SOME RÉNYI TYPE LIMIT THEOREMS FOR EMPIRICAL DISTRIBUTION FUNCTIONS[1]

By Miklós Csörgő

Princeton University

1. Summary. Let $F_n(x)$ denote the empirical distribution function of a random sample of size n drawn from a population having continuous distribution function $F(x)$. In Section 3 the limiting distribution of the supremum of the random variables $\{F_n(x) - F(x)\}/F_n(x)$, $|F_n(x) - F(x)|/F_n(x)$, $\{F_n(x) - F(x)\}/(1 - F(x))$, $|F_n(x) - F(x)|/(1 - F(x))$, $\{F_n(x) - F(x)\}/(1 - F_n(x))$, $|F_n(x) - F(x)|/(1 - F_n(x))$ is derived where sup is taken over suitable ranges of x respectively. Relevant tests and some combinations of them are also discussed briefly in Section 3.

2. Introduction. Let $\xi_1, \xi_2, \cdots, \xi_n$ be mutually independent random variables with the same continuous distribution function $F(x)$ and let $\xi_1^* < \xi_2^* < \cdots < \xi_n^*$ be their order statistics. Let us define the empirical distribution function

$$F_n(x) = 0 \quad \text{if} \quad x < \xi_1^*$$
$$(2.1) \qquad = k/n \quad \text{if} \quad \xi_k^* \leqq x < \xi_{k+1}^*$$
$$= 1 \quad \text{if} \quad \xi_n^* \leqq x.$$

In his paper [5] A. Rényi proves the following theorems:

Theorem 1.

$$\lim_{n\to\infty} P\{n^{\frac{1}{2}} \sup_{a \leqq F(x)} \{F_n(x) - F(x)\}/F(x) < y\}$$
$$(2.2) \qquad = (2/\pi)^{\frac{1}{2}} \int_0^{y(a/[1-a])^{\frac{1}{2}}} \exp(-t^2/2)\, dt,$$
if $y > 0$, $0 < a < 1$ *and zero otherwise*,
$$= \Phi(y\{a/[1-a]\}^{\frac{1}{2}}).$$

Theorem 2.

$$\lim_{n\to\infty} P\{n^{\frac{1}{2}} \sup_{a \leqq F(x)} |F_n(x) - F(x)|/F(x) < y\}$$
$$(2.3) \qquad = 4/\pi \sum_{k=0}^{\infty} \{(-1)^k/(2k+1)\} \exp\{-(2k+1)^2\pi^2(1-a)/8ay^2\},$$
if $y > 0$, $0 < a < 1$, *zero otherwise*,
$$= L(y\{a/[1-a]\}^{\frac{1}{2}}).$$

Received 12 November 1963; revised 15 May 1964.

[1] This research was supported in part by the U. S. Army Research Office (Durham) on a project at Princeton University.

THEOREM 3.

$$(2.4)\quad \lim_{n\to\infty} P\{n^{\frac{1}{2}} \sup_{a\leqq F(x)\leqq b} \{F_n(x) - F(x)\}/F(x) < y\} = N(y;\, a,\, b),$$

where $-\infty < y < +\infty$, $0 < a < b < 1$.

For the form of $N(\cdot)$, we refer the reader to (3.6) of [5].

THEOREM 4.

$$(2.5)\quad \lim_{n\to\infty} P\{n^{\frac{1}{2}} \sup_{a\leqq F(x)\leqq b} |F_n(x) - F(x)|/F(x) < y\} = R(y;\, a\, b)$$

if $y > 0$, $0 < a < b < 1$, *zero otherwise.*

For the form of $R(\cdot)$, we refer the reader to (3.7) of [5].

3. Some combinations of Rényi's theorems to provide symmetrical and consistent tests.

THEOREM 5. $1/F(x)$ *of Theorems* 1, 2, 3 *and* 4 *of Section* 2 *can be replaced by* $1/F_n(x)$ *and the same limit statements hold when* sup *is taken over all* x's *such that* $a \leqq F_n(x)$ *or* $a \leqq F_n(x) \leqq b$ *respectively in the appropriate theorems.*

PROOF. That when sup is taken over all x's such that $a \leqq F_n(x)$ or $a \leqq F_n(x) \leqq b$ instead of $a \leqq F(x)$ or $a \leqq F(x) \leqq b$ respectively in the appropriate theorems then the same limit statements hold is a part of the proofs of Rényi's original theorems.

To prove the validity of the replacement of $1/F(x)$ by $1/F_n(x)$ we could use Rényi's method of proof of his original theorems (this was done in [3]) or we can use the following extension of a theorem of Cramér suggested by the referee. First we state a simplified form of Cramér's theorem [2]: If an arbitrary sequence of random variables $\{X_n\}$ converges in distribution (in law) to X as $n \to \infty$, written as $X_n \to_L X$, and if another sequence of random variables $\{Y_n\}$ converges in probability (in measure) to 1 as $n \to \infty$, written $Y_n \to_p 1$, then $Z_n = X_nY_n \to_L X$ as $n \to \infty$. This theorem extends to the following situation: if $X_n(x) \to_L X$ uniformly in x and $Y_n(x) \to_p 1$ uniformly in x then $Z_n(x) = X_n(x)Y_n(x) \to_L X$ uniformly in x. In our case we put $X_n(x) = n^{\frac{1}{2}}\{F_n(x) - F(x)\}/F(x)$, $Y_n(x) = F(x)/F_n(x)$ say in Theorem 1 of Section 2. Then we know that $P(X_n(x) \leqq y) \to \Phi[y(a/(1-a))^{\frac{1}{2}}]$ of Theorem 1 uniformly in x over all x's such that $F_n(x) \geqq a$, $0 < a < 1$, as $n \to \infty$ by Theorem 1 and $Y_n(x) \to_p 1$ uniformly in x over all x's such that $F_n(x) \geqq a$, $0 < a < 1$, as $n \to \infty$ by the Glivenko-Cantelli theorem. Therefore $P(X_n(x)Y_n(x) < y) \to \Phi[y(a/(1-a))^{\frac{1}{2}}]$ uniformly in x as $n \to \infty$. Similar statements hold for Theorems 2, 3 and 4 of Section 2 and thus Theorem 5 is verified.

In his paper Rényi noted the possibility of replacing $F(x)$ by $1 - F(x)$ in the denominator of his random variables and also showed how a substitution of $-x$ for x gave the required result. This works for the two cases where absolute values of the differences are considered. In the other two cases a repetition of Rényi's proofs, mutatis mutandis, produces the required results. In particular, we have:

COROLLARY TO THEOREM 1.

$$\text{(3.1)}\quad \lim_{n\to\infty} P\{n^{\frac{1}{2}} \sup_{F(x)\leqq b} \{F_n(x) - F(x)\}/\{1 - F(x)\} < y\} = \Phi(y\{[1 - b]/b\}^{\frac{1}{2}})$$

where $\Phi(\cdot)$ *is as it was defined in Theorem* 1, *and* $0 < b < 1$.

COROLLARY TO THEOREM 2.

$$\text{(3.2)}\quad \lim_{n\to\infty} P\{n^{\frac{1}{2}} \sup_{F(x)\leqq b} |F_n(x) - F(x)|/\{1 - F(x)\} < y\} = L(y\{[1 - b]/b\}^{\frac{1}{2}})$$

where $L(\cdot)$ *is as it was defined in Theorem* 2, *and* $0 < b < 1$.

If in these two corollaries we put $b = 1 - a$, $0 < a < 1$, then their results coincide with those of Theorems 1 and 2 respectively.

COROLLARY TO THEOREM 3.

$$\text{(3.3)}\quad \begin{aligned} &\lim_{n\to\infty} P\{n^{\frac{1}{2}} \sup_{a\leqq F(x)\leqq b} \{F_n(x) - F(x)\}/\{1 - F(x)\} < y\} \\ &= \pi^{-1}\int_{-\infty}^{y\{(1-a)/a\}^{\frac{1}{2}}} \exp(-u^2/2)\cdot[\int_0^{h(y,u)} \exp(-t^2/2)\,dt]\,du, \\ &\textit{where } h(y, u) = [y\{(1 - a)/a\}^{\frac{1}{2}} - u]\cdot[(1 - b)(1 - a)/(b - a)]^{\frac{1}{2}}, \\ &= N'(y; a, b), \qquad -\infty < y + \infty, 0 < a < b < 1. \end{aligned}$$

COROLLARY TO THEOREM 4.

$$\text{(3.4)}\quad \lim_{n\to\infty} P\{n^{\frac{1}{2}} \sup_{a\leqq F(x)\leqq b} |F_n(x) - F(x)|/\{1 - F(x)\} < y\} = R'(y; a, b)$$

where $R'(\cdot)$ *is obtained from* $R(\cdot)$ *of Theorem* 4 *by inserting* $b = 1 - a$ *and* $a = 1 - b$.

We note here that $N'(\cdot)$ of (3.3) cannot be gained from $N(\cdot)$ of Theorem 3 this way.

A statement analogous to Theorem 5 is obvious here in connection with (3.1), (3.2), (3.3) and (3.4).

On the basis of these theorems symmetrical statistical tests can be constructed as follows. Let A_n^2 and B_n^2 be the events of Theorem 2 and its corollary respectively. Rényi showed in [5] that

$$\lim_{n\to\infty} P(A_n^2 \cap B_n^2) \geqq L(y\{a/[1 - a]\}^{\frac{1}{2}}) + L(y\{[1 - b]/b\}^{\frac{1}{2}}) - \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2y^2).$$

Similar statements can be made using the other theorems. Using similar notation in connection with Theorem 4 and its corollary we get for example

$$\lim_{n\to\infty} P(A_n^4 \cap B_n^4) \geqq R(y; a, b) + R'(y; a, b) - \varphi_2(y; a, b)$$

where $\varphi_2(y; a, b) = \lim_{n\to\infty} P\{n^{\frac{1}{2}} \sup_{a\leqq F(x)\leqq b} |F_n(x) - F(x)| < y\}$ and its form was derived by Manija and is given in [4].

Symmetrical tests giving more weight to deviations in the middle can be constructed on the basis of the following theorems.

Let

$$(3.5)\quad M_n = \max \{n^{\frac{1}{2}} \sup_{F(x)<\frac{1}{2}} (F_n(x) - F(x))/(1 - F(x)),\; n^{\frac{1}{2}} \sup_{\frac{1}{2}<F(x)} (F_n(x) - F(x))/F(x)\}$$

and

$$(3.6)\quad D_n = \max \{n^{\frac{1}{2}} \sup_{F(x)<\frac{1}{2}} |F_n(x) - F(x)|/(1 - F(x)),\; n^{\frac{1}{2}} \sup_{\frac{1}{2}<F(x)} |F_n(x) - F(x)|/F(x)\}.$$

Then, using Theorem 1 and its corollary (3.1), we have

THEOREM 6.

$$(3.7)\quad \lim_{n\to\infty} P\{M_n < y\} = \Phi^2(y), \qquad y > 0, \textit{ zero otherwise.}$$

And, using Theorem 2 and its corollary (3.2), we have

THEOREM 7.

$$(3.8)\quad \lim_{n\to\infty} P\{D_n < y\} = L^2(y), \qquad y > 0, \textit{ zero otherwise.}$$

PROOF OF THEOREM 6. (3.7) follows immediately from Theorem 1 and its corollary (3.1) if we can show that random variables in { } of (3.5) are asymptotically independent. Using the Glivenko-Cantelli theorem it can be easily shown that in the limit M_n of (3.7) can be replaced by

$$(3.9)\quad \max \{n^{\frac{1}{2}} \sup_{F_n(x)<\frac{1}{2}} (F_n(x) - F(x))/(1 - F(x)),\; n^{\frac{1}{2}} \sup_{\frac{1}{2}<F_n(x)} (F_n(x) - F(x))/F(x)\}$$

and the same limit theorem holds. From an adaptation of Rényi's proof of Theorem 1 to prove (3.1) directly, it follows that the asymptotic behavior of the first random variable of (3.9) is the same as that of

$$(3.10)\quad n^{\frac{1}{2}} \max_{1\leq k<\frac{1}{2}n} \textstyle\sum_k [(1 - \delta_k)/(n + 1 - k)]$$

and from Rényi's proof of Theorem 1 in [5] we conclude that the asymptotic behavior of the second random variable of (3.9) is the same as that of

$$(3.11)\quad n^{\frac{1}{2}} \max_{\frac{1}{2}n<k\leq n} \textstyle\sum_k [(\delta'_{n+1-k} - 1)/k]$$

where δ_k, $k = 1, \cdots, n$ of (3.10) are mutually independent, exponentially distributed random variables with mean value 1 and are defined as

$$(3.12)\quad \delta_k = (n - k + 1)\{\log [(1/(1 - \eta_k^*) - (1/(1 - \eta_{k-1}^*)]\},$$

and δ'_{n+1-k}, $k = 1, \cdots, n$, of (3.11) are also mutually independent, exponentially distributed random variables with mean value 1 and are defined as

$$(3.13)\quad \delta'_{n+1-k} = k(\log (1/\eta_k^*) - \log (1/\eta_{k+1}^*))$$

where, in both cases, $\eta_k^* = F(\xi_k^*)$, $k = 1, 2, \cdots, n$, with $\eta_0^* = 0$ and $\eta_{n+1}^* = 1$ by definition.

It follows then from mutual independence of random variables of (3.12) and that of (3.13) and from their definitions that the two sequences of variables $\{\delta_k \mid 1 \leqq k < n/2\}$ and $\{\delta'_{n+1-k} \mid n/2 < k \leqq n\}$ are independent and when put together they form a new sequence of mutually independent random variables. This implies that the random variables of (3.10) and (3.11) are also independent. From the independence of (3.10) and (3.11) it follows through (3.9) that random variables of M_n of (3.5) are aysmptotically independent and this terminates the proof of Theorem 6.

A similar argument shows that random variables of D_n of (3.6) are also independent asymptotically and this in turn implies Theorem 7.

Using Theorem 5 we can replace weight functions $1/F(x)$ and $1/(1 - F(x))$ by $1/F_n(x)$ and $1/(1 - F_n(x))$ respectively in Theorems 6 and 7.

Reasoning, similar to that of Chapman's in [1], can be used to prove consistency of above mentioned statistical tests.

4. Acknowledgments. I would like to take this opportunity to thank Professor Douglas G. Chapman and the referee for their valuable remarks and suggestions.

REFERENCES

[1] Chapman, Douglas G. (1958). A comparative study of several one-sided goodness-of-fit tests. *Ann. Math. Statist.* **29** 655–674.

[2] Cramér, H. (1946), *Mathematical Methods of Statistics*. Princeton Univ. Press.

[3] Csörgő, M. (1963). On some Kolmogorov-Smirnov-Rényi type theorems of probability. Unpublished doctoral dissertation. McGill University.

[4] Manija, G. M. (1949). Obobschenije kriterija A. N. Kolmogorova dlja otcenki zakona racpredelenija po empiricheskim dannym. *Dokl. Akad. Nauk SSSR* **69** 495–497.

[5] Rényi, Alfréd (1953). On the theory of order statistics. *Acta Math. Acad. Sci. Hungar.* **4** 191–231.