Midterm 2: Math 6266 (Zhilova)

Peter Williams

11/30/2017

Exercise 1 (The James-Stein estimator)

Let $X \sim N(\theta, \sigma^2 I_p)$ for some $\sigma^2 > 0$, $\theta \in R^p$; dimension ≥ 3; θ is an unknown true parameter. Denote the quadratic risk function as $R(\delta, \theta) = E_\theta(|\delta - \theta|)$, where $\delta = \delta(X)$ is some estimator of θ, and $|\cdot|$ is the ℓ_2-norm in R^p.

1. Calculate the quadratic risk for $\delta = X$

With $R(\theta, \delta) = R(\theta, X) = E[L(0, X)] = E||X - \theta||^2$. We can calculate the quadratic risk:

$$E||X - \theta||^2 = E(X - \theta)^\intercal (X - \theta) = E[X^\intercal X] - 2\theta^\intercal E[X] + \theta^\intercal \theta = E[X^\intercal X] - \theta^\intercal \theta = E[X^\intercal X] - ||\theta||^2$$

which for $X \sim N(\theta, \sigma^2 I_p)$, reduces to

$$E[X^\intercal X] - ||\theta||^2 = \sum_{i=1}^{p} E[X_i^2] - ||\theta||^2 = \sum_{i=1}^{p} (\theta_i^2 + \sigma^2) - ||\theta||^2 = p\sigma^2 + ||\theta||^2 - ||\theta||^2 = p\sigma^2$$

2. Let $R = p\sigma^2 + ||h(X)||^2 - 2\sigma\ trace(Dh(X))$, where $h = (h_1, ..., h_p)^\intercal : R^p \to R^p$ is a differentiable function, s.t. all necessary moments exist. $Dh(X)$ is a $p \times p$ matrix of partial derivatives: $\{Dh(x)\}_{i,j} = \frac{\partial}{\partial x_j} h_i(x)$ Show that $\hat{R}$ is an unbiased risk estimator for $\delta(X) = h(X)$, i.e.

$$R(\theta, X - h(X)) = E_\theta \hat{R}$$

(Hint: use Stein's identity)

3. Consider $h(X) = \frac{(p-2)\sigma^2}{||X||^2} X$ and the James-Stein estimator $X - h(X)$. Show that $R(\theta, \hat{\theta}_{JS}) < R(\theta, X)$, for all $\theta \in R^p$.

4. Now consider an i.i.d. sample $Y_1, ..., Y_n$ where $Y_i \sim N(\theta, \sigma^2 I_p)$. Denote $\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$. Calculate the risk $R(\theta, \bar{Y})$.

5. Consider the estimator $\hat{\theta}_{JS} = \bar{Y} - \frac{(p-2)\sigma^2}{||\bar{Y}||^2} \bar{Y}$. Show that $R(\theta, \hat{\theta}_{JS}) < R(\theta, \bar{Y})$ for all $\theta \in R^p$.

(Hint: Use that $Y \sim N(\theta, \frac{\sigma^2}{n} I_p)$).

Exercise 2

Consider the linear regression model $Y_i = X_i^\intercal \theta^* + \varepsilon_i$, $i = 1, ..., n$, the errors ε_i are $i.i.d.$,$E\varepsilon_i = 0$, $Var(\varepsilon_i) = \sigma^2 > 0$ The unknown true parameter $\theta^* \in R^p$. Assume that matrix $XX^\intercal = \sum_{i=1}^{n} X_i X_i^\intercal$ is not invertible, i.e. some of its eigenvalues equal to zero.

Derive the spectral representation of the model$Y = X^\intercal \theta^* + \varepsilon$ (this was done at a lecture), i.e. show that for some $Z, \xi, \eta^* \in R^p$ the model is equivalent to $Z = \lambda \eta^* + \xi$,

where $\lambda = diag\{\lambda_1, ..., \lambda_p\}$, and $\lambda_1 \geq ... \geq \lambda_p \geq 0$ are eigenvalues of $XX^\intercal$

Let $A = diag\{\alpha_1, ..., \alpha_p\}$ for some numbers $\alpha_1, ..., \alpha_p \in [0, 1]$. Let $\hat{\eta}_A = (\hat{\eta}_{A,1}, ..., \hat{\eta}_{A,p})^\intercal$, be a shrinkage estimator of $\hat{\eta}^* = (\eta_1^*, ..., \eta_p^*)^\intercal$

$$\hat{\eta}A, j = \begin{cases} \alpha_j \lambda_j^{-1} z_j, & \text{if } \lambda_j \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Find bias, variance and the quadratic risk of $\hat{\eta}A : R(\eta^*, \hat{\eta}A) = E(||\hat{\eta}A - \eta^*||^2)$

Exercise 3

Let $X_1, ..., X_n$ be real valued $i.i.d.$ random variables. Assume $E(|X_i|M) < \infty$ for some $M \geq 2$. Let $X_1^*, ..., X_n^*$ be a bootstrap sample based on the original data $X_1, ..., X_n$ and obtained by the Efron's bootstrap procedure, i.e.

$$P(X_j^* = X_i | \{X_i\}_{i=1}^n) = 1/n \quad \forall\ j = 1, ..., n$$

Show that for all integer $m \in [0, M]$

$$E(X_j^{*m} | \{X_i\}_{i=1}^n) \xrightarrow{P} E(X_1^m)\ for\ n \to \infty.$$

Show also that

$$Var(X_j^* | \{X_i\}_{i=1}^n) \xrightarrow{P} Var(X_1)\ for\ n \to \infty.$$

(Hint 1: Use the Weak Law of Large Numbers.)
(Hint 2: the 1-st bootstrap moment of X_j^* equals to $E(X_j^* | \{X_i\}_{i=1}^n) = \sum_{i=1}^n X_i/n$.)