Expectation

$Y=g(X): \mathbb{R}^1 \to \mathbb{R}^1$ (LNp.5-13~19, LNp.6-13~16)

- Recall. Expectation for univariate random variable.
- Theorem. For random variables $\mathbf{X}=(X_1, \ldots, X_n)$ with joint pmf $p_{\mathbf{X}}$/pdf $f_{\mathbf{X}}$, the *expectation* of a univariate random variable Y, where

$$Y=g(X_1, \ldots, X_n), \qquad g: \mathbb{R}^n \to \mathbb{R}^1,$$

is

by definition, discrete

$$E(Y) \overset{\text{by definition}}{=} \sum_{y\in\mathcal{Y}} y\, p_Y(y) \tag{1}$$

no need to use $p_Y(y)$

$$= \sum_{\mathbf{x}=(x_1,\ldots,x_n)\in\mathcal{X}} g(x_1,\ldots,x_n)\, p_{\mathbf{X}}(x_1,\ldots,x_n) \tag{2}$$

$$\equiv E[g(X_1,\ldots,X_n)]$$

$\equiv$ defined as; $\mathcal{X}$: range of $\mathbf{X}$

$\sum_{\mathbf{x}\in\mathcal{X}} |g(\mathbf{x})|\, p_{\mathbf{X}}(\mathbf{x}) < \infty$

if $X_1, \ldots, X_n$ are discrete and the sum converges absolutely, or

$$E(Y) \overset{\text{by definition}}{=} \int_{-\infty}^{\infty} y f_Y(y)\, dy \tag{3}$$

no need to use $f_Y(y)$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\ldots,x_n) f_{\mathbf{X}}(x_1,\ldots,x_n)\, dx_1\cdots dx_n \tag{4}$$

$$\equiv E[g(X_1,\ldots,X_n)]$$

$\equiv$ defined as; $\int_{\mathbb{R}^n}$

if Y and $X_1, \ldots, X_n$ are continuous and the integrals converges absolutely. $\int_{\mathbb{R}^n} |g(\mathbf{x})| f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} < \infty$

Proof. Like the univariate case. proof in LNp.5-15 & LNp.6-16 $g: \mathbb{R}^1 \to \mathbb{R}^1$

- Q: What if Y is discrete and $X_1, \ldots, X_n$ are continuous? (1)=(4)

  e.g. $Y=g(\mathbf{X})=\begin{cases} 1, & \text{if } \mathbf{X}\in A\subset\mathbb{R}^n \\ 0, & \text{otherwise}\end{cases}$

➢ Notation.

- Shorthand notation. Combine (1) and (3) by writing

Note: cdf is defined for any r.v.'s

$$E(Y)=\int_{-\infty}^{\infty} y\, dF_Y(y) = \begin{cases} \sum_{y\in\mathcal{Y}} y\, p_Y(y), & \text{for discrete case,} \\ \int_{-\infty}^{\infty} y\, f_Y(y)\, dy, & \text{for continuous case,}\end{cases}$$

F_Y: cdf of Y; $dF_Y(y) \approx F_Y(y)-F_Y(y-)$

Note: $dF_Y(y)/dy = f_Y(y) \Rightarrow dF_Y(y)=f_Y(y)\,dy$

$dF_Y(y) \approx F_Y(y+\frac{dy}{2}) - F_Y(y-\frac{dy}{2}) = P(y-\frac{dy}{2} < Y \le y+\frac{dy}{2})$ and let $dy \to 0$

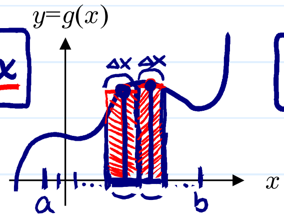
and combine (2) and (4) by writing

$$E[g(\mathbf{X})] = \int_{\mathbb{R}^n} g(\mathbf{x})\, dF_{\mathbf{X}}(\mathbf{x}) = \begin{cases} \sum_{\mathbf{x}\in\mathcal{X}} g(\mathbf{x})\, p_{\mathbf{X}}(\mathbf{x}), & \text{for discrete case,} \\ \int_{\mathbb{R}^n} g(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}, & \text{for continuous case.}\end{cases}$$

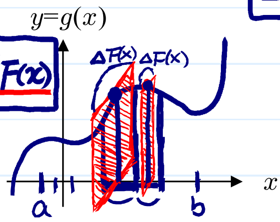
$F_{\mathbf{X}}$: joint cdf of $\mathbf{X}$

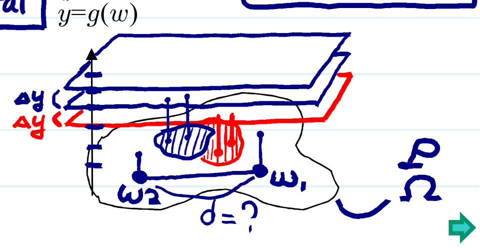
- Riemann-Stieltjes Integral. Lebesgue Integral; a r.v. $y=g(w)$; $\int_\Omega g(w)\, dP(w)$

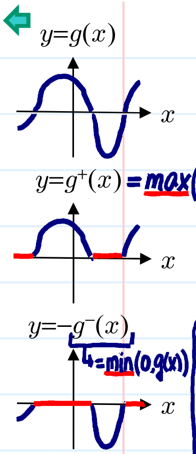
$\int_a^b g(x)\,dx$



$\int_a^b g(x)\,dF(x)$







For example, for non-negative g, and non-decreasing, right-continuous F, ← e.g. a cdf → $P((x_{i-1}, x_i])$ ← weight of $(x_{i-1}, x_i]$

$$\int_a^b g(x)\, dF(x) = \lim \sum_{i=1}^n g(x_i)[F(x_i) - F(x_{i-1})].$$

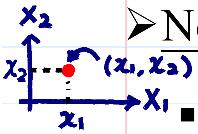
where the limit is taken over all $a=x_0<x_1<\cdots<x_n=b$ as $n\to\infty$ and $\max_{i=1,\ldots,n}(x_i - x_{i-1}) \to 0$.

$\int g(x)dF(x) = \int g^+(x)dF(x) - \int g^-(x)dF(x)$

[Recall. The integral of g over (a, b] is defined as cf.

$$\int_a^b g(x)\, dx = \lim \sum_{i=1}^n g(x_i)(x_i - x_{i-1}).]$$

length of $(x_{i-1}, x_i]$



➢Note. For their calculation, it is enough to know the marginal dist.

a projection

- $g(X_1, \ldots, X_n)=X_i \Rightarrow E[g(X_1, \ldots, X_n)]=E(X_i) \equiv \mu_{X_i}$.
- $g(X_1, \ldots, X_n)=(X_i-\mu_{X_i})^2 \Rightarrow E[g(X_1, \ldots, X_n)]=Var(X_i) \equiv \sigma^2_{X_i}$. $= E[(X_i-\mu_{X_i})^2] =$

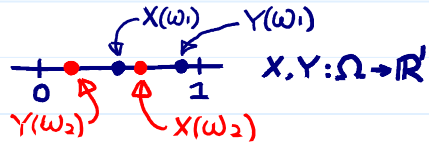
This is a fixed value, not random.

➢Example (Average distance between two points). Suppose that

jointly distributed → X, Y are i.i.d. ~ Uniform(0, 1).

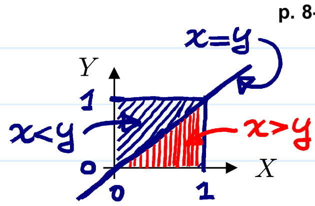
Let $D=|X-Y|$. Find $E(D)$.

$g(X,Y): \mathbb{R}^2 \to \mathbb{R}^1$



- The joint pdf of (X, Y) is

$$f(x,y) = \begin{cases} 1, & 0 \le x \le 1,\ 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$



Note. not necessary to derive the pdf of D

- $E(D) = \int_0^1\int_0^1 |x-y|\, dydx = \int_0^1 \left[\int_0^x (x-y)\, dy + \int_x^1 (y-x)\, dy\right] dx$

$$= \int_0^1 \left[-\tfrac{1}{2}(y-x)^2\Big|_{y=0}^{x} + \tfrac{1}{2}(y-x)^2\Big|_{y=x}^{1}\right] dx$$

$$= \int_0^1 \tfrac{1}{2}\left[x^2+(1-x)^2\right] dx = \tfrac{1}{6}\left[x^3-(1-x)^3\right]\Big|_{x=0}^{1} = \tfrac{1}{3}.$$

negative a_i's ⇒ difference

⊙Theorem (Mean of Sum). For jointly distributed r.v.'s $X_1, \ldots, X_n$ and constants $-\infty < a_0, a_1, \ldots, a_n < \infty$, ← a special form of g.

$$E(a_0+\underbrace{a_1X_1}_{=Y_1}+\cdots+\underbrace{a_nX_n}_{=Y_n}) = a_0+a_1\underbrace{E(X_1)}_{=E(Y_1)}+\cdots+a_n\underbrace{E(X_n)}_{=E(Y_n)}.$$

$g(X_1,\cdots,X_n)$

Note. They can be any r.v.'s with finite means (no additional condition is required)

exchange of E & Σ ← Why?

Proof. $E(a_0 + a_1X_1 + \cdots + a_nX_n)$

$$= \int_{\mathbb{R}^n} (a_0 + a_1x_1 + \cdots + a_nx_n)\, dF_{\mathbf{X}}(\mathbf{x})$$

$$= \int_{\mathbb{R}^n} a_0\, dF_{\mathbf{X}}(\mathbf{x}) + a_1\int_{\mathbb{R}^n} x_1\, dF_{\mathbf{X}}(\mathbf{x}) + \cdots + a_n\int_{\mathbb{R}^n} x_n\, dF_{\mathbf{X}}(\mathbf{x})$$

$$= a_0 + a_1E(X_1) + \cdots + a_nE(X_n).$$

Interchange of integration & summation ("=" guaranteed by absolute convergence)

$g_1(X_1,\cdots,X_n)=X_1$

$g_n(X_1,\cdots,X_n)=X_n$

only need the marginal dist. of $X_1,\cdots,X_n$, not necessary to know joint dist.

➤ Corollary. Suppose that $\mu = E(X_1) = \cdots = E(X_n)$. Let

e.g. $X_1, \ldots, X_n$ are identically distributed

$g(X_1, \cdots, X_n)$

a r.v.

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}, \rightarrow a_0 = 0,\ a_1 = a_2 = \cdots = a_n = \frac{1}{n}$$

then, $E(\bar{X}_n) = \mu$.

cf. definition of mean $= \underline{\underline{E}}(\underline{\underline{X}}) = \mu$

A={x≤y}

➤ Corollary. If X and Y are r.v.'s with finite means and

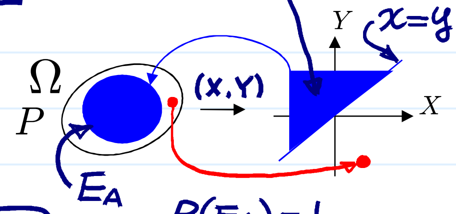
e.g. X:出生時身高 Y:10歲時身高

$$P(X \le Y) = 1,$$

"$X \le Y$" with probability one or almost surely

then $E(X) \le E(Y)$.

P("some property")=1 ⇔ "some property"



$P(E_A)=1$

$P(E_A^c)=0$

Proof. First, if Z is a random variable with finite mean and

$$P(Z \ge 0) = 1,$$

"$Z \ge 0$" with probability one or almost surely

then $E(Z) = \int_0^\infty z\, dF_Z(z) \ge 0$.

$F_Z(z) = 0$, when $z < 0$

For the general case, let $Z = Y - X$, then $Z \ge 0$ with probability one, and therefore, $0 \le E(Z) = E(Y - X) = E(Y) - E(X)$.

➤ Corollary. If $P(a \le X \le b) = 1$ for some constants a, b, then

$P(X - a \ge 0) = 1$ → $0 \le E(X - a) = E(X) - a$ → $a \le E(X) \le b$. ← intuition

- Theorem. If two random vectors $\mathbf{X}$ $(\in \mathbb{R}^m)$ and $\mathbf{Y}$ $(\in \mathbb{R}^n)$ are independent (i.e., $F_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = F_{\mathbf{X}}(\mathbf{x}) \times F_{\mathbf{Y}}(\mathbf{y})$, or

exchange of E & Σ

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) \times f_{\mathbf{Y}}(\mathbf{y}), \text{ or } p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x}) \times p_{\mathbf{Y}}(\mathbf{y}) \text{)},$$

cf.

then for $g: \mathbb{R}^m \to \mathbb{R}$ and $h: \mathbb{R}^n \to \mathbb{R}$,

exchange of E & Π

$$E[g(\mathbf{X}) \times h(\mathbf{Y})] = E[g(\mathbf{X})] \times E[h(\mathbf{Y})].$$

Note $g(\mathbf{X})$ & $h(\mathbf{Y})$ are independent. (LNp.7-24)

Proof. We only prove it for the continuous case:

$$E[g(\mathbf{X})h(\mathbf{Y})] = \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} g(\mathbf{x})h(\mathbf{y}) f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})\, d\mathbf{y}\, d\mathbf{x}$$

∵ independent

$$\overset{\vee}{=} \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} g(\mathbf{x})h(\mathbf{y}) f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y}\, d\mathbf{x}$$

product of function of x function of y

$$= \int_{\mathbb{R}^m} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \left[\int_{\mathbb{R}^n} h(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y}\right] d\mathbf{x}$$

This is a constant for x

$$= \left[\int_{\mathbb{R}^m} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}\right] \left[\int_{\mathbb{R}^n} h(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y}\right]$$

$$= E[g(\mathbf{X})] E[h(\mathbf{Y})].$$

independent: $E(X_1 \times X_2 \times \cdots \times X_n) = E(X_1)E(X_2)\cdots E(X_n)$ generalization

independent ⇒ uncorrelated, ~~⇐~~

➤ Corollary. For 2 independent r.v.'s X and Y,

exchange of E & Π

$$E(XY) = E(X) \times E(Y).$$

This is called "X and Y are uncorrelated"

Proof. Let $g(X) = X$ and $h(Y) = Y$.

check 3. in LNp.8-8

➢ **Q**: For independent r.v.'s X and Y, $E(X/Y) \neq E(X)/E(Y)$?

linear transformation like a+bY ← cf. → h: concave function (convex?)

$E[h(Y)] < h(E(Y))$

$E(\frac{X}{Y}) = E(X \cdot \frac{1}{Y})$ ∵ X & $\frac{1}{Y}$ are indep. (LNp.7-24)

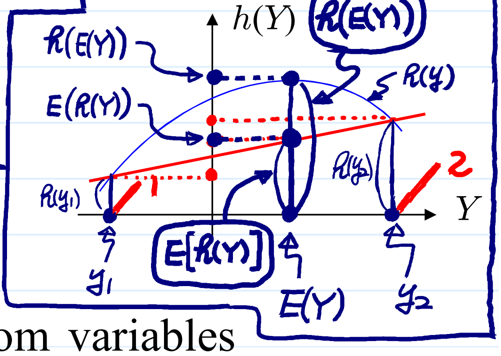
$= E(X) \cdot E(\frac{1}{Y}) \overset{?}{=} E(X) \cdot \frac{1}{E(Y)}$

➢ Note. $E[h(Y)] \neq h(E(Y))$ in general, e.g.,

Are h & E exchangable?

$E(1/Y) \neq 1/E(Y)$.

$h(y) = \frac{1}{y}$ – convex, $y>0$; concave, $y<0$



⊙ Covariance and Correlation between 2 random variables → joint distribution

Recall: mean & variance (defined in terms of E)

➢ Definition. Suppose that X and Y are two random variables with finite means μ_X, μ_Y and variances σ_X^2, σ_Y^2, respectively.

They can be calculated from the marginal distributions of X & Y.

1. Let $g(x, y) = (x-\mu_X)(y-\mu_Y)$, then $g: \mathbb{R}^2 \to \mathbb{R}^1$

$$Cov(X,Y) \equiv E[g(X,Y)]$$

$$= E[(X-\mu_X)(Y-\mu_Y)]$$

Its calculation needs the joint dist.

fixed value, not random

is called the *covariance* between X and Y, denoted by σ_{XY}. 共變數

2. The *correlation* (*coefficient*) between X and Y is defined as

相關係數

$$Cor(X,Y) = \sigma_{XY}/(\sigma_X \sigma_Y)$$

standard deviation (LNp.5-16, 6-14)

and denoted by ρ_{XY}. cov(X,Y)

3. X and Y are called *uncorrelated* if $\rho_{XY}=0$. $\overset{\text{iff}}{\longleftrightarrow}$ $\text{cov}(X,Y)=0$

- A special case of covariance:

$Cov(X, X) \overset{\nabla}{=} Var(X)$. — $E[(X-\mu_X)(X-\mu_X)] = E[(X-\mu_X)^2]$

Y=X

➢ Intuitive explanation of covariance and correlation

by definition $E[(X-\mu_X)(Y-\mu_Y)]$

▪ Covariance is the average value of the product of the deviation of X from its mean and the deviation of Y from its mean.

check the 1st graph in LNp.8-9

▪ Covariance is a measure of the *joint* variability of X and Y, or their degree of association.

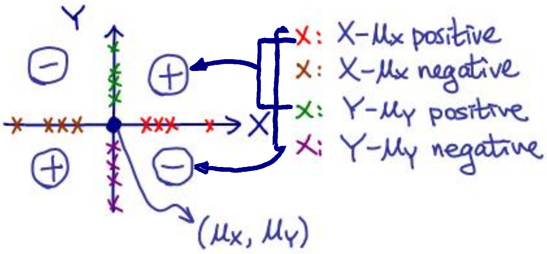
not necessarily a causal relationship

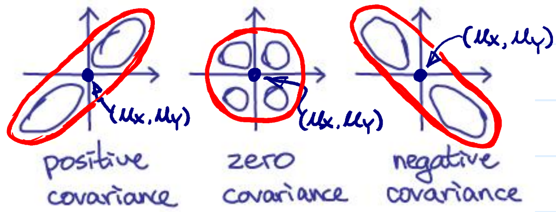
whether Y↑ (or Y↓) when X↑
e.g. X: height, Y: weight

▪ Positive Covariance and Negative Covariance

drawback: covariance depends on the unit/scale of X & Y
e.g. height: m → cm, |covariance| 10² times larger

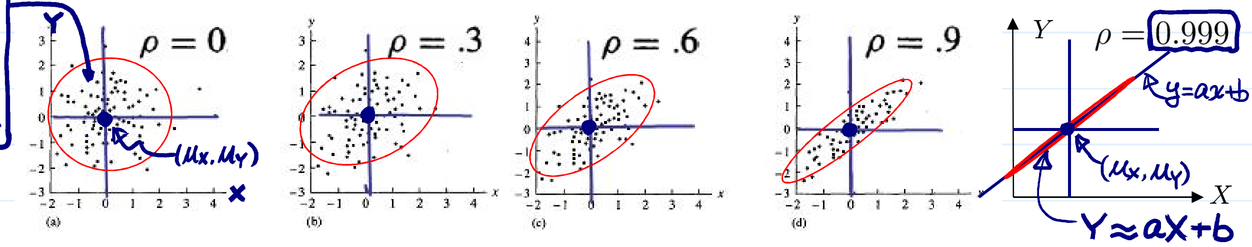
$(X-\mu_X)(Y-\mu_Y)$





- Correlation Coefficient is unit free. (why?) ← check its definition: $\frac{\sigma_{XY}}{\sigma_X \sigma_Y}$
- Correlation coefficient measures the strength of the *linear* relationship between X and Y.

the points reflect the joint distribution



Theorem. $Cov(X, Y) = E(XY) - \mu_X\mu_Y$. $\xleftrightarrow{\text{c.f.}}$ $Cov(X,X) = Var(X) = E(X^2) - \mu_X^2$

cov = 0 $\Leftrightarrow$ $E(XY) = E(X)E(Y)$

Proof. $Cov(X, Y) = E[(X-\mu_X)(Y-\mu_Y)]$

$$= E(XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y)$$

constants

$$= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y$$

$$= E(XY) - \mu_X\mu_Y - \cancel{\mu_Y\mu_X} + \cancel{\mu_X\mu_Y}.$$