Approximations to the Distribution Function of Sums of Independent Identically Distributed Random Variables

BY J. M. SHAPIRO

Unclassified

SCAMP Problem V−1/63 asks for the exact sampling distribution of

$$|X_1| + \ldots + |X_n|$$

when $X_1, \ldots, X_n$ are independent normal 0, 1 random variables. In this paper a general method is presented which in particular gives the answer to the above problem in the form of an infinite series. The series appears to be easily adaptable to computer solution.

INTRODUCTION

The problem presented in SCAMP Problem $V-1/63$ is to find the exact distribution function of $Y = |X_1| + \ldots + |X_n|$, where X_i is normal with mean 0 and variance 1 and $X_1, \ldots, X_n$ are independent. An approximation is asked for if the distribution function of Y is "complicated." It is not difficult to see that the probability density function of Y is given in terms of an $(n - 1)$ fold integral involving $n - 1$ elliptic integrals with variable limits of integration. It would thus appear fruitless to look for an exact solution to this problem (in terms of elementary functions).

In this paper, using known results, we present a method which in particular will give the distribution function of Y in the form of an infinite series so that any desired degree of accuracy can be obtained. The method is applicable to a large class of random variables X_i, and furnishes answers to the problems presented in the attached memorandum from R. McShane to D. Dribin, dated January 14, 1963.

DEFINITIONS

Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent identically distributed random variables with mean 0, variance 1, and common characteristic function $\phi(t)$. If ξ_1 has a finite moment of order s, we may consider for $r \leq s$,

$$\chi_r = \frac{1}{i^r}\left[\frac{d^r}{dt^r} \ln \phi(t)\right]_{t=0}$$

X_r is called the r^{th} semi-invariant of the random variable ξ_1. The semi-invariants are easily expressed in terms of the moments of the random variable (*cf.*, [1] p. 65). Thus:

$$\chi_1 = E[\xi_1] = 0$$
$$\chi_2 = E[\xi_1^2] = 1$$
$$\chi_3 = E[\xi_1^3]$$
$$\chi_4 = E[\xi_1^4] - 3$$
$$\chi_5 = E[\xi_1^5] - 10\,E[\xi_1^3]$$
$$\chi_6 = E[\xi_1^6] - 15\,E[\xi_1^4] - 10\,(E[\xi_1^3])^2 + 30$$

.
.
.

Define $P_k(-w)$ by the formal expression

$$e^{\sum_{k=1}^{\infty} \frac{\chi_{k+2}}{(k+2)!} (-w)^{k+2} z^k} = 1 + \sum_{k=1}^{\infty} P_k(-w) z^k .$$

That is, $P_k(-w)$ is the coefficient of z^k in the expansion of the left-hand side of the above expression in powers of z. It is easily seen that $P_k(-w)$ is a polynomial in w of degree $3k$ with coefficients depending on $x_3, x_4, \ldots, x_{k+2}$, that is on the moments of ξ_1. Thus

$$P_1(-w) = \frac{x_3}{6}(-w)^3$$

$$P_2(-w) = \frac{\chi_4}{24} w^4 + \frac{x_3^2}{72} w^6$$

. . . .

Let

$$\phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2}\, du$$

and let $\phi^{(r)}(y)$ denote its r^{th} derivative. Let $P_k(-\phi(y))$ be calculated by replacing w^r by $\phi^{(r)}(y)$ in $P_k(-w)$.

Define the polynomials $Q_k(y)$ by

$$\frac{1}{\sqrt{2\pi}} Q_k(y)\, e^{-y^2/2} = P_k(-\phi(y)) ,$$

so that

$$Q_1(y) = \frac{x_3}{6}(1 - y^2)$$

$$Q_2(y) = \frac{10x_3^2}{6!}y^5 + \frac{1}{8}\left(\frac{x_4}{3} - \frac{10x_3^2}{9}\right)y^3 + \left(\frac{5x_3^2}{24} - \frac{x_4}{8}\right)y$$

. . . .

THE THEOREM

Let $F_n(y)$ be the distribution function of

$$\frac{\xi_1 + \ldots + \xi_n}{\sqrt{n}}.$$

Let

$$E[\,|\,\xi_1\,|\,s] < \infty, \; (s \geq 3),$$

and suppose

$$\lim_{|t| \to \infty} \sup |\phi(t)| < 1. \tag{3.1}$$

Then

$$F_n(y) - \phi(y) =$$

$$\frac{e^{-y^2/2}}{\sqrt{2\pi}}\left(\frac{Q_1(y)}{n^{1/2}} + \frac{Q_2(y)}{n} + \ldots + \frac{Q_{s-2}(y)}{n^{(s-2)/2}}\right) + o\left(\frac{1}{n^{(s-2)/2}}\right)$$

uniformly in y.

For the proof of this theorem see [1] p. 220 ff. We point out that the condition (3.1) is satisfied for a large class of distributions. In fact, (*cf.*, [1] p. 222) the only way $\lim_{|t| \to \infty} \sup |\phi(t)|$ can be 1 for a non-lattice distribution is for all the variation of $F(x)$ (the distribution function of $\phi(t)$) to be concentrated in a set of (Lebesgue) measure zero.

Referring to the problem in the Introduction, let X_j be normal with mean 0 and variance 1. Let

$$\xi_j = \frac{|X_j| - \sqrt{\frac{2}{\pi}}}{\sqrt{1 - \frac{2}{\pi}}}.$$

Then ξ_j has mean 0 and variance 1. Furthermore, the moments of ξ_j are easily found (see SCAMP Problem V−1/63).

Now $F_n(y)$ is the distribution function of

$$\frac{\xi_1 + \ldots + \xi_n}{\sqrt{n}} = \frac{Y - n\sqrt{\frac{2}{\pi}}}{\sqrt{n}\,(1 - \frac{2}{\pi})}.$$

Since for these random variables we may let $s \to \infty$, we have (for fixed n) $F_n(y)$ expressed as an infinite series (depending only on the moments of $|X_1|$). Using computers, we should be able to get any desired degree of accuracy for $F_n(y)$ and hence for the distribution function of Y.

Clearly, this method may be used to obtain approximations to the distribution function of the sum of the absolute values of n random variables when the addend variables have finite moments of order $s \geq 3$ and (3.1) is satisfied. *Thus the questions presented in the attached memorandum are answered by this method.*

REFERENCE

[1] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables,* Translated by K. L. Chung, Addison-Wesley, 1954.