Chapter 12 Solutions

Andrew Wu
Wasserman: All of Statistics

July 4, 2025

Problem 12.1. In each of the following models, find the Bayes risk and the Bayes estimator, using squared error loss.

a) $X \sim \text{Binomial}(n, p), p \sim \text{Beta}(\alpha, \beta)$.

b) $X \sim \text{Pois}(\lambda), \lambda \sim \text{Gamma}(\alpha, \beta)$.

c) $X \sim N(\theta, \sigma^2)$ where σ^2 is known and $\theta \sim N(a, b^2)$.

Solution. Because we're using squared error loss, we know the Bayes estimator is

$$\widehat{\theta}(x) = \int \theta f(\theta|x) d\theta = \mathbb{E}(\theta|X).$$

a) We have

$$\begin{aligned} f(p|X) &\propto p^x (1-p)^{n-x} p^{\alpha-1} (1-p)^{\beta-1} \\ &\propto p^{x+\alpha-1} (1-p)^{n-x+\beta-1} \end{aligned}$$

so

$$p|X \sim \text{Beta}(\alpha + x, n - x + \beta).$$

It follows that

$$f(p|X) = \frac{\Gamma(\alpha+\beta+x)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} p^{\alpha+x-1} (1-p)^{\beta+n-x-1}.$$

Then we compute

$$\begin{aligned} \widehat{p}(x) &= \int_0^1 p \cdot f(p|X) dp \\ &= \int_0^1 \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} p^{\alpha+x+1-1} (1-p)^{\beta+n-x-1} dp \\ &= \int_0^1 \frac{\Gamma(\alpha+\beta+n+1)}{\Gamma(\alpha+x+1)\Gamma(\beta+n-x)} \cdot \frac{\Gamma(\alpha+x+1)}{\Gamma(\alpha+\beta+n+1)} \cdot \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)} p^{\alpha+x+1-1} (1-p)^{\beta+n-x-1} dp \\ &= \frac{\Gamma(\alpha+x+1)}{\Gamma(\alpha+x)} \cdot \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\beta+n+1)} \\ &= \frac{\alpha+x}{\alpha+\beta+n}, \end{aligned}$$

which is the Bayes' estimator.

Note alternatively that $\widehat{p}(x) = \mathbb{E}(p|X)$ means that once we know the posterior is $p|X \sim \text{Beta}(\alpha+x, n-x+\beta)$ that we can directly go to $\widehat{p}(x) = \frac{\alpha+x}{\alpha+x+n-x+\beta} = \frac{\alpha+x}{\alpha+\beta+n}$, because the Beta distribution mean is an established result.

Then the risk of $\widehat{p}$ is

$$\begin{aligned}
R(p,\widehat{p}) &= \mathbb{E}_p\left((\widehat{p}-p)^2\right) \\
&= \mathbb{V}_p(\widehat{p}) + \text{bias}_p^2(\widehat{p}) \\
&= \mathbb{V}_p\left(\frac{\alpha+x}{\alpha+\beta+n}\right) + \text{bias}_p^2\left(\frac{\alpha+x}{\alpha+\beta+n}\right) \\
&= \frac{1}{(\alpha+\beta+n)^2}\mathbb{V}_p(x) + \left(\mathbb{E}\left(\frac{\alpha+x}{\alpha+\beta+n}\right) - p\right)^2 \\
&= \frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{\alpha(1-p)-\beta p}{\alpha+\beta+n}\right)^2
\end{aligned}$$

and the Bayes risk is

$$r(f,\widehat{p}) = \int_0^1 \left[\frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{\alpha(1-p)-\beta p}{\alpha+\beta+n}\right)^2\right] \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}dp.$$

To evaluate this integral, we'll split it. First:

$$\begin{aligned}
\int_0^1 \frac{np(1-p)}{(\alpha+\beta+n)^2} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}dp &= \frac{n}{(\alpha+\beta+n)^2}\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha+1-1}(1-p)^{\beta+1-1}dp \\
&= \frac{n}{(\alpha+\beta+n)^2} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+2)} \cdot \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \cdot \frac{\Gamma(\beta+1)}{\Gamma(\beta)} \\
&= \frac{n}{(\alpha+\beta+n)^2} \cdot \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}.
\end{aligned}$$

Next: observe that

$$\left(\frac{\alpha(1-p)-\beta p}{\alpha+\beta+n}\right)^2 p^{\alpha-1}(1-p)^{\beta-1} = \frac{1}{(\alpha+\beta+n)^2}(\alpha^2(1-p)^2 - 2\alpha\beta p(1-p) + \beta^2 p^2)p^{\alpha-1}(1-p)^{\beta-1}$$

so we'll actually split the second integral into three separate integrals. We'll also remove the constant terms $\frac{1}{(\alpha+\beta+n)^2} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ and put them back later.
We have

$$\begin{aligned}
\int_0^1 \alpha^2(1-p)^2 p^{\alpha-1}(1-p)^{\beta-1}dp &= \alpha^2\int_0^1 p^{\alpha-1}(1-p)^{\beta+2-1}dp \\
&= \frac{\alpha^2\Gamma(\alpha)\Gamma(\beta+2)}{\Gamma(\alpha+\beta+2)},
\end{aligned}$$

and

$$\begin{aligned}
\int_0^1 -2\alpha\beta p(1-p)p^{\alpha-1}(1-p)^{\beta-1}dp &= -2\alpha\beta\int_0^1 p^{\alpha+1-1}(1-p)^{\beta+1-1}dp \\
&= -\frac{2\alpha\beta\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)},
\end{aligned}$$

and finally,

$$\begin{aligned}
\int_0^1 \beta^2 p^2 p^{\alpha-1}(1-p)^{\beta-1}dp &= \beta^2\int_0^1 p^{\alpha+2-1}(1-p)^{\beta-1}dp \\
&= \frac{\beta^2\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)}.
\end{aligned}$$

It follows that our second integral reduces to

$$\frac{\Gamma(\alpha+\beta)}{(\alpha+\beta+n)^2\Gamma(\alpha)\Gamma(\beta)}\left(\frac{\alpha^2\Gamma(\alpha)\Gamma(\beta+2)}{\Gamma(\alpha+\beta+2)}-\frac{2\alpha\beta\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}+\frac{\beta^2\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)}\right)$$

which further reduces to

$$\frac{\alpha^2\Gamma(\alpha)\Gamma(\beta+2)-2\alpha\beta\Gamma(\alpha+1)\Gamma(\beta+1)+\beta^2\Gamma(\alpha+2)\Gamma(\beta)}{(\alpha+\beta+n)^2\Gamma(\alpha)\Gamma(\beta)(\alpha+\beta)(\alpha+\beta+1)}.$$

It doesn't look like we can simplify, but we'll reduce all the Γ terms to multiples of $\Gamma(\alpha)$ and $\Gamma(\beta)$. This yields

$$\frac{\alpha^2\beta(\beta+1)\Gamma(\alpha)\Gamma(\beta)-2\alpha^2\beta^2\Gamma(\alpha)\Gamma(\beta)+\alpha(\alpha+1)\beta^2\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+n)^2\Gamma(\alpha)\Gamma(\beta)(\alpha+\beta)(\alpha+\beta+1)}$$

which further simplifies to

$$\frac{\alpha\beta}{(\alpha+\beta+n)^2(\alpha+\beta+1)}.$$

It follows that

$$\begin{aligned}
r(f,\widehat{p}) &= \frac{n}{(\alpha+\beta+n)^2}\cdot\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}+\frac{\alpha\beta}{(\alpha+\beta+n)^2(\alpha+\beta+1)} \\
&= \frac{\alpha\beta}{(\alpha+\beta+n)^2(\alpha+\beta+1)}\cdot\left(\frac{n}{\alpha+\beta}+1\right) \\
&= \frac{\alpha\beta}{(n+\alpha+\beta)(\alpha+\beta)(\alpha+\beta+1)}.
\end{aligned}$$

b) We have

$$\begin{aligned}
f(\lambda|X) &\propto \frac{e^{-\lambda}\lambda^x}{x!}\frac{1}{\Gamma(\alpha)\beta^\alpha}e^{-\lambda/\beta}\lambda^{\alpha-1} \\
&\propto e^{-\lambda-\lambda/\beta}\lambda^{\alpha+x-1} \\
&\propto e^{\frac{-\lambda}{\beta/(\beta+1)}}\lambda^{\alpha+x-1}
\end{aligned}$$

so

$$\lambda|X \sim \text{Gamma}\left(\alpha+x,\frac{\beta}{\beta+1}\right).$$

It follows that

$$f(\lambda|X)=\frac{1}{\Gamma(\alpha+x)\left(\frac{\beta}{\beta+1}\right)^{\alpha+x}}\cdot e^{\frac{-\lambda}{\beta/(\beta+1)}}\lambda^{\alpha+x-1}.$$

We could compute $\widehat{\lambda}(x)$ with some integrals, but instead, we could note that $\widehat{\lambda}(x)=\mathbb{E}(\lambda|X)=(\alpha+x)\left(\frac{\beta}{\beta+1}\right)$ through known properties of the Gamma distribution. This is our Bayes estimator.

Then the risk of $\widehat{\lambda}$ is

$$\begin{aligned}
R(\lambda, \widehat{\lambda}) &= \mathbb{V}_\lambda(\widehat{\lambda}) + \text{bias}^2_\lambda(\widehat{\lambda}) \\
&= \mathbb{V}_\lambda\left((\alpha + x)\left(\frac{\beta}{\beta+1}\right)\right) + \text{bias}^2_\lambda\left((\alpha + x)\left(\frac{\beta}{\beta+1}\right)\right) \\
&= \left(\frac{\beta}{\beta+1}\right)^2 \lambda\lambda + \left(\mathbb{E}\left((\alpha + x)\left(\frac{\beta}{\beta+1}\right)\right) - \lambda\right)^2 \\
&= \left(\frac{\beta}{\beta+1}\right)^2 \lambda + \left(\left(\frac{\beta}{\beta+1}\right)(\lambda + \alpha) - \lambda\right)^2 \\
&= \left(\frac{\beta}{\beta+1}\right)^2 \lambda + \left(-\frac{\lambda - \alpha\beta}{\beta+1}\right)^2 \\
&= \frac{\beta^2\lambda + (\lambda - \alpha\beta)^2}{(\beta+1)^2}
\end{aligned}$$

Then the Bayes risk is

$$r(f, \widehat{\lambda}) = \int_0^\infty \frac{\beta^2\lambda + (\lambda - \alpha\beta)^2}{(\beta+1)^2} \cdot \frac{1}{\beta^\alpha\Gamma(\alpha)} e^{-\lambda/\beta}\lambda^{\alpha-1} d\lambda.$$

Again we will separate this into multiple integrals. First:

$$\begin{aligned}
\int_0^\infty \frac{\beta^2\lambda}{(\beta+1)^2} \cdot \frac{1}{\beta^\alpha\Gamma(\alpha)} e^{-\lambda/\beta}\lambda^{\alpha-1} d\lambda &= \frac{\beta^2}{(\beta+1)^2} \int_0^\infty \frac{1}{\beta^\alpha\Gamma(\alpha)} e^{-\lambda/\beta}\lambda^{\alpha+1-1} d\lambda \\
&= \frac{\beta^2}{(\beta+1)^2} \int_0^\infty \alpha\beta \cdot \frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)} e^{-\lambda/\beta}\lambda^{\alpha+1-1} d\lambda \\
&= \alpha\beta \cdot \frac{\beta^2}{(\beta+1)^2}.
\end{aligned}$$

Next:

$$\begin{aligned}
\int_0^\infty \frac{\lambda^2}{(\beta+1)^2} \cdot \frac{1}{\beta^\alpha\Gamma(\alpha)} e^{-\lambda/\beta}\lambda^{\alpha-1} d\lambda &= \frac{1}{(\beta+1)^2} \int_0^\infty \frac{1}{\beta^\alpha\Gamma(\alpha)} e^{-\lambda/\beta}\lambda^{\alpha+2-1} d\lambda \\
&= \frac{1}{(\beta+1)^2} \int_0^\infty \alpha(\alpha+1)\beta^2 \cdot \frac{1}{\beta^{\alpha+2}\Gamma(\alpha+2)} e^{-\lambda/\beta}\lambda^{\alpha+2-1} d\lambda \\
&= \frac{\alpha(\alpha+1)\beta^2}{(\beta+1)^2},
\end{aligned}$$

and

$$\begin{aligned}
\int_0^\infty \frac{-2\alpha\beta\lambda}{(\beta+1)^2} \frac{1}{\beta^\alpha\Gamma(\alpha)} e^{-\lambda/\beta}\lambda^{\alpha-1} d\lambda &= -\frac{2\alpha\beta}{(\beta+1)^2} \int_0^\infty \frac{1}{\beta^\alpha\Gamma(\alpha)} e^{-\lambda/\beta}\lambda^{\alpha+1-1} d\lambda \\
&= -\frac{2\alpha\beta}{(\beta+1)^2} \int_0^\infty \alpha\beta \cdot \frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)} e^{-\lambda/\beta}\lambda^{\alpha+1-1} d\lambda \\
&= -\frac{2\alpha^2\beta^2}{(\beta+1)^2},
\end{aligned}$$

and finally,

$$\int_0^\infty \frac{(\alpha\beta)^2}{(\beta+1)^2} \cdot \frac{1}{\beta^\alpha\Gamma(\alpha)} e^{-\lambda/\beta}\lambda^{\alpha-1} d\lambda = \frac{(\alpha\beta)^2}{(\beta+1)^2}.$$

Thus

$$\begin{aligned}
r(f,\widehat{\lambda}) &= \frac{\alpha\beta^3}{(\beta+1)^2} + \frac{\alpha(\alpha+1)\beta^2}{(\beta+1)^2} - \frac{2\alpha^2\beta^2}{(\beta+1)^2} + \frac{\alpha^2\beta^2}{(\beta+1)^2} \\
&= \frac{\alpha\beta^2}{(\beta+1)^2} \cdot (\beta + \alpha + 1 - 2\alpha + \alpha) \\
&= \frac{\alpha\beta^2}{\beta+1}.
\end{aligned}$$

c) We have

$$\begin{aligned}
f(\theta|X) &\propto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\theta)^2\right) \cdot \frac{1}{b\sqrt{2\pi}} \exp\left(-\frac{1}{2b^2}(\theta-a)^2\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}(x-\theta)^2 - \frac{1}{2b^2}(\theta-a)^2\right) \\
&\propto \exp\left(-\frac{\frac{1}{\sigma^2}+\frac{1}{b^2}}{2}\left(\theta - \frac{\frac{x}{\sigma^2}+\frac{a}{b^2}}{\frac{1}{\sigma^2}+\frac{1}{b^2}}\right)^2\right)
\end{aligned}$$

so $\theta|X \sim N\left(\frac{\frac{x}{\sigma^2}+\frac{a}{b^2}}{\frac{1}{\sigma^2}+\frac{1}{b^2}}, \frac{1}{\frac{1}{\sigma^2}+\frac{1}{b^2}}\right)$.

Then $\widehat{\theta}(x) = \mathbb{E}(\theta|X) = \frac{\frac{x}{\sigma^2}+\frac{a}{b^2}}{\frac{1}{\sigma^2}+\frac{1}{b^2}} = \frac{xb^2+a\sigma^2}{b^2+\sigma^2}$. The risk of $\widehat{\theta}$ is

$$\begin{aligned}
R(\theta,\widehat{\theta}) &= \mathbb{V}_\theta\left(\frac{xb^2+a\sigma^2}{b^2+\sigma^2}\right) + \mathrm{bias}_\theta^2\left(\frac{xb^2+a\sigma^2}{b^2+\sigma^2}\right) \\
&= \left(\frac{1}{b^2+\sigma^2}\right)^2 \cdot b^4 \cdot \sigma^2 + \left(\frac{\theta b^2+a\sigma^2}{b^2+\sigma^2} - \theta\right)^2 \\
&= \frac{\sigma^2 b^4}{(b^2+\sigma^2)^2} + \left(\frac{(a-\theta)\sigma^2}{b^2+\sigma^2}\right)^2 \\
&= \frac{\sigma^2 b^4 + \sigma^4(a-\theta)^2}{(b^2+\sigma^2)^2}.
\end{aligned}$$

Then the Bayes risk is

$$\begin{aligned}
r(f,\widehat{\theta}) &= \int_{-\infty}^{\infty} \frac{\sigma^2 b^4 + \sigma^4(a-\theta)^2}{(b^2+\sigma^2)^2} \cdot \frac{1}{b\sqrt{2\pi}} \exp\left(-\frac{1}{2b^2}(\theta-a)^2\right) d\theta \\
&= \frac{\sigma^2 b^4}{(b^2+\sigma^2)^2} + \int_{-\infty}^{\infty} \frac{\sigma^4(a-\theta)^2}{(b^2+\sigma^2)^2} \cdot \frac{1}{b\sqrt{2\pi}} \exp\left(-\frac{1}{2b^2}(\theta-a)^2\right) d\theta.
\end{aligned}$$

But note that the remaining integral is actually equivalent to

$$\begin{aligned}
\frac{\sigma^4}{(b^2+\sigma^2)^2} \int_{-\infty}^{\infty} (a-\theta)^2 f(\theta) d\theta &= \frac{\sigma^4}{(b^2+\sigma^2)^2} \mathbb{E}[(a-\theta)^2] \\
&= \frac{\sigma^4}{(b^2+\sigma^2)^2} \mathbb{V}(\theta) \\
&= \frac{\sigma^4 b^2}{(b^2+\sigma^2)^2}.
\end{aligned}$$

So the Bayes risk is

$$\begin{aligned}
\frac{\sigma^2 b^4}{(b^2+\sigma^2)^2} + \frac{\sigma^4 b^2}{(b^2+\sigma^2)^2} &= \frac{(\sigma^2 b^2)(\sigma^2+b^2)}{(b^2+\sigma^2)^2} \\
&= \frac{\sigma^2 b^2}{b^2+\sigma^2}.
\end{aligned}$$

$\square$

Problem 12.2. Let $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ and suppose we estimate θ with loss function $L(\theta, \widehat{\theta}) = \frac{(\theta - \widehat{\theta})^2}{\sigma^2}$. Show that $\overline{X}$ is admissible and minimax.

Solution. The risk of $\widehat{\theta} = \overline{X}$ is

$$\begin{aligned}
R(\theta, \overline{X}) &= \mathbb{E}_\theta(L(\theta, \overline{X})) \\
&= \mathbb{E}_\theta\left(\frac{(\theta - \overline{X})^2}{\sigma^2}\right) \\
&= \frac{1}{\sigma^2}\left(\theta^2 - \mathbb{E}(2\theta\overline{X}) + \mathbb{E}(\overline{X}^2)\right) \\
&= \frac{1}{\sigma^2}\left(\theta^2 - 2\theta^2 + \mathbb{V}(\overline{X}) + \mathbb{E}(\overline{X})^2\right) \\
&= \frac{1}{\sigma^2}\left(-\theta^2 + \frac{\sigma^2}{n} + \theta^2\right) \\
&= \frac{1}{n}.
\end{aligned}$$

So $\overline{X}$ has constant risk. Now we just need to show it is admissible, and it will follow that $\overline{X}$ is minimax, too.

To show that $\overline{X}$ is admissible, we will consider a sequence of proper priors $\theta \sim N(0, a)$, which converge to the (improper) flat prior. We will show that their corresponding Bayes estimators converge to $\overline{X}$, then conclude that under some regularity conditions that the limit of admissible estimators is admissible.

First we find the posterior density. We have

$$\begin{aligned}
f(\theta|x^n) &\propto f(x^n|\theta)f(\theta) \\
&\propto \left(\prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \theta)^2\right)\right) \cdot \frac{1}{a\sqrt{2\pi}} \exp\left(-\frac{1}{2a^2} \cdot \theta^2\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \theta)^2 - \frac{1}{2a^2} \cdot \theta^2\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2}\left(n\theta^2 - 2\theta\sum_{i=1}^n x_i\right) - \frac{\theta^2}{2a^2}\right)
\end{aligned}$$

and upon completing the square we obtain

$$\theta|x^n \sim N\left(\frac{\sum_{i=1}^n x_i}{n + \frac{\sigma^2}{a^2}}, \frac{a^2\sigma^2}{na^2 + \sigma^2}\right).$$

It is known that the Bayes estimator is the posterior mean if the loss function is squared error loss. Multiplying by a constant doesn't change this; therefore the Bayes estimator is

$$\widehat{\theta} = \frac{\sum_{i=1}^n x_i}{n + \frac{\sigma^2}{a^2}}.$$

This estimator is admissible, and moreover, as $a \to \infty$, $\widehat{\theta} \to \overline{X}$. It follows that $\overline{X}$ is admissible and thus minimax. $\square$

Problem 12.3. Let $\Theta = \{\theta_1, \ldots, \theta_k\}$ be a finite parameter space. Prove that the posterior mode is the Bayes estimator under zero-one loss.

Solution. The loss function for zero-one loss is $L(\theta, \widehat{\theta}) = 0$ if $\theta = \widehat{\theta}$ and $L(\theta, \widehat{\theta}) = 1$ otherwise.

The Bayes estimator is the estimator that minimizes the posterior risk.

In this case, the posterior risk of $\widehat{\theta}$ is $r(\widehat{\theta}|x) = \sum_{i=1}^{k} L(\theta_i, \widehat{\theta}(x)) f(\theta_i|x)$. In choosing $\widehat{\theta}$, we can have no impact upon $f(\theta_i|x)$; indeed, the best we can do is to force $L(\theta_i, \widehat{\theta}(x)) = 0$ as much as possible. That is, we want to take $\widehat{\theta}$ to be equal to as many of the θ_is as possible, so the posterior mode is indeed the Bayes estimator under zero-one loss. ☐

Problem 12.4. Let $X_1, \ldots, X_n$ be a sample from a distribution with variance σ^2. Consider estimators of the form bS^2 where S^2 is the sample variance. Let the loss function for estimating σ^2 be

$$L(\sigma^2, \widehat{\sigma}^2) = \frac{\widehat{\sigma}^2}{\sigma^2} - 1 - \log\left(\frac{\widehat{\sigma}^2}{\sigma^2}\right).$$

Find the optimal value of b that minimizes the risk for all σ^2.

Solution. Note that $\mathbb{E}(S^2) = \sigma^2$. The risk is

$$\begin{aligned}
\mathbb{E}(L(\sigma^2, \widehat{\sigma}^2)) &= \mathbb{E}\left(\frac{\widehat{\sigma}^2}{\sigma^2} - 1 - \log\left(\frac{\widehat{\sigma}^2}{\sigma^2}\right)\right) \\
&= \mathbb{E}\left(\frac{bS^2}{\sigma^2} - 1 - \log\left(\frac{bS^2}{\sigma^2}\right)\right) \\
&= \mathbb{E}\left(\frac{bS^2}{\sigma^2}\right) - 1 - \mathbb{E}\left(\log b + \log S^2 - \log \sigma^2\right) \\
&= b - 1 - \log b - \mathbb{E}(\log S^2) + \log \sigma^2
\end{aligned}$$

so to minimize the risk, we ought to minimize $b - \log b - \mathbb{E}\left(\log S^2\right) + \log \sigma^2$. However, the only parts of the risk that involve b are $b - \log b$, and this is minimized when $b = 1$ after taking the derivative and setting it equal to 0. ☐

Problem 12.5. Let $X \sim \text{Binomial}(n, p)$ and suppose the loss function is

$$L(p, \widehat{p}) = \left(1 - \frac{\widehat{p}}{p}\right)^2$$

where $0 < p < 1$. Consider the estimator $\widehat{p}(X) = 0$. This estimator falls outside the parameter space $(0, 1)$ but we will allow this. Show that $\widehat{p}(X) = 0$ is the unique, minimax rule.

Solution. The maximum risk is $\overline{R}(\widehat{p}) = \sup_p R(p, \widehat{p})$. The risk of any estimator $\widehat{p}$ is

$$R(p, \widehat{p}) = \mathbb{E}_p\left(\left(1 - \frac{\widehat{p}}{p}\right)^2\right)$$

so when $\widehat{p} = \widehat{p}(X) = 0$, the risk is always 1 no matter the value of p. Thus we must show that with any other estimator $\widehat{p}$, the risk exceeds 1 for some value of p.

Note that any estimator $\widehat{p}$ is a function of X; that is, it maps inputs from $\{0, 1, \ldots, n\}$ to $(0, 1)$. Denote $\widehat{p}(i) = \widehat{p}_i$ for any estimator $\widehat{p}$. We split the problem into two cases.

Case 1: $\widehat{p}_0 \neq 0$. We claim that for some p, $\mathbb{E}_p\left(\left(1 - \frac{\widehat{p}}{p}\right)^2\right) > 1$. We have

$$\begin{aligned}
\mathbb{E}_p\left(\left(1 - \frac{\widehat{p}}{p}\right)^2\right) &= \sum_{i=0}^{n} \mathbb{P}(X = i)\left(1 - \frac{\widehat{p}_i}{p}\right)^2 \\
&> \mathbb{P}(X = 0)\left(1 - \frac{\widehat{p}_0}{p}\right)^2 \\
&= (1 - p)^n \left(1 - \frac{\widehat{p}_0}{p}\right)^2
\end{aligned}$$

and taking $p \to 0$, we see that $(1-p)^n \to 1$ and $\left(1 - \frac{\widehat{p}_0}{p}\right)^2 \to \infty$, so for some value of p the risk exceeds 1.

Case 2: $\widehat{p}_0 = 0$. Suppose k is the smallest positive integer for which $\widehat{p}_k \neq 0$. (If k does not exist, then $\widehat{p}$ is just always 0, which reduces to the estimator we were given.)

Then we have

$$\begin{aligned}
\mathbb{E}_p\left(\left(1 - \frac{\widehat{p}}{p}\right)^2\right) &= \sum_{i=0}^{n} \mathbb{P}(X = i)\left(1 - \frac{\widehat{p}_i}{p}\right)^2 \\
&\geq \sum_{i=0}^{k-1} \mathbb{P}(X = i) + \mathbb{P}(X = k)\left(1 - \frac{\widehat{p}_k}{p}\right)^2.
\end{aligned}$$

It thus suffices to show that $\mathbb{P}(X = k)\left(\frac{p-\widehat{p}_k}{p}\right)^2 > \mathbb{P}(X \geq k)$ for some value of p. But we have

$$\begin{aligned}
\mathbb{P}(X = k)\left(\frac{p - \widehat{p}_k}{p}\right)^2 &\approx \mathbb{P}(X = k)\left(\frac{\widehat{p}_k}{p}\right)^2 \\
&= \binom{n}{k} p^k (1-p)^{n-k} \left(\frac{\widehat{p}_k}{p}\right)^2 \\
&= \binom{n}{k} \widehat{p}_k^2 p^{k-2}.
\end{aligned}$$

when p is small. But

$$\mathbb{P}(X \geq k) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

and all the terms in the summation are of order p^k at least. Thus for sufficiently small values of p, we have $\mathbb{P}(X = k)\left(\frac{p-\widehat{p}_k}{p}\right)^2 > \mathbb{P}(X \geq k)$. We are done. $\square$