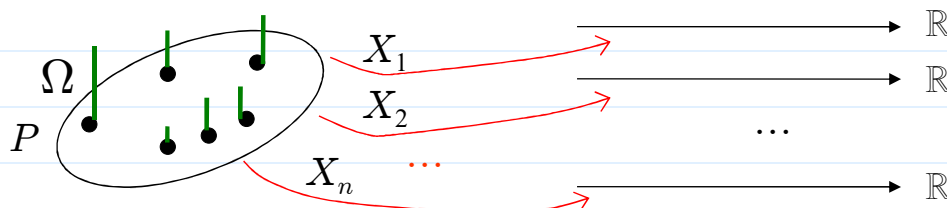
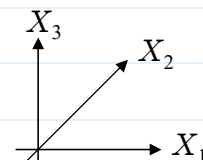


Jointly Distributed Random Variables

- Recall. In Chapters 4 and 5, focus on univariate random variable.

➤ However, often a single experiment will have more than one random variables which are of interest.



➤ Definition. Given a sample space Ω and a probability measure P defined on the subsets of Ω , random variables

$$X_1, X_2, \dots, X_n: \Omega \rightarrow \mathbb{R}$$

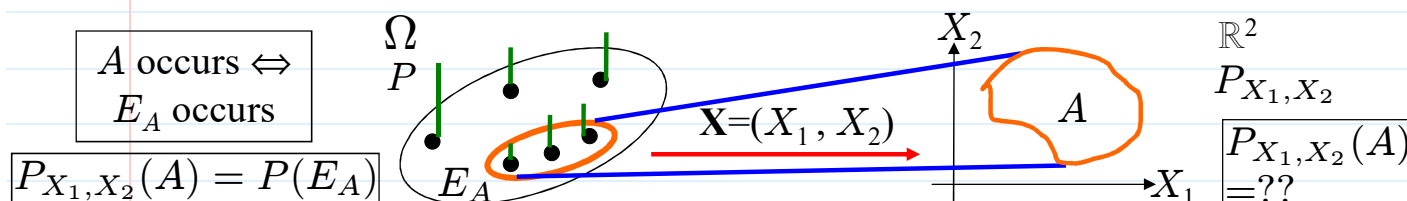
are said to be jointly distributed.

- We can regard n jointly distributed r.v.'s as a random vector $\mathbf{X} = (X_1, \dots, X_n): \Omega \rightarrow \mathbb{R}^n$.

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- Q: For $A \subset \mathbb{R}^n$, how to define the probability of $\{\mathbf{X} \in A\}$ from P ?

p. 7-2

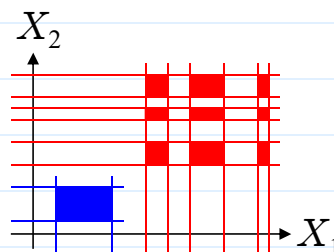


➤ For $A \subset \mathbb{R}^n$,

$$P_{X_1, \dots, X_n}(A) = P(\{\omega \in \Omega \mid (X_1(\omega), \dots, X_n(\omega)) \in A\})$$

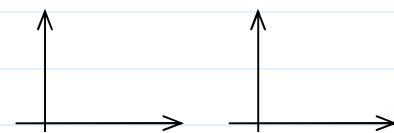
➤ For $A_i \subset \mathbb{R}$, $i=1, \dots, n$,

$$P_{X_1, \dots, X_n}(X_1 \in A_1, \dots, X_n \in A_n) = P(\{\omega \in \Omega \mid X_1(\omega) \in A_1\} \cap \dots \cap \{\omega \in \Omega \mid X_n(\omega) \in A_n\})$$



➤ Definition. The probability measure of \mathbf{X} ($P_{\mathbf{X}}$, defined on subsets of \mathbb{R}^n) is called the joint distribution of X_1, \dots, X_n . The probability measure of X_i (P_{X_i} , defined on subsets of \mathbb{R}) is called the marginal distribution of X_i .

- **Q:** Why need joint distribution? Why are marginal distributions not enough?



- Example (Coin Tossing, Toss a fair coin 3 times, LNp.5-3).

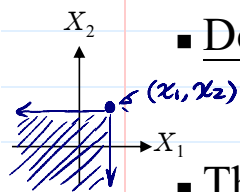
X_2 : # of head on 1 st toss	X_1 : total # of heads			
	0 (1/8)	1 (3/8)	2 (3/8)	3 (1/8)
0 (1/2)	1/8 [1/16]	2/8 [3/16]	1/8 [3/16]	0 [1/16]
1 (1/2)	0 [1/16]	1/8 [3/16]	2/8 [3/16]	1/8 [1/16]

- **blue numbers**: joint distribution of X_1 and X_2
- (black numbers): marginal distributions
- **[red numbers]**: joint distribution of another (X_1', X_2')
- Some findings:
 - When joint distribution is given, its corresponding marginal distributions are known, e.g.,
 - ◆ $P(X_1=i)=P(X_1=i, X_2=0)+P(X_1=i, X_2=1), i=0, 1, 2, 3.$

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- (X_1, X_2) and (X_1', X_2') have identical marginal distributions but different joint distributions.
 - ◆ When the marginal distributions are given, the corresponding joint distribution is still unknown. There could be many possible different joint distributions. (A special case: X_1, \dots, X_n are independent.)
 - Joint distribution offers more information, e.g.,
 - ◆ When not observing X_1 , the distribution of X_2 is:
 $P(X_2=0)=1/2, P(X_2=1)=1/2 \Rightarrow$ marginal distribution
 - ◆ When X_1 was observed, say $X_1=1$, the distribution of X_2 is: $P(X_2=0|X_1=1)=(2/8)/(3/8)=2/3$ and $P(X_2=1|X_1=1)=(1/8)/(3/8)=1/3 \Rightarrow$ the calculation requires the knowing of joint distribution
- We can characterize the joint distribution of \mathbf{X} in terms of its
 1. Joint Cumulative Distribution Function (joint cdf)
 2. Joint Probability Mass (Density) Function (joint pmf or pdf)
 3. Joint Moment Generating Function (joint mgf, Chapter 7)

➤ Joint Cumulative Distribution Function



■ Definition. The joint cdf of $\mathbf{X}=(X_1, \dots, X_n)$ is defined as

$$F_{\mathbf{X}}(x_1, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n).$$

■ Theorem. Suppose that $F_{\mathbf{X}}$ is a joint cdf. Then,

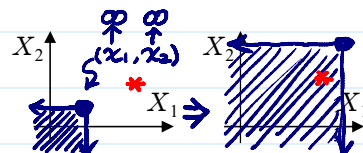
(i) $0 \leq F_{\mathbf{X}}(x_1, \dots, x_n) \leq 1$, for $-\infty < x_i < \infty, i=1, \dots, n$.

(ii) $\lim_{x_1, x_2, \dots, x_n \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_n) = 1$

Proof. Let $z_{im} \uparrow \infty, 1 \leq i \leq n$.

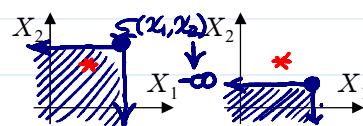
Let $A_m = (-\infty, z_{1m}] \times \dots \times (-\infty, z_{nm}]$.

Then, $A_m \uparrow \mathbb{R}^n \Rightarrow \lim P(A_m) = P(\mathbb{R}^n) = 1$.



(iii) For any $i \in \{1, \dots, n\}$,

$$\lim_{x_i \rightarrow -\infty} F_{\mathbf{X}}(x_1, \dots, x_n) = 0.$$

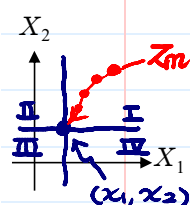


Proof. Let $z_{im} \downarrow -\infty$, for some i .

Let $A_m = (-\infty, x_1] \times \dots \times (-\infty, z_{im}] \times \dots \times (-\infty, x_n]$

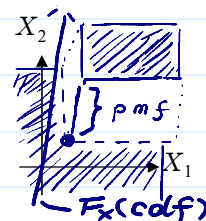
Then, $A_m \downarrow \emptyset \Rightarrow \lim P(A_m) = P(\emptyset) = 0$.

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

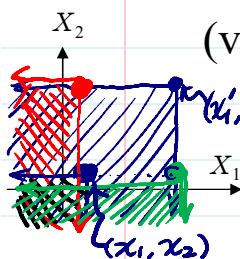


(iv) $F_{\mathbf{X}}$ is continuous from the right with respect to each of the coordinates, or any subset of them jointly, i.e., if $\mathbf{x}=(x_1, \dots, x_n)$ and $\mathbf{z}_m=(z_{1m}, \dots, z_{nm})$ such that $\mathbf{z}_m \downarrow \mathbf{x}$, then

$$F_{\mathbf{X}}(\mathbf{z}_m) \downarrow F_{\mathbf{X}}(\mathbf{x}).$$



(v) If $x_i \leq x'_i, i = 1, \dots, n$, then

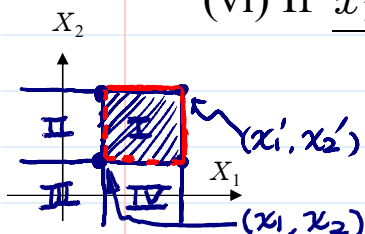


$$F_{\mathbf{X}}(x_1, \dots, x_n) \leq F_{\mathbf{X}}(t_1, \dots, t_n) \leq F_{\mathbf{X}}(x'_1, \dots, x'_n).$$

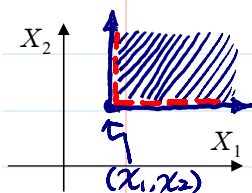
where $t_i \in \{x_i, x'_i\}, i = 1, 2, \dots, n$. When $n=2$, we have

$$F_{X_1, X_2}(x_1, x_2) \leq \left\{ \begin{array}{l} F_{X_1, X_2}(x_1, x'_2) \\ F_{X_1, X_2}(x'_1, x_2) \end{array} \right\} \leq F_{X_1, X_2}(x'_1, x'_2).$$

(vi) If $x_1 \leq x'_1$ and $x_2 \leq x'_2$, then



$$\begin{aligned} P(x_1 < X_1 \leq x'_1, x_2 < X_2 \leq x'_2) \\ &= F_{X_1, X_2}(x'_1, x'_2) - F_{X_1, X_2}(x_1, x'_2) \\ &\quad - F_{X_1, X_2}(x'_1, x_2) + F_{X_1, X_2}(x_1, x_2). \end{aligned}$$

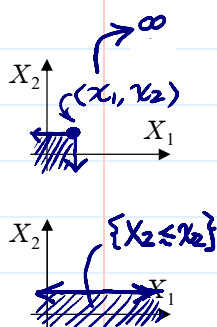


In particular, let $\underline{x}'_1 \uparrow \infty$ and $\underline{x}'_2 \uparrow \infty$, we get

$$\begin{aligned} P(x_1 < X_1 < \infty, x_2 < X_2 < \infty) \\ = 1 - F_{X_1}(x_1) - F_{X_2}(x_2) + F_{X_1, X_2}(x_1, x_2). \end{aligned}$$

(vii) The joint cdf of $\underline{X}_1, \dots, \underline{X}_k, k < n$, is

$$\begin{aligned} F_{X_1, \dots, X_k}(x_1, \dots, x_k) &= P(X_1 \leq x_1, \dots, X_k \leq x_k) \\ &= P(X_1 \leq x_1, \dots, X_k \leq x_k, \\ &\quad -\infty < X_{k+1} < \infty, \dots, -\infty < X_n < \infty) \\ &= \lim_{x_{k+1}, x_{k+2}, \dots, x_n \rightarrow \infty} F_{\mathbf{X}}(x_1, \dots, x_k, x_{k+1}, \dots, x_n). \end{aligned}$$



In particular, the marginal cdf of \underline{X}_1 is

$$\begin{aligned} F_{X_1}(x) &= P(X_1 \leq x) \\ &= \lim_{x_2, x_3, \dots, x_n \rightarrow \infty} F_{\mathbf{X}}(x, x_2, x_3, \dots, x_n). \end{aligned}$$

- Theorem. A function $F_{\mathbf{X}}(x_1, \dots, x_n)$ can be a joint cdf if $F_{\mathbf{X}}$ satisfies (i)-(v) in the previous theorem.

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

➤ Joint Probability Mass Function

- Definition. Suppose that $\underline{X}_1, \dots, \underline{X}_n$ are discrete random variables. The joint pmf of $\mathbf{X} = (X_1, \dots, X_n)$ is defined as

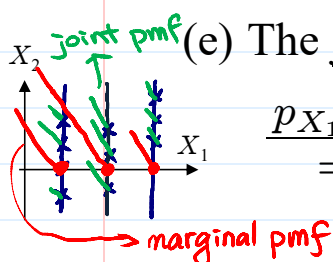
$$p_{\mathbf{X}}(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n).$$

- Theorem. Suppose that $p_{\mathbf{X}}$ is a joint pmf. Then,

- $p_{\mathbf{X}}(x_1, \dots, x_n) \geq 0$, for $-\infty < x_i < \infty, i = 1, \dots, n$.
- There exists a finite or countably infinite set $\mathcal{X} \subset \mathbb{R}^n$ such that $p_{\mathbf{X}}(x_1, \dots, x_n) = 0$, for $(x_1, \dots, x_n) \notin \mathcal{X}$.
- $\sum_{\mathbf{x} \in \mathcal{X}} p_{\mathbf{X}}(\mathbf{x}) = 1$, where $\mathbf{x} = (x_1, \dots, x_n)$.
- For $A \subset \mathbb{R}^n$, $P(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A \cap \mathcal{X}} p_{\mathbf{X}}(\mathbf{x})$.

(e) The joint pmf of $\underline{X}_1, \dots, \underline{X}_k, k < n$, is

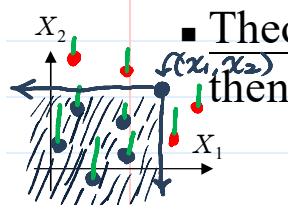
$$\begin{aligned} p_{X_1, \dots, X_k}(x_1, \dots, x_k) &= P(X_1 = x_1, \dots, X_k = x_k) \\ &= P(X_1 = x_1, \dots, X_k = x_k, \\ &\quad -\infty < X_{k+1} < \infty, \dots, -\infty < X_n < \infty) \\ &= \sum_{\substack{(x_1, \dots, x_n) \in \mathcal{X} \\ -\infty < x_{k+1} < \infty, \dots, -\infty < x_n < \infty}} p_{\mathbf{X}}(x_1, \dots, x_k, x_{k+1}, \dots, x_n). \end{aligned}$$



In particular, the marginal pmf of \underline{X}_1 is

$$\begin{aligned} p_{X_1}(x) &= P(X_1 = x) \\ &= \sum_{\substack{(x, x_2, \dots, x_n) \in \mathcal{X} \\ -\infty < x_2 < \infty, \dots, -\infty < x_n < \infty}} p_{\mathbf{X}}(x, x_2, x_3, \dots, x_n). \end{aligned}$$

■ Theorem. A function $p_{\mathbf{X}}(x_1, \dots, x_n)$ can be a joint pmf if $p_{\mathbf{X}}$ satisfies (a)-(c) in the previous theorem.



■ Theorem. If $F_{\mathbf{X}}$ and $p_{\mathbf{X}}$ are the joint cdf and joint pmf of $\underline{\mathbf{X}}$, then

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \sum_{\substack{(t_1, \dots, t_n) \in \mathcal{X} \\ t_1 \leq x_1, \dots, t_n \leq x_n}} p_{\mathbf{X}}(t_1, \dots, t_n).$$

To derive $p_{\mathbf{X}}$ from $F_{\mathbf{X}}$, take $n=2$ to illustrate:

$$\begin{aligned} p_{\mathbf{X}}(x_1, x_2) &= \lim_{m \rightarrow \infty} P \left(x_1 - \frac{1}{m} < X_1 \leq x_1 + \frac{1}{m}, x_2 - \frac{1}{m} < X_2 \leq x_2 + \frac{1}{m} \right) \\ &= \lim_{m \rightarrow \infty} \left[F_{\mathbf{X}}(x_1 + 1/m, x_2 + 1/m) - F_{\mathbf{X}}(x_1 + 1/m, x_2 - 1/m) \right. \\ &\quad \left. - F_{\mathbf{X}}(x_1 - 1/m, x_2 + 1/m) + F_{\mathbf{X}}(x_1 - 1/m, x_2 - 1/m) \right] \\ &= F_{\mathbf{X}}(x_1, x_2) - F_{\mathbf{X}}(x_1, x_2-) - F_{\mathbf{X}}(x_1-, x_2) + F_{\mathbf{X}}(x_1-, x_2-) \end{aligned}$$

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

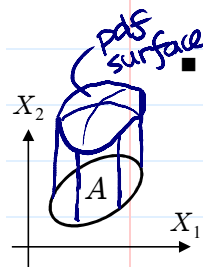
➤ Joint Probability Density Function

p. 7-10

■ Definition. A function $f_{\mathbf{X}}(x_1, \dots, x_n)$ can be a joint pdf if

(1) $f_{\mathbf{X}}(x_1, \dots, x_n) \geq 0$, for $-\infty < x_i < \infty, i=1, \dots, n$.

(2) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \cdots dx_n = 1$.



■ Definition. Suppose that $\underline{X}_1, \dots, \underline{X}_n$ are continuous r.v.'s.

The joint pdf of $\mathbf{X}=(X_1, \dots, X_n)$ is a function $f_{\mathbf{X}}(x_1, \dots, x_n)$ satisfying (1) and (2) above, and for any event $\underline{A} \subset \mathbb{R}^n$,

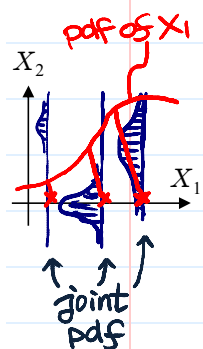
$$P(\mathbf{X} \in \underline{A}) = \int \cdots \int_{\underline{A}} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

■ Theorem. Suppose that $f_{\mathbf{X}}$ is the joint pdf of $\mathbf{X}=(X_1, \dots, X_n)$. Then, the joint pdf of $\underline{X}_1, \dots, \underline{X}_k, k < n$, is

$$\begin{aligned} f_{X_1, \dots, X_k}(x_1, \dots, x_k) \\ = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \dots, x_k, x_{k+1}, \dots, x_n) dx_{k+1} \cdots dx_n. \end{aligned}$$

In particular, the marginal pdf of \underline{X}_1 is

$$f_{X_1}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x, x_2, \dots, x_n) dx_2 \cdots dx_n.$$



- Theorem. If $F_{\mathbf{X}}$ and $f_{\mathbf{X}}$ are the joint cdf and joint pdf of \mathbf{X} , then

$$\frac{F_{\mathbf{X}}(x_1, \dots, x_n)}{= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, \dots, t_n) dt_1 \cdots dt_n, \text{ and}$$

$$f_{\mathbf{X}}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\mathbf{X}}(x_1, \dots, x_n).$$

at the continuity points of $f_{\mathbf{X}}$.

- Examples.

- Experiment. Two balls are drawn without replacement from a box with one ball labeled 1,
two balls labeled 2,
three balls labeled 3.

Let $X =$ label on the 1st ball drawn,

$Y =$ label on the 2nd ball drawn.

- The joint pmf and marginal pmfs of (X, Y) are

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

$p(x, y)$		X			$p_Y(y)$
		1	2	3	
Y	1	0	2/30	3/30	1/6
	2	2/30	2/30	6/30	2/6
	3	3/30	6/30	6/30	3/6
$p_X(x)$		1/6	2/6	3/6	

Q: The balls are drawn without replacement. Why do X (from 1st ball) and Y (from 2nd ball) have same marginal distributions?

- **Q:** $P(|X-Y|=1)=??$

$$P(|X-Y|=1) = P(X=1, Y=2) + P(X=2, Y=1) \\ + P(X=2, Y=3) + P(X=3, Y=2) = 8/15.$$

- **Q:** What are the joint pmf and marginal pmfs of (X, Y) if the balls are drawn with replacement (LNp. 4-6)?

$p(x, y)$		X			$p_Y(y)$
		1	2	3	
Y	1	1/36	2/36	3/36	1/6
	2	2/36	4/36	6/36	2/6
	3	3/36	6/36	9/36	3/6
$p_X(x)$		1/6	2/6	3/6	

➤ Multinomial Distribution

■ Recall. Partitions

- If $\underline{n} \geq 1$ and $\underline{n}_1, \dots, \underline{n}_m \geq 0$ are integers for which

$$\underline{n}_1 + \dots + \underline{n}_m = \underline{n},$$

then a set of \underline{n} elements may be partitioned into \underline{m} subsets of sizes $\underline{n}_1, \dots, \underline{n}_m$ in

$$\binom{\underline{n}}{\underline{n}_1, \dots, \underline{n}_m} = \frac{\underline{n}!}{\underline{n}_1! \times \dots \times \underline{n}_m!} \text{ ways.}$$

- Example (LNp.2-8) : MISSISSIPPI

$$\binom{11}{4,1,2,4} = \frac{11!}{4!1!2!4!}.$$

■ Example (Die Rolling).

- **Q**: If a balanced (6-sided) die is rolled 12 times,
 $P(\text{each face appears twice}) = ??$

- Sample space of rolling the die once (basic experiment):

$$\underline{\Omega}_0 = \{1, 2, 3, 4, 5, 6\}.$$

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- The sample space for the 12 trials is

$$\underline{\Omega} = \underline{\Omega}_0 \times \dots \times \underline{\Omega}_0 = \underline{\Omega}_0^{12}$$

An outcome $\underline{\omega} \in \underline{\Omega}$ is $\underline{\omega} = (\underline{i}_1, \underline{i}_2, \dots, \underline{i}_{12})$, where
 $\underline{1} \leq \underline{i}_1, \dots, \underline{i}_{12} \leq \underline{6}$.

- There are $\underline{6}^{12}$ possible outcomes in $\underline{\Omega}$, i.e., $\#\underline{\Omega} = \underline{6}^{12}$.

- Among all possible outcomes, there are $\binom{12}{2,2,2,2,2,2} = \frac{12!}{(2!)^6}$
of which each face appears twice.

- $P(\text{each face appears twice}) = \frac{12!}{(2!)^6} \left(\frac{1}{6}\right)^{12}.$

■ Generalization.

- Consider a basic experiment which can result in one of \underline{m} types of outcomes. Denote its sample space as

$$\underline{\Omega}_0 = \{1, 2, \dots, \underline{m}\}.$$

Let $\underline{p}_i = P(\text{outcome } i \text{ appears in a basic experiment})$,

then, (i) $\underline{p}_1, \dots, \underline{p}_m \geq 0$, and

(ii) $\underline{p}_1 + \dots + \underline{p}_m = 1.$

- Repeat the basic experiment n times. Then, the sample space for the n trials is

$$\underline{\Omega} = \Omega_0 \times \cdots \times \Omega_0 = \underline{\Omega}_0^n$$

Let $\underline{X}_i = \#$ of trials with outcome i , $i=1, \dots, m$,

Then, (i) $\underline{X}_1, \dots, \underline{X}_m: \underline{\Omega} \rightarrow \mathbb{R}$, and

$$(ii) \underline{X}_1 + \cdots + \underline{X}_m = n.$$

- The joint pmf of $\underline{X}_1, \dots, \underline{X}_m$ is

$$\begin{aligned} \underline{p_X}(x_1, \dots, x_m) &= P(\underline{X}_1 = x_1, \dots, \underline{X}_m = x_m) \\ &= \binom{n}{x_1, \dots, x_m} p_1^{x_1} \times \cdots \times p_m^{x_m}. \end{aligned}$$

for $\underline{x}_1, \dots, \underline{x}_m \geq 0$ and $\underline{x}_1 + \cdots + \underline{x}_m = n$.

Proof. The probability of any sequence with \underline{x}_i i 's is

$$p_1^{x_1} \times \cdots \times p_m^{x_m},$$

and there are

$$\binom{n}{x_1, \dots, x_m}$$

such sequences.

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- The distribution of a random vector $\underline{X}=(X_1, \dots, X_m)$ with the above joint pmf is called the multinomial distribution with parameters n, m , and $\underline{p}_1, \dots, \underline{p}_m$, denoted by Multinomial($n, m, \underline{p}_1, \dots, \underline{p}_m$).

- ◆ The multinomial distribution is called after the Multinomial Theorem:

$$\begin{aligned} & (a_1 + \cdots + a_m)^n \\ &= \sum_{\substack{x_i \in \{0, \dots, n\}; i=1, \dots, m \\ x_1 + \cdots + x_m = n}} \binom{n}{x_1, \dots, x_m} a_1^{x_1} \times \cdots \times a_m^{x_m}. \end{aligned}$$

- ◆ It is a generalization of the binomial distribution from 2 types of outcomes to m types of outcomes.

- Some Properties.

- ◆ Because $\underline{X}_i = n - (\underline{X}_1 + \cdots + \underline{X}_{i-1} + \underline{X}_{i+1} + \cdots + \underline{X}_m)$, and

$$\underline{p}_i = 1 - (\underline{p}_1 + \cdots + \underline{p}_{i-1} + \underline{p}_{i+1} + \cdots + \underline{p}_m),$$

WLOG, we can write

$$(\underline{X}_1, \dots, \underline{X}_{m-1}, \underline{X}_m) \rightarrow (\underline{X}_1, \dots, \underline{X}_{m-1}, n - (\underline{X}_1 + \cdots + \underline{X}_{m-1}))$$

◆ Marginal Distribution. Suppose that

$$(X_1, \dots, X_m) \sim \text{Multinomial}(n, \underline{m}, \underline{p}_1, \dots, \underline{p}_k, \underline{p}_{k+1}, \dots, \underline{p}_m).$$

For $1 \leq k < m$, the distribution of

$$(X_1, \dots, X_k, X_{k+1} + \dots + X_m)$$

is Multinomial $(n, k+1, \underline{p}_1, \dots, \underline{p}_k, \underline{p}_{k+1} + \dots + \underline{p}_m)$.

In particular, $X_i \sim \text{Binomial}(n, \underline{p}_i)$

◆ Mean and Variance.

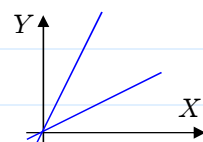
$$E(X_i) = np_i \text{ and } Var(X_i) = np_i(1-p_i)$$

for $i = 1, \dots, m$.

➤ Example.

■ Suppose that the joint pdf of 2 continuous r.v.'s (X, Y) is

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda(x+y)}, & x \geq 0, y \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$



Q: $P(Y \geq 2X \text{ or } X \geq 2Y) = ??$

■ The event $\{Y \geq 2X\} \cup \{X \geq 2Y\}$ is

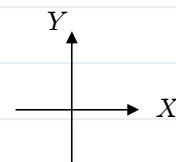
NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

■ So, $P(Y \geq 2X \text{ or } X \geq 2Y) = P(Y \geq 2X) + P(X \geq 2Y) = 2/3$ because^{p. 7-18}

$$\begin{aligned} P(Y \geq 2X) &= \int_0^\infty \left[\int_{2x}^\infty \lambda^2 e^{-\lambda(x+y)} dy \right] dx \\ &= \int_0^\infty -\lambda e^{-\lambda(x+y)} \Big|_{y=2x}^\infty dx = \int_0^\infty \lambda e^{-3\lambda x} dx \\ &= (-1/3) e^{-3\lambda x} \Big|_{x=0}^\infty = 1/3. \end{aligned}$$

and similarly, we can get $P(X \geq 2Y) = 1/3$ (**exercise**).

➤ Example. Consider two continuous r.v.'s X and Y .



■ Uniform Distribution over a region D . If $\underline{D} \subset \mathbb{R}^2$ and $0 < \underline{\alpha} = \text{Area}(D) < \infty$, then

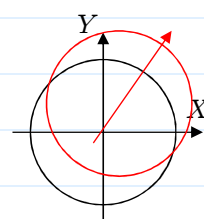
$$f(x, y) = \underline{c} \cdot \underline{1}_D(x, y)$$

is a joint pdf when $\underline{c} = 1/\underline{\alpha}$, called the uniform pdf over \underline{D} .

■ Let $\underline{D} = \{(x, y): x^2 + y^2 \leq 1\}$, then $\underline{\alpha} = \text{Area}(D) = \underline{\pi}$ and

$$f(x, y) = \frac{1}{\pi} \underline{1}_D(x, y)$$

is a joint pdf.

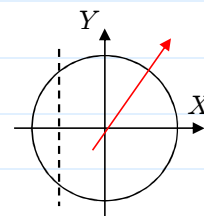


- Marginal distribution. The marginal pdf of X is

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi} \sqrt{1-x^2}$$

for $-1 \leq x \leq 1$, and $f_X(x)=0$, otherwise.

(exercise: Find the marginal distribution of Y .)

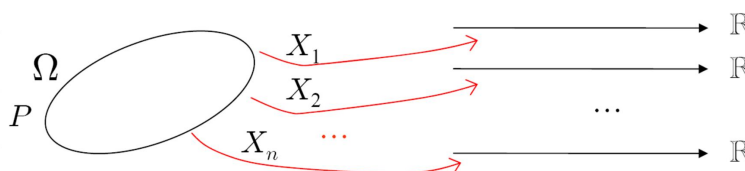


❖ Reading: textbook, Sec 6.1

Independent Random Variables

• Recall.

- If joint distribution is given, marginal distributions are known.
- The converse statement does not hold in general.
- However, when random variables are independent,
marginal distributions + independence \Rightarrow joint distribution.



NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- Definition. The n jointly distributed r.v.'s X_1, \dots, X_n are called (mutually) independent if and only if for any (measurable) sets $A_i \subset \mathbb{R}$, $i=1, \dots, n$, the events

$$\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$$

are (mutually) independent. That is,

$$\begin{aligned} &P(X_{i_1} \in A_{i_1}, X_{i_2} \in A_{i_2}, \dots, X_{i_k} \in A_{i_k}) \\ &= P(X_{i_1} \in A_{i_1}) \times P(X_{i_2} \in A_{i_2}) \times \dots \times P(X_{i_k} \in A_{i_k}), \end{aligned}$$

for any $1 \leq i_1 < i_2 < \dots < i_k \leq n$; $k=2, \dots, n$.

- If X_1, \dots, X_n are independent, for $1 \leq k < n$,

$$\begin{aligned} &P(X_{k+1} \in A_{k+1}, \dots, X_n \in A_n | X_1 \in A_1, \dots, X_k \in A_k) \\ &= P(X_{k+1} \in A_{k+1}, \dots, X_n \in A_n) \end{aligned}$$

provided that $P(X_1 \in A_1, \dots, X_k \in A_k) > 0$.

- In other words, the values of X_1, \dots, X_k do not carry any information about the distribution of X_{k+1}, \dots, X_n .

- Theorem (Factorization Theorem). The random variables $\mathbf{X}=(X_1, \dots, X_n)$ are independent if and only if one of the following conditions holds.

- (1) $F_{\mathbf{X}}(x_1, \dots, x_n) = F_{X_1}(x_1) \times \dots \times F_{X_n}(x_n)$, where $F_{\mathbf{X}}$ is the joint cdf of \mathbf{X} and F_{X_i} is the marginal cdf of X_i for $i=1, \dots, n$.
- (2) Suppose that X_1, \dots, X_n are discrete random variables.
 $p_{\mathbf{X}}(x_1, \dots, x_n) = p_{X_1}(x_1) \times \dots \times p_{X_n}(x_n)$, where $p_{\mathbf{X}}$ is the joint pmf of \mathbf{X} and p_{X_i} is the marginal pmf of X_i for $i=1, \dots, n$.
- (3) Suppose that X_1, \dots, X_n are continuous random variables.
 $f_{\mathbf{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n)$, where $f_{\mathbf{X}}$ is the joint pdf of \mathbf{X} and f_{X_i} is the marginal pdf of X_i for $i=1, \dots, n$.

Proof.

$$\begin{aligned}
 \text{independent} \Rightarrow (1). \quad F_{\mathbf{X}}(x_1, \dots, x_n) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) \\
 &= P(X_1 \in (-\infty, x_1], \dots, X_n \in (-\infty, x_n]) \\
 &= P(X_1 \in (-\infty, x_1]) \times \dots \times P(X_n \in (-\infty, x_n]) \\
 &= F_{X_1}(x_1) \times \dots \times F_{X_n}(x_n)
 \end{aligned}$$

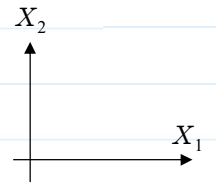
independent \Leftarrow (1). Out of the scope of this course so skip.

NTHU MATH 2810, 2024, Lecture Notes
 made by S.-W. Cheng (NTHU, Taiwan)

$$\begin{aligned}
 \text{independent} \Rightarrow (2). \quad p_{\mathbf{X}}(x_1, \dots, x_n) &= P(X_1 = x_1, \dots, X_n = x_n) \\
 &= P(X_1 \in \{x_1\}, \dots, X_n \in \{x_n\}) \\
 &= P(X_1 \in \{x_1\}) \times \dots \times P(X_n \in \{x_n\}) \\
 &= p_{X_1}(x_1) \times \dots \times p_{X_n}(x_n)
 \end{aligned}$$

(2) \Rightarrow (1).

$$\begin{aligned}
 F_{\mathbf{X}}(x_1, \dots, x_n) &= \sum_{\substack{(t_1, \dots, t_n) \in \mathcal{X} \\ t_1 \leq x_1, \dots, t_n \leq x_n}} p_{\mathbf{X}}(t_1, \dots, t_n) \\
 &= \sum_{\substack{(t_1, \dots, t_n) \in \mathcal{X} \\ t_1 \leq x_1}} \dots \sum_{\substack{(t_1, \dots, t_n) \in \mathcal{X} \\ t_n \leq x_n}} \frac{p_{X_1}(t_1) \times \dots \times p_{X_n}(t_n)}{1} \\
 &= \sum_{\substack{(t_1, \dots, t_n) \in \mathcal{X} \\ t_1 \leq x_1}} p_{X_1}(t_1) \times \dots \times \sum_{\substack{(t_1, \dots, t_n) \in \mathcal{X} \\ t_n \leq x_n}} p_{X_n}(t_n) = F_{X_1}(x_1) \times \dots \times F_{X_n}(x_n)
 \end{aligned}$$



(3) \Rightarrow (1).

$$\begin{aligned}
 F_{\mathbf{X}}(x_1, \dots, x_n) &= \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f_{\mathbf{X}}(t_1, \dots, t_n) dt_1 \dots dt_n \\
 &= \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} \frac{f_{X_1}(t_1) \times \dots \times f_{X_n}(t_n)}{1} dt_1 \dots dt_n \\
 &= \int_{-\infty}^{x_1} f_{X_1}(t_1) dt_1 \times \dots \times \int_{-\infty}^{x_n} f_{X_n}(t_n) dt_n = F_{X_1}(x_1) \times \dots \times F_{X_n}(x_n)
 \end{aligned}$$

$$(3) \Leftarrow (1).$$

$$\begin{aligned} f_{\mathbf{X}}(x_1, \dots, x_n) &= \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\mathbf{X}}(x_1, \dots, x_n). \\ &= \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{X_1}(x_1) \times \cdots \times F_{X_n}(x_n) \\ &= \frac{\partial}{\partial x_1} F_{X_1}(x_1) \times \cdots \times \frac{\partial}{\partial x_n} F_{X_n}(x_n) = \underline{f_{X_1}(x_1) \times \cdots \times f_{X_n}(x_n)} \end{aligned}$$

➤ Remark. It follows from the Multiplication Law (LNp.4-11) that

$$\begin{aligned} F_{\mathbf{X}}(x_1, \dots, x_n) &= P(X_1 \leq x_1, \dots, X_n \leq x_n) \\ &= P(X_1 \leq x_1) \quad \quad \quad (= F_{X_1}(x_1)) \\ &\quad \times P(X_2 \leq x_2 | X_1 \leq x_1) \quad \quad \quad \left(\stackrel{?}{=} P(X_2 \leq x_2) = F_{X_2}(x_2) \right) \\ &\quad \times P(X_3 \leq x_3 | X_1 \leq x_1, X_2 \leq x_2) \quad \quad \quad \left(\stackrel{?}{=} P(X_3 \leq x_3) = F_{X_3}(x_3) \right) \\ &\quad \times \cdots \\ &\quad \times P(X_n \leq x_n | X_1 \leq x_1, \dots, X_{n-1} \leq x_{n-1}) \left(\stackrel{?}{=} P(X_n \leq x_n) = F_{X_n}(x_n) \right) \end{aligned}$$

The independence can be established sequentially.

➤ Example. If $A_1, \dots, A_n \subset \Omega$ are independent events, then $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$ are independent random variables. For example,

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

$$\begin{aligned} &P(\mathbf{1}_{A_1} = 1, \mathbf{1}_{A_2} = 0, \mathbf{1}_{A_3} = 1) \\ &= P(A_1 \cap A_2^c \cap A_3) = P(A_1)P(A_2^c)P(A_3) \\ &= \underline{P(\mathbf{1}_{A_1} = 1)P(\mathbf{1}_{A_2} = 0)P(\mathbf{1}_{A_3} = 1)}. \end{aligned}$$

➤ Theorem. If $\mathbf{X} = (X_1, \dots, X_n)$ are independent and

$$\underline{Y_i} = \underline{g_i(X_i)}, \quad i=1, \dots, n,$$

then

$\underline{Y_1}, \dots, \underline{Y_n}$ are independent.

Proof.

Let $A_i(y) = \{x : g_i(x) \leq y\}, i=1, \dots, n$, then

$$\begin{aligned} F_{\mathbf{Y}}(y_1, \dots, y_n) &= P(Y_1 \leq y_1, \dots, Y_n \leq y_n) \\ &= P(X_1 \in A_1(y_1), \dots, X_n \in A_n(y_n)) \\ &= P(X_1 \in A_1(y_1)) \times \cdots \times P(X_n \in A_n(y_n)) \\ &= P(Y_1 \leq y_1) \times \cdots \times P(Y_n \leq y_n) \\ &= \underline{F_{Y_1}(y_1) \times \cdots \times F_{Y_n}(y_n)}. \end{aligned}$$

generalization

$$1 = i_0 < i_1 < \cdots < i_k = n$$

$$Y_1 = g_1(X_1, \dots, X_{i_1}),$$

$$Y_2 = g_2(X_{i_1+1}, \dots, X_{i_2}),$$

...

$$Y_k = g_k(X_{i_{k-1}+1}, \dots, X_{i_k}).$$

• Theorem. $\underline{\mathbf{X}}=(X_1, \dots, X_n)$ are independent if and only if there exist univariate functions $\underline{g_i(x)}$, $i=1, \dots, n$, such that

(a) when $\underline{X_1, \dots, X_n}$ are discrete r.v.'s with joint pmf $\underline{p_{\mathbf{X}}}$,
 $\underline{p_{\mathbf{X}}(x_1, \dots, x_n) \propto \underline{g_1(x_1)} \times \dots \times \underline{g_n(x_n)}, -\infty < x_i < \infty, i=1, \dots, n.}$

(b) when $\underline{X_1, \dots, X_n}$ are continuous r.v.'s with joint pdf $\underline{f_{\mathbf{X}}}$,
 $\underline{f_{\mathbf{X}}(x_1, \dots, x_n) \propto \underline{g_1(x_1)} \times \dots \times \underline{g_n(x_n)}, -\infty < x_i < \infty, i=1, \dots, n.}$

Sketch of proof for (b).

$$\begin{aligned} \underline{f_{X_1}(x_1)} &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \underline{f_{\mathbf{X}}(x_1, x_2, \dots, x_n)} \, dx_2 \dots dx_n \\ &\propto \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \underline{g_1(x_1) g_2(x_2) \dots g_n(x_n)} \, dx_2 \dots dx_n \propto \underline{g_1(x_1)}. \end{aligned}$$

$$\begin{aligned} \text{Similarly, } \underline{f_{X_2}(x_2)} &\propto \underline{g_2(x_2)}, \dots, \underline{f_{X_n}(x_n)} \propto \underline{g_n(x_n)} \\ \Rightarrow \underline{f_{X_1}(x_1) \dots f_{X_n}(x_n)} &\propto \underline{g_1(x_1) \dots g_n(x_n)} \\ \Rightarrow \underline{f_{\mathbf{X}}(x_1, \dots, x_n)} &\propto \underline{f_{X_1}(x_1) \dots f_{X_n}(x_n)} \\ \Rightarrow \underline{f_{\mathbf{X}}(x_1, \dots, x_n)} &= \underline{c} \cdot \underline{f_{X_1}(x_1) \dots f_{X_n}(x_n)} \\ &\text{for some constant } \underline{c}. \end{aligned}$$

NTHU MATH 2810, 2024, Lecture Notes
 made by S.-W. Cheng (NTHU, Taiwan)

$$\begin{aligned} \text{Because } \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \underline{f_{\mathbf{X}}(x_1, x_2, \dots, x_n)} \, \underline{dx_1 \dots dx_n} &= 1, \text{ and} \\ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \underline{f_{X_1}(x_1) \dots f_{X_n}(x_n)} \, \underline{dx_1 \dots dx_n} &= 1, \Rightarrow \underline{c} = 1. \end{aligned}$$

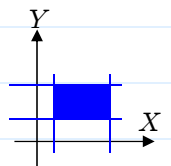
➤ Example.

■ If the joint pdf of $(\underline{X}, \underline{Y})$ is

$$\underline{f(x, y) \propto e^{-2x} e^{-3y}, \quad 0 < x < \infty, 0 < y < \infty,}$$

and $\underline{f(x, y)=0}$, otherwise, i.e.,

$$\underline{f(x, y) \propto e^{-2x} e^{-3y} \mathbf{1}_{(0, \infty)}(x) \mathbf{1}_{(0, \infty)}(y),}$$



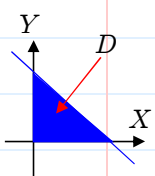
then \underline{X} and \underline{Y} are independent. Note that the region in which the joint pdf is nonzero can be expressed in the form $\{(x, y): \underline{x \in A}, \underline{y \in B}\}$.

■ Suppose that the joint pdf of $(\underline{X}, \underline{Y})$ is

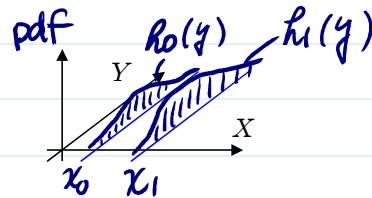
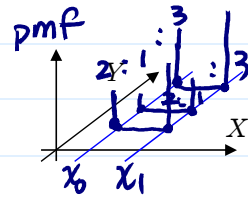
$$\underline{f(x, y) \propto \underline{xy}, \quad 0 < x < 1, 0 < y < 1, 0 < x + y < 1,}$$

and $\underline{f(x, y)=0}$, otherwise, i.e., $\underline{f(x, y) \propto xy \cdot \mathbf{1}_D(x, y)},$

X and Y are not independent.



➤ **Q:** For independent \underline{X} and \underline{Y} , how should their joint pdf/pmf look like?



$$\frac{h_1(y)}{h_0(y)} = \underline{\text{a constant}}$$

❖ **Reading:** textbook, Sec 6.2

Transformation

- **Q:** Given the joint distribution of $\underline{X}=(X_1, \dots, X_n)$, how to find the distribution of $\underline{Y}=(Y_1, \dots, Y_k)$, where

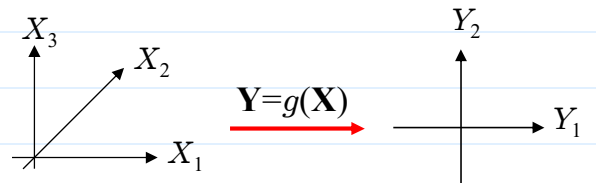
$$\underline{Y}_1 = \underline{g}_1(\underline{X}_1, \dots, \underline{X}_n) : \underline{\mathbb{R}}^n \rightarrow \underline{\mathbb{R}},$$

...

$$\underline{Y}_k = \underline{g}_k(\underline{X}_1, \dots, \underline{X}_n) : \underline{\mathbb{R}}^n \rightarrow \underline{\mathbb{R}},$$

denoted by

$$\underline{Y} = \underline{g}(\underline{X}), \underline{g} : \underline{\mathbb{R}}^n \rightarrow \underline{\mathbb{R}}^k.$$



NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

➤ The following methods are useful:

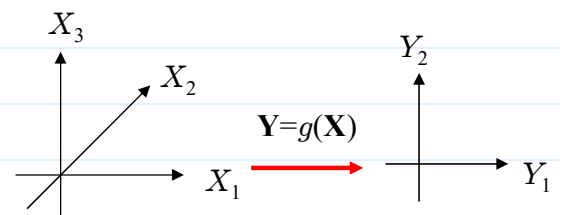
1. Method of Events (\rightarrow pmf)
2. Method of Cumulative Distribution Function
3. Method of Probability Density Function
4. Method of Moment Generating Function (chapter 7)

➤ Method of Events

- Theorem. The distribution of \underline{Y} is determined by the distribution of \underline{X} as follows: for any event $\underline{B} \subset \underline{\mathbb{R}}^k$,

$$P_{\underline{Y}}(\underline{Y} \in \underline{B}) = P_{\underline{X}}(\underline{X} \in \underline{A}),$$

where $\underline{A} = \underline{g}^{-1}(\underline{B}) \subset \underline{\mathbb{R}}^n$.



- Example. Let \underline{X} be a discrete random vector taking values

$$\underline{x}_i = (x_{1i}, x_{2i}, \dots, x_{ni}), i=1, 2, \dots,$$

(i.e., $\mathcal{X} = \{\underline{x}_1, \underline{x}_2, \underline{x}_3, \dots\}$) with joint pmf $p_{\underline{X}}$.

Then, $\underline{Y} = \underline{g}(\underline{X})$ is also a discrete random vector.

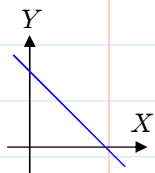
Suppose that \underline{Y} takes values on $\underline{y}_j, j=1, 2, \dots$. To determine the joint pmf of \underline{Y} , by taking $\underline{B}=\{\underline{y}_j\}$, we have

$$\underline{A} = \{\underline{x}_i \in \mathcal{X} : \underline{g}(\underline{x}_i) = \underline{y}_j\}$$

and hence, the joint pmf of \underline{Y} is

$$p_{\underline{Y}}(\underline{y}_j) = P_{\underline{Y}}(\{\underline{y}_j\}) = P_{\underline{X}}(\underline{A}) = \sum_{\underline{x}_i \in \underline{A}} p_{\underline{X}}(\underline{x}_i).$$

- Example. Let \underline{X} and \underline{Y} be random variables with the joint pmf $p(x, y)$. Find the distribution of $\underline{Z}=\underline{X}+\underline{Y}$.



$$\square \{Z=\underline{z}\} = \{(X, Y) \in \{(x, y): x+y=\underline{z}\}\}$$

$$p_Z(z) = P_Z(\{z\}) = P(X + Y = z) = \sum_{x \in \mathcal{X}_X} p(x, z - x).$$

- When \underline{X} and \underline{Y} are independent,

$$p(x, y) = p_X(x)p_Y(y),$$

So,

$$p_Z(z) = \sum_{x \in \mathcal{X}_X} p_X(x)p_Y(z - x).$$

which is referred to as the convolution of $p_{\underline{X}}$ and $p_{\underline{Y}}$.

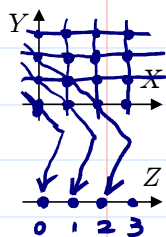
- (Exercise) $\underline{Z}=\underline{X}-\underline{Y}$

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- Theorem. If \underline{X} and \underline{Y} are independent, and

$$\underline{X} \sim \text{Poisson}(\lambda_1), \quad \underline{Y} \sim \text{Poisson}(\lambda_2),$$

then $\underline{Z} = \underline{X} + \underline{Y} \sim \text{Poisson}(\lambda_1 + \lambda_2)$.



Proof. For $z=0, 1, 2, \dots$, the pmf $p_Z(z)$ of \underline{Z} is

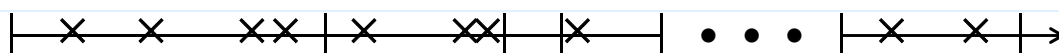
$$\begin{aligned} p_Z(z) &= \sum_{x=0}^z p_X(x)p_Y(z-x) = \sum_{x=0}^z \frac{e^{-\lambda_1} \lambda_1^x}{x!} \frac{e^{-\lambda_2} \lambda_2^{z-x}}{(z-x)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{z!} \left(\sum_{x=0}^z \frac{z!}{x!(z-x)!} \lambda_1^x \lambda_2^{z-x} \right) = \frac{e^{-(\lambda_1+\lambda_2)}}{z!} (\lambda_1 + \lambda_2)^z. \end{aligned}$$

- Corollary. If $\underline{X}_1, \dots, \underline{X}_n$ are independent, and

$$\underline{X}_i \sim \text{Poisson}(\lambda_i), \quad i=1, \dots, n,$$

then $\underline{X}_1 + \dots + \underline{X}_n \sim \text{Poisson}(\lambda_1 + \dots + \lambda_n)$.

Proof. By induction (exercise).



➤ Method of cumulative distribution function

1. In the $(\underline{X}_1, \dots, \underline{X}_n)$ space, find the region that corresponds to

$$\{Y_1 \leq y_1, \dots, Y_k \leq y_k\}.$$

2. Find $F_Y(y_1, \dots, y_k) = P(Y_1 \leq y_1, \dots, Y_k \leq y_k)$ by summing the joint pmf or integrating the joint pdf of $\underline{X}_1, \dots, \underline{X}_n$ over the region identified in 1.

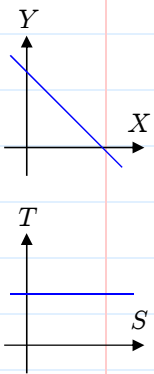
3. (for continuous case) Find the joint pdf of \underline{Y} by differentiating $F_Y(y_1, \dots, y_k)$, i.e.,

$$f_Y(y_1, \dots, y_k) = \frac{\partial^k}{\partial y_1 \dots \partial y_k} F_Y(y_1, \dots, y_k).$$

■ Example. \underline{X} and \underline{Y} are random variables with joint pdf $f(x, y)$. Find the distribution of $\underline{Z} = \underline{X} + \underline{Y}$.

□ $\{Z \leq z\} = \{(X, Y) \in \{(x, y): x+y \leq z\}\}$. So,

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X + Y \leq z) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y) dy dx \\ &= \int_{-\infty}^z \int_{-\infty}^{\infty} f(s, t-s) ds dt \quad \left(\text{set } \begin{cases} x = s \\ y = t-s \end{cases} \right) \end{aligned}$$



NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

$$\text{and } \underline{f_Z(z)} = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f(x, z-x) dx$$

□ When \underline{X} and \underline{Y} are independent,

$$\underline{f(x, y)} = \underline{f_X(x)} \underline{f_Y(y)}.$$

$$\begin{aligned} \text{So, } F_Z(z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y) dy dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x} f_Y(y) dy \right] f_X(x) dx \\ &= \int_{-\infty}^{\infty} F_Y(z-x) f_X(x) dx \end{aligned}$$

which is referred to as the convolution of $\underline{F_X}$ and $\underline{F_Y}$, and

$$\underline{f_Z(z)} = \int_{-\infty}^{\infty} \underline{f_X(x)} \underline{f_Y(z-x)} dx$$

which is referred to as the convolution of $\underline{f_X}$ and $\underline{f_Y}$.

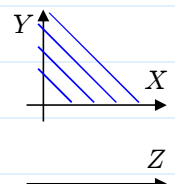
□ (exercise) $\underline{Z} = \underline{X} - \underline{Y}$.

■ Theorem. If \underline{X} and \underline{Y} are independent, and

$$\underline{X} \sim \text{Gamma}(\underline{\alpha}_1, \underline{\lambda}), \quad \underline{Y} \sim \text{Gamma}(\underline{\alpha}_2, \underline{\lambda}),$$

then

$$\underline{Z} = \underline{X} + \underline{Y} \sim \text{Gamma}(\underline{\alpha}_1 + \underline{\alpha}_2, \underline{\lambda}).$$



Proof. For $z \geq 0$,

$$\begin{aligned} f_Z(z) &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^z x^{\alpha_1-1} (z-x)^{\alpha_2-1} e^{-\lambda z} dx \\ &= \frac{\lambda^{\alpha_1+\alpha_2} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^1 z^{(\alpha_1-1)+(\alpha_2-1)+1} y^{\alpha_1-1} (1-y)^{\alpha_2-1} dy \\ &= \frac{\lambda^{\alpha_1+\alpha_2} z^{(\alpha_1+\alpha_2)-1} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \times \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}. \end{aligned}$$

and $f_Z(z) = 0$, for $z < 0$.

□ Corollary. If X_1, \dots, X_n are independent, and

$$X_i \sim \text{Gamma}(\alpha_i, \lambda), i=1, \dots, n,$$

then $X_1 + \dots + X_n \sim \text{Gamma}(\alpha_1 + \dots + \alpha_n, \lambda)$.

Proof. By induction (exercise).

□ Corollary. If X_1, \dots, X_n are independent, and

$$X_i \sim \text{Exponential}(\lambda), i=1, \dots, n,$$

then $X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda)$.

Proof. (exercise).

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

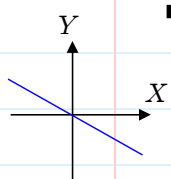
■ Theorem. If X_1, \dots, X_n are independent, and

$$X_i \sim \text{Normal}(\mu_i, \sigma_i^2), i=1, \dots, n,$$

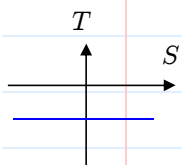
then $X_1 + \dots + X_n \sim \text{Normal}(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2)$.

Proof. (exercise).

■ Example. X and Y are random variables with joint pdf $f(x, y)$. Find the distribution of $Z=Y/X$.



$$\begin{aligned} \square \text{ Let } Q_z &= \{(x, y) : y/x \leq z\} \\ &= \{(x, y) : x < 0, y \geq zx\} \\ &\quad \cup \{(x, y) : x > 0, y \leq zx\} \end{aligned}$$



$$\begin{aligned} \text{then, } F_Z(z) &= \int \int_{Q_z} f(x, y) dx dy \\ &= \int_{-\infty}^0 \int_{zx}^{\infty} + \int_0^{\infty} \int_{-\infty}^{zx} f(x, y) dy dx \quad \left(\text{set } \begin{cases} x = s \\ y = st \end{cases} \right) \\ &= \int_{-\infty}^0 \int_{-\infty}^z + \int_0^{\infty} \int_{-\infty}^z f(s, st) |s| dt ds \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^z |s| f(s, st) dt ds \\ &= \int_{-\infty}^z \int_{-\infty}^{\infty} |s| f(s, st) ds dt \end{aligned}$$

and, $\underline{f_Z(z)} = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} |x| \underline{f(x, zx)} dx$

□ When X and Y are independent,

$$\underline{f(x, y)} = \underline{f_X(x)} \underline{f_Y(y)}.$$

So, $F_Z(z) = \int_{-\infty}^z \int_{-\infty}^{\infty} |s| \underline{f_X(s)} \underline{f_Y(st)} ds dt$

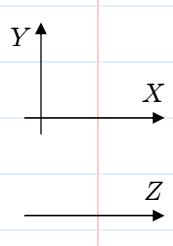
and, $\underline{f_Z(z)} = \int_{-\infty}^{\infty} |x| \underline{f_X(x)} \underline{f_Y(zx)} dx$

□ (exercise) Z=XY

□ If X and Y are independent,

$$\underline{X} \sim \underline{\text{exponential}(\lambda_1)}, \text{ and } \underline{Y} \sim \underline{\text{exponential}(\lambda_2)},$$

Let Z=Y/X. The pdf of Z is



$$\begin{aligned} \underline{f_Z(z)} &= \int_0^{\infty} x (\lambda_1 e^{-\lambda_1 x}) [\lambda_2 e^{-\lambda_2 (xz)}] dx \\ &= \frac{\lambda_1 \lambda_2 \Gamma(2)}{(\lambda_1 + \lambda_2 z)^2} \int_0^{\infty} \frac{(\lambda_1 + \lambda_2 z)^2}{\Gamma(2)} x^{2-1} e^{-(\lambda_1 + \lambda_2 z)x} dx \\ &= \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2 z)^2} \end{aligned}$$

for $z \geq 0$, and 0 for $z < 0$.

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

And, the cdf of Z is

$$\begin{aligned} F_Z(z) &= \int_0^z \underline{f_Z(t)} dt = \int_0^z \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2 t)^2} dt \\ &= -\frac{\lambda_1 \lambda_2}{\lambda_2} (\lambda_1 + \lambda_2 t)^{-1} \Big|_0^z = 1 - \frac{\lambda_1}{\lambda_1 + \lambda_2 z} \end{aligned}$$

for $z \geq 0$, and 0 for $z < 0$.

➤ Method of probability density function

■ Theorem. Let $\mathbf{X}=(X_1, \dots, X_n)$ be continuous random variables with the joint pdf $\underline{f_X(x_1, \dots, x_n)}$. Let

$$\underline{\mathbf{Y}}=(Y_1, \dots, Y_n)=\underline{g(\mathbf{X})},$$

where g is 1-to-1, so that its inverse exists and is denoted by

$$\underline{\mathbf{x}}=\underline{g^{-1}(\mathbf{y})}=\underline{\mathbf{w}(\mathbf{y})}=(\underline{w_1(\mathbf{y})}, \underline{w_2(\mathbf{y})}, \dots, \underline{w_n(\mathbf{y})}).$$

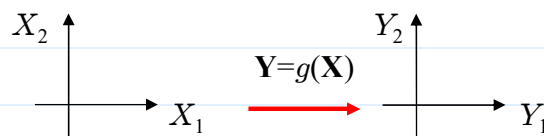
Assume w have continuous partial derivatives. Let

$$\underline{J} = \begin{vmatrix} \frac{\partial w_1(\mathbf{y})}{\partial y_1} & \frac{\partial w_1(\mathbf{y})}{\partial y_2} & \dots & \frac{\partial w_1(\mathbf{y})}{\partial y_n} \\ \frac{\partial w_2(\mathbf{y})}{\partial y_1} & \frac{\partial w_2(\mathbf{y})}{\partial y_2} & \dots & \frac{\partial w_2(\mathbf{y})}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial w_n(\mathbf{y})}{\partial y_1} & \frac{\partial w_n(\mathbf{y})}{\partial y_2} & \dots & \frac{\partial w_n(\mathbf{y})}{\partial y_n} \end{vmatrix}_{n \times n}$$

Then $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) \times |J|$,

for \mathbf{y} s.t. $\mathbf{y}=g(\mathbf{x})$ for some \mathbf{x} , and $f_{\mathbf{Y}}(\mathbf{y})=0$, otherwise.

(Q: What is the role of $|J|$?)



$$\begin{aligned} \text{Proof. } F_{\mathbf{Y}}(y_1, \dots, y_n) &= \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} f_{\mathbf{Y}}(t_1, \dots, t_n) dt_n \cdots dt_1 \\ &= \int \cdots \int_{\substack{(x_1, \dots, x_n): \\ g_1(x_1, \dots, x_n) \leq y_1 \\ \vdots \\ g_n(x_1, \dots, x_n) \leq y_n}} f_{\mathbf{X}}(x_1, \dots, x_n) dx_n \cdots dx_1. \end{aligned}$$

It then follows from an exercise in advanced calculus that

$$\begin{aligned} f_{\mathbf{Y}}(y_1, \dots, y_n) &= \frac{\partial^n}{\partial y_1 \cdots \partial y_n} F_{\mathbf{Y}}(y_1, \dots, y_n) \\ &= f_{\mathbf{X}}(w_1(\mathbf{y}), \dots, w_n(\mathbf{y})) \times |J|. \end{aligned}$$

□ Remark. When the dimensionality of \mathbf{Y} (denoted by k) is less than n , we can choose another $n-k$ transformations \mathbf{Z} such that

$$(\mathbf{Y}, \mathbf{Z}) = g(\mathbf{X})$$

satisfy the assumptions in above theorem.

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

By integrating out the last $n-k$ arguments in the joint pdf of (\mathbf{Y}, \mathbf{Z}) , the joint pdf of \mathbf{Y} can be obtained.

■ Example. X_1 and X_2 are random variables with joint pdf $f_{\mathbf{X}}(x_1, x_2)$. Find the distribution of $Y_1 = X_1 / (X_1 + X_2)$.

□ Let $Y_2 = X_1 + X_2$, then

$$\begin{aligned} x_1 &= y_1 y_2 \equiv w_1(y_1, y_2) \\ x_2 &= y_2 - y_1 y_2 \equiv w_2(y_1, y_2). \end{aligned}$$

$$\text{Since } \frac{\partial w_1}{\partial y_1} = y_2, \quad \frac{\partial w_1}{\partial y_2} = y_1, \quad \frac{\partial w_2}{\partial y_1} = -y_2, \quad \frac{\partial w_2}{\partial y_2} = 1 - y_1,$$

$$J = \begin{vmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{vmatrix} = y_2 - y_1 y_2 + y_1 y_2 = y_2, \text{ and } |J| = |y_2|.$$

$$\text{Therefore, } f_{\mathbf{Y}}(y_1, y_2) = f_{\mathbf{X}}(y_1 y_2, y_2 - y_1 y_2) |y_2|,$$

$$\text{and, } f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{\mathbf{Y}}(y_1, y_2) dy_2$$

$$= \int_{-\infty}^{\infty} f_{\mathbf{X}}(y_1 y_2, y_2 - y_1 y_2) |y_2| dy_2.$$

$$\begin{aligned} & (= \int_{-\infty}^{\infty} f_{X_1}(y_1 y_2) f_{X_2}(y_2 - y_1 y_2) |y_2| dy_2 \\ & \text{when } X_1 \text{ and } X_2 \text{ are independent}) \end{aligned}$$

- Theorem. If X_1 and X_2 are independent, and

$$\begin{array}{c} X_2 \\ \uparrow \\ \text{---} X_1 \end{array} \quad \begin{array}{c} X_1 \sim \text{Gamma}(\alpha_1, \lambda), \\ X_2 \sim \text{Gamma}(\alpha_2, \lambda), \end{array}$$

then $Y_1 = X_1 / (X_1 + X_2) \sim \text{Beta}(\alpha_1, \alpha_2)$.

Proof. For $x_1, x_2 \geq 0$, the joint pdf of \mathbf{X} is

$$\begin{array}{c} Y_2 \\ \uparrow \\ \text{---} Y_1 \end{array} \quad \begin{aligned} f_{\mathbf{X}}(x_1, x_2) &= \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} x_1^{\alpha_1-1} e^{-\lambda x_1} \times \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)} x_2^{\alpha_2-1} e^{-\lambda x_2} \\ &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x_1^{\alpha_1-1} x_2^{\alpha_2-1} e^{-\lambda(x_1+x_2)}. \end{aligned}$$

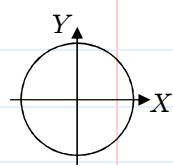
So, for $0 \leq y_1 \leq 1$,

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{-\infty}^{\infty} f_{X_1}(y_1 y_2) f_{X_2}(y_2 - y_1 y_2) |y_2| dy_2 \\ &= \int_0^{\infty} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} (y_1 y_2)^{\alpha_1-1} (y_2 - y_1 y_2)^{\alpha_2-1} e^{-\lambda y_2} \cdot y_2 dy_2 \\ &= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1} (1-y_1)^{\alpha_2-1} \\ &\quad \times \int_0^{\infty} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} y_2^{(\alpha_1+\alpha_2)-1} e^{-\lambda y_2} dy_2 \\ &= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1} (1-y_1)^{\alpha_2-1} \end{aligned}$$

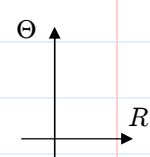
and $f_{Y_1}(y_1) = 0$, otherwise.

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- Example. Suppose that X and Y have a uniform distribution over the region $D = \{(x, y): x^2 + y^2 \leq 1\}$, i.e., their joint pdf is

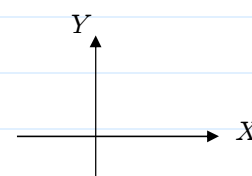


$$f_{X,Y}(x, y) = \frac{1}{\pi} \mathbf{1}_D(x, y).$$



Find the joint distribution of (R, Θ) and examine whether R and Θ are independent, where (R, Θ) is the polar coordinate representation of (X, Y) , i.e.,

$$\begin{aligned} X &= R \cos(\Theta) \equiv w_1(R, \Theta), \\ Y &= R \sin(\Theta) \equiv w_2(R, \Theta). \end{aligned}$$



$$\begin{aligned} \square \text{ Since } \frac{\partial w_1}{\partial r} &= \cos(\theta), & \frac{\partial w_1}{\partial \theta} &= -r \sin(\theta), \\ \frac{\partial w_2}{\partial r} &= \sin(\theta), & \frac{\partial w_2}{\partial \theta} &= r \cos(\theta), \end{aligned}$$

$$J = \begin{vmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{vmatrix} = r \cos^2(\theta) + r \sin^2(\theta) = r,$$

and $|J| = |r| = r$.

- For $0 \leq r \leq 1$ and $0 \leq \theta \leq 2\pi$, the joint pdf of (R, Θ) is

$$f_{R,\Theta}(r, \theta) = f_{X,Y}(r \cos(\theta), r \sin(\theta)) \times |J| = \frac{1}{\pi} r$$

and $f_{R,\Theta}(r, \theta) = 0$, otherwise.

□ By the theorem in LNp.7-25, $(\underline{R}, \underline{\Theta})$ are independent.

- Example. Let $\underline{X}_1, \dots, \underline{X}_n$ be independent and identically distributed (i.e., i.i.d.) exponential(λ). Let

$$\underline{Y}_i = \underline{X}_1 + \dots + \underline{X}_i, i = 1, \dots, n.$$

Find the distribution of $\underline{Y} = (Y_1, \dots, Y_n)$.

[Note. It has been shown that $\underline{Y}_i \sim \text{Gamma}(i, \lambda)$, $i = 1, \dots, n$.]

- The joint pdf of $\underline{X}_1, \dots, \underline{X}_n$ is

$$\begin{aligned} f_{\mathbf{X}}(x_1, \dots, x_n) &= \prod_{i=1}^n f_{X_i}(x_i) \\ &= \prod_{i=1}^n (\lambda e^{-\lambda x_i}) = \lambda^n e^{-\lambda(x_1 + \dots + x_n)}. \end{aligned}$$

for $0 \leq x_i < \infty$, $i = 1, \dots, n$.

- Since $x_1 = y_1 \equiv w_1(y_1, \dots, y_n)$,
 $x_2 = y_2 - y_1 \equiv w_2(y_1, \dots, y_n)$,
 \dots
 $x_n = y_n - y_{n-1} \equiv w_n(y_1, \dots, y_n)$,

we have

$$\frac{\partial w_i}{\partial y_j} = \begin{cases} 1, & \text{if } j = i, \\ -1, & \text{if } j = i - 1, \\ 0, & \text{otherwise,} \end{cases}$$

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

$$J = \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix} = 1, \text{ and } |J| = 1.$$

- For $0 \leq y_1 \leq y_2 \leq \dots \leq y_{i-1} \leq y_i \leq y_{i+1} \leq \dots \leq y_n < \infty$,

$$\begin{aligned} \underline{f}_{\mathbf{Y}}(y_1, \dots, y_n) &= \underline{f}_{\mathbf{X}}(y_1, y_2 - y_1, \dots, y_n - y_{n-1}) \times |J| \\ &= \lambda^n e^{-\lambda y_n}. \end{aligned}$$

and $\underline{f}_{\mathbf{Y}}(y_1, \dots, y_n) = 0$, otherwise.

- The marginal pdf of \underline{Y}_i is

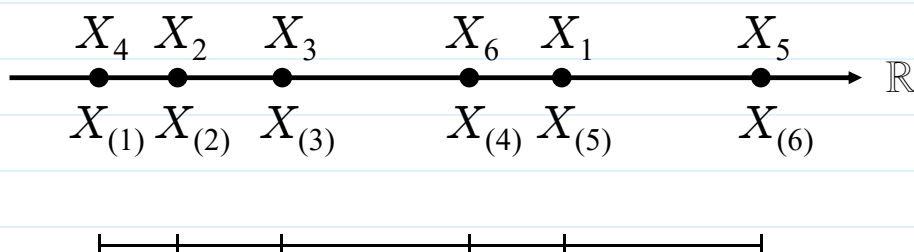
$$\begin{aligned} f_{Y_i}(y) &= \int_0^y \int_{y_1}^y \dots \int_{y_{i-2}}^y \int_y^\infty \int_{y_{i+1}}^\infty \dots \int_{y_{n-1}}^\infty \\ &\quad \lambda^n e^{-\lambda y_n} dy_n \dots dy_{y_{i+2}} dy_{i+1} dy_{i-1} \dots dy_2 dy_1 \\ &= \int_0^y \int_{y_1}^y \dots \int_{y_{i-2}}^y \lambda^i e^{-\lambda y} dy_{i-1} \dots dy_2 dy_1 \\ &= \lambda^i e^{-\lambda y} \frac{y^{i-1}}{(i-1)!}, \end{aligned}$$

for $y \geq 0$, and $\underline{f}_{Y_i}(y) = 0$, otherwise.

➤ Method of moment generating function.

- Based on the uniqueness theorem of moment generating function to be explained later in Chapter 7
- Especially useful to identify the distribution of sum of independent random variables.

• Order Statistics



➤ Definition. Let X_1, \dots, X_n be random variables. We sort the X_i 's and denote by

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

the order statistics. Using the notation,



NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)



$X_{(i)}$ = i th-smallest value in X_1, \dots, X_n , $i=1, 2, \dots, n$,

$X_{(1)}$ = min(X_1, \dots, X_n) is the minimum,

$X_{(n)}$ = max(X_1, \dots, X_n) is the maximum,

$R \equiv X_{(n)} - X_{(1)}$ is called range,

$S_j \equiv X_{(j)} - X_{(j-1)}$, $j=2, \dots, n$, are called j th spacing.

Q: What are the joint distributions of various order statistics and their marginal distributions?

➤ Definition. X_1, \dots, X_n are called i.i.d. (independent, identically distributed) with cdf F /pdf f /pmf p if the random variables X_1, \dots, X_n are independent and have a common marginal distribution with cdf F /pdf f /pmf p .

- Remark. In the discussion about order statistics, we only consider the case that X_1, \dots, X_n are i.i.d.

- Note. Although X_1, \dots, X_n are independent, their order statistics $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are not independent in general.

➤ Theorem. Suppose that X_1, \dots, X_n are i.i.d. with cdf F .

1. The cdf of $X_{(1)}$ is $1 - [1 - F(x)]^n$, and the cdf of $X_{(n)}$ is $[F(x)]^n$.
2. If \mathbf{X} are continuous and F has a pdf f , then the pdf of $X_{(1)}$ is $nf(x)[1 - F(x)]^{n-1}$, and the pdf of $X_{(n)}$ is $nf(x)[F(x)]^{n-1}$.

Proof. By the method of cumulative distribution function,

$$\begin{aligned} 1 - F_{X_{(1)}}(x) &= P(X_{(1)} > x) = P(X_1 > x, \dots, X_n > x) \\ &= P(X_1 > x) \cdots P(X_n > x) = [1 - F(x)]^n. \end{aligned}$$

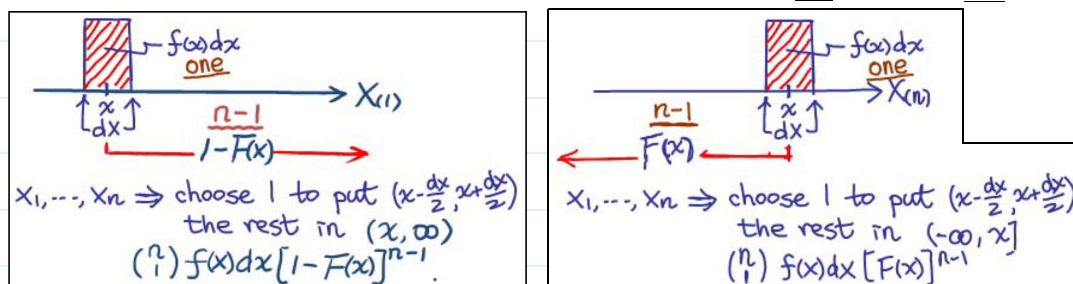
$$\begin{aligned} F_{X_{(n)}}(x) &= P(X_{(n)} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) \\ &= P(X_1 \leq x) \cdots P(X_n \leq x) = [F(x)]^n. \end{aligned}$$

$$\begin{aligned} f_{X_{(1)}}(x) &= \frac{d}{dx} F_{X_{(1)}}(x) \\ &= n[1 - F(x)]^{n-1} \left(\frac{d}{dx} F(x) \right) = nf(x)[1 - F(x)]^{n-1}. \end{aligned}$$

$$\begin{aligned} f_{X_{(n)}}(x) &= \frac{d}{dx} F_{X_{(n)}}(x) \\ &= n[F(x)]^{n-1} \left(\frac{d}{dx} F(x) \right) = nf(x)[F(x)]^{n-1}. \end{aligned}$$

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

■ Graphical interpretation for the pdfs of $X_{(1)}$ and $X_{(n)}$.



- Example. n light bulbs are placed in service at time $t=0$, and allowed to burn continuously. Denote their lifetimes by X_1, \dots, X_n , and suppose that they are i.i.d. with cdf F .

If burned out bulbs are not replaced, then the room goes dark at time

$$Y = X_{(n)} = \max(X_1, \dots, X_n).$$

- If $n=5$ and F is exponential with $\lambda = 1$ per month, then

$$F(x) = 1 - e^{-x}, \text{ for } x \geq 0, \text{ and } 0, \text{ for } x < 0.$$

- The cdf of Y is

$$F_Y(y) = (1 - e^{-y})^5, \text{ for } y \geq 0, \text{ and } 0, \text{ for } y < 0,$$

and its pdf is $5(1 - e^{-y})^4 e^{-y}$, for $y \geq 0$, and 0, for $y < 0$.

▣ The probability that the room is still lighted after two months is $P(Y > 2) = 1 - F_Y(2) = 1 - (1 - e^{-2})^5$.

➤ Theorem. Suppose that X_1, \dots, X_n are i.i.d. with pmf p /pdf f . Then, the joint pmf/pdf of $X_{(1)}, \dots, X_{(n)}$ is

$$\begin{aligned} p_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) \\ = n! \times p(x_1) \times \dots \times p(x_n), \end{aligned}$$

$$\text{or } f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) \\ = n! \times f(x_1) \times \dots \times f(x_n),$$

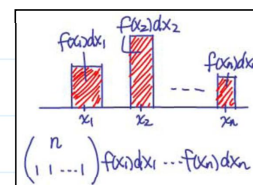
for $x_1 \leq x_2 \leq \dots \leq x_n$, and 0 otherwise.

Proof. For $x_1 \leq x_2 \leq \dots \leq x_n$,

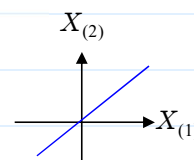
$$\begin{aligned} p_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) \\ = P(X_{(1)} = x_1, \dots, X_{(n)} = x_n) \\ = \sum_{\substack{(i_1, \dots, i_n): \\ \text{permutations of} \\ (1, \dots, n)}} P(X_1 = x_{i_1}, \dots, X_n = x_{i_n}) \\ = \sum_{\substack{(i_1, \dots, i_n): \\ \text{permutations of} \\ (1, \dots, n)}} p(x_1) \times \dots \times p(x_n) \\ = n! \times p(x_1) \times \dots \times p(x_n). \end{aligned}$$

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

$$\begin{aligned} f_{X_{(1)}, \dots, X_{(n)}}(x_1, \dots, x_n) dx_1 \cdots dx_n \\ \approx P\left(x_1 - \frac{dx_1}{2} < X_{(1)} < x_1 + \frac{dx_1}{2}, \dots, \right. \\ \left. x_n - \frac{dx_n}{2} < X_{(n)} < x_n + \frac{dx_n}{2}\right) \\ = \sum_{\substack{(i_1, \dots, i_n): \\ \text{permutations of} \\ (1, \dots, n)}} P\left(\frac{x_{i_1} - \frac{dx_{i_1}}{2}}{2} < X_1 < \frac{x_{i_1} + \frac{dx_{i_1}}{2}}{2}, \dots, \right. \\ \left. \frac{x_{i_n} - \frac{dx_{i_n}}{2}}{2} < X_n < \frac{x_{i_n} + \frac{dx_{i_n}}{2}}{2}\right) \\ \approx \sum_{\substack{(i_1, \dots, i_n): \\ \text{permutations of} \\ (1, \dots, n)}} f(x_1) \times \dots \times f(x_n) dx_1 \cdots dx_n \\ = n! \times f(x_1) \times \dots \times f(x_n) dx_1 \cdots dx_n. \end{aligned}$$



■ **Q:** Examine whether $X_{(1)}, \dots, X_{(n)}$ are independent using the Theorem in LNp.7-25.



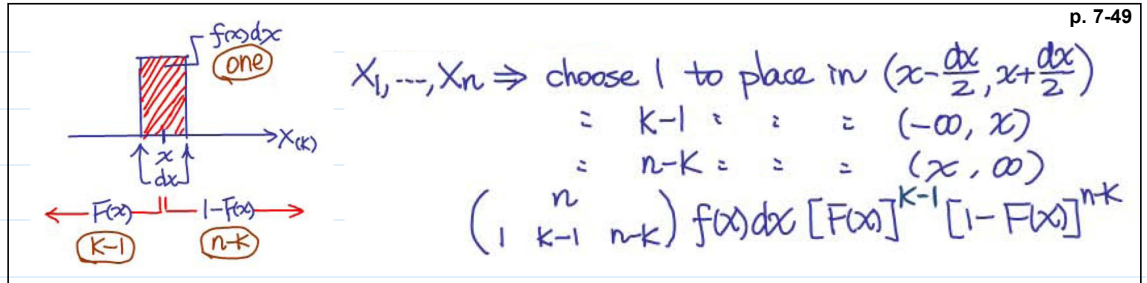
➤ Theorem. If X_1, \dots, X_n are i.i.d. with cdf F and pdf f , then

1. The pdf of the k^{th} order statistic $X_{(k)}$ is

$$\begin{aligned} f_{X_{(k)}}(x) \\ = \binom{n}{1, k-1, n-k} f(x) F(x)^{k-1} [1 - F(x)]^{n-k}. \end{aligned}$$

2. The cdf of $X_{(k)}$ is

$$F_{X_{(k)}}(x) = \sum_{m=k}^n \binom{n}{m} [F(x)]^m [1 - F(x)]^{n-m}.$$



$$F_{X_{(k)}}(x) = P(X_{(k)} \leq x)$$

$$= P(\text{at least } k \text{ of the } X_i\text{'s are } \leq x)$$

$$= \sum_{m=k}^n P(\text{exact } m \text{ of the } X_i\text{'s are } \leq x) \longrightarrow X_{(k)}$$

$$= \sum_{m=k}^n \binom{n}{m} [F(x)]^m [1-F(x)]^{n-m}$$

➤ Theorem. If X_1, \dots, X_n are i.i.d. with cdf F and pdf f , then

1. The joint pdf of $X_{(1)}$ and $X_{(n)}$ is

$$f_{X_{(1)}, X_{(n)}}(s, t) = \underline{n(n-1)f(s)f(t)[F(t) - F(s)]^{n-2}},$$

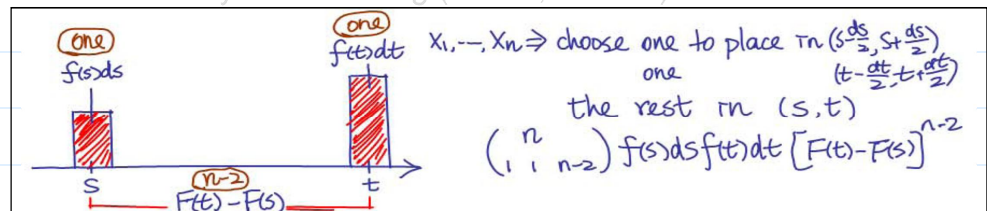
for $s \leq t$, and 0 otherwise.

2. The pdf of the range $R = X_{(n)} - X_{(1)}$ is

$$f_R(r) = \int_{-\infty}^{\infty} f_{X_{(1)}, X_{(n)}}(u, u+r) du,$$

for $r \geq 0$, and 0 otherwise.

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)



➤ Theorem. If X_1, \dots, X_n are i.i.d. with cdf F and pdf f , then

1. The joint pdf of $X_{(i)}$ and $X_{(j)}$, where $1 \leq i < j \leq n$, is

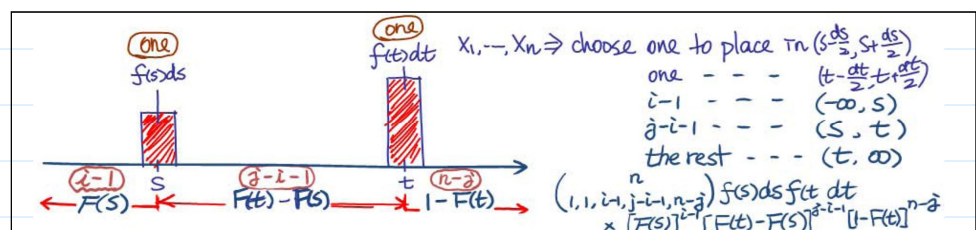
$$f_{X_{(i)}, X_{(j)}}(s, t) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} f(s)f(t) \times [F(s)]^{i-1} [F(t) - F(s)]^{j-i-1} [1-F(t)]^{n-j},$$

for $s \leq t$, and 0 otherwise.

2. The pdf of the j^{th} spacing $S_j = X_{(j)} - X_{(j-1)}$ is

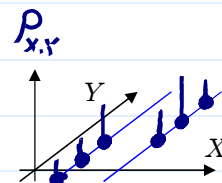
$$f_{S_j}(s) = \int_{-\infty}^{\infty} f_{X_{(j-1)}, X_{(j)}}(u, u+s) du,$$

for $s \geq 0$, and zero otherwise.



Conditional Distribution

- Definition. Let $\underline{X} (\in \mathbb{R}^n)$ and $\underline{Y} (\in \mathbb{R}^m)$ be discrete random vectors and $(\underline{X}, \underline{Y})$ have a joint pmf $p_{\underline{X}, \underline{Y}}(\underline{x}, \underline{y})$, then the conditional joint pmf of \underline{Y} given $\underline{X}=\underline{x}$ is defined as



$$\begin{aligned} p_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) &\equiv P(\{\underline{Y} = \underline{y}\} | \{\underline{X} = \underline{x}\}) = \frac{P(\{\underline{X} = \underline{x}, \underline{Y} = \underline{y}\})}{P(\{\underline{X} = \underline{x}\})} \\ &= \frac{p_{\underline{X}, \underline{Y}}(\underline{x}, \underline{y})}{p_{\underline{X}}(\underline{x})} = \frac{\text{joint}}{\text{marginal}} \end{aligned}$$

if $p_{\underline{X}}(\underline{x}) > 0$. The probability is defined to be zero if $p_{\underline{X}}(\underline{x}) = 0$.

➤ Some Notes.

- For each fixed \underline{x} , $p_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})$ is a joint pmf for \underline{y} , since

$$\sum_{\underline{y}} p_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = \frac{1}{p_{\underline{X}}(\underline{x})} \sum_{\underline{y}} p_{\underline{X}, \underline{Y}}(\underline{x}, \underline{y}) = \frac{1}{p_{\underline{X}}(\underline{x})} \times p_{\underline{X}}(\underline{x}) = 1.$$

- For an event B of \underline{Y} , the probability that $\underline{Y} \in B$ given $\underline{X}=\underline{x}$ is

$$P(\underline{Y} \in B | \underline{X} = \underline{x}) = \sum_{\underline{u} \in B} p_{\underline{Y}|\underline{X}}(\underline{u}|\underline{x}).$$

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- The conditional joint cdf of \underline{Y} given $\underline{X}=\underline{x}$ can be similarly defined from the conditional joint pmf $p_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})$, i.e.,

$$F_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = P(\underline{Y} \leq \underline{y} | \underline{X} = \underline{x}) = \sum_{\underline{u} \leq \underline{y}} p_{\underline{Y}|\underline{X}}(\underline{u}|\underline{x}).$$

- Theorem. Let $\underline{X}_1, \dots, \underline{X}_m$ be independent and

$$\underline{X}_i \sim \text{Poisson}(\lambda_i), \quad i=1, \dots, m.$$

Let $\underline{Y} = \underline{X}_1 + \dots + \underline{X}_m$, then

$$(\underline{X}_1, \dots, \underline{X}_m | \underline{Y} = n) \sim \text{Multinomial}(n, m, p_1, \dots, p_m),$$

where $p_i = \lambda_i / (\lambda_1 + \dots + \lambda_m)$ for $i=1, \dots, m$.



Proof. The joint pmf of $(\underline{X}_1, \dots, \underline{X}_m, \underline{Y})$ is

$$\begin{aligned} p_{\underline{X}, \underline{Y}}(x_1, \dots, x_m, n) &= P(\{X_1 = x_1, \dots, X_m = x_m\} \cap \{Y = n\}) \\ &= \begin{cases} P(X_1 = x_1, \dots, X_m = x_m), & \text{if } x_1 + \dots + x_m = n, \\ 0, & \text{if } x_1 + \dots + x_m \neq n. \end{cases} \end{aligned}$$

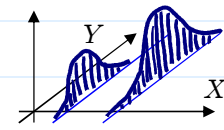
Furthermore, the distribution of \underline{Y} is Poisson($\lambda_1 + \dots + \lambda_m$), i.e.,

$$p_Y(n) = P(Y = n) = \frac{e^{-(\lambda_1 + \dots + \lambda_m)} (\lambda_1 + \dots + \lambda_m)^n}{n!}.$$

Therefore, for $\underline{x} = (x_1, \dots, x_m)$ where $x_i \in \{0, 1, 2, \dots\}$, $i = 1, \dots, m$, and $x_1 + \dots + x_m = n$, the conditional joint pmf of \underline{X} given $Y = n$ is

$$\begin{aligned} p_{\underline{X}|\underline{Y}}(\underline{x}|n) &= \frac{p_{\underline{X},Y}(x_1, \dots, x_m, n)}{p_Y(n)} = \frac{\prod_{i=1}^m \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!}}{\frac{e^{-(\lambda_1 + \dots + \lambda_m)} (\lambda_1 + \dots + \lambda_m)^n}{n!}} \\ &= \frac{n!}{x_1! \times \dots \times x_m!} \times \left(\frac{\lambda_1}{\lambda_1 + \dots + \lambda_m} \right)^{x_1} \times \dots \times \left(\frac{\lambda_m}{\lambda_1 + \dots + \lambda_m} \right)^{x_m}. \end{aligned}$$

- Definition. Let $\underline{X} (\in \mathbb{R}^n)$ and $\underline{Y} (\in \mathbb{R}^m)$ be continuous random vectors and $(\underline{X}, \underline{Y})$ have a joint pdf $f_{\underline{X},\underline{Y}}(\underline{x}, \underline{y})$, then the conditional joint pdf of \underline{Y} given $\underline{X} = \underline{x}$ is defined as



$$f_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) \equiv \frac{f_{\underline{X},\underline{Y}}(\underline{x}, \underline{y})}{f_{\underline{X}}(\underline{x})} = \frac{\text{joint}}{\text{marginal}},$$

if $f_{\underline{X}}(\underline{x}) > 0$, and 0 otherwise.

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

➤ Some Notes.

- $P(\underline{X} = \underline{x}) = 0$ for a continuous random vector \underline{X} .
- The justification of $f_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})$ comes from

$$\begin{aligned} P(\underline{Y} \leq \underline{y} | \underline{x} - (\Delta \underline{x}/2) < \underline{X} \leq \underline{x} + (\Delta \underline{x}/2)) \\ &= \frac{\int_{-\infty}^{\underline{y}} \int_{\underline{x} - (\Delta \underline{x}/2)}^{\underline{x} + (\Delta \underline{x}/2)} f_{\underline{X},\underline{Y}}(\underline{u}, \underline{v}) d\underline{u} d\underline{v}}{\int_{\underline{x} - (\Delta \underline{x}/2)}^{\underline{x} + (\Delta \underline{x}/2)} f_{\underline{X}}(\underline{t}) d\underline{t}} \\ &\approx \frac{\int_{-\infty}^{\underline{y}} f_{\underline{X},\underline{Y}}(\underline{x}, \underline{v}) |\Delta \underline{x}| d\underline{v}}{f_{\underline{X}}(\underline{x}) |\Delta \underline{x}|} = \int_{-\infty}^{\underline{y}} \frac{f_{\underline{X},\underline{Y}}(\underline{x}, \underline{v})}{f_{\underline{X}}(\underline{x})} d\underline{v} \end{aligned}$$

- For each fixed \underline{x} , $f_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})$ is a joint pdf for \underline{y} , since

$$\int_{-\infty}^{\infty} f_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) d\underline{y} = \frac{1}{f_{\underline{X}}(\underline{x})} \int_{-\infty}^{\infty} f_{\underline{X},\underline{Y}}(\underline{x}, \underline{y}) d\underline{y} = \frac{1}{f_{\underline{X}}(\underline{x})} \times f_{\underline{X}}(\underline{x}) = 1.$$

- For an event B of \underline{Y} , we can write

$$P(\underline{Y} \in B | \underline{X} = \underline{x}) = \int_B f_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) d\underline{y}.$$

- The conditional joint cdf of \underline{Y} given $\underline{X} = \underline{x}$ can be similarly defined from the conditional joint pdf $f_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})$, i.e.,

$$F_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = P(\underline{Y} \leq \underline{y} | \underline{X} = \underline{x}) = \int_{-\infty}^{\underline{y}} f_{\underline{Y}|\underline{X}}(\underline{t}|\underline{x}) d\underline{t}.$$

➤ Example. If X and Y have a joint pdf

$$f(x, y) = \frac{2}{(1+x+y)^3},$$

for $0 \leq x, y < \infty$, then

$$f_X(x) = \int_0^\infty f(x, y) dy = -\frac{1}{(1+x+y)^2} \Big|_0^\infty = \frac{1}{(1+x)^2},$$

for $0 \leq x < \infty$. So,

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{2(1+x)^2}{(1+x+y)^3},$$

$$\begin{aligned} \text{and, } P(Y > c | X = x) &= \int_c^\infty \frac{2(1+x)^2}{(1+x+y)^3} dy \\ &= -\frac{(1+x)^2}{(1+x+y)^2} \Big|_{y=c}^\infty = \frac{(1+x)^2}{(1+x+c)^2}. \end{aligned}$$

- Mixed Joint Distribution: Definition of conditional distribution can be similarly generalized to the case in which some random variables are discrete and the others continuous (see a later example).
- Theorem (Multiplication Law). Let \mathbf{X} and \mathbf{Y} be random vectors and (\mathbf{X}, \mathbf{Y}) have a joint pdf $f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$ /pmf $p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$, then

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

$$f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \times f_{\mathbf{X}}(\mathbf{x}).$$

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

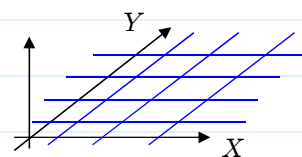
Proof. By the definition of conditional distribution.

- Theorem (Law of Total Probability). Let \mathbf{X} and \mathbf{Y} be random vectors and (\mathbf{X}, \mathbf{Y}) have a joint pdf $f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$ /pmf $p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$, then

$$p_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{x}=-\infty}^{\infty} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}), \quad \text{or}$$

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_{-\infty}^{\infty} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

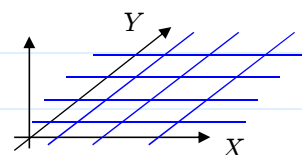
Proof. By the definition of marginal distribution and the multiplication law.



- Theorem (Bayes Theorem). Let \mathbf{X} and \mathbf{Y} be random vectors and (\mathbf{X}, \mathbf{Y}) have a joint pdf $f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$ /pmf $p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$, then

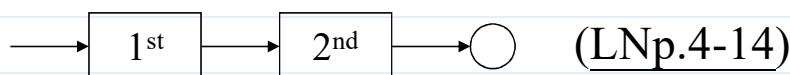
$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) p_{\mathbf{X}}(\mathbf{x})}{\sum_{\mathbf{x}=-\infty}^{\infty} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) p_{\mathbf{X}}(\mathbf{x})}, \quad \text{or}$$

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{\int_{-\infty}^{\infty} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}}.$$



Proof. By the definition of conditional distribution, multiplication law, and the law of total probability.

➤ Example.



- Suppose that $\underline{X} \sim \underline{\text{Uniform}}(0, 1)$, and

$(\underline{Y}_1, \dots, \underline{Y}_n | \underline{X} = x)$ are i.i.d. with Bernoulli(x), i.e.,

$$p_{\mathbf{Y}|\mathbf{X}}(y_1, \dots, y_n | x) = x^{y_1 + \dots + y_n} (1 - x)^{n - (y_1 + \dots + y_n)},$$

for $y_1, \dots, y_n \in \{0, 1\}$.

- By the multiplication law, for $y_1, \dots, y_n \in \{0, 1\}$ and $0 < x < 1$,

$$p_{\mathbf{Y}, \mathbf{X}}(y_1, \dots, y_n, x) = x^{y_1 + \dots + y_n} (1 - x)^{n - (y_1 + \dots + y_n)}.$$

- Suppose that we observed $\underline{Y}_1 = 1, \dots, \underline{Y}_n = 1$.

- By the law of total probability,

$$\begin{aligned} P(\underline{Y}_1 = 1, \dots, \underline{Y}_n = 1) &= p_{\mathbf{Y}}(1, \dots, 1) \\ &= \int_0^1 p_{\mathbf{Y}|\mathbf{X}}(1, \dots, 1 | x) f_X(x) dx \\ &= \int_0^1 x^n dx = \frac{1}{n+1} x^{n+1} \Big|_0^1 = \frac{1}{n+1}. \end{aligned}$$

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)

- And, by Bayes' Theorem,

$$\begin{aligned} f_{\underline{X}|\mathbf{Y}}(x | \underline{Y}_1 = 1, \dots, \underline{Y}_n = 1) \\ = \frac{p_{\mathbf{Y}|\mathbf{X}}(1, \dots, 1 | x) f_X(x)}{p_{\mathbf{Y}}(1, \dots, 1)} = \frac{(n+1)x^n}{1}. \end{aligned}$$

for $0 < x < 1$, i.e., $(\underline{X} | \underline{Y}_1 = 1, \dots, \underline{Y}_n = 1) \sim \underline{\text{Beta}}(n+1, 1)$.

(cf., marginal distribution of $\underline{X} \sim \underline{\text{Uniform}}(0, 1) = \underline{\text{Beta}}(1, 1)$.)

- If there were an $(n+1)^{\text{st}}$ Bernoulli trial \underline{Y}_{n+1} ,

$$\begin{aligned} P(\underline{Y}_{n+1} = 1 | \underline{Y}_1 = 1, \dots, \underline{Y}_n = 1) \\ = \frac{P(\underline{Y}_1 = 1, \dots, \underline{Y}_{n+1} = 1)}{P(\underline{Y}_1 = 1, \dots, \underline{Y}_n = 1)} = \frac{1/(n+2)}{1/(n+1)} = \frac{n+1}{n+2}. \end{aligned}$$

- (exercise) In general, it can be shown that

$$(\underline{X} | \underline{Y}_1 = y_1, \dots, \underline{Y}_n = y_n) \sim \underline{\text{Beta}}((y_1 + \dots + y_n) + 1, n - (y_1 + \dots + y_n) + 1).$$

- Theorem (Conditional Distribution & Independent). Let \mathbf{X} and \mathbf{Y} be random vectors and (\mathbf{X}, \mathbf{Y}) have a joint pdf $f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$ /pmf $p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$. Then, \mathbf{X} and \mathbf{Y} are independent, i.e.,

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

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

if and only if

$$\underline{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = p_{\mathbf{Y}}(\mathbf{y})}, \quad \text{or}$$

$$\underline{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = f_{\mathbf{Y}}(\mathbf{y})}.$$

Proof. By the definition of conditional distribution.

➤ Intuition.

- The 2 graphs about the joint pmf/pdf of independent r.v.'s in LNp.7-27
- $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ or $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ offers information about the distribution of \mathbf{Y} when $\mathbf{X}=\mathbf{x}$.

$p_{\mathbf{Y}}(\mathbf{y})$ or $f_{\mathbf{Y}}(\mathbf{y})$ offers information about the distribution of \mathbf{Y} when \mathbf{X} not observed.

❖ Reading: textbook, Sec 6.4, 6.5

NTHU MATH 2810, 2024, Lecture Notes
made by S.-W. Cheng (NTHU, Taiwan)