Concentration inequalities using higher moments information

Bar Light*

April 27, 2023

Abstract:

In this paper, we generalize and improve some fundamental concentration inequalities using information on the random variables’ higher moments. In particular, we 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 $p$. Importantly, 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. To prove our results, we derive novel upper bounds on the moment-generating function of a random variable that depend on the random variable’s first $p$ moments and show that these bounds satisfy appropriate convexity properties.

Keywords: Concentration inequalities, Hoeffding’s inequality, Bennett’s inequality, moment-generating function.

*Microsoft Research, NY, USA. e-mail: barlight@microsoft.com
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 and operations research (see Massart (2000)), the analysis of algorithms and machine learning theory (see Alon and Spencer (2004) and 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. Despite their importance and numerous generalizations\footnote{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), Vidyasagar (1999), Bousquet (2002), Bentkus (2004), Klein and Rö (2005), Kontorovich and Ramanan (2008), Fan et al. (2012), Junge and Zeng (2013), Pinelis (2014), Paulin (2015), Pelekis et al. (2015), Jiang et al. (2018), and Pepin (2021).}, there are not many improvements even for the basic case of sums of independent real-valued random variables (Pinelis, 2014), especially concentration bounds that use information on higher order moments and are given as a simple closed-form expression.

In this paper we generalize and 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 $p \geq 1$. 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 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

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

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, we show that the results in Talagrand (1995) can be easily adapted to our setting to obtain a concentration bound of optimal order that uses information about the random variables’ higher moments. In addition, our bounds can be generalized for martingales and other stochastic processes in a standard way.

To prove our concentration bounds we derive novel upper bounds on the random variable’s moment-generating function that depend on the random variable’s first $p$ moments. These bounds satisfy appropriate convexity properties that imply that we can derive a closed-form expression concentration bounds.

## 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 $\|X\|_p = \text{ess sup}_{\omega} |X(\omega)|$ 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. We say that $f$ is increasing if $f(x) \geq f(y)$ for all $x \geq y$.

For the rest of the paper, for every positive integer $p$, 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. We use the convention that $\sum_{j=0}^k a_j = 0$ whenever $k < 0$ so $T_1(x) = \exp(x)$. 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 \( p \) (see the proof of Theorem 1). 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 \( p \). Section 4 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)

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

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

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

(3)

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).

For a random variable \( X \) on \([0, b] \) let

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

be the right-hand side of inequality (2). The next proposition shows that for every even number \( p \) and \( s > 0 \) we have \( m_{X,s}(p) \geq m_{X,s}(p + 1) \). If, in addition, the random variable \( X \) is non-negative, then
we also have $m_{X,s}(p + 1) \geq m_{X,s}(p + 2)$, and hence, $m_{X,s}(p)$ is decreasing. Thus, for non-negative random variables, inequality (2) is tighter when $p$ increases. In particular, we have $m_{X,s}(2) \geq m_{X,s}(p)$ 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 and $s > 0$. The following statements hold:

(i) $m_{X,s}(p) \geq m_{X,s}(p + 1)$.

(ii) If $X \geq 0$ then $m_{X,s}(p + 1) \geq m_{X,s}(p + 2)$.

Note that 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 $p = 1$. 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 $p = 1$.

The upper bounds on the moment-generating function (2) are not optimal in the sense that there might be a smaller bound given the information on the random variable’s first $p$ moments. The optimal bound can be found by solving a linear program and is typically not given as a closed-form expression (see Pandit and Meyn (2006) for a discussion). The main advantage of our upper bounds is the fact that the derivative of the right-hand-side of inequality (2) with respect to $s$ is log-convex for non-negative random variables. This key convexity property is the main ingredient in deriving a closed-form Hoeffding type concentration bounds that depend on the random variables’ first $p$ moments (see the discussion after Theorem 2). For a proof of Lemma 1 see the proof of Theorem 2.

**Lemma 1** Let $p$ be a positive integer and suppose that $X$ is random variable on $[0, b]$. Then the derivative of $z(s)$ where $z(s)$ is the right-hand-side of inequality (2),

$$z(s) := \frac{\mathbb{E}(X^p)}{b^p} T_{p+1}(sb) + \mathbb{E}\left(\sum_{j=0}^{p-1} s^j X^j \right)$$

is log-convex on $(0, \infty)$, i.e., $\log(z^{(1)}(s))$ is a convex function on $(0, \infty)$.

### 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. 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 positive integer $p$, the bound is given as a simple closed-form expression that depends on the random variables’ first $p$ moments.

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

(i) For all \( t > 0 \) we have
\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( - \frac{2t^2}{\sum_{i=1}^{n} b_i^2 C_p \left( 4tb_i/D_n, b_i, \mu_i^1, \ldots, \mu_i^p \right)} \right)
\]
where
\[
C_p(y, b_i, \mu_i^1, \ldots, \mu_i^p) = \frac{\mu_i^p \exp(y) + \sum_{j=0}^{p-3} \frac{y^j}{j!} \left( b_i^p - j \mu_i^j + \mu_i^p \right)}{\mu_i^p \exp(y) + \sum_{j=0}^{p-2} \frac{y^j}{j!} \left( b_i^p - j + 1 \mu_i^j - \mu_i^p \right)}
\]
for \( i = 1, \ldots, n \) 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 \( C_1 = 1 \).

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
\[
\mathbb{P} \left( \sum_{i=1}^{n} Y_i - \mathbb{E} \left( \sum_{i=1}^{n} Y_i \right) \geq t \right) = \mathbb{P}(S_n - \mathbb{E}(S_n) \geq t)
\]
\[
\leq \exp \left( - \frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2 C_p \left( 4tb_i - a_i \right)/D_n, b_i - a_i, \mu_i^1, \ldots, \mu_i^p \right) \right)
\]
Note that \( \mu_i^k = \mathbb{E}(Y_i - a_i)^k \) for \( i = 1, \ldots, n \) and \( k = 1, \ldots, p \).

Applying the last inequality to \(-Y_i\) and using the union bound yield
\[
\mathbb{P} \left( \sum_{i=1}^{n} Y_i - \mathbb{E} \left( \sum_{i=1}^{n} Y_i \right) \geq t \right) \leq 2 \exp \left( - \frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2 C_p \left( t, Y_i, b_i - a_i \right) \right) \right)
\]
where
\[
C_p(t, Y_i, d_i) = \max \left\{ C_p \left( 4td_i/\sum_{i=1}^{n} (\lambda_i^2/\lambda_i)^2, d_i, \lambda_i^1, \ldots, \lambda_i^p \right), C_p \left( 4td_i/\sum_{i=1}^{n} (\mu_i^2/\mu_i)^2, d_i, \mu_i^1, \ldots, \mu_i^p \right) \right\}
\]
with \( \lambda_i^k = \mathbb{E}(Y_i - a_i)^k \) and \( \mu_i^k = \mathbb{E}(b_i - Y_i)^k \)

(ii) If \( X_1, \ldots, X_n \) are identically distributed then inequality (7) yields
\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq nt) \leq \exp \left( - \frac{2nt^2}{b_i^2 C_p \left( 4tb_i/d(X_i), b_i, \mu_i^1, \ldots, \mu_i^p \right) \right) \right)
\]

(iii) In some cases of interest only a bound on the random variables’ higher moments is a available. Theorem 2 part (i) holds also under the condition \( \mathbb{E}(X_i^k) \leq \mu_i^k \) for \( k = 2, \ldots, p \), \( i = 1, \ldots, n \) as long as \( (\mu_i^{d+1})^2 \leq \mu_i^{d+2} \mu_i^{d+1} \) for \( i = 1, \ldots, n \) and \( d = 1, \ldots, p - 2 \).

(iv) Our results can be extended in a standard way for martingales and other stochastic processes such as Markov chain (see Freedman (1975)). For the sake of brevity we omit the details.

The approach. We now discuss the sketch of the proof of Theorem 2 part (i). The full proof is in Section 4. Fix a positive integer \( p \). We start with a random variable \( X \) on \([0, b]\). Assume for
simplicity that \( b = 1 \). Suppose that we have \( \mathbb{E} \exp(yX) \leq v(y) \) for all \( y \geq 0 \) where \( v(y) \) is some bound on the moment generating function of \( X \). Let \( g(y) = \ln(v(y)) \). Then using Taylor’s theorem we can proceed as in the proof of Hoeffding’s inequality \(^{[1]}\) to show that
\[
\mathbb{E} \exp(sX - s\mathbb{E}(X)) \leq \exp(0.5s^2 \max_{0 \leq y \leq s} g''(y)).
\]
Let \( C(s) = \max_{0 \leq y \leq s} \frac{v''(y)}{v(y)} \). We have
\[
\max_{0 \leq y \leq s} g''(y) = \max_{0 \leq y \leq s} \frac{v''(y)}{v(y)} \left( 1 - \frac{v''(y) (v'(y))^2}{v(y)(v''(y))^2} \right) \leq \max_{0 \leq y \leq s} \frac{v''(y)}{v(y)} \left( 1 - \frac{v''(y)}{v(y) C(s)} \right) \leq 0.25 C(s)
\]
where the second inequality follows from the elementary inequality \( x(1-x) \leq \frac{1}{4} \) for all \( x > 0 \) and \( x < 1 \). Hence, an essential step in deriving a closed-form exponential bound on the moment generating function is to find a function \( v \) that induces a simple closed-form expression for the function \( C \). This is exactly where Theorem 1 is useful. Suppose that \( v(y) \) is the right-hand side of inequality (2) (for some \( p \)). Then \( v(y) \) provides a bound on the moment generating function that depends on the random variable’s first \( p \) moments. A key step in the proof of Theorem 2 is to use Lemma 1, i.e., to use the fact that \( v(y) \) is a log-convex function. This implies that \( v''(y) / v(y) \) is increasing so \( C(s) = (v''(y) / v(y) (s))^2 \) is given in a closed-form expression. This key step shows the usefulness of Theorem 1 for deriving closed-form concentration bounds using higher moments information. With this bound we can conclude that
\[
\mathbb{E} \exp(sX - s\mathbb{E}(X)) \leq \exp(s^2 C(s)/8).
\]
Applying the Chernoff bound and choosing a specific value for \( s \) proves Theorem 2 part (i).

**A simple special case.** The calculation of \( C_p \) in inequality (1) is immediate. For example, for \( p = 2 \) we have
\[
C_2(x, b_i, \mu_i^1, \mu_i^2) = \left( \frac{\mu_i^2 \exp(x)}{\mu_i^1 \exp(x) + b_i \mu_i^1 - \mu_i^2} \right)^2
\]
and for \( p = 3 \) we have
\[
C_3(x, b_i, \mu_i^1, \mu_i^2, \mu_i^3) = \left( \frac{\mu_i^3 \exp(x) + b_i \mu_i^2 - \mu_i^3}{\mu_i^3 \exp(x) + b_i^2 \mu_i^1 - \mu_i^3 + (b_i \mu_i^2 - \mu_i^3) x} \right)^2
\]
for all \( i = 1, \ldots, n \) and all \( x \geq 0 \).

The dependence of \( C_p \) on the first argument in Theorem 2 can be simplified. For a random variable \( X \) on \([0,1]\), a positive integer \( p \) and \( c > 0 \) let
\[
I_p(X, c) = \frac{T_p(c) \mathbb{E}(X^p) + \sum_{j=0}^{p-2} \mathbb{E}(X^{j+1}) c^j / j!}{T_{p-1}(c) \mathbb{E}(X^p) + \sum_{j=0}^{p-3} \mathbb{E}(X^{j+2}) c^j / j!}
\]
where \((x)_+ = \max(x,0)\). We have \( I_p \geq 1 \) for every positive integer \( p \). The function \( I_p \) can be interpreted as a measure for the usefulness of knowing the random variable’s first \( p \) moment given that we know the first \( p-1 \) moments. For example, \( I_1(X, c) = 1 \) for every random variable \( X \) and
\[
I_2(X, 1) = \frac{\exp(1) \mathbb{E}(X^2) + \mathbb{E}(X) - \mathbb{E}(X^2)}{\exp(1) \mathbb{E}(X^2)} = \frac{\exp(1) - 1}{\exp(1)} + \frac{1}{\exp(1) \mathbb{E}(X^2)} \mathbb{E}(X)
\]
equals 1 if knowing \( \mathbb{E}(X^2) \) is not useful at all given the knowledge of \( \mathbb{E}(X) \) (because \( \mathbb{E}(X^2) \) is bounded above by \( \mathbb{E}(X) \), then the highest second moment possible is \( \mathbb{E}(X) \) which yields \( I_2 = 1 \)) and is greater than 1 when the second moment provides useful information, i.e., the first two moments differ. Simi-
larly, \[
I_3(X, 1) = \frac{\exp(1)\mathbb{E}(X^3) + \mathbb{E}(X^2) + \mathbb{E}(X) - 2\mathbb{E}(X^3)}{\exp(1)\mathbb{E}(X^3) + \mathbb{E}(X^2) - \mathbb{E}(X^3)}
\]
equals \(I_2(X, 1)\) when the information about the third moment is not useful (i.e., \(\mathbb{E}(X^3) = \mathbb{E}(X^2)\)).

We now provide a concentration bounds that depend on \(I_p\) and simplify the concentration bounds in Theorem 2 when \(t\) is relatively small by using the fact that \(C_p\) is increasing in the first argument.

**Corollary 1** Let \(X_1, \ldots, X_n\) be independent random variables where \(X_i\) is a random variable on \([0, 1]\). Let \(S_n = \sum_{i=1}^n X_i\). Assume that \(\mathbb{E}(X_i^k) = \mathbb{E}(X_j^k) = \mathbb{E}(X^k) > 0\) for \(k = 1, \ldots, p\) and \(i, j = 1, \ldots, n\). Let \(c > 0\) and suppose that
\[
t \leq c \left( \frac{\mathbb{E}(X^2)}{2\mathbb{E}(X)} \right)^2.
\]
Then for every positive integer \(p\) we have
\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq nt) \leq \exp(-2nt^2(I_p(X, c))^2).
\]

**The missing factor in Hoeffding’s inequality.** Consider the function
\[
\theta(x) := \frac{1}{(2\pi)^{0.5}} \frac{1}{x} \frac{\exp(x^2/2)}{\int_x^{\infty} \exp(-u^2/2)du}
\]
Then the central limit theorem and the fact that (see [Talagrand (1995)])
\[
\frac{1}{(2\pi)^{0.5}(1 + x)} \leq \theta(x) \leq \frac{1}{(2\pi)^{0.5}x}
\]
for all \(x > 0\) imply that there is a missing factor in Hoeffding’s inequality [Talagrand 1995]. We use the results in [Talagrand 1995] together with Theorem 2 to derive concentration bounds of optimal order that depend on the random variables’ first \(p\) moments.

**Corollary 2** Let \(X_1, \ldots, X_n\) be independent random variables where \(\mathbb{E}(X_i) = 0\) and \(|X_i| \leq b\) for \(i = 1, \ldots, n\). Let \(S_n = \sum_{i=1}^n X_i\) and let \(p \geq 1\) be an integer.

There exists a large universal constant \(K\) such that for \(0 \leq t \leq \sigma^2/Kb\) we have
\[
\mathbb{P}(S_n \geq t) \leq \exp\left(-\frac{t^2}{2 \sum_{i=1}^n b_i^2 C_p(8tb_i/D_n, 2b_i, \mu_i^1, \ldots, \mu_i^p)}(\theta \left( \frac{t}{\sigma} \right) + \frac{Kb}{\sigma})\right)
\]
where \(\mu_i^k = \mathbb{E}(X_i + b_i)^k\) for \(i = 1, \ldots, n, k = 1, \ldots, p, \sigma^2 =: \mathbb{E}(\sum_{i=1}^n X_i)^2\).

**A limiting case.** Under the conditions of Theorem 2, when \(p\) tends to infinity then we need knowledge on all the moments. In other words, we need knowledge on the moment generating function of the random variables under consideration. The following Corollary provides an exponential bound for the case that \(p\) tends to infinity.

**Corollary 3** Under the conditions and notations of Theorem 2 we have
\[
\lim_{p \to \infty} C_p(x, b_i, \mu_i^1, \ldots, \mu_i^p) = \frac{1}{b_i^2} \left( \frac{\mathbb{E}(X_i^2)}{\mathbb{E}(X_i)} \exp(xX_i/b_i) \right)^2
\]
\[\text{Note that}
\theta \left( \frac{t}{\sigma} \right) + \frac{Kb}{\sigma} \leq K' \frac{\sigma}{t}
\]
for some constant \(K'\) so inequality 11 provides a concentration bound of optimal order.
for all $i = 1, \ldots, n$ and all $x \geq 0$. Theorem 2 implies that for all $t > 0$ we have
\[ P(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( - \frac{2t^2}{\sum_{i=1}^{n} \left( \frac{\mathbb{E}(X_i^2) \exp(4tX_i/D_n)}{\mathbb{E}(X_i) \exp(4tX_i/D_n)} \right)^2} \right). \quad (12) \]

Note that inequality (12) does not depend on $b_i$.

**Examples.** We now provide examples where our results significantly improve Hoeffding’s inequality. The first example studies a sum of uniform random variables and the second example studies confidence intervals.

**Example 1 (Uniform distribution).** Suppose that $X_1, \ldots, X_n$ are independent continuous uniform random variables on $[0, 1]$, i.e., $P(X_i \leq t) = t$ 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 Corollary 3 we have
\[ \lim_{p \to \infty} C_p(x, 1, \mu_1^1, \ldots, \mu_1^n) = \left( \frac{-2 + \exp(x)(2 - 2x + x^2)}{x(1 + \exp(x)(x-1))} \right)^2 := C_{\infty}(x). \]

Using the fact that $d(X_i) = \left( \mathbb{E} X_i^2 / \mathbb{E} X_i \right)^2 = 4/9$ inequality (12) yields
\[ P(S_n - \mathbb{E}(S_n) \geq nt) \leq \exp \left( - \frac{2nt^2}{C_{\infty}(9t)} \right). \quad (13) \]

In Figure 1 we plot the bound given in Hoeffding’s inequality (see Theorem 2 inequality (6)) for $P(S_n - \mathbb{E}(S_n) \geq nt)$ divided by the bound given in (13) as a function of $t$ on the interval $[0.1, 0.4]$ for $n = 40$. We see that the bound given in (13) significantly improves Hoeffding’s bound.

![Figure 1](image-url)

**Example 2 (Confidence intervals).** Suppose that we want to know how many independent and identically distributed samples $n$ are required to acquire a $(1 - \alpha)$ confidence interval around the mean
\[ \sum_{i=1}^{n} X_i/n \text{ of size } t. \] That is, suppose that we want to find an \( n \) such that
\[
\delta =: \mathbb{P} \left( \sum_{i=1}^{n} X_i/n \notin \left[ \mathbb{E} \left( \sum_{i=1}^{n} X_i/n \right) - t, \mathbb{E} \left( \sum_{i=1}^{n} X_i/n \right) + t \right] \right)
\]
is at most \( \alpha \) where \( X_i \) are independent and identically distributed random variables on \([a,b]\). Then from Remark 2 part (i) for every \( n \) that satisfies
\[
n \geq \frac{\ln(2/\alpha)(b-a)^2C_p(t,X_i,b-a)}{2t^2}
\]
we have \( \delta \leq \alpha \). This improves the confidence interval derived by Hoeffding’s inequality (which corresponds to \( C_p = 1 \)) by a factor of \( C_p \) that is given in closed-form. This factor depends on the information that the random variable’s higher moments provide. When the higher moments provide useful information and \( t \) is small then this factor is significant and we need less observations to achieve a significant level of \( \alpha \).

As an example suppose that \( X_i \) is a random variable on the interval \([0,1]\) with \( \mathbb{E}(X_i) = 1/2 \) and \( \mathbb{E}(X_i^2) = 1/3 \). Upon examination, it can be easily checked that \( C_2 < 1/2 \) holds for a small value of \( t \) (\( t < 0.02 \)). By applying Theorem 2, we find that fewer than half the samples needed when using Hoeffding’s bound are required to obtain a \((1 - \alpha)\) confidence interval using the generalized Hoeffding’s bound with \( p = 2 \).

### 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:

\[
\alpha_0 - \sum_{j=1}^{p} \alpha_j x^j = \exp(x). \tag{14}
\]

The solutions to equation (14) 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 (14) has a positive solution. We denote the non-empty set of positive solutions of equation (14) 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)).

Finding the positive solutions of the transcendental equation (14) 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 (14) 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 \( \mathbb{E}(X_i) = \mu_i^1 \), 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 some \( \mu_i^p \) for all \( k = 1, \ldots, p-1 \) and all \( i = 1, \ldots, n \).

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

\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( \max_{y \in G_{p-2}(\alpha_0, \ldots, \alpha_{p-2})} \left( \frac{t}{b} - \left( \frac{t}{b} + \frac{\mu^2}{b^2} \right) y + \sum_{j=2}^{p-1} \left( \frac{\mu^j}{b^j j!} - \frac{\mu^{j+1}}{b^{j+1} j!} \right) y^j \right) \right) \tag{15}
\]

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

(ii) If \( \mu^1 \geq 0 \) for every odd number \( j = 3, \ldots, p-1 \) (for example one can choose \( \mu^j = \max(\sum_{i=1}^n \mu_i^j, 0) \)) then \( G_{p-2} \) consists of one element and inequality (15) reduces to

\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( \frac{t}{b} - \left( \frac{t}{b} + \frac{\mu^2}{b^2} \right) y + \sum_{j=2}^{p-1} \left( \frac{\mu^j}{b^j j!} - \frac{\mu^{j+1}}{b^{j+1} j!} \right) y^j \right) \tag{16}
\]

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 (15) reduces to Bennett’s inequality:

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

\[
= \exp \left( -\frac{\mu^2}{b^2} \ln \left( \frac{tb}{\mu^2} + 1 \right) - \frac{bt}{\mu^2} \right). \tag{17}
\]

(iv) Suppose that \( p = 3, \alpha_1 \neq 0, \) and \( \mathbb{E}(\max(X_i^3, 0)) = \mu_i^3 \) for all \( i = 1, \ldots, n \). Then \( G_1(\alpha_0, \alpha_1) = \left\{ \frac{\alpha_0}{\alpha_1} - W(\exp(\frac{\alpha_0}{\alpha_1})/\alpha_1) \right\} \) consists of one element and inequality (15) reduces to

\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( \frac{t}{b} - \left( \frac{t}{b} + \frac{\mu^2}{b^2} \right) y + \left( \frac{\mu^2}{2b^2} - \frac{\mu^3}{2b^2} \right) y^2 \right) \tag{18}
\]

where \( y = \frac{\alpha_0}{\alpha_1} - W(\exp(\frac{\alpha_0}{\alpha_1})/\alpha_1) \) 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.
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 these bounds and their convexity properties, 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.

4 Proofs

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 \( T_p = T_{p-1} \) for all \( p \geq 2 \)).

We now show that \( g(x) \leq g(0) \) for all \( x < 0 \).

Suppose 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 \( T_{p+1}(x)/x^p \geq 1/p! \) 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 \( T_k^{(1)} = T_k \) 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

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

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

Note that

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

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

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

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

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

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

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

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 $m_{X,s}(p+1) \geq m_{X,s}(p+2)$ if and only if $bE X^{p+1} \geq E X^{p+2}$ 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.** We will use the following notations in proof. Let $X$ be a random variable on $[0, b]$ with $P(X > 0) > 0$. Denote $E(X^k) = \mu_k$ 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

$$C_p(x, b, \mu^1, \ldots, \mu^p) = \max_{0 \leq y \leq x} \frac{(v^{(2)}(y, b, \mu^1, \ldots, \mu^p))^2}{(v^{(1)}(y, b, \mu^1, \ldots, \mu^p))^2} \quad (19)$$

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

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

Thus, for $p \geq 2$ we have

$$v^{(1)}(0, b, \mu^1, \ldots, \mu^p) = \frac{\mu^1}{b} > 0 \text{ and } v^{(2)}(0, b, \mu^1, \ldots, \mu^p) = \frac{\mu^2}{b^2} > 0.$$ 

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 $y \geq 0$.

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 $p$ the function $v^{(1)}(y, b, \mu^1, \ldots, \mu^p)$ is log-convex in $y$ on $(0, \infty)$ (i.e., $\log(v^{(1)}(y, b, \mu^1, \ldots, \mu^p))$ is a convex function on $(0, \infty)$).

**Proof of Step 2.** Fix a positive integer $p$. Let

$$w(y) = T_p(y) + \frac{\mu^p}{b^p} \sum_{j=0}^{p-2} \frac{\mu^{j+1} y^j}{b^{j+1} j!} = \frac{\mu^p v^{(1)}(y, b, \mu^1, \ldots, \mu^p)}{\mu^p}.$$ 

To prove Step 2 it is enough to prove that $w$ is log-convex on $(0, \infty)$. 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 $j = 1, \ldots, p$. 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 $(0, \infty)$ if and only if $w^{(1)}/w$ is increasing on $(0, \infty)$. 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 $(0, \infty)$ 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 $w_0(y) = \exp(y) + \beta_{p-1}$ is log-convex on $(0, \infty)$.

Assume that $w_k$ is log-convex on $(0, \infty)$ for some integer $0 \leq k \leq p - 3$. We show that $w_{k+1}$ is log-convex on $(0, \infty)$. Log-convexity of $w_k$ implies that the function

$$\frac{w_k^{(1)}(y)}{w_k(y)} = \frac{\exp(y) + \sum_{j=0}^{k-1} \frac{y^j}{j!} \beta_{p-j+k}}{\exp(y) + \sum_{j=0}^{k} \frac{y^j}{j!} \beta_{p-1+j-k}}$$

(20)

is increasing on $(0, \infty)$. Using the fact that $w_{k+1}^{(1)} = w_k$ and applying Step 2 in the proof of Theorem

we conclude that the function

$$m(y) = \frac{w_k(y) - w_k(0)}{w_{k+1}(y) - w_{k+1}(0)} = \frac{\exp(y) + \sum_{j=0}^{k} \frac{y^j}{j!} \beta_{p-1+j-k} - (1 + \beta_{p-1-k})}{\exp(y) + \sum_{j=0}^{k+1} \frac{y^j}{j!} \beta_{p-2+j-k} - (1 + \beta_{p-2-k})}$$

is increasing on $(0, \infty)$. Thus, $m^{(1)}(y) \geq 0$ for all $y \in (0, \infty)$. That is,

$$w_{k+1}^{(2)}(y)w_{k+1}(y) - w_{k+1}^{(1)}(y)(1 + \beta_{p-2-k}) \geq (w_{k+1}^{(1)}(y))^2 - w_{k+1}^{(1)}(y)(1 + \beta_{p-1-k})$$

(21)

for all $y \in (0, \infty)$. We now show that $w_{k+1}^{(2)}(y)w_{k+1}(y) \geq (w_{k+1}^{(1)}(y))^2$. Because $w_{k+1}^{(2)}/w_{k+1}^{(1)}$ is increasing and positive (see [20]) we have

$$\frac{w_{k+1}^{(2)}(y)}{w_{k+1}^{(1)}(y)}(1 + \beta_{p-2-k}) \geq (1 + \beta_{p-1-k})$$

(22)

for all $y \in (0, \infty)$ if the last inequality holds for $y = 0$, i.e., if

$$(1 + \beta_{p-k})(1 + \beta_{p-2-k}) \geq (1 + \beta_{p-1-k})^2 \iff \left( \frac{b^k \mu^{p-k}}{\mu^p} \right) \left( \frac{b^{k+2} \mu^{p-2-k}}{\mu^p} \right) \geq \left( \frac{b^{k+1} \mu^{p-k-1}}{\mu^p} \right)^2 \iff \mu^{p-k} \mu^{p-k-2} \geq (\mu^{p-1})^2$$

which holds from Step 1. We conclude that inequality [22] holds. Using inequality [21] we have

$$w_{k+1}^{(2)}(y)w_{k+1}(y) - w_{k+1}^{(1)}(y) \geq w_{k+1}^{(2)}(y) - w_{k+1}^{(1)}(y)(1 + \beta_{p-2-k})$$

That is, $w_{k+1}^{(2)}(y)w_{k+1}(y) \geq (w_{k+1}^{(1)}(y))^2$ for all $y \in (0, \infty)$. We conclude that $w_{k+1}^{(2)}/w_{k+1}^{(1)}$ is increasing on $(0, \infty)$, 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

$$C_p(x, b, \mu^1, \ldots, \mu^p) = \frac{\exp(x) + \sum_{j=0}^{p-3} \frac{x^j}{j!} \left( \frac{b^p - j \mu^{j+1}}{\mu^p} - 1 \right)}{\exp(x) + \sum_{j=0}^{p-2} \frac{x^j}{j!} \left( \frac{b^{p-j-1} \mu^{j+1}}{\mu^p} - 1 \right)}$$

for all $x \geq 0$ where $C_p$ is given in Equation [19].

**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

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

15
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 \( x > 0 \) we have

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

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.** For \( s \geq 0 \) we have

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

**Proof of Step 5.** From Theorem \( \square \) for all \( s \geq 0 \) we have

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

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 \( v^{(1)}(0, b, \mu^1, \ldots, \mu^p) = \mathbb{E}(X)/b \). 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 \( g(y) = g(0) + yg^{(1)}(0) + 0.5y^2 g^{(2)}(z) \). Thus, using the fact that \( y = sb \) we have

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

where

\[
V(sb, b, \mu^1, \ldots, \mu^p) = \sup_{0 \leq z \leq y} g^{(2)}(z).
\]
Using $\mathbb{E}\exp(sX) \leq \exp(g(y))$ and Step 4 imply
\[
\mathbb{E}\exp(s(X - \mathbb{E}(X))) \leq \exp\left(\frac{s^2b^2V(sb, b, \mu^1, \ldots, \mu^p)}{2}\right) \\
\leq \exp\left(\frac{s^2b^2C_p(sb, b, \mu^1, \ldots, \mu^p)}{8}\right).
\]

**Step 6.** For all $t > 0$ we have
\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp\left(-\frac{t^2}{2 \sum_{i=1}^{n} b_i^2 C_p\left(\frac{4tb_i}{\sum_{j=1}^{n} d(X_j)}, b_i, \mu_i^1, \ldots, \mu_i^p\right)}\right).
\]

**Proof of Step 6.** Using independence, Step 5, and Markov’s inequality, a standard argument shows that:
\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp(-st)\mathbb{E}\exp(s(S_n - \mathbb{E}(S_n))) \\
= \exp(-st) \prod_{i=1}^{n} \mathbb{E}\exp(s(X_i - \mathbb{E}(X_i))) \\
\leq \exp(-st) \prod_{i=1}^{n} \exp\left(\frac{s^2b_i^2C_p(sb_i, b_i, \mu_i^1, \ldots, \mu_i^p)}{8}\right) \\
= \exp\left(-st + \frac{s^2}{8} \sum_{i=1}^{n} b_i^2 C_p(sb_i, b_i, \mu_i^1, \ldots, \mu_i^p)\right).
\]

Let
\[
s = \frac{4t}{\sum_{i=1}^{n} b_i^2 C_p\left(\frac{4tb_i}{\sum_{j=1}^{n} d(X_j)}, b_i, \mu_i^1, \ldots, \mu_i^p\right)}.
\]

Note that
\[
C_p(0, b_i, \mu_i^1, \ldots, \mu_i^p) = \left(\frac{\mu_i^2}{\mu_i^1 b_i}\right) = \frac{d(X_i)}{b_i^2}.
\]

Because $C_p$ is increasing in the first argument for all $y \geq 0$, we have
\[
C_p(y, b_i, \mu_i^1, \ldots, \mu_i^p) \geq \frac{d(X_i)}{b_i^2} > 0.
\]

Thus,
\[
sb_i = \frac{4tb_i}{\sum_{i=1}^{n} b_i^2 C_p\left(\frac{4tb_i}{\sum_{j=1}^{n} d(X_j)}, b_i, \mu_i^1, \ldots, \mu_i^p\right)} \leq \frac{4tb_i}{\sum_{j=1}^{n} d(X_j)}.
\]

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

We conclude that
\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} b_i^2 C_p\left(\frac{4tb_i}{\sum_{j=1}^{n} d(X_j)}, b_i, \mu_i^1, \ldots, \mu_i^p\right)}\right),
\]
which proves Step 6.
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 \( v^{(2)}(y, b, \mu_1, \ldots, \mu^p) \leq v^{(1)}(y, b, \mu_1, \ldots, \mu^p) \) for all \( y \geq 0 \).

We have \( v^{(2)}(y, b, \mu_1, \ldots, \mu^p) \leq v^{(1)}(y, b, \mu_1, \ldots, \mu^p) \) 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 \( b^{i+1} x^{i+2} \leq x^{i+1} b^{i+2} \). 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 \( v^{(2)}(y, b, \mu_1, \ldots, \mu^p) \leq v^{(1)}(y, b, \mu_1, \ldots, \mu^p) \) 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)).

**Proof of Corollary 1** Under the Corollary’s assumption we have \( 4t/d(X_i) \leq c \). Because the function \( C_p \) is increasing in the first argument (see the proof of Theorem 2) we have

\[
(I_p(X_i, c))^2 = \frac{1}{C_p(c, 1, \mathbb{E}(X_i), \ldots, \mathbb{E}(X_i^p))} \leq \frac{1}{C_p(4t/d(X_i), 1, \mathbb{E}(X_i), \ldots, \mathbb{E}(X_i^p))}.
\]

Combining the inequality above and Remark 4 part (ii) proves the result.

**Proof of Corollary 2** Theorem 3.3 in [Talagrand (1995)] shows that there exists a universal constant \( K \) such that for all \( t \leq \sigma^2/Kb \) we have

\[
\mathbb{P}(S_n \geq t) \leq \inf_{s \geq 0} \exp \left( \sum_{i=1}^{n} \ln \mathbb{E} \exp(s X_i) - st \right) \left( \theta \left( \frac{t}{\sigma} \right) + K b \right).
\]

The result now follows from Steps 5 and 6 in the proof of Theorem 2.

**Proof of Corollary 3** 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 \( 0 \leq \lim_{p \to \infty} b^{-p} \mu^p T_p(y) = \lim_{p \to \infty} T_p(y) = 0 \). Because \( Z \) is a random variable on
we can use the bounded convergence theorem to conclude that
\[ \lim_{p \to \infty} v^{(1)}(y, b, \mu^1, \ldots, \mu^p) = \lim_{p \to \infty} \left( \frac{\mu^p}{b^p} T_p(y) + \sum_{j=0}^{p-2} \frac{y^j \mu^{j+1}}{b^{j+1} j!} \right) = b^{-1} E Z \exp(y Z/b). \]

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

We conclude that
\[ \lim_{p \to \infty} \left( \frac{v^{(2)}(y, b, \mu^1, \ldots, \mu^p)}{v^{(1)}(y, b, \mu^1, \ldots, \mu^p)} \right)^2 = \left( \frac{b^{-2} E Z^2 \exp(y Z/b)}{b^{-1} E Z \exp(y Z/b)} \right)^2 = b^{-2} \left( \frac{E Z^2 \exp(x Z/b)}{E Z \exp(x Z/b)} \right)^2. \]

Using Step 3 in the proof of Theorem 2 part (i) yields
\[ \lim_{p \to \infty} C_p(x, b, \mu^1, \ldots, \mu^p) = \lim_{p \to \infty} \max_{0 \leq y \leq x} \left( \frac{v^{(2)}(y, b, \mu^1, \ldots, \mu^p)}{v^{(1)}(y, b, \mu^1, \ldots, \mu^p)} \right)^2 = \left( \frac{b^{-2} E Z^2 \exp(x Z/b)}{E Z \exp(x Z/b)} \right)^2. \]

which proves the result.

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
\[ \mathbb{E} \exp(s X_i) \leq \mu_i^p \left( \exp(s) - \sum_{j=0}^{p-1} \frac{s^j}{j!} \right) + 1 + \sum_{j=1}^{p-1} \frac{s^j \mu_i^j}{j!} \]
\[ \leq \exp \left( \mu_i^p \left( \exp(s) - \sum_{j=0}^{p-1} \frac{s^j}{j!} \right) + \sum_{j=1}^{p-1} \frac{s^j \mu_i^j}{j!} \right) \]
\[ = \exp \left( \mu_i^p T_{p+1}(s) + \sum_{j=2}^{p-1} \frac{s^j \mu_i^j}{j!} \right). \]

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

From the Chernoff bound and the fact that \(X_1, \ldots, X_n\) are independent random variables, for all \(t > 0\), we have
\[
P(S_n - \mathbb{E}(S_n) \geq t) \leq \inf_{s \geq 0} \exp(-st) \mathbb{E} \exp(s(S_n - \mathbb{E}(S_n))
\]
\[
= \inf_{s \geq 0} \exp(-st) \prod_{i=1}^{n} \mathbb{E} \exp(s(X_i - \mathbb{E}(X_i))
\]
\[
\leq \inf_{s \geq 0} \exp \left( -st + \sum_{j=2}^{p-1} \frac{s^j \mu_j^i}{j!} + \mu_p^p T_{p+1}(s) \right)
\]
\[
= \exp \left( -\sup_{s \geq 0} \left( st - \sum_{j=2}^{p-1} \frac{s^j \mu_j^i}{j!} - \mu_p^p T_{p+1}(s) \right) \right)
\]
\[
= \exp \left( -\mu_p^p \sup_{x \geq 0} h_p(x, t, \mu_2^i, \ldots, \mu_p^i) \right)
\]
where
\[
h_p(x, t, \mu_2^i, \ldots, \mu_p^i) = \frac{t}{\mu_p^p} x - \frac{1}{\mu_p^p} \sum_{j=2}^{p-1} \frac{x^j \mu_j^i}{j!} - T_{p+1}(x)
\]
\[
= 1 + \left( \frac{t}{\mu_p^p} + 1 \right) x - \sum_{j=2}^{p-1} \left( \frac{\mu_j^i}{\mu_p^p j!} - \frac{1}{j!} \right) x^j - \exp(x).
\]
Because \(h_p\) is continuous, \(h_p(0, t, \mu_2^i, \ldots, \mu_p^i) = 0\), and \(\lim_{x \to \infty} h_p(x, t, \mu_2^i, \ldots, \mu_p^i) = -\infty\), the function \(h_p\) has a maximizer. Let \(h_p^{(j)}\) the \(j\)th derivative of \(h_p\) with respect to \(x\).

Note that
\[
h_p^{(1)}(x, t, \mu_2^i, \ldots, \mu_p^i) = \frac{t}{\mu_p^p} + 1 - \sum_{j=1}^{p-2} \left( \frac{\mu_j^i + 1}{\mu_p^p j!} - \frac{1}{j!} \right) x^j - \exp(x)
\]
\[
= \alpha_0 - \sum_{j=1}^{p-2} \alpha_j x^j - \exp(x)
\]
Thus, \(h_p^{(1)}(0, t, \mu_2^i, \ldots, \mu_p^i) = \alpha_0 - \exp(0) > 0\) and \(h_p^{(1)}(x, t, \mu_2^i, \ldots, \mu_p^i) < 0\) for all \(x \geq \bar{x}\) for some large \(\bar{x}\). Because \(h_p^{(1)}\) is continuous we conclude that a maximizer \(y\) of \(h_p\) on \([0, \infty)\) satisfies \(h_p^{(1)}(y, t, \mu_2^i, \ldots, \mu_p^i) = 0\).
0, that is, $y \in G_{p-2}(\alpha_0, \ldots, \alpha_{p-2})$. Plugging $y$ into $h_p$ yields
\[
\left( \frac{t}{\mu^p} + 1 \right) y - \sum_{j=2}^{p-1} \left( \frac{\mu^j}{\mu^p j!} - \frac{1}{j!} \right) y^j + 1 = -\frac{t}{\mu^p} + \left( \frac{t}{\mu^p} + \frac{\mu^2}{\mu^p} \right) y - \sum_{j=2}^{p-1} \left( \frac{\mu^j}{\mu^p j!} - \frac{\mu^{j+1}}{j!} \right) y^j.
\]
In the first equality we used the fact that $h_p^{(1)}(y, t, \mu^2, \ldots, \mu^p) = 0$. Thus,
\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \exp \left( -\mu^p \left( -\frac{t}{\mu^p} + \left( \frac{t}{\mu^p} + \frac{\mu^2}{\mu^p} \right) y - \frac{1}{\mu^p} \sum_{j=2}^{p-1} \left( \frac{\mu^j}{j!} - \frac{\mu^{j+1}}{j!} \right) y^j \right) \right)
\]  
\begin{equation}
(25)
\end{equation}
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 (25) for the random variables $Y_1, \ldots, Y_n$ to conclude that for all $t > 0$ we have
\[
\mathbb{P}(S_n - \mathbb{E}(S_n) \geq t) \leq \mathbb{P} \left( \sum_{i=1}^n Y_i - \mathbb{E} \left( \sum_{i=1}^n Y_i \right) \geq \frac{t}{b} \right) \leq \exp \left( \max_{y \in G_{p-2}(\alpha_0, \ldots, \alpha_{p-2})} \left( t \left( \frac{t}{b^2} + \frac{\mu^2}{b^2} \right) - \sum_{j=2}^{p-1} \left( \frac{\mu^j}{j!} - \frac{\mu^{j+1}}{j!} \right) y^j \right) \right)
\]
where
\[
\alpha_0 = 1 + \frac{tb^{p-1}}{\mu^p} > 1 \text{ and } \alpha_j = \frac{tb^{p-j-1} \mu^{j+1}}{\mu^p j!} - \frac{1}{j!}
\]
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
\[
h_p^{(1)}(x, t, \mu^2, \ldots, \mu^p) = \frac{t}{\mu^p} - \frac{1}{\mu^p} \sum_{j=1}^{p-2} x^j \mu^{j+1} j! - T_p(x)
\]
so if $\mu^j \geq 0$ 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 $\alpha_0 = \exp(x) = \ln(\alpha_0)$. Thus, $G_2(\alpha_0) = \{y\}$ where
\[
y = \ln \left( 1 + \frac{tb}{\mu^2} \right).
\]
Plugging $y$ into inequality (16) 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 inequality (16) proves part (iv).\[\blacksquare\]

References

ALON, N. AND J. H. SPENCER (2004): The probabilistic method, John Wiley & Sons.
ANDERSON, G., M. VAMANAMURTHY, AND M. VUORINEN (1993): “Inequalities for quasiconformal mappings in space,” Pacific Journal of Mathematics, 160, 1–18.
BENNERT, G. (1962): “Probability inequalities for the sum of independent random variables,” Journal of the American Statistical Association, 57, 33–45.
BENTKUS, V. (2004): “On Hoeffding’s inequalities,” The Annals of Probability, 32, 1650–1673.
BOUCHERON, S., G. LUGOSI, AND P. MASSART (2013): Concentration inequalities: A nonasymptotic theory of independence, Oxford university press.
BOUSQUET, O. (2002): “A Bennett concentration inequality and its application to suprema of empirical processes,” Comptes Rendus Mathematique, 334, 495–500.
COHEN, A., Y. RABINOVICH, A. SCHUSTER, AND H. SHACHNAl (1999): “Optimal bounds on tail probabilities: a study of an approach,” in Advances in Randomized Parallel Computing, Springer, 1–24.
CORLESS, R. M., G. H. GONNET, D. E. HARE, D. J. JEFFREY, AND D. E. KNUTH (1996): “On the LambertW function,” Advances in Computational mathematics, 5, 329–359.
FAN, X., I. GRAMA, AND Q. LIU (2012): “Hoeffding’s inequality for supermartingales,” Stochastic Processes and their Applications, 122, 3545–3559.
FREEDMAN, D. A. (1975): “On tail probabilities for martingales,” the Annals of Probability, 100–118.
HOEFFDING, W. (1994): “Probability inequalities for sums of bounded random variables,” in The Collected Works of Wassily Hoeffding, Springer, 409–426.
JEBA, T. (2018): “A refinement of Bennett’s inequality with applications to portfolio optimization,” arXiv preprint arXiv:1804.05454.
JIANG, B., Q. SUN, AND J. FAN (2018): “Bernstein’s inequality for general Markov chains,” arXiv preprint arXiv:1805.10721.
JUNGE, M. AND Q. ZENG (2013): “Noncommutative bennett and rosenthal inequalities,” The Annals of Probability, 41, 4287–4316.
KLEIN, T. AND E. RIO (2005): “Concentration around the mean for maxima of empirical processes,” The Annals of Probability, 33, 1060–1077.
KONTOROVICH, L. A. AND K. RAMANAN (2008): “Concentration inequalities for dependent random variables via the martingale method,” The Annals of Probability, 36, 2126–2158.
Massart, P. (2000): “Some applications of concentration inequalities to statistics,” in *Annales de la Faculté des sciences de Toulouse: Mathématiques*, vol. 9, 245–303.

Mohri, M., A. Rostamizadeh, and A. Talwalkar (2018): *Foundations of machine learning*, MIT press.

Pandit, C. and S. Meyn (2006): “Worst-case large-deviation asymptotics with application to queueing and information theory,” *Stochastic processes and their applications*, 116, 724–756.

Paulin, D. (2015): “Concentration inequalities for Markov chains by Marton couplings and spectral methods,” *Electronic Journal of Probability*, 20.

Pelekis, C., J. Ramon, and Y. Wang (2015): “On the Bernstein-Hoeffding method,” *arXiv preprint arXiv:1503.02284*.

Pepin, B. (2021): “Concentration inequalities for additive functionals: A martingale approach,” *Stochastic Processes and their Applications*, 135, 103–138.

Pinelis, I. (1994): “Optimum bounds for the distributions of martingales in Banach spaces,” *The Annals of Probability*, 1679–1706.

——— (2014): “On the Bennett-Hoeffding inequality,” in *Annales de l’IHP Probabilités et statistiques*, vol. 50, 15–27.

Pinelis, I. and S. Utev (1990): “Exact exponential bounds for sums of independent random variables,” *Theory of Probability & Its Applications*, 34, 340–346.

Roussas, G. G. (1996): “Exponential probability inequalities with some applications,” *Lecture Notes-Monograph Series*, 303–319.

Scott, T. C., R. Mann, and R. E. Martinez II (2006): “General relativity and quantum mechanics: towards a generalization of the Lambert W function A Generalization of the Lambert W Function,” *Applicable Algebra in Engineering, Communication and Computing*, 17, 41–47.

Talagrand, M. (1995): “The missing factor in Hoeffding’s inequalities,” in *Annales de l’IHP Probabilités et statistiques*, vol. 31, 689–702.

Victor, H. (1999): “A general class of exponential inequalities for martingales and ratios,” *The Annals of Probability*, 27, 537–564.

Zheng, S. (2018): “An improved Bennett’s inequality,” *Communications in Statistics-Theory and Methods*, 47, 4152–4159.