Lesson 9
Graded Quiz • 12 min

Due Apr 6, 2:59 AM EDT

Congratulations! You passed!

TO PASS 75% or higher

Keep Learning

GRADE
100%

Your skill score is improving!

You've made progress in Statistics. View Skill Tracking.

Was this useful? Yes No

Statistics NEW SCORE

55

Intermediate

0 30 70

Lesson 9

LATEST SUBMISSION GRADE

100%

1. **For Questions 1-3, refer to the bus waiting time example from the lesson.** 1 / 1 point

Recall that we used the conjugate gamma prior for λ, the arrival rate in busses per minute. Suppose our prior belief about this rate is that it should have mean 1/20 arrivals per minute with standard deviation 1/5. Then the prior is $\mathrm{Gamma}(a, b)$ with $a = 1/16$.

- Find the value of b. Round your answer to two decimal places.

1.25

Correct
This is $b = 5/4$ which results in prior mean $a/b = 1/(16b) = 1/20$ and prior standard deviation $\sqrt{a}/b = 1/(\sqrt{16}b) = 1/5$.

Note that a prior expected arrival rate of 1/20 busses per minute is *not* equivalent to a prior expected wait time of 20 minutes per bus. Indeed, for random variables generally, $E(1/X) \neq 1/E(X)$.

2. Bus waiting times: 1 / 1 point

Suppose that we wish to use a prior with the same mean (1/20), but with effective sample size of one arrival. Then the prior for λ is $\mathrm{Gamma}(1, 20)$.

In addition to the original $Y_1 = 12$, we observe the waiting times for four additional busses: $Y_2 = 15$, $Y_3 = 8$, $Y_4 = 13.5$, $Y_5 = 25$.

Recall that with multiple (independent) observations, the posterior for λ is $\mathrm{Gamma}(\alpha, \beta)$ where $\alpha = a + n$ and $\beta = b + \sum y_i$.

- What is the posterior mean for λ? Round your answer to two decimal places.

0.06

Correct
This is the mean of the posterior distribution: $\mathrm{Gamma}(\alpha, \beta)$ with $\alpha = a + n = 1 + 5$ and $\beta = b + \sum y_i = 20 + 73.5$.

3. Bus waiting times: 1 / 1 point

- Continuing Question 2, use R or Excel to find the posterior probability that $\lambda < 1/10$? Round your answer to two decimal places.

answer to two decimal places.

0.90

Correct

There is a fairly high posterior probability that the arrival rate is less than 1/10 busses per minute, or equivalently that the average waiting time for a bus is greater than 10 minutes.

In R:

```
pgamma(q=1/10, shape=6, rate=93.5)
```

In Excel:

```
= GAMMA.DIST(1/10, 6, 1/93.5, TRUE)
```

where x=1/10, alpha=6, beta=1/93.5 (in Excel, beta is a shape parameter), cumulative=TRUE.

4. **For Questions 4-10, consider the following earthquake data:**

1 / 1 point

The United States Geological Survey maintains a list of significant earthquakes worldwide. We will model the rate of earthquakes of magnitude 4.0+ in the state of California during 2015. An iid exponential model on the waiting time between significant earthquakes is appropriate if we assume:

1. earthquake events are independent,
2. the rate at which earthquakes occur does not change during the year, and
3. the earthquake hazard rate does not change (i.e., the probability of an earthquake happening tomorrow is constant regardless of whether the previous earthquake was yesterday or 100 days ago).

Let Y_i denote the waiting time in days between the ith earthquake and the following earthquake. Our model is $Y_i \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$ where the expected waiting time between earthquakes is $E(Y) = 1/\lambda$ days.

Assume the conjugate prior $\lambda \sim \text{Gamma}(a, b)$. Suppose our prior expectation for λ is 1/30, and we wish to use a prior effective sample size of one interval between earthquakes.

- What is the value of a?

1

Correct

In the exponential-gamma model, a is the prior effective sample size.

5. Earthquake data:

1 / 1 point

- What is the value of b?

30

Correct

The prior mean is $a/b = 1/30$, and since we know the effective sample size $a = 1$, we have $b = 30$.

6. Earthquake data:

1 / 1 point

The significant earthquakes of magnitude 4.0+ in the state of California during 2015 occurred on the following dates (http://earthquake.usgs.gov/earthquakes/browse/significant.php?year=2015):

January 4, January 20, January 28, May 22, July 21, July 25, August 17, September 16, December 30.

- Recall that we are modeling the waiting times between earthquakes in days. Which of the following is our data vector?

- [ ] **y** = (3, 16, 8, 114, 60, 4, 23, 30, 105)
- [ ] **y** = (0, 0, 4, 2, 0, 1, 1, 3)

- [x] **y** = (16, 8, 114, 60, 4, 23, 30, 105)
- [ ] **y** = (3, 16, 8, 114, 60, 4, 23, 30, 105, 1)

Correct

There are eight intervals between the first and last event.

We are excluding four days of the year in which no events were observed. A more comprehensive model (e.g., censoring methods) would account for the fact that there were no major earthquakes Jan. 1 to Jan. 4 and Dec. 30 to Dec. 31. This is beyond the scope of the course.

7. Earthquake data:

1 / 1 point

- The posterior distribution is $\lambda \mid \mathbf{y} \sim \mathrm{Gamma}(\alpha, \beta)$. What is the value of α?

9

Correct

This is $\alpha = a + n = 1 + 8$.

8. Earthquake data:

1 / 1 point

- The posterior distribution is $\lambda \mid \mathbf{y} \sim \mathrm{Gamma}(\alpha, \beta)$. What is the value of β?

390

Correct

This is $\beta = b + \sum y_i = 30 + 360$.

9. Earthquake data:

1 / 1 point

- Use R or Excel to calculate the upper end of the 95% equal-tailed credible interval for λ, the rate of major earthquakes in events per day. Round your answer to two decimal places.

0.04

Correct

The full interval is (0.011, 0.040). Thus our posterior probability that $0.011 < \lambda < 0.040$ is 0.95.

The interval in terms of $1/\lambda$, the expected number of days between events is (24.7, 94.8). Note that although $E(1/\lambda) \neq 1/E(\lambda)$, we can take the reciprocals of quantiles since $P(X < q) = P(1/q < 1/X)$. Just remember that the lower end of the interval becomes the upper end and vise versa.

In R:

```
qgamma(p=0.975, shape=9, rate=390)
```

In Excel:

```
= GAMMA.INV(0.975, 9, 1/390)
```

where probability=0.975, alpha=9, and beta=1/390 (in Excel, beta is the shape parameter).

10. Earthquake data:

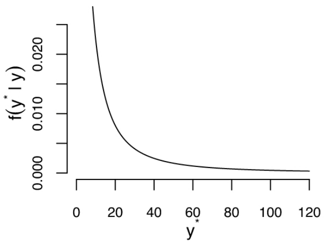
1 / 1 point

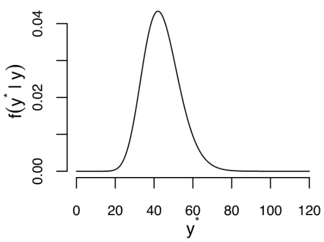
The posterior predictive density for a new waiting time y^* in days is:

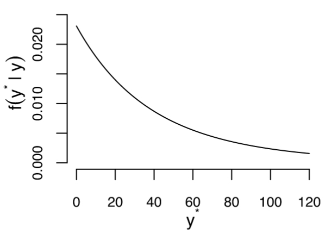
$$f(y^* \mid \mathbf{y}) = \int f(y^* \mid \lambda) \cdot f(\lambda \mid \mathbf{y}) d\lambda = \frac{\beta^\alpha \Gamma(\alpha+1)}{(\beta+y^*)^{\alpha+1}\Gamma(\alpha)} I_{\{y^* \geq 0\}} = \frac{\beta^\alpha \alpha}{(\beta+y^*)^{\alpha+1}} I_{\{y^* \geq 0\}}$$

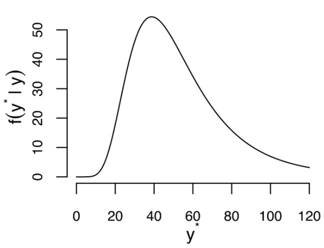
where $f(\lambda \mid \mathbf{y})$ is the $\mathrm{Gamma}(\alpha, \beta)$ posterior found earlier. Use R or Excel to evaluate this posterior predictive PDF.

- Which of the following graphs shows the posterior predictive distribution for y^*?









Correct

Given the data, this is the distribution of the waiting time (in days) between significant earthquakes. It turns out that the first significant 4.0+ magnitude earthquake in California in 2016 occurred on January 6.