STA 6360, Report 4.6

Carson Slater *Baylor University*

1

Work Problem 12 on p. 82 in *BDA3*.

Poisson regression model: expand the model of Exercise 2.13(a) by assuming that the number of fatal accidents in year t follows a Poisson distribution with mean $\alpha + \beta t$. You will estimate α and β following th example of the analysis in Section 3.7.

(a)

Use the data for 1970-1999 to construct a power prior to be used in the analysis of the data from 2000-2008. Refer to airline fatality example on p.70 for the likelihood structure and other details. Your goal is to provide a posterior analysis of λ, as in the example. Do this for power parameter values $a_0 = 0.25$, 0.5, and 0.75.

Posterior Analysis of λ Using Power Priors

Model and Prior Structure

We assume the number of fatal accidents in a given year, Y_i, follows a Poisson distribution:

$$Y_i \sim \text{Poisson}(\lambda), \quad i = 1970, \ldots, 2008.$$

The likelihood for the historical data (1970–1999) is:

$$L(\lambda; \text{data}_{\text{hist}}) = \prod_{i=1970}^{1999} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}.$$

The power prior incorporates historical data into the prior for λ:

$$\pi(\lambda \mid a_0) \propto L(\lambda; \text{data}_{\text{hist}})^{a_0} \cdot \pi_0(\lambda),$$

where: - $a_0 \in [0, 1]$ is the power parameter, - $\pi_0(\lambda)$ is the initial prior for λ (assume $\pi_0(\lambda) \propto 1$, a non-informative prior).
The posterior distribution for λ using the data from 2000–2008 and the power prior is:

$$\pi(\lambda \mid \text{data}_{\text{recent}}, \text{data}_{\text{hist}}, a_0) \propto [L(\lambda; \text{data}_{\text{hist}})^{a_0} \cdot \pi_0(\lambda)] \cdot L(\lambda; \text{data}_{\text{recent}}).$$

Data Summary

- Historical data (1970–1999): $\sum_{i=1970}^{1999} y_i = 727$ (total fatal accidents over 30 years).
- Recent data (2000–2008): $\sum_{i=2000}^{2008} y_i = 133$ (total fatal accidents over 9 years).
- Number of years:
  - Historical: $n_{\text{hist}} = 30$,
  - Recent: $n_{\text{recent}} = 9$.

Posterior Distribution for λ

The posterior distribution of λ is:

$$\pi(\lambda \mid \text{data}, a_0) \propto \lambda^{a_0 \sum_{i=1970}^{1999} y_i + \sum_{i=2000}^{2008} y_i} e^{-\lambda(a_0 n_{\text{hist}} + n_{\text{recent}})}.$$

This corresponds to a Gamma distribution:

$$\lambda \mid \text{data}, a_0 \sim \text{Gamma}\left(\alpha, \beta\right),$$

where:

$$\alpha = a_0 \sum_{i=1970}^{1999} y_i + \sum_{i=2000}^{2008} y_i + 1, \quad \beta = a_0 n_{\text{hist}} + n_{\text{recent}}.$$

Analysis for Different a_0 Values

For $a_0 = 0.25, 0.5, 0.75$, we calculate the posterior parameters:

$$\alpha = a_0 \cdot 727 + 133 + 1,$$
$$\beta = a_0 \cdot 30 + 9.$$

Table 1: Posterior Parameters for λ

a_0	α	β
0.25	$727 \cdot 0.25 + 133 + 1 = 317.75$	$30 \cdot 0.25 + 9 = 16.5$
0.50	$727 \cdot 0.50 + 133 + 1 = 498.5$	$30 \cdot 0.50 + 9 = 24$
0.75	$727 \cdot 0.75 + 133 + 1 = 679.25$	$30 \cdot 0.75 + 9 = 31.5$

Posterior Mean and Variance

The posterior mean and variance for λ are:

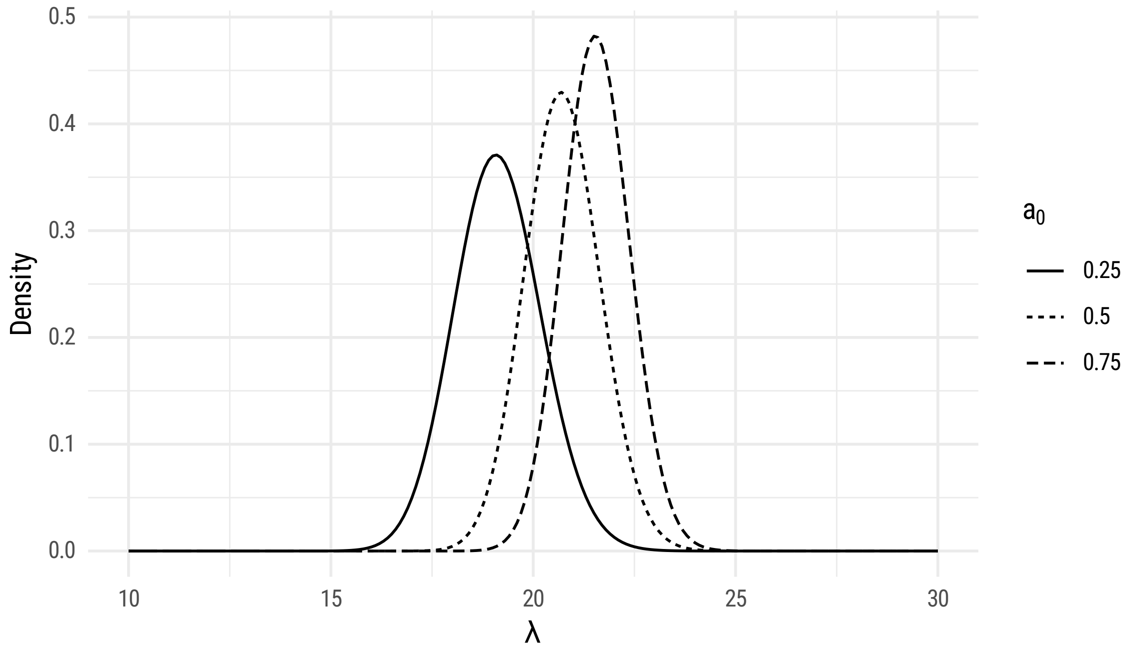
$$\text{Mean} = \frac{\alpha}{\beta}, \quad \text{Variance} = \frac{\alpha}{\beta^2}.$$

For each a_0:

$$\text{For } a_0 = 0.25: \quad \text{Mean} = \frac{317.75}{16.5}, \quad \text{Variance} = \frac{317.75}{16.5^2},$$
$$\text{For } a_0 = 0.50: \quad \text{Mean} = \frac{498.5}{24}, \quad \text{Variance} = \frac{498.5}{24^2},$$
$$\text{For } a_0 = 0.75: \quad \text{Mean} = \frac{679.25}{31.5}, \quad \text{Variance} = \frac{679.25}{31.5^2}.$$

Posterior Plots

Below is a plot of the posterior distributions for different a_0 values.



(b)

The former USSR did not provide data before 1986. Arguably, using the data before then to build a prior for the rate after 2000 is problematic. Use the data from 1986-1999 to form a power prior to be used with the data from 2000-2008. Again, provide a posterior analysis for λ. Do this for power parameter values $a_0 = 0.25, 0.5$, and 0.75. Did use of the pre–Soviet data matter?

Posterior Analysis of λ Using Power Priors (1986–1999 Data)

Model and Prior Structure

We assume the number of fatal accidents in a given year, Y_i, follows a Poisson distribution:

$$Y_i \sim \text{Poisson}(\lambda), \quad i = 1986, \ldots, 2008.$$

The likelihood for the historical data (1986–1999) is:

$$L(\lambda; \text{data}_{\text{hist}}) = \prod_{i=1986}^{1999} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}.$$

The power prior incorporates historical data into the prior for λ:

$$\pi(\lambda \mid a_0) \propto L(\lambda; \text{data}_{\text{hist}})^{a_0} \cdot \pi_0(\lambda),$$

where: - $a_0 \in [0, 1]$ is the power parameter, - $\pi_0(\lambda)$ is the initial prior for λ (assume $\pi_0(\lambda) \propto 1$, a non-informative prior).
The posterior distribution for λ using the data from 2000–2008 and the power prior is:

$$\pi(\lambda \mid \text{data}_{\text{recent}}, \text{data}_{\text{hist}}, a_0) \propto [L(\lambda; \text{data}_{\text{hist}})^{a_0} \cdot \pi_0(\lambda)] \cdot L(\lambda; \text{data}_{\text{recent}}).$$

Data Summary

- Historical data (1986–1999): $\sum_{i=1986}^{1999} y_i = 369$ (total fatal accidents over 14 years).
- Recent data (2000–2008): $\sum_{i=2000}^{2008} y_i = 133$ (total fatal accidents over 9 years).
- Number of years:
  - Historical: $n_{\text{hist}} = 14$,
  - Recent: $n_{\text{recent}} = 9$.

Posterior Distribution for λ

The posterior distribution of λ is:

$$\pi(\lambda \mid \text{data}, a_0) \propto \lambda^{a_0 \sum_{i=1986}^{1999} y_i + \sum_{i=2000}^{2008} y_i} e^{-\lambda(a_0 n_{\text{hist}} + n_{\text{recent}})}.$$

This corresponds to a Gamma distribution:

$$\lambda \mid \text{data}, a_0 \sim \text{Gamma}\left(\alpha, \beta\right),$$

where:

$$\alpha = a_0 \sum_{i=1986}^{1999} y_i + \sum_{i=2000}^{2008} y_i + 1, \quad \beta = a_0 n_{\text{hist}} + n_{\text{recent}}.$$

Analysis for Different a_0 Values

For $a_0 = 0.25, 0.5, 0.75$, we calculate the posterior parameters:

$$\alpha = a_0 \cdot 369 + 133 + 1,$$
$$\beta = a_0 \cdot 14 + 9.$$

Table 2: Posterior Parameters for λ (1986–1999 Data)

a_0	α	β
0.25	$369 \cdot 0.25 + 133 + 1 = 226.25$	$14 \cdot 0.25 + 9 = 12.5$
0.50	$369 \cdot 0.50 + 133 + 1 = 317.5$	$14 \cdot 0.50 + 9 = 16$
0.75	$369 \cdot 0.75 + 133 + 1 = 408.75$	$14 \cdot 0.75 + 9 = 19.5$

Posterior Mean and Variance

The posterior mean and variance for λ are:

$$\text{Mean} = \frac{\alpha}{\beta}, \quad \text{Variance} = \frac{\alpha}{\beta^2}.$$

For each a_0:

$$\text{For } a_0 = 0.25: \quad \text{Mean} = \frac{226.25}{12.5}, \quad \text{Variance} = \frac{226.25}{12.5^2},$$
$$\text{For } a_0 = 0.50: \quad \text{Mean} = \frac{317.5}{16}, \quad \text{Variance} = \frac{317.5}{16^2},$$
$$\text{For } a_0 = 0.75: \quad \text{Mean} = \frac{408.75}{19.5}, \quad \text{Variance} = \frac{408.75}{19.5^2}.$$

Did the Use of Pre–1986 Data Matter?

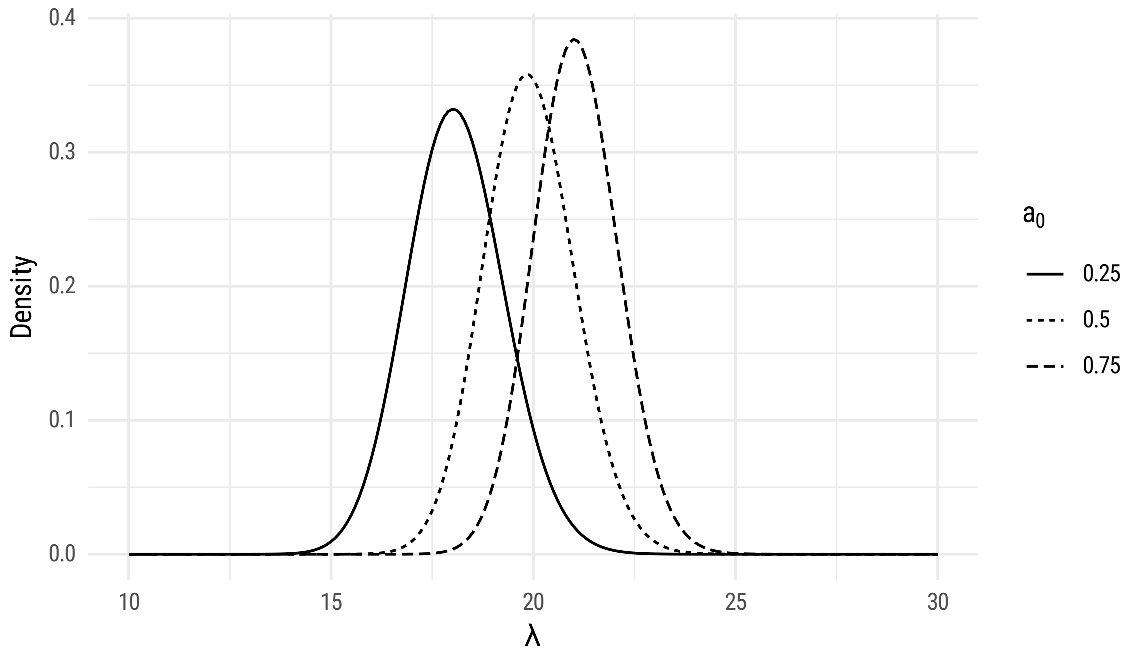
To assess the impact of pre–1986 data, we compare the posterior distributions derived from:

- Data from 1970–1999, and
- Data from 1986–1999.

Including pre–1986 data results in a larger posterior mean for λ due to the higher fatal accident rates in earlier years. The variance also decreases because of the increased historical sample size. By excluding pre–1986 data, the posterior reflects the lower fatality rates after 1986, particularly as air travel safety improved globally after the dissolution of the Soviet Union.

Posterior Plots

Below is a plot of the posterior distributions for different a_0 values.



(c)

Using the posterior predictive distribution corresponding to the model in (b), predict the number of deaths in 2009. Consult online sources for worldwide airline fatalities in 2009 for comparison to your prediction.

Perhaps the safest option would be to use the mean of the posterior predictive distribution, and I elect to use the one constructed with the a_0 in the power prior. This would be a $\text{gamma}(408.75, 19.5)$ distribution with mean $\frac{408.75}{19.5} = 20.96$.

(d)

Now, using a relatively $\text{gamma}(25, 1)$ prior for λ and the data from 1986 to 2008, obtain the posterior for λ and the posterior predictive distribution for the count in 2009. Compare your results to those obtained using the power prior with and without the pre–Soviet data.

Posterior Analysis of λ Using a Gamma$(25, 1)$ Prior

Prior

The prior distribution for λ is:

$$\lambda \sim \text{Gamma}(25, 1),$$

which has a mean of 25 and a variance of 25.

The likelihood for the data (1986–2008) is:

$$L(\lambda; \text{data}) = \prod_{i=1986}^{2008} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}.$$

The posterior distribution for λ is proportional to the product of the prior and the likelihood:

$$\pi(\lambda \mid \text{data}) \propto \lambda^{25-1} e^{-\lambda} \cdot \lambda^{\sum y_i} e^{-\lambda n},$$

where:

- $\sum y_i = 369 + 133 = 502$ (total fatal accidents from 1986–2008),
- $n = 14 + 9 = 23$ (number of years from 1986–2008).

Posterior Distribution

The posterior distribution for λ is:

$$\lambda \mid \text{data} \sim \text{Gamma}\left(\alpha, \beta\right),$$

where:

$$\alpha = 25 + \sum y_i = 25 + 502 = 527, \quad \beta = 1 + n = 1 + 23 = 24.$$

Thus:

$$\lambda \mid \text{data} \sim \text{Gamma}(527, 24).$$

The posterior mean and variance are:

$$\text{Mean} = \frac{\alpha}{\beta} = \frac{527}{24} \approx 21.96, \quad \text{Variance} = \frac{\alpha}{\beta^2} = \frac{527}{24^2} \approx 0.91.$$

Posterior Predictive Distribution for 2009

The posterior predictive distribution for the count in 2009, Y_{2009}, is:

$$Y_{2009} \sim \text{Poisson}(\lambda),$$

where λ is integrated over its posterior distribution. This leads to a negative binomial distribution:

$$Y_{2009} \sim \text{NegativeBinomial}\left(r = \alpha, p = \frac{\beta}{\beta + 1}\right).$$

The parameters for the negative binomial distribution are:

$$r = 527, \quad p = \frac{\beta}{\beta + 1} = \frac{24}{24 + 1} = \frac{24}{25} = 0.96.$$

The mean and variance of the posterior predictive distribution are:

$$\text{Mean} = \frac{r(1-p)}{p} = \frac{527(1 - 0.96)}{0.96} \approx 21.96,$$

$$\text{Variance} = \frac{r(1-p)}{p^2} = \frac{527(1 - 0.96)}{0.96^2} \approx 22.88.$$

Comparison with Power Priors

The results obtained using the $\text{Gamma}(25, 1)$ prior can be compared to those obtained using the power prior (both with and without the pre-Soviet data):

- **Power prior (with pre-Soviet data)**: This analysis leads to a higher posterior mean for λ, reflecting the influence of earlier years with higher accident rates. The variance is smaller due to the larger sample size.
- **Power prior (without pre-Soviet data)**: This analysis results in a lower posterior mean and a variance closer to the posterior derived from the $\text{Gamma}(25, 1)$ prior.
- **Gamma prior**: The mean posterior estimate of λ using the $\text{Gamma}(25, 1)$ prior is comparable to the results with the power prior (without pre-Soviet data). However, the posterior variance is smaller due to the informative nature of the prior.

Conclusion

Using the $\text{Gamma}(25, 1)$ prior provides results that are robust and reflective of the data from 1986–2008. Excluding pre-Soviet data in the power prior analysis aligns the posterior mean more closely with the results from the $\text{Gamma}(25, 1)$ prior. Including pre-Soviet data leads to a posterior that may overestimate the fatal accident rate due to the higher accident counts before 1986.