Tail Probabilities: Cumulative Distribution Function: Asymptotic Lower Bounds

Rama Murthy Garimella

January 6, 2025

Tail Probabilities: Cumulative Distribution Function: Asymptotic Lower Bounds

Garimella Rama Murthy,

Professor, Computer Science Department,

Mahindra University, Hyderabad,

Telangana, India

ABSTRACT

In this research paper, motivated by the Chebyshev inequality and its generalizations, we derive interesting "asymptotic lower bounds" (as the "sample" size tends to infinity) on the commom cumulative distribution function evaluated at a positive real number. We expect more results of such nature to be discovered and proved. Further, a practical application of these asymptotic bounds is briefly discussed.

1. Introduction:

Probability theory enables capturing certain types of uncertainities arising in static as well as dynamic phenomena. In modern probability theory, the concept of random variable and associated cumulative distribution function played a crucial role in capturing the probability of critical events. Markov focused attention on the tail probability of a non-negative random variable. He proved the so called "Markov inequality". Motivated by the essential argument of Markov, mathematicians such as Chebyshev, Bienayme and others generalized the essential idea and arrived at interesting inequalities [1].

The concept of Markov inquality, Chernoff bound led to the research area of large deviations theory. Many important results were derived and applied by probabilitsts. These theoretical results found many practical applications in engineering and other areas of human endeavor. In recent times, the area of "data science" received considerable attention. There are efforts to place the conceptual foundations of data sciene on a strong logical basis. In this direction, the authors of [3] made an important contribution. In his attempt to begin teaching the foundations of data science, the author took a closer look at the bounds on tail probability. He was specifically interested in a sequence of independent, identically distributed random variables and the associated bounds. One of the contributions of such efforts is the current research paper. The author is specifically interested in asymptotic bound on the cumulative distribution function of a “sample” (independent, identically distributed random variables) of random variables evaluated at an interesting point on the real line [2].

This research paper is organized as follows. In Section 2, several bounds motivated by Markov inequality, Chebyshev inequality are summarized. The asymptotic bounsd discovered and proved by the author are presented in Section 3. The research paper concludes in Section 4.

2. **Bounds Motivated by Markov Inequality, Chebyshev Inequality**:

Russian mathematician Markov proved the following inequality. He was a student of Russian mathematician Chebyshev. We now state the Markov and Chebyshev inequalities [1]:

Markov Inequality:

For any real valued random variable Y and positive constant ‘a’, we have that

$$Prob\{\, |Y| \geq a \,\} \leq \frac{E(|Y|)}{a}.$$

This inequality can be used to prov the following Chebyshev inequality

Chebyshev Inequality:

Let X be an integrable random variable with finite non-zero variance σ^2 (and thus finite expected value μ). Then, for any real number $k > 0$, we have that

$$Prob\{\, |X - \mu| \geq k\,\sigma \,\} \leq \frac{1}{k^2}.$$

Note: In view of the above inequality and Euler’s result, we have that

$$\sum_{k=1}^{\infty} Prob\{\, |X - \mu| \geq k\,\sigma \,\} \leq \frac{\pi^2}{6}.$$

Infact the equality is attaned for the example random variable provided in [1].

Various mathematicians such as Selberg, Birnbaum et. al proved interesting inequalities associated with tail probabilities motivated by the Chebyshev inequality. Infact, most of the related bounds such as Selbrg's inequality, Samuelson's inequality are motivated by the themes related to Chebyshev inequality.Detailed results are documented in [1]. They are not repeated here for brevity.

The author was motivated to derive inequalities related to the probabilities associated with a "sample" of size N (i.e. N independent, identically distributed random variables). The culmination of such efforts are documented in the following Section [2]. It is hoped that the results in the next section will enable future generations of reseachers to derive novel bounds related to the asymptotics of a "sample" of random variables.

3. Independent, Identically Distributed Random Variables: Asymptotic Lower Bounds:

(I) Consider a finite collection of independent, identically distributed random variables: $\{ X_1, X_2, \ldots, X_N \}$ (i.e. a "sample" of N random variables) with mean μ and variance σ^2.

Each random variable satisfies the following Chebyshev inequality i.e

$$Prob\{ |X_i - \mu| \geq k\,\sigma \} \leq \frac{1}{k^2} \quad for \;\; 1 \leq i \leq N.$$

Since, they are independent and identically distributed, the following equation follows

$$\prod_{i=1}^{N} Prob\{ |X_i - \mu| \geq k\,\sigma \} \leq \frac{1}{k^{2N}} \ .$$

Let $F(.)$ be the common cumulative distribution function of the independent identically distributed random variables. Hence, we have that

$$(F(k\sigma))^N = \tilde{F}(k\,\sigma, k\,\sigma, \ldots, k\,\sigma) \geq \left(1 - \frac{1}{k^{2N}} \right).$$

Thus, as N tends to infinity, we have

$$\lim_{n\to\infty} (\tilde{F}(k\,\sigma, k\,\sigma, \ldots, k\,\sigma))^N \geq \quad e^{-K^2}.$$

By independence of the random variables, we have that

$$\lim_{n\to\infty} (F(k\,\sigma))^{N^2} \geq \quad e^{-K^2}.$$

We expect the above result to be useful in studying the properties of a large sample of random variables. Since the Chebyshev inequality is SHARP (It is an equality for certain distribution [1]), the asymptotic lower bound for "sample" may not be improved upon for arbitrary probability distributions.

(II) We now consider, a family of tail bounds derived by Mitzenmacher and Upfal (by application of Markov inequality to the random variable $|X - E(x)|^n$). We now invoke the inequality satisfied by each random variable in the collection (sample size 'n') of independent, identically distributed random variables (considered above)

$$Prob\left\{ |X_i - \mu| \geq k\, E(|X_i - \mu|^n)^{\frac{1}{n}} \right\} \leq \frac{1}{k^n} \; for \;\; k > 0, n \geq 2 \; and \; 1 \leq i \leq n.$$

\ $Equivalently, we \;\; have \;\; that$

$$1 - \; Prob\left\{ |X_i - \mu| \geq k\, E(|X_i - \mu|^n)^{\frac{1}{n}} \right\} \geq 1 - \frac{1}{k^n} \; for \;\; k > 0, n \geq 2 \; and \; 1 \leq i \leq n.$$

Thus (denoting the common CDF of the independent random variables by F(.)) by raising both the sides of the inequality to the power n, we have that

$$F(\, k\, E(|X_i - \mu|^n)^{\frac{1}{n}} \;)^n \geq (\, 1 - \tfrac{1}{k^n})^n$$

Letting 'n' tend to infinity (on both the sides), we have that

$$\lim_{n \to \infty} (F\left(k\, E(|X_i - \mu|^n)^{\frac{1}{n}} \right))^n \geq \;\; e^{-k}.$$

Thus, we have the above asymptotic lower bound on the common CDF of I.I.D random variables.

Note: The above asymptotic bounds (derived based on Chebyshev inequality) cannot in general (remaining true for arbitrary distributions) be improved upon. In [1], an example is provided for which the bound provided by Chebyshev inequality is SHARP i.e. Chebyshev inequality is an equality. The example can easily be generalized to provide an example distribution for which the asymptotic bound provided in (II) is attained. Detailed generalized example is avoided for brevity.

Note: From Mitzenmacher and Upfal inequality, we have that

$$\sum_{k=1}^{\infty} Prob\left\{|X_i-\mu| \geq k\, E(|X_i-\mu|^n)^{\frac{1}{n}}\right\} \leq \sum_{k=1}^{\infty}\frac{1}{k^n} = Zeta(n),$$

$$where\ Zeta(.)\ is\ the\ Riemann\ (\,Euler\,)\ Zeta\ \ function.$$

(III) We now apply the Lyapunov inequality associated with a non-negative random variable Z to the Mitzenmacher-Upfal family of tail bounds. Given, non-negative random variable Z, Lyapunov inequality states that

$$0 < E(Z) \leq (E(Z^2)^{\frac{1}{2}} \leq \cdots . \leq (E(Z^m)^{\frac{1}{m}}\ \ for\ m \geq 2.$$

$$Hence, the\ \ mitzenmacher-$$

$$\geq upfal\ \ inequality\ \ leads\ \ to\ \ the\ \ following$$

$$\theta(n) = Prob\left\{|X-\mu| \geq k\, E(|X-\mu|^n)^{\frac{1}{n}}\right\} \leq \frac{1}{k^n}\ \ for\ \ k > 0, n \geq 2\,.$$

But, by Lyapunov inequality applied to the random variable $|X-\mu|$, we have that

$$Prob\left\{|X-\mu| \geq k\, E(|X-\mu|^n)^{\frac{1}{n}}\right\} \leq Prob\left\{|X-\mu| \geq k\, E(|X-\mu|^l)^{\frac{1}{l}}\right\} \leq \frac{1}{k^l}$$

$$for\ \ k > 0, n \geq 2\ \ and\ \ 1 \leq l < n\,.$$

Thus, it readily follows that

$$\theta(n) \leq \theta(n-1) \leq \theta(n-2) \ldots . \leq \theta(2) \leq \theta(1) \leq \frac{1}{k}\ for\ \ k > 0\ \ amd\ \ any\ \ 'n'.$$

- From the Chebyshev inequality, we have that

$$-\,Prob\{\,|X-\mu| \geq k\,\sigma\,\} \geq -\frac{1}{k^2}.$$

Let Z=|X-μ| be the non-zero random variable with zero mean. We have that its cumulative distribution function satisfies:

$$F_Z(\,k\,\sigma\,) \geq 1 - \frac{1}{k^2}\ \ for\ \ any\ \ integer\ \ k.$$

Hence, we readily have that

$$Prob\{\,k\,\sigma \leq Z \leq (k+1)\sigma\,\} = F_Z(\,(k+1)\sigma\,) - F_Z(\,k\,\sigma\,) \leq \frac{1}{k^2}$$

$$since\ \ F_Z(\,(k+1)\,\sigma\ \leq 1\,).$$

Summing for all the integer values of k, we have that

$$\sum_{k=1}^{\infty} Prob\{\,k\,\sigma \leq Z \leq (k+1)\sigma\,\} \leq \sum_{k=1}^{\infty}\frac{1}{k^2}.$$

Thus, using the Euler sum, we have that for any non-negative random variable Z, we have

$$Prob\{\ \sigma \leq Z < \infty\ \} \leq \frac{\pi^2}{6}.$$

In the spirit of the above result, using the Mitzenmacher-Upfal family of bounds, it readily follows that (for integer values of k also and integer n>1)

$$\sum_{k=1}^{\infty} Prob\{\ k\ m(n) \leq Z \leq (k+1)\ m(n)\ \} \leq \sum_{k=1}^{\infty} \frac{1}{k^n} = Zeta(n),\ \ where$$

$$m(n) = \ \ E(|Z-\mu|^n)^{\frac{1}{n}}$$

- We now provide a practical application, where the above lower bounds in (I), (II) find utility

  Consider an interconnection of finitely many systems with the times to failure being independent and distributed non-negative random variables $\{X_1,$ $X_2, \ldots, X_N\}$ with the common CDF being F(.). Let these systems be connected in PARALLEL. Thus, the time to failure of the TOTAL system happens when all the sub-systems (connected in parallel) fail. Let Z denote the random variable denoting the time to failure of TOTAL SYSTEM: It readily follows that

  Z = Max $\{ X_1, X_2, \ldots, X_N \}$. Hence we readily have that

$$F_Z(z) = \ \ (F(z))^N.$$

Thus, the asymptotic lower bound in (II) readily applies in this application. Also, the asymptotic lower bound proved in (I) above applies.

Note: The results in [1] for finite sample size can easily be utilized to arrive at the bounds when the sample size tends to infinity. For instance, given that *m* is the sample mean and '*s*' is the sample standard deviation, from [1], we have that

$$Prob(\ |X-m| \geq km\ ) \leq \left(1-\frac{1}{N}\right)\left(\frac{s^2}{k^2m^2}\right)+\frac{1}{N}.$$

Thus, as N tends to infinity, we have that

$$Prob(\ |X-m| \geq km\ ) \leq \left(\frac{s^2}{k^2m^2}\right).$$

Note: We expect the reasoning utilized in the above discussion to lead to novel and interesting inequalities and bounds

4. **Conclusion**:

   In this research paper, using the Chebyshev inequality, an asymptotic lower bound related to the Cumulative Distribution Function of a random variable is proved. Also, using Mitzenmacher-Upfal family of tail bounds, an interesting asymptotic lower bound is derived. Further, using Lyapunov inequality, Mitzenmacher-Upfal tail bounds are modified. We expect the results in the research paper to have theoretical as well as practical implications.

REFERENCES:

1. Chebyshev's Inequality….Wikipedia Page
2. G. Rama Murthy,"Transient and Equilibrium Analysis of Computer Networks: Finite Memory and Matrix Geometric Recursions", **PhD thesis, Purdue University, West Lafayette, USA, 1989.**
3. A. Blum, J. Hopcroft and R. Kannan,"Foundations of Data Science," Cambridge University Press