South African Statist. J. (2009) **43**, *195–218*

AN ADAPTIVE BAYESIAN ANALYSIS FOR BINOMIAL PROPORTIONS

Sonali Das[1]
Logistics and Quantitative Methods, CSIR Built Environment, P.O. Box 395,
Pretoria 0001, South Africa
E-mail: *sdas@csir.co.za*
and
Sourish Das
Statistical and Applied Mathematical Sciences Institute and Department of
Statistical Science, Duke University, Durham, NC, USA
E-mail: *sourish.das@stat.duke.edu*

Key words: Adaptive Bayes; Bayesian central limit theorem; Monte Carlo method; proportion; statistical power; ϕ-divergence.

Summary: We consider the problem of statistical inference of binomial proportions for non-matched, correlated samples, under the Bayesian framework. Such inference can arise when the same group is observed at a different number of times with the aim of testing the proportion of some trait. For example, say, we are interested to infer about the effectiveness of a certain intervention teaching strategy, by comparing proportion of 'proficient' teachers, before and after an intervention. The number of teachers may differ between the two measurement time points, due to any number of reasons, and thus can result in an unequal number of observations in two periods. For such nonmatched design, we develop an adaptive Bayesian method, and suggest a heuristic decision procedure to conduct statistical inference. We use the ϕ-divergence measure to quantify the perturbation of the posterior distribution of the proportion in different time points. We present a simulation study to compare the statistical power between the adaptive Bayesian method and the existing frequentist method. Our study and theoretical results indicate that under certain design, the adaptive Bayesian method outperforms the existing

[1] Corresponding author

AMS: 62F15

method. We administer the developed adaptive Bayesian method to two case studies.

1. Introduction

Inference on the probability of 'success' of the binomial distribution is one of the most important problems in statistical literature. Popular applications involve inferring the equality of proportions of success from two samples. Such samples can be either independent or matched pairs, over the two time points. A third scenario can arise when for the same group, the number of measurements is different at different time points. In such a scenario, the samples are not independent, but neither do they satisfy the matched pair criteria. As such, usual inference procedures for the proportion cannot be conducted under the independent sample assumptions, nor under the matched pair assumption. The authors have come across a number of such real case studies, and this work arises from a need to address this inference problem.

We briefly recapitulate some popular inference methods for binomial proportions. In the simplest case of two samples, two textbook scenarios for inferring the hypothesis of equality of binomial proportions occur when the samples are independent, or are matched. When the samples are independent, the Cochran-Mantel-Hansel type χ^2tests for equality of proportion are commonly used (Cochran, 1954, Mantel & Haenszel, 1959, Agresti and Caffo, 2000). For matched pairs, the McNemar (1947) χ^2 test, or the exact test by Liddell (1983) is available. To test for the difference of proportions for matched binary pairs, Lloyd (1990) proposed two methods to construct confidence intervals. In the Bayesian framework, Chen and Dey (1998) developed a model for correlated binary responses using a scale mixture of multivariate normal link functions, and implemented the model using the Markov Chain Monte Carlo (MCMC) method. Ghosh *et al.* (2000) developed a Bayesian hierarchical technique for binary matched pairs data, while Ghosh and Chen (2002) developed a general technique for inferring in a matched case-control setup. Berger (1985:457) provides a decision theoretic Bayesian stopping rule under a sequential analysis framework for binomial proportion.

In this paper, we develop an adaptive Bayesian method for binomial proportions, when neither assumptions of independence, nor matched-pair holds. The paper is organised as follows: In Section 2, we develop an adaptive Bayesian method for non-matched correlated binomial proportions under various non-informative conjugate Beta priors, as well as the Zellner's non-informative prior. In Section 3, we evaluate the distance between posteriors at two time points using the proposed method. In Section 4, we investigate the statistical power of the developed method. In Section 5, we present two case studies for inferring on non-matched correlated proportions using our method, and in Section 6, we conclude the paper.

2. An adaptive Bayesian method for binomial proportions

We first present two motivating case studies to motivate the development of the adaptive Bayesian method for binomial proportions. The first case study involves the inference on the proportion of a particular group of teachers rated into two categories – 'proficient' and 'non-proficient'. Teachers were observed and rated at two instances, before and after an intervention. At each instance, they were categorised as either 'proficient' or 'non-proficient', both before and after the intervention. The second case study is motivated by Webster *et al.* (2005), in which they are interested to look for indications of change in the proportion of category '4 and 5' hurricanes, in the Atlantic Ocean, during two time periods. In each period, there were a different number of hurricanes recorded, each of which could be a category '4 and 5' or lower. In both cases, the sample sizes were not the same at the different time points. Clearly, regular matched pairs analysis of proportions is not applicable here. Also, regular analysis of binomial proportions for more than one sample is not applicable as it does not take care of the time order dependence embedded in the data. With these case studies as background, we proceed to explain the new adaptive Bayesian method for binomial proportions from an information theory point of view.

2.1 An adaptive Bayesian method in the context of information processing

Suppose a process has been observed at k different time points, say $T_1, T_2, ..., T_k$. We observe the process n_i times at time point T_i, of which $s_i\,(\leq n_i)$ are categorised as success for a binary outcome (where $i = 1, 2, ..., k$). If we denote S_i as the random number of successes in time point T_i, then we assume $S_i \sim Binomial(n_i, p_{T_i}),\ i = 1, ..., k$. Hence the natural conjugate prior is $Beta(a_t, b_t)$.

Zellner's Information Processing Rule (ZIPR) (Zellner 1988, 2002) for statistical inference is based on the input and output information measure. A measure is considered 100% efficient if the output information is exactly equal to the input information. In our setup, the information processing between two time points is

$$\begin{aligned}\triangle\left[\widetilde{\pi}^{t}(p|s_t)\right] &= \text{Output information} - \text{Input information}\\ &= \int \widetilde{\pi}^{t}(p)\log\widetilde{\pi}^{t}(p)\,dp + \int\widetilde{\pi}^{t}(p)\log\left(h\left(s_t\right)\right)dp\\ &\quad -\int\widetilde{\pi}^{t}(p)\log f\left(s_t|p\right) - \int\widetilde{\pi}^{t}(p)\log\widetilde{\pi}^{t-1}\left(p|s_{t-1}\right)dp,\end{aligned}$$

where

$$\begin{aligned}h\left(s_t\right) &= \int f\left(s_t|p\right)\pi^{t-1}\left(p|s_t\right)dp\\ &= \begin{pmatrix} n_t\\ s_t\end{pmatrix}\frac{Beta\left(s_t+a_t n_t+b_t-s_t\right)}{Beta\left(a_t,b_t\right)},\end{aligned}$$

with $\widetilde{\pi}^{t}(p)$ being proper probability density at stage/time t, such that $\int\widetilde{\pi}^{t}(p)\,dp=1$.

Observe that $h(s_t)$ is free from the parameter p, and is the likelihood at stage/time t, while $\pi^{t-1}(p)$ is the prior distribution at stage/time t.

$$\begin{aligned}\triangle\left[\widetilde{\pi}^{t}(p|s_t)\right] &= \int\widetilde{\pi}^{t}(p)\log\left(\frac{\widetilde{\pi}^{t}(p)\,h\left(s_t\right)}{f\left(s_t|p\right)\pi^{t-1}(p)}\right)dp\\ &= \int\widetilde{\pi}^{t}(p)\log\left(\frac{\widetilde{\pi}^{t}(p)}{\pi^{t}(p)}\right)dp\end{aligned}$$

where $\pi^{t}(p)=\dfrac{f\left(s_t|p\right)\pi^{t-p}(p)}{h\left(s_t\right)}$ is the posterior distribution of p at stage/time t where we use $\pi^{t}(p)=\pi^{t-1}(p|s_t)$ as the prior. Hence, we are choosing the prior adaptively based on the posterior we observe in the previous stage. By Zellner's criterion, a rule is 100% efficient if $\triangle\left[\widetilde{\pi}^{t}(p|s_t)\right]=0$. Clearly, $\triangle\left[\widetilde{\pi}^{t}(p|s_t)\right]=0$ if $\widetilde{\pi}^{t}(p)=\pi^{t}(p)=\dfrac{f\left(s_t|p\right)\pi^{t-1}(p)}{h\left(s_t\right)}$. We thus summarise the above in the following theorem:

Theorem 2.1. *For binomial proportions, $\widetilde{\pi}^{t}(p)$ yields 100% efficient Zellner's Information Processing Rule (ZIPR) if one uses the posterior of previous stage/time point as the prior for the next stage/time point.*

2.2 Adaptive Bayesian method and Markov transition model

In this section we discuss the relation between adaptive Bayesian method and Markov transition model. Suppose $\pi^{t-1}(p|s_1) \sim Beta(a_{t-1}+s_{t-1},\ n_{t-1}+b_{t-1}-s_{t-1})$and is the posterior at stage $(t-1)$. Motivated from the previous section, we choose $\pi^{t-1}(p|s_1)$ as prior for stage t, i.e., $\pi^{t}(p) = \pi^{t-1}(p|s_1)$ or simply $\pi^{t}(p) \sim Beta(a_t, b_t)$. where $a_t = a_{t-1}+s_{t-1}$ and $b_t = n_{t-1}+b_{t-1}-s_{t-1}$. Therefore, with $S_t \sim Binomial(n_t, p)$, the posterior of p at time point t is

$$\pi^{t}(p|s_t) \sim Beta(a_t+s_t, n_t+b_t-s_t).$$

Observe that

$$\begin{aligned} E^{\pi^t}(p|s_t) &= \frac{a_t+s_t}{n_t+b_t+a_t} \\ &= \omega_t \frac{s_t}{n_t} + (1-\omega_t) E^{\pi^t}(p|s_{t-1},) \end{aligned} \quad (2.1)$$

where $\omega_t = \dfrac{n_t}{n_{t-1}+n_t+a_{t-1}+b_{t-1}}$. Observe, (2.1) has an intuitive form of a weighted average of the observed proportion and the expected proportion from the previous stage. Thus, (2.1) can be re-written as

$$E^{\pi^t}(p|s_t) = \alpha_t + \beta_t E^{\pi^{t-1}}(p|s_{t-1}),$$

where $\alpha_t = \omega_t \dfrac{s_t}{n_t}$ and $\beta_t = (1-\omega_t)$, for $t = 1, 2, ..., k$, which is a Markov transition model. Therefore, the adaptive Bayesian method developed in this paper provides solid justification for the one-step Markov transition process.

2.3 Prior at the initial stage

We now discuss the different choices of prior that we can use at the first stage. Note, if $a_1 = b_1 = 0.5$, we have the Jeffreys's (1961) non-informative prior, while if $a_1 = b_1 = 0$, we have the Novick-Hall's (1965) non-informative prior for binomial proportions.

The Zellner's (1977) non-informative prior in the first time point $\pi^1(p)$ is proportional to $p^p(1-p)^{1-p}$. For Zellner's prior, we do not need any hyper-parameters. The posterior for the second time point is then

$$\begin{aligned}\pi^1(p|s_1) &\propto \binom{n_1}{s_1} p^{s_1}(1-p)^{n_1-s_1} \times p^p(1-p)^{1-p} \\ &\propto p^{s_1+p}(1-p)^{n_1+1-(s-1+p)}.\end{aligned}$$

Following the same adaptive Bayesian idea from Section (2.1), we consider the posterior of the first time point as the prior for the second time point. That is, $\pi^2(p) = \pi^1(p|s_1)$, or equivalently, $\pi^2(p) \propto p^{s_1+p}(1-p)^{n_1+1-(s-1+p)}$. This gives the posterior for the second time point as

$$\pi^2(p|s_2) \propto p^{s_1+s_2+p}(1-p)^{n_1+n_2+1-(s-1+s_2+p)} \tag{2.2}$$

In the next section, we discuss different measures for quantifying the amount of information that propagates from one stage to another stage using this adaptive Bayesian method.

3. Distance between posterior of two time points

According to the Bayesian paradigm, the posterior distribution contains all the relevant information about the parameters. Therefore, any perturbation in the parameters should depend on the discrepancy between the posteriors in the two time points with the lag l. We use the ϕ-divergence measure (Ali and Silvey 1966; Csiszar 1967) to capture this discrepancy. We define the ϕ-divergence in our setup as

$$D_{\phi}^{(t,l)} = D\left(\pi^{t}\left(p|s_{t}\right),\ \pi^{t+l}\left(p|s_{t+l}\right)\right) = \int \phi\left(\frac{\pi^{t+l}\left(p|s_{t+l}\right)}{\pi^{t}\left(p|s_{t}\right)}\right)\times\pi^{t}\left(p|s_{t}\right)dp,$$

where ϕ is a convex function with $\phi(1) = 0$. Several choices of ϕ are given by Dey and Birmiwal (1994). Using the Monte Carlo technique, we can easily implement the Kullback-Leibler divergence measure (Kullback & Leibler, 1951), the χ^2-divergence measure and the Hellinger distance.

In our adaptive Bayesian method, the posterior distribution, under conjugate Beta prior, in the two time points is given by

$$\pi^{t}\left(p|s_{t}\right) \sim Beta\left(\sum_{t=1}^{t} s_i + a_1,\ \sum_{t=1}^{t} n_i + b_1 - \sum_{t=1}^{t} s_i\right),$$

and

$$\pi^{t+l}\left(p|s_{t+l}\right) \sim Beta\left(\sum_{t=1}^{t+l} s_i + a_1,\ \sum_{t=1}^{t+l} n_i + b_1 - \sum_{t=1}^{t+l} s_i\right),$$

We define the Kullback-Leibler divergence measure, under the conjugate prior, for our case as

$$\begin{aligned} D_{KL}^{(t,l)} &= E\left[-\log\frac{\pi^{t+l}\left(p|s_{t+l}\right)}{\pi^{t}\left(p|s_{t}\right)}\right] \\ &= E\left[-\log K^{(t,l)} p^{\sum_{i=t+1}^{t+l} s_i}\left(1-p\right)^{\sum_{i=t+1}^{t+l} n_i \sum_{i=t+1}^{t+l} s_i}\right], \quad (3.3) \end{aligned}$$

where

$$K^{(t,l)} = \frac{Beta\left(\sum_{i=1}^{t} s_i + a_1,\ \sum_{i=1}^{t} n_i + b_1 - \sum_{i=1}^{t} s_i\right)}{Beta\left(\sum_{i=1}^{t+l} s_i + a_1,\ \sum_{i=1}^{t+l} n_i + b_1 - \sum_{i=1}^{t+l} s_i\right)}.$$

The χ^2-divergence, under conjugate prior in our case is

$$D_{\chi^2}^{(t+l)} = E\left[\left(\frac{\pi^{t+l}\left(p|s_{t,\,l}\right)}{\pi^{t}\left(p|s_{t}\right)} - 1\right)^2\right]$$

$$= E\left[\left(K^{(t,l)} p^{\sum_{i=t+1}^{t+l} s_i} (1-p)^{\sum_{i=t+1}^{t+l} n_i - \sum_{i=t+1}^{t+l} s_i} - 1\right)^2\right]. \quad (3.4)$$

The Hellinger distance, under conjugate prior in our case is

$$D_{H}^{(t,l)} = 2 \times \left[1 - E\left(\sqrt{\frac{\pi^{t+l}\left(p|s_{t+l}\right)}{\pi^{t}\left(p|s_{t}\right)}}\right)\right]$$

$$= 2\times\left[1 - E\left(\sqrt{K^{(t,l)} p^{\sum_{i=t+1}^{t+l} s_i} (1-p)^{\sum_{i=t+1}^{t+l} n_i - \sum_{i=t+1}^{t+l} s_i}}\right)\right]. \quad (3.5)$$

The forms of $D_{KL}^{(t,l)}$, $D_{\chi^2}^{(t,l)}$ and $D_{H}^{(t,l)}$ under conjugate prior as given in (3.3), (3.4) and (3.5) are analytically intractable, and thus we calculate these distance measures using the Monte Carlo technique, or via the 'Integral' function of the R software.

Observe that for $t = 1$ and $l = 1$, we have the Kullback-Leibler divergence measure between time point 1 and time point 2 as

$$D_{KL}^{(1,1)} = E\left[-\log \frac{\pi^{2}\left(p|s_{2}\right)}{\pi^{1}\left(p|s_{1}\right)}\right]$$

$$= E\left[-\log\ K^{(1,1)} p^{s_2} \left(1-p\right)^{n_2 - s_2}\right],$$

where

$$K^{(1,1)} = \frac{Beta\left(s_1 + a_1,\ n_1 + b_1 - s_1\right)}{Beta\left(s_1 + s_2 + a_1,\ n_1 + n_2 + b_1 - \left(s_1 + s_2\right)\right)}.$$

Similarly, for $t = 1$ and $l = 2$, we have the Kullback-Leibler divergence measure between time point 1 and time point 3 as

$$\begin{aligned} D_{KL}^{(1,2)} &= E\left[-\log \frac{\pi^3(p|s_3)}{\pi^1(p|s_1)}\right] \\ &= E\left[-\log\ K^{(1,2)} p^{s_2+s_3}(1-p)^{(n_2+n_3)-(s_2+s_3)}\right], \end{aligned}$$

and for $t = 2$ and $l = 1$, we have the Kullback-Leibler divergence measure between time point 2 and time point 3 as

$$\begin{aligned} D_{KL}^{(2,1)} &= E\left[-\log \frac{\pi^3(p|s_3)}{\pi^2(p|s_2)}\right] \\ &= E\left[-\log\ K^{(2,1)} p^{s_3}(1-p)^{n_3+s_2}\right], \end{aligned}$$

where $K^{(1,\,2)}$ and $K^{(2,\,1)}$ are defined accordingly.

4. Statistical power of the adaptive Bayesian method

In this section, we investigate some properties of the developed method. We suggest a heuristic procedure to conclude whether the binomial proportion in the second time point is significantly different from the previous time point. Thus, it is important to verify the long-run behaviour of the Bayesian method, which leads us to investigate the statistical power of the heuristic procedure based on the adaptive Bayesian method, using the Monte Carlo technique.

4.1 Inferences based on the adaptive Bayesian method

Bayesian inference is determined via a loss function and the prior information. Therefore, we start with the loss function as given by Geisser (2006)

$$L(p,\ C) = \begin{cases} \alpha\,\mathcal{L}(C) - A, & \text{if } p \in C, \\ \alpha\,\mathcal{L}(C), & \text{otherwise,} \end{cases}$$

where $A > 0$, and $\mathcal{L}(C)$ is the Lebesgue measure of C. Under this loss function, it can be shown that the Bayes estimator (i.e., the minimum posterior expected loss) is a $100(1-\alpha)\%$ highest posterior density (HPD) interval for p, which is given by

$$C = \{p : \pi\{p|s\} \geq \pi_\alpha\},$$

where π_a is the largest constant such that $P(p \in C) \geq (1-\alpha)$. The length of C will then be an indicator of the statistical power of the method. Other than loss, statistical power in the Bayesian paradigm will also depend on the choice of the prior. Following the objective Bayesian philosophy as discussed by Berger (2006), one can use Jeffreys' non-informative prior. If there is a true shift in p, then it would be expected that the $100(1-\alpha)\%$ HPD interval from the second time point would not contain the posterior mean of p from the first time point. Hence we are motivated to define a heuristic decision procedure as follows: $d = \{Shift\ in\ p\}$ and $d^c = \{No\ shift\ in\ p\}$. We reject d if

$$E^{\pi^1}(p|s_1) \in C_2,$$

where C_2 is the $100(1-\alpha)\%$ HPD interval at the second time point, and accept d otherwise. We can easily extend this heuristic procedure for a general setup with $k(\geq 2)$ time points.

4.2 Simulation study

We develop a simple Monte Carlo algorithm to compute the statistical power of the heuristic decision procedure based on the adaptive Bayesian method developed in the previous sub-section. The algorithm is as follows:

Step 0: Initialise $count = 0$, and simulation size N.

Step 1: For given $(n_1,\ p_1)$ we generate a sample s_1 from $Binomial(n_1,\ p_1)$, where n_1 is the sample size at first time point, and p_1 is the true proportion of success.

Step 2: Given s_1 and $(a; b)$, we compute the posterior mean at the first time point as

$$E^{\pi^1}(p|s_1) = \frac{a + s_1}{n_1 + a + b},$$

where a and b are the hyper-parameters from the conjugate prior of the first time point.

Step 3: For given $(n_2,\ p_2)$, we generate a sample s_2 from $Binomial(n_2,\ p_2)$, where n_2 is the sample size at the second time point, and p_2 is the true proportion of success.

Step 4: We calculate the HPD interval C_2 for p for the second time point using the posterior of the second time point.

Step 5: If $E^{\pi^1}(p|s_1) \in C_2$ then $count = count + 1$. Go to Step 1.

Step 6: Repeat Steps 1–5 N times.

Step 7: Statistical power of the adaptive Bayesian method is $1 - \dfrac{count}{N}$.

We implement the above algorithm using the R-software. In Figures 1 and 2, we present the power function for some specific choices of sample sizes $n_1,\ n_2$, and p_1, for the entire range of $[0,\ 1]$ for p_2. For equal sample sizes $n_1 = n_2 = 15,\ 30,\ 50$ and $p_1 = 0.5,\ 0.25$, the performance of the two methods are similar. For equal sample size of $n_1 = n_2 = 15$ and $p_1 = 0.1,\ 0.9$, the performance of the regular frequentist method is better than the adaptive Bayesian method.

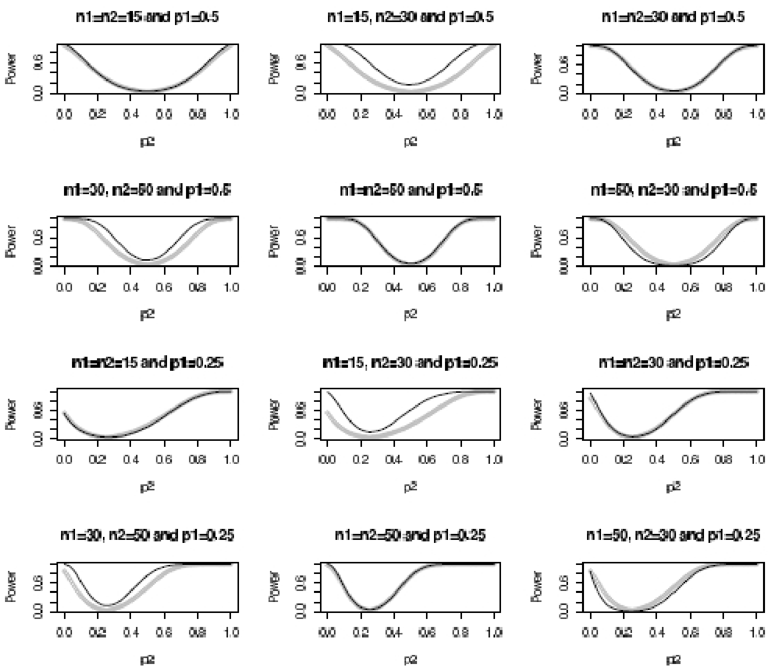

Figure 1 Simulation study of power function $(p_1 = 0.5,\ 0.25)$.
Note, Grey circles: Frequentist method, Black solid: Adaptive Bayesian method.

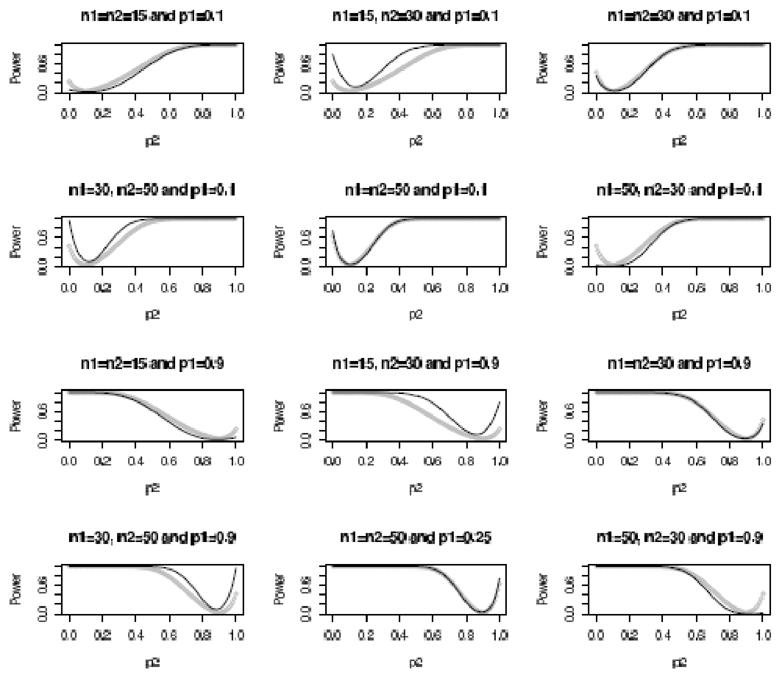

Figure 2 Simulation study of power function $(p_1 = 0.1,\ 0.9)$.
Grey circles: Frequentist method, Black solid: Adaptive Bayesian method.

In Figure 3, we present the exact power, calculated using the exact HPD interval (Chen *et al.*, 2002), and observe that the performance, for equal sample size cases $n_1 = n_2 = 15$ and $p_1 = 0.1,\ 0.9$, is similar in both the methods. When $n_2 > n_1$, the adaptive Bayesian method uniformly outperforms the frequentist method because the frequentist method uses the $min(n_1,\ n_2)$, forcing equality of sample size, thereby resulting in loss of information. When $n_2 < n_1$, the adaptive Bayesian method performs worse than the frequentist method. Hence, based on these studies, we recommend that one uses the adaptive Bayesian method when the sample size in the second time point is larger than the sample size in the first time point.

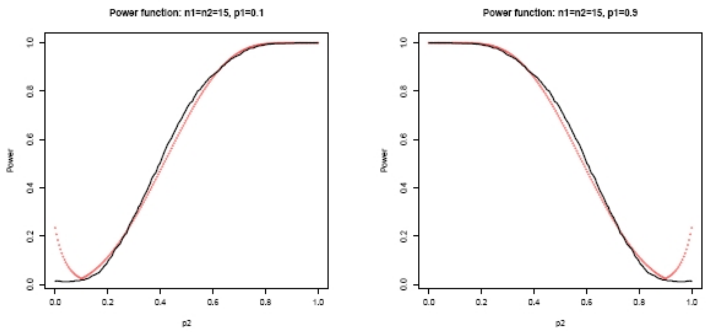

Figure 3 Exact power calculation for small sample size.
Note, Circles: Frequentist, Solid line: Adaptive Bayesian.

The theoretical justification of the above observation is provided via the following theorem:

Theorem 4.1. *Suppose* $S_t \sim Bin(n_t,\ p)$ *with prior on* $p = \pi^t \sim Beta(a_t,\ b_t)$, *and posterior* $\pi^t(p|s_t) \sim Beta(a_t + s_t,\ n_t + b_t - s_t)$. *Assume* $E(S_{t-1}) = \rho_t E(S_t)$, *where* ρ_t *is some known constant. If* $\rho_t n_t \geq n_{t-1}$, *then* $Var(p_t|S_{t-1}) \leq Var(p_{t-1}|S_{t-1})$, *while if* $\rho_t n_t \leq n_{t-1}$, *then* $Var(p_t|S_{t-1}) > Var(p_{t-1}|S_{t-1})$.

Proof. Assumption $E(S_{t-1}) = \rho_t\ E(S_t)$ implies

$$n_{t-1}p_{t-1} = \rho_t n_t p_t. \tag{4.6}$$

From (4.6), the posterior variance of $n_{t-1}p_{t-1}$ is given as

$$Var\left(p_{t-1}|S_{t-1}\right) = \rho_t^2 \frac{n_t^2}{n_{t-1}^2} Var\left(p_t|S_{t-1}\right). \tag{4.7}$$

Thus, when $\rho_t n_t \geq n_{t-1}$, then $Var(p_t|S_{t-1}) \leq Var(p_{t-1}|S_{t-1})$, while when $\rho_t n_t \leq n_{t-1}$, then $Var(p_t|S_{t-1}) > Var(p_{t-1}|S_{t-1})$. $\square$

In the following section, we present two case studies, where we apply the proposed adaptive Bayesian method to test for non-matched correlated binomial proportions where the sample size in each time point is different.

5. Case studies

5.1 Case study I: Evaluating teaching performance

In this study, data came from 2 time points, where the before and after intervention number of observations were unequal. The assumption of independence did not hold as the same group was evaluated at the two time points, while it was not strictly a matched pair problem. The Teachers for a New Era (TNE) project at the University of Connecticut conducted a study between 2005 and 2006 to evaluate the effectiveness of an intervention on the teaching quality in a particular grade of a participating school. The teachers were evaluated by 4 different raters into binary categories – 'proficient' and 'non-proficient', before and after the intervention. The teachers were rated blinded a different number of times, at the 2 time points. The total number of observations per teacher at the 2 time points being different, resulted in the proportion of 'proficient' teachers being calculated based on a different total number of observation. The data is summarised in Table 1.

Table 1 Summary of 4 raters evaluation at time points T_1 and T_2

	Rater 1		Rater 2		Rater 3		Rater 4	
	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2
Proficient	8	15	10	10	3	1	3	4
Non-proficient	19	19	9	14	11	24	12	32
Total	27	34	19	24	14	25	15	36

We implement the adaptive Bayesian method to the TNE data set using the conjugate Beta priors – namely, uniform, Novick-Hall, Zellner and Jeffreys. Due to space constraints, we are reporting results only from the Jeffreys prior in Table 2, though the results using the other mentioned priors are very similar. From Table 2, we observe that for Rater 3, the posterior mean of time point 1 is not contained in the 95% CI interval of the second time point. Also from Table 2, the 95% CI interval of the second time point of only Rater 3 lies entirely to the left of the posterior mean of the first time point. For all others, the 95% CI interval contains their respective posterior mean from the first time point.

We observe from Table 2 that the Kullback-Leibler divergence measure for Rater 3 is the largest, while Figure 4 reveals that for Raters 2, 3 and 4, the posterior density in the second time point (solid line) shifted to the left compared to the posterior density in the first time point (dotted line). This indicates that there is an agreement between the assessment of Raters 2, 3 and 4. Overall, we can say that based on the raters evaluation, there was a decrease in the proportion of proficient teachers after the intervention, with Rater 3 evaluation suggesting a significant drop.

Table 2 Posterior estimates of mean, standard error (s.e.), 95% CI and the Kullback-Leibler (KL) divergence for the 4 raters at two time points

		Post. mean	Post. s.e.	95% CI	KL divergence (+/−)
Rater 1	Time 1	0.304	0.085	(0.136, 0.471)	1.025 (+)
	Time 2	0.379	0.061	(0.259, 0.499)	
Rater 2	Time 1	0.525	0.109	(0.311, 0.739)	0.524 (−)
	Time 2	0.466	0.074	(0.322, 0.611)	
Rater 3	Time 1	0.250	0.105	(0.044, 0.456)	**2.238 (−)**
	Time 2	0.122	0.050	**(0.023, 0.221)**	
Rater 4	Time 1	0.235	0.100	(0.039, 0.431)	1.452 (−)
	Time 2	0.151	0.049	(0.055, 0.246)	

Note: '+/−' indicates a positive/negative shift of proficiency after intervention.

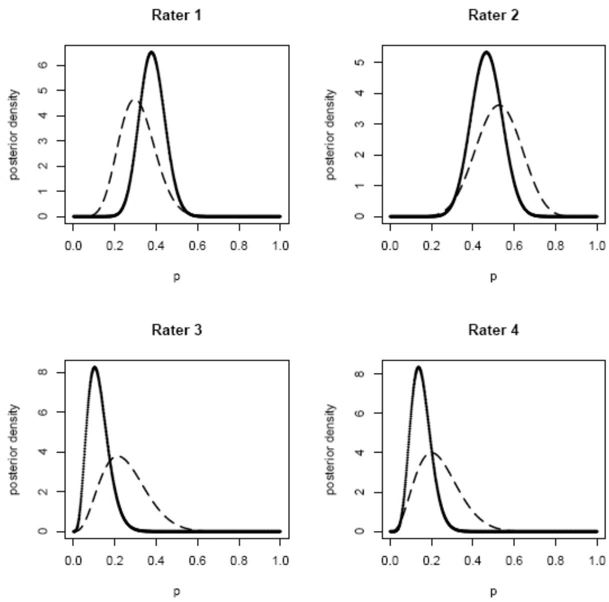

Figure 4 Posterior density of TNE data.
Note: Dotted lines: first time point; solid line: second time point.

5.2 Case study II: Hurricane data analysis

Webster *et al.* (2005) showed an increase in the number and proportions of hurricanes reaching categories '4 or 5'. In Table 3 we present the number of hurricanes in the Atlantic ocean during the period of 20 years between 1985 and 2004. Our objective is to check whether the proportion of category '4 and 5' hurricane has significantly gone up from the decade of 1985–1994 to 1995–2004. We thus propose to test the null hypothesis of $H_0 : p_1 = p_2$, where p_1 and p_2 correspond to the proportion of hurricanes in categories 4 or 5, in the two decades of 1985–1994 and 1995–2004 respectively. Clearly, a matched pair design is not conceivable for such a problem since the number of hurricanes in each of the decades vary. Also, the popular Fisher's exact test gives a p-value of 0.1367, which fails to reject the null hypothesis, and contradicts Webster *et al*. (2005). Moreover, the assumption of independence being naive in this case, we should not apply the Fisher's exact test.

Table 3 Hurricanes in Atlantic Ocean during 1985–2004

Hurricane category	1985–1994	1995–2004	Total
Cat 4 and 5	8	23	31
Cat 1, 2 and 3	41	55	96
Total	49	78	127

We thus analyse the hurricane data using our proposed adaptive Bayesian method. The strong Kullback-Leibler divergence measure of 3.021 indicates a change in proportions of category '4 and 5', over the two study periods. Also, from Table 4 and Figure 5 we observe that there is a right-ward shift of the posterior density of p during the latter decade, which indicates more category

'4 and 5' hurricanes during 1995–2004, compared to the previous decade, and agrees with Webster *et al.* (2005).

Table 4 Posterior estimates of mean proportion, standard error (s.e.) credible intervals at two time points for the Hurricane data

	Posterior		Credible interval		
	mean	s.e.	25%	50%	95%
Time 1	0.1700	0.5259	0.0803	0.1656	0.2846
Time 2	0.2461	0.0379	0.1757	0.2448	0.3240

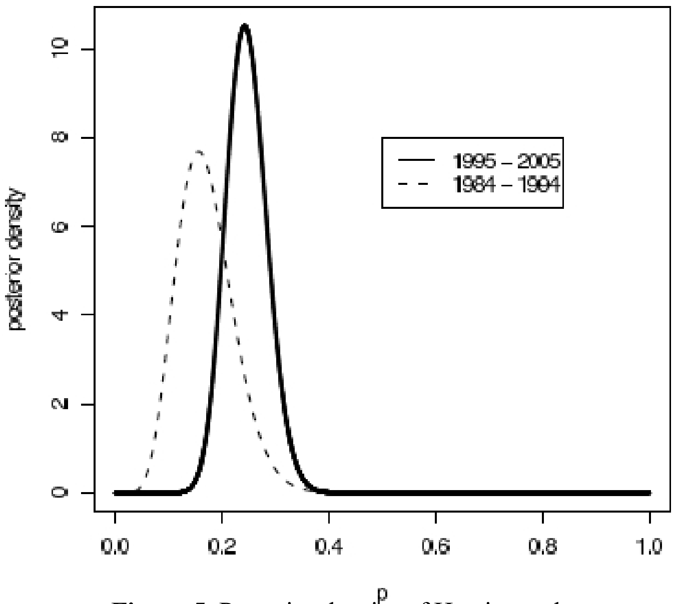

Figure 5 Posterior density of Hurricane data

6. Conclusion

The problem of inference for binomial proportions is important in statistics. Inference procedures for proportions from independent samples as well as from matched-paired samples are well established in the literature. However, for atypical cases where the same sample is tested for equality of proportions, at different time points, based on unequal numbers of observations, such inference methods do not suffice. In this paper, we develop an adaptive Bayesian method to address this problem. We evaluate the proposed method by comparing its performance to existing methods for detecting significant shifts in proportions between time points via the ϕ-divergence measure. We also make some recommendations for using the developed adaptive Bayesian method instead of the existing method in certain scenarios. We suggest a prescription, supported both by simulation studies and theoretical justification, that when the subsequent time point sample size is larger than the previous time point sample size, the adaptive Bayesian method outperforms the existing method. When the second time point sample size is smaller than the first time point sample size, the existing method outperforms the adaptive Bayesian method. For equal sample size scenarios, i.e., in the matched pair scenario, both methods perform equally well.

We apply the proposed method to two real case studies. In both case studies the inference problem could only be addressed via the developed method. The adaptive Bayesian method developed here does not consider any covariate information associated with the observations, which can be a natural extension of this paper.

Acknowledgement

This research was made possible, in part, by a grant from Carnegie Corporation of New York, the Ford Foundation, and the Annenberg Foundation. The authors wish to thank the Teachers for a New Era Project at the University of Connecticut for providing data and support for this project. The research was part supported by the CSIR, Pretoria. Sourish Das's research was part supported by NSF grant under DMS-0635449 to SAMSI. Further, we take this opportunity to express our gratitude to Prof. Dipak Dey, University of Connecticut, for his valuable comments. The statements made and views expressed are solely the responsibility of the authors.

References

AGRESTI, A. & CAFFO, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician*, **54**, 280–288.

AGRESTI, A. & COULL, B.A. (1998). Approximate is better than "Exact" for interval estimation of binomial proportions. *The American Statistician*, **52**, 119–126.

ALI, S.M. & SILVEY, S.D. (1966). A general class of coefficient of divergence of one distribution from another. *Journal of the Royal Statistical Society*, **286**, 131–142.

BANERJEE, M., CAPOZZOLI, M., MCSWEENEY, L. & SINHA, D. (1999). Beyond kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics*, **27**(1), 3–23.

BERGER, J.O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). Springer Series in Statistics, New York.

BERGER, J.O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, **1**(3), 385–402.

BROWN, L.D., CAI, T.T. & DASGUPTA, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, **16**(2), 101–133.

CHEN, M.-H. & DEY, D. (1998). Bayesian modeling of correlated binary responses via scale mixture of multivariate normal link functions. *Sankhyā Series A*, **60**, 322–343.

CHEN, M.H., SHAO, Q.M., & IBRAHIM, J.G. (2002). *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

COCHRAN, W.G. (1954). Some methods for strengthening the common χ^2 test. *Biometrics,* **10**, 417–451.

CSISZAR, I. (1967). Information type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, **2**, 105–113.

DEY, D.K. & BIRMIWAL, L.R. (1994). Robust Bayesian analysis using entropy and divergence measures. *Statist. Probab. Lett.*, **20**, 287–294.

EGUCHI, S. & COPAS, J. (2006). Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *Journal of Multivariate Analysis*, **97**(9), 2034–2040.

GEISSER, S. (2006). *Modes of parametric statistical inference*. Wiley Series in Probability and Statistics.

GHOSH, M. & CHEN, M.-H. (2002). Bayesian inference for matched case-control studies. *Sankhyā, Series B*, **2**(64), 107–127.

GHOSH, M., CHEN, M.-H., GHOSH, A. & AGRESTI, A. (2000). Hierarchical Bayesian analysis of binary matched pairs data. *Statista Sinica*, **10**, 647–657.

JEFFREYS, H. (1961). *Theory of probability* (3rd. ed.). Oxford University Press, London.

KULLBACK, S. & LEIBLER, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

LIAO, J.G. (1999). A hierarchical Bayesian model for combining multiple 2×2 tables using conditional likelihoods. *Biometrics*, **55**(1), 268–272.

LIDDELL, F.D. (1983). Simplified exact analysis of case-referent studies: matched pairs; dichotomous exposure. *Journal of Epidemiology and Community Health*, **37**, 82–84.

LLOYD, C.J. (1990). Confidence intervals from the difference between two correlated proportions. *Journal of the American Statistical Association*, **85**, 1154–1158.

MANTEL, N. & HAENSZEL, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, **22**, 719–748.

MCNEMAR, Q. (1947). Note on the sampling error of the differences between correlated proportions of percentages. *Psychometrika*, **12**, 153–157.

NEWCOMBE, R.G. (1998). Improved confidence interval for the difference between binomial proportions based in paired data. *Statistics in Medicine*, **17**, 2635–2650.

NOVICK, M.R. & HALL, W.J. (1965). A Bayesian indifference procedure. *Journal of the American Statistical Association*, **60**, 1104–1117.

WEBSTER, P.J., HOLLAND, G.J., CURRY, J.A. & CHANG, H.R. (2005). Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, **309**, 1844–1846.

ZELLNER, A. (1977). Maximal data information prior distributions. In *Methods in the applications of Bayesian methods*, A. Aykac & C. Brumat (eds.), 211–232 North-Holland, Amsterdam.

ZELLNER, A. (1988). Optimal information processing and Bayes' theorem. *The American Statistician*, **42**(4), 278–280.

ZELLNER, A. (2002). Information processing and Bayesian analysis. *Journal of Econometrics*, **107**, 41–50.

Manuscript received, 2008.12, revised, 2009.05, accepted, 2009.05.