

Marginal Distributions of Maximum-likelihood estimator in non-standard conditions

*Distribuzioni Marginali dello stimatore di massima verosimiglianza in
condizioni non-standard*

Marco Barnabani

Dipartimento di Statistica “G. Parenti”, Università di Firenze

Riassunto: Quando il vero parametro è un punto di frontiera, la distribuzione asintotica dello stimatore di massima verosimiglianza è data da quella di un vettore casuale che minimizza una funzione quadratica stocastica su un cono convesso, Λ , dove i coefficienti della funzione quadratica hanno una distribuzione normale multivariata. Se il cono è dato da $\Lambda = (\mathbb{R}^+)^2 \times \mathbb{R}^q$ la soluzione analitica è nota e dipende dalla regione nella quale cadono i coefficienti stocastici della funzione quadratica. Limitatamente al caso in cui tali coefficienti cadano nel primo ortante in questo lavoro si analizza la densità marginale delle componenti dello stimatore di massima verosimiglianza. Si mostra che tali funzioni di densità sono troncate normali o normali non troncate moltiplicate per delle “funzioni asimmetriche” (“skew-functions”) che modificano la simmetria della normalità.

Keywords: maximum likelihood estimator, convex cone, skew-function, truncated normal, marginal density function.

1. Introduction

The standard theory based on maximum likelihood estimators assumes that the true parameter, θ_0 , lies inside an open set in the parameter space $\Theta \subset \mathbb{R}^k$. This assumption allows one to make use of the fact that the first order conditions hold, at least asymptotically. When the true parameter lies on the boundary of the parameter space, the asymptotic properties of maximum likelihood estimator are no more valid, in particular the asymptotic distribution is not normal but a mixture of distributions. The problems connected to this type of “non-regularity” has been considered by several authors, Chernoff (1954), Moran (1971), Chant (1974), Shapiro (1985), Self and Liang (1987) and recently Andrews (1999). Let $l_n(\theta)$ denote the log-likelihood and let $\hat{\theta}_n = \arg \max_{\theta \in \Theta} l_n(\theta)$ be the maximum likelihood estimator over Θ . The main result of these authors is that under general conditions, if the shifted and rescaled parameter space, $n^{1/2}(\Theta - \theta_0)$, can be approximated by a convex cone, Λ , then $n^{1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \hat{T}$ with $\hat{T}$ such that $q(\hat{T}) = \inf_{T \in \Lambda} (T - Z)' B (T - Z)$ where $Z \sim N_k(0, B^{-1})$, B is the Fisher information matrix in an observation.

In sum the asymptotic distribution of $\hat{\theta}_n$ is given by that of a random vector $\hat{T}$ that minimizes a stochastic quadratic function over a convex cone Λ where the coefficients of the quadratic function have a multivariate normal distribution. Clearly the vector $\hat{T}$ is the projection of Z onto the convex cone Λ with respect to the metric B .

Although the structure of the solution is clear, the definition of Λ might make the solution rather arduous. Later on we shall confine to the relatively simple (polyhedral) cone given by $\Lambda = (\mathbb{R}^+)^2 \times \mathbb{R}^q$, $q + 2 = k$ with the first two components of θ_0 equal to

zero, which holds in many practical situations. Let $Z = [Z'_1, Z'_2]'$, $Z_1 \in \mathbb{R}^2$, $Z_2 \in \mathbb{R}^q$ be the partition of Z according to Λ . In this case the analytic solution of $\hat{T}$ is well known and may be written as (Self and Liang, 1987) $[Z'_1, Z'_2]'$, $[0, Z'_{(1)}]'$, $[Z_{(2)}[1], 0, Z'_{(2)}]'$, $[0', Z'_{(0)}]'$ depending on which of the four regions of the cone Z_1 falls. Here $Z_{(1)}$ is a vector whose i th component is $Z[i] - (b^{i1}/b^{11})Z[1]$, $i = 2, \dots, k$; $Z_{(2)}[1] = Z[1] - (b^{12}/b^{22})Z[2]$; $Z_{(2)}$ is a vector with i th component given by $Z[i] - (b^{i2}/b^{22})Z[2]$, $i = 3, \dots, k$; $Z_{(0)} = Z_2 - B^{21}(B^{11})^{-1}Z_1$, B^{ij} is a block of the matrix B^{-1} according to the partition of Z , b^{ij} is the (i, j) th element of B^{-1} and $Z[i]$ denotes the i th element of Z . The distribution of $\hat{T}$ has been studied in Moran (1971) but nothing is said about the marginal density functions of the estimator. Our result is that the marginal densities of the component of the maximum likelihood estimator are truncated or untruncated normal multiplied by "skewing functions" that perturb the symmetry of the normal density. Some of these components are "standard skew-normal" (Azzalini, 2005). In this paper we confine to the first component $[Z'_1, Z'_2]'$ which occurs when $Z_1 > 0$. We show that the marginal densities of the components of Z_1 are truncated normal, the components of Z_2 are normal, both multiplied by "skew functions" easy to evaluate using available computer algorithms.

2. Marginal Density Functions

The density of $Z/Z_1 > 0$ is clearly truncated normal,

$$f_Z(z, 0, B^{-1}, z_1 > 0) = \frac{\exp\left(-\frac{1}{2} z' B z\right) dz}{\iint_{0_2} \left[\int \cdots \int_{I_q} \exp\left(-\frac{1}{2} z' B z\right) dz_2 \right] dz_1}; \quad \begin{matrix} z_1 \in (\mathbb{R}^+)^2 \\ z_2 \in \mathbb{R}^q \end{matrix} \quad (1)$$

where $0_2 = \{z_1; 0 < z[i] < +\infty; i = 1, 2\}$ and $I_q = \{z_2; -\infty < z[i] < +\infty; i = 3, \dots, k\}$. To obtain an explicit solution of the marginal density functions we must compute first the denominator of (1). This can be done following two steps.

Step 1: we apply Theorem 10.5.1 of Graybill (1983) to the square brackets of (1) with $(1/2)z'Bz$ replaced by $(1/2)z'_2 B_{22} z_2 + z'_2 B_{21} z_1 + (1/2)z'_1 B_{11} z_1$ where B_{ij} is a block of the matrix B according to the partition of Z . We get, $(2\pi)^{(k-2)/2} |B_{22}|^{-1/2} \exp\left(-\frac{1}{2} z'_1 U z_1\right)$ with $U = B_{11} - B_{12} B_{22}^{-1} B_{21}$.

Step 2: Sheppard's result is applied to compute the double integral between 0 and $+\infty$ of the solution of Step 1. The denominator denoted with D is given by $D = (2\pi)^{k/2} |B|^{-1/2} \frac{1}{2} \left(1 - \frac{\arccos r_{12}}{\pi}\right)$ where r_{12} is the correlation between $Z[1]$ and $Z[2]$.

To derive the marginal density functions of $Z[i]$, $i = 1, \dots, k$ we consider the components of Z_1 and Z_2 separately. Let begin with $Z[1]$ the first component of Z_1 . We must solve the multiple integral, $\int_0^{+\infty} \left[\int \cdots \int_{I_q} \exp\left(-\frac{1}{2} z' B z\right) dz_2 \right] dz[2]$. A solution can be obtained applying Graybill's Theorem twice. First, to the expression in square brackets. Second, to the solution so obtained integrating between 0 and $+\infty$. By some algebra we can show that the marginal density of $Z[1]$ is given by,

$$f_{Z[1]}(u^{1/2} z[1]) = \frac{2u^{1/2} \phi(u^{1/2} z[1]) \Phi(\alpha z[1])}{1 - \frac{\arccos r_{12}}{\pi}} := \phi(u^{1/2} z[1]) g_1(z[1]); \quad z[1] > 0$$

where $(B^{11})^{-1} := \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $u = a_{11} - a_{12}(a_{22})^{-1}a_{21}$, $\alpha = -(a_{22})^{-1/2}a_{21}$. $\phi(\cdot)$ and

$\Phi(\cdot)$ are the standard normal pdf and the standard normal cumulative distribution function, respectively.

The marginal density of the second component of Z_1 , $Z[2]$, can be obtained in the same way once the matrix B has been modified changing the first row with the second, the first column with the second. We have the following result,

$$f_{Z[2]}(v^{1/2}z[2]) = \frac{2v^{1/2}\phi(v^{1/2}z[2])\Phi(\beta z[2])}{1 - \frac{\arccos r_{12}}{\pi}} := \phi(v^{1/2}z[2])g_2(z[2]); \quad z[2] > 0$$

where $v = a_{22} - a_{21}(a_{11})^{-1}a_{12}$, $\beta = -(a_{11})^{-1/2}a_{12}$.

Consider the marginal density of a component of the vector Z_2 . Without loss of generality let us derive the marginal density of the last component of Z_2 denoted with Z_k . With regard to this assume the following partitions of Z and B :

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} := \begin{bmatrix} Z_1 \\ Z_2^* \\ Z_k \end{bmatrix} := \begin{bmatrix} Z_1^* \\ (k-1) \times 1 \\ 1 \times 1 \end{bmatrix}$$

lumping in with Z^* any component different from the last component Z_k , and

$$B = \begin{bmatrix} B_{11} & b_{12} \\ (k-1) \times (k-1) & (k-1) \times 1 \\ b_{21} & b_{22} \\ 1 \times (k-1) & 1 \times 1 \end{bmatrix} \quad \text{with} \quad B_{11} = \begin{bmatrix} C_{11} & C_{12} \\ 2 \times 2 & 2 \times (k-3) \\ C_{21} & C_{22} \\ (k-3) \times 2 & (k-3) \times (k-3) \end{bmatrix}$$

To derive the marginal density of Z_k , f_{Z_k} , we must integrate out the remaining variables of the numerator of (1). Denote $I_{q-1} = \{z_2; -\infty < z[i] < +\infty; i = 3, \dots, k-1\}$, then by Graybill's theorem we have

$$\begin{aligned} & \iint_{0_2} \int \cdots \int_{I_{q-1}} \exp\left(-\frac{1}{2} z' B z\right) dz^* \\ &= \iint_{0_2} \int \cdots \int_{I_{q-1}} \exp\left[-\left(z^{*'} \frac{B_{11}}{2} z^* + z^{*'} b_{12} z_k + \frac{b_{22}}{2} z_k^2\right)\right] dz^* \\ &= \exp\left[-\frac{1}{2} (b_{22} - b_{21} B_{11}^{-1} b_{12}) z_k^2\right] \iint_{-c_1 z_k}^{+\infty} \int \cdots \int_{I_{q-1}} \exp\left(-\frac{1}{2} y' B_{11} y\right) dy \end{aligned}$$

with $c z_k = -B_{11}^{-1} b_{12} z_k = \begin{bmatrix} c_1 \\ c_2^* \end{bmatrix} z_k$ and $y = z^* - c = \begin{bmatrix} y_1 \\ y_2^* \end{bmatrix}$ according to the partition of Z^* . Moreover,

$$\begin{aligned} & \iint_{-c_1 z_k}^{+\infty} \int \cdots \int_{I_{q-1}} \exp\left(-\frac{1}{2} y' B_{11} y\right) dy \\ &= \iint_{-c_1 z_k}^{+\infty} \int \cdots \int_{I_{q-1}} \exp\left[-\left(y_2^{*'} \frac{C_{22}}{2} y_2^* + y_2^{*'} C_{21} y_1 + y_1' \frac{C_{11}}{2} y_1\right)\right] dy_2^* dy_1 \end{aligned}$$

Graybill's theorem applied to the integral with respect to y_2^* gives the following result

$$\exp\left(-\frac{1}{2} y_1' V y_1\right) (2\pi)^{\frac{k-3}{2}} |C_{22}|^{-\frac{1}{2}}; \quad V = C_{11} - C_{12} C_{22}^{-1} C_{21}$$

that must be integrated from $-c_1 z_k$ to $+\infty$. Therefore, the numerator of f_{Z_k} is given by

$$\exp \left[-\frac{1}{2} (b_{22} - b_{21} B_{11}^{-1} b_{12}) z_k^2 \right] (2\pi)^{\frac{k-3}{2}} |C_{22}|^{-\frac{1}{2}} \int_{-c_1 z_k}^{+\infty} \exp \left(-\frac{1}{2} y_1' V y_1 \right) d y_1$$

Some algebra applied to the ratio between this result and the denominator, D , gives the marginal density,

$$f_{Z_k} (w^{1/2} z_k) = \frac{2 w^{1/2} \phi (w^{1/2} z_k) \Phi_2 (c_1 z_k)}{1 - \frac{\arccos r_{12}}{\pi}} := \phi (w^{1/2} z_k) g_3 (z_k); \quad z_k \in \mathbb{R} \quad (2)$$

where $\Phi_2 (c_1 z_k) = (2\pi)^{-1} (\det V)^{1/2} \int_{-c_1 z_k}^{+\infty} \exp \left(-\frac{1}{2} y_1' V y_1 \right) d y_1$ and $w = b_{22} - b_{21} B_{11}^{-1} b_{12}$.

The marginal density of the i th component of Z_2 is still given by (2) once the matrix B has been modified changing the i th row with the k th row and the i th column with the k th column.

The functions $g_1 (z[1])$, $g_2 (z[2])$ and $g_3 (z_k)$ can be thought of as "skew functions" that perturb the symmetry of the normal density functions.

3. Conclusions

In this paper we analyzed the marginal density functions of $[Z_1' Z_2']' / Z_1 > 0$. We showed that the marginal densities of the components of Z_1 are truncated normal, those of the components of Z_2 are normal, both multiplied by "skew functions" which perturb the symmetry of the normality. These "skew functions" are easy to evaluate using available computer algorithms.

References

- Andrews D. W. K. (1999) Estimation when a Parameter is on a Boundary, *Econometrica*, 67, 1341–1383.
- Azzalini A. (2005) The Skew-normal Distribution and Related Multivariate Families, *Scandinavian Journal of Statistics*, 32, 159–188.
- Chant D. (1974) On Asymptotic Tests of Complete Hypotheses in Nonstandard Conditions, *Biometrika*, 61, 291–298.
- Chernoff H. (1954) On the Distribution of the Likelihood Ratio, *Annals of Mathematical Statistics*, 25, 573–578.
- Graybill F. A. (1983) *Matrices with Applications in Statistics*, Second Edition, Belmont, Wadsworth.
- Moran P. A. P. (1971) Maximum Likelihood Estimators in Nonstandard Conditions, *Proceedings of The Cambridge Philosophical Society*, Cambridge Philosophical Society, Cambridge UK, 70, 441–450.
- Self S. G., Liang K. Y. (1987) Asymptotic Properties of Maximum Likelihood Estimators and Likelihood ratio Tests under Nonstandard Conditions, *Journal of the American Statistical Association*, 82, 605–610.
- Shapiro A. (1985) Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures under Inequality Constraints, *Biometrika*, 72, 133–144