Function of $X_1, \dots, X_n$

- Corollary. Suppose that $X_1, \dots, X_n$ are uncorrelated and have same mean μ and variance σ^2. Let

relax i.i.d condition

$$\underbrace{S^2}_{\text{a r.v.}} = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-1}, \overset{\text{c.f.}}{\longleftrightarrow} \boxed{\text{definition of variance}} = E(X-\mu)^2 = \sigma^2$$

$E(\overline{X}_n)=\mu$ c.f.

then $E(S^2)=\sigma^2$. $\Rightarrow S^2 \xrightarrow{e} \sigma^2$

It can be shown that when $n\to\infty$ $Var(S^2)\to 0$

Proof. $(n-1)S^2 = \sum_{i=1}^n (X_i \overset{-\mu+\mu}{-} \overline{X}_n)^2$

$$\begin{aligned}
&= \sum_{i=1}^n [(X_i - \mu) - (\overline{X}_n - \mu)]^2 \\
&= \left[\sum_{i=1}^n (X_i - \mu)^2\right] + \left[\sum_{i=1}^n (\overline{X}_n - \mu)^2\right] \\
&\quad - 2(\overline{X}_n - \mu)\underbrace{\left[\sum_{i=1}^n (X_i - \mu)\right]}_{= n(\overline{X}_n - \mu)} \\
&= \left[\sum_{i=1}^n (X_i - \mu)^2\right] + n(\overline{X}_n - \mu)^2 - 2n(\overline{X}_n - \mu)^2 \\
&= \left[\sum_{i=1}^n (X_i - \mu)^2\right] - n(\overline{X}_n - \mu)^2.
\end{aligned}$$

Note. $\overline{X}_n \xrightarrow{e} \mu$ ← $E(\overline{X}_n)=\mu$, $Var(\overline{X}_n)=\sigma^2/n$

random

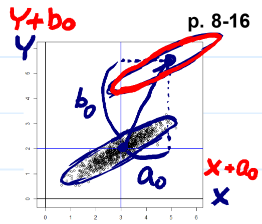
Therefore,

$$\cancel{(n-1)}E(S^2) = \left\{\sum_{i=1}^n \underbrace{E[(X_i-\mu)^2]}_{Var(X_i)=\sigma^2}\right\} - n\underbrace{E[(\overline{X}_n-\mu)^2]}_{Var(\overline{X}_n)}$$

$\because$ uncorrelated

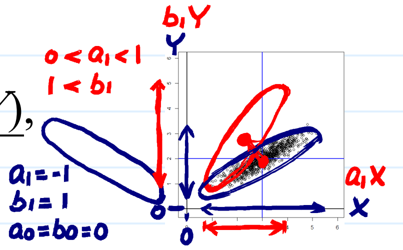
$$= n\sigma^2 - n\underbrace{Var(\overline{X}_n)}_{=\sigma^2/n \text{ (LNp.8-13)}} = \cancel{(n-1)}\sigma^2.$$

- Note. The previous three corollaries also hold if $X_1, \dots, X_n$ are "uncorrelated" is replaced by "independent." ← $\because$ "independence" implies "uncorrelated"



Recall 2nd Corollary in LNp.8-10

- Theorem (ρ of linear transformation).



$$Cor(\underbrace{a_0}_{\text{gone}}+a_1X,\ \underbrace{b_0}_{\text{gone}}+b_1Y)=\text{sign}(a_1b_1)\times Cor(X, Y),$$

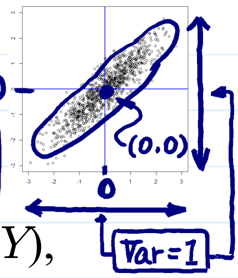
and

$$|Cor(a_0+a_1X,\ b_0+b_1Y)|=|Cor(X, Y)|,$$

Why?

i.e., $|\rho_{XY}|$ is invariant under location and scale changes.

standardization → 0



Thm in LNp.8-12

Proof. Let $S=a_0+a_1X$ and $T=b_0+b_1Y$, then

$$Cov(S, T)=Cov(a_0+a_1X,\ b_0+b_1Y)=a_1b_1Cov(X, Y),$$

$$Var(S)= a_1^2\,Var(X), \quad \text{and} \quad Var(T)= b_1^2\,Var(Y).$$

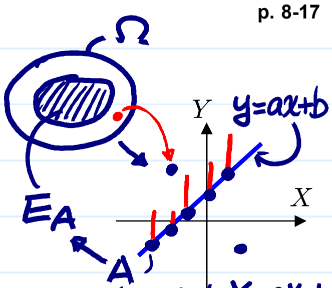
$$\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right) = \left[\frac{(X+a_0)-(\mu_X+a_0)}{\sigma_X}\right]\times\left[\frac{(Y+b_0)-(\mu_Y+b_0)}{\sigma_Y}\right] = \left(\frac{a_1X-a_1\mu_X}{|a_1|\sigma_X}\right)\times\left(\frac{b_1Y-b_1\mu_Y}{|b_1|\sigma_Y}\right)\times\text{sign}(a_1b_1)$$

Therefore,

$$\rho_{ST} = \frac{Cov(S,T)}{\sigma_S\sigma_T} = \frac{a_1b_1Cov(X,Y)}{|a_1||b_1|\sigma_X\sigma_Y} = \boxed{\frac{a_1b_1}{|a_1b_1|}}\rho_{XY}.$$

$\frac{a_1b_1}{|a_1b_1|} = \text{sign}(a_1b_1)$

$\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$ ➢ Theorem (some properties of ρ). cf. Cauchy-Schwarz inequality



$0 \le |\rho_{XY}| \le 1$ ⇔ (1) $-1 \le \rho_{XY} \le 1$. ($\Leftrightarrow |Cov(X,Y)| \le \sigma_X\sigma_Y$)

ρ is unit-free

(2) $\rho_{XY} = \pm 1$ if and only if there exist $a, b \in \mathbb{R}$ such that $P(Y=aX+b)=1$. $Y=aX+b$ almost surely

└ event $Y=aX+b$

$P(E_A)=1$

$P(E_A^c)=0$

(3) Furthermore, $\rho_{XY}=1$, if $a>0$ and $\rho_{XY}=-1$, if $a<0$.

Proof of (1).

柯西不等式 → Cauchy-Schwarz inequality $\underline{u}=(u_1,\cdots,u_n)$, $\underline{v}=(v_1,\cdots,v_n)$

$|\sum_i u_i v_i| = |\langle \underline{u}, \underline{v}\rangle| \le \|\underline{u}\|\cdot\|\underline{v}\| = \sqrt{\sum_i u_i^2}\sqrt{\sum_i v_i^2}$

$Y(\omega)\cdot\sqrt{p(\omega)}$, $X(\omega)\cdot\sqrt{p(\omega)}$, $\sum_i \to \sum_{\omega\in\Omega}$

$$0 \le Var\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right)$$

Thm in LNp. 8-13

$$= Var\left(\frac{X}{\boxed{\sigma_X}}\right) + Var\left(\frac{Y}{\boxed{\sigma_Y}}\right) + 2Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)$$

constants

$$= \frac{Var(X)}{\sigma_X^2} + \frac{Var(Y)}{\sigma_Y^2} + 2\,\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

$$= 1 + 1 + 2\rho_{XY} \quad \Rightarrow \rho_{XY} \ge -1.$$

$|E(XY)| \le \sqrt{E(X^2)}\sqrt{E(Y^2)}$

$|Cov(X,Y)| = |E[(X-\mu_X)(Y-\mu_Y)]| \le \sqrt{E[(X-\mu_X)^2]}\sqrt{E[(Y-\mu_Y)^2]} = \sigma_X\sigma_Y$

Similarly,

$$0 \le Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = 1 + 1 - 2\rho_{XY} \Rightarrow \rho_{XY} \le 1.$$

Proof of (2) and (3). We see from the proof of (1),

$$\rho_{XY} = 1 \Leftrightarrow Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = 0,$$

$Var(Z)=0 \Leftrightarrow Z=c$ almost surely, for a constant c

$$\Leftrightarrow \quad P\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = c\right) = 1,$$

where c is a constant.

$$\Leftrightarrow \quad P\left(Y = \boxed{\frac{\sigma_Y}{\sigma_X}}X + c\sigma_Y\right) = 1.$$

>0

<0

Similarly, $\rho_{XY} = -1 \Leftrightarrow P\left(Y = \boxed{-\frac{\sigma_Y}{\sigma_X}}X + c\sigma_Y\right) = 1.$

dist. unknown

• different transformation g ⇒ different information

• These expections are called parameters in statistics

• parameters can be estimated by r.v.'s (transformation of data), e.g., $\bar{X}_n \xrightarrow{p} \mu$, $S^2 \xrightarrow{p} \sigma^2$

• **Q**: How to use expectations to (roughly) characterize the distribution of random variables $X_1, \ldots, X_n$?

g: 1st order polynomials of $X_1, \cdots, X_n$

➢ $g(X_1, \ldots, X_n) = X_i \Rightarrow E[g(\mathbf{X})] = \mu_{X_i}$: *mean* of X_i.

➢ $g(X_1, \ldots, X_n) = (X_i - \mu_{X_i})^2 \Rightarrow E[g(\mathbf{X})] = \sigma^2_{X_i}$: *variance* of X_i.

g: 2nd order polynomials of $X_1, \cdots, X_n$

➢ $g(X_1, \ldots, X_n) = (X_i - \mu_{X_i})(X_j - \mu_{X_j})$ for $i \ne j$

$\Rightarrow E[g(\mathbf{X})] = \sigma_{X_iX_j}$: *covariance* of X_i and X_j.

➢ $g(X_1, \ldots, X_n) = [(X_i - \mu_{X_i})/\sigma_{X_i}][(X_j - \mu_{X_j})/\sigma_{X_j}]$ for $i \ne j$

$\Rightarrow E[g(\mathbf{X})] = \rho_{X_iX_j}$: *correlation coefficient* of X_i and X_j.

➢ Notes. $\mu_{X_i}, \sigma^2_{X_i}, \sigma_{X_iX_j}, \rho_{X_iX_j}$ are constants, not random

❖ **Reading**: textbook, Sec 7.1, 7.2, 7.4, 7.9

p. 8-19

Recall. Conditional dist. LNp.7-51~59

Conditional Expectation $\mathbf{X}\in\mathbb{R}^n, \mathbf{Y}\in\mathbb{R}^m$

- Recall. $p_{\cancel{\mathbf{Y}|\mathbf{X}}}(\mathbf{y}|\mathbf{x})$ or $f_{\cancel{\mathbf{Y}|\mathbf{X}}}(\mathbf{y}|\mathbf{x})$ is a pmf/pdf for $\mathbf{y}$ ($\mathbf{y}$: random, $\mathbf{x}$: fixed).
- Definition. For random vectors $\mathbf{X}$ and $\mathbf{Y}$, the *conditional expectation* of $Z=h(\mathbf{Y})$ given $\mathbf{X}=\mathbf{x}$, where h: $\mathbb{R}^m\to\mathbb{R}^1$, is

平均: $h(\mathbf{Y})$
權重: $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$

$$E_{\mathbf{Y}|\mathbf{X}}\left(\underbrace{h(\mathbf{Y})}_{Z}\Big|\,\mathbf{X}=\mathbf{x}\right)=\sum_{\mathbf{y}\in\mathcal{Y}} h(\mathbf{y})\, p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \sum_z z\, p_{Z|\mathbf{X}}(z|\mathbf{x})$$

(1)=(2) in LNp 8-1

in the discrete case, or,

平均: $h(\mathbf{Y})$
權重: $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$

$$E_{\mathbf{Y}|\mathbf{X}}\left(\underbrace{h(\mathbf{Y})}_{Z}\Big|\,\mathbf{X}=\mathbf{x}\right)=\int_{\mathbb{R}^m} h(\mathbf{y})\, f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\, d\mathbf{y} = \int_z z\, f_{Z|\mathbf{X}}(z|\mathbf{x})\,dz$$

(3)=(4) in LNp 8-1

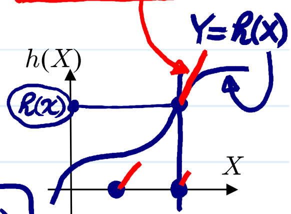
in the continuous case,

provided that the sum or integral converges absolutely.

➢ Some Notes.

- $E_{\mathbf{Y}|\mathbf{X}}(h(\mathbf{Y})\mid\mathbf{X}=\mathbf{x})$: a function of $\mathbf{x}$ and free of $\mathbf{Y}$.
- $E_{\mathbf{Y}|\mathbf{X}}[\,h(\mathbf{X})\mid\mathbf{X}=\mathbf{x}\,]=h(\mathbf{x})$.

cond. prob = 1



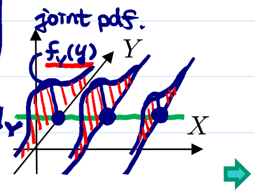
$E_{Y|X}[h(X,Y)|X=x]=\int h(x,y) f_{Y|X}(y|x)\,dy$

- If $\mathbf{X}$ and $\mathbf{Y}$ are independent, then

$\because f_{Y|X}(y|x)=f_Y(y)$
$p_{Y|X}(y|x)=p_Y(y)$

$$E_{\mathbf{Y}|\mathbf{X}}(h(\mathbf{Y})|\mathbf{X}=\mathbf{x})=E_{\mathbf{Y}}[h(\mathbf{Y})].$$

a constant line of x → μ_Y



p. 8-20

- Let $g(\mathbf{x})=E_{\mathbf{Y}|\mathbf{X}}[h(\mathbf{Y})|\mathbf{X}=\mathbf{x}]$, where g:$\mathbb{R}^n\to\mathbb{R}^1$, then we write

cf.

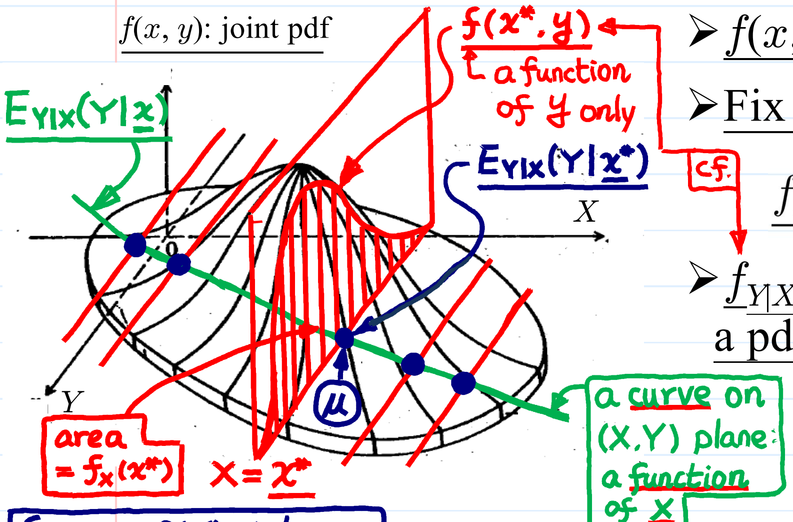
$$E_{\mathbf{Y}|\mathbf{X}}(h(\mathbf{Y})|\mathbf{X})$$

after / cf. / before

when $\mathbf{x}$ in g is replaced by $\mathbf{X}$ (a fixed value replaced by a r.v.).

- Notice that $g(\mathbf{X})$ is a random variable.

$f(x,y)$: joint pdf



$\int(y-\mu)\, f(x^*,y)\,dy=0$
$\Rightarrow \int y f(x^*,y)\,dy=\mu\int f(x^*,y)\,dy=\mu f_X(x^*)$
$\Rightarrow \mu=\int y\, f(x^*,y)/f_X(x^*)\,dy=E_{Y|X}(Y|x^*)$
$\Rightarrow f_{Y|X}(y|x^*)$ & $f(x^*,y)$ have same center of gravity

➢ $f(x,y)$: a joint pdf.

➢ Fix x^*, is $f(x^*,y)$ a pdf of y? i.e.,

cf.

$$f_X(x^*)=\int_{-\infty}^{\infty} f(x^*,y)\,dy \stackrel{?}{=} 1.$$

➢ $f_{Y|X}(y|x^*)=f(x^*,y)/f_X(x^*)$ is a pdf of y since

$$\frac{\int_{-\infty}^{\infty} f(x^*,y)\,dy}{f_X(x^*)}=1.$$

$h(Y)=Y$ center of gravity

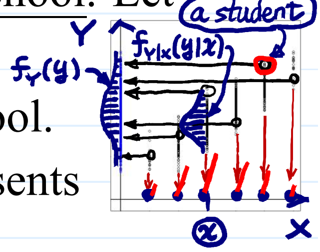
➢ $E_{Y|X}(Y|x^*)$: mean of $f_{Y|X}(y|x^*)$.

➢ Do it for any $x=x^*$, and get a function of $x \Rightarrow E_{Y|X}(Y|x)$

➢ Example. Sample a student from an elementary school. Let

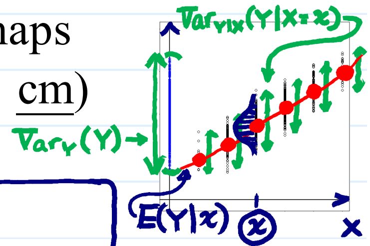
X=age (unit: year), Y=height (unit: cm)

of the student. **Population**: all students of the school.



Q: What's the source of their randomness?

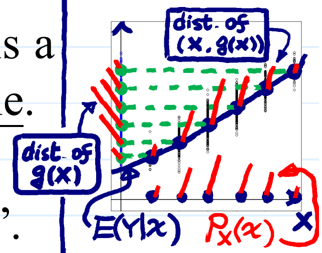
- $Y|X=x$: a random variable (unit: cm) that represents the height distribution of students with age=x.
- $g(x)=E_{Y|X}(Y|X=x)$ or $E_{Y|X}(Y|x)$: a *function* maps from age (unit: year) to average height (unit: cm) of students with age=x.



cf. after cf. before

Note. $E_{Y|X}(Y|x)$ is not a random variable.

- $g(X)=E_{Y|X}(Y|X)$: a random variable because it is a function of age X, where X is a random variable.



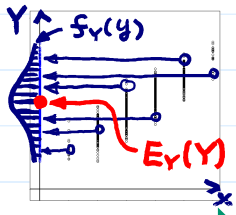
$g(x)$: 1-to-1 in this case $\Rightarrow P(E_{Y|X}(Y|X)=E_{Y|X}(Y|x))=P(X=x)$

$\boxed{R(Y)=[Y-g(x)]^2}$ $\quad E_{Y|X}[R(Y)|x]$

Note. $g(X)=E_{Y|X}(Y|X)$ is height, its unit is "cm".

- $Var_{Y|X}(Y|X=x)$ & $Var_{Y|X}(Y|X)$ defined similarly.
- $E_Y(Y)$: average height of ***all*** students; cf.

  $Var_Y(Y)$: variation of height of ***all*** students. cf.



- Theorem (Law of Total Expectation). For two random vectors $\mathbf{X}$ ($\in\mathbb{R}^m$) and $\mathbf{Y}$ ($\in\mathbb{R}^n$),

$E_{\mathbf{X},\mathbf{Y}}[R(\mathbf{Y})]=E_{\mathbf{Y}}E_{\mathbf{X}|\mathbf{Y}}$ $\quad E_{\mathbf{X}|\mathbf{Y}}[R(\mathbf{Y})]=R(\mathbf{Y})$

$$E_{\mathbf{X}}\{E_{\mathbf{Y}|\mathbf{X}}[h(\mathbf{Y})|\mathbf{X}]\}=E_{\mathbf{Y}}[h(\mathbf{Y})].$$

use the example in LN p.8-21 to realize the meaning of these terms.

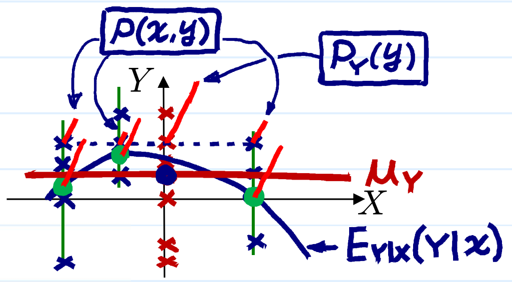
In particular, let $h(\mathbf{Y})=Y_i$, we have

$$E_{\mathbf{X}}[\underbrace{E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})}_{\equiv g(X)}]=\underbrace{E_{\mathbf{Y}}(Y_i)}_{=E_{Y_i}(Y_i)=E_{X,Y}(Y_i)}.$$

Y & $E_{Y|X}(Y|X)$ have same mean

Proof.

(only prove it for the continuous case)



$$
\begin{aligned}
&E_{\mathbf{X}}\{E_{\mathbf{Y}|\mathbf{X}}[h(\mathbf{Y})|\mathbf{X}]\}\\
&=\int_{\mathbb{R}^m} E_{\mathbf{Y}|\mathbf{X}}(h(\mathbf{Y})|\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}\\
&=\int_{\mathbb{R}^m}\left[\int_{\mathbb{R}^n} h(\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\, d\mathbf{y}\right] f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}\\
&=\int_{\mathbb{R}^n}\int_{\mathbb{R}^m} h(\mathbf{y}) \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{\cancel{f_{\mathbf{X}}(\mathbf{x})}} \cancel{f_{\mathbf{X}}(\mathbf{x})}\, d\mathbf{x}d\mathbf{y}\\
&=\int_{\mathbb{R}^n} h(\mathbf{y}) \left[\int_{\mathbb{R}^m} f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})\, d\mathbf{x}\right] d\mathbf{y}\\
&=\int_{\mathbb{R}^n} h(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y}\\
&=E_{\mathbf{Y}}[h(\mathbf{Y})].
\end{aligned}
$$

interchange dy & dx: $dydx \to dxdy$

$E_{X,Y}[R(Y)]$

E_Y ① $dx \leftrightarrow$ ② dy $\updownarrow$

$E_X E_{Y|X}$ ① dy $\updownarrow$ ② $dx \leftrightarrow$

multiplication law (LN p.7-55) $\Rightarrow = f_{X|Y}(x|y) \times f_Y(y)\,dx\,dy$ $\Rightarrow E_{X,Y}=E_Y E_{X|Y}$

generalization:

$E_{X,Y}[R(X,Y)]$

$=E_X E_{Y|X}[R(X,Y)|X]$

$=E_Y E_{X|Y}[R(X,Y)|Y]$