The Annals of Mathematical Statistics
1972, Vol. 43, No. 1, 351-353

ASYMPTOTIC NORMALITY OF SUMS OF MINIMA OF RANDOM VARIABLES

BY THOMAS HÖGLUND

University of Stockholm

Let $x_1, x_2, \cdots$ be independent and positive random variables with the common distribution function F. We show that if $\int_0^1 |F(x) - x/b| \times x^{-2}dx < \infty$ for some $0 < b < \infty$, then $\sum_{k=1}^n \min(x_1, \cdots, x_k)$ is asymptotically normal with expectation $b \log n$ and variance $b^2 2 \log n$.

Consider a sequence of independent and identically distributed positive random variables $x_1, x_2, \cdots$. Put $S_n = \sum_{k=1}^n \min(x_1, \cdots, x_k)$. In [2] Grenander has given a condition under which $S_n/\log n$ converges in probability to a certain limit. The convergence can be shown to be almost sure, [1]. In this note we will determine the asymptotic distribution of S_n under a slightly more restrictive condition than that of [2].

THEOREM. *Let $x_1, x_2, \cdots$ be independent and positive random variables with the common distribution function F. If $b > 0$ and $\int_0^1 |F(x) - x/b|x^{-2}dx < \infty$, then the distribution of*

$$(2b^2 \log n)^{-\frac{1}{2}}(S_n - b \log n)$$

tends to the normal distribution with zero expectation and unit variance.

PROOF. Suppose first that $x_1, x_2, \cdots$ are exponentially distributed with unit expectation and write

$$f_n(t) = Ee^{itS_n}$$

for $n \geqq 0$, where $S_0 = 0$. Given n, define the random variables $I_1', \cdots, I_n'$ and $I_1'', \cdots, I_n''$ by

$$I_1' = 1, \qquad I_k' = 1 \quad \text{if} \quad x_1, \cdots, x_{k-1} > x_k$$
$$= 0 \quad \text{otherwise}, \qquad k = 2, \cdots, n$$

and

$$I_k'' = 1 \quad \text{if} \quad x_{k+1}, \cdots, x_n > x_k,$$
$$= 0 \quad \text{otherwise}, \qquad k = 1, \cdots, n-1,$$

$I_n'' = 1$. The identity $\sum_1^n I_k' I_k'' = 1$ a.s. yields

$$f_n(t) = \sum_1^n E(e^{itS_n} I_k' I_k'') = \sum_1^n E(e^{it(n-k+1)x_k} E(e^{itS_{k-1}} I_k' I_k'' \mid x_k).$$

And since $x_1, \cdots, x_n$ are independent

$$E(e^{itS_{k-1}} I_k' I_k'' \mid x_k) = E(e^{itS_{k-1}} I_k' \mid x_k) E(I_k'' \mid x_k).$$

Received October 20, 1970.

Also

$$E(e^{itS_{k-1}}I_k' \mid x_k) = e^{itx_k(k-1)}E(e^{it(S_{k-1}-(k-1)x_k)} \mid I_k' = 1, x_k)P(I_k' = 1 \mid x_k) .$$

But because of the lack of memory of the exponential distribution

$$E(e^{it(S_{k-1}-(k-1)x_k)} \mid I_k' = 1, x_k) = f_{k-1}(t) .$$

Finally

$$E(I_k'' \mid x_k) = e^{-x_k(n-k)} \qquad \text{and} \qquad P(I_k' = 1 \mid x_k) = e^{-x_k(k-1)} .$$

If we collect these expressions and remember that

$$E(e^{itnx_k-(n-1)x_k}) = 1/n(1-it)$$

we obtain

$$f_n(t) = \sum_{k=0}^{n-1} f_k(t)/n(1-it) , \qquad n \geqq 1 .$$

Putting $u_n = \sum_0^n f_k(t)$ this reads

$$u_n - u_{n-1} = u_{n-1}/n(1-it) , \qquad n \geqq 1$$

so that $u_n = \prod_1^n (1 + 1/k(1-it))$.

Hence

$$f_n(t) = [1/n(1-it)] \prod_1^{n-1} (1 + 1/k(1-it)) \qquad \text{for } n \geqq 1 .$$

In order to find the asymptotic distribution in this case we note that

$$f_n(t) = \prod_1^n (1 - 1/k + 1/k(1-it)) = \prod_1^n (1 + it/k(1-it))$$

so that

$$f_n(t) = [\exp \textstyle\sum_1^n it/k(1-it)] \prod_1^n (1 + it/k(1-it)) \exp -it/k(1-it) .$$

Since $(1+z)e^{-z} = 1 + O(z^2)$ and $\sum_1^\infty 1/k^2$ converges, the product to the right equals $1 + O(t^2)$ uniformly in n. Hence, since $it/(1-it) = it - t^2 + O(t^3)$,

$$f_n(t) = [\exp (it - t^2) \textstyle\sum_1^n 1/k](1 + O(t^2) + O(t^3 \sum_1^n 1/k)) .$$

Finally, remembering that $\sum_1^n 1/k = \log n + O(1)$, we obtain

$$f_n(t(2 \log n)^{-\frac{1}{2}}) \exp - it(\tfrac{1}{2} \log n)^{\frac{1}{2}} \to e^{-t^2/2} \qquad \text{as } n \to \infty ,$$

so that the conclusion of the theorem is true in this special case.

Let us now deal with the case of a general distribution F satisfying the conditions of the theorem. Following Grenander: Let $\xi\nu$ be independent and rectangularly distributed over the interval $(0, 1)$ and put $\eta_n = \min(\xi_1, \cdots, \xi_n)$. Then our variables x_k can be represented as $\tilde{F}(\xi\nu)$, where

$$\tilde{F}(t) = \inf\{x \geqq 0 ; F(x) \geqq t\} .$$

Thus $(b^2 2 \log n)^{-\frac{1}{2}}(S_n - b \log n)$ is distributed as

$$(b^2 2 \log n)^{-\frac{1}{2}}(\textstyle\sum_1^n \tilde{F}(\eta_k) - b \log n) = (2 \log n)^{-\frac{1}{2}}(\sum_1^n \eta_k - \log n) + r_n .$$

Suppose that $r_n \to 0$ in probability under the condition of the theorem. Then since the exponential distribution with unit expectation satisfies this condition with $b = 1$, we may infer (think of an exponential F to the left in the identity above) that the conclusion of the theorem holds true also in the case of uniformly distributed variables. Once this is done the general case follows by inspection of the identity above once more (think of a general F to the left). Therefore it suffices to show that r_n tends to zero in probability under the condition of the theorem to complete the proof.

Put $y_n = 1$ if $\eta_n < \delta$, $=0$ otherwise $(0 < \delta < 1)$. Since

$$r_n = (b^2 2 \log n)^{-\frac{1}{2}}[\sum_1^n y_k(\tilde{F}(\eta_k) - b\eta_k) + \sum_1^n (1 - y_k)(\tilde{F}(\eta_k) - b\eta_k)]$$

and with probability one all but finitely many y_n equal one (note that the condition of the theorem implies that $y_n \uparrow 1$) r_n tends to zero in probability if

$$(\log n)^{-\frac{1}{2}} \sum_1^n y_k|\tilde{F}(\eta_k) - b\eta_k|$$

does. But since η_k has the density $k(1 - t)^{k-1}$ for $0 < t < 1$, the expectation of this nonnegative variable is dominated by

$$(\log n)^{-\frac{1}{2}} \int_0^\delta |\tilde{F}(t) - bt| \sum_1^\infty k(1 - t)^{k-1}dt = (\log n)^{-\frac{1}{2}} \int_0^\delta |\tilde{F}(t) - bt|t^{-2}dt\,.$$

However, for a properly chosen constant A

$$\int_0^\delta |\tilde{F}(t) - bt|t^{-2}dt \leqq \int_0^A |1/F(x) - b/x|dx = \int_0^A (bx/F(x))|F(x) - x/b|x^{-2}dx$$

which becomes obvious if we observe that

$$\int_0^\delta |\tilde{F}(t) - bt|t^{-2}dt = \int_0^\delta t^{-2}dt \int_{i(t)} dx$$

where $i(t) = [\min(bt, \tilde{F}(t)), \max(bt, \tilde{F}(t))]$, and then reverse the order of integration. Thus, if we show that $\int_0^1 |F(x) - x/b|x^{-2}dx < \infty$ implies that $x/F(x)$ is bounded in some interval $(0, \varepsilon)$, $\varepsilon > 0$, we have shown that r_n converges to zero in probability under the condition of the theorem. Choose $c > c' > b$. If $x \geqq cF(x)$ then since F is nondecreasing $y \geqq c'F(y)$ for all $y \in [\rho x, x]$, where $\rho = c'/c < 1$. Therefore, if each right neighborhood of the origin contains an $x > 0$ such that $x \geqq cF(x)$ we can choose a sequence $[\rho x_n, x_n]$, $n = 1, 2, \cdots$ of disjoint subintervals of $(0, 1)$ with the property that $x \geqq c'F(x)$ for $x \in \bigcup_{n=1}^\infty [\rho x_n, x_n]$. Hence

$$\begin{aligned}\int_0^1 |F(x) - x/b|x^{-2}dx &\geqq \sum_{n=1}^\infty \int_{\rho x_n}^{x_n} (1/b - 1/c')x^{-1}dx \\ &= \sum_{n=1}^\infty (1/b - 1/c') \log 1/\rho = \infty\,. \qquad \text{A contradiction.}\end{aligned}$$

REFERENCES

[1] Frank, O. (1966). Generalization of an inequality of Hájek and Rényi. *Skand. Aktuarietidskr.* 85–89.

[2] Grenander, U. (1965). A limit theorem for sums of minima of stochastic variables. *Ann. Math. Statist.* **36** 1041–1042.