Homework 3

Brian Cervantes Alvarez
June 2, 2024
ST 559 Bayesian Statistics

Question 1

Part A

Given the priors $\theta_A \sim \text{Gamma}(120, 10)$ and $\theta_B \sim \text{Gamma}(12, 1)$, we have strong prior information about θ_A because of the larger sample size and previous studies. On the other hand, the prior for θ_B is more uncertain, reflecting that we know less about strain B.

If we think that the tumor rates in strain B are influenced by those in strain A, we might consider a joint prior distribution that shows some kind of dependence between θ_A and θ_B. However, without clear evidence of such a relationship, it is reasonable to assume they are independent to keep things simple.

So, assuming $p(\theta_A, \theta_B) = p(\theta_A)p(\theta_B)$ makes sense without specific evidence showing a dependence between the two populations.

Part B

To calculate $\Pr(\theta_B < \theta_A \mid \mathbf{y}_A, \mathbf{y}_B)$, we can apply Monte Carlo sampling. The posterior distributions for θ_A and θ_B are given,

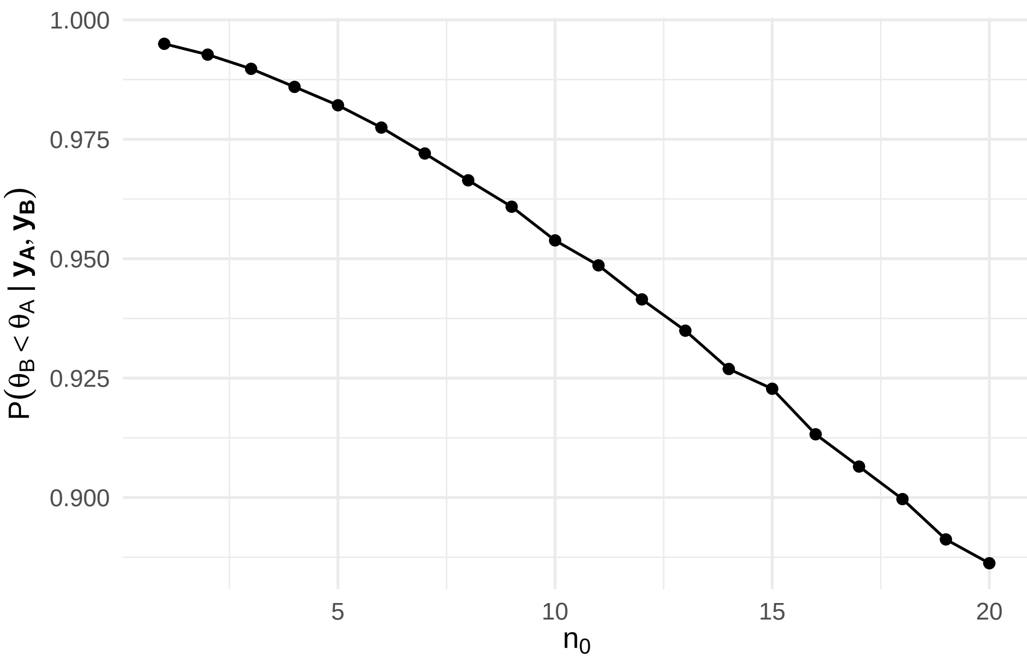
$$\theta_A \mid \mathbf{y}_A \sim \text{Gamma}(120 + \sum y_A, 10 + n_A) = \text{Gamma}(120 + 117, 10 + 10) = \text{Gamma}(237, 20)$$

$$\theta_B \mid \mathbf{y}_B \sim \text{Gamma}(12 + \sum y_B, 1 + n_B) = \text{Gamma}(12 + 113, 1 + 13) = \text{Gamma}(125, 14)$$

Next, we can draw samples from these posterior distributions and compute the proportion of samples where $\theta_B < \theta_A$. Using R, we can compute the simulation and get 0.9996

Part C

We perform a sensitivity analysis by varying the hyperparameter n_0 in the prior for θ_B. To be precise, we compute $\Pr(\theta_B < \theta_A \mid \mathbf{y}_A, \mathbf{y}_B)$ for different values of n_0 while keeping the prior for θ_A fixed. The prior for θ_B is updated to $\theta_B \sim \text{Gamma}(12 \times n_0, n_0)$. We evaluate n_0 ranging from 1 to 20 and plot the figure below.



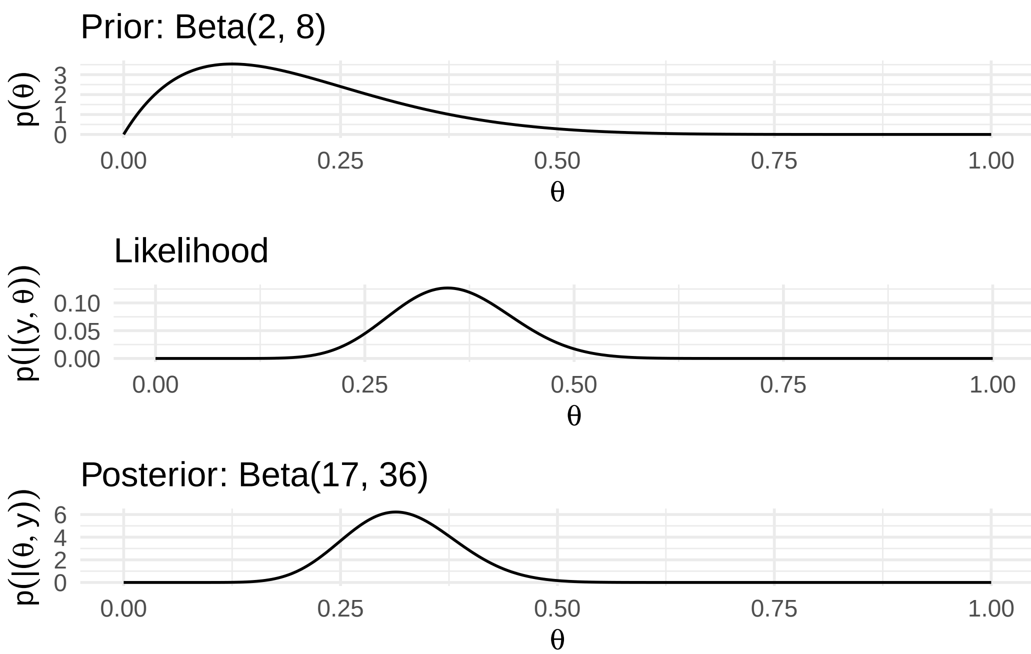
The sensitivity plot shows how $P(\theta_B < \theta_A \mid \mathbf{y}_A, \mathbf{y}_B)$ varies with different values of n_0. Hence, helps us understand conceptually that the choice of the hyperparameter in the prior distribution for θ_B matters.

Question 2

Part A

Table 1: Posterior statistics for Beta(2, 8) prior

Statistic	Value
Posterior Mean	0.3207547
Posterior Mode	0.3137255
Posterior SD	0.0635189
95% CI Lower	0.2032978
95% CI Upper	0.4510240

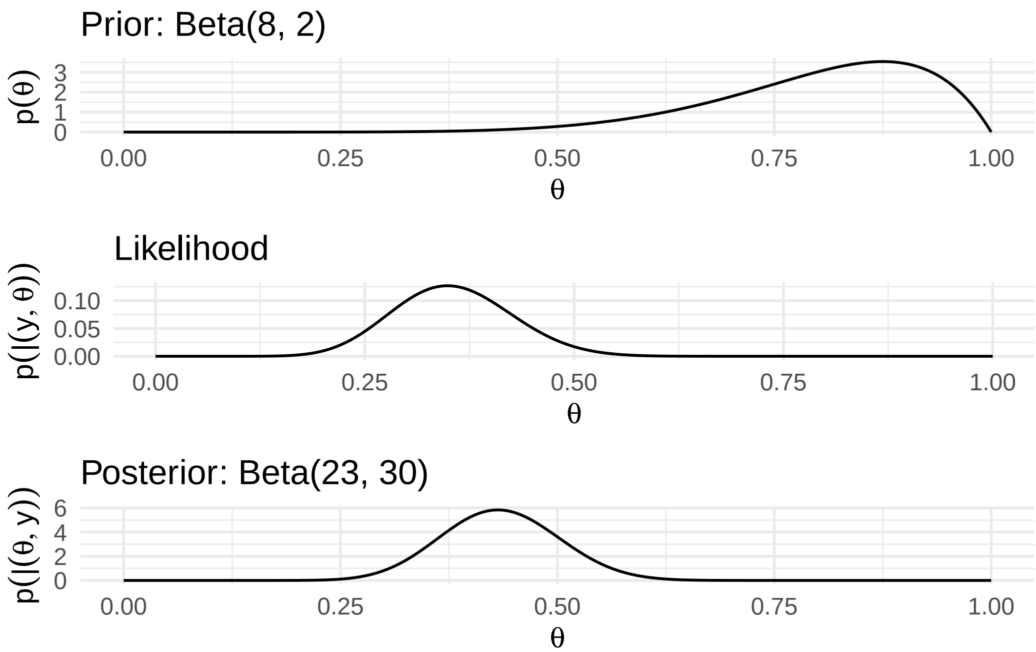


This prior suggests that the probability of teens reoffending after being released is quite low. It reflects a strong belief that most teens will not reoffend based on previous data or studies. Essentially, we are starting with the assumption that reoffending is unlikely.

Part B

Table 2: Posterior statistics for Beta(8, 2) prior

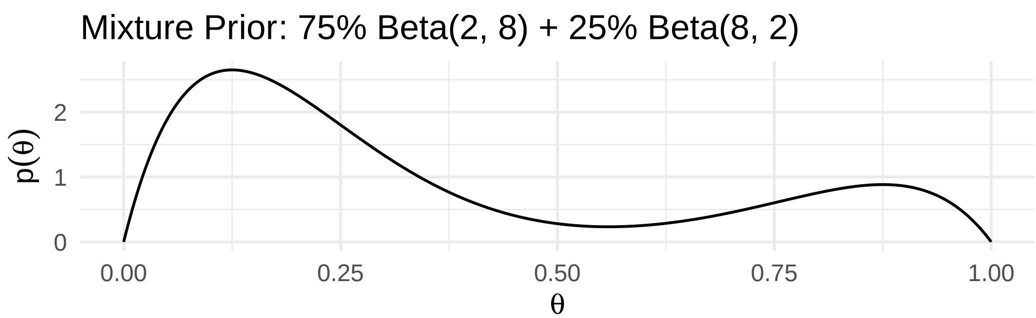
Statistic	Value
Posterior Mean	0.4339622
Posterior Mode	0.4313725
Posterior SD	0.0674453
95% CI Lower	0.3046956
95% CI Upper	0.5679528

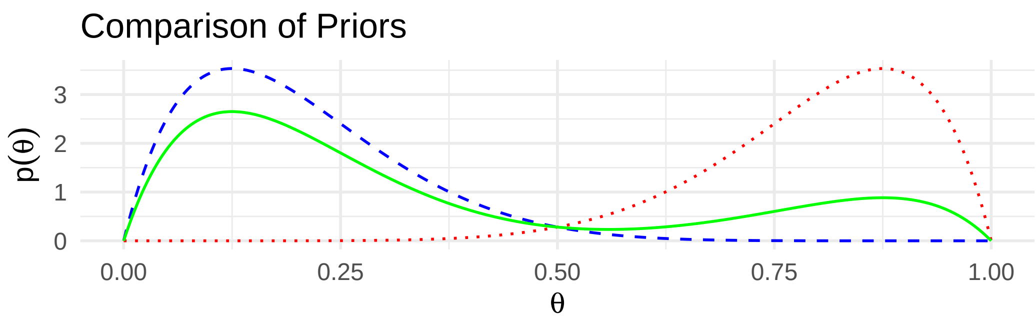


This prior indicates that the probability of teens reoffending is high. It shows a belief that a significant number of teens will reoffend after being released. We are beginning with the assumption that reoffending is more likely, possibly based on different past data or studies.

Part C

$$p(\theta) = \frac{1}{4}\frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}\left[3\theta(1-\theta)^7 + \theta^7(1-\theta)\right]$$





This prior combines two different beliefs: One that reoffending is unlikely and another that it is likely. It leans more towards the belief that reoffending is unlikely but still considers the possibility that it could be likely. This mixture approach represents a balanced view, taking into account both low and high chances of reoffending based on mixed evidence or diverse opinions.