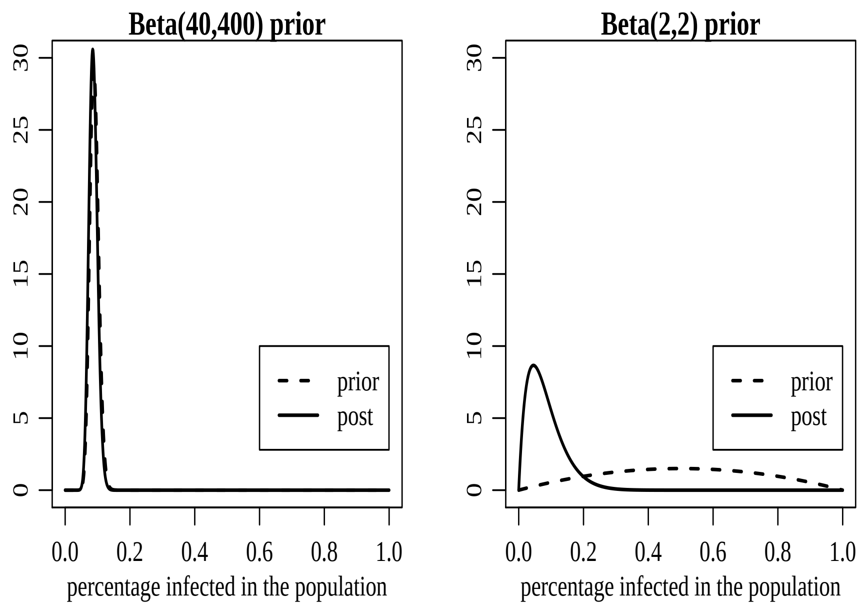
Introduction to Bayesian Data Analysis
Tutorial 1 - Solutions

[see `"Tutorial1.R"` for R code].

(1)



Under the Beta(40,400) prior, the prior and posterior are almost identical, with the mode at approximately 10%. This reflects the strong prior belief that θ=40/(40+400). In contrast, we have collected data on only n=20 units, and this sample size is small in comparison to the amount of prior information, hence the prior dominates the posterior.

Under the Beta(2,2) prior, we see the prior distribution is quite flat. We have some prior information that the infection rate is around 50% but we are quite uncertain about this guess (a weak prior). The data provides us with a lot more information, and we note the considerable change in our beliefs looking at the posterior distribution, which puts approximately zero probability on the infection rate being 50%, and the posterior mode is much less at $\approx$ 5%. Here, the likelihood dominates the posterior.

A Unif(0,1) prior is equivalent to a Beta(1,1) prior. The prior distribution is flat on the interval (0,1). In other words, there is no prior information and the posterior distribution of θ is completely driven by the likelihood.

Under the Unif(0.05,0.20) prior, then $p(\theta) = \frac{1}{0.15} \propto k$ where $\theta \in (0.05, 0.20)$, for some constant k. So

$$p(\theta|y) \propto \theta^y(1-\theta)^{n-y} \text{ for } \theta \in (0.15, 0.20)$$

Let's find the normalising constant c such that $\int_{0.05}^{0.20} c \cdot \theta^y(1-\theta)^{n-y} d\theta = 1$

$$c \times B(y+1, n-y+1) \int_{0.05}^{0.20} \frac{\theta^y (1-\theta)^{n-y}}{B(y+1, n-y+1)} d\theta = 1$$

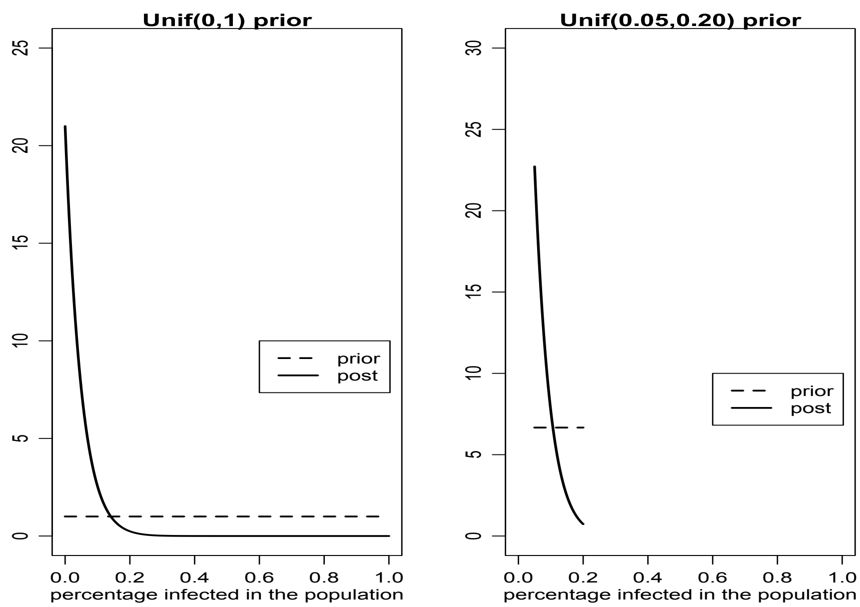
$$c \times B(y+1, n-y+1) \times [F^*(0.20) - F^*(0.05)] = 1$$

where $F^*(x)$ is the CDF of a Beta(y+1, n-y+1) distribution evaluated at x. Therefore

$$c = \frac{1}{B(y+1, n-y+1) \times [F^*(0.20) - F^*(0.05)]}$$

and

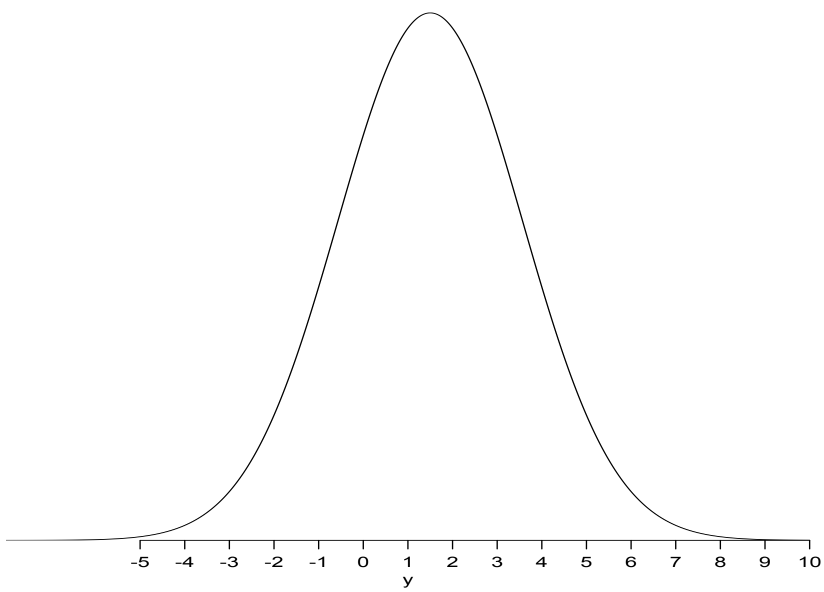
$$p(\theta|y) = \frac{\theta^y (1-\theta)^{n-y}}{B(y+1, n-y+1) \times [F^*(0.20) - F^*(0.05)]} \text{ for } \theta \in (0.05, 0.20)$$



(2) Conditional probability: suppose that if $\theta = 1$, the y has a normal distribution with mean 1 and standard deviation σ, and if $\theta = 2$, then y has a normal distribution with mean 2 and standard deviation σ. Also, suppose $\Pr(\theta = 1)$=0.5 and $\Pr(\theta = 2)$=0.5

(a)

$$\begin{aligned} p(y|\sigma^2 = 2^2) &= \sum_\theta p(y, \theta|\sigma^2 = 2^2) \\ &= \sum_\theta p(y|\theta, \sigma^2 = 2^2)p(\theta) \\ &= Pr(\theta = 1)p(y|\theta = 1, \sigma^2 = 2^2) + Pr(\theta = 2)p(y|\theta = 2, \sigma^2 = 2^2) \\ &= 0.5N(y|\theta = 1, \sigma^2 = 2^2) + 0.5N(y|\theta = 2, \sigma^2 = 2^2) \end{aligned}$$

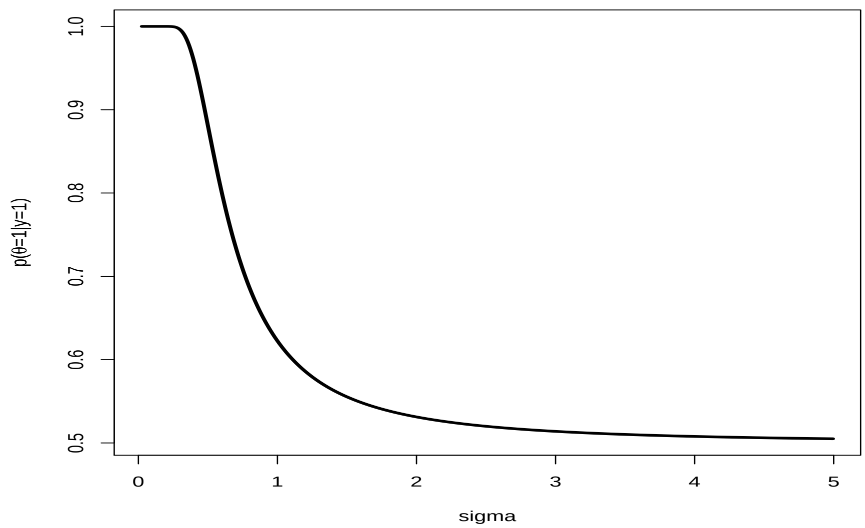


(b)

$$\begin{aligned} p(\theta = 1|y = 1) &= \frac{p(y = 1, \theta = 1)}{p(y = 1)} \\ &= \frac{0.5N(y = 1|\theta = 1, \sigma^2 = 2^2)}{0.5N(y|\theta = 1, \sigma^2 = 2^2) + 0.5N(y|\theta = 2, \sigma^2 = 2^2)} \\ &= 0.5312094 \end{aligned}$$

Note that the posterior probability that $\theta = 1$ is slightly higher than 0.5, given that we observed $y = 1$.

(c) The posterior density $Pr(\theta = 1|y = 1)$ decreases towards the prior probability ($Pr(\theta = 1) = 0.5$) as σ increases. That is, as the uncertainty in the observed data increases, this reduces the evidence to support the value $\theta = 1$ even though we observed $y = 1$, because we are assuming more variability in the sampling model.



(3) (a) The posterior distribution is $g(p=k|y) = \frac{g(y|p=k)g(p=k)}{g(y)}$.
We have:

$$\begin{aligned} g(y) &= \sum_p g(y|p)g(p) \\ &= \binom{60}{55} 1^{55} 0^5 \times 0.8 + \binom{60}{55} 0.975^{55} 0.025^5 \times 0.1 + \binom{60}{55} 0.95^{55} 0.05^5 \times 0.05 + \\ &\quad + \binom{60}{55} 0.925^{55} 0.075^5 \times 0.035 + \binom{60}{55} 0.90^{55} 0.10^5 \times 0.015 \end{aligned}$$

So $g(p=0|y) = \frac{\binom{60}{55} 1^{55} 0^5 \times 0.8}{g(y)} = 0$ and $g(p=0.10|y) = \frac{\binom{60}{55} 0.9^{55} 0.1^5 \times 0.015}{g(y)} = 0.165$.
That is, the posterior probability that all passengers show up is zero, and the posterior probability that 10% of passengers do not show up is 0.165.

(b) We need to sequentially update the posterior after the data from each flight. (Run a loop function in R). Each time, the prior is updated to be the posterior from the previous cycle, as in the following code:

```
g_p<-c(0.80, 0.10,0.05,0.035,0.015)
p<-c(0,0.025,0.05,0.075,0.10)

for (i in 1:10){
  g.post<-dbinom(y[i],n,1-p)*g_p
  g.post<-g.post/sum(g.post)
  g_p<-g.post #updated posterior becomes new prior
> g.post
[1] 0.000000e+00 1.261763e-07 1.407355e-01 8.445976e-01 1.466673e-02
}
```

The posterior mode is 7.5% of passengers do not show up.

(c) If the airline company maintains the same prior distribution $g(p)$, then the company would be overestimating the probability that all customers show up (p=0). By engaging in the practice of overbooking seats, given the past data above, the company can boost profits because the no-show rate is higher than that assumed in the prior.