Lecture 5

More prior and posteriors

Topics of the day:

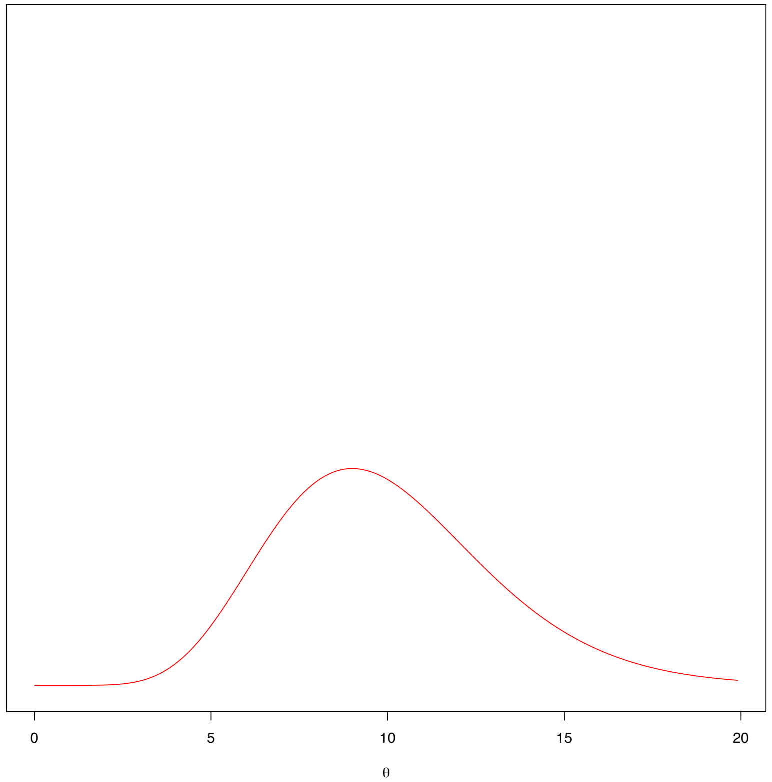
1. Further example of prior and posterior
2. Posterior given repeated independent measurements
3. Second half is exercise session

In the coming example we want to focus
prior to posterior,
and then
how todo predict from a Bayesian perspective

Example

- Suppose a hospital has around 200 beds occupied each day, and we want to know the underlying risk that a patient will be infected by MRSA (methicllin-resistant Staphylococcus aureus).
- Looking at the first six months of the year, we count y=20 infections in 40000 bed-days.
- Let θ be the risk of infection per 10000 bed-days.
- Since the infection rate given the number of bed-days are so low, a reasonable model could be that y is Poisson distributed.

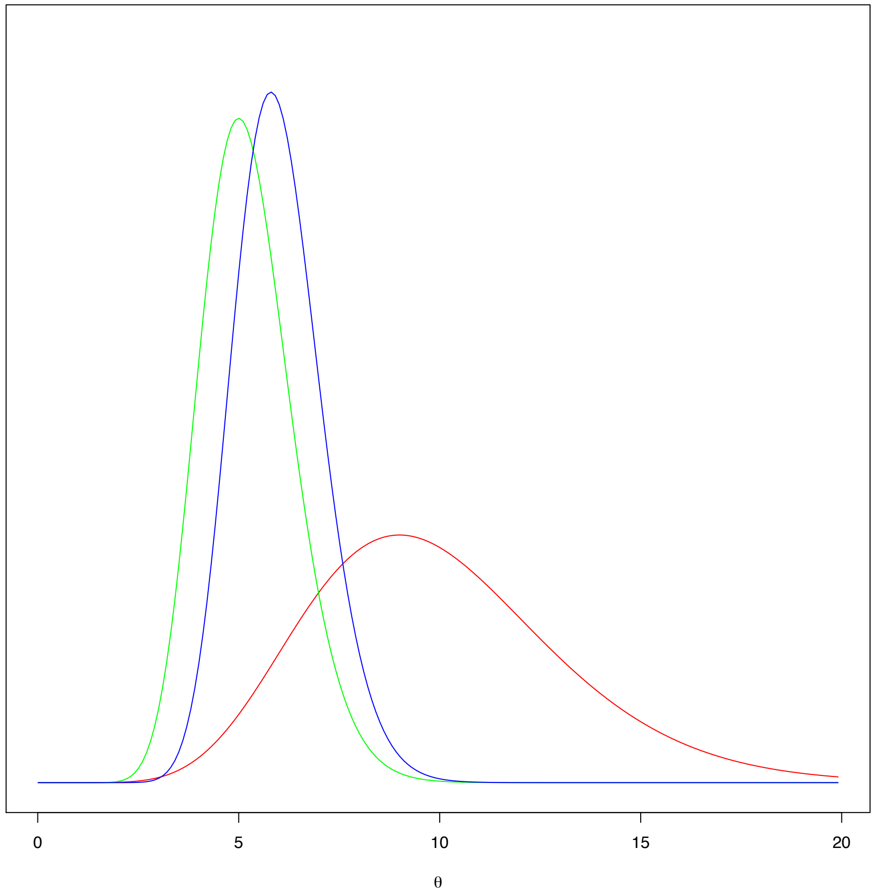
$$P(Y = k; \theta) = \frac{e^{-\theta}\theta^k}{k!}$$



Bayesian analysis

- Suppose we have previous information that around 5 to 17 persons per 10000 is usually gets the infection.
- We can express this as a prior

$$\theta \sim \Gamma(\alpha = 10, \beta = 1)$$



Posterior

Putting the prior

$$\theta \sim \Gamma(\alpha = 10, \beta = 1)$$

and the likelihood

$$P(Y = k; \theta) = \frac{e^{-\theta}\theta^k}{k!}$$

together we get the posterior

$$f(\theta; y) \propto \theta^y e^{-4\theta} \theta^{\alpha-1} e^{-\beta\theta} = \theta^{y+\alpha-1} e^{-(4+\beta)\theta}$$

$$\propto \Gamma(\theta; \alpha + y, \beta + 4) = \Gamma(\theta; 30, 5)$$

Now we want to create a 95% confidence intervall:

$$P(\theta \in [a,b]; y, \alpha, \beta) = 0.95$$

First we have:

$$P(\theta \in [a,b]; y, \alpha, \beta) = P(\theta \leq b; y, \alpha, \beta) - P(\theta \leq a; y, \alpha, \beta)$$

To find a and b we use R:

```
a = qgamma(0.025,30,5)

b = qgamma(0.975,30,5)
```

Suppose we want to know what is the probability of more then twelve MRSA cases in the next 10000 bed days?

Before we had any data

Given our prior this is

$$P(Y \geq 12; \alpha = 10, \beta = 1) = 1 - \sum_{k=0}^{12} \int_{\theta} P(Y = k; \theta)\Gamma(\theta, 10, 1) d\theta$$

After the data:

$$P(Y \geq 12; \alpha = 30, \beta = 5) = 1 - \sum_{k=0}^{12} \int_{\theta} P(Y = k; \theta)\Gamma(\theta, 30, 5) d\theta$$

How do we compute this complicated sum and integrals?

Monte Carlo method

Lets start with a integral that we can compute

$$P(\theta > 10; \alpha = 10, \beta = 1) = \int_{10}^{\infty} \Gamma(x, \alpha, \beta) dx$$

in R: `1-pgamma(10,10,1)`

But what do we want answer?

If we draw a random variable θ how often will it be over 10.

Monte Carlo method

So instead of computing the integral lets draw a lot of variables and see how often they are over 10.

In R: `thetas = rgamma(n=1000,10,1)`

In R: `mean(thetas>10)`

Monte Carlo method

Back to the more complicated example

$$P(Y = k; \alpha, \beta) = \int_\theta P(Y = k; \theta)\Gamma(\theta, \alpha, \beta) d\theta$$

This integral can be interpreted the same way.

We draw a θ from $\Gamma(\alpha, \beta)$

Given the θ we draw a Y from $Poisson(\theta)$

In R:

```
thetas = rgamma(n=1000,10,1)
Ys = rpois(1000, thetas)
mean(Ys==k)
```

Suppose we want to know what is the probability of more then twelve MRSA cases in the next 10000 bed days?

Before we had any data

$$P(Y \geq 12; \alpha = 10, \beta = 1) = 1 - \sum_{k=0}^{12} \int_{\theta} P(Y = k; \theta)\Gamma(\theta, 10, 1) d\theta$$

$$\approx 0.21$$

After the data:

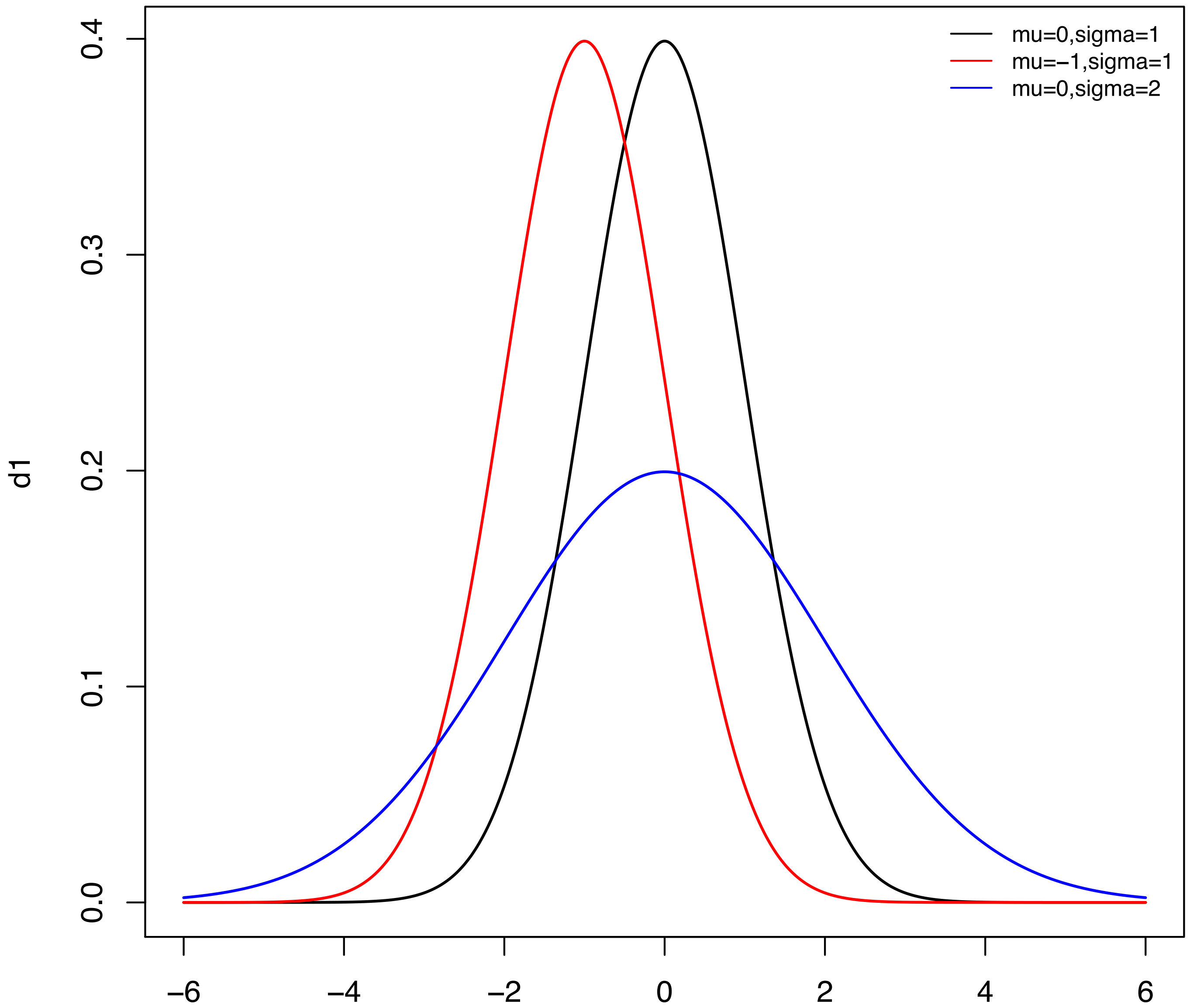
$$P(Y \geq 12; \alpha = 30, \beta = 5) = 1 - \sum_{k=0}^{12} \int_{\theta} P(Y = k; \theta)\Gamma(\theta, 30, 5) d\theta$$

$$\approx 0.01$$

In the next example we want to learn how to handle several independent variables

Normal distribution

$$f_X(a) = \frac{1}{2\pi\sigma} \exp\left(-\frac{1}{2\sigma^2}(a-\mu)^2 \right)$$



In the this example we will study the height of a group of individuals

We assume that a persons height is normal distributed:

$$Y \sim N(\mu, \sigma^2)$$

We assume that the variance is known:

$$\sigma^2 = 102$$

For the mean parameter we set a uniform prior:

$$f(\mu) \propto 1.$$

Suppose we observe n individuals
and we want the posterior distribution of μ

the likelihood:

$$f(Y_1, Y_2, \ldots, Y_n; \mu, \sigma^2) = \prod_{i=1}^{n} f(Y_i; \mu, \sigma^2)$$

$$\propto e^{-\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2\sigma^2}}$$

$$\propto e^{-\frac{(\mu - \frac{1}{n}\sum_{i=1}^{n} y_i)^2}{2n^{-1}\sigma^2}}$$

the likelihood time prior:

$$\begin{aligned} f(\mu; Y_1, Y_2, \ldots, Y_n, \sigma^2) &\propto f(Y_1, Y_2, \ldots, Y_n; \mu, \sigma^2) \times 1 \\ &\propto e^{-\frac{(\mu - \frac{1}{n}\sum_{i=1}^{n} y_i)^2}{2n^{-1}\sigma^2}} \\ &\propto N(\bar{y}, \frac{\sigma^2}{n}) \end{aligned}$$