9/2/2021. LECTURE 3.

1.3 SEQUENTIAL DATA ANALYSIS.

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

$$f(\theta \mid x,y) \propto \underbrace{f(x,y \mid \theta)}_{\text{likelihood is of the form data | model}} f(\theta)$$

Now,

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

so that

$$f(\theta \mid x,y) \propto f(y \mid x,\theta) \overbrace{f(x \mid \theta) f(\theta)}^{\text{As a function of } \theta\text{, this is proportional to } f(\theta \mid x)}$$

$$\propto \underbrace{f(y \mid x,\theta)}_{\text{likelihood for the data } y \text{ given observed } x \text{ (and the model)}} \underbrace{f(\theta \mid x)}_{\text{update of } \theta \text{ given } x}$$

[This is thus in the form "Posterior" $\propto$ "Likelihood" $\times$ "Prior"]

We can first update by x and then by y:

Prior	Likelihood	Posterior	
$f(\theta)$	$f(x \mid \theta)$	$f(\theta \mid x)$	Added the data x
$f(\theta \mid x)$	$f(y \mid x,\theta)$	$f(\theta \mid x,y)$	Added the data x,y.

$$\begin{tikzcd}
f(\theta) \arrow[rr, "f(x\mid\theta)", "\text{ADD } x"'] \arrow[d, "\text{Add } y", "f(y\mid\theta)"'] \arrow[drr, "\text{ADD } x,y", "f(x,y\mid\theta)"'] & & f(\theta \mid x) \arrow[d, "\text{Add } y", "f(y\mid\theta,x)"'] \\
f(\theta \mid y) \arrow[rr, "f(x\mid y,\theta)", "\text{Add } x"'] & & f(\theta \mid x,y)
\end{tikzcd}$$

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

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

[or $f(y \mid x, \theta) = f(y \mid \theta)$]

Thus,

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

[analogous to independent observations in the classical model].

1.4 CONJUGATE BAYESIAN UPDATES.

EXAMPLE : BETA - BINOMIAL.

Suppose $X \mid \theta \sim \text{Bin}(n, \theta)$ and we wish to specify a prior distribution for θ. Consider $\theta \sim \text{Beta}(\alpha, \beta)$ for $\alpha, \beta > 0$ known

[i.e. we specify some numerical values for these. An extension would be to treat α, β as random variables and specify prior distributions for those: idea of hierarchical model].

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

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

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

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

Note that $\Gamma(z+1) = z\Gamma(z)$ and $\Gamma(1) = 1$

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

[Note: we will typically utilise the Gamma function for positive reals rather than integers so you should use $\Gamma(z+1) = z\Gamma(z)$ rather than $\Gamma(z+1) = z!$ The former is the general case].

Similarly,

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

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

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

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

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

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

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

Similarly,

$$E(\theta^2) = \int_0^1 \theta^2 \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$