Statistica Sinica **10**(2000), 647-657

HIERARCHICAL BAYESIAN ANALYSIS OF BINARY MATCHED PAIRS DATA

Malay Ghosh*, Ming-Hui Chen, Atalanta Ghosh and Alan Agresti*

* *University of Florida, Worcester Polytechnic Institute and Lilly Corporate Center*

Abstract: The paper introduces a hierarchical Bayesian analysis of binary matched pairs data with noninformative prior distributions. Certain properties of the posterior distributions, including their propriety, are established. The Bayesian methods are implemented via Markov chain Monte Carlo integration techniques, and numerical illustrations are provided. For the logit link, the conditional and marginal maximum likelihood estimators of a treatment effect depend only on the off-main-diagonal elements of a 2×2 contingency table, and the same is true of McNemar's test. By contrast, the hierarchical Bayes estimators and subsequent analyses depend also on the main-diagonal elements in a natural way.

Key words and phrases: Conditional maximum likelihood estimator, Gibbs sampling, improper prior distribution, logit model, log-log model, Markov chain Monte Carlo, McNemar test, probit model.

1. Introduction

In many studies, especially in biomedical applications, data are collected from *matched pairs*. For example, in case-control studies, cases may be matched with controls on the basis of health or demographic characteristics. Both elements of the matched pairs sometimes refer to the same subject, such as measurements on the left and the right eyes or at two time points or for two treatments compared with a crossover experiment. For concreteness, we refer to the two components of the matched pair as treatments.

This paper considers the special case of a binary response. This response may depend not only on the treatment administered but also on a pair effect. Our analysis of such matched-pairs data is handled within the framework of one-parameter *item response* models. Let X_{ij} denote the binary response, say 0 or 1, of the jth observation within the ith pair, with $p_{ij} = P(X_{ij} = 1)$, $i = 1, \ldots, n$, $j = 1, 2$. The $\{p_{ij}\}$ are modeled as $p_{ij} = F(\theta_i + \alpha_j)$, where F is a distribution function. In this formulation the parameter θ_i for pair i is usually a nuisance parameter, while α_j represents the effect of the jth treatment. Three common choices of F are the standard logistic, normal, and extreme value distribution functions, the corresponding F^{-1} being the *logit*, *probit*, and *log-log* link functions. The parameter of interest is the treatment difference $\alpha_1 - \alpha_2$.

To our knowledge, Cox (1958) first introduced this model, with logistic F. This is a special case of the celebrated Rasch model (Rasch (1961)). Andersen (1973) showed that in this case the maximum likelihood estimator (MLE) of $\alpha_1 - \alpha_2$ is inconsistent. This is the famous Neyman-Scott phenomenon in which the number of nuisance parameters grows in direct proportion to the sample size. Various adjustments of the likelihood can attack this inconsistency problem, such as *modified profile likelihood* (Barndroff-Nielsen (1983)), *conditional profile likelihood* (Cox and Reid (1987)), and *adjusted profile likelihood* (McCullagh and Tibshirani (1990)). It is not clear, however, whether the resulting MLE of $\alpha_1 - \alpha_2$ based on these adjusted likelihoods is indeed consistent.

Andersen (1970, 1973) showed that for logistic F, if one finds the likelihood of $\alpha_1 - \alpha_2$ conditional on $T_i = X_{i1} + X_{i2}$ $(i = 1, \ldots, n)$, the sufficient statistics for the nuisance parameters, then the resulting *conditional* MLE of $\alpha_1 - \alpha_2$ is consistent. Breslow and Day (1980) provided an excellent account of the analysis of binary matched-pairs data via conditional logistic regression. However, for other links, sufficient statistics for the nuisance parameters do not exist. The usual approach in such cases is to treat the model as a mixed model, with random effects $\{\theta_i\}$. The method assigns a distribution function to $\{\theta_i\}$ and then integrates them out; the resulting likelihood for (α_1, α_2), referred to as the marginal likelihood (e.g., Kalbfleisch and Sprott (1970)), is then maximized to find the marginal MLE of $\alpha_1 - \alpha_2$.

Suppose that F is logistic and the $\{\theta_i\}$ are *i.i.d.* $N(\mu, \sigma^2)$. Then, from Neuhaus, Kalbfleisch and Hauck (1994), the marginal MLE of $\alpha_1 - \alpha_2$ is identical with its conditional MLE whenever the sample association between responses displays a nonnegative log odds ratio. Moreover, Neuhaus, Kalbfleisch and Hauck (1994) showed that this result applies for any continuous distribution for $\{\theta_i\}$, when a version of that distribution exists that is consistent with the sample data (i.e., when the model is saturated for the observed data). Let $n_{rs} = \sum_{i=1}^{n} I_{[x_{i1}=r, x_{i2}=s]}$ $(r = 0, 1; s = 0, 1)$. Then, this common estimate of $\alpha_1 - \alpha_2$ is $\log(n_{10}/n_{01})$.

Consider Table 1, for instance, based on the General Social Survey of 1989 in the United States. Subjects were asked whether government spending should increase or decrease on health spending and on law enforcement spending. The two responses for each subject form a matched pair. For the model with logit link, the estimated effect is $\log(25/14) = .580$. For each subject, the estimated odds of responding "increase" on health spending are $\exp(.580) = 1.79$ times the estimated odds of responding "increase" on law enforcement. With this analysis for the logit model, the estimate of the treatment effect and subsequent inference does not depend on n_{00} and n_{11}. Matched-pairs data usually display a positive association, with the majority of the observations falling in these two cells. In Table 1, for instance, 301 of the 340 observations make no contribution to the analysis. The same remark applies to McNemar's test of equality of matched proportions. It is a chi-squared approximation for the probability that a binomial random variable with 14+25=39 trials and parameter .5 takes value of at most 14 or at least 25. The exact two-sided binomial P-value equals .108.

Table 1. Opinions about government spending.

Health Spending	Law Enforcement Spending	
	Decrease	Increase
Decrease	9	14
Increase	25	292

The marginal ML approach is essentially an empirical Bayes (EB) approach. While the EB method is usually suitable for point estimation, it often leads to underestimation of the associated standard error due to failure to account for the uncertainty in σ^2. We propose instead a hierarchical Bayes (HB) procedure that assigns *i.i.d.* $N(0, \sigma^2)$ prior distributions for $\{\theta_i\}$ and an inverse gamma prior distribution for σ^2 at the second stage. This amounts to a multivariate t-prior distribution for $\{\theta_i\}$, which is known to be more robust than a multivariate normal prior distribution. Throughout, flat prior distributions are used for (α_1, α_2), independent of the prior distributions for $\{\theta_i\}$ and also independent among themselves. With this choice of prior distributions, the propriety of the joint posterior distribution for α_1, α_2, $\theta = (\theta_1, \ldots, \theta_n)$ is established for the logit, probit, and log-log links whenever $n_{01} \geq 1$ and $n_{10} \geq 1$.

Section 3 discusses implementation of the hierarchical Bayes procedure via Markov chain Monte Carlo (MC^2). Hierarchical Bayes estimators of $\alpha_1 - \alpha_2$ are compared with the marginal MLE's. Also, we study the sensitivity of the Bayesian analysis with respect to the choice of the parameters of the inverse gamma prior distribution of σ^2.

As mentioned above, for the logit link the traditional analyses ignore the main-diagonal counts of pairs making the same response for each component. The lack of contribution of these counts is counterintuitive to many data analysts and has been the basis of discussion at least back to Cochran (1950). By contrast, these main-diagonal entries influence the HB estimates in a very natural way, the estimated treatment effect diminishing as n_{00} or n_{11} increases.

2. The Hierarchical Bayes (HB) Approach

2.1. The model, priors and the propriety of posteriors

We first express the likelihood function of $\theta = (\theta_1, \ldots, \theta_n)$, α_1 and α_2 as

$$L(\theta, \alpha_1, \alpha_2) = \prod_{i=1}^{n} \prod_{j=1}^{2} \Big[F^{x_{ij}}(\theta_i + \alpha_j)\bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)\Big],$$

where $\bar{F} = 1 - F$. For prior distributions for θ and $\alpha = (\alpha_1, \alpha_2)$, marginally we assume that θ and α are mutually independent, with α uniform on R^2 and (conditional on σ^2) $\{\theta_i\}$ *i.i.d.* $N(0, \sigma^2)$. At the second stage, we assume that σ^2 has the inverse gamma probability density function (pdf)

$$\pi_2(\sigma^2) \propto (\sigma^2)^{-\frac{1}{2}l-1} \exp\Big(-\frac{a}{2\sigma^2}\Big), \; a > 0,$$

denoted by $\mathcal{IG}(\frac{1}{2}a, \frac{1}{2}l)$. This prior distribution is proper when $a > 0$ and $l > 0$, though later we also use improper prior distributions with $l = 0$.

Integrating with respect to σ^2, θ has the marginal multivariate t-prior distribution given by

$$\pi_1(\theta) \propto \Big(a + \sum_{i=1}^{n} \theta_i^2\Big)^{-\frac{1}{2}l}.$$

Letting x denote the vector of sample observations, the joint posterior distribution of θ and α is

$$\pi(\theta, \alpha|x) \propto L(\theta, \alpha)\Big(a + \sum_{i=1}^{n} \theta_i^2\Big)^{-\frac{1}{2}l}.$$

The following theorem provides a sufficient condition for the propriety of this posterior distribution, for the logit, probit, and log-log links.

Theorem 1. *Suppose* $n_{10} \geq 1$ *and* $n_{01} \geq 1$. *Then, for the logit, probit, and log-log links,* $\pi(\theta, \alpha|x)$ *is a proper pdf.*

The proof of the theorem is technical and is omitted. The reader is referred to Ghosh, Chen, Ghosh and Agresti (1997) for the details.

Remark 1. There is no loss of generality in choosing the prior means to be zero since the posterior distribution of $\alpha_1 - \alpha_2$ is invariant under the choice of this mean. Indeed, the result holds in a more general framework as we now see.

Suppose that, given σ^2, $\theta_1, \ldots, \theta_n$ are *i.i.d.*, and their common marginal *pdf* is

$$\pi(\theta) = \int_0^\infty \frac{1}{\sigma} g\left(\frac{\theta - \mu_0}{\sigma}\right) h(\sigma^2) d\sigma^2,$$

where g and h are *pdf*'s, and μ_0 is known. (In our case g is standard normal and h is inverse gamma.) Then with the reparameterization $\phi_i = \theta_i - \mu_0$ and $\beta_j = \alpha_j + \mu_0$ $(i = 1, \ldots, n;\ j = 1, 2)$, (θ, α) is one-to-one with (ϕ, β), where $\phi = (\phi_1, \ldots, \phi_n)$ and $\beta = (\beta_1, \beta_2)$. Also, the posterior distribution of β is

$$\pi(\beta|x) \propto \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L(\phi, \beta) \Big[\int_0^{\infty} (\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \phi_i^2 \Big) h(\sigma^2) d\sigma^2 \Big] d\phi.$$

Hence, $\alpha_1 - \alpha_2 = \beta_1 - \beta_2$ has the same posterior distribution regardless of μ_0.

One may wonder at this point about the possibility of using a prior distribution as above with μ_0 replaced by an unknown μ, then using a uniform (over the real line) prior distribution for μ independent of σ^2 at the second stage. However, the resulting posterior distribution is improper. The details are available in Ghosh, Chen, Ghosh and Agresti (1997).

Remark 2. For the logit link, the condition that $n_{10} \geq 1$ and $n_{01} \geq 1$ is also necessary and sufficient for the conditional MLE of $\alpha_1 - \alpha_2$, $\log(n_{10}/n_{01})$, to remain finite.

2.2. Properties of posteriors

Next we prove results that describe, as a function of $n_{10} - n_{01}$, the behavior of the posterior mean of $\alpha_1 - \alpha_2$ and the posterior probability that treatment 1 is better than treatment 2. The results hold for an arbitrary link F.

Theorem 2. *Consider the prior distribution* $\pi(\theta, \alpha, \sigma^2) \propto \sigma^{-n} [\prod_{i=1}^{n} g(\theta_i/\sigma)]$ $h(\sigma^2)$, *and let* $\rho = \alpha_1 - \alpha_2$.

(a) *For every* $t \geq 0$, $P(\rho > t|x) - P(\rho < -t|x)$ *has the same sign as* $n_{10} - n_{01}$;

(b) $P(\rho > 0|x) \gtreqless 1/2$ *according as* $n_{10} \gtreqless n_{01}$;

(c) $E(\rho|x)$ *has the same sign as* $n_{10} - n_{01}$.

Proof. With the transformation $u = \theta_i + \frac{1}{2}(\alpha_1 + \alpha_2)$, $\rho = \alpha_1 - \alpha_2$, and $\xi = \frac{1}{2}(\alpha_1 + \alpha_2)$, the posterior distribution of ρ is given by

$$\pi(\rho|x) = k(x) \int_0^{\infty} \int_{-\infty}^{\infty} a_{11}^{n_{11}}(\rho, \xi, \sigma) a_{00}^{n_{00}}(\rho, \xi, \sigma) a_{10}^{n_{10}}(\rho, \xi, \sigma) a_{01}^{n_{01}}(\rho, \xi, \sigma) \sigma^{-n} h(\sigma^2) d\xi d\sigma^2,$$

where $k(x)$ is the normalizing constant, and

$$a_{11}(\rho, \xi, \sigma) = \int_{-\infty}^{\infty} F(u + \frac{1}{2}\rho) F(u - \frac{1}{2}\rho) g\left(\frac{u - \xi}{\sigma}\right) du,$$

$$a_{00}(\rho, \xi, \sigma) = \int_{-\infty}^{\infty} \left[1 - F(u + \frac{1}{2}\rho)\right] \left[1 - F(u - \frac{1}{2}\rho)\right] g\left(\frac{u - \xi}{\sigma}\right) du,$$

$$a_{10}(\rho,\xi,\sigma)=\int_{-\infty}^{\infty}F(u+\frac{1}{2}\rho)\left[1-F(u-\frac{1}{2}\rho)\right]g\left(\frac{u-\xi}{\sigma}\right)du,$$
$$a_{01}(\rho,\xi,\sigma)=\int_{-\infty}^{\infty}\left[1-F(u+\frac{1}{2}\rho)\right]F(u-\frac{1}{2}\rho)g\left(\frac{u-\xi}{\sigma}\right)du.$$

This implies that

$$\pi(\rho|x)-\pi(-\rho|x)=k(x)\int_0^{\infty}\int_{-\infty}^{\infty}a_{11}^{n_{11}}(\rho,\xi,\sigma)a_{00}^{n_{00}}(\rho,\xi,\sigma)$$
$$\times\left[a_{10}^{n_{10}}(\rho,\xi,\sigma)a_{01}^{n_{01}}(\rho,\xi,\sigma)-a_{01}^{n_{10}}(\rho,\xi,\sigma)a_{10}^{n_{01}}(\rho,\xi,\sigma)\right]\sigma^{-n}h(\sigma^2)d\xi d\sigma^2. \quad (2.1)$$

(a) Now for $\rho>0$, $a_{10}(\rho,\xi,\sigma)>a_{01}(\rho,\xi,\sigma)$, which implies that $a_{10}^{n_{10}}(\rho,\xi,\sigma)$ $a_{01}^{n_{01}}(\rho,\xi,\sigma)\gtreqless a_{01}^{n_{10}}(\rho,\xi,\sigma)\ a_{10}^{n_{01}}(\rho,\xi,\sigma)$ according as $n_{10}\gtreqless n_{01}$. This proves that for $\rho>0$, $\pi(\rho|x)-\pi(-\rho|x)$ has the same sign as $n_{10}-n_{01}$. Then, since

$$P(\rho>t|x)-P(\rho<-t|x)=\int_t^{\infty}\pi(\rho|x)d\rho-\int_{-\infty}^{-t}\pi(\rho|x)d\rho$$
$$=\int_t^{\infty}[\pi(\rho|x)-\pi(-\rho|x)]d\rho,$$

result (a) follows.

(b) This follows from (a) by putting $t=0$.

(c) Write

$$E(\rho|x)=\int_0^{\infty}P(\rho>t|x)dt-\int_{-\infty}^{0}P(\rho<t|x)dt$$
$$=\int_0^{\infty}[P(\rho>t|x)-P(\rho<-t|x)]dt.$$

The result is now a consequence of (a).

Remark 3. The above result is intuitive. When $n_{10}=n_{01}$, the posterior distribution of $\rho=\alpha_1-\alpha_2$ is symmetric about 0, and the natural Bayes estimator of ρ equals 0. Using *i.i.d.* prior distributions $g\left(\frac{\theta-\xi}{\sigma}\right)$ for the $\{\theta_i\}$, one can see that the marginal likelihood $L(\rho,\xi,\sigma)$ then satisfies $L(\rho,\xi,\sigma)=L(-\rho,\xi,\sigma)$. If for fixed ξ and σ, $\log L(\rho,\xi,\sigma)$ is concave in ρ, then the MMLE of ρ also equals 0. Other than this special case, the MMLE and the posterior Bayes estimate usually do not agree. For the logit link, for instance, if $(n_{00}n_{11}/n_{01}n_{10})\geq 1$, the MMLE of ρ equals the conditional MLE and does not depend on n_{00} and n_{11}. The Bayes estimate of ρ, on the other hand, depends on all four cell counts.

It is also true in our Bayesian model that for fixed n_{10} and n_{01}, the magnitude of the difference in the treatment effects decreases in n_{00} and n_{11}. The result is intuitively reasonable; as the number of agreements increases, for a fixed number of disagreements, one naturally feels that the difference diminishes. The next theorem provides details.

Theorem 3. *Consider the prior distribution given in Theorem 2. Then for fixed n_{10} and n_{01},* (a) $P(|\rho| > t|x)$, (b) $E(|\rho|\ |x)$, *and* (c) $|P(\rho > 0|x) - 1/2|$ *are decreasing in n_{00} and n_{11}.*

Proof.

(a) For every $t \geq 0$,

$$\begin{aligned}P(|\rho| > t|x) &= \int_t^\infty \pi(\rho|x)d\rho + \int_{-\infty}^{-t} \pi(\rho|x)d\rho\\ &= \int_t^\infty [\pi(\rho|x) + \pi(-\rho|x)]d\rho\\ &= k(x)\int_t^\infty \int_0^\infty \int_{-\infty}^\infty \left[\sigma^{-1}a_{11}(\rho,\xi,\sigma)\right]^{n_{11}} \left[\sigma^{-1}a_{00}(\rho,\xi,\sigma)\right]^{n_{00}}\\ &\quad \times \Big\{\left[\sigma^{-1}a_{10}(\rho,\xi,\sigma)\right]^{n_{10}} \left[\sigma^{-1}a_{01}(\rho,\xi,\sigma)\right]^{n_{01}}\\ &\quad + \left[\sigma^{-1}a_{01}(\rho,\xi,\sigma)\right]^{n_{10}}\left[\sigma^{-1}a_{10}(\rho,\xi,\sigma)\right]^{n_{01}}\Big\}\, h(\sigma^2)d\xi d\sigma^2 d\rho.\end{aligned}$$

Since $\sigma^{-1}a_{rs}(\rho,\xi,\sigma) < \int_{-\infty}^{\infty} \sigma^{-1}g\left(\frac{u-\xi}{\sigma}\right) du = 1$ for all $r, s = 0, 1$, it follows that for fixed n_{10} and n_{01}, $P(|\rho| > t\ |x)$ decreases in n_{11} and in n_{00} for every $t \geq 0$.

(b) This follows from (a) by writing

$$E(|\rho||x) = \int_0^\infty P(|\rho| > t|x)dt.$$

(c) Note that

$$P(\rho > 0|x) - 1/2 = (1/2)\int_0^\infty [\pi(\rho|x) - \pi(-\rho|x)]d\rho. \tag{2.2}$$

From the proof of (a) in Theorem 1, $\pi(\rho|x) - \pi(-\rho|x)$ has the same sign for all $\rho > 0$. Using (2.1) for each ρ and the fact that $\sigma^{-1}a_{rr}(\rho,\xi,\sigma) < 1$ for r = 0,1 implies that the right-hand side of (2.2) is decreasing in absolute value as n_{00} or n_{11} increases.

3. Implementation and Illustration of the HB Method

The marginal posterior distribution of $\alpha_1 - \alpha_2$ is analytically intractable. However, it is easily calculated using the Gibbs sampling numerical integration technique. To this end, when we use the multivariate t-prior distribution for θ, with the standard parameter augmentation technique,

$$\pi(\theta_i \mid \theta_l\ (l \neq i), \alpha_1, \alpha_2, \sigma^2, x)$$

$$\propto \prod_{j=1}^{2} \left[F^{x_{ij}}(\theta_i + \alpha_j)\bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)\right] \exp\left(-\frac{1}{2\sigma^2}\theta_i^2\right), \quad (i = 1, \ldots, n); \quad (3.1)$$

$$\pi(\alpha_j \mid \alpha_k \ (k \neq j), \theta, \sigma^2, x) \propto \prod_{i=1}^{n} \left[F^{x_{ij}}(\theta_i + \alpha_j)\bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)\right], \ j = 1, 2; \quad (3.2)$$

$$\pi(\sigma^2 \mid \theta, \alpha_1, \alpha_2, x) \propto (\sigma^2)^{-\frac{n+l}{2}-1} \exp\left[-\frac{1}{2\sigma^2}\left(a + \sum_{i=1}^{n} \theta_i^2\right)\right]. \quad (3.3)$$

It is simple to simulate from the inverse gamma density (3.3). The conditionals (3.1) and (3.2) are not standard densities, but if F and $\bar{F}$ are both log-concave (which is true when F is increasing failure rate), they are log-concave. Then, one can use the *adaptive rejection sampling algorithm* of Gilks and Wild (1992) to generate samples from these full conditionals. To construct the posterior distribution of $\alpha_1 - \alpha_2$, it is convenient to use the one-to-one reparameterization from $(\theta, \alpha_1, \alpha_2)$ to (u, ρ, ξ).

Table 2. Characteristics of the posterior distribution of $|\rho|$, as a function of the main-diagonal counts, when $n_{10} = n_{01} = 5$.

(n_{00}, n_{11})	$E(\|\rho\| \mid x)$	$[V(\|\rho\| \mid x)]^{1/2}$	50th Percentile	90th Percentile	95th Percentile
Prior $l = 3,\ a = 10$					
(10, 10)	.531	.416	.448	.758	1.327
(15, 45)	.526	.400	.416	.712	1.274
(50, 50)	.492	.387	.412	.700	1.241
Prior $l = 5,\ a = 3$					
(10, 10)	.483	.374	.395	.666	1.165
(15, 45)	.470	.369	.383	.653	1.160
(50, 50)	.458	.360	.369	.629	1.095

We now numerically illustrate some results of the previous section for the logit link. Similar results occur for the probit and log-log links. We first illustrate Theorem 3 on characteristics of the posterior distribution of $|\rho|$. We fix $n_{10} = n_{01} = 5$ and choose two quite different prior distributions: one very diffuse ($l = 3, a = 10$, for which $E(\sigma^2) = 10$ and $V(\sigma^2) = \infty$), and one considerably more informative ($l = 5, a = 3$, for which $E(\sigma^2) = 1$ and $V(\sigma^2) = 2$). Since

$n_{10} = n_{01}$, the posterior distribution of ρ is symmetric about 0. The Gibbs sampling procedure, using every 15th iterate, produced 50,000 observations that showed negligible autocorrelation. Table 2 reports $E(|\rho| \mid x)$, $[V(|\rho| \mid x)]^{1/2}$, and various percentiles of the distribution of $|\rho|$ for these prior distributions and for various choices of n_{00} and n_{11}. The reported value of $E(|\rho| \mid x)$ has a simulation standard error of about .002. The table shows that $E(|\rho| \mid x)$ and the percentiles decrease as n_{00} and n_{11} increase. The main diagonal counts have an effect, and there is a substantive weakening of the estimated effect when those counts increase dramatically. The same table also provides the 50th, 90th and 95th percentiles of the distribution of $|\rho|$ for various choices of n_{00} and n_{11}. Since the posterior distribution of ρ is symmetric about 0, these are also the 75th, 95th and 97.5th percentiles of the posterior distribution of ρ.

Table 3. Characteristics of the posterior distribution of the treatment effect, as a function of the prior distribution, for Table 1.

Prior Parameters	$E(\rho\|x)$	$[V(\rho\|x)]^{1/2}$	$P(\rho < 0\|x)$	95% HPD Interval
1. $l = 5$, $a = 3$	.481	.293	.047	$(-.083, 1.064)$
2. $l = 4.0001$, $a = 2.0001$	.477	.292	.045	$(-.093, 1.054)$
2a. (2) with $n_{00} = 3, n_{11} = 117$	.639	.355	.027	$(-.032, 1.360)$
2b. (2) with $n_{00} = 27, n_{11} = 876$	.223	.200	.132	$(-.165, 0.615)$
3. $l = 0$, $a = 0.1$	.468	.289	.049	$(-.089, 1.045)$
4. $l = 0$, $a = 0.01$	.477	.281	.047	$(-.111, 1.051)$
5. $l = 0$, $a = 0.001$	.441	.277	.053	$(-.103, 0.987)$
6. Conditional, Marginal MLE	.580	.334	.054	$(-.074, 1.234)$

Table 3 shows results of HB analyses for various prior distributions for Table 1 on government spending. We consider whether there is any difference in the opinion distributions for the two types of spending and whether the conclusion is sensitive to the choice of prior distributions. Here, $E(\rho|x) > 0$, since $n_{10} - n_{01} > 0$. Table 3 provides $E(\rho|x)$, $[V(\rho|x)]^{1/2}$, $P(\rho < 0|x)$, and the 95% HPD interval for ρ for various choices of l and a for the prior distribution. The simulation standard error is about .003 for the reported posterior means.

The first two inverse gamma prior distributions for σ^2 in Table 3 have $E(\sigma^2) = 1$, but $V(\sigma^2) = 2$ in the first case and $V(\sigma^2) = 20,000$ in the second. The three prior distributions with $l = 0$ are improper. These prior distributions, with a close to 0, are close to $\pi(\sigma^2) \propto (\sigma^2)^{-1}$. The Bayes estimator of ρ and

$P(\rho < 0|x)$ are not very sensitive to the choice of prior parameters. An alternative prior structure would be to use independent t prior distributions for the $\{\theta_i\}$, rather than the multivariate t that results from our hierarchical approach. We did this for Table 1 and obtained similar results. For contrast, the table also reports results for the common conditional and marginal MLE, the interval in this case being the standard 95% Wald confidence interval and the entry in the $P(\rho < 0|x)$ column being instead the one-sided P-value. Although the Bayesian analyses do utilize the main diagonal counts, for these data the substantive results do not differ from those using standard methods.

To illustrate the potential effect of the main-diagonal counts on the Bayesian analysis, we used the second prior distribution ($l = 4.0, a = 2.0$) to analyze the data with two possible changes in those counts. The first case has smaller main diagonal counts ($n_{00} = 3, n_{11} = 117$) for which the odds ratio between the two responses is about 1.0, rather than the observed value of 7.5. The posterior mean is now substantially larger (.639 instead of .477), and $P(\rho < 0|x)$ is smaller (.027 instead of .045). The second case has main diagonal counts ($n_{00} = 27, n_{11} = 876$) that are triple the originals, corresponding to an odds ratio of 67.6. The posterior mean is now less than half the original value (.223 instead of .477), and the evidence of an effect is much smaller ($P(\rho < 0|x) =$ .132 instead of .045). The main diagonal counts can have considerable impact on the results (unlike in the traditional analyses), with increases in main-diagonal counts resulting in diminished estimated effects. It may be noted also that with $n_{00} = 27, n_{11} = 876$, for testing $H_0 : \rho \leq 0$ against $H_1 : \rho > 0$, while the posterior odds ratio increases significantly from .047 to .152, the one-sided P-value for McNemar's test remains the same at .054.

In summary, the hierarchical Bayesian analysis for matched pairs differs from the conditional ML and marginal ML approaches in using information on the number of pairs making the same response for each treatment. One might expect that this would be true of any Bayesian analysis. However, for a Bayesian model with Dirichlet prior distribution for the four multinomial counts $\{n_{rs}\}$, Altham (1971) showed that n_{00} and n_{11} affect the posterior distribution but not $P(\rho > 0|x)$.

Acknowledgements

Thanks are due to two referees for their very careful reading of the manuscript, and for many helpful suggestions. Ghosh's research was partially supported by the National Science Foundation under Grant Numbers SBR-9423996 and SBR-9810968. Chen's research was partially supported by the National Science Foundation under Grant Number DMS-9702172, and Agresti's research was partially supported by a grant from the National Institutes of Health.

References

Altham, P. M. E. (1971). The analysis of matched proportions. *Biometrika* **58**, 561-576.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *J. Roy. Statist. Soc. Ser. B* **32**, 283-301.

Andersen, E. B. (1973). Conditional inference for multiple-choice questionnaires. *British J. Math. Statist. Psych.* **26**, 31-41.

Barndorff-Nielsen, O. E. (1983). On a formula for the conditional distribution of the maximum likelihood estimator. *Biometrika* **70**, 343-365.

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research, Volume 1: The Analysis of Case-Control Studies.* International Agency for Research on Cancer, Lyon.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika* **37**, 256-266.

Cox, D. R. (1958). Two further applications of a model for binary regression. *Biometrika* **45**, 562-565.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **49**, 1-39.

Ghosh, M, Chen, M.-H, Ghosh, A. and Agresti, A. (1997). Hierarchical Bayesian analysis of binary matched pairs data. Technical Report No. 553, Department of Statistics, University of Florida.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41**, 337-348.

Kalbfleisch, J. D. and Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters (with discussion). *J. Roy. Statist. Soc. Ser. B* **32**, 175-208.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc. Ser. B* **52**, 325-344.

Neuhaus, J. M., Kalbfleisch, J. D. and Hauck, W. W. (1994). Conditions for consistent estimation in mixed-effects models for binary matched-pairs data. *Canad. J. Statist.* **22**, 139-148.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics* **4**, 321-334. University of California Press, Berkeley.

Department of Statistics, University of Florida, Gainesville, Florida 32611-8545, U.S.A.
E-mail: ghoshm@stat.ufl.edu

Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609-2280, U.S.A.
E-mail: mhchen@wpi.edu

Eli Lilly and Company, Lilly Corporate Center, Indianapolis, Indiana 46285-2233, U.S.A.
E-mail: ghoshatalanta@lilly.com

Department of Statistics, University of Florida, Gainesville, Florida 32611-8545, U.S.A.
E-mail: aa@stat.ufl.edu

(Received January 1998; accepted June 1999)