DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NIJMEGEN The Netherlands

Optimal Hoeffding-like Inequalities under a Symmetry Assumption

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

Report No. 0408 (June 2004)

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF NIJMEGEN
Toernooiveld 1
6525 ED Nijmegen
The Netherlands

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

This research was supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Ministry of Economic Affairs.

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 $X_1, ..., X_n$ 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 $S = X_1 + ... + X_n$.

The symmetry is understood to be w.r.t. the vertical axis. Therfore 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 ν on a σ-algebra $\mathcal{F}$ is said to be symmetric if*

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

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

$$\mathbb{P}\{X_1 + ... + X_n \geq nt\} \leq H(t). \tag{1}$$

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

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

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

$$\mathbb{P}\{X_1 + ... + X_n \geq nt\} \leq \left\{e^{-ht}\mathbb{E}e^{hX_1}\right\}^n. \tag{2}$$

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:

$$\mathbb{E}e^{hX_1} = \int_{-b}^{b} e^{hx} d\mathbb{P}_{X_1}(x). \tag{3}$$

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 μ supported in $[-b, b]$ a law ν 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

$$\mathbb{E}e^{hX_1} = \int_{-b}^{b} \frac{e^{hx} + e^{-hx}}{2} d\mathbb{P}_{X_1}(x) = \int_{0}^{b} \frac{e^{hx} + e^{-hx}}{2} d\nu(x). \tag{4}$$

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

$$\sup_{\nu \in V} \{ \int_{0}^{b} \cosh(hx) d\nu(x) \}. \tag{5}$$

where V is the collection of probability measures on $[0, b]$ with 2nd moment equal to σ^2. It will be shown that the maximizing measure is given by folding together the measure $\frac{\sigma^2}{2b^2}\delta_{-b}+$ $(1 - \frac{\sigma^2}{b^2})\delta_0 + \frac{\sigma^2}{2b^2}\delta_b$. At the same time we will find that the supremum is equal to

$$\frac{\cosh(bh) - 1}{b^2}\sigma^2 + 1. \tag{6}$$

So apparently we shall arrive at the following situation.

$$\mathbb{P}\{X_1 + ... + X_n \geq nt\} \leq \left(e^{-ht} \left\{ \frac{\cosh(bh) - 1}{b^2}\sigma^2 + 1 \right\} \right)^n. \tag{7}$$

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

$$h_0 = h_0(t, \sigma^2, b) = \frac{1}{b} \log \left[\frac{b^2 - \sigma^2}{\sigma^2} \frac{t}{b - t} + \sqrt{\frac{b + t}{b - t} + \left(\frac{b^2 - \sigma^2}{\sigma^2} \frac{t}{b - t} \right)^2} \right]. \tag{8}$$

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 $X_1, .., X_n$ 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 $i \in \{1, ..., n\}$. Then for $t \in [0, b]$ we have*

$$\mathbb{P}\{X_1 + ... + X_n \geq nt\} \leq \left(\frac{\sigma^2}{2b^2} V(t)^{\frac{t-b}{b}} + \frac{b^2 - \sigma^2}{b^2} V(t)^{\frac{t}{b}} + \frac{\sigma^2}{2b^2} V(t)^{\frac{t+b}{b}} \right)^n,$$

where $V(t)$ is short for

$$\frac{b - t}{b + t} \left(\sqrt{\frac{b + t}{b - t} + \left(\frac{b^2 - \sigma^2}{\sigma^2} \frac{t}{b - t} \right)^2} - \frac{b^2 - \sigma^2}{\sigma^2} \frac{t}{b - t} \right).$$

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 $Y_1, ..., Y_n$ be i.i.d. random variables s.t. $|Y_i - \mu| \leq b$, $\mathbb{E}(Y_i) = \mu$ and $\mathbb{V}(Y_i) = \sigma^2$. We can take $X_i = Y_i - \mu$ and apply the Theorem in order to find:

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

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.

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

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:

$$H_n(t, \sigma^2, b) = H_n(\frac{t}{b}, \frac{\sigma^2}{b^2}, 1).$$

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 σ^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 h the function $\sigma^2 \mapsto F(h, \sigma^2)$ is increasing. Furthermore, we have $\min_{h>0} F(h, \sigma^2) = F(h_0(\sigma^2), \sigma^2)$.
So, whenever $\sigma_1 < \sigma_2$ we obtain the claim by a simple computation.

$$\begin{aligned}
\min_{h>0} F(h, \sigma_1^2) &\leq F(h_0(\sigma_2^2), \sigma_1^2) \\
&\leq F(h_0(\sigma_2^2), \sigma_2^2) \\
&= \min_{h>0} F(h, \sigma_2^2).
\end{aligned}$$

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

$$\sup_{\nu \in V}\{\int_0^b \cosh(hx)d\nu(x)\} \tag{9}$$

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

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

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

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

Note that equality only occurs for $x \in \{0, b\}$.
Let $\nu \in V$ and use (11) to obtain the following.

$$\begin{aligned} \int_0^b \cosh(hx)d\nu(x) &\leq 1 + \tfrac{\cosh(bh)-1}{b^2}\sigma^2 \\ &= \int_0^b \cosh(xh)d((1 - \tfrac{\sigma^2}{b^2})\delta_0 + \tfrac{\sigma^2}{b^2}\delta_b). \end{aligned}$$

We conclude that the supremum exists and find its value.

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

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 $\alpha : x \mapsto 1 + \frac{\cosh(bh)-1}{b^2}x^2 - \cosh(xh)$ is nonnegative on $[0, b]$. From (12) we get $\int_0^b \alpha(x)d\nu_e = 0$. Therefore we must have that α is zero ν_e-a.s. Using this together with the remark made following (11), we infer

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

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.

$$H_t(h) = e^{-ht}\left(\frac{\cosh(bh)-1}{b^2}\sigma^2 + 1\right). \tag{13}$$

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

$$H_t^{'}(h) = \frac{1}{2}e^{-h(t+b)}\left(p(b-t)e^{2bh} - 2t(1-p)e^{bh} - p(b+t)\right). \tag{14}$$

After a small computation we find that $H_t{}^{'}(h) = 0$ only if

$$\left(e^{bh} - \frac{1-p}{p}\frac{t}{b-t}\right)^2 = \frac{b+t}{b-t} + \left(\frac{1-p}{p}\frac{t}{b-t}\right)^2. \tag{15}$$

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

$$e^{bh} = \frac{1-p}{p}\frac{t}{b-t} + \sqrt{\frac{b+t}{b-t} + \left(\frac{1-p}{p}\frac{t}{b-t}\right)^2}. \tag{16}$$

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

[1] N. Alon and V. D. Milman. Concentration of measure phenomena in the discrete case and the Laplace operator of a graph of a distribution function to the normal law. *Seminar on functional analysis*, Publ. Math. Univ. Paris VII(20):55–68, 1984.

[2] V. Bentkus. On the asymptotical behavior of the constant in the Berry–Esseen inequality. *J. Theoret. Probab.*, 7(2):211–224, 1994.

[3] V. Bentkus. An inequality for large deviation probabilities of sums of bounded i.i.d.r.v. *Lithuanian Math. J.*, 41(2):144–153, 2001.

[4] V. Bentkus. On measure concentration for separately Lipschitz functions in product spaces. *To appear in Israel J. Math.*, pages 321–332, 2002a.

[5] V. Bentkus. An inequality for tail probabilities of martingales with differences bounded from one side. *To appear in J. Theor. Probab.*, 2002b.

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

[7] V. Bentkus and K. Kirsha. Estimates for the closeness of a disribution function to the normal law. *Lithuanian Math. J.*, 29(4):321–332, 1989.

[8] V. Bentkus, G. Pap, and M. van Zuijlen. Confidence bounds for a parameter, October 2001. *Report series University of Nijmegen, The Netherlands*, 2001.

[9] V. Bentkus and M. van Zuijlen. Conservative confidence bounds for the mean. *To appear in Lithuanian Math. J.*, 2003.

[10] J. Bretagnolle. Statistique de Kolmogorov-Smirnov pour un enchantillon nonequireparti. *Colloq. Internat. CNRS*, 307:39–44, 1980.

[11] M.L. Eaton. A note on symmetric Bernoulli random variables. *Ann. Math. Stat.*, 41(4):1223–1226, 1970.

[12] M.L. Eaton. A probability inequality for linear combinations of bounded random variables. *Ann. Stat.*, 2(3):609–613, 1974.

[13] L. J. Gleser. On the distribution of the number of successes in independent trials. *Ann. Probab.*, 3(1):182–188, 1975.

[14] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *Amer. J. Math.*, 105(4):843–854, 1983.

[15] W. Hoeffding. On the distribution of the number of successes in independent trials. *Ann. Math. Statist.*, 27:713–721, 1956.

[16] W. Hoeffding. Probability inequalities for sums of bounded random variables. *JASA*, 58:13–30, 1963.

[17] I.A. Ibragimov. On the composition of unimodal distributions. *Theor. Probab. Appl.*, 1:255–260, 1956.

[18] S. Karlin and W. J. Studden. *Tchebycheff systems: With applications in analysis and statistics*, volume XV of *Pure and Applied Mathematics*. Wiley, New York-London-Sydney, 1966.

[19] J. Keilson and H. Gerber. Some results for discrete unimodality. *JASA*, 66(334):386–389, 1971.

[20] M. Ledoux. *Concentration of measure and logarithmic Sobolev inequalities*, volume 1709 of *Séminaire de Probabilités, XXXIII.* Springer, Berlin, 1999.

[21] A. W. Marshall and I. Olkin. *Inequalities: theory of majorization and its applications.* Mathematics in Science and Engineering, 143. Academic Press, New York-London, 1979.

[22] C. McDiarmid. On the method of bounded differences. *London Math. Soc. Lecture Note Ser*, 141:148–188, 1989.

[23] V. D. Milman. Random subspaces of proportional dimension of finite-dimensionalnormed spaces: approach through the isoperimetric inequality. *Lecture Notes in Math.*, 1166:106–115, 1985.

[24] V. D. Milman. The heritage of p. Lévy in geometrical functional analysis. *Ast érisque*, 157–158:273–301, 1988.

[25] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces.* Number 1200 in Lecture Notes in Mathematics. Springer, 1986.

[26] A. Ostrowski. Sur quelques applications des fonctions convexes et concave au sens de I. Schur. *J. Math. Pures Appl.*, 31:253–292, 1952.

[27] V. Paulauskas. Some comments on deviation probabilities for infinitely divisible random vectors. *To appear in Lithuanian Math. J.*, pages 253–292, 2002.

[28] V. V. Petrov. *Sums of independent random variables.* Springer-Verlag, New York-Heidelberg, 1975.

[29] I. Pinelis. Extremal probabilistic problems and hotelling's t^2 test under a symmetry assumption. *Ann. Stat.*, 22(4):357–368, 1994.

[30] I. Pinelis. Optimal tail comparison based on comparison of moments. *High dimensional probability*, 43:297–314, 1998.

[31] I. Pinelis. *Fractional sums and integrals of r-concave tails and applications to comparison probability inequalities.* Number 234 in Contemp. Math. Advances in stochastic inequalities (Atlanta, GA, 1997). Amer. Math. Soc., Providence, RI, 1999.

[32] I. Schur. über eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie. *Sitzber Berl. Math. Ges.*, 22:9–20, 1923.

[33] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics.* John Wiley & Sons, Inc., New York, 1986.

[34] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, 81:73–205, 1995.

[35] M. Talagrand. The missing factor in Hoeffding's inequalities. *Ann. Inst. H. Poincar é Probab. Statist.*, 31(4):689–702, 1995a.

Vidmantas Bentkus
Vilnius Institute of Mathematics
Akademijos 4
232600 Vilnius
Lithuania
e-mail: bentkus@ktl.mii.lt

Gudron Geuze
University of Nijmegen
Toernooiveld 1
6525 ED Nijmegen
The Netherlands
e-mail: gudrong@math.kun.nl

Martien van Zuijlen
University of Nijmegen
Toernooiveld 1
6525 ED Nijmegen
The Netherlands
e-mail: zuijlen@math.kun.nl