Q1

The likelihood function conditional on the data is:

$$\mathbb{P}(\mathbf{X}|p) = p^{\mathbf{1}^T\mathbf{X}}(1-p)^{N-\mathbf{1}^T\mathbf{X}}$$

Take the log-likelihood of the function and obtain the MLE of p^*:

$$l(p) = \mathbf{1}^T\mathbf{X}\log(p) + (N - \mathbf{1}^T\mathbf{X})\log(1-p)$$

$$\frac{\partial l}{\partial p} = \frac{\mathbf{1}^T\mathbf{X}}{p} - \frac{N-\mathbf{1}^T\mathbf{X}}{1-p} = 0$$

$$p^*N = \mathbf{1}^T\mathbf{X}$$

$$p^*_{MLE} = \frac{\mathbf{1}^T\mathbf{X}}{N} = \frac{1}{N}\sum_{i=1}^N x_i$$

Q2

The likelihood function of posterior is:

$$\mathbb{P}(p|\mathbf{X}) \propto \mathbb{P}(\mathbf{X}|p)\mathbb{P}(p) \propto p^{\mathbf{1}^T\mathbf{X}+\alpha-1}(1-p)^{N-\mathbf{1}^T\mathbf{X}+\beta-1}$$

We assume that $\alpha > \beta > 1$. Therefore, the posterior thus follows Beta distribution: Beta($\mathbf{1}^T\mathbf{X}+\alpha$, $N-\mathbf{1}^T\mathbf{X}+\beta$)

To maximize the posterior likelihood, we take the mode of the beta distribution:

$$p^*_{MAP} = \frac{\mathbf{1}^T\mathbf{X}+\alpha-1}{N+\alpha+\beta-2}.$$

Q3

(a)

Notice that, in expectation, $\mathbf{1}^T\mathbf{X} = Np^* = 0.99N$. However, each $x \in \{0,1\}$, so if $N < 100$, then $\hat{p}_{MLE} = 1$. If $N = 100$, in expectation, $\hat{p}_{MLE} = 0.99$, which is within 0.01 of p^*.

(b)

Take $\alpha = 7, \beta = 2$. Let $p^*_{MAP} = \frac{\mathbf{1}^T\mathbf{X}+\alpha-1}{N+\alpha+\beta-2} > 0.98$, we can obtain $N = 44$.

Q4

(a)

Use the similar arguments as Q3, we can see that $\hat{p}_{MLE}$ does not depend on the prior belief. Therefore, we obtain the same answer: $N = 100$.

(b)

Take $\alpha = 7, \beta = 2$. Let $p^*_{MAP} = \frac{\mathbf{1}^T\mathbf{X}+\alpha-1}{N+\alpha+\beta-2} < 0.02$, we can obtain $N = 294$.

Q5

In conclusion, when we do have the correct prior, $\hat{p}_{MAP}$ converges much faster to the true p^* than $\hat{p}_{MLE}$. However, if the prior is incorrect, then it takes extra samples to correct the prior belief. Therefore, $\hat{p}_{MLE}$ performs better.