Bayesian Statistics

Chapter 3. One Parameter Models

Hojin Yang

Department of Statistics
Pusan National University

Introduction

- A one-parameter model
- A class of sampling distributions that is indexed by a single unknown parameter.
- We discuss Bayesian inference for two one-parameter models: the binomial model and the Poisson model
  - Conjugate prior distributions
  - Predictive distributions
  - Confidence regions

3.1. Binomial Model

- Happiness data
- Each female of age 65 or over in the 1998 General Social Survey was asked whether or not they were generally happy.
- Let $Y_i = 1$ if respondent i reported being generally happy, and $Y_i = 0$ otherwise.
- $n = 129$ individuals are responses as being exchangeable
- For the total size N of the female senior citizen population, our beliefs about $Y_1, \ldots, Y_{129}$ are well approximated by
  - Our beliefs about $\theta = \sum_{i=1}^{N} Y_i/N$
  - Conditional on the θ, Y_i's are i.i.d. binary random variables with expectation θ

- This says that the probability for any potential outcome $\{y_1, \ldots, y_{129}\}$, conditional on θ, is given by

$$p(y_1, \ldots, y_{129}|\theta) = \theta^{\sum_{i=1}^{129} y_i}(1-\theta)^{129-\sum_{i=1}^{129} y_i}$$

- What remains to be specified is our prior distribution

- A uniform prior distribution
- The parameter θ is some unknown number between 0 and 1
- This condition implies that our density for θ must be the uniform density

$$p(\theta) = 1 \text{ for all } \theta \in [0, 1]$$

- Bayes' rule gives

$$\begin{aligned} p(\theta | y_1, \ldots, y_{129}) &= \frac{p(y_1, \ldots, y_{129} | \theta) p(\theta)}{p(y_1, \ldots, y_{129})} \\ &= p(y_1, \ldots, y_{129} | \theta) \times \frac{1}{p(y_1, \ldots, y_{129})} \\ &\propto p(y_1, \ldots, y_{129} | \theta) . \end{aligned}$$

- This says $p(\theta | \mathbf{y})$ and $p(\mathbf{y} | \theta)$ are proportional to each other as functions of θ

- This is because the posterior distribution is equal to $p(\mathbf{y}|\theta)$ divided by something that does not depend on θ
- This means that these two functions of θ have the same shape, but not necessarily the same scale.
- Data and posterior distribution
    - 129 individuals surveyed
    - 118 individuals report being generally happy (91%)
    - 11 individuals do not report being generally happy (9%)
- The probability of these data for a given value of a given value of θ

$$p(y_1, \ldots, y_{129}|\theta) = \theta^{118}(1-\theta)^{11}$$

- We will see the scale of $p(\theta|\mathbf{y})$ as well as the shape
- From Bayes' rule, we have

$$\begin{aligned} p(\theta|y_1,\ldots,y_{129}) &= \theta^{118}(1-\theta)^{11} \times p(\theta)/p(y_1,\ldots,y_{129}) \\ &= \theta^{118}(1-\theta)^{11} \times 1/p(y_1,\ldots,y_{129}) \end{aligned}$$

- We can calculate the scale (normalizing constant) using

$$\int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

- Let's recall what we know about $p(\theta|y_1,\dots,y_{129})$
  - $\int_0^1 p(\theta|y_1,\dots,y_{129}) = 1$ (integrate or sum to 1)
  - $\int_0^1 p(\theta|y_1,\dots,y_{129}) = \theta^{118}(1-\theta)^{11}/p(y_1,\dots,y_{129})$ (Bayes)
- Therefore

$$1 = \int_0^1 p(\theta|y_1,\dots,y_{129})\ d\theta$$

$$1 = \int_0^1 \theta^{118}(1-\theta)^{11}/p(y_1,\dots,y_{129})\ d\theta$$

$$1 = \frac{1}{p(y_1,\dots,y_{129})}\int_0^1 \theta^{118}(1-\theta)^{11}\ d\theta$$

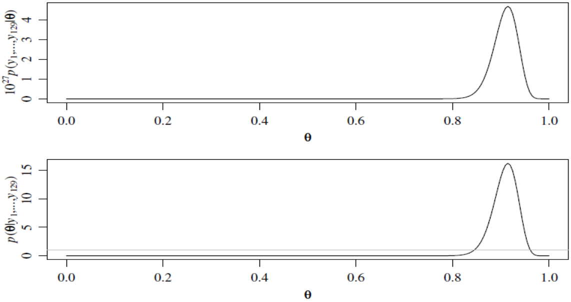
$$1 = \frac{1}{p(y_1,\dots,y_{129})}\frac{\Gamma(119)\Gamma(12)}{\Gamma(131)}$$

$$p(y_1,\dots,y_{129}) = \frac{\Gamma(119)\Gamma(12)}{\Gamma(131)}$$

- Putting everything together, we have

$$
\begin{aligned}
p(\theta|y_1,\ldots,y_{129}) &= \frac{\Gamma(131)}{\Gamma(119)\Gamma(12)}\theta^{118}(1-\theta)^{11} \\
&= \frac{\Gamma(131)}{\Gamma(119)\Gamma(12)}\theta^{119-1}(1-\theta)^{12-1}
\end{aligned}
$$

- This density for θ is called a beta distribution with parameters $a = 119$ and $b = 12$



- Beta distribution: unknown θ to be between 0 and 1

$$p(\theta) = \mathrm{dbeta}(\theta, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1} \quad \text{for } 0 \le \theta \le 1$$

- For such a random variable

$$\begin{aligned}
&\mathrm{mode}[\theta] = (a-1)/[(a-1)+(b-1)] \text{ if } a > 1 \text{ and } b > 1; \\
&\mathrm{E}[\theta] = a/(a+b); \\
&\mathrm{Var}[\theta] = ab/[(a+b+1)(a+b)^2] = \mathrm{E}[\theta] \times \mathrm{E}[1-\theta]/(a+b+1)
\end{aligned}$$

- For our happiness data

$$\begin{aligned}
\mathrm{mode}[\theta|y_1, \ldots, y_{129}] &= 0.915 \\
\mathrm{E}[\theta|y_1, \ldots, y_{129}] &= 0.908 \\
\mathrm{sd}[\theta|y_1, \ldots, y_{129}] &= 0.025
\end{aligned}$$

- Binomial distribution: r.v. $Y \in \{0, 1, \ldots, n\} \sim B(n, \theta)$ if

$$\Pr(Y = y|\theta) = \text{dbinom}(y, n, \theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, \quad y \in \{0, 1, \ldots, n\}$$

- For such a random variable

$$\begin{aligned} \mathrm{E}[Y|\theta] &= n\theta \\ \mathrm{Var}[Y|\theta] &= n\theta(1-\theta) \end{aligned}$$

- Having observed $Y = y$, posterior distribution of θ

$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta)p(\theta)}{p(y)} \\ &= \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}p(\theta)}{p(y)} \\ &= c(y)\theta^y(1-\theta)^{n-y}p(\theta) \end{aligned}$$

where $c(y)$ is a function of y and not of θ

- Posterior inference under a uniform prior distribution
- For the uniform distribution with $p(\theta) = 1$, we can find $c(y)$

$$1 = \int_0^1 c(y)\theta^y(1-\theta)^{n-y}\, d\theta$$
$$1 = c(y)\int_0^1 \theta^y(1-\theta)^{n-y}\, d\theta$$
$$1 = c(y)\frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)}$$

- The normalizing constant $c(y)$ is equal to a reciprocal
- We have

$$\begin{aligned} p(\theta|y) &= \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^y(1-\theta)^{n-y} \\ &= \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^{(y+1)-1}(1-\theta)^{(n-y+1)-1} \\ &= \mathrm{beta}(y+1, n-y+1) \end{aligned}$$

- Posterior distributions under beta prior distributions
- Suppose $\theta \sim \text{beta}(a, b)$ and $Y|\theta \sim B(n, \theta)$
- Observing $Y = y$

$$
\begin{aligned}
p(\theta|y) &= \frac{p(\theta)p(y|\theta)}{p(y)} \\
&= \frac{1}{p(y)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} \times \binom{n}{y} \theta^{y}(1-\theta)^{n-y} \\
&= c(n, y, a, b) \times \theta^{a+y-1}(1-\theta)^{b+n-y-1} \\
&= \text{dbeta}(\theta, a+y, b+n-y)
\end{aligned}
$$

- Second to last line says $p(\theta|y) \propto \theta^{a+y-1}(1-\theta)^{b+n-y-1}$
- This means that it has the same shape as the beta density

Definition: Conjugate

A class $\mathcal{P}$ of prior distributions for θ is called conjugate for a sampling model $p(y|\theta)$ if

$$p(\theta) \in \mathcal{P} \Rightarrow p(\theta|y) \in \mathcal{P}$$

- Conjugate prior distributions are very flexible and are computationally tractable
- For instance, we say that the class of beta priors is conjugate for the binomial sampling model

- Combining information

If $\theta|\{Y=y\} \sim \text{beta}(a+y, b+n-y)$, then

$$\mathrm{E}[\theta|y] = \frac{a+y}{a+b+n}, \ \text{mode}[\theta|y] = \frac{a+y-1}{a+b+n-2}, \ \mathrm{Var}[\theta|y] = \frac{\mathrm{E}[\theta|y]\mathrm{E}[1-\theta|y]}{a+b+n+1}$$

- The posterior expectation $E[\theta|y]$ is easily recognized as a combination of prior and data information:

$$\begin{aligned}
\mathrm{E}[\theta|y] &= \frac{a+y}{a+b+n} \\
&= \frac{a+b}{a+b+n}\frac{a}{a+b} + \frac{n}{a+b+n}\frac{y}{n} \\
&= \frac{a+b}{a+b+n} \times \text{prior expectation} + \frac{n}{a+b+n} \times \text{data average}
\end{aligned}$$

- This leads to the interpretation of a and b as “prior data”:
  - $a \approx$ prior number of 1’s
  - $b \approx$ prior number of 0’s
  - $a + b \approx$ prior sample size
- If sample size n is larger than our prior sample size $a + b$, then a majority of our information about θ should be coming from the data
- If $n \gg a + b$

$$\frac{a+b}{a+b+n} \approx 0 \ , \ \mathrm{E}[\theta|y] \approx \frac{y}{n} \ , \ \mathrm{Var}[\theta|y] \approx \frac{1}{n}\frac{y}{n}\left(1-\frac{y}{n}\right)$$

Prediction

- An important feature of Bayesian inference is the existence of a predictive distribution for new observations
- Let $y_1, \ldots, y_n$ be a sample of n binary r.v.s
- Let $\tilde{Y} \in \{0, 1\}$ be a new outcome from the same dist
- The predictive distribution of $\tilde{Y}$ is the conditional distribution of $\tilde{Y}$ given $\{Y_1 = y_1, \ldots, Y_n = y_n\}$, i.e.,

$$p(\tilde{Y} | Y_1 = y_1, \ldots, Y_n = y_n)$$

- The prediction probability for $\tilde{Y} = 1$

$$
\begin{aligned}
\Pr(\tilde{Y} = 1|y_1, \ldots, y_n) &= \int \Pr(\tilde{Y} = 1, \theta|y_1, \ldots, y_n)\ d\theta \\
&= \int \Pr(\tilde{Y} = 1|\theta, y_1, \ldots, y_n) p(\theta|y_1, \ldots, y_n)\ d\theta \\
&= \int \theta p(\theta|y_1, \ldots, y_n)\ d\theta \\
&= \mathrm{E}[\theta|y_1, \ldots, y_n] = \frac{a + \sum_{i=1}^{n} y_i}{a + b + n}
\end{aligned}
$$

- The prediction probability for $\tilde{Y} = 0$

$$
\Pr(\tilde{Y} = 0|y_1, \ldots, y_n) = 1 - \mathrm{E}[\theta|y_1, \ldots, y_n] = \frac{b + \sum_{i=1}^{n}(1 - y_i)}{a + b + n}
$$

- Notice two important things about the predictive distribution
    - The predictive distribution does not depend on any unknown quantities. If it did, we would not be able to use it to make predictions.
    - The predictive distribution depends on our observed data. In this distribution, $\tilde{Y}$ is not independent of $Y_1, \ldots, Y_n$. This is because observing $Y_1, \ldots, Y_n$ gives information about θ, which in turn gives information about $\tilde{Y}$.
        - It would be bad if $\tilde{Y}$ were independent of $Y_1, \ldots, Y_n$ - it would mean that we could never infer anything about the unsampled population from the sample cases.

- Example
- If $\theta \sim \text{beta}(1,1)$ and $Y|\theta \sim B(n,\theta)$
- $\theta|y \sim \text{beta}(1+Y, 1+n-Y)$

$$\Pr(\tilde{Y}=1|Y=y) = \mathrm{E}[\theta|Y=y] = \frac{2}{2+n}\frac{1}{2} + \frac{n}{2+n}\frac{y}{n}$$

where $Y = \sum_{i=1}^{n} Y_i$

Confidence Regions

Definition: Bayesian coverage

After observing the data $Y = y$, an interval $[L(y), U(y)]$ has 95% Bayesian coverage for θ if

$$Pr(L(y) \leq \theta \leq U(y) | Y = y) = .95$$

- The interpretation of this interval is that it describes our information about the location of the true value of θ after you have observed the data
- This is different from the frequentist interpretation of coverage probability, which describes the probability that the interval will cover the true value before the data are observed

Definition: Frequentist coverage

Before the data are gathered, a random interval $[L(y), U(y)]$ has 95% frequentist coverage for θ if

$$Pr(L(y) \leq \theta \leq U(y) | \theta) = .95$$

- The frequentist and Bayesian notions of coverage describe pre- and post-experimental coverage, respectively.
- For the 95% frequentist coverage, we can expect that 95% of your intervals contain the correct parameter value, while we can directly account for the parameter with the probability for the Bayesian coverage

Credible Interval (quantile-based interval)

- The easiest way to obtain a confidence interval is to use posterior quantiles.
- To make a $100 \times (1-\alpha)\%$ quantile-based ci. find numbers $\theta_{\alpha/2} < \theta_{1-\alpha/2}$ such that
  - $P(\theta < \theta_{2/\alpha} | Y = y) = \alpha/2$
  - $P(\theta > \theta_{1-\alpha/2} | Y = y) = \alpha/2$
- The numbers $\theta_{\alpha/2} < \theta_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ posterior quantiles of θ

$$\begin{aligned}
\Pr(\theta \in [\theta_{\alpha/2}, \theta_{1-\alpha/2}] | Y = y) &= 1 - \Pr(\theta \notin [\theta_{\alpha/2}, \theta_{1-\alpha/2}] | Y = y) \\
&= 1 - [\Pr(\theta < \theta_{\alpha/2} | Y = y) + \Pr(\theta > \theta_{1-\alpha/2} | Y = y)] \\
&= 1 - \alpha
\end{aligned}$$

- Binomial sampling and uniform prior
- Suppose out of $n = 10$ conditionally independent draws of a binary random variable we observe $Y = 2$ ones
- Using a uniform prior distribution
- The posterior distribution is $\theta|\{Y = 2\} \sim \text{beta}(1 + 2, 1 + 8)$
- The 95% posterior confidence interval can be obtained from the .025 and .975 quantiles of the posterior

```
> a<-1  ; b<-1   #prior
> n<-10 ; y<-2   #data

> qbeta( c(.025,.975), a+y,b+n-y)

[1] 0.06021773 0.51775585
```

Highest Posterior Density (HPD) Interval

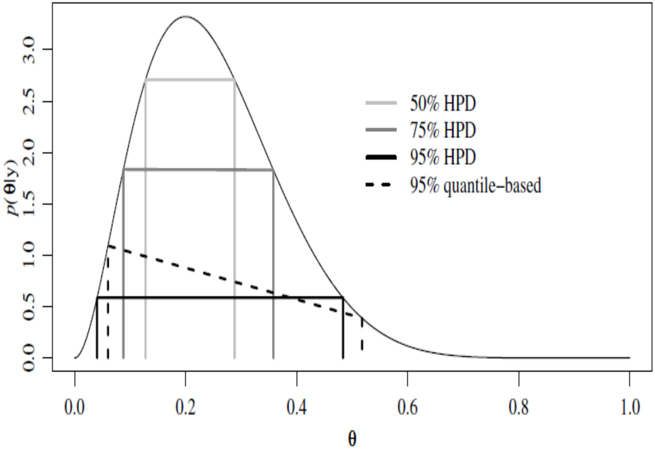
Definition: HPD region

A $100 \times (1-\alpha)\%$ HPD region consists of a subset of the parameter space $S(y) \subset \Theta$ such that

1.$Pr(\theta \in S(y)|Y = y) = 1 - \alpha$

2.If $\theta_a \in S(y)$ and $\theta_b \notin S(y),$ then $Pr(\theta_a|Y = y) > Pr(\theta_b|Y = y)$

- All points in an HPD region have a higher posterior density than points outside the region.
- HPD region might not be an interval if the posterior density is multimodal



- Highest posterior density regions of varying probability content. The dashed line is the 95% quantile-based interval
- The 95% HPD region is $[0.04, 0.048]$, which is narrower (more precise) than the credible interval

3.2. Poisson Model

- Some measurements, such as a person's number of children or number of friends, have values that are whole numbers
- In these cases our sample space is $\mathcal{Y} = \{0, 1, 2, \ldots\}$
- Perhaps the simplest probability model on Y is the Poisson model

- Poisson distribution: $Y \sim \text{Poi}(\theta)$

$$\Pr(Y = y|\theta) = \text{dpois}(y, \theta) = \theta^y e^{-\theta}/y! \;\; \text{for } y \in \{0, 1, 2, \ldots\}$$

- For such a random variable

$$\mathrm{E}[Y|\theta] = \theta$$
$$\mathrm{Var}[Y|\theta] = \theta$$

- Note that if $Y_1, \ldots, Y_n \sim \text{Poi}(\theta)$, then $\sum_{i=1}^n Y_i$ is a sufficient statistic
- Furthermore, $\sum_{i=1}^n Y_i|\theta \sim \text{Poi}(n\theta)$

- We will work with a class of conjugate prior distributions
- For the Poisson sampling model

$$
\begin{aligned}
p(\theta|y_1,\dots,y_n) &\propto p(\theta) \times p(y_1,\dots,y_n|\theta) \\
&\propto p(\theta) \times \theta^{\sum y_i} e^{-n\theta}.
\end{aligned}
$$

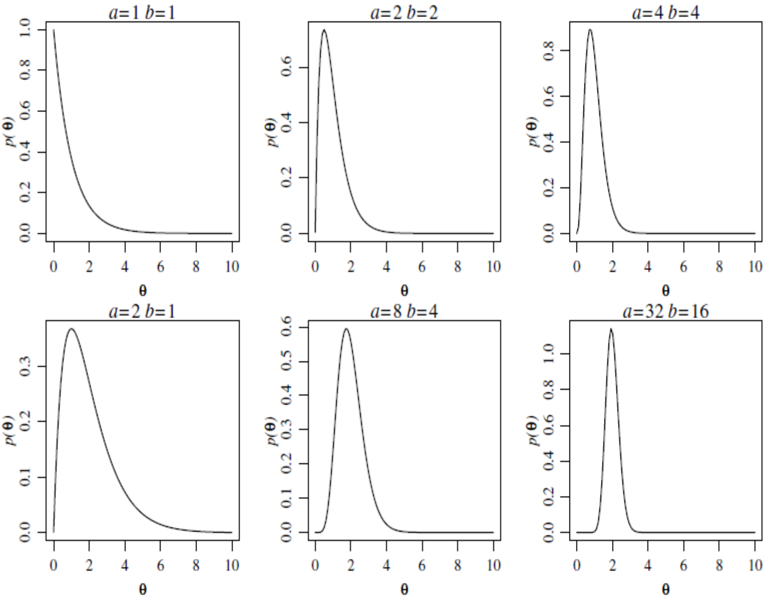
- This means that whatever our conjugate class of densities is, it will have $\theta^{c_1} e^{-c_2\theta}$ for numbers c_1 and c_2
- The simplest class of such densities includes the family of gamma distributions

- Gamma distribution: $\theta \sim G(a, b)$ if

$$p(\theta) = \text{dgamma}(\theta, a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}, \quad \text{for } \theta,\ a,\ b > 0.$$

- For such a random variable

$$\begin{aligned} \text{E}[\theta] &= a/b; \\ \text{Var}[\theta] &= a/b^2; \\ \text{mode}[\theta] &= \begin{cases} (a-1)/b & \text{if } a > 1 \\ 0 & \text{if } a \leq 1 \end{cases} \end{aligned}$$



a=1 b=1
a=2 b=2
a=4 b=4
a=2 b=1
a=8 b=4
a=32 b=16
p(θ)
θ

- Posterior distribution of θ
- Suppose $Y_1, \ldots, Y_n|\theta \sim \text{Poi}(\theta)$ and $p(\theta) \sim G(a, b)$
- Then

$$\begin{aligned} p(\theta|y_1,\ldots,y_n) &= p(\theta) \times p(y_1,\ldots,y_n|\theta)/p(y_1,\ldots,y_n) \\ &= \left\{\theta^{a-1}e^{-b\theta}\right\} \times \left\{\theta^{\sum y_i}e^{-n\theta}\right\} \times c(y_1,\ldots,y_n,a,b) \\ &= \left\{\theta^{a+\sum y_i-1}e^{-(b+n)\theta}\right\} \times c(y_1,\ldots,y_n,a,b) \end{aligned}$$

- We have confirmed the conjugacy

$$\left.\begin{array}{c} \theta \sim \text{gamma}(a,b) \\ Y_1,\ldots,Y_n|\theta \sim \text{Poisson}(\theta) \end{array}\right\} \Rightarrow \{\theta|Y_1,\ldots,Y_n\} \sim \text{gamma}(a+\sum_{i=1}^{n} Y_i, b+n)$$

- The posterior expectation of θ is a convex combination of the prior expectation and the sample average

$$\begin{aligned} \mathrm{E}[\theta|y_1,\ldots,y_n] &= \frac{a+\sum y_i}{b+n} \\ &= \frac{b}{b+n}\frac{a}{b} + \frac{n}{b+n}\frac{\sum y_i}{n} \end{aligned}$$

- b is interpreted as the number of prior observations

- a is interpreted as the sum of counts from b prior observations

- If $n \gg b$, then

$$\mathrm{E}[\theta|y_1,\ldots,y_n] \approx \bar{y}, \ \ \mathrm{Var}[\theta|y_1,\ldots,y_n] \approx \bar{y}/n$$

- Predictions about additional data can be obtained with the posterior predictive distribution:

$$
\begin{aligned}
p(\tilde{y}|y_1,\ldots,y_n) &= \int_0^\infty p(\tilde{y}|\theta,y_1,\ldots,y_n)p(\theta|y_1,\ldots,y_n)\,d\theta \\
&= \int p(\tilde{y}|\theta)p(\theta|y_1,\ldots,y_n)\,d\theta \\
&= \int \mathrm{dpois}(\tilde{y},\theta)\mathrm{dgamma}(\theta,a+\sum y_i,b+n)\,d\theta \\
&= \int \left\{\frac{1}{\tilde{y}!}\theta^{\tilde{y}}e^{-\theta}\right\}\left\{\frac{(b+n)^{a+\sum y_i}}{\Gamma(a+\sum y_i)}\theta^{a+\sum y_i-1}e^{-(b+n)\theta}\right\}\,d\theta \\
&= \frac{(b+n)^{a+\sum y_i}}{\Gamma(\tilde{y}+1)\Gamma(a+\sum y_i)}\int_0^\infty \theta^{a+\sum y_i+\tilde{y}-1}e^{-(b+n+1)\theta}\,d\theta
\end{aligned}
$$

- We know the gamma density

$$1 = \int_0^\infty \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \, d\theta \quad \text{for any values } a, b > 0$$

- This means that

$$\int_0^\infty \theta^{a-1} e^{-b\theta} \, d\theta = \frac{\Gamma(a)}{b^a} \quad \text{for any values } a, b > 0$$

- Now substitute in $a + \sum y_i + \tilde{y}$ of a and $b + n + 1$ of b

$$\int_0^\infty \theta^{a+\sum y_i+\tilde{y}-1} e^{-(b+n+1)\theta} \, d\theta = \frac{\Gamma(a + \sum y_i + \tilde{y})}{(b+n+1)^{a+\sum y_i+\tilde{y}}}$$

- After simplifying some of the algebra, this gives

$$p(\tilde{y}|y_1,\ldots,y_n) = \frac{\Gamma(a+\sum y_i+\tilde{y})}{\Gamma(\tilde{y}+1)\Gamma(a+\sum y_i)}\left(\frac{b+n}{b+n+1}\right)^{a+\sum y_i}\left(\frac{1}{b+n+1}\right)^{\tilde{y}}$$

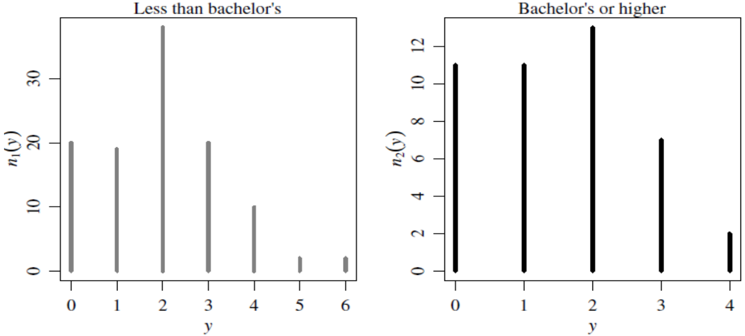
- This is a negative binomial dist, $NB(a+\sum y_i, b+n)$
- Such a random variable

$$\mathrm{E}[\tilde{Y}|y_1,\ldots,y_n] = \frac{a+\sum y_i}{b+n} = \mathrm{E}[\theta|y_1,\ldots,y_n];$$

$$\mathrm{Var}[\tilde{Y}|y_1,\ldots,y_n] = \frac{a+\sum y_i}{b+n}\frac{b+n+1}{b+n} = \mathrm{Var}[\theta|y_1,\ldots y_n]\times(b+n+1)$$

$$= \mathrm{E}[\theta|y_1,\ldots,y_n]\times\frac{b+n+1}{b+n}.$$

Example: Birth rates



- We will compare the women with college degrees to those without in terms of their numbers of children
- $Y_{1,1}, \ldots, Y_{n_1,1}$: # of children for the n_1 women without college degrees
- $Y_{1,2}, \ldots, Y_{n_2,2}$: # of children for the n_2 women with college degrees

- We will use the following sampling models:
  - $Y_{1,1}, \ldots, Y_{n_1,1}|\theta_1 \sim \text{Poi}(\theta_1)$
  - $Y_{1,2}, \ldots, Y_{n_2,2}|\theta_2 \sim \text{Poi}(\theta_2)$
- Group sums and means are as follows
  - $n_1 = 111$, $\sum_{i=1}^{n_1} Y_{i,1} = 217$, $\bar{Y}_1 = 1.95$
  - $n_2 = 44$, $\sum_{i=1}^{n_2} Y_{i,2} = 66$, $\bar{Y}_2 = 1.50$
- Prior distributions $\theta_1, \theta_2 \sim G(a = 2, b = 1)$
- Posterior distributions:

$$\theta_1|\{n_1 = 111, \sum Y_{i,1} = 217\} \sim \text{gamma}(2 + 217, 1 + 111) = \text{gamma}(219, 112)$$

$$\theta_2|\{n_2 = 44, \sum Y_{i,2} = 66\} \sim \text{gamma}(2 + 66, 1 + 44) = \text{gamma}(68, 45)$$

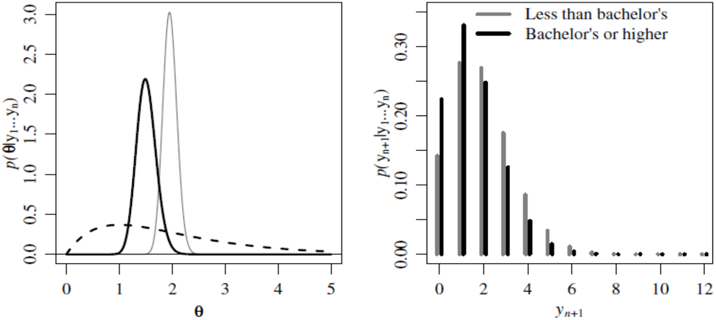
- Posterior means, modes and 95% quantile-based confidence intervals for θ_1 and θ_2 can be obtained from their gamma posterior distributions:

```
> a<-2 ; b<-1                # prior parameters
> n1<-111 ; sy1<-217         # data in group 1
> n2<-44  ; sy2<-66          # data in group 2

> (a+sy1)/(b+n1)             # posterior mean
[1] 1.955357
> (a+sy1-1)/(b+n1)           # posterior mode
[1] 1.946429
> qgamma( c(.025,.975),a+sy1,b+n1)    # posterior 95% CI
[1] 1.704943 2.222679

> (a+sy2)/(b+n2)
[1] 1.511111
> (a+sy2-1)/(b+n2)
[1] 1.488889
> qgamma( c(.025,.975),a+sy2,b+n2)
[1] 1.173437 1.890836
```

- Posterior densities for the population means of the two groups



- The posterior indicates substantial evidence $\theta_1 > \theta_2$
- $P(\theta_1 > \theta_2 | \sum_{i=1}^{n_1} Y_{i,1} = 217, \sum_{i=1}^{n_2} Y_{i,2} = 66) = 0.97$
- The posterior predictive distributions for $\tilde{Y}_1$ and $\tilde{Y}_2$ are both negative binomial distributions in the second panel