Some Basic Statistical Equations for Histogram Fitting

S.I. Redin
Budker Institute of Nuclear Physics, 11 Lavrentieva Avenue, Novosibirsk 630090, Russia and Department of Physics, Yale University, New Haven, CT 06520, USA

In this paper we consider fitting of a histogram, say a time distribution, with a fit function $f(\boldsymbol{x}, t)$, where $\boldsymbol{x} \equiv \{x_i\}$ is a vector of fit parameters. We start with the equation for statistical fluctuations of fit parameters as functions of fluctuations in the number of counts in the histogram channels, $\mathcal{N}_n$, and discuss the following topics:
- statistical errors and correlations of fit parameters;
- bias of fit parameters for $\chi^2 = \sum_n \frac{(f-\mathcal{N}_n)^2}{\sigma_n^2}$ minimization and for a binned likelihood function fit; and possible improvement of χ^2 fit;
- comparison of fit parameters obtained from fitting the full set of histogram channels with those obtained from fitting some subset of channels only (set-subset relations); set-subset relations for χ^2 values;
- improvement of errors of fit parameters due to additional (external) knowledge of linear combination[s] of those parameters;
- systematic shift of fit parameters due to neglected background.

Most of the presented equations are being used for the muon (g-2) experiment [1] data analysis.

1. STATISTICAL ERRORS AND CORRELATIONS

As a result of statistical fluctuations in individual histogram channels, minimization of

$$\chi^2 \equiv \sum_n \frac{\left(f(\boldsymbol{x}, t_n) - \mathcal{N}_n\right)^2}{\sigma_n^2} = \sum_n \frac{\left(f(\boldsymbol{x}, t_n) - \mathcal{N}_n\right)^2}{f(\boldsymbol{x}, t_n)} \quad (1)$$

gives a vector of "optimal" fit parameters $\boldsymbol{x}$, shifted with respect to the "true" value $\boldsymbol{x}_\circ$ by some $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_\circ$. Vector $\Delta\boldsymbol{x}$ is a function of fluctuations of the number of counts in a histogram's channels. Its elements can be derived from the χ^2 minimization requirement $\partial\chi^2/\partial x_i = 0$, which in general gives a system of nonlinear equations. The solution can be found by successive approximations $\Delta x_i = \Delta x_i^\circ + \Delta x_i^1 + \ldots$ with the leading approximation being

$$\Delta x_i^\circ = \sum_j \left(\mathcal{A}^{-1}\right)_{ij} \sum_n \frac{f_j'}{f}(\mathcal{N}_n - f) \quad (2)$$

which is of order $\frac{\mathcal{N}_n - f}{\mathcal{N}_n}$. The next-to-leading approximation Δx_i^1 is of order $\left(\frac{\mathcal{N}_n - f}{\mathcal{N}_n}\right)^2$, etc. Here

$$f_i' \equiv \frac{\partial f}{\partial x_i} \quad \text{and} \quad \mathcal{A}_{ij} = \sum_n \frac{f_i' f_j'}{f} \quad (3)$$

Important properties of fluctuations of the number of counts in a histogram's channels, $\mathcal{N}_n - f(t_n)$, are

$$\left\langle \mathcal{N}_n - f(t_n) \right\rangle = 0 \quad (4)$$
$$\left\langle \left(\mathcal{N}_n - f(t_n)\right) \left(\mathcal{N}_m - f(t_m)\right) \right\rangle = f(t_n)\, \delta_{nm} \quad (5)$$

where $\langle \ldots \rangle$ means average over an ensemble of similar histograms (ensemble average). Eqs. (2) to (5) are the basic elements in evaluation of various ensemble averages. The most fundumental is the correlation of fit parameters:

$$\begin{aligned} &\langle \Delta x_i\ \Delta x_j \rangle \approx \langle \Delta x_i^\circ\ \Delta x_j^\circ \rangle \\ &= \sum_{ab} \left(\mathcal{A}^{-1}\right)_{ia} \left(\mathcal{A}^{-1}\right)_{jb} \sum_{nm} \frac{f_a'}{f} \frac{f_b'}{f} \Big\langle (\mathcal{N}_n - f)(\mathcal{N}_m - f) \Big\rangle \\ &= \sum_{ab} \left(\mathcal{A}^{-1}\right)_{ia} \left(\mathcal{A}^{-1}\right)_{jb} \sum_n \frac{f_a' f_b'}{f} \\ &= \sum_{ab} \left(\mathcal{A}^{-1}\right)_{ia} \left(\mathcal{A}^{-1}\right)_{jb} \mathcal{A}_{ab} = \left(\mathcal{A}^{-1}\right)_{ij} \end{aligned} \quad (6)$$

As a specific, but most practical, case of eq. (6), one can immediately obtain equations for the statistical errors of fit parameters:

$$\sigma_i^2 \equiv \left\langle (\Delta x_i)^2 \right\rangle = \left(\mathcal{A}^{-1}\right)_{ii} \quad (7)$$

Equations (1) to (5) may be used to evaluate the mean (ensemble average) value of χ^2 itself:

$$\left\langle \chi^2 \right\rangle = N_{ch} - L \quad (8)$$

and the mean square variation of χ^2, $\sigma_{\chi^2}^2$, with respect to $\langle \chi^2 \rangle$:

$$\sigma_{\chi^2}^2 \equiv \left\langle \left(\chi^2 - \langle \chi^2 \rangle\right)^2 \right\rangle = 2N_{ch} - 2L \quad (9)$$

Here N_{ch} and L are number of histogram channels and number of fit parameters, respectively.

2. BIAS OF FIT PARAMETERS

Another important quantity is bias of fit parameters $\langle \Delta x_i \rangle$. Since ensemble averaging of Δx_i° in eq. (2) vanishes, $\langle \Delta x_i \rangle \approx \left\langle \Delta x_i^1 \right\rangle$. Ensemble averaging of Δx_i^1 for minimization of $\chi^2 \equiv \sum_n \frac{(\mathcal{N}_n - f)^2}{\sigma_n^2}$ gives:

$$\begin{aligned} \left\langle \Delta x_i^1 \right\rangle &= -\frac{1}{2} \sum_{jkl} \left(\mathcal{A}^{-1}\right)_{ij} \left(\mathcal{A}^{-1}\right)_{kl} \sum_n \frac{f'_j f''_{kl}}{f} \\ &+ \xi \sum_j \left(\mathcal{A}^{-1}\right)_{ij} \times \sum_n \frac{f'_j}{f} \\ &- \xi \sum_{jkl} \left(\mathcal{A}^{-1}\right)_{ij} \left(\mathcal{A}^{-1}\right)_{kl} \sum_n \frac{f'_j f'_k f'_l}{f^2} \end{aligned} \quad (10)$$

where $\xi = \frac{1}{2}$ for $\sigma_n^2 = f(\boldsymbol{x}_\circ, t_n)$ and $\xi = -1$ for $\sigma_n^2 = \mathcal{N}_n$. For a one parameter fit eq. (10) reads:

$$\left\langle \Delta x^1 \right\rangle = -\frac{1}{2} \sigma^4 \sum_n \frac{f' f''}{f} + \xi \sigma^2 \sum_n \frac{f'}{f} - \xi \sigma^2 \sum_n \frac{f'^3}{f} \quad (11)$$

The corresponding equations for the likelihood function fit are:

$$\left\langle \Delta x_i^1 \right\rangle = -\frac{1}{2} \sum_{jkl} \left(\mathcal{A}^{-1}\right)_{ij} \left(\mathcal{A}^{-1}\right)_{kl} \sum_n \frac{f'_j f''_{kl}}{f} \quad (12)$$

and

$$\left\langle \Delta x^1 \right\rangle = -\frac{1}{2} \sigma^4 \sum_n \frac{f' f''}{f} \quad (13)$$

which are the same as eqs. (10) and (11) with $\xi = 0$. Simple estimates show that the second term on the right side of eqs. (10), (11) in general supersedes other terms by a factor of order $N_{ch} \gg 1$. Thus the likelihood function fit, which does not contain such a term, generally has the smallest bias. The χ^2 fit, with $\sigma_n^2 = f$, has in general smaller bias (by a factor -2) than the fit with $\sigma_n^2 = \mathcal{N}_n$, though using the former is more complicated technically.

Detailed study reveals the reason for the differences in the biases as being the next-to-leading term in the Taylor expansion of likelihood and χ^2 functions in series $\left(\frac{f-\mathcal{N}_n}{\mathcal{N}_n}\right)^m$. As was shown in [2], simple modifications of the χ^2 function may reduce its bias to the level of bias for the likelihood function fit. Perhaps the simplest of such modifications is $\chi^2 = \left(\frac{(f-\mathcal{N}_n)^2}{\mathcal{N}_n} - \frac{2}{3}\frac{(f-\mathcal{N}_n)^3}{\mathcal{N}_n^2}\right)$ which is, in fact, nothing else but the first two terms of the Taylor expansion of the likelihood function (up to a factor of 2).

3. SET-SUBSET RELATIONS FOR THE χ^2 FIT

We denote $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to be vectors of fit parameters obtained from the χ^2 minimization fit for the full set of histogram channels Ω_1, and for some subset Ω_2, respectively. Denote $\boldsymbol{x}_\circ$ to be vector of “true” values of fit parameters, common for both Ω_1 and Ω_2, and $\Delta\boldsymbol{x}_1 \equiv \boldsymbol{x}_1 - \boldsymbol{x}_\circ$, $\Delta\boldsymbol{x}_2 \equiv \boldsymbol{x}_2 - \boldsymbol{x}_\circ$. Then

$$\Delta x_{1i} = \sum_j \left(\mathcal{A}_1^{-1}\right)_{ij} \sum_{n\in\Omega_1} \frac{f'_j}{f}(\mathcal{N}_n - f) \quad (14)$$

$$\Delta x_{2i} = \sum_k \left(\mathcal{A}_2^{-1}\right)_{ik} \sum_{m\in\Omega_2} \frac{f'_k}{f}(\mathcal{N}_m - f) \quad (15)$$

$$\text{where } \mathcal{A}_{1ij} = \sum_{n\in\Omega_1} \frac{f'_i\, f'_j}{f} \text{ and } \mathcal{A}_{2ik} = \sum_{m\in\Omega_2} \frac{f'_i\, f'_k}{f}. \quad (16)$$

The ensemble average of the difference $x_{1i} - x_{2i}$ vanishes in first approximation: $\langle x_{1i} - x_{2i} \rangle = \langle \Delta x_{1i} - \Delta x_{2i} \rangle \approx \langle \Delta x_{1i}^\circ \rangle - \langle \Delta x_{2i}^\circ \rangle = 0$. The mean square value of $x_{1i} - x_{2i}$ is:

$$\left\langle (x_{1i} - x_{2i})^2 \right\rangle = \sigma_{1i}^2 - 2\,\langle \Delta x_{1i}\, \Delta x_{2i} \rangle + \sigma_{2i}^2 \quad (17)$$

Since Ω_2 is a subset of Ω_1, the correlation term $\langle \Delta x_{1i}\, \Delta x_{2i} \rangle$ in eq. (17) is equal to σ_{1i}^2. That follows directly from eqs. (14) to (16) and the properties of basic fluctuations given in eqs. (4) and (5). Thus

$$\left\langle (x_{1i} - x_{2i})^2 \right\rangle = \sigma_{2i}^2 - \sigma_{1i}^2 \quad (18)$$

It is interesting to note that corresponding set-subset relation for the χ^2 values is

$$\left\langle \left(\chi_1^2 - \chi_2^2 - \left\langle \chi_1^2 - \chi_2^2 \right\rangle\right)^2 \right\rangle = 2N_{ch\,1} - 2N_{ch\,2} \quad (19)$$

which is the same as eq. (18) if one substitutes corresponding errors from eq. (9), although the real derivation of eq. (19)[1] is different and more complicated than that of eq. (18) for the fit parameters.

4. χ^2 FIT WITH INCORPORATION OF EXTERNAL LIMITED KNOWLEDGE OF FIT PARAMETERS

It might happen that some of the fit parameters are known with limited precision from other experiments, independently of our fit. If K linear combinations $F_k(\boldsymbol{x}) = \sum_{i=1} C_{ik} x_i$ $(k = 1, \ldots, K)$ are known to be

[1] We have such a derivation in [3]

$F_{*k} \pm \sigma_{F_k}$, it is reasonable to add $\sum_k \frac{(F(\boldsymbol{x})-F_{*k})^2}{\sigma^2_{F_k}}$ to the right side of eq. (1) and use the resulting expression for the χ^2 fit. In such a case eq. (2) becomes

$$\Delta x_i = \sum_j \left(\mathcal{A}_c^{-1}\right)_{ij} \left(\sum_n \frac{f'_j}{f} (\mathcal{N}_n - f) + \sum_k \frac{C_{jk}}{\sigma^2_{F_k}} \Delta F_{*k} \right) \quad (20)$$

where $(\mathcal{A}_c)_{ij} \equiv \mathcal{A}_{ij} + \sum_k \frac{C_{ik} C_{jk}}{\sigma^2_{F_k}}$ and $\Delta F_{*k} \equiv (F_k - F_{*k})$. The corresponding equation for the correlation matrix for such a case is

$$\langle \Delta x_i \; \Delta x_j \rangle = (\mathcal{A}_c)^{-1}_{ij} \quad (21)$$

5. SYSTEMATIC SHIFT OF FIT PARAMETERS DUE TO NEGLECTED BACKGROUND

Suppose we have some low level background $h(t)$ admixed to the data, which is otherwise unambiguously described by a multi-parameter function $f(\boldsymbol{x}; t)$. The background might be small enough to evade observation "by eye" or even to spoil χ^2 considerably. Nevertheless fitting the histogram with the function $f(\boldsymbol{x}; t)$ alone will give parameter values x_i, shifted with respect to the "true" values $x_{i\circ}$ by some $\delta x_i = x_i - x_{i\circ}$. These systematic shifts δx_i can be found from the χ^2 minimization requirement $\partial \chi^2 / \partial x_i = 0$:

$$\frac{\partial \chi^2}{\partial x_i} = 2 \sum_n \frac{f - \mathcal{N}_n}{f} \, f'_i \;\; \approx \;\; 2 \sum_{j=1}^{L} \delta x_j \sum_n \frac{f'_j \, f'_i}{f}$$

$$-2 \sum_n \frac{h}{f} \, f'_i \;\; = \;\; 2 \sum_{j=1}^{L} \mathcal{A}_{ij} \, \delta x_j - 2 \sum_n \frac{h}{f} \, f'_i = 0 \quad (22)$$

and hence

$$\delta x_i = \sum_j \left(\mathcal{A}^{-1}\right)_{ij} \times \sum_n \frac{h}{f} \, f'_j \quad (23)$$

with the same matrix $\mathcal{A}$ as in eq. (3). Here we use $f(\boldsymbol{x}, t_n) \approx f(\boldsymbol{x}_\circ, t_n) + \sum_j f'_j \, \delta x_j$ and $\mathcal{N}_n \approx f(\boldsymbol{x}_\circ, t_n) + h(t_n)$.

6. REPLACE SUMS BY INTEGRALS

For practical reasons, it is convenient to replace sums over the histogram's channels in the equations above by integrals in the following way:

$$\sum_n g(t_n) \;\; = \;\; \frac{1}{b} \sum_n g(t_n) \, \Delta t \;\; \approx \;\; \frac{1}{b} \int g(t) \, dt \quad (24)$$

where $b = \Delta t$ is the width of a histogram channel and g is an arbitrary combination of the fit function and its derivatives. In turn, b can be eliminated in favor of the total number of events, N_{tot}, via the equation

$$N_{tot} = \sum_n \mathcal{N}_n \approx \sum_n f(t_n) \approx \frac{1}{b} \int f(t) \, dt \quad (25)$$

Thus eq. (24) can be rewritten as

$$\sum_n g(t_n) \;\approx\; \frac{N_{tot}}{\int f(t) \, dt} \int g(t) \, dt \quad (26)$$

Replacing sums by integrals in many cases allows one to obtain *analytical* expressions for the important statistical quantities discussed in this paper and makes their analysis easy.

7. STATISTICAL EQUATIONS FOR THE MUON g-2 EXPERIMENT

For the muon g-2 experiment [1], the time distribution of decay electrons plays the central role. It may be approximated by the 5-parameter function $G(t) = N_\circ \, e^{-t/\tau} \, [1 + A \, \cos(\omega t + \phi)]$, where A, ω and ϕ are the amplitude, frequency and phase of $g-2$ oscillations, respectively; τ is the muon lifetime in the lab frame and $N_\circ$ is the normalization constant. The value of ω, obtained from the χ^2 fit, is used to evaluate the muon $g-2$ value.

Application of statistical equations, discussed in this paper, to the 5-parameter function fit $G(t)$ gives:

• statistical errors and correlations

$$\sigma_\omega \;\; = \;\; \frac{\sqrt{2}}{\tau \, A \, \sqrt{N_{tot}}} \quad (27)$$

$$\langle \Delta\omega \, \Delta\phi \rangle \;\; = \;\; -\frac{2}{\tau^2 \, A^2 \, N_{tot}} \; (t_s + \tau) \quad (28)$$

where t_s is the histogram start time. For an estimated $N_{tot} \approx 10^{10}$ total events in the whole experiment, $\sigma_\omega / \omega \sim 0.3 \cdot 10^{-6} = 0.3$ ppm. Correlations of ω with other parameters, except for the phase, vanish. In fact, $\langle \Delta\omega \, \Delta\phi \rangle$ also vanishes when $t_s = -\tau$, which may be achieved by appropriate choice of the time origin. This technical trick makes calculations easy.

• The correlation of frequency ω and phase ϕ allows one to use possible external knowledge of ϕ (at time t', with error σ_F) to improve the precision of ω. This would result in

$$\sigma_\omega \;\; = \;\; \sigma_{\omega\circ} \left(1 + \frac{\sigma_F^{-2}}{(\tau \sigma_{\omega\circ})^{-2} + \sigma_F^{-2}} \, \frac{(t' - t_s - \tau)^2}{\tau^2} \right)^{-1/2} \quad (29)$$

where $\sigma_{\omega\circ} = \frac{\sqrt{2}}{\tau A \sqrt{N_{tot}}}$ is the statistical error of ω from the χ^2 fit alone, see eq. (27).

• systematic shift due to neglected background $h(t)$:

$$\delta\omega \;\; = \;\; - \, \frac{2}{e \, N_\circ \, A \, \tau^3} \int \frac{t \, h(t) \, \sin(\omega t + \phi)}{1 + A \, \cos(\omega t + \phi)} \, dt \quad (30)$$

For $\delta\omega$ for a particular example of $h(t)$, see [4].

• bias of fit parameters (the leading second term in eq. (10) for $\Delta\omega$):

$$\frac{\Delta\omega}{\omega} \approx \frac{1}{2\omega}\left(\frac{\sqrt{2}}{\tau A\sqrt{N}}\right)^2 \frac{N_{ch}}{\Delta T} \times \int \frac{N_\circ e^{-t/\tau} A\, t\, \sin(\omega t+\phi)}{N_\circ e^{-t/\tau}[1+A\cos(\omega t+\phi)]}\, dt \approx \frac{N_{ch}}{\omega^2\,\tau^2\,A\,N_{tot}} \tag{31}$$

where ΔT is the total range of the histogram in time units. For $N_{tot} = 10^{10}$, $\Delta\omega/\omega$ is ~ 0.1 ppb, which is negligible compare to the statistical error ~ 0.3 ppm. However if for technical or other reasons one wants to split statistics into, say, 1000 parts, fit them separately and obtain the final result by averaging those 1000, the net result might be biased by 0.0001 ppm $\times\, 1000 =$ 0.1 ppm, which is not negligible.

• Set-subset relations: we use “standard” eqs. (18) and (19) for histogram start time change in the course of our systematic study, but we also have derived and use a more elaborate and specific equation for similar changes of the energy threshold (E_{thr}) of decay electrons,

$$\left\langle(\omega_1-\omega_2)^2\right\rangle = \sigma^2_{\omega 2} - \sigma^2_{\omega 1}\left(2\,\frac{A_1}{A_2}\,\cos(\phi_1-\phi_2)-1\right) \tag{32}$$

where $g-2$ amplitude A and phase ϕ are functions of E_{thr}. For the case of $A_1 = A_2$ and $\phi_1 = \phi_2$, eq. (32) coincides with eq. (18). For more details see [4].

References

[1] G.W. Bennett et al., Phys. Rev. Lett. **89**, 101804 (2002), and references therein.

[2] S.I. Redin, Internal *g*-2 note #418, 2002, unpublished.

[3] S.I. Redin, Internal *g*-2 note #410, 2002, unpublished.

[4] S.I. Redin et al., Proc. of the Conf. on Advanced Statistical Techniques in Particle Physics, p. 242-247, Durham, England, 2002.