The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/60231

Please be advised that this information was generated on 2019-02-02 and may be subject to change.
Optimal Hoeffding-like Inequalities
under a
Symmetry Assumption

V. Bentkus, G.D.C. Geuze, M.C.A. van Zuijlen

Report No. 0408 (June 2004)
Optimal Hoeffding-like Inequalities under a Symmetry Assumption

V. Bentkus, G.D.C. Geuze, M.C.A. van Zuijlen

Abstract

In this short paper we prove a Hoeffding-like inequality for the survival function of a sum of symmetric independent r.v. taking values in a segment \([-b, b]\) of the reals. The symmetric case is relevant to the auditing practice and is a handy case study for further investigations.

Keywords—Auditing, confidence upper bounds, tail probability upper bounds, Hoeffding inequalities, finite populations, bookkeeping.

1 Motivation

Large companies are by law obliged to publish annually a financial statement. This statement is subsequently examined and judged by an independent accountancy firm. An integral control (=audit) of every single (monetary) item in the statement is quite expensive and therefore usually omitted. Instead the external accountant (= auditor) will submit part of the company to a qualitative investigation. When too many (little) irregularities are met with or when the auditor seeks further quantitative support for his final judgement he can and often will resort to taking a representative sample of the relevant stated items. The auditor may then be interested in a confidence upper bound for a certain parameter. These bounds are based on the choice of a statistic e.g. the mean-estimator of the errors in the population (being the items in the statement). A \((1 - \alpha)\)-confidence upper bound can be constructed from the survival function of an appropriate statistic, cf [8, 9]. The quality of the confidence upper bound depends on the choice of the statistic and on the sharpness of the tail upper bound of the statistic. In this paper we construct a tail upper bound for the minimum variance unbiased linear estimator of the mean, under a symmetry assumption.

Until recently, the best known bounds for the tail probability of the sum of bounded i.i.d. random variables were the Hoeffding bounds, cf. [16]. These bounds have as a downside that they tend to be rather conservative. Different strategies can be used to circumvent this
aspect. One way is to choose a more appropriate function replacing the indicator function in the "Bernstein trick", cf. [5, 6] and see equation (2). Another approach is to consider a smaller class of laws, e.g. symmetric ones.

## 2 Notation and presentation

We begin by introducing some notation that will be used throughout the paper. For a fixed $n \in \mathbb{N}^*$ and $b \in (0, \infty)$, let $X_1, \ldots, X_n$ be i.i.d. random variables with a symmetric distribution and with $\mathbb{E}(X_i^2) = \sigma^2$, such that $-b \leq X_i \leq b$. We sometimes write for convenience’s sake $S = X_1 + \ldots + X_n$.

The symmetry is understood to be w.r.t. the vertical axis. Therefore the mean of such variables is zero. In section 3 we briefly illustrate that this is just a matter of translation.

**Definition 2.1** A law $\nu$ on a $\sigma$-algebra $\mathcal{F}$ is said to be symmetric if

$$\nu(A) = \nu(-A) \quad \text{for every} \quad A \in \mathcal{F}.$$  

Let $t \in [0, b]$. We will present a function $H$ such that the following inequality is sharp.

$$\mathbb{P}\{X_1 + \ldots + X_n \geq nt\} \leq H(t).$$  

The way we go about constructing this upper bound starts fairly the same as in [16]. Let us recapitulate a little. For the first nifty move let $h > 0$ and observe

$$\mathbb{P}\{X_1 + \ldots + X_n \geq nt\} = \mathbb{E}1_{\{S-nt>0\}} \leq \mathbb{E}e^{h(S-nt)}.$$  

The inequality is known as the "Bernstein trick". The last expression can be expanded, yielding the following.

$$\mathbb{P}\{X_1 + \ldots + X_n \geq nt\} \leq \{e^{-ht\mathbb{E}e^{hX_1}}\}^n.$$  

Eventually we want to minimize over $h$, but first we bring our symmetry property into play. Let us take a closer look at the expression at hand:

$$\mathbb{E}e^{hX_1} = \int_{-b}^{b} e^{hx} d\mathbb{P}_{X_1}(x).$$  

The device is to fold the symmetric distribution $\mathbb{P}_{X_1}$ together over the interval $[0, b]$, taking extra care of a possible mass at 0 such that this weight is counted only once. More explicitly we have to every symmetric law $\mu$ supported in $[-b, b]$ a law $\nu$ with support in $[0, b]$ (and vice versa) such that:

$$\nu(A) = \mu(-A \cup A) \quad (A \in \mathcal{B}([0, b])).$$
This operation preserves the second moment. Implementing this in formula (3) and at the same time adjusting the integrand to the symmetry gives

\[ E_{\mathbb{P}} \exp[hX_t] = \frac{\int_{-b}^{b} \frac{e^{hx} + e^{-hx}}{2} d\mathbb{P}_X(x)}{\int_{0}^{b} \frac{e^{hx} + e^{-hx}}{2} d\nu(x)}. \]  

We want to evaluate the best possible upper bound (after having established that it makes any sense at all) in this setting.

\[ \sup_{\nu \in V} \left\{ \int_{0}^{b} \cosh(hx) d\nu(x) \right\}. \]

where \( V \) is the collection of probability measures on \([0, b]\) with 2nd moment equal to \( \sigma^2 \). It will be shown that the maximizing measure is given by folding together the measure \( \frac{\sigma^2}{2b^2} \delta_{-b} + (1 - \frac{\sigma^3}{2b^2}) \delta_{0} + \frac{\sigma^3}{2b^2} \delta_{b} \). At the same time we will find that the supremum is equal to

\[ \frac{\cosh(bh) - 1}{b^2} \sigma^2 + 1. \]

So apparently we shall arrive at the following situation.

\[ P\{\sum_{i=1}^{n} X_i \geq nt\} \leq \left( e^{-ht} \left( \frac{\cosh(bh) - 1}{b^2} \sigma^2 + 1 \right) \right)^n. \]

Finally, we verify that the right hand side takes its minimum value w.r.t. \( h \) at \( h_0 \), which is

\[ h_0 = h_0(t, \sigma^2, b) = \frac{1}{b} \log \left[ \frac{b^2 - \sigma^2}{\sigma^2} \frac{t}{b - t} + \sqrt{\frac{b^2 - \sigma^2}{b - t} \left( \frac{b^2 - \sigma^2}{\sigma^2} \frac{t}{b - t} \right)^2} \right]. \]

Substituting this value in the right hand side of (7), we get the main result of this paper.

**Theorem 2.1** Let \( b > 0 \) and \( X_1, \ldots, X_n \) be i.i.d. random variables with a symmetric distribution, \( \mathbb{V}ar(X_1) = \sigma^2 \) and such that \( X_i \leq b \) for \( i \in \{1, \ldots, n\} \). Then for \( t \in [0, b] \) we have

\[ P\{\sum_{i=1}^{n} X_i \geq nt\} \leq \left( \frac{\sigma^2}{2b^2} V(t) + \frac{b^2 - \sigma^2}{b^2} V(t) \right)^n, \]

where \( V(t) \) is short for

\[ \frac{b - t}{b + t} \left( \frac{b^2 - \sigma^2}{b - t} \frac{t}{b + t} \right)^2 \left( \frac{b^2 - \sigma^2}{b - t} \frac{t}{b + t} \right)^2 \left( \frac{b^2 - \sigma^2}{b - t} \frac{t}{b + t} \right)^2 \]

We observe that after we used the “Bernstein trick” no further (strict) inequalities were met with and thus no losses were made. Therefore, improvement on the presented bound \( H \) can only be found in adjusting the premises of the Theorem or in alternatives for the “Bernstein trick”. In this sense our bound is optimal.
3 Discussion of the Result

The merit of the symmetry condition is that it models a real-life situation in auditing. Namely, the case when it is equally likely that an item contains an overstatement error of a certain size as that it contains an understatement error of the same size. This can easily be inferred from the definition of symmetry we use.

The symmetry can obviously be understood to be taken w.r.t. the mean. For \( n \in \mathbb{N}^* \) and \( b > 0 \), let \( Y_1, \ldots, Y_n \) be i.i.d. random variables s.t. \( |Y_i - \mu| \leq b \), \( E(Y_i) = \mu \) and \( V(Y_i) = \sigma^2 \). We can take \( X_i = Y_i - \mu \) and apply the Theorem in order to find:

\[
P\{Y_1 + \ldots + Y_n \geq n(t + \mu)\} \leq \left( \frac{\sigma^2}{2b^2} V(t) \frac{\mu b}{\sigma^2} + \frac{b^2 - \sigma^2}{b^2} V(t) \frac{\mu b}{\sigma^2} + \frac{\sigma^2}{2b^2} V(t) \frac{\mu b}{\sigma^2} \right)^n,
\]

where \( t \in [0, b] \) and \( V(t) \) is the same function as in Theorem 2.1.

Let’s take a closer look at the upper bound. We have constructed it setting some parameters fixed. We obtained for \( n \in \mathbb{N}^* \), \( b > 0 \), \( t \in [0, b] \) and \( \sigma^2 > 0 \) the following upper bound.

\[
H_n(t, \sigma^2, b) = \left( \frac{\sigma^2}{2b^2} V(t) \frac{\mu b}{\sigma^2} + \frac{b^2 - \sigma^2}{b^2} V(t) \frac{\mu b}{\sigma^2} + \frac{\sigma^2}{2b^2} V(t) \frac{\mu b}{\sigma^2} \right)^n.
\]

We will present some nice and important properties of the bound \( H \).

A useful fact is the scaling property w.r.t the variable \( b \). Let \( n \in \mathbb{N}^* \), \( b > 0 \), \( t \in [0, b] \) and \( \sigma^2 > 0 \) we have:

\[
H_n(t, \sigma^2, b) = H_n \left( \frac{t}{b}, \frac{\sigma^2}{b^2}, 1 \right).
\]

Hence, we may just as well take \( b = 1 \). This scaling property is particularly useful when we want to explore behavioural aspects of this bound e.g. by means of computer simulations.

Another observation is that \( H_n \) is an increasing function of the variance \( \sigma^2 \). Let us denote the right hand side of (7) as \( F(h, \sigma^2) \). Keeping \( b \) and \( t \) fixed and suppressing them in the notation, we observe that for fixed \( b \) the function \( \sigma^2 \mapsto F(h, \sigma^2) \) is increasing. Furthermore, we have \( \min_{h \geq 0} F(h, \sigma^2) = F(h_0(\sigma^2), \sigma^2) \).

So, whenever \( \sigma_1 < \sigma_2 \) we obtain the claim by a simple computation.

\[
\min_{h \geq 0} F(h, \sigma_1^2) \leq F(h_0(\sigma_2^2), \sigma_1^2) \leq F(h_0(\sigma_2^2), \sigma_2^2) = \min_{h \geq 0} F(h, \sigma_2^2).
\]

In a similar way it can be shown that \( H_n \) is decreasing in the variable \( t \). These monotonicity properties are useful when we want to apply the results of [8, 9] to obtain the corresponding confidence upper bound.

We obviously have that the bound improves when the sample is enlarged.

\[
H_n(t, \sigma^2, b) \geq H_{n+1}(t, \sigma^2, b).
\]
Some technical observations are also of interest. The function $t \mapsto H_n(t, \sigma^2, b)$ can best be extended to a function $\mathbb{R} \to \mathbb{R}$ by defining it zero beyond $t = b$ and equal to one for $t < 0$. One can easily compute that $H(0, \sigma^2, b) = 1$ and $t \mapsto H_n(t, \sigma^2, b)$ is at $t = b$ jump discontinuous, making a jump down of size $\frac{\sigma^2}{b^2}$.

It is appropriate to mention that W. Hoeffding in his celebrated paper [16] uses, after having applied the Bernstein trick, quite a different strategy to obtain his upper bounds. The comparison of his bound and ours is somewhat tedious. Of course as both bounds are optimal ours tends to be less conservative, because we consider a smaller class of probability laws. The heuristic that a supremum taken over fewer elements is smaller, seems valid.

4 The formal proofs and some explanations

We are going to evaluate the before mentioned least upper bound in this setting. Consider

$$\sup_{\nu \in V} \left\{ \int_0^b \cosh(hx) d\nu(x) \right\}$$

where $V(= V(\sigma^2))$ consists of all probability measures with second moment equal to $\sigma^2$ and supported in $[0, b]$. Here we introduce an auxiliary function

$$\gamma : x \mapsto \frac{\cosh(xh) - 1}{x^2}. \quad (10)$$

A quick glance at the power series of $\cosh$ reveals that this function $\gamma$ is strictly increasing on $[0, b]$, so we have $\gamma(0) \leq \gamma(x) \leq \gamma(b)$. Rewriting the second inequality renders for $x \in [0, b]$

$$\cosh(xh) \leq 1 + \frac{\cosh(bh) - 1}{b^2} x^2. \quad (11)$$

Note that equality only occurs for $x \in \{0, b\}$.

Let $\nu \in V$ and use (11) to obtain the following.

$$\int_0^b \cosh(hx) d\nu(x) \leq 1 + \frac{\cosh(bh) - 1}{b^2} \sigma^2$$

$$= b \int_0^b \cosh(xh) d\left((1 - \frac{\sigma^2}{b^2}) \delta_0 + \frac{\sigma^2}{b^2} \delta_b\right).$$

We conclude that the supremum exists and find its value.

$$\sup_{\nu \in V} \left\{ \int_0^b \cosh(hx) d\nu(x) \right\} = 1 + \frac{\cosh(bh) - 1}{b^2} \sigma^2. \quad (12)$$

Furthermore, a maximizing measure is given by $(1 - \frac{\sigma^2}{b^2}) \delta_0 + \frac{\sigma^2}{b^2} \delta_b$. We demonstrate that there is only one such measure. Let $\nu_e \in V$ be a maximizing element.
Observe that by (11) we have \( \alpha : x \mapsto 1 + \frac{\cosh(bh) - 1}{bh} x^2 - \cosh(xh) \) is nonnegative on \([0, b]\).

From (12) we get \( \int_0^b \alpha(x) d\nu_e = 0 \). Therefore we must have that \( \alpha \) is zero \( \nu_e \)-a.s. Using this together with the remark made following (11), we infer

\[
\text{supp}(\nu_e) \subseteq \{ x \in [0, b] \mid \alpha(x) = 0 \} = \{0, b\}.
\]

Thus \( \nu_e = (1 - p)\delta_0 + p\delta_b \) and by the 2nd moment condition: \( p = \frac{\sigma^2}{b^2} \).

After having exhibited the computational side to the supremum at hand we now turn to the upper bound \( H \). Some straightforward calculations will yield its asserted form.

We introduce the following shorthand notation.

\[
H_t(h) = e^{-ht} \left( \frac{\cosh(bh) - 1}{b^2} - \sigma^2 + 1 \right).
\]

(13)

We keep \( p = \frac{\sigma^2}{b^2} \). The derivative w.r.t. \( h \) is given by the following equation.

\[
H_t'(h) = \frac{1}{2} e^{-h(t+b)} \left( p(b-t)e^{2bh} - 2t(1-p)e^{bh} - p(b+t) \right).
\]

(14)

After a small computation we find that \( H_t'(h) = 0 \) only if

\[
\left( e^{bh} - \frac{1-p}{p} \frac{t}{b-t} \right)^2 = \frac{b+t}{b-t} + \left( \frac{1-p}{p} \frac{t}{b-t} \right)^2.
\]

(15)

Since the exponential function with a real argument only takes positive values, we must have:

\[
e^{bh} = \frac{1-p}{p} \frac{t}{b-t} + \sqrt{\frac{b+t}{b-t} + \left( \frac{1-p}{p} \frac{t}{b-t} \right)^2}.
\]

(16)

The last step is to conclude that \( H_t \) assumes its minimal value at \( h_0 \), cf. equation (8). Hereby we arrive at the conclusion of the Theorem.

The expression \( V(t) \) is reciprocal to the argument of the logarithm in \( h_0 \). We introduced \( V(t) \) merely for notational convenience.

References


[6] V. Bentkus. An inequality for tail probabilities of martingales with bounded differences, 
2002c.


[8] V. Bentkus, G. Pap, and M. van Zuijlen. Confidence bounds for a parameter, October 


statistics*, volume XV of *Pure and Applied Mathematics*. Wiley, New York-London-
Sydney, 1966.


