36700 – Probability and Mathematical Statistics

Spring 2019

Homework 7

Due Wednesday, April 10th at 12:40 PM

- All homework assignments shall be uploaded using Gradescope through the Canvas portal). Late submissions are not allowed.

Q1. Let X, Y be two random variables. Let $r(x) = \mathbb{E}(Y|X = x)$, and $f(x)$, $g(x)$ be an arbitrary functions.

(a) Use Q6 of HW2 to prove that (assuming all expectations are finite)

$$\mathbb{E}[(Y - r(X))g(X)] = 0\,.$$

(b) Use the previous part to prove that (again, assuming all expectations are finite)

$$\mathbb{E}(Y - f(X))^2 = \mathbb{E}(Y - r(X))^2 + \mathbb{E}(r(X) - f(X))^2\,.$$

As a result, $r(X)$ is the optimal predictor of Y based on X in terms of squared loss.

Q2. (Interpretation of least squares without linearity assumption) Let X, Y be two random variables with finite second moments. Let (β_0, β_1) be the minimizer of the following problem:

$$\min_{(b_0,b_1)} \mathbb{E}(Y - b_0 - b_1 X)^2\,.$$

Here, the interpretation of $\beta_0 + \beta_1 X$ is no longer the conditional expectation since we do not assume $r(x) = \mathbb{E}(Y|X = x)$ is linear. Instead, it s the best linear predictor of Y using X.

(a) Express β_0 and β_1 using moments of (X, Y).

(b) Compare your result in part (a) with the least squares estimates $(\hat{\beta}_0, \hat{\beta}_1)$. What kind of estimates is it (MLE, MoM, Plug-in, Bayes)?

Q3. In the simple regression problem, assume $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, where ϵ_i are iid $N(0, \sigma^2)$. Here the unknown parameter is $(\beta_0, \beta_1, \sigma^2)$.

(a) Find the MLE of $(\beta_0, \beta_1, \sigma^2)$.

(b) Find the (exact) distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ conditioning on X_i's (Hint: if $Z_i \sim N(\mu_i, \sigma_i^2)$ are independent, then $\sum_i a_i Z_i$ is normal with mean $\sum_i a_i \mu_i$ and variance $\sum_i a_i^2 \sigma_i^2$. Moreover, in the lecture we already derived the mean and variance of $\hat{\beta}_1$, and the mean of $\hat{\beta}_0$, which you can directly use as known results).

(c) Find the MSE (mean squared error) of the MLE (for β_0, β_1 only). Is the MLE consistent?

Q4. (Prediction intervals) In the same context as Q3, let's say we have a new independent data pair (X_0, Y_0) from the same model. But we only observe X_0 and would like to predict Y_0. Our point estimate of Y_0 is $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0$. Find a value δ such that $\mathbb{P}(\hat{y}_0 - \delta \le Y_0 \le \hat{y}_0 + \delta) = 1 - \alpha$. [Hint: Let $Y_0 = \beta_0 + \beta_1 X_0 + \epsilon_0$. It is equivalent to have $\mathbb{P}(-\delta \le \beta_0 + \beta_1 X_0 + \epsilon_0 - \hat{y}_0 \le \delta) = 1 - \alpha$. Use Q3 to show that $\beta_0 + \beta_1 X_0 + \epsilon_0 - \hat{y}_0$ is a normal variable with mean zero, and the desired δ can be determined by its variance and α.]

Q5. If we write

$$\mathbf{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_n \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix},$$

then the regression model can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

Verify that the LS estimate is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y},$$

where $\cdot^T$ denotes the matrix transpose. In other words, you need to show that the matrix form solution $\hat{\boldsymbol{\beta}}$ agrees with the LS solution given on page 3 of Lecture Note 16.

Optional problem. Even without normality assumption as in Q3 or the linearity assumption, one can still establish asymptotic normality of the least squares estimate $(\hat{\beta}_0, \hat{\beta}_1)$ using the CLT and delta method. In this exercise, we will use the setting of Q2, assuming a random design (consider $X_1, ..., X_n$ as iid random variables).

- Prove that
$$\sqrt{n}(\hat{\beta}_0 - \beta_0) \rightsquigarrow N(0, \sigma_0^2), \quad \sqrt{n}(\hat{\beta}_1 - \beta_1) \rightsquigarrow N(0, \sigma_1^2)$$
and find the asymptotic variances σ_0^2, σ_1^2 (expressed as functions of moments of X, Y).

- Indeed we can establish joint asymptotic normality of $(\hat{\beta}_0, \hat{\beta}_1)$. Show that
$$\sqrt{n}\begin{bmatrix} \hat{\beta}_0 - \beta_0 \\ \hat{\beta}_1 - \beta_1 \end{bmatrix} \rightsquigarrow N(\mathbf{0}, \Sigma)$$
where $\mathbf{0}$ is a zero vector of length 2, and find the asymptotic covariance matrix Σ.

- Using the previous part to construct a $(1-\alpha)$ confidence set for the two-dimensional vector (β_0, β_1).