Statistical and Mathematical Methods DS-GA 1002
, 2019

Recitation Session 10 Solutions

Problem

1. *Sample Average.* Suppose $\{X_1, \cdots, X_n\}$ are n independent identically distributed (they are of the same distribution) random variables with finite expectation μ, variance σ^2, and cdf $F_X(x)$. Denote the sample average as

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

a. Compute the expectation of $\overline{X}_n$.

b. Compute the variance of $\overline{X}_n$.

c. Use (a) and (b) to compute the mean and variance of $F_n(x)$ for any fixed x. The answer should be written in terms of $F_X(x)$.

Solution

a.

$$E(\overline{X}_n) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n}(E(X_i)) = \frac{1}{n}n\mu = \mu$$

b.

$$Var(\overline{X}_n) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n}(Var(X_i)) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

c.

$$E(X) = E(F_n(x)) = E(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}) = \frac{1}{n}\sum_{i=1}^{n} E(1_{\{X_i \leq x\}}) = \frac{1}{n}\sum_{i=1}^{n} F_X(x) = \frac{1}{n}nF_X(x) = F_X(x)$$

$$\begin{aligned}
Var(E) &= Var(F_n(x)) \\
&= Var(\frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \leq x\}}) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var(1_{\{X_i \leq x\}}) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} F_X(x)(1 - F_X(x)) \\
&= \frac{1}{n^2}nF_X(x)(1 - F_X(x)) \\
&= \frac{1}{n}F_X(x)(1 - F_X(x))
\end{aligned}$$

2. *Convergence.* Given the empirical cdf

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \le x\}$$

where $\{X_i\}_{i=1}^n$ are n i.i.d. random variables with cdf $F_X(x)$.

a. Show that for any fixed x, $F_n(x)$ converges to $F_X(x)$ in mean squares error as $n \to \infty$.

b. Use (a) to show that $F_n(x)$ converges to $F_X(x)$ in probability as $n \to \infty$.

Solution

a.

$$\mathrm{E}(F_n(x)) = F_X(x)$$
$$Var(F_n(x)) = F_X(x)(1 - F_X(x))$$

$$\mathrm{E}(F_n(x) - F_X(x))^2 = \mathrm{E}(F_n(x) - E(F_n(x)))^2 = Var(F_n(x)) = \frac{1}{n}F_X(x)(1 - F_X(x))$$

When $n \to \infty, \mathrm{E}(F_n(x) - F_X(x))^2 = \frac{1}{n}F_X(x)(1 - F_X(x)) \to 0$. Thus, for any fixed x, $F_n(x)$ converges to $F_X(x)$ in mean squares error as $n \to \infty$.

b.

$$\Pr(|F_n(x) - F_X(x)| \ge \epsilon) \le \frac{\mathrm{E}(F_n(x) - F_X(x))^2}{\epsilon^2} \to 0$$

Thus, $F_n(x)$ converges to $F_X(x)$ in probability as $n \to \infty$.

3. *Sample Variance.* v

Solution

a.

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 &= \sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2) \\
&= \sum_{i=1}^{n}(X_i^2) - \sum_{i=1}^{n}(2X_i\overline{X}_n - \overline{X}_n^2) \\
&= \sum_{i=1}^{n}(X_i^2) - (\overline{X}_n).\sum_{i=1}^{n}(2X_i - \overline{X}_n) \\
&= \sum_{i=1}^{n}(X_i^2) - \overline{X}_n.(2\sum_{i=1}^{n}X_i - \sum_{i=1}^{n}\overline{X}_n) \\
&= \sum_{i=1}^{n}(X_i^2) - \overline{X}_n.(2n(\frac{1}{n}\sum_{i=1}^{n}X_i) - n\overline{X}_i) \\
&= \sum_{i=1}^{n}(X_i^2) - \overline{X}_n.(2n\overline{X}_n - n\overline{X}_n) \\
&= \sum_{i=1}^{n}(X_i^2) - n(\overline{X}_n)^2
\end{aligned}$$

b.

$$\mathrm{E}(S_n^2) = \mathrm{E}(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2) = \frac{1}{n-1}\mathrm{E}[\sum_{i=1}^{n}(X_i^2) - n(\overline{X}_n)^2] = \frac{1}{n-1}\{\sum_{i=1}^{n}\mathrm{E}(X_i^2) - n\,\mathrm{E}(\overline{X}_n^2)\}$$

$$\mathrm{E}(X_i^2) = V(X_i^2) - (\mathrm{E}(X_i))^2 = \sigma^2 - \mu^2$$

$$\mathrm{E}(\overline{X}_n^2) = V(\overline{X}_n) - [\mathrm{E}(\overline{X}_n)]^2 = \frac{\sigma^2}{n} - \mu^2$$

$$\begin{aligned}
\mathrm{E}(S_n^2) &= \frac{1}{n-1}\{\sum_{i=1}^{n}\mathrm{E}(X_i^2) - n\,\mathrm{E}(\overline{X}_n^2)\} \\
&= \frac{1}{n-1}(n\cdot\mathrm{E}(X_i^2) - n\,\mathrm{E}(\overline{X}_n^2)) \\
&= \frac{1}{n-1}(n(\sigma^2 - \mu^2) - n(\frac{\sigma^2}{n} - \mu^2)) \\
&= \frac{1}{n-1}(n\sigma^2 - n\mu^2 - \sigma^2 + n\mu^2) \\
&= \frac{n-1}{n-1}\sigma^2 \\
&= \sigma^2
\end{aligned}$$

By weak law of large number:

$$\Pr(|S_n^2 - \mathrm{E}(S_n^2)| \geq \epsilon) \leq \frac{\mathrm{E}(S_n^2 - \mathrm{E}(S_n^2))^2}{\epsilon^2}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{n\to\infty} \mathrm{E}(X_i^2)$$

This is because, $\mathrm{E}\, X_i^4 < \infty$ so $Var(X_i^2) < \infty$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{n\to\infty} \mathrm{E}(X_i)$$

When n goes to ∞

$$S_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2\right) \xrightarrow{n\to\infty} \frac{n}{n-1}(\mathrm{E}(X_i^2) - (\mathrm{E}(X_i))^2) = \frac{n}{n-1} V(X_i) \xrightarrow{n\to\infty} \sigma^2$$

Thus, S_n^2 converges to σ^2 in probability.