A Likelihood Ratio Approach for Probabilistic Inequalities

Xinjia Chen
August 18, 2013

Abstract

We propose a new approach for deriving probabilistic inequalities based on bounding likelihood ratios. We demonstrate that this approach is more general and powerful than the classical method frequently used for deriving concentration inequalities such as Chernoff bounds. We discover that the proposed approach is inherently related to statistical concepts such as monotone likelihood ratio, maximum likelihood, and the method of moments for parameter estimation. A connection between the proposed approach and the large deviation theory is also established. We show that, without using moment generating functions, tightest possible concentration inequalities may be readily derived by the proposed approach. We have derived new concentration inequalities using the proposed approach, which cannot be obtained by the classical approach based on moment generating functions.

1 Introduction

A problem of ubiquitous significance in engineering and sciences is to obtain bounds for probabilities of random events. Formally, let \( E \) be an event defined in probability space \((\Omega, \Pr, \mathcal{F})\), where \( \Omega \) is the sample space, \( \Pr \) denotes the probability measure, and \( \mathcal{F} \) is the the \( \sigma \)-algebra. In many situations, it is desirable to obtain bounds for \( \Pr\{E\} \). For theoretical and practical purposes, the bounds are expected to be simple, insightful and as tight as possible. In general, the event \( E \) can be expressed in terms of a matrix-valued random variable \( X \). In particular, \( X \) can be a random vector or scalar. Clearly, the event \( E \) can be represented as \( \{X \in \mathcal{E}\} \), where \( \mathcal{E} \) is a certain set of deterministic matrices. In probability theory, one of the most frequently used approach for deriving inequalities for \( \Pr\{E\} \) is to bound the indicator function \( \mathbb{I}_{\{X \in \mathcal{E}\}} \) by a family of random variables having finite expectation and minimize the expectation. The central idea of this approach is to seek a family of bounding functions \( w(X, \vartheta) \) of \( X \), parameterized by \( \vartheta \in \Theta \), such that

\[
\mathbb{I}_{\{X \in \mathcal{E}\}} \leq w(X, \vartheta) \quad \text{for all } \vartheta \in \Theta.
\]

Here, the notion of inequality \( \mathbb{I} \) is that the inequality \( \mathbb{I}_{\{X(\omega) \in \mathcal{E}\}} \leq w(X(\omega), \vartheta) \) holds for every \( \omega \in \Omega \). As a consequence of the monotonicity of the mathematical expectation \( \mathbb{E}[\cdot] \),

\[
\Pr\{E\} = \mathbb{E}[\mathbb{I}_{\{X \in \mathcal{E}\}}] \leq \mathbb{E}[w(X, \vartheta)] \quad \text{for all } \vartheta \in \Theta.
\]
Minimizing the upper bound in (2) with respect to \( \vartheta \in \Theta \) yields
\[
\Pr\{E\} \leq \inf_{\vartheta \in \Theta} \mathbb{E}[w(X, \vartheta)].
\] (3)

Classical inequalities such as Chebyshev inequality, Markov inequality, Cantelli’s inequality, Bernstein’s inequalities, Chernoff bounds, Benett’s inequalities, Hoeffding’s inequalities can be derived by this approach with various bounding functions. To illustrate, the derivation of some fundamental inequalities is presented follows.

**Chebyshev inequality:** Let \( X \) be a random variable with mean \( \mu \) such that \( \mathbb{E}[|X - \mu|^s] < \infty \) for \( s \in \Theta \), where \( \Theta \) is a subset of positive real numbers. Let \( \varepsilon > 0 \). Define \( w(X, s) = \left( \frac{|X - \mu|^s}{\varepsilon} \right)^s \) for \( s \in \Theta \). Then, \( \mathbb{P}_{|X - \mu| \geq \varepsilon} \leq w(X, s) \) for \( s \in \Theta \). It follows from (3) that \( \Pr\{|X - \mu| \geq \varepsilon\} \leq \inf_{s \in \Theta} \frac{\mathbb{E}[|X - \mu|^s]}{\varepsilon^s} \), which is referred to as Chebyshev inequality in the particular case that \( \Theta = \{2\} \).

**Markov inequality:** Let \( X \) be a non-negative random variable such that \( \mathbb{E}[X^s] < \infty \) for \( s \in \Theta \), where \( \Theta \) is a subset of positive real numbers. Let \( \gamma > 0 \). Define \( w(X, s) = \left( \frac{X^s}{\gamma} \right)^s \) for \( s \in \Theta \). Then, \( \mathbb{P}_{X \geq \gamma} \leq w(X, s) \) for \( s \in \Theta \). It follows from (3) that \( \Pr\{X \geq \gamma\} \leq \inf_{s \in \Theta} \frac{\mathbb{E}[X^s]}{\gamma^s} \), which is referred to as Markov inequality in the particular case that \( \Theta = \{1\} \).

**Cantelli’s inequality:** Let \( X \) be a random variable with mean \( \mu \) and finite variance \( \sigma^2 > 0 \). Let \( \varepsilon > 0 \). Define \( w(X, s) = \left( \frac{X + s}{\varepsilon + \mu + s} \right)^2 \) for \( s > -(\varepsilon + \mu) \). Then, \( \mathbb{P}_{X - \mu \geq \varepsilon} \leq w(X, s) \) and \( \mathbb{E}[w(X, s)] = \frac{(\mu + s)^2 + \sigma^2}{(\varepsilon + \mu + s)^2} \) for all \( s > -(\varepsilon + \mu) \). It follows from (3) that \( \Pr\{X - \mu \geq \varepsilon\} \leq \inf_{s > -(\varepsilon + \mu)} \frac{(\mu + s)^2 + \sigma^2}{(\varepsilon + \mu + s)^2} = \frac{\sigma^2}{\varepsilon^2 + \sigma^2} \).

**Chernoff bound:** Let \( X \) be a random variable such that the moment generating function \( \phi(s) = \mathbb{E}[e^{sX}] \) is finite for \( s \in \Theta \), where \( \Theta \) is a subset of positive real numbers. Let \( \gamma \) be a real number. Define \( w(X, s) = e^{s(X - \gamma)} \) for \( s \in \Theta \). Then, \( \mathbb{P}_{X \geq \gamma} \leq w(X, s) \) for \( s \in \Theta \). It follows from (3) that \( \Pr\{X \geq \gamma\} \leq \inf_{s \in \Theta} \mathbb{E}[w(X, s)] = \inf_{s \in \Theta} e^{-\gamma s} e^{s\mathbb{E}[X]} \).

As can be seen from the above discussion, the central idea of this approach is to seek a family of bounding functions \( w(X, \vartheta) \), \( \vartheta \in \Theta \) such that the mathematical expectation \( \mathbb{E}[w(X, \vartheta)] \) is convenient for evaluation and minimization. We call this technique of deriving probabilistic inequalities as the mathematical expectation (ME) method, in view of the crucial role played by the mathematical expectation of bounding functions. The ME method is a very general approach. However, it has two drawbacks.

First, in some situations, the mathematical expectation \( \mathbb{E}[w(X, \vartheta)] \) may be intractable. For example, if \( X \) is a hypermetric random variable and one wishes to derive the Chernoff bound for \( \Pr\{X \geq \gamma\} \) with bounding function \( w(X, s) = e^{s(X - \gamma)} \), it is not possible to obtain an tractable expression for \( \mathbb{E}[w(X, s)] \) except some crude bounds. Similar difficulty is encountered if \( X \) is a generalized Poisson random variable. In the general case that \( X \) is matrix-valued or vector-valued, it may be more cumbersome to evaluate and minimize \( \mathbb{E}[w(X, \vartheta)] \).

Second, the ME method may not fully exploit the information of the underlying distribution, since the mathematical expectation is a only quantity of summary for the distribution. The issue may arise when the probability density or mass function of \( X \) is available.

In this paper, we would like to propose a more general approach for deriving probabilistic inequalities, aiming at overcoming the drawbacks of the ME method. Let \( f(.) \) denote the probability density function (pdf) or probability mass function (pmf) of \( X \). The primary idea of the proposed approach is to seek a family of pdf or pmf \( g(.; \vartheta) \), parameterized by \( \vartheta \in \Theta \), and a deterministic function \( \Lambda(\vartheta) \) of \( \vartheta \in \Theta \) such
that for all $\theta \in \Theta$, the indicator function $\mathbb{I}_{\{X \in \mathcal{E}\}}$ is bounded from above by the product of $\Lambda(\theta)$ and the likelihood ratio $\frac{g(\mathcal{E}, \theta)}{f(X, \theta)}$. Then, the probability $\Pr\{X \in \mathcal{E}\}$ is bounded from above by the infimum of $\Lambda(\theta)$ with respect to $\theta \in \Theta$. Due to the central role played by the likelihood ratio, this technique of deriving probabilistic inequalities is referred to as the likelihood ratio (LR) method. It can be demonstrated that the ME method is actually a special technique of the LR method. From the perspective of the LR method, the idea of the ME method is to construct a family of pdf or pmf $g(\cdot, \theta)$ via multiplying the pdf or pmf $f(\cdot)$ of $X$ by the bounding function $w(\cdot, \theta)$ and performing a normalization. In other words, the bounding function $w(\cdot, \theta)$ is used as a weight function to modify the distribution of $X$. The deterministic function $\Lambda(\theta)$ can be taken as $\mathbb{E}[w(X, \theta)]$. This approach for constructing the parameterized distribution $g(\cdot, \theta)$ and the function $\Lambda(\theta)$ is useful provided that the mathematical expectation $\mathbb{E}[w(X, \theta)]$ is tractable. In addition to this approach of constructing the parameterized distribution $g(\cdot, \theta)$ and the deterministic function $\Lambda(\theta)$, the LR method opens up another possible avenue. In many situations, since the pdf or pmf $f(\cdot)$ of $X$ belongs to a family of distributions parameterized by $\theta \in \Theta$, the desired parameterized distribution $g(\cdot, \theta)$ can be directly obtained by restricting the parameter space $\Theta$ to its subset $\Theta$. In many cases, by appropriately choosing the subset $\Theta$, the desired deterministic function $\Lambda(\theta)$ can obtained by the monotonicity or concavity of the likelihood ratio. In this way, the hassle of using a weight function is eliminated. Since this approach does not involve the weight function, the derivation of bounds for the probability $\Pr\{X \in \mathcal{E}\}$ can be extremely simple. Moreover, this approach may also be powerful in situations that the ME method is ineffective.

In the proposed LR method, an important step is to minimize the deterministic function $\Lambda(\theta)$ to obtain the tightest bound. We realize that this minimization process is relevant to the maximization of a likelihood function. This implies an inherent connection between the LR method for probabilistic inequalities and Fisher’s concept of maximum likelihood estimation of parameters. In situations that the maximum likelihood estimation is not mathematically convenient, we propose to use Pearson’s method of moments as an alternative technique to obtain simple bounds for $\Pr\{X \in \mathcal{E}\}$. Of course, the bounds obtained from the method of moments is relatively more conservative than their counterparts obtained from the maximum likelihood estimation.

We have established a link between the LR method and the large deviation theory. Under mild conditions, we have obtained explicit bounds for rate functions, which are tight for a wide class of distributions. We have applied the LR method to derive new concentration inequalities. In particular, we have derived bounds for exponential families. We show that the LR approach and the Berry-Essen inequality can be used together to derive tighter concentration inequalities. We have established inequalities for multivariate generalized hypergeometric and inverse hypergeometric distributions, Dirichlet distribution and matrix gamma distribution.

The remainder of the paper is organized as follows. In Section 2, we introduce the fundamentals of the LR method. Specially, we propose the general principle and establish the connection between the LR method and statistical parameter estimation. In Section 3, we apply the LR method to the development of concentration inequalities. In Section 4, we explore the connection between the LR method and the large deviation theory. In Section 5, we establish a unified approach for deriving concentration inequalities for exponential families. In particular, we demonstrate that tightest possible concentration inequalities can be derived without using moment generating functions. In Section 6, we apply the LR method to derive new concentration inequalities for the binomial distribution, hypergeometric distribution and generalized Poisson distribution. In Section 7, we apply the LR method to establish concentration inequalities for multivariate generalized hypergeometric and inverse hypergeometric distributions, Dirichlet distribution.
and matrix gamma distribution. Section 8 is the conclusion. Most proofs are given in Appendices.

Throughout this paper, we shall use the following notations. Let $\mathbb{R}$ denote the set of all real numbers. Let $\mathbb{Z}$ denote the set of all integers. Let $\mathbb{Z}^+$ denote the set of all nonnegative integers. Let $\mathbb{N}$ denote the set of all positive integers. Let $1_E$ denote the indicator function such that $1_E = 1$ if $E$ is true and $1_E = 0$ otherwise. We use the notation $\binom{t}{k}$ to denote a generalized combinatoric number in the sense that

$$\binom{t}{k} = \frac{\prod_{i=1}^{k}(t - i + 1)}{k!} = \frac{\Gamma(t + 1)}{\Gamma(k + 1)\Gamma(t - k + 1)}, \quad \binom{t}{0} = 1,$$

where $t$ is a real number and $k$ is a non-negative integer. We use $\overline{X}_n$ to denote the average of random variables $X_1, \cdots, X_n$, that is, $\overline{X}_n = \frac{\sum_{i=1}^{n}X_i}{n}$. The notation $\top$ denotes the transpose of a matrix. The trace of a matrix is denoted by $\text{tr}$. We use pdf and pmf to represent probability density function and probability mass function, respectively. The notation \(\Leftrightarrow\) means “if and only if”. The other notations will be made clear as we proceed.

2 Likelihood Ratio Method

In this section, we shall propose a new method for deriving probabilistic inequalities, referred to as the likelihood ratio (LR) method.

Let $E$ be an event which can be expressed in terms of matrix-valued random variable $X$, where $X$ is defined on the sample space $\Omega$ and $\sigma$-algebra $\mathcal{F}$ such that the true probability measure is one of two measures $\Pr$ and $\mathbb{P}_\vartheta$. Here, the measure $\Pr$ is determined by pdf or pmf $f(.)$. The measure $\mathbb{P}_\vartheta$ is determined by pdf or pmf $g(., \vartheta)$, which is parameterized by $\vartheta \in \Theta$. The subscript in $\mathbb{P}_\vartheta$ is used to indicate the dependence on the parameter $\vartheta$. Clearly, there exists a set, $\mathcal{E}$, of deterministic matrices of the same size as $X$ such that $E = \{X \in \mathcal{E}\}$. In many applications, it is desirable to obtain an upper bound for the probability $\Pr\{E\}$.

2.1 General Principle

The LR method is based on the following general result.

**Theorem 1** Assume that there exists a function $\Lambda(\vartheta)$ of $\vartheta \in \Theta$ such that

$$f(X) 1_{\{X \in \mathcal{E}\}} \leq \Lambda(\vartheta) g(X, \vartheta) \text{ for all } \vartheta \in \Theta.
\tag{4}$$

Then,

$$\Pr\{E\} \leq \inf_{\vartheta \in \Theta} \Lambda(\vartheta) \mathbb{P}_\vartheta\{E\} \leq \inf_{\vartheta \in \Theta} \Lambda(\vartheta).
\tag{5}$$

The notion of the inequality in (4) is that $f(X(\omega)) 1_{\{X(\omega) \in \mathcal{E}\}} \leq \Lambda(\vartheta) g(X(\omega), \vartheta)$ for every $\omega \in \Omega$. We call the function $\Lambda(\vartheta)$ in (1) as likelihood-ratio bounding function. Theorem 1 asserts that the probability of event $E$ is no greater than the likelihood ratio bounding function. In a comparison of (3) with (4), it can be seen that the distribution of $X$ is directly involved in (4), which indicates that the LR method allows for more direct and complete use of the distribution of $X$. As can be seen from Theorem 1, the key of the LR method is to find the likelihood ratio bounding function which is tight and amenable for minimization as well.

To establish Theorem 1 note that if $f(.)$ is a pdf, then it follows from (4) that

$$\Pr\{E\} = \int_{x \in \mathcal{E}} f(x) \, dx \leq \int_{x \in \mathcal{E}} \Lambda(\vartheta) g(x, \vartheta) \, dx = \Lambda(\vartheta) \int_{x \in \mathcal{E}} g(x, \vartheta) \, dx \text{ for all } \vartheta \in \Theta.$$
Similarly, if $f(\cdot)$ is a pmf, then
\[
\Pr\{E\} = \sum_{x \in \mathcal{E}} f(x) \leq \sum_{x \in \mathcal{E}} \Lambda(\vartheta) g(x, \vartheta) = \Lambda(\vartheta) \sum_{x \in \mathcal{E}} g(x, \vartheta) \quad \text{for all } \vartheta \in \Theta.
\]
This implies that
\[
\Pr\{E\} \leq \Lambda(\vartheta) \Pr_{\vartheta}\{E\} \leq \Lambda(\vartheta) \quad \text{for all } \vartheta \in \Theta. \tag{6}
\]
Since the inequalities in (6) hold for any $\vartheta \in \Theta$, minimizing the bounds with respect to $\vartheta \in \Theta$ yields (5).

### 2.2 Construction of Parameterized Distributions

In the sequel, we will discuss two approaches for constructing parameterized distributions $g(\cdot, \vartheta)$ that are required for the application of the LR method.

#### 2.2.1 Weight Function

A natural approach to construct parameterized distribution $g(\cdot, \vartheta)$ is to modify the pdf or pmf $f(\cdot)$ by multiplying it with a parameterized function and performing a normalization. Specifically, let $w(\cdot, \vartheta)$ be a non-negative function with parameter $\vartheta \in \Theta$ such that $E[w(X, \vartheta)] < \infty$ for all $\vartheta \in \Theta$, where the expectation is taken under the probability measure $\Pr$ determined by $f(\cdot)$. Define a family of distributions as
\[
g(x, \vartheta) = \frac{w(x, \vartheta) f(x)}{E[w(X, \vartheta)]}
\]
for $\vartheta \in \Theta$ and $x$ in the range of $X$. In view of its role in the modification of $f(\cdot)$ as $g(\cdot, \vartheta)$, the function $w(\cdot, \vartheta)$ is called a weight function. Note that
\[
f(X) w(X, \vartheta) = E[w(X, \vartheta)] g(X, \vartheta) \quad \text{for all } \vartheta \in \Theta. \tag{7}
\]
For simplicity, we choose the weight function such that the condition (\ref{cond1}) is satisfied. Combining (\ref{cond1}) and (7) yields
\[
f(X) \mathbb{1}_{\{X \in \mathcal{E}\}} \leq f(X) w(X, \vartheta) = E[w(X, \vartheta)] g(X, \vartheta) \quad \text{for all } \vartheta \in \Theta.
\]
Thus, the likelihood ratio bounding function can be taken as
\[
\Lambda(\vartheta) = E[w(X, \vartheta)] \quad \text{for } \vartheta \in \Theta.
\]
It follows from Theorem 1 that
\[
\Pr\{E\} \leq \inf_{\vartheta \in \Theta} \Lambda(\vartheta) \Pr_{\vartheta}\{E\} \leq \inf_{\vartheta \in \Theta} \Lambda(\vartheta).
\]
Thus, we have demonstrated that the ME method is actually a special technique of the LR method.

#### 2.2.2 Parameter Restriction

In Section 2.2.1, we have discussed an approach for constructing the parameterized distribution $g(\cdot, \vartheta)$ based on weight function. If the mathematical expectation of $E[w(X, \vartheta)]$ is not tractable, then this approach is not effective for deriving tight probabilistic inequalities. To overcome this problem, we propose another approach, which is based on the idea of restricting parameter space.

In many situations, the pdf or pmf $f(\cdot)$ of $X$ comes from a family of distributions parameterized by $\vartheta \in \Theta$. If so, then the parameterized distribution $g(\cdot, \vartheta)$ can be taken as the subset of pdf or pmf with
parameter $\vartheta$ contained in a subset $\Theta$ of parameter space $\Theta$. By appropriately choosing the subset $\Theta$, the deterministic function $\Lambda(\vartheta)$ may be readily obtained. To illustrate the idea, consider $\Pr\{X_n \geq z\}$, where $z \in [\frac{1}{2}, 1]$ and $X_n = \sum_{i=1}^{n} X_i$ with $X_1, \ldots, X_n$ being i.i.d. random variables uniformly distributed over interval $[0, 1]$. Let $X = [X_1, X_2, \ldots, X_n]^\top$ and $x = [x_1, x_2, \ldots, x_n]^\top$. The pdf of $X$ is easily seen to be

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x_i \leq 1 \text{ for } i = 1, \ldots, n, \\ 0 & \text{otherwise} \end{cases}$$

Note that the pdf $f(x)$ is contained in the family

$$g(x, \theta) = \begin{cases} [C(\theta)]^n \exp(\theta \sum_{i=1}^{n} x_i) & \text{if } 0 \leq x_i \leq 1 \text{ for } i = 1, \ldots, n, \\ 0 & \text{otherwise} \end{cases}$$

where

$$C(\theta) = \frac{1}{\int_0^1 e^{\theta x} dx} = \begin{cases} e^{-\theta} & \text{for } \theta \neq 0, \\ 1 & \text{for } \theta = 0 \end{cases}$$

with parameter $\theta \in \mathbb{R}$, that is, the parameter space $\Theta$ is the set of all real numbers. Letting $\theta = 0$ yields $f(x) = g(x, 0)$. If we restrict the parameter space as $\Theta = (0, \infty)$, then it can be readily seen that

$$f(X) \mathbb{I}_{\{X_n \geq z\}} \leq \Lambda(\theta) \ g(X, \vartheta) \quad \text{for all } \vartheta \in \Theta, 
$$

where

$$\Lambda(\vartheta) = \frac{[C(0)]^n \exp(nz \times 0)}{[C(\theta)]^n \exp(n\vartheta z)} = \left( e^{\theta} - 1 \right)^n \left( e^{-\vartheta z} \right)^n.$$

It follows from (8) and Theorem 1 that

$$\Pr\{X_n \geq z\} \leq \inf_{\vartheta \in (0, \infty)} \left( e^{\theta} - 1 \right)^n \left( e^{-\vartheta z} \right)^n = \left( e^{\nu} - 1 \right)^n \left( e^{-\nu z}\right),$$

where $\nu > 0$ is the unique number satisfying $z = 1 + \frac{1}{e^{\nu} - 1} = \frac{1}{\nu}$. 

### 2.3 Bounding Probability of Order Relation

Let $Z$ be a random matrix which can be expressed as a function of random matrix $X$. Let $z$ be a deterministic matrix in the range of $Z$. Let “$\prec$” denote a partial order relation on the range of $Z$. It is a frequent problem to obtain an upper bound for the probability $\Pr\{Z \prec z\}$. For this purpose, we seek a nonnegative function $\mathcal{M}(., \vartheta)$, parameterized by $\vartheta \in \Theta$, which satisfies the following requirements:

(i) For arbitrary parametric value $\vartheta \in \Theta$, $\mathcal{M}(u, \vartheta) \leq \mathcal{M}(z, \vartheta)$ if $u \prec z$, where $u$ is in the range of $Z$.

(ii) $f(X) \mathbb{I}_{\{Z \prec z\}} \leq \mathcal{M}(Z, \vartheta) \ g(X, \vartheta) \quad \text{for all } \vartheta \in \Theta.$

As a consequence of (i) and (ii), we can write (9) as

$$f(X) \mathbb{I}_{\{Z \prec z\}} \leq \mathcal{M}(z, \vartheta) \ g(X, \vartheta) \quad \text{for all } \vartheta \in \Theta.$$

Applying the general principle described by Theorem 1 we have

$$\Pr\{Z \prec z\} \leq \inf_{\vartheta \in \Theta} \mathcal{M}(z, \vartheta) \ P_{\vartheta}\{Z \prec z\} \leq \inf_{\vartheta \in \Theta} \mathcal{M}(z, \vartheta).$$

It should be noted that in many situations, especially in scenarios where the underlying distribution belongs to an exponential family, the likelihood ratio $\frac{f(X)}{g(X, \vartheta)}$ can be expressed as a function $\mathcal{M}(Z, \vartheta)$ which is monotone in $Z$. Such property is called monotone likelihood ratio property, which has been extensively explored in our paper [7] for deriving probabilistic inequalities.
2.3.1 Optimal Bound with Maximum-Likelihood Estimate

Under some conditions, a connection between the LR method and Fisher’s concept of maximum-likelihood estimation can be established as follows. Assume that

\[ f(X) = \mathcal{M}(Z, \vartheta) g(X, \vartheta) \quad \text{for all } \vartheta \in \Theta \]

and that the maximum-likelihood estimator of \( \vartheta \in \Theta \) exists and can be expressed as a function \( \varphi(Z) \) of \( Z \). That is,

\[ \max_{\vartheta \in \Theta} g(X, \vartheta) = g(X, \varphi(Z)). \]

Making use of (10) and (11), we have

\[ \min_{\vartheta \in \Theta} \mathcal{M}(Z, \vartheta) = \frac{f(X)}{\max_{\vartheta \in \Theta} g(X, \vartheta)} = \frac{f(X)}{g(X, \varphi(Z))} = \mathcal{M}(Z, \varphi(Z)). \]

Without loss of generality, assume that \( z \) is contained in the range of \( Z \). Hence,

\[ \min_{\vartheta \in \Theta} \mathcal{M}(z, \vartheta) = \mathcal{M}(z, \varphi(z)) \]

and it follows from Theorem 2 that

\[ \Pr\{Z < z\} \leq \mathcal{M}(z, \varphi(z)) \Pr_{\varphi(z)}\{Z < z\} \leq \mathcal{M}(z, \varphi(z)). \]

2.3.2 Approximately Optimal Bound with Method of Moments

As can be seen from Section 2.3.1, to make the bound as tight as possible, the minimization problem becomes finding the maximum-likelihood estimate for \( \vartheta \in \Theta \). In situations that the computation of the maximum-likelihood estimate is inconvenient, other estimation techniques may be used. In this direction, we propose to use Pearson’s method of moments, which may offer an estimate which is close to the maximum-likelihood estimate. Specifically, for \( z \) in the range of \( Z \), let \( \xi \) be a number in \( \Theta \) such that \( z = E_{\xi}[Z] \), where the expectation is taken with the underlying pdf or pmf \( g(\cdot, \xi) \). Hence, \( \xi \) is actually a function of \( z \). So, we write \( \xi = \xi(z) \). In many cases, \( \varphi(z) = \xi(z) \) or \( \varphi(z) \approx \xi(z) \). If \( \xi(z) \in \Theta \), then it follows from Theorem 2 that

\[ \Pr\{Z < z\} \leq \mathcal{M}(z, \xi(z)) \Pr_{\xi(z)}\{Z < z\} \leq \mathcal{M}(z, \xi(z)). \]

3 Concentration Inequalities

Let \( X_1, X_2, \cdots, X_n \) and \( X \) be matrix-valued random variables defined on the same sample space \( \Omega \) and \( \sigma \)-algebra \( \mathcal{F} \) such that the true probability measure is one of two measures \( \Pr \) and \( P_{\vartheta} \). Here, the measure \( \Pr \) is determined by pdf or pmf \( f(\cdot) \). The measure \( P_{\vartheta} \) is determined by pdf or pmf \( g(\cdot, \vartheta) \), which is parameterized by \( \vartheta \in \Theta \). Under each of these probability measures, the \( X \)’s are independent and identically distributed. Define \( \mathbf{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \) for \( n \in \mathbb{N} \). Let \( \prec \) denote a partial order relation defined on the range of \( \mathbf{X}_n \). Let \( \mathcal{X} \) denote the union of the ranges of \( \mathbf{X}_n \) for all \( n \in \mathbb{N} \). In this setting, we have the following results.

**Theorem 2** Assume that \( \rho(x, \vartheta) \) is a concave function of \( x \in \mathcal{X} \), with parameter \( \vartheta \in \Theta \), such that \( f(X) \leq \exp(\rho(X, \vartheta)) \) \( g(X, \vartheta) \) for all \( \vartheta \in \Theta \). Let \( z \in \mathcal{X} \) be a matrix such that for \( x \in \mathcal{X} \) satisfying \( x \prec z \), the inequality \( \rho(x, \vartheta) \leq \rho(z, \vartheta) \) holds for all \( \vartheta \in \Theta \). Then,

\[ \Pr\{\mathbf{X}_n \prec z\} \leq \inf_{\vartheta \in \Theta} \exp(n\rho(z, \vartheta)) \Pr_{\vartheta}\{\mathbf{X}_n \prec z\} \leq \inf_{\vartheta \in \Theta} \exp(n\rho(z, \vartheta)). \]
See Appendix A for a proof.
In the particular case that \( f(X) = \exp(\rho(X, \vartheta)) \) \( g(X, \vartheta) \) for all \( \vartheta \in \Theta \), the connection between the LR method and statistical parameter estimation can be established as follows.

To achieve the optimal bound, consider the minimization of

\[
\exp(\rho(z, \vartheta)) = \frac{f(z)}{g(z, \vartheta)} \quad \text{subject to} \quad \vartheta \in \Theta.
\]

Let \( \varphi(X) \) be the maximum-likelihood estimator of \( \vartheta \in \Theta \) such that

\[
\max_{\vartheta \in \Theta} g(X, \vartheta) = g(X, \varphi(X))
\]

and that the function \( \varphi(.) \) is defined on the range of \( X \). Assume that \( g \) and \( \varphi \) can be extended to the range of \( X_n \) such that

\[
\max_{\vartheta \in \Theta} g(X_n, \vartheta) = g(X_n, \varphi(X_n)).
\]

Note that

\[
\min_{\vartheta \in \Theta} f(X_n) = \frac{f(X_n)}{\max_{\vartheta \in \Theta} g(X_n, \vartheta)} = \frac{f(X_n)}{g(X_n, \varphi(X_n))}.
\]

Assume that \( z \) is contained in the range of \( X_n \). Then,

\[
\min_{\vartheta \in \Theta} \frac{f(z)}{g(z, \vartheta)} = \frac{f(z)}{g(z, \varphi(z))}
\]

and it follows from [13] that

\[
\Pr\{X_n < z\} \leq \left[ \frac{f(z)}{g(z, \varphi(z))} \right]^n \Pr_{\varphi(z)}\{X_n < z\} \leq \left[ \frac{f(z)}{g(z, \varphi(z))} \right]^n.
\]

In situations that the computation of the maximum-likelihood estimate is inconvenient, the method of moments may offer an estimate which is close to the maximum-likelihood estimate. Since there is only one parameter to be estimated, it suffices to use the first moment. Specifically, let \( \xi \) be a number in \( \Theta \) such that \( z = \mathbb{E}_\xi[X] \), where the expectation is taken with the underlying pdf or pmf \( g(., \xi) \). Hence, \( \xi \) is actually a function of \( z \). So, we write \( \xi = \xi(z) \). In many cases, \( \varphi(z) \approx \xi(z) \). If \( \xi(z) \in \Theta \), then it follows from [13] that

\[
\Pr\{X_n < z\} \leq \left[ \frac{f(z)}{g(z, \xi(z))} \right]^n \Pr_{\xi(z)}\{X_n < z\} \leq \left[ \frac{f(z)}{g(z, \xi(z))} \right]^n.
\]

### 4 A Link between LR Method and Large Deviation Theory

In this section, we will show that there exists a fundamental connection between the LR method and the theory of large deviations.

Let \( X_1, X_2, \ldots, X_n \) and \( X \) be random variables defined on the same sample space \( \Omega \) and \( \sigma \)-algebra \( \mathcal{F} \) such that the true probability measure is one of two measures \( \mathbb{P}_r \) and \( \mathbb{P}_g \). Here, the measure \( \mathbb{P}_r \) is determined by pdf or pmf \( f(.) \). The measure \( \mathbb{P}_g \) is determined by pdf or pmf \( g(.) \). Under each of these probability measures, the \( X \)'s are independent and identically distributed. Let \( \mathbb{E}[.] \) and \( \mathbb{E}_g[.] \) denote the mathematical expectations taken under probability measures \( \mathbb{P}_r \) and \( \mathbb{P}_g \), respectively. Let \( o(n) \) be a function of \( n \) such that \( \frac{o(n)}{n} \to 0 \) as \( n \to \infty \). In this setting, we have the following results.
Theorem 3 Let $\mu = \mathbb{E}[X]$ and $z = \mathbb{E}_0[X]$. Assume that $\mathbb{E}_0[|X - z|^2] < \infty$ and that $\mathbb{E}[e^{sX}] < \infty$ for $s$ in a neighborhood of 0. Let $\rho(\cdot)$ be a continuous function. The following assertions hold.

(I): If $z \geq \mu$, $\Pr\{X > z\} > 0$ and there exists $\delta > 0$ such that

$$\left[ \prod_{i=1}^{n} f(X_i) \right]_{\{z \leq X_n < z + \delta\}} \leq \exp (n\rho(X_n) + o(n)) \prod_{i=1}^{n} g(X_i)$$

for sufficiently large $n > 0$, then $\Pr\{X_n \geq z\} \leq \exp(n\rho(z))$ and $\lim_{n \to \infty} \frac{1}{n} \ln \Pr\{X_n \geq z\} = \rho(z)$.

(II): If $z \geq \mu$, $\Pr\{X > z\} > 0$ and there exists $\delta > 0$ such that and the following inequality

$$\left[ \prod_{i=1}^{n} g(X_i) \right]_{\{z \leq X_n < z + \delta\}} \leq \exp (-n\rho(X_n) + o(n)) \prod_{i=1}^{n} f(X_i)$$

hold for sufficiently large $n > 0$, then $\lim_{n \to \infty} \frac{1}{n} \ln \Pr\{X_n \geq z\} = \rho(z)$.

(III): If $z \leq \mu$, $\Pr\{X < z\} > 0$ and there exists $\delta > 0$ such that

$$\left[ \prod_{i=1}^{n} f(X_i) \right]_{\{z \leq X_n < z - \delta\}} \leq \exp (n\rho(X_n) + o(n)) \prod_{i=1}^{n} g(X_i)$$

for sufficiently large $n > 0$, then $\Pr\{X_n \leq z\} \leq \exp(n\rho(z))$ and $\lim_{n \to \infty} \frac{1}{n} \ln \Pr\{X_n \leq z\} = \rho(z)$.

(IV): If $z \leq \mu$, $\Pr\{X < z\} > 0$ and there exists $\delta > 0$ such that and the following inequality

$$\left[ \prod_{i=1}^{n} g(X_i) \right]_{\{z \geq X_n > z - \delta\}} \leq \exp (-n\rho(X_n) + o(n)) \prod_{i=1}^{n} f(X_i)$$

hold for sufficiently large $n > 0$, then, $\lim_{n \to \infty} \frac{1}{n} \ln \Pr\{X_n \leq z\} = \rho(z)$.

(V): If $\rho(\cdot)$ is a concave continuous function such that $f(X) \leq \exp(\rho(X)) g(X)$, then

$$\Pr\{X_n \geq z\} \leq \exp(n\rho(z)) \quad \text{provided that } z \geq \mu \text{ and } \Pr\{X > z\} > 0,$$

$$\Pr\{X_n \leq z\} \leq \exp(n\rho(z)) \quad \text{provided that } z \leq \mu \text{ and } \Pr\{X < z\} > 0.$$

(VI): If $\rho(\cdot)$ is a linear continuous function such that $f(X) = \exp(\rho(X)) g(X)$, then

$$\lim_{n \to \infty} \frac{1}{n} \ln \Pr\{X_n \geq z\} = \rho(z) \quad \text{provided that } z \geq \mu \text{ and } \Pr\{X > z\} > 0,$$

$$\lim_{n \to \infty} \frac{1}{n} \ln \Pr\{X_n \leq z\} = \rho(z) \quad \text{provided that } z \leq \mu \text{ and } \Pr\{X < z\} > 0.$$

See Appendix B for a proof. The assertions (V) and (VI) indicate that, under certain conditions, the probabilistic inequalities obtained from the LR method coincide with the Chernoff bounds derived from moment generating functions.

5 A Unified Theory for Exponential Families

Our main objective for this section is to develop a unified theory for bounding the tail probabilities of exponential families. A multivariate exponential family has pdf or pmf of the form

$$f(x, \theta) = v(x) \exp (\eta(\theta) \ast u(x) - \zeta(\theta)), \quad \theta \in \Theta,$$

where $x$ and $\theta$ are column vectors, $\eta(\theta)$ is column vector function of $\theta$, $u(x)$ is a column vector function of $x$, $\zeta(\theta)$ is a scalar function of $\theta$, and $v(x)$ is a scalar function of $x$. The symbol “$\ast$” denotes the inner
product. Assume that $\eta(\theta)$ and $\zeta(\theta)$ are differentiable functions of $\theta \in \Theta$. Let $X$ be a random vector having a pdf or pmf in the exponential family \[15\]. It can be shown that

$$\eta'(\theta) \mathbb{E}_\theta[u(X)] = \zeta'(\theta),$$

where $\eta'(\theta)$ is the derivative matrix such that the element in the $i$th row and $j$th column is equal to $\frac{\partial \eta(\theta)}{\partial \theta_i}$. The subscript in $\mathbb{E}_\theta$ indicates that the expectation is taken under the probability measure determined by a pdf or pmf with parameter $\theta$. Let $\Pr\{E \mid \theta\}$ denote the probability of event $E$, where the probability measure is determined by a pdf or pmf with parameter $\theta$. For two vectors $V_1$ and $V_2$, we write $V_1 \leq V_2$ if each element of $V_1$ is no greater than the corresponding element of $V_2$. Similarly, we write $V_1 \geq V_2$ if each element of $V_1$ is no less than the corresponding element of $V_2$. In this setting, we have the following results.

**Theorem 4** Let $X_1, X_2, \ldots, X_n$ be i.i.d. random vectors with common pdf or pmf parameterized by $\theta \in \Theta$ as \[15\]. Define $Z = \frac{1}{n} \sum_{i=1}^n u(X_i)$ and $\rho(z, \vartheta) = [\eta(\theta) - \eta(\vartheta)]*z - \zeta(\theta) + \zeta(\vartheta)$, where $\theta, \vartheta \in \Theta$ and $z$ is a vector of the same size as $Z$. Define $\mathcal{A}_\theta = \{\vartheta \in \Theta : \eta(\vartheta) \leq \eta(\theta)\}$ and $\mathcal{B}_\theta = \{\vartheta \in \Theta : \eta(\vartheta) \geq \eta(\theta)\}$. Then,

$$\Pr\{Z \leq z \mid \theta\} \leq \inf_{\vartheta \in \mathcal{A}_\theta} \exp(n\rho(z, \vartheta)) \Pr\{Z \leq z \mid \vartheta\} \leq \inf_{\vartheta \in \mathcal{A}_\theta} \exp(n\rho(z, \vartheta)), \quad (19)$$

$$\Pr\{Z \geq z \mid \theta\} \leq \inf_{\vartheta \in \mathcal{B}_\theta} \exp(n\rho(z, \vartheta)) \Pr\{Z \geq z \mid \vartheta\} \leq \inf_{\vartheta \in \mathcal{B}_\theta} \exp(n\rho(z, \vartheta)). \quad (20)$$

It should be noted that Theorem 4 involves no moment generating function in the derivation of probabilistic inequalities. To find explicit expressions for the bounds, the method of moments may be useful. Specifically, if a vector $\alpha \in \mathcal{A}_\theta$ can be obtained such that $\mathbb{E}_\alpha[u(X)] = z$ or equivalently, $\eta'(\alpha) z = \zeta'(\alpha)$, then it follows from \[10\] that

$$\Pr\{Z \leq z \mid \theta\} \leq \exp(n\rho(z, \alpha)) \Pr\{Z \leq z \mid \alpha\} \leq \exp(n\rho(z, \alpha)),$$

Similarly, if a vector $\alpha \in \mathcal{B}_\theta$ can be obtained such that $\eta'(\alpha) z = \zeta'(\alpha)$, then it follows from \[20\] that

$$\Pr\{Z \geq z \mid \theta\} \leq \exp(n\rho(z, \alpha)) \Pr\{Z \geq z \mid \alpha\} \leq \exp(n\rho(z, \alpha)).$$

To prove Theorem 4, it suffices to make use of Theorem 4 and the observation that

$$\prod_{i=1}^n f(X_i, \theta) \mathbb{I}_{\{Z \leq z\}} \leq \exp(n\rho(z, \vartheta)) \prod_{i=1}^n f(X_i, \vartheta) \quad \text{for } \vartheta \in \mathcal{A}_\theta;$$

$$\prod_{i=1}^n f(X_i, \theta) \mathbb{I}_{\{Z \geq z\}} \leq \exp(n\rho(z, \vartheta)) \prod_{i=1}^n f(X_i, \vartheta) \quad \text{for } \vartheta \in \mathcal{B}_\theta.$$

In the above discussion, we have developed a general method for deriving probabilistic inequalities for the multivariate case. In the remainder of this section, we will discuss the univariate case. A univariate exponential family is a set of probability distributions whose pdf or pmf can be expressed in the form

$$f(x, \theta) = v(x) \exp(\eta(\theta)u(x) - \zeta(\theta)), \quad \theta \in \Theta \quad (21)$$

where $u(x), v(x), \eta(\theta)$, and $\zeta(\theta)$ are known functions. Assume that $\eta(\theta)$ and $\zeta(\theta)$ are differentiable functions of $\theta \in \Theta$. Let $X$ be a random variable having a pdf or pmf in the exponential family \[21\]. Then,

$$\eta'(\theta) \mathbb{E}_\theta[u(X)] = \zeta'(\theta), \quad \theta \in \Theta.$$
Clearly, the expectation of \( u(X) \) is a function of \( \theta \). Let this function be denoted by \( \mu(\theta) \).

For the exponential family described above, we have the following results.

**Theorem 5** Let \( X \) be a random variable with pdf or pmf parameterized by \( \theta \in \Theta \) as \( f(X) \). Define \( Z = \frac{1}{n} \sum_{i=1}^{n} u(X_i) \), where \( X_1, \ldots, X_n \) are i.i.d. samples of \( X \). Let \( \alpha \in \Theta \) and \( z = \mu(\alpha) \). Define \( \rho(z, \theta) = |\eta(\theta) - \eta(\theta_0)|z - \zeta(\theta) + \zeta(\theta_\theta) \) for \( \theta, \theta_\theta \in \Theta \). Let \( c \) denote the absolute constant in the Berry-Essen inequality. Define

\[
\Delta = \min \left\{ \frac{1}{2}, \frac{c}{\sqrt{n}} \mathbb{E} ||u(X) - z||^2 \right\},
\]

where the expectation is taken under the probability measure determined by the pdf or pmf \( f(x, \alpha) \). The following assertions hold.

(I): If \( \mu(\theta) \leq z \) and \( \Pr\{u(X) < z \mid \theta\} > 0 \), then

\[
\Pr\{Z \geq z \mid \theta\} \leq \exp(n\rho(z, \alpha)), \quad \lim_{n \to \infty} \frac{1}{n} \Pr\{Z \geq z \mid \theta\} = \rho(z, \alpha).
\]

(II): If \( \mu(\theta) \geq z \) and \( \Pr\{u(X) > z \mid \theta\} > 0 \), then

\[
\Pr\{Z \leq z \mid \theta\} \leq \exp(n\rho(z, \alpha)), \quad \lim_{n \to \infty} \frac{1}{n} \Pr\{Z \leq z \mid \theta\} = \rho(z, \alpha).
\]

(III): If \( \eta(\theta) \leq \eta(\alpha) \), then \( \Pr\{Z \geq z \mid \theta\} \leq \exp(n\rho(z, \alpha)) \times \Pr\{Z \geq z \mid \alpha\} \leq (\frac{1}{\theta} + \Delta) \exp(n\rho(z, \alpha)). \)

(IV): If \( \eta(\theta) \geq \eta(\alpha) \), then \( \Pr\{Z \leq z \mid \theta\} \leq \exp(n\rho(z, \alpha)) \times \Pr\{Z \leq z \mid \alpha\} \leq (\frac{1}{\theta} + \Delta) \exp(n\rho(z, \alpha)). \)

The assertions (I) and (II) of Theorem 5 follow immediately from the assertions (V) and (VI) of Theorem 4. The assertions (III) and (IV) of Theorem 5 are direct consequences of Theorem 4 and the Berry-Essen inequality. The famous Berry-Essen inequality \[3, 10\] asserts the following:

Let \( Y_1, Y_2, \ldots \) be i.i.d. samples of random variable \( Y \) such that \( \mathbb{E}[Y] = 0 \), \( \mathbb{E}[Y^2] > 0 \), and \( \mathbb{E}||Y||^3 < \infty \). Also, let \( F_n \) be the cumulative distribution function (cdf) of \( \frac{\sum_{i=1}^{n} Y_i}{\sqrt{n} \mathbb{E}[Y]^2} \), and \( \Phi \) the cdf of the standard normal distribution. Then, there exists a positive constant \( c \) such that for all \( y \) and \( n \),

\[
|F_n(y) - \Phi(y)| \leq \frac{c}{\sqrt{n}} \frac{\mathbb{E}[||Y||^3]}{\mathbb{E}[Y^2]^2}.
\]

Recently, Tyurin \[16\] has shown that \( c < 0.4785 < \frac{1}{7} \).

### 6 Univariate Probabilistic Inequalities

In this section, we shall apply the LR method to derive inequalities for some important univariate random variables.

#### 6.1 Bernoulli Distribution

A Bernoulli random variable, \( X \), of mean value \( p \in (0, 1) \) has a pmf

\[
f(x, p) = \Pr\{X = x\} = p^x(1 - p)^{1-x}, \quad x \in \{0, 1\},
\]

which belongs to the exponential family \( f(x, p) = v(x) \exp(\eta(p)u(x) - \zeta(p)) \), \( x \in \{0, 1\}, \ p \in (0, 1) \), where

\[
u(x) = x, \quad v(x) = 1, \quad \eta(p) = \ln\frac{p}{1 - p}, \quad \zeta(p) = \ln\frac{1}{1 - p}.
\]

11
Note that
\[
E[X] = p, \quad E[|X - p|^2] = p(1 - p), \quad E[|X - p|^3] = |p^2 + (1 - p)^2|p(1 - p)
\]
and
\[
\frac{E[|X - p|^3]}{E^2[|X - p|^2]} = \frac{p^2 + (1 - p)^2}{\sqrt{p(1 - p)}}.
\]
Making use of these facts, the assertions (III) and (IV) of Theorem 5, we have the following results.

**Theorem 6** Let \(X_1, \ldots, X_n\) be i.i.d. samples of Bernoulli random variable \(X\) of mean value \(p \in (0, 1)\). Define \(\mathcal{M}(z, p) = \left(\frac{z}{n}\right)^z \left(\frac{1-z}{n-z}\right)^{1-z}\) for \(z \in (0, 1)\) and \(p \in (0, 1)\). Then,
\[
\Pr \{X_n \leq z\} \leq \left(\frac{1}{2} + \Delta\right) [\mathcal{M}(z, p)]^n \quad \text{for } z \in (0, p),
\]
\[
\Pr \{X_n \geq z\} \leq \left(\frac{1}{2} + \Delta\right) [\mathcal{M}(z, p)]^n \quad \text{for } z \in (p, 1),
\]
where \(\Delta = \min \left\{ \frac{1}{2}, \frac{G(z^2 + (1-z)^2)}{\sqrt{n}(z-1)} \right\}\) with \(G\) being the Berry-Esseen constant.

Clearly, \(\Delta \to 0\) as \(n \to \infty\). For moderate to large \(n\), the factor \(\frac{1}{2} + \Delta\) is close to \(\frac{1}{2}\). This implies that Theorem 6 has reduced the existing Chernoff bounds for the tail probabilities of a binomial distribution by a factor of \(\frac{1}{2}\). This improvement is due to the application of the LR method. Moreover, the LR method offers an extremely proof for the existing Chernoff bounds in the case of a binomial distribution. Specifically, making use of Theorem 1 and the observation that
\[
\left[ \prod_{i=1}^{n} f(X_i, p) \right] I(X_n \leq z) \leq [\mathcal{M}(z, p)]^n \left[ \prod_{i=1}^{n} f(X_i, z) \right] \quad \text{for } z \in (0, p),
\]
we have \(\Pr \{X_n \leq z\} \leq [\mathcal{M}(z, p)]^n\) for \(z \in (0, p)\). On the other hand, making use of Theorem 1 and the observation that
\[
\left[ \prod_{i=1}^{n} f(X_i, p) \right] I(X_n \geq z) \leq [\mathcal{M}(z, p)]^n \left[ \prod_{i=1}^{n} f(X_i, z) \right] \quad \text{for } z \in (p, 1),
\]
we have \(\Pr \{X_n \geq z\} \leq [\mathcal{M}(z, p)]^n\) for \(z \in (p, 1)\). In sharp contrast to the classical method, such arguments involve no moment generating function.

### 6.2 Hypergeometric Distribution

The hypergeometric distribution can be described by the following model. Consider a finite population of \(N\) units, of which there are \(R\) units having a certain attribute. Draw \(n\) units from the whole population by sampling without replacement. Let \(X\) denote the number of units having the attribute found in the \(n\) draws. Then, \(X\) is a random variable possessing a hypergeometric distribution such that
\[
\Pr\{X = x\} = \begin{cases} 
\frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} & \text{for } x \in \{0, 1, \cdots, n\} \text{ such that } x \leq R \text{ and } n-x \leq B, \\
0 & \text{for } x \in \{0, 1, \cdots, n\} \text{ such that } x > R \text{ or } n-x > B,
\end{cases}
\]
where \(B = N - R\). The mean of \(X\) is \(\mu = E[X] = \frac{nR}{N}\). Since the moment generating function of the hypergeometric distribution is intractable, it is difficult to derive tight inequalities for the tail distributions based on Chernoff bounds. However, such difficulties can be overcome by the LR method. In this direction, we have the following results.
Theorem 7 Let $X$ be a random variable possessing a hypergeometric distribution defined by [24]. Let $r$ and $b$ be nonnegative integers such that $r \leq R$, $b \leq B$ and $r + b = n$. Define $\hat{R} = \min\{N, \lfloor (N + 1)\frac{r}{n} \rfloor \}$ and $\hat{B} = N - \hat{R}$. Then,

$$\Pr\{X \leq r\} \leq \frac{\binom{R}{r} \binom{B}{b}}{\binom{R + B}{r + b}} \text{ for } r \leq \mu,$$

and

$$\Pr\{X \geq r\} \leq \frac{\binom{R}{r} \binom{B}{b}}{\binom{R + B}{r + b}} \text{ for } r \geq \mu.$$ 

(23)

(24)

See Appendix C for proof. Actually, in August 2010, Chen had derived inequalities (23) and (24) based on the LR method and applied them in [4, page 54, inequalities (35), (36)] and [5, page 20, inequalities (15), (16)] for developing multistage sampling schemes for statistical inference of the population proportion $p$.

6.3 Generalized Poisson Distribution

A random variable $X$ is said to possess a generalized Poisson distribution if

$$\Pr\{X = x\} = \frac{\lambda (\lambda + x \alpha)^{x-1} e^{-\lambda - x \alpha}}{x!}, \quad x = 0, 1, 2, \ldots,$$

where $\lambda > 0$ and $0 \leq \alpha < 1$. This distribution is sometimes called Consul’s generalized Poisson distribution (see, [9] and the references therein). It can be shown that the expectation of $\alpha \in [0, \lambda]$.

$$\Pr \{X \leq r\} \leq \frac{\binom{B}{r} \binom{B}{b}}{\binom{B + b}{r}}$$

and

$$\Pr \{X \geq r\} \leq \frac{\binom{B}{r} \binom{B}{b}}{\binom{B + b}{r}}$$

for $r \leq \mu$.

(25)

(26)

(27)

Theorem 8 Let $X_1, \ldots, X_n$ be i.i.d. random variables possessing the generalized Poisson distribution described by (25). Then,

$$\Pr\{X_n \leq z\} \leq \frac{\lambda}{(1 - \alpha)(\lambda + z \alpha)} \left[ \frac{\lambda}{z} + \alpha \right]^z \frac{e^{(1 - \alpha)z}}{e^{\lambda}}$$

provided that $0 < z \leq \frac{\lambda}{1 - \alpha}$,

$$\Pr\{X_n \geq z\} \leq \frac{\lambda}{(1 - \alpha)(\lambda + z \alpha)} \left[ \frac{\lambda}{z} + \alpha \right]^z \frac{e^{(1 - \alpha)z}}{e^{\lambda}}$$

provided that $z \geq \frac{\lambda}{1 - \alpha}$.

Moreover,

$$\Pr\{X_n \leq z\} \leq \frac{\lambda (\nu + z \alpha)}{\nu (\lambda + z \alpha)} \left[ \frac{\lambda}{\nu} + \alpha \right]^z \frac{e^{\nu \gamma}}{e^{\lambda}}$$

provided that $0 < z \leq \frac{\lambda}{1 - \alpha + \frac{\alpha}{\nu \lambda}}$,

$$\Pr\{X_n \geq z\} \leq \frac{\lambda (\nu + z \alpha)}{\nu (\lambda + z \alpha)} \left[ \frac{\lambda}{\nu} + \alpha \right]^z \frac{e^{\nu \gamma}}{e^{\lambda}}$$

provided that $z \geq \frac{\lambda}{1 - \alpha + \frac{\alpha}{\nu \lambda}}$.

(28)

(29)

where $\nu = \frac{1}{2}(1 - \alpha)z + \sqrt{(1 - \alpha)^2 z^2 + 4z \alpha / n}$. Furthermore, if $\alpha = 0$, then

$$\Pr\{X_n \geq z\} \leq \left( \frac{1}{2} + \Delta \right) \left( \frac{\lambda e^2}{z^2 e^{\lambda}} \right)^n$$

for $z \geq \lambda$,

$$\Pr\{X_n \leq z\} \leq \left( \frac{1}{2} + \Delta \right) \left( \frac{\lambda e^2}{z^2 e^{\lambda}} \right)^n$$

for $0 < z \leq \lambda$.

(30)

(31)

where $\Delta = \min \left\{ \frac{1}{2}, \frac{C}{\sqrt{n}} (3 + \frac{1}{2})^{\frac{3}{2}} \right\}$ with $C$ being the Berry-Essen constant.

See Appendix D for a proof.
6.4 Gamma Distribution

In probability theory and statistics, a random variable $X$ is said to have a gamma distribution if its pdf is of the form

$$f(x) = \frac{x^{k-1}}{\Gamma(k)\theta^k} \exp \left( -\frac{x}{\theta} \right) \quad \text{for} \quad 0 < x < \infty,$$

(32)

where $\theta > 0$, $k > 0$ are referred to as the scale parameter and shape parameter respectively. By letting

$$u(x) = \frac{x}{k}, \quad v(x) = x^{k-1}k^{k-1}, \quad \eta(\theta) = -\frac{k}{\theta}, \quad \zeta(\theta) = k \ln \theta,$$

we can write $f(x) = v(x)\exp \left( \eta(\theta)u(x) - \zeta(\theta) \right)$, which can be seen to be an exponential family. The moment generating function of $X$ is $\phi(s) = \mathbb{E}[e^{sx}] = (1 - \theta s)^{-k}$ for $s < \frac{1}{\theta}$. It can be shown by induction that

$$\frac{d^{\ell+1}\phi(s)}{ds^{\ell+1}} = (k + \ell)\theta \frac{d^{\ell}\phi(s)}{ds^{\ell}} \quad \text{for} \quad \ell = 0, 1, 2, \ldots.$$

Therefore,

$$\mathbb{E}[|X - k\theta|^2] = k\theta^2, \quad \mathbb{E}[|X - k\theta|^4] = \sum_{i=0}^{4} \binom{4}{i} (-k\theta)^i \mathbb{E}[X^{4-i}] = 3k(k + 2)\theta^4,$$

and by Lyapunov’s inequality,

$$\frac{\mathbb{E}[|X - k\theta|^3]}{\mathbb{E}[|X - k\theta|^2]} < \frac{\mathbb{E}[|X - k\theta|^4]}{\mathbb{E}[|X - k\theta|^2]} = \left( 3 + \frac{6}{k} \right)^{\frac{3}{2}}.$$

Making use of these facts, the assertions (III) and (IV) of Theorem 5, we have the following results.

**Theorem 9** Let $X_1, \ldots, X_n$ be i.i.d. samples of Gamma random variable $X$ possessing a probability density function (32). Then,

$$\Pr \{ X_n \leq \rho k\theta \} \leq \left( \frac{1}{2} + \Delta \right) \left[ \rho \exp \left( 1 - \rho \right) \right]^{kn} \quad \text{for} \quad 0 < \rho \leq 1,$$

$$\Pr \{ X_n \geq \rho k\theta \} \leq \left( \frac{1}{2} + \Delta \right) \left[ \rho \exp \left( 1 - \rho \right) \right]^{kn} \quad \text{for} \quad \rho \geq 1,$$

where $\Delta = \min \left\{ \frac{1}{2}, \frac{\mathcal{C}}{\sqrt{n}} \left( 3 + \frac{6}{k} \right)^{\frac{3}{2}} \right\}$ with $\mathcal{C}$ being the Berry-Essen constant.

7 Multivariate Probabilistic Inequalities

In this section, we shall apply the LR method to derive inequalities for multivariate random variables. We need to define two partial order relations $\prec$ and $\succ$ for vectors as follows.

Let $\kappa \in \mathbb{N}$. For vectors $x = [x_0, x_1, \ldots, x_\kappa]^T$ and $y = [y_0, y_1, \ldots, y_\kappa]^T$, we write $x \prec y$ if

$$x_i \leq y_i \quad \text{for} \quad i = 1, \ldots, \kappa.$$

Similarly, we write $x \succ y$ if

$$x_i \geq y_i \quad \text{for} \quad i = 1, \ldots, \kappa.$$

These partial order relations will be used in Sections 7.1, 7.2 and 7.3.
7.1 Multivariate Generalized Hypergeometric Distribution

In probability theory, random variables $X_0, X_1, \cdots, X_κ$ are said to possess a multivariate hypergeometric distribution if

$$\Pr\{X_i = x_i, i = 0, 1, \cdots, κ\} = \begin{cases} 0 & \text{if } x_i \geq 1 + C_i \text{ for some } i \in \{0, 1, \cdots, κ\}, \\ \frac{\Pi_{i=0}^κ (C_i)}{C_i^{x_i}} & \text{otherwise} \end{cases}$$

where $C_i \in \mathbb{N}, x_i \in \mathbb{Z}^+$ for $i = 0, 1, \cdots, κ$ and

$$N = \sum_{i=0}^κ C_i \geq \sum_{i=0}^κ x_i = n.$$ 

Actually, under mild restrictions, the multivariate hypergeometric distribution can be generalized by allowing $N$ and $C_i$ to be real numbers. More formally, random variables $X_0, X_1, \cdots, X_n$ are said to possess a multivariate generalized hypergeometric distribution if

$$\Pr\{X_i = x_i, i = 0, 1, \cdots, κ\} = \begin{cases} 0 & \text{if } x_i \geq 1 + C_i > 1 \text{ for some } i \in \{0, 1, \cdots, κ\}, \\ \frac{\Pi_{i=0}^κ (C_i)}{C_i^{x_i}} & \text{otherwise} \end{cases}$$

(33)

where $C_i \in \mathbb{R}, x_i \in \mathbb{Z}^+$ for $i = 0, 1, \cdots, κ$ and

$$N = \sum_{i=0}^k C_i \neq 0, \quad \sum_{i=0}^κ x_i = n, \quad \frac{n-1}{N} < 1, \quad \frac{C_i}{N} > 0 \quad \text{for } i = 0, 1, \cdots, k.$$ 

By an urn model approach, we have shown in [6] that (33) indeed defines a distribution.

The multivariate generalized hypergeometric distribution includes many important distributions as special cases. Clearly, the multivariate hypergeometric distribution is obtained from the multivariate generalized hypergeometric distribution by restricting $C_i, i = 0, 1, \cdots, k$ and $N$ as positive integers. The multivariate negative hypergeometric distribution is obtained from the multivariate generalized hypergeometric distribution by taking $C_i, i = 0, 1, \cdots, k$ and $N$ as negative integers. The multinomial distribution is obtained from the multivariate generalized hypergeometric distribution by letting $N \rightarrow \infty$ under the constraint that $\frac{C_i}{N}, i = 0, 1, \cdots, k$ converge to positive numbers sum to 1. Moreover, the multivariate Pólya-Eggenberger distribution [15] can be accommodated as a special case of the multivariate generalized hypergeometric distribution.

Let $X_0, X_1, \cdots, X_κ$ be random variables possessing a multivariate generalized hypergeometric distribution defined by (33). It can be shown that the means of these variables are given as

$$\mu_i = \mathbb{E}[X_i] = \frac{nC_i}{N}, \quad i = 0, 1, \cdots, κ.$$ 

Define vectors

$$\mathbf{x} = [x_0, x_1, \cdots, x_κ]^T, \quad \mathbf{X} = [X_0, X_1, \cdots, X_κ]^T, \quad \mathbf{μ} = [μ_0, μ_1, \cdots, μ_κ]^T.$$ 

Clearly, $\mathbb{E}[\mathbf{X}] = \mathbf{μ}$. We use $f(\mathbf{x})$ to denote the probability mass function defined by (33), that is,

$$f(\mathbf{x}) = \Pr\{X_i = x_i, i = 0, 1, \cdots, κ\}.$$ 

In this setting, we have established the following results.
Theorem 10  Let \( z_i, i = 0, 1, \ldots, k \) be nonnegative integers such that \( \sum_{i=0}^{k} z_i = n \). Define \( z = [z_0, z_1, \ldots, z_k]^\top \) and \( \hat{C}_i = N \frac{z_i}{n} \) for \( i = 0, 1, \ldots, k \). Assume that \( f(z) > 0 \). The following assertions hold.

(I):

\[
\Pr\{X < z\} \leq \prod_{i=0}^{k} \frac{C_i}{\binom{N}{z_i}} \quad \text{provided that } z < \mu. \tag{34}
\]

(II):

\[
\Pr\{X > z\} \leq \prod_{i=0}^{k} \frac{C_i}{\binom{N}{z_i}} \quad \text{provided that } z > \mu. \tag{35}
\]

(III):

\[
\prod_{i=0}^{k} \frac{C_i}{\binom{N}{z_i}} < \left[ \frac{(n)}{\binom{N}{x}} \prod_{i=0}^{k} \frac{C_i}{\binom{N}{z_i}} \right] \left( \frac{e^2}{2\pi} \right)^{k+1} \left( \frac{N-n}{N} \right)^{N-n+k+1} \prod_{i=0}^{k} \frac{C_i}{C_i - z_i} \quad \text{provided that } C_i - z_i \geq 1 \text{ and } \hat{C}_i - z_i \geq 1 \text{ for } i = 0, 1, \ldots, k. \tag{36}
\]

(IV):

\[
\left( \frac{N-n}{N} \right)^{N-n+k+1} \prod_{i=0}^{k} \frac{C_i}{C_i - z_i} \xrightarrow{N \to \infty} 0
\]

as \( N \to \infty \) under the constraint that \( \frac{C_i}{N} \to \alpha \) and \( \frac{z_i}{n} \to \theta_i, \frac{\hat{C}_i}{n} \to \beta_i \) for \( i = 0, 1, \ldots, k \), where \( \theta_i, \beta_i \) are positive real numbers less than 1 such that \( \theta_i > \alpha \beta_i \) for \( i = 0, 1, \ldots, k \) and that \( \theta_i \neq \beta_i \) for some \( i \in \{0, 1, \ldots, k\} \).

See Appendix E for a proof.

As \( N \to \infty \) under the constraint that \( \frac{C_i}{N} \to p_i, i = 0, 1, \ldots, k \),

\[
\prod_{i=0}^{k} \frac{C_i}{\binom{N}{z_i}} = \binom{n}{x} \prod_{\ell=0}^{k} \frac{C_i - \ell + 1}{\binom{N}{\ell} - 1} \to \binom{n}{x} \prod_{i=0}^{k} p_i^{x_i},
\]

where

\[
\binom{n}{x} = \frac{n!}{\prod_{i=0}^{x_i} i!}
\]

is the multinomial coefficient. This implies that the multivariate generalized hypergeometric distribution converges to the multinomial distribution

\[
\Pr\{X = x\} = \binom{n}{x} \prod_{i=0}^{k} p_i^{x_i}. \tag{37}
\]

Accordingly,

\[
\prod_{i=0}^{k} \frac{C_i}{\binom{N}{z_i}} = \prod_{i=0}^{k} \frac{z_i}{\hat{C}_i - \ell + 1} = \prod_{i=0}^{k} \frac{n \hat{C}_i - n(\ell-1)}{\hat{C}_i - \ell + 1} = \prod_{i=0}^{k} \frac{\mu_i - n(\ell-1)}{z_i - n(\ell-1)} \to \prod_{i=0}^{k} \left( \frac{\mu_i}{z_i} \right),
\]

where

\[
\mu_i = \mathbb{E}[X_i] = np_i
\]

with \( p_i, i = 0, 1, \ldots, k \) being positive real numbers sum to 1. As before, let \( X = [X_0, X_1, \ldots, X_k]^\top \) and \( \mu = [\mu_0, \mu_1, \cdots, \mu_k]^\top \). As applications of (34) and (35) of Theorem 10, we have the following results.
Corollary 1 Let \( z_i, i = 0, 1, \cdots, k \) be nonnegative integers such that \( \sum_{i=0}^{k} z_i = n \). Let \( z = [z_0, z_1, \cdots, z_k]^\top \).

Then,

\[
\Pr\{X \prec z\} \leq k \prod_{i=0}^{k} \left( \frac{\mu_i}{z_i} \right)^{z_i} \quad \text{provided that } z \prec \mu, \tag{38}
\]

and

\[
\Pr\{X \succ z\} \leq k \prod_{i=0}^{k} \left( \frac{\mu_i}{z_i} \right)^{z_i} \quad \text{provided that } z \succ \mu. \tag{39}
\]

It should be noted that Corollary 1 is a multivariate generalization of the Chernoff bounds for binomial distributions.

### 7.2 Multivariate Generalized Inverse Hypergeometric Distribution

Let \( \kappa \) be a positive integer. In probability theory, random variables \( X_1, \cdots, X_\kappa \) are said to possess a multivariate inverse hypergeometric distribution if

\[
\Pr\{X_i = x_i, i = 1, \cdots, \kappa\} = \begin{cases} 
0 & \text{if } n + 1 + N \geq 1 + \sum_{i=1}^{\kappa} C_i \text{ or } x_i \geq 1 + C_i \text{ for some } i \in \{1, \cdots, \kappa\}, \\
\frac{1}{n \prod_{i=0}^{k} \binom{C_i}{x_i}} & \text{otherwise},
\end{cases}
\]

where \( C_i \in \mathbb{N} \), \( x_i \in \mathbb{Z}_+ \) for \( i = 1, \cdots, k \) and

\[
C_0 \in \mathbb{N}, \quad \gamma \in \mathbb{N}, \quad x_0 = \gamma \leq C_0, \quad n = \sum_{i=0}^{k} x_i, \quad N = \sum_{i=0}^{k} C_i.
\]

Actually, under mild restrictions, the multivariate inverse hypergeometric distribution can be generalized by allowing \( N \) and \( C_i \) to be real numbers. More formally, random variables \( X_1, \cdots, X_\kappa \) are said to possess a multivariate generalized inverse hypergeometric distribution if

\[
\Pr\{X_i = x_i, i = 1, \cdots, \kappa\} = \begin{cases} 
0 & \text{if } n + 1 + N \geq 1 + \sum_{i=1}^{\kappa} C_i \text{ or } x_i \geq 1 + C_i > 1 \text{ for some } i \in \{1, \cdots, \kappa\}, \\
\frac{1}{n \prod_{i=0}^{k} \binom{C_i}{x_i}} & \text{otherwise},
\end{cases}
\]

where \( C_i \in \mathbb{R} \), \( x_i \in \mathbb{Z}_+ \) for \( i = 1, \cdots, k \),

\[
C_0 \in \mathbb{R}, \quad \gamma \in \mathbb{N}, \quad x_0 = \gamma, \quad n = \sum_{i=0}^{k} x_i, \quad N = \sum_{i=0}^{k} C_i \neq 0, \quad \frac{C_0}{N} > \frac{\gamma - 1}{N},
\]

and \( \frac{C_0}{N} > 0 \) for \( i = 1, \cdots, k \). By an urn model approach, we have shown in [6] that (40) indeed defines a distribution.

Let \( X_1, \cdots, X_\kappa \) be random variables possessing a multivariate generalized inverse hypergeometric distribution defined by (40). Let \( X_0 = \gamma \). It can be shown that the means of these variables are given as

\[
\mu_i = \mathbb{E}[X_i] = \frac{\gamma C_i}{C_0}, \quad i = 0, 1, \cdots, \kappa.
\]

Define vector

\[
X = [X_0, X_1, \cdots, X_\kappa]^\top.
\]

Let the probability mass function defined by (40) be denoted by \( f(x) \), that is,

\[
f(x) = \Pr\{X_i = x_i, i = 0, 1, \cdots, \kappa\},
\]

where \( x = [x_0, x_1, \cdots, x_\kappa]^\top \). In this setting, we have established the following results.
Theorem 11  Let $z_0 = \gamma$ and $z_i$, $i = 1, \cdots, k$ be nonnegative integers. Define $n = \sum_{i=0}^{k} z_i$ and $z = [z_0, z_1, \cdots, z_k]^T$. Define $\bar{\mu}_i = \frac{n z_i}{N}$ and $\bar{C}_i = N \frac{z_i}{n}$ for $i = 0, 1, \cdots, k$. Define $\bar{\mu} = [\bar{\mu}_0, \bar{\mu}_1, \cdots, \bar{\mu}_k]^T$. Assume that $\frac{n}{n-1} < 1$ and $f(z) > 0$. Then, the following assertions hold.

(I):

$$\Pr\{X < z\} \leq \prod_{i=0}^{k} \left( C_i \begin{pmatrix} C_i \cr z_i \end{pmatrix} \right) \quad \text{provided that } z < \bar{\mu}. \tag{41}$$

(II):

$$\Pr\{X > z\} \leq \prod_{i=0}^{k} \left( \frac{C_i}{z_i} \right) \quad \text{provided that } z > \bar{\mu}. \tag{42}$$

(III):

$$\prod_{i=0}^{k} \frac{C_i}{z_i} < \left( \frac{n}{N} \right)^n \prod_{i=0}^{k} \left( \frac{C_i}{x_i} \right)^{z_i} \left( \frac{e^2}{2\pi} \right)^{k+1} \left( \frac{N-n}{N} \right)^{N-n+k+1} \prod_{i=0}^{k} \left( \frac{C_i}{z_i} \right)^{X_i} \quad \text{provided that } C_i - z_i \geq 1 \text{ and } \hat{C}_i - z_i \geq 1 \text{ for } i = 0, 1, \cdots, k.$$

(IV):

$$\left( \frac{N-n}{N} \right)^{N-n+k+1} \prod_{i=0}^{k} \left( \frac{C_i}{z_i} \right)^{X_i} \rightarrow 0 \quad \text{as } N \rightarrow \infty \text{ under the constraint that } \frac{n}{N} \rightarrow \alpha \text{ and } \frac{C_i}{N} \rightarrow \theta_i, \frac{z_i}{n} \rightarrow \beta_i \text{ for } i = 0, 1, \cdots, k, \text{ where } \theta_i, \beta_i \text{ and } \alpha \text{ are positive numbers less than 1 such that } \theta_i > \alpha \beta_i \text{ for } i = 0, 1, \cdots, k \text{ and that } \theta_i \neq \beta_i \text{ for some } i \in \{0, 1, \cdots, k\}.$$

See Appendix F for a proof.

As $N \rightarrow \infty$ under the constraint that $\frac{C_i}{N} \rightarrow p_i$ for $i = 0, 1, \cdots, k$,

$$\frac{\gamma n}{n} \prod_{i=0}^{k} \left( \frac{C_i}{N} \right) = \gamma \left( \frac{n}{n} \right) \left( \prod_{i=0}^{k} \prod_{\ell=1}^{x_i} \frac{C_i - \ell + 1}{N - \ell + 1} \right) \rightarrow \gamma \left( \frac{n}{x} \right) \prod_{i=0}^{k} p_i^{x_i}.$$

This implies that the multivariate generalized inverse hypergeometric distribution converges to the negative multinomial distribution

$$\Pr\{X_i = x_i, \ i = 1, \cdots, k\} = \frac{\gamma \left( \frac{n}{x} \right)}{n} \prod_{i=0}^{k} p_i^{x_i}. \tag{43}$$

Accordingly,

$$\prod_{i=0}^{k} \left( \frac{C_i}{z_i} \right) = \prod_{i=0}^{k} \prod_{\ell=1}^{z_i} \frac{C_i - \ell + 1}{z_i - \ell + 1} = \prod_{i=0}^{k} \prod_{\ell=1}^{z_i} \frac{n C_i - n(\ell - 1)}{n(\ell - 1) - \ell} = \prod_{i=0}^{k} \prod_{\ell=1}^{z_i} \frac{n p_i - n(\ell - 1)}{n(\ell - 1) - \ell} \rightarrow \prod_{i=0}^{k} \left( \frac{n p_i}{z_i} \right)^{z_i}.$$

Define

$$X = [X_0, X_1, \cdots, X_k]^T$$

with $X_1, \cdots, X_k$ possessing a negative multinomial distribution defined by (43). As applications of (41) and (42) of Theorem 11 we have the following results.
Corollary 2 Let \( z_0 = \gamma \) and \( z_i, i = 1, \ldots, k \) be nonnegative integers. Define \( n = \sum_{i=0}^{k} z_i \) and \( z = [z_0, z_1, \ldots, z_k]^\top \). Define \( \mu = [\mu_0, \mu_1, \ldots, \mu_k]^\top \), where \( \mu_i = np_i, i = 0, 1, \ldots, \kappa \). Then,

\[
\Pr\{X < z\} \leq \prod_{i=0}^{k} \left( \frac{\hat{\mu}_i}{z_i} \right)^{z_i} \quad \text{provided that } z < \hat{\mu},
\]

and

\[
\Pr\{X \succeq z\} \leq \prod_{i=0}^{k} \left( \frac{\hat{\mu}_i}{z_i} \right)^{z_i} \quad \text{provided that } z \succeq \hat{\mu}.
\]

7.3 Dirichlet Distribution

Let \( \kappa \) be a positive integer. Random variables \( X_0, X_1, \ldots, X_\kappa \) are said to possess a Dirichlet distribution if they have a probability density function

\[
f(x, \alpha) = \begin{cases} \frac{1}{B(\alpha)} \prod_{i=0}^{\kappa} x_i^{\alpha_i-1} & \text{for } x_i \in \mathbb{R}^+, i = 0, 1, \ldots, \kappa \text{ such that } \sum_{i=0}^{\kappa} x_i = 1, \\ 0 & \text{else} \end{cases}
\]

where \( x = [x_0, x_1, \ldots, x_\kappa]^\top \), \( \alpha_0, \alpha_1, \ldots, \alpha_\kappa \) are positive real numbers, \( \alpha = [\alpha_0, \alpha_1, \ldots, \alpha_\kappa]^\top \), and

\[
B(\alpha) = \prod_{i=0}^{\kappa} \Gamma(\alpha_i) / \Gamma(\sum_{i=0}^{\kappa} \alpha_i).
\]

The means of \( X_0, X_1, \ldots, X_\kappa \) are

\[
\mu_i = \mathbb{E}[X_i] = \frac{\alpha_i}{\sum_{\ell=0}^{\kappa} \alpha_\ell} \quad \text{for } i = 0, 1, \ldots, \kappa.
\]

Define vectors

\[
X = [X_0, X_1, \ldots, X_\kappa]^\top, \quad \mu = [\mu_0, \mu_1, \ldots, \mu_\kappa]^\top.
\]

Let \( \mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n \) be independent random vectors possessing the same distribution as \( X \). Define \( \overline{\mathcal{X}}_n = \sum_{i=1}^{n} \mathcal{X}_i/n \). We have the following result.

Theorem 12 Let \( z_\ell, \ell = 0, 1, \ldots, \kappa \) be positive real numbers such that \( \sum_{\ell=0}^{k} z_\ell = 1 \). Define \( z = [z_0, z_1, \ldots, z_\kappa]^\top \). Then,

\[
\Pr\{\overline{\mathcal{X}}_n < z\} \leq \left[ \frac{B(\hat{\alpha})}{B(\alpha)} \prod_{i=1}^{\kappa} \left( \frac{z_i}{\hat{\alpha}_i} \right)^{z_i} \right]^n \quad \text{provided that } z \prec \mu,
\]

where \( \hat{\alpha} = [\hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_\kappa]^\top \) with \( \hat{\alpha}_i = \frac{\alpha_i z_i}{z_0} \) for \( i = 0, 1, \ldots, \kappa \).

See Appendix [C] for a proof.

7.4 Matrix Gamma Distribution

In statistics, a matrix gamma distribution is a generalization of the gamma distribution to positive-definite matrices (see, [11] and the references therein). It is a more general version of the Wishart distribution. The pdf of a matrix gamma distribution is

\[
f(x) = \frac{|\Sigma|^{-\alpha}}{\beta^{\alpha n} \Gamma_p(\alpha)} |x|^{n-(p+1)/2} \exp \left( -\frac{1}{\beta} \text{tr}(\Sigma^{-1} x) \right), \quad \alpha > 0, \beta > 0
\]

(47)
where \( \mathbf{x} \) and \( \Sigma \) are positive definite matrices of size \( p \times p \). Here, the notation \( "|.|" \) denotes the determinant function. Let \( \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n \) be i.i.d. positive definite random matrices possessing the same pdf defined by (47). Define
\[
\mathbf{X}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i, \quad \mu = \alpha \beta \Sigma.
\]
Clearly, \( E[\mathbf{X}_n] = \mu \). We write \( \mathbf{x} \prec \mathbf{y} \) if \( \mathbf{y} - \mathbf{x} \) is positive definite. Similarly, we write \( \mathbf{x} \succ \mathbf{y} \) if \( \mathbf{x} - \mathbf{y} \) is positive definite. In this setting, we have the following results.

**Theorem 13**
\[
\Pr\{\mathbf{X}_n \succ z\} \leq \left[ \left( \frac{e^{\alpha \beta}}{\alpha \beta} \right)^{\frac{p}{\alpha}} \left( \frac{|z|}{|\Sigma|}\right)^{\frac{1}{\alpha}} \exp \left( -\frac{1}{\beta} \text{tr}(\Sigma^{-1} z) \right) \right]^n \quad \text{for } z \succ \mu,
\]
\[
\Pr\{\mathbf{X}_n \prec z\} \leq \left[ \left( \frac{e^{\alpha \beta}}{\alpha \beta} \right)^{\frac{p}{\alpha}} \left( \frac{|z|}{|\Sigma|}\right)^{\frac{1}{\alpha}} \exp \left( -\frac{1}{\beta} \text{tr}(\Sigma^{-1} z) \right) \right]^n \quad \text{for } z \prec \mu.
\]

In particular, \( \Pr\{\mathbf{X}_n \succ q\mu\} \leq \left[ q \exp(1 - q) \right]^n \) for \( q \geq 1 \) and \( \Pr\{\mathbf{X}_n \prec q\mu\} \leq \left[ q \exp(1 - q) \right]^n \) for \( 0 \leq q \leq 1 \).

See Appendix H for a proof.

## 8 Conclusion

We have developed the LR method for deriving probabilistic inequalities. We have established fundamental connections among the LR method, the theory of large deviations and statistical concepts such as likelihood ratio, maximum likelihood, and the method of moments for parameter estimation. The LR method overcomes the limitations of the classical approach based on mathematical expectation. The LR method may provide easy derivation of concentration inequalities in situations that moment generating functions are not readily tractable. We have applied the LR method to obtain a wide spectrum of new concentration inequalities.

### A Proof of Theorem 2

By the assumption that \( f(\mathbf{X}) \leq \exp(\rho(\mathbf{X}, \vartheta)) g(\mathbf{X}, \vartheta) \) for all \( \vartheta \in \Theta \), we have
\[
\prod_{i=1}^{n} f(\mathbf{X}_i) \leq \exp \left( \sum_{i=1}^{n} \rho(\mathbf{X}_i, \vartheta) \right) \prod_{i=1}^{n} g(\mathbf{X}_i, \vartheta) \quad \text{for all } \vartheta \in \Theta.
\]
(48)

By the assumption that \( \rho(\mathbf{x}, \vartheta) \) is a concave function of \( \mathbf{x} \), we have
\[
\exp \left( \sum_{i=1}^{n} \rho(\mathbf{X}_i, \vartheta) \right) \leq \exp \left( n \rho(\mathbf{X}_n, \vartheta) \right) \quad \text{for all } \vartheta \in \Theta.
\]
(49)

Combining (48) and (49) yields
\[
\prod_{i=1}^{n} f(\mathbf{X}_i) \leq \exp(n \rho(\mathbf{X}_n, \vartheta)) \prod_{i=1}^{n} g(\mathbf{X}_i, \vartheta)
\]
(50)
for all $\vartheta \in \Theta$. Multiplying (50) by the indicator function $I_{(X_n \prec z)}$ yields

$$\prod_{i=1}^{n} f(X_i) I_{(X_n \prec z)} \leq \exp(n\rho(X_n, \vartheta)) \prod_{i=1}^{n} g(X_i, \vartheta) I_{(X_n \prec z)} \quad \text{for all } \vartheta \in \Theta. \quad (51)$$

Since $z$ is a matrix such that $\rho(x, \vartheta) \leq \rho(z, \vartheta)$ for all $\vartheta \in \Theta$ provided that $x \prec z$, it follows from (51) that

$$\prod_{i=1}^{n} f(X_i) I_{(X_n \prec z)} \leq \exp(n\rho(z, \vartheta)) \prod_{i=1}^{n} g(X_i, \vartheta) I_{(X_n \prec z)} \quad \text{for all } \vartheta \in \Theta,$$

which implies that

$$\prod_{i=1}^{n} f(X_i) I_{(X_n \prec z)} \leq \exp(n\rho(z, \vartheta)) \prod_{i=1}^{n} g(X_i, \vartheta) \quad \text{for all } \vartheta \in \Theta. \quad (52)$$

Applying (52) and Theorem 4 yields (53). This completes the proof of the theorem.

**B Proof of Theorem 3**

We need some preliminary results. The results in the following lemma are due to Petrov [14].

**Lemma 1** Let $X$ be a non-degenerate random variable with mean $\mu = \mathbb{E}[X]$ and moment generating function $\phi(s) = \mathbb{E}[e^{sX}]$ for $s \in (-a, b)$, where $a$ and $b$ are $\infty$ or positive real numbers. Let $\psi(s) = \ln \phi(s)$. Let $X_1, X_2, \cdots$ be i.i.d. samples of $X$. Let $X_n = \frac{\sum_{i=1}^{n} X_i}{n}$ for $n \in \mathbb{N}$. Let $\alpha > \mu$ be a real number such that there exists $\tau_\alpha \in (0, b)$ satisfying $\psi'(\tau_\alpha) = \alpha$. Let $\gamma(\alpha) = \alpha \tau_\alpha - \psi(\tau_\alpha)$. The following assertions hold.

(I) If $X$ is non-lattice valued, then

$$\Pr\{X_n \geq \alpha\} = \frac{\exp(-n\gamma(\alpha))}{\tau_\alpha \sqrt{2\pi n \psi''(\tau_\alpha)}}[1 + o(1)].$$

(II) If $X$ is a lattice-valued random variable with span $\nu$, then

$$\Pr\{X_n \geq \alpha\} = \frac{\nu}{\sqrt{2\pi n \psi''(\tau_\alpha)}} \frac{\exp(-n\gamma(\alpha))}{1 - \exp(-\nu \tau_\alpha)}[1 + o(1)].$$

**Lemma 2** For $z \geq \mu$ and $\Pr\{X > z\} > 0$,

$$\lim_{n \to \infty} \frac{1}{n} \ln \frac{\Pr\{z \leq X_n < z + \varepsilon\}}{\Pr\{X_n \geq z\}} = 0 \quad (53)$$

provided that $\varepsilon > 0$ is sufficiently small.

**Proof.** To show the lemma, we need to show that $\Pr\{X_n \geq z\} \geq \Pr\{z \leq X_n < z + \varepsilon\} > 0$ for large enough $n$ and that (53) holds for small enough $\varepsilon > 0$. Since

$$\ln \frac{\Pr\{z \leq X_n < z + \varepsilon\}}{\Pr\{X_n \geq z\}} = \ln \left(1 - \frac{\Pr\{X_n \geq z + \varepsilon\}}{\Pr\{X_n \geq z\}}\right),$$

to show (53), it suffices to show that

$$\lim_{n \to \infty} \frac{\Pr\{X_n \geq z + \varepsilon\}}{\Pr\{X_n \geq z\}} = 0 \quad (54)$$
holds for small enough \( \varepsilon > 0 \).

If \( z = \mu \), then

\[
\frac{\Pr\{X_n \geq z + \varepsilon\}}{\Pr\{X_n \geq z\}} = \frac{\Pr\{X_n \geq \mu + \varepsilon\}}{\Pr\{X_n \geq \mu\}} \leq \frac{\Pr\{X_n - \mu \geq \varepsilon\}}{\Pr\{\frac{\sqrt{n}(X_n - \mu)}{\sigma} \geq 0\}},
\]

(55)

where \( \sigma > 0 \) is the standard deviation of \( X \). Making use of (55) and the observation that \( \lim_{n \to \infty} \Pr\{|X_n - \mu| \geq \varepsilon\} = 0 \) due to the weak law of large numbers and that \( \lim_{n \to \infty} \Pr\{\frac{\sqrt{n}(X_n - \mu)}{\sigma} \geq 0\} = \frac{1}{2} \) as a result of the central limit theorem, we have that \( \Pr\{X_n \geq z\} \geq \Pr\{z \leq X_n < z + \varepsilon\} > 0 \) for large enough \( n \) and that (55) holds. In the sequel, we will show (54) by restricting \( \varepsilon \) to be greater than \( \mu \).

Define \( \varphi(s) = \mathbb{E}[e^{s(X-z)}] \) and

\[
b = \sup\{c \geq 0 : \varphi(s) < \infty \text{ for any } s \in [0, c]\}.
\]

(56)

Invoking the assumption that \( \varphi(s) = \mathbb{E}[e^{sX}] < \infty \) for \( s \) in a neighborhood of 0, we have that either \( b = \infty \) or \( b \) is a positive number. Hence, \( \varphi(s) < \infty \) and \( \varphi(s) < \infty \) for \( 0 < s < b \). We claim that \( \lim_{s \uparrow b} \varphi(s) = \infty \).

To show the claim, note that, as a consequence of the assumption that \( \Pr\{X > z\} > 0 \), there exists a number \( y \) greater than \( z \) such that \( \Pr\{X > y\} > 0 \). This implies that \( \varphi(s) \geq e^{s(y-z)} \Pr\{X > y\} \) for \( s > 0 \) such that \( \phi(s) < \infty \). Therefore, it must be true that either \( \lim_{s \to \infty} \varphi(s) = \infty \) or there exists a positive number \( \gamma > 0 \) such that \( \lim_{s \uparrow \gamma} \varphi(s) = \infty \). This fact together with the definition (56) for \( b \) imply that the claim is true.

Note that \( \varphi(s) = \psi(s) - sz \), where \( \psi(s) = \ln \phi(s) \). Since \( \varphi(s) \) is a convex function of \( s \in (0, b) \) and

\[
\varphi(0) = 0, \quad \varphi'(0) = \psi'(0) - z = \mu - z < 0, \quad \lim_{s \uparrow b} \varphi(s) = \infty,
\]

there must exist a unique number \( \tau_z \in (0, b) \) such that \( \varphi'(\tau_z) = \psi'(\tau_z) - z = 0 \). Let \( \zeta = \frac{\tau_z + b}{2} \). Then, \( \psi'(\zeta) > z \), since \( \zeta > \tau_z \) and \( \psi(.) \) is convex. Let \( \alpha \) be a number such that \( z \leq \alpha < \psi'(\zeta) \). Since \( \mu < z = \psi'(\tau_z) \leq \alpha < \psi'(\zeta) \), there must exist a unique number \( \tau_\alpha \in (0, b) \) such that \( \psi'(\tau_\alpha) = \alpha \). Clearly, \( \tau_\alpha \) is a function of \( \alpha \). Define \( \gamma(\alpha) = \alpha \tau_\alpha - \psi(\tau_\alpha) \) for \( \alpha \) such that \( z \leq \alpha < \psi'(\zeta) \). The derivative of \( \gamma(\alpha) \) with respect to \( \alpha \) is

\[
\gamma'(\alpha) = \frac{d}{d\alpha} [\alpha \tau_\alpha - \psi(\tau_\alpha)] = \tau_\alpha + \alpha \frac{d \tau_\alpha}{d\alpha} - \psi'(\tau_\alpha) \frac{d \tau_\alpha}{d\alpha} = \tau_\alpha + \alpha \frac{d \tau_\alpha}{d\alpha} - \alpha \frac{d \tau_\alpha}{d\alpha} = \tau_\alpha > 0,
\]

which implies that \( \gamma(\alpha) \) is strictly increasing with respect to \( \alpha \) such that \( z \leq \alpha < \psi'(\zeta) \). From now on, we restrict \( \varepsilon \) to be positive and less than \( \psi'(\zeta) - z \). Since \( 0 < \varepsilon < \psi'(\zeta) - z \), we have that \( \tau_{z+\varepsilon} \) exists and that

\[
0 < \tau_z < \tau_{z+\varepsilon} < b, \quad 0 < \varphi(z) < \varphi(z+\varepsilon).
\]

(57)

If \( X \) is non-lattice valued, then according to assertion (I) of Lemma \( \text{II} \) we have

\[
\Pr\{X_n \geq z\} = \frac{\exp(-n\gamma(z))}{\tau_z \sqrt{2\pi n \psi''(\tau_z)}} [1 + o(1)],
\]

(58)

\[
\Pr\{X_n \geq z + \varepsilon\} = \frac{\exp(-n\gamma(z + \varepsilon))}{\tau_{z+\varepsilon} \sqrt{2\pi n \psi''(\tau_{z+\varepsilon})}} [1 + o(1)].
\]

(59)

It follows from (57), (58) and (59) that \( \Pr\{X_n \geq z\} \geq \Pr\{z \leq X_n < z + \varepsilon\} > 0 \) for large enough \( n \) and that

\[
\frac{\Pr\{X_n \geq z + \varepsilon\}}{\Pr\{X_n \geq z\}} = \frac{\tau_z}{\tau_{z+\varepsilon}} \sqrt{\frac{\psi''(\tau_z)}{\psi''(\tau_{z+\varepsilon})}} \left[\frac{\exp(\gamma(z))}{\exp(\gamma(z + \varepsilon))}\right]^n [1 + o(1)] \to 0
\]

22
as $n \to \infty$. This shows \(54\) in the case that $X$ is non-lattice valued.

If $X$ is a lattice-valued random variable with span $\nu$, then according to assertion (II) of Lemma 1, we have

$$
\Pr\{X_n \geq z\} = \frac{\nu}{1 - \exp(-\nu \tau_z)} \frac{\exp(-n \gamma(z))}{\sqrt{2\pi n \nu'}} [1 + o(1)], \quad (60)
$$

$$
\Pr\{X_n \geq z + \varepsilon\} = \frac{\nu}{1 - \exp(-\nu \tau_{z+\varepsilon})} \frac{\exp(-n \gamma(z + \varepsilon))}{\sqrt{2\pi n \nu'}} [1 + o(1)]. \quad (61)
$$

It follows from (57), (60) and (61) that $\Pr\{X_n \geq z\} \geq \Pr\{z \leq X_n < z + \varepsilon\} > 0$ for large enough $n$ and that

$$
\frac{\Pr\{X_n \geq z + \varepsilon\}}{\Pr\{X_n \geq z\}} = \frac{1 - \exp(-\nu \tau_z)}{1 - \exp(-\nu \tau_{z+\varepsilon})} \frac{\psi'(\tau_z)}{\psi'(\tau_{z+\varepsilon})} \frac{\exp(\gamma(z))}{\exp(\gamma(z + \varepsilon))}^n \to 0
$$

as $n \to \infty$. This shows (54) in the case that $X$ is a lattice-valued random variable. The proof of the lemma is thus completed.

\[\square\]

**Lemma 3**

$$
\lim_{n \to \infty} \frac{1}{n} \ln \Pr_g\{z \leq X_n < z + \varepsilon\} = 0 \quad \text{for any} \ \varepsilon > 0.
$$

**Proof.** For notational simplicity, let $\sigma = \sqrt{\mathbb{E}_g[|X - z|^2]}$. If $\sigma = 0$, then $\Pr_g\{z \leq X_n < z + \varepsilon\} = 1$ and the lemma is obviously true. So, it remains to show the lemma for the case that $0 < \sigma < \infty$.

Since the random variable $X$ has mean $z$ and finite variance $\sigma^2 > 0$ associated with pdf or pmf $g(.)$, it follows from the central limit theorem that $\sqrt{n(X_n - z)}$ converges to a Gaussian random variable with zero mean and unit variance. Therefore, there exists an integer $m > 0$ such that

$$
\frac{1}{3} < \Pr_g\left\{\frac{\sqrt{n}(X_n - z)}{\sigma} \geq 0\right\} < \frac{2}{3}
$$

for any integer $n$ greater than $m$. Since the variance $\sigma^2$ is finite, it follows from the Chebyshev inequality that

$$
\Pr_g\{X_n \geq z + \varepsilon\} \leq \Pr_g\{X_n \geq \varepsilon\} \leq \frac{\sigma^2}{n \varepsilon^2}.
$$

Hence, for $n > \max\{m, \frac{3 \sigma^2}{\varepsilon^2}\}$, we have

$$
\Pr_g\{z \leq X_n < z + \varepsilon\} \leq \Pr_g\{X_n \geq z\} = \Pr_g\left\{\frac{\sqrt{n}(X_n - z)}{\sigma} \geq 0\right\} < \frac{2}{3}
$$

and

$$
\Pr_g\{z \leq X_n < z + \varepsilon\} = \Pr_g\{X_n \geq z\} - \Pr_g\{X_n \geq z + \varepsilon\}
$$

$$
= \Pr_g\left\{\frac{\sqrt{n}(X_n - z)}{\sigma} \geq 0\right\} - \Pr_g\{X_n \geq z + \varepsilon\}
$$

$$
> \frac{1}{3} - \frac{\sigma^2}{n \varepsilon^2} > 0.
$$

It follows that

$$
\frac{1}{n} \ln \left(\frac{1}{3} - \frac{\sigma^2}{n \varepsilon^2}\right) < \frac{1}{n} \ln \Pr_g\{z \leq X_n < z + \varepsilon\} < \frac{1}{n} \ln \frac{2}{3}
$$

23
for $n > \max\{m, \frac{3\varepsilon^2}{\sigma^2}\}$. Thus, $\lim_{n \to \infty} \frac{1}{n} \ln P_g\{z \leq X_n < z + \varepsilon\} = 0$ for any $\varepsilon > 0$. \hfill \Box

**Lemma 4**

\[
\limsup_{n \to \infty} \frac{1}{n} \ln \Pr\{z \leq X_n < z + \varepsilon\} \leq \sup_{y \in [z, z+\varepsilon)} \rho(y), \tag{62}
\]

\[
\liminf_{n \to \infty} \frac{1}{n} \ln \Pr\{z \leq X_n < z + \varepsilon\} \geq \inf_{y \in [z, z+\varepsilon)} \rho(y) \tag{63}
\]

provided that $\varepsilon > 0$ is sufficiently small.

**Proof.** We only prove the lemma for the case that $f(.)$ is a pmf, since the proof for the case that $f(.)$ is a pdf is similar. For simplicity of notations, let $x_n$ denote a realization of $X_n$. Let $0 < \varepsilon < \delta$. From (14), we have

\[
\Pr\{z \leq X_n < z + \varepsilon\} = \sum_{z \leq x_n < z + \varepsilon} \prod_{i=1}^{n} f(x_i)
\]

\[
\leq \sum_{z \leq x_n < z + \varepsilon} \exp(n\rho(x_n) + o(n)) \times \prod_{i=1}^{n} g(x_i)
\]

\[
\leq \sup_{y \in [z, z+\varepsilon)} \exp(n\rho(y) + o(n)) \sum_{z \leq x_n < z + \varepsilon} \prod_{i=1}^{n} g(x_i)
\]

\[
= \sup_{y \in [z, z+\varepsilon)} \exp(n\rho(y) + o(n)) \mathbb{P}_g\{z \leq X_n < z + \varepsilon\}. \tag{64}
\]

From the assumption that $\mathbb{E}_g[X] = z$ and $\mathbb{E}_g[|X - z|^2] < \infty$, according to Lemma 3, we have that

\[
\lim_{n \to \infty} \frac{1}{n} \ln \mathbb{P}_g\{z \leq X_n < z + \varepsilon\} = 0. \tag{65}
\]

Combining (64) and (65) yields (62).

From (65), we have

\[
\Pr\{z \leq X_n < z + \varepsilon\} = \sum_{z \leq x_n < z + \varepsilon} \prod_{i=1}^{n} f(x_i)
\]

\[
\geq \sum_{z \leq x_n < z + \varepsilon} \exp(n\rho(x_n) - o(n)) \times \prod_{i=1}^{n} g(x_i)
\]

\[
\geq \inf_{y \in [z, z+\varepsilon)} \exp(n\rho(y) - o(n)) \sum_{z \leq x_n < z + \varepsilon} \prod_{i=1}^{n} g(x_i)
\]

\[
= \inf_{y \in [z, z+\varepsilon)} \exp(n\rho(y) - o(n)) \mathbb{P}_g\{z \leq X_n < z + \varepsilon\}. \tag{66}
\]

Combining (64) and (66) yields (63). \hfill \Box

We are now in a position to prove the theorem. From the proof of Lemma 2, we know that

\[
\Pr\{\overline{X}_n \geq z\} \geq \Pr\{z \leq X_n < z + \varepsilon\} > 0 \tag{67}
\]
if \( \varepsilon > 0 \) is sufficiently small and \( n \) is sufficiently large. Therefore, in the sequel, we can restrict \( \varepsilon > 0 \) to be small enough and \( n \) to be large enough so that (67) holds.

To show assertion (I) of Theorem 3, note that
\[
\ln \Pr \{ \overline{X}_n \geq z \} = \ln \Pr \{ z \leq \overline{X}_n < z + \varepsilon \} - \ln \frac{\Pr \{ z \leq \overline{X}_n < z + \varepsilon \}}{\Pr \{ \overline{X}_n \geq z \}}. \tag{68}
\]

By virtue of (68), Lemma 2, and (62) of Lemma 4, we have
\[
\limsup_{n \to \infty} \frac{1}{n} \ln \Pr \{ \overline{X}_n \geq z \} = \limsup_{n \to \infty} \frac{1}{n} \ln \Pr \{ z \leq \overline{X}_n < z + \varepsilon \} - \limsup_{n \to \infty} \frac{1}{n} \ln \frac{\Pr \{ z \leq \overline{X}_n < z + \varepsilon \}}{\Pr \{ \overline{X}_n \geq z \}}
\]
\[
= \limsup_{n \to \infty} \frac{1}{n} \ln \Pr \{ z \leq \overline{X}_n < z + \varepsilon \} - \sup_{y \in [z, z + \varepsilon]} \rho(y)
\]
for sufficiently small \( \varepsilon > 0 \). By the assumption that \( \rho(y) \) is continuous in \( y \), it must be true that
\[
\lim_{n \to \infty} \frac{1}{n} \ln \Pr \{ \overline{X}_n \geq z \} \leq \limsup_{n \to \infty} \frac{1}{n} \ln \Pr \{ \overline{X}_n \geq z \} \leq \rho(z). \tag{69}
\]

It follows from (69) and Cramer-Chernoff theorem that
\[
\Pr \{ \overline{X}_n \geq z \} \leq \exp \left( n \lim_{m \to \infty} \frac{1}{m} \ln \Pr \{ \overline{X}_m \geq z \} \right) \leq \exp(n \rho(z)).
\]

This completes the proof of the assertion (I) of Theorem 3.

To show assertion (II) of Theorem 3, making use of (68), Lemma 2 and (63) of Lemma 4, we have
\[
\liminf_{n \to \infty} \frac{1}{n} \ln \Pr \{ \overline{X}_n \geq z \} = \liminf_{n \to \infty} \frac{1}{n} \ln \Pr \{ z \leq \overline{X}_n < z + \varepsilon \} - \liminf_{n \to \infty} \frac{1}{n} \ln \frac{\Pr \{ z \leq \overline{X}_n < z + \varepsilon \}}{\Pr \{ \overline{X}_n \geq z \}}
\]
\[
= \liminf_{n \to \infty} \frac{1}{n} \ln \Pr \{ z \leq \overline{X}_n < z + \varepsilon \} - \inf_{y \in [z, z + \varepsilon]} \rho(y)
\]
for sufficiently small \( \varepsilon > 0 \). By the assumption that \( \rho(y) \) is continuous in \( y \), it must be true that
\[
\lim_{n \to \infty} \frac{1}{n} \ln \Pr \{ \overline{X}_n \geq z \} \geq \liminf_{n \to \infty} \frac{1}{n} \ln \Pr \{ \overline{X}_n \geq z \} \geq \rho(z). \tag{70}
\]
Combining (69) and (70) yields \( \lim_{n \to \infty} \frac{1}{n} \ln \Pr \{ \overline{X}_n \geq z \} = \rho(z) \). This completes the proof of the assertion (II) of Theorem 3.

The proof of assertion (III) is similar to that of assertion (I). The proof of assertion (IV) is similar to that of assertion (II).

To show assertion (V), note that \( \prod_{i=1}^{n} f(X_i) \leq \exp(n \rho(\overline{X}_n)) \prod_{i=1}^{n} g(X_i) \), as a consequence of the assumption that \( \rho(\cdot) \) is a concave function such that \( f(X) \leq \exp(\rho(X)) \) \( g(X) \). Hence, applying the established assertions (I) and (III), we have that assertion (V) is true.

To show assertion (VI), note that \( \prod_{i=1}^{n} f(X_i) = \exp(n \rho(\overline{X}_n)) \prod_{i=1}^{n} g(X_i) \), as a consequence of the assumption that \( \rho(\cdot) \) is a linear function such that \( f(X) = \exp(\rho(X)) \) \( g(X) \). Hence, applying the established assertions (II) and (IV), we have that assertion (VI) is true. This completes the proof of the theorem.

### C Proof of Theorem 7

Let the pmf \( \Pr \{ X = x \} \) defined by (22) be denoted by \( f(x) \). Let \( g(x, \vartheta) \) denote the pmf \( \Pr \{ X = x \} \) defined by (22) with \( R \) replaced by \( \vartheta \in \{ 0, 1, \cdots, N \} \). Define
\[
\mathcal{M}(x, \vartheta) = \frac{f(x)}{g(x, \vartheta)}
\]
for \( \vartheta \in \{0, 1, \cdots, N\} \) and \( x \in \{0, 1, \cdots, n\} \) such that \( \vartheta + n - N \leq x \leq \vartheta \). Note that
\[
\mathcal{M}(x, \vartheta) = \frac{\binom{R}{x} (N-R)}{\binom{\vartheta}{x} (N-\vartheta)}.
\]

Define
\[
\mathcal{A} = \{ \vartheta \in \mathbb{Z}^+ : r \leq \vartheta \leq R \}, \quad \mathcal{B} = \{ \vartheta \in \mathbb{Z}^+ : R \leq \vartheta \leq r + N - n \}
\]
for \( r \in \{0, 1, \cdots, n\} \) and \( R \in \{0, 1, \cdots, N\} \) such that \( R + n - N \leq r \leq R \). Clearly, both \( \mathcal{A} \) and \( \mathcal{B} \) are nonempty as a consequence of \( R + n - N \leq r \leq R \). We need to establish some preliminary results.

**Lemma 5** Assume that \( r \in \{0, 1, \cdots, n\} \), \( R \in \{0, 1, \cdots, N\} \) and that \( R + n - N \leq r \leq R \). Then, \( f(X) \mathbb{1}_{\{X \leq r\}} \leq \mathcal{M}(r, \vartheta) g(X, \vartheta) \) holds for all \( \vartheta \in \mathcal{A} \). Similarly, \( f(X) \mathbb{1}_{\{X \geq r\}} \leq \mathcal{M}(r, \vartheta) g(X, \vartheta) \) holds for all \( \vartheta \in \mathcal{B} \).

**Proof.** Note that \( \mathcal{M}(r, \vartheta) \) is well-defined for \( \vartheta \in \mathcal{A} \) and \( r \in \{0, 1, \cdots, n\} \) such that \( R + n - N \leq r \leq R \). It is easy to see that \( f(x) \mathbb{1}_{\{x \leq r\}} \leq \mathcal{M}(r, \vartheta) g(x, \vartheta) \) holds for \( x \in \{0, 1, \cdots, r\} \) such that \( f(x) = 0 \). To show the first assertion, it suffices to show that \( f(x) \mathbb{1}_{\{x \leq r\}} \leq \mathcal{M}(r, \vartheta) g(x, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \{0, 1, \cdots, r\} \) such that \( f(x) > 0 \). Clearly, \( f(x) > 0 \) implies \( R + n - N \leq x \leq R \). Hence, for \( \vartheta \in \mathcal{A} \) and \( x \in \{0, 1, \cdots, r\} \) such that \( f(x) > 0 \), it must be true that \( R + n - N \leq x \leq \vartheta \leq R \). To show the second assertion, we can verify that
\[
\frac{\mathcal{M}(x+1, \vartheta)}{\mathcal{M}(x, \vartheta)} = \frac{(R-x)(N-\vartheta-n+x+1)}{(\vartheta-x)(N-R-n+x+1)} \geq 1
\]
for \( \vartheta \in \mathcal{A} \) and \( x \in \{0, 1, \cdots, r\} \) such that \( R + n - N \leq x \leq \vartheta \). This implies that for \( \vartheta \in \mathcal{A} \), \( \mathcal{M}(x, \vartheta) \) is increasing with respect to \( x \in \{0, 1, \cdots, r\} \) such that \( R + n - N \leq x \leq \vartheta \). Consequently, \( \frac{f(x)}{g(x, \vartheta)} = \mathcal{M}(x, \vartheta) \leq \mathcal{M}(r, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \{0, 1, \cdots, r\} \) such that \( f(x) > 0 \). Thus, we have shown that for \( r \in \{0, 1, \cdots, n\} \) such that \( R + n - N \leq r \leq R \), the inequality \( f(X) \mathbb{1}_{\{X \leq r\}} \leq \mathcal{M}(r, \vartheta) g(X, \vartheta) \) holds for all \( \vartheta \in \mathcal{A} \).

The second assertion can be shown in a similar manner. Note that \( \mathcal{M}(r, \vartheta) \) is well-defined for \( \vartheta \in \mathcal{B} \) and \( r \in \{0, 1, \cdots, n\} \) such that \( R + n - N \leq r \leq R \). It is easy to see that \( f(x) \mathbb{1}_{\{x \geq r\}} \leq \mathcal{M}(r, \vartheta) g(x, \vartheta) \) holds for \( x \in \{r, r+1, \cdots, n\} \) such that \( f(x) = 0 \). For \( \vartheta \in \mathcal{B} \) and \( x \in \{r, r+1, \cdots, n\} \) such that \( f(x) > 0 \), it must be true that \( R + n - N \leq \vartheta + n - N \leq x \leq R \leq \vartheta \) and \( \mathcal{M}(x, \vartheta) > 0 \). Making use of (71), we can verify that \( \frac{f(x)}{g(x, \vartheta)} = \mathcal{M}(x, \vartheta) \leq \mathcal{M}(r, \vartheta) \) holds for \( \vartheta \in \mathcal{B} \) and \( x \in \{r, r+1, \cdots, n\} \) such that \( x < R \). Consequently, \( \frac{f(x)}{g(x, \vartheta)} = \mathcal{M}(x, \vartheta) \leq \mathcal{M}(r, \vartheta) \) holds for \( \vartheta \in \mathcal{B} \) and \( x \in \{r, r+1, \cdots, n\} \) such that \( f(x) > 0 \). It follows that for \( r \in \{0, 1, \cdots, n\} \) such that \( R + n - N \leq r \leq R \), the inequality \( f(X) \mathbb{1}_{\{X \geq r\}} \leq \mathcal{M}(r, \vartheta) g(X, \vartheta) \) holds for all \( \vartheta \in \mathcal{B} \). This completes the proof of the lemma.

\[\square\]

Applying Theorem 11 and Lemma 5, we have the following results.

**Lemma 6** Assume that \( r \in \{0, 1, \cdots, n\} \), \( R \in \{0, 1, \cdots, N\} \) and that \( R + n - N \leq r \leq R \). Then,
\[
\Pr\{X \leq r\} \leq \frac{\binom{R}{x} (N-R)}{\binom{\vartheta}{x} (N-\vartheta)} \quad \text{for} \quad \vartheta \in \mathcal{A},
\]
\[
\Pr\{X \geq r\} \leq \frac{\binom{R}{x} (N-R)}{\binom{\vartheta}{x} (N-\vartheta)} \quad \text{for} \quad \vartheta \in \mathcal{B}.
\]

26
Lemma 7 Let \( \hat{R} = \min \{ N, [(N+1)\frac{r}{n}] \} \). Then, \( \hat{R} \in \mathcal{A} \) provided that \( \frac{r}{n} \leq \frac{R}{N} \). Similarly, \( \hat{R} \in \mathcal{B} \) provided that \( \frac{r}{n} \geq \frac{R}{N} \).

Proof. First, we shall show \( \hat{R} \in \mathcal{A} \) under the assumption that \( \frac{r}{n} \leq \frac{R}{N} \). Clearly, \( r \leq \hat{R} \). To show \( \hat{R} \in \mathcal{A} \), it remains to show \( \hat{R} \leq R \) by considering the cases that \( r = n \) and \( 0 \leq r < n \). In the case of \( r = n \), we have \( \hat{R} = R = N \). In the case of \( 0 \leq r < n \), we have \( \hat{R} = [(N+1)\frac{r}{n}] \) and \( Nr + r < nR + n \). Hence, \( (N+1)\frac{r}{n} < R + 1 \) and it follows that \( \hat{R} \leq R \). This proves that \( \hat{R} \in \mathcal{A} \) provided that \( \frac{r}{n} \leq \frac{R}{N} \).

Next, we shall show \( \hat{R} \in \mathcal{B} \) under the assumption that \( \frac{r}{n} \geq \frac{R}{N} \). We need to consider the cases that \( r = n \) and \( 0 \leq r < n \). In the case of \( r = n \), we have \( \hat{R} = R + N = N \geq R \). In the case of \( 0 \leq r < n \), we have \( \hat{R} = [(N+1)\frac{r}{n}] \) and \( Nr + r \geq nR \), which imply that \( \hat{R} \geq R \). Moreover, we observe that

\[
(N+1)\frac{r}{n} - (r + N - n + 1) = -\left(1 - \frac{r}{n}\right)(N + 1 - n) < 0
\]

and hence \( \hat{R} = [(N+1)\frac{r}{n}] \leq r + N - n \).

Finally, Theorem 7 follows from Lemmas 6 and 7.

D Proof of Theorem 8

Define \( Y = \sum_{i=1}^{n} X_i \) and \( \theta = n\lambda \). Using the property of parameter additivity of the generalized Poisson distribution, we have that \( Y \) possesses the following distribution:

\[
f(y) = \Pr\{Y = y\} = \frac{\theta(\theta + y\alpha)^{y-1}e^{-\theta - y\alpha}}{y!}, \quad y = 0, 1, 2, \ldots. \tag{72}
\]

It can be shown that the mean of \( Y \) is \( \mu = \mathbb{E}[Y] = \frac{\theta}{1 - \alpha} \). Let \( \alpha \in (0, 1) \) be fixed. Define

\[
g(y, \vartheta) = \frac{\partial(y + y\alpha)^{y-1}e^{-\theta - y\alpha}}{y!}, \quad y = 0, 1, 2, \ldots
\]

and

\[
\mathcal{M}(y, \vartheta) = \frac{f(y)}{g(y, \vartheta)} = \frac{\theta e^{-\theta}}{\vartheta e^{-\vartheta}} \left( \frac{\theta + y\alpha}{\vartheta + y\alpha} \right)^{y-1} = \frac{\theta e^{-\theta}}{\vartheta e^{-\vartheta}} \exp(h(y, \vartheta)),
\]

where \( h(y, \vartheta) = (y - 1)\ln \frac{\vartheta + y\alpha}{\vartheta + y\alpha} \). Straightforward but tedious calculation shows that

\[
\frac{\partial h(y, \vartheta)}{\partial y} = \frac{\alpha(\vartheta - \theta)(y - 1)}{(\vartheta + y\alpha)(\vartheta + y\alpha)} + \ln \frac{\vartheta + y\alpha}{\vartheta + y\alpha}
\]

and

\[
\frac{\partial^2 h(y, \vartheta)}{\partial y^2} = \frac{\alpha(\vartheta - \theta)}{(\vartheta + y\alpha)(\vartheta + y\alpha)} \left( \frac{\theta + \alpha}{\vartheta + y\alpha} + \frac{\vartheta + \alpha}{\vartheta + y\alpha} \right).
\]

Let \( 0 < \vartheta \leq \theta \). Observing that \( \lim_{y \to \infty} \frac{\partial h(y, \vartheta)}{\partial y} = 0 \) and \( \frac{\partial^2 h(y, \vartheta)}{\partial y^2} < 0 \) for \( y \geq 0 \), we have that \( \frac{\partial h(y, \vartheta)}{\partial y} > 0 \) for \( y \geq 0 \). Thus, \( h(y, \vartheta) \) is increasing with respect to \( y \geq 0 \). This shows that for \( 0 < \vartheta \leq \theta \), \( \mathcal{M}(y, \vartheta) \) is increasing with respect to \( y \geq 0 \). It follows that

\[
f(Y) I\{Y \leq y\} \leq \mathcal{M}(y, \vartheta) g(Y, \vartheta) \quad \text{for } 0 < \vartheta \leq \theta.
\]
Therefore, invoking Theorem 1, we have

\[
\Pr\{X_n \leq z\} = \Pr\{Y \leq nz\} \leq \inf_{\theta \leq \theta_0} \left\{ \frac{\theta e^{-\theta}}{\theta + n z \alpha} \right\}^{n z - 1}. \tag{73}
\]

To show (26), we use the method of moments. Specifically, we seek \( \omega \) such that \( \mathbb{E}_\omega[Y] = nz \), that is, \( \frac{\partial}{\partial \omega} = nz \), from which we obtain \( \theta = nz(1 - \alpha) \). As a consequence of \( z \geq \frac{\lambda}{1 - \alpha} \), such \( \omega \) is no greater than \( \theta \). It follows from (73) that

\[
\Pr\{X_n \leq z\} \leq \frac{\theta e^{-\theta}}{\theta + \theta + n z \alpha} \left( \frac{\theta + n z \alpha}{\theta + n z \alpha} \right)^{n z - 1},
\]

with \( \omega = nz(1 - \alpha) \). Substituting \( \omega \) into the right side of this inequality yields (26).

To show (27), we use the method of maximum likelihood to find the tightest bound. For this purpose, we define \( L(z, \omega) = \omega - \ln \omega - (nz - 1) \ln(\omega + n z \alpha) \) and attempt to minimize the function \( L(z, \omega) \) subject to \( 0 < \omega \leq \theta \). Note that \( \frac{\partial L(z, \omega)}{\partial \omega} = 1 - \frac{1}{\omega} - \frac{n z - 1}{\omega + n z \alpha} \) and

\[
\frac{\partial^2 L(z, \omega)}{\partial \omega^2} = \frac{1}{\omega^2} - \frac{1}{(\omega + n z \alpha)^2} + \frac{n z}{(\omega + n z \alpha)^2} > 0.
\]

To make \( \frac{\partial L(z, \omega)}{\partial \omega} = 0 \), it suffices to have \( \omega^2 - (1 - \alpha)nz\omega - n z \alpha = 0 \). This equation has a unique nonnegative solution, which is

\[
\omega = \frac{1}{2} \left( 1 - \alpha \right) n z + \sqrt{\left( 1 - \alpha \right) n z + 4 n z \alpha} = n \nu.
\]

It can be checked that \( \frac{\partial L(z, \omega)}{\partial \omega} |_{\omega = \theta} = 1 - \frac{1}{\theta} - \frac{n z - 1}{\theta + n z \alpha} \geq 0 \) provided that \( z \leq \frac{\lambda}{1 - \alpha} \). Hence, \( n \nu \leq \theta \) and it follows from (73) that

\[
\Pr\{X_n \leq z\} \leq \frac{\theta e^{-\theta}}{n \nu e^{-n \nu}} \left( \frac{\theta + n z \alpha}{\nu + n z \alpha} \right)^{n z - 1} = \frac{\lambda(\nu + z \alpha)}{\nu(\lambda + z \alpha)} \left[ \left( \frac{\lambda + z \alpha}{\nu + z \alpha} \right)^z e^\nu \right]^n
\]

for \( 0 < z \leq \frac{\lambda}{1 - \alpha} \). This establishes (28).

To show (27) and (29), let \( \omega \geq \theta \). Noting that \( \lim_{y \to \infty} \frac{\partial h(y, \omega)}{\partial y} = 0 \) and \( \frac{\partial^2 h(y, \omega)}{\partial y^2} > 0 \) for \( y \geq 0 \), we have that \( \frac{\partial h(y, \omega)}{\partial y} < 0 \) for \( y \geq 0 \). Thus, \( h(y, \omega) \) is decreasing for \( y \geq 0 \). This shows that for \( \omega \geq \theta \), \( \mathbb{M}(y, \omega) \) is decreasing with respect to \( y \geq 0 \). It follows that

\[
f(Y) \mathbb{1}_{Y \geq \theta} \leq \mathbb{M}(y, \omega)g(Y, \theta) \quad \text{for } \omega \geq \theta.
\]

Hence, invoking Theorem 4, we have

\[
\Pr\{X_n \geq z\} = \Pr\{Y \geq nz\} \leq \inf_{\theta \geq \theta_0} \left\{ \frac{\theta e^{-\theta}}{\theta + n z \alpha} \right\}^{n z - 1}. \tag{74}
\]

To show (27), we use the method of moments. Specifically, we seek \( \omega \in [\theta, \infty) \) such that \( \mathbb{E}_\omega[Y] = nz \), that is, \( \frac{\partial}{\partial \omega} = nz \), from which we obtain \( \theta = nz(1 - \alpha) \). As a consequence of \( z \geq \frac{\lambda}{1 - \alpha} \), such \( \theta \) is no less than \( \theta \). It follows from (74) that

\[
\Pr\{X_n \geq z\} \leq \frac{\theta e^{-\theta}}{\theta + n z \alpha} \left( \frac{\theta + n z \alpha}{\theta + n z \alpha} \right)^{n z - 1}
\]

with \( \omega = nz(1 - \alpha) \). Substituting \( \omega \) into the right side of this inequality yields (27).
To show (29), we use the method of maximum likelihood to find the tightest bound. For this purpose, it suffices to minimize the function $L(z, \theta)$ subject to $\vartheta \geq \theta$. It can be verified that \( \frac{\partial L(z, \vartheta)}{\partial \vartheta} |_{\vartheta = \theta} = 1 - \frac{z - 1}{\theta + n z} \leq 0 \) provided that $z \geq \frac{\lambda}{\theta + n z}$. Hence, $n \nu \geq \theta$ and it follows from (74) that

\[
\Pr\{X_n \geq z\} \leq \frac{\theta e^{-\theta}}{n \nu e^{-n \nu}} \left( \frac{\theta + n z \alpha}{n \nu + n z \alpha} \right)^{n z - 1} = \frac{\lambda (\nu + z \alpha)}{\nu (\lambda + z \alpha)} \left[ \left( \frac{\lambda + z \alpha}{\nu + z \alpha} \right)^{\nu} \right]^{n}
\]

for $z \geq \frac{\lambda}{1 - \alpha + \frac{\lambda}{\theta}}$. This establishes (29).

In the special case that $\alpha = 0$, the generalized Poisson distribution reduces to the standard Poisson distribution in the exponential family with pmf

\[
f(x) = \Pr\{X = x\} = \frac{\lambda^x e^{-\lambda}}{x!} = v(x) \exp (\eta(\lambda) u(x) - \zeta(\lambda)), \quad x \in \{0, 1, 2, \cdots\}
\]

where

\[
u(x) = x, \quad \eta(\lambda) = \ln \lambda, \quad \zeta(\lambda) = \lambda.
\]

The moment generating function is $\phi(s) = \mathbb{E}[e^{sX}] = e^{-\lambda} \exp(\lambda e^s)$. Using $\mathbb{E}[X^\ell] = \frac{d^\ell \phi(s)}{ds^\ell} |_{s=0}$ for $\ell = 1, 2, \cdots$, we have

\[
\mathbb{E}[X] = \lambda, \quad \mathbb{E}[X^2] = \lambda(1 + \lambda), \quad \mathbb{E}[X^3] = \lambda^2 + (1 + \lambda)^2 \lambda, \quad \mathbb{E}[X^4] = \lambda^2 + (1 + \lambda)(3\lambda^2 + (1 + \lambda)^2 \lambda).
\]

Hence, $\mathbb{E}[|X - \lambda|^2] = \lambda$ and $\mathbb{E}[|X - \lambda|^4] = \mathbb{E}[X^4 - 4X^3\lambda + 6X^2\lambda^2 - 4X\lambda^3 + \lambda^4] = \lambda(3\lambda + 1)$. Using these facts and Lyapunov’s inequality, we have

\[
\frac{\mathbb{E}[|X - \lambda|^3]}{\mathbb{E}[|X - \lambda|^2]} < \frac{\mathbb{E}[|X - \lambda|^4]}{\mathbb{E}[|X - \lambda|^2]} = \left( 3 + \frac{1}{\lambda} \right)^{3/4}.
\]

Making use of this inequality, the assertions (III) and (IV) of Theorem 5 we can establish (30) and (31).

### E Proof of Theorem 10

We need to define some quantities. Define

$$
\mathcal{C} = [c_0, c_1, \cdots, c_k]^\top, \quad \hat{\mathcal{C}} = [\hat{c}_0, \hat{c}_1, \cdots, \hat{c}_k]^\top,
$$

$$
\mathcal{X} = \left\{ [x_0, x_1, \cdots, x_k]^\top : x_i \in \mathbb{Z}^+, \ i = 0, 1, \cdots, k \text{ and } \sum_{i=0}^k x_i = n \right\}
$$

and

$$
\mathcal{X} = \left\{ [\vartheta_0, \vartheta_1, \cdots, \vartheta_k]^\top : \vartheta_i \in \mathbb{R}, \ \frac{\vartheta_i}{N} > 0 \text{ for } i = 0, 1, \cdots, k \text{ and } \sum_{i=0}^k \vartheta_i = N \right\}.
$$

Let $g(x, \vartheta)$, where $x = [x_0, x_1, \cdots, x_k]^\top \in \mathcal{X}$ and $\vartheta = [\vartheta_0, \vartheta_1, \cdots, \vartheta_k]^\top \in \mathcal{X}$, denote the pmf defined by (33) with $\mathcal{C}_i$ replaced by $\vartheta_i$ for $i = 0, 1, \cdots, k$. Define $\mathcal{M}(x, \vartheta) = \frac{g(x)}{g(x, \vartheta)}$ for $x \in \mathcal{X}$ and $\vartheta \in \mathcal{X}$ such that $g(x, \vartheta) > 0$. Define $\mathcal{A} = \{ x \in \mathcal{X} : f(x) > 0 \}$ and

$$
\mathcal{B} = \{ \vartheta \in \mathcal{X} : |\vartheta| \geq |\mathcal{C}|, \ g(x, \vartheta) > 0 \}, \quad \mathcal{B} = \{ \vartheta \in \mathcal{X} : |\vartheta| \geq |\mathcal{C}|, \ g(z, \vartheta) > 0 \}.
$$

**Lemma 8** Assume that $f(z) > 0$. Then, $f(X)|_{X \leq z} \leq \mathcal{M}(z, \vartheta)g(X, \vartheta)$ holds for all $\vartheta \in \mathcal{A}$. Similarly, $f(X)|_{X > z} \leq \mathcal{M}(z, \vartheta)g(X, \vartheta)$ holds for all $\vartheta \in \mathcal{B}$. 
Proof. First, we shall show the first assertion that \( f(X) \mathbb{1}_{\{X \prec \vartheta\}} \leq \mathcal{M}(\vartheta) g(X, \vartheta) \) holds for all \( \vartheta \in \mathcal{A} \).

Since \( f(\vartheta) = g(\vartheta, C) \), the set \( \mathcal{A} \) is nonempty as a consequence of the assumption that \( f(\vartheta) > 0 \). Hence, \( \mathcal{M}(\vartheta) \) is well-defined for \( \vartheta \in \mathcal{A} \) and \( \vartheta \in \mathcal{A} \). Note that \( f(\vartheta) \mathbb{1}_{\{X \prec \vartheta\}} \leq \mathcal{M}(\vartheta) g(x, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( f(x) = 0 \). To show the first assertion, it suffices to show that \( f(\vartheta) \mathbb{1}_{\{X \prec \vartheta\}} \leq \mathcal{M}(\vartheta) g(x, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \). We need to consider two cases as follows.

Case (I): \( N > 0 \).

Case (II): \( N < 0 \).

In Case (I), as a consequence of \( N > 0 \), we have \( \vartheta_i > 0, C_i > 0 \) for \( i = 0, 1, \cdots, \kappa \). Consider \( x \in \mathcal{X} \) such that \( x \prec \vartheta \). Note that \( \prod_{l=0}^{x_0} (\vartheta_l - \ell + 1) \geq \prod_{l=1}^{x_0} (\vartheta_l - \ell + 1) > 0 \) because of \( x \prec \vartheta \) and \( g(z, \vartheta) > 0 \). Moreover, \( \prod_{l=0}^{x_0} (\vartheta_l - \ell + 1) \geq \prod_{l=1}^{x_0} (C_0 - \ell + 1) > 0 \) because of \( \vartheta_0 \geq C_0 \) and \( f(x) > 0 \). This implies that \( g(x, \vartheta) > 0 \) for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( x \prec \vartheta \). Since \( f(x) > 0 \), we have that

\[
\mathcal{M}(x, \vartheta) = \frac{f(x)}{g(x, \vartheta)} = \prod_{l=0}^{x_0} \frac{(\vartheta_l - \ell + 1)}{C_l - \ell + 1} \prod_{l=1}^{x_0} \frac{(C_i - \ell + 1)}{\vartheta_l - \ell + 1} \prod_{l=1}^{x_0} \frac{(C_i - \ell + 1)}{\vartheta_l - \ell + 1}
\]

is well-defined and positive for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( x \prec \vartheta \). Consequently, we have a meaningful ratio

\[
\frac{\mathcal{M}(z, \vartheta)}{\mathcal{M}(x, \vartheta)} = \left( \prod_{l=0}^{x_0} \frac{\vartheta_l - \ell + 1}{C_l - \ell + 1} \right) \left( \prod_{l=1}^{x_0} \frac{C_i - \ell + 1}{\vartheta_l - \ell + 1} \right) \quad \text{for } \vartheta \in \mathcal{A} \text{ and } x \in \mathcal{X} \text{ such that } x \prec \vartheta.
\]

(75)

Since \( \vartheta_0 > C_0 \) and \( \vartheta_i < C_i \) for \( i = 1, \cdots, \kappa \), it follows from (75) that \( \mathcal{M}(x, \vartheta) \leq \mathcal{M}(z, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( x \prec \vartheta \).

In Case (II), as a consequence of \( N < 0 \), we have \( \vartheta_i < 0, C_i < 0 \) for \( i = 0, 1, \cdots, \kappa \). Consider \( x \in \mathcal{X} \) such that \( x \prec \vartheta \). Note that for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \),

\[
\mathcal{M}(x, \vartheta) = \prod_{l=0}^{x_0} \frac{\vartheta_l - \ell + 1}{C_l - \ell + 1} \prod_{l=1}^{x_0} \frac{C_i + \ell - 1}{\vartheta_l - \ell + 1} = \prod_{l=1}^{x_0} \frac{C_i + \ell - 1}{\vartheta_l - \ell + 1} \prod_{l=1}^{x_0} \frac{C_i - \ell + 1}{\vartheta_l - \ell + 1}
\]

is well-defined and positive. Moreover, for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( x \prec \vartheta \), we have a meaningful ratio

\[
\frac{\mathcal{M}(z, \vartheta)}{\mathcal{M}(x, \vartheta)} = \left( \prod_{l=0}^{x_0} \frac{|\vartheta_l| + \ell - 1}{|C_l| + \ell - 1} \right) \left( \prod_{l=1}^{x_0} \frac{|C_i| + \ell - 1}{|\vartheta_l| + \ell - 1} \right).
\]

(76)

Since \( |\vartheta_0| > |C_0| \) and \( |\vartheta_i| < |C_i| \) for \( i = 1, \cdots, \kappa \), it follows from (76) that \( \mathcal{M}(x, \vartheta) \leq \mathcal{M}(z, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( x \prec \vartheta \). Therefore, in both cases, we have that

\[
\frac{f(x)}{g(x, \vartheta)} = \mathcal{M}(x, \vartheta) \leq \mathcal{M}(z, \vartheta)
\]

holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( x \prec z \). Thus, we have shown that \( f(x) \mathbb{1}_{\{X \prec z\}} \leq \mathcal{M}(z, \vartheta) g(x, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \). This completes the proof of the first assertion of the lemma.

Next, we shall show the second assertion that \( f(X) \mathbb{1}_{\{X \succ z\}} \leq \mathcal{M}(z, \vartheta) g(X, \vartheta) \) holds for all \( \vartheta \in \mathcal{B} \).

Since \( f(z) = g(z, C) \), the set \( \mathcal{B} \) is nonempty as a consequence of the assumption that \( f(z) > 0 \). Hence, \( \mathcal{M}(z, \vartheta) \) is well-defined for \( z \in \mathcal{X} \) and \( \vartheta \in \mathcal{B} \). Note that \( f(x) \mathbb{1}_{\{X \succ z\}} \leq \mathcal{M}(z, \vartheta) g(x, \vartheta) \) holds for \( \vartheta \in \mathcal{B} \) and \( x \in \mathcal{X} \) such that \( f(x) = 0 \). To show the second assertion, it suffices to show that \( f(x) \mathbb{1}_{\{X \succ z\}} \leq \mathcal{M}(z, \vartheta) g(x, \vartheta) \) holds for \( \vartheta \in \mathcal{B} \) and \( x \in \mathcal{X} \). We need to consider two cases as follows.

Case (I): \( N > 0 \).
Case (II): $N < 0$.

In Case (I), as a consequence of $N > 0$, we have $0 < \vartheta_0 \leq C_0$ and $\vartheta_i \geq C_i > 0$ for $i = 1, \cdots, \kappa$. Consider $x \in \mathcal{X}$ such that $x \succ z$. Note that $\prod_{i=1}^{k} \prod_{\ell=z_i+1}^{C_i} (\vartheta_i - \ell + 1) \geq \prod_{i=1}^{k} \prod_{\ell=z_i+1}^{C_i} (C_i - \ell + 1) > 0$ because of $\vartheta > C$ and $f(x) > 0$. Moreover, $\prod_{i=1}^{\kappa} (\vartheta_0 - \ell + 1) \geq \prod_{i=1}^{\kappa} (\vartheta_0 - \ell + 1) > 0$ because of $x \succ z$ and $g(z, \vartheta) > 0$. This implies that $g(x, \vartheta) > 0$ for $\vartheta \in \mathcal{B}$ and $x \in \mathcal{X}$ such that $x \succ z$. Since $f(x) > 0$, we have that

$$\mathcal{M}(x, \vartheta) = \frac{f(x)}{g(x, \vartheta)} = \prod_{i=0}^{k} \frac{\mathcal{C}_i}{\vartheta_i} = \prod_{i=0}^{k} \frac{C_i - \ell + 1}{\vartheta_i - \ell + 1}$$

is well-defined and positive for $\vartheta \in \mathcal{B}$ and $x \in \mathcal{X}$ such that $x \succ z$. Consequently, we have a meaningful ratio

$$\frac{\mathcal{M}(x, \vartheta)}{\mathcal{M}(z, \vartheta)} = \left( \prod_{\ell=z_0+1}^{\vartheta_0} \frac{\vartheta_0 - \ell + 1}{C_0 - \ell + 1} \right) \left( \prod_{i=1}^{\kappa} \prod_{\ell=z_i+1}^{C_i} \frac{C_i - \ell + 1}{\vartheta_i - \ell + 1} \right)$$

for $\vartheta \in \mathcal{B}$ and $x, z \in \mathcal{X}$ such that $x \succ z$. (77)

It can be seen from (77) that $\mathcal{M}(x, \vartheta) \leq \mathcal{M}(z, \vartheta)$ holds for $\vartheta \in \mathcal{B}$ and $x, z \in \mathcal{X}$ such that $x \succ z$.

In Case (II), as a consequence of $N < 0$, we have $\vartheta_i < 0$, $C_i < 0$ for $i = 1, \cdots, \kappa$. Consider $x \in \mathcal{X}$ such that $x \succ z$. Note that for $\vartheta \in \mathcal{B}$ and $x \in \mathcal{X}$,

$$\mathcal{M}(x, \vartheta) = \prod_{i=0}^{k} \prod_{\ell=1}^{C_i} \frac{C_i - \ell + 1}{\vartheta_i - \ell + 1} = \prod_{i=0}^{k} \prod_{\ell=z_i+1}^{C_i} \frac{C_i - \ell + 1}{\vartheta_i - \ell + 1}$$

is well-defined and positive. Moreover, for $\vartheta \in \mathcal{B}$ and $x \in \mathcal{X}$ such that $x \succ z$, we have a meaningful ratio

$$\frac{\mathcal{M}(x, \vartheta)}{\mathcal{M}(z, \vartheta)} = \left( \prod_{\ell=z_0+1}^{\vartheta_0} \frac{\vartheta_0 - \ell + 1}{C_0 - \ell + 1} \right) \left( \prod_{i=1}^{\kappa} \prod_{\ell=z_i+1}^{C_i} \frac{C_i - \ell + 1}{\vartheta_i - \ell + 1} \right).$$

Since $|\vartheta_0| < |C_0|$ and $|\vartheta_i| > |C_i|$ for $i = 1, \cdots, \kappa$, it follows from (78) that $\mathcal{M}(x, \vartheta) \leq \mathcal{M}(z, \vartheta)$ holds for $\vartheta \in \mathcal{B}$ and $x \in \mathcal{X}$ such that $x \succ z$. Therefore, in both cases, we have that

$$\frac{f(x)}{g(x, \vartheta)} = \mathcal{M}(x, \vartheta) \leq \mathcal{M}(z, \vartheta)$$

holds for $\vartheta \in \mathcal{B}$ and $x \in \mathcal{X}$ such that $x \succ z$. Thus, we have shown that $f(x) \mathbb{I}_{\{x \succ z\}} \leq \mathcal{M}(z, \vartheta) g(x, \vartheta)$ holds for $\vartheta \in \mathcal{B}$ and $x \in \mathcal{X}$. This completes the proof of the second assertion of the lemma.

Applying Theorem II and Lemma 8, we have the following results.

**Lemma 9** Assume that $f(z) > 0$. Then,

$$\Pr\{X < z\} \leq \prod_{i=0}^{\kappa} \frac{C_i}{\vartheta_i} \quad \text{for} \quad \vartheta \in \mathcal{A},$$

$$\Pr\{X \succ z\} \leq \prod_{i=0}^{\kappa} \frac{C_i}{\vartheta_i} \quad \text{for} \quad \vartheta \in \mathcal{B}.$$

**Lemma 10** $g(z, \hat{C}) > 0$ for $z \in \mathcal{X}$.
Proof. Recall that \( \tilde{C} = [\tilde{C}_0, \tilde{C}_1, \cdots, \tilde{C}_\kappa]^T \) with \( \tilde{C}_i = N \frac{z_i}{n} \) for \( i = 0, 1, \cdots, \kappa \).

In the case of \( N < 0 \), we have \( \tilde{C}_i \leq 0 \) for \( i = 0, 1, \cdots, \kappa \), which implies that \( 1 + \tilde{C}_i \leq 1 \) for \( i = 0, 1, \cdots, \kappa \). Hence, \( g(z, \tilde{C}) > 0 \) for \( z \in \mathcal{X} \) if \( N < 0 \).

In the case of \( N > 0 \), we have \( 0 < n < N + 1 \) and \( \tilde{C}_i \geq 0 \) for \( i = 0, 1, \cdots, \kappa \). For \( i = 0, 1, \cdots, \kappa \), we have \( z_i < \tilde{C}_i + 1 = 1 \) if \( z_i = 0 \); and \( z_i = \frac{N}{n} \tilde{C}_i < \tilde{C}_i + \frac{N}{n} = \tilde{C}_i + 1 \) if \( z_i > 0 \). Hence, \( g(z, \tilde{C}) > 0 \) for \( z \in \mathcal{X} \) if \( N > 0 \).

\[ \square \]

Lemma 11 \( |\tilde{C}| < |C| \Leftrightarrow z < \mu \text{ and } |\tilde{C}| > |C| \Leftrightarrow z > \mu \).

Proof. In the case of \( N > 0 \), we have \( C_i > 0 \) and \( \tilde{C}_i \geq 0 \) for \( i = 0, 1, \cdots, \kappa \). Hence,
\[
|\tilde{C}| < |C| \Leftrightarrow \tilde{C} < C \Leftrightarrow z < \mu, \quad |\tilde{C}| > |C| \Leftrightarrow \tilde{C} > C \Leftrightarrow z > \mu.
\]

In the case of \( N < 0 \), we have \( C_i < 0 \) and \( \tilde{C}_i \leq 0 \) for \( i = 0, 1, \cdots, \kappa \). Hence,
\[
|\tilde{C}| < |C| \Leftrightarrow \tilde{C} < C \Leftrightarrow z < \mu, \quad |\tilde{C}| > |C| \Leftrightarrow \tilde{C} > C \Leftrightarrow z > \mu.
\]

\[ \square \]

We are now in a position to prove the theorem. Clearly, the assertions (I) and (II) of Theorem 10 follow from Lemmas 9, 10, and 11.

To show assertion (III) of Theorem 10, we need to use the following inequalities
\[
\sqrt{2\pi x^{x-1/2}e^{-x}} < \Gamma(x) \leq e^{x-1/2}e^{-x}, \quad x \geq 1.
\] (79)

This result is established by Mortici and Chen [13] page 70, eq. (1.3)]. Making use of (79) and the fact that \( \Gamma(x+1) = x\Gamma(x) \), we have
\[
\sqrt{2\pi x^{x+1}e^{-x}} < \Gamma(x+1) \leq e^{x+1/2}e^{-x} \quad \text{for all real number } x \geq 1.
\] (80)

Using (80) and the assumption that \( C_i - z_i \geq 1 \) and \( \tilde{C}_i - z_i \geq 1 \) for \( i = 0, 1, \cdots, k \), we have
\[
\frac{C_i}{\tilde{C}_i} = \frac{\Gamma(C_i+1)}{\Gamma(C_i)} \cdot \frac{\Gamma(C_i - z_i + 1)}{\Gamma(C_i - z_i)} = \frac{e^{\frac{C_i}{2}}}{2\pi (C_i - z_i)^{C_i - z_i + \frac{1}{2}}} \frac{(C_i - z_i)^{C_i - z_i + \frac{1}{2}}}{\tilde{C}_i^{C_i + \frac{1}{2}}} \leq \frac{e^{\frac{C_i}{2}}}{2\pi (C_i - z_i)^{C_i - z_i + \frac{1}{2}}} \frac{(N - z_i)^{C_i - z_i + \frac{1}{2}}}{(\frac{N}{n})^{C_i + \frac{1}{2}}}
\]
for \( i = 0, 1, \cdots, k \). Making use of this inequality and the fact that
\[
\sum_{i=0}^{k} z_i = n, \quad \sum_{i=0}^{k} (\tilde{C}_i - z_i + \frac{1}{2}) = N - n + k + \frac{1}{2},
\]
we have
\[
\sum_{i=0}^{k} z_i = n, \quad \sum_{i=0}^{k} (\tilde{C}_i - z_i + \frac{1}{2}) = N - n + k + \frac{1}{2},
\]

32
we have that (36) holds.

to show assertion (IV) of Theorem 10, we observe that

$$\frac{1}{\mathcal{N}} \ln \left[ \left( \frac{\mathcal{N} - n}{\mathcal{N}} \right)^{N-n} \prod_{i=0}^{k} \left( \frac{C_i}{C_i - z_i} \right)^{C_i - z_i} \right] \to (1 - \alpha) \ln(1 - \alpha) + \sum_{i=0}^{k} (\theta_i - \alpha \beta_i) \ln \frac{\theta_i}{\theta_i - \alpha \beta_i} \quad (81)$$

as \( \mathcal{N} \to \infty \) under the constraint that \( \frac{\mathcal{N}}{n} \to \alpha \in (0, 1) \) and \( \frac{\mathcal{N}}{n} \to \beta_i \in (0, 1) \) with \( \theta_i > \alpha \beta_i \) for \( i = 0, 1, \cdots, k \).

Using the weighted AM-GM inequality, we have

$$\prod_{i=0}^{k} \left( \frac{\theta_i}{\theta_i - \alpha \beta_i} \right)^{\frac{\theta_i - \alpha \beta_i}{1 - \alpha}} \leq \sum_{i=0}^{k} \left( \frac{\theta_i}{\theta_i - \alpha \beta_i} \right)^{\frac{\theta_i - \alpha \beta_i}{1 - \alpha}} = \frac{1}{1 - \alpha} \quad (82)$$

We claim that (82) does not hold with equality. To prove the claim, note that if (82) holds with equality, then there exists a number \( \nu \) such that

$$\frac{\theta_i}{\theta_i - \alpha \beta_i} = \frac{1}{\nu}, \quad i = 0, 1, \cdots, k.$$

It follows that

$$\frac{1}{1 - \alpha} = \prod_{i=0}^{k} \left( \frac{\theta_i}{\theta_i - \alpha \beta_i} \right)^{\frac{\theta_i - \alpha \beta_i}{1 - \alpha}} = \prod_{i=0}^{k} \left( \frac{1}{\nu} \right)^{\frac{\theta_i - \alpha \beta_i}{1 - \alpha}} = \frac{1}{\nu},$$

which implies that \( \nu = 1 - \alpha \) and thus

$$\frac{\theta_i}{\theta_i - \alpha \beta_i} = \frac{1}{1 - \alpha}, \quad i = 0, 1, \cdots, k.$$

Consequently, \( \theta_i = \beta_i \) for \( i = 0, 1, \cdots, k \). This contradicts the assumption that \( \theta_i \neq \beta_i \) for some \( i \in \{0, 1, \cdots, \kappa\} \). This proves our claim and thus

$$\prod_{i=0}^{k} \left( \frac{\theta_i}{\theta_i - \alpha \beta_i} \right)^{\frac{\theta_i - \alpha \beta_i}{1 - \alpha}} < \frac{1}{1 - \alpha},$$

which can be written as

$$(1 - \alpha) \ln(1 - \alpha) + \sum_{i=0}^{k} (\theta_i - \alpha \beta_i) \ln \frac{\theta_i}{\theta_i - \alpha \beta_i} < 0. \quad (83)$$

As a result of (81) and (83), we have that

$$\left[ \left( \frac{\mathcal{N} - n}{\mathcal{N}} \right)^{N-n} \frac{1}{\mathcal{N}} \prod_{i=0}^{k} \left( \frac{C_i}{C_i - z_i} \right)^{C_i - z_i} \right]^\mathcal{N} \to 0.$$

This leads to the truth of assertion (IV). The proof of the theorem is thus completed.

**F Proof of Theorem 11**

We need to define some quantities. Define

$$\mathcal{C} = [C_0, C_1, \cdots, C_\kappa]^\top, \quad \tilde{\mathcal{C}} = [\tilde{C}_0, \tilde{C}_1, \cdots, \tilde{C}_\kappa]^\top,$$
\( X = \left\{ [x_0, x_1, \ldots, x_\kappa]^T : x_0 = \gamma, x_i \in \mathbb{Z}^+, i = 1, \ldots, \kappa \text{ and } \frac{1}{N} \sum_{i=0}^\kappa x_i < 1 + \frac{1}{N} \right\} \)

and

\[ \Theta = \left\{ [\vartheta_0, \vartheta_1, \ldots, \vartheta_\kappa]^T : \vartheta_i \in \mathbb{R}, \vartheta_0 > \gamma - \frac{1}{N}, \vartheta_i \frac{1}{N} > 0 \text{ for } i = 0, 1, \ldots, \kappa \text{ and } \sum_{i=0}^\kappa \vartheta_i = N \right\} \]

Let \( g(x, \vartheta) \), where \( x = [x_0, x_1, \ldots, x_\kappa]^T \in X \) and \( \vartheta = [\vartheta_0, \vartheta_1, \ldots, \vartheta_\kappa]^T \in \Theta \), denote the pmf defined by (10) with \( C_i \) replaced by \( \vartheta_i \) for \( i = 0, 1, \ldots, \kappa \). Define \( M(x, \vartheta) = \frac{f(x, \vartheta)}{g(x, \vartheta)} \) for \( x \in X \) and \( \vartheta \in \Theta \) such that \( g(x, \vartheta) > 0 \). Define \( \mathcal{X} = \{ x \in X : f(x) > 0 \} \) and

\[ \mathcal{A} = \{ \vartheta \in \Theta : |\vartheta| < |\vartheta|, g(z, \vartheta) > 0 \}, \quad \mathcal{B} = \{ \vartheta \in \Theta : |\vartheta| > |\vartheta|, g(z, \vartheta) > 0 \}. \]

**Lemma 12** Assume that \( f(z) > 0 \). Then, \( f(X)^{\mathcal{A}}(X < z) \leq M(z, \vartheta)g(X, \vartheta) \) holds for all \( \vartheta \in \mathcal{A} \). Similarly, \( f(X)^{\mathcal{B}}(X > z) \leq M(z, \vartheta)g(X, \vartheta) \) holds for all \( \vartheta \in \mathcal{B} \).

**Proof.** First, we shall show that \( f(X)^{\mathcal{A}}(X < z) \leq M(z, \vartheta)g(X, \vartheta) \) holds for all \( \vartheta \in \mathcal{A} \). Since \( f(z) = g(z, C) \), the set \( \mathcal{A} \) is nonempty as a consequence of the assumption that \( f(z) > 0 \). Hence, \( M(z, \vartheta) \) is well-defined for \( z \in \mathcal{X} \) and \( \vartheta \in \mathcal{A} \). Note that \( f(x)^{\mathcal{A}}(X < z) \leq M(z, \vartheta)g(x, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( f(x) = 0 \). To show the first assertion, it suffices to show that \( f(x)^{\mathcal{A}}(X < z) \leq M(z, \vartheta)g(x, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \). We need to consider two cases as follows.

Case (I): \( N > 0 \).
Case (II): \( N < 0 \).

In Case (I), as a consequence of \( N > 0 \), we have \( \vartheta_i > 0, C_i > 0 \) for \( i = 0, 1, \ldots, \kappa \). Consider \( x \in \mathcal{X} \) such that \( x < z \). Note that \( \prod_{i=1}^k \prod_{\ell=1}^{x_i} (\vartheta_i - \ell + 1) \geq \prod_{i=1}^k \prod_{\ell=1}^{x_i} (\vartheta_i - \ell + 1) > 0 \) because of \( x < z \) and \( g(z, \vartheta) > 0 \). Hence, \( g(x, \vartheta) > 0 \) for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( x < z \). Since \( f(x) > 0 \), we have that

\[ M(x, \vartheta) = \frac{f(x)}{g(x, \vartheta)} = \prod_{i=1}^k \frac{(C_i)}{(\vartheta_i)} = \prod_{i=1}^k \frac{C_i - \ell + 1}{\vartheta_i - \ell + 1} \]

is well-defined and positive for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( x < z \). Hence, we have a meaningful ratio

\[ \frac{M(z, \vartheta)}{M(x, \vartheta)} = \prod_{i=1}^k \frac{C_i - \ell + 1}{\vartheta_i - \ell + 1} \]

for \( x \in \mathcal{X} \) such that \( x < z \). Since \( \vartheta_i \leq C_i \) for \( i = 1, \ldots, \kappa \), it follows from (84) that \( M(z, \vartheta) \leq M(z, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{X} \) such that \( x < z \).

In Case (II), as a consequence of \( N < 0 \), we have \( \vartheta_i < 0, C_i < 0 \) for \( i = 0, 1, \ldots, \kappa \). Note that

\[ M(x, \vartheta) = \prod_{i=0}^k \frac{C_i - \ell + 1}{\vartheta_i - \ell + 1} = \prod_{i=0}^k \frac{C_i + \ell - 1}{\vartheta_i + \ell - 1} > 0 \quad \text{for } x \in \mathcal{X} \]

and

\[ \frac{M(z, \vartheta)}{M(x, \vartheta)} = \prod_{i=1}^k \frac{C_i + \ell - 1}{\vartheta_i + \ell - 1} \]

for \( x \in \mathcal{X} \) such that \( x < z \). (85)

Since \( |\vartheta_i| \leq |C_i| \) for \( i = 1, \ldots, \kappa \), it follows from (84) that \( M(x, \vartheta) \leq M(z, \vartheta) \) holds for \( x \in \mathcal{X} \) such that \( \vartheta \in \mathcal{A} \) and \( x < z \). Therefore, in both cases, we have that

\[ \frac{f(x)}{g(x, \vartheta)} = M(x, \vartheta) \leq M(z, \vartheta) \]
holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{I} \) such that \( x \prec z \). Thus, we have shown that \( f(x)\mathbb{I}_{x\prec z} \leq \mathcal{M}(z, \vartheta)g(x, \vartheta) \) holds for \( \vartheta \in \mathcal{A} \) and \( x \in \mathcal{I} \). This completes the proof of the first assertion of the lemma.

The second assertion of the lemma can be shown in a similar manner. Since \( f(z) = g(z, C) \), the set \( B \) is nonempty as a consequence of the assumption that \( f(z) > 0 \). Hence, \( \mathcal{M}(z, \vartheta) \) is well-defined for \( z \in \mathcal{I} \) and \( \vartheta \in B \). Note that \( f(x)\mathbb{I}_{x\succ z} \leq \mathcal{M}(z, \vartheta)g(x, \vartheta) \) holds for \( \vartheta \in B \) and \( x \in \mathcal{I} \) such that \( f(x) = 0 \). To show the second assertion, it suffices to show that

\[
\mathcal{M}(z, \vartheta)g(x, \vartheta) \text{ holds for } \vartheta \in \mathcal{B} \text{ and } x \in \mathcal{I}.
\]

We need to consider two cases as follows.

Case (I): \( N > 0 \).

Case (II): \( N < 0 \).

In Case (I), as a consequence of \( N > 0 \), we have \( \vartheta_i > 0 \), \( C_i > 0 \) for \( i = 0, 1, \ldots, \kappa \). Consider \( x \in \mathcal{I} \) such that \( x \succ z \). Note that \( \prod_{i=1}^k \prod_{\ell=1}^{\ell_i}(\vartheta_i - \ell + 1) \geq \prod_{i=1}^k \prod_{\ell=1}^{\ell_i}(C_i - \ell + 1) \) because of \( x \succ z \) and \( f(x) > 0 \). Hence, \( g(x, \vartheta) > 0 \) for \( \vartheta \in B \) and \( x \in \mathcal{I} \) such that \( x \succ z \). Since \( f(x) > 0 \), we have that

\[
\mathcal{M}(x, \vartheta) = \frac{f(x)}{g(x, \vartheta)} = \frac{\prod_{i=0}^k \prod_{\ell=1}^{\ell_i} \frac{C_i - \ell + 1}{\vartheta_i - \ell + 1}}{\prod_{i=0}^k \prod_{\ell=1}^{\ell_i} \vartheta_i - \ell + 1}
\]

is well-defined and positive for \( \vartheta \in B \) and \( x \in \mathcal{I} \) such that \( x \succ z \). Hence, we have a meaningful ratio

\[
\frac{\mathcal{M}(x, \vartheta)}{\mathcal{M}(z, \vartheta)} = \frac{\prod_{i=1}^k \prod_{\ell=z_i+1}^{z_i} \frac{C_i + \ell - 1}{|\vartheta_i| + \ell - 1}}{\prod_{i=1}^k \prod_{\ell=1}^{\ell_i} \vartheta_i - \ell + 1}
\]

for \( x \in \mathcal{I} \) such that \( x \succ z \). Since \( \vartheta_i \geq C_i \) for \( i = 1, \ldots, \kappa \), it follows from (86) that \( \mathcal{M}(x, \vartheta) \leq \mathcal{M}(z, C) \) holds for \( \vartheta \in B \) and \( x \in \mathcal{I} \) such that \( x \succ z \).

In Case (II), as a consequence of \( N < 0 \), we have \( \vartheta_i < 0 \), \( C_i < 0 \) for \( i = 0, 1, \ldots, \kappa \). Note that

\[
\mathcal{M}(x, \vartheta) = \frac{\prod_{i=0}^k \prod_{\ell=1}^{\ell_i} \frac{C_i - \ell + 1}{\vartheta_i - \ell + 1}}{\prod_{i=1}^k \prod_{\ell=z_i+1}^{z_i} \vartheta_i - \ell + 1}
\]

and

\[
\frac{\mathcal{M}(x, \vartheta)}{\mathcal{M}(z, \vartheta)} = \frac{\prod_{i=1}^k \prod_{\ell=1}^{\ell_i} \frac{|C_i| + \ell - 1}{|\vartheta_i| + \ell - 1}}{\prod_{i=1}^k \prod_{\ell=z_i+1}^{z_i} \vartheta_i - \ell + 1}
\]

for \( x \in \mathcal{I} \) such that \( x \succ z \). Since \( |\vartheta_i| \geq |C_i| \) for \( i = 1, \ldots, \kappa \), it follows from (87) that \( \mathcal{M}(x, \vartheta) \leq \mathcal{M}(z, \vartheta) \) holds for \( \vartheta \in B \) and \( x \in \mathcal{I} \) such that \( x \succ z \). Therefore, in both cases, we have that

\[
\frac{f(x)}{g(x, \vartheta)} = \mathcal{M}(x, \vartheta) \leq \mathcal{M}(z, \vartheta)
\]

holds for \( \vartheta \in B \) and \( x \in \mathcal{I} \) such that \( x \succ z \). Thus, we have shown that \( f(x)\mathbb{I}_{x\succ z} \leq \mathcal{M}(z, \vartheta)g(x, \vartheta) \) holds for \( \vartheta \in B \) and \( x \in \mathcal{I} \). This completes the proof of the second assertion of the lemma.

Applying Theorem \ref{thm:main} and Lemma \ref{lem:main} we have the following results.

**Lemma 13** Assume that \( f(z) > 0 \). Then,

\[
\Pr\{X \prec z\} \leq \prod_{i=0}^k \frac{C_i}{\vartheta_i z_i} \text{ for } \vartheta \in \mathcal{A},
\]

\[
\Pr\{X \succ z\} \leq \prod_{i=0}^k \frac{C_i}{\vartheta_i z_i} \text{ for } \vartheta \in \mathcal{B}.
\]

**Lemma 14** \( g(z, \hat{C}) > 0 \) for \( z \in \mathcal{I} \).
**Proof.** Recall that \( \hat{\mathbf{C}} = [\hat{C}_0, \hat{C}_1, \ldots, \hat{C}_\kappa]^T \) with \( \hat{C}_i = \mathcal{N} \frac{\mathbf{X}}{\hat{\vartheta}^i} \) for \( i = 0, 1, \ldots, \kappa \).

In the case of \( \mathcal{N} < 0 \), we have \( \hat{C}_i \leq 0 \) for \( i = 0, 1, \ldots, \kappa \), which implies that \( 1 + \hat{C}_i \leq 1 \) for \( i = 0, 1, \ldots, \kappa \). Hence, \( g(z, \hat{\mathbf{C}}) > 0 \) for \( z \in \mathcal{X} \) if \( \mathcal{N} < 0 \).

In the case of \( \mathcal{N} > 0 \), as a consequence of \( \frac{n - 1}{\mathcal{N}} < 1 \), we have \( 0 < n < \mathcal{N} + 1 \) and \( \hat{C}_i \geq 0 \) for \( i = 0, 1, \ldots, \kappa \). For \( i = 0, 1, \ldots, \kappa \), we have \( z_i < \hat{C}_i + 1 = 1 \) if \( z_i = 0 \); and \( z_i = \frac{\hat{C}_i}{\mathcal{N}} \hat{C}_i + \frac{\hat{C}_i}{\mathcal{N}} \hat{C}_i + \frac{\hat{C}_i}{n} \leq \hat{C}_i + 1 \) if \( z_i > 0 \). Hence, \( g(z, \hat{\mathbf{C}}) > 0 \) for \( z \in \mathcal{X} \) if \( \mathcal{N} > 0 \).

\( \square \)

**Lemma 15** \( |\hat{\mathbf{C}}| < |\mathbf{C}| \Leftrightarrow z < \hat{\mu} \) and \( \hat{\mathbf{C}} \succ |\mathbf{C}| \Leftrightarrow z \succ \hat{\mu} \).

**Proof.** In the case of \( \mathcal{N} > 0 \), we have \( \mathbf{C}_i > 0 \) and \( \hat{C}_i \geq 0 \) for \( i = 0, 1, \ldots, \kappa \). Hence,

\[
|\hat{\mathbf{C}}| < |\mathbf{C}| \Leftrightarrow \hat{\mathbf{C}} < \mathbf{C} \Leftrightarrow z < \hat{\mu}, \quad |\hat{\mathbf{C}}| > |\mathbf{C}| \Leftrightarrow \hat{\mathbf{C}} > \mathbf{C} \Leftrightarrow z > \hat{\mu}.
\]

In the case of \( \mathcal{N} < 0 \), we have \( \mathbf{C}_i < 0 \) and \( \hat{C}_i \leq 0 \) for \( i = 0, 1, \ldots, \kappa \). Hence,

\[
|\hat{\mathbf{C}}| < |\mathbf{C}| \Leftrightarrow \hat{\mathbf{C}} > \mathbf{C} \Leftrightarrow z < \hat{\mu}, \quad |\hat{\mathbf{C}}| > |\mathbf{C}| \Leftrightarrow \hat{\mathbf{C}} < \mathbf{C} \Leftrightarrow z > \hat{\mu}.
\]

\( \square \)

We are now in a position to prove the theorem. Clearly, the assertions (I) and (II) of Theorem 11 follow from Lemmas 13, 14 and 15. The assertions (III) and (IV) of Theorem 11 can be shown by the same arguments as that of the counterparts of Theorem 10.

**G Proof of Theorem 12**

Define \( \vartheta_0 = \alpha_0 \),

\[
\Theta = \{ [\vartheta_0, \vartheta_1, \ldots, \vartheta_\kappa]^T : 0 < \vartheta_i \leq \alpha_i, \ i = 1, \ldots, \kappa \}
\]

and

\[
\rho(x, \vartheta) = \ln \frac{B(\vartheta)}{B(\alpha)} + \sum_{i=1}^{\kappa} (\alpha_i - \vartheta_i) \ln x_i.
\]

It can be seen that \( \rho(x, \vartheta) \) is a concave function of \( x \) such that \( \rho(x, \vartheta) \leq \rho(z, \vartheta) \) for all \( \vartheta \in \Theta \) provided that \( x \prec z \). Moreover, \( f(X, \alpha) = \exp(\rho(X, \vartheta))f(X, \vartheta) \) for all \( \vartheta \in \Theta \). Making use of these facts and Theorem 2, we have

\[
\Pr(\mathbf{X}_n \prec z) \leq \inf_{\vartheta \in \Theta} \left[ \frac{B(\vartheta)}{B(\alpha)} \prod_{i=1}^{\kappa} \frac{z_i^{\alpha_i}}{\vartheta_i^{\alpha_i}} \right]^n.
\]

To derive an explicit bound, we shall use the method of moments by seeking \( \vartheta \) such that \( z = E_{\vartheta} [X] \). From this equation, we have

\[
z_i = \frac{\vartheta_i}{\sum_{\ell=0}^{\kappa} \vartheta^\ell}, \quad i = 1, \ldots, \kappa,
\]

which implies that \( \frac{\vartheta_i}{z_i} \) are equal for all \( i = 1, \ldots, \kappa \). It follows that

\[
z_i = \frac{1}{\vartheta_i} + \sum_{\ell=0}^{\kappa} \frac{\vartheta^\ell}{\vartheta_i^\ell} = \frac{1}{\vartheta_i} + \sum_{\ell=0}^{\kappa} \frac{\vartheta^\ell}{z_i^\ell}, \quad i = 1, \ldots, \kappa,
\]

36
from which we obtain
\[
\frac{\alpha_0}{\vartheta_i} = \frac{1}{z_i} - \sum_{\ell=0}^{\kappa} \frac{z_\ell}{z_i}, \quad i = 1, \ldots, \kappa.
\]
Hence,
\[
\vartheta_i = \frac{\alpha_0 z_i}{1 - \sum_{\ell=1}^{\kappa} z_\ell} = \frac{\alpha_0 z_i}{z_0} = \alpha_i, \quad i = 1, \ldots, \kappa.
\]
Noting that \(z_0 \geq \mu_0 = \frac{\alpha_0}{\sum_{\ell=0}^{\kappa} \alpha_\ell}\), we have \(\frac{\alpha_0}{z_0} \leq \sum_{\ell=0}^{\kappa} \alpha_\ell\) and thus
\[
\alpha_i = \vartheta_i \leq z_i \sum_{\ell=0}^{\kappa} \alpha_\ell \leq \mu_i \sum_{\ell=0}^{\kappa} \alpha_\ell = \alpha_i, \quad i = 1, \ldots, \kappa.
\]
This establishes \(46\). The proof of the theorem is thus completed.

H Proof of Theorem \(13\)

Define
\[
g(x, \vartheta) = |\vartheta|^{-\alpha} |x|^{-(p+1)/2} \exp \left( -\frac{1}{\beta} \text{tr}(\vartheta^{-1} x) \right),
\]
where \(\vartheta \succ \Sigma\). Define
\[
\rho(x, \vartheta) = \alpha \ln \frac{|\vartheta|}{|\Sigma|} - \frac{1}{\beta} \text{tr}(\Sigma^{-1} - \vartheta^{-1})|x|.
\]
Then, \(\rho(x, \vartheta) = \frac{f(x)}{g(x, \vartheta)}\) is a linear function of \(x\) such that \(f(X) = \exp(\rho(X, \vartheta)) g(X, \vartheta)\). Clearly, \(\Sigma^{-1} \succ \vartheta^{-1}\). Let \(A = \sqrt{\Sigma^{-1} - \vartheta^{-1}}\) and \(B = \sqrt{y - x}\). Note that
\[
\text{tr}\left[ (\Sigma^{-1} - \vartheta^{-1}) (x - z) \right] = \text{tr}(AABB) = \text{tr}(BAAB) = \text{tr}\left[ (AB)^\top (AB) \right] \geq 0
\]
for \(x \succ z\). It follows that \(\rho(x, \vartheta) \leq \rho(z, \vartheta)\) for \(x \succ z\). Making use of these facts and Theorem \(2\) we have
\[
\Pr\{X_n \succ z\} \leq \inf_{\vartheta \succ \Sigma} \left[ \frac{f(z)}{g(z, \vartheta)} \right]^n.
\]
Now let \(\vartheta = \frac{\alpha}{\alpha_0} \Sigma\). Since \(z \succ \alpha \Sigma\), we have \(\vartheta = \frac{\alpha}{\alpha_0} \succ \Sigma\). Note that
\[
\frac{f(z)}{g(z, \frac{\alpha}{\alpha_0} \Sigma)} = \left( \frac{\alpha}{\alpha_0} \right)^{p\alpha} \left( \frac{|z|}{|\Sigma|} \right)^\alpha \exp \left( -\frac{1}{\beta} \text{tr}(\Sigma^{-1} z) \right).
\]
It follows from \(88\) that
\[
\Pr\{X_n \succ z\} \leq \left[ \left( \frac{\alpha}{\alpha_0} \right)^{p\alpha} \left( \frac{|z|}{|\Sigma|} \right)^\alpha \exp \left( -\frac{1}{\beta} \text{tr}(\Sigma^{-1} z) \right) \right]^n \quad \text{for } z \succ \alpha \Sigma.
\]
Taking \(z = \rho_0 \alpha \Sigma\), we have
\[
\Pr\{X \succ \rho_0 \alpha \Sigma\} \leq \left[ \left( \frac{\alpha}{\alpha_0} \right)^{p\alpha} \left( \frac{|\rho_0 \alpha \Sigma|}{|\Sigma|} \right)^\alpha \exp \left( -\frac{1}{\beta} \text{tr}(\Sigma^{-1} \rho_0 \alpha \Sigma) \right) \right]^n = \left[ \rho \exp(1 - \varrho) \right]^{p\alpha}
\]
for \(\varrho \geq 1\). In a similar manner, we can show that
\[
\Pr\{X_n \prec z\} \leq \left[ \left( \frac{\alpha}{\alpha_0} \right)^{p\alpha} \left( \frac{|z|}{|\Sigma|} \right)^\alpha \exp \left( -\frac{1}{\beta} \text{tr}(\Sigma^{-1} z) \right) \right]^n \quad \text{for } z \prec \alpha \Sigma,
\]
and \(\Pr\{X \prec \rho_0 \alpha \Sigma\} \leq \left[ \rho \exp(1 - \varrho) \right]^{p\alpha}\) for \(\varrho \in (0, 1]\). This completes the proof of the theorem.
References

[1] S. Bernstein, *Theory of Probability*, Moscow, 1927.

[2] G. Bennett, “Probability inequalities for the sum of independent random variables,” *J. Amer. Statist. Assoc.*, vol. 57, pp. 33–35, 1962.

[3] A. C. Berry, “The accuracy of the Gaussian approximation to the sum of independent variates,” *Trans. Amer. Math. Soc.*, vol. 49, no. 1, pp. 122–139, 1941.

[4] X. Chen, “A new framework of multistage estimation,” [arXiv:0809.1241v20] [math.ST], August 2010.

[5] X. Chen, “A new framework of multistage hypothesis tests,” [arXiv:0809.3170v15] [math.ST], August 2010.

[6] X. Chen, “An urn model approach for deriving multivariate generalized hypergeometric distributions,” submitted for publication, 2013.

[7] X. Chen, “New probabilistic inequalities from monotone likelihood ratio property,” [arXiv:1010.3682v1] [math.PR], October 2010.

[8] H. Chernoff, “A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations,” *Ann. Math. Statist.*, vol. 23, pp. 493–507, 1952.

[9] P. C. Consul, *Generalized Poisson Distribution*, Dekker, 1989.

[10] C. G. Esseen, “On the Liapunoff limit of error in the theory of probability,” *Ark. Mat. Astron. Fys.*, vol. A28, no. 9, pp. 1-19, 1942.

[11] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, Chapman and Hall/CRC, 1999.

[12] W. Hoeffding, “Probability inequalities for sums of bounded variables,” *J. Amer. Statist. Assoc.*, vol. 58, pp. 13–29, 1963.

[13] C. Mortici and C. P. Chen, “New sharp double inequalities for bounding the Gamma and Digamma function,” *Analele Universității de Vest, Timișoara Seria Matematică – Informatică*, vol. XLIX, No. 2, pp. 69–75, 2011.

[14] V. V. Petrov, “On the probabilities of large deviations for sums of independent random variables,” *Theory of Probability and Its Applications*, vol. X, No. 2, pp. 287–298, 1965.

[15] H. S. Steyn, “On discrete multivariate probability functions,” *Proceedings, Koninklijke Nederlandse Akademie Van Wetenschappen, Series A*, vol. 54, pp. 23–30, 1951.

[16] Ilya Tyurin, “An improvement of upper estimates of the constants in the Lyapunov theorem,” *Russian Mathematical Surveys*, vol. 65, pp. 201–202, 2010.