NOTES

A LIMIT THEOREM FOR SUMS OF MINIMA OF STOCHASTIC VARIABLES

By Ulf Grenander[1]

University of Stockholm

1. Summary. We consider a sequence of independent and identically distributed positive stochastic variables $x_1, x_2, x_3, \cdots$ with the distribution function $F(x)$. Let y_n be the smallest of the values taken by the n first of these variables and $S_n = y_1 + y_2 + \cdots + y_n$. It is then shown that $S_n/\log n$ tends in probability to the value $F = \lim_{t \downarrow 0} t/F(t)$ assumed to exist as a finite or infinite number.

2. The main result. The limit relation can be formulated in the following simple way.

Theorem. *Consider the independent and non-negative stochastic variables* $x_1, x_2, x_3, \cdots$ *with the common distribution function* $F(x)$. *Then the expression*

$$(1/\log n)[x_1 + \min(x_1, x_2) + \min(x_1, x_2, x_3) + \cdots + \min(x_1, x_2, x_3, \cdots, x_n)]$$

converges in probability to the value $F = \lim_{t \downarrow 0} t/F(t)$ *assumed to exist as a finite or infinite number.*

Proof. Let us start by examining the special case when the variables x_n have a rectangular distribution on the interval $(0, 1)$. The variable $y_n = \min(x_1, x_2, \cdots, x_n)$ then has the frequency function $n(1-u)^{n-1}$ for $0 < u < 1$ and the mean value $1/n + 1$. The ratio $R_n = S_n/\log n$ has the mean value

$$ER_n = (1/\log n)(\tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \cdots + (1/n + 1))$$

so that $\lim_{n \to \infty} ER_n = 1$. Now we consider the joint distribution of y_n and y_{n+k}, $k > 0$. Writing $u = \min(x_1, x_2, \cdots, x_n)$ and $v = \min(x_{n+1}, x_{n+2}, \cdots, x_{n+k})$. We have

$$y_n = u$$

$$y_{n+k} = \min(u, v).$$

Since u and v have the joint frequency function

$$n(1-u)^{n-1} \cdot k(1-v)^{k-1}$$

in the unit square $0 < u, v < 1$ we get the mixed second order moment

Received 11 January 1965.

[1] Temporarily with the IBM Watson Research Center, Yorktown Heights, New York.

$$Ey_ny_{n+k} = \int_{u=0}^{1} nu(1-u)^{n-1}\left[\int_{v=0}^{u} vk(1-v)^{k-1}\,dv + \int_{v=u}^{1} uk(1-v)^{k-1}\,dv\right] du.$$

Elementary calculations reduce this expression to

$$Ey_ny_{n+k} = \{1/[(k+1)(n+1)]\} - \{n/[(k+1)(n+k+1)(n+k+2)]\}.$$

The covariance is then $C_{n,n+k} = \mathrm{Cov}\,(y_n, y_{n+k}) = \{n/[(n+1)(n+k+1)\cdot(n+k+2)]\}$. The variance is, by direct calculation, of the same form $C_{n,n} = \{n/[(n+1)^2(n+2)]\}$. This gives us the variance of the ratio R_n

$$\mathrm{Var}\,(R_n) = (\log n)^{-2} \sum_{\nu=1}^{n} \{\nu/[(\nu+1)^2(\nu+2)]\} + 2(\log n)^{-2} \sum_{\nu=1}^{n}\sum_{k=1}^{n-\nu} \{\nu/[(\nu+1)(\nu+k+1)(\nu+k+2)]\} \leqq (\log n)^{-2} \sum_{\nu=1}^{\infty} (\nu+1)^{-2} + 2(\log n)^{-2} \sum_{\nu=1}^{n} (\nu+2)^{-1} \leqq (\log n)^{-2}\ [\text{constant} + 2\log n].$$

This shows that the variance of R_n tends to zero as n tends to infinity so that R_n tends to 1 in the mean and hence in probability.

Let us now deal with the case of a general distribution $F(x)$ satisfying the conditions of the theorem and with $0 < F < \infty$. Let ξ_ν be independent and rectangularly distributed over the interval $(0, 1)$ and introduce the function

$$G(t) = \inf\,\{x \geqq 0 \mid F(x) \geqq t\},$$

where we use the right continuous normalization of the distribution function $F(t)$. Then our variables x_ν can be represented as $G(\xi_\nu)$. Given a small neighborhood $(0, \epsilon)$ of zero it is clear that $\eta_n = \min\,(\xi_1, \xi_2, \cdots, \xi_n)$ takes values in $(0, \epsilon)$ with large probability if n is large. Note that η_n is non-increasing in n. But then the S_n can be enclosed between bounds of the form

$$[F \pm \delta]\cdot S_n',$$

where $S_n' = \eta_1 + \eta_2 + \cdots + \eta_n$. But we know that $S_n'/\log n$ tends to 1 in probability so that the first part of the theorem is proved. The cases $F = 0$ or $F = \infty$ can be dealt with by making small modifications of the original distribution function $F(x)$.

3. Acknowledgment. This problem originated in a study of network problems. I am grateful to F. Knight for some helpful suggestions. He has also shown that a direct extension of the proof can be made in order to establish convergence with probability one. O. Frank has informed me that convergence with probability one can be proved by using an extended form of the Hájek-Rényi inequality.