

**AN ASYMPTOTIC EXPANSION
 OF A BETA-TYPE INTEGRAL AND ITS APPLICATION TO
 PROBABILITIES OF LARGE DEVIATIONS¹**

J. C. FU AND R. WONG

ABSTRACT. An asymptotic expansion is obtained for an incomplete beta-type integral, which arises in the study of probabilities of large deviations. The expansion obtained yields large deviation results for binomial, quantile, and related probabilities. Our approach is based on a generalized version of Laplace's method.

1. Introduction. In this paper, we are concerned with the asymptotic behavior of the integral

$$B_n = \frac{\Gamma(n+1)}{\Gamma(k_n)\Gamma(n-k_n+1)} \int_{a_n}^1 t^{k_n-1}(1-t)^{n-k_n} dt \quad (1.1)$$

as $n \rightarrow \infty$. In (1.1), we assume that k_n and a_n are real-valued functions of n satisfying

$$k_n/n = p + O(1/n) \quad (1.2)$$

and

$$n^k |a_n - a| \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad \text{for each } k = 1, 2, \dots, \quad (1.3)$$

where $0 < p < a < 1$, $c > 0$ and $\rho > 0$.

The above integral occurs in the study of probabilities of large deviations. The expansion obtained in this paper can be used to yield large deviation results for binomial, negative binomial and related probabilities, e.g., quantile probabilities. A brief discussion of these applications is included in §3.

To evaluate B_n asymptotically, one naturally attempts to use the method of Laplace [6, p. 80] or its generalizations [6, p. 331]. However, this method and its generalizations do not apply directly to this case. A modification of this method is given in §2, by means of which an asymptotic expansion is obtained for B_n .

2. Derivation of asymptotic expansion. To simplify some of the expressions which occur, we let I_n denote the integral in (1.1), i.e.,

$$I_n = \int_{a_n}^1 t^{k_n-1}(1-t)^{n-k_n} dt, \quad (2.1)$$

and put

$$\lambda(t) = t^p(1-t)^{1-p}. \quad (2.2)$$

Received by the editors January 12, 1979 and, in revised form, July 18, 1979.

AMS (MOS) subject classifications (1970). Primary 41A60, 60F10; Secondary 60F15.

Key words and phrases. Asymptotic expansion, Laplace's method, large deviation, sample quantile.

¹This work was supported in part by the National Research Council of Canada under grants NRC A-9216 and NRC A-7359.

Note that $0 < p < a < 1$ and that $\lambda(t)$ is positive and strictly decreasing in the interval $(p, 1)$. Furthermore, $0 < \lambda(p) < 1$.

The main result of this paper is stated in the following theorem.

THEOREM. *As $n \rightarrow \infty$, we have*

$$B_n \sim \frac{\Gamma(n+1)}{\Gamma(k_n)\Gamma(n-k_n+1)} [\lambda(a)]^n \left\{ \frac{c_{1,n}}{n} + \frac{c_{2,n}}{n^2} + \dots \right\}, \tag{2.3}$$

where the coefficients $c_{i,n}$ are bounded functions of n . In particular,

$$c_{1,n} = \frac{a}{a-p} \left(\frac{1-a}{a} \right)^{np-k_n+1} \tag{2.4}$$

and

$$c_{2,n} = \frac{a}{(a-p)^2} \left(\frac{1-a}{a} \right)^{np-k_n+1} \left(\frac{p(a-1)}{a-p} - np + k_n - 1 \right). \tag{2.5}$$

PROOF. Let $b = (1+a)/2$ and assume that n is sufficiently large so that all a_n lie to the right of p and to the left of b . Set

$$\begin{aligned} I_n &= \left(\int_{a_n}^a + \int_a^b + \int_b^1 \right) t^{k_n-1} (1-t)^{n-k_n} dt \\ &= I_{n,1} + I_{n,2} + I_{n,3}. \end{aligned} \tag{2.6}$$

From (1.3), we have

$$I_{n,1} = O[\lambda^n(a)\exp(-cn^p)], \tag{2.7}$$

and from (1.2) it follows that

$$I_{n,3} = O[\lambda^{n-\zeta}(b)] \tag{2.8}$$

for some $\zeta > 0$. Clearly

$$B_n = \frac{\Gamma(n+1)}{\Gamma(k_n)\Gamma(n-k_n+1)} [I_{n,1} + I_{n,2} + I_{n,3}]. \tag{2.9}$$

Estimates of $I_{n,1}$ and $I_{n,3}$ are given in (2.7) and (2.8). We only need to consider the integral

$$I_{n,2} = \int_a^b t^{k_n-1} (1-t)^{n-k_n} dt = \int_a^b \psi(t;n) e^{-nh(t)} dt, \tag{2.10}$$

where

$$h(t) = -p \log t - (1-p) \log(1-t) \tag{2.11}$$

and

$$\psi(t;n) = t^{k_n-np-1} (1-t)^{np-k_n}. \tag{2.12}$$

Note that for $a < t < b$, we have

$$h'(t) = (t-p)/t(1-t) > 0. \tag{2.13}$$

Therefore, the minimum of $h(t)$ occurs at $t = a$. We rewrite (2.10) as

$$I_{n,2} = e^{-nh(a)} \int_a^b G(t;n) e^{-nh'(a)(t-a)} dt \tag{2.14}$$

with

$$G(t; n) = \psi(t; n) \exp \left\{ -n \left[\frac{1}{2} h''(a)(t-a)^2 + \dots \right] \right\}. \quad (2.15)$$

The Taylor expansion of $G(t; n)$ at $t = a$ gives

$$G(t; n) = \sum_{j=0}^{m-1} \frac{1}{j!} G^{(j)}(a; n)(t-a)^j + \frac{1}{m!} G^{(m)}(\xi; n)(t-a)^m, \quad (2.16)$$

where m is any positive integer and $a < \xi < b$. Note that the numbers

$$M_j = \sup_{a < t < b} |\psi^{(j)}(t; n)|, \quad j = 0, 1, \dots, m, \quad (2.17)$$

are all finite and bounded in n . Furthermore, by using the inequality $x^l \exp(-x^2) < l!$ for $x > 0$ and $l = 0, 1, 2, \dots$, it is easy to see that

$$G^{(j)}(t; n) = O(n^{j/2}), \quad j = 0, 1, \dots, m, \quad (2.18)$$

where the O -symbol is independent of n and t for $a < t < b$. Inserting (2.16) into (2.14) and integrating term by term, we get

$$I_{n,2} = e^{-nh(a)} \left[\sum_{j=0}^{m-1} \frac{1}{j!} G^{(j)}(a; n) \int_a^b (t-a)^j \exp(-nh'(a)(t-a)) dt + \delta_{n,m} \right], \quad (2.19)$$

where the remainder satisfies

$$\begin{aligned} |\delta_{n,m}| &= O\left(n^{m/2} \int_a^b (t-a)^m \exp(-nh'(a)(t-a)) dt\right) \\ &= O(n^{-(m+2)/2}), \quad \text{as } n \rightarrow \infty. \end{aligned} \quad (2.20)$$

For each $j = 0, 1, 2, \dots$, there exists a positive number ε_j such that

$$\int_a^b (t-a)^j \exp(-nh'(a)(t-a)) dt = \frac{j!}{[nh'(a)]^{j+1}} + O(\exp(-\varepsilon_j n)), \quad (2.21)$$

as $n \rightarrow \infty$. Hence (2.19) can be rewritten as

$$I_{n,2} = e^{-nh(a)} \left\{ \sum_{j=0}^{m-1} G^{(j)}(a; n) [nh'(a)]^{-j-1} + O(n^{-(m+2)/2}) \right\}. \quad (2.22)$$

The O -terms from (2.21) are now included in that appearing in (2.22). Note that $e^{-h(a)} = \lambda(a)$ and that from (2.15)

$$G(a; n) = \psi(a; n), \quad G'(a; n) = \psi'(a; n), \quad (2.23)$$

and

$$G''(a; n) = \psi''(a; n) - nh''(a)\psi(a; n). \quad (2.24)$$

Taking the first three terms in the expansion (2.22) yields

$$I_{n,2} \sim [\lambda(a)]^n \left\{ \frac{\psi(a; n)}{nh'(a)} + \frac{\psi'(a; n)}{n^2 [h'(a)]^2} - \frac{h''(a)\psi(a; n)}{n^2 [h'(a)]^3} + \dots \right\}. \quad (2.25)$$

A straightforward calculation gives

$$h''(a) = \frac{p(1 - a)^2 + (1 - p)a^2}{a^2(1 - a)^2} \tag{2.26}$$

and

$$\psi'(a; n) = \left(\frac{1 - a}{a}\right)^{np - k_n + 1} \left(\frac{a + k_n - np - 1}{a(1 - a)^2}\right). \tag{2.27}$$

Thus, in terms of the coefficients $c_{1,n}$ and $c_{2,n}$ given in (2.4) and (2.5), we have

$$I_{n,2} \sim [\lambda(a)]^n \left\{ \frac{c_{1,n}}{n} + \frac{c_{2,n}}{n^2} + \dots \right\}, \text{ as } n \rightarrow \infty. \tag{2.28}$$

In view of the estimates (2.7) and (2.8) and the fact that $\lambda(b) < \lambda(a)$, the same is true of I_n . The final result (2.3) now follows from (2.9).

3. Applications. Let (X_1, \dots, X_n) be n independent identically distributed random variables with a common cumulative distribution function $F(x)$. We assume that $F(x)$ is continuous and strictly increasing in a finite or infinite interval $[c, d]$ with $F(c) = 0$ and $F(d) = 1$. Let $(X_{[1]}^{(n)}, \dots, X_{[n]}^{(n)})$ be the order statistics of (X_1, \dots, X_n) . For $0 < p < 1$, we denote by $\xi_p = F^{-1}(p)$ the population p -quantile, and by $Y_p^{(n)} = X_{[k]}^{(n)}$, $k = [np] + 1$, the sample p -quantile of the sample (X_1, \dots, X_n) .

From the stochastic convergence of the sample quantile to the population quantile (i.e. $Y_p^{(n)} \rightarrow \xi_p$ in probability for every $0 < p < 1$), it follows that the large deviation probabilities

$$\alpha_n(\epsilon, p) = P(Y_p^{(n)} \geq \xi_p + \epsilon) \tag{3.1}$$

tend to zero as $n \rightarrow \infty$, for every fixed $\epsilon > 0$ and $0 < p < 1$. Motivated by the needs in statistical inferences, several authors ([2, p. 25], [3], [5] and [7]) have recently investigated the rate of convergence of sample quantiles. In particular, they show that $\alpha_n(\epsilon, p)$ decays exponentially as $n \rightarrow \infty$. In this section we shall give even more precise asymptotic information than those in the above-mentioned references.

It is well known that the distribution function of the sample p -quantile $Y_p^{(n)}$ is given by

$$\Phi_{kn}(x) = \frac{n!}{(k - 1)!(n - k)!} \int_0^{F(x)} t^{k-1}(1 - t)^{n-k} dt, \tag{3.2}$$

where $k = [np] + 1$; see [4, p. 376, equation (10.3.8)]. Hence it follows that the tail probability $\alpha_n(\epsilon, p)$ of the sample p -quantile is

$$\begin{aligned} \alpha_n(\epsilon, p) &= P(Y_p^{(n)} \geq \xi_p + \epsilon) \\ &= \frac{n!}{(k - 1)!(n - k)!} \int_a^1 t^{k-1}(1 - t)^{n-k} dt, \end{aligned} \tag{3.3}$$

where $a = F(\xi_p + \epsilon)$. The right-hand side of (3.3) is exactly the integral B_n given in (1.1) with $a_n = a$ and $k_n = [np] + 1$ for all n . Thus the theorem of the preceding

section yields

$$\alpha_n(\varepsilon, p) \sim \frac{(n-1)!}{(k-1)!(n-k)!} [\lambda(a)]^n \frac{a}{a-p} \left(\frac{1-a}{a}\right)^{np-[np]} \quad (3.4)$$

with $k = [np] + 1$. By a straightforward calculation using the Stirling formula, we have

$$\alpha_n(\varepsilon, p) \sim \frac{b_n}{\sqrt{2\pi np(1-p)}} e^{-nK(a,p)}, \quad (3.5)$$

where

$$K(a, p) = p \log \frac{p}{a} + (1-p) \log \frac{1-p}{1-a} \quad (3.6)$$

and

$$b_n = a \left(\frac{1-p}{a-p}\right) \left(\frac{p(1-a)}{a(1-p)}\right)^{np-[np]}. \quad (3.7)$$

Note that in terms of the function $\lambda(t)$ given in (2.2), we have $K(a, p) = \log \lambda(p) - \log \lambda(a)$. Since $\lambda(t)$ is strictly decreasing in $(p, 1)$, it follows that $K(a, p) > 0$.

Special cases of (3.5) have been obtained previously by Book [3], Fu [5] and Steinebach [7]. Since

$$\alpha_n(\varepsilon, p) = \sum_{j=0}^{k-1} \binom{n}{j} a^j (1-a)^{n-j}, \quad (3.8)$$

the conclusion (3.5) is available also from results in Bahadur [1]. In view of this identity, we remark that our theorem in §2 also gives an asymptotic expansion for the binomial probability given in (3.8).

ACKNOWLEDGEMENT. We would like to thank the referee for some helpful suggestions.

REFERENCES

1. R. R. Bahadur, *Some approximations to the binomial distribution function*, Ann. Math. Statist. **31** (1960), 43–54.
2. ———, *Some limit theorems in statistics*, Regional Conf. Ser. Appl. Math., Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1970.
3. S. A. Book, *Large deviation probabilities for order statistics*, Naval. Res. Logist. Quart. **18** (1971), 521–523.
4. M. Fisz, *Probability theory and mathematical statistics*, 3rd ed., Wiley, New York, 1963.
5. J. C. Fu, *The rate of convergence of consistent point estimators*, Ann. Statist. **3** (1975), 234–240.
6. F. W. J. Olver, *Asymptotics and special functions*, Academic Press, New York, 1974.
7. J. Steinebach, *Exponential convergence properties of linear estimators under exponential and uniform distribution*, Metrika **24** (1977), 137–161.

DEPARTMENT OF STATISTICS, UNIVERSITY OF MANITOBA, WINNIPEG R3T 2N2, MANITOBA, CANADA

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MANITOBA, WINNIPEG R3T 2N2, MANITOBA, CANADA