Chapter 3 Discrete Random Variables and Their Distributions

3.1. <u>Random Variables</u>

Random Variable
A *random variable* is a function of an outcome,

$$X = f(\omega).$$

In other words, it is a quantity that depends on chance.

Discrete Random Variable
A random variable that assumes countable values is called a *discrete random variable*.
<u>Example:</u>

1. The number of jobs submitted to a printer.
2. The number of errors, error-free modules, number of failed components.
3. The number of heads obtained in three tosses of a coin.

<u>Note.</u> Discrete variables don't have to be integer. This variable assumes a finite number.

Continuous Random Variable
A random variable that can assume any value contained in one or more intervals is called a *continuous random variable*.
<u>Example:</u>

1. The software installation time, code execution time, connection time.
2. The time taken to complete an examination.
3. The weight, height, voltage, temperature, distance.

<u>Note.</u> Rounding a continuous random variable, say, to the nearest integer makes it discrete. Sometimes we can see mixed random variables that are discrete on some range of values and continuous elsewhere.

3.2. <u>Probability Distribution of A Discrete Random Variable</u>

Probability Mass Function
Collection of all the probabilities related to a discrete random variable X is the distribution of X. The function

$$P_X(x) = P(X = x)$$

is the *probability mass function*, or ***pmf***.

<u>*Two* characteristics of a probability mass function.</u>

1. $0 \le P_X(x) \le 1$ for each value of x
2. $\sum_{x\in\Omega} P_X(x) = 1$

Cumulative Distribution Function

The *cumulative distribution function*, or ***cdf*** of a random variable X is defined as

$$F_X(x) = P(X \le x) = \sum_{y \le x} P(y).$$

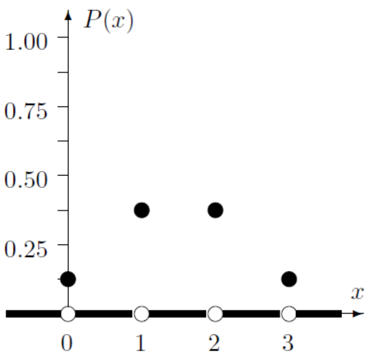
The set of possible values of X is called the *support* of the distribution F.

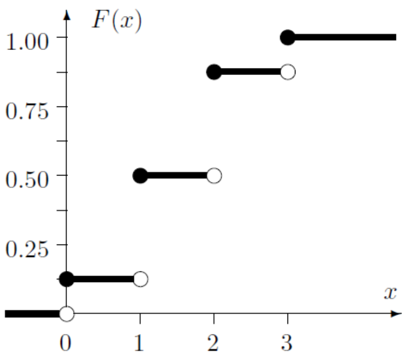
<u>*Some* characteristics of a cumulative distribution function.</u>

1. $F_X(x)$ is a non-decreasing function of *x*.
2. $F_X(x)$ is continuous from the right-hand side.
3. $F(-\infty) = \lim_{x \to -\infty} F_X(x) = 0$, $\quad F(\infty) = \lim_{x \to \infty} F_X(x) = 1$
4. $0 \le F_X(x) \le 1$

<u>*Example 3.1.*</u>

Consider an experiment of tossing 3 fair coins and counting the number of heads. Let X be the number of heads obtained. The probability mass function $P_X(x)$ and the cumulative distribution function $F_X(x)$ of X are shown in the following figure. White circles denote excluded points.





Example 3.2.

According to a survey, 60% of all students at a large university suffer from math anxiety. Two students are randomly selected from this university. Let X denote the number of students in this sample who suffer from math anxiety, compute the probability mass function (*pmf*) and the cumulative distribution function (*cdf*) of X. Then, draw a graph for its *pmf* and *cdf* .

Solution:

Example 3.3.

For each of the following, determine the constant k so that $P_X(x)$ satisfies the conditions of being a *pmf* for a random variable X, and then develop the probability distribution of X.

(a) $P_X(x) = \dfrac{x}{k}$ for $x = 2, 3, 4, 5.$

(b) $P_X(x) = k(x+1)^2$ for $x = 0, 2, 4, 6.$

Solution:

Example 3.4.

A program consists of two modules. The number of errors X_1 in the first module has the *pmf* $P_1(x)$, and the number of errors X_2 in the second module has the *pmf* $P_2(x)$, independently of X_1, where

x	$P_1(x)$	$P_2(x)$
0	0.5	0.7
1	0.3	0.2
2	0.1	0.1
3	0.1	0

Find the *pmf* and *cdf* of $Y = X_1 + X_2$, the total number of errors.

Solution:

3.3. Distribution of a Random Vector

Often we deal with several random variables simultaneously. We may look at the size of a RAM and the speed of a CPU, the price of a computer and its capacity, temperature and humidity, technical and artistic performance, etc.

Joint Distribution and Marginal Distributions

If X and Y are random variables, then the pair (X,Y) is a *random vector*. Its distribution is called the *joint distribution* of X and Y. Individual distributions of If X and Y are then called the *marginal distributions*.

Similarly to a single variable, the *joint distribution* of a vector is a collection of probabilities for a vector (X,Y) to take value (x,y). Recall that two vectors are equal,

$$(X,Y)=(x,y),$$

if $X=x$ and $Y=y$. This "and" means the intersection, therefore, the *joint probability mass function* of X and Y is

$$P(x,y)=P\big((X,Y)=(x,y)\big)=P(X=x\cap Y=y).$$

Again, $\{(X,Y)=(x,y)\}$ are exhaustive and mutually exclusive events for different pairs (x,y), therefore,

$$\sum_x\sum_y P(x,y)=1.$$

Addition Rule

The joint distribution of (X,Y) carries the complete information about the behavior of this random vector. In particular, the marginal probability mass functions of X and Y can be obtained from the joint *pmf* by the *Addition Rule*, namely,

$$P_X(x)=P(X=x)=\sum_y P_{(X,Y)}(x,y)$$
$$P_Y(y)=P(Y=y)=\sum_x P_{(X,Y)}(x,y)$$

That is, to get the marginal *pmf* of one variable, we add the joint probabilities over all values to the other variable.

In general, the joint distribution cannot be computed from marginal distributions because they carry no information about interrelations between random variables. For example, marginal distributions cannot tell whether variables X and Y are independent or dependent.

Independence of Random Variables

Random variables X and Y are *independent* if

$$P_{(X,Y)}(x,y)=P_X(x)P_Y(y)$$

for *all* values of x and y. This means, events $\{X=x\}$ and $\{Y=y\}$ are independent for all x and y; in other words, variables X and Y take their values independently of each other. To prove dependence, we only need to present one counterexample, a pair (x,y) with

$$P_{(X,Y)}(x,y)\neq P_X(x)P_Y(y).$$

<u>Example 3.5.</u>

A program consists of two modules. The number of errors (X) in the first module and the number of errors (Y) in the second module have the joint distribution, $P(0,0)=P(0,1)=P(1,0)=0.2$, $P(1,1)=P(1,2)=P(1,3)=0.1$, $P(0,2)=P(0,3)=0.05$.

Find

(a) the marginal distributions of X and Y,

(b) the probability of no errors in the first module,

(c) the distribution of the total number of errors in the program.

(d) Determine if errors in the two modules occur independently.

Solution:

(a)

$P_{(X,Y)}(x,y)$		y				$P_X(x)$
		0	1	2	3	
x	0					
	1					
$P_Y(y)$						

3.4. Expectation and Variance

Expectation

The *expectation* of a discrete random variable X is its *mean*, the average value and is computed as

$$E(X) = \sum_x x \cdot P_X(x).$$

Properties of expectations.

For any random variables X and Y and any non-random numbers a, b and c:

1. $E(X+Y) = E(X) + E(Y)$
2. $E(aX) = aE(X)$
3. $E(c) = c$
4. $E(aX + bY + c) = aE(X) + bE(Y) + c$
5. $E(XY) = E(X)E(Y)$ for *independent* X and Y

Example 3.6.

Reconsider *Example 3.5.*, compute $E(X)$, $E(Y)$ and the expected total number of errors.

Solution:

Variance and Standard Deviation

The *variance* of a discrete random variable X is defined as the expected squared deviation from the mean, namely,

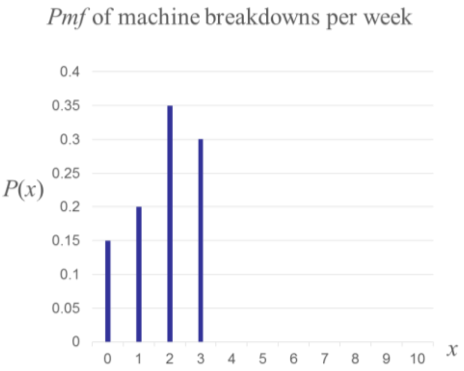
$$Var(X) = E(X - E(X))^2 = E(X^2) - [E(X)]^2$$
$$= \sum_x (x-\mu)^2 P_X(x) = \sum_x [x^2 \cdot P_X(x)] - \mu^2$$

The *standard deviation* of a discrete random variable X is given by

$$Std(X) = \sqrt{Var(X)}.$$

<u>*Example 3.7.*</u>

The following bar graph depicts the probability distribution of the number of breakdowns per week for a machine based on past data.



Find the mean and standard deviation of the number of breakdowns per week for this machine.

Solution:

Let X denote the number of breakdowns for this machine in a given week.

x	0	1	2	3	
$P_X(x)$					
$x \cdot P_X(x)$					$\sum x \cdot P_X(x) =$
$x^2 \cdot P_X(x)$					$\sum x^2 \cdot P_X(x) =$

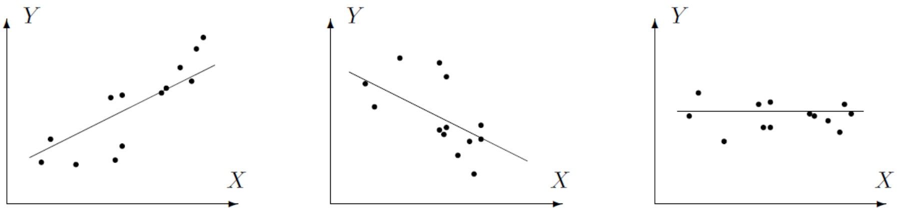
3.5. Covariance and Correlation

Covariance

Covariance $\sigma_{XY} = Cov(X,Y)$ is defined as

$$\begin{aligned} Cov(X,Y) &= E\left[(X - E(X))(Y - E(Y))\right] \\ &= E(XY) - E(X)E(Y) \end{aligned}.$$

It summarizes interrelation of two random variables.



(a) $\mathrm{Cov}(X,Y) > 0$ (b) $\mathrm{Cov}(X,Y) < 0$ (c) $\mathrm{Cov}(X,Y) = 0$

Correlation

Correlation coefficient between variables X and Y is defined as

$$\rho = \frac{Cov(X,Y)}{Std(X) \times Std(Y)}, \quad -1 \le \rho \le 1.$$

It measures the association of two random variables.

Properties of variances and covariances.

For any random variables X,Y,Z,W and any non-random numbers a,b,c,d :

1. $Cov(X,Y) = Cov(Y,X)$
2. $\rho(X,Y) = \rho(Y,X)$
3. $Var(aX+b) = a^2 Var(X)$
4. $Var(aX+bY+c) = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X,Y)$
5. $Cov(aX+b, cY+d) = ac Cov(X,Y)$
6. $Cov(aX+bY, cZ+dW) = ac Cov(X,Z) + ad Cov(X,W) + bc Cov(Y,Z) + bd Cov(Y,W)$
7. $\rho(aX+b, cY+d) = \rho(X,Y)$

For *independent* X and Y,

8. $Cov(X,Y) = 0$
9. $Var(X+Y) = Var(X) + Var(Y)$

3.6. Application to Finance

Chebyshev's inequality shows that in general, higher variance implies higher probabilities of *large deviations*, and this increases the *risk* for a random variable to take values far from its expectation.

This finds a number of immediate applications. Here we focus on evaluating risks of financial deals, allocating funds, and constructing optimal portfolios. This application is intuitively simple. The same methods can be used for the optimal allocation of computer memory, CPU time, customer support, or other resources.

Example 3.8.

We would like to invest \$10,000 into shares of companies XX and YY. Shares of XX cost \$20 per share. The market analysis shows that their expected return is \$1 per share with a standard deviation of \$0.5. Shares of YY cost \$50 per share, with an expected return of \$2.50 and a standard deviation of \$1 per share, and returns from the two companies are independent. In order to maximize the expected return and minimize the risk (standard deviation or variance), is it better to invest (A) all \$10,000 into XX, (B) all \$10,000 into YY, or (C) \$5,000 in each company?

Solution:

Let X be the actual (random) return from each share of XX

Y be the actual return from each share of YY

(A) With \$10000, we can buy 500 shares of XX which cost \$20/share, collecting a profit of $A = 500X$

(B) With \$10000, we can buy 200 shares of YY which cost \$50/share, collecting a profit of $B = 200Y$

(C) With \$5000 each, we can buy 250 shares of XX which cost \$20/share and 100 shares of YY which cost \$50/share, collecting a profit of $C = 250X + 100Y$

In terms of the expected return, all three portfolios are equivalent.

i.e. $1/20$ or $2.5/50$, each share of each company is expected to a return of 5%

Example 3.9.

Suppose now that the individual stock returns X and Y are no longer independent. If the correlation coefficient is $\rho = 0.4$, how will it change the results of the previous example? What if they are negatively correlated with $\rho = -0.2$?

Solution:

Only the volatility of portfolio C changes due to the correlation coefficient.

For $\rho = 0.4$,

Now if $\rho = -0.2$,

The risk of portfolio C increase due to positive correlation of the two stocks. When X and Y are positively correlated, low values of X are likely to accompany low values of Y; therefore, the probability of the overall low return is higher, increasing the risk of the portfolio.

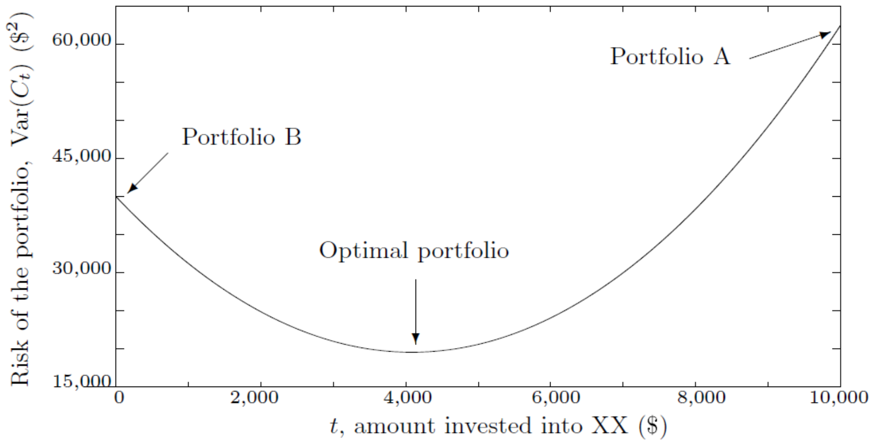
Conversely, negative correlation means that low values of X are likely to be compensated by high values of Y , and vice versa. Thus, the risk is reduced. Diversified portfolios consisting of negatively correlated components are the least risky.

<u>*Example 3.10.*</u>
So, after all, with \$10000 to invest, what is the most optimal portfolio consisting of shares of XX and YY, given their correlation coefficient of $\rho = -0.2$?

Suppose t dollars are invested into XX and $(10000 - t)$ dollars into YY, with the resulting profit is C_t. This amounts for $t/20$ shares of XX and $(10000-t)/50 = 200 - t/50$ shares of YY. Plans A and B correspond to $t = 10000$ and $t = 0$.

Solution:

The following figure shows the variance of a diversified portfolio.



The minimum of this variance is found at

Thus, for the most optimal portfolio, we should invest \$4081.63 into XX and the remaining \$5919.37 into YY. Then we achieve the smallest possible risk (variance) of (\$^{2})19592, measured, as we know, in squared dollars

3.7. Families of Discrete Distributions – The Bernoulli Distribution

A random variable with two possible values, 0 and 1, is called a ***Bernoulli variable***, its distribution is ***Bernoulli distribution***. Any experiment in which only two outcomes (mutually exclusive and exhaustive) are possible, say, "success" or "failure", is called a ***Bernoulli trial***.

Let *p* denote the $P(\textit{success})$ and *q* denote the $P(\textit{failure})$ on each trial, the probability mass function of a discrete random variable *X* follows a Bernoulli distribution is given by

$$P(X=x)=p^x q^{1-x} \text{ for } x=0,1$$

where $q=1-p$.

Mean and Variance

Suppose $X \sim B(p)$. Then

$$\mu=p \text{ and } \sigma^2=pq.$$

Example:

Trial	"Success"	"Failure"
1. Toss a coin	H	T
2. Roll a die	Six	1, 2, 3, 4, 5
3. Fire a riffle	Hit	Missed

3.8. Families of Discrete Distributions – The Binomial Distribution

A variable described as the number of successes in a sequence of independent Bernoulli trials has ***Binomial distribution***. Its parameters are *n*, the number of trials, and *p*, the probability of success.

Example:

Trial	"Success"	"Failure"	Binomial Experiment	Binomial Random Variable *X*
1. Toss a coin	H	T	Toss 5 times	# of H obtained
2. Roll a die	Six	1, 2, 3, 4, 5	Roll 10 times	# of sixes obtained
3. Fire a riffle	Hit	Missed	Fire 7 times	# of hits obtained

Suppose a discrete random variable *X* follows a Binomial distribution. Then the probability of exactly *x* successes in *n* trials is given by

$$P(X=x)=\binom{n}{x}p^x q^{n-x} \text{ for } x=0,1,2,\ldots,n$$

where $q=1-p$.

Mean and Variance

Suppose $X \sim Bin(n,p)$. Then

$$\mu=np \text{ and } \sigma^2=npq.$$

Note.

Suppose $X \sim Bin(n,p)$ and $Y=n-X$. Then $Y \sim Bin(n,1-p)$.

Example 3.11.
As part of a business strategy, randomly selected 20% of new internet service subscribers receive a special promotion from the provider. A group of 10 neighbors signs for the service. What is the probability that at least 4 of them get a special promotion?
Solution:

Example 3.12.
Suppose that $X \sim Bin(n=20, p=0.3)$. Find

(a) $P(X=6)$

(b) $P(4 \le X < 12)$

(c) $P(X > 2)$

(d) $E(X)$ and $Var(X)$

Example 3.13.
Suppose $X \sim Bin(n=10, p=0.8)$ and $Y = 10 - X$, $Y \sim Bin(n=10, p=0.2)$. Find

(a) $P(X \le 7)$

(b) $P(Y \ge 3)$

3.9. Families of Discrete Distributions – The Geometric Distribution

The number of Bernoulli trials needed to get the first success has ***Geometric distribution***. Geometric random variables can take any integer value from 1 to infinity, because one needs at least 1 trial to have the first success, and the number of trials needed is not limited by any specific number. The only parameter is *p*, the probability of "success".

The probability mass function of a random variable *X* follows a Geometric distribution is given by

$$P(X = x) = pq^{x-1} \text{ for } x = 1, 2, \ldots$$

where $q = 1 - p$.

Mean and Variance

Suppose a discrete random variable *X* follows a Geometric distribution, $X \sim Geo(p)$. Then

$$\mu = \frac{1}{p} \text{ and } \sigma^2 = \frac{q}{p^2}.$$

An additional important property of the geometric distribution is that it is *memoryless*. It does not remember how many times you had tried before when you try the next time. In mathematical notation:

$$P(X = n + k \mid X > n) = P(X = k).$$

Example 3.14.

Some biology students were checking eye color in a large number of fruit flies. For the individual fly, suppose that the probability of white eyes is $\frac{1}{4}$ and the probability of red eyes is $\frac{3}{4}$, and that we may treat these observations as independent Bernoulli trials. Find the probability that the first fly with white eyes is the forth fly.

Solution:

3.10. Families of Discrete Distributions – The Negative Binomial Distribution

In a sequence of independent Bernoulli trials, the number of trials needed to obtain *k* successes has ***Negative Binomial distribution***. It has two parameters, *k*, the number of successes, and *p*, the probability of success.

The Negative Binomial probability mass function is

$$P(X = x) = P\left\{\text{the } x^{th} \text{ trial results in the } k^{th} \text{ success}\right\}$$

$$= P\left\{\begin{matrix} (k-1) \text{ successes in the first } (x-1) \text{ trials,} \\ \text{and the last trial is a success} \end{matrix}\right\}.$$

Hence,

$$P(X = x) = \binom{x-1}{k-1} p^k q^{x-k} \text{ for } x = k, k+1, \ldots$$

where $q = 1 - p$. This formula accounts for the probability of k successes, the remaining $(x-k)$ failures, and the number of outcomes – sequences with the k^{th} success coming on the x^{th} trial. Note that with $k = 1$, it becomes Geometric.

Mean and Variance

Suppose a discrete random variable *X* follows a Negative Binomial distribution, $X \sim NB(k, p)$. Then

$$\mu = \frac{k}{p} \text{ and } \sigma^2 = \frac{kq}{p^2}.$$

<u>*Example 3.15.*</u>

Suppose that during practice a basketball player can make a free throw 80% of the time. Furthermore, assume that a sequence of free-throw shooting can be thought of as independent Bernoulli trials. Let *X* equal the minimum number of free throws that this player must attempt to make a total of 10 shots. Then,

The *pmf* of *X* is

The mean, variance and standard deviation of *X* are, respectively,

And we have, for example,

Example 3.16.

In a recent production, 5% of certain electronic components are defective. We need to find 7 non-defective components for our 7 new computers. Components are tested until 7 non-defective ones are found. What is the probability that more than 10 components will have to be tested?

Solution:

3.11. Families of Discrete Distributions –The Poisson Distribution

The Poisson distribution is a discrete probability distribution that applies to occurrences of some events over ***a specified interval***. The random variable X is *the number of occurrences* of the events in an interval. The interval can be time, distance, area, or some similar unit.

The following are **examples** of discrete random variable for which the Poisson probability distribution can be applied.

1. The number of telemarketing phone calls received by a household during a given day.
2. The number of mistakes typed on a given page.
3. The number of customers entering a grocery store during a one-hour interval.

Poisson Probability Distribution

The probability distribution of X follows a Poisson distribution in an interval is

$$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!} \text{ for } x=0,1,2,\dots$$

where λ is the mean number of occurrences in that interval.

Mean and Variance

Suppose $X \sim \wp oi(\lambda)$, then $\mu=\lambda=\sigma^2$.

Example 3.17.

A washing machine in a laundry shop breaks down an average of three times per month. Find the probability that during the next month this machine will have

(a) exactly two breakdowns
(b) at most one breakdown
(c) at least one breakdown

Solution:

Note.
The intervals for λ and X must be equal. If they are not, the mean λ should be **redefined** to make them equal. The following example will show how to redefine λ and X to make them equal.

Example 3.18.
Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day. What is the probability that
(a) more than 8 new accounts will be initiated today?
(b) more than 16 accounts will be initiated within 2 days?
Solution:

Example 3.19.
On average, two new accounts are opened per day at Bank ABC.
(a) Find the probability that on a given day, the number of new accounts opened at this bank will be at most 3.
(b) Determine the average number of new accounts opened for the past 5 working days at Bank ABC.
Solution:

3.12. Families of Discrete Distributions – The Poisson approximation of Binomial Distribution

Poisson distribution can be effectively used to approximate Binomial probabilities when the number of trials *n* is large, and the probability of success *p* is small. Such an approximation is adequate, say, for $n \geq 30$ and $p \leq 0.05$, and it becomes more accurate for larger *n*.

Suppose $X \sim Bin(n, p)$, if $n \geq 30$ and $p \leq 0.05$, then $\lambda = np$, then

$$X \approx \wp oi\left(\lambda = np\right).$$

When *p* is large $\left(p \geq 0.95\right)$, the Poisson approximation is applicable too. The probability of failure $q = 1 - p$ is small in this case. Then, we can approximate the number of failures, which is also Binomial, by a Poisson distribution.

Example 3.20.
Ninety-seven percent of electronic messages are transmitted with no error. What is the probability that out of 200 messages, at least 195 will be transmitted correctly?
Solution: