STAT 713 sp 2023 Lec 05 slides

Bayes estimators

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Suppose we observe data $\mathbf{X}$ with distribution F depending on $\theta \in \Theta$.

Two paradigms:

- *Frequentist*: Treat θ as a fixed constant and write $\mathbf{X} \sim F(\mathbf{X}; \theta)$.
- *Bayesian*: Treat θ as a random variable and write $\mathbf{X}|\theta \sim F(\mathbf{X}|\theta)$.

The hierarchical model

Let $\mathbf{X} = (X_1, \dots, X_n)$ be our data and let θ be a rv in $\Theta \subset \mathbb{R}$. Assume

$$\begin{aligned} \mathbf{X}|\theta &\sim f(\mathbf{x}|\theta) \\ \theta &\sim \pi(\theta), \end{aligned}$$

where

- $f(\cdot|\theta)$ is the joint pmf/pdf of $\mathbf{X}$, conditional on θ (*data distribution*).
- $\pi(\cdot)$ is the marginal pmf/pdf of the parameter θ (*prior distribution*).

$\pi(\theta)$ prior — belief about θ before observing $\underset{\sim}{X}$.

$f(\underset{\sim}{x};\theta)$ data dist

$\pi(\theta|\underset{\sim}{X})$: posterior — beliefs about θ after observing $\underset{\sim}{X}$

Posterior distribution

The *posterior distribution* of θ is the distribution of θ conditional on the data $\mathbf{X}$.

If $f(\mathbf{X}|\theta)$ is the joint pdf/pmf of $\mathbf{X}|\theta$, the conditional pdf/pmf of $\theta|\mathbf{X}$ is given by

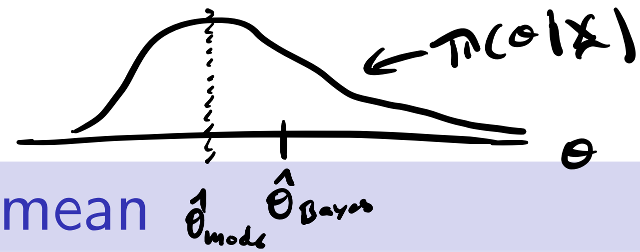
$$\pi(\theta|\mathbf{X}) = \begin{cases} \dfrac{f(\mathbf{X}|\theta)\pi(\theta)}{\int_\Theta f(\mathbf{X}|\tilde{\theta})\pi(\tilde{\theta})d\tilde{\theta}} & \text{if } \theta \text{ is continuous} \\[2ex] \dfrac{f(\mathbf{X}|\theta)\pi(\theta)}{\sum_{\tilde{\theta}\in\Theta} f(\mathbf{X}|\tilde{\theta})\pi(\tilde{\theta})} & \text{if } \theta \text{ is discrete.} \end{cases}$$

Exercise: Derive the above.

Given $X|\theta \sim f(\underset{\sim}{x}|\theta)$, $\theta \sim \pi(\theta)$ then if θ is continuous ...

$$\underline{\pi(\theta|\underset{\sim}{x})} = \frac{f(\underset{\sim}{x},\theta)}{f(\underset{\sim}{x})} = \frac{f(\underset{\sim}{x}|\theta)\pi(\theta)}{\int_\Theta f(\underset{\sim}{x}|\theta')\pi(\theta')\,d\theta'}$$

$\pi(\theta|\underset{\sim}{X})$ posterior



Bayesian estimation with posterior mean

A typical Bayesian estimator of θ is the *posterior mean* of θ, which is

$$\hat{\theta}_{\text{Bayes}} = \mathbb{E}[\theta|\mathbf{X}] = \begin{cases} \int_\Theta \theta \cdot \pi(\theta|\mathbf{X})d\theta & \text{if } \theta \text{ is continuous} \\ \sum_{\theta\in\Theta} \theta \cdot \pi(\theta|\mathbf{X}) & \text{if } \theta \text{ is discrete.} \end{cases}$$

Could also use the posterior mode or median.

Exercise: Show that the above estimator is the solution to

$$\hat{\theta}_{\text{Bayes}} = \underset{a}{\text{argmin}}\ \mathbb{E}[(\theta - a)^2|\mathbf{X}]. = \mathbb{E}[\theta|\underset{\sim}{X}]$$

By the same arguments as below,

Generally: If Y is a rv with $\mathbb{E}Y = \mu$ $\mathrm{Var}\,Y = \sigma^2$.

$$\underset{a}{\operatorname{argmin}}\ \mathbb{E}(Y-a)^2$$

$$= \underset{a}{\operatorname{argmin}}\ \mathbb{E}\left(\left(Y - \mathbb{E}Y\right) + \left(\mathbb{E}Y - a\right)\right)^2$$

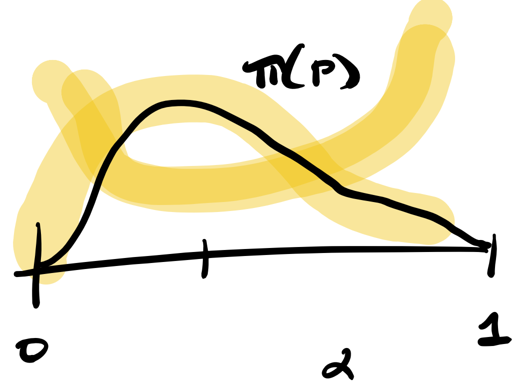
$$= \underset{a}{\operatorname{argmin}}\ \mathbb{E}\left\{ \left(Y - \mathbb{E}Y\right)^2 + 2\left(Y - \mathbb{E}Y\right)\left(\mathbb{E}Y - a\right) + \left(\mathbb{E}Y - a\right)^2 \right\}$$

$$= \underset{a}{\operatorname{argmin}} \left[\cancel{\mathbb{E}\left(Y - \mathbb{E}Y\right)^2} + 2\,\overbrace{\mathbb{E}\left(Y - \mathbb{E}Y\right)}^{0}\underbrace{\left(\mathbb{E}Y - a\right)}_{\text{const}} + \left(\mathbb{E}Y - a\right)^2 \right]$$

$$= \underset{a}{\operatorname{argmin}}\ \left(\mathbb{E}Y - a\right)^2$$

$$= \mathbb{E}Y.$$

Select 1 spectator of TBB game to shoot 10 free throws at ½-time.
Let Y = # free throws made.



Exercise: Suppose we have

$$Y|p \sim \text{Binomial}(n, p)$$

10

$$\boxed{p \sim \text{Beta}(\alpha, \beta)} \rightarrow \mathbb{E}p = \frac{\alpha}{\alpha+\beta}$$

select one spectator

for some $\alpha > 0$ and $\beta > 0$.

1. Find the posterior distribution of $p|Y$.
2. Find an expression for $\hat{p}_{\text{Bayes}} = \mathbb{E}[p|Y]$.
3. Suppose $n = 10$ and $Y = 3$ is observed. Find the posterior mean of p under
   - $\alpha = 1$, $\beta = 1$
   - $\alpha = 4$, $\beta = 4$
   - $\alpha = 4$, $\beta = 10$

$$p(y|p) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

① $$\pi(p|y) = \frac{p(y|p)\,\pi(p)}{p(y)}$$

$$p(y) = \int_0^1 \underbrace{p(y|p)\,\pi(p)}_{p(y,p)}\,dp$$

$$= \int_0^1 \binom{n}{y} p^y (1-p)^{n-y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\,dp$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\binom{n}{y}\frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)} \underbrace{\int_0^1 \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} p^{(y+\alpha)-1}(1-p)^{(n-y+\beta)-1}\,dp}_{=1}$$

$$\boxed{p_Y(y) = \binom{n}{y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}}$$ Beta-Binomial

$$\pi(p|y) = \frac{\cancel{\binom{n}{y}} p^y (1-p)^{n-y} \cancel{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}} p^{\alpha-1}(1-p)^{\beta-1}}{\cancel{\binom{n}{y}}\cancel{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}\frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}}$$

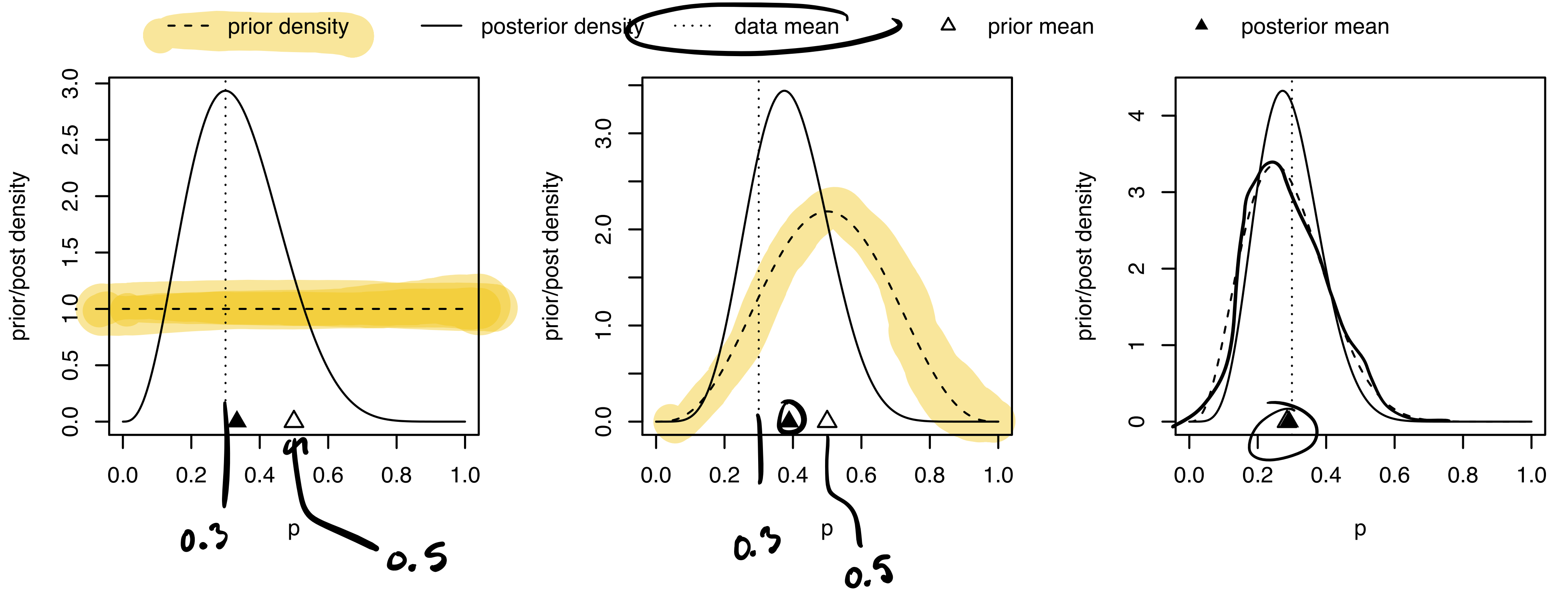
$$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} p^{(y+\alpha)-1}(1-p)^{(n-y+\beta)-1}$$

$\mathbb{E}p = \frac{\alpha}{\alpha+\beta} \leftarrow$ Best guess of p before observing data

$$p|y \sim \text{Beta}(y+\alpha,\ n-y+\beta)$$

② $$\hat{p}_{Bayes} = \mathbb{E}[p|y] = \frac{y+\alpha}{y+\alpha+n-y+\beta} \overset{\text{algebra}}{=} \underbrace{\frac{y}{n}}_{\text{data mean}}\left(\frac{n}{n+\alpha+\beta}\right) + \underbrace{\frac{\alpha}{\alpha+\beta}}_{\text{prior mean}}\left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)$$

- Suppose $n = 10$ and $Y = 3$
  - $\alpha = 1$, $\beta = 1$
  - $\alpha = 4$, $\beta = 4$
  - $\alpha = 4$, $\beta = 10$



"proportional to"

$$\pi(\theta|\underset{\sim}{x}) = \frac{f(\underset{\sim}{x}|\theta)\pi(\theta)}{\int_\Theta f(x|\theta')\pi(\theta')d\theta'} \propto f(\underset{\sim}{x}|\theta)\pi(\theta) \propto g(\theta)$$

just a normalizing constant.

Shortcut to finding the posterior distribution

1. Find $g(\theta)$ such that $f(\mathbf{X}|\theta)\pi(\theta) \propto g(\theta)$.
2. Then $\pi(\theta|\mathbf{X}) = C \cdot g(\theta)$, where $C = [\int_\Theta g(\theta)d\theta]^{-1}$.

Can sometimes find C in Step 2 by identifying the distribution.

Exercise: Find the posterior distribution in the previous example with this strategy.

$$\pi(p \mid y) = \frac{p(y \mid p)\, \pi(p)}{p(y)}$$

$$\propto \cancel{\binom{n}{y}} p^{y} (1-p)^{n-y} \cancel{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}} p^{\alpha-1} (1-p)^{\beta-1}$$

$$\propto p^{(y+\alpha)-1} (1-p)^{(n-y+\beta)-1}$$

$\propto$ the pdf of the Beta$(y+\alpha,\ n-y+\beta)$ dist.

Exercise: Assume for $\mathbf{X} = (X_1, \ldots, X_n)$ the hierarchical model

$$\mathbf{X}|\lambda \overset{\text{ind}}{\sim} \text{Poisson}(\lambda)$$
$$\lambda \sim \text{Gamma}(\alpha, \beta)$$

for some $\alpha > 0$ and $\beta > 0$.

1. Find the posterior distribution of $\lambda|\mathbf{X}$.
2. Find an expression for $\hat{\lambda}_{\text{Bayes}} = \mathbb{E}[\lambda|\mathbf{X}]$.
3. Under $\alpha = 4$ and $\beta = 5$ and a sample size of $n = 3$, compute the posterior mean of $\lambda|\mathbf{X}$ when $\bar{X}_n = 10, 15, 30$.

1

$$p(\underset{\sim}{x}|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$\pi(\lambda) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \lambda^{\alpha-1} e^{-\lambda/\beta}$$

$$\pi(\lambda \mid \underset{\sim}{X}) \propto p(\underset{\sim}{X} \mid \lambda)\, \pi(\lambda)$$

$$= \left(\prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} \right) \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \lambda^{\alpha-1} e^{-\lambda/\beta}$$

$$= \frac{e^{-n\lambda} \lambda^{\sum x_i = n\bar{x}_n}}{\prod_{i=1}^{n} x_i!} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \lambda^{\alpha-1} e^{-\lambda/\beta}$$

$$\propto \cdot\ \lambda^{n\bar{x}_n + \alpha - 1} e^{-\lambda / \left(\frac{1}{\beta} + n\right)^{-1}}$$

$\propto$ the pdf of $\text{Gamma}\left(n\bar{x}_n + \alpha,\ \frac{\beta}{1+n\beta}\right)$

$$\left(\frac{1}{\beta} + n\right)^{-1} = \frac{1}{\frac{1}{\beta} + n} = \frac{\beta}{1+n\beta}$$

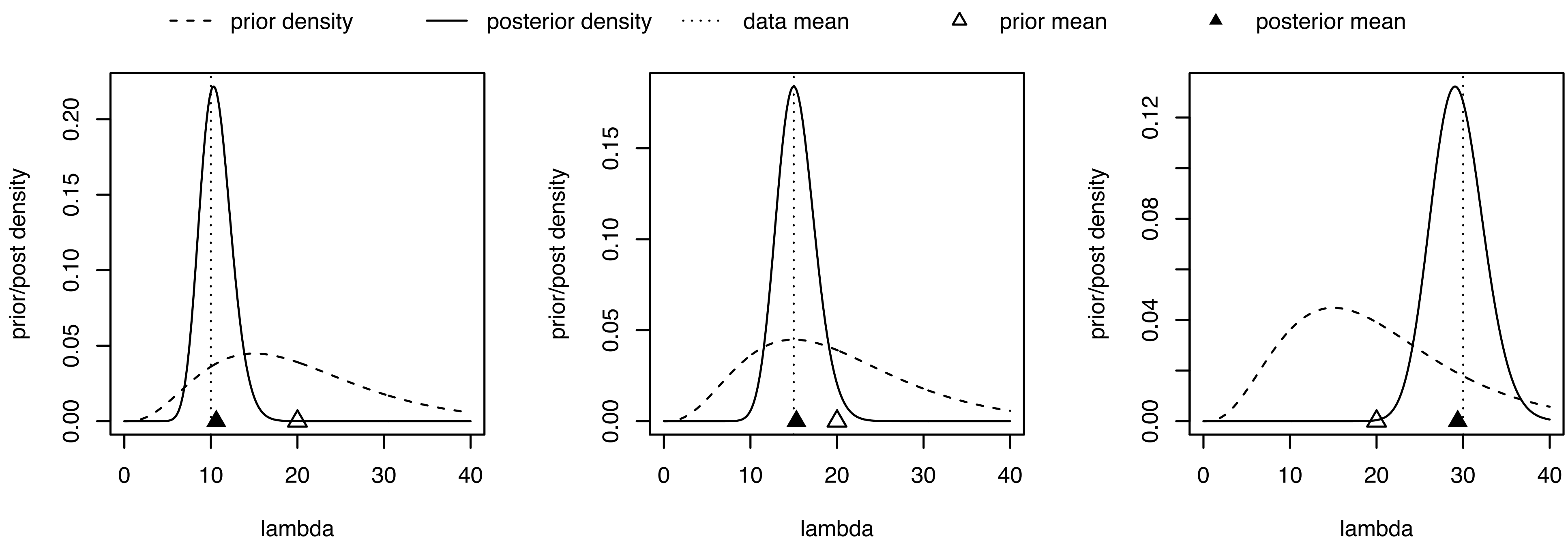
② $\hat{\lambda}_{\text{Bayes}} = \mathbb{E}[\lambda \mid \underset{\sim}{X}] = (n\bar{x}_n + \alpha)\left(\frac{\beta}{1+n\beta}\right)$

prior mean

$$= \bar{X}_n \left(\frac{n\beta}{1+n\beta}\right) + \alpha\beta \left(\frac{1}{1+n\beta}\right)$$

data mean

$\mathbb{E}\lambda = \alpha\beta$



prior density
posterior density
data mean
prior mean
posterior mean
prior/post density
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
lambda
0.00
0.05
0.10
0.15
0.04
0.08
0.12

$$\pi(\theta|\underset{\sim}{x}) \propto f(\underset{\sim}{x}|\theta)\,\pi(\theta) = \underbrace{g(T(\underset{\sim}{x});\theta)\,h(\underset{\sim}{x})}_{\text{factorization thm}} \cdot \pi(\theta) \propto g(T(\underset{\sim}{x});\theta)\,\pi(\theta)$$

Theorem (Posterior distribution depends on data via a suff. stat.)

If, given θ, $T(\mathbf{X})$ is a sufficient statistic for θ, then the distribution of $\theta|\mathbf{X}$ is the same as the distribution of $\theta|T(\mathbf{X})$.

Exercise: Show why using the factorization theorem.

This means we have $\pi(\theta|\mathbf{X}) \propto f_T(T(\mathbf{X})|\theta)\pi(\theta)$, provided $T(\mathbf{X})$ is suff.

This can simplify calculations.

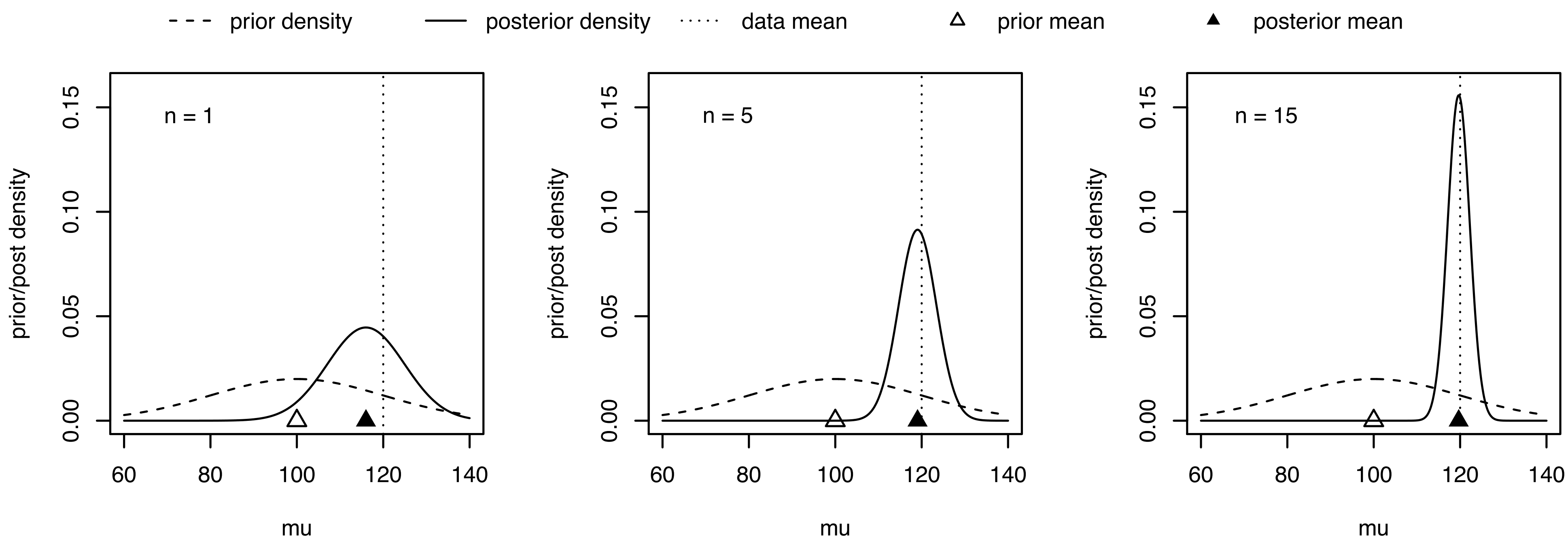
Exercise: Assume for $\mathbf{Y} = (Y_1, \dots, Y_n)$ the hierarchical model

$$\mathbf{Y}|\mu \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$$
$$\mu \sim \text{Normal}(\mu_0, \tau^2)$$

for some $\mu_0 \in \mathbb{R}$ and $\tau^2 > 0$, where σ^2 is a known constant.

1. Find the posterior distribution of $\mu|\mathbf{Y}$.
2. Find an expression for $\hat{\mu}_{\text{Bayes}} = \mathbb{E}[\mu|\mathbf{Y}]$.
3. Let $\sigma = 10$, $\tau = 20$, and $\mu_0 = 100$. Suppose that under sample sizes of $n = 1, 5, 15$, the sample mean $\bar{Y}_n = 120$ is observed. Compute the posterior mean of $\mu|\mathbf{Y}$ in each case.

$$\hat{\mu}_{\text{Bayes}} = \bar{Y}_n \left(\frac{\tau^2}{\tau^2 + \sigma^2/n} \right) + \mu_0 \left(\frac{\sigma^2/n}{\tau^2 + \sigma^2/n} \right)$$



prior density
posterior density
data mean
prior mean
posterior mean
n = 1
n = 5
n = 15
prior/post density
0.00
0.05
0.10
0.15
60
80
100
120
140
mu

Conjugacy

In the Bayesian setup, a *conjugate prior* is a prior distribution under which the posterior distribution belongs to the same family of distributions.

Examples we have seen:

- In Beta-Binomial model, posterior was Beta.
- In Gamma-Poisson model, posterior was Gamma.
- In Normal-Normal model, posterior was Normal.

We do not *need* to choose a conjugate prior; it just makes hand calculations easier.

Conjugacy mostly for textbook examples. In real life we use 🖥️ 🖥️ 🖥️ power.

$$p(\underset{\sim}{x};\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{n\bar{x}_n}}{\prod_{i=1}^{n} x_i!} \qquad \pi(\lambda) = \frac{1}{\lambda_{max}} \mathbb{1}(0<\lambda<\lambda_{max})$$

Exercise: Assume for $\mathbf{X} = (X_1, \ldots, X_n)$ the hierarchical model

$$\mathbf{X}|\lambda \overset{\text{ind}}{\sim} \text{Poisson}(\lambda)$$
$$\lambda \sim \text{Uniform}(0, \underline{\lambda_{\text{max}}})$$

for some $\lambda_{\text{max}} > 0$.

1. Find the posterior density of $\lambda|\mathbf{X}$.
2. Is the prior a conjugate prior? NO
3. What happens to the posterior distribution of $\lambda|\mathbf{X}$ when $\boxed{\lambda_{\text{max}} \to \infty}$?

① $$\pi(\lambda|\underset{\sim}{x}) \propto p(\underset{\sim}{x};\lambda)\cdot\pi(\lambda) = \frac{e^{-n\lambda}\lambda^{n\bar{x}_n}}{\prod_{i=1}^{n} x_i!} \frac{1}{\lambda_{max}} \mathbb{1}(0<\lambda<\lambda_{max})$$

$$= C \cdot \lambda^{(n\bar{x}_n+1)-1} e^{-n\lambda} \mathbb{1}(0<\lambda<\lambda_{max})$$

$$C = \frac{1}{\prod x_i!}\frac{1}{\lambda_{max}}$$

$$\pi(\lambda \mid \underset{\sim}{x}) = \frac{\lambda^{(n\bar{x}_n+1)-1} e^{-n\lambda}}{\int_0^{\lambda_{max}} (\lambda')^{(n\bar{x}_n+1)-1} e^{-n\lambda'} \, d\lambda'} \mathbb{1}(0 < \lambda < \lambda_{max})$$

Consider finding

$$\hat{\lambda}_{Bayes} = \mathbb{E}[\lambda \mid \underset{\sim}{X}] = \int_0^{\lambda_{max}} \lambda \cdot \frac{\lambda^{(n\bar{x}_n+1)-1} e^{-n\lambda}}{\int_0^{\lambda_{max}} (\lambda')^{(n\bar{x}_n+1)-1} e^{-n\lambda'} \, d\lambda'} \, d\lambda$$