Chapter 5

Poisson-Gamma model

5.1 Introduction

In this lesson we study the Poisson-Gamma combination of likelihood and prior.

If we have a random sample $y_1, y_2, ..., y_n$ where each y_i is believed to be generated by a Poisson distribution with parameter θ, and if we consider as prior distribution for θ a Gamma distribution with parameters a, b, then the posterior distribution of θ is also a Gamma with parameters $(a + \sum y_i, \quad b + n)$. The predictive distribution for a future observation is Negative Binomial with parameters That is, $a + \sum y_i, b + n$.

$$\begin{aligned} \theta &\sim \text{Gamma}(a, b) & (5.1)\\ y_1, y_2, ..., y_n \mid \theta &\sim \text{Poisson}(\theta) & (5.2)\\ \theta \mid y_1, y_2, ..., y_n &\sim \text{Gamma}\left(a + \sum y_i, \quad b + n\right) & (5.3)\\ y_{n+1} \mid y_1, y_2, ..., y_n &\sim \text{NegBinomial}\left(a + \sum y_i, b + n\right) & (5.4) \end{aligned}$$

In this lesson, we want to get acquainted with all these distributions, see what we can obtain from them with R and, along the way, construct the posterior and predictive distributions with R for one example. You will have to do three exercises on your own, using what describe below.

Please, read section 3.2 of Hoff's book and these notes as required reading.

5.2 The Poisson random variable Poisson(θ)

.

A discrete random variable Y is said to be a Poisson r.v. with parameter θ if the probability distribution of a single observation y is

$$P(Y = y \mid \theta) = \begin{cases} \frac{e^{-\theta}\theta^y}{y!} & \text{for } y = 0, 1, 2, ..., n \\ 0 & \text{elsewhere} \end{cases} \quad (5.5)$$

For such a random variable,

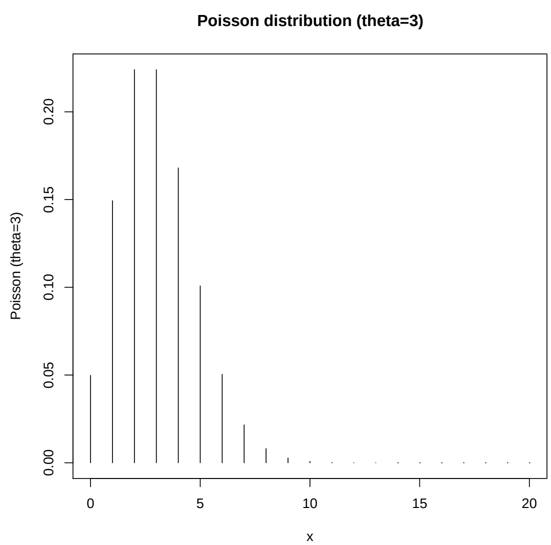
$$E[Y|\theta] = \theta \qquad Var[Y|\theta] = \theta$$

Poisson r.v.'s are, for example, counts in areas, volumes or time. For example, we may model the number of flaws in a square yard of textile, the number of bacterial colonies in a cubic centimeter of water, or the number of times a machine fails in the course of a workday.

R will not give you the mean and expected value directly. There is not a function for that in R (unless you use simulated data from the Poisson, in which case you would use the mean() function.)

Let's see how we can summarize things of the Poisson distribution with R. Suppose we have a Poisson with $\theta = 3$.

```
######### Things R can give you from a Poisson(theta=3)#######

##############  Obtaining the probabilities for each X ###########
  x=seq(0,20,by=1)   # create x=0,1,2,....
   x        # view x
   [1]  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20

 p.x= dpois(x, 3 )     # Find Prob(X=x) for x=0,1,2,......
 p.x    # view p.x
 [1] 4.978707e-02 1.493612e-01 2.240418e-01 2.240418e-01 1.680314e-01
 [6] 1.008188e-01 5.040941e-02 2.160403e-02 8.101512e-03 2.700504e-03
[11] 8.101512e-04 2.209503e-04 5.523758e-05 1.274713e-05 2.731529e-06
[16] 5.463057e-07 1.024323e-07 1.807629e-08 3.012715e-09 4.756919e-10
[21] 7.135379e-11

 ########### Plotting the Poisson(theta=3) ##########

 plot(x,p.x,type="h",ylab="Poisson (theta=3)",main="Poisson distribution (theta=3)")
```



{fig:poisson3} Figure 5.1: Poisson distribution with mean 3

```
####### find probabilities for specific ranges of X ###############

 p.x.0=dpois(0,3)   # just Prob(X=0)
 [1] 0.04978707
```

```
 ppois(4,3)   #   Prob(X <= 4), at most 4
 [1] 0.8152632
  1-ppois(3,3)  # Prob(X>= 4), at least 4
[1] 0.3527681

######## Quantiles ####################

qpois(0.5, 3)   # Find the median
[1] 3
qpois(c(0.025, 0.975),3) # find x's for a 95 percent interval
[1] 0 7
 qpois(0.9,3)   # Find the x that is a 90th percentile.
 [1] 5
```

5.2.1 Functions of a Poisson Random Variable

Quite often, we are interested not in the random variable per se, but in a function of the random variable. In that case, it is better to use simulation of random numbers from the Poisson distribution and work with the function of these random numbers. For example, suppose we are interested in $Y = a + bX$, where X is Poisson (lambda=3). Then we draw random numbers from Poisson of lambda =3 and then apply the function to them and summarize the distribution of the latter empirically.

```
########  Draw random numbers from a Poisson and ###########
#####  work with functions of those random variables ###########

random.numbers=rpois(1000,3) # draw 1000 random numbers

function.rn = 3*random.numbers+10  # we are interested in this function

########### Empirical summaries of the empirical distribution of the #####
######## function.rn numbers #########################

hist(function.rn)   # gives the empirical distribution of the function



mean(function.rn)   # gives the empirical mean
[1] 19.198
var(function.rn)  # gives the variance of the numbers given by function.rn
[1] 28.03283
sd(function.rn) # gives the standard deviation
[1] 5.294604
quantile(function.rn,0.9)    # find the x that is the 90th percentile
90\%
 25
quantile(function.rn,c(0.025,0.975))
2.5\%  97.5\%
   10    31
median(function.rn)    # find the median
```

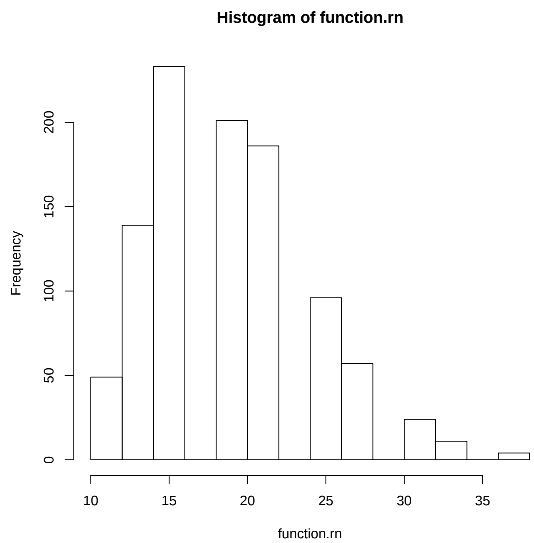


Figure 5.2: histogram of y=3*Poisson(3)+10

{fig:functionp

```
[1] 19
n=length(function.rn)      # sample size
n
[1] 1000

sum(function.rn <30)/n  # empirical probability that the function is less than 30
[1] 0.961
```

5.2.2 Exercise to do on your own

For a certain manufacturing industry, the number of industrial accidents averages 5 per week. Use R to answer the following questions.

(a) What is the mean and variance of the number of industrial accidents?

If the average is 5, the variance is also 5

(b) Find the probability that 2 accidents will occur in a given week.

If X is the number of accidents per week, $P(X = 2)$ is given by

```
dpois(2,5)
[1] 0.08422434
```

(c) Find the probability that there are at most 6 accidents in the given week.

$P(X \leq 6)$ is given by

```
ppois(6,5)
[1] 0.7621835
```

(d) Find the probability that there are at least 3 accidents in a given week.

$P(X \geq 3) = 1 - Prob(X \leq 2)$

```
1-ppois(2,5)
[1] 0.875348
```

(e) Each accident costs the industry 3000 dollars. Find the distribution of the cost function and give its mean and variance.

Let $Y = 3000X$ be the cost function. We generate random numbers from the Poisson(5) and then apply the function to it and do the histogram.

```
 x=rpois(1000,5)
 y=3000*x
 mean(y)
[1] 14934
 var(y)
[1] 49941586
 hist(y,main="histogram of Y=3000X")
```

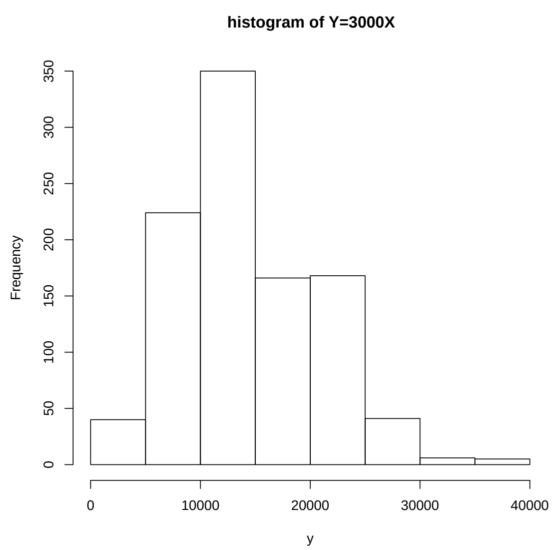


Figure 5.3: histogram of y=3000*Poisson(5) {fig:functionp

5.3 Likelihood function of a sample of data when each observation is Poisson

For a random sample $Y = (Y_1, Y_2, ..., Y_n)$ of iid observations from a Poisson distribution, the joint probability distribution of the data or likelihood function is

$$p(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n|\theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} \quad (5.6)$$

$$= \prod_{i=1}^{n} \frac{1}{y_i!} \theta^{\sum_{i=1}^{n} y_i} e^{-n\theta} \quad (5.7)$$

$$\propto \theta^{\sum_{i=1}^{n} y_i} e^{-n\theta} \quad (5.8)$$

$$(5.9)$$

where $\sum_{i=1}^{n} y_i$ is a sufficient statistic (i.e., it tells us all the information available in the data). The likelihood depends on θ once you have observed the data.

The likelihood is a Poisson distribution with parameter $n\theta$. This is proved in a Probability class. We will not prove it here.

To plot the likelihood, you need to have data. The likelihood is a function of θ. It tells us the value of θ that is most likely to have generated the data. Suppose that each y_i is considered to be Poisson(θ) and we have a random sample of $n = 10$ observations, with the following values: $1, 2, 2, 2, 4, 6, 4, 1, 4, 5$ The $\sum_{i=1}^{n} y_i = 31$ The likelihood function is then

$$p(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n \mid \theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} \quad (5.10)$$

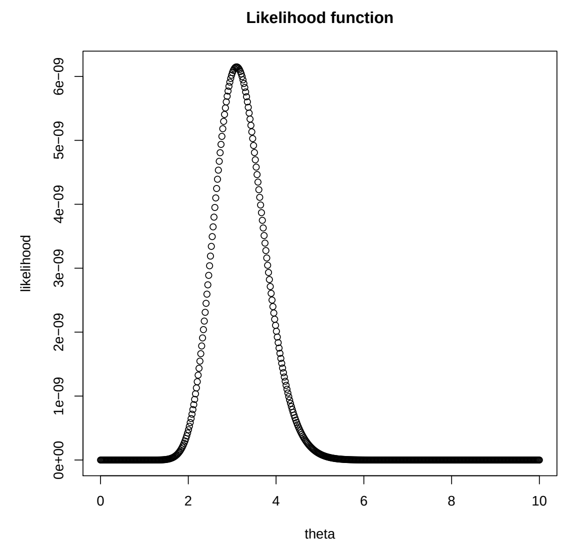
$$= \prod_{i=1}^{n} \frac{1}{y_i} \theta^{31} e^{-10\theta} \quad (5.11)$$

$$\propto \theta^{31} e^{-10\theta} \quad (5.12)$$

$$(5.13)$$

We can plot the likelihood function using R. We will need to do the following:

```
data=c(1, 2, 2, 2, 4, 6, 4, 1, 4, 5)  # enter only when n is small.

  prod.data=1/ prod(factorial(data))   # gives 1/ product of y_i !
 sum.data=sum(data)    # gives the sum y_i

 theta=seq(0,10,length=500)  #generate thetas
 likelihood=rep(0,length(theta))   # create space to put the likelihood
 for(i in 1:length(theta) )     # loop to compute the likelihood for each theta
 {
  likelihood[i]=prod.data*(theta[i]^sum.data)*exp(-length(data)*theta[i])
  likelihood      # return the likelihood when running this function
  }

 plot(theta, likelihood, main="Likelihood function")  # plot the likelihood function



theta[likelihood==max(likelihood)]  # which theta maximizes the likelihood
[1] 3.106212
```



ihood.poisson}

Figure 5.4: Likelihood function when y Poisson

Be careful with this program. Always clear the workspace after your run it.

Look at the graph you obtain. It shows that the theta that maximizes the likelihood is very close to 3. We saw that it is $\theta = 3.106$.

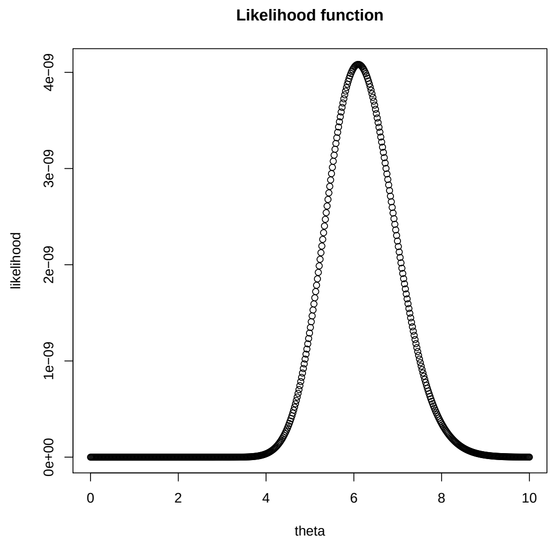
You will do the following example.

5.3.1 Exercise to do on your own

The industry mentioned earlier is trying to figure out the average number of accidents per week. The industry, in reality, does not know what that average is. The industry only believes that the number of accidents per week is Poisson with unknown θ. To find out, the industry collects the following random sample: 8, 6, 5, 5, 8, 7, 5, 6, 5, 6

Plot the likelihood function and find the value of θ that maximizes the likelihood. Use R commands given above.

```
data=c(8, 6, 5, 5, 8, 7, 5, 6, 5, 6)  # enter only when n is small.

  prod.data=1/ prod(factorial(data))   # gives 1/ product of y_i !
  prod.data
[1] 1.576906e-30

 sum.data=sum(data)    # gives the sum y_i
 [1] 61

 theta=seq(0,10,length=500)  #generate thetas
 likelihood=rep(0,length(theta))   # create space to put the likelihood
 for(i in 1:length(theta) )     # loop to compute the likelihood for each theta
 {
  likelihood[i]=prod.data*(theta[i]^sum.data)*exp(-length(data)*theta[i])
  likelihood      # return the likelihood when running this function
  }

 plot(theta, likelihood, main="Likelihood function")  # plot the likelihood function
```



od.poisson-1}

Figure 5.5: Another Likelihood function when y Poisson

```
theta[likelihood==max(likelihood)]  # which theta maximizes the likelihood
[1] 6.092184
```

It appears that the value of θ that maximizes the likelihood is 6.092.

5.4 Conjugate Prior for the Poisson Sampling Model

We will work first with a class of conjugate prior distributions that will make posterior calculations simple. One of the prior distributions used for the θ is the Gamma density with parameters a and b or

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

for $\theta, a, b > 0$. For such a distribution, $E[\theta] = \frac{a}{b}$; $Var[\theta] = \frac{a}{b^2}$; $mode[\theta] = \begin{cases} (a-1)/b & if a > 1 \\ 0 & \text{if } a \leq 1 \end{cases}$.

Comparing $P(y \mid \theta)$ with $p(\theta)$ reveals that the prior density is, in some sense, equivalent to a total count of $a - 1$ in b prior observations.

The Gamma is a continuous density function. The following R functions give us the density, cumulative probabilities, quantiles and random numbers for θ:

```
 dgamma(theta, a, b)   # density value for a theta
  pgamma(theta_1, a, b)  #  Prob(theta < theta_1)
  qgamma(0.95, a,b)  # 95th percentile of theta
  rgamma(1000,a,b)  # draw 1000 random numbers from the
```

5.4.1 Exercise to do on your own

In the industry example that we were looking at earlier, the industry believes that the average number of accidents is around 2 per year. Based on past evidence, they consider that a Gamma(4,2) is a good approximation to their beliefs based on information available.

Summarize this prior distribution with R. I.e., find the prior mean, prior variance, prior median, 95% prior probability interval. Draw also 1000 random numbers from this distribution and do a histogram of it. Plot the prior distribution.

```
theta=seq(0,10,length=500)
plot(theta,dgamma(theta,4,2), type='l',main="Gamma(4,2) prior")
```

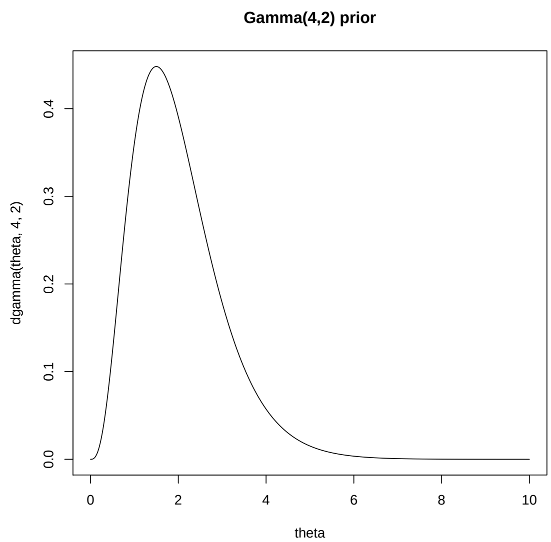


Figure 5.6: Gamma prior distribution {fig:gamma4and

```
 qgamma(0.5,4,2)   # Find the prior median
[1] 1.836030         # this is the prior median
 qgamma(c(0.025,0.975),4,2)  # Find a prior 95\% interval
[1] 0.5449327 4.3836365    $ this is the interval
 mean.my.gamma=4/2   # prior mean
 var.my.gamma=4/2^2   # prior variance
 mean.my.gamma      # view prior mean
[1] 2                      # this is the prior mean
 var.my.gamma        # view prior variance
[1] 1                     # this is the prior variance
```

5.5 Posterior distribution

With this conjugate prior distribution, the posterior distribution is a Gamma density function

$$p(\theta \mid y) \sim G(a + \sum y_i, b + n)$$

Thus the posterior expectation of θ given the data is

$$E[\theta \mid y] = \frac{a + \sum y_i}{b + n}$$

and the variance is

$$Var[\theta \mid y] = \frac{a + \sum y_i}{(b + n)^2}$$

There are several rearrangements of this posterior expectations that help give more meaning to it. See Hoff's Section 3.2.

5.5.1 Exercise to do on your own

Given the prior and likelihood of the last two examples, use R to summarize the posterior distribution of θ. Give posterior mean of θ , posterior standard deviation, posterior 95% probability interval. Compare them with the same quantities from the prior distrubution. Plot the posterior distribution, and in the same graph, plot the likelihood and the prior distribution. Use the

```
lines
```

command twice after the plot command with different

```
lty
```

options.

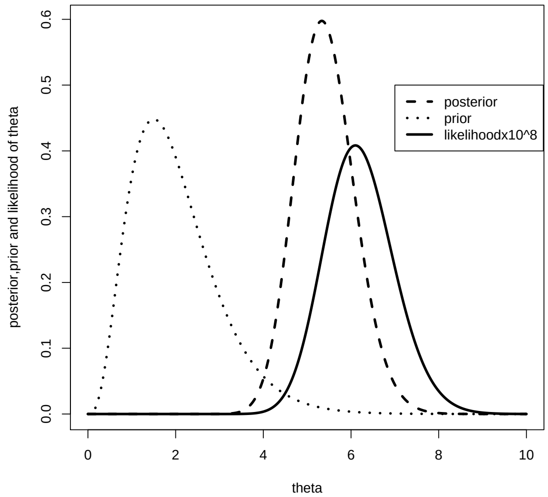
You may use code used in Outline 3, section 3.

```
posterior=dgamma(theta,(sum.data+4),(10+2))  # posterior is Gamma(61+4, 12) =gamma(65,12)
 pdf('priorpostpoisson.pdf')
 plot(theta,posterior,type="l",ylab="posterior,prior and likelihood of theta",lty=2,lwd=3)
 lines(theta,prior,lty=3,lwd=3)
 lines(theta,likelihood*10^8,lty=1,lwd=3)
 legend(7,0.5,c("posterior","prior","likelihoodx10^8"), lty=c(2,3,1),lwd=c(3,3,3))
 dev.off()
```

To find the posterior mean, variance and standard deviation, we type

```
 post.mean=65/12   # compute posterior mean
 > post.mean       # view posterior mean
[1] 5.416667       # this is the posterior mean

 post.var=65/12^2   #compute posterior variance
 post.sd=sqrt(post.var)  #compute posterior standard deviation
 post.sd    #view posterior standard deviation
[1] 0.6718548     # this is the posterior standard deviation.

 qgamma(c(0.025,0.975),65,12)   # find the posterior 95\% interval
[1] 4.180469 6.810548  # These are the lower and upper bound of the interval
```

	Mean	Standard Deviation	95% interval
Prior	2	1	(0.5449, 4.3836)
Posterior	5.416	0.6718	(4.1804, 6.8105)



rpostpoisson}

Figure 5.7: Prior, likelihood and posterior distributions for the rate of accidents per week (theta)

5.6 Predictive distributions

To make inferences about an unknown observable, often called predictive inferences, we follow a similar logic. Before the data $y_1, y_2, ..., y_n$ are considered, the distribution of the unknown but observable data is

$$p(y_1, y_2, ..., y_n) = \int p(y_1, y_2, ..., y_n, \theta) d\theta = \int p(\theta) p(y_1, y_2, ..., y_n \mid \theta)$$

This is often called the marginal distribution of the vector of data $y = (y_1, y_2, ..., y_n)$. But a more informative name is the *prior predictive distribution*: prior because it is not conditional on a previous observation of the process, and predictive because it is the distribution for a quantity that is observable.

After the vector of data y has been observed, we can predict an unknown observable, $\widetilde{y}$ from the same process. For example, $y = (y_1, y_2, ..., y_n)$ may be the vector of recorded number of accidents per week observed over 10 weeks and $\widetilde{y}$ is the yet to be recorded count of accidents in a future week. The distribution of $\widetilde{y}$ is called the *posterior predictive distribution*, posterior because it is conditional on the observed vector y and predictive because it is a prediction for an observable $\widetilde{y}$.

$$\begin{aligned}
p(\widetilde{y} \mid y) &= \int p(\widetilde{y}, \theta \mid y) d\theta & (5.14)\\
&= \int p(\widetilde{y} \mid \theta, y) p(\theta \mid y) d\theta & (5.15)\\
&= \int p(\widetilde{y} \mid \theta) p(y) d\theta & (5.16)
\end{aligned}$$

The second and third lines display the posterior predictive distribution as an average of conditional predictions over the posterior distribution of θ. The last equation follows because y and $\widetilde{y}$ are conditionally independent given θ in this model.

5.6.1 Prior prediction in the Gamma-Poisson model

With conjugate families, like the Gamma-Poisson, the known form of the prior and posterior densities can be used to find the marginal distribution, $p(y)$, using the formula

$$p(y) = \frac{p(y \mid \theta)p(\theta)}{p(\theta \mid y)}$$

For instance, the Poisson model for a single observation, y, has prior predictive distribution

$$\begin{aligned} p(y) &= \frac{\text{Poisson}(y \mid \theta)\,\text{Gamma}(\theta \mid a, b)}{\text{Gamma}(\theta \mid a+y, 1+b)} \\ &= \frac{\text{Gamma}(a+y)b^a}{\text{Gamma}(a) \quad (y!) \quad (1+b)^{a+y}} \end{aligned}$$

which reduces to

$$p(y) = \begin{pmatrix} a+y-1 \\ y \end{pmatrix} \left(\frac{b}{b+1}\right)^a \left(\frac{1}{b+1}\right)^y .$$

This is known as the negative binomial distribution with parameters (a, b).

The above derivation shows that the negative binomial distribution is a mixture of Poisson distributions with rates, θ, that follow the gamma distribution:

$$\text{NegBinomial}(y|a, b) = \int \text{Poisson}(y \mid \theta)\,\text{Gamma}(\theta \mid a, b)d\theta$$

The negative binomial distribution is considered a robust alternative to the Poisson distribution. It has thicker tails, so it is considered a better candidate than the Poisson when there are rare events.

5.6.2 Posterior prediction in the Gamma-Poisson model

We usually have more than one observation in our sample. Predictions about additional data can be obtained with the posterior predictive distribution:

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

for $\widetilde{y} \in 0, 1, 2.......$ This is a negative binomial distribution with parameters $(a + \sum y_i, b + n)$, for which

$$E[\widetilde{y} \mid y_1, ..., y_n] = \frac{a + \sum y_i}{b+n} = E[\theta \mid y_1, ..., y_n]$$

;

$$\begin{aligned} Var[\widetilde{y} \mid y_1, ..., y_n] &= \left(\frac{a + \sum y_i}{b+n}\right)\left(\frac{b+n+1}{b+n}\right) && (5.17) \\ &= Var[\theta \mid y_1, ..., y_n](b+n+1) && (5.18) \\ &= E[\theta \mid y_1, ..., y_n]\frac{b+n+1}{b+n}. && (5.19) \end{aligned}$$

Let's try to obtain a deeper understanding of this formula for the predictive variance. The predictive variance is to some extent a measure of our posterior uncertainty about a new sample $\tilde{Y}$ from the population. Uncertainty about $\tilde{Y}$ stems from uncertainty about the population and the variability in sampling from the population. For large n, uncertainty about θ is small ($\frac{(b+n+1}{b+n} \approx 1$) and uncertainty about $\widetilde{Y}$ stems primarily from sampling variability, which for the Poisson model is equal to θ. For small n, uncertainty in $\widetilde{Y}$ also includes the uncertainty in θ, and so the total uncertainty is larger than just the sampling variability ($\frac{b+n+1}{b+n} > 1$).

In R, there is a function for the negative binomial. To do the usual things, you have the commands:

```
rnbinom(n, mu,  size )   #  generate n random numbers
drbinom(x,mu,size)    # density of x
prbinom(x0, mu, size)  # prob(X <=x0) since it is continuous the = is irrelevant
q(quantile, mu, size)  # obtain quantiles
```

$mu = \frac{a+\sum y_i}{b+n}$, $size = a + \sum y_i$

You plug in the size and the mu,making sure you type their names.

The following result might be useful later: in general, the variance of the neg binomial is $mu + mu^2/size$, so if you know the variance, then $size = mu^2/(variance - mu)$.

5.6.3 Exercise to do on your own

Given the findings for the posterior distribution, the industry we are working on in this lesson is interested in predicting the number of accidents next week. Use the predictive distribution

```
y=seq(0,20,by=1)
predictive=dnbinom(y,mu=(65/12),size=65)
 plot(y,predictive,type="l",ylab="Posterior predictive distribution")

 qnbinom(c(0.025,0.975),mu=(65/12),size=65)
[1]  1 11
```

There is 95% probability that the number of accidents is between 1, and 11 next week.

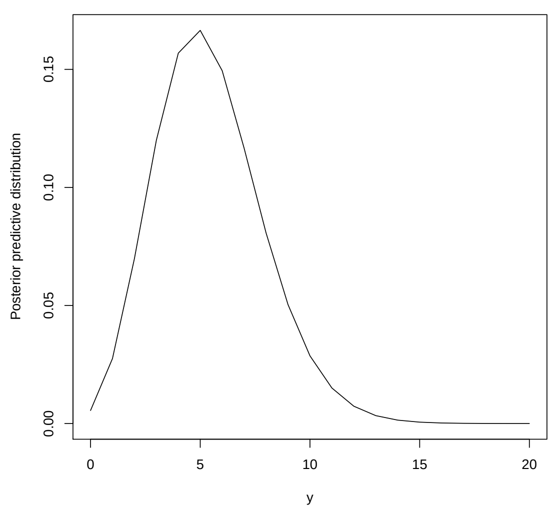


Figure 5.8: Predicted number of accidents next week {fig:predictiv

```
pred.mean=65/12
 pred.mean
```

```
[1] 5.416667     # this is the expected  number of accidents next week


 pred.var=pred.mean*13/12
 pred.var
 5.868056   # this is the variance of the predicted number of accidents next week

 pred.sd=sqrt(pred.var)
 pred.sd
[1] 2.422407    # this is the standard deviation of the predicted accidents next week
```

5.7 Poisson model parameterized in terms of rate and exposure

In many applications, it is convenient to extend the Poisson model for data points $y_1, y_2, ..., y_n$ to the form

$$y_i \sim \text{Poisson}(x_i\theta),$$

where the values x_i are known positive values of an explanatory variable, x, and θ is the unknown parameter of interest. In epidemiology, the parameter θ is often called the *rate*, and x_i is called the *exposure* of the *ith* unit. This model is not exchangeable in the $y_i'x$ but is exchangeable in the pairs $(x, y)_i$. The likelihood for θ in the extended Poisson model is

$$p(y|\theta) \propto \theta^{\sum_{i=1}^n y_i} e^{\theta \sum_{i=1}^n x_i}$$

(ignoring factors that do not depend on θ, and so the gamma distribution for θ is conjugate. With prior distribution

$$\theta \sim \text{Gamma}(a, b),$$

the resulting posterior distribution is

$$\theta|y \sim \text{Gamma}\left(a + \sum_{i=1}^n y_i, b + \sum_{i=1}^n x_i\right)$$

5.7.1 Estimating a rate from Poisson data: an idealized example

Suppose the causes of death are reviewed in detail for a city in the United States for a single year. It is found that 3 persons, out of a population of 200,000, died of asthma, giving an estimated asthma mortality rate in the city of 1.5 cases per 100,000 persons per year. A Poisson sampling model is often used for epidemiological data of this form. The Poisson model derives from an assumption of exchangeability among all small intervals of exposure. Under the Poisson model, the distribution of y, the number of deaths in a city of 200,000 in one year, may be expressed as Poisson(2θ), where θ represents the true underlying long-term asthma mortality rate in our city (measured in cases per 100,000 persons per year). In the above notation, $y = 3$ is a single observation with exposure $x = 2.0$ (since θ is defined in units of 100,000 people) and unknown rate θ. We can use knowledge about asthma mortality rates around the world to construct a prior distribution for θ and then combine the datum $y = 3$ with that prior distribution to obtain a posterior distribution.

What is a sensible prior distribution for θ? Review of asthma mortality rates around the world suggest that mortality rates above 1.5 per 100,000 people are rare in Western countries, with typical asthma mortality rates around 0.6 per 100,000. Trial-and-error exploration of the properties of the gamma distribution, the conjugate prior family for this problem, reveals that a Gamma(3.0, 5.0) density provides a plausible prior density for the asthma mortality rate in this example if we assume exchangeability between this city and other cities and this year and other years. The mean of this prior density is 0.6 (with a mode of 0.4) and 97.5% of the mass of the density lies below 1.44. In practice, specifying a prior mean sets the ratio of the two gamma parameters and then changing the shape parameter by trial and error allows us to match the prior knowledge about the tail of the distribution.

With the prior distribution and data described, the posterior distribution for θ is Gamma(6.0,7.0), which has mean 0.86 –substantial shrinkage has occurred toward the prior distribution. The posterior probability that the long-term death rate from asthma in our city is more than 1.0 per 100,000 per year, computed from the gamma posterior density, is 0.3.

5.8 Example

The table below gives the number of fatal accidents and deaths on scheduled airline flights per year over a ten-year period.

(a) Assume that the number of fatal accidents in each year are independent with a Poisson(θ) distribution. Set a prior distribution for θ and determine the posterior distribution based on the data from 1976 through 1985. Under this model, give a 95% predictive interval for the number of fatal accidents in 1986.

(b) Assume that the numbers of fatal accidents in each year follow independent Poisson distributions with a constant rate and an exposure in each year proportional to the number of passenger miles flown. Set a prior distribution for θ and determine the posterior distribution based on the data for 1976-1985. (Estimate the number of passenger miles flown in each year by dividing the appropriate columns of the table and ignoring round-off errors.) Give a 95% predictive interval for the number of fatal accidents in 1986 under the assumption that 8000 million passenger miles are flown that year.

(c) Repeat (a) above, replacing "fatal accidents" with "passenger deaths." (homework question)

(d) Repeat (b) above, replacing "fatal accidents" with "passenger deaths." (Homework questions)

(e) In which of the cases (a)-(d) above does the Poisson model seem more or less reasonable? Why? Discuss based on general principles, without specific reference to the numbers in TAble

Year	Fatal accidents	Passenger deaths	Death rate
1976	24	734	0.19
1977	25	516	0.12
1978	31	754	0.15
1979	31	877	0.16
1980	22	814	0.14
1981	21	362	0.06
1982	26	764	0.33
1983	20	809	0.13
1984	16	223	0.03
1985	22	1066	0.15

I need to assess the prior Gamma distribution first. Since it is much easier to think of observables than of parameters of distributions, and since the prior predictive distribution is negative binomial with the same parameters as the prior, I am going to use the prior predictive distribution's mean and variance to solve for the parameters, based on my beliefs about those two quantities.

The conjugate prior for a Poisson likelihood, $Y \sim Poisson(\theta)$, is a gamma prior, $\theta \sim$ Gamma(a, b). Choose a, b to match your prior beliefs about the mean and variance on fatal airline accidents worldwide per year. For example, I believe there might be around 50 give or take 40 fatal airline accidents per year, i.e, $E(Y) = 50$, $Var(Y) = 40^2$. Note that the marginal distribution of Y is negative binomial $NegBinomial(a, b)$. Matching mean and variance of a NegBinomial with 50 and 40^2, respectively, I find $a = 1.6, \quad , b = 0.03$. Alternatively, you could have guessed an a-priori mean and variance of θ and matched with moments of the gamma. It's probably easier though to think about an observable (Y), rather than an unobservable parameter (θ).

(a) The $\sum y_i = 238$ and $n = 10$. The posterior distribution is $P(\theta \mid y) \sim \text{Gamma}(1.6{+}238, 0.03{+}10) = \text{Gamma}(239.6, 10.03)$. The predictive distribution $p(y_{n+1} \mid y) = \int p(y_{n+1} \mid \theta)p(\theta \mid y)d\theta = \int \text{Poisson}(y_{n+1} \mid \theta)\,\text{Gamma}(\theta \mid a, b))d\theta$. This is a NegBinomial(239.6, 10.03).

(b) Let X equal passenger miles. In the table, X=pass deaths/death rate (in 100 mio miles). $p(\theta \mid y) = Ga(a_2, b_2)$, $a_2 = a + \sum y_i = 239.6$, $b_2 = b + \sum x_i = 57158.72$. To find the posterior predictive $p(y_{n+1} \mid y)$ use:

$$
\begin{aligned}
p(y_{n+1} \mid y) &= \int p(y_{n+1} \mid \theta)p(\theta \mid y)d\theta & (5.20)\\
&= \int Poi(y_{n+1} \mid x_{n+1}\theta)Ga(\theta \mid a_2, b_2)d\theta & (5.21)\\
&= \int Poi(y_{n+1} \mid \eta)Ga(\eta \mid a_2, b_2/x_{n+1} & (5.22)\\
&= Nbin(y_{n+1} \mid a_2, b_2/x_{n+1}) & (5.23)\\
&= NBin(y_{n+1} \mid 239.6, 7.144) & (5.24)
\end{aligned}
$$

The third equation comes from a change of variables from θ to $\eta = x_{n+1}\theta$. Note that if $\theta \sim Ga(a, b)$, then $\eta = r\theta \sim Ga(a, b/r)$. The fourth equation is true by theoretical results on the Poisson. The desired predictive interval is the central 95% interval for a $NBin(a_2, b_2/x_{n+1})$. We can find that it is (22,23,...., 43).

5.9 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 to those without in terms of their numbers of children. Let $Y_{1,1}, Y_{2,1}, ..., Y_{n_1,1}$ denote the number of children for the n_1 women without college degrees and $Y_{1,2}, Y_{2,2}, ..., Y_{n_2,2}$ be the data for women with degrees. For this example, we will use the following sampling models:

$$
\begin{aligned}
Y_{1,1}, Y_{2,1}, ..., Y_{n_1,1} &\sim \text{i. i. d. Poisson}(\theta_1) & (5.25)\\
Y_{1,2}, Y_{2,2}, ..., Y_{n_2,2} &\sim \text{i. i. d. Poisson}(\theta_2) & (5.26)
\end{aligned}
$$

Emprirical distributions for the data can be seen in Hoff's. Group sums and means are as follows:
Less than bachelor's: $n_1 = 111, \sum_{i=1}^{n_1} Y_{i,1} = 217, \bar{Y}_1 = 1.95$
Bachelor's or higher: $n_2 = 44, \sum_{i=1}^{n_2} Y_{i,2} = 66, \bar{Y}_2 = 1.5$
In the case where $\{\theta_1, \theta_2\} \sim \text{Gamma}(a = 2, b = 1)$, we have the following posterior distributions:

$$
\begin{aligned}
\theta_1 \mid n_1 = 111, \sum Y_{i,1} = 217 &\sim \text{Gamma}(2 + 217, 1 + 111) = \text{Gamma}(219, 112) & (5.27)\\
\theta_2 \mid n_2 = 44, \sum Y_{i,2} = 66 &\sim \text{Gamma}(2 + 66, 1 + 44) = \text{Gamma}(68, 45) & (5.28)
\end{aligned}
$$

Posterior means, models 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 ; s1<-217    # data in group 1
n2<-44  ; s2<-66     # data in group 2
```

```
(a+s1)/(b+n1)        # posterior mean
(a+s1-1)/(b+n1)      # posterior mode
qgamma( c(.025,.975),a+s1,b+n1)   # posterior 95\% CI

(a+s2)/(b+n2)
(a+s2-1)/(b+n2)
qgamma( c(.025,.975),a+s2,b+n2)

-------
th1_mc<-rgamma(100000,a+s1,b+n1)

th2_mc<-rgamma(100000,a+s2,b+n2)

mean(th1_mc>th2_mc)

y1_mc<-rpois(1000000,th1_mc)
y2_mc<-rpois(1000000,th2_mc)
mean(y1_mc>y2_mc)
mean(y1_mc>=y2_mc)
mean(y1_mc==y2_mc)
```

Posterior densities for the population means of the two groups are shown in ...... The posterior indicates substantial evidence that $\theta_1 > \theta_2$. For example (I guess using the fact that a sum of gammas is gammas, or simply subtracting the generated gamma random numbers, $Pr(\theta_1 > \theta_2 \mid \sum Y_{i,1} = 217, \sum Y_{i,2} = 66) = 0.97$.

Now consider two randomly sampled individuals, one from each of the two populations. To what extent do we expect the one without the bachelor's degree to have more children than the other? We can calculate the relevant probabilities exactly: The posterior predictive distributions for $\tilde{Y}_1$ and $\tilde{Y}_2$ are both negative binomial distributions and are both plotted in...

```
y =0:10
dnbinom(y, size=(a+sy1), mu=(a+sy1)/(b+n1))
[1] 1.427473e  ?01 2.766518e ?01 2.693071e ?01 1.755660e?01
[5] 8.622930e  ?02 3.403387e ?02 1.124423e ?02 3.198421e?03
[9] 7.996053e  ?04 1.784763e ?04 3.601115e?05
dnbinom(y, size=(a+sy2), mu=(a+sy2)/(b+n2))
[1] 2.243460e?01 3.316420e?01 2.487315e?01 1.261681e?01
[5] 4.868444e?02 1.524035e?02 4.030961e?03 9.263700e?04
[9] 1.887982e?04 3.465861e?05 5.801551e?06
```

Notice that there is much more overlap between these two distributions than between the posterior distributions of θ_1 and θ_2. For example, $Pr(\tilde{Y}_1 > \tilde{Y}_2 \mid \sum Y_{i,1} = 217, \sum Y_{i,2}) = 0.22$. The distinction between the events $\{\theta_1 > \theta_2\}$ and $\tilde{Y}_1 > \tilde{Y}_2$ is extremely important: Strong evidence of a difference between two populations does not mean that the difference itself is large.