Statistics and Sampling Distributions

In a typical statistical problem, we have a random variable whose underlying probability distribution is known, but its parameters are unknown; e.g.,

$$X \sim Exp(\mu)$$

$$X \sim N(\mu, \sigma^2)$$

$$X \sim Poisson(\lambda)$$

* We represent the parameters of these distributions by θ; e.g.,

$$X \sim Exp(\theta)$$

$$X \sim N(\theta_1, \theta_2)$$

$$X \sim Poisson(\theta)$$

* We call θ a parameter of the distribution.

* The goal is to estimate the unknown θ based on the available data or observations.

* A statistic (singular) is a function of the observations which summarizes (reduces) the data.

* Let us assume that

sample size — read as "parametrized by θ"

$$X_1, X_2, \dots, X_n \sim f(x;\theta)$$

The sequence of random variables $X_1, X_2, \dots, X_n$ is called a sample, where each of the random variables is called observation.

* Once the sample is realized, we denote the numerical values taken by each observation using lower-case Roman letters.

$$x_1, x_2, \dots, x_n$$

these are realized observations once the sample is drawn

* Note that a statistic is always a function of the samples:

$$T = g(X_1, X_2, \dots, X_n)$$

e.g.,

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$X_{(n)} = \max(X_1, X_2, \dots, X_n)$$

* Estimator vs. Estimate:

While we call T an estimator of θ, we call its realization t an estimate of θ.

* Once the sample is drawn, then t is called the realization of the statistics T, where

$$t = g(x_1, x_2, \dots, x_n)$$

Note: A sample is random if $X_1, X_2, \dots, X_n \overset{iid}{\sim} f(x;\theta)$

Order Statistics

★ Let $X_1, X_2, \cdots, X_n$ be a random sample from the distribution $f(x;\theta)$.

Then, $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ represents the order statistic associated to the random sample.

Q. Find the sampling distribution of maximum order statistics.

Let $Y = \max(X_1, X_2, \cdots, X_n)$

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(\max(X_1, X_2, \cdots, X_n) \leq y) \\
&= P(X_1 \leq y, X_2 \leq y, \cdots, X_n \leq y) \\
&\overset{\text{independence}}{=} P(X_1 \leq y) P(X_2 \leq y) \cdots P(X_n \leq y) \\
&= F_X(y) \cdot F_X(y) \cdots F_X(y) \\
&= (F_X(y))^n
\end{aligned}
$$

$$\therefore \quad F_Y(y) = (F_X(y))^n$$

$$f_Y(y) = n\, f_X(y)\, (F_X(y))^{n-1}$$

Exp. $X_1, X_2, \dots, X_n \overset{iid}{\sim} \text{Uniform}(0,1)$

$$f_Y(y) = ny^{n-1}$$

alternatively

$$f_{X_{(n)}}(x) = nx^{n-1}$$