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 $\bar{X}_{n+1} = \frac{X_{n+1} + n\bar{X}_n}{n+1}$.

   Solution.

$$\begin{aligned}
\bar{X}_{n+1} &= \frac{1}{n+1}\sum_{i=1}^{n+1} X_i \\
&= \frac{1}{n+1}\left(X_{n+1} + \sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n+1}\left(X_{n+1} + n\bar{X}\right)
\end{aligned}$$

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

   Solution. We proceed by using the solution to part (a) and further manipulating the equality:

$$\begin{aligned}
nS_{n+1}^2 &= \frac{n}{(n+1)-1}\sum_{i=1}^{n+1}\left(X_i - \bar{X}_{n+1}\right)^2 \\
&= \sum_{i=1}^{n+1}\left(X_i - \frac{X_{n+1} + n\bar{X}_n}{n+1}\right)^2 \quad \text{(using part (a))} \\
&= \sum_{i=1}^{n+1}\left(X_i - \frac{X_{n+1}}{n+1} - \frac{n\bar{X}_n}{n+1}\right)^2 \\
&= \sum_{i=1}^{n+1}\left[\left(X_i - \bar{X}_n\right) - \left(\frac{X_{n+1}}{n+1} - \frac{\bar{X}_n}{n+1}\right)\right]^2 \quad (\pm\bar{X}_n) \\
&= \sum_{i=1}^{n+1}\left[\left(X_i - \bar{X}_n\right)^2 - 2\left(X_i - \bar{X}_n\right)\left(\frac{X_{n+1} - \bar{X}_n}{n+1}\right) + \frac{1}{(n+1)^2}\left(X_{n+1} - \bar{X}_n\right)^2\right] \\
&= \sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 + \left(X_{n+1} - \bar{X}_n\right)^2 - 2\frac{\left(X_{n+1} - \bar{X}_n\right)^2}{n+1} + \frac{n+1}{(n+1)^2}\left(X_{n+1} - \bar{X}_n\right)^2 \\
&= (n-1)S_n^2 + \frac{n}{n+1}\left(X_{n+1} - \bar{X}_n\right)^2.
\end{aligned}$$

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.

Solution. We need to see if $\mathbb{E}[\widehat{\sigma}_n^2] = \sigma^2$

$$\begin{aligned}
\mathbb{E}[\widehat{\sigma}_n^2] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n + \mu - \mu)^2\right] && (\pm\mu) \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}((X_i - \mu) - (\bar{X}_n - \mu))^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X}_n - \mu) + (\bar{X}_n - \mu)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i - \mu)^2] - \frac{2}{n}\mathbb{E}\left[\sum_{i=1}^{n}(\bar{X}_n - \mu)(X_i - \mu)\right] + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(\bar{X}_n - \mu)^2]
\end{aligned}$$

We can deal with each of these terms separately. For the first term, we use the fact that the samples are identically distributed to obtain

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i - \mu)^2] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X - \mu)^2] = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \sigma^2$$

The second term is the most tricky, but can be determined easily by manipulating the sum:

$$\begin{aligned}
-\frac{2}{n}\mathbb{E}\left[\sum_{i=1}^{n}(\bar{X}_n - \mu)(X_i - \mu)\right] &= -\frac{2}{n}\mathbb{E}\left[(\bar{X}_n - \mu)\sum_{i=1}^{n}(X_i - \mu)\right] \\
&= -2\mathbb{E}\left[(\bar{X}_n - \mu)\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right] \\
&= -2\mathbb{E}\left[(\bar{X}_n - \mu)(\bar{X}_n - \mu)\right] \\
&= -2\left(\frac{\sigma^2}{n}\right) && \text{(variance of sample mean)}
\end{aligned}$$

Finally, the last term is simply the variance of the sample mean:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(\bar{X}_n - \mu)^2] = \mathbb{E}[(\bar{X}_n - \mu)^2] = \frac{\sigma^2}{n}$$

Putting these three parts together, we obtain:

$$\mathbb{E}[\widehat{\sigma}^2] = \sigma^2 - \frac{2\sigma^2}{n} + \frac{\sigma^2}{n} = \left(\frac{n-1}{n}\right)\sigma^2$$

(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:

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

Solution. Consider a class of linear functions $g(x) = cx$, c is some constant. We have that:

$$\mathbb{E}[g(\widehat{\sigma}_n^2)] = \mathbb{E}[c\widehat{\sigma}_n^2] = c\mathbb{E}[\widehat{\sigma}_n^2] = c\left(\frac{n-1}{n}\right)\sigma^2$$

To satisfy, $\mathbb{E}[g(\widehat{\sigma}_n^2)] = \sigma^2$, we can choose $c = \frac{n}{n-1}$. This is called *Bessel's correction* and yields the sample variance.

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

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

Solution. The proof is straightforward and simply uses the linearity property of expectation and the definition of variance:

$$\begin{aligned}
\mathbb{E}[\widehat{\sigma}_n^2] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(X_i - \mu)^2\right] \\
&= \mathbb{E}\left[(X - \mu)^2\right] = \sigma^2
\end{aligned}$$

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.

Solution. The proof is straightforward and uses the given condition $\sum_{i=1}^{n} w_i = 1$ and the linearity property of expectations:

$$\begin{aligned}
\mathbb{E}[\widehat{\mu}_n] &= \mathbb{E}\left[\sum_{i=1}^{n} w_i X_i\right] \\
&= \sum_{i=1}^{n} w_i \mathbb{E}[X_i] \\
&= \sum_{i=1}^{n} w_i \mu = \mu
\end{aligned}$$

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

Solution. Recall the variance of the unweighted sample mean is $\mathbb{V}[\bar{X}_n] = \frac{\sigma^2}{n}$. For the weighted sample mean we have

$$\begin{aligned}
\mathbb{V}[\widehat{\mu}_n] &= \mathbb{V}\left[\sum_{i=1}^{n} w_i X_i\right] \\
&= \sum_{i=1}^{n} \mathbb{V}\left[w_i X_i\right] && \text{(by independence)} \\
&= \sum_{i=1}^{n} w_i^2 \mathbb{V}[X_i] && (\mathbb{V}[aX] = a^2\mathbb{V}[X]) \\
&= \sigma^2\left(\sum_{i=1}^{n} w_i^2\right)
\end{aligned}$$

It is not difficult to see that the minimum variance of $\widehat{\mu}_n$ is obtained by setting $w_i = \frac{1}{n}$ for all i and so:

$$\mathbb{V}[\bar{X}] \leq \mathbb{V}[\widehat{\mu}_n]$$

The maximum variance is achieved when only one of the samples have nonzero weight. In that case, $\mathbb{V}[\widehat{\mu}_n] = \sigma^2$.

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.

Solution. Using $\mathbb{E}[\bar{X}_n]=\mu$ and $\mathbb{V}\left[\bar{X}_n\right]=\sigma^2/n$, we obtain

$$\mathbb{E}\left[\frac{\sqrt{n}\left(\bar{X}_n-\mu\right)}{\sigma}\right]=\frac{\sqrt{n}}{\sigma}\mathbb{E}\left[(\bar{X}_n-\mu)\right]=\frac{\sqrt{n}}{\sigma}(\mu-\mu)=0.$$
$$\mathbb{V}\left[\frac{\sqrt{n}\left(\bar{X}_n-\mu\right)}{\sigma}\right]=\frac{n}{\sigma^2}\mathbb{V}\left[(\bar{X}_n-\mu)\right]=\frac{n}{\sigma^2}\mathbb{V}\left[\bar{X}\right]=\frac{n}{\sigma^2}\frac{\sigma^2}{n}=1.$$

5. Let $X_1, X_2, \ldots$ be a sequence of random variables that converges in probability to a constant a, that is $X_n \to a$ 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 $Y_i=\sqrt{X_i}\to\sqrt{a}$ and $Y_i'=a/X_i\to 1$ converge in probability.

Solution. The key for this problem is to try and rewrite the convergence in probability statement for each of these transformed random variables as a statement about the convergence in probability of X_n to a. To demonstrate, let us show $Y_i\to\sqrt{a}$ in probability. For any $\epsilon>0$,

$$\begin{aligned}\mathbb{P}\left(\left|\sqrt{X_n}-\sqrt{a}\right|>\epsilon\right)&=\mathbb{P}\left(\left|\sqrt{X_n}-\sqrt{a}\right|\left|\sqrt{X_n}+\sqrt{a}\right|>\epsilon\left|\sqrt{X_n}+\sqrt{a}\right|\right)\\&=\mathbb{P}\left(|X_n-a|>\epsilon\left|\sqrt{X_n}+\sqrt{a}\right|\right)\\&\leq\mathbb{P}\left(|X_n-a|>\epsilon\sqrt{a}\right)\to 0,\end{aligned}$$

as $n\to\infty$, since $X_n\to a$ in probability. Thus $\sqrt{X_n}\to\sqrt{a}$ in probability. Now, let's show that $Y_i'\to 1$ in probability. For any $\epsilon>0$, we have that:

$$\begin{aligned}\mathbb{P}\left(\left|\frac{a}{X_n}-1\right|\leq\epsilon\right)&=\mathbb{P}\left(\frac{a}{1+\epsilon}\leq X_n\leq\frac{a}{1-\epsilon}\right)\\&=\mathbb{P}\left(a-\frac{a\epsilon}{1+\epsilon}\leq X_n\leq a+\frac{a\epsilon}{1-\epsilon}\right)\\&\geq\mathbb{P}\left(a-\frac{a\epsilon}{1+\epsilon}\leq X_n\leq a+\frac{a\epsilon}{1+\epsilon}\right)\qquad\text{Use: }\left(a+\frac{a\epsilon}{1+\epsilon}<a+\frac{a\epsilon}{1-\epsilon}\right)\\&=\mathbb{P}\left(|X_n-a|\leq\frac{a\epsilon}{1+\epsilon}\right)\to 1,\end{aligned}$$

as $n\to\infty$, since $X_n\to$ a in probability. Thus $a/X_n\to 1$ in probability.