STA 360/602L: Module 2.8

Gamma-Poisson model II; finding conjugate distributions

Dr. Olanrewaju Michael Akande

Posterior predictive distribution

- What is the posterior predictive distribution for the Gamma-Poisson model?
- Let $a_n = a + \sum y_i$ and $b_n = b + n$.
- We have

$$
\begin{aligned}
p(y_{n+1}|y_{1:n}) &= \int p(y_{n+1}|\theta)\pi(\theta|y_{1:n})\, d\theta \\
&= \int \mathrm{Po}(y_{n+1}|\theta)\mathrm{Ga}(\theta|a_n, b_n)\, d\theta \\
&= \dots \\
&= \dots \\
&= \frac{\Gamma(a_n + y_{n+1})}{\Gamma(a_n)\Gamma(y_{n+1}+1)} \left(\frac{b_n}{b_n+1}\right)^{a_n} \left(\frac{1}{b_n+1}\right)^{y_{n+1}}
\end{aligned}
$$

which is the negative binomial distribution, $\text{Neg-binomial}\left(a_n, \frac{1}{b_n+1}\right)$.

Negative binomial distribution

- Originally derived as the number of successes in a sequence of independent $\text{Bernoulli}(p)$ trials before r failures occur.

- The negative binomial distribution $\text{Neg-binomial}\,(r, p)$ is parameterized by r and p and the pmf is given by

$$\Pr[Y = y | r, p] = \binom{y + r - 1}{y} (1-p)^r p^y; \quad y = 0, 1, 2, \ldots; \quad p \in [0, 1].$$

- Starting with this, the distribution can be extended to allow $r \in (0, \infty)$ as

$$\Pr[Y = y | r, p] = \frac{\Gamma(y + r)}{\Gamma(y+1)\Gamma(r)} (1-p)^r p^y; \quad y = 0, 1, 2, \ldots; \quad p \in [0, 1].$$

- Some properties:
  - $\mathbb{E}[\theta] = \dfrac{pr}{1-p}$
  - $\mathbb{V}[\theta] = \dfrac{pr}{(1-p)^2}$

Posterior predictive distribution

- The negative binomial distribution is an over-dispersed generalization of the Poisson.

- What does over-dispersion mean?

- In marginalizing θ out of the Poisson likelihood, over a gamma distribution, we obtain a negative-binomial.

- For $(y_{n+1}|y_{1:n}) \sim \text{Neg-binomial}\left(a_n, \dfrac{1}{b_n+1}\right)$, we have
  - $\mathbb{E}[y_{n+1}|y_{1:n}] = \dfrac{a_n}{b_n} = \mathbb{E}[\theta|y_{1:n}] =$ posterior mean, and
  - $\mathbb{V}[y_{n+1}|y_{1:n}] = \dfrac{a_n(b_n+1)}{b_n^2} = \mathbb{E}[\theta|y_{1:n}]\left(\dfrac{b_n+1}{b_n}\right)$,

  so that variance is larger than the mean by an amount determined by b_n, which takes the over-dispersion into account.

Predictive uncertainty

- Note that as the sample size n increases, the posterior density for θ becomes more and more concentrated.

$$\mathbb{V}[\theta|y_{1:n}] = \frac{a_n}{b_n^2} = \frac{a + \sum_i y_i}{(b+n)^2} \approx \frac{\bar{y}}{n} \to 0.$$

- Also, recall that $\mathbb{V}[y_{n+1}|y_{1:n}] = \mathbb{E}[\theta|y_{1:n}] \left(\frac{b_n + 1}{b_n} \right)$.

- As we have less uncertainty about θ, the inflation factor

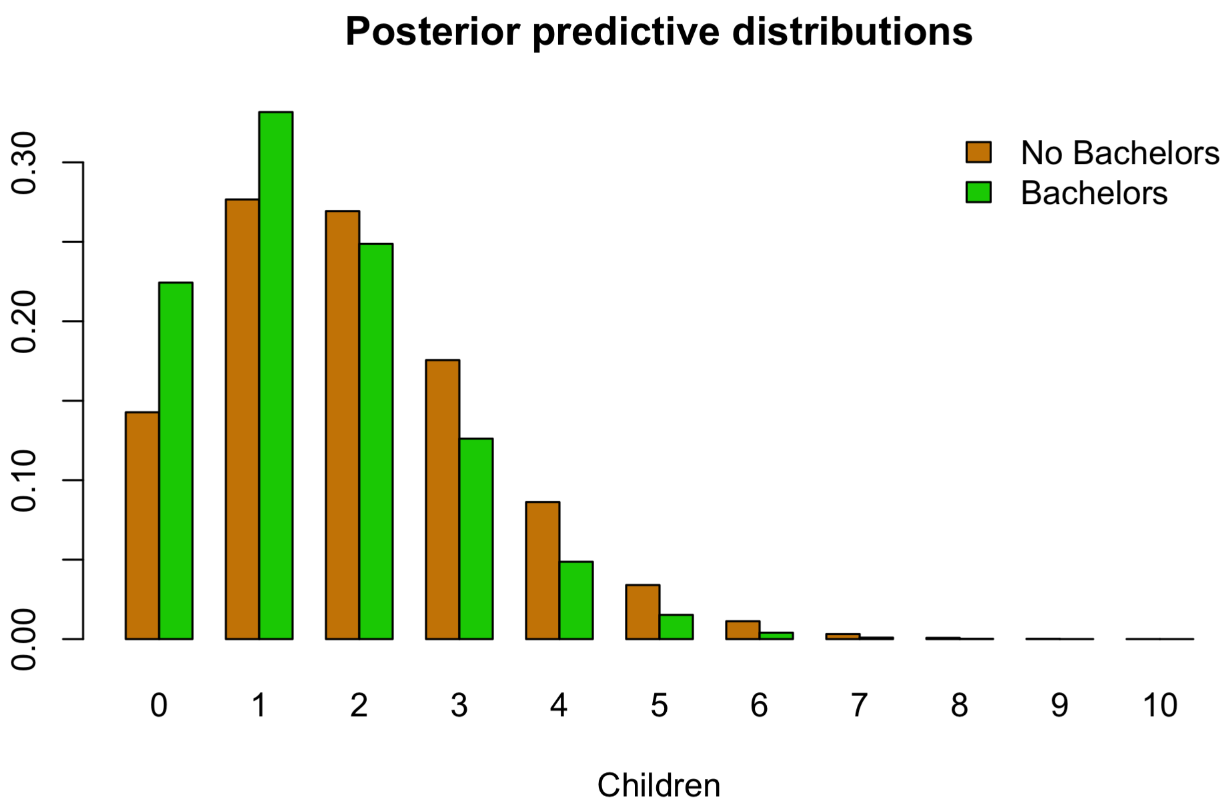
$$\frac{b_n + 1}{b_n} = \frac{b + n + 1}{b + n} \to 1$$

and the predictive density $f(y_{n+1}|y_{1:n}) \to \mathrm{Po}(\bar{y})$.

- Of course, in smaller samples, it is important to inflate our predictive intervals to account for uncertainty in θ.

Back to birth rates

- Let's compare the posterior predictive distributions for the two groups of women.



Poisson model in terms of rate

- In many applications, it is often convenient to parameterize the Poisson model a bit differently. One option takes the form

$$y_i \sim \text{Po}(x_i\theta); \quad i = 1, \ldots, n.$$

  where x_i represents an explanatory variable and θ is once again the population parameter of interest. The model is not exchangeable in the y_i's but is exchangeable in the pairs $(x, y)_i$.

- In epidemiology, θ is often called the population "rate" and x_i is called the "exposure" of unit i.

- When dealing with mortality rates in different counties for example, x_i can be the population n_i in county i, with $\theta =$ the overall mortality rate.

- The gamma distribution is still conjugate for θ, with the resulting posterior taking the form

$$\pi(\theta|\{x_i, y_i\}) : \theta|\{x_i, y_i\} \sim \text{Ga}(a + \sum_i y_i, b + \sum_i x_i).$$

BDA example: asthma mortality rate

- Consider an example on estimating asthma mortality rates for cities in the US.
- Since actual mortality rates can be small on the raw scale, they are often commonly estimated per 100,000 or even per one million.
- To keep it simple, let's use "per 100,000" for this example.
- For inference, ideally, we collect data which should basically count the number of asthma-related deaths per county.
- Note that inference is by county here, so county is indexes observations in the sample.
- Since we basically have count data, a Poisson model would be reasonable here.

Asthma mortality rate

- Since each city would be expected to have different populations, we might consider the sampling model:

$$y_i \sim \mathrm{Po}(x_i\theta); \quad i = 1, \ldots, n.$$

  where

  - x_i is the "exposure" for county i, that is, population of county i is $x_i \times 100,000$; and
  - θ is the unknown "true" city mortality rate per 100,000.

- Suppose

  - we pick a city in the US with population of 200,000;
  - we find that 3 people died of asthma, i.e., roughly 1.5 cases per 100,000.

- Thus, we have one single observation with $x_i = 2$ and $y_i = 3$ for this city.

Asthma mortality rate

- Next, we need to specify a prior. What is a sensible prior here?
- Perhaps we should look at mortality rates around the world or in similar countries.
- Suppose reviews of asthma mortality rates around the world suggest rates above 1.5 per 100,000 are very rare in Western countries, with typical rates around 0.6 per 100,000.
- Let's try a gamma distribution with $\mathbb{E}[\theta] = 0.6$ and $\Pr[\theta \geq 1.5]$ very low!
- A few options here, but let's go with $\mathrm{Ga}(3, 5)$, which has $\mathbb{E}[\theta] = 0.6$ and $\Pr[\theta \geq 1.5] \approx 0.02$.
- Using trial-and error, explore more options in R!

Asthma mortality rate

- Therefore, our posterior takes the form

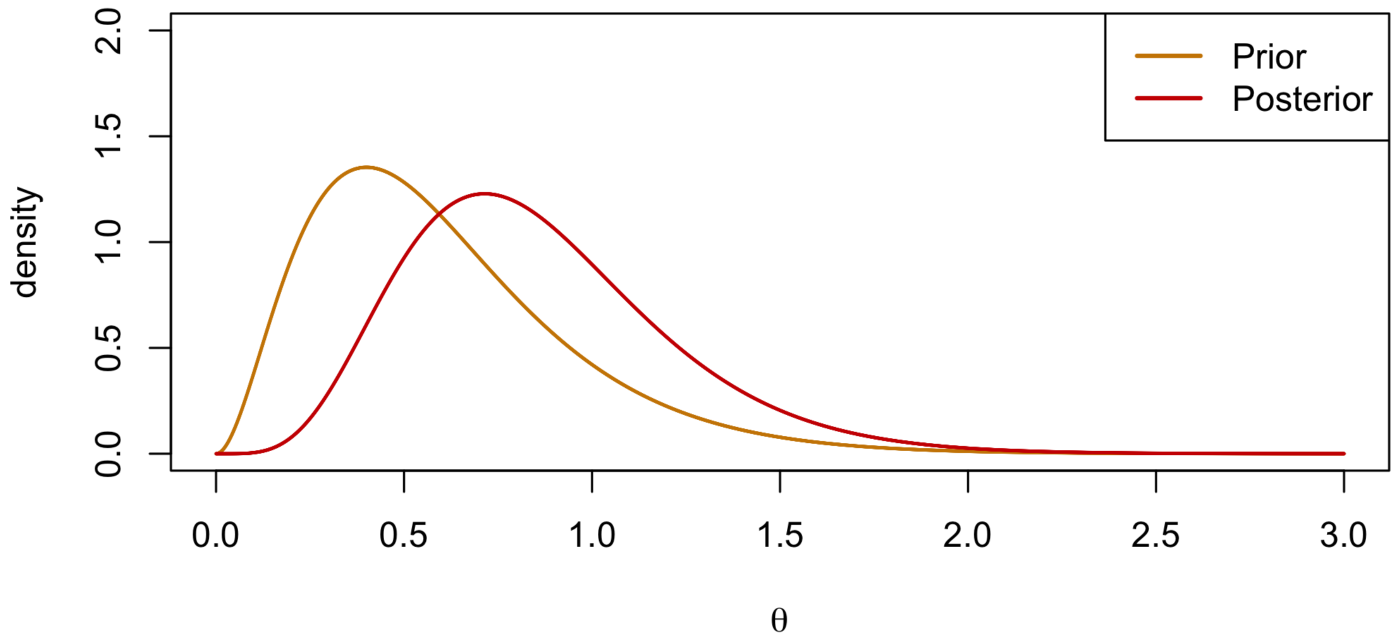
$$\pi(\theta|\{x_i, y_i\}) : \theta|\{x_i, y_i\} \sim \mathrm{Ga}(a + \sum_i y_i, b + \sum_i x_i)$$

  which is actually

$$\pi(\theta|x, y) = \mathrm{Ga}(a + y, b + x) = \mathrm{Ga}(3 + 3, 5 + 2) = \mathrm{Ga}(6, 7).$$

- $\mathbb{E}[\theta|x, y] = 6/7 = 0.86$ so that we expect less than 1 (0.86 to be exact) asthma-related deaths per 100,000 people in this city.
- In fact, the posterior probability that the long term death rate from asthma in this city is more than 1 per 100,000, $\Pr[\theta > 1|x, y]$, is 0.3.
- Also, $\Pr[\theta \leq 2|x, y] = 0.99$, so that there is very little chance that we see more than 2 asthma-related deaths per 100,000 people in this city.
- Use `pgamma` in R to compute the probabilities.

Prior vs posterior



Posterior is to the right of the prior since the data suggests higher mortality rates are more likely than the prior suggests. However, we only have one data point!

Finding conjugate distributions

Finding conjugate distributions

- In the conjugate examples we have looked at so far, how did we know the prior distributions we chose would result in conjugacy?

- Can we figure out the family of distributions that would be conjugate for arbitrary densities?

- Let's explore this using the exponential distribution. The exponential distribution is often used to model "waiting times" or other random variables (with support $(0, \infty)$) often measured on a time scale.

- If $y \sim \text{Exp}(\theta)$, we have the pdf

$$p(y|\theta) = \theta e^{-y\theta}; \quad y > 0.$$

where θ is the rate parameter, and $\mathbb{E}[y] = 1/\theta$.

- Recall, if $Y \sim \text{Ga}(1, \theta)$, then $Y \sim \text{Exp}(\theta)$. What is $\mathbb{V}[y]$ then?

- Let's figure out what the conjugate prior for this density would look like (to be done on the board).

What's next?

Move on to the readings for the next module!