ESE 531: Statistical Learning and Inference

Homework 1: Properties of Random Samples

1. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ be the sample mean and sample variance, respectively, of $X_1, \ldots, X_n$. Then suppose another observation, X_{n+1}, becomes available. Establish the following recursion relations for sample means and sample variances.

   (a) Show that $\bar{X}_{n+1} = \frac{X_{n+1} + n\bar{X}_n}{n+1}$.

   (b) Show that $nS_{n+1}^2 = (n-1)S_n^2 + \left(\frac{n}{n+1}\right)\left(X_{n+1} - \bar{X}_n\right)^2$

2. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_n$. The empirical variance is another estimator of the population variance defined as

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_n - \bar{X}_n)^2,$$

   (a) Show that $\widehat{\sigma}_n^2$ is a biased estimator of the population variance σ^2. Determine the exact bias.

   (b) Propose a correction to the empirical variance that removes the bias in the estimator. In other words, find a meaningful function $g(\cdot)$ that guarantees:

$$\mathbb{E}[g(\widehat{\sigma}_n^2)] = \sigma^2$$

   (c) Show that if the population mean is known, then the empirical variance is unbiased. That is, show the following estimator is unbiased:

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

3. Let $w_1, \ldots, w_n$ define a set of weights such that $w_i \geq 0$ and $\sum_{i=1}^{n} w_i = 1$. The weighted sample mean and variance are defined as follows:

$$\widehat{\mu}_n = \sum_{i=1}^{n} w_i X_i$$

   (a) Show that $\widehat{\mu}_n$ is and unbiased estimator for the population mean.

   (b) Compare the variance of the weighted sample mean $\widehat{\mu}_n$ with the unweighted sample mean $\bar{X}_n$. Which has smaller variance?

4. Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2. Show that

$$\mathbb{E}\left[\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}\right] = 0$$

$$\mathbb{V}\left[\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}\right] = 1$$

   Thus, the normalization of $\bar{X}_n$ in the Central Limit Theorem gives random variables that have the same mean and variance as the limiting $\mathcal{N}(0,1)$ distribution.

5. Let $X_1, X_2, \ldots$ be a sequence of random variables that converges in probability to a constant a, that is $X_n \rightarrow a$ in probability. Assume that $\mathbb{P}(X_i > 0) = 1$ for all i (i.e., the random variables X_i are positive). Verify that the sequences defined by $Y_i = \sqrt{X_i} \rightarrow \sqrt{a}$ and $Y_i' = a/X_i \rightarrow 1$ converge in probability.