36-705 Intermediate Statistics Homework #5
Solutions

October 20, 2018

Problem 1

(a) Since there is only one parameter to estimate and \( E[X_i] = \frac{1}{p} \), the method of moments estimator \( \hat{p}_{MOM} \) is given by:

\[
\frac{1}{\hat{p}_{MOM}} = \frac{1}{n} \sum_{i=1}^{n} X_i \implies \hat{p}_{MOM} = \frac{n}{\sum_{i=1}^{n} X_i}.
\]

To find the maximum likelihood estimator \( \hat{p}_{mle} \) calculate the likelihood function:

\[
L(p) = \prod_{i=1}^{n} p(1 - p)^{X_i-1} = p^n (1 - p)^{\sum_{i=1}^{n} X_i - n}.
\]

Hence the log likelihood function is:

\[
l_n(p) = n \log(p) + \left( \sum_{i=1}^{n} X_i - n \right) \log(1 - p).
\]

Its first derivative equals to:

\[
\frac{\partial}{\partial p} l_n(p) = \frac{n}{p} - \frac{\sum_{i=1}^{n} X_i - n}{1 - p}.
\]

Setting it equal to 0 we get the unique solution,

\[
\frac{n}{p} - \frac{\sum_{i=1}^{n} X_i - n}{1 - p} = 0 \implies \frac{p}{n} = \frac{1 - p}{\sum_{i=1}^{n} X_i - n} \implies p \sum_{i=1}^{n} X_i = n \implies \hat{p}_{mle} = \frac{n}{\sum_{i=1}^{n} X_i}.
\]

This is a global maximum because the second derivative is negative for every \( p \)

\[
\frac{\partial^2}{\partial p^2} l_n(p) = -\frac{n}{p^2} - \frac{\sum_{i=1}^{n} X_i - n}{(1 - p)^2} < 0.
\]

(b) Notice that the \( \hat{p}_{MOM} \) and \( \hat{p}_{mle} \) are convex functions of \( \sum_{i=1}^{n} X_i \). Therefore using Jensen’s inequality,

\[
E[\hat{p}_{mle}] = E[\hat{p}_{MOM}] \geq \frac{n}{E[\sum_{i=1}^{n} X_i]} = p.
\]

Since the estimators are strictly convex functions of \( \sum_{i=1}^{n} X_i \), the inequality is also strict. Hence, both the estimators are biased.
Problem 2

Since there are two parameters to estimate and

\[ \mathbb{E}[X_i] = \frac{a + b}{2}, \]

\[ \mathbb{E}[X_i^2] = \text{Var}(X_i) + \mathbb{E}[X_i]^2 = \frac{(b - a)^2}{12} + \frac{(b + a)^2}{4}. \]

Hence the method of moments estimators \( \hat{a}_{MOM} \) and \( \hat{b}_{MOM} \) are given by:

\[ \hat{a}_{MOM} + \hat{b}_{MOM} = 2 \frac{1}{n} \sum_{i=1}^{n} X_i = 2 \bar{X}, \]

\[ (\hat{b}_{MOM} - \hat{a}_{MOM})^2 = 12 \left( \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 \right) = \frac{12}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2. \]

Therefore,

\[ \hat{b}_{MOM} = \frac{1}{2} \left( 2\bar{X} + \sqrt{\frac{12}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2} \right) = \bar{X} + \sqrt{\frac{3}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}, \]

\[ \hat{a}_{MOM} = 2\bar{X} - \hat{b}_{MOM} = \bar{X} - \sqrt{\frac{3}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}. \]

To find the maximum likelihood estimators \( \hat{a}_{mle} \) and \( \hat{b}_{mle} \) calculate the likelihood function:

\[ \mathcal{L}(a, b) = \prod_{i=1}^{n} \frac{1}{b - a} I\{a \leq X_i \leq b\} = \frac{1}{(b - a)^n} I\{a \leq X_{(1)}, b \geq X_{(n)}\}, \]

where \( X_{(1)} = \min_i X_i \) and \( X_{(n)} = \max_i X_i \). This function continuously increases as \( a \) increases to \( X_{(1)} \) and is then equal to zero for \( a \in (X_{(1)}, \infty) \). This means that the supremum occurs at \( a = X_{(1)} \). Similarly, the likelihood function is zero for \( b \in (-\infty, X_{(n)}) \), then at \( X_{(n)} \) jumps to a non-zero positive quantity and decreases asymptotically to zero as \( b \to \infty \). Hence the supremum occurs at \( b = X_{(n)} \). That is, the MLE is given by \( \hat{a} = X_{(1)} \) and \( \hat{b} = X_{(n)} \).

Problem 3

(a) Since there is only one parameter to estimate and \( \mathbb{E}[X_1] = \lambda \), the method of moments estimator \( \hat{\lambda}_m \) is given by:

\[ \hat{\lambda}_m = \frac{1}{n} \sum_{i=1}^{n} X_i \]

To find the maximum likelihood estimator \( \hat{\lambda}_{mle} \) calculate the log likelihood function:

\[ l_n(\lambda) = -n\lambda + \sum_{i=1}^{n} X_i \log \lambda - \sum_{i=1}^{n} \log X_i! \]

Its first derivative equals to:

\[ \frac{\partial}{\partial \lambda} l_n(\lambda) = -n + \frac{1}{\lambda} \sum_{i=1}^{n} X_i \]
Setting it equal to 0 we get the unique solution \( \hat{\lambda}_{mle} = \frac{1}{n} \sum_{i=1}^{n} X_i \). This is a global maximum because the second derivative is negative for every \( \lambda \)

\[
\frac{\partial^2}{\partial \lambda^2} l_n(\lambda) = -\frac{1}{\lambda^2} \sum_{i=1}^{n} X_i < 0
\]

To get the Fisher information we use the second derivative to get:

\[
\mathcal{I}_n(\lambda) = -\mathbb{E} \left[ \frac{\partial^2}{\partial \lambda^2} l_n(\lambda) \right] = \frac{1}{\lambda^2} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{\lambda}.
\]

(b) By \( E(X) = \lambda \), we get estimator,

\[
\hat{\lambda}_m = \frac{1}{n} \sum_{i=1}^{n} X_i.
\]

By \( V(X) = E(X^2) - (E(X))^2 = \lambda \), we get an unbiased estimator,

\[
\hat{\lambda}_v = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.
\]

Note that we use factor \( \frac{1}{n-1} \) instead of \( \frac{1}{n} \) so that the estimator is unbiased.

The variance for \( \hat{\lambda}_m \) is given by,

\[
V(\hat{\lambda}_m) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{\lambda}{n}.
\]

To get the variance for \( \hat{\lambda}_v \), we use the following two facts,

1. the variance for a sample variance is given by

\[
Var(S_n) = \frac{\mu_4}{n} - \frac{\mu_2^2(n-3)}{n(n-1)},
\]

where \( \mu_k \) is the \( k \)-the center moment \( E(X - E(X))^k \).

2. For Poisson distribution, \( \mu_2 = \lambda, \mu_4 = \lambda(1+3\lambda) \).

So the variance for \( \hat{\lambda}_v \) is given by,

\[
V(\hat{\lambda}_v) = \frac{\mu_4}{n} - \frac{\mu_2^2(n-3)}{n(n-1)} = \frac{\lambda(1+3\lambda)}{n} - \frac{\lambda^2(n-3)}{n(n-1)} = \frac{\lambda}{n} + \frac{2\lambda^2}{(n-1)},
\]

which is strictly bigger than \( V(\hat{\lambda}_m) \).
Problem 4

(a) To find the maximum likelihood estimator \( \hat{\theta}_{mle} \) calculate the likelihood function:

\[
L(\theta) = \prod_{i=1}^{n} p(X_i; \theta) = \prod_{i=1}^{n} \theta^{-2} I\{\theta \leq X_i\} = \theta^n \prod_{i=1}^{n} X_i^{-2} I\{\theta \leq X_{(1)}\},
\]

where \( X_{(1)} = \min_i X_i \). The likelihood function continuously increases as \( \theta \) increases to \( X_{(1)} \) and is then drops to zero for \( \theta \in (X_{(1)}, \infty) \). This means that the supremum occurs at \( \theta = X_{(1)} \). Therefore the MLE is given by, \( \hat{\theta}_{mle} = X_{(1)} \).

(b) Since there is only one parameter to estimate and

\[
E[X_i] = \int_{\theta}^\infty x \theta x^{-2} dx = \theta \int_{\theta}^{\infty} x^{-1} dx = \infty,
\]

that is it does not exist. Hence the method of moments estimator also does not exist.

Problem 5

(a) \( X \sim \text{Poisson}(\lambda), \lambda \sim \text{Gamma}(\alpha, \beta) \).

\[
p(x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad \pi(\lambda) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda/\beta}.
\]

Therefore,

\[
\pi(\lambda|x) \propto \frac{e^{-\lambda} \lambda^x}{x!} \frac{1}{\beta^\alpha \Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda/\beta} \propto e^{-\lambda(1+\frac{1}{\beta})} \lambda^{x+\alpha-1} = e^{-\lambda \frac{\alpha+1}{\beta}} \lambda^{x+\alpha-1}.
\]

Therefore, \( \lambda|x \sim \text{Gamma}(x + \alpha, \frac{\beta}{\beta+1}) \). Since we are considering the squared loss , by Theorem 16.7 in Lecture 16, the Bayes estimator is given by,

\[
\hat{\lambda} = E[\lambda|X] = \frac{\beta}{\beta+1} (X + \alpha).
\]

Remember from Lecture notes 16 that the posterior risk of an estimator is given by,

\[
r(\hat{\lambda}|x) = \int L(\lambda, \hat{\lambda}) \pi(\lambda|x) d\lambda = \int (\lambda - \hat{\lambda})^2 \pi(\lambda|x) d\lambda = \text{Var}(\lambda|x),
\]

since \( \hat{\lambda} = E[\lambda|X] \).

Now, the Bayes risk is given by,

\[
B_\pi(\hat{\lambda}) = \int r(\hat{\lambda}|x) m(x) dx = \int r(\hat{\lambda}|x) \int p(x|\lambda) \pi(\lambda) d\lambda dx = E_{\pi(\lambda)}[E_X[\text{Var}(\lambda|X)]].
\]
Therefore, the Bayes risk in this case is,
\[
\mathbb{E}_\pi(\lambda) \left[ \mathbb{E}_X \left[ \mathbb{E} \left[ (\lambda - \hat{\lambda})^2 \mid X \right] \right] \right] = \mathbb{E}_\pi(\lambda) \left[ \mathbb{E}_X \left[ \text{Var}(\lambda \mid X) \right] \right] \\
= \mathbb{E}_\pi(\lambda) \left[ \mathbb{E}_X \left[ (X + \alpha) \left( \frac{\beta}{\beta + 1} \right)^2 \right] \right] \\
= \mathbb{E}_\pi(\lambda) \left[ (\lambda + \alpha) \left( \frac{\beta}{\beta + 1} \right)^2 \right] \\
= \left( \alpha \beta + \alpha \right) \left( \frac{\beta}{\beta + 1} \right)^2 = \frac{\alpha \beta^2}{\beta + 1}.
\]

Alternate way of finding the risk: We could also find the Bayes risk by computing,
\[
B_\pi(\hat{\lambda}) = \mathbb{E}_\pi(\lambda) \left[ R(\lambda, \hat{\lambda}) \right].
\]

In this case,
\[
R(\lambda, \hat{\lambda}) = \mathbb{E} \left[ (\lambda - \hat{\lambda})^2 \right] = \mathbb{E} \left[ \left( \frac{\beta}{\beta + 1} (X + \alpha) - \lambda \right)^2 \right] \\
= \mathbb{E} \left[ \left( \frac{\beta}{\beta + 1} (X - \lambda) - \frac{1}{\beta + 1} (\lambda - \alpha \beta) \right)^2 \right] \\
= \left( \frac{\beta}{\beta + 1} \right)^2 \text{Var}(X) - 2 \frac{\beta}{\beta + 1} \frac{1}{\beta + 1} (\lambda - \alpha \beta) \mathbb{E}[X - \lambda] + \frac{1}{(\beta + 1)^2} (\lambda - \alpha \beta)^2 \\
= \left( \frac{\beta}{\beta + 1} \right)^2 \lambda + \frac{1}{(\beta + 1)^2} (\lambda - \alpha \beta)^2.
\]

Therefore,
\[
B_\pi(\hat{\lambda}) = \mathbb{E}_\pi(\lambda) \left[ R(\lambda, \hat{\lambda}) \right] \\
= \left( \frac{\beta}{\beta + 1} \right)^2 \mathbb{E}_\pi(\lambda) \left[ \lambda \right] + \frac{1}{(\beta + 1)^2} \mathbb{E}_\pi(\lambda) \left[ (\lambda - \alpha \beta)^2 \right] \\
= \left( \frac{\beta}{\beta + 1} \right)^2 \alpha \beta + \frac{1}{(\beta + 1)^2} \text{Var}(\lambda) \\
= \left( \frac{\beta}{\beta + 1} \right)^2 \alpha \beta + \frac{1}{(\beta + 1)^2} \alpha \beta^2 \\
= \frac{\alpha \beta^3 + \alpha \beta^2}{(\beta + 1)^2} = \frac{\alpha \beta^2}{\beta + 1}.
\]

(b) \( X \sim N(\theta, \sigma^2) \), where \( \sigma^2 \) is known and \( \theta \sim N(a, b^2) \).
\[
p(x \mid \theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} (x-\theta)^2}, \quad \pi(\theta) = \frac{1}{\sqrt{2\pi} b} e^{-\frac{1}{2b^2} (\theta-a)^2}.
\]
Therefore,
\[
\pi(\theta|x) \propto \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\theta)^2} \frac{1}{\sqrt{2\pi}b} e^{-\frac{1}{2b^2}(\theta-a)^2}
\]
\[
\propto \exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{1}{\sigma^2} + \frac{1}{b^2}\right) - 2\theta\left(\frac{x}{\sigma^2 + \frac{a}{b^2}}\right)\right]\right\}
\]
\[
\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2} + \frac{1}{b^2}\right)\left(\theta - \frac{x}{\sigma^2 + \frac{a}{b^2}}\right)^2\right\}.
\]

Therefore, \(\theta|x \sim N\left(\frac{x\frac{b^2}{\sigma^2 + b^2} + a\frac{\sigma^2}{\sigma^2 + b^2}}{\frac{\sigma^2}{\sigma^2 + b^2} + \frac{b^2}{\sigma^2 + b^2}}, \frac{\sigma^2}{\frac{\sigma^2}{\sigma^2 + b^2} + \frac{b^2}{\sigma^2 + b^2}}\right)\). Since we are considering the squared loss, by Theorem 16.7 in Lecture 16, the Bayes estimator is given by,
\[
\hat{\theta} = \mathbb{E}[\theta|X] = X \frac{b^2}{\sigma^2 + b^2} + a \frac{\sigma^2}{\sigma^2 + b^2}.
\]

Similar to the previous case, we can find the Bayes risk by considering the posterior variance of \(\lambda|X\) which is given by,
\[
\mathbb{E}\left[(\theta - \hat{\theta})^2|X\right] = \text{Var}(\theta|X) = \frac{\sigma^2b^2}{\sigma^2 + b^2}.
\]

Therefore, the Bayes risk is given by,
\[
B_\pi(\hat{\theta}) = \mathbb{E}_\pi(\theta) \left[\mathbb{E}_X\left[\text{Var}(\theta|X)\right]\right] = \mathbb{E}_\pi(\theta) \left[\mathbb{E}_X\left[\frac{\sigma^2b^2}{\sigma^2 + b^2}\right]\right] = \frac{\sigma^2b^2}{\sigma^2 + b^2}.
\]

**Alternate way of finding the risk:** We could also find the Bayes risk by computing,
\[
B_\pi(\hat{\theta}) = \mathbb{E}_\pi(\theta) \left[R(\theta, \hat{\theta})\right].
\]

In this case,
\[
R(\theta, \hat{\theta}) = \mathbb{E}\left[\left(\frac{X - \frac{b^2}{\sigma^2 + b^2} + a\frac{\sigma^2}{\sigma^2 + b^2} - \theta}{\frac{\sigma^2}{\sigma^2 + b^2} + \frac{b^2}{\sigma^2 + b^2}}\right)^2\right]
\]
\[
= \mathbb{E}\left[\left((X - \theta) \frac{b^2}{\sigma^2 + b^2} + (a - \theta) \frac{\sigma^2}{\sigma^2 + b^2}\right)^2\right]
\]
\[
= \text{Var}(X) \left(\frac{b^2}{\sigma^2 + b^2}\right)^2 + (a - \theta)^2 \left(\frac{\sigma^2}{\sigma^2 + b^2}\right)^2
\]
\[
= \sigma^2 \left(\frac{b^2}{\sigma^2 + b^2}\right)^2 + (a - \theta)^2 \left(\frac{\sigma^2}{\sigma^2 + b^2}\right)^2.
\]
Therefore the Bayes risk is given by,

\[ B_\pi(\hat{\theta}) = \mathbb{E}_\pi(\theta) \left[ R(\theta, \hat{\theta}) \right] \]

\[ = \mathbb{E}_\pi(\theta) \left[ \sigma^2 \left( \frac{b^2}{\sigma^2 + b^2} \right)^2 + (a - \theta)^2 \left( \frac{\sigma^2}{\sigma^2 + b^2} \right)^2 \right] \]

\[ = \sigma^2 \left( \frac{b^2}{\sigma^2 + b^2} \right)^2 + \mathbb{E}_\pi(\theta) \left[ (a - \theta)^2 \right] \left( \frac{\sigma^2}{\sigma^2 + b^2} \right)^2 \]

\[ = \sigma^2 \left( \frac{b^2}{\sigma^2 + b^2} \right)^2 + b^2 \left( \frac{\sigma^2}{\sigma^2 + b^2} \right)^2 = \frac{\sigma^2 b^2}{\sigma^2 + b^2}. \]

Problem 6

(a) \( X_1, \ldots, X_n \sim N(\mu, 1) \). To find the Cramer-Rao lower bound, we first find the likelihood function.

\[ L(\mu) = \left( \frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum_{i=1}^{n} (X_i - \mu)^2}. \]

We then calculate the log likelihood function:

\[ l_n(\mu) = -n \log \left( \sqrt{2\pi} \right) - \frac{1}{2} \sum_{i=1}^{n} (X_i - \mu)^2. \]

Its first derivative equals to:

\[ \frac{\partial}{\partial \mu} l_n(\mu) = \sum_{i=1}^{n} (X_i - \mu). \]

Since we want the Fisher’s information for \( \mu^2 \), we find

\[ \frac{\partial}{\partial \mu^2} l_n = \frac{\partial}{\partial \mu} \frac{\partial}{\partial \mu} l_n = \frac{\sum_{i=1}^{n} (X_i - \mu)}{2\mu} \]

Hence we get the Fisher information as:

\[ I_n(\mu^2) = \mathbb{E} \left[ \left( \frac{\sum_{i=1}^{n} (X_i - \mu)}{2\mu} \right)^2 \right] = \frac{1}{4\mu^2} \mathbb{E} \left[ \left( \sum_{i=1}^{n} X_i - n\mu \right)^2 \right] = \frac{1}{4\mu^2} \text{Var} \left( \sum_{i=1}^{n} X_i \right) = \frac{n}{4\mu^2}. \]

Therefore the Cramer Rao lower bound is given by, \( \frac{4\mu^2}{n} \), that is,

\[ \text{Var} \left( \hat{\mu}^2 \right) \geq \frac{4\mu^2}{n}. \]

(b) Let \( \hat{\mu}^2 = \overline{X}_n^2 - \frac{1}{n} \). Then,

\[ \text{Var} \left( \hat{\mu}^2 \right) = \text{Var} \left( \overline{X}_n^2 - \frac{1}{n} \right) = \text{Var} \left( \overline{X}_n^2 \right) = \mathbb{E} \left[ \overline{X}_n^4 \right] - \left( \mathbb{E} \left[ \overline{X}_n^2 \right] \right)^2. \]
Now
\[ \mathbb{E}\left[ X_n^2 \right] = \mathbb{E}\left[ X_n (X_n - \mu) \right] + \mu^2 = \frac{1}{n} \mathbb{E}[1] + \mu^2 = \frac{1}{n} + \mu^2. \]

In the second equality we used Stein's identity with the random variable \( X_n \sim N(\mu, 1/n) \) and the function \( g(X_n) = X_n \).

In the derivation of \( \mathbb{E}\left[ X_n^4 \right] \) below, we are going to use Stein's identity again with the random variable \( X_n \sim N(\mu, 1/n) \) and the functions \( X_n^3 \) and \( X_n^4 \). So,
\[
\begin{align*}
\mathbb{E}\left[ X_n^4 \right] &= \mathbb{E}\left[ X_n^3 (X_n - \mu) \right] + \mu \mathbb{E}\left[ X_n^3 \right] \\
&= \frac{1}{n} \mathbb{E}\left[ 3X_n^2 \right] + \mu \mathbb{E}\left[ X_n (X_n - \mu) \right] + \mu^2 \mathbb{E}\left[ X_n^2 \right] \\
&= \frac{3}{n} \left( \frac{1}{n} + \mu^2 \right) + \mu \left( \frac{1}{n} \mathbb{E}[2X_n] + \mu^2 \left( \frac{1}{n} + \mu^2 \right) \right) \\
&= \frac{3}{n} \left( \frac{1}{n} + \mu^2 \right) + \frac{2}{n} \mu^2 + \mu^2 \left( \frac{1}{n} + \mu^2 \right) \\
&= \frac{6\mu^2}{n} + \mu^4 + \frac{3}{n^2}.
\end{align*}
\]

So,
\[
\text{Var}(\hat{\mu}) = \frac{6\mu^2}{n} + \mu^4 + \frac{3}{n^2} - \left( \frac{1}{n} + \mu^2 \right)^2 = \frac{4\mu^2}{n} + \frac{2}{n^2} > \frac{4\mu^2}{n}.
\]

This shows that the estimator has a higher variance than the Cramer-Rao lower bound.

**Problem 7**

An estimator as a function of \( X \), \( \hat{\theta} = \hat{\theta}(X) \), is unbiased if,
\[
\frac{1}{p} = E(\hat{\theta}), \quad \text{for all } 0 < p < 1.
\]

Write out expectation explicitly, and we have,
\[
\frac{1}{p} = E(\hat{\theta}(X)) = \sum_{k=0}^{n} \hat{\theta}(X = k) P(X = k) = \sum_{k=0}^{n} \hat{\theta}(k) \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for all } 0 < p < 1.
\]

However, this equation cannot hold for all \( 0 < p < 1 \). Because as \( p \) goes to zero, the left side \( \frac{1}{p} \) is going to infinity, whereas the right side is going to zero.

**Problem 8**

We have that \( X_1, \ldots, X_n \sim \text{Ber}(p) \) and the prior is given by \( p \sim \text{Beta}(1, b) \). That is, we have,
\[
p(X_1, \ldots, X_n; p) = \prod_{i=1}^{n} p^{X_i} (1-p)^{1-X_i}, \quad \text{and } \pi(p) = \frac{(1-p)^{b-1}}{B(1, b)},
\]

8
where $B(1,b) = \frac{\Gamma(1) \Gamma(b)}{\Gamma(b+1)}$. We can now find the posterior distribution of $p|X_1, \ldots, X_n$ as:

$$
\pi(p|X_1, \ldots, X_n) \propto p(X_1, \ldots, X_n; p) \pi(p) \\
\propto \prod_{i=1}^n p^{X_i}(1-p)^{1-X_i} (1-p)^{b-1} \\
= p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i} (1-p)^{b-1} \\
= p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i+b-1},
$$

which is proportional to the density of Beta($\sum_{i=1}^n X_i + 1, n - \sum_{i=1}^n X_i + b$). Therefore,

$$p|X_1, \ldots, X_n \sim \text{Beta}(\sum_{i=1}^n X_i + 1, n - \sum_{i=1}^n X_i + b).
$$

For the rest of the question, we do the calculations under squared loss.

(a) Since we are considering the squared loss, by Theorem 16.7 in Lecture 16, the Bayes estimator is given by the posterior mean,

$$\hat{p} = \mathbb{E}[p|X_1, \ldots, X_n] = \frac{\sum_{i=1}^n X_i + 1}{\sum_{i=1}^n X_i + n - \sum_{i=1}^n X_i + b} = \frac{\sum_{i=1}^n X_i + 1}{n + b + 1}.
$$

(b) The risk of the Bayes estimator can then be computed as

$$R(p, \hat{p}) = \mathbb{E}[(p - \hat{p})^2] = \text{Var}(\hat{p}) + (\text{bias}_p(\hat{p}))^2 \\
= \text{Var}\left(\frac{\sum_{i=1}^n X_i + 1}{n + b + 1}\right) + \left(\mathbb{E}\left[\frac{\sum_{i=1}^n X_i + 1}{n + b + 1}\right] - p\right)^2 \\
= \frac{np(1-p)}{(n + b + 1)^2} + \left(\frac{np + 1}{n + b + 1} - p\right)^2 \\
= \frac{np(1-p)}{(n + b + 1)^2} + \left(\frac{pb + p - 1}{n + b + 1}\right)^2 \\
= \frac{np(1-p) + (pb + p - 1)^2}{(n + b + 1)^2}.
$$

(c) If we plug in $p = 1$, then the risk of the Bayes estimator becomes:

$$R(p, \hat{p}) = \frac{\hat{b}^2}{(n + b + 1)^2}.
$$

Now from Lecture 16 page 4 (Example 16.4) and Lecture 17 page 4 (Example 17.3 and Theorem 17.2), we know that the minimax estimator in this case is given by,

$$\hat{p} = \frac{\sum_{i=1}^n X_i + \sqrt{n/4}}{n + \sqrt{n}}
$$

and its risk is given by,

$$R(p, \hat{p}) = \frac{n}{4(n + \sqrt{n})^2}.$$
As $b$ gets larger, for fixed $n$, $R(p, \hat{p}) \to 1 > \frac{n}{4(n + \sqrt{n})^2} = R(p, \hat{p})$. Especially when it gets larger than $\sqrt{n}$,

$$R(p, \hat{p}) > \frac{n}{(n + \sqrt{n} + 1)^2} > R(p, \hat{p}).$$

That is, the risk of the Bayes estimator becomes larger than the risk of the minimax estimator.

(d) To find the maximum risk of the Bayes estimator, we just have to find the maximum of the numerator. So it is enough to just take the derivative of the numerator of the risk with respect to $p$ and set it to zero to find the $p$ that maximizes the risk.

Let

$$g(p) = np(1 - p) + (pb + p - 1)^2 = np - np^2 + p^2(b + 1)^2 - 2p(b + 1) + 1.$$

Then,

$$g'(p) = n - 2np + 2p(b + 1)^2 - 2(b + 1).$$

Setting this equal to zero, we get

$$g'(p) = 0 \implies 2p((b + 1)^2 - n) = 2(b + 1) - n \implies p = \frac{2(b + 1) - n}{2((b + 1)^2 - n)}.$$

So the maximizer of the Bayes risk is given by

$$p_n = \frac{2(b + 1) - n}{2((b + 1)^2 - n)}.$$

Hence the ratio of the maximum risk of the Bayes estimator to the minimax rate is given by,

$$\text{Ratio} = \frac{np_n(1 - p_n) + (p_n b + p_n - 1)^2}{(n + b + 1)^2} = \frac{4(np_n(1 - p_n) + (p_n b + p_n - 1)^2)(n + \sqrt{n})^2}{n(n + b + 1)^2}.$$

Now notice that

$$\lim_{n \to \infty} p_n = \lim_{n \to \infty} \frac{2(b + 1) - n}{2((b + 1)^2 - n)} = \frac{1}{2}.$$ 

Therefore, as $n \to \infty$,

$$\lim_{n \to \infty} \text{Ratio} = \lim_{n \to \infty} \frac{4np_n(1 - p_n)n^2}{n^3} = \frac{1}{4} = 1.$$