5/03/2022. LECTURE 11: NOTES MADE ON THE VISUALISER.

EXAMPLE. (BERNOULLI LIKELIHOOD).

$$f(x \mid \theta) = \theta^{x}(1-\theta)^{1-x} \qquad x \in \{0,1\}$$
$$= \exp\{x \log \theta + (1-x)\log(1-\theta)\}$$
$$= \exp\left\{\left(\log \frac{\theta}{1-\theta}\right)x + \log(1-\theta)\right\}.$$

A member of the regular 1-parameter exponential family with

$$\phi_1(\theta) = \log\frac{\theta}{1-\theta}, \quad u_1(x) = x, \quad g(\theta) = \log(1-\theta)$$
$$h(x) = 0 = \log 1$$

PROPOSITION.

If $X_1, \dots, X_n$ is an exchangeable sequence such that given a regular k-parameter $EF_k(\cdot)$

$$f(x_1, \dots, x_n) = \int_{\Theta} \left\{\prod_{i=1}^{n} EF_k(x_i \mid g, h, u, \phi, \theta)\right\} f(\theta)$$

[Exchangeable likelihood]

then $t_n = t_n(X_1, \dots, X_n) = \left[n, \sum_{i=1}^{n} u_1(X_i), \dots, \sum_{i=1}^{n} u_k(X_i)\right]$ $n = 1, 2, \dots$ is a sequence of sufficient statistics.

Proof.

$$\prod_{i=1}^{n} EF_k(x_i \mid g, h, u, \phi, \theta) = \prod_{i=1}^{n} \exp\left\{\sum_{j=1}^{k} \phi_j(\theta) u_j(x_i) + 1\, g(\theta) + h(x_i)\right\}$$

[Regular so support doesn't depend on θ]

$$= \exp\left\{\sum_{j=1}^{k} \phi_j(\theta)\left(\sum_{i=1}^{n} u_j(x_i)\right) + n\, g(\theta) + \sum_{i=1}^{n} h(x_i)\right\}$$

$$= \underbrace{\exp\left\{\sum_{j=1}^{k} \phi_j(\theta)\left(\sum_{i=1}^{n} u_j(x_i)\right) + n\, g(\theta)\right\}}_{\tilde{g}(t_n, \theta)} \underbrace{\exp\left\{\sum_{i=1}^{n} h(x_i)\right\}}_{\text{function of } x = (x_1, \dots, x_n) \text{ only}}$$

$$= \tilde{g}(t_n, \theta) \exp\left\{ \sum_{i=1}^{n} h(x_i) \right\}$$

Thus, t_n is sufficient for $X = (X_1, \dots, X_n)$ for learning about θ.

[Sometimes omit explicit reference to n in t_n to focus on "interesting" functions of $x_1, \dots, x_n$. In this case, we have $(k+1)$-sufficient statistics: the sample size n and the k "interesting" functions $\sum_{i=1}^{n} u_j(x_i)$ $j = 1, \dots, k$]

EXAMPLE. (Bernoulli distribution continued).

$$f(x_i | \theta) = \exp\left\{ \left(\log \frac{\theta}{1-\theta} \right) x_i + \log(1-\theta) \right\}$$

If $X_1, \dots, X_n$ is an exchangeable sequence with $X_i | \theta \sim \text{Bernoulli}(\theta)$ then:

$$t_n = \left[n, \sum_{i=1}^{n} u_1(x_i) \right] = \left[n, \sum_{i=1}^{n} x_i \right]$$

is a sequence of sufficient statistics.

2.3.1 Exponential families and conjugate priors.

If $f(x|\theta)$ is a member of a (regular) k-parameter exponential family it is easy to see that a conjugate prior can be found in the $(k+1)$-parameter exponential family over θ.

[Note: $x = (x_1, \dots, x_n)$ possibly].

$$f(x|\theta) = \exp\left\{ \sum_{j=1}^{k} \phi_j(\theta) u_j(x) + g(\theta) + h(x) \right\}$$

Regard as a function of θ

$$= \exp\left\{ \sum_{j=1}^{k} u_j(x) \phi_j(\theta) + 1\, g(\theta) + h(x) \right\}$$

[Exchangeable: $\exp\left\{ \sum_{j=1}^{k} \left(\sum_{i=1}^{n} u_j(x_i) \right) \phi_j(\theta) + n\, g(\theta) + \sum_{i=1}^{n} h(x_i) \right\}$]

$$\propto \exp\left\{ \sum_{j=1}^{k} u_j(x) \phi_j(\theta) + 1\, g(\theta) \right\}$$

Take $f(\theta) = \exp\left\{ \sum_{j=1}^{k} a_j \phi_j(\theta) + d\, g(\theta) + c(a, d) \right\}$

where $c(a,d) = -\log \int_{\Theta} \exp\left\{ \sum_{j=1}^{k} a_j \phi_j(\theta) + d\, g(\theta) \right\} d\theta$ is the normalising constant and $a = (a_1, \dots, a_k)$. Then:

$$f(\theta|x) = \exp\left\{ \sum_{j=1}^{k} \underbrace{(a_j + u_j(x))}_{\tilde{a}_j} \phi_j(\theta) + \underbrace{(d+1)}_{\tilde{d}} g(\theta) + c(\tilde{a}, \tilde{d}) \right\}$$

where $c(\tilde{a}, \tilde{d}) = -\log \int_{\Theta} \exp\left\{ \sum_{j=1}^{k} \tilde{a}_j \phi_j(\theta) + \tilde{d}\, g(\theta) \right\} d\theta$ is the normalising constant, $\tilde{a} = (\tilde{a}_1, \dots, \tilde{a}_k)$ $\tilde{a}_j = a_j + u_j(x)$, $\tilde{d} = d + 1$

Thus, $f(\theta|x)$ and $f(\theta)$ are from the same family giving conjugacy. In this case, with $(k+1)$-parameters, we talk about the NATURAL CONJUGATE PRIOR.

[Likelihood iid $x = (x_1, \dots, x_n)$

$$f(\theta|x) = \exp\left\{ \sum_{j=1}^{k} \underbrace{\left(a_j + \sum_{i=1}^{n} u_j(x_i)\right)}_{\hat{a}_j} \phi_j(\theta) + \underbrace{(d+n)}_{\hat{d}} g(\theta) + c(\hat{a}, \hat{d}) \right\}$$

$$\hat{a}_j = a_j + \sum_{i=1}^{n} u_j(x_i) \qquad \hat{d} = d + n$$

Note the addition of the sufficient statistics and the $(k+1)$-parameters.]

EXAMPLE (Bernoulli concluded).

$$f(x|\theta) = \exp\left\{ \log\left(\frac{\theta}{1-\theta}\right) x + \log(1-\theta) \right\}$$

$$f(\theta) \propto \exp\left\{ a_1 \log\left(\frac{\theta}{1-\theta}\right) + d \log(1-\theta) \right\}$$

$$= \left(\frac{\theta}{1-\theta}\right)^{a_1} (1-\theta)^{d}$$

$$= \theta^{a_1} (1-\theta)^{d-a_1}$$

Take $a_1(\alpha, \beta) = \alpha - 1$ $d(\alpha, \beta) = \beta + \alpha - 2$ to obtain the familiar representation

$$f(\theta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

i.e. $\theta \sim \text{Beta}(\alpha, \beta)$.