Chapter 3. Discrete Random Variables and Their Probability Distributions

2.11 Definition of random variable

3.1 Definition of a discrete random variable

3.2 Probability distribution of a discrete random variable

3.3 Expected value of a random variable or a function of a random variable

3.4-3.8 Well-known discrete probability distributions

- Discrete uniform probability distribution
- Bernoulli probability distribution
- Binomial probability distribution
- Geometric probability distribution
- Hypergeometric probability distribution
- Poisson probability distribution

3.9 Moments and Moment generating functions (see Chapter 6)

3.11 Tchebysheff's Theorem

2.11 Definition of random variable

(Example : opinion poll)
In an opinion poll, we decide to ask 100 people whether they agree or disagree with a certain issue.
Suppose we are interested in *the number of people who agree out of 100*.
If we record a "1" as for agree and "0" for disagree, then the sample space S for this experiment has sample points, each an ordered string of 1s and 0s of length 100. It is tedious to list all sample points.
Our interest is *the number of people who agree out of 100*. If we define a variable Y=number of 1s recorded out of 100, the (sample) space for Y is .

It frequently occurs that we are mainly interested in some functions of the outcomes as opposed to the outcome itself. In the example what we do is to define a new variable Y, the quantity of our interest. In statistics, Y is called a *random variable*.

(Def 2.12) A *random variable* (RV) Y is a real-valued function(mapping) from S into (not onto) R, a set of real number,

$$Y : S \to R \text{ where } y = Y(s) \text{ and } s \in S.$$

[note]

- Use late-alphabet capital letters (e.g., X, Y, Z) for RVs
- The *support* of Y is the set of possible values of Y, $\{y \in R : Y(s) = y, s \in S\}$.
- The different roles of (capital) Y and (lowercase) y(=a particular value that a RV Y may assume).

(Example) Toss of a coin
1) What is the S?
2) We are interested in Y = number of tail. What is the support of Y?
3) $Y : S \to R$

[Note] $Y : S \to R$ where $y = Y(s)$ and $s \in S$.

- Y is a variable that is a function of the sample points in S.
- Mapping or function: For each $s \in S$, there exists one and only one y such that $y = Y(s)$:
- One assigns a real number denoting the value of Y to each point in S :
  $\{Y = y\} = \{s : Y(s) = y, s \in S\}$ is the numerical event assigned the number y.
- Y partitions S into subsets so that points within a subset are all assigned the same value of Y. These subsets are mutually exclusive since no point is assigned two different numerical values.
- $P(Y = y)$ is the sum of the probabilities of the sample points that are assigned the value y.

(Example : opinion poll)
In an opinion poll, we ask four people whether they agree or disagree with a certain issue. Suppose we are interested in the number of people who agree out of four. We record a "1" as for agree and "0" for disagree.
i) Identify the sample points in S,
ii) Assign a value of Y to each sample point,
iii) identify the sample points associated with each value of the random variable Y.
iv) Compute probabilities for each value of Y.

3.1 Definition of a discrete r.v.

(Def 3.1) A random variable Y is said to be *discrete* if the support of Y is countable (either finite or pairable with the positive integers)

(Revisited opinion poll example)
The event of interest is $Y = \{$ the number of people who agree with a certain issue $\}$. Since the observed value of Y must be between zero and 100, sample size, Y takes on only a finite number of values and then is *discrete*.

(Example) common example: Any integer-valued Y is discrete.

3.2 Probability distribution of a discrete random variable

Every discrete random variable, Y, a *probability mass function* (or probability distribution) that gives the probability that Y is exactly equal to some value.

(Def 3.2 and 3.3) The probability that a discrete Y takes on the value y, $P(y) = P(Y = y)$, is a *probability mass function(p.m.f.)* (or probability distribution) of Y

- The expression $(Y = y)$: the set of all points in S assigned the value y by the random variable Y
- $P(Y = y)$: the *sum of the probabilities of all sample points in* S that are assigned the value y
- $P(y)$: represented by a formula, a table or a graph

(Example) A supervisor in a manufacturing plant has two men and three women working for him. He wants to choose two workers for a special job, and decides to select the two workers at random. Let Y denote the number of women in his selection. Find the probability distribution for Y and represent it by a table, or a graph and formula.

(Theorem 3.1)
For $p(y)$ for a discrete Y, the following must be true:

1. $0 \leq p(y) \leq 1$ for every y in the support of Y.
2. $\sum_y p(y) = 1$
3. $P(Y \in B) = \sum_{y \in B} p(y)$ where $B \subset R$.

(Example) $p(y) = c(y+1)^2$, $y = 0, 1, 2, 3$.
Determine c such that $p(y)$ is a discrete probability function.
Also find the probability distribution for Y, and represent it by a table and a graph.

(Def) *Cumulative Distribution Function*
For a discrete variable Y and real number a, the *cumulative distribution function* for Y is

$$F_Y(a) = P(Y \leq a) = \sum_{\text{all } y \leq a} p(y)$$

(Example) For discrete Y, $p(y)$ is defined over $y = -2, -1, 0, 1, 2, \ldots, 10$.

1) $F_Y(2) =$
2) $F_Y(6) =$
3) $P(2 \leq Y \leq 6) =$

3.3 The expected value of a r.v. or a function of a r.v.

The probability distribution for a r.v. Y: *theoretical model* for real distribution of data associated with a real population.

[Note] Given n observed samples $y_1, \ldots, y_n$, how one can describe the distribution of the data?

- measures of central tendency
  - Sample mean, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ for the unknown population mean : μ
- measures of dispersion or variation
  - Sample variance, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ and Sample standard deviation, $s = \sqrt{s^2}$ for the unknown population variance and standard deviation: σ^2 and σ

Our interest : characteristics of the probability distribution (p.m.f.) $p(y)$ for a discrete Y such as the mean and the variance for a discrete Y.

(Def 3.4) Let Y be a discrete r.v. with the probability mass function $p(y)$. Then, the expected value (mean) of Y, $E(Y)$, is defined to be

$$E(Y) = \sum_{y} yp(y).$$

How about the expected value of a function of a r.v. Y like Y^2?

(Theorem 3.2)
Let Y be a discrete r.v. with the probability mass function $p(y)$ and $g(Y)$ be a real-valued function of Y . Then the expected value of $g(Y)$ is given by

$$E(g(Y)) = \sum_{y} g(y)p(y).$$

(example) Roll one die; let X be the number obtained. Find $E(X)$ and $E(X^2)$.

Four useful expectation theorems

Assume that Y is a discrete r.v. with $p(y)$.

(Theorem 3.3)
Let Y be a discrete r.v. with $p(y)$ and c be a constant. Then $E(c) = c$.
(Proof)

(Theorem 3.4)
Let Y be a discrete r.v. with $p(y)$, $g(Y)$ be a function of Y, and let c be a constant. Then

$$E[cg(Y)] = cE[g(Y)]$$

(Theorem 3.5)
Let Y be a discrete r.v. with $p(y)$ and $g_1(Y), g_2(Y), \ldots, g_k(Y)$ be k functions of Y. Then,

$$E[g_1(Y) + g_2(Y) + \cdots + g_k(Y)]$$
$$= E[g_1(Y)] + E[g_2(Y)] + \cdots + E[g_k(Y)].$$

(Def 3.5) The variance of a discrete Y is defined to be the expected value of $(Y - \mu)^2$. That is,

$$V(Y) = E[(Y-\mu)^2] = \sum_y (Y-\mu)^2 p(y)$$

where $\mu = E(Y)$.
The standard deviation of Y is the positive square root of $V(Y)$, $\sqrt{V(Y)}$.

(Theorem 3.6)
Let Y be a discrete r.v. with $p(y)$. Then

$$\begin{aligned} V(Y) &= E[(Y-\mu)^2] = \sum_y (Y-\mu)^2 p(y) \\ &= \sum_y Y^2 p(y) - \mu^2 = E(Y^2) - \mu^2. \end{aligned}$$

where $\mu = E(Y)$.

(Example 3.2)

(Example) Let Y have $p(y) = \frac{y}{10}$, $y = 1, 2, 3, 4$.

Find $E(Y)$, $E(Y^2)$, $E(Y(5 - Y))$ and σ^2.

(Exercise 3.33) For constants a and b,
1) $E(aY + b) = aE(Y) + b = a\mu + b$
2) $Var(aY + b) = a^2 Var(Y) = a^2\sigma^2$
When a=1,   .
When a=-1 and b=0,   .

(Exercise) Let $\mu = E(Y)$ and $\sigma^2 = Var(Y)$.

Find $E\left(\frac{Y-\mu}{\sigma}\right)$ and $E\left(\frac{Y-\mu}{\sigma}\right)^2$.

In practice many experiments exhibit similar characteristics and generate random variables with the same types of probability distribution.

It is important to know the probability distributions, means and variances for random variables associated with common types of experiments.

Note that a probability distribution for a r.v. Y has the (unknown) constant(s) that determine its specific form, called *parameters*.

3.4-1 The discrete uniform random variable

(Def) A random variable Y is said to have a *discrete uniform distribution with the parameter* m if and only if $p(y) = \frac{1}{m}$ where $y = 1, 2, \ldots, m$.

(Theorem) Let Y be a discrete uniform random variable. Then,

$$\mu = E(Y) = \frac{m+1}{2} \text{ and } \sigma^2 = V(Y) = \frac{m^2-1}{12}.$$

(Question) Does p(y) in (Def) above satisfy the necessary properties in (Theorem 3.1)?

3.4-2 The Bernoulli random variable

The Bernoulli random variable is related to *Bernoulli experiment*.

1. The experiment results in one of two outcomes (concerned with the r.v. Y of interest). We call one outcome a success and the other a failure (success is merely a name for one of the two outcomes).

2. The probability of success is equal to p and then the probability of a failure is equal to $q(= 1 - p)$.

3. The random variable of interest is Y, the outcome itself (Let $Y = 1$ when the outcome is a success and $Y = 0$ when the outcome is a failure)

(Def) A random variable Y is said to have a *Bernoulli distribution* with the parameter p if and only if

$$p(y) = p^y q^{1-y},$$

where $q = 1 - p$, $y = 0, 1$ and $0 \leq p \leq 1$.

(Example) Toss a die one time. Let Y be a random variable indicating that one observes a number 6. The probability distribution of Y, $p(y)$, is

(Question) Does p(y) in (Def) satisfy the necessary properties in (Theorem 3.1)?

(Theorem) Let Y be a bernoulli random variable with success probability p. Then,

$$\mu = E(Y) = p \text{ and } \sigma^2 = V(Y) = pq$$

(Example) Suppose one tosses a die three times independently. Let Y be the number of times one observes a number 6. The probability distribution of Y, $p(y)$, is

(Answer) use the Binomial probability distribution.

3.4-3 The Binomial random variable

The Binomial random variable is related to *binomial experiments*

(Def 3.6)

1. The experiment consists of n identical and independent trials.

2. Each trial results in one of two outcomes (concerned with the r.v. Y of interest). We call one outcome a success S and the other a failure F. Here, success is merely a name for one of the two possible outcomes on a single trial of an experiment.

3. The probability of success on a single trial is equal to p and remains the same from trial to trial. The probability of a failure is equal to $q(= 1 - p)$.

4. The random variable of interest is Y, the number of successes observed during the n trials.

(Example 3.5) Reading

Before giving the definition of the *binomial p.m.f.*, try to derive it by using the sample-point approach.

How?
1. Each sample point in the sample space can be characterized by an n-tuple involving the letters S and F, corresponding to success and failure: Ex) $\underbrace{SSFSFSFFFFSS\ldots SF}_{\text{n positions(n trials)}}$.

The letter in the i-th position(proceeding from left to right) indicates the outcome of the i-th trial.

2. Consider a particular sample point corresponding to y successes and contained in the numerical event $Y = y$.

Ex) $\underbrace{SSSS\ldots SSS}_{\text{y}}\ \underbrace{FFF\ldots FF}_{\text{n-y}}$.

This sample point represents the intersection of n *independent* events in which there were y successes followed by $(n - y)$ failures.

3. If the probability of success(S) is p, this probability is unchanged from trial to trial because the trials are independent. So, the probability of the sample point in 2 is

Ex) $\underbrace{pppp\ldots ppp}_{\text{y terms}}\ \underbrace{qqq\ldots qq}_{\text{n-y terms}} = p^y q^{n-y}$.

Every other sample point in the event $Y = y$ can be represented as an n-tuple containing y S's and (n-y) F's in some order. Any such sample point also has probability $p^y q^{n-y}$..

4. The number of distinct n-tuples that contain y S's and $(n - y)$ F's is

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}.$$

5. The event $(Y = y)$ is made up of $\binom{n}{y}$ sample points, each with probability $p^y q^{n-y}$, and that the binomial probability distribution is

$$p(y) = \binom{n}{y} p^y q^{n-y}.$$

(Def 3.7) A random variable Y is said to have a *binomial distribution* with the parameters n trials and success probability p (in the binomial experiment) (i.e., $Y \sim b(n,p)$) if and only if

$$p(y) = \binom{n}{y} p^y q^{n-y},$$

where $q = 1-p$, $y = 0, 1, 2, \ldots, n$ and $0 \leq p \leq 1$.

How about $Y \sim b(1,p)$?

(Question) Does p(y) in (Def 3.7) satisfy the necessary properties in (Theorem 3.1)?

(Example 3.7) Suppose that a lot of 5000 electrical fuses contains 5% defectives. If a sample of five fuses is tested, find the probability of observing at least one defective.

(Exercise 3.39) A complex electronic system is built with a certain number of backup components in its subsystems. One subsystem has four identical components, each with a probability of .2 of failing in less than 1000 hours. The system will operate if any two of the four components are operating. Assume that the components operate independently.

(a) Find the probability that exactly two of the four components last longer than 1000 hours.

(b) Find the probability that the subsystem operates longer than 1000 hours.

(Theorem 3.7) Let Y be a binomial random variable based on n trials and success probability p. Then,

$$\mu = E(Y) = np \text{ and } \sigma^2 = V(Y) = npq$$

(Proof)

(Example 3.7) Mean and Variance

(Exercise 3.39) Mean and Variance

3.5 The Geometric random variable

The geometric random variable is related to the following experiments

1. The experiment consists of identical and independent trials, but **the number of trials is not fixed**.

2. Each trial results in one of two outcomes (concerning with the r.v, Y), a success and a failure.

3. The probability of success on a single trial is equal to p and remains the same from trial to trial. The probability of a failure is equal to $q(= 1 - p)$.

4. However, the random variable of interest Y is the number of the trial on which the first success occurs, not the number of successes that occur in n trials.

   So, the experiment could end with the first trial if a success is observed on the very first trial, or the experiment could go on indefinitely!!.

(Def 3.8) A random variable Y is said to have a *geometric probability distribution* with the parameter p, success probability (i.e., $Y \sim Geo(p)$) if and only if

$$p(y) = q^{y-1}p,$$

where $q = 1 - p$, $y = 1, 2, 3 \ldots$, and $0 \leq p \leq 1$.

(Question) Does p(y) in (Def 3.8) satisfy the necessary properties in (Theorem 3.1)?

(Exercise 3.67) Suppose that 30% of the applicants for a certain industrial job possess advanced training in computer programming. Applicants are interviewed sequentially and are selected at random from the pool. Find the probability that the first applicant with advanced training in programming is found on the fifth interview.

(Example) A basket player can make a free throw 60% of the time. Let X be the minimum number of free throws that this player must attempt to make first shot. What is $P(X = 5)$?

(Theorem 3.8)

Let Y be a ~~binomial~~ random variable with a geometric distribution,

$$\mu = E(Y) = \frac{1}{p} \text{ and } \sigma^2 = V(Y) = \frac{1-p}{p^2}$$

(Exercise 3.67 and Example above) Mean and Variance

[Memoryless property]

- CDF of $Y \sim Geo(p)$: $F_Y(a) = P(Y \leq a) =$
- $P(Y > a + b \mid Y > a) = P(Y > b)$: given that the first success has not yet occurred, the probability of the number of additional trials does not depend on how many failures has been observed.

3.7 The Hypergeometric random variable

The hypergeometric random variable is related to the following experiments

1. In the population of N elements there are elements of two distinct types (success and failure)

2. Among N elements r elements can be classified as success and $N-r$ elements can be classified as failture.

3. A sample of size n is randomly selected without replacement from a population of N elements

4. The random variable of interest is Y , the number of success in the sample

(Example) A bowl contains N chips, of which N_1 are white, and N_2 are green chips. Randomly select n chips from the bowl without replacement. Let Y be the number of white chips chosen. What is $P(Y=y)$?

(Def 3.10) A random variable Y is said to have a *hypergeometric probability distribution* with the parameters N and$r$ if and only if

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}},$$

where y is an integer $0, 1, 2, \ldots, n$, subject to the restrictions $y \le r$ and $n - y \le N - r$.

(Question) Does p(y) in (Def 3.10) satisfy the necessary properties in (Theorem 3.1)?
(Hint) Use the following facts :

$$\binom{a}{b} = 0 \quad \text{if} \quad b > a, \sum_{i=0}^{n} \binom{r}{i}\binom{N-r}{n-i} = \binom{N}{n}.$$

(Theorem 3.10)
Let Y be a random variable with a hypergeometric distribution,

$$\mu = E(Y) = \frac{nr}{N} \text{ and}$$
$$\sigma^2 = V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

(Exercise 3.103) A warehouse contains ten printing machines, four of which are defective. A company selects five of the machines at random, thinking all are in working condition. What is the probability that all five of the machines are nondefective?

[Relationship between Binomial distribution and Hypergeometric distribution]

When N is large, n is relatively small and r/N is held constant and equal to p, the following holds:

$$\lim_{N\to\infty} \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} = \binom{n}{y} p^y (1-p)^{n-y}$$

where $r/N = p$.

We learned the following discrete random variable and their probability distributions (p.m.f.):

1) Discrete uniform probability distribution

2) Bernoulli probability distribution

3) Binomial probability distribution

4) Geometric probability distribution

5) Hypergeometric probability distribution

The experiments in 2)-5) has two outcomes concerned with the r.v. Y, for example "success" and "failure".

Now we will learn how to model *counting* data (number of times a particular event occurs) : Poisson r.v. and its probability distribution.

3.8 The Poisson random variable

The Poisson r.v. often provides a good model for the probability distribution of the number Y of (rare) events that occur in a fixed space, time interval, volume, or any other dimensions.

(Example) the number of automobile accidents, or other types of accidents in a given unit of time.
(Example) the number of prairie dogs found in a square mile of prairie

(Def 3.11) A r.v. Y is said to have a *Poisson probability distribution* with the parameter λ (i.e., $Y \sim Poisson(\lambda)$) if and only if

$$p(y) = \frac{\lambda^y}{y!}e^{-\lambda},$$

where $y = 0, 1, 2, \ldots,$ and $\lambda > 0$ (λ does not have to be an integer, but $Y does$).

Here, λ (rate)= average of (rare) events that occur in a fixed space, time interval, volume, or any other dimensions (i.e., number of occurrences per that unit of dimension).

(Question) Does p(y) in (Def 3.11) satisfy the necessary properties in (Theorem 3.1)?

(Hint) Use the following fact :

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots = \sum_{y=0}^{\infty} \frac{\lambda^y}{y!}.$$

(Theorem 3.11)
Let Y be a random variable with a poisson distribution,

$$\mu = E(Y) = \lambda, \sigma^2 = V(Y) = \lambda.$$

(Example) If $Y \sim Poi(\lambda)$ and $\sigma^2 = 3$, $P(Y = 2)$?

(Example) Suppose $Y \sim Poi(\lambda)$ so that $3P(Y = 1) = P(Y = 2)$. Find $P(Y = 4)$?

(Example) The mean of a poisson r.v. Y is $\mu = 9$. Compute $P(\mu - 2\sigma < Y < \mu + 2\sigma)$.

(Example)The average number of homes sold by the X Realty company is 3 homes per day. What is the probability that exactly 4 homes will be sold tomorrow?

(Exercise 3.127) The number of typing errors made by a typist has a poisson distribution with an average of four errors per page. If more than four errors appear on a given page, the typist must retype the whole page. What is the probability that a certain page does not have to be typed?

(Example) Suppose that Y is the number of accidents in a 4 hours window, and the number of accidents per hour is 3. Then, what is $P(Y = 2)$?

Average

[Relationship between Binomial distribution and Poisson distribution]

Suppose Y is a Binomial r.v. with parameters n (total number of trials) and p (probability of success). For large n and small p such that $\lambda = np$, the following approximation can be used :

$$\lim_{n\to\infty} P(Y=y) = \lim_{n\to\infty} \binom{n}{y} p^y (1-p)^{n-y} = \frac{\lambda^y}{y!} e^{-\lambda}$$

This approximation is quite accurate if either $n \geq 20$ and $p \leq 0.05$ or $n \geq 100$ and $p \leq 0.10$.

(Example) $Y \sim b(100, 0.02)$

(a) $P(Y \leq 3) =$ 0.859

(b) Using Poisson distribution, approximate $P(Y \leq 3)$.

0.857

3.11 Tchebysheff's Theorem

How one can approximate the probability that the r.v. Y is within a certain interval?

- Use Empirical Rule if the probability distribution of Y is approximately bell-shaped.

- The interval with endpoints,

  · $(\mu - \sigma, \mu + \sigma)$ contains approximately ___ % of the measurements.

  · $(\mu - 2\sigma, \mu + 2\sigma)$ contains approximately ___ % of the measurements.

  · $(\mu - 3\sigma, \mu + 3\sigma)$ contains approximately ____ % of the measurements.

  E.G.) suppose that the scores on STAT515 midterm exam have *approximately* a bell-shaped curve with $\mu = 80$ and $\sigma = 5$. Then

  · approximately 68% of the scores are between ___ and ___ ,

  · approximately 95% of the scores are between ___ and ___,

  · almost all of the scores are between ___ and ___.

However, how one can approximate the probability that the r.v. Y is within a certain interval when the shape of its probability distribution is not bell-shaped?

(Theorem 4.13)[Tchebysheff's Theorem]

Let Y be a r.v. with $\mu = E(Y) < \infty$ and $\sigma^2 = Var(Y) < \infty$. Then, for any $k > 0$,

$$P(\mid Y - \mu \mid < k\sigma) \geq 1 - \frac{1}{k^2} \quad \text{or} \quad P(\mid Y - \mu \mid \geq k\sigma) \leq \frac{1}{k^2}$$

Even if the exact distributions are unknown for r.v. of interest, knowledge of the associated means and standard deviations permits us to determine a (meaningful) lower bounds for the probability that the r.v. Y falls in an intervals $\mu \pm k\sigma$.

(Example 3.28) The number of customers per day at a sales counter Y, has been observed for a long period of time and found to have mean 20 and standard deviation 2. The probability distribution of Y is not known. What is the probability that tomorrow Y will be greater than 16 but less than 24?