SMALL GROUPS and QUESTIONNAIRES

L.F. Lemmens
Universiteit Antwerpen Departement Fysica & StatUa
CGBU415
Groenenborgerlaan 171 B 2020 Antwerpen, Belgium
(lucien.lemmens at ua.ac.be)
March 31, 2008

Most administrations want to have surveys on the quality of services provided by their officials. The questionnaire technique is often used. For a number of items one asks the respondents to indicate how strongly they agree or disagree with a given statement. Usually several items form a dimension – a name given to an essential part of the service – and the survey of the dimensions reports a summary of the attitude of the respondents. This summary is used to evaluate the performance of the official and can have consequences for promotion. Because these surveys can have consequences, they are contested when the groups are small: a classical analysis rarely avoids the use of the central limit theorem or the law of large numbers. In bayesian statistics, however, the inverse probability problem is readily solved given the likelihood function of the problem and prior density and the evidence follows as usual from normalization.

Assume that a respondent can take 6 attitudes for an item, the information we want to obtain is then $\{N_i \quad i \in [1, \cdots, 6]\}$ with $\sum_i N_i = N$ where N is the number of possible respondents in the complete group. The information we obtain in the questionnaire is $\{n_i \quad i \in [1, \cdots, 6]\}$ with $\sum_i n_i = n$ where n is the number of respondents for that item. The knowledge about $\{n_i\}$ will be used to guess the $\{N_i\}$ This model has a multivariate hypergeometric density with the N_i as parameters. The prior starts from an educated guess that predicts the $\{N_i\}$ without using results from the questionnaire on that item. The most convenient density is a multinomial with given $\{p_i\}$ where the p_i indicate the plausibility that a respondent takes the attitude i. Combining this setting for an item to a model for a dimension we can use the posterior of the first item as a prior for the second item and so on. This leads, for the dimension, to a Dirichlet-model, that belongs to the exponential family. Hence the results are obtained by upgrading, avoiding numerical integration for the calculation of the evidence.

Although the statistical analysis is computationally simple, there are a lot of surveys to be analyzed and communicated to decision makers. A relatively simple R–code was written to automize the analysis and decision theoretical arguments are used to implement a representation of the uncertainties on the data graphically.