9th problem set for Probability and Statistics — April 29/30

Summary

- We examine a sequence of i.i.d. random variables with the same distribution, e.g., $Geom(\theta)$, $U(0, \theta)$, where θ is a parameter.
- We write $X_1, \ldots, X_n \sim F_\theta$, called a **random sample** from F_θ (parametric model).
- We measure $X_1 = x_1, \ldots,$ and want to estimate θ.
- $\hat{\theta}$ ... some method to estimate θ using the measured data $(X_1, \ldots, X_n)$, called an *estimator*.
- $\widehat{m}_r(\theta) = \frac{1}{n}\sum_{i=1}^{n} X_i^r$ ... r-th sample moment, a random variable, a function of our observed sample (i.e., a statistic).
- **Bias:** $\mathbb{E}_\theta(\hat{\theta} - \theta)$ ... θ true parameter, $\hat{\theta}$ our estimate (random variable as it depends on observed data).
- Estimator is **unbiased:** $bias = 0$ for all $\theta \in \Theta$
- Estimator is **asymptotically unbiased:** bias converges to 0, i.e., $\mathbb{E}_\theta(\hat{\theta}) \to \theta$ for all $\theta \in \Theta$
- Estimator is **consistent:** $\hat{\theta} \xrightarrow{P} \theta$: for all $\varepsilon > 0$ and all $\theta \in \Theta$, $P(|\hat{\theta} - \theta| > \varepsilon) \to 0$
- **MSE (Mean Square Error):** $\mathbb{E}_\theta((\hat{\theta} - \theta)^2)$
- Theorem: $MSE = bias(\hat{\theta})^2 + \mathrm{var}(\hat{\theta})$.

Estimators and their properties

1. For the exponential distribution $Exp(\vartheta)$, $\vartheta \in \Theta = (0, \infty)$ consider the estimator

$$\hat{\vartheta} = 1/\bar{X}_n = n/(X_1 + \cdots + X_n),$$

and recall that $\mathbb{E}(X) = 1/\vartheta$. Is it unbiased? Hint: you will need to use the Gamma distribution.

2. Consider the family of estimators for the cdf defined, for each $x \in \mathbb{R}$, by

$$\hat{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x),$$

where $I(X_i \le x)$ is 1 if $X_i \le x$, and 0 otherwise. Compute its bias, variance and MSE. Is $\hat{F}_n(x)$ consistent?

3. Prove that $\hat{S}_n^2$ (the corrected sample variance, i.e. the sample variance times $n/(n-1)$) is a consistent estimator. More generally, if an estimator is unbiased and has a vanishing variance (with n going to infinity) then show that the estimator is consistent.

4. Assume a sample of continuous random variables: $X_1, X_2, \ldots, X_n$, where $E[X_i] = \mu$, $\mathrm{var}[X_i] = \sigma^2 > 0$. Consider the following estimators: $\hat{\mu}_{1,n} = X_n$, $\hat{\mu}_{2,n} = \frac{1}{n+1}\sum_{i=1}^{n} X_i$.

(a) Are $\hat{\mu}_{1,n}$ and $\hat{\mu}_{2,n}$ unbiased?
(b) Are $\hat{\mu}_{1,n}$ and $\hat{\mu}_{2,n}$ consistent?

5. Consider a sample of random variables: $X_1, X_2, \ldots, X_n$, where $n > 10$, $E[X_i] = \mu$, $\text{var}[X_i] = \sigma^2 > 0$ and the estimator $\hat{\mu}_n = \frac{1}{n-10} \sum_{i=11}^{n} X_i$. Then calculate:

(a) The bias of $\hat{\mu}_n$.
(b) The variance of $\hat{\mu}_n$.
(c) The MSE of of $\hat{\mu}_n$.

6. In a TV-show the host picks independently n random reals numbers uniformly from $[0, \theta]$ (where θ is known only to the host), and reveals them to the players. Based on the sample, the players have to guess θ. The first player, guided by LLN, guesses θ to be twice the sample mean, whereas the second player guesses θ to be the maximum value of the sample. Decide for each estimator whether it is consistent and unbiased. Calculate the MSE of both and compare them.