24/03/2022. LECTURE 14: NOTES MADE ON THE VISUALISER.

The use of an improper prior may be a convenient tool to minimise the role of the prior. However, the posterior is not always guaranteed to be proper.

EXAMPLE.

Consider $X \mid \theta \sim \text{Bin}(n, \theta)$ and $f(\theta) \propto \theta^{-1}(1-\theta)^{-1}$

(Haldane's Prior)

Interpret this as the improper "Beta(0,0)"

Thus, $f(\theta \mid x) \propto \theta^{x}(1-\theta)^{n-x}\,\theta^{-1}(1-\theta)^{-1}$

$$= \theta^{x-1}(1-\theta)^{n-x-1}$$

i.e. $\theta \mid x \sim \text{Beta}(x, n-x)$

Note that $E(\theta \mid X) = \frac{x}{n}$, the maximum likelihood estimate.

(which motivates the prior being noninformative). However, if $x = 0$ or $n = 1$ then the posterior is improper.

3. COMPUTATION.

Practical implementation of Bayesian methods requires substantial computation. This is required, essentially, to compute summaries of the posterior distribution

So far, we have worked with tractable (e.g. conjugate) models but in reality we need to work with distributions that are:

a) Mathematically complex
b) Often high dimensional

Interested in summaries of the posterior

$$f(\theta|x) = c\, g(\theta)$$

where $g(\theta) \propto f(x|\theta) f(\theta)$ and c is the normalising constant (or constant of integration). In many cases, c may be difficult to compute as the required integral cannot be carried out analytically.

EXAMPLE.

Suppose $X|\theta \sim N(\theta, \sigma^2)$ where σ^2 is known. The natural conjugate prior is $\theta \sim N(\mu_0, \sigma_0^2)$. Suppose instead that $\theta \sim t_\nu(\mu_0, \sigma_0^2)$, the (non-central) t-distribution with ν degrees of freedom. [similar shape to the normal but with heavier tails]

$$f(\theta) \propto \left(1 + \frac{1}{\nu}\left(\frac{\theta - \mu_0}{\sigma_0}\right)^2\right)^{-\frac{(\nu+1)}{2}}$$

[Example in the on-line notes has $t_\nu(0,1)$]

$$f(\theta|x) = c \exp\left\{-\frac{1}{2\sigma^2}(\theta^2 - 2x\theta)\right\}\left(1 + \frac{1}{\nu}\left(\frac{\theta - \mu_0}{\sigma_0}\right)^2\right)^{-\frac{(\nu+1)}{2}}$$

$$= c\, g(\theta)$$

We cannot find $\int_{-\infty}^{\infty} g(\theta)\, d\theta$ analytically. We need computational techniques to proceed. Many of these compute summaries of $f(\theta|x)$ without requiring c.

3.1 Normal Approximations.

We consider a normal approximation to the posterior, utilising a Taylor series expansion of $\log f(\theta|x)$ around the mode $\tilde{\theta}$

[Mode of $\log f(\theta|x)$ is the same as the mode of $f(\theta|x)$].

Suppose $f(\theta|x)$ is unimodal and roughly symmetric. For simplicity, we'll assume θ is univariate.

Recall the Taylor series expansion of $h(\theta)$ around θ_0

$$h(\theta) = h(\theta_0) + (\theta - \theta_0)\frac{d}{d\theta}h(\theta)\Big|_{\theta=\theta_0} + \dots + \frac{(\theta-\theta_0)^r}{r!}\frac{d^r}{d\theta^r}h(\theta)\Big|_{\theta=\theta_0} + \dots$$

Let $\log f(\theta|x) = h(\theta)$ and $\theta_0 = \tilde{\theta}$ (the mode). We have:

As $\tilde{\theta}$ is the mode of $f(\theta|x)$, is the mode of $\log f(\theta|x)$

$$\log f(\theta|x) = \log f(\tilde{\theta}|x) + (\theta - \tilde{\theta})\frac{d}{d\theta}\log f(\theta|x)\Big|_{\theta=\tilde{\theta}}$$

then $\frac{d}{d\theta}\log f(\theta|x)\Big|_{\theta=\tilde{\theta}} = 0$

$$+ \frac{(\theta-\tilde{\theta})^2}{2}\frac{d^2}{d\theta^2}\log f(\theta|x)\Big|_{\theta=\tilde{\theta}} + \dots$$

So, approximately,

$$\log f(\theta|x) = \log f(\tilde{\theta}|x) + \frac{(\theta-\tilde{\theta})^2}{2}\frac{d^2}{d\theta^2}\log f(\theta|x)\Big|_{\theta=\tilde{\theta}}$$

Let $I(\theta|x) = \cancel{-\frac{d^2}{d\theta^2}\log f(\theta|x)}$, the OBSERVED INFORMATION

$$= -\frac{d^2}{d\theta^2}\log f(\theta|x)$$

[$\tilde{\theta}$ is a maximum of $\log f(\theta|x)$ so $I(\tilde{\theta}|x)$ is non-negative]

Thus, approximately,

$$\log f(\theta|x) = \log f(\tilde{\theta}|x) - \frac{(\theta-\tilde{\theta})^2}{2} I(\tilde{\theta}|x)$$

↑ Observed information evaluated at $\tilde{\theta}$

$$\Rightarrow f(\theta|x) \propto \exp\left(-\frac{I(\tilde{\theta}|x)}{2}(\theta-\tilde{\theta})^2\right)$$

i.e. Approximately, $\theta|x \sim N(\tilde{\theta}, I^{-1}(\tilde{\theta}|x))$

This approximation can be generalised to the case when θ is multivariate.

Note: If we have $f(\theta|x) = c\, g(\theta)$ then $\log f(\theta|x) = \log c + \log g(\theta)$

so that $\dfrac{d^r}{d\theta^r} \log f(\theta|x) = \dfrac{d^r}{d\theta^r} \log g(\theta)$ for $r \geq 1$

Therefore, we don't need to know c for this approximation.

EXAMPLE.

Let $X = (X_1, \dots, X_n)$ where the X_is are exchangeable with $X_i | \lambda \sim Po(\lambda)$. If $f(\lambda) \propto 1$, find the normal approximation about the mode to the posterior.

$$f(\lambda|x) \propto f(x|\lambda) f(\lambda) \qquad x = (x_1, \dots, x_n)$$

$$= \prod_{i=1}^{n} \lambda^{x_i} e^{-\lambda} = \lambda^{n\bar{x}} e^{-n\lambda}$$

We recognise this as a kernel of a Gamma distribution

$$\lambda | x \sim Gamma(n\bar{x}+1, n)$$

The mode of $Gamma(\alpha, \beta)$ is $\dfrac{\alpha - 1}{\beta}$ (for familiar families, mode is on the formula sheet)

Thus, for $\lambda|x$, the mode is $\bar{x}$. Now,

$$\frac{d^2}{d\lambda^2} \log f(\lambda|x) = \frac{d^2}{d\lambda^2} \left\{ n\bar{x} \log \lambda - n\lambda \right\}$$

$$= \frac{d}{d\lambda} \left\{ \frac{n\bar{x}}{\lambda} - n \right\}$$

verify that $\tilde{\lambda} = \bar{x}$

$$= -\frac{n\bar{x}}{\lambda^2}.$$

The observed information at $\tilde{\lambda} = \bar{x}$ is

$$I(\tilde{\lambda} \mid \lambda) = I(\bar{x} \mid \lambda) = -\frac{d^2}{d\lambda^2} \log f(\lambda \mid x) \Big|_{\lambda = \bar{x}}$$

$$= \frac{n}{\bar{x}}.$$

Thus, approximately, $\lambda \mid x \sim N(\bar{x}, \bar{x}/n)$

sanity check: variance decreases in n.