Regression: ch. 14

X & Y are jointly distributed w/

w/ $f_{X,Y}(x,y)$ joint pdf.

$\boxed{\text{Def. 14.1}}$

$$\mu_{Y|X=x} = E(Y|X=x)$$

is the regression eq. for Y given X.

$\mu_{Y|x}$ is a funct. of x

e.g. $\mu_{Y|x} = \alpha + \beta x \quad \alpha, \beta \in \mathbb{R}$

$\mu_{Y|x} = c \cdot e^{rx} \quad c, r \in \mathbb{R}$

$\boxed{\text{Linear}}$

See Math 421 notes on Conditional expectation & distributions. $\boxed{\S 3.7}$

$\boxed{\S 4.8}$ e.g. $f_{Y|X=x}(y|x) = \dfrac{f(x,y)}{f_X(x)}$

is the conditional pdf of Y given $X=x$

ex] $f(x,y) = 2x + 3y^2$ on $[0,1]^2$

In general:

$$\mu_{Y|x} = E(Y \mid X = x) = \int_0^1 y \cdot f_{Y|X=x}(y|x)\, dy$$

$$= \frac{\int_0^1 y \cdot f(x,y)\, dy}{\int_0^1 f(x,y)\, dy}$$

$$= \frac{\int_0^1 y \cdot (2x + 3y^2)\, dy}{\int_0^1 2x + 3y^2\, dy}$$

$$= \frac{\left[xy^2 + \frac{3}{4}y^4\right]_0^1}{\left[2xy + y^3\right]_0^1} = \frac{x + \frac{3}{4}}{2x+1}$$

Thus the regression eq. for Y given $X = x$

$$\boxed{\mu_{Y|x} = \frac{x + 3/4}{2x+1}}$$

Non-linear funct. of x

Note: If we had actual (X,Y) data, the regression function is not necessarily a "good fit". We'd need to investigate the conditional variance

$$Var(Y \mid X = x) = E(Y^2 \mid x) - \left[E(Y \mid x)\right]^2.$$

Linear Regression - §14.2

$$\mu_{Y|X=x} = \alpha + \beta x$$

Let μ_1, μ_2 be means &
σ_1^2, σ_2^2 variances of
X & Y

Remember: X & Y are
jointly distributed

$\boxed{\S 4.6} \to$ Recall Covariance

$$\text{Cov}(X, Y) = E\big((X-\mu_1)(Y-\mu_2)\big)$$
$$= \sigma_{12}$$

Correlation Coef.: $\rho = \dfrac{\sigma_{12}}{\sigma_1 \cdot \sigma_2}$

"rho" $\nearrow$

$\boxed{\text{Thm } 14.1}$ If $\mu_{Y|x}$ is linear in x, then:

$$\boxed{\mu_{Y|x} = \mu_2 + \rho \cdot \frac{\sigma_2}{\sigma_1}(x - \mu_1)}$$

Proof of Thm 14.1

$$\mu_{Y|x} = \alpha + \beta x = \frac{\int_{-\infty}^{\infty} y \cdot f(x,y)\,dy}{f_X(x)}$$

so:

★ $\alpha \cdot f_X(x) + \beta \cdot x \cdot f_X(x) = \int_{-\infty}^{\infty} y \cdot f(x,y)\,dy$

integrate w.r.t. x

$$\alpha + \beta \cdot E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y \cdot f(x,y)\,dy\,dx$$

$$= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} y \cdot f_Y(y)\,dy$$

$$= E(Y)$$

Thus $\alpha + \beta \cdot \mu_1 = \mu_2$

mult. ★ by x before integrating:

$$\alpha \cdot \underbrace{E(X)}_{\mu_1} + \beta \cdot \underbrace{E(X^2)}_{\sigma_1^2 + \mu_1^2} = \underbrace{E(XY)}_{\sigma_{12} + \mu_1 \mu_2}$$

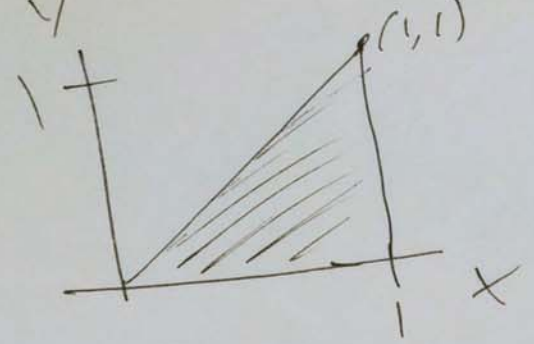
Then have 2 eq. for α and β

get: $\alpha = \mu_2 - \frac{\sigma_1^2}{\sigma_2^2}\mu_1$

$\beta = \frac{\sigma_{12}}{\sigma_1^2}$

★ a bit of algebra gives Thm 14.1

ex:



joint pdf:

$f(x,y) = 2$ on this triangle

Marginals: $f_X(x) = \int_0^x 2\,dy = 2x \qquad x \in (0,1)$

$f_Y(y) = \int_y^1 2\,dx = 2(1-y) \qquad y \in (0,1)$

So: Regr. y given x is:

$$\mu_{Y|x} = \frac{\int_0^x y \cdot 2\,dy}{2x} = \frac{[y^2]_0^x}{2x} = \frac{x^2}{2x} = \boxed{\frac{x}{2}}$$

Thus $\alpha = 0$, $\beta = \frac{1}{2}$

but must also have $\mu_{Y|x} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$

$\mu_1 = E(X) = \int_0^1 x \cdot 2x\,dx = \frac{2}{3}$

$\mu_2 = E(Y) = \int_0^1 y \cdot 2(1-y)\,dy = \frac{1}{3}$

$E(X^2) = \int_0^1 x^2 \cdot 2x\,dx = \frac{1}{2}$

$E(Y^2) = \int_0^1 y^2 \cdot 2(1-y)\,dy = \frac{1}{6}$

$E(XY) = \int_0^1 \int_0^x xy \cdot 2\,dy\,dx = \int_0^1 x \cdot [y^2]_0^x\,dx = \int_0^1 x^3\,dx$

$= \frac{1}{4}$

Thus $\sigma_1^2 = \sigma_2^2 = \frac{1}{18}$

$$\rho = \frac{\sigma_{12}}{\sigma_1 \cdot \sigma_2} = \frac{\frac{1}{4} - \frac{2}{3}\cdot\frac{1}{3}}{\frac{1}{18}} = \frac{1}{2}$$

So: $\mu_{Y|x} = \frac{1}{3} + \frac{1}{2} \cdot \frac{\sqrt{1/18}}{\sqrt{1/18}}\left(x - \frac{2}{3}\right) = \frac{1}{2}x$ ✓ both eq.'s same!