# Example Class 6

Meng Xuran

Department of Statistics and Actuarial Science,
The University of Hong Kong

October 21, 2022

# Jeffreys Prior

- The Jeffreys Prior is proportional to the square root of the determinant of the Fisher information matrix

$$f(\theta) \propto |I(\theta)|^{\frac{1}{2}}$$

$$I(\theta) = -E\left\{\frac{\partial^2 \log L(\theta|y)}{\partial \theta^2}\right\}$$

- It is invariant to re-parametrization.
- That is, the relative probability assigned to a volume of a probability space using a Jeffreys prior will be the same regardless of the parameterization used to define the Jeffreys prior.

# Jeffreys Prior (Parameterization)

- For a monotone variable transformation $\eta = h(\theta), \theta = h^{-1}(\eta)$
- Note that

$$I(\eta) = -\mathbb{E}\left(\frac{\partial^2 \log f(x|\eta)}{\partial \eta^2}\right) = -\mathbb{E}\left[\frac{\frac{\partial^2}{\partial \eta^2} f(x|\eta)}{f(x|\eta)} - \left(\frac{\frac{\partial}{\partial \eta} f(x|\eta)}{f(x|\eta)}\right)^2\right] = \mathbb{E}\left(\frac{\partial \log f(x|\eta)}{\partial \eta}\right.$$

where

$$\mathbb{E}\left(\frac{\frac{\partial^2}{\partial \eta^2} f(x|\eta)}{f(x|\eta)}\right) = \int \frac{\frac{\partial^2}{\partial \eta^2} f(x|\eta)}{f(x|\eta)} f(x|\eta) dx = \frac{\partial^2}{\partial \eta^2} \int f(x|\eta) dx = 0.$$

- So we have

$$\begin{aligned} |I(\eta)| &= \left|\mathbb{E}\left(\frac{\partial^2 \log f(x|\eta)}{\partial \eta^2}\right)\right| \\ &= \mathbb{E}\left(\frac{\partial \log f(x|h(\theta))}{\partial \eta}\right)^2 = \mathbb{E}\left(\frac{\partial \log f(x|h(\theta))}{\partial \theta} \frac{\partial \theta}{\partial \eta}\right)^2 \\ &= |I(\theta)| \left|\frac{\partial \theta}{\partial \eta}\right|^2. \end{aligned}$$

- We thus have $|I(\eta)|^{1/2} = |I(\theta)|^{1/2} \left|\frac{\partial \theta}{\partial \eta}\right|$.

# How to Borrow Information from Historical Data

- Power Prior ($\gamma$ is the power parameter):

$$\pi(\theta|D_0,\gamma) \propto \underbrace{L(\theta|D_0)^{\gamma}}_{\substack{\text{discounted}\\ \text{historical}\\ \text{information}}} \quad \underbrace{\pi(\theta)}_{\substack{\text{prior information}\\ \text{unrelated to}\\ \text{historical data}}}$$

- Posterior:

$$f(\theta|D_0, D, \gamma) \propto L(\theta|D)\pi(\theta|D_0,\gamma) \propto L(\theta|D)L(\theta|D_0)^{\gamma}\pi(\theta)$$

- $\gamma$ weights the historical data $D_0$ relative to the current data $D$
- With a flexible $\gamma$,

$$\pi(\theta,\gamma|D_0) \propto \frac{L(\theta|D_0)^{\gamma}\pi(\theta)}{\int L(\theta|D_0)^{\gamma}\pi(\theta)d\theta}\pi(\gamma)$$

# Example (Normal Case)

- Historical data $\boldsymbol{x}_0 = (x_{01}, \ldots, x_{0n_0})'$
- Current data $\boldsymbol{x} = (x_1, \ldots, x_n)'$, $x_i \overset{i.i.d}{\sim} N(\mu, \sigma^2)$
- For simplicity, we assume that the variance $\sigma^2$ is known.
- Power prior with fixed $\gamma$

$$\pi^{PP}(\mu|D_0, \gamma) \propto L(\mu|D_0)^{\gamma}\pi_0(\mu)$$

- With a flat prior $\pi_0(\mu) \propto 1$,

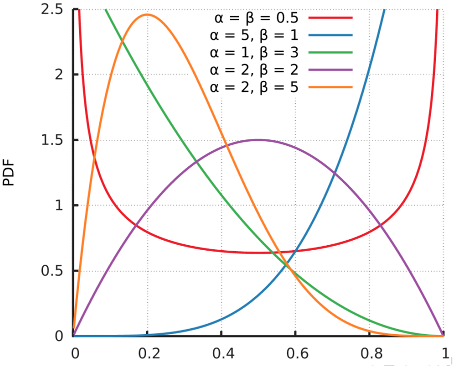
$$\pi^{PP}(\mu|D_0, \gamma) \propto e^{-\frac{\gamma \sum_{i=1}^{n_0}(x_{0i}-\mu)^2}{2\sigma^2}} \sim N\left(\bar{x}_0, \frac{\sigma^2}{n_0\gamma}\right)$$

- The posterior distribution has the form,

$$p^{PP}(\mu|D, D_0, \gamma) \propto e^{-\frac{\gamma \sum_{i=1}^{n_0}(x_{0i}-\mu)^2}{2\sigma^2}} e^{-\frac{\sum_{i=1}^{n_0}(x_i-\mu)^2}{2\sigma^2}} \sim N\left(\frac{\gamma n_0\bar{x}_0 + n\bar{x}}{\gamma n_0 + n}, \frac{\sigma^2}{n_0\gamma + n}\right)$$

# How to Borrow Information from Historical Data

- If we specify $\pi(\gamma)$ as a Beta$(a, b)$ distribution for fixed positive hyperparameters $a$ and $b$
- $(a = 5, b = 1)$ would strongly encourage borrowing
- $(a = 1, b = 5)$ would strongly discourage it
- $(a = b = 1)$ would be agnostic on the subject, essentially letting the data determine the degree of borrowing.



# Example (Normal Case)

- Modified power prior (flexible $\gamma$) Page 4

$$\pi^{MPP}(\mu, \gamma | D_0) \propto \frac{L(\mu|D_0)^{\gamma}\pi(\mu)}{\int L(\mu|D_0)^{\gamma}\pi(\mu)d\mu}\pi(\gamma)$$

- With a Beta(a,b) prior on $\gamma$,

$$\pi^{MPP}(\mu, \gamma | D_0) \propto \frac{e^{-\frac{(\mu-\bar{x}_0)^2}{2\sigma^2/(n_0\gamma)}}}{\int e^{-\frac{(\mu-\bar{x}_0)^2}{2\sigma^2/(n_0\gamma)}}d\mu}\pi(\gamma) \propto \sqrt{\gamma}e^{-\frac{n_0\gamma(\mu-\bar{x}_0)^2}{2\sigma^2}}\gamma^{a-1}(1-\gamma)^{b-1}$$

$$\propto N\left(\mu|\bar{x}_0, \frac{\sigma^2}{n_0\gamma}\right)\text{Beta}(\gamma|a,b)$$

- Posterior

$$p^{MPP}(\mu, \gamma | D_0, D) \propto \pi^{MPP}(\mu, \gamma | D_0)L(\mu|D)$$

# Example (Normal Case)

Posterior

$$p^{MPP}(\mu,\gamma|D_0,D) \propto \pi^{MPP}(\mu,\gamma|D_0)L(\mu|D)$$

$$\begin{aligned}
p^{MPP}(\mu,\gamma|D_0,D) &\propto \gamma^{1/2} e^{-\frac{n_0\gamma(u-\bar{x}_0)}{2\sigma^2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} \gamma^{a-1}(1-\gamma)^{b-1} \\
&\propto \gamma^{1/2}\gamma^{a-1}(1-\gamma)^{b-1} e^{-\left[\frac{\left(u-\frac{n_0\gamma\bar{x}_0+n\bar{x}}{n_0\gamma+n}\right)^2}{2\sigma^2/(n_0\gamma+n)} + \frac{n_0\gamma\bar{x}_0^2-\frac{(n_0\gamma\bar{x}_0+n\bar{x})^2}{n_0\gamma+n}}{2\sigma^2}\right]} \\
&\propto \gamma^{1/2}\gamma^{a-1}(1-\gamma)^{b-1} e^{-\frac{\left(u-\frac{n_0\gamma\bar{x}_0+n\bar{x}}{n_0\gamma+n}\right)^2}{2\sigma^2/(n_0\gamma+n)}} e^{-\frac{nn_0\gamma(\bar{x}-\bar{x}_0)^2}{2\sigma^2(n+n_0\gamma)}} \\
p^{MPP}(\mu|D_0,D,\gamma) &\sim N\left(\frac{\gamma n_0\bar{x}_0+n\bar{x}}{\gamma n_0+n}, \frac{\sigma^2}{n_0\gamma+n}\right) \\
p^{MPP}(\gamma|D_0,D) &\sim N\left(\bar{x}|\bar{x}_0, \frac{\sigma^2(n+n_0\gamma)}{nn_0\gamma}\right)\mathrm{Beta}(\gamma|a,b)
\end{aligned}$$

# Commensurate Power Prior

- The hierachical power prior do not directly parametrize the commensurability of the historical and new data since the posterior of $\gamma$ is free of the current data $D$.
- Commensurate Power Prior: assume different parameters in the historical and current data.

$$\pi(\theta, \gamma, \tau | D_0) \propto \int \frac{L(\theta_0|D_0)^{\gamma}\pi(\theta_0)}{\int L(\theta_1|D_0)^{\gamma}\pi(\theta_1)d\theta_1}\pi(\theta|\theta_0, \tau)d\theta_0 \cdot \pi(\gamma|\tau)\pi(\tau)$$

- The prior of $\theta$ has the mean $\theta_0$ and precision $\tau$ and the power parameter $\gamma$ depends on $\tau$. Thus, $\tau$ parameterizes commensurability between the historical and current data.
- For example, $\theta \sim N(\theta_0, 1/\tau), \gamma \sim \text{Beta}(g(\tau), 1)$

# Example (Normal Case)

- Commensurate Power Prior:
  $\pi(\mu_0) \propto 1, \pi(\mu|\mu_0, \tau) \sim N(\mu_0, 1/\tau), \pi(\gamma|\tau) \sim \text{Beta}(g(\tau), 1)$

$$\pi^{CPP}(\mu, \gamma, \tau|D_0) \propto \int \frac{L(\mu_0|D_0)^\gamma \pi(\mu_0)}{\int L(\mu_0|D_0)^\gamma \pi(\mu_0)} \pi(\mu|\mu_0, \tau) d\mu_0 \cdot \pi(\gamma|\tau)\pi(\tau)$$

$$\sim N\left(\mu|\bar{x}_0, \frac{1}{\tau} + \frac{\sigma^2}{\gamma n_0}\right) \text{Beta}(g(\tau), 1)\pi(\tau)$$

$$p^{CPP}(\mu|\gamma, \tau, D_0, D) \sim N\left(\frac{\gamma n_0 \tau \sigma^2 \bar{x}_0 + n(\gamma n_0 + \sigma^2 \tau)\bar{x}}{\gamma n_0 \tau \sigma^2 + n(\gamma n_0 + \sigma^2 \tau)}, \frac{(\gamma n_0 + \sigma^2 \tau)\sigma^2}{\gamma n_0 \tau \sigma^2 + n(\gamma n_0 + \sigma^2 \tau)}\right)$$

$$p^{CPP}(\gamma, \tau|D, D_0) \sim N\left(\bar{x}|\bar{x}_0, \frac{\sigma^2}{n} + \frac{1}{\tau} + \frac{\sigma^2}{\gamma n_0}\right) \text{Beta}(g(\tau), 1)\pi(\tau)$$

# Hierarchical Commensurate Prior

- General case of commensurate priors: weighting the influence of prior information relative to its consistency with data from the concurrent study.

$$\pi(\theta|\theta_0, D_0, \tau) \propto L(\theta_0|D_0)p(\theta|\theta_0, \tau)\pi_0(\theta)$$

- Note that the full conditional posterior distribution for $\theta_0$ would be independent of the current data. Thus, $\theta_0$ should be integrated out of the prior when the $\int L(\theta_0|D_0)p(\theta|\theta_0, \tau)d\theta_0$ is tractable.
- When $\tau \to 0$, $\pi(\theta|\theta_0, D_0, \tau) \to \pi_0(\theta)$.
- When $\tau \to \infty$, $\pi(\theta|\theta_0, D_0, \tau) \to L(\theta|D_0)\pi_0(\theta)$

# Example (Normal Case)

- Hierarchical Commensurate Prior: consider $\pi_0(\mu) \propto 1$

$$\pi^{HCP}(\mu, \tau | D_0) = \int L(\mu_0 | D_0) \pi(\mu | \mu_0, \tau) d\mu_0 \pi_0(\mu) \pi(\tau)$$
$$\propto N\left(\mu | \bar{x}_0, \frac{1}{\tau} + \frac{\sigma^2}{n_0}\right) \pi(\tau)$$

# Example (Binary Case)

- Historical data: $y_0, n_0$
- Current data: $y, n, y \sim \text{Bin}(n, \theta)$
- Power prior with a fixed $\gamma$ and a Beta$(a, b)$ prior on $\theta$:

$$\begin{aligned}\pi^{PP}(\theta|D_0, \gamma) \propto L(\theta|D_0)^\gamma \pi_0(\theta) &\propto \theta^{\gamma y_0}(1-\theta)^{\gamma(n_0-y_0)}\theta^{a-1}(1-\theta)^{b-1} \\ &\sim \text{Beta}(a+\gamma y_0, b+\gamma(n_0-y_0))\end{aligned}$$

- Posterior:

$$p^{PP}(\theta|D, D_0, \gamma) \sim \text{Beta}(a+y+\gamma y_0, b+(n-y)+\gamma(n_0-y_0))$$

# Example (Binary Case w/ Modified Power Prior)

- Modified power prior (with a Beta$(c, d)$ on $\gamma$):

$$
\begin{aligned}
\pi^{MPP}(\theta, \gamma | D_0) \propto & \frac{\Gamma(a+b+\gamma n_0)}{\Gamma(a+\gamma y_0)\Gamma(b+\gamma(n_0-y_0))} \theta^{\gamma y_0 + a - 1}(1-\theta)^{\gamma(n_0-y_0)+b-1} \\
& \gamma^{c-1}(1-\gamma)^{d-1} \\
& \propto \text{Beta}(\theta | a+\gamma y_0, b+\gamma(n_0-y_0))\text{Beta}(\gamma | c, d)
\end{aligned}
$$

- Posterior:

$$
\begin{aligned}
p^{MPP}(\theta | D, D_0, \gamma) &\sim \text{Beta}(a+\gamma y_0 + y, b+\gamma(n_0-y_0)+n_0) \\
p^{MPP}(\gamma | D, D_0) &\propto \frac{\Gamma(a+b+\gamma n_0)}{\Gamma(a+\gamma y_0)\Gamma(b+\gamma(n_0-y_0))} \\
& \cdot \frac{\Gamma(a+\gamma y_0+y)\Gamma(b+\gamma(n_0-y_0)+(n-y))}{\Gamma(a+b+\gamma n_0+n)} \\
& \cdot \text{Beta}(\gamma | c, d)
\end{aligned}
$$