The Annals of Probability
1977, Vol. 5, No. 6, 913–925

LARGE DEVIATION PROBABILITIES FOR SAMPLES FROM A FINITE POPULATION

BY J. ROBINSON

University of Sydney

Let X_n be the standardized mean of s observations obtained by simple random sampling from the n numbers $a_{n1}, \cdots, a_{nn}$ and let b_n be the maximum deviation of these numbers from their mean. If b_n tends to zero then the distribution function of X_n tends uniformly to the normal distribution function. However this approximation is not adequate at the tails of the distribution. Here we obtain limit theorems for $P(X_n > x)$ in the two cases when $x = o(b_n^{-1})$ and $x = O(b_n^{-1})$. These are related to similar results for sums of independent random variables.

1. Introduction. Let $\{a_{nk} : k = 1, \cdots, n, n = 2, 3, \cdots\}$ be a triangular array of real numbers and suppose that $\sum_{k=1}^n a_{nk} = 0$ and $\sum_{k=1}^n a_{nk}^2 = 1$. Let $(R_{n1}, \cdots, R_{nn})$ be a random vector taking each of the $n!$ permutations of $(1, \cdots, n)$ with equal probability and for $s_n < n$, let

$$X_n = (p_n q_n)^{-\frac{1}{2}} \sum_{k=1}^{s_n} a_{nR_{nk}} ,$$

where $p_n = s_n/n$ and $q_n = 1 - p_n$. Write $F_n(x) = P(X_n \leqq x)$. Then if $b_n = \max_{1 \leqq k \leqq n} |a_{nk}|$, Hájek (1960) has shown that $F_n(x)$ converges to $\Phi(x)$, where $\Phi(x)$ is the distribution function of a standardized normal variate, if and only if b_n tends to zero. Erdös and Rényi (1959) proved the sufficiency of this result using an analytically tractable form of the characteristic function of X_n and Bikelis (1969) used this form to obtain a bound on the rate of convergence. This form is exploited here to obtain limit theorems for large deviation probabilities of the type described for sums of independent random variables in Feller (1966, Chapter XVI), Ibragimov and Linnik (1971, Chapter 6, 8) and Petrov (1975, Chapter VIII).

Let $Q_n(u + iv)$ be the complex moment generating function of X_n defined by

$$\begin{aligned} Q_n(u + iv) &= \int_{-\infty}^{\infty} \exp\{(u + iv)x\}\, dF_n(x) \\ &= \binom{n}{s}^{-1} \sum{}^{*} \exp\{(u + iv)(a_{k_1} + \cdots + a_{k_s})(pq)^{-\frac{1}{2}}\} \qquad (1) \\ &= (2\pi B_{ns})^{-1} \int_{-\pi}^{\pi} \prod_{k=1}^{n} (q + pe^{(u+iv)a_k(pq)^{-\frac{1}{2}}+\alpha+i\theta}) e^{-s(\alpha+i\theta)}\, d\theta \end{aligned}$$

for any α, where here and in the sequel we omit the subscript n from a_{nk}, s_n, p_n and q_n, $\sum^*$ denotes summation over all choices of $1 \leqq k_1 < k_2 < \cdots < k_s \leqq n$ and $B_{ns} = \binom{n}{s} p^s q^{n-s}$. For since

$$\begin{aligned} \int_{-\pi}^{\pi} e^{im\theta} &= 2\pi , \quad & m &= 0 \\ &= 0 , & m &= \pm 1, \pm 2, \cdots , \end{aligned}$$

Received April 15, 1976; revised February 8, 1977.

AMS 1970 *subject classifications*. Primary 60F10; Secondary 62G20.

Key words and phrases. Large deviations, limit theorems, asymptotic efficiency, sampling without replacement, permutation tests.

we can write the product behind the integral as a sum and only terms appearing in the sum given by $\sum^*$ do not vanish.

In the next section we will obtain an approximation for $Q_n(u + iv)$ of a form which will subsequently be used for asymptotic large deviation results. Here a saddlepoint approximation is used for the integrand in (1). Section 3 contains the main result giving an estimate of the relative error of the approximation of $1 - F_n(x)$ by $1 - \Phi(x)$ for $x = o(b_n^{-1})$. Then in Section 4, we use Lemma 1 and a general result of Plachky and Steinebach (1975) to obtain, under some restrictions on the a_k, the limit of $n^{-1} \log [1 - F_n(x_n)]$ for very large deviations, when $x_n = O(n^{\frac{1}{2}})$. This result is compared to the earlier result of Stone (1969). In Section 5 we consider the relationship of these limit theorems for simple random sampling to similar results for Poisson sampling.

2. The moment generating function. Let

$$K(z) = \log (pe^{qz} + qe^{-pz}) ,$$

for $z = u + iv$, where log denotes the principal value of the logarithm. Then since

$$|pe^{qz} + qe^{-pz}|^2 = (pe^{qu} + qe^{-pu})^2 \left[1 - \frac{2pqe^{(q-p)u}(1 - \cos v)}{(pe^{qu} + qe^{-pu})^2}\right], \tag{2}$$

and

$$\frac{pqe^{(q-p)u}}{(pe^{qu} + qe^{-pu})^2} \leq \tfrac{1}{4} \tag{3}$$

with equality only when $u = \log (q/p)$, it is readily seen that $K(z)$ is an analytic function of z in the region $|u| < C$ and $|v| < \pi - \varepsilon$ for any $0 < \varepsilon < \pi$ and $0 < C < \infty$. In the sequel B will denote a positive quantity depending only on p but which may change with each occurrence. For any fixed $0 < p^* < 1$, if $p^* < p < 1 - p^*$, then B will be bounded. Write $K'(z), K''(z), K'''(z)$ for the first, second and third derivatives of $K(z)$ in the domain of analyticity. In the sequel $\sum$ will always denote summation over the subscript k from 1 to n.

LEMMA 1. *For any fixed $C > 0$, if $|u| < Cb_n^{-1}(pq)^{-\frac{1}{2}}$, then there exists $\varepsilon > 0$ and $B > 0$, depending only on p, such that for $|v| < \varepsilon b_n^{-1}(pq)^{\frac{1}{2}}$,*

$$Q_n(u + iv) = B_{ns}^{-1}(2\pi \textstyle\sum K_k'')^{-\frac{1}{2}} e^{\sum K_k + ivm_n - \frac{1}{2}v^2\sigma_n^2}(1 + R) \tag{4}$$

where

$$m_n = (pq)^{-\frac{1}{2}} \textstyle\sum a_k K_k' , \qquad \sigma_n^2 = (pq)^{-1}[\sum a_k^2 K_k'' - (\sum a_k K_k'')^2 / \sum K_k''] , \tag{5}$$

where K_k, K_k' and K_k'' are the values of $K(x)$, $K'(x)$ and $K''(x)$ evaluated at $x = ua_k(pq)^{-\frac{1}{2}} + \alpha_n(u)$, where $\alpha_n(u)$ is the unique real solution of the equation

$$\textstyle\sum K'[ua_k(pq)^{-\frac{1}{2}} + \alpha] = 0 \tag{6}$$

and where

$$|R| < Bb_n(|v|^3 + B)e^{\frac{1}{4}v^2\sigma_n^2} . \tag{7}$$

PROOF. In the integral given in (1), let $\theta = \phi n^{-\frac{1}{2}}$ and write $\zeta_k = ua_k(pq)^{-\frac{1}{2}} + \alpha$ and $\xi_k = va_k(pq)^{-\frac{1}{2}} + \phi n^{-\frac{1}{2}}$. Then

$$Q_n(u + iv) = I_1 + I_2$$

where

$$I_1 = (2\pi B_{ns} n^{\frac{1}{2}})^{-1} \int_{-2\varepsilon n^{\frac{1}{2}}}^{2\varepsilon n^{\frac{1}{2}}} \exp\{\sum K(\zeta_k + i\xi_k)\}\, d\phi \tag{8}$$

and

$$I_2 = (2\pi B_{ns} n^{\frac{1}{2}})^{-1} \int_{2\varepsilon n^{\frac{1}{2}} < |\phi| < \pi n^{\frac{1}{2}}} \prod_{k=1}^{n} [qe^{-p(\zeta_k + i\xi_k)} + pe^{q(\zeta_k + i\xi_k)}]\, d\phi\,, \tag{9}$$

where $0 < \varepsilon < \frac{1}{2}\pi$ will be chosen later. For any $C > 0$, if $-2C < x < 2C$, then we can find $\delta > 0$, depending only on p, such that

$$K''(x) = pqe^{(q-p)x}(pe^{qx} + qe^{-px})^{-2} > \delta \tag{10}$$

and for $-\frac{1}{2}\pi < y < \frac{1}{2}\pi$,

$$|K'''(x + iy)| < B\,.$$

Then using a complex version of the Taylor series with remainder (see, for example, Copson (1935, pages 72–73)) for the integrand E in (8), we have

$$\begin{aligned} E &= \exp \sum \{K(\zeta_k) + i\xi_k K'(\zeta_k) - \tfrac{1}{2}\xi_k^2 K''(\zeta_k) + \tfrac{1}{6}\theta_1|\xi_k|^3 K''(\zeta_k)\} \\ &= \exp\{\sum K(\zeta_k) + iv(pq)^{-\frac{1}{2}} \sum a_k K'(\zeta_k) + i\phi n^{-\frac{1}{2}} \sum K'(\zeta_k) \\ &\quad - \tfrac{1}{2}v^2(pq)^{-1} \sum a_k^2 K''(\zeta_k) - v\phi(npq)^{-\frac{1}{2}} \sum a_k K''(\zeta_k) \\ &\quad - \tfrac{1}{2}\phi^2 n^{-1} \sum K''(\zeta_k) + \tfrac{1}{6}\theta_1 \sum |\xi_k|^3 K''(\zeta_k)\}\,, \end{aligned} \tag{11}$$

where θ_1 is some quantity with $|\theta_1| < B\delta^{-1}$, for $|\zeta_k| = |ua_k(pq)^{-\frac{1}{2}} + \alpha| < 2C$.

For any fixed u with $|u| < Cb_n^{-1}(pq)^{\frac{1}{2}}$, $\sum K'(ua_k(pq)^{-\frac{1}{2}} + \alpha)$ is negative when $\alpha < -C$ and positive when $\alpha > C$. Further, it is strictly monotone since $K''(ua_k(pq)^{-\frac{1}{2}} + \alpha) > 0$. So (6) has a unique solution $\alpha_n = \alpha_n(u)$, say, with $-C < \alpha_n(u) < C$ for $|u| < Cb_n^{-1}(pq)^{\frac{1}{2}}$. The integral in (1) is unchanged by taking this particular choice for α, so we can rewrite the integrals I_1 and I_2 using this choice of α. Then, using K_k, K_k', K_k'' as defined in the statement of the lemma, we have from (11) and (6),

$$\begin{aligned} E = \exp\Big\{&\sum K_k - \tfrac{1}{2}\Big\{\phi + \frac{n^{\frac{1}{2}}(pq)^{-\frac{1}{2}}v \sum a_k K_k''}{\sum K_k''}\Big\}^2 \frac{\sum K_k''}{n} \\ &+ iv(pq)^{-\frac{1}{2}} \sum a_k K_k' \\ &- \tfrac{1}{2}(pq)^{-1}v^2[\sum a_k^2 K_k''] - (\sum a_k K_k'')^2/\sum K_k'']\Big\}(1 + R_1)\,, \end{aligned} \tag{12}$$

where

$$R_1 = \exp\{\tfrac{1}{6}\theta_1 \sum |\xi_k|^3 K_k''\} - 1\,. \tag{13}$$

We can choose $\varepsilon' > 0$, such that

$$\begin{aligned} \tfrac{1}{6}\theta_1 \sum |\xi_k|^3 K_k'' &< \tfrac{1}{4} \sum \xi_k^2 K_k'' \\ &= \tfrac{1}{4}\Big[\Big\{\phi + \frac{n^{\frac{1}{2}}(pq)^{-\frac{1}{2}}v \sum a_k K_k''}{\sum K_k''}\Big\}^2 \frac{\sum K_k''}{n} + v^2\sigma_n^2\Big]\,, \end{aligned} \tag{14}$$

for $|\xi_k| < \varepsilon'$. Also, since $\delta < K_k'' \leqq \frac{1}{4}$, when $|\zeta_k| < 2C$, as shown in (3) and (10) and since $|\theta_1| < B\delta^{-1}$ and $b_n \geqq n^{-\frac{1}{2}}$, it follows from the c_r-inequality (Loève (1963), page 155) that

$$(15) \qquad |\tfrac{1}{6}\theta_1 \sum |\xi_k|^3 K_k''| \leqq \frac{B\delta^{-1}b_n}{24(pq)^{\frac{3}{2}}} (|v| + |\psi|)^3 \\ \leqq Bb_n(|v|^3 + |\psi|^3) \, .$$

So for $|\xi_k| < \varepsilon'$, using (14) and (15) in (13), we have

$$(16) \qquad |R_1| < Bb_n(|v|^3 + |\psi|^3) \exp\left\{\tfrac{1}{4}\left[\left(\psi + \frac{n^{\frac{1}{2}}(pq)^{-\frac{1}{2}}v \sum a_k K_k''}{\sum K_k''}\right)^2 \times \frac{\sum K_k''}{n} + v^2\sigma_n^{\ 2}\right]\right\} .$$

Let $\varepsilon = \varepsilon'/3$ and consider $|v| < \varepsilon b_n^{-1}(pq)^{\frac{1}{2}}$. If $|\psi| < 2\varepsilon n^{\frac{1}{2}}$ then $|\xi_k| < 3\varepsilon = \varepsilon'$. To estimate I_1 we need to integrate E over the range $|\psi| < 2\varepsilon n^{\frac{1}{2}}$. Since

$$(17) \qquad (\sum a_k K_k'')^2 \leqq (\sum a_k^2 K_k'')(\sum K_k'')$$

and from (3) and (10), $\delta < K_k'' \leqq \frac{1}{4}$,

$$(18) \qquad \left|\frac{n^{\frac{1}{2}}(pq)^{-\frac{1}{2}}v \sum a_k K_k''}{\sum K_k''}\right| \leqq \frac{n^{\frac{1}{2}}(pq)^{-\frac{1}{2}}|v|}{2(n\delta)^{\frac{1}{2}}} \leqq B|v| \, .$$

So we have, from the c_r-inequality (Loève (1963, page 155))

$$(19) \qquad |\psi|^3 \leqq 2^2\left|\psi + \frac{n^{\frac{1}{2}}(pq)^{-\frac{1}{2}}v \sum a_k K_k''}{\sum K_k''}\right|^3 + B|v|^3 \, .$$

So from (16) and (19),

$$(20) \qquad \left|\int_{-2\varepsilon n^{\frac{1}{2}}}^{2\varepsilon n^{\frac{1}{2}}} R_1 \exp\left\{-\tfrac{1}{2}\left(\psi + \frac{n^{\frac{1}{2}}(pq)^{-\frac{1}{2}}v \sum a_k K_k''}{\sum K_k''}\right)^2 \frac{\sum K_k''}{n}\right\} d\psi\right| \\ \leqq Bb_n(|v|^3 + B)e^{\frac{1}{4}v^2\sigma_n^2} \, .$$

Also for $|v| < \varepsilon b_n^{-1}(pq)^{\frac{1}{2}}$,

$$(21) \qquad \left|\frac{n^{\frac{1}{2}}(pq)^{-\frac{1}{2}}v \sum a_k K_k''}{\sum K_k''}\right| < \varepsilon n^{\frac{1}{2}} \, ,$$

so, since $\Phi(x) < x^{-1}\exp(-\frac{1}{2}x^2)$ for $x > 0$,

$$(22) \qquad \left|\int_{|\psi|>2\varepsilon n^{\frac{1}{2}}} \exp\left\{-\tfrac{1}{2}\left(\psi + \frac{n^{\frac{1}{2}}(pq)^{-\frac{1}{2}}v \sum a_k K_k''}{\sum K_k''}\right)^2 \frac{\sum K_k''}{n}\right\} d\psi\right| \\ \leqq B(\sum K_k'')^{-\frac{1}{2}} \exp(-\tfrac{1}{2}\varepsilon^2 \sum K_k'') \\ \leqq B(\sum K_k'')^{-\frac{1}{2}} \exp(-\tfrac{1}{4}v^2\sigma_n^{\ 2} - \tfrac{1}{4}\varepsilon^2\delta n) \\ \leqq Bb_n(\sum K_k'')^{-\frac{1}{2}}e^{-\frac{1}{4}v^2\sigma_n^2} \, ,$$

where we have used the inequalities $K_k'' > \delta$ and

$$(23) \qquad \sigma_n^{\ 2} \leqq (pq)^{-1} \sum a_k^2 K_k'' \leqq (pq)^{-1}b_n^{\ 2} \sum K_k'' \, .$$

Further,

$$(24)\qquad \int_{-\infty}^{\infty} \exp\left\{-\tfrac{1}{2}\left(\psi + \frac{n^{\frac{1}{2}}(pq)^{-\frac{1}{2}}v \sum a_k K_k''}{\sum K_k''}\right)^2 \frac{\sum K_k''}{n}\right\} d\psi = (2\pi n/\sum K_k'')^{\frac{1}{2}} .$$

So, combining the results (20), (22) and (24) in the integral of E over the range $|\psi| < 2\varepsilon n^{\frac{1}{2}}$, we have

$$(25)\qquad I_1 = B_{ns}^{-1}(2\pi \sum K_k'')^{-\frac{1}{2}} e^{\sum K_k + ivm_n - \frac{1}{2}v^2\sigma_n^2}(1 + R_2)$$

where

$$(26)\qquad |R_2| < Bb_n(|v|^3 + B)e^{\frac{1}{4}v^2\sigma_n^2} .$$

To estimate I_2 for $|v| < \varepsilon b_n^{-1}(pq)^{\frac{1}{2}}$, we note that $2\varepsilon n^{\frac{1}{2}} \leqq |\psi| \leqq \pi n^{\frac{1}{2}}$ and so $|\xi_k| > \varepsilon$. Also from (2)

$$(27)\qquad |qe^{-p(\zeta_k+i\xi_k)} + pe^{q(\zeta_k+i\xi_k)}|^2 = e^{2K(\zeta_k)}[1 - 2K''(\zeta_k)(1 - \cos \xi_k)]$$

For $\varepsilon < |\xi_k| < 2\pi - \varepsilon$,

$$(28)\qquad 1 - \cos \xi_k \geqq 1 - \cos \varepsilon \geqq \varepsilon^2/2 - \varepsilon^4/24 \geqq \varepsilon^2/3 ,$$

for $\varepsilon < 2$. So using (27) and (28) and putting $\alpha = \alpha_n(u)$, the solution of (6), we have

$$(29)\qquad \prod_{k=1}^{n} |qe^{-p(\zeta_k+i\xi_k)} + pe^{q(\zeta_k+i\xi_k)}| \leqq \exp[\sum K_k - \tfrac{1}{3}\varepsilon^2 \sum K_k''] \leqq \exp[\sum K_k - \tfrac{1}{4}v^2\sigma_n^2 - \tfrac{1}{12}\varepsilon^2\delta n]$$

where we have used the inequalities $K_k'' > \delta$ and (23). Using the estimate (29) in (9) we have, for $|v| < \varepsilon b_n^{-1}(pq)^{\frac{1}{2}}$,

$$(30)\qquad |I_2| < B_{ns}^{-1} \exp[\sum K_k - \tfrac{1}{4}v^2\sigma_n^2 - \tfrac{1}{12}\varepsilon^2\delta n] < Bb_n B_{ns}^{-1}(2\pi \sum K_k'')^{-\frac{1}{2}} \exp[\sum K_k - \tfrac{1}{4}v^2\sigma_n^2] .$$

Combining (25), (26) and (30) leads immediately to the result (4) with the inequality (7).

REMARK. It is worthwhile noticing that in the particular case when $p = q$ and the a_k are symmetric, that is when $\sum a_k^{2j+1} = 0$ for $j = 1, 2, \cdots$, the solution of (6) is $\alpha = 0$.

3. The relative error of the normal approximation. Let V_n be an associated distribution for F_n, defined by

$$(31)\qquad dV_n(x) = [Q_n(u)]^{-1}e^{ux}\, dF_n(x)$$

and let $G_n(x)$ be the distribution function of a normal variate with mean m_n and variance σ_n^2. Then the characteristic functions of V_n and G_n will be denoted by $Q_V(v) = Q_n(u + iv)/Q_n(u)$ and $Q_G(v) = \exp(ivm_n - \frac{1}{2}v^2\sigma_n^2)$, respectively. First we show that G_n gives a uniform approximation for V_n.

LEMMA 2. *For any $C > 0$ and $|u| < Cb_n^{-1}(pq)^{\frac{1}{2}}$, there exists $B > 0$, depending* *only on p, such that*

$$\sup_x |V_n(x) - G_n(x)| < Bb_n . \tag{32}$$

PROOF. We will use the well-known inequality (see, for example, Feller (1966), page 501), $\sup_x |V_n(x) - G_n(x)| < \int_{-T}^{T} |v|^{-1}|Q_V(v) - Q_G(v)|\, dv + 12m^*/\pi T$, where $m^* = \sup_x |G'(x)| = (2\pi\sigma_n^2)^{-\frac{1}{2}}$ and we take $T = \varepsilon b_n^{-1}(pq)^{\frac{1}{2}}$, with ε chosen as in Lemma 1. From Lemma 1, we have for $|v| < \varepsilon b_n^{-1}(pq)^{\frac{1}{2}}$

$$Q_V(v) = \exp[ivm_n - \tfrac{1}{2}v^2\sigma_n^2]\,\frac{1+R}{1+R'} ,$$

where R is given in (4) and satisfies the inequality (7) while R' is the particular value of R taken when $v = 0$ and so

$$|R'| < Bb_n .$$

We will assume $|R'| < \frac{1}{2}$, since otherwise we can choose B in (32) so that the inequality is trivial. So

$$\begin{aligned} |Q_V(v) - Q_G(v)| &< e^{-\frac{1}{2}v^2\sigma_n^2}\,\frac{|R - R'|}{|1 + R'|} \\ &< Bb_n(|v|^3 + B)e^{-\frac{1}{4}v^2\sigma_n^2} . \end{aligned} \tag{33}$$

Now since $\delta < K_k''$,

$$\begin{aligned} \sigma_n^2 &= (pq)^{-1} \sum (a_k - \sum a_k K_k''/\sum K_k'')^2 K_k'' \\ &> \delta(pq)^{-1} \sum (a_k - \sum a_k K_k''/\sum K_k'')^2 \geqq \delta(pq)^{-1} . \end{aligned}$$

So using (33),

$$\int_{-T}^{-1} + \int_1^T |v|^{-1}|Q_V(v) - Q_G(v)|\, dv < Bb_n \cdot \tag{34}$$

For $|v| < 1$, we notice that

$$|Q_V(v) - Q_G(v)| \leqq |v| \sup_{0 \leqq t \leqq u} |Q_V'(t) - Q_G'(t)| . \tag{35}$$

From (8)

$$\frac{dI_1}{dv} = [2\pi B_{ns} n^{\frac{1}{2}}]^{-1} \int_{-2\varepsilon n^{\frac{1}{2}}}^{2\varepsilon n^{\frac{1}{2}}} \sum ia_k(pq)^{-\frac{1}{2}}K'(\zeta_k + i\xi_k)e^{\sum K(\zeta_k + i\xi_k)}\, d\psi , \tag{36}$$

where the notation is as in Lemma 1 and the integral is evaluated at $\alpha = \alpha_n(u)$. Now at $\alpha = \alpha_n(u)$, we use a Taylor expansion to obtain

$$\begin{aligned} &i(pq)^{-\frac{1}{2}} \sum a_k K'(\zeta_k + i\xi_k) \\ &\quad = i[\sum a_k K_k' + \sum a_k i\xi_k K_k'' + \theta_2 \sum |a_k|\,|\xi_k|^2 K_k''] , \\ &\quad = im_n - v\sigma_n^2(pq)^{\frac{1}{2}} - \left(\psi + \frac{n^{\frac{1}{2}}v(pq)^{-\frac{1}{2}} \sum a_k K_k''}{\sum K_k''}\right) \sum a_k K_k'' \\ &\qquad + \theta_3 b_n(|\psi|^2 + |v|^2) , \end{aligned} \tag{37}$$

where θ_2 and θ_3 are bounded by positive quantities depending only on p. Now expanding the exponent as in Lemma 1, and integrating as we did to obtain

(25), we obtain from (36) and (37),

$$\frac{dI_1}{dv} = B_{ns}^{-1}(2\pi \sum K_k'')^{-\frac{1}{2}} e^{\sum K_k}[(im_n - v\sigma_n^2)\exp[ivm_n - \tfrac{1}{2}v^2\sigma_n^2] + R_3] \tag{38}$$

where, since $|v| < 1$,

$$|R_3| < Bb_n . \tag{39}$$

Further,

$$\frac{dI_2}{dv} = (2\pi B_{ns} n^{\frac{1}{2}})^{-1} \int_{2\varepsilon n^{\frac{1}{2}} \leq |\phi| < \pi n^{\frac{1}{2}}} \sum ia_k(pq)^{\frac{1}{2}}(e^{q(\zeta_k+i\xi_k)} - e^{-p(\zeta_k+i\xi_k)}) \times \prod_{k'\neq k} (pe^{q(\zeta_k+i\xi_k)} + qe^{-p(\zeta_k+i\xi_k)})\, d\phi .$$

In the same way as we obtained (30), we see that for $|v| < 1$,

$$\left|\frac{dI_2}{dv}\right| < Bb_n B_{ns}^{-1}(2\pi \sum K_k'')^{-\frac{1}{2}} e^{\sum K_k} . \tag{40}$$

Now since $Q_V(v) = Q_n(u + iv)/Q_n(u)$, it follows from (38), (39) and (40), that

$$|Q_V'(v) - Q_G'(v)| < Bb_n . \tag{41}$$

So from (35) and (41),

$$\int_{-1}^{1} |v|^{-1} Q_V(v) - Q_G(v)\,|\,dv < Bb_n . \tag{42}$$

Now the lemma follows from (34) and (42).

Now we can state and prove the main result giving an estimate of the ratio of the tail probability of the distribution of X_n to the corresponding tail probability of its normal approximation.

THEOREM 1. *Suppose* $b_n \to 0$ *as* $n \to \infty$ *and* $1 < x_n = o(b_n^{-1})$, *then*

$$1 - F_n(x_n) = [1 - \Phi(x_n)] e^{x_n^2 \lambda_n(x_n b_n)}[1 + O(x_n b_n)] , \tag{43}$$

as $n \to \infty$, *where* $\lambda_n(t)$ *is a power series which is majorized by a power series with coefficients not depending on n and convergent in some circle, so* $\lambda_n(t)$ *converges uniformly for sufficiently small* $|t|$. *If* $\alpha = \alpha_n(u)$ *is the solution of equation* (6) *and* $m_n = m_n(u)$ *is defined in* (5), *then we can define u as the unique real root of the equation*

$$b_n m_n(u) = t . \tag{44}$$

Then $\lambda_n(t)$ *is defined by*

$$t^2\lambda_n(t) = b_n^2[\sum K_k - (pq)^{-\frac{1}{2}} u \sum a_k K_k' + \tfrac{1}{2}(pq)^{-1}(\sum a_k K_k')^2] \tag{45}$$

where K_k, K_k' *are as defined in Lemma* 1. *In particular*, $\lambda_n(t) = \lambda_{1n} t + \lambda_{2n} t^2 + \cdots$, *with*

$$b_n\lambda_{1n} = \frac{q-p}{6(pq)^{\frac{1}{2}}} \sum a_k^3 , \tag{46}$$

$$b_n\lambda_{2n} = \frac{1-6pq}{24pq} \sum a_k^4 - \frac{(q-p)^2}{8npq} - \frac{(q-p)^2}{8pq}(\sum a_k^3)^2, \cdots .$$

PROOF. From (31),

$$1 - F_n(x_n) = Q_n(u) \int_{x_n}^{\infty} e^{-uy}\, dV_n(y) .$$

We use the approximation of Lemma 2 for $V_n(x)$ and the approximation of Lemma 1 for $Q_n(u)$. Then for $u > 0$,

$$(47) \qquad 1 - F_n(x_n) = B_{ns}^{-1}(2\pi \sum K_k'')^{-\frac{1}{2}} e^{\Sigma K_k - um_n} \times [(2\pi)^{-\frac{1}{2}} \int_{(x_n - m_n)/\sigma_n}^{\infty} e^{-uy\sigma_n - \frac{1}{2}y^2}\, dy + O(b_n)][1 + O(b_n)] .$$

Consider the function

$$f^{(n)}(z, \zeta) = \sum K(za_k(pq)^{-\frac{1}{2}} + \zeta) ,$$

as a function of the two complex variables $z = u + iv$ and $\zeta = \alpha + i\theta$, with a power series expansion in z and ζ about $(0, 0)$, which converges for $|z| < b_n^{-1}(pq)^{\frac{1}{2}}$ and $|\zeta| < 1$, for then $|(pq)^{-\frac{1}{2}}za_k + \zeta| < 2 < \pi$. The series is

$$f^{(n)}(z, \zeta) = \tfrac{1}{2}(z^2 + n\zeta^2 pq) + \tfrac{1}{6}\gamma_3(z^3 \sum a_k^3 + 3z^2\zeta(pq)^{\frac{1}{2}} + n\zeta^3(pq)^{\frac{3}{2}}) + \cdots ,$$

where $\gamma_3 = (q - p)(pq)^{-\frac{1}{2}}$, $\gamma_4 = (1 - 6pq)/pq, \cdots$ are the standardized cumulants of the binomial distribution. This series is majorized by the series

$$g(z, \zeta) = \tfrac{1}{2}(z^2 + n\zeta^2 pq) + \tfrac{1}{6}|\gamma_3|(z^3 b_n + z^2\zeta(pq)^{\frac{1}{2}} + n\zeta^3(pq)^{\frac{3}{2}}) + \cdots$$

obtained by replacing γ_r by $|\gamma_r|$ and $\sum a_k^r$ by b_n^{r-2}. $g(z, \zeta)$ is convergent for $|zb_n| < (pq)^{\frac{1}{2}}$ and $|\zeta| < 1$. Also the partial derivatives of $f^{(n)}(z, \zeta)$ with respect to z and ζ, denoted $f_z^{(n)}(z, \zeta)$ and $f_\zeta^{(n)}(z, \zeta)$, respectively, are majorized by the corresponding partial derivatives of $g(z, \zeta)$, which are also, of course, convergent for $|zb_n| < (pq)^{\frac{1}{2}}$ and $|\zeta| < 1$.

$f_\zeta^{(n)}(z, \zeta) = \sum K'(za_k(pq)^{-\frac{1}{2}} + \zeta)$ is an analytic function of z and ζ with a power series expansion in z and ζ about $(0, 0)$ and with the coefficient of ζ in this expansion equal to pq. Further, $f_\zeta^{(n)}(z, \zeta)$ is majorized by $g_\zeta(z, \zeta)$. So following the methods of Hille (1959, Volume I, pages 269–272), we see that the equation

$$(48) \qquad f_\zeta^{(n)}(z, \zeta) = 0$$

has a unique solution $\zeta = \zeta_n(z)$, which is an analytic function of z with a power series expansion majorized by the solution $\zeta = \gamma(z)$ of the equation

$$g_\zeta(z, \zeta) = 0 .$$

Now $\gamma(z)$ converges for $|zb_n| < c(pq)^{\frac{1}{2}}$, for some $c > 0$ not depending on n, and in this circle $|\gamma(z)| < 1$. So $\zeta_n(z)$ converges in this circle and is bounded by 1 there.

So $f_z^{(n)}[z, \zeta_n(z)]$ is an analytic function of z in the circle $|zb_n| < c(pq)^{\frac{1}{2}}$, with the coefficient of z in its power series expansion equal to 1 and it is majorized by $g_z[z, \gamma(z)]$ in this circle. Consider the equation

$$(49) \qquad t = f_z^{(n)}(z, \zeta_n(z)) = (pq)^{-\frac{1}{2}} \sum a_k K'[za_k(pq)^{-\frac{1}{2}} + \zeta_n(z)] .$$

This equation has a unique solution $z = z_n(t)$ for small enough $|t|$ and this solution is majorized by the solution $z = k(t)$ of the equation $t = g_z(z, \gamma(z))$. $k(t)$ converges and $|k(t)b_n| < c(pq)^{\frac{1}{2}}$, for $|t| < c'$, for some sufficiently small $c' > 0$, not depending on n. So $z_n(t)$ converges and $|z_n(t)b_n| < c(pq)^{\frac{1}{2}}$ in this circle with centre $t = 0$ and radius not depending on n. If

$$\lambda_n(t) = f^{(n)}(z_n(t), \zeta_n[z_n(t)]) - z_n(t)t + \tfrac{1}{2}t^2 , \tag{50}$$

then $\lambda_n(t)$ is uniformly convergent for $|t| < c'$.

For $z = u$, a real number, $\zeta_n(u)$, the solution of (48), is equal to $\alpha_n(u)$ defined in Lemma 1. Further, for x_n real, the equation

$$b_n x_n = b_n m_n(u) = (pq)^{-\frac{1}{2}} b_n \sum a_k K'(u a_k (pq)^{-\frac{1}{2}} + \alpha_n(u)) ,$$

has a unique real root $u_n = u_n(x_n b_n) = z_n(x_n b_n)$, the solution of (49), for $x_n b_n$ sufficiently small. Then $\lambda_n(t)$ in (50) is the same as $\lambda_n(t)$ in (45) for real t.

Now returning to (47), we have, after putting $x_n = m_n(u_n)$,

$$\begin{aligned} 1 - F_n(x_n) = {} & B_{ns}^{-1}(2\pi \sum K_k'') e^{\sum K_k - u_n x_n + \frac{1}{2}x_n^2}[1 - \Phi(x_n)] \\ & \times \left[\frac{\rho(u_n \sigma_n)}{\rho(x_n)} + \frac{O(b_n)}{\rho(x_n)}\right] \end{aligned} \tag{51}$$

where $\rho(v) = e^{\frac{1}{2}v^2}[1 - \Phi(v)]$. If $(pq)^{\frac{1}{2}}\alpha_n(u) = c_0 + c_1 u + c_2 u^2 + \cdots$, is defined by (6), then

$$\begin{aligned} 0 = (pq)^{-\frac{1}{2}} \sum K_k' = {} & n\alpha_n(pq)^{\frac{1}{2}} + \tfrac{1}{2}\gamma_3(u^2 + n\alpha_n^2 pq) \\ & + \tfrac{1}{6}\gamma_4[u^3 \sum a_k^3 + 3u^2\alpha_n(pq)^{\frac{1}{2}} + n\alpha_n^3(pq)^{\frac{3}{2}}] + \cdots . \end{aligned}$$

So $c_0 = c_1 = 0$ and

$$0 = u^2[\tfrac{1}{2}\gamma_3 + nc_2] + u^3[\tfrac{1}{6}\gamma_4 \sum a_k^3 + nc_3] + \cdots .$$

Thus $c_2 = -\frac{1}{2}\gamma_3/2n$, $c_3 = -\gamma_4 \sum a_k^3/6n$, $\cdots$. So

$$\alpha_n(u) = -\tfrac{1}{2}u^2\gamma_3/2n - \gamma_4 \sum a_k^3/6n + \cdots . \tag{52}$$

Also

$$\begin{aligned} x_n &= (pq)^{-\frac{1}{2}} \sum a_k K_k' \\ &= (pq)^{-\frac{1}{2}} \sum a_k K'[(pq)^{-\frac{1}{2}}\{u a_k - \gamma_3 u^2/2n - u^3\gamma_4 \sum a_k^3/6n + \cdots\}] \\ &= u + \tfrac{1}{2}u^2\gamma_3 \sum a_k^3 + \tfrac{1}{6}u_3(\gamma_4 \sum a_k^4 - 3\gamma_3^2/n) + \cdots . \end{aligned} \tag{53}$$

Inverting this power series gives

$$\begin{aligned} u_n = {} & x_n - \tfrac{1}{2}x_n^2\gamma_3 \sum a_k^3 \\ & + x_n^3[\tfrac{1}{2}\gamma_3^2(\sum a_k^3)^2 - \tfrac{1}{6}\gamma_4 \sum a_k^4 + \gamma_3^2/2n] + \cdots . \end{aligned} \tag{54}$$

Also if $\lambda_n(t)$ is defined as in (45), then

$$\begin{aligned} x_n^2\lambda_n(x_n b_n) &= \sum K_k - u_n x_n + \tfrac{1}{2}x_n^2 \\ &= x_n^3 b_n \lambda_{1n} + x_n^4 b_n^2 \lambda_{2n} + \cdots , \end{aligned}$$

where $\lambda_{1n}, \lambda_{2n}, \cdots$ are given in (46).

Now

$$\frac{1}{v} - \frac{1}{v^3} < \rho(v) < \frac{1}{v},$$

so

(55) $$x_n b_n < b_n/\rho(x_n) < x_n b_n(1 - x_n^{-2})^{-1}$$

and

(56) $$\frac{x_n}{u_n \sigma_n}\left(1 - \frac{1}{u_n^2\sigma_n^2}\right) < \frac{\rho(u_n\sigma_n)}{\rho(x_n)} < \frac{x_n}{u_n\sigma_n}\left(1 - \frac{1}{x_n^2}\right)^{-1}.$$

Also from (54), for $x_n b_n$ small,

$$u_n = x_n[1 + O(x_n b_n)],$$

so from (52)

$$\alpha_n(u_n) = O(x_n^2 b_n^2)$$

and so from the form of K'' given in (10),

$$\begin{aligned}\sigma_n^2 &= (pq)^{-1}[\sum a_k^2 K_k'' - (\sum a_k K_k'')^2/\sum K_k''] \\ &= 1 + O(x_n b_n).\end{aligned}$$

Thus

$$u_n \sigma_n = x_n[1 + O(x_n b_n)]$$

and so from (56)

(57) $$\frac{\rho(u_n\sigma_n)}{\rho(x_n)} = 1 + O(x_n b_n).$$

From Stirling's formula

$$B_{ns} = (2\pi n)^{\frac{1}{2}}[1 + O(n^{-\frac{1}{2}})],$$

so using this and the form of K'' given in (10),

(58) $$B_{ns}^{-1}(2\pi \sum K_k'')^{-\frac{1}{2}} = 1 + O(x_n b_n).$$

Now using (55), (57) and (58) in (51), we have

$$1 - F_n(x_n) = e^{x_n^2\lambda_n(x_n b_n)}[1 - \Phi(x_n)][1 + O(x_n b_n)],$$

where $x_n^2\lambda_n(x_n b_n) = \sum K_k - u_n x_n + \frac{1}{2}x_n^2$, as required.

4. Results for very large deviations. Some recent work summarized in Bahadur (1971) has been concerned with results for large deviation probabilities for $x = O(n^{\frac{1}{2}})$. In this section we will obtain results of this type for the distribution of X_n. Results of this type were obtained by Stone (1967, 1968, 1969) by different methods. Particular cases were also studied by Sievers (1969) and Hoadley (1967). We will derive the main result of Stone (1969) as a consequence of our result at the end of this section.

In this section we are interested in limit results rather than in approximations, so we will assume that the sequences $n^{\frac{1}{2}}b_n$ and p_n converge to $b > 0$ and λ with $0 < \lambda < 1$, respectively.

THEOREM 2. *Write* $u = t(npq)^{\frac{1}{2}}$ *and put*

$$h^{(n)}(t, \alpha) = n^{-1} \sum K(ta_k n^{\frac{1}{2}} + \alpha) .$$

Assume that $p_n \to \lambda$, *for* $0 < \lambda < 1$, *that* $b_n n^{\frac{1}{2}} \to b > 0$, *and that* $h^{(n)}(t, \alpha)$ *converges pointwise as a function of* t *and* α *to* $h(t, \alpha)$, *which we assume has continuous partial derivatives* $h_t(t, \alpha)$ *and* $h_\alpha(t, \alpha)$ *with respect to* t *and* α. *Then, for any constant* $C >$ 0 *and any sequence* $\{x_n\}_{n=1,2,\cdots}$ *such that* $x_n(pq/n)^{\frac{1}{2}} \to a \in \{h_t(t, \alpha(t)) : \alpha(t)$ *is the solution of* $h_\alpha(t, \alpha) = 0$, *for* $0 < t < C\}$,

$$\lim_{n\to\infty} n^{-1} \log [1 - F_n(x_n)] = h(\tau, \alpha(\tau)) - \tau h_t(\tau, \alpha(\tau)) , \tag{59}$$

where $\alpha(t)$ *is the unique solution of the equation*

$$h_\alpha(t, \alpha) = 0 \tag{60}$$

and τ *is the unique solution of the equation*

$$h_t(t, \alpha(t)) = a . \tag{61}$$

PROOF. For any $C > 0$, $\alpha = \alpha_n(t(npq)^{\frac{1}{2}})$, is the unique solution of $h_\alpha^{(n)}(t, \alpha) =$ 0, for $0 < t < C$, as obtained in Lemma 1. Also $h_{\alpha\alpha}^{(n)}(t, \alpha) > \delta > 0$, for $0 <$ $t < C$ and $|\alpha| < C$, where δ is a function of p only and is bounded away from 0 since p is, for large enough n. So, since $h_\alpha^{(n)}(t, \alpha)$ is increasing as a function of α for each t and since $h_\alpha(t, \alpha)$ exists, $h_\alpha^{(n)}(t, \alpha)$ converges to $h_\alpha(t, \alpha)$ and since $h_{\alpha\alpha}^{(n)}(t, \alpha) > \delta > 0$, for $|t| < C$ and $|\alpha| < C$, $h_\alpha(t, \alpha)$ is strictly increasing. So (60) has a unique solution, $\alpha = \alpha(t)$, and $\alpha_n(t(npq)^{\frac{1}{2}})$ converges to $\alpha(t)$.

From Lemma 1, the moment generating function of F_n is Q_n and so from (4), (7) and (58),

$$n^{-1} \log Q_n(u) = h^{(n)}[t, \alpha_n(t(npq)^{\frac{1}{2}})] + O(n^{-1}) . \tag{62}$$

Now the theorem of Plachky and Steinebach (1975), together with (62) and the fact that $h^{(n)}(t, \alpha_n[t(npq)^{\frac{1}{2}}]) \to h(t, \alpha(t))$, imply that (59) holds, where τ is the solution of equation (61).

Let G_n be the empirical distribution function with jumps of $1/n$ at points $n^{\frac{1}{2}}a_k$. Assume that G_n converges weakly to a distribution function G. Then

$$\begin{aligned} h^{(n)}(t, \alpha) &= n^{-1} \sum K(tn^{\frac{1}{2}}a_k + \alpha) \\ &\to \int \log (pe^{q(tx+\alpha)} + qe^{-p(tx+\alpha)})\, dG(x) \\ &= h(t, \alpha) , \end{aligned} \tag{63}$$

say. So

$$h_t(t, \alpha) = \int \frac{pqx(e^{q(tx+\alpha)} - e^{-p(tx+\alpha)})}{pe^{q(tx+\alpha)} + qe^{-p(tx+\alpha)}}\, dG(x) \tag{64}$$

and

$$h_\alpha(t, \alpha) = \int \frac{pq(e^{q(tx+\alpha)} - e^{-p(tx+\alpha)})}{pe^{q(tx+\alpha)} + qe^{-p(tx+\alpha)}}\, dG(x) . \tag{65}$$

Equating the right-hand side of (65) to 0, we can reduce the equation to

$$\lambda = \int \frac{dG(x)}{1 + Re^{-Hx}} \tag{66}$$

where $R = qe^{-\alpha}/p$ and $H = t$. Also equating the right-hand side of (64) to a, we can reduce the equation to

$$\int \frac{x\, dG(x)}{1 + Re^{-Hx}} = a\,, \tag{67}$$

after noticing that $\int x\, dG(x) = 0$. Further, (63) can be put in the form

$$\int \log\,(e^{Hx} + R)\, dG(x) + \log p + q\alpha\,. \tag{68}$$

So if $x_n(pq/n)^{\frac{1}{2}} \to a$ we have, from Theorem 2 and (68),

$$\lim_{n\to\infty} n^{-1} \log\,[1 - F_n(x_n)] = \int \log\,(e^{Hx} + R)\, dG(x) - Ha + (1-\lambda)\log R + \lambda \log \lambda + (1-\lambda)\log\,(1-\lambda)$$

where (R, H) is the solution of (66) and (67). This is just the result of Theorem 3.1 of Stone (1969).

5. Relationship to sampling with replacement. Let $V_1, \cdots, V_n$ be independently identically distributed random variables taking values 0 and 1 with probabilities p and q, respectively, and define

$$Y_n = (pq)^{-\frac{1}{2}} \sum V_k a_k\,.$$

Then Y_n and X_n are related as sums obtained from the same set of $a_1, \cdots, a_n$ by Poisson sampling and by simple random sampling, respectively. Some of the proofs of the asymptotic normality of X_n, for example Hájek (1960), used the fact that X_n and Y_n are asymptotically equivalent and Stone (1969) used a large deviation result for Y_n to develop the result discussed in Section 4. Y_n is the statistic used for one sample nonparametric tests. So a comparison of the limit results in the two cases is worthwhile.

The cumulant generating function of Y_n is $\sum K(ua_k(pq)^{\frac{1}{2}})$. If we define

$$m_n^* = m_n^*(u) = (pq)^{-\frac{1}{2}} \sum a_k K'(ua_k(pq)^{-\frac{1}{2}})\,,$$

then a result analogous to Theorem 1 holds, with

$$1 - F_n^*(x_n) = e^{x_n^2 \lambda_n^*(x_n b_n)}[1 - \Phi(x_n)][1 + O(x_n b_n)]$$

where F_n^* is the distribution function of Y_n and λ_n^* is defined by

$$t^2\lambda_n^*(t) = b_n^2[\sum K(ua_k(pq)^{-\frac{1}{2}}) - um_n^* + \tfrac{1}{2}m_n^{*2}]$$

and u is the unique real root of the equation

$$b_n m_n^*(u) = t\,.$$

Further, a result analogous to Theorem 2 will apply to the limit of $n^{-1}\log\,[1 -$

$F_n^*(x_n)]$, if for $t = u(npq)^{-\frac{1}{2}}$, $n^{-1} \sum K(ta_k n^{\frac{1}{2}})$ converges to $h(t)$, which is continuous and has a continuous derivative. In the particular case when $p = q$ and the a_k are symmetric, we see from the remark in Section 2, that $\alpha_n \equiv 0$, so $\lambda_n(t) \equiv \lambda_n^*(t)$ and $\alpha(t) \equiv 0$, so $h(t) = h(t, 0)$, where $\alpha(t)$ and $h(t, \alpha)$ are defined in Section 4. So in this case, the limit results are the same. This remarkable fact was noted for the case of very large deviations by Sievers (1969) and Stone (1967), in the special case of the Wilcoxon one and two sample statistics.

Acknowledgment. I wish to thank the referee for very detailed comments on the paper which have led to a number of improvements.

REFERENCES

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics.* Regional Conference Series in Applied Mathematics, **8**, SIAM.

BIKELIS, A. (1969). On the estimation of the remainder term in the central limit theorem for samples from finite populations (in Russian). *Studia Sci. Math. Hungar.* **4** 345–354.

COPSON, E. T. (1935). *An Introduction to the Theory of Functions of a Complex Variable.* Oxford Univ. Press, London.

ERDÖS, P. and RÉNYI, A. (1959). On the central limit theorem for samples from a finite population. *Publ. Math. Inst. Hungar. Acad. Sci.* **4** 49–61.

FELLER, W. (1966). *An Introduction to Probability Theory and its Applications,* **2**. Wiley, New York.

FELLER, W. (1969). Limit theorems for probabilities of large deviations. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** 1–20.

HÁJEK, J. (1960). Limiting distributions in simple random sampling for a finite population. *Publ. Math. Inst. Hungar. Acad. Sci.* **5** 361–374.

HILLE, E. (1959). *Analytic Function Theory.* Chelsea, New York.

HOADLEY, A. B. (1967). On the probability of large deviations of functions of several empirical cdf's. *Ann. Math. Statist.* **38** 360–381.

IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolter-Noordhoff, The Netherlands.

LOÈVE, M. (1955). *Probability Theory.* Van Nostrand, Princeton.

PETROV, V. V. (1975). *Sums of Independent Random Variables.* Springer-Verlage, Berlin.

PLACHKY, D. and STEINEBACH, J. (1975). A theorem about probabilities of large deviations with an application to queueing theory. *Period. Math. Hungar.* **6** 343–345.

SIEVERS, G. L. (1969). On the probability of large deviations and exact slopes. *Ann. Math. Statist.* **40** 1908–1921.

STONE, M. (1967). Extreme tail probabilities for the null distribution of the two-sample Wilcoxon statistic. *Biometrika* **54** 629–640.

STONE, M. (1968). Extreme tail probabilities for sampling without replacement and exact Bahadur efficiency of the two sample normal scores test. *Biometrika* **55** 371–375.

STONE, M. (1969). Approximations to extreme tail probabilities for sampling without replacement. *Proc. Cambridge Philos. Soc.* **66** 587–606.

DEPARTMENT OF MATHEMATICAL STATISTICS
UNIVERSITY OF SYDNEY
2006, NEW SOUTH WALES
AUSTRALIA