Outline

- Exponential Likelihood, Gamma Prior, & Prof Smedley
- Prior Influence on Posterior
- Comparing 95% intervals
- Conjugate Priors
- Non-Conjugate Priors & WinBUGS
- Normal Model: Estimate Mean, Variance Known
- Shrinkage and the Minnesota Radon Example
- Start reading Lynch Ch 4 – on course website!
Exponential Distribution

- Last week we saw that if \( x_1, ..., x_n \sim \text{Expon}(\lambda) \), i.e. each \( x_i \) is indep., with density \( f(x|\lambda) = \lambda e^{-\lambda x} \), then

\[
L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} = \lambda^n e^{-\lambda n\bar{x}}
\]

\[
\hat{\lambda}_{MLE} = n/\sum_{i=1}^{n} x_i = 1/\bar{X}
\]

and

\[
SE(\hat{\lambda}) = \sqrt{1/I(\hat{\lambda})} = \sqrt{\hat{\lambda}^2/n} = \hat{\lambda}/\sqrt{n}
\]

MLE point estimate and interval for Prof. Smedley

- We saw that three randomly-chosen students from Prof. Smedley’s class took \( x_1 = 3 \), \( x_2 = 10 \) and \( x_3 = 8 \) days to learn boxplots

- The MLE and SE are:

\[
\hat{\lambda} = 1/\bar{x} = 1/7 = 0.014
\]

\[
I(\lambda) = 3/\lambda^2
\]

\[
SE(\hat{\theta}) = \hat{\lambda}/\sqrt{3} = (1/7)/\sqrt{3} = 0.082
\]

- and so a rough 95% confidence interval for \( \lambda \) would be \([1/7 - 2(1/7)/\sqrt{3}, 1/7 + 2(1/7)/\sqrt{3}]\), or \((-0.022, 0.308)\)
How does this look as a Bayesian problem?

- Still have \( L(\lambda) = \lambda^n e^{-\lambda n \bar{x}} \)
- Need a prior distribution for \( \lambda \). We’ll start with a gamma distribution:

\[
\text{Gamma}(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda \beta}
\]

- Since (posterior) \( \propto \) (likelihood) \( \times \) (prior),

\[
f(\lambda|x's) \propto (\lambda^n e^{-\lambda n \bar{x}})(\lambda^{\alpha-1} e^{-\lambda \beta}) = \lambda^{(\alpha+n)-1} e^{-\lambda (\beta+n \bar{x})}
\]

Prior influence on Posterior

- Here are plots for \( n = 3, \bar{x} = 7 \):
  - \( \alpha=1, \beta=2 \)
  - \( \alpha=4, \beta=8 \)
  - \( \alpha=20, \beta=20 \)
- We see the “shrinkage” idea again: the posterior is “between” the likelihood and the prior
- We see that the location and spread of the prior influences the location and spread of the posterior distribution
Bayesian point estimate and interval for Prof. Smedley

- We will take $\alpha=4$, $\beta=8$, as an example
- There are formulae, but we will simulate

```r
> n <- 3
> xbar <- 7
> alpha <- 4
> beta <- 8
> nsim <- 1000
> simdata <- rgamma(nsim, alpha+n, beta+n*xbar)
> quantile(simdata, c(0.025, 0.5, 0.975))
  2.5%        50%      9.75%
0.0975616  0.2308562  0.4435332
```

So a point estimate would be 0.23, and a 95% interval would be (0.098, 0.444)

---

Comparing the MLE and Bayes intervals

- MLE interval: (-0.022, 0.308)
  - Left endpoint can be unrealistic
  - "In 95% of experiments, the procedure we used would produce a CI that contains the true value of $\lambda$"
  - CI = “confidence interval”
- Bayesian interval: (0.098, 0.444)
  - Endpoints always in the parameter space
  - $P[0.098 < \lambda < 0.444 | \text{data}] = 0.95$
  - CI = “credible interval”
Choosing priors... Conjugate priors

- For the binomial
  - Beta prior: \( f(p|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \)
  - Binomial likelihood: \( L(p) \propto p^k (1-p)^{n-k} \)
  - Beta posterior: \( f(p|\text{data}) = Beta(p|\alpha+k, \beta+n-k) \)

- For the exponential
  - Gamma prior: \( f(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta} \)
  - Exponential likelihood: \( L(\lambda) = \lambda^n e^{-\lambda n\bar{x}} \)
  - Gamma posterior: \( f(\lambda|\text{data}) = Gamma(\lambda|\alpha+n, \beta+n\bar{x}) \)

- Conjugate prior: iff posterior is in same “family”

Non-conjugate priors & JAGS

- Conjugate priors make life easy
  - Even with no formulae, using \( \text{rbeta()} \) or \( \text{rgamma()} \) to simulate from the posterior was easy!

- For exponential model, a non-conjugate choice:
  - Prior: log-normal: \( f(\lambda|\mu, \sigma) = \frac{1}{\lambda\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(\log \lambda - \mu)^2}{2\sigma^2} \right\} \)
  - Likelihood: exponential: \( L(\lambda) = \lambda^n e^{-\lambda n\bar{x}} \)
  - Posterior: \( f(\lambda|\text{data}) \propto \lambda^n e^{-\lambda n\bar{x}} \frac{1}{\lambda\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(\log \lambda - \mu)^2}{2\sigma^2} \right\} \)

- JAGS, WinBUGS are programs for simulating from the posterior, no matter what prior!
Normal Model: Estimate $\mu$, with $\sigma^2$ known, One Observation $x \sim N(\mu, \sigma^2)$

- Easy to “see” conjugate prior

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$f(\mu) = \frac{1}{\sqrt{2\pi}\tau_0} e^{-\frac{1}{2\tau_0^2}(\mu-\mu_0)^2}$$

$$f(\mu|x) \propto f(x|\mu)f(\mu) \propto \exp \left\{ -\frac{1}{2} \left[ \frac{(x-\mu)^2}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\tau_0^2} \right] \right\}$$

- Posterior must be normal for $\mu$ (quadratic in $\mu$!); to identify it, complete the square...

The exponent of $p(\mu|x)$ looks like -1/2 times

$$\frac{(x-\mu)^2}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\tau_0^2} = \frac{\tau_0^2 + \sigma^2}{\tau_0^2\sigma^2} \left[ \frac{\mu^2 - 2x\mu\tau_0^2 + 2\mu\mu_0\sigma^2}{\tau_0^2 + \sigma^2} + \frac{x^2\tau_0^2 + \mu_0^2\sigma^2}{\tau_0^2 + \sigma^2} \right]$$

$$- \frac{\tau_0^2 + \sigma^2}{\tau_0^2\sigma^2} \left[ \left( \mu - \frac{x\tau_0^2 + \mu_0\sigma^2}{\tau_0^2 + \sigma^2} \right)^2 + \text{junk}(x, \sigma^2, \mu_0, \tau_0^2) \right]$$

$$= \frac{1}{\tau_1^2} (\mu - \mu_1)^2 + \text{(known junk)}$$

so that $\mu|x \sim N(\mu_1, \tau_1^{-2})$, where

$$\tau_1^2 = \frac{\tau_0^2\sigma^2}{\tau_0^2 + \sigma^2} = \frac{1}{1/\sigma^2 + 1/\tau_0^2}$$

$$\mu_1 = \frac{x\tau_0^2 + \mu_0\sigma^2}{\tau_0^2 + \sigma^2} = \left( \frac{\tau_0^2}{\tau_0^2 + \sigma^2} \right) x + \left( \frac{\sigma^2}{\tau_0^2 + \sigma^2} \right) \mu_0$$
The exponent of \( p(\mu | x) \) looks like \(-1/2\) times

\[
\frac{(x - \mu)^2}{\sigma^2} + \frac{(\mu - \mu_0)^2}{\tau_0^2} = \frac{\tau_0^2 + \sigma^2}{\tau_0^2 \sigma^2} \left[ \mu^2 - \frac{2x\mu\tau_0^2 + 2\mu\mu_0\sigma^2}{\tau_0^2 + \sigma^2} + \frac{x^2\tau_0^2 + \mu_0^2\sigma^2}{\tau_0^2 + \sigma^2} \right] - \frac{\tau_0^2 + \sigma^2}{\tau_0^2 \sigma^2} \left[ \left( \mu - \frac{x\tau_0^2 + \mu_0\sigma^2}{\tau_0^2 + \sigma^2} \right)^2 + \text{junk}(x, \sigma^2, \mu_0, \tau_0^2) \right] = \frac{1}{\tau_1^2} (\mu - \mu_1)^2 + \text{(known junk)}
\]

so that \( \mu | x \sim N(\mu_1, \tau_1^{-2}) \), where

\[
\tau_1^2 = \frac{\tau_0^2 \sigma^2}{\tau_0^2 + \sigma^2} = \frac{1}{1/\sigma^2 + 1/\tau_0^2}
\]

\[
\mu_1 = \frac{x\tau_0^2 + \mu_0\sigma^2}{\tau_0^2 + \sigma^2} = \left( \frac{\tau_0^2}{\tau_0^2 + \sigma^2} \right) x + \left( \frac{\sigma^2}{\tau_0^2 + \sigma^2} \right) \mu_0
\]

---

**n Observations** \( x_i \sim N(\mu, \sigma^2) \)

- Since

\[
p(x_1, \ldots, x_n | \mu) = N(x_1, \ldots, x_n | \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2} (y_i - \mu)^2}
\]

\[
\propto N(\bar{x} | \mu, \sigma^2/n) \equiv \frac{1}{\sqrt{2\pi\sigma^2/n}} e^{-\frac{1}{2\sigma^2/n} (y - \mu)^2}
\]

we can apply the results for one observation

- \( p(x_1, \ldots, x_n | \mu) \propto N(\bar{x} | \mu, \sigma^2/n), \sigma_n^2 = \sigma^2/n \)
- \( p(\mu) = N(\mu | \mu_0, \tau_0^2) \)
- \( p(\mu | \text{data}) = N(\mu | \mu_n, \tau_n^2) \) where

\[
\tau_n^2 = \frac{1}{1/\sigma_n^2 + 1/\tau_0^2} = \frac{1}{n/\sigma^2 + 1/\tau_0^2}
\]

\[
\mu_n = \frac{\bar{x}/\sigma_n^2 + \mu_0/\tau_0^2}{1/\sigma_n^2 + 1/\tau_0^2} = \left( \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n} \right) \bar{x} + \left( \frac{\sigma^2/n}{\tau_0^2 + \sigma^2/n} \right) \mu_0
\]
Normal Mean, Example

- Suppose we know \( \sigma = 12 \), we look at \( n = 169 \) IQ scores, and we find \( \bar{x} = 100 \).
- We use as prior \( N(\mu_0, \tau_0^2) \) with \( \mu_0 = 90, \tau_0^2 = 4 \).
- Shrinkage determined by
  \[
  \mu_n = \left( \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n} \right) \bar{x} + \left( \frac{\sigma^2/n}{\tau_0^2 + \sigma^2/n} \right) \mu_0
  \]
- \( \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n} \) is the **reliability**
- \( n \) larger \( \Rightarrow \) reliability larger \( \Rightarrow \) less shrinkage

---

Minnesota Radon Example

- Emphasize Distribution Structure
  
  Level 2: \( \mu_j \overset{iid}{\sim} N(\mu_0, \tau_0^2) \)
  
  Level 1: \( y_{j[i]} \overset{indep}{\sim} N(\mu_j, \sigma^2) \)

- Emphasize Bayesian point of view (more later!)
  
  Prior: \( \mu_j \overset{iid}{\sim} N(\mu_0, \tau_0^2) \)

  Likelihood: \( y_{j[i]} \overset{indep}{\sim} N(\mu_j, \sigma^2) \)

- Emphasize two-stage (multistage) sampling
  
  Mean radon across MN

  County-level differences from grand mean

  Individual house levels
In each county $i$ with $n_i$ houses, the posterior mean radon level will be

$$
\mu_{i,\text{post}} = \left( \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n_i} \right) \bar{y}_i + \left( \frac{\sigma^2/n_i}{\tau_0^2 + \sigma^2/n_i} \right) \mu_0
$$

- When $n_i$ large, $\mu_{i,\text{post}} \approx \bar{y}_i$
- When $n_i$ small, $\mu_{i,\text{post}} \approx \mu_0$

In the figure, the grand mean is $\mu_0$.

In each county $i$ with $n_i$ houses, posterior mean is

$$
\mu_{i,\text{post}} = \left( \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n_i} \right) \bar{y}_i + \left( \frac{\sigma^2/n_i}{\tau_0^2 + \sigma^2/n_i} \right) \mu_0
$$

- When $n_i$ large, $\mu_{i,\text{post}} \approx \bar{y}_i$
- When $n_i$ small, $\mu_{i,\text{post}} \approx \mu_0$
Summary

- Exponential Likelihood, Gamma Prior, & Prof Smedley
- Prior Influence on Posterior
- Comparing 95% intervals
- Conjugate Priors
- Non-Conjugate Priors & WinBUGS
- Normal Model: Estimate Mean, Variance Known
- Shrinkage and the Minnesota Radon Example
- Start reading Lynch Ch 4 – on course website!