A tiny note on hypothesis testing: from particle physics to cosmology?

1 Introduction

Consider the probability density function (PDF) $p(x)$ of a random variable X. For instance, X could be (M, C) with M being the invariant mass and C the angle in the center-of-mass (C.O.M.) frame. We consider a source experiment, and define the "events number density" $n(x)$, that is, the distribution normalized to the total number of expected events for the experiment, N:

$$n(x) = Np(x)$$

This implies that the integral of $n(x)$ over the entire space of x must be equal to N:

$$\int dx\, n(x) = N$$

Given a reference model R, we can predict the theoretical differential cross-section and apply observational effects, obtaining the distribution:

$$n(x|R)$$

normalized with the condition:

$$\int dx\, n(x|R) = N(R)$$

Now, consider a second model w, which will predict a different distribution $n(x|w)$. We parametrize it as:

$$n(x|w) = n(x|R)e^{f(x,w)}$$

2 Likelihood

To write the likelihood, suppose we consider binned data. The bins are usually considered independent, and the likelihood $L(R|D)$ is a product of these likelihoods:

$$L_{\text{bin}}(R|D) = \prod_i \frac{n(x_i|R)^{n_i}}{n_i!} e^{-n(x_i|R)}$$

where $n(x_i|R)$ is the theoretical prediction of the event number in the bin i (corresponding to variables x_i) under the model R, and n_i are the observed data in bin i.

If instead we have unbinned data, the likelihood is:

$$L_{\text{un}}(D|R) = \prod_{i=1}^{N} P(x_i|R)$$

where $P(x_i|R)$ is the probability of obtaining the observed event x_i under model R, and N is the total number of observations in the dataset.

Now, N is not fixed; it actually follows a Poisson distribution with mean $\bar{N}(R)$, which depends on the model. This defines what is called extended likelihood:

$$L(D|R) = \frac{\bar{N}(R)^{N_{\text{obs}}}}{N_{\text{obs}}!} e^{-\bar{N}(R)} \prod_{i=1}^{N_{\text{obs}}} P(x_i|R)$$

Here, N_{obs} is the actual observed number of events. We can rewrite L by defining the expected number of events observed with x_i variables as

$$n(x_i|R) = \bar{N}(R)P(x_i|R)$$

so that

$$L(D|R) = \frac{e^{-\bar{N}(R)}}{N_{\text{obs}}!} \prod_{i=1}^{N_{\text{obs}}} n(x_i|R)$$

3 Normal Limit

We now consider the following variable, which is the standard score for the Poisson distribution (since $\sigma_N(R)/\sqrt{\bar{N}(R)} = \sqrt{\bar{N}(R)}$)

$$y = \frac{N_{\text{obs}} - \bar{N}(R)}{\sqrt{\bar{N}(R)}}$$

In the limit $N_{\text{obs}} \to \infty$, $\bar{N}(R) \to \infty$, y fixed, the Poisson distribution approaches a standard normal distribution for y. Thus:

$$L(D|R) = \frac{1}{\sqrt{2\pi\bar{N}(R)}} \exp\left(-\frac{(N_{\text{obs}} - \bar{N}(R))^2}{2\bar{N}(R)}\right) \prod_{i=1}^{N_{\text{obs}}} P(x_i|R)$$

For the power spectrum in cosmology, however, our data are always binned and there are many events in each single bin. So we go back to

$$L_{\text{bin}}(D|R) = \prod_i \frac{n(x_i|R)^{n_i}}{n_i!} e^{-n(x_i|R)}$$

Here n_i are the events in each bin, and $n(x_i|R)$ is the expected value of events in each bin. In cosmology, these correspond, respectively, to the observed values of the power spectrum P_i and the theoretical predictions $P(k_i)$ in wavenumber bin k_i. In the central limit, the binned likelihood is:

$$L(D|R)_{\rm bin} \approx \prod_i \frac{1}{\sqrt{2\pi\, n(x_i|R)}} \exp\left(-\frac{(n_i - n(x_i|R))^2}{2\, n(x_i|R)}\right) =$$
$$= \frac{1}{\prod_i \sqrt{2\pi\, n(x_i|R)}} \exp\left(-\sum_i \frac{(n_i - n(x_i|R))^2}{2\, n(x_i|R)}\right)$$

This is a Gaussian likelihood with a diagonal covariance matrix scaling as $n(x_i|R)$. The difference w.r.t. cosmology is that the covariance is only approximately diagonal, and it scales as $P(k_i)^2$ instead of $P(k_i)$. Also, practically all the time we do data analysis with fixed covariance.

4 Log-Likelihood Ratio

Let us now construct the usual log-likelihood ratio for goodness-of-fit tests. For Poisson statistics, we have:

$$\ln L_{\rm bin}(D|R) = \sum_i \ln L_i = \sum_i \left[n_i \ln n(x_i|R) - \ln(n_i!) - n(x_i|R)\right]$$

For another model w, parametrized by $n(x_i|w) = n(x_i|R)e^{f(x_i,w)}$, we have

$$\ln L_{\rm bin}(D|w) = \sum_i \left[n_i \ln n(x_i|R) + n_i f_i(w) - \ln(n_i!) - n(x_i|R)e^{f_i(w)}\right]$$

The log-likelihood ratio becomes

$$\ln \frac{L_{\rm bin}(D|w)}{L_{\rm bin}(D|R)} = \sum_i \left[n_i f_i(w) - n(x_i|R)(e^{f_i(w)} - 1)\right]$$

This is, in fact, the loss function that is used to train the neural network approximating the function $f(w)$. Suppose instead we consider a Gaussian, binned, likelihood, with fixed diagonal covariance $\sigma_i^2 \delta_{ij}$. Then

$$\ln L_c(D|R) = -\frac{1}{2} \sum_i \left[\ln(2\pi\sigma_i^2) + \frac{(n(x_i|R) - n_i)^2}{\sigma_i^2}\right]$$

If the new model is parametrized by $n(x_i|w) = n(x_i|R)e^{f(x_i,w)}$, the log-likelihood ratio is:

$$\ln \frac{L_c(D|w)}{L_c(D|R)} = \frac{1}{2} \sum_i \frac{n(x_i|R)}{\sigma_i^2} \left[2n_i(e^{f_i(w)} - 1) - n(x_i|R)(e^{2f_i(w)} - 1)\right]$$

This looks different than the Poisson case. However, an interesting to note is that there is a maximum in f_i (for each term) for both the log-likelihoods, and in fact it is at the same point $f_i(w) = \ln \frac{n_i}{n(x_i|R)}$.

4.1 Non-diagonal covariance

In the generic case of a non-diagonal covariance (but still fixed, not depending on parameters), we get

$$\ln L_c(D|R) = -\frac{1}{2}\left[\ln\det(2\pi C) + \sum_{i,j}(n(x_i|R) - n_i)C^{-1}_{ij}(n(x_j|R) - n_j)\right] .$$

It is easy to get the log-likelihood ratio:

$$\ln\frac{L_c(D|w)}{L_c(D|R)} = \frac{1}{2}\sum_{i,j}\frac{n(x_i|R)}{2}\left[(e^{f_i(w)}C^{-1}_{ij} - C^{-1}_{ij})n_j - (e^{f_i(w)}C^{-1}_{ij}e^{f_j(w)} - C^{-1}_{ij})n(x_j|R)\right] .$$