1 Random Variables and Distributions

Random variables are the main link between probability and statistics. In statistics we observe numbers (or data) as the result of an experiment, and a random variable links the numbers to the probability structure of the experiment.

Definition of random variable

Let the sample space of an experiment be $\mathcal{S}$. A *random variable* is a mapping, or function, from $\mathcal{S}$ to the real number line.

We will use capital letters late in the alphabet, such as X, Y and Z, to denote random variables. If X is a random variable, then X associates with each $s \in \mathcal{S}$ a real number $X(s)$.

The notation X and $X(s)$ parallels the notation we often see in math classes, where

- f is used to denote a function, and
- $f(x)$ is the value of f at argument x.

So,

$$X \Longleftrightarrow f \quad \text{and} \quad X(s) \Longleftrightarrow f(x).$$

A value of X will often be denoted x, i.e., $X(s) = x$.

Example 12 Suppose that a coin is tossed four times. Then there are $2^4 = 16$ possible sequences of tosses (such as $HTHT$). Let X be the number of heads until the first tail. For our example sequence, the mapping is $X(HTHT) = 1$. We can form a table of the mapping for all possible sequences:

$$X(TTTT) = 0 \quad X(TTTH) = 0 \quad X(TTHT) = 0 \quad X(TTHH) = 0$$

$$X(THTT) = 0 \quad X(THTH) = 0 \quad X(THHT) = 0 \quad X(THHH) = 0$$

$$X(HTTT) = 1 \quad X(HTTH) = 1 \quad X(HTHT) = 1 \quad X(HTHH) = 1$$

$$X(HHTT) = 2 \quad X(HHTH) = 2 \quad X(HHHT) = 3 \quad X(HHHH) = 4$$

Let X be a random variable defined on a sample space $\mathcal{S}$, and let A be some subset of the real numbers. We then define $P(X \in A)$ by

$$P(X \in A) = P\left(\{s \in \mathcal{S} : X(s) \in A\}\right).$$

The probabilities so defined by all relevant subsets A is called the *probability distribution* of X.

We use $P(X = x)$ as a shorthand for $P(X \in \{x\})$.

Example 13 In the experiment of Example 12, we assume that the coin is a fair coin. Then each outcome has probability $\frac{1}{16}$.

For example, we have

$$P(X = 1) = P(\{HTTT, HTTH, HTHT, HTHH\}) = \frac{4}{16} = 0.25.$$

Similar reasoning yields:

x	0	1	2	3	4
$P(X = x)$	0.5	0.25	0.125	0.0625	0.0625

Any other probability of interest concerning the random variable X may be determined from these probabilities.

There are two main types of random variables: *discrete* and *continuous*.

Remember, X is a mapping from $\mathcal{S}$ to some subset of the real numbers.

The domain of the mapping is $\mathcal{S}$, and we'll call the range R_X.

If R_X is countable, then X is a *discrete* random variable. If R_X is not countable, then X is a *continuous* random variable. When X is continuous, R_X is usually an interval or a union of disjoint intervals.

When $\mathcal{S}$ is countable, then X *must* be discrete, while if $\mathcal{S}$ is uncountable, then X can be either discrete or continuous.

2 Discrete Random Variables

The *probability function (or probability mass function)* of a discrete random variable is a function p_X defined by

$$p_X(x) = P(X = x) \quad \text{for every real number } x.$$

Write the range of X as $R_X = \{x_1, x_2, \ldots\}$. Then

$$\sum_{i=1}^{\infty} p_X(x_i) = 1.$$

For any subset A of real numbers, we may express $P(X \in A)$ as

$$P(X \in A) = \sum_{x \in A \cap R_X} p_X(x).$$

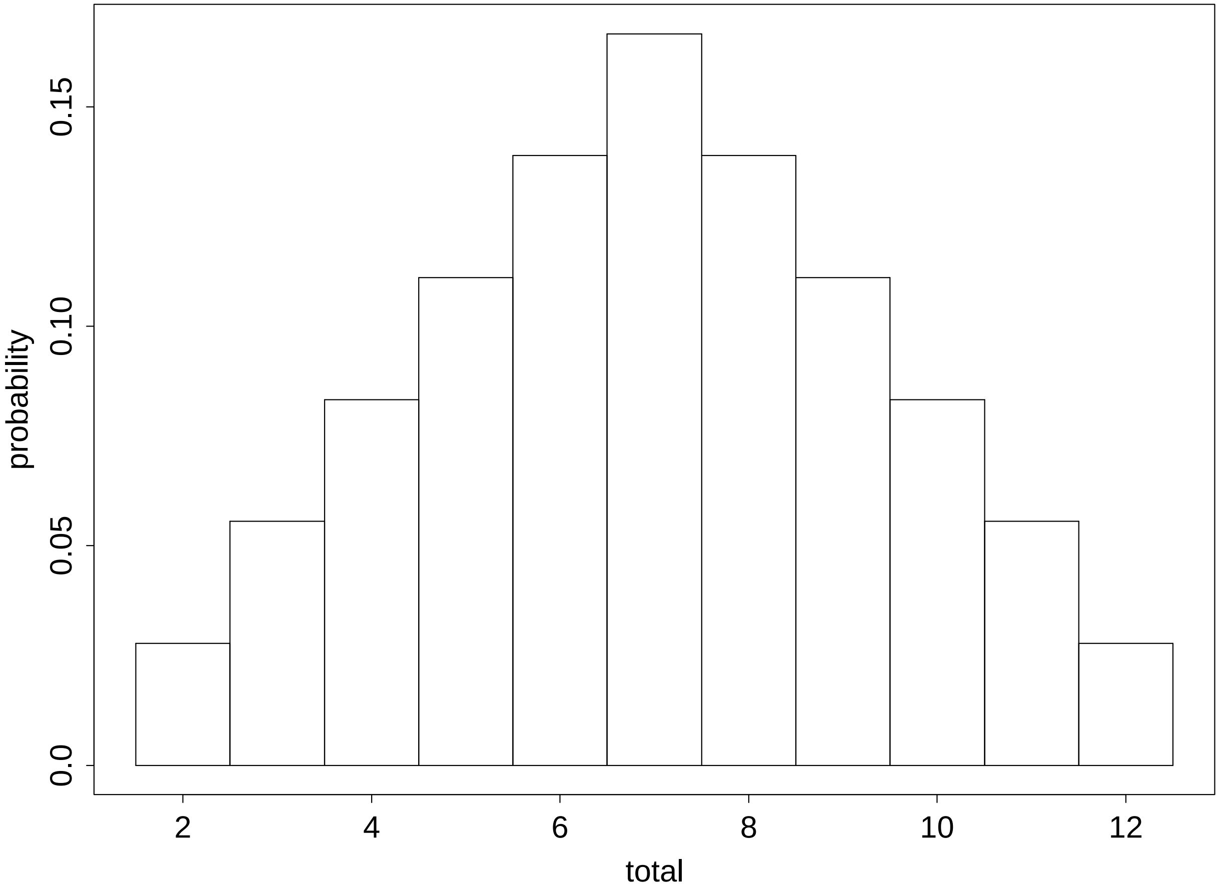
Example 4 revisited Consider again our dice experiment. If the dice are balanced, then the probability of each of the 36 different outcomes is the same. In this case, for each (i, j)

$$P(\{(i, j)\}) = \frac{1}{36}. \qquad \text{Why?}$$

We define the random variable X to be total on the two dice. We can compute the probability mass function for X from the probabilities on the original sample space. For example,

$$p_X(6) = P(X = 6) = P(\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}) = \frac{5}{36}.$$

We can likewise find the probability of any other possible total. A graph of the probability mass function is given below.



0.0
0.05
0.10
0.15
probability
2
4
6
8
10
12
total

Example 4 continued

Suppose now we let $A = \{x : x \leq 4\}$. We are interested in $P(X \in A)$. We note that

$$R_X = \{2, 3, 4, \ldots, 11, 12\} \text{ and } A \cap R_X = \{2, 3, 4\}.$$

We have two ways of finding $P(X \in A)$.

1. We can use the original probability space:

$$P(X \in A) = P(\{(1,1),(1,2),(2,1),(1,3),(2,2),(3,1)\}) = \frac{6}{36} = \frac{1}{6}$$

2. We can use the probability mass function of X:

$$P(X \in A) = \sum_{x=2}^{4} p_X(x) = p_X(2)+p_X(3)+p_X(4) = \frac{1}{36}+\frac{2}{36}+\frac{3}{36} = \frac{6}{36} = \frac{1}{6}$$

A function that is often useful in defining and proving properties of a random variable is an *indicator function*.

The *indicator function* of an event A is defined as

$$I_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \notin A \end{cases}$$

Example: **Exercise 2.1.6 (modified a little)** Let $\mathcal{S} = \{1, 2, 3, 4\}$, $X = I_{\{1,2\}}$, $Y = I_{\{2,3\}}$, $Z = I_{\{3,4\}}$, and $W = X - 2Y + Z$.

$$W(1) = X(1) - 2Y(1) + Z(1) = 1 - 2(0) + 0 = 1$$

$$W(2) = X(2) - 2Y(2) + Z(2) = 1 - 2(1) + 0 = -1$$

$$W(3) = X(3) - 2Y(3) + Z(3) = 0 - 2(1) + 1 = -1$$

$$W(4) = X(4) - 2Y(4) + Z(4) = 0 - 2(0) + 1 = 1$$

We'll discuss several probability mass functions (pmfs) that are important in statistical applications. These are the pmfs for the **Bernoulli distribution**, **discrete uniform distribution**, the **binomial distribution**, the **negative binomial distribution**, the **Poisson distribution**, and the **hypergeometric distribution**.

2.1 Bernoulli distribution

The simplest discrete random variable X takes on only two values, 0 and 1. Suppose that $0 < \theta < 1$. The probability mass function (pmf) of X is

$$\begin{aligned} p_X(1) &= \theta \\ p_X(0) &= 1 - \theta \\ p_X(x) &= 0, \qquad \text{otherwise.} \end{aligned}$$

We can also write the pmf as

$$p_X(x) = \begin{cases} \theta^x (1-\theta)^{1-x} & \text{if } x = 0 \text{ or } x = 1 \\ 0, & \text{otherwise.} \end{cases}$$

2.2 Discrete uniform distribution

The discrete uniform probability mass function p_X is defined for a positive integer k by

$$p_X(x) = \begin{cases} 1/k, & x = 1, 2, \ldots, k \\ 0, & \text{otherwise.} \end{cases}$$

This probability function is the distribution of a random variable X that has range $\{1, \ldots, k\}$ and is equally likely to take on any value in this range.

The uniform distribution arises in the analysis of *ranks* in statistics. Suppose that k numbers are drawn randomly from an *infinite* population (to be defined later).

Let R be the rank of the first number drawn among all k numbers. So, $R = 1$ if the first number drawn is the smallest one, $R = 2$ if the first number drawn is the next to the smallest, and so on.

It turns out that R has a discrete uniform distribution in this case.

2.3 Binomial distribution

Binomial experiment

1. Observe a sequence of n trials, where n is fixed in advance.
2. Each trial results in one of two possible outcomes; call them "success" and "failure" (S and F).
3. The trials are independent of each other.
4. The probability of S on any one trial is θ where $0 < \theta < 1$. (Note: θ remains the same from trial to trial.)

Define X to be the number of successes among the n trials of a binomial experiment. Then the probability mass function of X has the following form:

$$p_X(x) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x}, & x = 0, 1, 2, \ldots, n \\ 0, & \text{otherwise.} \end{cases}$$

This pmf is called the *binomial pmf*.

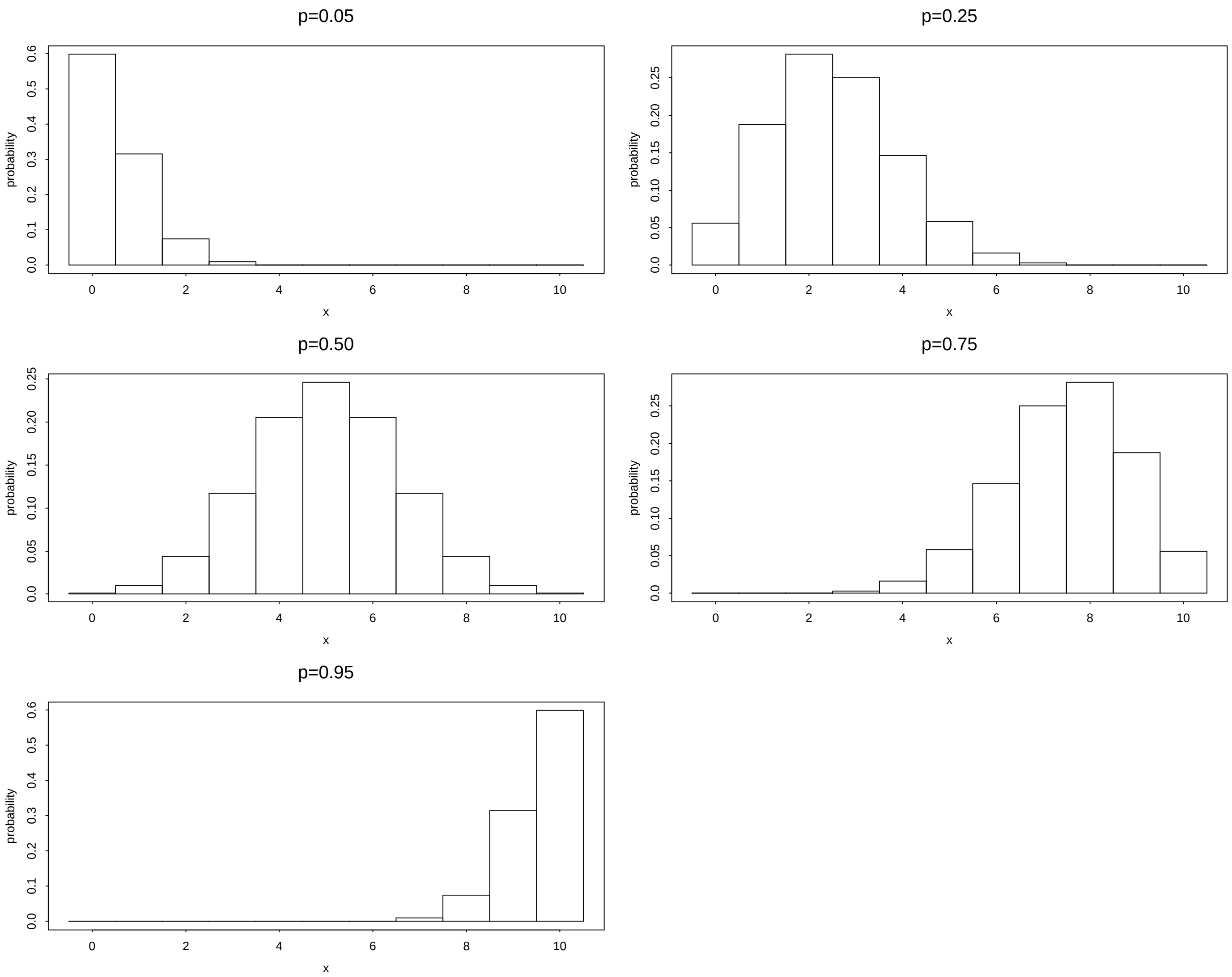
We can use ideas from Chapter 1 to prove that p_X is the pmf of the number of successes X in a binomial experiment.

There is a statistical application for the binomial distribution in sampling from a finite population. Suppose a population consists of N items, M of which are defective and $N - M$ nondefective.

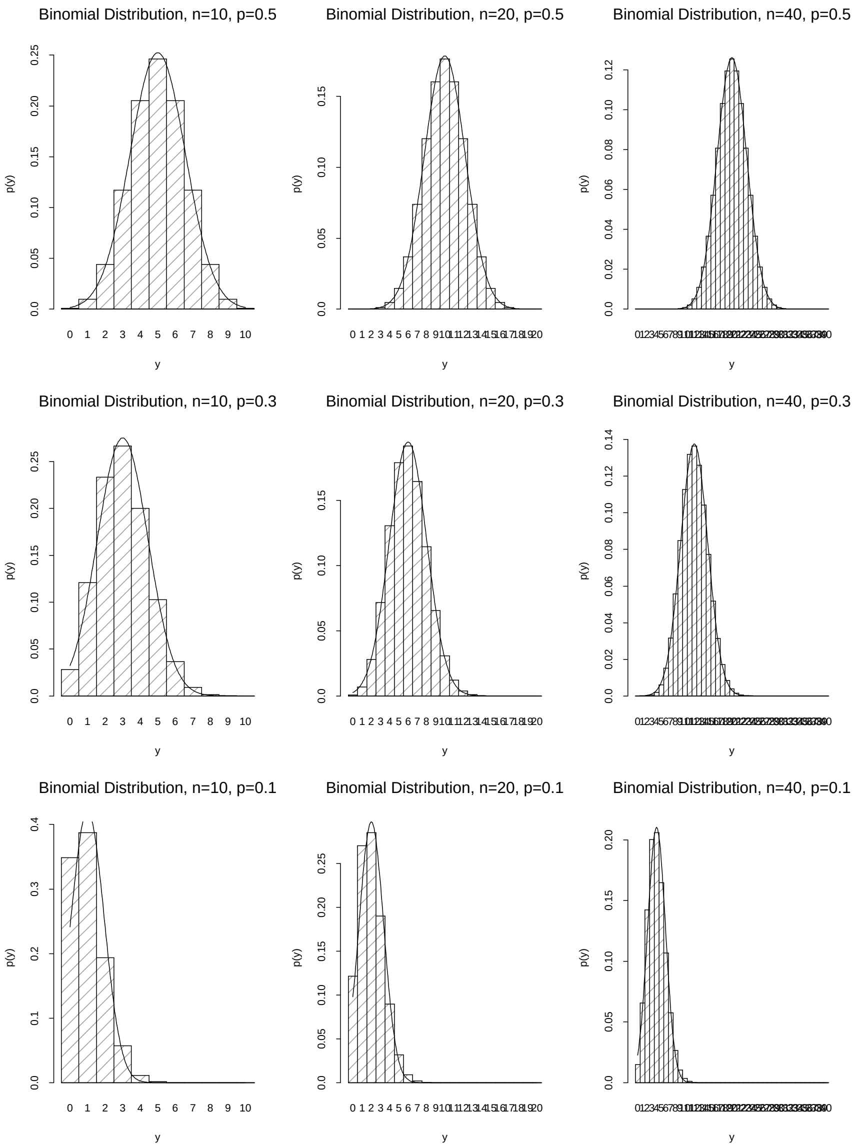
Suppose we randomly select n items from the population with replacement. Let X be the number of defective items among the n selected. Then X has the binomial distribution with $\theta = M/N$.

To argue that the distribution is binomial, just verify the conditions of the binomial experiment in this setting.

Various Binomial Distributions for $n = 10$



Some Binomial Distributions for $n = 10, 20$ and 40



2.4 Negative Binomial Probability Distribution

Consider the following simple experiment:

We flip a coin until we have tossed 5 heads.

Note:

1. The experiment consists of a sequence of independent trials.
2. Each trial is identical and can result in one of two outcomes (S or F).
3. The probability of S equals θ ($0 < \theta < 1$) for each trial.
4. We continue the experiment until r successes have been observed.

Any experiment meeting all of the above conditions is called a

Negative Binomial Experiment.

If Y is our random variable representing *the number of failures obtained before obtaining* r *successes*, then Y has a negative binomial distribution:

$$Y \sim \text{Negative Binomial}(r, \theta)$$

$$\begin{array}{lcl} r & = & \text{number of } S \\ \theta & = & \text{probability of } S \end{array}$$

The probability mass function of a negative binomial rv is given by

$$p_Y(y) = P(Y = y) = \binom{r-1+y}{r-1} \theta^r (1-\theta)^y, \qquad y = 0, 1, 2, \ldots$$

- If $r = 1$ and $X =$ the number of failures until the first success, we have a *geometric* distribution with pmf

$$P(X = x) = p_X(x) = \theta(1-\theta)^x, \qquad x = 0, 1, 2, 3, \ldots$$

- Some books define the negative binomial distribution as *the number of trials it takes to obtain the* r^{th} *success*. This changes the formulas a little.

2.5 Hypergeometric Distribution

The hypergeometric distribution applies to "sampling without replacement" from a finite population containing two types of items.

We have a population of size N containing M defective items and $N - M$ nondefective items. We randomly select n items ($n \leq N$) *without replacement*.

Define X to be the number of defectives in the sample. Then X has the probability mass function

$$p_X(x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & x = M_1, \ldots, M_2, \\ 0 & \text{otherwise,} \end{cases}$$

where $M_1 = \max(0, n - (N - M))$, and $M_2 = \min(n, M)$.

This is the pmf of the *hypergeometric distribution*.

2.6 Poisson Distribution

Consider these random variables:

- Number of phone calls received per hour by AAA emergency service.
- Number of customers logging onto Amazon Prime in a 5 minute interval.
- Number of trees in an area of forest.
- Number of bacteria in a culture.

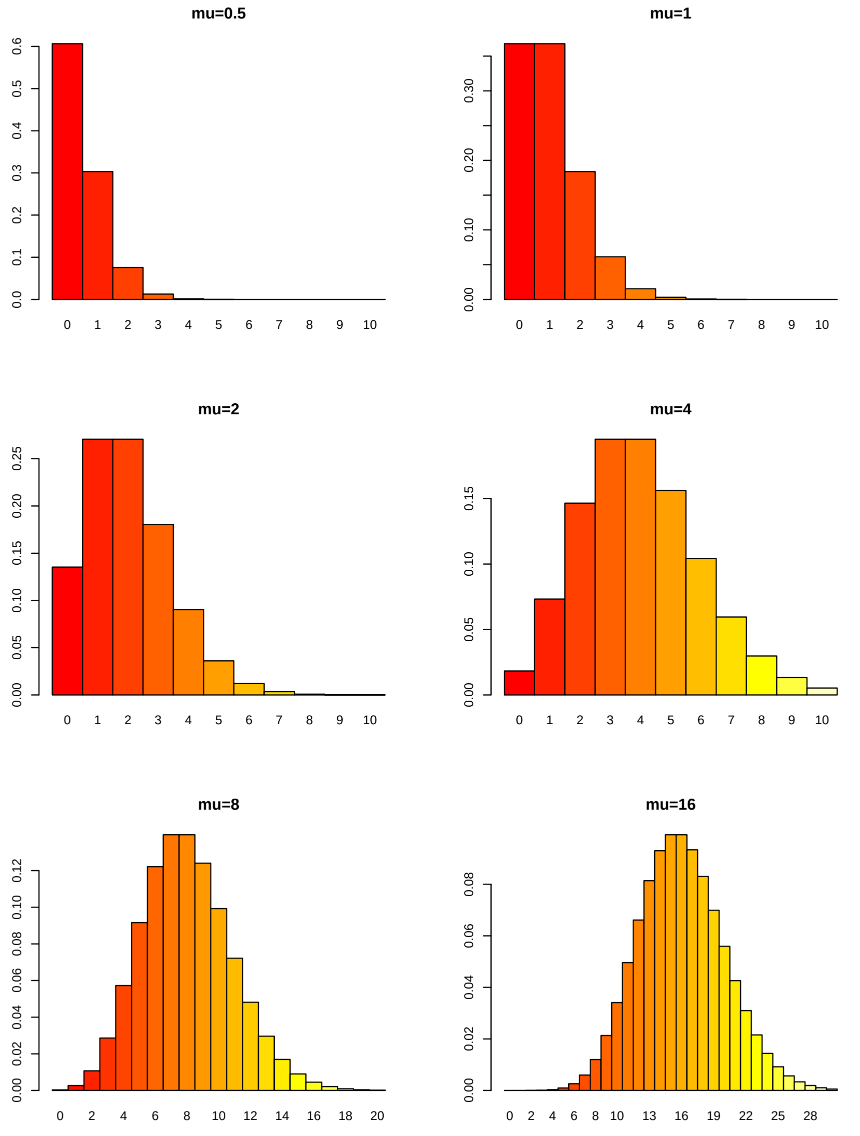
A random variable X, the number of successes occurring during a given time interval or in a specified region, is called a *Poisson* random variable. The corresponding distribution of

$$Y \sim \text{Poisson}(\lambda)$$

where λ is the rate for the given time or area, has pmf

$$p_Y(y) = P(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0, 1, 2, \ldots, \; \lambda > 0.$$

Some Poisson Distributions



2.7 Some Relationships among the Distributions

- If we sample without replacement and n is small relative to N and M, we can approximate the hypergeometric distribution by using the binomial distribution with $\theta = \frac{M}{N}$:

$$X \sim \text{Hypergeometric}(N, M, n) \;\rightarrow\; X \sim \text{Binomial}(n, \theta = \frac{M}{N})$$

- Let X be a binomial random variable with probability distribution $X \sim \text{Binomial}(n, \theta)$. When $n \to \infty$ and $\theta \to 0$ and $\lambda = n\theta$ remains fixed at $\lambda > 0$, then

$$X \sim \text{Binomial}(n, \theta) \rightarrow X \sim \text{Poisson}(\lambda = n\theta)$$

As a rule of thumb, this approximation can be safely applied if:

$$n \geq 100 \qquad \theta \leq .01 \qquad n\theta \leq 20$$