CH 5. Multivariate Probability Distributions

Jeongyoun Ahn

KAIST

Univariate

↓

Bivariate

↓

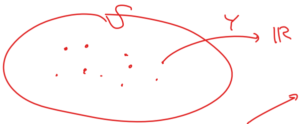
Multivariate

↓

High-dimension

Bivariate data

$\to (Y_1, Y_2) \to$ must be from the same Sample space!!



$P(ht < 170, wt > 50) = P(\underbrace{\{\omega : ht(\omega) < 170\} \cap \{\omega : wt(\omega) > 50\}}_{\text{Both} \subset S})$

- e.g) (height, weight), (age, income), and so on
- Purpose of collecting bivariate data
  - Are the variables related?
  - What form of relationship is indicated by the data?: linear, quadratic, etc.
  - Can we quantify the strength of their relationship?: strong, weak
  - Can we predict one variable from the other? $\to$ Conditional Expectation
- Can be generalized to multivariate data: (age, education level, income)

Joint (or Bivariate) Probability Distribution

CDF
MGF
PDF
PMF

Joint PMF for discrete r.v.

Intersection : $P(\{Y_1 = y_1\} \cap \{Y_2 = y_2\})$

- $p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2), \quad -\infty < y_1, y_2 < \infty.$
- 1 $p(y_1, y_2) \geq 0$ for all y_1, y_2.
- $\sum_{y_1, y_2} p(y_1, y_2) = 1.$

Marginal Distribution of Y_2

$p(1) = 6/8$
$p(3) = 2/8$

Condition. of Y_2 given $Y_1 = 1$.

$p(1 \mid 1) = 1$
$p(3 \mid 1) = 0$

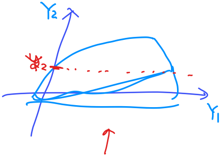
Eg 1 Toss 3 coins

Y_1 : # H , Y_2 : $|\#H - \#T|$

$Y_2 \backslash Y_1$	0	1	2	3
1	0	3/8	3/8	0
3	1/8	0	0	1/8

Joint PDF $f(y_1, y_2)$

- $f(y_1, y_2) \geq 0$ for all y_1, y_2.
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 = 1.$



Eg 2

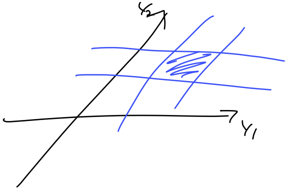
$f(y_1, y_2) = 24 y_1 y_2$

$0 \leq y_1 \leq 1$. $0 \leq y_2 \leq 1$.

$y_1 + y_2 \leq 1$

Joint CDF

$$F(y_1, y_2) = P(Y_1 \le y_1, Y_2 \le y_2)$$



$$F(-\infty, -\infty) = 0$$
$$F(-\infty, y_2) = 0 = F(y_1, -\infty)$$
$$F(\infty, \infty) = 1$$
$$F(\infty, y_2) = F(y_2) = P(Y_2 \le y_2)$$

$$P(a < Y_1 \le b,\ c < Y_2 \le d) = F(b, d) - F(b, c) - F(a, d) + F(a, c)$$

$\boxed{\text{MGF}}$

$\underline{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$ $\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$

$$m(t_1, t_2) = E\left(e^{t_1 Y_1 + t_2 Y_2}\right) = E\left(e^{\langle \underline{t}, \underline{Y} \rangle}\right)$$

$$E(Y_1^2 Y_2) = \ldots$$

Eg 2

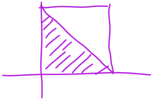
$f(y_1, y_2) = 24 y_1 y_2$ $y_1 + y_2 \le 1$

$\int_0^1 \int_0^1 24 y_1 y_2 \, dy_1 dy_2$

$Y_1 + Y_2 \le 1$
$Y_1 - Y_2 > .5$

$P(Y_1 - Y_2 > .5) =$

$P(Y_1 Y_2 < .5) =$



Marginal and Conditional Probability Distributions

Just the univariate density, emphasizing that we look one variable only.

Y_1, Y_2: continuous r.v. with $\boxed{f(y_1, y_2)}$

Marginal density of Y_1 (or Y_2)

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)dy_2 \quad \text{and} \quad f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)dy_1.$$

Must not depend on Y_2

Conditional density of Y_1 given $Y_2 = y_2$

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

: a ft of Y_2 Y_1

provided that $f_2(y_2) > 0$.