COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

STAT 4224/5224

Bayesian Statistics

Dobrin Marchev

Poisson Distribution

Poisson distribution is used to model counts, like number of friends, number of customers at a service center in each hour, and so on. That is, the sample space is countably infinite:

$$\mathcal{X} = \{0, 1, 2, \dots \}$$

Recall that X is said to have a Poisson distribution with parameter $\theta > 0$ if

$$p(x;\theta) = \frac{\theta^x e^{-\theta}}{x!}, x = 0, 1, 2, \dots$$

It can be shown that when $X \sim \text{Poisson}(\theta)$ we have

$$\text{E}(X \mid \theta) = \text{Var}(X \mid \theta) = \theta$$

Poisson Likelihood

Suppose that $X_1, \ldots, X_n \mid \theta \sim \text{Poisson}(\theta)$

Then the likelihood is

$$p(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{x_i!} \theta^{x_i} e^{-\theta}$$

$$= \frac{1}{\prod_{i=1}^{n} x_i!} \theta^{\sum_{i=1}^{n} x_i} e^{-n\theta} = c(x_1, \ldots, x_n) \theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}$$

The likelihood (and posterior) depend only on $\sum_{i=1}^{n} X_i$, making it a sufficient statistic (like in the Binomial model).

It can be shown that

$$\sum_{i=1}^{n} X_i \,|\theta \sim \text{Poisson}(n\theta)$$

Conjugate Prior

The posterior distribution for the Poisson model is

$$f(\theta|x_1, \dots, x_n) = \frac{p(x_1, \dots, x_n|\theta)\pi(\theta)}{p(x_1, \dots, x_n)}$$

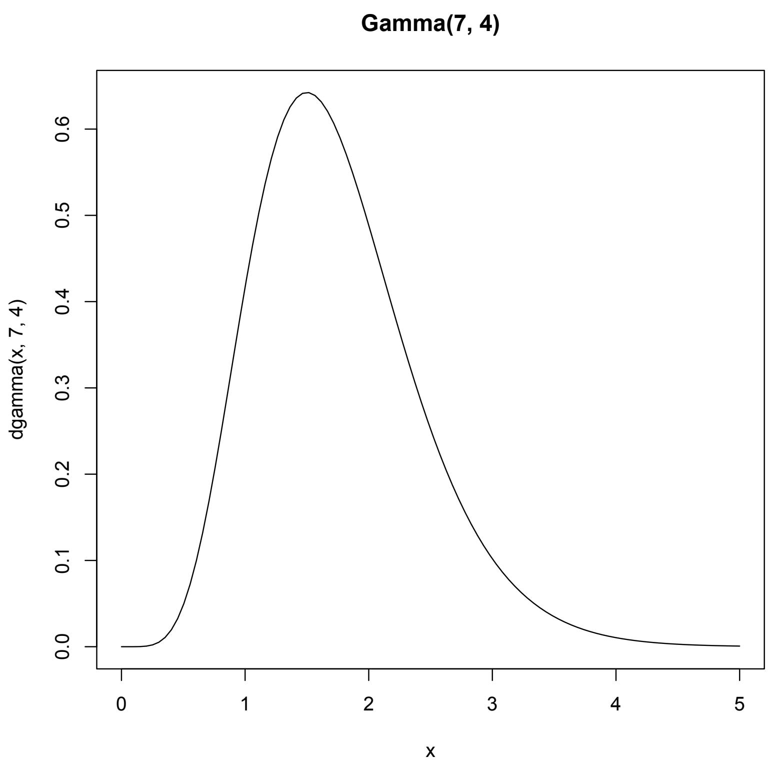
$$\propto p(x_1, \dots, x_n|\theta)\pi(\theta) \propto \pi(\theta)\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}$$

Therefore, in order that the posterior is in the same family as the prior, we need $\pi(\theta)$ to include terms like $\theta^{c_1} e^{-c_2\theta}$ for some numbers c_1 and c_2. The simplest such density is the Gamma distribution:

$$\theta \sim \Gamma(a, b)$$

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}, \theta > 0$$

where a, $b > 0$ are the hyperparameters.



Aside: Gamma Distribution Properties

- Let $X \sim \Gamma(a, b)$
- $E(X) = \frac{a}{b}$
- $Var(X) = \frac{a}{b^2}$
- $mode(X) = \begin{cases} \frac{a-1}{b}, & a > 1 \\ 0, 0 < a \leq 1 \end{cases}$

It is available in R with the `dgamma` function.

Posterior Distribution

Suppose that

$$X_1, \ldots, X_n \mid \theta \sim \text{Poisson}(\theta)$$
$$\theta \sim \Gamma(a, b)$$

Then the posterior distribution is

$$f(\theta|x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n|\theta)\pi(\theta)}{p(x_1, \ldots, x_n)}$$
$$\propto \left(\theta^{a-1}e^{-b\theta}\right)\theta^{\sum_{i=1}^{n} x_i}e^{-n\theta} \propto \theta^{a+\sum_{i=1}^{n} x_i-1}e^{-(b+n)\theta}$$

Obviously, this corresponds to a Gamma distribution. Therefore,

$$\theta|X_1, \ldots X_n \sim \Gamma\left(a + \sum_{i=1}^{n} X_i\,, b + n\right)$$

Posterior Mean

Estimation and prediction proceed in a manner like that in the binomial model. For example,

$$E(\theta|x_1, \dots x_n) = \frac{a + \sum_{i=1}^{n} x_i}{b+n} = \frac{b}{b+n} \cdot \frac{a}{b} + \frac{n}{b+n} \cdot \frac{\sum_{i=1}^{n} x_i}{n}$$

which is again a weighted average of the prior mean and sample mean. If $n >> b$, then

$$E(\theta|x_1, \dots x_n) \approx \bar{x}$$

Exercise: Derive the posterior variance and argue that if $n >> b$, we have $\mathrm{Var}(\theta|x_1, \dots x_n) \approx \frac{\bar{x}}{n}$

Prediction

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

$$f(x_{new}|x_1, \dots, x_n) = \int_0^\infty f(x_{new}, \theta|x_1, \dots, x_n)d\theta$$

$$= \int_0^\infty p(x_{new}|\theta, x_1, \dots, x_n)f(\theta|x_1, \dots, x_n)d\theta$$

$$= \int_0^\infty p(x_{new}|\theta)f(\theta|x_1, \dots, x_n)d\theta$$

$$= \int_0^\infty \left(\frac{1}{x_{new}!}\theta^{x_{new}}e^{-\theta}\right)\left[\frac{(b+n)^{a+\sum_{i=1}^n x_i}}{\Gamma\left(a+\sum_{i=1}^n x_i\right)}\theta^{a+\sum_{i=1}^n x_i-1}e^{-(b+n)\theta}\right]d\theta$$

Prediction (Continued)

Posterior predictive distribution is:

$$f(x_{new}|x_1, \ldots, x_n)$$

$$= \int_0^\infty \left(\frac{1}{x_{new}!}\theta^{x_{new}} e^{-\theta}\right)\left[\frac{(b+n)^{a+\sum_{i=1}^n x_i}}{\Gamma\left(a+\sum_{i=1}^n x_i\right)}\theta^{a+\sum_{i=1}^n x_i - 1} e^{-(b+n)\theta}\right] d\theta$$

$$= \frac{(b+n)^{a+\sum_{i=1}^n x_i}}{\Gamma(x_{new}+1)\Gamma\left(a+\sum_{i=1}^n x_i\right)}\int_0^\infty \theta^{a+\sum_{i=1}^n x_i + x_{new} - 1} e^{-(b+n+1)\theta} d\theta$$

From the properties of the Gamma distribution:

$$\int_0^\infty \theta^{a+\sum_{i=1}^n x_i + x_{new} - 1} e^{-(b+n+1)\theta} d\theta = \frac{\Gamma(a+\sum_{i=1}^n x_i + x_{new})}{(b+n+1)^{a+\sum_{i=1}^n x_i + x_{new}}}$$

That is,

$$f(x_{new}|x_1, \ldots, x_n)$$

$$= \frac{\Gamma(a+\sum_{i=1}^n x_i + x_{new})}{\Gamma(x_{new}+1)\Gamma\left(a+\sum_{i=1}^n x_i\right)}\left(\frac{b+n}{b+n+1}\right)^{a+\sum_{i=1}^n x_i}\left(\frac{1}{b+n+1}\right)^{x_{new}}$$

Prediction (Continued)

Posterior predictive distribution is:

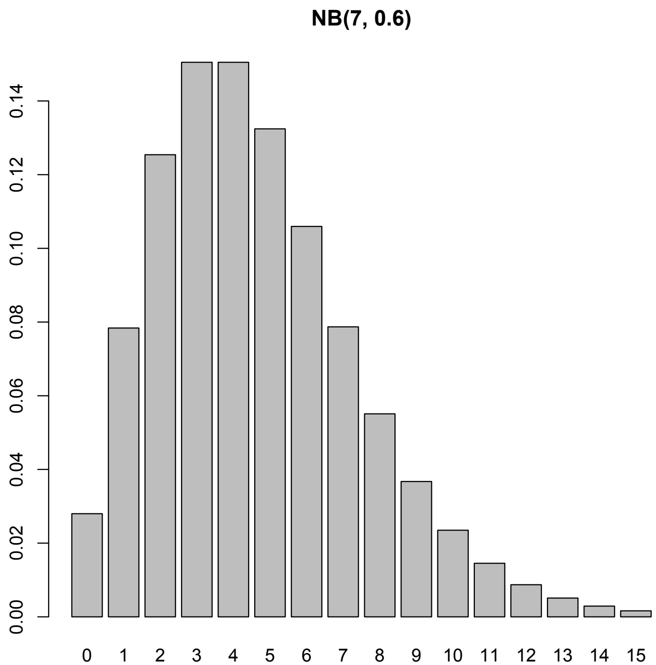
$$f(x_{new}|x_1,\dots,x_n)$$
$$=\frac{\Gamma(a+\sum_{i=1}^n x_i + x_{new})}{\Gamma(x_{new}+1)\Gamma\left(a+\sum_{i=1}^n x_i\right)}\left(\frac{b+n}{b+n+1}\right)^{a+\sum_{i=1}^n x_i}\left(\frac{1}{b+n+1}\right)^{x_{new}}$$

for $x_{new}=0,1,2,\dots$

Note that this is the negative binomial distribution with parameters $a+\sum_{i=1}^n x_i$ and $\frac{b+n}{b+n+1}$. This means that:

$$E(x_{new}|x_1,\dots,x_n)=\left(a+\sum_{i=1}^n x_i\right)\frac{\frac{1}{b+n+1}}{\frac{b+n}{b+n+1}}=\frac{a+\sum_{i=1}^n x_i}{b+n}$$
$$=E(\theta|x_1,\dots,x_n)$$

$$Var(x_{new}|x_1,\dots,x_n)=\frac{a+\sum_{i=1}^n x_i}{b+n}\cdot\frac{b+n+1}{n}$$



Aside: Negative Binomial Distribution Properties

- Let $X \sim \mathrm{NB}(r, p)$
- Meaning: number of failures until r successes
- $p(x) = \binom{x+r-1}{x}(1-p)^x p^r, x = 0,1,2,\ldots$
- $E(X) = r\frac{1-p}{p}$
- $Var(X) = r\frac{1-p}{p^2}$
- It is available in R with the `dnbinom` function.

Exercise 1

An oil company conducts a geological study that indicates that an exploratory oil well should have a 20% chance of striking oil. What is the probability that the third strike comes on the seventh well drilled?

Answer: 0.049

Example 1: Birth rates (p. 48)

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. We have the model:

$$X_{1,1}, \ldots, X_{n_1,1} | \theta_1 \sim Poisson(\theta_1)$$
$$X_{1,2}, \ldots, X_{n_2,2} | \theta_2 \sim Poisson(\theta_2)$$

where Group 1 is women without college degree, and Group 2 is women with degrees. Here are the sample summaries:

$$n_1 = 111, \sum_{i=1}^{n_1} x_{i,1} = 217, \bar{x}_1 = 1.95$$
$$n_2 = 44, \sum_{i=1}^{n_2} x_{i,2} = 66, \bar{x}_2 = 1.5$$

Example 1 Solution

Consider independent priors

$$\theta_1, \theta_2 \sim \Gamma(2, 1)$$

Notice this gives average of 2 births per woman, which is kind of historically correct before the 1970s.

Then the posterior distributions are:

$$\theta_1 | x_{1,1}, \dots, x_{n_{1,1}} \sim \Gamma(219, 112)$$
$$\theta_2 | x_{1,2}, \dots, x_{n_{2,2}} \sim \Gamma(68, 45)$$

Show R code.

Exponential Family

- The exponential family is a set of distributions that possess a certain common form of their pdf's.

- It includes many of the important distributions studied separately, like normal, Bernoulli, …

- It is widely used in more complex statistical models (like GLM) and can accommodate additional data types, like counts and intervals.

Exponential Family

Definition: A *one-parameter exponential family* model is any model whose densities can be expressed as

$$f(x|\phi) = h(x)c(\phi)e^{q(\phi)t(x)}$$

where ϕ is the parameter and $t(x)$ is sufficient statistic.

To verify if some family of distributions is of exponential type, we must be able to identify the functions c(), h(), t() and q().

Example 2: Bernoulli

Recall the Bernoulli pmf:

$$f(x|p) = p^x(1-p)^{1-x} = (1-p)e^{x\log\frac{p}{1-p}}$$

so, in this case

$$c(p) = (1-p), h(x) = 1, t(x) = x, q(p) = \log\frac{p}{1-p}$$

Exercise: Verify the Poisson density is also of exponential type and identify the functions in it.

Multivariate Definition

Let $X = (X_1, \dots, X_n)$ be a random vector with joint pdf

$$f(x;\theta) = c(\theta)h(\boldsymbol{x})e^{\sum_{j=1}^{k} t_j(\boldsymbol{x})q_j(\theta)}$$

where θ is k-parameter vector.

Such distribution is said to be in a k-parameter exponential family.

Example 3

Let $X_1, \ldots, X_n$ be a random sample from N(μ, σ). Then

$$f(\boldsymbol{x};\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

$$= (2\pi)^{-\frac{n}{2}} \sigma^{-n} e^{n\mu^2} e^{-\frac{1}{2\sigma^2}\left(\underbrace{\sum_{i=1}^{n} x_i^2}_{t_2(\boldsymbol{x})} - 2\mu \underbrace{\sum_{i=1}^{n} x_i}_{t_1(\boldsymbol{x})}\right)}$$

Exercise 3

Let $X_1, \ldots, X_n$ be a random sample from $\Gamma(a, b)$.
Write the likelihood in such a form to show that it belongs to the two-parameter exponential family.

Exponential Family and Conjugate Priors

Consider the one-parameter exponential family distribution

$$f(x|\phi) = h(x)c(\phi)e^{\phi t(x)}$$

with prior distribution:

$$\pi(\phi|n_0, t_0) = \kappa(n_0, t_0)c(\phi)^{n_0}e^{n_0 t_0 \phi}$$

Then the posterior is:

$$f(\phi|x_1, \dots, x_n) \propto c(\phi)^{n_0+n}e^{\phi[n_0 t_0 + \sum_{i=1}^{n} t(x_i)]}$$
$$\propto \pi(\phi|n_0 + n, n_0 t_0 + n\bar{t}(x))$$

where $\bar{t}(x) = \frac{\sum_{i=1}^{n} t(x_i)}{n}$

n_0 can be interpreted as prior “sample size”

t_0 can be interpreted as “prior guess” of $t(x)$

It can be shown that $E[t(X)] = t_0$

Example 4

Consider again the Binomial model

$$p(x|\theta) = \theta^x(1-\theta)^{1-x} = \left(\frac{\theta}{1-\theta}\right)^x (1-\theta)$$

$$= e^{\phi x}\left(1+e^{\phi}\right)^{-1}$$

where $\phi = \log\left(\frac{\theta}{1-\theta}\right)$ is the log-odds aka natural parameter.

Conjugate prior is:

$$\pi(\phi|n_0,t_0) \propto \left(1+e^{\phi}\right)^{-n_0} e^{n_0 t_0 \phi}$$

where t_0 is the prior probability that $X = 1$.

Using change of variables, it can be shown that

$$\pi(\theta|n_0,t_0) \propto \theta^{n_0 t_0 - 1}(1-\theta)^{n_0(1-t_0)-1}$$

A weakly informative prior distribution can be obtained by setting t_0 equal to our prior expectation and $n_0 = 1$.

Exercise 4

Redo Example 4 for the Poisson model.

Practice

Suppose $X_1, \ldots X_n \sim Exp(\theta)$.

a) Find conjugate prior for θ
b) Find the corresponding posterior distribution
c) Show that the reciprocal of posterior mean can be written as a weighted average of the reciprocal of prior mean and the MLE of θ.

Practice Solution

We have that the likelihood is

$$f(x_1, \ldots, x_n|\theta) = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$$

Thus, the conjugate prior is Gamma(a, b)

$$\pi(\theta) \propto \theta^{a-1} e^{-b\theta}$$

which gives us posterior distribution:

$$f(\theta|x_1, \ldots, x_n) \propto \theta^{n+a-1} e^{-\theta\left(\sum_{i=1}^{n} x_i + b\right)}$$

That is,

$$\theta|x_1, \ldots, x_n \sim \Gamma\left(n + a, \sum_{i=1}^{n} x_i + b\right)$$

Practice Solution

The posterior mean of

$$\theta|x_1, \dots, x_n \sim \Gamma\left(n + a, \sum_{i=1}^{n} x_i + b\right)$$

is

$$E(\theta|x_1, \dots, x_n) = \frac{n + a}{\sum_{i=1}^{n} x_i + b}$$

$$\hat{\theta}_{post}^{-1} = \frac{n}{n + a}\frac{\sum_{i=1}^{n} x_i}{n} + \frac{a}{n + a}\frac{b}{a} = \frac{n}{n + a}\hat{\theta}_{MLE}^{-1} + \frac{a}{n + a}\hat{\theta}_{prior}^{-1}$$

Exercise 5

Problem 3.9 on p. 230

3.9 Galenshore distribution: An unknown quantity Y has a Galenshore(a, θ) distribution if its density is given by

$$p(y) = \frac{2}{\Gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2}$$

for $y > 0$, $\theta > 0$ and $a > 0$. Assume for now that a is known. For this density,

$$\mathrm{E}[Y] = \frac{\Gamma(a+1/2)}{\theta \Gamma(a)}, \quad \mathrm{E}[Y^2] = \frac{a}{\theta^2}.$$

a) Identify a class of conjugate prior densities for θ. Plot a few members of this class of densities.
b) Let $Y_1, \ldots, Y_n \sim$ i.i.d. Galenshore(a, θ). Find the posterior distribution of θ given $Y_1, \ldots, Y_n$, using a prior from your conjugate class.
c) Write down $p(\theta_a | Y_1, \ldots, Y_n)/p(\theta_b | Y_1, \ldots, Y_n)$ and simplify. Identify a sufficient statistic.
d) Determine $\mathrm{E}[\theta | y_1, \ldots, y_n]$.