Assignment 2

Yufei Yin

1. (2 points)

Suppose that we have observed $y = (y_1, ..., y_n)$, with the y_i independent and $y_i|\lambda \sim Poisson(\lambda)$. If we specify a prior distribution $\lambda \sim Gamma(\alpha, \beta)$, show that $\lambda|y \sim Gamma(\alpha + \sum_{i=1}^n y_i, \beta + n)$.

prior distribution

Gamma(α, β)

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

likelihood function

$P(y|\lambda) = \prod_{i=1}^n \frac{\lambda^{y_i} e^{-\lambda}}{y!} \propto e^{-n\lambda} \lambda^{\sum_{i=1}^n y_i}$

posterior distribution

$P(\lambda|y) \propto P(y|\lambda) \times P(\lambda) = (e^{-n\lambda} \lambda^{\sum_{i=1}^n y_i}) \times \lambda^{\alpha-1} e^{-\beta\lambda} = \lambda^{\alpha + \sum_{i=1}^n y_i - 1} e^{-(\beta+n)\lambda}$

$\lambda^{\alpha + \sum_{i=1}^n y_i - 1} e^{-(\beta+n)\lambda}$ is the "kernel" of Gamma($\alpha + \sum_{i=1}^n y_i$, $\beta + n$).

2. (2 points)

When David Beckham joined the LA Galaxy soccer team, he scored one goal in his first two games. Suppose that the manager of LA Galaxy is interested in the the number of goals Beckham would score in each of the remaining games. We model the number of goals, y_i, Beckham scores in a game using a Poisson distribution with parameter λ:

$$y_i | \lambda \sim Poisson(\lambda).$$

For a prior distribution on λ, we use the fact that while playing in Madrid a year before joining LA Galaxy, Beckham scored 3 goals in 22 games (or, $3/22 = 0.14$ goals per game on average). We therefore choose a Gamma prior with mean around 0.14 and variance large enough to reflect our uncertainty. Specifically, we choose $\lambda \sim$ Gamma(1.4, 10).

(a) Provide a point estimate for the expected number of goals Beckham will score in a game using the posterior mean, $E(\lambda|y)$. (Note that this does not have to be an integer.)

Because David Beckham scored one goal in his first two games, $y_1 + y_2 = 1$ and n = 2.

Using the result from Question 1

Posterior distribution

Gamma($\alpha + \sum_{i=1}^{n} y_i$, $\beta + n$) = Gamma(1.4+1, 10+2) = Gamma(2.4, 12)

Posterior mean

$E(\lambda|y) = \frac{2.4}{12} = 0.2$

So the expected number of goals Beckham will score in a game using the posterior mean, $E(\lambda|y)$ is 0.2.

(b) Give a 95% equal-tailed posterior interval for λ.

Posterior distributioin is Gamma(2.4, 12).

```
lower <- qgamma(p = 0.025, shape = 2.4, rate = 12)
upper <- qgamma(p = 0.975, shape = 2.4, rate = 12)
c(lower, upper)
```

```
## [1] 0.03152509 0.52088466
```

A 95% equal-tailed posterior interval for λ is (0.03152509, 0.52088466).

3. (2 points)

Chapter 2, Exercise 8 from Bayesian Data Analysis, 3rd Edition (Gelman et al.), which is linked to from the course Canvas page.

8. Normal distribution with unknown mean: a random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is $\bar{y} = 150$ pounds. Assume the weights in the population are normally distributed with unknown mean θ and known standard deviation 20 pounds. Suppose your prior distribution for θ is normal with mean 180 and standard deviation 40.

   (a) Give your posterior distribution for θ. (Your answer will be a function of n.)

$\bar{y} = 150$, $\sigma^2 = 20^2$, $\mu_0 = 180$, $\tau_0^2 = 40^2$

Prior Distribution

$N(\mu_0, \tau_0^2) = N(180, 40^2)$

Posterior distribution

For n observations

$N(\mu_n, \tau_n^2)$ distribution with

$$\mu_n = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad ; \quad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$$

$\mu_n = \frac{\frac{180}{40^2} + \frac{150n}{20^2}}{\frac{1}{40^2} + \frac{n}{20^2}}$

$\tau_n^2 = \frac{1}{\frac{1}{40^2} + \frac{n}{20^2}}$

So Posterior distribution is $N\left(\frac{\frac{180}{40^2} + \frac{150n}{20^2}}{\frac{1}{40^2} + \frac{n}{20^2}}, \frac{1}{\frac{1}{40^2} + \frac{n}{20^2}}\right)$.

(b) A new student is sampled at random from the same population and has a weight of $\tilde{y}$ pounds. Give a posterior predictive distribution for $\tilde{y}$. (Your answer will still be a function of n.)

Posterior predictive distribution

$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta \propto \int exp(-\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2)exp(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2)d\theta$

$exp(-\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2)exp(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2)$ is the exponential of a quadratic function of $(\tilde{y}, \theta)$, So p($\tilde{y}|y$) is also a Normal distribution.

$E(\tilde{y}|y) = E(E(\tilde{y}|\theta, y)|y) = E(\theta|y) = \mu_n$

$Var(\tilde{y}|y) = E(Var(\tilde{y}|\theta, y)|y) + Var(E(\tilde{y}|\theta, y)|y) = E(\sigma^2|y) + Var(\theta|y) = \sigma^2 + \tau_n^2$

$\tilde{y}|y \sim N(\mu_n, \sigma^2 + \tau_n^2)$

So Posterior predictive distribution is $N(\frac{\frac{180}{40^2}+\frac{150n}{20^2}}{\frac{1}{40^2}+\frac{n}{20^2}}, 20^2 + \frac{1}{\frac{1}{40^2}+\frac{n}{20^2}})$

(c) For n = 10, give a 95% posterior interval for θ and a 95% posterior predictive interval for $\tilde{y}$.

```
# Initialize
mu_0 <- 180
tau_0 <- 40
sig <- 20
y.bar <- 150
n <- 10
```

```
# Calculation
mu_n <- (mu_0/tau_0^2+n*y.bar/sig^2)/(1/tau_0^2+n/sig^2)
tau_n <- sqrt(1/(1/tau_0^2+n/sig^2))
```

```
# 95% posterior interval
lower <- qnorm(p = 0.025, mean = mu_n, sd = tau_n)
upper <- qnorm(p = 0.975, mean = mu_n, sd = tau_n)
c(lower, upper)
```

```
## [1] 138.4879 162.9755
```

The 95% posterior interval for θ is (138.4879, 162.9755).

```
# 95% posterior predictive interval
lower <- qnorm(p = 0.025, mean = mu_n, sd = sqrt(tau_n^2+sig^2))
upper <- qnorm(p = 0.975, mean = mu_n, sd = sqrt(tau_n^2+sig^2))
c(lower, upper)
```

```
## [1] 109.6648 191.7987
```

The 95% posterior predictive interval for $\tilde{y}$ is (109.6648, 191.7987).

(d) Do the same for n = 100.

```
n <- 100
```

```
# update
mu_n <- (mu_0/tau_0^2+n*y.bar/sig^2)/(1/tau_0^2+n/sig^2)
tau_n <- sqrt(1/(1/tau_0^2+n/sig^2))
```

```
# 95% posterior interval
lower <- qnorm(p = 0.025, mean = mu_n, sd = tau_n)
upper <- qnorm(p = 0.975, mean = mu_n, sd = tau_n)
c(lower, upper)
```

```
## [1] 146.1598 153.9899
```

The 95% posterior interval for θ is (146.1598, 153.9899).

```
# 95% posterior predictive interval
lower <- qnorm(p = 0.025, mean = mu_n, sd = sqrt(tau_n^2+sig^2))
upper <- qnorm(p = 0.975, mean = mu_n, sd = sqrt(tau_n^2+sig^2))
c(lower, upper)
```

```
## [1] 110.6805 189.4691
```

The 95% posterior predictive interval for $\tilde{y}$ is (110.6805, 189.4691).

4. (2 points)

Chapter 2, Exercise 12 from Bayesian Data Analysis, 3rd Edition.

12. Jeffreys' prior distributions: suppose $y|\theta \sim \text{Poisson}(\theta)$. Find Jeffreys' prior density for θ, and then find α and β for which the $\text{Gamma}(\alpha, \beta)$ density is a close match to Jeffreys' density.

Jeffrey's prior:

$$P(\theta) \propto \sqrt{J(\theta)}$$

where $J(\theta)$ is the fisher information:

$$J(\theta) = -E\left(\frac{d^2 log P(y|\theta)}{d\theta^2}|\theta\right)$$

$P(y|\theta) = \frac{\theta^y e^{-\theta}}{y!}$

$log P(y|\theta) = y log\theta - \theta - log(y!)$

$\frac{d log P(y|\theta)}{d\theta} = \frac{y}{\theta} - 1 - 0$

$\frac{d^2 log P(y|\theta)}{d\theta^2} = -\frac{y}{\theta^2}$

$J(\theta) = -E\left(\frac{d^2 log P(y|\theta)}{d\theta^2}|\theta\right) = -E(-\frac{y}{\theta^2}) = \frac{1}{\theta^2}E(y) = \frac{1}{\theta}$

So the Jeffrey's prior density for θ is $P(\theta) \propto \sqrt{J(\theta)} = \sqrt{\frac{1}{\theta}} = \theta^{-\frac{1}{2}}$

$\theta^{-\frac{1}{2}} = \theta^{(\frac{1}{2}-1)}e^0$ is the "kernel" of $\text{Gamma}(\frac{1}{2}, 0)$, so $\alpha = \frac{1}{2}$, $\beta = 0$.

5. (2 points)

Chapter 2, Exercise 17 from Bayesian Data Analysis, 3rd Edition.

17. Posterior intervals: unlike the central posterior interval, the highest posterior interval is *not* invariant to transformation. For example, suppose that, given σ^2, the quantity nv/σ^2 is distributed as χ_n^2, and that σ has the (improper) noninformative prior density $p(\sigma) \propto \sigma^{-1}$, $\sigma > 0$.

(a) Prove that the corresponding prior density for σ^2 is $p(\sigma^2) \propto \sigma^{-2}$.

Consider a one-to-one transformation of the parameter $\phi = h(\theta)$. Then, by transformation of variables, the prior density $p(\theta)$ is equivalent to the following prior density on ϕ:

$$p(\phi) = p(\theta) \left| \frac{d\theta}{d\phi} \right| = p(\theta)|h'(\theta)|^{-1}$$

$p(\sigma^2) = p(\sqrt{\sigma^2})/(2\sqrt{\sigma^2}) = p(\sigma)\frac{1}{2}\sigma^{-1}$

Because $p(\sigma) \propto \sigma^{-1}$, $p(\sigma^2) = p(\sigma)\frac{1}{2}\sigma^{-1} \propto \sigma^{-2}$.

(b) Show that the 95% highest posterior density region for σ^2 is not the same as the region obtained by squaring the endpoints of a posterior interval for σ.

pivotal quantity

$$T = \frac{nv}{\sigma^2}\Big|\sigma^2 \sim \chi_n^2$$

Which is a special case of $\text{Gamma}(\frac{n}{2}, \frac{1}{2})$.

Change of variable method

$y = g(T) = T\sigma^2$, $g^{-1}(y) = \frac{y}{\sigma^2}$, $\frac{dg^{-1}(y)}{dy} = \frac{1}{\sigma^2}$

likelihood of σ^2

$p(y|\sigma^2) = p(g^{-1}(y)|\sigma^2)\left|\frac{dg^{-1}(y)}{dy}\right| = \text{Gamma}(T|\frac{n}{2}, \frac{1}{2}) \cdot \frac{1}{\sigma^2}$

$= \frac{\frac{1}{2}^{\frac{n}{2}} \cdot (\frac{nv}{\sigma^2})^{\frac{n}{2}-1} \cdot e^{-\frac{nv}{2\sigma^2}}}{\Gamma(\frac{n}{2})} \cdot \frac{1}{\sigma^2}$

$\propto (\sigma^{-2})^{\frac{n}{2}-1} \cdot e^{-\frac{nv}{2\sigma^2}} \cdot \sigma^{-2} = \sigma^{-n} \cdot e^{-\frac{nv}{2\sigma^2}}$

$p(y|\sigma^2) \propto \sigma^{-n} \cdot e^{-\frac{nv}{2\sigma^2}}$

$p(y|\sigma) \propto \sigma^{-n} \cdot e^{-\frac{nv}{2\sigma^2}}$

prior distribution

$p(\sigma) \propto \sigma^{-1}$

$p(\sigma^2) \propto \sigma^{-2}$

posterior distribution

$p(\sigma^2|y) \propto p(y|\sigma^2)p(\sigma^2) = \sigma^{-n} \cdot e^{-\frac{nv}{2\sigma^2}} \cdot \sigma^{-2} = (\sigma^2)^{-\frac{n}{2}-1} \cdot e^{-\frac{nv}{2\sigma^2}}$

$p(\sigma|y) \propto p(y|\sigma)p(\sigma) = \sigma^{-n} \cdot e^{-\frac{nv}{2\sigma^2}} \cdot \sigma^{-1} = (\sigma)^{-n-1} \cdot e^{-\frac{nv}{2\sigma^2}}$

Proof by contradiction

posterior distribution is univariate, HPD interval is unique.

Let $[\sqrt{a}, \sqrt{b}]$ be 95% HPD interval for σ:

$(\sqrt{a})^{-n-1} \cdot e^{-\frac{nv}{2(\sqrt{a})^2}} = (\sqrt{b})^{-n-1} \cdot e^{-\frac{nv}{2(\sqrt{b})^2}}$

After simplification, we get:

$$a^{(-\frac{n}{2}-\frac{1}{2})} \cdot e^{-\frac{nv}{2a}} = b^{(-\frac{n}{2}-\frac{1}{2})} \cdot e^{-\frac{nv}{2b}} \tag{1}$$

If [a, b] is the 95% HPD interval for σ^2:

$$a^{\left(-\frac{n}{2}-1\right)} \cdot e^{-\frac{nv}{2a}} = b^{\left(-\frac{n}{2}-1\right)} \cdot e^{-\frac{nv}{2b}} \tag{2}$$

Using $\frac{(1)}{(2)}$, we can get:

$$a^{\frac{1}{2}} = b^{\frac{1}{2}} \Rightarrow a = b$$

So [a, b] is not 95% HPD interval for σ^2.