REFERENCES

[1] H. Pollard, "Distribution functions containing a Gaussian factor," *Proc. Am. Math. Soc.*, Vol. 4 (1953), pp. 578–582.
[2] G. Szego, *Orthogonal Polynomials*, American Mathematical Society New York, 1939.
[3] S. Bochner, "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse," *Math. Annalen*, Vol. 108 (1933).
[4] W. H. Young, "On multiple integration by parts and the second theorem of the mean," *Proc. London Math. Soc.*, Vol. 16 (1917), p. 276.

A CERTAIN CLASS OF TESTS OF FIT[1]

By Lionel Weiss

University of Oregon

1. Summary and introduction. Suppose $X_1, X_2, \cdots, X_n$ are known to be independently and identically distributed, each with the density function $f(x)$, with $\int_0^1 f(x)\,dx = 1$. Let $Y_1 \leqq Y_2 \leqq \cdots \leqq Y_n$ be the ordered values of $X_1, X_2, \cdots, X_n$, and define $W_1 = Y_1$, $W_2 = Y_2 - Y_1, \cdots, W_n = Y_n - Y_{n-1}$, and $W_{n+1} = 1 - Y_n$, so that $W_1 + \cdots + W_{n+1} = 1$. Finally, define $Z_1, \cdots, Z_{n+1}$ as the ordered values of $W_1, \cdots, W_{n+1}$, so that $0 \leqq Z_1 \leqq Z_2 \leqq \cdots \leqq Z_{n+1}$, with $Z_1 + \cdots + Z_{n+1} = 1$. We are going to test the hypothesis that $f(x) = 1$ for $0 < x < 1$, and we are going to consider only tests based on $Z_1, Z_2, \cdots, Z_n$. The intutitive justification for this is that, roughly speaking, deviations from the hypothesis on any part of the unit interval are treated alike. Several authors have discussed tests based on $Z_1, \cdots, Z_n$. (See references [1], [2], [3].)

If u is a number greater than unity, it is shown that the test of the form "reject the hypothesis if $Z_1^u + \cdots + Z_{n+1}^u > K$" is consistent against a very wide class of alternatives. When $u = 2$, the resulting test has some desirable properties with respect to alternatives with linear density functions.

2. The distribution of $Z_1, \cdots, Z_n$. It is easily seen that $P[Z_i = Z_j$ for any $i \neq j]$ is equal to zero. We want to find the joint density function $h(z_1, \cdots, z_n)$ of $Z_1, \cdots, Z_n$. The joint density function of $W_1, \cdots, W_n$ is equal to $n!\, f(w_1)f(w_1 + w_2) \cdots f(w_1 + w_2 + \cdots + w_n)$ in the region $w_i \geqq 0$, $w_1 + \cdots + w_n \leqq 1$, and is equal to zero elsewhere. Let $\{j(1), j(2), \cdots, j(n+1)\}$ be any permutation of the first $n + 1$ integers, and let $\sum_p$ denote summation over all the $(n+1)!$ permutations. Given any set of numbers $0 < z_1 < z_2 < \cdots < z_n < 1 - (z_1 + \cdots + z_n)$, we denote by $Q[j(1), j(2), \cdots, j(n+1)]$ the conditional probability that $W_i = z_{j(i)}$ for $i = 1, \cdots, n+1$, given that $Z_i = z_i$ for $i = 1, \cdots, n+1$. It is understood that if $j(i) = n + 1$, then $z_{j(i)} = 1 -$

Received August 22, 1955; revised June 20, 1956.

[1] Research under contract with the Office of Naval Research.

$(z_1 + \cdots + z_n)$. For each set of values $z_1, \cdots, z_n$ for which $h(z_1, \cdots, z_n)$ is positive, we have

$$Q[j(1), \cdots, j(n+1)] = \frac{n!f(z_{j(1)})f(z_{j(1)} + z_{j(2)}) \cdots f(z_{j(1)} + \cdots + z_{j(n)})}{h(z_1, \cdots, z_n)}.$$

Since $\sum_p Q[j(1), \cdots, j(n+1)] = 1$, for each set of values $z_1, \cdots, z_n$ for which $h(z_1, \cdots, z_n)$ is positive, we have $h(z_1, \cdots, z_n) = n! \sum_p f(z_{j(1)})f(z_{j(1)} + z_{j(2)}) \cdots f(z_{j(1)} + \cdots + z_{j(n)})$. Now let D be the region in $(z_1, \cdots, z_n)$-space where the following three conditions are satisfied:

(1) $0 \leqq z_1 \leqq z_2 \leqq \cdots \leqq z_n \leqq 1 - (z_1 + \cdots + z_n)$,

(2) $h(z_1, \cdots, z_n) = 0$,

(3) $n! \sum_p f(z_{j(1)})f(z_{j(1)} + z_{j(2)}) \cdots f(z_{j(1)} + \cdots + z_{j(n)}) > 0$.

Then D must be of measure zero. For if D is of positive measure, then

$$\int \cdots_D \int n! \sum_p f(z_{j(1)})f(z_{j(1)} + z_{j(2)}) \cdots f(z_{j(1)} + \cdots + z_{j(n)})\, dz_1 \cdots dz_n > 0,$$

which implies that $P[(W_1, \cdots, W_n) \text{ in } D] > 0$, which in turn implies that $P[(Z_1, \cdots, Z_n) \text{ in } D] > 0$, which is a contradiction. Therefore we have shown that if $0 \leqq z_1 \leqq z_2 \leqq \cdots \leqq z_n \leqq 1 - (z_1 + \cdots + z_n)$, then

$$h(z_1, \cdots, z_n) = n! \sum_p f(z_{j(1)})f(z_{j(1)} + z_{j(2)}) \cdots f(z_{j(1)} + \cdots + z_{j(n)}), \tag{2.1}$$

while $h(z_1, \cdots, z_n)$ is zero for other values of $z_1, \cdots, z_n$. We note that when $f(x) = 1$, the right-hand side of (2.1) is equal to $n!(n + 1)!$.

3. Properties of the power of tests based on $Z_1, \cdots, Z_n$. Let $r(x)$ be a given bounded measurable function of x satisfying the conditions

$$\int_0^1 r(x)\, dx = 0, \quad \int_0^1 r^2(x)\, dx > 0.$$

Then for δ small enough in absolute value, $1 + \delta r(x)$ is a density function on $(0, 1)$. For any given measurable region R in $(z_1, \cdots, z_n)$-space, we denote by $M(R, \delta)$ the probability that $(Z_1, \cdots, Z_n)$ will fall in R, assuming the density of the original observations is equal to $1 + \delta r(x)$. In what follows, we shall always assume that R is a subset of the region

$$0 \leqq z_1 \leqq \cdots \leqq z_n \leqq 1 - (z_1 + \cdots + z_n).$$

For any given region R, we have

$$\left.\frac{dM(R, \delta)}{d\delta}\right]_{\delta=0} = n! \int \cdots_R \int \sum_p \sum_{k=1}^n r(z_{j(1)} + \cdots + z_{j(k)})\, dz_1 \cdots dz_n, \tag{3.1}$$

$$\left.\frac{d^2M(R, \delta)}{d\delta^2}\right]_{\delta=0} = 2n! \int \cdots_R \int \sum_p \sum_{1 \leqq k < L \leqq n}\sum r(z_{j(1)} + \cdots + z_{j(k)}) \cdot r(z_{j(1)} + \cdots + z_{j(L)})\, dz_1 \cdots dz_n. \tag{3.2}$$

Equations (3.1) and (3.2) follow easily when $f(x)$ in (2.1) is replaced by $1 + \delta r(x)$, and the result is expressed as a polynomial in δ.

4. The case of linear $r(x)$. The integrands in (3.1) and (3.2) are complicated for many functions $r(x)$. However, when $r(x) = x - \frac{1}{2}$, we have (remembering $z_{n+1} = 1 - (z_1 + \cdots + z_n)$) that the integrand in (3.1) is identically equal to zero, while the integrand in (3.2) is equal to

$$(4.1) \qquad \frac{(n-1)(n+2)!}{24}(z_1^2 + \cdots + z_{n+1}^2) - \frac{(n-1)(n+1)!}{12}.$$

Suppose we are testing the hypothesis that $f(x) = 1$ against alternatives of the form $f(x) = 1 + \delta(x - \frac{1}{2})$, with given level of significance α. We are going to consider only tests based on $Z_1, \cdots, Z_n$, so that our critical region will be a region in $(Z_1, \cdots, Z_n)$-space. We want to find the critical region R satisfying the following three conditions:

$$(1) \qquad M(R, 0) = \alpha,$$

$$(2) \qquad \left.\frac{dM(R, \delta)}{d\delta}\right]_{\delta=0} = 0,$$

$$(3) \qquad \left.\frac{d^2M(R, \delta)}{d\delta^2}\right]_{\delta=0} \text{ is a maximum.}$$

In the terminology of [6], this region R would be called an "unbiassed critical region of type A" for testing the hypothesis that $\delta = 0$. We know that in the present case, condition (2) is automatically satisfied by any region R, since the integrand in (3.1) is identically zero. But then a very simple application of the Neyman-Pearson lemma shows that the desired region R is given by

$$\frac{2n!\{(n-1)(n+2)!(z_1^2 + \cdots + z_{n+1}^2)/24 - (n-1)(n+1)!/12\}}{n!(n+1)!} \geqq K(\alpha),$$

where $K(\alpha)$ is a properly chosen constant. Equivalently, R is given by

$$z_1^2 + \cdots + z_{n+1}^2 \geqq k(\alpha),$$

where $k(\alpha)$ is a properly chosen constant.

5. Consistency of the proposed test. In this section we prove that the test described in Section 4 is one of a class of tests, any one of which is consistent against a wide class of alternatives. First we need some lemmas.

LEMMA 1. *If $g(x)$, the common density of $X_1, \cdots, X_n$, has at most a finite number of discontinuities, and if $R_n(t)$ denotes the proportion of the values*

$$Z_1, \cdots, Z_{n+1}$$

which are not greater than $t/(n+1)$, while $S(t)$ denotes $1 - \int_0^1 g(x) \exp\{-t\, g(x)\}\, dx$, and $V(n)$ denotes $\sup_{t \geqq 0} |R_n(t) - S(t)|$, then $V(n)$ converges to zero with probability one as n increases.

PROOF. This is proved in [4].

Now we introduce the following notation. Let u be any positive number, while Y_n shall denote $\Gamma(n + u + 1)/\Gamma(n + 2)[Z_1^u + \cdots + Z_{n+1}^u]$. Let $g(x)$ denote the common density of $X_1, \cdots, X_n$, and define $J(g; u)$ as

$$\Gamma(u + 1) \int_{[x:g(x)>0]} [g(x)]^{1-u} \, dx.$$

$J(g; u)$ may fail to exist (that is, be infinite).

LEMMA 2. *If $J(g; u)$ is finite, then given any positive numbers ϵ, δ, there is a positive integer $N(\epsilon, \delta)$ such that*

$$P[Y_n > J(g; u) - \epsilon \text{ simultaneously for all } n > N(\epsilon, \delta)] > 1 - \delta.$$

If $J(g; u)$ fails to exist, then given any positive numbers B, δ, there is a positive integer $M(B, \delta)$ such that

$$P[Y_n > B \text{ simultaneously for all } n > M(B, \delta)] > 1 - \delta.$$

PROOF. In the notation of Lemma 1, we have

$$Z_1^u + \cdots + Z_{n+1}^u = (n + 1)^{1-u} \int_0^\infty t^u \, dR_n(t),$$

and therefore $Y_n = \Gamma(n + u + 1)/\Gamma(n + 2)(n + 1)^{1-u} \int_0^\infty t^u \, dR_n(t)$. Now we choose any positive number T and hold it fixed until further notice. We have $Y_n \geqq \Gamma(n + u + 1)/\Gamma(n + 2)(n + 1)^{1-u} \int_0^T t^u \, dR_n(t)$. As n increases, the coefficient of the integral in this last expression approaches unity, and from now on we shall treat it as unity, and it will be seen that this does not affect our conclusion. We have $\int_0^T t^u \, dR_n(t) = T^u R_n(T) - u \int_0^T t^{u-1} R_n(t) \, dt$, and by Lemma 1, the expression on the right of this equality approaches the following with probability one:

$$T^u \left[1 - \int_0^1 g(x) \exp\{-Tg(x)\} \, dx\right] - u \int_0^T t^{u-1} \, dt + u \int_0^1 g(x) \left\{\int_0^T t^{u-1} \exp[-tg(x)] \, dt\right\} dx,$$

which equals

$$-\int_0^1 T^u g(x) \exp[-Tg(x)] \, dx + u \int_{[x:g(x)>0]} [g(x)]^{1-u} \int_0^{Tg(x)} r^{u-1} e^{-r} \, dr \, dx.$$

But by taking T large enough, this last expression can be made arbitrarily close to $J(g; u)$ if its exists, or it can be made arbitrarily large if $J(g; u)$ fails to exist. This proves Lemma 2.

LEMMA 3. *If the common density of $X_1, \cdots, X_n$ is uniform on $(0, 1)$, then Y_n converges stochastically to $\Gamma(u + 1)$ as n increases.*

PROOF. This is proved directly from the discussion on page 245 of [5].

LEMMA 4. *If $u > 1$, and if $g(x)$ is positive almost everywhere on* (0, 1) *and differs from unity on a subset of* (0, 1) *of positive measure, then $J(g; u) > \Gamma(u + 1)$.*

PROOF. For convenience, we omit the limits of integration, which are always zero and one throughout this proof. Hölder's inequality states that if $p > 0$, $q > 0$, and $p + q = pq$, then

$$\left| \int r(x)s(x)\, dx \right| \leq \left(\int |r(x)|^p\, dx \right)^{1/p} \left(\int |s(x)|^q\, dx \right)^{1/q},$$

with equality holding if and only if $|r(x)|^p = K\,|s(x)|^q$ almost everywhere, where K is a constant, and either $r(x)s(x) \geqq 0$ almost everywhere or $r(x)s(x) \leqq 0$ almost everywhere. Applying this inequality with $r(x) = [g(x)]^{(u-1)/u}$, $s(x) = [g(x)]^{-1+1/u}$, $p = u/(u - 1)$, and $q = u$, the lemma follows immediately.

THEOREM. *Suppose it is known that $X_1, X_2, \cdots$ are independently and identically distributed, and it is desired to test the hypothesis that the common distribution is the uniform distribution over* (0, 1). *For a given level of significance, α $(0 < \alpha < 1)$, a given number $u > 1$, and a given positive integer n, let $T(\alpha; n; u)$ denote the test of the hypothesis described as follows: Reject the hypothesis if and only if at least one of the following occurs:*

(1) *At least one of the values $X_1, \cdots, X_n$ falls outside the open interval* (0, 1),

(2) $X_i = X_j$ *for some integers i, j with $1 \leqq i < j \leqq n$,*

(3) $Z_1^u + Z_2^u + \cdots + Z_{n+1}^u \geqq K(\alpha; n; u)$,

where $K(\alpha; n; u)$ is a constant chosen to give the proper level of significance. Then the sequence of tests $\{T(\alpha; n; u), T(\alpha; n + 1; u), \cdots\}$ is consistent against any alternative common distribution function $G(x)$ with at least one of the following properties:

(1′) $G(0) > 0$,

(2′) $G(1) < 1$,

(3′) *$G(x)$ has at least one positive saltus,*

(4′) $G(0) = 0$, $G(1) = 1$, *$G(x)$ is absolutely continuous with derivative $g(x)$, and $g(x)$ differs from unity on a subset of* (0, 1) *of positive measure and has at most a finite number of discontinuities, and a finite number of oscillations.*

PROOF. If $G(x)$ has property (1′) or (2′), specification (1) of $T(\alpha; n; u)$ proves consistency. If $G(x)$ has property (3′), specification (2) of $T(\alpha; n; u)$ proves consistency. If property (4′) is possessed by $G(x)$, we distinguish two cases, according to whether or not $g(x)$ is positive almost everywhere on (0, 1).

CASE 1: $g(x)$ is positive almost everywhere on (0, 1). We may express specification (3) of the test in terms of Y_n defined above. For large n, Lemma 3 tells us that specification (3) of the test is essentially $Y_n > \Gamma(u + 1)$. But Lemmas 2 and 4 guarantee that under $G(x)$ the probability is high that Y_n will be greater than $\Gamma(u + 1)$ if n is large.

CASE 2: $g(x)$ is zero on a subset of (0, 1) of positive measure. Since $g(x)$ has at most a finite number of discontinuities, a point w in the interior of (0, 1) can be found such that $g(x)$ is continuous in a neighborhood of w, $g(w) = 0$, and any neighborhood of w contains a set of positive measure on which $g(x) = 0$. Since

$g(x)$ has a finite number of oscillations, this implies that there is an interval of positive length Δ in the interior of $(0, 1)$ on which $g(x) = 0$. But then the largest of the values $Z_1, \cdots, Z_{n+1}$ is certainly no smaller than Δ; therefore Y_n is certainly no smaller than $\Delta^u\Gamma(n + u + 1)/\Gamma(n + 2)$, and this last expression approaches infinity as n increases.

REFERENCES

[1] B. F. KIMBALL, "Some basic theorems for developing tests of fit for the case of nonparametric probability distribution functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 540–548.

[2] B. SHERMAN, "A random variable related to the spacing of sample values," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 339–361.

[3] M. GREENWOOD, "The statistical study of infectious diseases," *J. Roy. Stat. Soc.*, Vol. 109 (1946), pp. 85–110.

[4] L. WEISS, "The stochastic convergence of a function of sample successive differences," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 532–536.

[5] D. A. DARLING, "On a class of problems related to the random division of an interval," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 239–253.

[6] J. NEYMAN AND E. S. PEARSON, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Memoirs*, Vol. 1 (1936), pp. 1–37.

ON THE PROBABILITY OF LARGE DEVIATIONS FOR SUMS OF BOUNDED CHANCE VARIABLES

BY HARRY WEINGARTEN

Bureau of Ships, Navy Department

1. Summary. The following theorems are proved.

THEOREM 1. *If $x_1, x_2, \cdots$ satisfy $-1 \leqq x_n \leqq a$, $a \leqq 1$ and*

$$E(x_n \mid x_1, \cdots, x_{n-1}) \leqq -u \max (|x_n| \mid x_1, \cdots, x_{n-1}),$$

$0 < u < 1$, *then for any positive t,*

$$\Pr \{x_1 + \cdots + x_n \geqq t \text{ for some } n\} \leqq \theta^t,$$

where θ is the positive root (other than $\theta = 1$) of

$$\frac{a + u}{a + 1}\theta^{a+1} - \theta^a + \frac{1 - u}{a + 1} = 0. \tag{1}$$

This choice of θ is the best possible.

THEOREM 2. *If $x_1, x_2, \cdots$ satisfy $|x_n| \leqq 1$ and $E(x_n \mid x_1, \cdots, x_{n-1}) = 0$, then for all $N > 0$,*

$$\Pr\left\{\left|\frac{x_1 + \cdots + x_n}{n}\right| \geqq \epsilon \text{ for some } n \geqq N\right\} \leqq 2\varphi^N,$$

Received November 4, 1955. (Revised January 3, 1956).