Probability Density

What is Density?

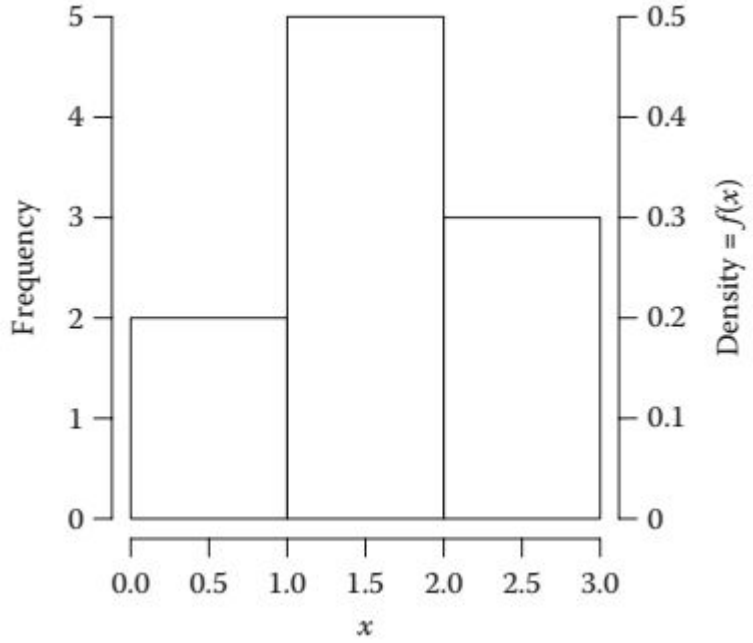
Density is the output generated by the pdf function

A pdf function is the result of the function when x varies.

A pdf is best illustrated using a histogram.

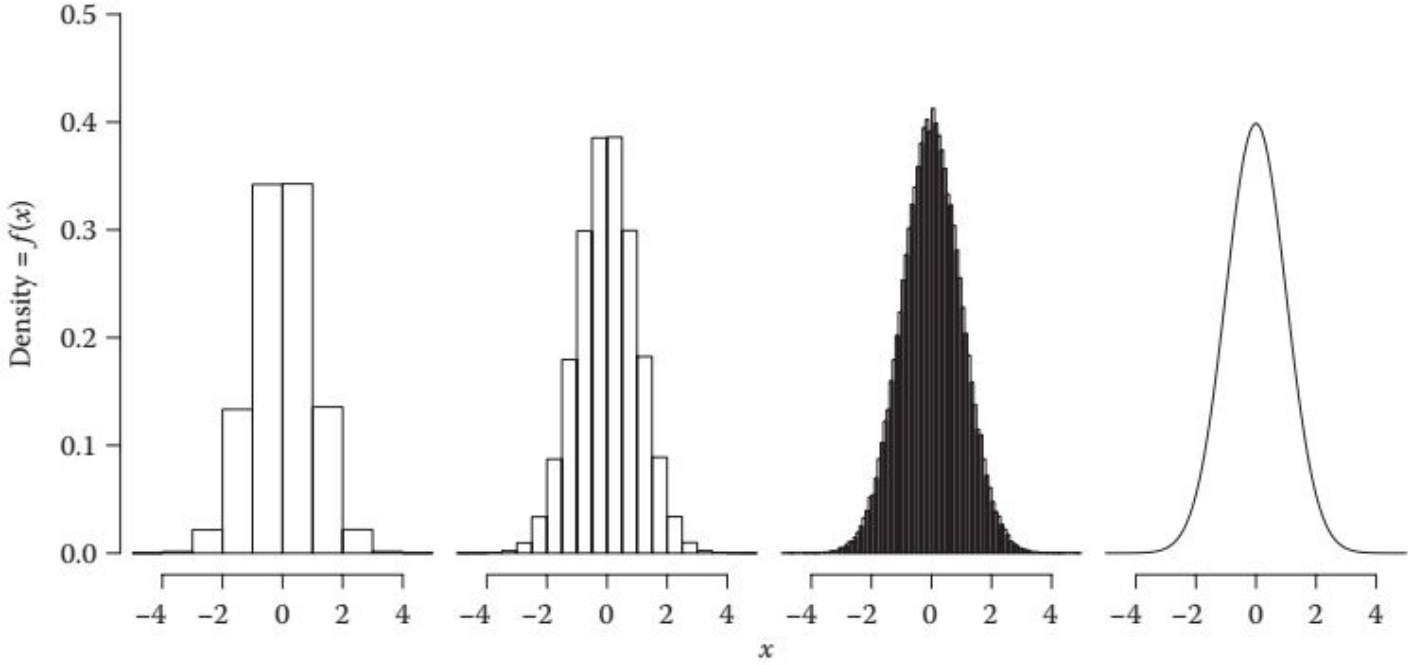
If the numerical value is discrete, the height of the bar at a given values of x, is the probability of x.

If the numerical value is continuous, this not possible.



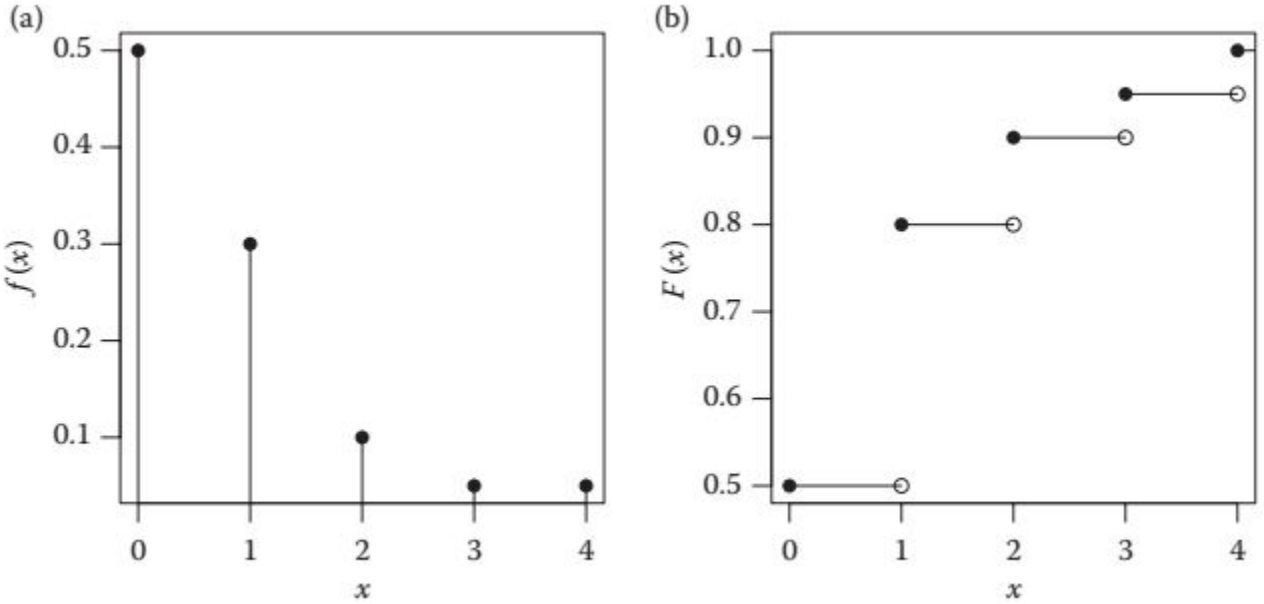
Density of Continuous values

As we shrink the size of the bins, the discrete values will be more continuous.



Cumulative Density Function (CDF)

The cdf is simply the sum of the density upto x.



Bernoulli Distribution

- Probability of success for a single binary trial (the output is binary)
- π = probability of success, x = number of times you want x to occur, in this case 1 or 0.

$$f(x) = \pi^x (1-\pi)^{1-x}$$

For a coin toss π is 0.5, for a dice roll, π is 0.167

Binomial Distribution

- Is a combination of Bernoulli trials that are independent and identically distributed

$$f(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x},$$

R commands for statistical distributions

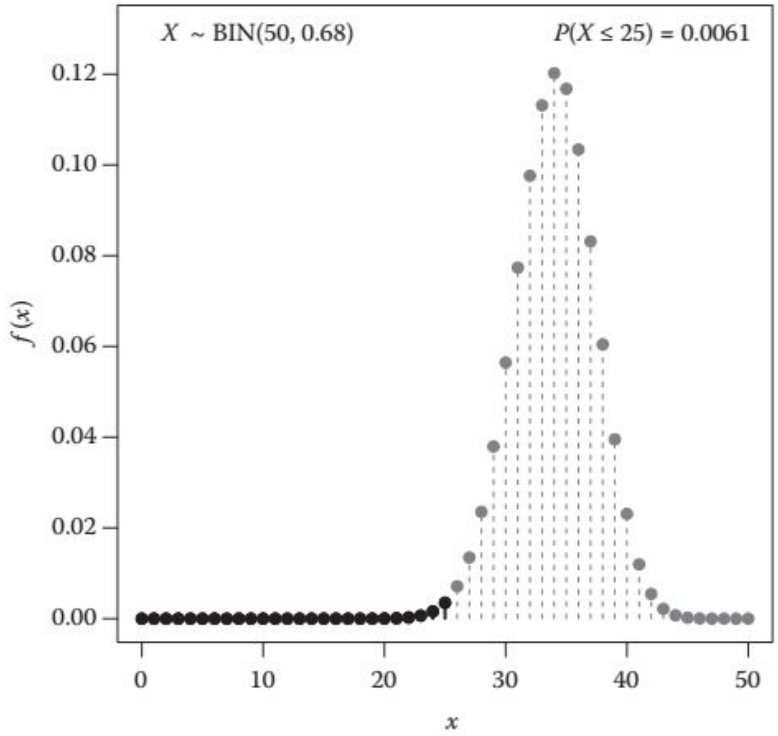
rbinom - generate random numbers from a binomial distribution

dbinom - density (height) of distribution at x

pbinom - cumulative density of distribution at x

qbinom - quantile of x in the defined distribution

X ~ BIN(50,0.68)


X ~ BIN(50, 0.68)
P(X ≤ 25) = 0.0061
0.12
0.10
0.08
0.06
0.04
0.02
0.00
f(x)
0
10
20
30
40
50
x

Exercise

During the industrial revolution in England, London become covered in a lot of black soot from burning coal.

Five years beforehand, the proportion of white moths that could be found in London was 87%. However black moths gained a survival advantage as the air became more polluted.

If you were to sample moths in London 25 years into the industrial revolution, and you found that 35 out of 50 moths were white, how likely would that be if the population had remained the same?

You can compute the probability of finding exactly 35/50

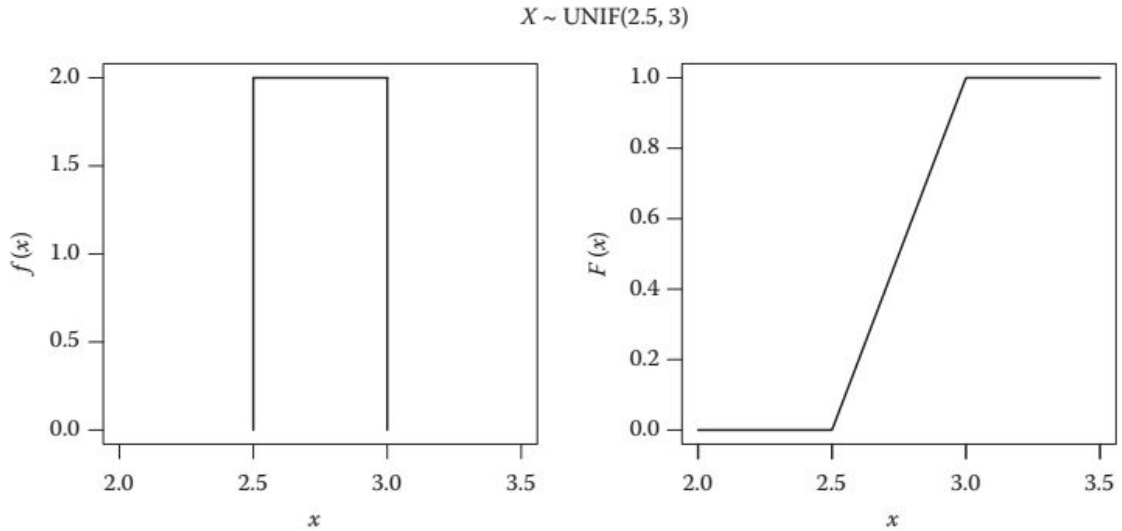
First write the equation and then use the r functions to calculate the probability.

Common continuous pdf - Uniform distribution

The probability is the same no matter what value.

a is the lower limit and b is the upper.

$$f(x) = \frac{1}{b-a},$$



Common continuous pdf - Normal distribution

The expected value is the average

The function is symmetric around the mean

Produces output that is continuous and independen

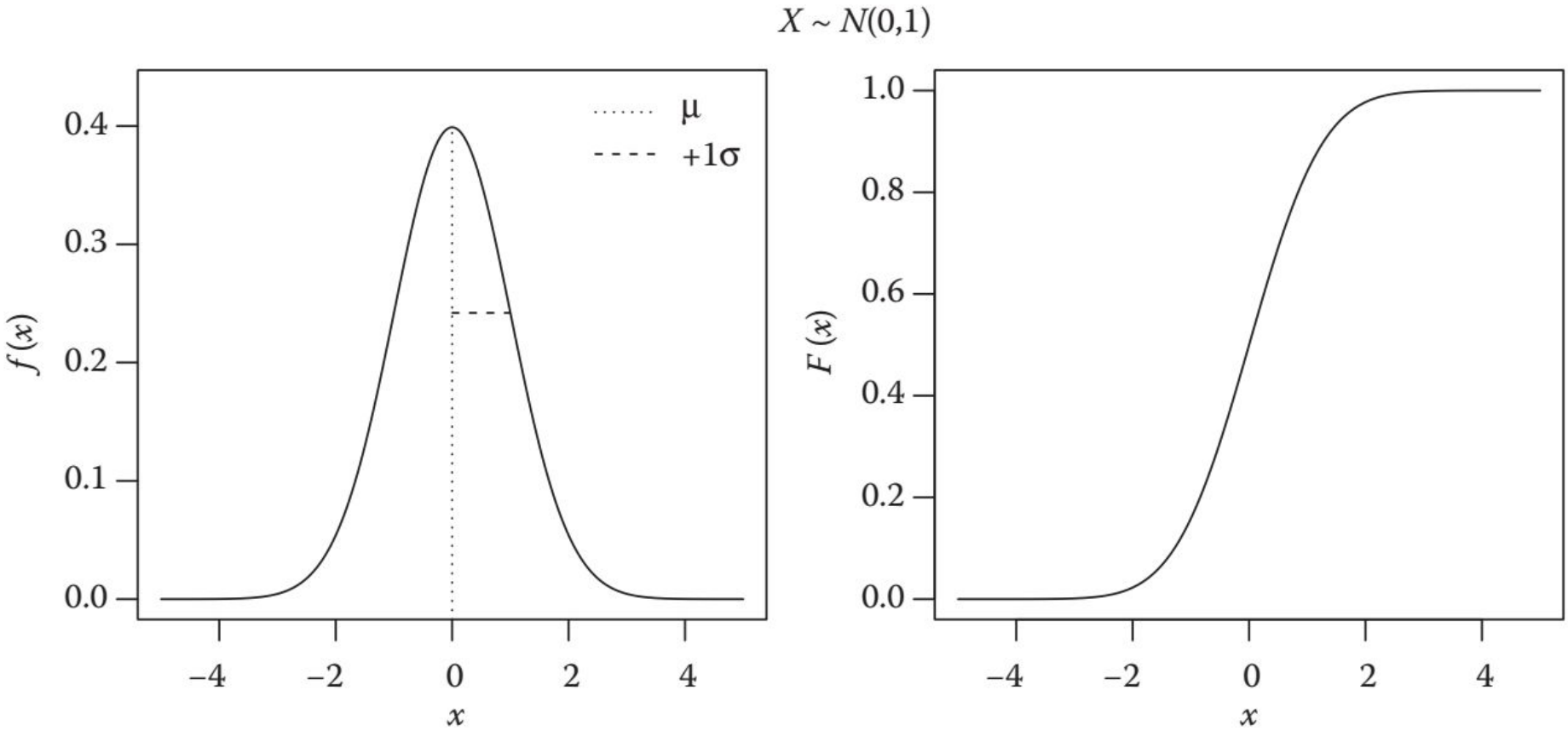
Two parameters:

μ = mean

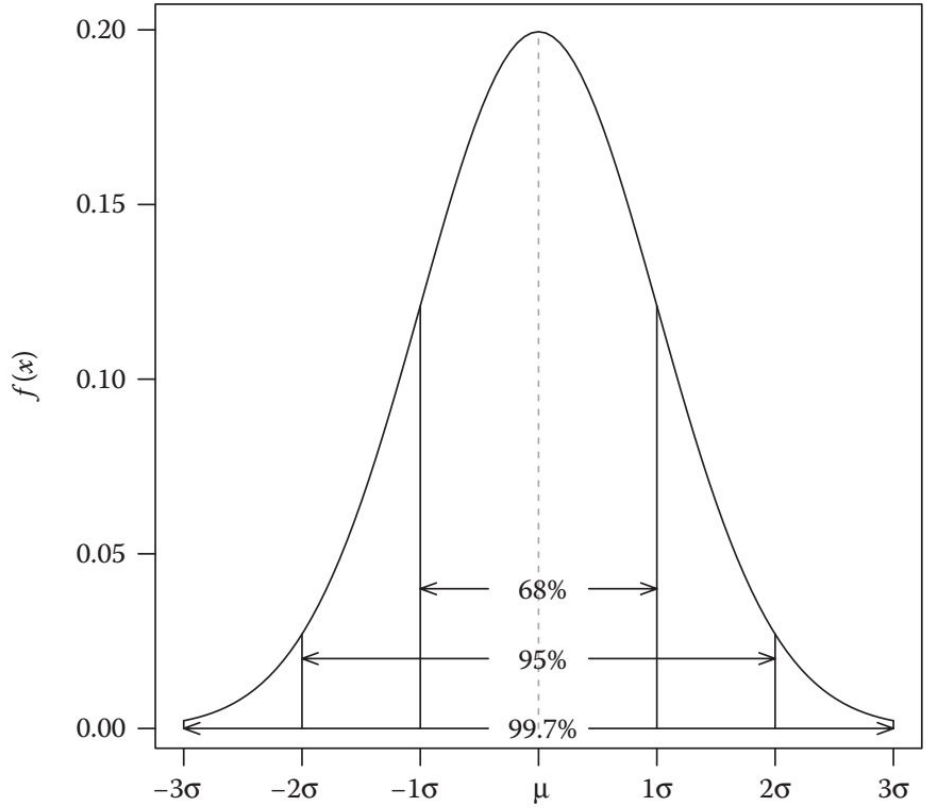
σ = standard deviation (square root of variance)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Standard normal pdf (mean=0, sd = 1) (Z-distribution)



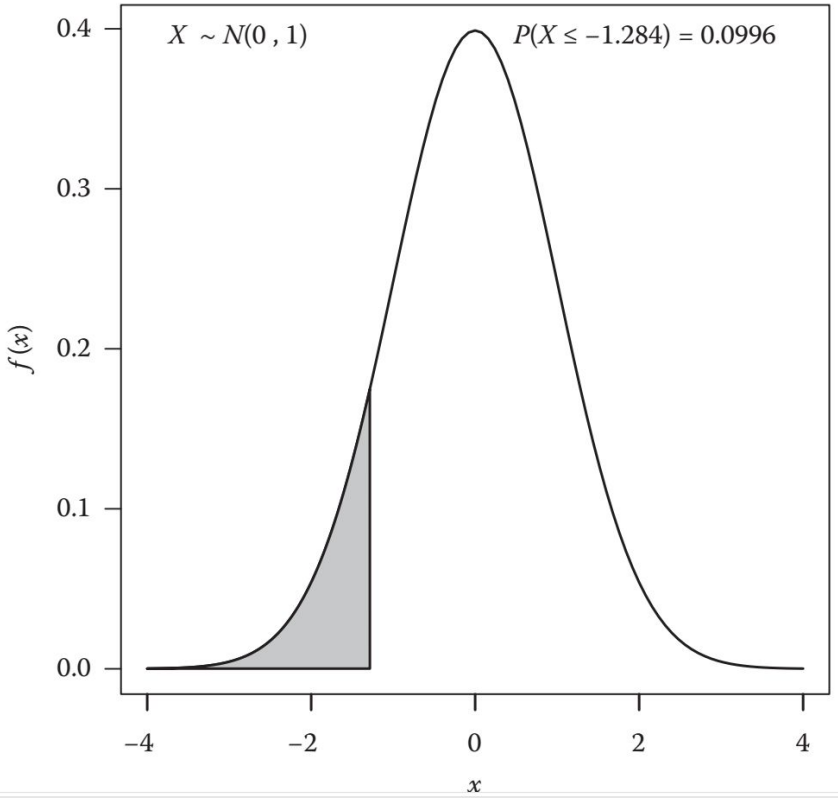
Normal distribution properties.



$$Z = \frac{X - \mu}{\sigma}$$

Normal Distribution Example

In a summary for the center for disease control, McDowell (2005) reported that the height of non-Hispanic white women in the United States over 20 years of age was approximately normally distributed with a mean of 162.9 cm and a standard deviation of 7.71 cm. Given this, what proportion of these women are ≤153 cm tall?



The average weight of an adult female greyhound is 63 pounds, with a standard deviation of 8 pounds. What proportion of female greyhounds weigh less than or equal to 55 pounds?

What proportion weigh between 60 and 65 pounds?

It can safely be said that 75% weigh no more than what amount?

Discrete Distributions - Poisson

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

Similar to binomial distribution

Counts the number of successes in a trial

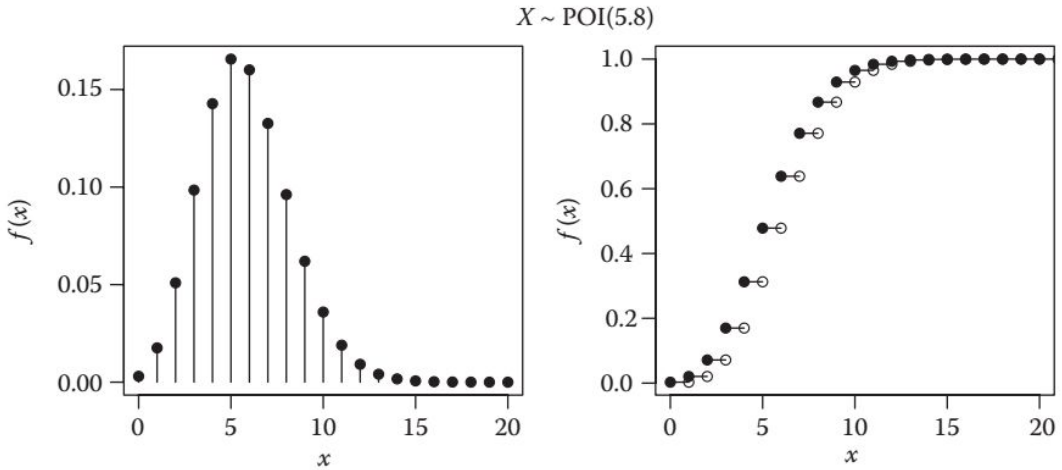
But no limit on number of successes since it is over time or space.

If n is large and p is low,

approximates binomial.

One parameter λ

which is mean and variance



Hypergeometric Distribution

$$f(x) = \binom{M}{x}\binom{N-M}{n-x} \Big/ \binom{N}{n}$$

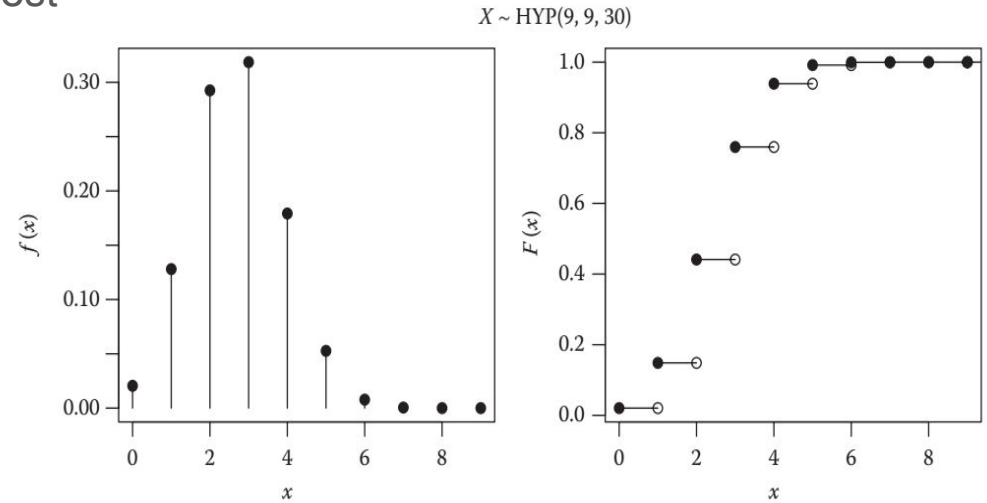
Probability of x successes over n trials without replacement.

Binomial is with replacement.

n is the number of trials

M is the number of items of group of interest

N is the number of items selected



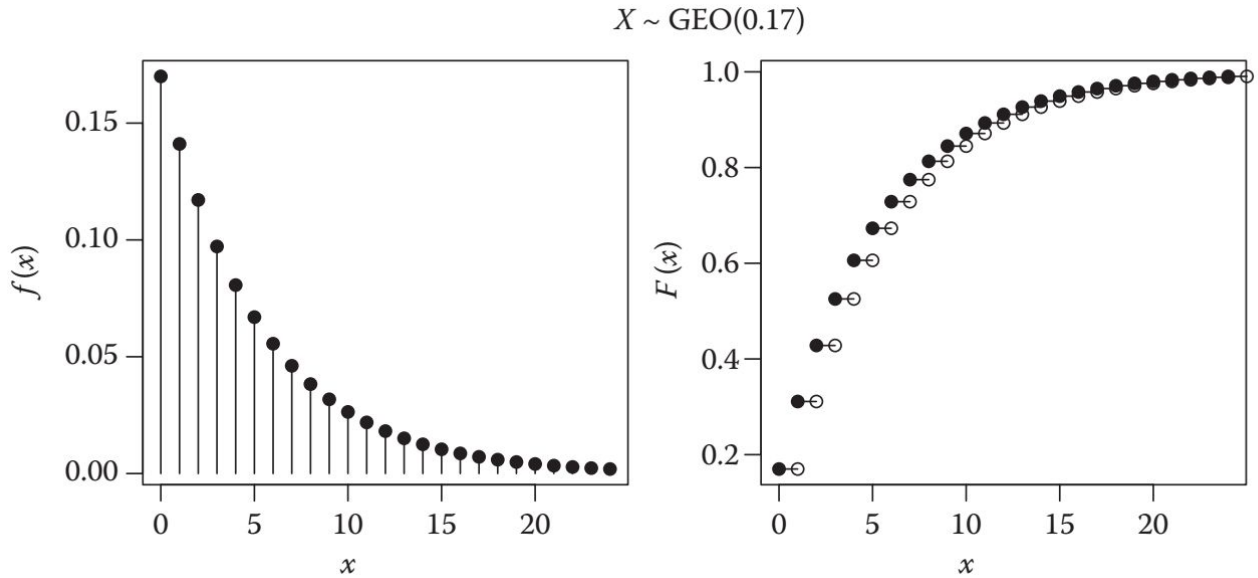
Hypergeometric example

The Greater Yellowstone Area (GYA) is a contiguous set of natural ecosystems located in the North-Central Rocky Mountains that comprise an area of 80,000 km2 . Garrott et al. (2011) compiled data collected from 18 GYA mountain goat herds over 35 years. Using this information, Flesch and Garrott (2011) found that the long-term kid:adult ratio of the herds was approximately $M/(N-M) = 0.27$, while the average population size was 1927 individuals. Recently, 88 goats were randomly sampled from the GYA without replacement (sampled goats were kept in a holding pen by a survey team) and 32 were found to be kids. Assuming the (unlikely) independence of kid and adult captures, what is the probability of this outcome?

Geometric distribution

Number of failure in Bernoulli trial before success.

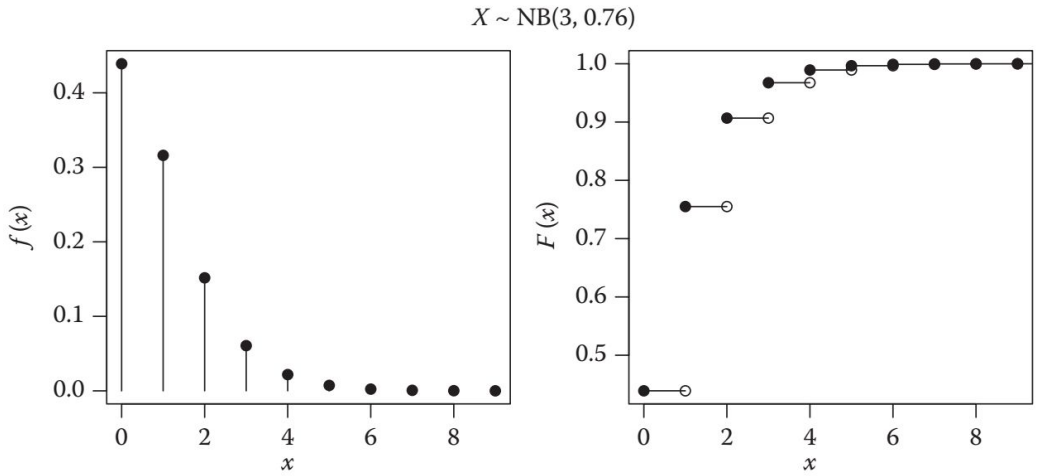
$$f(x) = \pi(1-\pi)^x$$



Negative Binomial Distribution

for RNA seq

Probability of x Bernoulli failures before the rth success.



$$f(x) = \binom{x+r-1}{r-1} \pi^r (1-\pi)^x$$

Negative Binomial Example

The successful hatching probability for glaucous-winged gull (Larus glaucescens) eggs in Glacier Bay Alaska is about 0.76 (Zador and Piatt 2007; see Figure 3.15). Assuming the independence of hatches, what is the probability that a gull nest site will fail once before successfully hatching three eggs?