

# STAT 35920

Robert  
Winter

## HW #3

- 1a. Sample  $n > 1$  observations  $(y_1, \dots, y_n)$  w/ observed mean  $\bar{y}$ . Assume the likelihood for (joint dist.) of  $(y_1, \dots, y_n)$  is prop. to

$$\prod_{i=1}^n \exp\left\{-\frac{1}{2}\tau(y_i - \mu)^2\right\}$$

- The prior for mean  $\mu$  is  $\mu \sim N(\mu_0, \sigma_0^2)$ .  
 Let  $\tau_0 = \frac{1}{\sigma_0^2}$ . Assume  $\tau, \tau_0, \mu_0$  all known.  
 Derive  $p(\bar{y}|\mu)$  +  $p(\mu|\bar{y})$ .

- We'll write  $\tau = \frac{1}{\sigma^2}$ .
- Since  $L(y_1, \dots, y_n | \mu, \tau) \propto \prod_{i=1}^n \exp\left\{-\frac{1}{2}\tau(y_i - \mu)^2\right\}$ , it follows that for each  $i \in \{1, \dots, n\}$ ,  $f(y_i | \mu, \tau) \propto \exp\left\{-\frac{1}{2}\tau(y_i - \mu)^2\right\}$ , so  $y_i \sim N(\mu, \frac{1}{\tau}) \quad \forall i = 1, \dots, n$ .
- Since any linear combination of normal RVs is normal,  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  has normal distribution.
- $E[\bar{y}] = E\left[\frac{1}{n} \sum_{i=1}^n y_i\right] = \frac{1}{n} \sum_{i=1}^n E[y_i] = \frac{1}{n} \cdot n\mu = \mu$ .
- $\text{Var}(\bar{y}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n y_i\right) = \frac{1}{n^2} \sum_{i=1}^n \frac{1}{\tau} = \frac{1}{n^2} \cdot \frac{n}{\tau} = \frac{1}{n\tau}$ , where we've used that  $y_i \perp\!\!\!\perp y_j \quad \forall i \neq j$ . We know this is the case since there are no covariance terms in the likelihood function.

- So,  $\bar{y} | \mu \sim N\left(\mu, \frac{1}{n\tau}\right)$ , and

$$p(\bar{y} | \mu) = \frac{1}{\sqrt{2\pi \cdot \frac{1}{n\tau}}} \exp\left\{-\frac{1}{2 \cdot \frac{1}{n\tau}} (\bar{y} - \mu)^2\right\}$$

$$\Rightarrow p(\bar{y} | \mu) = \frac{\sqrt{n\tau}}{\sqrt{2\pi}} \exp\left\{-\frac{n\tau}{2} (\bar{y} - \mu)^2\right\}$$

Then  $p(\mu | \bar{y}) \propto p(\bar{y} | \mu) \pi(\mu)$ .

$$\begin{aligned}
 &= \frac{\sqrt{\tau_n}}{\sqrt{2\pi}} e^{-\frac{\tau_n}{2}(\bar{y} - \mu)^2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\mu - \mu_0)^2} \\
 &\propto \exp \left\{ -\frac{1}{2}\tau_n(\bar{y}^2 - 2\bar{y}\mu + \mu^2) - \frac{1}{2\sigma^2}(\mu^2 - 2\mu_0\mu + \mu_0^2) \right\} \\
 &= \exp \left\{ -\frac{1}{2}\tau_n\bar{y}^2 + \tau_n\bar{y}\mu - \frac{1}{2}\tau_n\mu^2 - \frac{1}{2}\tau_0\mu^2 + \tau_0\mu_0\mu - \frac{1}{2}\tau_0\mu_0^2 \right\} \\
 &= \exp \left\{ \mu^2 \left( -\frac{1}{2}\tau_n - \frac{1}{2}\tau_0 \right) + \mu (\tau_n\bar{y} + \tau_0\mu_0) \right. \\
 &\quad \left. \cdot \exp \left\{ -\frac{1}{2}\tau_n\bar{y}^2 - \frac{1}{2}\tau_0\mu_0^2 \right\} \right\} \\
 &\propto \exp \left\{ -\frac{1}{2}[\mu^2(\tau_n + \tau_0) - 2\mu(\tau_n\bar{y} + \tau_0\mu_0)] \right\} \\
 &= \exp \left\{ -\frac{1}{2}(\tau_n + \tau_0) \left[ \mu^2 - \frac{2(\tau_n\bar{y} + \tau_0\mu_0)}{\tau_n + \tau_0} \mu + \left( \frac{2(\tau_n\bar{y} + \tau_0\mu_0)}{\tau_n + \tau_0} \right)^2 \right] \right\} \\
 &= \exp \left\{ -\frac{1}{2}(\tau_n + \tau_0) \left[ \mu^2 - \frac{2(\tau_n\bar{y} + \tau_0\mu_0)}{\tau_n + \tau_0} \mu + \left( \frac{2(\tau_n\bar{y} + \tau_0\mu_0)}{\tau_n + \tau_0} \right)^2 - \left( \frac{\tau_n\bar{y} + \tau_0\mu_0}{\tau_n + \tau_0} \right)^2 \right] \right\} \\
 &= \exp \left\{ -\frac{1}{2}(\tau_n + \tau_0) \left[ \mu - \frac{\tau_n\bar{y} + \tau_0\mu_0}{\tau_n + \tau_0} \right]^2 \right\} \\
 &\quad \cdot \exp \left\{ -\frac{1}{2}(\tau_n + \tau_0) \cdot - \left( \frac{\tau_n\bar{y} + \tau_0\mu_0}{\tau_n + \tau_0} \right)^2 \right\} \\
 &\propto \exp \left\{ -\frac{1}{2}(\tau_n + \tau_0) \left( \mu - \frac{\tau_n\bar{y} + \tau_0\mu_0}{\tau_n + \tau_0} \right)^2 \right\}
 \end{aligned}$$

Thus,

$$\mu | \bar{y} \sim N \left( \frac{\tau_n\bar{y} + \tau_0\mu_0}{\tau_n + \tau_0}, \frac{1}{\tau_n + \tau_0} \right),$$

and

$$p(\mu | \bar{y}) = \frac{1}{\sqrt{2\pi \cdot \frac{1}{\tau_n + \tau_0}}} \exp \left\{ -\frac{1}{2(\frac{1}{\tau_n + \tau_0})} \left( \mu - \frac{\tau_n\bar{y} + \tau_0\mu_0}{\tau_n + \tau_0} \right)^2 \right\}$$

$$p(\mu | \bar{y}) = \frac{\sqrt{\tau_n + \tau_0}}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(\tau_n + \tau_0) \left( \mu - \frac{\tau_n\bar{y} + \tau_0\mu_0}{\tau_n + \tau_0} \right)^2 \right\}$$

1b. Likelihood  $y|\theta \sim \text{Bin}(n, \theta)$ ; prior  $\theta \sim \text{Beta}(a, b)$ .

For a new obs  $y_{\text{new}}$ , what is the posterior predictive distribution  $p(y_{\text{newly}})$ ? What is posterior predictive mean  $E[y_{\text{newly}}]$ ?

$$\begin{aligned} p(y_{\text{newly}}) &= \int_0^1 p(y_{\text{new}}|\theta) p(\theta|y) d\theta \\ &= \int_0^1 p(y_{\text{new}}|\theta) \cdot p(\theta|y) d\theta \\ &= \int_0^1 (y_{\text{new}})^{\theta} (1-\theta)^{n-y} \cdot p(\theta|y) d\theta \end{aligned}$$

$$\begin{aligned} p(\theta|y) &\propto \pi(\theta) L(y|\theta) \\ &= \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} \cdot \binom{n}{y} \theta^y (1-\theta)^{n-y} \\ &\propto \theta^{a+y-1} (1-\theta)^{b+n-y-1} \end{aligned}$$

so  $\theta|y \sim \text{Beta}(a+y, b+n-y)$ .

Thus,

$$\begin{aligned} p(y_{\text{newly}}) &= \int_0^1 (y_{\text{new}})^{\theta} (1-\theta)^{n-y} \cdot \frac{1}{B(a+y, b+n-y)} \theta^{a+y-1} (1-\theta)^{b+n-y-1} d\theta \\ &= (y_{\text{new}})^n \frac{1}{B(a+y, b+n-y)} \int_0^1 \theta^{a+y+y_{\text{new}}-1} (1-\theta)^{b+2n-y-y_{\text{new}}-1} d\theta \\ &\propto (y_{\text{new}})^n \cdot B(a+y+y_{\text{new}}, b+2n-y-y_{\text{new}}), \end{aligned}$$

so

$$y_{\text{newly}} \sim \text{BetaBin}(n, a+y, b+n-y)$$

Thus,  $P(y_{\text{newly}}) = \binom{n}{y_{\text{new}}} \frac{B(a+y+y_{\text{new}}, b+2n-y-y_{\text{new}})}{B(a+y, b+n-y)}$

So, using the expectation of the Beta-Binomial distribution,

$$E[y_{\text{newly}}] = \frac{n(a+y)}{(a+y)+(b+n-y)}$$

$$E[y_{\text{newly}}] = \frac{n(a+y)}{a+b+n}$$

2. Assume  $y \sim N(\mu, \sigma^2)$ . Sps  $\mu$  &  $\tau = \sigma^{-2}$  both unknown,  
 & joint prior is  $\pi(\mu, \tau) = \pi(\mu | \tau)\pi(\tau)$  w/  
 $\mu | \tau \sim N(\mu_0, \tau^{-1}\sigma_0^2)$ ,  
 $\tau \sim \text{Gamma}(\frac{a_0}{2}, \frac{b_0}{2})$ .

Derive & recognize  $p(\mu | \tau, y)$ ,  $p(\mu | y)$ ,  $p(\tau | y)$ .

$$\begin{aligned}
 p(\mu | \tau, y) &\propto \pi(\mu | \tau)L(y | \mu, \tau) \\
 &= \frac{1}{\sqrt{2\pi\tau^{-1}\sigma_0^2}} e^{-\frac{1}{2\tau^{-1}\sigma_0^2}(\mu - \mu_0)^2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y - \mu)^2} \\
 &\propto \exp\left\{-\frac{1}{2\tau^{-1}\sigma_0^2}(\mu - \mu_0)^2 - \frac{1}{2\sigma^2}(y - \mu)^2\right\} \\
 &= \exp\left\{-\frac{1}{2\tau^{-1}\sigma_0^2}(\mu^2 - 2\mu_0\mu + \mu_0^2) - \frac{1}{2\sigma^2}(y^2 - 2y\mu + \mu^2)\right\} \\
 &= \exp\left\{-\frac{\mu^2}{2\tau^{-1}\sigma_0^2} + \frac{\mu_0\mu}{\tau^{-1}\sigma_0^2} - \frac{\mu_0^2}{2\tau^{-1}\sigma_0^2} - \frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right\} \\
 &= \exp\left\{\mu^2\left(\frac{-1}{2\tau^{-1}\sigma_0^2} - \frac{1}{2\sigma^2}\right) + \mu\left(\frac{\mu_0}{\tau^{-1}\sigma_0^2} + \frac{y}{\sigma^2}\right)\right\} \\
 &\cdot \exp\left\{-\frac{\mu_0^2}{2\tau^{-1}\sigma_0^2} - \frac{y^2}{2\sigma^2}\right\} \\
 &\propto \exp\left\{\mu^2\left(\frac{-1}{2\tau^{-1}\sigma_0^2} - \frac{\sigma_0^2}{2\tau^{-1}\sigma_0^2}\right) + \mu\left(\frac{\mu_0}{\tau^{-1}\sigma_0^2} + \frac{y\sigma_0^2}{\tau^{-1}\sigma_0^2}\right)\right\} \\
 &= \exp\left\{\left(\frac{-1 - \sigma_0^2}{2\tau^{-1}\sigma_0^2}\right)\left[\mu^2 + \mu\left(\frac{\mu_0 + y\sigma_0^2}{\tau^{-1}\sigma_0^2}\right)\right] / \left(\frac{-1 - \sigma_0^2}{2\tau^{-1}\sigma_0^2}\right)\right\} \\
 &= \exp\left\{-\frac{(1 + \sigma_0^2)}{2\tau^{-1}\sigma_0^2}\left[\mu^2 + \mu \cdot \frac{\mu_0 + y\sigma_0^2}{\tau^{-1}\sigma_0^2} \cdot \frac{2\tau^{-1}\sigma_0^2}{-(1 + \sigma_0^2)}\right]\right\} \\
 &= \exp\left\{-\frac{(1 + \sigma_0^2)}{2\tau^{-1}\sigma_0^2}\left[\mu^2 - 2\mu\left(\frac{\mu_0 + y\sigma_0^2}{1 + \sigma_0^2}\right)\right]\right\} \\
 &= \exp\left\{-\frac{(1 + \sigma_0^2)}{2\tau^{-1}\sigma_0^2}\left[\mu^2 - 2\mu\left(\frac{\mu_0 + y\sigma_0^2}{1 + \sigma_0^2}\right) + \left(\frac{\mu_0 + y\sigma_0^2}{1 + \sigma_0^2}\right)^2 - \left(\frac{\mu_0 + y\sigma_0^2}{1 + \sigma_0^2}\right)^2\right]\right\} \\
 &= \exp\left\{-\frac{(1 + \sigma_0^2)}{2\tau^{-1}\sigma_0^2}\left[\mu - \left(\frac{\mu_0 + y\sigma_0^2}{1 + \sigma_0^2}\right)\right]^2\right\} \exp\left\{-\frac{(1 + \sigma_0^2)}{2\tau^{-1}\sigma_0^2} \cdot -\left(\frac{\mu_0 + y\sigma_0^2}{1 + \sigma_0^2}\right)^2\right\} \\
 &\propto \exp\left\{\frac{-1}{2\left(\frac{\tau^{-1}\sigma_0^2}{1 + \sigma_0^2}\right)}\left[\mu - \left(\frac{\mu_0 + y\sigma_0^2}{1 + \sigma_0^2}\right)\right]^2\right\}
 \end{aligned}$$



That is,

$$\mu | \tau, y \sim N\left(\frac{\mu_0 + y\sigma_0^2}{1 + \sigma_0^2}, \frac{\tau^{-1}\sigma_0^2}{1 + \sigma_0^2}\right)$$

So,

$$p(\mu | \tau, y) = \frac{1}{\sqrt{2\pi \cdot \frac{\tau^{-1}\sigma_0^2}{1 + \sigma_0^2}}} \exp\left\{-\frac{1}{2\left(\frac{\tau^{-1}\sigma_0^2}{1 + \sigma_0^2}\right)} \left(\mu - \frac{\mu_0 + y\sigma_0^2}{1 + \sigma_0^2}\right)^2\right\}$$

To find  $p(\tau|y)$ , we begin by finding  $p(y, \tau|z)$ .

$$\begin{aligned}
 p(y, \tau|z) &\propto L(y|z, \mu, \tau) \pi(\mu|\tau) \pi(\tau) \\
 &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \cdot \frac{1}{\sqrt{2\pi\tau^{-1}\sigma_0^2}} e^{-\frac{1}{2\tau^{-1}\sigma_0^2}(\mu-\mu_0)^2} \\
 &\quad \cdot \frac{(b_0/2)^{\alpha_0/2}}{\Gamma(\alpha_0/2)} \tau^{\frac{\alpha_0}{2}-1} e^{-b_0\tau/2} \\
 &= \frac{\tau^{1/2}}{\sqrt{2\pi}} e^{-\frac{1}{2}\tau(y-\mu)^2} \cdot \frac{\tau^{1/2}}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}\tau(\mu-\mu_0)^2} \\
 &\quad \cdot \frac{(b_0/2)^{\alpha_0/2}}{\Gamma(\alpha_0/2)} \tau^{\frac{\alpha_0}{2}-1} e^{-b_0\tau/2} \\
 &\propto \tau^{\frac{\alpha_0}{2} + \frac{1}{2} - 1} \tau^{\frac{1}{2}} e^{-b_0\tau/2} \exp\left\{-\frac{\tau}{2}(y-\mu)^2\right\} \exp\left\{-\frac{\tau}{2\sigma_0^2}(\mu-\mu_0)^2\right\} \\
 &= \tau^{\frac{\alpha_0}{2} + \frac{1}{2} - 1} \tau^{\frac{1}{2}} e^{-\frac{b_0\tau}{2}} \exp\left\{-\frac{\tau}{2}(y-\mu)^2 - \frac{\tau}{2\sigma_0^2}(\mu-\mu_0)^2\right\}
 \end{aligned}$$

Then

$$p(\tau|y) \propto \int_{-\infty}^{\infty} \tau^{\frac{\alpha_0}{2} - \frac{1}{2}} \tau^{\frac{1}{2}} e^{-\frac{b_0\tau}{2}} \exp\left\{-\frac{\tau}{2}(y-\mu)^2 - \frac{\tau}{2\sigma_0^2}(\mu-\mu_0)^2\right\} d\mu$$

$$= \tau^{\frac{\alpha_0}{2} - \frac{1}{2}} \tau^{\frac{1}{2}} e^{-\frac{b_0\tau}{2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{\tau}{2}(y-\mu)^2 - \frac{\tau}{2\sigma_0^2}(\mu-\mu_0)^2\right\} d\mu$$

$$= \int_{-\infty}^{\infty} \exp\left\{-\frac{\tau}{2}(y-\mu)^2 - \frac{\tau}{2\sigma_0^2}(\mu-\mu_0)^2\right\} d\mu$$

$$= \int_{-\infty}^{\infty} \exp\left\{-\frac{\tau}{2}(y^2 - 2\mu y + \mu^2) - \frac{\tau}{2\sigma_0^2}(\mu^2 - 2\mu\mu_0 + \mu_0^2)\right\} d\mu$$

$$= \int_{-\infty}^{\infty} \exp\left\{\mu^2\left(-\frac{\tau}{2} - \frac{\tau}{2\sigma_0^2}\right) + \mu\left(\frac{-\tau}{2} \cdot -2y - \frac{\tau}{2\sigma_0^2} \cdot -2\mu_0\right) + \left(-\frac{\tau}{2}y^2 - \frac{\tau}{2\sigma_0^2}\mu_0^2\right)\right\} d\mu$$

$$= \int_{-\infty}^{\infty} \exp\left\{-\mu^2\left(\frac{\tau\sigma_0^2 + \tau}{2\sigma_0^2}\right) + \mu\left(\tau y + \frac{\tau\mu_0}{\sigma_0^2}\right) - \left(\frac{\tau}{2}y^2 + \frac{\tau\mu_0^2}{2\sigma_0^2}\right)\right\} d\mu$$

$$= \int_{-\infty}^{\infty} \exp\left\{-\left[\underbrace{\mu^2\left(\frac{\tau\sigma_0^2 + \tau}{2\sigma_0^2}\right)}_{:= \alpha} - \underbrace{\mu\left(\frac{\tau y + \tau\mu_0}{\sigma_0^2}\right)}_{:= \beta} + \underbrace{\left(\frac{\tau\sigma_0^2 y^2 + \tau\mu_0^2}{2\sigma_0^2}\right)}_{:= \gamma}\right]\right\} d\mu$$

$$= \sqrt{\frac{\pi}{\alpha}} \exp\left\{\frac{\beta^2}{4\alpha} - \gamma\right\}$$

$$= \sqrt{\frac{\pi}{\left(\frac{\tau\sigma_0^2 + \tau}{2\sigma_0^2}\right)}} \exp\left\{\frac{1}{4} \left(\frac{\tau y + \tau\mu_0}{\sigma_0^2}\right)^2 \cdot \frac{2\sigma_0^2}{\tau\sigma_0^2 + \tau} - \frac{\tau\sigma_0^2 y^2 + \tau\mu_0^2}{2\sigma_0^2}\right\}$$

$$= \sqrt{\frac{2\pi\sigma_0^2}{\tau(\sigma_0^2 + 1)}} \exp\left\{\frac{1}{2} \cdot \frac{\tau^2(y + \mu_0)^2}{\sigma_0^4} \cdot \frac{\sigma_0^2}{\tau(\sigma_0^2 + 1)} - \frac{\tau\sigma_0^2 y^2 + \tau\mu_0^2}{2\sigma_0^2}\right\}$$

$$\begin{aligned}
&= \tau^{-1/2} \sqrt{\frac{2\pi\sigma_0^2}{\sigma_0^2 + 1}} \exp \left\{ \frac{\tau(y\sigma_0^2 + \mu_0)^2}{2\sigma_0^2(\sigma_0^2 + 1)} - \frac{\tau(\sigma_0^2 y^2 + \mu_0^2)}{2\sigma_0^2} \right\} \\
&\propto \tau^{-1/2} \exp \left\{ \frac{\tau}{2\sigma_0^2} \left( \frac{(y\sigma_0^2 + \mu_0)^2}{\sigma_0^2 + 1} - \frac{(\sigma_0^2 + 1)(\sigma_0^2 y^2 + \mu_0^2)}{\sigma_0^2 + 1} \right) \right\} \\
&= \tau^{-1/2} \exp \left\{ \frac{\tau}{2\sigma_0^2} \left( y^2 \sigma_0^4 + 2\mu_0 y \sigma_0^2 + \mu_0^2 - (\sigma_0^4 y^2 + \mu_0^2 \sigma_0^2 + \sigma_0^2 y^2 + \mu_0^2) \right) \right\} \\
&= \tau^{-1/2} \exp \left\{ \frac{\tau}{2(\sigma_0^2 + 1)} \left( \frac{2\mu_0 y \sigma_0^2 - \mu_0^2 \sigma_0^2 - \sigma_0^2 y^2}{\sigma_0^2} \right) \right\} \\
&= \tau^{-1/2} \exp \left\{ \frac{\tau}{2(\sigma_0^2 + 1)} \left( \frac{-\sigma_0^2(y^2 - 2\mu_0 y + \mu_0^2)}{\sigma_0^2} \right) \right\} \\
&= \tau^{-1/2} \exp \left\{ \frac{-\tau}{2(\sigma_0^2 + 1)} (y - \mu_0)^2 \right\}
\end{aligned}$$

So,

$$\begin{aligned}
p(\tau|y) &\propto \tau^{\frac{a_0}{2} - \frac{1}{2}} \tau^{\frac{1}{2}} e^{-\frac{b_0\tau}{2}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{\tau}{2}(y - \mu)^2 - \frac{\tau}{2\sigma_0^2} (\mu - \mu_0)^2 \right\} d\mu \\
&\propto \tau^{\frac{a_0}{2} - \frac{1}{2}} \tau^{\frac{1}{2}} e^{-\frac{b_0\tau}{2}} \tau^{-\frac{1}{2}} \exp \left\{ -\frac{\tau}{2(\sigma_0^2 + 1)} (y - \mu_0)^2 \right\} \\
&= \tau^{\frac{a_0}{2} - \frac{1}{2}} \exp \left\{ -\tau \left[ \frac{b_0}{2} + \frac{1}{2(\sigma_0^2 + 1)} (y - \mu_0)^2 \right] \right\},
\end{aligned}$$

which is the kernel of a Gamma distribution. So,

$$\boxed{\tau|y \sim \text{Gamma}\left(\frac{a_0}{2} + \frac{1}{2}, \frac{b_0}{2} + \frac{(y - \mu_0)^2}{2(\sigma_0^2 + 1)}\right)}$$

Thus,

$$p(\tau|y) = \frac{\left( \frac{b_0}{2} + \frac{(y - \mu_0)^2}{2(\sigma_0^2 + 1)} \right)^{\frac{a_0}{2} + \frac{1}{2}}}{\Gamma\left(\frac{a_0}{2} + \frac{1}{2}\right)} \tau^{\frac{a_0}{2} - \frac{1}{2}} \exp \left\{ -\tau \left[ \frac{b_0}{2} + \frac{(y - \mu_0)^2}{2(\sigma_0^2 + 1)} \right] \right\}$$

with support  $(0, \infty)$

So,

$$\begin{aligned}
 p(\mu|y) &= \int_0^\infty p(\mu|\tau, y) p(\tau|y) d\tau \\
 &= \int_0^\infty \frac{1}{2\pi \cdot \frac{\tau^{-1}\sigma_0^2}{1+\sigma_0^2}} \exp\left\{-\frac{1}{2(\frac{\tau^{-1}\sigma_0^2}{1+\sigma_0^2})} \left(\mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2}\right)^2\right\} \\
 &\quad \cdot \frac{\left(\frac{b_0}{2} + \frac{(y-\mu_0)^2}{2(\sigma_0^2+1)}\right)^{\frac{a_0}{2} + \frac{1}{2}}}{\Gamma\left(\frac{a_0}{2} + \frac{1}{2}\right)} \tau^{\frac{a_0}{2} - \frac{1}{2}} \exp\left\{-\tau\left[\frac{b_0}{2} + \frac{(y-\mu_0)^2}{2(\sigma_0^2+1)}\right]\right\} d\tau \\
 &\propto \int_0^\infty \frac{1+\sigma_0^2}{2\pi\tau^{-1}\sigma_0^2} \cdot \tau^{\frac{a_0}{2} - \frac{1}{2}} \cdot \exp\left\{-\frac{1+\sigma_0^2}{2\tau^{-1}\sigma_0^2} \left(\mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2}\right)^2\right. \\
 &\quad \left.- \tau\left[\frac{b_0}{2} + \frac{(y-\mu_0)^2}{2(\sigma_0^2+1)}\right]\right\} d\tau \\
 &\propto \int_0^\infty \tau^{\frac{1}{2}} \cdot \tau^{\frac{a_0}{2} - \frac{1}{2}} \exp\left\{-\tau\left[\left(\frac{1+\sigma_0^2}{2\sigma_0^2}\right)\left(\mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2}\right)^2\right.\right. \\
 &\quad \left.\left.+ \left(\frac{b_0}{2} + \frac{(y-\mu_0)^2}{2(\sigma_0^2+1)}\right)\right]\right\} d\tau \\
 &= \int_0^\infty \underbrace{\tau^{\frac{a_0}{2}}}_{:= t^\beta} \exp\left\{-t\left[\left(\frac{1+\sigma_0^2}{2\sigma_0^2}\right)\left(\mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2}\right)^2 + \left(\frac{b_0}{2} + \frac{(y-\mu_0)^2}{2(\sigma_0^2+1)}\right)\right]\right\} dt \\
 &\quad := \alpha \\
 &= \frac{\Gamma(\beta+1)}{\alpha^{\beta+1}} \\
 &= \frac{\Gamma\left(\frac{a_0}{2} + 1\right)}{\left[\left(\frac{1+\sigma_0^2}{2\sigma_0^2}\right)\left(\mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2}\right)^2 + \left(\frac{b_0}{2} + \frac{(y-\mu_0)^2}{2(\sigma_0^2+1)}\right)\right]^{\frac{a_0}{2}+1}} \\
 &\propto \left[\left(\frac{1+\sigma_0^2}{2\sigma_0^2}\right)\left(\mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2}\right)^2 + \left(\frac{b_0}{2} + \frac{(y-\mu_0)^2}{2(\sigma_0^2+1)}\right)\right] - \left(\frac{a_0+2}{2}\right) \\
 &= \left\{ \left(\frac{b_0}{2} + \frac{(y-\mu_0)^2}{2(\sigma_0^2+1)}\right) \left[ 1 + \frac{\left(\frac{1+\sigma_0^2}{2\sigma_0^2}\right)\left(\mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2}\right)^2}{\left(\frac{b_0}{2} + \frac{(y-\mu_0)^2}{2(\sigma_0^2+1)}\right)} \right] \right\} - \left(\frac{a_0+1+1}{2}\right)
 \end{aligned}$$

$$\begin{aligned}
& \propto \left[ 1 + \left( \frac{1+\sigma_0^2}{2\sigma_0^2} \right) \left( \mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2} \right)^2 \left( \frac{2(\sigma_0^2+1)}{b_0(\sigma_0^2+1) + (y-\mu_0)^2} \right) \right]^{-\left(\frac{(a_0+1)+1}{2}\right)} \\
& = \left[ 1 + \frac{(\sigma_0^2+1)^2}{b_0\sigma_0^2(\sigma_0^2+1) + \sigma_0^2(y-\mu_0)^2} \left( \mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2} \right)^2 \right]^{-\left(\frac{(a_0+1)+1}{2}\right)} \\
& = \left[ 1 + \frac{1}{a_0+1} (a_0+1) \left( \frac{(\sigma_0^2+1)^2}{b_0\sigma_0^2(\sigma_0^2+1) + \sigma_0^2(y-\mu_0)^2} \right) \left( \mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2} \right)^2 \right]^{-\left(\frac{(a_0+1)+1}{2}\right)} \\
& = \left[ 1 + \frac{1}{a_0+1} \left( \frac{(a_0+1)(\sigma_0^2+1)^2}{b_0\sigma_0^2(\sigma_0^2+1) + \sigma_0^2(y-\mu_0)^2} \right) \left( \mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2} \right)^2 \right]^{-\left(\frac{(a_0+1)+1}{2}\right)} \\
& = \left[ 1 + \frac{1}{a_0+1} \cdot \frac{\left( \mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2} \right)^2}{\frac{b_0\sigma_0^2(\sigma_0^2+1) + \sigma_0^2(y-\mu_0)^2}{(a_0+1)(\sigma_0^2+1)^2}} \right]^{-\left(\frac{(a_0+1)+1}{2}\right)}
\end{aligned}$$

which is the kernel of a location-scale t distribution. So,

$$M|y \sim LSt\left(\frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2}, \frac{b_0\sigma_0^2(\sigma_0^2+1) + \sigma_0^2(y-\mu_0)^2}{(a_0+1)(\sigma_0^2+1)^2}, a_0+1\right)$$

$$\begin{aligned}
& \text{Thus, } P(z|y) = \frac{\Gamma\left(\frac{a_0}{2}+1\right) \left[ 1 + \frac{1}{a_0+1} \left( \frac{(a_0+1)(\sigma_0^2+1)^2}{b_0\sigma_0^2(\sigma_0^2+1) + \sigma_0^2(y-\mu_0)^2} \right) \left( \mu - \frac{\mu_0 + y\sigma_0^2}{1+\sigma_0^2} \right) \right]^{-\left(\frac{a_0}{2}+1\right)}}{\Gamma\left(\frac{a_0+1}{2}\right) \sqrt{\pi(a_0+1)} \cdot \frac{b_0\sigma_0^2(\sigma_0^2+1) + \sigma_0^2(y-\mu_0)^2}{(a_0+1)(\sigma_0^2+1)^2}}
\end{aligned}$$

w/ support  $(-\infty, \infty)$

3a.  $X \sim \text{Bin}(50, p)$ ; observe  $x=30$ . Derive the posterior distribution for the prior

$$p \sim \frac{1}{2} \{ \text{Beta}(10, 20) + \text{Beta}(20, 10) \}$$

The prior PDF is

$$\begin{aligned}\pi(p) &= \frac{1}{2} \left( \frac{1}{B(10, 20)} p^{10-1} (1-p)^{20-1} \right) + \frac{1}{2} \left( \frac{1}{B(20, 10)} p^{20-1} (1-p)^{10-1} \right), p \in [0, 1] \\ &= \frac{1}{2} \cdot \frac{1}{B(10, 20)} p^9 (1-p)^{19} + \frac{1}{2} \cdot \frac{\Gamma(20+10)}{\Gamma(20)\Gamma(10)} p^{19} (1-p)^9 \\ &= \frac{1}{2} \cdot \frac{1}{B(10, 20)} p^9 (1-p)^{19} + \frac{1}{2} \cdot \frac{1}{B(10, 20)} p^{19} (1-p)^9 \\ &= \frac{1}{2B(10, 20)} [p^9 (1-p)^{19} + p^{19} (1-p)^9]\end{aligned}$$

The likelihood is

$$f(x|p) = \binom{50}{x} p^x (1-p)^{50-x}$$

So, the posterior PDF is

$$\begin{aligned}f(p|x) &\propto \pi(p) f(x|p) \\ &= \frac{1}{2B(10, 20)} [p^9 (1-p)^{19} + p^{19} (1-p)^9] \cdot \binom{50}{x} p^x (1-p)^{50-x} \\ &\propto p^x (1-p)^{50-x} \left[ \frac{1}{2} p^9 (1-p)^{19} + \frac{1}{2} p^{19} (1-p)^9 \right] \\ &= \frac{1}{2} p^{9+x} (1-p)^{50-x+19} + \frac{1}{2} p^{19+x} (1-p)^{50-x+9} \\ &= \frac{1}{2} p^{9+x} (1-p)^{69-x} + \frac{1}{2} p^{19+x} (1-p)^{59-x},\end{aligned}$$

which is the average of the kernels of two Beta distributions, so

$$\begin{aligned}p|x &\sim \frac{1}{2} \text{Beta}(9+x+1, 69-x+1) + \frac{1}{2} \text{Beta}(19+x+1, 59-x+1) \\ &= \frac{1}{2} \text{Beta}(10+x, 70-x) + \frac{1}{2} \text{Beta}(20+x, 60-x)\end{aligned}$$

Since we observe  $x=30$ ,

$$\begin{aligned}p|x=30 &\sim \frac{1}{2} \text{Beta}(10+30, 70-30) + \frac{1}{2} \text{Beta}(20+30, 60-30) \\ \Rightarrow p|x=30 &\sim \frac{1}{2} \text{Beta}(40, 40) + \frac{1}{2} \text{Beta}(50, 30)\end{aligned}$$

So, the posterior PDF of  $p$  is

$$f(p|x=30) = \frac{1}{2B(40, 40)} p^{39} (1-p)^{39} + \frac{1}{2B(50, 30)} p^{49} (1-p)^{29}$$

with support  $p \in [0, 1]$ .



3b. In general, for data  $(x, n)$  w/  $n$  fixed &  $x \sim \text{Bin}(n, p)$ , if the prior is  $p \sim \sum_{i=1}^K w_i \text{Beta}(a_i, b_i)$  w/  $\sum_{i=1}^K w_i = 1$ ,  $w_i > 0 \forall i$ , derive the posterior distribution in closed form.

Now the prior PDF is

$$\pi(p) = \sum_{i=1}^K w_i \cdot \frac{1}{B(a_i, b_i)} p^{a_i-1} (1-p)^{b_i-1}$$

The likelihood function is

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

So, the posterior PDF is

$$\begin{aligned} f(p|x) &\propto \pi(p) f(x|p) \\ &= \sum_{i=1}^K w_i \cdot \frac{1}{B(a_i, b_i)} p^{a_i-1} (1-p)^{b_i-1} \cdot \binom{n}{x} p^x (1-p)^{n-x} \\ &\propto p^x (1-p)^{n-x} \sum_{i=1}^K w_i \frac{1}{B(a_i, b_i)} p^{a_i-1} (1-p)^{b_i-1} \\ &= \sum_{i=1}^K w_i \frac{1}{B(a_i, b_i)} p^{a_i+x-1} (1-p)^{b_i+n-x-1} \end{aligned}$$

which is a weighted average of the PDFs of Beta distributions, so

$$p|x \sim \sum_{i=1}^K w_i \text{Beta}(a_i+x, b_i+n-x)$$

So, the posterior PDF of  $p$  is

$$f(p|x) = \sum_{i=1}^K w_i \cdot \frac{1}{B(a_i+x, b_i+n-x)} p^{a_i+x-1} (1-p)^{b_i+n-x-1}$$

with support  $p \in [0, 1]$

4. Bayesian linear regression.  $\mathbf{Y} = (Y_1, \dots, Y_n)$ ,

$\mathbf{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$ ,  $\beta = (\beta_0, \beta_1, \dots, \beta_{k-1})$ ,

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1,k-1} \\ 1 & X_{21} & \dots & X_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{n,k-1} \end{bmatrix}.$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{k-1} X_{ik-1} + \varepsilon_i, \quad \varepsilon_i \sim \text{iid } N(0, \sigma^2)$$

$$\text{OLS: } \hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}, \quad \hat{\sigma}^2 = \frac{1}{n-k} (\mathbf{y} - \mathbf{X}\hat{\beta})^\top (\mathbf{y} - \mathbf{X}\hat{\beta}).$$

Improper priors:  $p(\beta) \propto 1, \quad \beta \in \mathbb{R}^k$

$$p(\sigma^2) \propto \frac{1}{\sigma}$$

Derive  $p(\beta | \mathbf{X}, \mathbf{Y})$  +  $p(\sigma^2 | \mathbf{X}, \mathbf{Y})$  up to normalizing constant  
+ name them.

Recall from lecture notes that we may write the likelihood function as

$$\begin{aligned} L(\beta, \sigma^2 | \mathbf{Y}, \mathbf{X}) &\propto \sigma^{-n} \exp \left\{ -\frac{1}{2\sigma^2} [\hat{\sigma}^2(n-k) + (\beta - \hat{\beta})^\top \mathbf{X}^\top \mathbf{X} (\beta - \hat{\beta})] \right\} \\ &= (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \hat{\sigma}^2 (n-k) \right\} \exp \left\{ -\frac{1}{2\sigma^2} (\beta - \hat{\beta})^\top \mathbf{X}^\top \mathbf{X} (\beta - \hat{\beta}) \right\} \end{aligned}$$

The joint prior for  $(\beta, \sigma^2)$  is

$$p(\beta, \sigma^2) = p(\beta | \sigma^2) p(\sigma^2)$$

$$= p(\beta) p(\sigma^2) \quad \text{since the marginal priors}$$

$$\propto 1 \cdot \frac{1}{\sigma}$$

indicate  $\beta \perp\!\!\!\perp \sigma^2$

Then

$$\begin{aligned} p(\beta, \sigma^2 | \mathbf{Y}, \mathbf{X}) &\propto p(\beta, \sigma^2) L(\beta, \sigma^2 | \mathbf{Y}, \mathbf{X}) \\ &= \frac{1}{\sigma} \cdot (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \hat{\sigma}^2 (n-k) \right\} \exp \left\{ -\frac{1}{2\sigma^2} (\beta - \hat{\beta})^\top \mathbf{X}^\top \mathbf{X} (\beta - \hat{\beta}) \right\} \\ &= (\sigma^2)^{-(n+1)/2} \exp \left\{ -\frac{1}{2\sigma^2} \hat{\sigma}^2 (n-k) \right\} \exp \left\{ -\frac{1}{2\sigma^2} (\beta - \hat{\beta})^\top \mathbf{X}^\top \mathbf{X} (\beta - \hat{\beta}) \right\} \end{aligned}$$

So,

$$\begin{aligned} p(\sigma^2 | \mathbf{Y}, \mathbf{X}) &\propto \int_{\mathbb{R}^k} p(\beta, \sigma^2 | \mathbf{Y}, \mathbf{X}) d\beta \\ &= \int_{\mathbb{R}^k} (\sigma^2)^{-(n+1)/2} \exp \left\{ -\frac{1}{2\sigma^2} \hat{\sigma}^2 (n-k) \right\} \exp \left\{ -\frac{1}{2\sigma^2} (\beta - \hat{\beta})^\top \mathbf{X}^\top \mathbf{X} (\beta - \hat{\beta}) \right\} d\beta \\ &= (\sigma^2)^{-(n+1)/2} \exp \left\{ -\frac{1}{2\sigma^2} \hat{\sigma}^2 (n-k) \right\} \int_{\mathbb{R}^k} \exp \left\{ -\frac{1}{2\sigma^2} (\beta - \hat{\beta})^\top \mathbf{X}^\top \mathbf{X} (\beta - \hat{\beta}) \right\} d\beta \end{aligned}$$

$$= (\sigma^2)^{-\frac{(n+1)}{2}} \sqrt{\frac{\pi^k}{\det(\frac{1}{2\sigma^2} X^T X)}} \exp\left\{-\frac{1}{2\sigma^2} \hat{\sigma}^2(n-k)\right\}$$

$$= (\sigma^2)^{-\frac{(n+1)}{2}} \sqrt{\frac{\pi^k}{(\frac{1}{2\sigma^2})^k \det(X^T X)}} \exp\left\{-\frac{1}{2\sigma^2} \hat{\sigma}^2(n-k)\right\}$$

$$= (\sigma^2)^{-\frac{(n+1)}{2}} \sqrt{\frac{(2\pi)^k (\sigma^2)^k}{\det(X^T X)}} \exp\left\{-\frac{1}{2\sigma^2} \hat{\sigma}^2(n-k)\right\}$$

$$\propto (\sigma^2)^{-(n+1)/2} \cdot (\sigma^2)^{k/2} \exp\left\{-\frac{1}{2\sigma^2} \hat{\sigma}^2(n-k)\right\}$$

$$= (\sigma^2)^{\frac{1}{2}(-n-1+k)} \exp\left\{-\frac{1}{2\sigma^2} \hat{\sigma}^2(n-k)\right\}$$

$$= (\sigma^2)^{-(n-k-1)/2+1} \exp\left\{-\frac{1}{2\sigma^2} \hat{\sigma}^2(n-k)\right\}$$

$$= (\sigma^2)^{-(1+\frac{n-k-1}{2})} \exp\left\{-\frac{(n-k-1)(n-k)\hat{\sigma}^2}{2\sigma^2(n-k-1)}\right\},$$

which is the PDF of an inverse-scaled  $\chi^2$

distribution, so

$$\sigma^2 | Y, X \sim \text{IS} \chi^2(n-k-1, \frac{(n-k)\hat{\sigma}^2}{n-k-1})$$

Also, letting  $\tau := \frac{1}{\sigma^2}$

$$p(\beta | Y, X) \propto \int_0^\infty (\sigma^2)^{-(n+1)/2} \exp\left\{-\frac{1}{2\sigma^2} \hat{\sigma}^2(n-k)\right\} \exp\left\{-\frac{1}{2\sigma^2} (\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})\right\} d\sigma^2$$

$$\begin{aligned} & \left( \begin{aligned} \tau &= \frac{1}{\sigma^2} = (\sigma^2)^{-1} \\ d\tau &= \frac{1}{\sigma^4} d\sigma^2 \Rightarrow d\sigma^2 = -\sigma^4 d\tau = -\tau^{-2} d\tau \end{aligned} \right) \quad \sigma^2 \rightarrow 0 \Rightarrow \tau \rightarrow \infty \\ & \sigma^2 \rightarrow \infty \Rightarrow \tau \rightarrow 0 \end{aligned}$$

$$= \int_0^\infty \tau^{-(n+1)/2} \exp\left\{-\frac{\tau}{2} [\hat{\sigma}^2(n-k) + (\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})]\right\} \cdot -\tau^{-2} d\tau$$

$$= \int_0^\infty \tau^{(n+1)/2-2} \exp\left\{-\frac{\tau}{2} [\hat{\sigma}^2(n-k) + (\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})]\right\} d\tau$$

$$= \int_0^\infty \underbrace{\tau^{\frac{1}{2}(n-3)}}_{:= \tau^{\beta}} \exp\left\{-\tau \cdot \frac{1}{2} [\hat{\sigma}^2(n-k) + (\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})]\right\} d\tau$$

$$:= \alpha$$

$$= \frac{\Gamma(\frac{1}{2}(n-3)+1)}{[\frac{1}{2} \hat{\sigma}^2(n-k) + \frac{1}{2} (\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})]^{\frac{1}{2}(n-3)+1}}$$

$$\propto \left[ \frac{\hat{\sigma}^2(n-k)}{2} + \frac{(\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})}{2} \right]^{-(\frac{n}{2} - \frac{3}{2} + 1)}$$

$$= \left[ \frac{\hat{\sigma}^2(n-k)}{2} \left( 1 + \frac{(\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})}{\hat{\sigma}^2(n-k)} \right) \right]^{-(\frac{n}{2} - \frac{3}{2})}$$

$$\propto \left[ 1 + \frac{(\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})}{\hat{\sigma}^2(n-k)} \right]^{-(n-1)/2}$$

$$= \left[ 1 + \frac{(n-k-1)}{(n-k-1)} \frac{(\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})}{\hat{\sigma}^2(n-k)} \right]^{-\frac{1}{2}[(n-k-1)+k]}$$

$$= \left[ 1 + \frac{1}{n-k-1} \frac{(\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})}{\hat{\sigma}^2(n-k)/(n-k-1)} \right]^{-\frac{1}{2}[(n-k-1)+k]}$$

which is the kernel of a multivariate t distribution, so

$$\beta | Y, X \sim MVT\left(\hat{\beta}, \frac{\hat{\sigma}^2(X^T X)^{-1}(n-k)}{n-k-1}, n-k-1\right)$$

5. n identical light bulbs. After L hours of lighting, k have gone out w/ lighting times  $y_1, y_2, \dots, y_k$ , all  $< L$ . Remaining  $(n-k)$  bulbs still on, but expt ends.  $y_i \sim \text{iid Exp}(\theta)$ .  $Z_i := \mathbb{I}(y_i > L)$ . Write likelihood fcn, as a fcn of  $\theta$ , for observed data  $(y_1, z_1), \dots, (y_n, z_n)$ . Find a conjugate prior for  $\theta$  + its posterior.

- Since  $y_i \sim \text{iid Exp}(\theta)$ , each  $y_i$  has PDF

$$f(y) = \begin{cases} \theta e^{-\theta y}, & y \in [0, \infty) \\ 0, & \text{o/w} \end{cases}$$

- We decompose f into survival + hazard functions:

$$f(y) = \theta \cdot e^{-\theta y}$$

$$= \lambda(y) \cdot S(y),$$

$$\text{where } \lambda(y) = \theta, S(y) = e^{-\theta y}.$$

- Then the likelihood function is

$$\begin{aligned} \mathcal{L}(\theta | y_i's, z_i's) &= \prod_{i=1}^n \lambda(y_i)^{1-z_i} S(y_i) \\ &= \prod_{i=1}^n \theta^{1-z_i} e^{-\theta y_i} \\ &= \theta^{\sum_{i=1}^n (1-z_i)} e^{\sum_{i=1}^n -\theta y_i} \\ &= \theta^{n - \sum_{i=1}^n z_i} e^{-\theta \sum_{i=1}^n y_i} \\ &= \theta^{n - (n-k)} e^{-\theta \sum_{i=1}^n y_i} \end{aligned}$$

$$\Rightarrow \mathcal{L}(\theta | y_i's, z_i's) = \theta^k \exp(-\theta \sum_{i=1}^n y_i)$$

- If we treat  $y_i = L$  for those bulbs that were censored, we can further simplify:

$$\begin{aligned} \mathcal{L}(\theta | y_i's, z_i's) &= \theta^k e^{-\theta [\sum_{i=1}^k y_i + \sum_{i=k+1}^n L]} \\ &= \theta^k e^{-\theta \sum_{i=1}^k y_i} e^{-\theta L \sum_{i=k+1}^n} \end{aligned}$$

$$\mathcal{L}(\theta | y_i's, z_i's) = \theta^k \exp(-\theta \sum_{i=1}^k y_i) \exp(-\theta L(n-k))$$

- By inspection, suppose  $\Theta \sim \text{Gamma}(\alpha, \beta)$ , so

$$\pi(\theta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, & \theta \in (0, \infty) \\ 0, & \text{o/w} \end{cases}$$

- Then

$$\begin{aligned} p(\theta | y_i's, z_i's) &\propto \pi(\theta) L(\theta | y_i's, z_i's) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \cdot \theta^k e^{-\theta \sum_{i=1}^n y_i} \\ &\propto \theta^{\alpha+k-1} e^{-(\beta + \sum_{i=1}^n y_i)\theta}, \end{aligned}$$

so

$$\theta | y_i's, z_i's \sim \text{Gamma}(\alpha+k, \beta + \sum_{i=1}^n y_i).$$

- Since the posterior distribution of  $\theta$  is also Gamma, a conjugate prior for  $\theta$  is  $\theta \sim \text{Gamma}(\alpha, \beta)$ , w/ posterior distribution  $\theta | y_i's, z_i's \sim \text{Gamma}(\alpha+k, \beta + \sum_{i=1}^n y_i)$

6. N race cars.  $N \sim \text{Geom}(\lambda)$  w/  $E[N] = 100$ :

$$h(N) = \frac{1}{100} \left( \frac{99}{100} \right)^{N-1},$$

$N \in \mathbb{N}$ . Observe  $x = 203$ ,

$$f(x|N) = \begin{cases} \frac{1}{N}, & x \in \{1, \dots, N\} \\ 0, & x > N \end{cases}$$

6a. Find  $h(N|x)$ ,  $E[N|x]$ ,  $\text{sd}(N|x)$ .

$$\begin{aligned} h(N|x) &\propto f(x|n)h(N) \\ &= \begin{cases} \frac{1}{N} \cdot \frac{1}{100} \left( \frac{99}{100} \right)^{N-1}, & x \leq N \\ 0 \cdot \frac{1}{100} \left( \frac{99}{100} \right)^{N-1}, & x > N \end{cases} \\ &\propto \begin{cases} \frac{1}{N} \left( \frac{99}{100} \right)^{N-1}, & N \geq x \\ 0, & N < x \end{cases} \\ &= \begin{cases} \frac{c}{N} \left( \frac{99}{100} \right)^{N-1}, & N \geq x \\ 0, & N < x \end{cases} \quad \text{for some } c \in \mathbb{R}. \end{aligned}$$

In particular,

$$h(N|x=203) = \begin{cases} \frac{c}{N} \left( \frac{99}{100} \right)^{N-1}, & N \geq 203 \\ 0, & N < 203 \end{cases}.$$

$c$  must solve  $\sum_{n=203}^{\infty} \frac{c}{n} \left( \frac{99}{100} \right)^{n-1} = 1$ .

$$\begin{aligned} \sum_{n=203}^{\infty} \frac{c}{n} \left( \frac{99}{100} \right)^{n-1} &= c \sum_{n=203}^{\infty} \frac{1}{n} \left( \frac{99}{100} \right)^n \left( \frac{99}{100} \right)^{-1} \\ &= \frac{100}{99} c \sum_{n=203}^{\infty} \frac{1}{n} \left( \frac{99}{100} \right)^n \end{aligned}$$

$$\begin{aligned} \text{Note that } \int_0^{0.99} \sum_{n=203}^{\infty} t^{n-1} dt &= \sum_{n=203}^{\infty} \int_0^{0.99} t^{n-1} dt \quad (\text{by Fubini's/Tonelli's Thms}) \\ &= \sum_{n=203}^{\infty} \frac{1}{n} t^n \Big|_0^{0.99} \\ &= \sum_{n=203}^{\infty} \left( \frac{1}{n} \left( \frac{99}{100} \right)^n - \frac{1}{n}(0) \right) \\ &= \sum_{n=203}^{\infty} \frac{1}{n} \left( \frac{99}{100} \right)^n. \end{aligned}$$

But for  $t \in [0, \frac{99}{100}]$ ,

$$\begin{aligned} \sum_{n=203}^{\infty} t^{n-1} &= \sum_{m=0}^{\infty} t^{m+203-1} \\ &= \sum_{m=0}^{\infty} t^m t^{202} \\ &= \frac{t^{202}}{1-t} \end{aligned}$$



• So,  $\sum_{n=203}^{\infty} \frac{1}{n} \left(\frac{99}{100}\right)^n = \int_0^{0.99} \sum_{n=203}^{\infty} t^{n-1} dt$   
 $= \int_0^{0.99} \frac{t^{202}}{1-t} dt$

$$\approx 0.047 \text{ (via WolframAlpha).}$$

• Thus,  $\sum_{n=203}^{\infty} \frac{1}{n} \left(\frac{99}{100}\right)^{n-1} \approx \frac{100}{99} C (0.047) \approx 1$   
 $\Rightarrow C \approx \frac{99}{100} / 0.047$   
 $\Rightarrow C \approx 21.253$

• So,

$$h(N|x=203) \approx \begin{cases} \frac{21.253}{N} \left(\frac{99}{100}\right)^{N-1}, & N \geq 203 \\ 0, & \text{o/w} \end{cases}$$

•  $E[N|x=203] = \sum_{n=203}^{\infty} n h(n|x=203)$   
 $\approx \sum_{n=203}^{\infty} n \cdot \frac{21.253}{n} \left(\frac{99}{100}\right)^{n-1}$   
 $= 21.253 \sum_{n=203}^{\infty} \left(\frac{99}{100}\right)^{n-1}$   
 $= 21.253 \sum_{m=0}^{\infty} \left(\frac{99}{100}\right)^{203+m-1}$   
 $= 21.253 \sum_{m=0}^{\infty} \left(\frac{99}{100}\right)^m \left(\frac{99}{100}\right)^{202}$   
 $= 21.253 \left(\frac{99}{100}\right)^{202} \sum_{m=0}^{\infty} \left(\frac{99}{100}\right)^m$   
 $= 21.253 \left(\frac{99}{100}\right)^{202} \cdot \frac{1}{1 - \frac{99}{100}}$

$\Rightarrow E[N|x=203] \approx 279.089$

•  $Var(N|x=203) = E[N^2|x=203] - E[N|x=203]^2$

•  $E[N^2|x=203] = \sum_{n=203}^{\infty} n^2 h(n|x=203)$   
 $\approx \sum_{n=203}^{\infty} n^2 \cdot \frac{1}{n} (21.253) \left(\frac{99}{100}\right)^{n-1}$   
 $= 21.253 \sum_{n=203}^{\infty} n \left(\frac{99}{100}\right)^{n-1}$

• Note that  $\sum_{n=203}^{\infty} n \left(\frac{99}{100}\right)^{n-1} = \sum_{m=0}^{\infty} (m+203) \left(\frac{99}{100}\right)^{203+m-1}$   
 $= \sum_{m=0}^{\infty} (m+203) \left(\frac{99}{100}\right)^{m-1} \left(\frac{99}{100}\right)^{203}$   
 $= \left(\frac{99}{100}\right)^{203} \left[ \sum_{m=0}^{\infty} m \left(\frac{99}{100}\right)^{m-1} + 203 \sum_{m=0}^{\infty} \left(\frac{99}{100}\right)^m \right]$   
 $= \left(\frac{99}{100}\right)^{203} \left[ \sum_{m=0}^{\infty} m \left(\frac{99}{100}\right)^{m-1} + 203 \cdot \frac{1}{1 - \frac{99}{100}} \right]$

• Now note that for  $x \in (0, 1)$ ,

$$\sum_{m=0}^{\infty} m x^{m-1} = \sum_{m=0}^{\infty} \frac{d}{dx} x^m$$
 $= \frac{d}{dx} \sum_{m=0}^{\infty} x^m$ 
 $= \frac{d}{dx} \left( \frac{1}{1-x} \right)$

since the series converges uniformly

$$= \frac{1}{(1-x)^2},$$

$$\text{So } \sum_{m=0}^{\infty} m \left(\frac{99}{100}\right)^{m-1} = \frac{1}{(1-0.99)^2} = 10,000,$$

$$\text{So } \sum_{n=203}^{\infty} n \left(\frac{99}{100}\right)^{n-1} = \left(\frac{99}{100}\right)^{203} [10,000 + 20,300] \\ \approx 3965.667,$$

$$\text{So } E[N^2|x=203] \approx (21.253)(3965.667) \\ \approx 84,284.730.$$

. Thus,

$$\text{Var}(N|x=203) \approx 84,284.730 - (279.089)^2 \\ \Rightarrow \text{Var}(N|x=203) \approx 6394.333$$

$$\Rightarrow \text{sd}(N|x=203) \approx \sqrt{6394.333}$$

$$\Rightarrow \text{sd}(N|x=203) \approx 79.965$$

7. Poisson sampling model  $f(n|\lambda) = \text{Pois}(\lambda)$ . Find Jeffrey's prior for  $\lambda$ .

$$\cdot f(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!}$$

$$\cdot \Rightarrow \log f(n|\lambda) = \log\left(\frac{\lambda^n e^{-\lambda}}{n!}\right)$$

$$\begin{aligned}&= \log(\lambda^n) + \log(e^{-\lambda}) - \log(n!) \\&= n \log \lambda - \lambda - \log(n!)\end{aligned}$$

$$\cdot \Rightarrow \frac{\partial \log f(n|\lambda)}{\partial \lambda} = \frac{n}{\lambda} - 1$$

$$\cdot \Rightarrow \frac{\partial^2 \log f(n|\lambda)}{\partial \lambda^2} = -\frac{n}{\lambda^2}$$

$$\cdot \Rightarrow I(\lambda) = -\mathbb{E}\left[\frac{\partial^2 \log f(n|\lambda)}{\partial \lambda^2}\right]$$

$$= -\mathbb{E}\left[-\frac{n}{\lambda^2}\right]$$

$$= \frac{1}{\lambda^2} \mathbb{E}[n]$$

$$= \frac{1}{\lambda^2} \cdot \lambda$$

$$= \frac{1}{\lambda}$$

$$\cdot \Rightarrow \pi_J(\lambda) \propto I(\lambda)^{1/2}$$
$$= (1/\lambda)^{1/2}$$

Thus, Jeffrey's prior is

$$\pi_J(\lambda) \propto \frac{1}{\sqrt{\lambda}}$$

Note that this is an improper prior.

8. Con.  $\Theta$  is prob of heads.  $H_1: \Theta < \frac{1}{2}$ ,  $H_2: \Theta = \frac{1}{2}$ ,  
 $H_3: \Theta > \frac{1}{2}$ .  $p(H_2) = \frac{1}{2}$ ,  $p(H_1) = p(H_3) = \frac{1}{4}$ .  
 $\Theta | H_1 \sim U(0, \frac{1}{2})$ ,  $\Theta | H_3 \sim U(\frac{1}{2}, 1)$ . Observe  $n=1$  toss.  
 $x_1 \in \{0, 1\}$  indicator of heads.

8a. Find  $p(x_1=1 | H_1)$ ,  $p(x_1=1 | H_2)$ ,  $p(x_1=1 | H_3)$ .

$$\cdot h(\Theta | H_1) = \begin{cases} \frac{1}{\frac{1}{2}-0}, & \Theta \in (0, \frac{1}{2}) \\ 0, & \text{o/w} \end{cases} = \begin{cases} 2, & \Theta \in (0, \frac{1}{2}) \\ 0, & \text{o/w} \end{cases}$$

$$\cdot h(\Theta | H_2) = \begin{cases} 1, & \Theta = \frac{1}{2} \\ 0, & \text{o/w} \end{cases}$$

$$\cdot h(\Theta | H_3) = \begin{cases} \frac{1}{1-\frac{1}{2}}, & \Theta \in (\frac{1}{2}, 1) \\ 0, & \text{o/w} \end{cases} = \begin{cases} 2, & \Theta \in (\frac{1}{2}, 1) \\ 0, & \text{o/w} \end{cases}$$

$$\begin{aligned} \cdot p(x_1=1 | H_1) &= \int_{\Theta \in H_1} p(x_1=1 | \Theta) h(\Theta | H_1) d\Theta \\ &= \int_0^{1/2} \Theta \cdot 2 d\Theta \\ &= 2 \cdot \frac{1}{2} \Theta^2 \Big|_0^{1/2} \\ &= \left(\frac{1}{2}\right)^2 - 0^2 \\ &= \frac{1}{4} \end{aligned}$$

$$\begin{aligned} \cdot p(x_1=0 | H_1) &= \int_{\Theta \in H_1} p(x_1=0 | \Theta) h(\Theta | H_1) d\Theta \\ &= \int_0^{1/2} (1-p(x_1=1 | \Theta)) 2 d\Theta \\ &= 2 \int_0^{1/2} (1-\Theta) d\Theta \\ &= 2 \left(\Theta - \frac{1}{2}\Theta^2\right) \Big|_0^{1/2} \\ &= 2 \left[\left(\frac{1}{2} - \frac{1}{2}\left(\frac{1}{2}\right)^2\right) - (0 - \frac{1}{2}0^2)\right] \\ &= 2 \left[\frac{1}{2} - \frac{1}{8}\right] \\ &= \frac{3}{4} \end{aligned}$$

$$\cdot \text{So, } p(x_1 | H_1) = \begin{cases} \frac{1}{4}, & x_1=1 \\ \frac{3}{4}, & x_1=0 \end{cases}$$

and in particular,  $p(x_1=1 | H_1) = \frac{1}{4}$

$$\begin{aligned} \cdot p(x_1=1 | H_2) &= p(x_1=1 | \theta=\frac{1}{2}) h(\theta=\frac{1}{2} | H_2) \\ &= (\frac{1}{2})(1) \\ &= \frac{1}{2} \end{aligned}$$

$$\begin{aligned} \cdot p(x_1=0 | H_2) &= p(x_1=0 | \theta=\frac{1}{2}) h(\theta=\frac{1}{2} | H_2) \\ &= (1 - p(x_1=1 | \theta=\frac{1}{2}))(1) \\ &= (1 - \frac{1}{2}) \\ &= \frac{1}{2} \end{aligned}$$

• So,

$$p(x_1 | H_2) = \begin{cases} \frac{1}{2}, & x_1=1 \\ \frac{1}{2}, & x_1=0 \end{cases}$$

and in particular,

$$p(x_1=1 | H_2) = \frac{1}{2}$$

$$\begin{aligned} \cdot p(x_1=1 | H_3) &= \int_{\theta \in H_3} p(x_1=1 | \theta) h(\theta | H_3) d\theta \\ &= \int_{1/2}^1 \theta \cdot 2d\theta \\ &= 2 \cdot \frac{1}{2} \theta^2 \Big|_{1/2}^1 \\ &= (1^2 - (\frac{1}{2})^2) \\ &= \frac{3}{4} \end{aligned}$$

$$\begin{aligned} \cdot p(x_1=0 | H_3) &= \int_{\theta \in H_3} p(x_1=0 | \theta) h(\theta | H_3) d\theta \\ &= \int_{1/2}^1 (1 - p(x_1=1 | \theta)) 2d\theta \\ &= 2 \int_{1/2}^1 (1 - \theta) d\theta \\ &= 2 (\theta - \frac{1}{2}\theta^2) \Big|_{1/2}^1 \\ &= 2 [(1 - \frac{1}{2}(1)^2) - (\frac{1}{2} - \frac{1}{2}(\frac{1}{2})^2)] \\ &= 2 [\frac{1}{2} - \frac{1}{2} + \frac{1}{8}] \\ &= \frac{1}{4} \end{aligned}$$

• So,

$$p(x_1 | H_3) = \begin{cases} \frac{3}{4}, & x_1=1 \\ \frac{1}{4}, & x_1=0 \end{cases}$$

and in particular,

$$p(x_1=1 | H_3) = \frac{3}{4}$$

8b. Find  $p(x_1=1)$ .

• By the Law of Total Probability,

$$\begin{aligned} p(x_1=1) &= p(x_1=1|H_1)p(H_1) + p(x_1=1|H_2)p(H_2) \\ &\quad + p(x_1=1|H_3)p(H_3) \\ &= \frac{1}{4}\left(\frac{1}{4}\right) + \frac{1}{2}\left(\frac{1}{2}\right) + \left(\frac{3}{4}\right)\left(\frac{1}{4}\right) \\ &= \frac{1}{16} + \frac{1}{4} + \frac{3}{16} \\ &= \frac{1}{2} \end{aligned}$$

• Similarly,

$$\begin{aligned} p(x_1=0) &= p(x_1=0|H_1)p(H_1) + p(x_1=0|H_2)p(H_2) \\ &\quad + p(x_1=0|H_3)p(H_3) \\ &= \frac{3}{4}\left(\frac{1}{4}\right) + \frac{1}{2}\left(\frac{1}{2}\right) + \frac{1}{4}\left(\frac{1}{4}\right) \\ &= \frac{3}{16} + \frac{1}{4} + \frac{1}{16} \\ &= \frac{1}{2} \end{aligned}$$

• So,

$$P(x_1) = \begin{cases} \frac{1}{2}, & x_1=1 \\ \frac{1}{2}, & x_1=0 \end{cases}$$

and in particular,

$$P(x_1=1) = \frac{1}{2}$$

9. Rayleigh dist. w/ PDF  $f(x|\delta) = \delta x e^{-\delta x^2/2} I_{(0,\infty)}(x)$ .  
 $x = \{x_1, \dots, x_n\}$  is a realization of a random samp.  
from this model. Prior dist. of  $\delta$  is  $\text{Gamma}(a, b)$ .

9a. Find posterior  $h(\delta|x)$ . Name?  $E[\delta|x]$ ?  $\text{Var}(\delta|x)$ ?

- $\pi(\delta) = \frac{b^a}{\Gamma(a)} \delta^{a-1} e^{-b\delta}$
- $h(\delta|x) \propto \prod_{i=1}^n \delta x_i e^{-\delta x_i^2/2} \cdot \frac{b^a}{\Gamma(a)} \delta^{a-1} e^{-b\delta}$   
 $= \delta^n \cdot \prod_{i=1}^n x_i \cdot e^{-\frac{\delta}{2} \sum_{i=1}^n x_i^2} \cdot \frac{b^a}{\Gamma(a)} \delta^{a-1} e^{-b\delta}$   
 $\propto \delta^{n+a-1} e^{-b\delta - \frac{1}{2}\delta \sum_{i=1}^n x_i^2}$   
 $= \delta^{n+a-1} e^{-\delta(b + \frac{1}{2} \sum_{i=1}^n x_i^2)}$

• Thus,

$$\delta|x \sim \text{Gamma}(a+n, b + \frac{1}{2} \sum_{i=1}^n x_i^2)$$

and

$$h(\delta|x) = \frac{(b + \frac{1}{2} \sum_{i=1}^n x_i^2)^{a+n}}{\Gamma(a+n)} \delta^{a+n-1} e^{-\delta(b + \frac{1}{2} \sum_{i=1}^n x_i^2)}$$

• Using the moments of the Gamma distribution,

$$E[\delta|x] = \frac{a+n}{b + \frac{1}{2} \sum_{i=1}^n x_i^2}$$

and

$$\text{Var}(\delta|x) = \frac{a+n}{(b + \frac{1}{2} \sum_{i=1}^n x_i^2)^2}$$

9b. Find Jeffrey's prior & corresponding posterior dist conditional on one obs.  $x$  from Rayleigh model.

$$\cdot f(x|\delta) = \delta x e^{-\delta x^2/2}$$

$$\Rightarrow \log f(x|\delta) = \log(\delta x e^{-\delta x^2/2}) \\ = \log \delta + \log x - \frac{1}{2} \delta x^2$$

$$\Rightarrow \frac{\partial \log f(x|\delta)}{\partial \delta} = \frac{1}{\delta} - \frac{1}{2} x^2$$

$$\Rightarrow \frac{\partial^2 \log f(x|\delta)}{\partial \delta^2} = \frac{-1}{\delta^2}$$

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

$$= - \mathbb{E}\left[-\frac{1}{\delta^2}\right]$$

$$= \frac{1}{\delta^2}$$

$$\Rightarrow h_J(\delta) = (1/\delta^2)^{1/2} = \frac{1}{\delta}$$

That is, Jeffrey's prior is

$$h_J(\delta) = \frac{1}{\delta}$$

Note that this is an improper prior.

$$\begin{aligned} \cdot f(\delta|x) &\propto h_J(\delta) \cdot f(x|\delta) \\ &= \frac{1}{\delta} \cdot \delta x e^{-\delta x^2/2} \\ &= x e^{-\delta x^2/2} \\ &\propto e^{-\frac{1}{2} x^2 \delta} \end{aligned}$$

Thus,  $\delta|x \sim \text{Exp}(\frac{1}{2}x^2)$ , and

$$f(\delta|x) = \frac{1}{2} x^2 e^{-\frac{1}{2} x^2 \delta}$$