Lecture 1 : Statistical Review

Def: A random variable, x, is a variable ~~whose~~ which is an outcome of an experiment determined in part by chance. The p.d.f. $f(x)$ gives the probability that the variable assumes the value x.

$$\mu_x \equiv E(x) = \int x f(x)\, dx \quad \text{or} \quad \sum_i x_i f(x_i)$$

$$V(x) \equiv \sigma_x^2 = E(x-\mu_x)^2$$

$$= \int (x-\mu_x)^2 f(x)\, dx \quad \text{or} \quad \sum_i (x_i-\mu_x)^2 f(x)$$

Note:

$$\sigma_x^2 = E(x^2 - 2\mu_x x + \mu_x^2)$$

$$= E(x^2) - 2\mu_x E(x) + \mu_x^2 = E(x^2) - 2\mu_x^2 + \mu_x^2$$

$$= E(x^2) - \mu_x^2$$

Property:

1. $E(\sum_i a_i X_i) = \sum_i a_i E(x_i) = \sum_i a_i \mu_{x_i}$

~~$V(ax) = a^2 V(X)$~~ $V(a+bX) = b^2 V(X)$

2. If $y = h(x)$ nonlinear, $E(y) \neq h(E(x))$.

3. $E(x_i x_2 \cdots x_n) = E(x_1)E(x_2)\cdots E(x_n)$
iff x_i's are independent.

We want to estimate parameter α from a random sample. Properties of the estimator $\hat{\alpha}$ are

Unbiasedness $\quad E(\hat{\alpha}) = \alpha$

Efficiency $\quad$ i) $E(\hat{\alpha}) = \alpha$
ii) $V(\hat{\alpha}) \leq V(\alpha')$ where α' is any other unbiased estimator

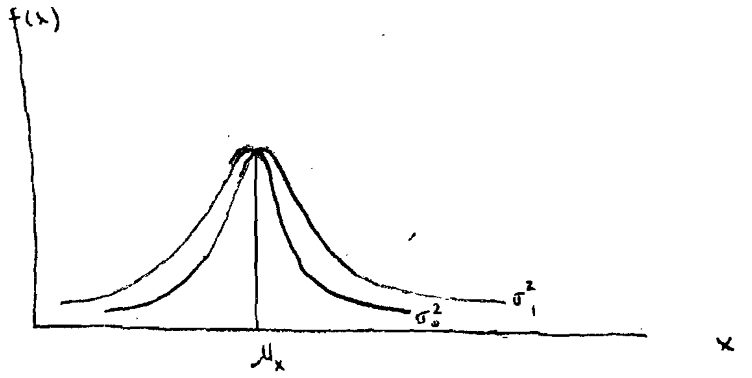
Consistency $\quad \lim\limits_{N \to \infty} \text{Prob}(|\hat{\alpha} - \alpha| > \varepsilon) = 0$ for small ε.

or in shorthand

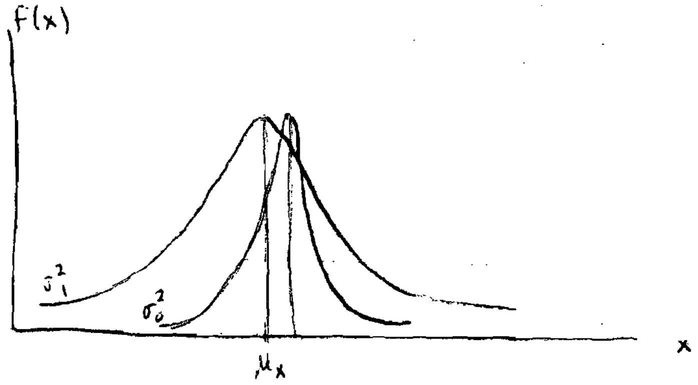
$$\text{Plim } \hat{\alpha} = \alpha$$

Consistency is a large sample property which is concerned with $N \to \infty$. Unbiasedness & efficiency are small sample characteristics re: replications of given finite sampling.

Unbiasedness / Efficiency



$\sigma_1^2 > \sigma_0^2$

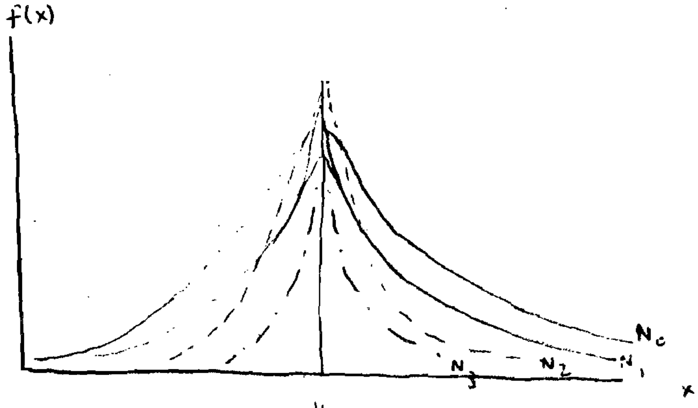


$\sigma_0^2 < \sigma_1^2$

$E(x_0) \neq \mu_x$

$E(x_1) = \mu_x$

Consistency



$N_0 < N_1 < N_2 < N_3$

The probability mass centers on μ_x as $N \to \infty$.

Mean Square Error

$$MSE(\hat\theta) = E(\hat\theta-\theta)^2 = E\Big(\overbrace{\hat\theta - E(\hat\theta)}^{a} + \overbrace{E(\hat\theta) - \theta}^{b}\Big)^2$$

$$= \overbrace{E(\hat\theta - E(\hat\theta))^2}^{a} + \overbrace{\big(E(\hat\theta)-\theta\big)^2}^{b} + \overbrace{2E\Big[\big(\hat\theta - E(\hat\theta)\big)\big(E(\hat\theta)-\theta\big)\Big]}^{ab}$$

$$= E(\hat\theta - E(\hat\theta))^2 + (E(\hat\theta)-\theta)^2$$

$$MSE(\hat\theta) = \text{Variance}(\hat\theta) + \text{Bias}(\hat\theta)^2$$

Note: An efficient (unbiased) estimator may have a larger MSE than a biased estimator.

Variance of Sums

Let $\{x\}$ be uncorrelated r.v.'s. Then let $y = \sum a_i x_i$:

$$\sigma_y^2 = \sum_i a_i^2 \sigma_i^2$$

Proof:

$$\mu_y = \sum a_i \mu_i$$

$$\sigma_y^2 = E(y-\mu_y)^2 = E\big(a_1(x_1-\mu_1) + \cdots + a_n(x_n-\mu_n)\big)^2$$

$$\sigma_y^2 = E\left(a_1^2 (x_1-\mu_1)^2 + \cdots + a_n^2 (x_n-\mu_n)^2 + \text{cross products} \right)$$

$$= \sum_i a_i^2 \sigma_i^2$$

qed.

<u>Covariance</u> : Consider r.v. x, y with μ_x, μ_y. Then

$$\sigma_{xy} = E(x-\mu_x)(y-\mu_y).$$

i) relates only to linear dependence

ii) $\sigma_{xy} = E(xy) - \mu_x \mu_y$.

$$\hat{\sigma}_{xy} = \frac{1}{T-1} \sum_t^T (x_t - \bar{x})(y_t - \bar{y}) \qquad \bar{x} = \frac{\sum_t x_t}{T}$$

$$\bar{y} = \frac{\sum_t y_t}{T}$$

i) $E(\hat{\sigma}_{xy}) = \sigma_{xy}$

ii) $\text{plim } \hat{\sigma}_{xy} = \sigma_{xy}$

<u>Correlation Coefficient</u>

Measures the strength of association, normalized to be invariant to units of measurement, i.e. by variances.

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \qquad \sigma_x = \left[E(x-\mu_x)^2\right]^{\frac{1}{2}}$$

$$\sigma_y = \left[E(y-\mu_y)^2\right]^{\frac{1}{2}}$$

Notes:

1. Let $x' = \lambda x$. Then $\mu_{x'} = \lambda \mu_x$

$$\sigma_{x'y} = E\,\lambda(x-\mu_x)(y-\mu_y) = \lambda \sigma_{xy}$$

$$\sigma_{x'}^2 = E\,(\lambda(x-\mu_x)(y-\mu_y))^2 = \lambda^2 \sigma_x^2 \Rightarrow \sigma_{x'} = \lambda \sigma_x$$

$$\therefore \quad \rho_{x'y} = \frac{\sigma_{x'y}}{\sigma_{x'}\sigma_y} = \frac{\lambda\sigma_{xy}}{\lambda\sigma_x\sigma_y} = \rho_{xy}.$$

2. $-1 \le \rho_{xy} \le 1$

$$\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} \qquad \hat{\sigma}_x = \left[\frac{1}{T-1}\sum_t (x_t - \bar{x})^2\right]^{\frac{1}{2}}$$

Note that

$$\text{plim } \bar{x} = \mu_x$$
$$\text{plim } \bar{y} = \mu_y$$

so

$$\text{plim } \hat{\sigma}_x = \sigma_x$$
$$\text{plim } \hat{\sigma}_y = \sigma_y$$
$$\text{plim } \hat{\sigma}_{xy} = \sigma_{xy}$$

$\therefore$ intuitively it follows that $\text{plim } \hat{\rho}_{xy} = \rho_{xy}$. But since ρ_{xy} is nonlinear function, $\hat{\rho}_{xy}$ is not an unbiased estimator of ρ_{xy}.