Lecture 2. Multivariate Distributions

Multivariate distribution

for f two arguments. "and"

- $F_{X,Y}(x,y) = P\{X \leq x, Y \leq y\}$ is joint cdf for two random variables
- X and Y are said to have *joint* pdf $f_{X,Y}$: $f_{X,Y}(x,y) \geq 0$ for all $x, y \in \mathbb{R}$ and $F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)dtds$ $\quad f_{x,y}(x,y) = \frac{\partial^2 F_{xy}}{\partial x \partial y}$
- marginal distribution (distribution of one variable) → can figure out
  $F_X(x) = P\{X \leq x\} = F_{X,Y}(x,\infty)$ and $x \leq \infty$
- for continuous: $F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(s,t)dtds$ pdf for x only
  $f_X(s) = \int_{-\infty}^{+\infty} f(s,t)dt$

Two distbn of two var's known: but can't get one, need to know if they truly related or not

Conditional distribution

how can I report Y distn given I know x

- If X and Y are discrete, *conditional* is defined as

A B joint

$$\mathbb{P}_{Y|X}(y|X=x) \overset{def}{=} \frac{\mathbb{P}[(Y=y) \text{ and } (X=x)]}{\mathbb{P}(X=x)}$$

B

for values of x with $\mathbb{P}(X=x) > 0$.

- *conditional* pdf of Y given $X=x$ (for x such that $f_X(x) > 0$):

joint on top
prob. of cond.

$$\boxed{f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x)}$$

cond defn.

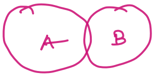
to shrink

- measure theory is used to define conditional in more complicated cases

know its pdf: non -ve; integrate wrt y.

$$\int f_{y|x}(y|x)\,dy = \int \frac{f_{x,y}(x,y)}{f_x(x)}\,dy = \frac{1}{f_x(x)} \int f_{x,y}(x,y)\,dy$$

going from joint to marginal



$P(A|B) = \frac{P(A \cap B)}{P B}$

Conditional distribution — is a random var — have mean/variances.

$$f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x)$$

- conditional distribution characterizes how Y is distributed for a given $X = x$
- $\mathbb{P}\{Y \in A|X = x\} = \int_A f_{Y|X}(y|x)dy$.
- $\mathbb{E}[Y|X = x] = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x)dy$

Conditional distribution

- properties of conditional expectation
  1. $\mathbb{E}[g(X)Y|X = x] = g(x)\mathbb{E}[Y|X = x]$ now x there is realised
  2. *the law of iterated expectations*: weight gender : eg: avg height of whole class (Y) → avg heights by gender - then avg. it

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

.

$$E[E[Y|X]] \overset{\text{means}}{=} \int E[Y|X=x] f_x(x)\, dx$$

$$= \int \int y f_{y|x}(y|x)\, dy\, f_x(x)\, dx$$

$$= \iint y \frac{f_{x,y}(x,y)}{\cancel{f_x(x)}}\, dy\, \cancel{f_x(x)}\, dx$$

$$= EY$$

Empty slide

□

Independence

- Random variables X and Y are said to be *independent* if
$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$
- it implies $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
- for conditional distribution it means $f_{Y|X}(y|x) = f_Y(y)$ for all $x \in \mathbb{R}$
- If X and Y are independent, then $g(X)$ and $f(Y)$ are also independent for any functions
- if X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

new random variable

$$\mathbb{E}[XY] = \iint xy\, f_{x,y}(x,y)\,dx\,dy$$

$$= \iint xy\, f_x(x) f_y(y)\,dx\,dy$$

$$= \iint [x f_x(x)] \cdot [y f_y(y)]\,dx\,dy$$

$$= \underbrace{\int x f(x)\,dx}_{\mathbb{E}X} \cdot \underbrace{\int y f_y(y)\,dy}_{\mathbb{E}Y}$$

no info about each other

then covariance = 0

bec

$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

Covariance * (on avg):

deviation of x from its mean. ψ_1

covariance $cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$

- ☐ $cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- ☐ if X and Y are independent, then $cov(X, Y) = 0$ → not vice versa
- ☒ $cov(aX + c, bY + d) = ab \cdot cov(X, Y)$ for random variables X and Y and constants a, b, c and d — but when multiply wrt. const. Δ in scale.
- ☐ $cov(X, Y) = cov(Y, X)$

const.

$$\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$$

random

const.

$$= \mathbb{E}[XY - \mathbb{E}X \cdot Y - X \cdot \mathbb{E}Y + \mathbb{E}X \cdot \mathbb{E}Y]$$

push Exp. to have linearity

$$= \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y - \mathbb{E}X \cdot \mathbb{E}Y + \mathbb{E}X \cdot \mathbb{E}Y$$

$$= \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y$$

$Var(X) = cov(X, X)$

Covariance

corr 1, −1 : perfectly linearly corr

* PSet !

- ☐ Cauchy-Schwarz: $|cov(X,Y)| \le \sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}$
- ☐ $\mathbb{V}ar(X+Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y) + 2cov(X,Y)$
- ☐ if $X_1, ..., X_n$ are independent, then $\mathbb{V}ar(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{V}ar(X_i)$

$corr(X,Y) = \frac{cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

$|corr(X,Y)| \le 1$

$(X - aY)$, wts the variance $Var(X - aY)$

$= E(X - EX - a(Y - EY))^2$

$= \underbrace{E(X-EX)^2}_{Var(X)} - \underbrace{2aE(X-EX)(Y-EY)]}_{2a\, cov(X,Y)} + \underbrace{a^2 E(Y-EY)^2}_{a^2 Var(Y)}$

$= Var(X) - 2a\,cov(X,Y) + a^2 Var(Y)$

OLS coef

$a = \frac{cov(X,Y)}{Var(Y)} \xrightarrow{\text{plug in}} = Var(X) - 2\frac{cov(X,Y)}{Var(Y)} \cdot cov(X,Y) + \frac{cov^2(X,Y)}{Var^2(Y)} \cdot Var(Y)$

proved the Cauchy. keep in econometrics.

$\therefore = Var(X) - \frac{cov^2(X,Y)}{Var(Y)} \ge 0$

$cov(X,Y)^2 \le Var(X)\,Var(Y)$

Empty slide

- $\operatorname{var}(X+Y) = E\left([X-EX) + (Y-EY)]^2\right)$

push E $= E\left([X-EX]^2 + [Y-EY]^2 + 2[X-EX]\,(Y-EY]\right) =$

$= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{cov}(X,Y)$ [some more spread out, smaller, narrow

If X & Y independent: $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$ yes. cov is 0.

Normal Random Variables

. Multivariate gaussians. given by μ

$n \times n$

- Let Σ be a positive definite $n \times n$ matrix.
- Let $\underset{}{\overset{\text{mean}}{\mu}}$ be $n \times 1$ vector.
- Definition: $X \sim N(\mu, \Sigma)$ if X is continuous and its pdf is given by

n random var $n \times 1$

$1 \times n$ $n \times n$ $n \times 1$ scalar

$$\text{joint pdf} \rightarrow f_X(x) = \frac{\exp(-(x-\mu)^T \Sigma^{-1} (x-\mu)/2)}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}}$$

for any $n \times 1$ vector x.

Normal Random Variables

$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$ $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$ $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{1n} \\ & \Sigma_{22} & \\ & \ddots & \Sigma_{nn} \end{bmatrix}$

if $X \sim N(\mu, \Sigma)$, then

var cov matrix

on diagonal $\Sigma_{ii} = Var(X_i)$

$\Sigma_{ij} = cov(X_i, X_j)$

- ☐ $\Sigma_{ij} = cov(X_i, X_j)$ for any $i, j = 1, ..., n$ where $X = (X_1, ..., X_n)^T$
- ☐ $\mu_i = \mathbb{E}[X_i]$ for any $i = 1, ..., n$
- ☐ any subset of components of X is normal as well
- ☐ $X_i \sim N(\mu_i, \Sigma_{ii})$

any component of Gaussian is Gaussian

Empty slide

$\square$

Normal Random Variables

A, B are constants

- if X and Y are jointly normal and uncorrelated then they are independent

$X \sim N(\mu, \Sigma)$

$\underset{k\times 1}{Y} = \underset{k\times n}{A}\underset{n\times 1}{X} + \underset{k\times 1}{B}$

- linear transformation of normal is normal
- if $X \sim N(\mu, \Sigma)$ is $n \times 1$ normal vector, and A is a fixed $k \times n$ full-rank matrix with $k \leq n$, then
  $Y = AX$ is a normal $k \times 1$ vector: $Y \sim N(A\mu, A\Sigma A^T)$

Claim: Y is gaussian so write $EY = E(AX + B)$

$= \underbrace{A\mu}_{a} + B$

$\underset{k\times k}{\operatorname{var}(Y)} = \underset{k\times n}{A}\,\underset{n\times n}{\Sigma}\,\underset{n\times k}{A^T}$

Normal Random Variables

corresp. "means" & variances'   when two rv's correlated give high corr, more info

X_1, X_2 are vectors

- if

$$X = \begin{bmatrix} \underbrace{X_1}_{k_1 \times 1} \\ \underbrace{X_2}_{k_2 \times 1} \end{bmatrix} \sim N\left(\begin{bmatrix} \underbrace{\mu_1}_{k_1 \times 1} \\ \underbrace{\mu_2}_{k_2 \times 1} \end{bmatrix}, \begin{bmatrix} \underbrace{\Sigma_{11}}_{k_1 \cdot k_1} & \underbrace{\Sigma_{12}}_{k_1 \cdot k_2} \\ \underbrace{\Sigma_{21}}_{k_2 \times k_1} & \underbrace{\Sigma_{22}}_{k_2 \cdot k_2} \end{bmatrix} \right)$$

now, x_2 is realised

then $X_1|X_2 = x_2 \sim N(\tilde{\mu}, \tilde{\Sigma})$ with

$$\tilde{\mu} = \underbrace{\mu_1}_{k_1 \cdot 1} + \underbrace{\Sigma_{12}}_{k_1 \cdot k_2}\underbrace{\Sigma_{22}^{-1}}_{k_2 \cdot k_2}\underbrace{(x_2 - \mu_2)}_{k_2 \cdot 1} \rightarrow \text{formula for mean.}$$

$$\text{var} \quad \underline{\tilde{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}$$

- if X_1 and X_2 are both random variables, then

scalars.

spl. case formula

$$\star \quad \mathbb{E}[X_1|X_2 = x_2] = \mu_1 + \underbrace{\frac{cov(X_1, X_2)}{\mathbb{V}ar(X_2)}(x_2 - \mu_2)}_{\text{correction/given info}}$$

↑ expected

|| OLS formula

Empty slide

χ^2-distribution

in tests with table

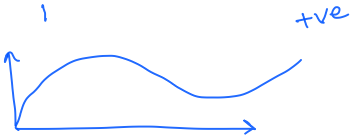
- If $X_1, \dots, X_n$ are i.i.d. $N(0,1)$

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2$$

is called a χ^2 random variable with n degrees of freedom.

$$E\chi^2_n = E\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} \underbrace{E x_i^2}_{1} = n$$

$$\mathrm{Var}(\chi^2_n) = 2n$$



Gaussian family

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

If $X_1, ..., X_n$ are iid random variables with $N(\mu, \sigma^2)$ distribution, then idpen. indenticaly disbuted.

avg (1) $\overline{X}_n \sim N(\mu, \sigma^2/n)$

(2) $\overline{X}_n$ and s_n^2 are independent,

(3) $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$

to prove.

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$: avg is guassion, is a linear trans, normal.

$E\bar{X} = E\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum E x_i = \frac{1}{n}\sum_{1}^{n}\mu = \mu$

$Var(\bar{X}) = Var\left(\frac{1}{n}\sum x_i\right) = \frac{1}{n^2} Var\left(\sum_{1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i)$

$= \frac{n\sigma^2}{n^2} = \sigma^2/n$

Empty slide