How Accurate Is Your Gaussian\Gamma Approximation?

Fayez Alejawey

***Abstract*—Dealing with random variables and their distributions in Telecommunication systems is mandatory for their analysis, where almost every system has a random behaviour that can be interpreted to a known distribution which can be achieved by applying the moment matching process. But even if we find our approximation for such a system there is an error produced in such a process. This letter will provide a simple and accurate tool to use for measuring the accuracy for one of the most common distributions that commonly used to represent a behaviour of a system which are the Gaussian and Gamma distributions. Furthermore, the concept of developing it can be extended to encompass other distributions. After developing the target tool, it will be applied to a real world application which is the distributed massive MIMO.**

***Index Terms*—Gaussian Distribution, Gamma Distribution, Edgeworth Series Expansion, Laguerre Series Expansion, MIMO**

I. Introduction

Every aspect in any Telecommunication system analysis must involve studying the randomness behaviour of it and this randomness can be characterized with a distribution. Moment matching is the process that links the unknown distribution with a known distribution (such as Gaussian, Gamma, ...etc) if and only if the first and second moments from the unknown distribution are matched to the first and second moments of the known distribution.

Applying the moment matching process guaranteed that the analyzed randomness is approximately behaved such as the matched distribution, but it is not telling that it has the same behaviour of the matched distribution. So, when we are dealing with approximations, error measurements always take a place to insure that this approximation is accurate enough for such a case.

Gaussian and Gamma distributions appear very frequently as a matched distributions for the most of the applications in Telecommunication. This letter is interested in measuring the accuracy for such approximations by proposing a tool that depends only on calculating the first four moments with a high accuracy of assessing.

II. Accuracy of The Gaussian Approximation

The Edgeworth series expansion can be considered as an improvement to the central limit theorem, where this improvement came from the fact that the error of such an approximation is controlled. Therefore, in this letter the use of such an expansion is adopted when measuring the accuracy for the Gaussian approximation.

Let V be the random variable we want to approximate and suppose after applying the moment matching process it suggests that a Gaussian distribution is a good approximation for it. Now, based on the original random variable we define the standardized random variable X as

$$X = \frac{V - v_1}{\sqrt{v_2}} \tag{1}$$

where v_1 and v_2 are the mean and variance for V, respectively.

Edgeworth expansion for the CDF of the normalized Gaussian random variable X can be written as

$$\begin{aligned} F_X(x) = \Phi_X(x) - [\frac{\mu_3}{6}(x^2-1) - \frac{(\mu_4-3)}{24}(x^3-3x) \\ - \frac{\mu_3^2}{72}(x^5-10x^3+15x)]\phi_X(x) + ... \end{aligned} \tag{2}$$

where $\Phi_X(x) = \int_{-\infty}^{x} \phi_X(\zeta)\, d\zeta$, $\phi_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, μ_3 and μ_4 are the CDF, the PDF, the third central moment and the fourth central moment of the normalized Gaussian random variable, respectively.

As X is tending to have the Gaussian shape, the error terms will tend to zero. This letter will consider the leading error terms which are

$$\begin{aligned} \Psi = \mid \frac{\mu_3}{6}(x^2-1)\phi_X(x) \mid + \mid \frac{(\mu_4-3)}{24}(x^3-3x)\phi_X(x) \mid \\ + \mid \frac{\mu_3^2}{72}(x^5-10x^3+15x)\phi_X(x) \mid \end{aligned} \tag{3}$$

In fact, Higher terms of the expansion are accompanied by large denominator which will allow us to neglect them without affecting the tool's accuracy. Notice, by using (3) we can measure the accuracy of our approximation, but the intervention of x variable will make the measurement hard. Therefore, this letter proposes to take the upper bound to this equation to have a simpler one to evaluate without affecting the accuracy of the tool.

In order to get the upper bound of (3) we need to get the extrema points for each term and plugin back the point that achieves the maximum value for each term. Table 1, will summarize the optimization process needed.

TABLE 1: Summarization of The Optimization Process

$f(x)$	$\mid \frac{\mu_3}{6}(x^2-1)\phi_X(x) \mid$	$\mid \frac{(\mu_4-3)}{24}(x^3-3x)\phi_X(x) \mid$	$\mid \frac{\mu_3^2}{72}(x^5-10x^3+15x)\phi_X(x) \mid$
Extrema points	0 $\sqrt{3}$ $-\sqrt{3}$	0.74196 $-$ 0.74196 2.33441 $-$ 2.33441	$-$0.61671 0.61671 $-$ 1.88918 1.88918 $-$ 3.32426 3.32426
Extrema substitution	0.067 0.03 0.03	0.023 0.023 0.006 0.006	0.032 0.032 0.014 0.014 0.002 0.002
Max value	0.067	0.023	0.032

Based on Table 1, (3) can be reformulated to

$$\Psi \leq 0.067 \mid \mu_3 \mid +0.023 \mid \mu_4 - 3 \mid +0.032 \mid \mu_3^2 \mid \quad (4)$$

Which is as mentioned above, a simpler inequality to measure the accuracy of an approximation by simply calculating the constants μ_3 and μ_4.

III. Accuracy of The Gamma Approximation

The Laguerre series expansion is an expression for the PDF of the sum of I.I.D random variables where it has the form of infinite series containing Laguerre polynomials.

Although the Central Limit theorem allows the PDF of the sum of I.I.D random variables to be approximated by a Gaussian distribution under the constrained of large number of them, but if this constraint does not satisfied the theorem fails to provide accurate results.

Laguerre series is an exact expansion for channels that suffer from Nakagami fading (e.g. studying of SIMO, MISO and MIMO systems) in compared to the Edgeworth series, which is exact for Gaussian channels.

Let V be the random variable we want to approximate and suppose after applying the moment matching process it suggests that a Gamma distribution is a good approximation for it. Now, based on the original random variable we define the standardized random variable X as

$$X = \beta V \quad (5)$$

where $\beta = \frac{v_1}{v_2}$ and v_1, v_2 are the mean and variance for V, respectively.

Laguerre expansion for the PDF of the normalized Gamma random variable X can be written as

$$f_X(x) = \frac{x^{m-1}e^{-x}}{\Gamma(m)}[1 + \zeta_3 L_3^{(m)}(x) + \zeta_4 L_4^{(m)}(x) + ...] \quad (6)$$

Where $m = \frac{v_1^2}{v_2}$, $\zeta_3 = \frac{\Gamma(m)}{3!\Gamma(m+3)}(\mu_3 - 2m)$, $\zeta_4 = \frac{\Gamma(m)}{4!\Gamma(m+4)}(\mu_4 - 12\mu_3 - 3m^2 + 18m)$, $L_3^{(m)}(x) = x^3 - 3(m+2)x^2 + 3(m+2)(m+1)x - (m+2)(m+1)m$ and $L_4^{(m)}(x) = x^4 - 4(m+3)x^3 + 6(m+3)(m+2)x^2 - 4(m+3)(m+2)(m+1)x + (m+3)(m+2)(m+1)m$. After integrating (6) with respect to x, the CDF of the expansion is

$$\begin{aligned} F_X(x) = F_m(x) &+ m(m+1)(m+2)[(m+3)\zeta_4 - \zeta_3]F_m(x) \\ &+ m(m+1)(m+2)[3\zeta_3 - 4(m+3)\zeta_4]F_{m+1}(x) \\ &+ m(m+1)(m+2)[-3\zeta_3 + 6(m+3)\zeta_4]F_{m+2}(x) \\ &+ m(m+1)(m+2)[\zeta_3 - 4(m+3)\zeta_4]F_{m+3}(x) \\ &+ m(m+1)(m+2)[(m+3)\zeta_4]F_{m+4}(x) + ... \end{aligned} \quad (7)$$

Where $F_m(x)$ is the CDF of the Gamma random variable.

From (7), the error which can be considered to measure the degree of accuracy needed is the first six terms after the Gamma CDF term, where with some mathematical manipulations, the error formula can be interpreted to

$$\Psi = \mid \alpha\Theta_1 + \gamma\Theta_2 \mid \quad (8)$$

where $\alpha = \frac{\mu_3 - 2m}{3!}$, $\gamma = \frac{\mu_4 - 12\mu_3 - 3m^2 + 18m}{4!}$, $\Theta_1 = -F_m(x) + 3F_{m+1}(x) - 3F_{m+2}(x) + F_{m+3}(x)$ and $\Theta_2 = F_m(x) - 4F_{m+1}(x) + 6F_{m+2}(x) - 4F_{m+3}(x) + F_{m+4}(x)$. Notice that the error formula contains only the third and fourth central moments which is as mentioned above, also notice that the error for the Gamma distribution depends on two parameters which are m and x, so for a given m we can upper bound the error formula for the adopted system. Table 2 summarizes some upper bounds for known values of m.

TABLE 2: Upper Bound of $\mid \Theta_1 \mid$ & $\mid \Theta_2 \mid$ For Different m

m	1	2	4	6	8	≥ 10
$\mid \Theta_1 \mid$	0.168	0.077	0.031	0.019	0.013	≤ 0.01
$\mid \Theta_2 \mid$	0.132	0.051	0.021	0.011	0.007	≤ 0.005

IV. Distributed Massive MIMO Application

One of the interested applications where Gamma distribution appears on it, is distributed massive MIMO system. This letter will apply the results gained on such a system in order to measure the accuracy of the Gamma approximation.

Let us have a massive MIMO system with the following criteria

- n_{tx} & n_{rx} represents the number of transmit antennas and the number of receive antennas, respectively.
- $n_{rx} >> n_{tx}$
- Transmitters & Receivers have single antennas and can be located anywhere.
- All receivers are connected to a single base station, or in other words, all transmitted bits are jointly detected.

At single time instance the received signal can be formulated as

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w} \tag{9}$$

Where each of which is

- y = $[y_1 ... y_{n_{rx}}]^T$ the received signal vector.
- H = $\begin{pmatrix} h_{11} & .. & h_{1*n_{tx}} \\ . & .. & . \\ . & .. & . \\ h_{n_{rx}*1} & .. & h_{n_{rx}*n_{tx}} \end{pmatrix}$ the channel matrix, where the channel coefficients within it are independent complex normal (complex Gaussian) random variables with zero mean and P_{nk} variance ($E[\mid h_{nk} \mid^2] = P_{nk}$), where P_{nk} is the received power at antenna n from transmitter k which known as the power profile.
- s = $[s_1 ... s_{n_{tx}}]^T$ the normalized transmitted symbols. In other words, $E[\mid s_k \mid^2] = 1$, for all symbols.
- w = $[w_1 ... w_{n_{rx}}]^T$ the AWGN vector with $E[w_k] = 0$ and $E[\mid w_k \mid^2] = \sigma^2$, for all received noises.

Under the following conditions for a MIMO system, the SNR for it can be approximated as

$$\bar{\gamma_k} = \rho \sum_{n=1}^{n_{rx}} \mid h_{nk} \mid^2 \tag{10}$$

Where $\rho = \frac{n_{rx} - n_{tx} + 1}{n_{rx}\sigma^2}$. (10) accuracy will be checked via the tool developed above for the Gamma distribution, but in order to proceed with the analysis we must calculate the first four moments for (10) which are

- $E[\bar{\gamma_k}] = \rho \sum_{n=1}^{n_{rx}} P_{nk}$
- $E[(\bar{\gamma_k} - E[\bar{\gamma_k}])^2] = \rho^2 \sum_{n=1}^{n_{rx}} {P_{nk}}^2$
- $E[(\bar{\gamma_k} - E[\bar{\gamma_k}])^3] = 2\rho^3 \sum_{n=1}^{n_{rx}} {P_{nk}}^3$
- $E[(\bar{\gamma_k} - E[\bar{\gamma_k}])^4] = \rho^4 [6 \sum_{n=1}^{n_{rx}} {P_{nk}}^4 + 3(\sum_{n=1}^{n_{rx}} {P_{nk}}^2)^2]$

Secondly, we will normalize the Gamma random variable to apply the developed tool on it, so we use (5), where $\beta = \frac{v_1}{v_2} = \frac{\sum_{n=1}^{n_{rx}} P_{nk}}{\rho \sum_{n=1}^{n_{rx}} {P_{nk}}^2}$. Finally, subtitute ρ within the first four moments with $\frac{\sum_{n=1}^{n_{rx}} P_{nk}}{\sum_{n=1}^{n_{rx}} {P_{nk}}^2}$ to have the moments for X.

Now, let us consider three scenarios applied for massive MIMO system with 100 receive antennas. The first one is that all antennas are co-located such that $P_{nk} = 10dB$, where in the second one antennas are located in two clusters such that $P_{nk} = 10dB$ for the first 50 receive antennas and $P_{nk} = 12dB$ for the remaining ones. The final scenario is that a single user is located close to the first antenna with $P_{nk} = 10dB$, where the others are too far with $P_{nk} = 0dB$.

The variances for each scenario can be evaluated as follows

- $E[P_{nk}] = \frac{\sum_{n=1}^{100} 10}{100} = 10dB$ and $E[(P_{nk} - E[P_{nk}])^2] = \frac{\sum_{n=1}^{100}(10-10)^2}{100} = 0dB$
- $E[P_{nk}] = \frac{\sum_{n=1}^{50} 10 + \sum_{n=51}^{100} 12}{100} = 11dB$ and $E[(P_{nk} - E[P_{nk}])^2] = \frac{\sum_{n=1}^{50}(10-11)^2 + \sum_{n=51}^{100}(12-11)^2}{100} = 1dB$
- $E[P_{nk}] = \frac{\sum_{n=1}^{1} 10 + \sum_{n=2}^{100} 0}{100} = 0.1dB$ and $E[(P_{nk} - E[P_{nk}])^2] = \frac{\sum_{n=1}^{1}(10-0.1)^2 + \sum_{n=2}^{100}(0-0.1)^2}{100} = 0.99dB$

Applying the derived formula for such scenarios will get the following results

- $\alpha = 0, \gamma = 0 and \Psi = 0$ for the first scenario.
- $\alpha = 1.103, \gamma = -21 and \Psi \leq 8.84*10^{-4}$ where m taken as 96 for the second scenario.
- $\alpha = 40, \gamma = 87 and \Psi \leq 4.7 * 10^{-2}$ where $m = 60$ for the last scenario.

After observing these scenarios, we can say that the Gamma approximation is more accurate when the power profile has a small variance compared to it's mean value.

V. Conclusion

A new simple and accurate tool was developed that can measure the accuracy of the Gaussian and Gamma approximations where it depends only on the first four moments. Specifically, the Edgeworth series was used to develop the formula that measures the accuracy of the Gaussian approximation, where we use the Laguerre series to develop the formula that measures the Gamma approximation accuracy. After that, the developed tool was applied on a massive MIMO system to insure its validity. Finally, it good to be mentioned that the steps adopted here can be extended to develop a new tool that encompasses other distributions.

References

[1] Peter J. Smith, Rajitha Senanayake, and Pawel A. Dmochowski, "How Accurate is Your Gaussian/gamma Approximation?" in Wireless Communications, 2018.

[2] https://www.wikipedia.org/

[3] U. Amarakone and N. Ekanayake, "Laguerre Series Expansion for the Density Function of Sum of Random Variables and its Application to Analysis of EGC Diversity Systems", Faculty of Engineering, University of Peradeniya, Peradeniya, December 2009.