***Recall:***

The deviation of an observation from its mean:

x –

How far the data point from the mean?

A function of the mean value of X

g(X) = (X - )2

We are not interesting in negative values, we interested in distance so we take square

Variance:

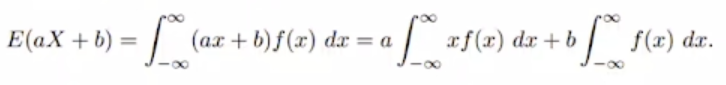
= Var(X) = E[g(X)] = E[(X – )2]

Standard deviation: Positive square root of the variance

***Theorem 4.5:***

* If a and b are constants, then (both for continuous and discrete cases)
  + E(aX + b) = aE(X) + b

Proof:

* By the definition of expected value,
* 
* The first integral on the right is E(X) and the second integral equals 1. Therefore we have,
* 

Corollary 4.1: Setting a = 0, we see that E(b) = b. Nothing changes so you can expect b as result.

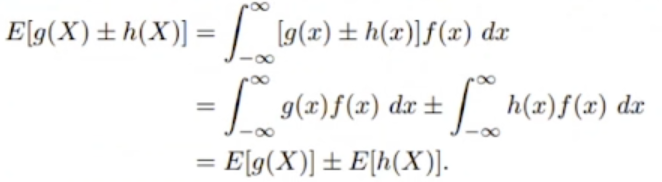
Corollary 4.2: Setting b = 0, we see that E(aX) = aE(X)

Corollary means conclusion out of the theorem.

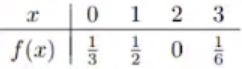
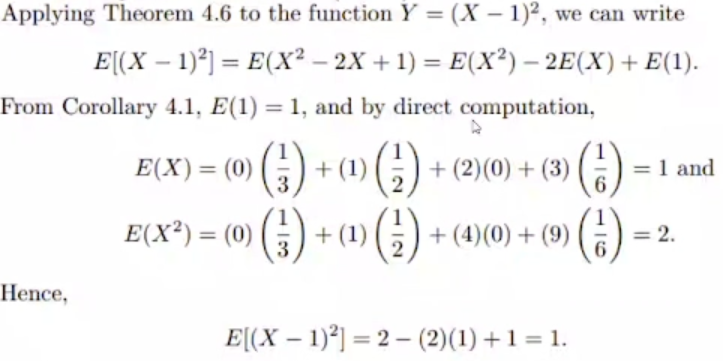
***Theorem 4.6:***

* The expected value of the sum or difference of 2 or more functions of a random variable X is the sum or difference of the expected values of the functions. That is, (both for continuous and discrete cases)
* 

Proof:

* By definition,
* 

Example:

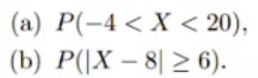
* Let X be a random variable with probability distribution as follows:
* 
* Find the expected value of Y = (X – 1)2
* 

***Theorem 4.10 (Chebyshev’s Theorem):***

* The probability that any random variable X will assume a value within k standard deviations of the mean is at least 1 – 1/k2.
* That is,
* Text

  Description automatically generated with medium confidence
  + is the mean. means move little bit left and move little bit right at same amount.
  + If k is small, you will get closer to mean.
  + If you go away from the mean (big k), probability of X value being in that area is getting lower.
* For k = 2, theorem states that the random variable X has a probability of at least 1 – 1/22 = 3/4 of falling within 2 standard deviations of the mean.
  + Probability of X been in the interval is 3/4 at least.
* That is, three-fourths or more of the observations of any distribution lie in the interval
* Similarly, the theorem says that at least eight-ninths of the observations of any distribution fall in the interval

Example:

* A random variable X has a mean = 8, a variance = 9, and an unknown probability distribution.
* Find:
  + 
* Text, letter

  Description automatically generated

Now we will see some special distributions both for discrete and continuous cases.

Any data set doesn’t have to have the proper or known distribution. Sometimes distribution function may be so weird and unknown.

Sometimes people try to fit set of data into certain distribution.

People sometimes manipulate data to fit into special distributions.

**Discrete** Probability Distributions: BERNOULLI

An experiment often consists of **repeated trials**, each with **two possible outcomes** (for example, tossing a coin) that may be labeled success or failure.

The most obvious application deals with the testing of items as they come off an assembly line, where each trial may indicate a defective or a nondefective item.

We may choose to define either outcome as a success.

The process is referred to as a Bernoulli process. Each trial is called a Bernoulli trial.

Bernoulli Process

Strictly speaking, the Bernoulli process must possess the following properties:

* The experiment consists of **repeated trials**.
* Each trial results in an outcome that may be classified as a **success or a failure**.
* The probability of success, denoted by p, remains constant from trial to trial.
  + The probability of failure is denoted by q and it is 1-p.
  + q and p are independent. You cannot get both of them at the same time.
* The repeated trials are independent.



Consider the set of Bernoulli trials where 3 items are selected at random from a manufacturing process, inspected, and classified as defective or nondefective.

A defective item is designated a success.

The # of successes is a random variable X assuming integer values from 0 through 3.

The 8 possible outcomes and the corresponding values of X are



There are 3 trials. In each trial, we count # of defectives. Since the items are selected independently and we assume that the process produces 25% defectives, we have

Text

Description automatically generated with medium confidenceWe have independent events so we can multiply probabilities separately.

Similar calculations yield the probabilities for the other possible outcomes. The probability distribution of X is therefore

A picture containing text, clock

Description automatically generated

The number X of successes in n Bernoulli trials is called a binomial random variable.

The probability distribution of this discrete random variable is called the binomial distribution, and its values will be denoted by b(x; n, p) since they depend on the # of trials and the probability of a success on a given trial.

* x: variable
* n: # of trials
* p: probability of the success

Thus, for the probability distribution of X, the # of defectives is:

A picture containing diagram

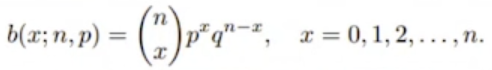
Description automatically generated

We were considering getting 2 defective ones.

Binomial Distribution

A Bernoulli trial can result in a success with probability p and a failure with probability q = 1-p.

Then the probability distribution of the binomial random variable X, the # of successes in n independent trials, is (n is # of trials)

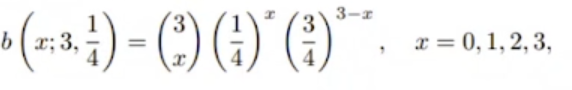


🡪 There are n independent trials. If we are trying to find # of success, max # of success is   
 the max # of trials because we tried n times and you get success n times. All in all, this  
 means n trials and x success.

px 🡪 x is the # of success. You multiply p for all your successes.

qn-x 🡪 n-x is the # of failure. You multiply q for all your failures.

Note that when n = 3 and p = 1/4 , the probability distribution of X, the # of defectives, may be written as



CHECK PAGE 747 OF THE BOOK FOR THE TABLE

If n is 1, you get either 1 success or 0 success. r is the number of success.

Example:

* The probability that a certain kind of component will survive a shock test is 3/4.
* Find the probability that exactly 2 of the next 4 components tested survive.
* A picture containing text

  Description automatically generated
* You can also check the table in the book. n is 4 and r is 2. Probability of success is 3/4 which is 0.75. It is not listed but you can say that value is between 0.3483 and 0.1808

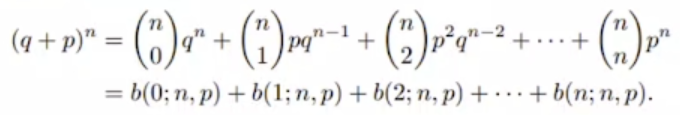
Where Does the Name Binomial Come From?

Name binomial comes from binomial expension. When you try to find some certain power of a sum of 2 different values, we could compute the coefficient of each term using combination. Each term is corresponding to binomial distribution function.

A picture containing graphical user interface

Description automatically generated

The binomial distribution derives its name from the fact that the n+1 terms in the binomial expansion of (q+p)n correspond to the various values of b(x; n, p) for x = 0, 1, 2, …, n. That is,



Since p + q = 1, we see that

A picture containing text, watch

Description automatically generated

a condition that must hold for any probability distribution.

Frequently, we are interested in problems where it is necessary to find P(X < r) or P(a <= X <= b). If something like P(X >= r) comes, then we can use 1-P(X<r) which is same.

Binomial sums

A picture containing text, watch, gauge

Description automatically generated

are given in table in book for n = 1, 2, …, 20 for selected values of p from 0.1 to 0.9. We illustrate the use of table with the following example.

Example 5.2:

* The probability that a patient recovers from a rare blood disease is 0.4.
* If 15 people are known to have contracted this disease, what is the probability that
  + (a) at least 10 survive
  + (b) from 3 to 8 survive
  + (c) exactly 5 survive
* Text, letter

  Description automatically generated
* Since this is discrete case, C is valid, you can get value at 1 certain point. You can also do C with following formula:
  + A close-up of a calculator

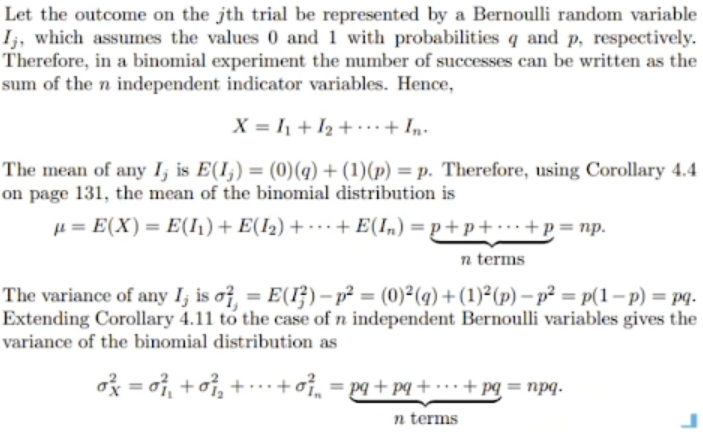
    Description automatically generated with low confidence
  + x=5, n=15, p=0.4, q=0.6

The Mean and Variance

Theorem 5.1:

* The mean and variance of the binomial distribution b(x; n, p) are
  + = np and = npq

Proof:

* 

Example 5.4:

* It is conjectured that an impurity exists in 30% of all drinking wells in a certain rural community.
* In order to gain some insight into the true extent of the problem, it is determined that some testing is necessary.
* It is too expensive to test all of the wells in the area, so 10 are randomly selected for testing.
  + (a) Using the binomial distribution, what is the probability that exactly 3 wells have the impurity, assuming that the conjecture is correct?
  + (b) What is the probability that more than 3 wells are impure?
* Text

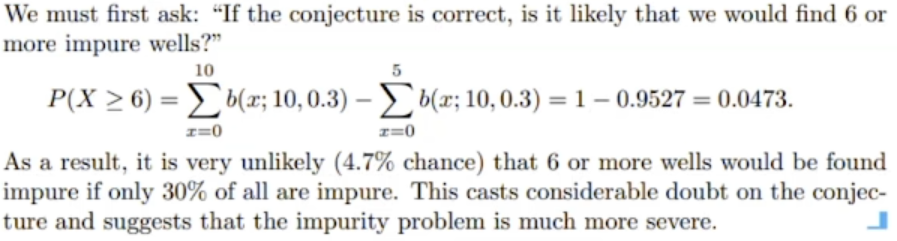
  Description automatically generated
* For b 🡪 P(X>3) = 1 – P(X<=3)

Example:

* Find the mean and the variance of the binomial random variable of example 5.2, and then use Chebyshev’s theorem (on page 137) to interpret the interval
* Text

  Description automatically generated

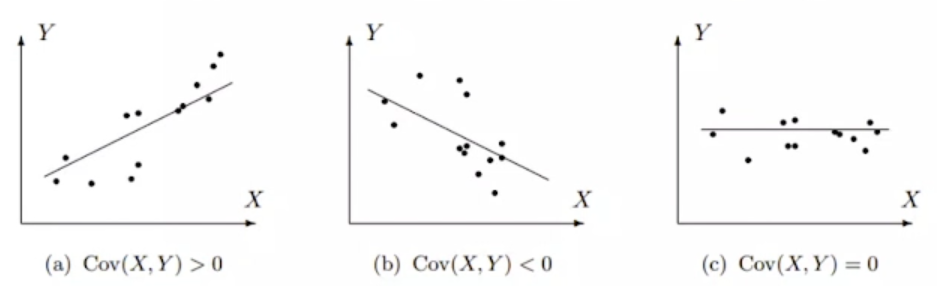
Example:

* Consider the situation of example 5.4. The notion that 30% of the wells are impure is merely a conjecture put forth by the area water board.
* Suppose 10 wells are randomly selected and 6 are found to contain the impurity.
* What does this imply about the conjecture? Use a probability statement.
* 

Covariance and Correlation

Expectation, variance, and standard deviation characterize the distribution of a single random variable.

Now we introduce measures of association of 2 random variables.



*FIGURE 3.5: Positive, negative, and zero covariance*

Change of one random variable affects other one. Sometimes they affect each other, sometimes don’t.

***Definition***

Covariance = Cov(X, Y) is defined as

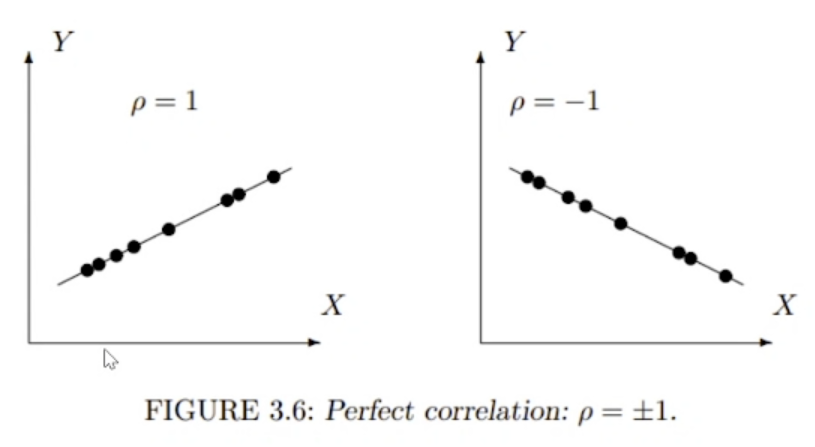
* Cov(X, Y) = E{(X – EX)(Y – EY)} = E(XY) – E(X)E(Y)

It summarizes interrelation of 2 random variables.

Correlation coefficient between variables X and Y is defined as

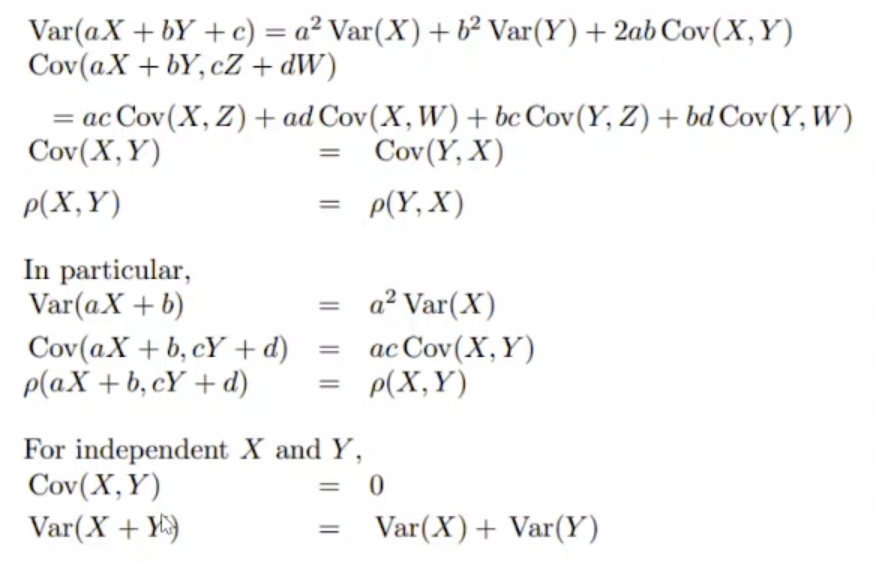
StdX: Standard deviation of X  
StdY: Standard deviation of Y

tells us the rate of change, how steep is the line.



This is really suspicious. If you collect data from real life, this never happens. In real data, there are always some error.

Properties of Variances and Covariances



***Notation***

Text

Description automatically generated with medium confidence

Bernoulli Distribution

If P(1) = p is the probability of a success, then P(0) = q = 1 - p is the probability of a failure.

We can then compute the expectation and variance as

Text, letter

Description automatically generated

P(x) is distribution function. Same as f(x). They wanted to show as P(x).

Text

Description automatically generated with medium confidence

Bernoulli distribution is special case. When it is repeated, it becomes binomial distribution.

Binomial Distribution

Number of trials is many times. Trial is same thing, it is still Bernoulli trial. Here it is repeated in a sequence.

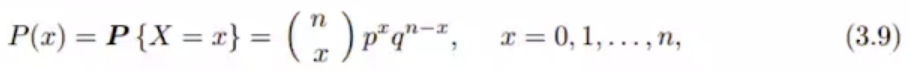
Bernoulli trial: If you are flipping a coin, label one of head and tail as success and other one failure.  
Binomial: If you flip the coin 10 times.

When you repeat the Bernoulli, you still have p (probability of success) and q (probability of failure).

Definition:

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

Binomial probability mass function is (n trial, x success)



Text

Description automatically generated with low confidence

At most x can be n, number of trials.

Example:

* An exciting computer game is released.
* 60% of players complete all the levels.
* Among 30% of completed players will then buy an advanced version of the game.
* Among 15 users, what is the expected # of people who will buy the advanced version?
  + ***You either buy or don’t buy so 2 choices. Independent cases.***
  + ***Discrete because we are counting people.***
* What is the probability that at least 2 people will buy it?
* Text

  Description automatically generated

Geometric Distribution

Again, consider a sequence of independent Bernoulli trials. Each trial results in a “success” or a “failure”.

Definition 3.12:

* The # of Bernouilli trials needed to get the 1st success has Geometric distribution
* We continue until we get the 1st success, then we stop

Example:

* A search engine goes through a list of sites looking for a given key phrase. Suppose the search terminates as soon as the key phrase is found. The # of sites visited is Geometric.

Example:

* A hiring manager interviews candidates, one by one, to fill a vacancy. The # of candidates interviewed until one candidate receives an offer has Geometric distribution.

Geometric probability mass function (pmf) has the form:

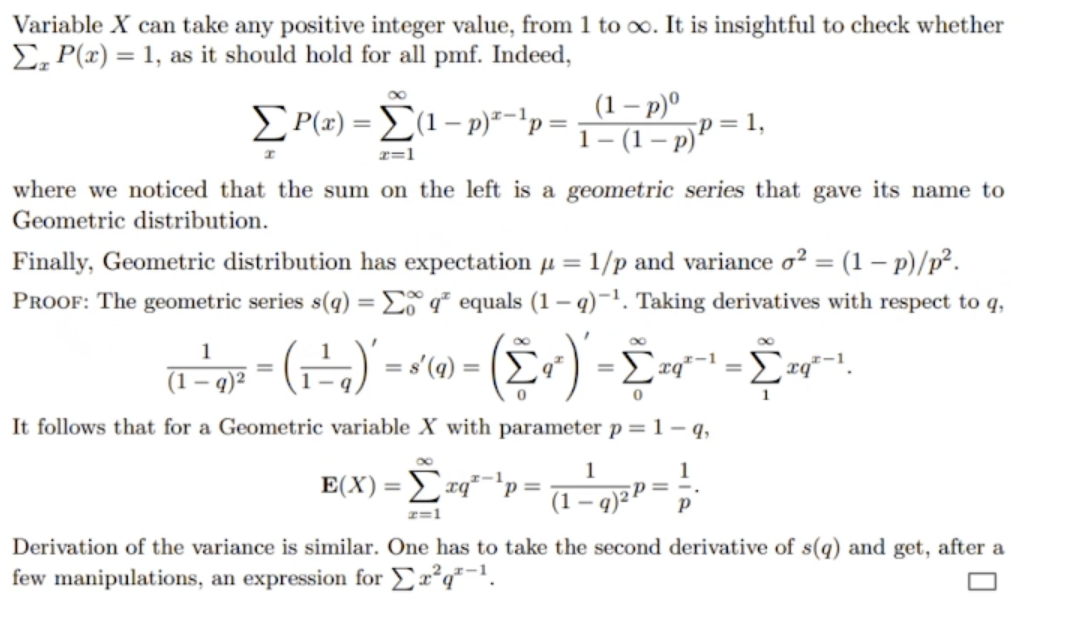


which is the probability of (x-1) failures followed by one success.

Comparing with (3.9), there is no number of combinations in this formula because only one outcome has the first success coming on the x-th trial.

If you try x times, you will fail x-1 times bc at x-th trial you success and finish. Probability of failure is 1-p.

We succeed 1 times so it is p1.



In geometric series, equation holds only if that p is a number between 0 and 1. That is the case in probability.

For example, tossing a coin. p is 1/2. Expected value of success is 1/(1/2) which is 2. So you can expect to get success in second flip.

Text

Description automatically generated

This little summaries will be available in exams.

BERNOULLI DISTRIBUTION

BINOMIAL DISTRIBUTION 🡪 repeated version of Bernoulli

GEOMETRIC DISTRIBUTION 🡪 repeat the process until you get the success