Combining measurements

Soumen Halder

January 4, 2023

Overview

References and documentation

References

- Statistical Methods for Data Analysis in Particle Physics (Luca Lista) - Chapter 6
- Lyons, L., Gibaut, D., Clifford, P.: How to combine correlated estimates of a single physical quantity. Nucl. Inst. Methods A270, 110–117 (1988)
- A. Valassi, Combining correlated measurements of several different physical quantities. Nucl. Instr. Meth. A 500, 391 (2003)

Documentation

- Reproduced the last paper in this repository

Introduction

The inputs

Suppose we have,

- n experimental results, denoted as $y_i = y_1,\ y_2,\ y_3\ ...\ y_n$
- Covariance matrix of the measurements $M_{ij} = \text{cov}(y_i,y_j)$ is a $n \times n$ matrix
- N observables, $X_\alpha = X_1, X_2,\ X_3\ ...\ X_N$,so it's obvious $n = \sum_{\alpha=1}^{N} n_\alpha \geq N$
- The link between measurement y_i and the observables X_α is denoted by a $n \times N$ matrix,

$$U_{i\alpha} = \begin{cases} 1, & \text{if } y_i \text{ is a measurement of} X_\alpha \\ 0, & \text{if } y_i \text{ is a not measurement of} X_\alpha \end{cases}$$

Basic theory (contd..)

Example 1

- Measurement of Branching fraction in e and τ decay channel in two different experiments A and B.
  $\hat{\mathcal{B}}^e_A = (10.50 \pm 1)\%$, $\hat{\mathcal{B}}^e_B = (13.50 \pm 3)\%$, $\hat{\mathcal{B}}^\tau_A = (9.50 \pm 3)\%$, $\hat{\mathcal{B}}^\tau_B = (14.00 \pm 3)\%$
- Here $y_i = \{\hat{\mathcal{B}}^e_A, \hat{\mathcal{B}}^e_B, \hat{\mathcal{B}}^\tau_A, \hat{\mathcal{B}}^\tau_B\}$, with $i \in [1,4]$ and $n = 4$
- Covariance matrix is diagonal matrix (**in absence of any correlation between measurements**),
  $M = 10^{-4}\text{diag}(1^2, 3^2, 3^2, 3^2)$
- One set of observable of interest may be $X_\alpha = \{\mathcal{B}^e, \mathcal{B}^\tau\}$ with $\alpha \in [1,2]$, $N = 2$, and $n_1 = 2$, $n_2 = 2$ with the link matrix will then be $U = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}^T$
- Another set of observables can be $X_\alpha = \{\mathcal{B}^\ell\}$ assuming lepton universality. Here $\alpha = 1$, $N = 1$, $n_1 = 1$ with $U = (1,1,1,1)^T$

Example 2

- Measurement of Branching fraction of $B \to K^*\gamma$ in four decay channels of K^*.
  $\hat{\mathcal{B}}^0_{K^+\pi^-} = (4.5 \pm 0.3 \pm 0.2) \times 10^{-5}$, $\hat{\mathcal{B}}^0_{K^0_S\pi^0} = (4.4 \pm 0.9 \pm 0.6) \times 10^{-5}$, $\hat{\mathcal{B}}^+_{K^+\pi^0} = (4.5 \pm 0.5 \pm 0.4) \times 10^{-5}$,
  $\hat{\mathcal{B}}^+_{K^0_S\pi^+} = (4.4 \pm 0.6 \pm 0.4) \times 10^{-5}$,
- Here $y_i = \{\hat{\mathcal{B}}^0_{K^+\pi^-}, \hat{\mathcal{B}}^0_{K^0_S\pi^0}, \hat{\mathcal{B}}^+_{K^+\pi^0}, \hat{\mathcal{B}}^+_{K^0_S\pi^+}\}$, with $i \in [1,4]$ and $n = 4$
- Covariance matrix is diagonal matrix (**in absence of any correlation between measurements**),
  $M = 10^{-10}\text{diag}(0.3^2 + 0.2^2, 0.9^2 + 0.6^2, 0.3^2 + 0.4^2, 0.9^2 + 0.4^2)$
- One set of observable of interest may be $X_\alpha = \{\mathcal{B}^0, \mathcal{B}^+\}$ with $\alpha \in [1,2]$, $N = 2$, and $n_1 = 2$, $n_2 = 2$ the link matrix will then be $U = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}^T$
- Another set of observables can be $X_\alpha = \{\mathcal{B}^{+/0}\}$ assuming isospin symmetry. Here $\alpha = 1$, $N = 1$, $n_1 = 1$ with $U = (1,1,1,1)^T$

Basic theory (contd..)

The desired outputs

- **Observable Estimation:** $\hat{x}_\alpha = \lambda_{\alpha i} y_i$ as estimation of the observable X_α
- **Covariance matrix of measured observables:** as cov($\hat{x}_\alpha$,$\hat{x}_\beta$) $= \lambda_{\alpha i} M_{ij} \lambda_{\beta j} = (\lambda M \lambda^T)_{\alpha\beta}$
  - The ref. says $\lambda = (U^T M^{-1} U)^{-1} (U^T M^{-1})$, or in index notation $\lambda_{\alpha i} = \sum_{\beta=1}^{N} (U^T M^{-1} U)^{-1}_{\alpha\beta} (U^T M^{-1})_{\beta i}$.
  - Puting that in covariance matrix expression we get, cov($\hat{x}_\alpha$,$\hat{x}_\beta$)= $(U^T M^{-1} U)^{-1}_{\alpha\beta}$
- **Decomposition of covariance matrix to statistical and systematics:** Suppose the covariance of measurements can be written as sum of statistical and systematic uncertainty like $M_{ij} = M_{ij}^{\text{stat}} + M_{ij}^{\text{sys}}$. The covariance matrix of observables can also be decomposed as,

$$\text{cov}(\hat{x}_\alpha, \hat{x}_\beta) = (\lambda M^{\text{stat}} \lambda^T)_{\alpha\beta} + (\lambda M^{\text{sys}} \lambda^T)_{\alpha\beta}$$

Example 1

- First we have to calculate the weight matrix, using the formulae above we get $\lambda = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$
- $\hat{x} = \{10.80\%, 11.75\%\}$
- Covariance matrix of measured observables is, Cov$\{\mathcal{B}^e, \mathcal{B}^\tau\} = \begin{pmatrix} 9 \times 10^{-5} & 0 \\ 0 & 45 \times 10^{-5} \end{pmatrix}$

Remarks

- The covariance matrix, shown in example is block diagonal i.e no correlation between measurements of inter-observables. In simple words there are no correlation between measurements of let's say same experiment (eg. $\hat{\mathcal{B}}^{e}_{A}$ and $\hat{\mathcal{B}}^{\tau}_{A}$). As a consequence of this fact we can actually decompose the whole problem into two independent problems.
- If we look into each diagnoal block of the block-diagonal covariance matrix we can actually see they are also diagonal diag (σ_1^2, σ_2^2). For such simple cases, $\{\lambda\} = \{\frac{1/\sigma_1^2}{1/\sigma_1^2+1/\sigma_2^2}, \frac{1/\sigma_2^2}{1/\sigma_1^2+1/\sigma_2^2}\}$ and $\text{cov}(y_1, y_2) = \frac{1}{\sqrt{1/\sigma_1^2+1/\sigma_2^2}}$
- For non-zero correlation of two measurements, $M = \begin{pmatrix} \sigma_1^2 & \sigma_c^2 \\ \sigma_c^2 & \sigma_2^2 \end{pmatrix}$ the weight factors are,

  $\lambda = \{\frac{\frac{1}{\sigma_1^2-\sigma_c^2}}{\frac{1}{\sigma_1^2-\sigma_c^2}+\frac{1}{\sigma_2^2-\sigma_c^2}}, \frac{\frac{1}{\sigma_2^2-\sigma_c^2}}{\frac{1}{\sigma_1^2-\sigma_c^2}+\frac{1}{\sigma_2^2-\sigma_c^2}}\}$ and $\text{cov}(x_1, x_2) = \frac{\sigma_1^2\sigma_2^2-\sigma_c^4}{\sigma_1^2+\sigma_2^2-2\sigma_4^c}$

  - If we can decompose M as $M = \begin{pmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix} + \begin{pmatrix} \sigma_c^2 & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 \end{pmatrix}$ then, $\{\lambda\} = \{\frac{1/\sigma_a^2}{1/\sigma_a^2+1/\sigma_b^2}, \frac{1/\sigma_2^b}{1/\sigma_a^2+1/\sigma_b^2}\}$ and $\text{cov}(y_1, y_2) = \frac{1}{\sqrt{1/\sigma_a^2+1/\sigma_b^2}} + \sigma_c^2$. But this is too much assumption, since we are assuming 100% correlation with same absolute uncertainty for both the quantity (For our case relative uncertainties are same for correlated systematic uncertainty, not absolute uncertainty).

Finding correlation matrix for two measurements

- Suppose there are two measurements , $y_1 = \hat{y}_1 \pm \sigma_1, \hat{y}_2 \pm \sigma_2$. The degree of correlation between these measurements (correlation coefficient) is ρ, then the covariance matrix will be, $M = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

But I thought cross terms are addition of two uncertainty!!

- First of all let's see what is covariance matrix? A covariance matrix between two measurements are defined as, $\begin{pmatrix} \mathrm{Var}(y_1) & \mathrm{Cov}(y_1, y_2) \\ \mathrm{Cov}(y_1, y_2) & \mathrm{Var}(y_1) \end{pmatrix}$, with $\mathrm{Var}(y_1) = E[y_1^2] - (E[y_1])^2$, and $\mathrm{cov}(y_1 y_2) = E[y_1 y_2] - (E[y_1]E[y_2])$
- But how the "Uncertainty added for 100% correlation concepts come"?
  - Suppose $f = f(y_1, y_2)$
  - From propagation of error we get $\sigma_f^2 = \left(\frac{\partial f}{\partial y_1}\right)^2 \sigma_1^2 + \left(\frac{\partial f}{\partial y_2}\right)^2 \sigma_2^2 + 2\frac{\partial f}{\partial y_1}\frac{\partial f}{\partial y_2}\mathrm{cov}(y_1, y_2)|_{\hat{y}_1, \hat{y}_2}$, we define correlation coefficient ρ as $\mathrm{cov}(\sigma_1, \sigma_2) = \rho\sigma_1\sigma_2$
    - for $f = y_1 + y_2$, $\sigma_f^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$. For $\rho = 100\%$, $\sigma_f = \sigma_1 + \sigma_2$, and for $\rho = 0$, $\sigma_f = \sqrt{\sigma_1^2 + \sigma_2^2}$
    - For $f = y_1 y_2$, $\sigma_f^2/f^2 = \sigma_1^2/y_1^2 + \sigma_2^2/y_2^2 + 2\rho\sigma_1/y_1\sigma_2/y_2$. For $\rho = 100\%$, $\sigma_f/f = \sigma_1/y_1 + \sigma_2/y_2$, and for $\rho = 0$, $\sigma_f/f = \sqrt{\sigma_1^2/y_1^2 + \sigma_2^2/y_2^2}$

So it should be clear now that, **in the off-diagonal element of covariance matrix we are putting only the covariance term. Hence there are no question of full square to come**. For simple case $f = y_1 + y_2$, $\sigma_f^2 = \sum_{ij} M_{ij}$

Finding correlation matrix for two measurements (contd..)

Let's review the Example 2 again. We have (for simplicity ignoring 10^{-5}),

- $\hat{\mathcal{B}}^0_{K^+\pi^-} = (4.5(b_1) \pm 0.3(\text{stat}) \pm 0.2(\text{syst})) \times 10^{-5}$
- $\hat{\mathcal{B}}^0_{K^0_S\pi^0} = (4.4(b_2) \pm 0.9(\text{stat}) \pm 0.6(\text{syst})) \times 10^{-5}$,

Statistical covariance matrix

Statistical component are uncorrelated to each other, So $M_{\text{stat}} = \begin{pmatrix} 0.09 & 0 \\ 0 & 0.81 \end{pmatrix}$

Systematics covariance matrix

Some of the sources of systematic uncertainties are correlated. Let's take an example of relative systematics uncertainty (in %) table of two sources.

Systematic source	$K^{*0} \to K^+\pi^-$	$K^{*0} \to K^0_S\pi^0$
γ selection	r_1 (0.3)	r_1 (0.3)
K^+-pid	r_2 (0.8)	–
K^0_S-selection	–	r_3 (2.4)

- Needless to say diagonal components of covariance matrix will be $b_1\sqrt{r_1^2 + r_2^2}$, $b_2\sqrt{r_1^2 + r_3^2}$
- Only γ-selection systematics will contribute to the cross term. The absolute uncertainties for this source are $\sigma_1 = b_1 r_1$ and $\sigma_2 = b_2 r_1$. We are using the same systematics (provided by performance group) for both measurements. So they are 100% correlated, on the other hand $\rho = 1$. Hence the cross term will be $\rho\sigma_1\sigma_2 = b_1 b_2 r_1^2$

So the systematic component of covariance matrix is, So $M_{\text{sys}} = 10^{-4}\begin{pmatrix} b_1^2(r_1^2 + r_2^2) & b_1 b_2 r_1^2 \\ b_1 b_2 r_1^2 & b_2^2(r_1^2 + r_3^2) \end{pmatrix}$

So the total covariance matrix will be $M = M_{\text{stat}} + M_{\text{sys}}$

Is that all?

No!!!!!!!!

Considering inter-observables correlation

- In the previous discussion we have considered correlation between two measurements which contributes to a single observables.
- But there are correlation between measurements contributing to different observables, like $\hat{\mathcal{B}}^0_{K^+\pi^-}$, $\hat{\mathcal{B}}^+_{K^0_S\pi^+}$.
  $N_{B\bar{B}}$, photon selection, π^0/η-veto, tracking, Kaon-pid are examples of such sources
- Considering such correlation the covariance matrix for problem under consideration is neither block-diagonal nor diagonal. So we have to really deploy the main algorithm, and find the weights.

Calculating weights for Example 2

Source_id	Source	1	2	3	4	correlation
0	No.-of-BB-events	1.6	1.6	1.60	1.6	1234
1	Photon-selection	0.3	0.3	0.30	0.3	1234
2	pi0-eta_veto	3.8	3.8	3.80	3.8	1234
3	Pion_identification	0.6	0.0	0.00	0.6	14
4	Kaon_identification	0.8	0.0	0.80	0.0	13
5	KS0_reconstruction	0.0	2.4	0.00	2.4	24
6	pi0_selection	0.0	3.4	3.40	0.0	23
7	Tracking_efficiency	1.4	1.4	0.70	1.4	1234
8	MVA_selection	2.0	6.0	2.00	4.0	None
9	MC_statistics	0.2	0.5	0.30	0.3	None
10	PDF_shape_parameters	1.0	6.4	2.75	1.0	None
11	Misreconstructed_signal	1.5	7.0	5.30	2.8	None

Figure: Systematics table for the example 2. The columns 1, 2, 3, 5 represent four channels. The correlation columns represent the channels that are affected by the corresponding systematic source

We are calculating V_{ij} = If the relative systematics uncertainty due to source k, for mode i, is r_{ki} then,

- $M_{\text{sys},ii} = 10^{-4} \times b_i^2 \sum_{\text{k}=0}^{11} r_{ki}^2$
- $M_{\text{sys},ij} = 10^{-4} \times b_i b_j \sum_{\text{k}=0}^{11} r_{ki}^2$ (if k in the correlation columns) where b_s is central value of $\mathcal{B}$ for mode s

Calculating weights for example 2 (Contd..)

The inputs

$$y = \begin{pmatrix} 4.5 \\ 4.4 \\ 5.0 \\ 5.4 \end{pmatrix}$$

$$U = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

$$M_{\text{sys}} = \begin{pmatrix} 0.0553635 & 0.037719 & 0.040995 & 0.0471663 \\ 0.037719 & 0.31475488 & 0.064108 & 0.05894856 \\ 0.040995 & 0.064108 & 0.17380625 & 0.047466 \\ 0.0471663 & 0.05894856 & 0.047466 & 0.1460916 \end{pmatrix}$$

$$M_{\text{stat}} = \begin{pmatrix} 0.09 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.81 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.25 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.36 \end{pmatrix}$$

Calculating weights for example 2 (Contd..)

The outputs

$$\begin{pmatrix} \mathcal{B}^0 \\ \mathcal{B}^+ \end{pmatrix} = \begin{pmatrix} 4.48852787423507 \\ 5.18171567716517 \end{pmatrix}$$

$$\text{cov}(\mathcal{B}^+, \mathcal{B}^0) = \begin{pmatrix} 0.135632615003812 & 0.0453979008426059 \\ 0.0453979008426059 & 0.253908753899837 \end{pmatrix}$$

$$\text{cov}^{\text{stat}}(\mathcal{B}^+, \mathcal{B}^0) = \begin{pmatrix} 0.0811124133699011 & -1.80464355460596 \cdot 10^{-5} \\ -1.80464355460596 \cdot 10^{-5} & 0.148755178004303 \end{pmatrix}$$

$$\text{cov}^{\text{syst}}(\mathcal{B}^+, \mathcal{B}^0) = \begin{pmatrix} 0.054520201633911 & 0.045415947278152 \\ 0.045415947278152 & 0.105153575895534 \end{pmatrix}$$

Finally,

- $\mathcal{B}^0 = 4.48 \pm 0.28(\text{stat.}) \pm 0.23(\text{stst.}) \times 10^{-5}$
- $\mathcal{B}^+ = 5.18 \pm 0.38(\text{stat.}) \pm 0.32(\text{stst.}) \times 10^{-5}$

What if we ignore inter-observables correlation?

The outputs

$$\begin{pmatrix} \mathcal{B}^0 \\ \mathcal{B}^+ \end{pmatrix} = \begin{pmatrix} 4.49098968211063 \\ 5.18029012803338 \end{pmatrix}$$

$$\mathrm{cov}(\mathcal{B}^+, \mathcal{B}^0) = \begin{pmatrix} 0.135664388359571 & 0.0 \\ 0.0 & 0.254180170358464 \end{pmatrix}$$

$$\mathrm{cov}^{\mathrm{stat}}(\mathcal{B}^+, \mathcal{B}^0) = \begin{pmatrix} 0.0810881523612081 & 0.0 \\ 0.0 & 0.148560861598515 \end{pmatrix}$$

$$\mathrm{cov}^{\mathrm{syst}}(\mathcal{B}^+, \mathcal{B}^0) = \begin{pmatrix} 0.054576235998363 & 0.0 \\ 0.0 & 0.105619308759949 \end{pmatrix}$$

Finally,

- $\mathcal{B}^0 = 4.48 \pm 0.28(\text{stat.}) \pm 0.23(\text{stst.}) \times 10^{-5}$
- $\mathcal{B}^+ = 5.18 \pm 0.38(\text{stat.}) \pm 0.32(\text{stst.}) \times 10^{-5}$

- The final result and uncertainty does not vary much, but the correlation is zero here. But since we do not need to provide any correlation between these measurements, we can simply use two independent combination method
- But if anyone try to combine these, assuming isospin symmetry then he will end up with wrong results if he ignores the correlation