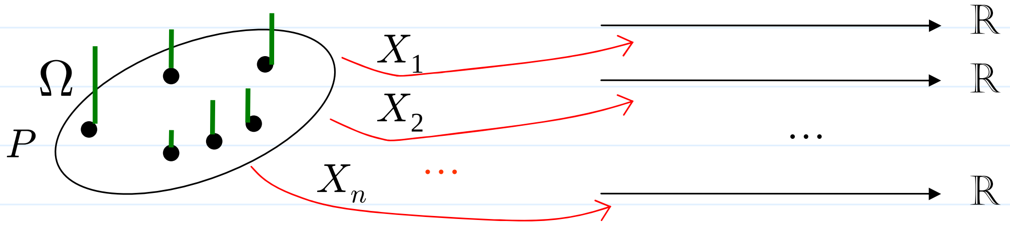
Jointly Distributed Random Variables

- Recall. In Chapters 4 and 5, focus on *univariate* random variable.
  - ➢However, often a single experiment will have more than one random variable which is of interest.



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

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

are said to be *jointly distributed*.

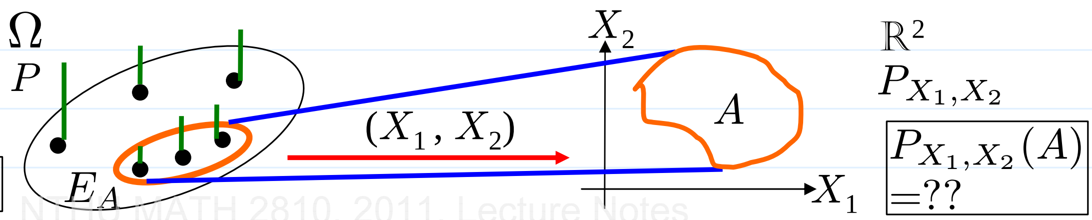
    - We can regard n jointly distributed r.v.'s as a *random vector*

$$\mathbf{X}=(X_1, \ldots, X_n): \Omega \rightarrow \mathbb{R}^n.$$

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

A occurs ⇔ E_A occurs

$P_{X_1,X_2}(A) = P(E_A)$

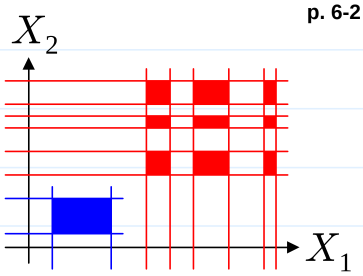


- ➢For $A\subset\mathbb{R}^n$,

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

- ➢For $A_i\subset\mathbb{R}$, i=1, …, n,

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



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

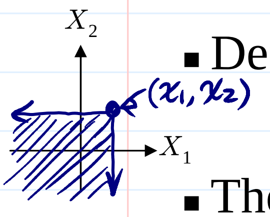
- **Q**: Why need joint distribution? Why are marginal distributions not enough?
  - ➢Example (Coin Tossing, LNp.4-2).

X_2: # of head on 1st 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
- [read numbers]: joint distribution of another (X_1', X_2')
- Some findings:
  - When joint distribution is given, its corresponding marginal distributions are known, e.g.,
    - $P(X_1=i)=P(X_1=i, X_2=0)+P(X_1=i, X_2=1)$, $i=0, 1, 2, 3$.
  - (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: $X_1, \ldots, X_n$ 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



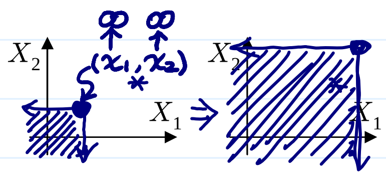
- Definition. The joint cdf of $\mathbf{X}=(X_1, \ldots, X_n)$ is defined as

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

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

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

(ii) $\lim_{x_1, x_2, \cdots, x_n \to \infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = 1$



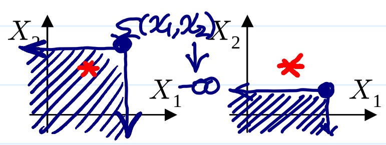
Proof. Let $z_{im} \uparrow \infty$, $1 \le i \le n$.

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

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

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

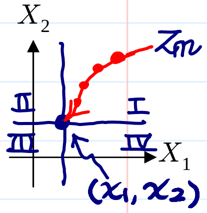
$$\lim_{x_i \to -\infty} F_{\mathbf{X}}(x_1, \ldots, x_n) = 0.$$



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

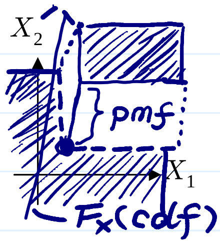
Let $A_m = (-\infty, x_1) \times \cdots \times (-\infty, z_{im}) \times \cdots \times (-\infty, x_n)$.

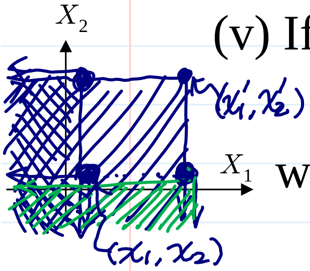
Then, $A_m \downarrow \emptyset \Rightarrow \lim P(A_m) = P(\emptyset) = 0$.



(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 $\mathbf{x}=(x_1, \ldots, x_n)$ and $\mathbf{z}_m=(z_{1m}, \ldots, z_{nm})$ such that $\mathbf{z}_m \downarrow \mathbf{x}$, then

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



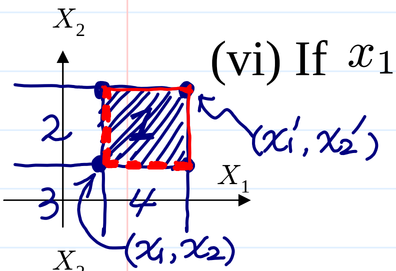


(v) If $x_i \leq x_i', i = 1, \ldots, n$, then

$$F_{\mathbf{X}}(x_1, \ldots, x_n) \leq F_{\mathbf{X}}(t_1, \ldots, t_n) \leq F_{\mathbf{X}}(x_1', \ldots, x_n').$$

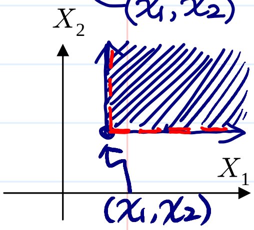
where $t_i \in \{x_i, x_i'\}, i = 1, 2, \ldots, n$. When n=2, we have

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



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

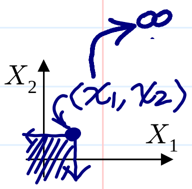
$$\begin{aligned} & P(x_1 < X_1 \leq x_1', x_2 < X_1 \leq x_2') \\ = \quad & F_{X_1,X_2}(x_1', x_2') - F_{X_1,X_2}(x_1, x_2') \\ & -F_{X_1,X_2}(x_1', x_2) + F_{X_1,X_2}(x_1, x_2). \end{aligned}$$



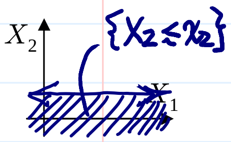
In particular, let $x_1' \uparrow \infty$ and $x_2' \uparrow \infty$, we get

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

(vii) The joint cdf of $X_1, \ldots, X_k$, $k < n$, is



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



In particular, the marginal cdf of X_1 is

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

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

Joint Probability Mass Function

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

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

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

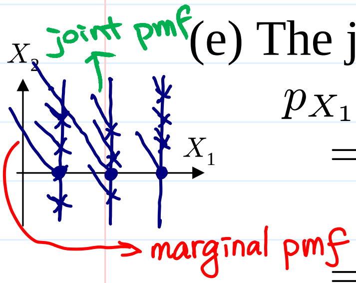
(a) $p_{\mathbf{X}}(x_1, \ldots, x_n) \geq 0, \quad$ for $-\infty < x_i < \infty$, $i = 1, \ldots, n$.

(b) There exists a finite or countably infinite set $\mathcal{X} \subset \mathbb{R}^n$ such that $p_{\mathbf{X}}(x_1, \ldots, x_n) = 0, \quad \text{for } (x_1, \ldots, x_n) \notin \mathcal{X}.$

(c) $\sum_{\mathbf{x} \in \mathcal{X}} p_{\mathbf{X}}(\mathbf{x}) = 1, \quad \text{where } \mathbf{x} = (x_1, \ldots, x_n).$

(d) For $A \subset \mathbb{R}^n$,

$$P(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A \cap \mathcal{X}} p_{\mathbf{X}}(\mathbf{x}).$$



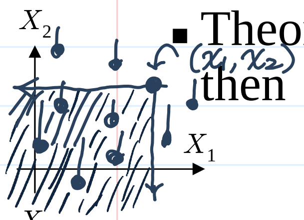
(e) The joint pmf of $X_1, \ldots, X_k$, $k < n$, is

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

In particular, the marginal cdf of X_1 is

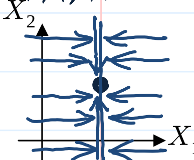
$$\begin{aligned} p_{X_1}(x) &= P(X_1 = x) \\ &= \sum_{\substack{(x, x_2, \ldots, x_n) \in \mathcal{X} \\ -\infty < x_2 < \infty, \ldots, -\infty < x_n < \infty}} p_{\mathbf{X}}(x, x_2, x_3, \ldots, x_n). \end{aligned}$$

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

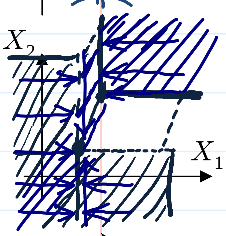


- Theorem. If $F_{\mathbf{X}}$ and $p_{\mathbf{X}}$ are the joint cdf and joint pmf of $\mathbf{X}$, then

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \sum_{\substack{(t_1,\ldots,t_n) \in \mathcal{X} \\ t_1 \le x_1, \ldots, t_n \le x_n}} p_{\mathbf{X}}(t_1, \ldots, t_n), \quad \text{and}$$



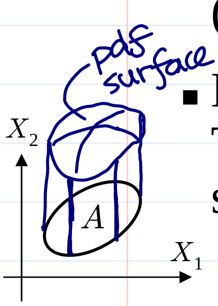
$$p_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{x}) - F_{\mathbf{X}}(\mathbf{x}-), \quad \text{where } \mathbf{x} = (x_{10}, \ldots, x_{n0}), \text{ and}$$



$$\begin{aligned} F^{(1)}_{\mathbf{X}}(x_2, \ldots, x_n) &\equiv F_{\mathbf{X}}(x_{10}, x_2, \ldots, x_n) - F_{\mathbf{X}}(x_{10}-, x_2, \ldots, x_n) \\ F^{(2)}_{\mathbf{X}}(x_3, \ldots, x_n) &\equiv F^{(1)}_{\mathbf{X}}(x_{20}, x_3, \ldots, x_n) - F^{(1)}_{\mathbf{X}}(x_{20}-, x_3, \ldots, x_n) \\ \cdots &\equiv \cdots \\ F_{\mathbf{X}}(\mathbf{x}) - F_{\mathbf{X}}(\mathbf{x}-) &\equiv F^{(n-1)}_{\mathbf{X}}(x_{n0}) - F^{(n-1)}_{\mathbf{X}}(x_{n0}-) \end{aligned}$$

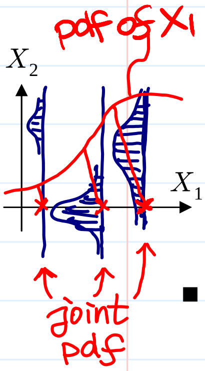
➢Joint Probability Density Function

- Definition. A function $f_{\mathbf{X}}(x_1, \ldots, x_n)$ can be a joint pdf if
  (1) $f_{\mathbf{X}}(x_1, \ldots, x_n) \ge 0$, for $-\infty < x_i < \infty$, $i=1, \ldots, n$.
  (2) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1.$



- Definition. Suppose that $X_1, \ldots, X_n$ are continuous r.v.'s. The joint pdf of $\mathbf{X}=(X_1, \ldots, X_n)$ is a function $f_{\mathbf{X}}(x_1, \ldots, x_n)$ satisfying (1) and (2) above, and for any event $A \subset \mathbb{R}^n$,

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



- Theorem. Suppose that $f_{\mathbf{X}}$ is the joint pdf of $\mathbf{X}=(X_1, \ldots, X_n)$. Then, the joint pdf of $X_1, \ldots, X_k$, $k < n$, is

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

In particular, the marginal pdf of X_1 is

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

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

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

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

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

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

    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

$p(x, y)$		X = 1	2	3	$p_Y(y)$
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: When balls drawn without replacement, why do X and Y have same marginal distributions?

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

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

- Multinomial Distribution
  - Recall. Partitions
    - If $n \geq 1$ and $n_1, \ldots, n_m \geq 0$ are integers for which

    $$n_1 + \cdots + n_m = n,$$

    then a set of n elements may be partitioned into m subsets of sizes $n_1, \ldots, n_m$ in

$$\binom{n}{n_1,\cdots,n_m} = \frac{n!}{n_1!\times\cdots\times n_m!} \quad \text{ways.}$$

- Example: 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(each face appears twice)=??
  - Sample space of rolling the die once (basic experiment):

    $$\Omega_0 = \{1, 2, 3, 4, 5, 6\}.$$

  - The sample space for the 12 trials is

    $$\Omega = \Omega_0 \times \cdots \times \Omega_0 = \Omega_0^{12}$$

    An outcome $\omega \in \Omega$ is $\omega = (i_1, i_2, \ldots, i_{12})$, where $1 \le i_1, \ldots, i_{12} \le 6$.
  - There are 6^{12} possible outcomes in Ω, i.e., $\#\Omega = 6^{12}$.
  - 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(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 m types of outcomes. Denote its sample space as

    $$\Omega_0 = \{1, 2, \ldots, m\}.$$

    Let $p_i = P$(outcome i appears),

    then, (i) $p_1, \ldots, p_m \ge 0$, and
    (ii) $p_1 + \cdots + p_m = 1$.
  - Repeat the basic experiment n times. Then, the sample space for the n trials is

    $$\Omega = \Omega_0 \times \cdots \times \Omega_0 = \Omega_0^n$$

    Let X_i = # of trials with outcome i, $i=1, \ldots, m$,

    Then, (i) $X_1, \ldots, X_m: \Omega \to \mathbb{R}$, and
    (ii) $X_1 + \cdots + X_m = n$.
  - The joint pmf of $X_1, \ldots, X_m$ is

    $$\begin{aligned} p_{\mathbf{X}}(x_1,\ldots,x_m) &= P(X_1 = x_1, \ldots, X_m = x_m) \\ &= \binom{n}{x_1,\cdots,x_m} p_1^{x_1} \times \cdots \times p_m^{x_m}. \end{aligned}$$

    for $x_1, \ldots, x_m \ge 0$ and $x_1 + \cdots + x_m = n$.

Proof. The probability of any sequence with x_i i's is

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

and there are

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

such sequences.

- The distribution of a random vector **X**=(X_1, … , X_m) with the above joint pmf is called the *multinomial* distribution with parameters n, m, and p_1, …, p_m, denoted by Multinomial(n, m, p_1, …, p_m).
  - The multinomial distribution is called after the Multinomial Theorem:

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

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

- Some Properties.
  - Because $X_i = n - (X_1+\cdots+X_{i-1}+X_{i+1}+\cdots+X_m)$, and

$$p_i = 1 - (p_1+\cdots+ p_{i-1}+p_{i+1}+\cdots+p_m),$$

wlog, we can write

$$(X_1,\ldots, X_{m-1}, X_m) \rightarrow (X_1,\ldots, X_{m-1}, n-(X_1+\cdots+X_{m-1}))$$

  - Marginal Distribution. Suppose that

$$(X_1, \ldots , X_m) \sim \text{Multinomial}(n, m, p_1, \ldots, p_m).$$

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

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

is Multinomial(n, k+1, p_1, …, p_k, $p_{k+1}+\cdots+p_m$).

In particular, $X_i \sim$ Binomial(n, p_i)

  - Mean and Variance.

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

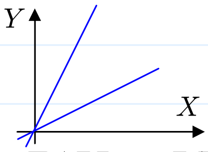
for i = 1, …, m.

➢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: $P(Y\geq 2X$ or $X\geq 2Y)$=??

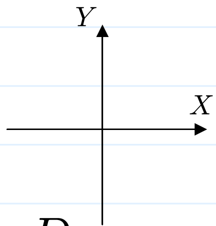


- The event $\{Y\geq 2X\}\cup\{X\geq 2Y\}$ is
- So, $P(Y\geq 2X$ or $X\geq 2Y)=P(Y\geq 2X)+P(X\geq 2Y)=2/3$ because

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

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 $D\subset\mathbb{R}^2$ and $0<\alpha=\text{Area}(D)<\infty$, then

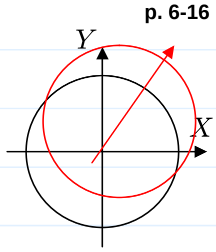
$$f(x,y)=c\cdot\mathbf{1}_D(x,y)$$

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

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

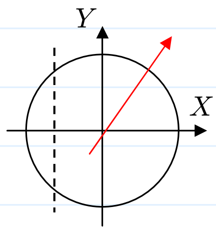
$$f(x,y)=\frac{1}{\pi}\mathbf{1}_D(x,y)$$

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.

➢When the joint distribution is given, the marginal distributions are known.

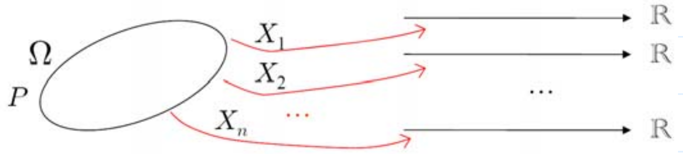
➢The converse statement does not hold in general.

➢However, when random variables are independent,

marginal distributions + independence ⇒ joint distribution.

- Definition. The random variables $X_1, \ldots, X_n$ are called (*mutually*) *independent* if and only if for *any* (measurable) sets $A_i \subset \mathbb{R}$, $i=1, \ldots, n$, the events

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



are independent. That is,

$$P(X_{i_1} \in A_{i_1}, X_{i_2} \in A_{i_2}, \cdots, X_{i_k} \in A_{i_k})$$
$$= P(X_{i_1} \in A_{i_1}) \times P(X_{i_2} \in A_{i_2}) \times \cdots \times P(X_{i_k} \in A_{i_k}),$$

for *any* $1 \le i_1 < i_2 < \cdots < i_k \le n$; $k=2, \ldots, n$.

  - If $X_1, \ldots, X_n$ are independent, for $1 \le k < n$,

$$P(X_{k+1} \in A_{k+1}, \ldots, X_n \in A_n | X_1 \in A_1, \ldots, X_k \in A_k)$$
$$= P(X_{k+1} \in A_{k+1}, \ldots, X_n \in A_n)$$

  provided that $P(X_1 \in A_1, \ldots, X_k \in A_k) > 0$. In other words, $X_1, \ldots, X_k$ do not carry information about $X_{k+1}, \ldots, X_n$.

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

  (1) $F_{\mathbf{X}}(x_1, \ldots, x_n) = F_{X_1}(x_1) \times \cdots \times F_{X_n}(x_n)$, where $F_{\mathbf{X}}$ is the joint cdf of $\mathbf{X}$ and F_{X_i} is the marginal cdf of X_i for $i=1,\ldots,n$.

  (2) Suppose that $X_1, \ldots, X_n$ are discrete random variables. $p_{\mathbf{X}}(x_1, \ldots, x_n) = p_{X_1}(x_1) \times \cdots \times p_{X_n}(x_n)$, where $p_{\mathbf{X}}$ is the joint pmf of $\mathbf{X}$ and p_{X_i} is the marginal pmf of X_i for $i=1,\ldots,n$.

  (3) Suppose that $X_1, \ldots, X_n$ are continuous random variables. $f_{\mathbf{X}}(x_1, \ldots, x_n) = f_{X_1}(x_1) \times \cdots \times f_{X_n}(x_n)$, where $f_{\mathbf{X}}$ is the joint pdf of $\mathbf{X}$ and f_{X_i} is the marginal pdf of X_i for $i=1,\ldots,n$.

Proof.

independent $\Rightarrow$ (1).
$$F_{\mathbf{X}}(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n)$$
$$= P(X_1 \in (-\infty, x_1], \ldots, X_n \in (-\infty, x_n])$$
$$= P(X_1 \in (-\infty, x_1]) \times \cdots \times P(X_n \in (-\infty, x_n])$$
$$= F_{X_1}(x_1) \times \cdots \times F_{X_n}(x_n)$$

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

---

independent $\Rightarrow$ (2).
$$p_{\mathbf{X}}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$
$$= P(X_1 \in \{x_1\}, \ldots, X_n \in \{x_n\})$$
$$= P(X_1 \in \{x_1\}) \times \cdots \times P(X_n \in \{x_n\})$$
$$= p_{X_1}(x_1) \times \cdots \times p_{X_n}(x_n)$$

(2) ⇒ (1).

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

$$= \sum_{\substack{(t_1,\ldots,t_n)\in\mathcal{X}\\ t_1\le x_1}} \cdots \sum_{\substack{(t_1,\ldots,t_n)\in\mathcal{X}\\ t_n\le x_n}} p_{X_1}(t_1)\times\cdots\times p_{X_n}(t_n)$$

$$= \sum_{\substack{(t_1,\ldots,t_n)\in\mathcal{X}\\ t_1\le x_1}} p_{X_1}(t_1)\times\cdots\times \sum_{\substack{(t_1,\ldots,t_n)\in\mathcal{X}\\ t_n\le x_n}} p_{X_n}(t_n) = F_{X_1}(x_1)\times\cdots\times F_{X_n}(x_n)$$

(3) ⇒ (1).

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

$$= \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_{X_1}(t_1)\times\cdots\times f_{X_n}(t_n)\,dt_1\cdots dt_n$$

$$= \int_{-\infty}^{x_1} f_{X_1}(t_1)\,dt_1\times\cdots\times\int_{-\infty}^{x_n} f_{X_n}(t_n)\,dt_n = F_{X_1}(x_1)\times\cdots\times F_{X_n}(x_n)$$

(3) ⇐ (1).

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

$$= \frac{\partial^n}{\partial x_1\cdots\partial x_n}F_{X_1}(x_1)\times\cdots\times F_{X_n}(x_n)$$

$$= \frac{\partial}{\partial x_1}F_{X_1}(x_1)\times\cdots\times\frac{\partial}{\partial x_n}F_{X_n}(x_n) = f_{X_1}(x_1)\times\cdots\times f_{X_n}(x_n)$$

➢Remark. It follows from the Multiplication Law (LNp.3-7)

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

that the independence can be established *sequentially*.

➢Example. If A_1, …, A_n are independent events, then $\mathbf{1}_{A_1},\ldots,\mathbf{1}_{A_n}$, are independent random variables. For example,

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

➢Example. If $\mathbf{X}=(X_1, \ldots, X_n)$ are independent and

$$Y_i = g_i(X_i),\ i=1, \ldots, n,$$

then $Y_1, \ldots, Y_n$ are independent.

generalization

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

$$\begin{aligned} Y_1 &= g_1(X_1,\ldots,X_{i_1}),\\ Y_2 &= g_2(X_{i_1+1},\ldots,X_{i_2}),\\ &\cdots\\ Y_k &= g_k(X_{i_{k-1}+1},\ldots,X_{i_k}). \end{aligned}$$

Proof. Let $A_i(y) = \{x : g_i(x) \le y\}$, i=1, …, n, then

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

- Theorem. $\mathbf{X}=(X_1, \ldots, X_n)$ are independent if and only if there exist univariate functions $g_i(x)$, i=1, …, n, such that
  - (a) when $X_1, \ldots, X_n$ are discrete r.v.'s with joint pmf $p_{\mathbf{X}}$,
    $p_{\mathbf{X}}(x_1, \ldots, x_n) \propto g_1(x_1)\times\cdots\times g_n(x_n)$, $-\infty<x_i<\infty$, $i=1,\ldots,n$.
  - (b) when $X_1, \ldots, X_n$ are continuous r.v.'s with joint pdf $f_{\mathbf{X}}$,
    $f_{\mathbf{X}}(x_1, \ldots, x_n) \propto g_1(x_1)\times\cdots\times g_n(x_n)$, $-\infty<x_i<\infty$, $i=1,\ldots,n$.

Sketch of proof for (b).

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

Similarly, $f_{X_2}(x_2) \propto g_2(x_2), \ldots, f_{X_n}(x_n) \propto g_n(x_n)$

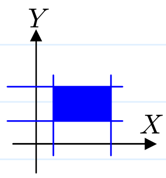
$$\begin{aligned} &\Rightarrow\ f_{X_1}(x_1)\cdots f_{X_n}(x_n) \propto g_1(x_1)\cdots g_n(x_n) \\ &\Rightarrow\ f_{\mathbf{X}}(x_1,\ldots,x_n) \propto f_{X_1}(x_1)\cdots f_{X_n}(x_n) \\ &\Rightarrow\ f_{\mathbf{X}}(x_1,\ldots,x_n) = c\cdot f_{X_1}(x_1)\cdots f_{X_n}(x_n) \\ &\qquad \text{for some constant } c. \end{aligned}$$

Because

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1,x_2,\ldots,x_n)\,dx_1\cdots dx_n = 1, \text{ and}$$

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1}(x_1)\cdots f_{X_n}(x_n)\,dx_1\cdots dx_n = 1, \quad \Rightarrow \boldsymbol{c=1}.$$

➢Example.



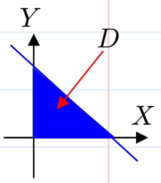
- If the joint pdf of (X, Y) is

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

and $f(x, y)$=0, otherwise, i.e.,

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

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

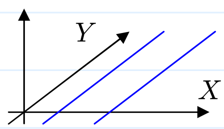


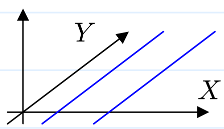
- Suppose that the joint pdf of (X, Y) is

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

and $f(x, y)$=0, otherwise, i.e., $f(x,y) \propto xy \cdot \mathbf{1}_D(x,y)$,
X and Y are not independent.

➢ **Q**: For independent X and Y, how should their joint pdf/pmf look like?



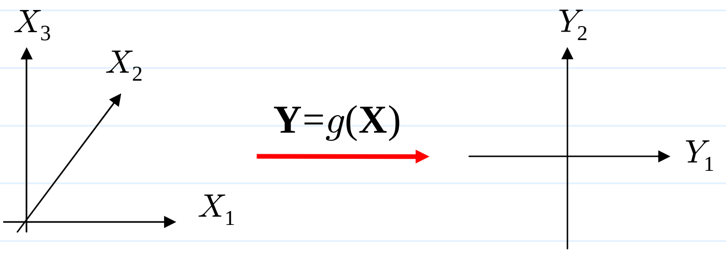


❖ **Reading**: textbook, Sec 6.2

Transformation

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

$$Y_1=g_1(X_1, \ldots, X_n),$$
$$\ldots,$$
$$Y_k=g_k(X_1, \ldots, X_n),$$



denoted by $\mathbf{Y}=\boldsymbol{g}(\mathbf{X})$, $\boldsymbol{g}:\mathbb{R}^n \to \mathbb{R}^k$.

➢ The following methods are useful:

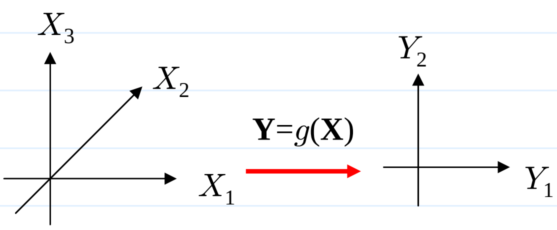
1. Method of Events
2. Method of Cumulative Distribution Function
3. Method of Probability Density Function
4. Method of Moment Generating Function (chapter 7)

➢ Method of Events

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

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

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



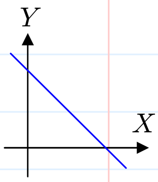
- Example. Let $\mathbf{X}$ be a discrete random vector taking values $\mathbf{x}_i=(x_{1i}, x_{2i}, \ldots, x_{ni})$, i=1, 2, …, with joint pmf $p_{\mathbf{X}}$. Then, $\mathbf{Y}=\boldsymbol{g}(\mathbf{X})$ is also a discrete random vector. Suppose that $\mathbf{Y}$ takes values on $\mathbf{y}_j$, j= 1, 2, …. To determine the joint pmf of $\mathbf{Y}$, by taking $B=\{\mathbf{y}_j\}$, we have

$$A = \{\mathbf{x}_i : \boldsymbol{g}(\mathbf{x}_i) = \mathbf{y}_j\}$$

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

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

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



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

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

  - When X and Y are independent,

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

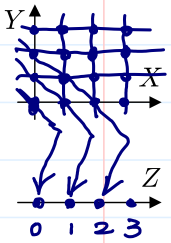
So,

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p_X(x)p_Y(z-x).$$

which is referred to as the *convolution* of p_X and p_Y.

  - (Exercise) $Z=X-Y$

- Theorem. If X and Y are independent, and $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$, then



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

Proof. For $z=0, 1, 2, \ldots$, the pmf $p_Z(z)$ of Z is

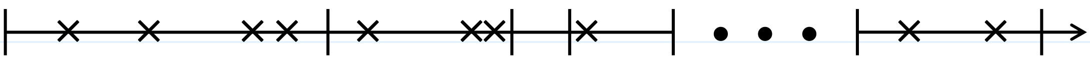
$$p_Z(z) = \sum_{x=0}^{z} p_X(x)p_Y(z-x) = \sum_{x=0}^{z} \frac{e^{-\lambda_1}\lambda_1^x}{x!}\frac{e^{-\lambda_2}\lambda_2^{z-x}}{(z-x)!}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{z!}\left(\sum_{x=0}^{z} \frac{z!}{x!(z-x)!}\lambda_1^x\lambda_2^{z-x}\right) = \frac{e^{-(\lambda_1+\lambda_2)}}{z!}(\lambda_1+\lambda_2)^z.$$

  - Corollary. If $X_1, \ldots, X_n$ are independent, and $X_i \sim \text{Poisson}(\lambda_i)$, $i=1, \ldots, n$, then

$$X_1+\cdots+X_n \sim \text{Poisson}(\lambda_1+\cdots+\lambda_n).$$

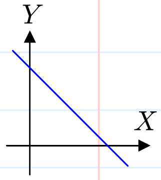
  Proof. By induction (exercise).

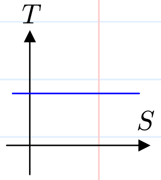


➢ Method of cumulative distribution function

1. In the $(X_1, \ldots, X_n)$ space, find the region that corresponds to $\{Y_1 \le y_1, \ldots, Y_k \le y_k\}$.
2. Find $F_{\mathbf{Y}}(y_1, \ldots, y_k) = P(Y_1 \le y_1, \ldots, Y_k \le y_k)$ by summing the joint pmf or integrating the joint pdf of $X_1, \ldots, X_n$ over the region identified in 1.
3. (for continuous case) Find the joint pdf of $\mathbf{Y}$ by differentiating $F_{\mathbf{Y}}(y_1, \ldots, y_k)$, i.e.,

$$f_{\mathbf{Y}}(y_1, \ldots, y_k) = \frac{d^k}{dy_1 \cdots dy_k} F_{\mathbf{Y}}(y_1, \ldots, y_k).$$





- Example. X and Y are random variables with joint pdf $f(x, y)$. Find the distribution of $Z=X+Y$.
  - $\{Z \le z\} = \{(X, Y) \in \{(x, y): x+y \le z\}\}$. So,

$$F_Z(z) = P(Z \le z) = P(X+Y \le z)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)\, dydx$$
$$= \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(s,t-s)\, dsdt \quad \left(\text{set } \begin{cases} x = s \\ y = t-s \end{cases}\right)$$

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

  - When X and Y are independent,

$$f(x, y) = f_X(x) f_Y(y).$$

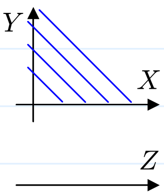
So, $F_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_X(x) f_Y(y)\, dydx$
$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-x} f_Y(y)\, dy\right] f_X(x)\, dx$$
$$= \int_{-\infty}^{\infty} F_Y(z-x) f_X(x)\, dx$$

which is referred to as the *convolution* of F_X and F_Y, and

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

which is referred to as the *convolution* of f_X and f_Y.

  - (exercise) $Z=X-Y$.



- Theorem. If X and Y are independent, and $X \sim \text{Gamma}(\alpha_1, \lambda)$, $Y \sim \text{Gamma}(\alpha_2, \lambda)$, then

$$Z = X + Y \sim \text{Gamma}(\alpha_1+\alpha_2, \lambda).$$

Proof. For $z \ge 0$,

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

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

  - Corollary. If $X_1, \ldots, X_n$ are independent, and $X_i \sim \text{Gamma}(\alpha_i, \lambda)$, $i=1, \ldots, n$, then

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

Proof. By induction (exercise).

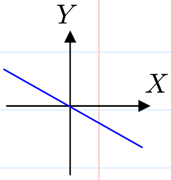
  - (exercise) Corollary. If $X_1, \ldots, X_n$ are independent, and $X_i \sim \text{Exponential}(\lambda)$, $i=1, \ldots, n$, then

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

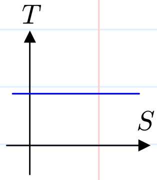
- (exercise) Theorem. If $X_1, \ldots, X_n$ are independent, and $X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$, $i=1, \ldots, n$, then

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

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



  - Let
$$\begin{aligned} Q_z &= \{(x,y) : y/x \le z\} \\ &= \{(x,y) : x<0, y \ge zx\} \\ &\qquad \cup \{(x,y) : x>0, y \le zx\} \end{aligned}$$



then,
$$\begin{aligned} F_Z(z) &= \int\int_{Q_z} f(x,y)\,dxdy \\ &= \int_{-\infty}^{0}\int_{xz}^{\infty} + \int_{0}^{\infty}\int_{-\infty}^{xz} f(x,y)\,dydx \\ &= \int_{-\infty}^{0}\int_{z}^{-\infty} + \int_{0}^{\infty}\int_{-\infty}^{z} s f(s,st)\,dtds \\ &= \int_{-\infty}^{0}\int_{-\infty}^{z} (-s) f(s,st)\,dtds \\ &\quad + \int_{0}^{\infty}\int_{-\infty}^{z} s f(s,st)\,dtds \\ &= \int_{-\infty}^{z}\int_{-\infty}^{\infty} |s| f(s,st)\,dsdt \end{aligned}$$

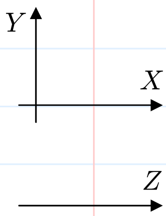
$$\left(\text{set } \begin{cases} x = s \\ y = st \end{cases}\right)$$

$(= \int_{-\infty}^{z}\int_{-\infty}^{\infty} |s| f_X(s) f_Y(st)\,dtds$ when X and Y are independent$)$

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

$(= \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz)\,dx$ when X and Y are independent$)$

  - (exercise) $Z=XY$
  - If X and Y are independent, $X \sim \text{exponential}(\lambda_1)$, $Y \sim \text{exponential}(\lambda_2)$, and $Z=Y/X$. The pdf of Z is



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

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

And, the cdf of Z is

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

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

➢Method of probability density function

- Theorem. Let $\mathbf{X}=(X_1, \ldots, X_n)$ be continuous random variables with the joint pdf $f_{\mathbf{X}}$. Let

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

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

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

Assume w have continuous partial derivatives, and let

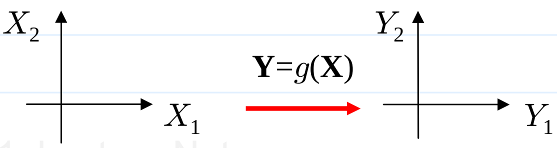
$$J = \begin{vmatrix} \frac{\partial w_1(\mathbf{y})}{\partial y_1} & \frac{\partial w_1(\mathbf{y})}{\partial y_2} & \cdots & \frac{\partial w_1(\mathbf{y})}{\partial y_n} \\ \frac{\partial w_2(\mathbf{y})}{\partial y_1} & \frac{\partial w_2(\mathbf{y})}{\partial y_2} & \cdots & \frac{\partial w_2(\mathbf{y})}{\partial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

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 $|J|$?)



Proof.

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

It then follows from an exercise in advanced calculus that

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

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

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

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

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

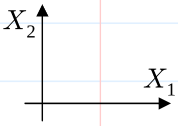
Therefore, $f_{\mathbf{Y}}(y_1, y_2) = f_{\mathbf{X}}(y_1 y_2, y_2 - y_1 y_2)|y_2|$,

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

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

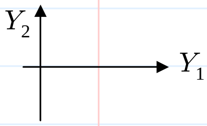
$(= \int_{-\infty}^{\infty} f_{X_1}(y_1 y_2) f_{X_2}(y_2 - y_1 y_2)|y_2|\, dy_2$ when X_1 and X_2 are independent$)$

- Theorem. If X_1 and X_2 are independent, and $X_1 \sim \text{Gamma}(\alpha_1, \lambda)$, $X_2 \sim \text{Gamma}(\alpha_2, \lambda)$, 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



$$f_{\mathbf{X}}(x_1, x_2) = \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} x_1^{\alpha_1 - 1} e^{-\lambda x_1} \times \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)} x_2^{\alpha_2 - 1} e^{-\lambda x_2}$$
$$= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} e^{-\lambda(x_1+x_2)}.$$

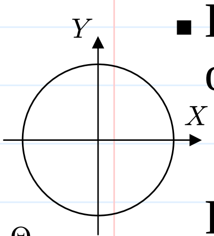
So, for $0 \leq y_1 \leq 1$,

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{X_1}(y_1 y_2) f_{X_2}(y_2 - y_1 y_2)|y_2|\, dy_2$$

$$= \int_0^{\infty} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} (y_1 y_2)^{\alpha_1 - 1} (y_2 - y_1 y_2)^{\alpha_2 - 1} e^{-\lambda y_2} \cdot y_2\, dy_2$$
$$= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1 - 1}(1-y_1)^{\alpha_2 - 1}$$
$$\times \int_0^{\infty} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} y_2^{(\alpha_1+\alpha_2)-1} e^{-\lambda y_2}\, dy_2.$$

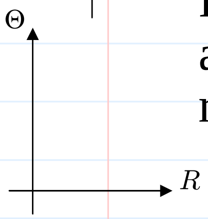
and $f_{Y_1}(y_1) = 0$, otherwise.

- 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.,



$$X = R\cos(\Theta) \equiv w_1(R, \Theta),$$
$$Y = R\sin(\Theta) \equiv w_2(R, \Theta).$$

- Since $\frac{\partial w_1}{\partial r} = \cos(\theta)$, $\frac{\partial w_1}{\partial \theta} = -r\sin(\theta)$,
$\frac{\partial w_2}{\partial r} = \sin(\theta)$, $\frac{\partial w_2}{\partial \theta} = r\cos(\theta)$,

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

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

- For $0 \le r \le 1$ and $0 \le \theta \le 2\pi$, 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.

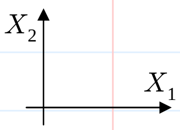
- By the theorem in LNp.6-21, (R, Θ) are independent.

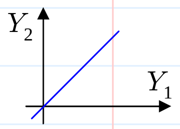
- Example. Let $X_1, \ldots, X_n$ be independent and identically distributed exponential(λ). Let

$$Y_i = X_1 + \cdots + X_i,\ i = 1, \ldots, n.$$

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

(Note. It has been shown that $Y_i \sim$ Gamma(i, λ), i=1, …, n.)





- The joint pdf of $X_1, \ldots, X_n$ is

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

for $0 \le x_i < \infty$, i=1, …, n.

- Since

$$\begin{aligned} x_1 &= y_1 \equiv w_1(y_1, \ldots, y_n), \\ x_2 &= y_2 - y_1 \equiv w_2(y_1, \ldots, y_n), \\ &\cdots \\ x_n &= y_n - y_{n-1} \equiv w_n(y_1, \ldots, y_n), \end{aligned}$$

we have

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

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

- For $0 \le y_1 \le \cdots \le y_{i-1} \le y_i \le y_{i+1} \le \cdots \le y_n < \infty$,

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

and $f_{\mathbf{Y}}(y_1, \ldots, y_n) = 0$, otherwise.

- The marginal pdf of Y_i is

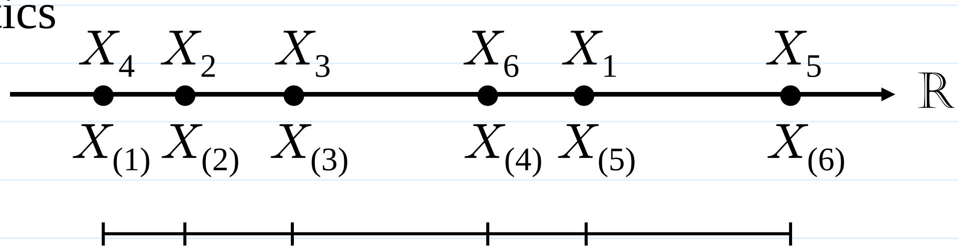
$$\begin{aligned} & f_{Y_i}(y) \\ &= \int_0^y \int_{y_1}^y \cdots \int_{y_{i-2}}^y \int_y^\infty \int_{y_{i+1}}^\infty \cdots \int_{y_{n-1}}^\infty \lambda^n e^{-\lambda y_n}\, dy_n \cdots dy_{i+2} dy_{i+1} dy_{i-1} \cdots dy_2 dy_1 \\ &= \int_0^y \cdots \int_{y_{i-2}}^y \lambda^i e^{-\lambda y}\, dy_{i-1}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

for $y \ge 0$, and $f_{Y_i}(y) = 0$, 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 $X_1, \ldots, X_n$ be random variables. We sort the X_i's and denote by $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ the *order statistics*. Using the notation,

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

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

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

$S_j \equiv X_{(j)} - X_{(j-1)}$, j=2, …, n, are called *spacings*.

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

➢ Definition. $X_1, \ldots, X_n$ are called *i.i.d.* (**i**ndependent, **i**dentically **d**istributed) with cdf F/pdf f/pmf p if random variables $X_1, \ldots, X_n$ are independent and have a common marginal distribution with cdf F/pdf f/pmf p.

- Remark. For order statistics, we only consider the case that $X_1, \ldots, X_n$ are i.i.d.
  - Note. Although $X_1, \ldots, X_n$ are independent, their order statistics $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ are not independent in general.

➢ Theorem. Suppose that $X_1, \ldots, X_n$ 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 **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,

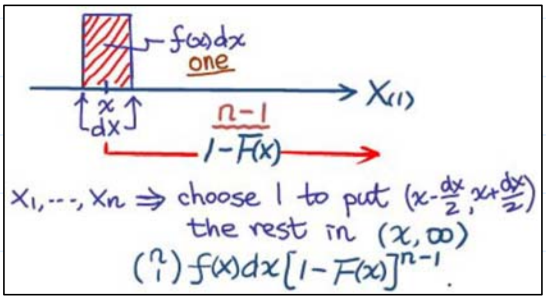
$$\begin{aligned} 1 - F_{X_{(1)}}(x) &= P(X_{(1)} > x) = P(X_1 > x, \ldots, X_n > x) \\ &= P(X_1 > x) \cdots P(X_n > x) = [1 - F(x)]^n. \end{aligned}$$

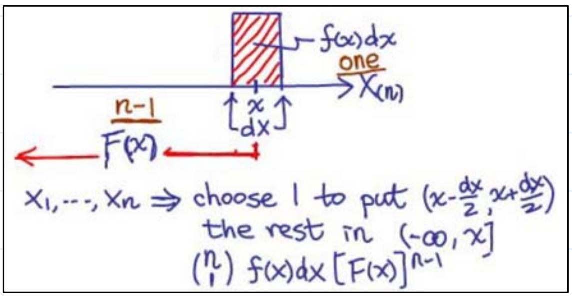
$$F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(X_1 \le x, \ldots, X_n \le x)$$
$$= P(X_1 \le x) \cdots P(X_n \le x) = [F(x)]^n.$$

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

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

- 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 $X_1, \ldots, X_n$, 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
$$Y = \max(X_1, \ldots, X_n).$$
  - If n=5 and F is exponential with $\lambda = 1$ per month, then

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

  - The cdf of Y is
$$F_Y(y) = (1 - e^{-y})^5, \text{ for } y \ge 0, \text{ and } 0, \text{ for } y < 0,$$
and its pdf is $5(1 - e^{-y})^4 e^{-y}$, for $y \ge 0$, and 0, for $y < 0$.
  - The probability that the room is still lighted after two months is $P(Y > 2) = 1 - F_Y(2) = 1 - (1 - e^{-2})^5$.

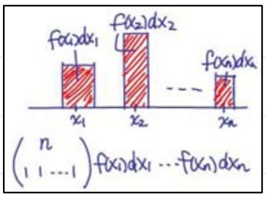
➢ Theorem. Suppose that $X_1, \ldots, X_n$ are i.i.d. with pdf f/pmf p. Then, the joint pmf/pdf of $X_{(1)}, \ldots, X_{(n)}$ is

$$p_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n! \times p(x_1) \times \cdots \times p(x_n),$$
$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, \ldots, x_n) = n! \times f(x_1) \times \cdots \times f(x_n),$$

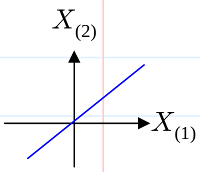
for $x_1 \le x_2 \le \cdots \le x_n$, and 0 otherwise.

Proof. For $x_1 \le x_2 \le \cdots \le x_n$,

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



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



- **Q**: Examine whether $X_{(1)}, \ldots, X_{(n)}$ are independent using the Theorem in LNp.6-21.

➢Theorem. If $X_1, \ldots, X_n$ are i.i.d. with cdf F and pdf f, then

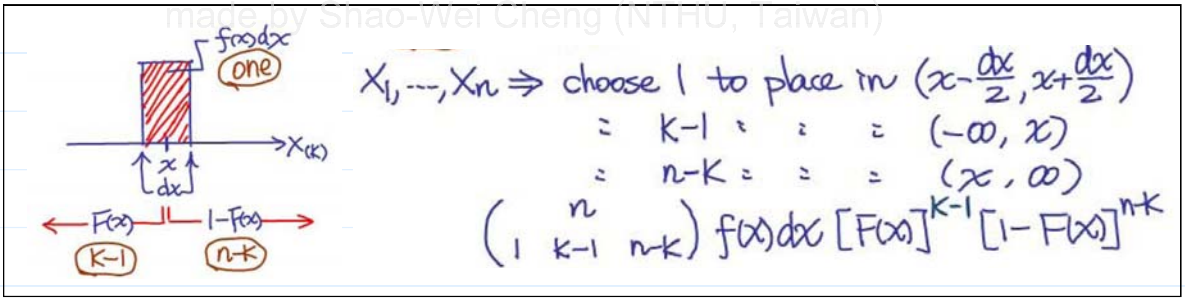
1.The pdf of the k^{th} order statistic $X_{(k)}$ is

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

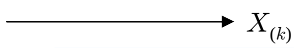
2.The cdf of $X_{(k)}$ is

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

Proof.



$$\begin{aligned} F_{X_{(k)}}(x) &= P(X_{(k)}\le x) \\ &= P(\text{at least } k \text{ of the } X_i\text{'s are} \le x) \\ &= \sum_{m=k}^{n} P(\text{exact } m \text{ of the } X_i\text{'s are} \le x) \\ &= \sum_{m=k}^{n}\binom{n}{m}[F(x)]^m[1-F(x)]^{n-m} \end{aligned}$$



➢Theorem. If $X_1, \ldots, X_n$ are i.i.d. with cdf F and pdf f, then

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

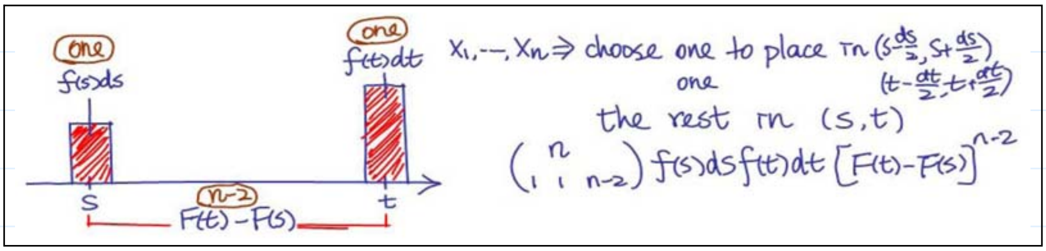
$$f_{X_{(1)},X_{(n)}}(s,t)=n(n-1)f(s)f(t)[F(t)-F(s)]^{n-2},$$

for $s \le t$, 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 $r \ge 0$, and 0 otherwise.



➢Theorem. If $X_1, \ldots, X_n$ are i.i.d. with cdf F and pdf f, then

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

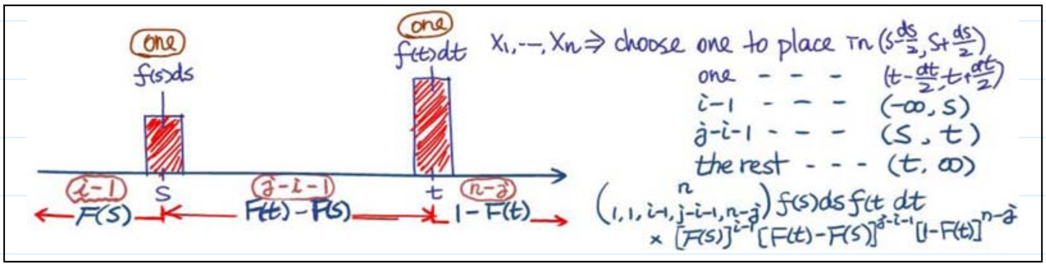
$$f_{X_{(i)},X_{(j)}}(s,t) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} f(s) f(t) \times [F(s)]^{i-1}[F(t)-F(s)]^{j-i-1}[1-F(t)]^{n-j},$$

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

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

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

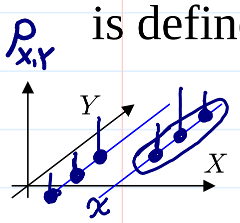
for $s \ge 0$, and zero otherwise.



❖ **Reading**: textbook, Sec 6.3, 6.6, 6.7

Conditional Distribution

- Definition. Let **X** and **Y** be discrete random vectors and (**X**, **Y**) have a joint pmf $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$, then the *conditional joint pmf* of **Y** given **X**=**x** is defined as



$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \equiv P(\{\mathbf{Y}=\mathbf{y}\}|\{\mathbf{X}=\mathbf{x}\}) = \frac{P(\{\mathbf{X}=\mathbf{x}, \mathbf{Y}=\mathbf{y}\})}{P(\{\mathbf{X}=\mathbf{x}\})} = \frac{p_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})}$$

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

➢Some Notes.

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

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

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

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

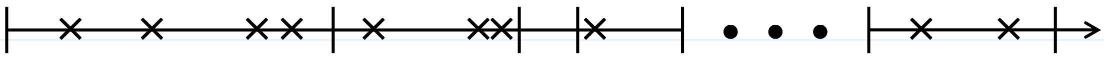
- The *conditional joint cdf* of **Y** given **X**=**x** can be similarly defined from the conditional joint pmf $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$, i.e.,

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

➢Theorem. Let $X_1, \ldots, X_m$ be independent and $X_i \sim$Poisson(λ_i), i=1, …, m. Let $Y=X_1+\cdots+X_m$, then

$$(X_1, \ldots, X_m | Y=n) \sim \text{Multinomial}(n, m, p_1, \ldots, p_m),$$

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



Proof. The joint pmf of $(X_1, \ldots, X_m, Y)$ is

$$p_{\mathbf{X},Y}(x_1,\ldots,x_m,n) = P(\{X_1=x_1,\ldots,X_m=x_m\}\cap\{Y=n\})$$

$$= \begin{cases} P(X_1=x_1,\ldots,X_m=x_m), & \text{if } x_1+\cdots+x_m=n, \\ 0, & \text{if } x_1+\cdots+x_m\neq n. \end{cases}$$

Furthermore, the distribution of Y is Poisson($\lambda_1+\cdots+\lambda_m$), i.e.,

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

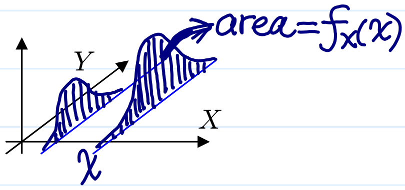
Therefore, for $\mathbf{x}=(x_1, \ldots, x_m)$ wheres $x_i\in\{0,1,2,\ldots\}$, i=1, …, m, and $x_1+\cdots+x_m=n$, the conditional joint pmf of $\mathbf{X}$ given $Y=n$ is

$$p_{\mathbf{X}|Y}(\mathbf{x}|n) = \frac{p_{\mathbf{X},Y}(x_1,\ldots,x_m,n)}{p_Y(n)} = \frac{\prod_{i=1}^m \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}}{\frac{e^{-(\lambda_1+\cdots+\lambda_m)}(\lambda_1+\cdots+\lambda_m)^n}{n!}}$$

$$= \frac{n!}{x_1!\times\cdots\times x_m!}\times\left(\frac{\lambda_1}{\lambda_1+\cdots+\lambda_m}\right)^{x_1}\times\cdots\times\left(\frac{\lambda_m}{\lambda_1+\cdots+\lambda_m}\right)^{x_m}.$$

- Definition. Let **X** and **Y** be continuous random vectors and (**X**, **Y**) have a joint pdf $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$, then the *conditional joint pdf* of **Y** given **X**=**x** is defined as

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \equiv \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})},$$



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

➢Some Notes.

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

$$P\left(\mathbf{Y}\leq\mathbf{y}\,|\,\mathbf{x}-(\Delta\mathbf{x}/2)<\mathbf{X}\leq\mathbf{x}+(\Delta\mathbf{x}/2)\right)$$

$$= \frac{\int_{-\infty}^{\mathbf{y}}\int_{\mathbf{x}-\frac{\Delta\mathbf{x}}{2}}^{\mathbf{x}+\frac{\Delta\mathbf{x}}{2}} f_{\mathbf{X},\mathbf{Y}}(\mathbf{u},\mathbf{v})\,d\mathbf{u}d\mathbf{v}}{\int_{\mathbf{x}-\frac{\Delta\mathbf{x}}{2}}^{\mathbf{x}+\frac{\Delta\mathbf{x}}{2}} f_{\mathbf{X}}(\mathbf{t})\,d\mathbf{t}}$$

$$\approx \frac{\int_{-\infty}^{\mathbf{y}} f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{v})\Delta\mathbf{x}\,d\mathbf{v}}{f_{\mathbf{X}}(\mathbf{x})\Delta\mathbf{x}} = \int_{-\infty}^{\mathbf{y}} \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})}\,d\mathbf{y}$$

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

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

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

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

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

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

➢Example. If X and Y have a joint pdf

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

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

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

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

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

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

- Mixed Distribution: The 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

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

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 $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})$/or a joint 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 X ~ Uniform(0, 1), and $(Y_1, \ldots, Y_n|X=x)$ are i.i.d. with Bernoulli(x), i.e.,
$$p_{\mathbf{Y}|X}(y_1, \ldots, y_n|x) = x^{y_1+\cdots+y_n}(1-x)^{n-(y_1+\cdots+y_n)},$$
for $y_1, \ldots, y_n \in \{0, 1\}$.
- By the multiplication law, for $y_1, \ldots, y_n \in \{0, 1\}$ and $0<x<1$,
$$p_{\mathbf{Y},X}(y_1, \ldots, y_n, x) = x^{y_1+\cdots+y_n}(1-x)^{n-(y_1+\cdots+y_n)}.$$
- Suppose that we observed $Y_1=1, \ldots, Y_n=1$. By the law of total probability, $P(Y_1 = 1, \ldots, Y_n = 1) = p_{\mathbf{Y}}(1, \ldots, 1)$
$$= \int_0^1 p_{\mathbf{Y}|X}(1, \ldots, 1|x) f_X(x)\, dx$$
$$= \int_0^1 x^n\, dx = \frac{1}{n+1}x^{n+1}\Big|_0^1 = \frac{1}{n+1}.$$

- And, by Bayes' Theorem,
$$f_{X|\mathbf{Y}}(x|Y_1 = 1, \ldots, Y_n = 1)$$
$$= \frac{p_{\mathbf{Y}|X}(1, \ldots, 1|x) f_X(x)}{p_{\mathbf{Y}}(1, \ldots, 1)} = (n+1)x^n.$$
for $0<x<1$, i.e., $(X|Y_1=1, \ldots, Y_n=1)$ ~ Gamma(n+1, 1).
- If there were an (n+1)st Bernoulli trial Y_{n+1},
$$P(Y_{n+1} = 1|Y_1 = 1, \ldots, Y_n = 1)$$
$$= \frac{P(Y_1 = 1, \ldots, Y_{n+1} = 1)}{P(Y_1 = 1, \ldots, Y_n = 1)} = \frac{1/(n+2)}{1/(n+1)} = \frac{n+1}{n+2}.$$
- (exercise) In general, it can be shown that
$(X|Y_1=y_1, \ldots, Y_n=y_n)$ ~ Gamma($(y_1+\cdots+y_n)$+1, n–$(y_1+\cdots+y_n)$+1).

• Theorem (Independent). Let **X** and **Y** be random vectors and (**X**, **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, **X** and **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

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

Proof. By the definition of conditional distribution.

- intuition.
  - the 2 graphs in LNp.6-23
  - $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ (or $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$) offers information of $\mathbf{Y}$ when $\mathbf{X} = \mathbf{x}$; $p_{\mathbf{Y}}(\mathbf{y})$ (or $f_{\mathbf{Y}}(\mathbf{y})$) offers information of $\mathbf{Y}$ when $\mathbf{X}$ not observed.

❖ **Reading**: textbook, Sec 6.4, 6.5