REVSTAT – Statistical Journal
Volume 3, Number 1, June 2005, 61–75

PREDICTION OF TREATMENTS EFFECTS IN A BIASED ALLOCATION MODEL

Authors: Fernando J.M. Magalhães
– ISCAP / Inst. Politécnico do Porto, Univ. Portucalense and CEAUL, R. Prof. Jaime Lopes Amorim, 4465-111 S. Mamede de Infesta, Portugal

Ian R. Dunsmore
– Department of Probability and Statistics, University of Sheffield, Sheffield S3 7RH, U.K.

David J. Robson
– Department of Probability and Statistics, University of Sheffield, Sheffield S3 7RH, U.K.

Received: April 2004 Revised: November 2004 Accepted: February 2005

Abstract:

- Robbins and Zhang [15] provide consistent estimators of multiplicative treatment effects under a biased treatment allocation scheme, and illustrate their methodology within Poisson and binomial models. Here we use predictive criteria to assess the differential treatment effects, and develop predictive distributions for the Poisson errors in variables models. With a hierarchical prior structure, various approximations are investigated, and an illustrative example is included.

Key-Words:

- *biased allocation; errors in variables; Gibbs sampling; Laplace approximation; Poisson model; predictive distributions; treatment effect.*

AMS Subject Classification:

- 62F15, 62F10, 62P10, 62E15, 62E17, 62E20.

1. INTRODUCTION

Robbins and Zhang ([13], [14], [15]) consider the estimation of a multiplicative treatment effect under biased allocation. For example, with a slight change from their notation to allow for generalization, suppose that within a Poisson errors in variables model (θ_i, X_i, Y_i), $i=1,2,...,n$, are independent random vectors such that

(1.1)
$$\begin{aligned}
&\textbf{(i)} \quad \text{given } \theta_i, X_i \text{ is } Po(\theta_i)\,;\\
&\textbf{(ii)} \quad \text{given } \beta_1,\ \beta_2,\ \theta_i \text{ and } x_i,\\
&\qquad\qquad Y_i \text{ is } Po(\beta_1\theta_i) \text{ if treatment } T_1 \text{ is used}\,;\\
&\qquad\qquad Y_i \text{ is } Po(\beta_2\theta_i) \text{ if treatment } T_2 \text{ is used}\,;\\
&\textbf{(iii)} \quad \text{given } a,\ T_1 \text{ is used if } x_i < a \ \text{ and } \ T_2 \text{ is used if } x_i \geq a\ ;
\end{aligned}$$

where $Po(\mu)$ represents a Poisson distribution with mean μ. No distributional assumptions about the θ_i's are made, and their values are not observed. The unknown parameters β_1 and β_2 could be thought of as multiplicative treatment effects. An alternative parameterization would be through logarithmic link functions with additive treatment effects.

Robbins and Zhang [15] discuss two scenarios for this model. The first concerns the number of accidents at road junctions. Suppose that X_i counts the number of night accidents during year 1 at junction i, $i=1,2,...,n$. Extra lights are installed at the beginning of year 2 at those junctions for which $x_i \geq a$, with no change being made to the light system at other junctions. Then, Y_i records the number of night accidents at junction i during year 2. Of particular interest is whether or not the extra lights reduce the frequency of night accidents.

The second, more controversial, application is in the context of clinical trials in which the allocation of treatments is based on the screening variable X. Robbins and Zhang [15] then seek to estimate the differential treatment effects based on this biased allocation of treatments to patients.

In both situations Robbins and Zhang [15] consider the problem as one of estimation, and take the difference $\beta_2 - \beta_1$ or the ratio β_2/β_1 as a measure of the differential treatment effect. Based on data $(\boldsymbol{x}^n, \boldsymbol{y}^n) = \{(x_i, y_i):\ i=1,2,...,n\}$, they derive the following consistent estimates for β_1 and β_2:

(1.2)
$$\beta_{1,n} = \frac{\sum_{i=1}^{n} y_i\, I(x_i < a)}{\sum_{i=1}^{n} x_i\, I(x_i < a+1)}\ , \qquad \beta_{2,n} = \frac{\sum_{i=1}^{n} y_i\, I(x_i \geq a)}{\sum_{i=1}^{n} x_i\, I(x_i \geq a+1)}\ ,$$

where I represents the indicator function. They suggest the use of the central limit theorem to obtain confidence intervals for β_1, β_2, $\beta_2 - \beta_1$ or β_2/β_1

with coverage probabilities that tend to 0.95, say, as $n \to \infty$, but omit details. They also note that the Poisson assumption for the conditional distribution of Y_i in (ii) of (1.1) is not required for the consistency of the estimates in (1.2).

Godambe and Kunte [6] provide an alternative semi-parametric solution for the estimation of β_1 and β_2 through the use of optimum estimating functions (Godambe and Thompson, [7]). Their model does not require the Poisson assumptions in (1.1) for X_i or Y_i, but only the mean value specifications. However, they do require an additional assumption, namely that

(**iv**) given x_i the mean value of θ_i is $f(x_i)$,
f being a specified function of x_i .

The assumption that the θ_i are (unobservable) random variables distinguishes the model from one in which they are unknown parameters. In this latter case, one might then consider the θ_i as "incidental" parameters, following the terminology of Neyman and Scott [10], as opposed to the "structural" parameters β_1 and β_2. Kiefer and Wolfowitz [8] discuss problems of consistency with maximum likelihood estimation in such cases, and illustrate how these may be overcome if the θ_i are independent chance variables with a common distribution, as in here.

We develop here, in Section 2, an approach to treatment comparisons based on predictive criteria, which perhaps seem more relevant for answering, in the medical context for example, the question "Which of the two treatments do I give to the next patient?". The approach extends the models used in Dunsmore and Robson [2] for other Poisson errors in variables models. We concentrate attention on the outcomes $Y_{n+1,1}$ and $Y_{n+1,2}$ from separate applications of the two treatments, T_1 and T_2, applied to patient $n+1$, and seek to make predictions for future values $y_{n+1,1}$ and $y_{n+1,2}$ based on $(\boldsymbol{x}^n, \boldsymbol{y}^n)$ and x_{n+1}.

An illustrative example is provided in Section 3, and extensions to several treatments and other distributional models are discussed briefly in sections 4 and 5.

2. POISSON PREDICTIVE MODELS

2.1. Predictive distribution

Consider the Poisson errors in variables model specified in (1.1). Suppose further that for a future individual, labelled by $n+1$, we observe x_{n+1} from a $Po(\theta_{n+1})$ distribution and model potential outcomes from the two treatments through

$$(2.1) \qquad Y_{n+1,1} \text{ is } Po(\beta_1\theta_{n+1}) ; \qquad Y_{n+1,2} \text{ is } Po(\beta_2\theta_{n+1}) .$$

The dependence between the outcomes for this individual from the two treatments is modelled through the common (unobserved) θ_{n+1}.

Such an individual will only be given one of the two treatments, and the predictive paradigm suggests that the choice centres around properties of the joint predictive function $p\big(y_{n+1,1}, y_{n+1,2} \,|\, x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big)$, or perhaps considerations of $Y_{n+1,2} - Y_{n+1,1}$.

We denote the treatment given to an individual, for $i = 1, 2, ..., n$, by

$$\delta_{ij} = \begin{cases} 1, & \text{if individual } i \text{ gets treatment } T_j \,, \\ 0, & \text{otherwise} \,, \end{cases}$$

so that $\delta_{i1} + \delta_{i2} = 1$ for each i; and let, for $j = 1, 2$,

$$\begin{aligned} n_j &= \sum_{i=1}^{n} \delta_{ij} \;; \qquad & n &= n_1 + n_2 \;; \\ S_{xj} &= \sum_{i=1}^{n} \delta_{ij}\, x_i \;; \qquad & T_x &= \sum_{j=1}^{2} S_{xj} \;= \sum_{i=1}^{n} x_i \;; \\ S_{yj} &= \sum_{i=1}^{n} \delta_{ij}\, y_i \;; \qquad & T_y &= \sum_{j=1}^{2} S_{yj} \;= \sum_{i=1}^{n} y_i \;. \end{aligned}$$

The maximum likelihood estimates of the parameters are given by

$$\hat{\theta}_i = \frac{x_i + y_i}{1 + \sum_{j=1}^{2} \delta_{ij}\, \hat{\beta}_j}, \quad i=1,2,...,n; \quad \hat{\theta}_{n+1} = x_{n+1}; \quad \hat{\beta}_j = \frac{S_{yj}}{S_{xj}},\ j=1,2 \;. \tag{2.2}$$

In this notation the Robbins and Zhang [15] estimates (1.2) based on $(\boldsymbol{x}^n, \boldsymbol{y}^n)$ are given by

$$\beta_{1,n} = \frac{S_{y1}}{S_{x1} + aN_a} \;, \qquad \beta_{2,n} = \frac{S_{y2}}{S_{x2} - aN_a} \;, \tag{2.3}$$

where N_a are the number of x_i's equal to a. The additional information provided by x_{n+1} could be used to amend these estimates to

$$\begin{aligned} \beta^*_{1,n} &= \left(1 + \frac{1}{n}\right) \frac{S_{y1}}{S_{x1} + aN_a + x_{n+1}\, I\,(x_{n+1} < a+1)} \;, \\ \beta^*_{2,n} &= \left(1 + \frac{1}{n}\right) \frac{S_{y2}}{S_{x2} - aN_a + x_{n+1}\, I\,(x_{n+1} \geq a+1)} \;. \end{aligned} \tag{2.4}$$

2.2. Plug-in estimates

Predictive approaches within the classical framework typically involve plug-in estimates, pivotal statistics or tolerance regions. A simple plug-in estimate for the probability function of $Z = Y_{n+1,2} - Y_{n+1,1}$, for example, would be given by

$$P(Z=z) \;=\; \sum_{i=\max(0,-z)}^{\infty} \prod_{j=1}^{2} \frac{(\hat{\beta}_j\,\hat{\theta}_{n+1})^{i+(j-1)z}\,\exp(-\hat{\beta}_j\,\hat{\theta}_{n+1})}{\big(i+(j-1)z\big)!}\;. \tag{2.5}$$

Other such estimates are available if specification of the underlying distribution for the θ_i's is provided. As an illustration, we take $p(\theta_i)$ to be $\gamma\exp(-\gamma\theta_i)$ with unknown parameter $\gamma > 0$. The model specification in (1.1) then reduces to

$$p\big(x_i, y_i \mid \beta_1, \beta_2, \gamma\big) \;=\; \frac{(x_i+y_i)!}{x_i!\;y_i!}\;\frac{\gamma\,\beta_1^{\delta_{i1}y_i}\,\beta_2^{\delta_{i2}y_i}}{\big(1+\gamma+\delta_{i1}\beta_1+\delta_{i2}\beta_2\big)^{x_i+y_i+1}}\;, \tag{2.6}$$

whilst Godambe and Kunte's [6] condition (iv) above is satisfied by $f(x) = (x+1)/(\gamma+1)$. Both Robbins and Zhang [15] and Godambe and Kunte [6] consider this case. The former demonstrate that their estimates $\beta_{1,n}$ and $\beta_{2,n}$ in (1.2) compete well with the maximum likelihood estimates based on $(\boldsymbol{x}^n, \boldsymbol{y}^n)$ from this fully parametric model; whilst the latter's solution coincides with them.

With the additional information from x_{n+1}, the maximum likelihood estimates are now given by

$$\breve{\beta}_1 = (1+\breve{\gamma})\,\frac{S_{y1}}{S_{x1}+n_1}\;;\quad \breve{\beta}_2 = (1+\breve{\gamma})\,\frac{S_{y2}}{S_{x2}+n_2}\;;\qquad \breve{\gamma} = \frac{n+1}{T_x + x_{n+1}}\;. \tag{2.7}$$

A simple plug-in estimate for the probability function of $Z = Y_{n+1,2} - Y_{n+1,1}$ would then be given by

$$P(Z=z) \;=\; \sum_{i=\max(0,-z)}^{\infty} \frac{(z+2i)!}{i!\,(z+i)!}\;\frac{\breve{\gamma}\,\breve{\beta}_1^{i}\,\breve{\beta}_2^{i}}{\big(\breve{\gamma}+\breve{\beta}_1+\breve{\beta}_2\big)^{z+2i+1}}\;. \tag{2.8}$$

2.3. Hierarchical prior structure

Within a Bayesian framework for the model specified by (1.1) and (2.1), the central feature is the predictive function $p\big(y_{n+1,1}, y_{n+1,2} \,|\, x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big)$ given by

$$\int \prod_{j=1}^{2} \big\{p(y_{n+1,j} \,|\, \beta_j, \theta_{n+1})\big\}\; p\big(\theta_{n+1}, \beta_1, \beta_2 \,|\, x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big)\; d\theta_{n+1}\, d\beta_1\, d\beta_2\;, \tag{2.9}$$

where $p\big(\theta_{n+1}, \beta_1, \beta_2 \,|\, x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big)$ is the posterior density function. Notice here that $\theta^n = (\theta_1, \theta_2, ..., \theta_n)$ behaves in the same way as a nuisance parameter, and we only require the posterior distribution of $(\theta_{n+1}, \beta_1, \beta_2)$ — or indeed only of $(\beta_1\theta_{n+1}, \beta_2\theta_{n+1})$.

Following the ideas in Gelfand and Smith [5], we adopt a Bayesian hierarchical prior structure. At the first stage we take

$$p\big(\theta^n, \theta_{n+1}, \beta_1, \beta_2 \mid \gamma, \eta_1, \eta_2\big) \;=\; \prod_{i=1}^{n+1} p(\theta_i \,|\, \gamma) \;\prod_{j=1}^{2} p(\beta_j \,|\, \eta_j) \;,$$

whilst at the second stage we assume

$$p(\gamma, \eta_1, \eta_2) \;=\; p(\gamma) \prod_{j=1}^{2} p(\eta_j) \;.$$

An appropriate structure here would be of the form

$$\theta_i \sim Ga(k, \gamma) \;, \qquad \beta_j \sim Ga(g_j, \eta_j) \;,$$
$$\gamma \sim Ga(\ell, m) \;, \qquad \eta_j \sim Ga(u_j, v_j) \;,$$

for $i = 1, 2, ..., n, n{+}1$ and $j = 1, 2$, where $Ga(a, b)$ represents a gamma distribution with density proportional to $\theta^{a-1}\exp(-b\theta)$, $\theta > 0$, and where k, g_1, g_2, ℓ, m, u_1, v_1, u_2 and v_2 are known constants. Gaver and O'Muircheartaigh [3] and Gelfand and Smith [5] suggest, in a similar framework, that k, g_1 and g_2 might be treated as tuning parameters or estimated in an empirical Bayes spirit. Notice that the distributional assumptions about θ^n and θ_{n+1} in Section 2.2 are a special case of the above.

The posterior density function $p\big(\theta^n, \theta_{n+1}, \beta_1, \beta_2, \gamma, \eta_1, \eta_2 \mid x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big)$ is proportional to

$$\begin{aligned} &\prod_{i=1}^{n}\left[\exp\left\{-\theta_i\left(1+\sum_{j=1}^{2}\delta_{ij}\,\beta_j+\gamma\right)\right\}\theta_i^{x_i+y_i+k}\right] \;\times \\ &\qquad\times\;\; \exp\big\{-\theta_{n+1}(1{+}\gamma)\big\}\,\theta_{n+1}^{x_{n+1}+k}\,\gamma^{(n+1)k+\ell}\,\exp\{-m\gamma\} \;\times \\ &\qquad\times\;\; \prod_{j=1}^{2}\Big\{\beta_j^{S_{yj}+g_j}\exp(-\eta_j\beta_j)\,\eta_j^{g_j+u_j}\exp(-v_j\eta_j)\Big\} \;. \end{aligned} \tag{2.10}$$

Eliminating θ^n, η_1 and η_2 we have that $p\big(\theta_{n+1}, \beta_1, \beta_2, \gamma \,|\, x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big)$ is proportional to

$$\frac{\exp\big\{-\theta_{n+1}(1{+}\gamma)\big\}\,\theta_{n+1}^{x_{n+1}+k}\;\prod_{j=1}^{2}\big\{\beta_j^{S_{yj}+g_j}\big\}\,\gamma^{(n+1)k+\ell}\exp\{-m\gamma\}}{\prod_{j=1}^{2}\big\{(1+\beta_j+\gamma)^{W_j}\,(v_j+\beta_j)^{g_j+u_j}\big\}} \;, \tag{2.11}$$

where $W_j = S_{xj} + S_{yj} + k\,n_j,\ j=1,2$. Elimination of γ cannot be undertaken explicitly, but we find that $p\big(y_{n+1,1}, y_{n+1,2} \,|\, x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big)$ is proportional to

$$\int \frac{\Gamma(x_{n+1}+y_{n+1,1}+y_{n+1,2}+k)}{y_{n+1,1}!\ y_{n+1,2}!} \times \times \frac{\prod_{j=1}^{2}\{\beta_j^{S_{yj}+y_{n+1,j}+g_j}\}}{(1+\beta_1+\beta_2+\gamma)^{x_{n+1}+y_{n+1,1}+y_{n+1,2}+k}} \times \times \frac{\gamma^{(n+1)k+\ell}\exp\{-m\gamma\}}{\prod_{j=1}^{2}\{(1+\beta_j+\gamma)^{W_j}\,(v_j+\beta_j)^{g_j+u_j}\}}\, d\beta_1\, d\beta_2\, d\gamma \tag{2.12}$$

for $y_{n+1,1} = 0, 1, \ldots$ and $y_{n+1,2} = 0, 1, \ldots$. The joint predictive probability function may then be found numerically through three dimensional integration techniques.

Although no simple analytical form is available for (2.12) here, it is possible to obtain the marginal (but dependent) predictive probabilities in the case of vague second stage priors ($\ell, m, u_1, v_1, u_2, v_2 \to 0$) explicitly, namely, for $j=1,2$,

$$p\big(y_{n+1,j} \,|\, x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big) \propto \frac{B(x_{n+1}+y_{n+1,j}+k,\ W_j)}{B(y_{n+1,j}+1,\ S_{ij}-1)}\ , \quad y_{n+1,j} = 0, 1, \ldots, \tag{2.13}$$

and these can easily be compared graphically.

2.4. Approximations

Alternatively, we might consider approximations through the use of, for example, posterior normality assumptions, Gibbs sampling or Laplace approximations to evaluate the predictive probabilities.

Noting that, as the sample size increases, the number of parameters in our model also increases, we surmise that problems may arise over assumptions of asymptotic normality of the overall posterior distribution, especially for the usual asymptotic normal approximation for the full posterior distribution in (2.10) — see, for example, Bernardo and Smith ([1, pp. 285–97]). A better result is likely from following O'Hagan's ([11, pp. 208]) suggestion of using a normal approximation for the reduced posterior $p\big(\theta_{n+1}, \beta_1, \beta_2, \gamma \,|\, x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big)$ alone, based on the posterior mode and modal dispersion matrix. We do not pursue this approach here, but further details can be found in Magalhães [9].

2.4.1. Gibbs sampling

The conditional distributions of $\theta_{n+1},\ \beta_1,\ \beta_2$ and γ follow from (2.11) in a straightforward manner, and, using rejection sampling with t iterations in each cycle, we obtain M random samples

$$\Big(\theta^{(t)}_{n+1(\ell)}, \beta^{(t)}_{1(\ell)}, \beta^{(t)}_{2(\ell)}, \gamma^{(t)}_{(\ell)}\Big)\ , \qquad \ell = 1, 2, ..., M\ .$$

The prediction function (2.9) can then be estimated using

$$\hat{p}\big(y_{n+1,1}, y_{n+1,2} \mid x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big) \;=\; \frac{1}{M}\sum_{\ell=1}^{M}\prod_{j=1}^{2}\frac{\mu_{j\ell}^{\,y_{n+1,j}}\ e^{-\mu_{j\ell}}}{y_{n+1,j}!}\ , \tag{2.14}$$

where $\mu_{j\ell} = \beta^{(t)}_{j\ell}\,\theta^{(t)}_{n+1(\ell)}$, $j=1,2$. If interest lies, say, in $Z = Y_{n+1,2} - Y_{n+1,1}$, we then need to derive the predictive distribution of Z.

Notice that, although it is necessary to generate values of $\gamma^{(t)}_{(\ell)}$ in this Gibbs routine, the values of this hyperparameter are not required further for our prediction problem.

2.4.2. Laplace approximation

Since the joint predictive probability function in (2.9) is a posterior expectation, which may be written, in generic form, as

$$E\big\{g(\psi)\,|\,data\big\} \;=\; \frac{\displaystyle\int g(\psi)\,L(data)\,p(\psi)\ d\psi}{\displaystyle\int L(data)\,p(\psi)\ d\psi} \;=\; \frac{\displaystyle\int \exp\{-nh^*(\psi)\}\ d\psi}{\displaystyle\int \exp\{-nh(\psi)\}\ d\psi}\ ,$$

we may also use the Laplace approximation method; see, for example, Bernardo and Smith ([1, pp. 340–5]). In the posterior expectation above, ψ represents an unknown parameter and $L(data)$ is the likelihood function. Also, functions $h(\psi)$ and $h^*(\psi)$ are defined such that

$$-nh(\psi) = \ln p(\psi) + \ln L(data) \quad \text{ and } \quad -nh^*(\psi) = \ln g(\psi) + \ln p(\psi) + \ln L(data)\,.$$

Again, we present the results for the special case of vague second stage priors. A good approximation for (2.9) is given by

$$\frac{1}{y_{n+1,1}!}\,\frac{1}{y_{n+1,2}!}\left(\frac{\sigma^*}{\tilde{\sigma}}\right)\exp\Big[-n\Big\{h^*(\theta^*_{n+1}, \beta^*_1, \beta^*_2, \gamma^*) - h(\tilde{\theta}_{n+1}, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\gamma})\Big\}\Big]\ , \tag{2.15}$$

where

$$-nh(\theta_{n+1},\beta_1,\beta_2,\gamma) \;=\; -\theta_{n+1}(1+\gamma) + (x_{n+1}+k)\,\ln\theta_{n+1} + (n+1)\,k\,\ln\gamma + \sum_{j=1}^{2}\Big\{S_{yj}\ln\beta_j - W_j\ln(1+\beta_j+\gamma)\Big\}\;,$$

$$-nh^*(\theta_{n+1},\beta_1,\beta_2,\gamma) \;=\; = \; -nh(\theta_{n+1},\beta_1,\beta_2,\gamma) + \sum_{j=1}^{2}\Big\{y_{n+1,j}(\ln\theta_{n+1}+\ln\beta_j) - \theta_{n+1}\,\beta_j\Big\}$$

and where $\tilde{\theta}_{n+1}$, $\tilde{\beta}_1$, $\tilde{\beta}_2$, $\tilde{\gamma}$ and θ^*_{n+1}, β^*_1, β^*_2, γ^* are the modes of $-h$ and $-h^*$, respectively. The former are given by

$$\begin{aligned} \tilde{\theta}_{n+1} \;&=\; \frac{(x_{n+1}+T_x)\,(x_{n+1}+k)}{x_{n+1}+T_x+(n+1)k}\;, \qquad \tilde{\gamma} = \frac{(n+1)k}{x_{n+1}+T_x}\;, \\ \tilde{\beta}_j \;&=\; \frac{\{x_{n+1}+T_x+(n+1)k\}S_{yj}}{(x_{n+1}+T_x)\,(W_j-S_{yj})}\;, \quad j=1,2\;; \end{aligned} \tag{2.16}$$

whilst the latter are found iteratively from

$$\theta_{n+1}\left(1+\sum_{j=1}^{2}\beta_j+\gamma\right) \;=\; x_{n+1}\;+\sum_{j=1}^{2}y_{n+1,j}+k\;,$$

$$\beta_j\left(\theta_{n+1}+\frac{W_j}{1+\beta_j+\gamma}\right) \;=\; y_{n+1,j}+S_{yj}\;, \qquad j=1,2\;,$$

$$\gamma\left(\theta_{n+1}+\sum_{j=1}^{2}\frac{W_j}{1+\beta_j+\gamma}\right) \;=\; (n+1)k\;.$$

Finally, $\tilde{\sigma}$ and σ^* are the square roots of the inverse of the determinants of the appropriate hessian matrices of second order derivatives, namely

$$\tilde{\sigma} \;=\; \left|n\,\nabla^2 h(\tilde{\theta}_{n+1},\tilde{\beta}_1,\tilde{\beta}_2,\tilde{\gamma})\right|^{-\frac{1}{2}}\;,$$

$$\sigma^* \;=\; \left|n\,\nabla^2 h^*(\theta^*_{n+1},\beta^*_1,\beta^*_2,\gamma^*)\right|^{-\frac{1}{2}}\;.$$

Full details can be found in Magalhães [9].

3. ILLUSTRATION

In order to illustrate the different approximations, we consider the data shown in Table 1.

Table 1: Simulated data set of size $n = 20$.

x_i	δ_{i1}	y_i	x_i	δ_{i1}	y_i
7	0	3	11	0	22
8	0	8	6	0	13
9	0	1	9	0	10
13	0	16	6	0	10
5	1	1	10	0	16
2	1	2	17	0	16
13	0	12	3	1	0
4	1	2	2	1	2
6	0	12	2	1	1
7	0	4	8	0	11

These $n = 20$ data values were simulated from models with $\beta_1 = 0.3$ and $\beta_2 = 1.4$, with $a = 6$, and for a random selection of θ_i values. Note that $\hat{\beta}_1 = 0.44$ and $\hat{\beta}_2 = 1.25$ from (2.2), whilst the equivalent Robbins and Zhang [15] are $\beta_{1,n} = 0.22$ and $\beta_{2,n} = 1.46$ from (2.3).

Predictions are given for $y_{21,1}$ and $y_{21,2}$ corresponding to $x_{21} = 4$. The amended maximum likelihood estimates are now $\breve{\beta}_1 = 0.38$ and $\breve{\beta}_2 = 1.29$ from (2.7), whilst the Robbins and Zhang values are updated to $\beta^*_{1,n+1} = 0.21$ and $\beta^*_{2,n+1} = 1.53$ from (2.4).

In the analyses we assume a vague second stage prior $(\ell, m, u_1, v_1, u_2, v_2 \to 0)$. For such a case, specification of g_1 and g_2 is not necessary. We take $k = 6$ based on matching the first two marginal moments of the X_i's; see Dunsmore and Robson [2].

A clear picture emerges if we consider the marginal predictive functions for $y_{21,1}$ and $y_{21,2}$ separately. Figures 1 and 2 show the approximations from the two methods together with the exact forms from (2.13). Clearly, the Gibbs and Laplace methods provide excellent approximations to the exact distribution. Also shown in Figures 1 and 2 are the marginal predictive functions with the posterior normal approximations mentioned in Section 2.4. Normal approximation 1 refers to the Bernardo & Smith [1] approach and normal approximation 2 refers to O'Hagan's [11] approach. The anticipated problems with a sample size of only 20 manifest themselves, although perhaps not surprisingly O'Hagan's [11]

suggestion, based on only four parameters, seems superior to the more usual posterior normal approximation, based on 26 parameters. Figures 1 and 2 also show that the predictive approach leads to more disperse distributions than the ones obtained through the plug-in method. This fact is not surprising because the predictive approach incorporates uncertainty about the parameters.

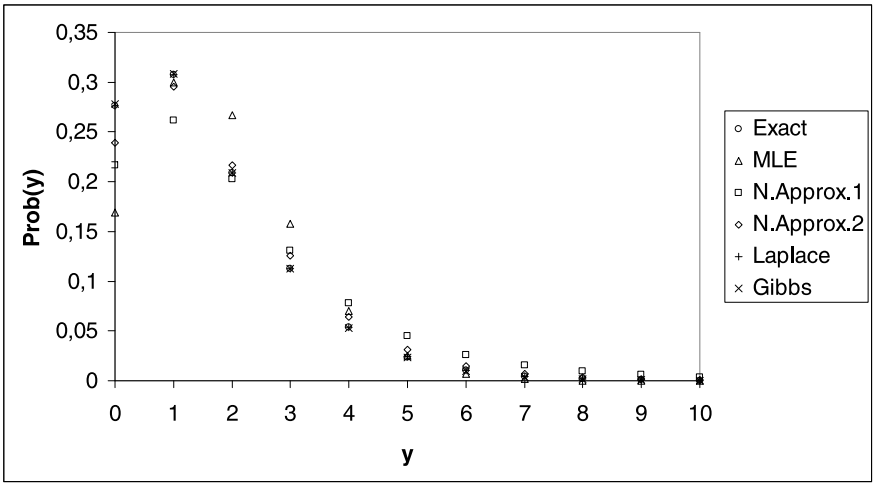

Figure 1: Comparison of the predictive functions $p\big(y_{21,1} \,|\, x_{21}, \boldsymbol{x}^{20}, \boldsymbol{y}^{20}\big)$ from the four approximations with the exact form in (2.13).

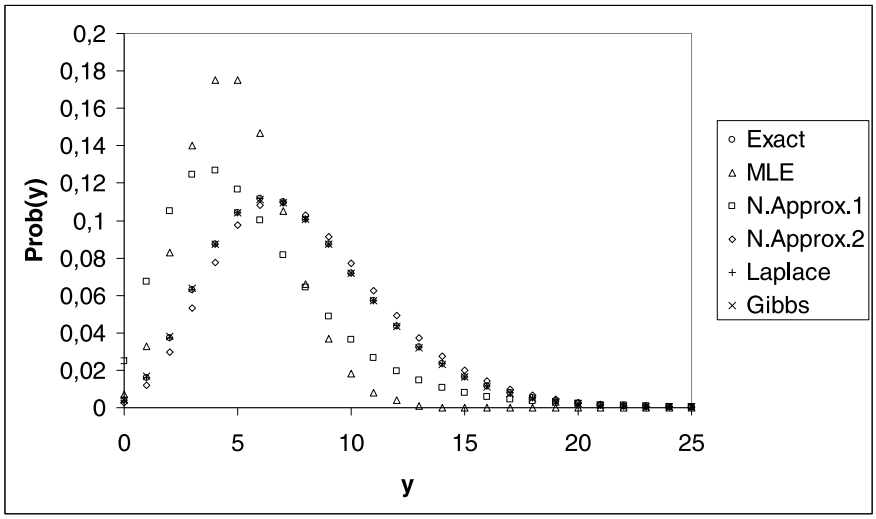

Figure 2: Comparison of the predictive functions $p(y_{21,2} \mid x_{21}, \boldsymbol{x}^{20}, \boldsymbol{y}^{20})$ from the four approximations with the exact form in (2.13).

We may conclude that the Gibbs and Laplace methods lead to excellent results when compared to the exact predictive distribution. The speed of the Laplace method, in comparison to Gibbs sampling, is a strong point in its favour.

Figure 3 compares the predictive functions for $Z = Y_{21,2} - Y_{21,1}$ for the Laplace method and the plug-in method of (2.5), and illustrates the unsatisfactory nature of the latter.

Similar conclusions were drawn in several other simulations.

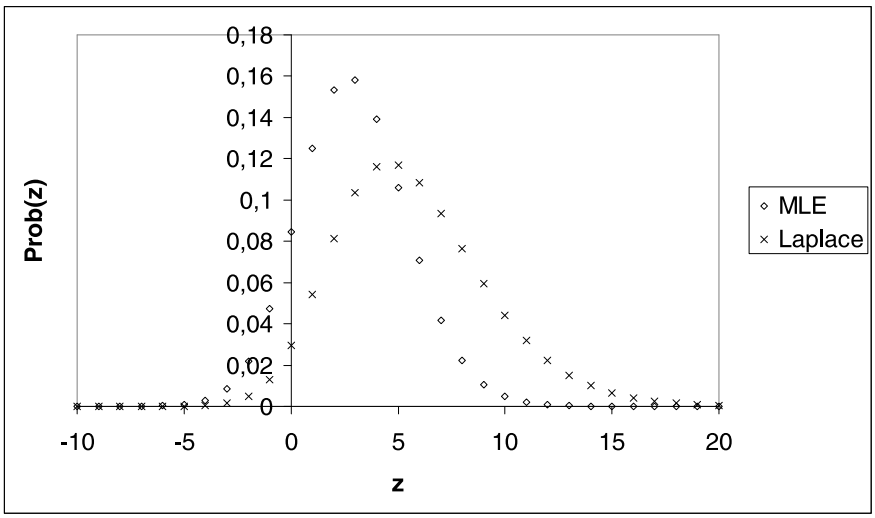

Figure 3: Comparison of the predictive functions for $Z = Y_{21,2} - Y_{21,1}$ from the Laplace method and the plug-in form in (2.8).

4. GENERALISATION TO $J \geq 2$ TREATMENTS

The models can be extended to the case of J treatments in a straightforward manner. Suppose that we can define mutually disjoint and exhaustive subsets C_1, C_2, ..., C_J of the non negative integers, such that treatment T_j is used for individual I if $x_i \in C_j$. We assume that Y_i is $Po(\beta_j \theta_i)$ if treatment T_j is used for individual i, $(i = 1, 2, ..., n+1;\ j = 1, 2, ..., J)$. Notice that the identification of subsets through cut-off points $a_1 < a_2 < ... < a_{J-1}$ is only one possible partition.

Robbins and Zhangs [15] method generalises (1.2) above to give consistent estimates

$$\beta_{jn} = \frac{\sum_{i=1}^{n} y_i \, I(x_i \in C_j)}{\sum_{i=1}^{n} x_i \, I(x_i - 1 \in C_j)}$$

for $j = 1, 2, ..., J$. Similarly, with an obvious extension of the notation for δ_{ij}, n_j, S_{xj}, S_{yj}, T_x, T_y, the maximum likelihood estimates corresponding to (2.2) generalise in a simple way.

Numerical integration for the joint predictive probability

$$p\big(y_{n+1,1}, y_{n+1,2}, ..., y_{n+1,J} \mid x_{n+1}, \boldsymbol{x}^n, \boldsymbol{y}^n\big) \ ,$$

corresponding to (2.9) becomes impractical, but the Gibbs and Laplace methods provide approximations. Full details are again given in Magalhães [9].

Of interest now might be predictive probabilities associated with $\max(y_{n+1,1},\ y_{n+1,2},\ ...,\ y_{n+1,J})$. Within the Gibbs framework, one way of deriving these would be to consider the $y_{n+1,1},\ y_{n+1,2},\ ...,\ y_{n+1,J}$ as missing data and within each cycle to generate values of $y_{n+1,j}$ from a $Po(\beta_j\theta_{n+1})$ distribution, $j = 1, 2, ..., J$.

From the resulting samples $\big(y^{(t)}_{n+1,1(\ell)},\ y^{(t)}_{n+1,2(\ell)},\ ...,\ y^{(t)}_{n+1,J(\ell)}\big)$ it is then straightforward to approximate the probability that treatment T_j, say, provides the maximum response.

5. CONCLUSIONS

We have developed Bayesian predictive models for a Poisson errors in variables situation in which there are simple, multiplicative effects. Whilst standard numerical integration techniques, here in three dimensions, might be suitable for the determination of the appropriate predictive distributions, we have found that Laplace approximation and Gibbs sampling can provide alternative and reliable approaches. The use of the posterior normal approximations can be suspect because of the high dimensionality of the parameters, although O'Hagan's [11] approach improves matters somewhat.

Robbins and Zhang [15] also consider estimation in a binomial model, whilst Robbins [12] discusses the exponential case. Similar predictive frameworks can be developed for these situations. For example, in the model specification in Section 2 we might replace the Poisson assumptions (i) and (ii) by

(**i**) given r and θ_i, X_i is $Bi\left(r, \dfrac{\theta_i}{1+\theta_i}\right)$;

(**ii**) given s $\beta_1, \beta_2, \theta_i$ and x_i,

$$Y_i \text{ is } Bi\left(s, \frac{\beta_1\theta_i}{1+\beta_1\theta_i}\right) \text{ if treatment } T_1 \text{ is used ;}$$

$$Y_i \text{ is } Bi\left(s, \frac{\beta_2\theta_i}{1+\beta_2\theta_i}\right) \text{ if treatment } T_2 \text{ is used .}$$

Here, the odds ratios are θ_i for X_i and $\beta_1\theta_i$ or $\beta_2\theta_i$ for Y_i. Details of the predictive distributions for the binomial and exponential cases can be found in Magalhães [9].

ACKNOWLEDGMENTS

The authors acknowledge the support for F.J.M. Magalhães from Instituto Politécnico do Porto and Universidade Portucalense, Portugal, during his period of postgraduate studies at the University of Sheffield.

REFERENCES

[1] Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian Theory*, Wiley, Chichester.

[2] Dunsmore, I.R. and Robson, D.J. (1997). Prediction in some Poisson errors in variables models, *Scand. J. Statist.*, **24**, 543–554.

[3] Gaver, D.P. and O'Muircheartaigh, I.G. (1987). Robust empirical Bayes analyses of event rates, *Technometrics*, **29**, 1–15.

[4] Gelfand, A.E.; Hills, S.E.; Racine-Poon, A. and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling, *J. Amer. Statist. Ass.*, **85**, 972–985.

[5] Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities, *J. Amer. Statist. Ass.*, **85**, 398–409.

[6] Godambe, V.P. and Kunte, S. (1993). Optimal estimation under biased allocation of treatments, *Biometrika*, **80**, 797–806.

[7] Godambe, V.P. and Thompson, M.E. (1989). An extension of quasi-likelihood estimation (with discussion), *J. Statist. Plan. Inf.*, **22**, 137–172.

[8] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Ann. Math. Statist.*, **27**, 887–906.

[9] Magalhães, F.J.M. (1997). *Prediction in some errors in variables in variables models*, Unpublished Ph.D. Thesis, University of Sheffield.

[10] Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations, *Econometrica*, **16**, 1–32.

[11] O'Hagan, A. (1994). *Bayesian Inference: Kendall's Advanced Theory of Statistics, Volume 2B.*, Edward Arnold, London.

[12] Robbins, H. (1988). *The u, v method of estimation.* In "Statistical Decision Theory and Related Topics IV, **1**" (S.S. Gupta and J.O. Berger, Eds.), Springer-Verlag, New York, 265–270.

[13] Robbins, H. and Zhang, Cun-Hui (1988). Estimating a treatment effect under biased sampling, *Proc. Natl. Acad. Sci. USA*, **85**, 3670–3672.

[14] Robbins, H. and Zhang, Cun-Hui (1989). Estimating the superiority of a drug to a placebo when all and only those patients at risk are treated with the drug, *Proc. Natl. Acad. Sci. USA*, **86**, 3003–3005.

[15] Robbins, H. and Zhang, Cun-Hui (1991). Estimating a multiplicative treatment effect under biased allocation, *Biometrika*, **78**, 349–354.