SciaNews
Biometric Society E:\WORD\CLARE\ARTICLE.DOC June 27, 1995

Joint Meeting of the International Biometric Society (ENAR)

The Joint Meeting of the International Biometric Society (ENAR) was held in Birmingham, Alabama from March 26-29, 1995. This conference consisted of sessions covering various areas of bio-statistics as well as a programme of several continuing education courses. Some of these presentations will be summarized in this article, as well as an overview of the course *"Applied Bayesian Statistics"*.

One of the educational courses offered at this meeting was an introduction into the fundamental concepts of Applied Bayesian Statistics. Bayesian statistics has had a resurgence in popularity because it is practical for solving problems, conceptually simple and provides a framework for combining multiple sources of information. In brief, Bayesian statistics involves a *prior distribution* and a *likelihood function*. The *prior distribution* represents information about the parameter of interest which is known **prior** to observing the data. In contrast, the *likelihood function* represents information on observed data. The cornerstone of Bayesian statistics comes from Bayes' Theorem which provides a method for combining the *prior distribution* with the *likelihood function*. The result is known as the *posterior distribution*.

This one day course featured two speakers, Robert E. Kass and Larry Wasserman, from Carnegie Mellon University. Although the course covered several topics, the emphasis was on current methods for calculating features of the *posterior distribution*. Unless the *posterior distribution* belongs to some common family of distributions, estimating features (e.g. mean) of the *posterior* is often difficult with the difficulty increasing in higher dimensions. In some situations, it is possible to obtain an asymptotic approximation. However, when the *posterior distribution* is intractable, one of the simplest methods is to perform *posterior simulations*.

In *posterior simulation*, a sample of observations is drawn from the *posterior*, and this sample is then used to characterize features of the *posterior distribution*. For example, in order to obtain an 100(1-)% confidence interval, one would simply sort the sample data and then cut off /2 % of the data from each side. Two of the more traditional methods used in *posterior simulation* are *direct simulation* and *importance sampling*. With the former method, one draws samples directly from the *posterior distribution*. If it is not possible to sample from the *posterior*, one obtains a sample from a density function (called the importance function) from which it is easier to obtain observations. The general idea is to sample from the "wrong" density and correct the estimate by using a weighted sample. This method forms the basis of *importance sampling*. However, the following should be kept in mind : the importance function should be easy to sample from, should have thicker tails that the *posterior distribution* and should approximate the *posterior distribution*.

Among the current methods discussed in posterior simulation were *Gibbs sampling*, *Data augmentation* and the *Metropolis/Hastings* method. *Gibbs sampling* involves iteratively drawing from one thing while holding the other fixed. For example, if f(x,y) is the function from which a sample is to be drawn, arbitrary starting values are first selected for x0 and y0. For each iteration, xi is drawn from f(x|yi-1) and then yi from f(y|xi). This allows the next value of x to be generated while holding the previous generated value of y as fixed. The resulting observations after performing the iteration N times is then a sample from f(x,y). In some problems, data is often incomplete or missing. Data augmentation involves adding the incomplete data as part of the estimation problem. In general, data augmentation simplifies Gibbs sampling by introducing latent variables.. The final method discussed, known as the *Metropolis/Hastings* method, involves

drawing a random sequence of points in such a way that the distribution of the points is approximately the posterior distribution.

The remainder of the course briefly touched on *Hierarchical Models*. *Hierarchical models* refer to models in which the family of data densities is indexed by the parameter of interest and the parameter itself is assumed to be distributed according to a parametric family of densities indexed by some other parameter. Such models are used when there is obvious variability among values of the parameter - e.g. when the parameter varies from subject-to-subject. Other topics which were not discussed but well-documented in course notes included Bayes Factor and Sensitivity Analysis. The latter in particular is an interesting area since it deals with the effects of using different priors in the same dataset. In a sense it determines the robustness of the analysis due to mis-specification of the prior distribution.

In conclusion, the course on Applied Bayesian Statistics was well presented, informative and well-documented with notes and references directed towards further reading.

The session on *"New Designs for Dose-Response Studies"* addressed the fallacies of traditional methods used in Phase I clinical trials and motivated the need for alternative designs. There is general agreement that current phase I designs do not achieve the objectives of the trial such as estimating the MTD, minimizing the number of subjects in the trial and avoiding over/under treatment of subjects. Of the three papers presented, two described design approaches based on the Bayesian framework while the remaining paper discussed an adaptation of randomized up-and-down designs.

In his presentation, J. O'Quigley (University of California at San Diego) briefly described his continuing work on what is known as the Continual Reassessment Method (CRM). The CRM is a Bayesian approach to estimating the maximum tolerated dose, thereby efficiently allocating patients. As each patient is entered into the trial, the CRM reassesses the maximum tolerated dose and treats the next patient at the dose level which the current estimate indicates to be the most appropriate. In a recent paper comparing a *modified* version of the CRM to the traditional "3+3" methods, the authors concluded that the behaviour of the CRM was inferior to traditional methods. The following modifications had been made : the first patient was always treated at dose level 1; skipping dose levels between patients was prohibited and the trial is stopped when the recommended dose level for the next patient was a dose at which a pre-specified number of patients (K) had already been treated. Based on their simulations, a trial under the CRM would require the trial duration to be prolonged due to the CRM's single patient entry accrual. Furthermore, the paper indicated that more patients were treated at higher dose levels when allocated according to the CRM scheme. O'Quigley commented on these conclusions and stressed that the results were inconsistent with his own findings. In response to the CRM allocating patients to higher dose levels, compared to standard methods, O'Quigley questioned the methods of the comparison. The paper reported that the CRM allocated 12% of patients to higher dose levels, compared to the traditional method's 7%. Based on these figures, the article concluded that the CRM would treat 70% ((12% - 7%)/7% = 70%) more patients at higher dose levels. With regard to the CRM prolonging the trial, O'Quigley presented simulation results which showed that the average number of patients used in CRM trials would be similar to that of the standard method. In conclusion, O'Quigley stressed that the original CRM had been modified by the authors and did not represent a valid comparison between the designs.

The second paper presented by R.K. Tsutakawa and D. Sun (University of Missouri) discussed some design considerations in response experiments when a logistic model is implemented within a Bayesian framework. One concern is the predicted posterior variance which remains an unreliable part of design selection. That is, the estimated posterior variance may be far greater than the expected posterior variance. Tsutakawa and Sun suggest that this can be partially overcome by minimizing the expected posterior variance. In order to protect again unexpected large posterior variances (reflecting uncertainty in the estimates of the posterior distribution), a

penalty function is introduced into the design.

N. Flournay et al. (American University) described an adaptive design for sequentially allocating dose levels to patients. The response is taken to be binary (toxic, non-toxic) and the design involves what is known as the "biased coin rule" (BCR) which is used to assign dose levels to patients around a target quantile. The target toxicity is first defined and a random probability is assigned to the coin with the bias as a function of the target toxicity.

The discussion led by Peter F. Thall (University of Texas) further stressed the objectives of Phase I trials and the inadequacies of traditional designs which he referred to as *"idiot designs"*. With respect to Tsutakawa and Sun's design, he commented that it is only applicable to animal experiments or in non-life threatening trials. As for the CRM and BCR, he stated that both designs perform reasonably well and certainly out-perform the traditional methods based on his own research. However, in his final statement, he strongly recommend the CRM when conducting a Phase I trial.

The session on *"Exact tests for Categorical Data"* included a paper addressing two common analyses of 2x2 tables. For such tables, tests of homogeneity and/or Mc Nemar's test (for matched pairs data) are usually conducted. The exact test for homogeneity is normally Fisher's exact test, based on the hypergeometric distribution. The corresponding unconditional test is based on the binomial distribution. In the case of matched pairs data, the exact test is based on the binomial distribution, conditioning on the marginal totals, while the unconditional test utilizes the multinomial distribution. Compared to unconditional tests, exact conditional tests eliminate nuisance parameters and usually generalize to give exact inference.

R. Berger (North Carolina State University) described more powerful unconditional tests for these two situations (homogeneity and matched pairs) and compared his approach with the current exact conditional and unconditional tests. Berger proposed that instead of maximizing over the entire parameter space of the nuisance parameter (as done in the unconditional test), one could maximize over a reduced set of values. The reduced parameter space is a 100(1 -)% confidence interval for the parameter, where is a specified error probability of the interval. The p-value computed from this approach would be adjusted upwards by . As a result, the value of should be fairly small, or alternatively, the confidence interval should have a high coverage probability - usually 99.9% or higher. This approach offers several advantages : computations are reduced because of the smaller set, values of the nuisance parameter can be ignored if not supported by the data and the test is often uniformly more powerful than the usual exact unconditional test. Based on a comparison between his method and current conditional/unconditional tests, Berger drew the following conclusions : the two unconditional tests were generally more powerful than the conditional test and both unconditional tests were comparable in terms of performance, although in some situations, the proposed method was often better. This approach appears to make computations more feasible and could potentially be extended to handle larger contingency tables.

In other presentations, J. Forster (University of Southampton) described applicability of Markov chain Monte Carlo methods, such as Gibbs sampling and the Metropolis/Hastings algorithm, in exact inference. Such methods are used to generate a sample from the exact distribution when the distribution cannot be enumerated. The empirical distribution is then used as an approximation. Forster demonstrated this approach using a logistic regression example.

K. Hirif reviewed published applications of the fast fourier transform (FFT) - a numerical computational algorithm for evaluating polynomials - in computing exact distributions and significance levels for multinomial data. In his discussion the applications were critiqued and advantages/limitations of the FFT approach were presented. He stressed that applications using the FFT method, and published in journals, should be more stringently reviewed and that further work on its applicability is necessary.

In the contributed papers for clinical trials, M. Tan and X. Xiong (Cleveland Clinic Foundation) alternatively discussed monitoring of clinical trials via a sequential conditional probability ratio test. The formulation of the method, steps in design and the reasons for using this approach were explained. Tan proposes to initially design the trial as a (reference) fixed sample test (RFST) using the sequential conditional probability ratio test (SCPRT). Once this is done, the test can be sequentialized by conditioning on what could be achieved with the RFST. Computation of boundaries in this design were briefly outlined. These boundaries are based on defining a maximum discordant probability, which is the essential feature of the SPRT. The maximum discordant probability is the probability that the test statistic crosses the upper boundary before the lower boundary, given that the test statistic ends at the cutoff point for the RFST. Once the boundaries are computed from this value, the resulting power function and expected sample size are checked. If these are not satisfactory, the maximum discordant probability should be re-specified and the boundaries again computed. This process continues until the power function and the expected sample size are acceptable. The achievements of such an approach include a power function identical to that of the RFST, a maximum sample size which does not exceed that of the RFST but is close to a sample size produced by Wald's SPRT, flexibility in the number and size of groups, and no restriction in un-planned interim analyses. This is in contrast to the weakness of current sequential methods which have superficial power when the trial is terminated early. Furthermore, current sequential designs do not agree with RFST in terms of operating characteristics.

C. Chuang-Stein & D. DM. Tong (The Upjohn Company) reviewed three approaches for comparing several treatments with a control when analyzing binary response data. The null hypothesis is that the treatment proportions are equal to that of the control group. The three methods compared were : Asymptotic Freeman-Tukey acceptance region, Binomial acceptance region and Dunnett's acceptance region. Freeman-Tukey's acceptance region is based on an angular transformation of the observed proportions. The binomial acceptance region is based on the binomial distributions estimated under the joint null hypothesis. Finally, with Dunnett's approach, the data is treated as continuous and analyzed using Dunnett's procedure for multiple comparisons. The three procedures were evaluated by simulation. Based on these simulations, the following findings were observed : (1) overall type-I error rate remains acceptable for all three approaches, (2) when the proportion and sample size per treatment group are both small, the Freeman-Tukey approach tends to be conservative, (3) when the proportion is greater than or equal to 0.35 and the sample size is small/moderate, Dunnett's procedure is liberal, and (4) the binomial approach can be liberal when the sample size is equal to 10 and the proportion is greater than or equal to 0.40. Tong recommended that the binomial approach should be used when the proportion exceeds 0.40 and the sample size is between 15 and 25. With smaller sample sizes and a proportion greater than 0.40, the Freeman-Tukey approach should be used. Dunnett's approach is recommended when the sample size is small to moderate and the rate is between 0.05 and 0.30. In all other cases, the Freeman-Tukey approach is sufficient and provides control in the overall Type-I error rate.