Jinguo Lian
Math437 Notes
January 14, 2016

2 Discrete Random Variables

2.1 Basic Concepts

2.1.1 Random Variables

Example 1. *If you toss a coin five times. This is a random experiment and the sample space can be written as*

$$S = \{TTTTT, TTTTH, ..., HHHHH\}$$

Note that here the sample space S has $2^5 = 32$ elements. Suppose that in this experiment, we are interested in the number of heads. We can define a random variable X whose value is the number of observed heads. The value of X will be one of 0,1,2,3,4 or 5 depending on the outcome of the random experiment.

Definition 2. *Random Variables: A random variable X is a function from the sample space to the real numbers.*

$$X : S \to \mathbb{R}$$

The range of a random variable X, shown by Range(X) or R_X, is the set of possible values of X.

Example 3. *Find the range for each of the following random variables.*
I toss a coin 100 times. Let X be the number of heads I observe.
I toss a coin until the first heads appears. Let Y be the total number of coin tosses.
The random variable T is defined as the time (in hours) from now until the next earthquake occurs in a certain city.
Solution:
The random variable X can take any integer from 0 to 100, so $R_X = \{0, 1, 2, ..., 100\}$.
The random variable Y can take any positive integer, so $R_Y = \{1, 2, 3, ...\} = \mathbb{N}$.
The random variable T can in theory get any positive real number, so $R_T = [0, \infty)$.

2.1.2 Discrete Random Variables

There are two important classes of random variables that we discuss in this book: discrete random variables and continuous random variables. We will discuss discrete random variables in this section.

Definition 4. *X is a discrete random variable, if its range is countable.*

2.1.3 Probability Mass Function (PMF)

If X is a discrete random variable then its range R_X is a countable set, so, we can list the elements in R_X. In other words, we can write

$$R_X = \{x_1, x_2, x_3, ...\}$$

Note that here $x_1, x_2, x_3, ...$ are possible values of the random variable X. While random variables are usually denoted by capital letters, to represent the numbers in the range we usually use lowercase letters such as x, x_1, y, z, etc. For a discrete random variable X, we are interested in knowing the probabilities of $X = x_k$. Note that here, the event $A = \{X = x_k\}$ is defined as the set of outcomes s in the sample space S for which the corresponding value of X is equal to x_k. In particular, $A = \{s \in S | X(s) = x_k\}$. The probabilities of events $\{X = x_k\}$ are formally shown by the probability mass function (pmf) of X.

Definition 5. *Let X be a discrete random variable with range $R_X = \{x_1, x_2, x_3, ...\}$ (finite or countably infinite). The function*

$$P_X(x_k) = P(X = x_k), \text{ for } k = 1, 2, 3, ...,$$

assigns a probability to each value x_k of the random variable X, such that

1. *$P(x_k) \geq 0$ for all $x_k \in R_X$.*
2. *$\sum_{x_k \in R_X} P(x_k) = 1$.*

The probability function is called the probability mass function (PMF) of X.

Thus, the PMF is a probability measure that gives us probabilities of the possible values for a random variable. While the above notation is the standard notation for the PMF of X, it might look confusing at first. The subscript X here indicates that this is the PMF of the random variable X. Thus, for example, $P_X(1)$ shows the probability that X=1. To better understand all of the above concepts, let's look at some examples.

Example 6. *I toss a fair coin twice, and let X be defined as the number of heads I observe. Find the range of X, R_X, as well as its probability mass function P_X.*

Solution: *Here, our sample space is given by*

$$S = \{HH, HT, TH, TT\}$$

The number of heads will be 0, 1 or 2. Thus $R_X = \{0, 1, 2\}$. Since this is a finite (and thus a countable) set, the random variable X is a discrete random variable. Next, we need to find PMF of X. The PMF is defined as

$$P_X(k) = P(X = k) \text{ for } k = 0, 1, 2$$

We have

$$P_X(0) = P(X = 0) = P(TT) = \frac{1}{4}$$

,

$$P_X(1) = P(X = 1) = P(HT, TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

,

$$P_X(2) = P(X = 2) = P(HH) = \frac{1}{4}$$

.

Although the PMF is usually defined for values in the range, it is sometimes convenient to extend the PMF of X to all real numbers. If $x \notin R_X$, we can simply write $P_X(x) = P(X = x) = 0$. Thus, in general we can write

$$P_X(x) = \begin{cases} P(X = x) & \text{if } x \text{ is in } R_X \\ 0 & \text{otherwise} \end{cases}$$

To better visualize the PMF, we can plot it. The Figure shows the PMF of the above random variable X. As we see, the random variable can take three possible values 0,1 and 2. The figure also clearly indicates that the event X=1 is twice as likely as the other two possible values. The Figure can be interpreted in the following way: If we repeat the random experiment (tossing a coin twice) a large number of times, then about half of the times we observe X=1, about a quarter of times we observe X=0, and about a quarter of times we observe X=2.

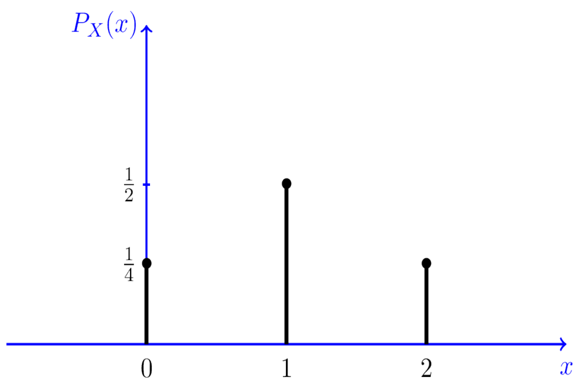

Figure 1: PMF for random Variable X in Example

For discrete random variables, the PMF is also called the probability distribution. Thus, when asked to find the probability distribution of a discrete random variable X, we can do this by finding its PMF. The phrase distribution function is usually reserved exclusively for the cumulative distribution function CDF (as defined later in the book). The word distribution, on the other hand, in this book is used in a broader sense and could refer to PMF, probability density function (PDF), or CDF.

Example 7. *I have an unfair coin for which $P(H) = p$, where $0 < p < 1$. I toss the coin repeatedly until I observe a heads for the first time. Let Y be the total number of coin tosses. Find the distribution of Y.*

Solution: *First, we note that the random variable Y can potentially take any positive integer, so we have $R_Y = N = \{1, 2, 3, ...\}$. To find the distribution of Y, we need to find*

$$P_Y(k) = P(Y = k) \text{ for } k = 1, 2, 3, ...$$

We have

$$P_Y(1) = P(Y = 1) = P(H) = p$$

,

$$P_Y(2) = P(Y = 2) = P(TH) = (1 - p)p$$

,

$$P_Y(3) = P(Y = 3) = P(TTH) = (1 - p)^2 p$$

,

$$....$$

$$....$$

$$....$$

$$P_Y(k) = P(Y = k) = P(TT...TH) = (1 - p)^{k-1} p$$

. *Thus, we can write the PMF of Y in the following way*

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

for $y = 1, 2, 3, ...$

Remark 8. *Properties of PMF:*

$$0 \leq P_X(x) \leq 1 \text{ for all } x$$

;

$$\sum_{x \in R_X} P_X(x) = 1$$

;

$$\text{for any set } A \subset R_X, P(X \in A) = \sum_{x \in A} P_X(x)$$

.

2.1.4 The Cumulative Distribution Function

Definition 9. *The cumulative distribution function $F(x)$ for X is a function giving the probability that the random variable X is less than or equal to x, for every value x. For a discrete random variable, the cumulative distribution function is found by summing up the probabilities. That is, $F_X(y) = P(X \leq y) = \sum_{x \leq y} p(x)$.*

Remark 10.

1. $\sum_{i=1}^{n} ar^i = \frac{a(1-r^{n+1})}{1-r}$
2. $\sum_{i=1}^{\infty} ar^i = \frac{a}{1-r}$, *if* $|r| < 1$.
3. *If the range of a discrete random variable X consists of the values $x_1 < x_2 < \cdots < x_n$ then $P_X(x_1) = F_X(x_1)$ and $P_X(x_i) = F_X(x_i) - F_X(x_{i-1}), i = 2, 3, \cdots, n$.*

2.2 Expected Value of a Discrete Random Variable

Remark 11.

1. $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$, *if* $|x| < 1$.
2. $\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$, *if* $|x| < 1$.

Definition 12.

let the range of a discrete random variable X be a sequence of numbers $x_1, x_2, \cdots, x_k$, and let $P_X(x)$ be the corresponding probability mass function. Then the expected value of X is

$$E(X) = x_1 P_X(x_1) + x_2 P_X(x_2) + \cdots + x_k P_X(x_k)$$

The expected value of X is also known as the mean value. The expected value of the random variable X is often denoted by the Greek letter μ, $E(X) = \mu$.

Remark 13.

1. *For any constant a and random variable X, $E(aX) = aE(X)$.*
2. *For any constants a, b and random variable X, $E(aX + b) = aE(X) + b$.*
3. *If X is a discrete random variable with range R_X and pmf $P(x)$, then the expected value of any function $g(X)$ is computed by*

$$E(g(X)) = \sum_{x \in R_X} g(x) P_X(x)$$

4. *If $g(x) = x^n$ then we call $E(X^n) = \sum_x x^n P_X(x)$ the n^{th} moment about the origin of X or the n^{th} raw moment. Thus, $E(X)$ is the first moment of X.*
5. *In our definition of expectation the set R_X can be countably infinite. It is possible to have a random variable with undefined expectation.*

Definition 14. *The mode of a probability function is the value of x which has the highest probability $P_X(x)$.*

Example 15. *The mode of the probability function for the number of claim is $x = 0$. as the table clearly shows.*

Number of claims (x)	*0*	*1*	*2*	*3*
$P_X(x)$	*.72*	*.22*	*.05*	*.01*

2.3 Variance and Standard Deviation

Definition 16.

The expected squared distance between the random variable and its mean is called the ***variance of the random variable****. The positive square root of the variance is called the* ***standard deviation of the random variable****. If σ_X denotes the standard deviation then the variance is given by the formula*

$$Var(X) = \sigma_X^2 = E[(X - E(X))^2 = \sum (x - \mu)^2 \cdot P_X(x) = \sigma_X^2$$

$$\sigma_X = \sqrt{Var(X)}$$

Remark 17.

1. *The variance of a random variable is typically calculated using the following formula*

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

2. *If X is a discrete random variable then for any constant a, we have $Var(aX) = a^2 Var(X)$.*

3. *Note that the units of Var(X) is the square of the units of X. This motivates the definition of the standard deviation $\sigma_X = \sqrt{Var(X)}$ which is measured in the same units as X. Also $\sigma_{aX} = |a|\sigma_X$.*

4. *If X is a discrete random variable then for any constants a and b we have $Var(aX+b) = a^2 Var(X)$.*

Theorem 18. *Let c be a constant and let X be a random variable with mean $E(X)$ and variance $Var(X) < \infty$. Then*

1. $g(c) = E[(X - c)^2] = Var(X) + (c - E(X))^2$

2. *$g(c)$ is minimized at $c = E(X)$.*

2.3.1 Comparing two Stocks

Example 19. *Forecast: The value of each stocks will increase by 5% if the national economy stays as it is. If the economic outlook improves, Stock A will increase in value by 10% and Stock B will increase in value by 15%. If the economic outlook deteriorated, Stock A will decrease in value by 10% and Stock B will decrease in value by 15%. You believe that probability for the future states of the economy are given by the following table*

State of the economy	*Deteriorate*	*unchanged*	*improve*
Probability	*.20*	*.60*	*.20*

This information enables you to craete probability function tables for the return on each of the two stocks.

% Change in value of Stock A: a	*-.10*	*0.05*	*+.10*
Probability: P(a)	*.20*	*.60*	*.20*

% Change in value of Stock B: a	*-.15*	*0.05*	*+.15*
Probability: P(B)	*.20*	*.60*	*.20*

since $E(A) = E(B) = 0.03$, $Var(A) = 0.0046$, $Var(B) = 0.0096$, *therefore* $\sigma_A \approx 0.068$ *and* $\sigma_B \approx 0.098$. *The standard deviation of the riskier stock is higher.*

2.3.2 Z-scores; Chebychev's Theorem

Definition 20. *For any possible value x of a random variable, the z-scores is*

$$z = \frac{x - \mu}{\sigma}$$

The z-scores measures the distance of x from $\mu = E(x)$ *in standard deviation units.*

Definition 21. *We say that a value x of the random variable X is within K standard deviations of the mean if* $|z| \leq k$.

Theorem 22. *For any random variable X, the probability that X is within k standard deviations of the mean is at least* $1 - \frac{1}{k^2}$.

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

Example 23. *The score for each student of a quiz is entered into a spreadsheet and a histogram created from the data. Suppose the histogram is bell-shaped, symmetrical about mean of 50 and st. dev of 10. What can we say about someone with a score of 50?*

Solution: *The z-score is* $(50 - 50)/10 = 0$. *Interpretation: student score is 0 distance (in units of standard deviations) from the mean, so the student has scored average.*

60? The z-score is $(60 - 50)/10 = 1$. *Interpretation: student has scored above average - a distance of 1 standard deviation above the mean.*

69.6? The z-score is $(69.6 - 50)/10 = 1.96$. *Interpretation: student has scored above average - a distance of 1.96 above the average score. Now, you might say after these examples that the z-score hasn't told you anything you can't see without doing any calculations. But the z-score can tell you a bit more. Because not only can it say whether a score is above or below the mean by so many st. dev., but it can tell you what proportion of the data is below or above a particular score. In other words you can make statements such as "so and so got X marks in a quiz. The corresponding Z-score is Y, and we can say that only Z per cent of students scored higher". And this is where the z-table comes in. Looking back the final example. The z-score was 1.96. Now if we look in the z-table we find that 2.5 per cent of the scores is above 1.96. So we can say that in the population this student did better than 97.5 per cent of students, or that only 2.5 per cent of students scored higher. What about for the student with z-score of zero - what proportion of students did better?*

Comparing scores from different normal distributions using the z-score

Example 24. *Suppose a student sits 2 exams, getting 55 in a verbal test and 60 in a numerical reasoning test. The class scores for each exam are normally distributed. For the verbal test, the mean is 50 and standard deviation 5; for the numerical test, the mean is 50 and standard deviation is 12.*

Now it is plain to see that the student did above average for each test, and did better at numerical reasoning. How did this student perform relative to everyone else? We can answer this by calculating the z-score.

The z-score for the verbal test is $(55 - 50)/5 = 1$.

The z-score for the numerical test is $(60 - 50)/12 = 0.83$

Since the z-score for the verbal test is larger than for the numerical test, the student did better in the verbal than in the numerical test compared to everyone else. Another way to see this is that z-score of 1 for verbal implies about 16 per cent did better at verbal; a z-score of 0.83 for numerical implies about 20 per cent did better at numerical.

2.3.3 Population and Sample Statistics

The difference between a population and a sample can be illustrated by returning to our probability function for the number of claims X filed by a policyholder with a large insurance company.

Number of claims (x)	0	1	2	3
P(x)	.72	.22	.05	.01

This is the probability function for all policyholders of the company-the entire population of policyholders.

$$\text{The mean } \mu = \sum x \cdot P(x) = .35$$

$$\text{The standard deiation } \sigma = \sqrt{\sum (x-\mu)^2 \cdot P(x)} = .622495$$

suppose the company had n=100,000 policyholders and had compiled the above table by looking at all records to obtain the following table:

Number of claims (x)	0	1	2	3
Number of policyholders with x claims(f)	72000	22000	5000	1000
P(x)	.72	.22	.05	.01

Therefore we have the following definition:

Definition 25.

$$\mu = \frac{1}{n} \sum f \cdot x$$

$$\sigma = \sqrt{\frac{1}{n} \sum f \cdot (x-\mu)^2}$$

In many case, it is not possible to gather complete data on an entire population. Then people who need information might take a sample of records to get an estimate of the mean and standard deviation of the population.

Suppose an analyst does not know the true values of μ and σ for the entire company population. She pick a sample of n=10 policyholder records from the company files, and find the following number on the records

Number of claims (x)	0	1	2
Number of policyholders with x claims(f)	7	2	1
P(x)	.70	.20	.10

$$\overline{x} = \frac{1}{n}(7(0) + 2(1) + 1(2)) = 0.40$$

$$s = \sqrt{\frac{1}{9}[7(0-0.40)^2 + 2(1-0.40)^2 + 1(2-0.40)^2]} \approx .699206$$

Definition 26.

$$\textit{Sample mean } \overline{x} = \frac{1}{n}\sum f \cdot x$$

$$\textit{Standard deviation } s = \sqrt{\frac{1}{n-1}\sum f \cdot (x-\overline{x})^2}$$

Statistics: unknown μ, σ and distribution, use $\overline{x}$ and s to estimate μ and σ and distribution.

Probability: Know μ, σ and distribution to solve problems for population.

2.4 Common Discrete Random Variables

As it turns out, there are some specific distributions that are used over and over in practice, thus they have been given special names. There is a random experiment behind each of these distributions. Since these random experiments model a lot of real life phenomenon, these special distributions are used frequently in different applications. That's why they have been given a name and we devote a section to study them. We will provide PMFs for all of these special random variables, but rather than trying to memorize the PMF, you should understand the random experiment behind each of them. If you understand the random experiments, you can simply derive the PMFs when you need them. Although it might seem that there are a lot of formulas in this section, there are in fact very few new concepts. Do not get intimidated by the large number of formulas, look at each distribution as a practice problem on discrete random variables.

2.4.1 Bernoulli Distribution

A Bernoulli random variable is a random variable that can only take two possible values, usually 0 and 1. This random variable models random experiments that have two possible outcomes, sometimes referred to as "success" and "failure." Here are some examples:

You take a pass-fail exam. You either pass (resulting in X=1) or fail (resulting in X=0).

You toss a coin. The outcome is ether heads or tails.
A child is born. The gender is either male or female.
Formally, the Bernoulli distribution is defined as follows:

Definition 27. *A random variable X is said to be a Bernoulli random variable with parameter p, shown as $X \sim Bernoulli(p)$, if its PMF is given by*

$$P_X(x) = \begin{cases} p & \textit{for } x = 1 \\ 1-p & \textit{for } x = 0 \\ 0 & \textit{otherwise} \end{cases}$$

where $0 < p < 1$.

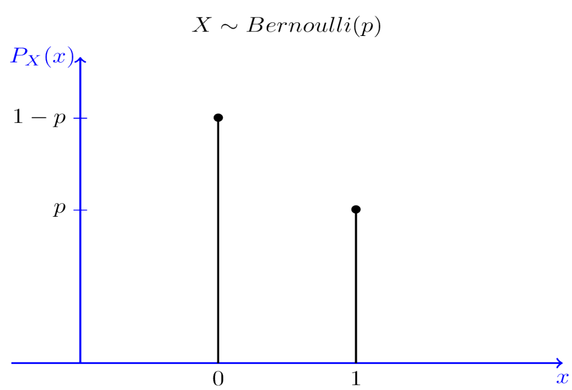

Figure 2: PMF of a Bernoulli(p) random variable.

Remark 28.
1. *If $X \sim Bernoulli(p)$, $E(X) = 1 \cdot p + 0 \cdot (1-p) = p$.*
2. *If $X \sim Bernoulli(p)$, $Var(X) = (1-p)^2 p + (0-p)^2(1-p) = p(1-p)$.*

A Bernoulli random variable is associated with a certain event A. If event A occurs (for example, if you pass the test), then X=1; otherwise X=0. For this reason the Bernoulli random variable, is also called the indicator random variable. In particular, the indicator random variable IA for an event A is defined by

$$I_A = \begin{cases} 1 & \text{if the event } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

The indicator random variable for an event A has Bernoulli distribution with parameter p=P(A), so we can write

$$I_A \sim Bernoulli(P(A))$$

.

2.4.2 Geometric Distribution

The random experiment behind the geometric distribution is as follows. Suppose that I have a coin with $P(H) = p$. I toss the coin until I observe the first heads. We define X as the total number of coin tosses in this experiment. Then X is said to have geometric distribution with parameter p. In other words, you can think of this experiment as repeating independent Bernoulli trials until observing the first success.

The range of X here is $R_X = \{1, 2, 3, ...\}$.

$$P_X(k) = P(X = k) = (1-p)^{k-1}p, \text{ for } k = 1, 2, 3, ...$$

We usually define $q = 1 - p$, so we can write $P_X(k) = pq^{k-1}$, for $k = 1, 2, 3, ...$ To say that a random variable has geometric distribution with parameter p, we write $X \sim Geometric(p)$. More formally, we have the following definition:

Definition 29. *A random variable X is said to be a geometric random variable with parameter p, shown as $X \sim Geometric(p)$, if its PMF is given by*

$$P_X(k) = \begin{cases} p(1-p)^{k-1} & \textit{for } k = 1, 2, 3, ... \\ 0 & \textit{otherwise} \end{cases}$$

where $0 < p < 1$.

Remark 30. 1. *If $X \sim Geometric(p)$,*

$$E(X) = p + 2pq + 3pq^2 + ... + kpq^{k-1} + ... = p(1 + 2q + 3q^2 + 4q^3 + ...)$$

$$= p(q + q^2 + q^3 + q^4 + ...)' = p\left(\frac{1}{1-q}\right)' = \frac{p}{(1-q)^2}$$

2. *If $X \sim Geometric(p)$, $Var(X) = \frac{q}{p^2}$ the proof is complex.*

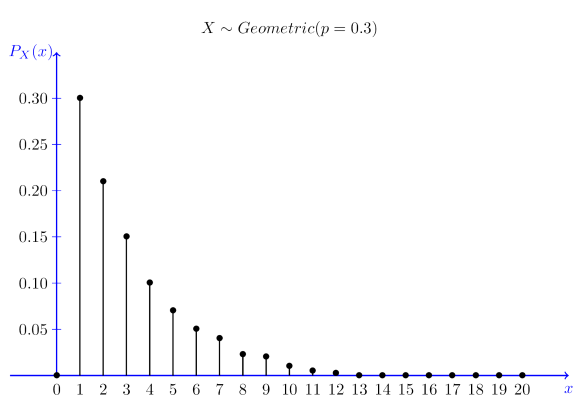

Figure 3: PMF of a Geometric(0.3) random variable.

2.4.3 Binomial Distribution

The random experiment behind the binomial distribution is as follows. Suppose that I have a coin with $P(H) = p$. I toss the coin n times and define X to be the total number of heads that I observe. Then X is binomial with parameter n and p, and we write $X \sim Binomial(n, p)$. The range of X in this case is $R_X = \{0, 1, 2, ..., n\}$. As we have seen in previous Section, the PMF of X in this case is given by binomial formula

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ for } k = 0, 1, 2, ..., n.$$

Definition 31. *A random variable X is said to be a binomial random variable with parameters n and p, shown as $X \sim Binomial(n, p)$, if its PMF is given by*

$$P_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

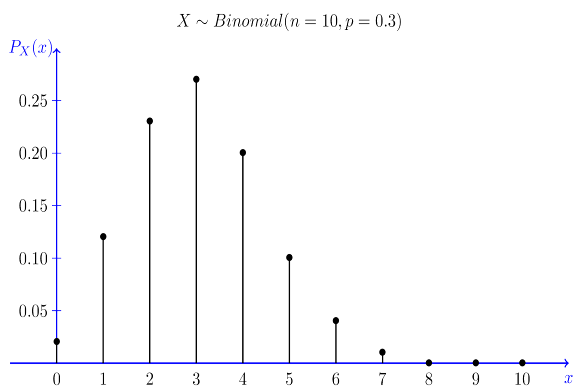

Figure 4: PMF of a Binomial(10,0.3) random variable.

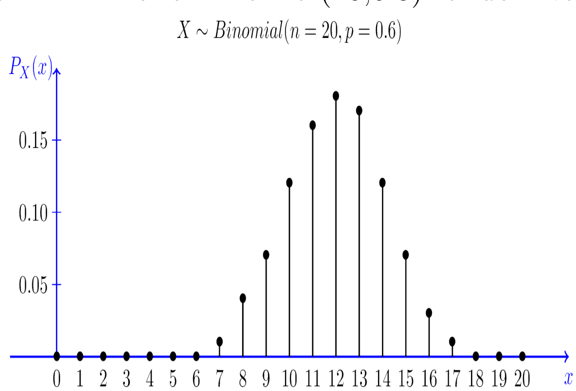

Figure 5: PMF of a Binomial(20,0.6) random variable.

Remark 32. *Binomial random variable as a sum of Bernoulli random variables. Here is a useful way of thinking about a binomial random variable. Note that a Binomial(n,p) random* *variable can be obtained by n independent coin tosses. If we think of each coin toss as a Bernoulli(p) random variable, the Binomial(n,p) random variable is a sum of n independent Bernoulli(p) random variables. This is stated more precisely in the following lemma.*

Lemma 33. *If $X_1, X_2, ..., X_n$ are independent Bernoulli(p) random variables, then the random variable X defined by $X = X_1 + X_2 + ... + X_n$ has a Binomial(n,p) distribution.*

Remark 34. *1. If $X \sim Binomial(p)$, $E(X) = np$.*

2. If $X \sim Binomial(p)$, $Var(X) = np(1-p)$.

To generate a random variable $X \sim Binomial(n,p)$, we can toss a coin n times and count the number of heads. Counting the number of heads is exactly the same as finding $X_1 + X_2 + ... + X_n$, where each X_i is equal to one if the corresponding coin toss results in heads and zero otherwise. This interpretation of binomial random variables is sometimes very helpful. Let's look at an example.

Example 35. *Let $X \sim Binomial(n,p)$ and $Y \sim Binomial(m,p)$ be two independent random variables. Define a new random variable as $Z = X + Y$. Find the PMF of Z.*

Solution: *Since $X \sim Binomial(n,p)$, we can think of X as the number of heads in n independent coin tosses, i.e., we can write*

$$X = X_1 + X_2 + ... + X_n$$

, where the X_i's are independent Bernoulli(p) random variables. Similarly, since $Y \sim Binomial(m,p)$, we can think of Y as the number of heads in m independent coin tosses, i.e., we can write

$$Y = Y_1 + Y_2 + ... + Y_m$$

, where the Yj's are independent Bernoulli(p) random variables. Thus, the random variable $Z = X + Y$ will be the total number of heads in n+m independent coin tosses:

$$Z = X + Y = X_1 + X_2 + ... + X_n + Y_1 + Y_2 + ... + Y_m$$

, where the X_i's and Y_j's are independent Bernoulli(p) random variables. Thus, Z is a binomial random variable with parameters $m+n$ and p, i.e., Binomial$(m+n, p)$. Therefore, the PMF of Z is

$$P_Z(k) = \begin{cases} \binom{m+n}{k} p^k (1-p)^{m+n-k} & \text{for } k = 0, 1, 2, 3, ..., m+n \\ 0 & \text{otherwise} \end{cases}$$

Negative Binomial (Pascal) Distribution

Suppose that I have a coin with $P(H) = p$. I toss the coin until I observe m heads, where $m \in \mathbb{N}$. We define X as the total number of coin tosses in this experiment. Then X is said to have Pascal distribution with parameter m and p. We write $X \sim Pascal(m,p)$. Note that $Pascal(1,p) = Geometric(p)$. Note that by our definition the range of X is given by $R_X = \{m, m+1, m+2, m+3, ...\}$.

Let us derive the PMF of a $Pascal(m,p)$ random variable X. Suppose that I toss the coin until I observe m heads, and X is defined as the total number of coin tosses in this experiment. To find the probability of the event $A = \{X = k\}$, we argue as follows. By definition, event A can be written as $A = B \cap C$, where B is the event that we observe $m-1$ heads (successes) in the first $k-1$ trials, and C is the event that we observe a heads in the k^{th} trial. Note that B and C are independent events because they are related to different independent trials (coin tosses). Thus we can write

$$P(A) = P(B \cap C) = P(B)P(C).$$

Now, we have $P(C) = p$. Note also that $P(B)$ is the probability that I observe observe m1 heads in the k1 coin tosses. This probability is given by the binomial formula, in particular

$$P(B) = \binom{k-1}{m-1} p^{m-1}(1-p)^{\left((k-1)-(m-1)\right)} = \binom{k-1}{m-1} p^{m-1}(1-p)^{k-m}.$$

Definition 36. *A random variable X is said to be a Pascal random variable with parameters m and p, shown as $X \sim Pascal(m, p)$, if its PMF is given by*

$$P_X(k) = \begin{cases} \binom{k-1}{m-1} p^m (1-p)^{k-m} & \text{for } k = m, m+1, m+2, m+3, ... \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

Remark 37. *Negative Binomial Distribution is m copies of Geometric distribution, therefore,*

1. $E(X) = \frac{mq}{p}$
2. $Var(X) = \frac{mq}{p^2}$

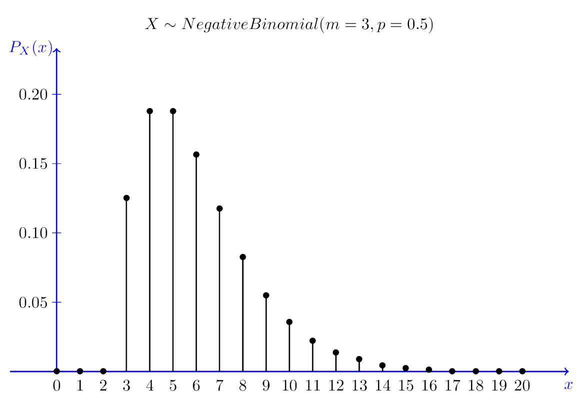

Figure 6: Pascal(3,0.5) (negative binomial) random variable.

2.4.4 Hypergeometric Distribution

Here is the random experiment behind the hypergeometric distribution. You have a bag that contains b blue marbles and r red marbles. You choose $k \leq b+r$ marbles at random (without replacement). Let X be the number of blue marbles in your sample. By this definition, we have $X \leq \min(k, b)$. Also, the number of red marbles in your sample must be less than or equal to r, so we conclude $X \geq \max(0, k - r)$. Therefore, the range of X is given by

$$R_X = \{\max(0, k - r), \max(0, k - r) + 1, \max(0, k - r) + 2, ..., \min(k, b)\}.$$

To find $P_X(x)$, note that the total number of ways to choose k marbles from b+r marbles is $\binom{b+r}{k}$. The total number of ways to choose x blue marbles and $k-x$ red marbles is $\binom{b}{x}\binom{r}{k-x}$. Thus, we have

$$P_X(x) = \frac{\binom{b}{x}\binom{r}{k-x}}{\binom{b+r}{k}}, \qquad \text{for } x \in R_X.$$

Definition 38. *A random variable X is said to be a Hypergeometric random variable with parameters b,r and k, shown as $X \sim Hypergeometric(b, r, k)$, if its range is*

$$R_X = \{\max(0, kr), \max(0, kr) + 1, \max(0, kr) + 2, ..., \min(k, b)\}$$

, and its PMF is given by

$$P_X(x) = \begin{cases} \frac{\binom{b}{x}\binom{r}{k-x}}{\binom{b+r}{k}} & \textit{for } x \in R_X \\ 0 & \textit{otherwise} \end{cases}$$

Remark 39.

1. $E(X) = k\frac{b}{b+r}$
2. $Var(X) = k(\frac{b}{b+r})(1 - \frac{b}{b+r})(\frac{b+r-k}{b+r-1})$

2.4.5 Poisson Distribution

The Poisson distribution is one of the most widely used probability distributions. It is usually used in scenarios where we are counting the occurrences of certain events in an interval of time or space. In practice, it is often an approximation of a real-life random variable. Here is an example of a scenario where a Poisson random variable might be used. Suppose that we are counting the number of customers who visit a certain store from 1pm to 2pm. Based on data from previous days, we know that on average $\lambda = 15$ customers visit the store. Of course, there will be more customers some days and fewer on others. Here, we may model the random variable X showing the number customers as a Poisson random variable with parameter $\lambda = 15$. Let us introduce the Poisson PMF first, and then we will talk about more examples and interpretations of this distribution.

Definition 40. *A random variable X is said to be a Poisson random variable with parameter λ, shown as $X \sim Poisson(\lambda)$, if its range is $R_X = \{0, 1, 2, 3, ...\}$, and its PMF is given by*

$$P_X(k) = \begin{cases} \frac{e^{-\lambda}\lambda^k}{k!} & \textit{for } k \in R_X \\ 0 & \textit{otherwise} \end{cases}$$

Before going any further, let's check that this is a valid PMF. First, we note that $P_X(k) \geq 0$ for all k. Next, we need to check $\sum_{k \in R_X} P_X(k) = 1$. To do that, let us first remember the Taylor series for e^x, $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. Now we can write

$$\sum_{k \in R_X} P_X(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!}$$
$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$
$$= e^{-\lambda} e^{\lambda} = 1$$

Remark 41. $E(X) = Var(X) = \lambda$.

Example 42. *A busy intersection is the scene of many traffic accidents. An analyst studies data on the accidents and concludes that accidents occurs there at "an average rate of $\lambda = 2$ per month". This does not means there are 2 accidents in each month. In any given number of accidents X in a month is a random variable. The Possion distribution can be used to find the probabilities $P(X = k)$ in terms of k and λ, the average rate.*

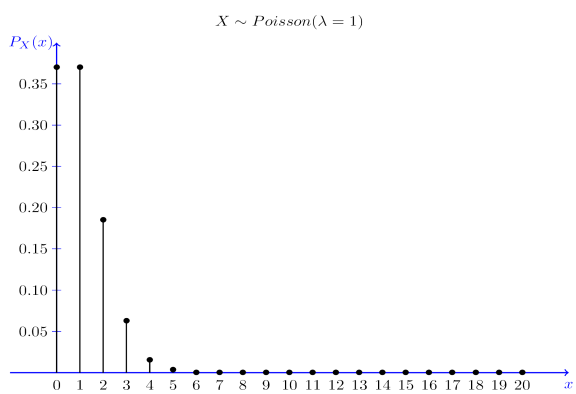

Figure 7: PMF of a Poisson(1) random variable.

2.4.6 Poisson as an approximation for binomial

The Poisson distribution can be viewed as the limit of binomial distribution. Suppose $X \sim Binomial(n, p)$ where n is very large and p is very small. In particular, assume that $\lambda = np$ is a positive constant. We show that the PMF of X can be approximated by the PMF of a $Poisson(\lambda)$ random variable. The importance of this is that Poisson PMF is much easier to compute than the binomial. Let us state this as a theorem.

Assumption 1: The probability of exactly one accident in a small time interval of length is approximately λt.

Assumption 2: Accident occur independently in time intervals which do not intersect.

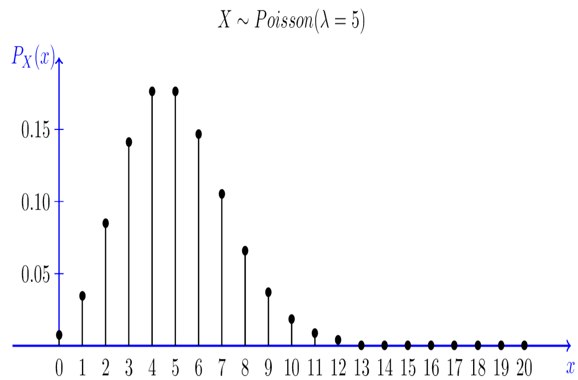

Figure 8: PMF of a Poisson(5) random variable.

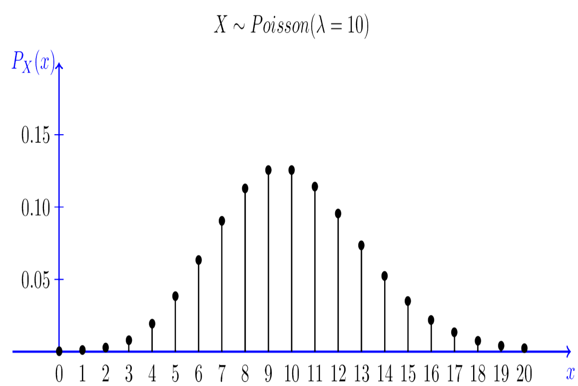

Figure 9: PMF of a Poisson(10) random variable.

Theorem 43. *Let $X \sim Binomial(n, p = \frac{\lambda}{n})$, where $\lambda > 0$ is fixed. Then for any $k \in \{0, 1, 2, ...\}$, we have*

$$\lim_{n\to\infty} P_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

2.4.7 The Discrete Uniform Distribution

When we roll a single fair die and observe the number X that came up, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$ and each of the outcomes was equally likely with probability 1/6. The random variable X is said to have a discrete uniform distribution on 1,2,,3,4,5,6.

Definition 44 (Discrete Uniform Distribution on 1,2,...,n)**.**

1. $P(x) = \frac{1}{n}, x = 1, 2, 3, 4, ..., n$
2. $E(X) = \frac{n+1}{2}$
3. $Var(X) = \frac{n^2-1}{12}$.

2.5 Applications for Discrete Random Variables

2.5.1 Y=f(X) in general

Lemma 45. *1.* $E(aX+b) = a \cdot E(X) + b$

2. $Var(aX+b) = a^2 \cdot Var(X)$

3. $E(f(X)) = \sum_x f(x) \cdot P(x)$

Example 46. *Let the random variable X have the distribution below.*

x	-1	0	1
$P(x)$	.20	.60	.20

$$E(X) = \sum_x x \cdot P(x) = -1 \cot(0.20) + 0 \cdot (0.60) + 1 \cdot (0.20) = 0$$

If $y = f(x) = x^2$*, then the distribution for* $f(x)$*,*

$y = f(x) = x^2$	1	0
$P(y)$	.40	.60

$$E(Y) = \sum_y y \cdot P(y) = 0.60(0) + 0.40(1) = 0.40$$

$f(x) = x^2$	1	0	1
$P(x)$	.20	.60	.20

$$E(f(X)) = \sum_x f(x) \cdot P(x) = 1 \cdot (0.20) + 0 \cdot (0.60) + 1 \cdot (0.20) = 0.40$$

2.5.2 Another Way to Calculate the Variance of a Random Variable

Remember that we define the variance of a random variable X by

$$Var(X) = E[(X-\mu)^2] = \sum (x-\mu)^2 \cdot P(x)$$

It is very easy to prove that

Theorem 47.

$$Var(X) = E(X^2) - \mu^2 = E(X^2) - (E(X))^2$$

2.5.3 Moment and Moment Generating Function

Definition 48. *Let X be a distance random variable. The moment generating function, denoted* $M_X(t)$*, is defined by*

$$M_X(t) = E(e^{tX} = \sum e^{tx} \cdot P(x)$$

$$M_X^{(n)}(0) = \sum x^n \cdot P(x) = E(X^n) \text{ is called the } n^{th} \text{ Moment of X.}$$

$$M_X'(0) = E(X) \text{ is the mean of the random variable X, called the first Moment.}$$

$$M_X''(0) = E(X^2) \text{ could be used to calculate Variance, called the second Moment.}$$

2.5.4 Moment generating Function for Discrete Random Variable

Remark 49. *If $X \sim Binomial(1, p)$,*

x	*0*	*1*
e^{tx}	*1*	e^t
P(x)	q	p

$$M_X(t) = E(e^{tX}] = 1 \cdot q + p \cdot e^t = q + pe^t$$

Remark 50. *If $X \sim Binomial(2, p)$,*

x	*0*	*1*	*2*
e^{tx}	*1*	e^t	e^{2t}
P(x)	q^2	*2pq*	p^2

$$M_X(t) = E(e^{tX}] = (q + pe^t)^2$$

In general, If $X \sim Binomial(n, p)$, then $M_X(t) = E(e^{tX}] = (q + pe^t)^n$

Lemma 51.

If $X \sim Poisson(\infty, \lambda)$, then $M_X(t) = E(e^{tX}) = e^{\lambda(e^t - 1)}$

If $X \sim Geometric(\infty, p)$, then $M_X(t) = E(e^{tX}) = \dfrac{p}{1 - qe^t}$

If $X \sim Nbinomial(r, p)$ (X is the number of observation of r^{th} success), then

$$M_X(t) = E(e^{tX}] = \left(\frac{p}{1 - qe^t}\right)^r$$

A useful identity $M_{aX+b}(t) = E(e^{(aX+b)t}) = e^{bt}E(e^{aXt}) = e^{bt}M_{aX}(t)$

Example 52. *Suppose X is Poisson with $\lambda = 2$. Let $Y = 3X + 5$. Then $M_X(t) = e^{2(e^t - 1)}$ and*

$$M_Y(t) = e^{5t}M_{3X}(t) = e^{5t}e^{2(e^{3t} - 1)}$$

2.5.5 The Cumulative Distribution Function

Remark 53.

1. *if X is a random variable then the cumulative distribution function (abbreviated c.d.f) is the function $F_X(x) = P(X \leq x)$.*
2. *F is a nondecreasing function, that is, if $a < b$ then $F_X(a) \leq F_X(b)$.*
3. *F is continuous from the right. That is, $\lim_{x \to b^+} F_X(t) = F_X(b)$.*
4. $\lim_{b \to -\infty} F_X(b) = 0, \lim_{b \to \infty} F_X(b) = 1$.
5. *For any random variable X and any real number a, we have $P(X > a) = 1 - F_X(a)$.*

6. *For any random variable X and any real number a, we have* $P(X < a) = \lim_{n\to\infty} F_X(a - \frac{1}{n}) = F(a^-)$.

7. $P(X \geq a) = 1 - \lim_{n\to\infty} F(a - \frac{1}{n}) = 1 - F(a^-)$

8. *If* $a < b$ *then* $P(a < X \leq b) = F_X(b) - F_X(a)$.

9. *If* $a < b$ *then* $P(a \leq X < b) = F_X(b^-) - F_X(a^-)$.

10. *If* $a < b$ *then* $P(a \leq X \leq b) = F_X(b) - \lim_{n\to\infty} F_X(a - \frac{1}{n}) = F_X(b) - F_X(a^-)$.

11. *If* $a < b$ *then* $P(a < X < b) = F_X(b^-) - F_X(a)$.

Example 54. 1. *If $X \sim Binomial(n, p)$, then cumulative distribution function is given by*

$$\begin{cases} 0, & x < 0, \\ \sum_{k=0}^{[x]} C(n,k)p^k q^{n-k}, & 0 \leq x \leq n \\ 1, & x > n. \end{cases}$$

where [x] is the ***floor function****.*

2. *If $X \sim Geometric(k, p)$, then cdf of X given by*

$$F(x) = Pr(X \leq x) = \begin{cases} 0, & x < 1, \\ 1 - (1-p)^{[x]}, & x \geq 1. \end{cases}$$