Chapter 2 Discrete Distributions

- Random Variables of the Discrete Type
  - Uniform Distribution
  - Hypergeometric Distribution
- Mathematical Expectation
- Moment Generating Function
- Bernoulli Trials and the Binomial Distribution
- Geometric and Negative Binomial Distribution
- The Poisson Distribution

2.1 Random Variables of the Discrete Type

- Def 2.1-1: Random Variable X (abbreviated by r.v.)
  - a **function** that maps the possible outcomes of an experiment to real numbers.（把實驗結果轉成實數）
  - i.e., assign to each element s in S one and only one real number
  - Notice that: *a function assigns one and only one number in the range (output) to each number in the domain (input)*
  - $X: S \rightarrow R$, where S is the set of all outcomes of an experiment, and R is the set of real numbers.

- The space of X is the set of real numbers S_X, where

  $$S_X = \{x : X(s) = x , s \in S\}$$

  From now on, S_X is replaced by S in this class

Eexample: Drawing in the Yearend Party

- *X* and *Y* map some outcomes to different real numbers.
- However, the spaces of *X* and *Y*
  - are identical
  - both are {0, 10000, 50000, 100000}.
- The probability functions of *X* and *Y* are also equal.

$Prob(X{=}10{,}000) = Prob(Y{=}10{,}000) = 0.01$

$Prob(X{=}50{,}000) = Prob(Y{=}50{,}000) = 0.01$

$Prob(X{=}100{,}000) = Prob(Y{=}100{,}000) = 0.01$

$Prob(X{=}0) = Prob(Y{=}0) = 0.97$

Probability Mass Function

(Probability Function)

- The probability mass function (p.m.f.) of a discrete random variable X is defined to be

$$P_X(k) \equiv Prob(X = k) = \sum_{q \in Q_k} Prob(q),$$

where Q_k contains all outcomes that are mapped to k by random variable X.

- In the previous example of drawing,

$$P_X(10{,}000) = Prob(X = 10{,}000)$$

$$= \sum_{\substack{<10,i,j> \\ i \neq 10,\, j \neq 10 \\ i \neq j}} Prob(<10, i, j>)$$

$$= \sum_{\substack{<10,i,j> \\ i \neq 10,\, j \neq 10 \\ i \neq j}} \frac{1}{100 \times 99 \times 98} = 0.01.$$

$$P_X(k) \equiv Prob(X = k) = \sum_{q \in Q_k} Prob(q),$$

where Q_k contains all outcomes that are mapped to k by random variable X.

- In fact, the p.m.f. of a random variable is defined on a set of events of the experiment conducted.
  - $f(x) \equiv P(X=x)$ as shown in the Hogg's textbook
- In the previous drawing example, the set of outcomes that are mapped to 10,000 by X is an event.

Properties of Probability Mass Function

$$P_X(k) \equiv Prob(X = k) = \sum_{q \in Q_k} Prob(q),$$

- The p.m.f. of a random variable X satisfies the following three properties:

$$(1)\ \ P_X(x) > 0\ \ ,\ \ x \in S\ :\ \ the\ \ space\ \ of\ \ X.$$

$$(2)\ \ \sum_{x_i \in S} P_X(x_i) = 1\ .$$

$$(3)\ \ Prob(A) = \sum_{x_j \in A} P_X(x_j)\ ,\ \ where\ \ A \subseteq S.$$

課本將以上的Properties視為定義，如下一頁：

For a random variable X of the discrete type, the probability $P(X = x)$ is frequently denoted by $f(x)$, and this function $f(x)$ is called the **probability mass function**. Note that some authors refer to $f(x)$ as the probability function, the frequency function, or the probability density function. In the discrete case, we shall use "probability mass function," and it is hereafter abbreviated pmf.

Let $f(x)$ be the pmf of the random variable X of the discrete type, and let S be the space of X. Since $f(x) = P(X = x)$ for $x \in S$, $f(x)$ must be nonnegative for $x \in S$, and we want all these probabilities to add to 1 because each $P(X = x)$ represents the fraction of times x can be expected to occur. Moreover, to determine the probability associated with the event $A \in S$, we would sum the probabilities of the x values in A. This leads us to the following definition.

?

Definition 2.1-2

The pmf $f(x)$ of a discrete random variable X is a function that satisfies the following properties:

(a) $f(x) > 0, \qquad x \in S;$

(b) $\displaystyle\sum_{x \in S} f(x) = 1;$

(c) $\displaystyle P(X \in A) = \sum_{x \in A} f(x), \qquad \text{where } A \subset S.$

課本第29-30頁

?

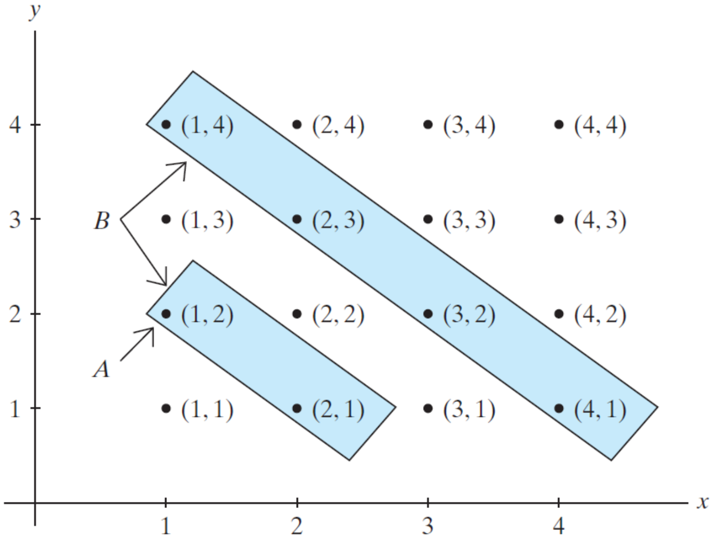

Figure 1.3-2 Dice example

Example 1.3-4 A pair of fair four-sided dice is rolled and the sum is determined. Let A be the event that a sum of 3 is rolled, and let B be the event that a sum of 3 or a sum of 5 is rolled. In a sequence of rolls, the probability that a sum of 3 is rolled before a sum of 5 is rolled can be thought of as the conditional probability of a sum of 3 given that a sum of 3 or 5 has occurred; that is, the conditional probability of A given B is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{2/16}{6/16} = \frac{2}{6}.$$

Note that for this example, the only outcomes of interest are those having a sum of 3 or a sum of 5, and of these six equally likely outcomes, two have a sum of 3. (See Figure 1.3-2 and Exercise 1.3-13.)

Def: Cumulative Distribution Function

(Probability Distribution Function)

(Distribution Function)

- For a random variable X, we define its cumulative distribution function F as

$$F_X(t) = Prob(X \le t)$$

-- Cumulative Distribution Function, CDF

Properties of a Cumulative Distribution Function

$$1.\ \lim_{t\to\infty} F_X(t) = 1.$$

$$2.\ \lim_{t\to-\infty} F_X(t) = 0.$$

$$3.\ F_X(w) \geq F_X(t),\ if\ w \geq t.$$

- Is there any function that satisfies these conditions not a distribution function?
- *Any function that satisfies these conditions above can be a distribution function.*

Example 1.3-4

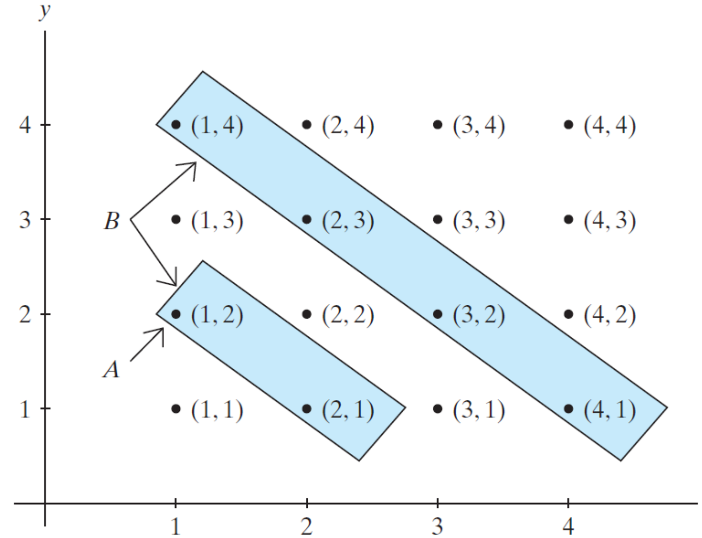

Figure 1.3-2 Dice example

- Assume that we toss a 4-sided die twice. Then, we have 16 possible outcomes：

$$\left\{ \begin{array}{l} (1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(2,4), \\ (3,1),(3,2),(3,3),(3,4),(4,1),(4,2),(4,3),(4,4) \end{array} \right\}$$

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

- Let random variable X be the sum of the outcome. Then,

$$\text{Prob}(X=2)=\frac{1}{16}\ , \quad \text{Prob}(X=3)=\frac{2}{16}$$

$$\text{Prob}(X=4)=\frac{3}{16}\ , \quad \text{Prob}(X=5)=\frac{4}{16}$$

$$\text{Prob}(X=6)=\frac{3}{16}\ , \quad \text{Prob}(X=7)=\frac{2}{16}$$

$$\text{Prob}(X=8)=\frac{1}{16}.$$

$$F_X(5)=\text{Prob}(X\le 5)=\frac{1}{16}+\frac{2}{16}+\frac{3}{16}+\frac{4}{16}=\frac{5}{8}.$$

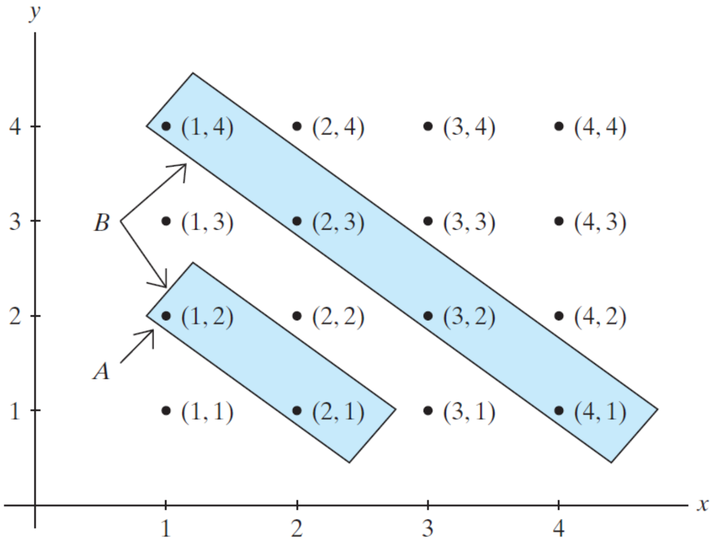

Figure 1.3-2 Dice example

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

Example 2.1-3 Roll a fair four-sided die twice, and let X be the maximum of the two outcomes. The outcome space for this experiment is $S_0 = \{(d_1, d_2) : d_1 = 1, 2, 3, 4; d_2 = 1, 2, 3, 4\}$, where we assume that each of these 16 points has probability 1/16. Then $P(X = 1) = P[(1, 1)] = 1/16$, $P(X = 2) = P[\{(1, 2), (2, 1), (2, 2)\}] = 3/16$, and similarly $P(X = 3) = 5/16$ and $P(X = 4) = 7/16$. That is, the pmf of X can be written simply as

$$f(x) = P(X = x) = \frac{2x - 1}{16}, \qquad x = 1, 2, 3, 4. \tag{2.1-1}$$

We could add that $f(x) = 0$ elsewhere; but if we do not, the reader should take $f(x)$ to equal zero when $x \notin S = \{1, 2, 3, 4\}$.

$$\left\{\begin{matrix}(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(2,4), \\ (3,1),(3,2),(3,3),(3,4),(4,1),(4,2),(4,3),(4,4)\end{matrix}\right\}$$

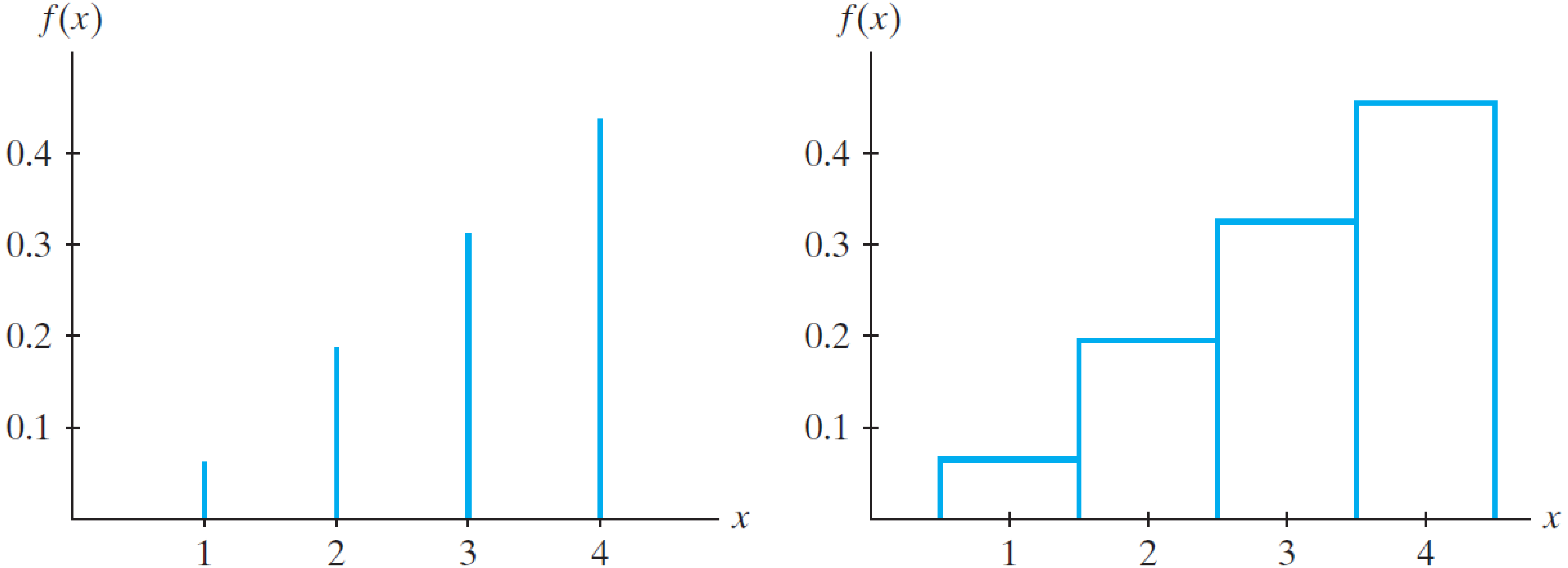

Figure 2.1-1 Line graph and probability histogram

Operations of Random Variables

- Let X and Y be two random variables defined on the same outcome space of an experiment.
- Then, we can define a new random variable $Z = f(X,Y)$.

Operations of Random Variables

- For example, in the example of drawing, if Edward and Grace are husband and wife, then we can define a new random variable $Z=X+Y$.
- We have

$X(<30, 10, *>) = 50,000$

$Y(<30, 10, *>) = 10,000$

$Z(<30, 10, *>) = 60,000$

?

Uniform distribution

$$f(x) = 1/m, \qquad x = 1, 2, \ldots, m$$

Example 2.1-2 The cast of a die: $f(x) = 1/6$

Example 2.1-2 Let the random experiment be the cast of a die. Then the outcome space associated with this experiment is $S = \{1, 2, 3, 4, 5, 6\}$, with the elements of S indicating the number of spots on the side facing up. For each $s \in S$, let $X(s) = s$. The space of the random variable X is then $\{1, 2, 3, 4, 5, 6\}$.

If we associate a probability of 1/6 with each outcome, then, for example, $P(X = 5) = 1/6$, $P(2 \le X \le 5) = 4/6$, and $P(X \le 2) = 2/6$ seem to be reasonable assignments, where, in this example, $\{2 \le X \le 5\}$ means $\{X = 2, 3, 4, \text{ or } 5\}$ and $\{X \le 2\}$ means $\{X = 1 \text{ or } 2\}$.

Hypergeometric distribution （超幾何分布）

-- select n objects from N_1+N_2 objects
(e.g., red chips and blue chips)

$$\boxed{x \le n,\ x \le N_1, \text{ and } n - x \le N_2}$$

the probability of selecting exactly x red chips is ？？

$$f(x) = P(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}$$

若n=1，超幾何分布可以簡化為？？？？

Hypergeometric distribution

-- select n objects from N_1+N_2 objects
(e.g., red chips and blue chips)

$$f(x) = P(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}$$

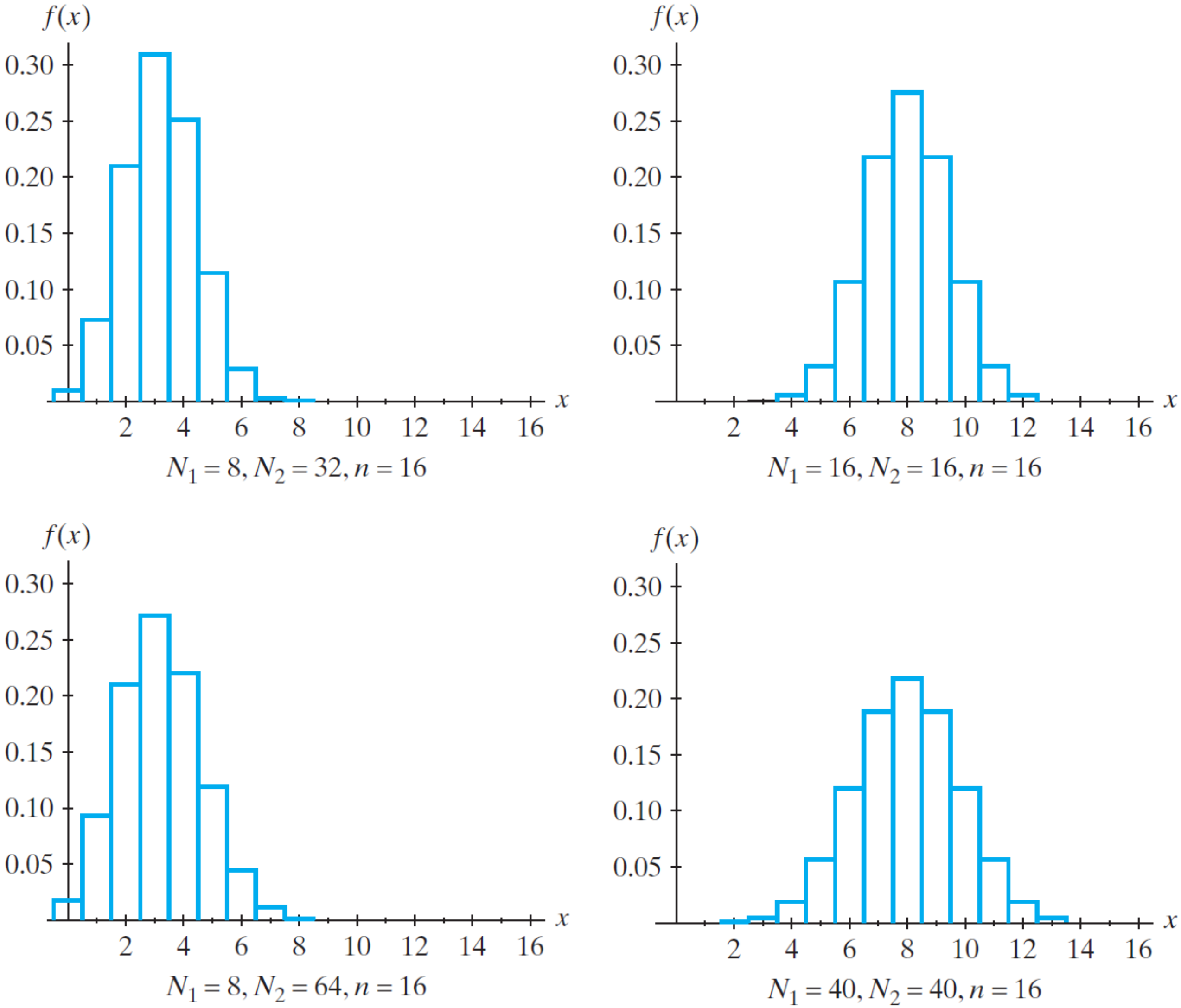

Figure 2.1-2 Hypergeometric probability histograms

Chapter 2 Discrete Distributions

- Random Variables of the Discrete Type
  - Uniform Distribution
  - Hypergeometric Distribution
- Mathematical Expectation
- Moment Generating Function
- Bernoulli Trials and the Binomial Distribution
- Geometric and Negative Binomial Distribution
- The Poisson Distribution

2.2 Mathematical Expectation

Expected Value of a Discrete Random Variable

- Let X be a discrete random variable and S be its space. Then, the expected value of X is

$$E[X] \equiv \sum_{x_i \in S} P_X(x_i)\, x_i$$

- μ is a widely used symbol for expected value.

Example 2.2-1

- 1 – pay $1
- 2 – pay $1
- 3 – pay $1
- 4 – pay $2
- 5 – pay $2
- 6 – pay $3

Charge $?

Example 2.2-1

- 1 – pay \$1
- 2 – pay \$1
- 3 – pay \$1
- 4 – pay \$5
- 5 – pay \$5
- 6 – pay \$35

Charge \$?

Expected Value of a Function of a Random Variable

- Let X be a random variable and *u*(.) be a function. Then, the expected value of random variable $Y = u(X)$ is equal to

$$E[Y] = E[u(X)] \overset{?}{\equiv} \sum_{x_i \in S} u(x_i) P_X(x_i)$$

consistency

In textbook, $E[u(X)] \equiv \sum_{x \in S} u(x) f(x)$

Expected Value of a Function of a Random Variable

- Proof：

$$E[Y] = \sum_{y_i \in S'} P_Y(y_i) y_i \ \ , \ where\ S'\ is\ the\ space\ of\ Y.$$

$$= \sum_{y_i \in S'} \mathrm{Prob}(Y = y_i)\, y_i$$

$$= \sum_{y_i \in S'} \sum_{\substack{all\ x_j \\ such\ that \\ u(x_j) = y_i}} \mathrm{Prob}(X = x_j)\, u(x_j)$$

$$= \sum_{x_j \in S} P_X(x_j) u(x_j).$$

[Ross,9e] p.133

Definition 2.2-1

If $f(x)$ is the pmf of the random variable X of the discrete type with space S, and if the summation

$$\sum_{x \in S} u(x)f(x), \qquad \text{which is sometimes written} \qquad \sum_{S} u(x)f(x),$$

exists, then the sum is called the **mathematical expectation** or the **expected value** of $u(X)$, and it is denoted by $E[u(X)]$. That is,

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

As a proposition is other books, e.g., Ross.

Remark The usual definition of mathematical expectation of $u(X)$ requires that the sum converge absolutely; that is,

$$\sum_{x \in S} |u(x)|\, f(x)$$

converges and is finite. However, in this book, each $u(x)$ is such that the convergence is absolute, and we do not burden the student with this additional requirement. Moreover, sometimes $E[u(X)]$ is called, more simply, the expectation of $u(X)$. ◆

There is another important observation that must be made about the consistency of this definition. Certainly, this function $u(X)$ of the random variable X is itself a random variable, say Y. Suppose that we find the p.m.f. of Y to be $g(y)$ on the support S_1. Then $E(Y)$ is given by the summation

$$\sum_{y \in S_1} y\, g(y).$$

In general, it is true that

$$\sum_{x \in S} u(x) f(x) = \sum_{y \in S_1} y\, g(y);$$

that is, the same expectation is obtained by either method. We do not prove this general result but only illustrate it in the following example.

see Example 2.2-2 in the text book （next page）

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

$$\boxed{\sum_{x \in S_X} u(x)f(x) = \sum_{y \in S_Y} yg(y),}$$

Example 2.2-2 Let the random variable X have the pmf

$$f(x) = \frac{1}{3}, \qquad x \in S_X,$$

where $S_X = \{-1, 0, 1\}$. Let $u(X) = X^2$. Then

$$\boxed{\sum_{x \in S_X} u(x)f(x)} \qquad E(X^2) = \sum_{x \in S_X} x^2 f(x) = (-1)^2\left(\frac{1}{3}\right) + (0)^2\left(\frac{1}{3}\right) + (1)^2\left(\frac{1}{3}\right) = \frac{2}{3}.$$

$$\boxed{\sum_{y \in S_Y} y g(y)}$$

However, the support of the random variable $Y = X^2$ is $S_Y = \{0, 1\}$ and

$$P(Y = 0) = P(X = 0) = \frac{1}{3},$$
$$P(Y = 1) = P(X = -1) + P(X = 1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

That is,

$$g(y) = \begin{cases} \frac{1}{3}, & y = 0, \\ \frac{2}{3}, & y = 1; \end{cases}$$

and $S_Y = \{0, 1\}$. Hence,

$$\mu_Y = E(Y) = \sum_{y \in S_Y} y\, g(y) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{2}{3}\right) = \frac{2}{3},$$

which again illustrates the preceding observation. ■

- Example:
  - let X correspond to the outcome of tossing a fair die once.
  - Then, $P_x(1)=P_x(2)=P_x(3)=P_x(4)=P_x(5)=P_x(6)=1/6$. and $E[X]=$ ……

- If we are concerned about the difference between the observed outcome and the mean.
  - Define $Y=|X-E[X]|$
  - then $E[Y] =$?

- Example:
  - let X correspond to the outcome of tossing a fair die once.
  - Then, $P_x(1)=P_x(2)=P_x(3)=P_x(4)=P_x(5)=P_x(6)=1/6$. and $E[X]=7/2=3.5$
- If we are concerned about the difference between the observed outcome and the mean.
  - Define $Y=|X-E[X]|$
  - $\boxed{\begin{aligned} E[\,|X-E[X]|\,] &= (5/2)*(1/6)+(3/2)*(1/6)+(1/2)*(1/6)+\ldots \\ &= 3/2 \end{aligned}}$

$$\boxed{\sum_{x \in S_X} u(x) f(x)}$$ Method 1

Method 2:

$P_Y(1/2)=1/3$, $P_Y(3/2)=1/3$, $P_Y(5/2)=1/3$.

$$\boxed{\sum_{y \in S_Y} y g(y)}$$

$$E[Y] = (1/2)*(1/3)+(3/2)*(1/3)+(5/2)*(1/3)$$
$$= 3/2$$

$$\boxed{\sum_{x \in S_X} u(x) f(x)}$$

Method 1

$$\boxed{\begin{aligned} E[\,|X-E[X]|\,] &= (5/2)*(1/6)+(3/2)*(1/6)+(1/2)*(1/6)+\ldots \\ &= 3/2 \end{aligned}}$$

Same!

Theorem 2.2-1

Theorems about the Expected Value

(a) If c is a constant, $E[c] = c.$

(b) If c is a constant and u is a function,

$$E[c\,u(X)] = cE[u(X)]$$

(c) If c_1 and c_2 are constants

and u_1 and u_2 are functions,

then $E[c_1\,u_1(X) + c_2\,u_2(X)] = c_1 E[u_1(X)] + c_2 E[u_2(X)].$

-- a linear operator, *a distributed operator*

Theorems about the Expected Value

- Proof of （a）：

Trivial.

- Proof of （b）：

$$E\left[cu(X)\right] = \sum_{x_i \in S} cu(x_i) P_X(x_i), \qquad \textit{where } S \textit{ is the space of } X$$

$$\textit{and } P_X(x) \textit{ is the p.m.f of } X.$$

$$= c \sum_{x_i \in S} u(x_i) P_X(x_i)$$

$$= cE\left[u(X)\right]$$

Theorems about the Expected Value

- Pr

$$E\left[c_1u_1(X)+c_2u_2(X)\right]=\sum_{x_i\in S}\left[c_1u_1(x_i)+c_2u_2(x_i)\right]P_X(x_i)$$
$$=\sum_{x_i\in S}c_1u_1(x_i)P_X(x_i)+\sum_{x_i\in S}c_2u_2(x_i)P_X(x_i)$$
$$=c_1E\left[u_1(X)\right]+c_2E\left[u_2(X)\right].$$

- Ar

$$E\left[\sum_{i=1}^{k}c_iu_i(X)\right]=\sum_{i=1}^{k}c_iE\left[u_i(X)\right].$$

Example 2.2-4

- E[X] is the value of b that minimizes E[(X-b)^2]

Example 2.2-4 Let $u(x) = (x - b)^2$, where b is not a function of X, and suppose $E[(X - b)^2]$ exists. To find that value of b for which $E[(X - b)^2]$ is a minimum, we write

$$\begin{aligned} g(b) = E[(X - b)^2] &= E[X^2 - 2bX + b^2] \\ &= E(X^2) - 2bE(X) + b^2 \end{aligned}$$

because $E(b^2) = b^2$. To find the minimum, we differentiate $g(b)$ with respect to b, set $g'(b) = 0$, and solve for b as follows:

$$\begin{aligned} g'(b) = -2E(X) + 2b &= 0, \\ b &= E(X). \end{aligned}$$

Since $g''(b) = 2 > 0$, the mean of X, $\mu = E(X)$, is the value of b that minimizes $E[(X - b)^2]$. ■

Variance of Random Variable

- The variance of a random variable is defined to be E[$(X$-$\mu)^2$] and is typically denoted by σ^2 or Var(X) .
- For a discrete random variable X,

$$
\begin{aligned}
Var[X] = E\left[(X-\mu)^2\right] &= E\left[X^2 - 2\mu X + \mu^2\right] \\
&= E\left[X^2\right] - 2\mu E[X] + \mu^2 \\
&= E\left[X^2\right] - \mu^2.
\end{aligned}
$$

- σ is normally called the standard deviation.

Variance of aX+b

- Let X be a random variable with mean μ_X and variance σ_X^2.
- Let Y= aX+b, where a and b are constants.

Then,

$$E[Y] = E[aX + b] = aE[X] + b = a\mu_X + b$$

$$Var[Y] = E\left[(Y - \mu_y)^2\right]$$

$$= E\left[(aX + b - a\mu_X - b)^2\right]$$

$$= E\left[a^2(X - \mu_X)^2\right] = a^2 E\left[(X - \mu_X)^2\right] = a^2 \sigma_X{}^2.$$

Variance of Random Variable

- The variance of a random variable measures the deviation of its distribution from the mean.
- For example, in one drawing, Robert has 0.1% of chance to win \$100,000, while in another drawing, he has 0.01% of chance to win \$1,000,000.

- The expected amounts of award in these two drawings are equal.

  $0.001 * 100000 = 100$

  $0.0001 * 1000000 = 100$

- However, their variances are different.

  $0.001 * (100000 - 100)^2$
  $\quad + 0.999 * (0 - 100)^2 = 9{,}990{,}000$

  $0.0001 * (1000000 - 100)^2$
  $\quad + 0.9999 * (0 - 100)^2 = 99{,}990{,}000$

- In many distributions, the *mean* and *variance* together uniquely determine the parameters of the random variables.
  - The parameters that determine the probability distribution of the random variables

Moment of a Distribution

- Let X be a random variable and k be a positive integer.
- If $$E[X^k] = \sum_{x_i \in S} x_i^{\,k} P_X(x_i)$$
  is finite,
  then it is called the k^{th} moment of the distribution about origin.
- In addition, $E[(X-b)^k]$ is called the k^{th} moment of the distribution about b.

Chapter 2 Discrete Distributions

- Random Variables of the Discrete Type
  - Uniform Distribution
  - Hypergeometric Distribution
- Mathematical Expectation
- Moment Generating Function
- Bernoulli Trials and the Binomial Distribution
- Geometric and Negative Binomial Distribution
- The Poisson Distribution

The Moment-Generating Function

Definition 2.3-1

Let X be a random variable of the discrete type with pmf $f(x)$ and space S. If there is a positive number h such that

$$E(e^{tX}) = \sum_{x \in S} e^{tx} f(x)$$

exists and is finite for $-h < t < h$, then the function defined by

$$M(t) = E(e^{tX})$$

is called the **moment-generating function of** X (or of the distribution of X). This function is often abbreviated as mgf.

- Generating Property
- Uniqueness Property

The Moment-Generating Function

- Let X and Y be two discrete random variables defined on the same space S.

If $E\left[e^{tX}\right] = E\left[e^{tY}\right]$,

then the probability mass functions of X and Y are equal.

- Insight of the argument above：

Assume that S= ｛$s_1, s_2, \ldots$｝ contains only positive integers.

Then, we have

$$P_X(s_1)e^{ts_1} + P_X(s_2)e^{ts_2} + \ldots$$
$$= P_Y(s_1)e^{ts_1} + P_Y(s_2)e^{ts_2} + \ldots \quad .$$

(by mathematical transform theory)

Therefore, $P_X(s_1) = P_Y(s_1)$, i.e. X and Y have the same p.m.f.

Example 2.3-5 If X has the m.g.f.

$$M(t) = e^t\left(\frac{3}{6}\right) + e^{2t}\left(\frac{2}{6}\right) + e^{3t}\left(\frac{1}{6}\right),$$

then the probabilities are

$$P(X = 1) = \frac{3}{6}, \qquad P(X = 2) = \frac{2}{6}, \qquad P(X = 3) = \frac{1}{6}.$$

We can write this, if we choose to do so, by saying X has the p.m.f.

$$f(x) = \frac{4 - x}{6}, \qquad x = 1, 2, 3.$$

▲

Example 2.3-6 Suppose we are given that the m.g.f. of X is

$$M(t) = \frac{e^t/2}{1 - e^t/2}, \qquad t < \ln 2.$$

Until we expand $M(t)$, we can not detect the coefficients of $e^{b_i t}$. Recalling

$$(1 - z)^{-1} = 1 + z + z^2 + z^3 + \cdots, \qquad -1 < z < 1,$$

we have that

$$\frac{e^t}{2}\left(1 - \frac{e^t}{2}\right)^{-1} = \frac{e^t}{2}\left(1 + \frac{e^t}{2} + \frac{e^{2t}}{2^2} + \frac{e^{3t}}{2^3} + \cdots\right)$$

$$= (e^t)\left(\frac{1}{2}\right) + (e^{2t})\left(\frac{1}{2}\right)^2 + (e^{3t})\left(\frac{1}{2}\right)^3 + \cdots,$$

when $e^t/2 < 1$ and thus $t < \ln 2$. That is,

$$P(X = x) = \left(\frac{1}{2}\right)^x,$$

when x is a positive integer, or, equivalently, the p.m.f. of X is,

$$f(x) = \left(\frac{1}{2}\right)^x, \qquad x = 1, 2, 3, \cdots.$$

▲

$$M'(t) = \sum_{x \in S} x e^{tx} f(x),$$

$$M''(t) = \sum_{x \in S} x^2 e^{tx} f(x), \quad \dashrightarrow \quad M^{(r)}(t) = \sum_{x \in S} x^r e^{tx} f(x).$$

Setting $t = 0$, we see that

$$M'(0) = \sum_{x \in S} x f(x) = E(X),$$

$$M''(0) = \sum_{x \in S} x^2 f(x) = E(X^2),$$

and, in general,

$$M^{(r)}(0) = \sum_{x \in S} x^r f(x) = E(X^r).$$

Moment-Generating Function

- Let $M_X(t)$ be the m.g.f of a discrete random variable X.

$$\frac{d^k M_X(t)}{dt^K} = \sum_{x_i \in S} x_i^{\,k} e^{tx_i} P_X(x_i).$$

Furthermore,

$$\frac{d^k M_X(0)}{dt^K} = \sum_{x_i \in S} x_i^{\,k} P_X(x_i). = E\left[X^k\right].$$

- In particular,

$$\boxed{\mu_X = M_X{}'(0)\, and\, \sigma^2 = M_X{}''(0) - \left[M_X{}'(0)\right]^2.}$$