17/03/2022. LECTURE 12: NOTES MADE ON THE VISUALISER.

Natural conjugate prior

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

is a member of the $(k+1)$-parameter exponential family over θ. To distinguish these parameters from θ, the parameter of interest, we term them hyperparameters.

e.g. The natural conjugate prior of $X|\theta \sim \text{Bernoulli}(\theta)$ is the $\text{Beta}(\alpha, \beta)$ so α, β are the hyperparameters.

Notice that it is possible to have a $(k+1+s)$-parameter exponential family conjugate prior,

$$\text{e.g. } f(\theta) = \exp\left\{ \sum_{j=1}^{k} a_j \phi_j(\theta) + d\, g(\theta) + \sum_{r=1}^{s} e_r E_r(\theta) + c(a,d,e) \right\}$$

is also conjugate to the k-parameter EF likelihood.

Why are conjugate priors useful?

1). Ease the inferential process given data

Posterior is straightforward to calculate - just revise the hyperparameters.

2) Ease the burden on the prior specification

Reduces to specifying the hyperparameters either directly or by the same number of independent distributional summaries (e.g. means, variances, ...)

Note: in order for a conjugate family to exist, likelihood $\prod_{i=1}^{n} f(x_i|\theta)$ must only include a finite number of different functions of $x=(x_1,\dots,x_n)$ for n arbitrarily large.

i.e. must generate a finite number of sufficient statistics (fixed as $n \to \infty$)

i.e. typically, the likelihood is a member of a regular exponential family

i.e. typically, only regular EFs exhibit conjugacy.

[Pitman - Koopman - Darmois Theorem]

[Uniform with a Pareto prior is a notable exception: Question Sheet Three, Q4. Natural generalisation of conjugate priors to mixtures of EFs can be done - Question 2e of 2020/1 exam. $f(\theta) = \sum_{i=1}^{m} \alpha_i f_i(\theta)$ where each $f_i(\theta)$ is a member of the conjugate family].

2.4 Noninformative prior distributions.

In many situations, there is a desire for a prior distribution that is guaranteed to play a minimal role in the posterior.

The prior density is "flat" or "noninformative" to "let the data speak for themselves".

One idea is to represent ignorance using a uniform prior distribution,

$$f(\theta) \propto 1$$

so that $f(\theta|x) \propto f(x|\theta)$. However, if the support of θ is not finite then is no such proper distribution. This might not be a problem if the scaled (or standardised) likelihood has a finite integral over θ,

$$f(\theta|x) = \frac{f(x|\theta)}{\underbrace{\int_\theta f(x|\theta)\,d\theta}_{\text{No guarantee that this is finite.}}}$$

DEFINITION (Improper prior).
The specification $f(\theta) \propto f^*(\theta)$ is said to be an IMPROPER PRIOR if

$$\int_\theta f^*(\theta)\,d\theta = \infty$$

A major drawback to representing ignorance by a default uniform prior is that this is NOT invariant to the choice of parameterisation.
If I choose $f_\theta(\theta) \propto 1$ then if $\lambda = \frac{1}{\theta}$ we have:

$$f_\lambda(\lambda) = \left|\frac{\partial\theta}{\partial\lambda}\right| f_\theta\left(\frac{1}{\lambda}\right) \propto \left|-\frac{1}{\lambda^2}\right| \times 1 = \frac{1}{\lambda^2}.$$

which is not (improper) uniform [e.g. θ is the variance, $\lambda = 1/\theta$ is the precision] A prior that is invariant to the choice of parametrisation is the JEFFREYS PRIOR.

2.4.1 THE JEFFREYS PRIOR.

The Jeffreys prior for a parameter θ relative to a likelihood $f(x|\theta)$ is

$$f(\theta) \propto \{\det I(\theta)\}^{1/2}$$

where $I(\theta)$ is the FISHER INFORMATION MATRIX and $\det I(\theta)$ is the determinant of $I(\theta)$. [Sometimes written $|I(\theta)|$]

Thus, if $\theta = (\theta_1, \dots, \theta_k)$ is a k-dimensional parameter then $I(\theta)$ is the $(k \times k)$-matrix with (i,j)th entry:

$$I(\theta)_{ij} = E\left[\frac{\partial}{\partial\theta_i}\log f(x|\theta)\frac{\partial}{\partial\theta_j}\log f(x|\theta)\,\middle|\,\theta\right]$$

$$= -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(x|\theta)\,\middle|\,\theta\right]$$

In most cases, we will have $k=1$ so that $I(\theta)$ is a scalar

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\,\middle|\,\theta\right] \leftarrow$$

This tends to be useful for proving theoretical results

$$= -E\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\,\middle|\,\theta\right]$$

so that $f(\theta) \propto \sqrt{I(\theta)}$.

EXAMPLE.

Suppose that $X_1,\dots,X_n$ are exchangeable with $X_i|\theta \sim N(\theta,\sigma^2)$ where σ^2 is known. Find the Jeffreys prior for θ.

$$f(x|\theta) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\theta)^2\right\}$$

where $x=(x_1,\dots,x_n)$

$$\frac{\partial^2}{\partial\theta^2}\log f(x|\theta) = \frac{\partial^2}{\partial\theta^2}\left\{-\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\theta)^2\right\}$$

$$= \frac{\partial}{\partial\theta}\left\{\frac{1}{\sigma^2}\sum_{i=1}^n(x_i-\theta)\right\} = -\frac{n}{\sigma^2}.$$

Thus, $I(\theta) = -E\left(-\frac{n}{\sigma^2}\,\middle|\,\theta\right) = \frac{n}{\sigma^2}$

The Jeffreys prior is thus $f(\theta) \propto \sqrt{\frac{n}{\sigma^2}} \propto 1$.

Note that this is improper but $\theta | x \sim N(\bar{x}, \sigma^2/n)$ which is proper.

We can think of $f(\theta) \propto 1$ as being

$$\lim_{\sigma_0^2 \to \infty} N(\mu_0, \sigma_0^2), \text{ the conjugate prior.}$$