On the error rate for asymptotic chi-squared distributions in the lattice case

BY J. ROUSSEAU AND I. M. SKOVGAARD

University of Paris V and CREST and The Royal Veterinary and Agricultural University

In this paper we study the order of error of the Edgeworth approximation to the distribution of statistics that allow for an expansion with a quadratic leading term, in the case of independent identically distributed lattice random variables. Matthes (1975) proves that the error $O(n^{-p/(p+1)})$ is obtained for smooth and bounded convex sets. We extend his result to cover likelihood ratio tests and other asymptotically quadratic statistics, essentially allowing for nuisance parameters and for non-sufficiency of the score statistic. In mathematical terms the extension is to sets that are only conditionally convex and bounded. Thus, we allow the quadratic leading term in the asymptotic expansion of the statistic of interest not to be positive definite and higher order terms to depend on components that are not in the quadratic form. Like Matthes (1975), we obtain the order $O(n^{-p/(p+1)})$ for the error term.

1. Introduction and main result. Asymptotic expansions of probabilities of sets for lattice random variables are difficult to obtain because of the jumps in the distribution. In one dimension the jumps in the distribution have probabilities

Received

AMS 2000 subject classifications. Primary 62E20; secondary 60E10, 11H06, 11P21

Key words and phrases. chi-squared approximation, Edgeworth expansion, lattice distribution, likelihood ratio, order of error

of order $1/\sqrt{n}$, and continuous approximations ignoring the precise location of the lattice points will only be correct to this order of error. For this situation it may be feasible to apply some kind of continuity correction improving the order, but only because the position of the lattice points relative to the boundary of the set may be identified. If the dimension is higher than one even a halfspace, bounded by a hyperplane, is difficult to handle. If the hyperplane follows the lattice directions the situation is as in the one-dimensional case, but in other cases the order of the error is unknown. For sets with curved boundary the situation may, perhaps surprisingly, be slightly better, essentially because the boundary of the set cannot hit quite as many lattice points as in the linear case.

The chi-squared approximation to the distribution of the likelihood ratio test statistic in a parametric statistical model is a primary example of a situation with an approximately elliptical boundary. If the set is expanded beyond the first order the set changes slightly in the leading variable, but more awkwardly further variables appear in the expansion, namely derivatives of the log likelihood function with respect to the nuisance parameters, and derivatives of higher order. The chi-squared approximation is known to be correct apart from an error of order $O(n^{-1})$ in well-behaved continuous models for independent replications. When the data are discrete the situation is much more complicated, with error ranging from $O(n^{-1/2})$ in the simplest one-dimensional lattice cases to the same order as for the continuous case in the very best situations of higher dimensional tests, see Götze (2000).

Let p denote the degrees of freedom for the chi-squared test, or equivalently the dimension of the parameter of interest. Esseen (1945) proved that the error of the normal approximation for a mean of lattice random variables in $\mathbb{R}^p$ is $O(n^{-p/(p+1)})$ for the probability of centered ellipsoids. Matthes (1975) extended this result to a uniform bound for any class of convex sets with uniformly bounded curvatures (above and below) instead of the ellipsoids. Because the convex sets are not necessarily symmetric Matthes' result dealt with the approximation provided by the second order Edgeworth expansion. His result was based on a mathematical result on Fourier transforms and convex sets by Herz (1961). The best available general result for the chi-squared approximation to the likelihood ratio test and HPD (highest posterior density) regions gives the error $O(n^{-p/(p+2)} \log n)$ when the log-likelihood derivatives are lattice random variables, see Rousseau (2002).

In the present paper we prove that $O(n^{-p/(p+1)})$ can be obtained also for the case of independent replications of a lattice random variable when the expansion of the test statistic can be given in terms of linear functions of the lattice random variable and the leading term is quadratic. In the cases of likelihood ratio tests or modified likelihood ratio tests the linear statistics are derivatives of the log likelihood function statistic, and the leading term is quadratic in the score statistic for the parameters of interest.

A linear combination of lattice random variables is not in general itself a lattice random variable. Therefore the leading terms of these statistics can be considered either as quadratic forms of possibly non-lattice, but not strongly non-lattice, random vectors or as degenerate quadratic forms of lattice random vectors; that is, quadratic forms of a projection of the lattice random vector. In neither case can we apply the existing results on continuous approximations of sums or the error bounds previously obtained for lattice random variables.

To prove the result we use the technique of Matthes (1975) which again is based on the important papers by Herz (1961) and Esseen (1945). We concentrate on confidence regions that are, to first order, ellipsoids, possibly degenerate, because they are of relevance to statistics, although the technique can be applied to more general sets as indicated in the discussion section. The same comment applies to our restriction to statistics which have a limiting asymptotic chi-squared distribution.

To be more precise consider independent and identically distributed random variables $Y_1, \dots, Y_n$ supported on a lattice in $\mathbb{R}^d$; without loss of generality we take the lattice to be the shifted d-dimensional integer lattice $y_0 + \mathbb{Z}^d$ where $y_0 \in \mathbb{R}^d$ is fixed. Furthermore let $EY_i = 0$ and assume that Y_i has non-degenerate variance, Γ say, and finite moments of order 4. Let A be a $p \times d$ matrix of rank p and consider a statistic of the form

$$W = W(X) = (AX)^T \Sigma^{-1} AX + \frac{1}{\sqrt{n}} Q(X) + R_n(X), \tag{1}$$

where $X = \sum_i Y_i/\sqrt{n}$, $\Sigma = \mathrm{Var}(AX) = A\Gamma A^T$, $Q(X)$ is an odd polynomial in X, and for any constant $C > 0$ there exists a $t < 1/(p+1)$ and a $c > 0$ such that $R_n(X)$ is bounded by cn^{-1+t} whenever $\|X\| < C\sqrt{\log n}$. Our main result is the following.

Theorem 1. *With W as above we have*

$$|P(W \leq w) - F_p(w)| = O(n^{-p/(p+1)}),$$

uniformly in $w_1 < w < w_2$ for any fixed positive w_1 and w_2, where F_p denotes the distribution function of the chi-squared distribution with p degrees of freedom.

The proof, given in the next section, employs Matthes' result to sets in p coordinates of X conditional on the remaining coordinates. Relative to a straightforward application of Matthes' result two technical difficulties arise: the conditional distributions must be represented in Matthes' framework of sums of independent replications and the error bound obtained for the probabilities of the conditional sets must be uniform in the conditioning variables, thus allowing the error to remain of the same order when integrated.

It is a reasonable conjecture that the uniformity of the approximation extends to all w. Technically, for large w, this is because the possible increase in the lattice versus continuous approximation error for unbounded sets is compensated by the small probabilities for large observations. For small w the scaling result, Theorem 3, in Matthes (1975) may provide the extension.

Theorem 1 covers many typical applications of the likelihood ratio test to lattice random variables. For example, if $X = (X_1, \ldots, X_d)$ are multinomially distributed with probabilities $(p_1(\theta), \ldots, p_d(\theta))$, the log likelihood function and all its derivatives are linear in X, and the log likelihood ratio test statistic for any smooth hypothesis within any smooth model is of the form (1), see, for example, Lawley (1956) and Chandra & Ghosh (1979).

For the same reason the result applies to the log likelihood ratio statistic in linear or non-linear logistic regression if only it fits into the framework of independent replications. Similarly, log-linear Poisson models and other smooth, linear and non-linear, models are covered.

As an alternative to the log likelihood ratio test statistic other asymptotically equivalent statistic, such as the Pearson chi-squared statistic for multinomial models, similarly admit an expansion of the form 1. This also applies to HPD regions and other modified likelihood ratio statistics, see DiCiccio & Stern (1994).

2. Proof of the theorem. Without loss of generality we may assume that $R_n(x) = 0$. If the theorem holds under this condition it also holds in the stated form, because we may squeeze the set $\{W(x) \leq w\}$ between two similar sets without the $R_n(x)$ term and with w replaced by $w - \delta_n$ and $w + \delta_n$, say, with $\delta_n = O(n^{-1+t})$.

When the chi-squared approximation is uniformly valid without the $R_n(x)$ term, the result follows. Thus, we assume in the sequel that $R_n(x) = 0$.

First we split X into two parts, $X = (X_1, X_2)$ with $X_1 \in \mathbb{R}^p$ and $X_2 \in \mathbb{R}^{d-p}$. For our purpose we have to select p linearly independent columns of A, and without loss of generality we assume these to be the first. Then X_1 is simply the first p coordinates of X and

$$AX = A_1 X_1 + A_2 X_2,$$

where A_1 is the $p \times p$ matrix consisting of the first p columns of A and A_2 contains the remaining $d-p$ columns. Note that A_1 has full rank, p, and hence is invertible. For fixed w let $K_n = \{x \in \mathbb{R}^d : W(x) \leq w\}$ and consider also the conditional sets $K_n(x_2) = \{x_1 \in \mathbb{R}^p : (x_1, x_2) \in K_n\}$. Similarly, defining $W_0(x)$ as the leading term, $(Ax)^T \Sigma^{-1}(Ax)$, of W we consider the corresponding sets $K_0 = \{x \in \mathbb{R}^d : W_0(x) \leq w\}$ and $K_0(x_2) = \{x_1 \in \mathbb{R}^p : (x_1, x_2) \in K_0\}$.

The set $K_0(x_2)$ is of the form

$$(X_1 + b(x_2))^T B(X_1 + b(x_2)) \leq w,$$

where $b(x_2) = A_1^{-1} A_2 x_2$ and $B = A_1^T \Sigma^{-1} A_1$ is a positive definite symmetric matrix. Thus, to first order the set $K_n(x_2)$ is the ellipsoid $\{x_1^T B x_1 \leq w\}$ parallelly translated by the vector $-b(x_2)$ which is linear in x_2.

The probabilities of the full lattice vector X may be expanded in an Edgeworth expansion to second order, see Bhattacharya & Rao (1986, Theorem 22.1), as

$$P_n(X = x) = n^{-d/2} \varphi_\Gamma(x) \left(1 + \frac{1}{\sqrt{n}} H_{3,\Gamma}(x)\right) + O(n^{-1}),$$

where the error is uniform in x, φ_Γ denotes the d-dimensional normal density function with variance Γ, and $H_{3,\Gamma}(x)$ is an odd polynomial of order 3.

A considerable convenience is that we may limit attention to values of x bounded in Euclidean norm by $C\sqrt{\log n}$ for some sufficiently large $C > 0$, because the probability of the complement of this set is $o(n^{-1})$ when fourth moments are finite. In the sequel we therefore assume x to be within this range without further mention.

The desired probability is

$$P(W \leq w) = P_1 + P_2 + P_3 + P_4 + P_5 + O(n^{-1}),$$

where

$$P_1 = \int_{K_0} \varphi_\Gamma(x)\, dx = F_p(w),$$

$$P_2 = \left(\int_K - \int_{K_0}\right) \varphi_\Gamma(x)\, dx,$$

$$P_3 = n^{-d/2} \sum_K{}^{*} \varphi_\Gamma(x) - \int_K \varphi_\Gamma(x)\, dx,$$

$$P_4 = n^{-1/2} \int_{K_0} \varphi_\Gamma(x) H_{3,\Gamma}(x)\, dx = 0,$$

$$P_5 = n^{-1/2} \left(n^{-d/2} \sum_K{}^{*} \varphi_\Gamma(x) H_{3,\Gamma}(x) - \int_{K_0} \varphi_\Gamma(x) H_{3,\Gamma}(x)\, dx \right),$$

where $\sum^*$ refers to summation over lattice points for x, that is, over points in $L_n = \sqrt{n} y_0 + \frac{1}{\sqrt{n}} \mathbb{Z}^d$, and the error term stems from the Edgeworth expansion over the set K.

The term P_4 vanishes because the set K_0 is symmetric and the integrand is an odd function.

Next, $P_2 = O(n^{-1})$ because the polynomial $Q(x)$ is odd and K_0 is symmetric in x so the contribution of order $n^{-1/2}$ vanishes, leaving an error of order $O(n^{-1})$.

The term P_3 is our main concern for which we will use the result from Matthes (1975).

Also, $P_5 = O(n^{-1})$ which is seen by considering the conditional sets (given x_2). The discrete sum over such a set is equal to the Riemann approximation to the integral of the same function on the same set. As this function is the Gaussian density function (times a polynomial) and the set is convex, the error of the Riemann approximation is of order $n^{-1/2}$. This error is uniform over all convex sets and is therefore uniform in w, see Bhattacharya and Rao (1986, Corollary 3.2).

Now consider P_3. Let $\Gamma_2 = \operatorname{Var} X_2$ and consider the conditional Gaussian density $\varphi(x_1|x_2)$ of X_1 given X_2; that is $\varphi(x_1|x_2) = \varphi_{\mu_2(x_2),\Gamma_{1.2}}(x_1)$, the Gaussian density with mean $\mu_2(x_2)$ which is linear in x_2 and variance $\Gamma_{1.2} = \operatorname{Var} X_1 - \operatorname{Cov}(X_1, X_2)(\operatorname{Var} X_2)^{-1} \operatorname{Cov}(X_2, X_1)$. Decompose P_3 as

$$P_3 = P_{3,1} + P_{3,2} = n^{-(d-p)/2} \sum_{L_{2,n}} \varphi_{\Gamma_2}(x_2) P_{3,1}(x_2)$$
$$+ n^{-(d-p)/2} \sum_{L_{2,n}} \varphi_{\Gamma_2}(x_2) g(x_2) - \int_{\mathbb{R}^{d-p}} \varphi_{\Gamma_2}(x_2) g(x_2)\, dx_2,$$

with

$$P_{3,1}(x_2) = n^{-p/2} \sum_{L_{1,n} \cap K_n(x_2)} \varphi(x_1|x_2) - \int_{K_n(x_2)} \varphi(x_1|x_2)\, dx_1,$$

where $g(x_2) = \int_{K_n(x_2)} \varphi(x_1|x_2)\, dx_1$ is twice continuously differentiable, and $L_{1,n}$ is the lattice $\sqrt{n} y_{01} + \mathbb{Z}^p/\sqrt{n}$, where the vector y_{01} consists of the first p coordinates of y_0.

To handle $P_{3,1}(x_2)$ define the shifted variable $\tilde{x}_1 = x_1 - \mu_2(x_2)$ and the correspondingly shifted lattice $\tilde{L}_{1,n} = \sqrt{n} y_{01} - \mu_2(x_2) + \mathbb{Z}^p/\sqrt{n}$ and the shifted set $\tilde{K}_n(x_2) = K_n(x_2) - \mu_2(x_2)$. Then

$$P_{3,1}(x_2) = n^{-p/2} \sum_{\tilde{L}_{1,n} \cap \tilde{K}_n(x_2)} \varphi_{0,\Gamma_{1.2}}(\tilde{x}_1) - \int_{\tilde{K}_n(x_2)} \varphi_{0,\Gamma_{1.2}}(\tilde{x}_1) d\tilde{x_1}$$

is now written in terms of the density function $\varphi_{0,\Gamma_{1.2}}$ which is independent of n.

Ideally we would apply Matthes' result at this stage to show that the lattice sum of the Gaussian point probabilities over $K_n(x_2)$ is approximated to the prescribed order by the integral. Matthes' result, however, applies only to a standardized sum of independent replications of a lattice random variable, so we need to show that the conditional Gaussian density may arise as the leading term of an Edgeworth expansion for such a standardized sum. Unfortunately, this may not be true because the statistic must have support on the lattice $\tilde{L}_{1,n}$ which depends awkwardly on n. To solve the problem we make another, very small, shift of the lattice so that zero becomes a lattice point. This will change the mean of the Gaussian density slightly, but only by a magnitude that will give an error of sufficiently small order.

Let $\rho_n(x_2) = \sqrt{n}y_{01} - \mu_2(x_2) - [\sqrt{n}y_{01} - \mu_2(x_2)]$, where the square brackets denote coordinate-wise integer rounding, and define $x_1^\star = \tilde{x}_1 - \rho_n(x_2)/\sqrt{n}$, $K_n^\star(x_2) = \tilde{K}_n(x_2) - \rho_n(x_2)$, and $L_{0,n} = \mathbb{Z}^p/\sqrt{n}$. This leads to

$$\begin{aligned} P_{3,1}(x_2) = n^{-p/2} & \sum_{L_{0,n}\cap K_n^\star(x_2)} \varphi_{0,\Gamma_{1.2}}(x_1^\star + \rho_n(x_2)/\sqrt{n}) \\ & - \int_{K_n^\star(x_2)} \varphi_{0,\Gamma_{1.2}}(x_1^\star + \rho_n(x_2)/\sqrt{n})\, dx_1^\star \\ = n^{-p/2} & \sum_{L_{0,n}\cap K_n^\star(x_2)} \varphi_{0,\Gamma_{1.2}}(x_1^\star)\left(1 - \frac{\rho_n(x_2)^T\Gamma_{1.2}^{-1}x_1^\star}{\sqrt{n}}\right) \\ & - \int_{K_n^\star(x_2)} \varphi_{0,\Gamma_{1.2}}(x_1^\star)\left(1 - \frac{\rho_n(x_2)^T\Gamma_{1.2}^{-1}x_1^\star}{\sqrt{n}}\right) dx_1^\star + O(n^{-1}). \end{aligned}$$

The right hand side of the above equation can thus be written as two differences between sums and integrals. The first is a first order difference, the second corresponds to the rounding error term. This latter term is equal to $n^{-1/2}$ times the error of the Riemann approximation of the integral of $\varphi_{0,\Gamma_{1.2}}(z)\rho_n(x_2)^T\Gamma_{1.2}^{-1}z$ over a convex set. Using the same argument as for P_5, we obtain a term of order $O(n^{-1})$ uniformly over w and x_2, since x_2 only appears in $K_n^\star$, for which we can use the uniformity over convex sets, and in $\rho_n(x_2)$ which is bounded by $1/2$.

To deal with the first order difference, we apply Matthes' (1975) result. Using Lemma 1, see the appendix, we can construct n independent replications of lattice random vectors Z_i, $i = 1, ..., n$ with zero mean, covariance matrix $\Gamma_{1.2}$, null third order cumulants and finite fourth order cumulants. Then using Bhattacharya and Rao (1986, Theorem 22.1) to expand the point probabilities, $p_n(z)$, of the standardized sum, we have

$$\sum_{L_{0,n}\cap K_n^\star(x_2)} p_n(z) - n^{-p/2} \sum_{L_{0,n}\cap K_n^\star(x_2)} \varphi_{0,\Gamma_{1.2}}(z) = O(n^{-1}),$$

uniformly over x_2, as x_2 only appears in the set $K_n^\star(x_2)$. Consequently

$$P_{3,1}(x_2) = \sum_{L_{0,n}\cap K_n^\star(x_2)} p_n(z) - \int_{K_n^\star(x_2)} \varphi_{0,\Gamma_{1.2}}(x_1^\star)\, dx_1^\star + O(n^{-1}).$$

When n is large enough, $K_n^\star(x_2)$ is a convex set for each $|x_2| \le C\log n$; it is the translation of the ellipsoid $\{x_1 : x_1^T B x_1 \le w\}$ by the vector $-(b + \mu_2 + \rho_n)(x_2)$ plus a perturbation $Q(x_1^\star + \mu_2(x_2) + \rho_n(x_2), x_2)/\sqrt{n}$, which is a polynomial in $x_1^\star$. Hence it has uniformly bounded curvature from above and below and its recentered support function is infinitely differentiable, and uniformly bounded in x_2. Re-centering does not change Matthes's result, since the set enters the proof only through the absolute value of the characteristic function of the indicator of the set. This is obviously independent of the origin. We can therefore apply Matthes' Theorem 1 (1975); apart from terms of order $O(n^{-1})$,

$$P_{3,1}(x_2) = \left| \sum_{L_{0,n}\cap K_n^\star(x_2)} p_n(z) - \int_{K_n^\star(x_2)} \varphi_{0,\Gamma_{1.2}}(x_1^\star)\, dx_1^\star \right| \le c(\beta_4)^{3p/(4(p+1))} n^{-p/(p+1)},$$

where β_4 is the sum of the fourth moments of the coordinates of Z_i and hence is independent of x_2. The constant c is independent of n and x_2. We thus obtain

$$P_{3,1} = n^{-(d-p)/2} \sum_{L_{2,n}} \varphi_{\Gamma_2}(x_2) P_{3,1}(x_2) = O(n^{-p/(p+1)}).$$

Now consider the second part of P_3, namely $P_{3,2}$. Since $K_n(x_2)$ is smooth, $g(x_2)$ is twice continuously differentiable and is bounded by 1. Thus, uniformly in x_2,

$$n^{-(d-p)/2} \sum_{L_2} \varphi_{\Gamma_2}(x_2) g(x_2) - \int_{\mathbb{R}^{d-p}} \varphi_{\Gamma_2}(x_2) g(x_2) = O(n^{-1}),$$

which completes the proof.

3. Discussion. It is important to notice that our result gives only an upper bound for the error of the chi-squared approximation but does not rule out the possibility that the error could be of lower order. Recent results by Götze (2000) and Bentkus & Götze (2001) indicate that for classes of sets including exact ellipsoids the order $O(n^{-1})$ is achievable for dimension $p \geq 5$. This extends previous results by Bentkus & Götze (1997, 1999) where similar results for $p \geq 9$ were obtained. In our setting their results correspond to the situation where the matrix A in the definition of the statistic $W(X) \sim (AX)^T \Sigma^{-1}(AX)$ is of full rank and without a term of order $O(n^{-1/2})$ in $W(X)$, see equation (1). From an intuitive point of view one would expect the inclusion of the extra components, due to the nuisance parameters of dimension $d-p$, to have some smoothing effect rather than adding to the discreteness error. It is unclear, however, whether such a smoothing effect improves the order of the approximation in particular cases.

Frydenberg & Jensen (1989) simulate some examples of the likelihood ratio test for multinomial random variables to investigate the order of error empirically. In particular, they compare with the Bartlett correction which, in the continuous case, would improve the error from order $O(n^{-1})$ to $O(n^{-2})$. If the error due to discreteness is of order $O(n^{-1})$ it would give some meaning, asymptotically, to apply the Bartlett correction to a discrete model since it might remove part of the dominating error. The empirical findings of Jensen & Frydenberg indicate that this is not the case, but their examples are low-dimensional ($p \leq 3$) and therefore do not challenge the conjecture of a general error of order $O(n^{-1})$ for $p \geq 5$. Moreover, a Bartlett correction would still correct the expectation of any smooth function of $W_n(X)$. Even though it might not improve the order of the approximation of the discrete probabilities by its continuous counterparts, it would re-center the approximation to a better order of accuracy. In the special case of the one-dimensional binomial Brown *et al.* (2002) prove that re-centering diminishes the coverage error of confidence regions.

We have concentrate on approximate, possibly degenerate ellipsoids. However, as Matthes's (1975) result is valid for any bounded convex set, our result remains valid if $W(X)$ in equation (1) is a more general (smooth) function, provided that the conditional sets, $K_n(x_2) = \{x_1 : W(x_1, x_2) \leq w\}$, are convex for sufficiently large n and satisfy Matthes' conditions of bounded curvature. Under this framework, the technique that we have used to prove Theorem 1 can be applied directly, although we would have to add the second term, of order $1/\sqrt{n}$, in the Edgeworth expansion if the set is not symmetric.

Actually, the condition on independent and identically distributed random vectors is not necessary either. Indeed, it is enough to assume that there exists a suitable pointwise Edgeworth expansion on lattice points for the distribution of x, as the key point in our proof was the use of Matthes's result on the vector z that is constructed from the pointwise Edgeworth expansion. So assuming that we have such an Edgeworth expansion on a lattice say $L/\sqrt{n}$, it is possible to construct a vector z as a renormalised sum of independent and identically distributed lattice vector $z_i \in L$ with zero mean and the same covariance matrix. Then z would have the same Edgeworth expansion to the first order as our original statistic and we would be able to use Matthes's result as above to obtain the order $O(n^{-p/(p+1)})$. One possible application of such an extension is the logistic regression with lattice covariates. In fact, logistic regression would already be covered by Theorem 1 in the case of systematic replications of the covariates.

Appendix: Construction of a lattice distribution with certain moments. As a technical tool we need to prove the existence of a distribution on the integer lattice, $\mathbb{Z}^p$, having zero mean and a given non-singular covariance matrix, Λ, say.

LEMMA 1. *For any positive definite symmetric matrix, Λ, of size $p \times p$ there exists a probability distribution on the integer lattice $\mathbb{Z}^p$ with mean zero and variance matrix Λ. Furthermore, there exist symmetric distributions of this kind with moments of all orders.*

PROOF. Let V be a p-dimensional Gaussian random variable with $EV = 0$ and $\operatorname{Var} V = \Lambda$. Then, for any positive number k define the lattice random variable

$$W_k = \begin{cases} 0 & \text{with probability } 1 - 1/k^2 \\ [kY] & \text{with probability } 1/k^2 \end{cases},$$

where $[kY]$ denotes the integer rounding of kY, that is $kY = [kY] + R_k(Y)$ with all coordinates of $R_k(Y)$ bounded by $1/2$. Since the distribution of W_k is symmetric we have $EW_k = 0$. Further,

$$\begin{aligned} \operatorname{Var} W_k &= \frac{1}{k^2}\left(k^2 \operatorname{Var} Y - k \operatorname{Cov}(Y, R_k(Y)) - k \operatorname{Cov}(R_k(Y), Y) + \operatorname{Var} R_k(Y)\right) \\ &= \operatorname{Var} Y + O(1/k), \end{aligned}$$

as $k \to \infty$, thus implying that we can approximate any non-singular variance, Λ, as closely as we want by a distribution on the integer lattice. Since the variance of a distribution mixture is the mixture of the variances, the set of achievable variance matrices for lattice distributions is convex. Thus, the target Λ, can be enclosed in a cube of which we may approximate the corners, and appropriate mixing yields a lattice distribution with mean zero and covariance matrix Λ. The distribution constructed in this way is symmetric and has moments of all orders.

Although the result is not valid if the condition that the mean is zero is removed, the distribution may be shifted to have mean on any other lattice point. And by non-singular linear transformation the result trivially extends to any non-degenerate lattice, meaning that we may construct a lattice distribution with mean on any lattice point and with any non-singular variance matrix.

REFERENCES

Bentkus, V. & Götze, F. (1997). On the lattice point problem for ellipsoids. *Acta Arith.* **80**, 101–125.

Bentkus, V. & Götze, F. (1999). Lattice point problems and distribution of values of quadratic forms. *Ann. Math.* **150**, 977–1027.

Bentkus, V. & Götze, F. (2001). Lattice points in multidimensional bodies. *Forum Math.* **13**, 149–225.

Bhattacharya, R. N. & Rao, R. R. (1986). *Normal Approximation and Asymptotic Expansion.* 2nd ed. Wiley, New York.

Brown, L. D., Cai, T. & DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Statist.* **30**, 160-201.

Chandra, T. K. & Ghosh, J. K. (1979). Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-squared variables. *Sankhya Ser. A* **41**, 22–47.

DiCiccio, T. & Stern, S. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *J. R. Statist. Soc. Ser. B* **56**, 397–408.

Esseen, C.-G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.* **77**, 1–125.

Frydenberg, M. & Jensen, J. L. (1989). Is the 'improved likelihood ratio statistic' really improved in the discrete case? *Biometrika* **76**, 655–661.

Götze, F. (2000). Lattice point problems and values of quadratic forms. Preprint 00–103, Faculty of Mathematics, University of Bielefeld. (Available from `www.mathematik.uni-bielefeld.de/~goetze`).

Herz, C. S. (1961). Fourier transforms related to convex sets. *Ann. Math.* **75**, 81–92.

Lawley, D. N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika* **43**, 295–303.

Matthes, T. K. (1975). The multivariate central limit theorem for regular convex sets. *Ann. Probab.* **3**, 503–515.

Rousseau, J. (2002). Asymptotic properties of HPD regions in the discrete case. *J. Multivariate Anal.*, in press.

Department of Statistics
University of Paris V
45 Rue des Saint Pères
75006 Paris, France
E-mail: rousseau@ensae.fr

Department of Mathematics and Physics
The Royal Veterinary and Agricultural University
Thorvaldsensvej 40
DK-1871 Frederiksberg C, Denmark
E-mail: ims@kvl.dk