STATISTICS IN MEDICINE, VOL. 14, 1057–1063 (1995)

BAYESIAN PREDICTIVE APPROACH FOR INFERENCE ABOUT PROPORTIONS

BRUNO LECOUTRE

URA 1378, Analyse et Modélisation Stochastique, C.N.R.S. et Université de Roven (Mathématiques), 76821 Mont-Saint-Aignan Cedex, France

GÉRARD DERZKO

SANOFI Recherche, 371 rue du Professeur Joseph Blayac 34184 Montpellier Cedex 4, France

AND

JEAN-MARIE GROUIN

Equipe ERIS, Laboratoire de Psychologie, Université de Roven BP108, 76134 Mont-Saint-Aignan Cedex, France

SUMMARY

This paper investigates the Bayesian procedures for comparing proportions. These procedures are especially suitable for accepting (or rejecting) the equivalence of two population proportions. Furthermore the Bayesian predictive probabilities provide a natural and flexible tool in monitoring trials, especially for choosing a sample size and for conducting interim analyses. These methods are illustrated with two examples where antithrombotic treatments are administrated to prevent further occurrences of thromboses.

INTRODUCTION

Several papers have shown the interest of the Bayesian predictive approach in monitoring clinical trials. The idea is to compute, on the basis of the available information, the chances of obtaining a prespecified conclusion (for instance, a significant result at level α, a prespecified confidence or Bayesian credibility interval) with future observations. Predictive probabilities provide a powerful and natural tool for determining the required sample size in a projective clinical trial (see Berry[1]) and for allowing sensible decision making in interim analyses (see Choi and Pepple[2]).

Many authors have laid stress on the interest of the Bayesian approach in decision making. In this paper, it will be emphasized that Bayesian procedures have also an important contribution to inference and data analysis. Frequency data analysis will be developed within this framework. Exact Bayesian methods for one or two binomial samples will be presented. These methods are especially suitable for assessing the equivalence of two proportions or the non-inferiority of one proportion with respect to the other. Bayesian solutions will be applied to the variables most commonly used to compare two proportions, namely their difference, ratio, or odds-ratio. Predictive Bayesian procedures will be investigated. These procedures require heavy computations which have long been an impedement to their use, but are now affordable. They open new interesting perspectives for monitoring clinical trials, which are sketched in this paper.

Two examples, where antithrombotic treatments were administered to prevent further occurrences of thromboses, serve as illustration of the procedures. Data were analysed with 'standard'

CCC 0277–6715/95/101057–07

Bayesian procedures, that is, based on a *non-informative* prior distribution (according to Jeffreys), but the procedures can, in a straightforward way, be extended to situations where external information must be included through (informative) conjugate priors.

The basics of the Bayesian methodology will not be presented here. Several handbooks are suitable for this purpose. As an introduction the reader is referred to Novick and Jackson,[3] Lee,[4] Rouanet *et al.*[5] For the conceptual background, see Berger[6] and Robert.[7]

EXAMPLES: DESIGN, QUESTIONS OF INTEREST

Example 1: Inference about one proportion

The patients under study were post-MI (myocardial infarction) patients, treated with a low molecular weight heparin (LMWH) as a prophylaxis of an intracardial left ventricular thrombosis (ILVT). Because of the limited knowledge available on drug potential efficacy, a two-stage sequential design was used. This was aimed at abandoning further development as early as possible if the drug was likely to be not effective, and at estimating its efficacy if it turned out to be promising.

The sample sizes for the two stages were chosen so as to minimize the expected total sample size, and were computed along the lines of a method developed by Simon[8] in the context of oncology. It was considered that $p_1 = 0{\cdot}15$ was the thrombosis rate below which the drug would not be very attractive, and that $p_2 = 0{\cdot}30$ was the rate above which the drug would be of no interest. An adaptation of Simon's method showed that $n = 20$ patients in the first stage and 39 patients in the second stage were required if $\alpha = 0{\cdot}05$ and $\beta = 0{\cdot}15$.

The sampling distribution is assumed to be binomial, with parameter φ. The maximum sample size has been chosen for additional data, so that a total sample size $N = 20 + 39$ is envisioned. It follows from the hypotheses that three regions of interest are to be selected: (i) $\varphi < 0{\cdot}15$ defines a potentially interesting drug; (ii) $\varphi > 0{\cdot}30$ defines a failure; (iii) $0{\cdot}15 < \varphi < 0{\cdot}30$ is a no-decision region.

Note that Simon's method allows for designing a small early phase activity study based on explicit hypotheses. Its decision-oriented conception, however, is often misleading for the clinician, who tends to confuse rejecting the bad case ($\varphi > 0{\cdot}30$) and accepting the good case ($\varphi < 0{\cdot}15$). The trouble is that a decision (to accept or to reject the drug) is taken at the second stage, even if the observed proportion falls in the no-decision region [0·15, 0·30]. What a clinician actually needs is to evaluate at any stage of the trial the probability of some specified regions of interest and the ability for a future sample to support and corroborate findings already observed. The Bayesian analysis addresses this problem.

Example 2: Inference about two proportions

Hull *et al.*[9] reported on a study where a short-course (5 days) treatment with continuous intravenous heparin was compared to the conventional 10-day course in patients suffering acute proximal venous thrombosis. From the observed rates of 7 out of 99 in the short-course treatment, and 7 out of 100 in the long-course treatment, they computed a two-sided 95 per cent confidence interval for the difference $[-0{\cdot}073, 0{\cdot}075]$, and concluded that the short-course treatments were equally effective. Corey[10] criticized this conclusion, arguing that the data did not provide sufficient evidence of equivalence, as the upper boundary of a one-sided 95 per cent confidence interval for the thrombosis rate ratio was 2·7.

The Bayesian analysis offers simple and flexible techniques for thoroughly assessing the outcome of this study, and the alternative use of a sequential design.

SOME TECHNICAL RESULTS FOR BAYESIAN ANALYSIS

One binomial sample

For a binomial sample of size $n = 20$, let a be the observed 'number of successes' and let $b = n - a$. Assuming for φ a prior beta distribution with parameter a_0 and b_0, the posterior is again a beta distribution with parameters $a_0 + a$ and $b_0 + b$.

Here the Jeffreys' non-informative prior $a_0 = b_0 = 1/2$ (that is, a uniform prior for $\sin^{-1}\sqrt{\varphi}$) is used. This last solution is a compromise between the two 'extreme' choices also proposed as non-informative priors: $a_0 = b_0 = 0$ and $a_0 = b_0 = 1$ (that is, uniform for φ). External information could be alternatively incorporated into the analysis, by choosing suitable a_0 and b_0.

Given the posterior distribution beta $(a_0 + a, b_0 + b)$, the predictive probability of observing k successes in a future sample of size N is obtained from a beta-binomial distribution (see Lee[4]).

Two independent binomial samples

The procedure for the case of one single sample is extended to the case of two independent binomial samples with parameters φ_1 and φ_2. Assuming two respective marginal independent prior distributions, beta $(a_{1,0}, b_{1,0})$ and beta $(a_{2,0}, b_{2,0})$, the above results apply to each parameter, φ_1 and φ_2. Both the marginal posterior distributions and the marginal predictive distributions are independent. Here the non-informative prior $a_{1,0} = b_{1,0} = a_{2,0} = b_{2,0} = 1/4$ is chosen. Statements about derived parameters, such as $\varphi_1 - \varphi_2$, φ_1/φ_2, etc., can be obtained by a simple numerical method using the incomplete beta function: see for instance Novick and Jackson[3] (pages 338–342).

ILLUSTRATION: EXAMPLE 1

Standard Bayesian analysis for φ

A *guarantee* (or credibility level) $\gamma = 0{\cdot}90$ is selected for Bayesian statements. At the end of the first stage, the *standard* Bayesian posterior distribution of φ is determined and the corresponding probabilities associated with the regions of interest are computed. Three examples of posterior distributions are shown in Figure 1, corresponding to the observed rates 1/20, 4/20 and 10/20.

If the Bayesian guarantee obtained is less than γ, no conclusion can be drawn at the end of the first stage, and the second stage must be carried out. For instance, this occurs if the observed thrombosis rate is 4/20, since $\Pr(\varphi < 0{\cdot}15|4/20) = 0{\cdot}251$, $\Pr(0{\cdot}15 < \varphi < 0{\cdot}30|4/20) = 0{\cdot}584$, and $\Pr(\varphi > 0{\cdot}30|4/20) = 0{\cdot}165$.

If the Bayesian probability associated with a given region is more than γ, a conclusion is available based upon *the data from the first stage*. For instance, an observed thrombosis rate of 10/20 leads to a conclusion of inefficacy, since $\Pr(\varphi > 0{\cdot}30|10/20) = 0{\cdot}971$. On the contrary, an observed thrombosis rate of 1/20 leads to a conclusion of efficacy, since $\Pr(\varphi < 0{\cdot}15|10/20) = 0{\cdot}907$. The issue is then to know whether the future data (39 additional patients) could invalidate this conclusion.

The (potential) Bayesian posterior probability based upon *all of the data* can be computed for each possible result at the second stage. Table I summarizes the different possible outcomes at the end of each stage, as a function of the observed number of thromboses.

Future sample: predictive probabilities

Let k be the number of thromboses in the future sample ($0 \leqslant k \leqslant 39$). If 10/20 is observed at the first stage, the conclusion of inefficacy will be confirmed (with a guarantee of at least 0·90) for

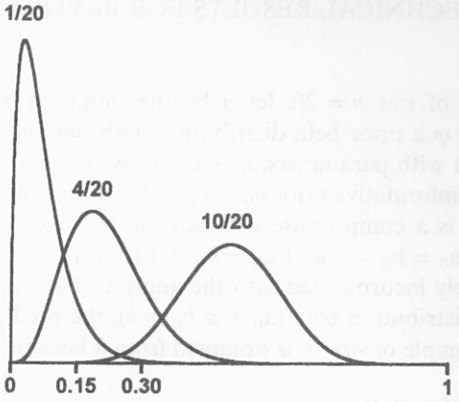

Figure 1. Posterior distributions of φ for the observed rates 1/20, 4/20 and 10/20

Table I. Range of observed numbers of thromboses giving the corresponding conclusion

Conclusion at guarantee $\gamma = 0{\cdot}90$	First stage ($n = 20$)	Second stage ($N = 59$)
$\varphi < 0{\cdot}15$	[0, 1]	[0, 5]
$0{\cdot}15 < \varphi < 0{\cdot}30$	/	/
$\varphi > 0{\cdot}30$	[9, 20]	[23, 59]

k satisfying $10 + k \geqslant 23$, that is, $13 \leqslant k \leqslant 39$. The predictive Bayesian probability $\Pr(k \geqslant 13 | 10/20)$ (*conditional on the sampled data*) of observing such a value of k can be easily computed. It is equal to 0·910, hence the second stage is likely to confirm the conclusion obtained at the end of the first stage. If so, it is not necessary to sample additional data.

If 1/20 is observed at the first stage, the conclusion of efficacy will be confirmed for $k \leqslant 4$. The predictive probability $\Pr(k \leqslant 4 | 1/20)$ is only 0·786, and thus the second stage must be carried out in order to accumulate more evidence of efficacy.

Remark: Evaluating the sample size

The predictive approach can also be used to evaluate if a given sample size is appropriate for a conclusion of efficacy, for example. Such an evaluation is based upon the *prior* predictive probability of observing a thrombosis rate of at most 5/59. Consider the sampling probabilities $\Pr(k \leqslant 5 | \varphi)$ for different values of φ:

$\Pr(k \leqslant 5 | \varphi = 0{\cdot}01) = 0{\cdot}99997$, $\Pr(k \leqslant 5 | \varphi = 0{\cdot}02) = 0{\cdot}999$, $\Pr(k \leqslant 5 | \varphi = 0{\cdot}03) = 0{\cdot}992$,

$\Pr(k \leqslant 5 | \varphi = 0{\cdot}04) = 0{\cdot}970$, $\Pr(k \leqslant 5 | \varphi = 0{\cdot}05) = 0{\cdot}926$, $\Pr(k \leqslant 5 | \varphi = 0{\cdot}06) = 0{\cdot}858$,

$\Pr(k \leqslant 5 | \varphi = 0{\cdot}07) = 0{\cdot}770$, $\Pr(k \leqslant 5 | \varphi = 0{\cdot}08) = 0{\cdot}667$, $\Pr(k \leqslant 5 | \varphi = 0{\cdot}09) = 0{\cdot}559$.

This sampling probability is greater than 0·90 when $\varphi \leqslant 0{\cdot}05$. If the true value of φ is assumed to be less than 0·05, $N = 59$ is therefore a reasonable choice. Furthermore, when external information can be expressed through a prior distribution, the associated predictive probability can be used instead of the sampling probabilities. For instance, a prior distribution beta(8·5, 192·5), which gives *a priori* $\Pr(\varphi < 0{\cdot}061) = 0{\cdot}90$, yields the predictive probability 0·940.

Table II. Bayesian statements for φ_1 and φ_2 at guarantee $\gamma = 0{\cdot}90$

Joint statement	$\varphi_1 < 0{\cdot}119$ and $\varphi_2 > 0{\cdot}035$
Absolute equivalence	$\lvert\varphi_1 - \varphi_2\rvert < 0{\cdot}060$ $0{\cdot}42 < \varphi_1/\varphi_2 < 2{\cdot}37$ $0{\cdot}40 < (\varphi_1/1 - \varphi_1)/(\varphi_2/1 - \varphi_2) < 2{\cdot}52$
Relative efficacy	$\varphi_1 - \varphi_2 < 0{\cdot}047$ $\varphi_1/\varphi_2 < 1{\cdot}97$ $(\varphi_1/1 - \varphi_1)/(\varphi_2/1 - \varphi_2) < 2{\cdot}07$

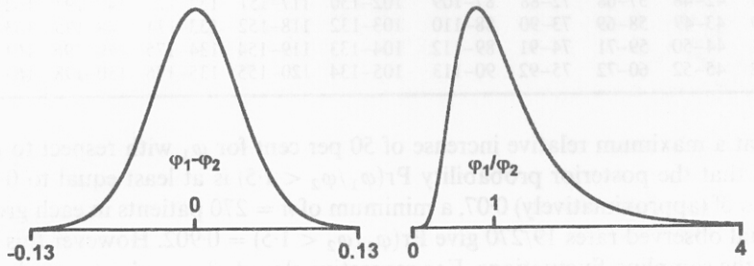

Figure 2. Posterior distributions of $\varphi_1 - \varphi_2$ and φ_1/φ_2

Note that the prior distribution is used here in order to evaluate the sample size, but is not included in the data analysis itself.

ILLUSTRATION: EXAMPLE 2

Standard Bayesian analysis

The design involves two independent binomial samples with parameters φ_1 (short-course) and φ_2 (long-course). A joint probability statement is, in a way, the best summary of the posterior distribution. For instance, the posterior probability that both $\varphi_1 < 0{\cdot}131$ and $\varphi_2 > 0{\cdot}030$ is 0·95. However, a statement that deals with the comparison of the two courses directly would be preferable. This is solved in the Bayesian approach, since the distribution of any derived parameter of interest can be obtained from the joint posterior distribution of (φ_1, φ_2). Therefore, the main classical criteria for equivalence, that is, $\varphi_1 - \varphi_2$, φ_1/φ_2, and $(\varphi_1/1 - \varphi_1)/(\varphi_2/1 - \varphi_2)$, can be dealt with. It may be argued here that the study does not aim to demonstrate absolute equivalence (for instance $\lvert\varphi_1 - \varphi_2\rvert < \varepsilon$ or $1/\rho < \varphi_1/\varphi_2 < \rho$), but rather to conclude to the relative efficacy of shorter therapy. Hence a one-sided condition, such as $\varphi_1 - \varphi_2 < \varepsilon$ or $\varphi_1/\varphi_2 < \rho$ $(\rho > 1)$, appears to be more suitable.

The corresponding Bayesian statements at guarantee $\gamma = 0{\cdot}90$ are given in Table II. Figure 2 shows the posterior distributions of $\varphi_1 - \varphi_2$ and φ_1/φ_2.

A suitable design for assessing the efficacy of the short-course treatment

The Bayesian analysis clearly supports Corey's reserves. A statement such as $\varphi_1/\varphi_2 < 1{\cdot}97$ may hardly be regarded as a sufficient demonstration of the effectiveness of the short-course treatment. In other words more data are necessary to obtain acceptable evidence. For simplicity's sake, we will use a single criterion, but all the procedures could apply in the same way to any criteria.

Table III. The statement $\Pr(\varphi_1/\varphi_2 | < 1{\cdot}5) \geqslant 0{\cdot}90$ holds for all pairs (k_1, k_2) where k_2 is the first number and k_1 is inferior or equal to the second numbers

1– 0	16–15	31–33	46–53	61–73	76– 94	91–115	106–135	121–157	136–178	151–199	166–221
2– 0	17–16	32–35	47–54	62–75	77– 95	92–116	107–137	122–158	137–179	152–201	167–222
3– 1	18–17	33–36	48–56	63–76	78– 97	93–117	108–138	123–159	138–180	153–202	168–223
4– 2	19–18	34–37	49–57	64–77	79– 98	94–119	109–140	124–161	139–181	154–203	169–211
5– 3–	20–20	35–39	50–58	65–79	80– 99	95–120	110–141	125–162	140–183	155–205	170–226
6– 4	21–21	36–40	51–60	66–80	81–101	96–121	111–142	126–164	141–185	156–206	171–228
7– 5	22–22	37–41	52–61	67–81	82–102	97–123	112–144	127–165	142–186	157–208	172–229
8– 6	23–23	38–43	53–62	68–83	83–103	98–124	113–145	128–166	143–188	158–209	173–231
9– 7	24–25	39–44	54–64	69–84	84–105	99–126	114–147	129–168	144–189	159–211	174–232
10– 8	25–26	40–45	55–65	70–86	85–106	100–127	115–148	130–169	145–191	160–212	175–233
11– 9	26–27	41–46	56–66	71–87	86–108	101–128	116–149	131–171	146–192	161–213	176–235
12–10	27–28	42–48	57–68	72–88	87–109	102–130	117–151	132–172	147–193	162–215	177–236
13–11	28–30	43–49	58–69	73–90	88–110	103–132	118–152	133–174	148–195	163–216	178–238
14–12	29–31	44–50	59–71	74–91	89–112	104–133	119–154	134–175	149–196	164–218	179–239
15–14	30–32	45–52	60–72	75–92	90–113	105–134	120–155	135–176	150–198	165–219	180–241

Assume that a maximum relative increase of 50 per cent for φ_1 with respect to φ_2 is allowed. This requires that the posterior probability $\Pr(\varphi_1/\varphi_2 < 1{\cdot}5)$ is at least equal to 0·90. For equal observed rates of (approximatively) 0·07, a minimum of $n = 270$ patients in each group is needed, since two equal observed rates 19/270 give $\Pr(\varphi_1/\varphi_2 < 1{\cdot}5) = 0{\cdot}902$. However this does not take into account the sampling fluctuations. For respective closed observed rates, such as 20/270 and 17/270, the posterior probability $\Pr(\varphi_1/\varphi_2 < 1{\cdot}5)$ falls to 0·776.

It appears here that, even in assuming *a priori* $\varphi_1 = \varphi_2$ with an expected common value of about 0·07, a reasonable choice for n must be at least 1000. The subset of future pairs of results (k_1, k_2), for which the posteriori probability $\Pr(\varphi_1/\varphi_2 < 1{\cdot}5)$ is at least 0·90, must be determined. The range of considered results for the long-course treatment can be restricted to the subset [0, 180], with a negligible loss of accuracy upon the predictive probability for any value of $\varphi_1 = \varphi_2$ ranking from 0 to 0·125. For such a value the overall sampling probability of the neglected set is indeed always less than 10^{-6}. Table III gives the subset of the corresponding pairs (k_1, k_2). Consider the sampling probabilities of observing such a pair for different (common) values of φ_1 and φ_2.

$$\varphi_1 = \varphi_2 = 0{\cdot}06\text{: } 0{\cdot}836, \quad \varphi_1 = \varphi_2 = 0{\cdot}07\text{: } 0{\cdot}881, \quad \varphi_1 = \varphi_2 = 0{\cdot}08\text{: } 0{\cdot}914,$$

$$\varphi_1 = \varphi_2 = 0{\cdot}09\text{: } 0{\cdot}938, \quad \varphi_1 = \varphi_2 = 0{\cdot}10\text{: } 0{\cdot}956.$$

Owing to the large samples needed, interim analyses appear here to be highly desirable and can be conducted, for instance, after each inclusion of 100 pairs of patients. More particularly these analyses are essential for deciding to stop the trial early, either if the short-course treatment turns out to be ineffective or if a conclusion of efficacy may already be reached with a smaller number of patients.

Examples of interim analyses

The process is the same as in the one sample case. Given two interim observed rates a_1/n_1 and a_2/n_2, the set of future pairs such as the posterior probability $\Pr(\varphi_1/\varphi_2 < 1{\cdot}5 | \text{all of the data})$ is at least 0·90 is deduced from Table III, and the corresponding predictive probability given the available data computed.

Consider for instance an interim analysis at $n_1 = n_2 = 500$; each of the two results $(a_1 = 52, a_2 = 50)$ and $(a_1 = 37, a_2 = 35)$ states the relative efficacy of the short-course treatment

for the observed data, as demonstrated by $\Pr(\varphi_1/\varphi_2 < 1{\cdot}5 | a_1 = 52,\ a_2 = 50) = 0{\cdot}974$ and $\Pr(\varphi_1/\varphi_2 < 1{\cdot}5 | a_1 = 37, a_2 = 35) = 0{\cdot}937$. The predictive probability that this conclusion should be confirmed by two additional samples of 500 patients is 0·928 in the first case, but only 0·811 in the second.

CONCLUSION

These two examples illustrate several interesting features of the Bayesian approach. In the framework of a definitely decisional trial (first example), there is also a need for estimation. This is shown by Gehan's[11] former method, in which the first stage was decisional and the second stage aimed at assessing precision. Bayesian methodology enables the probabilities of the prespecified regions of interest to be obtained. Moreover, if the drug shows early signs of being effective, or ineffective, the predictive approach allows a confirmatory next stage to be set up.

In a clearly undersized demonstrative trial (second example), the predictive approach enables the experimentation to be properly extended to an adequate size step by step, in an actual sequential perspective.

REFERENCES

1. Berry, D. A. 'Experimental design for drug development: A Bayesian approach', *Journal of Biopharmaceutical Statistics*, **1**, 81–102 (1991).
2. Choi, S. C. and Pepple, P. A. 'Monitoring clinical trials based on predictive probability of significance', *Biometrics*, **45**, 317–323 (1989).
3. Novick, M. R. and Jackson, P. H. *Statistical Methods for Educational and Psychological Research*, McGraw-Hill, New York, 1973.
4. Lee, P. M. *Bayesian Statistics: An Introduction*, Oxford University Press, New York, 1991.
5. Rouanet, H., Lecoutre, M.-P., Bert, M.-C., Lecoutre, B. and Bernard, J.-M. *L'inférence Statistique dans la Démarche du Chercheur*, Peter Lang, Berne, 1991.
6. Berger, G. O. *Statistical Decision Theory and Bayesian Analysis*, Springer Verlag, New York, 1985.
7. Robert, C. *L'Analyse Statistique Bayésienne*, Economica, Paris, 1992.
8. Simon, R. 'Optimal two stages design for phase II clinical trials', *Controlled Clinical Trials*, **10**, 1–10 (1989).
9. Hull, R. D., Raskob, G. E., Rosenbloom, D., Panju, A. A., Edwards, P. B., Ginsberg, J. S., Hirsh, J., Martin, G. J. and Green, D. 'Heparin for 5 days as compared with 10 days in the initial treatment of proximal venous thrombosis', *The New England Journal of Medicine*, **322**, 1260–1264 (1990).
10. Corey, G. A. 'Correspondence: Heparin in the initial treatment of venous thrombosis', *The New England Journal of Medicine*, **323**, 830 (1990).
11. Gehan, E. A. 'The determination of the number required in a follow-up trial of a new chemotherapeutic agent', *Journal of Chronic Diseases*, **13**, 346–353 (1961).