22/02/2022. LECTURE 5: NOTES MADE ON THE VISUALISER.

Last time: θ_c probability of successful coin toss, $\theta_c \sim \text{Beta}(20,20)$

θ_p probability of pin-up on a single toss, $\theta_p \sim \text{Beta}(2,2)$

In each case, suppose I observe $x = 4$ in $n = 10$ tosses (mle is 0.4). Then:

$$\theta_c \mid x \sim \text{Beta}(20+4, 20+\overset{6}{\cancel{10}}) \sim \text{Beta}(24, 26)$$
$$\theta_p \mid x \sim \text{Beta}(2+4, 2+6) \sim \text{Beta}(6, 8)$$

We plot the prior, posterior and standardised likelihood,

$$\frac{f(x \mid \theta)}{\int_a f(x \mid a)\, da}$$

[i.e. scaled to integrate to 1 over θ]

[In this case, $f(x \mid \theta) \propto \theta^4 (1-\theta)^6 = \theta^{5-1}(1-\theta)^{7-1}$ and so the standardized likelihood has the same curve as Beta(5,7)].

θ_c - Strong prior. (see figure 1 at the end)

Posterior density closely follows the prior. The mode is shifted to the left towards the likelihood. Posterior density (slightly) taller and narrower: variance reduced.

$$\mathbb{E}(\theta_c) = 0.5 \qquad \mathbb{E}(\theta_c \mid X=4) = \frac{24}{24+26} = 0.48$$

$$\text{Var}(\theta_c) = 0.006 \qquad \text{Var}(\theta_c \mid X=4) = 0.005$$

θ_p - WEAK PRIOR. (see figure 2 at the end)

Posterior density much more follows the (standardised) likelihood. Most of the posterior information is coming from the likelihood i.e. the data.

Variance of the posterior is vastly reduced (in relation to the prior).

$[\ \mathbb{E}(\theta_p) = 0.5,\quad \mathbb{E}(\theta_p \mid X=4) = 0.43$

$\mathrm{Var}(\theta_p) = 0.05,\quad \mathrm{Var}(\theta_p \mid X=4) = 0.016\ ]$

DEFINITION (KERNEL OF A DENSITY).

For a RV X with density $f(x)$ if $f(x)$ can be written in the form $cq(x)$ where c is a constant not depending upon x then any such $q(x)$ is a KERNEL of the density $f(x)$.

EXAMPLE.

$\theta \sim \mathrm{Beta}(\alpha, \beta)$ $\quad q(\theta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}$ is a kernel.

Spotting kernels of distributions can be very useful in computing posterior distributions.

EXAMPLE.

Let $X \mid \theta \sim N(\theta, \sigma^2)$ so that the mean θ is unknown and the variance σ^2 is known. Suppose $\theta \sim N(\mu_0, \sigma_0^2)$ where μ_0 and σ_0^2 are known. Find the posterior distribution of θ given $X = x$.

$$f(\theta) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left\{ -\frac{1}{2\sigma_0^2} (\theta - \mu_0)^2 \right\}$$

$$\underset{\text{wrt } \theta}{\propto} \exp\left\{ -\frac{1}{2\sigma_0^2} (\theta - \mu_0)^2 \right\} \quad \text{(a kernel of } N(\mu_0, \sigma_0^2))$$

$$\propto \exp\left\{ -\frac{1}{2\sigma_0^2} (\theta^2 - 2\mu_0\theta) \right\} \quad \text{(kernel in its simplest form)}$$

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{1}{2\sigma^2} (x - \theta)^2 \right\}$$

$$\underset{\text{wrt } \theta}{\propto} \exp\left\{ -\frac{1}{2\sigma^2} (\theta^2 - 2x\theta) \right\}$$

Viewed as a function of θ, this looks like a normal kernel - a kernel of $N(x, \sigma^2)$

$$f(\theta|x) \propto f(x|\theta) f(\theta)$$

$$\propto \exp\left\{ -\frac{1}{2\sigma^2} (\theta^2 - 2x\theta) \right\} \exp\left\{ -\frac{1}{2\sigma_0^2} (\theta^2 - 2\mu_0\theta) \right\}$$

$$= \exp\left\{ -\frac{1}{2} \left[\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \right] \left(\theta^2 - 2 \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \right)^{-1} \left(\frac{\mu_0}{\sigma_0^2} + \frac{x}{\sigma^2} \right) \theta \right) \right\}$$

KERNEL OF A NORMAL.

I recognise this as a kernel of a Normal distribution and so

$$\theta|x \sim N(\mu_1, \sigma_1^2)$$

where: $\sigma_1^2 = \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2} \right)^{-1}$ i.e. $\frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}$

[Precision is the inverse of the variance]. Posterior Precision = Prior Precision + Data Precision

$$\mu_1 = \left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}\right)^{-1}\left(\frac{\mu_0}{\sigma_0^2} + \frac{x}{\sigma^2}\right)$$

$$= \alpha\mu_0 + (1-\alpha)x \quad \text{when} \quad \alpha = \frac{1/\sigma_0^2}{1/\sigma_0^2 + 1/\sigma^2}$$

The posterior mean is a WEIGHTED AVERAGE of the prior mean, μ_0, and the data (mle) x weighted according to their corresponding precisions.

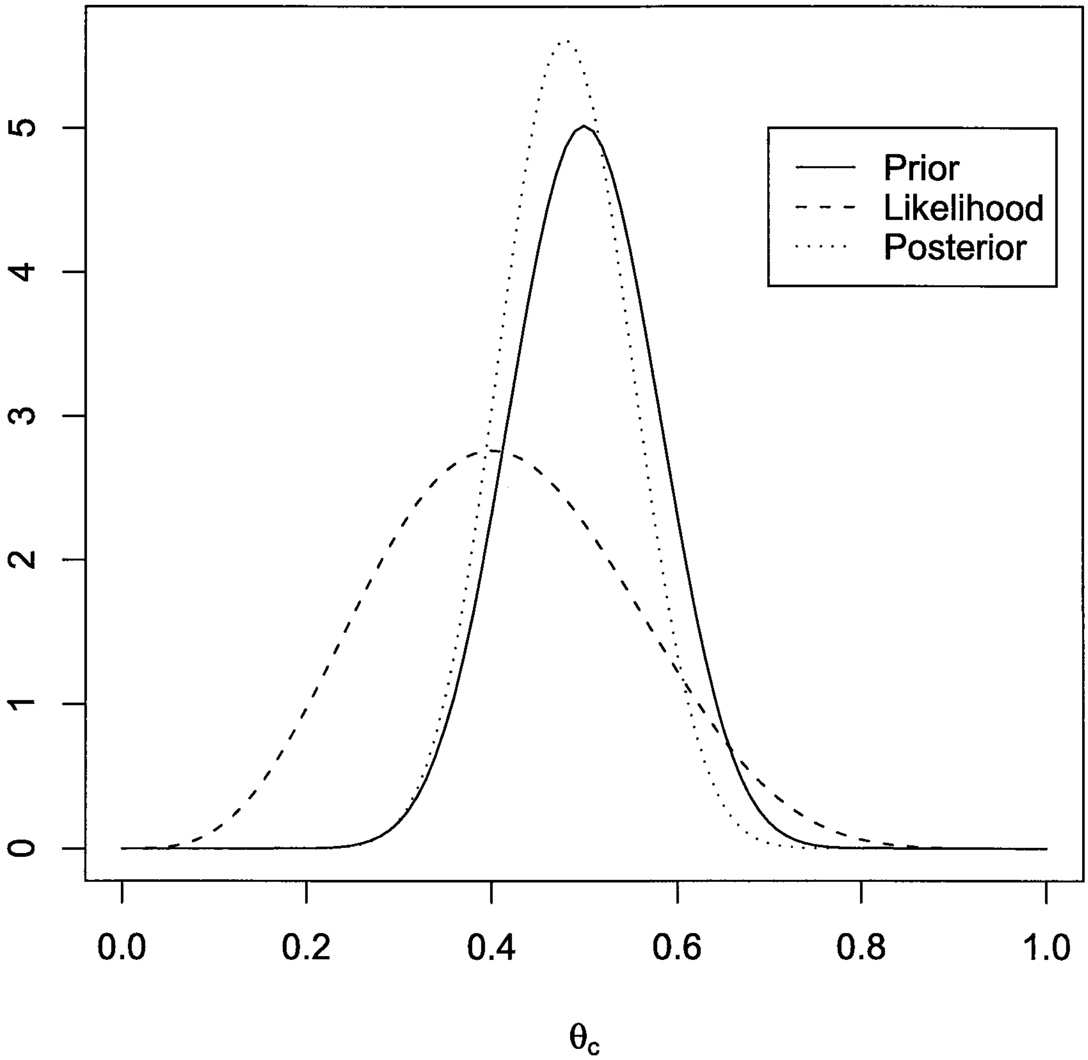

Figure 1: Prior to posterior for Bernoulli trials model, $n = 10$, $x = 4$. The likelihoods are standardised. Tossing coins. Prior $\theta_c \sim Beta(20, 20)$, posterior $\theta_c \,|\, x \sim Beta(24, 26)$.

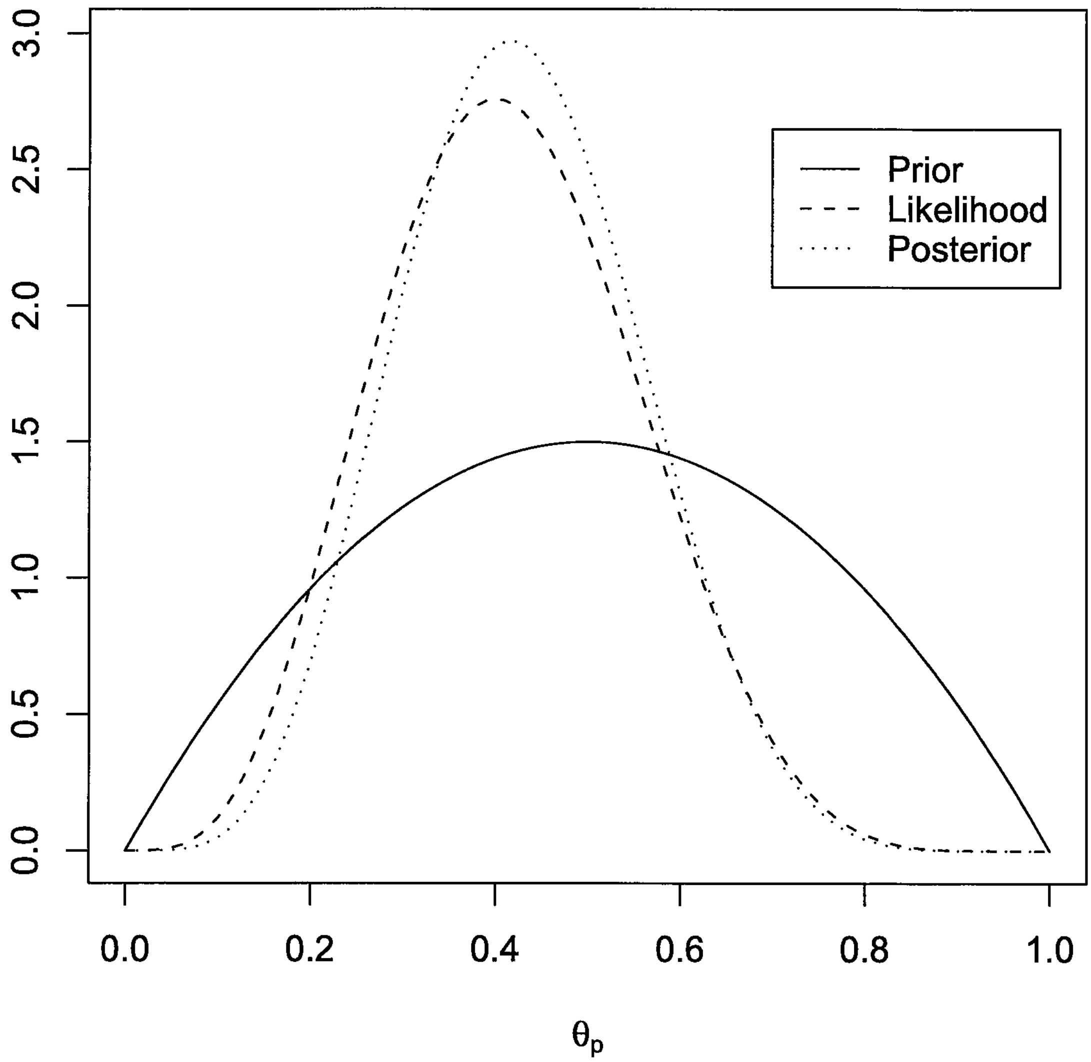

Figure 2: Prior to posterior for Bernoulli trials model, $n = 10$, $x = 4$. The likelihoods are standardised. Tossing drawing pins. Prior $\theta_p \sim Beta(2,2)$, posterior $\theta_p \,|\, x \sim Beta(6,8)$.