15/02/2022. LECTURE 3: NOTES MADE ON THE VISUALISER.

Last time: $f(\theta|x) \propto f(x|\theta) f(\theta)$

Typically, θ and X are continuous

[e.g. $X|\theta \sim N(\theta, \sigma^2)$ σ^2 known, θ continuous]

or θ is continuous and X discrete

[e.g. $X|\theta \sim Bin(n, \theta)$, θ continuous]

In exceptional cases, θ could be discrete.

The Bayesian method comprises of the following principle steps:

① Prior.

How? ← Obtain prior density $f(\theta)$. This expresses our knowledge about θ prior to observing the data.

② Likelihood.

How? ← Obtain likelihood function $f(x|\theta)$. This step simply describes the process giving rise to data x in terms of θ.

[Similar for both schools]

③ Posterior.

Hard in non-trivial cases ← Apply Bayes' theorem to derive the posterior $f(\theta|x)$. Expresses our knowledge about θ after observing x.

④ Inference

What criteria might we use? [Decision Theory] ← Derive appropriate inference statements from the posterior distribution.

e.g. Point estimates, interval estimates, prob. of a hypothesis,

1.3 Sequential Data Analysis.

Suppose we have two sources of data x and y. We can add the data sequentially.

$$f(\theta \mid x, y) \underset{\text{wrt } \theta}{\propto} f(x, y \mid \theta)\, f(\theta)$$

Now,

$$f(x, y \mid \theta) = f(y \mid x, \theta)\, f(x \mid \theta)$$

so that

$$f(\theta \mid x, y) \propto f(y \mid x, \theta)\, \underbrace{f(x \mid \theta)\, f(\theta)}_{\text{"likelihood of } x\text{"} \times \text{ prior } \propto f(\theta \mid x)}$$

$$\propto f(y \mid x, \theta)\, f(\theta \mid x)$$

"likelihood of y" × "prior" (given data x)

We can first update by x and then y

Prior	Likelihood	Posterior	
$f(\theta)$	$f(x \mid \theta)$	$f(\theta \mid x)$	(added data x)
$f(\theta \mid x)$	$f(y \mid x, \theta)$	$f(\theta \mid x, y)$	(added data x, y)

$$\begin{tikzcd}
f(\theta) \arrow[r, "f(x \mid \theta)", "\text{(add } x\text{)}"'] \arrow[d, "f(y \mid \theta)", "\text{(add } y\text{)}"'] \arrow[dr, "f(x, y \mid \theta)", "\text{(add } x, y\text{)}"'] & f(\theta \mid x) \arrow[d, "f(y \mid \theta, x)", "\text{(add } y\text{)}"'] \\
f(\theta \mid y) \arrow[r, "f(x \mid y, \theta)", "\text{(add } x\text{)}"'] & f(\theta \mid x, y)
\end{tikzcd}$$

Note that if X and Y are conditionally independent given θ $[(X \perp\!\!\!\perp Y) \mid \theta]$ then:

$$f(x,y \mid \theta) = f(x \mid \theta) f(y \mid \theta) \quad [\text{or } f(y \mid x, \theta) = f(y \mid \theta)]$$

Thus,

$$f(\theta \mid x,y) \propto f(y \mid \theta) f(\theta \mid x)$$

$$\propto \underbrace{f(x \mid \theta) f(y \mid \theta)}_{\text{multiply the likelihoods}} f(\theta)$$

[compare with IID observations]

1.4 Conjugate Bayesian update.

Example: Beta - Binomial.

Suppose $X \mid \theta \sim Bin(n, \theta)$ and we wish to specify a prior distribution for θ

Consider $\theta \sim Beta(\alpha, \beta)$ for $\alpha, \beta > 0$ known.

[i.e. We state numerical values for these. We could extend the model further and specify probability distributions for α, β: hierarchical model].

Thus, for $0 \leq \theta \leq 1$,

$$f(\theta) = \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

where the BETA FUNCTION $B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and

$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} \, dt$ is the GAMMA FUNCTION

USE THIS!

Note that $\boxed{\Gamma(z+1) = z\,\Gamma(z)}$ and $\Gamma(1) = 1$.

[Thus, if $z \in \mathbb{Z}^+$ then $\Gamma(z+1) = z!$]

Similarly $\quad B(\alpha, \beta) = \int_0^1 \theta^{\alpha-1} (1-\theta)^{\beta-1}\, d\theta.$

Note: $$\mathbb{E}(\theta) = \int_0^1 \theta \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}\, d\theta$$

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 \theta^{(\alpha+1)-1} (1-\theta)^{\beta-1}\, d\theta$$

$$= \frac{B(\alpha+1, \beta)}{B(\alpha,\beta)}$$

$$= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$= \frac{\alpha\,\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$= \frac{\alpha}{\alpha+\beta}. \qquad \left[\, E(1-\theta) = \frac{\beta}{\alpha+\beta} \,\right]$$