11/3/22

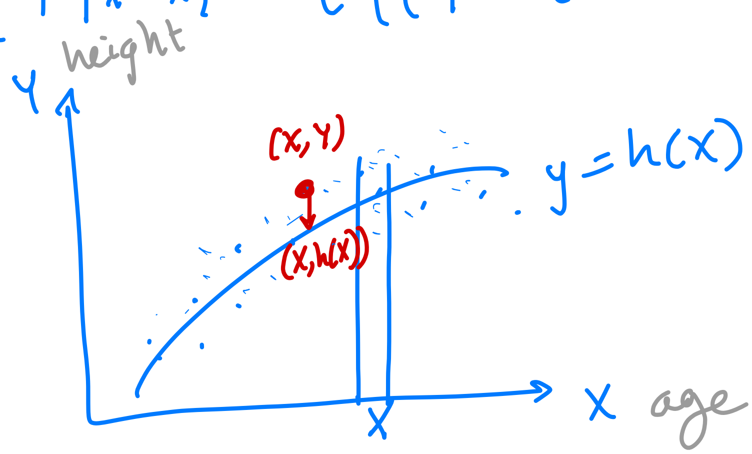
Conditional distribution

$[Y|X] \sim p(y|x)$ discrete

$f(y|x)$ continuous

$$E(Y|X=\overset{\text{fixed}}{x}) = \boxed{\int y\, f(y|x)\,dy = h(x)}$$

$$Var(Y|X=\overset{\text{fixed}}{x}) = E((Y-h(x))^2 \,|\, X=x)) = \int (y-h(x))^2 f(y|x)\,dy$$



2D

$n \to \infty$

$\frac{1}{n}\sum y_i \to E(Y)$ (1)

$\frac{1}{n}\sum h(x_i) \to E(h(X))$

linear: $X \sim N(0,1)$

$[Y|X=x] \sim N(\rho x, 1-\rho^2)$

$h(x) = \rho x$

"Adam"

(1) $E_X \underbrace{E_{Y|X}(Y|X)}_{h(X)} = E(Y)$

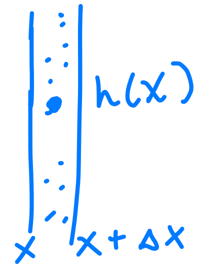
$$E(\underset{\text{capital}}{h(X)}) = \int \underset{\text{lowercase}}{\boxed{h(x)}} f(x)\,dx = \int \boxed{\int y\, f(y|x)\,dy}\, f(x)\,dx = E(Y)$$

★ see plot

$$Y \to h(X)$$

$$E(Y) = E(h(X))$$

summation is unchanged
↳ when converting Y to $h(X)$, not changing sum w/in each slice



"EVE formula"

$$var(Y) = E(Var(Y|X)) + Var(\underbrace{E(Y|X)}_{h(X)})$$

w/in group　　　between group

variance reduction by conditioning

$$\varepsilon = Y - h(x)$$

$$Y = \underset{\text{regression px}}{h(x)} + \underset{\text{error}}{\varepsilon}$$

$$E(\varepsilon) = E(Y - h(x)) = E(Y) - E(h(x)) = 0$$

$$E(\varepsilon g(x)) = E_X E(\varepsilon g(x) | X)$$

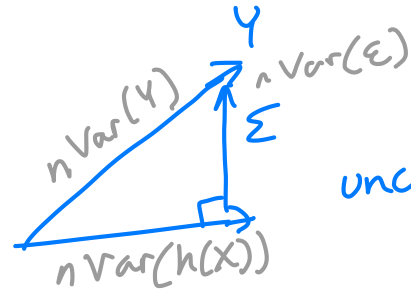
$$E(\varepsilon g(x) | X = x) = g(x) E(\varepsilon | X = x) = g(x) E(Y - h(X) | X = x)$$

$$= 0$$

$$cov(\varepsilon, g(x)) = E(\underbrace{\cancel{\varepsilon g}(x)}_{0}) - E(\underbrace{\cancel{\varepsilon) E(g}}_{0}(x)) = 0$$

	X	Y	$h(X)$	ε
1				
$\vdots$				
i	x_i	y_i	$h(x_i)$	ε_i
$\vdots$				
n				

n-dimensional



uncorrelated

$$\mathrm{Var}(Y) = \mathrm{Var}(h(X)+\varepsilon) = \mathrm{Var}(h(X)) + \mathrm{Var}(\varepsilon) + 2\,\mathrm{Cov}(h(X),\varepsilon)$$

$$= \mathrm{var}(E(Y|X)) + E_X \underbrace{E\left((Y-h(X))^2 \,|\, X\right)}_{\mathrm{var}(Y|X)}$$

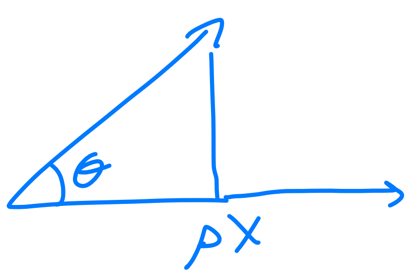
linear: $E(Y|X=x) = \rho x$

$$\mathrm{var}(Y|X=x) = 1-\rho^2$$

$$E(Y) = E\,E(Y|X) = E(\rho X) = \rho\, E(X) = 0$$

$$\mathrm{var}(Y) = E(\mathrm{var}(Y|X)) + \mathrm{var}(E(Y|X))$$

$$= 1-\rho^2+\rho^2 = 1$$



$$Y = \rho X + \varepsilon$$

$$\varepsilon \perp X \qquad \varepsilon \sim N(0, 1-\rho^2)$$

$$\mathrm{cov}(X,Y) = \mathrm{cov}(X, \rho X + \varepsilon)$$

Recall

(1) $\mathrm{cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y)$

(2) $\mathrm{Cov}(X+Y, Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$

Then $Cov(X,Y) = \rho\, Cov(X,X) + \cancel{Cov(X, \varepsilon)} = \rho$

$= Corr(X,Y)$

"Regression" original meaning

$E(Y \mid X = x) = \rho x$ $\quad \rho < 1$

Y: son's height, x: father's height

regression towards mean

ex.) Michael Jordan's son isn't as tall → son is towards mean

↳ but ε will continue to cause fluctuation

↳ which caused Jordan to be taller than his dad

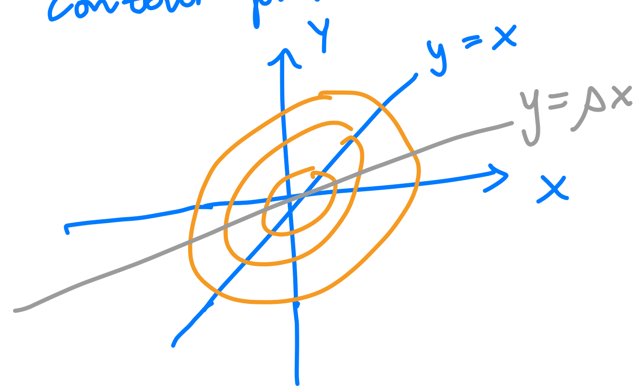
otherwise $Y = X + \varepsilon$

$Var(Y) = Var(X) + Var(\varepsilon)$

bivariate normal

$$\begin{bmatrix} X \\ Y \end{bmatrix} = N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

contour plot:



Conditional Covariance

$$\mathrm{Cov}(X, Y \mid Z = z) = E\big((X - h(z))(Y - g(z)) \mid Z = z\big)$$
$$= \int (x - h(z))(y - g(z)) f(x, y \mid z)\, dx\, dy$$

$$E(X \mid Z = z) = h(z)$$
$$E(Y \mid Z = z) = g(z)$$

(3) $\mathrm{Cov}(X, Y) = E_Z(\mathrm{Cov}(X, Y \mid Z)) + \mathrm{Cov}(E(X \mid Z), E(Y \mid Z))$

proof:
$$X = h(Z) + \varepsilon$$
$$Y = g(Z) + \delta$$

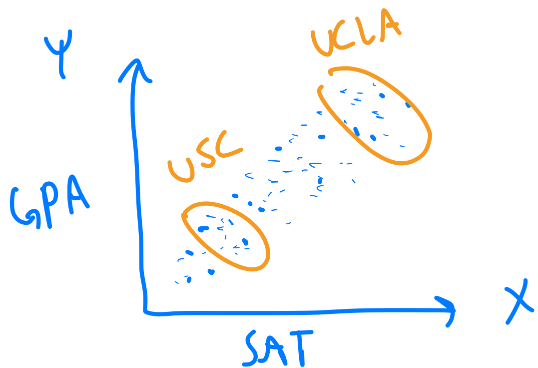
$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(h(Z) + \varepsilon, g(Z) + \delta)$$
$$= \mathrm{Cov}(h(Z), g(Z)) + \mathrm{Cov}(\varepsilon, \delta) + \mathrm{Cov}(h(Z), \delta) + \mathrm{Cov}(\varepsilon, g(Z))$$

since ε is uncorrelated

$$= \mathrm{Cov}(E(X \mid Z), E(Y \mid Z)) + E\,\underbrace{E((X - h(Z))(Y - g(Z)) \mid Z)}_{\mathrm{Cov}(X, Y \mid Z)}$$

ex.) X: SAT
Y: GPA
Z: school

$\mathrm{Cov}(X, Y \mid Z) < 0$



Conditional Independence

Recall $X \perp Y$

$$f(x,y) = f_X(x) f_Y(y)$$

$X \perp Y \mid Z$

$$f(x,y \mid z) = f(x \mid z) f(y \mid z)$$

$$f(x,y) = \int f(x,y,z)\, dz = \int f(x,y \mid z) f_Z(z)\, dz$$

$$= \int f(x \mid z) f(y \mid z) f_Z(z)\, dz$$

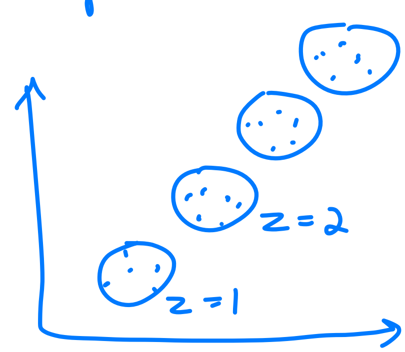
$$X = Z + \varepsilon$$
$$Y = Z + \delta$$

ex.) Z : time it takes from office to bus stop near home

ε : person A time from bus stop to their home

δ : person B time from bus stop to their home

ex.) $\underset{\text{child}}{X} = \overset{\text{parent}}{Z} + \varepsilon$



ex.) Discrete case

Y: health

X: smoking habit < pipe, cigarette } seem to have strong correlation

Z: age → confounding variable

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ X & & Y \end{tikzcd}$$

* older people tend to be less healthy & more likely to smoke from a pipe

causal effect

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ X \arrow[rr] & & Y \end{tikzcd} \longrightarrow \text{joint distribution } p(x, y, z)$$

say that health is measured on a scale from 1-100

	pipe $X=0$	cigarette $X=1$	Z
1	?	90	
⋮	70	?	
⋮	?	80	
⋮		?	
n	60		
	$\bar{X}_0$	$\bar{X}_1$	

$\longrightarrow$ 80 - ?

counterfactual

structural equation

let $Z = h_Z(\varepsilon_Z)$

$X = h_X(Z, \varepsilon_X)$

$Y = h_Y(X, Z, \varepsilon_Y)$

assume $\varepsilon_Z, \varepsilon_X, \varepsilon_Y$ are ind.

counterfactual analysis

$$\begin{tikzcd} & Z = h_Z(\varepsilon_Z) \arrow[d] \\ X \arrow[r] & Y = h_Y(x, z, \varepsilon_Y) \end{tikzcd} \Longrightarrow \tilde{p}(y, z \mid do(x))$$

do 1 : $X \leftarrow 1$

do 0 : $X \leftarrow 0$

causal effect: $E(Y \mid X \leftarrow 1) - E(Y \mid X \leftarrow 0)$

back-door

or $\tilde{p}(y \mid do(x)) = \sum_z p(y \mid x, z)\, p(z)$

vs $p(y \mid x) = \sum_z p(y, z \mid x) = \sum_z p(y \mid x, z)\, p(z \mid x)$