Notes on means, medians and Gaussian tails

J.-R Chazottes

Notes on means, medians and Gaussian tails

J.-R. Chazottes *1

1CPHT, CNRS, Institut Polytechnique de Paris, France

Dated: April 18, 2022

Abstract

In the course of thinking about Gaussian concentration bounds for $X - \mathbb{E}(X)$, we arrived at the basic question of replacing $\mathbb{E}(X)$ by a median of X. Not surprinsingly, having a Gaussian concentration bound for X about its mean is equivalent to having it for any of its median. The question is: Can we find simple relations between the involved constants? This is indeed what is asked in exercise 2.14 p. 53 in [4], not exactly for Gaussian concentration bounds but for Gaussian tail bounds, which is equivalent (see below). We also solve an exercise from [1]. (This note is not intended to be published and we acknowledge F. Redig for email exchanges on that matters.)

1 Median versus expectation

Here we solve exercise 2.14 p. 53 in [4].

Given a scalar random variable X, suppose that there are positive constants c_1, c_2 such that

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq c_1 \,\mathrm{e}^{-c_2 t^2},\ t \geq 0. \tag{1}$$

Observe that we must have $c_1 \geq 1$ since we can let $t \downarrow 0$: the rhs tends to c_1 while the lhs is a priori ≤ 1. One says that X has "Gaussian tails around its mean".

A basic consequence of (1) is that

$$\mathrm{Var}(X) \leq \frac{c_1}{c_2}\cdot \tag{2}$$

Indeed we have

$$\mathrm{Var}(X) = \int_0^\infty 2t\,\mathbb{P}(|X - \mathbb{E}(X)| \geq t)\,\mathrm{d}t \leq c_1 \int_0^\infty 2t\,\mathrm{e}^{-c_2 t^2}\,\mathrm{d}t = \frac{c_1}{c_2}\cdot$$

*Email: `jeanrene@cpht.polytechnique.fr`

Before proceeding, let us recall some basic facts about medians. A median m_X is any number such that $\mathbb{P}(X \geq m_X) \geq 1/2$ and $\mathbb{P}(X \leq m_X) \geq 1/2$. In general it is ***not unique***. The basic example illustrating this is when X takes the values 0 or 1 with probability $1/2$. Then any $0 < m < 1$ is a median.
If X has a cumulative distribution function which is continuous and whose support is an interval, then m_X is uniquely defined.
The Cauchy distribution is an example of a random variable without a first moment but with a unique median which is 0. (Hence medians always exist, contrarily to expected values.)
Before going on, we would like to recall the nice general bound

$$|\mathbb{E}(X) - m_X| \leq \sqrt{\mathrm{Var}(X)} \tag{3}$$

which holds for any random variable such that $\mathbb{E}(X^2) < \infty$. The proof is based on the fact that a median minimizes the function $m \mapsto \mathbb{E}(|X - m|)$ on $\mathbb{R}$ (note that in fact this charaterization works even for a non-integrable random variable if one writes the minimization problem in an appropriate way). Given this fact, we then have

$$\begin{aligned} |m_X - \mathbb{E}(X)| = |\mathbb{E}(m_X - X)| \leq \mathbb{E}(|X - m_X|) \leq \mathbb{E}(|X - \mathbb{E}(X)|) \\ \leq \sqrt{\mathbb{E}[(X - \mathbb{E}(X))^2]} \leq \sqrt{\mathrm{Var}(X)}. \end{aligned}$$

(The first inequality is by Jensen's inequality for $x \mapsto |x|$, the second one comes from the characterization of a median given above, and the third is by Cauchy-Schwarz inequality.) So we proved (3).

1.1 Gaussian tails around the mean implies Gaussian tails around the median

Theorem 1.1. *Suppose that* (1) *holds. Then for any median* m_X *one has*

$$\mathbb{P}(|X - m_X| \geq t) \leq c_3\, \mathrm{e}^{-c_4 t^2},\ t \geq 0 \tag{4}$$

with $c_3 = 2c_1$ *and* $c_4 = c_2/4$.

Before giving the proof, observe that we must have $c_3 \geq 1$, since we can let $t \downarrow 0$ in (4). Note also that Wainwright asks to prove this result with $c_3 = 4c_1$ and $c_4 = c_2/8$, so we obtain slightly better constants.
Proof. Put $\delta := |\mathbb{E}(X) - m_X|$. (Of course, we assume that $\delta > 0$, otherwise there is nothing to prove.)
The main observation is that, by the very definition of a median, and using the assumption, we have

$$\frac{1}{2} \leq \mathbb{P}(|X - \mathbb{E}(X)| \geq \delta\,) \leq c_1\, \mathrm{e}^{-c_2 \delta^2}\,.$$

(Indeed, assume that $m_X < \mathbb{E}(X)$. Then the event $\{|X - \mathbb{E}(X)| \geq \delta\}$ contains the event $\{X \leq m_X\}$ which has probability at least $1/2$. The case $m_X > \mathbb{E}(X)$ is similar.) Hence

$$2c_1 \,\mathrm{e}^{-c_2\delta^2} \geq 1.$$

Now, if we take any $t < 2\delta$ we have

$$1 \leq 2c_1 \,\mathrm{e}^{-c_2\delta^2} = 2c_1 \,\mathrm{e}^{-c_2(2\delta)^2/4} \leq 2c_1 \,\mathrm{e}^{-c_2t^2/4}\,.$$

Hence in this regime, (4) is (trivially) true with $c_3 = 2c_1$ and $c_4 = c_2/4$.[1] Now we consider $t \geq 2\delta$. Since

$$|X - \mathbb{E}(X)| \geq |X - m_X| - \delta \tag{5}$$

by the reverse triangle inequality ($|a-b| \geq ||a| - |b|| \geq |a| - |b|$, $a, b \in \mathbb{R}$), we have

$$\begin{aligned}
\mathbb{P}\left(|X - m_X| \geq t\right) &= \mathbb{P}\left(|X - m_X| \geq \frac{t}{2} + \frac{t}{2}\right)\\
&\leq \mathbb{P}\left(|X - m_X| \geq \frac{t}{2} + \delta\right) \quad (\text{since}\ \ \delta \leq t/2)\\
&\leq \mathbb{P}\left(|X - \mathbb{E}(X)| \geq \frac{t}{2}\right) \quad (\text{by}\ \ (5))\\
&\leq c_1 \,\mathrm{e}^{-c_2t^2/4} \quad (\text{by}\ \ (1)).
\end{aligned}$$

Therefore we proved (4) with $c_3 = 2c_1$ and $c_4 = c_2/4$. $\square$

1.2 Gaussian tails around the median implies Gaussian tails around the mean

Theorem 1.2. *Assume that* (4) *holds. Then* (1) *holds with* $c_1 = 2c_3$ *and* $c_2 = c_4/16$.

Before giving the proof, we mention that Wainwright asks for $c_1 = 2c_3$ and $c_2 = c_4/4$.

Proof. By Markov's inequality and the elementary inequality $\mathrm{e}^{\lambda|x|} \leq \mathrm{e}^{\lambda x} + \mathrm{e}^{-\lambda x}$, $x \in \mathbb{R}$, we have for all $t \geq 0$

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq t\big) \leq \inf_{\lambda > 0} \frac{\mathbb{E}\big(\,\mathrm{e}^{\lambda(X - \mathbb{E}(X))}\,\big) + \mathbb{E}\big(\,\mathrm{e}^{-\lambda(X - \mathbb{E}(X))}\,\big)}{\mathrm{e}^{\lambda t}}. \tag{6}$$

The trick is to introduce an independent copy Y of X, and thanks to Jensen's inequality to write

$$\mathbb{E}\big(\,\mathrm{e}^{\lambda(X - \mathbb{E}(X))}\,\big) = \mathbb{E}_X\big(\,\mathrm{e}^{\lambda(X - \mathbb{E}_Y(Y))}\,\big) \leq \mathbb{E}_{X,Y}\big(\,\mathrm{e}^{\lambda(X - Y)}\,\big)$$

[1] Of course, we can take $t < \tau\delta$ for some $\tau \geq 1$ to be fixed, but this doesn't seem to lead to better constants.

where the notation $\mathbb{E}_X$ is to precise with respect to which probability distribution we integrate, and $\mathbb{E}_{X,Y}$ denotes integration with respect to the product of the probability distributions of X and Y. When no confusion can arise, we simply write $\mathbb{E}$ for $\mathbb{E}_{X,Y}$, etc. Now we use that $X-Y$ has all its odd moments which are equal to 0, hence

$$\mathbb{E}\big(\,\mathrm{e}^{\lambda(X-Y)}\,\big)=1+\sum_{k=1}^{\infty}\frac{\lambda^{2k}}{(2k)!}\,\mathbb{E}\big[(X-Y)^{2k}\big]. \tag{7}$$

(We assume momentarily that integrability holds; it will be an easy consequence of what follows.) We have

$$\begin{aligned}\mathbb{E}\big[(X-Y)^n\big]&=\int_0^\infty \mathbb{P}\big(|X-Y|^n\geq u\big)\,\mathrm{d}u\\&=n\int_0^\infty v^{n-1}\mathbb{P}\big(|X-Y|\geq v\big)\,\mathrm{d}v\\&\leq 2\,n\int_0^\infty v^{n-1}\mathbb{P}\Big(|X-m_X|\geq\frac{v}{2}\Big)\,\mathrm{d}v.\end{aligned}$$

Let us explain how to obtain the inequality. We use two basic facts:

$$|X-Y|=|X-m_X-(Y-m_X)|\leq|X-m_X|+|Y-m_X|$$

and

$$\mathbb{P}\big(|X-Y|<t\big)\geq\mathbb{P}\Big(|X-m_X|<\frac{t}{2}\Big)\,\mathbb{P}\Big(|Y-m_X|<\frac{t}{2}\Big)$$

where we use independence. Passing to the complement, we thus obtain

$$\begin{aligned}\mathbb{P}\big(|X-Y|\geq t\big)&\leq 2\,\mathbb{P}\Big(|X-m_X|\geq\frac{t}{2}\Big)-\Big(\mathbb{P}\Big(|X-m_X|\geq\frac{t}{2}\Big)\Big)^2\\&\leq 2\,\mathbb{P}\Big(|X-m_X|\geq\frac{t}{2}\Big).\end{aligned}$$

We now use (4) to obtain

$$\begin{aligned}\mathbb{E}\big[(X-Y)^n\big]&\leq 2\,n\,c_3\int_0^\infty v^{n-1}\,\mathrm{e}^{-\frac{c_4v^2}{4}}\,\frac{\sqrt{2\pi\times\frac{2}{c_4}}}{\sqrt{2\pi\times\frac{2}{c_4}}}\,\mathrm{d}v\\&=2\,n\,c_3\sqrt{\frac{\pi}{c_4}}\int_{-\infty}^{+\infty}|v|^{n-1}\,\frac{\mathrm{e}^{-\frac{c_4v^2}{4}}}{\sqrt{2\pi\times\frac{2}{c_4}}}\,\mathrm{d}v\\&=c_3\,n\left(\frac{2}{\sqrt{c_4}}\right)^n\Gamma\left(\frac{n}{2}\right)\end{aligned}$$

where we used that if Z is Gaussian random variable with mean 0 and variance $2/c_4$ then

$$\mathbb{E}\big(|Z|^{n-1}\big)=\left(\frac{2}{c_4}\right)^{\frac{n-1}{2}}\frac{2^{\frac{n-1}{2}}\,\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}}$$

where Γ is the Gamma function. Hence we get from (7)

$$\begin{aligned}
\mathbb{E}\big(\mathrm{e}^{\lambda(X-Y)}\big) &\leq 1 + c_3 \sum_{k=1}^{\infty} \left(\frac{2\lambda}{\sqrt{c_4}}\right)^{2k} \frac{\Gamma(k)}{\Gamma(2k)} \\
&\leq 1 + c_3 \sum_{k=1}^{\infty} \left(\frac{4\lambda^2}{c_4}\right)^{k} \frac{1}{k!} \\
&\leq c_3\,\mathrm{e}^{\frac{4\lambda^2}{c_4}}
\end{aligned}$$

where in the first inequality we used the identity $\Gamma(2k) = (2k-1)!$, in the second one once again this identity, and the bound $1/((2k-1)\cdots k) \leq 1/k!$, and in the last one we used that $c_3 \geq 1$. The same argument works for $\mathbb{E}\big[\,\mathrm{e}^{-\lambda(X-\mathbb{E}(X))}\,\big]$, therefore we obtain from (6)

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq t\big) \leq 2\,c_3 \inf_{\lambda>0} \mathrm{e}^{\frac{4\lambda^2}{c_4} - \lambda t} = c_1\,\mathrm{e}^{-c_2 t^2}$$

where $c_1 = 2c_3$ and $c_2 = c_4/16$. (Wainwright asks to show that $c_2 = c_4/4$.) $\square$

Remark 1.1. *We didn't use the fact that we started from concentration around a median. If one assumes that (4) holds for some constant a instead of m_X, the proof is exactly the same. On the contrary, when we proved how to go from Gaussian tails around the mean to Gaussian tails around a median, we did use a property of medians.*

Remark 1.2. *Using (2), which results from (1), and (3) we get*

$$|\mathbb{E}(X) - m_X| \leq \sqrt{\frac{c_1}{c_2}}\,.$$

But we can use directly (1) to deduce that $|\mathbb{E}(X) - m_X| \leq t_0$, where t_0 is such that $c_1\,\mathrm{e}^{-c_2 t_0^2} = 1/2$, which is $t_0 = \sqrt{\log(2c_1)/c_2}$. Not surprisingly, $t_0 < \sqrt{\frac{c_1}{c_2}}$.

Remark 1.3. *In Proposition 1.8 p. 10 in [2], there is a result saying that if you have Gaussian tails around some real number a_X (which of course can be $\mathbb{E}(X)$) then you have Gaussian tails around any median of X, and vice-versa. It doesn't seem that the method used therein gives precise relations between the involved constants (but this is not the goal). Besides that, he doesn't only deal with Gaussian tails.*

2 Exercise 2.2 p. 46 in [1]

The assumption is (4). The question is: Prove that

$$|\mathbb{E}(X) - m_X| \leq \min\left(\frac{\sqrt{\pi}c_3}{2\sqrt{c_4}}, \sqrt{\frac{c_3}{c_4}}\right). \tag{8}$$

(Remember that $c_3 \geq 1$.) As above when we discussed the distance between the mean and the median for a square-integrable random variable (without involving concentration), we start with

$$|m_X - \mathbb{E}(X)| = |\mathbb{E}(m_X - X)| \leq \mathbb{E}(|X - m_X|).$$

Then using (4) we have

$$\begin{aligned}\mathbb{E}(|X - m_X|) &= \int_0^\infty \mathbb{P}(|X - m_X| \geq t)\,\mathrm{d}t \leq c_3 \int_0^\infty \mathrm{e}^{-c_4 t^2}\,\mathrm{d}t \\ &= \frac{c_3}{\sqrt{2c_4}} \int_0^\infty \mathrm{e}^{-u^2/2}\,\mathrm{d}u = \frac{c_3}{\sqrt{2c_4}} \frac{\sqrt{2\pi}}{2}.\end{aligned}$$

For the other bound, we want to use (3), so we have to prove that $\mathrm{Var}(X) \leq c_3/c_4$. We have

$$\begin{aligned}\mathbb{E}\big(|X - m_X|^2\big) &= \int_0^\infty \mathbb{P}\big(|X - m_X|^2 \geq t\big)\,\mathrm{d}t = \int_0^\infty \mathbb{P}\big(|X - m_X| \geq \sqrt{t}\,\big)\,\mathrm{d}t \\ &\leq \int_0^\infty c_3\,\mathrm{e}^{-c_4 t}\,\mathrm{d}t = \frac{c_3}{c_4}.\end{aligned}$$

This proves that X is square integrable, and we conclude by observing that $\mathrm{Var}(X) = \mathrm{Var}(X - m_X) \leq \mathbb{E}\big(|X - m_X|^2\big)$. Hence $|m_X - \mathbb{E}(X)| \leq \sqrt{c_3/c_4}$. Gathering the two bounds we thus proved (8).

3 Gaussian concentration bound and Gaussian tail bound

Usually one proves (1) by proving the following bound that we call the Gaussian concentration bound:

$$\mathbb{E}\big(\,\mathrm{e}^{\lambda(X - \mathbb{E}(X))}\,\big) \leq \mathrm{e}^{c\lambda^2},\ \forall \lambda \in \mathbb{R} \tag{9}$$

where $c > 0$ is independent of λ. Let us derive (1) from (9) for completeness. For any $\lambda > 0$ we have by Markov's inequality

$$\begin{aligned}\mathbb{P}(X - \mathbb{E}(X) \geq t) = \mathbb{P}(\exp(\lambda(X - \mathbb{E}(X)) \geq \exp(\lambda t)) &\leq \mathrm{e}^{-\lambda t}\,\mathbb{E}\big(\,\mathrm{e}^{\lambda(X - \mathbb{E}(X))}\,\big) \\ &\leq \mathrm{e}^{-\lambda t + c\lambda^2} \quad \text{(by (1))}.\end{aligned}$$

Now we can minimize the bound over $\lambda > 0$ and find out

$$\mathbb{P}(X - \mathbb{E}(X) \geq t) \leq \mathrm{e}^{-\frac{t^2}{4c}}.$$

We can repeat these estimates with $-X$ in place of X to find

$$\mathbb{P}(-X - \mathbb{E}(-X) \geq t) \leq \mathrm{e}^{-\frac{t^2}{4c}}$$

hence by a union bound we finally get

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq 2\,\mathrm{e}^{-\frac{t^2}{4c}},\ \forall t \geq 0.$$

Therefore we obtain (1) with $c_1 = 2$ and $c_2 = 1/(4c)$.

Going from (1) to (9) can be done as in [1, Theorem 2.1 p. 25].

References

[1] S. Boucheron, G. Lugosi, P. Massart. *Concentration inequalities.* A nonasymptotic theory of independence. Oxford University Press, 2013.

[2] M. Ledoux. *The concentration of measure phenomenon.* Mathematical Surandom variableeys and Monographs, 89. American Mathematical Society, Providence, RI, 2001.

[3] R. Vershynin. *High-dimensional probability.* An introduction with applications in data science. Cambridge Series in Statistical and Probabilistic Mathematics, 47. Cambridge University Press, Cambridge, 2018.

[4] M. Wainwright. *High-dimensional statistics.* A non-asymptotic viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics, 48. Cambridge University Press, Cambridge, 2019.