The Annals of Statistics
1985, Vol. 13, No. 1, 204–205

DISCUSSION

BY C. L. MALLOWS

AT&T Bell Laboratories

We show that it is helpful to generalize Vardi's problem, estimating an unknown distribution from selection-biased samples, by allowing the sample sizes to be random. The problem then assumes an elegant symmetry, leading to a transparent demonstration of the convergence of Vardi's algorithm.

The following discussion brings out more of the structure of Vardi's (1984) problem, and explains why his algorithm converges.

Consider a multinomial specification, with $s \times t$ cells, and associated probabilities

$$q_{ij} = \alpha_i \beta_j w_{ij} / \sum_{ij} \alpha_i \beta_i w_{ij} \quad 1 \le i \le s, \quad i \le j \le t$$

where $w = \{w_{ij}\}$ is a matrix of known nonnegative constants. The vectors $\alpha = (\alpha_i)$ and $\beta = (\beta_j)$ are nonnegative parameters. Clearly both α and β are identifiable only up to scalar multiples. (We consider below how to resolve these ambiguities.)

Suppose N independent realizations of this specification result in cell frequencies (n_{ij}), with row sums (n_{i+}) and column sums (n_{+j}). We have the following alternative factorizations of the likelihood $L = P\{(n_{ij}) \mid N\ \alpha, \beta, W\}$. The notations are explained in the following paragraphs.

$$L = P\{(n_{i+}), (n_{+j}) \mid N, \alpha, \beta, W\} P\{(n_{ij}) \mid (n_{i+}), (n_{+j}), W\} \tag{1}$$

$$= P\{(n_{i+}) \mid N, \alpha^*\} \prod_{i=1}^{s} P\{(n_{ij}) \mid n_{i+}, \beta, W\} \tag{2}$$

$$= P\{(n_{+j}) \mid N, \beta^*\} \prod_{j=1}^{t} P\{(n_{ij}) \mid n_{+j}, \alpha, W\}. \tag{3}$$

The factorization (1) expresses the fact that $\{(n_{i+}), (n_{+j})\}$ is sufficient for $\{\alpha, \beta\}$. The second factor is

$$\prod_{ij} \left(\frac{w_{ij}^{n_{ij}}}{n_{ij}!} \right) \Big/ \prod_{ij} \left(\frac{w_{ij}^{n_{ij}}}{n_{ij}!} \right)$$

where the sum is over all realizations (n_{ij}) that have the given marginal totals $(n_{i+}), (n_{+j})$.

In (2), the first factor is a s-cell multinomial specification, with probabilities

$$\alpha_i^* = \alpha_i^*(\alpha, \beta, W) = \alpha_i \sum_j \beta_j w_{ij} / \sum_{ij} \alpha_i \beta_j w_{ij} \quad 1 \le i \le s. \tag{4}$$

The remaining factor is the product of s independent t-cell multinomials, the ith of these having probabilities

$$\beta_j w_{ij} / \sum_j \beta_j w_{ij} \quad 1 \le j \le t \tag{5}$$

which are independent of α.

Received January 1984; revised May 1984.
AMS 1980 *subject classification.* 62F10.
Key words and phrases. Multinomial, selection bias.

In (3), the first factor is a t-cell multinomial specification, with probabilities

$$\beta_j^* = \beta_j^*(\alpha, \beta, W) = \beta_j \textstyle\sum_i \alpha_i w_{ij} / \sum_{ij} \alpha_i \beta_j w_{ij} \quad 1 \le j \le t. \tag{6}$$

The remaining factor is the product of t independent s-cell multinomials, the jth of these having probabilities

$$\alpha_i w_{ij} / \textstyle\sum_i \alpha_i w_{ij} \quad 1 \le i \le s \tag{7}$$

which are independent of β.

Vardi works with the second factor in (2), treating β (with the standardization $\sum_j \beta_j = 1$) as an unknown probability distribution. He obtains his maximum likelihood estimate (MLE) (which we shall call $\hat{\beta}_V$) by maximizing this factor. We observe that the above factorizations of L give three alternative ways of expressing the same MLE:

$$(\hat{\alpha}, \hat{\beta}) \text{ maximize (1)}$$

$$(\hat{\alpha}^*, \hat{\beta}) \text{ maximize (2)} \quad (\hat{\beta} = \hat{\beta}_V)$$

$$(\hat{\alpha}, \hat{\beta}^*) \text{ maximize (3)},$$

and these maximizers are mutually consistent, so that

$$\hat{\alpha}^* = \alpha^*(\hat{\alpha}, \hat{\beta}, W), \quad \hat{\beta}^* = \beta^*(\hat{\alpha}, \hat{\beta}, W).$$

Thus to compute $\hat{\beta}_V$ we can choose for convenience to maximize (3) with respect to (α, β^*) and then use (6) to determine $\hat{\beta}$ from $(\hat{\alpha}, \hat{\beta}^*)$.

This is in effect what Vardi does. The second factor in (3) has logarithm $\phi = \sum_i n_{i+} \ln \alpha_i - \sum_j n_{+j} \ln(\sum_i \alpha_i w_{ij}) +$ constant. Let us standardize by fixing $\alpha_s = n_{s+}$; then

$$\frac{\partial \phi}{\partial \alpha_i} = \frac{n_{i+}}{\alpha_i} - \sum_j \frac{n_{+j} w_{ij}}{\sum_i \alpha_i w_{ij}} \quad 1 \le i \le s - 1. \tag{8}$$

Writing A_i for n_{i+}/α_i, the equation $\partial\phi/\partial\alpha_i = 0$ is equivalent to Vardi's $H_i = 1$, $1 \le i \le s - 1$.

We can now see why Vardi's iteration works. From (8), we see that for any fixed (nonnegative) values of $\alpha_i (i \ne i_0)$, $\partial\phi/\partial\alpha_{i_0}$ has exactly one zero, so that ϕ is uniquely maximized when α_{i_0} makes $\partial\phi/\partial\alpha_{i_0}$ vanish. Thus at each step of Vardi's algorithm, ϕ is increased. Once $\hat{\alpha}$ is known, since $\hat{\beta}_j^*$ is trivially n_{+j}/N, (6) gives $\hat{\beta}_j$ proportional to $n_{+j}/\sum_i \hat{\alpha}_i w_{ij}$ where the standardizing factor can be determined from the condition $\sum \hat{\beta}_j = 1$.

REFERENCES

VARDI, Y. (1984). Empirical distributions in selection bias models. *Ann. Statist.* **13** 178–203.

AT&T BELL LABORATORIES
MURRAY HILL, NEW JERSEY 07974