

A Bayesian Approach to Subset Analysis with Application to Binomial Count Data

*Un approccio bayesiano all'analisi dei sottogruppi
con applicazione a dati binomiali*

Stefano Conti, Fulvio Spezzaferri

Dipartimento di Statistica, Probabilità e Statistiche Applicate

Università di Roma "La Sapienza"

s.conti@uniroma1.it,

fulvio.spezzaferri@uniroma1.it

Riassunto: Nel quadro schematico di una prova clinica, i pazienti sotto esame siano preventivamente raccolti in classi omogenee secondo requisiti prognostici e/o demografici; avendo dunque estrapolato tra i vari sottogruppi uno di diretto interesse, a questo resta improntata l'indagine sulla portata dell'azione terapeutica, in forma di problema di test d'ipotesi. Il responso conclusivo viene infine adattato alla sensibilità tendenziale esibita verso il trattamento da parte dei restanti sottogruppi; a tale scopo si è ricorso, secondo l'impostazione bayesiana, ad un modello di regressione gerarchica, in cui effetti sperimentali di verso concorde sono supposti scambiabili. La descrizione delle modalità di intervento dei sottogruppi di supporto nel processo decisionale, relativo a quello d'interesse, viene riferita in particolare all'analisi di dati binari (incidenza di eventi clinici).

Keywords: Bayes factor, Bayesian inference, Borrowing strength, Exchangeability, Hierarchical models, Subset analysis.

1. Introduction

The main purpose of a controlled clinical trial relies in assessing if, and to what extent, an experimental therapy is appropriate in contrasting a certain pathology, as compared to a standard treatment or a placebo. Usually, paired with measurements of an appropriate outcome of clinical interest, records of prognostic and/or demographic factors are made available for each patient undertaking the trial. A *subset analysis* is then aimed at detecting and examining possible treatment differences arising from distinct clusters of homogeneous patients, besides the effect possibly exhibited by the whole set of individuals under examination.

The field of subset analysis has been traditionally dealt with by means of frequentist techniques (Byar, 1985, among the others). More recently, the subject has also been explored by adopting a Bayesian perspective (in particular Dixon and Simon, 1991). In the present paper a different Bayesian approach is proposed, inspired by a methodology described by De Santis and Spezzaferri (1998): given $m \geq 2$ distinct subgroups of patients randomized to either control or treatment arms, a specific subgroup of relevant interest is chosen (say, with no loss of generality, the first one). Based on the corresponding data, an hypothesis testing procedure is then performed only on the selected patients, in order to assess the direction of the local treatment effect. The final decision relative to such subgroup is then adapted according to how the remaining subgroups react to the therapeutic prescription. Such re-weighting procedure seems sensible, because a subset analysis is usually attempted when a promising treatment doesn't show significant evidence of a

global effect; any local finding, however, may possibly be biased due to outlying or influential data. In this sense, involving the other subgroups in the decision process reveals useful in validating the inference about the one of explicit interest.

In section 2, after formulating appropriate assumptions, the usual tool for Bayesian hypothesis testing, that is the Bayes factor, is derived and discussed. In section 3 the reliability of the technique is described with reference to binary response data.

2. Description of the Problem and Bayesian Solution

Suppose outcome data y_{rit} are collected from each individual $r = 1, \dots, n_{it}$ in subgroup $i = 1, \dots, m$, either under treatment ($t = 1$) or control ($t = 0$). Assume the iid property holds between vectors $\mathbf{y}_i$ of measurements from each subgroup i , given nuisance and interest parameters β_i and ϑ_i , respectively; therefore, the full likelihood of the parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$ and $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_m)$ is given by $f(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\vartheta}) = \prod_{i=1}^m f(\mathbf{y}_i | \beta_i, \vartheta_i)$, $f(\mathbf{y}_i | \beta_i, \vartheta_i)$ being the conditional pmf, or pdf, of $\mathbf{y}_i$. The hypothesis testing problem under debate is then of the form

$$\begin{cases} H_0) & A_1^+ \\ H_1) & A_1^- \end{cases}, \quad (1)$$

$A_i^+ = \{\vartheta_i \in \mathbb{R} \mid \vartheta_i > 0\}$ and $A_i^- = \{\vartheta_i \in \mathbb{R} \mid \vartheta_i < 0\}$ representing respectively an improvement or worsening of patients' health in subgroup i , due to therapy prescription, each with prior probability $p(A_i^+) \in (0, 1)$ and $p(A_i^-) = 1 - p(A_i^+)$. Indeed, events relative to different subgroups are supposed to be unrelated.

Assuming the nuisance parameters $\boldsymbol{\beta}$ independent of each other given the treatment effects $\boldsymbol{\vartheta}$, the following hypotheses hold about the latter:

mixture: the prior probability mass of ϑ_i is split between independent continuous distributions on $\mathbb{R}^+$ and $\mathbb{R}^-$: if $\mathbb{I}_C$ is the indicator function of some event C ,

$$\pi(\vartheta_i) = p(A_i^+)\pi(\vartheta_i \mid A_i^+)\mathbb{I}_{A_i^+} + p(A_i^-)\pi(\vartheta_i \mid A_i^-)\mathbb{I}_{A_i^-}$$

independence: knowledge of how a subgroup of patients reacts to the treatment doesn't influence the prior opinion regarding the value of any other local effect: if $\star$ indicates either + or - symbols, it results

$$\pi(\vartheta_i \mid A_{i'}^\star) = \pi(\vartheta_i), \quad i' \neq i$$

exchangeability: treatment effects having the same direction are conditionally independent, given a certain hyper-parameter:

$$\pi\left(\boldsymbol{\vartheta} \mid \bigcap_{i=1}^m A_i^\star\right) = \int \prod_{i=1}^m \pi(\vartheta_i \mid \mu^\star, A_i^\star) p(\mu^\star) d\mu^\star.$$

The Bayes factor of H_0 vs. H_1 , namely the ratio of posterior to prior odds relative to A_1^+ , can then be shown to be

$$BF_1(\mathbf{y}) = BF_1(\mathbf{y}_1) \frac{p(\bigcup_{i=2}^m A_i^+ \mid \mathbf{y}_{(1)}) \mathcal{H}^+(\mathbf{y}) + p(\bigcap_{i=2}^m A_i^- \mid \mathbf{y}_{(1)}) \mathcal{H}^-(\mathbf{y})}{p(\bigcup_{i=2}^m A_i^- \mid \mathbf{y}_{(1)}) \mathcal{H}^-(\mathbf{y}) + p(\bigcap_{i=2}^m A_i^+ \mid \mathbf{y}_{(1)}) \mathcal{H}^+(\mathbf{y})}, \quad (2)$$

where $BF_1(\mathbf{y}_1) = f(\mathbf{y}_1 | A_1^+) / f(\mathbf{y}_1 | A_1^-)$ is the Bayes factor for (1) based only on data $\mathbf{y}_1$ and $\mathbf{y}_{(1)} = (\mathbf{y}_2, \dots, \mathbf{y}_m)$. The evidence provided on a local basis by $BF_1(\mathbf{y}_1)$ is thus emphasized or reappraised according to a ratio of weighted posterior averages between 1 and in turn the correction factors

$$\mathcal{H}^*(\mathbf{y}) \equiv \frac{\int f(\mathbf{y}_1 | \vartheta_1, A_1^*) \pi(\vartheta_1 | A_1^*, \bigcup_{i=2}^m A_i^*, \mathbf{y}_{(1)}) d\vartheta_1}{\int f(\mathbf{y}_1 | \vartheta_1, A_1^*) \pi(\vartheta_1 | A_1^*) d\vartheta_1},$$

which are the only statistics in (2) depending on the whole sample. For instance, the term $\mathcal{H}^+(\cdot)$ is the Bayes factor needed to discriminate amongst the hypotheses of exchangeability and independency between the treatment effect of interest, and at least one of the remaining local ones, all being of positive sign. Values of such coefficient greater than 1 thus indicate that positive values of the parameter ϑ_1 which are plausible in view of the marginal likelihood $f(\mathbf{y}_1 | \vartheta_1, A_1^+)$ receive even stronger support *a posteriori* from some of the other subgroups rather than *a priori*. The role of the factor $\mathcal{H}^-(\cdot)$ is to be interpreted in an analogous fashion, the local effects involved now being negative.

3. An Application to Binomial Data

The validity of the tools introduced in the previous section can be better appreciated if referred to a specific example. Suppose the following hierarchical model applies:

$$\left\{ \begin{array}{ll} [y_{rit} | \beta_i, \vartheta_i] & \sim \text{Bin}(n_{it}, \beta_i + \vartheta_i t) & \text{ind., } r = 1, \dots, n_{it} \\ [\beta_i | \vartheta_i] & \sim \mathcal{U}(\max\{0, -\vartheta_i\}, \min\{1, 1 - \vartheta_i\}) & \text{ind., } i = 1, \dots, m \\ [|\vartheta_i| | \mu^*, A_i^*] & \sim \mathcal{B}(h\mu^*, h(1 - \mu^*)) & \text{ind., } i = 1, \dots, m \\ \mu^+, \mu^- & \sim \mathcal{U}(0, 1) & \text{ind.} \end{array} \right.,$$

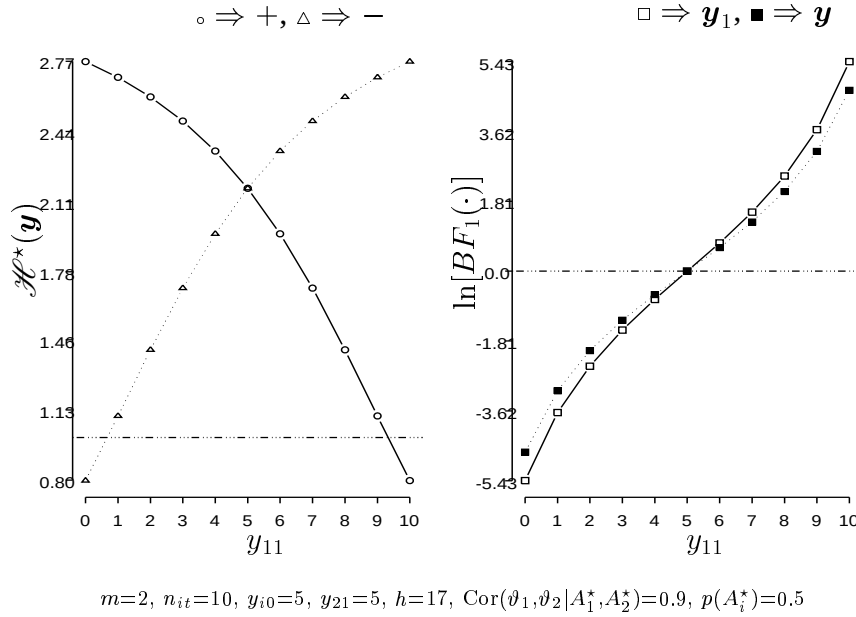
where $h > 0$ is fixed according to the prior correlation $\text{Cor}(\vartheta_i, \vartheta_{i' \neq i} | A_i^*, A_{i'}^*) = \frac{h+1}{h+3}$.

Besides algebraic technicalities (see also Consonni and Veronese, 1995), with the aid of basic MC integration tools the diagrams in 1 were built, according to caption's specifications. The factor $\mathcal{H}^+(\cdot)$ (figure 1, left panel) exhibits a steeper slope as $y_{11} - y_{10}$ grows, according on how well values of ϑ_1 plausible in light of $f(\mathbf{y}_1 | \vartheta_1, A_1^+)$ match the relatively considerable weight placed by subgroup 2 to quite small ϑ_1 positive values. Not surprisingly, its negative counterpart behaves exactly in a symmetric way. The monotonic behaviour of $\ln[BF_1(\mathbf{y}_1)]$ (figure 1, right panel) for increasing values of the observed treatment difference $y_{11} - y_{10}$ looks sensible, as expected in problems like (1). The interaction between the “positive” and “negative” evidence from $\mathbf{y}_2$, re-scaled by the posterior weights in (2), finally results in slightly reducing the emphasis credited to A_1^- by $\mathbf{y}_1$ alone, being as subgroup 2 mostly supports $\vartheta_2 \cong 0$ values; similarly, the stronger the evidence provided by subgroup 1 in favour of A_1^+ , the heavier the charge on it due to the weaker unbalance of $\mathbf{y}_2$ towards ϑ_2 values of corresponding magnitude of the implicit ϑ_1 .

4. Conclusions

In the present work, a Bayesian procedure has been outlined for adapting inferences about local treatment effects in clinical trials, in light of the general pattern emerging from the remaining patients' profiles. The method relies on an assumption of exchangeability of

Figure 1: Bayes factors on Binomial counts



such effects, leading to a meaningful isolation of the contribution provided by the various subgroups to both the competing hypotheses in (1); an artificial example on Binomial count data illustrates the ability of the technique in partially protecting against conclusions laid down by extreme data.

Further developments aim in examining normal likelihoods for large scale trials, and in introducing the more realistic assumption of partial exchangeability between the effects, in order to stress the importance of homogeneous subgroups of individuals.

References

- Byar D.P. (1985) Assessing apparent treatment–covariate interactions in randomized clinical trials, *Statistics in Medicine*, 4, 255–263.
- Consonni G. and Veronese P. (1995) A Bayesian method for combining results from several binomial experiments, *Journal of the American Statistical Association*, 90, 935–944.
- De Santis F. and Spezzaferri F. (1998) Bayes factors and hierarchical models, *Journal of Statistical Planning and Inference*, 74, 323–342.
- Dixon D.O. and Simon R. (1991) Bayesian subset analysis, *Biometrics*, 47, 871–881.