$$[|\phi_m(x_{-m}, \cdots, x_0) - \phi_n(x_{-n}, \cdots, x_0)| > \epsilon]$$

has the same measure as the set $[|\phi_m(x_m, \cdots, x_0) - \phi_n(x_n, \cdots, x_0)| > \epsilon]$. Hence $\phi_n(x_{-n}, \cdots, x_0)$ converges in probability and Theorem 1 applies.

The above theorem is an extension of the Hewitt-Savage zero-one law for symmetric sets, as the following theorem makes clear.

THEOREM 5. *Let $x_0, x_1, \cdots$ be a sequence of identically distributed, independent random variables and f any integrable function on the process such that f is invariant under finite permutations of the coordinates. Then f is a.s. constant.*

PROOF. Let $\phi_n(x_0, \cdots, x_n) = E(f \mid x_0, \cdots, x_n)$. Then $\phi_n(x_n, \cdots, x_0) = \phi_n(x_0, \cdots, x_n)$ by the symmetry of f and the $\phi_n(x_0, \cdots, x_n)$ sequence forms a martingale which converges a.s. to f. The conclusion follows from Theorem 3.

REFERENCES

[1] E. HEWITT AND L. J. SAVAGE, "Symmetric measures on Cartesian products," *Trans. Amer. Math. Soc.*, Vol. 80 (1955), pp. 470–501.
[2] C. KRAFT, "Some conditions for consistency and uniform consistency of statistical procedures," *Univ. of California Publ. Stat.*, Vol. 2, No. 6 (1955), pp. 125–141.
[3] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons, New York, 1953.
[4] M. LOÈVE, *Probability Theory*, D. Van Nostrand, New York, 1955.

A TEST OF FIT FOR MULTIVARIATE DISTRIBUTIONS[1]

BY LIONEL WEISS

Cornell University

1. Summary and introduction. Suppose X is a chance variable taking values in k-dimensional Euclidean space. That is, $X = (Y_1, \cdots, Y_k)$, where Y_i is a univariate chance variable. The joint distribution of $(Y_1, \cdots, Y_k)$ has density $f(y_1, \cdots, y_k)$, say.

We shall call a function $h(y_1, \cdots, y_k)$ "piecewise continuous" if it is everywhere bounded, and k-dimensional Euclidean space can be broken into a finite number of Borel-measurable subregions, such that in the interior of each subregion $h(y_1, \cdots, y_k)$ is continuous, and the set of all boundary points of all subregions has Lebesgue measure zero.

We assume that $f(y_1, \cdots, y_k)$ is piecewise continuous. Let $h(y_1, \cdots, y_k)$ be some given nonnegative piecewise continuous function, and let $X_1, \cdots, X_n$ be independent chance variables, each with the density $f(y_1, \cdots, y_k)$. Choose a nonnegative number t, and for each i, construct a k-dimensional sphere with center at $X_i = (Y_{i1}, \cdots, Y_{ik})$ and of k-dimensional volume

$$\frac{th(Y_{i1}, \cdots, Y_{ik})}{n}.$$

Such a sphere will be called "of type s" if it contains exactly s of the $(n - 1)$

Received August 8, 1957.
[1] Research sponsored by the Office of Naval Research.

points $X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n$. Let $R_n(t; s)$ denote the proportion of the n spheres which are of type s.

For typographical simplicity, we denote the vector $(y_1, \cdots, y_k)$ by y. Let $S(t; s)$ denote the multiple integral

$$(t^s/s!) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h^s(y) f^{s+1}(y) \exp\{-th(y)f(y)\}\, dy_1 \cdots dy_k.$$

It is shown that $R_n(t; s)$ converges stochastically to $S(t; s)$ as n increases. This result is then used to construct a test of the hypothesis that the unknown density $f(y)$ is equal to a given density $g(y)$.

2. Proof of the convergence of $R_n(t; s)$. We define the chance variable Z_i to be equal to one if the sphere centered at X_i is of type s, and to be equal to zero otherwise. $R_n(t; s) = (1/n)(Z_1 + \cdots + Z_n)$.

Let $V(v; y)$ denote the probability assigned by the density $f(y)$ to the sphere of volume v and center at y. In any closed region in which $f(y)$ is continuous, $V(v; y)$ can be written as $vf(y) + v\epsilon(y; v)$, where $\epsilon(y; v)$ approaches zero as v approaches zero, uniformly in y throughout the region. Clearly, $E\{Z_i\}$ is equal to

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{(n-1)!}{s!(n-1-s)!} \left[V\left(\frac{th(y)}{n}; y\right)\right]^s \cdot \left[1 - V\left(\frac{th(y)}{n}; y\right)\right]^{n-1-s} f(y)\, dy_1 \cdots dy_k. \tag{2.1}$$

The region of integration can be broken into a finite number of subregions, in the interior of each of which $f(y)$ and $h(y)$ are continuous. A closed subset of each such subregion can be found so that the measure of the set of points in k-dimensional space outside these closed subsets is arbitrarily small. Within each such closed subset, we may write

$$V\left(\frac{th(y)}{n}; y\right) = \frac{th(y)}{n} f(y) + \frac{th(y)}{n} \epsilon\left(y; \frac{th(y)}{n}\right),$$

where $\epsilon(y; th(y)/n)$ approaches zero as n increases, uniformly in y in the closed subset. Then it follows easily that the multiple integral (2.1) converges to $S(t; s)$ as n increases, so that $E\{R_n(t; s)\}$ approaches $S(t; s)$ as n increases.

To complete the proof that $R_n(t; s)$ converges stochastically to $S(t; s)$, we shall show that Var $\{R_n(t; s)\}$ approaches zero as n increases. Var $\{R_n(t; s)\}$ is equal to

$$\frac{1}{n^2} \sum_{i=1}^{n} \text{Var}\{Z_i\} + \frac{1}{n^2} \sum_{i \neq j} \sum \text{Cov}\{Z_i, Z_j\}.$$

Since $\{Z_i\}$ are uniformly bounded variables, $(1/n^2)\sum \text{Var}\{Z_i\}$ approaches zero as n increases. Therefore, to show that Var $\{R_n(t; s)\}$ approaches zero, it will suffice to show that $E\{Z_i Z_j\}$ approaches $[S(t; s)]^2$ as n increases, since this implies that Cov $\{Z_i, Z_j\}$ approaches zero. $E\{Z_i Z_j\}$ is equal to

$$
\begin{aligned}
(2.2)\qquad & \int \underset{R_1}{\cdots} \int \frac{(n-2)!}{s!s!(n-2-2s)!} \left[V\left(\frac{th(y)}{n}; y\right)\right]^s \left[V\left(\frac{th(z)}{n}; z\right)\right]^s \\
& \cdot \left[1 - V\left(\frac{th(y)}{n}; y\right) - V\left(\frac{th(z)}{n}; z\right)\right]^{n-2-2s} \\
& \cdot f(y)f(z)\, dy_1 \cdots dy_k\, dz_1 \cdots dz_k \\
& + \int \underset{R_2}{\cdots} \int Q(y, z) f(y) f(z)\, dy_1 \cdots dy_k\, dz_1 \cdots dz_k,
\end{aligned}
$$

where R_1 is the region in (y, z) space such that the k-dimensional sphere of volume $th(y)/n$ centered at y does not intersect the k-dimensional sphere of volume $th(z)/n$ centered at z; R_2 is the remainder of (y, z) space; and $Q(y, z)$ is the conditional probability that the spheres around X_i and X_j are both of type s, given that $X_i = y$ and $X_j = z$. Clearly, the second integral in (2.2) approaches zero as n increases, and the first approaches $[S(t; s)]^2$. This completes the proof that $R_n(t; s)$ converges stochastically to $S(t; s)$ as n increases.

3. Application to multivariate tests of fit. Suppose the density of X, $f(y_1, \cdots, y_k)$ say, is unknown, and the hypothesis to be tested is that almost everywhere over a given region R, $f(y_1, \cdots, y_k) = g(y_1, \cdots, y_k)$, where $g(y_1, \cdots, y_k)$ is a given piecewise continuous function, $g(y_1, \cdots, y_k) \geqq B > 0$ at every point of R, and

$$1 \geqq \int \underset{R}{\cdots} \int g(y_1, \cdots, y_k)\, dy_1 \cdots dy_k > 0.$$

The hypothesis says nothing about $f(y_1, \cdots, y_k)$ outside the region R.

To construct a test of this hypothesis, we apply the result of Section 2 with $t = 1$, $s = 1$, and $h(y_1, \cdots, y_k) = 1/g(y_1, \cdots, y_k)$ for $(y_1, \cdots, y_k)$ in R, $h(y_1, \cdots, y_k) = 0$ elsewhere. Then

$$S(1; 1) = \int \underset{R}{\cdots} \int f(y) \frac{f(y)}{g(y)} \exp\left\{\frac{-f(y)}{g(y)}\right\} dy_1 \cdots dy_k.$$

Using the fact that the function ue^{-u} takes on its absolute maximum at $u = 1$, we find that

$$S(1; 1) \leqq e^{-1} \int \underset{R}{\cdots} \int f(y)\, dy_1 \cdots dy_k,$$

with equality holding if and only if $g(y) = f(y)$ almost everywhere on R where $f(y) > 0$. Denote by $Q(n)$ the proportion of the observed points $X_1, X_2, \cdots, X_n$ that fall in the region R. $Q(n)$ converges stochastically to

$$\int \underset{R}{\cdots} \int f(y)\, dy_1 \cdots dy_k$$

as n increases. Thus, if the hypothesis is true, $R_n(1; 1)$ converges stochastically to

$$e^{-1} \int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k,$$

and $Q(n)$ converges stochastically to

$$\int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k.$$

Conversely, if $R_n(1; 1)$ converges stochastically to

$$e^{-1} \int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k$$

and $Q(n)$ converges stochastically to

$$\int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k,$$

then

$$S(1; 1) = e^{-1} \int \underset{R}{\cdots} \int f(y)\, dy_1 \cdots dy_k,$$

so the hypothesis is true.

For a given n, we define the following test T_n of the hypothesis. Accept the hypothesis if and only if

$$\left| R_n(1; 1) - e^{-1} \int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k \right| < A_n$$

and

$$\left| Q(n) - \int \underset{R}{\cdots} \int g(y)\, dy_1 \cdots dy_k \right| < B_n,$$

where A_n, B_n are numbers chosen to give the desired level of significance, and it may (and will) be assumed that A_n and B_n both approach zero as n increases. From the discussion above, it is clear that the sequence of tests $\{T_n\}$ is consistent against any piecewise continuous alternative $f(y)$. To set the exact values of A_n, B_n the joint distribution of $Q(n)$ and $R_n(1; 1)$ would be required, but this distribution is unknown, although the author conjectures that it is asymptotically normal. However, given the function $g(y)$, the region R, and an alternative $f(y)$, the integrals (2.1) and (2.2) can be evaluated, at least approximately, and then Chebyshev inequalities can be used to give an upper bound to the level of significance and a lower bound to the power, for a given choice of A_n and B_n.

There are other consistent tests for the hypothesis under discussion: the chi-square test and obvious extensions of the univariate Kolmogorov-Smirnov and von Mises tests. A comparison of the power functions of these tests would be of great interest. Almost nothing is known of the small-sample power of any of these tests. The large-sample power of the chi-square test is known. It is the author's conjecture that the limiting joint distribution of $Q(n)$ and $R_n(1; 1)$ is bivariate normal under the alternatives as well as under the hypothesis. If this conjecture could be proved, the asymptotic power of the proposed test would be known.

TABLES FOR OBTAINING NON-PARAMETRIC TOLERANCE LIMITS

By Paul N. Somerville[1]

General Analysis Corporation, Sierra Vista, Arizona

The general consideration of non-parametric tolerance limits had its origin with Wilks [10]. Wilks showed that for continuous populations, the distribution of P, the proportion of the population between two order statistics from a random sample, was independent of the population sampled, and was in fact a function only of the particular order statistics chosen. Wald [9] and Tukey [8] extended the method to multivariate populations, Tukey being responsible for the term "statistically equivalent block." Their work was extended further by Fraser [2], [3]. Murphy [4] presented graphs of minimum probable coverage by sample blocks determined by order statistics of a sample from a population with a continuous but unknown c.d.f. This note extends the results of Murphy, and tabularizes the results in a manner which eliminates or minimizes interpolation, particularly with respect to m, in a large number of cases. The form of Table I parallels the tables of Eisenhart, Hastay and Wallis [1] "Tolerance Factors for Normal Distributions."

Let P represent the proportion of the population between the r^{th} smallest and the s^{th} largest value in a random sample of n from a population having a continuous but unknown distribution function. Table I gives the largest value of $m = r + s$ such that we have confidence of at least that 100 P percent of the population lies between the r^{th} smallest and s^{th} largest in the sample. Note, that we may choose any $r, s \geqq 0$ such that $r + s = m$. We must, of course, decide upon the values of r and s independently of the observations in the sample. We obtain one-sided confidence intervals when we use $r = 0$ or $s = 0$ for a given m. The values of m are the largest such that

$$\gamma \leqq I_{1-P}(m, n - m + 1)$$

where I is the incomplete Beta function tabulated in [5] and [7].

Received March 21, 1957; revised January 15, 1958.

[1] Part of this work was done while the author was a guest worker at the National Bureau of Standards.