Lecture note on statistics for classes given at the BND school

Lydia Brenner

September 29, 2025

Disclaimer: These lecture notes are prepared for my own use during the lectures, not for printing. The goal of these notes is to point out some key points in between the different sessions of throwing items and learning by doing.
These notes have been cleaned up upon request of participants, but are by no means a neat collection.

1 Introduction

We start with a Likelihood function

$$L(H) = P(x|H) = the\, likelihood\, function\, of\, H \tag{1}$$

note the following

- we consider the data as fixes
- this is not a pdf for H

Using this we can get to Neyman-Pearson Lemma, e.g. likelihood ratio

$$\frac{P(x|H_0)}{P(x|H_1)} \leq c_\alpha \tag{2}$$

where c_α is a constant chosen to give a test of the desired size. e.g. 5σ
Gives you an optimal scalar test statistic.

2 Estimation/fitting

Let's make it one step more difficult. Assume we have a dataset x with n data points $x = \{x_1, ..., x_n\}$, and unknown parameters a creating a joined pdf $P(x, a)$.

We want to know

- What is the value of a according to the data
- What is the uncertainty on that value
- Does the resulting $P(x, a)$ describe the data

This is called 'estimation' by statisticians and 'fitting' by physicists.

An Estimator is a function of all the x_i which returns some value for a ; $\hat{a}(x_1, x_2, ...x_n)$. There is no 'correct' estimator. You would like an estimator to be

- Consistent: $\hat{a}(x) \to a$ for $n \to \infty$
- Unbiased: $<\hat{a}> = a$
- Efficient: $V(\hat{a}) = <\hat{a}^2> - <\hat{a}>^2$ is small
- Invariant under reparametrisation: $\widehat{f(a)} = f(\hat{a})$
- Convenient

But no estimator is perfect, and these requirements are self-contradictory.

3 Maximum Likelihood estimation

The ML estimator: To estimate a using data $x_1, x_2, ..., x_n$, find the value of a for which the total log likelihood $\sum lnP(x_i; a)$ is maximum. 3 types of problem

1. Differentiate, set to zero, solve the equation(s) algebraically
2. Differentiate, set to zero, solve the equation(s) numerically
3. Maximise numerically

Things to note

- No deep justification for ML estimation, except that it works well
- These are not 'the most likely values' of a. They are the values of a for which the values of x are most likely
- The logs make the total a sum, easier to handle than a product
- Remember a minus sign if you use a minimiser

Going back to the list of requirement above

- Consistent: Almost always
- Unbiased; It is biassed. But the bias usually falls like 1/N
- Efficient: In the large N limit ML saturates the Minimum Variance Bound, and you can't do better than that
- Invariant under reparametrisation: clearly.
- Convenient: Usually

Minimum Variance bound: If $\hat{a}$ is unbiased the variance V limits the efficiency

$$V(\hat{a}) \geq <(\frac{\partial lnL}{\partial a})^2>^{-1} = <-\frac{\partial^2 lnL}{\partial a^2}>^{-1} \tag{3}$$

Side note: Same as likelihood featuring in Bayes' theorem, though emphasis here is that L is Likelihood for all measurements of sample.

With some trivial math you can show that

$$<(\frac{\partial lnL}{\partial a})^2> + <\frac{\partial^2 lnL}{\partial a^2}> = 0 \;\; Fisher\, information \tag{4}$$

4 Goodness of fit

Let's fit data points; Suppose your data is a set of x_i, y_i pairs with predictions $y_i = f(x_i|a)$ with x_i known precisely, y_i measured with Gaussian errors σ_i. Does the model $f(x;a)$ provide a good description of the y_i? Naïvely each term in $\chi^2 sum \approx 1$ more precisely

$$p(\chi^2, n) = \frac{1}{2^{n/2}\Gamma(n/2)} \chi^{n/2-1} e^{-\chi^2/2} \tag{5}$$

as n-dimensional Gaussion integrated over hypersphere.
This is quatified by p-value: probability that, if the model is true, χ^2 would be this large or larger.
Note: p-values apply for any test statistic.

So what are the possible reasons for a large χ^2

- bad theory
- bad data
- errors underestimated
- unsuspected negative correlation between data point (not often likely)
- bad luck

and the reasons for a small χ^2

- errors overestimated
- unsuspected positive correlation between data points (more likely)
- good luck

Note: although $-\frac{1}{2}\chi^2$ is a log likelihood, $-2lnL$ is NOT a χ^2 and it tells you nothing about the goodness of your fit.
Note2: Wilks' theorem says it does for differences in similar models. Useful for comparisons, but not absolute.

This leaves 4 ways of fitting data

- **Full maximum likelihood** Write down the likelihood and maximise $\sum_j lnP(x_j, a)$ where j runs over all events.
  Slow for large data samples and no goodness of fit.
- **Binned maximum likelihood** Put data in a histogram and maximise the log of the Poisson probabilities $\sum_i n_i lnf_i - f_i$, where i runs over all bins $f_i = NP(x_i)w$. Do NOT forget the bind width w.
  Quicker but lose info from any structure smaller than bin size.
- **Put data in a histogram and minimise χ^2 (Pearson)** Assume Poisson distributions are approximated by Gaussians. $\chi^2 = \sum_i (n_i - f_i)^2/f_i$
  Do not use if bin contents are small, but you do get a goodness of fit
- **Put data in a histogram and minimise χ^2 (Neyman)** $\chi^2 = \sum_i (n_i - f_i)^2/n_i$. This makes the algebra and fitting a lot easier, but introduces bias as downwards fluctuations get more weight and disaster if any $n_i = 0$.

The methods are not equivalent so choose carefully.

5 Systematics

Systematics are important now that statistics is no longer the limiting factor. What is a systematic error?
option 1: Bevington
Systematic error: reproducible inaccuracy introduced by faulty equipment, calibration, or technique.
option 2: Orear
Systematic effects is a general category which includes effects such as background, scanning efficiency, energy resolution, variation of counter efficiency with beam position, and energy, dead time, etc. The uncertainty in the estimation of such a systematic effect is called a systematic error.
Note: These are contradictory
Orear is RIGHT, Bevington is WRONG, So are a lot of other books and websites. An error is not a mistake... Bevington is describing Systematic mistakes Orear is describing Systematic uncertainties - which are 'errors' in the way we use the term.
Avoid using 'systematic error' and always use 'uncertainty' or 'mistake'? Probably impossible. But should always know which you mean.

So there are several ways to quantify different types of systematic uncertainties. My favourite is this:

- The Good ; Well controlled calibrations or your own measurements
- The Bad ; Other peoples measurements, model assumptions or poor analysis strategies
- The Ugly ; Different theory predictions or too fe parameters in your model

Looking at this in a more systematic (haha) or mathematical way we can quantify this as:

- Uncertainty in an explicit continuous parameter
  E.g. Uncertainty in efficiency, background, luminosity, branching ratio, cross section. Straight forward way to handle it. Standard combination of errors formula and algebra.
- Uncertainty in an implicit continuous parameter
  E.g. MC tuning parameter. No easy algebra.... Vary one parameter at a time by $\pm 1\sigma$ and look what happens to your analysis result.*
- Discrete uncertainties
  E.g. Model choices. If one model is preferred you can take $M_1 \pm |M_1 - M_2|$, if they're equally good you can take $\frac{M_1+M_2}{2} \pm |\frac{M_1-M_2}{2}|$, for N models you can take $\bar{M} \pm \sqrt{\frac{N}{N-1}(\bar{M^2} - \bar{M}^2)}$. **

*Note If the effect isn't equal in both directions, you need asymmetric errors, which gives it's own problems.
Note2 Your analysis results will have errors due to e.g. MC statistics. Some people add these (in quadrature). This is wrong. Technically correct thing to do is subtract them in quadrature, but this is not advised. Alternatively use more points regularly spaced. Alternatively use more points chosen at random according to Gaussian distribution
**Note Do not push these too hard, if the difference is not small you have a problem no matter what you do!. Study your model differences instead.

5.1 Confusions and take home messages

What's the difference between?

1. Evaluating implicit systematic errors: vary lots of parameters, see what happens to the result, and include in systematic error
2. Checks: vary lots of parameters, see what happens to the result, and don't include in systematic error

Questions to ask yourself:

- Are you expecting to see an effect? If so, it's an evaluation, if not, it's a check
- Do you clearly know how much to vary them by? If so, it's an evaluation. If not, it's a check.

So here are the rules to follow (take home messages) - Thanks to Roger Barlow

- Thou shalt never say 'systematic error' when thou meanest 'systematic effect' or 'systematic mistake'.
- Thou shalt know at all times whether what thou performest is a check for a mistake or an evaluation of an uncertainty.
- Thou shalt not incorporate successful check results into thy total systematic error and make thereby a shield to hide thy dodgy result.
- Thou shalt not incorporate failed check results unless thou art truly at thy wits' end.
- Thou shalt not add uncertainties on uncertainties in quadrature. If they are larger than chick-enfeed thou shalt generate more Monte Carlo until they shrink to become so.
- Thou shalt say what thou doest, and thou shalt be able to justify it out of thine own mouth; not the mouth of thy supervisor, nor thy colleague who did the analysis last time, nor thy local statistics guru

Do these, and thou shalt flourish, and thine analysis likewise.

6 Coverage

A frequentist property... That we should all use and understand!
The concept is in essence quite simple; If we give a confidence interval of x%, does it also mean that in x% of the time your result falls inside the confidence interval given if the experiment was repeated infinitely often. If the answer to this is yes, or it's even more, then you have *coverage*, yay! Is the answer no, then you do not. Obviously you should always have coverage, and you should reevaluate your uncertainties to figure out why you don't have this.
Note: This also means that if you move a little bit away from the minimum that you expect, the uncertainties don't change greatly and you can check this with toys.