach to Bayesian Graduation. *T*ransactions of Society of Actuaries, 44, 55-76.

Crook, J. F., & Good, I. J. (1982). The powers and strengths of tests for multinomials and contingency tables. *J*ournal of the American Statistical Association, 77, 793-802. https://doi.org/10.1080/01621459.1982.10477890

Lee, Y., & Nelder, J. A. (1996) Hierarchical Generalized Linear Models (with discussion), *J*ournal of the Royal Statistical Society, Ser. B, 58, 619-678.

Leonard, T. (1972). Bayesian Methods for Binomial Data. *B*iometrika. 59, 581-589. https://doi.org/10.1093/biomet/59.3.581

Leonard, T. (1976). Some Alternative Approaches to Multiparameter Estimation. *B*iometrika, 63, 69-76. https://doi.org/10.1093/biomet/63.1.69

Leonard, T., & Novick, M. R. (1986). Bayesian Full Rank Marginalization for Two-way Contingency Tables. *J*ournal of Educational Statistics, 11, 33-56. https://doi.org/10.2307/1164847

Liu, C. (2000). Estimation of Discrete Distributions with a class of Simple Constrains. *J*ournal of the American Statistical Association, 95, 109-120. https://doi.org/10.1080/01621459.2000.10473907

O'Hagan, A., & Leonard, T. (1976). Bayes Estimation Subject to Uncertainty about Parameter Constraints. *B*iometrika, 63, 201-203. https://doi.org/10.1093/biomet/63.1.201

Warn, D. E., Thompson, S. G., & Spiegelhalter, D. J. (2002). Bayesian Random Effects Meta-analysis of Trials with Binary Outcomes: Methods for the Absolute Risk Difference and Relative Risk Scales. *S*tatistics in Medicine, 21, 1601-1623. https://doi.org/10.1002/sim.1189

West, M., & Aguilar, O. (1997). Studies of Quality Monitor Time Series: The VA Hospital System. *R*eport for the VA Management Science Group. Bedford, MA.

West, M., Aguilar, O., & Lourdes, V. (1998). *V*A Hospital Quality Monitors: 1988-1997. Bedford, MA.

West, M., & Harrison, P. J. (1989). *B*ayesian Forecasting and Dynamic Models. New York: Springer-Verlag.

Yang, T. Y., & Swartz, T. B. (2005). Applications of Binary Segmentation to the Estimation of Quantal Response Curve and Spatial Intensity. *B*iometrical Journal, 47, 489-501. https://doi.org/10.1002/bimj.200310136

Appendices

Appendix 1: A Simple Approximation

Let $\theta|\xi \sim \text{Beta}\{\gamma\xi, \gamma(1-\xi)\}$, where $\xi \sim \text{Beta}(b_0, b_1)$, with $b_0 = \lambda\eta$ and $b_1 = \lambda(1-\eta)$. For fixed γ, λ, and η, we consider the approximation

$$\xi|\theta \sim \text{Beta}\{(\gamma+1)\theta + b_0, (\gamma+1)(1-\theta) + b_1\} \tag{18}$$

to the conditional distribution of ξ given θ. In Figure 3 we compare the corresponding approximate and exact densities, for the choices $\eta = 0.3$, $\gamma = 10$ and $\lambda = 11$, so that $b_0 = 3.3$ and $b_1 = 7.7$, and for six different values (0.05, 0.25, 0.40, 0.60, 0.75, and 0.95) of θ. The approximate (dotted) curves are close to the corresponding exact (solid) curves, unless θ is very different from λ. It is also possible to show that they substantially increase in accuracy as b_0, b_1 or γ increases. Some slight algebraic rearrangement of (18) justifies the approximation in (8) and a modest extension suggests the approximation in (12).

The approximation in (18)may be motivated by noting that, given ξ, $\tilde{y} = (\gamma+1)\theta$ possesses mean $\tilde{n}\xi$ and variance $\tilde{n}\xi(1-\xi)$ where $\tilde{n} = \gamma + 1$. By matching first two moments, we see that when $\tilde{n}$ is an integer, a specified value of $\tilde{y}$ provides similar information regarding ξ as if $\tilde{y}$ represented the realization of a $\text{BIN}(\xi, \tilde{n})$ variate. This indicates the plausibility of the discrete approximation, $(\gamma+1)\theta|\xi \sim \text{BIN}(\xi, \tilde{n})$ to the continuous exact distribution. Subject to this approximation, the conjugate analysis for the binomial distribution, then tells that $\xi|\tilde{y} \sim \text{Beta}\{b_0 + \tilde{y}, b_1 + \tilde{n} - \tilde{y}\}$, which is equivalent to (18)

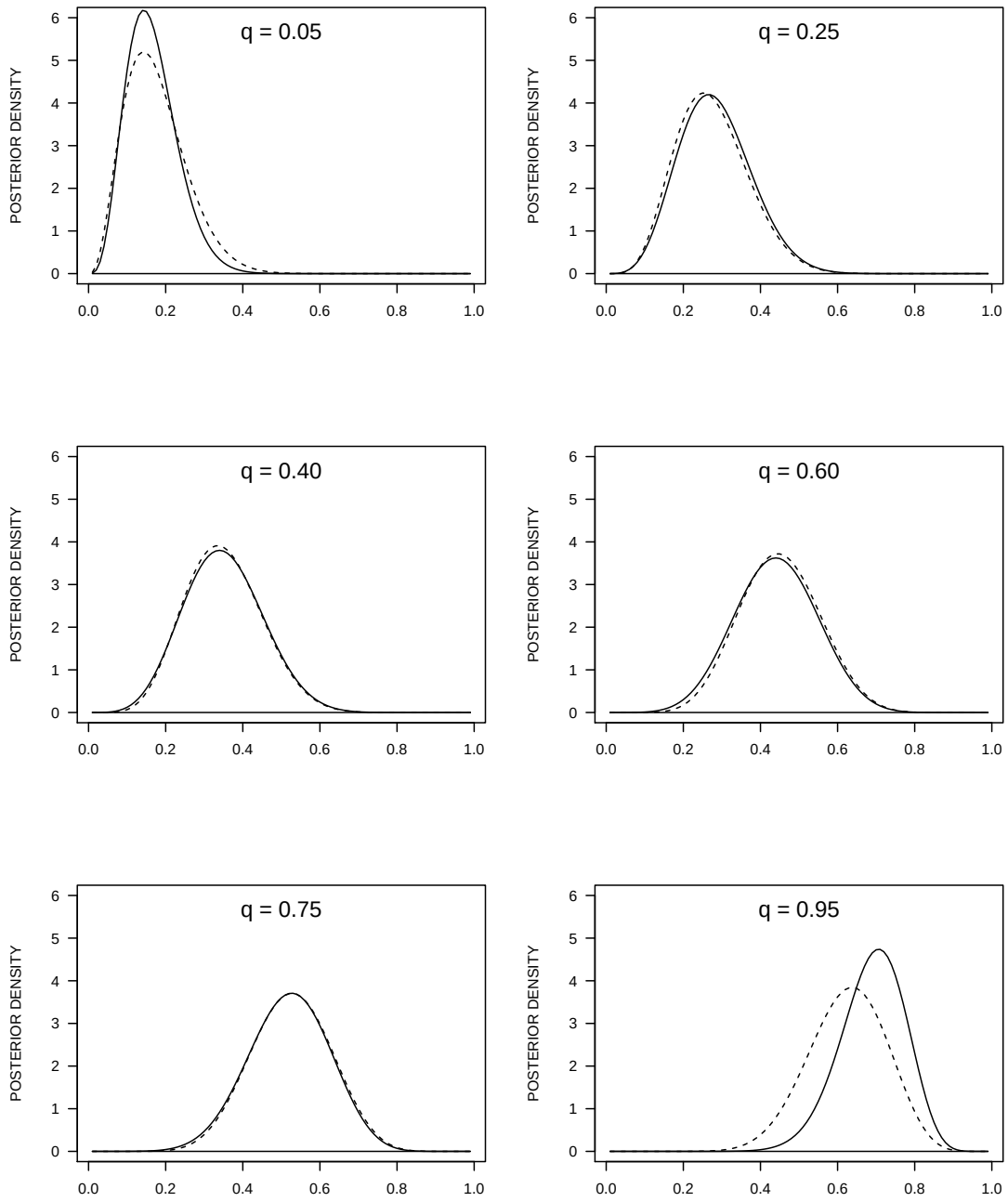

FIGURE 3: BETA APPROXIMATIONS

Our derivation is not however as convincing as the numerical comparisons. The result certainly needs to be inferred and subsequently numerically validated in situations when $\tilde{n}$ is not an integer.

Appendix 2: Posterior Computations

We employ standard Metropolis algorithm/MCMC procedures based upon successive simulations from the following conditional distributions, which all refer to the joint distribution of the θ_i, ξ_i, γ, λ, and η, conditional on the observed data:

(D1) Given the ξ_i and γ, the θ_i are independent and beta distributed, with respective sample sizes $n_i + \gamma$ and means in (4).

(When $n_i = 0$, θ_i possesses a beta distribution with sample size γ and mean ξ_i.)

(D2) Given the θ_i, γ, η, and λ, an approximate joint distribution for the ξ_i constrains the m independent distributions in (8) to the region (2).

(D3) The distribution of γ, given the ξ_i, but unconditional upon the θ_i, assigns probabilities $\pi_1^*, \pi_2^*, \ldots, \pi_k^*$ to the points $g_1, g_2, \ldots, g_k$, where

$$\pi_i^* \propto \pi_i l^*(h_i|\boldsymbol{\xi}, \mathbf{y}) \quad , \tag{19}$$

for $i = 1, 2, \ldots, m$, with $\pi_1, \pi_2, \ldots, \pi_k$ denoting the corresponding prior probabilities, and

$$l^*(\gamma|\boldsymbol{\xi}, \mathbf{y}) = \prod_{k=1}^{m} l^*(\gamma|\xi_k, y_k) \quad , \tag{20}$$

where the contributions to the product on the right hand side of (20) are defined in (3). It is essential to refer to (19) rather than posterior probabilities for γ, given the ξ_i and θ_i, in order to avoid insurmountable instabilities in the posterior computations.

(D4) The distribution of λ, given η and the ξ_i, assigns probabilities $\delta_1^*, \delta_2^*, \ldots, \delta_l^*$ to the points $g_1, g_2, \ldots, g_l$, where

$$\delta_i^* \propto \delta_i \tilde{l}(\eta, g_i|\boldsymbol{\xi}) \quad ,$$

for $i = 1, 2, \ldots, l$, with $\delta_1, \delta_2, \ldots, \delta_l$ denoting the corresponding prior probabilities, and $\tilde{l}(\eta, \lambda|\boldsymbol{\xi})$ defined in (15).

(D5) The distribution of η, given λ and the ξ_i, may be approximated by the beta distribution in (12).

The simulations from D2 can be made effectively exact. The constrained beta approximations can be handled by successive sampling from truncated beta distributions. When generating values for η, just simulate from the approximate distribution in (12). This conditional distribution can be highly concentrated, for large λ, about its mean in (13) and the corresponding exact density in (14) can be highly peaked around a slightly different location. Acceptance sampling for η can therefore lead to a high rejection rate. However, subject to our minor approximation, all posterior quantities of interest can be calculated in standard fashion.

About 200,000 successive simulations on all parameters are recommended for good practical accuracy, after an initial burn-in period of about 1,000 simulations. Good starting values in D1 are $\gamma = n^*$, our baseline prior estimate, and $\xi_i = p_i$ for $i = 1, 2, \ldots, m$. Increasing the numbers k and l of grid points too much will not necessarily provide completely exact representations of Bayesian inferences under a continuous prior distribution. The errors of our discrete approximation to a continuous posterior distribution will confound with the errors of simulation.