Robert Morris

CSCI Data Analytics

Stability of Statistical Estimators

notes: Sample mean $\rightarrow \hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

"Biased" Sample Variance $\rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ $(\sigma^2 = E[(x-\mu)^2])$

"Unbiased" Sample Variance $\rightarrow S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

Defn: The bias of an estimator $\hat{\theta}$ is $\text{bias}(\hat{\theta}) = E\hat{\theta} - \theta$

Defn: The estimator $\hat{\theta}$ is unbiased if $\text{bias}(\hat{\theta}) = 0$

$E(\sum x_i) = \sum E(x_i)$, $E(cX) = cE(X)$,

$Var(X) = E(X^2) - [E(X)]^2$

$E(X^2) = \sigma^2 + \mu^2$

$Var(\bar{X}) = E(\bar{X}^2) - [E(\bar{X})]^2$

$E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$

1. Prove theoretically that the sample variance, s^2, is unbiased.

$E(S^2) = E\left[\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(X_i-\bar{X})^2\right]$

$= \left(\frac{1}{n-1}\right)E\left[\sum(X_i-\bar{X})^2\right]$

$= \left(\frac{1}{n-1}\right)E\left[\sum((X_i-\bar{x})(X_i-\bar{x}))\right]$

$= \left(\frac{1}{n-1}\right)E\left[\sum(X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right] = \left(\frac{1}{n-1}\right)E\left[\sum X_i^2 - \overset{\text{constants}}{\sum 2X_i\bar{X}} + \sum\bar{X}^2\right]$

note: $\bar{X} = \frac{\sum X_i}{n} = n\bar{X}$

$= \left(\frac{1}{n-1}\right)E\left[\sum X_i^2 - 2\bar{X}\sum X_i + n\bar{X}^2\right]$

$= \left(\frac{1}{n-1}\right)E\left[\sum X_i^2 - 2\bar{X}n\bar{X} + n\bar{X}^2\right] = E\left[\sum X_i^2 - n\bar{X}^2\right]$

$-2n\bar{X}^2 + 1n\bar{X}^2 = \sum E(X_i^2) - E(n\bar{X}^2)$

$= \left(\frac{1}{n-1}\right)\sum E(X_i^2) - nE(\bar{X}^2)$

note: $E(x_i^2) = \sigma^2 + \mu^2$
$E(\bar{x}^2) = \frac{\sigma^2}{n} + \mu^2$

$= \left(\frac{1}{n-1}\right)\sum(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)$

$= \left(\frac{1}{n-1}\right) n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2} \rightarrow = \sigma^2(n-1)\left(\frac{1}{n-1}\right)$

since $\left(\frac{1}{n-1}\right)$ and $(n-1) = 1$, we are left with σ^2

$\therefore$ Hence $E(S^2)$ is an unbiased estimator of σ^2

2. Prove theoretically that the sample variance, s^2, is consistent.

$Var(S^2) = var\left(\frac{\sum(x_i-\bar{x})^2}{n-1}\right) = \frac{1}{(n-1)^2}Var\left(\sum(x_i-\bar{x})^2\right) = \frac{2\sigma^4}{n-1}$

(n becomes very large) $\lim_{n\to\infty} Var(S^2) = 0 \;\therefore\; S^2$ is consistant

★ note: $\sum \rightarrow$ all cases of $\sum = \sum_{i=1}^{n}$

note: $\bar{X} = \frac{\sum x_i}{n} = n\bar{X} = \sum x_i$