Dhaka Univ. J. Sci. 71(1): 76-81, 2023 (January)
ISSN 1022-2502, E-ISSN 2408-8528
DOI: https://doi.org/10.3329/dujs.v71i1.65276

Evaluating the Accuracy of Chebyshev's Inequality for Probability Calculation: A Simulation Study

Tasmiah Afrin Emma, Anamul Haque Sajib* and Sabina Sharmin

Department of Statistics, University of Dhaka, Dhaka-1000, Bangladesh

(Received: 15 November 2022; Accepted: 31 January 2023)

Abstract

This paper aims to evaluate the accuracy of probability calculation using Chebyshev's inequality based on simulation study. We consider symmetric $\left(\text{Normal}\left(3,1.5^2\right), \text{Laplace}\left(3,2\right), \text{Beta}\left(7,7\right), t_5\right)$, positively skewed, negatively skewed $\left(\chi^2_5, \text{Beta}\left(3,8\right), \text{Gamma}\left(5,1\right)\right)$ (Beta (7, 2)), Exponential (5) and Uniform $(0,1)$ distributions, $f_X(x)$ in our simulation study to measure the performance of Chebyshev's inequality. We then calculate $\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right)$ for $X \sim f_X\left(x\right)$, $\mu = E\left(X\right)$ and $\sigma^2 = Var\left(X\right)$, and compare this with the approximated probability obtained from Chebyshev's inequality to measure the accuracy of Chebyshev's inequality. From our simulation study, it is observed that loss due to using Chebyshev's inequality for probability calculation is the least and the maximum when $f_X(x)$ is the Exponential and the Beta (symmetric) distributions, respectively for $k \ge 2.5$. Moreover, the value of $\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right)$ calculated using Chebyshev's inequality is underapproximated for all the probability distributions considered in the study.

Keywords: Chebyshev's inequality; probability distribution, symmetric, positively and negatively skewed distributions.

I. Introduction

The Chebyshev's inequality is one of the most widely used inequalities in Statistics which can be used to calculate the probability of a random variable X when its distribution function is unknown. More specifically, it can be used to calculate $\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right)$ provided that $\mu = E\left(X\right)$ and $\sigma^2 = Var\left(X\right) < \infty$ of a random variable X exist. In other words, calculating $\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right)$ does not require knowing the shape of the distribution as long as $E\left(X\right)$ and $Var\left(X\right)$ exist. This non-parametric assumption made this inequality so popular and useful in different disciplines.

Bienayme[1] (1853) first introduced Chebyshev's inequality but he did not provide its mathematical proof during that time. Later in 1867, a Russian mathematician Pafnuty Chebyshev's provided its mathematical proof[2] and therefore, this inequality is named after Pafnuty Chebyshev's.

Although Chebyshev's inequality is widely applicable, it has two main issues as far as probability calculation is concerned. Firstly, when the lower bound on the probability of $\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right)$ for a known $X \sim f_X\left(x\right)$ is considered to be p, then to achieve the same probability Chebyshev's inequality requires a wider interval compared to the interval of $\left(\mu - k\sigma,\ \mu + k\sigma\right)$. For example, 95% observations lie within 1.96σ from the mean of X where $X \sim N\left(\mu, \sigma^2\right)$ while Chebyshev's inequality requires 4.47σ distance from the mean of X to produce the same probability, which is about 2.29 times higher. This interval which is 4.47σ distance from the mean of X is termed as Chebyshev's greater-than-95% interval[3]. In other words, for a known $X \sim f_X\left(x\right)$ the value of $\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right)$ is much higher than the lower bound on probability calculated using Chebyshev's inequality for the same interval. That is, calculating the value of $\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right)$ using Chebyshev's inequality is underapproximated. For example, when $X \sim N\left(\mu, \sigma^2\right)$ then $\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right) \approx 0.95$ while it is at least 0.75 under Chebyshev's inequality and the loss of information incurred due to using Chebyshev's inequality in this case is 21.05%. The amount of loss incurred due to using Chebyshev's inequality for probability calculation varies and it depends on the type of parent distribution of the random variable[4].

To the best of our knowledge, we have not found any study in the literature which covered this issue. In other words, there is no study in the literature which investigates the performance of Chebyshev's inequality for probability calculation. In this paper, we mainly focus on quantifying the amount of loss incurred because of using Chebyshev's inequality for probability calculation.

The rest of the paper is organized as follows. In section 2, the methodology of the paper is presented briefly. In section 3, a brief overview of the simulation settings is presented. In section 4, we discuss the results of our study and the paper concludes in section 5.

II. Methodology

In this section, we discuss the methodologies used in this paper briefly. More specifically, how to calculate $\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right)$ for both the known and unknown form of $f_X\left(x\right)$ is discussed here. When the form of $f_X\left(x\right)$ is unknown, Chebyshev's inequality is used to find $\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right)$ provided that $E\left(X\right)$ and finite variance $\sigma^2 = Var\left(X\right)$ exist. Let X be a random variable having finite mean $\mu = E\left(X\right)$ and finite variance $\sigma^2 = Var\left(X\right)$. Then for any real number $k > 1$, the Chebyshev's inequality takes the form

$$\Pr\left(\mu - k\sigma \le X \le \mu + k\sigma\right) \ge 1 - \frac{1}{k^2}.$$

This inequality proves to be impractical for $k \le 1$. The calculated probabilities of $\Pr(\mu - k\sigma \le X \le \mu + k\sigma\ )$

*Author of correspondence, e-mail: sajibstat@du.ac.bd

express the minimum proportion of data that fall within or beyond k standard deviations from the mean for the given data set. In other words, they characterize the dispersion of the data from its mean. In fact, when actual probability distributions are known they provide tighter bounds compared to the inequality.

If a random variable X follows a Uniform (0, 1), then $\mu = E(X) = 0.5$ and $\sigma^2 = Var(X) = 12^{-1}$. Mathematically, it is shown that

$$\Pr\left(|X-\mu| < k\sigma\right) = \Pr\left(\left|X-\frac{1}{2}\right| < \frac{k}{\sqrt{12}}\right)$$

$$= \int_{0.5-\frac{k}{\sqrt{12}}}^{0.5+\frac{k}{\sqrt{12}}} dx = \frac{k}{\sqrt{3}},$$

where $k \le \sqrt{3}$. This implies that, for different values of k, the exact probabilities of X falling within k standard deviations of the mean is $\frac{k}{\sqrt{3}}$ when the random variable originates from a Uniform(0, 1). When X follows an exponential $\left(\frac{1}{\lambda}\right)$ then $\mu = E(X) = \lambda$ and $\sigma^2 = Var(X) = \lambda^2$. Then, $\Pr\left(|X-\mu| < k\sigma\right)$ can be determined as

$$\Pr\left(|X-\mu| < k\sigma\right) = \Pr\left(|X-\lambda| < k\lambda\right)$$

$$= \int_{\lambda-k\lambda}^{\lambda+k\lambda} \frac{1}{\lambda} e^{-\frac{1}{\lambda}x} dx$$

$$= e^{k-1} - e^{-(k+1)},$$

where $(\lambda - k\lambda) \ge 0$. However, for some probability distributions considered in this paper deriving such mathematical forms for obtaining exact probabilities is not quite as straight-forward and often proven to be computationally tedious. In such situations, we use statistical package R to calculate the required probability.

III. Simulation Settings

We conduct a simulation study in this paper to compare the performance of Chebyshev's inequality for probability calculation. In our simulation study, different known symmetric, positively skewed and negatively skewed continuous probability distributions are considered to compare the performance of Chebyshev's inequality. More specifically, Normal(3, 1.5^2), Laplace(3, 2) (Double Exponential), Beta(7, 7), and t_5 distributions are considered to cover symmetric distributions while χ^2_5, Gamma (5, 1), and Beta(3, 8) distributions and Beta(7, 2) distribution are considered to cover positively skewed and negatively skewed distributions, respectively.

Table 1. Mean and Variance of $X \sim f_X(x)$.

Distribution	Parameter	Mean	Variance
Exponential	λ^{-1} (rate)	λ	λ^2
χ^2	ϑ	ϑ	2ϑ
Uniform	a, b	$0.5(a+b)$	$12^{-1}(b-a)^2$
Normal	μ, σ^2	μ	σ^2
Gamma	$\alpha(shape)$, $\beta(scale)$	$\alpha\beta$	$\alpha\beta^2$
Laplace	$\mu(location)$, $\beta(scale)$	μ	$2\beta^2$
Beta	$\alpha(shape)$, $\beta(scale)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
t	ϑ	0 ($\vartheta > 1$)	$\frac{\vartheta}{\vartheta-2}$ ($\vartheta > 2$)

Furthermore, we also considered Uniform(0, 1) and Exponential(5) distributions. Then for the known probability distributions the exact probability of $\Pr(\mu - k\sigma \le X \le \mu + k\sigma)$ is compared with the probability of that interval obtained using the Chebyshev's inequality pretending that the form of $f_X(x)$ is not known but mean and variance exist. The mean and variance of the probability distributions considered in the study are presented in Table 1.

IV. Results and Discussions

This section presents the results of our comparative study for calculating $\Pr(\mu - k\sigma \le X \le \mu + k\sigma)$ using the exact form of $f_X(x)$ and Chebyshev's inequality (pretending that we do not know the exact form of $f_X(x)$ but the mean and variance are known to us). The exact probabilities of $\Pr(\mu - k\sigma \le X \le \mu + k\sigma)$ obtained from the probability distributions are displayed in Table 2 along with the corresponding probabilities obtained using Chebyshev's inequality for different values of k.

Table 2 illustrates that when mean and finite variance exist, Chebyshev's inequality generally provides a poorer bound as compared to what may be obtained if the original distribution of the random variable X is known. From Table 2, it is evident that, for any particular k-value, $Pr(\mu - k\sigma \le X \le \mu + k\sigma)$ is higher when directly obtained from the probability distributions compared to the probability values obtained using Chebyshev's inequality. That is, Chebyshev's inequality underapproximates the values of $Pr(\mu - k\sigma \le X \le \mu + k\sigma)$ for the distributions taken under consideration.

Table 2. The values of $Pr(\mu - k\sigma \leq X \leq \mu + k\sigma)$ when $X \sim f_X(x)$ is Uniform(0, 1), Exponential(5), Normal(3, 1.5^2), Laplace(3, 2), Beta (7, 7), t_5, χ^2_5, Gamma(5, 1), Beta(3, 8), and Beta(7, 2) distributions and using Chebyshev's inequality for different values of .

Distribution	k										
	1	*1.1*	*1.5*	$\sqrt{3}$	*2*	*2.5*	*3*	*3.5*	*4*	*5*	*10*
Uniform(0,1)	0.5774	0.6351	0.8660	1.0000	-	-	-	-	-	-	-
Exponential (5)	0.8647	0.8775	0.9179	0.9349	0.9502	0.9698	0.9817	0.9889	0.9933	0.9975	0.9999
Symmetric											
Normal (3, 1.5^2)	0.6827	0.7287	0.8664	0.9167	0.9545	0.9876	0.9973	0.9995	0.9999	0.9999	1.0000
Laplace(3, 2)	0.7569	0.7889	0.8801	0.9137	0.9409	0.9708	0.9856	0.9929	0.9965	0.9992	0.9999
Beta(7, 7)	0.6657	0.7136	0.8617	0.9176	0.9594	0.9931	0.9996	0.9999	1.0000	1.0000	1.0000
t_5	0.7468	0.7852	0.8894	0.9244	0.9507	0.9767	0.9883	0.9937	0.9964	0.9987	0.9999
Positively Skewed											
Gamma(5, 1)	0.7007	0.7507	0.8926	0.9346	0.9588	0.9801	0.9907	0.9958	0.9981	0.9997	1.0000
Beta(3, 8)	0.6699	0.7207	0.8767	0.9305	0.9630	0.9864	0.9959	0.9991	0.9999	0.9999	1.0000
χ^2_5	0.7236	0.7788	0.9155	0.9372	0.9547	0.9757	0.9872	0.9933	0.9966	0.9991	0.9999
Negatively Skewed											
Beta(7, 2)	0.6764	0.7311	0.8966	0.9368	0.9579	0.9821	0.9935	0.9981	0.9996	0.9999	1.0000
Chebyshev's Inequality	0.0000	0.1736	0.5556	0.6667	0.7500	0.8400	0.8889	0.9184	0.9375	0.9600	0.9900

Moreover, it is observed that for $k < \sqrt{3}$ Exponential(5) distribution and for $k \geq 2.5$ Beta (symmetric) distribution provide bounds of $Pr(\mu - k\sigma \leq X \leq \mu + k\sigma)$ that are higher than those obtained from the rest of the distributions. For $k < 1.5$ Uniform$(0,1)$ distribution, for $1.5 < k < 2.5$ Laplace(3,2), and for $k \geq 2.5$ Exponential(5) distribution give probabilities that are comparatively nearer to those obtained from Chebyshev's Inequality. In addition, it is observed that in the case of Uniform$(0,1)$ distribution, for $k > \sqrt{3}$ and $X \geq 0$, the values of $Pr(\mu - k\sigma \leq X \leq \mu + k\sigma)$ are greater than 1 which contradicts an important axiom of probability (i.e., probability of an event is always between 0 and 1). Thus, Chebyshev's Inequality does not demonstrate much use in case of a Uniform distribution. For $k > 4$ almost all of the observations are encompassed within the range $\mu \pm k\sigma$.

To put the idea of tighter bounds into perspective, the bounds of $Pr(\mu - k\sigma \leq X \leq \mu + k\sigma)$ are plotted in Figure 1. Upon closer look at Figure 1, it is revealed that the probabilities

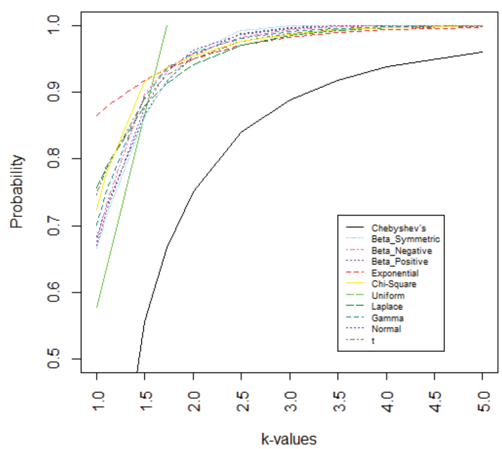


Fig.1. Values of $Pr(\mu - k\sigma \leq X \leq \mu + k\sigma)$ when $X \sim f_X(x)$ is Uniform(0, 1), Exponential(5), Normal(3, 1.5^2), Laplace(3, 2), Beta(7, 7), t_5, χ^2_5, Gamma(5, 1), Beta(3, 8), and Beta(7, 2) distributions and using Chebyshev's inequality.

obtained when $X \sim f_X(x)$ is an Exponential(5) distribution are farthest from those obtained from Chebyshev's inequality for $k < \sqrt{3}$ but nearest for $k \ge 2.5$. Again when $X \sim f_X(x)$ is a symmetric Beta distribution (i.e., Beta(7, 7)) for $k \ge 2.5$ the probabilities are farthest from those obtained using Chebyshev's Inequality. Furthermore, Figure 1 shows that the probabilities obtained using the exact knowledge of probability distributions remain way above those obtained from Chebyshev's inequality, indicating the value of $\Pr(\mu - k\sigma \le X \le \mu + k\sigma)$ using Chebyshev's inequality is generally under approximated.

The greater distance of the probabilities obtained using the exact knowledge of probability distributions compared to those obtained using Chebyshev's inequality demonstrate a greater loss of accuracy of Chebyshev's inequality when it is used (distribution of the random variable X is not known). For better visualization the probabilities are plotted for different scenarios in Figure 2. More specifically, we have considered four different scenarios in Figure 2: (a) Exponential and Uniform, (b) symmetric, (c) positively skewed and (d) negatively skewed distributions. Figure 2(b) displays that in the case of symmetric distributions, the probabilities of Chebyshev's inequality are comparatively near to those obtained from a symmetric Beta distribution for $k \le 1.5$, from a Laplace(3, 2) distribution for $1.5 < k < 4$ and from a t distribution for $k \ge 4$.

Again, the Chebyshev's inequality probabilities are at a greater distance from those obtained from a Laplace distribution for $k \le 1.1$, from a t distribution for $1.1 < k < 2$, and from a symmetric Beta distribution for $k \ge 2$. Similarly, Figure 2(c) shows that for positively skewed distributions, the probabilities of Chebyshev's inequality are comparatively near to those obtained from a positively skewed Beta distribution for $k < 2$ and from a Chi-square distribution for $k \ge 2$. Again they farthest from those obtained from a Chi-square distribution for $k < 2$ and a positively skewed Beta distribution for $k \ge 2$. Table 3 provides clear insight regarding the loss of accuracy that ensues when Chebyshev's inequality is used which is the graphically shown in Figure 3. At $k = 1$, there is 100% loss of accuracy when Chebyshev's inequality is used -

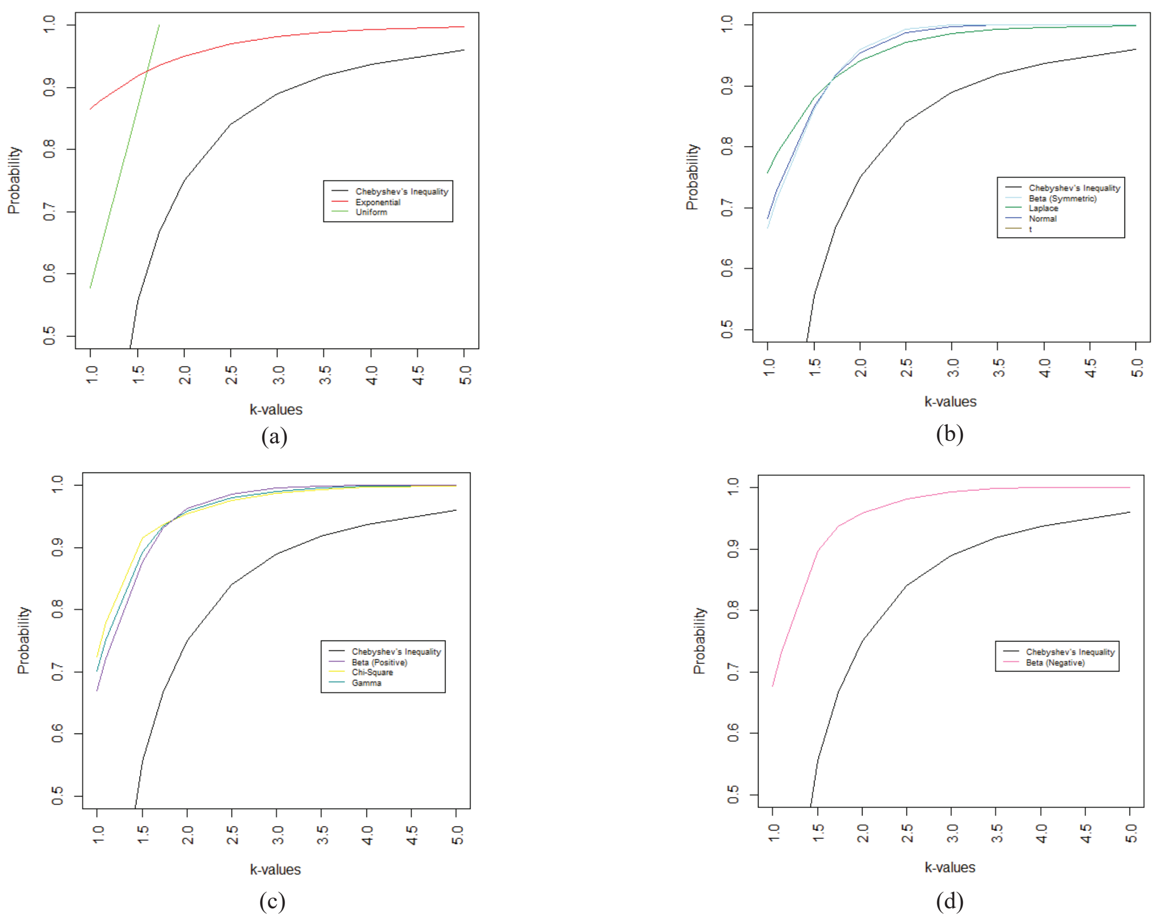


Fig. 2. Values of $Pr(\mu - k\sigma \le X \le \mu + k\sigma)$ when $X \sim f_X(x)$ is (a) Uniform(0, 1) and Exponential(5) distributions, (b) Symmetric distributions (Normal(3, 1.5^2), Laplace(3, 2), Beta(7, 7), t_5), (c) Positively skewed distributions (χ^2_5, Gamma(5, 1), Beta(3, 8), and (d) Negatively skewed distribution (Beta(7, 2)) and using Chebyshev's inequality.

Table 3. Loss of accuracy in absolute difference (and Loss of accuracy in %) due to Chebyshev's inequality when $X \sim f_X(x)$ is Uniform(0, 1), Exponential(5), Normal($3, 1.5^2$), Beta (7, 7), Laplace(3, 2), t_5, χ^2_5, Gamma(5, 1), Beta (3, 8), and Beta (7, 2) distributions.

Distribution	*k*									
	1	*1.1*	*1.5*	$\sqrt{3}$	*2*	*2.5*	*3*	*3.5*	*4*	*5*
Uniform (0, 1)	0.5774 *(100)*	0.4615 *(72.67)*	0.3105 *(35.85)*	0.3333 *(33.33)*	-	-	-	-	-	-
Exponential (5)	0.8647 *(100)*	0.7040 *(80.22)*	0.3624 *(39.48)*	0.2683 *(28.69)*	0.2002 *(21.07)*	0.1298 *(13.38)*	0.0928 *(9.45)*	0.0705 *(7.13)*	0.0558 *(5.61)*	0.0375 *(3.76)*
Symmetric										
Normal ($3, 1.5^2$)	0.6827 *(100)*	0.5551 *(76.18)*	0.3108 *(35.88)*	0.2501 *(27.28)*	0.2045 *(21.42)*	0.1476 *(14.94)*	0.1084 *(10.87)*	0.0812 *(8.12)*	0.0624 *(6.24)*	0.0399 *(3.99)*
Laplace (3, 2)	0.7569 *(100)*	0.6154 *(78.00)*	0.3246 *(36.88)*	0.2470 *(27.03)*	0.1909 *(20.29)*	0.1309 *(13.48)*	0.0967 *(9.82)*	0.0745 *(7.51)*	0.0590 *(5.92)*	0.0392 *(3.92)*
Beta (7, 7)	0.6657 *(100)*	0.5401 *(75.70)*	0.3061 *(35.53)*	0.2509 *(27.35)*	0.2094 *(21.83)*	0.1531 *(15.41)*	0.1107 *(11.07)*	0.0816 *(8.16)*	0.0625 *(6.25)*	0.0400 *(4.00)*
t_5	0.7468 *(100)*	0.6116 *(77.90)*	0.3339 *(37.54)*	0.2577 *(27.88)*	0.2007 *(21.11)*	0.1367 *(13.99)*	0.0994 *(10.06)*	0.0753 *(7.58)*	0.0589 *(5.91)*	0.0387 *(3.87)*
Positively Skewed										
Gamma (5, 1)	0.7007 *(100)*	0.5771 *(76.88)*	0.3371 *(37.76)*	0.2679 *(28.67)*	0.2088 *(21.78)*	0.1401 *(14.30)*	0.1018 *(10.28)*	0.0774 *(7.77)*	0.0606 *(6.07)*	0.0397 *(3.97)*
Beta (3, 8)	0.6699 *(100)*	0.5471 *(75.92)*	0.3212 *(36.63)*	0.2638 *(28.35)*	0.2130 *(22.11)*	0.1464 *(14.84)*	0.1071 *(10.75)*	0.0807 *(8.08)*	0.0624 *(6.24)*	0.0399 *(3.99)*
χ^2_5	0.7236 *(100)*	0.6053 *(77.72)*	0.3600 *(39.32)*	0.2705 *(28.87)*	0.2047 *(21.44)*	0.1357 *(13.91)*	0.0983 *(9.96)*	0.0750 *(7.55)*	0.0591 *(5.93)*	0.0391 *(3.92)*
Negatively Skewed										
Beta (7, 2)	0.6764 *(100)*	0.5576 *(76.26)*	0.3411 *(38.04)*	0.2701 *(28.83)*	0.2079 *(21.70)*	0.1421 *(14.47)*	0.1046 *(10.53)*	0.0797 *(7.99)*	0.0621 *(6.21)*	0.0399 *(3.99)*

which shows the inequality is impractical for $k \le 1$. There is a gradual decrease in the loss of accuracy with an increase in the values of k. Table 3 shows that loss of accuracy when $X \sim f_X(x)$ is Exponential distribution is the least (as observed from the small absolute difference of the probabilities as well as the low percentage of loss of accuracy) and highest when $X \sim f_X(x)$ is a symmetric Beta distribution (as observed from the greater percentage of loss of accuracy and absolute difference of the probabilities) for $k \ge 2.5$. Normal distribution shows the second highest loss of accuracy after symmetric Beta distribution for $k \ge 2.5$. While for $k \ge 2.5$ we see a general pattern of loss of accuracy, for $k < 2.5$ no such general pattern is easily discernible.

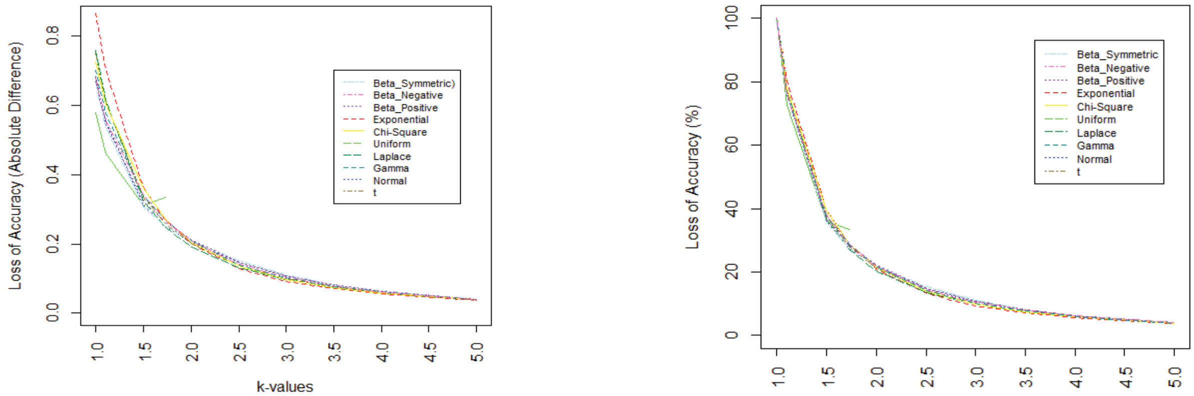
In case of symmetric distributions, for $k \le 1.5$ symmetric Beta distribution, for $1.5 < k < 4$ Laplace distribution and for $k \ge 4$ t-distribution show the least amount of loss while for $k \ge 2$ symmetric Beta distribution show the greatest amount of loss. Again we may further classify the symmetric distributions with respect to their nature of symmetry. Normal, t, and Beta distributions have more of a bell-shaped symmetry while Laplace distribution has a sharper peak. For bell-shaped symmetric distributions, t distribution shows lower loss of accuracy when $k > 2$.

In case of positively skewed distributions, Beta distribution shows lower amount of loss for $k < 2$ and Chi-square distribution shows lower amount of loss for $k \ge 2$. On the contrary, Beta distribution shows a greater amount of loss for $k \ge 2$ while Chi-square distribution shows a greater amount of loss for $k < 2$.

Figure 4 is constructed for 2.5≤ k ≤5 to get a clearer view of the general pattern of loss of accuracy. Beyond 2.5 Exponential distribution generates the least percentage of loss while symmetric Beta distribution generates the greatest percentage of loss.

V. Conclusion

Taking into account all the findings, the study concludes that generally loss of accuracy incurred due to Chebyshev's inequality is the least when data came from Exponential distribution and highest when data came from Beta (symmetric) distribution for $k \geq 2.5$. However, no general pattern was apparent for $k < 2.5$. While this paper showed the results obtained from continuous probability distributions, similar type of results can be showed in case of discrete distributions by considering the cases of symmetric, positively skewed, and negatively skewed distributions.



(a)

(b)

Fig. 3. Loss of Accuracy (absolute difference) and Loss of Accuracy (%) due to Chebyshev's inequality

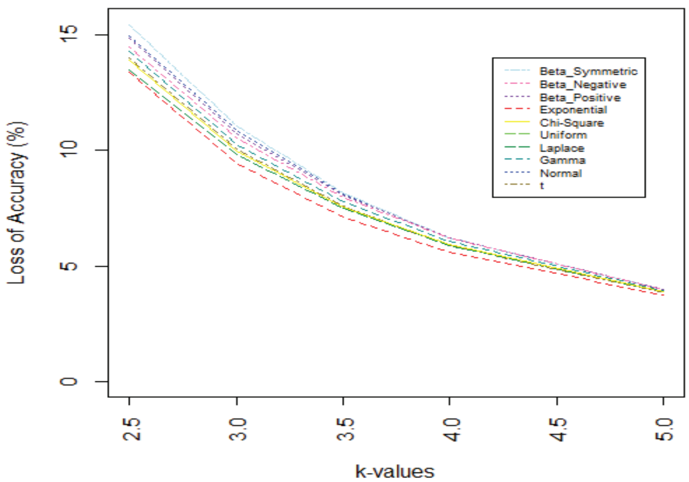


Fig. 4. Loss of Accuracy (%) due to Chebyshev's inequality for $2.5 \leq k \leq 5$

For all the probability distributions considered in the study, the Chebyshev's inequality underapproximates the values of $\Pr(\mu - k\sigma \leq X \leq \mu + k\sigma)$. These findings enable us to understand that Chebyshev's inequality provides us with approximations as the results obtained are rather conservative as well as crude. But while it does not provide as sharp a bound as obtained when the actual distribution of the random variable is known, the inequality offers at least some information regarding how spread the dataset is when the actual distribution is unknown. Consequently, Chebyshev's inequality renders such universal utility since it can be used for any probability distribution as long as we can define the mean and the variance.

References

1. Bienaymé, I. J. 1853. Considérations àl›appui de la découverte de Laplace. *Comptes Rendus de l'Académie des Sciences*. **37**, 309–324.
2. Tchebichef, P. (1867). "Des valeurs moyennes". *Journal de Mathématiques Pures et Appliquées*. **12(2)**, 177–184.
3. Wu, J. C., C. L. Wilson, and J. C. Wu, 2005. *Using Chebyshev's Inequality to Determine Sample Size in Biometric Evaluation of Fingerprint Data*. US Department of Commerce, National Institute of Standards and Technology.
4. Casella, G. and R. L. Berger, 2021. *Statistical Inference*. Cengage Learning, Second edition, 135.