A NOTE ON THE CONVOLUTION OF UNIFORM DISTRIBUTIONS[1]

By Edwin G. Olds

Carnegie Institute of Technology

1. Introduction. Some time ago, when Dr. Acton and the present author were preparing a paper [1] on the combination of tolerances, the question arose as to the distribution of the sum of rectangular random variables having unequal bases. (For equal bases, the distribution has been known since Laplace.) As Acton pointed out, the distribution can be obtained by operational calculus. However, it seems useful to outline a derivation requiring only the well known formula for the probability density function for the sum of two random variables. In addition, this note gives several other results which may be needed in statistical quality control.

2. The distribution of the sum. Let x_i be independent random variables with probability density functions

$$f_i(x_i) = [\varepsilon(x_i) - \varepsilon(x_i - a_i)]/a_i \quad (a_i > 0; i = 1, 2, \cdots, n),$$

where $\varepsilon(x - c)$ is unity for $x \geqq c$ and zero elsewhere. Let $s = \sum_1^n x_i$ and let $f_n(s)$ and $F_n(s)$ represent the probability density function and cumulative distribution function of s respectively. Then it will be proved that

$$\begin{aligned} f_n(s) = \Big[s^{n-1}\varepsilon(s) &- \sum_1^n (s - a_i)^{n-1}\varepsilon(s - a_i) \\ &+ \sum_{i<j} (s - a_i - a_j)^{n-1}\varepsilon(s - a_i - a_j) - \cdots \\ &+ (-1)^n \left(s - \sum a_i\right)^{n-1}\varepsilon\left(s - \sum a_i\right)\Big] \Big/ \Big[(n-1)!\prod a_i\Big], \end{aligned} \tag{1}$$

and

$$\begin{aligned} F_n(s) = \Big[s^{n}\varepsilon(s) &- \sum_1^n (s - a_i)^{n}\varepsilon(s - a_i) \\ &+ \sum_{i<j} (s - a_i - a_j)^{n}\varepsilon(s - a_i - a_j) - \cdots \\ &+ (-1)^n \left(s - \sum a_i\right)^{n}\varepsilon\left(s - \sum a_i\right)\Big] \Big/ \Big[n!\prod a_i\Big]. \end{aligned} \tag{2}$$

The proof is by induction. Using the convolution formula, ([2] p. 191), we have, in our notation,

---

[1] Presented at the Annual Meeting of the Institute of Mathematical Statistics at Chicago, December 29, 1950.

Some revision of this paper was made under partial support from a U. S. Air Force contract with the Department of Mathematics of Carnegie Institute of Technology.

$$f_2(s) = \int_{-\infty}^{+\infty} f_1(s - t) f_2(t)\, dt$$

$$(3) \qquad = \frac{1}{a_1a_2}\left\{\int_{+\infty}^{+\infty} \varepsilon(t)\varepsilon(s - t)\, dt - \int_{-\infty}^{+\infty} \varepsilon(t)\varepsilon(s - t - a_1)\, dt\right.$$

$$\left. - \int_{-\infty}^{+\infty} \varepsilon(t - a_2)\varepsilon(s - t)\, dt + \int_{-\infty}^{+\infty} \varepsilon(t - a_2)\varepsilon(s - t - a_1)\, dt\right\}.$$

The first integral within the braces is zero for $t < 0$ and for $t > s > 0$ and is unity between zero and $s \geqq 0$ so the effective limits are zero and s. Likewise the effective limits for the second integral are zero and $s - a_1$. After replacing $t - a_2$ by t in the last two integrals and making obvious changes of limits, we get

$$(4) \qquad f_2(s) = \frac{1}{a_1a_2}\left\{\int_0^{s} \varepsilon(t)\varepsilon(s - t)\, dt - \int_0^{s-a_1} \varepsilon(t)\varepsilon(s - t - a_1)\, dt\right.$$

$$\left. - \int_0^{s-a_2} \varepsilon(t)\varepsilon(s - t - a_2)\, dt - \int_0^{s-a_1-a_2} \varepsilon(t)\varepsilon(s - t - a_1 - a_2)\, dt\right\}.$$

Finally,

$$(5) \quad f_2(s) = [s\varepsilon(s) - (s - a_1)\varepsilon(s - a_1) - (s - a_2)\varepsilon(s - a_2) + (s - a_1 - a_2)\varepsilon(s - a_1 - a_2)]/(a_1a_2).$$

To complete the induction we need only assume that (1) holds for $n = k$ and show that it then is true for $n = k + 1$. Using the same method for combining the density functions of $s = \sum_1^k x_i$ and x_{k+1} as was used above for x_1 and x_2, this presents no difficulty. Also it is easy to show that (2) is a direct consequence of (1).

3. Asymptotic normality. For use in the remaining sections it is noted that the constants of the distribution of s are:

$$(6) \quad \text{mean, } \mu_s = \tfrac{1}{2}\sum a_i; \text{ variance, } \sigma_s^2 = \sum a_i^2/12; \text{ skewness, } \gamma_1 = 0;$$
$$\text{excess, } \gamma_2 = -\tfrac{6}{5}\sum a_i^4/(\sum a_i^2)^2.$$

The matter of convergence of the classically normed sum to the Gaussian distribution with zero mean and unit variance can be settled easily by using the well known Lindeberg condition, the sufficiency of which, as Loève [3] notes, was established by Lindeberg and by P. Lévy, and the necessity by Feller. (For discussion and references see Loève's paper.)

As the second part of the solution of the classical central limit problem Loève ([3], p. 326) states the theorem:

NC holds and $\max_{k \leqq n} \dfrac{\sigma(x_k)}{\sigma(s_n)} \to 0$ *if, and only if, for every* $\epsilon > 0$

$$\sum_{k=1}^{n} \frac{1}{\sigma^2(s_n)} \int_{|x|>\epsilon\sigma(s_n)} x^2\, dF_k(x + Ex_k) \to 0.$$

For our case

$$(7) \qquad \sigma(x_k) = \frac{a_k}{\sqrt{12}}, \qquad \sigma(s_n) = \sigma_s = \sqrt{\frac{\sum_1^n a_i^2}{12}}, \qquad Ex_k = \tfrac{1}{2}a_k.$$

In order to specialize the above theorem to our use, we first establish the following:

LEMMA. *The Lindeberg condition holds if, for* $k \leqq n$,

$$(8) \qquad \frac{\max a_k}{\sqrt{\sum_1^n a_i^2}} \rightarrow 0.$$

To prove this lemma it is sufficient to note, first, that each term of the Lindeberg sum is identically zero whenever $\epsilon\sigma(s_n)$ is greater than $\frac{1}{2}a_k$, and second, that the condition imposed in the lemma implies the existence of an N for any $\epsilon > 0$ such that for all $n > N$

$$(9) \qquad \frac{a_k}{\sqrt{\sum_1^n a_i^2}} < \frac{\epsilon}{\sqrt{3}}.$$

The following theorem[2] can now be established.

THEOREM. *A necessary and sufficient condition for the asymptotic normality of a sum of independent rectangular random variables is*

$$(10) \qquad \lim_{\substack{n\rightarrow\infty \\ (k \leqq n)}} \frac{\max a_k}{\sqrt{\sum_1^n a_i^2}} = 0.$$

To prove the sufficiency of this condition we note that by virtue of the above lemma, the Lindeberg condition is satisfied and then note, from the quoted theorem, that the Lindeberg condition implies normal convergence. For necessity, we note that if the condition fails then the Lindeberg condition must fail for, otherwise, the quoted theorem would lead to a contradiction.

Of course the condition on max a_k implies that $(a_k/\sigma_s) \rightarrow 0$ and that $\sigma_s \rightarrow \infty$. Thus any sum for which $a_i = ra_{i+1}$ will converge to normality only if $r = 1$, since, for $r > 1$, $(a_n/\sigma_s) \nrightarrow 0$, and for $r < 1$, σ_s is bounded.

4. Percentage outside three-sigma limits. For statistical quality control there is considerable interest in knowing the percentage of a distribution outside of the limits $\mu \pm 3\sigma$. For any particular sum of rectangulars this percentage can be calculated from equation [2] above. Often the total range of nonzero probability for the sum will not exceed 6σ so that the required probability will be zero.

[2] The author is grateful to the referee for suggesting a slightly different form of this theorem as an improvement of the author's original treatment, which used Lyapunov's Theorem.

It is easy to verify that this condition will hold whenever $a_{i+1} = ra_i$ and either $0 \leqq r \leqq 0.5$ or $r \geqq 2$.

When the range for the sum is greater than 6σ, an approximation to the required percentage can be obtained from an Edgeworth series. (For discussion, see [2], pp. 221–231.) Let x be the standardized variable $(s - \mu)/\sigma_s$, $\Phi(x)$ the normal distribution function, $\phi(x)$ the normal density function, and $\phi^{(i)}(x)$ its ith derivative; then, following Cramér, we have approximately

$$f(x) = \phi(x) - \frac{\gamma_1}{3!}\phi^{(3)}(x) + \frac{\gamma_2}{4!}\phi^{(4)}(x) \frac{10\gamma_1^2}{6!}\phi^{(6)}(x). \tag{11}$$

Then, integrating and substituting the pertinent values from (6) above, we have

$$F(x) = \Phi(x) - \frac{\phi^{(3)}(x)}{20}\left[\frac{\sum a_i^4}{(\sum a_i^2)^2}\right]. \tag{12}$$

Since the lower three-sigma limit for s corresponds to $x = -3$ we have, finally, for the approximate percentage below this limit

$$F(-3) = 0.00135 - 0.004\,[\textstyle\sum a_i^4/(\sum a_i^2)^2], \tag{13}$$

where the bracketed quantity takes its minimum value, n^{-1}, when all of the a's are equal.

The multiplier of the bracket has been rounded off for easy use. A better value for it is 0.0039885. Using this instead of 0.004 in (13), and making a comparison of (2) and (13) when $a_i = 1$ $(i = 1, 2, \cdots, 8)$ and $a_9 = 2$, the result from the former formula is 0.000694 and from the latter 0.000685. Using 0.004, (13) gives the approximate value 0.00068.

5. An application. The natural tolerance limits for a controlled process often are taken as $\mu \pm 3\sigma$. Let us suppose that the individual components are symmetrically distributed originally and then are symmetrically truncated by inspection with bases, a_i. Birnbaum [4] has proved that the distribution of the sum of the truncated variables is "more peaked" about the mean than the distribution of the sum of rectangular variables with the same bases, where "more peaked" means less probability for values more than any arbitrary distance from the mean. Thus, as Birnbaum points out for the case of equal truncations, the distribution of the sum of rectangulars can be used to get an upper bound for useful probabilities required for the sum of the truncated variables. For symmetric but unequal truncations, an upper bound to the percentage outside natural tolerance limits can be calculated by using formula (2) above.

REFERENCES

[1] F. S. Acton and E. G. Olds, "Tolerances—additive or Pythagorean?" *Industrial Quality Control*, Vol. 5 (1948), pp. 6–12.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] M. Loève, "Fundamental limit theorems of probability theory," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 321–338.

[4] Z. W. Birnbaum, "On random variables with comparable peakedness," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 76–80.