

ON THE EXPECTED ABSOLUTE VALUE OF A BIVARIATE NORMAL DISTRIBUTION

S. Reza H. Shojaie, Mina Aminghafari and Adel Mohammadpour*

Department of Statistics, Faculty of Mathematics and Computer Science,
Amirkabir University of Technology (Tehran Polytechnic)

Abstract

The expected absolute value of a bivariate normal distribution is calculated analytically, numerically, and through simulation. These solution methods may be introduced to undergraduate students so they will become familiar with their advantages.

Key Words and Phrases. Expected Absolute Value, Bivariate Normal, Simulation, Numerical Integration.

AMS 2000 Subject Classifications. 65C60

*Corresponding author (e-mail: adel@aut.ac.ir)

1 Introduction

Introductory courses in probability and statistics include joint distribution and calculation of bivariate expectation. Over several years of teaching experience, the difficulty of calculating complex integrals is observed for most students. When an expectation of a function of bivariate normal distribution (X, Y) , such as $E(|XY|)$, is calculated, it is difficult and the students are not confident of the result and consequently discouraging. From a pedagogical point of view, it is important to help students to use another way for calculating such an expectation instead of direct integration. In this work, we propose three different methods to calculate expectations of joint random variables using an interesting example. The first method is direct one, i.e. calculating analytically. This is a more difficult way for undergraduate students and a lot of mistakes may occur. In some cases, analytic solutions may not exist, they may be very difficult, or they cannot be calculated by computer algebra softwares. The second method is numerical calculation. When analytic solutions fail, a numerical method to approximate integrals may be used. Implementation of numerical methods are not easy for statistics students and computer numerical softwares are usually used by them. However, the third method, using simulation, is much simpler than the two previous ones. In this case, simulating realizations of a bivariate normal distribution, (x_i, y_i) , $i = 1, \dots, n$, and approximate $E(|XY|)$ by $\sum_{i=1}^n |x_i y_i|/n$ with the aid of the weak law of large number is proposed.

In the next three sections, expected absolute value of a bivariate normal distribution are calculated, analytically, numerically, and through simulation, respectively.

2 Analytical method

Let (X, Y) be jointly distributed according to the bivariate normal distribution. Without loss of generality, it is assumed that $E(X) = E(Y) = 0$. Considering that, $Var(X) = \sigma^2$, $Var(Y) = \gamma^2$, and $Corr(X, Y) = \rho$, then probability density function (PDF) of (X, Y) has the following form.

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma\gamma\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{x^2}{\sigma^2} + \frac{y^2}{\gamma^2} - 2\rho\frac{xy}{\sigma\gamma} \right) \right\}, \quad (1)$$

for all $x, y \in \mathbb{R}$, $\sigma, \gamma \in \mathbb{R}^+$, and $\rho \in [-1, 1]$. This PDF can be reconstructed in another way. Let Z_1, Z_2, Z_3 be independent standard normal random variables. It can be shown that PDF

of (X, Y) and that of

$$\left(\sigma(\sqrt{1-\rho}Z_1 - \sqrt{\rho}Z_3), \gamma(\sqrt{1-\rho}Z_2 - \sqrt{\rho}Z_3)\right),$$

are equal for $\rho \in [0, 1]$. Now, $(E|XY|)$ can be calculated by one of the following three ways:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |xy| f_{X,Y}(x, y) dx dy \tag{2}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\sigma(\sqrt{1-\rho}z_1 - \sqrt{\rho}z_3) \gamma(\sqrt{1-\rho}z_2 - \sqrt{\rho}z_3)| f_{Z_1}(z_1) f_{Z_2}(z_2) f_{Z_3}(z_3) dz_1 dz_2 dz_3 \tag{3}$$

$$\int_0^{\infty} \int_0^{\infty} xy f_{|X|,|Y|}(x, y) dx dy \tag{4}$$

Integral in (2) is calculated by Nabeya (1951). Triplet integral in (3) seems simpler than (2) by the independence of Z_i s, however it cannot be calculated analytically by a computer algebra software. Also we can obtain the absolute moments of a bivariate normal distribution using bivariate density of the absolute normal distribution, i.e. (4).

The rest of this section deals with calculating (4). The cumulative distribution function of $(|X|, |Y|)$ is given by

$$F_{|X|,|Y|}(x, y) = F_{X,Y}(x, y) - F_{X,Y}(x, -y) - F_{X,Y}(-x, y) + F_{X,Y}(-x, -y), \tag{5}$$

where $F_{X,Y}$ denotes the cumulative distribution function of (X, Y) . The PDF of $(|X|, |Y|)$ can be obtained by the mixed partial derivatives of (5) with respect to x and y as

$$f_{|X|,|Y|}(x, y) = f_{X,Y}(x, y) + f_{X,Y}(x, -y) + f_{X,Y}(-x, y) + f_{X,Y}(-x, -y). \tag{6}$$

By substituting (1) into (6) and after simplification, we arrive at PDF $(|X|, |Y|)$,

$$f_{|X|,|Y|}(x, y) = \frac{2}{\pi\sigma\gamma\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma^2} + \frac{y^2}{\gamma^2}\right)\right\} \cosh\left(\frac{\rho xy}{(1-\rho^2)\sigma\gamma}\right), \tag{7}$$

for all $x, y \in \mathbb{R}^+$. This result is a little different than Abdel-Hameed and Sampson (1978) expression. One can show that the appeared equation in Abdel-Hameed and Sampson (1978) p. 1363 and Tong (1990) p. 100, are not PDF's and cannot be corrected by a constant multiplier, but it may be modified by multiplier $\sqrt{(4-\rho^2)/(1-\rho^2)}/2$.

In calculating $E(|XY|)$, the PDF of an absolute bivariate normal random vector can be used as follows:

$$\begin{aligned}
E(|XY|) &= \int_0^\infty \int_0^\infty \frac{2xy}{\pi\sigma\gamma\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma^2} + \frac{y^2}{\gamma^2}\right)\right\} \\
&\quad \times \cosh\left(\frac{\rho xy}{(1-\rho^2)\sigma\gamma}\right) dx dy \\
&= \frac{\sigma\gamma}{\pi} \int_0^{\pi/2} \int_0^\infty (1-\rho^2)^{\frac{3}{2}} \frac{u}{(\rho \sin 2\theta - 1)^2} \left(\exp\{u\} \right. \\
&\quad \left. + \exp\left\{-\frac{\rho \sin 2\theta + 1}{\rho \sin 2\theta - 1} u\right\}\right) du d\theta \tag{8} \\
&= \frac{\sigma\gamma}{\pi} (1-\rho^2)^{\frac{3}{2}} \int_0^{\pi/2} \left\{ \frac{\sin 2\theta}{(\rho \sin 2\theta - 1)^2} + \frac{\sin 2\theta}{(\rho \sin 2\theta + 1)^2} \right\} d\theta \\
&= \frac{\sigma\gamma}{\pi} (1-\rho^2)^{\frac{3}{2}} \frac{d}{d\rho} \int_0^{\pi/2} \left\{ \frac{1}{1-\rho \sin 2\theta} - \frac{1}{1+\rho \sin 2\theta} \right\} d\theta \tag{9} \\
&= \frac{\sigma\gamma}{\pi} (1-\rho^2)^{\frac{3}{2}} \frac{d}{d\rho} \left\{ \frac{2}{\sqrt{1-\rho^2}} \arcsin \rho \right\} \\
&= \frac{2\sigma\gamma}{\pi} (\rho \arcsin \rho + \sqrt{1-\rho^2}), \tag{10}
\end{aligned}$$

where (8) is obtained with polar coordinate and an appropriate change of variable and (9) can be solved using relations (1.624) and (2.551) in Gradshteyn and Ryzhik (2007) pages 57 and 171, respectively. This integral can also be calculated by a computer algebra software such as *Mathematica* with the following commands:

```

Integrate[2 * x * y / (Pi * sigma * gamma * Sqrt[1 - rho^2]) * Cosh[rho * x * y / ((1 - rho^2) * sigma * gamma)] *
Exp[(x^2 / sigma^2 + y^2 / gamma^2) / (-2 + 2 * rho^2)], {x, 0, Infinity}, {y, 0, Infinity},
Assumptions -> {-1 < rho < 1, sigma > 0, gamma > 0}]

```

The reader can refer to Wellin et al. (2005) for more information about *Mathematica* commands.

3 Numerical method

When a problem cannot be solved analytically, numerical methods are employed. Numerical integration is the numerical approximation of the integral of a function. For a function of two variables it is equivalent to finding an approximation to the volume under the surface. For example, integral (3) can be solved by *Mathematica* (for fixed parameters) by **NIntegrate**

command. However, there are several popular software with numerical calculating ability, such as MATLAB, S-PLUS or R. The open source software, R is used in this study. For more details about R, the readers are referred to Everitt (2005).

In the following, R functions for calculating (2) and (4) will be introduced, respectively. The first function, `f1`, supposes that (X, Y) has a bivariate normal distribution with a mean 0, standard deviations `g` and `s`, and correlation `r`.

```
> f1=function(g,s,r){
+   integrate(function(y) {
+     sapply(y,function(y) {
+       integrate(function(x) {
+         sapply(x,function(x) abs(x*y)*exp(-((x/g)^2+(y/s)^2-2*r*x*y/(s*g))
+           /((2*(1-r^2)))/(2*g*s*pi*sqrt(1-r^2)))
+       }, -Inf, Inf)$value })
+     }, -Inf, Inf)}
> f1(1,1,0.5)
0.7179956 with absolute error < 1.0e-07
> f1(2,2,0.5)
2.871980 with absolute error < 3e-05
```

The second function, `f2`, supposes that (X, Y) has an absolute bivariate normal distribution. It has the same structure as `f1` with the following integrand over $[0, \infty) \times [0, \infty)$.

```
x*y*(exp(-((x/g)^2+(y/s)^2)/(2*(1-r^2))+r*(x/g)*(y/s)/(1-r^2))+exp(-((x
/g)^2+(y/s)^2)/(2*(1-r^2))-(x/g)*(y/s)*r/(1-r^2)))/(pi*g*s*sqrt(1-r^2)))
> f2(1,1,0.5)
0.7179956 with absolute error < 9e-08
> f2(2,2,0.5)
2.871982 with absolute error < 2.9e-06
```

The results of this function are approximately equal to the results of `f1`, and its exact value in (10).

4 Simulation method

An alternative method to find $E(|XY|)$ can be simulation. Let (X_i, Y_i) , $i = 1, \dots, n$ be a random sample of size n from a bivariate normal distribution. It is well known that, by the weak law of large number

$$\frac{1}{n} \sum_{i=1}^n |X_i Y_i| \longrightarrow E(|XY|),$$

in probability. Therefore, to calculate an approximation of $E(|XY|)$, realizations of the bivariate normal distribution are simulated and its sample mean absolute value is calculated. Here is an example of R code. A sample of observations having a size of 10^6 , from the desired bivariate normal distribution are simulated, e.g., with means 0, variances 1, and correlation 0.5 and sample mean absolute value is calculated as follows:

```
> library(MASS)           # loading package
> xy=mvrnorm(1000000,c(0,0),matrix(c(1,0.5,0.5,1),2))
                           # simulate bivariate normal observations
> mean(abs(xy[, 1]*xy[, 2])) # calculate sample mean absolute value
0.717534
```

The exact value, using (10), where $\rho = 0.5$ and $\sigma = \gamma = 1$, is equal to 0.7179956 which shows that simulation can be applied as well.

Acknowledgments. The authors would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions to improve the manuscript.

References

- [1] Abdel-Hameed, M. and Sampson, A. R. (1978), Positive Dependence of the Bivariate and Trivariate Absolute Normal, t , χ^2 and F Distributions, *The Annals of Statistics*, 6, 1360-1368.
- [2] Everitt, B. S. (2005), *An R and S-PLUS companion to multivariate analysis*, New York: Springer-Verlag.

- [3] Gradshteyn, I. S. and Ryzhik, I. M. (2007), *Table of Integrals, Series, and Products* (7th ed.), Amsterdam: Elsevier/Academic Press.
- [4] Nabeya, S. (1951), Absolute Moments in 2-dimensional Normal Distributions, *Annals of the Institute of Statistical Mathematics*, 3, 2-6.
- [5] Tong, Y. L. (1990), *The Multivariate Normal Distribution*, New York: Springer-Verlag.
- [6] Wellin, P. R., Gaylord, R. J., and Kamin, S. N. (2005), *An Introduction to Programming with Mathematica* (3rd ed.), New York: Cambridge University Press.