Lecture 3: Discrete Random Variables and Probability Distributions

MATH 697

Omidali Jazi

September 11, 2018

McGill University

Goals for this Chapter

- Learn about discrete random variables
  - Probability distributions
  - Expected value, variance, and standard deviation
- Define specific distributions
  - Binomial
  - Geometric
  - Negative Binomial
  - Hypergeometric
  - Poisson
- Define and apply Tchebycheff's Theorem

Introduction

- This chapter is concerned with the definitions of random variables, distribution functions $F(y)$, probability/density functions $f(y)$, and the development of the concepts necessary for carrying out calculations for a probability model using these entities.
- The concept of a random variable allows us to pass from the experimental outcomes themselves to a numerical function of the outcomes.

Introduction

There are two fundamentally different types of random variables in this course:

(i) discrete random variables

(ii) continuous random variables

We examine the basic properties and discuss the most important examples of discrete variables. We then focuses on continuous random variables.

Some Definitions

Definition 1 (Random Variable (rv))

A random variable X on Ω is a function from the sample space Ω to the set $\mathbb{R}$ of all real numbers denoted by

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

Let R_X denote the range of X.

They are random because they take on random values based on the outcome of the experiment.

Random Variables

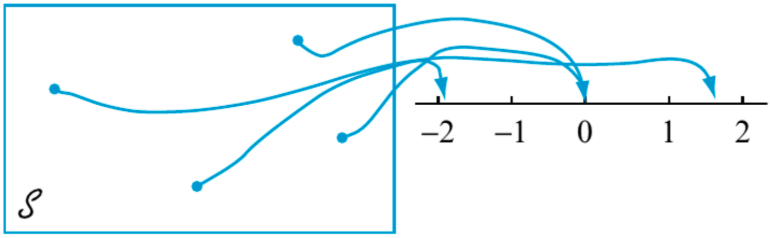


Figure 3.1 A random variable

Some Definitions

Definition 2

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

Example and Notation

- Example: If Y represents the outcome of a six-sided die, it can only take on values 1, 2, 3, 4, 5, and 6.
- Capital letters denote random variables, typically X, Y, and Z and small letters denote the value of a random variable, typically x, y, and z. They are the outcomes of the experiment.
- We write $P(Y = y)$ which in words is "the probability that random variable Y takes on the value y".
- Every random variable is associated with a probability distribution that specifies the possible r.v. values and the probability each value will occur.

Definition 3

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

- Note that $p(y) \geq 0$ for all y, but the probability distribution only assigns nonzero probabilities to a countable number of y values.
  It is convention that any value y not explicitly assigned a positive probability is understood to be zero; $p(y) = 0$.

Example

- Basically, the probability distribution of the random variable Y tells us how the total probability of 1 is "distributed" among all the possible values that Y can have.

- Example: Let the random variable Y be the sum of a pair of dice: $p(y) = (6 - |y - 7|)/36$
  The table is the probability distribution of y.

y	2	3	4	5	6	7	8	9	10	11	12
$P(Y=y)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Example

Example 4

A recruiting manager has six internship applications, three men and three women. He has to choose two for a special project. So as not to play favorites, he decides to select the two at random. Let Y denote the number of women selected. Find the probability distribution for Y.

Example: Results Summary

- We can represent the result as a formula:

$$p(y) = \frac{\begin{pmatrix} 3 \\ y \end{pmatrix} \begin{pmatrix} 3 \\ 2-y \end{pmatrix}}{\begin{pmatrix} 6 \\ 2 \end{pmatrix}}$$

  for $y = 0, 1, 2$

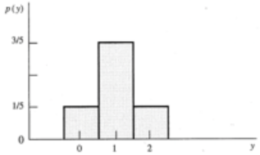
- As shown below, the results can be summarized in either a table or a probability histogram.

Example: Results Summary

Table 3.1 Probability distribution for Example 3.1

y	$p(y)$
0	1/5
1	3/5
2	1/5

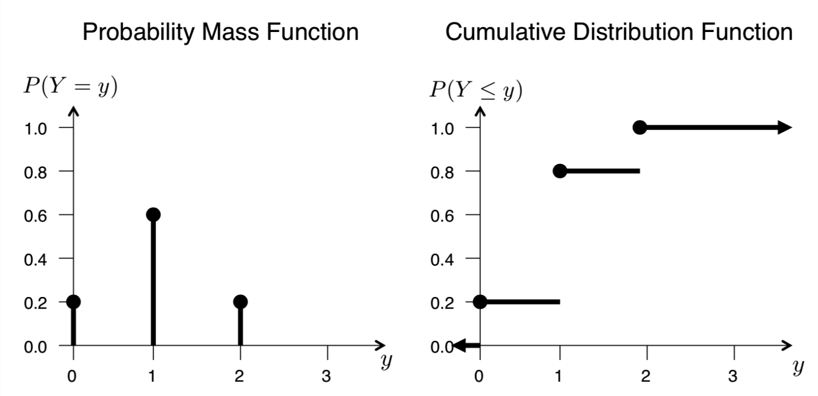
FIGURE 3.1
Probability histogram for Table 3.1



Some Common Terminology

- For discrete random variables, the probability distribution is often called the probability mass function (pmf).
- The notation for a pmf is $p(y) = P(Y = y)$.
- The probability distribution for continuous random variables is called the probability density function (pdf) which will be discussed later.
- The expression $F(y) = P(Y \leq y)$ is often referred to as cumulative distribution function (CDF).

Graphically Depicting the PMF and CDF



Probability Axioms Redux

Theorem 5

For any discrete probability distribution, the following must be true:

1. $p(y) \geq 0$ *for all* y.
2. $\sum_y p(y) = 1$ *where the summation is over all values of* y *with nonzero probability. Note these are just the Probability Axioms from the previous Chapter applied to random variables.*

Expected Value

Definition 6

Let Y be a discrete r.v. with probability function $p(y)$. Then the expected value of Y, $E(Y)$ is defined by

$$E(Y) = \sum_y yp(y)$$

To be precise about the definition, the expected value only exists if the above sum is absolutely convergent, that is

$$\sum_y |y|\, p(y) < \infty$$

Expected Value, continued

- We often use μ as shorthand to denote the expected value. i.e, $E(Y) = \mu$.
- Example: Consider a random variable Y with the following probability distribution:

Table 3.2 Probability distribution for *Y*

y	$p(y)$
0	1/4
1	1/2
2	1/4

Then

$$E(Y) = \sum_{y} yp(y) = 0 \times (1/4) + 1 \times (1/2) \times 2 \times (1/4) = 1$$

Expected Value of a Function of a R.V.

Theorem 7

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

$$E[g(Y)] = \sum_{y} g(y)p(y)$$

Variance & Standard Deviation of a R.V.

Definition 8

For a random variable Y with mean $E(Y) = \mu$, the variance of Y is defined as the expected value of $(Y - \mu)^2$. That is,

$$V(Y) = E[(Y - \mu)^2]$$

- The variance is literally the average squared deviation of Y from its mean.
  It's a measure of how variable a random variable is
  The bigger the variance, the more "spread out" (around the mean) the values are that the r.v. can take on.

The Standard Deviation of a R.V.

- The standard deviation of Y is the positive square root of $V(Y)$.
- Conventional notation often denotes the varianceas $V(Y) = \sigma^2$.
  Hence, the standard deviation is often denoted by σ.
- Note that the variance is in the mean's units squared.
  In comparison, the standard deviation is in the same units as the mean.
  This often makes the standard deviation easier to understand and interpret.

The Expected Value of a Constant is a Constant

Theorem 9

Let Y be a discrete r.v. with probability function $p(y)$ and let c be a constant. Then $E(c) = c$.

Theorem 10

Let Y be a discrete r.v. with probability function $p(y)$, let $g(Y)$ be a function of Y, and let c be a constant. Then

$$E[cg(Y)] = cE[g(Y)]$$

Continued

Theorem 11

Let Y be a discrete r.v. with probability function $p(y)$, and let $g_1(Y), ..., g_k(Y)$ be k functions of Y. Then

$$E[g_1(Y) + ...g_k(k)] = E[g_1(Y)] + ... + E[g_k(Y)]$$

A Theorem Simplifying Variance Calcs

Theorem 12

Let Y be a discrete r.v. with probability function $p(y)$ and mean μ. Then

$$V(Y) = \sigma^2 = E[(Y-\mu)^2] = E[Y^2] - \mu^2$$

Special Probability Distributions

- Up to this point, we have calculated various probability distributions that are unique to a particular problem or scenario
- However, there are some probability distributions that apply in many situations
- We are going to learn about a whole bunch of these now. They will come up over and over again in this curriculum and in your future life

Binomial Experiments

The binomial distribution applies when there is:

- A sequence of independent and identical trials.
- Each trial can result in only one of two outcomes.
- Examples:
  - A sequence of helmet tests where each shot is either a penetration or not.
  - A medical treatment where each person is either cured or not.
  - A retention bonus experiment where each sailor either reenlists or not.

Binomial Experiments

A binomial experiment has 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” or “failure”.
3. The probability of success on any trial is p and remains the same from trial to trial. (the probability of failure is $1–p$)
4. The trials are independent.
5. The random variable of interest (Y) is the number of successes out of the n trials.

The Binomial Distribution

Definition 13

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

$$p(y) = \begin{pmatrix} n \\ y \end{pmatrix} p^y (1-p)^{n-y}$$

where $y = 0, ..., n$ and $0 \leq p \leq 1$

Example: if the experiment is all failures then

$$p(0) = \begin{pmatrix} n \\ 0 \end{pmatrix} p^0 (1-p)^n = (1-p)^n$$

Mean and Variance of a Binomial R.V

Theorem 14

Let Y be a binomial random variable based on n trials and success probability p. Then

$$E(Y) = np$$

$$V(Y) = np(1 - p)$$

Binomial Distribution Example

Example 15 (Exxon)

Exxon has just bought a large tract of land in northern Quebec, with the hope of finding oil. Suppose they think that the probability that a test hole will result in oil is .2. Assume that Exxon decides to drill 7 test holes. What is the probability that

1. Exactly 3 of the test holes will strike oil?
2. At most 2 of the test holes will strike oil?
3. Between 3 and 5 (including 3 and 5) of the test holes will strike oil?
4. What are the mean and standard deviation of the number of test holes which strike oil.

Binomial Distribution Exxon Example (R code)

```r
# 1
dbinom(x = 3, size = 7, prob = 0.2)
```

```
## [1] 0.114688
```

```r
# 2
pbinom(q = 2, size = 7, prob = 0.2, lower.tail = TRUE)
```

```
## [1] 0.851968
```

```r
# 3
dbinom(x = 3, size = 7, prob = 0.2) +
  dbinom(x = 4, size = 7, prob = 0.2) +
  dbinom(x = 5, size = 7, prob = 0.2)
```

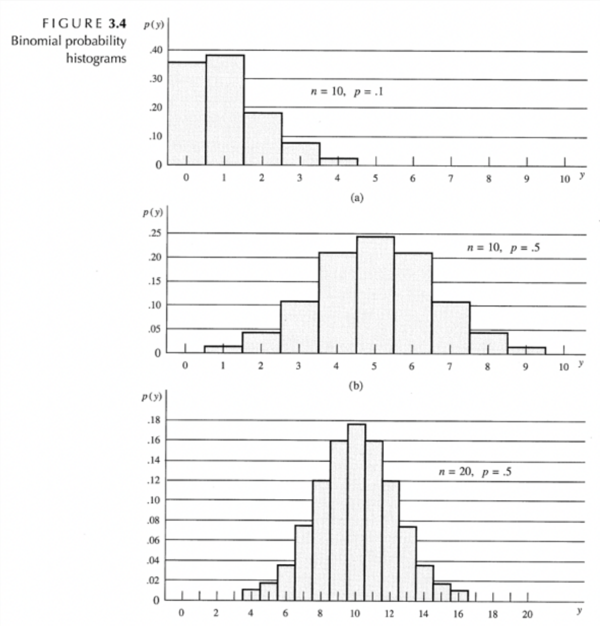
```
## [1] 0.1476608
```

Example (exercise)

Example 16

- Data show that 30% of people recover from a particular illness without treatment.
  - Ten ill people are selected at random to receive a new treatment and 9 recover after the treatment.
  - Assuming the medication is worthless (i.e., has no effect on the illness),
- what is the probability that prior to treatment we would have expected to observe that at least 9 of the 10 treated people recover?

Example Binomial Distributions



Reading Table

Table 1 Binomial Probabilities

Tabulated values are $P(Y \le a) = \sum_{y=0}^{a} p(y)$. (Computations are rounded at third decimal place.)

(a) $n = 5$

	p													
a	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	a
0	.951	.774	.590	.328	.168	.078	.031	.010	.002	.000	.000	.000	.000	0
1	.999	.977	.919	.737	.528	.337	.188	.087	.031	.007	.000	.000	.000	1
2	1.000	.999	.991	.942	.837	.683	.500	.317	.163	.058	.009	.001	.000	2
3	1.000	1.000	1.000	.993	.969	.913	.812	.663	.472	.263	.081	.023	.001	3
4	1.000	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226	.049	4

(b) $n = 10$

	p													
a	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	a
0	.904	.599	.349	.107	.028	.006	.001	.000	.000	.000	.000	.000	.000	0
1	.996	.914	.736	.376	.149	.046	.011	.002	.000	.000	.000	.000	.000	1
2	1.000	.988	.930	.678	.383	.167	.055	.012	.002	.000	.000	.000	.000	2
3	1.000	.999	.987	.879	.650	.382	.172	.055	.011	.001	.000	.000	.000	3
4	1.000	1.000	.998	.967	.850	.633	.377	.166	.047	.006	.000	.000	.000	4
5	1.000	1.000	1.000	.994	.953	.834	.623	.367	.150	.033	.002	.000	.000	5
6	1.000	1.000	1.000	.999	.989	.945	.828	.618	.350	.121	.013	.001	.000	6
7	1.000	1.000	1.000	1.000	.998	.988	.945	.833	.617	.322	.070	.012	.000	7
8	1.000	1.000	1.000	1.000	1.000	.998	.989	.954	.851	.624	.264	.086	.004	8
9	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.972	.893	.651	.401	.096	9

Geometric Probability Distribution

- Imagine (like with the binomial) you have a situation with a sequence of independent trials, where there are only two outcomes (success and failure), and the trials have probability of success p
- Now, if we let Y be the number of trials until the first success (where we won't fix the number of trials in advance), then Y has a geometric distribution.
  Example: The number of huanted animals before the first diseased animal has a geometric distribution.

Deriving the Geometric Distribution

Definition 17

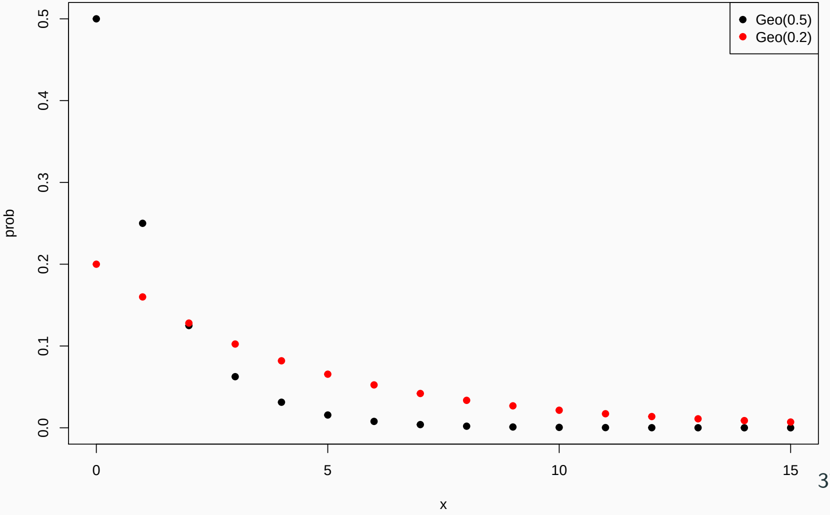
A random variable Y is said to have a geometric distribution with success probability p if and only if

$$p(y) = p(1-p)^{y-1}$$

for $y = 1, 2, ...$ and $0 \leq p \leq 1$

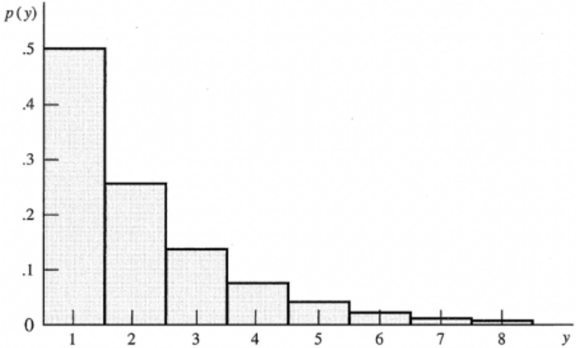
The Geometric Distribution

```
plot(0:15, dgeom(x = 0:15, prob = 0.5),pch = 19, xlab="x", ylab="prob")
points(0:15, dgeom(0:15, prob = 0.20), pch = 19, col = "red")
legend("topright", c("Geo(0.5)", "Geo(0.2)"), col = c("black","red"), pch=19)
```



Example of a Geometric Distribution

FIGURE 3.5 The geometric probability distribution, $p = .5$



Mean and Variance of a Geometric R.V

Theorem 18

Let Y be a geometric random variable with success probability p. Then

$$\mu = E(Y) = 1/p$$

and

$$\sigma^2 = V(Y) = (1-p)/p^2$$

Example

Example 19

Assume the probability an engine fails during any one-hour period is constant at p = 0.02. Find the probability that the engine continues operating for more than two hours.

Solution:

$P(Y > 2) = 1 - p(1) - p(2) = 1 - 0.02 - 0.02 \times 0.98 = 0.96$

Shortcut: $P(Y > y) = (1 - p)^y$

Negative Binomial Probability Distribution

- Assume you have a situation with only two outcomes (success and failure) and constant probability of success p.
- If Y is the number of trials until the rth success ($r > 1$), then Y has a negative binomial distribution.

Deriving the Negative Binomial Distribution

Definition 20

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

$$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$$

for $y = r, r+1, ...$ and $0 \leq p \leq 1$

Mean and Variance of a Negative Binomial Random Variable

Theorem 21

If Y is a r.v. with a negative binomial distribution, then the mean is

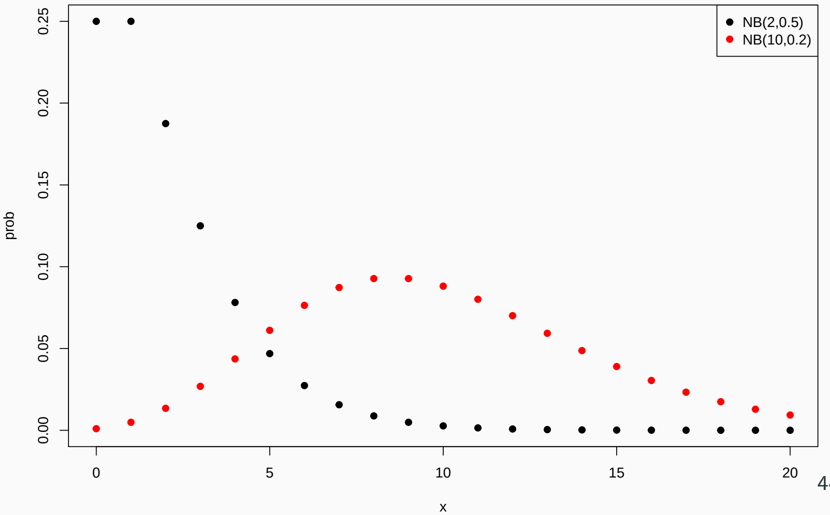
$$E[Y] = r/p$$

and the variance is

$$V(Y) = r(1-p)/p^2$$

The Negative Binomial Distribution

```
plot(0:20, dnbinom(x = 0:20, size = 2, prob = 0.5),pch = 19, xlab="x", ylab="pr
points(0:20, dnbinom(0:20, size = 10, prob = 0.5), pch = 19, col = "red")
legend("topright", c("NB(2,0.5)", "NB(10,0.2)"), col = c("black","red"), pch=19
```



Example

Example 22

A geological study indicates that wells drilled in a particular region should strike oil with probability $p = 0.2$. Find the probability that the third oil strike comes on the fifth well drilled.

Solution: $p(5) = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \times 0.2^3 \times 0.8^2 = 0.03$

Hypergeometric Probability Distribution

To apply the binomial distribution to some finite population of size N, the sample size n must be much smaller than N. We are going to sample without replacement.

- There are ${}_NC_n$ ways to choose a sample of size n out of a population of size N.
- So, the probability of choosing any one particular sample $E_i \in S$ is $P(E_i) = \frac{1}{\left(\begin{array}{c} N \\ n \end{array} \right)}$.
- Now, if there are r red elements in the population, then the probability of picking y red elements out of the r is ${}_rC_y$.

Hypergeometric Distribution Derivation

- If there are r red elements then there must be $N - r$ black elements
  And, if you selected y red then there must be $n - y$ black elements in the sample.
  Thus, there are ${}_{N-r}C_{n-y}$ ways to select the black elements.
- Using the mn rule, it thus follows that there elements are ${}_{r}C_{y} \times {}_{N-r}C_{n-y}$ ways to choose y red and $n - y$ black elements from a population of size N containing r red elements.

The Hypergeometric Distribution

Definition 23

A random variable Y has a hypergeometric distribution if and only if

$$p(y) = \frac{\begin{pmatrix} r \\ y \end{pmatrix} \begin{pmatrix} N-r \\ n-y \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

where y is an integer on either $0, 1, ..., n$ if $r \geq n$ or $0, 1, ..., r$ if $r < n$, Subject to the restrictions that $y \leq r$ and $n - y \leq N - r$.

That is, the restrictions are $n \leq N$, $r \leq N$ and $y \leq min(n, r)$

Mean and Variance of a Hypergeometric Random Variable

Theorem 24

If Y is a r.v. with a hypergeometric distribution, then the mean is

$$E[Y] = \frac{nr}{N}$$

and the variance is

$$V(Y) = n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1}$$

Poisson Probability Distribution

- The Poisson distribution is often a good model for the number of rare events Y that occur in space or time.
  "Rare" means that the probability that two events happen in the same (very small) increment of space or time is zero (or essentially zero).

- Useful for modeling the number of events that occur per unit space or time.
  Example: Number of accidents/incidents, equipment failures, arrivals, defects, etc.

- Can also be used to approximate the binomial.

The Poisson Distribution

Definition 25

A random variable Y has a Poisson distribution if and only if

$$P(Y = y) = e^{-\lambda}\frac{\lambda^y}{y!}$$

for $y = 0, 1, 2, \ldots$ and $\lambda > 0$.

- It's related to the binomial, as we just saw. But in the binomial n is finite and fixed. In the Poisson there is no n per se.
  Instead, we are looking at infinitesimal regions or time slices where at most one event can occur.
- The parameter λ is the average number of events per unit.

Mean and Variance of a Poisson Random Variable

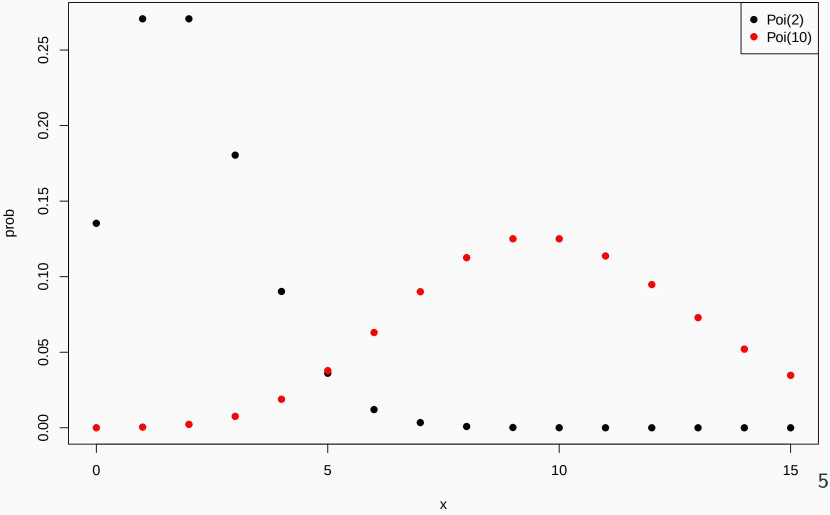
Theorem 26

If Y is a r.v. with a Poisson distribution with parameter λ, then the mean and variance are

$$E(Y) = V(Y) = \lambda$$

The Poisson Distribution

```r
plot(0:15, dpois(x = 0:15, lambda = 2),pch = 19, xlab="x", ylab="prob")
points(0:15, dpois(0:15, lambda = 10), pch = 19, col = "red")
legend("topright", c("Poi(2)", "Poi(10)"), col = c("black","red"), pch=19)
```



Poisson is a Limiting Form of the Binomial

- For $\lambda = np$:

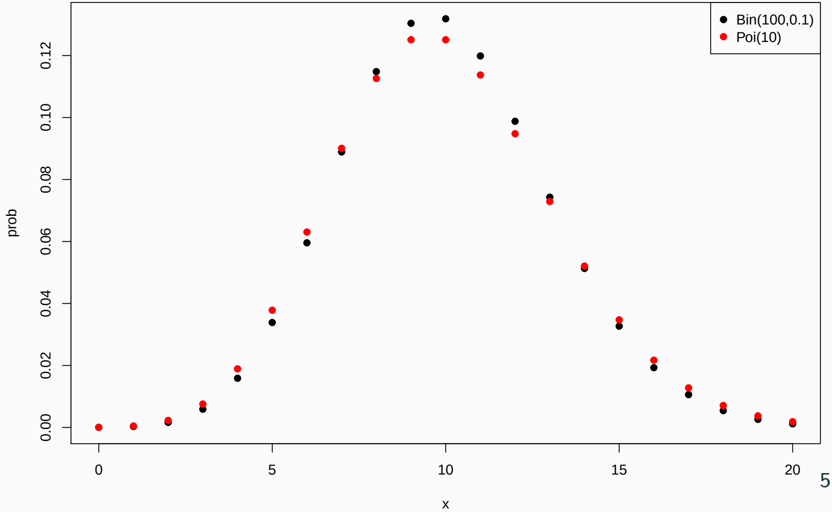
$$\begin{aligned}
\lim_{n\to\infty} \binom{n}{y} p^y (1-p)^{n-y} &= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-y+1)}{y!} \left(\frac{\lambda}{n}\right)^y \left(1-\frac{\lambda}{n}\right)^{n-y} \\
&= \lim_{n\to\infty} \frac{\lambda^y}{y!} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-y} \frac{n(n-1)\cdots(n-y+1)}{n^y} \\
&= \frac{\lambda^y}{y!} \lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-y} \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{y-1}{n}\right)
\end{aligned}$$

- Now, note that $\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$ and all other terms to the right of it have a limit of 1
- Thus,

$$\lim_{n\to\infty} \binom{n}{y} p^y (1-p)^{n-y} = \frac{\lambda^y}{y!} e^{-\lambda}$$

Poisson approximation to the Binomial

```
plot(0:20, dbinom(x = 0:20, size = 100, p=0.1), pch = 19, xlab="x", ylab="prob"
points(0:20, dpois(0:20, lambda = 100*0.1), pch = 19, col = "red")
legend("topright", c("Bin(100,0.1)", "Poi(10)"), col = c("black","red"), pch=19
```



Poisson Processes

- A Poisson process is a stochastic process which counts the number of events per unit of space or time
- In a Poisson process with mean number of occurrences λ per unit, the number of occurrences Y in a units has a Poisson distribution with mean $a\lambda$.
- A key assumption with Poisson processes is that the number of occurrences in disjoint intervals is independent.

Difference Between Binomial and Poisson R.V.s

- Binomial:
  Event is the number of "successes" in a sequence of n independent trials.
  Example: number of operational units out of n sampled.

- Poisson:
  Events occur per unit basis (per unit time, per unit area, per unit volume, etc.)
  Example: number of defectives per unit.

Tchebysheff's Theorem

- The empirical rule says that approximately 95% of observations from a symmetric (or "bell shaped") probability distribution will be in the interval $\mu \pm 2\sigma$.
  It's a useful rule of thumb for assessing whether an observation is "unusual".
- chebysheff's Theorem provides a conservative lower bound for the probability that an observation falls in the interval $\mu \pm k\sigma$.
  It applies to any probability distribution.

Tchebysheff's Theorem

Theorem 27

Let Y be a r.v. with mean μ and finite variance σ^2. Then, for any constant $k > 0$.

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

Equivalently

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

Tchebysheff's Theorem Explained

$$\underbrace{P(\mu - k\sigma \le Y \le \mu + k\sigma)}_{\substack{\text{True probability that } Y \text{ is} \\ \text{within } k \text{ standard deviations of} \\ \text{the mean}}} \ge \underbrace{1 - \frac{1}{k^2}}_{\substack{\text{Lower bound} \\ \text{on the true} \\ \text{probability}}}$$

$$\underbrace{P(Y \le \mu - k\sigma \text{ or } Y \ge \mu + k\sigma)}_{\substack{\text{True probability that } Y \text{ is} \\ \text{outside of } k \text{ standard} \\ \text{deviations from the mean}}} \le \underbrace{\frac{1}{k^2}}_{\substack{\text{Upper bound} \\ \text{on the true} \\ \text{probability}}}.$$

Example

Example 28

The number of customers per day at a sales counter, Y has mean 20 and standard deviation 2. The probability distribution is unknown. What can be said about the probability that tomorrow $16 < Y < 24$.

Solution:

$$P(16 < Y < 24) = P(|Y - 20| < 4) = P(|Y - \mu| < 2\sigma)$$
$$\geq 1 - 1/2^2 = 0.75$$

conclusion: At least 75% of the customers per day at the sale counter are in this interval.

Example (exercise)

Example 29

On average, a supply depot gets 156 requisitions per day with a standard deviation of 22. What is an upper bound on the probability that the number of requisitions in one day will be will be more than 3 standard deviations away from the mean.

What We Have Just Learned

- Learned about discrete random variables
  Probability distributions
  Expected value, variance, and standard deviation
- Defined specific distributions.
  Binomial
  Geometric
  Negative Binomial
  Hypergeometric
  Poisson
- Defined and applied Tchebycheff's Theorem