Chapter 2

Discrete Distributions

2.1 Random Variables of the Discrete Type

An outcome space S is difficult to study if the elements of S are not numbers. However, we can associate each element/outcome of S with a number.

Def. *Given a random experiment with an outcome space S, a function X that assigns one and only one real number $X(s) = x$ to each element s in S is called a* **random variable**. *The* **space** *of X is the set of real numbers $\{x \mid X(s) = x,\ s \in S\}$.*

When the elements of S are real numbers, the identical function $X(s) = s$ is itself a random variable with the space S.

Ex (Ex 2, p52, scanned file)

With random variables, the probabilities of outcomes/events in S may be studied by the probabilities of random variables (e.g. $P(X = x)$, $P(a < X \leq b)$, etc.)

If the outcome space S has finitely many elements or countable many elements, i.e. the elements of S can be listed in ordered:

$$S = \{s_1, s_2, \cdots\},$$

then S is called a **discrete outcome space**, and a random variable X is called a random variable of the **discrete type**.

For a random variable X of discrete type, the probability $P(X = x)$ is frequently denoted by $f(x)$, and is called the **probability mass function** (p.m.f.).

Def. *A* **probability mass function** *(p.m.f.) $f(x)$ of a discrete random variable X is a function that satisfies the following properties: (Let S be the space of r.v. X.)*

1. $f(x) > 0, \quad x \in S;$
2. $\sum_{x\in S} f(x) = 1$
3. $P(X \in A) = \sum_{x\in A} f(x)$, *where* $A \subset S$.

We denote $f(x) = 0$ when $x \notin S$.

Ex Ex 3, p.54 (scanned file)

The p.m.f. may be described by **bar graph** or **probability histogram** (Fig 2.1-1, p.54, scanned file).

Hypergeometric distribution: Consider a collection of $N = N_1 + N_2$ objects; N_1 of them belong to class 1; N_2 of them belong to class 2. A collection of n objects is selected from these N objects without replacement. The probability that exactly x objects of the n objects are in class 1, and $n - x$ objects are in class 2, is

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

Ex Ex 4-Ex 5, p55-p56 (scanned file)

Homework

§2.1 1, 7, 9, 11, 13

Attachment: Scanned textbook pages of Section 2-1

2.2 Mathematical Expectation

Ex 1 (p.61, scanned file)

Def. *Let $f(x)$ be the p.m.f. of the r.v. X of the discrete type with space S (i.e. $f : S \to \mathbf{R}$). Let $u(x)$ be a function with space S (i.e. $u : S \to \mathbf{R}$). The sum*

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

is called the **mathematical expectation** *or the* **expected value** *of the function $u(X)$.*

The definition of $E[u(X)]$ requires that the sum converges absolutely, that is,

$$\sum_{x \in S} |u(x)| f(x) \quad \text{converges to a finite number.}$$

Ex Let $f(x)$ be the p.m.f. of the r.v. X. Then

$$E(X) = \sum_{x \in S} x f(x), \qquad E(X^n) = \sum_{x \in S} x^n f(x).$$

Ex 2 (p.63, scanned file)

Thm 2.1. *The mathematical expectation E satisfies the following properties:*

1. *$E(c) = c$ for constant c;*
2. *$E[cu(X)] = cE[u(X)]$ for constant c and function $u(X)$.*
3. *$E[c_1 u_1(X) + c_2 u_2(X)] = c_1 E[u_1(X)] + c_2 E[u_2(X)]$ for constants c_1 and c_2, and functions $u_1(X)$ and $u_2(X)$.*

Ex (Exercise 2.2-2, p.66) Let the r.v. X have the p.m.f.

$$f(x) = \frac{(|x| + 1)^2}{9}, \qquad x = -1, 0, 1.$$

Compute $E(X)$, $E(X^2)$, and $E(3X^2 - 2X + 4)$.

Homework
§2.2 1, 3, 7

Attachment: Scanned textbook pages of Section 2-2

2.3 The Mean, Variance, and Standard Deviation

Let X be a random variable with space $S = \{u_1, u_2, \cdots, u_k\}$. Let $f : S \to [0, 1]$ be the p.m.f. of X. The **mean** of X (or of its distribution) is

$$\mu := E[X] = \sum_{x \in S} x f(x) = u_1 f(u_1) + u_2 f(u_2) + \cdots + u_k f(u_k).$$

The **variance** of X (or of its distribution) is

$$\begin{aligned} \mathrm{Var}(X) &:= E[(X-\mu)^2] = \sum_{x \in S} (x-\mu)^2 f(x) \\ &= (u_1-\mu)^2 f(u_1) + (u_2-\mu)^2 f(u_2) + \cdots + (u_k-\mu)^2 f(u_k). \end{aligned}$$

The **standard deviation** of X is

$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{E[(X-\mu)^2]}.$$

One can show that

$$\begin{aligned} \sigma^2 = \mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \\ &= \sum_{x \in S} x^2 f(x) - \mu^2. \end{aligned}$$

Ex 1, p69 (scanned file.)

The variance and standard deviation measure how dispersed the data are from the mean.

Ex 3, p70 (scanned file.)

Ex 4, p70, the mean and the variance of a uniform distribution.

The r**th moment** of the distribution X about the origin is

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

The r**th moment** of the distribution X about b is

$$E[(X-b)^r] = \sum_{x \in S} (x-b)^r f(x).$$

The r**th factorial moment** of the distribution X is

$$E[(X)_r] = E[X(X-1)(X-2)\cdots(X-r+1)].$$

If we perform a random experiment n times and obtain n observed values of the random variable — say, $x_1, x_2, \cdots, x_n$. This collection is called a **sample**. We actually create a probability distribution with weight $1/n$ on each of these x values, called the **empirical distribution**.

The **sample mean** of the empirical distribution is

$$\bar{x} := \frac{1}{n} \sum_{i=1}^{n} x_i.$$

It gives an estimation of $\mu = E[X]$.

The variance of the empirical distribution is

$$v = E[(X-\bar{x})^2] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

HOWEVER, $\text{Var}(X) = \sigma^2$ is better estimated by the **sample variance**

$$s^2 := \left(\frac{n}{n-1}\right) v = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Here s is the **sample standard deviation**.

Ex (Exercise 2.3-10, p76, see scanned file.)

Homework
§2.3 1, 3, 5, 9, 15

Attachment: Scanned textbook pages of Section 2-3

2.4 Bernoulli Trials and the Binomial Distribution

A **Bernoulli experiment** is an experiment with only two outcomes — say, success or failure (e.g. female or male, nondefective or defective).

Let p to denote the probability of success, and $q = 1 - p$ the probability of failure. Let the r.v. X denote as follow:

$$X(\text{success}) = 1, \qquad X(\text{failure}) = 0.$$

The p.m.f. of X is

$$f(1) = p, \qquad f(0) = q = 1 - p.$$

We say that X has a **Bernoulli distribution**.

The expected value of X is

$$\mu = E(X) = 1(p) + 0(1 - p) = p.$$

The variance of X is

$$\begin{aligned} \sigma^2 = \mathrm{Var}(X) &= E[(X - \mu)^2] \\ &= (1 - p)^2(p) + (0 - p)^2(1 - p) \\ &= p(1 - p). \end{aligned}$$

The standard deviation of X is

$$\sigma = \sqrt{p(1 - p)}.$$

Ex 1, p78 (file).

A sequence of **Bernoulli trials** occurs when

1. A Bernoulli (success-failure) experiment is performed n times.
2. The trials are independent.
3. The probability of success on each trial is a constant p; the probability of failure is $q = 1 - p$.

Ex 4, p79 (file).

In a sequence of n Bernoulli trials, let the random variable X denote the number of successes in n trials. The space of X is $\{0, 1, \cdots, n\}$. If x successes occur, then we have x successes in n trials and the remaining $n-x$ trials are failures. The p.m.f. of X is

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

These probabilities are called binomial probabilities, and the r.v. X is said to have a **binomial distribution**. A binomial distribution will be denoted by $b(n, p)$. The numbers n and p are called the **parameters** of the binomial distribution; they correspond to the number n of independent trials and the probability p of success on each trial.

Ex 6, Fig 2.4-1, p80-81 (file)

Ex 7, p80 (file).

For a random variable X, the **cumulative distribution function** (or the **distribution function**) is defined as:

$$F(x) := P(X \leq x), \qquad -\infty < X < \infty.$$

Ex (Fig 2.4-2 on p82, file) The distribution function for the $b(10, 0.8)$ distribution.

Ex 9, p82 (file).

Let X be the r.v. with the binomial distribution $b(n, p)$. The sum of p.m.f. over the space of X is 1:

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1 \qquad \text{(true for all } n\text{)}.$$

It can be applied to compute the mean of X.

Prop 2.2 (Binomial distribution $X \sim b(n,p)$)**.**

p.m.f.:	$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
Mean $\mu = E(X)$*:*	$\mu = np$
Variance $\sigma^2 = E(X^2) - \mu^2$*:*	$\sigma^2 = np(1-p)$.
Standard deviation:	$\sigma = \sqrt{np(1-p)}$
Distribution function:	*Table II in the appendix*

Ex 10 and 11, p84-85 (file).

Homework
§2.4 1, 3, 5, 11, 15, 21

Attachment: Scanned textbook pages of Section 2-4

2.5 The Moment-Generating Function

The mean, variance, and standard deviation are important characteristics of a distribution. However, it may be difficult to compute $E(X)$ or $E(X^2)$ for some distributions (e.g. binomial distribution.)

We define a new function of t that will help use generate the moments of a distribution.

Def. *Let X be a r.v. of discrete type with p.m.f. $f(x)$ and space S. The function of t:*

$$M(t) := E[e^{tX}] = \sum_{x\in S} e^{tx} f(x), \qquad -h < t < h,$$

is called the **moment-generating function of** X *(abbreviated as m.g.f.), provided that $M(t)$ converges for $-h < t < h$.*

Prop 2.3. *Properties of m.g.f. $M(t)$ of X:*

1. *If the space $S = \{b_1, b_2, b_3, \cdots\}$, then*

$$M(t) = e^{tb_1} f(b_1) + e^{tb_2} f(b_2) + e^{tb_3} f(b_3) + \cdots$$

Thus, the coefficient of e^{tb_i} is the probability $f(b_i) = P(X = b_i)$.

2. *Two random variables have the same m.g.f. if and only if they have the same p.m.f.*

3. *The derivatives of $M(t)$ at 0 generate all $E[X^r]$:*

$$M(t) = E[e^{tX}] = \sum_{x\in S} e^{tx} f(x), \qquad M(0) = E[1] = 1;$$

$$M'(t) = E[Xe^{tX}] = \sum_{x\in S} xe^{tx} f(x), \qquad M'(0) = E[X];$$

$$M''(t) = E[X^2e^{tX}] = \sum_{x\in S} x^2e^{tx} f(x), \qquad M''(0) = E[X^2];$$

$$\vdots$$

$$M^{(r)}(t) = E[X^re^{tX}] = \sum_{x\in S} x^re^{tx} f(x), \qquad M^{(r)}(0) = E[X^r].$$

Ex 1, p90: Suppose X has the m.g.f.

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

Describe the space and the p.m.f. of X.

Ex 2, p90: Describe X if the r.v. X has the m.g.f.

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

Ex 3, p91: Use the m.g.f. of X to find the mean, the variance, and the standard deviation of the binomial distribution $X \sim b(n,p)$ with p.m.f.:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \cdots, n,$$

(Answer: $M(t) = [(1-p) + pe^t]^n$. See scanned file.)

. . . .

Let r be a fixed positive integer. Let the r.v. X denote the number of Bernoulli trials needed to observe the r-th successes. Then $P(X = x)$ is the probability to see the r-th success in the x-th trial, that is, $(r-1)$ successes in the first $(x-1)$ trials and success in the last trial. The p.m.f. of X is

$$\begin{aligned} g(x) = P(X = x) &= \binom{x-1}{r-1} p^{r-1} (1-p)^{(x-1)-(r-1)} \cdot p \\ &= \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad x = r, r+1, \cdots . \end{aligned}$$

We say that X has a **negative binomial distribution** (parameters r, p).

Binomial distribution:	Fix the number of Bernoulli trials. Observe the number of successes.
Negative binomial distribution:	Fix the number of successes. Observe the number of Bernoulli trials needed.

Prop 2.4 (Negative binomial distribution X)**.**

p.m.f.: $g(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \qquad x = r, r+1, \cdots$

m.g.f.: $M(t) = \frac{(pe^t)^r}{[1-(1-p)e^t]^r}$ *where* $(1-p)e^t < 1$.

Mean: $\mu = E(X) = \frac{r}{p}$

Variance: $\sigma^2 = \frac{r(1-p)}{p^2}$.

The negative binomial distribution for $r = 1$ is called a **geometric distribution**, since the p.m.f. of X consists of terms of a geometric series, namely

$$g(x) = p(1-p)^{x-1}, \qquad x = 1, 2, 3, \cdots .$$

. . . .

Ex 4, p93 (scanned file)

Ex (HW 16, p98) (scanned file)

Ex 5, p94 (scanned file)

Ex 6, p95 (scanned file)

Homework
§2.5 1, 7, 9, 15

Attachment: Scanned textbook pages of Section 2-5

2.6 The Poisson Distribution

Def. *In counting the number of changes that occur in a time interval, we have an* **approximate Poisson process** *with parameter* $\lambda > 0$ *if:*

1. *The number of changes occurring in nonoverlapping intervals are independent.*

2. *The probability of exactly one change occurring in a sufficiently short interval of length* h *is approximately* λh.

3. *The probability of two or more changes occurring in a sufficiently short interval is essentially zero.*

Let the r.v. X denote the number of changes that occur in an interval of length 1. To get $P(X = x)$, we divide the unit interval into n subintervals with equal length $1/n$. We have a sequence of n Bernoulli trials with probability $p \approx \lambda/n$. Therefore,

$$P(X = x) \approx \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}.$$

Take limit for $n \to \infty$:

$$P(X = x) = \lim_{n\to\infty} \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x e^{-\lambda}}{x!}.$$

Prop 2.5 (**Poisson distribution** X with parameter $\lambda > 0$)**.**

p.m.f.:	$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, \cdots$
m.g.f.:	$M(t) = E(e^{tX}) = e^{\lambda(e^t - 1)}$
Mean:	$\mu = E(X) = \lambda$
Variance:	$\sigma^2 = E(X^2) - \mu^2 = \lambda$
Distribution function:	*Table III in the appendix*

Ex 1, 2, 3, p101-103, Table III in the appendix (scanned file)

If events in a Poisson process occur at a mean rate of λ per unit, then the number of occurrences—say, X—in the interval of length t has the Poisson p.m.f.

$$f(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, \qquad x = 0, 1, 2, \cdots$$

It just likes that we treat the interval of length t as if it were the "unit interval" with mean λt instead of λ.

Ex 4, 5, p103-104, Table III in the appendix (scanned file)

When n is large and p is small, the binomial distribution $b(n, p)$ may be approximated by the Poisson distribution with $\lambda = np$.

Ex Fig 2.6-3, p105 (scanned file)

Homework
§2.6 1, 7, 9, 15

Attachment: Scanned textbook pages of Section 2-6