The Three Sigma Rule

Friedrich PUKELSHEIM

For random variables with a unimodal Lebesgue density the 3σ rule is proved by elementary calculus. It emerges as a special case of the Vysochanskiĭ–Petunin inequality, which in turn is based on the Gauss inequality.

KEY WORDS: Bienaymé–Chebyshev inequality; Gauss inequality; Vysochanskiĭ–Petunin inequality

1. INTRODUCTION

Let X be a real random variable with mean μ and variance σ^2. The celebrated inequality of Bienaymé (1853) and Chebyshev (1867) says that X falls outside the interval with center μ and radius $r > 0$ at most with probablity by σ^2/r^2. This is a rough, general bound, and there have been many efforts to find additional assumptions on X which lead to particular, tighter bounds. Savage (1961) reviews probability inequalities of this type, and includes an excellent bibliography.

One such assumption is unimodality, that is, the distribution of X permits a Lebesgue density f that is nondecreasing up to a mode ν and nonincreasing thereafter. At the mode f may be infinite, or there may be more than a single mode. For this class the bound becomes $(4/9)\left(\sigma^2/r^2\right)$ for all $r > 1.63\,\sigma$, more than halving the Bienaymé–Chebyshev bound. The case $r = 3\sigma$ is the 3σ rule, that the probability for X falling away from its mean by more than 3 standard deviations is at most 5 percent,

$$\mathrm{P}(|X - \mu| \geq 3\sigma) \leq \frac{4}{81} < 0.05.$$

The bound for unimodal distributions is due to Vysochanskiĭ and Petunin (1980). In a follow-up paper the authors (1983) mention that their result extends to an arbitrary center $\alpha \in \mathbb{R}$, rather than being restricted to the mean $\alpha = \mu$. If $\alpha = \nu$ is a mode

Friedrich Pukelsheim is Professor, Institut für Mathematik, Universität Augsburg, 86135 Augsburg, Germany.

of f, then a forerunner is the Gauss (1821) inequality. In this note we review the Gauss inequality and present a streamlined proof of the Vysochanskiĭ–Petunin inequality, just drawing on elementary calculus.

Ulin (1953) treats the problem as one of finding an extremal member in the class of unimodel distribution functions. This approach is perfected by Dharmadikhari and Joag-dev (1985; 1988, p. 29) who use a Choque representation for a general unimodal distribution, and then concentrate on degenerate distributions which are the extreme points of the convex set of all unimodal distributions. They also obtain an extension of the Vysochanskiĭ–Petunin inequality using higher than second moments.

For the Gauss inequality, the extension to higher moments is due to Winckler (1866, p. 21), see also Krüger (1897). Although these references are quoted in the textbook by Helmert (1907, p. 345) they seem to have largely gone unnoticed, in favor of Meidell (1922), Camp (1922; 1923), Narumi (1923).

Throughout this note we fix some radius $r > 0$. Let X be a real random variable. We assume that the distribution of X admits a Lebesgue density f which is unimodal with a mode $\nu \in \mathbb{R}$, that is, f is nondecreasing on $(-\infty, \nu)$ and nonincreasing on (ν, ∞).

2. THE GAUSS INEQUALITY

The Gauss inequality bounds the probability for a deviation from a mode ν. We present three proofs. The first expands on the arguments of Gauss (1821, Art. 10, pp. 10–11). Let Φ be the distribution function of X. Gauss sets out with the *inverse* function Ψ of Φ, calculates the tangent to Ψ at $\Phi(r)$, and then returns to the original random variable by means of the probability transform.

In the second proof we adapt this argument to the original distribution function Φ itself, by investigating the tangent to $1 - \Phi$ at r. The third proof is from Cramér (1946, p. 256), first studying integrands that are step functions and then extending the result to more general integrands. Cramér's approach is simpler, and also points towards the proof of the Vysochanskiĭ–Petunin inequality.

Gauss Inequality. With $\tau^2 = \mathrm{E}\left[(X - \nu)^2\right]$ the expected squared deviation from the mode ν, we have

$$\mathrm{P}\big(|X - \nu| \geq r\big) \leq \begin{cases} \dfrac{4\tau^2}{9r^2} & \text{for all } r \geq \sqrt{4/3}\tau, \\[2ex] 1 - \dfrac{r}{\sqrt{3}\tau} & \text{for all } r \leq \sqrt{4/3}\tau. \end{cases} \tag{1}$$

Proof. Prelude. The function $g(z) = f(\nu + z) + f(\nu - z)$ is on $(0, \infty)$ a Lebesgue density of the nonnegative random variable $Z = |X - \nu|$. The unimodality assumption makes g nonincreasing on $(0, \infty)$, with a mode 0. If the left hand side in (1) is 0 or if $\tau^2 = \infty$, then (1) is evidently true. Otherwise we have $g(r) > 0$, and $\tau^2 = \int_0^\infty z^2 g(z)\, dz < \infty$.

Theme (Gauss 1821). Upon setting $a = \sup\{z \geq 0 : g(z) > 0\} \in (0, \infty]$, the distribution function

$$\Phi(x) = \int_0^x g(z)\, dz : (0, a) \to (0, 1)$$

is strictly isotonic and differentiable, with derivative $\Phi'(x) = g(x)$. Its inverse

$$\Psi(y) = \Phi^{-1}(y) : (0, 1) \to (0, a)$$

then is also strictly isotonic and differentiable. The derivative $\Psi'(y) = 1/g\big(\Psi(y)\big)$ is nondecreasing, whence the function Ψ is convex.

Because of $\Phi(r) < 1$ we have $r \in (0, a)$. From $g(z) \geq g(r)$ for all $z \in (0, r)$ we obtain $\Phi(r) = \int_0^r g(z)\, dz \geq rg(r) > 0$. The number

$$\epsilon = 1 - \frac{rg(r)}{\Phi(r)} \in [0, 1)$$

is the key quantity, measuring how much the rectangle of height $g(r)$ and base running from 0 to r differs in size from the area under g up to r. With these preparations the tangent of Ψ at $\Phi(r)$ is

$$L(y) = \frac{1}{g(r)} y - \frac{\Phi(r)}{g(r)} + r = \frac{1}{g(r)} \big(y - \epsilon\Phi(r)\big).$$

Convexity of Ψ implies that L bounds Ψ from below, $L(y) \leq \Psi(y)$ for all $y \in (0, 1)$.

Setting $y_0 = \epsilon\Phi(r)$, we get $\big(L(y)\big)^2 \leq \big(\Psi(y)\big)^2$ for all $y \in (y_0, 1)$. Hence the integral $\int_{y_0}^1 \big(L(y)\big)^2\, dy$ has a value less than or equal to the integral $\int_{y_0}^1 \big(\Psi(y)\big)^2\, dy$. The latter is bounded from above by $\int_0^1 \big(\Psi(y)\big)^2\, dy = \mathrm{E}[\big(\Psi(Y)\big)^2] = \mathrm{E}[Z^2] = \tau^2$, where the random variable Y is taken to be uniformly distributed on $(0, 1)$.

Therefore τ^2 is bounded from below by the former integral,

$$\frac{1}{\big(g(r)\big)^2} \int_{y_0}^1 \big(y - y_0\big)^2\, dy = \frac{\big(1 - y_0\big)^3}{3\big(g(r)\big)^2} = \frac{r^2}{3\big(\Phi(r)\big)^2} \frac{\big(1 - \epsilon\Phi(r)\big)^3}{\big(1 - \epsilon\big)^2} = B(\epsilon),$$

say. The derivative of B as a function of $\epsilon \in (-\infty, 1)$ is

$$B'(\epsilon) = \frac{r^2 \big(1 - \epsilon\Phi(r)\big)^2}{3\Phi(r)(1 - \epsilon)^3} \left(\epsilon - 3 + \frac{2}{\Phi(r)}\right),$$

and changes sign around its zero $\epsilon_0 = 3 - 2/\Phi(r) < 1$. It follows that the global minimum of B is $B(\epsilon_0) = (9r^2/4)(1 - \Phi(r))$. But $\tau^2 \geq B(\epsilon_0)$ is the same as

$$P(|X - \nu| \geq r) = 1 - \Phi(r) \leq \frac{4\tau^2}{9r^2}, \quad \text{for all } r > 0. \tag{2}$$

Variation 1. Gauss' approach also applies to the distribution function Φ directly, without a detour via its inverse Ψ. The function $1 - \Phi(x)$ is convex since the derivative $-g(x)$ is nondecreasing. The tangent to $1 - \Phi$ at r is

$$L(x) = -g(r)x + g(r)r + 1 - \Phi(r) = -g(r)\left(x - \frac{1 - \epsilon\Phi(r)}{g(r)}\right).$$

Convexity entails $L(\sqrt{x}) \leq 1 - \Phi(\sqrt{x}) = P(Z > \sqrt{x}) = P(Z^2 > x)$, for all $x > 0$. Setting $x_0 = (1 - \epsilon\Phi(r))/g(r)$, the integral $\int_0^{x_0^2} L(\sqrt{x})\,dx$ has a value less than or equal to $\int_0^{x_0^2} P(Z^2 > x)\,dx$. The latter is estimated by $\int_0^\infty P(Z^2 > x)\,dx = E[Z^2] = \tau^2$. Therefore τ^2 is bounded from below by the former integral which is found to be equal to $B(\epsilon)$. Again $\tau^2 \geq B(\epsilon_0)$ establishes (2).

Variation 2 (Cramér 1946). Yet another proof of (2) is as follows. First we consider uniform densities $g(z) = \frac{1}{s}1_{(0,s)}(z)$. From $g(r) > 0$ we get $s > r$, whence (2) is equivalent to

$$s - r \leq \frac{4}{9r^2}\int_0^s z^2\,dz. \tag{3}$$

The difference of the two sides is nonnegative, $4s^3/(27r^2) - s + r = 4(s - 3r/2)^2(s + 3r)/(27r^2) \geq 0$.

Now we turn to an arbitrary nonincreasing density g on $(0, \infty)$, and define

$$s = r + \frac{1}{g(r)}\int_r^\infty g(z)\,dz.$$

Thus the rectangle with height $g(r)$ and base running from r to s has the same area $g(r)(s - r)$ that lies under the tail of g from r on. Hence (3) yields

$$\int_r^\infty g(z)\,dz = g(r)\left(s - r\right) \leq \frac{4}{9r^2}g(r)\int_0^s z^2\,dz.$$

The product $g(r)\int_0^s z^2\,dz$ splits into three terms,

$$g(r)\int_0^r z^2\,dz + \int_r^s z^2\left(g(r) - g(z)\right)dz + \int_r^s z^2 g(z)\,dz.$$

The first term is estimated by $g(r) \int_0^r z^2 \, dz \leq \int_0^r z^2 g(z) \, dz$, using $g(r) \leq g(z)$ for all $z \in (0, r)$. In the second term we have $g(r) - g(z) \geq 0$ for all $z \in (r, s)$. This leads to the estimate

$$
\begin{aligned}
\int_r^s z^2 \Big(g(r) - g(z) \Big) \, dz &\leq s^2 \int_r^s \Big(g(r) - g(z) \Big) \, dz \\
&= s^2 \left(g(r) \big(s - r \big) - \int_r^s g(z) \, dz \right) \\
&= s^2 \int_s^\infty g(z) \, dz \\
&\leq \int_s^\infty z^2 g(z) \, dz.
\end{aligned}
$$

The estimates of the first two terms plus the third term as is sum to τ^2, again establishing (2).

Coda. Quoting from Gauss (1821, p. 11), (1) *is easily concluded* from (2). Indeed, the convex function $1 - \Phi$ does not cut across any cord between the point $(0; 1) \in \mathbb{R}^2$ and any one of the points $\big(r; 4\tau^2/(9r^2) \big) \in \mathbb{R}^2$ that lie on and above its graph, respectively. The steepest cord provides the tightest upper bound for $1 - \Phi$. This determines $r = \sqrt{4/3}\,\tau$, see Figure 1. The proof is complete.

Figure 1. Among the cords through the point $(0; 1)$ and any point of the graph of the bounding function $(4/9)(r/\tau)^{-2}$, the one with $r/\tau = \sqrt{4/3}$ has a steepest slope.

3. THE VYSOCHANSKIĬ–PETUNIN INEQUALITY

The Vysochanskiĭ–Petunin inequality replaces the center ν, a mode, by an arbitrary center $\alpha \in \mathbb{R}$. Our presentation follows Vysochanskiĭ and Petunin (1980, pp. 28–34; 1983, p. 28), condensing some details without jeopardizing the conceptual view that, if a problem can be stated in terms of calculus, then it can be solved in the same terms.

Vysochanskiĭ–Petunin Inequality. With $\rho^2 = \mathrm{E}\left[(X - \alpha)^2\right]$ the expected squared deviation from an arbitrary point $\alpha \in \mathbb{R}$, we have

$$
\mathrm{P}\bigl(|X - \alpha| \geq r\bigr) \leq
\begin{cases}
\dfrac{4\rho^2}{9r^2} & \text{for all } r \geq \sqrt{8/3}\,\rho, \\[2ex]
\dfrac{4\rho^2}{3r^2} - \dfrac{1}{3} & \text{for all } r \leq \sqrt{8/3}\,\rho.
\end{cases}
\tag{4}
$$

Proof. Preamble. We reduce the problem as follows. We take $\alpha = 0$ since otherwise we switch to $X - \alpha$. If some mode of f is zero, $\nu = 0$, then the Gauss inequality proves the assertion. Otherwise we restrict attention to $\nu > 0$, since for $\nu < 0$ we study $-X$ instead. If $r \leq \rho$ then $4\rho^2/(3r^2) - 1/3 \geq 1$ proves (4); hence we assume $r > \rho$. This entails $\int_0^r f(x)\,dx > 0$. Else f vanishes on $(0, r)$, as well as on $(-\infty, 0)$ since it is nondecreasing up to $\nu > 0$; but $\int_r^\infty f(x)\,dx = 1$ contradicts the assumption $r > \rho$,

$$
r^2 = r^2 \int_r^\infty f(x)\,dx \leq \int_r^\infty x^2 f(x)\,dx = \rho^2.
$$

In summary, we consider $\alpha = 0 < \nu$ and $r > \rho$. This forces $\int_0^r f(x)\,dx > 0$.

The proof distinguishes two cases, essentially (but not quite so) whether the mode ν lies near the origin or far away. To be precise, we introduce the quantities

$$
\begin{aligned}
h &= \frac{1}{r} \int_0^r f(x)\,dx \in (0, \infty), \\
p &= \inf\bigl\{x \in (0, r) : f(x) \geq h\bigr\}, \\
q &= \sup\bigl\{x \in (0, r) : f(x) \geq h\bigr\},
\end{aligned}
$$

and discriminate between the two cases wheather $q < r$ and $q = r$.

Both cases use the inequality

$$
h \int_0^q x^2\,dx \leq \int_0^q x^2 f(x)\,dx.
\tag{5}
$$

To establish (5) we start from

$$\int_0^p \Big(h - f(x)\Big)\, dx \le \left(\int_0^p + \int_q^r\right) \Big(h - f(x)\Big)\, dx = \int_p^q \Big(f(x) - h\Big)\, dx.$$

Then we estimate $\int_0^p x^2 \Big(h - f(x)\Big)\, dx \le p^2 \int_0^p \Big(h - f(x)\Big)\, dx \le p^2 \int_p^q \Big(f(x) - h\Big)\, dx \le \int_p^q x^2 \Big(f(x) - h\Big)\, dx$. A rearrangement of terms leads to (5).

Large radius case $r > q$. In this case we have $0 < \nu \le q < r$. Let A be the area below h and above f over the interval (q, r),

$$A = \int_q^r \Big(h - f(x)\Big)\, dx \in (0, \infty).$$

The area A is reallocated over the interval $(0, q/k)$, for $k = 1, 2, \ldots$, see Figure 2.

The resulting function

$$f_k(x) = \begin{cases} h + \dfrac{A}{q/k} 1_{(0, q/k)}(x) & \text{for all } x \in (0, q), \\[2mm] f(x) & \text{otherwise,} \end{cases}$$

Figure 2. Large radius case $r > q$. The area A is reallocated uniformly over the interval $(0, q/k)$ to generate the unimodal densities f_k (bold line). This case results in the bound $4\rho^2/(9r^2)$.

is a Lebesgue density that is unimodal around 0. We apply the Gauss inequality (2) to the density $g_k(z) = f_k(z) + f_k(-z)$, and set $\tau_k^2 = \int_0^\infty z^2 g_k(z)\,dz$. Since f coincides with f_k outside $(-r, r)$, we obtain

$$\mathrm{P}(|X| \geq r) = \int_r^\infty g_k(z)\,dz \leq \frac{4\tau_k^2}{9r^2}.$$

The second moment $\tau_k^2 = \int_{-\infty}^\infty x^2 f_k(x)\,dx$ is further estimated using (5),

$$\tau_k^2 = \left(\int_{-\infty}^0 + \int_q^\infty \right) x^2 f(x)\,dx + h \int_0^q x^2\,dx + \frac{A}{q/k} \int_0^{q/k} x^2\,dx \leq \rho^2 + \frac{A}{3} \left(\frac{q}{k} \right)^2.$$

Now $k \to \infty$ proves $\mathrm{P}(|X| \geq r) \leq 4\rho^2/\left(9r^2\right)$, in the case $r > q$.

Small radius case $r = q$. In this case we have $f(-x) \leq f(x)$ for all $x \in (0, r)$. Hence we get $\int_{-r}^0 f(x)\,dx = \int_0^r f(-x)\,dx \leq \int_0^r f(x)\,dx$, see Figure 3.

We now introduce $t = \frac{1}{h} \int_{-r}^0 f(x)\,dx \in [0, r]$, and complement (5) with the inequality

$$h \int_{-t}^0 x^2\,dx \leq \int_{-r}^0 x^2 f(x)\,dx. \tag{6}$$

To establish (6) we use $h - f(x) \geq 0$ for all $x \in (-t, 0)$, and estimate

$$\int_{-t}^0 x^2 \left(h - f(x) \right) dx \leq t^2 \left(ht - \int_{-t}^0 f(x)\,dx \right) = t^2 \int_{-r}^{-t} f(x)\,dx \leq \int_{-r}^{-t} x^2 f(x)\,dx.$$
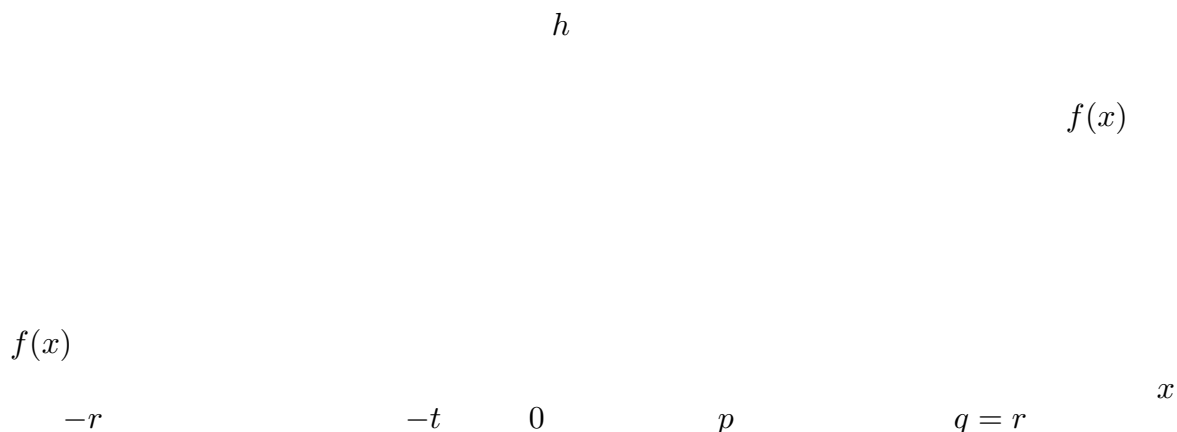
A rearrangement of terms leads to (6).

h

$f(x)$

$f(x)$

x

$-r \qquad\qquad -t \qquad 0 \qquad\qquad p \qquad\qquad q = r$

Figure 3. *Small radius case* $r = q$. *The rectangle with height* h *and base running from* $-t$ *to* 0 *has the same area that lies under* f *between* $-r$ *and* 0. *This case results in the bound* $4\rho^2/(3r^2) - 1/3$.

Finally we utilize (5) and (6) to obtain

$$\rho^2 \geq \int_{|x| \geq r} x^2 f(x)\, dx + h \int_{-t}^{r} x^2\, dx \geq r^2 \int_{|x| \geq r} f(x)\, dx + \frac{h}{3}\left(r^3 + t^3\right).$$

The last term is further estimated by

$$\frac{h}{3}\left(r^3 + t^3\right) = \frac{h}{3}\left(\left(\frac{r}{2} - t\right)^2 + \frac{3r^2}{4}\right)\left(r + t\right) \geq \frac{r^2}{4}\int_{-r}^{r} f(x)\, dx.$$

In $\rho^2 \geq r^2\, \mathrm{P}(|X| \geq r) + \left(r^2/4\right)\left(1 - \mathrm{P}(|X| \geq r)\right)$ we solve for $\mathrm{P}(|X| \geq r)$ and find $\mathrm{P}(|X| \geq r) \leq 4\rho^2/(3r^2) - 1/3$, in the case $r = q$.

Conclusion. The two cases combine into $\mathrm{P}(|X| \geq r) \leq \max\{4\rho^2/(9r^2),\, 4\rho^2/(3r^2) - 1/3\}$, which is the same as (4). The proof is complete.

Although a deviation from the mean looks like being a very particular case, $\alpha = \mu$, none of the above arguments simplify. For small radii, $r^2 < \sigma^2$, the bound $4\sigma^2/(3r^2) - 1/3$ exceeds 1 and is useless, as is the Bienaymé–Chebyshev bound σ^2/r^2. For a mode $\alpha = \nu$ the Gauss bound $4\tau^2/(9r^2)$ is tighter than $4\tau^2/(3r^2) - 1/3$ for all $r < \sqrt{8/3}\,\tau$. For $r < \sqrt{4/3}\,\tau$ none of these beat the Gauss bound $1 - r/\left(\sqrt{3}\,\tau\right)$. See Figure 4.
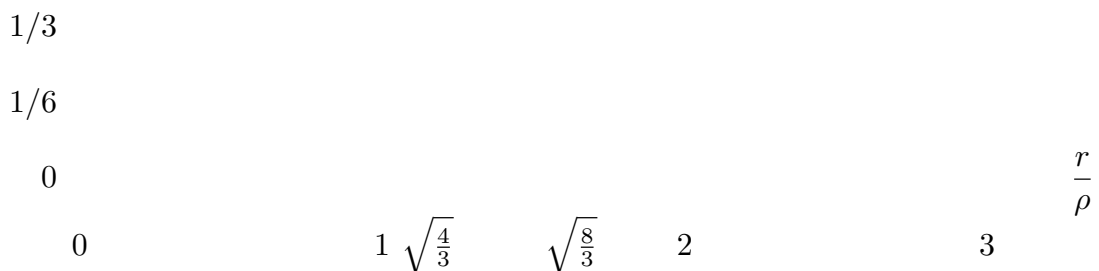

Figure 4. Bounds for $\mathrm{P}(|X - \alpha| \geq r)$, *in terms of* $\rho^2 = \mathrm{E}[(X - \alpha)^2]$. *Bottom line: Gauss (1821), for unimodal distributions, centered at a mode* $\alpha = \nu$. *Bold middle line: Vysochanskiĭ and Petunin (1980; 1983), for unimodal distributions, centered anywhere,* $\alpha \in \mathbb{R}$. *Top line: Bienaymé (1853) and Chebyshev (1867), for any distribution, centered at the mean,* $\alpha = \mu$.

REFERENCES

Bienaymé, J. (1853). Considérations à l'Appui de la Découverte de Laplace sur la Loi de Probabilité dans la Méthode des Moindres Carrés. *Compte Rendu des Séances de l'Académie des Sciences Paris 37,* 309–324.

Camp, B.H. (1922). A new Generalization of Tchebycheff's Statistical Inequality. *Bulletin of the American Mathematical Society 28,* 427–432.

―――― (1923). Note on Professor Narumi's Paper. *Biometrika 15,* 421–423.

Chebyshev [Tchebychef], P.L. (1867). Des Valeurs Moyennes. *Journal de Mathématiques Pures et Appliquées, 2 Série, 12,* 177–184. Also in: *Œuvres de P.L. Tchebychef, Tome 1,* Chelsea Publishing Company New York, pp. 687–694.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton, NJ: Princeton University Press.

Dharmadhikari, S. and Joag-dev, K. (1985). The Gauss–Tchebyshev Inequality for Unimodal Distributions. *Theory of Probability and its Applications 30,* 867–871.

―――― (1988). *Unimodality, Convexity, and Applications.* New York: Academic Press.

Gauss, C. F. (1821). Theoria Combinationis Observationum Erroribus Minimis Obnoxiae, Pars Prior. *Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores 5,* Also in *Werke, Band 4,* 1–93 (in Latin); German summary *ibidem,* 95–100.

Helmert, F.R. (1907). *Die Ausgleichungsrechnung nach der Methode der Kleinsten Quadrate, mit Anwendungen auf die Geodäsie, die Physik und die Theorie der Messinstrumente. Zweite Auflage.* Leipzig: Teubner.

Krüger, L. (1897). Ueber einen Satz der Theoria Combinationis. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch–physikalische Klasse, Heft 2,* 146–157.

Meidell, B. (1922). Sur un Problème du Calcul des Probabilités et les Statistique Mathématiques. *Comptes Rendus Hebdomaires des Séances de l'Académie des Sciences Paris 173,* 806–808.

Narumi, S. (1923). On Further Inequalities with Possible Application to Problems in the Theory of Probability. *Biometrika 15,* 245–253.

Savage, I.R. (1961). Probability Inequalities of the Tchebycheff Type. *Journal of Research of the National Bureau of Standards—B. Mathematics and Mathematical Physics 65B,* 211–222.

Ulin, B. (1953). An Extremal Problem in Mathematical Statistics. *Skandinavisk Aktuarietidskrift 36,* 158–167.

Vysochanskiĭ, D.F. and Petunin, Yu. I. (1980). Justification of the 3σ Rule for Unimodal Distributions. *Theory of Probability and Mathematical Statistics 21,* 25–36.

―――― (1983). A Remark on the Paper "Justification of the 3σ Rule for Unimodal Distributions". *Theory of Probability and Mathematical Statistics 27,* 27–29.

Winckler, A. (1866). Allgemeine Sätze zur Theorie der unregelmäßigen Beobachtungsfehler. *Sitzungsberichte der mathematisch–naturwissenschaftlichen Classe der kaiserlichen Akademie der Wissenschaften zu Wien 53(2),* 6–41.