ASYMMETRIC B-FACTORY NOTE #24

AUTHOR: Frank Porter
DATE: January 25, 1990
TITLE: Dilution Factor in CP Violation Measurement

THE PHYSICS PROGRAM OF A HIGH-LUMINOSITY ASYMMETRIC B-FACTORY AT SLAC
INTEROFFICE MEMORANDUM
CALIFORNIA INSTITUTE OF TECHNOLOGY
January 25, 1990

FROM: Frank Porter        EXTENSION: 6777        MAIL CODE: 356-48

SUBJECT: Dilution Factor in CP Violation Measurement

I have observed some confusion on the nature of the "dilution factor" used in considering the power of various methods of measuring CP violation via mixing in the B meson system. In this note, I attempt to shed further light on this matter with a simple analytic statistical analysis. Further illumination is possible via Monte Carlo simulations (for example, to determine the actual form of the maximum likelihood p.d.f for small sample sizes or for varying resolutions). While this would be easy to do, I hope that this analytic treatment is better suited to the present purpose, that of making a basic statistical point.

We are concerned here with the measurement of CP violation via mixing in the B\bar{B} system, where one B is used to tag the state of the other B at some time. Let \( z = \Delta m / \Gamma \) and \( t = (t_2 - t_1) / \tau \) be the time difference between the B decays, in units of the lifetime. Let \( A \) be the CP-violation asymmetry parameter we are interested in determining (i.e., \( A = \sin 2\phi \), where \( \phi = \alpha, \beta, \) or \( \gamma \), the angles of the unitarity triangle). For the purpose of this discussion, we assume that our detector measures the time \( t \) with perfect resolution, and that \( z \) is well-known. Thus, we draw our measurements of \( t \) from the p.d.f.:

\[
p(t; A) = \frac{1}{2} e^{-|t|} (1 + A \sin z t),
\]

where \( t \in (-\infty, \infty) \), and the parameter \( A \) may lie anywhere between -1 and 1.

The quantity we wish to learn about is \( A \), so we wish to form a reasonable estimate for \( A \), given a set of measurements \( \{t_1, t_2, \ldots, t_N\} \). A very simple estimator is obtained by counting the number of measurements where \( t > 0 \), \( n_+ \), and the number of times \( t < 0 \), \( n_- \), and forming the quantity:

\[
\hat{A}_\pm = d^{-1} n_+ - n_-.
\]

This estimator is equivalent to assuming a peculiar resolution function, in which we can tell the order of the decays, but cannot determine how far apart they are. The quantity "\( d \)" which we sometimes refer to as the "dilution factor", is given by:

\[
d = \left( \frac{n_+ - n_-}{N} \right) / A = \frac{x}{1 + x^2}.
\]

Thus, by definition, \( \hat{A}_\pm \) is an unbiased estimator for \( A \). The standard deviation of this

\[\text{**A nice reference for the statistics encountered here is the set of lecture notes by G. P. Yost, } LBL-16993 \text{ Rev. II, UCPPG/85/06, 84/HENP/3, June 1985.} \]
The estimator, treating it as a binomial process, is:

\[ \delta \tilde{A}_{\pm} = d^{-1} \sqrt{\frac{(1 - d^2 A^2)}{N}}. \tag{4} \]

An alternative estimator is the maximum likelihood estimator. This is the value, \( \tilde{A}_{ml} \), of \( A \) which maximizes the likelihood function:

\[ \mathcal{L} \left( \{t_i, i = 1, \ldots, N\}; A \right) = \prod_{i=1}^{N} e^{-|t_i|} (1 + A \sin xt_i). \tag{5} \]

We would like to learn the standard deviation of this estimator, and to compare it with our earlier example. We will do this by calculating the Rao-Cramér-Fréchet lower bound, and then argue that, at least asymptotically, the maximum likelihood estimator achieves this bound.

For an unbiased estimator, the variance is bounded from below according to:

\[ \delta^2 \hat{A} \geq \frac{1}{\left\langle \left( \frac{\partial}{\partial A} \sum_{i=1}^{N} \log p(t_i; A) \right)^2 \right\rangle}. \tag{6} \]

Since the events are independent, the expectation value in this expression may be written in the form \( NI_1(A) \), where

\[ I_1(A) = \left\langle \left( \frac{\partial}{\partial A} \log p(t; A) \right)^2 \right\rangle = \left\langle \left( \frac{\sin xt}{1 + A \sin xt} \right)^2 \right\rangle. \tag{7} \]

Performing the required integration\(^\dagger\) yields, assuming our estimator achieves this bound,

\[ \delta \tilde{A}_{ml} = \frac{1}{\sqrt{N}} \left\{ \sum_{k=1}^{\infty} A^{2(k-1)} \frac{x^{2k}(2k)!}{\prod_{r=1}^{k} [1 + (2r)^2] [1 + (4r)^2] \cdots [1 + (2k)^2]} \right\}^{-1/2}. \tag{8} \]

While the maximum likelihood estimator may not be efficient for finite datasets (since an unbiased efficient estimator may not exist), we know that under fairly mild conditions, which are met in this case (unless \( A = \pm 1 \)), the maximum likelihood estimator is asymptotically unbiased and efficient, so Eqn. (8) may be used for the standard deviation of \( \tilde{A}_{ml} \) in the large \( N \) limit. For the purpose of the present discussion, this is adequate.

\(^\dagger\) Making use of Gradshteyn and Ryzhik, Eqn. 3.895.1.
In Figure 1, I show the standard deviation of the maximum likelihood estimator according to Eqn. (8) (with the root $N$ factored out), as a function of the true value of the asymmetry parameter for different values of $x$. The difference between $x = 3$ and $x = 10$ is very small.

This note was motivated largely by confusion over the difference in power between the time-integrated estimator and the non-integrated (maximum likelihood) estimator. Therefore, in Figure 2, I show the ratio of the variance for the integrated method to that for the maximum likelihood method, again as a function of the asymmetry parameter and for different values of $x$. The variance ratio is chosen here because the required luminosity for the observation of $CP$ violation in a channel such as $J/\psi K^0_S$ scales like one over the variance. We note that the $x = 10$ curve is way off scale on this graph. This is readily understood, since for large $x$ there are many oscillations in a lifetime, and $n_+ \approx n_-$. In comparison, the maximum likelihood method makes use of the time-dependent (phase) information which is available. In other words, it is well-known that estimators using binned data are typically not sufficient, and that the maximum likelihood estimator is sufficient if a sufficient estimator exists.

Finally, I remark again that these calculations have been made in the limit of a detector with perfect time resolution. In either method, the variance will increase as the resolution worsens; the integrated method because $t > 0$ and $t < 0$ events get mixed up, and the maximum likelihood method because the time-dependence gets smeared (including the above effect).