J. Appl. Prob. **24**, 875–887 (1987)
Printed in Israel

SADDLEPOINT EXPANSIONS FOR CONDITIONAL DISTRIBUTIONS

IB M. SKOVGAARD,* *Royal Veterinary and Agricultural University, Copenhagen*

Abstract

A saddlepoint expansion is given for conditional probabilities of the form $P\{\bar{Y} \geqq y \mid \bar{X} = x\}$ where $(\bar{X}, \bar{Y})$ is an average of n independent bivariate random vectors. A more general version, corresponding to the conditioning on a $p-1$-dimensional linear function of a p-dimensional variable is also included. A separate formula is given for the lattice case. The expansion is a generalization of the Lugannani and Rice (1980) formula, which reappears if $\bar{X}$ and $\bar{Y}$ are independent. As an example an approximation to the hypergeometric distribution is derived.

CONDITIONAL PROBABILITY; HYPERGEOMETRIC DISTRIBUTION; LATTICE VARIABLE; TAIL PROBABILITY; UNIFORM SADDLEPOINT EXPANSION

1. Introduction

The saddlepoint method used to drive asymptotic approximations to integrals of a certain type is known to give remarkably good approximations. In Daniels (1954) it was shown that this technique was applicable to the problem of approximating densities of sums of independent random variables. In fact, to apply the resulting approximation, it is necessary to know the cumulant generating function for the statistic, the density of which is to be approximated, whereas it is immaterial in this sense whether it is a sum of independent random variables. However, in the case of an average, $\bar{X}$ say, of n independent replications, it is known that the relative error to the density of $\bar{x}$ is $O(n^{-1})$ as n tends to ∞, uniformly for $\bar{x}$ in a bounded set. For comparison the Edgeworth expansions typically keep a relative error of order $O(n^{-k/2})$ uniformly only within sets for which $\bar{x}/\sqrt{n}$ grow slowly. In this sense the saddlepoint expansions are large-deviation expansions. Similar expansions were derived by a different method in Esscher (1932) for distribution functions, except that the expansion was in powers of $n^{-1/2}$. By employing a technique outlined in Bleistein (1966), Lugannani and Rice (1980) derived

Received 31 July 1986; revision received 9 September 1986.

* Postal address: Royal Veterinary and Agricultural University, Department of Mathematics, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark.

saddlepoint expansions for distribution functions of a one-dimensional statistic. A review of this and related methods is given in Daniels (1987).

For a conditional density it is straightforward to obtain a large-deviation expansion, simply by approximating the numerator and denominator separately by a saddlepoint expansion. This is the so-called double saddlepoint approximation, see Barndorff-Nielsen and Cox (1979). It is easy to see that this approximation keeps the same properties in terms of the relative error uniformly within sets of large deviations as do the single saddlepoint expansions. It is the purpose of the present paper to derive a saddlepoint expansion for the conditional distribution function of one coordinate of a random vector given the others. A solution to this problem is simple if the conditional cumulant generating function is tractable for further calculations, but this may not always be the case. However, this possibility should be kept in mind as a preferable method whenever feasible, because the expansion derived below is based on a saddlepoint expansion of a multivariate integral, which may be less accurate than for one-dimensional integrals. Also the possibility of combining the two methods is worth considering, i.e. to calculate the conditional cumulant generating function given some of the coordinates directly, and then apply the formula given below to the conditional probability given the remaining conditioning coordinates. The calculation of the expansion for the conditional distribution function requires knowledge of the cumulant generating function for the distribution for the entire vector random variable under study, and the solution of two saddlepoint equations, one for this vector random variable and one for the vector of conditioning coordinates. The expression is given in Section 2 for the expansion of $P\{\bar{Y} \geqq \tilde{y} \mid \bar{X} = \tilde{x}\}$ where $(\bar{X}, \bar{Y})$ is the average of n independent replications of bivariate random vectors, and for the p-dimensional case where we condition on a linear function of dimension $p-1$. We stick to the case of n independent replications to clarify the orders of terms, although the approximation may be used for other cases as well. Formally this is done merely by taking $n = 1$. We only state the first-order expansion; further terms may be obtained as described in Bleistein (1966), and they will be in orders of integer powers of n^{-1} relative to the main term uniformly in large deviation sets. Section 3 contains an outline of the proof based on inversion of characteristic functions. A brief sketch of another proof, which is somewhat simpler, is also given, because the method may be of some general interest. This consists of a saddlepoint expansion for the integral of Daniels' saddlepoint approximation to the density, and involves a technique from Bleistein (1966) for a saddlepoint near an endpoint of integration. The two methods lead to the same expansion, just as for unconditional probabilities where they both lead to the Lugannani and Rice expansion. However, the second method holds for the continuous case only, whereas the first one generalizes to the lattice case also, by modifications as in Daniels (1987), as is shown in Section 4. The second method of proof may be compared to the expansion obtained by

applying the method of Laplace to the integral of the conditional saddlepoint density approximation. This method yields an expansion in powers of $n^{-1/2}$, and for the unconditional case it was shown by Robinson (1982) that it has an asymptotic behaviour similar to the expansion in Esscher (1932). For further discussion of these methods, see Daniels (1987). Finally, in Section 5 we apply the method to an example, namely the hypergeometric distribution, for which the approximation turns out to give an excellent agreement with the exact values. Since it can hardly be considered a problem to obtain exact values for the hypergeometric distributions, except possibly for magnitudes for which the normal approximation gives satisfactory values, this distribution is included merely as an example. More important applications may be to conditional tests in exponential family models, or to other statistical problems of conditional inference.

2. The expansion for the conditional distribution

Consider first the bivariate continuous case. Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent identically distributed bivariate random variables with density $f(x, y)$ at (x, y) and cumulant generating function κ given by

$$\kappa(s, u) = \ln \int\int f(x, y)\exp\{sx + uy\}du\, ds, \qquad s, u \in \mathbb{C}, \tag{2.1}$$

which we shall assume exists in a neighbourhood of $(0, 0)$. Let $\bar{X} = n^{-1}\Sigma X_i$, $\bar{Y} = n^{-1}\Sigma Y_i$ and fix a point $(\tilde{x}, \tilde{y})$. We want to expand the conditional probability $P\{\bar{Y} \geqq \tilde{y} \mid \bar{X} = \tilde{x}\}$. Denote the derivatives of κ as follows:

$$\begin{gathered} \dot{\kappa}_s(s, u) = \frac{\partial}{\partial s}\kappa(s, u), \quad \dot{\kappa}_u(s, u) = \frac{\partial}{\partial u}\kappa(s, u), \\ \ddot{\kappa}_{ss}(s, u) = \frac{\partial^2}{\partial s^2}\kappa(s, u), \end{gathered} \tag{2.2}$$

etc., and let $\ddot{\kappa}$ denote the 2×2 matrix of second derivatives. We need the saddlepoint, $\tilde{s}_0$ say, corresponding to $\bar{X} = \tilde{x}$, defined by the saddlepoint equation

$$\dot{\kappa}_s(\tilde{s}_0, 0) = \tilde{x}, \qquad \tilde{s}_0 \in \mathbb{R} \tag{2.3}$$

and the bivariate saddlepoint, $(\tilde{s}, \tilde{u})$ say, corresponding to $(\bar{X}, \bar{Y}) = (\tilde{x}, \tilde{y})$, defined by

$$\dot{\kappa}_s(\tilde{s}, \tilde{u}) = \tilde{x}, \quad \dot{\kappa}_u(\tilde{s}, \tilde{u}) = \tilde{y}, \qquad \tilde{s}, \tilde{u} \in \mathbb{R}. \tag{2.4}$$

The approximation now becomes

$$P\{\bar{Y} \geqq \tilde{y} \mid \bar{X} = \tilde{x}\}$$

$$\sim 1 - \Phi(\sqrt{n}\tilde{w}) + \phi(\sqrt{n}\tilde{w})\{[\ddot{\kappa}_{ss}(\tilde{s}_0, 0)]^{1/2}/(\sqrt{n}\tilde{u}\,|\ddot{\kappa}(\tilde{s}, \tilde{u})|^{1/2}) - (\sqrt{n}\tilde{w})^{-1}\}, \tag{2.5}$$

where Φ and ϕ are the standard normal distribution and density functions respectively, $|\ddot{\kappa}(\tilde{s}, \tilde{u})|$ means the determinant of $\ddot{\kappa}(\tilde{s}, \tilde{u})$ and

$$\tilde{w} = \operatorname{sign}(\tilde{u})\{2[\tilde{s}\tilde{x} + \tilde{u}\tilde{y} - \kappa(\tilde{s}, \tilde{u})] - 2[\tilde{s}_0\tilde{x} - \kappa(\tilde{s}_0, 0)]\}^{1/2}. \tag{2.6}$$

The quantity $\sqrt{n}\tilde{w}$ is the signed square root of minus twice the log likelihood ratio statistic, based on the observation $(\bar{X}, \bar{Y}) = (\tilde{x}, \tilde{y})$, for the hypothesis $u = 0$ in the generated exponential family with densities $f(x, y)\exp\{sx + uy\}$ parametrized by (s, u). The error of approximation (2.5) is $O(n^{-1})$ relative to the main term as $n \to \infty$, uniformly for $(\tilde{x}, \tilde{y}) \in K$, where K is some fixed compact neighbourhood of $(E\bar{X}, E\bar{Y})$. As this set K we may take any compact neighbourhood of $(E\bar{X}, E\bar{Y})$ which is interior to the set of points $(\dot{\kappa}_s(s, u), \dot{\kappa}_u(s, u))$ for which $\kappa(s, u)$ exists.

If X_i and Y_i are independent the approximation reduces to the one obtained by Lugannani and Rice (1980).

The generalization of (2.5) to conditional probabilities of the form $P\{\bar{X}_p \geqq \tilde{x}_p \mid (\bar{X}_1, \cdots, \bar{X}_{p-1}) = (\tilde{x}_1, \cdots, \tilde{x}_{p-1})\}$ is quite trivial as well as its proof. Stated in somewhat more generality, although trivially reduced to the case mentioned above, the multivariate version takes the following form, which may be useful for applications. Let $X_1, \cdots, X_n$ be independent identically distributed random vectors in $\mathbb{R}^p$, let A be a $p \times (p-1)$ matrix of rank $p-1$ and $b \in \mathbb{R}^p$ a vector which is linearly independent of the columns of A. With $\bar{X} = n^{-1}\Sigma X_i$ and $\bar{Y} = b'\bar{X}$ we get the approximation analogues to (2.5),

$$P\{\bar{Y} \geqq y \mid A'\bar{X} = \tilde{a}\}$$
$$\sim 1 - \Phi(\sqrt{n}\tilde{w}) + \phi(\sqrt{n}\tilde{w})\{|A'\ddot{\kappa}(\tilde{t}_0)A|^{1/2}/(\sqrt{n}\tilde{u}\,|\ddot{\kappa}(\tilde{t})|^{1/2}|(A, b)|) - (\sqrt{n}\tilde{w})^{-1}\}, \tag{2.7}$$

where $\ddot{\kappa}$ is the $p \times p$ matrix of second derivatives of the cumulant generating function, (A, b) is the $p \times p$ matrix that equals A with b appended as the last column, and the saddlepoints $\tilde{t}_0 = A\tilde{s}_0$ and $\tilde{t} = A\tilde{s} + b\tilde{u}$, both in $\mathbb{R}^p$, are given by the equations

$$A'\dot{\kappa}(A\tilde{s}_0) = \tilde{a}, \quad \tilde{s}_0 \in \mathbb{R}^{p-1}, \tag{2.8}$$

where

$$\dot{\kappa}(t) = \frac{\partial}{\partial t}\kappa(t),$$

and

$$A'\dot{\kappa}(\tilde{t}) = \tilde{a}, \quad b'\dot{\kappa}(\tilde{t}) = \tilde{y}, \tag{2.9}$$

which is equivalent to $\dot{\kappa}(\tilde{t}) = \tilde{x}$, where $\tilde{x}$ is the unique point with $A'\tilde{x} = \tilde{a}$, $b'\tilde{x} = \tilde{y}$. Finally $\tilde{w}$ is

$$\tilde{w} = \operatorname{sign}(\tilde{u})\{2[\tilde{t}\tilde{x} - \kappa(\tilde{t})] - 2[\tilde{s}_0\tilde{a} - \kappa(\tilde{t}_0)]\}^{1/2}. \tag{2.10}$$

If, in particular the columns of (A, b) are orthonormal, the determinant $|(A, b)|$ is 1, and the approximation is recognized as a generalization of (2.5).

The above expressions are all for the case of continuous variables. We return to the lattice case in Section 4.

As in the Lugannani and Rice formula the singularity at $\tilde{u} = 0$ corresponding to $\tilde{w} = 0$ is removable, such that the expressions (2.5) and (2.7) are analytic throughout their domain and should be replaced by their limiting value at $\tilde{u} = 0$, which in the general form corresponding to (2.7) is

$$\begin{aligned} &P\{\bar{Y} \geqq b'\dot{\kappa}(A\tilde{s}_0) \,|\, A'\bar{X} = \tilde{a}\} \\ &\quad \sim \frac{1}{2} - \left(\frac{1}{2\pi n}\right)^{1/2} \{\tfrac{1}{6}\kappa^{(3)}(v^3)/(v'\ddot{\kappa}v)^{3/2} + \tfrac{1}{2}\operatorname{tr}\{A(A'\ddot{\kappa}A)^{-1}A'[\kappa^{(3)}(v)]\}/(v'\ddot{\kappa}v)^{1/2}\}, \end{aligned} \tag{2.11}$$

where the derivatives $\ddot{\kappa}$ and $\kappa^{(3)}$ are evaluated at $\tilde{t}_0 = A\tilde{s}_0$, $v \in \mathbb{R}^p$ is the vector $b - A(A'\ddot{\kappa}A)^{-1}A'\ddot{\kappa}b$, while $\kappa^{(3)}(v^3)$ is the three-fold product $\Sigma\Sigma\Sigma\kappa^{(3)}_{ijk}v_iv_jv_k$, where (v_i) and $\kappa^{(3)}_{ijk}$ are the coordinates of v and $\kappa^{(3)}$, respectively, and finally $\kappa^{(3)}(v)$ is the $p \times p$ matrix with (i, j)th coordinate $\Sigma\kappa^{(3)}_{ijk}v_k$. If $A'\bar{X}$ and $b'\bar{X}$ are independent the expression reduces to $\frac{1}{2} - \frac{1}{6}\lambda_3(2\pi n)^{-1/2}$, where λ_3 is the standardized third cumulant of $b'X_i$, in agreement with the Lugannani and Rice formula. The expression (2.11) is identical to the one obtained by the mixed Edgeworth saddlepoint approximation at this particular value, see Barndorff-Nielsen and Cox (1979); just as the Lugannani and Rice formula reduces to the Edgeworth expansion when evaluated at the mean.

It is often convenient for applications to restate the expression (2.7) in terms of the line of support of the conditional distribution. Thus let $c \in \mathbb{R}^p$ be a non-zero vector satisfying $A'c = 0$. Then, if we let $\tilde{x}$ denote the point in $\mathbb{R}^p$ such that $A'\tilde{x} = \tilde{a}$ and $b'\tilde{x} = \tilde{y}$, the conditional distribution of $\bar{X}$ is located on $\tilde{x} + rc$, $r \in \mathbb{R}$, and we may write $\bar{X} = \tilde{x} + \bar{R}c$, where $\bar{R} = (\bar{Y} - \tilde{y})/(b'c)$. Now approximation (2.7) becomes

$$\begin{aligned} &P\{\bar{R} \geqq 0 \,|\, A\bar{X} = A\tilde{x}\} \sim 1 - \Phi(\sqrt{n}\tilde{w}) \\ &\quad + \phi(\sqrt{n}\tilde{w})\{|\ddot{\kappa}(\tilde{t}_0)|^{1/2}(c'\ddot{\kappa}^{-1}(\tilde{t}_0)c)^{1/2}/(\sqrt{n}(c'\tilde{t})|\ddot{\kappa}(\tilde{t})|^{1/2}) - (\sqrt{n}\tilde{w})^{-1}\}, \end{aligned} \tag{2.12}$$

where the sign of w should be redefined to equal that of $(c'\tilde{t})$, to take into account the possible change of 'direction' if $b'c < 0$. The saddlepoint equation (2.8) for $\tilde{t}_0$ is in this setting more conveniently stated as the solution to $A'\dot{\kappa}(\tilde{t}_0) = A\tilde{x}$ with the restriction $c'\tilde{t}_0 = 0$, while the saddlepoint $\tilde{t}$ from (2.9) is

simply given by $\dot{\kappa}(\tilde{t}) = \tilde{x}$. The limit of (2.12) at $c'\hat{t} = 0$ is identical to (2.11), except that v should be replaced by $(\ddot{\kappa}(\tilde{t}_0))^{-1}c$.

3. Conditions and proofs

We shall prove the expansion for the bivariate case for which the method is more transparent and, except for notation, practically identical to the one for the multivariate case. The following three conditions are required for the validity of the proof.

Conditions.

(I) Some positive power of the characteristic function for (X_i, Y_i) is integrable.

(II) The cumulant generating function κ for (X_i, Y_i) exists in a neighbourhood of $(0, 0)$.

(III) The variance matrix $\ddot{\kappa}(0, 0)$ is non-singular.

The last condition obviously causes no loss of generality, because we may restrict attention to a subspace if it does not hold.

In the proof below we shall outline the method but not pay much attention to the error term or to the possibility of deriving higher order expansions. That the error is of the form described in Section 2 follows from the methods being used, for which we refer to Daniels (1954), Bleistein (1966) and, in particular, Olver (1974), Chapters 3–4.

Proof. The inversion formula for the density $\tilde{f}$ of $(\tilde{X}, \tilde{Y})$ is

$$\tilde{f}(\tilde{x}, \tilde{y}) = \left(\frac{n}{2\pi i}\right)^2 \int_{-i\infty}^{i\infty} \int_{-i\infty}^{i\infty} \exp\{n[\kappa(s, u) - s\tilde{x} - u\tilde{y}]\} ds\, du \tag{3.1}$$

from which we obtain the integral, $Q(\tilde{y} \mid \tilde{x})$ say, given by

$$\begin{aligned} Q(\tilde{y} \mid \tilde{x}) &= \int_{\tilde{y}}^{\infty} f(\tilde{x}, \bar{y}) d\bar{y} \\ &= \left(\frac{n}{2\pi i}\right)^2 \int_{c-i\infty}^{c+i\infty} \int_{-i\infty}^{i\infty} \exp\{n[\kappa(s, u) - s\tilde{x} - u\tilde{y}]\} ds \frac{du}{nu}, \end{aligned} \tag{3.2}$$

where $c > 0$ indicates a transformation of the path of integration to avoid the singularity at $u = 0$. In the case $c < 0$ we should subtract a certain quantity from the left-hand side, but for simplicity we stick to the case $c > 0$ in the sequel. For the integral in s in (3.2) for fixed u we now use a standard saddlepoint approximation, exactly as for the usual approximation to densities as in Daniels (1954). The saddlepoint, $\tilde{s}_u$ say, is defined by the equation

$$\dot{\kappa}_s(\tilde{s}_u, u) = \tilde{x}, \quad \tilde{s}_u \in \mathbb{R}, \tag{3.3}$$

and the approximation to (3.2) becomes

$$Q(\tilde{y} \mid \tilde{x}) \sim \left(\frac{n}{2\pi}\right)^{1/2} \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \exp\{nh(u)\}(\ddot{\kappa}_{ss}(\tilde{s}_u, u))^{-1/2} \frac{du}{u}, \tag{3.4}$$

where

$$h(u) = \kappa(\tilde{s}_u, u) - \tilde{s}_u\tilde{x} - u\tilde{y}. \tag{3.5}$$

By differentiation of (3.3) with respect to u we see that $d\tilde{s}_u/du = -(\ddot{\kappa}_{ss}(\tilde{s}_u, u))^{-1}\ddot{\kappa}_{su}(\tilde{s}_u, u)$, and hence that the first two derivatives of h are

$$\dot{h}(u) = \dot{\kappa}_u(\tilde{s}_u, u) - \tilde{y}, \tag{3.6}$$

$$\ddot{h}(u) = \ddot{\kappa}_{uu}(\tilde{s}_u, u) - (\ddot{\kappa}_{su}(\tilde{s}_u, u))^2/\ddot{\kappa}_{ss}(\tilde{s}_u, u). \tag{3.7}$$

To approximate (3.4) we use another saddlepoint expansion, modified as described in Bleistein (1966) to deal with the singularity at $u = 0$, and, in fact, quite similar to the one used to obtain the Lugannani and Rice formula; see Daniels (1986). The saddlepoint $\tilde{u} \in \mathbb{R}$ is obtained by equating (3.6) to zero, which leads to the equations (2.4). We introduce the new variable w by the equations

$$\tfrac{1}{2}(w - \tilde{w})^2 = h(u) - h(\tilde{u}), \tag{3.8}$$

$$\tfrac{1}{2}\tilde{w}^2 = h(0) - h(\tilde{u}), \quad \operatorname{sign}(\tilde{w}) = \operatorname{sign}(\tilde{u}), \tag{3.9}$$

such that $w = 0$ when $u = 0$, and $w = \tilde{w}$ corresponds to $u = \tilde{u}$. We now choose $c = \tilde{u}$ in (3.4), assume $\tilde{u} > 0$ to avoid technicalities, and rewrite (3.4) as

$$Q(\tilde{y} \mid \tilde{x}) \sim \left(\frac{n}{2\pi}\right)^{1/2} \exp\{nh(\tilde{u})\} \frac{1}{2\pi i} \int_{\tilde{w}-i\infty}^{\tilde{w}+i\infty} \exp\left\{\frac{n}{2}(w - \tilde{w})^2\right\} g(w) \frac{dw}{w}, \tag{3.10}$$

where

$$g(w) = \frac{du}{dw}\frac{w}{u}\{\ddot{\kappa}_{ss}(\tilde{s}_u, u)\}^{-1/2} \tag{3.11}$$

is analytic with a removable singularity at $w = 0$. In fact, we have

$$g(0) = \{\ddot{\kappa}_{ss}(\tilde{s}_0, 0)\}^{-1/2}, \quad g(\tilde{w}) = \tilde{w}\tilde{u}^{-1}|\ddot{\kappa}(\tilde{s}, \tilde{u})|^{-1/2}, \tag{3.12}$$

where we have derived the value of dw/du at $u = \tilde{u}$ by differentiating (3.8) twice with respect to u, which gives

$$(w - \tilde{w})\frac{d^2w}{du^2} + \left(\frac{dw}{du}\right)^2 = \ddot{h}(u), \tag{3.13}$$

and the calculation is finished by insertion of (3.7) and taking $u = \tilde{u}$, $w = \tilde{w}$. Now, following Bleistein (1966), the integral in (3.10) is approximated by the integral obtained by replacing $g(w)$ with the linear function $g(0) + w\tilde{w}^{-1}(g(\tilde{w}) - g(0))$, which agrees with g at the two 'critical points'

$w = 0$ and $w = \hat{w}$. Bleistein showed by a partial integration that the remainder then is an integral of the same form, but with a factor n^{-1}. Thus (3.10) is approximated by

$$
\begin{aligned}
Q(\tilde{y} \mid \tilde{x}) \sim & \left(\frac{n}{2\pi}\right)^{1/2} g(0)\exp\{nh(0)\}\frac{1}{2\pi i}\exp\left\{-\frac{n}{2}\hat{w}^2\right\} \\
& \times \int_{\hat{w}-i\infty}^{\hat{w}+i\infty} \exp\left\{\frac{n}{2}(w-\hat{w})^2\right\}\{w^{-1}+\hat{w}^{-1}(g(\hat{w})/g(0)-1)\}dw \\
= & \left(\frac{n}{2\pi}\right)^{1/2} \{\ddot{\kappa}_{ss}(\tilde{s}_0, 0)\}^{-1/2}\exp\{-n(\tilde{s}_0\tilde{x}-\kappa(\tilde{s}_0, 0))\} \\
& \times [1-\Phi(\sqrt{n}\hat{w})+\phi(\sqrt{n}\hat{w})(\sqrt{n}\hat{w})^{-1}(g(\hat{w})/g(0)-1)],
\end{aligned}
\tag{3.14}
$$

where the final expression is noticed to be valid also when $\hat{u} < 0$, because to the order considered the same quantity should then be subtracted from this as from $Q(\tilde{y} \mid \tilde{x})$. The factor in front of the square brackets in the last expression is exactly the first-order saddlepoint approximation to the marginal density of $\tilde{X}$ at $\tilde{x}$. Hence the result follows by insertion.

A second method of proof starts from the first-order saddlepoint approximation to the bivariate density

$$
\tilde{f}(\tilde{x}, \tilde{y}) \sim \frac{n}{2\pi}|\ddot{\kappa}(\tilde{s}_y, \tilde{u}_y)|^{-1/2}\exp\{-n[\tilde{s}_y\tilde{x}+\tilde{u}_y\tilde{y}-\kappa(\tilde{s}_y, \tilde{u}_y)]\}, \tag{3.15}
$$

where $(\tilde{s}_y, \tilde{u}_y)$ is the bivariate saddlepoint corresponding to $(\tilde{x}, \tilde{y})$. To obtain an approximation to the integral $Q(\tilde{y} \mid \tilde{x})$ of $f(\tilde{x}, \tilde{y})$ from $\tilde{y}$ to ∞, we use a saddlepoint approximation to the integral of the right side of (3.15), modified as described in Bleistein (1966), Section 5 or Temme (1982), to take into account the endpoint of integration which may be close to the saddlepoint. This method appears a bit easier than the one given in the proof above, but it does not easily generalize to the lattice case as does the other method by slight modifications to be discussed in Section 4. To obtain higher-order terms, in the continuous case, by the second method based on integration of (3.15), we have to include higher-order terms in the expansion of $\tilde{f}(\tilde{x}, \tilde{y})$ as well as in the expansion of the integral. For such expansions it is probably easier to use the result in Temme (1982), which is adapted exactly to integrals of this type.

4. The lattice case

With modifications quite analogous to those described in Daniels (1987) for the Lugannani and Rice formula, the proof given in Section 3 carries over to the case when Y_i is a lattice random variable. The conditioning variable X_i may be continuous or lattice, but for the moment consider the bivariate case

corresponding to Formula (2.5), except that we assume that $\mathbb{Z}^2$ is a minimal lattice for (X_i, Y_i). Let $n\tilde{x}$ and $n\tilde{y}$ be integers, then Formula (2.5) may be replaced by

$$P\{\bar{Y} \geqq \tilde{y} \mid \bar{X} = \tilde{x}\} \sim 1 - \Phi(\sqrt{n}\tilde{w}) + \phi(\sqrt{n}\tilde{w})\{[\ddot{\kappa}_{ss}(\tilde{s}_0, 0)]^{1/2}/(\sqrt{n}(1 - \exp(-\tilde{u}))|\ddot{\kappa}(\tilde{s}, \tilde{u})|^{1/2}) - (\sqrt{n}\tilde{w})^{-1}\}, \tag{4.1}$$

where the quantities $\tilde{w}$, $\tilde{s}_0$ and $(\tilde{s}, \tilde{u})$ are still as given in (2.3), (2.4) and (2.6). It may be preferable, for reasons discussed in Daniels (1987) to introduce a continuity correction and hence define

$$\hat{y} = \tilde{y} - \tfrac{1}{2}n^{-1}, \tag{4.2}$$

and the corresponding bivariate saddlepoint $(\hat{s}, \hat{u})$ by Equations (2.4) with $\tilde{y}$ replaced by $\hat{y}$. Furthermore $\hat{w}$ is a redefinition of $\tilde{w}$ in (2.6) with $\tilde{s}$, $\tilde{u}$ and $\tilde{y}$ replaced by $\hat{s}$, $\hat{u}$ and $\hat{y}$, respectively, whereas $\tilde{s}_0$ is unchanged. With this continuity correction, Formula (4.1) is replaced by

$$P\{\bar{Y} \geqq \tilde{y} \mid \bar{X} = \tilde{x}\} \sim 1 - \Phi(\sqrt{n}\hat{w}) + \phi(\sqrt{n}\hat{w})\{[\ddot{\kappa}_{ss}(\tilde{s}_0, 0)]^{1/2}/(\sqrt{n}\,2\sinh(\tfrac{1}{2}\hat{u})|\ddot{\kappa}(\hat{s}, \hat{u})|^{1/2}) - (\sqrt{n}\hat{w})^{-1}\}. \tag{4.3}$$

The expressions (4.1) and (4.3) are unchanged for the case when X_i is continuous, but Y_i still lattice. The asymptotic properties of the approximations (4.1) and (4.3) are the same as for the continuous case, i.e. the relative error is still $O(n^{-1})$ uniformly for $(\tilde{x}, \tilde{y})$ in a compact set. Also the conditions for validity, given in Section 3, are still the same except that Condition I is replaced by the condition that the lattice considered is a minimal lattice for the distribution in question. If there is a continuous component, such as X_i when X_i is continuous and Y_i lattice, this must still fulfil Condition I.

Concerning the proof the only change, except for minor trivial changes in the inversion formula, stems from the replacement in (3.2) of the integral from $\tilde{y}$ to ∞ by a sum, where ny runs from $n\tilde{y}$ to ∞. This summation results in the factor $(1 - \exp(-u))^{-1}$ instead of $(nu)^{-1}$ on the right side of (3.2), and this change corresponds exactly to the change from the approximation in (2.5) to the one in (4.1). Details concerning the continuity correction are found in Daniels (1987). A further advantage of the continuity correction is that it takes us 'away from the boundary' of the support, where the saddlepoint is undefined, except when the desired probability is trivially equal to 1.

Let us now turn to the general formulation where $\bar{X}$ is a p-dimensional random vector, and we consider the conditional distribution of $\bar{Y} = b'\bar{X}$ given $A'\bar{X}$ as in Formula (2.7). We assume that the conditional distribution of $\bar{Y}$ given $A'\bar{X} = \tilde{a}$ is a lattice distribution and that b has been scaled such that $n\bar{Y}$

has minimal lattice $y_0 + \mathbb{Z}$, where y_0 is some constant that may depend on $\tilde{a}$. If $\tilde{y}$ is a lattice point in this conditional distribution of $\bar{Y}$, the saddlepoint approximation corresponding to (4.3) now becomes

$$P\{\bar{Y} \geqq \tilde{y} \mid A'\bar{X} = \tilde{a}\} \sim 1 - \Phi(\sqrt{n}\hat{w}) + \phi(\sqrt{n}\hat{w})\{|A'\ddot{\kappa}(\tilde{t}_0)A|^{1/2}/(\sqrt{n}\, 2\sinh(\tfrac{1}{2}\hat{u})|\ddot{\kappa}(\hat{t})|^{1/2}|(A,\ b)|) - (\sqrt{n}\hat{w})^{-1}\} \tag{4.4}$$

with notation as in Formula (2.7), except for the continuity correction $\hat{y} = \tilde{y} - \frac{1}{2}n^{-1}$ and corresponding changes in the definitions (2.9) and (2.10) leading to $\hat{t}$ and $\hat{w}$ in place of $\tilde{t}$ and $\tilde{w}$, while $\hat{u}$ is still given by the relation $\hat{t} = A\hat{s} + b\hat{u}$. Whether $A'\bar{X}$ is lattice or continuous or a mixture of the two is immaterial for the approximation (4.4). The setup considered here trivially covers the case when $X_i \in \mathbb{Z}^p$ and we condition on a $(p-1)$-dimensional linear function of $\bar{X}$, in which case the remaining component will be either degenerate or lattice. The limit of the expression in (4.4) as $\hat{u} \to 0$ is identical to the expression (2.11). This is not so for the uncorrected version corresponding to (4.1).

As for the continuous case the analogue of version (2.12) of the formula may be useful. Thus, let $c \in \mathbb{R}^p$ satisfy $A'c = 0$, and let us further require that $b'c = 1$, such that $n\bar{X}$ may be written as $n\bar{X} = n\tilde{x} + n\bar{Z}c$, with $\mathbb{Z}$ as minimal lattice for $n\bar{Z} = n(\bar{Y} - \tilde{y})$. The continuity correction included in (4.4) becomes $\hat{x} = \tilde{x} - \frac{1}{2}n^{-1}c$, and the two saddlepoints are given by the equations $\dot{\kappa}(\hat{t}) = \hat{x}$ and $A\dot{\kappa}(\tilde{t}_0) = A\tilde{x}$ with the restriction $c'\tilde{t}_0 = 0$. Then Formula (4.4) takes the form

$$P\{\bar{Z} \geqq 0 \mid A\bar{X} = A\tilde{x}\} \sim 1 - \Phi(\sqrt{n}\hat{w}) + \phi(\sqrt{n}\hat{w})\{|\ddot{\kappa}(\tilde{t}_0)|^{1/2}(c'\ddot{\kappa}^{-1}(\tilde{t}_0)c)^{1/2}/(\sqrt{n}\, 2\sinh(\tfrac{1}{2}c'\hat{t})|\ddot{\kappa}(\hat{t})|^{1/2}) - (\sqrt{n}\hat{w})^{-1}\}, \tag{4.5}$$

where $\hat{w}$ is the continuity corrected form of (2.10), i.e.

$$\hat{w} = \operatorname{sign}(c'\hat{t})\{2[\hat{t}\hat{x} - \kappa(\hat{t})] - 2[\tilde{t}_0\tilde{x} - \kappa(\tilde{t}_0)]\}^{1/2}. \tag{4.6}$$

The corresponding restatement of the uncorrected version corresponding to (4.1) is apparent from the above expressions.

5. Example: The hypergeometric distribution

The hypergeometric distribution is a one-dimensional distribution with distribution function

$$H_{n,r,N}(k) = \sum_{j=m}^{k} \binom{r}{j}\binom{N-r}{n-j}\Big/\binom{N}{n}, \qquad m = \max\{0, n + r - N\} \tag{5.1}$$

where $k \leqq r, n$ and $r, n \leqq N$ and all numbers are non-negative integers. An

attempt to apply the Lugannani and Rice formula leads to a complicated equation involving the hypergeometric function. The distribution may, however, be represented as a conditional distribution constructed from four independent Poisson random variables, and hence we may apply the result in Section 4, more specifically we shall use the continuity corrected version (4.5).

Let (X_{ij}), $i, j = 1, 2$, be four independent Poisson random variables with $EX_{ij} = 1$. Considered as a two by two table the marginal totals are $X_{i\cdot} = X_{i1} + X_{i2}$ and $X_{\cdot j} = X_{1j} + X_{2j}$, while $X_{\cdot\cdot} = \Sigma\Sigma X_{ij}$ is the total. The conditional distribution of X_{11} given the marginal totals is of the form (5.1) with $N = X_{\cdot\cdot}$, $n = X_{1\cdot}$ and $r = X_{\cdot 1}$. Thus we consider this conditional distribution, or equivalently the conditional distribution of the vector $X = (X_{11}, X_{12}, X_{21}, X_{22})'$ given the marginal totals.

Fix a point $(\tilde{x}_{ij})$ and let $\tilde{x}_{i\cdot}$, $\tilde{x}_{\cdot j}$, $\tilde{x}_{\cdot\cdot}$ denote the corresponding margins, and consider the conditional probability of $\{X_{11} \geqq \tilde{x}_{11}\}$ given these. The conditional distribution of X is concentrated on the set $\{\tilde{x} + zc; z \in \mathbb{Z}\}$, where c is the vector $(1, -1, -1, 1)'$, and the set is a minimal lattice for the distribution. Notice that in the notation we have chosen, the number of replications from the previous sections is $n = 1$. The continuity correction replaces $\tilde{x} = (\tilde{x}_{11}, \tilde{x}_{12}, \tilde{x}_{21}, \tilde{x}_{22})'$ by the vector $\hat{x} = \tilde{x} - \frac{1}{2}c$. The cumulant generating function and its first two derivatives are

$$\begin{aligned} \kappa(t) &= \sum_{i,j} (\lambda_{ij} - 1), \quad \lambda_{ij} = \exp(t_{ij}), \\ \dot{\kappa}(t) &= (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})', \\ \ddot{\kappa}(t) &= \operatorname{diag}(\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}), \end{aligned} \tag{5.2}$$

where $t = (t_{11}, t_{12}, t_{21}, t_{22})'$. Thus the saddlepoint $\hat{t}$, cf. (4.5), has coordinates $\hat{t}_{ij} = \ln \hat{\lambda}_{ij}$, say, where $\hat{\lambda}_{ij} = \hat{x}_{ij}$. The other saddlepoint $\tilde{t}_0$ with coordinates $\ln \tilde{\lambda}_{ij}$, say, must satisfy $c'\tilde{t}_0 = 0$ which is equivalent to $\tilde{\lambda}_{11}\tilde{\lambda}_{22}/(\tilde{\lambda}_{12}\tilde{\lambda}_{21}) = 1$, and it is easily seen that the solution is

$$\tilde{\lambda}_{ij} = \tilde{x}_{i\cdot}\tilde{x}_{\cdot j}/\tilde{x}_{\cdot\cdot}, \tag{5.3}$$

which in statistical jargon is 'the expected value' of X_{ij} in the model of independence between rows and columns. It is now a matter of insertion to obtain the expression corresponding to (4.5) for which we note that

$$\begin{aligned} &|\ddot{\kappa}(\hat{t}_0)|^{1/2} = (\Pi\tilde{x}_{i\cdot})(\Pi\tilde{x}_{\cdot j})/\tilde{x}_{\cdot\cdot}^2 \\ &|\ddot{\kappa}(\hat{t})|^{1/2} = \left(\prod_{i,j} \hat{x}_{ij}\right)^{1/2}, \\ &c'\hat{t} = \ln(\hat{x}_{11}\hat{x}_{22}/(\hat{x}_{12}\hat{x}_{21})) \\ &(c'\ddot{\kappa}^{-1}(\hat{t}_0)c)^{1/2} = \left(\sum_{i,j} \tilde{\lambda}_{ij}^{-1}\right)^{1/2}, \\ &\hat{w} = \operatorname{sign}(c'\hat{t})\left\{2\left[\sum_{i,j} \hat{x}_{ij} \ln \hat{x}_{ij} - \sum_i \tilde{x}_{i\cdot} \ln \tilde{x}_{i\cdot} - \sum_j \tilde{x}_{\cdot j} \ln \tilde{x}_{\cdot j} + \tilde{x}_{\cdot\cdot} \ln \tilde{x}_{\cdot\cdot}\right]\right\}^{1/2}, \end{aligned} \tag{5.4}$$

which, inserted in (4.5), give the approximation

$$P\{X_{11} \geqq \tilde{x}_{11} \mid X_{1\cdot} = \tilde{x}_{1\cdot}, X_{2\cdot} = \tilde{x}_{2\cdot}, X_{\cdot 1} = \tilde{x}_{\cdot 1}\}$$
$$\sim 1 - \Phi(\hat{w}) + \phi(\hat{w}) \left\{ \left(\frac{\tilde{x}_{1\cdot}\tilde{x}_{2\cdot}\tilde{x}_{\cdot 1}\tilde{x}_{\cdot 2}}{\tilde{x}_{\cdot\cdot}\hat{x}_{11}\hat{x}_{12}\hat{x}_{21}\hat{x}_{22}} \right)^{1/2} (2 \sinh(\tfrac{1}{2}c'\hat{t}))^{-1} + \hat{w}^{-1} \right\}. \tag{5.5}$$

The approximation is only undefined when one of the marginal totals is 0 in which case the probability is 1. The quantity $\hat{w}$ is the signed square root of minus twice the log likelihood ratio statistic for the test of independence in the 2×2 table, except that it has been calculated with a continuity correction.

TABLE 1
Hypergeometric distribution. Approximation (5.5) to $P\{X_{11} \geqq x_{11} \mid x_{1\cdot}, x_{2\cdot}, x_{\cdot 1}\}$ compared to the exact value, for four independent Poisson variables (X_{ij}) with the same mean

x_{11} / x_{21}	x_{12} / x_{22}	Exact probability	Approximation (5.5)	Relative error (%)
85	5			
75	35	1.505×10^{-6}	1.501×10^{-6}	−0.3
14	6			
8	12	0.05548	0.05541	−0.1
5	3			
1	9	0.03167	0.03127	−1.3
5	1			
1	5	0.04004	0.03929	−1.9
6	0			
0	6	1.082×10^{-3}	0.976×10^{-3}	−9.9

Some numerical examples of the approximation (4.5) to $P\{X_{11} \geqq x_{11} \mid x_{1\cdot}, x_{2\cdot}, x_{\cdot 1}\}$ are given in Table 1. It is seen that for these examples the relative error never exceeds 10%. Because the example is merely included to give an impression of the quality of the approximation, it is not compared to alternative approximations. An extensive study of various kinds of normal, binomial and Poisson approximations to the hypergeometric distribution, is given in Molenaar (1970). If approximations are needed, however, it is clearly an advantage to have a single approximation that works well over the major range of parameters and distributions.

References

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

BLEISTEIN, N. (1966) Uniform asymptotic expansions of integrals with stationary point near algebraic singularity. *Comm. Pure Appl. Math.* **19**, 353–370.

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

DANIELS, H. E. (1987) Tail probability approximations. *Internat Statist. Rev.* **55**, 37–48.

ESSCHER, F. (1932) On the probability function in the collective theory of risk. *Skand. Akt. Tidsskr.* **15**, 175–195.

LUGANNANI, R. AND RICE, S. O. (1980) Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Prob.* **12**, 475–490.

MOLENAAR, W. (1970) *Approximations to the Poisson, Binomial and Hypergeometric Distribution Functions*. Mathematical Centre Tracts, No. 31. Mathematisch Centrum, Amsterdam.

OLVER, F. W. J. (1974) *Asymptotics and Special Functions*. Academic Press, New York.

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

TEMME, N. M. (1982) The uniform asymptotic expansion of integrals related to cumulative distribution functions. *Siam J. Math. Anal.* **13**, 239–253.