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.

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

\[
\Pr (|X| \leq a \sigma) \geq 1 - \frac{4}{9a^2},
\]

which is valid in the tail of the distribution, i.e., for \( a \) large enough such that \( \Pr (|X| > a \sigma) < \frac{1}{9} \), very closely resembles the inequality of Chebychev given below. The inequality of Chebychev below involves the mean \( \mu = E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx \) and the variance by \( \sigma^2 = \text{Var}(X) = E[X^2] - \mu^2 \). In 1867, Chebychev [6] has proved that

\[
\Pr (|X - \mu| \leq a \sigma) \geq 1 - \frac{1}{a^2}, \quad a > 0.
\]

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

\[
\int_{-\infty}^{\infty} |x - \mu|^2 f_X(x) \, dx \leq \int_{-\infty}^{\infty} |x|^2 f_X(x) \, dx = \sigma^2.
\]
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:

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

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

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

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

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

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

$$\Pr[|X| \leq a\sigma] = \Pr[X \leq a\sigma] - \Pr[X \leq -a\sigma] = \Pr[Z \leq a] - \Pr[Z \geq a] = 1 - 2\Pr[Z > a],$$

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 $g_a(x)$ is, for $x < 1$, and equal to

$$g_a(x) = \frac{\alpha}{x^{\alpha+1}}, \quad x > 1.$$

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

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

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

$$\sigma^2 = E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) \, dx = 2 \int_{0}^{\infty} \frac{1}{2} g_a(1+x) \, dx = \int_{1}^{\infty} \frac{\alpha(x-1)^2}{x^{\alpha+1}} \, dx = \frac{2}{\alpha} \left(1 + \frac{1}{\alpha - 1}\right),$$

since $\int_{1}^{\infty} \frac{\alpha x^\beta}{x^{\alpha+1}} \, dx = \alpha/\sqrt{\alpha - \beta}$ 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:

$$\Pr[|X| \leq a\sigma] = 2 \int_{0}^{a} \frac{1}{2} g_a(1+x) \, dx = \int_{1}^{a} \frac{3}{(x+1)^2} \, dx = 1 - \frac{2}{(1+a)^2}.$$
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 \( (m > \frac{1}{2}) \), 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 value that in case (i), the uniform distribution, the lower bound of Gauss gives the exact value.

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 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

\[
\text{Var}(\bar{X}_n) = \text{Var}(\frac{X_1 + X_2 + \cdots + X_n}{n}) = \frac{1}{n^2} \text{Var}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n} \text{Var}(X_1) + \cdots + \text{Var}(X_n) = \frac{1}{n} \sigma^2 = \frac{\sigma^2}{n}.
\]

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

\[
\Pr(|\bar{X}_n| > \varepsilon) = 1 - \Pr(|\bar{X}_n| < \varepsilon) = 1 - \Pr\left(|\bar{X}_n| < \frac{\sigma}{\sqrt{n}} \varepsilon \sqrt{n} \right) \leq \frac{4a^2}{9\varepsilon^4 n} \to 0.
\]

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

\[
\frac{1}{(\varepsilon \sqrt{n})^2} \leq \frac{\sigma^2}{\varepsilon^2 n}.
\]

which also converges to 0. Hence, for theoretical purposes, the advantage of the factor \( \frac{3}{8} \) 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 \( \chi \) 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 \( h(x) = F_x(x) - F_x(-x), x \geq 0 \). 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 \( h(x) = 0 \), 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

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

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

\[
E[g(X)] = \int_{0}^{1} g(F_{x}^{-1}(u)) du,
\]

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 \( x \),

\[
\{U \leq F_X(x)\} = \{F_{x}^{-1}(U) \leq x\},
\]

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

\[
\Pr[U \leq u] = \int_{0}^{u} dx = u, \quad 0 < u < 1,
\]

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

\[
\Pr[U \leq F_X(x)] = F_X(x).
\]
Combining (g) and (10) gives:
\[
\Pr[F_X^{-1}(U) \leq x] = F_X(x),
\]
so that \(X\) and \(F_X^{-1}(U)\) are equal in distribution. Thus, also the expectations \(E[g(X)]\) and \(E[g(F_X^{-1}(U))]\) are equal, for any function \(g\). Invoking the general definition (8), we find again
\[
E[g(X)] = \int_0^1 g(F_X^{-1}(u))^2 \, du = \int_0^1 g(F_X^{-1}(u))^2 \, du.
\]
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 \(a \leq \sqrt{3}\), as was shown in Example 1, where we derived that \(m = a/\sqrt{3}\) for \(a \leq \sqrt{3}\). Since the uniform distribution function is a linear function on its support (see (g)), 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 2). From the basic identity \(F_X(F_X(y)) = y\), 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 \(u^* = F_X(a\sigma) - a\sigma f_X(a\sigma)\).

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
\[
\sigma^2 = E[X^2] = \int_0^1 (F_X^{-1}(u))^2 \, du = 2 \int_{\frac{1}{2}}^1 (F_X^{-1}(u))^2 \, du.
\]
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 \(u^*\) of the tangent with the \(u\)-axis satisfies \(u^* \geq \frac{1}{2}\), so that the following inequalities are satisfied (note that we first use that \(u^* \geq \frac{1}{2}\) and secondly that \(F_X^{-1}(u) \geq \frac{u - u^*}{\sigma f_X(a\sigma)}\) when the inequalities are performed the other way around, the reasoning is false),
\[
2 \int_{\frac{1}{2}}^1 (F_X^{-1}(u))^2 \, du \geq 2 \int_{u^*}^1 (F_X^{-1}(u))^2 \, du \geq 2 \int_{u^*}^1 \left(\frac{u}{f_X(a\sigma)} - u^*\right)^2 \, du.
\]
A simple computation gives
\[
2 \int_{u^*}^1 \left(\frac{u}{f_X(a\sigma)} - u^*\right)^2 \, du = \frac{2}{3(f_X(a\sigma))^2} \cdot [1 - F_X(a\sigma)]^3.
\]
After combining (11), (12) and (13), we end up with
\[
\sigma^2 \geq \frac{2}{3(f_X(a\sigma))^2} \cdot [1 - F_X(a\sigma)]^3.
\]
Let \(z = u - u^* = a\sigma f_X(a\sigma)\) and recall that \(m = F_X(a\sigma) - F_X(-a\sigma) = 2F_X(a\sigma) - 1\). Substitution in (14) yields
\[
\frac{2a^2\sigma^2}{3z^2} \left[1 - \frac{m}{2} + z\right]^3 \leq \sigma^2. \tag{15}
\]
Define the function \(G(z)\) by the left-hand side of (15). Obviously \(z = u - u^* > 0\). On the other hand \(z \leq m/2\), since by hypothesis \(f_X\) is non-increasing on \((0, \infty)\), so that for \(x > 0\),
\[
x f_X(x) \leq \int_0^x f_X(y) \, dy = F_X(x) - F_X(0) = F_X(x) - F_X(0) = f_X(0) = \frac{1}{2},
\]
and if we take \(x = a\sigma\), we obtain:
\[a\sigma f_X(a\sigma) = F_X(a\sigma) - F_X(-a\sigma) \leq \frac{1}{2}\] or \(z \leq m/2\).

In order to find the minimum value of \(G(z)\) on the interval \((0, m/2)\), we compute the derivative
\[
G'(z) = \frac{2a^2\sigma^2}{3z^2} \left[1 - \frac{m}{2} + z\right]^2 \left[1 - \frac{m}{2} + z\right].
\]
The minimum of \(G\) is attained at \(z = 1 - m\), when \(1 - m \leq m/2\), or equivalently for \(m \geq \frac{3}{2}\), and in the point \(z = m/2\), when \(1 - m > m/2\) or \(m < \frac{3}{4}\). Substitution of \(z = 1 - m\), which corresponds to \(m \geq \frac{3}{2}\), gives
\[
\frac{9}{4} a^2(1 - m) \leq 1 \quad \text{or} \quad a \leq \frac{2}{\sqrt{3}}.
\]
For \(m < \frac{3}{4}\), we obtain
\[
\frac{2a^2\sigma^2}{3(m/2)^2} \leq \frac{1}{2} \quad \text{or} \quad a \leq \frac{m\sqrt{3}}{2}.
\]
This yields (1) and (2), since for \(m = \frac{3}{4}\) we have \(m\sqrt{3} = \frac{3}{4}\) \(\sqrt{3}\).