$$Y \sim \text{Bin}(n, \theta)$$

$$P(\theta) \sim \text{Beta}(\alpha, \beta)$$

$$E[\theta] = \frac{\alpha}{\alpha+\beta}$$

$$\Longrightarrow P(\theta | y) \sim \text{Beta}(\alpha + y, n - y + \beta)$$

(Conjugacy)

pseudo success

pseudo fails

$$E[\theta | y] = \frac{\alpha + y}{n + \alpha + \beta} = w\frac{y}{n} + (1-w)\frac{\alpha}{\alpha+\beta}$$

"posterior mean"

$$w = \frac{n}{n+\alpha+\beta}$$

The Poisson Distribution

- Model for count data, $\{0, 1, \ldots, \}$
- Applications:
  - # of meteorites entering solar system.
  - # of patients $\to$ hospital
  - # of neurons firing.

$$Y \sim \text{Pois}(\lambda), \quad E[Y] = \text{Var}(Y) = \lambda$$

Poisson model

Assume $Y_1, \ldots, Y_n$ are n i.i.d. $\mathrm{Pois}(\lambda)$

$$P(y_1, \ldots, y_n | \lambda) = \prod_{i=1}^{n} P(y_i | \lambda)$$

$$= \prod^{n} \frac{\lambda^{y_i} e^{-\lambda}}{\cancel{y_i!}}$$

$$\hat{\lambda}_{MLE} = \frac{\sum y_i}{n}$$

$$\propto \lambda^{\sum y_i} e^{-n\lambda}$$

$$\propto L(\lambda)$$

$Y_i \overset{ind}{\sim} \text{Pois}(\lambda V_i)$ ← Known Data

Unknown param.

λ is expected counts per unit "time"

V_i is length of "time" for obs. i.

$$P(Y_1, \ldots Y_n \mid \lambda, V_1, \ldots V_n) =$$

$$\prod_{i=1}^{n} P(Y_i \mid \lambda, V_i) = \prod^{n} \frac{e^{-(\lambda V_i)} (\lambda V_i)^{y_i}}{y_i!}$$

$$= \frac{e^{-\lambda \sum^n V_i} \lambda^{\sum y_i} \cancel{\left(\prod V_i^{y_i}\right)}}{\cancel{\prod y_i!}} \propto e^{-\lambda \sum V_i} \lambda^{\sum y_i}$$

$$= L(\lambda)$$

$$\ell(\lambda) = \log L = -\lambda \sum V_i + \sum y_i \log(\lambda)$$

$$\ell' = -\sum v_i + \frac{\sum y_i}{\lambda} = 0$$

$$\hat{\lambda}_{MLE} = \frac{\sum y_i}{\sum v_i} \begin{array}{l} \leftarrow \text{total counts} \\ \leftarrow \text{total "time"} \end{array}$$

$$L(\lambda) \propto e^{-\lambda \sum v_i} \lambda^{\sum y_i}$$

$$P(\lambda) \propto e^{-\lambda k_1} \lambda^{k_2}$$

$$\text{Gamma}(a, b) = \frac{b^a}{\Gamma(a)} e^{-b\lambda} \lambda^{a-1}$$

normalizing constant

$$y_i \overset{ind}{\sim} \text{Pois}(\lambda v_i)$$

$$p(\lambda) \sim \text{Gam}(a, b)$$

$$P(\lambda \mid y_1, \dots y_n) \propto L(\lambda)\, P(\lambda)$$

$$= e^{-\lambda \sum v_i} \lambda^{\sum y_i} e^{-b\lambda} \lambda^{a-1} \times \text{consts}$$

$$\propto e^{-\lambda \underbrace{(\sum v_i + b)}_{b_{new}}} \lambda^{\overbrace{\sum y_i + a}^{}-1}$$

$$\lambda^{\underbrace{\sum y_i + a}_{a_{new}} - 1}$$

$$P(\lambda \mid y_1, \dots y_n) \sim \text{Gam}(\sum y_i + a, \sum v_i + b)$$

Gamma is conjugate for Pois.

Poisson model with exposure

- Often times we include an “exposure” term in the Poisson model:

Known constant: "how long"

$$p(y_i \mid \nu_i \lambda) = (\nu_i \lambda)^y e^{\nu_i \lambda} / y_i!$$

- How many cars do we expect to pass an intersection in one hour? How many in two hours?
  - If we model the distribution as Poisson, we expect twice as many in two hours as in one hours.
- Homework: exposure is the length of the chapter

$$Y_i \sim \text{Pois}(\lambda \nu_i)$$

$$P(Y_i \mid \underbrace{\lambda \nu_i}) \propto \frac{(\lambda \nu_i)^{y_i} e^{-\nu_i \lambda}}{y_i!}$$

expected counts per unit "time"

Poisson model example

- In a particular county 3 people out of a population of 100,000 died of asthma
- Assume a Poisson sampling model with rate λ (units are rate of deaths per 100,000 people)
- How do we specify a prior distribution for λ?
- How would our Bayesian estimate for λ differ?

$y_i \sim \text{Pois}(\lambda \nu_i)$

λ = Rate of death per 100,000

$\hat{\lambda}_{MLE}$: 3 (per 100K)

$\nu_i = 1$

$P(\lambda \mid y = 3)$ needed.

Choose $P(\lambda)$. How?

- Air Quality of county.
- Previous years' data
- Info from other counties.
- Regress on more info?

$P(\lambda \mid y_1, \ldots)$

Conjugate Prior for the Poisson

Assume n i.i.d observations of a $\text{Poisson}(\lambda)$

$$\begin{aligned} p(\lambda \mid y_1, \ldots y_n) &\propto L(\lambda) \times p(\lambda) \\ &\propto \lambda^{\sum y_i} e^{-n\lambda} \times p(\lambda) \end{aligned}$$

- A prior distribution for λ should have support on $\mathbb{R}^+$, the positive real line
- Bayesian definition of sufficiency: $p(\lambda \mid s, y_1, \ldots y_n) = p(\lambda \mid s)$
  - For the Poisson, $\sum y_i$ is sufficient
- Can we find a density of the form $p(\lambda) \propto \lambda^{k_1} e^{k_2 \lambda}$?

Conjugate Prior for the Poisson

Assume n i.i.d observations of a $\text{Poisson}(\lambda)$

$$\begin{aligned} p(\lambda \mid y_1, \ldots y_n) &\propto L(\lambda) \times p(\lambda) \\ &\propto \lambda^{\sum y_i} e^{-n\lambda} \times p(\lambda) \end{aligned}$$

- A prior distribution for λ should have support on $\mathbb{R}^+$, the positive real line
- Bayesian definition of sufficiency: $p(\lambda \mid s, y_1, \ldots y_n) = p(\lambda \mid s)$
  - For the Poisson, $\sum y_i$ is sufficient
- Can we find a density of the form $p(\lambda) \propto \lambda^{k_1} e^{k_2 \lambda}$?

- $\text{Gamma}(a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$

The Gamma distribution Gam(a, b)

Useful properties of the Gamma distribution:

- $\boxed{E[\lambda] = a/b}$
- $\text{Var}[\lambda] = a/b^2$
- $\text{mode}[\lambda] = (a-1)/b$ if $a > 1, 0$ otherwise
- In R: `dgamma`, `rgamma`, `pgamma`, `qgamma`

The Gamma distribution

Mean 1

Mean 2

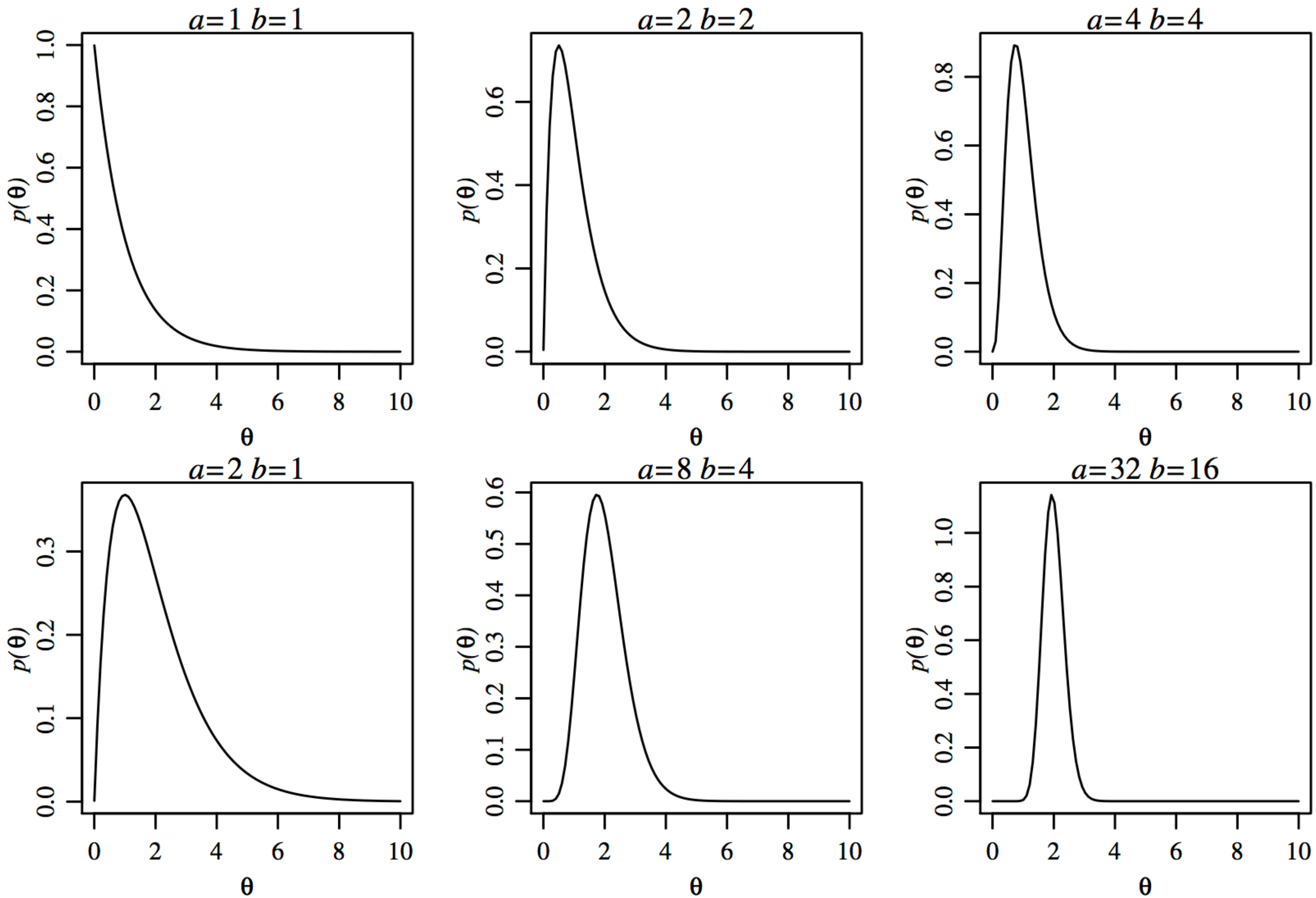


Fig. 3.8. Gamma densities.

$$P(\lambda \mid y_1, \dots y_n) \sim \mathrm{Gam}\Big(\underbrace{\sum y_i}_{\text{total counts}} + \underbrace{a}_{\text{pseudo counts}}, \underbrace{\sum v_i}_{\text{total "time"}} + \underbrace{b}_{\text{pseudo time}}\Big)$$

$$p(\lambda) \sim \mathrm{Gam}(a, b)$$

$$E[\lambda] = \frac{a}{b} \leftarrow \text{pseudo counts per time}$$

prior <u>mean</u>

$$E[\lambda \mid y_1, \dots y_n] = \frac{\sum y_i + a}{\sum v_i + b} =$$

posterior mean

$$\frac{\sum v_i}{\sum v_i} \frac{\sum y_i}{\sum v_i + b} + \frac{b}{b} \frac{a}{\sum v_i + b} =$$

$$\underbrace{\frac{\sum v_i}{\sum v_i + b}} \underbrace{\frac{\sum y_i}{\sum v_i}} + \frac{b}{\sum v_i + b} \underbrace{\frac{a}{b}}$$

prior

$w \quad \lambda_{MLE} \quad (1-w)\ \lambda_{\text{mean}}$

$$E[\lambda | y_1, \dots y_n] = w\hat{\lambda}_{MLE} + (1-w)\lambda_{\text{prior mean}}$$

$$w = \frac{\sum v_i}{\sum v_i + b}$$

The posterior in the Poisson-Gamma model

Assume one observation with $y_i \sim \text{Pois}(\lambda \nu_i)$ where ν_i is the exposure

$$
\begin{aligned}
p(\lambda \mid y_i) &\propto L(\lambda) \times p(\lambda) \\
&\propto (\lambda \nu_i)^{y_i} e^{-\lambda \nu_i} \times \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \\
&\propto (\lambda)^{y_i + a - 1} e^{-(b + \nu_i)\lambda}
\end{aligned}
$$

$$
p(\lambda \mid y, a, b) = \text{Gamma}(y_i + a, b + \nu_i)
$$

What is the posterior distribution for n observations, y_1, … y_n, with exposures $\nu_1 \ldots \nu_n$?

Poisson model example

λ: deaths per 100K/year

$= y$

- In a particular county 3 people out of a population of 100,000 died of asthma

$\nu = 1$

- Assume a Poisson sampling model with rate λ
  - Units are rate of deaths per 100,000 people/year
- Experts know that typical rates of asthma mortality in the US are closer to 0.6 per 100,000
- Let's choose a Gamma distribution with a mean of 0.6 and appropriate uncertainty.

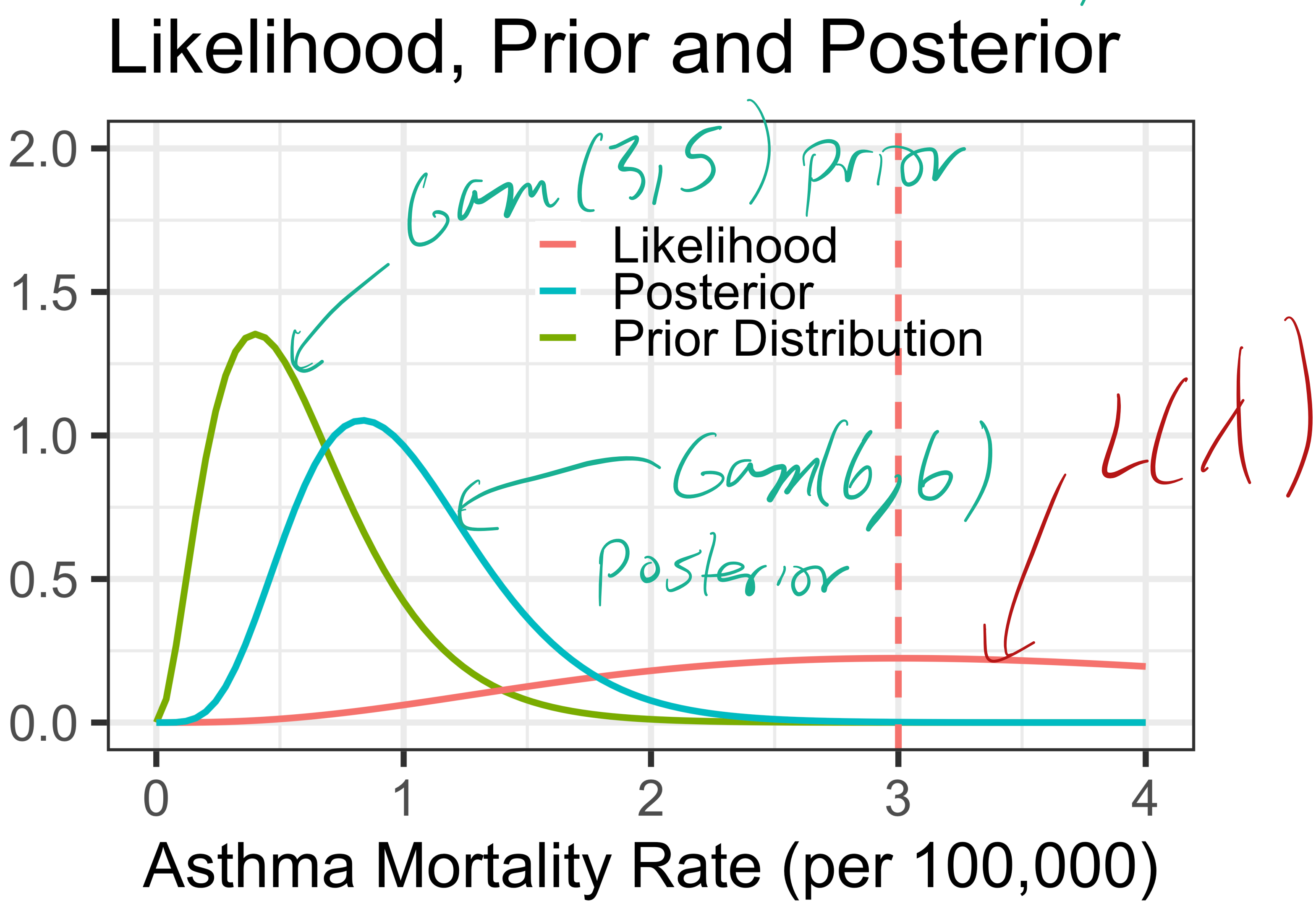
$\frac{3}{5} = .6$

$P(\lambda) \sim Gam(3, 5)$

Possible Gamma prior distributions

Asthma Mortality

$Gam(\sum y_i + a, \sum v_i + b)$



Using Gamma(3, 5) prior distribution

The posterior mean

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

- $w \to 1$ as $n \to \infty$ (data dominates prior)
- b can be interpreted as the number of *prior* observations
  - Analogous to n or total prior exposure
- a can be interpreted as the sum of the counts from prior total exposure of b
  - Analogous to $\sum_i y_i$

Summary

- The Beta distribution
  - Conjugate prior for Binomial likelihood
- The Gamma distribution
  - Conjugate prior for the Poisson likelihood
- Pseudo-counts interpretations of conjugate prior distributions

- Post. mean is weighted avg of MLE & prior mean.