NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

A PROBLEM IN ESTIMATION

By Joseph F. Daly

The Catholic University of America

Several recent psychological studies in the field of memory testing [1], [2], [3] have suggested the following problem. Let each individual E in our population be characterized by the variates $y^1, \cdots, y^p; y^{p+1}, \cdots, y^{p+t}$ $(p > t)$. Suppose, however, that circumstances make it impossible for us to observe the last t variates. For example, we may think of $y^1, \cdots, y^p$ as an individual's scores on a battery of tests, and think of $y^{p+1}, \cdots, y^{p+t}$ as measures of certain psychological characteristics which, though affecting the individual's performance, are not subject to direct observation. To make up for this, assume that we have a theory which tells us that if $y^{p+1}, \cdots, y^{p+t}$ are held constant, then the observable y's are dependent upon them according to a specified regression equation

$$y^i = x^i_\mu y^\mu, \qquad (i = 1, \cdots, p; \mu = p + 1, \cdots, p + t).$$

Somewhat more precisely, we assume the distribution laws

$$(1) \quad f(y^1, \cdots, y^{p+t}) = (2\pi)^{-\frac{1}{2}(p+t)} \mid A_{rs} \mid^{\frac{1}{2}} \exp \{-\tfrac{1}{2}A_{rs}(y^r - a^r)(y^s - a^s)\},$$

(where $r, s = 1, \cdots, p + t$, and repeated indices are to be summed according to the usual convention) and

$$(2) \quad f(y^1, \cdots, y^p \mid y^{p+1}, \cdots, y^{p+t}) = (2\pi\sigma^2)^{-\frac{1}{2}p} \exp\left\{-\frac{1}{2\sigma^2}\sum_i (y^i - x^i_\mu y^\mu)^2\right\}.$$

The x^i_μ are supposed to be known, but except for the conditions imposed by (1) and (2) nothing is known about the quantities A_{rs}, a^r, and σ^2. Having observed the test scores y^i_α $(\alpha = 1, \cdots, N)$ obtained by N individuals E_α drawn at random from the population, we wish to estimate the values $y^{p+1}_\alpha, \cdots, y^{p+t}_\alpha$ corresponding to each E_α, and the essential parameters in the distribution law (1), particularly the variances and covariances of $y^{p+1}, \cdots, y^{p+t}$.

We can easily find optimum estimates of the y^μ_α by applying the method of maximum likelihood to the function (2) after substituting for the y^i the scores y^i_α obtained by the individual in question. Thus if we write

$$v_{\mu\nu} = x_\mu^i x_\nu^i, \qquad \| v^{\mu\nu} \| = \| v_{\mu\nu} \|^{-1},$$

(assuming thereby that the rank of the matrix $\| x_\mu^i \|$ is t) we have

$$\hat{y}_\alpha^\mu = v^{\mu\nu} x_\nu^i y_\alpha^i. \tag{3}$$

These estimates are unbiased in the sense that the expected value of $\hat{y}^\mu$ calculated from the distribution law (2) is y^μ.

But when we come to estimate the variances and covariances involved in (1), the procedure is less straightforward. Under the present circumstances we cannot use the expression

$$\frac{1}{N-1} \sum_\alpha (y_\alpha^\mu - \bar{y}^\mu)(y_\alpha^\nu - \bar{y}^\nu), \tag{4}$$

for the sample covariance of y^μ and y^ν. We might, of course, try substituting the estimates $\hat{y}_\alpha^\mu$ from (3) for the unknown y_α^μ in (4). But this expedient will in general produce a biased estimate. Denoting the required covariance by $A^{\mu\nu}$ (the element in the appropriate position in the inverse of the matrix $\| A_{rs} \|$), we find as a matter of fact that the expected value of (4) when the y_α^μ are replaced by their estimates $\hat{y}_\alpha^\mu$ is

$$A^{\mu\nu} + \sigma^2 v^{\mu\nu}. \tag{5}$$

This bias may or may not be important in any given case. But it can conceivably be quite serious if the $A^{\mu\nu}$ are relatively small, especially if such expressions are employed in the usual way to estimate the correlation coefficient rather than the covariance.

Perhaps the most logical way to attack the problem is through the joint distribution of $y^1, \cdots, y^p$ alone, obtainable by integrating the undesirable variates $y^{p+1}, \cdots, y^{p+t}$ out of (1). We therefore consider

$$f(y^1, \cdots, y^p) = (2\pi)^{-\frac{1}{2}p} \,|\, \bar{A}_{ij} \,|^{\frac{1}{2}} \exp\{-\tfrac{1}{2}\bar{A}_{ij}(y^i - a^i)(y^j - a^j)\}, \tag{6}$$

where

$$\bar{A}_{ij} = A_{ij} - A_{i\mu} B^{\mu\nu} A_{\nu j}, \qquad \| B^{\mu\nu} \| = \| A_{\mu\nu} \|^{-1}.$$

Moreover, when account is taken of (2), we find that we must have

$$A_{ij} = \frac{\delta_{ij}}{\sigma^2} \qquad A_{i\mu} = -\frac{x_\mu^i}{\sigma^2} \qquad a^i = x_\mu^i a^\mu$$

(δ_{ij} being Kronecker's delta). If we now form the likelihood function $\prod_{\alpha=1}^{N} f(y_\alpha^1, \cdots, y_\alpha^p)$ from (6) for our sample, and set its derivatives with respect to the a^μ, σ^2, and the $B^{\mu\nu}$, equal to zero, we arrive, after some simplification, at the equations

$$a^\mu = v^{\mu\nu} x_\nu^i \bar{y}^i = \frac{1}{N} \sum \hat{y}_\alpha^\mu, \qquad \text{[cf. (3)]}$$

$$
(7) \qquad
\begin{aligned}
&\left\{A^{ij} - \frac{1}{N}\sum_{\alpha} (y^i_\alpha - x^i_\mu a^\mu)(y^j_\alpha - x^j_\nu a^\nu)\right\} \delta_{ij} = 0,\\
&\left\{A^{ij} - \frac{1}{N}\sum_{\alpha} (y^i_\alpha - x^i_\mu a^\mu)(y^j_\alpha - x^j_\nu a^\nu)\right\} x^i_\sigma x^j_\tau = 0,\\
&A^{ij} = \sigma^2 \delta^{ij} + x^i_\mu A^{\mu\nu} x^j_\nu,
\end{aligned}
$$

for determining the maximum likelihood estimates. The first of equations (7) is already solved for the a^μ, and the solution of the simultaneous equations for the remaining essential parameters yields the estimates

$$
(8) \qquad \hat{\sigma}^2 = \frac{1}{N(p-t)} \sum_{\alpha,i} (y^i_\alpha - x^i_\mu \hat{y}^\mu_\alpha)^2
$$

$$
(9) \qquad \hat{A}^{\mu\nu} = \frac{1}{N}\sum_{\alpha} (\hat{y}^\mu_\alpha - \hat{a}^\mu)(\hat{y}^\nu_\alpha - \hat{a}^\nu) - v^{\mu\nu}\hat{\sigma}^2.
$$

A considerable amount of algebraic manipulation is required to put the solutions in the form given above; but since the results are about what one would expect in view of (5), we omit the details. As is often the case, some bias remains in the "optimum" estimates (9). However, this can be eliminated by writing $N - 1$ in place of N. The estimate (8) of σ^2 is unbiased as it stands.

REFERENCES

[1] T. V. Moore, "The analysis of association by its equational constants," *Aspects of the New Scholastic Philosophy*, New York, Benziger Brothers, 1932, pp. 181–225.

[2] Regis Holland, "The development of logical and rote memory," *Studies in Psychology and Psychiatry*, the Catholic University of America, Vol. 4, no. 8 (1940).

[3] Helen deSales Forrest, "Correlations between the constants in the curve of learning," *Studies in Psychology and Psychiatry*, the Catholic University of America, Vol. 5, no. 1 (1941).

CONFIDENCE LIMITS FOR AN UNKNOWN DISTRIBUTION FUNCTION

By A. Kolmogoroff

Moscow, U.S.S.R.

Let $x_1, x_2, \cdots, x_n$ be mutually independent random variables following the same distribution law

$$
(1) \qquad P\{x_i \leq \xi\} = F(\xi).
$$

A recent paper by A. Wald and J. Wolfowitz[1] deals with the problem of using

[1] A. Wald and J. Wolfowitz, "Confidence limits for distribution functions," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 105–118.