Math 6266. Fall 2017. Midterm 2.

Exercise 1 (The James-Stein estimator)**.** Let $X \sim \mathcal{N}(\theta, \sigma^2 \boldsymbol{I}_p)$ for some $\sigma^2 > 0$, $\theta \in \mathbf{R}^p$; dimension $p \geq 3$; θ is an unknown true parameter. Denote the quadratic risk function as $R(\delta, \theta) = \mathbf{E}_\theta(\|\delta - \theta\|^2)$, where $\delta = \delta(X)$ is some estimator of θ, and $\|\cdot\|$ is the ℓ_2-norm in $\mathbf{R}^p$.

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

2. Let $\hat{R} \stackrel{\text{def}}{=} p\sigma^2 + \|h(X)\|^2 - 2\sigma^2 \text{trace}(Dh(X))$, where $h = (h_1, \ldots, h_p)^\top : \mathbf{R}^p \mapsto \mathbf{R}^p$ 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 $\delta(X) = X - h(X)$, i.e.

   $$R(\theta, X - h(X)) = \mathbf{E}_\theta \hat{R}.$$

   (Hint: use Stein's identity)

3. Consider

   $$h(X) = \frac{(p-2)\sigma^2}{\|X\|^2} X$$

   and the James-Stein estimator

   $$\hat{\theta}_{JS} = X - h(X).$$

   Show that $R(\theta, \hat{\theta}_{JS}) < R(\theta, X)$ for all $\theta \in \mathbf{R}^p$.

4. Now consider an i.i.d. sample $Y_1, \ldots, Y_n$, where $Y_i \sim \mathcal{N}(\theta, \sigma^2 \boldsymbol{I}_p)$. Denote $\overline{Y} \stackrel{\text{def}}{=} n^{-1} \sum_{i=1}^n Y_i$. Calculate the risk $R(\theta, \overline{Y})$.

5. Consider the estimator

   $$\tilde{\theta}_{JS} = \overline{Y} - \frac{(p-2)\sigma^2/n}{\|\overline{Y}\|^2} \overline{Y}.$$

   Show that $R(\theta, \tilde{\theta}_{JS}) < R(\theta, \overline{Y})$ for all $\theta \in \mathbf{R}^p$.
   (Hint: Use that $\overline{Y} \sim \mathcal{N}(\theta, \sigma^2 n^{-1} \boldsymbol{I}_p)$)

References:

- Lecture note "The James-Stein Phenomenon" by Prof. Michael Jordan:

  `https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjv58a-8OLXAhXGZCYKHaX3Bc8QFggmMAA&url=https%3A%2F%2Fpiazza.com%2Fclass_profile%2Fget_resource%2Fhzdbtb6jdr56q1%2Fi2kz4qj4x102b1&usg=AOvVaw3i8Zmw8Sq6oU`

- A "non-technical" introduction with a real data example by Prof. Richard J. Samworth: `https://pdfs.semanticscholar.org/7eeb/d55f569395544f2b5d367d6aee614901d2c1.pdf`

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

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

$$Z = \Lambda \eta^* + \xi,$$

where $\Lambda = \mathrm{diag}\{\lambda_1, \dots, \lambda_p\}$ and $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ are eigenvalues of $XX^\top$.

Let $A \stackrel{\text{def}}{=} \mathrm{diag}\{\alpha_1, \dots, \alpha_p\}$ for some numbers $\alpha_1, \dots, \alpha_p \in [0, 1]$ Let $\hat{\eta}_A = (\hat{\eta}_{A,1}, \dots, \hat{\eta}_{A,p})^\top$ be a shrinkage estimator of $\eta^* = (\eta_1^*, \dots, \eta_p^*)^\top$:

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

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

Reference: Chapters 4.2.4, 4.7, 4.8 in the textbook `https://link.springer.com/book/10.1007/978-3-642-39909-1`

Exercise 3. Let $X_1, \dots, X_n$ be real valued i.i.d. random variables. Assume $\mathbf{E}(|X_i|^M) < \infty$ for some $M \geq 2$. Let $X_1^*, \dots, X_n^*$ be a bootstrap sample based on the original data $X_1, \dots, X_n$ and obtained by the Efron's bootstrap procedure, i.e.

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

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

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

Show also that

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

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

References:

- A concise introductory text:
  `http://galton.uchicago.edu/~eichler/stat24600/Handouts/bootstrap.pdf`
- Lecture notes by Prof. Peter Hall:
  `http://anson.ucdavis.edu/~peterh/sta251/bootstrap-lectures-to-may-16.pdf`