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

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

$Var(aX) = a^2 Var(x)$

If sample is iid, $Var(\bar{X}) = Var(X)$

CoVariance: $Cov(X+Y, Z) = Cov(X,Z) + Cov(Y,Z)$

$Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$

$= E[XY] - E[X]E[Y]$

$E(X(Y - E[Y])) = E[Y(X - E[X])]$

$Cov(a, X) = 0$ if a is a constant

$Cov(X,Y)$ is 0 if X is balanced neg and pos

$Cov(X,Y)$ is 0 if $X \perp\!\!\!\perp Y$

Independence

If $X \perp\!\!\!\perp Y$, then $E[XY] = E[X]E[Y]$ by dist and EDO

discrete: $P(X,Y) = P(X) \cdot P(Y)$

continuous: $f(x,y) = f(x) \cdot f(y)$

similarly $P(X|Y) = P(x)$ AND $P(Y|X) = P(y)$

LAW of Iterated Expectations

$E[E[Y|X]] = E[Y]$, sum up all $E[Y|X]$ for all x to get avg. value of Y

Also, $E[E[Y|X]|Z] = E[Y|Z]$

Samples: iid if $\rightarrow E[X_i] = E[X_j]$

independent: all $X_i \perp\!\!\!\perp X_j$, all $Cov(X_i, X_j) = 0$

identically dist: all $X_i \sim F_x$, all $Var(X_i) = Var(X_j)$

When sample is iid, an estimator g for target θ is

unbiased: if $E[g] = \theta$ (like $E[\bar{X}] = E[X]$)

consistent: if $g \rightarrow \theta$ as $n \rightarrow \infty$

Strong Law of Large Numbers: $\bar{X}$ is consistent when X_n is iid

Central Limit Theorem

If sample is iid AND $Var[X] < \infty$, then:

$\sqrt{n}(\bar{X} - E[X]) \xrightarrow{d} N(0, Var[X])$

dist of ↑ will begin to resemble ↑

Linear Transformation on dist:

if $Y \sim N(M, \sigma^2)$, then $a + bY \sim N(a + bM, b^2\sigma^2)$

i.e. $\bar{X} \sim (E[X], \frac{1}{n} Var[X])$ as $n \rightarrow \infty$

Ordinary Least Squares

loss function is squared deviation of line $\alpha_0 + \alpha_1 X$ from Y (predicted, real)

$L_{BLP}(\alpha_0, \alpha_1) \equiv (Y - \alpha_0 - \alpha_1 X)^2$ - linear

$L_{BLA}(\alpha_0, \alpha_1) = (E[Y|X] - \alpha_0 - \alpha_1 X)^2$ - non linear

$\alpha_0 = E[Y] - \alpha_1 E[X]$

$\alpha_1 = Cov(X,Y) / Var(X)$ $\Big\} \; Y = \alpha_0 + \alpha_1 X + \epsilon$

ϵ is difference between Y and function, i.e. error

Expected Value:

discrete: $\sum_{x \in supp(x)} x \cdot P(x)$

continuous: $\int_{x \in supp(x)} x \cdot f_x(x)\, dx$

Distributions

Given a joint distribution, a marginal distribution can be found by summing or integrating over other

Dist is $P(X)$ for discrete and $f(X)$ for continuous

$Cov(X, E[Y|X]) = Cov(X, Y)$

when integral, go $\int x \int y \; dy\, dx$

end: determined by solving ec. model

ex: not determined by economic model

mean independence is when $E[U|X] = 0$. Independence > mean

if $E[U|X]$ varies with x, then $U \not\perp\!\!\!\perp X$

if $Var[U|X]$ varies with X, then $U \not\perp\!\!\!\perp X$

if supp(U) varies with X, then $U \not\perp\!\!\!\perp X$

or $E[U|X] = U$ / $E[X|U] = E[X]$

Conditional: $P(X|Y)$

$P(X|Y) = \frac{P(X,Y)}{P(Y)}$

Quick:

r.v. is a function mapping from sample space to a number

sample collection of r.v.

$E[X]$: not random

$E[X|Y]$: random

$E[X|Y=3]$: not random

$E[X|X] = X$

$Var[X|Y=2]$: not random

PLLN: $P(|X - E[X]| > \epsilon) = 0$

Log rules:

$\ln(ab) = \ln(a) + \ln(b)$

$\ln(\frac{a}{b}) = \ln(a) - \ln(b)$

$\ln(a^b) = b \ln a$

causal: effect of X and only X

statistical: do X and Y move together

Identification: Recovering parameters from population

Estimation: Recovering parameters from sample

endogenous: A choice, comes from economic model. in econometric model, means $E[U|X] \neq 0$ (mean independent)

exogenous: random, affects Y but $\perp\!\!\!\perp X, v$. Like rainfall

ex. Y is Income, X is years in school, U is Ability

Y is House Price, X is Safety, U is Hood Quality

OVB leads to overestimates of effect of X on Y

Ommitted Variable Bias:

If W is an ommitted variable, then $U = \gamma W + v$

If $Cov(X, W) = 0$ then OVB $= 0$.

OVB Problems:

1. $W \not\perp\!\!\!\perp X$: creates biased coefficients
2. W directly affects Y: incomplete

if model is $Y = \beta_0 + \beta_1 X + U$

The intercept of β_0 can always be adjusted such that $E[U] = 0$. We can shift and not affect β_1. $Cov(X, U)$ may not be 0. $Cov(X, \epsilon)$ is always 0 so ϵ and U can be different. However, if $\beta_0 = \alpha_0$ and $\beta_1 = \alpha_1$, then $U = \epsilon$