Expectation

- Recall. Expectation for univariate random variable. — LNp.4-11~16, LNp.5-10~13
- 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

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

by definition.

is
$$E(Y) \equiv \sum_{y\in\mathcal{Y}} y\, p_Y(y) \quad (1)$$
$$\equiv \sum_{\mathbf{x}=(x_1,\ldots,x_n)\in\mathcal{X}} g(x_1,\ldots,x_n)\, p_{\mathbf{X}}(x_1,\ldots,x_n) \quad (2)$$
$$\equiv E[g(X_1,\ldots,X_n)]$$

no need to calculate $P_Y(y)$

i.e. $\sum |g(\underline{x})| P_X(\underline{x}) < \infty$

if $X_1, \ldots, X_n$ are discrete and the sum $\boxed{\text{converges absolutely}}$, or

by definition

$$E(Y) \equiv \int_{-\infty}^{\infty} y f_Y(y)\, dy \quad (3)$$
$$\equiv \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\ldots,x_n) f_{\mathbf{X}}(x_1,\ldots,x_n)\, dx_1\cdots dx_n \quad (4)$$
$$\equiv E[g(X_1,\ldots,X_n)]$$

no need to calculate $f_Y(y)$

i.e. $\int |g(\underline{x})| f_X(\underline{x}) d\underline{x} < \infty$

if Y and $X_1, \ldots, X_n$ are continuous and the integrals $\boxed{\text{converges absolutely}}$.

Proof. Like the univariate case. — LNp.4-13, LNp.5-11

e.g. $Y=\begin{cases}1, & \underline{X}\in A\subset\mathbb{R}^n\\ 0, & \text{o.w.}\end{cases}$

- **Q**: What if Y is discrete and $X_1, \ldots, X_n$ are continuous?

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

$F_Y(y)-F_Y(y-)=\Delta F_Y(y)$

cdf.

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

and combine (2) and (4) by writing

Note: $\frac{dF_Y(y)}{dy}=f_Y(y) \Rightarrow dF_Y(y)=f_Y(y)dy$

joint cdf.

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

  - Riemann-Stieltjes Integral. For example, for non-negative g,

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

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

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

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

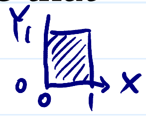
- Note.

Note: this is a fixed value, not random.

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

➢Example (Distance between two points). Suppose that

$g(X,Y)$ X, Y are i.i.d. ~ Uniform(0, 1).



Let $D=\boxed{|X-Y|}$. Find $E(D)$.

- The joint pdf of (X, Y) is

Note: not necessary to derive the pdf of D

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

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

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

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

- Theorem (Mean of Sum). For r.v.'s $X_1, \ldots, X_n$ and constants $-\infty < a_0, a_1, \ldots, a_n < \infty$,

→ Note: no additional requirement for $X_1, \cdots, X_n$.

exchange of Σ & E

$$E(a_0+a_1X_1+\cdots+a_nX_n) = a_0+a_1E(X_1)+\cdots+a_nE(X_n).$$

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

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

$$= \int_{\mathbb{R}^n} a_0\,dF_{\mathbf{X}}(\mathbf{x}) + a_1\int_{\mathbb{R}^n} x_1\,dF_{\mathbf{X}}(\mathbf{x})$$ $g_1(x_1,\cdots,x_n)=x_1$

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

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

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

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

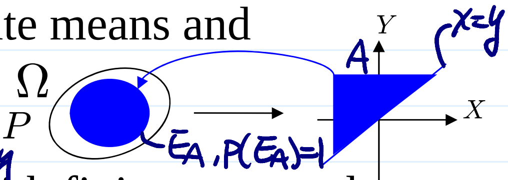
then, $E(\overline{X}_n)=\mu$.

➢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 prob. one, almost surely

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



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

$P(Z\ge 0)=1$, ⇒ $Z\ge 0$, with prob. one, almost surely ⇒ $P(Z<0)=0$.

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

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

→ $P(X-a\ge 0)=1$ ⇒ $\boxed{a\le E(X)}\le b$.

- 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

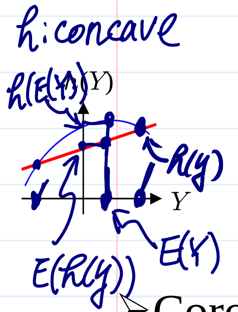
can be relaxed to uncorrelated

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

then for $g:\mathbb{R}^m\to\mathbb{R}$ and $h:\mathbb{R}^n\to\mathbb{R}$, → Note: $(g(X), h(Y))$ are indep. (LNp.6-20 Example)

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

Proof. We only prove it for the continuous case:



$$\begin{aligned}
E[g(\mathbf{X})h(\mathbf{Y})] &= \int_{\mathbb{R}^m}\int_{\mathbb{R}^n} g(\mathbf{x})h(\mathbf{y})f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})\,d\mathbf{y}d\mathbf{x}\\
&\overset{\because \text{ indep}}{=} \int_{\mathbb{R}^m}\int_{\mathbb{R}^n} g(\mathbf{x})h(\mathbf{y})f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{Y}}(\mathbf{y})\,d\mathbf{y}d\mathbf{x}\\
&= \int_{\mathbb{R}^m} g(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\underbrace{\left[\int_{\mathbb{R}^n} h(\mathbf{y})f_{\mathbf{Y}}(\mathbf{y})\,d\mathbf{y}\right]}_{\text{This is a constant for } X.}d\mathbf{x}\\
&= \left[\int_{\mathbb{R}^m} g(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\,d\mathbf{x}\right]\left[\int_{\mathbb{R}^n} h(\mathbf{y})f_{\mathbf{Y}}(\mathbf{y})\,d\mathbf{y}\right]\\
&= E[g(\mathbf{X})]E[h(\mathbf{Y})].
\end{aligned}$$

➢ Corollary. For 2 independent r.v.'s X and Y, $E(XY)=E(X)E(Y)$.

proof: $g(x)=x$, $h(y)=y$

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

$E(\frac{X}{Y}) = E[X\cdot\frac{1}{Y}] = E(X)\cdot E(\frac{1}{Y}) \overset{?}{=} E(X)\cdot\frac{1}{E(Y)}$ ∵ X & $\frac{1}{Y}$ indep by an example LNp.6-20

➢ Note. $E[h(Y)] \neq h(E(Y))$ in general, e.g., $E(1/Y) \neq 1/E(Y)$. $h(y)=1/y$

• Covariance and Correlation between 2 random variables

Recall mean variance

➢ Definition. Suppose that X and Y are two random variables with finite means μ_X, μ_Y and variances $\sigma_X{}^2$, $\sigma_Y{}^2$, respectively.

can be calculated from the marginal distribution of X & Y

1. Let $g(x, y)=(x-\mu_X)(y-\mu_Y)$, then

$$Cov(X,Y) \equiv E[g(X,Y)] = E[(X-\mu_X)(Y-\mu_Y)]$$

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

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

$$Cor(X,Y) = \sigma_{XY}/(\sigma_X\sigma_Y)$$ standard deviation (LNp.4-14, 5-12)

and denoted by ρ_{XY}.

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

▪ A special case of covariance: $Cov(X, X) = Var(X)$.
$= E[(X-\mu_X)(X-\mu_X)] = E[(X-\mu_X)^2]$

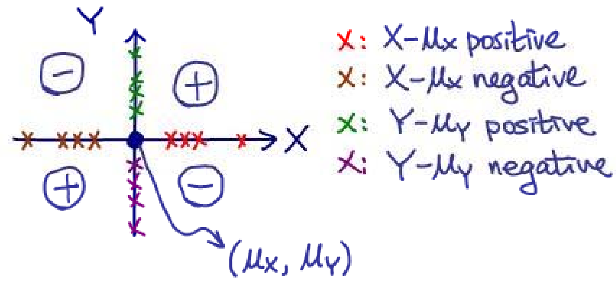
➢ Intuitive explanation of covariance and correlation

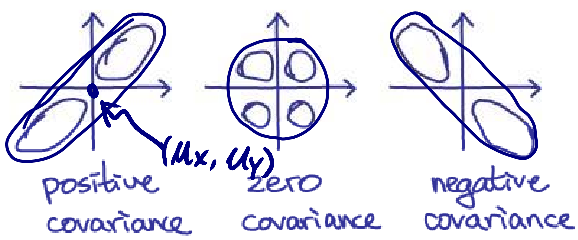
▪ Covariance is a measure of the joint variability of X and Y, or their degree of association. — whether $Y\uparrow$ (or $Y\downarrow$) when $X\uparrow$
e.g. X: height, Y: weight.

▪ Covariance is the average value of the product of the deviation of X from its mean and the deviation of Y from its mean. — drawback: covariance depends on the units/scales of X & Y, e.g. height: m → cm, 10^2 larger

definition

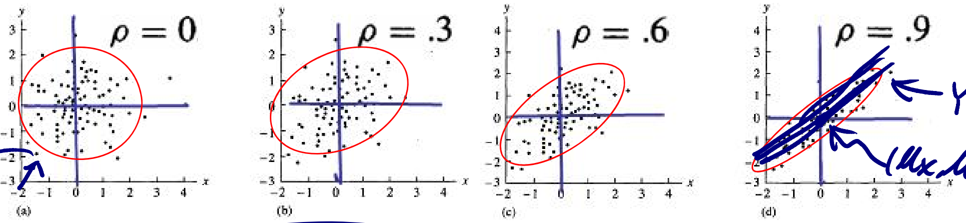
▪ Positive Covariance and Negative Covariance





- Correlation Coefficient is unit free.
- Correlation coefficient measures the strength of the linear relationship between X and Y.

why? check definition



joint pdf.

$Y \approx aX+b$

(μ_X, μ_Y)

1/3

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

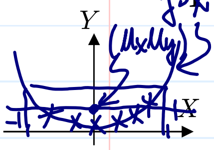
Proof.
$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= E(XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y) \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y \\ &= E(XY) - \mu_X\mu_Y - \cancel{\mu_Y\mu_X} + \cancel{\mu_X\mu_Y}. \end{aligned}$$

constants

- Corollary. If X and Y are independent, then $Cov(X, Y)$=0, i.e. X and Y are uncorrelated.

corollary LNp.7-5

Proof. When X, Y are independent, $E(XY)=E(X)E(Y)=\mu_X\mu_Y$.



uncorrelated is a weaker property than indep.

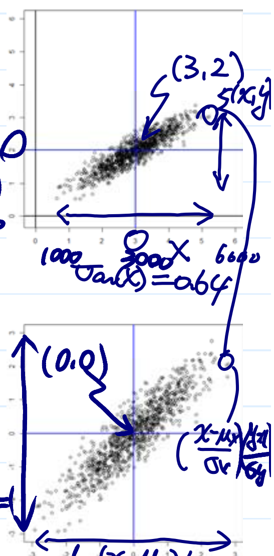
- However, the converse statement is not necessarily true (e.g., let X~Uniform(–1, 1) and $Y=X^2$, then $Cov(X, Y)$=0, but X and Y are not independent). $EXY = \{ EX^3 = \int_{-1}^{1} x^3 dx = 0$

- Corollary.

standardization (標準化, LNp.5-27)

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

Proof. By definition.



Q: Why cor unit free?

➢Example. If $(X_1, \ldots, X_m)$ ~ Multinomial(n, m, $p_1, \ldots, p_m$), then

$$Cov(X_i, X_j) = -np_ip_j, \quad \text{for } 1 \le i \ne j \le m.$$

- Because $(X_1, X_2, X_3+\cdots+X_m)$ ~ Multinomial(n, 3, p_1, p_2, $p_3+\cdots+p_m$), and

$$X_3+\cdots+X_m = n-X_1-X_2,$$
$$p_3+\cdots+p_m = 1-p_1-p_2,$$

$2 \le x_1+x_2 \le n$

we have

$0 \le x_1+x_2 \le n$

$$E(X_1X_2) = \sum x_1x_2 \binom{n}{x_1, x_2, n-x_1-x_2} p_1^{x_1}p_2^{x_2}(1-p_1-p_2)^{n-x_1-x_2}$$

$0 \le x_1+x_2 \le n$, $x_1 \ne 0$, $x_2 \ne 0$

$$= \sum \cancel{x_1}\cancel{x_2} \frac{n!}{\cancel{x_1!}\cancel{x_2!}(n-x_1-x_2)!} p_1^{x_1}p_2^{x_2}(1-p_1-p_2)^{n-x_1-x_2}$$

$(x_1-1)!(x_2-1)!$

$$= n(n-1)p_1p_2\left[\sum \frac{(n-2)!}{(x_1-1)!(x_2-1)!(n-x_1-x_2)!} \times p_1^{x_1-1}p_2^{x_2-1}(1-p_1-p_2)^{n-x_1-x_2}\right]$$

$0 \le (x_1-1)+(x_2-1) \le n-2$

$$= n(n-1)p_1p_2.$$

joint pmf of multinomial$(n-2, 3, p_1, p_2, 1-p_1-p_2)$

- WLOG, we can get $E(X_iX_j) = n(n-1)p_ip_j$, for $i \neq j$.

  Therefore, $$Cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j) = n(n-1)p_ip_j - (np_i)(np_j) = -np_ip_j.$$

- And, for $i \neq j$,

$$Cor(X_i, X_j) = \frac{-np_ip_j}{\sqrt{np_i(1-p_i)}\sqrt{np_j(1-p_j)}} = -\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}.$$

why negative?

• Expectations for Sums of Random Variables

➢ Notation. In the following, let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be r.v.'s and $-\infty<a_0,a_1,\ldots,a_n,b_0,b_1,\ldots,b_m<\infty$ are constants.

➢ Recall. $E(a_0+a_1X_1+\cdots+a_nX_n) = a_0+a_1E(X_1)+\cdots+a_nE(X_n)$.

➢ Theorem (covariance of two sums).

$$Cov(a_0 + a_1X_1 + \cdots + a_nX_n, b_0 + b_1Y_1 + \cdots + b_mY_m) = \sum_{i=1}^n \sum_{j=1}^m a_ib_jCov(X_i, Y_j) = [a_1 \cdots a_n]\left[Cov(X_i,Y_j)\right]\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

a_0, b_0 are gone

Proof. Let $S = a_0+a_1X_1+\cdots+a_nX_n$ and $T = b_0+b_1Y_1+\cdots+b_mY_m$, then

$$S - E(S) = \sum_{i=1}^n a_i(X_i - \mu_{X_i}),$$

$E(S) = a_0+a_1\mu_{X_1}+\cdots+a_n\mu_{X_n}$

$$T - E(T) = \sum_{j=1}^m b_j(Y_j - \mu_{Y_j}),$$

$$[S - E(S)][T - E(T)] = \sum_{i=1}^n \sum_{j=1}^m a_ib_j(X_i - \mu_{X_i})(Y_j - \mu_{Y_j}).$$

$X_1 \cdots X_n$; $Y_1 \cdots Y_m$; $Cov(X_i,Y_j)$

Therefore, $$Cov(S, T) = E\{[S - E(S)][T - E(T)]\} = \sum_{i=1}^n \sum_{j=1}^m a_ib_jE[(X_i - \mu_{X_i})(Y - \mu_{Y_j})] = \sum_{i=1}^n \sum_{j=1}^m a_ib_jCov(X_i, Y_j).$$

$[a_1,\cdots,a_n]\left[Cov(X_i,X_j)\right]\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ symmetric $n\times n$

➢ Theorem (variance of sum). $Cov(a_0+a_1X_1+\cdots+a_nX_n, a_0+a_1X_1+\cdots+a_nX_n)$ ||

1/5

$$Var(a_0 + a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^n \sum_{j=1}^n a_ia_jCov(X_i, X_j) = \sum_{i=1}^n a_i^2Var(X_i) + 2\sum_{1\leq i<j\leq n} a_ia_jCov(X_i, X_j).$$

Proof. $Cov(X_i, X_i)=Var(X_i)$. Why? (1) $X_1\approx X_2$, $Var(X_1+X_2)\approx Var(2X_1)=4Var(X_1)$ (2) $X_1\approx +X_2$. $Var(X_1-X_2)=0=2Var(X_1)-2Var(X_1)$

mean of sum of r.v. LN p.7-3

- Corollary. If $X_1, \ldots, X_n$ are uncorrelated, then c.f.

  $Cov(X_i,X_j)=0$, $\forall i,j$.

$$Var(\cancel{a_0} + a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^n a_i^2Var(X_i).$$

Y_1, Y_n; square; $= Var(Y_i)$

- Corollary. If $X_1, \ldots, X_n$ are uncorrelated and

  $\frac{X_1+\cdots+X_n}{n} \Leftarrow a_0=0, a_1=\cdots=a_n=\frac{1}{n}$ $Var(X_1)=\cdots = Var(X_n) \equiv \boxed{\sigma^2 < \infty}$ ← Var exists.

  then $Var(\overline{X}_n) = \sigma^2/n$ ≈ 0 when $n\uparrow\infty$ → i.e. $\overline{X}_n \approx C_n$ when n is large enough

  a constant

not assume i.i.d

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

  c.f. definition of variance of X

$$S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-1},$$

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

then $\boxed{E(S^2)=\sigma^2}$

Proof.

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

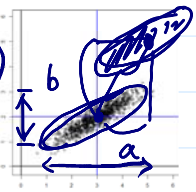
$$
\begin{aligned}
(n-1)S^2 &= \textstyle\sum_{i=1}^{n}[(X_i-\mu)-(\overline{X}_n-\mu)]^2 \\
&= \left[\textstyle\sum_{i=1}^{n}(X_i-\mu)^2\right]+\left[\textstyle\sum_{i=1}^{n}(\overline{X}_n-\mu)^2\right] \\
&\quad -2(\overline{X}_n-\mu)\boxed{\left[\textstyle\sum_{i=1}^{n}(X_i-\mu)\right]} = n(\overline{X}_n-\mu)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

$\boxed{\text{Note: } E(\overline{X}_n)=\mu,\ Var(\overline{X}_n)=\frac{\sigma^2}{n}}$

Therefore

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

- Note. The previous three corollaries also hold if $X_1, \ldots, X_n$ are independent ($\because$ "indep" implies "uncorrelated")



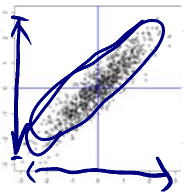
➢Theorem (ρ of linear transformation).

$Cor(a_0+a_1X_1,\ b_0+b_1Y_1)=\text{sign}(a_1b_1)\times Cor(X_1, Y_1)$,

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

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

→ why? check corollary in LN p. 7-8.



Proof. Let $S=a_0+a_1X_1$ and $T=b_0+b_1Y_1$, then

$$Cov(S, T)=Cov(a_0+a_1X_1,\ b_0+b_1Y_1)=a_1b_1Cov(X_1, Y_1),$$

$$Var(S)= a_1^2\, Var(X_1), \quad \text{and} \quad Var(T)= b_1^2\, Var(Y_1).$$

Therefore,

$$\rho_{ST}=\frac{Cov(S,T)}{\sigma_S\sigma_T}=\frac{a_1b_1Cov(X_1,Y_1)}{|a_1||b_1|\sigma_X\sigma_Y}=\frac{a_1b_1}{|a_1b_1|}\rho_{XY}.$$

$= \text{sign}(a_1b_1)$

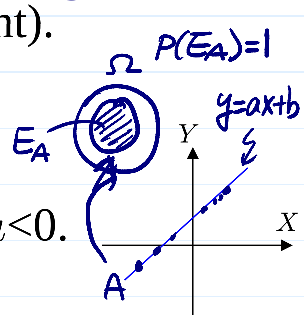
➢Theorem (some properties of correlation coefficient).

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

(2) $\rho_{XY} = \pm 1$ if and only if $P(Y=aX+b)=1$.

→ $Y=aX+b$ almost surely.

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



Proof of (1).

$$
\begin{aligned}
0 \le Var\left(\frac{X}{\sigma_X}+\frac{Y}{\sigma_Y}\right) & \quad \text{constants} \\
&= Var\left(\frac{X}{\sigma_X}\right)+Var\left(\frac{Y}{\sigma_Y}\right)+2Cov\left(\frac{X}{\sigma_X},\frac{Y}{\sigma_Y}\right) \\
&= \frac{Var(X)}{\sigma_X^2}+\frac{Var(Y)}{\sigma_Y^2}+2\frac{Cov(X,Y)}{\sigma_X\sigma_Y} \\
&= 1+1+2\rho_{XY} \quad \Rightarrow \rho_{XY} \ge -1.
\end{aligned}
$$

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),

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

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

where c is a constant.

$$\Leftrightarrow\; P\left(Y = \underbrace{\frac{\sigma_Y}{\sigma_X}}_{>0} X + c\sigma_Y\right) = 1.$$

Note: $Var(Z)=0$ if and only if $P(Z = \text{a constant}) = 1$ i.e. Z = a constant almost surely

Similarly, $\rho_{XY} = -1 \;\Leftrightarrow\; P\left(Y = \underbrace{-\frac{\sigma_Y}{\sigma_X}}_{<0} X + c\sigma_Y\right) = 1.$

- **Q**: How to use expectations to (roughly) characterize random variables $X_1, \ldots, X_n$?
  - $g(X_1, \ldots, X_n) = X_i \Rightarrow E[g(\mathbf{X})] = \mu_{X_i}$: *mean* of X_i. ← g: 1st order polynomial of $x_1, \ldots, x_n$
  - $g(X_1, \ldots, X_n) = (X_i - \mu_{X_i})^2 \Rightarrow E[g(\mathbf{X})] = \sigma^2_{X_i}$: *variance* of X_i.
  - $g(X_1, \ldots, X_n) = (X_i - \mu_{X_i})(X_j - \mu_{X_j})$ for $i \neq j$

    $\Rightarrow E[g(\mathbf{X})] = \sigma_{X_iX_j}$: *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 $i \neq j$

    $\Rightarrow E[g(\mathbf{X})] = \rho_{X_iX_j}$: *correlation coefficient* of X_i and X_j. g: 2nd order polynomial of $x_1, \ldots, x_n$
  - Notes. $\mu_{X_i}, \sigma^2_{X_i}, \sigma_{X_iX_j}, \rho_{X_iX_j}$ are constants, not r.v.'s.

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

Conditional Expectation ← conditional distribution LNp.6-44~51

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

$$E_{Y|X} \leftarrow E(h(\mathbf{Y})|\mathbf{X} = \mathbf{x}) = \sum_{\mathbf{y}\in\mathcal{Y}} h(\mathbf{y}) p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}),$$

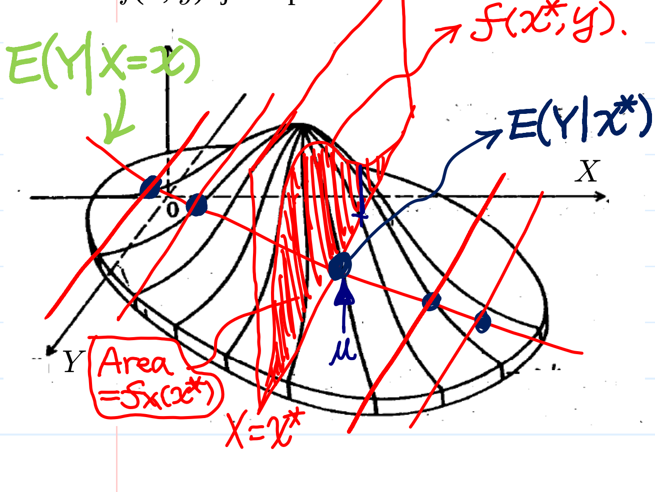
in the discrete case, or, $= \sum_z z\, P_{Z|X}(z|x)$

$$E_{Y|X} \leftarrow E(h(\mathbf{Y})|\mathbf{X} = \mathbf{x}) = \int_{\mathbb{R}^m} h(\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\, d\mathbf{y},$$

in the continuous case, provided that the sum or integral $= \int_z z f_{Z|X}(z|x)dz$ converges absolutely. ← i.e. exists

$\int (y-\mu) \cdot f(x^*, y)dy = 0 \Rightarrow \int y f(x^*, y)dy = \mu \int f(x^*, y)dy = \mu \cdot f_X(x^*)$

$f(x, y)$: joint pdf



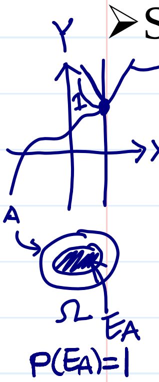
- $f(x, y)$: a joint pdf.
- Fix x^*, is $f(x^*, y)$ a pdf of y? i.e.,

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

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

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

- $E(Y|x^*)$: mean of $f_{Y|X}(y|x^*)$. $h(Y)=Y$
- Do it for any $x=x^*$, and get a function of $x \Rightarrow E(Y|x)$

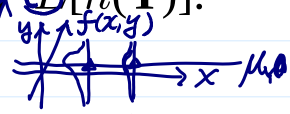


➢Some Notes.

It's a constant line for x.

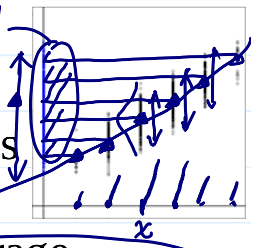
- $E(h(\mathbf{Y})|\mathbf{X}=\mathbf{x})$ is a function of $\mathbf{x}$ and is free of $\mathbf{Y}$.
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent, then $\overset{E_{Y|X}}{E}(h(\mathbf{Y})|\mathbf{X}=\mathbf{x})=\overset{E_Y}{E}[h(\mathbf{Y})]$.

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



- $E[\overset{Y}{h(\mathbf{X})}|\mathbf{X}=\mathbf{x}]=h(\mathbf{x})$.
- Let $g(\mathbf{x})=E[h(\mathbf{Y})|\mathbf{X}=\mathbf{x}]$, where $g:\mathbb{R}^n\to\mathbb{R}$, then we write $E(h(\mathbf{Y})|\mathbf{X})$ when $\mathbf{x}$ (a fixed value) replaced by $\mathbf{X}$ (a r.v.) in g.
  - Notice that $g(\mathbf{X})$ is a random variable.

E[Y|X]



➢Example. X=age (unit=year), Y=height (unit=cm)

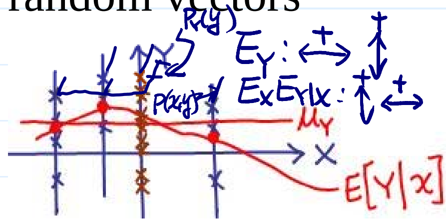
Q: What's the source of their randomness?

- $Y|X=x$: a random variable (unit=cm) that represents the height distribution of people with age=x. E[Y|x]
- $E(Y|X=x)$: a function maps from age (year) to average height (cm) of people with age=x. It is not a random variable.
- $E(Y|X)$: a random variable because it is a function of age, where age is treated as random. Notice that the unit of $E(Y|X)$ is "cm". $P(E[Y|X]=E[Y|x])=P(X=x)$
- $Var(Y|X=x)$ and $Var(Y|X)$ can be similarly defined.
- $E(Y)$: average height of *all* people; c.f.
  $Var(Y)$: variation of height of *all* people

- Theorem (Law of Total Expectation). For two random vectors $\mathbf{X}$ and $\mathbf{Y}$,

$$\overset{E_{XY}[h(Y)]}{\parallel}$$

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



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

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

use the example given in LNp.7-15 to realize what the terms mean.

Proof. (only prove it for the continuous case)

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

$$\parallel$$
$$E_XE_{Y|X}$$

generalization

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

➢Example. If a sample of n balls is drawn without replacement from a box containing R red balls, W white balls, and $N–R–W$ blue balls. Let

generalization of hypergeometric (similar to the generalization from binomial to multinomial)

$\boxed{X}$= # of red balls in the sample,

$\boxed{Y}$= # of white balls in the sample, then, the joint pmf of (X, Y) is

$$p_{X,Y}(x,y) = \frac{\binom{R}{x}\binom{W}{y}\binom{N-R-W}{n-x-y}}{\binom{N}{n}},$$

Find $E(Y)$.

Sol. Because $Y|X=x \sim$ hypergeometric$(n-x, N-R, W)$,

$$g(x) \equiv E(Y|X=x) = (n-x)[W/(N-R)].$$ LNp.4-35

Because $X \sim$ hypergeometric$(n, N, R) \Rightarrow E(X)=n(R/N)$, and then

$$\begin{aligned} E(Y) &= E_X[E_{Y|X}(Y|X)] = E_X[g(X)] \\ &= E_X\left[(n-X)\frac{W}{N-R}\right] = \frac{W}{N-R}[n - E_X(X)] \\ &= \frac{W}{N-R}\left(n - n\frac{R}{N}\right) = n\frac{W}{N}. \end{aligned}$$

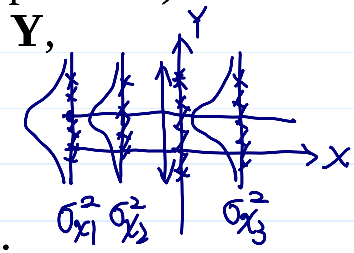
Note that $Y \sim$ hypergeometric$(n, N, W) \Rightarrow E(Y)=n(W/N)$.

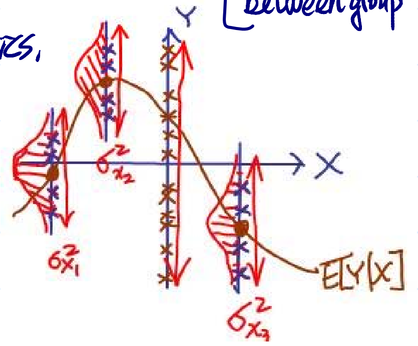
"within group."
"between group"

The concept leads to the "Analysis of Variance (ANOVA)"

Theorem (Variance Decomposition). For is statistics, two random vectors $\mathbf{X}$ and $\mathbf{Y}$,

$Var_{\mathbf{Y}}(Y_i)$
$= Var_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]$
$+ E_{\mathbf{X}}[Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]$.





Proof. $Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{x}) = E_{\mathbf{Y}|\mathbf{X}}\{[Y_i - E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{x})]^2|\mathbf{x}\}$

Note $Var(Z)=E(Z^2)-[E(Z)]^2$

$\quad\quad \stackrel{\vee}{=} E_{\mathbf{Y}|\mathbf{X}}(Y_i^2|\mathbf{x}) - [E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{x})]^2,$

and, $E_{\mathbf{X}}[Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]$

$\quad\quad = E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i^2|\mathbf{X})] - E_{\mathbf{X}}\{[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]^2\}.$

Also, $Var_{\mathbf{X}}[\boxed{E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})}] \equiv g(X)$

$\quad\quad \stackrel{\vee}{=} E_{\mathbf{X}}\{[\boxed{E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})}]^2\} - \{E_{\mathbf{X}}[\boxed{E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})}]\}^2.$

Now, $Var_{\mathbf{Y}}(Y_i) = E_{\mathbf{Y}}(Y_i^2) - [E_{\mathbf{Y}}(Y_i)]^2$

Law of Total expectation

$$\begin{aligned} &= E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i^2|\mathbf{X})] - \{E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]\}^2 \\ &= E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i^2|\mathbf{X})] - E_{\mathbf{X}}\{[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]^2\} \\ &\quad + E_{\mathbf{X}}\{[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]^2\} - \{E_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]\}^2 \\ &= E_{\mathbf{X}}[Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})] + Var_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]. \end{aligned}$$

➢Corollary.

- $Var_{\mathbf{Y}}(Y_i) \geq E_{\mathbf{X}}[Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]$ and the equality holds if and only if $E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X}) = E_{\mathbf{Y}}(Y_i)$ with probability one.

  $Var[E(Y|X)]=0 \Rightarrow E[Y|X]$ is a constant over x & $E_X E_{Y|X}(Y|X) = \mu_Y$.

- $Var_{\mathbf{Y}}(Y_i) \geq Var_{\mathbf{X}}[E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})]$ and the equality hold if and only if $Var_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})=0$ ($\Rightarrow Y_i = E_{\mathbf{Y}|\mathbf{X}}(Y_i|\mathbf{X})$) with probability one.

  $E[Var(Y|X)]=0.$

❖ **Reading**: textbook, Sec 7.5

Moment Generating Function

- Definition (Moment and Central Moment). If a random variable X has a cdf F_X, then

$$\mu_k \equiv E(X^k) = \int_{-\infty}^{\infty} x^k \, dF_X(x), \quad k = 1, 2, 3, \ldots,$$

are called the k^{th} *moments* of X provided that the integral converges absolutely, and

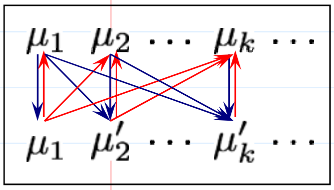
$$\mu'_k \equiv E[(X-\mu_X)^k] = \int_{-\infty}^{\infty} (x-\mu_X)^k \, dF_X(x), \quad k = 1, 2, 3, \ldots,$$

are called k^{th} *moment about the mean* μ_X or *central moment* of X provided that the integral converges absolutely.

➤ Some Notes.

- $\mu'_k = E[(X-\mu_X)^k] = E\left[\sum_{i=0}^{k} \binom{k}{i} (-\mu_X)^{n-i} X^i\right]$

$$= \sum_{i=0}^{k} \binom{k}{i} (-\mu_X)^{n-i} E(X^i) = \sum_{i=0}^{k} \binom{k}{i} (-\mu_X)^{n-i} \mu_i.$$

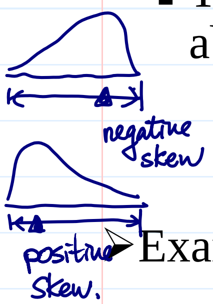


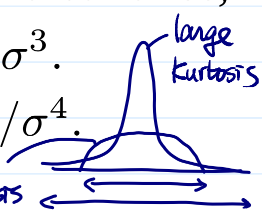
and,
$$\begin{aligned} \mu_k &= E(X^k) = E\{[(X-\mu_X)+\mu_X]^k\} \\ &= \sum_{i=0}^{k} \binom{k}{i} (\mu_X)^{n-i} E[(X-\mu_X)^i] \\ &= \sum_{i=0}^{k} \binom{k}{i} (\mu_X)^{n-i} \mu'_i. \end{aligned}$$

In particular,
$$\begin{aligned} E(X) &= \mu_X = \mu_1, \quad \text{and}, \\ Var(X) &= \sigma_X^2 = \mu_2 - \mu_1^2 = \mu'_2. \end{aligned}$$

$E(X^2) - [E(X)]^2$

- The (central) moments give a lot of useful information about the distribution, e.g., in addition to mean and variance,
  - Skewness (a measure of the asymmetry): μ'_3/σ^3.
  - Kurtosis (a measure of the "peakedness"): μ'_4/σ^4.

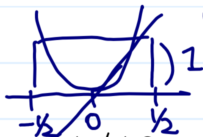


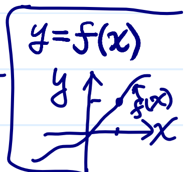


➤ Example (Uniform). If $X \sim$ Uniform(0, 1), then

$$\mu_k = \int_0^1 x^k \, dx = \frac{1}{k+1},$$

therefore,
$$\begin{aligned} \mu_X &= \mu_1 = 1/2, \quad \text{and}, \\ \sigma_X^2 &= \mu_2 - \mu_1^2 = 1/3 - (1/2)^2 = 1/12. \end{aligned}$$





And,
$$\begin{aligned} \mu'_k &= \int_0^1 (x-1/2)^k \, dx = \frac{1}{k+1}\left[(1/2)^{k+1} - (-1/2)^{k+1}\right] \\ &= \begin{cases} 0, & k \text{ is odd}, \\ \frac{1}{(k+1)2^k}, & k \text{ is even}. \end{cases} \end{aligned}$$

how to characterize a distribution?
① pdf/pmf
② cdf
③ mgf

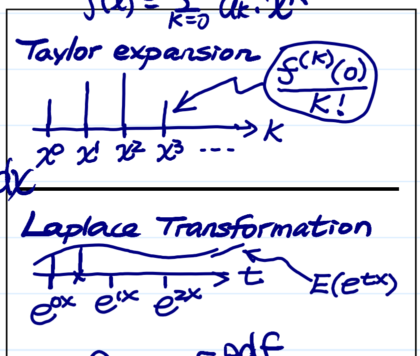
⊙ Definition (Moment Generating Function). If X is a random variable with the cdf F_X, then

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \, dF_X(x),$$

$f_X(x)dx$

is called the *moment generating function* (mgf) of X provided that the integral converges absolutely in some non-degenerate interval of t.

$f(x) = \sum_{k=0}^{\infty} a_k \cdot x^k$



$f(x)$: { pdf, pmf

➢ Some Notes.

- The mgf is a function of the variable t.
- The mgf may only exist for some particular values of t.

i.e. not for all $t \in \mathbb{R}$.

➢ Example.

- If X is a discrete r.v. taking on values x_i with probability p_i, i=1, 2, 3, …, then $M_X(t) = \sum_{i=1}^{\infty} e^{tx_i} p_i$.
- If $X \sim$ Poisson(λ), then for $-\infty < t < \infty$,

$$M_X(t) = \sum_{x=0}^{\infty} \left(e^{tx} \times \frac{e^{-\lambda}\lambda^x}{x!} \right)$$

$$= e^{-\lambda} \left(e^{\lambda e^t} \right) \sum_{x=0}^{\infty} \underbrace{\frac{e^{-(\lambda e^t)}(\lambda e^t)^x}{x!}}_{\text{pmf of Poisson}(\lambda e^t)} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}.$$

- If $X \sim$ Exponential(λ), then for $t < \lambda$,

$$M_X(t) = \int_0^{\infty} e^{tx} \times \lambda e^{-\lambda x} \, dx$$

$$= \lambda \left(\frac{1}{\lambda - t} \right) \int_0^{\infty} \underbrace{(\lambda - t) e^{-(\lambda - t)x}}_{\text{pdf of exponential}(\lambda - t)} \, dx = \frac{\lambda}{\lambda - t},$$

and $M_X(t)$ does not exist for $t \ge \lambda$.

this must be >0

- A list of some mgfs (exercise)
  - If $X \sim$ Binomial(n, p), use binomial expansion (LNp.4-20)

$$M_X(t) = (1 - p + pe^t)^n, \text{ for } t < -\log(1-p).$$

  - If $X \sim$ Negative Binomial(r, p), use negative binomial expasion (LNp.4-25)

$$M_X(t) = \left[\frac{pe^t}{1-(1-p)e^t} \right]^r, \text{ for } t < -\log(1-p).$$

  - If $X \sim$ Uniform(α, β), $M_X(t) = \frac{e^{\beta t} - e^{\alpha t}}{t(\beta - \alpha)}$.

use $e^u = \sum_{k=0}^{\infty} \frac{u^k}{k!}$ & use STO

  - If $X \sim$ Gamma(α, λ), use STO

$$M_X(t) = \left(\frac{\lambda}{\lambda - t} \right)^{\alpha}, \text{ for } t < \lambda.$$

  - If $X \sim$ Beta(α, β), $M_X(t) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha + r}{\alpha + \beta + r} \right) \frac{t^k}{k!}$

use STO

  - If $X \sim$ Normal(μ, σ^2), $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$.

Ee^{tx} mgf pdf/pmf cdf $\frac{d}{dx}$, $F(x)-F(x-)$

⊙ Theorem (Uniqueness Theorem). Suppose that the mgfs $M_X(t)$ and $M_Y(t)$ of random variables X and Y exist for all $|t| < h$ for some $h > 0$. If

$$M_X(t) = M_Y(t),$$

for $|t| < h$, then

$$F_X(z) = F_Y(z)$$

for all $z \in \mathbb{R}$, where F_X and F_Y are the cdfs of X and Y, respectively.

does not mean X=Y

i.e. an open interval containing zero

$-h$ 0 h t

distribution mgf Ee^{tx} one-to-one

Proof. Skipped (by the uniqueness theorem of Laplace transform.)

Fourier transformation

➢ Application of the uniqueness theorem

- When a moment generating function exists, there is a unique distribution corresponding to that mgf.
- This allows us to use mgfs to find distributions of transformed random variables in some cases.
- This technique is most commonly used for linear combinations of independent random variables

Why? check the Thms in LN p. 7-24~25

➢ Example. If $M_X(t) = p_1 e^{a_1 t} + \cdots + p_k e^{a_k t}$, where $p_1+\cdots+p_k=1$, then X is a discrete r.v. and its pmf is

$$p_X(x) = \begin{cases} p_i, & \text{for } x = a_i,\ i = 1, \ldots, k, \\ 0, & \text{otherwise.} \end{cases}$$

- Theorem (Moments and MGF). If $M_X(t)$ exist for $|t|<h$ for some $h>0$, then

$$M_X(0)=1,$$

and,

$$M_X^{(k)}(0) = \mu_k, \quad k = 1, 2, 3, \ldots$$

This explains why it's called moment generating function

Proof. First,

$$M_X(0) = \int_{-\infty}^{\infty} e^{0\cdot x}\, dF_X(x) = \int_{-\infty}^{\infty} 1\, dF_X(x) = 1.$$

$F(x)\Big|_{-\infty}^{\infty}$ ॥

$$M_X'(0) = \frac{d}{dt} M_X(t)\Big|_{t=0} = \left[\frac{d}{dt}\int_{-\infty}^{\infty} e^{tx}\, dF_X(x)\right]\Bigg|_{t=0}$$
$$= \int_{-\infty}^{\infty} \left(\frac{d}{dt} e^{tx}\Big|_{t=0}\right) dF_X(x) = \int_{-\infty}^{\infty} \left(xe^{tx}\big|_{t=0}\right) dF_X(x)$$
$$= \int_{-\infty}^{\infty} x \cdot 1\, dF_X(x) = E(X) = \mu_1.$$

$$\cdots = \cdots$$

$$M_X^{(k)}(0) = \frac{d^k}{dt^k} M_X(t)\Bigg|_{t=0} = \left[\frac{d^k}{dt^k}\int_{-\infty}^{\infty} e^{tx}\, dF_X(x)\right]\Bigg|_{t=0}$$
$$= \int_{-\infty}^{\infty} \left(\frac{d^k}{dt^k} e^{tx}\Bigg|_{t=0}\right) dF_X(x) = \int_{-\infty}^{\infty} \left(x^k e^{tx}\big|_{t=0}\right) dF_X(x)$$
$$= \int_{-\infty}^{\infty} x^k \cdot 1\, dF_X(x) = E(X^k) = \mu_k.$$

➢ Example. If X ~ Exponential(λ), then $M_X(t) = \frac{\lambda}{\lambda - t}$.

Because
$$M_X^{(k)}(t) = \frac{k!\lambda}{(\lambda-t)^{k+1}},$$
we get
$$\mu_k = M_X^{(k)}(0) = \frac{k!}{\lambda^k}.$$

We can use kth moments to calculate mean, variance, skewness, kurtosis, central moments, ---

- Theorem (MGF for linear transformation). For constants a and b,

$$M_{a+bX}(t) = e^{at} M_X(bt).$$

can be use to identify the dist of a+bX from the dist. of X.

Proof. $M_{\underbrace{a+bX}_{Y}}(t) = E[e^{t(a+bX)}] = e^{at} E[e^{(bt)X}] = e^{at} M_X(bt)$.

- Theorem (MGF for sum of independent r.v.'s). If $X_1, \ldots, X_n$ are independent each with mgfs $M_1(t), \ldots, M_n(t)$, respectively, then the mgf of $S = X_1 + \cdots + X_n$ is

$$\mathbb{R}^1 \to \mathbb{R}^1: \boxed{M_S}(t) = M_1(t) \times \cdots \times M_n(t). \quad \text{---- } (*)$$

Proof. $M_S(t) = E(e^{tS}) = E[e^{t(X_1+\cdots+X_n)}]$

$$= E(e^{tX_1} \times \cdots \times e^{tX_n}) = E(e^{tX_1}) \times \cdots \times E(e^{tX_n})$$

$$= M_1(t) \times \cdots \times M_n(t).$$

$\because$ independent

Note: Geo. is a special case of negative binomial $r=1$

➢Example. If $X_1, \ldots, X_n$ are i.i.d. ~ Geometric(p), then $S = X_1 + \cdots + X_n$ ~ Negative Binomial(n, p).

use convolution (LNp.6-25) to prove it

Proof. $M_S(t) = M_{X_1}(t) \times \cdots \times M_{X_n}(t)$

$$= \frac{pe^t}{1-(1-p)e^t} \times \cdots \times \frac{pe^t}{1-(1-p)e^t} = \left[\frac{pe^t}{1-(1-p)e^t}\right]^n.$$

c.f.

➢Example. If $X_1, \ldots, X_n$ are independent and

$X_i \sim$ Normal(μ_i, σ_i^2), for $i=1, \ldots, n$.

convolution approach LNp.6-27~29

Let $S = a_0 + a_1X_1 + \cdots + a_nX_n$, then

$$S \sim \text{Normal}(a_0 + a_1\mu_1 + \cdots + a_n\mu_n, a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2).$$

Proof. $M_S(t) = e^{a_0 t} \times \prod_{i=1}^n e^{\mu_i(a_i t) + \frac{\sigma_i^2 (a_i t)^2}{2}}$

$$= e^{\overbrace{(a_0 + a_1\mu_1 + \cdots + a_n\mu_n)}^{\mu'=}t + \frac{\overbrace{(a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2)}^{=\sigma'^2} t^2}{2}}.$$

- Definition (Joint Moment Generating Function). For random variables $X_1, \ldots, X_n$, their joint mgf is defined as

$$\mathbb{R}^n \to \mathbb{R}^1: \boxed{M_{X_1,\ldots,X_n}}(t_1, \ldots, t_n) = E(e^{t_1 X_1 + \cdots + t_n X_n})$$

provided that the expectation exists.

different t_i's

c.f. (*) in LNp.7-25.

➢Example. If $X_1, \ldots, X_m \sim$ Multinomial(n, m, p_1, …, p_m),

The multinomial expansion in LNp.6-13

$$M_{X_1,\ldots,X_m}(t_1, \ldots, t_m)$$

$$= \sum_{x_1+\cdots+x_m=n} e^{t_1x_1+\cdots+t_mx_m} \binom{n}{x_1, \ldots, x_m} p_1^{x_1} \cdots p_m^{x_m}$$

$$= \sum_{x_1+\cdots+x_m=n} \binom{n}{x_1, \ldots, x_m} (p_1e^{t_1})^{x_1} \cdots (p_m e^{t_m})^{x_m}$$

$$= (p_1 e^{t_1} + \cdots + p_m e^{t_m})^n.$$

relationship between joint mgf & marginal mgf

- Some Properties of Joint mgf

➢ $M_{X_1}(t) = M_{X_1,X_2,\ldots,X_n}(t, 0, \ldots, 0)$.

➢ uniqueness theorem

compare it with the (*) in LNp.7-25

➢ $X_1, \ldots, X_n$ are independent if and only if

Note: same property holds for cdf, pdf/pmf

$$M_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = M_{X_1}(t_1) \times \cdots \times M_{X_n}(t_n).$$

different t_i's

➢ $\frac{\partial^{k_1+\cdots+k_n}}{\partial t_1^{k_1} \cdots \partial t_n^{k_n}} M_{X_1,\ldots,X_n}(0, \ldots, 0) = E(X_1^{k_1} \times \cdots \times X_n^{k_n})$.

❖ **Reading**: textbook, Sec 7.7