Midterm Practice Solutions

Problem 1

1.

- population: all products produced by the machine.
- sample: selected products produced by the machine.
- estimand: the expected time needed for the machine to produce a product.

2.

```
for (i in 1:n) {
  draw y_i from Expo(\lambda)
}
```

3.

The likelihood function:

$$L(\lambda; y_1, \cdots, y_n) = \lambda^n e^{-\lambda \sum_{i=1}^n y_i}$$

Taking log of the likelihood function to get the log-likelihood:

$$l(\lambda; y_1, \cdots, y_n) = n \log \lambda - \lambda \sum_{i=1}^n y_i$$

4.

Taking derivative of log-likelihood function:

$$l'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n y_i$$

Setting $l'(\lambda) = 0$ and solve it for λ:

$$\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^n y_i}$$

5.

Gamma. This can be demonstrated by showing that both prior and posterior can be written in the form $\lambda^{\alpha-1} e^{-\beta\theta}$ for appropriate values of α and β

6.

$$
\begin{aligned}
p(\lambda \mid y_1, \cdots, y_n) \propto\ & \lambda^n e^{-\lambda \sum_{i=1}^n y_i} \lambda^{\alpha-1} e^{-\beta\lambda} \\
& \propto \lambda^{\alpha+n-1} e^{-(\sum_{i=1}^n y_i + \beta)\lambda}
\end{aligned}
$$

Therefore,

$$
\lambda \mid y_1, \cdots, y_n \sim Gamma(\alpha + n,\ \beta + \sum_{i=1}^n y_i)
$$

7.

- posterior mean: $\frac{\alpha+n}{\beta+\sum_{i=1}^n y_i}$
- posterior mode: $\frac{\alpha+n-1}{\beta+\sum_{i=1}^n y_i}$

8.

```r
y <- c(2, 30, 40, 10, 24, 14, 6, 8, 20, 35)
```

For the prior, choose $\alpha = 100$ and $\beta = 5$. (Any other reasonable choice will also work.)

```r
a <- 100
b <- 5
a_post <- a + length(y)
b_post <- b + sum(y)
# posterior mean #
a_post / b_post
```

```
## [1] 0.5670103
```

```r
# posterior mode #
(a_post - 1) / b_post
```

```
## [1] 0.5618557
```

9.

- 95% Central interval: find the 0.025% quantile of $Gamma(\alpha + n,\ \beta + \sum_{i=1}^n y_i)$ as the lower bound, and the 0.975% quantile of $Gamma(\alpha + n,\ \beta + \sum_{i=1}^n y_i)$ as the upper bound. Can use the `qgamma` function to do this.
- 95% HPD interval: find the interval $C = \{\theta : p(\theta \mid y) \geq k\}$, where k is the largest number such that $\int_{\theta : p(\theta|y) \geq k} p(\theta \mid y) d\theta = 1 - \alpha$ and $p(\theta \mid y)$ is the density of $Gamma(\alpha + n,\ \beta + \sum_{i=1}^n y_i)$.

Both are regions for which the posterior probability that θ is in the region is equal to 0.95.

10.

- The HPD interval is shorter.
- There is 95% of the chance that λ is within the 95% credible interval.
- For the credible interval, the parameter is random while the endpoints are fixed given the parameter; For the frequentist confidence interval, the parameter is fixed while the endpoints are random.

11.

$$
\begin{aligned}
p(\tilde{y} \mid y) &= \int p(\tilde{y} \mid \lambda) p(\lambda \mid y) d\lambda \\
&= \int \lambda e^{-\lambda \tilde{y}} \frac{(\sum_i y_i + \beta)^{n+\alpha}}{\Gamma(n+\alpha)} \lambda^{n+\alpha-1} e^{-(\sum_i y_i + \beta)\lambda} d\lambda \\
&= \frac{(\sum_i y_i + \beta)^{n+\alpha}}{\Gamma(n+\alpha)} \int \lambda^{n+\alpha+1-1} e^{(\sum_i y_i + \tilde{y} + \beta)\lambda} d\lambda \\
&= \frac{(\sum_i y_i + \beta)^{n+\alpha}}{\Gamma(n+\alpha)} \frac{\Gamma(n+\alpha+1)}{(\sum_i y_i + \tilde{y} + \beta)^{n+\alpha+1}} \\
&= \frac{(n+\alpha)(\sum_i y_i + \beta)^{n+\alpha}}{(\sum_i y_i + \tilde{y} + \beta)^{n+\alpha+1}}, \quad \tilde{y} > 0
\end{aligned}
$$

An alternative to doing calculus is to draw samples from the posterior predictive distribution and approximate the distribution (or an posterior summaries) using Monte Carlo. Pseudocode:

```
for(s in 1:nsamples) {
  1. Sample lambda_s from a Gamma(n + alpha, sum(y) + beta)
  2  Sample ypred_s from an Expo(lambda_s)
}
return ypreds
```

Problem 2

1. Poisson-Gamma Practice

(a)

$$
L(\lambda | y_1, y_2, \ldots, y_n, \nu_1, \nu_2, \ldots, \nu_n) = \prod_{i=1}^{n} (\nu_i \lambda)^{y_i} e^{-\nu_i \lambda} / y_i! \propto \lambda^{\sum y_i} e^{-\lambda \sum \nu_i}
$$

(b)

$$
p(\lambda | y) \propto L(\lambda) * p(\lambda) \propto \lambda^{\sum y_i} e^{-\lambda \sum \nu_i} \times \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \propto \lambda^{\sum y_i + a - 1} e^{-\lambda(\sum \nu_i + b)}
$$

So the posterior is a $Gamma(\sum y_i + a, \sum \nu_i + b)$

(c)

b can be interpreted as the number of prior observations and a can be interpreted as the sum of the counts from prior total exposure of b.

(d)

$$
\begin{aligned}
E[\lambda \mid y_1, ...y_n] &= \frac{a + \sum y_i}{b + \sum \nu_i} \\
&= \frac{b}{b + \sum \nu_i} \frac{a}{b} + \frac{\sum \nu_i}{b + \sum \nu_i} \frac{\sum y_i}{\sum \nu_i} \\
&= (1 - w)\frac{a}{b} + w\hat{\lambda}_{\text{MLE}}
\end{aligned}
$$

2. Beta-Bernoulli practice

(a)

$$L(\theta) \propto \theta^{\sum y_i}(1-\theta)^{n-\sum y_i}$$

(b)

$$p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$$
$$p(\theta \mid y_1, y_2, \ldots, y_n) = \theta^{\sum y_i + a - 1}(1-\theta)^{n - \sum y_i + b - 1}$$

So the posterior is a $Beta(\sum y_i + a, n - \sum y_i + b)$

(c)

We can interpret a as prior total number of success and b as prior total number of failures.

(d)

$$E[\theta \mid y_1, y_2, \ldots, y_n] = \frac{\sum y_i + a}{n + a + b} = \frac{\sum y_i}{n} \times \frac{n}{n + a + b} + \frac{a}{a + b} \times \frac{a + b}{n + a + b}$$

Multiple Choice Practice

1. a b c
2. a c e
3. c
4. Beta Gamma
5. a
6. c
7. b c
8. a
9. b