ABSTRACTS OF PAPERS

(Presented on September 10, 1940, at the Hanover meeting of the Institute)

Contributions to the Theory of the Representative Method of Sampling. WILLIAM G. MADOW, Washington, D. C.

The theory of representative sampling may be regarded as a dual sampling process; the first of which consists in the sampling of different random variables and the second of which consists in repeating several times the experiments associated with each of the different random variables. It follows that while the theory of sampling from finite populations without replacement may be required for the first process, the second leads directly into the theory of sampling from infinite populations. There is, however, one difference. Although the usual theory is concerned with the evaluation of fiducial or confidence limits for parameters the theory of sampling is concerned with the evaluation of fiducial or confidence limits for, say, the mean of a sample of N, when n, $(N \geq n)$, of the values are known.

It is thus possible to use the usual theories of estimation in obtaining estimates of the parameters and to allow the effects of subsampling process to show themselves in the different values of the fiducial limits. It is shown that the limits obtained are almost identical with those obtained by the theory of sampling from a finite population. Distributions of the statistics used in these limits are derived.

Besides these results, the theory is extended to the theory of sampling vectors, and conditions are stated under which the "best" allocation of the number in a sample among several strata is proportional to the kth roots of the generalized variance of a random vector having k components.

A Generalization of the Law of Large Numbers. HILDA GEIRINGER, Bryn Mawr.

Let $V_1(x)$, $V_2(x)$, $\cdots$, $V_n(x)$ be n probability distributions which are not supposed to be independent and let $F(x_1, x_2, \cdots, x_n)$ be a "statistical function" of n observations in the sense of v. Mises,—$V_i(x)$ $(i = 1, 2, \cdots n)$ indicating as usual the probability of getting a result $\leqq x$ at the ith observation—. Then *it can be proved* that under fairly general conditions $F(x_1, x_2, \cdots, x_n)$ *converges stochastically toward its "theoretical value"*; or in other words, that under these general conditions *a great class of statistics* $F(x_1, x_2, \cdots, x_n)$ *is "consistent"* in the sense of R. A. Fisher.

Well known particular cases of this theorem result if (a) we take for $F(x_1, x_2, \cdots, x_n)$ the *average* $(x_1 + x_2 + \cdots + x_n)/n$ of the n observations, (b) we assume that the $V_i(x)$ are *independent* distributions.

On the Problem of Two Samples from Normal Populations with Unequal Variances. S. S. WILKS, Princeton University.

Suppose O_{n_1} and O_{n_2} are samples of n_1 and n_2 elements from normal populations π_1 and π_2 respectively. Let a_1, σ_1^2 and a_2, σ_2^2 be the means and variances of π_1 and π_2 and let O_{n_1} and O_{n_2} have means $\bar{x}_1$ and $\bar{x}_2$ and variances s_1^2 and s_1^2 (unbiased estimates of σ_1^2, σ_2^2) respectively. It is shown that there exists no function (Borel measurable) of $\bar{x}_1$, $\bar{x}_2$, s_1^2, s_2^2, $a_1 - a_2$ independent of σ_1 and σ_2, having its probability law independent of the four population parameters. It is therefore impossible to obtain exact confidence limits for $a_1 - a_2$ corresponding to a given confidence coefficient. Functions of the four parameters and four statistics are devised from which one can set up confidence limits for $a_1 - a_2$ with associated confidence coefficient inequalities.

Experimental Determination of the Maximum of an Empirical Function. HAROLD HOTELLING, Columbia University.

In physical and economic experimentation to determine the maximum of an unknown function, for example of a monopolist's profit as a function of price, or of the magnetic permeability of an alloy as a function of its composition, the characteristic procedure is to perform experiments with chosen values of the argument x, each of which then yields an observation, subject to error, on the corresponding functional value $y = f(x)$. The values of x need, however, to be chosen on the basis of earlier experiments in order to make the determination efficient. The experimentation properly proceeds, therefore, in successive stages, with the values used at each stage determined with the help of the earlier work. The question what distribution of x as a function of previous results should be used is discussed in this paper on the basis of various hypotheses regarding the function, and further criteria. In particular, a conflict is shown to exist under some conditions between the criterion of minimum sampling variance and that calling for absence of bias.

Asymptotically Shortest Confidence Intervals. ABRAHAM WALD, Columbia University.

Let $f(x, \theta)$ be the probability density function of a variate x involving an unknown parameter θ. Denote by $x_1, \cdots, x_n$ n independent observations on x and let $C_n(\theta)$ be a positive function of θ such that the probability that $\left| \frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_{\alpha=1}^{n} \log f(x_\alpha, \theta) \right| \leq C_n(\theta)$ is equal to a constant β under the assumption that θ is the true value of the parameter. Denote by $\theta'(x_1, \cdots, x_n)$ the root in θ of the equation $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_{\alpha} \log f(x_\alpha, \theta) = C_n(\theta)$ and by $\theta''(x_1, \cdots, x_n)$ the root of $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} \sum_{\alpha} \log f(x_\alpha, \theta) = -C_n(\theta)$. Under some weak assumptions on $f(x, \theta)$ the interval $\delta_n(x_1, \cdots, x_n) = [\theta'(x_1, \cdots, x_n), \theta''(x_1, \cdots, x_n)]$ is in the limit with $n \rightarrow \infty$ a shortest unbiased confidence interval[1] of θ corresponding to the confidence coefficient β. This confidence interval is identical with that given by S. S. Wilks in his paper "Shortest average confidence intervals from large samples," *The Annals of Mathematical Statistics*, Sept. 1938. Wilks has shown that $\delta_n(x_1, \cdots, x_n)$ is asymptotically shortest in the average compared with all confidence intervals computed on the basis of statistics belonging to a certain class C. In the present paper it has been proved that the confidence interval in question is asymptotically shortest compared with *any* arbitrary unbiased confidence interval, without any restriction to a certain class of functions.

Reduction of Certain Composite Statistical Hypotheses. GEORGE W. BROWN, R. H. Macy and Co., New York.

The results obtained make it possible to reduce a large class of composite statistical hypotheses to equivalent simple hypotheses. The fundamental theorem established states essentially that if two distributions give rise, in sampling, to the same distribution of the

[1] For the definition of a shortest unbiased confidence interval see the paper by J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc.* (1937).

set of differences between observations, then one distribution must be a translation of the other, subject to a condition requiring that the characteristic function of one of the distributions be such that any interior intervals of zeros be not too large. The result is established by means of the functional equation $\varphi(t_1)\varphi(t_2)\varphi(-t_1 - t_2) = \psi(t_1)\psi(t_2)\psi(-t_1 - t_2)$ relating the characteristic functions. Similar results are obtained for scale, and combination of location and scale, and the corresponding situations in multivariate distributions. This type of uniqueness theorem permits one to reduce a composite hypothesis involving an unknown location parameter (or scale, or both) to an equivalent simple hypothesis.

Conception of Equivalence in the Limit of Tests and Its Application to Certain λ- and χ^2-Tests. J. NEYMAN, University of California.

Denote by E a system of observable variables and by N the number of independent observations of those variables to be used for testing a certain statistical hypothesis H against a set Ω of admissible simple hypotheses h. Let further $T_1(N)$ and $T_2(N)$ be two different tests of H using the same number N of observations. Consider the probability $P_N(h)$ calculated on any admissible simple hypothesis h, of the two tests, contradicting themselves.

Definition: If, whatever be $h \epsilon \Omega$, the probability $P_N(h)$ tends to zero as N is indefinitely increased, then the two tests are said to be equivalent in the limit.

Consider a number s of series of independent trials and denote by $E_{i1}, E_{i2}, \cdots, E_{im_i}$ all the m_i possible and mutually exclusive outcomes of each of the trials forming the ith series. Let p_{ij} be the probability of E_{ij}, n_i the total number of trials in the ith series, and n_{ij} the number of these which give the outcome E_{ij}.

Suppose that it is desired to test a composite hypothesis H concerning all the probabilities p_{ij} and consisting of the assumption that any one of them is a given linear function of some t independent parameters θ_k, so that

$$p_{ij} = a_{ij0} + a_{ij1}\theta_1 + \cdots + a_{ijt}\theta_t \tag{1}$$

where the coefficients a_{ijk} are known. The main result of the paper is then that the λ-test of the above hypothesis H, tested against the set Ω of alternatives ascribing to the p_{ij} any non-negative values, is equivalent in the limit to the test consisting of rejecting H when the minimum of the expression

$$\chi^2 = \sum_{i=1}^{s} \sum_{j=1}^{m_i} \frac{(n_{ij} - n_i p_{ij})^2}{n_{ij}} \tag{2}$$

calculated with respect to unrestricted variation of the θ's, exceeds the tabled value of χ_ϵ^2 corresponding to the chosen level of significance ϵ and to the number of degrees of freedom $\sum_{i=1}^{s} m_i - s - t$.

It will be noticed that the expression (2) differs from the usual χ^2 in the denominator of each term.

As an example of the application of the test based on (2), consider the case where M varieties of sugar beet are tested for resistance to a certain disease in an experiment arranged in N randomized blocks. Denote by n the number of beets selected at random for inspection from each plot and by n_{ij} the number of those of the ith variety from the plot in the jth block which are found to be infected. Denote further by p_{ij} the proportion of infected beets of the ith variety in the plot in the jth block. The hypothesis that the effects of variety and of block are additive is expressed by $p_{ij} = p + V_i + B_j$ with $\Sigma V_i = \Sigma B_j = 0$. To test this hypothesis we may use (2) which in this particular case reduces itself to

$$\chi^2 = \sum_{i=1}^{M} \sum_{j=1}^{N} w_{ij}(q_{ij} - p - V_i - B_j)^2 \tag{3}$$

with $w_{ij} = n^3/\{n_{ij}(n - n_{ij})\}$, $q_{ij} = n_{ij}/n$. The minimum χ_0^2 of χ^2 is found by solving a set of equations which are linear in p, V_i, B_j and the comparison of χ_0^2 with the tabled value corresponding to $(M - 1)(N - 1)$ degrees of freedom will tell us whether we are likely to be very wrong in assuming additivity or not. In the favorable case we may next proceed similarly to test another hypothesis that there is no differentiation between the varieties, so that $V_1 = V_2 = \cdots = V_M = 0$.

Empirical Comparison of the "Smooth" Test for Goodness of Fit with the Pearson's χ^2 Test.

J. Neyman, University of California.

In a previous publication[2] the author has deduced a test for goodness of fit, described as the "smooth test" or the ψ^2 test, applicable to cases where the hypothesis tested H is simple. The test is so devised as to be particularly sensitive to departures from H which are "smooth" in the sense explained in detail in the publication quoted. Whether the test so devised does present any advantage over the usual χ^2 test depends on how frequently we meet, in practice, cases where the hypotheses alternative to the one tested are actually smooth.

The present investigation was undertaken with the object of obtaining some information on this point. For that purpose a number of cases described in the literature where there was a question of testing that some observable variable x follows some perfectly specified distribution $p(x)$ were analyzed. Of all such cases, the ones where there were *a priori* theoretical reasons to believe that $p(x)$ could not possibly represent the true distribution of x and, at the most, it could be considered as only an approximation to the true distribution were selected.

It was assumed that the departures from the hypothetical distributions are typical of those that may be met in practice when no definite information as to the actual state of affairs is available. The hypothesis of goodness of fit was tested both by means of the χ^2 and by the fourth order smooth test. Out of the 130 cases studied the two tests were in perfect agreement eight times. Out of the remaining 122 cases the smooth test proved to be more sensitive than the χ^2 in 70 cases and the χ^2 better than the smooth test in 52 cases. We may further compare the tests by counting those cases where one of them detected the falsehood of the hypothesis tested at a given level of significance while the other failed to do so. At the level of significance .05 the χ^2 test rejected the hypothesis tested 13 times, while P_{ψ^2} was $>.05$. The reverse was true in 17 cases. At the level of significance .01 the corresponding figures are 5 and 14, again in favor of the smooth test.

[2] J. Neyman, " 'Smooth Test' for Goodness of Fit." *Skandinavisk Aktuarietidskrift*, 1937, pp. 149–199.