03/03/2022. LECTURE 8 - NOTES MADE ON THE VISUALISER.

[$\mathbb{E}(Z) = \mathbb{E}(\mathbb{E}(Z|\theta))$ and condition everywhere on X]

1. $$\mathbb{E}(Z|X) = \mathbb{E}(\mathbb{E}(Z|\theta, X)|X)$$
$$= \underbrace{\mathbb{E}(\underbrace{\mathbb{E}(Z|\theta)}_{\text{Calculate using } f(z|\theta)\text{, giving a function of } \theta}|X)}_{\text{Calculate using } f(\theta|x) \text{ as } \mathbb{E}(Z|\theta) = g(\theta)} \text{ if } (Z \perp\!\!\!\perp X)|\theta$$

[$\mathrm{Var}(Z) = \mathrm{Var}(\mathbb{E}(Z|\theta)) + \mathbb{E}(\mathrm{Var}(Z|\theta))$. See question sheet one for a proof. Use this result and condition everywhere on X]

2. $$\mathrm{Var}(Z|X) = \mathrm{Var}(\mathbb{E}(Z|\theta, X)|X) + \mathbb{E}(\mathrm{Var}(Z|\theta, X)|X)$$
$$= \underbrace{\mathrm{Var}(\underbrace{\mathbb{E}(Z|\theta)}_{\text{Calculate using } f(z|\theta)}|X)}_{\text{Calculate using } f(\theta|x)} + \underbrace{\mathbb{E}(\underbrace{\mathrm{Var}(Z|\theta)}_{\text{Calculate using } f(z|\theta)}|X)}_{\text{Calculate using } f(\theta|x)} \text{ if } (Z \perp\!\!\!\perp X)|\theta$$

as $\mathbb{E}(Z|\theta) = g(\theta)$
$\mathrm{Var}(Z|\theta) = h(\theta)$

EXAMPLE: BINOMIAL - BETA.

$$\left.\begin{aligned} \theta &\sim \mathrm{Beta}(\alpha, \beta) \\ X|\theta &\sim \mathrm{Bin}(n, \theta) \end{aligned}\right\} \quad \theta|x \sim \mathrm{Beta}(a, b) \quad a = \alpha + x, \quad b = \beta + n - x$$

$Z|\theta \sim \mathrm{Bin}(m, \theta)$ and $(X \perp\!\!\!\perp Z)|\theta$.

$$\mathbb{E}(Z|X) = \mathbb{E}(m\theta|X) \quad \text{as } Z|\theta \sim \mathrm{Bin}(m, \theta) \text{ so } \mathbb{E}(Z|\theta) = m\theta$$
$$= m\,\mathbb{E}(\theta|X)$$
$$= \frac{ma}{a+b} \quad \text{as } \theta|x \sim \mathrm{Beta}(a, b)$$
$$= \frac{m(\alpha + x)}{\alpha + \beta + n}.$$

Note the explicit dependence on the observed value of x.

$$\mathrm{Var}(Z|X) = \mathrm{Var}(mQ\,|X) + \mathbb{E}(mQ(1-Q)|X)$$

as $Z|Q \sim \mathrm{Bin}(m, Q)$ so $\mathbb{E}(Z|Q) = mQ$

$\mathrm{Var}(Z|Q) = mQ(1-Q)$

$$= m^2\,\mathrm{Var}(Q|X) + m\left[\mathbb{E}(Q|X) - \mathbb{E}(Q^2|X)\right]$$

$$= \frac{m^2 ab}{(a+b)^2(a+b+1)} + m\left[\frac{a}{a+b} - \frac{a(a+1)}{(a+b)(a+b+1)}\right]$$

where $a = \alpha + x$, $b = \beta + n - x$

Note: X and Z are NOT INDEPENDENT and nor would we expect them to be. If we don't know Q then we would expect observing $X = x$ will be informative for Z.

Consider such a model from a classical perspective.

We would view X and Z as a random sample where each Bernoulli trial was IID. As we see from the prediction of Z given $X = x$ this is slightly misleading: they are only independent conditional on the parameter Q.

2.2 EXCHANGEABILITY.

One of the key differences between the classical and Bayesian schools is that observations that the former treat as IID are treated as EXCHANGEABLE by the latter.

The concept of EXCHANGEABILITY was introduced by BRUNO DE FINETTI in the 1930s.

DEFINITION (FINITE EXCHANGEABILITY)

The RVs $X_1, \dots, X_n$ are judged to be FINITELY EXCHANGEABLE if the joint density function satisfies:

$$f(x_1, \dots, x_n) = f(x_{\pi(1)}, \dots, x_{\pi(n)})$$

outcome of X_1 outcome of X_n outcome of X_1 outcome of X_n

for all permutations π defined on the set $\{1, \dots, n\}$

EXAMPLE.

X_1 and X_2 are finitely exchangeable if $f(x_1, x_2) = f(x_2, x_1)$ $\forall x_1, x_2$. [so if X_1, X_2 discrete $\mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}(X_1 = x_2, X_2 = x_1)$ $\forall x_1, x_2$]

X_1, X_2, X_3 finitely exchangeable if

$$f(x_1, x_2, x_3) = f(x_1, x_3, x_2) = f(x_2, x_1, x_3) = f(x_2, x_3, x_1)$$
$$= f(x_3, x_1, x_2) = f(x_2, x_2, x_1)$$

In essence, exchangeability captures the notion that only the values matter and not the order in which the observations occur: the labels are uninformative.

NB. Finitely exchangeable implies that

$$f(x_1, \dots, x_m) = f(x_{\pi(1)}, \dots, x_{\pi(m)})$$

for all $m \leq n$

[e.g. $n = 3$ $f(x_1, x_2) = \int_{x_3} f(x_1, x_2, x_3)\, dx_3$ $f(x_2, x_1, x_3)$

$= \int_{x_3} f(x_2, x_1, x_3)\, dx_3$ by exchangeability

$$= f(x_2, x_1). \quad]$$

Exchangeability is a stronger statement than identically distributed (joints included) but weaker than independence.

IID observations are exchangeable:

$$f(x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i)$$

Exchangeable observations need not be independent:

EXAMPLE:

X_1, X_2 are exchangeable RVs with joint density,

$$f(x_1, x_2) = \frac{3}{2}(x_1^2 + x_2^2) \qquad \begin{array}{l} 0 < x_1 < 1 \\ 0 < x_2 < 1 \end{array}$$

However, they are not independent.