CHAPTER 3

Random Variable

A rule that associate a number to each outcome of an experiment (or each outcome in S) is random variable.

Bernoulli random variable: Any random variable whose only possible values are 0 and 1

Example: Give three examples of Bernoulli random variables.

There is two different types of random variable:

- **Discrete random variable:** Possible values are integer.
- **Continuous random variable:** Possible values consist of an entire interval on the number line.

Example: Three automobiles are selected at random, and each is categorized as having a diesel (S) or nondiesel (F) engine. If X=the number of cars among the three with diesel engine, list each outcome in S and its associated X value.

Probability Distribution for Discrete Random Variables

The probability distribution of X determine how the total probability is distributed among the values of X. For showing probability distribution can use a formula, graph, or table. The probability distribution or probability mass function for discrete random variable $p(x) = P(X = x)$ has two conditions:

1. $p(x) \geqslant 0$
2. $\sum_{\text{all possible} x} p(x) = 1$

Examples:

• Airline sometimes overbook flights. Suppose that for a plane with 50 seats, 55 passengers have tickets. Define the random variable Y as the number of ticketed passengers who actually show up for the flight. The probability mass function of Y appears in the accompanying table.

y	45	46	47	48	49	50	51	52	53	54	55
$p(y)$	.05	.10	.12	.14	.25	.17	.06	.05	.03	.02	.01

a. What is the probability that flight will accommodate all ticketed passengers who show up?
b. What is the probability that not all ticketed passengers who show up can be accommodate?
• An automobile service facility specializing in engine tune-ups knows that 45% of all tune-ups are done on four cylinder automobiles, 40% on six cylinder automobiles, and 15% on eight-cylinder automobiles. Let $X=$ the number of cylinders on the next car to be tuned. What is the pmf of x?

A Parameter of a Probability Distribution

Suppose $p(x)$ depends on a quantity that can be assigned any of a number of possible values, with each different value determining a different probability distribution. Such quantity is called **parameter** of the distribution. The collection of all probability distributions for different values of the parameter is called a **family** of probability distributions.

$$p(x;\alpha) = \begin{cases} 1-\alpha & \text{if } x=0 \\ \alpha & \text{if } x=1 \\ 0 & \text{otherwise.} \end{cases}$$

Example: Starting at fixed time, we observe that the gender of each newborn child until a boy (B) is born. Let $p = P(B)$, and define the random variable X by X=number of birth observed, then

$$p(x) = \begin{cases} (1-p)^{x-1}p & x = 1,2,3,\cdots \\ 0 & \text{otherwise.} \end{cases}$$

The Cumulative Distribution Function

The **cumulative distribution function** (cdf) $F(x)$ of a discrete random variable X with pmf $p(x)$ is defined for every number x by

$$F(x) = P(X \leqslant x) = \sum_{y:y\leqslant x} p(y)$$

For any number x, $F(x)$ is the probability that the observed value of X will be at most x. Cumulative distribution function for random variable in above example is:

$$F(x) = \begin{cases} 1-(1-p)^{[x]} & x \geqslant 1 \\ 0 & x < 1. \end{cases}$$

Example: The pmf of Y is

y	1	2	3	4
$p(y)$	.4	.3	.2	.1

Obtain the cdf of Y and show it by graph.

Based on definition for cdf, for any two number a and b with $a \leqslant b$.

$$P(a \leqslant X \leqslant b) = F(b) - F(a^-)$$

where a^- represent the largest possible X value that is strictly less than a. In particular, if the only possible values are integers and if a and b are integers, then

$$P(a \leqslant X \leqslant b) = F(b) - F(a-1)$$

Taking $a = b$ yields $P(X = a) = F(a) - F(a-1)$.

Example: An insurance company offers its policyholders a number of different premium payment options. For a randomly selected policyholder, let X = the number of months between successive payments. The cdf of X is as follow

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.3 & 1 \leqslant x < 3 \\ 0.4 & 3 \leqslant x < 4 \\ 0.45 & 4 \leqslant x < 6 \\ 0.60 & 6 \leqslant x < 12 \\ 1 & 12 \leqslant x \end{cases}$$

a. What is the pmf of X?
b. Using just the cdf, compute $P(3 \leqslant X \leqslant 6)$ and $P(4 \leqslant X)$.

Expected Values of Discrete Random Variable

$$E(X) = \mu_x = \sum_{x \in D} x p(x) \quad D \text{ is set possible values of } x.$$

Expected value for a function $h(x)$ is $E[h(x)] = \sum h(x)p(x)$
Expected value for a linear function is $E(aX + b) = aE(X) + b$, therefore for any constant

- $E(aX) = aE(X)$
- $E(X + b) = E(X) + b$

The Variance of Random Variable

$$V(X) = \sum (x-\mu)^2 p(x) = E[(X-\mu)^2].$$

Also

$$V(X) = E(X^2) - [E(X)]^2 = \left[\sum x^2 p(x)\right] - \mu^2.$$

The standard deviation of X is $\sigma_x = \sqrt{\sigma_x^2}$.

The variance of a function $h(x)$ is $V[h(x)] = \sigma_{h(x)} = \sum \Big(h(x) - E[h(x)]\Big)^2 p(x)$.

Variance for a linear function is $V(aX+b) = a^2\sigma_x^2$ and $\sigma_{aX+b} = |a|\sigma_x$.

Therefore

- $\sigma_{ax}^2 = a^2\sigma_x^2$
- $\sigma_{x+b}^2 = \sigma_x^2$.

Example: The random variable X has following pmf

x	0	1	2	3	4
$p(x)$	0.08	0.15	0.45	0.27	0.05

Compute

a. $E(X)$

b. $V(X)$

c. The standard deviation of X.

✠ Binomial Distribution

A **binomial experiment** is one that has these five characteristics:

1. The experiment consists of n identical trials.
2. Each trial results in one of two outcomes. The one outcome is called a success S, and the other a failure, F.
3. The probability of success on a single trial is equal to p and probability of failure is equal to $(1-p) = q$.
4. The trials are independent.
5. We are interested in X, the number of successes observed during the n trials, for $X = 0, \cdots, n$.

Example: Determine whether the following experiments are binomial

- Check 100 births to find the proportion of boys.
- A shipment contains 30 computer and 2 of them are defective, a purchaser wants to check 3 of them to reject or accept the shipment.
- In a population, there are 500,000 licensed drivers, of whom 400,000 are insured, a sample of 10 drivers is chosen without replacement.

When the sample came from a large population, the probability of success p stayed about the same from trial to trial.
Rule of thumb: If the sample size is small relative to the population size such that n is at most 5% of the population size in without replacement sampling, the experiment follows **binomial** if satisfies the other conditions.

⧫ The Binomial Probability Distribution

A binomial experiment consists of n identical trials with probability of success p on each trial. Because the pmf of a binomial rv X depends on the two parameters n and p, the pmf is denoted by $b(x; n, p)$. The probability of x successes in n trials is equal to

$$b(x; n, p) = \begin{cases} C_x^n p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} & x = 0, 1, 2, \cdots n \\ 0 & \text{otherwise.} \end{cases}$$

⧫ Mean and Standard Deviation for the Binomial Probability Distribution

The random variable x, the number of successes in n trials, has a probability distribution with

$$\begin{aligned} \text{Mean}: &\quad \mu = np \\ \text{Variance}: &\quad \sigma^2 = npq \\ \text{Standard deviation}: &\quad \sigma = \sqrt{npq} \end{aligned}$$

Example: A marksman hits a target 80% of the time. He fires five shots at the target. What is the probability that exactly 3 shots hit the target? What is the probability that more than 3 shots hit the target?

⧫ Cumulative Probability Tables

You can use the **cumulative probability tables** to find probabilities for selected binomial distributions.

- Find the table for the correct value of n.
- Find the column for the correct value of p.
- The row marked "x" gives the cumulative probability, $P(X \leqslant x) = P(X = 0) + \cdots + P(X = x)$.

Example: Let x be a binomial random variable with $n = 20$ and $p = 0.1$.
a. Calculate $P(x \leqslant 4)$.
b. Calculate the mean and standard deviation of the random variable x.
c. Calculate the interval $\mu \pm \sigma$, $\mu \pm 2\sigma$, and $\mu \pm 3\sigma$. Find the probability that an observation will fall into each of these intervals.

♦ Hypergeometric Distribution

A bowl contains M red balls and $N - M$ white balls, for a total of N balls in the bowl. Select n balls from the bowl and record x the number of red balls. If define a success to be a red ball, then x is a hypergeometric random variable.

♦ The Hypergeometric Probability Distribution

A population contains M successes and $N-M$ failures. The probability of exactly x successes in a random sample of size n is

$$P(X = x) = h(x; n, M, N) = \frac{C_x^M C_{n-x}^{N-M}}{C_n^N} \qquad \max(0, n - N + M) \leqslant x \leqslant \min(n, M)$$

The mean and variance of a hypergeometric random variable x are

$$\mu = n(\frac{M}{N})$$
$$\sigma^2 = n(\frac{M}{N})(\frac{N-M}{N})(\frac{N-n}{N-1})$$

Example: A candy dish contains five blue and three red candies. A child reaches up and selects three candies without looking.
a. What is probability that there are two blue and one red candies in the selection?
b. What is the probability that the candies are all red?
c. What is the probability that the candies are all blue?

✠ The Negative Binomial Distribution

The negative binomial is based on experiment satisfying the following conditions:

1. The experiment consists of a sequence of independent trials.
2. Each trial can result in either success (S) or a failure (F).
3. The probability of success is constant from trial to trial, so $P(S \text{ on trial } i) = p$ for $i = 1, 2, \cdots$.
4. The experiment continuous (trials are performed) until a total of r successes have been observed, where r is a specified positive integer.

The random variable of interest is X = the number of failures that precede the rth success. The pmf of the negative binomial rv X with parameters r = number of successes and $p = P(\text{success})$ is

$$nb(x; r, p) = C_{r-1}^{x+r-1} p^r (1-p)^x \quad x = 0, 1, 2, \cdots$$

If X is a negative binomial rv with pmf $nb(x; r, p)$, then

$$E(X) = \frac{r(1-p)}{p} \quad V(X) = \frac{r(1-p)}{p^2}$$

Examples:

• An instructor who taught two sections of engineers statistics last term, the first with 20 students and the second with 30, decided to assign a term project. After all projects had been turned in, the instructor randomly order them before grading. Consider the first 15 graded projects.

a. What is the probability that exactly 10 of these are from the second section?
b. What is the probability that at least 10 of these are from the second section?
c. What is the probability that at least 10 of these are from the same section?
d. What is the mean value and standard deviation of the number among these 15 that are from the second section?
e. What are the mean value and standard deviation of the number of projects not among these first 15 that are from the second section?

• A family decides to have children until it has three children of the same gender. Assuming $P(B) = P(G) = 0.5$, what is the pmf of X = the number of children in the family?

✠ Poisson Distribution

The **Poisson random variable** x is a model for data that present the number of occurrences of a specified event in a given unit of time or space.

Examples:

- The number of calls received by a switchboard during a given period of time.
- The number of machine breakdowns in a day.
- The number of traffic accidents at a given intersection during a given time period.

♦ The Poisson Probability Distribution

Let λ be the average number of times that an event occurs in a certain period of time or space. A random variable X is said to have a **Poisson distribution** with parameter λ ($\lambda > 0$) if the pmf of X is

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

The mean and standard deviation of the Poisson random variable X are

$$\text{Mean}: \quad E(X) = \lambda$$
$$\text{Variance}: \quad V(X) = \lambda$$

Example: Suppose pulses arrive at the counter at an average rate of six per minute, what is the probability that in a 0.5-min interval at least one pulse is received?

♦ Cumulative Probability Tables

You can use the **cumulative probability tables** to find probabilities for selected Poisson distributions.

- Find the column for the correct value of λ.
- The row marked "k" gives the cumulative probability, $P(x \leqslant k) = P(x = 0) + \cdots + P(x = k)$

♦ The Poisson Approximation to the Binomial Distribution

The Poisson probability distribution provides a simple, easy-to-compute, and accurate approximation to binomial probabilities when n is large and $\lambda = np$ is small, preferably with $n > 50$ and $np < 5$, i.e.

$$b(x;n,p) \to p(x,\lambda) \ \text{ when } \ n \to \infty, p \to 0$$

Examples:

1. The number X of people entering the intensive care unit at the particular hospital on any one day has a Poisson probability distribution with mean equal to five persons per day.
   a. What is the probability that the number of people entering the intensive care unit one particular day is two? Less than or equal to two?
   b. Is it likely that X will exceed 10? Explain.

2. Sporadic outbreaks of E.coli have occurred at a rate of 2.5 per 100,000 for period of two years.
   a. What is the probability that at most five cases of E.coli per 100,000 are reported in a given year?
   b. What is the probability that more than five cases of E.coli per 100,000 are reported in a given year?

Suggested Exercises from Chapter 3: 7, 11, 13, 17, 23, 29, 39, 47, 49, 55, 57, 65, 69, 71, 73, 79, 81, 85, 95, 97, 101, 103, 109,