UNIVERSITY OF TEXAS AT AUSTIN

HW Assignment 1

Prerequisite material.

Please, provide your **complete solutions** to the following problems. Final answers only, even if correct will earn zero points for those problems.

Problem 1.1. (5 points) Provide the definition of the *bias* of an estimator. What does it mean for the estimator to be *unbiased*? What about *biased*?

Solution: Let $\hat{\theta}$ be a point estimator for the parameter θ. The *bias* of this estimator is defined as

$$bias(\hat{\theta}) = \mathbb{E}_\theta[\hat{\theta}] - \theta$$

We say that the estimator is *unbiased* if its *bias* is equal to zero. Otherwise, the estimator is said to be *biased*.

Problem 1.2. (5 points) Provide the definition of the *mean squared error (MSE)* of an estimator.

Solution: Let $\hat{\theta}$ be a point estimator for the parameter θ. The *mean squared error* of this estimator is defined as

$$\mathbb{E}_\theta[(\hat{\theta} - \theta)^2].$$

Problem 1.3. (10 points) Show that, for a point estimator $\hat{\theta}$,

$$MSE[\hat{\theta}] = Var_\theta[\hat{\theta}] + (bias(\hat{\theta}))^2.$$

Solution:

$$\begin{aligned}
MSE[\hat{\theta}] &= \mathbb{E}_\theta[(\hat{\theta} - \theta)^2] \\
&= \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}] + \mathbb{E}_\theta[\hat{\theta}] - \theta)^2] \\
&= \mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}])^2] + 2\mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}])(\mathbb{E}_\theta[\hat{\theta}] - \theta)] + (\mathbb{E}_\theta[\hat{\theta}] - \theta)^2 \\
&= Var_\theta[\hat{\theta}] + 2\mathbb{E}_\theta[(\hat{\theta} - \mathbb{E}_\theta[\hat{\theta}])(\mathbb{E}_\theta[\hat{\theta}] - \theta)] + (bias(\hat{\theta}))^2
\end{aligned}$$

The middle term is obviously zero which completes the proof.

Problem 1.4. (10 points) The Pareto distribution with parameters α and θ has the distribution function

$$F(x) = 1 - \left(\frac{\theta}{x + \theta}\right)^\alpha.$$

For integer k, its k^{th} moment is

$$\mathbb{E}[X^k] = \frac{\theta^k k!}{(\alpha - 1)\dots(\alpha - k)}$$

A random variable X has a two-parameter Pareto distribution with parameters $\alpha = 4$ and θ (unknown, and to be estimated). Let $\hat{\theta} = 3X$ be our proposed estimator for the θ parameter, based on a random sample consisting of a single measurement. Find the mean squared error of this estimator.

Solution:

$$\begin{aligned}MSE_{\hat{\theta}}(\theta) &= Var_\theta[\hat{\theta}] + (bias(\hat{\theta}))^2\\ &= Var_\theta[3X] + (\mathbb{E}_\theta[3X]-\theta)^2\\ &= 9Var_\theta[X] + (3\mathbb{E}_\theta[X]-\theta)^2\\ &= 9\cdot\frac{4\theta^2}{3^2\cdot 2} - (3\cdot\frac{\theta}{3}-\theta)^2\\ &= 2\theta^2.\end{aligned}$$

Problem 1.5. (10 points) The gamma distribution with parameters α and β has mean $\alpha\beta$ and variance $\alpha\beta^2$.

Let the random variable X have the Gamma distribution with parameters $\alpha = 3$ and θ unknown (and to be estimated). A proposed estimator for the parameter θ based on a single observation X_1 of the above distribution is $\hat{\theta} = \frac{1}{3}X_1$. What is the **mean-squared error** of this estimator?

Solution:

$$\begin{aligned}MSE[\hat{\theta}] &= \mathbb{E}_\theta[(\hat{\theta}-\theta)^2]\\ &= \mathbb{E}_\theta\left[(\tfrac{1}{3}X_1-\theta)^2\right]\\ &= \frac{1}{9}\mathbb{E}_\theta[(X_1)^2] - \frac{2}{3}\theta\mathbb{E}_\theta[X_1] + \theta^2\\ &= \frac{1}{9}(\alpha\theta^2 + (\alpha\theta)^2) - \frac{2}{3}(\alpha\theta)\theta + \theta^2\\ &= \frac{1}{9}(3\theta^2 + 9\theta^2) - \frac{2}{3}(3\theta)\theta + \theta^2 = \frac{1}{3}\theta^2.\end{aligned}$$

Problem 1.6. (10 points) Let Y_1, Y_2 be a random sample from the exponential distribution with the unknown parameter θ. The estimator $\hat{\theta}_2 = cY_{(1)}$ for θ is proposed. Find the constant c such that $\hat{\theta}_2$ is an unbiased estimator of θ.

Solution: $Y_{(1)}$ is the first order statistic of the random sample (Y_1, Y_2) so it can be written as

$$Y_{(1)} = \min(Y_1, Y_2).$$

Obviously, the support of the random variable $Y_{(1)}$ is $[0,\infty)$. For minima of families of random variables, it's more expedient to consider the survival function rather than the cumulative distribution function. As a reminder, the survival function $S_X : \mathbb{R} \to [0,1]$ of a random variable X is defined as

$$S_X(x) = 1 - F_X(x) \tag{1.1}$$

where F_X denotes the cumulative distribution function of the random variable X.

Then, we have for every $y > 0$,

$$S_{Y_{(1)}}(y) = \mathbb{P}[Y_{(1)} > y] = \mathbb{P}[\min(Y_1, Y_2) > y] = \mathbb{P}[Y_1 > y, Y_2 > y] = \mathbb{P}[Y_1 > y]\mathbb{P}[Y_2 > y].$$

By definition, for $Y \sim Exponential(mean = \theta)$, we have

$$F_Y(y) = 1 - e^{-y/\theta} \quad \Rightarrow \quad S_Y(y) = e^{-y/\theta} \quad \text{for } y > 0.$$

Hence,

$$S_{Y_{(1)}}(y) = \mathbb{P}[Y_1 > y]\mathbb{P}[Y_2 > y] = e^{-y/\theta}e^{-y/\theta} = e^{-y/(\theta/2)}.$$

We can recognize the distribution of $Y_{(1)}$ as exponential with mean $\theta/2$. Therefore, our unbiasedness condition becomes

$$\theta = \mathbb{E}[\hat{\theta}_2] = \mathbb{E}[cY_{(1)}] = c\mathbb{E}[Y_{(1)}] = c\left(\frac{\theta}{2}\right) \quad \Rightarrow \quad c = 2.$$