



Tail Probabilities: Cumulative Distribution Function: Asymptotic Lower Bounds

Rama Murthy Garimella

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

December 18, 2024

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 common cumulative distribution function evaluated at a positive real number. We expect more results of such nature to be discovered and proved.

1. Introduction:

Probability theory enables capturing certain types of uncertainties 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 inequality, Chernoff bound led to the research area of large deviations theory. Many important results were derived and applied by probabilists. 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 science 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 bounds 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 prove 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

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

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: $\{X_1, X_2, \dots, X_N\}$ (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

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

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

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

Thus, as N tends to infinity, we have

$$\lim_{n \rightarrow \infty} (\tilde{F}(k\sigma, k\sigma, \dots, k\sigma))^N \geq e^{-k^2}.$$

By independence of the random variables, we have that

$$\lim_{n \rightarrow \infty} (F(k\sigma))^{N^2} \geq e^{-k^2}.$$

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)

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

\ Equivalently, we have that

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

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

$$F(k E(|X_i - \mu|^n)^{\frac{1}{n}})^n \geq \left(1 - \frac{1}{k^n}\right)^n$$

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

$$\lim_{n \rightarrow \infty} (F(k E(|X_i - \mu|^n)^{\frac{1}{n}}))^n \geq e^{-k}.$$

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

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