Chapter 2
One Parameter Models

A one parameter model is a class of sampling distribution that is indexed by a single unknown parameter. Here we discuss Bayesian inference for one parameter models: the binomial model and the Poisson model. In addition to being useful statistical tools, these models also provides a simple environment within which we can learn the basic of Bayesian data analysis, including conjugate prior distributions and predictive distributions.

2.1 The Poisson model

A random variable $Y$ has a Poisson distribution with mean $\theta$ if

$$p(Y = y|\theta) = dpois(y, \theta) = \exp(-\theta)\theta^y/y!$$

(2.1)

for such random variable,

$$E(y|\theta) = \theta$$

$$V(y|\theta) = \theta$$

The Poisson family of distribution has a ‘mean–variance relationship’ because if one poisson distribution has a larger mean than another, it will have a larger variance as well.

Features of Poisson distribution

If we have sampling model,

$$y_1, \ldots, y_n|\theta \sim P(\theta)$$
2.1 The Poisson model

\[ p(y_1, \ldots, y_n | \theta) = \frac{\theta^y \exp(-\theta)}{y!} \]

prior,

\[ \theta \sim \text{gamma}(a, b) \]

\[ p(\theta) = d\text{gamma}(\theta, a, b) \]

\[ = \frac{b^a}{a} \theta^{(a-1)} \exp(-b\theta) \]

then, posterior distribution of \( \theta \),

\[ \theta | y_1, \ldots, y_n \sim \text{gamma}(a + \sum_{i=1}^{n} y_i, b + n) \]

Since, posterior distribution is also gamma distribution as prior distribution, so we have confirmed the conjugacy of the gamma family for the Poisson sampling model.

**Posterior mean**

\[ E(\theta | y_1, \ldots, y_n) = \frac{a + \sum y_i}{b + n} = \frac{a}{b + n} - \frac{b}{b + n} + \frac{n}{b + n} \sum y_i \] (2.2)

where \( b \) is interpreted as the number of prior observations, and \( a \) is interpreted as the sum of counts from \( b \) prior observation. It is evident from Equation (2.2) that for large \( n \) posterior mean and posterior variances are,

\[ E(\theta | y_1, \ldots, y_n) \approx \bar{y} \] (2.3)

\[ V(\theta | y_1, \ldots, y_n) \approx \frac{\bar{y}}{n} \] (2.4)

**Posterior predictive distribution**

The predictive distribution is already introduced in Section 1.2. It is the posterior distribution of new observation \( \bar{Y} \), which is the conditional distribution of \( \bar{Y} \) given \( (Y_1 = y_1, \ldots, Y_n = y_n) \). For the Poisson model posterior predictive distribution is

\[ p(\bar{y} | y_1, \ldots, y_n) = \frac{\Gamma(a + \sum y_i + \bar{y})}{\Gamma(\bar{y} + 1) \Gamma(a + \sum y_i)} \frac{1}{(b + n + 1)^{a + \sum y_i}} \frac{\bar{y}^b}{\Gamma(b + n + 1)} \]
2.2 Example: Birth rates

This is a negative binomial distribution with parameters \((a+y_i, b+n)\), for which,

\[
E(\bar{y}|y_1, \ldots, y_n) = \frac{a + \sum y_i}{b + n}
\]

\[
= E(\theta|y_1, \ldots, y_n)
\]  \hfill (2.5)

\[
V(\bar{y}|y_1, \ldots, y_n) = \frac{a + \sum y_i b + n + 1}{b + n} \cdot \frac{b + n}{b + n}
\]

\[
= V(\theta|y_1, \ldots, y_n) \cdot b + n + 1
\]

\[
= E(\theta|y_1, \ldots, y_n) \cdot \frac{b + n + 1}{b + n}
\]  \hfill (2.6)

2.2 Example: Birth rates

Over the course of the 1990s the General Social Survey gathered data on the educational attainment and number of children of 155 women who were 40 years of age at the time of their participation in the survey. These women were in their 20s during the 1970s, a period of historically low fertility rates in the United States. In this example we will compare the women with college degrees and to those without in terms of their numbers of children. Let \(Y_{1,1}, \ldots, Y_{n_1,1}\) denote the numbers of children for the \(n_1\) women without college degrees and \(Y_{1,2}, \ldots, Y_{n_2,2}\) be the data for women with degrees. For this example we will use the following sampling models:

\[y_{1,1}, \ldots, y_{n_1,1} \mid \theta_1 \sim iid \text{ Poisson}(\theta_1)\]

\[y_{1,2}, \ldots, y_{n_2,2} \mid \theta_2 \sim iid \text{ Poisson}(\theta_2)\]

Less than bachelor’s: \(n_1 = 111, \sum_{i=1}^{n_1} Y_{1,i} = 217, \bar{Y}_1 = 1.95\)

Bachelor’s or higher: \(n_2 = 44, \sum_{i=1}^{n_2} Y_{1,i} = 66, \bar{Y}_2 = 1.50\)

In the case where \((\theta_1, \theta_2) \sim iid \text{ gamma}(a = 2, b = 1)\), we have the following posterior distributions:

\[\theta_1 \mid (n_1 = 111, \sum Y_{1,1} = 217) \sim \text{ gamma}(2 + 217, 1 + 111) = \text{ gamma}(219, 112)\]

\[\theta_1 \mid (n_2 = 44, \sum Y_{1,2} = 66) \sim \text{ gamma}(2 + 66, 1 + 44) = \text{ gamma}(68, 45)\]

2.2.1 R-Codes for computation as well as plotting

R-codes required for construction of exact and simulated posterior and prior densities. Output of these codes is reported in Figure 2.1.
2.2 Example: Birth rates

```
########## codes for creating figure for posterior densities ##########
par(mar=c(3,3,1,1), mgp=c(1.75,.75,0))
par(mfrow=c(1,2))
a<-2
b<-1
s1<-217
s2<-66
n1<-111
n2<-44
xtheta<-seq(0,5,length=1000)
plot(xtheta,dgamma(xtheta,a+s1,b+n1),type="l",xlab=expression(theta),
ylab=expression(paste(italic("p"),theta,"|",y[1],"...",y[n],"")),
sep=""),lwd=3)
lines(xtheta,dgamma(xtheta,a+s2,b+n2),lty=2,lwd=3)
lines(xtheta,dgamma(xtheta,a,b),lty=3,lwd=3)
legend(2.5,3,legend=c("without degree","with degree","prior"),
lty=c(1,2,3),lwd=c(3,3,3),bty="n")
title("Exact densities")

########## commands for simulations ##########
a<-2
b<-1
s1<-217
s2<-66
n1<-111
n2<-44
theta1<-rgamma(n=1000,a+s1,b+n1)
theta2<-rgamma(n=1000,a+s2,b+n2)
theta<-rgamma(n=1000,a,b)
plot(density(theta1),type="l",xlab=expression(theta),ylab="Density",
xlim=c(0,5),main="",lwd=3)
lines(density(theta2),lty=2,lwd=3)
lines(density(theta),lty=3,lwd=3)
legend(2.5,3,legend=c("without degree","with degree","prior"),
lty=1:3,lwd=c(3,3,3))
title("Simulated densities")
```
2.3 Binomial model

A random variable $Y$ is said to follow binomial $(n, \theta)$ distribution if,

$$p(Y = y | \theta) = \text{dbinom}(y, n, \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

for binomial $(n, \theta)$ random variables,

$$E(Y|\theta) = n\theta$$  \hspace{1cm} (2.7)  
$$V(Y|\theta) = n\theta(1 - \theta)$$  \hspace{1cm} (2.8)

Posterior inference under a uniform prior distribution

If we have sampling model,

$$y_1, \ldots, y_n | \theta \sim B(n, \theta)$$

$$\binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

Figure 2.1: Posterior and prior distributions of mean birth rates. In the left panel of this figure exact densities are plotted whereas in the right panel same densities are plotted by simulating 1000 observations from each of density. The similarity of the figures in the two panels shows that simulation tool is a very powerful tool to analyze numeric as well as graphic features of any posterior density.
prior,

\[ p(\theta) = 1 \]

then the resulting posterior distribution of \( \theta \) is,

\[ \theta | y \sim \text{beta}(y + 1, n - y + 1) \]  

(2.9)

**Posterior distribution under beta prior distributions**

We have sampling model,

\[ Y \sim B(n, \theta) \]

\[ Y \sim \binom{n}{y} \theta^y (1 - \theta)^{n-y} \]

\[ \theta \sim \text{beta}(1, 1) \]

then,

\[ \theta | Y = y \sim \text{beta}(1 + y, 1 + n - y) \]

So, we conclude that a beta prior distribution and a binomial sampling model leads to a beta posterior distribution. To reflect this, we say that the class of beta priors is conjugate for the binomial sampling model.

### 2.4 Example: 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 let \( Y_i = 0 \) otherwise. If we lack information distinguishing these \( n = 129 \) individuals we may treat their responses as being exchangeable.

**Data and posterior distribution**

- 129 individuals surveyed
- 118 individual report being generally happy
- 11 individual do not report being generally happy

Therefore, our sampling model

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

prior

\[ p(\theta) = 1 \]
2.4 Example: Happiness data

Then, the posterior distribution is

\[ p(\theta | y_1, \ldots, y_{129}) = \theta^{a+y}(1 - \theta)^{b+n-y} / \binom{a+b}{y} \]

This density for \( \theta \) is called a beta distribution with parameter \( a = 119 \) and \( b = 12 \). Now, if \( Y \sim \text{binomial}(n, \theta) \)

\( \theta \sim \text{beta}(1, 1) \) (uniform)

then, \( \theta | Y \sim \text{beta}(1 + y, 1 + n - y) \)

For any \( a, b \ i.e \), prior, \( \theta \sim \text{beta}(a, b) \)

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

2.4.1 R-codes

R-code for the construction of prior and posterior distribution for binomial sampling model, which can be shown in Figure 2.2

```r
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0),oma=c(0,0,.5,0))
par(mfrow=c(2,2))
theta<-seq(0,1,length=100)
a<-1; b<-1
n<-5 ; y<-1
plot(theta,dbeta(theta,a+y,b+n-y),type="l",ylab=
expression(paste(italic("p(\theta)",theta,"|y)",sep="")), xlab=expression(theta; lwd=3)
mtext(expression(paste("beta(1,1) prior, ", italic("n"),"=5 ", italic(sum(y[i]))","=1",sep="")))),
lines(theta,dbeta(theta,a,b),type="l",col="gray",lwd=3)
legend(.45,2.4,legend=c("prior","posterior"),lwd=c(3,3),
col=c("gray","black"), bty="n")
a<-1 ; b<-1
n<-100; y<-20
plot(theta,dbeta(theta,a+y,b+n-y),type="l",ylab=
expression(paste(italic("p(\theta)",theta,"|y)",sep=""))),
```

```r
depth(119,12)
```
Figure 2.2: Beta posterior distributions under two different sample sizes and under two prior distributions.

```r
xlab=expression(theta),lwd=3)
mtext(expression(paste("beta(1,1) prior, \"", italic("n"),"=100 \", italic(sum(y[i])),"=20",sep="")),)
lines(theta,dbeta(theta,a,b),type="l",col="gray",lwd=3)
a<-3; b<-2
n<-5 ; y<-1
plot(theta,dbeta(theta,a+y,b+n-y),type="l",ylab=expression(paste(italic("p(\"), theta,"|y")",sep="")), xlab=expression(theta),lwd=3)
mtext(expression(paste("beta(3,2) prior, \"", italic("n"),"=5 \", italic(sum(y[i])),"=1",sep="")),)
lines(theta,dbeta(theta,a+b),type="l",col="gray",lwd=3)
```
Figure 2.3: Prior and posterior distributions of binomial sampling model by simulating 1000 samples using Monte Carlo Simulation. Figure 2.3 shows that there is a little difference between the exact and simulated densities. Prior and posterior inference from simulation become more close to exact density by increasing the number of samples drawn from the posterior density.

#########Commands for simulation##############

```r
par(mar=c(3,3,1,1),mgp=c(1.75,.75,.0),oma=c(0,0,.5,0))
par(mfrow=c(2,2))
a<-1; b<-1; n<-5; y<-1
thetal<-rbeta(10000,a+y,b+n-y)
theta<-rbeta(10000,a,b)
plot(density(thetal),type="l",xlab=expression(theta),ylab="Density",
    lwd=3,main="")
lines(density(theta),type="l",col="gray",lwd=3)
a<-1; b<-1; n<-100; y<-20
thetal<-rbeta(10000,a+y,b+n-y)
theta<-rbeta(10000,a,b)
plot(density(thetal),type="l",xlab=expression(theta),ylab="Density",
    lwd=3,main="")
lines(density(theta),type="l",col="gray",lwd=3)
a<-3; b<-2; n<-5; y<-1
thetal<-rbeta(10000,a+y,b+n-y)
theta<-rbeta(10000,a,b)
plot(density(thetal),type="l",xlab=expression(theta),ylab="Density",
    lwd=3,main="")
lines(density(theta),type="l",col="gray",lwd=3)
```

2.4 Example: Happiness data
2.5 Discussion and further references

The notation of conjugacy for classes of prior distributions was developed in Raiffa and Schlaifer (1961). Most authors refer to intervals of high posterior probability as credible intervals as opposed to confidence intervals. Doing so fails to recognize that Bayesian intervals do have frequentist coverage probabilities, often being very close to the specified Bayesian coverage level (Welch and Peers, 1963; Hartigan, 1966; Severini, 1991). Some authors suggest that accurate frequentist coverage can be a guide for the construction of prior distributions (Tibshirani, 1989; Sweeting, 1999, 2001). Also Kass and Wasserman (1996) for a review of formal methods for selecting prior distributions.