CIVL 7012/8012

Discrete Distributions

Definitions

A *random variable* is a "function" that associates a unique numerical value with every outcome of an experiment.

A ***probability distribution*** is a "function" that defines the probability of occurrence of every possible value that a random variable can take on.

Probability Distributions

There are two general types of probability distributions:

- ***Discrete***
- ***Continuous***

A ***discrete random variable*** can only take on discrete (i.e., specific) values.

A ***continuous random variable*** takes on continuous values (i.e., real values).

Properties of Discrete Distributions

The ***probability mass function*** (PMF) gives the probability that the random variable X will take on a value of x when the experiment is performed (also referred as prob. density function, PDF):

$$p(x) = \mathrm{P}(X = x)$$

By definition, *p(x)* is always a number between zero and one:

$$0 \leq p(x) \leq 1$$

and, since every trial must have *exactly one* outcome,

$$\sum_{x} p(x) = 1$$

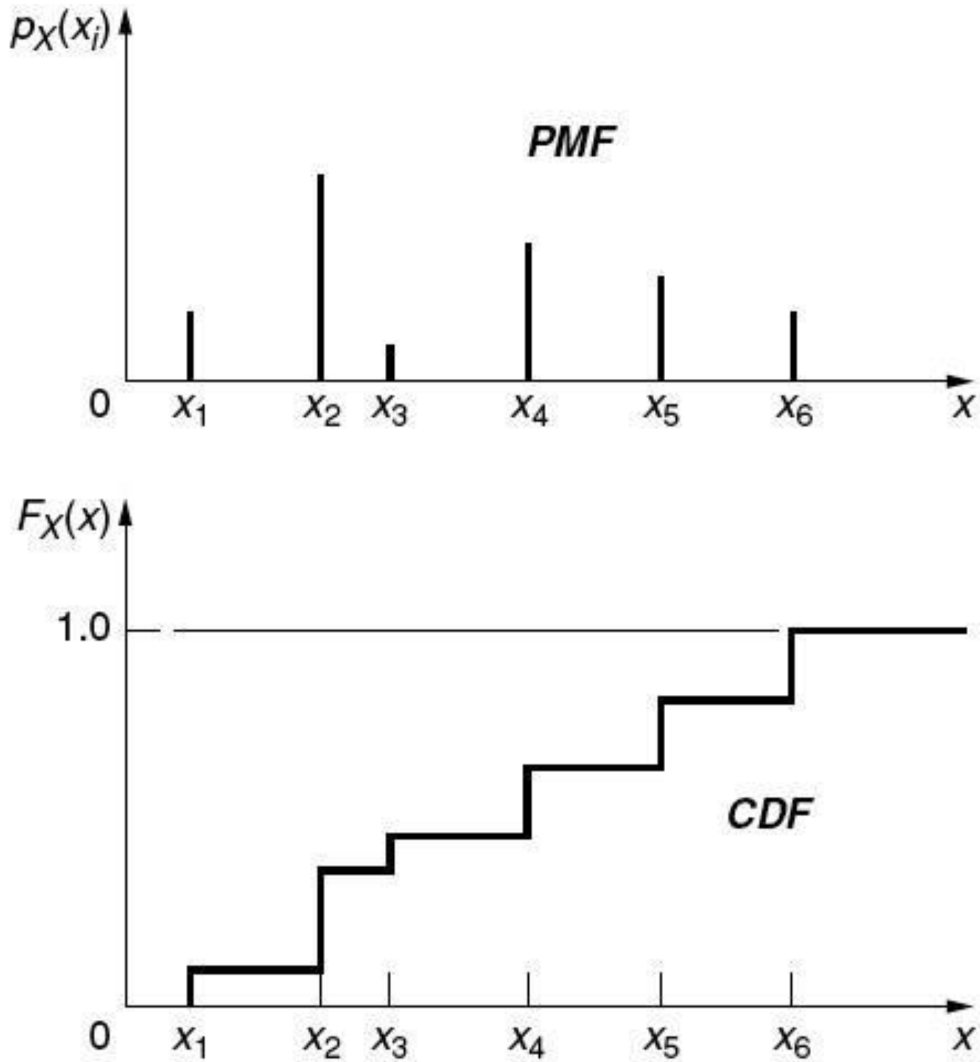


Figure 3.2a Discrete probability distribution.

Properties of Discrete Distributions

The ***cumulative distribution function (CDF)*** gives the probability that t random variable X will take on a value less than or equal to x when the experiment is performed:

$$F(x) = \mathrm{P}(X \le x) = \sum_{x_i \le x} p(x_i)$$

The ***expected value*** of a random variable is the probability-weighted average of the possible outcomes:

$$E(X) = \mu_x = \sum_x x \cdot p(x)$$

Properties of Discrete Distributions

The ***variance*** of a probability distribution is a measure of the amount of variability in the distribution of the random variable, X, about its expected value.

Mathematically, the variance is just the probability-weighted average of the squared deviations:

$$V(X) = \sigma_x^2 = \sum_x (x - \mu_x)^2 \, p(x)$$

We can also calculate the variance as:

$$V(X) = E\left(X^2\right) - \left[E(X)\right]^2$$

Bernoulli Trials

To be considered a Bernoulli trial, an experiment must meet three criteria:

1. There must be only 2 possible outcomes.
2. Each outcome must have an invariant probability of occurring. The probability of success is usually denoted by p, and the probability of failure is denoted by $q = 1 - p$.
3. The outcome of each trial is completely independent of the outcome of any other trials.

Discrete Distributions

Binomial Distribution – Probability of **exactly X successes in n trials**.

Negative Binomial – Probability that it will take **exactly n trials to produce exactly X successes**.

Geometric Distribution – Probability that it will take **exactly n trials to produce exactly one success**. (Special case of the negative binomial).

Hypergeometric Distribution – Probability of **exactly X successes in a sample of size n drawn without replacement**.

Poisson Distribution – Probability of **exactly X successes in a "unit" or continuous interval**.

Binomial Distribution

Gives the probability of *exactly x successes in n trials*

A. Requirements

There must be x successes and $(n-x)$ failures in the n trials, but the order in which the successes and failures occur is immaterial.

B. Mathematical Relationships

1. General Equation:

$$p(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}$$

where

n = the number of trials

x = the number of successes

p = the probability of a success for any given trial

$q = 1 - p$ = the probability of a failure for any given trial

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$ = the Binomial Coefficient

Binomial Distribution

2. Expectation – the expected (mean) number of successes in n trials

$$E(X) = np$$

3. Variance – the expected sum of the squared deviations from the mean

$$V(X) = npq$$

Binomial Distribution Shapes

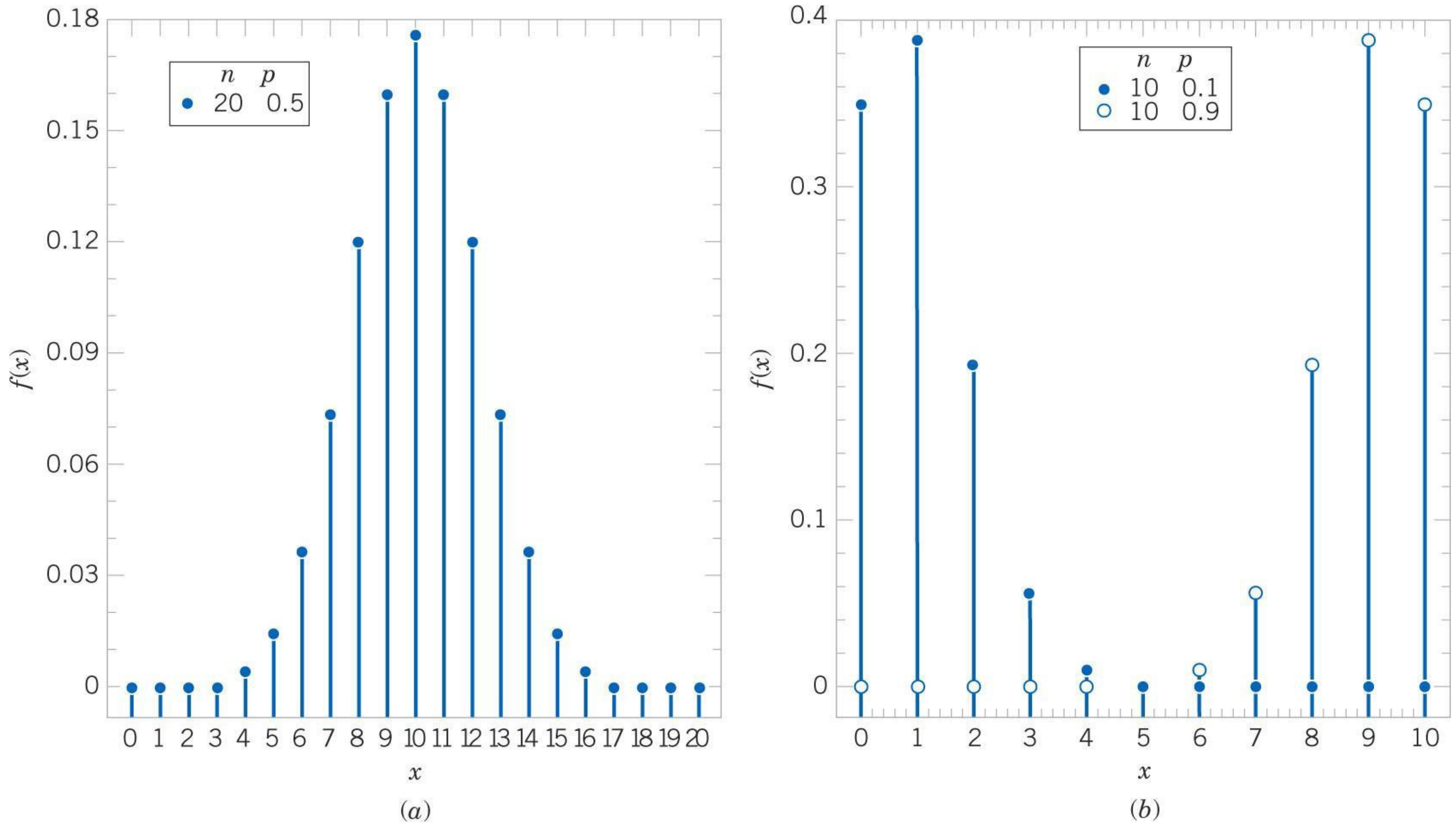


Figure 3-8 Binomial Distributions for selected values of *n* and *p*. Distribution (a) is symmetrical, while distributions (b) are skewed. The skew is right if *p* is small.

Examples:

1. The probability that a certain kind of component will survive a given shock test is ¾. Find the probability that exactly two of the next four components tested survive.

2. The drainage system of a city has been designed for a rainfall intensity that will be exceeded on an average once in 50 years. What is the probability that the city will be flooded at most 2 out of 10 years?

Negative Binomial Distribution

Gives the probability that it will take *exactly n trials* to produce *exactly x successes.*

A. Requirements

The last (*n*th) trial must be a success, otherwise the *x*th success actually occurred on an earlier trial. This means that we must have *(x – 1)* successes in the first *(n – 1)* trials plus success on the *n*th trial.

B. Mathematical Relationships

1. General Equation: $$p(n) = \binom{n-1}{x-1} p^x q^{n-x}$$

Our textbook: $f(x) = C_{r-1}^{x-1} p^r (1-p)^{x-r}$ for $x = r, r+1, r+2 \ldots$ (3-11)

Negative Binomial Distribution

2. Expectation – the expected number of trials to produce *x* successes

$$E(N) = \frac{x}{p}$$

3. Variance – the expected sum of the squared deviations from the mean.

$$V(N) = \frac{xq}{p^2}$$

Negative Binomial Graphs

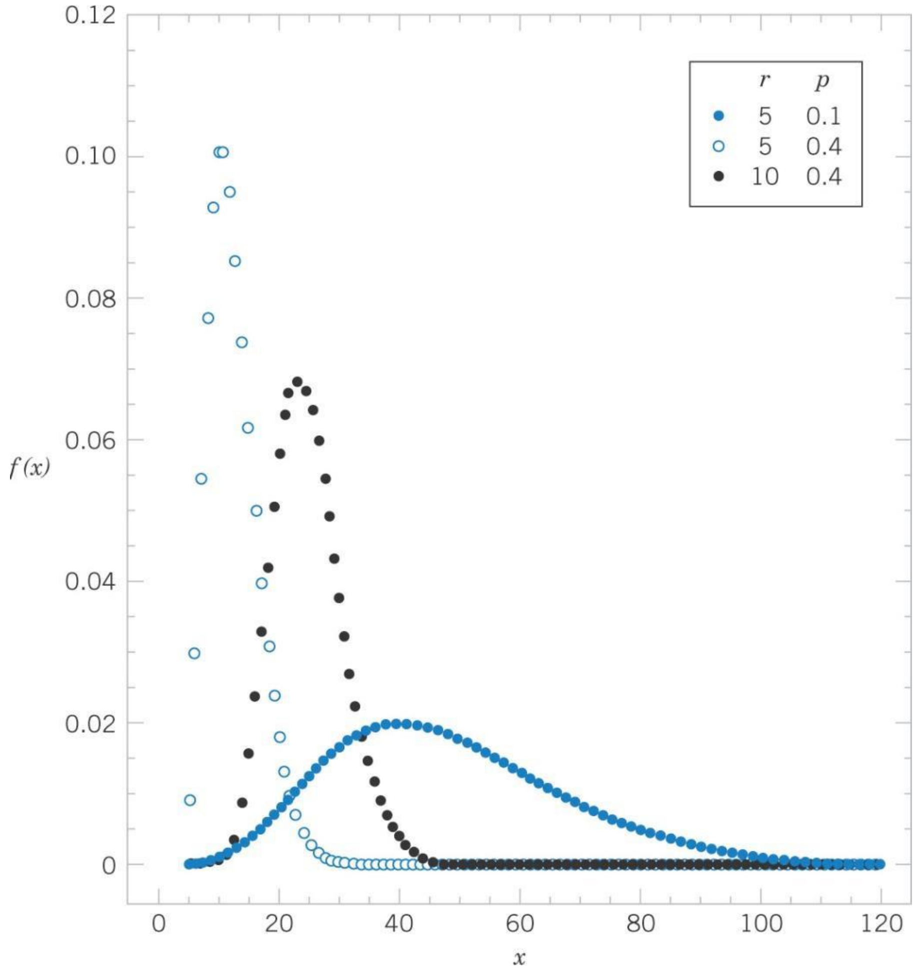


Figure 3-10 Negative binomial distributions for 3 different parameter combinations.

Example

Cotton linters used in the production of rocket propellant are subjected to a nitration process that enables the cotton fibers to go into solution. The process is 90% effective in that the material produced can be shaped as desired in a later processing stage with probability 0.9. What is the probability that exactly 20 lots will be produced in order to obtain the third defective lot?

Geometric Distribution

Gives the probability that it will take *exactly n trials* to produce the first success.

A. Requirements

The first success must occur on the nth trial, so we must have $(n-1)$ failures in the first $(n-1)$ trials plus success on the nth trial.

B. Mathematical Relationships

1. General Equation

Substituting ($x = 1$) into the Negative Binomial equation:

$$p(n) = \binom{n-1}{1-1} p^1 q^{n-1} \quad \Rightarrow \quad p(n) = \frac{(n-1)!}{0!(n-1)!}\, p\, q^{n-1}$$

or:

$$p(n) = pq^{n-1}$$

Geometric Distribution

2. Expectation – the expected number of trials to produce x successes

$$E(N) = \frac{1}{p}$$

3. Variance – the expected sum of the squared deviations from the mean.

$$V(N) = \frac{q}{p^2}$$

Example

In a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

Recap

- Binomial distribution:
  - Fixed number of trials (n).
  - Random number of successes (x).

- Negative binomial distribution:
  - Random number of trials (x).
  - Fixed number of successes (r).

- Because of the reversed roles, a negative binomial can be considered the opposite or negative of the binomial.

Hypergeometric Distribution

- This distribution is fundamentally different from binomial distributions. The trials for hypergeometric are NOT INDEPENDENT. (Items are selected without replacement). Useful for acceptance sampling, electronic testing, quality assurance.

- Given a set of N items, K of which have a trait of interest ("successes"), we wish to select a sample of n items without replacement from the set N. The variable x represents the number of successes in the n items.

Hypergeometric Distribution

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \text{ where } x = \max(0, n+K-N) \text{ to } \min(K, n) \qquad (3\text{-}13)$$

A set of *N* objects contains:

- *K* objects classified as success
- *N* - *K* objects classified as failures

A sample of size *n* objects is selected without replacement from the *N* objects, where:

- $K \le N$ *and* $n \le N$

Hypergeometric Distribution

- If X is a hypergeometric random variable with parameters N, K, and n, then

$$\mu = E(X) = np \quad \text{and} \quad \sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right) \qquad (3\text{-}14)$$

where $p = K/N$

and $\left(\frac{N-n}{N-1}\right)$ is the finite population correction factor.

σ^2 approaches the binomial variance as n/N becomes small.

Hypergeometric Graphs

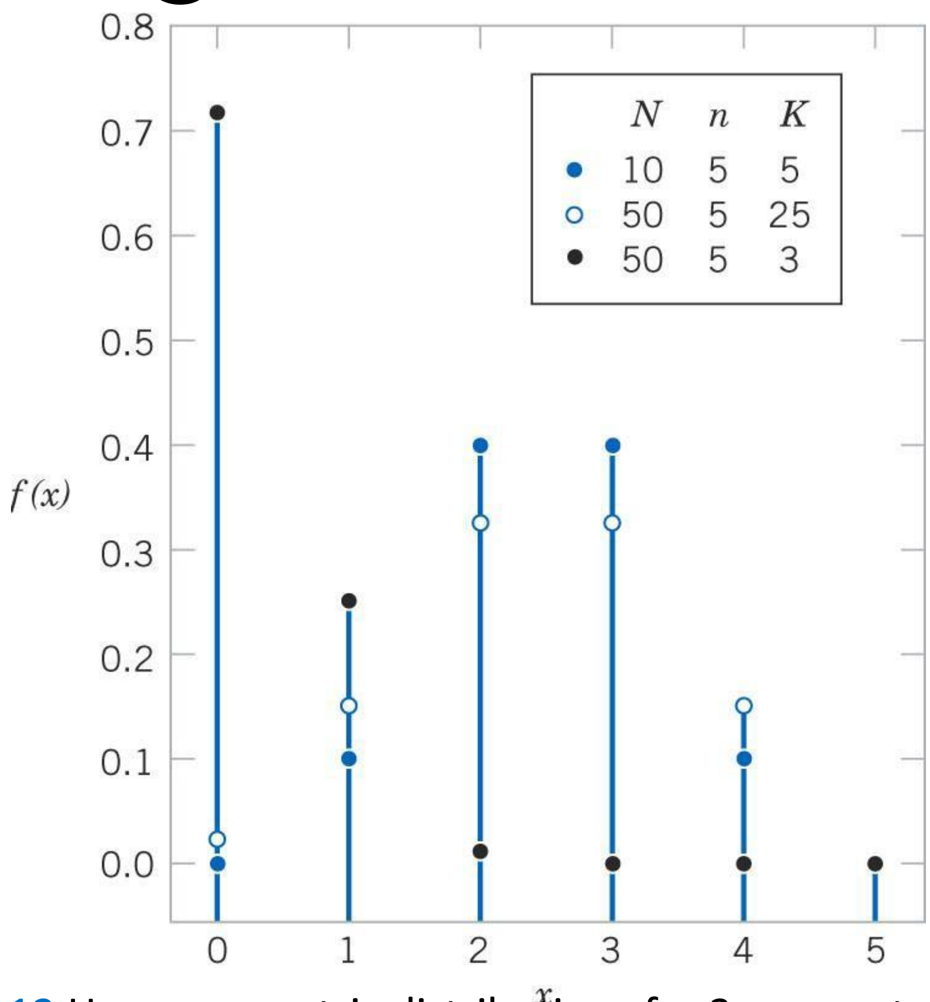
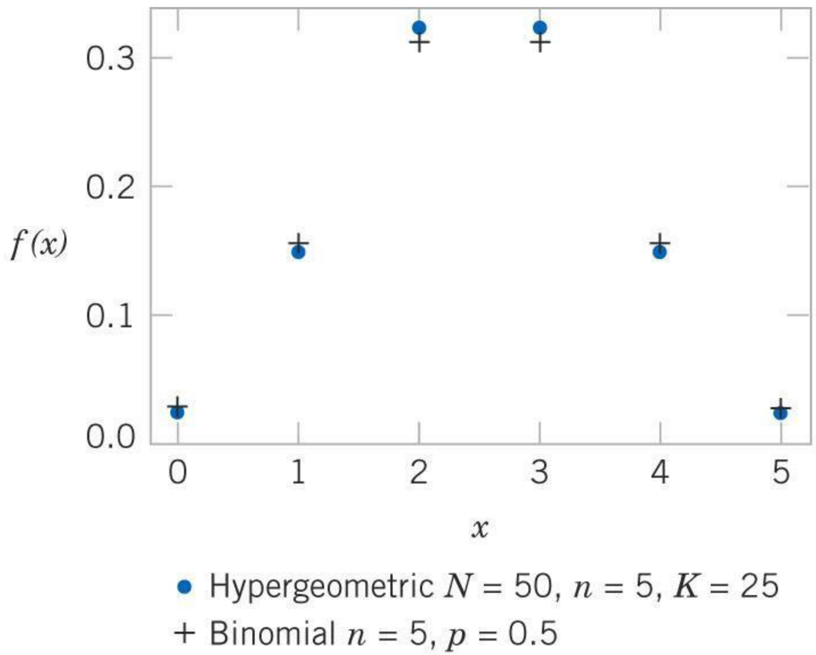


Figure 3-12 Hypergeometric distributions for 3 parameter sets of *N*, *K*, and *n*.

Hypergeometric & Binomial Graphs



	0	1	2	3	4	5
Hypergeometric probability Binomial probability	0.025 0.031	0.149 0.156	0.326 0.312	0.326 0.312	0.149 0.156	0.025 0.031

Figure 3-13 Comparison of hypergeometric and binomial distributions.

Example

- A lot o 1,000 items contains 100 defective and 900 non-defective. A sample of 20 items are taken in random without replacement. What is the probability that the sample contains two or fewer defective items.

Poisson Distribution

Gives the probability of *exactly* x *successes* over some *continuous interval* of time or space.

A. Mathematical Relationships

1. General Equation:

$$p(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

x = the number of successes in a (continuous) unit

λ = the average number of successes per unit

Note: The "unit" for both λ and x *must be the same*.

2. Expectation – the expected number of successes in a unit

$$E(X) = \lambda$$

3. Variance:

$$V(X) = \lambda$$

Poisson Graphs

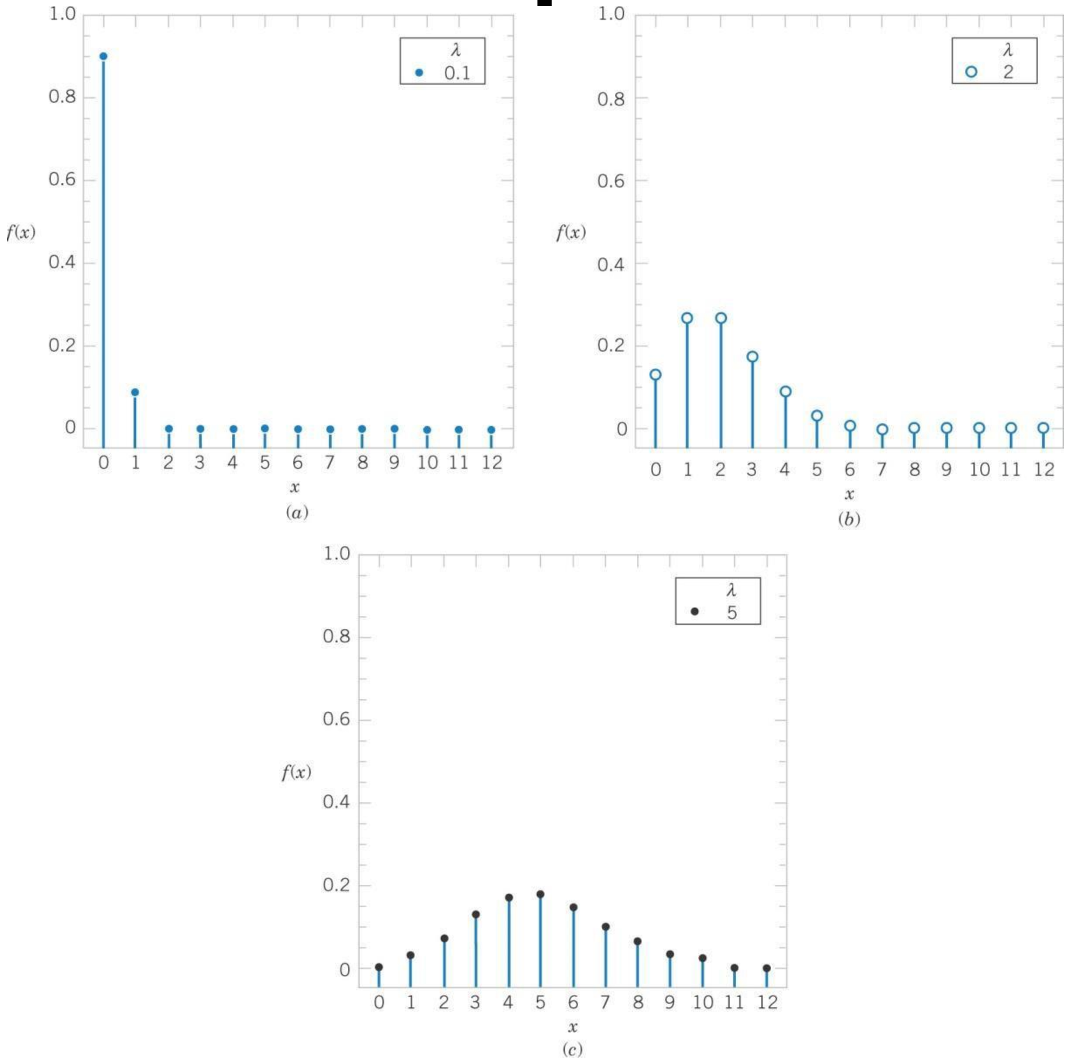


Figure 3-14 Poisson distributions for λ = 0.1, 2, 5.

Example

From records of the past 50 years, it is observed that tornadoes occur in a particular area an average of two times a year. What is the probability of no tornadoes in the next year? What is the probability of exactly 2 tornadoes next year? What is the probability of no tornadoes in the next 50 years?

The Binomial – Poisson Connection

The binomial distribution gives the probability of exactly x successes in n trials. If the probability of success (p) for any given trial is small and the number of trials (n) is large, the binomial distribution can be approximated by a Poisson distribution with $\lambda = np$.

Why? Intuitively, if p is small and n is large, the long strings of "failures" between the infrequent "successes" start to look like continuous intervals rather than discrete events.

Generally speaking, the Poisson distribution provides a pretty good approximation of the binomial distribution as long as $n > 20$ and $np < 5$.