Chapter 1: Basic Concepts

STAT2602A Probability and statistics II
(2024-2025 1st Semester)

Contents

1.1 Discrete distribution

- **Definition:** Let x be a realization of a discrete random variable $X \in \mathcal{R}$. Then,

$$f(x) = \mathrm{P}(X = x)$$

is the discrete probability density function (discrete p.d.f.) or probability mass function of X.

- **Theorem 1.1 (Discrete univariate probability density function)** A discrete univariate probability density function has the following properties:
  (1) $f(x) > 0$ for $x \in S$;
  (2) $\sum_{x \in S} f(x) = 1$;
  (3) $\mathrm{P}(X \in A) = \sum_{x \in A} f(x)$, where $A \subset S$.

  Then,

$$f(x) = \mathrm{P}(X = x)$$

is the probability density function (p.d.f.) of X.

Based on the p.d.f. $f(x)$, we further define cumulative distribution function.

1.1 Discrete distribution

- **Definition:** we define the cumulative distribution function (c.d.f.) $F(x)$ of the discrete random variable X by

$$F(x) = \mathrm{P}(X \leq x) = \sum_{s \in S \text{ and } s \leq x} f(s).$$

  Note that $F(x)$ is a step function on $\mathcal{R}$ and the height of a step at x, $x \in S$, equals the probability $f(x)$

From Theorem 1.1, we can obtain the following about $F(x)$

- **Theorem 1.2** (Discrete cumulative distribution function) A discrete univariate cumulative distribution function has the following properties:
  (1) $0 \leq F(x) \leq 1$ for $x \in \mathcal{R}$;
  (2) $F(x)$ is a nondecreasing function of x;
  (3) $F(\infty) = 1$ and $F(-\infty) = 0$.

Remark 1.1 The discrete p.d.f. $f(x)$ and the c.d.f. $F(x)$ are one-to-one corresponding. We can first define the c.d.f. $F(x)$, and then define the discrete p.d.f. $f(x)$ by

$$f(x) = F(x) - F(x-) \quad \text{for } x \in S.$$

1.1 Discrete distribution

- Example: A discrete random variable X with p.d.f. $f(x) = P(X = x) = x/6$ for $x = 1, 2, 3$

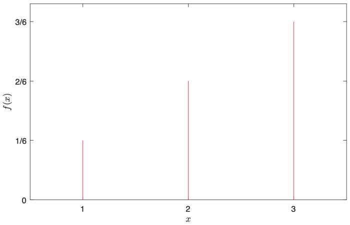


Figure 1: p.d.f

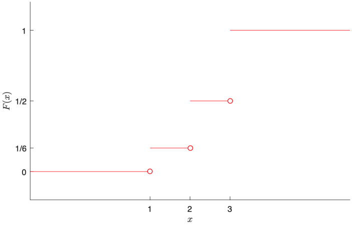


Figure 2: c.d.f.

1.1 Discrete distribution

- **Property 1.1** Two discrete random variables X and Y are independent if and only if $F(x, y) = F_X(x)F_Y(y)$ for all $(x, y) \in S$, where F is joint distribution of X and Y, and F_X (or F_Y) is the marginal distribution of X (or Y).
- **Property 1.2** Let X and Y be two independent discrete random variables. Then,
  (a) for arbitrary countable sets A and B,
  $$\mathrm{P}(X \in A, Y \in B) = \mathrm{P}(X \in A)\mathrm{P}(Y \in B);$$
  (b) for any real functions $g(\cdot)$ and $h(\cdot)$, $g(X)$ and $h(Y)$ are independent.

1.2 Continuous distribution

- **Definition:** Let $X \in \mathcal{R}$ be a continuous random variable. The probability of X lies in an interval $(a, b]$ is

$$\mathrm{P}(a < X \leq b) = \int_a^b f(x)dx$$

  for some non-negative function $f(\cdot)$. We call $f(x)$ the p.d.f. of the continuous random variable X.

- **Theorem 1.3** (Continuous univariate probability density function) A continuous univariate probability density function has the following properties:
  (1) $f(x) \geq 0$ for $x \in \mathcal{R}$;
  (2) $\int_{\mathcal{R}} f(x)dx = 1$;
  (3) $\mathrm{P}(X \in A) = \int_A f(x)dx$ for $A \subset \mathcal{R}$.

Based on the p.d.f. $f(x)$, we further define cumulative distribution function of continuous random variable.

1.2 Continuous distribution

- **Definition:** We define the cumulative distribution function of continuous random variable X by

$$F(x) = \mathrm{P}(X \leq x) = \int_{-\infty}^{x} f(s)ds,$$

  which also satisfies Theorem 1.2. From the fundamental theorems of calculus, we have $F'(x) = f(x)$ if exists. Since there are no steps or jumps in a continuous c.d.f., it must be true that $\mathrm{P}(X = b) = 0$ for all real values of b.

- *Example 1.1* (**Uniform distribution**) A random variable X has a uniform distribution if

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

  Briefly, we say that $X \sim U(a, b)$.

1.2 Continuous distribution

- **Property 1.3** If F is a continuous c.d.f. and $X \sim U(0,1)$, then $Y = F^{-1}(X) \sim F$.

Proof.

$$\mathrm{P}(Y \leq y) = \mathrm{P}(F^{-1}(X) \leq y) = \mathrm{P}(X \leq F(y)) = F(y).$$

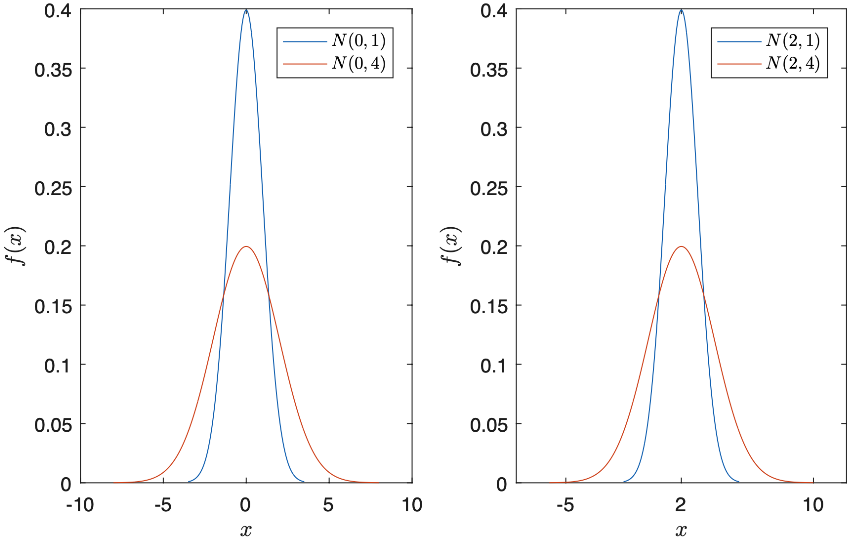
Note that this property helps us to generate a random variable from certain distribution. $\square$

- *Example 1.2* (Normal distribution) A random variable X has a normal distribution if

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \text{for } x \in \mathcal{R},$$

where $\mu \in \mathcal{R}$ is the location parameter and $\sigma > 0$ is the scale parameter. Briefly, we say that $X \sim \mathrm{N}(\mu, \sigma^2)$.

1.2 Continuous distribution



1.2 Continuous distribution

- *Example 1.2 (con't)* Further, $Z = (X - \mu)/\sigma \sim \mathrm{N}(0, 1)$ (the standard normal distribution), and the c.d.f. of Z is typically denoted by $\Phi(x)$, where

$$\Phi(x) = \mathrm{P}(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{s^2}{2}\right] ds.$$

  Numerical approximations for $\Phi(x)$ have been well tabulated in practice.

1.2 Continuous distribution

- **Property 1.4** If the p.d.f. of a continuous random variable X is $f_X(x)$ for $x \in \mathcal{R}$, the p.d.f. of $Y = aX + b$ for $a > 0$ is $f_Y(x) = \frac{1}{a} f_X(\frac{x-b}{a})$ for $x \in \mathcal{R}$.
- **Quick proof of Property 1.4** Let $F_X(x)$ be the c.d.f. of X. Then, the c.d.f. of Y is

$$F_Y(x) = \mathrm{P}(Y = aX+b \leq x) = \mathrm{P}\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right)$$

for $x \in \mathcal{R}$. Hence,

$$f_Y(x) = F_Y'(x) = \frac{1}{a} F_X'\left(\frac{x-b}{a}\right) = \frac{1}{a} f_X\left(\frac{x-b}{a}\right).$$

This completes the proof. $\square$

1.2 Continuous distribution

- **Property 1.5** Two continuous random variables X and Y are independent if and only if

$$F(x,y) = F_X(x)F_Y(y) \quad \text{for all } (x,y) \in R^2.$$

- **Property 1.6** Let X and Y be two independent continuous random variables. Then,
(a) for arbitrary intervals A and B,

$$\mathrm{P}(X \in A, Y \in B) = \mathrm{P}(X \in A)\mathrm{P}(Y \in B);$$

(b) for any real functions $g(\cdot)$ and $h(\cdot)$, $g(X)$ and $h(Y)$ are independent.

1.3 Empirical distribution

- **Brief Introduction:** Suppose that $X \sim F(x)$ is a random variable resulting from a random experiment. Repeat this experiment n independent times, we get n random variables $X_1, \cdots, X_n$ associated with these outcomes. The collection of these random variables is called a sample from a distribution with c.d.f. $F(x)$ (or p.d.f. $f(x)$). The number n is called the sample size.

As all random variables in a sample follow the same c.d.f. as X, we expect that they can give us the information about the c.d.f of X. Next, we are going to show that the empirical distribution of $\{X_1, \cdots, X_n\}$ is close to $F(x)$ in some probability sense.

1.3 Empirical distribution

- **Definition:** The empirical distribution of $\{X_1, \cdots, X_n\}$ is defined as

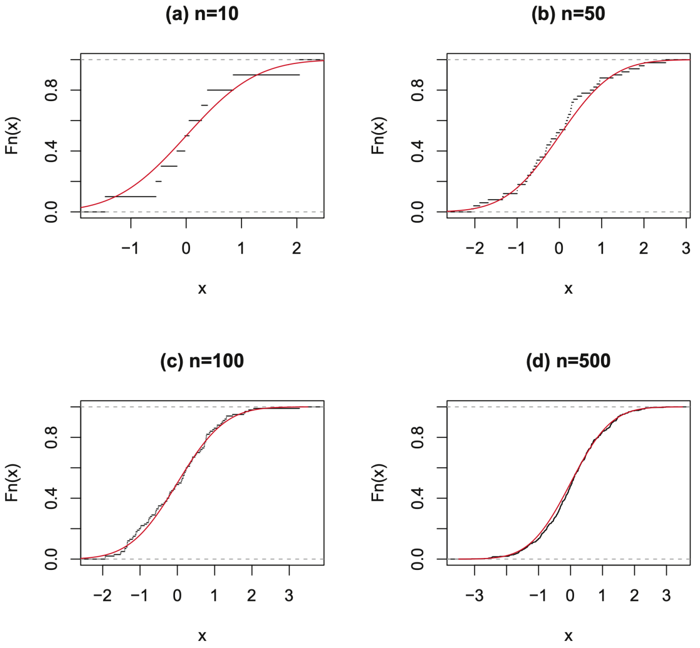
$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathrm{I}(X_k \le x)$$

for $x \in \mathcal{R}$, where $\mathrm{I}(A)$ is an indicator function such that $\mathrm{I}(A) = 1$ if A holds and $\mathrm{I}(A) = 0$ otherwise.

$F_n(x)$ assigns the probability $1/n$ to each X_k, and we can check that it satisfies Theorem 1.2 (please do it by yourself). Since $F_n(x)$ is the relative frequency of the event $X \le x$, it is an approximation of the probability $\mathrm{P}(X \le x) = F(x)$. Thus, the following result is expected.

1.3 Empirical distribution

- **Theorem 1.4** As $n \to \infty$, $\sup_{x \in \mathcal{R}} |F_n(x) - F(x)| \to 0$ almost surely (a.s.).
  Remark The proof of aforementioned theorem is omitted. A rough explanation of this theorem could be found in the lecture notes. More details about convergence almost surely can be found in the book *Probability: Theory and examples* by Rick Durrett.
- $F_n(x)$ provides an estimate of the c.d.f. $F(x)$ for each realization $x_1, \ldots, x_n$.
- **Example:** In the below figure, the black step function is the empirical distribution function $F_n(x)$ based on a realization $x_1, \ldots, x_n$ with $X_i \sim N(0, 1)$. The red solid line is the c.d.f. $\Phi(x)$ of $N(0, 1)$.


(a) n=10
(b) n=50
(c) n=100
(d) n=500
Fn(x)
x
0.0
0.4
0.8
−3
−2
−1
0
1
2
3

1.3 Empirical distribution

- **Example 1.3** Let X denote the number of observed heads when four coins are tossed independently and at random. Recall that the distribution of X is $\mathrm{B}(4, 1/2)$. One thousand repetitions of this experiment (actually simulated on the computer) yielded the following results:

Number of heads	Frequency
0	65
1	246
2	358
3	272
4	59

1.3 Empirical distribution

- **Example 1.3 (con't)** This information above determines the following empirical distribution function:

x	$F_{1000}(x)$	x	$F_{1000}(x)$
$(-\infty, 0)$	0.000	$[2, 3)$	0.669
$[0, 1)$	0.065	$[3, 4)$	0.941
$[1, 2)$	0.311	$[4, \infty)$	1.000

  The graph of the empirical distribution function $F_{1000}(x)$ and the theoretical distribution function $F(x)$ for the binomial distribution are very close (please check it by yourself).

- **Example 1.4** The following numbers are a random sample of size 10 from some distribution:

  $$-0.49,\ 0.90,\ 0.76,\ -0.97,\ -0.73,\ 0.93,\ -0.88,\ -0.75,\ 0.88,\ 0.96.$$

  (a) Write done the empirical distribution; (b) use the empirical distribution to estimate $\mathrm{P}(X \leq -0.5)$ and $\mathrm{P}(-0.5 \leq X \leq 0.5)$.

1.3 Empirical distribution

- **Example 1.4 Solution** Order the random sample:

  $-0.97, -0.88, -0.75, -0.73, -0.49, 0.76, 0.88, 0.90, 0.93, 0.96.$

  Then, the empirical distribution function $F_{10}(x)$ is as follows:

x	$F_{10}(x)$	x	$F_{10}(x)$
$(-\infty, -0.97)$	0.0	$[-0.49, 0.76)$	0.5
$[-0.97, -0.88)$	0.1	$[0.76, 0.88)$	0.6
$[-0.88, -0.75)$	0.2	$[0.88, 0.90)$	0.7
$[-0.75, -0.73)$	0.3	$[0.90, 0.93)$	0.8
$[-0.73, -0.49)$	0.4	$[0.93, 0.96)$	0.9
		$[0.96, \infty)$	1.0

  Thus, $\mathrm{P}(X \leq -0.5) = F(-0.5) \approx F_{10}(-0.5) = 0.4$ and $\mathrm{P}(-0.5 \leq X \leq 0.5) = F(0.5) - F(-0.5) \approx F_{10}(0.5) - F_{10}(-0.5) = 0.5 - 0.4 = 0.1.$ $\square$

1.3 Empirical distribution

- **Question:** How to estimate the p.d.f. $f(x)$
  Explanation: Apply relative frequency histogram. For the discrete random variable X, we can estimate $f(x) = \mathrm{P}(X = x)$ by the relative frequency of occurrences of x. That is,

$$f(x) \approx f_n(x) = \frac{\sum_{k=1}^{n} \mathrm{I}(X_k = x)}{n}.$$

- **Reuse of example 1.3** The relative frequency of observing $x = 0, 1, 2, 3$ or 4 is listed in the second column, and it is close to the value of $f(x)$, which is the p.d.f of $\mathrm{B}(4, 1/2)$.

x	$f_{1000}(x)$	$f(x)$
0	0.065	0.0625
1	0.246	0.2500
2	0.358	0.3750
3	0.272	0.2500
4	0.059	0.0625

  By increasing the value of n, the difference between $f_n(x)$ and $f(x)$ will become small.

1.3 Empirical distribution - Relative frequency histogram

The following points explain how to define relative frequency histogram and its properties.

- **Class intervals** For the continuous random variable X, we first define the so-called class intervals. Choose an integer $l \geq 1$, and a sequence of real numbers $c_0, c_1, \cdots, c_l$ such that $c_0 < c_1 < \cdots < c_l$. The class intervals are

$$(c_0, c_1],\ (c_1, c_2], \cdots, (c_{l-1}, c_l].$$

Roughly speaking, the class intervals are a non-overlapped partition of the interval $[X_{\min}, X_{\max}]$. As $f(x) = F'(x)$, we expect that when c_{j-1} and c_j are close,

$$f(x) \approx \frac{F(c_j) - F(c_{j-1})}{c_j - c_{j-1}} \quad \text{for } x \in (c_{j-1}, c_j],\ \ j = 1, 2, \cdots, l.$$

1.3 Empirical distribution - Relative frequency histogram

Note that

$$F(c_j) - F(c_{j-1}) = \mathrm{P}(X \in (c_{j-1}, c_j]) \approx \frac{\sum_{k=1}^{n} \mathrm{I}(X_k \in (c_{j-1}, c_j])}{n}$$

is the relative frequency of occurrences of $X_k \in (c_{j-1}, c_j]$. Thus, we can approximate $f(x)$ by

$$f(x) \approx h_n(x) = \frac{\sum_{k=1}^{n} \mathrm{I}(X_k \in (c_{j-1}, c_j])}{n(c_j - c_{j-1})},$$

for $x \in (c_{j-1}, c_j]$, $j = 1, 2, \cdots, l$.
We call $h_n(x)$ the relative frequency histogram.
Clearly, the way that we define the class intervals is not unique, and hence the value of $h_n(x)$ is not unique. When the sample size n is large and the length of the class interval is small, $h_n(x)$ is expected to be a good estimate of $f(x)$.

1.3 Empirical distribution - Relative frequency histogram

- **Properties of $h_n(x)$**
  (i) $h_n(x) \geq 0$ for all x;
  (ii) The total area bounded by the x axis and below $h_n(x)$ equals one, i.e.,
  $$\int_{c_0}^{c_l} h_n(x)dx = 1;$$
  (iii) The probability for an event A, which is composed of a union of class intervals, can be estimated by the area above A bounded by $h_n(x)$, i.e.,
  $$\mathrm{P}(A) \approx \int_A h_n(x)dx.$$

1.3 Empirical distribution - Relative frequency histogram

- *Example 1.5* A random sample of 50 college-bound high school seniors yielded the following high school cumulative GPA's.

3.77	2.78	3.40	2.20	3.26
3.00	2.85	2.65	3.08	2.92
3.69	2.83	2.75	3.97	2.74
2.90	3.38	2.38	2.71	3.31
3.92	3.29	4.00	3.50	2.80
3.57	2.84	3.18	3.66	2.86
2.81	3.10	2.84	2.89	2.59
2.95	2.77	3.90	2.82	3.89
2.83	2.28	3.20	2.47	3.00
3.78	3.48	3.52	3.20	3.30

  (a) Construct a frequency table for these 50 GPA's using 10 intervals of equal length with $c_0 = 2.005$ and $c_{10} = 4.005$.
  (b) Construct a relative frequency histogram for the grouped data.
  (c) Estimate $f(3)$ and $f(4)$.

1.3 Empirical distribution - Relative frequency histogram

- **Example 1.5 Solution** (a) and (b). The frequency and the relative frequency histogram based on the class intervals are given in the following table:

C.I.	F	R.F. H.	C.I.	F	R.F. H.
(2.005, 2.205]	1	0.1	(3.005, 3.205]	5	0.5
(2.205, 2.405]	2	0.2	(3.205, 3.405]	6	0.6
(2.405, 2.605]	2	0.2	(3.405, 3.605]	4	0.4
(2.605, 2.805]	7	0.7	(3.605, 3.805]	4	0.4
(2.805, 3.005]	14	1.4	(3.805, 4.005]	5	0.5

C.I. = Class Interval; F. = Frequency; R.F.H. = Relative Frequency Histogram

(c) As $3 \in (2.805, 3.005]$ and $4 \in (3.805, 4.005]$,

$$f(3) \approx h_{50}(3) = \frac{14}{50 \times (3.005 - 2.805)} = 1.4,$$

$$f(4) \approx h_{50}(4) = \frac{5}{50 \times (4.005 - 3.805)} = 0.5.$$

1.4 Expectation

- **Definition:** (*Expectation of a discrete statistic*) If $u(X)$ is a function of a discrete random variable X that has a p.d.f. $f(x)$, then

$$\mathrm{E}[u(X)] = \sum_{x \in S} u(x) f(x),$$

  where the summation is taken over all possible values of x. If $\mathrm{E}[u(X)]$ exists, it is called the mathematical expectation (or expected value) of $u(X)$.
- *Remark 1.2* $\mathrm{E}[u(X)]$ exists if $\sum_{x \in S} |u(x)| f(x) < \infty$.
- **Uncorrelation:** Two random variables X_1 and X_2 are uncorrelated, if $\mathrm{Cov}(X_1, X_2) = 0$, where

$$\mathrm{Cov}(X_1, X_2) = \mathrm{E}(X_1 X_2) - \mathrm{E}(X_1)\mathrm{E}(X_2)$$

  is the covariance of X_1 and X_2.

1.4 Expectation

- **Property 1.7** Let X be a discrete random variable with finite mean $\mathrm{E}(X)$, and let a and b be constants. Then,
(i) $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$;
(ii) if $\mathrm{P}(X = b) = 1$, then $\mathrm{E}(X) = b$;
(iii) if $\mathrm{P}(a < X \leq b) = 1$, then $a < \mathrm{E}(X) \leq b$;
(iv) if $g(X)$ and $h(X)$ have finite mean, then

$$\mathrm{E}(g(X) + h(X)) = \mathrm{E}(g(X)) + \mathrm{E}(h(X)).$$

- **Property 1.8** If $X \geq 0$ takes integer values, then

$$\mathrm{E}(X) = \sum_{x=1}^{\infty} \mathrm{P}(X \geq x) = \sum_{x=0}^{\infty} \mathrm{P}(X > x)$$

.

1.4 Expectation

- **Definition:** (*Expectation of a continuous statistic*) If $u(X)$ is a function of a continuous random variable that has a p.d.f. $f(x)$, then
$$\mathrm{E}[u(X)] = \int_{\mathcal{R}} u(x)f(x)\mathrm{d}x.$$
If $\mathrm{E}[u(X)]$ exists, it is called the mathematical expectation (or expected value) of $u(X)$.
- *Remark 1.3* $\mathrm{E}[u(X)]$ exists if
$$\int_{\mathcal{R}} |u(x)|f(x)\mathrm{d}x < \infty.$$

1.4 Expectation

- *Example 1.6* Let X have the $\mathrm{N}(\mu, \sigma^2)$ distribution. Then,

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \int_{-\infty}^{\infty} \frac{s\sigma + \mu}{\sqrt{2\pi}\sigma} \exp\left[-\frac{s^2}{2}\right] d(s\sigma + \mu) \quad \text{(letting } s = (x-\mu)/\sigma\text{)} \\
&= \int_{-\infty}^{\infty} \frac{s\sigma}{\sqrt{2\pi}} \exp\left[-\frac{s^2}{2}\right] ds + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{s^2}{2}\right] ds.
\end{aligned}$$

The first integrand is an odd function, and so the integral over $\mathcal{R}$ is zero. The second integrand is one by some algebra. Hence, $\mathrm{E}(X) = \mu$.

1.4 Expectation

- **Property 1.9** Let X be a continuous random variable, a and b be constants, and g and h be functions. Then,
(i) if $g(X)$ and $h(X)$ have finite mean then

$$\mathrm{E}(ag(X) + bh(X)) = a\mathrm{E}(g(X)) + b\mathrm{E}(h(X));$$

(ii) if $\mathrm{P}(a \leq X \leq b) = 1$, then $a \leq \mathrm{E}(X) \leq b$;
(iii) if h is non-negative, then for $a > 0$,
$\mathrm{P}(h(X) \geq a) \leq \mathrm{E}(h(X)/a)$;
(iv) if g is convex, then $g(\mathrm{E}(X)) \leq \mathrm{E}(g(X))$.
- **Property 1.10** Let X be a non-negative random variable with c.d.f. F, p.d.f f, and finite expected value $\mathrm{E}(X)$. Then,

$$\mathrm{E}(X) = \int_0^\infty (1 - F(x))dx.$$

1.4 Expectation

- **Property 1.11** Let a, b, c, and d be constants. Then,
  (i) $E(X^2) = 0$ if and only if $\mathrm{P}(X = 0) = 1$;
  (ii) $\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$;
  (iii) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$;
  (iv) if X and Y are independent,
  $\mathrm{E}(h(X)g(Y)) = \mathrm{E}(h(X))\mathrm{E}(g(Y))$ provided that
  $\mathrm{E}(h(X)) < \infty$ and $\mathrm{E}(g(Y)) < \infty$.
  (v) $-1 \leq \rho(X, Y) \leq 1$;
  (vi) $|\rho(X, Y)| = 1$ if and only if $\mathrm{P}(X = aY + b) = 1$ for some constants a and b;
  (vii) $\rho(aX + b, cY + d) = \mathrm{sgn}(ac)\rho(X, Y)$, where $\mathrm{sgn}(x)$ denotes the sign of x;
  (viii) if X and Y are independent, $\rho(X, Y) = 0$.

- **Property 1.12** (Cauchy-Schwarz inequality) If $\mathrm{E}(X^2)\mathrm{E}(Y^2) < \infty$, then

$$\mathrm{E}(XY) \leq \sqrt{\mathrm{E}(X^2)\mathrm{E}(Y^2)}.$$

Quick Proof: Without loss generality, we assume that $\mathrm{E}(Y^2) > 0$. Note that

$$0 \leq \mathrm{E}\left[(X\mathrm{E}(Y^2) - Y\mathrm{E}(XY))^2\right] = \mathrm{E}(Y^2)\left[\mathrm{E}(X^2)\mathrm{E}(Y^2) - (\mathrm{E}(XY))^2\right].$$

Hence, the conclusions holds. $\square$