Pedagogical Notes on Error Estimation Using Maximum Likelihood Methods

With Applications to BaBar CP Analysis

R. Cahn

LBNL

August 2000

1 Introduction

Unbinned maximum likelihood is a convenient means of determining parameters of a fit to data. It also is a useful tool in establishing analytic estimations and approximations. We here provide elementary demonstrations of the well-known relations for the anticipated uncertainty in a fit for a parameter and for the anticipated bias. In addition, we address the question of systematic errors and provide a less well known result, the variance in the systematic error.

We suppose we are to determine a quantity A by fitting with a dataset $\{t_i\}$. The is, in addition a parameter p, which is determined externally, i.e. without using the dataset. For simplicity, we assume that there is a single distribution $f(A,p;t)$. It is normalized so that

$$\int dt f(A,p;t) = 1 \tag{1}$$

independent of A and p. Then as the number of data points approaches infinity, maximizing the log-likelihood function

$$\mathcal{L} = \sum_i \ln f(A,p;t_i) \tag{2}$$

will find values for A that converge to the true value, $\overline{A}$. We assume that the underlying data are described by

$$\overline{f} \equiv f(\overline{A},\overline{p},t) \tag{3}$$

For the finite dataset $\{t_i\}$ we expand the log-likelihood around the true value $\overline{A}$:

$$\begin{aligned}
\mathcal{L} &= \sum_i \ln f(A, p; t_i) \\
&= \sum_i [\ln f(\overline{A}, p; t_i) + \delta A \frac{\partial}{\partial A} \ln f(A, p; t_i) + \\
&\quad + \tfrac{1}{2}\delta A^2 \frac{\partial^2}{\partial A^2} \ln f(A, p; t_i) + + \tfrac{1}{6}\delta A^3 \frac{\partial^3}{\partial A^3} \ln f(A, p; t_i) + \ldots \\
&= \mathcal{E} + \delta A \mathcal{F} + \tfrac{1}{2}\delta A^2 \mathcal{G} + \tfrac{1}{6}\delta A^3 \mathcal{H} \ldots
\end{aligned} \tag{4}$$

where all partial derivatives are evaluated at $(\overline{A}, \overline{p})$ and where

$$\begin{aligned}
\mathcal{E} &= \sum_i \ln f(\overline{A}, p; t_i) \\
\mathcal{F} &= \sum_i \frac{\partial}{\partial A} \ln f(\overline{A}, p; t_i) = \sum_i F_i \\
\mathcal{G} &= \sum_i \frac{\partial^2}{\partial A^2} \ln f(\overline{A}, p; t_i) = \sum_i G_i \\
\mathcal{H} &= \sum_i \frac{\partial^3}{\partial A^3} \ln f(\overline{A}, p; t_i) = \sum_i H_i
\end{aligned} \tag{5}$$

Let us also write

$$\begin{aligned}
\mathcal{F} &= N\overline{F} + \sum_i (F_i - \overline{F}) = N\overline{F} + \sum_i \delta F \\
\mathcal{G} &= N\overline{G} + \sum_i (G_i - \overline{G}) = N\overline{G} + \sum_i \delta G \\
\mathcal{H} &= N\overline{H} + \sum_i (H_i - \overline{H}) = N\overline{H} + \sum_i \delta H
\end{aligned} \tag{6}$$

Note that

$$\overline{F} = \int dt \overline{f} F = \int dt \frac{\partial}{\partial A} f(A, p; t) = 0 \tag{7}$$

by virtue of the uniform normalization condition. It is good to keep in mind that $\mathcal{F}, \mathcal{G}$, and $\mathcal{H}$ are functions of the entirety of the dataset $\{t_i\}$, while $F, G,$

and H, (and δF, δG, and δH) are functions of a single argument t only. The quantities $\overline{F}, \overline{G}$, and $\overline{H}$ are not functions of t at all.

In many cases, there is not a single distribution f that describe the data but a collection, f_α. Each event falls into a category $\alpha, \beta, \ldots$. If the fraction of events in category α is ϵ_α, then we need to write expressions like

$$\mathcal{F} = N \sum_\alpha \epsilon_\alpha \overline{F}^\alpha + \sum_{i,\alpha} (F_i^\alpha - \overline{F}^\alpha) = N\overline{F} + \sum_{i,\alpha} \delta F^\alpha \tag{8}$$

2 Bias in A

Now δA is determined by maximizing $\mathcal{L}$:

$$\frac{\partial \mathcal{L}}{\partial A} = 0 \to \mathcal{F} + \delta A \mathcal{G} + \tfrac{1}{2} \delta A^2 \mathcal{H} = 0 \tag{9}$$

If we include the final term only as a perturbation, we have

$$\delta A = -\frac{\mathcal{F}}{\mathcal{G}} - \tfrac{1}{2} \frac{\mathcal{F}^2 \mathcal{H}}{\mathcal{G}^3} \tag{10}$$

Remember that δA, $\mathcal{F}$, $\mathcal{G}$, and $\mathcal{H}$ are functions of the dataset $\{t_i\}$. To determine the expectation value of δA we integrate over all the observed times, weighting with the underlying distribution, $\overline{f}$:

$$\langle \delta A \rangle = \prod_i \int dt_i \overline{f}(t_i) \delta A \tag{11}$$

Now

$$\begin{aligned} \left\langle \frac{\mathcal{F}}{\mathcal{G}} \right\rangle &= \left\langle \frac{\overline{F} + \frac{1}{N} \sum \delta F}{\overline{G} + \frac{1}{N} \sum \delta G} \right\rangle \\ &= \left\langle \frac{1}{\overline{G}} \frac{1}{N} \sum \delta F \left(1 - \frac{1}{N\overline{G}} \sum \delta G + \ldots \right) \right\rangle \\ &= -\frac{1}{\overline{G}^2} \frac{1}{N} \langle \delta F \delta G \rangle \end{aligned} \tag{12}$$

where

$$\langle \delta F \delta G \rangle = \int dt \overline{f} (F - \overline{F})(G - \overline{G}) \tag{13}$$

We have used the relations $\overline{F} = 0$, $\langle \delta F \rangle = 0$, $\langle \delta G \rangle = 0$ and

$$\begin{aligned} \prod_i \int dt_i \overline{f}(t_i) \sum_j \delta F(t_j) \sum_k \delta G(t_k) &= \prod_i \int dt_i \overline{f}(t_i) \sum_j \delta F(t_j) \delta G(t_j) \\ &= N \int dt \overline{f} \delta F(t) \delta G(t) \end{aligned} \tag{14}$$

.

Proceeding analogously, we find the leading terms in the expectation value of δA, i.e. the bias in A:

$$\langle \delta A \rangle = \frac{1}{N\overline{G}^2} \langle \delta F \delta G \rangle - \frac{1}{2N} \frac{\langle \delta F^2 \rangle}{\overline{G}^3} \overline{H} \tag{15}$$

This corresponds to Eq. 7.3.1 of Eadie et al., *Statistical Methods in Experimental Physics*. If there are several categories, we need to make the replacements

$$\begin{aligned} \overline{G} &\rightarrow \sum_\alpha \epsilon_\alpha \overline{G}^\alpha \\ \langle \delta F \delta G \rangle &\rightarrow \sum_\alpha \epsilon_\alpha \langle \delta F^\alpha \delta G^\alpha \rangle \end{aligned} \tag{16}$$

and so on.

3 Variance of A

Now lets calculate the variance in δA. We shall find that $\langle \delta A^2 \rangle = \mathcal{O}(1/N)$, so we can ignore $\langle \delta A \rangle^2$.

We compute directly

$$\left\langle \delta A^2 \right\rangle = \left\langle \frac{\mathcal{F}^2}{\mathcal{G}^2} + \frac{\mathcal{F}^3 \mathcal{H}}{\mathcal{G}^4} \right\rangle \tag{17}$$

The first term in the brackets on the right gives

$$\begin{aligned} \left\langle \frac{\mathcal{F}^2}{\mathcal{G}^2} \right\rangle &= \left\langle \frac{\sum \delta F \sum \delta F}{N\overline{G}^2 + \ldots} \right\rangle \\ &= \frac{1}{N\overline{G}^2} \left\langle \delta F^2 \right\rangle + \mathcal{O}(1/N^2) \end{aligned} \tag{18}$$

On the other hand,

$$\left\langle \frac{\mathcal{F}^3\mathcal{H}}{\mathcal{G}^4} \right\rangle \approx \left\langle \frac{\sum \delta F \sum \delta F \sum \delta F}{N^4 \overline{G}^4} N\overline{H} \right\rangle = \mathcal{O}(1/N^2) \tag{19}$$

so that our answer, to leading order is

$$\left\langle \delta A^2 \right\rangle = \frac{1}{N} \frac{\langle \delta F^2 \rangle}{\overline{G}^2} \tag{20}$$

This is, in fact, the familiar result. Note that

$$\begin{aligned} \left\langle \delta F^2 \right\rangle &= \int dt f \left(\frac{1}{f} \frac{\partial f}{\partial A} \right)^2 \\ \langle G \rangle &= \int dt f \left[-\frac{1}{f^2} \left(\frac{\partial f}{\partial A} \right)^2 + \frac{1}{f} \frac{\partial^2 f}{\partial A^2} \right] = -\left\langle \delta F^2 \right\rangle \end{aligned} \tag{21}$$

so

$$\left\langle \delta A^2 \right\rangle = \frac{1}{N \int dt \frac{1}{f} \left(\frac{\partial f}{\partial A} \right)^2}, \tag{22}$$

the familiar result.

4 Expected uncertainty from single experiment

From a single dataset we cannot only find an estimate of A, but also of its uncertainty. The usual rule is

$$\sigma^{-2} = -\frac{\partial^2 \mathcal{L}}{\partial A^2} = -(\mathcal{G} + \mathcal{H}\delta A + \ldots) \tag{23}$$

This is to be evaluated where $\partial \mathcal{L}/\partial A = 0$, so we find that the expectation value is

$$\left\langle \sigma^{-2} \right\rangle = -\left\langle \mathcal{G} - \frac{\mathcal{F}}{\mathcal{G}} \mathcal{H} + \ldots \right\rangle \tag{24}$$

The leading contribution is

$$\left\langle \sigma^{-2} \right\rangle = -N\overline{G} \tag{25}$$

so that indeed

$$\left\langle \sigma^2 \right\rangle = \left\langle \delta A^2 \right\rangle \tag{26}$$

to leading order in N. This is reassuring: the σ we derive from a single dataset is, on average, the σ we would get from a large ensemble.

5 Variance of expected uncertainty

To determine the variance of σ we use

$$\sigma^2 = \mathcal{G}^{-1}(1 + \frac{\mathcal{F}}{\mathcal{G}^2}\mathcal{H} + \ldots) \tag{27}$$

and calculate the expectation value:

$$\begin{aligned} \left\langle \sigma^2 \right\rangle &= \frac{1}{N(-\overline{G})} \left\langle (1 + \frac{1}{N\overline{G}} \sum \delta G)^{-1} \left(1 + \frac{\overline{H} \sum \delta F (1 + \frac{1}{N\overline{H}} \sum \delta H)}{\overline{G}^2 (1 + \frac{1}{N\overline{G}} \sum \delta G)^2} \right) \right\rangle \\ &= \frac{1}{N(-\overline{G})} \left[1 - \frac{3}{N\overline{G}^3} \left\langle \delta F \delta G \right\rangle \overline{H} + \frac{1}{N\overline{G}^2} \left\langle \delta G^2 \right\rangle + \frac{\left\langle \delta F \delta H \right\rangle}{N\overline{G}^2} \right] \end{aligned} \tag{28}$$

Simlarly, for σ we find

$$\langle \sigma \rangle = \frac{1}{N^{1/2}(-\overline{G})^{1/2}} \left(1 - \frac{5}{4} \frac{\left\langle \delta F \delta G \right\rangle \overline{H}}{\overline{G}^3} + \frac{3}{8N\overline{G}^2} \left\langle \delta \overline{G}^2 \right\rangle + \frac{1}{2N\overline{G}^2} \left\langle \delta F \delta H \right\rangle \right) \tag{29}$$

Thus

$$\left\langle \sigma^2 \right\rangle - \langle \sigma \rangle^2 = \frac{1}{4N^2(-\overline{G})^3} \left\langle \delta G^2 \right\rangle - \frac{1}{2N^2\overline{G}^4} \left\langle \delta F \delta G \right\rangle \overline{H} \tag{30}$$

6 Example from CP Asymmetry

Consider the CP asymmetry as measured in $B \to \psi K_S$. We use as distributions

$$f_{\pm} = \frac{1}{4}\Gamma e^{-\Gamma|t|}(1 \pm DA \sin \Delta m t) \tag{31}$$

where we ignore smearing due to Δz resolution. We then have

$$\begin{aligned} F_{\pm} &= \tfrac{\partial}{\partial A}\ln f_{\pm} &= \frac{\pm D\sin\Delta mt}{1\pm DA\sin\Delta mt} \\ G_{\pm} &= \tfrac{\partial^2}{\partial A^2}\ln f_{\pm} &= \frac{-D^2\sin^2\Delta mt}{(1\pm DA\sin\Delta mt)^2} \\ H_{\pm} &= \tfrac{\partial^3}{\partial A^3}\ln f_{\pm} &= \frac{\pm 2D^3\sin^3\Delta mt}{(1\pm DA\sin\Delta mt)^3} \end{aligned} \tag{32}$$

We have to combine the $\pm$ pieces appropriately. For example

$$\begin{aligned} \overline{F} &= \int_{-\infty}^{\infty} dt\frac{1}{4}\Gamma e^{-\Gamma|t|}\left\{+D\sin\Delta mt - D\sin\Delta mt\right\} = 0 \\ \overline{G} &= \int_{-\infty}^{\infty} dt\frac{1}{4}\Gamma e^{-\Gamma|t|}(-D^2\sin^2\Delta mt)\left\{\frac{1}{1+DA\sin\Delta mt}+\frac{1}{1-DA\sin\Delta mt}\right\} \\ &= \int_0^{\infty} dt\Gamma e^{-\Gamma t}\frac{-D^2\sin^2\Delta mt}{1-D^2A^2\sin^2\Delta mt} \\ \overline{H} &= -4\int_0^{\infty} dt\Gamma e^{-\Gamma t}\frac{D^4A\sin^4\Delta mt}{(1-D^2A^2\sin^2\Delta mt)^2} \end{aligned} \tag{33}$$

To lowest order in A,

$$\begin{aligned} \overline{F} &= 0 \\ \overline{G} &= \frac{2x^2}{1+4x^2}(-D^2) \\ \overline{H} &= \left[\frac{2x^2}{1+4x^2}-\frac{2x^2}{1+16x^2}\right](-4D^4A) \end{aligned} \tag{34}$$

We also find

$$\begin{aligned} \langle\delta F\delta G\rangle &= \int_0^{\infty} dt\Gamma e^{-\Gamma t}\frac{D^4A\sin^4\Delta mt}{(1-D^2A^2\sin^2\Delta mt)^2} \approx -\frac{1}{4}\overline{H} \\ \left\langle\delta G^2\right\rangle &\approx \int_0^{\infty} dt\Gamma e^{-\Gamma t}D^4(\sin^2\Delta mt - \frac{2x^2}{1+4x^2})^2 \\ &\approx D^4 2x^2\left[\frac{1+2x^2}{(1+4x^2)^2}-\frac{1}{1+16x^2}\right] \end{aligned} \tag{35}$$

Category	events	ϵ	D	ϵD^2	ϵD^4
ℓ	14	0.117	0.808	0.076	0.0499
K	63	0.525	0.606	0.193	0.0708
NT1	14	0.117	0.666	0.052	0.0230
NT2	29	0.242	0.338	0.028	0.0032
Total	120	1.00		0.349	0.147

Table 1: Data used in the Osaka presentation on $\sin 2\beta$. We normalized to the number of tagged events.

Combining these results we find the bias

$$\langle \delta A \rangle = \frac{1}{4}\frac{\overline{H}}{N\overline{G}^2} \approx -\frac{2}{N}A \tag{36}$$

This is truly a very small correction compared to the statistical uncertainty, which is of order $1/\sqrt{N}$.

7 Numerical Examples

We take $\tau = 1.548$ ps, $\Delta m = 0.472$ ps^{-1}, so $x = 0.73$. The data used in the Osaka presentation of $\sin 2\beta$ are shown in Table 1.

In our simple example we have ignored the Δz resolution. In this limit and with $A = 0.12$ we have the values

$$\begin{aligned} \overline{F} &= 0 \\ \overline{G} &= 0.341 \times (-0.349) = -0.119 \\ \overline{H} &= -0.016 \\ \left\langle \delta G^2 \right\rangle &= 0.113 \times 0.147 = 0.0166 \\ \langle \delta F \delta G \rangle &= 0.004 \end{aligned} \tag{37}$$

Using these numbers we compute

$$\langle \sigma \rangle = \frac{1}{\sqrt{-N\overline{G}}} = 0.264$$

$$\sqrt{\langle \sigma^2 \rangle - \langle \sigma \rangle^2} \quad = \quad 0.0134 \tag{38}$$

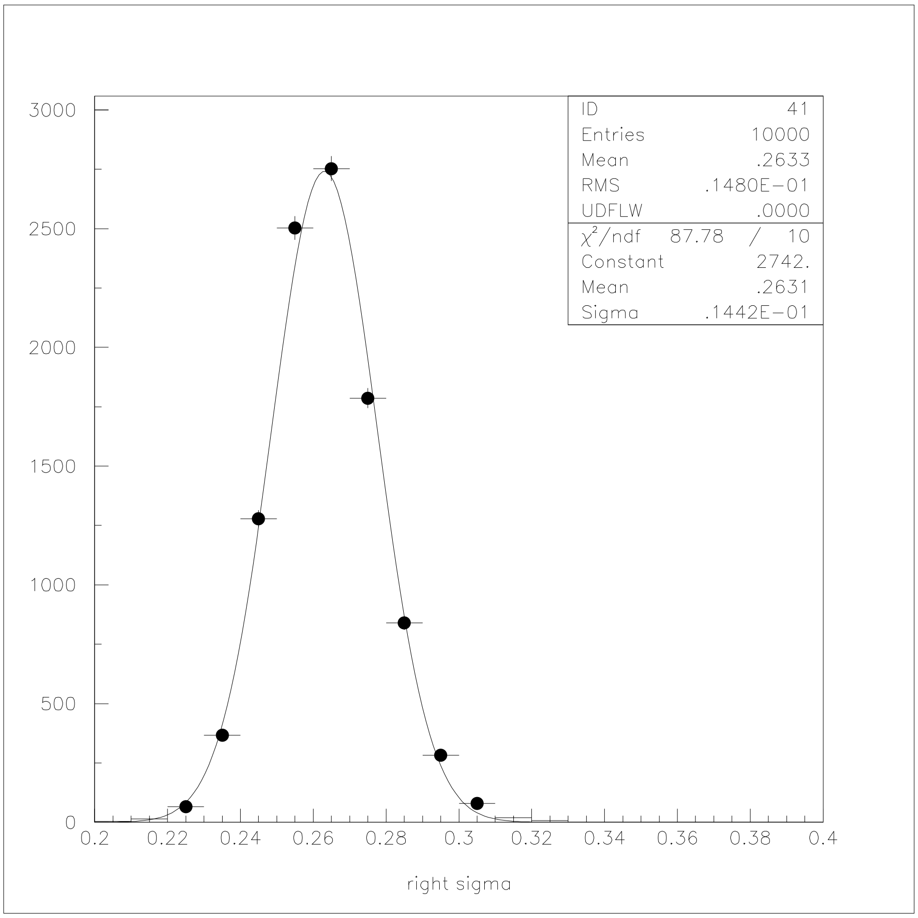

Figure 1: Monte Carlo results using the same distribution of 120 among the various tagging categories, but with times chosen randomly in a fashion consistent with A=0.12. The Δz resolution has been made essentially perfect.

8 Systematic Errors

We turn now to the question of systematic error. Suppose we use the wrong value $p = \overline{p} + \delta p$ for the input parameter. There will be a shift in the value of A deduced from the maximum likelihood fit:

$$\frac{\partial \delta A}{\partial p}\delta p = -\frac{\partial}{\partial p}\left[\frac{\mathcal{F}}{\mathcal{G}} + \tfrac{1}{2}\frac{\mathcal{F}^2\mathcal{H}}{\mathcal{G}^3}\right]\delta p \tag{39}$$

Note that $\langle \partial\mathcal{F}/\partial p\rangle \neq \partial\langle\mathcal{F}\rangle/\partial p$ and in particular $\langle \partial\mathcal{F}/\partial p\rangle \neq 0$. Instead, we find

$$\left\langle \frac{\partial \mathcal{F}}{\partial p}\right\rangle = -N\int dt \frac{1}{f}\frac{\partial f}{\partial p}\frac{\partial f}{\partial A} \tag{40}$$

Thus to leading order, the systematic shift in A is

$$\begin{aligned}\left\langle \frac{\partial \delta A}{\partial p}\right\rangle \delta p &= -\left\langle \frac{\partial}{\partial p}\frac{\mathcal{F}}{\mathcal{G}}\right\rangle \delta p \\ &= -\delta p \frac{\int dt \frac{1}{f}\frac{\partial f}{\partial p}\frac{\partial f}{\partial A}}{\int dt \frac{1}{f}\frac{\partial f}{\partial A}\frac{\partial f}{\partial A}}\end{aligned} \tag{41}$$

The term involving $\mathcal{H}$ can be shown to be of higher order because of the presence of two factors of $\mathcal{F}$.

While this gives us the asymptotic shift in δA, there are of course fluctuations around this value. To determine the variance of δA consider

$$\left\langle\left(\frac{\partial \delta A}{\partial p}\right)^2\right\rangle - \left\langle\left(\frac{\partial \delta A}{\partial p}\right)\right\rangle^2 = \left\langle\left[\frac{\partial}{\partial p}\frac{\mathcal{F}}{\mathcal{G}}\right]^2\right\rangle - \left\langle\left[\frac{\partial}{\partial p}\frac{\mathcal{F}}{\mathcal{G}}\right]\right\rangle^2 \tag{42}$$

The leading terms are obtained by the replacements $\mathcal{F} \to (1/N)\sum \delta F$, $\mathcal{G} \to \overline{G}$. In this way we find

$$\left\langle\left(\frac{\partial \delta A}{\partial p}\right)^2\right\rangle - \left\langle\left(\frac{\partial \delta A}{\partial p}\right)\right\rangle^2 = \frac{1}{N}\left\{\left\langle\left[\frac{\partial}{\partial p}\frac{F}{\overline{G}}\right]^2\right\rangle - \left\langle\left[\frac{\partial}{\partial p}\frac{F}{\overline{G}}\right]\right\rangle^2\right\} \tag{43}$$

Once again, the other term can be shown to be less important.

9 Example: Error in Bias

Following Roucelle, LeDiberder, and Roos [1], we consider the error induced by a mistake in the bias, i.e. in the shift in the Δz measurement. A shift from zero is expected to result from confusion of the secondary (charm) decay vertex with the primary (b) decay vertex. The question here is what systematic error arises if we have incorrectly asigned the bias to be used in analyzing the CP data, based on the measurement of the bias in the TagMix sample.

For simplicity we consider a case in which the true bias is zero, but where we mistakenly believe the bias is δt. We indicate the distributions for B^0 and $\overline{B}^0$ tags by

$$\begin{aligned} f_\pm &= \frac{1}{4}\Gamma e^{-\Gamma|t|}(1 \pm DA\sin[\Delta m(t+\delta t)]) \\ &\equiv \frac{1}{4}\Gamma e^{-\Gamma|t|}(1 \pm DAs) \end{aligned} \tag{44}$$

In terms of these,

$$\begin{aligned} F_\pm &= \tfrac{\partial}{\partial A} f_\pm &= \frac{\pm Ds}{1 \pm DAs} \\ G_\pm &= \tfrac{\partial^2}{\partial A^2} f_\pm &= \frac{-D^2 s^2}{(1 \pm DAs)^2} \end{aligned} \tag{45}$$

To find the shift in A we need to calculate

$$\delta A = -\left\langle \frac{\partial}{\partial \delta t}\frac{\mathcal{F}}{\mathcal{G}} \right\rangle \tag{46}$$

Both $\mathcal{F}$ and $\mathcal{G}$ are nominally of order N, since each is a sum of N terms. However $\langle(\rangle F_+ + F_-) = 0$. The result is the the leading contribution in N should come from the term in which the derivative acts on $\mathcal{F}$. Explicitly,

$$\begin{aligned} \left\langle \frac{\partial A}{\partial \delta t} \right\rangle &= -\left\langle \frac{1}{\mathcal{G}}\frac{\partial \sum F_\pm}{\partial \delta t} - \sum \frac{F_\pm}{\mathcal{G}^2}\sum\frac{\partial G_\pm}{\partial \delta t} \right\rangle \\ &= -\left\langle \frac{1}{\mathcal{G}}\frac{\partial \sum F_\pm}{\partial \delta t} \right\rangle + \mathcal{O}(1/N) \\ &= -\left\langle \frac{1}{\overline{G}}\frac{\partial \sum F}{\partial \delta t} \right\rangle \end{aligned} \tag{47}$$

Here we have used a cryptic notation: $\overline{G}$ means $\overline{G}_+ + \overline{G}_-$ and similarly for F.

Now we compute

$$\begin{aligned}
\frac{\partial F_\pm}{\partial \delta t} &= \frac{\pm D\Delta m c}{(1\pm DAs)^2} \\
\left\langle \frac{\partial F_\pm}{\partial \delta t} \right\rangle &= \int_{-\infty}^{\infty} dt \frac{1}{4}\Gamma e^{-\Gamma|t|} \frac{\pm D\Delta m c}{(1\pm DAs)^2} \\
&= \pm\tfrac{1}{2} D\Delta m \frac{1}{1+x^2} + \mathcal{O}(DA)
\end{aligned} \tag{48}$$

We see that with equal contributions from the B^0 and $\overline{B}^0$, the total contribution will vanish (and not just to order DA). The consequence will be that the leading term in the systematic error will be of order δt^2. However, with a finite data set, we expect fluctuations in the number of B^0 and $\overline{B}^0$.

If there are unequal numbers of B^0 and $\overline{B}^0$ events we can use in place of the expectation of likelihood function

$$\begin{aligned}
F &\rightarrow N_{B^0} F_+ + N_{\overline{B}^0} F_- \\
\left\langle \frac{\partial \sum F}{\partial \delta t} \right\rangle &\rightarrow N_{B^0} \left\langle \frac{\partial F_+}{\partial \delta t} \right\rangle + N_{\overline{B}^0} \left\langle \frac{\partial F_-}{\partial \delta t} \right\rangle
\end{aligned} \tag{49}$$

and similarly for G. In this way we find

$$\begin{aligned}
\left\langle \frac{\partial F}{\partial \delta t} \right\rangle &= \int_{-\infty}^{\infty} dt \frac{1}{4}\Gamma e^{-\Gamma|t|} (N_{B^0} - N_{\overline{B}^0}) \Delta m D c \\
&= \frac{(N_{B^0} - N_{\overline{B}^0})\Delta m D}{2(1+x^2)} \\
\langle G \rangle &= -D^2 \frac{x^2 (N_{B^0} + N_{\overline{B}^0})}{1+4x^2}
\end{aligned} \tag{50}$$

and

$$\delta A = -\frac{1+4x^2}{2x(1+x^2)} \frac{\Gamma \delta t}{D} \frac{(N_{B^0} - N_{\overline{B}^0})}{(N_{B^0} + N_{\overline{B}^0})} \tag{51}$$

Such an effect might not be small if the error in the bias is large. The Osaka $\sin 2\beta$ paper cited an uncertainty in the bias of the wider Gaussian of 0.5 ps so that $\Gamma \delta t \approx 1/3$. With $N_{B^0} = 70, N_{\overline{B}^0} = 50$, and taking 1/4 of the effect since the wider Gaussian accounts for 25% of the resolution function, we find a shift of 0.033, using $D = 0.6$.

There is an additional contribution of order δt^2, which is present even with equal numbers of B^0 and $\overline{B}^0$. We need to expand

$$\delta A = \frac{\partial \delta A}{\partial \delta t}\delta t + \tfrac{1}{2}\frac{\partial^2 \delta A}{\partial \delta t^2}\delta t^2 + \dots \tag{52}$$

Assuming $N_{B^0} = N_{\overline{B}^0}$, the first term on the right vanishes. The leading term in N is

$$\frac{\partial^2 \delta A}{\partial \delta t^2} = -\frac{1}{\overline{\overline{G}}}\left\langle \frac{\partial^2 F}{\partial \delta t^2}\right\rangle \tag{53}$$

Now

$$\frac{\partial^2 F_\pm}{\partial \delta t^2} = [\mp s - DAs^2 - 2DAc^2]\frac{D\Delta m^2}{(1 \pm DAs)^3} \tag{54}$$

Expanding in DA we find

$$\begin{aligned}
\left\langle \frac{\partial^2 F_\pm}{\partial \delta t^2}\right\rangle &= \int_{-\infty}^{\infty} dt \frac{\Gamma}{4} e^{-\Gamma|t|}(1 \pm DAs)D\Delta m^2(1 \mp 3DAs)(\mp s - DAs^2 - 2DAc^2) \\
\left\langle \frac{\partial^2 F_\pm}{\partial \delta t^2}\right\rangle &= \int_{-\infty}^{\infty} dt \frac{\Gamma}{4} e^{-\Gamma|t|}(DAs^2 - 2DAc^2)D\Delta m^2 \\
\left\langle \frac{\partial^2 F}{\partial \delta t^2}\right\rangle &= -2D^2 A\Delta m^2 \frac{1+x^2}{1+4x^2}
\end{aligned} \tag{55}$$

Using

$$\langle G\rangle = -D^2\frac{2x^2}{1+4x^2} \tag{56}$$

we find

$$\tfrac{1}{2}\frac{\partial^2 \delta A}{\partial \delta t^2}\delta t^2 = -\tfrac{1}{2}(1+x^2)(\Gamma\delta t)^2 A \tag{57}$$

in agreement with the result of Ref. [1].

Next, let us consider the point emphasized by LeDiberder: systematic errors have not only an asymptotic component, but another component that falls with increasing N. As we demonstrated above

$$\left\langle \left(\frac{\partial \delta A}{\partial p}\right)^2\right\rangle - \left\langle \left(\frac{\partial \delta A}{\partial p}\right)\right\rangle^2 = \frac{1}{N}\left\{\left\langle \left[\frac{\partial}{\partial p}\frac{F}{\overline{\overline{G}}}\right]^2\right\rangle - \left\langle \left[\frac{\partial}{\partial p}\frac{F}{\overline{\overline{G}}}\right]\right\rangle^2\right\} \tag{58}$$

Now we have seen that if $N_{B^0} = N_{\overline{B}^0}$

$$\begin{aligned} \langle F \rangle &= 0 \\ \left\langle \frac{\partial F}{\partial \delta t} \right\rangle &= 0 \end{aligned} \tag{59}$$

In fact

$$\left\langle \frac{\partial G}{\partial \delta t} \right\rangle = 0 \tag{60}$$

too. Now we are interested in

$$\left(\frac{\partial}{\partial \delta t} \frac{\sum F}{\sum G} \right)^2 = \frac{(\frac{\partial}{\partial \delta t} \sum F)^2}{(\sum G)^2} - 2 \frac{(\frac{\partial}{\partial \delta t} \sum F) \sum F (\frac{\partial}{\partial \delta t} \sum G)}{(\sum G)^3} + \frac{\sum F \sum F (\sum \frac{\partial}{\partial \delta t} G)(\sum \frac{\partial}{\partial \delta t} G)}{(\sum G)^4} \tag{61}$$

The vanishing of $\sum F$, $\frac{\partial}{\partial \delta t} \sum F$, and $\frac{\partial}{\partial \delta t} \sum G$ means that the leading term in N is simply

$$\left\langle \left(\frac{\partial}{\partial \delta t} \frac{\sum F}{\sum G} \right)^2 \right\rangle \to \frac{\left\langle (\frac{\partial F}{\partial \delta t})^2 \right\rangle}{\overline{G}^2} \tag{62}$$

Now

$$\begin{aligned} \left\langle \sum (\frac{\partial}{\partial \delta t} F)^2 \right\rangle &\to \left\langle \sum [(\frac{\partial}{\partial \delta t} F_+)^2 + (\frac{\partial}{\partial \delta t} F_-)^2] \right\rangle \\ \left\langle (\sum G)^2 \right\rangle &\to \left\langle (\sum [G_+ + G_-])^2 \right\rangle \end{aligned} \tag{63}$$

while

$$\begin{aligned} \left\langle \left(\frac{\partial}{\partial \delta t} F_\pm \right)^2 \right\rangle &\to \left\langle D^2 \Delta m^2 c^2 \right\rangle = D^2 \Delta m^2 \frac{1 + 2x^2}{1 + 4x^2} \\ \langle G_\pm \rangle &\to \left\langle -D^2 s^2 \right\rangle = \frac{2x^2 D^2}{1 + 4x^2} \end{aligned} \tag{64}$$

and we have worked to lowest order in A.

In this artificial limit in which there are equal numbers of B^0 and $\overline{B}^0$,

$$\left\langle \left(\frac{\partial \delta A}{\partial p} \right)^2 \right\rangle - \left\langle \left(\frac{\partial \delta A}{\partial p} \right) \right\rangle^2 = \frac{1}{4N} \frac{\Delta m^2}{D^2 x^4} (1 + 2x^2)(1 + 4x^2) \tag{65}$$

The rms value of δ is thus

$$\langle \delta A \rangle^{1/2} \approx \frac{1}{2\sqrt{N}} \frac{1}{Dx} \sqrt{(1 + x^2)(1 + 4x^2)} \Gamma \delta t \tag{66}$$

References

[1] C. Roucelle, F. LeDiberder, and L. Roos, *Uncertainty on* $\sin 2\beta$ *and the wrong tag fractioin due to a systematical shift on* Δz.