Canad. Math. Bull. Vol. **29** (2), 1986

ASYMPTOTIC APPROXIMATION OF AN INTEGRAL INVOLVING THE NORMAL DISTRIBUTION*

BY

J. P. McCLURE AND R. WONG†

ABSTRACT. An asymptotic approximation is obtained, as $k \to \infty$, for the integral

$$I(k) = \int_{-\infty}^{\infty} [\Phi(x) + 1 - \Phi(x + L)]^{k-1} d\Phi(x),$$

where Φ is the cumulative distribution function for a standard normal random variable, and L is a positive constant. The problem is motivated by a question in statistics, and an outline of the application is given. Similar methods may be used to approximate other integrals involving the normal distribution.

1. **Introduction**. Suppose $X_1, \ldots, X_k$ are k independent standard normal random variables, which have been put in natural order: $X_1 \leq X_2 \leq \ldots \leq X_k$. For a fixed number $L \geq 0$, let $N(L, k)$ be the number of "gaps" $g_j = X_{j+1} - X_j$ ($j = 1, \ldots, k - 1$) satisfying $g_j \geq L$. Then $N(L, k)$ is itself a random variable. In 1949, Tukey [2] showed that the mean or expected value $EN(L, k)$ of $N(L, k)$ satisfies

$$(1.1) \qquad 1 + EN(L, k) = k \int_{-\infty}^{\infty} [\Phi(x) + 1 - \Phi(x + L)]^{k-1} d\Phi(x),$$

where

$$(1.2) \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

is the cumulative distribution function for a standard normal random variable. Recently, Professor I. Olkin raised the question of the behavior of $EN(L, k)$ or, equivalently, the behavior of the integral

$$(1.3) \qquad I(k) = \int_{-\infty}^{\infty} [\Phi(x) + 1 - \Phi(x + L)]^{k-1} d\Phi(x)$$

as $k \to \infty$. The purpose of this paper is to provide the first two terms of an asymptotic expansion for $I(k)$ as $k \to \infty$. Our result implies

Received by the editors July 6, 1984.

AMS Subject Classification: 41A60, 62E20.

*This research was partially supported by Natural Sciences and Engineering Research Council of Canada grants A-7359 and A-8069.

†I. W. Killam Research Fellow.

(1.4) $$kI(k) \sim 1 + \frac{2e^{-L^2/2}}{e^{L\sqrt{2\log k}}}, \quad \text{as } k \to \infty,$$

when L is positive. From (1.1), an asymptotic approximation for $EN(L, k)$ follows, and in particular, we see that $EN(L, k) \to 0$ as $k \to \infty$, if $L > 0$. This may seem surprising compared with $EN(0, k) = k - 1$ (either from (1.1), or directly from the definition), but is actually quite reasonable, when one thinks of the values of the random variables X_j clustering around the mean value zero, so that the gaps should get smaller as the number of variables increases.

An integral somewhat similar to (1.3) is

(1.5) $$J(k) = \int_{-\infty}^{\infty} xe^{-x^2}\left[\frac{1 + \theta(x)}{2}\right]^k dx,$$

where

(1.6) $$\theta(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du.$$

Methods similar to those employed in the following sections lead to the result

(1.7) $$J(k) = \frac{\sqrt{\pi}\log(k+1)}{k+1} - \frac{\sqrt{\pi}}{4}\frac{\log\log(k+1)}{(k+1)\sqrt{\log(k+1)}} + 0\left(\frac{1}{(k+1)\log(k+1)}\right),$$

as $k \to \infty$. The details of this last result can be found in [3].

2. **A Transcendental Equation**. For convenience, let us write

(2.1) $$\Psi(x) = \Phi(x) + 1 - \Phi(x + L).$$

Later in our derivation (§4), we shall make the change of variable

(2.2) $$\Psi(x) = e^{-t}$$

in the integral $I(k)$. For this reason we need the following result.

LEMMA 1. *For small positive t the real roots of equation (2.2) are given by*

(2.3) $$x = \sqrt{-2\log t} + 0\left\{\frac{\log(-2\log t)}{\sqrt{-2\log t}}\right\}$$

(2.4) $$x = -\sqrt{-2\log t} - L + 0\left\{\frac{\log(-2\log t)}{\sqrt{-2\log t}}\right\}.$$

PROOF. We begin with the well-known formula

(2.5) $$\Phi(x) = 1 - \frac{e^{-x^2/2}}{\sqrt{2\pi}\,x}\left[1 + 0\left(\frac{1}{x^2}\right)\right],$$

valid for large positive x. It is straightforward to show that

$$\Psi(x) = 1 - \frac{e^{-x^2/2}}{\sqrt{2\pi}\,x}\left[1 + 0\left(\frac{1}{x^2}\right)\right],$$

also valid for large positive x. Hence, equation (2.2) can be written as

$$\frac{e^{-x^2/2}}{\sqrt{2\pi}\,x}\left[1 + 0\left(\frac{1}{x^2}\right)\right] = 1 - e^{-t}$$

or

$$\frac{e^{-x^2/2}}{\sqrt{2\pi}\,x}\left[1 + 0\left(\frac{1}{x^2}\right)\right] = t[1 + 0(t)], \tag{2.6}$$

as $x \to +\infty$ and $t \to 0^+$. The last equation gives

$$-x^2 - \log 2\pi - 2\log x + 0\left(\frac{1}{x^2}\right) = 2\log t + 0(t). \tag{2.7}$$

When x is large, the left-hand side is dominated by the first term. By the same reasoning, the right-hand side is also dominated by its leading term. Thus it follows that

$$x^2 \sim -2\log t \qquad (t \to 0^+), \tag{2.8}$$

and

$$x \sim \sqrt{-2\log t} \qquad (t \to 0^+).$$

This is the first approximation to the positive root. To improve this result, we set

$$x^2 = -2\log t + \epsilon(t). \tag{2.9}$$

Note that by (2.8), we have

$$\frac{\epsilon(t)}{-2\log t} \to 0 \qquad (t \to 0^+). \tag{2.10}$$

From (2.7), it also follows that

$$\begin{aligned} 2\log t - \epsilon(t) - \log 2\pi - \log(-2\log t + \epsilon(t)) \\ = 2\log t + o(1). \end{aligned}$$

Hence, as $t \to 0^+$,

$$\begin{aligned} \epsilon(t) &= -\log(-2\log t) - \log 2\pi - \log\left(1 - \frac{\epsilon(t)}{2\log t}\right) + o(1) \\ &= -\log(-2\log t) - \log 2\pi + o(1), \end{aligned} \tag{2.11}$$

in view of (2.10). Substituting (2.11) in (2.9) yields

$$x^2 = (-2\log t)\left[1 + 0\left(\frac{\log(-2\log t)}{-2\log t}\right)\right] \qquad (t \to 0^+),$$

https://doi.org/10.4153/CMB-1986-028-x Published online by Cambridge University Press

and

$$x = \sqrt{-2 \log t} + 0\left\{\frac{\log(-2 \log t)}{\sqrt{-2 \log t}}\right\} \qquad (t \to 0^+), \tag{2.12}$$

thus proving (2.3).

To obtain the negative root, we note that

$$\Phi(x) = 1 - \Phi(-x).$$

This together with (2.5) gives

$$\Phi(x) = -\frac{e^{-x^2/2}}{\sqrt{2\pi}\, x}\left[1 + 0\left(\frac{1}{x^2}\right)\right]$$

as $x \to -\infty$. From (2.1), we also have

$$\Psi(x) = 1 + \frac{e^{-(x+L)^2/2}}{\sqrt{2\pi}\,(x + L)}\left[1 + 0\left(\frac{1}{x^2}\right)\right]$$

as $x \to -\infty$. Equation (2.2) can thus be written in the form

$$-\frac{e^{-(x+L)^2/2}}{\sqrt{2\pi}\,(x + L)}\left[1 + 0\left(\frac{1}{x^2}\right)\right] = 1 - e^{-t}$$

or

$$-\frac{e^{-(x+L)^2/2}}{\sqrt{2\pi}\,(x + L)}\left[1 + 0\left(\frac{1}{x^2}\right)\right] = t[1 + 0(t)],$$

as $x \to -\infty$ and $t \to 0^+$. The last equation is equivalent to (2.6), except that $-(x + L)$ now replaces x. Thus, by (2.12),

$$-(x + L) = \sqrt{-2 \log t} + 0\left\{\frac{\log(-2 \log t)}{\sqrt{-2 \log t}}\right\} \qquad (t \to 0^+),$$

or equivalently

$$x = -\sqrt{-2 \log t} - L + 0\left\{\frac{\log(-2 \log t)}{\sqrt{-2 \log t}}\right\} \qquad (t \to 0^+).$$

This completes the proof of the lemma. □

3. **A Basic Integral**. In establishing our final result, we will encounter the integral

$$I_2^+(k) = \int_0^{c_1} e^{-kt - L\sqrt{-2 \log t}}\, dt \tag{3.1}$$

twice, where c_1 depends on k and is explicitly given by

$$c_1 = \frac{1}{\sqrt{k}}. \tag{3.2}$$

The following lemma gives the behavior of this integral for large values of k.

LEMMA 2. *As $k \to +\infty$, we have*

$$(3.3) \qquad I_2^+(k) = \frac{1}{k} e^{-L\sqrt{2 \log k}} \left[1 + 0\left(\frac{\log \log k}{\sqrt{\log k}} \right) \right].$$

PROOF. We split the interval of integration $(0, c_1)$ into $(0, a)$ (a, b) and (b, c_1), where

$$a = \frac{1}{k \log k} \text{ and } b = \frac{\log k}{k}.$$

For $0 < t \leq a$, $e^{-kt} \leq 1$ and $\sqrt{2 \log k} \leq \sqrt{-2 \log t} < \infty$. Hence, the integrand of $I_2^+(k)$ is dominated by $e^{-L\sqrt{2 \log k}}$, and

$$(3.4) \qquad \int_0^a e^{-kt - L\sqrt{-2 \log t}}\, dt \leq \frac{1}{k \log k} e^{-L\sqrt{2 \log k}}.$$

For $b \leq t \leq c_1$, $e^{-kt} \leq k^{-1}$ and $\sqrt{\log k} \leq \sqrt{-2 \log t} \leq \sqrt{2 \log k}$. Thus, similarly, the integrand of $I_2^+(k)$ is dominated by $k^{-1} e^{-L\sqrt{\log k}}$ and

$$(3.5) \qquad \int_b^{c_1} e^{-kt - L\sqrt{-2 \log t}}\, dt \leq \frac{1}{k^{3/2}} e^{-L\sqrt{\log k}}.$$

Finally, we consider the integral over the interval (a, b). Making the change of variable $kt = \tau$ gives

$$(3.6) \qquad \int_a^b e^{-kt - L\sqrt{-2 \log t}}\, dt = \frac{1}{k} \int_{1/\log k}^{\log k} e^{-\tau - L\sqrt{2 \log k - 2 \log \tau}}\, d\tau.$$

For $1/\log k \leq \tau \leq \log k$, we have by the binominal theorem

$$\sqrt{2 \log k - 2 \log \tau} = \sqrt{2 \log k} + 0\left(\frac{\log \log k}{\sqrt{\log k}} \right).$$

From (3.6), it follows that

$$(3.7) \qquad \int_a^b e^{-kt - L\sqrt{-2 \log t}}\, dt = \frac{1}{k} e^{-L\sqrt{2 \log k}} \int_{1/\log k}^{\log k} e^{-\tau} \left[1 + 0\left(\frac{\log \log k}{\sqrt{\log k}} \right) \right] d\tau.$$

Clearly,

$$\int_{1/\log k}^{\log k} e^{-\tau} d\tau = 0(1).$$

Thus (3.7) gives

$$(3.8) \qquad \int_a^b e^{-kt - L\sqrt{-2 \log t}}\, dt = \frac{1}{k} e^{-L\sqrt{2 \log k}} \left[1 + 0\left(\frac{\log \log k}{\sqrt{\log k}} \right) \right].$$

Note that as $k \to +\infty$,

$$\frac{1}{k^{1/2}}e^{(\sqrt{2}-1)L\sqrt{\log k}} = o\left(\frac{\log\log k}{\sqrt{\log k}}\right).$$

The desired result (3.3) now follows from (3.4), (3.5) and (3.8).□

4. **Proof of (1.4)**. Let us first rewrite (1.3) as

$$(4.1) \qquad I(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}[\Psi(x)]^{k-1}e^{-x^2/2}dx.$$

From (2.1) and (1.2), it is clear that

$$\Psi'(x) = \frac{1}{\sqrt{2\pi}}[e^{-x^2/2} - e^{-(x+L)^2/2}].$$

Thus $\Psi(x)$ has exactly one critical point, which is located at $x = -L/2$, and is increasing in $[-L/2, \infty)$ and decreasing in $(-\infty, -L/2]$. Furthermore, $\Psi(x) \to 1$ as $x \to \pm\infty$. The graph of $\Psi(x)$ is shown in the figure below.

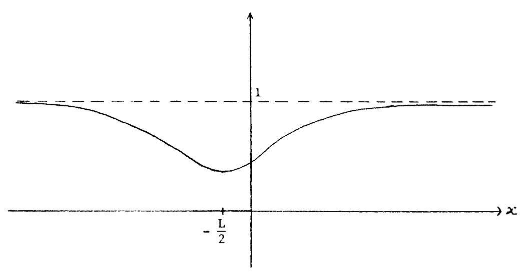

FIG. 1. *The function* $\Psi(x)$

Since the maximum of $\Psi(x)$ does not occur at a finite point, the well-known method of Laplace [1, p. 80] does not apply. Put

$$(4.2) \qquad I^+(k) = \frac{1}{\sqrt{2\pi}}\int_{-L/2}^{\infty}\Psi(x)^{k-1}e^{-x^2/2}dx$$

and

$$(4.3) \qquad I^-(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-L/2}\Psi(x)^{k-1}e^{-x^2/2}dx.$$

Then

$$(4.4) \qquad I(k) = I^+(k) + I^-(k).$$

We shall first treat the integral $I^+(k)$. For fixed c in $(-L/2, \infty)$, we have

$$(4.5) \qquad I^+(k) = \frac{1}{\sqrt{2\pi}}\int_{-L/2}^{c}\Psi(x)^{k-1}e^{-x^2/2}dx + \frac{1}{\sqrt{2\pi}}\int_{c}^{\infty}\Psi(x)^{k-1}e^{-x^2/2}dx.$$

Observe that

$$\frac{1}{\sqrt{2\pi}}\int_{-L/2}^{c}\Psi(x)^{k-1}e^{-x^2/2}dx \leq \Psi(c)^{k-1} = e^{-c_1(k-1)}, \tag{4.6}$$

where

$$c_1 = -\log \Psi(c) > 0. \tag{4.7}$$

In the second integral on the right-hand side of (4.5), we shall make the change of variable given in (2.2), from which we have

$$\Psi'(x)\frac{dx}{dt} = -e^{-t}$$

and

$$\frac{dx}{dt} = -\frac{\sqrt{2\pi}\,e^{-t}}{e^{-x^2/2} - e^{-(x+L)^2/2}} = -\frac{\sqrt{2\pi}e^{-t}}{e^{-x^2/2}(1 - e^{-xL-L^2/2})}. \tag{4.8}$$

Hence,

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{c}^{\infty}\Psi(x)^{k-1}e^{-x^2/2}dx &= \int_0^{c_1} e^{-kt}\frac{1}{1 - e^{-xL-L^2/2}}dt \\ &= \int_0^{c_1} e^{-kt}\left[1 + e^{-xL-L^2/2} + \frac{e^{-2xL-L^2}}{1 - e^{-xL-L^2/2}}\right]dt.\end{aligned} \tag{4.9}$$

Clearly

$$\int_0^{c_1} e^{-kt}dt = \frac{1}{k}(1 - e^{-kc_1}), \tag{4.10}$$

and

$$\int_0^{c_1} e^{-kt-xL-L^2/2}dt = e^{-L^2/2}\int_0^{c_1} e^{-kt-xL}dt \equiv e^{-L^2/2}I_1^+(k). \tag{4.11}$$

Observe that equation (2.3) gives

$$e^{-xL} = e^{-L\sqrt{-2\log t}}\left[1 + 0\left(\frac{\log(-2\log t)}{\sqrt{-2\log t}}\right)\right], \tag{4.12}$$

as $t \to 0^+$. Inserting this in (4.11), we obtain

$$\begin{aligned}I_1^+(k) &= \int_0^{c_1} e^{-kt-L\sqrt{-2\log t}}\left[1 + 0\left(\frac{\log(-2\log t)}{\sqrt{-2\log t}}\right)\right]dt \\ &= I_2^+(k)\left[1 + 0\left(\frac{\log(-2\log c_1)}{\sqrt{-2\log c_1}}\right)\right],\end{aligned} \tag{4.13}$$

provided that c_1 is sufficiently small. The last condition is automatically satisfied, if we take $c_1 = 1/\sqrt{k}$ as in (3.2). With this choice of c_1, equation (4.13) becomes

https://doi.org/10.4153/CMB-1986-028-x Published online by Cambridge University Press

$$(4.14)\qquad I_1^+(k) = I_2^+(k)\left[1 + 0\left(\frac{\log(\log k)}{\sqrt{\log k}}\right)\right],$$

as $k \to +\infty$.

Now observe that as $x \to +\infty$,

$$\frac{e^{-2xL - L^2}}{1 - e^{-xL - L^2/2}} = 0(e^{-2xL}).$$

Hence, for sufficiently large c and for $x \geqslant c$, we have

$$(4.15)\qquad \frac{e^{-2xL - L^2}}{1 - e^{-xL - L^2/2}} = 0(e^{-2cL}).$$

Since equation (4.7) is equivalent to equation (2.2) with c and c_1 replacing x and t, respectively, we obtain from (4.12) and (3.2)

$$(4.16)\qquad e^{-2cL} = e^{-2L\sqrt{\log k}}\left[1 + 0\left(\frac{\log\log k}{\sqrt{\log k}}\right)\right].$$

Coupling (4.15) and (4.16) gives

$$\frac{e^{-2xL - L^2}}{1 - e^{-xL - L^2/2}} = 0(e^{-2L\sqrt{\log k}})$$

for $t < c_1 = k^{-1/2}$, and

$$(4.17)\qquad \int_0^{c_1} e^{-kt}\frac{e^{-2xL - L^2}}{1 - e^{-xL - L^2/2}}\,dt = 0(k^{-1}e^{-2L\sqrt{\log k}}).$$

By a combination of the results in (4.5), (4.6), (4.9), (4.10), (4.11), (4.14) and (4.17), we arrive at

$$(4.18)\qquad \begin{aligned} I^+(k) &= \frac{1}{k} + e^{-L^2/2}I_2^+(k)\left[1 + 0\left(\frac{\log\log k}{\sqrt{\log k}}\right)\right] \\ &\quad + 0(e^{-\sqrt{k}}) + 0(k^{-1}e^{-2L\sqrt{\log k}}). \end{aligned}$$

Substituting (3.3) in (4.18), and noting that

$$e^{-L(2-\sqrt{2})\sqrt{\log k}} = o\left(\frac{\log\log k}{\sqrt{\log k}}\right)$$

and

$$e^{-\sqrt{k}} = o\left(\frac{\log\log k}{k\sqrt{\log k}}e^{-L\sqrt{2\log k}}\right),$$

as $k \to +\infty$, we obtain

$$(4.19)\qquad I^+(k) = \frac{1}{k} + \frac{1}{k}e^{-L^2/2 - L\sqrt{2\log k}}\left[1 + 0\left(\frac{\log\log k}{\sqrt{\log k}}\right)\right],$$

as $k \to +\infty$.

https://doi.org/10.4153/CMB-1986-028-x Published online by Cambridge University Press

Now we turn to the consideration of $I^-(k)$, defined in (4.3). As in (4.5), we have, for $c \in (L/2, \infty)$,

$$(4.20)\quad I^-(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-c} \Psi(x)^{k-1}e^{-x^2/2}dx + \frac{1}{\sqrt{2\pi}}\int_{-c}^{-L/2} \Psi(x)^{k-1}e^{-x^2/2}dx,$$

and

$$(4.21)\qquad \frac{1}{\sqrt{2\pi}}\int_{-c}^{-L/2} \Psi(x)^{k-1}e^{-x^2/2}dx \leq \Psi(-c)^{k-1} = e^{-c_1(k-1)},$$

where

$$c_1 = -\log \Psi(-c) > 0.$$

In the first integral on the right-hand side of (4.20), we again make the change of variable given in (2.2). From (4.8), it is then easily seen that

$$(4.22)\qquad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-c} \Psi(x)^{k-1}e^{-x^2/2}dx = \int_0^{c_1} e^{-kt}\frac{e^{xL+L^2/2}}{1-e^{xL+L^2/2}}dt$$
$$= \int_0^{c_1} e^{-kt+xL+L^2/2}\left[1 + \frac{e^{xL+L^2/2}}{1-e^{xL+L^2/2}}\right]dt.$$

By equation (2.4) in Lemma 1,

$$xL + \frac{1}{2}L^2 = -L\sqrt{-2\log t} - \frac{1}{2}L^2 + 0\left\{\frac{\log(-2\log t)}{\sqrt{-2\log t}}\right\}.$$

as $t \to 0^+$. We again take $c_1 = k^{-1/2}$. *Then, for* $0 < t \leq c_1 = k^{-1/2}$,

$$\frac{\log(-2\log t)}{\sqrt{-2\log t}} = 0\left(\frac{\log\log k}{\sqrt{\log k}}\right)$$

and

$$e^{xL+L^2/2} = 0(e^{-L\sqrt{\log k}}).$$

Together with these estimates, equation (4.22) gives

$$(4.23)\qquad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-c} \Psi(x)^{k-1}e^{-x^2/2}dx = e^{-L^2/2}\int_0^{c_1} e^{-kt-L\sqrt{-2\log t}}\,dt$$
$$\cdot\left[1 + 0\left(\frac{\log\log k}{\sqrt{\log k}}\right)\right][1 + 0(e^{-L\sqrt{\log k}})].$$

The integral on the right-hand side of (4.23) has already been evaluated in Lemma 2. Since $e^{-c_1(k-1)} = 0(e^{-\sqrt{k}})$, we have, upon combining (3.3), (4.20), (4.21) and (4.23),

$$(4.24)\qquad I^-(k) = \frac{1}{k}e^{-L^2/2-L\sqrt{2\log k}}\left[1 + 0\left(\frac{\log\log k}{\sqrt{\log k}}\right)\right],$$

as $k \to +\infty$.

Coupling (4.19) and (4.24) gives the desired result

$$I(k) = \frac{1}{k} + \frac{2}{k} e^{-L^2/2 - L\sqrt{2 \log k}} \left[1 + 0\left(\frac{\log \log k}{\sqrt{\log k}}\right)\right], \tag{4.25}$$

as $k \to \infty$, from which (1.4) follows.

Acknowledgement. We are grateful to Professor F. W. J. Olver for communicating this problem to us, and to Professor I. Olkin for helpful correspondence about the background to the problem.

References

1. F. W. J. Olver, *Asymptotics and special functions*, Academic Press, New York, 1974.
2. J. W. Tukey, *Comparing individual means in the analysis of variance*, Biometrics, **5** (1949), pp. 99–114.
3. R. Wong, *A generalization of Watson's lemma*, Ph.D. Thesis, University of Alberta, 1969.

Department of Mathematics and Astronomy
University of Manitoba
Winnipeg, Canada R3T 2N2