ECS315 2012/1 Part III.2 Dr.Prapun

8.3 Families of Discrete Random Variables

Many physical systems can be modeled by the same or similar random experiments and random variables. In this subsection, we present the analysis of several discrete random variables that frequently arise in applications.[27]

Definition 8.20. X is ***uniformly distributed*** on a finite set S if

$$p_X(x) = P[X = x] = \begin{cases} \frac{1}{|S|}, & x \in S, \\ 0, & \text{otherwise}, \end{cases}$$

- We write $X \sim \mathcal{U}(S)$ or $X \sim \text{Uniform}(S)$.
- Read "X is uniform on S" or "X is a uniform random variable on set S".
- The pmf is usually referred to as the uniform discrete distribution.
- Simulation: When the support S contains only consecutive integers[28], it can be generated by the command `randi` in `MATLAB` (R2008b).

[27] As mention in 7.11, we often omit a discussion of the underlying sample space of the random experiment and directly describe the distribution of a particular random variable.

[28] or, with minor manipulation, only uniformly spaced numbers

Example 8.21. X is uniformly distributed on $1, 2, \ldots, n$ if

In `MATLAB`, X can be generated by `randi(10)`.

Example 8.22. Uniform pmf is used when the random variable can take finite number of "equally likely" or "totally random" values.

- Classical game of chance / classical probability
- Fair gaming devices (well-balanced coins and dice, well-shuffled decks of cards)

Example 8.23. Roll a fair dice. Let X be the outcome.

Definition 8.24. X is a ***Bernoulli*** random variable if

$$p_X(x) = \begin{cases} 1-p, & x = 0, \\ p, & x = 1, \\ 0, & \text{otherwise}, \end{cases} \qquad p \in (0,1)$$

- Write $X \sim \mathcal{B}(1, p)$ or $X \sim \text{Bernoulli}(p)$
- X takes only two values: 0 or 1

Definition 8.25. X is a ***binary*** random variable if

$$p_X(x) = \begin{cases} 1-p, & x = a, \\ p, & x = b, \\ 0, & \text{otherwise}, \end{cases} \qquad p \in (0,1), \quad b > a.$$

- X takes only two values: a or b

Definition 8.26. X is a ***binomial*** random variable with size $n \in \mathbb{N}$ and parameter $p \in (0,1)$ if

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x \in \{0,1,2,\ldots,n\}, \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

- Write $X \sim \mathcal{B}(n,p)$ or $X \sim \text{binomial}(p)$.
    - Observe that $\mathcal{B}(1,p)$ is Bernoulli with parameter p.
- To calculate $p_X(x)$, can use `binopdf(x,n,p)` in `MATLAB`.
- Interpretation: X is the number of successes in n independent Bernoulli trials.

Example 8.27. An optical inspection system is to distinguish among different part types. The probability of a correct classification of any part is 0.98. Suppose that three parts are inspected and that the classifications are independent.

(a) Let the random variable X denote the number of parts that are correctly classified. Determine the probability mass function of X. [15, Q3-20]

(b) Let the random variable Y denote the number of parts that are incorrectly classified. Determine the probability mass function of Y.

Solution:

(a) X is a binomial random variable with $n = 3$ and $p = 0.98$. Hence,

$$p_X(x) = \begin{cases} \binom{3}{x} 0.98^x (0.02)^{3-x}, & x \in \{0,1,2,3\}, \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

In particular, $p_X(0) = 8 \times 10^{-6}$, $p_X(1) = 0.001176$, $p_X(2) = 0.057624$, and $p_X(3) = 0.941192$. Note that in `MATLAB`, these probabilities can be calculated by evaluating `binopdf(0:3,3,0.98)`.

(b) Y is a binomial random variable with $n = 3$ and $p = 0.02$. Hence,

$$p_Y(y) = \begin{cases} \binom{3}{y} 0.02^y (0.98)^{3-y}, & y \in \{0, 1, 2, 3\}, \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

In particular, $p_Y(0) = 0.941192$, $p_Y(1) = 0.057624$, $p_Y(2) = 0.001176$, and $p_Y(3) = 8 \times 10^{-6}$. Note that in MATLAB, these probabilities can be calculated by evaluating `binopdf(0:3,3,0.02)`.

Alternatively, note that there are three parts. If X of them are classified correctly, then the number of incorrectly classified parts is $n - X$, which is what we defined as Y. Therefore, $Y = 3 - X$. Hence, $p_Y(y) = P[Y = y] = P[3 - X = y] = P[X = 3 - y] = p_X(3 - y)$.

Example 8.28. Daily Airlines flies from Amsterdam to London every day. The price of a ticket for this extremely popular flight route is $75. The aircraft has a passenger capacity of 150. The airline management has made it a policy to sell 160 tickets for this flight in order to protect themselves against no-show passengers. Experience has shown that the probability of a passenger being a no-show is equal to 0.1. The booked passengers act independently of each other. Given this overbooking strategy, what is the probability that some passengers will have to be bumped from the flight?

Solution: This problem can be treated as 160 independent trials of a Bernoulli experiment with a success rate of $p = 9/10$, where a passenger who shows up for the flight is counted as a success. Use the random variable X to denote number of passengers that show up for a given flight. The random variable X is binomial distributed with the parameters $n = 160$ and $p = 9/10$. The probability in question is given by

$$P[X > 150] = 1 - P[X \leq 150] = 1 - F_X(150).$$

In MATLAB, we can enter `1-binocdf(150,160,9/10)` to get 0.0359. Thus, the probability that some passengers will be bumped from any given flight is roughly 3.6%. [22, Ex 4.1]

Definition 8.29. A geometric random variable X is defined by the fact that for some constant $\beta \in (0,1)$,

$$p_X(k+1) = \beta \times p_X(k)$$

for all $k \in S$ where S can be either $\mathbb{N}$ or $\mathbb{N} \cup \{0\}$.

(a) When its support is $\mathbb{N} = \{1, 2, \dots\}$,

$$p_X(x) = \begin{cases} (1-\beta)\,\beta^{x-1}, & x = 1, 2, \dots \\ 0, & \text{otherwise.} \end{cases}$$

- Write $X \sim \mathcal{G}_1(\beta)$ or $\text{geometric}_1(\beta)$.
- In `MATLAB`, to evaluate $p_X(x)$, use `geopdf(x-1,1-`β`)`.
- Interpretation: X is the number of trials required in a Bernoulli trials to achieve the first success.

  In particular, in a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until the first success. Then X is a geometric random variable with parameter $\beta = 1 - p$ and

$$\begin{aligned} p_X(x) &= \begin{cases} (1-\beta)\,\beta^{x-1}, & x = 1, 2, \dots \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} p(1-p)^{x-1}, & x = 1, 2, \dots \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

(b) When its support is $\mathbb{N} \cup \{0\}$,

$$\begin{aligned} p_X(x) &= \begin{cases} (1-\beta)\,\beta^{x}, & x = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} p(1-p)^{x}, & x = 0, 1, 2, \dots \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

- Write $X \sim \mathcal{G}_0(\beta)$ or $\text{geometric}_0(\beta)$.
- In `MATLAB`, to evaluate $p_X(x)$, use `geopdf(x,1-`β`)`.
- Interpretation: X is the number of failures in a Bernoulli trials before the first success occurs.

8.30. In 1837, the famous French mathematician Poisson introduced a probability distribution that would later come to be known as the Poisson distribution, and this would develop into one of the most important distributions in probability theory. As is often remarked, Poisson did not recognize the huge practical importance of the distribution that would later be named after him. In his book, he dedicates just one page to this distribution. It was Bortkiewicz in 1898, who first discerned and explained the importance of the Poisson distribution in his book Das Gesetz der Kleinen Zahlen (*The Law of Small Numbers*).

Definition 8.31. X is a ***Poisson*** random variable with ***parameter*** $\alpha > 0$ if

$$p_X(x) = \begin{cases} e^{-\alpha}\frac{\alpha^x}{x!}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

- In `MATLAB`, use `poisspdf(x,alpha)`.
- Write $X \sim \mathcal{P}(\alpha)$ or Poisson(α).
- We will see later in Example 9.7 that α is the "average" or expected value of X.
- Instead of X, Poisson random variable is usually denoted by Λ. The parameter α is often replaced by $\lambda\tau$ where λ is referred to as the ***intensity/rate parameter*** of the distribution

Example 8.32. The first use of the Poisson model is said to have been by a Prussian (German) physician, Bortkiewicz, who found that the annual number of late-19th-century Prussian (German) soldiers kicked to death by horses fitted a Poisson distribution [6, p 150],[3, Ex 2.23][29].

[29] I. J. Good and others have argued that the Poisson distribution should be called the Bortkiewicz distribution, but then it would be very difficult to say or write.

Example 8.33. The number of hits to a popular website during a 1-minute interval is given by $N \sim \mathcal{P}(\alpha)$ where $\alpha = 2$.

(a) Find the probability that there is at least one hit between 3:00AM and 3:01AM.

(b) Find the probability that there are at least 2 hits during the time interval above.

8.34. One of the reasons why Poisson distribution is important is because many natural phenomenons can be modeled by ***Poisson processes***.

Definition 8.35. A ***Poisson process*** (PP) is a random arrangement of "marks" (denoted by "$\times$" below) on the time line.

The "marks" may indicate the arrival times or occurrences of event/phenomenon of interest.

Example 8.36. Examples of processes that can be modeled by ***Poisson process*** include

(a) the sequence of times at which lightning strikes occur or mail carriers get bitten within some region

(b) the emission of particles from a radioactive source

(c) the arrival of

- telephone calls at a switchboard or at an automatic phone-switching system
- urgent calls to an emergency center
- (filed) claims at an insurance company
- incoming spikes (action potential) to a neuron in human brain

(d) the occurrence of

- serious earthquakes
- traffic accidents
- power outages

in a certain area.

(e) page view requests to a website

8.37. It is convenient to consider the Poisson process in terms of customers arriving at a facility.

We focus on a type of Poisson process that is called *homogeneous Poisson process*.

Definition 8.38. For ***homogeneous Poisson process***, there is only one parameter that describes the whole process. This number is call the ***rate*** and usually denoted by λ.

Example 8.39. If you think about modeling customer arrival as a Poisson process with rate $\lambda = 5$ customers/hour, then it means that during any fixed time interval of duration 1 hour (say, from noon to 1PM), you expect to have about 5 customers arriving in that interval. If you consider a time interval of duration two hours (say, from 1PM to 3PM), you expect to have about $2 \times 5 = 10$ customers arriving in that time interval.

8.40. One important fact which we will revisit later is that, for a homogeneous Poisson process, the number of arrivals during a time interval of duration T is a Poisson random variable with parameter $\alpha = \lambda T$.

Example 8.41. Examples of Poisson *random variables*:

- #photons emitted by a light source of intensity λ [photons/second] in time τ
- #atoms of radioactive material undergoing decay in time τ
- #clicks in a Geiger counter in τ seconds when the average number of click in 1 second is λ.
- #dopant atoms deposited to make a small device such as an FET
- #customers arriving in a queue or workstations requesting service from a file server in time τ
- Counts of demands for telephone connections in time τ
- Counts of defects in a semiconductor chip.

Example 8.42. Thongchai produces a new hit song every 7 months on average. Assume that songs are produced according to a Poisson process. Find the probability that Thongchai produces more than two hit songs in 1 year.

8.43. ***Poisson approximation*** of Binomial distribution: When p is small and n is large, $\mathcal{B}(n,p)$ can be approximated by $\mathcal{P}(np)$

(a) In a large number of independent repetitions of a Bernoulli trial having a small probability of success, the total number of successes is approximately Poisson distributed with parameter $\alpha = np$, where n = the number of trials and p = the probability of success. [22, p 109]

(b) More specifically, suppose $X_n \sim \mathcal{B}(n, p_n)$. If $p_n \to 0$ and $np_n \to \alpha$ as $n \to \infty$, then

$$P[X_n = k] = \binom{n}{k} p_n^k (1 - p_n)^{n-k} \to e^{-\alpha} \frac{\alpha^k}{k!}.$$

Example 8.44. Consider $X_n \sim \mathcal{B}(n, 1/n)$.

Example 8.45. Recall that Bortkiewicz applied the Poisson model to the number of Prussian cavalry deaths attributed to fatal horse kicks. Here, indeed, one encounters a very large number of trials (the Prussian cavalrymen), each with a very small probability of "success" (fatal horse kick).

8.46. Summary:

$X \sim$	Support S_X	$p_X(x)$
Uniform $\mathcal{U}_n$	$\{1, 2, \ldots, n\}$	$\frac{1}{n}$
$\mathcal{U}_{\{0,1,\ldots,n-1\}}$	$\{0, 1, \ldots, n-1\}$	$\frac{1}{n}$
Bernoulli $\mathcal{B}(1, p)$	$\{0, 1\}$	$\begin{cases} 1-p, & x = 0 \\ p, & x = 1 \end{cases}$
Binomial $\mathcal{B}(n, p)$	$\{0, 1, \ldots, n\}$	$\binom{n}{x} p^x (1-p)^{n-x}$
Geometric $\mathcal{G}_0(\beta)$	$\mathbb{N} \cup \{0\}$	$(1-\beta)\beta^x$
Geometric $\mathcal{G}_1(\beta)$	$\mathbb{N}$	$(1-\beta)\beta^{x-1}$
Poisson $\mathcal{P}(\alpha)$	$\mathbb{N} \cup \{0\}$	$e^{-\alpha}\frac{\alpha^x}{x!}$

Table 3: Examples of probability mass functions. Here, $p, \beta \in (0, 1)$. $\alpha > 0$. $n \in \mathbb{N}$

8.4 Some Remarks

8.47. Sometimes, it is useful to define and think of pmf as a vector $\underline{p}$ of probabilities.

When you use MATLAB, it is also useful to keep track of the values of x corresponding to the probabilities in p. This can be done via defining a vector $\underline{x}$.

Example 8.48. For $\mathcal{B}\left(3, \frac{1}{3}\right)$, we may define

$$\underline{x} = [0, 1, 2, 3]$$

and

$$\begin{aligned} \underline{p} &= \left[\binom{3}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^3, \binom{3}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2, \binom{3}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^1, \binom{3}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^0\right] \\ &= \left[\frac{8}{27}, \frac{4}{9}, \frac{2}{9}, \frac{1}{27}\right] \end{aligned}$$

8.49. At this point, we have a couple of ways to define probabilities that are associated with a random variable X

(a) We can define $P[X \in B]$ for all possible set B.

(b) For discrete random variable, we only need to define its pmf $p_X(x)$ which is defined as $P[X = x] = P[X \in \{x\}]$.

(c) We can also define the cdf $F_X(x)$.

Definition 8.50. If $p_X(c) = 1$, that is $P[X = c] = 1$, for some constant c, then X is called a ***degenerated*** random variable.

9 Expectation and Variance

Two numbers are often used to summarize a probability distribution for a random variable X. The mean is a measure of the center or middle of the probability distribution, and the variance is a measure of the dispersion, or variability in the distribution. These two measures do not uniquely identify a probability distribution. That is, two different distributions can have the same mean and variance. Still, these measures are simple, useful summaries of the probability distribution of X.

9.1 Expectation of Discrete Random Variable

The most important characteristic of a random variable is its expectation. Synonyms for expectation are expected value, mean, and first moment.

The definition of expectation is motivated by the conventional idea of numerical average. Recall that the numerical average of n numbers, say $a_1, a_2, \ldots, a_n$ is

$$\frac{1}{n}\sum_{k=1}^{n} a_k.$$

We use the average to summarize or characterize the entire collection of numbers $a_1, \ldots, a_n$ with a single value.

Example 9.1. Consider 10 numbers: 5, 2, 3, 2, 5, -2, 3, 2, 5, 2.

The average is

$$\frac{5+2+3+2+5+(-2)+3+2+5+2}{10} = \frac{27}{10} = 2.7.$$

We can rewrite the above calculation as

$$-2 \times \frac{1}{10} + 2 \times \frac{4}{10} + 3 \times \frac{2}{10} + 5 \times \frac{3}{10}$$

Definition 9.2. Suppose X is a discrete random variable, we define the **expectation** (or ***mean*** or ***expected value***) of X by

$$\mathbb{E}X = \sum_x x \times P[X = x] = \sum_x x \times p_X(x). \tag{15}$$

In other words, The expected value of a discrete random variable is a weighted mean of the values the random variable can take on where the weights come from the pmf of the random variable.

- Some references use m_X or μ_X to represent $\mathbb{E}X$.
- For conciseness, we simply write x under the summation symbol in (15); this means that the sum runs over all x values in the support of X. (Of course, for x outside of the support, $p_X(x)$ is 0 anyway.)

9.3. Analogy: In mechanics, think of point masses on a line with a mass of $p_X(x)$ kg. at a distance x meters from the origin.

In this model, $\mathbb{E}X$ is the center of mass (the balance point).

This is why $p_X(x)$ is called probability mass function.

Example 9.4. When $X \sim$ Bernoulli(p) with $p \in (0, 1)$,

Note that, since X takes only the values 0 and 1, its expected value p is "never seen".

9.5. Interpretation: The expected value is in general not a typical value that the random variable can take on. It is often helpful to interpret the expected value of a random variable as the ***long-run average value*** of the variable over many independent repetitions of an experiment

Example 9.6. $p_X(x) = \begin{cases} 1/4, & x = 0 \\ 3/4, & x = 2 \\ 0, & \text{otherwise} \end{cases}$

Example 9.7. For $X \sim \mathcal{P}(\alpha)$,

$$\begin{aligned}\mathbb{E}X &= \sum_{i=0}^{\infty} i e^{-\alpha} \frac{(\alpha)^i}{i!} = \sum_{i=1}^{\infty} e^{-\alpha} \frac{(\alpha)^i}{i!} i + 0 = e^{-\alpha} (\alpha) \sum_{i=1}^{\infty} \frac{(\alpha)^{i-1}}{(i-1)!} \\ &= e^{-\alpha} \alpha \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = e^{-\alpha} \alpha e^{\alpha} = \alpha.\end{aligned}$$

Example 9.8. For $X \sim \mathcal{B}(n,p)$,

$$\begin{aligned}\mathbb{E}X &= \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^{n} i \frac{n!}{i!\,(n-i)!} p^i (1-p)^{n-i} \\ &= n \sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!\,(n-i)!} p^i (1-p)^{n-i} = n \sum_{i=1}^{n} \binom{n-1}{i-1} p^i (1-p)^{n-i}\end{aligned}$$

Let $k = i - 1$. Then,

$$\mathbb{E}X = n \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} (1-p)^{n-(k+1)} = np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

We now have the expression in the form that we can apply the binomial theorem which finally gives

$$\mathbb{E}X = np(p + (1-p))^{n-1} = np.$$

We shall revisit this example again using another approach in Example 10.34.

Example 9.9. ***Pascal's wager***: Suppose you concede that you don't know whether or not God exists and therefore assign a 50 percent chance to either proposition. How should you weigh these odds when deciding whether to lead a pious life? If you act piously and God exists, Pascal argued, your gain–eternal happiness–is infinite. If, on the other hand, God does not exist, your loss, or negative return, is small–the sacrifices of piety. To weigh these possible gains and losses, Pascal proposed, you multiply the probability of each possible outcome by its payoff and add them all up, forming a kind of average or expected payoff. In other words, the mathematical expectation of your return on piety is one-half infinity (your gain if God exists) minus one-half a small number (your loss if he does not exist). Pascal knew enough about infinity to know that the answer to this calculation is infinite, and thus the expected return on piety is infinitely positive. Every reasonable person, Pascal concluded, should therefore follow the laws of God. [14, p 76]

- Pascals wager is often considered the founding of the mathematical discipline of game theory, the quantitative study of optimal decision strategies in games.

9.10. Technical issue: Definition (15) is only meaningful if the sum is well defined.

The sum of infinitely many nonnegative terms is always well-defined, with $+\infty$ as a possible value for the sum.

- ***Infinite Expectation***: Consider a random variable X whose pmf is defined by

$$p_X(x) = \begin{cases} \frac{1}{cx^2}, & x = 1, 2, 3, \dots \\ 0, & \text{otherwise} \end{cases}$$

  Then, $c = \sum_{n=1}^{\infty} \frac{1}{n^2}$ which is a finite positive number ($\pi^2/6$). However,

$$\mathbb{E}X = \sum_{k=1}^{\infty} k p_X(k) = \sum_{k=1}^{\infty} k \frac{1}{c} \frac{1}{k^2} = \frac{1}{c} \sum_{k=1}^{\infty} \frac{1}{k} = +\infty.$$

Some care is necessary when computing expectations of signed random variables that take infinitely many values.

- The sum over countably infinite many terms is not always well defined when both positive and negative terms are involved.

- For example, the infinite series $1 - 1 + 1 - 1 + \dots$ has the sum 0 when you sum the terms according to $(1-1)+(1-1)+\cdots$, whereas you get the sum 1 when you sum the terms according to $1 + (-1 + 1) + (-1 + 1) + (-1 + 1) + \cdots$.

- Such abnormalities cannot happen when all terms in the infinite summation are nonnegative.

It is the convention in probability theory that $\mathbb{E}X$ should be evaluated as

$$\mathbb{E}X = \sum_{x \geq 0} x p_X(x) - \sum_{x<0} (-x) p_X(x),$$

- If at least one of these sums is finite, then it is clear what value should be assigned as $\mathbb{E}X$.
- If both sums are $+\infty$, then no value is assigned to $\mathbb{E}X$, and we say that $\mathbb{E}X$ is **undefined**.

Example 9.11. Undefined Expectation: Let

$$p_X(x) = \begin{cases} \frac{1}{2cx^2}, & x = \pm 1, \pm 2, \pm 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

Then,

$$\mathbb{E}X = \sum_{k=1}^{\infty} k p_X(k) - \sum_{k=-\infty}^{-1} (-k) p_X(k).$$

The first sum gives

$$\sum_{k=1}^{\infty} k p_X(k) = \sum_{k=1}^{\infty} k \frac{1}{2ck^2} = \frac{1}{2c} \sum_{k=1}^{\infty} \frac{1}{k} = \frac{\infty}{2c}.$$

The second sum gives

$$\sum_{k=-\infty}^{-1} (-k) p_X(k) = \sum_{k=1}^{\infty} k p_X(-k) = \sum_{k=1}^{\infty} k \frac{1}{2ck^2} = \frac{1}{2c} \sum_{k=1}^{\infty} \frac{1}{k} = \frac{\infty}{2c}.$$

Because both sums are infinite, we conclude that $\mathbb{E}X$ is undefined.

9.12. More rigorously, to define $\mathbb{E}X$, we let $X^+ = \max\{X, 0\}$ and $X^- = -\min\{X, 0\}$. Then observe that $X = X^+ - X^-$ and that both X^+ and X^- are nonnegative r.v.'s. We say that a random variable X ***admits an expectation*** if $\mathbb{E}X^+$ and $\mathbb{E}X^-$ are not both equal to $+\infty$. In which case, $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$.

9.2 Function of a Discrete Random Variable

Given a random variable X, we will often have occasion to define a new random variable by $Y \equiv g(X)$, where $g(x)$ is a real-valued function of the real-valued variable x. More precisely, recall that a random variable X is actually a function taking points of the sample space, $\omega \in \Omega$, into real numbers $X(\omega)$. Hence, we have the following definition

Definition 9.13. The notation $Y = g(X)$ is actually shorthand for $Y(\omega) := g(X(\omega))$.

- The random variable $Y = g(X)$ is sometimes called **derived** random variable.

Example 9.14. Let

$$p_X(x) = \begin{cases} \frac{1}{c}x^2, & x = \pm 1, \pm 2 \\ 0, & \text{otherwise} \end{cases}$$

and

$$Y = X^4.$$

Find $p_Y(y)$ and then calculate $\mathbb{E}Y$.

9.15. For discrete random variable X, the pmf of a derived random variable $Y = g(X)$ is given by

$$p_Y(y) = \sum_{x: g(x) = y} p_X(x).$$

Note that the sum is over all x in the support of X which satisfy $g(x) = y$.

Example 9.16. A "binary" random variable X takes only two values a and b with

$$P[X = b] = 1 - P[X = a] = p.$$

X can be expressed as $X = (b - a)I + a$, where I is a Bernoulli random variable with parameter p.

9.3 Expectation of a Function of a Discrete Random Variable

Recall that for discrete random variable X, the pmf of a derived random variable $Y = g(X)$ is given by

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x).$$

If we want to compute $\mathbb{E}Y$, it might seem that we first have to find the pmf of Y. Typically, this requires a detailed analysis of g which can be complicated, and it is avoided by the following result.

9.17. Suppose X is a discrete random variable.

$$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x).$$

This is referred to as the **law/rule of the lazy/unconcious statistician** (LOTUS) [23, Thm 3.6 p 48],[9, p. 149],[8, p. 50] because it is so much easier to use the above formula than to first find the pmf of Y. It is also called **substitution rule** [22, p 271].

Example 9.18. Back to Example 9.14. Recall that

$$p_X(x) = \begin{cases} \frac{1}{c}x^2, & x = \pm 1, \pm 2 \\ 0, & \text{otherwise} \end{cases}$$

(a) When $Y = X^4$, $\mathbb{E}Y =$

(b) $\mathbb{E}[2X-1]$

9.19. Caution: A frequently made *mistake* of beginning students is to set $\mathbb{E}[g(X)]$ equal to $g(\mathbb{E}X)$. In general, $\mathbb{E}[g(X)] \neq g(\mathbb{E}X)$.

(a) In particular, $\mathbb{E}\left[\frac{1}{X}\right]$ is not the same as $\frac{1}{\mathbb{E}X}$.

(b) An exception is the case of a linear function $g(x) = ax + b$. See also (9.22).

Example 9.20. For $X \sim \text{Bernoulli}(p)$,

(a) $\mathbb{E}X = p$

(b) $\mathbb{E}\left[X^2\right] = 0^2 \times (1-p) + 1^2 \times p = p \neq (\mathbb{E}X)^2$.

Example 9.21. Continue from Example 9.7. Suppose $X \sim \mathcal{P}(\alpha)$.

$$\mathbb{E}\left[X^2\right] = \sum_{i=0}^{\infty} i^2 e^{-\alpha}\frac{\alpha^i}{i!} = e^{-\alpha}\alpha\sum_{i=0}^{\infty} i\frac{\alpha^{i-1}}{(i-1)!} \tag{16}$$

We can evaluate the infinite sum in (16) by rewriting i as $i-1+1$:

$$\begin{aligned}\sum_{i=1}^{\infty} i\frac{\alpha^{i-1}}{(i-1)!} &= \sum_{i=1}^{\infty}(i-1+1)\frac{\alpha^{i-1}}{(i-1)!} = \sum_{i=1}^{\infty}(i-1)\frac{\alpha^{i-1}}{(i-1)!} + \sum_{i=1}^{\infty}\frac{\alpha^{i-1}}{(i-1)!}\\ &= \alpha\sum_{i=2}^{\infty}\frac{\alpha^{i-2}}{(i-2)!} + \sum_{i=1}^{\infty}\frac{\alpha^{i-1}}{(i-1)!} = \alpha e^{\alpha} + e^{\alpha} = e^{\alpha}(\alpha+1).\end{aligned}$$

Plugging this back into (16), we get

$$\mathbb{E}\left[X^2\right] = \alpha(\alpha+1) = \alpha^2 + \alpha.$$

9.22. Some Basic Properties of Expectations

(a) For $c \in \mathbb{R}$, $\mathbb{E}[c] = c$

(b) For $c \in \mathbb{R}$, $\mathbb{E}[X + c] = \mathbb{E}X + c$ and $\mathbb{E}[cX] = c\mathbb{E}X$

(c) For constants a, b, we have $\mathbb{E}[aX + b] = a\mathbb{E}X + b$.

Definition 9.23. Some definitions involving expectation of a function of a random variable:

(a) ***Absolute moment***: $\mathbb{E}\left[|X|^k\right]$, where we define $\mathbb{E}\left[|X|^0\right] = 1$

(b) ***Moment***: $m_k = \mathbb{E}\left[X^k\right]$ = the k^{th} moment of X, $\quad k \in \mathbb{N}$.

- The first moment of X is its expectation $\mathbb{E}X$.
- The second moment of X is $\mathbb{E}\left[X^2\right]$.

9.4 Variance and Standard Deviation

An average (expectation) can be regarded as one number that summarizes an entire probability model. After finding an average, someone who wants to look further into the probability model might ask, "How typical is the average?" or, "What are the chances of observing an event far from the average?" A measure of **dispersion/deviation/spread** is an answer to these questions wrapped up in a single number. (The opposite of this measure is the **peakedness**.) If this measure is small, observations are likely to be near the average. A high measure of dispersion suggests that it is not unusual to observe events that are far from the average.

Example 9.24. Consider your score on the midterm exam. After you find out your score is 7 points above average, you are likely to ask, "How good is that? Is it near the top of the class or somewhere near the middle?".

Example 9.25. In the case that the random variable X is the random payoff in a game that can be repeated many times under identical conditions, the expected value of X is an informative measure on the grounds of the law of large numbers. However, the information provided by $\mathbb{E}X$ is usually not sufficient when X is the random payoff in a nonrepeatable game.

Suppose your investment has yielded a profit of \$3,000 and you must choose between the following two options:

- the first option is to take the sure profit of \$3,000 and
- the second option is to reinvest the profit of \$3,000 under the scenario that this profit increases to \$4,000 with probability 0.8 and is lost with probability 0.2.

The expected profit of the second option is

$$0.8 \times \$4,000 + 0.2 \times \$0 = \$3,200$$

and is larger than the \$3,000 from the first option. Nevertheless, most people would prefer the first option. The downside ***risk*** is too big for them. A measure that takes into account the aspect of risk is the *variance* of a random variable. [22, p 35]

9.26. The most important ***measures of dispersion*** are the standard deviation and its close relative, the variance.

Definition 9.27. ***Variance***:

$$\operatorname{Var} X = \mathbb{E}\left[(X - \mathbb{E}X)^2\right]. \tag{17}$$

- Read "the variance of X"
- *Notation*: D_X, or $\sigma^2(X)$, or σ_X^2, or $\mathbb{V}X$ [23, p. 51]
- In some references, to avoid confusion from the two expectation symbols, they first define $m = \mathbb{E}X$ and then define the variance of X by

$$\operatorname{Var} X = \mathbb{E}\left[(X - m)^2\right].$$

- We can also calculate the variance via another identity:

$$\operatorname{Var} X = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2$$

- The units of the variance are squares of the units of the random variable.

9.28. Basic properties of variance:

- $\operatorname{Var} X \geq 0$.
- $\operatorname{Var} X \leq \mathbb{E}\left[X^2\right]$.
- $\operatorname{Var}[cX] = c^2 \operatorname{Var} X$.
- $\operatorname{Var}[X + c] = \operatorname{Var} X$.
- $\operatorname{Var}[aX + b] = a^2 \operatorname{Var} X$.

Definition 9.29. ***Standard Deviation***: $\sigma_X = \sqrt{\operatorname{Var}[X]}$.

- It is useful to work with the standard deviation since it has the same units as $\mathbb{E}X$.
- Informally we think of outcomes within $\pm\sigma_X$ of $\mathbb{E}X$ as being in the center of the distribution. Some references would informally interpret sample values within $\pm\sigma_X$ of the expected value, $x \in [\mathbb{E}X - \sigma_X, \mathbb{E}X + \sigma_X]$, as "typical" values of X and other values as "unusual".
- $\sigma_{aX+b} = |a|\,\sigma_X$.

9.30. In finance, standard deviation is a key concept and is used to measure the ***volatility*** (risk) of investment returns and stock returns.

It is common wisdom in finance that diversification of a portfolio of stocks generally reduces the total risk exposure of the investment. We shall return to this point in Example 10.51.

Example 9.31. Continue from Example 9.24. If the standard deviation of exam scores is 12 points, the student with a score of +7 with respect to the mean can think of herself in the middle of the class. If the standard deviation is 3 points, she is likely to be near the top.

Example 9.32. Suppose $X \sim \text{Bernoulli}(p)$.

(a) $\mathbb{E}\left[X^2\right] = 0^2 \times (1-p) + 1^2 \times p = p$.

(b) $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = p - p^2 = p(1-p)$.
Alternatively, if we directly use (17), we have

$$\begin{aligned}\operatorname{Var} X = \mathbb{E}\left[(X-\mathbb{E}X)^2\right] &= (0-p)^2 \times (1-p) + (1-p)^2 \times p \\ &= p(1-p)(p+(1-p)) = p(1-p).\end{aligned}$$

Example 9.33. Continue from Example 9.7 and Example 9.21. Suppose $X \sim \mathcal{P}(\alpha)$. We have

$$\operatorname{Var} X = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2 = \alpha^2 + \alpha - \alpha^2 = \alpha.$$

Therefore, for Poisson random variable, the expected value is the same as the variance.

Example 9.34. Consider the two pmfs shown in Figure 5. The random variable X with pmf at the left has a smaller variance than the random variable Y with pmf at the right because more probability mass is concentrated near zero (their mean) in the graph at the left than in the graph at the right. [9, p. 85]

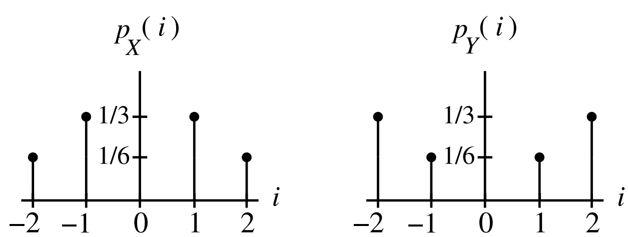

Figure 5: Example 9.34 shows that a random variable whose probability mass is concentrated near the mean has smaller variance. [9, Fig. 2.9]

9.35. We have already talked about variance and standard deviation as a number that indicates spread/dispersion of the pmf. More specifically, let's imagine a pmf that shapes like a bell curve. As the value of σ_X gets smaller, the spread of the pmf will be smaller and hence the pmf would "look sharper". Therefore, the probability that the random variable X would take a value that is far from the mean would be smaller.

The next property involves the use of σ_X to bound "the tail probability" of a random variable.

9.36. *Chebyshev's Inequality*:

$$P\left[|X - \mathbb{E}X| \geq \alpha\right] \leq \frac{\sigma_X^2}{\alpha^2}$$

or equivalently

$$P\left[|X - \mathbb{E}X| \geq n\sigma_X\right] \leq \frac{1}{n^2}$$

- Useful only when $\alpha > \sigma_X$

Example 9.37. If X has mean m and variance σ^2, it is sometimes convenient to introduce the normalized random variable

$$Y = \frac{X - m}{\sigma}.$$

Definition 9.38. *Central Moments*: A generalization of the variance is the nth central moment which is defined to be

$$\mu_n = \mathbb{E}\left[(X - \mathbb{E}X)^n\right].$$

(a) $\mu_1 = \mathbb{E}\left[X - \mathbb{E}X\right] = 0$.

(b) $\mu_2 = \sigma_X^2 = \operatorname{Var} X$: the second central moment is the variance.