Theorems In Statistics

Animesh Renanse

May 11, 2020

1 The Triads of Statistics

The three important theorems of mathematical statistics are[1].

1.1 Strong Law of Large Numbers

Given that $\{X_n\}$ *is a sequence of i.i.d. Random Variables and define* $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$*, then we have the following result:*

$$\left\{\overline{X_n}\right\} \to \mu.$$

almost surely/with probability 1.

Proof: Since we need to show that $\left\{\overline{X_n}\right\} \to \mu$ with probability 1, or, we can show that:

$$P\left[\mid \overline{X_n} - \mu \mid > \epsilon\right] \to 0$$

Now we can Chebyshev's Inequality to get:

$$P\left[\mid \overline{X_n} - \mu \mid > \epsilon\right] \leq \frac{Var(\overline{X_n})}{\epsilon^2}.$$

Note that $Var(\overline{X_n})$ is equal to $\frac{\sigma^2}{n}$, therefore, we can write that:

$$P\left[\mid \overline{X_n} - \mu \mid > \epsilon\right] \leq \frac{\sigma^2}{n\epsilon^2}.$$

To that $\frac{\sigma^2}{n\epsilon^2} \to 0$, therefore $P\left[\mid \overline{X_n} - \mu \mid\right] \to 0$. Hence Proved.

1.2 Central Limit Theorem

Given that $\{X_n\}$ *be a sequence of i.i.d. Random variables with finite* μ *and finite* σ^2*, it can be shown that,*

$$P\left(\frac{\sqrt{n}\left(\overline{X_n} - \mu\right)}{\sigma} \leq a\right) \to \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

OR,

$$\frac{\sqrt{n}\left(\overline{X_n} - \mu\right)}{\sigma} \to N(0,1).$$

1.3 Sampling Distributions

Let $X_1, X_2, \ldots X_n$ *be i.i.d. Random variables, it can then be shown that*

$$\overline{X_n} \sim N(\mu, \frac{\sigma^2}{n}).$$

and

$$\frac{(n-1)\,S^2}{\sigma^2} \to \chi^2_{n-1}.$$

[1]As per my thinking

where, $\overline{X_n}$ *and* S^2 *are*

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

and

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X_n}\right).$$

In other words, one can say that Sample Mean follows a Normal Distribution with mean as that of Population itself and variance exactly equal to $\frac{\sigma^2}{n}$ *and the Sample Variance follows a* χ^2_{n-1} *distribution.*

2 Estimation

2.1 Unbiased Estimators

A statistic $T(X)$ *is said to be unbiased estimator of* $g(\theta)$ *if,*

$$\mathbb{E}_\theta\left[T(X)\right] = g(\theta).$$

for all θ *in* Θ.

Also, if $\mathbb{E}_\theta\left[T(X)\right] = g(\theta) + b(\theta)$*, then we can say that* $\mathbb{E}_\theta[T(X)]$ *is a biased estimator of* $g(\theta)$ *.*

2.2 Consistent Estimator

A statistic $T_n(\theta) = T(\theta)$ *is said to be consistent estimator for* $g(\theta)$ *if for every* $\epsilon > 0$,

$$P\left[\mid T_n - g(\theta) \mid > \epsilon\right] \to 0.$$

which is, in the long run, the value of estimator and the estimand comes infinitesimally close to each other.

Note that checking whether any given estimator for $g(\theta)$ is consistent or not is not an easy task, to help us with it we have some useful theorems:

Theorem 1:- *If* $\mathbb{E}\left[T_n\right] = \theta_n \to \theta$ *and* $Var(T_n) = \sigma_n^2 \to 0$ *as* $n \to \infty$ *then* T_n *is a consistent estimator for* θ.

Proof:

We can write,

$$\mid T_n - \theta \mid \leq \mid T_n - \theta_n \mid + \mid \theta_n - \theta \mid.$$

$$P\left(\mid T_n - \theta \mid > \epsilon\right) \leq P\left(\mid T_n - \theta_n \mid + \mid \theta_n - \theta \mid > \epsilon\right)$$

or,

$$P\left(\mid T_n - \theta \mid > \epsilon\right) \leq P\left(\mid T_n - \theta_n \mid > \epsilon - \mid \theta_n - \theta \mid\right).$$

Using Chebyshev's Inequality,

$$P\left(\mid T_n - \theta_n \mid > \epsilon\right) \leq P\left(\mid T_n - \theta_n \mid > \epsilon - \mid \theta_n - \theta \mid\right) \leq \frac{\sigma_n^2}{(\epsilon - \mid \theta_n - \theta \mid)^2} \to 0.$$

Hence proved.