Simulation Schemes Justification
One Parameter Models

Statistics 220

Spring 2005

Simulation - Joint and Marginal Distributions

Joint Distribution:

Want to simulate X, Y from $p(x, y)$

- Sample x_i from $p(x); i = 1, \ldots, m$
- Sample y_i from $p(y|x_i); i = 1, \ldots, m$

Justification that this scheme actually draws from the joint distribution:

The joint empirical CDF of $(x_i, y_i); i = 1, \ldots, m$ is

$$\hat{P}(x, y) = \frac{1}{m} \sum_{i=1}^{m} I(x_i \leq x, y_i \leq y) = \frac{1}{m} \sum_{i=1}^{m} I(x_i \leq x) I(y_i \leq y)$$

The expected value of the ECDF is

$$E[\hat{P}(x,y)] = P[X \leq x, Y \leq y] = P(x,y)$$

since

$$\begin{aligned} E[I(x_i \leq x)I(y_i \leq y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x_i \leq x)I(y_i \leq y)p(x_i)p(y_i|x_i)dy_i dx_i \\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} p(x_i, y_i)dy_i dx_i \\ &= P[X \leq x, Y \leq y] \end{aligned}$$

The ECDF is an unbiased estimate of the CDF.

In addition

$$\mathrm{Var}(\hat{P}(x,y)) = \frac{P(x,y)(1-P(x,y))}{m} \longrightarrow 0$$

as $m \to \infty$, which implies $\hat{P}(x,y) \longrightarrow P(x,y)$ in probability (WLLN).

Marginal Distribution:

Want to simulate Y based on $p(x,y)$

- Sample x_i from $p(x); i = 1, \ldots, m$
- Sample y_i from $p(y|x_i); i = 1, \ldots, m$
- Keep only $y_i; i = 1, \ldots, m$

Justification that this scheme actually draws from the marginal distribution $p(y)$:

The empirical CDF of $y_i; i = 1, \ldots, m$ is

$$\hat{P}(y) = \frac{1}{m} \sum_{i=1}^{m} I(y_i \leq y)$$

The expected value of the ECDF is

$$E[\hat{P}(y)] = P[Y \leq y] = P(y)$$

since

$$\begin{aligned} E[I(y_i \le y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(y_i \le y) p(x_i) p(y_i|x_i) dy_i dx_i \\ &= \int_{-\infty}^{y}\int_{-\infty}^{\infty} p(x_i, y_i) dx_i dy_i \\ &= \int_{-\infty}^{y}\int_{-\infty}^{\infty} p(y_i) dy_i \\ &= P[Y \le y] \end{aligned}$$

Similarly to before

$$\mathrm{Var}(\hat{P}(y)) = \frac{P(y)(1 - P(y))}{m} \longrightarrow 0$$

as $m \to \infty$, which implies $\hat{P}(y) \longrightarrow P(y)$ in probability (WLLN).

Other One Parameter Models

1. Poisson

Example: Prussian Cavalry Fatailities Due to Horse Kicks

10 Prussian cavalry corp were monitored for 20 years (200 Corp-Years) and the number of fatalities due to horse kicks were recorded

$x = \#$ Deaths	Number of Corp-Years with x Fatalities
0	109
1	65
2	22
3	3
4	1

Let $y_i, i = 1, \ldots, 200$ be the number of deaths in observation i.

Assume that $y_i|\theta \overset{iid}{\sim} Poisson(\theta)$. (This has been shown to be a good description for this data). Then the MLE for θ is

$$\hat{\theta} = \bar{y} = \frac{122}{200} = 0.61$$

This can be seen from

$$p(y|\theta) = \prod_{i=1}^{200} \frac{1}{y_i!} \theta^{y_i} e^{-\theta} \propto \theta^{\sum y_i} e^{-n\theta} = \theta^{n\bar{y}} e^{-n\theta}$$

Instead lets take a Bayesian approach. For a prior, lets use $\theta \sim Gamma(\alpha, \beta)$

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

Note that this is a conjugate prior for θ.

The posterior density satisfies

$$p(\theta|y) \propto \theta^{n\bar{y}} e^{-n\theta} \theta^{\alpha-1} e^{-\beta\theta} = \theta^{n\bar{y}+\alpha-1} e^{-(n+\beta)\theta}$$

which is proportional to a $Gamma(\alpha + n\bar{y}, \beta + n)$ density

The mean and variance of a $Gamma(\alpha, \beta)$ are

$$E[\theta] = \frac{\alpha}{\beta} \qquad \mathrm{Var}(\theta) = \frac{\alpha}{\beta^2}$$

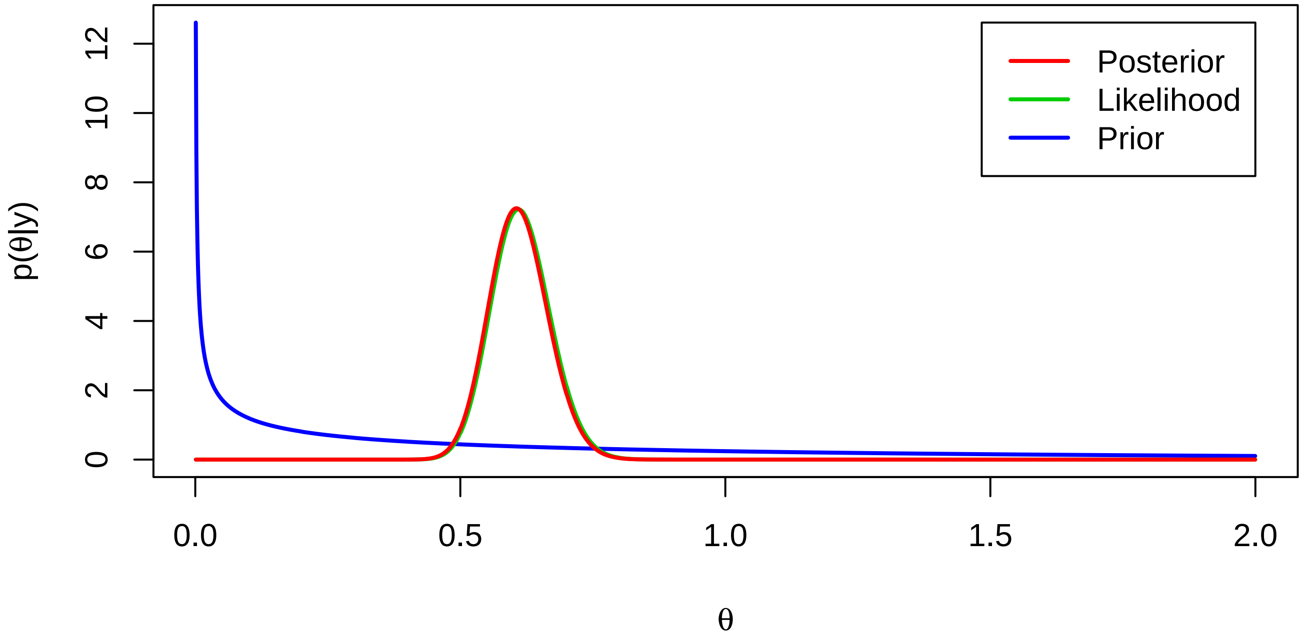
So the posterior mean and variance in this analysis are

$$E[\theta|y] = \frac{\alpha + n\bar{y}}{\beta + n} \qquad \mathrm{Var}(\theta|y) = \frac{\alpha + n\bar{y}}{(\beta + n)^2}$$

Similarly to before, the posterior mean is a weighted average of the prior mean and the MLE (weights β and n).

Lets examine the posteriors under different prior choices

$$n = 200, \bar{y} = 0.61, \alpha = \beta = 0.5$$

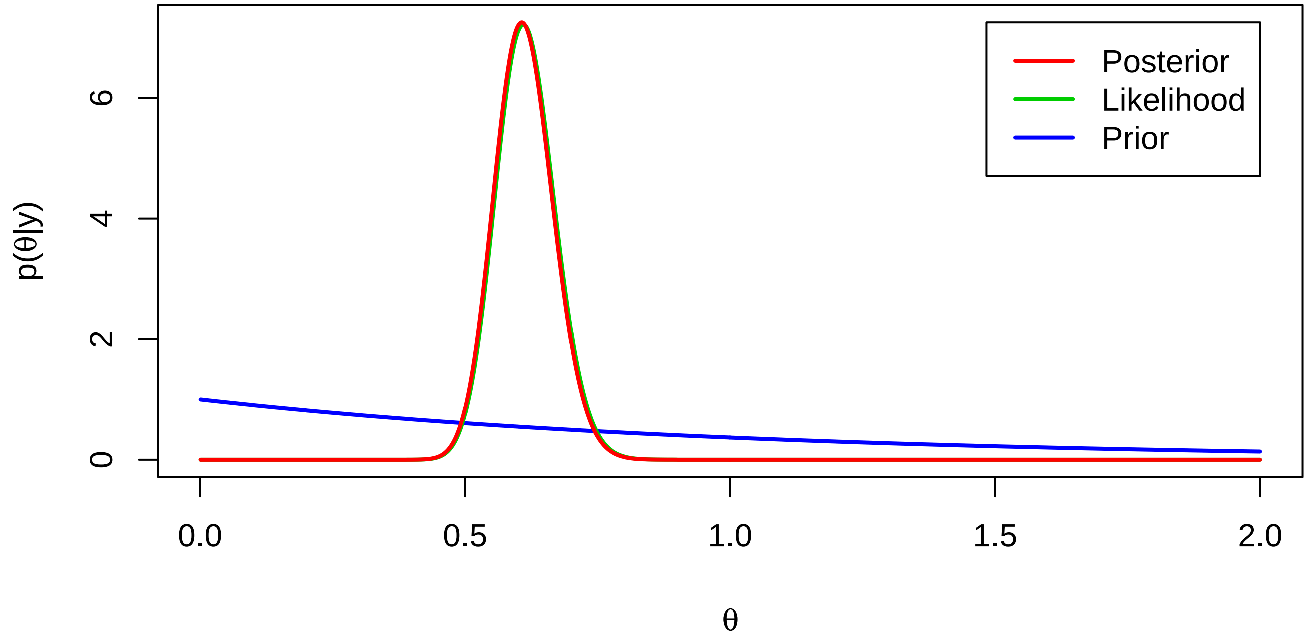


$$E[\theta] = 1 \qquad \hat{\theta} = 0.61 \qquad E[\theta|y] = 0.611$$

$$\mathrm{Var}(\theta) = 2 \qquad \mathrm{Var}(\theta|y) = 0.0030$$

$$\mathrm{SD}(\theta) = 1.412 \qquad \mathrm{SD}(\theta|y) = 0.055$$

$n = 200, \bar{y} = 0.61, \alpha = \beta = 1$

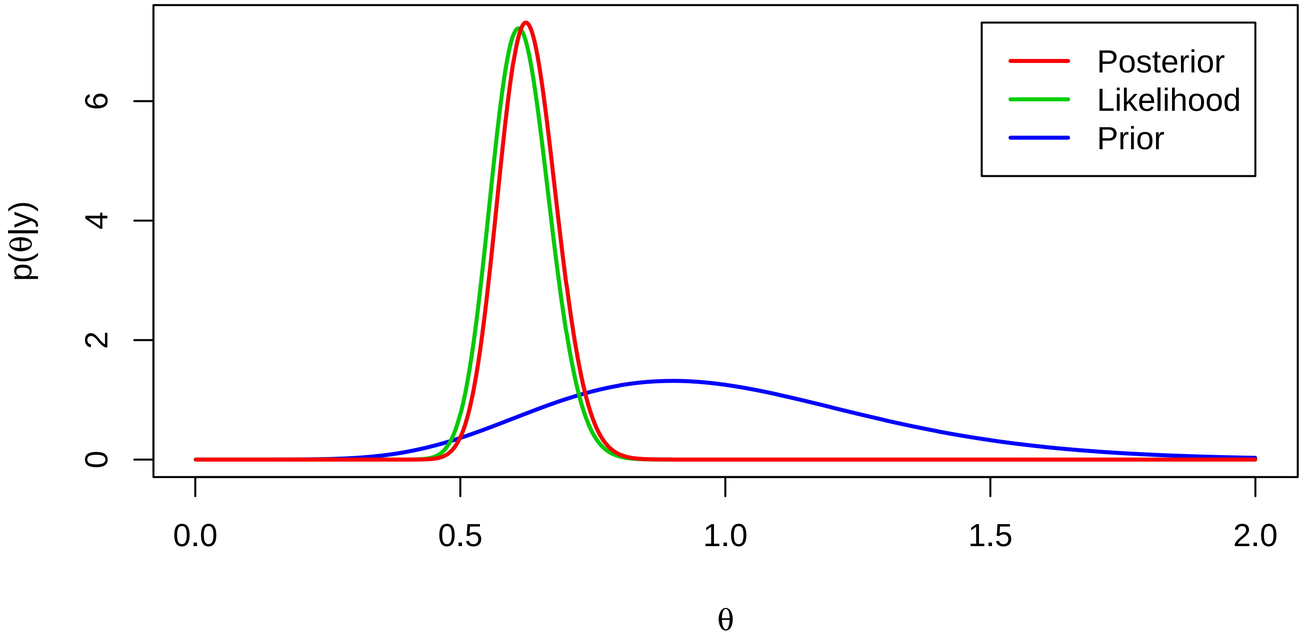


$$E[\theta] = 1 \qquad \hat{\theta} = 0.61 \qquad E[\theta|y] = 0.612$$

$$\mathrm{Var}(\theta) = 1 \qquad \mathrm{Var}(\theta|y) = 0.0030$$

$$\mathrm{SD}(\theta) = 1 \qquad \mathrm{SD}(\theta|y) = 0.055$$

$n = 200, \bar{y} = 0.61, \alpha = \beta = 10$

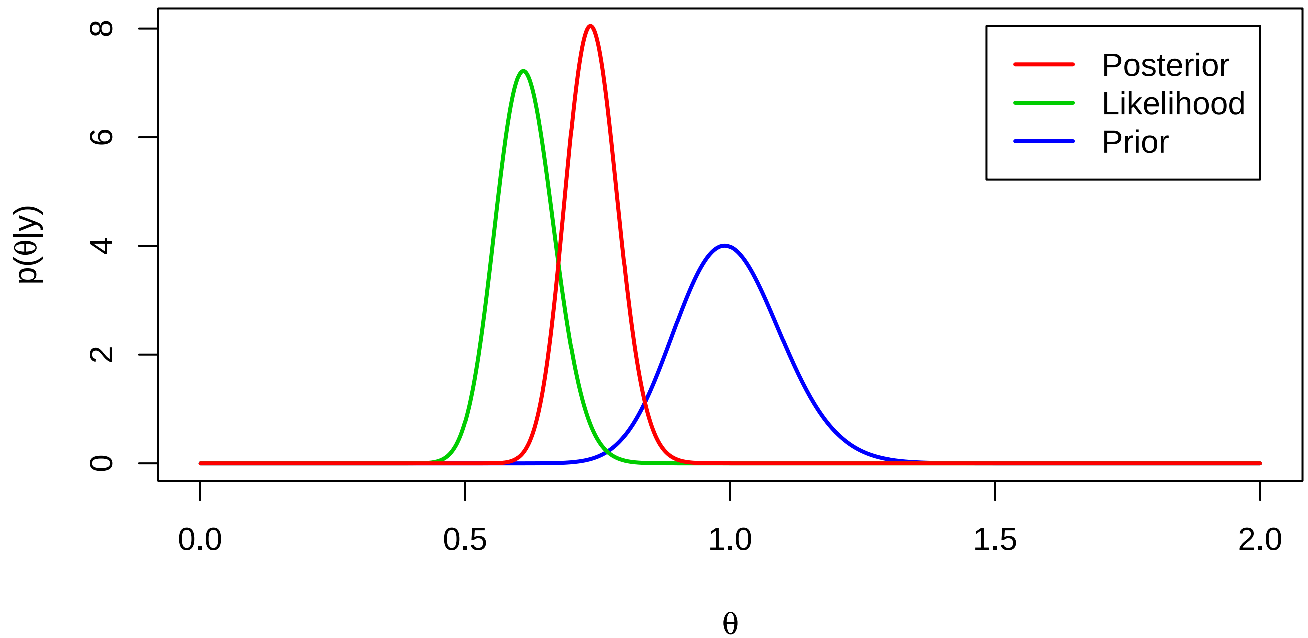


$$E[\theta] = 1 \qquad \hat{\theta} = 0.61 \qquad E[\theta|y] = 0.629$$

$$\mathrm{Var}(\theta) = 0.1 \qquad \mathrm{Var}(\theta|y) = 0.0030$$

$$\mathrm{SD}(\theta) = 0.316 \qquad \mathrm{SD}(\theta|y) = 0.055$$

$n = 200, \bar{y} = 0.61, \alpha = \beta = 100$



$$E[\theta] = 1 \qquad \hat{\theta} = 0.61 \qquad E[\theta|y] = 0.74$$

$$\mathrm{Var}(\theta) = 0.01 \qquad \mathrm{Var}(\theta|y) = 0.0025$$

$$\mathrm{SD}(\theta) = 0.1 \qquad \mathrm{SD}(\theta|y) = 0.050$$

One way to think of the gamma prior in this case is that you have a data set with β observations with and observed Poisson count of α.

Note that the Gamma distribution can be parameterized many ways.

Often the scale parameter form $\lambda = \frac{1}{\beta}$ is used.

Also it can be parameterized in terms of mean, variance, and coefficient of variation (only two are needed).

This gives some flexibility in thinking about the desired form of the prior for a particular model.

In the example, I fixed the mean at 1 and let the variance decrease.

Marginal data distribution for Poisson-Gamma Model

The marginal distribution of a single observation (also known as the prior predictive distribution) in this model is

$$
\begin{aligned}
p(y) &= \frac{p(y|\theta)p(\theta)}{p(\theta|y)} \\
&= \frac{\frac{1}{y!}\theta^y e^{-\theta} \frac{\beta^\alpha}{\Gamma(\alpha+y)} \theta^{(\alpha+y)-1} e^{-(\beta+1)\theta}}{\frac{(\beta+1)^{\alpha+y}}{\Gamma(\alpha+y)} \theta^{\alpha+y-1} e^{-\beta\theta}} \\
&= \binom{\alpha+y-1}{y} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^y \qquad y = 0, 1, \ldots
\end{aligned}
$$

Which is the Negative Binomial distribution (Neg-Bin(α, β)).

The mean and variance are

$$E[y] = \frac{\alpha}{\beta} \qquad \mathrm{Var}(y) = \frac{\alpha}{\beta^2}(\beta + 1) = \frac{\alpha}{\beta}\frac{\beta + 1}{\beta}$$

Note that the variance of the Neg-Bin with mean $\frac{\alpha}{\beta}$ is greater than the variance of a Poisson with the same mean (Var $= \frac{\alpha}{\beta}$). This distribution can be used for count data which is more dispersed than you would expect with the Poisson distribution (like the Beta-Binomial).

2. Exponential

Example: Air Conditioning Failures in a Boeing 720 (Proschan, 1963)

For plane 7910 (there are 13 planes in the complete dataset), the times between failures (in hours) are 74, 57, 48, 29, 502, 12, 70, 21, 29, 386, 59, 27, 153, 26, 326 (n = 15)

Assume that $y_i|\theta \overset{iid}{\sim} Exp(\theta)$ where $\theta = \frac{1}{E[y|\theta]}$ is often referred to as the rate parameter (i.e. number of events per unit time). Note that the exponential distribution is a special case of the gamma ($Exp(\theta) = Gamma(1, \theta)$).

$$p(y|\theta) = \prod_{i=1}^{n} \theta e^{-y_i\theta} = \theta^n e^{-n\bar{y}\theta}$$

which gives the MLE as

$$\hat{\theta} = \frac{1}{\bar{y}}$$

For the example, $\hat{\theta} = 0.00825$ (about 8 failures for every 1000 hours of flight time). Note that $\frac{1}{\hat{\theta}} = 121.27$ hours is the average time between failures.

A conjugate prior for the exponential distribution is the gamma. If $\theta \sim Gamma(\alpha, \beta)$, then the posterior is

$$p(\theta|y) \propto \theta^n e^{-n\bar{y}\theta} \theta^{\alpha-1} e^{\beta\theta} = \theta^{\alpha+n-1} e^{\beta+n\bar{y}}$$

which is proportional to a $Gamma(\alpha + n, \beta + n\bar{y})$ density.

This gamma prior can be thought of as $\alpha - 1$ exponential observations totalling β.

What prior for this example:

For the example, we can use some of the other planes to develop a gamma prior for plane 7910. I took 4 of the planes and calculated the MLEs of θ for each of them. The average of these was about 0.009 with a standard deviation of 0.003. The Gamma distribution with this mean and standard deviation has

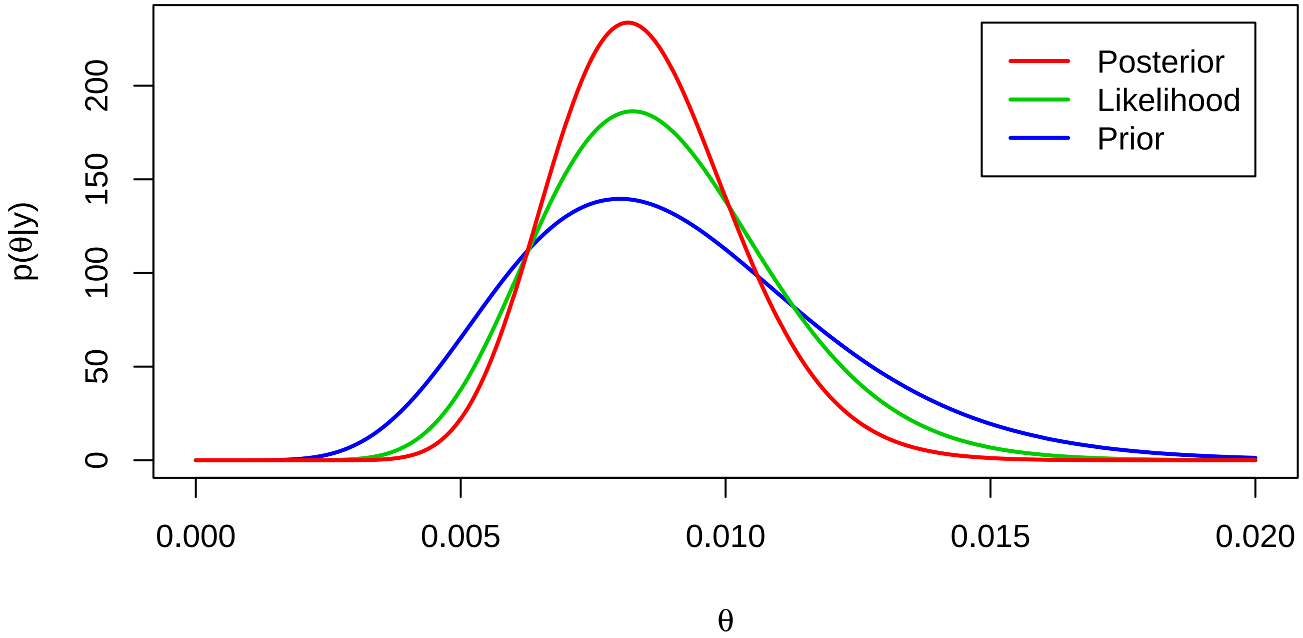
$$\alpha = 9 \qquad \beta = 1000$$

(Note: this is a complete hack)

As this is a complete hack, lets also use a less informative prior to see how dependent on our answer is on our prior choice.

Since the above prior corresponds to 8 observations, lets use corresponding to half as many observations ($\alpha = 5$), with half as much time ($\beta = 500$)

$n = 15, \bar{y} = 121.27, \alpha = 9, \beta = 1000$

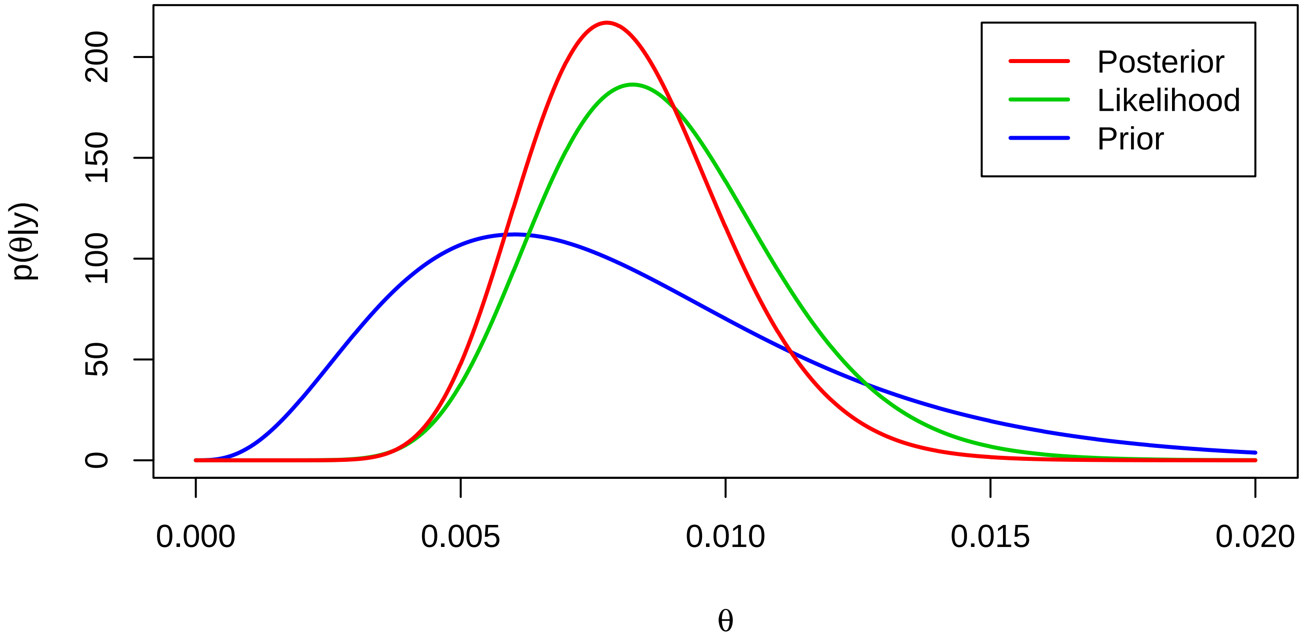


$$E[\theta] = 0.009 \qquad \hat{\theta} = 0.00825 \qquad E[\theta|y] = 0.00851$$

$$\mathrm{Var}(\theta) = 0.000009 \qquad \mathrm{Var}(\theta|y) = 0.000003$$

$$\mathrm{SD}(\theta) = 0.003 \qquad \mathrm{SD}(\theta|y) = 0.00173$$

$n = 15, \bar{y} = 121.27, \alpha = 5, \beta = 500$



$$E[\theta] = 0.01 \qquad \hat{\theta} = 0.00825 \qquad E[\theta|y] = 0.00862$$

$$\mathrm{Var}(\theta) = 0.0000200 \qquad \mathrm{Var}(\theta|y) = 0.0000037$$

$$\mathrm{SD}(\theta) = 0.00447 \qquad \mathrm{SD}(\theta|y) = 0.00193$$

Suppose that we are interested in $\mu = \frac{1}{\theta} = E[y|\theta]$, the expected time between breakdowns. It is easy to get the expected value of $\frac{1}{\theta}$ in the conjugate prior case.

If $x \sim Gamma(\alpha, \beta)$

$$E\left[\frac{1}{x}\right] = \int_0^\infty \frac{1}{x}\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} = \frac{\beta}{\alpha - 1}$$

(Note that its not the reciprocal of $E[x]$.)

Similarly

$$\mathrm{Var}\left(\frac{1}{x}\right) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$$

For the example,

Prior	$E\left[\frac{1}{\theta}\middle|y\right]$	$\mathrm{Var}\left(\frac{1}{\theta}\middle|y\right)$	$\mathrm{SD}\left(\frac{1}{\theta}\middle|y\right)$
$\alpha = 9, \beta = 1000$	122.56	682.83	26.13
$\alpha = 5, \beta = 500$	122.05	827.60	28.77

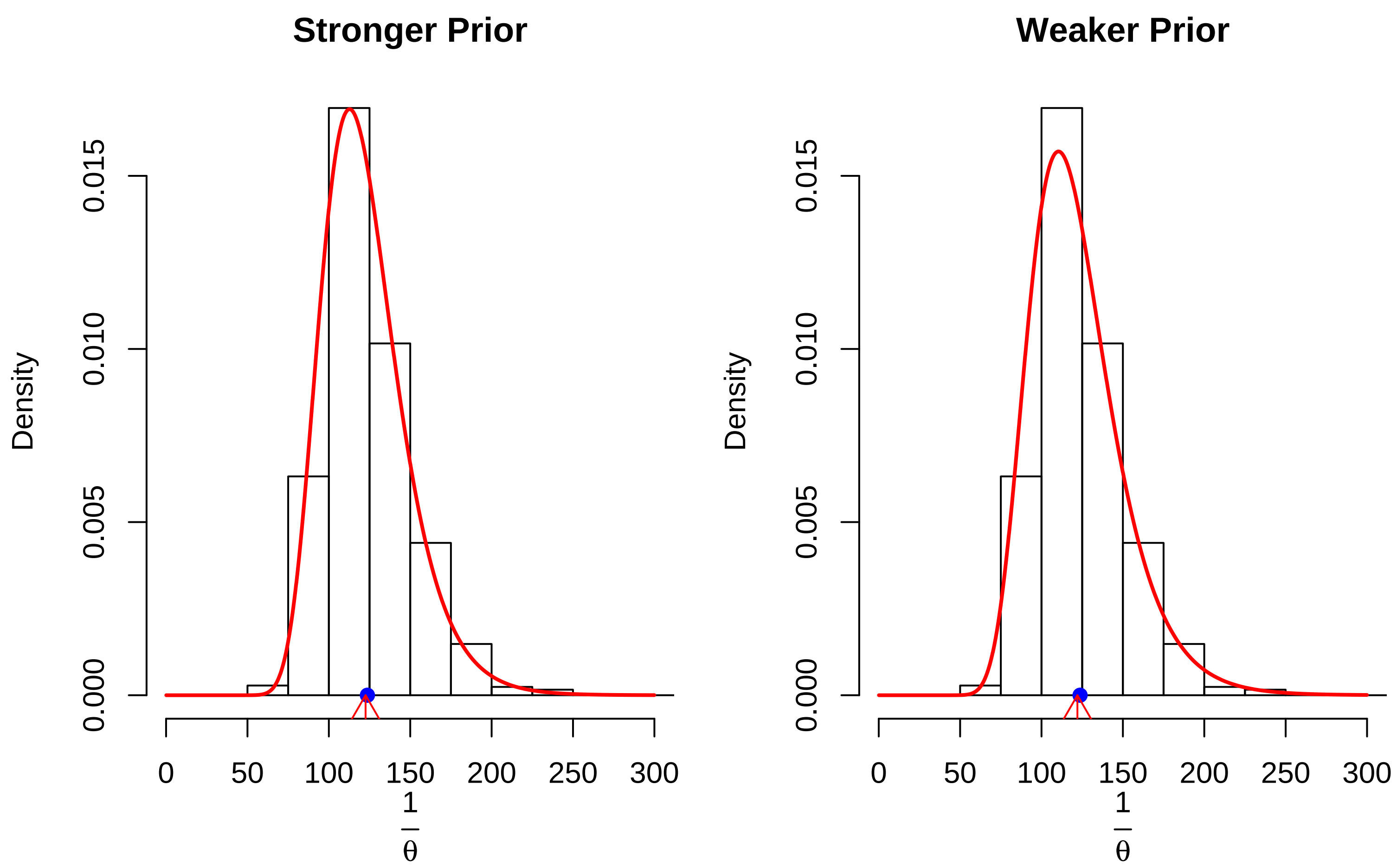
In fact, we know more about distribution of $\frac{1}{\theta}|y$. It has an inverse gamma distribution. So plotting the density, etc isn't difficult.

However we can answer many questions by simulation.

Let $\theta_i \overset{iid}{\sim} p(\theta|y); i = 1, \ldots, m$ and suppose we are interested in $\lambda = f(\theta)$ for some function $f(\cdot)$. Then

$$f(\theta_i) = \lambda_i \overset{iid}{\sim} p(\lambda|y); i = 1, \ldots, m$$

So for example, $\bar{\lambda}$ is an unbiased estimate of $E\left[\frac{1}{\theta}\middle|y\right]$



Prior	$E\left[\frac{1}{\theta}\middle|y\right]$	$\hat{E}\left[\frac{1}{\theta}\middle|y\right]$	$\mathrm{SD}\left(\frac{1}{\theta}\middle|y\right)$	$\widehat{\mathrm{SD}}\left(\frac{1}{\theta}\middle|y\right)$
$\alpha = 9, \beta = 1000$	122.56	123.68	26.13	27.00
$\alpha = 5, \beta = 500$	122.05	122.10	28.77	27.57