Z. Wahrscheinlichkeitstheorie verw. Geb. 4, 217—221 (1965)

On Asymptotic Sampling Theory for Distributions Approaching the Uniform Distribution*

By

LIONEL WEISS

1. Introduction and summary

Suppose $X_1, \ldots, X_n$ are independent and identically distributed random variables, the common probability density function being $f_n(x)$, where $f_n(x) = 1 + \frac{r(x)}{n^\delta}$ for $0 < x < 1$, where $\int_0^1 r(x)\,dx = 0$ and $|r''(x)| < D < \infty$ for $0 \leqq x \leqq 1$, and δ is some positive constant. Strictly speaking, $f_n(x)$ is not defined when $x = 0$ or 1, but when we write $f_n(0)$ we shall mean $1 + \frac{r(0)}{n^\delta}$, and when we write $f_n(1)$ we shall mean $1 + \frac{r(1)}{n^\delta}$. $\int_{-\infty}^{x} f_n(u)\,du$ is denoted by $F_n(x)$.

Let $Y_1 \leqq Y_2 \leqq \cdots \leqq Y_n$ denote the ordered values of $X_1, \ldots, X_n$, and let Y_0 denote 0, Y_{n+1} denote 1. Define V_i as $Y_i - Y_{i-1}$ for $i = 1, \ldots, n+1$. Let $g_n(v_1, \ldots, v_n)$ denote the joint probability density function for $V_1, \ldots, V_n$.

Let $U_1, \ldots, U_{n+1}$ be independent identically distributed random variables, each with probability density function e^{-u} for $u > 0$, zero for $u < 0$. Define W_i as $\frac{U_i}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]}$ for $i = 1, \ldots, n+1$, T_n' as $\sum_{i=1}^{n+1} W_i$, and Z_i as $\frac{W_i}{T_n'}$ for $i = 1, \ldots, n+1$. Let $h_n(z_1, \ldots, z_n)$ denote the joint probability density function for $Z_1, \ldots, Z_n$.

For each n, let R_n be any measurable set in n-dimensional space. We shall prove the following

Theorem.

$$\lim_{n\to\infty} \left| \int_{R_n}\cdots\int g_n(v_1, \ldots, v_n)\,dv_1 \cdots dv_n - \int_{R_n}\cdots\int h_n(v_1, \ldots, v_n)\,dv_1 \cdots dv_n \right| = 0 .$$

Applications of this theorem to finding the asymptotic power of tests of fit are discussed.

2. Proof of the theorem

It is easily seen that $g_n(v_1, \ldots, v_n) = n!\, f_n(v_1) f_n(v_1 + v_2) \cdots f_n(v_1 + \cdots + v_n)$ for $v_i > 0$ $(i = 1, \ldots, n)$ and $v_1 + \cdots + v_n < 1$, and is zero otherwise.

Let $Q_n(z_1, \ldots, z_n)$ denote $\sum_{i=1}^{n} \left\{ f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right] - f_n[F_n^{-1}(1)] \right\} z_i + f_n[F_n^{-1}(1)]$. A standard calculation shows that $h_n(z_1, \ldots, z_n) = n!\,[Q_n(z_1, \ldots, z_n)]^{-n-1} \prod_{i=1}^{n+1} f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]$ for $z_i > 0$ $(i = 1, \ldots, n)$ and $z_1 + \cdots + z_n < 1$, and is zero otherwise.

Now we investigate $\log \frac{g_n(V_1, \ldots, V_n)}{h_n(V_1, \ldots, V_n)}$ over the region $V_i > 0$ $(i = 1, \ldots, n)$

* Research supported by National Science Foundation Grant No. NSF-G22585.

and $V_1 + \cdots + V_n < 1$. Recalling that $Y_i = V_1 + \cdots + V_i$ for $i = 1, \ldots, n$, we find that $\log \frac{g_n(V_1, \ldots, V_n)}{h_n(V_1, \ldots, V_n)}$ can be written as the sum of the following three terms:

$$-\log f_n[F_n^{-1}(1)] \tag{2.1}$$

$$(n+1)\log Q_n(V_1, \ldots, V_n) \tag{2.2}$$

$$\sum_{i=1}^{n} \log \frac{f_n(Y_i)}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]}. \tag{2.3}$$

By our assumptions about $f_n(x)$, for all sufficiently large n we have $F_n^{-1}(1) = 1$, and from now on we assume that n is large enough to make this so. Since $f_n(1)$ approaches 1 as n increases, the expression (2.1) approaches zero as n increases.

We have

$$\begin{aligned} Q_n(V_1, \ldots, V_n) &= f_n[F_n^{-1}(1)] + \sum_{i=1}^{n} \left\{ f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right] - f_n\left[F_n^{-1}\left(\frac{i+1}{n+1}\right)\right]\right\} Y_i \\ &= f_n(1) - \frac{1}{n+1} \sum_{i=1}^{n} \frac{f_n'\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]} Y_i - \frac{1}{2(n+1)^2} \sum_{i=1}^{n} r(i, n)\, Y_i \end{aligned} \tag{2.4}$$

where

$$r(i, n) = \frac{1}{f_n^2[\theta_i(n)]} \left\{ f_n''[\theta_i(n)] - \frac{f_n'[\theta_i(n)]}{f_n[\theta_i(n)]} \right\},$$

$\theta_i(n)$ being some value in (0, 1). By our assumptions about $f_n(x)$, $r(n) = \max_{1 \leq i \leq n} |r(i, n)|$ approaches zero as n increases. Since $|Y_i| \leq 1$ for all i, the third term in (2.4) is less in absolute value than $\frac{r(n)}{2(n+1)}$. The second term in (2.4) can be written

$$\begin{aligned} &-\frac{1}{n+1} \sum_{i=1}^{n} \frac{f_n'\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]} \left\{ Y_i - F_n^{-1}\left(\frac{i}{n+1}\right)\right\} - \\ &-\frac{1}{n+1} \sum_{i=1}^{n} \frac{f_n'\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]} F_n^{-1}\left(\frac{i}{n+1}\right). \end{aligned} \tag{2.5}$$

By an elementary computation similar to that used in proving the lemma in [2], and using the fact that $\sup_{0<x<1} |f_n'(x)|$ approaches zero as n increases, we find that the second term in (2.5) can be written as

$$-\int_0^1 \frac{f_n'[F_n^{-1}(x)]}{f_n[F_n^{-1}(x)]} F_n^{-1}(x)\, dx + \frac{s(n)}{n+1} \tag{2.6}$$

where $|s(n)|$ approaches zero as n increases. Making the change of variable $y = F_n^{-1}(x)$ in (2.6), we find that the second term in (2.5) is equal to $1 - f_n(1) + \frac{s(n)}{(n+1)}$. Therefore

$$\begin{aligned} Q_n(V_1, \ldots, V_n) = 1 &- \frac{1}{n+1} \sum_{i=1}^{n} \frac{f_n'\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]} \left\{ Y_i - F_n^{-1}\left(\frac{i}{n+1}\right)\right\} - \\ &- \frac{1}{2(n+1)^2} \sum_{i=1}^{n} r(i, n)\, Y_i + \frac{s(n)}{n+1}. \end{aligned} \tag{2.7}$$

In [3] it was shown that for any positive ε, $\max_{1 \leq i \leq n} \left| Y_i - F_n^{-1}\left(\frac{i}{n+1}\right) \right| n^{1/2-\varepsilon}$ converges stochastically to zero as n increases. Since $\sup_{0<x<1} \frac{|f_n'(x)|}{f_n(x)}$ is less than $\frac{K}{n^\delta}$ for sufficiently large n, where K is a finite constant, it follows that

$$n^{1/2+\delta-\varepsilon} \left| \frac{1}{n+1} \sum_{i=1}^{n} \frac{f_n'\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]} \left\{ Y_i - F_n^{-1}\left(\frac{i}{n+1}\right) \right\} \right| \tag{2.8}$$

converges stochastically to zero as n increases. Then, using (2.7), it follows that (2.2) can be written as

$$-\sum_{i=1}^{n} \frac{f_n'\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]} \left\{ Y_i - F_n^{-1}\left(\frac{i}{n+1}\right) \right\} + \Delta_n \tag{2.9}$$

where Δ_n converges stochastically to zero as n increases.

We can write (2.3) as

$$\sum_{i=1}^{n} \log \left\{ \frac{f_n\left\{F_n^{-1}\left(\frac{i}{n+1}\right) + \left[Y_i - F_n^{-1}\left(\frac{i}{n+1}\right)\right]\right\}}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]} \right\} = \sum_{i=1}^{n} \log \times$$

$$\times \left\{ \frac{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right] + f_n'\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]\left\{Y_i - F_n^{-1}\left(\frac{i}{n+1}\right)\right\} + \frac{1}{2} f_n''[\theta(i,n)]\left\{Y_i - F_n^{-1}\left(\frac{i}{n+1}\right)\right\}^2}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]} \right\}$$

where $\theta(i, n)$ is some value in $(0, 1)$. Expanding this last expression, using the property of $\max_{1 \leq i \leq n} \left| Y_i - F_n^{-1}\left(\frac{i}{n+1}\right) \right|$ stated above, and the fact that

$$\sup_{0<x<1} \left\{ \frac{|f_n'(x)| + |f_n''(x)|}{f_n(x)} \right\} < \frac{K_1}{n^\delta},$$

where K_1 is a finite constant, for sufficiently large n, we find that (2.3) can be written as

$$\sum_{i=1}^{n} \frac{f_n'\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]}{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]} \left\{ Y_i - F_n^{-1}\left(\frac{i}{n+1}\right) \right\} + \Delta_n' \tag{2.10}$$

where Δ_n' converges stochastically to zero as n increases.

Collecting the information we have developed about the expressions (2.1), (2.2), and (2.3), we see that we have proved that $\log \frac{g_n(V_1, \ldots, V_n)}{h_n(V_1, \ldots, V_n)}$ converges stochastically to zero as n increases, or equivalently, that $\frac{g_n(V_1, \ldots, V_n)}{h_n(V_1, \ldots, V_n)}$ converges stochastically to unity. If we define $S_n(V_1, \ldots, V_n)$ by the equation

$$\frac{g_n(V_1, \ldots, V_n)}{h_n(V_1, \ldots, V_n)} = 1 + S_n(V_1, \ldots, V_n)$$

then $S_n(V_1, \ldots, V_n)$ converges stochastically to zero as n increases. This means that there are two sequences of positive values, $\{\varepsilon_n\}$, $\{\delta_n\}$, such that $\lim_{n\to\infty} \varepsilon_n = \lim_{n\to\infty} \delta_n = 0$, and

$$\underset{|S_n(v_1,\ldots,v_n)|<\varepsilon_n}{\int\cdots\int} g_n(v_1,\ldots,v_n)\,dv_1\cdots dv_n > 1-\delta_n\,. \tag{2.11}$$

Now we have

$$\left|\underset{|S_n(v_1,\ldots,v_n)|<\varepsilon_n}{\int\cdots\int} g_n(v_1,\ldots,v_n)\,dv_1\cdots dv_n - \underset{|S_n(v_1,\ldots,v_n)|<\varepsilon_n}{\int\cdots\int} h_n(v_1,\ldots,v_n)\,dv_1\cdots dv_n\right|$$

$$\leqq \underset{|S_n(v_1,\ldots,v_n)|<\varepsilon_n \text{ and } h_n(v_1,\ldots,v_n)>0}{\int\cdots\int} h_n(v_1,\ldots,v_n)\,|S_n(v_1,\ldots,v_n)|\,dv_1\cdots dv_n < \varepsilon_n$$

and using (2.11), we find that

$$\underset{|S_n(v_1,\ldots,v_n)|<\varepsilon_n}{\int\cdots\int} h_n(v_1,\ldots,v_n)\,dv_1\cdots dv_n > 1-\delta_n-\varepsilon_n\,. \tag{2.12}$$

Now, for each n let $b_n(v_1,\ldots,v_n)$ be a measurable function of $v_1,\ldots,v_n$, satisfying $|b_n(v_1,\ldots,v_n)| < B < \infty$ for all $v_1,\ldots,v_n$. We have

$$\left|\underset{g_n(v_1,\ldots,v_n)>0}{\int\cdots\int} b_n(v_1,\ldots,v_n)\,g_n(v_1,\ldots,v_n)\,dv_1\cdots dv_n - \underset{h_n(v_1,\ldots,v_n)>0}{\int\cdots\int} b_n(v_1,\ldots,v_n)\,h_n(v_1,\ldots,v_n)\,dv_1\cdots dv_n\right|$$

$$\leqq B\underset{|S_n(v_1,\ldots,v_n)|<\varepsilon_n \text{ and } h_n(v_1,\ldots,v_n)>0}{\int\cdots\int} h_n(v_1,\ldots,v_n)\,|S_n(v_1,\ldots,v_n)|\,dv_1\cdots dv_n + B\delta_n + B(\delta_n+\varepsilon_n)$$

$$\leqq B\varepsilon_n + B\delta_n + B(\delta_n+\varepsilon_n)\,,$$

using (2.11) and (2.12). The theorem now follows by defining $b_n(v_1,\ldots,v_n)$ to be unity if $(v_1,\ldots,v_n)$ is in R_n, and zero otherwise.

3. Applications

A common problem is that of testing the hypothesis that the common unknown distribution of the independent random variables $X_1,\ldots,X_n$ is the uniform distribution over (0, 1). For a given test of this hypothesis, it would be of interest to know its asymptotic power against a sequence of alternatives $\{f_n(x)\}$, where $f_n(x) = 1 + \frac{r(x)}{n^\delta}$. The theorem proved above enables us to study this power by studying functions of independent exponentially distributed variables.

As an example, we discuss the test which rejects the hypothesis when $\sum_{i=1}^{n+1} V_i^2$ is greater than $\frac{2}{n} + 2k(\alpha)n^{-3/2}$, where $k(\alpha)$ satisfies the equation

$$\frac{1}{\sqrt{2\pi}}\int_{k(\alpha)}^{\infty} e^{-1/2\,t^2}\,dt = \alpha\,.$$

This test is discussed in [4], where it is shown that its asymptotic level of significance is α. Here we want to discuss its power against alternatives with probability density function $f_n(x)$.

Let A_n' denote $\sum_{i=1}^{n+1} W_i^2$, and let $c_n(b)$ denote $\sum_{i=1}^{n+1} \left\{f_n\left[F_n^{-1}\left(\frac{i}{n+1}\right)\right]\right\}^{-b}$. A straightforward calculation shows that for positive integral b,

$$\frac{c_n(b)}{n+1} = 1 + \frac{b(b-1)}{2n^{2\delta}} \int_0^1 r^2(y)\,dy + \frac{K_n(b)}{n^{1+\delta}}$$

where $|K_n(b)| < K(b) < \infty$. Let A_n denote

$$\frac{A_n' - 2c_n(2)}{\sqrt{20\,c_n(4)}} \text{ and } T_n \text{ denote } \frac{T_n' - c_n(1)}{\sqrt{c_n(2)}}.$$

The central limit theorem shows that asymptotically A_n and T_n have a joint normal distribution with zero means, unit variances, and covariance $\frac{4}{\sqrt{20}}$. The test being considered rejects when $\frac{A_n'}{(T_n')^2} > \frac{2}{n} + 2k(\alpha)n^{-3/2}$. Expressing A_n' and T_n' in terms of A_n and T_n, and using the expressions for $c_n(b)$ given above, we find that the test rejects when

$$\frac{\sqrt{20}}{2} A_n - 2T_n > k(\alpha) - n^{1/2-2\delta} \int_0^1 r^2(y)\,dy + \Delta_n$$

where Δ_n converges stochastically to zero as n increases. From the asymptotic joint distribution of A_n and T_n given above, it follows that $\frac{\sqrt{20}}{2} A_n - 2T_n$ has asymptotically a standard normal distribution. Then, letting $\Phi(v)$ denote $\frac{1}{\sqrt{2\pi}} \int_v^\infty e^{-1/2 t^2}\,dt$, we find that the asymptotic power of the test against $f_n(x)$ is

$$\Phi\left[k(\alpha) - n^{1/2-2\delta} \int_0^1 r^2(y)\,dy\right].$$

4. Relation to earlier work

In [1], Rényi studied the distribution of the ordered observations from a population with distribution function $F(x)$ in terms of the distribution of functions of independent exponential variables. Since $F(x)$ is not approaching the uniform distribution as n increases, the Theorem of the present paper does not hold for all sequences $\{R_n\}$ of measurable sets in n-dimensional space.

References

[1] Rényi, A.: On the theory of order statistics. Acta math. Acad. Sci. Hungar. **4**, 191—227 (1953).

[2] Weiss, Lionel: On the asymptotic distribution of an estimate of a scale parameter. Naval Res. Logist. Quart. **10**, 1—9 (1963).

[3] — Tests of fit in the presence of nuisance location and scale parameters. Ann. math. Statistics **28**, 1016—1020 (1957).

[4] — The asymptotic power of certain tests of fit based on sample spacings. Ann. math. Statistics **28**, 783—786 (1957).

Department of Industrial Engineering
and Administration
Cornell University
Ithaca, N. Y. (USA)

(Received September 1, 1964)