Improving Bennett’s and Hoeffding’s inequalities using higher moments information

Bar Light*

June 20, 2020

Abstract:

We derive upper bounds on the moment-generating function of a random variable that depends on the random variable’s first $p$ moments. We use these bounds to generalize and improve the classical Hoeffding’s and Bennett’s inequalities for the case where there is some information on the random variables’ first $p$ moments for every positive integer $p$. Our generalized Hoeffding’s inequality is tighter than Hoeffding’s inequality and is given in a simple closed-form expression for every positive integer $p$. Hence, the generalized Hoeffding’s inequality is easy to use in applications. Our generalized Bennett’s inequality is given in terms of the generalized Lambert $W$-function and is tighter than Bennett’s inequality.

Keywords: Concentration inequalities, Hoeffding’s inequality, Bennett’s inequality.

*Graduate School of Business, Stanford University, Stanford, CA 94305, USA. e-mail: barl@stanford.edu
1 Introduction

Concentration inequalities provide bounds on the probability that a random variable differs from some value, typically the random variable’s expected value (see Boucheron et al. (2013) for a textbook treatment of concentration inequalities). Besides their importance in probability theory, concentration inequalities are an important mathematical tool in statistics (see Massart (2000)), machine learning theory (see Mohri et al. (2018)) and many other fields. Two of the most important and useful concentration inequalities are Hoeffding’s inequality (Hoeffding, 1994) and Bennett’s inequality (Bennett, 1962). These are inequalities that bound the probability that the sum of independent random variables differs from its expected value. The bound derived in Hoeffding’s inequality holds for bounded random variables and uses information on the random variables’ first moment. The bound derived in Bennett’s inequality holds for random variables that are bounded from above and uses information on the random variables’ first and second moments.

In this paper we generalize and significantly improve Bennett’s and Hoeffding’s inequalities. We provide bounds that use information on the random variables’ higher moments. More precisely, we provide bounds on the probability that the sum of independent random variables differs from its expected value where the bounds depend on the random variables’ first $p$ moments for every integer $p \geq 1$. We provide two families of concentration inequalities, one that generalizes Hoeffding’s inequality and one that generalizes Bennett’s inequality. Importantly, the bounds that we derive are tighter than Bennett’s and Hoeffding’s inequalities and are given as closed-form expressions in most cases. In our generalized Hoeffding’s inequality, our bounds hold for bounded random variables and are given as simple closed-form expressions (see Theorem 2) for every integer $p \geq 1$. For any integer $p \geq 1$ the bound uses information on the random variables’ first $p$ moments and is always tighter than Hoeffding’s inequality. We also show, using a numerical example, that our bound can significantly improve Hoeffding’s inequality. In our generalized Bennett’s inequality, our bounds hold for random variables that are bounded from above. For $p = 3$, our bound is given in a closed-form expression in terms of the Lambert $W$-function. This bound uses information on the random variables’ first three moments and is tighter than Bennett’s inequality. For $p > 3$ our

\[^{1}\text{There are many extensions and generalizations of Hoeffding’s and Bennett’s inequalities. For example see Freedman (1975), Pinelis (1994), Talagrand (1995), Roussas (1996), Cohen et al. (1999), Victor et al. (1999), Bousquet (2002), Bentkus et al. (2004), Klein et al. (2005), Kontorovich et al. (2008), Fan et al. (2012), Junge et al. (2013), Pinelis (2014), and Pelekis et al. (2015).}\]
bounds are given in terms of the generalized Lambert W-function (see Theorem 3).

For every positive integer $p$, independent random variables $X_1, \ldots, X_n$ such that $\mathbb{P}(X_i \in [a_i, b_i] = 1)$, and all $t > 0$, our generalized Hoeffding’s inequality is given by

$$
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( -\frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2 C_p(t, X_i)} \right)
$$

where $S_n = \sum_{i=1}^{n} X_i$ and $C_p(t, X_i)$ is a function that depends on $t$, on the first $p$ moments of $X_i$, and on $X_i$’s support: $[a_i, b_i]$. We show that for every positive integer $p$ we have $C_p \leq 1$. Thus, our generalized Hoeffding’s inequality is tighter than Hoeffding’s inequality which corresponds to $p = 1$ and $C_1 = 1$. We provide a simple closed-form expression for the function $C_p$ for any integer $p \geq 1$. For example, suppose that the support of a random variable $X$ is $[0, b]$ for some $X = X_i$, $i = 1, \ldots, n$. Then $C_p(t, X)$ is given by

$$
C_p(t, X) = \left( \frac{\mathbb{E}^{\mathbb{P}} \exp(y) + \sum_{j=0}^{p-3} \frac{y^j}{j!} (b^{p-j-2} \mathbb{E}X^{j+2} - \mathbb{E}X^p)}{\mathbb{E}^{\mathbb{P}} \exp(y) + \sum_{j=0}^{p-2} \frac{y^j}{j!} (b^{p-j-1} \mathbb{E}X^{j+1} - \mathbb{E}X^p)} \right)^2
$$

where $y = 4tb / \sum_{i=1}^{n} d(X_i)$ and $d(X_i) = (\mathbb{E}X_i^2 / \mathbb{E}X_i)^2$ (see Theorem 2). We note that our generalized Hoeffding’s bounds are exponential bounds, and hence, these bounds are not optimal in the sense that there is a missing factor in those bounds (see Talagrand (1995)). However, in many applications it is convenient to use exponential bounds that are given in closed-form expressions such as the bound given in Hoeffding’s inequality and in our generalized Hoeffding’s inequality. Therefore, we believe that our bounds will be useful for future research and applications.

2 Main results

In this section we state our main results. In Section 2.1 we derive upper bounds on the moment-generating function of a random variable that is bounded from above. In Section 2.2 we derive our generalized Hoeffding’s inequalities. In Section 2.3 we derive our generalized Bennett’s inequalities.

We first introduce some notations.

Throughout the paper we consider a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A random variable $X$ is a measurable real-valued function from $\Omega$ to $\mathbb{R}$. We denote the expectation of a random variable
on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ by $\mathbb{E}$. For $1 \leq p \leq \infty$ let $L^p := L^p(\Omega, \mathcal{F}, \mathbb{P})$ be the space of all random variables $X : \Omega \to \mathbb{R}$ such that $\|X\|_p$ is finite, where $\|X\|_p = (\int_\Omega |X(\omega)|^p \mathbb{P}(d\omega))^{1/p}$ for $1 \leq p < \infty$ and $\|X\|_\infty = \sup_{\omega \in \Omega} |X(\omega)|$ for $p = \infty$. We say that $X$ is a random variable on $[a, b]$ for some $a < b$ if $\mathbb{P}(X \in [a, b]) = 1$.

For $k \geq 1$, we denote by $f^{(k)}$ the $k$th derivative of a $k$ times differentiable function $f : [a, b] \to \mathbb{R}$ and for $k = 0$ we define $f^{(0)} := f$. As usual, the derivatives at the extreme points $f^{(k)}(a)$ and $f^{(k)}(b)$ are defined by taking the left-side and right-side limits, respectively.

For the rest of the paper we define

$$T_p(x) := \exp(x) - \sum_{j=0}^{p-2} \frac{x^j}{j!}$$

to be the Taylor remainder of the exponential function of order $p - 2$ at the point 0. The function $T_p$ plays an important role in our analysis.

### 2.1 Upper bounds on the moment-generating function

In this section we provide upper bounds on the moment-generating function of a random variable that is bounded from above.

We show that

$$\frac{T_{p+1}(x)}{T_{p+1}(b)} \leq \frac{\max(x^p, 0)}{b^p}$$

(1)

for all $x \leq b$, $b > 0$ and every positive integer $p$. This bound on the ratio of the Taylor remainders is the key ingredient in deriving the upper bounds on the moment-generating function. The proof of Bennett’s inequality uses inequality (1) with $p = 2$ to bound the moment-generating function (see Boucheron et al. (2013)). We use inequality (1) to provide upper bounds on the moment-generating function using information on the random variable’s first $p$ moments for every positive integer $p$.

The Appendix contains the proofs not presented in the main text.

**Theorem 1** Let $X \in L^{p-1}$ be a random variable on $(-\infty, b]$ for some $b > 0$ where $p$ is a positive
integer. For all \( s \geq 0 \) we have

\[
\mathbb{E} \exp(sX) \leq \frac{\mathbb{E} \max(X^p, 0)}{b^p} \left( \exp(sb) - \sum_{j=0}^{p-1} \frac{s^j b^j}{j!} \right) + \mathbb{E} \left( \sum_{j=0}^{p-1} \frac{s^j X^j}{j!} \right)
\]

(2)

Theorem 1 provides a unified approach for seemingly independent bounds on the moment-generating function that were derived in previous literature and used to prove concentration inequalities.

For \( p = 2 \), and for a random variable \( X \) on \((-\infty, b]\), Theorem 1 yields the inequality

\[
\mathbb{E} \exp(sX) \leq \frac{\mathbb{E} X^2}{b^2} (\exp(sb) - 1 - sb) + 1 + s\mathbb{E}(X)
\]

which is fundamental in proving Bennett’s inequality (see Bennett (1962)). For \( p = 3 \), denoting \( \mu^3 = \mathbb{E} \max(X^3, 0) \), we have

\[
\frac{\mu^3}{b^3} T_4(sb) + \mathbb{E} \left( \sum_{j=0}^2 \frac{s^j X^j}{j!} \right) = \frac{\mu^3}{b^3} T_3(sb) + 1 + s\mathbb{E}(X) + \frac{s^2}{2} \left( \mathbb{E} X^2 - \frac{\mu^3}{b} \right)
\]

\[
\leq \exp \left( \frac{\mu^3}{b^3} T_3(sb) + s\mathbb{E}(X) + \frac{s^2}{2} \left( \mathbb{E} X^2 - \frac{\mu^3}{b} \right) \right)
\]

The last inequality follows from the elementary inequality \( 1 + x \leq \exp(x) \) for all \( x \in \mathbb{R} \). Thus, Theorem 1 implies

\[
\mathbb{E} \exp(sX) \leq \exp \left( \frac{\mathbb{E} \max(X^3, 0)}{b^3} T_3(sb) + s\mathbb{E}(X) + \frac{s^2}{2} \left( \mathbb{E} X^2 - \frac{\mathbb{E} \max(X^3, 0)}{b^3} \right) \right)
\]

which is proved in Theorem 2 in Pinelis and Utev (1990).

Let

\[
m_X(p) := \frac{\mathbb{E} \max(X^p, 0)}{b^p} T_{p+1}(sb) + \mathbb{E} \left( \sum_{j=0}^{p-1} \frac{s^j X^j}{j!} \right)
\]

be the right-hand side of inequality (2). The next proposition shows that for every even number \( p \) we have \( m_X(p) \geq m_X(p + 1) \). If, in addition, the random variable \( X \) is non-negative, then we
also have \( m_X(p+1) \geq m_X(p+2) \), and hence, \( m_X(p) \) is decreasing. Thus, for non-negative random variables, inequality (2) is tighter when \( p \) increases. In particular, we have \( m_X(2) \geq m_X(p) \) for every integer \( p \geq 3 \), i.e., the bound on the moment-generating function given in inequality (2) is tighter than Bennett’s bound (3) for every integer \( p \geq 3 \) when \( X \) is non-negative.

**Proposition 1** Let \( X \in L^p \) be a random variable on \( (-\infty, b] \). Let \( p \geq 2 \) be an even number. The following statements hold:

(i) \( m_X(p) \geq m_X(p + 1) \).

(ii) If \( X \geq 0 \) then \( m_X(p + 1) \geq m_X(p + 2) \).

The upper bound on the moment-generating function (2) is optimal in the sense that there exists a random variable that achieves equality. Even for \( p = 1 \) there exists a random variable that achieves equality in (2). For example, a Bernoulli random variable that yields 1 with probability \( q \) and 0 with probability \( 1 - q \) achieves equality in (2) for \( p = 1 \). Note that for the Bernoulli random variable all the moments are equal to \( q \) which is the highest value that the higher moments can have given that the first moment equals \( q \) and the support is \([0, 1]\). Thus, higher moments do not provide any useful information and for every integer \( p > 1 \) inequality (2) reduces to the case of \( p = 1 \).

For any random variable \( X \) on \([0, b]\) and \( s \geq 0 \) Theorem 1 implies that

\[
\mathbb{E} T_{p+1}(sX) \leq \frac{\mathbb{E} X^p}{b^p} T_{p+1}(sb)
\]

where \( p \) is a positive integer. The last inequality can be easily applied to any bounded random variable. For example, suppose that \( Y \) is a random variable on \([a, b]\) for some \( a < b \). Defining the random variable \( X = Y - a \) and using inequality (2) yields the following upper bound on the moment-generating function:

\[
\mathbb{E} \exp(s(Y - a)) \leq \frac{\mathbb{E}(Y - a)^p}{(b - a)^p} \left( \exp(s(b - a)) - \sum_{j=0}^{p-1} \frac{s^j(b - a)^j}{j!} \right) + \mathbb{E} \left( \sum_{j=0}^{p-1} \frac{s^j(Y - a)^j}{j!} \right).
\]
2.2 Concentration inequalities: Hoeffding type inequalities

In this section we derive Hoeffding type concentration inequalities that provide exponential bounds on the probability that the sum of independent bounded random variables differs from its expected value. We improve Hoeffding’s inequality by using information on the random variables’ first \( p \) moments and by using a refined upper bound on the moment-generating function of a bounded random variable (see Theorem 1). We derive a tighter bound than the standard Hoeffding’s bound for every integer \( p \geq 2 \) (see Theorem 2 part (ii)). Importantly, for every \( p \) the bound is given as a simple closed-form expression that depends on the random variables’ first \( p \) moments. Thus, the bound can be easily used in applications. We also provide a bound for the case that \( p \) tends to infinity. This bound depends on all of the random variables’ moments (see Theorem 2 part (iii)).

**Theorem 2** Let \( X_1, \ldots, X_n \) be independent random variables where \( X_i \) is a random variable on \([0, b_i], b_i > 0\). Let \( S_n = \sum_{i=1}^{n} X_i \). Let \( p \geq 1 \) be an integer. Denote \( \mathbb{E}(X_i^k) = \mu_i^k > 0 \) for all \( k = 1, \ldots, p \) and all \( i = 1, \ldots, n \). Let \( D_n = \sum_{i=1}^{n} d(X_i) \) where \( d(X_i) = \left( \frac{\mu_i^2}{\mu_i^1} \right)^2 \).

(i) For all \( t > 0 \) we have

\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( -\frac{2t^2}{\sum_{i=1}^{n} b_i^2 C_p \left( \frac{\mu_i^p}{\mu_i^0}, \frac{\mu_i^1}{\mu_i^1}, \ldots, \frac{\mu_i^p}{\mu_i^p} \right)} \right),
\]

where

\[
C_p \left( y, b_i, \mu_i^1, \ldots, \mu_i^p \right) = \left( \frac{\exp(y) + \sum_{j=0}^{p-3} \frac{y^j}{j!} \left( \frac{\mu_i^{p-j} - \mu_i^{j+2}}{\mu_i^1} - 1 \right)}{\exp(y) + \sum_{j=0}^{p-2} \frac{y^j}{j!} \left( \frac{\mu_i^{p-j-1} - \mu_i^{j+1}}{\mu_i^1} - 1 \right)} \right)^2
\]

for all \( i = 1, \ldots, n \) and all \( y > 0 \).

(ii) For every integer \( p \geq 1 \) we have \( 0 < C_p \leq 1 \). Thus, inequality (4) is tighter than Hoeffding’s inequality:

\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( -\frac{2t^2}{\sum_{i=1}^{n} b_i^2} \right)
\]

which corresponds to \( p = 1 \) and \( C_1 = 1 \).

(iii) When \( p \) tends to infinity we have

\[
\lim_{p \to \infty} C_p(x, b_i, \mu_i^1, \ldots, \mu_i^p) = \frac{1}{b_i^2} \left( \frac{\mathbb{E}X_i^2 \exp(xX_i/b_i)}{\mathbb{E}X_i \exp(xX_i/b_i)} \right)^2
\]
for all $i = 1, \ldots, n$ and all $x \geq 0$. Using part (i) implies that for all $t > 0$ we have

$$
\Pr(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( - \frac{2t^2}{\sum_{i=1}^{n} \left( \frac{b_i}{d(X_i)} \right) \left( \frac{2\mu_i}{D_n} \right)} \right). \tag{7}
$$

Note that inequality (7) does not depend on $b_i$.

**Remark 1** (i) Theorem 2 can be easily applied to bounded random variables that are not necessarily positive. If $Y_i$ is a random variable on $[a_i, b_i]$ and $Y_1, \ldots, Y_n$ are independent, we can define the random variables $X_i = Y_i - a_i$ on $[0, b_i - a_i]$ and use Theorem 2 to conclude that

$$
\Pr \left( \sum_{i=1}^{n} Y_i - \mathbb{E} \left( \sum_{i=1}^{n} Y_i \right) \geq t \right) = \Pr(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( - \frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2 C_p \left( \frac{4t(b_i - a_i)/D_n}{b_i - a_i, \mu_i, \ldots, \mu_p} \right)} \right).
$$

Note that $\mu_i = \mathbb{E}(Y_i - a_i)^k$ for all $i = 1, \ldots, n$ and all $k = 1, \ldots, p$.

(ii) If $X_1, \ldots, X_n$ are identically distributed then inequality (4) yields

$$
\Pr(S_n - \mathbb{E}(S_n) \geq nt) \leq \exp \left( - \frac{2nt^2}{b_i^2 C_p \left( \frac{4tb_i}{d(X_i)} \right)} \right). \tag{8}
$$

We now discuss the sketch of the proof of Theorem 2 part (i). The full proof is in the Appendix.

Fix a positive integer $p$. We start with a random variable $X$ on $[0, b]$. Assume for simplicity that $b = 1$. From Theorem 1 we have $\mathbb{E} \exp(sX) \leq v(y)$ where $v(y)$ is the right-hand side of inequality (2). Let $g(y) = \ln(v(y))$. Then using Taylor’s theorem we can show (using a standard argument) that $\mathbb{E} \exp(sX - s\mathbb{E}(X)) \leq \exp(0.5 s^2 \max_{0 \leq y \leq s} g^{(2)}(y))$. Let $C(s) = \max_{0 \leq y \leq s} (v^{(2)}(y)/v^{(1)}(y))^2$. A key step in the proof of Theorem 2 is to show that $v^{(1)}$ is a log-convex function, and hence, $v^{(2)}/v^{(1)}$ is increasing which implies that $C(s) = (v^{(2)}(s)/v^{(1)}(s))^2$ is given in a closed-form expression. We have

$$
\max_{0 \leq s \leq s} g^{(2)}(y) = \max_{0 \leq y \leq s} \frac{v^{(2)}(y)}{v(y)} \left( 1 - \frac{v^{(2)}(y)}{v(y)} \frac{v^{(1)}(y)}{v(y)} \right)^2 \leq \max_{0 \leq y \leq s} \frac{v^{(2)}(y)}{v(y)} \left( 1 - \frac{v^{(2)}(y)}{v(y)} C(s) \right) \leq 0.25 C(s)
$$

where the second inequality follows from the elementary inequality $x(1-x/z) \leq 0.25z$ for all $z > 0$. 

8
and $x > 0$. With this bound we can conclude that $\mathbb{E} \exp(sX - s\mathbb{E}(X)) \leq \exp(s^2C(s)/8)$. Applying the Chernoff bound and choosing a specific value for $s$ proves Theorem 1 part (i).

The calculation of $C_p$ in inequality (4) is immediate. For example, for $p = 2$ we have

$$C_2(x, b_i, \mu_1^i, \mu_2^i) = \left( \frac{\mu_1^2 \exp(x)}{\mu_2^2 \exp(x) + b_i \mu_1^i - \mu_2^i} \right)^2$$

and for $p = 3$ we have

$$C_3(x, b_i, \mu_1^i, \mu_2^i, \mu_3^i) = \left( \frac{\mu_3^2 \exp(x) + b_i \mu_2^i - \mu_3^i}{\mu_3^2 \exp(x) + b_i \mu_2^i - \mu_3^i + (b_i \mu_2^i - \mu_3^i)x} \right)^2$$

for all $i = 1, \ldots, n$ and all $x \geq 0$.

We now provide a numerical example where the results in Theorem 2 significantly improve Hoeffding’s inequality.

**Example 1** (i) Suppose that $X_1, \ldots, X_n$ are independent continuous uniform random variables on $[0, 1]$, i.e., $\mathbb{P}(X_i \leq t) = t$ for $0 \leq t \leq 1$. In this case, a straightforward calculation shows that

$$\mathbb{E}X_i \exp(sX_i) = \frac{\exp(s)(s-1) + 1}{s^2} \quad \text{and} \quad \mathbb{E}X_i^2 \exp(sX_i) = \frac{\exp(s)(s^2 - 2s + 2) - 2}{s^3}.$$

Using Theorem 2 part (iii) we have

$$\lim_{p \to \infty} C_p(x, 1, \mu_1^i, \ldots, \mu_p^i) = \left( \frac{-2 + \exp(x)(2 - 2x + x^2)}{x(1 + \exp(x)(x-1))} \right)^2 = C_\infty(x).$$

Using the fact that $d(X_i) = (\mathbb{E}X_i^2/\mathbb{E}X_i)^2 = 4/9$ inequality (7) yields

$$\mathbb{P}(S_n - \mathbb{E}(S_n) \geq nt) \leq \exp \left( -\frac{2nt^2}{C_\infty(9t)} \right). \quad (9)$$

In Figure 1 we plot the bound given in Hoeffding’s inequality (see Theorem 2 inequality (6)) for $\mathbb{P}(S_n - \mathbb{E}(S_n) \geq nt)$ divided by the bound given in (9) as a function of $t$ on the interval $[0.1, 0.4]$ for $n = 40$. We see that the bound given in (9) significantly improves Hoeffding’s bound.
Figure 1: Comparing Hoeffding’s inequality and inequality (9): $X_i$ is a uniform random variable on $[0, 1]$ for $i = 1, \ldots, 40$. The plot describes the ratio of the right-hand side of inequality (6) (Hoeffding’s inequality) to the right-hand side of inequality (9) for $n = 40$.

2.3 Concentration inequalities: Bennett type inequalities

In this section we derive Bennett type concentration inequalities that provide bounds on the probability that the sum of independent and bounded from above random variables differs from its expected value. The bounds depend on the random variables’ first $p$ moments and are given in terms of the generalized Lambert $W$-function (Scott et al., 2006). For real numbers $\alpha_i$, $i = 0, \ldots, p$, $\alpha_0 > 1$, consider the one dimensional transcendental equation:

$$\alpha_0 - \sum_{j=1}^{p} \alpha_j x^i = \exp(x). \quad (10)$$

The solutions to equation (10) are a special case of the generalized Lambert $W$-function (Scott et al., 2006). Because $\alpha_0 > 1$ it is easy to see that equation (10) has a positive solution. We denote the non-empty set of positive solutions of equation (10) by $G_p(\alpha_0, \ldots, \alpha_p)$. The bounds given in Theorem 3 depend on the elements of the set $G_p(\alpha_0, \ldots, \alpha_p)$ where $\alpha_i$ depends on the random variables’ moments. When $p = 0$ the set $G_0(\alpha_0)$ consists of one element $\ln(\alpha_0)$. When $p = 1$ and assuming that $\alpha_1 > 0$, the set $G_1(\alpha_0, \alpha_1)$ consists of one element that is given in terms of the Lambert $W$-function. Recall that for $x \geq 0$, $y \exp(y) = x$ holds if and only if $y = W(x)$ where $W$ is the principal branch of the Lambert $W$-function (see Corless et al. (1996)). Because $\alpha_0 > 1$ and
assuming \( \alpha_1 > 0 \), the unique positive solution to the equation \( \exp(x) = \alpha_0 - \alpha_1 x \) is given by

\[
\frac{\alpha_0}{\alpha_1} - W\left(\frac{\exp(\alpha_0/\alpha_1)}{\alpha_1}\right)
\]

(see Corless et al. (1996)). We leverage this observation to derive a bound on the probability that the sum of independent random variables differs from its expected value when we use information on the random variables’ first three moments. The bound is given as a closed-form expression in terms of the \( W \)-Lambert function (see Theorem 3 part (iv)). We show that this bound is tighter than the bound given in Bennett’s inequality (see Proposition 2). We provide an example of the magnitude of improvement (see Example 2).

Finding the positive solutions of the transcendental equation (10) for \( p \geq 2 \) can be done using a computer program. It involves solving an exponential polynomial equation of order \( p \) that has at least one positive solution. When the random variables have non-negative moments we show that the transcendental equation (10) has a unique positive solution (see Theorem 3 part (ii)).

**Theorem 3** Let \( X_1, \ldots, X_n \) be independent random variables on \((-\infty, b]\) for some \( b > 0 \) and let \( S_n = \sum_{i=1}^{n} X_i \). Let \( p \geq 2 \) be an integer and assume that \( X_i \in L^p \) for all \( i = 1, \ldots, n \). Denote \( \mathbb{E}(X_i) = \mu_i \), and assume that \( \mathbb{E}(X_i^k) \leq \mu_i^k \) and \( 0 < \mathbb{E}(\max(X_i^p, 0)) \leq \mu_i^p \) for all \( k = 1, \ldots, p-1 \) and all \( i = 1, \ldots, n \).

(i) For all \( t > 0 \) we have

\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp\left(-\frac{t}{b} - \frac{\mu_2}{b^2} y + \sum_{j=2}^{p-1} \left(\frac{\mu^j}{b^j j!} - \frac{\mu^j+1}{b^{j+1} j!} y^j\right)\right)
\]

where

\[
\alpha_0 = 1 + \frac{t b^{p-1}}{\mu^p} > 1 \quad \text{and} \quad \alpha_j = \frac{b^{p-j-1} \mu^{j+1}}{\mu^p j!} - \frac{1}{j!}
\]

for all \( j = 1, \ldots, p-2 \) and \( \mu^k = \sum_{i=1}^{n} \mu_i^k \) for all \( k = 1, \ldots, p \).

(ii) If \( \mu^j \geq 0 \) for every odd number \( j \geq 3 \) then \( G_{p-2} \) consists of one element and inequality (11) reduces to

\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp\left(-\frac{t}{b} - \frac{\mu_2}{b^2} y + \sum_{j=2}^{p-1} \left(\frac{\mu^j}{b^j j!} - \frac{\mu^j+1}{b^{j+1} j!} y^j\right)\right)
\]
where \( y \) is the unique element of \( G_{p-2} \), i.e., \( y \) is the unique positive solution of the equation \( \alpha_0 - \sum_{j=1}^{p-2} \alpha_j x^j = \exp(x) \).

(iii) Suppose that \( p = 2 \). Then \( G_0(\alpha_0) = \{\ln(\alpha_0)\} \) consists of one element and inequality (11) reduces to Bennett’s inequality:

\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( - \left( \frac{t}{b} - \frac{t + \mu^2}{b^2} \right) \ln \left( \frac{tb}{\mu^2} + 1 \right) \right)
\]

(reduce to Bennett’s inequality:)

\[
= \exp \left( - \frac{\mu^2}{b^2} \left( \frac{bt}{\mu^2} + 1 \right) \ln \left( \frac{bt}{\mu^2} + 1 \right) - \frac{bt}{\mu^2} \right).
\]

(iv) Suppose that \( p = 3 \), \( \alpha_1 \neq 0 \), and \( \mathbb{E}(\max(X_3^i, 0)) = \mu_3^i \) for all \( i = 1, \ldots, n \). Then \( G_1(\alpha_0, \alpha_1) = \left\{ \frac{\alpha_0}{\alpha_1} - W\left( \frac{\exp(\alpha_0/\alpha_1)}{\alpha_1} \right) \right\} \) consists of one element and inequality (11) reduces to

\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( - \left( \frac{t}{b} - \frac{t + \mu^2}{b^2} \right) y + \left( \frac{\mu^2}{2b^2} - \frac{\mu_3^3}{2b^3} \right) y^2 \right)
\]

where \( y = \frac{\alpha_0}{\alpha_1} - W\left( \frac{\exp(\alpha_0/\alpha_1)}{\alpha_1} \right) \) and \( W \) is the Lambert \( W \)-function.

The proof of Theorem 3 consists of three steps. In the first step we bound the moment-generating function of a random variable \( X \) that is bounded from above using the first \( p \) moments of \( X \). We use Theorem 1 to prove the first step. In the second step we derive an exponential bound on the moment-generating function using the elementary inequality \( 1 + x \leq \exp(x) \) for all \( x \in \mathbb{R} \). We note that in some cases this inequality is loose and and so the second step may potentially be improved (for example see Jebara (2018) and Zheng (2018)). In the third step we apply the Chernoff bound to derive the concentration inequality.

In applications, it is more convenient to use inequality (12) than inequality (11). Using a Taylor series approximation, one can easily calculate the unique and positive solution to the equation \( \alpha_0 - \sum_{j=1}^{p-2} \alpha_j x^j = \exp(x) \). To use inequality (12) when there is information on the random variables’ first \( p \) moments, we can choose a non-negative \( \mu^j \) for every odd number \( j \geq 3 \). This is the essence of Corollary 1. For \( p \geq 4 \), Corollary 1 can be used instead of Theorem 3 part (i). The proof of Corollary 1 follows immediately from part (ii) of Theorem 3.

**Corollary 1** Assume that the notations and conditions of Theorem 3 hold. Suppose that \( p \geq 2 \).
For every odd number $j \geq 3$ such that $j \neq p$ let $\mu^j = \max(\sum_{i=1}^{n} \mu^j_i, 0)$. Then

$$
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( - \left( \frac{t}{b} - \left( \frac{t}{b^2} + \frac{\mu^2}{b^2} \right) y + \sum_{j=2}^{p-1} \left( \frac{\mu^j}{b^j j!} - \frac{\mu^j+1}{b^{j+1} j!} \right) y^j \right) \right)
$$

where $y$ is the unique positive solution of the equation $\alpha_0 - \sum_{j=1}^{p-2} \alpha_j x^j = \exp(x)$.

The next proposition shows that the concentration inequality (14) that we derive in Theorem 3 part (iv) is always tighter than Bennett’s inequality.

**Proposition 2** Assume that the conditions of Theorem 3 hold. Assume that $\mathbb{E}(X_i^2) = \mu^2_i$ and $\mathbb{E}(\max(X_i^3, 0)) = \mu^3_i$ for all $i = 1, \ldots, n$. Then the right-hand side of inequality (14) is smaller than the right-hand side of inequality (13). That is, inequality (14) is tighter than Bennett’s inequality.

We provide a numerical example that compares inequality (14) which uses information on the random variables’ third moment to Bennett’s inequality (13). For simplicity, we consider a one-sided concentration inequality for exponential random variables.

**Example 2** Let $G_1, \ldots, G_n$ be independent exponential random variables with rate 1, i.e., $G_i$ is a random variable on $[0, \infty)$ and $\mathbb{P}(G_i \leq x) = 1 - \exp(-x)$ for all $i = 1, \ldots, n$ and $x \geq 0$. Define the random variables $X_i = \mathbb{E}(G_i) - G_i = 1 - G_i$ on $(-\infty, 1]$. We have $\mu^1_i = 0$, $\mu^2_i = 1$, and

$$
\mu^3_i = \int_{0}^{\infty} \max((1-x)^3, 0) \exp(-x) dx = \frac{6}{\exp(1)} - 2
$$

for all $i = 1, \ldots, n$. In Figure 2 we plot the bound for $\mathbb{P}(S_n - \mathbb{E}(S_n) \geq nt)$ given in Bennett’s inequality (see Theorem 3 part (iii)) divided by the bound derived in Theorem 3 part (iv) as a function of $t$ for $n = 30$. We use the program Mathematica (Wolfram, 2020) to plot Figure 2 where the Lambert $W$-function is implemented.

## 3 Conclusions

We provide upper bounds on the moment-generating function of a random variable that is bounded from above using information on the random variable’s higher moments (see Theorem 1). Using
Figure 2: Comparing Bennett’s inequality and inequality (14): $X_i = 1 - G_i$ for $i = 1, \ldots, 30$ where $G_i$ is an exponential random variable with rate 1. The plot describes the ratio of the right-hand side of inequality (13) (that bounds $\mathbb{P}(S_n - \mathbb{E}(S_n) \geq nt)$) to the right-hand side of inequality (14) for $n = 30$.

these bounds we generalize and improve Hoeffding’s inequality (see Theorem 2) and Bennett’s inequality (see Theorem 3) for the case that some information on the random variables’ higher moments is available. Our bounds are simple to use and are given as closed-form expressions in most cases. Our results can be extended in a standard way to martingales and to their maximal functions. Other inequalities and results that use Hoeffding’s or Bennett’s inequalities can also be improved using our results.

4 Appendix

4.1 Proofs of the results in Section 2.1

Proof of Theorem 1. Clearly Theorem 1 holds for $s = 0$. Fix $s > 0$, $b > 0$ and a positive integer $p$. Consider the function $g(x) := T_{p+1}(x)/x^p$ on $(-\infty, \infty)$ where we define

$$g(0) := \frac{1}{p!} = \lim_{x \to 0} g(x).$$

The proof proceeds with the following steps:

Step 1. We have $g(x) \leq g(0)$ for all $x < 0$.

Proof of Step 1. First note that for $x \leq 0$ we have $T_p \leq 0$ if $p$ is an even number and $T_p \geq 0$ if $p$ is an odd number (to see this note that $T_1(x) = \exp(x) \geq 0$, $T_p(0) = 0$ and $T_p^{(1)} = T_{p-1}$ for all
We now show that \( g(x) \leq g(b) \) for all \( x \leq b \).

Suppose first that \( x < 0 \). If \( p \) is an even number then \( T_{p+1}(x)/x^p \leq 1/p! \) if and only if \( T_{p+1}(x) - x^p/p! \leq 0 \). The last inequality is equivalent to \( T_{p+2}(x) \leq 0 \) which holds because \( p \) is an even number. Similarly, if \( p \) is an odd number then \( T_{p+1}(x)/x^p \leq 1/p! \) if and only if \( T_{p+2}(x) \geq 0 \) which holds because \( p \) is an odd number. Thus, \( g(x) \leq g(0) \) for all \( x < 0 \).

**Step 2.** Let \( f, k : [a, b) \to \mathbb{R} \) be continuously differentiable functions such that \( k^{(1)}(x) \neq 0 \) for all \( x \in (a, b) \). If \( f^{(1)}/k^{(1)} \) is increasing on \((a, b)\) then \((f(x) - f(a))/(k(x) - k(a))\) is increasing in \( x \) on \((a, b)\).

**Proof of Step 2.** Step 2 is known as the L’Hospital rule for monotonicity. For a proof see Lemma 2.2 in Anderson et al. (1993).

**Step 3.** The function \( g \) is increasing on \((0, y)\) for all \( y > 0 \).

**Proof of Step 3.** Let \( y > 0 \) and note that the function \( T_1(x)/p! = \exp(x)/p! \) is increasing on \((0, y)\). Using Step 2 with \( f(x) = \exp(x) \) and \( k(x) = p!x \) implies that the function

\[
\frac{\exp(x) - 1}{p!x} = \frac{T_2(x)}{p!x}
\]

is increasing on \((0, y)\). Applying again Step 2 and using the facts that \( T_{k+1}^{(1)} = T_k \) and \( T_k(0) = 0 \) for all \( k = 2, \ldots \) implies that the function \( T_k(x)/(x^{k-1}p!/(k-1)! \) is increasing in \( x \) on \((0, y)\) for all \( k = 2, \ldots \). Choosing \( k = p + 1 \) shows that \( g \) is increasing on \((0, y)\).

**Step 4.** We have

\[
T_{p+1}(sx) \leq \frac{\max(x^p, 0)}{bp} T_{p+1}(sb)
\]

for all \( x \leq b \).

**Proof of Step 4.** Step 3 shows that \( g \) is an increasing function on \((0, b)\). Hence, \( g(x) \leq g(b) \) for all \( x \in (0, b) \). Because \( g \) is a continuous function we have \( g(0) \leq g(b) \). Using Step 1 implies that \( g(x) \leq g(b) \) for all \( x \leq b \).

Let \( x \leq b \) and assume \( x \neq 0 \). Multiplying each side of the inequality \( g(sx) \leq g(sb) \) by the positive number \( \max(x^p, 0) \) yields

\[
\max(x^p, 0) T_{p+1}(sx) \leq \frac{\max(x^p, 0)}{bp} T_{p+1}(sb)
\]
Note that

\[ T_{p+1}(sx) \leq \frac{\max(x^p, 0)}{x^p} T_{p+1}(sx). \]

The last inequality holds as equality if \( x > 0 \) or if \( p \) is an even number. If \( x < 0 \) and \( p \) is an odd number, then \( T_{p+1}(sx) \leq 0 \) (see Step 1), so the last inequality holds. We conclude that

\[ T_{p+1}(sx) \leq \frac{\max(x^p, 0)}{b^p} T_{p+1}(sb) \]

for all \( x \leq b \).

To prove Theorem 1 apply Step 4 to conclude that

\[ \exp(sx) \leq \frac{\max(x^p, 0)}{b^p} T_{p+1}(sb) + \sum_{j=0}^{p-1} \frac{s^j x^j}{j!} \]

for all \( x \leq b \). Taking expectations in both sides of the last inequality proves Theorem 1. ■

**Proof of Proposition 1.** Let \( p \geq 2 \) be an even number.

(i) We have

\[ \frac{\mathbb{E}X^p}{b^p} T_{p+1}(sb) + \mathbb{E} \left( \sum_{j=0}^{p-1} \frac{s^j X^j}{j!} \right) \geq \frac{\mathbb{E} \max(X, 0)^{p+1}}{b^{p+1}} T_{p+2}(sb) + \mathbb{E} \left( \sum_{j=0}^{p} \frac{s^j X^j}{j!} \right) \]

\[ \iff \frac{\mathbb{E}X^p}{b^p} \left( T_{p+1}(sb) - \frac{s^p b^p}{p!} \right) \geq \frac{\mathbb{E} \max(X, 0)^{p+1}}{b^{p+1}} T_{p+2}(sb) \]

\[ \iff b \mathbb{E}X^p \geq \mathbb{E} \max(X, 0)^{p+1} \]

which holds for a random variable \( X \) on \(( -\infty, b ] \) and an even number \( p \) because \( bx^p \geq \max(x, 0)^{p+1} \)

for all \( x \leq b \).

(ii) Similarly to part (i) we have \( m_X(p+1) \geq m_X(p+2) \) if and only if \( b \mathbb{E}X^{p+1} \geq \mathbb{E}X^{p+2} \) which holds for a non-negative random variable because \( bx^{p+1} \geq x^{p+2} \) for all \( 0 \leq x \leq b \). ■

4.2 Proofs of the results in Section 2.2

**Proof of Theorem 2.** Let \( p \geq 2 \) be an integer. We will use the following notations in proof. Let \( X \) be a random variable on \([0, b] \). Denote \( \mathbb{E}(X^k) = \mu^k \) for all \( k = 1, \ldots, p \).
For every integer \( p \geq 1 \) we define the function

\[
v(y, b, \mu_1, \ldots, \mu_p) := \frac{\mu_p}{b^p} T_{p+1}(y) + \sum_{j=0}^{p-1} \frac{y^j \mu_j}{b^j j!}.
\]

For all \( x \geq 0 \) we define the function

\[
C_p(x, b, \mu_1, \ldots, \mu_p) = \max_{0 \leq y \leq x} \frac{\left(v^{(2)}(y, b, \mu_1, \ldots, \mu_p)\right)^2}{\left(v^{(1)}(y, b, \mu_1, \ldots, \mu_p)\right)^2}.
\]

We denote by \( v^{(k)}(y, b, \mu_1, \ldots, \mu_p) \) the \( k \)th derivative of \( v \) with respect to its first argument. A straightforward calculation shows that

\[
v^{(k)}(y, b, \mu_1, \ldots, \mu_p) = \frac{\mu_p}{b^p} T_{p+1-k}(y) + \sum_{j=0}^{p-1-k} \frac{\mu^{j+k} y^j}{b^{j+k} j!}.
\]

Thus,

\[
v^{(1)}(0, b, \mu_1, \ldots, \mu_p) = \frac{\mathbb{E}(X)}{b} > 0 \text{ and } v^{(2)}(0, b, \mu_1, \ldots, \mu_p) = \frac{\mathbb{E}(X^2)}{b^2} > 0.
\]

Because \( v^{(2)} \) and \( v^{(1)} \) are increasing in the first argument as the sum of increasing functions, we conclude that \( v^{(2)}(y, b, \mu_1, \ldots, \mu_p) \) and \( v^{(1)}(y, b, \mu_1, \ldots, \mu_p) \) are positive for every \( y \in [0, x] \) and all \( x > 0 \).

The proof proceeds with the following steps:

**Step 1.** We have \( \mu^{d+2} \mu^d \geq (\mu^{d+1})^2 \) for every positive integer \( d \).

**Proof of Step 1.** Let \( d \) be a positive integer. From the Cauchy-Schwarz inequality for the (positive) random variables \( X^{(d+2)/2} \) and \( X^{d/2} \) we have

\[
\mathbb{E}X^{d/2}X^{(d+2)/2} \leq \sqrt{\mathbb{E}X^d \mathbb{E}X^{d+2}}.
\]

That is, we have \( \mu^{d+2} \mu^d \geq (\mu^{d+1})^2 \) which proves Step 1.

**Step 2.** For every positive integer \( p \) and all \( x > 0 \) the function \( v^{(1)}(y, b, \mu_1, \ldots, \mu_p) \) is log-convex in \( y \) on \((0, x)\) (i.e., \( \log(v^{(1)}(y, b, \mu_1, \ldots, \mu_p)) \) is a convex function on \((0, x))\).
Proof of Step 2. Fix a positive integer $p$ and $x > 0$. Let
\[
w(y) := T_p(y) + \frac{b^p}{\mu^p} \sum_{j=0}^{p-2} \frac{\mu^{j+1} y^j}{b^{j+1} j!} = \frac{b^p v^{(1)}(y, b, \mu^1, \ldots, \mu^p)}{\mu^p}.
\]

To prove Step 2 it is enough to prove that $w$ is log-convex on $(0, x)$. Note that
\[
w(y) = \exp(y) + \sum_{j=0}^{p-2} \frac{y^j}{j!} \beta_{j+1}
\]
where
\[
\beta_j = \frac{b^p - j \mu^j}{\mu^p} - 1
\]
for $j = 1, \ldots, p$. We have $\beta_j \geq 0$ for all $j = 1, \ldots, p - 1$. To see this note that $x^j b^p - j \geq x^p$ for all $x \in [0, b]$ so taking expectations implies that $\beta_j \geq 0$.

$w$ is log-convex on $(0, x)$ if and only if $w^{(1)}/w$ is increasing on $(0, x)$. For every integer $k = 0, \ldots, p - 2$ define the function
\[
w_k(y) = \exp(y) + \sum_{j=0}^{k} \frac{y^j}{j!} \beta_{p-1+j-k}
\]
and note that $w_{p-2} = w$. By construction we have $w_k^{(1)} = w_{k-1}$. We now show that $w_k$ is log-convex on $(0, x)$ for all $k = 0, \ldots, p - 2$. The proof is by induction.

For $k = 0$ the function
\[
\frac{w_0^{(1)}(y)}{w_0(y)} = \frac{\exp(y)}{\exp(y) + \beta_{p-1}}
\]
is increasing because $\beta_{p-1} \geq 0$ and the function $x/(x + d)$ is increasing in $x$ on $[0, \infty)$ when $d \geq 0$. We conclude that the function $w_0 = \exp(y) + \beta_{p-1}$ is log-convex on $(0, x)$.

Assume that $w_k$ is log-convex on $(0, x)$ for some integer $0 \leq k \leq p - 3$. We show that $w_{k+1}$ is log-convex on $(0, x)$. Log-convexity of $w_k$ implies that the function
\[
\frac{w_k^{(1)}(y)}{w_k(y)} = \frac{\exp(y) + \sum_{j=0}^{k-1} \frac{y^j}{j!} \beta_{p+j-k}}{\exp(y) + \sum_{j=0}^{k} \frac{y^j}{j!} \beta_{p-1+j-k}}
\]
is increasing on $(0, x)$. Using the fact that $w_{k+1}^{(1)} = w_k$ and applying Step 2 in the proof of Theorem
1 we conclude that the function
\[ m(y) := \frac{w_k(y) - w_k(0)}{w_{k+1}(y) - w_{k+1}(0)} = \frac{\exp(y) + \sum_{j=0}^{k} w_j \beta_{p-1+j-k} - (1 + \beta_{p-1-k})}{\exp(y) + \sum_{j=0}^{k+1} w_j \beta_{p-2+j-k} - (1 + \beta_{p-2-k})} \]
is increasing on \((0, x)\). Thus, \(m^{(1)}(y) \geq 0\) for all \(y \in (0, x)\). That is,
\[ w_k^{(2)}(y)w_{k+1}(y) - w_k^{(1)}(y)(1 + \beta_{p-2-k}) \geq (w_k^{(1)}(y))^2 - w_k^{(1)}(y)(1 + \beta_{p-1-k}) \tag{16} \]
for all \(y \in (0, x)\). We now show that \(w_k^{(2)}(y)w_{k+1}(y) \geq (w_k^{(1)}(y))^2\). Because \(w_{k+1}/w_k^{(1)}\) is increasing and positive (see (15)) we have
\[ \frac{w_k^{(2)}(y)}{w_k^{(1)}(y)}(1 + \beta_{p-2-k}) \geq (1 + \beta_{p-1-k}) \tag{17} \]
for all \(y \in (0, x)\) if the last inequality holds for \(y = 0\), i.e., if
\[ (1 + \beta_{p-k})(1 + \beta_{p-2-k}) \geq (1 + \beta_{p-1-k})^2 \iff \left( \frac{b^k \mu^{p-k}}{\mu_p} \right) \left( \frac{b^{k+2} \mu^{p-2-k}}{\mu_p} \right) \geq \left( \frac{b^{k+1} \mu^{p-k-1}}{\mu_p} \right)^2 \]
\[ \iff \mu^{p-k} \mu^{p-k-2} \geq (\mu^{p-k-1})^2 \]
which holds from Step 1. We conclude that inequality (17) holds. Using inequality (16) we have
\[ w_k^{(2)}(y)w_{k+1}(y) - (w_k^{(1)}(y))^2 \geq w_k^{(2)}(y)(1 + \beta_{p-2-k}) - w_k^{(1)}(y)(1 + \beta_{p-1-k}) \geq 0. \]
That is, \(w_{k+1}(y)w_{k+1}(y) \geq (w_k^{(1)}(y))^2\) for all \(y \in (0, x)\). We conclude that \(w_k^{(1)}/w_{k+1}\) is increasing on \((0, x)\), i.e., \(w_{k+1}\) is log-convex. This shows that \(w_k\) is log-convex for all \(k = 0, \ldots, p - 2\). In particular, \(w_{p-2} := w\) is log-convex which proves Step 2.

**Step 3.** We have
\[ C_p(x, b, \mu^1, \ldots, \mu^p) = \left( \frac{\exp(x) + \sum_{j=0}^{p-3} x^j \frac{b^{p-j-2} \mu^j + 2}{\mu^j p} - 1}{\exp(x) + \sum_{j=0}^{p-2} x^j \frac{b^{p-j-1} \mu^j + 1}{\mu^j p} - 1} \right)^2. \]
for all \(x \geq 0\).

**Proof of Step 3.** Let \(x \geq 0\). From Step 2 the function \(v^{(1)}(y, b, \mu^1, \ldots, \mu^p)b^p/\mu^p := w(y)\) is
log-convex on $(0, x)$ where $w$ is defined in the proof of Step 2. This implies that

$$C_p(x, b, \mu^1, \ldots, \mu^p) = \max_{0 \leq y \leq x} \left( \frac{(v(2)(y, b, \mu^1, \ldots, \mu^p))^2}{(v(1)(y, b, \mu^1, \ldots, \mu^p))^2} \right) = \max_{0 \leq y \leq x} \left( \frac{w(1)(y)}{w(y)} \right)^2 = \left( \frac{w(1)(x)}{w(x)} \right)^2$$

which proves Step 3.

**Step 4.** For all $x \geq 0$ we have

$$\max_{0 \leq y \leq x} \left( \frac{v(2)(y, b, \mu^1, \ldots, \mu^p)}{v(y, b, \mu^1, \ldots, \mu^p)} - \frac{(v(1)(y, b, \mu^1, \ldots, \mu^p))^2}{(v(y, b, \mu^1, \ldots, \mu^p))^2} \right) \leq \frac{1}{4} C_p(x, b, \mu^1, \ldots, \mu^p).$$

**Proof of Step 4.** For all $x \geq 0$ we have

$$\max_{0 \leq y \leq x} \left( \frac{v(2)(y, b, \mu^1, \ldots, \mu^p)}{v(y, b, \mu^1, \ldots, \mu^p)} - \frac{(v(1)(y, b, \mu^1, \ldots, \mu^p))^2}{(v(y, b, \mu^1, \ldots, \mu^p))^2} \right) = \max_{0 \leq y \leq x} \left( \frac{v(2)(y, b, \mu^1, \ldots, \mu^p)}{v(y, b, \mu^1, \ldots, \mu^p)} \left(1 - \frac{v(2)(y, b, \mu^1, \ldots, \mu^p)(v(1)(y, b, \mu^1, \ldots, \mu^p))^2}{v(y, b, \mu^1, \ldots, \mu^p)(v(2)(y, b, \mu^1, \ldots, \mu^p))^2} \right) \right) \leq \frac{1}{4} C_p(x, b, \mu^1, \ldots, \mu^p).$$

The first inequality follows from the definition of $C_p$ and because $v(2) > 0$ and $v(1) > 0$. The second inequality follows from the elementary inequality $x(1 - x/z) \leq 0.25z$ for all $z > 0$ and $x > 0$.

**Step 5.** We have

$$E \exp(s(X - E(X)) \leq \exp \left( \frac{s^2 b^2 C_p(sb, b, \mu^1, \ldots, \mu^p)}{8} \right)$$

**Proof of Step 5.** From Theorem 1 for all $s \geq 0$ we have

$$E \exp(sX) \leq \frac{EX^p}{b^p} T_{p+1}(sb) + \mathbb{E} \left( \sum_{j=0}^{p-1} \frac{s^j X^j}{j!} \right) = v(y, b, \mu^1, \ldots, \mu^p)$$
where \( y = sb \geq 0 \). Define the function

\[
g(y) = \ln \left( v(y, b, \mu^1, \ldots, \mu^p) \right).
\]

Clearly \( v \) is a positive function so the function \( g : \mathbb{R}_+ \to \mathbb{R} \) is well defined. Note that \( \mathbb{E}\exp(sX) \leq \exp(g(y)) \). Recall that

\[
v^{(1)}(0, b, \mu^1, \ldots, \mu^p) = \frac{\mathbb{E}(X)}{b} > 0 \quad \text{and} \quad v^{(2)}(0, b, \mu^1, \ldots, \mu^p) = \frac{\mathbb{E}(X^2)}{b^2} > 0.
\]

Because \( v(0, b, \mu^1, \ldots, \mu^p) = 1 \) we have \( g(0) = \ln(1) = 0 \). We have

\[
g^{(1)}(y) = \frac{v^{(1)}(y, b, \mu^1, \ldots, \mu^p)}{v(y, b, \mu^1, \ldots, \mu^p)}.
\]

Thus, \( g^{(1)}(0) = \mathbb{E}(X)/b \). Differentiating again yields

\[
g^{(2)}(y) = \frac{v^{(2)}(y, b, \mu^1, \ldots, \mu^p)}{v(y, b, \mu^1, \ldots, \mu^p)} - \frac{(v^{(1)}(y, b, \mu^1, \ldots, \mu^p))^2}{(v(y, b, \mu^1, \ldots, \mu^p))^2}.
\]

From Taylor’s theorem for all \( y \geq 0 \) there exists a \( z \in [0, y] \) such that \( g(y) = g(0) + yg^{(1)}(0) + 0.5y^2g^{(2)}(z) \). Thus, using the fact that \( y = sb \) we have

\[
g(y) = g(0) + yg^{(1)}(0) + 0.5y^2g^{(2)}(z) = s\mathbb{E}(X) + 0.5s^2b^2g^{(2)}(z) \leq s\mathbb{E}(X) + 0.5s^2b^2V(sb, b, \mu^1, \ldots, \mu^p)
\]

where

\[
V(y, b, \mu^1, \ldots, \mu^p) = \sup_{0 \leq z \leq y} g^{(2)}(z).
\]

Using \( \mathbb{E}\exp(sX) \leq \exp(g(y)) \) and Step 4 imply

\[
\mathbb{E}\exp(s(X - \mathbb{E}(X))) \leq \exp \left( \frac{s^2b^2V(sb, b, \mu^1, \ldots, \mu^p)}{2} \right) \leq \exp \left( \frac{s^2b^2C_p(sb, b, \mu^1, \ldots, \mu^p)}{8} \right)
\]

21
Step 6. For all $t > 0$ we have

$$
\mathbb{P}(S_n - E(S_n) \geq t) \leq \exp \left( -\frac{t^2}{2 \sum_{i=1}^{n} b_i^2 C_p \left( \frac{4tb_i}{\sum_{i=1}^{n} d(X_i)}, b_i, \mu_i^1, \ldots, \mu_i^p \right)} \right).
$$

Proof of Step 6. Using independence, Step 5, and Markov's inequality, a standard argument shows that:

$$
\mathbb{P}(S_n - E(S_n) \geq t) \leq \exp(-st) \mathbb{E} \exp(s(S_n - E(S_n)))
$$

$$
= \exp(-st) \prod_{i=1}^{n} \mathbb{E} \exp(s(X_i - E(X_i)))
$$

$$
\leq \exp(-st) \prod_{i=1}^{n} \exp \left( \frac{s^2 b_i^2 C_p(s b_i, b_i, \mu_i^1, \ldots, \mu_i^p)}{8} \right)
$$

$$
= \exp \left( -st + \frac{s^2}{8} \sum_{i=1}^{n} b_i^2 C_p(s b_i, b_i, \mu_i^1, \ldots, \mu_i^p) \right)
$$

Let

$$
s = \frac{4t}{\sum_{i=1}^{n} b_i^2 C_p \left( \frac{4tb_i}{\sum_{i=1}^{n} d(X_i)}, b_i, \mu_i^1, \ldots, \mu_i^p \right)}
$$

Note that

$$
C_p(0, b_i, \mu_i^1, \ldots, \mu_i^p) = \left( \frac{\mu_i^2}{\mu_i^1 b_i} \right)^2 = \frac{d(X_i)}{b_i^2}
$$

Because $C_p$ is increasing in the first argument we have

$$
C_p(y, b_i, \mu_i^1, \ldots, \mu_i^p) \geq \frac{d(X_i)}{b_i^2} > 0
$$

for all $y \geq 0$. Thus,

$$
\frac{4tb_i}{\sum_{i=1}^{n} b_i^2 C_p \left( \frac{4tb_i}{\sum_{i=1}^{n} d(X_i)}, b_i, \mu_i^1, \ldots, \mu_i^p \right)} \leq \frac{4tb_i}{\sum_{i=1}^{n} d(X_i)}.
$$
Using again the fact that $C_p$ is increasing in the first argument implies
\[-st + \frac{s^2}{8} \sum_{i=1}^{n} b_i^2 C_p(s b_i, b_i, \mu_1^i, \ldots, \mu_p^i) \leq -st + \frac{s^2}{8} \sum_{i=1}^{n} b_i^2 C_p \left( \frac{4tb_i}{\sum_{i=1}^{n} d(X_i)}, b_i, \mu_1^i, \ldots, \mu_p^i \right) = -\frac{2t^2}{\sum_{i=1}^{n} b_i^2 C_p \left( \frac{4tb_i}{\sum_{i=1}^{n} d(X_i)}, b_i, \mu_1^i, \ldots, \mu_p^i \right)}.
\]

We conclude that
\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( -\frac{2t^2}{\sum_{i=1}^{n} b_i^2 C_p \left( \frac{4tb_i}{\sum_{i=1}^{n} d(X_i)}, b_i, \mu_1^i, \ldots, \mu_p^i \right)} \right),
\]
which proves Step 6.

Combining Steps 3 and 6 proves part (i).

(ii) Let $X$ be a random variable on $[0, b]$. Denote $\mathbb{E}(X^k) = \mu^k$ for all $k = 1, \ldots, p$. Clearly $0 < C_p$ because $v^{(2)}$ and $v^{(1)}$ are positive functions (see part (i)).

We show that $v^{(2)}(y, b, \mu_1^1, \ldots, \mu^p) \leq v^{(1)}(y, b, \mu_1^1, \ldots, \mu^p)$ for all $y \geq 0$.

We have $v^{(2)}(y, b, \mu_1^1, \ldots, \mu^p) \leq v^{(1)}(y, b, \mu_1^1, \ldots, \mu^p)$ if and only if
\[
\frac{\mu^p}{b^p} T_{p-1}(y) + \sum_{j=0}^{p-3} \frac{y^j \mu_{j+2}}{b^{j+2} j!} \leq \frac{\mu^p}{b^p} T_p(y) + \sum_{j=0}^{p-2} \frac{y^j \mu_{j+1}}{b^{j+1} j!}.
\]

The last inequality holds if and only if
\[
\frac{\mu^p y^{p-2}}{b^p (p-2)!} + \sum_{j=0}^{p-3} \frac{y^j \mu_{j+2}}{b^{j+2} j!} - \sum_{j=0}^{p-2} \frac{y^j \mu_{j+1}}{b^{j+1} j!} \leq 0
\]
\[
\iff \sum_{j=0}^{p-2} \frac{y^j \mu_{j+2}}{b^{j+2} j!} - \sum_{j=0}^{p-2} \frac{y^j \mu_{j+1}}{b^{j+1} j!} \leq 0.
\]

To see that the last inequality holds let $0 \leq x \leq b$. We have $b^{j+1} x^{j+2} \leq x^{j+1} b^{j+2}$. Taking expectations and multiplying by $y^j/j!$ show that
\[
\frac{y^j \mu_{j+2}}{b^{j+2} j!} \leq \frac{y^j \mu_{j+1}}{b^{j+1} j!}
\]
for all $1 \leq j \leq p-2$ and all $y \geq 0$. We conclude that $v^{(2)}(y, b, \mu_1^1, \ldots, \mu^p) \leq v^{(1)}(y, b, \mu_1^1, \ldots, \mu^p)$
for all \( y \geq 0 \). Thus,

\[
C_p(x, b, \mu^1, \ldots, \mu^p) = \max_{0 \leq y \leq x} \frac{(v^{(2)}(y, b, \mu^1, \ldots, \mu^p))^2}{(v^{(1)}(y, b, \mu^1, \ldots, \mu^p))^2} \leq 1
\]

which immediately implies that inequality (4) is tighter than Hoeffding’s inequality which corresponds to \( C_p = 1 \) (when \( p = 1 \) the argument above shows that \( v^2 = v^1 \) so \( C_1 = 1 \) and we derive Hoeffding’s inequality (6)).

(iii) Let \( Z \) be a random variable on \([0, b]\). Denote \( \mathbb{E}(Z^k) = \mu^k \) for all \( k = 1, \ldots \) and let \( y \in [0, x] \) for some \( x \geq 0 \).

First note that \( 0 \leq \lim_{p \to \infty} b^{-p} \mu^p T_p(y) \leq \lim_{p \to \infty} T_p(y) = 0 \). In addition, for every \( z \in [0, b] \) we have

\[
\lim_{p \to \infty} b^{-1} z \exp(yz/b) - b^{-1} z \sum_{j=0}^{p-2} \frac{y^j z^j}{b^j j!} = 0.
\]

Because \( Z \) is a random variable on \([0, b]\) we can use the bounded convergence theorem to conclude that

\[
\lim_{p \to \infty} v^{(1)}(y, b, \mu^1, \ldots, \mu^p) = \lim_{p \to \infty} \frac{\mu^p}{b^p} T_p(y) + \sum_{j=0}^{p-2} \frac{y^j \mu^{j+1}}{b^{j+1} j!} = b^{-1} \mathbb{E} Z \exp(yZ/b).
\]

Similarly,

\[
\lim_{p \to \infty} v^{(2)}(y, b, \mu^1, \ldots, \mu^p) = \lim_{p \to \infty} \frac{\mu^p}{b^p} T_{p-1}(y) + \sum_{j=0}^{p-3} \frac{y^j \mu^{j+2}}{b^{j+2} j!} = b^{-2} \mathbb{E} Z^2 \exp(yZ/b).
\]

We conclude that

\[
\lim_{p \to \infty} \left( \frac{v^{(2)}(y, b, \mu^1, \ldots, \mu^p)}{v^{(1)}(y, b, \mu^1, \ldots, \mu^p)} \right)^2 = \left( \frac{b^{-2} \mathbb{E} Z^2 \exp(yZ/b)}{b^{-1} \mathbb{E} Z \exp(yZ/b)} \right)^2 = \frac{\mathbb{E} Z^2 \exp(yZ/b)}{\mathbb{E} Z \exp(yZ/b)}
\]
Using Step 3 in the proof of part (i) yields

\[
\lim_{p \to \infty} C_p(x, b_i, \mu_1, \ldots, \mu_p) = \lim_{p \to \infty} \max_{0 \leq y \leq x} \left( \frac{v(2)(y, b_i, \mu_1, \ldots, \mu_p)}{v(1)(y, b_i, \mu_1, \ldots, \mu_p)} \right)^2
\]

\[
= \lim_{p \to \infty} \left( \frac{v(2)(x, b_i, \mu_1, \ldots, \mu_p)}{v(1)(x, b_i, \mu_1, \ldots, \mu_p)} \right)^2
\]

\[
= b^{-2} \left( \frac{\mathbb{E}Z^2 \exp(xZ/b)}{\mathbb{E}Z \exp(xZ/b)} \right)^2
\]

which proves part (iii). •

4.3 Proofs of the results in Section 2.3

**Proof of Theorem 3.** (i) Let \( s \geq 0 \) and let \( p \geq 2 \) be an integer. We first assume that \( b = 1 \) so that \( X_i \) is a random variable on \(( -\infty, 1 \] for all \( i = 1, \ldots, n \).

For any random variable \( X_i \) on \(( -\infty, 1 \] we have

\[
\mathbb{E} \exp(sX_i) \leq \mu_i^p \left( \exp(s) - \sum_{j=0}^{p-1} \frac{s^j}{j!} \right) + 1 + \sum_{j=1}^{p-1} \frac{s^j \mu_i^j}{j!}
\]

\[
\leq \exp \left( \mu_i^p \left( \exp(s) - \sum_{j=0}^{p-1} \frac{s^j}{j!} \right) + \sum_{j=1}^{p-1} \frac{s^j \mu_i^j}{j!} \right)
\]

\[
= \exp \left( \mu_i^p T_{p+1}(s) + \sum_{j=1}^{p-1} \frac{s^j \mu_i^j}{j!} \right)
\]

The first inequality follows from Theorem 1 and the fact that \( T_{p+1}(s) \geq 0 \) for \( s \geq 0 \). The second inequality follows from the elementary inequality \( 1 + x \leq e^x \) for all \( x \in \mathbb{R} \). Thus,

\[
\mathbb{E} \exp(s(X_i - \mu_1^1)) \leq \exp \left( \mu_i^p T_{p+1}(s) + \sum_{j=2}^{p-1} \frac{s^j \mu_i^j}{j!} \right)
\]
and
\[
\prod_{i=1}^{n} \mathbb{E} \exp(s(X_i - \mathbb{E}(X_i))) \leq \prod_{i=1}^{n} \exp \left( \mu_i^p T_{p+1}(s) + \sum_{j=2}^{p-1} \frac{s^j \mu_j^i}{j!} \right) \\
= \exp \left( \mu^p T_{p+1}(s) + \sum_{j=2}^{p-1} \frac{s^j \mu_j^p}{j!} \right)
\]

From the Chernoff bound and the fact that \(X_1, \ldots, X_n\) are independent random variables, for all \(t > 0\), we have

\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \inf_{s \geq 0} \exp(-st) \mathbb{E} \exp(s(S_n - \mathbb{E}(S_n)))
\]

\[
= \inf_{s \geq 0} \exp(-st) \prod_{i=1}^{n} \mathbb{E} \exp(s(X_i - \mathbb{E}(X_i)))
\]

\[
\leq \inf_{s \geq 0} \exp \left( -st + \sum_{j=2}^{p-1} \frac{s^j \mu_j^p}{j!} + \mu^p T_{p+1}(s) \right)
\]

\[
= \exp \left( -\sup_{s \geq 0} \left( st - \sum_{j=2}^{p-1} \frac{s^j \mu_j^p}{j!} - \mu^p T_{p+1}(s) \right) \right)
\]

\[
= \exp \left( -\mu^p \sup_{x \geq 0} h_p(x, t, \mu^2, \ldots, \mu^p) \right)
\]

where

\[
h_p(x, t, \mu^2, \ldots, \mu^p) = \frac{t}{\mu^p} x - \frac{1}{\mu^p} \sum_{j=2}^{p-1} \frac{x^j \mu_j^p}{j!} - T_{p+1}(x)
\]

\[
= 1 + \left( \frac{t}{\mu^p} + 1 \right) x - \sum_{j=2}^{p-1} \left( \frac{\mu_j^p}{\mu^p j!} - \frac{1}{j!} \right) x^j \exp(x)
\]

Because \(h_p\) is continuous, \(h_p(0, t, \mu^2, \ldots, \mu^p) = 0\), and \(\lim_{x \to \infty} h_p(x, t, b, \mu^2, \ldots, \mu^p) = -\infty\), the function \(h_p\) has a maximizer. Let \(h_p^{(j)}\) the \(j\)th derivative of \(h_p\) with respect to \(x\).
Note that
\[
    h_p^{(1)}(x, t, \mu^2, \ldots, \mu^p) = \frac{t}{\mu^p} + 1 - \sum_{j=1}^{p-2} \frac{\mu^{j+1}}{\mu^p j!} x^j - \exp(x)
\]
\[
    = \alpha_0 - \sum_{j=1}^{p-2} \alpha_j x^j - \exp(x)
\]

Thus, \( h_p^{(1)}(0, t, \mu^2, \ldots, \mu^p) = \alpha_0 - \exp(0) > 0 \) and \( h_p^{(1)}(x, t, \mu^2, \ldots, \mu^p) < 0 \) for all \( x \geq \overline{x} \) for some large \( \overline{x} \). Because \( h_p^{(1)} \) is continuous we conclude that the maximizer \( y \) of \( h_p \) on \([0, \infty)\) satisfies \( h_p^{(1)}(y, t, \mu^2, \ldots, \mu^p) = 0 \), that is, \( y \in G_{p-2}(\alpha_0, \ldots, \alpha_{p-2}) \). Plugging \( y \) into \( h_p \) yields
\[
    \left( \frac{t}{\mu^p} + 1 \right) y - \sum_{j=2}^{p-1} \left( \frac{\mu^j}{\mu^p j!} - \frac{1}{j!} \right) y^j + 1 - \exp(y)
\]
\[
    = \left( \frac{t}{\mu^p} + 1 \right) y - \sum_{j=2}^{p-1} \left( \frac{\mu^j}{\mu^p j!} - \frac{1}{j!} \right) y^j - t \mu^0 + \sum_{j=1}^{p-2} \left( \frac{\mu^{j+1}}{\mu^p j!} - \frac{1}{j!} \right) y^j
\]
\[
    = -t \frac{\mu^0}{\mu^p} + \left( \frac{t}{\mu^p} + t \mu^2 \right) y - \frac{1}{\mu^p} \sum_{j=2}^{p-1} \left( \frac{\mu^j}{j!} - \frac{\mu^{j+1}}{j!} \right) y^j.
\]

In the first equality we used the fact that \( h_p^{(1)}(y, t, \mu^2, \ldots, \mu^p) = 0 \). Thus,
\[
    \mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( -\mu^p \left( -t \frac{\mu^0}{\mu^p} + \left( \frac{t}{\mu^p} + \mu^2 \right) y - \frac{1}{\mu^p} \sum_{j=2}^{p-1} \left( \frac{\mu^j}{j!} - \frac{\mu^{j+1}}{j!} \right) y^j \right) \right)
\]
\[
    = \exp \left( -\max_{y \in G_{p-2}(\alpha_0, \ldots, \alpha_{p-2})} \left( -t + \left( t + \mu^2 \right) y - \frac{1}{\mu^p} \sum_{j=2}^{p-1} \left( \frac{\mu^j}{j!} - \frac{\mu^{j+1}}{j!} \right) y^j \right) \right)
\]
(20)

which proves part (i) for the case that \( b = 1 \). Now suppose that \( b \neq 1 \) and \( X_i \leq b \) for some \( b > 0 \). Define the random variable \( Y_i = X_i/b \) and note that \( Y_i \leq 1 \) and \( \mathbb{E}Y_i^k \leq \mu_i^k / b^k \). Thus, we can apply inequality (20) for the random variables \( Y_1, \ldots, Y_n \) to conclude that for all \( t > 0 \) we have
\[
    \mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) = \mathbb{P} \left( \sum_{i=1}^{n} Y_i - \mathbb{E} \left( \sum_{i=1}^{n} Y_i \right) \geq \frac{t}{b} \right)
\]
\[
    \leq \exp \left( -\max_{y \in G_{p-2}(\alpha_0, \ldots, \alpha_{p-2})} \left( \frac{t}{b} - \left( \frac{t}{b} + \mu^2 \right) y + \frac{1}{b^p} \sum_{j=2}^{p-1} \left( \frac{\mu^j}{b^j j!} - \frac{\mu^{j+1}}{b^{j+1} j!} \right) y^j \right) \right)
\]
where 
\[ \alpha_0 = 1 + \frac{tb^{p-1}}{\mu^p} > 1 \quad \text{and} \quad \alpha_j = \frac{tb^{p-j-1}\mu_{j+1}}{\mu^{p}j!} - \frac{1}{j!} \]
for all \( j = 1, \ldots, p - 2 \). This proves part (i).

(ii) Suppose for simplicity that \( b = 1 \) (as in part (i) part (ii) holds for any \( b > 0 \) when it holds for \( b = 1 \)). Note that

\[ h_p^{(1)}(x, t, \mu^2, \ldots, \mu^p) = \frac{t}{\mu^p} - \frac{1}{\mu^p} \sum_{j=1}^{p-2} x^j \mu_{j+1} - T_p(x) \]

so if \( \mu^2 \geq 0 \) for every odd number \( j \geq 3, j \neq p \), then \( h_p^{(1)} \) is strictly decreasing on \((0, \infty)\). Hence, there is a unique positive solution for the equation \( h_p^{(1)}(x, t, \mu^2, \ldots, \mu^p) = 0 \) which implies that the set \( G_{p-2}(\alpha_0, \ldots, \alpha_{p-2}) \) consists only one element (see the proof of part (i)).

(iii) Assume that \( p = 2 \). Then the unique solution to the equation \( \alpha_0 = \exp(x) \) is \( \ln(\alpha_0) \). Thus, \( G_2(\alpha_0) = \{y\} \) where

\[ y = \ln \left( 1 + \frac{tb}{\mu^2} \right). \]

Plugging \( y \) into equation (3) proves part (iii).

(iv) Assume that \( p = 3 \). From part (ii) \( G_3 \) consists of one element. Note that \( bx^2 \geq \max(x^3, 0) \) for all \( x \leq b \). Thus, \( b\mu_i^2 \geq \mu_i^3 \) for all \( i = 1, \ldots, n \). Hence, \( \alpha_1 \) is non-negative. Because \( \alpha_0 > 1 \) and \( \alpha_1 > 0 \) (if \( \alpha_1 = 0 \) we get Bennett’s inequality as in part (iii)), \( G_3(\alpha_0, \alpha_1) = \{y\} \) where \( y \) is the unique and positive solution to the equation \( \exp(x) = \alpha_0 - \alpha_1 x \) that is given by

\[ y = \frac{\alpha_0}{\alpha_1} - W \left( \frac{\exp(\alpha_0/\alpha_1)}{\alpha_1} \right) \]

where \( W \) is the Lambert \( W \)-function (see Corless et al. (1996)). Plugging \( y \) into equation (3) proves part (iv). ■

**Proof of Proposition 2.** As in Theorem 3 we denote \( \sum_{i=1}^n \mathbb{E}(\max(X_i^3, 0)) = \mu^3 \) and \( \sum_{i=1}^n \mathbb{E}X_i^2 = \mu^2 \). We can assume without loss of generality that \( b = 1 \) so that \( X_i \) is a random variable on \((-\infty, 1]\) for all \( i = 1, \ldots, n \) (see the proof of Theorem 3 part (i)).
Let
\[ \mu^p h_p(x, t, \mu^2, \ldots, \mu^p) := tx - \sum_{j=2}^{p-1} \frac{x^j \mu^j}{j!} - \mu^p T_{p+1}(x) \]
for \( p = 2 \) and \( p = 3 \) (see the proof of Theorem 3 for the definition of \( h_p \)).

For all \( x \geq 0 \), we have
\[ \mu^2 h_2(x, t, \mu^2) \leq \mu^3 h_3(x, t, \mu^2, \mu^3) \]
\[ \Longleftrightarrow tx - \mu^2 T_3(x) \leq tx - \frac{x^2 \mu^2}{2} - \mu^3 T_4(x) \]
\[ \Longleftrightarrow \mu^3 T_4(x) \leq \mu^2 T_4(x) \]
\[ \Longleftrightarrow \mu^3 \leq \mu^2 \]
which holds because \( \max(x^3, 0) \leq x^2 \) for all \( x \leq 1 \). We conclude that \( \mu^2 h_2 \leq \mu^3 h_3 \). Thus,
\[ \exp(-\mu^2 \sup_{x \geq 0} h_2(x, t, \mu^2)) \geq \exp(-\mu^3 \sup_{x \geq 0} h_3(x, t, \mu^2, \mu^3)). \]

From the proof of Theorem 3 the left-hand side of the last inequality equals the right-hand side of inequality (13) and the right-hand side of the last inequality equals the right-hand side of inequality (14). ■

References


