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

$$\begin{aligned} \sum P(H_i, X) &= \sum P(H_i|X) \cdot P(X) && (5) \\ &= P(X) \sum P(H_i|X) && (6) \\ &= P(X) && (7) \end{aligned}$$

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:

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

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

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

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

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

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 $X = X_0$:

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

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

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

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

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

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

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

Note that for X within $\pm\sigma$, the probability is 67%, for $\pm 2\sigma$ that is 95%, etc.