Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

Generalized Markov Inequality: Extensions, Numerical Illustrations, and Multivariate Chernoff Bounds

Nouara Lazri[1,2], Ahlem Djebar[2]

[1]Higher School of Management Sciences Annaba, Algeria

lazri.nouara@essg-annaba.dz

[2]LaPS laboratory, Badji Mokhtar -Annaba University, Box 12, Annaba, 23000 Algeria

ahlem.djebar@univ-annaba.dz

Article History:

Received: 12-01-2025

Revised: 15-02-2025

Accepted: 01-03-2025

Abstract : The classical Markov inequality provides a simple yet powerful bound on tail probabilities of non-negative random variables. In this paper, we explore a generalization of the Markov inequality that employs convex, non-decreasing functions to yield more flexible and often tighter bounds. We present three distinct proofs of the generalized inequality, including one based on Jensen's inequality. Several convex functions such as $\phi(x) = x^2$ and $\phi(x) = e^{\lambda x}$ are examined, and their impact on the tightness of probabilistic bounds is illustrated through detailed numerical examples. Furthermore, we extend the analysis to the multivariate setting and derive a version of the Chernoff bound for vector-valued random variables. These results are particularly relevant in areas such as large deviations, risk theory, and high-dimensional machine learning, where sharp tail bounds play a critical role. Theoretical insights are complemented with numerical illustrations to highlight practical implications.

Keywords : Markov inequality; Chernoff bound; Multivariate Chernoff Bound; Portfolio Risk Assessment.

Introduction

Probability inequalities are central to the study of stochastic processes, statistical inference, and theoretical computer science. Among the most elementary and powerful tools is the classical *Markov inequality*, which provides a bound on the probability that a non-negative random variable exceeds a certain threshold in terms of its expected value. Specifically, if X is a nonnegative random variable and $a > 0$, then:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

This inequality is a cornerstone of probability theory and forms the basis for more refined tools such as Chebyshev's inequality, Chernoff bounds, and various concentration inequalities [1, 2].

A natural and important generalization of the Markov inequality involves replacing the identity function $x \mapsto x$ with a non-decreasing convex function ϕ, leading to the *generalized Markov inequality*:

$$P(X \geq a) \leq \frac{E[\phi(X)]}{\phi(a)} .$$

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

This form allows for tighter and more flexible bounds and has been extensively used in areas such as large deviations theory [3], risk management [5], and information theory [4].

The purpose of this paper is multifold. We begin by presenting three distinct proofs of the generalized Markov inequality, including one based on Jensen's inequality and others rooted in fundamental properties of convex functions. We then explore how different choices of ϕ, such as $\phi(x) = x^2$ and $\phi(x) = e^{\lambda x}$, yield bounds of varying tightness. These are illustrated with detailed numerical examples to highlight their practical implications.

Beyond the univariate case, we extend the discussion to the multivariate setting, where we derive a multivariate version of the Chernoff bound using convex analysis and optimization. Such bounds are essential in high dimensional statistics and machine learning, particularly in the analysis of generalization, robustness, and large deviations of vector-valued random variables.

Statement of the Generalized Markov Inequality

Proposition 1. Let X be a non-negative random variable, and let $\phi : [0, \infty) \to [0, \infty)$ be a convex and non-decreasing function. Then, for any $a > 0$ such that $\phi(a) > 0$, the following inequality holds:

$$P(X \geq a) \leq \frac{E[\phi(X)]}{\phi(a)} . \qquad (1)$$

Proof 1: Using Jensen's Inequality

Proposition 2. *Under the assumptions of Proposition 1, the inequality can*

be proven via Jensen's inequality.

Proof. Let $\mathrm{I}_{\{X \geq a\}}$ denote the indicator function of the event $\{X \geq a\}$.

Then:

$$E[\phi(X)] \geq E[\phi(X) \cdot \mathrm{I}_{\{X \geq a\}}].$$

Since ϕ is non-decreasing and convex, on the event $\{X \geq a\}$, we have $\phi(X) \geq \phi(a)$, hence:

$$E[\phi(X)] \geq \phi(a) \cdot P(X \geq a).$$

Dividing both sides by $\phi(a) > 0$ gives the desired result:

$$P(X \geq a) \leq \frac{E[\phi(X)]}{\phi(a)}.$$

□

Proof 2: Decomposition via Conditioning

Proposition 3. *The inequality in Proposition 1 also follows from a conditional decomposition of expectation.*

Proof.

Decompose the expectation over disjoint events:

$$E[\phi(X)] = E[\phi(X) \mid X < a] \cdot P(X < a) + E[\phi(X) \mid X \geq a] \cdot P(X \geq a).$$

Since ϕ is non-decreasing and $X \geq a$ on the second event,

$$E[\phi(X) \mid X \geq a] \geq \phi(a).$$

Therefore,

$$E[\phi(X)] \geq \phi(a) \cdot P(X \geq a),$$

which again yields:

$$P(X \geq a) \leq E[\phi(X)]\, \phi(a)\,.$$

Proof 3: Direct Inequality via Indicator Function

Proposition 4. *The inequality in Proposition 1 can also be obtained using a simple bound on* $\phi(X)$.

Proof.

We directly observe that:

$$\phi(X) \geq \phi(a) \cdot \mathrm{I}_{\{X \geq a\}},$$

since ϕ is non-decreasing. Taking expectations:

$$E[\phi(X)] \geq E[\phi(a) \cdot \mathrm{I}_{\{X \geq a\}}] = \phi(a) \cdot P(X \geq a).$$

Dividing both sides by $\phi(a)$ gives the required inequality:

$$P(X \geq a) \leq \frac{E[\phi(X)]}{\phi(a)}.$$

□

Remark.

Each of the above proofs highlights a different perspective: Jensen's inequality emphasizes convexity, conditioning showcases probabilistic decomposition, and the indicator function approach provides a direct algebraic bound.

Multivariate Chernoff Bound

In probability theory and statistical learning, the Chernoff bound is a powerful exponential inequality that provides tight upper bounds on the tail probabilities of sums or functions of random variables. The classical (univariate) Chernoff bound is widely used in analyzing the performance of randomized algorithms, information theory, and risk management. In multivariate settings, the inequality generalizes to vectors of random variables, allowing for simultaneous control of tail probabilities in multiple dimensions (see Theorem 1).

Theorem 1 (Multivariate Chernoff Bound)**.**

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top \in \mathbb{R}^d$ *be a random vector, and suppose its moment generating function*

$$M_{\boldsymbol{X}}(\boldsymbol{\lambda}) = E\left[e^{\boldsymbol{\lambda}^\top \boldsymbol{X}}\right]$$

is finite for all $\boldsymbol{\lambda} \in \mathbb{R}^d$. *Then, for any vector* $\mathbf{a} \in \mathbb{R}^d$,

$$P(\boldsymbol{X} \geq \boldsymbol{a}) \leq inf_{\boldsymbol{\lambda} \succ \mathbf{0}} \frac{M_{\boldsymbol{X}}(\boldsymbol{\lambda})}{e^{\boldsymbol{\lambda}^\top \boldsymbol{a}}}$$

where $\mathbf{X} \succeq \mathbf{a}$ *means* $X_i \geq a_i$ *for all i* = 1, . . . , *d, and* $\boldsymbol{\lambda} \succ \mathbf{0}$ *means* $\lambda_i > 0$ *for all i.*

Proof.

Let $\lambda \in \mathbb{R}^d$ be such that $\lambda_i > 0$ for all *i*. Consider the event $\mathbf{X} \succeq \mathbf{a}$, i.e., $X_i \geq a_i$ for all *i*. On this event, we have :

$\boldsymbol{\lambda}^\top \boldsymbol{X} \geq \boldsymbol{\lambda}^\top \boldsymbol{a}$.

Therefore,

$P(\boldsymbol{X} \geq \boldsymbol{a}) = P\left(e^{\boldsymbol{\lambda}^\top \boldsymbol{X}} \geq e^{\boldsymbol{\lambda}^\top \boldsymbol{a}}\right)$.

Now, apply Markov's inequality to the non-negative random variable $e^{\boldsymbol{\lambda}^\top \boldsymbol{X}}$:

$P\left(e^{\boldsymbol{\lambda}^\top \boldsymbol{X}} \geq e^{\boldsymbol{\lambda}^\top \boldsymbol{a}}\right) \leq \frac{E\left[e^{\boldsymbol{\lambda}^\top \boldsymbol{X}}\right]}{e^{\boldsymbol{\lambda}^\top \boldsymbol{a}}} = \frac{M_{\boldsymbol{X}}(\boldsymbol{\lambda})}{e^{\boldsymbol{\lambda}^\top \boldsymbol{a}}}$

Since this inequality holds for any $\boldsymbol{\lambda} \succ \mathbf{0}$, we can minimize the right-hand side to obtain the tightest bound:

$$P(\boldsymbol{X} \geq \boldsymbol{a}) \leq inf_{\boldsymbol{\lambda} \succ \mathbf{0}} M_{\boldsymbol{X}}(\boldsymbol{\lambda})\, e^{\boldsymbol{\lambda}^\top \boldsymbol{a}}.$$

This completes the proof.

This result is particularly useful when dealing with rare events in high dimensions, such as the probability that multiple components of a random vector simultaneously exceed their respective thresholds. Direct computation of such joint tail probabilities is often intractable, especially when dependencies exist between components. The Chernoff bound circumvents this by converting the tail probability problem into an optimization over exponential moments

The key idea behind the Chernoff method is based on Markov's inequality :

$\mathrm{P}(\mathbf{X} \succeq \mathbf{a}) = \mathrm{P}\left(e^{\boldsymbol{\lambda}^\top \mathbf{X}} \geq e^{\boldsymbol{\lambda}^\top \boldsymbol{a}}\right) \leq \frac{E\left[e^{\boldsymbol{\lambda}^\top \boldsymbol{X}}\right]}{e^{\boldsymbol{\lambda}^\top \boldsymbol{a}}}$,

for any $\boldsymbol{\lambda} \succ \mathbf{0}$. The tightest bound is then obtained by minimizing this ratio over all such vectors $\boldsymbol{\lambda}$.

The multivariate Chernoff bound finds use in several area such as:

Finance: bounding joint loss probabilities in multi-asset portfolios.

Reliability: evaluating failure probabilities in redundant systems.

Machine Learning: analyzing the generalization error of vector-valued predictors.

Information Theory: bounding decoding error probabilities for vector codes.

Application: Portfolio Risk Assessment

In financial risk management, the joint behavior of multiple asset returns is crucial for understanding extreme market scenarios.

Let $\boldsymbol{X} = (X_1, X_2, \dots, X_d)^\top \in \mathbb{R}^d$ represent the random vector of returns for d financial assets. Suppose an investor allocates capital according to a weight vector $\mathbf{w} \in \mathbb{R}^d$, where

$$\sum_{i=1}^{d} w_i = 1 \; and \; w_i \geq 0$$

The portfolio return is given by $R = \boldsymbol{w}^\top \boldsymbol{X}$.

In risk-sensitive applications such as stress testing or regulatory compliance (e.g., Basel III), it is important to bound the probability that all asset returns exceed a specified threshold. Specifically, for a threshold vector $\mathbf{a} \in \mathbb{R}^d$, we are interested in estimating the joint tail probability

P($\mathbf{X} \succeq \mathbf{a}$) = P($X_1 \geq a_1, \dots, X_d \geq a_\mathrm{d}$).

Assuming the moment generating function $M_\mathrm{X}(\boldsymbol{\lambda})$ = E[$e^{\boldsymbol{\lambda}^\top \mathbf{X}}$] is finite for

$\boldsymbol{\lambda} \in \mathbb{R}^d$, the multivariate Chernoff bound provides the inequality:

$$P(\boldsymbol{X} \geq \boldsymbol{a}) \leq inf_{\boldsymbol{\lambda} > \mathbf{0}} \frac{M_{\boldsymbol{X}}(\boldsymbol{\lambda})}{e^{\boldsymbol{\lambda}^\top \boldsymbol{a}}}$$

This offers a conservative yet computationally tractable upper bound on rare-event probabilities.

Numerical Illustration: Bivariate Normal Portfolio

Consider a portfolio composed of two assets, where the return vector

$\boldsymbol{X} = (X_1, X_2)^\top$ follows a bivariate normal distribution:

$$\boldsymbol{X} \sim N2\left(\mu = \begin{pmatrix} \mathbf{0.05} \\ \mathbf{0.04} \end{pmatrix}, \Sigma = \begin{pmatrix} 0.01 & 0.002 \\ 0.002 & 0.008 \end{pmatrix}\right).$$

We aim to compute an upper bound on the probability: P($X_1 \geq 0.06$, $X_2 \geq 0.05$),

using the multivariate Chernoff bound with a trial vector $\boldsymbol{\lambda} = (40, 40)^\top$.

For a multivariate normal distribution, the moment generating function is:

$M_{\boldsymbol{X}}(\boldsymbol{\lambda}) = exp\left(\boldsymbol{\lambda}^\top \mu + \frac{1}{2} \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda}\right)$.

Substituting into the bound, we have:

$$P(\boldsymbol{X} \geq \boldsymbol{a}) \leq \exp\left(\boldsymbol{\lambda}^\top(\mu - \boldsymbol{a}) + \frac{1}{2} \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda}\right),$$

with $\boldsymbol{a} = (0.06, 0.05)^\top$. Computing:

$$\boldsymbol{\lambda}^{\top}(\mu - \boldsymbol{a}) = 40(-0.01) + 40(-0.01) = -0.8,$$

$$\boldsymbol{\lambda}^{\top}\Sigma\boldsymbol{\lambda} = (40, 40)\begin{pmatrix} 0.01 & 0.002 \\ 0.002 & 0.008 \end{pmatrix}\begin{pmatrix} 40 \\ 40 \end{pmatrix} = 35.2.$$

Hence,

$$P(\boldsymbol{X} \geq \boldsymbol{a}) \leq exp(-0.8 + 17.6) = exp(16.8) \approx 1.95 \times 10^{7}.$$

Since this bound exceeds 1, it is uninformative in this case. However, for more extreme thresholds (e.g., $\boldsymbol{a} = (0.08, 0.07)^{\top}$), the bound becomes more meaningful:

$\boldsymbol{\lambda}^{\top}(\mu - \boldsymbol{a})$ =40(-0.03) $+ 40(-0.03) = -2.4 \Rightarrow P(\boldsymbol{X} \geq \boldsymbol{a}) \leq exp(15.2) \approx 4.02 \times 10^{6}$.

The multivariate Chernoff bound provides a tractable method to estimate joint tail probabilities. Although the bound may be loose for modest deviations, it becomes valuable in stress testing scenarios where evaluating the probability of rare joint exceedances is crucial.

Example: Multivariate Normal Distribution

Consider a random vector $\boldsymbol{X} = (X_1, X_2)^{\top}$ following a multivariate normal distribution:

$$\boldsymbol{X} \sim N(\mu, \Sigma),$$

where the mean vector is:

$$\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix} ,$$

and the covariance matrix is:

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix} .$$

The moment generating function (MGF) of the multivariate normal distribution is given by:

$$M_{\boldsymbol{X}}(\boldsymbol{\lambda}) = exp\,(\boldsymbol{\lambda}^{\top}\mu + \frac{1}{2}\boldsymbol{\lambda}^{\top}\Sigma\boldsymbol{\lambda}),$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^{\top}$ is a vector of parameters, and the expectation is taken over the random vector **X**.

We are interested in calculating the upper bound for the probability $P(X_1 \geq 1.5, X_2 \geq 2)$ using the **multivariate Chernoff bound**. The Chernoff bound is:

$P(\boldsymbol{X} \geq \boldsymbol{a}) \leq inf_{\boldsymbol{\lambda} \succ \boldsymbol{0}} \frac{M_{\boldsymbol{X}}(\boldsymbol{\lambda})}{e^{\boldsymbol{\lambda}^{\top}\boldsymbol{a}}}$,

where $\boldsymbol{a} = \begin{pmatrix} 1.5 \\ 2 \end{pmatrix}$ is the threshold vector.

To apply the Chernoff bound, we need to minimize the following expression:

$\frac{exp\,(\boldsymbol{\lambda}^{\top}\mu + \frac{1}{2}\boldsymbol{\lambda}^{\top}\Sigma\boldsymbol{\lambda})}{e^{\boldsymbol{\lambda}^{\top}\boldsymbol{a}}}$.

In practice, this optimization is typically solved using numerical methods (e.g., gradient descent or convex optimization techniques).

Particular case. Let us choose $\boldsymbol{\lambda} = (1, 1)^T$ and calculate the Chernoff bound.

First, we compute the exponential part of the Chernoff bound:

$e^{\boldsymbol{\lambda}^\top \boldsymbol{a}} = e^{1.5+2} = e3.5 \approx 33.115$

Next, we calculate the MGF at $\boldsymbol{\lambda} = (1, 1)^T$:

$$M_{\boldsymbol{X}}(1,1) = exp\left((1,1)^\top \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{1}{2}(1,1)^\top \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix} (1,1)^\top \right).$$

This simplifies to:

$$M_{\boldsymbol{X}}(1,1) = exp\,(1 + 2 + \tfrac{1}{2}(1 + 2 + 0.5 + 0.5)) = exp\,(3 + 2) = exp(5).$$

Thus, the Chernoff bound is:

$$P(X_1 \geq 1.5, X_2 \geq 2) \leq \frac{exp(5)}{e3.5} = exp(5) = 33.115$$

This gives us the upper bound on the probability. In practice, numerical optimization can be used to obtain tighter bounds by adjusting $\boldsymbol{\lambda}$.

Conclusion and Perspectives

In this study, we reviewed the generalized Markov inequality in various ways, showing its importance in probability theory. By looking at different convex functions, we showed that the generalization can provide tighter bounds than the traditional form. Practical examples confirmed the usefulness of these bounds, particularly regarding the tail behavior of distributions.

We also developed a multivariate Chernoff bound to analyze the joint behavior of random vectors in high-dimensional spaces. This has applications in areas like risk aggregation and machine learning.

Future research could explore using data-dependent convex functions for optimized bounds, implementing these bounds in learning algorithms, and studying generalized inequalities under dependency structures. This shows that the generalized Markov inequality is an important tool in modern probability analysis.

References

[1] S. M. Ross, *Stochastic Processes*, John Wiley & Sons, 2nd edition, 1996.
[2] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, Oxford University Press, 2013.
[3] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer-Verlag, 2nd edition, 1998.
[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, WileyInterscience, 2nd edition, 2006.
[5] P. Embrechts, A. J. McNeil, and D. Straumann, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, 2005.