THE MULTIVARIATE SADDLEPOINT APPROXIMATION TO THE DISTRIBUTION OF ESTIMATORS. A GENERAL APPROACH

JUAN CARLOS ABRIL(*), MARÍA DE LAS MERCEDES ABRIL and CARLOS I. MARTÍNEZ

Universidad Nacional de Tucumán. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Tucumán, Argentina

jabril@herrera.unt.edu.ar mabril@face.unt.edu.ar cimartinez@herrera.unt.edu.ar

ABSTRACT

We develop the theory of multivariate saddlepoint approximations. Our treatment differs from the one in Barndorff-Nielsen and Cox (1979, equation (4.7)) in two aspects:

- Our results are satisfied for random vectors that are not necessarily sums of independent and identically distributed random vectors, and
- We consider that the sample is taken from any distribution, not necessarily a member of the exponential family of densities.

We also show the relationship with the corresponding multivariate Edgeworth approximations whose general treatment was developed by Durbin in 1980, emphasizing that the basic assumptions that support the validity of both approaches are essentially similar.

Key words: Approximate distributions; Asymptotic expansions; Edgeworth approximations; Saddlepoint approximations.

1. Multivariate saddlepoint expansions

The main purpose of this paper is to present the theory of multivariate saddlepoint expansions. These were introduced in statistics in a pioneering work of Daniels (1954) for the univariate case. Barndorff-Nielsen and Cox (1979) studied the multivariate case, but basically his focus is centred on applications to sums of independent and identically distributed vectors and when the underlying distributions are members of the exponential family.

Let $h_n(\mathbf{x},\boldsymbol{\theta})$ be the density of the random vector

$$\mathbf{X}_n = n^{-1/2}\{\mathbf{S}_n - E(\mathbf{S}_n)\}, \qquad (1)$$

where $\mathbf{S}_n$ m-dimensional random vector. We will derive, and also find the error bound of the multivariate saddlepoint approximation to the density of $h_n(\mathbf{x},\boldsymbol{\theta})$.

The characteristic function of $\mathbf{X}_n$ is

$$\phi_{\mathbf{X}_n}(\mathbf{z}) = e^{-i\frac{\mathbf{z}'}{\sqrt{n}}E(\mathbf{S}_n)}\phi_n\left(\frac{\mathbf{z}}{\sqrt{n}}\right),$$

where ϕ_n is the characteristic function of $\mathbf{S}_n$.

In what follows we need to introduce the following three assumptions:

- **Assumption 1:** If n is large enough, $|\phi_n(\mathbf{z})|$ is integrable over $\Re^m$ and if δ_1 is an arbitrary positive constant the limit of

$$n^{\frac{r}{2}-1}\int_{B_{\delta_1\sqrt{n}}}\left|\phi_n\left(\mathbf{z}/\sqrt{n}\right)\right|d\mathbf{z}, \qquad (2)$$

as $n \to \infty$, $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$ is zero, where $B_{\delta_1\sqrt{n}}$ is the region that contains $\hat{\mathbf{z}}$ that is, the roots of the following equation $\nabla\{\psi_{X_n}(\mathbf{z}) - i\mathbf{z}'\mathbf{x}\} = \mathbf{0}$.

- **Assumption 2:** The derivatives

$$\frac{\partial^j \psi_n(\mathbf{z})}{\partial \mathbf{z}^j} \qquad (3)$$

exist for $\mathbf{z}$ in neighbourhoods of the origin and of $\hat{\mathbf{z}}$. This also applies for $\boldsymbol{\theta}$ in a neighbourhood of $\boldsymbol{\theta}_0$ and for $j=1,\ldots,r$. Besides, the limit of $n^{-1}\frac{\partial^r \psi_n(\mathbf{z})}{\partial \mathbf{z}^r}$ exists when $n \to \infty$, $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$ in neighbourhoods of the origin and of $\hat{\mathbf{z}}$.

- **Assumption 3**: The derivatives

$$\frac{\partial^j \psi_n(\mathbf{z})}{\partial \mathbf{z}^j} = O(n) \qquad (4)$$

uniformly for $\boldsymbol{\theta}$ in a neighbourhood of $\boldsymbol{\theta}_0$; for $\mathbf{z}$ in neighbourhoods of the origin and of $\hat{\mathbf{z}}$.

These three assumptions were stated in a way that they contain assumptions 2, 3 and 4 given in Durbin (1980) in his general treatment of multivariate Edgeworth expansions. In effect, if we take $\hat{\mathbf{z}}$ not as root of an equation but equal to the null vector $\mathbf{0}$, everything leads us to what was discussed by Durbin (1980).

With this in mind we can formulate a theorem, present in the full-length paper, that will lead us to the following approximation to the density $h_n(\mathbf{x},\boldsymbol{\theta})$ of $\mathbf{X}_n$ defined in (1).

$$\hat{h}_n(\mathbf{x},\boldsymbol{\theta}) = \frac{1}{(2\pi)^{m/2}}|\mathbf{B}|^{-1/2}\exp\{\psi_{\mathbf{X}n}(\hat{\mathbf{z}}) - i\hat{\mathbf{z}}'\mathbf{x}\} \qquad (5)$$

where $\mathbf{B}$ is a matrix such that $\mathbf{B} = \mathbf{B}'$. Formula (5) is called the *saddlepoint approximation* to the density $h_n(\mathbf{x},\boldsymbol{\theta})$. The error bound of approximation is $O(n^{-1})$ in this case.

2. Error bound of the multivariate saddlepoint approximation

Let us take, not the density $h_n(\mathbf{x},\boldsymbol{\theta})$ but a proper member of the exponential family or the conjugate family as stated in Kinchin (1949)

$$f_n(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) = \frac{e^{i\boldsymbol{\lambda}'\mathbf{x}}h_n(\mathbf{x},\boldsymbol{\theta})}{\phi_{\mathbf{X_n}}(\boldsymbol{\lambda})}, \qquad (6)$$

where the function $\phi_{\mathbf{X}_n}$ was defined at the beginning of the previous section. Clearly, $h_n(\mathbf{x},\boldsymbol{\theta})$ depends on the parameter vector $\boldsymbol{\theta}$ of dimension q and the density function $f_n(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda})$ depends on the parameter $\boldsymbol{\gamma} = (\boldsymbol{\theta},\boldsymbol{\lambda})'$ of dimension $q+n$.

Formula (6) can be written in the more familiar form

$$f_n(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) = \exp\{i\mathbf{z}'\mathbf{x} - \alpha(\mathbf{x},\boldsymbol{\theta}) - \beta(\boldsymbol{\lambda})\}. \qquad (7)$$

It is important to see that when $\boldsymbol{\lambda} = \mathbf{0}$ then $f(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda},\mathbf{0}) = h_n(\mathbf{x},\boldsymbol{\theta})$.

Suppose that we are interested in the density $f_n(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}_0)$ for some particular value $\boldsymbol{\lambda}_0$ of $\boldsymbol{\lambda}$. Then, for any $\boldsymbol{\lambda}$ we have that

$$\begin{aligned} f_n(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}_0) &= \exp\left\{\beta(\boldsymbol{\lambda}) - \beta(\boldsymbol{\lambda}_0) + i(\boldsymbol{\lambda}-\boldsymbol{\lambda}_0)'\mathbf{x}\right\} f_n(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) \\ &= \exp\left\{\psi_{\mathbf{X}_n}(\boldsymbol{\lambda}) - \psi_{\mathbf{X}_n}(\boldsymbol{\lambda}_0) + i(\boldsymbol{\lambda}-\boldsymbol{\lambda}_0)'\mathbf{x}\right\} \\ &\quad \times f_n(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\lambda}) \end{aligned} \qquad (8)$$

Therefore, an approximation to $f_n(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda}_0)$ can be obtained via an approximation based on the Edgeworth expansion of $f_n(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\lambda})$ for any $\boldsymbol{\lambda}$ satisfying the assumptions made in Theorem 1 of Durbin (1980).

On the other hand, when we consider $\boldsymbol{\lambda}_0 = \mathbf{0}$ and $\boldsymbol{\lambda} = \hat{\mathbf{z}}$ we may see that

$$f_n(\mathbf{x}, \boldsymbol{\theta}, \mathbf{0}) = h_n(\mathbf{x}, \boldsymbol{\theta}) = \exp\{\psi_{\mathbf{X}_n}(\hat{\mathbf{z}}) - i\hat{\mathbf{z}}'\mathbf{x}\} f_n(\mathbf{x}, \boldsymbol{\theta}, \hat{\mathbf{z}}). \tag{9}$$

In this formula, if we replace $f_n(\mathbf{x}, \boldsymbol{\theta}, \hat{\mathbf{z}})$ on the right hand side for the corresponding Edgeworth expansion, the result will lead to the saddlepoint approximation and its corresponding error bound of approximation of the density $h_n(\mathbf{x}, \boldsymbol{\theta})$. This is stated on the following theorem.

- **Theorem:** Suppose that there is an integer $r = 4$ such that assumptions 1, 2 and 3 stated above hold. Then there is a neighbourhood $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta_2$ of $\boldsymbol{\theta}_0$ such that

$$h_n(\mathbf{x}, \boldsymbol{\theta}) = \hat{h}_n(\mathbf{x}, \boldsymbol{\theta})\{1 + O(n^{-1})\} \tag{10}$$

uniformly in $\mathbf{x}$ and in $\boldsymbol{\theta}$ for $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta_2$ where δ_2 is a suitably chosen positive constant independent of n and $\hat{h}_n(\mathbf{x}, \boldsymbol{\theta})$ is the saddlepoint approximation already defined to the density $h_n(\mathbf{x}, \boldsymbol{\theta})$ of the random vector $\mathbf{X}_n$.

3. Approximation to the estimators distribution

Usually one is interested in the saddlepoint approximation $\hat{g}(\mathbf{t}, \boldsymbol{\theta})$ to the distribution $g(\mathbf{t}, \boldsymbol{\theta})$ of the estimator $\mathbf{T}_n = n^{-1}\mathbf{S}_n$ for the parameter vector $\boldsymbol{\theta}$, where $\mathbf{S}_n$ satisfies assumptions 1, 2 and 3 that we made before. What we may see is that maximum likelihood estimators often satisfy these assumptions. Using a result stated in the full-length paper we can have that in these cases $\mathbf{X}_n = n^{-1/2}\{\mathbf{T}_n - E(\mathbf{T}_n)\}$.

If we also use some other expressions present at the full-length paper we may see that

$$\begin{aligned} g(\mathbf{t}, \boldsymbol{\theta}) &= \left(\frac{n}{2\pi}\right)^{m/2} |\mathbf{B}|^{-1/2} \exp\{\psi_{\mathbf{X}_n}(\hat{\mathbf{z}}) - in^{-1/2}\hat{\mathbf{z}}'\{\mathbf{t} - E(\mathbf{T}_n)\}\} \\ &\quad \times \{1 + O(n^{-1})\} \\ &= \hat{g}(\mathbf{t}, \boldsymbol{\theta})\{1 + O(n^{-1})\}, \end{aligned} \tag{11}$$

uniformly in $\mathbf{t}$ for $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta_2$, with

$$\hat{g}(\mathbf{t}, \boldsymbol{\theta}) = \left(\frac{n}{2\pi}\right)^{m/2} |\mathbf{B}|^{-1/2} \exp\{\psi_{\mathbf{X}_n}(\hat{\mathbf{z}}) - in^{-1/2}\hat{\mathbf{z}}'\{\mathbf{t} - E(\mathbf{T}_n)\}\}. \tag{12}$$

3.1 Integration over the sample space

What we need for practical applications is an error bound of an integral over an appropriate region. The fact that the error term in (11) is proportional and uniform in $\mathbf{t}$ for $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta_2$, enables us to establish such a bound in a simple way.

Since the cumulants $\mathbf{S}_n$ of orders up to four exist and are of order n, the fourth moment of $\mathbf{S}_n$ around its mean is of order n^2. Also we can see that since $\mathbf{T}_n = n^{-1}\mathbf{S}_n$, it follows that the fourth moment of $\mathbf{T}_n$ around $\boldsymbol{\theta}$ is of order n^{-2}. A constant C_1 therefore exists such that, when $\boldsymbol{\theta}_0$

is the true value of $\boldsymbol{\theta}$, $E\left(\left\|\mathbf{T}_n-\boldsymbol{\theta}_0\right\|^4\right)<C_1 n^{-2}$ for a sufficiently large n. By a Chebychev-type argument we have that, for any δ^2, if A_{δ_2} is the region $\left\|\mathbf{t}-\boldsymbol{\theta}_0\right\| \geq \delta_2$, then

$$C_1 n^{-2} > \int_{A_{\delta_2}} \left\|\mathbf{t}-\boldsymbol{\theta}_0\right\|^4 g(\mathbf{t},\boldsymbol{\theta}) d\mathbf{t} + \int_{\bar{A}_{\delta_2}} \left\|\mathbf{t}-\boldsymbol{\theta}_0\right\|^4 g(\mathbf{t},\boldsymbol{\theta}) d\mathbf{t}$$

$$> \delta_2^4 \Pr\left(\left\|\mathbf{t}-\boldsymbol{\theta}_0\right\| \geq \delta_2\right),$$

Now let's write (11) in the following form

$$\hat{g}(\mathbf{t},\boldsymbol{\theta})\left(1-\frac{C_2}{n}\right) < g(\mathbf{t},\boldsymbol{\theta}) < \hat{g}(\mathbf{t},\boldsymbol{\theta})\left(1+\frac{C_2}{n}\right) \quad (13)$$

uniformly in $\mathbf{t}$ for $\left\|\mathbf{t}-\boldsymbol{\theta}_0\right\| < \delta_2$, where C_2 does not depend on n or $\mathbf{t}$, n is sufficiently large and $\hat{g}(\mathbf{t},\boldsymbol{\theta})$ is given in (12).

Suppose, on the other hand, that $\hat{g}(\mathbf{t},\boldsymbol{\theta})$ is integrable and let $\{B_n\}$ be a sequence of Borel sets in $\Re^m$ such that $\Pr(\mathbf{T}_n \in B_n)$ converges to a positive limit. Using (13) and the following result

$$\Pr\left(\left\|T_n-\boldsymbol{\theta}_0\right\| \geq \delta_2\right) = O\left(n^{-2}\right) \quad (14)$$

we can easily show that the integral of $\hat{g}(\mathbf{t},\boldsymbol{\theta})$ over the region $\left\|\mathbf{t}-\boldsymbol{\theta}_0\right\| \geq \delta_2$ is $O\left(n^{-1}\right)$ and hence

$$\int_{B_n} \hat{g}(\mathbf{t},\boldsymbol{\theta}) d\mathbf{t} = \int_{B_n} g(\mathbf{t},\boldsymbol{\theta}) d\mathbf{t} + O\left(n^{-1}\right). \quad (15)$$

The error in (15) is uniform in B_n because it is less than αn^{-1} in absolute valor for a sufficiently large n where α is a suitable positive constant that does not depend on the sequence $\{B_n\}$.

From a practical point of view (15) is the most important result since it demonstrates that the basic approximation $\hat{g}(\mathbf{t},\boldsymbol{\theta})$ can be integrated for inference purposes with an error that is of order n^{-1} at most.

However, the result has been formulated in a way that to a mathematical statistician may seem somewhat eccentric. For problems such as this, the customary practice is to consider instead of $\mathbf{T}_n$ a function such as $\mathbf{W}_n = (\mathbf{T}_n - \boldsymbol{\theta})\sqrt{n}$ which has a nondegenerate limiting distribution. One would then seek to show that for an arbitrary Borel set A in a space $\Re^m$ the probability that $\mathbf{W}_n$ takes values in A is the integral of the approximating density over A plus an error of order say n^{-1}. The reason we decided to base our results on $\mathbf{T}_n$ rather than $\mathbf{W}_n$ is that we want the basic formulae to be suitable in the cases where n is finite. The applied worker wants an approximation to the density of the estimator he is actually using, not some transformation that has been introduced purely for mathematical convenience. The formulation adopted here does not entail in our view any real loss of rigor or generality. The statement $\mathbf{W}_n \in A$ is readily transformable to a statement $\mathbf{T}_n \in B_n$ and the analogue of the requirement that $\Pr(\mathbf{T}_n \in B_n)$ converges to a positive limit is just the requirement that the limit of $\Pr(\mathbf{W}_n \in A)$ is positive.

In all cases of practical importance, $\mathbf{T}_n$ is either the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ or is asymptotically equivalent to it in the sense that $\mathbf{T}_n = \hat{\boldsymbol{\theta}} + O_p\left(n^{-1}\right)$.

4. Renormalization

Daniels (1956) pointed out that when the constant term in the saddlepoint approximation is adjusted to make the integral over the whole space equal to unity, the order of magnitude of the error is often reduced from n^{-1} to $n^{-3/2}$. He called this process ***renormalization***. We show that the same result holds for the case of (11) and we obtain the following result

$$g(\mathbf{t},\boldsymbol{\theta}) = \tilde{g}(\mathbf{t},\boldsymbol{\theta})\{1 + O_x(n^{-3/2})\}. \qquad (16)$$

where $\tilde{g}(\mathbf{t},\boldsymbol{\theta}) = K_n(\boldsymbol{\theta})\hat{g}(\mathbf{t},\boldsymbol{\theta})$ is the renormalized density and $K_n(\boldsymbol{\theta})$ comes from the following expression developed in the full length paper

$$K_n^{-1} = \int \hat{g}(\mathbf{t},\boldsymbol{\theta})d\mathbf{t}$$

where the integral is over the space $\Re^m$. If we work on (16) we will find out that the error term is not uniform in $\mathbf{x}$. Here we use the we use the notation $O_x(n^{-q})$ introduced by Durbin(1980) to denote a quantity which $O(n^{-q})$ uniformly for all $\mathbf{x}$.

What we finally see is that if $\mathbf{x}$ is held fixed, then $\tilde{g}(\mathbf{t},\boldsymbol{\theta})$ has a proportional error of order $n^{-3/2}$.

5. References

Abril, J. C. (1984). Un estudio de las aproximaciones a las densidades con énfasis en las expansiones de Edgeworth. *Estadística - Journal of the Inter-American Statistical Institute*, 36, 109-24.

Abril, J.C. (1985). *Asymptotic expansions for time series problems with applications to moving average models*. Ph.D. Thesis, The London School of Economics and Political Science, University of London, England.

Abril, J.C. (1987). The approximate densities of some quadratic forms of stationary random variables. *Journal of Time Series Analysis*. 8, 249-59.

Abril, J.C. (2004). *Modelos para el Análisis de las Series de Tiempo*. Ediciones Cooperativas, Buenos Aires.

Abril, J. C., Abril, M. de las M. and Martínez, C. I. (2007). Aproximaciones a la distribución del estimador del coeficiente en modelos AR(1) no estacionarios. *Estadística - Journal of the Inter-American Statistical Institute*, 59, 87-102.

Abril, J. C., Abril, M. de las M. and Martínez, C. I. (2008). La aproximación de punto de ensilladura a la distribución del coeficiente de correlación serial en el caso de raíz unitaria. *Actas del VIII CLATSE (Congreso Latinoamericano de Sociedades de Estadística),* Montevideo, Uruguay.

Abril, J. C., Abril, M. de las M. and Martínez, C. I. (2008). Edgeworth approximation to the distribution of the autocorrelation coefficient in the presence of unit roots. *Pakistan Journal of Statistics*, 24(3), 193-206.

Abril, J.C. and Santillan, M.R. (1998). Approximation to the finite sample distribution of a general estimator of the coefficient in a AR(1) model. *Pak. J. Statist.* 14(1), 11-23.

Barndorff-Nielsen, O and Cox, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. R. Statist. Soc. B*,41, 279-312.

Bhattacharya, R.N. and Rao, R.R. (1976). *Normal approximations and asymptotic expansions*. John Wiley and Sons. New York.

Cox, D. R. (1948). A note on the asymptotic distribution of the range. *Biometrika, 35*, 311-15.

Daniels, H. E. (1954). Saddlepoint approximations in Statistics. *Ann. Math. Statist., 25*, 631-50.

Daniels, H. E. (1956), The approximate distribution of serial correlation coefficients, *Biometrika, 43,* 169-85.

Durbin, J. (1980a), Approximations for densities of sufficient estimators, *Biometrika, 67*, 311-33.

Durbin, J. (1980b), The approximate distribution of partial serial correlation coefficients calculated from residuals from regression on Fourier series, *Biometrika, 67,* 335-49.

Edgeworth, F. Y. (1904). The Law of Errors. *Trans. Camb. Phil. Soc., 20,* 36 and 113 (with a non - printed Appendix in *T. C. P. S* but handed over with reprints).

Fowler, R. H. (1936). *Statistical Mechanics*. Cambridge University Press. Cambridge.

Jeffreys, H. (1948). *Theory of Probability*. Oxford University Press. Oxford.

Khinchin, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*. Dover Publications. New York.

Leipnik, R. B. (1947). Distribution of the serial correlation in a circularly correlated universe. *Ann. Math. Statist., 18*, 80-7.

Phillips, P. C. B. (1977a). Approximations to some finite sample distributions associated with a first order stochastic difference equation. *Econometrica, 45*, 463- 85.

Phillips, P. C. B. (1977b). A general theorem in the theory of asymptotic expansions as approximations to the finite sample distributions of econometric estimators. *Econometrica, 45,* 1517-34.

Phillips, P. C. B. (1978). Edgeworth and saddlepoint approximations in a first order autoregression. *Biometrika, 65*, 91-8.

Stuart, A. y Ord, J. K. (1987). *Kendall's Advance Theory of Statistics, Vol. 1.* Charles Griffin and Company Limited. London.

Watson, G. N. (1948). *Theory of Bessel Functions.* Cambridge University Press. Cambridge.