- Equation (6.47) shows explicitly what is going on and why the result is consistent with the way we have modelled the uncertainties. In fact we have performed two independent calibrations: one of the offset and one of μ_1. The best estimate of the true value of the "zero" Z is the weighted average of the two measured offsets.
- The new uncertainty of μ_2 [see Eq. (6.45)] is a combination of σ_2 and the uncertainty of the weighted average of the two offsets. Its value is smaller than it would be with only one calibration and, obviously, larger than that due to the sampling fluctuations alone:

$$\sigma_2 \leq \sqrt{\sigma_2^2 + \frac{\sigma_1^2\,\sigma_Z^2}{\sigma_1^2 + \sigma_Z^2}} \leq \sqrt{\sigma_2^2 + \sigma_Z^2}\,. \tag{6.48}$$

6.12 The Gauss derivation of the Gaussian

It might be interesting to end this chapter in a historical vein, looking at how Gauss arrived at the distribution function which now carries his name. [68] Note that the Gaussian function was already known before Gauss, describing the asymptotical behavior of the binomial distribution, in a purely probabilistic context. The Gauss derivation arose in a more inferential framework and, indeed, Gauss used what we would nowadays call Bayesian reasoning.

Gauss's problem, expressed in modern terms, was: what is the more general form of the likelihood such that the maximum of the posterior of μ is equal to the arithmetic average of the observed values (and the function has some 'good' mathematical properties)?

In solving his problem, Gauss first derived a formula for calculating the probability of hypotheses given some observations had been made, under the assumption of equal prior probability of the hypotheses. In practice, he reobtained Bayes theorem (without citing Bayes) in the case of uniform prior. Note that the concept of prior (*"ante eventum cognitum"*) [4] was very clear and natural to him, opposed to the concept of posterior (*"post eventum cognitum"*). Then moving from discrete hypotheses to continuous observations x_i and true value μ (using our terminology), he looked for the functional form of φ, which describes the probability of obtaining x_i from μ (the likelihood, in our terms). Considering the observations to be

[4] All quotes in Latin are from Ref. [68].

independent, the joint distribution of the sample $\boldsymbol{x}$ is then given by

$$f(\boldsymbol{x}\,|\,\mu) = \varphi(x_1-\mu)\cdot\varphi(x_2-\mu)\cdot\;\cdots\;\cdot\varphi(x_n-\mu)\,. \tag{6.49}$$

At this point, two hypotheses enter.

(1) All values of μ are considered *a priori* (*"ante illa observationes"*) equally likely (*"... aeque probabilia fuisse"*).
(2) The maximum *a posteriori* (*"post illas observationes"*) is given by $\mu = \overline{x}$, arithmetic average of the n observed values.

The first hypothesis gives

$$f(\mu\,|\,\boldsymbol{x}) \propto f(\boldsymbol{x}\,|\,\mu) = \varphi(x_1-\mu)\cdot\varphi(x_2-\mu)\cdot\;\cdots\;\cdot\varphi(x_n-\mu)\,. \tag{6.50}$$

To use the second condition, he imposed that the first derivative of the posterior is null for $\mu = \overline{x}$:

$$\left.\frac{\mathrm{d}f(\mu\,|\,\boldsymbol{x})}{\mathrm{d}\mu}\right|_{\mu=\overline{x}} = 0 \implies \left(\frac{\mathrm{d}}{\mathrm{d}\mu}\prod_i \varphi(x_i-\mu)\right)_{\mu=\overline{x}} = 0\,, \tag{6.51}$$

i.e.

$$\sum_i \frac{\varphi'(x_i-\overline{x})}{\varphi(xi-\overline{x})} = 0\,, \tag{6.52}$$

where φ' stands for the derivative of ϕ with respect to μ. Calling ψ the function φ'/φ and indicating with $z_i = x_i - \overline{x}$ the differences from the average, which have to follow the constraint $\sum_i z_i = 0$, we have

$$\begin{cases} \sum_i \psi(z_i) = 0 \\ \sum_i z_i = 0 \end{cases}. \tag{6.53}$$

Since this relation must hold independently of n and the values of z_i, the functional form of $\psi(z)$ has to satisfy the following constraint:

$$\frac{1}{z}\,\psi(z) = k\,, \tag{6.54}$$

where k is a constant (note that the limit $z \to 0$ is not a problem, for the derivative of φ at $z = 0$ vanishes and the condition $\psi(z)/z = k$ implies that numerator and denominator have to tend to zero with the same speed). It follows that

$$\frac{\mathrm{d}\varphi}{\varphi} = k\,z\,\mathrm{d}z\,,$$

i.e.

$$\varphi(z) \propto e^{\frac{k}{2}\,z^2} = e^{-h^2\,z^2}\,, \tag{6.55}$$

where Gauss replaced $k/2$ by $-h^2$ to make its negative sign evident, because φ is required to have a maximum in $z = 0$. Normalizing the function dividing by its integral from $-\infty$ to ∞, an integral acknowledged to be due to Laplace (*"ab ill. Laplace inventum"*), he finally gets the 'Gauss' error function (*"functio nostra fiet"*):

$$\varphi(z) = \frac{h}{\sqrt{\pi}}\,e^{-h^2\,z^2}\,. \tag{6.56}$$