Chapter 3 Discrete Random Variables and Their Probability Distributions

* 3.1 Basic Definition

*A* ***random variable*** *is a real-valued function defined over a sample space.*

*A random variable Y is said to be* ***discrete*** *if it can assume only a finite or countably infinite number of distinct values.*

A set of elements is countably infinite if the elements in the set can be put into one-to-one correspondence with the positive integers.

* 3.2 The Probability Distribution for a Discrete Random Variable

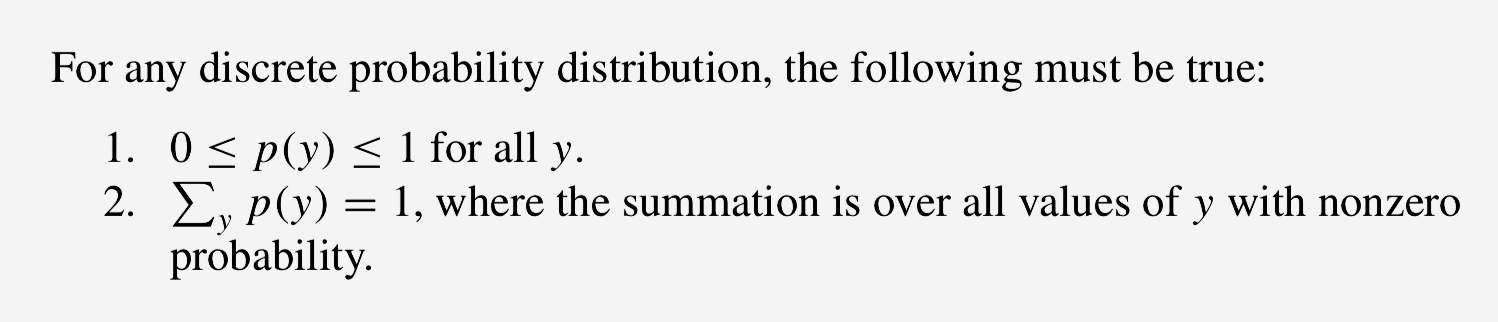
We use uppercase letter, such as Y , to denote a random variable and a lowercase letter, such as y, to denote a particular value that a random variable may assume.

***The probability that Y takes on the value y, P(Y = y),*** *is defined as the sum of the probabilities of all sample points in S that are assigned the value y. We will sometimes denote P(Y = y) by p(y).*

Because p(y) is a function that assigns probabilities to each value y of the random variable Y, it is sometimes called the probability function for Y.

***The probability distribution for a discrete variable Y*** *can be represented by a formula, a table, or a graph that provides p(y) = P(Y = y) for all y.*

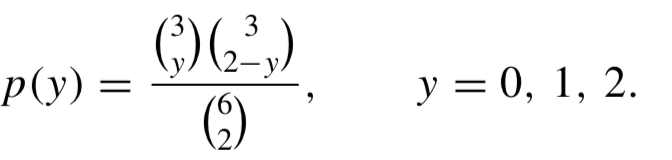
<Theorem>



Ex. 3 men and 3 women. Select two workers at random. Let Y denote the number of women in his selection. Find the probability distribution for Y .







function:

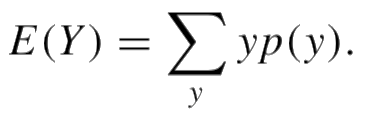
graph:





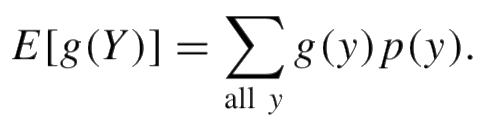
* 3.3 The Expected Value of a Random Variable or Function of a Random Variable

*Let Y be a discrete random variable with the probability function p(y). Then the* ***expected value of Y, E(Y),*** *if defined to be*



If p(y) is an accurate characterization of the population frequency distribution, then E (Y ) = μ, the **population mean**.

*Let Y be a discrete random variable with probability function p(y) and g(Y) be a real-valued function of Y . Then the expected value of g(Y ) is given by*

**

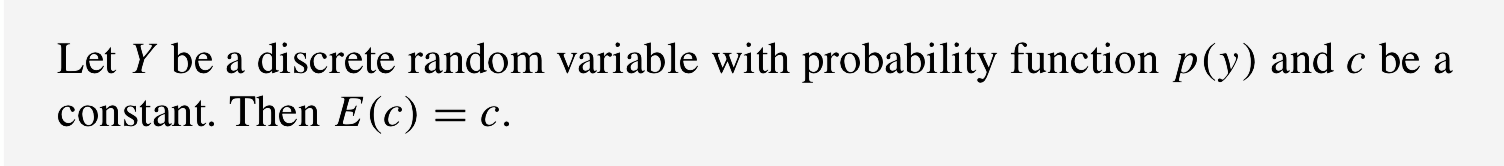
*If Y is a random variable with mean E(Y) = μ, the* ***variance of a random variable Y*** *is defined to be the expected value of (Y − μ)2. That is,*

*V (Y ) = E [(Y − μ)2 ].*

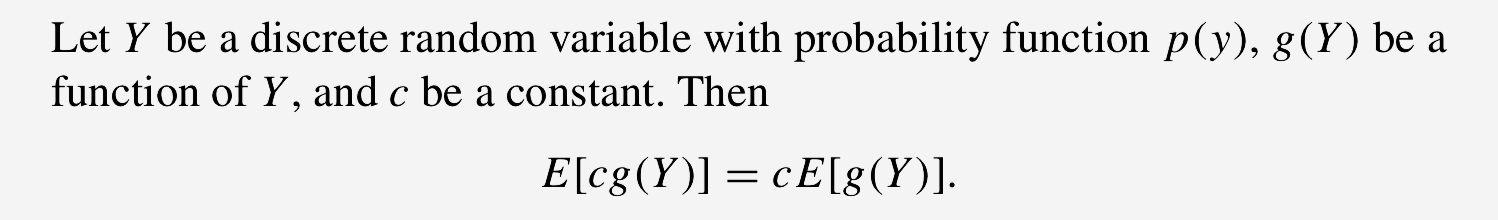
*The* ***standard deviation*** *of Y is the positive square root of V (Y ).*

If p(y) is an accurate characterization of the population frequency distribution (and to simplify notation, we will assume this to be true), then E(Y) = μ, V(Y) = σ2, the population variance, and σ is the population standard deviation.

<Theorem 1>



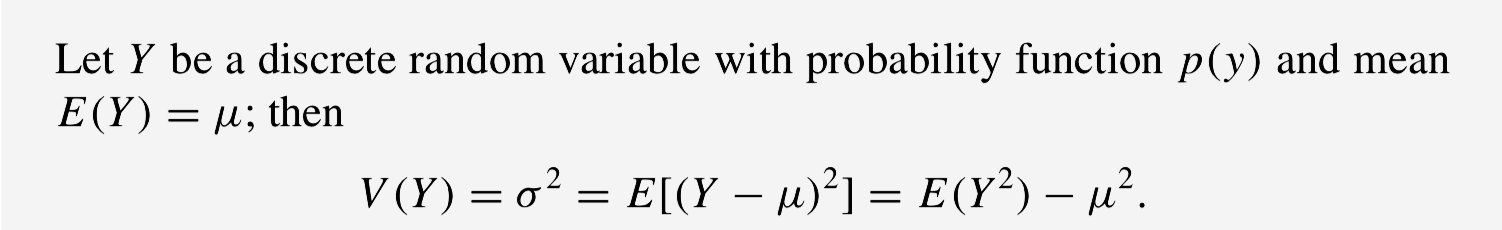
<Theorem 2>



<Theorem 3>



<Theorem 4>



* 3.4 The Binomial Probability Distribution

*A* ***binomial experiment*** *possesses the following properties:*

*1. The experiment consists of a fixed number, n, of identical trials.*

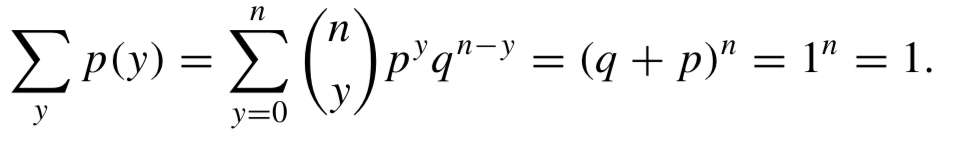
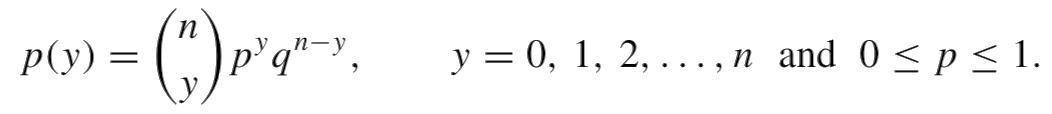
*2. Each trial results in one of two outcomes: success, S, or failure, F.*

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

*4. The trials are independent.*

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

*A random variable Y is said to have a* ***binomial distribution*** *based on n trials with success probability p if and only if*

**

Note that

<Theorem>

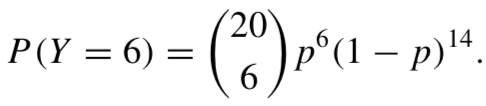


1. 

2. E (Y 2 ) = E [Y (Y − 1)] + μ = n(n − 1)p2 + np

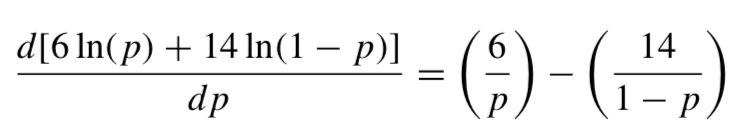
Ex. Suppose that we survey 20 individuals working for a large company and ask each whether they favor implementation of a new policy regarding retirement funding. If, in our sample, 6 favored the new policy, find an estimate for p, the true but unknown proportion of employees that favor the new policy.

If Y denotes the number among the 20 who favor the new policy, it is reasonable to conclude that Y has a binomial distribution with n = 20 for some value of p. Whatever the true value for p, we conclude that the probability of observing 6 out of 20 in favor of the policy is



We will use as our estimate for p the value that maximizes the probability of observing the value that we actually observed (6 in favor in 20 trials). How do we find the value of p that maximizes P(Y = 6)?

Because ( ) is a constant (relative to p) and ln(w) is an increasing function of w, the value of p that maximizes P(Y=6)= ( )p(1−p) is the same as the value of p that maximizes ln[p6(1− p)14]=[6ln(p)+14ln(1− p)].

If we take the derivative of [6 ln( p) + 14 ln(1 − p)] with respect to p, we obtain 

Setting this derivative equal to 0 and solving, we obtain p = 6/20. Because the second derivative of [6 ln( p) + 14 ln(1 − p)] is negative when p = 6/20, it follows that [6ln(p) + 14ln(1 − p)] [and P(Y = 6)] is *maximized* when p = 6/20. Our estimate for p, based on 6 “successes” in 20 trials is therefore 6/20.

* 3.5 The Geometric Probability Distribution

*A random variable Y is said to have a* ***geometric probability distribution*** *if and only if*

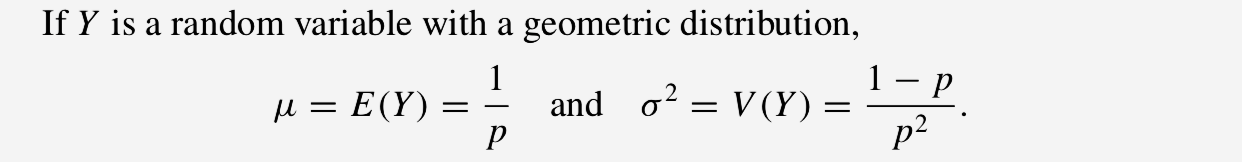
*p(y)=qy−1 p, y=1,2,3,..., 0≤p≤1.*

The geometric random variable Y is the number of the trial on which the first success occurs.

Note that some authors define random variable Y\* =the number of failures before the first success.

ie. Y\*=Y-1

<Theorem>



Ex. Suppose that we interview successive individuals working for the large company discussed in Example 3.10 and stop interviewing when we find the first person who likes the policy. If the fifth person interviewed is the first one who favors the new policy, find an estimate for p, the true but unknown proportion of employees who favor the new policy.

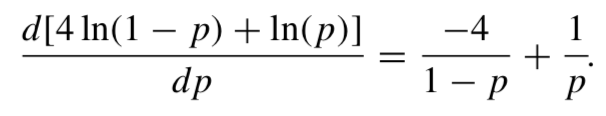
If Y denotes the number of individuals interviewed until we find the first person who likes the new retirement plan, it is reasonable to conclude that Y has a geometric distribution for some value of p. Whatever the true value for p, we conclude that the probability of observing the first person in favor of the policy on the fifth trial is

P(Y = 5) = (1 − p)4 p.

We will use as our estimate for p the value that maximizes the probability of observing the value that we *actually observed* (the first success on trial 5).

To find the value of p that maximizes P(Y = 5), we again observe that the value of p that maximizes P(Y = 5) = (1− p)4p is the same as the value of p that maximizes ln[(1 − p)4 p] = [4 ln(1 − p) + ln( p)].

If we take the derivative of [4 ln(1 − p) + ln( p)] with respect to p, we obtain



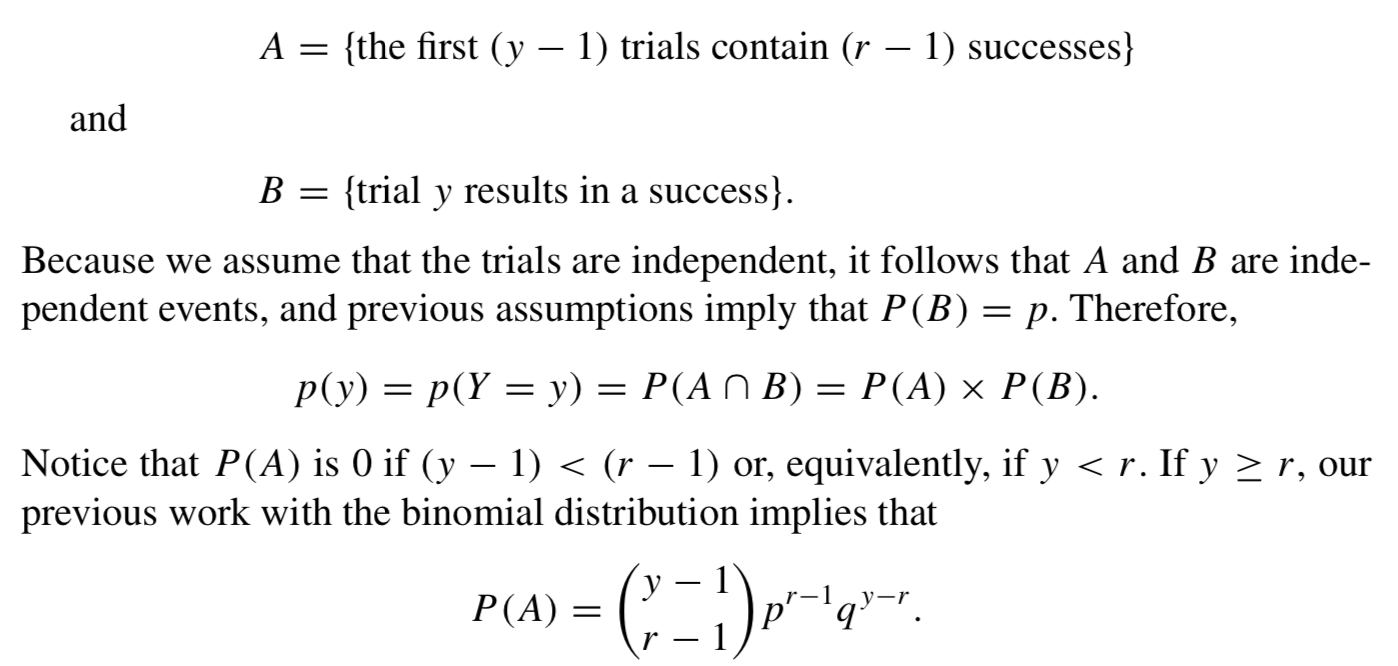
Setting this derivative equal to 0 and solving, we obtain p = 1/5.

Because the second derivative of [4 ln(1 − p) + ln( p)] is negative when p = 1/5, it follows that [4ln(1 − p) + ln(p)] [and P(Y = 5)] is maximized when p = 1/5.

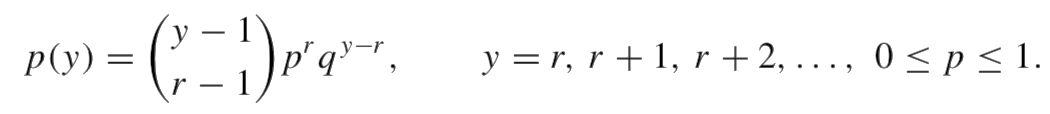
Our estimate for p, based on observing the first success on the fifth trial is 1/5.

* 3.6 The Negative Binomial Probability Distribution

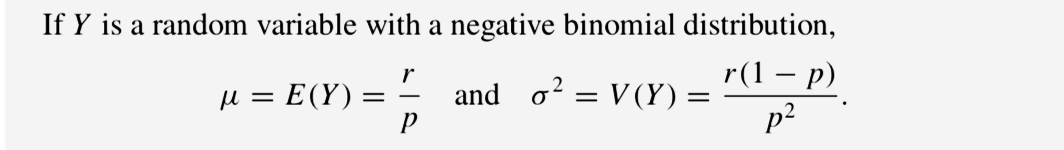
Random variable Y equal to the number of the trial on which the rth success occurs (r = 2, 3, 4, etc.) is the negative binomial distribution.



*A random variable Y is said to have* ***a negative binomial probability distribution*** *if and only if*

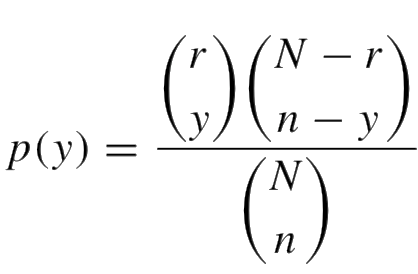


<Theorem>



* 3.7 The Hypergeometric Probability Distribution

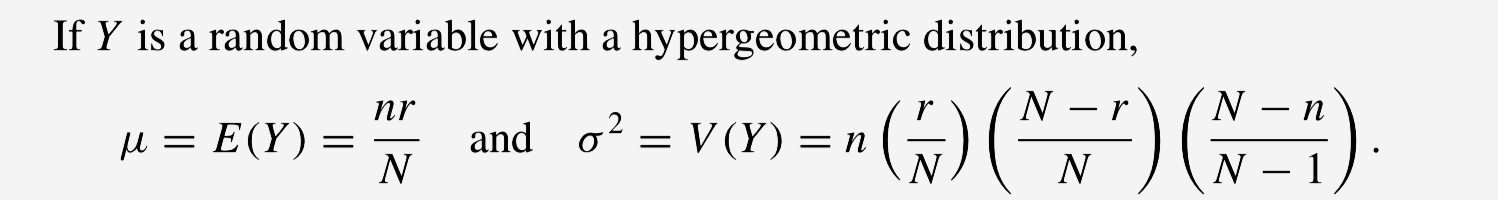
*A random variable Y is said to have a* ***hypergeometric probability distribution*** *if and only if*

**

*where y is an integer 0, 1, 2,...,n, subject to the restrictions y ≤ r and n − y ≤ N − r.*

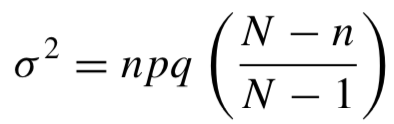
Suppose that a population of size N consists of r units with the attribute and N − r without. If a sample of size n it taken, without replacement, and Y is the number of items with the attribute in the sample

<Theorem>

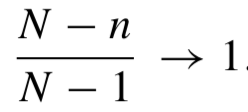


If we define p= r/N and q=1−p= (N−r)/N, we can re-express the mean and variance of the hypergeometric as

μ = np



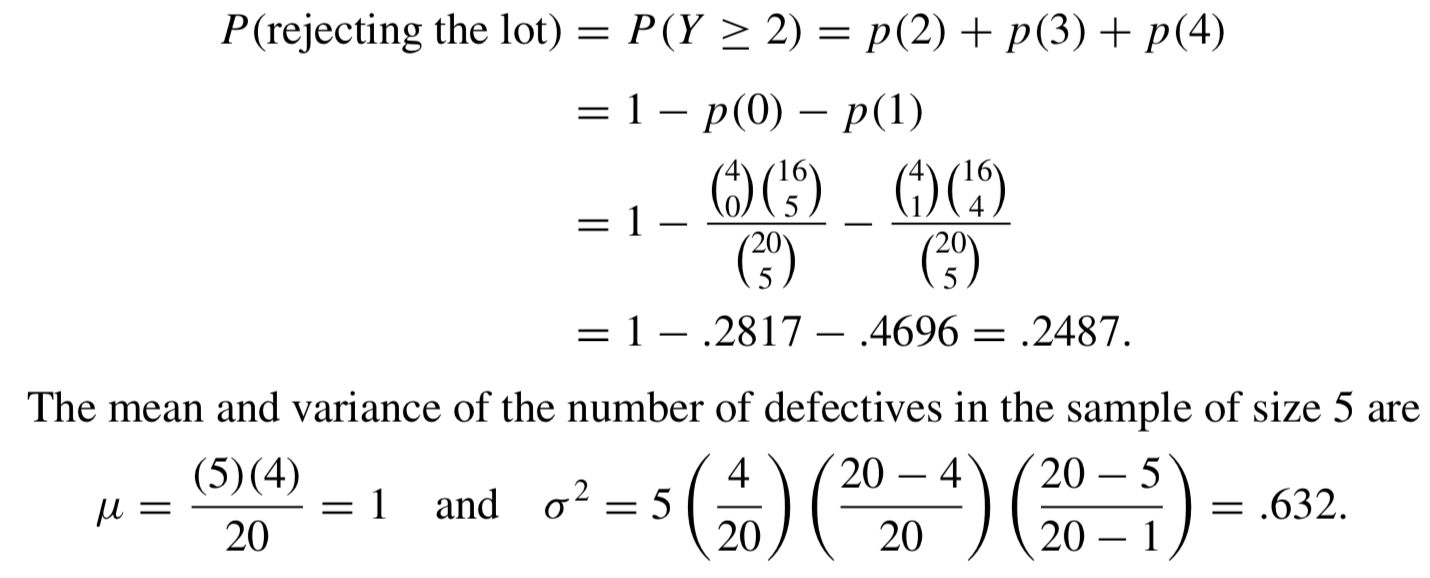
You can view the factor (N-n)/(N-1) in V (Y ) as an adjustment that is appropriate when n is large relative to N . For fixed n, as N → ∞,



Ex. An industrial product is shipped in lots of 20. Testing to determine whether an item is defective is costly, and hence the manufacturer samples his production rather than using a 100% inspection plan. A sampling plan, constructed to minimize the number of defectives shipped to customers, calls for sampling five items from each lot and rejecting the lot if more than one defective is observed. (If the lot is rejected, each item in it is later tested.) If a lot contains four defectives, what is the probability that it will be rejected? What is the expected number of defectives in the sample of size 5? What is the variance of the number of defectives in the sample of size 5?

Let Y equal the number of defectives in the sample. Then N = 20, r = 4, and n = 5.

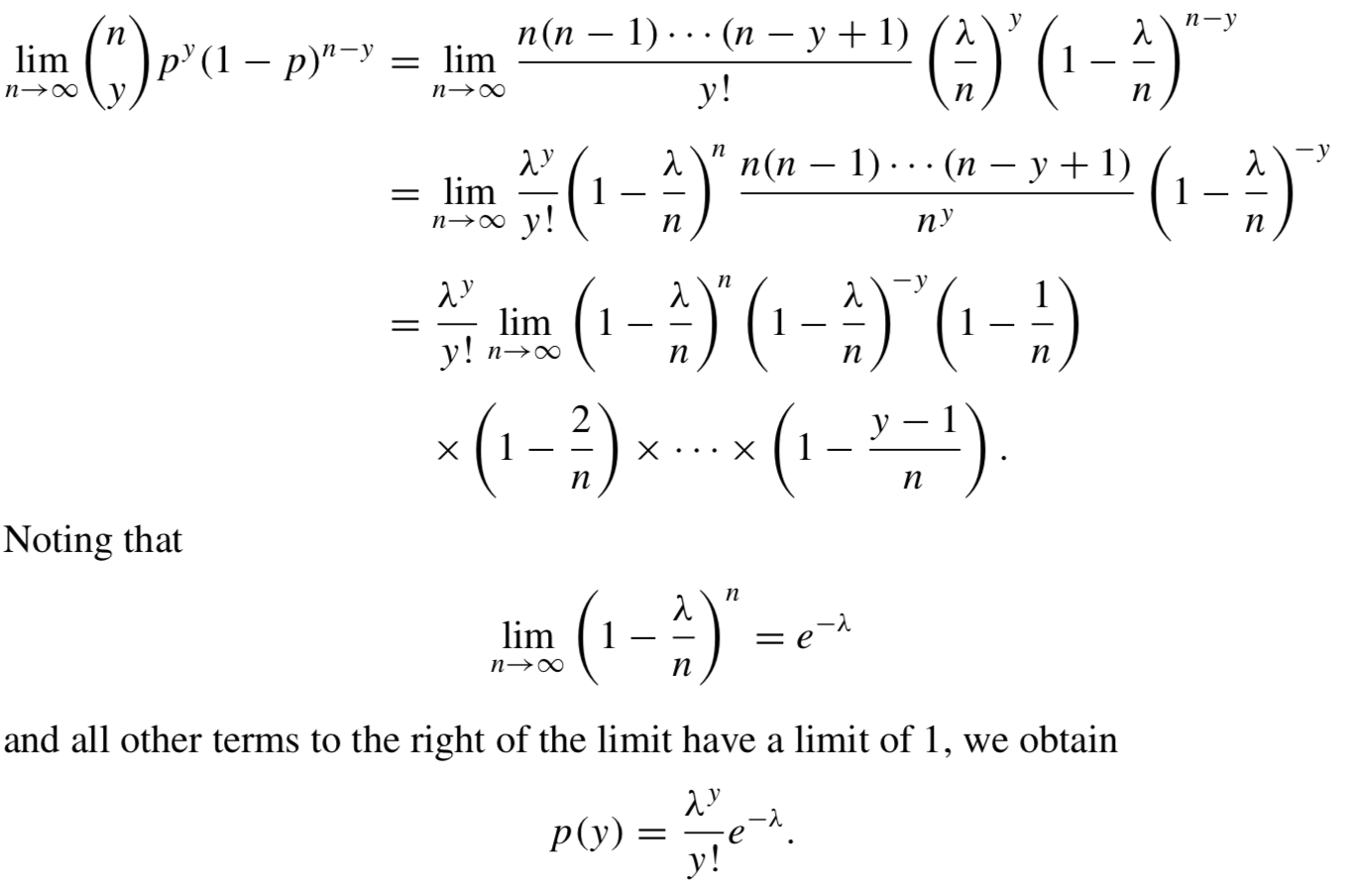
The lot will be rejected if Y = 2, 3, or 4. Then



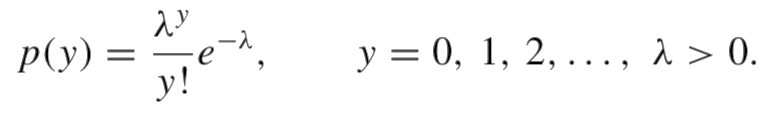
For a fixed fraction defective p = r/N, the hypergeometric probability function converges to the binomial probability function as N becomes large.

* 3.8 The Poisson Probability Distribution

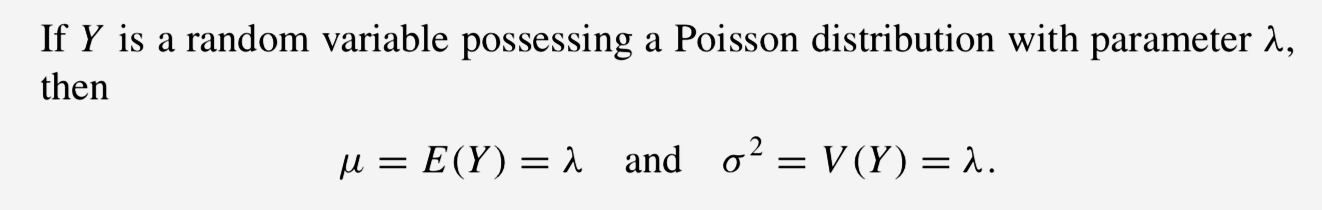
The Poisson probabilities can be used to approximate their binomial counterparts for large n, small p, and λ = np less than, roughly, 7.



*A random variable Y is said to have a* ***Poisson probability distribution*** *if and only if*



<Theorem>



If we observe a **Poisson process** and λ is the mean number of occurrences per unit (length, area, etc.), then Y = the number of occurrences in a units has a Poisson distribution with mean aλ.

A key assumption in the development of the theory of Poisson process is independence of the numbers of occurrences in disjoint intervals (areas, etc.).

Ex. Industrial accidents occur according to a Poisson process with an average of three accidents per month. During the last two months, ten accidents occurred. Does this number seem highly improbable if the mean number of accidents per month, μ, is still equal to 3? Does it indicate an increase in the mean number of accidents per month?

The number of accidents in two months, Y , has a Poisson probability distribution with mean λ⋆ = 2(3) = 6.



The empirical rule tells us that we should expect Y to take values in the interval μ ± 2σ with a high probability.

Notice that μ + 2σ = 6 + (2)(2.45) = 10.90. The observed number of acci- dents, Y = 10, does not lie more than 2σ from μ, but it is close to the boundary. Thus, the observed result is not highly improbable, but it may be sufficiently improbable to warrant an investigation.

* 3.9 Moments and Moment-Generating functions

*The kth* ***moment of a random variable Y taken about the origin*** *is E(Yk) and denoted by μ’k.*

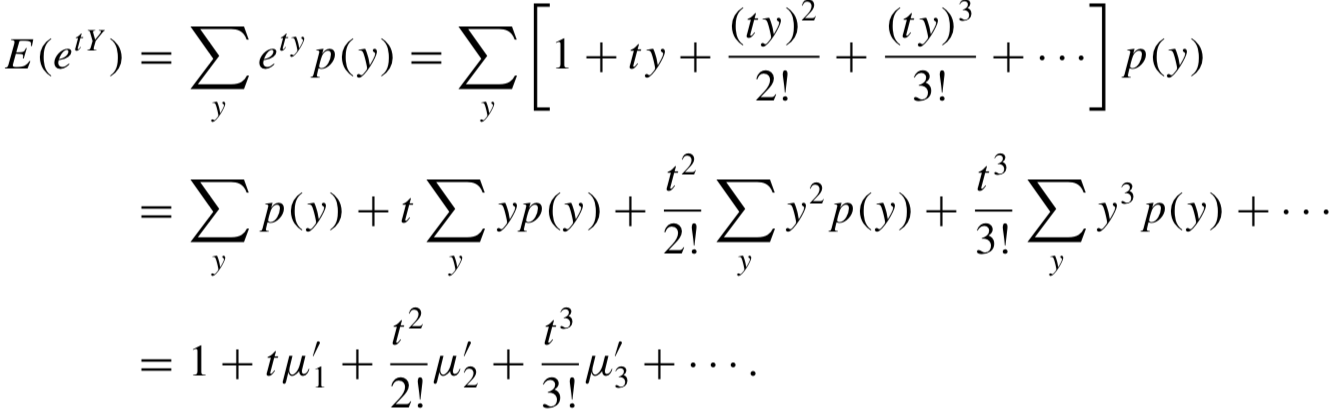
*Notice in particular that the first moment about the origin, is E (Y ) = μ’1 = μ and that μ’2 = E(Y2) is employed to find σ2*

*The* ***kth moment of a random variable Y taken about its mean****, or the* ***kth central moment of Y****, is defined to be E[(Y − μ)k] and is denoted by μk.*

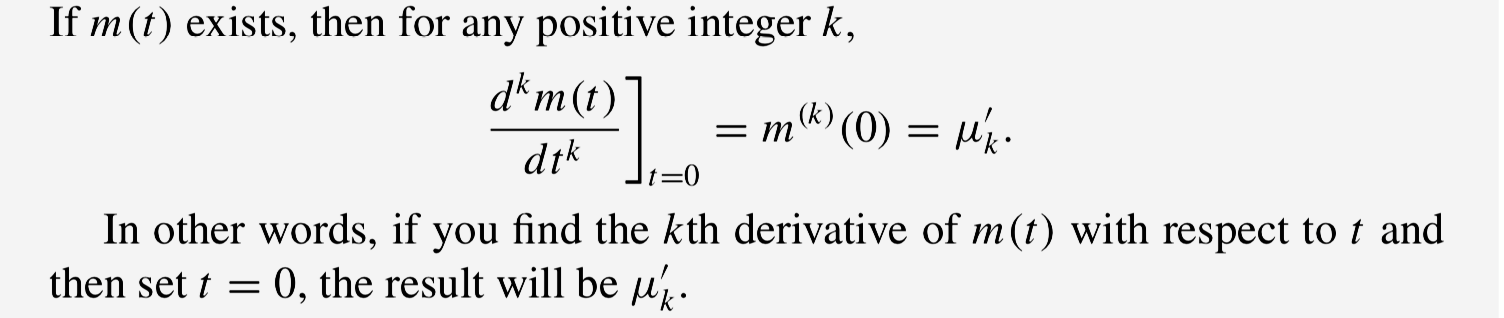
In particular, σ2 = μ2.

*The* ***moment-generating function m(t) for a random variable Y*** *is defined to be m(t) = E(etY ).*

*We say that a moment-generating function for Y exists if there exists a positive constant b such that m(t) is finite for |t| ≤ b*



<Theorem>



Ex1. Find the moment-generating function m (t ) for a Poisson distributed random variable with mean λ.

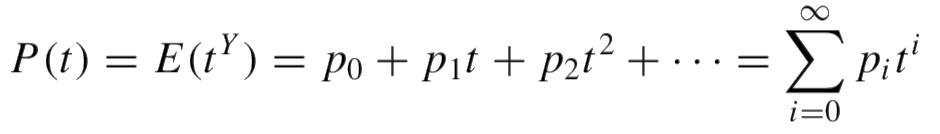
Ex2. Use the moment-generating function of Ex1and Theorem 3.12 to find the mean, μ, and variance, σ2.

Ex3. Suppose that Y is a random variable with moment-generating function m Y (t ) = e3.2(et −1). What is the distribution of Y ?

In Ex1, we showed that the moment-generating function of a Poisson distributed random variable with mean λ is m(t) = eλ(et−1). Because moment-generating functions are unique, Y must have a Poisson distribution with mean 3.2.

* Probability-Generating Functions

*Let Y be* ***an integer-valued random variable*** *for which P(Y = i) = pi, where i = 0, 1, 2, . . . .* ***The probability-generating function P (t ) for Y*** *is defined to be*

**

*for all values of t such that P(t) is finite.*

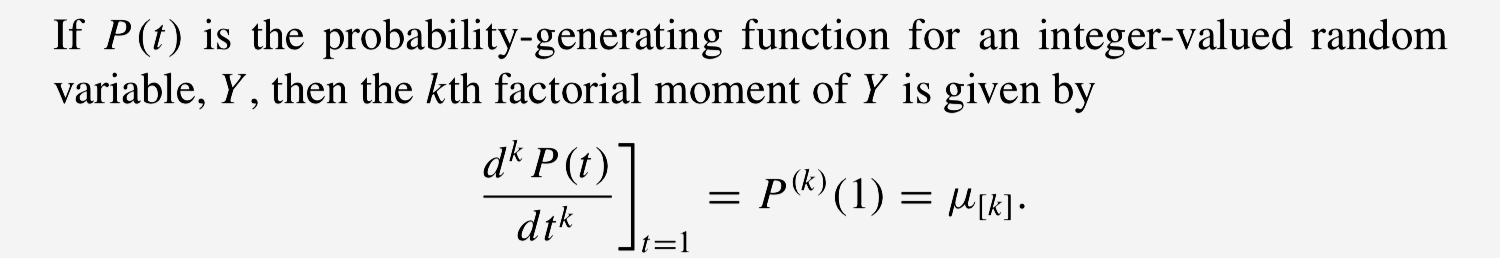
*The* ***kth factorial moment*** *for a random variable Y is defined to be*

**

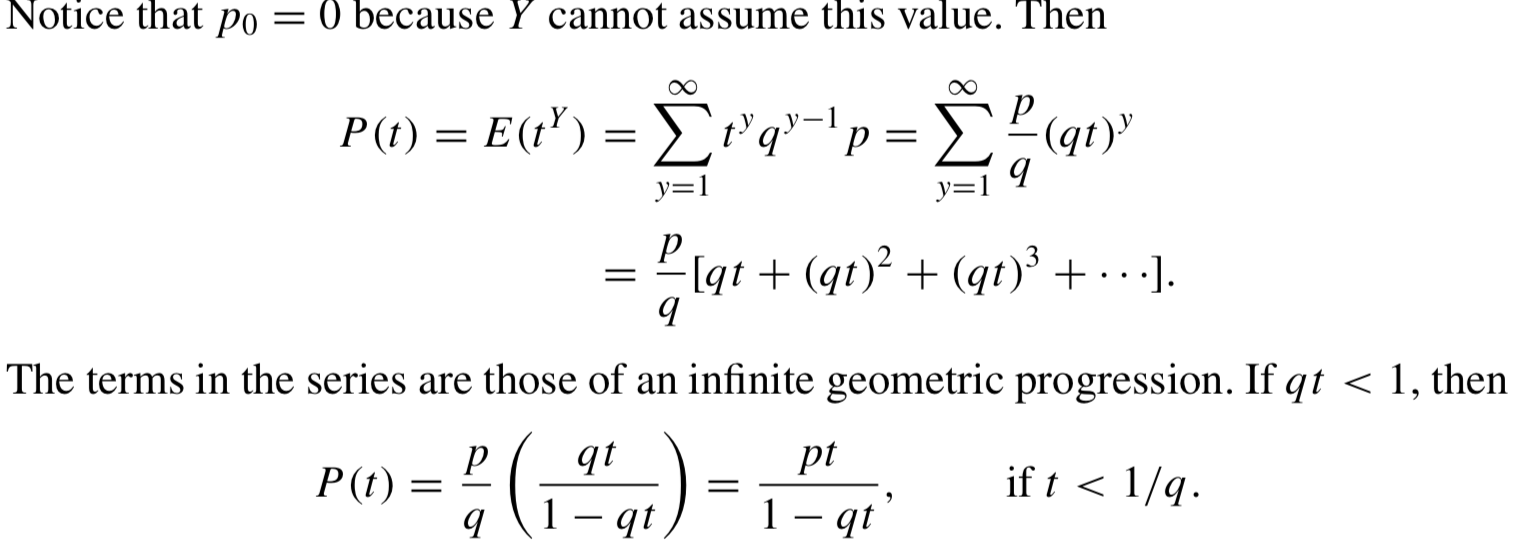
*where k is a positive integer.*

Notice that μ[1] = E (Y ) = μ. The second factorial moment, μ[2] = E [Y (Y − 1)], was useful in finding the variance for binomial, geometric, and Poisson random variables.

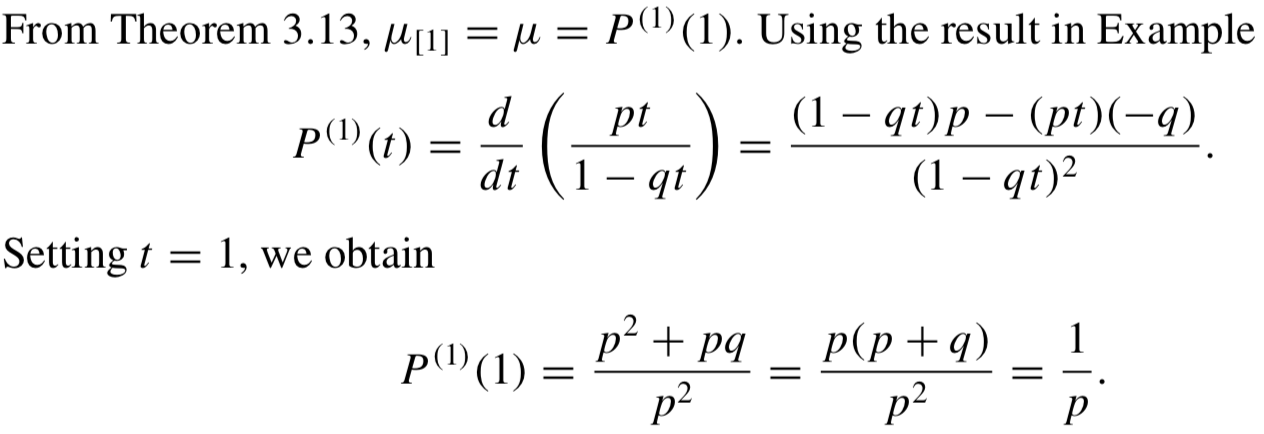
<Theorem>



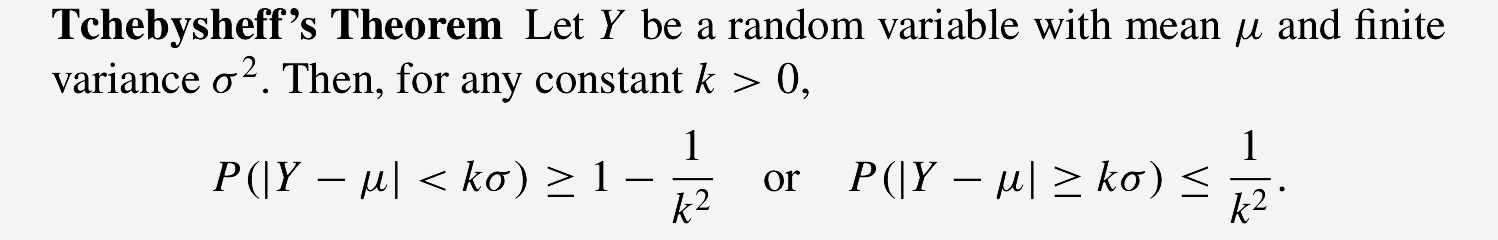
Ex1. Find the probability-generating function for a geometric random variable.



Ex2. Use P(t) in Ex1, to find the mean of a geometric random variable.



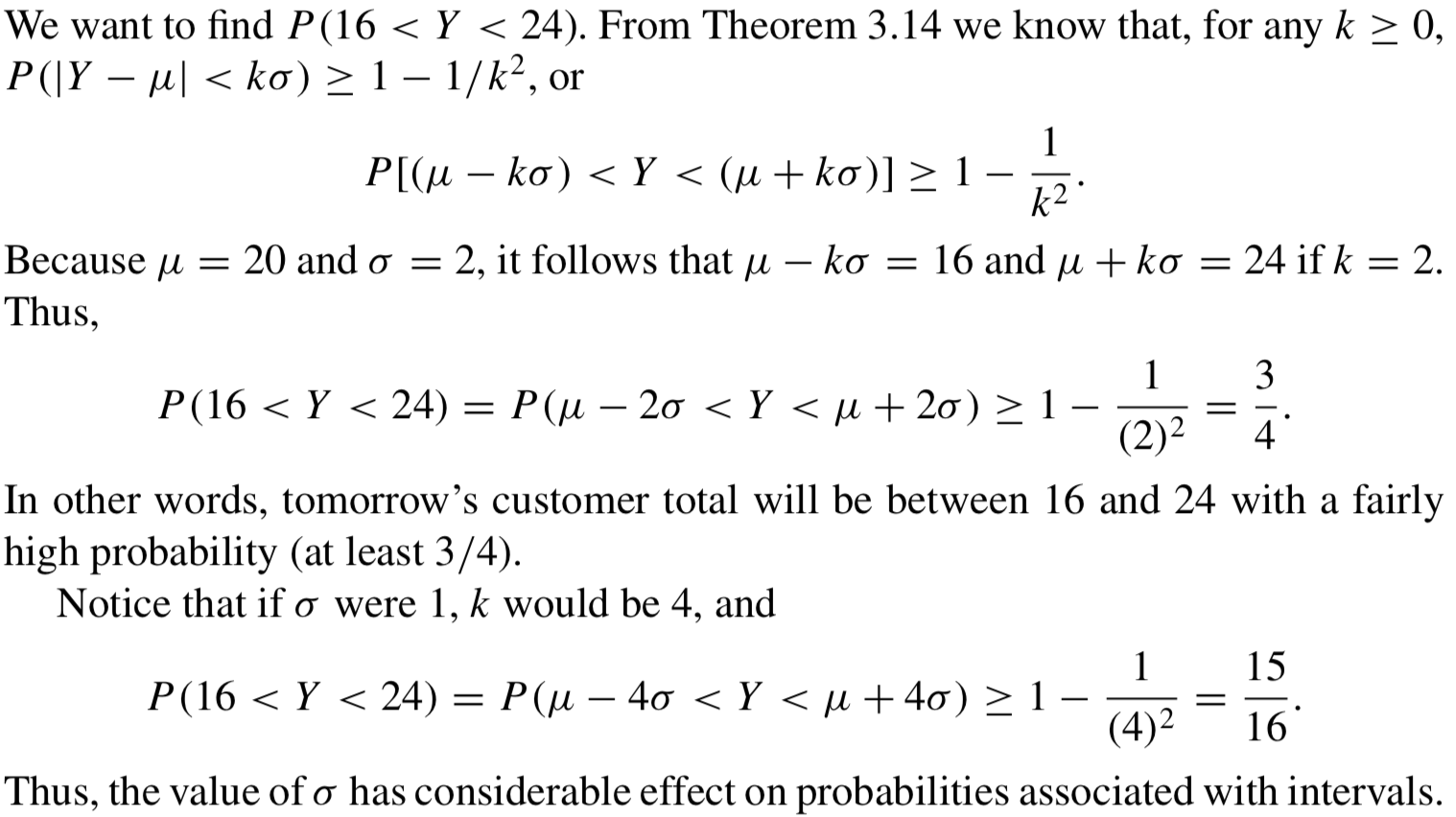
* 3.11 Tchebysheff’s Theorem



First, the result applies for any probability distribution.

Second, the results of the theorem are very conservative in the sense that the actual probability that Y is in the interval μ ± kσ usually exceeds the lower bound for the probability, 1 − 1/k2, by a considerable amount.

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



* 
* E(Y2)= E[Y(Y−1)]+E(Y)

