THE CHINESE UNIVERSITY OF HONG KONG, SHENZHEN
SCHOOL OF SCIENCE AND ENGINEERING
STA2002 Probability and Statistics II
Zhaoyuan LI
Lecture 1 Notes (Review of Probability and Statistics I)

Counting

1. Multiplication Principle

2. Selection of r from n **distinct** objects:

	With replacement	Without replacement
Ordered	n^r	${}_nP_r = \dfrac{n!}{(n-r)!}$
Unordered	${}_nH_r = \dbinom{n+r-1}{r} = \dfrac{(n+r-1)!}{r!(n-1)!}$	${}_nC_r = \dbinom{n}{r} = \dfrac{n!}{r!(n-r)!}$

- Classify the following cases into the above 4 groups:
  (a) Arrangement of r objects taken from a set of n **distinguishable** objects **with replacement**. (e.g., number of possible codes on a combination lock.)
  (b) Arrangement of r objects taken from a set of n **distinguishable** objects **without replacement** (**permutation** of n objects taken r at a time). (e.g., number of possible outcomes (winners of first, second and third prizes) in a contest.)
  (c) Selection of r objects **without regard to order** from n **distinguishable** objects **with replacement**. (e.g., number of possible results from rolling three dices.)
  (d) Selection of r objects **without regard to order** from n **distinguishable** objects **without replacement** (**combination** of n objects taken r at a time). (e.g., number of possible poker hands from a deck of cards.)

3. Arrangement of n objects, with r **distinct** types. In other words, there are n_1 objects of type 1, n_2 objects of type 2, ..., n_r objects of type r.

(e.g., number of different letter arrangements can be formed using the letters STATISTICS.)

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!} \tag{1}$$

4. Partition of n indistinguishable objects into r distinct groups where some groups can be empty.

$$ {}_rH_n = \binom{r+n-1}{n} = \frac{(r+n-1)!}{n!(r-1)!} \tag{2}$$

5. The counting formulae we have seen so far can be found as the coefficients in the following identities:

   (a) Binomial Theorem

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r} \tag{3}$$

   (b) Multinomial Theorem

$$(x_1 + x_2 + \cdots + x_r)^n = \sum_{\substack{(n_1, \ldots, n_r) \\ n_1 + \cdots + n_r = n}} \frac{n!}{n_1! n_2! \cdots n_r!} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r} \tag{4}$$

   (c) Negative Binomial Theorem

$$\frac{1}{(1-x)^n} = \sum_{r=0}^{\infty} \binom{n+r-1}{r} x^r \tag{5}$$

Probability

1. The Mathematical Theory of Probability

   (a) De Morgan's Law

   $$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c \quad \text{and} \quad \left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

   (b) Some Terminologies

      i. Mutually exclusive: $A_1, A_2, \ldots, A_n$ are mutually exclusive if $A_i \cap A_j = \varnothing$ for all $i \neq j$.

      ii. Exhaustive: $A_1, A_2, \ldots, A_n$ are exhaustive if $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$.

      iii. Partition: $A_1, A_2, \ldots, A_n$ is called a partition if the events are mutually exclusive and exhaustive.

      iv. Complement: The complement of event A is the collection of outcomes not in A, i.e., $A^c = \Omega \backslash A$.

   (c) Kolmogorov's Axiom

      i. $\Pr(A) \geq 0$ for any event A.

      ii. $\Pr(\Omega) = 1$.

      iii. For any sequence of mutually exclusive events $A_1, A_2, \ldots,$

      $$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i) \tag{6}$$

      (Countable additivity).

   (d) Inclusion-Exclusion Principle

   $$\begin{aligned} \Pr(A_1 \cup \cdots \cup A_n) &= \sum_{i=1}^{n} \Pr(A_i) - \sum_{i_1 < i_2} \Pr(A_{i_1} \cap A_{i_2}) + \cdots \\ &\quad + (-1)^{n+1} \sum_{i_1 < i_2 < \cdots < i_n} \Pr(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}) \\ &= \sum_{j=1}^{n} (-1)^{j+1} \sum_{i_1 < i_2 < \cdots < i_j} \Pr\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_j}\right) \end{aligned} \tag{7}$$

   where n can also be ∞.

(e) Conditional Probability:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad \text{where } \Pr(B) > 0$$

2. Law of Total Probability

   (a) If $0 < \Pr(B) < 1$, then $\Pr(A) = \Pr(A|B)\Pr(B) + \Pr(A|B^c)\Pr(B^c)$ for any A.

   (b) If $B_1, B_2, \ldots, B_k$ are mutually exclusive and exhaustive events (i.e., a partition of the sample space), then for any event A,

$$\Pr(A) = \sum_{j=1}^{k} \Pr(A|B_j)\Pr(B_j) \tag{8}$$

   where k can also be ∞.

3. Bayes' Theorem (Bayes' Rule, Bayes' Law)
   For any two events A and B with $\Pr(A) > 0$ and $\Pr(B) > 0$,

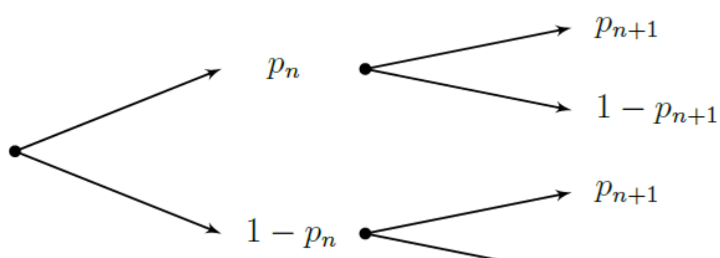
$$\Pr(B|A) = \Pr(A|B)\frac{\Pr(B)}{\Pr(A)}. \tag{9}$$

4. Bayes' Theorem
   If $B_1, B_2, \ldots, B_k$ are mutually exclusive and exhaustive events (i.e., a partition of the sample space), and A is any event with $\Pr(A) > 0$, then for any B_j,

$$\Pr(B_j|A) = \frac{\Pr(A|B_j)\Pr(B_j)}{\Pr(A)} = \frac{\Pr(B_j)\Pr(A|B_j)}{\sum_{i=1}^{k}\Pr(B_i)\Pr(A|B_i)} \tag{10}$$

   where k can also be ∞.

5. Recurrence Relation
   Draw a tree diagram to consider the relationship between two consecutive terms in the probability sequence.



If $p_{n+1} = a+bp_n$ (in other words, $p_n = a+bp_{n-1}$), then $p_n = \frac{a\left(1-b^{n-1}\right)}{1-b} + b^{n-1}p_1$.

Discrete distributions

1. Random Variables

   (a) Basic Concepts

      i. A random variable (r.v.) is a measurable function between a sample space (domain) and state space (range).

      ii. A function (with some requirement) $X : \Omega \ni \omega \mapsto X(\omega) \in X(\Omega)$ defined on the sample space $\Omega = \{\omega\}$ is called a random variable.

   (b) Distribution - Law of a Random Variable
   is a probabilistic one, called the probability distribution of a random variable. There are two criteria for any real-valued function $f(x)$ to be a valid probability mass function (pmf) or probability density function (pdf).

      i. $f(x) > 0$ for any $x \in X(\Omega)$.

      ii. $\sum_{x \in X(\Omega)} p(x) = 1$ or $\int_{x \in X(\Omega)} f(x)\mathrm{d}x = 1$.

   (c) Real-valued Random Variables
   There are three groups of real-valued random variables:

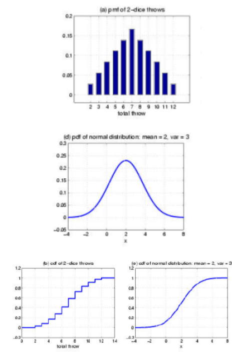
      i. Discrete Random Variables (Its state space is a discrete subset in $\mathbb{R}$.)

      ii. Continuous Random Variables (Its state space is a continuous subset in $\mathbb{R}$.)

      iii. Partially Discrete and Partially Continuous Random Variables

   The distribution of a discrete/continuous random variable is called its probability mass/density function because of a superficial difference in mathematical treatments and graphical representation.

	Discrete	Continuous
Random variable	$X:\ \Omega \ni \omega \mapsto X(\omega) \in X(\Omega) \subset \mathbb{R}$	
Probability mass function (pmf)	$p_X(x) = \Pr(X = x)$	N/A
Probability density function (pdf)	N/A	$f_X(x) = \lim\limits_{dx \to 0} \frac{\Pr\{X \in (x - dx, x]\}}{dx}$
Cumulative distribution function (cdf)	$F_X(x) = \Pr(X \le x) = \begin{cases} \sum_{t \le x} p_X(t), & \text{for discrete } X; \\ \int_{-\infty}^{x} f_X(t)dt, & \text{for continuous } X. \end{cases}$	
Expectation	$\mathrm{E}(X) = \begin{cases} \sum_{x \in X(\Omega)} x \cdot p_X(x), & \text{for discrete } X; \\ \int_{x \in X(\Omega)} x \cdot f_X(x)dx, & \text{for continuous } X. \end{cases}$	



2. Discrete random variables

   (a) Definition of pmf

$$p_X(x) = \Pr(X = x) = \Pr(\{\omega : X(\omega) = x\}), \text{ where } x \in \mathbb{R} \quad (11)$$

   (b) Properties of pmf
   The state space of a discrete random variable X must be countable. Denote the state space of X by $S = \{x_1, x_2, x_3, \ldots\}$,

      i. $0 \le p_X(x_k) \le 1, k = 1, 2, \ldots$
      ii. If $x \ne x_k$ for all $k = 1, 2, \ldots$, then $p_X(x) = 0$.
      iii. $\sum_k p_X(x_k) = 1$.
      iv. $F_X(x) = \sum_{x_k \le x} p_X(x_k)$

   Remark 1: $\Pr(E) = 0 \Leftrightarrow E = \varnothing$.
   Remark 2: If X is a continuous random variable, then $\Pr(X = x) = 0$.

   (c) Expectation

      i. $\mathrm{E}(X) = \sum_{x \in X(\Omega)} x p_X(x)$.
      ii. $\mathrm{E}[g(X)] = \sum_{x \in X(\Omega)} g(x) p_X(x)$.

   (d) Mean and Variance

      i. The mean of X is denoted by $\mu = \mathrm{E}(X)$.

ii. The variance of X is denoted by $\sigma^2 = \mathrm{Var}(X) = \mathrm{E}\left\{[X - \mathrm{E}(X)]^2\right\} = \mathrm{E}\left(X^2\right) - [\mathrm{E}(X)]^2$.

(e) Moment and Moment Generating Function

i. Moment.
Let r be a positive integer. The r-th moment of a r.v. X is $E(X^r)$. The r-th moment of X about b is $E((X-b)^r)$.

ii. Moment generating function.
The moment generating function (mgf) of a r.v. X is defined as $M_X(t) = \mathrm{E}\left(e^{tX}\right)$ if it exists. The domain of $M_X(t)$ is the range of real number t such that the expectation $\mathrm{E}\left(e^{tX}\right)$ is finite. Also

$$M_X^{(r)}(0) = \left.\frac{\partial^r M_X(t)}{\partial t^r}\right|_{t=0} = \mathrm{E}\left(X^r\right). \tag{12}$$

3. Some common discrete distributions for random variable X:

Distribution	Support	pmf	E(X)	Var(X)	Description
Bernoulli Ber(p)	$\{0,1\}$	$p^x(1-p)^{1-x}$	p	$p(1-p)$	One trial; 1 if success, 0 if failure
Binomial B(n,p)	$\{0,1,2,\ldots,n\}$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$	Number of success(es) in n Bernoulli trials
Geometric Geo(p)	$\{1,2,3,\ldots\}$	$(1-p)^{x-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	Number of Bernoulli trials needed to get one success
Negative Binomial (Pascal) NB(r,p)	$\{r,r+1,r+2,\ldots\}$	$\binom{x-1}{r-1}(1-p)^{x-r}p^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	Number of Bernoulli trials needed to get r success(es)
Hypergeometric Hyp(N,m,n)	$\{\max[n-(N-m),0], \ldots, \min(n,m)\}$	$\frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$	$\frac{nm}{N}$	$\frac{n(N-n)}{N-1}\frac{m}{N}\frac{N-m}{N}$	Number of type 1 objects chosen when n objects are chosen from N ones, of which m are of type 1
Poisson Po(λ)	$\{0,1,2,\ldots\}$	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ	Approximates binomial for large n, but $\lambda = np$ for not large n
Uniform DU($1,2,\ldots,n$)	$\{1,2,\ldots,n\}$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	All the numbers $1,2,\ldots,n$ are equally probable

Continuous distributions

1. Cumulative distribution function (cdf)

   (a) Definition

   $$F_X(x) = \Pr(X \leq x) = \Pr(\{\omega : X(\omega) \leq x\}), \text{ where } x \in (-\infty, +\infty) \tag{13}$$

   (b) Properties

      i. $0 \leq F_X(x) \leq 1$.
      ii. If $x_1 < x_2$, then $F_X(x_1) \leq F_X(x_2)$.
      iii. $\lim_{x \to +\infty} F_X(x) = F_X(+\infty) = 1$.
      iv. $\lim_{x \to -\infty} F_X(x) = F_X(-\infty) = 0$.
      v. $\lim_{x \to a^+} F_X(x) = F_X(a^+) = F_X(a)$, where $a^+ = \lim_{\epsilon \to 0^+}(a + \epsilon)$.
      vi. $\Pr(a < X \leq b) = F_X(b) - F_X(a)$.
      vii. $\Pr(X > a) = 1 - F_X(a)$.
      viii. $\Pr(X < b) = F_X(b^-)$, where $b^- = \lim_{\epsilon \to 0^+}(b - \epsilon)$.

2. Continuous Random Variables

   (a) Definition of pdf

   $$\begin{aligned} f_X(x) &= \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} \\ &= \lim_{\Delta x \to 0} \frac{\Pr(X \leq x + \Delta x) - \Pr(X \leq x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Pr(X < x + \Delta x) - \Pr(X \leq x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Pr(X \leq x + \Delta x) - \Pr(X < x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Pr(X < x + \Delta x) - \Pr(X < x)}{\Delta x} \end{aligned} \tag{14}$$

   (b) Properties of pdf

      i. $f_X(x) \geq 0$.
      ii. $\int_{-\infty}^{+\infty} f_X(x) \mathrm{d}x = 1$.

iii. $F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t$.

iv.

$$\Pr(a < X \leq b) = \Pr(a \leq X \leq b) = \Pr(a \leq X < b) = \Pr(a < X < b) = \int_a^b f(x)\mathrm{d}x =$$
$$F_X(b) - F_X(a). \tag{15}$$

(c) Expectation (Continuous)

$$\begin{aligned} \mathrm{E}(X) &= \int_{X(\Omega)} x f(x)\mathrm{d}x \\ \mathrm{E}[g(X)] &= \int_{X(\Omega)} g(x) f(x)\mathrm{d}x \end{aligned} \tag{16}$$

(d) Gamma Function and Beta Function

$$\begin{aligned} \Gamma(\alpha) &= \int_0^\infty e^{-x} x^{\alpha-1}\mathrm{d}x, \quad \alpha > 0 \\ \mathrm{B}(\alpha, \beta) &= \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \end{aligned} \tag{17}$$

(e) Properties of the Gamma Function

i. $\Gamma(1) = 1$

ii. For $\alpha > 1, \Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$

iii. For any integer $n \geq 1, \Gamma(n) = (n-1)!$

iv. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

3. Some Common Continuous Distributions

(a) Uniform Distribution

Let $X \sim U(a, b)$, where $b > a$. The probability density function (pdf) of X is

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a < x < b \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

$\mu = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}$.

(b) Exponential Distribution

Let $X \sim Exp(\lambda)$, where $\lambda > 0$. The pdf of X is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

$\mu = \frac{1}{\lambda}, \sigma^2 = \frac{1}{\lambda^2}$.

A exponential random variable can be used to describe random time elapsing between unpredictable events.

(c) Gamma Distribution

Let $X \sim \Gamma(\alpha, \lambda)$, where $\alpha, \lambda > 0$. The pdf of X is

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

$\mu = \frac{\alpha}{\lambda}, \sigma^2 = \frac{\alpha}{\lambda^2}$.

Exponential distribution is a special case of gamma distribution with $\alpha = 1$, i.e., $Exp(\lambda) = \Gamma(1, \lambda)$.

Chi-squared distribution with r degrees of freedom (where r is a positive integer) is a special case of gamma distribution, i.e., $\chi^2(r) \equiv \Gamma\left(\frac{r}{2}, \frac{1}{2}\right)$.

(d) Chi-squared Distribution

Let $X \sim \chi^2(r)$, where r (a positive integer) is the number of degrees of freedom. The pdf of X is

$$f(x) = \begin{cases} \frac{1}{\Gamma\left(\frac{r}{2}\right) 2^{\frac{r}{2}}} x^{\frac{r}{2}-1} e^{-\frac{x}{2}}, & \text{for } x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

$\mu = r, \sigma^2 = 2r$.

(e) Normal Distribution

Let $X \sim N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ is the mean and $\sigma^2 > 0$ is the variance. The pdf of X is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty. \tag{22}$$

Some important points:

i. A normal distribution is symmetric about its mean. That is, if $X \sim N(\mu, \sigma^2)$, then

$$\Pr(X \leq \mu - x) = \Pr(X \geq \mu + x). \tag{23}$$

In particular, $\Phi(x) = 1 - \Phi(-x)$, where $\Phi(x)$ is the cumulative distribution function (cdf) of the standard normal distribution.

ii. Transformation into the standard normal random variable $Z \sim N(0, 1)$:

$$Z \text{ -score } = \frac{x - \mu}{\sigma}. \tag{24}$$

iii. Relationship between normal and chi-squared distribution:
If $X \sim N(0, 1)$, then $X^2 \sim \chi^2(1)$.
If $Z_1, \ldots, Z_k$ are independent random variables, each with a standard normal distribution, then by definition $Z_1^2 + \cdots + Z_k^2$ has a $\chi^2(k)$ distribution.

(f) Beta Distribution
Let $X \sim Beta(\alpha, \beta)$, where $\alpha, \beta > 0$. The pdf of X is

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & \text{for } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

$\mu = \frac{\alpha}{\alpha+\beta}, \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

4. Other properties of distribution

(a) Markov's inequality
If X is a non-negative random variable with finite mean $E(X)$, then for any $c > 0$,

$$\Pr(X \geq c) \leq \frac{\mathrm{E}(X)}{c} \tag{26}$$

Proof:
Note that since $X \geq 0$,

$$c\mathbf{1}_{\{X \geq c\}} \leq X. \tag{27}$$

The inequality follows by taking expectation on both sides.

(b) Chebyshev's inequality

If the random variable X has finite mean μ and finite variance σ^2, then for any real number $k > 0$,

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}. \tag{28}$$

Proof:

By Markov's inequality,

$$\Pr\left\{(X - \mu)^2 \geq k^2\sigma^2\right\} \leq \frac{\mathrm{E}\left[(X - \mu)^2\right]}{k^2\sigma^2} = \frac{1}{k^2}. \tag{29}$$

(c) Transformation of pdf

Let X be a continuous random variable on distributed on a space S with pdf $f_X(x)$. Let $Y = g(x)$ where g is a function such that g^{-1} exists. Then the pdf of Y can be obtained by

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right| \quad y \in g(S). \tag{30}$$

(d) Quantiles of a distribution

The β-quantile of a probability distribution function $F_X$of a random variable X is defined to be

$$F_X^{-1}(\beta) = \inf\left\{x \in \mathbb{R} : F_X(x) \geq \beta\right\}. \tag{31}$$

Multivariate distribution

1. Joint and Marginal Distributions

   (a) Let $X_1, \ldots, X_n$ be random variables defined on the same sample space Ω. The joint distribution function of $(X_1, \ldots, X_n)$ is defined by

   $$F(x_1, \ldots, x_n) = \Pr(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) \tag{32}$$

   The distribution function F_X of each X_i is called the marginal distribution function of X_i.

   (b) For **discrete** random variables $X_1, \ldots, X_n$, the joint probability mass function of $(X_1, \ldots, X_n)$ is

   $$p(x_1, \ldots, x_n) = \Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n). \tag{33}$$

   The mass function $p_{X_i}(x) = \Pr(X_i = x)$ of each X_i is called the marginal probability mass function of X_i:

   $$p_{X_i}(x) = \sum_{u_1} \cdots \sum_{u_{i-1} u_{i+1}} \cdots \sum_{u_n} p(u_1, \ldots, u_{i-1}, x, u_{i+1}, \ldots, u_n). \tag{34}$$

   (c) Random variables $X_1, \ldots, X_n$ are (jointly) continuous if their joint distribution function F satisfies

   $$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \ldots, u_n)\, \mathrm{d}u_n \cdots \mathrm{d}u_1, \tag{35}$$

   for some non-negative function $f : \mathbb{R}^n \to [0, \infty)$. The function f is called the joint probability density function of $(X_1, \ldots, X_n)$.The pdf of X_i is called the marginal pdf of X_i.

   $$f_{X_i}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \ldots, u_{i-1}, x, u_{i+1}, \ldots, u_n)\, \mathrm{d}u_n \cdots \mathrm{d}u_{i+1} \mathrm{d}u_{i-1} \cdots \mathrm{d}u_1. \tag{36}$$

   (d) If a joint distribution function F possesses all partial derivatives at $(x_1, \ldots, x_n)$, then the joint pdf is

   $$f(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \ldots, x_n) \tag{37}$$

2. Independence of Random Variables
   Random variables $X_1, \ldots, X_n$ are **independent** if and only if their joint pmf (pdf) or cdf is equal to the product of their marginal pmfs (pdfs) or cdfs, i.e.,

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) \quad \text{or} \quad F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \tag{38}$$

   Proposition Random variables X and Y are independent if and only if

   (a) the supports of X and Y do not depend on each other, and

   (b) $f(x, y)$ can be factorized as $g(x)h(y)$.

   This proposition applies to both discrete and continuous random variables and can be generalized to multivariate cases.

3. Expectation of Function of Random Variables
   For random variables $X_1, \ldots, X_n$ with joint pmf or pdf $p(x_1, \ldots, x_n)$ or $f(x_1, \ldots, x_n)$, if $u(X_1, \ldots, X_n)$ is a function of these random variables, then the expectation of this function is defined as

$$\begin{aligned} \mathrm{E}[u(X_1, \ldots, X_n)] &= \sum_{x_1} \cdots \sum_{x_n} u(x_1, \ldots, x_n)\, p(x_1, \ldots, x_n) \quad \text{for discrete case} \\ \mathrm{E}[u(X_1, \ldots, X_n)] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_n)\, f(x_1, \ldots, x_n)\, \mathrm{d}x_n \cdots \mathrm{d}x_1 \quad \text{for continuous cas} \end{aligned} \tag{39}$$

   Key properties
   If X and Y are independent, then

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y) \text{ and } M_{X+Y}(t) = M_X(t)M_Y(t). \tag{40}$$

4. Covariance
   Let X and Y be random variables with mean μ_X and μ_Y respectively. The covariance between X and Y, denoted by σ_{XY}, is defined as

$$\begin{aligned} \mathrm{Cov}(X, Y) &= \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) \end{aligned} \tag{41}$$

   Properties:

(a) $\mathrm{Cov}(aX+c, bY+d) = ab\,\mathrm{Cov}(X,Y)$.
In general, $\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i + c_i, \sum_{j=1}^{N} b_j Y_j + d_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \,\mathrm{Cov}\left(X_i, Y_j\right)$.

(b) $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

(c) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$.
In general, $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) + 2\sum_{i<j} \mathrm{Cov}\left(X_i, X_j\right)$.

(d) If X and Y are independent, then $Cov(X,Y) = 0$.
However, $Cov(X,Y) = 0$ **does not imply** X and Y are independent. (**Very important!**)

5. Correlation Coefficient
Let X and Y be two random variables. The correlation coefficient between X and Y, denoted as $\rho_{X,Y}$ (or simply ρ) is defined as

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}. \tag{42}$$

Properties:

(a) $-1 \leq \rho \leq 1$.

(b) ρ is invariant under linear transformations of X and Y. That is,

$$\mathrm{Corr}(aX+c, bY+d) = \mathrm{sgn}(ab)\,\mathrm{Corr}(X,Y). \tag{43}$$

The sign and the magnitude of ρ reveal the direction and strength of the linear relationship between X and Y.

6. Sum of Independent Random Variables
Let $X_1, X_2, \ldots, X_n$ be independent random variables, $Y = \sum_{i=1}^{n} a_i X_i$, where a_i's are constants. Then

(a) $\mathrm{E}(Y) = \sum_{i=1}^{n} a_i \mathrm{E}\left(X_i\right)$.

(b) $\mathrm{Var}(Y) = \sum_{i=1}^{n} \mathrm{Var}\left(a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \,\mathrm{Var}\left(X_i\right)$.

(c) $M_Y(t) = \prod_{i=1}^{n} M_{X_i}\left(a_i t\right)$.

Conditional Distributions

1. Conditional Distribution and Conditional Expectation
   For any two events E and F, the conditional probability of E given F is defined by

$$\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)} \text{ provided that } \Pr(F) > 0. \tag{44}$$

   Let (X, Y) be a discrete bivariate random vector with joint pmf $\Pr(X = x, Y = y) = p(x, y)$ and marginal pmfs $p_X(x)$ and $p_Y(y)$.
   The conditional pmf of Y given that $X = x$ is the function of y denoted by $p_{Y|X}(y|x)$, where $p_X(x) > 0$

$$\begin{aligned} p_{Y|X}(y|x) &= \Pr(Y = y|X = x) \\ &= \frac{\Pr(Y = y, X = x)}{\Pr(X = x)} \\ &= \frac{p(x, y)}{p_X(x)} \end{aligned} \tag{45}$$

   If X is independent of Y , then the conditional pmf becomes

$$\begin{aligned} p_{Y|X}(y|x) &= \frac{p(x, y)}{p_X(x)} \\ &= \frac{p_X(x)p_Y(y)}{p_X(x)} \\ &= p_Y(y) \end{aligned} \tag{46}$$

   For continuous random variables, the conditional distributions are defined as:

$$\begin{aligned} f_{Y|X}(y|x) &= \tfrac{f(x,y)}{f_X(x)} \quad \text{provided that } f_X(x) > 0 \\ f_{X|Y}(x|y) &= \tfrac{f(x,y)}{f_Y(y)} \quad \text{provided that } f_Y(y) > 0 \end{aligned} \tag{47}$$

2. Definitions and Formulas:

   (a) Conditional distribution function of Y given $X = x$:

$$F_{Y|X}(y|x) = \Pr(Y \le y|X = x) = \begin{cases} \sum_{i \le y} p_{Y|X}(i|x), & \text{for discrete case} \\ \int_{-\infty}^{y} f_{Y|X}(t|x)\mathrm{d}t, & \text{for continuous case.} \end{cases} \tag{48}$$

(b) Conditional expectation of $g(Y)$ given $X = x$:

$$\mathrm{E}[g(Y)|X=x] = \begin{cases} \sum_i g(i) p_{Y|X}(i|x), & \text{for discrete case} \\ \int_{-\infty}^{\infty} g(y) f_{Y|X}(y|x) \mathrm{d}y, & \text{for continuous case.} \end{cases} \tag{49}$$

(c) Conditional mean of Y given $X = x$: $E(Y|X = x)$.

(d) Conditional variance of Y given $X = x$:

$$\mathrm{Var}(Y|X=x) = \mathrm{E}\left\{[Y - \mathrm{E}(Y|X=x)]^2 | X = x\right\} = \mathrm{E}\left(Y^2|X=x\right) - [\mathrm{E}(Y|X=x)]^2. \tag{50}$$

(e) Computing expectations by conditioning:

$$\begin{aligned} \mathrm{E}[u(X)] &= \mathrm{E}\{\mathrm{E}[u(X)|Y]\} \\ \mathrm{Var}(X) &= \mathrm{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}[\mathrm{E}(X|Y)] \end{aligned} \tag{51}$$

Transformation of Multivariate Distributions

Let $X_1, X_2, \ldots, X_n$ be jointly distributed continuous random variables with joint probability density function $f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$. Also, let $Y_i = g_i(X_1, X_2, \ldots, X_n), i = 1, 2, \ldots, n$ for some function g's which satisfy the following conditions:

1. The transformation from X's to Y's is one-one correspondence.
2. The function g's have continuous partial derivatives at all points $(x_1, x_2, \ldots, x_n)$ and the $n \times n$ **Jacobian determinant** is non-zero, i.e.,

$$J(x_1, x_2, \ldots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix} \neq 0 \tag{52}$$

at all points $(x_1, x_2, \ldots, x_n)$.

Then the joint pdf of $Y_1, Y_2, \ldots, Y_n$ is given by the following formula:

$$f_{\mathbf{Y}}(y_1, y_2, \ldots, y_n) = f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) \times |J(x_1, x_2, \ldots, x_n)|^{-1}, \tag{53}$$

where $x_i = h_i(y_1, y_2, \ldots, y_n), i = 1, 2, \ldots, n$.

In fact, for easier calculation,the Jacobian determinant $J(x_1, x_2, \ldots, x_n)$ can be determined:

$$|J(x_1, x_2, \ldots, x_n)|^{-1} = \begin{vmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} & \cdots & \frac{\partial h_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \frac{\partial h_n}{\partial y_2} & \cdots & \frac{\partial h_n}{\partial y_n} \end{vmatrix} \tag{54}$$

where h's are the inverse transformations $x_i = h_i(y_1, y_2, \ldots, y_n), i = 1, 2, \ldots, n$.

Some Important Transformations:

1. $Z = X + Y$
   Discrete case: $p_Z(z) = \Pr(X+Y = z) = \sum_x p(x, z-x) = \sum_y p(z-y, y)$

Continuous case:

$$\begin{aligned}
F_Z(z) &= \Pr(Z \le z) = \Pr(X + Y \le z) \\
&= \Pr(Y \le z - X) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y) \mathrm{d}y \mathrm{d}x \\
&= \Pr(X \le z - Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f(x, y) \mathrm{d}x \mathrm{d}y \\
f_Z(z) &= F'(z) = \int_{-\infty}^{\infty} f(x, z - x) \mathrm{d}x = \int_{-\infty}^{\infty} f(z - y, y) \mathrm{d}y
\end{aligned} \tag{55}$$

2. $Z = X - Y$
   Discrete case: $p_Z(z) = \Pr(X - Y = z) = \sum_x p(x, x - z) = \sum_y p(z + y, y)$
   Continuous case:

$$\begin{aligned}
F_Z(z) &= P(Z \le z) = P(X - Y \le z) \\
&= P(Y \le X - z) = \int_{-\infty}^{\infty} \int_{-\infty}^{x-z} f(x, y) \mathrm{d}y \mathrm{d}x \\
&= P(X \le z + Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{z+y} f(x, y) \mathrm{d}x \mathrm{d}y \\
f_Z(z) &= F'(z) = \int_{-\infty}^{\infty} f(x, x - z) \mathrm{d}x = \int_{-\infty}^{\infty} f(z + y, y) \mathrm{d}y
\end{aligned} \tag{56}$$

3. $Z = XY$
   Discrete case: $p_Z(z) = \Pr(XY = z) = \sum_x p(x, x/z) = \sum_y p(y/z, y)$
   Continuous case:

$$\begin{aligned}
F_Z(z) &= \Pr(Z \le z) = \Pr(XY \le z) \\
&= \Pr(Y \le z/X, X > 0) + \Pr(Y \ge z/X, X < 0) \\
&= \int_0^{\infty} \int_{-\infty}^{z/x} f(x, y) \mathrm{d}y \mathrm{d}x + \int_{-\infty}^{0} \int_{z/x}^{\infty} f(x, y) \mathrm{d}y \mathrm{d}x \\
f_Z(z) &= F'(z) = \int_0^{\infty} f(x, z/x) \frac{1}{x} \mathrm{d}x - \int_{-\infty}^{0} f(x, z/x) \frac{1}{x} \mathrm{d}x \\
&= \int_{-\infty}^{\infty} f(x, z/x) \left| \frac{1}{x} \right| \mathrm{d}x = \int_{-\infty}^{\infty} f(z/y, y) \left| \frac{1}{y} \right| \mathrm{d}y
\end{aligned} \tag{57}$$

4. $Z = \frac{X}{Y}$
   Discrete case: $p_Z(z) = \Pr(X/Y = z) = \sum_x p(x, xz) = \sum_y p(yz, y)$

Continuous case:

$$
\begin{aligned}
F_Z(z) &= \Pr(Z \leq z) = \Pr(X/Y \leq z) \\
&= \Pr(X \leq zY, Y > 0) + \Pr(X \geq zY, Y < 0) \\
&= \int_0^{\infty} \int_{-\infty}^{zy} f(x,y)\mathrm{d}x\mathrm{d}y + \int_{-\infty}^{0} \int_{zy}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y \\
f_Z(z) &= F'(z) = \int_0^{\infty} f(zy,y)y\mathrm{d}y - \int_{-\infty}^{0} f(zy,y)y\mathrm{d}y \\
&= \int_{-\infty}^{\infty} f(x,x/z)\frac{|x|}{z^2}\mathrm{d}x = \int_{-\infty}^{\infty} f(zy,y)|y|\mathrm{d}y
\end{aligned}
\tag{58}
$$