

© Copyright 2004-2019 by Jenny A. Baglivo. All Rights Reserved.

2	MATH442601 2 Notebook 2	3
2.1	Introduction	3
2.1.1	Definitions	3
2.2	Discrete Random Variables	4
2.2.1	Probability Distribution, PMF, CDF	4
2.2.2	Discrete Uniform Distribution	6
2.2.3	Hypergeometric Distribution	7
2.2.4	Bernoulli Experiments, Bernoulli and Binomial Distributions	8
2.2.5	Simple Random Samples, Binomial Approximation, Survey Analysis	9
2.2.6	Geometric and Negative Binomial Distributions	10
2.2.7	Poisson Limit Theorem, Poisson Distribution	13
2.3	Continuous Random Variables	16
2.3.1	Probability Distribution, PDF and CDF	16
2.3.2	Quantiles, Percentiles	19
2.3.3	Continuous Uniform Distribution	22
2.3.4	Exponential Distribution	24
2.3.5	Euler Gamma Function, Gamma Distribution	27
2.3.6	Poisson Processes, Revisited	28
2.3.7	Normal (Gaussian) Distribution	32
2.3.8	Cauchy Distribution	37
2.3.9	Transforming Random Variables, CDF Method	39

List of Tables

1	Standard normal cumulative probabilities, $\Phi(z) = P(Z \leq z)$, when $z \geq 0$	34
---	---	----

2 MATH442601|2 Notebook 2

This notebook is concerned with discrete and continuous random variables. The notes correspond to material covered in Chapter 2 of the Rice textbook.

2.1 Introduction

Researchers use random variables to describe the numerical results of experiments.

2.1.1 Definitions

1. A *random variable* is a function from the sample space of an experiment to the real numbers. The *range* of a random variable is the set of values the random variable assumes. Random variables are usually denoted by capital letters (X, Y, Z, \dots) and their values by lower case letters (x, y, z, \dots).
2. If the range of a random variable is a finite or countably infinite set, then the random variable is said to be *discrete*; if the range is an interval or a union of intervals, then the random variable is said to be *continuous*; otherwise, the random variable is said to be of *mixed* type.

Example 1. You toss a fair coin eight times and record the sequence of heads and tails. Let X be the number of heads in the sequence, and Y be the number of heads minus the number of tails. Then

- X is a discrete random variable with range $\mathcal{R}_X = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$, and
- Y is a discrete random variable with range $\mathcal{R}_Y = \{-8, -6, -4, -2, 0, 2, 4, 6, 8\}$.

Y can be written as a linear function of X as follows (please complete):

Example 2. Let X be the height of an individual measured in inches with infinite precision. Then X is a continuous random variable whose range is the positive real numbers.

If Y is the height measured in feet with infinite precision, then Y is a continuous random variable whose range is the positive real numbers, and $Y = \frac{1}{12}X$.

2.2 Discrete Random Variables

A discrete random variable is one whose range is a finite set or a countably infinite set.

2.2.1 Probability Distribution, PMF, CDF

If X is discrete, then $P(X = x)$ is the probability that an outcome has value x .

The *probability distribution* of X is the collection of probabilities $P(X = x)$, for all $x \in \mathbf{R}$.

“ $X = x$ ” is a convenient shorthand for the event “the random variable X takes value x .”
Similarly, “ $X \leq x$ ” is a convenient shorthand for the event “the random variable X takes a value that is at most x ,” the notation “ $a < X < b$ ” is a convenient shorthand for the event “the random variable X takes a value strictly between a and b ,” and so forth.

Probability mass function. If X is discrete, then the *probability mass function* (PMF) (also called the *frequency function* (FF)) of X is defined as follows:

$$p(x) = P(X = x) \text{ for all real numbers } x.$$

Probability density functions satisfy the following properties:

1. **Nonnegative:** $p(x) \geq 0$ for all real numbers x .
2. **Unit Sum:** $\sum_{x \in \mathcal{R}} p(x) = 1$, where \mathcal{R} is the range of X .

Note that the events “ $X = x$ ” for $x \in \mathcal{R}$ are pairwise disjoint sets whose union is the sample space of the experiment. Events must have nonnegative probability (property 1), and the sum of their probabilities must be the probability of the sample space (property 2).

Exercise. Let X be the number of heads obtained in eight tosses of a fair coin.

Find a general formula for the probability mass function of X . Be sure to specify the value of your function for each real number.

Cumulative distribution function. The *cumulative distribution function* (CDF) of the discrete random variable X is defined as follows:

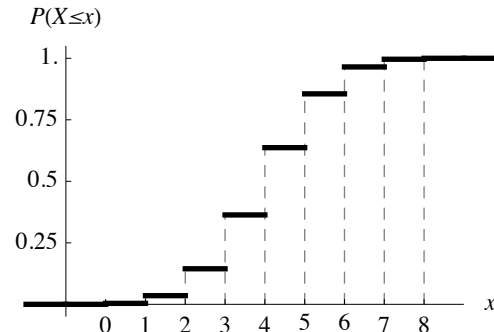
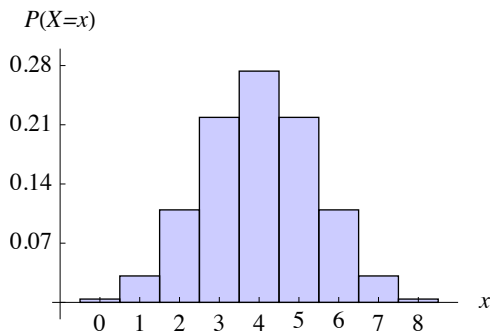
$$F(x) = P(X \leq x) \text{ for all real numbers } x.$$

Cumulative distribution functions satisfy the following properties:

1. Infinite Limits: $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$.
2. Nondecreasing: If $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
3. Right Continuous: $\lim_{x \rightarrow a^+} F(x) = F(a)$ for each real number a .

Note that $F(x)$ represents cumulative probability, with limits 0 and 1 (property 1). Cumulative probability increases with increasing x (property 2), and has discrete jumps at values of x in the range of the random variable (property 3).

Example 1, continued. Let X be the number of heads in eight tosses of a fair coin.



1. Left Plot: The left plot is a *probability histogram* of the PMF of X .

A probability histogram is constructed as follows:

For each x in the range of X , a rectangle with base the interval $[x - 0.50, x + 0.50]$ and with height $p(x)$ is drawn. The total area represented in the plot is 1.

2. Right Plot: The right plot shows the CDF of X .

$F(x)$ is a step function, with steps of height $p(x)$ for each x in the range.

Since the range of X is the finite set $\{0, 1, 2, \dots, 8\}$,

$$F(x) = 0 \text{ when } x < 0 \text{ and } F(x) = 1 \text{ when } x \geq 8.$$

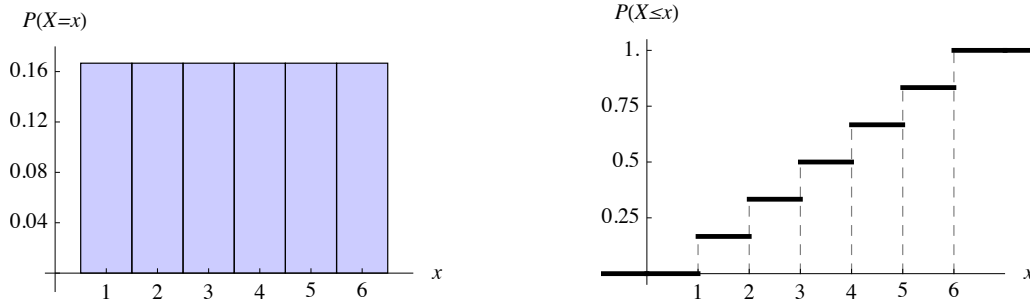
2.2.2 Discrete Uniform Distribution

Let n be a positive integer. The random variable X is said to be a *discrete uniform random variable*, or to have a *discrete uniform distribution*, with parameter n when its probability mass function is as follows:

$$p(x) = \frac{1}{n} \quad \text{when } x \in \{1, 2, \dots, n\}; \text{ and } 0 \text{ otherwise.}$$

For example, suppose that you roll a fair six-sided die and let X be the number of dots on the top face. Then X is a discrete uniform random variable with $n = 6$.

The PDF and CDF of X are displayed below:



Example (Hand et al, Chapman & Hall, 1994, page 98). Is a fair die really fair?

That's the question that R. Wolf tried to answer when he rolled a "fair" die 20,000 times and recorded the number of dots on the top face each time. The following table summarizes the results of his experiment, where the expected frequency is

$$\text{Expected Frequency} = 20000 p(x) = 20000 \left(\frac{1}{6}\right) \approx 3333.33 \quad \text{for each } x,$$

and the relative error is the following ratio:

$$\text{Relative Error} = \frac{\text{Observed Frequency} - \text{Expected Frequency}}{\text{Expected Frequency}}.$$

Note that the relative error has been converted to a percentage.

	$x = 1$	$x = 2$	$x = 3$	$x = 4$	$x = 5$	$x = 6$
Observed Frequency	3407	3631	3176	2916	3448	3422
Expected Frequency	3333.33	3333.33	3333.33	3333.33	3333.33	3333.33
Relative Error (%)	2.21%	8.93%	-4.72%	-12.52%	3.44%	2.66%

The rather large relative errors call the fairness of the die into question.

2.2.3 Hypergeometric Distribution

Let n , M , and N be integers with $0 < M < N$ and $0 < n < N$. The random variable X is said to be a *hypergeometric random variable*, or to have a *hypergeometric distribution*, with parameters n , M , and N , when its probability mass function is as follows:

$$p(x) = P(X = x) = \frac{\binom{M}{x} \times \binom{N-M}{n-x}}{\binom{N}{n}},$$

when x is an integer between $\max(0, n + M - N)$ and $\min(n, M)$, and 0 otherwise.

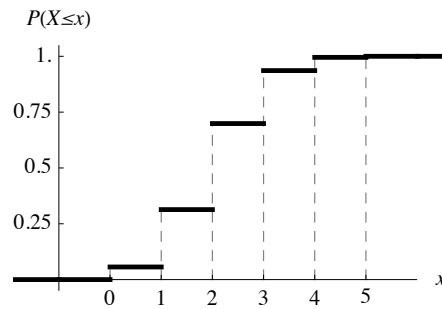
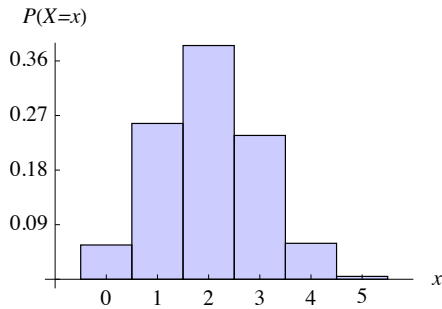
Hypergeometric distributions are used to model *urn experiments*, where N is the number of objects in the urn, M is the number of “special” objects, n is the size of the subset chosen from the urn and X is the number of special objects in the chosen subset. If the choice of each subset is equally likely, then X has a hypergeometric distribution.

Example. A group of twenty-five first graders is to be randomly assigned to two classes: a class of 10 students to be taught by Mrs Smith and a class of 15 students to be taught by Mr Jones. Assume that each choice of rosters is equally likely.

Five of the first graders are close friends. Let X be the number of close friends assigned to Mrs Smith’s class. Then X has a hypergeometric distribution, where

$$n = \underline{\hspace{2cm}}, M = \underline{\hspace{2cm}}, \text{ and } N = \underline{\hspace{2cm}}.$$

The PMF and CDF of X are displayed below:



- The probability that all 5 friends are assigned to the same class is

- The probability that exactly 4 of the 5 friends are assigned to the same class is

2.2.4 Bernoulli Experiments, Bernoulli and Binomial Distributions

A *Bernoulli experiment* is a random experiment with two possible outcomes. The outcome of chief interest is often called “success” and the other outcome “failure.” Let

$$p = P(\text{Success}) \quad \text{and} \quad 1 - p = P(\text{Failure}).$$

Let X be the number of successes in n repeated trials of a Bernoulli experiment with success probability p . Then X is said to be a *binomial random variable*, or to have a *binomial distribution*, with parameters n and p . The probability mass function of X is as follows:

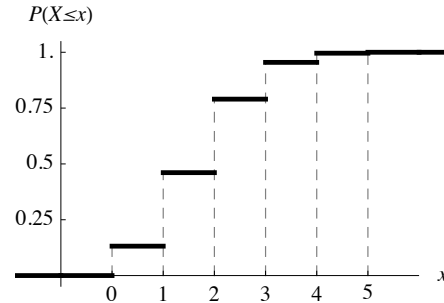
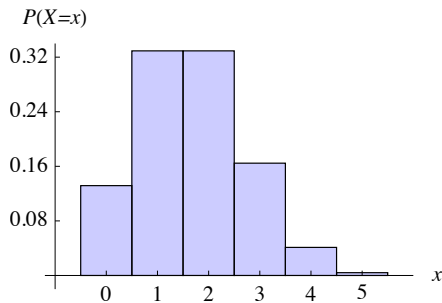
$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad \text{when } x \in \{0, 1, 2, \dots, n\}; \text{ and } 0 \text{ otherwise.}$$

A *Bernoulli distribution* is the special case when $n = 1$.

Example. You roll a fair six-sided die five times and observe the number of dots on the top face each time. Let X be the number of times you observe either a 1 or a 4 in the five trials. Then X is a binomial random variable with parameters

$$n = \underline{\hspace{2cm}}, \text{ and } p = \underline{\hspace{2cm}}.$$

The PMF and CDF of X are displayed below:



- The probability of observing at least 3 successes in the 5 trials is

- The probability of observing at least one success and at least one failure in the 5 trials is

2.2.5 Simple Random Samples, Binomial Approximation, Survey Analysis

Suppose that an urn contains N objects.

A *simple random sample* of size n is a sequence of n objects chosen *without* replacement from the urn, where each choice of sequence is equally likely.

Let M be the number of special objects in the urn and X be the number of special objects in a simple random sample of size n . Then X has a hypergeometric distribution with parameters n, M, N . Further, if N is very large, then binomial probabilities can be used to approximate hypergeometric probabilities.

Theorem (Binomial Approximation). If N is large and $p = M/N$ is not extreme, then the binomial distribution with parameters n and $p = M/N$ can be used to approximate the hypergeometric distribution with parameters n, M, N . Specifically,

$$P(x \text{ special objects}) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \approx \binom{n}{x} p^x (1-p)^{n-x} \text{ for each } x.$$

To see this,

1. First note that
$$\frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = \binom{n}{x} \frac{M P_x N-M P_{n-x}}{N P_n}.$$

(Write both sides as ratios of factorials and simplify; see notebook 1 for more details.)

2. Further, note that, when $n < 0.01N$, and $0.05 < \frac{M}{N} < 0.95$,

$$\frac{M P_x N-M P_{n-x}}{N P_n} \approx \frac{M^x (N-M)^{n-x}}{N^n}.$$

Since last expression equals $(M/N)^x ((N-M)/N)^{n-x} = p^x (1-p)^{n-x}$, the result follows.

Sampling with/without replacement. If X is the number of special objects in a sequence of n objects chosen *with* replacement from the urn and if the choice of each sequence is equally likely, then X has a binomial distribution with parameters n and $p = M/N$.

The theorem says that, under certain conditions, the model where sampling is done *with* replacement can be used to approximate the model where sampling is done *without* replacement.

Survey analysis. Simple random samples are used in surveys. If the survey population is small, then hypergeometric distributions are used to analyze the results. If the survey population is large, then binomial distributions are used to analyze the results, even though each person’s opinion is solicited at most once.

For example, suppose that a surveyor is interested in determining the level of support for a proposal to change the local tax structure, and decides to choose a simple random sample of size 10 from the registered voter list.

- If there are a total of 120 registered voters, one-third of whom support the proposal, then the probability that exactly 3 of the 10 chosen voters support the proposal is

$$P(X = 3) = \frac{\binom{40}{3} \binom{80}{7}}{\binom{120}{10}} \approx 0.27.$$

- If there are thousands of registered voters, one-third of whom support the proposal, then the probability is

$$P(X = 3) \approx \binom{10}{3} \left(\frac{1}{3}\right)^3 \left(\frac{2}{3}\right)^7 \approx 0.26.$$

Note that you do not need to know the exact number of registered voters when using the “large sample” approximation of the probability.

2.2.6 Geometric and Negative Binomial Distributions

Let X be the trial number of the r^{th} success in a sequence of mutually independent Bernoulli trials with success probability p . Then X is said to be a *negative binomial random variable*, or to have a *negative binomial distribution*, with parameters r and p .

The probability mass function of X is as follows:

$$p(x) = \binom{x-1}{r-1} (1-p)^{x-r} p^r \quad \text{when } x \in \{r, r+1, r+2, \dots\}; \text{ and } 0 \text{ otherwise.}$$

Note that, for each x , $p(x)$ is the probability of the event

“Exactly $x - r$ failures and r successes in x trials, with the last trial a success.”

The *geometric distribution* corresponds to the special case where $r = 1$. If X has a geometric distribution, then its probability mass function is as follows:

$$p(x) = (1-p)^{x-1} p, \quad \text{when } x \in \{1, 2, 3, \dots\}; \text{ and } 0 \text{ otherwise.}$$

Exercise. Let X be the trial number of the first success in a sequence of independent Bernoulli trials with success probability p . Find simplified formulas for

- (a) the cumulative probability $P(X \leq x)$, where x is a nonnegative integer, and
- (b) the upper tail probability $P(X > x)$, where x is a nonnegative integer.

Recall:

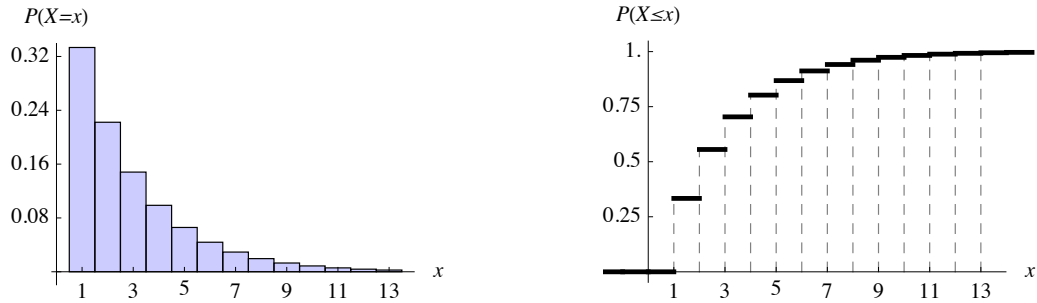
$$(1) \sum_{i=0}^{n-1} ar^i = a + ar + ar^2 + \cdots + ar^{n-1} = a \left(\frac{1-r^n}{1-r} \right) \text{ when } r \neq 1.$$

$$(2) \sum_{i=0}^{\infty} ar^i = \lim_{n \rightarrow \infty} \left(\sum_{i=0}^{n-1} ar^i \right) = \frac{a}{1-r} \text{ when } |r| < 1.$$

Exercise. A Bernoulli experiment consists of rolling a fair six-sided die and recording an S (for success) if 1 or 4 dots appear on the top face, and an F (for failure) otherwise.

- (a) Let X be the trial number of the first success in a sequence of independent trials of the experiment. Then X has a geometric distribution with parameter $p = \underline{\hspace{2cm}}$.

The PMF and CDF of X are displayed below:

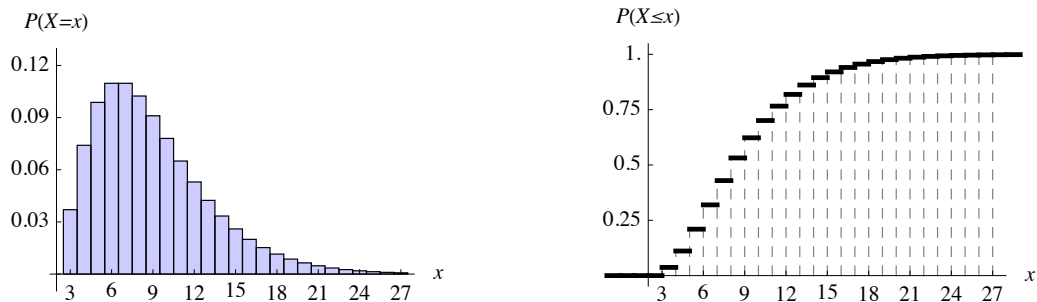


- The probability of observing 5 or more failures before the first success is

- (b) Let X be the trial number of the third success in a sequence of independent trials of the experiment. Then X has a negative binomial distribution with parameters

$$r = \underline{\hspace{2cm}} \text{ and } p = \underline{\hspace{2cm}}.$$

The PMF and CDF of X are displayed below:



- The probability of observing 4 or fewer failures before the third success is

2.2.7 Poisson Limit Theorem, Poisson Distribution

The following limit theorem, proven by S.Poisson in the 1830's, can be used to estimate binomial probabilities when the number of trials is large and the probability of success is small.

Theorem (Poisson Limit Theorem). Let λ ("lambda") be a fixed positive real number, let n be an integer greater than λ , and let $p = \frac{\lambda}{n}$. Then

$$\lim_{n \rightarrow \infty} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{when } x = 0, 1, 2, \dots$$

To see this, let $p = \lambda/n$. Then

$$\binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \left(\frac{\lambda^x}{x!}\right) \binom{n}{x} \left(\frac{1 - (\lambda/n)}{1 - (\lambda/n)}\right)^{n-x}.$$

(Please complete this proof. Note that $\lim_{n \rightarrow \infty} \left(1 + \frac{r}{n}\right)^n = e^r$ for any real number r .)

Poisson approximation. The theorem tells us that if n is large and $\lambda = np$ is moderate, then the binomial probability can be approximated as follows:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \approx e^{-\lambda} \frac{\lambda^x}{x!}.$$

Note that $\lambda = np$ represents the average number of successes in n independent trials.

For example, if $n = 10000$, $p = \frac{1}{2500}$, $x = 3$ (and $\lambda = np = 4$), then

$$P(X = 3) = \binom{10000}{3} \left(\frac{1}{2500}\right)^3 \left(\frac{2499}{2500}\right)^{9997} \approx 0.195386 \quad \text{and} \quad e^{-4} \frac{4^3}{3!} \approx 0.195367.$$

The values are very close.

Poisson distribution. Let λ be a positive real number. The random variable X is said to be a *Poisson random variable*, or to have a *Poisson distribution*, with parameter λ if its probability mass function is as follows:

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{when } x = 0, 1, 2, \dots, \text{ and } 0 \text{ otherwise.}$$

Exercise. Use the Maclaurin series for e^r to demonstrate that $\sum_{x=0}^{\infty} p(x) = 1$.

Poisson process. Events occurring in time or in space are said to be generated by an (approximate) *Poisson process* with rate λ when the following three conditions are satisfied:

1. **Independence:** The number of events occurring in disjoint subintervals of time or subregions of space are independent of one another.
2. **Proportionality:** The probability of one event occurring in a sufficiently small subinterval of time or subregion of space is proportional to the size of the subinterval or subregion. If h is the size of the subinterval or subregion, then the probability is λh .
3. **Unlimited Number of Events:** Any number of events are possible within a single interval of time or region of space. But, the probability of two or more events occurring in a sufficiently small subinterval or subregion is virtually 0.

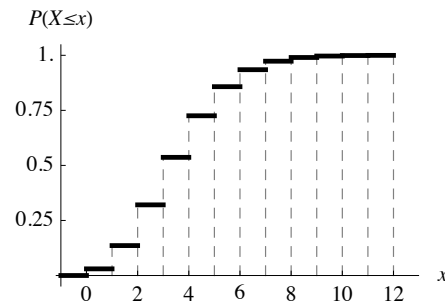
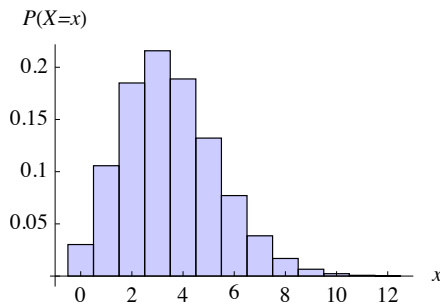
In this definition, λ represents the average number of events per unit time or space.

If events follow an (approximate) Poisson process and X is the number of events observed in one unit of time or space, then X has a Poisson distribution with parameter λ .

Poisson distributions have been used to model many real world situations such as: the number of angina attacks a patient suffers in a certain period of time, the number of industrial accidents occurring in a certain period of time, and the number of leukemia cases occurring within a certain distance of a toxic waste site.

Example. Suppose that cars passing a certain intersection in the middle of a work day follow a Poisson process, with an average of 7 cars per hour. Let X be the number of cars passing in a thirty-minute period. Then X has a Poisson distribution with $\lambda = \underline{\hspace{2cm}}$.

The PMF and CDF of X are displayed below:



- The probability of observing 4 or fewer cars in a thirty-minute period in the middle of a workday is

2.3 Continuous Random Variables

A continuous random variable is one whose range is an interval or a union of intervals.

2.3.1 Probability Distribution, PDF and CDF

The probability distribution of a continuous random variable is given in terms of its cumulative distribution function, or its probability density function.

Cumulative distribution function. If X is continuous, then the *cumulative distribution function* (CDF) of X is defined as follows:

$$F(x) = P(X \leq x) \text{ for all real numbers } x.$$

Cumulative distribution functions satisfy the following properties:

1. Infinite Limits: $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$.
2. Nondecreasing: If $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
3. Continuous: $F(x)$ is a continuous function for all real numbers x .

Note that $F(x)$ represents cumulative probability, with limits 0 and 1 (property 1). Cumulative probability increases with increasing x (property 2). For continuous random variables, the CDF is a continuous function (property 3).

Probability density function. If X is continuous, then the *probability density function* (PDF) of X is defined as follows:

$$f(x) = \frac{d}{dx} F(x) \text{ whenever the derivative exists.}$$

Probability density functions satisfy the following properties:

1. Nonnegativity: $f(x) \geq 0$ whenever f exists.
2. Unit Integral: $\int_{\mathcal{R}} f(x) dx = 1$, where \mathcal{R} is the range of X .

Note that $f(x)$ represents the rate of change of probability. The rate of change of probability must be nonnegative (property 1). Since $f(x)$ is the derivative of the CDF, the area under the graph of f must be 1 (property 2).

Computing probabilities using the PDF. If \mathcal{R} is the range of X and I is an interval, then the probability of the event “the value of X is in the interval I ” is obtained by finding the area under the density curve for $x \in I \cap \mathcal{R}$:

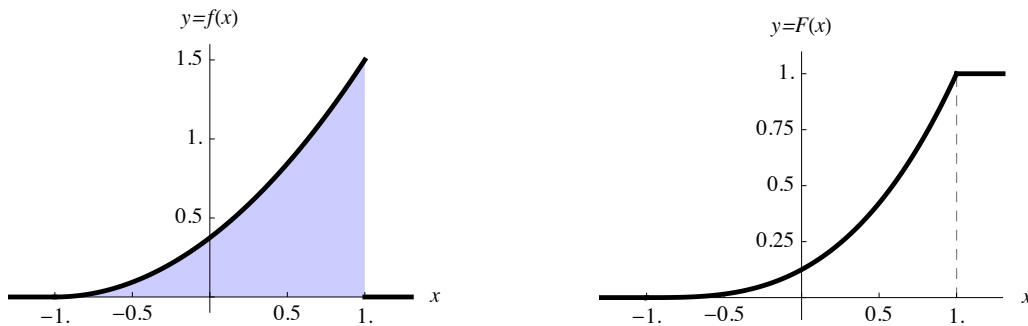
$$P(X \in I) = \int_{I \cap \mathcal{R}} f(x) dx.$$

Probability density functions do not always exist. When a PDF exists, then probabilities over intervals can be found using integration. In addition, if $a \in \mathcal{R}$, then $P(X = a) = 0$ since the area under the density curve and over an interval of length zero is zero.

Exercise. Let X be the continuous random variable with PDF as follows:

$$f(x) = \frac{3}{8}(x+1)^2 \quad \text{when } x \in [-1, 1]; \text{ and } 0 \text{ otherwise.}$$

The PDF and CDF of X are displayed below:



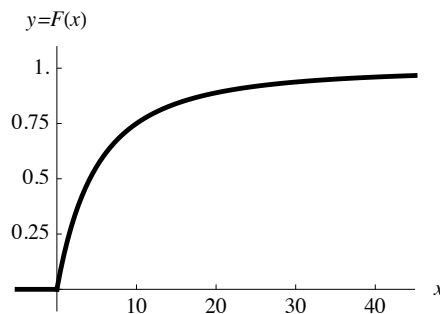
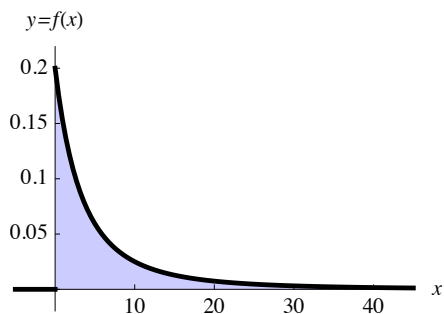
- (a) Completely specify the cumulative distribution function of X . Be sure to define your function for all real numbers x .

- (b) Find $P(X \geq 0)$. You can either use the CDF you calculated in part (a) or find the integral directly.

Exercise. Let X be the continuous random variable with PDF as follows:

$$f(x) = \frac{200}{(10+x)^3} \quad \text{when } x \in [0, \infty); \text{ and } 0 \text{ otherwise.}$$

The PDF and CDF of X are displayed below:



- (a) Completely specify the cumulative distribution function of X . Be sure to define your function for all real numbers x .

(b) Find $P(X \geq 8)$. You can either use the CDF you calculated in part (a) or find the integral directly.

(c) Suppose instead that the PDF of X is

$$f(x) = \frac{c}{(10+x)^4} \quad \text{when } x \in [0, \infty); \text{ and } 0 \text{ otherwise,}$$

where c is some positive constant. Find the value of c .

2.3.2 Quantiles, Percentiles

Let X be a continuous random variable, and p be a proportion satisfying $0 < p < 1$.

The p^{th} *quantile* (or $100p^{\text{th}}$ *percentile*) of the X distribution is the point, x_p , satisfying the equation

$$P(X \leq x_p) = p, \text{ if it exists.}$$

To find x_p , solve the equation $F(x) = p$ for x .

Exercise. In each case, find a general formula for the p^{th} quantile, where $p \in (0, 1)$.

(a) When the PDF of X is $f(x) = 3(x + 1)^2/8$ when $x \in [-1, 1]$, and 0 otherwise.

(b) When the PDF of X is $f(x) = 200/(10 + x)^3$ when $x \in [0, \infty)$, and 0 otherwise.

Median, quartiles, interquartile range. Important special cases are

1. **Median:** The median of the X distribution is its 50th percentile.

We use the notation $Med(X)$ to denote the median.

2. **Quartiles:** The quartiles of the X distribution are its 25th, 50th, and 75th percentiles.

In addition, the *interquartile range (IQR)* is the length of the interval containing the middle 50% of the X distribution. We use the notation $IQR(X)$ to denote its interquartile range.

The median is a measure of the center of a continuous distribution, and the interquartile range is a measure of the spread of the distribution.

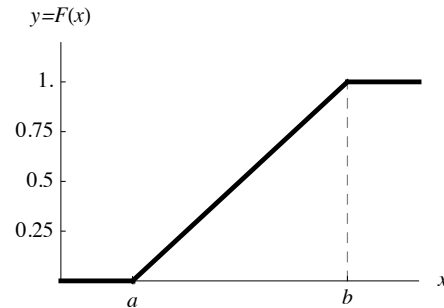
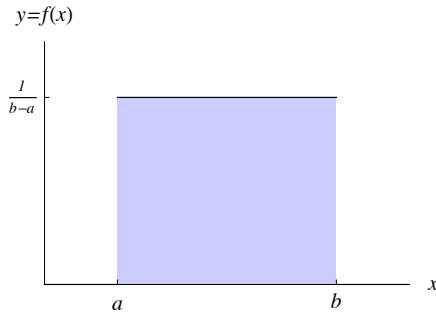
Exercise. Consider again the random variables from the last exercise. In each case, find the median of X and the interquartile range of X .

2.3.3 Continuous Uniform Distribution

Let a and b be real numbers with $a < b$. The random variable X is said to be a *uniform random variable*, or to have a *uniform distribution*, on the interval $[a, b]$ when its probability density function is as follows:

$$f(x) = \frac{1}{b-a} \quad \text{when } x \in [a, b]; \text{ and } 0 \text{ otherwise.}$$

General forms of the PDF and CDF of X are shown below:



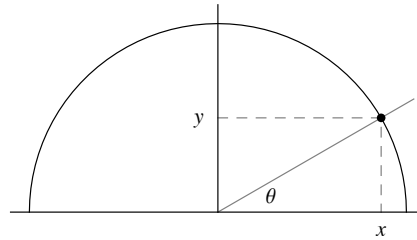
Exercise. Let X be a uniform random variable on the interval $[a, b]$.

- (a) Demonstrate that the probability that X lies in a subinterval of $[a, b]$ depends only on the length of the subinterval.
- (b) Use your answer to part (a) to find the median and interquartile range of X .

Exercise. Recall that positive angles are measured counterclockwise from the positive x -axis, and negative angles are measured clockwise from the positive x -axis.

Let Θ be a random angle in the interval $[0, \pi]$.

Given $\theta \in [0, \pi]$, construct a ray through the origin at angle θ and let (x, y) be the point where the ray intersects the half-circle of radius 1, as illustrated in the plot to the right.



Assume Θ is a continuous uniform random variable on the interval $[0, \pi]$. Find

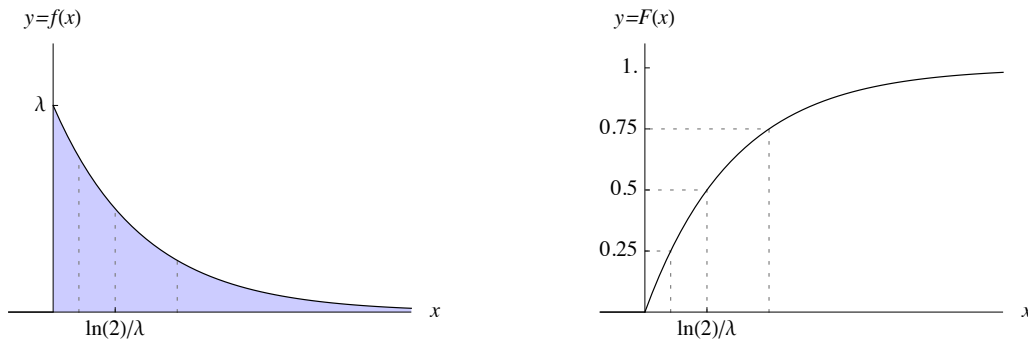
- (a) The probability that the x -coordinate of the point of intersection is less than $1/2$.
- (b) The probability that the y -coordinate of the point of intersection is less than $1/2$.

2.3.4 Exponential Distribution

Let λ be a positive real number. The random variable X is said to be an *exponential random variable*, or to have an *exponential distribution*, with parameter λ when its PDF is as follows:

$$f(x) = \lambda e^{-\lambda x} \quad \text{when } x \in [0, \infty); \text{ and } 0 \text{ otherwise.}$$

General forms of the PDF and CDF of X are shown below:



Note that the quartiles of the distribution have been marked in each plot.

Exercise. Let X be an exponential random variable with parameter λ .

- (a) Completely specify the CDF of X .
- (b) Find a simple formula for $P(X > x)$, where x is a positive real number.

- (c) Find a general formula for the p^{th} quantile of the X distribution, where $0 < p < 1$.
- (d) Use your answer to part (c) to verify that $Med(X) = \ln(2)/\lambda$ and $IQR(X) = \ln(3)/\lambda$.

Exercise. A trucker drives between fixed locations in Los Angeles and Phoenix. The duration, in hours, of a round trip has an exponential distribution with parameter $1/20$. Determine

- (a) The median round trip time.
- (b) The probability that the round trip exceeds 25 hours.
- (c) The probability a round trip takes at most 40 hours given that it exceeds 25 hours.

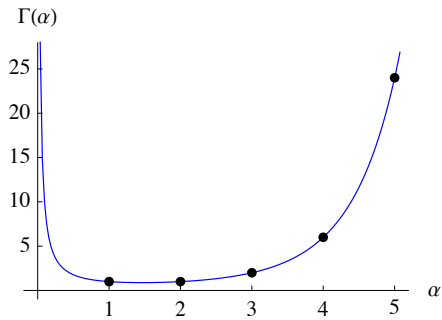
2.3.5 Euler Gamma Function, Gamma Distribution

Let α be a positive real number. The *Euler gamma function* is defined as follows:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$$

The gamma function is said to *interpolate* the factorials since, if $\alpha = r$ is a positive integer, then

$$\Gamma(r) = (r - 1)!$$



You should be able to verify that $\Gamma(r) = (r - 1)!$ using integration-by-parts.

Gamma distribution. Let α and λ be positive real numbers. The continuous random variable X is said to be a *gamma random variable*, or to have a *gamma distribution*, with parameters α and λ when its PDF is as follows:

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad \text{when } x \in (0, \infty); \text{ and } 0 \text{ otherwise.}$$

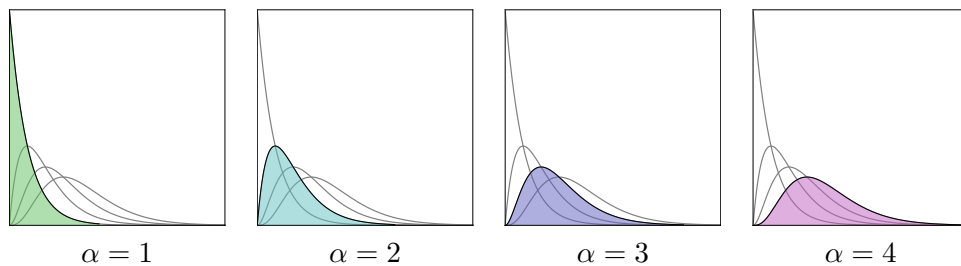
Notes:

1. If $\alpha = r$ is a positive integer, then the PDF of X is

$$f(x) = \frac{\lambda^r}{(r - 1)!} x^{r-1} e^{-\lambda x} \quad \text{when } x \in (0, \infty); \text{ and } 0 \text{ otherwise.}$$

Further, if $r = 1$, then X has an exponential distribution with parameter λ .

2. The parameter λ is known as the *scale* parameter of the gamma distribution, and the parameter α is known as the *shape* parameter of the distribution. As you vary α , the shape of the distribution changes, as illustrated below:



Note that the exponential distribution ($\alpha = 1$) has a fixed shape. And, when $\alpha > 1$, the distribution peaks at $(\alpha - 1)/\lambda$.

The gamma family of distributions is used in survival analysis and reliability analysis.

In both cases, we are concerned with the time until a catastrophic event occurs. Catastrophic events include the time until a person dies, the time until a person has a relapse of a serious illness, the time until a person has a car accident, the time until a machine breaks down.

1. *In Survival Analysis*, an exponential distribution can be used to model the time until a catastrophic event occurs when the population is fairly “homogeneous” (for example, all males born in the U.S. who are currently 20 to 25 years old).

When the population is “heterogeneous”, then the more general gamma distribution is used. The exponential distribution forms the basis of what actuaries call “life table analysis”, where lifetime information is tabled according to different groups defined by age, gender, occupation, and so forth.

2. *In Reliability Analysis*, an exponential distribution can be used to model the time until a machine breaks down when the machine itself is “simple” (not too many subsystems).

When the machine is “complicated”, then the more general gamma distribution is used.

2.3.6 Poisson Processes, Revisited

Suppose that events occur according to a Poisson process with rate λ , where λ represents the average number of events per unit time. Let X be the time you observe the r^{th} event.

- When $r = 1$, X has an exponential distribution with parameter λ ;
- When $r > 1$, X has a gamma distribution with parameters $\alpha = r$ and λ .

To see this,

- (i) Suppose you observe the process for t units of time, and let X be the time you observe the r^{th} event, starting from time 0.

Then X is a continuous random variable whose range is $(0, \infty)$. Further,

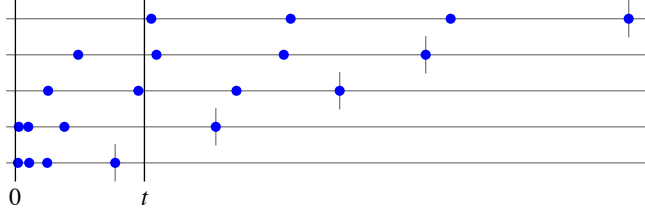
$$P(X > t) = P(\text{Fewer than } r \text{ events occur in the interval } [0, t]).$$

(You observe the r^{th} event after time t if and only if you observe fewer than r events in the time interval from 0 to t . Equivalent statements must have equal probabilities.)

- (ii) The probability on the right above can be written as follows:

$$P(\text{Fewer than } r \text{ events occur}) = \sum_{y=0}^{r-1} P(\text{Exactly } y \text{ events occur}).$$

The following graph illustrates the situation when $r = 4$:



In the graph, the dots represent the observed times of the first 4 events starting from time point 0, and the short vertical segments represent the times that the 4th event is observed. There are 5 possibilities to consider:

1. all four events occur before time point t (bottom row),
2. exactly 3 of the 4 events occur before time point t (second from bottom);
3. exactly 2 of the 4 events occur before time point t (third from bottom);
4. exactly 1 of the 4 events occur before time point t (fourth from bottom); and
5. all events occur after time point t (top row).

(iii) Let Y equal the number of events occurring in the fixed interval $[0, t]$. Then

$$P(X > t) = P(Y < r) = \sum_{y=0}^{r-1} P(Y = y).$$

Further, Y is a Poisson random variable with parameter λt .

(iv) The CDF of X can now be written in terms of the distribution of Y :

$$F(t) = P(X \leq t) = 1 - \sum_{y=0}^{r-1} P(Y = y) = 1 - \sum_{y=0}^{r-1} e^{-\lambda t} \frac{(\lambda t)^y}{y!} = 1 - e^{-\lambda t} \left(\sum_{y=0}^{r-1} \frac{(\lambda t)^y}{y!} \right),$$

and the PDF of X , $f(t) = \frac{d}{dt} F(t)$, can be obtained using the product rule:

$$\begin{aligned} \frac{d}{dt} \left[1 - e^{-\lambda t} \left(\sum_{y=0}^{r-1} \frac{\lambda^y t^y}{y!} \right) \right] &= \lambda e^{-\lambda t} \left(\sum_{y=0}^{r-1} \frac{\lambda^y t^y}{y!} \right) - e^{-\lambda t} \left(\sum_{y=1}^{r-1} \frac{\lambda^y y t^{y-1}}{y!} \right) \\ &= \lambda e^{-\lambda t} \left[\left(\sum_{y=0}^{r-1} \frac{\lambda^y t^y}{y!} \right) - \left(\sum_{y=1}^{r-1} \frac{\lambda^{y-1} t^{y-1}}{(y-1)!} \right) \right] \\ &= \lambda e^{-\lambda t} \left[\frac{\lambda^{r-1} t^{r-1}}{(r-1)!} \right] \\ &= \frac{\lambda^r}{(r-1)!} t^{r-1} e^{-\lambda t} \end{aligned}$$

Since the PDF of X is the same as the PDF of a gamma random variable with parameters $\alpha = r$ and λ , X has a gamma distribution with parameters $\alpha = r$ and λ .

Note that step (iv) implies that if r is a positive integer, then the CDF of a gamma random variable with parameters $\alpha = r$ and λ is

$$F(x) = P(X \leq x) = 1 - \sum_{y=0}^{r-1} e^{-\lambda x} \frac{(\lambda x)^y}{y!} \quad \text{when } x > 0 \text{ and } 0 \text{ otherwise,}$$

and the upper tail probability is $P(X > x) = \sum_{y=0}^{r-1} e^{-\lambda x} \frac{(\lambda x)^y}{y!}$ when $x > 0$.

In summary, there are three distributions related to Poisson processes over time:

1. The Poisson distribution with parameter λ , is used to represent the number of events occurring in one unit of time,
2. The exponential distribution with parameter λ , is used to represent the time until the first event (or the time between successive events) of a Poisson process, where λ is the average number of events in one time unit, and
3. The gamma distribution with parameters $\alpha = r$ and λ , is used to represent the time until the r^{th} event of a Poisson process, where λ is the average number in one time unit.

Example (Hand et al, Chapman & Hall, 1994) Researchers in Great Britain studied the occurrences of major earthquakes worldwide over the period of 1900 to 1980. They determined that the average time between successive events was approximately 425 days and that events followed an approximate Poisson process with rate

$$\lambda = \frac{1}{425} \quad (\text{or 1 event every 425 days, on average}).$$

Let X equal the time, in days, between successive major earthquakes, and assume that X has an exponential distribution with $\lambda = \frac{1}{425}$. Then

- The probability that the time between successive major earthquakes is between 200 and 625 days is

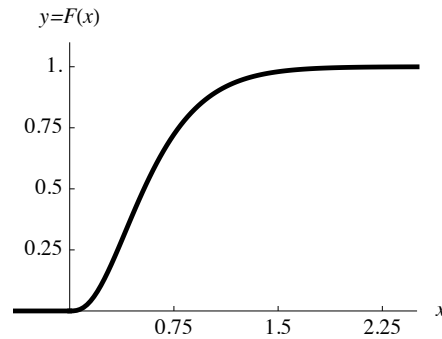
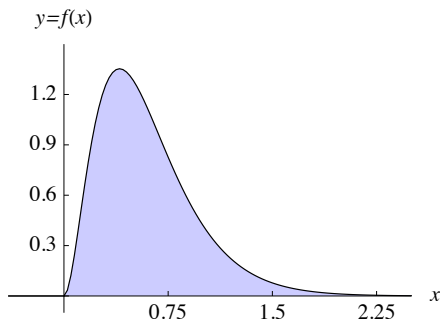
- The median time between successive major earthquakes is

Example. Suppose that cars passing a certain intersection in the middle of a work day follow a Poisson process, with an average of 5 cars per hour.

Let X be the time in hours until the third car passes. Then X has a gamma distribution with parameters

$$\alpha = r = \underline{\hspace{1cm}} \text{ and } \lambda = \underline{\hspace{1cm}}.$$

The PDF and CDF of X are displayed below:



- The probability that three cars will pass the intersection in one hour or less is

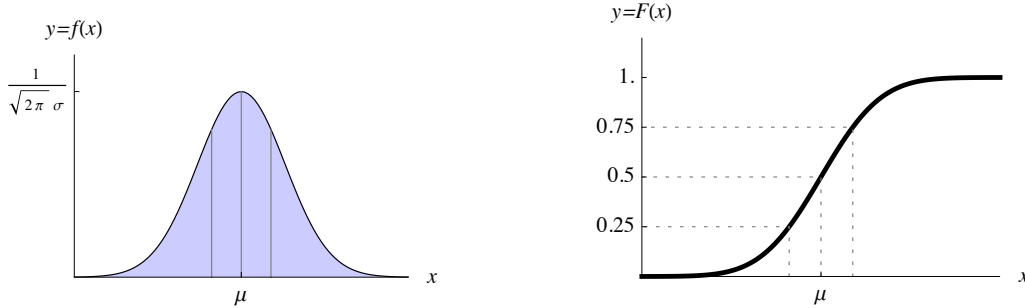
- The probability that it takes more than 45 minutes for three cars to pass the intersection is

2.3.7 Normal (Gaussian) Distribution

Let μ be a real number and σ be a positive real number. The continuous random variable X is said to be a *normal random variable*, or to have a *normal distribution*, with mean μ and standard deviation σ when its PDF and CDF are as follows:

1. Normal PDF: $f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-(x-\mu)^2/(2\sigma^2)}$ for all real numbers x .
2. Normal CDF: $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ for all real numbers x ,

where Φ is an important function of applied mathematics that we will discuss below. Here are general forms of the PDF and CDF of X :



Note that I have marked the locations of the quartiles of the distribution.

Gaussian distribution. The normal distribution is also called the *Gaussian distribution*, in honor of the mathematician Carl Friedrich Gauss.

Standard normal distribution, phi-functions. The *standard normal distribution* is the normal distribution with $\mu = 0$ and $\sigma = 1$. It is common practice to use the letter Z to represent the standard normal random variable.

Special letters are also used for the PDF and CDF of the standard normal random variable Z . Namely,

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad \text{and} \quad \Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

for all real numbers z .

Since $\phi(z)$ has no exact antiderivative, the values of $\Phi(z)$ must be approximated. Table 1 (page 34) gives values of the standard normal CDF when $z \geq 0$, where

$$z = \text{Row Value} + \text{Column Value}$$

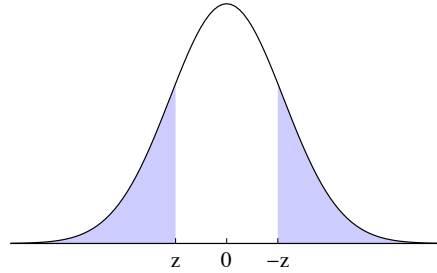
to two decimal places of accuracy. Cumulative probabilities are given with four decimal place accuracy.

When z is negative, we use the fact that the normal density curve is symmetric to find cumulative probabilities.

Specifically, if $z < 0$, then

$$\Phi(z) = 1 - \Phi(-z),$$

as illustrated on the right.

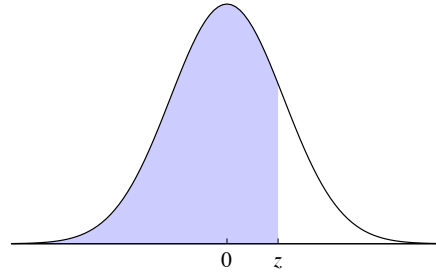


Exercise. Use Table 1 (page 34) to find $P(Z \leq 0.53)$, $P(Z \leq -1.32)$, and $P(-1 < Z < 1.5)$.

Quantiles of the standard normal distribution. Let z_p be the p^{th} quantile of the standard normal distribution. That is, z_p is the point satisfying the equation $\Phi(z_p) = p$.

Exercise. Use Table 1 (page 34) to find (approximately) the 20th, 40th, 60th and 80th percentiles of the standard normal distribution.

Table 1: *Standard normal cumulative probabilities, $\Phi(z) = P(Z \leq z)$, when $z \geq 0$.*



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Quantiles of the normal distribution. Let x_p be the p^{th} quantile of the normal distribution with mean μ and standard deviation σ . Then,

$$x_p = \mu + z_p\sigma,$$

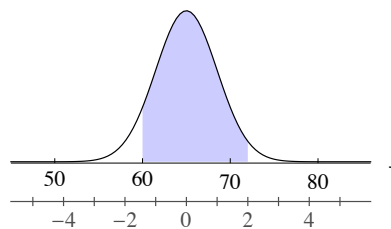
where z_p is the p^{th} quantile of the standard normal distribution.

Exercise. Let X be a normal random variable with mean μ and standard deviation σ .

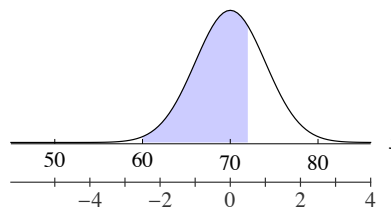
Use Table 1 (page 34) to demonstrate that the median of the X distribution is μ and the interquartile range is approximately 1.34σ .

Exercise (Agresti & Franklin, 2007, page 306). Distributions of heights for adult men and for adult women are often well-approximated by normal distributions. In North America, for example,

- (1) *Women's Heights:* the distribution of heights for adult women is well-approximated by a normal distribution with mean 65 inches and standard deviation 3.5 inches.



- (2) *Men's Heights:* the distribution of heights for adult men is well-approximated by a normal distribution with mean 70 inches and standard deviation 4 inches.



In each case, I drew a “ z -axis,” where $z = (x - \mu)/\sigma$, under the x -axis, and highlighted the area under the curve between 5 feet (60 inches) and 6 feet (72 inches).

Use the normal models to

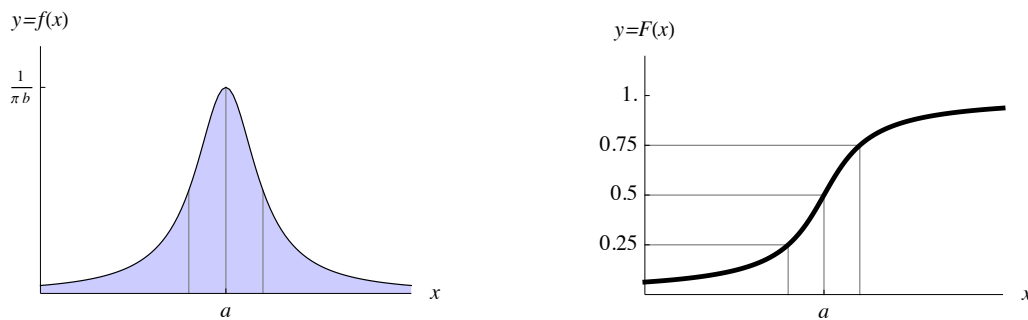
- (a) Find the probability that
 - (1) an adult women in North America is between 5 and 6 feet tall.
 - (2) an adult man in North America is between 5 and 6 feet tall.
- (b) Find the 20th, 40th, 60th and 80th percentiles of each distribution.

2.3.8 Cauchy Distribution

Let a be a real number and b be a positive real number. The continuous random variable X is said to be a *Cauchy random variable*, or to have a *Cauchy distribution*, with center a and spread b when its PDF and CDF are as follows:

1. *Cauchy PDF*: $f(x) = \frac{b}{\pi(b^2 + (x - a)^2)}$ for all real numbers x .
2. *Cauchy CDF*: $F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left(\frac{x - a}{b} \right)$ for all real numbers x .

General forms of the PDF and CDF of X are shown below:



Note that I have marked the locations of the quartiles of the distribution.

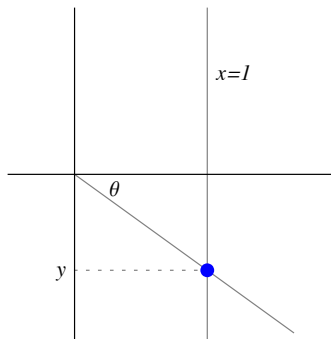
Exercise. Let X be a Cauchy random variable with center a and spread b . Find a general formula for the p^{th} quantile of the X distribution. Use your formula to demonstrate that the median of the X distribution is a and the interquartile range is $2b$.

Exercise. Recall that positive angles are measured counterclockwise from the positive x -axis, and negative angles are measured clockwise from the positive x -axis.

Let Θ be a random angle in the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$.

Given $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, construct a ray through the origin at angle θ and let $(1, y)$ be the point where the ray intersects the line $x = 1$, as illustrated in the plot to the right.

Assume Θ is a continuous uniform random variable on the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$, and let Y be the y -coordinate of the point of intersection as described above.



- (a) Demonstrate that $P(Y \leq y) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(y)$ for all y .
- (b) Use your answer to part (a) to conclude that Y has a Cauchy distribution.

2.3.9 Transforming Random Variables, CDF Method

Let X be a continuous random variable and $Y = g(X)$, where g is a continuous function. Our goal is to determine the Y distribution. The first steps will be to compute

$$P(Y \leq y) = P(g(X) \leq y)$$

as an expression in y , and to find $\frac{d}{dy}P(Y \leq y)$, for each y in the range of Y .

Exercise. Let X be the continuous random variable with PDF

$$f(x) = \frac{x}{2} \quad \text{when } x \in (0, 2); \text{ and } 0 \text{ otherwise.}$$

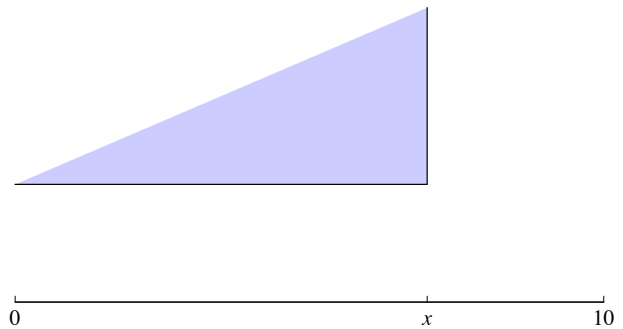
Further, let Y be the reciprocal of X , that is, $Y = 1/X$.

- (a) Find the range of Y .
- (b) Use the X distribution to write $P(Y \leq y)$ as an expression in y , and find the derivative of the expression with respect to y , $\frac{d}{dy}P(Y \leq y)$.
- (c) Completely specify the PDF of Y .

Exercise. Consider a wire of length 10 inches, with a coordinate system as shown on the right.

Given $x \in (0, 10)$, imagine bending the wire at position x by 90° to form 2 sides of a right triangle and let y be the area of that triangle.

Let X be an arbitrary bending point and Y be the area of the resulting triangle.



Assume that X is a continuous uniform random variable on the interval $(0, 10)$. Completely specify the PDF of the Y distribution.

Exercise. Let Z be the standard normal random variable, and let $Y = Z^2$. Completely specify the PDF of Y .