A revised version of this paper has been published in
***American Mathematical Monthly*, 122(8):786-789, 2015.**

An exponential inequality for symmetric random variables

Raphaël Cerf
ENS Paris

and

Matthias Gorny
Université Paris Sud and ENS Paris

A central question in the study of sums of independent identically distributed random variables is to control the probabilities that they deviate from their typical values. These probabilities are usually hard to compute, instead the standard strategy is to rely on inequalities. The most fundamental inequality of this sort is the celebrated Chebyshev inequality (see for instance [4]). Numerous refinements and extensions of this inequality have been worked out. It turns out that, for independent variables admitting an exponential moment, the speed of deviation from the typical behavior is of exponential order in the number of variables. One seeks therefore exponential inequalities. Two famous inequalities are the Chebyshev exponential inequality and Hoeffding's inequality [5]. These inequalities help to design statistical tests and to compute intervals of confidence associated to a random estimator.

We present here a very simple exponential inequality. This inequality holds for any symmetric distribution, and does not require any integrability condition.

Theorem. *Let $n \geq 1$ and let $X_1, \ldots, X_n$ be n independent identically distributed symmetric real-valued random variables. For any $x, y > 0$, we have*

$$\mathbb{P}\big(X_1 + \cdots + X_n \geq x,\, X_1^2 + \cdots + X_n^2 \leq y\big) < \exp\left(-\frac{x^2}{2y}\right).$$

If we apply this inequality with nx, ny instead of x, y, we obtain an inequality controlling the first two empirical moments of the sequence $X_1, \cdots, X_n$, with an upper bound of exponential order in n:

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n}{n} \geq x,\, \frac{X_1^2 + \cdots + X_n^2}{n} \leq y\right) < \exp\left(-\frac{nx^2}{2y}\right).$$

With the classical theory of large deviations [3], we usually obtain exponential inequalities of this type, but unfortunately they are valid for n large and it is a difficult task to quantify how large n has to be. During the last decades, probabilists and statisticians have been trying to find non asymptotic inequalities, which are valid for any $n \geq 1$ and with explicit constants [1]. Most of these inequalities are based on the phenomenon of concentration of measure and they require a strong control on the tail of the distribution of X_1, typically the existence of an exponential moment. The inequality we present here is a deviation inequality which is valid for all $n \geq 1$. We suppose that the distribution is symmetric, and no further integrability condition is required. Initially, this inequality was obtained through classical results of large deviations within Cramér's theory and an inequality on the rate function derived to study a Curie–Weiss model of self–organized criticality [2]. Since the statement of the exponential inequality is very simple, we looked for an elementary proof, and it is this proof we present here.

Let us prove the theorem. We suppose that $\mathbb{P}(X_1 = 0) < 1$, otherwise the inequality of the theorem is immediate. Let $n \geq 1$ and $x, y > 0$. We set

$$S_n = X_1 + \cdots + X_n \,, \qquad T_n = X_1^2 + \cdots + X_n^2 \,.$$

Let $s, t > 0$. We have

$$\begin{aligned}\mathbb{P}\big(S_n \geq x,\, T_n \leq y\big) &= \mathbb{P}\left(sS_n \geq sx,\, -tT_n \geq -ty\right)\\ &\leq \mathbb{P}\big(sS_n - tT_n \geq sx - ty\big)\\ &\leq \mathbb{P}\Big(\exp\left(sS_n - tT_n\right) \geq \exp\left(sx - ty\right)\Big)\,.\end{aligned}$$

We recall one of the most classical stochastic inequality.

Markov's inequality. If X is a non–negative random variable, then

$$\forall \lambda > 0 \qquad \mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}\,.$$

Using Markov's inequality and the fact that $X_1, \dots, X_n$ are i.i.d., we get

$$\begin{aligned}\mathbb{P}\big(S_n \geq x,\, T_n \leq y\big) &\leq \exp\big(-sx + ty\big)\,\mathbb{E}\Bigg(\prod_{i=1}^n \exp\left(sX_i - tX_i^2\right)\Bigg)\\ &= \exp\big(-sx + ty\big)\Bigg(\mathbb{E}\Big(\exp\left(sX_1 - tX_1^2\right)\Big)\Bigg)^n\,.\end{aligned}$$

The distribution of X_1 is symmetric thus

$$\begin{aligned}\mathbb{E}\Big(\exp\left(sX_1 - tX_1^2\right)\Big) &= \mathbb{E}\Big(\exp\left(-sX_1 - tX_1^2\right)\Big)\\ &= \frac{1}{2}\Big(\mathbb{E}\Big(\exp\left(sX_1 - tX_1^2\right)\Big) + \mathbb{E}\Big(\exp\left(-sX_1 - tX_1^2\right)\Big)\Big)\\ &= \mathbb{E}\Big(\cosh(sX_1)\,\exp\left(-tX_1^2\right)\Big)\,.\end{aligned}$$

We choose now $t = s^2/2$. We have the inequality

$$\forall u \in \mathbb{R}\backslash\{0\} \qquad \cosh(u)\,\exp\left(-u^2/2\right) < 1\,.$$

Since $\mathbb{P}(X_1 = 0) < 1$, the above inequality implies that

$$\mathbb{E}\Big(\cosh(sX_1)\,\exp\left(-s^2X_1^2/2\right)\Big) \,<\, 1\,,$$

whence also

$$\mathbb{P}\big(S_n \geq x,\, T_n \leq y\big) \,<\, \exp\big(-sx + s^2y/2\big)\,.$$

We finally choose $s = x/y$ and we obtain the desired inequality.

Link with Hoeffding's inequality. Let us suppose that the random variables $X_1, \dots, X_n$ are bounded by a deterministic constant $c > 0$:

$$\mathbb{P}\big(|X_i| \leq c\big) \;=\; 1\,, \qquad 1 \leq i \leq n\,.$$

Taking $y = c^2$ in our inequality, we obtain

$$\forall x > 0 \qquad \mathbb{P}\left(\frac{X_1 + \cdots + X_n}{n} \geq x\right) < \exp\left(-\frac{nx^2}{2c^2}\right)\,.$$

This way we recover the upper bound given by Hoeffding's inequality (specialized to the case of bounded symmetric i.i.d. random variables, and with a strict inequality). Yet the proof of our inequality is simpler than the proof of Hoeffding's inequality. In the case of Hoeffding's inequality, the key exponential estimate relies on a second order expansion, Taylor's formula and the boundedness hypothesis. In our case, the key exponential estimate relies on the inequality $\cosh(u) \leq \exp(-u^2/2)$ and the symmetry hypothesis. Conversely, as pointed out by the Referee, Hoeffding's inequality can be used to recover our inequality via a conditioning argument, as follows. Let $\varepsilon_1, \dots, \varepsilon_n$ be n independent Bernoulli random variables which are independent of $X_1, \dots, X_n$ and such that

$$\mathbb{P}\big(\varepsilon_i = -1\big) \;=\; \mathbb{P}\big(\varepsilon_i = 1\big) \;=\; \frac{1}{2}\,, \qquad 1 \leq i \leq n\,.$$

We set $T_n = X_1^2 + \cdots + X_n^2$. Since $X_1, \dots, X_n$ are symmetric, we have

$$\mathbb{P}\big(X_1 + \cdots + X_n \geq x,\, T_n \leq y\big) \;=\; \mathbb{P}\big(\varepsilon_1 X_1 + \cdots + \varepsilon_n X_n \geq x,\, T_n \leq y\big)\,.$$

We condition next with respect to $X_1, \dots, X_n$. Denoting by $\mathbb{1}_A$ the indicator function of an event A and by $\mathbb{E}$ the expectation, the above expression can be rewritten as

$$\mathbb{E}\big(\mathbb{1}_{\varepsilon_1 X_1 + \cdots + \varepsilon_n X_n \geq x}\, \mathbb{1}_{T_n \leq y}\big) \;=\; \mathbb{E}\Big(\mathbb{P}\big(\varepsilon_1 X_1 + \cdots + \varepsilon_n X_n \geq x \,\big|\, X_1, \dots, X_n\big) \mathbb{1}_{T_n \leq y}\Big)\,.$$

We apply Hoeffding's inequality to the random variables $\varepsilon_1, \dots, \varepsilon_n$ and we get

$$\mathbb{P}\big(\varepsilon_1 X_1 + \cdots + \varepsilon_n X_n \geq x \,\big|\, X_1, \dots, X_n\big) \;\leq\; \exp\left(-\frac{x^2}{2(X_1^2 + \cdots + X_n^2)}\right)\,.$$

Plugging this inequality in the previous expectation and using the condition $T_n \leq y$, we obtain the exponential inequality stated in our theorem, except that the inequality is large and not strict.

Application to the non symmetric case. Let $n \geq 1$ and let $X_1, \dots, X_n$, $X_1', \dots, X_n'$ be $2n$ independent identically distributed real-valued random variables. For any $x, y > 0$, we have

$$\mathbb{P}\left(\begin{array}{c} \dfrac{X_1 + \cdots + X_n}{n} - \dfrac{X_1' + \cdots + X_n'}{n} \geq x \\[2ex] \dfrac{{X_1}^2 + \cdots + {X_n}^2}{n} + \dfrac{{X_1'}^2 + \cdots + {X_n'}^2}{n} \leq y \end{array}\right) < \exp\left(-\frac{nx^2}{4y}\right)\,.$$

The nice feature of this inequality is that it does not require any hypothesis at all on the common distribution of the random variables. This inequality is a simple consequence of the theorem, applied to the random variables $Y_i = X_i - X_i'$, $1 \leq i \leq n$, which are again symmetric.

References

[1] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence.* Oxford University Press, 2013.

[2] Raphaël Cerf and Matthias Gorny. A Curie-Weiss model of Self-Organized Criticality. *The Annals of Probability*, to appear, 2013.

[3] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability.* Springer-Verlag, 2010.

[4] William Feller. *An introduction to probability theory and its applications. Vol. I and II.* John Wiley & Sons Inc., New York, 1968 and 1971.

[5] Hoeffding's inequality, Wikipedia.