Homework 2

Instructor: Hedibert Freitas Lopes
Course: STP 598 Advanced Bayesian Statistical Learning (Class # 31199)
Semester: Spring 2022
Due date: 1:30pm, February 28th, 2022.

1 Risk analysis

Recall that the risk of an estimator $\hat{\theta}$ is given by

$$R(\theta, \hat{\theta}) = E_\theta[L(\theta, \hat{\theta})] = \int L(\theta, \hat{\theta}) p(x|\theta) dx,$$

while the maximum risk is $\overline{R}(\hat{\theta}) = \sup_\theta R(\theta, \hat{\theta})$, and the Bayes risk is

$$r(\pi, \hat{\theta}) = \int_\Theta R(\theta, \hat{\theta}) \pi(\theta) d\theta,$$

where π is a prior for θ. Assume that the loss function is squared error, so the risk is just the mean squared error (MSE):

$$R(\theta, \hat{\theta}) \;\; = \;\; E_\theta[(\hat{\theta} - \theta)^2] = E_\theta[(\hat{\theta} - E_\theta(\hat{\theta}))^2 + (E_\theta(\hat{\theta}) - \theta)^2] = V_\theta(\hat{\theta}) + \text{bias}^2_\theta(\hat{\theta}).$$

Now, let $X_1, \ldots, X_n$ be, conditionally on θ, independent Bernoulli(θ), for $\theta \in (0, 1)$. Consider squared error loss and two estimators of θ:

$$\hat{\theta}_1 = \frac{X_1 + \cdots + X_n}{n} \quad \text{and} \quad \hat{\theta}_2 = \frac{X_1 + \cdots + X_n + \alpha}{\alpha + \beta + n},$$

where α and β are positive constants.

a) Show that

$$R(\theta, \hat{\theta}_1) = \frac{\theta(1-\theta)}{n}$$

b) Show that

$$R(\theta, \hat{\theta}_2) = \frac{n\theta(1-\theta)}{(\alpha+\beta+n)^2} + \left(\frac{n\theta + \alpha}{\alpha + \beta + n} - \theta\right)^2$$

For c), d) and e), assume that $\alpha = \beta = \sqrt{n/4}$.

c) Graphically show that neither estimator uniformly dominates the other. Try $n = 1, 10, 50$ to see how the risk functions behave as n increases.

d) Show that the maximum risks are

$$\overline{R}(\hat{\theta}_1) = \frac{1}{4n} \quad \text{and} \quad \overline{R}(\hat{\theta}_2) = \frac{n}{4(n+\sqrt{n})^2},$$

so, based on the maximum risk, $\hat{\theta}_2$ is a better estimator. However, when n is large, $R(\hat{\theta}_1)$ has smaller risk except for a small region in the parameter space near $\theta = 1/2$, where the risk of $\hat{\theta}_1$ is maximum.

e) Show that the Bayes risks are

$$r(\pi, \hat{\theta}_1) = \frac{1}{6n} \quad \text{and} \quad r(\pi, \hat{\theta}_2) = \frac{n}{4(n+\sqrt{n})^2},$$

when π is the uniform prior in the interval $(0, 1)$. For large n (larger than or equal to 20), $\hat{\theta}_1$ is a better estimator. This corroborates with the graphical inspection obtained in c).

2 Stein's Paradox

Suppose that $X \sim N(\theta, 1)$ and consider estimating θ with squared error loss. We know that $\hat{\theta}(X) = X$ is admissible. Now consider estimating two, unrelated quantities $\theta = (\theta_1, \theta_2)$ and supposed that $X_1 \sim N(\theta_1, 1)$ and $X_2 \sim N(\theta_2, 1)$ independently, with loss

$$L(\theta, \hat{\theta}) = (\theta_1 - \hat{\theta}_1)^2 + (\theta_2 - \hat{\theta}_2)^2.$$

Not surprisingly, $\hat{\theta}(X) = X$ is again admissible where $X = (X_1, X_2)$. Now consider the generalization to k normal means. Let $\theta = (\theta_1, \dots, \theta_k)$, $X = (X_1, \dots, X_k)$ with $X_i \sim N(\theta_i, 1)$ (independent) and loss

$$L(\theta, \hat{\theta}) = (\theta_1 - \hat{\theta}_1)^2 + \cdots + (\theta_k - \hat{\theta}_k)^2.$$

Stein astounded everyone when he proved that if $k \geq 3$, then $\hat{\theta}(X) = X$ is inadmissible. It can be shown that the **James-Stein estimator**

$$\hat{\theta}^S = (\hat{\theta}_1^S, \dots, \hat{\theta}_k^S)$$

has smaller risk, where

$$\hat{\theta}_i^S(X) = \left(1 - \frac{k-2}{X_1^2 + \cdots + X_k^2}\right)^+ X_i,$$

where $(z)^+ = max\{0, z\}$. This estimator shrinks the X_i's towards 0. The message is that, when estimating many parameters, there is great value in shrinking the estimates.

Computer Experiment: Compare the risk of the MLE and the James-Stein estimator by simulation. Try various values of k and various vectors θ. Summarize your results.