10/2/2022. LECTURE 2: NOTES MADE ON THE VISUALISER.

Two Bernoulli trials:

① Toss a coin n times and observe x heads, parameter θ_c

② Toss a drawing pin n times and observe x pin-ups, parameter θ_p

Problems treated identically by classical statistics.

* Is this sensible ? *

I have lots of experience of tossing coins and these are well known to have propensities near to $1/2$.

$\Rightarrow$ I have some knowledge about θ_c

Equally, I have little knowledge about the properties of drawing pins

$\Rightarrow$ I don't know much about θ_p.

Shouldn't I take these differences into account somehow ?

Classical approach provides NO SCOPE for taking into account different prior knowledge. It treats θ_c and θ_p as fixed unknown constants.

Bayesian approach: reflect our prior knowledge about θ_c and θ_p by specifying probability distributions for them.

For a general parameter θ,

- specify a PRIOR DISTRIBUTION $f(\theta)$ for θ
- combine with the LIKELIHOOD $f(x|\theta)$ to obtain
- POSTERIOR DISTRIBUTION $f(\theta|x)$ for θ given x.

using Bayes' Theorem:

$$\overbrace{f(\theta|x)}^{\text{POSTERIOR}} = \frac{f(x|\theta)f(\theta)}{\boxed{f(x)}} \propto \overbrace{f(x|\theta)}^{\text{LIKELIHOOD}}\overbrace{f(\theta)}^{\text{PRIOR}}$$

Function of θ

does not depend on θ

WRT θ

Bayesian analysis is concerned with distributions of θ and how they change in the light of new information (typically data)

With this approach, we can answer the question

"What value of θ is most likely given the data?"

rather than the question

"What value of θ makes the data most likely to occur?"

- distribution $f(x|\theta)$ irrelevant to Bayesian after data x has been observed.
- for the classicist, $f(x|\theta)$ is the ONLY distribution available.

1. THE BAYESIAN METHOD.

Note: use $f(\cdot)$ to represent density function irrespective of whether RVs are discrete or continuous.

Generally we make no distinction as to whether variables are univariate or multivariate.

1.1 Bayes' theorem

Let X and Y be RVs with joint density $f(x,y)$ [If we need to make the RVs explicit, can write $f_{X,Y}(x,y)$]. The marginal density of Y, $f(y)$, is

$$f(y) = \int_X f(x,y)\, dx$$

[NB. X could be multivariate: e.g. $X = (X_1, X_2)$ then

$$f(y) = \int_{X_1} \int_{X_2} f(x_1, x_2, y)\, dx_1\, dx_2 \quad]$$

The CONDITIONAL DISTRIBUTION of Y given $X = x$ is, for $f(x) > 0$,

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

so that, for example,

$$f(y) = \int_X f(y|x) f(x)\, dx$$

which is the THEOREM OF TOTAL PROBABILITY.

X and Y are INDEPENDENT if and only if, for all x and y

$$f(x,y) = f(x)f(y) \qquad [\text{Often write } X \perp\!\!\!\perp Y]$$

[An equivalent condition is that, for $f(x) > 0$,

$$f(y|x) = f(y)$$

so that nothing further can be learnt about Y by observing X (and vice versa)]

If Z is a third RV then X and Y are CONDITIONALLY INDEPENDENT given $Z = z$ if and only if, for all x, y, z

$$f(x, y|z) = f(x|z) f(y|z)$$ [written $(X \perp\!\!\!\perp Y) | Z$]

[NB. $$f(y|x, z) = \frac{f(x,y,z)}{f(x,z)} = \frac{f(x,y|z)}{f(x|z)} = f(y|z)$$

when $(X \perp\!\!\!\perp Y) | Z$

Having observed $Z = z$, nothing further can be learnt about Y by observing X]

Bayes' theorem states that, for $f(x) > 0$

$$f(y|x) = \frac{f(x|y) f(y)}{f(x)} = \frac{f(x|y) f(y)}{\int_Y f(x|y) f(y)\, dy}.$$

2 Bayes' theorem for parametric inference.

In a Bayesian analysis, we treat parameter θ as a RV and thus specify a density function $f(\theta)$. If we have data x then:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)} = \frac{f(x|\theta)f(\theta)}{\underbrace{\boxed{\int_\theta f(x|\theta)f(\theta)\,d\theta}}_{\text{normalising constant}}}$$

$$\propto f(x|\theta)f(\theta)$$

i.e. POSTERIOR $\propto$ PRIOR $\times$ LIKELIHOOD.