MATH-1550H: Introduction to Probability
Notes

Notation

Important Sets

- Natural numbers: $\mathbb{N} = \{1, 2, 3, \dots\}$.
- Integers: $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$.
- Rational numbers: $\mathbb{Q} = \left\{\frac{a}{b} | a, b \in \mathbb{Z}, b \neq 0\right\}$.
- Real numbers: All decimal expansions of the form $a_1 a_2 \dots a_n . d_1 d_2 d_3 \dots$ (made up of single digit whole numbers from 0 to 9) where $n \in \mathbb{N}$,

$$\mathbb{R} = \{a_1 a_2 \dots a_n . d_1 d_2 d_3 \dots | n \in \mathbb{N}, a_i, d_j \in \{0, 1, \dots, 9\}\}$$

- Complex numbers: $\mathbb{C} = \{a + bi | a, b \in \mathbb{R}, i^2 = -1\}$.

Greek Alphabet

(with common English pronunciations)

A, α - alpha (al-fah)	B, β - beta (bay-tah)
Γ, γ - gamma (gam-mah)	Δ, δ - delta (del-tah)
E, ϵ, (or ε) - epsilon (ep-si-lon)	Z, ζ - zeta (zay-tah)
H, η - eta (ay-tah)	Θ, θ - theta (thay-tah)
I, ι - iota (eye-oh-tah)	K, κ - kappa (ka-pah)
Λ, λ - lambda (lam-dah)	M, μ - mu (mew)
N, ν - nu (new)	Ξ, ξ - xi (ksigh, or ksee)
O, o - omicron (oh-mi-cron)	Π, π - pi (pie)
P, ρ - rho (row)	Σ, σ - sigma (sig-mah)
T, τ- tau (tow as in cow)	Υ, υ - upsilon (oop-si-lon)
Φ, ϕ, (or φ) - phi (fie as in hi, or fee)	X, χ - chi (ki as in hi, or kee)
Ψ, ψ - psi (sigh, psigh, or psee)	Ω, ω - omega (oh-may-gah)

Contents

Chapter 1

Combinatorial Methods

1.1 Counting

1.1 Theorem (Counting Rule for Compound Events)**.** If a process/operation/choice consists of two steps where the first can be done in n_1 ways and the second can be done in n_2 ways, then the entire process can be done in $n_1 \cdot n_2$ ways.

1.2 Example. How many different meal options can be made from a choice of 13 appetizers and 25 main dishes?

Solution. There are $13 \times 25 = 325$ different possible meals. □

1.3 Example. How many outfits can be made with 23 shirts and 14 pairs of pants?

Solution. There are $23 \times 14 = 322$ different possible outfits (even though some may not go very well together). □

1.4 Example (Cartesian Product of Sets)**.** If A and B are sets, we may form a new set

$$A \times B = \{(x, y) | x \in A, y \in B\}$$

of all (ordered) pairs of elements from from A and B.

If A and B are finite and have respectively m and n elements, then $A \times B$ has mn elements.

1.5 Theorem (General Counting Rule for Compound Events)**.** If a process consists of k steps where each can be done in n_i ways (for $i = 1, 2, \ldots, k$) then the entire process can be done in $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ ways.

1.6 Example (Cartesian Product of Sets)**.** Generalize the Cartesian product to k sets, $A_1, \ldots, A_k$, by

$$A_1 \times \cdots \times A_k = \{(x_1, \ldots, x_k) | x_i \in A_i \text{ for } i \in \{1, \ldots k\}\}.$$

Thus if $A_1, \ldots, A_k$ are finite and have $A_1, \ldots, A_k$ elements respectively, then $A_1, \ldots, A_k$ has $n_1 \cdots n_k$ elements.

1.7 Example. A room number in a certain university building is an ordered triple $(f, h, n) \in F \times H \times N$ where

$$F = \{\text{B1}, 1, 2\}, \quad H = \{\text{A,B}\}, \quad N = \{1, 2, 3, 4\}.$$

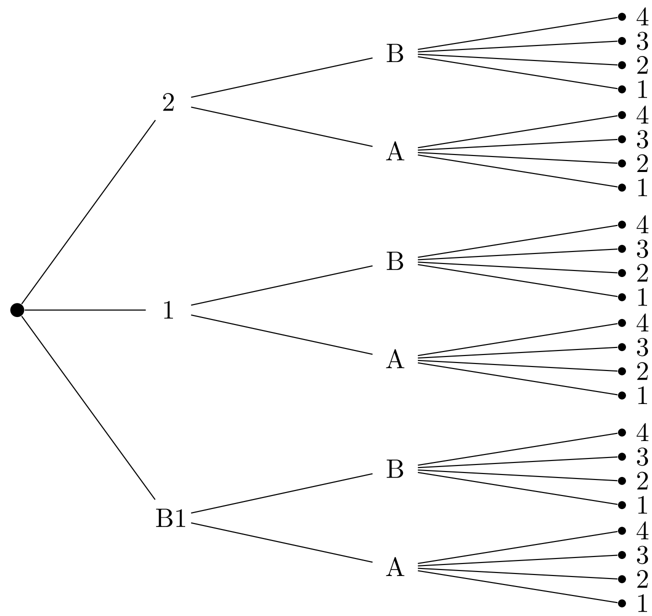
How many different room numbers are there?

Solution. There are

$$3 \times 2 \times 4 = 24$$

different room numbers.

The following *tree diagram* demonstrates the general counting rule for this example.



□

1.2 Permutations

1.8 Definition. An ordered arrangement of all elements of a set S, in which no element occurs more than once, is called a *permutation* of S. Another way to define this is that a permutation of set S is a bijection from S to itself.

1.9 Example. Let $S = \{a, b, c\}$. How many permutations of S are there?

Solution. The possible ordered arrangements of S can be listed as

$$abc, \quad acb, \quad bac, \quad bca, \quad cab, \quad cba$$

Without repetition, there are three choices for the first letter, leaving two for the second letter, and one remaining for the third letter. The general counting rule gives $3 \cdot 2 \cdot 1 = 6$ permutations. □

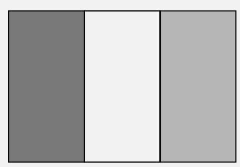
1.10 Theorem. The number of permutations of n distinct objects is

$$n! = n \cdot (n-1) \cdot \ldots \cdot 3 \cdot 2 \cdot 1.$$

We define $0! = 1$. The notation "$n!$" is read "n factorial".

1.11 Exercise. Let $\mathcal{S} = \{R, O, Y, G, B, I, V\}$ be the set of seven rainbow colours.

1. How many permutations of $\mathcal{S}$ are there?
2. How many different flags, composed of three vertical bars with distinct colours (as depicted below) can be made from $\mathcal{S}$?



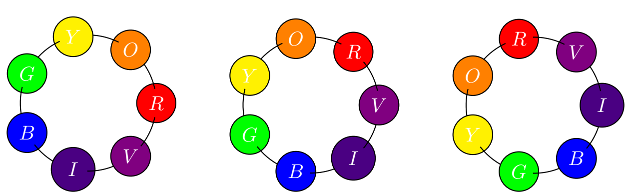
1.12 Theorem. The number of permutations of n distinct objects taken r at a time is

$$_nP_r = \frac{n!}{(n-r)!}$$

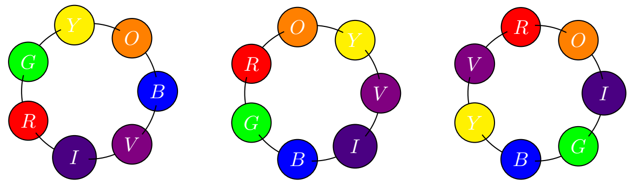
Circular Permutations

1.13 Example. How many different circular tile patterns, such as the ones below, can be made with the colours from $\mathcal{S} = \{R, O, Y, G, B, I, V\}$? (no two dots are the same colour)

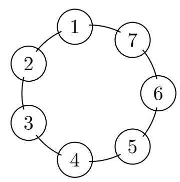
These patterns are considered to be the same since they differ by a rotation:



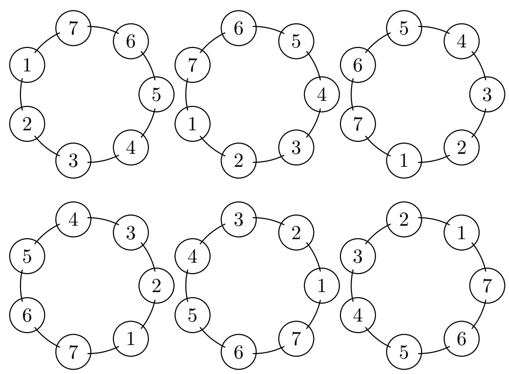
These patterns are different since they do not differ by a rotation:



Solution. Take any permutation of the 7 colours and arrange it around the circle a shown, starting at the top and going around counterclockwise (we have used numbers instead of colours for convenience).



Suppose we arrange all 7! permutations of the colours in this way. If we listed these, the same arrangement as the one above would appear 7 times but rotated in $2\pi/7$ angle increments. Here are the other 6:



Since each circular arrangement appears 7 times, we divide the total number of 7! permutations by 7 to get the number of different circular permutation. Thus there are

$$\frac{7!}{7} = 6! = 720$$

different tile patterns we could make.

This strategy can also be used to count the number of different coloured-bead necklaces that can be made with 7 different colours of beads, using one of each colour. However, when a bead necklace is flipped over it reverses the order of the beads. We then consider two necklace patterns to be the same if they differ by a rotation (just as above) or if the beads appear in the reverse order. It follows that there are

$$\frac{7!}{7 \cdot 2} = \frac{6!}{2} = 360$$

different bead necklaces that can be made using 7 different colours of beads, using one of each colour. □

The following theorem can be used to count the number of different tile patterns such as in the example above. It also counts the number of different ways to seat people at a round table, so that two seating arrangements are considered the same when every person has the same neighbor to their right (and left) even though their position at the table may differ by a rotation. The proof of this theorem follows from the example above.

1.14 Theorem (Circular Permutations)**.** The number of permutations of n distinct objects arranged in a circle is $(n-1)!$.

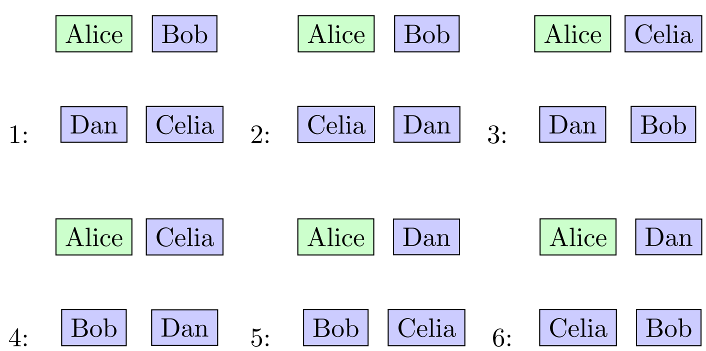
1.15 Example. How many different ways can we seat 4 people, Alice, Bob, Celia and Dan, at a round table? (Again "different" means they don't differ by a rotation.)

Solution. By the theorem we see that there are

$$(4-1)! = 3! = 6$$

different ways to seat them.

Here is another way to view the solution to this problem. Fix Alice's position at the table, and then arrange the other 3 people at the remaining 3 seats. Fixing Alice's position prevents repeating arrangements by rotation. There are $3! = 6$ ways to arrange the 3 others at the 3 seats, as seen below.

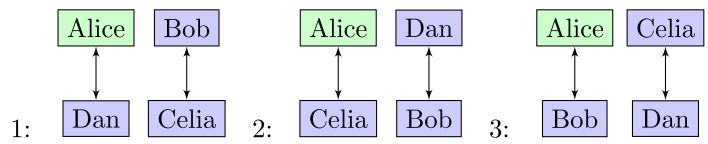


□

Round-Robin Tournament

A round-robing tournament is one where each person (or each team) plays every other team exactly once. To organize a round-robin tournament, think of seating people at long rectangular table, with the same number of chairs on the two long sides. Fix one position at the table (say at the corner) and cycle the other players clockwise. Matches occur between people sitting across from one another. If there is an odd number of players, have one empty seat for a "rest" match.

1.16 Example. Here is an example of a round-robin tournament with four players, and hence three rounds:



Permutations with Repeated Elements

1.17 Example. How many different permutations of the word "sassy" are there?

Solution. Here is a list of the different "words" we can make with these letters.

asssy, assys, asyss, aysss, sassy
sasys, sayss, ssasy, ssays, sssay
sssya, ssyas, ssysa, syass, sysas
syssa, yasss, ysass, yssas, ysssa

In order to count these without having to list them all, index each "s" to get the set

$$\{\mathrm{a},\mathrm{s}_1,\mathrm{s}_2,\mathrm{s}_3,\mathrm{y}\}.$$

There are 5! permutations of these 5 distinct objects. In each permutation, the 3! arrangements of $\mathrm{s}_1,\mathrm{s}_2,\mathrm{s}_3$ give the same word. e.g. the following are the same:

$$\mathrm{s}_1\mathrm{as}_2\mathrm{s}_3\mathrm{y},\ \mathrm{s}_1\mathrm{as}_3\mathrm{s}_2\mathrm{y},\ \mathrm{s}_2\mathrm{as}_1\mathrm{s}_3\mathrm{y},\ \mathrm{s}_2\mathrm{as}_3\mathrm{s}_1\mathrm{y},\ \mathrm{s}_3\mathrm{as}_1\mathrm{s}_2\mathrm{y},\ \mathrm{s}_3\mathrm{as}_2\mathrm{s}_1\mathrm{y}$$

Therefore there are $\frac{5!}{3!} = 20$ different permutations of the word "sassy"

□

1.18 Theorem (Permutations with Repeated Elements)**.** The number of permutations of n objects of which n_1 are of one kind, n_2 are of a second kind, ..., n_k are of a kth kind and $n = n_1 + n_2 + \cdots + n_k$ is

$$\frac{n!}{n_1! \cdot n_2! \cdot \cdots \cdot n_k!}.$$

1.19 Exercise. How many different 6-digit numbers can be formed with the single digits $1, 2, 2, 3, 3, 3$?

1.3 Combinations

1.20 Definition. A *combination* of n objects taken r at a time is any subset of size r taken from a set of size n. The order of the selection does not matter.

Note that in a locker "combination" the order does matter, making it a permutation.

1.21 Example. A photographer is allowed to choose three photos to display at an upcoming art exhibit. How many possible arrangements can be made from a selection of six photos.

Solution. Choosing the pictures in order, there are ${}_6P_3 = 6 \cdot 5 \cdot 4 = 120$ permutations. Each set of three appears $3! = 6$ times but in a different ordering. Therefore if we ignore order (i.e. consider different orderings of the same 3 pictures to be the same exhibit) there are $\frac{120}{6} = 20$ ways to choose three photos from six. □

1.22 Theorem. The number of ways to choose r objects from n distinct objects is

$$\binom{n}{r} = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!}$$

for $r = 0, 1, \ldots, n$. The notation $\binom{n}{r}$, also written ${}_nC_r$, is read "n choose r".

1.23 Example. In the card game *cribbage*, one pair scores two points. How many points are scored with four nines?

$$9\spadesuit \quad 9\heartsuit \quad 9\diamondsuit \quad 9\clubsuit$$

Solution. There are

$$\binom{4}{2} = \frac{4!}{2!2!} = \frac{24}{4} = 6$$

possible pairs of nines for a total of $2 \cdot 6 = 12$ points.

□

1.4 Partitions

1.24 Definition. A *partition* of a set S is an indexed collection of nonempty subsets $S_1, \ldots, S_k$ of S such that

(i) $S = S_1 \cup \cdots \cup S_k$, and

(ii) $S_i \cap S_j = \emptyset$ for all $i, j \in \{1, \ldots, k\}$, $i \neq j$.

(in general a partition may be infinite).

The order within each subset S_i does not matter, but the ordering of the subsets themselves does matter.

1.25 Example. There are 12 partitions of $\{a, b, c, d\}$ into three subsets with two, one and one elements respectively:

$$\{a,b\},\{c\},\{d\} \quad \{a,b\},\{d\},\{c\}, \quad \{a,c\},\{b\},\{d\}, \quad \{a,c\},\{d\},\{b\}$$

$$\{a,d\},\{b\},\{c\} \quad \{a,d\},\{c\},\{b\}, \quad \{b,c\},\{a\},\{d\}, \quad \{b,c\},\{d\},\{a\}$$

$$\{b,d\},\{a\},\{c\} \quad \{b,d\},\{c\},\{a\}, \quad \{c,d\},\{a\},\{b\}, \quad \{c,d\},\{b\},\{a\}$$

1.26 Exercise. Offices A and B have 6 desks each, and offices C and D have 4 desks each. In how many ways can 20 grad students be put into these 4 office offices? (it doesn't matter who sits at what desk within an office)

1.27 Theorem. The number of ways in which n distinct objects can be partitioned into k subsets, with n_1 objects in the first subset, n_2 in the second,..., n_k in the kth, is

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}.$$

*note $n_1 + \cdots + n_k = n$.

1.5 Binomial Coefficients

Powers of a binomial $x + y$, are computed using properties of real numbers (distributivity, associativitiy, commutativity). e.g.

$$\begin{aligned}(x+y)^2 &= (x+y)(x+y)\\ &= (x+y)x + (x+y)y\\ &= xx + yx + xy + yy\\ &= x^2 + 2xy + y^2.\end{aligned}$$

Expanding $(x + y)^n$ in this way for large n is impractical. Instead we can compute the coefficients of each $x^k y^{n-k}$ term in the result with counting techniques.

1.28 Example. Here is the expansion of $(x + y)^3$,

$$\begin{aligned}(x+y)^3 &= (x+y)(x+y)(x+y)\\ &= xxx + xxy + xyx + xyy + yxx + yxy + yyx + yyy\\ &= x^3 + 3x^2y + 3xy^2 + y^3.\end{aligned}$$

Each term in the second line is obtained by choosing one of either x or y from each of the 3 factors. Below, the letters in boxes are chosen to form the product:

$$\begin{array}{lcl}
xxx & \leftrightsquigarrow & (\boxed{x}+y)(\boxed{x}+y)(\boxed{x}+y), \\
xxy & \leftrightsquigarrow & (\boxed{x}+y)(\boxed{x}+y)(x+\boxed{y}), \\
xyx & \leftrightsquigarrow & (\boxed{x}+y)(x+\boxed{y})(\boxed{x}+y), \\
yxx & \leftrightsquigarrow & (x+\boxed{y})(\boxed{x}+y)(\boxed{x}+y), \\
xyy & \leftrightsquigarrow & (\boxed{x}+y)(x+\boxed{y})(x+\boxed{y}) \\
yxy & \leftrightsquigarrow & (x+\boxed{y})(\boxed{x}+y)(x+\boxed{y}) \\
yyx & \leftrightsquigarrow & (x+\boxed{y})(x+\boxed{y})(\boxed{x}+y) \\
yyy & \leftrightsquigarrow & (x+\boxed{y})(x+\boxed{y})(x+\boxed{y})
\end{array}$$

In summary, we choose k factors (of the three) to provide y to get a $x^{3-k}y^k$ term for $k = 0, 1, 2, 3$. For example, there are

$$\binom{3}{2} = 3$$

ways to obtain an xy^2 term by choosing y from two of the factors and x from the remaining one.

1.29 Theorem (The Binomial Theorem)**.** For $n \in \mathbb{N}$

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^{n-r} y^r.$$

The numbers $\binom{n}{r}$ are called *binomial coefficients*.

Choosing r things from n things indirectly chooses $n - r$ things to leave behind. This demonstrates the next theorem.

1.30 Theorem. For $n \in \mathbb{N}$ and $r = 0, 1, \ldots, n$

$$\binom{n}{r} = \binom{n}{n-r}.$$

1.31 Exercise. Verify this theorem by using the definitions of the right and left sides of the equation, and show that

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^{n-r} y^r = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}.$$

Binomial coefficients can be arranged in the formation below which is known as Pascal's Triangle. The first seven rows of Pascal's triangle are shown below.

$$\begin{array}{ccccccccccccc}
&&&&&&\binom{0}{0}&&&&&& \\
&&&&&\binom{1}{0}&&\binom{1}{1}&&&&& \\
&&&&\binom{2}{0}&&\binom{2}{1}&&\binom{2}{2}&&&& \\
&&&\binom{3}{0}&&\binom{3}{1}&&\binom{3}{2}&&\binom{3}{3}&&& \\
&&\binom{4}{0}&&\binom{4}{1}&&\binom{4}{2}&&\binom{4}{3}&&\binom{4}{4}&& \\
&\binom{5}{0}&&\binom{5}{1}&&\binom{5}{2}&&\binom{5}{3}&&\binom{5}{4}&&\binom{5}{5}& \\
\binom{6}{0}&&\binom{6}{1}&&\binom{6}{2}&&\binom{6}{3}&&\binom{6}{4}&&\binom{6}{5}&&\binom{6}{6}
\end{array}$$

Each row of the triangle contains lists all $\binom{n}{r}$ for a fixed n and $r = 0, \ldots, n$. Evaluating these reveals an interesting pattern; each number in the triangle is the sum of the two adjacent numbers in row above.

$$\begin{array}{ccccccccccccc}
&&&&&&1&&&&&& \\
&&&&&1&&1&&&&& \\
&&&&1&&2&&1&&&& \\
&&&1&&3&&3&&1&&& \\
&&1&&4&&6&&4&&1&& \\
&1&&5&&10&&10&&5&&1& \\
1&&6&&15&&20&&15&&6&&1
\end{array}$$

The next theorem proves that this pattern in Pascal's Traingle is not a coincidence.

1.32 Theorem. For $n \in \mathbb{N}$ and $r = 0, 1, \ldots, n-1$

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}.$$

Proof. Proving this theorem by algebraic manipulation is straightforward, however we will demonstrate another proof strategy here which is useful in combinatorics. We will equate coefficients of like powers of x for two equal polynomial expressions.

Consider,

$$(1+x)^n = (1+x)(1+x)^{n-1} = (1+x)^{n-1} + x(1+x)^{n-1}.$$

Both sides are polynomials in x, and two polynomials are equal if and only they have the same coefficients, so we may equate coefficients of x^r for any $r = 0, 1, \ldots, n$.

On the left, the coefficient of x^r is $\binom{n}{r}$ (by binomial theorem).

On the right, the coefficient of x^r in $(1+x)^{n-1}$ is $\binom{n-1}{r}$, and in $x(1+x)^{n-1}$, the coefficient on x^r is $\binom{n-1}{r-1}$.

Thus $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$.

□

It makes sense to define $\binom{n}{r} = 0$ when $r > n$. Using this we have the following identity.

1.33 Theorem. For $m, n, k \in \mathbb{N}$,

$$\sum_{r=0}^{k} \binom{m}{r}\binom{n}{k-r} = \binom{m+n}{k}.$$

Proof. Using the same proof strategy as above, consider

$$(1+x)^{m+n} = (1+x)^m(1+x)^n$$

viewing left and right hand sides as polynomials in x. Now find the coefficient of x^k on both sides. □

1.34 Exercise. Write the details for the proof of this theorem.

Multinomial Coefficients

The expansion of $(x_1 + x_2 + \cdots + x_k)^n$, a generalization of a binomial called a *multinomial*, is the sum of all monomials $x_1^{r_1} x_2^{r_2} \cdots x_k^{r_k}$ where $r_1 + \cdots + r_k = n$. For example

$$\begin{aligned}(x_1+x_2+x_3)^4 = {} & x_1^4 + 4x_1^3x_2 + 4x_1^3x_3 + 6x_1^2x_2^2 + 12x_1^2x_2x_3 + 6x_1^2x_3^2 + 4x_1x_2^3 \\ & + 12x_1x_2^2x_3 + 12x_1x_2x_3^2 + x_2^4 + 4x_2^3x_3 + 6x_2^2x_3^2 + 4x_2x_3 + x_3^4\end{aligned}$$

1.35 Theorem. The coefficient of $x_1^{r_1} x_2^{r_2} \cdots x_k^{r_k}$ in $(x_1 + x_2 + \cdots + x_k)^n$ is

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! \cdot r_2! \cdot \ldots \cdot r_k!}.$$

Proof. Each exponent is a different partition of n into k subsets and so the theorem on partitions applies. □

1.36 Remark. Notice that the multinomial coefficient

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! \cdot r_2! \cdot \ldots \cdot r_k!},$$

which we interpreted as counting the number of partitions of n objects into k subsets, also gives the number of words that can be made with n letters of which there are r_i copies of letter x_i, for each $i = 1, \ldots, k$.

1.37 Exercise. What the is coefficient of $x_1^2 x_2^2 x_4^3$ in the expansion of

$$(x_1 + x_2 + x_3 + x_4)^7?$$

1.38 Exercise. Give the expansion of $(x_1 + x_2 + x_3 + x_4)^4$.

Chapter 2

Probability

Mathematics is used to model real world phenomena. Here are two classes of model.

Deterministic model (ideal situation): Predicts the outcome of an experiment with certainty based on given initial conditions. e.g. velocity of a falling object

$$v = gt.$$

Probabilistic, or stochastic, model (randomness): When the same initial conditions can lead to a variety of outcomes, these models provide a value (probability) to the possible outcomes. e.g. rolling a die results in one of six numbers facing up.

⚀ ⚁ ⚂ ⚃ ⚄ ⚅

Assigning each outcome the value $\frac{1}{6}$ is one way to model this.

Classical Probability Concept

When there are N possible (equally likely) outcomes of which n are considered successful, then the *probability* of a success is the ratio $\frac{n}{N}$.

2.1 Example. Some examples of the classical probability concept: The probability of,

- tossing tails with a balanced coin: $\frac{1}{2}$
- drawing an ace from deck of cards: $\frac{4}{52}$
- rolling either 3 or 5 with a six-sided die: $\frac{2}{6}$
- rolling a total of 1 with a pair of dice: 0
- landing on red in roulette: $\frac{18}{38}$

These values describe the frequency of a successful outcome; the proportion of time the event occurs in the long run.

2.1 Sample Spaces

The set of all possible outcomes of an experiment is called the *sample space*.

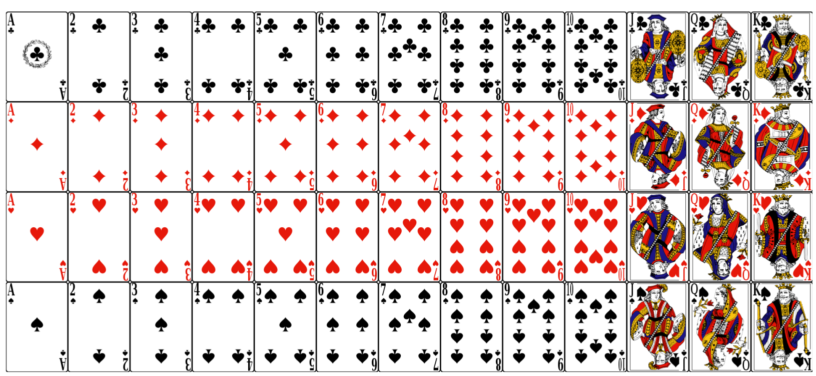
2.2 Example.

Experiment	Sample space
single coin toss	$\{H, T\}$
roll of two dice	$\{(d_1, d_2) \| d_1, d_2 \in \{1, 2, 3, 4, 5, 6\}\}$
sum of two dice	$\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
drawing a card	all (r, s) with $r \in \{2, \ldots, 10, J, Q, K, A\}$ and $s \in \{\spadesuit, \heartsuit, \diamondsuit, \clubsuit\}$
three coin tosses	$\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

Deck of Cards

Playing cards will be used in several examples, and it will be good to be familiar with them. A standard deck of 52 playing cards is organized into

- 4 suits: clubs - $\clubsuit$, diamonds - $\diamondsuit$, hearts - $\heartsuit$, spades - $\spadesuit$
- 13 ranks: A (ace), 2, 3, 4, 5, 6, 7, 8, 9, 10, J (jack), Q (queen), K (king)



2.3 Example (Blackjack)**.** In the card game Blackjack (a.k.a. Twenty-one), aces are worth 1 or 11, jacks, queens, and kings are worth 10, and all other cards are worth their face value (2-9). According to the rules of classic probability, what is the probability of drawing two cards from a deck of 52 whose sum is 21? One example would be the hand $J\spadesuit\ A\heartsuit$.

All previous examples had finite sample spaces. The following experiments have infinite sample spaces.

2.4 Example. Some examples of experiments with infinite sample spaces.

- Tossing a coin until heads is reached:

$$\{H, TH, TTH, TTTH, TTTTH, \dots\}$$

- Playtime for two AA alkaline batteries in a Wii remote:

$$\{t \text{ hours} | t \in [0, 50]\}$$

- Actual length of a non-pointed heavy hex bolt of nominal length 5 inches, and nominal size 3/8 inches:

$$\{\ell \text{ inches} | \ell \in [4.9, 5.06]\}$$

Continuous and Discrete Sample Spaces

There is an important distinction between the sample spaces in the previous example; the outcomes of the first example (coin toss) may be listed, whereas the outcomes in the other two belong to a continuum of values.

Discrete sample space: has only finitely many, or a countably infinite number of elements.

Continuous sample space: is an interval in $\mathbb{R}$, or a product of intervals lying in $\mathbb{R}^n$.

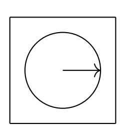
The important distinction is how probabilities are assigned.

2.2 Events

While individual elements of a sample space are called *outcomes*, subsets of a sample space are called *events*. If the outcome of an experiment lies in an event, we say that event has occurred.

2.5 Example. Experiment: Tossing a coin three times.
Sample space: $\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$
Event A: Getting at least two heads $\{HHH, HHT, HTH, THH\}$
Event B: Getting exactly two tails $\{HTT, THT, TTH\}$
Event C: Getting two consecutive heads $\{HHH, HHT, THH\}$
Event D: Getting three consecutive heads $\{HHH\}$

2.6 Example. Experiment: Spinning a probability spinner.



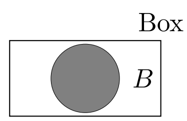
$$\text{Sample space: } \{\theta \text{ degrees} | \theta \in [0, 360)\}$$

Event A: Landing between 90 and 180 degrees, $[90, 180]$
Event B: Landing either between 45 and 90 degrees or between 270 and 315 degrees, $[45, 90] \cup [270, 315]$
Event C: Landing precisely on 180 degrees.

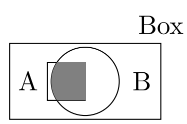
2.7 Example. Experiment: Dropping a pencil head first into a rectangular box. Sample space: All points on the bottom of the box.

Event A:

Event B:



Event $A \cap B$:



(The event occurs when the pencil lands in shaded region.)

Union, Intersection, Complement

Let A and B be events in sample space S; i.e. A and B are subsets of a set S.

The *union* of A and B is the set of outcomes that is in either A or B or both. Symbolically,

$$A \cup B = \{x \in S | x \in A \text{ or } x \in B\}.$$

The *intersection* of A and B is the set of outcomes that is in both A and B.

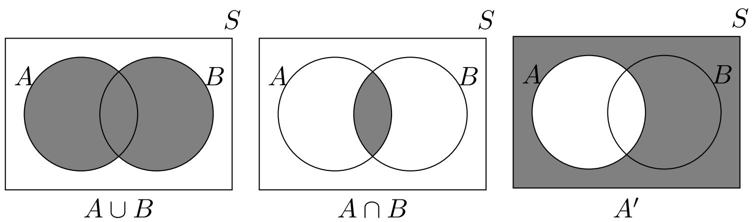
$$A \cap B = \{x \in S | x \in A \text{ and } x \in B\}.$$

The *complement* of A in S is the set of outcomes in S that are not in A.

$$A' = \{x \in S | x \notin A\} = S \setminus A.$$

Venn Diagrams

A *Venn diagram* is a visual depiction of subsets of some "universal" set. Subsets are represented (typically) by disks lying within a rectangle representing the universal set. Sets of interest are represented by shaded regions.

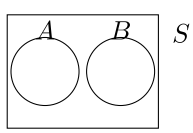


In the pencil dropping example, Venn diagrams gave a literal representation of the events in that experiment, but these representations can be used in more general situations to help visualize relationships between different subsets of a sample space.

Mutually Exclusive Events

A set which has no elements is called the *empty set*, denoted $\emptyset$.

2.8 Example. Experiment: Rolling two dice.
Event A: Rolling at least one six,
$A = \{(d_1, 6), (6, d_2) | d_1, d_2 \in \{1, 2, 3, 4, 5, 6\}\}$
Event B: Sum of dice equals 4,
$B = \{(1, 3), (2, 2), (3, 1)\}$.
Event C: Rolling at least one six and having a sum of 4,
$C = A \cap B = \emptyset$.



$A \cap B = \emptyset$

Sets with empty intersection are called *disjoint*, and the events in this case are called *mutually exclusive*.

Algebra of Sets

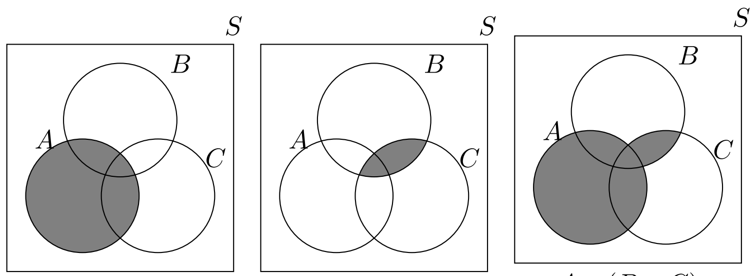
Let A, B and C be subsets of a universal set S.

- Idempotent laws:
  $A \cup A = A, \qquad A \cap A = A$
- Associative laws:
  $(A \cup B) \cup C = A \cup (B \cup C), \qquad (A \cap B) \cap C = A \cap (B \cap C)$

- Commutative laws:
  $A \cup B = B \cup A, \qquad A \cap B = B \cap A$
- Distributive laws:
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C), \quad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- Identity laws:
  $A \cup \emptyset = A, \quad A \cup S = S, \quad A \cap S = A, \quad A \cap \emptyset = \emptyset$
- Complement laws:
  $(A')' = A, \quad A \cup A' = S, \quad A \cap A' = \emptyset, \quad S' = \emptyset, \quad \emptyset' = S$
- DeMorgan's Laws:
  $(A \cup B)' = A' \cap B', \quad (A \cap B)' = A' \cup B'$

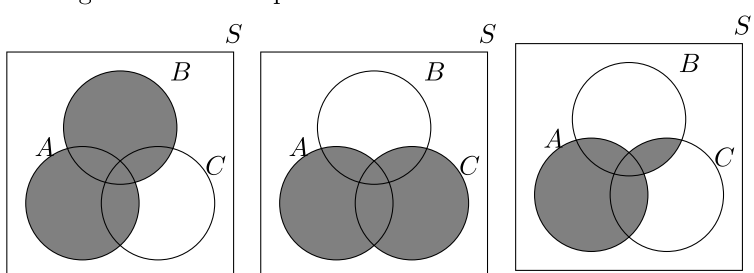
2.9 Example. Use Venn diagrams to verify the distributive law $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Solution. The left side of the equation is the union of A and B:



$A \qquad B \cap C \qquad A \cup (B \cap C)$

The right side of the equation is the intersection of $A \cup B$ and $A \cup C$.



$A \cup B \qquad A \cup C \qquad (A \cup B) \cap (A \cup C)$

Since the resulting diagrams are the same, we conclude that the equation holds. □

While Venn diagrams are useful for intuition, they should not be used for a rigorous proof. Below is a proof of the distributive law using definitions of the sets involved.

Proof. Let $x \in A \cup (B \cap C)$. Then $x \in A$ or $x \in (B \cap C)$.

- If $x \in A$ then $x \in A \cup B$ and $x \in A \cup C$, so $x \in (A \cup B) \cap (A \cup C)$.
- If $x \in B \cap C$ then $x \in B$ and $x \in C$ so $x \in A \cup B$ and $x \in A \cup C$, and thus $x \in (A \cup B) \cap (A \cup C)$.

This shows that if $x \in A \cup (B \cap C)$ then $x \in (A \cup B) \cap (A \cup C)$, and hence $A \cup (B \cap C) \subset (A \cup B) \cap (A \cup C)$.

Now suppose $y \in A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. Then $y \in A \cup B$ and $y \in A \cup C$.

- If $y \notin A$ we must have $y \in B$ and $y \in C$, so $y \in B \cap C$.
- Otherwise $y \in A$.

In either case $y \in A \cup (B \cap C)$, and it follows that $(A \cup B) \cap (A \cup C) \subset A \cup (B \cap C)$.

$A \cup (B \cap C) \subset (A \cup B) \cap (A \cup C)$ and $(A \cup B) \cap (A \cup C) \subset A \cup (B \cap C)$ implies $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

□

2.10 Exercise. Use Venn diagrams to verify DeMorgan's Laws.

$$(A \cup B)' = A' \cap B', \quad (A \cap B)' = A' \cup B'$$

2.3 The Probability of an Event

A *probability*, or *probability measure*, is a function P which maps events in the sample space S to real numbers.

In order to assign probabilities in a meaningful way, P must satisfy the following called the *postulates* (or *axioms*) *of probability*.

P1. The probability of any event A in S is a nonnegative real number, i.e. $P(A) \geq 0$.

P2. $P(S) = 1$.

P3. If $A_1, A_2, A_3, \ldots,$ is a finite or infinite sequence of (pairwise) mutually exclusive events in S then

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

(P is *countably additive*)

Interpreting a probability as a frequency, or a proportion of time, it makes sense that $P(A) \geq 0$; in fact we will show that $0 \leq P(A) \leq 1$ for any event A.

P2 says that the probability that the outcome of the experiment lies in S must be assigned value 1. Since this is certain to happen, we interpret $P(A) = 1$ as "A happens 100 percent of the time."

P3 is for consistency. For example, if events A_1 and A_2 share no common outcomes, then the probability that either event occurs, $P(A_1 \cup A_2)$, is the sum of their individual probabilities.

2.11 Remark. A technical detail has been overlooked in the postulates of probability presented above. In a discrete sample space S, an "event" can be any subset of S, however in the continuous case one has to be more careful about which subsets of S are allowed as events. A precise definition for these allowable events comes in a course on *measure theory*. In this course we won't require that level of detail; i.e. the subsets we assign probabilities to will be allowable events.

2.12 Example. Consider the experiment of rolling a single 6-sided die. The sample space is,

$$S = \{⚀,\ ⚁,\ ⚂,\ ⚃,\ ⚄,\ ⚅\}$$

Each outcome in S is its own event, call these $A_1, \ldots, A_6$. Events $A_1, \ldots, A_6$ are mutually exclusive, and any event E in S is a union of these, for example let $E = A_2 \cup A_4 \cup A_5$. By the classical probability concept, $P(E) = \frac{3}{6}$ (successes/number of outcomes), and $P(A_i) = \frac{1}{6}$ for each i. This P satisfies the postulates of probability:

- $P(B) \geq 0$ for any $B \subset S$.
- $P(S) = \frac{6}{6} = 1$.
- P3 is satisfied: by example $P(E) = \frac{3}{6} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = P(A_2) + P(A_4) + P(A_5)$.

2.13 Exercise.

$$S = \{⚀,\ ⚁,\ ⚂,\ ⚃,\ ⚄,\ ⚅\}$$

Suppose we assigned probabilities to this die rolling experiment in a different way. Using the same notation as before, for any event B specify that

$$P(B) = \sum_{A_i \in B} P(A_i), \text{ and}$$

$$P(A_1) = \tfrac{1}{2}, P(A_2) = \tfrac{1}{4}, P(A_3) = \tfrac{1}{8},$$
$$P(A_4) = 0, P(A_5) = \tfrac{1}{16}, P(A_6) = \tfrac{1}{16}$$

Are the postulates of probability still satisfied?

2.14 Exercise. An experiment has four possible outcomes A, B, C, D that are mutually exclusive. Explain why the following assignments of probabilities are not permissible.

(a) $P(A) = 0.12, P(B) = 0.63, P(C) = 0.45, P(D) = -0.20$

(b) $P(A) = \frac{9}{120}, P(B) = \frac{45}{120}, P(C) = \frac{27}{120}, P(D) = \frac{46}{120}$

2.15 Theorem. If A is an event in a discrete sample space S, then $P(A)$ is the sum of the probabilities of the individual outcomes (elements) of A.

(*Theorem assumes P is a probability measure, and hence satisfies the postulates)

2.16 Example. Experiment: Tossing a coin three times.
Sample space: $\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$
Event A: Getting at least two heads $\{HHH, HHT, HTH, THH\}$
Event B: Getting exactly two tails $\{HTT, THT, TTH\}$
Event C: Getting two consecutive heads $\{HHH, HHT, THH\}$
Event D: Getting three consecutive heads $\{HHH\}$

2.17 Exercise. Assuming a *balanced* coin, i.e. equal likely heads or tails, what are the probabilities of the events in the example above?

2.18 Exercise.

$S = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$

Suppose our six sided die is weighted so that each odd number is twice as likely to occur than each even number.

What is the probability of rolling a number greater than 3?

What if instead each even number is four times as likely to occur than each odd number?

Infinite Discrete Sample Spaces

When a sample space has countably infinite outcomes, probabilities must be assigned via some rule/formula as opposed to listing them individually.

2.19 Example. Tossing a coin until heads is reached:

$$S = \{H, TH, TTH, TTTH, TTTH, TTTTH, TTTTTH, \dots\}.$$

If A_i is the event of i flips, then $P(A_i) = \frac{1}{2^i}$ defines a probability on S (assuming countable additivity). From the geometric series formula we get P2:

$$P(S) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} \frac{1}{2^i} = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1.$$

Brief note on infinite series:

- Sequence: countably infinite list of real numbers; $s_1, s_2, s_3, \ldots$.

$$\text{e.g.} \quad 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, \ldots$$

- Geometric sequence: terms occur in a common ratio r; $a, ar, ar^2, ar^3 \ldots$

$$\text{e.g.} \quad 4, \tfrac{4}{3}, \tfrac{4}{9}, \tfrac{4}{27}, \tfrac{4}{81}, \ldots\ldots$$

- Partial sum of a sequence: $S_n = \displaystyle\sum_{i=1}^{n} s_i = s_1 + s_2 + \cdots + s_n$.

- Series: Limit of partial sums (if it exists); $\displaystyle\lim_{n\to\infty} S_n = \sum_{i=1}^{\infty} s_i$.

- Partial sum geometric sequence: $G_n = a + ar + \cdots + ar^n$.

$$\begin{aligned}(1-r)G_n &= (1-r)(a + ar + ar^2 + \cdots + ar^n) \\ &= (a + ar + ar^2 + \cdots + ar^n) \\ &\quad - (ar + ar^2 + ar^3 + \cdots + ar^{n+1}) \\ &= a - ar^{n+1}\end{aligned}$$

  So $G_n = \frac{a - ar^{n+1}}{(1-r)}$ (for $r \neq 1$) .

- If $-1 < r < 1$ then $\lim_{n\to\infty} r^n = 0$, and so it follows that

$$\sum_{i=0}^{\infty} ar^i = \lim_{n\to\infty} G_n = \frac{a}{1-r}.$$

- In the coin flipping example above, $a = \frac{1}{2}$ and $r = \frac{1}{2}$.

2.20 Theorem. If an experiment has N equally likely outcomes and A is an event made up of n of those outcomes then

$$P(A) = \frac{n}{N}.$$

2.21 Example. A five-card poker hand dealt from a deck of 52 playing cards is said to be a full house if it consists of three of a kind and a pair. For example:

$$8\spadesuit, J\spadesuit, 8\heartsuit, 8\clubsuit, J\diamondsuit$$

If all the five-card hands are equally likely, what is the probability of being dealt a full house?

Solution. The number of different full house hands similar to the one above, any three 8's and any two Jacks, is $\binom{4}{3}\binom{4}{2}$. (Ignores order in which they are dealt).

There are 13 possible choices for the three-of-a-kind card, leaving 12 possibilities for the two-of-a-kind card. So the total number of full house hands is $13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}$.

In total there are $\binom{52}{5}$ equally likely outcomes.

The probability of getting any full house is:

$$\frac{13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}}{\binom{52}{5}} = \frac{13 \cdot 4 \cdot 12 \cdot 6}{\left(\frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}\right)} = \frac{3744}{2598960} \approx 0.00144$$

□

Rules of Probability

2.22 Theorem. Let S be a sample space with probability measure P, and let A and B be events in S. Then

1. $P(A) + P(A') = 1$, or equivalently $P(A') = 1 - P(A)$.
2. $P(\emptyset) = 0$.
3. If $A \subset B$ then $P(A) \le P(B)$.
4. $0 \le P(A) \le 1$.
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(think of the Venn diagrams)

Rule 5, and its generalizations, are call the *inclusion-exclusion principle*.

2.23 Example. What is the probability that at least two people out of a group of r people have the same birthday for $r \le 365$ (ignoring leap years)?

Let A be the event that at least two people have the same birthday. There are 365^r possible birthday arrangements for r people.

There are ${}_{365}P_r = 365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365 - r + 1)$ ways that they could have distinct birthdays so $P(A') = \frac{365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365 - r + 1)}{365^r}$

This means that $P(A) = 1 - P(A') = 1 - \frac{365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365 - r + 1)}{365^r}$.

For $r = 23$, $P(A) \approx 0.507$.

2.24 Example. Suppose the probabilities are $0.86, 0.35$, and 0.29, respectively, that a family owns a laptop computer, a desktop computer, or both kinds. What is the probability that a family owns either or both kinds of computer and what is the probability that a family owns neither?

Let A be the event that family owns a laptop and B the event that a family owns a desktop. Then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.86 + 0.35 - 0.29 = 0.92.$$

The probability that a family owns neither is

$$P\left((A \cup B)'\right) = 1 - P(A \cup B) = 1 - 0.92 = 0.08.$$

2.25 Theorem. If A, B and C are any three events in sample space S, then

$$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

Again, this is known as the *inclusion-exclusion principle*. Similar rules can be obtained for any finite number of sets.

2.4 Conditional Probability

2.26 Example. Mayoral candidate Alice receives 56 percent of the entire vote, but only 47 percent of the female vote.

Let $P(A)$ be the probability that a randomly selected person has voted for Alice, and let $P(A|F)$ denote the probability that a randomly selected female has voted for Alice. So

$$P(A) = 0.56 \quad \text{and} \quad P(A|F) = 0.47$$

Value $P(A|F)$ is called the *conditional probability of A relative to F*, or the *conditional probability of A given F*.

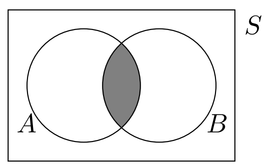
2.27 Example. Let A be the event of rolling 8 with two dice; Then $P(A) = \frac{5}{36} \approx 0.1389$.

Suppose we are given that the roll of die 1 is 3. Knowing this (i.e. given that this event has occurred), what is the probability of rolling an 8?

Let $B = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$; the event that die 1 is 3. Since these outcomes are equally likely prior to knowing die 1 is 3, they are still equally likely given that B has occurred. So given B has occurred, each have probability $\frac{1}{6}$ (the other 30 outcomes have probability 0).

Therefore the probability of rolling an 8 given that die 1 is a 3 is $P(A|B) = \frac{1}{6} \approx 0.1667$

In the example above, if B occurs, then in order for A to occur, the outcome must lie in both A and B. Thus $A \cap B$ becomes the event of interest, and B is considered the new sample space.



The conditional probability of A given B is the probability of $A \cap B$ relative to the probability of B.

2.28 Definition (Conditional Probability)**.** If A and B are events in S and $P(B) \neq 0$, then the *conditional probability* of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

2.29 Example. Results of a survey of 50 car dealerships:

	Good service under warranty	Poor service under warranty
In business 10 years or more	16	4
In business less than 10 years	10	20

If a person randomly chooses one dealership, what is the probability...

(a) ...they get one who provides good service under warranty?

(b) ...they get one who provides good service under warranty if they select from dealers in business 10 years or more?

Solution. Assume all choices are equally likely. Let G be the event of getting good service and T the event that a dealer has been in business 10 years or more. Let $n(A)$ denote the number of elements in event A.

Then probability of getting good service is

$$P(G) = \frac{n(G)}{n(S)} = \frac{16+10}{50} = 0.52.$$

Restricting the random selection to T, we have $n(T) = 16 + 4 = 20$ and $n(G \cap T) = 16$, so

$$P(G|T) = \frac{P(G \cap T)}{P(T)} = \frac{\frac{n(G\cap T)}{n(S)}}{\frac{n(T)}{n(S)}} = \frac{\frac{16}{50}}{\frac{20}{50}} = \frac{16}{20} = 0.80.$$

□

2.30 Example. A coin is tossed twice. Assuming all outcomes in the sample space

$$S = \{HH, HT, TH, TT\}$$

are equally likely, what is the probability that both flips land on heads given that...

(a) ...the first flip is heads?

(b) ...at least one flip is heads?

Solution. Let

$$A = \{HH\} \text{ - event that both flips are heads,}$$

$$B = \{HH, HT\} \text{ even that first flip is heads,}$$

$$C = \{HH, HT, TH\} \text{ - event that at least one flip is heads.}$$

(a)

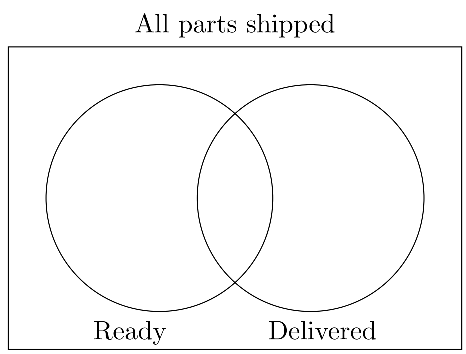
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{HH\})}{P(B)} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2} = 0.5.$$

(b)

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(\{HH\})}{P(C)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \approx 0.333.$$

□

2.31 Example. A manufacturer of airplane parts knows from past experience that the probability is 0.80 that an order will be ready for shipment on time, and it is 0.72 that an order will be ready for shipment on time and will also be delivered on time. What is the probability that such an order will be delivered on time given that it was ready for shipment on time?



Conditional Probability Multiplication Rule

2.32 Theorem (Multiplication Rule)**.** If A and B are events in S and $P(A) \neq 0$, then

$$P(A \cap B) = P(A) \cdot P(B|A).$$

i.e. The probability that both A and B will occur is the product of the probability of A and the probability of B given A.

2.33 Example. A pot contains 8 red balls and 4 green balls. We draw 2 balls without replacement. If each ball has an equally likely chance of being chosen, what is the probability that both balls are red? (*without replacement means that the first ball is not returned to the pot before the second ball is drawn*)

Solution. Let R_1 be the event that ball 1 is red, and R_2 be the event that ball 2 is red. Then $R_1 \cap R_2$ is the event that both are red.

Then using the multiplication rule

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2|R_1).$$

For ball 1, $P(R_1) = \frac{8}{12}$.

Since ball 1 is chosen red we have 7 remaining red balls and 4 green balls and so $P(R_2|R_1) = \frac{7}{11}$. Thus

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2|R_1) = \frac{8}{12} \cdot \frac{7}{11} = \frac{14}{33} \approx 0.4242.$$

Since the the outcomes are equally likely, we could also have computed the probability as the number of successful outcomes over total number of outcomes:

$$P(R_1 \cap R_2) = \frac{\binom{8}{2}}{\binom{12}{2}} = \frac{28}{66} = \frac{14}{33} \approx 0.4242.$$

□

2.34 Example. Find the probabilities of randomly drawing two aces in succession from an ordinary deck of 52 playing cards if we sample...

(a) without replacement.

(b) with replacement.

Solution. We will present a couple of different ways one can arrive at the same answer.

(a) Let A_1 be the event that card 1 is an ace, and A_2 be the event that card 2 is an ace. Then $A_1 \cap A_2$ is the event that both are aces.

Since there are 4 aces (and outcomes are equally likely) $P(A_1) = \frac{4}{52}$.

Given that the first card drawn is an ace, the probability of drawing an ace for card 2, $P(A_2|A_1)$, is $\frac{3}{51}$.

By the multiplication rule we have:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1) = \frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221} \approx 0.0045$$

without replacement.

Here is another way to look at this problem: View the sample space S as the $52 \cdot 51$ possible outcomes, keeping track of the order in which the cards are drawn; e.g $(A\heartsuit, K\clubsuit)$ is different from $(K\clubsuit, A\heartsuit)$.

Event A_1 has $4 \cdot 51$ outcomes; all those in S where the first card is one of the four aces.

Event A_2 has $51 \cdot 4$ outcomes, and event $A_1 \cap A_2$ has $4 \cdot 3$ outcomes. Thus

$$P(A_1 \cap A_2) = \frac{12}{52 \cdot 51} \approx 0.0045.$$

Of course we could have computed the same probability as

$$P(A_1 \cap A_2) = \frac{\binom{4}{2}}{\binom{52}{2}} = \frac{6}{1326} \approx 0.0045.$$

(counting this way ignores order drawn) with replacement.

(b) The sample space is different from part (a), but we will use the same notation for events.

Again since there are 4 aces $P(A_1) = \frac{4}{52}$.

The probability of drawing an ace for card 2 given that card 1 is an ace $P(A_2|A_1)$, is also $\frac{4}{52}$, since that ace was put back.

By the multiplication rule we have:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1) = \frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169} \approx 0.0059$$

Alternative view: In this case the sample space S has $52 \cdot 52$ possible outcomes, again keeping track of the order in which the cards are drawn.

Event A_1 has $4 \cdot 52$ outcomes, event A_2 has $52 \cdot 4$ outcomes, and event $A_1 \cap A_2$ has $4 \cdot 4$ outcomes. Thus

$$P(A_1 \cap A_2) = \frac{16}{52 \cdot 52} \approx 0.0059.$$

□

Since $A \cap B \cap C = (A \cap B) \cap C)$ we have by the multiplication rule

$$P((A \cap B) \cap C)) = P(A \cap B) \cdot P(C|A \cap B).$$

Applying the multiplication rule again to $P(A \cap B)$ gives the following.

2.35 Theorem. If A, B and C are events in S and $P(A \cap B) \neq 0$, then

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B).$$

2.36 Example. A bushel of 126 apples contains 15 rotten ones. If three apples are drawn chosen at random, what is the probability that all three are rotten?

If A_1, A_2, A_3 are the events that the first, second and third (resp.) choice is rotten, then

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \\ &= \frac{15}{126} \cdot \frac{14}{125} \cdot \frac{13}{124} \\ &= \frac{2730}{1953000} \\ &\approx 0.0014. \end{aligned}$$

2.5 Independent Events

2.37 Example. Suppose a coin is tossed twice. What is the probability of getting tails on the second toss given that the first toss was tails?

Surely the outcome of the second toss doesn't depend on the what has previously come up. Indeed if $S = \{HH, HT, TH, TT\}$, $T_1 = \{TH, TT\}$, and $T_2 = \{HT, TT\}$ are the events of getting tails on flips 1, and 2 respectively, then

$$P(T_2|T_1) = \frac{P(T_1 \cap T_2)}{P(T_1)} = \frac{P(\{TT\})}{P(T_1)} = \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2} = P(T_2).$$

Similarly we see that $P(T_1|T_2) = P(T_1)$. In this case events T_1 and T_2 are called *independent*.

Replacing $P(B|A)$ with $P(B)$ in the multiplication rule gives us the formal definition of when A and B are considered independent events.

2.38 Definition. Events A and B are called *independent* if and only if

$$P(A \cap B) = P(A) \cdot P(B).$$

They are otherwise called *dependent*. (we allow $P(A) = 0$ or $P(B) = 0$)

2.39 Example. In the initial coin toss example, the probability of getting two consecutive tails is

$$P(T_1 \cap T_2) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(T_1) \cdot P(T_2).$$

2.40 Example. Suppose a coin a tossed 3 times. Let
$A = \{HHH, HHT\}$ - first two are H
$B = \{HHT, HTT, THT, TTT\}$ - third always T
$C = \{HTT, THT, TTH\}$ - exactly two T

(a) Show that A and B are independent.

(b) Show that B and C are dependent.

2.41 Example. In the example of drawing two aces without replacement

$$P(A_1 \cap A_2) = \frac{12}{52 \cdot 51} \neq \frac{4 \cdot 51}{52 \cdot 51} \cdot \frac{51 \cdot 4}{52 \cdot 51} = P(A_1) \cdot P(A_2).$$

We see that events A_1 and A_2 are dependent.

However, in drawing two aces with replacement

$$P(A_1 \cap A_2) = \frac{16}{52 \cdot 52} = \frac{4 \cdot 52}{52 \cdot 52} \cdot \frac{52 \cdot 4}{52 \cdot 52} = P(A_1) \cdot P(A_2)$$

events A_1 and A_2 (now considered in the replacement sample space) are independent.

2.42 Theorem. If A and B are independent then so are A and B'.

Proof. Since $A = (A \cap B) \cup (A \cap B')$ we have

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B') \quad \text{(mutally exclusive events)} \\ &= P(A) \cdot P(B) + P(A \cap B') \quad (A, B \text{ independent}). \end{aligned}$$

Rearrange this equation to get

$$\begin{aligned} P(A \cap B') &= P(A) - P(A) \cdot P(B) \\ &= P(A) \cdot (1 - P(B)) \\ &= P(A) \cdot P(B'). \end{aligned}$$

□

2.43 Definition. Events $A_1, A_2, \ldots, A_k$ are *independent* if and only if the probability of the intersection of any $2, 3, \ldots, k$ of these is equal to the product of their individual probabilities.

2.44 Example. Three events A_1, A_2, A_3 are independent if and only if

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$$

$$\begin{aligned} P(A_1 \cap A_3) &= P(A_1) \cdot P(A_3) \\ P(A_2 \cap A_3) &= P(A_2) \cdot P(A_3) \\ P(A_1 \cap A_2 \cap A_3) &= P(A_1) \cdot P(A_2) \cdot P(A_3) \end{aligned}$$

2.45 Exercise. Let $S = \{a, b, c, d\}$ be the sample for space for an experiment with equally likely outcomes and define events

$$A = \{a, d\}, \quad B = \{b, d\}, \quad C = \{c, d\}.$$

Show that A, B, C are pairwise independent, but not independent.

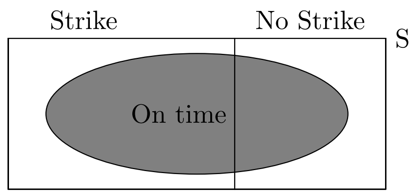
2.46 Exercise. Find the probabilities of getting

(a) three heads in three random tosses of a balanced coin;

(b) four sixes and then another number in five random rolls of a balanced die.

2.6 Rule of Total Probability

2.47 Example. The completion of a construction job may be delayed because of a strike. The probabilities are 0.60 that there will be a strike, 0.85 that the construction job will be completed on time if there is no strike, and 0.35 that the construction job will be completed on time if there is a strike. What is the probability that the construction job will be completed on time?



Solution.

Let A be the event that the job will be completed on time, B the event of a strike, therefore B' is the event of no strike.

We want $P(A)$ and are given

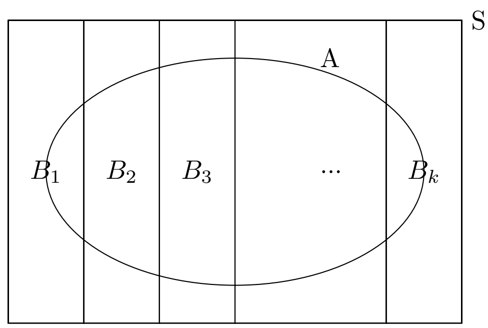
$$P(B) = 0.60, \quad P(A|B) = 0.35, \quad P(A|B') = 0.85.$$

Using the fact that $A = (A \cap B) \cup (A \cap B')$ (union of mutually exclusive events) and the multiplicative rule, we have

$$P(A) = P(A \cap B) + P(A \cap B') = P(B) \cdot P(A|B) + P(B') \cdot P(A|B').$$

Thus $P(A) = (0.6)(0.35) + (0.4)(0.85) = 0.55$. □

We can generalize the idea above to obtain a formula for the probability of any event, given that we have a partition of our sample space into event of known probability.



(a partition of a set S is a collection of pairwise disjoint subsets whose union is S)

2.48 Theorem (Rule of Total Probability)**.** Suppose events $B_1, B_2, \ldots B_k$ form a partition of the sample space S, and $P(B_i) \neq 0$ for $i = 1, \ldots, k$. Then for any event A in S,

$$P(A) = \sum_{i=1}^{k} P(B_i) \cdot P(A|B_i).$$

2.49 Example. Three machine M_1, M_2 and M_3 produce respectively 40, 10 and 50 percent of the items in a factory. The percentage of defective items produced by each respective machine is 2, 3 and 4 percent. Find the probability that a randomly selected item from the factory is defective.

Let D denote the event that a randomly selected item is defective, then

$$\begin{aligned}P(D) &= P(M_1)\cdot P(D|M_1) + P(M_2)\cdot P(D|M_2) + P(M_3)\cdot P(D|M_3)\\ &= (0.40)(0.02) + (0.10)(0.03) + (0.50)(0.04)\\ &= 0.031.\end{aligned}$$

Bayes' Theorem

With reference to the example above one might ask: If the randomly selected item is defective, what is the probability that the item was produced by

(a) machine M_1,

(b) machine M_2, or

(c) machine M_3?

This question is answered by *Bayes' Theorem*.

2.50 Theorem (Bayes' Theorem)**.** Suppose events $B_1, B_2, \ldots B_k$ form a partition of the sample space S, and $P(B_i) \neq 0$ for $i = 1, \ldots, k$. Then for any event A in S with $P(A) \neq 0$

$$P(B_r|A) = \frac{P(B_r)\cdot P(A|B_r)}{\sum_{i=1}^{k} P(B_i)\cdot P(A|B_i)}$$

for $r = 1, \ldots, k$.

Proof.

$$\begin{aligned}P(B_r|A) &= \frac{P(B_r\cap A)}{P(A)} \quad \text{(by definition)}\\ &= \frac{P(B_r)\cdot P(A|B_r)}{P(A)} \quad \text{(multiplication rule)}\\ &= \frac{P(B_r)\cdot P(A|B_r)}{\sum_{i=1}^{k} P(B_i)\cdot P(A|B_i)} \quad \text{(rule of total probability)}\end{aligned}$$

□

2.51 Example. Three machine M_1, M_2 and M_3 produce respectively 40, 10 and 50 percent of the items in a factory. The percentage of defective items produced by each respective machine is 2, 3 and 4 percent. If the randomly selected item is defective, what is the probability that the item was produced by

(a) machine M_1,

(b) machine M_2, or

(c) machine M_3?

Solution. Using Bayes' Theorem,

(a)

$$P(M_1|D) = \frac{P(M_1) \cdot P(D|M_1)}{P(D)} = \frac{(0.40)(0.02)}{0.031} \approx 0.2581$$

Parts (b) and (c) are similar. □

Chapter 3

Probability Distributions and Densities

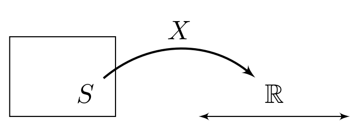
The raw outcomes of a given probability experiment can contain a wealth of information, however we may only be interested in a few specific attributes.

For example

- In rolling two dice, we might only care about the sum of the outcome, and not the individual values.
- In a random sample of bottled water, we might want to know the volume of a certain chemical, but not the price.
- In a randomly chosen family, we might want to know their joint income, but not their address or hobbies.

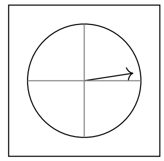
3.1 Random Variables

Let S be a sample space with a probability measure. A *random variable* is a function $X : S \to \mathbb{R}$, which maps the outcomes in the sample space to real numbers. The output of a random variable is something we can measure.



Random variables are defined when we want to focus on a particular property of the outcomes of an experiment. More than one random variable can be defined for a given sample space.

Capital letters, e.g. "X", denote random variables, and their lower case letter, e.g. "x", for particular values that X can take.



3.1 Example. Earlier we mentioned the experiment of spinning a probability spinner, and gave sample space $\{\theta \text{ degrees} | \theta \in [0, 360)\}$.

However, the actual sample space could include more information, such as multiple rotations, angular velocity at time t, elapsed time, the colour it landed on, etc.

A random variable focuses on one property of the outcome that can be assigned a real number. Examples of random variables:

- X_1: resting position (degrees), outputs values in $[0, 360)$.
- X_2: resting position (radians), outputs values in $[0, 2\pi)$.
- X_3: angle of rotation (radians), outputs values in $(-\infty, \infty)$.
- X_4: number of full rotations, can take values $0, 1, 2, 3, \ldots$.
- X_5: points for each coloured space; e.g. 1-red, 2-blue,...

3.2 Example. Two socks are selected at random and removed in succession from a drawer containing five brown socks and three green socks.

List the elements of the sample space, the corresponding probabilities, and the corresponding values x of the random variable X, where X is the number of brown socks selected.

Solution.

Element of sample space	Probability	x
BB	$\frac{20}{56}$	2
BG	$\frac{15}{56}$	1
GB	$\frac{15}{56}$	1
GG	$\frac{6}{56}$	0

We write: $P(X = 2) = \frac{20}{56}$, $P(X = 1) = \frac{30}{56}$, $P(X \leq 1) = \frac{36}{56}$. □

3.3 Exercise. Three balls are randomly chosen (without replacement) from a bag of 20 balls numbered 1-20. We bet that at least one of the numbers drawn is as large, or larger than 17. What is the probability of winning the bet?

Outcomes in the sample space are subsets of three numbered balls, and they are all equally likely to occur.

Let random variable X denote the largest number of the three selected. Thus X takes values $3, 4, \ldots, 20$, and we want $P(X \geq 17)$.

By the rule of equal probability we have for $i = 3, \ldots, 20$,

$$P(X = i) = \frac{\overbrace{\binom{1}{1}}^{\text{ball } i} \cdot \overbrace{\binom{i-1}{2}}^{\text{any two} < i}}{\binom{20}{3}} = \frac{\binom{i-1}{2}}{1140}.$$

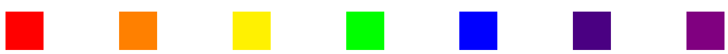
$$P(X \geq 17) = P(X = 17) + P(X = 18) + P(X = 19) + P(X = 20).$$

$$= \frac{120}{1140} + \frac{136}{1140} + \frac{153}{1140} + \frac{171}{1140} \approx 0.5088$$

Discrete Random Variable

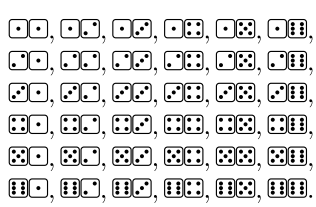
Recall that the set of all possible output values of a function is called its *range*.

If the range of a random variable X is a finite or countably infinite set, then we say that X is a *discrete random variable*.



In contrast, a *continuous random variables* is one whose range is a continuum of values, like an interval or a union of intervals in $\mathbb{R}$. We will deal with this type later. The important difference to notice is in how the probabilities are assigned.

3.2 Probability Distributions



3.4 Example. Experiment: Rolling two dice

Let random variable X denote the sum of a roll. The range of X is $\{2, 3, \ldots, 12\}$.

Knowing that each outcome in the sample space has probability $\frac{1}{36}$, we can automatically find the probability that X takes on any value in its range. e.g. $P(X = 7) = \frac{6}{36}$, $P(X = 11) = \frac{2}{36}$.

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

This information is summarized in the table.

It is sometimes preferable to have a formula describing these probabilities instead of simply listing them individually. i.e. we would like an algebraic expression which gives $P(X = x)$ for each value x in the range of random variable X. In this case the probabilities are given by the expression

$$f(x) = \frac{6 - |x - 7|}{36}.$$

(verify this by substitution) We may not always be able to obtain an expression like this.

If X is a discrete random variable, the function f given by

$$f(x) = P(X = x)$$

for each x in the range of X, is called the *probability distribution of* X.

3.5 Theorem. A function f is allowable as a probability distribution for discrete random variable X if and only if its values, $f(x)$, satisfy

1. $f(x) \geq 0$ for any x,
2. $\sum_x f(x) = 1$, (sum taken over all x in the range of X)

3.6 Example. Let X be the number heads obtained in tossing a balanced coin 4 times.

(a) What is the range of X?

(b) What is $P(X = x)$ for each x in the range of X?

(c) Find a formula for the probability distribution of X.

Solution. (a) The range of X is $\{0, 1, 2, 3, 4\}$.

(b) Individual probabilities:

x	$P(X = x)$
0	1/16
1	4/16
2	6/16
3	4/16
4	1/16

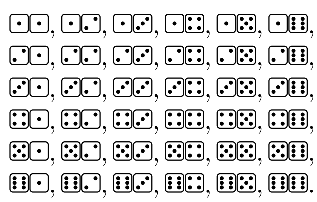
(this is also called the probability distribution of X)

(c) The probability distribution of X is also given by the formula

$$f(x) = \frac{\binom{4}{x}}{16}$$

for $x = 0, 1, 2, 3, 4$.

□



3.7 Example. Return to the dice rolling experiment.

Let Y be the maximum that either die shows in a single roll; $Y(a, b) = \max(a, b)$, for example $Y(3, 5) = 5$.

(a) What is the range of Y?

(b) What is $P(Y = y)$ for each y in the range of Y?

(c) Find a formula for the probability distribution of Y.

Solution. (a) The range of Y is $\{1, 2, 3, 4, 5, 6\}$.

(b) Probability distribution:

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

(c) The probability distribution of Y is also given by

$$g(y) = \frac{2y-1}{36}$$

for $y = 1, 2, 3, 4, 5, 6.$

□

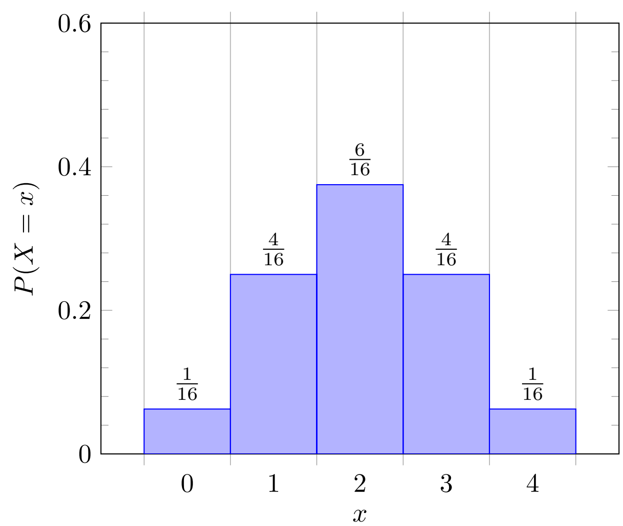
3.8 Exercise. Check whether the function given by

$$f(x) = \frac{x+2}{25},$$

for $x = 1, 2, 3, 4, 5$ can serve as the probability distribution of a discrete random variable.

Probability Histogram

Probability distributions for a random variable, say X, may be represented graphically by means of a probability histogram.



Each rectangle corresponds to a value for X, its height is $P(X = x)$, and its width is 1, so that the area of each rectangle equals $P(X = x)$. The total area of the histogram is 1.

(The histogram above is for the number of heads in 4 coin flips.)

3.3 Cumulative Distribution (Discrete)

In many problems we are interested in the probability that the value of a random variable is less than or equal to (or "at most") some real number x. i.e. $P(X \leq x)$.

If X is a discrete random variable with probability distribution f, the function given by

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

for $x \in (-\infty, \infty)$, is called the *cumulative distribution of* X.

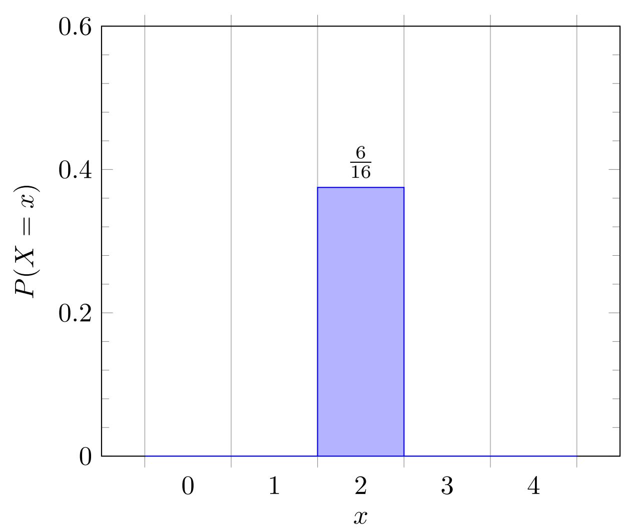
While x can be any real number, the t values in the sum are restricted to values in the domain of X. This is sometimes called the *distribution function.*

Sometimes the probability for a random variable is defined by the cumulative distribution function.

3.9 Example. The following probability histograms demonstrate the difference between a random variable's probability distribution "$f(x)$" and its cumulative distribution "$F(x)$".

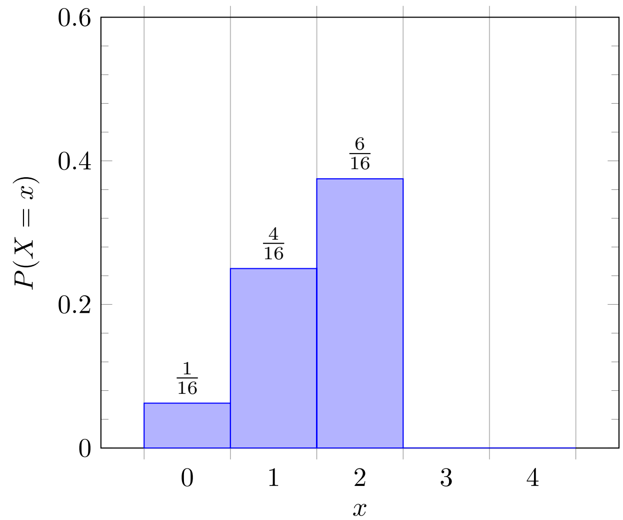
Probability distribution:

$$f(2) = \frac{6}{16}$$



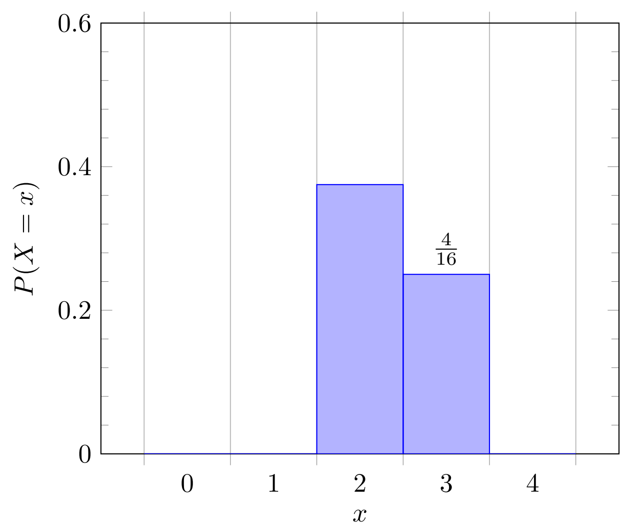
Cumulative distribution:

$$F(2) = \frac{11}{16}$$



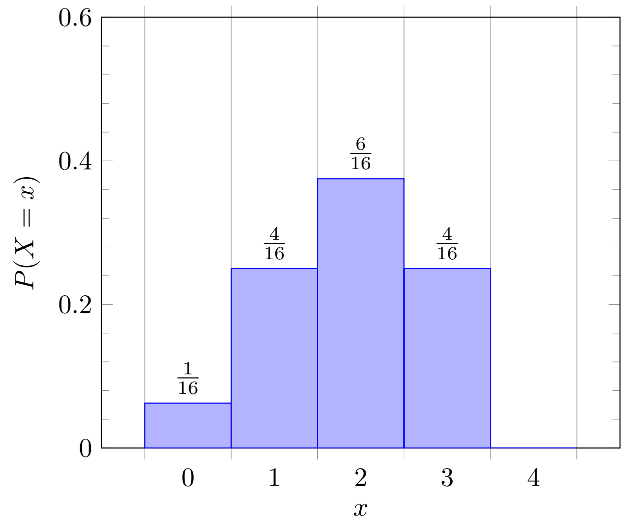
Probability distribution:

$$f(3) = \tfrac{4}{16}$$



Cumulative distribution:

$$F(3) = \tfrac{11}{16}$$

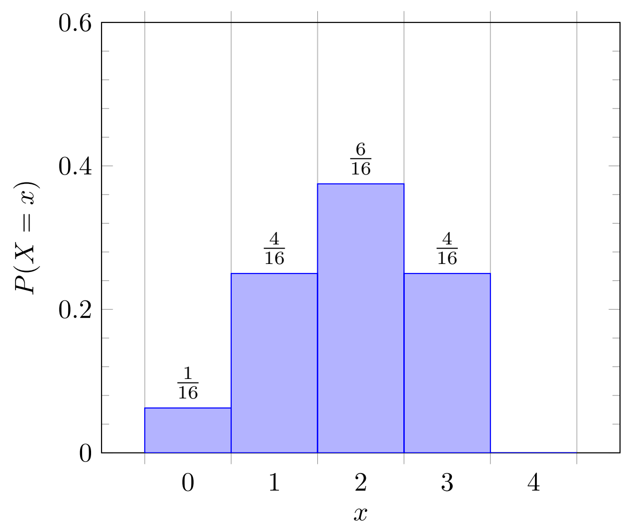


Note that the cumulative distribution function $F(x)$ is defined for all real x, not only those with nonzero probability. For example

$$F(3.6) = P(X \leq 3.6) = P(X \leq 3) + P(3 < X \leq 3.6) = P(X \leq 3) = F(3),$$

since $P(3 < X \leq 3.6) = 0$.

$$F(3.6) = \frac{11}{16}$$



3.10 Theorem. The cumulative distribution $F(x)$ satisfies

1. $F(-\infty) = 0$ and $F(\infty) = 1$.

2. If $a < b$ then $F(a) \leq F(b)$ for any $a, b \in \mathbb{R}$.

3.11 Example. Find the cumulative distribution (or the distribution function) for X in the brown green sock example.

Element of sample space	Probability	x
BB	$\frac{20}{56}$	2
BG	$\frac{15}{56}$	1
GB	$\frac{15}{56}$	1
GG	$\frac{6}{56}$	0

Solution. First note that the probability distribution f is given by

$$f(x) = \begin{cases} \frac{20}{56} & \text{for } x = 2 \\ \frac{30}{56} & \text{for } x = 1 \\ \frac{6}{56} & \text{for } x = 0 \end{cases}$$

We have

$$\begin{aligned} F(0) = f(0) &= \frac{6}{56} \\ F(1) = f(0) + f(1) &= \frac{6}{56} + \frac{30}{56} = \frac{36}{56} \\ F(2) = f(0) + f(1) + f(2) &= \frac{36}{56} + \frac{20}{56} = \frac{56}{56} \end{aligned}$$

and so

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{6}{56} & \text{for } 0 \leq x < 1 \\ \frac{36}{56} & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2 \end{cases}$$

□

3.12 Exercise. Suppose a random variable X has range $\{1, 2, 3, 4\}$. Define f by

$$f(1) = \frac{1}{4}, \quad f(2) = \frac{1}{2} \quad f(3) = \frac{1}{8}, \quad f(4) = \frac{1}{8}$$

(a) Show that f is a valid probability distribution for X.

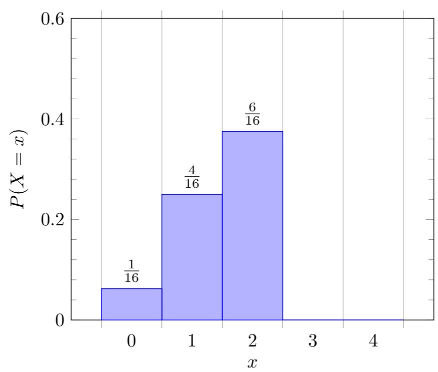
(b) Give the cumulative distribution (or the distribution function) for X.

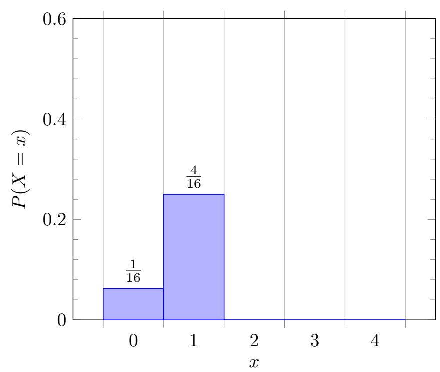
3.13 Theorem. If the range of a random variable X consists of the values $x_1 < x_2 < \cdots < x_n$, then $f(x_1) = F(x_1)$ and

$$f(x_i) = F(x_i) - F(x_{i-1})$$

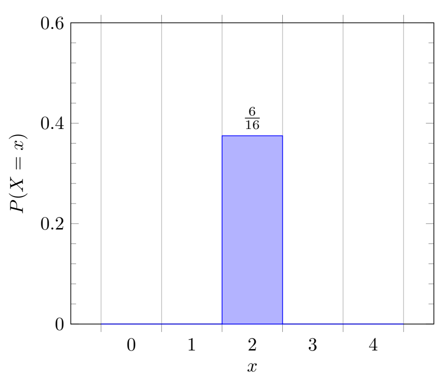
for $i = 2, 3, \ldots, n$.

3.14 Example. Below are the histograms representing $F(2)$ and $F(1)$.





Subtracting these give $f(2)$; i.e. $F(2) - F(1) = f(2)$.



3.15 Exercise. The cumulative distribution for a discrete random variable X is given by

$$F(x) = \begin{cases} 0 & \text{for } x < -2 \\ \frac{4}{18} & \text{for } -2 \leq x < -1 \\ \frac{7}{18} & \text{for } -1 \leq x < 0 \\ \frac{12}{18} & \text{for } 0 \leq x < 1 \\ 1 & \text{for } x \geq 1 \end{cases}$$

Find the probability distribution for X.

3.4 Continuous Random Variables

3.16 Example. On a 100 km stretch of rural road we are concerned with the possibility that a deer might cross. We are interested in the probability that it will occur at a given location or stretch of the road. The sample space for this experiment consists of all points in the interval from 0-100.

Suppose the probability that a deer crosses in any particular stretch of road is the length of that section divided by 100; this assumes that any point on the 100 km stretch has a equally likely chance of being crossed at, and it is guaranteed that the deer will cross at some point.

From point a to point b with $0 \leq a \leq b \leq 100$, is the interval $[a, b]$ and its length is given by $b - a$. Thus its probability is

$$P([a, b]) = \frac{b - a}{100}.$$

The probability of any two or more non overlapping intervals can be found by summing the probabilities of the connected components. Thus the probability measure proposed here has nonnegative values, assigns the entire sample space a probability of 1, and is countably additive; hence it satisfies our postulates of probability. We have taken the sample space to be any interval on this stretch of road, and the random variable X here is the function that assigns that interval to a real number in the interval $[0, 100]$. This is an example of a continuous random variable. We can give the probability that X lies within an interval by

$$P(a \leq X \leq b) = \frac{b - a}{100}$$

for $a < b$. Notice that the probability that X is any single point is zero.

3.5 Probability Density Function

In the case of a continuous random variable, probabilities cannot simply be assigned to every individual outcome as is done with a discrete random variable.

Therefore a continuous random variable must be accompanied by a *probability density function* in order to compute probabilities.

3.17 Definition. A positive valued function f defined on $\mathbb{R}$ is called a *probability density function* for continuous random variable X, if and only if

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

for any $a, b \in \mathbb{R}$ with $a \leq b$. These are also called "p.d.f's" for short.

Note that $f(r)$ does not give the probability that $X = r$.

3.18 Example. In the deer crossing example, the p.d.f. for X is $f(x) = \frac{1}{100}$.

For example

$$\begin{aligned} P(35 \leq X \leq 50) &= \int_{35}^{50} \frac{1}{100}\, dx \\ &= \frac{x}{100}\bigg|_{35}^{50} \\ &= \frac{50 - 35}{100} \\ &= \frac{15}{100}. \end{aligned}$$

The next Theorem comes from properties of definite integrals.

3.19 Theorem. Let X be a continuous random variable. If $a, b \in \mathbb{R}$ with $a \leq b$ then

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b).$$

From the postulates of probability we obtain the following.

3.20 Theorem. A function f can serve as a probability density function for X only if it satisfies

1. $f(x) \geq 0$ for all $x \in \mathbb{R}$.
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1$.

3.21 Exercise. Consider the function

$$f(x) = \begin{cases} 3x^2 & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Show that f is permissible as a probability density function.

(b) Use f to compute $P(0.1 < x < 0.5)$.

(c) Sketch the graph of f and indicate the area which represents the probability in (b).

3.22 Example. If X has probability density function

$$f(x) = \begin{cases} k \cdot e^{-3x} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

determine an appropriate $k \in \mathbb{R}$ and compute $P(0.5 \leq X \leq 1)$.

Solution. Solve for k using condition 2. from the theorem.

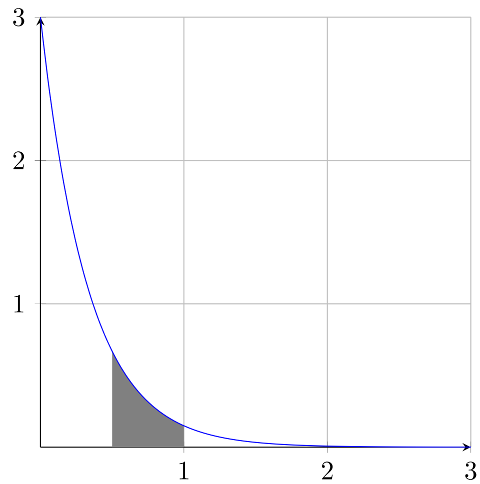
$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{\infty} k \cdot e^{-3x}\,dx \\ &= \lim_{c\to\infty} k\,\frac{e^{-3x}}{(-3)}\Big|_0^c \\ &= \lim_{c\to\infty} k\frac{e^{-3c}}{(-3)} - k\frac{e^{-3(0)}}{(-3)} \\ &= \frac{k}{3}. \qquad \text{(since } \lim_{r\to\infty} e^{-r} = 0) \end{aligned}$$

Thus $k = 3$. Now we can compute

$$P(0.5 \leq X \leq 1) = \int_{0.5}^{1} f(x)\,dx = \int_{0.5}^{1} 3e^{-3x}\,dx = -e^{-3x}\Big|_{0.5}^{1}$$

$$= -e^{-3} - (-e^{-1.5}) \approx 0.1733$$

Below is a plot of $3e^{-3x}$.



The shaded area is $P(0.5 \le X \le 1)$. □

3.6 Cumulative Distribution (Continuous)

Let X be a continuous random variable with probability density function f. Then the function

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

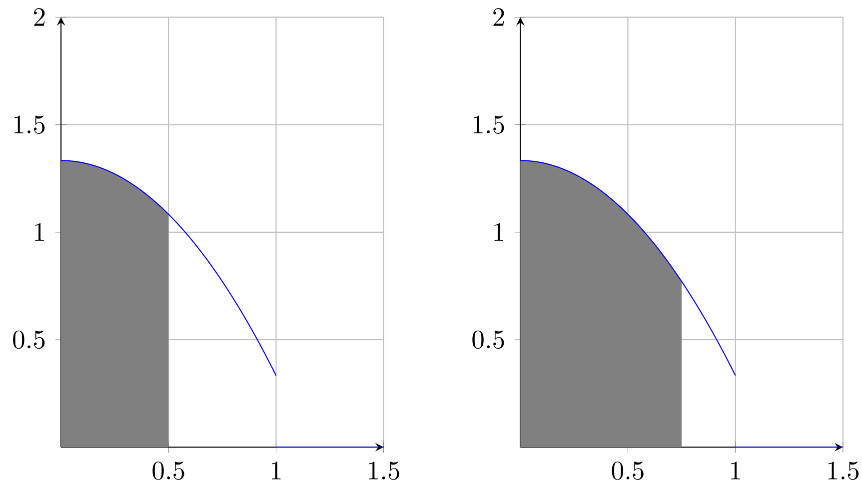
for all $x \in \mathbb{R}$, is called the *cumulative distribution function of* X.

3.23 Example. Random variable X with p.d.f. (plotted in blue)

$$f(x) = \begin{cases} -x^2 + \frac{4}{3} & \text{for } 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

Cumulative distribution $F(x) = \int_{-\infty}^{x} f(t)\,dt$.

The shaded areas in the graphs below represent values $F(0.5)$ and $F(0.75)$ respectively.



From the properties of integrals we have the following.

3.24 Theorem. If continuous random variable X has probability density function $f(x)$ and cumulative distribution function $F(x)$ then

$$P(a \leq X \leq b) = F(b) - F(a)$$

for any $a, b \in \mathbb{R}$ with $a \leq b$, and

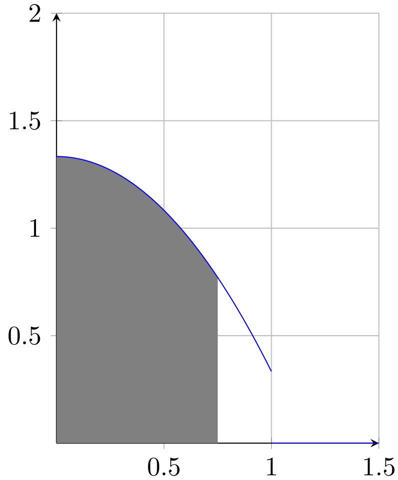
$$f(x) = \frac{d}{dx} F(x)$$

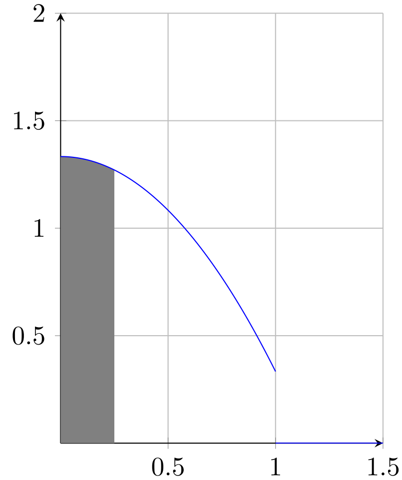
where derivative exists.

3.25 Example. In the previous example we had probability density $f(x) = -x^2 + \frac{4}{3}$ for $0 \leq x \leq 1$ and 0 elsewhere. By the theorem we have,

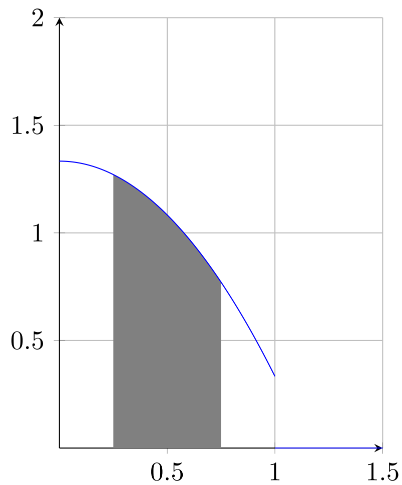
$$P(0.25 \leq X \leq 0.75) = F(0.75) - F(0.25)$$

We can visualize the result by subtracting the shaded areas under the curve in the plots of the cumulative distribution.





In the figures above, subtract the shaded area on the right from the shaded area on the left to get the shaded are shown below



The cumulative distribution function is

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} -t^2 + \frac{4}{3}\,dt = -\frac{t^3}{3} + \frac{4t}{3}\bigg|_{0}^{x} = -\frac{x^3}{3} + \frac{4x}{3},$$

for $0 \leq x \leq 1$; $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$. Its derivative is the probability density function

$$\frac{d}{dx}F(x) = \frac{d}{dx}\left(-\frac{x^3}{3} + \frac{4x}{3}\right) = -x^2 + \frac{4}{3} = f(x).$$

3.26 Example. Find the cumulative distribution function $F(x)$ for

$$f(x) = \begin{cases} 3e^{-3x} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

and use it to evaluate $P(0.5 \le X \le 1)$ via the theorem above.

Solution. For $x > 0$ we have

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{0}^{x} 3e^{-3t}\, dt = -e^{-3t}\Big|_0^x = -e^{-3x} + 1.$$

For $x \le 0$, $f(x) = 0$, and so

$$F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ 1 - e^{-3x} & \text{for } x > 0 \end{cases}$$

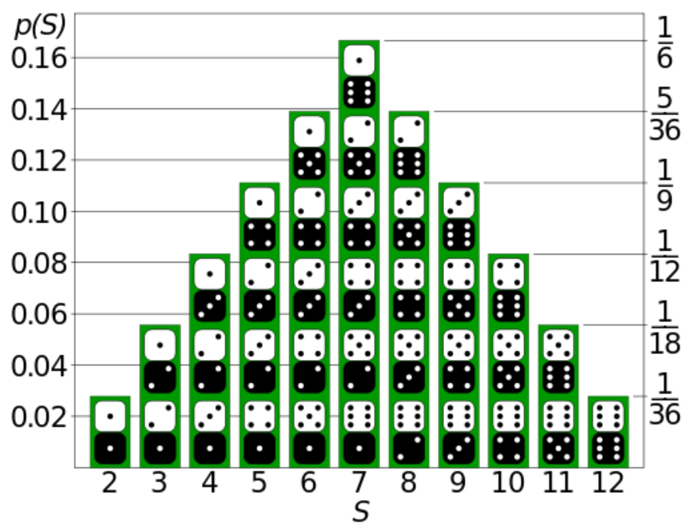
Then using the theorem

$$P(0.5 \le X \le 1) = F(1) - F(0.5) = (1 - e^{-3}) - (1 - e^{-1.5}) \approx 0.1733.$$

□

3.7 Multivariate Distributions

Histogram showing the probability distribution for the roll of two dice:

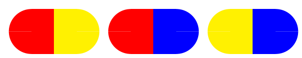


We now consider the case when two or more random variables are defined on the same (joint) sample space. We start with the *bivariate* case, that is when two random variables X and Y are defined for a common sample space.

For example X could be the sum of rolling two dice, and Y could be the product.

Write $P(X = x, Y = y)$ for the probability of the intersection of events $X = x$ and $Y = y$.

3.27 Example.



Two caplets are randomly selected from a bottle containing 3 asprin, 2 sedative, and 4 laxitive. Let X be the number of asprin, and Y be the number of sedative drawn (of the two). Find the probabilities associated to each possible pair of values for X and Y.

The possible pairs for X, Y are: $(0, 0), (1, 0), (0, 1), (2, 0), (1, 1), (0, 2)$.

There are $\binom{9}{2} = 36$ different possible two-pill selections that can be drawn.

The number of different ways to draw x asprin, y sedative, and therefore $2 - x - y$ laxitive (where $0 \leq x + y \leq 2$) is

$$\binom{3}{x}\binom{2}{y}\binom{4}{2-x-y}.$$

Thus, for the (x, y) pairs above,

$$P(X = x, Y = y) = \frac{\binom{3}{x}\binom{2}{y}\binom{4}{2-x-y}}{36}$$

We summarize the probabilities in a table

y \ x	0	1	2
0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$
1	$\frac{8}{36}$	$\frac{6}{36}$	
2	$\frac{1}{36}$		

3.8 Joint Probability Distributions

3.28 Definition. If X and Y are discrete random variables, the function

$$f(x, y) = P(X = x, Y = x)$$

for each pair (x, y) in the range of X and Y is called the *joint probability distribution of X and Y*.

3.29 Theorem. A bivariate function f can serve as a joint probability distribution for discrete random variables X and Y if and only if

1. $f(x, y) \geq 0$.
2. $\sum_x \sum_y f(x, y) = 1$, where the sums are taken over all possible pairs (x, y).

3.30 Example.

	$x=0$	$x=1$	$x=2$
y = 0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$
y = 1	$\frac{8}{36}$	$\frac{6}{36}$	
y = 2	$\frac{1}{36}$		

To verify the theorem for the caplet example, note that all values are positive and

$$\sum_x \sum_y f(x, y) = f(0, 0) + f(1, 0) + f(0, 1) + f(2, 0) + f(1, 1) + f(0, 2)$$

$$= \frac{6}{36} + \frac{12}{36} + \frac{3}{36} + \frac{8}{36} + \frac{6}{36} + \frac{1}{36} = 1.$$

3.31 Exercise. Suppose the joint probability distribution of discrete random variables X and Y is given by

$$f(x, y) = c(x^2 + y^2)$$

for all pairs (x, y) with $x = -1, 0, 1, 3$ and $y = -1, 2, 3$. Find the value of $c \in \mathbb{R}$.

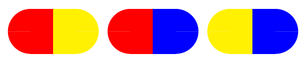
3.9 Joint Cumulative Distribution (Discrete)

3.32 Definition. If X and Y are discrete random variables, with joint probability distribution f, then the function

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{s \leq x} \sum_{t \leq y} f(s, t)$$

defined for all $x, y \in \mathbb{R}$, is called the *joint cumulative distribution of X and Y*, or the *joint distribution function*.

3.33 Example.



Below is the joint distribution of X and Y for the caplet example.

y \ x	0	1	2
0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$
1	$\frac{8}{36}$	$\frac{6}{36}$	
2	$\frac{1}{36}$		

Let $F(x, y)$ be the joint cumulative distribution of the caplet example. Find $F(2.3, 1.1)$.

Solution. To find $F(2.3, 1.1) = P(X \leq 2.3, Y \leq 1.1)$ we must sum the probabilities $f(x, y)$ over all pairs (x, y) in the range of X and Y with $x \leq 2.3$ and $y \leq 1.1$.

The pairs included here are $(0, 0), (1, 0), (0, 1), (2, 0), (1, 1)$.

Therefore

$$F(2.3, 1.1) = f(0, 0) + f(0, 1) + f(1, 0) + f(1, 1) + f(2, 0)$$
$$\frac{6}{36} + \frac{8}{36} + \frac{12}{36} + \frac{6}{36} + \frac{3}{36} = \frac{35}{36}$$

□

As with the single variable case we have the following properties.

3.34 Theorem. If $F(x, y)$ is the joint cumulative distribution for discrete random variables X and Y then

1. $F(-\infty, -\infty) = 0$
2. $F(\infty, \infty) = 1$
3. If $a \leq c$ and $b \leq d$ then $F(a, b) \leq F(c, d)$.

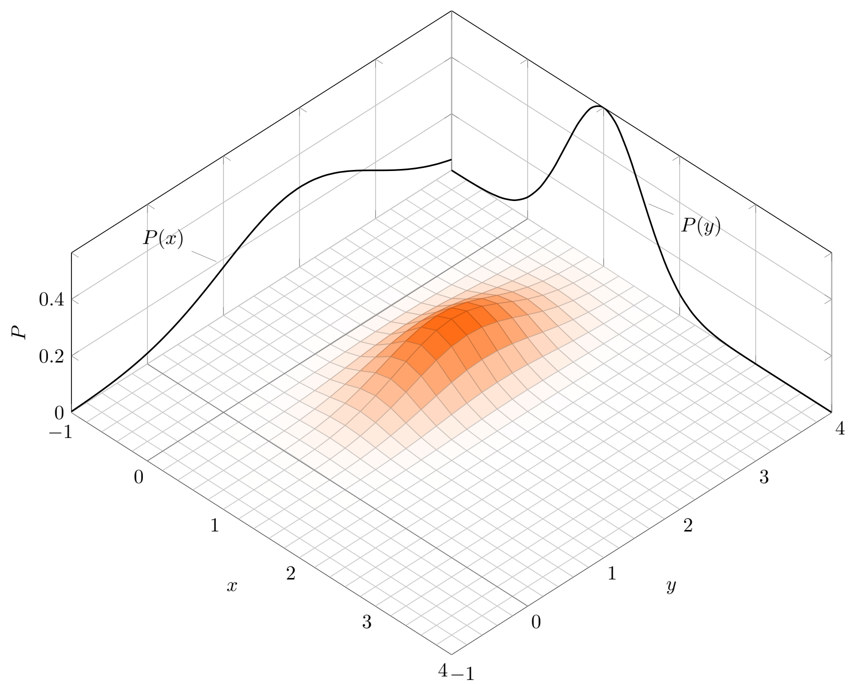
3.10 Joint Probability Density Function

3.35 Definition. We say that random variables X and Y are *jointly continuous* if there exists a function $f(x, y)$ defined for all $x, y \in \mathbb{R}$, such that

$$P((X,Y) \in A) = \iint_{(x,y)\in A} f(x,y)\,dx\,dy$$

for any region A in the xy-plane. The function $f(x, y)$ is called the *joint probability density function of X and Y.*

Plot of a 2-variable joint density function:



3.36 Theorem. A bivariate function f can serve as a joint probability density function of a pair of continuous random variables X and Y if it satisfies:

1. $f(x, y) \geq 0$ for all $x, y \in \mathbb{R}$.
2. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1.$

3.37 Example. Let

$$f(x,y) = \begin{cases} \frac{3}{5}x(y+x) & \text{for } 0 < x < 1, 0 < y < 2 \\ 0 & \text{elsewhere} \end{cases}$$

1. Verify that f can serve as a probability density function for two jointly continuous random variables X and Y.

2. For $A = \{(x, y)|0 < X < \frac{1}{2}, 1 < Y < 2\}$ find $P((X, Y) \in A)$.

Solution.

$$f(x,y) = \begin{cases} \frac{3}{5}x(y+x) & \text{for } 0 < x < 1, 0 < y < 2 \\ 0 & \text{elsewhere} \end{cases}$$

We see that $f(x, y) \geq 0$ for all $0 < x < 1, 0 < y < 2$.

Next we integrate over the entire plane. First we see that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ dx\ dy = \int_0^2\int_0^1 \frac{3}{5}x(y+x)\ dx\ dy$$

since $f(x, y) = 0$ for all other regions.

We proceed by integrating first with respect to x, treating y as constant:

$$\begin{aligned}\int_0^2\left(\int_0^1 \frac{3}{5}(yx+x^2)\ dx\right)\ dy &= \frac{3}{5}\int_0^2\left(y\frac{x^2}{2}+\frac{x^3}{3}\bigg|_0^1\ dx\right)\ dy \\ &= \frac{3}{5}\int_0^2 \frac{y}{2}+\frac{1}{3}\ dy\end{aligned}$$

Finally integrate with respect to y.

$$\begin{aligned}\frac{3}{5}\int_0^2 \frac{y}{2}+\frac{1}{3}\ dy &= \frac{3}{5}\left(\frac{y^2}{4}+\frac{y}{3}\bigg|_0^2\right) \\ &= \frac{3}{5}\left(\frac{4}{4}+\frac{2}{3}\right) \\ &= 1.\end{aligned}$$

Therefore $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ dx\ dy = 1$ as required.

$$
\begin{aligned}
P((X,Y) \in A) &= \int_1^2 \int_0^{\frac{1}{2}} \frac{3}{5} x(y+x)\, dx\, dy \\
&= \frac{3}{5} \int_0^2 \left(y\frac{x^2}{2} + \frac{x^3}{3}\Big|_0^{\frac{1}{2}} dx \right) dy \\
&= \frac{3}{5} \int_1^2 \frac{y}{8} + \frac{1}{24}\, dy \\
&= \frac{3}{5} \left(\frac{y^2}{16} + \frac{y}{24}\Big|_1^2 \right) \\
&= \frac{3}{5} \left(\frac{4}{16} + \frac{2}{24} - \frac{1}{16} + \frac{1}{24} \right) \\
&= \frac{11}{80}
\end{aligned}
$$

□

3.11 Joint Cumulative Distribution (Continuous)

3.38 Definition. If X and Y are jointly continuous random variables, with joint probability density f, the function given by

$$F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t)\, ds\, dt$$

for $x, y \in \mathbb{R}$, is called the *joint cumulative distribution function of* X *and* Y (or simply the *joint distribution function*).

As with the discrete case we have that

- $F(-\infty, -\infty) = 0$
- $F(\infty, \infty) = 1$
- If $a \le c$ and $b \le d$ then $F(a,b) \le F(c,d)$.

It also follows that $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$ is the joint probability density function.

3.39 Example. If the joint probability density function of X and Y is given by

$$f(x,y) = \begin{cases} 4xy & \text{for } 0 < x < 3, 0 < y < \frac{1}{3} \\ 0 & \text{elsewhere} \end{cases}$$

find $F(x,y)$.

Solution. To find $F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t)\, ds\, dt$ we must consider different regions in the plane where $f(x,y)$ is defined.

If either $x < 0$ or $y < 0$ then $f(x, y) = 0$ and so $F(x, y) = 0$.

If $0 < x < 3$ and $0 < y < \frac{1}{3}$ then

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)\, ds\, dt = \int_{0}^{y} \int_{0}^{x} 4st\, ds\, dt = x^2 y^2$$

If $x \geq 3$ and $0 < y < \frac{1}{3}$ then

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)\, ds\, dt = \int_{0}^{y} \int_{0}^{3} 4st\, ds\, dt = 9y^2$$

If $0 < x < 3$ and $y \geq \frac{1}{3}$ then

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)\, ds\, dt = \int_{0}^{\frac{1}{3}} \int_{0}^{x} 4st\, ds\, dt = \frac{x^2}{9}$$

Finally if $x \geq 3$ and $y \geq \frac{1}{3}$ then

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)\, ds\, dt = \int_{0}^{\frac{1}{3}} \int_{0}^{3} 4st\, ds\, dt = 1$$

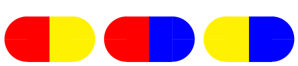
In summary

$$F(x, y) = \begin{cases} 0 & \text{for } x < 0 \text{ or } y < 0 \\ x^2 y^2 & \text{for } 0 < x < 3, 0 < y < \frac{1}{3} \\ 9y^2 & \text{for } x \geq 3, 0 < y < \frac{1}{3} \\ \frac{x^2}{9} & \text{for } 0 < x < 3, y \geq \frac{1}{3} \\ 1 & \text{for } x \geq 3, y \geq \frac{1}{3} \end{cases}$$

□

Joint probability distributions/densities can be defined similarly for three or more random variables, though we won't spend time on this here.

3.12 Marginal Distributions



Returning to the caplet example, sum the rows and columns of the table of probabilities.

		x = 0	1	2	
	0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$	$\frac{21}{36}$
y	1	$\frac{8}{36}$	$\frac{6}{36}$		$\frac{14}{36}$
	2	$\frac{1}{36}$			$\frac{1}{36}$
		$\frac{15}{36}$	$\frac{18}{36}$	$\frac{3}{36}$	

The column sums are the probabilities that $X = 0, 1, 2$ respectively, and the row sums are probabilities that $Y = 0, 1, 2$ respectively. i.e. the column totals are the probability distribution for X: for $x = 0, 1, 2$

$$g(x) = P(X = x) = \sum_{y=0}^{2} f(x, y),$$

and the row totals are the probability distribution for Y: for $y = 0, 1, 2$

$$h(y) = P(Y = y) = \sum_{x=0}^{2} f(x, y).$$

We summarize the marginal distributions as follows:

Marginal distribution for X:

$$g(0) = \frac{15}{36}, \quad g(1) = \frac{18}{36}, \quad g(2) = \frac{3}{36}.$$

Marginal distribution for Y:

$$h(0) = \frac{21}{36}, \quad h(1) = \frac{14}{36}, \quad h(2) = \frac{1}{36}.$$

Note that the probability distribution for X (alone) is given by

$$P(X = x) = \frac{\binom{3}{x}\binom{6}{2-x}}{\binom{9}{2}}$$

(for $x = 0, 1, 2$), which is the same as the marginal distribution for X. For example,

$$P(X = 1) = \frac{\binom{3}{1}\binom{6}{1}}{\binom{9}{2}} = \frac{18}{36} = g(1).$$

3.40 Definition. If X and Y are discrete random variables, and $f(x, y)$ is their joint probability distribution, then the function

$$g(x) = \sum_y f(x, y)$$

is called the *marginal distribution of* X and the function

$$h(y) = \sum_x f(x, y)$$

is called the *marginal distribution of* Y. The sums are over all values of either y or x respectively.

3.13 Marginal Density

3.41 Definition. If X and Y are jointly continuous random variables, and $f(x, y)$ is their joint probability density function, then

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

is called the *marginal density of* X and the function

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

is called the *marginal density of* Y. These functions are defined for all $x \in \mathbb{R}$,$y \in \mathbb{R}$ respectively.

3.42 Example. Find the marginal densities of X and Y given their joint probability density

$$f(x, y) = \begin{cases} \frac{2}{3}(x + 2y) & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Solution. Marginal density of X: For $0 < x < 1$,

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^1 \frac{2}{3}(x + 2y)\, dy = \frac{2}{3}(x + 1)$$

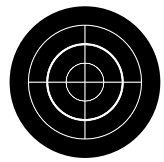
and $g(x) = 0$ otherwise.

Marginal density of Y: For $0 < y < 1$,

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^1 \frac{2}{3}(x + 2y)\, dx = \frac{1}{3}(1 + 4y)$$

and $h(y) = 0$ otherwise. □

Joint marginal distributions can be defined in the case of 3 or more random variables, though we won't pursue this.



3.43 Example. A circular biathlon target for prone position has a diameter of 45mm. Suppose that each point on the target has an equally likely probability of being hit by a shot.

Let $(0, 0)$ be the centre of the target, and define random variables X and Y, so that (X, Y) denotes the coordinates (in millimetres) of the shot fired.

The joint density function for X and Y is then, for some constant k,

$$f(x, y) = \begin{cases} k & \text{for } x^2 + y^2 \leq (22.5)^2 \\ 0 & \text{elsewhere} \end{cases}$$

It follows that $k = \frac{1}{(22.5)^2\pi}$, which is 1 over the area of the circle.

To find the marginal density for X, integrate over all y values:

$$\begin{aligned} x^2 + y^2 \leq (22.5)^2 \Rightarrow &\quad y^2 \leq (22.5)^2 - x^2 \\ \Rightarrow &\quad -\sqrt{(22.5)^2 - x^2} \leq y \leq \sqrt{(22.5)^2 - x^2} \end{aligned}$$

Thus

$$\begin{aligned} g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy &= \int_{-\sqrt{(22.5)^2 - x^2}}^{\sqrt{(22.5)^2 - x^2}} \frac{1}{(22.5)^2\pi}\, dy \\ &= \frac{2\sqrt{(22.5)^2 - x^2}}{(22.5)^2\pi} \end{aligned}$$

The marginal density of X can be used to find the probability the shot will land any horizontal distance x from the centre, regardless of its vertical position.

Notice that $g(x)$ is largest when $x = 0$ and gets smaller as x gets near the boundary of the target.

*This example uses more multivariable calculus than you are expected to know for this course, and was only presented as an application of the theory we have learned.

3.14 Conditional Distributions

Recall: Conditional probability of event A given event B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (P(B) \neq 0)$$

In terms of random variables: If A is the event $X = x$ and B is the event $Y = y$ then

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

For discrete random variables with joint probability distribution $f(x, y)$ we have

$$P(X = x|Y = y) = \frac{f(x, y)}{h(y)},$$

where $h(y) \neq 0$ is the marginal distribution of Y. This prompts the following definition.

3.44 Definition. If X and Y are discrete random variables with joint probability distribution $f(x, y)$, and respective marginal distributions $g(x)$ and $h(y)$, the function

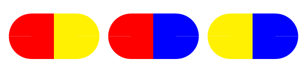
$$f(x|y) = \frac{f(x, y)}{h(y)}$$

is called the *conditional distribution of* X *given* $Y = y$, provided $h(y) \neq 0$. The function

$$w(y|x) = \frac{f(x, y)}{g(x)}$$

is called the *conditional distribution of* Y *given* $X = x$, provided $g(x) \neq 0$.

3.45 Example.



		x = 0	1	2	
y	0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$	$\frac{21}{36}$
	1	$\frac{8}{36}$	$\frac{6}{36}$		$\frac{14}{36}$
	2	$\frac{1}{36}$			$\frac{1}{36}$
		$\frac{15}{36}$	$\frac{18}{36}$	$\frac{3}{36}$	

Returning to the caplet example: The conditional distribution of X given $Y = 1$ is, $f(x|1) = \frac{f(x,1)}{h(1)}$.

Its values are:

$$f(0|1) = \frac{f(0,1)}{h(1)} = \frac{\frac{8}{36}}{\frac{14}{36}} = \frac{8}{14}$$
$$f(1|1) = \frac{f(1,1)}{h(1)} = \frac{\frac{6}{36}}{\frac{14}{36}} = \frac{6}{14}$$
$$f(2|1) = \frac{f(2,1)}{h(1)} = \frac{0}{\frac{14}{36}} = 0$$

In words, what do these probabilities mean?

3.15 Conditional Density

We now extend the notion of conditional distribution for joint discrete random variables to *conditional density* for jointly continuous random variables.

3.46 Definition. For jointly continuous random variables X and Y with joint density $f(x, y)$, and marginal densities $g(x)$ and $h(y)$, the function

$$f(x|y) = \frac{f(x, y)}{h(y)}$$

is called the *conditional density of* X *given* $Y = y$, provided $h(y) \neq 0$. The function

$$w(y|x) = \frac{f(x, y)}{g(x)}$$

is called the *conditional density of* Y *given* $X = x$, provided $g(x) \neq 0$.

3.47 Example. Let X and Y be jointly continuous random variables with joint probability density given by

$$f(x, y) = \begin{cases} \frac{3}{5}x(y + x) & \text{for } 0 < x < 1, 0 < y < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find conditional probability of Y given $X = x$.

(b) Find $P(0 < Y < 1|X = 0.75)$.

Solution. (a) First we need the marginal density function for X:

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^2 \frac{3}{5}x(y + x)\, dy = \frac{3}{5}\left(\frac{xy^2}{2} + x^2 y\right)\Big|_0^2$$

thus $g(x) = \frac{6}{5}(x+x^2)$. The conditional density function for $0 < x < 1, 0 < y < 2$ is then

$$w(y|x) = \frac{f(x, y)}{g(x)} = \frac{\frac{3}{5}x(y + x)}{\frac{6}{5}(x + x^2)} = \frac{y + x}{2 + 2x}.$$

(b) Using what we just found,

$$P(0 < Y < 1|X = 0.75) = \int_0^1 w(y|0.75)\,dy = \int_0^1 \frac{y+0.75}{3.5}\,dy$$

$$= \frac{y^2}{7} + \frac{3y}{14}\bigg|_0^1 = \frac{5}{14}.$$

□

3.48 Exercise. Let X and Y be jointly continuous random variables with joint probability density given by

$$f(x,y) = \begin{cases} 4xy & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $P(0 < X < 0.5|Y = 0.5)$.

For joint distributions/densities of three or more joint random variables, we can define conditional probabilities in various ways, consistent with the two variable case. For example, if we have four random variables X_1, X_2, X_3, X_4, with joint distribution/density $f(x_1, x_2, x_3, x_4)$, one can define functions

$$f(x_1|x_2,x_3,x_4) = \frac{f(x_1,x_2,x_3,x_4)}{m(x_2,x_3,x_4)},$$

$$f(x_2,x_3|x_1,x_4) = \frac{f(x_1,x_2,x_3,x_4)}{m(x_1,x_4)},$$

$$f(x_1,x_2,x_4|x_3) = \frac{f(x_1,x_2,x_3,x_4)}{m(x_3)},$$

where the denominators are marginal distributions/densities. This gives many ways to analyse probabilities in an experiment when several random variables are defined.

3.16 Independent Random Variables

Just as we defined the concept of independent events, we may speak of independent random variables.

3.49 Definition. If random variables X and Y have joint probability distribution (or density) $f(x, y)$ and marginal distributions (resp. densities) $g(x)$ and $h(y)$, then we say X and Y are *independent* if and only if

$$f(x,y) = g(x) \cdot h(y).$$

(In the higher dimensional cases we require that the joint distribution/density be the product of the individual marginal distributions/densities)

3.50 Exercise. Let X and Y be jointly continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} 6e^{-2x}e^{-3y} & \text{for } 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Show that X and Y are independent random variables.

Chapter 4

Mathematical Expectation

4.1 Example. Suppose you are at a casino that has a dice game which costs \$1000 for a single roll of two 6-sided dice. You win \$5,555 by rolling a 7 and lose your money otherwise.

Do you think it is worthwhile to play this game? Could you expect to come out ahead by repeatedly playing this game?

4.2 Example. Suppose a university fundraiser sells 10,000 raffle tickets at a dollar apiece with a grand prize of \$5,000, a second prize of \$1,000 and two third place prizes of \$500 each.

Do you think your ticket is worth \$1? How much do you think it is worth? In other words, how much can you "expect" to win in this raffle?

4.1 The Expected Value of a Discrete Random Variable

4.3 Definition. If X is a discrete random variable and $f(x)$ is the value of its probability distribution at x, the *expected value* of X (or *expectation of* X) is defined

$$E(X) = \sum_x x \cdot f(x).$$

where the sum is over all x in the range of X.

The sum must be defined in order for the expected value to have meaning.

In the first example of the dice game, let random variable X be the amount of money won on each roll. The range of X is $\{0, 5555\}$.

Since the probability of rolling a 7 is $\frac{1}{6}$ we have $P(X = 5555) = \frac{1}{6}$ and therefore $P(X = 0) = \frac{5}{6}$.

The expected value is

$$E(X) = 0 \cdot \left(\frac{5}{6}\right) + 5555 \cdot \left(\frac{1}{6}\right) \approx 925.83.$$

This analysis shows that this is a losing game, because our expected value is less than the cost to play. i.e. in the long run, we can expect to lose money.

Raffle Ticket

In the raffle ticket example we let X denote the possible winnings for our raffle ticket. Typically once a ticket is drawn it is not replaced to be drawn again, so the range of X is $\{0, 500, 1000, 5000\}$. (This ignores the cost of the ticket.)

Four tickets will be drawn for the four prizes and there is an equally likely chance of $\frac{1}{10000}$ for each prize. Therefore $P(X = 0) = \frac{9996}{10000}$, $P(X = 500) = \frac{2}{10000}$, $P(X = 1000) = \frac{1}{10000}$, $P(X = 5000) = \frac{1}{10000}$.

The expected value of X is $E(X) = 0 \cdot \left(\frac{9996}{10000}\right) + 500 \cdot \left(\frac{2}{10000}\right) + 1000 \cdot \left(\frac{1}{10000}\right) + 5000 \cdot \left(\frac{1}{10000}\right) = 0.70$.

Hypothetically, by playing the raffle repeatedly, we expect to win \$0.70 on average; therefore losing money with the \$1 cost. We could place a value of \$0.70 for our ticket.

4.4 Example. In 2022, the Heart and Stroke Lottery sold/distributed 168,000 tickets for their annual lottery, where there were 68,879 prizes (of a variety of cash values) to be won with at total value of \$5,375,499.89. This is a charity event and tickets cost \$100 each.

How much could a person "expect" to win with a purchase of a ticket?

To answer this question we may need more information.

Firstly, in this lottery, tickets are replaced after being drawn so that one ticket can win multiple times. This increases the range of X. For example, the top two grand prizes are \$1,000,000 each, meaning we include both 1000000 and 2000000 in the range of X (as well as 5,375,499.89, for the event of winning all prizes).

Second we need a list of all the different prizes, their quantities and their cash values. (there are several non-cash prizes)

It can be quite a chore to compute expected value since different combinations of prizes can add up to the same value.

Raffle with Replacement

Solution. Returning to the previous raffle example, let's compute the expected value of a single ticket with only three prize draws of \$5,000, \$1,000, \$500 each, but now tickets are replaced each time to allow for multiple wins. (Again 10000 tickets sold.)

Let X be the total prize money won. The range of X is $\{0, 500, 1000, 1500, 5000, 5500, 6000, 6500\}$. Then

$$P(X=0) = \left(\tfrac{9999}{10000}\right)^3,$$
$$P(X=500) = \left(\tfrac{9999}{10000}\right)^2 \left(\tfrac{1}{10000}\right),$$
$$P(X=1000) = \left(\tfrac{9999}{10000}\right) \left(\tfrac{1}{10000}\right) \left(\tfrac{9999}{10000}\right),$$
$$P(X=1500) = \left(\tfrac{9999}{10000}\right) \left(\tfrac{1}{10000}\right)^2,$$
$$P(X=5000) = \left(\tfrac{1}{10000}\right) \left(\tfrac{9999}{10000}\right)^2,$$
$$P(X=5500) = \left(\tfrac{1}{10000}\right) \left(\tfrac{9999}{10000}\right) \left(\tfrac{1}{10000}\right),$$
$$P(X=6000) = \left(\tfrac{1}{10000}\right)^2 \left(\tfrac{9999}{10000}\right),$$
$$P(X=6500) = \left(\tfrac{1}{10000}\right)^3$$

The expected value of X is

$$\begin{aligned}
&E(X) \\
&= 0 \cdot \left(\frac{9999}{10000}\right)^3 + 500 \cdot \left(\frac{9999}{10000}\right)^2 \left(\frac{1}{10000}\right) \\
&\quad + 1000 \cdot \left(\frac{9999}{10000}\right)^2 \left(\frac{1}{10000}\right) + 1500 \cdot \left(\frac{9999}{10000}\right) \left(\frac{1}{10000}\right)^2 \\
&\quad + 5000 \cdot \left(\frac{9999}{10000}\right)^2 \left(\frac{1}{10000}\right) + 5500 \cdot \left(\frac{9999}{10000}\right) \left(\frac{1}{10000}\right)^2 \\
&\quad + 6000 \cdot \left(\frac{9999}{10000}\right) \left(\frac{1}{10000}\right)^2 + 6500 \cdot \left(\frac{1}{10000}\right)^3 \\
&= 0.65
\end{aligned}$$

□

4.5 Example.

7	\$	$\overline{BAR}$
$\overline{BAR}$	$\overline{BAR}$	7
🔔	♡	\$

A slot machine has three wheels with 20 symbols on each wheel. There is one 7, two $\overline{BAR}$, and three 🔔 on each wheel. It costs \$0.25 to play. Payouts:

Wheel 1	Wheel 2	Wheel 3	
7	7	7	\$500
$\overline{BAR}$	$\overline{BAR}$	$\overline{BAR}$	\$100
🔔	🔔	🔔	\$20

(all other permutations lose).

What is the expected value of this game? (ignore cost to play)

4.2 The Expected Value of a Continuous Random Variable

4.6 Definition. If X is a continuous random variable and $f(x)$ is its probability density function, the *expected value* of X is defined

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx.$$

The integral must exist in order for the expected value to have meaning.

4.7 Example. If a contractor's profit on a construction job can be looked upon as a continuous random variable having probability density

$$f(x) = \begin{cases} \frac{1}{18}(x+1) & \text{for } -1 < x < 5 \\ 0 & \text{otherwise} \end{cases}$$

where the units are in \$1,000, what is her expected profit?

Solution. The expected value of X, where X denotes the contractor's profit in \$1,000's, is

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot f(x)\, dx \\ &= \int_{-1}^{5} x \cdot \frac{1}{18}(x+1)\, dx \\ &= \frac{1}{18}\left(\frac{x^3}{3} + \frac{x^2}{2}\right)\Bigg|_{-1}^{5} \\ &= \frac{1}{18}\left(\frac{125}{3} + \frac{25}{2} - \frac{(-1)}{3} - \frac{1}{2}\right) \\ &= 3. \end{aligned}$$

Therefore the expected profit is $3 \cdot \$1,000 = \$3,000$. □

4.3 Properties Expected Value

4.8 Theorem. If X is a discrete random variable with probability distribution $f(x)$, the expected value of $g(X)$ is given by

$$E(g(X)) = \sum_{x} g(x) \cdot f(x).$$

If X is a continuous random variable with probability density function $f(x)$, the expected value of $g(X)$ is given by

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx.$$

4.9 Example. Let X be a random variable that takes the values $-1, 0, 1$, and has probability distribution given by

$$f(-1) = 0.2, \quad f(0) = 0.5, \quad f(1) = 0.3.$$

Find $E(X^2)$.

Solution. Before using the theorem, let's find $E(X^2)$ directly. We view X^2 as a new random variable which we'll call Y.

The range of Y is $\{0, 1\}$ and it has probability distribution

$$P(Y = 0) = P(X = 0) = f(0) = 0.5$$

$$P(Y = 1) = P(X = 1) + P(X = -1) = f(1) + f(-1) = 0.5$$

Then

$$E(X^2) = E(Y) = 0 \cdot P(Y = 0) + 1 \cdot P(Y = 1) = 0 \cdot (0.5) + 1 \cdot (0.5) = 0.5.$$

We can find this using the theorem as well: Let $g(X) = X^2$, and let $x_1 = -1, x_2 = 0, x_3 = 1$. Then according the theorem

$$\begin{aligned} E(g(X)) &= \sum_{i=1}^{3} g(x_i) f(x_i) \\ &= g(x_1)f(x_1) + g(x_2)f(x_2) + g(x_3)f(x_3) \\ &= g(-1)f(-1) + g(0)f(0) + g(1)f(1) \\ &= (-1)^2 \cdot (0.2) + (0)^2 \cdot (0.5) + (1)^2 \cdot (0.3) \\ &= 0.5. \end{aligned}$$

Note that: $E(X^2) = 0.5 \neq (E(X))^2 = 0.01$

□

4.10 Example. Suppose X has probability density

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the expected value of $g(X) = e^{3X/4}$.

Solution. By our theorem

$$
\begin{aligned}
E(g(x)) &= \int_{-\infty}^{\infty} g(x)f(x)\,dx \\
&= \int_{0}^{\infty} e^{3x/4} \cdot e^{-x}\,dx \\
&= \int_{0}^{\infty} e^{-x/4}\,dx \\
&= -4e^{-x/4}\Big|_{0}^{\infty} \\
&= 4.
\end{aligned}
$$

□

A useful special case of the theorem is:

4.11 Theorem. If a and b are constants, then

$$E(aX + b) = aE(X) + b.$$

In particular if $a = 0$, then $E(b) = b$ and if $b = 0$ then $E(aX) = aE(X)$.

Proof. To prove this (for the discrete case) let $g(X) = aX + b$. Then

$$
\begin{aligned}
E(aX + b) &= E(g(x)) \\
&= \sum_x g(x) \cdot f(x) \\
&= \sum_x (ax + b) \cdot f(x) \\
&= \sum_x (ax \cdot f(x) + b \cdot f(x)) \\
&= \sum_x ax \cdot f(x) + \sum_x b \cdot f(x) \\
&= a \sum_x x \cdot f(x) + b \sum_x f(x) \\
&= aE(X) + b,
\end{aligned}
$$

since $\sum_x f(x) = 1$. (see text for continuous case) □

7	\$	$\overline{BAR}$
$\overline{BAR}$	$\overline{BAR}$	7
🔔	♡	\$

4.12 Example. Returning to our slot machine example, we chose our random variable X to be the expected payout, and not the expected profit. Then we calculated the expected payout. But suppose we want the expected profit.

If Y is our expected profit then Y has range $\{-0.25, 19.75, 99.75, 499.75\}$

So then $P(Y = y) = P(X = y + 0.25)$, and

$$\begin{aligned} E(Y) = (-0.25) \cdot P(X = 0) + (19.75) \cdot P(X = 20) \\ + (99.75) \cdot P(X = 100) + (499.75) \cdot P(X = 500) \end{aligned}$$

On the other hand since $Y = g(X) = X - 0.25$, we can compute

$$E(Y) = E(X - 0.25) = E(X) - 0.25$$

using the property above (which, in this case, is a nicer calculation).

We can extend extend the theorem above to more expressions:

4.13 Theorem. If $c_1, c_2, \ldots, c_n$ are constants, then

$$E\left(\sum_{i=1}^{n} c_i g_i(X)\right) = \sum_{i=1}^{n} c_i E(g_i(X)),$$

where the g_i are functions.

(try proving this for the continuous case)

Proof. Suppose X has p.d.f. $f(x)$. To apply the theorem let $h(x) = \sum_{i=1}^{n} c_i g_i(X)$. Then

$$\begin{aligned} E(h(X)) &= \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx \\ &= \int_{-\infty}^{\infty} \left(\sum_{i=1}^{n} c_i g_i(X)\right) \cdot f(x)\, dx \\ &= \sum_{i=1}^{n} c_i \int_{-\infty}^{\infty} g_i(X) \cdot f(x)\, dx \\ &= \sum_{i=1}^{n} c_i E(g_i(X)). \end{aligned}$$

□

4.4 Multivariate Expected Value

Suppose X and Y are random variables with a joint probability distribution/density $f(x, y)$. Then $Z = g(X, Y)$ is a random variable defined by the function g depending on X and Y. The expected value of Z may be computed in the following way.

4.14 Theorem. With notation as above if X and Y are discrete random variables, then

$$E(g(X,Y)) = \sum_x \sum_y g(x,y) \cdot f(x,y)$$

where sums are taken over x and y in the ranges of X and Y respectively. In the continuous case,

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \cdot f(x,y)\, dx\, dy$$

4.15 Example. Returning again to the caplet example.

Let $Z = X+Y$. Then Z is the random variable which gives the total number of asprin or sedative when two caplets are drawn.

What is the expected value of Z?

y \ x	0	1	2
0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$
1	$\frac{8}{36}$	$\frac{6}{36}$	
2	$\frac{1}{36}$		

Solution. First note that $P(Z \geq 3) = 0$ as $f(2,1) = f(1,2) = f(2,2) = 0$.

$$\begin{aligned}
&E(X+Y) \\
&= \sum_{x=0}^{2} \sum_{y=0}^{2} (x+y) \cdot f(x,y) \\
&= (0+0) \cdot \frac{6}{36} + (0+1) \cdot \frac{8}{36} + (0+2) \cdot \frac{1}{36} \\
&\quad + (1+0) \cdot \frac{12}{36} + (1+1) \cdot \frac{6}{36} + (2+0) \cdot \frac{3}{36} \\
&= \frac{40}{36} \\
&\approx 1.1111.
\end{aligned}$$

□

4.16 Example. If the joint probability density of X and Y is given by,

$$f(x,y) = \begin{cases} x+y & \text{for } 0<x<1, 0<y<1 \\ 0 & \text{otherwise} \end{cases}$$

find the expected value of XY.

Solution.

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot f(x,y)\,dx\,dy \\ &= \int_0^1\int_0^1 xy\cdot(x+y)\,dx\,dy \\ &= \int_0^1\int_0^1 x^2y + xy^2\,dx\,dy \\ &= \int_0^1 \frac{x^3y}{3} + \frac{x^2y^2}{2}\Big|_0^1\,dy \\ &= \int_0^1 \frac{y}{3} + \frac{y^2}{2}\,dy \\ &= \frac{y^2}{6} + \frac{y^3}{6}\Big|_0^1 \\ &= \frac{1}{3}. \end{aligned}$$

□

4.5 Moments

4.17 Definition. The *rth moment about the origin* of a random variable X is defined as the expected value of X^r. In the discrete case this is,

$$E(X^r) = \sum_x x^r \cdot f(x),$$

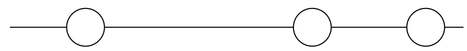
and in the continuous case this is

$$E(X^r) = \int_{-\infty}^{\infty} x^r \cdot f(x)\,dx$$

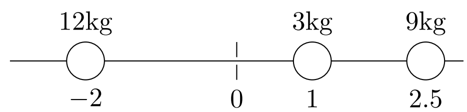
for $r = 0, 1, 2, 3, \ldots$. (here $f(x)$ is the probability distribution/density of X)

The Latin origin of the word is the verb "to move." In physics a moment is the product of the distance to some point from the origin (raised to the power r), times a physical quantity such as mass, force, or charge.

For example, suppose there are three point masses sitting on a massless rigid beam 6m in length.



Object 1 lies 2m left of centre and has a mass of 12kg, objects 2 and 3 lie respectively 1m and 2.5m right of centre and have respective masses 3kg and 9 kg.



Let $x_1 = -2, x_2 = 1, x_3 = 2.5$ be the positions of the objects, and let $f(x)$ be the mass distribution (i.e. $f(x_i)$ is the mass of object i). The *rth moments of mass* of this physical system are

$$\sum_{i=1}^{3} x_i^r \cdot f(x_i)$$

0th moment of mass: $\sum_{i=1}^{3} x_i^0 \cdot f(x_i) = (1)(12) + (1)(3) + (1)(9) = 24$.
1st moment of mass: $\sum_{i=1}^{3} x_i \cdot f(x_i) = (-2)(12) + (1)(3) + (2.5)(9) = 1.5$.
2nd moment of mass: $\sum_{i=1}^{3} x_i^2 \cdot f(x_i) = (4)(12) + (1)(3) + (6.25)(9) = 107.25$.

The 0th moment of mass is the total mass, 24kg.

The 1st moment of mass divided by the total mass is the *centre of mass*, or the balance point of the beam which is $\frac{1.5\text{kg}\cdot\text{m}}{24\text{kg}} = 0.0625$m right of centre.

The 2nd moment of mass 107.25 kg·m^2, is the *moment of inertia*, which gives the amount of torque (in kg m^2/s^2) required to cause an angular acceleration of 1 rad/s^2, around the line $y = 0$.

In a similar way, moments of random variables describe their probability distribution.

The Mean of a Distribution

The 0th moment about the origin of a random variable X, is equal to 1 since,

$$E(X^0) = \sum_x x^0 \cdot f(x) = \sum_x f(x) = 1$$

in the discrete case, and in the continuous case

$$E(X^0) = \int_{-\infty}^{\infty} x^0 \cdot f(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

The 1st moment about the origin is the expected value:

$$E(X) = \sum_x x \cdot f(x), \qquad E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Because of its importance, it is called the *mean of the distribution of* X (or the *mean of* X) and is denoted simply by μ.

In every day language, "mean" and "average" (of a set of values) are used interchangeably to mean the sum of all values dived by the number of values; example, average test score.

We can view the test score as a discrete random variable X, for the experiment of selecting a test at random from the class. In this case some test scores may appear repeatedly (e.g. 5 students may have scored 80%, and 3 scored 75%), so the probability of drawing particular scores may not be equally likely. The expected value of X, also called the mean of X, is the sum of each score times their respective probabilities.

However, assuming an equally likely chance of any test being drawn, the probability of drawing a particular score is the number of tests with that score divided by the total number of tests, and it follows that the mean and average, for equally likely outcomes, is the same.

The rth Moment About the Mean

4.18 Definition. Let X be a random variable with probability distribution/density $f(x)$ having mean μ. The *rth moment about the mean* of X is defined as the expected value of $(X-\mu)^r$.

$$\text{discrete:} \qquad E((X-\mu)^r) = \sum_x (x-\mu)^r \cdot f(x),$$

$$\text{continuous:} \qquad E((X-\mu)^r) = \int_{-\infty}^{\infty} (x-\mu)^r \cdot f(x)\, dx$$

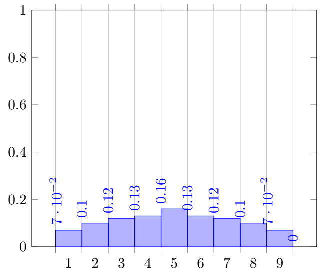
Note that $E((X-\mu)^0) = 1$ and $E((X-\mu)) = 0$ (provided μ exists).

Variance and Standard Deviation

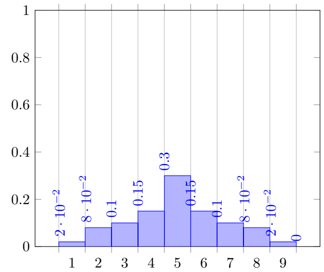
The 2nd moment about the mean, $E((X-\mu)^2)$, is called the *variance of the distribution for* X, or simply the *variance of* X, and is denoted by σ^2, or $\mathrm{var}(X)$.

The positive square root of the variance, σ, is called the *standard deviation*.

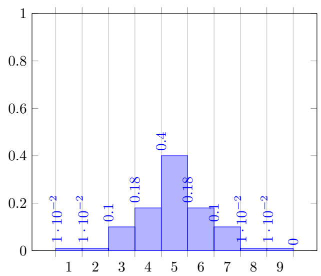
These values describe the dispersion of the probability distribution of X.



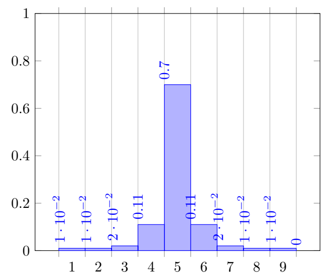
$\mu = 5, \sigma^2 = 5.26$



$\mu = 5, \sigma^2 = 3.18$



$\mu = 5, \sigma^2 = 1.66$



$\mu = 5, \sigma^2 = 0.88$

4.19 Example. Let X and Y be discrete random variables with the following distributions

x	$P(X = x)$
1	0
2	1/4
3	1/2
4	0
5	0
6	1/4

y	$P(Y = y)$
1	1/4
2	0
3	0
4	1/2
5	1/4
6	0

Show that these distributions have the same mean and variance.

Solution. For X:

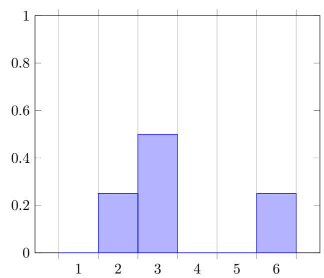
$$\mu = E(X) = 1 \cdot 0 + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{2} + 4 \cdot 0 + 5 \cdot 0 + 6 \cdot \frac{1}{4} = \frac{7}{2}$$

$$\begin{aligned} \sigma^2 = E((X-\mu)^2) &= \left(1 - \frac{7}{2}\right)^2 \cdot 0 + \left(2 - \frac{7}{2}\right)^2 \cdot \frac{1}{4} + \left(3 - \frac{7}{2}\right)^2 \cdot \frac{1}{2} \\ &+ \left(4 - \frac{7}{2}\right)^2 \cdot 0 + \left(5 - \frac{7}{2}\right)^2 \cdot 0 + \left(6 - \frac{7}{2}\right)^2 \cdot \frac{1}{4} = \frac{9}{4} \end{aligned}$$

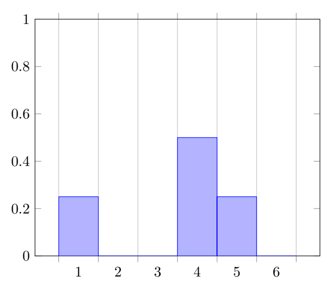
For Y:

$$\mu = E(Y) = 1 \cdot \frac{1}{4} + 2 \cdot 0 + 3 \cdot 0 + 4 \cdot \frac{1}{2} + 5 \cdot \frac{1}{4} + 6 \cdot 0 = \frac{7}{2}$$

$$\sigma^2 = E((Y-\mu)^2) = \left(1-\frac{7}{2}\right)^2 \cdot \frac{1}{4} + \left(2-\frac{7}{2}\right)^2 \cdot 0 + \left(3-\frac{7}{2}\right)^2 \cdot 0$$
$$+ \left(4-\frac{7}{2}\right)^2 \cdot \frac{1}{2} + \left(5-\frac{7}{2}\right)^2 \cdot \frac{1}{4} + \left(6-\frac{7}{2}\right)^2 \cdot 0 = \frac{9}{4}$$



$P(X = x)$



$P(Y = y)$

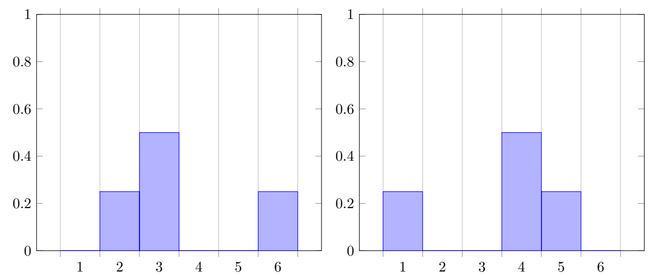
In both cases the mean is $\frac{7}{2}$ and the variance is $\frac{9}{4}$.

Let us now compute the 3rd moment about the mean. For X:

$$\mu_3 = E((X-\mu)^3) = \left(1-\frac{7}{2}\right)^3 \cdot 0 + \left(2-\frac{7}{2}\right)^3 \cdot \frac{1}{4} + \left(3-\frac{7}{2}\right)^3 \cdot \frac{1}{2}$$
$$+ \left(4-\frac{7}{2}\right)^3 \cdot 0 + \left(5-\frac{7}{2}\right)^3 \cdot 0 + \left(6-\frac{7}{2}\right)^3 \cdot \frac{1}{4} = 3$$

For Y:

$$\mu_3 = E((Y-\mu)^3) = \left(1-\frac{7}{2}\right)^3 \cdot \frac{1}{4} + \left(2-\frac{7}{2}\right)^3 \cdot 0 + \left(3-\frac{7}{2}\right)^3 \cdot 0$$
$$+ \left(4-\frac{7}{2}\right)^3 \cdot \frac{1}{2} + \left(5-\frac{7}{2}\right)^3 \cdot \frac{1}{4} + \left(6-\frac{7}{2}\right)^3 \cdot 0 = -3$$



$P(X = x)$ $P(Y = y)$

The 3rd moment about the mean is used to describe the symmetry/skewness of the graph about the mean. The distribution on the left has 3rd moment about the mean, $E((X-\mu)^3)$,equal to positive 3, and the distribution on the right has 3rd moment about the mean equal to negative 3.

□

4.20 Example. Let random variable X be the number of points on a regular 6-sided die. Compute mean and variance of X.

Solution. The mean is

$$\mu = E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

The variance is

$$\begin{aligned}\sigma^2 = E((X-\mu)^2) &= (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6} \\ &\quad + (4-3.5)^2 \cdot \frac{1}{6} + (5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6} \\ &= \frac{17.5}{6} \\ &\approx 2.9167\end{aligned}$$

□

Using properties of expected values we have

$$\begin{aligned}E((X-\mu)^2) &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2E(\mu X) + E(\mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 \\ &= E(X^2) - \mu^2\end{aligned}$$

We summarize this in a theorem.

4.21 Theorem.

$$\sigma^2 = E(X^2) - \mu^2$$

4.22 Example. Redo previous die rolling problem with theorem.

Solution. We first need to find the mean μ:

$$\mu = E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

Next we find $E(X^2)$:

$$\begin{aligned} E(X^2) &= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} \\ &= \frac{91}{6} \\ &\approx 15.1667 \end{aligned}$$

(this saves subtracting the mean in each term)

Then using the theorem, the variance is

$$\sigma^2 = E(X^2) - \mu^2 = \frac{91}{6} - (3.5)^2 \approx 2.9167.$$

□

4.23 Theorem. If X has variance σ^2, then for constants a and b

$$\mathrm{var}(aX + b) = a^2\sigma^2.$$

Proof. Let $Y = aX + b$, and let $\mu = E(X)$. Then

$$E(Y) = E(aX + b) = aE(X) + b = a\mu + b.$$

For the variance

$$\begin{aligned} \mathrm{var}(Y) &= E((Y - (a\mu + b))^2) \\ &= E((aX + b - a\mu - b)^2) \\ &= E((aX - a\mu)^2) \\ &= E(a^2X^2 - 2a^2X\mu + a^2\mu^2) \\ &= a^2E(X^2) - 2a^2\mu E(X) + a^2\mu^2 \\ &= a^2(E(X^2) - 2\mu^2 + \mu^2) \\ &= a^2(E(X^2) - \mu^2) \\ &= a^2\sigma^2 \end{aligned}$$

□

4.6 Chebyshev's Theorem

The next important theorem shows how σ describes the spread of the probability distribution.

4.24 Theorem (Chebyshev's Theorem)**.** Let X be a random variable with mean μ and standard deviation σ. Then for any $k > 0$,

$$P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

In words, the probability that values for X lie within k standard deviations of the mean is at least $1 - \frac{1}{k^2}$.

$$|X-\mu| < k\sigma \quad \Leftrightarrow \quad -k\sigma < X - \mu < k\sigma \quad \Leftrightarrow \quad \mu - k\sigma < X < \mu + k\sigma$$

Proof. Proof for continuous case: Using the definition of variance we have

$$\begin{aligned}
\sigma^2 = E((X-\mu)^2) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \\
&= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)\,dx \\
&\quad + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\,dx \\
&\geq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\,dx
\end{aligned}$$

For $x \leq \mu - k\sigma$ or $x \geq \mu - k\sigma$ we have $(x-\mu)^2 \geq k^2\sigma^2$ so

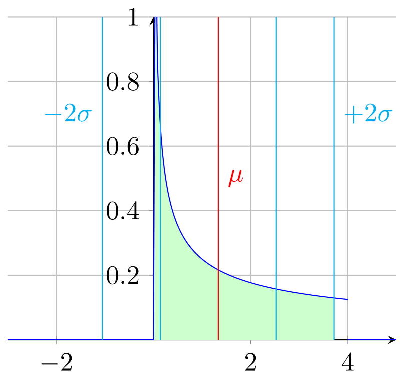
$$\begin{aligned}
\sigma^2 &\geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)\,dx \\
&= k^2\sigma^2 \int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx \\
\Rightarrow \frac{1}{k^2} &\geq \int_{-\infty}^{\mu-k\sigma} f(x)\,dx + \int_{\mu+k\sigma}^{\infty} f(x)\,dx
\end{aligned}$$

The integrals on the right are $P(X \leq \mu - k\sigma)$ and $P(X \geq \mu + k\sigma)$ or combined $P(|X-\mu| \geq k\sigma)$. Thus

$$P(|X-\mu| < k\sigma) = 1 - P(|X-\mu| \geq k\sigma) \geq 1 - \frac{1}{k^2}.$$

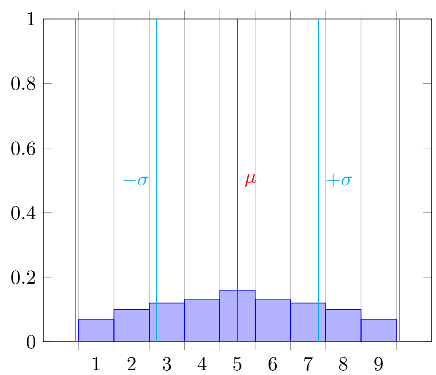
(similar proof for discrete case) □

4.25 Example. $f(x) = \begin{cases} \frac{1}{4}x^{-\frac{1}{2}} & \text{for } 0 < x \leq 4 \\ 0 & \text{otherwise} \end{cases}$

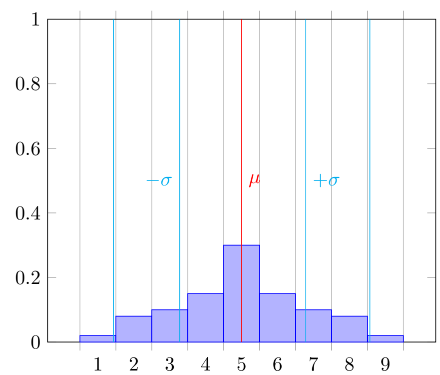


$\mu = \frac{4}{3} \approx 1.3333$, $\sigma = \frac{8\sqrt{5}}{15} \approx 1.1926$, $P(|X-\mu| < 2\sigma) \approx 0.9642$.

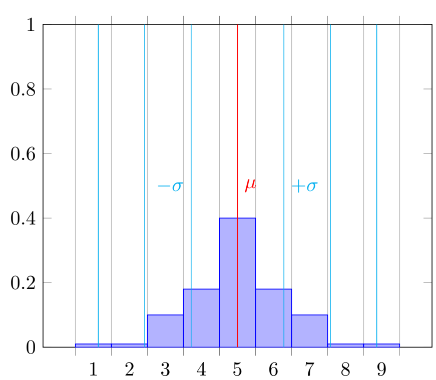
By Chebyshev: $P(|X-\mu| < 2\sigma) \geq 1 - \frac{1}{2^2} = \frac{3}{4}$.



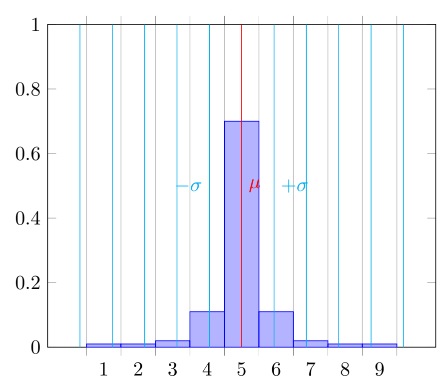
$\mu = 5, \sigma^2 = 5.26$



$\mu = 5, \sigma^2 = 3.18$



$\mu = 5, \sigma^2 = 1.66$



$\mu = 5, \sigma^2 = 0.88$

4.26 Exercise. The mean score of an exam is 70, with a standard deviation of 5. At least what percentage of the data set lies between 40 and 100?

4.27 Exercise. The mean age of a flight attendant is 40, with a standard deviation of 8. At least what percent of the data set lies between 20 and 60?

Maclaurin series

From calculus, the Maclaurin series (also known as the Taylor series about 0) of the function e^x is given by

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \dots \quad = \sum_{i=0}^{\infty} \frac{1}{i!}x^i.$$

This approximates values of e^x for different x, e.g. for $x = 1$

$$\sum_{i=0}^{4} \frac{1}{i!}x^i = 1 + (1) + \frac{1}{2}(1)^2 + \frac{1}{6}(1)^3 + \frac{1}{24}(1)^4 \approx 2.708333$$

The more terms we use the better the approximation:

$$1 + (1) + \frac{1}{2}(1)^2 + \frac{1}{6}(1)^3 + \frac{1}{24}(1)^4 + \frac{1}{120}(1)^5 + \frac{1}{720}(1^6) \approx 2.71806$$

Note that

$$e = 2.71828182845904523536028747135266249775724709369995957496696762772407663035354759457138$$

(its decimal representation does not end, and has no repeating pattern).

The *Maclaurin series* (where it exists) for a function $f(x)$ is

$$\sum_{i=0}^{\infty} \frac{f^{(i)}(0)}{i!}x^i$$

where $f^{(i)}(0)$ is the ith derivative of f evaluated at 0.

For a Maclaurin series to exist all derivatives involved must exist. Second, the infinite sum has to converge for the x values we plug in. The range of x values where it does converge is called its interval of convergence.

A function's Maclaurin series gives better approximations (i.e. requires fewer terms) for x values which are close to 0.

The derivative of a Maclaurin series can be found by taking the derivatives of the individual terms in the sum (where defined).

For this course we do not need to know the theory of series expansions for functions.

What we do need to know is the Maclaurin Series for e^x,

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \dots \quad = \sum_{i=0}^{\infty} \frac{1}{i!}x^i,$$

and term-by-term differentiation,

$$\frac{d}{dx}\left(\sum_{i=0}^{\infty} f_i(x)\right) = \sum_{i=0}^{\infty} \frac{d}{dx}\left(f_i(x)\right),$$

in order to talk about *moment generating functions*.

4.7 Moment Generating Functions

4.28 Definition. The *moment generating function of a random variable* X, where it exists, is given by

$$\text{discrete case:} \quad M_X(t) = E(e^{tX}) = \sum_x e^{tx} \cdot f(x)$$

$$\text{continuous case:} \quad M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)\, dx$$

where $f(x)$ is the probability distribution/density of X.

We will see why this name is appropriate. Expanding the expression for $M_X(t)$,

$$\begin{aligned} M_X(t) &= \sum_x e^{tx} \cdot f(x) \\ &= \sum_x \left(1 + (tx) + \frac{1}{2!}(tx)^2 + \frac{1}{3!}(tx)^3 + \dots\right) \cdot f(x) \\ &= \sum_x f(x) + (tx)f(x) + \frac{(tx)^2}{2!}f(x) + \frac{(tx)^3}{3!}f(x) + \dots \\ &= \sum_x f(x) + t\sum_x xf(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \frac{t^3}{3!}\sum_x x^3 f(x) + \dots \end{aligned}$$

we see the rth moments about the origin appearing in the terms of the series.

$$M_X(t) = \sum_x f(x) + t\sum_x xf(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \frac{t^3}{3!}\sum_x x^3 f(x) + \dots$$

To extract the ith moment, we take the ith derivative with respect to t, and evaluate at $t = 0$.

For example, to get the 2nd moment:

$$\begin{aligned} &\frac{d^2}{dt^2}M_X(t)\Big|_{t=0} \\ &= \frac{d^2}{dt^2}\left(\sum_x f(x) + t\sum_x xf(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \frac{t^3}{3!}\sum_x x^3 f(x) + \dots\Big|_{t=0}\right) \end{aligned}$$

Take 2nd derivative of each term with respect to t,

$$= \left(0 + 0 + \sum_x x^2 f(x) + t\sum_x x^3 f(x) + \frac{t^2}{2}\sum_x x^4 f(x) + \dots\Big|_{t=0}\right)$$

Letting $t = 0$ gives

$$\frac{d^2}{dt^2}M_X(t)\Big|_{t=0} = \sum_x x^2 f(x) = E(X^2).$$

4.29 Example. Let X be a discrete random variable with distribution $f(x) = \frac{1}{8}\binom{3}{x}$ for $x = 0, 1, 2, 3$.

The moment generating function for X is

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{3} e^{tx} \cdot \left(\frac{1}{8}\binom{3}{x}\right) \\ &= \frac{1}{8}\left(e^0\binom{3}{0} + e^t\binom{3}{1} + e^{2t}\binom{3}{2} + e^{3t}\binom{3}{3}\right) \\ &= \frac{1}{8}(1 + e^t)^3 \end{aligned}$$

To find the mean (1st moment about the origin):

$$\frac{d}{dt}M_X(t)\Big|_{t=0} = \frac{d}{dt}\frac{1}{8}(1 + e^t)^3\Big|_{t=0} = \frac{3}{8}(1 + e^t)^2 e^t\Big|_{t=0} = \frac{3}{8}2^2 = \frac{3}{2}$$

Since we already have the moment generating function $M_X(t) = \frac{1}{8}(1 + e^t)^3$, we can quickly compute other moments.

The second moment about the origin, $E(X^2)$:

$$\begin{aligned}
\frac{d^2}{dt^2} M_X(t)\Big|_{t=0} &= \frac{d^2}{dt^2}\frac{1}{8}(1+e^t)^3\Big|_{t=0} \\
&= \frac{d}{dt}\frac{3}{8}(1+e^t)^2 e^t\Big|_{t=0} \\
&= \frac{6}{8}(1+e^t)e^{2t} + \frac{3}{8}(1+e^t)^2 e^t\Big|_{t=0} \\
&= 3.
\end{aligned}$$

These two could now be used to find the variance, $\sigma^2 = E(X^2) - \mu^2$. Continue this process to find higher order moments.

Properties of Moment Generating Functions

We will only briefly touch on moment generating functions for now, and come back to them when we need to use them.

One advantage of knowing the moment generating function for a random variable, is that it can be used to find the moment generating function for related random variables via the following theorem.

4.30 Theorem. If $M_X(t)$ is the moment generating function for X and a and b are nonzero constants, then

1. $$M_{X+a}(t) = e^{at} \cdot M_X(t)$$

2. $$M_{bX}(t) = M_X(bt)$$

3. $$M_{\frac{X+a}{b}}(t) = e^{\frac{a}{b}t} \cdot M_X\left(\frac{t}{b}\right)$$

4.8 Bivariate Moments

Product Moments about the Origin

We have already discussed the expected value of a bivariate function $g(X, Y)$, where $E(g(X, Y)) = \sum_x \sum_y g(x, y) \cdot f(x, y)$.

The following is a special case of this:

4.31 Definition. The *rth and sth product moment about the origin* of X and Y is the expected value of X^rY^s:

$$\text{discrete:} \qquad E(X^rY^s) = \sum_x \sum_y x^r y^s \cdot f(x,y)$$

$$\text{continuous:} \qquad E(X^rY^s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s \cdot f(x,y)\, dx\, dy$$

for $r = 0, 1, 2, \ldots$, $s = 0, 1, 2, \ldots$.

We will denote $E(X)$ by μ_X and $E(Y)$ by μ_Y.

Product Moments about the Means

4.32 Definition. The *rth and sth product moment about the mean* of X and Y is the expected value of $(X - \mu_X)^r(Y - \mu_Y)^s$:

$$\text{discrete:} \qquad E((X-\mu_X)^r(Y-\mu_Y)^s) = \sum_x \sum_y (x-\mu_X)^r(y-\mu_Y)^s \cdot f(x,y)$$

$$\text{continuous:} \qquad E((X-\mu_X)^r(Y-\mu_Y)^s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_X)^r(y-\mu_Y)^s \cdot f(x,y)\, dx\, dy$$

Covariance

4.33 Definition. The 1st and 1st product moment about the means of X and Y is called the *covariance* of X and Y. It is commonly denoted σ_{XY}, or $\text{cov}(X, Y)$.

Summary:

$$\mu_X = E(X) = \sum_x \sum_y x \cdot f(x,y)$$

$$\mu_Y = E(Y) = \sum_x \sum_y y \cdot f(x,y)$$

$$\text{cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = \sum_x \sum_y (x-\mu_X)(y-\mu_Y) \cdot f(x,y)$$

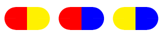
(discrete case shown here, continuous case is analogous)
The covariance of describes the relationship between X and Y.

If there is a high probability that large X values and large Y values appear together, then the covariance is positive (or small X with small Y). On the other hand large/small X values occurring with small/large Y values is more likely, then the covariance will be negative. (again, only discrete shown above)

Just as with σ^2 for single variable case, we have a "shortcut" formula:

4.34 Theorem.

$$\text{cov}(X, Y) = \sigma_{XY} = E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y$$



		x: 0	1	2
	0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$
y	1	$\frac{8}{36}$	$\frac{6}{36}$	
	2	$\frac{1}{36}$		

4.35 Example. In the caplet example, find the covariance of X and Y.

Solution. We start by finding μ_X and μ_Y:

$$\mu_X = E(X) = \sum_x \sum_y x \cdot f(x, y) = \sum_x x \sum_y f(x, y) = \sum_x x \cdot g(x)$$

$$\mu_Y = E(Y) = \sum_x \sum_y y \cdot f(x, y) = \sum_y y \sum_x f(x, y) = \sum_y y \cdot h(y)$$

where $g(x)$ and $h(y)$ are the marginal distributions of x and y resp.

		x: 0	1	2	
	0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$	$\frac{21}{36}$
y	1	$\frac{8}{36}$	$\frac{6}{36}$		$\frac{14}{36}$
	2	$\frac{1}{36}$			$\frac{1}{36}$
		$\frac{15}{36}$	$\frac{18}{36}$	$\frac{3}{36}$	

Thus

$$\mu_X = 0 \cdot \frac{15}{36} + 1 \cdot \frac{18}{36} + 2 \cdot \frac{3}{36} = \frac{24}{36}$$
$$\mu_Y = 0 \cdot \frac{21}{36} + 1 \cdot \frac{14}{36} + 2 \cdot \frac{1}{36} = \frac{16}{36}$$

Using $\sigma_{XY} = E(XY) - \mu_X \mu_Y$, we now need $E(XY)$:

$$\begin{aligned} &= \sum_{x=0}^{2} \sum_{y=0}^{2} (xy) \cdot f(x,y) \\ &= (0 \cdot 0) \cdot \frac{6}{36} + (0 \cdot 1) \cdot \frac{8}{36} + (0 \cdot 2) \cdot \frac{1}{36} \\ &\quad + (1 \cdot 0) \cdot \frac{12}{36} + (1 \cdot 1) \cdot \frac{6}{36} + (2 \cdot 0) \cdot \frac{3}{36} \\ &= \frac{6}{36}. \end{aligned}$$

So we have $\sigma_{XY} = \frac{6}{36} - (\frac{24}{36})(\frac{16}{36}) = -\frac{7}{54} \approx -0.1296$. □

Covariance and Independence

Recall that joint random variables X and Y are independent if and only if $f(x,y) = g(x) \cdot h(y)$; i.e. their joint distribution is the product of the marginal distributions. From this it follows that:

4.36 Theorem. If X and Y are independent, then

$$E(XY) = E(X) \cdot E(Y)$$

and $\sigma_{XY} = 0$. (exercise the proof)

This theorem is only a one-way implication as seen in the next example.

4.37 Example. Let X and Y be discrete random variables with joint distribution given by,

		x = -1	x = 0	x = 1
y	-1	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$
	0	0	0	0
	1	$\frac{1}{6}$	0	$\frac{1}{6}$

Find $\text{cov}(X,Y)$ and determine whether X and Y are independent.

To find $\text{cov}(X,Y)$ first we need the marginal distributions, $g(x)$ and $h(y)$.

		$x=-1$	$x=0$	$x=1$	
y	-1	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{2}{3}$
	0	0	0	0	0
	1	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{3}$
		$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

Then we can find μ_X and μ_Y:

$$\mu_X = \sum_x x \cdot g(x) = (-1) \cdot \frac{1}{3} + (0) \cdot \frac{1}{3} + (1) \cdot \frac{1}{3} = 0$$

$$\mu_Y = \sum_y y \cdot h(y) = (-1) \cdot \frac{2}{3} + (0) \cdot \frac{1}{3} + (1) \cdot \frac{1}{3} = -\frac{1}{3}$$

Next we need $E(XY) = \sum_x \sum_y xy \cdot f(x,y)$:

$= ((-1) \cdot (-1)) \cdot \frac{1}{6} + ((-1) \cdot 0) \cdot 0 + ((-1) \cdot 1) \cdot \frac{1}{6} + (0 \cdot (-1)) \cdot \frac{1}{3} + (0 \cdot 0) \cdot 0 + (0 \cdot 1) \cdot 0 + (1 \cdot (-1)) \cdot \frac{1}{6} + (1 \cdot 0) \cdot 0 + (1 \cdot 1) \cdot \frac{1}{6} = 0.$

So $\text{cov}(X,Y) = E(XY) - \mu_X \mu_Y = 0 - (0)(-\frac{1}{3}) = 0.$

The random variables are not independent however, as can be seen by the example that $f(-1,-1) = \frac{1}{6} \neq g(-1) \cdot h(-1) = \frac{2}{9}$.

		$x=0$	$x=1$	$x=2$	
y	0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$	$\frac{21}{36}$
	1	$\frac{8}{36}$	$\frac{6}{36}$		$\frac{14}{36}$
	2	$\frac{1}{36}$			$\frac{1}{36}$
		$\frac{15}{36}$	$\frac{18}{36}$	$\frac{3}{36}$	

Are the random variables X and Y of the caplet example independent?

No. We found that $\text{cov}(X,Y) = -\frac{7}{54} \neq 0$. Therefore they could not be independent.

4.38 Example. Let X and Y be jointly continuous random variables with joint density

$$f(x) = \begin{cases} \frac{3x^2y}{28} & \text{for } 1 < x < 2, 1 < y < 3 \\ 0 & \text{otherwise} \end{cases}$$

Find the covariance of X and Y and determine whether they are independent.

4.9 Conditional Expectations

Earlier we defined conditional probability $f(x|y)$, for joint random variables X and Y, we can also talk about conditional expectation.

4.39 Definition. Let X and Y have probability distribution/density $f(x,y)$, and let $u(X)$ be some function of X. The *conditional expected value of* $u(X)$ *given* $Y = y$ is

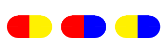
$$\text{discrete case:} \quad E(u(X)|Y=y) = \sum_x u(x) \cdot f(x|y)$$

$$\text{continuous case:} \quad E(u(X)|Y=y) = \int_x u(x) \cdot f(x|y)\,dx$$

The special case $E(X|Y=y)$ is called the *conditional mean of* X *given* $Y = y$.

(*we may write* $E(u(X)|Y=y)$ *more compactly as* $E(u(X)|y)$)

4.40 Example. In the caplet example:



		$x = 0$	$x = 1$	$x = 2$	
y	0	$\frac{6}{36}$	$\frac{12}{36}$	$\frac{3}{36}$	$\frac{21}{36}$
	1	$\frac{8}{36}$	$\frac{6}{36}$		$\frac{14}{36}$
	2	$\frac{1}{36}$			$\frac{1}{36}$
		$\frac{15}{36}$	$\frac{18}{36}$	$\frac{3}{36}$	

Find the expected value (or conditional mean) of X given that $Y = 1$.

Solution. By definition,

$$E(X|1) = \sum_x x f(x|1).$$

Recall that $f(x|y) = \frac{f(x,y)}{h(y)}$, and so

$$f(0|1) = \frac{8}{14}, \quad f(1|1) = \frac{6}{14}, \quad f(2|1) = 0.$$

Therefore we have

$$E(X|1) = 0 \cdot \frac{8}{14} + 1 \cdot \frac{6}{14} + 2 \cdot 0 = \frac{6}{14} \approx 0.4286.$$

□

4.41 Example. Let X be the amount a salesperson person spends on gas in a day, and Y be the amount of money for which they are reimbursed. The joint density of X and Y is

$$f(x,y) = \begin{cases} \frac{1}{25}\left(\frac{20-x}{x}\right) & \text{for } 10 < x < 20, \frac{x}{2} < y < x \\ 0 & \text{otherwise} \end{cases}$$

(gives the probability (density) that they will be reimbursed y dollars after spending x dollars)

Find, $f(y|x)$, the conditional probability of Y given $X = x$, and use it to find the probability of being reimbursed at least \$8 given that \$12 we spent. What is the expected reimbursement given that \$12 were spent?

4.10 Expectation by Conditioning

Let $E(X|Y)$ denote the conditional expectation of X given an arbitrary value for y; i.e. $E(X|Y)$ is the function of Y which outputs $E(X|Y = y)$ for whatever y we choose. This makes $E(X|Y)$ itself a random variable, and it can be helpful in computing expected values in certain situations.

4.42 Theorem. Let X and Y be joint random variables. Then

$$E(X) = E(E(X|Y))$$

This says

$$\text{discrete case:} \quad E(X) = \sum_y E(X|Y = y)h(y)$$

$$\text{continuous case:} \quad E(X) = \int_{-\infty}^{\infty} E(X|Y = y)h(y)\, dy$$

where $h(y)$ is the marginal distribution/density for Y.

Proof. Let $f(x, y)$ be the joint distribution for discrete random variables X and Y, and $g(x)$, $h(y)$ their respective marginal distributions. Then

$$\begin{aligned}\sum_y E(X|Y=y)h(y) &= \sum_y \sum_x xf(x|y)h(y) \\ &= \sum_y \sum_x x\frac{f(x,y)}{h(y)}h(y) \\ &= \sum_y \sum_x xf(x,y) \\ &= \sum_x x \sum_y f(x,y) \\ &= \sum_x xg(x) \\ &= E(x)\end{aligned}$$

(Continuous case similar) □

4.43 Example. You are wandering lost in a cave and arrive at a spot with three tunnel entrances. Tunnel 1 leads to the exit after walking for 3 hours, tunnel 2 leads back to this same spot after 5 hours of walking, and tunnel 3 leads back to this same spot after 7 hours of walking. If you return to this spot, the 3 tunnel entrances are indistinguishable, and you will proceed by randomly choosing one of the 3 tunnels. What is your expected length of time needed to exit the cave?

Solution. Let X be the number of hours until the exit is reached, and Y be the tunnel you choose first. By the theorem

$$E(X) = E(E(X|Y)) = \sum_{y=0}^{2} E(X|Y=y)h(y) = \frac{1}{3}\sum_{y=0}^{2} E(X|Y=y)$$

where $h(y) = \frac{1}{3}$ as each tunnel is equally likely. Now

$$E(X|Y=1) = 3,\ E(X|Y=2) = 5 + E(X),\ [E(X|Y=1) = 7 + E(X)$$

since tunnel 2 or 3 returns to the same spot after 5 or 7 hours (respectively) and puts you in the position you started. Then

$$E(X) = \frac{1}{3}(3 + 5 + E(X) + 7 + E(X)),$$

from which we can solve $E(X) = 15$. □

Chapter 5

Special Probability Distributions

This chapter presents some commonly used probability distributions for discrete random variables.

Having a pre-determined probability distribution to model a chance experiment prevents from having to re-derive its properties each time (e.g. mean and variance).

The models presented depend on *parameters*; input values which tailor the probability distribution to the particular example.

In some cases the values of the distribution for a range of parameters are recorded in a table which can be used to evaluate probabilities, rather than computing the sums directly, (or integrating in the case of continuous random variables). This is not only a convenience, in some cases it may be impractical to compute such values on the spot, or impossible if, for example, no exact expression exists for an integral.

5.1 The Discrete Uniform Distribution

Suppose a random variable X has a finite range of k values, $\{x_1, x_2, \ldots, x_k\}$. Then X has *discrete uniform distribution* if

$$f(x) = \frac{1}{k}$$

for each $x \in \{x_1, x_2, \ldots, x_k\}$. In other words each outcome is equally likely. (example: rolling a balanced die)

Our only parameter in this case is k. For the discrete uniform distribution:

$$\mu = \sum_{i=1}^{k} x_i f(x_i) = \frac{\sum_{i=1}^{k} x_i}{k}.$$

$$\sigma^2 = \sum_{i=1}^{k} (x_i - \mu)^2 f(x_i) = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k} = \frac{\sum_{i=1}^{k} x_i^2}{k} - \left(\frac{\sum_{i=1}^{k} x_i}{k} \right)^2$$

5.2 The Bernoulli Distribution

Consider an experiment with two possible outcomes, either success or failure. (example: single coin toss)

Assign random variable X the value 1 for success and 0 for failure.

If the probability of success is θ, then the probability of a failure is $1 - \theta$.

In this case X is called a *Bernoulli random variable* and has *Bernoulli distribution* given by

$$f(x; \theta) = \theta^x (1 - \theta)^{1-x} \quad \text{for } x = 0, 1.$$

An experiment with a Bernoulli distribution is referred to as a *Bernoulli trial.*

5.1 Example. If X is a Bernoulli random variable show that

$$\mu = \theta,$$

and

$$\sigma^2 = \theta(1 - \theta)$$

5.3 The Binomial Distribution

Now consider an experiment with repeated trials, in which the outcome of each trial is either a success or failure, and the trials are independent.

Random variable X will denote the number of successes, the probability of success is known to be θ, and n is the given number of trials in the experiment.

Then X has *binomial distribution* which is given by

$$b(x; n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad \text{for } x = 0, 1, \ldots, n.$$

Random variable X is called a *binomial random variable* if and only if it has this distribution.

The Bernoulli distribution is the special case of the binomial distribution when $n = 1$; a single trial experiment.

5.2 Example. Some examples of binomial random variables:

- Number of heads in 35 flips of a coin with 0.63 probability of heads and 0.37 probability of tails.

$$P(17 \text{ heads}) = b(17; 35, 0.63) = \binom{35}{17}(0.63)^{17}(0.37)^{18} \approx 0.02973.$$

- There is a 6.6% chance that a person has O- blood type. In a selection of 20 people what is the probability that 5 of them will have O- blood.

$$P(5 \text{ people}) = b(5; 20, 0.066) = \binom{20}{5}(0.066)^{5}(0.934)^{15} \approx 0.006972.$$

5.3 Example. Find the probability that seven of ten people will recover from a disease if we assume independence and the probability of 0.80 that any one of them will recover from the disease.

$$b(7; 10, 0.80) = \binom{10}{7}(0.80)^{7}(0.20)^{3} \approx 0.2013.$$

5.4 Theorem.

$$b(x; n, \theta) = b(n - x; n, 1 - \theta)$$

For example,

$$b(7; 11, 0.75) = b(4; 11, 0.25) = 0.1721$$

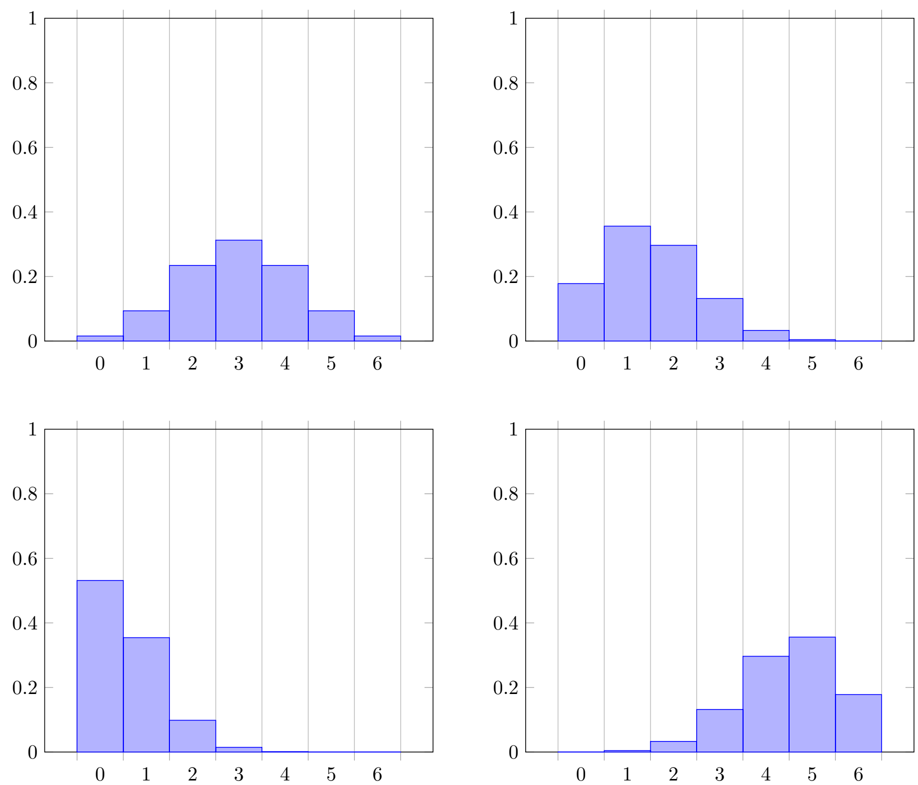
(prove this theorem as an exercise)

This table gives the binomial distribution when $n = 6$ for different values of θ up to 0.50.

x	$\binom{n}{x}$	0.01	0.05	0.10	0.15	0.20	0.25
0	1	0.9415	0.7351	0.5314	0.3771	0.2621	0.1780
1	6	0.0571	0.2321	0.3543	0.3993	0.3932	0.3560
2	15	0.0014	0.0305	0.0984	0.1762	0.2458	0.2966
3	20		0.0021	0.0146	0.0415	0.0819	0.1318
4	15		0.0001	0.0012	0.0055	0.0154	0.0330
5	6			0.0001	0.0004	0.0015	0.0044
6	1					0.0001	0.0002

x	$\binom{n}{x}$	0.30	0.35	0.40	0.45	0.49	0.50
0	1	0.1176	0.0754	0.0467	0.0277	0.0176	0.0156
1	6	0.3025	0.2437	0.1866	0.1359	0.1014	0.0938
2	15	0.3241	0.3280	0.3110	0.2780	0.2436	0.2344
3	20	0.1852	0.2355	0.2765	0.3032	0.3121	0.3125
4	15	0.0595	0.0951	0.1382	0.1861	0.2249	0.2344
5	6	0.0102	0.0205	0.0369	0.0609	0.0864	0.0938
6	1	0.0007	0.0018	0.0041	0.0083	0.0183	0.0156

5.5 Example.



For each graph of $b(x; n, \theta)$ we have $n = 6$. Determine which of these has $\theta = 0.1, 0.25, 0.5,$ and 0.75

5.6 Example. Screws produced by a certain company have a 0.01 probability of begin defective (independently of one another). The company sells the screws in packaged of 10 and will replace the package if more than 1 screw is defective in the package.

What is the probability that more than 1 screw is defective?

Solution. Let X be the number of defective screws in a package. Then X has binomial distribution with $n = 10$ and $\theta = 0.01$; i.e. the probability of having x defective screws in a package is

$$b(x; 10, 0.01) = \binom{10}{x}(0.01)^x(0.99)^{10-x}$$

The probability that more than 1 screw is defective is

$$P(X \geq 2) = \sum_{x=2}^{10}\binom{10}{x}(0.01)^x(0.99)^{10-x}$$

However, it is easier to calculate as

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1)$$

$$= 1 - \binom{10}{0}(0.01)^0(0.99)^{10} - \binom{10}{1}(0.01)^1(0.99)^9 \approx 0.04.$$

□

Moments of the Binomial Distribution

5.7 Theorem. The mean and variance of the binomial distribution:

$$\mu = n\theta, \quad \sigma^2 = n\theta(1-\theta)$$

See text for direct proof, or use:

5.8 Theorem. The moment generating function for a random variable with the binomial distribution $b(x; n, \theta)$ is

$$M_X(t) = (1 + \theta(e^t - 1))^n.$$

Proof.

$$\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{x=0}^{n}(e^{tx}) \cdot \binom{n}{x}\theta^x(1-\theta)^{n-x} \\
&= \sum_{x=0}^{n}\binom{n}{x}(e^t\theta)^x(1-\theta)^{n-x} \\
&= (e^t\theta + (1-\theta))^n \quad \text{(binomial expansion thereom)} \\
&= (1 + \theta(e^t - 1))^n.
\end{aligned}$$

□

Finding the mean from the moment generating function:

$$\begin{aligned}
\mu &= \frac{d}{dt} M_X(t)\Big|_{t=0} \\
&= \frac{d}{dt}(1+\theta(e^t-1))^n\Big|_{t=0} \\
&= n(1+\theta(e^t-1))^{n-1}\cdot(\theta e^t)\big|_{t=0} \\
&= n(1+\theta(e^0-1))^{n-1}\cdot(\theta e^0) \\
&= n\theta.
\end{aligned}$$

Next we want the second moment about the origin:

$$\begin{aligned}
E(X^2) &= \frac{d^2}{dt^2} M_X(t)\Big|_{t=0} \\
&= \frac{d^2}{dt^2}(1+\theta(e^t-1))^n\Big|_{t=0} \\
&= \frac{d}{dt} n\theta e^t(1+\theta(e^t-1))^{n-1}\Big|_{t=0} \\
&= n\theta e^t(1+\theta(e^t-1))^{n-1} \\
&\quad + n(n-1)\theta e^t(1+\theta(e^t-1))^{n-1}\cdot(\theta e^t)\big|_{t=0} \\
&= n\theta + n(n-1)\theta^2
\end{aligned}$$

Finally,

$$\sigma^2 = E(X^2) = \mu^2 = n\theta + n(n-1)\theta^2 - (n\theta)^2 = n\theta - n\theta^2 = n\theta(1-\theta).$$

5.9 Theorem. Let X be a binomial random variable and let $Y = \frac{X}{n}$. Then

$$E(Y) = \theta, \quad \sigma_Y^2 = \frac{\theta(1-\theta)}{n}.$$

Random variable Y denotes the proportion of successes in n trials.

By Chebyshev's Theorem, with $C = k\sigma$, or $k = \frac{C}{\sigma}$ we have

$$P(|Y-\theta| < C) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{(\frac{C^2}{\sigma^2})} = 1 - \frac{\theta(1-\theta)}{C^2 n}$$

($\mu = \theta$ in this case)

Thus for any value of $C > 0$ we have

$$P\left(\left|\frac{X}{n} - \theta\right| < C\right) \geq 1 - \frac{\theta(1-\theta)}{C^2 n}.$$

When n is large, the fraction on the right side gets small, and so

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - \theta\right| < C\right) = 1.$$

But this holds for any $C > 0$, no matter how small.

In this case, the more trials we perform, the more likely it is that the proportion of successes will be close to the probability of a success θ.

Example: In repeatedly flipping a balanced coin, the more flips we perform (n), the more likely that the proportion of heads obtained ($\frac{X}{n}$) will be 0.5 (θ).

5.4 The Negative Binomial Distribution

The Binomial distribution gives the probability of getting x successes in n trials.

Suppose we want to know the probability that the kth success occurs precisely on trial x.

For the kth success to occur on the xth trial, there must be exactly $k-1$ success on the first $x-1$ trials, and consequently $x-1-(k-1) = x-k$ failures.

If θ is the probability of a success on a given trial then the probability of getting $k-1$ successes in $x-1$ trails is

$$b(k-1; x-1, \theta) = \binom{x-1}{k-1}\theta^{k-1}(1-\theta)^{x-k}.$$

Then the probability that the kth success is on trial x is

$$\theta \cdot b(k-1; x-1, \theta) = \binom{x-1}{k-1}\theta^{k}(1-\theta)^{x-k}.$$

A random variable X has *a negative binomial random variable* if and only if it has *negative binomial distribution*,

$$b^*(k-1; x-1, \theta) = \binom{x-1}{k-1}\theta^{k}(1-\theta)^{x-k}$$

for $x = k, k+1, k+2, \ldots$.

It follows that

5.10 Theorem.

$$b^*(x; k, \theta) = \frac{k}{x} \cdot b(k, x, \theta)$$

So table values for the binomial distribution can be used to find values for the negative binomial distribution.

Mean and Variance of the Negative Binomial Distribution

We can use the theorem to compute the mean and variance from that of the binomial distribution. They are

$$\mu = \frac{k}{\theta}, \quad \sigma^2 = \frac{k}{\theta}\left(\frac{1}{\theta} - 1\right)$$

The Geometric Distribution

The special case when $k = 1$ (first success appears in trial x) is called the *geometric distribution*:

$$g(x;\theta) = b^*(x;1,\theta) = \theta(1-\theta)^{x-1}.$$

Chapter 6

Special Probability Densities

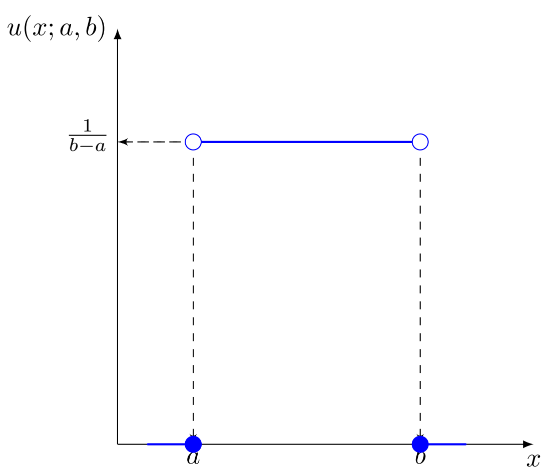
Just as was done in Chapter 5, we now present some common probability densities in the case of a continuous random variable.

6.1 The Uniform Distribution

A continuous random variable X is said to have *uniform distribution* if and only if its probability density function is given by

$$u(x; a, b) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

This means that if (a_1, b_1) and (a_2, b_2) are two intervals of equal length inside of (a, b), then $P(a_1 < X < b_2) = P(a_2 < X < b_2)$.



6.2 The Normal Distribution

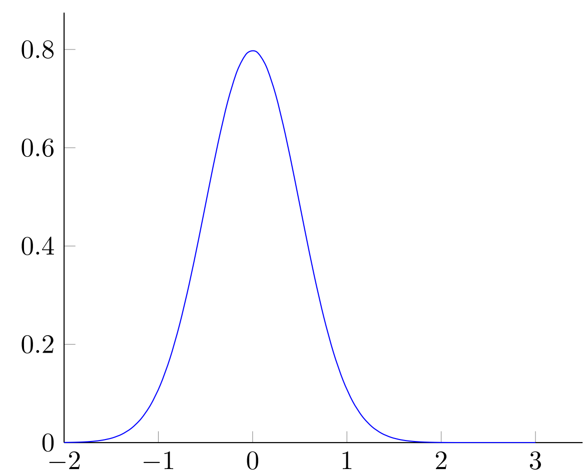
A continuous random variable X has *normal distribution*, and is called a *normal random variable* if and only if its probability density is given by

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

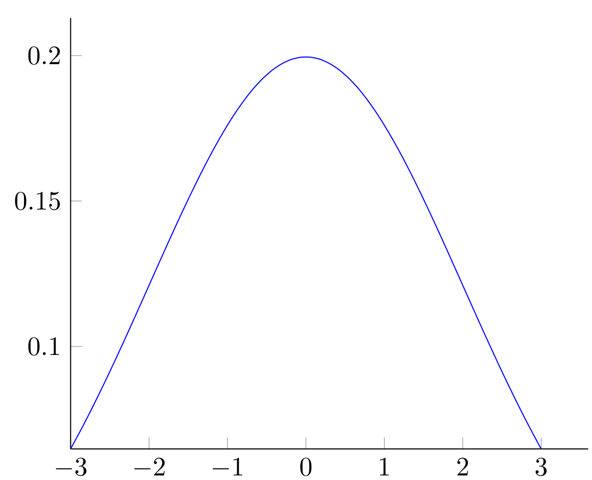
for all $x \in \mathbb{R}$, where $\sigma > 0$.

Showing that this function integrates to 1 over $\mathbb{R}$ requires a trick involving a change of variables to polar coordinates (found in a multivariable calculus course); we will omit this here and accept that this is a valid probability density for any μ and any $\sigma > 0$.

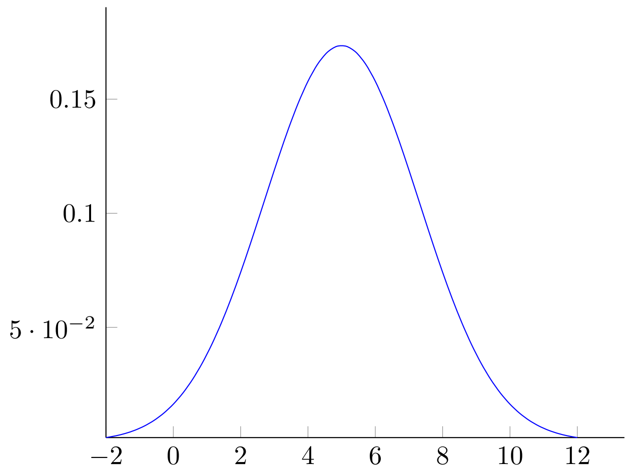
Plot of the Normal Distribution when $\mu = 0$, $\sigma = 0.5$:



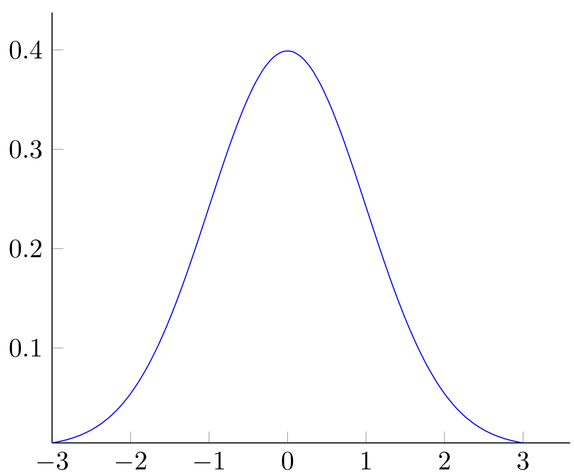
Plot of the Normal Distribution when $\mu = 0$, $\sigma = 2$:



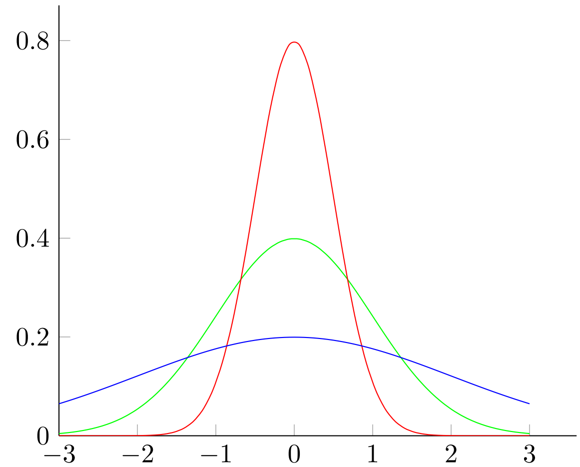
Plot of the Normal Distribution when $\mu = 5$, $\sigma = 2.3$:



Plot of the Normal Distribution when $\mu = 0$, $\sigma = 1$:



Plot of Normal Distributions with $\mu = 0$:



Red - $\sigma = 0.5$, Green - $\sigma = 1$, Blue - $\sigma = 2$

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

One thing to notice about these graphs is their "bell" shape. There is a higher probability density in the middle, which rapidly decreases as we move outward.

The next thing to notice is that these graphs have symmetry about the value μ, which is due to the $\left(\frac{x-\mu}{\sigma}\right)^2$ in the exponent.

It also appears that the σ parameter controls the dispersion of the probability.

Indeed, μ and σ are the mean and standard deviation of a normally distributed random variable X, as we will see.

Moment Generating Function

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \, dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}\left(-2xt\sigma^2 + (x-\mu)^2\right)} \, dx \end{aligned}$$

In the exponent we have:

$$-2xt\sigma^2 + (x-\mu)^2 = -2xt\sigma^2 + x^2 - 2x\mu + \mu^2 = x^2 - 2x(\mu + t\sigma^2) + \mu^2.$$

Completing the square gives:

$$x^2 - 2x(\mu + t\sigma^2) + \mu^2 = (x - (\mu + t\sigma^2))^2 - 2\mu t\sigma^2 - t^2\sigma^4.$$

This allows us to write

$$M_X(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2} \left(\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x-(\mu+t\sigma^2)}{\sigma}\right)^2} dx \right) = e^{\mu t + \frac{1}{2}t^2\sigma^2}$$

Mean and Variance

The moment generating function for normally distributed random variable X is:

$$M_X(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2}$$

Show that the mean and variance of X are indeed μ and σ^2.

First derivative:

$$\frac{d}{dt}M_X(t) = \frac{d}{dt}e^{\mu t + \frac{1}{2}t^2\sigma^2} = e^{\mu t + \frac{1}{2}t^2\sigma^2} \cdot (\mu + \sigma^2 t)$$

Second derivative:

$$\begin{aligned}\frac{d^2}{dt^2}M_X(t) &= \frac{d}{dt}\left(\mu e^{\mu t + \frac{1}{2}t^2\sigma^2} + \sigma^2 t e^{\mu t + \frac{1}{2}t^2\sigma^2}\right) \\ &= \mu e^{\mu t + \frac{1}{2}t^2\sigma^2} \cdot (\mu + \sigma^2 t) + \sigma^2 e^{\mu t + \frac{1}{2}t^2\sigma^2} \\ &\quad + \sigma^2 t e^{\mu t + \frac{1}{2}t^2\sigma^2} \cdot (\mu + \sigma^2 t)\end{aligned}$$

Setting $t = 0$ in both gives:

$$\frac{d}{dt}M_X(t) = \mu \qquad \frac{d^2}{dt^2}M_X(t) = \mu^2 + \sigma^2$$

Therefore the mean, $E(X)$, is μ and the variance is $E(X^2) - \mu^2 = \sigma^2$.

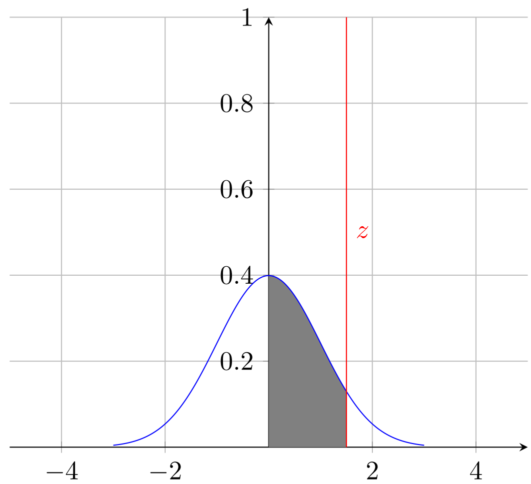
The Standard Normal Distribution

The normal distribution with $\mu = 0$ and $\sigma = 1$ is called the *standard normal distribution*.

$$n(x; 0, 1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

Probabilities for the standard normal distribution may be found by way of a table of "pre-calculated" probabilities.

For example, the table below gives $P(0 \le X \le z)$ for various z values. Graphically this looks like,



Standard Normal Table

The following table contains values for probabilities $P(0 \leq Z < z)$, where Z has standard normal distribution.

Table III: Standard Normal Distribution

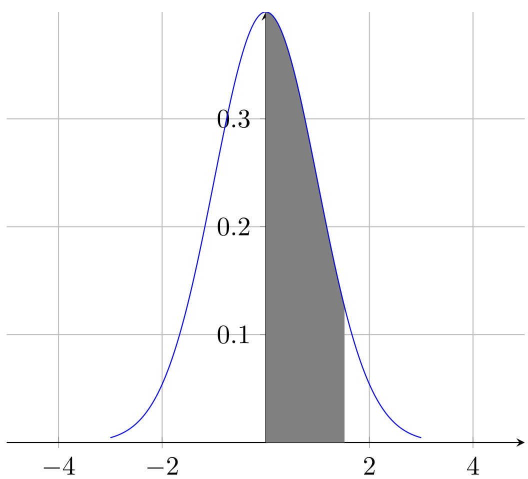
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4988
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Here is an example showing how to find $P(0 \leq Z < 1.52)$ from the table:

Table III: Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4988
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

A plot of the standard normal density function, where the shaded region is $P(0 \leq Z < 1.52)$:

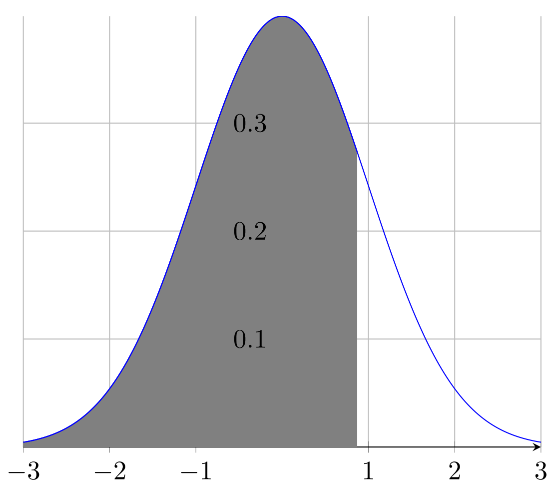


$$P(0 \leq X \leq 1.52)$$
$$= 0.4357$$

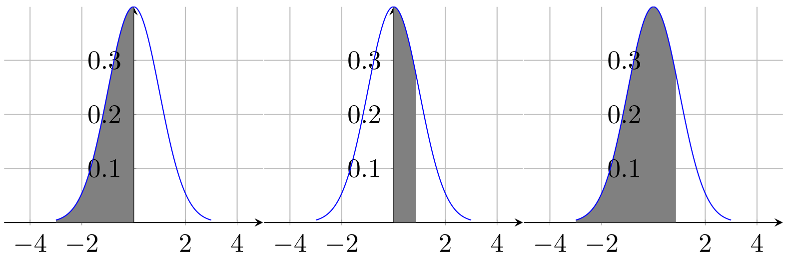
We noted earlier that μ, in this case 0, is the midpoint of the graph. Thus to find $P(X \leq z)$, we look up our value of z in the table, then add 0.5.

Table III: Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4988
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990



$$P(X \leq 0.87) = 0.3087 + 0.5 = 0.8087$$

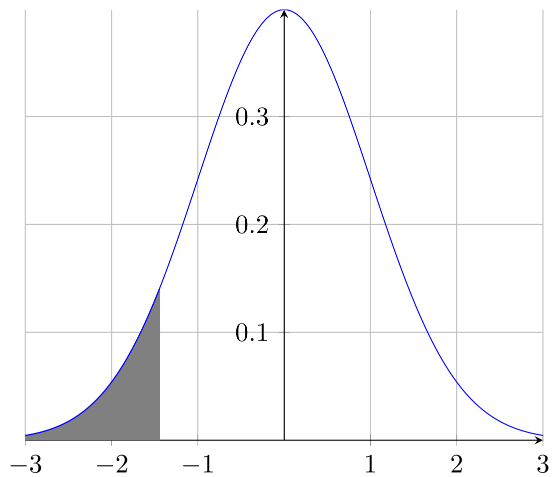


Adding the first two areas gives the third.

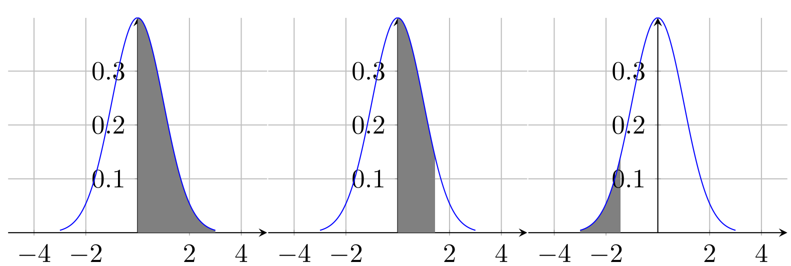
If $z < 0$, we find $P(X \leq z)$ by finding $0.5 - P(X \leq |z|)$.

Table III: Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4988
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990



$$P(X \leq -1.44) = 0.5 - 0.4251 = 0.0749$$

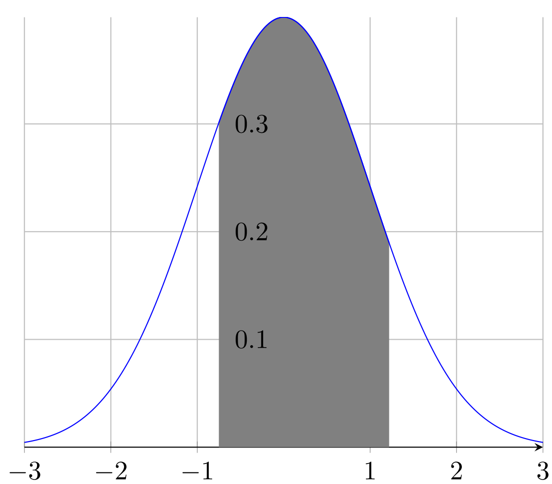


The difference of the first two areas is equal to the third.

If X has standard normal distribution, find $P(-0.75 \leq X \leq 1.22)$

Table III: Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4988
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

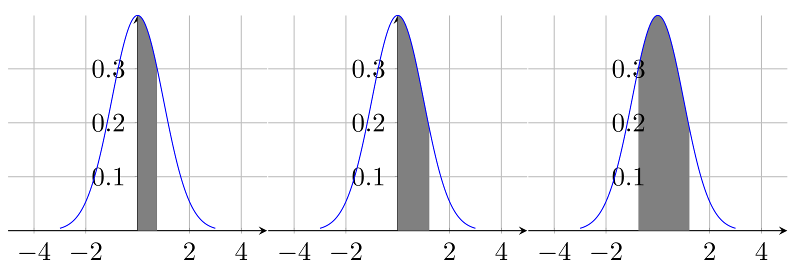


To find $P(-0.75 \leq X \leq 1.22)$, add $P(0 \leq X \leq 0.75)$ to $P(0 \leq X \leq 1.22)$

Table III: Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4988
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

$$P(-0.75 \leq X \leq 1.22) = 0.2734 + 0.3888 = 0.6622$$



$$P(0 \leq X \leq 0.75) + P(0 \leq X \leq 1.22) = P(-0.75 \leq X \leq 1.22)$$

The sum of the first two areas equals the third. (This uses the symmetry of the graph)

A couple of rules when using the table:

- For z values not found on the table we may simply choose the closest value.
- If our z value is exactly the midpoint between two z values on the table, then we can average the two probabilities.

6.1 Theorem. If X has a normal distribution with mean μ and standard deviation σ then

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

is a random variable having the standard normal distribution.

This allows us to compute probabilities for non-standard normal distributions with the standard normal table.

Proof. Let $Z = \frac{X-\mu}{\sigma}$. First note that

$$x_1 < X < x_2 \quad \Leftrightarrow \quad z_1 = \frac{x_1 - \mu}{\sigma} < Z < \frac{x_2 - \mu}{\sigma} = z_2$$

Then, using the substitution rule for integrals,

$$\begin{aligned} P(x_1 < X < x_2) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}(z)^2} \, dz \\ &= P(z_1 < Z < z_2). \end{aligned}$$

Therefore $P(x_1 < X < x_2) = P\left(\frac{x_1-\mu}{\sigma} < Z < \frac{x_2-\mu}{\sigma}\right)$, and we are able to look this up on the table. □

Non-standard Normal Distribution

6.2 Example. Let X be a continuous random variable with normal distribution $n(x; 70, 4)$; i.e. $\mu = 70, \sigma = 4$. Find

- $P(68 \leq X \leq 74)$

  By the theorem

$$\begin{aligned} P(68 \leq X \leq 74) &= P\left(\frac{68 - 70}{4} \leq Z \leq \frac{74 - 70}{4}\right) \\ &= P(-0.5 \leq Z \leq 1) \end{aligned}$$

Then, by the symmetry in the graph,

$$\begin{aligned} P(-0.5 \leq Z \leq 1) &= P(Z \leq 0.5) + P(Z \leq 1) \\ &= 0.1915 + 0.3413 = 0.5328 \end{aligned}$$

6.3 The Normal Approximation of the Binomial Distribution

If X is a random variable with binomial distribution $b(x; n, \theta)$, then the normal distribution $n(x; n\theta, \sqrt{n\theta(1-\theta)})$, with mean $n\theta$ and standard deviation $\sqrt{n\theta(1-\theta)}$, gives an approximation of the binomial distribution. (see text)

To use this approximation, we need to “convert” the discrete binomial random variable to the continuous case. Here $P(X = k)$ will be a approximated with the normal distribution by integrating from $k - 0.5$ to $k + 0.5$. This is called the *continuity correction*.

6.3 Example. Find the probability of getting 6 heads in 16 flips of a balanced coin. (binomial distribution)

Approximate this with the normal distribution.

Bibliography

[1] Irwin Miller, and Marylees Miller, John E. Freund's Mathematical Statistics with Applications, Sixth Edition. Pearson Prentice Hall Inc. (1998).

[2] Sheldon Ross, A First Course in Probability, Seventh Edition. Pearson Prentice Hall Inc (2006).

Index