I225E Statistical Signal Processing

2. Review of Probability Theory

MAWALIM and UNOKI

candylim@jaist.ac.jp and unoki@jaist.ac.jp

School of Information Science

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY,
HOKURIKU 1990

Review of probability theory

- Probability theory
  Sample space, Borel set, conditional probability, Bayes' theorem
- Random variables
- Distribution functions, density functions
- Joint distributions
- Moments
- Characteristic functions
- Law of large numbers
- Central limit theorem

1. Review of Probability

All possible outcomes that may result from a trial (experiment or observation) are known ***a priori***.

However, it is impossible to predict which outcome to occur.

- Trial ***S***: Doing experiment or observation
- Sample Point ω: Individual outcome that results from each trial
- Sample Space $\boldsymbol{\Omega}$: All sets of sample points
- Event ***A***: Subset of sample space

When $\omega \in \boldsymbol{\Omega}$ for $\omega \in \boldsymbol{A}$, we say event A took place.

[Example]

- Trial ***S***: Throw a dice
- Sample Point ω : 1, 2, 3, …
- Sample Space $\boldsymbol{\Omega}$: {1, 2, 3, 4, 5, 6}
- Event ***A***: Odd {1,3,5} and Even {2,4,6}

- Trial ***S***: Twice coin-toss
- Sample Point ω : head (h) or tail (t)
- Sample Space $\boldsymbol{\Omega}$: {hh, ht, th, tt}
- Event ***A***: only one head showed

Event

(E1) Complementary Event: $\boldsymbol{A}^C = \{\omega \in \boldsymbol{\Omega}: \omega \notin \boldsymbol{A}\}$

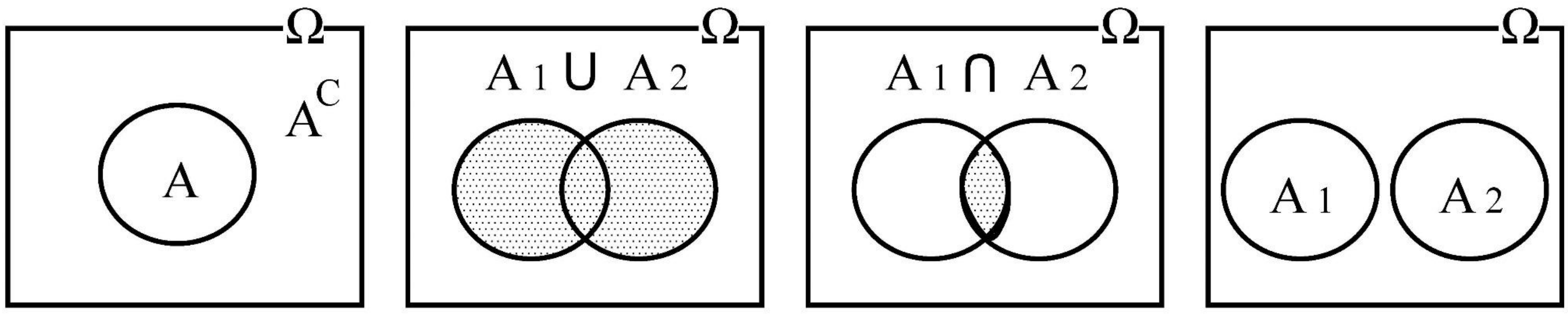
(E2) Sum Event (Union, OR):

$$\boldsymbol{A}_1 \cup \boldsymbol{A}_2 = \{\omega \in \boldsymbol{\Omega}: \omega \in \boldsymbol{A}_1 \text{ or } \omega \in \boldsymbol{A}_2\}$$ ("$\boldsymbol{A}_1$ or $\boldsymbol{A}_2$")

(E3) Product Event (Intersection, AND):

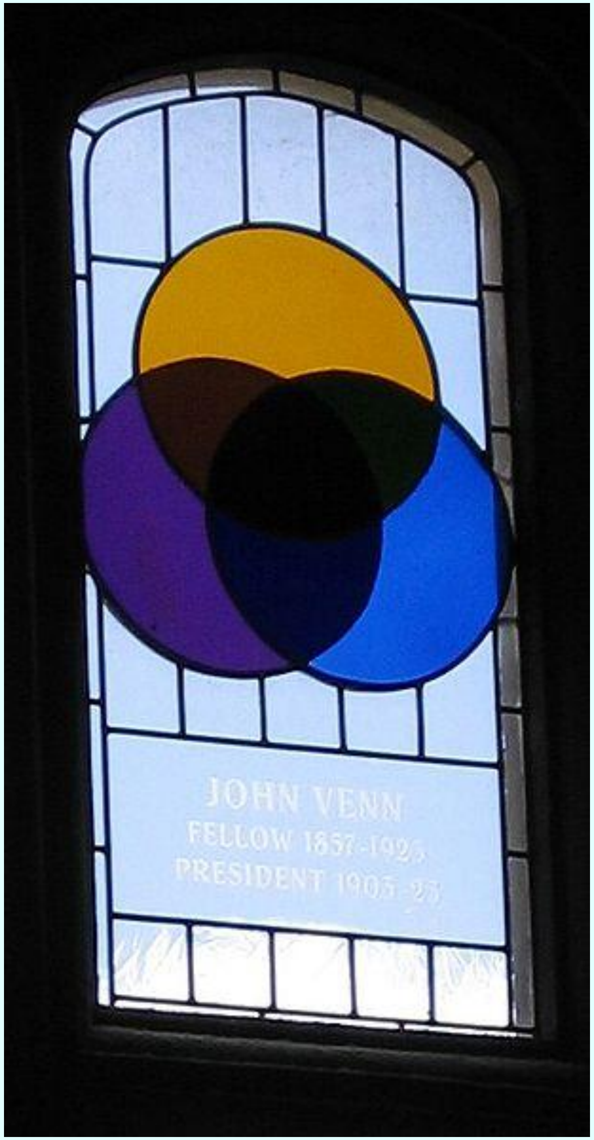
$$\boldsymbol{A}_1 \cap \boldsymbol{A}_2 = \{\omega \in \boldsymbol{\Omega}: \omega \in \boldsymbol{A}_1 \text{ and } \omega \in \boldsymbol{A}_2\}$$ ("$\boldsymbol{A}_1$ and $\boldsymbol{A}_2$")

(E4) Exclusive Event: $\boldsymbol{A}_1 \cap \boldsymbol{A}_2 = \boldsymbol{0}$



(1) (2) (3) (4)

Draw Venn diagrams!



Borel set: *B*

A Borel set ***B*** is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement.

(B1) $\boldsymbol{\Omega} \in \boldsymbol{B}$.

(B2) If $\boldsymbol{A} \in \boldsymbol{B}$ then $\boldsymbol{A}^C \in \boldsymbol{B}$.

(B3) If $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots \in \boldsymbol{B}$ then $\bigcup_{i=1}^{\infty} \boldsymbol{A}_i \; (= \boldsymbol{A}_1 \cup \boldsymbol{A}_2 \cup \cdots) \in \boldsymbol{B}$.

From (B1)-(B3),

(B4) $\boldsymbol{0} \in \boldsymbol{B}$.

(B5) If $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots \in \boldsymbol{B}$, then $\bigcap_{i=1}^{\infty} \boldsymbol{A}_i \; (= \boldsymbol{A}_1 \cap \boldsymbol{A}_2 \cap \cdots) \in \boldsymbol{B}$.

Borel set: Example

- Dice throwing

$$\boldsymbol{\Omega} = \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}\}$$

$$\begin{aligned}\boldsymbol{B} = \{&\emptyset, \{\mathbf{1}\}, \cdots, \{\mathbf{6}\},\\ &\{\mathbf{1}, \mathbf{2}\}, \cdots \{\mathbf{5}, \mathbf{6}\},\\ &\{\mathbf{1}, \mathbf{2}, \mathbf{3}\}, \cdots, \{\mathbf{4}, \mathbf{5}, \mathbf{6}\},\\ &\{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}\}, \{\mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}\},\\ &\{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}\}, \cdots \{\mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}\},\\ &\{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}\}\}\end{aligned}$$

Borel set = “a set of all possible events”

Probability

- Probability of event $\boldsymbol{A}$: $P(\boldsymbol{A})$
  mapping from event $\boldsymbol{A}$ to some numbers $P(\boldsymbol{A})$.
- **Properties of** $P(\boldsymbol{A})$ (Axioms of Probability theory)

  (P1) $0 \leq P(\boldsymbol{A}) \leq 1$

  (P2) $P(\boldsymbol{\Omega}) = 1$

  (P3) If $\boldsymbol{A_1}, \boldsymbol{A_2}, \ldots$ are exclusive events, then

  $P(\cup_{i=1}^{\infty} \boldsymbol{A}_i) = \sum_{i=1}^{\infty} P(\boldsymbol{A}_i)$ (Complete Additiveness)

- **Definition of Probability Space**:

  Sample space $\boldsymbol{\Omega}$, Borel set ***B***, and probability *P* define probability space.

Basic properties

(P4) $P(\mathbf{0}) = 0$

[Proof] Assuming $\boldsymbol{A}_i = \mathbf{0}\ (i = 1, 2, \cdots)$, then

$$\bigcup_{i=1}^{\infty} \boldsymbol{A}_i = \bigcup_{i=1}^{\infty} \mathbf{0} = \mathbf{0}.$$

On the other hand, from the assumption,

$$\boldsymbol{A}_i \cap \boldsymbol{A}_j = \mathbf{0} \cap \mathbf{0} = \mathbf{0}\ (i \neq \mathrm{j}).$$

According to Axiom (P3),

$$P(\mathbf{0}) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(\mathbf{0}).$$

Since $\mathbf{0} \in \boldsymbol{\Omega}$, $P(\mathbf{0}) \geq 0$ (due to Axiom (P1)). The above equation holds only if $P(\mathbf{0}) = 0$.

(P5) If events $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots$ are **mutually exclusive**, then

$$P(\cup_{i=1}^{n} \boldsymbol{A}_i) = \sum_{i=1}^{n} P(\boldsymbol{A}_i).$$

[Proof] Setting $\boldsymbol{A}_{n+1} = \boldsymbol{A}_{n+2}, = \cdots = \mathbf{0}$,

$\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n, \boldsymbol{A}_{n+1}, \cdots$ are mutually exclusive.

Moreover, $\cup_{i=1}^{n} \boldsymbol{A}_i = \cup_{i=1}^{\infty} \boldsymbol{A}_i$.

According to Axiom (P3) and Property (P4),

$$P(\cup_{i=1}^{n} \boldsymbol{A}_i) = P(\cup_{i=1}^{\infty} \boldsymbol{A}_i)$$

$$= \sum_{i=1}^{n} P(\boldsymbol{A}_i) + \sum_{i=n+1}^{\infty} P(\boldsymbol{A}_i)$$

$$= \sum_{i=1}^{n} P(\boldsymbol{A}_i) + \sum_{i=n+1}^{\infty} P(\mathbf{0}) = \sum_{i=1}^{n} P(\boldsymbol{A}_i)$$

(P6) $P(\boldsymbol{A}^C) = 1 - P(\boldsymbol{A})$

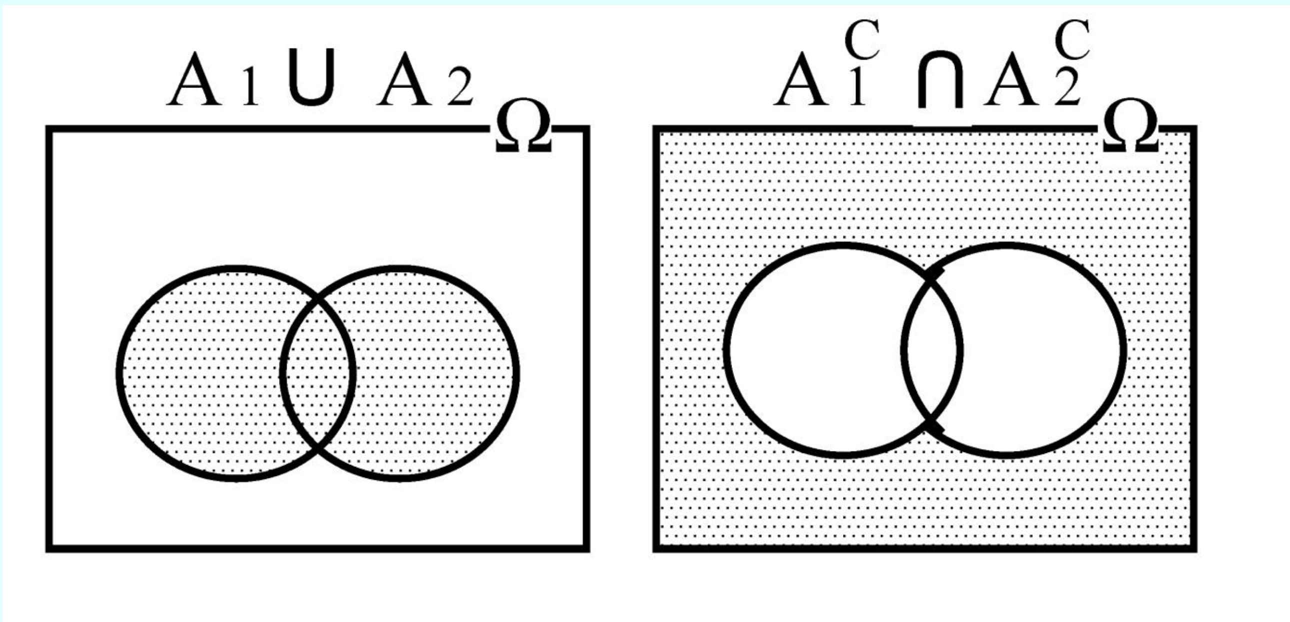
(P7) $P(\cup_i \boldsymbol{A}_i) = 1 - P(\cap_i \boldsymbol{A}_i^C)$

De Morgan's laws:

$$(\cap_i \boldsymbol{A}_i)^C = \cup_i \boldsymbol{A}_i^C,\ (\cup_i \boldsymbol{A}_i)^C = \cap_i \boldsymbol{A}_i^C$$

De Morgan's laws in case of 2 sets

$$(\boldsymbol{A}_1 \cup \boldsymbol{A}_2)^C = \boldsymbol{A}_1^C \cap \boldsymbol{A}_2^C$$

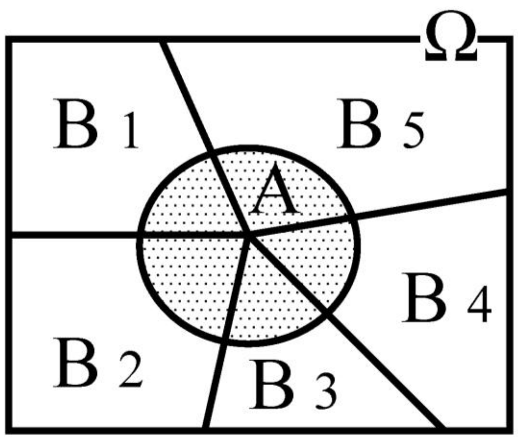


(P8) If sequence of mutually exclusive events, $\boldsymbol{B}_1, \boldsymbol{B}_2, \cdots$, is such that :

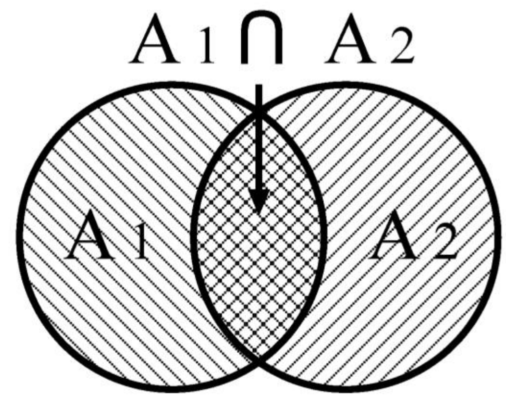
$\cup_i \boldsymbol{B}_i = \boldsymbol{\Omega}$, then $P(\boldsymbol{A}) = \sum_i P(\boldsymbol{A} \cap B_i)$.

(P9) If $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are not exclusive, then $P(\boldsymbol{A}_1 \cup \boldsymbol{A}_2) = P(\boldsymbol{A}_1) + P(\boldsymbol{A}_2) - P(\boldsymbol{A}_1 \cap \boldsymbol{A}_2)$.

(8)



(9)



(P10) For any $\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3$,

$$\begin{aligned}
&\mathrm{P}(\boldsymbol{A}_1 \cup \boldsymbol{A}_2 \cup \boldsymbol{A}_3) \\
&= P(\boldsymbol{A}_1) + P(\boldsymbol{A}_2) + P(\boldsymbol{A}_3) \\
&- \{P(\boldsymbol{A}_1 \cap \boldsymbol{A}_2) + P(\boldsymbol{A}_2 \cap \boldsymbol{A}_3) + P(\boldsymbol{A}_3 \cap \boldsymbol{A}_1)\} \\
&\qquad\qquad + P(\boldsymbol{A}_1 \cap \boldsymbol{A}_2 \cap \boldsymbol{A}_3)
\end{aligned}$$

(P11) General case: Denoting

$S_m = \sum_{i_1 \leq i_2 \leq \cdots \leq i_m} P(\boldsymbol{A}_{i_1} \cap \boldsymbol{A}_{i_2} \cap \cdots \cap \boldsymbol{A}_{i_m})$, then
$P(\bigcup_{i=1}^{n} \boldsymbol{A}_i) = S_1 - S_2 + S_3 - S_4 + \cdots + (-1)^{n-1} S_n$.

Note: Additive terms on the right side are subject to all combination of S_n choosing *n* events from $1, 2, \cdots, n$.
The combination number is ${}_nC_m = \binom{n}{m} = \frac{n!}{(n-m)!m!}$.

[Exercise] Consider an experiment of drawing two cards at random from a bag containing four cards marked with the integers 1 through 4.

(a) Find the sample space S1 of the experiment if the first card is replaced before the second is drawn.

(b) Find the sample space S2 of the experiment if the first card is not replaced.

[Answer]

(a) The sample space S1 contains 16 ordered pairs (i, j), $1 \le i \le 4$, $1 \le j \le 4$, where the first number indicates the first number drawn. Thus,

$$S_1 = \begin{Bmatrix} (1,1) & (1,2) & (1,3) & (1,4) \\ (2,1) & (2,2) & (2,3) & (2,4) \\ (3,1) & (3,2) & (3,3) & (3,4) \\ (4,1) & (4,2) & (4,3) & (4,4) \end{Bmatrix}$$

[Answer]

(b) The sample space S2 contains 12 ordered pairs (i, j), i ≠ j, 1 ≤ i ≤ 4, 1 ≤ j ≤ 4, where the first number indicates the first number drawn. Thus,

$$S_2 = \begin{Bmatrix} (1,2) & (1,3) & (1,4) \\ (2,1) & (2,3) & (2,4) \\ (3,1) & (3,2) & (3,4) \\ (4,1) & (4,2) & (4,3) \end{Bmatrix}$$

[Exercise]

Given that $P(A) = 0.9$, $P(B) = 0.8$, and $P(A \cap B) = 0.75$, find

(a) $P(A \cup B)$

(b) $P(A \cap \overline{B})$

(c) $P(\overline{A} \cap \overline{B})$.

[Answer]

(a) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$= 0.9 + 0.8 - 0.75 = 0.95$$

(b) $P(A \cup \bar{B}) = P(A) - P(A \cap B)$

$$= 0.9 - 0.75 = 0.15$$

(c) By Morgan's Law:

$$P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B)$$

$$= 1 - 0.95 = 0.05$$

[Exercise]

Prove that

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

[Answer]

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

$$= P(C \cap A \cap B) = P(A \cap B \cap C)$$

Continuity of Probability

(P12) Consider infinite sequence of events $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots$ such that $\boldsymbol{A}_1 \subset \boldsymbol{A}_2 \subset \cdots$. For $\boldsymbol{A} = \bigcup_{i=1}^{\infty} \boldsymbol{A}_i$,

$$P(\boldsymbol{A}) = \lim_{i \to \infty} P(\boldsymbol{A}_i)$$

(P13) Consider infinite sequence of events $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots$ such that $\boldsymbol{A}_1 \supset \boldsymbol{A}_2 \supset \cdots$. For $\boldsymbol{A} = \bigcap_{\boldsymbol{i}=\boldsymbol{1}}^{\infty} \boldsymbol{A}_i$,

$$P(\boldsymbol{A}) = \lim_{i \to \infty} P(\boldsymbol{A}_i)$$

Conditional Probability

Conditional Probability

First event has an influence on probability of next event.

B: First event

A: Next event

Conditional probability $P(\boldsymbol{A}|\boldsymbol{B})$ of event ***A*** assuming event ***B*** is

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

Properties of Conditional Probability

(C1) If we fix $\boldsymbol{B}$ and denote as $P(\boldsymbol{A}|\boldsymbol{B}) \equiv P^*(\boldsymbol{A})$,

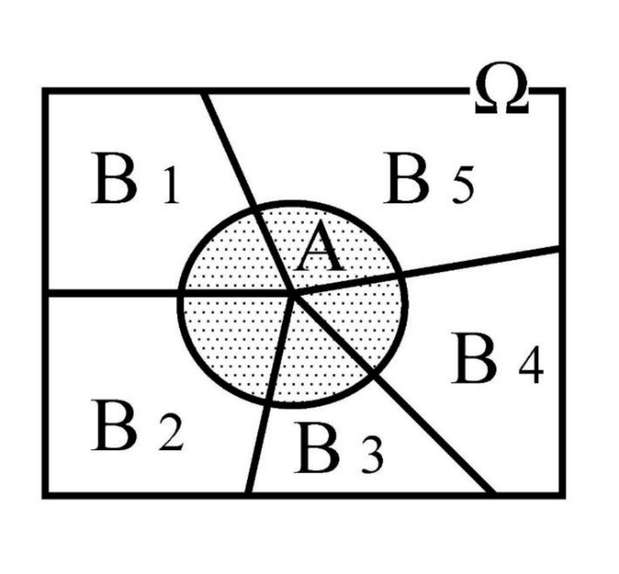
(P1) - (P13) hold for $P^*(\boldsymbol{A})$.

(C2) $P(\boldsymbol{A} \cap \boldsymbol{B}) = P(\boldsymbol{A}|\boldsymbol{B})P(\boldsymbol{B})$,

(C3) If events $\boldsymbol{B}_1, \boldsymbol{B}_2, \cdots$ are mutually exclusive and $\bigcup_i \boldsymbol{B}_i = \boldsymbol{\Omega}$, then, for any event $\boldsymbol{A}$,

$$P(\boldsymbol{A}) = \sum_i P(\boldsymbol{B}_i)P(\boldsymbol{A}|\boldsymbol{B}_i)$$

(Total Probability Theorem)



(C4) If $P(\boldsymbol{A}) \geq 0$,

$$P(\boldsymbol{B}_i|\boldsymbol{A}) = \frac{P(\boldsymbol{A}|\boldsymbol{B}_i)P(\boldsymbol{B}_i)}{\sum_j P(\boldsymbol{A}|\boldsymbol{B}_j)P(\boldsymbol{B}_j)}$$

(Bayes' Theorem)

[Exercise] Company C buys the same product from three companies (B_1, B_2, B_3). The purchase ratio is $B_1 = 0.5$, $B_2 = 0.3$, and $B_3 = 0.2$. Probability that a new product is broken within one year is given for each company as $B_1 = 0.015$, $B_2 = 0.025$, and $B_3 = 0.035$.

- Suppose now that one product is broken within one year of purchase. We denote this event by *A*. What is the probability that the broken product was purchased from company B_1, B_2, or B_3.

[Answer] $P(B_1) = 0.5, P(B_2) = 0.3, P(B_3) = 0.2$
$P(A|B_1) = 0.015, P(A|B_2) = 0.025,$
$P(A|B_3) = 0.035$

From Total Probability Theorem (C3), $P(A) = \sum_{i=1}^{3} P(A|B_i)P(B_i) = 0.0075 + 0.0075 + 0.007$
$= 0.022$

Using Bayes' Theorem (C4)

$$P(B_1|A) = \frac{0.0075}{0.022} \approx 0.34, P(B_2|A) = \frac{0.0075}{0.022} \approx 0.34$$

$$P(B_3|A) = \frac{0.0070}{0.022} \approx 0.32$$

Independency

(1) If events ***A*** and ***B*** are **mutually independent**, then

$$P(\boldsymbol{A} \cap \boldsymbol{B}) = P(\boldsymbol{A})P(\boldsymbol{B}).$$

(2) Necessary and sufficient condition for mutual independence of *n* events, $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n$, is that, for a set of arbitrarily chosen events, $\boldsymbol{A}_{i(1)}, \boldsymbol{A}_{i(2)}, \cdots, \boldsymbol{A}_{i(k)}$, the following holds

$$P\left(\boldsymbol{A}_{i(1)} \cap \boldsymbol{A}_{i(2)} \cap \cdots \cap \boldsymbol{A}_{i(k)}\right)$$
$$= P\left(\boldsymbol{A}_{i(1)}\right)P\left(\boldsymbol{A}_{i(2)}\right) \cdots P\left(\boldsymbol{A}_{i(k)}\right)$$

[Exercise]

Show that if three events A, B, and C are independent, then A and $(B \cup C)$ are independent.

[Exercise]

At a college:

54% of students are female

25% of students are majoring in engineering.

15% of female students are majoring in engineering.

Event E = student is majoring in engineering

Event F = student is female

- Are events E and F independent? Use probabilities to justify your conclusion.
- Find $P(E \cap F)$

2. Random variable

Definition:

If real valued function $\boldsymbol{X}(\omega)$ $(\omega \in \boldsymbol{\Omega})$ defined on probability space satisfies $\{\omega\text{:}\, \boldsymbol{X}(\omega) \leq x\} \in \boldsymbol{B}$ for any real value *x*, $\boldsymbol{X}(\boldsymbol{\Omega})$ is called ***random variables***.

Example

From a box with many balls with different color, pick up one ball. One obtains a coupon card corresponding to the color of the ball.

Color (Even) ω	White	Green	Yellow	Blue	Red
Coupon (Random variable) $X(\boldsymbol{\Omega})$	500	1000	2000	4000	6000
Probability	0.72	0.15	0.1	0.02	0.01

P(white)=$\{\omega\text{: } X(\boldsymbol{\Omega}) \leq 500\}$

3. Distribution function

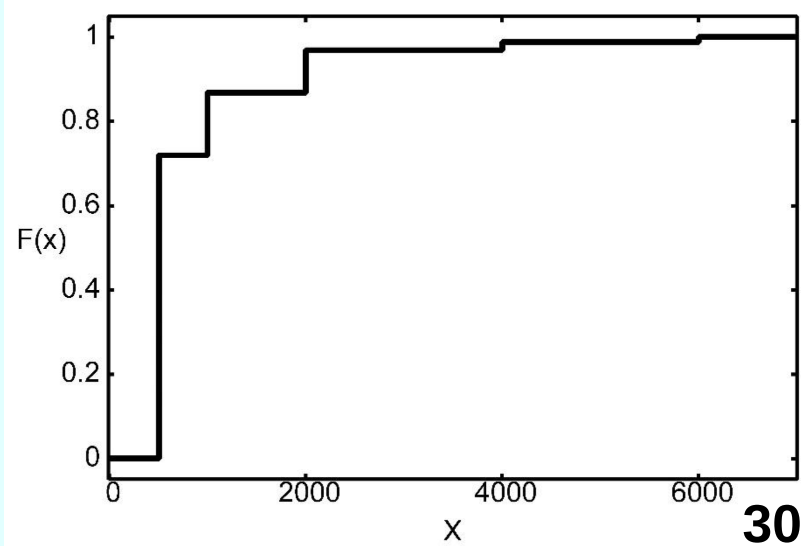
- **Definition**:

$$F_{\boldsymbol{X}}(x) = P(\{\omega: \boldsymbol{X}(\omega) \le x\}), -\infty \le x \le \infty$$

Left-hand-side can be simply described as *F(x)*
Right-hand-side as $P(\boldsymbol{X} \le x)$.

- **Example**

$$\begin{aligned} &F(2000) \\ &\quad = P(\{\omega: X(\omega) \le 2000\}) \\ &\quad = P(\{\text{Yellow} \cup \text{Green} \cup \text{White}\}) \\ &\quad = 0.72 + 0.15 + 0.1 \\ &\quad = 0.97 \end{aligned}$$



Properties of the distribution function

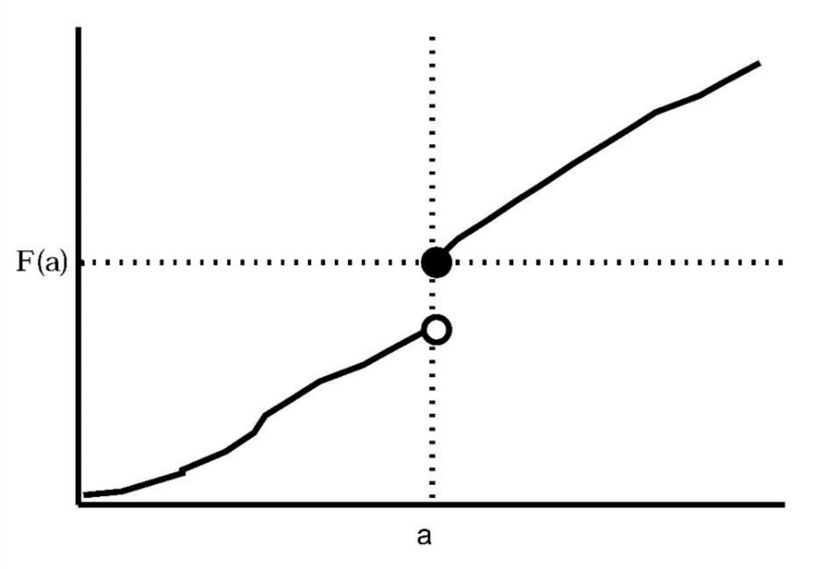
(PD1) Non-decreasing

If $a < b$,

then $F(a) \leq F(b)$.

(PD2) Continuous from the right

$$\lim_{x \to a+0} F(x) = F(a)$$



(PD3) Limit

$$\lim_{x \to -\infty} F(x) = 0$$

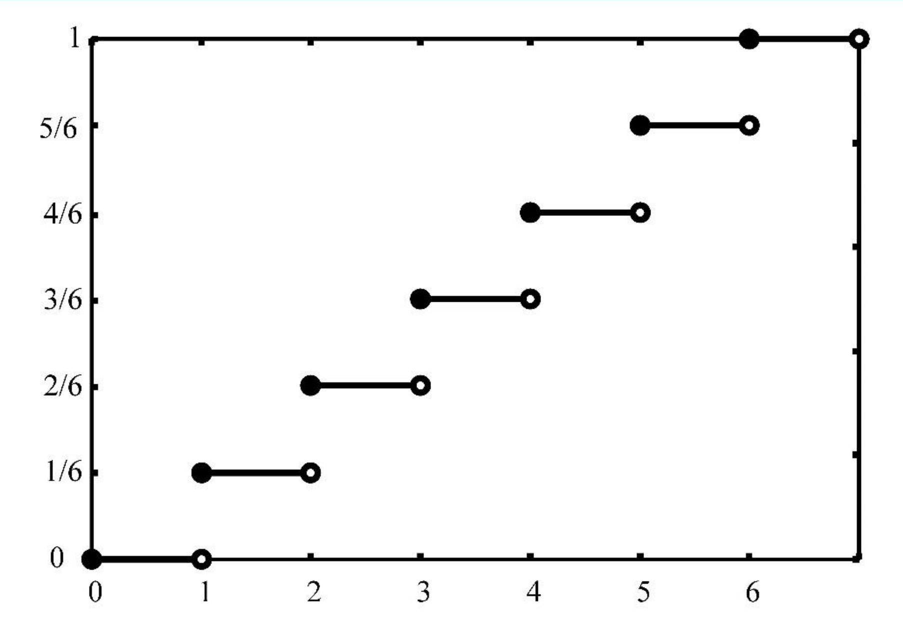
$$\lim_{x \to \infty} F(x) = 1$$

[Example]

- Trial ***S***: Throw dice
- Sample Point ω: 1, 2, 3, …
- Sample Space $\boldsymbol{\Omega}$: {1, 2, 3, 4, 5, 6}

- $F(\{1 \leq \omega < 2\}) = \frac{1}{6}$
- $F(2) = \frac{1}{3}$
- $F(3) = \frac{1}{2}$

  …
- $F(6) = 1$



[Exercise]

Consider the experiment of tossing a coin three times. Let X be the random variables giving the number of heads obtained. We assume that the tosses are independent and the probability of a head is p.

(a) What is the range of X?

(b) Find the probabilities P(X = 0), P(X = 1), P(X = 2), and P(X = 3).

[Answer]

(a) Since we are tossing the coin three times, the number of heads we can get can be 0, 1, 2, or 3. Therefore, the range of X is {0, 1, 2, 3}.

(b) If P(H) = p, then P(T) = 1 − p. Since the tosses are independent, we have

$$P(X=0)=P[\{TTT\}]=(1-p)^3$$

$$P(X=1)=P[\{HTT\}]+P[\{THT\}]+P[\{TTH\}]=3(1-p)^2p$$

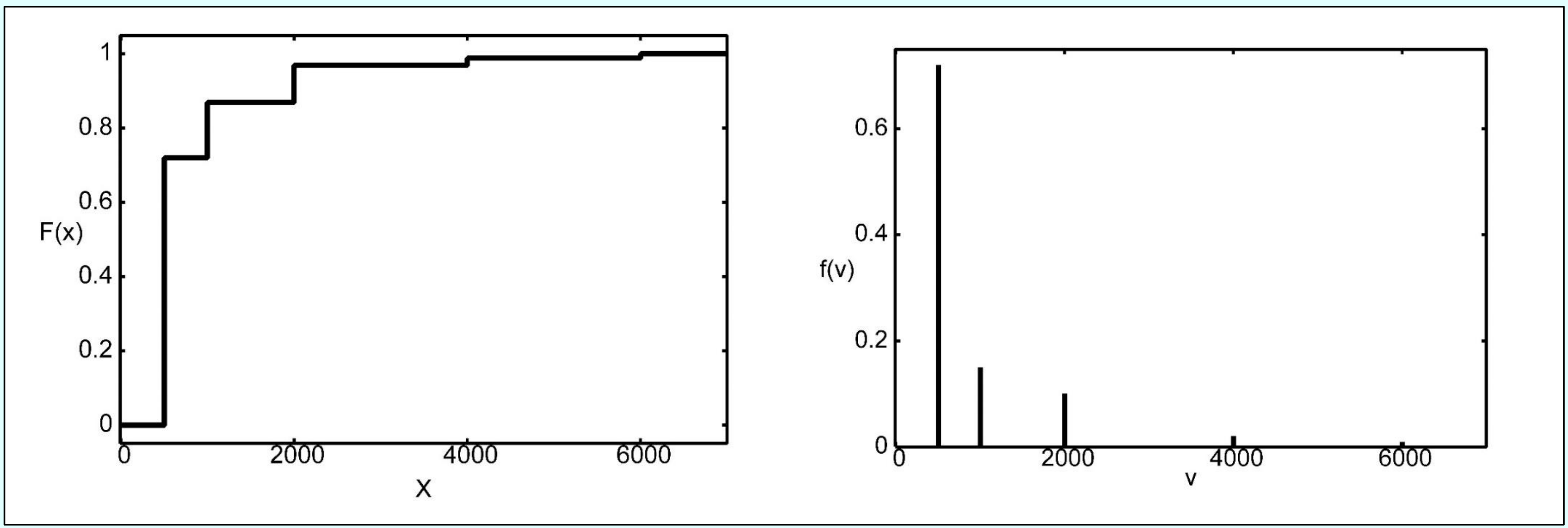
$$P(X=2)=P[\{HHT\}]+P[\{HTH\}]+P[\{THH\}]=3p^2(1-p)$$

$$P(X=3)=P[\{HHH\}]=p^3$$

4. Discrete-type Distribution Function

- Possible values $\boldsymbol{V} = \{v_0, v_1, v_2, \cdots\}$ that ***X*** can take is finite or countable. Taking $\varepsilon > 0$ in a way that no elements except v_k are included in the interval $[v_k - \varepsilon, v_k]$, the **probability mass function (PMF)** (probability that ***X*** takes a value of):

$$f(v_k) = P(X = v_k) \text{ is } f(v_k) = F(v_k) - F(v_k - \varepsilon).$$



- **Properties**

(DT1) $F(x) = \sum_{v_k \leq x} f(v_k)$

(DT2) For all k, $f(v_k) \geq 0$

(DT3) $\sum_k f(v_k) = 1$

- **Example of discrete-type distribution function**

(Ex1) **Bernoulli distribution**: $B(1; p), 0 < p < 1$

Random variable that can take only two values (i.e., biased coin)

$$V = \{0, 1\}$$

$$f(0) = 1 - p$$

$$f(1) = p$$

(Ex2) **Binominal distribution**: $B(n;p), 0 < p < 1$

Consider a lottery that has *p* percent of winning tickets. Try the lottery for *n* times. Probability to draw the winning ticket for k times yields binominal distribution. Probability to win *k* times and loose $(n-k)$ times is $p^k = (1-p)^{n-k}$. Combination of choosing *k* objects among *n* is ${}_nC_k = \binom{n}{k}$.

Hence, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$

(Ex3) **Poisson distribution**: $P_o(\lambda)$

Binominal distribution in the limit of $n \to \infty,\ p \to 0$,

$$np = \lambda,\ f(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Stirling's formula $n! \approx \sqrt{2\pi n}\ n^n e^{-n}$

[Exercise]

I roll two dice and observe two numbers X and Y.

a) Find sample space of X and Y, and the PMFs of X and Y

b) Find $P(X = 2, Y = 6)$

c) Find $P(X > 3|Y = 2)$

5. Continuous-type Distribution Function

- **Distribution function:**

  $F(x)$ can be described as

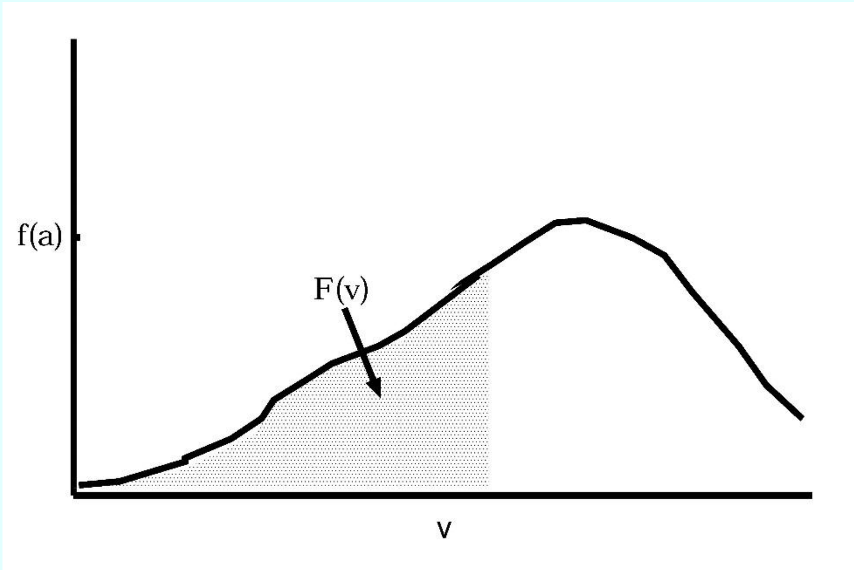
$$F(x) = \int_{-\infty}^{x} f(v)dv$$

- **Probability density** is

$$f(x) = \frac{d}{dx}F(x)$$

(CT1) $f(x) > 0, -\infty < x < \infty$

(CT2) $\int_{-\infty}^{\infty} f(x)dx = 1$



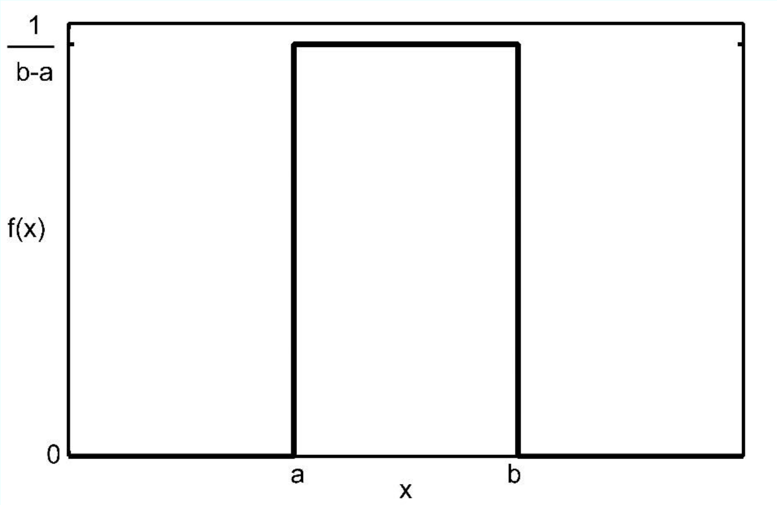
Example of continuous-type distribution function

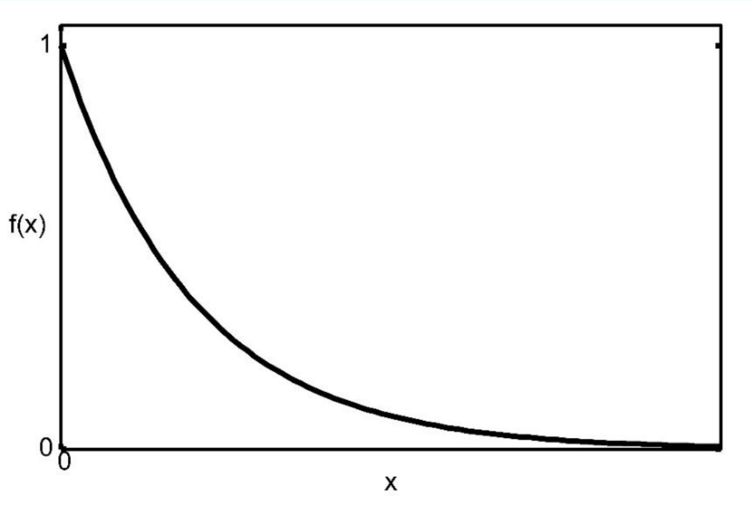
(Ex1) **Uniform distribution**: $U(a,b), a < b$

$$U(a,b) = f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

(Ex2) **Exponential distribution**: $E_x(\alpha), \alpha > 0$

$$E_x(\alpha) = f(x) = \begin{cases} \alpha e^{-\alpha x} & x \ge 0 \\ 0, & x < 0 \end{cases}$$





(Ex3) **Gamma distribution**: $G(\alpha,\nu)$

$$G(\alpha,\nu) = f(x) = \begin{cases} \frac{1}{\Gamma(\nu)}\alpha^{\nu}x^{\nu-1}e^{-\alpha x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

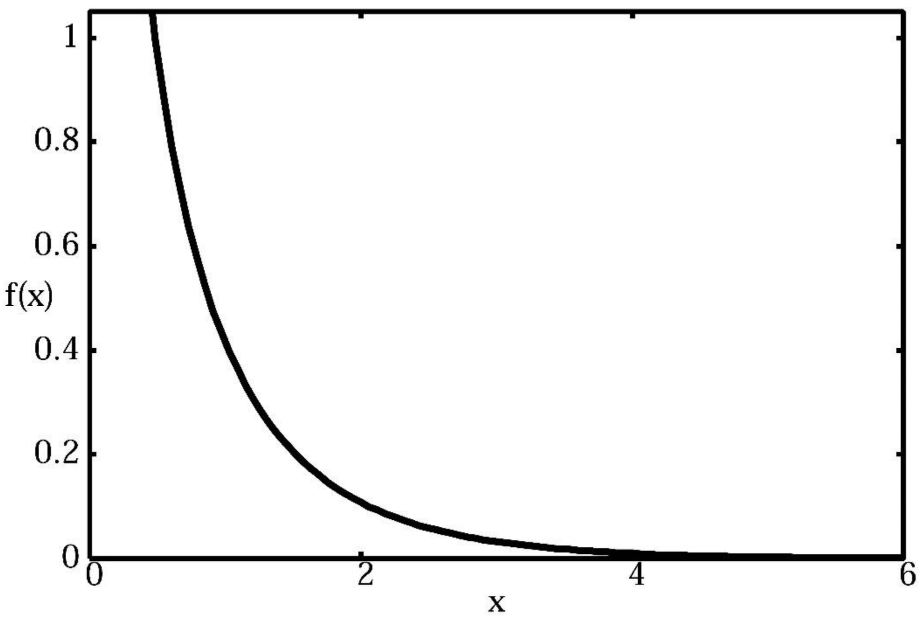
where $\Gamma(\nu) = \int_0^{\infty} t^{\nu-1}e^{-t}dt$.

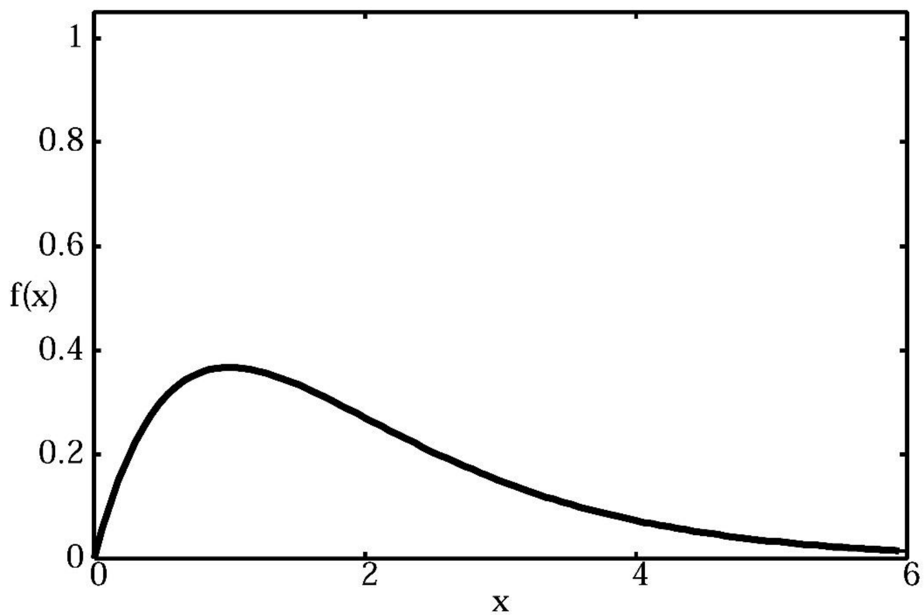
- Properties of Γ :
  - $\Gamma(\nu) = (\nu-1)\Gamma(\nu-1),\ \Gamma(1) = 1$
  - $\Gamma(1/2) = \sqrt{\pi}$

  Especially, if *n* is an integer,
  - $\Gamma(n) = (n-1)!$
  - $\Gamma\left(n+\frac{1}{2}\right) = \left(n-\frac{1}{2}\right)\left(n-\frac{3}{2}\right)\left(n-\frac{5}{2}\right)\cdots\frac{1}{2}\sqrt{\pi}$

When $\nu = 1$, equivalent to $E_x(\alpha)$.

- $G\left(\frac{1}{2}, \frac{n}{2}\right)$: χ^2 (chi-square) distribution with *n* degree of freedom (often used for ANOVA in statistics).



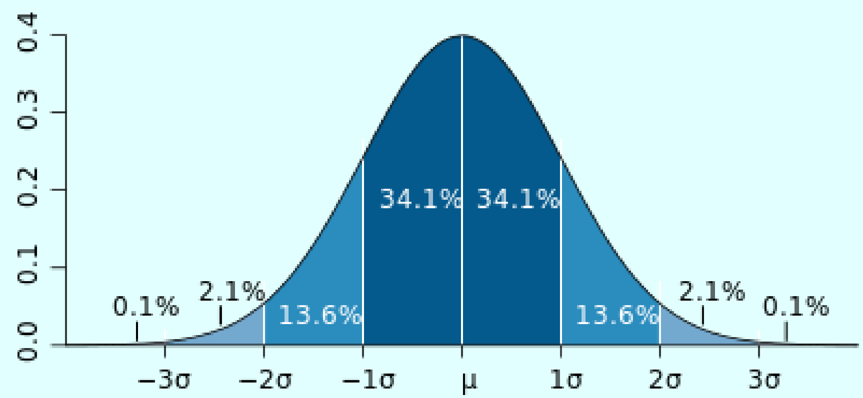


(Ex4) **Gaussian distribution**:

$$N(\mu, \sigma^2), -\infty < \mu < \infty, \sigma > 0$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], -\infty < x < \infty$$

- In case of $N(0,1)$, **standard normal distribution**
- Probability density: $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right]$
- Distribution function: $\boldsymbol{\Phi}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}t^2\right] dt$
- Properties of $\phi(x)$
  - $\phi(x) = \phi(-x)$
  - $\boldsymbol{\Phi}(x) = 1 - \boldsymbol{\Phi}(-x)$
  - $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$



[Exercise] Suppose $\boldsymbol{X}$ is a random variable of $N(0,1)$. With respect to $\boldsymbol{Y}=\boldsymbol{X}^2$, find its distribution function $F_{\boldsymbol{Y}}(y)$ and the probability density $f_{\boldsymbol{Y}}(y)$.

[Answer] If $y<0, F_{\boldsymbol{Y}}(y)=0$, $f_{\boldsymbol{Y}}(y)=0$.

If $y \geq 0$,

$$F_{\boldsymbol{Y}}(y)=P(\boldsymbol{Y}\leq y)=P(\boldsymbol{X}^2\leq y)=P(-\sqrt{y}\leq \boldsymbol{X}\leq \sqrt{y})$$

$$=\int_{-\sqrt{y}}^{\sqrt{y}}\phi(x)dx=2\int_0^{\sqrt{y}}\phi(x)dx,$$

$$f_{\boldsymbol{Y}}(y)=\frac{dF_{\boldsymbol{Y}}(y)}{dy}=\frac{d}{dy}\left\{2\int_0^{\sqrt{y}}\phi(x)dx\right\}=2\phi\left(\sqrt{y}\right)\frac{1}{2\sqrt{y}}$$

$$=\frac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}\exp\left(-\frac{1}{2}y\right)=G\left(\frac{1}{2},\frac{1}{2}\right)$$

This is χ^2 distribution with $n=1$ degrees of freedom.

6. Joint Distribution

- Consider more than n random variables $\boldsymbol{X}_1(\omega), \boldsymbol{X}_2(\omega), \cdots, \boldsymbol{X}_n(\omega)$ simultaneously, and examine their interrelation.
- With regard to event $\{\omega: \boldsymbol{X}_1(\omega) \le x_1, \boldsymbol{X}_2(\omega) \le x_2, \cdots, \boldsymbol{X}_n(\omega) \le x_n\} = \bigcap_{k=1}^{n} \{\omega: \boldsymbol{X}_k(\omega) \le x_k\}$ the distribution function is defined as

$$F(x_1, x_2, \cdots, x_n) = P(\bigcap_{k=1}^{n} \{\omega: \boldsymbol{X}_k(\omega) \le x_k\})$$

Case of $n = 2$: $F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(u, v) du dv$

- Marginal distribution function of $\boldsymbol{X}_1$: $F_1(x) = F(x, +\infty)$
- Marginal distribution function of $\boldsymbol{X}_1$: $F_2(x) = F(+\infty, x)$
- Conditional distribution function given

$$\boldsymbol{X}_2 < c: F(x, c)/F(+\infty, c)$$

[Example] Two-Dimensional Gaussian Distribution

Density function

$$f(u,v)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}Q(u,v)\right\}$$

where

$$Q(u,v)=\frac{1}{1-\rho^2}\left\{\left(\frac{u-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{u-\mu_1}{\sigma_1}\right)\left(\frac{v-\mu_2}{\sigma_2}\right)+\left(\frac{v-\mu_2}{\sigma_2}\right)^2\right\}$$

$$\rho=\frac{C}{\sigma_1\sigma_2},$$

$C=E\{(u-\mu_1)(v-\mu_2)\}$ (**Covariance**)

7. Independency of random variables

Consider *n* random variables $\boldsymbol{X}_{\mathbf{1}}(\omega), \boldsymbol{X}_2(\omega), \ \cdots, \boldsymbol{X}_n(\omega)$. If their joint distribution function $F(x_1, x_2, \cdots, x_n)$ is equal to the product of distribution functions of each random variable

$$F(x_1, x_2, \cdots, x_n) = F(x_1)F(x_2)\cdots F(x_n).$$

- *n* random variables are mutually independent.

- The same holds for probability density as

$$f(x_1, x_2, \cdots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

Case of 2-dimentional normal distribution function.
Suppose $\rho = 0$ (covariance is zero).

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$
$$= f_1(x_1) f_2(x_2)$$

where

$$f_i(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right]\right\} \quad (i = 1, 2)$$

8. Mean

- **Discrete-type Distribution Function**:

  When probability function $f(v_k)$ is given by

  $$f(v_k) = P(\boldsymbol{X} = v_k),$$

  $$E(\boldsymbol{X}) = \sum_k v_k f(v_k)$$

  is called ***mean*** or ***expected value***.

- **Remark**

  Strictly speaking, the mean exists when

  $$\sum_k |v_k| f(v_k) < \infty$$

- **Example Dice**

  $$1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$

Continuous-type Distribution Function:

When the probability function of $\boldsymbol{X}$ is given by $f(x)$,

$$E(\boldsymbol{X}) = \int_{-\infty}^{\infty} x f(x) dx$$

Remark

Strictly speaking, the mean exists when

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty$$

Example

(Ex1) **Bernoulli distribution** $B(1; p)$

$E(\boldsymbol{X}) = 0 \times f(0) + 1 \times f(1) = 0 \times 1(1-p) + 1 \times p = p$

[Exercise]

Let $X \sim Uniform(a, b)$. Find $E(\boldsymbol{X})$.

[Answer]

$$E(\boldsymbol{X}) = \frac{(a+b)}{2}$$

(Ex2) **Binomial distribution** $B(n;p)$

$$\begin{aligned}
E(\boldsymbol{X}) &= \textstyle\sum_{k=0}^{n} kf(k) = \sum_{k=1}^{n} kf(k) \\
&= \textstyle\sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \textstyle\sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \textstyle\sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \textstyle\sum_{l=0}^{m} \binom{m}{l} p^l (1-p)^{m-l} \\
&= np
\end{aligned}$$

(Ex3) **Poisson distribution** $P_0(\lambda)$

$$E(\boldsymbol{X}) = \sum_{k=0}^{\infty} kf(k) = \sum_{k=1}^{\infty} kf(k)$$

$$= \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = \lambda\sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!}$$

$$\{\text{substituting } m = k - 1\}$$

$$= \lambda\sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!}$$

$$\left\{\text{from } \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!} = \sum_{m=0}^{\infty} f(m) = 1\right\}$$

$$= \lambda$$

(Ex4) **Uniform distribution** $U(a,b)$

$$E(\boldsymbol{X}) = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

(Ex5) **Exponential distribution** $E_x(\alpha)$

$$E(\boldsymbol{X}) = \int_0^\infty x\alpha e^{-\alpha x}dx$$

$$= \left[x\alpha\frac{1}{-\alpha}e^{-\alpha x}\right]_0^\infty - \int_0^\infty \alpha\frac{1}{-\alpha}e^{-\alpha x}dx$$

$$= \int_0^\infty e^{-\alpha x}dx = \left[\frac{1}{-\alpha}e^{-\alpha x}\right]_0^\infty$$

$$= \frac{1}{\alpha}$$

(Ex6) **Gamma distribution** $G(\alpha, \nu)$

$$E(\boldsymbol{X}) = \int_0^\infty x \frac{1}{\Gamma(\nu)} \alpha^\nu x^{\nu-1} e^{-\alpha x} dx = \int_0^\infty \frac{1}{\Gamma(\nu)} t^\nu e^{-t} \frac{dt}{\alpha}$$

$$= \frac{1}{\alpha\Gamma(\nu)} \int_0^\infty t^\nu e^{-t} dt = \frac{\Gamma(\nu+1)}{\alpha\Gamma(\nu)} = \frac{\nu}{\alpha}$$

(Ex7) **Normal distribution** $N(\mu, \sigma^2)$

$$E(\boldsymbol{X}) = \int_{-\infty}^\infty x \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

$$= \int_{-\infty}^\infty (\mu + \sigma t) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} t^2\right] dt$$

$$= \mu \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} t^2\right] dt$$

$$+ \int_{-\infty}^\infty \frac{\sigma t}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} t^2\right] dt = \mu$$

Properties of the Mean Value

(M1) For any real-value function $g(x)$, $g(\boldsymbol{X})$ is also a random variable. Mean value for $g(\boldsymbol{X})$ is given as follows.

- Case of discrete-type random variable

$$E\left(g(\boldsymbol{X})\right) = \sum_k g(v_k) f(v_k)$$

- Case of continuous-type random variable

$$E\left(g(\boldsymbol{X})\right) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

(M2) Conditional expectation

- Case of discrete-type random variable

$$f(v|y) = P(\boldsymbol{X} = v|\boldsymbol{Y} = y) \frac{P(\boldsymbol{X}=v,\boldsymbol{Y}=y)}{P(\boldsymbol{Y}=y)}$$

yields $$E(\boldsymbol{X}|\boldsymbol{Y} = y) = \sum_k v_k f(v_k|y)$$

$$E(\boldsymbol{X}) = \sum_y E(\boldsymbol{X}|\boldsymbol{Y} = y) P(\boldsymbol{Y} = y)$$

■ Case of continuous-type random variable

$$f(v|y) = \frac{f(x,y)}{f(y)}$$

yields $$E(\boldsymbol{X}|\boldsymbol{Y}=y) = \int_{-\infty}^{\infty} x f(x|y)\, dx$$

$$E(\boldsymbol{X}) = \int_{-\infty}^{\infty} E(\boldsymbol{X}|\boldsymbol{Y}=y) f(y)\, dy$$

(M3) To unify the framework of computing the expected value for discrete-type and continuous-type random variables,

$$E(\boldsymbol{X}) = \int_{-\infty}^{\infty} x dF(x)$$

(**Reimann-Stieltjes integral**)

(a) $E(a\boldsymbol{X}+b)=aE(\boldsymbol{X})+b$

(b) When joint distribution function of $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ is given by $F(x_1, x_2, \cdots, x_n)$,

$$E\big(g(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n)\big) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \cdots, x_n) dF(x_1, x_2, \cdots, x_n)$$

(c) $E(\boldsymbol{X}_1+\boldsymbol{X}_2+\cdots+\boldsymbol{X}_n)=E(\boldsymbol{X}_1)+E(\boldsymbol{X}_2)+\cdots+E(\boldsymbol{X}_n)$

(d) If $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ are mutually independent,

$$E(\boldsymbol{X}_1\boldsymbol{X}_2\cdots\boldsymbol{X}_n)=E(\boldsymbol{X}_1)E(\boldsymbol{X}_2)\cdots E(\boldsymbol{X}_n)$$

This can be extended to

$$E(g_1(\boldsymbol{X}_1)g_2(\boldsymbol{X}_2)\cdots g_n(\boldsymbol{X}_n)) = E(g_1(\boldsymbol{X}_1))E(g_2(\boldsymbol{X}_2))\cdots E(g_n(\boldsymbol{X}_n))$$

9. Variance

Definition

Suppose mean of random variable $\boldsymbol{X}$ is given by $E(\boldsymbol{X}) = \mu$, $(\boldsymbol{X} - \mu)^2$ gives also a random variable. Mean of $(\boldsymbol{X} - \mu)^2$ is called ***variance*** of $\boldsymbol{X}$, which is denoted as $Var(\boldsymbol{X})$.

$$
\begin{aligned}
Var(\boldsymbol{X}) = E\{(\boldsymbol{X} - \mu)^2\} = \int_{-\infty}^{\infty} (x-\mu)^2 dF(x) \\
= \sum_k (v_k - \mu)^2 f(v_k) \\
= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx
\end{aligned}
$$

Remark

$$\sum_k |v_k|^2 f(v_k) < \infty, \int_{-\infty}^{\infty} |x|^2 f(x) dx < \infty$$

Standard deviation

$$\sqrt{Var(\boldsymbol{X})}$$

Properties of variance

(V1) $Var(\boldsymbol{X}) \geq 0$

(V2) $Var(\boldsymbol{X}) = E(\boldsymbol{X}^2) - \{E(\boldsymbol{X})\}^2$

[Proof] $Var(\boldsymbol{X}) = E(\boldsymbol{X}^2 - 2\mu\boldsymbol{X} + \mu^2)$

$$= E(\boldsymbol{X}^2) - 2E(\mu\boldsymbol{X}) + E(\mu^2)$$

$$= E(\boldsymbol{X}^2) - 2\mu E(\boldsymbol{X}) + \mu^2$$

$$= E(\boldsymbol{X}^2) - 2\mu\mu + \mu^2 = E(\boldsymbol{X}^2) - \mu^2$$

(V3) $Var(a\boldsymbol{X} + b) = a^2 Var(\boldsymbol{X})$

[Proof] $Var(a\boldsymbol{X} + b) = E\left\{\left((a\boldsymbol{X} + b) - (a\mu + b)\right)^2\right\}$

$$= E\{a^2(\boldsymbol{X} - \mu)^2\}$$

$$= a^2 E\{(\boldsymbol{X} - \mu)^2\}$$

$$= a^2 Var(\boldsymbol{X})$$

Properties of variance

(V4) $Var(\boldsymbol{X}_1 + \boldsymbol{X}_2) = Var(\boldsymbol{X}_1) + Var(\boldsymbol{X}_2) + 2Cov(\boldsymbol{X}_1, \boldsymbol{X}_2)$

Covariance: $Cov(\boldsymbol{X}_1, \boldsymbol{X}_2) = E\{(\boldsymbol{X}_1 - \mu_1)(\boldsymbol{X}_2 - \mu_2)\}$

Correlation coefficient: $r(\boldsymbol{X}_1, \boldsymbol{X}_2) = \frac{Cov(\boldsymbol{X}_1, \boldsymbol{X}_2)}{\sqrt{Var(\boldsymbol{X}_1)Var(\boldsymbol{X}_2)}}$

If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are mutually independent, $Cor(\boldsymbol{X}_1, \boldsymbol{X}_2) = 0$.

$\mathrm{r}(\boldsymbol{X}_1, \boldsymbol{X}_2) = 0$: No correlation

(V5) $Var(\boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots \boldsymbol{X}_n) =$

$$\sum_{i=1}^{n} Var(\boldsymbol{X}_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cor(\boldsymbol{X}_i, \boldsymbol{X}_j)$$

In particular, if $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ are mutually independent,

$$Var(\boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_n) = \sum_{i=1}^{n} Var(\boldsymbol{X}_i)$$

[Exercise]

Let X be a continuous random variable with PDF

$$f_x(x) = \begin{cases} x^2(2x + \frac{3}{2}), & 0 < x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

If $Y = \frac{2}{X} + 3$, Find $Var(Y)$.

[Answer]

$$Var(Y) = Var\left(\frac{2}{X} + 3\right) = 4Var\left(\frac{1}{X}\right) = \frac{71}{36}$$

10. Moment

Definition

- *m*-th central moment (moment about mean)

$$\mu_m = E\{(\boldsymbol{X} - \mu)^m\}$$

- *m*-th moment about zero

$$\mu'_m = E\{\boldsymbol{X}^m\}$$

Remark

- If there exists an *m*-th moment, there exists all the moments lower than *m*-th order.
- If there exists no *m*-th moment, there exists no moment higher than *m*-th order.

Table of distribution functions

Name of the distribution and range of the parameters	Probability density function	Mean and Variance	Characteristic function
Bernoulli distribution $B(1;p)$ $(0 < p < 1, q = 1-p)$	$p^k q^{1-k}$ $k = 0, 1$	p, pq	$pe^{jt} + q$
Binominal distribution $B(n;p)$ (n: Integer, $0 < p < 1, q = 1-p$)	$\binom{n}{k} p^k q^{n-k}$ $k = 0, 1, 2, \cdots, n$	np, npq	$(pe^{jt} + q)^n$
Poisson distribution $Po(\lambda)$ $(\lambda > 0)$	$e^{-\lambda} \frac{\lambda^k}{k!}$ $k = 1, 2, \cdots$	λ, λ	$\exp[\lambda(e^{jt} - 1]$
Uniform distribution $U(a,b)$ $(-\infty < a < b < \infty)$	$\frac{1}{b-a}$ $a \leq x \leq b$	$(a+b)/2$ $(b-a)^2/12$	$\frac{e^{jbt} - e^{jat}}{j(b-a)t}$
Exponential distribution $Ex(\alpha)$ $(\alpha > 0)$	$\alpha e^{-\alpha x}$ $x \geq 0$	$1/\alpha, 1/\alpha^2$	$(1 - \frac{jt}{\alpha})^{-1}$
Gamma distribution $G(\alpha, \nu)$ $(\alpha, \nu > 0)$	$\frac{1}{\Gamma(\nu)} \alpha^\nu x^{\nu-1} e^{-\alpha x}$ $x \geq 0$	ν/α ν/α^2	$(1 - \frac{jt}{\alpha})^{-\nu}$
Normal distribution $N(\mu, \sigma^2)$ $(-\infty < \mu < \infty, \sigma > 0)$	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$ $-\infty < x < \infty$	μ σ^2	$\exp[j\mu t - \frac{\sigma^2}{2} t^2]$

11. Characteristic Function

- **Definition** (Characteristic Function)

$$\varphi(t) \equiv E\left(e^{it\boldsymbol{X}}\right) = \int_{-\infty}^{\infty} e^{itx} dF(\boldsymbol{X})$$

- **Remark**

$$\begin{aligned} |\varphi(t)| &= \left|\int_{-\infty}^{\infty} e^{itx} dF(\boldsymbol{X})\right| \\ &\leq \int_{-\infty}^{\infty} \left|e^{itx}\right| dF(\boldsymbol{X}) \\ &= \int_{-\infty}^{\infty} dF(\boldsymbol{X}) = 1 \end{aligned}$$

Hence, characteristic function exists for arbitrary distribution.

Properties of characteristic function

(CF1) Moments can be obtained from derivatives of the characteristic function.

$$\varphi^{(m)}(t) = E\left\{(i\boldsymbol{X})^m e^{it\boldsymbol{X}}\right\}$$

$$\varphi^{(m)}(0) = E\{(i\boldsymbol{X})^m\}$$

Hence

$$\mu'_m = E(\boldsymbol{X}^m) = \frac{\varphi^{(m)}(0)}{i^m}$$

$$E(\boldsymbol{X}) = \frac{\varphi'(0)}{i}$$

$$Var(\boldsymbol{X}) = \frac{\varphi''(0)}{i^2} - \left(\frac{\varphi'(0)}{i}\right)^2$$

(CF2) If there exist central moment lower than *m*-th order μ'_k $(1 \le k \le m)$, characteristic function $\varphi(t)$ can be expanded around $t = 0$ into Taylor series:

$$\varphi(t) = \sum_{k=0}^{\infty} \frac{1}{k!} \varphi^{(k)}(0) t^k = 1 + \sum_{k=1}^{m} \frac{(it)^k}{k!} \mu'_k + O(t^{m+1})$$

(CF3) Suppose random variables $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ are mutually independent and characteristic function of $\boldsymbol{X}_i$ is given by $\varphi_i(t)$. Then, characteristic function of the following variable

$$S = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_n$$

is given by

$$\varphi_S(t) = \varphi_1(t)\, \varphi_2(t) \cdots \varphi_n(t)$$

(CF4) Consider two random variables $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Suppose their distribution functions are given respectively by $F_1(x)$ and $F_2(x)$ and their characteristic functions are given by $\varphi_1(t)$ and $\varphi_2(t)$.

The necessary and sufficient condition of $F_1(x) = F_2(x)$ is $\varphi_1(t) = \varphi_2(t)$.

Probability density function is obtained by inversion formula.

- Discrete-type distribution function:

$$f(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikx}\varphi(t)dt$$

- Continuous-type distribution function:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)dt$$

(CF5) (Continuity theorem)

Suppose $F_1(x), F_2(x), \cdots$ represent a sequence of distributions and $\varphi_1(t), \varphi_2(t), \cdots$ represent a sequence of corresponding characteristic functions. If the series $\varphi_1(t), \varphi_2(t), \cdots$ converges for any constant t

$$\lim_{n\to\infty} \varphi_n(t) = \varphi(t),$$

and the convergence value $\varphi(t)$ is continuous at $t = 0$, $\varphi(t)$ gives the characteristic function.

For a distribution function $F(x)$ that corresponds to $\varphi(t)$.

$$\lim_{n\to\infty} F_n(x) = F(x).$$

(CF6) Characteristic function for joint distribution of random variables $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ is defined by

$$\varphi(t_1, t_2, \cdots, t_n) = E\{\exp[i(t_1\boldsymbol{X}_1 + t_2\boldsymbol{X}_2 + \cdots t_n\boldsymbol{X}_n)]\}$$

Necessary and sufficient condition for mutual independency of $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ is given by

$$\varphi(t_1, t_2, \cdots, t_n) = \varphi_1(t_1)\varphi_2(t_2)\cdots\varphi_n(t_n).$$

Where:

$$\varphi_i(t_i) = \varphi_1(0, \cdots, 0, t_i, 0, \cdots, 0)\ .$$

12. Application of characteristic function

Law of Large Numbers

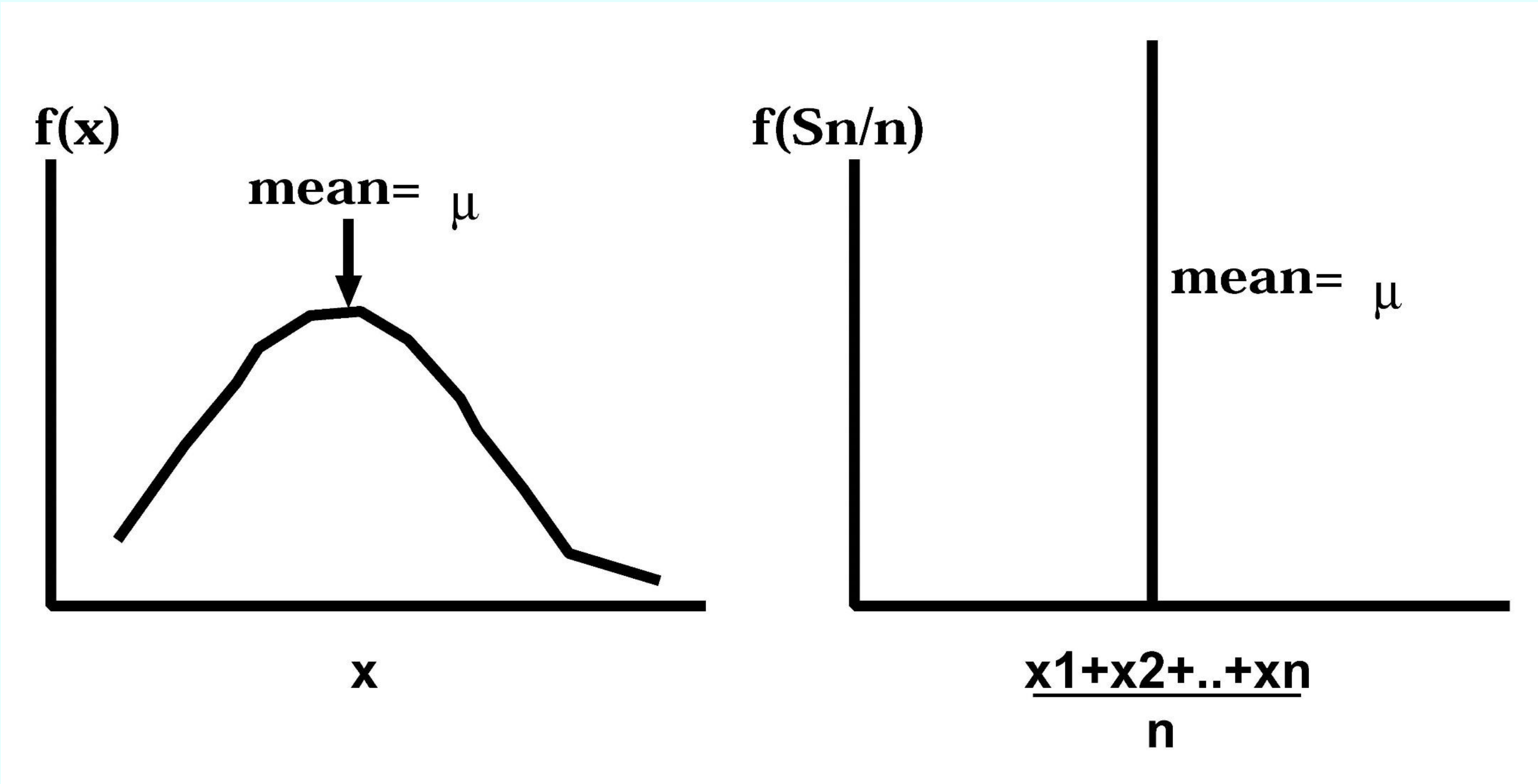
Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ be mutually independent random variables which obey the same distribution. Suppose that the distribution has a finite expectation value $\mu = E(\boldsymbol{X}_j)$, where its characteristic function is given by $\varphi(t)$.

Let

$$S_n = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_n$$

Then for any $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) = 0$$


f(x)
mean= μ
x
f(Sn/n)
mean= μ
x1+x2+..+xn
n

- [Proof] Let characteristic functions of $\boldsymbol{X}_i$ and $\frac{S_n}{n}$ be $\varphi(t)$ and $\varphi_n(t)$, respectively. $\varphi_n(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n$

From (CF2), Taylor's expansion of $\varphi(t)$ around $t = 0$ yields $\varphi(t) = 1 + i\mu t + O(t^2)$. On the other hand, Taylor's expansion of $e^{i\mu t}$ around $t = 0$ gives $e^{i\mu t} = 1 + i\mu t + O(t^2)$.

Hence, $\varphi(t) = e^{i\mu t} + O(t^2)$. This leads to

$$\varphi_n(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n = \left[e^{\frac{i\mu t}{n}} + O(t^2)\right]^n \to e^{i\mu t} \ (n \to \infty)$$

$e^{j\mu t}$ represents characteristic function of a random variable that takes a value of μ with a probability of 1. From the continuity theorem, $\varphi_n(t) \to e^{i\mu t}, \quad \frac{S_n}{n} \to \mu \ \ (n \to \infty)$.

Central Limit Theorem

$\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n$ be a sequence of *n* independent and identically distributed random variables, each having finite values of expectation μ and variance σ^2 (> 0).

Let S_n^* be

$$S_n^* = \sum_{j=1}^{n} \frac{X_j - \mu}{\sqrt{n}\sigma}.$$

Then, as *n* approaches infinity $(n \to \infty)$, S_n^* will converge in distribution to the standard normal distribution $N(0,1)$.

Example of central limit theorem: Summation of uniform random variables