(X, Y). two dimensional random vector with some joint distribution.

$$\rho_{XY} = \frac{\mathbb{E}[XY] - \mathbb{E}X\,\mathbb{E}Y}{\sqrt{Var(X) \cdot Var(Y)}}$$

$$= \frac{\mathbb{E}[(\underline{X - \mathbb{E}X})(\underline{Y - \mathbb{E}Y})]}{\sqrt{Var(X)\, Var(Y)}} \quad *$$

$\boxed{\rho_{XY}}$ is a measure of the linear association between X and Y: in other words, how well can Y be approximated by a linear function of X?

Y : response

X : predictor

$$* = \mathbb{E}\left[\left(\frac{X - \mu_X}{\sigma_X}\right)\left(\frac{Y - \mu_Y}{\sigma_Y}\right)\right]$$

$\sigma_X = s.d(X)$, $\sigma_Y = s.d(Y)$

Linear Model idea: How well can we predict Y by a linear fn of X i.e $a + bX$?

Idea of proof: Use a generic linear predictor of Y based on X, say $\phi(X) = tX$ where $t \in \mathbb{R}$, and find best t.

$$h(t) = \mathbb{E}\left[(Y - tX)^2\right]$$

$$= \mathbb{E}[Y^2] - t\,2\,\mathbb{E}(XY) + t^2\,\mathbb{E}(X^2)$$

Set $h'(t) = 0$

i.e $-2\,\mathbb{E}(XY) + 2t\,\mathbb{E}(X^2) = 0$

so $$t_{min} = \frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}$$ → regression co-efficient

$0 \leq h(t_{min})$

$$= E(Y^2) - \frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)}\,2\,\mathbb{E}(XY) + \frac{\mathbb{E}(XY)^2}{(\mathbb{E}(X^2))^2}\cdot \mathbb{E}X^2$$

$$= \mathbb{E}(Y^2) - \frac{(\mathbb{E}(XY))^2}{\mathbb{E}(X^2)}$$

$\mathbb{E}X^2 = 1$

$\mathbb{E}Y^2 = 1$

$$= 1 - \rho_{XY}^2 \cdot \left(\because \rho_{XY} = \mathbb{E}(XY) \right)$$

squared error of prediction of the best linear predictor is precisely $1 - \rho_{XY}^2$. Since $1 - \underline{\rho_{XY}^2} \geq 0$ we have $|\rho_{XY}| \leq 1$.

(a) ρ_{XY} measures linear association in the sense that large $|\rho_{XY}|$ produces small squared prediction error

(b) $|\rho_{XY}| = 1 \Rightarrow \mathbb{E}[(Y - \underline{t_{min} X})^2] = 0$

$\Rightarrow Y - t_{min} X = 0$ w.p. 1

$\Rightarrow Y = t_{min} X$

In general what you have is

$$\frac{Y - \mu_Y}{\sigma_Y} = t_{min} \frac{X - \mu_X}{\sigma_X}$$

when $|\rho| = 1$

$$t_{min} = \frac{\mathbb{E}(XY)}{\mathbb{E}(X^2)} = \rho \ . \text{ i.e}$$

$$\frac{Y - \mu_Y}{\sigma_Y} = \pm 1 \cdot \frac{X - \mu_X}{\sigma_Y}$$

Conditional Expectation: A Central Notion in Probability and Statistics.

(X, Y): want to talk about

$\mathbb{E}[Y|X]$: regression function

Fact: among all predictors of Y based on X, $\mathbb{E}[Y|X]$ is the best.

We'll revert to the comfortable scenario where (X, Y) has either a joint mass function (discrete case) or a joint density function (continuous case)

(X, Y) continuous. Joint density: $f(x,y)$

$$\underline{f_{Y|X=x}} = \frac{f(x,y)}{f_X(x)}$$

$$f_{X|Y=y} = \frac{f(x,y)}{f_Y(y)}$$

$f_X(x)$

$= \int_{\mathbb{R}} f(x,y)\,dy$

$f_Y(y)$

$= \int_{\mathbb{R}} f(x,y)\,dx$

<u>First define:</u>

$\boxed{\mathbb{E}[Y \mid X = t]}$ = Expectation of a random variable distributed with density $f_{Y|X=t}(y)$.

$$\overset{\shortparallel}{\int y \cdot \frac{f(t,y)}{f_X(t)}\,dy}$$

$\longleftarrow$ a pure function of t, say $\Xi(t)$

We now define a random variable:

$$\mathbb{E}[Y|X] = \xi(X):$$

where $\xi(t) = \int y \frac{f(t,y)}{f_X(t)}\, dy$

Similarly $\mathbb{E}[X|Y] = \tilde{\xi}(Y)$

where $\tilde{\xi}(t) = \mathbb{E}[X|Y=t]$

$$= \int x \frac{f(x,t)}{f_Y(t)}\, dx.$$

Proposition:

$$\mathbb{E}Y = \mathbb{E}[\mathbb{E}[Y|X]].$$

$$= \mathbb{E}[\xi(X)]$$

$$\mathbb{E}X = \mathbb{E}[\mathbb{E}(X|Y)] = \mathbb{E}[\tilde{\xi}(Y)]$$

Proof:

$$\mathbb{E}Y = \mathbb{E}[h(X,Y)], \text{ where } h(x,y) = y$$

$$= \int\int h(x,y) f(x,y)\, dy\, dx$$

$$= \int \left(\int y\, f(x,y)\, dy \right) dx$$

$$= \int \left[\int y \frac{f(x,y)}{f_X(x)}\, dy \right] f_X(x)\, dx$$

$$= \xi(x).$$

$$= \int \xi(x) f_X(x)\, dx$$

$$= \mathbb{E}[\xi(X)] = \mathbb{E}[\mathbb{E}[Y|X]].$$

That does it

How about $\mathbb{E}[g(Y) \mid X]$.

again, recipe is similar.

Let $\psi(t) = \mathbb{E}[g(\underline{Y}) \mid X = t]$

$$= \int g(y) \frac{f(t, y)}{f_X(t)} \, dy$$

Then define: $\mathbb{E}[g(Y) \mid X] = \psi(X)$.

Check: $\mathbb{E}[g(Y)] = \mathbb{E}[\mathbb{E}[g(Y) \mid X]]$

More important facts:

Consider $w(X, Y) \longrightarrow$ some generic function of (X, Y)

$w(x, y) = x^2 \cdot y^3 + x^5$

Want to define:

$\mathbb{E}[w(X, Y) \mid X = t] \longleftarrow$

Define this as the expectation of

② $w(u,v)$ where (u,v) has the same distribution as $(x,y)\,|\,x=t$.

$(x,y)\,|\,x=t$ $\quad x\,|\,x=t = t$ (constant) degenerate

$$y\,|\,x=t \sim \frac{f(y,t)}{f_x(t)}$$

So the distribution of (u,v) is simply determined by the constant t and the conditional of $y\,|\,x=t$.

and so $w(u,v)$ under this conditional is just $w(\underline{t},v)$ where v has the same distribution as $y\,|\,x=t$.

so $\mathbb{E}[w(u,v)] = \int w(t,y)\frac{f(y,t)}{f_x(t)}\,dy$

$$\underbrace{\mathbb{E}[w(x,y)\,|\,x=t]}_{=\beta(t)} = \int w(t,y)\frac{f(y,t)}{f_x(t)}\,dy$$

Again $\mathbb{E}[\underline{w(x,y)}]$

$$= \mathbb{E}\left[\mathbb{E}[w(x,y) \mid x]\right]$$

where $\mathbb{E}[w(x,y) \mid x] = \beta(x)$

<u>a generally useful property:</u>

$$\mathbb{E}[\underline{h(x,y)\, g(x)} \mid \underline{x}] \leftarrow$$

$$= \underline{g(x)}\, \underline{\underline{\mathbb{E}[h(x,y) \mid x]}}$$

Conditional variance of $y \mid x$.

(a) Define conditional variance of $y \mid x=t$

as $\mathrm{Var}(u)$ where $\underline{\underline{u}} \sim \left.\dfrac{f(t,y)}{f_X(t)}\right\}$

$$\mathrm{Var}(y \mid x=t) = \underline{\underline{\mathrm{Var}}}(u)$$

$$u \sim \left.\frac{f(t,y)}{f_X(t)}\right\}.$$

$$Var(u) = \mathbb{E}\left[(u - (\mathbb{E}u))^2\right]$$

$$= \underline{\mathbb{E}(u^2)} - (\mathbb{E}u)^2$$

$$= \underline{\int y^2 \frac{f(t,y)}{f_X(t)}\, dy} - \left(\int y \frac{f(t,y)}{f_X(t)}\, dy\right)^2$$

$$= \mathbb{E}\left[Y^2 \mid X=t\right] - \left(\mathbb{E}(Y \mid X=t)\right)^2 \quad (1)$$

$$= \mathbb{E}\left[\left(Y - \mathbb{E}(Y \mid X=t)\right)^2 \mid X=t\right] \quad (2)$$

So:

$\underline{Var(Y \mid X=t)}$ = equivalently as (1) or (2)

"$\mathcal{K}(t)$

Define: $\underline{Var(Y \mid X)} = \mathcal{K}(X)$.

Important equality writing unconditional variance in terms of conditional quantities:

$$\mathrm{Var}(y) = \mathbb{E}\left[\underline{\mathrm{Var}(y|x)}\right] + \mathrm{Var}\left[\mathbb{E}(y|x)\right]$$

Proof:

$$\mathrm{Var}(y) = \mathbb{E}\left[(y - \mathbb{E}y)^2\right]$$

$$= \mathbb{E}\left[\left(\underline{y - \mathbb{E}(y|x)} + \underline{\mathbb{E}(y|x) - \mathbb{E}y}\right)^2\right]$$

$$= \mathbb{E}\left[(y - \mathbb{E}(y|x))^2\right]$$

$$+ \mathbb{E}\left[(\mathbb{E}(y|x)) - \mathbb{E}y\right]^2$$

0

$$+ \mathbb{E}\left[2(y - \mathbb{E}(y|x))(\mathbb{E}(y|x) - \mathbb{E}y)\right]$$

$$= \text{Term 1} + \text{Term 2} + \text{Term 3}$$

Term 3 = 0

Term 2:

$$\mathbb{E}\left[\mathbb{E}(y|x) - \mathbb{E}y\right]^2$$

$$= \mathbb{E}\left[\mathbb{E}(y|x) - \mathbb{E}(\mathbb{E}(y|x))\right]^2$$

$$= Var[\mathbb{E}[y|x]]$$

Term 1:

$$= \mathbb{E}\left[\mathbb{E}\left[(y - \mathbb{E}(y|x))^2 \mid x\right]\right]$$

Claim: $\mathbb{E}\left[(y - \mathbb{E}(y|x))^2 \mid x\right]$

$$= Var[y|x]$$

Needs to be formally justified..

=