Notes on Model Fitting to Data

Anirban Das
(Dated: December 11, 2019)

Suppose, we want to explain an observation of some spectrum using a theory model. The data is given by an array d_i. The corresponding theory predictions are $d_i^{\rm th}$. This $d_i^{\rm th}$ includes both signal and background, $d_i^{\rm th} \equiv s_i + b_i$. Therefore the likelihood for this theory is

$$-2\ln\mathcal{L} = \sum_i \frac{(d_i - d_i^{\rm th})^2}{\sigma_i^2}\,. \tag{1}$$

Here σ_i are the total systematic and statistical errors of the data. This is valid only if the data d_i are independent. If this is not the case and they have a covariance matrix $\mathbf{V}$ then

$$-2\ln\mathcal{L} = (\mathbf{d} - \mathbf{d}^{\rm th})^T\mathbf{V}^{-1}(\mathbf{d} - \mathbf{d}^{\rm th})\,. \tag{2}$$

Now if the we define the *null hypothesis* likelihood $\mathcal{L}_0$, i.e., when the data is explained with the background only

$$-2\ln\mathcal{L}_0 = \sum_i \frac{(d_i - b_i)^2}{\sigma_i^2}\,, \tag{3}$$

then the *test statistics* TS is defined as

$$TS = -2\ln\left(\frac{\mathcal{L}_0}{\mathcal{L}}\right)\,. \tag{4}$$

Basically, TS is the difference between the χ^2s of the fits with and without signal. It can be calculated for a range of points in the parameter space of the theory. The point having the greatest TS is expected to be best description of the data.

We could just minimize χ^2 to find the best-fit point. Why bother computing TS?

TS is typically used for source detection in a skymap of, let's say gamma-ray. For example, Fermi collaboration uses this method to detect sources on the sky. In this case, the TS formula is modified to the following

$$TS = -2\ln\left(\frac{\mathcal{L}_{\rm max,0}}{\mathcal{L}_{\rm max}}\right) \tag{5}$$

where $\mathcal{L}_{\rm max}$ is the maximum likelihood found by minimizing χ^2 varying all parameters of the model. The TS is then the relative measurement of fitting with and without a source at a particular position in the sky[1].

Small number of counts

If the count of particles in each data bin is small then the normal distribution cannot be used for likelihood calculation. The more appropriate function is the Poisson distribution,

$$P(n;\mu) = \frac{\mu^n \exp^{-\mu}}{n!}\,. \tag{6}$$

[1] Fermi likelihood overview.

Therefore

$$\mathcal{L} = \prod_i \frac{(d_i^{\text{th}})^{d_i}}{d_i!} \exp(-d_i^{\text{th}}) = \exp(-N_{\text{exp}}) \prod_i \frac{(d_i^{\text{th}})^{d_i}}{d_i!} \,, \tag{7}$$

where N_{exp} is the total expected count in all bins[2].

[2] See this for more about binning issues.