Gerard Hooghiemstra
Delft Institute of Applied Mathematics
Delft University of Technology
g.hooghiemstra@tudelft.nl

Piet Van Mieghem
Network Architectures and Services
Delft University of Technology
p.f.a.vanmieghem@tudelft.nl

Research

An inequality of Gauss

In this article Gerard Hooghiemstra and Piet van Mieghem explain why the Gauss inequality is overshadowed by the Chebychev inequality. Furthermore, they present a proof of the Gauss inequality in modern notation.

In his general analysis on measurement errors [4], Gauss treats random errors in a probabilistic manner, which is, with respect to today's standards, surprisingly modern. Our interest here is a particularly general inequality involving probabilities, stated and proved by Gauss in [4], that does not seem to be well-known in the stochastic community. The Gauss inequality is, for instance, not mentioned in either of the two introductory volumes [2–3] written by William Feller. The main goal of this note is to explain why this inequality is overshadowed by the well-known Chebychev inequality and to present (a slightly modified) proof of the Gauss inequality.

We start by restating Gauss' inequality in modern notation. We consider a random variable X having a density f_X, which is symmetric around 0 (i.e., $f_X(-x) = f_X(x)$, $\forall x > 0$), and which is non-increasing for $x > 0$. Moreover, we assume a finite second moment $E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx < \infty$. By symmetry, the first moment (also called the mean μ) satisfies $\mu = E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx = 0$, so that the variance of X satisfies:

$$\sigma^2 = \text{Var}[X] = E[(X-\mu)^2] = E[X^2].$$

Theorem 1 (Gauss [4]). *Consider a random variable X having a density f_X, which is symmetric around 0 (i.e., $f_X(-x) = f_X(x)$, $\forall x > 0$), and which is non-increasing for $x > 0$. Moreover, we assume that $\sigma^2 = E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx < \infty$. Defining for $a > 0$*

$$m = \Pr[|X| \le a\sigma],$$

it holds that

$$\text{if} \quad m \le \frac{2}{3} \quad \text{then} \quad a \le m\sqrt{3}, \tag{1}$$

$$\text{if} \quad m > \frac{2}{3} \quad \text{then} \quad a \le \frac{2}{3\sqrt{1-m}}, \tag{2}$$

At first glance, the two inequalities (1) and (2) provide little insight. After rearranging, (1) and (2) can be rewritten as,

$$\text{if} \quad m \le \frac{2}{3} \quad \text{then} \quad m \ge \frac{a}{\sqrt{3}}, \tag{3}$$

$$\text{if} \quad m > \frac{2}{3} \quad \text{then} \quad m \ge 1 - \frac{4}{9a^2}. \tag{4}$$

The conclusions in (3) and (4) are somewhat peculiar, since the magnitude of m ($m \le \frac{2}{3}$ or $m > \frac{2}{3}$) is needed, before the respective statement gives a lower bound for m. The conclusion in (4) is

$$\Pr[|X| \le a\sigma] \ge 1 - \frac{4}{9a^2}, \tag{5}$$

which is valid in the tail of the distribution, i.e., for a large enough such that $\Pr[|X| > a\sigma] < \frac{1}{3}$, very closely resembles the inequality of Chebychev given below. The inequality of Chebychev below involves the mean $\mu = E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$ of X, which, in general, is unequal to 0.

Chebychev's inequality

We assume that X has a finite second moment. Let us denote the mean by $\mu = E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$ and the variance by $\sigma^2 = \text{Var}[X] = E[(X-\mu)^2] = E[X^2] - \mu^2$. In 1867, Chebychev [6] has proved that

$$\Pr[|X-\mu| \le a\sigma] \ge 1 - \frac{1}{a^2}, \quad a > 0. \tag{6}$$

The proof of Chebychev's inequality [3, p. 151] or [5, p. 103] needs a few lines only:

1. apply the Markov inequality to

$$Y = ((X - \mu)/\sigma)^2 \geq 0,$$

where

2. the Markov inequality is given by

$$\begin{aligned} E[Y] &= \int_0^\infty y f_Y(y)\, dy \\ &\geq \int_a^\infty y f_Y(y)\, dy \\ &\geq a \int_a^\infty f_Y(y)\, dy \\ &= a \Pr[Y > a]. \end{aligned}$$

Comparison of the two inequalities

Before we proceed with the comparison of the inequalities of Gauss and Chebychev, let us first treat some examples.

Example 1. Let X possess a uniform distribution on the interval $(-s, s)$, i.e.,

$$f_X(x) = \begin{cases} \frac{1}{2s}, & -s < x < s, \\ 0, & |x| \geq s. \end{cases}$$

The distribution function $F_X(x)$, defined by $F_X(x) = \Pr[X \leq x]$, for real numbers x, reads:

$$F_X(x) = \begin{cases} 0, & x \leq -s, \\ \int_{-s}^x \frac{1}{2s} dt = \frac{x+s}{2s}, & -s < x < s, \\ 1, & x \geq s. \end{cases} \quad (7)$$

In this example, $\mu = E[X] = 0$ and $\sigma^2 = \int_{-s}^s \frac{x^2}{2s}\, dx = s^2/3$, so that $\sigma = s/\sqrt{3}$. By straightforward calculation, we have

$$\begin{aligned} &\Pr[|X| \leq a\sigma] \\ &= F_X(\min\{s, a\sigma\}) - F_X(-\min\{s, a\sigma\}) \\ &= \min\{1, a/\sqrt{3}\}. \end{aligned}$$

Example 2. We perform the same computations for X, now having a normal distribution with parameters $\mu = 0$ and $\sigma^2 = E[X^2]$. The probability distribution function

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2\sigma}\, dt,$$

cannot be expressed in elementary functions, but the specific probabilities can be found from tables of the standard normal distribution. An accurate series for the inverse $F_X^{-1}(x)$ exists [5, p. 44]. Let Z have a standard normal distribution, i.e., a normal distribution with parameters $\mu = 0$ and $\sigma^2 = 1$, then

$$\begin{aligned} &\Pr[|X| \leq a\sigma] \\ &= \Pr[X \leq a\sigma] - \Pr[X \leq -a\sigma] \\ &= \Pr[Z \leq a] - \Pr[Z \leq -a] \\ &= 1 - 2\Pr[Z > a], \end{aligned}$$

where the probability $\Pr[Z > a]$ can be found in many places, for instance, in [1, Table B.1, p. 432].

Example 3. As a third example, we take a symmetric distribution with heavy tails. Roughly speaking, a distribution has a heavy tail, if the survival function $\Pr[|X| > t]$ decays polynomially in t. A well-known example is the Pareto distribution [1, p. 63]. A random variable X is said to have a Pareto distribution with parameter $\alpha > 0$, if its probability density $g_\alpha(x) = 0$ is, for $x < 1$, and equal to

$$g_\alpha(x) = \frac{\alpha}{x^{\alpha+1}}, \quad \text{for} \quad x \geq 1.$$

To satisfy the conditions of Theorem 1, we make the density f_X symmetric by defining,

$$f_X(x) = \begin{cases} \frac{1}{2} g_\alpha(1 + x), & x \geq 0, \\ \frac{1}{2} g_\alpha(1 - x), & x \leq 0. \end{cases}$$

Rather than computing the distribution function F_X, we instead derive $\Pr[|X| \leq a\sigma]$ directly from the density f_X. By construction, $E[X] = 0$ and the second moment is

$$\begin{aligned} \sigma^2 = E[X^2] &= \int_{-\infty}^\infty x^2 f_X(x)\, dx \\ &= 2\int_0^\infty x^2 \frac{1}{2} g_\alpha(1 + x)\, dx \\ &= \int_1^\infty \frac{\alpha(x-1)^2}{x^{\alpha+1}}\, dx \\ &= \frac{2}{(\alpha - 1)(\alpha - 2)}, \end{aligned}$$

since $\int_1^\infty \frac{\alpha x^\beta}{x^{\alpha+1}}\, dx = \alpha/(\alpha - \beta)$ for $\alpha > \beta$. Hence, we need to require that $\alpha > 2$ in order to have a finite variance $E[X^2] < \infty$. We shall take $\alpha = 3$ (and hence $\sigma = 1$) and find by integration:

$$\begin{aligned} \Pr[|X| \leq a\sigma] &= 2\int_0^a \frac{1}{2} g_3(1 + x)\, dx \\ &= \int_0^a \frac{3}{(x+1)^4}\, dx \\ &= 1 - \frac{1}{(1+a)^3}. \end{aligned}$$

In Table 1, we present $\Pr[|X| \leq a\sigma]$ for the distributions of Examples 1,2 and 3 and compare for various values of a the lower bounds of Gauss (5) and Chebychev (6), respectively. In the table, the abbreviation 'n.r.' stands for 'not relevant', since for the corresponding values of a in all three cases, we have $m > \frac{2}{3}$.

The lower bound (5) of Gauss is in all examples tighter than (6) of Chebychev. However, two remarks are in order: (i) the lower bound of Gauss is only valid under more stringent conditions, but more importantly (ii) we must know, whether $m = \Pr(|X| < a\sigma)$ is larger or smaller than $\frac{2}{3}$, which is not straightforward at all.

a	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$	3
$\Pr[\lvert X\rvert < a\sigma]$ for Example 1.	0.289	0.577	0.866	1	1	1
$\Pr[\lvert X\rvert < a\sigma]$ for Example 2.	0.383	0.683	0.866	0.954	0.988	0.997
$\Pr[\lvert X\rvert < a\sigma]$ for Example 3.	0.704	0.875	0.936	0.963	0.977	0.984
Lower bound from Gauss ($m \leq \frac{2}{3}$) (cf. (3))	0.289	0.577	n.r.	n.r.	n.r.	n.r.
Lower bound from Gauss ($m > \frac{2}{3}$) (cf. (4))	−0.778	0.556	0.802	0.889	0.929	0.951
Lower bound from Chebychev (cf. (6))	−3	0	0.556	0.750	0.840	0.889

Table 1 Comparison table.

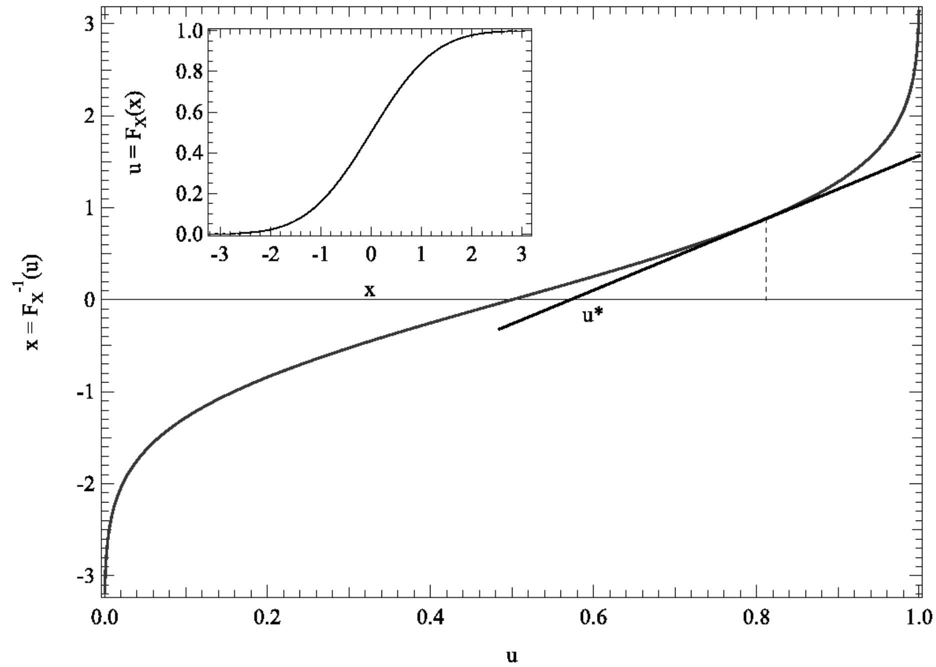

Figure 1 A distribution function $u{=}F_X(x)$ (in the inset) of a random variable X that is symmetric around $x{=}0$ and its inverse function $x{=}F_X^{-1}(u)$.

One might consider to approximate m by the lower bound, because we know that m falls in between the lower bound and 1. This approximation is rather crude, however in the tail ($m > \frac{2}{3}$), the lower bound of Gauss is definitely better than that of Chebychev. Also note that in case (i), the uniform distribution, the lower bound of Gauss gives the exact values for $m = Pr(|X| < a\sigma)$, for $m < \frac{2}{3}$.

For the theoretical value of the lower bounds, we consider an important application, namely the weak law of large numbers [3, p. 234]. Informally, the weak law of large numbers states that the average of repetitive and independent measurements converges (in probability) to the mean of the distribution. Indeed, for a series of repetitive and independent measurements $X_1, X_2, \ldots, X_n$ with density satisfying the conditions of Theorem 1, the mean of the underlying distribution is 0 and

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

converges to 0, in the sense that for each $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr[|\bar{X}_n| > \varepsilon] = 0.$$

This follows directly from the inequality (5) of Gauss, since

$$\begin{aligned}\text{Var}\,[\bar{X}_n] &= \text{Var}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\ &= \frac{1}{n^2}\text{Var}\,[X_1 + X_2 + \cdots + X_n] \\ &= \frac{1}{n^2}\{\text{Var}\,[X_1] + \cdots + \text{Var}\,[X_n]\} \\ &= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.\end{aligned}$$

Indeed, we find that, for $n \to \infty$, and with $a = \varepsilon\sqrt{n}/\sigma$,

$$\begin{aligned}\Pr[|\bar{X}_n| > \varepsilon] &= 1 - \Pr[|\bar{X}_n| < \varepsilon] \\ &= 1 - \Pr\left[|\bar{X}_n| < \frac{\sigma}{\sqrt{n}}\varepsilon\frac{\sqrt{n}}{\sigma}\right] \\ &\le \frac{4\sigma^2}{9\varepsilon^2 n} \to 0.\end{aligned}$$

However, the same conclusion can be drawn by applying the Chebychev inequality (6), in which case the upper bound is replaced by

$$\frac{1}{\left(\varepsilon\frac{\sqrt{n}}{\sigma}\right)^2} = \frac{\sigma^2}{\varepsilon^2 n},$$

which also converges to 0. Hence, for theoretical purposes, the advantage of the factor $\frac{4}{9}$ in Gauss' inequality (5) compared to Chebychev's inequality (6) is unimportant and is washed out entirely by the fact that Chebychev's inequality holds under the single condition that X must have a finite second moment. We believe that this explains why Gauss' inequality (5) is barely known in the stochastic community.

Proof of the Gauss inequality

In this section we present a proof of the Gauss inequality in modern notation. In his proof [4] in Latin (translated to English in [5, pp. 111–112]), Gauss uses the inverse of the function h defined by $h(x) = F_X(x) - F_X(-x)$, $x \ge 0$. It is slightly easier to concentrate on the inverse function F_X^{-1}, which we define below. Since, in the framework of Theorem 1, we exclusively work with continuous distribution functions and since these functions are by definition non-decreasing, we can define

$$F_X^{-1}(u) = \inf\{x : F(x) = u\}, \quad 0 < u < 1,$$

and, on intervals where $f_X(x) = 0$, or similarly, where $F_X(x)$ is constant, we take the left-endpoint of that interval.

The general definition of the expectation of a function g of X is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)dF_X(x). \tag{8}$$

After the substitution $x = F_X^{-1}(u)$ or $u = F_X(x)$ and $du = dF_X(x) = f_X(x)dx$, we obtain

$$E[g(X)] = \int_0^1 g\left(F_X^{-1}(u)\right)du,$$

from which the mean

$$\mu = E[X] = \int_0^1 F_X^{-1}(u)\,du$$

and the second moment

$$E[X^2] = \int_0^1 (F_X^{-1}(u))^2\,du$$

follows. A probabilistic way to obtain the same result is as follows. Let U be a uniform random variable on $(0, 1)$, then for all real numbers x,

$$\{U \le F_X(x)\} = \{F_X^{-1}(U) \le x\}. \tag{9}$$

For a random variable with a uniform distribution on $(0, 1)$, we have

$$\Pr[U \le u] = \int_0^u dx = u, \quad 0 < u < 1,$$

so that substitution of $u = F_X(x)$ yields

$$\Pr[U \le F_X(x)] = F_X(x). \tag{10}$$

Combining (9) and (10) gives:

$$\Pr[F_X^{-1}(U) \le x] = F_X(x),$$

so that X and $F_X^{-1}(U)$ are equal in distribution. Thus, also the expectations $E\left[g(X)\right]$ and $E\left[g\left(F_X^{-1}(U)\right)\right]$ are equal, for any function g. Invoking the general definition (8), we find again

$$E\left[g(X)\right] = \int_0^1 g\left(F_X^{-1}(u)\right) dF_U(u) = \int_0^1 g\left(F_X^{-1}(u)\right) du.$$

After this preparation, we start with the proof. Since Gauss assumed that f_X is symmetric around 0 and that $f_X(x)$ is non-increasing for $x > 0$, the function $u = F_X(x)$ is concave for $x > 0$. As a consequence and also illustrated in Figure 1, the inverse function $x = F_X^{-1}(u)$ is convex for $u \in [\frac{1}{2}, 1]$.

The idea of the proof is that, for the uniform distribution on a symmetric interval around zero, the inequality (1) is sharp for $a \le \sqrt{3}$, as was shown in Example 1, where we derived that $m = a/\sqrt{3}$ for $a \le \sqrt{3}$. Since the uniform distribution function is a linear function on its support (see (7)), we will replace $F_X^{-1}(u)$ on a sub-interval of $[\frac{1}{2}, 1]$ by the tangent to the function $F_X^{-1}(u)$ in the point $u = F_X(a\sigma)$, where a is any positive real number (see Figure 1). From the basic identity $F_X^{-1}(F_X(y)) = y$, we find that $(F_X^{-1})'(F_X(y))f_X(y) = 1$. Hence, the equation of the tangent at $u = F_X(a\sigma)$ reads

$$x - a\sigma = \frac{1}{f_X(a\sigma)}(u - F_X(a\sigma)).$$

The intersection of the tangent to the function $F_X^{-1}(u)$ at $u = F_X(a\sigma)$ with the u-axis is given by $u^\star = F_X(a\sigma) - a\sigma f_X(a\sigma)$.

Now, by symmetry of $f_X(x)$, the relation $F_X(x) = 1 - F_X(-x)$, $x > 0$, holds, so that $F_X^{-1}(\frac{1}{2} + u) = -F_X^{-1}(\frac{1}{2} - u)$, $\frac{1}{2} < u < 1$, and as a consequence

$$\sigma^2 = E[X^2] = \int_0^1 (F_X^{-1}(u))^2\, du = 2\int_{\frac{1}{2}}^1 (F_X^{-1}(u))^2\, du. \tag{11}$$

Since F_X^{-1} is convex on $[\frac{1}{2}, 1]$, the tangent does not intersect the graph of $F_X^{-1}(u)$, and the intersection $u^\star$ of the tangent with the u-axis satisfies $u^\star \ge \frac{1}{2}$, so that the following inequalities are satisfied (note that we first use that $u^\star \ge \frac{1}{2}$ and secondly that $F_X^{-1}(u) \ge \frac{u-u^\star}{f_X(a\sigma)}$; when the inequalities are performed the other way around, the reasoning is false),

$$2\int_{\frac{1}{2}}^1 (F_X^{-1}(u))^2\, du \ge 2\int_{u^\star}^1 (F_X^{-1}(u))^2\, du \ge 2\int_{u^\star}^1 \left(\frac{u - u^\star}{f_X(a\sigma)}\right)^2 du. \tag{12}$$

A simple computation gives

$$2\int_{u^\star}^1 \left(\frac{u - u^\star}{f_X(a\sigma)}\right)^2 du = \frac{2}{3(f_X(a\sigma))^2} \cdot \left[1 - F_X(a\sigma) + a\sigma f_X(a\sigma)\right]^3. \tag{13}$$

After combining (11), (12) and (13), we end up with

$$\sigma^2 \ge \frac{2}{3(f_X(a\sigma))^2} \cdot \left[1 - F_X(a\sigma) + a\sigma f_X(a\sigma)\right]^3. \tag{14}$$

Let $z = u - u^\star = a\sigma f_X(a\sigma)$ and recall that $m = F_X(a\sigma) - F_X(-a\sigma) = 2F_X(a\sigma) - 1$. Substitution in (14) yields

$$\frac{2a^2\sigma^2}{3z^2}\left[\frac{1-m}{2} + z\right]^3 \le \sigma^2. \tag{15}$$

Define the function $G(z)$ by the left-hand side of (15). Obviously $z = u - u^\star > 0$. On the other hand $z \le m/2$, since by hypothesis f_X is non-increasing on $(0, \infty)$, so that for $x > 0$,

$$x f_X(x) \le \int_0^x f_X(y)\, dy = F_X(x) - F_X(0) = F_X(x) - \tfrac{1}{2} \qquad (F_X(0) = \tfrac{1}{2}),$$

and if we take $x = a\sigma$, we obtain: $a\sigma f_X(a\sigma) \le F_X(a\sigma) - \frac{1}{2}$ or $z \le m/2$.

In order to find the minimum value of $G(z)$ on the interval $(0, m/2]$, we compute the derivative

$$G'(z) = \frac{2a^2\sigma^2}{3z^2}\left[\frac{1-m}{2} + z\right]^2 \left[1 - \frac{1-m}{z}\right].$$

The minimum of G is attained at $z = 1 - m$, when $1 - m \le m/2$, or equivalently for $m \ge \frac{2}{3}$, and in the point $z = m/2$, when $1 - m > m/2$ or $m < \frac{2}{3}$. Substitution of $z = 1 - m$, which corresponds to $m \ge \frac{2}{3}$, gives

$$\frac{9}{4}a^2(1-m) \le 1 \quad \text{or} \quad a \le \frac{2}{3\sqrt{1-m}}.$$

For $m < \frac{2}{3}$, we obtain

$$\frac{2a^2\sigma^2}{3(m/2)^2} \cdot \frac{1}{8} \le \sigma^2 \quad \text{or} \quad a \le m\sqrt{3}.$$

This yields (1) and (2), since for $m = \frac{2}{3}$ we have $m\sqrt{3} = \frac{2}{3\sqrt{1-m}} = \frac{2}{3}\sqrt{3}$.

References

1 F.M. Dekking, C. Kraaikamp, H. P. Lopuhaä and L.E. Meester, *A Modern Introduction to Probability and Statistics*, Springer, London, 2005.

2 W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, John Wiley & Sons, New York, 3rd edition, 1970.

3 W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2. John Wiley & Sons, New York, 2nd edition, 1971.

4 C.F. Gauss, Theoria combinationis observationum erroribus minimus obnoxiae (pars prior), *Gauss Werke* 4, Februari 1821, pp. 3–26.

5 P. Van Mieghem, *Performance Analysis of Complex Networks and Systems*, Cambridge University Press, Cambridge, 2014.

6 P. Tchebichef, Des valeurs moyennes, *Journal de mathématiques pures et appliquées*, 22(2) (1867), 177–184.