Lecture Notes Statistic for Astronomy

Mick Veldhuis

September 22, 2019

Contents

1 Statistics basics

1.1 Cox's rules

The first rule states that the probability of something happening, let's call the event X, plus that of X not happening is one:

$$P(X) + P(\bar{X}) = 1 \tag{1}$$

The second rule (the product rule) states that the probability of events X and Y happening is given by:

$$P(X, Y) = P(X|Y) \cdot P(Y) \tag{2}$$

Where $P(X|Y)$ is the probability of X given Y happening.

1.2 Bayes' theorem

Bayes' theorem gives us the probability of the hypothesis H being true given the data D.

$$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D)} \tag{3}$$

Where $P(H|D)$ is called the posterior, $P(D|H)$ the likelihood, $P(H)$ the prior, and $P(D)$ the evidence.

1.3 Marginalization

Suppose you have a range of hypotheses $\{H_i\}$, where $i = 0, 1, 2, 3, \ldots$. Then

$$\sum_{i=0}^{N-1} P(H_i) = 1 \tag{4}$$

Then suppose we have some nuisance parameter X (quantities of no intrinsic interest, that sadly enter our analyses), then

$$\sum P(H_i, X) = \sum P(H_i|X) \cdot P(X) \tag{5}$$

$$= P(X) \sum P(H_i|X) \tag{6}$$

$$= P(X) \tag{7}$$

Which could also be written as

$$P(X) = \int_{-\infty}^{\infty} P(X, Y) \, dY \tag{8}$$

Suppose we now have some continuous case. Then we dive into the realm of probability density functions:

$$\text{pdf}(X, Y = y) = \lim_{\delta y \to 0} \frac{P(X, y \le Y < y + \delta y)}{\delta y} \tag{9}$$

Then the probability that the value of Y lies between y_1 and y_2 is given by

$$P(X, y_1 \le Y < y_2) = \int_{y_1}^{y_2} \text{pdf}(X, Y) \, dY \tag{10}$$

Where

$$\int_{-\infty}^{\infty} P(Y|X) \, dY = 1 \tag{11}$$

2 Parameter estimation

2.1 The best estimate

The best estimate is the value, say X_0, such that the posterior is at a maximum. So

$$\left. \frac{dP}{dX} \right|_{X=X_0} = 0 \quad \text{and} \quad \left. \frac{d^2 P}{dX^2} \right|_{X=X_0} < 0 \tag{12}$$

The reliability of our best estimate is then given by the width of the posterior about $X = X_0$. To find this width we take the Taylor expansion of the natural logarithm of the posterior about $X = X_0$:

$$L = \ln\big[P(X|\{data\})\big] \quad \Rightarrow \quad L = L(X_0) + \frac{1}{2}\frac{d^2P}{dX^2}\bigg|_{X_0}(X - X_0)^2 + \ldots \quad (13)$$

Since the first term of the expansion is constant we can discard it. Thus we only care about the quadratic term, ignoring higher order terms. Such that

$$[P(X|\{data\}) \approx A\exp\left[\frac{1}{2}\frac{d^2P}{dX^2}\bigg|_{X_0}(X - X_0)^2\right] \quad (14)$$

For A is some normalization constant. We basically approximated the posterior by the Gaussian (or normal) distribution:

$$P(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (15)$$

Such that $X = X_0 \pm \sigma$, where

$$\sigma = \left(-\frac{d^2P}{dX^2}\bigg|_{X_0}\right)^{-1/2} \quad (16)$$

Note that for X within $\pm\sigma$, the probability is 67%, for $\pm 2\sigma$ that is 95%, etc. Also note that in the case of an asymmetric posterior the best estimate is not really given by X_0 (also called the mode), but the actual value is more likely to be on either side of the mode. We then use the mean (or expectation value) since it takes the asymmetry into account:

$$E(X) = \langle X \rangle = \int X\, P(X|\{data\})\, dX \quad (17)$$

Then there's also the median (denoted by $\mu_{1/2}$), an estimate of the middle of a distribution, which can sometimes be a reasonable approximation of the best estimate of a distribution with significant outliers. The median is such that half of the points will lie above and below it.

2.2 Average of the Gaussian distribution

If $\sigma =$ constant

For a Gaussian distribution the best estimate μ is just the arithmetic average of the data points $\{x_k\}$:

$$\mu_0 = \frac{1}{N}\sum_{k=1}^{N} x_k \quad (18)$$

And our confidence in the reliability of μ_0 is taken into account as follows, such that the value of μ is given to be

$$\mu = \mu_0 \pm \frac{\sigma}{\sqrt{N}} \quad (19)$$

3

If $\sigma \neq$ constant

In this case we get a weighted average

$$\mu_0 = \sum_k w_k x_k \bigg/ \sum_k w_k, \text{ where } w_k = \frac{1}{\sigma_k^2} \tag{20}$$

where σ_k^2 is the variance of data point k. Such that μ is given by

$$\mu = \mu_0 \pm \left(\sum_k w_k \right)^{-1/2} \tag{21}$$

2.3   The variance

The variance of the posterior is the expectation value of the square of the deviations from the mean

$$\text{Var}(X) = \left\langle (X - \mu)^2 \right\rangle = \int (X - \mu)^2 P(X|\{data\}) \, dX \tag{22}$$

Which for the Gaussian distribution reduces to

$$\left\langle (X - \mu)^2 \right\rangle = \sigma^2 \tag{23}$$

Note σ is called the standard deviation, and when $\mu = 0$ it is called the root-mean square uncertainty.

2.4   The central limit theorem

It is the basic idea that any peaked, symmetric distribution with tails that asymptote to zero will eventually converge to a Gaussian will eventually converge to a Gaussian distribution for $N \to \infty$. It is important to note that it will not work for all distributions, but we will get to that.

2.5   Probability distributions

2.6   Binomial distribution

2.7   Poisson distribution

2.8   Lorentzian / Cauchy distribution