Feldman-Cousins Confidence Intervals in Electron EDM Experiments

Elizabeth Petrik, Harvard University
DRAFT 2: ACME Gen. I Long Result Paper

August 13, 2014

A classical confidence interval is a natural choice for reporting the result of an EDM measurement. The confidence interval's inclusion or exclusion of the value $d_e = 0$ suggests whether the result is consistent or inconsistent, respectively, with the Standard Model. The confidence level represents an objective measure of the probability that the interval includes the unknown true value of the EDM, d_e^{true}. Since no statistically significant EDM has yet been observed, the recent custom has been for electron EDM experiments to report an upper limit at the 90% confidence level [5, 7]. The proper interpretation of such limits is that if the experiment were performed a large number of times, and the confidence interval were computed in the *same way* for each experimental trial, d_e^{true} would fall within the interval 90% of the time.

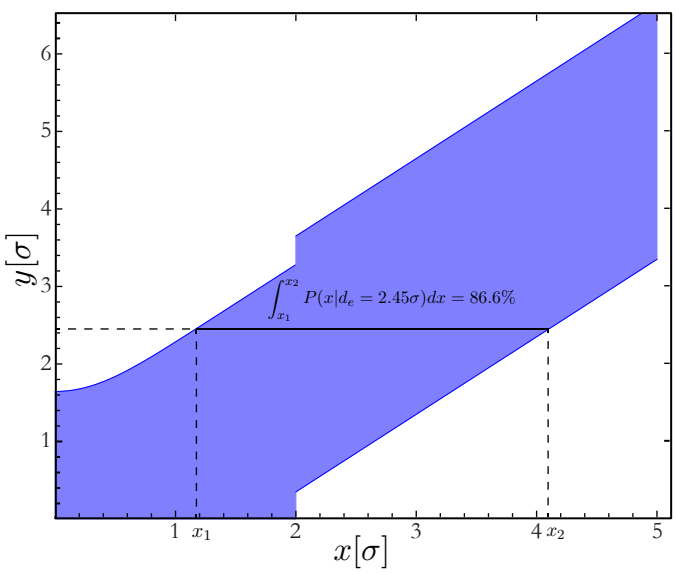

Figure 1: Flip-flopping "90% confidence" band described in the text. These intervals are not true 90% confidence intervals because if the true value of the EDM falls between about 1.6σ and 3.7σ, fewer than 90% of experimental trials will contain the true value. An example of this undercoverage is shown in the plot for the case $d_e^{\text{true}} = 2.45\sigma$.

Feldman and Cousins have pointed out that in order for this interpretation to be valid, the confidence interval construction must be independent of the result of the measurement [4]. If the procedure for constructing 90% confidence intervals is chosen contingent upon the measurement outcome, the resulting intervals may "undercover," i.e., fail to include the true value more than 10% of the time. For example, suppose we were to report a two-sided central confidence interval if our measured EDM, d_e^{meas}, fell two or more standard deviations from zero and an upper bound based on a folded Gaussian distribution otherwise. The resulting confidence band would be that shown in Fig. 1. The confidence interval for a particular measurement outcome is given by the vertical range of the shaded region at the position of the measured value $x = d_e^{\text{meas}}$. If

the true value of the EDM were in the range $1.6\sigma \leq d_e^{\text{true}} \leq 3.7\sigma$, our claimed "90% confidence intervals" would include the true value in fewer than 90% of our experimental trials. Feldman and Cousins term this inconsistent approach "flip-flopping."

We would like to indicate whether we have found a statistically significant EDM by constructing confidence intervals that include $d_e = 0$ for small values of d_e^{meas} (relative to the measurement uncertainty) and exclude $d_e = 0$ for large d_e^{meas}. In order to avoid flip-flopping, we must choose a confidence interval construction that consistently unifies these two limits. We use the Feldman-Cousins method described in Ref. [4], which is popular in low-signal, high-background measurements in astrophysics and particle physics (e.g., Ref. [1–3, 6]). We apply this method to a model with Gaussian statistics, in which the magnitude of the measured EDM $d_e^{\text{meas}} = |\vec{d}_e^{\text{meas}}|$ is sampled from a folded Gaussian distribution $P(d_e^{\text{meas}}|d_e^{\text{true}})$ centered on the true value of the EDM, d_e^{true}, with a standard deviation given by the quadrature sum of the statistical and systematic uncertainties, $\sigma = \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}$:

$$P(d_e^{\text{meas}}|d_e^{\text{true}}) = \frac{1}{\sigma\sqrt{2\pi}} \left(\exp\left[-\frac{(d_e^{\text{meas}} - d_e^{\text{true}})^2}{2\sigma^2} \right] + \exp\left[-\frac{(d_e^{\text{meas}} + d_e^{\text{true}})^2}{2\sigma^2} \right] \right). \tag{1}$$

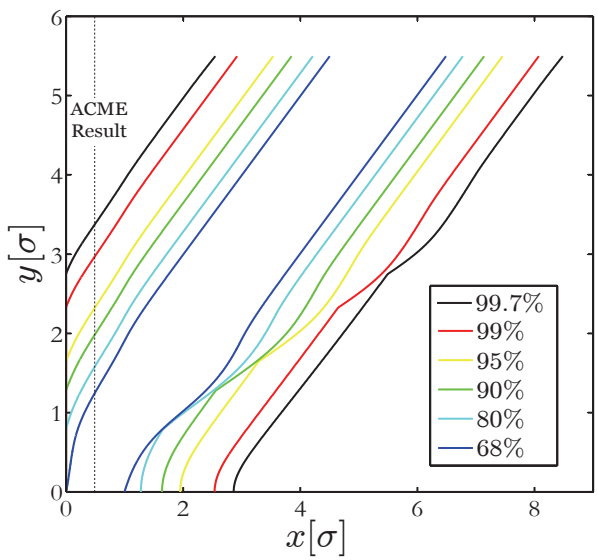

Figure 2: Feldman-Cousins confidence bands constructed as described in the text for a variety of confidence levels. Each pair of lines indicates the upper and lower bounds of the confidence band associated with each confidence level. To the left of the x-intercepts, the lower bounds are zero. Confidence bands are plotted as a function of the possible measured central values d_e^{meas} in units of standard deviation σ, and our result is plotted as a vertical dotted line. The y-value of the point at which our result line intersects with each of the colored lines gives the upper limit of our measurement at different confidence levels.

The general strategy of the Feldman-Cousins approach is to include in the confidence bands the measurement outcomes that maximize the likelihood of each possible true EDM value. Our procedure is a numerical calculation performed as follows:

1. Construct the confidence bands on a Cartesian plane, of which the x-axis represents the possible values of d_e^{meas}, and the y-axis represents the possible values of d_e^{true}. Divide the plane into a fine grid with x-intervals of width Δ_x and y-intervals of height Δ_y. We will consider only the discrete possible values $x_i = i\Delta_x$ and $y_j = j\Delta_y$, where the index $i(j)$ runs from 0 to $n_x(n_y)$.

2. For all values of i, maximize $P(x_i|y_j)$ with respect to y_j. Label the maximum points $y^{\text{max},i}$.

3. For some value of j, say $j = 0$ compute the likelihood ratio $R(x_i) = P(x_i|y_j)/P(x_i|y^{\text{max},i})$ for every value of i.

4. We will now construct the "horizontal acceptance band" at y_j. Include values of x_i in the horizontal acceptance band in descending order of $R(x_i)$. Stop adding values when the cumulative probability reaches the desired confidence level of 90%, i.e., $\sum_{x_i} P(x_i|y_j)\Delta_x = 0.9$.

5. Repeat steps (3)–(4) for all values of j.

6. To determine the reported confidence interval, draw a vertical line on your plot at $x = d_e^{\text{meas}}$. The 90% confidence interval is the region where the line intersects the constructed confidence band.

The plot in Fig. 2 was generated using the prescription above at a number of different confidence levels. Note that the 90% confidence intervals switch from upper bounds to two-sided confidence intervals when the value of d_e^{meas} becomes larger than 1.64σ, the same level of statistical significance required to reject the value $d_e = 0$ with 90% confidence in a Gaussian distribution.

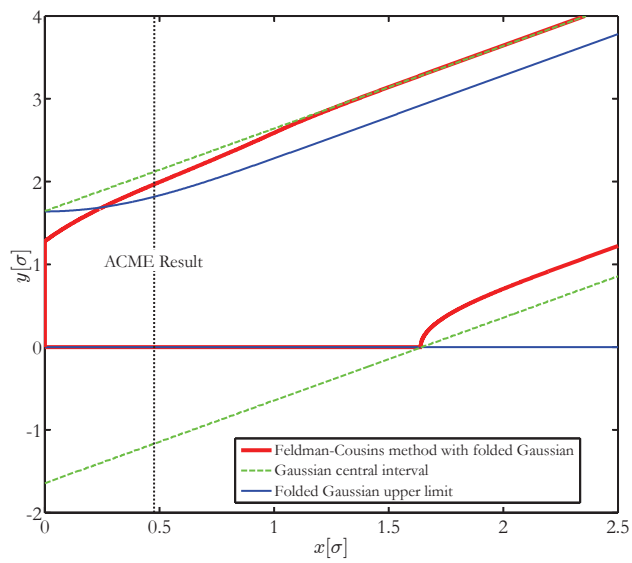

Figure 3: Comparison between 90% confidence intervals computed using three different methods. Confidence bands are plotted as a function of the possible measured central values d_e^{meas} in units of standard deviation σ, and our result is plotted as a vertical dotted line. The y-values of the points at which our result line intersects with the upper and lower line for each method give the upper and lower bounds of three possible 90% confidence intervals for our measurement. The valid confidence interval for our result is the Feldman-Cousins interval, which we chose before unblinding.

In our experiment, we found $d_e^{\text{meas}} = 0.458\sigma$ with $\sigma = 4.48 \times 10^{-29}\ e \cdot \text{cm}$. In our confidence interval construction, this corresponds to an upper bound of $d_e < 1.95\sigma = 8.7 \times 10^{-29}\ e \cdot \text{cm}$ at 90% confidence. A comparison between three different 90% confidence interval constructions for small values of d_e^{meas} is shown in Fig. 3: a strict upper bound computed using a folded Gaussian, a two-sided Gaussian central confidence band, and the Feldman-Cousins approach described here, which unifies upper limits and two-sided intervals. For our measurement outcome, the Feldman-Cousins intervals yield a few-percent larger EDM limit than the folded Gaussian upper bound would have.

References

[1] A. Aguilar, L. B. Auerbach, R. L. Burman, D. O. Caldwell, E. D. Church, A. K. Cochran, J. B. Donahue, A. Fazely, G. T. Garvey, R. M. Gunasingha, R. Imlay, W. C. Louis, R. Majkic, A. Malik, W. Metcalf, G. B. Mills, V. Sandberg, D. Smith, I. Stancu, M. Sung, R. Tayloe, G. J. VanDalen, W. Vernon, N. Wadia, D. H. White, and S. Yellin. Evidence for neutrino oscillations from the observation of ν_e appearance in a ν_μ beam. *Phys. Rev. D*, 64:112007, Nov 2001.

[2] M. Apollonio, A. Baldini, C. Bemporad, E. Caffau, F. Cei, Y. Dclais, H. de Kerret, B. Dieterle, A. Etenko, J. George, G. Giannini, M. Grassi, Y. Kozlov, W. Kropp, D. Kryn, M. Laiman, C.E. Lane, B. Lefivre, I. Machulin, A. Martemyanov, V. Martemyanov, L. Mikaelyan, D. Nicol, M. Obolensky, R. Pazzi, G. Pieri, L. Price, S. Riley, R. Reeder, A. Sabelnikov, G. Santin, M. Skorokhvatov, H. Sobel, J. Steele, R. Steinberg, S. Sukhotin, S. Tomshaw, D. Veron, and V. Vyrodov. Limits on neutrino oscillations from the {CHOOZ} experiment. *Physics Letters B*, 466(24):415 – 430, 1999.

[3] R. Bernabei, P. Belli, R. Cerulli, F. Montecchia, M. Amato, G. Ignesti, A. Incicchitti, D. Prosperi, C.J. Dai, H.L. He, H.H. Kuang, and J.M. Ma. Search for {WIMP} annual modulation signature: results from dama/nai-3 and dama/nai-4 and the global combined analysis. *Physics Letters B*, 480(12):23 – 31, 2000.

[4] Gary J. Feldman and Robert D. Cousins. Unified approach to the classical statistical analysis of small signals. *Physical Review D*, 57(7):3873–3889, April 1998.

[5] J J Hudson, D M Kara, I J Smallman, B E Sauer, M R Tarbutt, and E A Hinds. Improved measurement of the shape of the electron. *Nature*, 473(7348):493–6, May 2011.

[6] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project. Measurements of and from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565, 1999.

[7] B. Regan, Eugene Commins, Christian Schmidt, and David DeMille. New Limit on the Electron Electric Dipole Moment. *Physical Review Letters*, 88(7):18–21, February 2002.