1 Inference on the probability of success

The purpose of this writing is to illustrate the basic principles of Bayesian inference through a concrete example. We consider an experimental process that results in seven successes and three failures:

SSFSSFSSSF

The probability of success θ in each experiment is unknown and it is constant across experiments. In the classical (frequentist) approach, θ is assumed to have a fixed value that needs to be estimated from the available information. By contrast, under the Bayesian approach, the unknown success probability θ is considered random and it is characterized by a probability distribution. Once a new piece of information becomes available, our prior beliefs about θ (expressed through a prior distribution) are updated to new set of beliefs, also expressed through a probability distribution (called the posterior distribution). The relationship between the prior and the posterior distributions is given by the Bayes' rule.

The rigorous formulation of the experimental process that results in the observations SSFSSFSSSF is as follows. We assume that $Y_1, Y_2, \ldots, Y_{10}$ are independent and identically distributed Bernoulli random variables with probability of success θ. We write $Y_i \sim \mathsf{Ber}(\theta)$ for $i = 1, 2, \ldots, 10$. That is

$$\mathsf{Pr}(Y_i = 1 \mid \theta) = \theta, \quad \mathsf{Pr}(Y_i = 0 \mid \theta) = 1 - \theta.$$

or, equivalently,

$$\mathsf{Pr}(Y_i = y \mid \theta) = \theta^y (1-\theta)^{1-y}, \text{ for } y \in \{0, 1\}.$$

Here 1 stands for success and 0 stands for failure. In our case we see $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$, $Y_9 = 1$, $Y_{10} = 0$. Thus the probability of observing SSFSSFSSSF given θ is

$$\begin{aligned}
\mathsf{Pr}(Y_1, \ldots, Y_{10} \mid \theta) &= \mathsf{Pr}(Y_1 \mid \theta) \cdot \mathsf{Pr}(Y_2 \mid \theta) \cdot \ldots \cdot \mathsf{Pr}(Y_{10} \mid \theta) \\
&= \theta^{Y_1 + \ldots + Y_{10}} (1-\theta)^{10 - (Y_1 + \ldots + Y_{10})}, \\
&= \theta^7 (1-\theta)^3.
\end{aligned}$$

The MLE of θ is determined by maximizing the log-likelihood

$$l(\theta) = \log \mathsf{Pr}(Y_1, \ldots, Y_{10} \mid \theta) = 7 \log(\theta) + 3 \log(1-\theta).$$

which gives

$$\widehat{\theta}_{MLE} = \frac{7}{10}.$$

That is, the MLE of θ is the number of successes divided by the number of experiments (trials). Table 1 gives the MLEs of θ given the sequential accumulation of data from additional experiments: S, SS, SSF, ..., SSFSSFSSSF. We see that the frequentist estimates of the probability of success increase with each success and decrease with each failure. Table 1 gives the corresponding Bayesian estimates of θ. We see that, as the number of trials increases, the frequentist and the Bayesian estimates become very close. In other words, as more information becomes available, it seems that it no longer matters whether one chooses to be frequentist

Data	MLE of θ	Posterior mean of θ	Absolute value of the difference between the MLE and the posterior mean of θ
S	1	0.667	0.223
SS	1	0.75	0.250
SSF	0.667	0.6	0.067
SSFS	0.75	0.667	0.083
SSFSS	0.8	0.714	0.086
SSFSSF	0.667	0.625	0.042
SSFSSFS	0.714	0.667	0.047
SSFSSFSS	0.75	0.7	0.050
SSFSSFSSS	0.778	0.727	0.051
SSFSSFSSSF	0.7	0.667	0.033

Table 1: Comparison of frequentist estimates of the probability of success θ (MLEs) and Bayesian estimates of θ (posterior means).

or Bayesian. While this is in essence true (and this is where your intuition should be), the manner in which *uncertainty* is expressed is still very different in the frequentist and Bayesian thinking. The sections that follow give concrete examples of how our beliefs are changed by the accumulation of information and how uncertainty is expressed through probability distributions.

2 Observed data: S ($Y_1 = 1$)

Before seeing any data, we do not have any prior knowledge about what the probability of success θ should be. That is, we know that θ must be between 0 and 1, but we do not have any preference for one value in the interval $(0, 1)$ versus another. We express our current beliefs through a uniform distribution for θ:

$$\theta \sim \mathsf{Uni}(0,1), \quad \mathsf{Pr}(\theta) = 1, \text{ for } \theta \in (0,1).$$

Please see the left panel of Figure 1. Under the uniform distribution, the mean of θ is

$$\mathsf{E}[\theta] = \int_0^1 \theta \, d\theta = 0.5$$

That is, before seeing any data, we expect θ to be right in the middle of the interval $(0, 1)$. When we observe $Y_1 = 1$, we use the Bayes rule to update our beliefs and determine the posterior distribution of θ:

$$\mathsf{Pr}(\theta \mid Y_1) \quad \propto \quad \mathsf{Pr}(Y_1 \mid \theta)\mathsf{Pr}(\theta).$$

This equation means: the posterior distribution is proportional with the likelihood times the prior. Since we have observed a success $Y_1 = 1$, the corresponding likelihood is:

$$\mathsf{Pr}(Y_1 = 1 \mid \theta) = \theta,$$

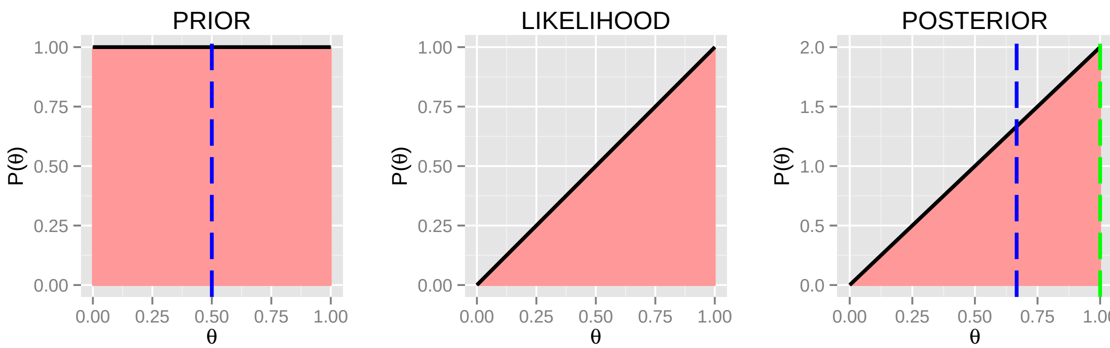

Figure 1: *Left panel*: prior distribution of θ before observing any data. The dotted blue line gives the prior mean. *Middle panel*: likelihood associated with a success S ($Y_1 = 1$). *Right panel*: posterior distribution of θ after observing S ($Y_1 = 1$). The dotted blue line gives the posterior mean and the dotted green line gives the MLE.

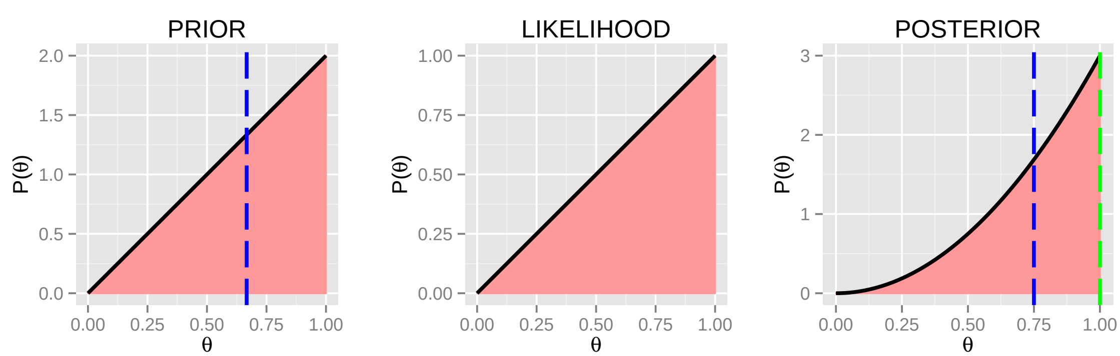

Figure 2: *Left panel*: prior distribution of θ before observing $Y_2 = 1$. The dotted blue line gives the prior mean. *Middle panel*: likelihood associated with a success S ($Y_2 = 1$). *Right panel*: posterior distribution of θ after observing SS ($Y_1 = 1$, $Y_2 = 1$). The dotted blue line gives the posterior mean and the dotted green line gives the MLE.

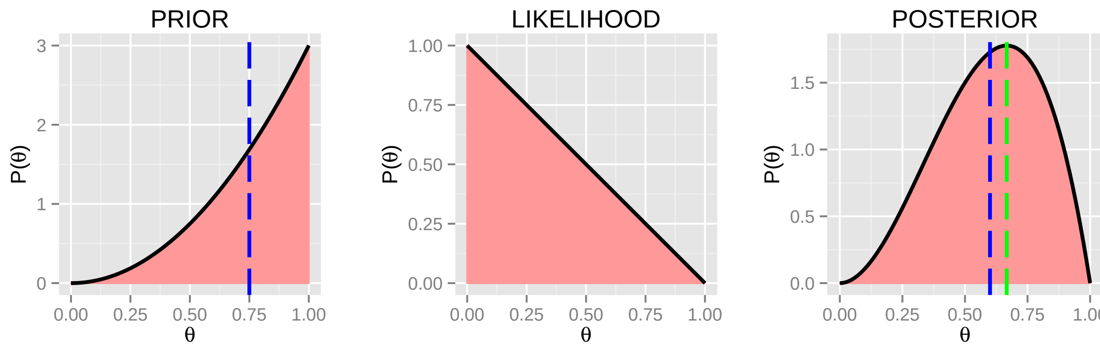

Figure 3: *Left panel*: prior distribution of θ before observing $Y_3 = 0$. The dotted blue line gives the prior mean. *Middle panel*: likelihood associated with a failure ($Y_3 = 0$). *Right panel*: posterior distribution of θ after observing SSF ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$). The dotted blue line gives the posterior mean and the dotted green line gives the MLE.

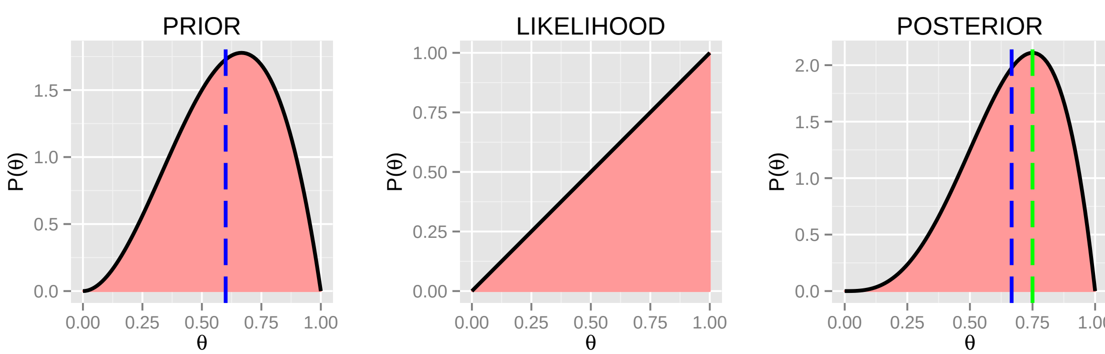

Figure 4: *Left panel*: prior distribution of θ before observing $Y_4 = 1$. The dotted blue line gives the prior mean. *Middle panel*: likelihood associated with a success ($Y_4 = 1$). *Right panel*: posterior distribution of θ after observing SSFS ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$). The dotted blue line gives the posterior mean and the dotted green line gives the MLE.

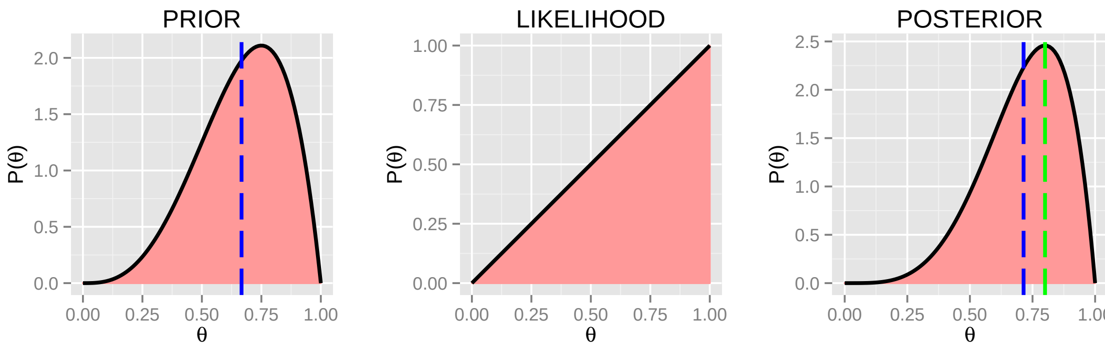

Figure 5: *Left panel*: prior distribution of θ before observing $Y_5 = 1$. The dotted blue line gives the prior mean. *Middle panel*: likelihood associated with a success ($Y_5 = 1$). *Right panel*: posterior distribution of θ after observing SSFSS ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$). The dotted blue line gives the posterior mean and the dotted green line gives the MLE.

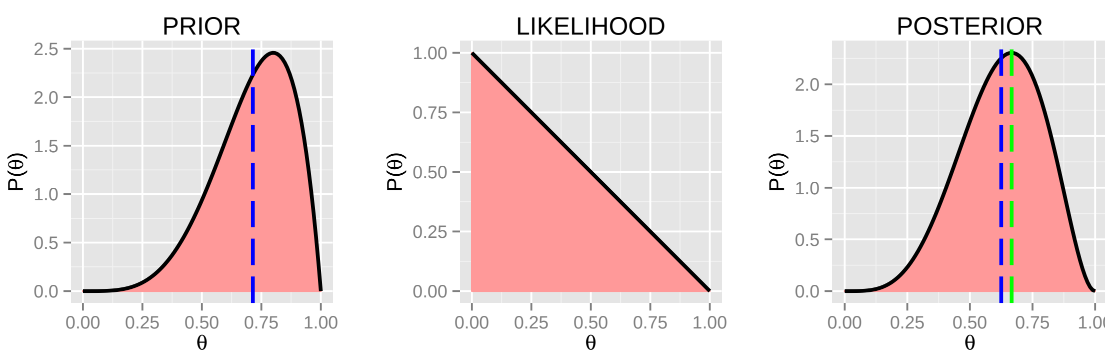

Figure 6: *Left panel*: prior distribution of θ before observing $Y_6 = 0$. The dotted blue line gives the prior mean. *Middle panel*: likelihood associated with a failure ($Y_6 = 0$). *Right panel*: posterior distribution of θ after observing SSFSSF ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$). The dotted blue line gives the posterior mean and the dotted green line gives the MLE.

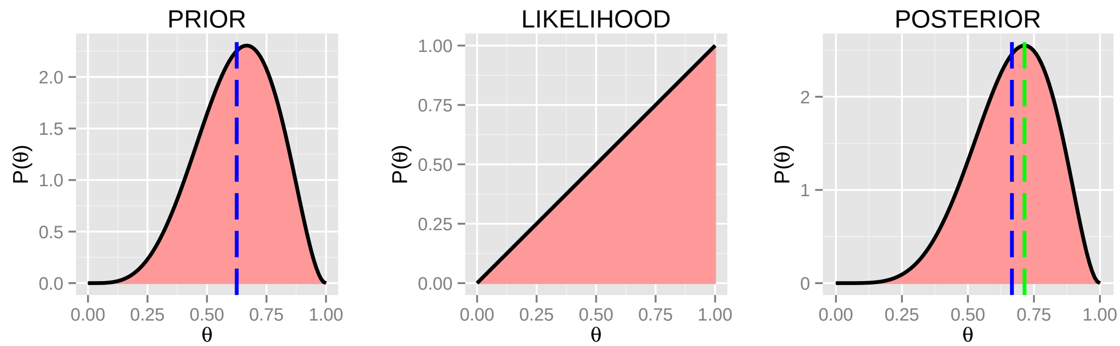

Figure 7: *Left panel*: prior distribution of θ before observing $Y_7 = 1$. The dotted blue line gives the prior mean. *Middle panel*: likelihood associated with a success ($Y_7 = 1$). *Right panel*: posterior distribution of θ after observing SSFSSFS ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$). The dotted blue line gives the posterior mean and the dotted green line gives the MLE.

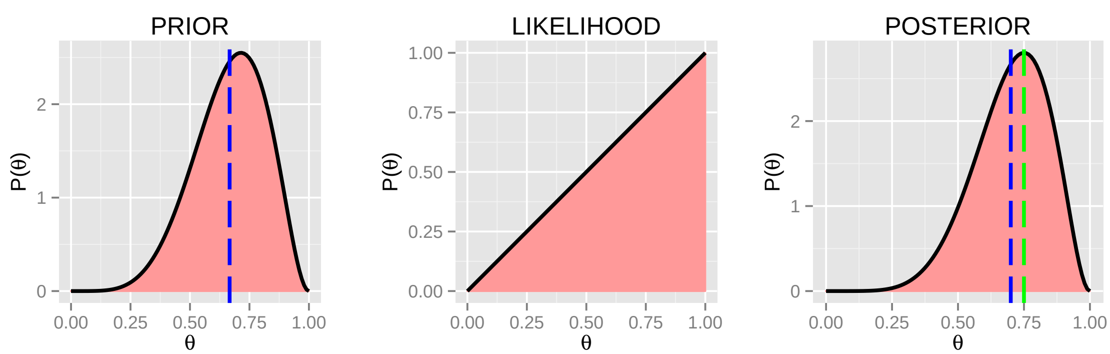

Figure 8: *Left panel*: prior distribution of θ before observing $Y_8 = 1$. The dotted blue line gives the prior mean. *Middle panel*: likelihood associated with a success ($Y_8 = 1$). *Right panel*: posterior distribution of θ after observing SSFSSFSS ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$). The dotted blue line gives the posterior mean and the dotted green line gives the MLE.

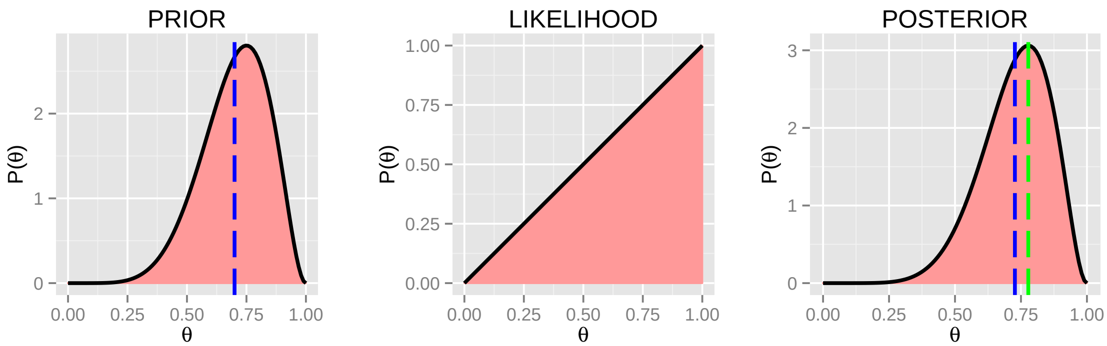

Figure 9: *Left panel*: prior distribution of θ before observing $Y_9 = 1$. The dotted blue line gives the prior mean. *Middle panel*: likelihood associated with a success ($Y_9 = 1$). *Right panel*: posterior distribution of θ after observing SSFSSFSSS ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$, $Y_9 = 1$). The dotted blue line gives the posterior mean and the dotted green line gives the MLE.

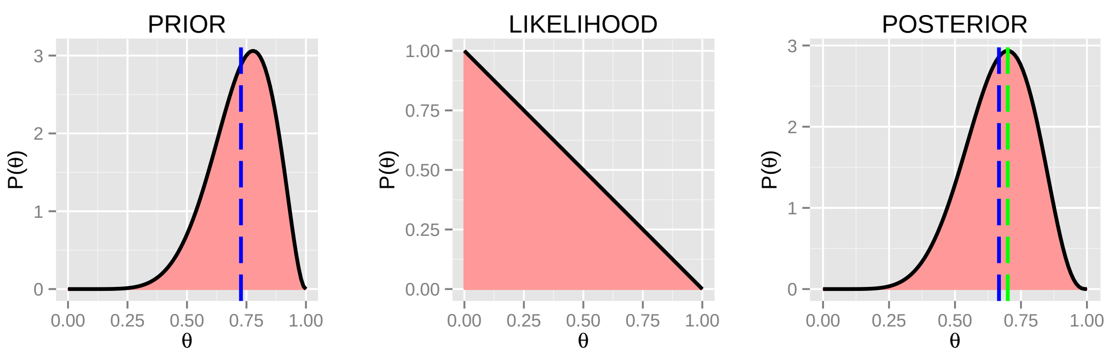

Figure 10: *Left panel*: prior distribution of θ before observing $Y_{10} = 0$. The dotted blue line gives the prior mean. *Middle panel*: likelihood associated with a failure ($Y_{10} = 0$). *Right panel*: posterior distribution of θ after observing SSFSSFSSSF ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$, $Y_9 = 1$, $Y_{10} = 0$). The dotted blue line gives the posterior mean and the dotted green line gives the MLE.

which implies that the posterior distribution of θ is proportional with

$$\mathsf{Pr}(\theta \mid Y_1) \quad \propto \quad \theta.$$

What is missing here is the normalizing constant that assures the posterior distribution of θ integrates to 1 (the area under the curve in the right panel of Figure 1 is 1):

$$\mathsf{Pr}(Y_1) = \int_0^1 \theta \, d\theta = \left.\frac{\theta^2}{2}\right|_0^1 = 0.5$$

The constant $\mathsf{Pr}(Y_1)$ is called the *marginal likelihood* or the *integrated likelihood*. The complete expression for the posterior distribution of θ is:

$$\mathsf{Pr}(\theta \mid Y_1) = \left(\int_0^1 \theta \, d\theta \right)^{-1} \theta = 2\theta, \text{ for } \theta \in (0, 1).$$

The posterior mean of θ is calculated as follows

$$\mathsf{E}[\theta \mid Y_1] = \int_0^1 \theta \mathsf{Pr}(\theta \mid Y_1) \, d\theta = \frac{2}{3}.$$

Therefore, after observing one success, we change our belief about the expected value of θ from 0.5 (prior mean) to 0.667 (posterior mean). By comparison, the MLE of θ given $Y_1 = 1$ is 1 – see Table 1. Compare the shape of the prior and posterior distributions for θ given the observed data S in Figure 1.

3 Observed data: SS ($Y_1 = 1$, $Y_2 = 1$)

After observing $Y_1 = 1$, our knowledge about θ is expressed through the distribution

$$\mathsf{Pr}(\theta \mid Y_1) = 2\theta, \text{ for } \theta \in (0, 1).$$

This distribution is the prior for θ before observing $Y_2 = 1$. We use the Bayes rule to update our beliefs after observing $Y_2 = 1$:

$$\mathsf{Pr}(\theta \mid Y_1, Y_2) \quad \propto \quad \mathsf{Pr}(Y_2 \mid \theta)\mathsf{Pr}(\theta \mid Y_1).$$

The likelihood associated with $Y_2 = 1$ is $\mathsf{Pr}(Y_2 = 1 \mid \theta) = \theta$, hence

$$\mathsf{Pr}(\theta \mid Y_1, Y_2) \quad \propto \quad 2\theta^2.$$

The normalizing constant (marginal likelihood) of this posterior distribution is

$$\mathsf{Pr}(Y_2 \mid Y_1) = \int_0^1 2\theta^2 \, d\theta = \frac{2}{3}.$$

Thus the complete expression for the posterior distribution of θ is:

$$\mathsf{Pr}(\theta \mid Y_1, Y_2) = \left(\int_0^1 2\theta^2 \, d\theta \right)^{-1} 2\theta^2 = 3\theta^2, \text{ for } \theta \in (0, 1).$$

The posterior mean of θ is calculated as follows

$$\mathsf{E}[\theta \mid Y_1, Y_2] = \int_0^1 \theta \mathsf{Pr}(\theta \mid Y_1, Y_2)\, d\theta = \frac{3}{4}.$$

Therefore, after observing two successes, we change our belief about the expected value of θ from 0.667 (prior mean after observing one success) to 0.75 (posterior mean after observing the second success). By comparison, the MLE of θ given $Y_1 = 1$, $Y_2 = 1$ is 1 – see Table 1. Compare the shape of the prior and posterior distributions for θ given the observed data SS in Figure 2.

4 Observed data: SSF ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$)

After observing $Y_1 = 1$, $Y_2 = 1$, our knowledge about θ is expressed through the distribution

$$\mathsf{Pr}(\theta \mid Y_1, Y_2) = 3\theta^2, \text{ for } \theta \in (0, 1).$$

This distribution is the prior for θ before observing $Y_3 = 0$. We use the Bayes rule to update our beliefs after observing $Y_3 = 0$:

$$\mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3) \quad \propto \quad \mathsf{Pr}(Y_3 \mid \theta)\mathsf{Pr}(\theta \mid Y_1, Y_2).$$

The likelihood associated with $Y_3 = 0$ is $\mathsf{Pr}(Y_3 = 0 \mid \theta) = 1 - \theta$, hence

$$\mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3) \quad \propto \quad 3\theta^2(1 - \theta).$$

The normalizing constant (marginal likelihood) of this posterior distribution is

$$\mathsf{Pr}(Y_3 \mid Y_1, Y_2) = \int_0^1 3\theta^2(1 - \theta)\, d\theta = \frac{1}{4}.$$

Thus the complete expression for the posterior distribution of θ is:

$$\mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3) = \left(\int_0^1 3\theta^2(1 - \theta)\, d\theta\right)^{-1} 3\theta^2(1 - \theta) = 12\theta^2(1 - \theta), \text{ for } \theta \in (0, 1).$$

The posterior mean of θ is calculated as follows

$$\mathsf{E}[\theta \mid Y_1, Y_2, Y_3] = \int_0^1 \theta \mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3)\, d\theta = \frac{3}{5}.$$

Therefore, after observing SSF, we change our belief about the expected value of θ from 0.75 (prior mean after observing SS) to 0.6 (posterior mean after observing SSF). By comparison, the MLE of θ given $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$ is $\frac{2}{3}$ – see Table 1. Compare the shape of the prior and posterior distributions for θ given the observed data SSF in Figure 3.

5 Observed data: SSFS ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$)

After observing $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$ our knowledge about θ is expressed through the distribution

$$\mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3) = 12\theta^2(1-\theta), \text{ for } \theta \in (0,1).$$

This distribution is the prior for θ before observing $Y_4 = 1$. We use the Bayes rule to update our beliefs after observing $Y_4 = 1$:

$$\mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3, Y_4) \quad \propto \quad \mathsf{Pr}(Y_4 \mid \theta)\mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3).$$

The likelihood associated with $Y_4 = 1$ is $\mathsf{Pr}(Y_3 = 1 \mid \theta) = \theta$, hence

$$\mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3, Y_4) \quad \propto \quad 12\theta^3(1-\theta).$$

The normalizing constant (marginal likelihood) of this posterior distribution is

$$\mathsf{Pr}(Y_4 \mid Y_1, Y_2, Y_3) = \int_0^1 12\theta^3(1-\theta)\, d\theta = \frac{12}{20}.$$

Thus the complete expression for the posterior distribution of θ is:

$$\mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3, Y_4) = \left(\int_0^1 12\theta^3(1-\theta)\, d\theta\right)^{-1} 12\theta^3(1-\theta) = 20\theta^3(1-\theta), \text{ for } \theta \in (0,1).$$

The posterior mean of θ is calculated as follows

$$\mathsf{E}[\theta \mid Y_1, Y_2, Y_3, Y_4] = \int_0^1 \theta \mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3, Y_4)\, d\theta = \frac{2}{3}.$$

Therefore, after observing SSFS, we change our belief about the expected value of θ from 0.6 (prior mean after observing SSF) to 0.667 (posterior mean after observing SSFS). By comparison, the MLE of θ given $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$ is $\frac{3}{4}$ – see Table 1. Compare the shape of the prior and posterior distributions for θ given the observed data $SSFS$ in Figure 4.

6 Observed data: SSFSS ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$)

After observing $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$ our knowledge about θ is expressed through the distribution

$$\mathsf{Pr}(\theta \mid Y_1, Y_2, Y_3, Y_4) = 20\theta^3(1-\theta), \text{ for } \theta \in (0,1).$$

This distribution is the prior for θ before observing $Y_5 = 1$. We use the Bayes rule to update our beliefs after observing $Y_5 = 1$:

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_4, Y_5) \quad \propto \quad \mathsf{Pr}(Y_5 \mid \theta)\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_4).$$

The likelihood associated with $Y_5 = 1$ is $\mathsf{Pr}(Y_5 = 1 \mid \theta) = \theta$, hence

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_5) \quad \propto \quad 20\theta^4(1-\theta).$$

The normalizing constant (marginal likelihood) of this posterior distribution is

$$\mathsf{Pr}(Y_5 \mid Y_1, \ldots, Y_4) = \int_0^1 20\theta^4(1-\theta)\, d\theta = \frac{20}{30}.$$

Thus the complete expression for the posterior distribution of θ is:

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_5) = \left(\int_0^1 20\theta^4(1-\theta)\, d\theta\right)^{-1} 20\theta^4(1-\theta) = 30\theta^4(1-\theta), \text{ for } \theta \in (0,1).$$

The posterior mean of θ is calculated as follows

$$\mathsf{E}[\theta \mid Y_1, \ldots, Y_5] = \int_0^1 \theta \mathsf{Pr}(\theta \mid Y_1, \ldots, Y_5)\, d\theta = \frac{5}{7}.$$

Therefore, after observing SSFSS, we change our belief about the expected value of θ from 0.667 (prior mean after observing SSFS) to 0.714 (posterior mean after observing SSFSS). By comparison, the MLE of θ given $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$ is $\frac{4}{5}$ – see Table 1. Compare the shape of the prior and posterior distributions for θ given the observed data $SSFSS$ in Figure 5.

7 Observed data: SSFSSF ($Y_1 = 1, Y_2 = 1, Y_3 = 0, Y_4 = 1, Y_5 = 1, Y_6 = 0$)

After observing $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$ our knowledge about θ is expressed through the distribution

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_5) = 30\theta^4(1-\theta), \text{ for } \theta \in (0,1).$$

This distribution is the prior for θ before observing $Y_6 = 0$. We use the Bayes rule to update our beliefs after observing $Y_6 = 0$:

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_5, Y_6) \quad \propto \quad \mathsf{Pr}(Y_6 \mid \theta)\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_5).$$

The likelihood associated with $Y_6 = 0$ is $\mathsf{Pr}(Y_6 = 0 \mid \theta) = 1 - \theta$, hence

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_6) \quad \propto \quad 30\theta^4(1-\theta)^2.$$

The normalizing constant (marginal likelihood) of this posterior distribution is

$$\mathsf{Pr}(Y_6 \mid Y_1, \ldots, Y_5) = \int_0^1 30\theta^4(1-\theta)^2\, d\theta = \frac{2}{7}.$$

Thus the complete expression for the posterior distribution of θ is:

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_6) = \left(\int_0^1 30\theta^4(1-\theta)^2\, d\theta\right)^{-1} 30\theta^4(1-\theta)^2 = 105\theta^4(1-\theta)^2, \text{ for } \theta \in (0,1).$$

The posterior mean of θ is calculated as follows

$$\mathsf{E}[\theta \mid Y_1, \ldots, Y_6] = \int_0^1 \theta \mathsf{Pr}(\theta \mid Y_1, \ldots, Y_6)\, d\theta = \frac{105}{168}.$$

Therefore, after observing SSFSSF, we change our belief about the expected value of θ from 0.714 (prior mean after observing SSFSS) to 0.625 (posterior mean after observing SSFSSF). By comparison, the MLE of θ given $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$ is $\frac{4}{6}$ – see Table 1. Compare the shape of the prior and posterior distributions for θ given the observed data $SSFSSF$ in Figure 6.

8 Observed data: SSFSSFS ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$)

After observing $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$ our knowledge about θ is expressed through the distribution

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_6) = 105\theta^4(1-\theta)^2, \text{ for } \theta \in (0, 1).$$

This distribution is the prior for θ before observing $Y_7 = 1$. We use the Bayes rule to update our beliefs after observing $Y_7 = 1$:

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_6, Y_7) \quad \propto \quad \mathsf{Pr}(Y_7 \mid \theta)\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_6).$$

The likelihood associated with $Y_7 = 1$ is $\mathsf{Pr}(Y_7 = 1 \mid \theta) = \theta$, hence

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_7) \quad \propto \quad 105\theta^5(1-\theta)^2.$$

The normalizing constant (marginal likelihood) of this posterior distribution is

$$\mathsf{Pr}(Y_7 \mid Y_1, \ldots, Y_6) = \int_0^1 105\theta^5(1-\theta)^2\, d\theta = \frac{105}{168}.$$

Thus the complete expression for the posterior distribution of θ is:

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_7) = \left(\int_0^1 105\theta^5(1-\theta)^2\, d\theta\right)^{-1} 105\theta^5(1-\theta)^2 = 168\theta^5(1-\theta)^2, \text{ for } \theta \in (0, 1).$$

The posterior mean of θ is calculated as follows

$$\mathsf{E}[\theta \mid Y_1, \ldots, Y_7] = \int_0^1 \theta \mathsf{Pr}(\theta \mid Y_1, \ldots, Y_7)\, d\theta = \frac{2}{3}.$$

Therefore, after observing SSFSSFS, we change our belief about the expected value of θ from 0.625 (prior mean after observing SSFSSF) to 0.667 (posterior mean after observing SSFSSFS). By comparison, the MLE of θ given $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$ is $\frac{5}{7}$ – see Table 1. Compare the shape of the prior and posterior distributions for θ given the observed data $SSFSSFS$ in Figure 7.

9 Observed data: SSFSSFSS ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$)

After observing $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$ our knowledge about θ is expressed through the distribution

$$\Pr(\theta \mid Y_1, \ldots, Y_7) = 168\theta^5(1-\theta)^2, \text{ for } \theta \in (0,1).$$

This distribution is the prior for θ before observing $Y_8 = 1$. We use the Bayes rule to update our beliefs after observing $Y_8 = 1$:

$$\Pr(\theta \mid Y_1, \ldots, Y_7, Y_8) \quad \propto \quad \Pr(Y_8 \mid \theta)\Pr(\theta \mid Y_1, \ldots, Y_7).$$

The likelihood associated with $Y_8 = 1$ is $\Pr(Y_8 = 1 \mid \theta) = \theta$, hence

$$\Pr(\theta \mid Y_1, \ldots, Y_8) \quad \propto \quad 168\theta^6(1-\theta)^2.$$

The normalizing constant (marginal likelihood) of this posterior distribution is

$$\Pr(Y_8 \mid Y_1, \ldots, Y_7) = \int_0^1 168\theta^6(1-\theta)^2 \, d\theta = \frac{168}{252}.$$

Thus the complete expression for the posterior distribution of θ is:

$$\Pr(\theta \mid Y_1, \ldots, Y_8) = \left(\int_0^1 168\theta^6(1-\theta)^2 \, d\theta\right)^{-1} 168\theta^6(1-\theta)^2 = 252\theta^6(1-\theta)^2, \text{ for } \theta \in (0,1).$$

The posterior mean of θ is calculated as follows

$$\mathrm{E}[\theta \mid Y_1, \ldots, Y_8] = \int_0^1 \theta \Pr(\theta \mid Y_1, \ldots, Y_8) \, d\theta = 0.7.$$

Therefore, after observing SSFSSFSS, we change our belief about the expected value of θ from 0.667 (prior mean after observing SSFSSFS) to 0.7 (posterior mean after observing SSFSSFSS). By comparison, the MLE of θ given $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$ is $\frac{6}{8}$ – see Table 1. Compare the shape of the prior and posterior distributions for θ given the observed data $SSFSSFSS$ in Figure 8.

10 Observed data: SSFSSFSSS ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$, $Y_9 = 1$)

After observing $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$ our knowledge about θ is expressed through the distribution

$$\Pr(\theta \mid Y_1, \ldots, Y_8) = 252\theta^6(1-\theta)^2, \text{ for } \theta \in (0,1).$$

This distribution is the prior for θ before observing $Y_9 = 1$. We use the Bayes rule to update our beliefs after observing $Y_9 = 1$:

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_8, Y_9) \quad \propto \quad \mathsf{Pr}(Y_9 \mid \theta)\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_8).$$

The likelihood associated with $Y_9 = 1$ is $\mathsf{Pr}(Y_9 = 1 \mid \theta) = \theta$, hence

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_9) \quad \propto \quad 252\theta^7(1-\theta)^2.$$

The normalizing constant (marginal likelihood) of this posterior distribution is

$$\mathsf{Pr}(Y_9 \mid Y_1, \ldots, Y_8) = \int_0^1 252\theta^7(1-\theta)^2 \, d\theta = \frac{252}{360}.$$

Thus the complete expression for the posterior distribution of θ is:

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_9) = \left(\int_0^1 252\theta^7(1-\theta)^2 \, d\theta\right)^{-1} 252\theta^7(1-\theta)^2 = 360\theta^7(1-\theta)^2, \text{ for } \theta \in (0,1).$$

The posterior mean of θ is calculated as follows

$$\mathsf{E}[\theta \mid Y_1, \ldots, Y_9] = \int_0^1 \theta \mathsf{Pr}(\theta \mid Y_1, \ldots, Y_9) \, d\theta = 0.727.$$

Therefore, after observing SSFSSFSSS, we change our belief about the expected value of θ from 0.7 (prior mean after observing SSFSSFSS) to 0.727 (posterior mean after observing SSFSSFSSS). By comparison, the MLE of θ given $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$, $Y_9 = 1$ is $\frac{7}{9}$ – see Table 1. Compare the shape of the prior and posterior distributions for θ given the observed data $SSFSSFSSS$ in Figure 9.

11 Observed data: SSFSSFSSSF ($Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$, $Y_9 = 1$, $Y_{10} = 0$)

After observing $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$, $Y_9 = 1$ our knowledge about θ is expressed through the distribution

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_9) = 360\theta^7(1-\theta)^2, \text{ for } \theta \in (0,1).$$

This distribution is the prior for θ before observing $Y_{10} = 0$. We use the Bayes rule to update our beliefs after observing $Y_{10} = 0$:

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_9, Y_{10}) \quad \propto \quad \mathsf{Pr}(Y_{10} \mid \theta)\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_9).$$

The likelihood associated with $Y_{10} = 0$ is $\mathsf{Pr}(Y_{10} = 0 \mid \theta) = 1 - \theta$, hence

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_{10}) \quad \propto \quad 360\theta^7(1-\theta)^3.$$

The normalizing constant (marginal likelihood) of this posterior distribution is

$$\mathsf{Pr}(Y_{10} \mid Y_1, \ldots, Y_9) = \int_0^1 360\theta^7(1-\theta)^3 \, d\theta = \frac{360}{1320}.$$

Thus the complete expression for the posterior distribution of θ is:

$$\mathsf{Pr}(\theta \mid Y_1, \ldots, Y_{10}) = \left(\int_0^1 360\theta^7(1-\theta)^3 \, d\theta\right)^{-1} 360\theta^7(1-\theta)^3 = 1320\theta^7(1-\theta)^3, \text{ for } \theta \in (0,1).$$

The posterior mean of θ is calculated as follows

$$\mathsf{E}[\theta \mid Y_1, \ldots, Y_{10}] = \int_0^1 \theta \mathsf{Pr}(\theta \mid Y_1, \ldots, Y_{10}) \, d\theta = 0.667.$$

Therefore, after observing SSFSSFSSSF, we change our belief about the expected value of θ from 0.727 (prior mean after observing SSFSSFSSS) to 0.667 (posterior mean after observing SSFSSFSSSF). By comparison, the MLE of θ given $Y_1 = 1$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 0$, $Y_7 = 1$, $Y_8 = 1$, $Y_9 = 1$, $Y_{10} = 0$ is $\frac{7}{10}$ – see Table 1. Compare the shape of the prior and posterior distributions for θ given the observed data $SSFSSFSSSF$ in Figure 10.

Appendix: R code for generating the figures in this document

```
require(ggplot2)

# Multiple plot function
#
# ggplot objects can be passed in ..., or to plotlist (as a list of ggplot objects)
# - cols:   Number of columns in layout
# - layout: A matrix specifying the layout. If present, 'cols' is ignored.
#
# If the layout is something like matrix(c(1,2,3,3), nrow=2, byrow=TRUE),
# then plot 1 will go in the upper left, 2 will go in the upper right, and
# 3 will go all the way across the bottom.
#
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  require(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)
  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])
  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                      layout.pos.col = matchidx$col))
    }
```

```
  }
}

x = seq(from=0,to=1,length.out=100)
pdf("plotSSFSSFSSSF.pdf",width=9,height=3)
thetaPrior = 0.727
myfunc = function(x) {360*x^7*(1-x)^2}
y = unlist(lapply(x,FUN=myfunc))
f <- ggplot(data.frame(x = x,y = y), aes(x))+
     geom_area(aes(y = y), fill="#FF9999", colour="#FF9999",show_guide=FALSE)+
     geom_line(aes(x = x,y = y),colour="black",size=1)+
     xlab(expression(theta))+
     ylab(expression(paste('P(',theta,')')))+
     ggtitle('PRIOR')+
     geom_vline(xintercept = thetaPrior,color="blue",linetype="longdash",size=1)+
     coord_fixed(ratio = max(x)/max(y))

myfunc = function(x) {1-x}
y = unlist(lapply(x,FUN=myfunc))
g <- ggplot(data.frame(x = x,y = y), aes(x))+
     geom_area( aes(y = y), fill="#FF9999", colour="#FF9999",show_guide=FALSE)+
     geom_line(aes(x = x,y = y),colour="black",size=1)+
     xlab(expression(theta))+
     ylab(expression(paste('P(',theta,')')))+
     ggtitle('LIKELIHOOD')+
     coord_fixed(ratio = max(x)/max(y))

thetaMLE = 7/10
thetaPost = 2/3
myfunc = function(x) {1320*x^7*(1-x)^3}
y = unlist(lapply(x,FUN=myfunc))
h <- ggplot(data.frame(x = x,y = y), aes(x))+
     geom_area( aes(y = y), fill="#FF9999", colour="#FF9999",show_guide=FALSE)+
     geom_line(aes(x = x,y = y),colour="black",size=1)+
     xlab(expression(theta))+
     ylab(expression(paste('P(',theta,')')))+
     ggtitle('POSTERIOR')+
     geom_vline(xintercept = thetaPost,color="blue",linetype="longdash",size=1)+
     geom_vline(xintercept = thetaMLE,color="green",linetype="longdash",size=1)
     coord_fixed(ratio = max(x)/max(y))

multiplot(f,g,h,cols=3)
dev.off()
```

Listing 1: Code for figures