Statistical bias of fit parameters and bias reduction techniques

S.I.Redin
Budker Institute, Novosibirsk, Russia

- How to estimate statistical bias of fit parameters
- Which method, likelihood function or χ^2, is better
- What's better: χ^2 with $\sigma = \sqrt{\mathcal{N}}$ or that with $\sigma = \sqrt{f}$
- Possible improvements of the χ^2 method
  (bias reduction techniques)

likelihood function and functions $\chi^2_{\sigma_n=\sqrt{\mathcal{N}_n}}$ and $\chi^2_{\sigma_n=\sqrt{f}}$

Probability of experimental distribution to match theoretical model, given by function $f(\boldsymbol{x},t)$, can be written as

$$\mathcal{P} = \prod_n \frac{[f(\boldsymbol{x},t_n)]^{\mathcal{N}_n}}{\mathcal{N}_n!}\ \exp\left[-f(\boldsymbol{x},t_n)\right] \approx \prod_n \frac{1}{\sqrt{2\pi\mathcal{N}_n}}\ \left(\frac{f(\boldsymbol{x},t_n)}{\mathcal{N}_n}\right)^{\mathcal{N}_n}\exp\left[\mathcal{N}_n - f(\boldsymbol{x},t_n)\right]$$

Here we use Poisson distribution of events (counts) in histogram channels and Stirling's formula for $\mathcal{N}_n!$ We use following definitions:

- n is n^{th} channel of a histogram
- $\mathcal{N}_n$ is number of events in the n^{th} channel
- $N = \sum_n \mathcal{N}_n$ is total number of events in the histogram
- $f(\boldsymbol{x},t) \equiv f$ is a fit function
- $\boldsymbol{x} \equiv \{x_i\}$ is a vector of fit paremeters
- $f(\boldsymbol{x},t_n)$ is a fit function value for the n^{th} channel of histogram

---

Introduce log-likelihood function $\mathcal{L} \equiv -\ln \mathcal{P}$:

$$\mathcal{L} = \sum_n \left(\frac{1}{2}\ln 2\pi + \frac{1}{2}\ln\mathcal{N}_n + \mathcal{N}_n \ln\frac{\mathcal{N}_n}{f} + f - \mathcal{N}_n\right) \Rightarrow \boxed{\sum_n \left(\mathcal{N}_n \ln\frac{\mathcal{N}_n}{f} + f - \mathcal{N}_n\right)}$$

Decompose $2\mathcal{L}$ into series on powers of small value $(f-\mathcal{N}_n)/\mathcal{N}_n$:

$$2\mathcal{L} = \sum_n \left(\frac{(f-\mathcal{N}_n)^2}{\mathcal{N}_n} - \frac{2}{3}\frac{(f-\mathcal{N}_n)^3}{\mathcal{N}_n^2} + \frac{1}{2}\frac{(f-\mathcal{N}_n)^4}{\mathcal{N}_n^3} - ...\right)$$

Introduce function $\chi^2 = \boxed{\sum_n \frac{\left(f(\boldsymbol{x};t_n) - \mathcal{N}_n\right)^2}{\sigma_n^2}}$, with 2 choices for σ_n :

$$\chi^2_{\sigma_n=\sqrt{\mathcal{N}_n}} = \sum_n \frac{(f-\mathcal{N}_n)^2}{\mathcal{N}_n}$$

$$\chi^2_{\sigma_n=\sqrt{f}} = \sum_n \frac{(f-\mathcal{N}_n)^2}{f} = \sum_n \frac{(f-\mathcal{N}_n)^2}{\mathcal{N}_n + (f-\mathcal{N}_n)} =$$

$$= \sum_n \left(\frac{(f-\mathcal{N}_n)^2}{\mathcal{N}_n} - \frac{(f-\mathcal{N}_n)^3}{\mathcal{N}_n^2} + \frac{(f-\mathcal{N}_n)^4}{\mathcal{N}_n^3} - ...\right)$$

Fluctuations of fit parameters versus fluctuations of $\mathcal{N}_n$

As a result of statistical fluctuations in individual histogram channels, optimization of $\mathcal{L} = \sum_n \left(\mathcal{N}_n \ln \frac{\mathcal{N}_n}{f} + f - \mathcal{N}_n \right)$ gives vector of "optimal" fit parameters $\boldsymbol{x}$, shifted with respect to the "true" value $\boldsymbol{x}_\circ$ by some $\Delta \boldsymbol{x}$: $\boldsymbol{x} = \boldsymbol{x}_\circ + \Delta \boldsymbol{x}$. Elements of vector $\Delta \boldsymbol{x}$ can be found from the $\mathcal{L}$ optimization requirement $\partial \mathcal{L} / \partial x_j = 0$:

$$
\begin{aligned}
0 &= \frac{\partial \mathcal{L}}{\partial x_j} = \sum_n f'_j \left(-\frac{\mathcal{N}_n}{f} + 1 \right) = \sum_n f'_j \frac{f - \mathcal{N}_n}{f} \approx \sum_n f'_j \frac{f_\circ + \left(\sum_i f'_i \Delta x_i \right) - \mathcal{N}_n}{f_\circ} = \\
&= \sum_i \left(\sum_n \frac{f'_i f'_j}{f_\circ} \right) \Delta x_i - \left(\sum_n \frac{f'_j}{f_\circ} (\mathcal{N}_n - f_\circ) \right) = 0
\end{aligned}
$$

and hence

$$
\boxed{\Delta x_i = \sum_j \left(\mathcal{A}^{-1} \right)_{ij} \sum_n \frac{f'_j}{f_\circ} (\mathcal{N}_n - f_\circ)}\,, \qquad \text{where} \qquad \boxed{\mathcal{A}_{ij} = \sum_n \frac{f'_i f'_j}{f_\circ}}
$$

Here we use:

- $f'_j \equiv \frac{\partial f}{\partial x_j}$ (notation)
- $f(\boldsymbol{x}) = f(\boldsymbol{x}_\circ + \Delta \boldsymbol{x}) \approx f_\circ + \sum_i f'_i \Delta x_i$, where $f_\circ \equiv f(\boldsymbol{x}_\circ, t)$
- $\langle \mathcal{N}_n - f_\circ \rangle = 0$
- $\langle (\mathcal{N}_n - f_\circ)^2 \rangle \equiv \sigma_n^2 = f_\circ$
- $\langle (\mathcal{N}_n - f_\circ)(\mathcal{N}_m - f_\circ) \rangle = 0$ for $n \neq m$

From that we have:

- $\langle \Delta x_i \rangle = 0$, since $\langle \mathcal{N}_n - f_\circ \rangle = 0$
- $\langle \Delta x_i (\mathcal{N}_m - f_\circ) \rangle = \sum_j \left(\mathcal{A}^{-1} \right)_{ij} f'_j$

By $\langle ... \rangle$ we denote average over ensemble of similar measurements (ensemble average)

Statistical bias of fit parameters for log-likelihood function fit

For $\mathcal{L}$ minimization in the next-to-leading approximation we have:

$$0 = \frac{\partial \mathcal{L}}{\partial x_j} = \sum_n f'_j \, \frac{f - \mathcal{N}_n}{f} = \sum_n \frac{f'_j}{f} \, (f - \mathcal{N}_n) = \sum_n \frac{f'_j + \sum_i f''_{ji} \Delta x_i + \ldots}{f_\circ + \sum_i f'_i \Delta x_i + \ldots} \times$$

$$\times \left(f_\circ - \mathcal{N}_n + \sum_i f'_i \, \Delta x_i + \frac{1}{2} \sum_{ik} f''_{ik} \, \Delta x_i \, \Delta x_k + \ldots \right)$$

We search for solution in form of successive approximations:

$\Delta x_i = \Delta x_i^\circ + \Delta x_i^1 + \ldots$. Then for Δx_i^1 we have equation:

$$\sum_n \frac{f'_i}{f_\circ} \left[\sum_p f'_p \, \Delta x_p^1 + \frac{1}{2} \sum_j \sum_k f''_{jk} \, \Delta x_j^\circ \, \Delta x_k^\circ + \right.$$

$$\left. + \left(- \frac{\sum_j f'_j \, \Delta x_j^\circ}{f_\circ} + \frac{\sum_j f''_{ij} \, \Delta x_j^\circ}{f'_i} \right) \left((f_\circ - \mathcal{N}_n) + \sum_k f'_k \, \Delta x_k^\circ \right) \right] = 0$$

which has solution: $\Delta x_p^1 = \sum_i \left(\mathcal{A}^{-1} \right)_{pi} \times$

$$\times \left[- \sum_j \sum_n \frac{f'_i \, f'_j}{f_\circ^2} (\mathcal{N}_n - f_\circ) \, \Delta x_j^\circ + \sum_{jk} \sum_n \frac{f'_i \, f'_j \, f'_k}{f_\circ^2} \Delta x_j^\circ \, \Delta x_k^\circ + \right.$$

$$+ \sum_j \sum_n \frac{f''_{ij}}{f_\circ} (\mathcal{N}_n - f_\circ) \, \Delta x_j^\circ - \sum_{jk} \sum_n \frac{f''_{ij} \, f'_k}{f_\circ} \Delta x_j^\circ \, \Delta x_k^\circ -$$

$$\left. - \frac{1}{2} \sum_{jk} \sum_n \frac{f'_i \, f''_{jk}}{f_\circ} \Delta x_j^\circ \, \Delta x_k^\circ \right]$$

$$\left\langle \Delta x_p^1 \right\rangle = - \frac{1}{2} \sum_{ijk} \left(\mathcal{A}^{-1} \right)_{pi} \left(\mathcal{A}^{-1} \right)_{jk} \sum_n \frac{f'_i \, f''_{jk}}{f_\circ} \neq 0$$

For a single parameter fit:

$\mathcal{L}$: $$\left\langle \Delta x^1 \right\rangle = - \frac{1}{2} \sigma^4 \sum_n \frac{f' \, f''}{f_\circ}$$

For $\chi^2_{\sigma_n = \sqrt{\mathcal{N}_n}}$: $$\left\langle \Delta x^1 \right\rangle = - \frac{1}{2} \sigma^4 \sum_n \frac{f' \, f''}{f_\circ} - \sigma^2 \sum_n \frac{f'}{f_\circ} + \sigma^4 \sum_n \frac{f'^{\,3}}{f_\circ^2}$$

For $\chi^2_{\sigma_n = \sqrt{f}}$: $$\left\langle \Delta x^1 \right\rangle = - \frac{1}{2} \sigma^4 \sum_n \frac{f' \, f''}{f_\circ} + \frac{1}{2} \sigma^2 \sum_n \frac{f'}{f_\circ} - \frac{1}{2} \sigma^4 \sum_n \frac{f'^{\,3}}{f_\circ^2}$$

Estimates of bias of fit parameters for the fit function $\quad G(t) = N_\circ e^{-t/\tau}(1 + A\cos(\omega t + \phi))$

For analytical evaluations and numerical estimates we replace sums by integrals: $\sum_n(\ldots) \approx \frac{1}{b}\int(\ldots)\,dt = \frac{N}{\int f\,dt}\int(\ldots)\,dt$
where b is bin width and $N = \sum_n \mathcal{N}_n \approx \sum_n f(\boldsymbol{x}, t_n) \approx \frac{1}{b}\int f\,dt$ is total number of events in the histogram.

$$
\begin{aligned}
\langle\Delta\omega\rangle_1 &\approx \frac{1}{2}\left(\frac{\sqrt{2}}{\tau A\sqrt{N}}\right)^4 \times \frac{N}{\int_{-\tau}^{\infty} N_\circ e^{-t/\tau}[1 + A\cos(\omega t + \phi)]\,dt} \times \\
&\times \int_{-\tau}^{t_{max}} \frac{N_\circ e^{-t/\tau} A\,t\,\sin(\omega t + \phi) \times N_\circ e^{-t/\tau} A\,t^2\,\cos(\omega t + \phi)}{N_\circ e^{-t/\tau}[1 + A\cos(\omega t + \phi)]}\,dt \approx \\
&\approx \frac{2}{\tau^4 A^2 N} \times \frac{1}{e\tau} \times \frac{1}{2\omega}\left[-e^{-t/\tau} t^3 \cos 2(\omega t + \phi)\right]\Big|_{-\tau}^{t_{max}} \sim \frac{1}{\omega\tau^2 A^2 N}
\end{aligned}
$$

$$
\begin{aligned}
\langle\Delta\omega\rangle_2 &\approx \frac{1}{2}\left(\frac{\sqrt{2}}{\tau A\sqrt{N}}\right)^2 \frac{1}{b}\int_{-\tau}^{t_{max}} \frac{N_\circ e^{-t/\tau} A\,t\,\sin(\omega t + \phi)}{N_\circ e^{-t/\tau}[1 + A\cos(\omega t + \phi)]}\,dt \approx \\
&\approx \frac{1}{\omega^2\tau^2 A N b}\left[-\omega t\,\cos(\omega t + \phi)\right]\Big|_{-\tau}^{t_{max}} \sim \frac{t_{max} + \tau}{\omega\tau^2 A N b} \quad = \quad \frac{N_{ch}}{\omega\tau^2 A N}
\end{aligned}
$$

$$
\begin{aligned}
\langle\Delta\omega\rangle_3 &\approx \frac{1}{2}\left(\frac{\sqrt{2}}{\tau A\sqrt{N}}\right)^4 \times \frac{N}{\int_{-\tau}^{\infty} N_\circ e^{-t/\tau}[1 + A\cos(\omega t + \phi)]\,dt} \times \\
&\times \int_{-\tau}^{t_{max}} \frac{\left(N_\circ e^{-t/\tau} A\,t\,\sin(\omega t + \phi)\right)^3}{\left(N_\circ e^{-t/\tau}[1 + A\cos(\omega t + \phi)]\right)^2}\,dt \approx \\
&\approx \frac{2}{\tau^4 A N} \times \frac{1}{e\tau} \times \frac{3}{4\omega}\left[-e^{-t/\tau} t^3 \cos(\omega t + \phi)\right]\Big|_{-\tau}^{t_{max}} \sim \frac{3}{2\omega\tau^2 A N}
\end{aligned}
$$

where $(t_{max}+\tau)/b = N_{ch} \sim 3600$ is the typical number of histogram channels

Possible improvement of χ^2 fit

It's straight forward to verify that some linear combination of results of $\chi^2_{\sigma_n=\sqrt{\mathcal{N}_n}}$ and $\chi^2_{\sigma_n=\sqrt{f}}$ fits, namely:

$$x_{comb} \;\equiv\; \frac{1}{3}\,(x)_{\sigma_n=\sqrt{\mathcal{N}_n}} \;+\; \frac{2}{3}\,(x)_{\sigma_n=\sqrt{f}}$$

has same bias as the log-likelihood function fit.

Note that similar combination of functions $\chi^2_{\sigma_n=\sqrt{\mathcal{N}_n}}$ and $\chi^2_{\sigma_n=\sqrt{f}}$ gives :

$$\frac{1}{3}\chi^2_{\sigma_n=\sqrt{\mathcal{N}_n}} + \frac{2}{3}\chi^2_{\sigma_n=\sqrt{f}} = \sum_n \left[\frac{(f-\mathcal{N}_n)^2}{\mathcal{N}_n} - \frac{2}{3}\,\frac{(f-\mathcal{N}_n)^3}{\mathcal{N}_n^2} + \frac{2}{3}\,\frac{(f-\mathcal{N}_n)^4}{\mathcal{N}_n^3} - \ldots\right]$$

which coincides with decomposition of $2\mathcal{L}$ in two lowest terms. That gives a hint for possible improvement of χ^2 fit : one should use a function which coincides with decomposition of $2\mathcal{L}$ in two (or more) lowest terms.
Examples:

$$\chi^2_{corr} \;=\; \sum_n \left[\frac{(f-\mathcal{N}_n)^2}{\mathcal{N}_n} - \frac{2}{3}\,\frac{(f-\mathcal{N}_n)^3}{\mathcal{N}_n^2}\right]$$

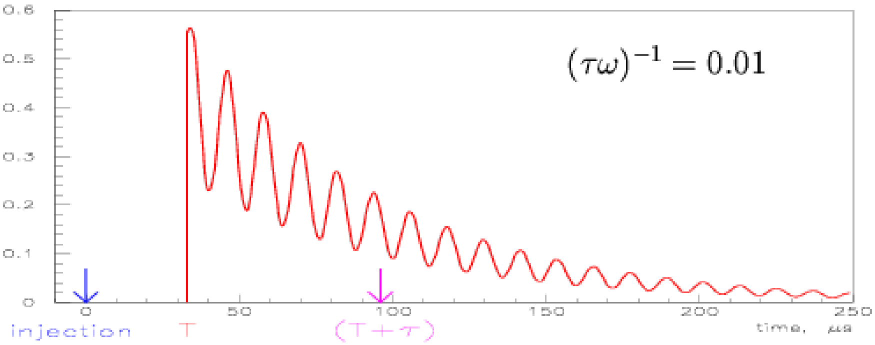
$$\chi^2_{\sigma_n^2=\mathcal{N}_n^{1/3}\times f^{2/3}} \;=\; \sum_n \left[\frac{(f-\mathcal{N}_n)^2}{\mathcal{N}_n} - \frac{2}{3}\,\frac{(f-\mathcal{N}_n)^3}{\mathcal{N}_n^2} + \frac{5}{6}\,\frac{(f-\mathcal{N}_n)^4}{\mathcal{N}_n^3} - \ldots\right]$$

Direct calculations show that indeed χ^2_{corr} and $\chi^2_{\sigma_n^2=\mathcal{N}_n^{1/3}\times f^{2/3}}$ fits have same bias as the log-likelihood function fit.

Bias of ω for muon g-2 experiment at BNL

Time distribution of high energy electrons ($E > E_{thr}$) from muon decays:

$$G(t) = N_\circ\, e^{-t/\tau}[1 + A\cos(\omega t + \phi)]$$



$$\sigma_{N_\circ} = \frac{N_\circ}{\sqrt{N}}\sqrt{(T/\tau+1)^2+1} \qquad \sigma_\tau = \frac{\tau}{\sqrt{N}} \qquad \sigma_A = \frac{\sqrt{2}}{\sqrt{N}} \qquad \boxed{\sigma_\omega = \frac{\sqrt{2}}{\tau A\sqrt{N}}} \qquad \sigma_\tau = \frac{\sqrt{2}}{A\sqrt{N}}\sqrt{(T/\tau+1)^2+1}$$

For $b = 0.15\,\mu$s, $A = 0.4$ and $N = 10^6$ decay electrons:

$$\frac{\sigma_\omega}{\omega} = \frac{\sqrt{2}}{\tau\omega A\sqrt{N}} = 38 \text{ ppm}$$

$$\left\langle\frac{\Delta\omega}{\omega}\right\rangle_2 = \frac{1}{\omega}\frac{\sigma_\omega^2}{2}\sum_n \frac{1}{G}\frac{\partial G}{\partial\omega} \approx \frac{1}{2\omega}\left(\frac{\sqrt{2}}{\tau A\sqrt{N}}\right)^2 \frac{1}{b}\int_{-\tau}^{T_{max}} \frac{N_\circ e^{-t/\tau} A\, t\, \sin(\omega t+\phi)}{N_\circ e^{-t/\tau}[1 + A\cos(\omega t+\phi)]}\, dt \approx 1 \text{ ppm}$$

$$\left\langle\frac{\Delta\omega}{\omega}\right\rangle_1 = \frac{1}{\omega}\frac{\sigma^4}{2}\sum_n \frac{1}{G}\frac{\partial G}{\partial\omega}\frac{\partial^2 G}{\partial\omega^2} \approx \frac{1}{2\omega}\left(\frac{\sqrt{2}}{\tau A\sqrt{N}}\right)^4 \times \frac{N}{\int_{-\tau}^{\infty} N_\circ e^{-t/\tau}[1 + A\cos(\omega t+\phi)]\, dt} \times$$
$$\times \int_{-\tau}^{\infty} \frac{N_\circ e^{-t/\tau} A\, t\, \sin(\omega t+\phi) \times N_\circ e^{-t/\tau} A\, t^2\, \cos(\omega t+\phi)}{N_\circ e^{-t/\tau}[1 + A\cos(\omega t+\phi)]}\, dt \approx 0.4 \text{ ppb}$$

Suppose we have $N = 10^9$ events and want to

- split them into 1000 parts, $N = 10^6$ in each;
- fit these 1000 parts separately;
- find ω as weighted average

then we'll have ~ 1 ppm bias, which is now comparable with statistical error 38 ppm $\times\sqrt{10^6/10^9} = 1.2$ ppm.