The Annals of Probability
1990, Vol. 18, No. 2, 727–753

ON APPROXIMATING PROBABILITIES FOR SMALL AND LARGE DEVIATIONS IN $\mathbb{R}^d$ [1]

BY J. ROBINSON, T. HÖGLUND, L. HOLST AND M. P. QUINE

University of Sydney, Royal Institute of Technology, Stockholm, Uppsala University and University of Sydney

A unified approach to approximations of probabilities for sums of n independent random vectors in $\mathbb{R}^d$ is presented based on the Edgeworth expansion of exponentially shifted vectors together with explicit bounds on the errors. Weak conditions are given under which the error bounds may be written as simple order terms in n. These results are used in particular to examine approximations to conditional probabilities giving a general method of approximation for these. A number of important special cases are discussed and examined numerically.

0. Introduction. This paper extends in several ways the unified central limit theorem of Höglund (1979). We deal with the distribution of the sum S_n of n independent (but not necessarily identically distributed) random vectors in $\mathbb{R}^d$ with the first $d_0 \le d$ dimensions being lattice with span 1. The probability $P(S_n \in x - B)$ for a Borel set $B \subset \mathbb{R}^d$ is approximated by the Edgeworth expansion of the random variables, exponentially shifted with respect to an arbitrary parameter θ. Taking $\theta = 0$ is shown to yield the usual Edgeworth approximations. Choosing $\theta = \hat{\theta}$ to make the mean of the exponentially shifted sum equal to x corresponds to a choice of a saddlepoint of the standardized complex cumulant generating function and yields in many cases very accurate approximations of both densities and tail probabilities connected with both large and small deviations.

Approximations for densities which hold under rather strong conditions have been obtained by Barndorff-Nielsen and Cox (1979), extending the methods of Daniels (1954) for continuous random variables and Blackwell and Hodges (1959) for lattice random variables. We use methods related to those of Stone (1965, 1967) and also von Bahr (1967a, b) and Efron and Truax (1968) to obtain approximations under relatively weak conditions. These results extend those of Robinson (1982) for a particular case. They can also be used to give Edgeworth and saddlepoint approximations in a number of special cases considered in the literature (in particular, the case of conditional probabilities), where in general only limit theorems or Edgeworth approximations were obtained; these include Erdös and Rényi (1959), Holst (1979), Morris (1975), Quine (1985), Quine and Robinson (1984) and Robinson (1978).

The principal theorem is stated and proved in Section 1 giving a general approximation for $P(S_n \in x - B)$ based on the Edgeworth expansion and an

Received September 1987; revised April 1989.
[1]Part of this research was performed with the help of ARGC Grant A6/85/15642.
AMS 1980 *subject classifications*. Primary 60F05, 60F10; secondary 62E20.
Key words and phrases. Edgeworth expansions, saddlepoint approximations, large deviations, limit theorems, conditional probabilities.

explicit bound for the error. In Section 2, sufficient conditions are given so that the theorem yields local and integral Edgeworth expansions and saddlepoint approximations with errors of the appropriate order. Ratios of these are used to obtain approximations for conditional probabilities. Section 3 discusses some one- and two-dimensional cases and Section 4 discusses numerical techniques which can be used to implement the approximations in practice, illustrating the methods with reference to examples including an occupancy problem and some problems on the conditional distributions of some linear combinations of independent random variables which yield approximations to some nonparametric test statistics.

1. The principal result. In Section 1.1 we give the formal approximation and in Section 1.2 we derive an explicit bound for the error.

1.1. *The approximation.* Let $X_1, \dots, X_n$ be independent random vectors in $\mathbb{R}^d$, let $S_n = \sum_{i=1}^n x_i$ and let μ_n be the probaiblity measure of S_n. Write

$$\kappa_n(\theta) = \log Ee^{\theta \cdot S_n}, \tag{1.1}$$

where $\theta \cdot S_n$ denotes the inner product of θ and S_n and let $\Theta_n = \{\theta\colon \kappa_n(\theta) < \infty\}$. Then certainly $0 \in \Theta_n$. Put

$$\mu_{n\theta}(dy) = e^{-\kappa_n(\theta)+\theta\cdot y}\mu_n(dy), \tag{1.2}$$

the probability measure of the associated random variable $S_{n\theta}$, which is the sum of the independent associated variables $X_{1\theta}, \dots, X_{n\theta}$. Assume $\kappa_n(\theta)$ has two derivatives (defined formally if $\theta = 0$ only), and let

$$m_n(\theta) = \kappa_n'(\theta) \tag{1.3}$$

and

$$V_n(\theta) = \kappa_n''(\theta) \tag{1.4}$$

be the mean vector and covariance matrix of $\mu_{n\theta}$. We will assume that for $\theta \in \operatorname{int}(\Theta_n)$, $V_n(\theta)$ is positive definite. For any Borel set $B \subset \mathbb{R}^d$ let

$$\nu_{n\theta}(B) = \mu_{n\theta}(m_n(\theta) + B), \tag{1.5}$$

where $a + B = \{x\colon x - a \in B\}$. Then $\nu_{n\theta}$ has mean 0 and covariance matrix $V_n(\theta)$.

For any Borel set $B \subset \mathbb{R}^d$,

$$\mu_n(x - B) = \int \chi_B(x-y)\mu_n(dy), \tag{1.6}$$

where $\chi_B(u) = 1$ if $u \in B$ and 0 otherwise. Thus

$$\begin{aligned} \mu_n(x - B) &= e^{\kappa_n(\theta)}\int \chi_B(x-y)e^{-\theta\cdot y}\mu_{n\theta}(dy) \\ &= e^{\kappa_n(\theta)-\theta\cdot x}I(\theta, B, w), \end{aligned} \tag{1.7}$$

where $w = x - m_n(\theta)$ and

$$I(\theta, B, w) = \int \chi_B(w-u) e^{\theta\cdot(w-u)} \nu_{n\theta}(du). \tag{1.8}$$

In the sequel we will assume that the first d_0 variables are lattice with span 1 and that $B = \{y: y = (0, y_1),\ y_1 \in B_1 \subset \mathbb{R}^{d_1}\}$, where we use the subscripts 0 to denote the first d_0 variables and 1 to denote the last $d_1 = d - d_0$ variables. So, for the first d_0 variables, all results in this section are local.

We will approximate $I(\theta, B, w)$ by

$$e_{s-3}(\theta, B, w) = \int_{\mathbb{R}^{d_1}} \chi_B(w-u) e^{\theta\cdot(w-u)} e_{s-3}(u, \nu_{n\theta})\, du_1, \tag{1.9}$$

if $d_1 > 0$ and $e_{s-3}(\theta, B, w) = e_{s-3}(w, \nu_{n\theta})$ if $d_1 = 0$, where $e_k(u, \nu_{n\theta})$ is the formal Edgeworth expansion for $\nu_{n\theta}$ of order k. Formulas for this expansion are given in Bhattacharya and Ranga Rao (1976), Section 7, and in Barndorff-Nielsen and Cox (1979); however, we give the result here in a convenient notation for the examples given later. Let ν represent a d-vector of nonnegative integers and let $\kappa^*_{\nu n}$ be the νth cumulant of $Y_n^*(\theta) = V_n(\theta)^{-1/2} Y_n(\theta)$, where $Y_n(\theta)$ is a random vector with probability measure $\nu_{n\theta}$. Usually it will be convenient to take $V_n(\theta)^{-1/2}$ as a lower triangular matrix. Further, write $H_m(x)\varphi(x) = (-1)^m \varphi^{(m)}(x)$, where $\varphi(x) = (2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$, so $H_m(x)$ is the usual Hermite–Chebyshev polynomial of degree m. Then we can write

$$e_k(y, \nu_{n\theta}) = \left(1 + \sum_{l=1}^{k} Q_{ln}(y^*)\right)(2\pi)^{-d/2} \det V_n(\theta)^{-1/2} e^{-1/2|y^*|^2}, \tag{1.10}$$

where $y^* = V_n(\theta)^{-1/2} y$, $|y^*|^2 = y^*\cdot y^*$ and

$$Q_{ln}(y^*) = \sum_{m=1}^{l} \frac{1}{m!} \sum{}^{*} \sum{}^{**} \frac{\kappa^*_{\nu_1 n} \cdots \kappa^*_{\nu_m n}}{\nu_1! \cdots \nu_m!} H_{I_1}(y_1^*) \cdots H_{I_d}(y_d^*), \tag{1.11}$$

where Σ^* denotes the sum over all m-tuples of positive integers $(j_1, \ldots, j_m)$ satisfying $j_1 + \cdots + j_m = l$, Σ^{**} denotes the sum over all m-tuples $(\nu_1, \ldots, \nu_m)$, with $\nu_i = (\nu_{1i}, \ldots, \nu_{di})$, satisfying $(\nu_{1i} + \cdots + \nu_{di} = j_i + 2,\ i = 1, \ldots, m)$, and $I_h = \nu_{h1} + \cdots + \nu_{hm}$, $h = 1, \ldots, d$. The use of standardized variables in (1.10) and (1.11) allows a convenient general expression agreeing, in particular, with the formulas of Barndorff-Nielsen and Cox (1979) for the first two terms.

Note that if $x = m_n(\theta)$ has a solution $\hat\theta = \hat\theta_n = \hat\theta_n(x)$ and if B_1 is a fixed compact set, then, taking $s = 3$, we obtain for $\mu_n(x - B)$ the saddlepoint approximation

$$e^{\kappa_n(\hat\theta) - \hat\theta\cdot x}(2\pi)^{-d/2}\left(\det V_n(\hat\theta)\right)^{-1/2} \operatorname{vol}(B_1).$$

Special cases of higher-order approximations are given in Section 3.

1.2. *The error bound.* First we approximate the convolution of $I(\theta, B, w)$ with a smoothing density over the last d_1 variables. Let

$$V(\theta, B, T, w) = \int I(\theta, B, w - u) K_T(du) \tag{1.12}$$

and

$$e_{s-3}(\theta, B, T, w) = \int e_{s-3}(\theta, B, w - u) K_T(du), \tag{1.13}$$

where, writing $u_1 = (u_{11}, \dots, u_{1d_1})$,

$$k(u_1) = \prod_{i=1}^{d_1} \left[(1 - \cos u_{1i}) / \pi u_{1i}^2 \right], \tag{1.14}$$

$K_T(B) = \int_{B_1} T^{d_1} k(T u_1)\, du_1$ for $d_1 > 0$ and $K_T(B) = 1$, otherwise. Then the characteristic function of this probability measure is

$$\hat{K}_T(\xi) = \begin{cases} \prod_{i=1}^{d_1} (1 - |\xi_{1i}|/T), & -T \le \xi_{1i} \le T,\ i = 1, \dots, d_1, \\ 0, & \text{otherwise}. \end{cases} \tag{1.15}$$

So

$$\hat{V}(\theta, B, T, \xi) = \hat{K}_T(\xi) \hat{\chi}_{\theta, B}(\xi) \hat{\nu}_{n\theta}(\xi), \tag{1.16}$$

where $\chi_{\theta, B}(u) = \chi_B(u) e^{\theta \cdot u}$, $\hat{\chi}_{\theta, B}$ is the Fourier transform of this function, if it exists, and $\hat{\nu}_{n\theta}$ is the characteristic function of $\nu_{n\theta}$. If $d_1 = 0$ then $B = \{0\}$ and $\chi_{\theta, B}(u) = 1$ if $u = 0$, 0 otherwise. $\hat{\chi}_{\theta, B}$ exists if

$$\hat{\chi}_{\theta, B}(0) = \int \chi_B(u) e^{\theta \cdot u}\, du_1 < \infty. \tag{1.17}$$

This is so if $\mathrm{vol}(B_1) < \infty$ and also for some sets B such that $\theta \cdot u < 0$ for $u \in B$. Also $\hat{\chi}_{\theta, B}(0) = 0$ if $\mathrm{vol}(B_1) = 0$ for $d_1 > 0$.

We will approximate $\hat{\nu}_{n\theta}(\xi)$ and use the inversion theorem for Fourier transforms. In the sequel C, c will denote constants which may be different on each occurrence. From Theorem 9.11 of Bhattacharya and Ranga Rao (1976), and using their notation, if $\nu_{n\theta}$ is the measure of a sum of n independent random vectors with finite sth moments,

$$\hat{\nu}_{n\theta}(\xi) = \hat{e}_{s-3}(\xi, \nu_{n\theta}) + R_{s-2}(\xi, \nu_{n\theta}), \tag{1.18}$$

where

$$|R_{s-2}(\xi, \nu_{n\theta})| \le C \eta_s(\theta) n^{-(s-2)/2} \left[\left| V_n(\theta)^{1/2} \xi \right|^s + \left| V_n(\theta)^{1/2} \xi \right|^{3(s-2)} \right] \times e^{-\xi \cdot V_n(\theta) \xi / 4}, \tag{1.19}$$

for $|V_n(\theta)^{1/2}\xi| \le C\eta_s(\theta)^{-1/s}n^{(s-2)/2s}$, where

$$\eta_s(\theta) = n^{(s-2)/2}\sum_{j=1}^{n} E\left|V_n(\theta)^{-1/2}(X_{j\theta} - EX_{j\theta})\right|^s. \tag{1.20}$$

Note that if $\theta \in \operatorname{int}(\Theta_n)$, then certainly these moments are finite, although $\eta_s(\theta)$ may depend on n and may not be uniformly bounded. Also, from Theorem 8.9 of Bhattacharya and Ranga Rao (1976), we have

$$|\hat{\nu}_{n\theta}(\xi)| < e^{-\xi\cdot V_n(\theta)\xi/4} \tag{1.21}$$

for

$$\left|V_n(\theta)^{1/2}\xi\right| < \tfrac{3}{4}l_{3n}(\theta)^{-1} \tag{1.22}$$

and so for

$$\left|V_n(\theta)^{1/2}\xi\right| < \tfrac{3}{4}\eta_3(\theta)^{-1}n^{1/2} < \tfrac{3}{4}l_{3n}(\theta)^{-1},$$

where

$$l_{3n}(\theta) = \sup_{|t|=1}\sum_{j=1}^{n} E\left|t\cdot V_n(\theta)^{-1/2}(X_{j\theta} - EX_{j\theta})\right|^3. \tag{1.23}$$

Further, if

$$\begin{aligned} q_{n\theta}(T) = \sup\Big\{|\hat{\nu}_{n\theta}(\xi)|: \left|V_n(\theta)^{1/2}\xi\right| > \tfrac{3}{4}\eta_3(\theta)^{-1}n^{1/2},\ |\xi_i| < \pi,\\ i = 1,\dots,d_0,\ |\xi_i| < T,\ i = d_0+1,\dots,d\Big\}, \end{aligned} \tag{1.24}$$

where we note that $q_{n\theta} = q_{n\theta}(T)$ does not depend on T if $d_1 = 0$, then, since

$$\begin{aligned} &|V(\theta,B,T,w) - e_{s-3}(\theta,B,T,w)| \\ &\quad = (2\pi)^{-d}\left|\int e^{-i\xi\cdot w}\hat{K}_T(\xi)\hat{\chi}_{\theta,B}(\xi)\left[\hat{\nu}_{n\theta}(\xi) - \hat{e}_{s-3}(\xi,\nu_{n\theta})\right]d\xi\right|, \end{aligned} \tag{1.25}$$

where the integral is over $(-\pi,\pi)$ for $\xi_1,\dots,\xi_{d_0}$ and over $(-T,T)$ for $\xi_{d_0+1},\dots,\xi_d$, we have the following result.

LEMMA 1.

$$\begin{aligned} \Delta(\theta,B,T) &= \sup_w|V(\theta,B,T,w) - e_{s-3}(\theta,B,T,w)| \\ &\le C\hat{\chi}_{\theta,B}(0)\left[\eta_s(\theta)(\det V_n(\theta))^{-1/2}n^{-(s-2)/2} + T^{d_1}q_{n\theta}(T)\right]. \end{aligned} \tag{1.26}$$

The theorem will follow immediately from this and the following smoothing lemma whose proof is given in Appendix A; related smoothing lemmas appear in Bhattacharya and Ranga Rao (1976) and von Bahr (1967a).

LEMMA 2. *If* $|\theta| < C$ *and* $\varepsilon = c/T$, *where* $K_T(S(\varepsilon)) = 1 - \alpha > 2/3$, *for* $S(\varepsilon) = \{y\colon\ y = (0, y_1) \in \mathbb{R}^d,\ |y_1| < \varepsilon\}$, *then writing* $\delta(\theta, B, w) = I(\theta, B, w) - e_{s-3}(\theta, B, w)$, $B_\varepsilon = \{x + y\colon x \in B,\ y \in S(\varepsilon)\}$ *and* $B_{-\varepsilon} = ((B^c)_\varepsilon)^c$, *we have*

$$\sup_w |\delta(\theta, B, w)| \le (1 - 3\alpha)^{-1}\Big\{\max(\Delta(\theta, B_\varepsilon, T), \Delta(\theta, B_{-\varepsilon}, T)) + (1 + \alpha) e^*_{s-3}(\nu_{n\theta}) \hat{\chi}_{\theta, B_{2\varepsilon} \setminus B_{-2\varepsilon}}(0)\Big\}, \tag{1.27}$$

where

$$e^*_{s-3}(\nu_{n\theta}) = \sup_u |e_{s-3}(u, \nu_{n\theta})| \le C(\det V_n(\theta))^{-1/2}(1 + \eta_s(\theta) n^{-1/2}).$$

THEOREM 1. *If* $|\theta| < C$, $x \in \mathbb{Z}^{d_0} \times \mathbb{R}^{d_1}$ *and* $\varepsilon = c/T$,

$$\begin{aligned} &\left|\mu_n(x - B) - e^{\kappa_n(\theta) - \theta\cdot x} e_{s-3}(\theta, B, x - m_n(\theta))\right| \\ &\quad \le e^{\kappa_n(\theta) - \theta \cdot x}(\det V_n(\theta))^{-1/2} \\ &\quad\quad \times C\Big[\hat{\chi}_{\theta, B_{2\varepsilon}}(0)\big(\eta_s(\theta) n^{-(s-2)/2} + (\det V_n(\theta))^{1/2} T^{d_1} q_{n\theta}(T)\big) \\ &\quad\quad\quad + \hat{\chi}_{\theta, B_{2\varepsilon}\setminus B_{-2\varepsilon}}(0)\Big]. \end{aligned} \tag{1.28}$$

2. General conditions and expansions. This section gives conditions under which the bound of the theorem may be used to give simple order terms for asymptotic approximations.

2.1. *Direct Edgeworth expansions.* Consider the case $\theta = 0$, and consider a triangular array $\{X_{1n}, \dots, X_{nn}\}$ of independent random vectors in $\mathbb{R}^d$ with $S_n = \sum_{i=1}^n X_{in}$, with the same notation as in Section 1 for $X_1, \dots, X_n$, applied to $X_{1n}, \dots, X_{nn}$, and

$$ES_n = 0, \qquad \mathrm{Cov}(S_n) = V_n = V_n(0).$$

We will consider the following conditions:

(E.1) There exists a positive-definite matrix V such that

$$n^{-1} V_n \to V \quad \text{as } n \to \infty.$$

(E.2) $\eta_s(0) < C$.
(E.3) $q_{n0}(n^{(s-3)/2}) < Cn^{-\lambda}$, or if $d_1 = 0$, $q_{n0} < Cn^{-\lambda}$, where

$$\lambda = d_1(s - 3)/2 + (s - 2)/2 + d/2.$$

Let $\mathscr{B}(\varepsilon)$ be the class of Borel sets $B \subset \mathbb{R}^d$ such that

$$\mathrm{vol}((B_\varepsilon)_1) < C\,\mathrm{vol}(B_1) \quad \text{and} \quad \mathrm{vol}((B_\varepsilon \setminus B_{-\varepsilon})_1) < C\varepsilon\,\mathrm{sur}(B_1), \tag{2.1}$$

where vol and sur are volume and surface area; for notational convenience we will take $\mathrm{vol}(B_1) = 1$ and $\mathrm{sur}(B_1) = 0$ if $d_1 = 0$ and $\mathrm{sur}(B_1) = 1$ if $d_1 = 1$. Then, using the fact that $\hat{\chi}_{0,B}(0) = \mathrm{vol}(B_1)$, and taking $T = n^{(s-3)/2}$ in Theorem 1, we obtain the following local result.

COROLLARY 2.1. *If* $B \in \mathscr{B}(n^{-(s-3)/2})$ *and conditions* (E.1), (E.2) *and* (E.3) *hold, then*

$$\left|\mu_n(x-B) - \int_{\mathbb{R}^{d_1}} \chi_{x-B}(u) e_{s-3}(u, \mu_n)\, du_1\right| = \big(\mathrm{sur}(B_1) n^{1/2} + \mathrm{vol}(B_1)\big) O\big(n^{-d/2-(s-2)/2}\big). \tag{2.2}$$

To obtain an integral version of this result, we may use the device in Corollary 3.3 of Skovgaard (1986) to obtain the following corollary.

COROLLARY 2.2. *If* (E.1), (E.2) *and* (E.3) *hold*

$$\left|\mu_n(x-B) - \int_{\mathbb{R}^{d_1}} \chi_{x-B}(u) e_{s-3}(u, \mu_n)\, du_1\right| = O\big((\log n) n^{-d_0/2-(s-2)/2}\big). \tag{2.3}$$

REMARK 1. If $X_1, X_2, \dots$ are iid, $EX_1 = 0$, $\mathrm{Cov}(X_1) = V$, which is positive definite, if $E|X_1|^s < \infty$ and if

$$|Ee^{it\cdot X_1}| < 1 - \delta \quad \text{for some } \delta > 0,\ c < t < Cn^{(s-3)/2}, \tag{2.4}$$

then conditions (E.1), (E.2) and (E.3) are easily shown to be satisfied and the corollaries above hold under these conditions. If in addition X_1 has an integrable characteristic function and f_n is the density of μ_n with respect to the direct product of the counting measure on $\mathbb{Z}^{d_0}$ and the Lebesgue measure on $\mathbb{R}^{d_1}$, then, taking $B = S(n^{-(s-2)/2})$ and $T = n^{-(s-2)/2}$ in Theorem 1, we have

$$\left|\mu_n^{(x-B)} - \int_{\mathbb{R}^{d_1}} \chi_{x-B}(u) e_{s-3}(u, \mu_n)\, du_1\right| = \mathrm{vol}(B_1) O(n^{-d/2-(s-2)/2}),$$

since $\hat{\chi}_{0, B_{2\varepsilon} \setminus B_{-2\varepsilon}}(0) < C \mathrm{vol}(B_1)$ and $\hat{\chi}_{0, B_{2\varepsilon}}(0) < C \mathrm{vol}(B_1)$. Dividing both sides here by $\mathrm{vol}(B_1)$ and using the fact that f_n has bounded derivatives for $n > 2$, we can see that

$$|f_n(x) - e_{s-3}(x, \mu_n)| = O(n^{-d/2-(s-2)/2}). \tag{2.5}$$

REMARK 2. If $d = d_1 = 1$ and B is any finite interval, then $|\hat{\chi}_{0,B}(\xi)| < 2/|\xi|$; so using this bound in (1.25) we can replace the bound on the right in (1.26) by

$$C\big[\eta_s(\theta) n^{-(s-2)/2} + q_{n\theta}(T) \log\big(Tn^{-1/2}\eta_3(\theta)\big)\big].$$

For B a half-line, we may bound the difference in $B \cap (-n, n)$ as above and bound it in $(-n, n)^c$ by $O(n^{-(s-2)/2})$ using a Chebyshev inequality. Thus under conditions (E.1)–(E.3), Corollary 2.2 holds with $\log n$ replaced by 1. When $d > 1$ it is possible to improve the bound in (2.3) using the methods of von Bahr (1967a).

REMARK 3. We could include some more general cases in the nonidentically distributed situation by defining a standardizing sequence $\rho_n \to \infty$ as $n \to \infty$ and adjusting (E.1)–(E.3) appropriately. For example, if X_{in} are independent Bernoulli random variables with $p_i = i^{-1}$, then it is necessary to take $\rho_n = \log n$.

REMARK 4. Suppose we wish to obtain an asymptotic approximation to

$$P(S_{1n} \in x_1 - B_1 | S_{0n} = x_0) = P(S_n \in x - B)/P(S_{0n} = x_0) \tag{2.6}$$

for $|x_0| < Cn^{1/2}$. Under the conditions of Corollary 2.1, both elements in this ratio have expansions of the form given in (2.1), so an approximation is given by

$$\int_{x_1 - B_1} e_{s-3}(u, \mu_n)/e_{s-3}(x_0, \mu_{0n})\, du_1,$$

where μ_{0n} is the probability measure of S_{0n}. If we replace $V_n^{-1/2}$ by Δ, where Δ is a lower triangular matrix such that $\Delta^T\Delta = V_n$ and $u^* = \Delta u$, then u_0^* depends only on u_0 while u^*_1 depends on both u_1 and u_0. So the integrand is approximated by

$$(2\pi)^{-d_1/2}(\det V_n^{11})^{1/2} e^{-|u_1^*|^2/2}\left(1 + \sum_{l=1}^{s-3} Q_{ln}(u^*)\right)\Bigg/\left(1 + \sum_{l=1}^{s-3} Q_{ln}^{(0)}(x_0^*)\right),$$

where $u_0^* = x_0^*$ and $Q_{ln}^{(0)}(x_0^*)$ is defined as in (1.11) by restricting attention to S_{0n} and V_n^{11} is the $d_1 \times d_1$ lower right-hand submatrix of V_n^{-1}.

2.2. *Saddlepoint approximations.* The other choice of θ of most interest is $\hat{\theta}_n = \hat{\theta}_n(x)$, the value of θ at which

$$\kappa_n(\theta) - \theta \cdot x \tag{2.7}$$

is minimized; that is the solution of

$$m_n(\hat{\theta}_n) = \kappa_n'(\hat{\theta}_n) = x. \tag{2.8}$$

This corresponds to a saddlepoint in the complex cumulant generating function and permits us to obtain approximation theorems with relative error $O(n^{-(s-2)/2})$ throughout an appropriate range of x values in many cases.

First we will give some general conditions.

(S.1) There exists a convex, compact $K \subset \operatorname{int}(\Theta_n)$ for all n, such that $\operatorname{int}(K)$ is nonempty.

(S.2) There exist positive-definite matrices $V(\theta)$ such that as $n \to \infty$,

$$n^{-1}V_n(\theta) \to V(\theta), \quad \text{uniformly for } \theta \in K. \tag{2.9}$$

(S.3) $\eta_s(\theta) < C$ for all $\theta \in K$.

(S.4) $q_{n\theta}(n^{(s-3)/2}) < Cn^{-\lambda}$ for all $\theta \in K$, or if $d_1 = 0$, $q_{n\theta} < Cn^{-\lambda}$, where $\lambda = d_1(s-3)/2 + (s-2)/2 + d/2$.

REMARK 5. If $X_1, X_2, \dots$ are iid with probability measure μ which is not concentrated on an affine subspace of $\mathbb{R}^d$, then $n^{-1}\kappa_n(\theta) = \kappa(\theta) = \log Ee^{\theta \cdot X_1}$, which is strictly convex on $\operatorname{int}(\Theta)$, where here $\Theta = \Theta_n$. The conditions above are all satisfied if

$$|E\exp((\theta + it)\cdot X_1)/E\exp(\theta\cdot X_1)| < 1-\delta, \tag{2.10}$$

for $\theta \in \operatorname{int}(\Theta)$, some $\delta > 0$, and $c < |t| < Cn^{(s-3)/2}$.

Under conditions (S.1)–(S.4), for $n^{-1}x \in K_n^* = \{n^{-1}\kappa_n'(\theta)\colon \theta \in K\}$, (2.8) has a unique solution, $\hat\theta_n$, say, since (S.2) ensures that $n^{-1}\kappa_n(\theta)$ is strictly convex for all $\theta \in K$ and n large enough, so $n^{-1}\kappa_n'(\theta)$ is a one-to-one map from K to K_n^*. Now we can obtain the following result.

COROLLARY 2.3. *If conditions* (S.1)–(S.4) *hold, if B is convex and included in a compact neighborhood of* 0 *and if $n^{-1}x \in K_n^*$, then*

$$\begin{aligned}\mu_n(x-B) &= e^{\kappa_n(\hat\theta_n)-\hat\theta_n\cdot x}\big[e_{s-3}(\hat\theta_n, B, 0) \\ &\qquad + (\operatorname{sur}(B_1)n^{1/2} + \operatorname{vol}(B_1))O(n^{-d/2-(s-2)/2})\big].\end{aligned}\tag{2.11}$$

If, in addition, $\sup\{|y|\colon y\in B\} < Cn^{-(s-3)/2}$, *then*

$$\begin{aligned}\mu_n(x-B) &= \frac{e^{\kappa_n(\hat\theta_n)-\hat\theta_n\cdot x}}{(2\pi)^{d/2}[\det \hat V_n]^{1/2}}\Bigg[\operatorname{vol}(B_1)\Bigg(1+\sum_{l=1}^{s-3}\hat Q_{ln}(0)\Bigg) \\ &\qquad + (\operatorname{sur}(B_1)n^{1/2} + \operatorname{vol}(B_1))O(n^{-(s-2)/2})\Bigg],\end{aligned}\tag{2.12}$$

where $\hat Q_{ln}(u)$ are defined as in (1.11) *with the cumulants calculated at $\hat\theta_n$ and we write $\hat V_n$ for $V_n(\hat\theta_n)$.*

REMARK 6. If in addition to the conditions of Remark 5, X_1 has an integrable characteristic function, then, as in Remark 1, we can see that

$$f_n(x) = e_{s-3}(x, \nu_{n\hat\theta})(1 + O(n^{-(s-2)/2})). \tag{2.13}$$

For B convex but not necessarily compact, we can use the method of Skovgaard (1986) to obtain a result as in (2.4) where the term $(\operatorname{sur}(B_1)n^{1/2} + \operatorname{vol}(B_1))$ is replaced by 1 and the order term is $O((\log n)n^{-d_0/2-(s-2)/2})$. However, in the important special case when $d_1 = 1$ and $B = (-\infty, 0]$, this can be improved.

COROLLARY 2.4. *If* $d_1 = 1$, $B_1 = (-\infty, 0]$, $n^{-1}x \in K_n^*$ *and* $x_1 > ES_{1n}$, *and if conditions* (S.1)–(S.4) *hold, then*

$$\mu_n(\{0\} \times [x, \infty)) = e^{\kappa_n(\hat{\theta}_n) - \hat{\theta}_n \cdot x}\Big[e_{s-3}(\hat{\theta}_n, B, 0) + (\hat{\theta}_{1n}^{-1} + n^{1/2})O(n^{-d/2-(s-2)/2})\Big]. \tag{2.14}$$

PROOF. Take $\varepsilon = cn^{-(s-3)/2}$. Since $x_1 > ES_{1n}$, $\hat{\theta}_{1n} > 0$, so

$$\hat{\chi}_{\hat{\theta}_n, B_{2\varepsilon}}(0) = \int_{-2\varepsilon}^{\infty} e^{-\hat{\theta}_{1n}u_1}\,du_1 = \hat{\theta}_{1n}^{-1}e^{2\varepsilon\hat{\theta}_{1n}} \le C\hat{\theta}_{1n}^{-1}$$

and similarly $\hat{\chi}_{\hat{\theta}_n, B_{2\varepsilon} \setminus B_{-2\varepsilon}}(0) < C\varepsilon$. So the result follows immediately from Theorem 1. □

REMARK 7. Consider approximations for $P(S_{1n} \in x_1 - B_1 | S_{0n} = x_0)$, using approximations for the numerator and denominator of $P(S_n \in x - B)/P(S_0 = x_0)$ obtained from Corollary 2.3. If $\kappa'_{0n}(\theta)$ denotes the vector of partial derivatives of $\kappa_n(\theta)$ with respect to θ_0 only, then under the conditions (S.1)–(S.4),we can choose $\hat{\theta}_n$ as in (2.8) and let $\tilde{\theta}_{0n}$ be the unique solution of

$$\kappa'_{0n}(\theta) = x_0. \tag{2.15}$$

Then if $\tilde{\theta}_n = (0, \tilde{\theta}_{0n})$, the approximation is

$$\frac{e^{\kappa_n(\hat{\theta}_n) - \hat{\theta}_n \cdot x - \kappa_n(\tilde{\theta}_n) + \tilde{\theta}_{0n} \cdot x_0} e_{s-3}(\hat{\theta}_n, B, 0)}{(2\pi)^{-d_0/2}\big[\det \tilde{V}_{00n}\big]^{-1/2}\big(1 + \Sigma_{l=1}^{s-3}\tilde{Q}_{0ln}(0)\big)},$$

where $\tilde{Q}_{0ln}(0)$ is defined as in (1.11) on $\mathbb{R}^{d_0}$ with cumulants calculated at $\tilde{\theta}_{0n}$ and $\tilde{V}_{00n}$ is the upper left-hand $d_0 \times d_0$ submatrix of $V_n(\tilde{\theta}_n)$.

3. One- and two-dimensional examples.

3.1. *One-dimensional lattice approximations.* Consider the case $d_0 = 1$, $d_1 = 0$, $B = \{0\}$, $\varepsilon = 0$, and suppose $X, X_1, X_2, \dots$ are iid. As in Remark 5, (S.1)–(S.3) are satisfied. The requirement that X has span 1 ensures $|q_{n\theta}| < e^{-cn}$ for $\theta \in K$, $n \ge 1$, so that (S.4) is also satisfied. For a local result, if x is such that $\hat{\theta}_n \in K$, where $\hat{\theta}_n$ solves $m_n(\hat{\theta}_n) = x$, taking $s = 6$, and bearing in mind (1.10), one obtains, writing $\hat{\theta}$ for $\hat{\theta}_n$ and $\hat{V}_n$ for $V_n(\hat{\theta})$,

$$\mu_n(\{x\}) = \frac{e^{\kappa_n(\hat{\theta}) - \hat{\theta}m_n(\hat{\theta})}}{(2\pi\hat{V}_n)^{1/2}}\big[1 + \hat{Q}_{2n}(0) + O(n^{-2})\big], \tag{3.1}$$

where $\hat{Q}_{2n}(0) = (3\kappa_{4n}^* - 5\kappa_{3n}^{*2})/24$, a result originally obtained by Blackwell and Hodges (1959).

For a tail result, take $m_n(\hat{\theta}) = x$, $w = y - x$ to get

$$\mu_n([x,\infty)) = \sum_{y=x}^{\infty} \mu_n(\{y\})$$

$$(3.2)\qquad = \sum_{w=0}^{\infty} e^{\kappa_n(\hat{\theta}) - \hat{\theta}(w+x)} \hat{V}_n^{-1/2} \varphi\big(w\hat{V}_n^{-1/2}\big)$$

$$\times\big[1 + \hat{Q}_{1n}\big(w\hat{V}_n^{-1/2}\big) + \hat{Q}_{2n}\big(w\hat{V}_n^{-1/2}\big) + O(n^{-3/2})\big].$$

Now for $x > 0$, $e^{-x} = 1 - x + O(x^2)$ as $x \to 0$, so for $k \ge 0$,

$$(3.3)\quad \sum_{w=0}^{\infty} w^k e^{-\hat{\theta}w - w^2/2\hat{V}_n} = \sum_{w=0}^{\infty} w^k e^{-\hat{\theta}w} - \frac{1}{2\hat{V}_n} \sum_{w=0}^{\infty} w^{k+2} e^{-\hat{\theta}w} + O(n^{-2}),$$

and using this in (3.2) gives

$$\mu_n([x,\infty)) = \frac{e^{\kappa_n(\hat{\theta}) - \hat{\theta}x}}{\big(2\pi\hat{V}_n\big)^{1/2}(1 - e^{-\hat{\theta}})}$$

$$(3.4)\qquad \times\big[1 + \tfrac{1}{8}\big(\kappa_{4n}^* - 5\kappa_{3n}^{*2}/3\big) - \tfrac{1}{2}\pi_2(\hat{\theta})/\hat{V}_n$$

$$-\kappa_{3n}^*\pi_1(\hat{\theta})/2\hat{V}_n^{1/2} + O(n^{-2})\big],$$

where

$$(3.5)\quad \pi_1(\theta) = e^{-\theta}/(1 - e^{-\theta}), \qquad \pi_2(\theta) = e^{-\theta}(1 + e^{-\theta})/(1 - e^{-\theta})^2,$$

again as obtained by Blackwell and Hodges (1959).

In practice, the approximation (3.3) [and hence (3.4)] is not very good for moderate values of n (e.g., 20 or 50), whereas the local result (3.1) gives a good approximation. However, good approximations can be obtained for tail probabilities either by adding up (3.1) [which involves solving a lot of equations $m_n(\hat{\theta}) = x$] or by calculating from (3.2). This point is discussed further in Section 4.

3.2. *The one-dimensional integral result.* For ease of notation, we will restrict attention in this section to the case of iid random variables, so $d_1 = 1$, $d_0 = 0$ and we will consider $B = (-\infty, 0]$. The results in the direct Edgeworth case when $\theta = 0$ are well known so we will consider only the saddlepoint approximation.

The assumptions which are required to ensure that conditions (S.1)–(S.4) hold are discussed in Remark 5 of Section 2.2. These are that $\kappa(\theta) = \log Ee^{\theta X_1} < \infty$ in an open neighborhood of 0 and that (2.10) (with the restriction to $d = 1$) holds. For the rest of this section we will assume that (2.10) holds; let $X_{1\theta}, X_{2\theta}, \dots$ be the associated random variables and write

$$(3.6)\qquad EX_{1\theta} = m(\theta) = \kappa'(\theta), \qquad \operatorname{var}(X_{1\theta}) = V_n(\theta)/n = \kappa''(\theta).$$

From Corollary 2.4 we obtain, writing $\hat{\theta}$ for $\hat{\theta}_n$,

$$(3.7)\quad \mu_n(x-B) = e^{\kappa_n(\hat{\theta})-\hat{\theta}x}\big[e_{s-3}(\hat{\theta},B,0) + (n^{1/2}+\hat{\theta}^{-1})O(n^{-(s-1)/2})\big].$$

Here we have

$$(3.8)\quad \begin{aligned} e_{s-3}(\hat{\theta},B,0) &= (2\pi\hat{V}_n)^{-1/2}\int_0^{\infty} e^{-\hat{\theta}u-u^2/2\hat{V}_n}\left(1+\sum_{l=1}^{s-3}\hat{Q}_{ln}(u/\hat{V}_n^{1/2})\right)du \\ &= (2\pi)^{-1/2}e^{\hat{\theta}^2\hat{V}_n/2}\int_{\hat{\theta}\hat{V}_n^{1/2}}^{\infty} e^{-v^2/2}\left(1+\sum_{l=1}^{s-3}\hat{Q}_{ln}(v-\hat{\theta}\hat{V}_n^{1/2})\right)dv \\ &= \tau(\hat{\theta}\hat{V}_n^{1/2})\left(1+\sum_{l=1}^{s-3}\hat{P}_{ln}(\hat{\theta}\hat{V}_n^{1/2})\right), \end{aligned}$$

where

$$(3.9)\quad \tau(\lambda) = (2\pi)^{-1/2}e^{\lambda^2/2}\int_{\lambda}^{\infty}e^{-v^2/2}\,dv$$

and

$$(3.10)\quad \tau(\lambda)\hat{P}_{ln}(\lambda) = (2\pi)^{-1/2}e^{\lambda^2/2}\int_{\lambda}^{\infty}e^{-v^2/2}\hat{Q}_{ln}(v-\lambda)\,dv.$$

Here $\hat{Q}_{ln}(u)$ is a linear combination of Hermite–Chebyshev polynomials $H_k(u)$ with coefficients depending on the standardized cumulants of $S_{n\hat{\theta}}$. We will show in Appendix B that

$$(3.11)\quad \begin{aligned} \tau(\lambda)\rho_k(\lambda) &= (2\pi)^{-1/2}e^{\lambda^2/2}\int_{\lambda}^{\infty}H_k(v-\lambda)e^{-v^2/2}\,d\lambda \\ &\quad +(-\lambda)^k\tau(\lambda) + (2\pi)^{-1/2}\sum_{j=0}^{k-1}H_j(0)(-\lambda)^{k-j-1}. \end{aligned}$$

So $\hat{P}_{ln}(\lambda)$ is the same linear combination of $\rho_k(\lambda)$ as $\hat{Q}_{ln}(u)$ is of $H_k(u)$. In particular,

$$(3.12)\quad \hat{P}_{1n}(\lambda) = \kappa_{3n}^*\rho_3(\lambda)/6,$$

$$(3.13)\quad \hat{P}_{2n}(\lambda) = \kappa_{4n}^*\rho_4(\lambda)/24 + \kappa_{3n}^{*2}\rho_6(\lambda)/72,$$

where $\kappa_{3n}^* = n\kappa^{(3)}(\hat{\theta})/\hat{V}_n^{3/2}$ and $\kappa_{4n}^* = n\kappa^{(4)}(\hat{\theta})/\hat{V}_n^2$.

3.3. *Conditional lattice approximations.* Take $d_1 = 0$, $d_0 = 2$, $B = \{0\}$, $\varepsilon = 0$, $X_1, X_2, \dots$ iid with $EX_{01} = 0$. Note that the natural notation $X_j =$

(X_{0j}, X_{1j}) used here differs from that used elsewhere. As in Section 3.1, $|q_{n\theta}| < e^{-cn}$ for $\theta \in K$, $n \geq 1$. If $\hat{\theta} = \hat{\theta}_n = m_n^{-1}(x) \in K$, the theorem gives, with $s = 6$,

$$\mu_n(\{x\}) = \frac{e^{\kappa_n(\hat{\theta}) - \hat{\theta}\cdot m_n(\hat{\theta})}}{2\pi\left(\det \hat{V}_n\right)^{1/2}} \times\left[1 + \hat{Q}_{1n}(w^*) + \hat{Q}_{2n}(w^*) + \hat{Q}_{3n}(w^*) + O(n^{-2})\right], \tag{3.14}$$

where w^* is the standardized variable $w^* = \hat{V}_n^{-1/2} w$ and $w = x - m_n(\hat{\theta}_n)$; in particular taking $w = 0$ gives a local bivariate result. Choosing $x = EX_{11} = 0$ in (3.2) gives

$$\mu_{0n}(\{0\}) = \frac{1}{\left(2\pi\tilde{V}_{00n}\right)^{1/2}}\left[1 + \tilde{Q}_{02n}(0) + O(n^{-2})\right], \tag{3.15}$$

where $\tilde{V}_{00n} = V_{00n}(0)$ and $\tilde{Q}_{02n}(0) = (3\tilde{\kappa}_4 - 5\tilde{\kappa}_3^2)/24$, the $\tilde{\kappa}$'s being the standardized cumulants of $X_{01} + \cdots + X_{0n}$.

Thus, with $\hat{\theta} = m_n^{-1}(0, x_1)$,

$$P\{S_{1n} = x_1 | S_{0n} = 0\} = \frac{e^{\kappa_n(\hat{\theta}) - \hat{\theta}\cdot m_n(\hat{\theta})}}{\left(2\pi \det \hat{V}_n/\tilde{V}_{00n}\right)^{1/2}} \times\left[1 + \hat{Q}_{2n}(0) - \tilde{Q}_{02n}(0) + O(n^{-2})\right]. \tag{3.16}$$

In the same way, with the same $\hat{\theta}$, incorporating $\hat{Q}_{3n}(v^*)$ in the order term,

$$P(S_{1n} \geq x_1 | S_{0n} = 0) = \sum_{v_1=0}^{\infty} \frac{e^{\kappa_n(\hat{\theta}_n) - \hat{\theta}_{1n}(v_1 + x_1) - \frac{1}{2}v_1^{*2}}}{\left(2\pi \det \hat{V}_n/\tilde{V}_{00n}\right)^{1/2}}, \times\left[1 + \hat{Q}_{1n}(v^*) + \hat{Q}_{2n}(v^*) - \tilde{Q}_{02n}(0) + O(n^{-3/2})\right], \tag{3.17}$$

where $v^* = (0, v_1/\sqrt{\hat{V}_n^{11}})$, and $\hat{V}_n^{11}$ is the lower right-hand element of $\hat{V}_n^{-1}$, giving

$$P(S_{1n} \geq x_1 | S_{0n} = 0) = \frac{e^{\kappa_n(\hat{\theta}) - \hat{\theta}_{1n}x_1}}{\left(2\pi \det \hat{V}_n/V_{00n}\right)^{1/2}(1 - e^{-\hat{\theta}_{1n}})} \times\Big[1 - \tfrac{1}{2}\pi_2(\hat{\theta}_1)/\hat{V}_n^{11} + \hat{Q}_{1n}(0)\pi_1(\hat{\theta}_1)/\sqrt{\hat{V}_n^{11}} + \hat{Q}_{2n}(0) - \tilde{Q}_{02n}(0) + O(n^{-3/2})\Big]. \tag{3.18}$$

The approximations implicit in (3.17) and (3.18) are subject to the same difficulties described at the end of Section 3.1; this point is amplified in Section 4.3.

3.4. *The one-dimensional integral conditional result.* Suppose $X_1, X_2, \ldots$ are iid random vectors in $\mathbb{R}^2$ with $d_0 = 1$, $d_1 = 1$, $B_1 = (-\infty, 0]$ and $x_1 > ES_{1n}$.

For the case $\theta = 0$, from Remarks 1 and 2 of Section 2.1, we obtain the following result:

$$P(S_{1n} \geq x_1 | S_{0n} = x_0) = \int_{x_1}^{\infty} e_{s-3}(u, \mu_n)/e_{s-3}(x_0, \mu_{0n})\, du_1 + O(n^{-(s-2)/2}). \tag{3.19}$$

The integral in this case is easily calculated as an expansion in Hermite–Chebyshev polynomials.

The assumptions required to ensure that (S.1)–(S.4) hold are discussed in Remark 5 of Section 2.2. Under these conditions, we have, writing $\hat{\theta}$ for $\hat{\theta}_n$, from Corollary 2.4,

$$\begin{aligned} P(S_{1n} \geq x_1 | S_{0n} = x_0) = {} & \frac{e^{\kappa_n(\hat{\theta}) - \hat{\theta}_0 x_0 - \hat{\theta}_1 x_1}}{(2\pi)^{-1/2} \tilde{V}_{00n}^{-1/2}} \\ & \times \left[e_{s-3}(\hat{\theta}, B, 0) \Big/ \left(1 + \sum_{l=1}^{s-3} \tilde{Q}_{0ln}(0)\right) \right. \\ & \left. + (n^{1/2} + |\hat{\theta}_1|^{-1}) O(n^{-s/2}) \right]. \end{aligned} \tag{3.20}$$

Here

$$\begin{aligned} e_{s-3}(\hat{\theta}, B, 0) = {} & \left(2\pi \det \hat{V}_n^{1/2}\right)^{-1} \\ & \times \int_0^{\infty} e^{-\hat{\theta} u_1 - u_1^2 \hat{V}_n^{11}/2} \left(1 + \sum_{l=1}^{s-3} \hat{Q}_{ln}(0, u_1)\right) du_1 \\ = {} & (2\pi)^{-1/2} \hat{V}_{00n}^{-1/2} \tau\left(\hat{\theta}\sqrt{\hat{V}_n^{11}}\right) \left(1 + \sum_{l=1}^{s-3} \hat{P}_{ln}\left(\hat{\theta}\sqrt{\hat{V}_n^{11}}\right)\right), \end{aligned} \tag{3.21}$$

where $\tau(\lambda)$ is defined in Section 3.2 and here

$$\tau(\lambda) \hat{P}_{ln}(\lambda) = (2\pi)^{-1/2} e^{\lambda^2/2} \int_{\lambda}^{\infty} e^{-v^2/2} \hat{Q}_{ln}(0, v - \lambda)\, dv. \tag{3.22}$$

It is worthwhile noting that, for example,

$$\begin{aligned} \hat{Q}_{1n}(u_1, 0) &= \frac{1}{3!} \sum_{i=0}^{3} \binom{3}{i} \kappa^*_{(3-i,i)n} H_{3-i}(u_1) H_i(0) \\ &= \frac{1}{6} (\kappa^*_{03n} H_3(u_1) - 3\kappa^*_{21n} H_1(u_1)) \end{aligned} \tag{3.23}$$

so

$$\hat{P}_{1n}(\lambda) = \tfrac{1}{6}\Big(\kappa^*_{30n}\rho_3\Big(\hat{\theta}\sqrt{\hat{V}_n^{11}}\Big) - 3\kappa^*_{21n}\rho_1\Big(\hat{\theta}\sqrt{\hat{V}_n^{11}}\Big)\Big). \tag{3.24}$$

These formulas are obtained from (1.10) and (1.11) and are used in Section 4 to obtain numerical results. The details of their use will be deferred to that section.

3.5. *Linear combinations.* Consider $Z, Z_1, Z_2, \ldots$ iid integer-valued random variables with span 1 having

$$\kappa(u) = \log E(e^{uZ}), \qquad \kappa'(0) = E(Z) = \alpha, \qquad \kappa''(0) = \operatorname{Var}(Z) > 0. \tag{3.25}$$

Let $a_1, \ldots, a_n$ be real numbers with

$$a_1 + \cdots + a_n = 0, \qquad a_1^2 + \cdots + a_n^2 = n. \tag{3.26}$$

Set $X_k = (Z_k, a_k Z_k)$ and $S_n = X_1 + \cdots + X_n$ with

$$\kappa_n(\theta) = \log E(e^{\theta \cdot S_n}) = \sum_{k=1}^n \kappa(\theta_0 + \theta_1 a_k) \tag{3.27}$$

finite for $\theta \in \Theta_n$.

In this section approximations for $P(S_{0n} = r,\ S_{1n} \geq t)$ are given when $r = \alpha n$, $t/\sqrt{n} > c > 0$. Note that $E(S_{0n}) = r$, $E(S_{1n}) = 0$. First some lemmas will be proved.

Consider associated random variables $Z_{1\theta}, \ldots, Z_{n\theta}$ with

$$\log E(e^{uZ_{k\theta}}) = \kappa(u + \theta_0 + \theta_1 a_k) - \kappa(\theta_0 + \theta_1 a_k),$$

and the corresponding $X_{1\theta}, \ldots, X_{n\theta}$ and $S_{n\theta}$. We have

$$m_n(\theta) = E(S_{n\theta}) = \sum_{k=1}^n \kappa'(\theta_0 + \theta_1 a_k)\begin{pmatrix} 1 \\ a_k \end{pmatrix}, \tag{3.28}$$

$$V_n(\theta) = \operatorname{Cov}(S_{n\theta}) = \sum_{k=1}^n \kappa''(\theta_0 + \theta_1 a_k)\begin{pmatrix} 1 & a_k \\ a_k & a_k^2 \end{pmatrix}. \tag{3.29}$$

In particular, for $\theta = 0$,

$$E(S_n) = \begin{pmatrix} r \\ 0 \end{pmatrix}, \qquad \operatorname{Cov}(S_n) = n\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\operatorname{Var}(Z). \tag{3.30}$$

Introduce the distribution function $F_n(x) = \#\{k\colon a_k \leq x\}/n$ having mean 0 and variance 1. We will let $n \to \infty$, so we are in fact considering double arrays of a's, but for notational convenience this will not be indicated by an extra index. Note that the distribution of the Z's does not depend on n. We assume

$$F_n \to F \quad \text{as } n \to \infty, \qquad \int dF = 1, \qquad \int x^2\, dF = 1. \tag{A.1}$$

In most applications this is sufficiently general; the "a-scores" are given by some distribution F with mean 0 and variance 1. The following regularity condition is also assumed.

(A.2) There exists a convex, compact $K \subset \operatorname{int}(\Theta_n)$ for all n, with $0 \in \operatorname{int}(K)$ and

(a)
$$h_n(\theta) = \frac{1}{n}\kappa_n(\theta) = \int \kappa(\theta_0 + \theta_1 x)\, dF_n \to h(\theta)$$
$$= \int \kappa(\theta_0 + \theta_1 x)\, dF,$$

(b)
$$h'_n(\theta) \to h'(\theta) = \int \begin{pmatrix} 1 \\ x \end{pmatrix} \kappa'(\theta_0 + \theta_1 x)\, dF,$$

(c)
$$h''_n(\theta) \to h''(\theta) = \int \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix} \kappa''(\theta_0 + \theta_1 x)\, dF,$$

as $n \to \infty$, for all $\theta \in K$.

Note that $h'_n(\theta) = m_n(\theta)/n$ and $h''_n(\theta) = V_n(\theta)/n$.

LEMMA 3.1. *If* (A.1) *and* (A.2) *hold, then*

(i) h_n, h *are strictly convex in* K,
(ii) $K_n^* = h'_n(K)$, $K^* = h'(K)$ *are compact sets,*
(iii) $h'_n(0) = (r/n, 0) \to h'(0) \in \operatorname{int}(K_n^* \cap K^*)$,
(iv) h'_n, h' *are one-to-one,*
(v) $\det((1/n)V_n(\theta)) \to \det h''(\theta) > 0$, $n \to \infty$, $\theta \in K$.

PROOF. (i) Since $\operatorname{Var}(Z) > 0$ we have $\kappa''(\theta) > 0$. Hence, as F is nondegenerate, it follows that there is strict inequality in the Cauchy–Schwarz inequality:

$$\left(\int x\kappa''(\theta_0 + \theta_1 x)\, dF\right)^2 \le \int x^2 \kappa''(\theta_0 + \theta_1 x)\, dF \int \kappa''(\theta_0 + \theta_1 x)\, dF.$$

Thus $h''(\theta)$ is a positive-definite matrix when $\theta \in K$, and h is strictly convex. Strict convexity of h_n is proved analogously. (ii)–(iv) follow from the continuity of h'_n, h', the strict convexity of h_n, h and as $0 \in \operatorname{int}(K)$. In Lemma 3.1(v) just note that h_n, h are strictly convex and

$$h''_n(\theta) = \frac{1}{n} V_n(\theta) \to h''(\theta). \qquad \square$$

LEMMA 3.2. *If*

(A.3)
$$\left| n^{-1} \sum_{k=1}^{n} a_k^j \prod{}^* \kappa^{(l_i)}(\theta_0 + \theta_1 a_k) \right| < C,$$

for all $\theta \in K$, $j = 0, 2, 4, 6$, *where* $\prod^*$ *denotes a product over a set of values of* l_i *satisfying* l_i *even,* $l_i \ge 2$, $\sum_i l_i = 6$, *and if* (A.1) *and* (A.2) *hold, then* $\eta_6(\theta) \le C$, *for* $\theta \in K$.

PROOF. For n large enough we need simply notice that using Lemma 3.1(v) and the functional relation of moments and cumulants we can find $\eta_6(\theta)$ by a linear combination of the quantities on the left in the condition (A.3). □

REMARK. Similarly any $\eta_s(\theta)$, for even s, can be bounded. For odd s note that $\eta_s(\theta)$ is bounded when $\eta_{s+1}(\theta)$ is bounded.

LEMMA 3.3. *If* (A.1)–(A.3) *hold, then*

$$q_{n\theta}(T) \le \sup\{|\hat{\nu}_{n\theta}(\xi)|: \xi_0^2 + \xi_1^2 > C, |\xi_0| < \pi, |\xi_1| < T\}, \tag{3.31}$$

for some $C > 0$ *and all* $\theta \in K$.

PROOF. $\eta_3(\theta)$ is bounded according to Lemma 3.2 and the remark above. Hence by Lemma 3.1(v) it follows that there exists a $C > 0$ such that

$$\left\{\xi: \left|V_n(\theta)^{1/2}\xi\right| > \tfrac{3}{4}\eta_3(\theta)^{-1}n^{1/2}\right\} \subset \{\xi: \xi_0^2 + \xi_1^2 > C\},$$

which proves the assertion. □

For $M > 0$ and $\zeta > 0$ consider the set of a_k with $|a_k| < M$. Consider a condition due to Albers, Bickel and van Zwet (1976). Let $\gamma(M, \zeta)$ denote the Lebesgue measure λ of the ζ-neighborhood of this set, that is,

$$\gamma(M, \zeta) = \lambda\{x: \exists_k |x - a_k| < \zeta, |a_k| < M\}.$$

LEMMA 3.4. *If*

(A.4) *positive numbers* C, δ, M *exist with*

$$n^{-1}\sum_{k=1}^{n} a_k^4 \le C, \qquad \gamma(M, \zeta) \ge \delta n \zeta, \quad \textit{for some } \zeta \ge n^{-3/2}\log n,$$

then there exist positive numbers b, B, c^* *depending on* n *and* $a_1, \dots, a_n$ *only through* c, C, δ, M *such that for all* $\theta \in K$,

$$\sup\{|\hat{\nu}_{n\theta}(\xi)|: |\xi_0| < \pi, |\xi_1| \le bn, \xi_0^2 + \xi_1^2 > c\} \le Bn^{-c^* \log n}.$$

PROOF. Consider the characteristic function

$$\hat{\mu}_{n\theta}(\xi) = E(e^{i\xi\cdot S_{n\theta}}) = \exp\left(\sum_{k=1}^{n}\left(\kappa(\theta_0 + \theta_1 a_k + i(\xi_0 + \xi_1 a_k)) - \kappa(\theta_0 + \theta_1 a_k)\right)\right).$$

Hence

$$\begin{aligned} |\hat{\nu}_{n\theta}(\xi)| = |\hat{\mu}_{n\theta}(\xi)| &= \prod_{k=1}^{n} |E(\exp((\theta_0 + \theta_1 a_k + i(\xi_0 + \xi_1 a_k))Z - \kappa(\theta_0 + \theta_1 a_k)))| \\ &= \prod_{k=1}^{n}\left|\sum_j p_{j\theta k}\exp(i(\xi_0 + \xi_1 a_k)j)\right|, \end{aligned}$$

where

$$p_{j\theta k} = \exp((\theta_0 + \theta_1 a_k)j - \kappa(\theta_0 + \theta_1 a_k))P(Z = j).$$

With the symmetrized distribution

$$p^s_{j\theta k} = \sum_l p_{l\theta k} p_{j+l,\theta k},$$

we get

$$|\hat{\nu}_{n\theta}(\xi)|^2 = \prod_{k=1}^n \left(\sum_j p^s_{j\theta k} \cos((\xi_0 + \xi_1 a_k)j) \right).$$

Since the Z's have span 1, it follows that there exists an integer $r \geq 1$ such that the n-fold convolution of $p^s_{\cdot\theta k}$ has positive mass on 1 for $n \geq r$. We may assume without loss of generality that $r = 1$. If $|a_k| < M$ there exists $0 < \varepsilon' < \frac{1}{2}$ with $\varepsilon' < p^s_{1\theta k} < 1 - \varepsilon'$ for all n. For all $\theta \in K$ we have

$$c \leq \frac{1}{n} \operatorname{var}\left(\sum_{k=1}^n a_k z_{k\theta} \right) = \frac{1}{n} \sum_{k=1}^n a_k^2 \kappa''(\theta_0 + \theta_1 a_k) \leq C$$

according to Lemma 3.1(v). Using this, a slight modification of the proof on page 114 in Albers, Bickel and van Zwet (1976), gives

$$|\hat{\nu}_{n\theta}(\xi)| \leq \prod_{k=1}^n (1 + 2p^s_{1\theta k}(\cos(\xi_0 + \xi_1 a_k) - 1)) \leq n^{-c^* \log n},$$

for all $\theta \in K$, $|\xi_0| < \pi$, $|\xi_1| < bn$ and $\xi_0^2 + \xi_1^2 > c > 0$. □

In order to approximate $P(S_{0n} = r,\ S_{1n} \geq t)$ we will use Theorem 1 with $\theta = \hat{\theta}_n$ given by the saddlepoint, that is the solution of

$$E(S_{n\theta}) = m_n(\theta) = (r, t),$$

provided there is a solution.

LEMMA 3.5. *If* (A.1) *and* (A.2) *hold and* $(r/n, t/n) \in K_n^*$, *then*

(i) *there exists a unique* $\hat{\theta}_n \in K$ *with* $m_n(\hat{\theta}_n) = (r, t)$,

(ii) *for* $t/n = \beta > 0$, $c \leq \hat{\theta}_{1n} \leq C$,

(iii) *for* $t/n^{1/2} = \gamma > 0$, $c \leq n^{1/2}\hat{\theta}_{1n} \leq C$.

REMARK. As $0 \in \operatorname{int}(K)$ and $m_n(0) = (r, 0)$, there exists a solution $\hat{\theta}_n \in K$ for t sufficiently small.

PROOF. (i) follows immediately from the definition of K_n^*. As

$$E(S_{1n\theta}) = \sum_{k=1}^n a_k \kappa'(\theta_0) = 0 \quad \text{for } \theta = (\theta_0, 0),$$

$$\frac{\partial}{\partial\theta} E(S_{1n\theta}) = \sum_{k=1}^n a_k^2 \kappa''(\theta_0 + \theta_1 a_k) > 0,$$

we have $E(S_{1n\theta}) > 0$ if and only if $\theta_1 > 0$. From this and (A.2), (ii) and (iii) easily follow. □

The preceding lemmas show that (A.1)–(A.4) imply (S.1)–(S.4). Thus the following result is a consequence of Corollary 2.4.

PROPOSITION 3.6. *If* (A.1)–(A.4) *hold*, $t/n^{1/2} > c > 0$, *and* $(r/n, t/n) \in K_n^*$, *then*

$$\left|P(S_{0n} = r,\ S_{1n} \geq t) - e^{\kappa_n(\hat{\theta}_n) - \hat{\theta}_{0n} r - \hat{\theta}_{1n} t} e_2\left(\hat{\theta}_n, \{0\} \times (-\infty, 0], 0\right)\right|$$
$$\leq C n^{-3/2} e^{\kappa_n(\hat{\theta}_n) - \hat{\theta}_{0n} r - \hat{\theta}_{1n} t}.$$

4. Applications and numerical methods.

4.1. *Examples.* In this section some specific examples will be discussed. In the first three subsections we consider linear combinations where the probabilities of interest are of the form

$$P\left(\sum_{k=1}^n a_k Z_k \geq t \,\middle|\, \sum_{k=1}^n Z_k = r\right) = P\left(\sum_{k=1}^n a_k Z_k \geq t,\ \sum_{k=1}^n Z_k = r\right) \Big/ P\left(\sum_{k=1}^n Z_k = r\right).$$

We can approximate this by approximating the numerator and denominator separately. In the fourth subsection we consider paired comparisons where conditioning is not required, then we consider an occupancy problem obtaining an approximation for a conditional probability and finally a vacancy problem where we condition with respect to a continuous variable.

4.1.1. *Sampling without replacement.* Let the Z's be iid Bernoulli distributed with mean $0 < \alpha = r/n < 1$, that is,

$$P(Z = 1) = 1 - P(Z = 0) = \alpha.$$

The conditional distribution of $S_{1n} = \sum_{k=1}^n a_k Z_k$ given $S_{0n} = \sum_{k=1}^n Z_k = r$ is the same as that of a simple random sample of size r drawn without replacement from the numbers $a_1, \dots, a_n$. We have

$$\kappa(u) = \log(\alpha e^u + 1 - \alpha).$$

Assuming that (A.1) holds, then it is easy to verify that (A.2) and (A.3) also hold for any compact set $K \subset \mathbb{R}^2$. For $0 < \alpha < 1$ and $t/n = \beta > 0$ not too large there exists a solution $\theta = \hat{\theta}_n$ to

$$E(S_{n\theta}) = m_n(\theta) = (r, t),$$

which in this case is

$$\sum_{k=1}^{n} \alpha e^{\theta_0+\theta_1 a_k}/(\alpha e^{\theta_0+\theta_1 a_k} + 1 - \alpha) = r,$$

$$\sum_{k=1}^{n} \alpha a_k e^{\theta_0+\theta_1 a_k}/(\alpha e^{\theta_0+\theta_1 a_k} + 1 - \alpha) = t.$$

Under assumption (A.4) Proposition 3.6 provides an approximation for $P(S_{0n} = r, S_{1n} \geq t)$. As S_{0n} is binomial with parameters n and r/n,

$$P(S_{0n} = r) = \binom{n}{r} r^r (n-r)^{n-r}/n^n,$$

which is easily calculated directly or by Stirling's formula.

Assumption (A.4) involves one moment condition, that is the boundedness of $n^{-1}\sum_{k=1}^{n} a_k^4$, and one condition on the "discreteness" of the a's. The Wilcoxon statistic corresponds to taking $a_k = (k - (n+1)/2)(12/(n^2-1))^{1/2}$, which satisfies (A.4), so Proposition 3.6 can be used. On the other hand for n even and $a_1 = \cdots = a_{n/2} = -1$, $a_{n/2+1} = \cdots = a_n = 1$, corresponding to the median test the "discreteness" condition is not satisfied. In this case Proposition 3.6 is not applicable but an approximation with $s = 3$ can be obtained or we could use a lattice approximation.

4.1.2. *Sampling with replacement.* Let the Z's be Poisson distributed with mean $\alpha = r/n > 0$. The conditional distribution of the Z's given their sum $S_{0n} = r$ is a multinomial distribution. Hence the conditional distribution of S_{1n} given $S_{0n} = r$ is the same as that of a simple random sample drawn with replacement from $a_1, \dots, a_n$. Here

$$\kappa(u) = \alpha(e^u - 1).$$

Assuming that

$$\int e^{ux}\, dF_n \to \int e^{ux}\, dF, \qquad n \to \infty,$$

for all u, then it is easy to verify that (A.1)–(A.3) hold for any compact set $K \subset \mathbb{R}^2$. For $t/n = \beta > 0$ not too large the equations for determining $\theta = \hat{\theta}_n$,

$$\alpha \sum_{k=1}^{n} e^{\theta_0+\theta_1 a_k} = r,$$

$$\alpha \sum_{k=1}^{n} a_k e^{\theta_0+\theta_1 a_k} = t$$

have a unique solution. Approximations for $P(S_{1n} \geq t | S_{0n} = r)$ can be obtained as in the case with sampling without replacement. S_{1n} is a statistic obtained from a bootstrap sample of $a_1, \dots, a_n$. This result could also be obtained by considering r iid samples from F_n.

4.1.3. *The geometric distribution.* Let the Z's be geometric, that is,

$$P(Z=j)=pq^{j}, \qquad j=0,1,2,\ldots,$$

with $q=1-p$, $\alpha=E(Z)=q/p$, $0<p<1$, and

$$\kappa(u)=\log p-\log(1-qe^{u}),$$

which is finite for $u<-\log q$. Hence $\kappa_n(\theta)$ is finite only in the sector

$$\{(\theta_0,\theta_1):(-\log q-\theta_0)/a_{1:n}<\theta_1<(-\log q-\theta_0)/a_{n:n}\},$$

where

$$a_{1:n}=\min_{1\le k\le n} a_k<0, \qquad a_{n:n}=\max_{1\le k\le n} a_k>0.$$

Assume that there exist constants $b_1<0<b_2$ such that for all n,

$$b_1<a_{1:n}<0<a_{n:n}<b_2.$$

This implies that the supports of F_n are contained in a bounded fixed interval. Assuming that $F_n\to F$ then it follows that (A.1)–(A.3) are all satisfied for any K contained in the sector

$$\{(\theta_0,\theta_1):(-\log q-\theta_0)/b_1<\theta_1<(-\log q-\theta_0)/b_2\}.$$

Approximations can be obtained as in the other examples provided $\beta=t/n>0$ is not too large.

We note that the conditional distribution of the Z's given their sum $S_{0n}=r$ can be written

$$P(Z_1=j_1,\ldots,Z_n=j_n|S_{0n}=r)=\frac{1}{\binom{n+r-1}{r}},$$

for $j_1+\cdots+j_n=r$, $j_k=0,1,\ldots$. The same distribution occurs in connection with the nonparametric two-sample problem. In this case $r, n-1$ are the sample sizes of the first and second sample from the same continuous distribution.

The distribution above also occurs in connection with Pólya sampling from an urn containing initially n balls, all of different colors. A drawn ball is replaced together with one more of the same color. The probability of getting $j_1,\ldots,j_n$ of the different colors in r drawings is $1/\binom{n+r-1}{r}$.

4.1.4. *Randomized paired comparisons.* Let $Z_1,\ldots,Z_n$ be defined as in Section 4.1.1 with $\alpha=\frac{1}{2}$. Then the distribution of $S_n=\sum_{k=1}^{n}a_kZ_k$ is of interest. For example, if $a_k=k$, this is the one-sample Wilcoxon statistic. To obtain results for this, we can use the approximations of Section 3.2 if we show the conditions of Sections 2.1 and 2.3 hold. Similar considerations to those of Section 4.1 will yield an expansion as in (3.8) with $s=5$. This is used in Section 4.3 to obtain numerical results for the one-sample Wilcoxon statistic with $n=10$.

4.1.5. *An empty cell statistic.* As an example of the approximations in Section 3.3 consider the following occupancy scheme. Let $X_{01} + \alpha$ be Poisson with mean $\alpha = r/n$, and $X_{11} = I(X_{01} = \alpha)$, so that the conditional distribution of S_{1n} given $S_{0n} = 0$ is that of the number of empty cells when r balls are distributed uniformly at random into n cells. In this case

$$\kappa_n(\theta) = n\left[\log(e^{\theta_1} + e^{\alpha \exp(\theta_0)} - 1) - \alpha(1 + \theta_0)\right].$$

4.1.6. *A covering problem.* In each of the conditional probabilities considered before, the conditioning has been taken with respect to a lattice variable. However, the results permit a wider application both to conditioning with respect to a variable with a density or to conditioning with respect to the event that a variable lies in some small interval. We will illustrate the first of these by considering the distribution of the vacancy after placing a number of intervals randomly on (0, 1).

Let $U_1, \dots, U_n$ be spacings of $n - 1$ independent uniform random variables on (0, 1). For a given number $a_n \in (0, 1)$, the vacancy V_n is defined by

$$nV_n = n\sum_{i=1}^{n} (U_i - a_n)_+ = \sum_{i=1}^{n} (nU_i - \alpha)_+,$$

where $(x)_+ = x$ if $x > 0$, 0 otherwise, and $na_n = \alpha$, a constant. If $Z, Z_1, Z_2, \dots$ are iid exponential with mean 1, then the distribution of nV_n is the same as the conditional distribution of $S_{1n} = \Sigma_{i=1}^{n}(Z_i - \alpha)_+$ given $S_{2n} = \Sigma_{i=1}^{n} Z_i = n$. For this case

$$\begin{aligned}\kappa(\theta) &= Ee^{\theta_1 Z + \theta_2 (Z - \alpha)}\\ &= (1 - \theta_1)^{-1}(1 - e^{-\alpha(1-\theta_1)}) + (1 - \theta_1 - \theta_2)^{-1} e^{-\alpha(1-\theta_1)},\end{aligned}$$

for $\theta_1 + \theta_2 < 1$.

We can solve (2.15) using this $\kappa(\theta)$, and use the results of Remark (7) to obtain an approximation for $P(V_n > e^{-\alpha} + x | \Sigma_{i=1}^{n} Z_i = n)$. This will be exactly of the form given in (3.20), although that equation was derived for conditioning with respect to a lattice random variable.

4.2. *A general computing method using* MACSYMA. MACSYMA is a symbolic operator program which enables general calculation of derivatives and algebraic manipulations together with numerical calculations. It can be used to give a general program to calculate the asymptotic approximations used here using a variety of forms for $\kappa_n(\theta)$. This has been done in the univariate and bivariate cases to obtain a number of examples.

The general program consists essentially of three "blocks" or subroutines:

A. A block for obtaining partial derivatives of $\kappa_n(\theta)$.

B. An algorithm for iterative solution $\hat{\theta}$ of $\kappa_n'(\theta) = x$ using $\kappa_n'(\theta)$ and $\kappa_n''(\theta)$, obtained numerically.

TABLE 1
Relative errors of tail probabilities as percentages

	A		B		C		D		E	
	a c	b d	a c	b d	a c	b d	a c	b d	e g	f d
0.0001			−18	−30	−8	−31	5	−8	2	8
			−292	993	102	308	150	800	5	200
0.0005			1	−4	3	−11	0	−12	−1	4
			−30	319	41	83	57	175	1	90
0.0010	−6	−35	0	4	2	−11	0	−10		
	34	190	−13	230	28	53	37	101		
0.0050	−6	−21	−1	−3	0	−10	−1	−10	−3	7
	4	47	4	88	6	4	16	38	2	46
0.0100	−1	−12	0	−2	0	−6	0	−9		
	2	28	4	57	2	−5	8	14		
0.0500	0	−8	0	−1	0	−7	0	−5	−9	13
	0	−2	2	16	−1	−13	1	0	3	36
0.1000	1	−5	0	−1	0	−5	0	−4		
	0	−4	1	9	−1	−10	0	−4		
0.2000	0	−4	0	0	0	−4	−1	−3	−78	30
	0	−4	0	4	−1	−7	−1	−3	0	23

[a]Saddlepoint $s = 5$.
[b]Saddlepoint $s = 3$.
[c]Edgeworth $s = 5$.
[d]Normal approximation.
[e]Discrete saddlepoint (3.18) $s = 6$.
[f]Discrete saddlepoint (3.18) $s = 4$.
[g]Discrete saddlepoint (3.17).

C. A block giving:

(a) Formulas for Hermite–Chebyshev polynomials, $H_k(y)$, for $\rho_k(\lambda)$ of Section 3.2 and for $\pi_k(\theta)$ of Section 3.1.

(b) Formulas for Q_{ln} from (1.11) in terms of standardized cumulants.

(c) Formulas for approximations as used in (3.8), (3.17)–(3.9) and (3.21) and (3.22); these cover the cases given in Section 3.5 and, in particular, the special cases described in Section 4.1.

These "blocks" are now used for the calculation. We input the function $\kappa_n(\theta)$ and use block A to obtain derivatives of $\kappa_n(V^{-1/2}\theta)$, for arbitrary V. We now input parameters (such as $n, a_1, \dots, a_n$) and using $V = I$, obtain $\hat{\theta}$ from block B using the first two derivatives of $\kappa_n(\theta)$. Then $\hat{\theta}$ is used to obtain $\hat{V} = V_n(\hat{\theta})$ and derivatives of $\kappa_n(\hat{V}^{-1/2}\theta)$ are evaluated at $\hat{\theta}$. Finally, block C is used to obtain the approximations.

4.3. *Numerical approximations.* The approximations were calculated and compared with the exact probabilities in a number of cases. To illustrate the accuracy of the methods, Table 1 gives the relative error as a percentage for probabilities as close as possible to a set of selected values from 0.0001 to 0.2. We calculated these values for the following cases:

A. One-sample Wilcoxon statistic, $n = 10$ (Section 4.1.4).

B. Two-sample Wilcoxon statistic with two samples of sizes 8, 8, from negative binomial, $n = 9$, $p = 9/17$ (Section 4.1.3).

C. As in case B, from binomial with $n = 16$, $p = \frac{1}{2}$ (Section 4.1.1).

D. Bootstrap sample using Wilcoxon scores, from Poisson with parameter 1 and $n = 10$ (Section 4.1.2).

E. Empty cell statistic, $n = 20$, $\alpha = 1$ (Section 4.1.5).

Table 1 gives results for both saddlepoint and Edgeworth approximations with $s = 3$ and $s = 5$ (in the lattice case as in case E, these can be taken to be $s = 4$ and $s = 6$).

Inspection of Table 1 shows that the saddlepoint approximation with $s = 5$, that is, with an error of order $n^{-3/2}$, gives excellent results throughout the range in all cases, except case E, which is discussed below. In the examples given and in a number of others calculated but not given here, these approximations have relative errors which are remarkably small. This contrasts with the much larger relative errors in the tails for both the normal approximation and the Edgeworth expansion with $s = 5$, and is generally a noticeable improvement on the saddlepoint approximation with $s = 3$. It is, of course, expected that the improvement in the extreme tails will be significant, but it is of interest that the saddlepoint method gives results at least as good as the Edgeworth throughout the range. It is worthwhile comparing the results in cases B and C with those of Stone (1969), who used a less accurate large deviation result (for example B) and found that it generally gave poor results.

In case E, the empty cell statistic, the methods of Section 3.3 are used. The saddlepoint technique based on (3.19), which uses an expansion of the exponential terms, gives good approximations in the extreme tails but fails in less extreme regions. However, if (3.18) is used the approximations are accurate throughout the range. We also calculated local results and obtained tail probabilities by summing these, as suggested by the methods of Daniels (1954). These give results even more accurate than (g) in the table, however, the calculations required for each tail probability are extensive.

APPENDIX A

Proof of Lemma 2. For $y \in S(\varepsilon)$,

$$\delta(\theta, B_\varepsilon, w - y) = \int \chi_{B_\varepsilon}(w - y - u) e^{\theta \cdot (w-y-u)} (\nu_{n\theta}(du) - e_{s-3}(u, \nu_{n\theta})\, du_1)$$

$$\geq e^{-\theta \cdot y} \delta(\theta, B, w) - e^*_{s-3}(\nu_{n\theta}) e^{\varepsilon|\theta|} \hat{\chi}_{\theta, B_{2\varepsilon} \setminus B}(0).$$

Thus

$$\int_{S(\varepsilon)} \delta(\theta, B_\varepsilon, w - y) k(Ty_1) T^{d_1} \, dy_1$$

$$\geq \delta(\theta, B, w) \int_{S(\varepsilon)} e^{-\theta \cdot y} k(Ty_1) T^{d_1} \, dy_1$$

$$- (1 - \alpha) e^{\varepsilon|\theta|} e^*_{s-3}(\nu_{n\theta}) \hat{\chi}_{\theta, B_{2\varepsilon} \setminus B}(0)$$

$$= \delta(\theta, B, w)(1 - \alpha)\kappa - (1 - \alpha) e^{\varepsilon|\theta|} e^*_{s-3}(\nu_{n\theta}) \hat{\chi}_{\theta, B_{2\varepsilon} \setminus B}(0),$$

where $e^{-\varepsilon|\theta|} \leq \kappa \leq e^{\varepsilon|\theta|}$. Also, for $y \notin S(\varepsilon)$,

$$\delta(\theta, B_\varepsilon, w - y) \geq \delta(\theta, B, w - y) - e^*_{s-3}(\nu_{n\theta}) \hat{\chi}_{\theta, B_\varepsilon \setminus B}(0)$$

$$\geq - \sup_z |\delta(\theta, B, z)| - e^*_{s-3}(\nu_{n\theta}) \hat{\chi}_{\theta, B_{2\varepsilon} \setminus B}(0).$$

Multiplying by $k(Ty_1)T^{d_1}$, integrating over $S(\varepsilon)^c$, combining this with the integral over $S(\varepsilon)$ and taking the supremum over w gives

$$\Delta(\theta, B_\varepsilon, T) \geq (1 - \alpha)\kappa \sup_w \delta(\theta, B, w) - \alpha \sup_z |\delta(\theta, B, z)|$$

$$- \left[(1 - \alpha) e^{\varepsilon|\theta|} + \alpha\right] e^*_{s-3}(\nu_{n\theta}) \hat{\chi}_{\theta, B_{2\varepsilon} \setminus B}(0).$$

Thus if we choose ε small enough to make $(1 - \alpha)\kappa > 1 - 2\alpha$ and $(1 - \alpha)e^{\varepsilon|\theta|} < 1$,

$$(1 - 2\alpha) \sup_w \delta(\theta, B, w) \leq \Big\{\Delta(\theta, B_\varepsilon, T) + \alpha \sup_z |\delta(\theta, B, z)|$$

$$+ (1 + \alpha) e^*_{s-3}(\nu_{n\theta}) \hat{\chi}_{\theta, B_{2\varepsilon} \setminus B}(0)\Big\}.$$

Similarly, we can show that

$$-(1 - 2\alpha) \inf_w \delta(\theta, B, w) \leq \Big\{\Delta(\theta, B_{-\varepsilon}, T) + \alpha \sup_z |\delta(\theta, B, z)|$$

$$+ (1 + \alpha) e^*_{s-3}(\nu_{n\theta}) \hat{\chi}_{\theta, B \setminus B_{-2\varepsilon}}(0)\Big\}.$$

Combining these, we obtain (1.27).

APPENDIX B

Proof of (3.11). Since $(-1)^k H_k(u)\varphi(u) = \varphi^{(k)}(u)$, it is easy to show that

$$H_k(u) = u H_{k-1}(u) - H'_{k-1}(u).$$

So

$$\begin{aligned}
&\int_\lambda^\infty H_k(v-\lambda)e^{-v^2/2}\,dv \\
&\quad = \int_\lambda^\infty H_{k-1}(v-\lambda)\frac{d}{dv}(-e^{-v^2/2})\,dv \\
&\qquad - \lambda\int_\lambda^\infty H_{k-1}(v-\lambda)e^{-v^2/2}\,dv - \int_\lambda^\infty H'_{k-1}(v-\lambda)e^{-v^2/2}\,dv \\
&\quad = H_{k-1}(0)e^{-\lambda^2/2} - \lambda\int_\lambda^\infty H_{k-1}(v-\lambda)e^{-v^2/2}\,dv \\
&\quad = \sum_{l=1}^{k} H_{k-l}(0)(-\lambda)^{l-1}e^{-\lambda^2/2} + (-\lambda)^k\int_\lambda^\infty e^{-v^2/2}\,dv.
\end{aligned}$$

That is,

$$\tau(\lambda)\rho_k(\lambda) = (-\lambda)^k\tau(\lambda) + (2\pi)^{-1/2}\sum_{j=0}^{k-1} H_j(0)(-\lambda)^{k-j-1}.$$

REFERENCES

ALBERS, W., BICKEL, P. J. and VAN ZWET, W. R. (1976). Asymptotic expansions for the power of distribution free tests in the one-sample problem. *Ann. Statist.* **4** 108–156.

BARNDORFF-NIELSEN, O. and COX, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 279–312.

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

BLACKWELL, D. and HODGES, J. L. (1959). The probability in the extreme tail of a convolution. *Ann. Math. Statist.* **30** 1113–1120.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25** 631–650.

EFRON, B. and TRUAX, D. (1968). Large deviations theory in exponential families. *Ann. Math. Statist.* **39** 1402–1424.

ERDÖS, P. and RÉNYI, A. (1959). On the central limit theorem for samples from a finite population. *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* **4** 49–61.

HÖGLUND, T. (1979). A unified formulation of the central limit theorem for small and large deviations from the mean. *Z. Wahrsch. Verw. Gebiete* **49** 105–117.

HOLST, L. (1979). Two conditional limit theorems with applications. *Ann. Statist.* **7** 551–557.

MORRIS. C. (1975). Central limit theorems for multinomial sums. *Ann. Statist.* **3** 165–188.

QUINE, M. P. (1985). Remainder term estimates in a conditional central limit theorem for integer-valued random variables. *J. Austral. Math. Soc. Ser. A* **39** 75–85.

QUINE, M. P. and ROBINSON, J. (1984). Normal approximations to sums of scores based on occupancy numbers. *Ann. Probab.* **12** 794–804.

ROBINSON, J. (1978). An asymptotic expansion for samples from a finite population. *Ann. Statist.* **6** 1005–1011.

ROBINSON, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *J. Roy. Statist. Soc. Ser. B* **44** 91–101.

SKOVGAARD, I. M. (1986). On multivariate Edgeworth expansions. *Internat. Statist. Rev.* **54** 169–186.

STONE, C. (1965). A local limit theorem for nonlattice multi-dimensional distribution functions. *Ann. Math. Statist.* **36** 546–551.

STONE, C. (1967). On local and ratio limit theorems. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **2** 217–224. Univ. California Press.

STONE, M. (1969). Approximations to extreme tail probabilities for sampling without replacement. *Math. Proc. Cambridge Philos. Soc.* **66** 587–606.
VON BAHR, B. (1967a). Multi-dimensional integral limit theorems. *Ark. Mat.* **7** 71–88.
VON BAHR, B. (1967b). Multi-dimensional integral limit theorems for large deviations. *Ark. Mat.* **7** 89–99.

J. ROBINSON
DEPARTMENT OF MATHEMATICAL STATISTICS
THE UNIVERSITY OF SYDNEY
NEW SOUTH WALES 2006
AUSTRALIA

T. HÖGLUND
ROYAL INSTITUTE OF TECHNOLOGY
S-100 44 STOCKHOLM
SWEDEN

L. HOLST
ROYAL INSTITUTE OF TECHNOLOGY
S-100 44 STOCKHOLM
SWEDEN

M. P. QUINE
DEPARTMENT OF MATHEMATICAL STATISTICS
THE UNIVERSITY OF SYDNEY
NEW SOUTH WALES 2006
AUSTRALIA