Bruce Hoadley, Bell Laboratories, Holmdel

Discussion of Paper by Bruce Hill

I will focus my discussion on the contingency table example because I believe it has the greatest practical value. In particular, I want to compare Hill's model with the superpopulation model.

Let R_j $j = 1,\ldots,K$ be the number of population units in cell j. So $\sum_j R_j = N$ which I assume is known. The vector $\underline{R} = (R_1,\ldots,R_K)$ can be regarded as the unknown parameter in the problem. The data consists of $\underline{y} = (y_1,\ldots,y_K)$, where y_j is the number of sample units in cell j. The sampling distribution of $\underline{y}$ is multivariate hypergeometric, so the Bayesian's problem boils down to the specification of a prior distribution $Pr\{\underline{R}\}$ for $\underline{R}$.

In the superpopulation model, the finite population is regarded as a sample from an infinite superpopulation for which the proportion of units in cell j is Q_j, and $\underline{Q}$ has some prior distribution. In this case, $Pr\{\underline{R}\}$ is a compound multinomial distribution. Furthermore, if $\underline{Q}$ has a Dirichlet distribution, then the posterior distribution of $\underline{R} - \underline{y}$ is compound multinomial and the problem is solved.

The question is: How does Hill's formulation differ from the above? Hill lets M be the number of cells with positive R_j, $\underline{X}$ be the vector of coordinate of $\underline{R}$ which have positive entries (e.g. if $\underline{R} = (0,5,4,0,2)$, then $\underline{X} = (2,3,5)$), and $\underline{L} = (R_{X_{(1)}},\ldots,R_{X_{(M)}})$. There is a one-to-one correspondence between $\underline{R}$ and $(M,\underline{X},\underline{L})$, so a distribution on $\underline{R}$ can be defined by a distribution on $(\underline{M},\underline{X},\underline{L})$. Hill does this by specifying $Pr\{\underline{M}\}$, $P\{\underline{X}|M\}$, and $Pr\{\underline{L}|M,\underline{X}\} = Pr\{\underline{L}|M\}$, where $\underline{L}$ is exchangeable and independent of $\underline{X}$.

Hill pays particular attention to the case where $Pr\{\underline{X}|M\}$ is a uniform distribution. When this is true, $\underline{R}$ is exchangeable, so that the ER_j's are all equal. So in a sense, Hill's model is less general than the superpopulation model where the ER_j's can be different.

In the case where $Pr\{\underline{L}|M\}$ is also a uniform distribution, Hill works out useful expressions for the posterior distribution of $\underline{R}$. If in addition, it was assumed that

$$Pr\{M\} = \binom{K}{M}\binom{N-1}{M-1} \Big/ \binom{N+K-1}{K-1},$$

then $Pr\{\underline{R}\}$ would be uniform, which is a special case of the superpopulation model. This points out a sense in which Hill's formulation is more general; namely, for the superpopulation model, $Pr\{M\}$ has a very special form, whereas in Hill's model it is completely arbitrary.

Two questions that are unresolved in my mind are: (i) What is the practical value of being able to specify $Pr\{M\}$ arbitrarily? (ii) What are some other choices (besides uniform for $Pr\{\underline{X}|M\}$ and $Pr\{\underline{L}|M\}$) that lead to useable expressions for the posterior distribution of $\underline{R}$?

Discussion of Paper by M. R. Novick and P. H. Jackson

Useful insight into the Bayesian method for recovery of collateral information can be obtained by plotting the Bayes and least squares estimates of β_1 for the various colleges. By doing this we see that the Bayes estimates amount to a smoothing of the least squares estimates. The remarkable thing is that the smoothing is more pronounced for the 25% sample than for the 100% sample. One wonders whether this is a property of the method, or just a coincidence. Presumably the Bayesian method is a device for dividing the total between college variation in the least squares estimates of the $\tilde{\beta}$'s into a component due to estimation error and a component due to the between college variation in the $\tilde{\beta}$'s themselves. If so, then I would not expect that the second component would necessarily be underestimated just because the sample size is small.

Another insight from the plot is that the Bayes estimates of β_1 do not always track the least squares estimates. The reason for this is that the components of $\tilde{\beta}$ are being smoothed jointly.

The paper emphasizes that Bayes estimates are much better than the independent least squares estimates in the 25% sample case. The plot suggests that the Bayes estimates would not be much better than the pooled least squares estimates in this same case. Perhaps the Bayesian model should be compared with several alternative classical models.

(The papers by Hill and Novick, discussed above by Mr. Hoadley, were not sent for inclusion in this Proceedings volume.)