A note on regions of given probability for the extended skew-normal distribution

Published version:

DOI:10.1080/03610926.2013.788710

(Article begins on next page)

Noname manuscript No.
(will be inserted by the editor)

A note on regions of given probability of the extended skew-normal distribution.

Antonio Canale

Received: December 2012 / Accepted: March 2013

Abstract The construction of regions with assigned probability p and minimum geometric measure has theoretical and practical interests, such as the construction of tolerance regions. Following Azzalini (2001) and exploiting the normal approximation of the extended skew-normal distribution when some of its parameters go to infinity, we discuss an approach for the construction of regions with assigned probability p for the bivariate extended skew-normal distribution.

Keywords skew-normal · tolerance regions · normal approximation

1 Introduction

Usual normality assumptions are unrealistic in a lot of concrete situations and for this reason the construction of asymmetric distributions, like the skew-normal of Azzalini (1985), has received more and more attention in the last two decades. Recently the skew-normal has been successfully applied in many fields including finance (Adcock, 2010), spatial analysis (Zhang and El-Shaarawi, 2010), survival analysis (Callegaro and Iacobelli, 2012), and insurance (Vernic, 2006).

Here we deal with the construction of regions with assigned probability p and minimum geometric measure. Such regions are very important in practice for the construction of tolerance regions, i.e. random regions with applications in quality control, industry, medicine, and environmental monitoring among others. See the book of Krishnamoorthy and Mathew (2009) for a recent discussion on the topic. The present note extends the work of Azzalini (2001) in which the author studies the construction of regions with given probability p and minimum volume under the assumption of skew-normality. A well known result for the normal distribution is that if X is a d-dimensional normal with correlation matrix Ω and c_p is the p-th quantile of a χ^2_d distribution, then

$$R_N = \{x : x^T \Omega^{-1} x \leq c_p\} \tag{1}$$

is the region with minimum volume and probability p. Our generalization substitutes the normal distribution with the extended skew-normal (ESN) one. The ESN family of distributions is introduced in the seminal

A. Canale
Department of Economics and Statistics, University of Torino and Collegio Carlo Alberto
E-mail: antonio.canale@unito.it

paper of Azzalini (1985) and it has been extensively studied in literature (Arnold et al., 1993; Arnold and Beaver, 2000; Capitanio et al., 2003; Canale, 2011). A d-dimensional random variable Z is distributed as a ESN with position vector $\xi \in \mathbb{R}^d$, positive definite $d \times d$ scale matrix Ω, shape vector $\alpha \in \mathbb{R}^d$, and truncation $\tau \in \mathbb{R}$, written $Z \sim ESN(\xi, \Omega, \alpha, \tau)$ if its probability density function (pdf) is

$$f_{ESN}(z) = \phi_d\left(z - \xi; \Omega\right) \Phi\left(\alpha_0(\tau) + \alpha^T \omega^{-1}(z - \xi)\right) / \Phi(\tau), \tag{2}$$

where $\alpha_0(\tau) = \tau(1 + \alpha^T \Omega^{-1} \alpha)^{1/2}$, ω is the diagonal matrix formed by standard deviations of Ω, $\phi_d(\cdot)$ is the d-dimensional standard normal pdf and $\Phi(\cdot)$ is the cumulative distribution function (cdf) of a standard normal. If $\tau = 0$, the density belongs to the skew-normal (SN) family of distribution, i.e. $Z \sim SN(\xi, \Omega, \alpha)$.

To determine regions with given probability and minimum volume, the work of Azzalini (2001), that we aim to extend here, considers the region

$$R_{SN} = \{z : f_{SN}(z) \geq f_0\}, \tag{3}$$

and discusses how to choose the value f_0 in function of p and of the parameters of the model.

In Section 2 the results of Azzalini (2001) are shortly recalled while in Section 3 regions of the form

$$R_{ESN} = \{z : f_{ESN}(z) \geq f_0\}, \tag{4}$$

are introduced. Unfortunately the construction of such regions for the ESN distribution has a clear and intrinsic bottleneck: if $Z \sim ESN(\xi, \Omega, \alpha, \tau)$, the quadratic form $Z^T \Omega^{-1} Z$ is not distributed as χ^2_d when $\tau \neq 0$. Section 3.1 bypasses this problem using a new result on the approximation of ESN densities when $|\tau| \to \infty$ while in Section 3.2 a direct modification of the approach of Azzalini (2001) is discussed.

2 Bivariate regions for the SN model

Rewriting the quadratic form $Z^T \Omega^{-1} Z$ when Z is normally distributed as

$$2 \log \phi_d(z; \Omega) - d \log(2\pi) - \log |\Omega|,$$

Azzalini (2001) suggests to consider the analogous expression for the skew-normal case and to let

$$R_{SN} = \{z : 2 \log f_{SN}(z) \geq -c_p - d \log(2\pi) - \log |\Omega|\} \tag{5}$$

as a candidate solution to the problem.

Empirical inspections of rule (5) suggest that an additive correction term h should be inserted with such h depending on α. Simulations and graphical inspections of the relations between $(\log(e^{h/2} - 1))^{-1}$ and $\alpha^* = \sqrt{\alpha^T \Omega \alpha}$, suggest a proportional relation of ratio -0.6478 leading to define

$$\hat{h} = 2 \log(1 + \exp\{-1.544/\alpha^*\}) \tag{6}$$

and hence to correct equation (5) with

$$R_{SN} = \{z : 2 \log f_{SN}(z) \geq -c_p - d \log(2\pi) - \log |\Omega| + \hat{h}\}. \tag{7}$$

See Azzalini (2001) for further insights and details.

Table 1 Nominal and actual values of the coverage probability using rule (10) when $\alpha_1 = 2$, $\alpha_2 = 6$, $\omega = -0.5$, and (a) $\tau = -10$, (b) $\tau = -2$, (c) $\tau = 1$, (d) $\tau = 4$.

p	0.99	0.975	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.025	0.01
$\tilde{p}$ (a)	0.989	0.973	0.946	0.893	0.790	0.687	0.489	0.292	0.193	0.097	0.049	0.024	0.010
$\tilde{p}$ (b)	0.959	0.928	0.886	0.818	0.702	0.598	0.411	0.240	0.158	0.078	0.039	0.020	0.008
$\tilde{p}$ (c)	0.993	0.982	0.962	0.922	0.836	0.746	0.549	0.339	0.228	0.114	0.058	0.030	0.012
$\tilde{p}$ (d)	0.990	0.975	0.950	0.901	0.801	0.701	0.501	0.299	0.199	0.099	0.050	0.025	0.010

3 Bivariate regions for the ESN model

3.1 Limiting cases: $\tau \to \pm\infty$

Let further assume, without loss of generality, that $\xi = (0, 0)$ and that the matrix Ω has unit diagonal and off diagonal elements equal to ω.

Canale (2011) showed that for $\tau \to \pm\infty$ the pdf of the scalar ESN approaches that of a normal with suitable parameters. Exploiting this behavior we approximate the density of the bivariate ESN with its limit normal density. Particularly for $\tau \to +\infty$ the following approximate result holds

$$Z \dot{\sim} N_2(0, \Omega), \tag{8}$$

while for $\tau \to -\infty$

$$Z + \delta\tau \dot{\sim} N_2(0, \tilde{\Omega}), \ \tilde{\Omega} = \begin{pmatrix} 1 - \delta_1^2 & \delta_1\delta_2 - \omega \\ \delta_1\delta_2 - \omega & 1 - \delta_2^2 \end{pmatrix}, \tag{9}$$

where $\delta = (1 + \alpha^T \Omega \alpha)^{-1/2} \alpha^T \Omega$, and $\delta = (\delta_1, \delta_2)^T$. The derivation of (8) and (9) is of independent interest and it is reported in the Appendix.

Regions with approximate probability p can be constructed using previous results and the rule

$$R_{ESN} = \{z : 2 \log f_i(z) \geq -c_p - d \log(2\pi) - \log |\Omega_i|\} \tag{10}$$

where $f_i(\cdot)$ and Ω_i are suitable normal pdf and matrices related to expressions (8) and (9). Henceforth we will refer to rule (10) using the convention to use approximation (8) for $\tau > 0$ and (9) for $\tau < 0$.

To test the performance of rule (10) we conducted a simulation experiment similar to the one of Azzalini (2001). Various parameters combinations have been selected and, for each choice, 10^6 replicated samples have been generated from the given ESN distribution. Applying rule (10) to the set of p values

$$p = (0.99, 0.975, 0.95, 0.90, 0.80, 0.70, 0.50, 0.30, 0.20, 0.10, 0.05, 0.025, 0.01),$$

we obtained a vector of observed relative frequencies $\tilde{p}$ reported in Table 1. Figure 1 shows the main features of the results. Circles are for the points $(c_p, c_{\tilde{p}})$, where $c_{\tilde{p}}$ denotes the quantile function of a χ^2_2 evaluated at $\tilde{p}$.

As expected, rule (10) performs well for large $|\tau|$ while it has poor performance for τ close to zero. Ideally the points in Figure 1 should lie on the dashed line corresponding to the identity function as in panel (a) and (d). Evidently rule (10) can be used only for large values of $|\tau|$.

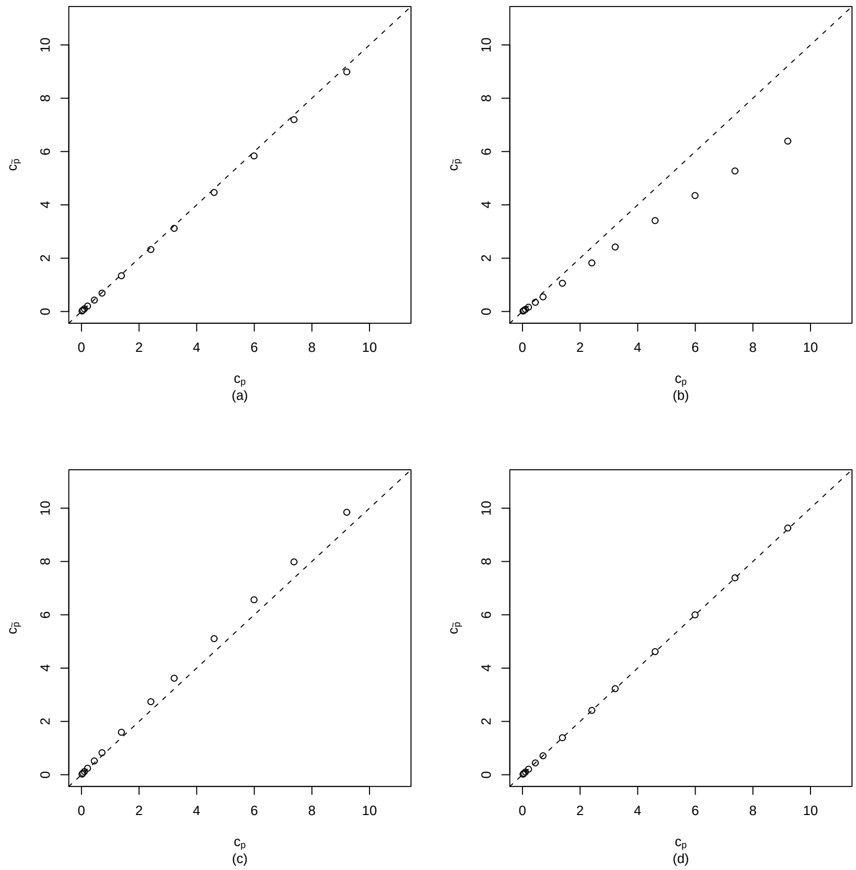

Fig. 1 Actual versus nominal values of the probability p using rule (10) transformed to the quantile scale. $\alpha_1 = 2$, $\alpha_2 = 6$, $\omega = -0.5$, and (a) $\tau = -10$, (b) $\tau = -2$, (c) $\tau = 1$, and (d) $\tau = 4$.

3.2 Cases for small $|\tau|$

In order to generalize rule (5), we applied it to 1250 samples drawn from ESN with different combinations of α and τ parameters. Figure 2 reports the results.

As for the SN we observe a relation

$$c_{\tilde{p}} = h + c_p$$

where here $h = h(\alpha^*, \tau)$ is a function of both α^* and τ. Figure 3 (a) shows the value of h versus α^* for the 1250 simulated samples. Different point shapes represent different values of τ.

The relation between h and (α, τ) seems to be approximate by (6) times a coefficient k depending on τ, i.e.

$$\hat{h} = k(\tau) 2 \log(1 + \exp\{-1.544/\alpha^*\}). \quad (11)$$

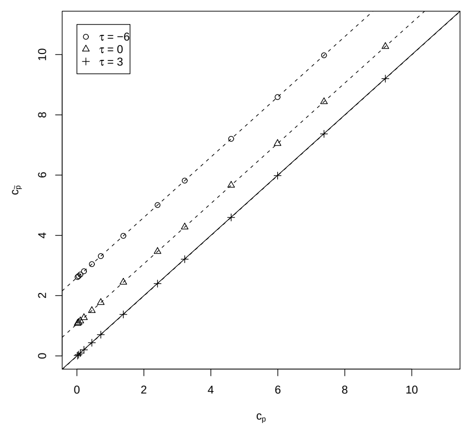

Fig. 2 Actual versus nominal values of the probability p, transformed to the quantile scale using rule (5) for $\alpha_1 = 2$, $\alpha_2 = 6$, $\omega = -0.5$ and for $\tau = -6$ (circles), $\tau = 0$ (triangles), and $\tau = 3$ (crosses)

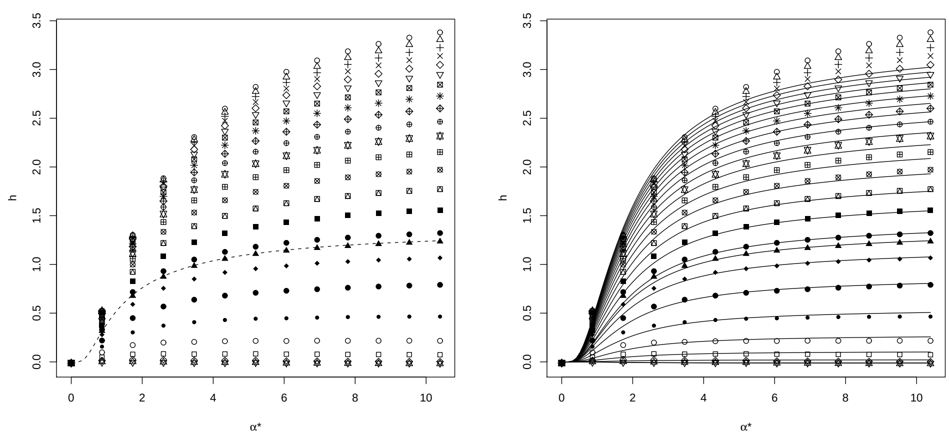

Fig. 3 Observed values of h plotted versus α^* and (a) interpolating function (6), (b) interpolating function (11) calculated for a grid of values of τ between -5 and 5, from top to the bottom.

To obtain an approximation for $k(\tau)$, we computed the ratio between $h(\tau_i)$ and $h(0)$ for every α^* of Figure 3. Then for each τ we compute k as

$$k(\tau_i) = \frac{1}{l} \sum_{j=1}^{l} \frac{h(\tau_i, \alpha_j^*)}{h(0, \alpha_j^*)}, \tag{12}$$

where l is the number of α^* values, here $l = 25$.

To get an expression for $k(\tau)$ we plotted the values obtained through (12) against τ. Several attempts have been made to find an accurate approximation, leading to

$$\hat{k} = \frac{\pi^2}{4}\zeta_2(\tau)^2, \tag{13}$$

where $\zeta_i(\cdot)$ is the ith derivative of $\zeta_0 = \log(2\Phi(\cdot))$. Note that for $\tau = 0$, $\hat{k} = 1$.

Using (13) we obtain Figure 3 (b) which gives a resonable approximation for negative τ and an excellent approximation for positive τ. This behavior can be explained noting that h goes to zero for $\tau \to +\infty$, i.e. when the density approaches the normal. For some negative values of τ, expression (11) over/under-estimates h for small and large values of α^* respectively. Figure 4 and Table 2 show the results obtained with (13) for the same samples of Figure 1.

Table 2 Nominal values of probability p versus actual values using rule (11) when $\alpha_1 = 2$, $\alpha_2 = 6$ and $\omega = -0.5$ for the same samples used in Figure 1 with (a) $\tau = -10$, (b) $\tau = -2$, (c) $\tau = 1$, and (d) $\tau = 4$.

p	0.990	0.975	0.950	0.900	0.800	0.700	0.500	0.300	0.200	0.100	0.050	0.025	0.010
$\tilde{p}$ (a)	0.991	0.979	0.958	0.915	0.834	0.752	0.590	0.431	0.351	0.272	0.232	0.213	0.201
$\tilde{p}$ (b)	0.987	0.969	0.940	0.883	0.775	0.673	0.482	0.303	0.219	0.135	0.094	0.074	0.062
$\tilde{p}$ (c)	0.992	0.979	0.958	0.914	0.824	0.731	0.540	0.330	0.219	0.099	0.039	0.008	0.000
$\tilde{p}$ (d)	0.990	0.974	0.950	0.899	0.799	0.698	0.499	0.299	0.200	0.101	0.050	0.025	0.010

4 Discussion

The approach discussed in Section 3.2 is more accurate than the approach of Section 3.1 for small values of τ. Despite this general behavior, the method introduced in Section 3.1 has dramatically better performance for large $|\tau|$, namely when a quadratic form of Z is approximately chi-squared distributed. From a practical viewpoint the choice between the former or the latter approach has to be done on a case by case basis. Specifically, for moderate or large positive τ (say $\tau \geq 3$) one should use (10) and approximation (8), for large negative τ (say $\tau \leq -6$) rule (10) and approximation (9), and the rule discussed in Section 3.2 otherwise. Note that these limits are merely qualitative.

Concerning the generalization for d other than 2 some problems may arise in using the approaches previously discussed. For $d = 1$ the required region is just an interval and we can obtain the solution via numerical methods. For $d > 2$, we investigated suitable modifications of rule (7) and behaviors similar to those described in Section 3.2 were noticed for $d = 3, 4$. On the other side, the normal approximation used in Section 3.1 is available in closed form only for $d = 2$. The existence of such an approximation for $d > 2$ is, in fact, very likely since it arises from the truncation representation of the ESN extensively discussed by Arnold et al. (1993). However the lack of an explicit expression does not allow us to study approaches similar to that of Section 3.1 for $d > 2$. Clearly our contribution to the topic discussed by Azzalini (2001) is still merely approximate, however it can be of interest at least from a practical viewpoint and we hope it can stimulate further studies involving the ESN model.

Acknowledgements This work is part of the master thesis of the author. The author would like to thank his advisor, professor Adelchi Azzalini, for the comments he kindly offered in early versions of the manuscript.

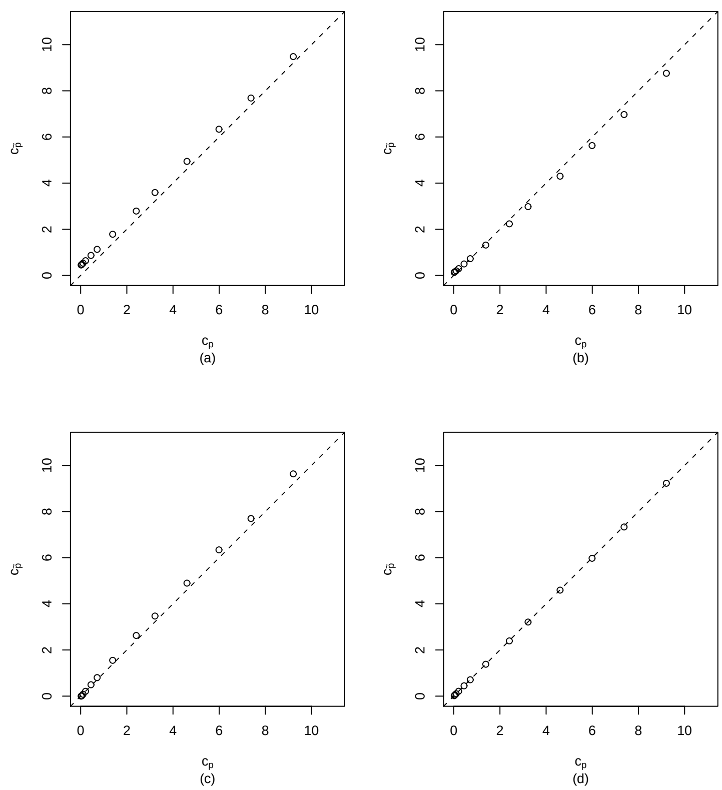

Fig. 4 Actual versus nominal values of the probability p using rule (11) transformed to the quantile scale. $\alpha_1 = 2$, $\alpha_2 = 6$, $\omega = -0.5$, and (a) $\tau = -10$, (b) $\tau = -2$, (c) $\tau = 1$, and (d) $\tau = 4$.

References

Adcock, C. J. (2010). Asset pricing and portfolio selection based on the multivariate extended skew-student-t distribution. *Annals of Operations Research*, 176:221–234.

Arnold, B. C. and Beaver, R. J. (2000). Hidden truncation model. *Sankhyā, series A*, 62:22–35.

Arnold, B. C., Beaver, R. J., Groeneveld, R. A., and Meeker, W. Q. (1993). The non truncated marginal of a truncated bivariate normal distribution. *Psychometrika*, 58:471–488.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, 12:171–178.

Azzalini, A. (2001). A note on regions of given probability of the skew-normal distribution. *Metron*, 59(3-4):27–34.

Callegaro, A. and Iacobelli, S. (2012). The Cox shared frailty model with log-skew-normal frailties. *Statistical Modelling*, 12:399–418.

Canale, A. (2011). Statistical aspects of the scalar extended skew-normal distribution. *Metron*, LXIX(3):279–295.

Capitanio, A., Azzalini, A., and Stanghellini, E. (2003). Graphical models for skew-normal variates. *Scand. J. Statist.*, 30:129–144.

Krishnamoorthy, K. and Mathew, T. (2009). *Statistical Tolerance Regions: Theory, Applications, and Computation*. Wiley Series in Probability and Statistics. Wiley.

Vernic, R. (2006). Multivariate skew-normal distributions with applications in insurance. *Insurance: Mathematics and Economics*, 38:413–426.

Zhang, H. and El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics*, 21(1):33–47.

Appendix

Details on how to obtain equations (8) and (9) are discussed below. These limit expressions are of independent interest since they provide a generalization for $d = 2$ of some results discussed by Canale (2011). Say $f_{ESN}(z; \Omega, \alpha, \tau)$ the bivariate ESN density function. For $\tau \to +\infty$, one easily gets

$$\lim_{\tau\to+\infty} f_{ESN}(z;\Omega,\alpha,\tau) = \lim_{\tau\to+\infty} \phi_2(z;\Omega)\Phi(\tau(1+\alpha^T\Omega\alpha)^{1/2}+\alpha^T z)/\Phi(\tau) = \phi_2(z;\Omega).$$

For $\tau \to -\infty$, using de l'Hospital Theorem, we have

$$\begin{aligned}
\lim_{\tau\to+\infty} f_{ESN}(z;\Omega,\alpha,\tau) &= \lim_{\tau\to-\infty} \phi_2(z;\Omega)\Phi(\tau(1+\alpha^T\Omega\alpha)^{1/2}+\alpha^T z)/\Phi(\tau) \\
&= (1+\alpha^T\Omega\alpha)^{1/2} \lim_{\tau\to-\infty} \phi_2(z;\Omega)\phi(\tau(1+\alpha^T\Omega\alpha)^{1/2}+\alpha^T z)/\phi(\tau) \\
&= \frac{1}{(1-\delta^T\Omega^{-1}\delta)^{1/2}} \lim_{\tau\to-\infty} \phi_2(z;\Omega)\phi(\tau(1+\alpha^T\Omega\alpha)^{1/2}+\alpha^T z)/\phi(\tau) \\
&= \frac{1}{(1-\delta^T\Omega^{-1}\delta)^{1/2}} \lim_{\tau\to-\infty} \frac{1}{2\pi|\Omega|} \exp\left\{-\frac{1}{2} z^T\Omega^{-1}z\right\} \times \\
&\quad \exp\left\{-\frac{1}{2}\left(\tau(1+\alpha^T\Omega\alpha)^{1/2}+\alpha^T z\right)^2\right\} \exp\left\{\frac{\tau^2}{2}\right\} \\
&= \frac{1}{(1-\delta^T\Omega^{-1}\delta)^{1/2}} \frac{1}{2\pi|\Omega|} \lim_{\tau\to-\infty} \exp\left\{-\frac{1}{2}\mathrm{A}\right\},
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{A} &= z^T\Omega^{-1}z + \left(\tau(1+\alpha^T\Omega\alpha)^{1/2}+\alpha^T z\right)^2 - \tau^2 \\
&= z^T\Omega^{-1}z + \tau^2\alpha^T\Omega\alpha + (\alpha^T z)^2 + 2\tau\alpha^T z(1+\alpha^T\Omega\alpha)^{1/2} \\
&= z^T\Omega^{-1}z + \tau^2\frac{\delta^T\Omega^{-1}\delta}{1-\delta^T\Omega^{-1}\delta} + \frac{(\delta^T\Omega^{-1}z)^2}{1-\delta^T\Omega^{-1}\delta} + 2\tau\frac{\delta^T\Omega^{-1}z}{1-\delta^T\Omega^{-1}\delta} \\
&= \frac{1}{1-\delta^T\Omega^{-1}\delta}\Big(z^T\Omega^{-1}z + (z^T\Omega^{-1}z\delta^T\Omega^{-1}\delta) + \tau^2\delta^T\Omega^{-1}\delta + \\
&\qquad (\delta^T\Omega^{-1}z)^2 + 2\tau\delta^T\Omega^{-1}z\Big).
\end{aligned}$$

Explicitly computing each quantity in equation (14) and letting $\delta = (\delta_1, \delta_2)^T$, $z = (z_1, z_2)^T$ we can recollect a quadratic form in z. Recalling the expressions for the scalar case presented by Canale (2011) we let $\mu = (-\delta_1\tau, -\delta_2\tau)^T$. This allow us to obtain

$$\tilde{\Omega} = \begin{pmatrix} 1-\delta_1^2 & \delta_1\delta_2-\omega \\ \delta_1\delta_2-\omega & 1-\delta_2^2 \end{pmatrix},$$

and to conclude that

$$\lim_{\tau\to+\infty} f_{ESN}(z+\delta\tau;\Omega,\alpha,\tau) = \phi_2(z;\tilde{\Omega}).$$