DSAI 510: Basic Concepts of Probability Theory

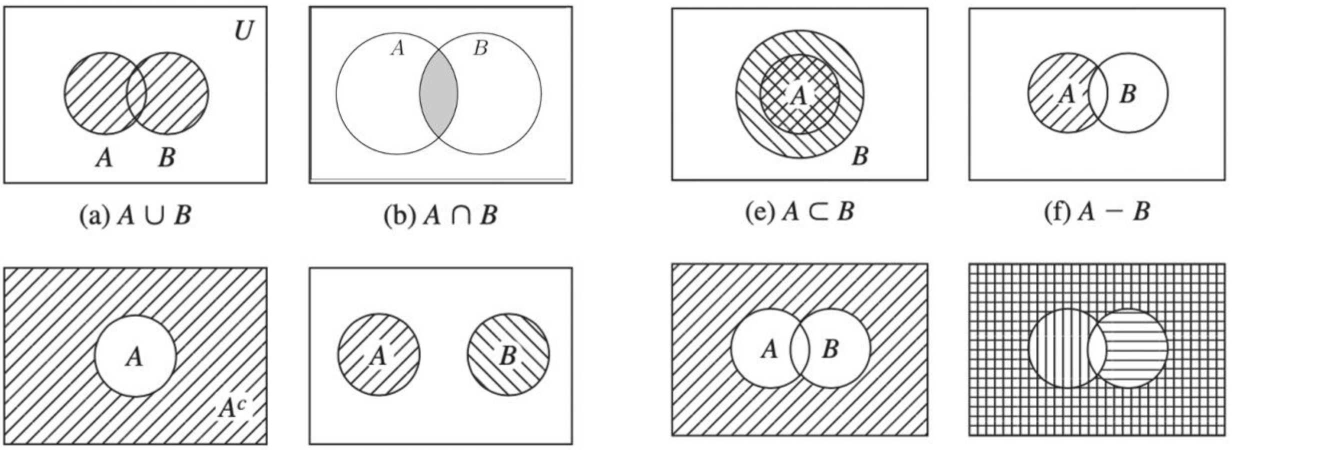
A **random experiment** is an experiment in which the outcome varies in an unpredictable fashion when the experiment is repeated under the same conditions. A **random experiment** is specified by stating an experimental procedure and a set of one or more measurements or observations.

Experiment E_1: Pick one number at random between 1 and 6.
Experiment E_2: Toss a coin twice and note the sequence of heads and tails.
Experiment E_3: Toss a coin twice and note the number of heads.
Experiment E_4: Toss an unfair coin twice and note the number of heads.
(Observations in E_2 and E_3 are not the same despite the apparent similarity!)

Sample space: Set of all possible outcomes. It depends on the how the experiment is designed. For the experiments mentioned above, sample space for each is
S_1={2,3,4,5}, S_2= {HH,TT,TH,HT}, S_3= {0,1,2}, S_4= {0,1,2}

Event: Possible outcome of an experiment. Subsets of sample space S.
Example: Getting A={HT} in experiment E_2, "Toss a coin twice and note the sequence of heads and tails."

Review of set theory:
Sample space S is the universal set. ∪ is the "union" sign, ∩ is the "intersection" sign and A^c means the "complement" of the set A.
Venn diagrams:



(a) $A \cup B$ (b) $A \cap B$ (e) $A \subset B$ (f) $A - B$

(c) A^c (d) $A \cap B = \varnothing$ (g) $(A \cup B)^c$ (h) $A^c \cap B^c$ =square shaded part

Rule for combination of two probabilities

$$\boxed{P[A \cup B] = P[A] + P[B] - P[A \cap B]}$$

The last term is needed to avoid double counting. You can understand the necessity of the last term by looking at the Venn diagram (b).

Example: In a certain school, 50% of students are part of the Math Club, while 40% are members of the Science Club. If 20% of students are members of both clubs, what's the probability that a randomly selected student is a member of either the Math Club or the Science Club or both?

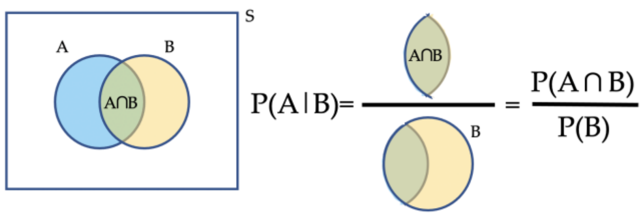
The probability that a student is part of the Math Club, $P(M)$, is 50% or 0.50.
The probability that a student is part of the Science Club, $P(S)$, is 40% or 0.40.
The probability that a student is part of both clubs, $P(M \cap S)$, is 20% or 0.20.

Using the formula for the union of two sets:
$P(M \cup S)=P(M)+P(S)-P(M \cap S) = 0.50+0.40-0.20=0.70$

So, there's a 70% chance that a randomly selected student is a member of either the Math Club, the Science Club, or both.

2.4 Conditional probability

What's the probability of the events happening within the set $A \cap B$ when B has already been happened? B happened means everything other than B did not happen so B^c ("B compliment"=Anything outside the set B) is dropped from the sample space. So, our new sample space is S=B. (Also notice that B includes a part of A, i.e., $A \cap B$). So the conditional probability of event A given that B has already occurred $P(A|B)$ is given by

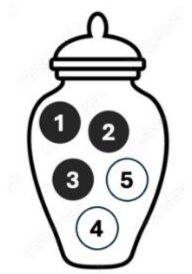


Conditional probability

P(A|B): The probability of A happening given B occurs.
E.g., P(I'm with umbrella | it rains): If we know it rains, what is the probability of Ayşe having her umbrella with her?

P(A∩B): The probability of A and B happening together.
E.g., P(I'm with umbrella ∩ it rains): What is the probability of Ayşe having her umbrella with her and it rains.

Example: A single ball is selected from an urn where balls have numbers and black or white colors as such: S={1b, 2b, 3b, 4w, 5w}. Assuming four outcomes are equally likely, find P[A|B] and P[A|C] for the following events:

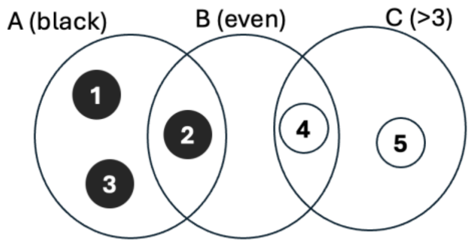


A={1b, 2b, 3b}, "black ball selected".

B={2b, 4w}, "even-numbered ball selected".

C={4w, 5w}, "number of ball is greater than 3".

<u>Solution:</u>



P[A]=3/5, P[B]=2/5, P[C]=2/5.

P[A∩B]=1/5, P[B∩C]=1/5, P[A∩C]=0.

- If selected ball is black, what is the probability that it is "2", i.e., P(B|A) ?

  P(B|A)= P[B ∩ A]/P[A]=(1/5) / (3/5) = 1/3

- If selected ball is black, what is the probability that it is greater than 3, i.e., P(C|A) ?

  P(C|A)= P[C ∩ A]/P[A]=0 / (3/5) = 0

  A & C are **mutually exclusive** events, if A happens, we know for sure C does not happen, and vice versa.

Bayes' Rule

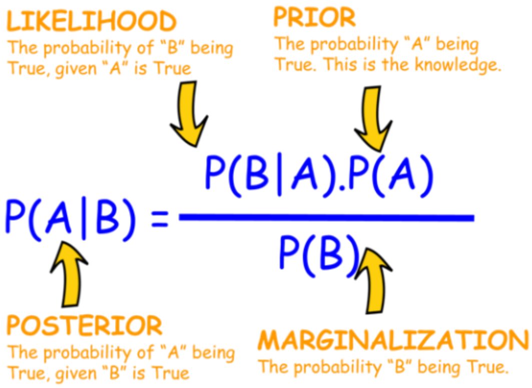
Remember the conditional probability formula:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{or} \quad P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)}$$

where in the last step we used $P(B \cap A) = P(A \cap B)$. So, from the second formula, we get

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

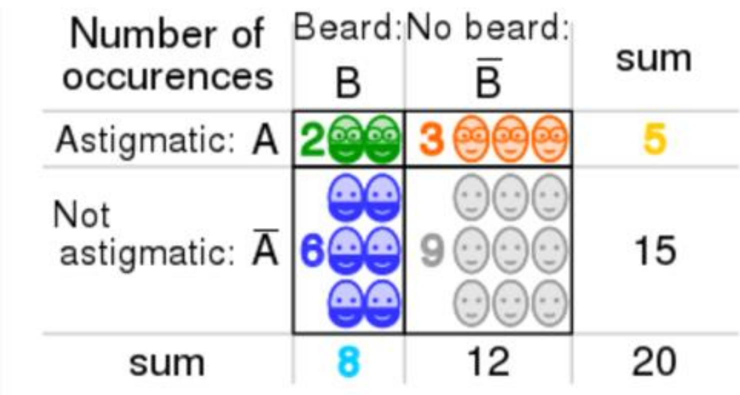
Now substitute this into the numerator in the first formula, we got the Bayes' Rule (or Bayes' Theorem):



Bayes' rule is very useful because it allows us to connect the "prior" (known before) probabilities $P[B_j]$ to the "posterior" (known after) probabilities.

Example:

Let's say we have 20 people, with and without astigmatism and beard. Let's say the question is what's the conditional probability of a person to be astigmatic if that person has beard. For that, we need to know the prior P[A], i.e., the probability of a person (among 20 people) to be astigmatic as well as P[B] and the likelihood P[B|A] (probability of having beard if that person is astigmatic). From the table, P[A]=5/20, P[B]=8/20 and P[B|A]=2/5.



Number of occurences	Beard: B	No beard: $\bar{B}$	sum
Astigmatic: A	2	3	5
Not astigmatic: $\bar{A}$	6	9	15
sum	8	12	20

So, to find P[A|B]

$$P[A|B] = \frac{P[B|A]\,P[A]}{P[B]} = \frac{\frac{2}{5}\cdot\frac{5}{20}}{\frac{8}{20}} = \frac{1}{4}$$

The result P[A|B]=1/4 can be verified from the table.

(Note: If you have the table, of course, you can tell what P[A|B] is right away. But imagine the situation where we don't have the table but we know P[A], P[B] and P[B|A]; we can calculate P[A|B] from the Bayes' rule.).

Independence of Events

Does one event affect the other event?

1. Independent events:

   Draw one ball (event 1), put it back, draw another ball (event 2). The event 1 and even 2 are independent since we put the drawn ball back.

   $P[A\cap B]=P[A]P[B]$

   So, $P[A|B]= P[A\cap B]/P[B] = P[A]P[B] /P[B] = P[A]$. In other words, B happening or not does not affect A, so conditional probability $P[A|B]$ is equal to $P[A]$.

2. Dependent events:

   Draw one ball (event 1), put it away and draw another ball (event 2). These two events are dependent. For example, probability of getting red ball in event 2 depends on what color of ball is drawn in event 1.

   $P[A\cap B]$ does not factorize as $P[A]$ and $P[B]$, so $P[A|B]$ is not necessarily equal to $P[A]$.

   Notes:

- The intersection $P[A\cap B]$ is about co-occurrence of the two events, not related to dependence or independence of events. $P[A\cap B]$ is shown as the intersection area on the Venn diagrams.
- Dependence/independence is about if one event affects the other one, so if they are correlated or not. Dependence/independence is not shown on the Venn diagram.

	$P(A\cap B) \neq 0$ 	$P(A\cap B) = 0$ 
Dependent (Events affect each other)	A: rain in Fatih today B: rain in Zeytinburnu today $P(A\|B)=P(A\cap B)/P(B)$	Single coin flip where A: tail and B: head. Mutually exclusive, so dependent. $P(A\cap B)=0$ $P(A\|B)=0$
Independent (Events don't affect each other)	A: eating rice today B: rain in Seattle today $P(A\cap B)=P(A)P(B)$ $P(A\|B)=P(A)$	**Not possible** (when $P(A\cap B) = 0$, they are mutually independent, they affect each other, so can't be independent).

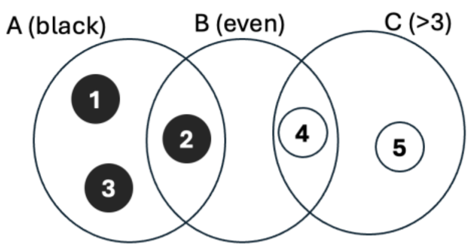
Examples:

(a) Independent events: (When two events don't affect each other.)
A: Choose one number <8.
B: Choose another number >2.

(b) Dependent events: (When two events affect each other.)
Choose one number.
A: That number is <8.
B: That number is >8.
(mutually exclusive, i.e., $P[A \cap B]=0$)

(c) Dependent events: (When two events affect each other.)
Choose one number.
A: That number is <8.
B: That number is >2.

Let's consider the previous question of black and white balls.



Let's check if A and B are independent events. <u>Remember, we draw only one ball.</u>

$P(B|A)= P[B \cap A]/P[A]=(1/5) / (3/5) = 1/3$

If A and B are independent events, $P(B|A)$ should be equal to $P[A]P[B]$.

$P[A]=3/5$, $P[B]=2/5$ -> $P[A]P[B]=6/25$. This is **not** equal to $P(B|A)$ which is $1/3$. So, these two events are dependent!

But how are these two events dependent? A: blackness. B: evenness. Since we draw one ball and that ball will have both blackness and evenness values, it makes these properties/events correlated, i.e., dependent.

Discrete Random Variables

A random variable is a numerical quantity whose value is determined by the outcome of a random process or experiment.

Example: Total number of heads in three coin toss.

Sample space S={HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

The random variable X counts the number of heads in each possible event (outcome).

ζ:	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(\zeta)$:	3	2	2	2	1	1	1	0

X is then a random variable taking on values in the set $S_X = \{0, 1, 2, 3\}$.

Example 3.2 A Betting Game

A player pays \$1.50 to play the following game: A coin is tossed three times and the number of heads X is counted. The player receives \$1 if $X = 2$ and \$8 if $X = 3$, but nothing otherwise. Let Y be the reward to the player. Y is a function of the random variable X and its outcomes can be related back to the sample space of the underlying random experiment as follows:

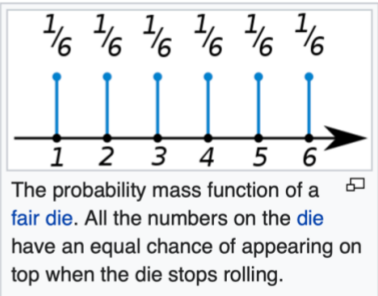
ζ:	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(\zeta)$:	3	2	2	2	1	1	1	0
$Y(\zeta)$:	8	1	1	1	0	0	0	0

Y is then a random variable taking on values in the set $S_Y = \{0, 1, 8\}$.

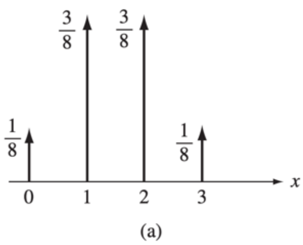
The example above shows that a function of a random variable Y(X) produces another random variable Y.

Probability Mass Function

Probability mass function (pmf) of a discrete random variable *X* gives the probability of each event.
Example 1 (left figure): pmf of a fair die.
Example 2 (right figure): pmf of the experiment of number of heads in three coin toss of fair dice.



The probability mass function of a fair die. All the numbers on the die have an equal chance of appearing on top when the die stops rolling.



(a)

Note that the probabilities in a pmf always sums up to 1, i.e., $\sum_i p_i = 1$.

Expected Value (Mean) of Discrete Random Variable

The expected value or mean of a discrete random variable X is defined by

$$\sum_i x_i p_i = \mu$$

In the example of the experiment of number of heads in three coin toss, x_i's take (0,1,2,3) and p_i's take (1/8, 3/8, 3/8, 1/8), respectively. The mean in this example becomes

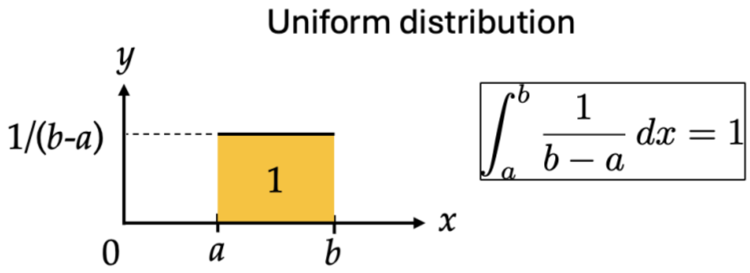
$$0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1.5$$

Notice that the mean doesn't have to be a number that exists in the sample space, e.g., the mean is 1.5 here but there's no number on the die for 1.5.

Probability Density Function of a continuous random variable:

In the continuous case, instead of "probability mass function", we have "probability density function" (pdf).

For example, the pdf of a continuous uniform random variable between a and b is given by f(x)=1/(b-a) (We may show pdf as f(x), p(x) or pdf(x)).



Let's choose the parameters of the uniform distribution f(x)=1/(b-a) as a=3 and b=7 for the following example. Since there are infinitely many numbers within the interval [0,1], questions such as "What's the probability of getting 4.2384424." are meaningless. Rather, we ask questions like "what's the probability of getting a number between 4.38 and 4.39?" That probability P[4.38≤x≤4.39] is given by the integral of the pdf f(x)=1/(b-a) with the appropriate limits

$$\mathrm{P}[4.38 \le \mathrm{x} \le 4.39] = \int_{x_1}^{x_2} f(x)\,dx = \int_{4.38}^{4.39} \frac{1}{7-3}\,dx = 0.0025.$$

To generalize, then, the probability is given as the integral of pdf f(x)

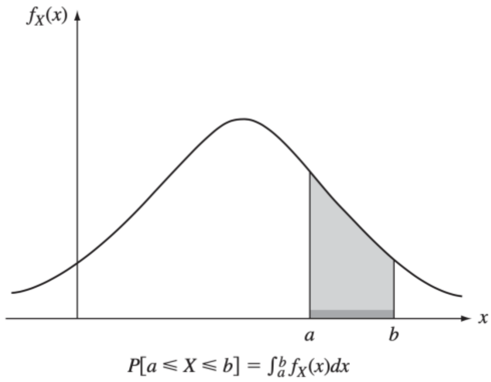
$$\mathrm{P}[x_1 \le \mathrm{x} \le x_2] = \int_{x_1}^{x_2} f(x)\,dx$$

Notice that f(x) is "probability density"; so its value at any point like f(x=2.33) doesn't give us the probability. What we can interpret as probability is the "probability mass". To find "probability mass", you'd have to integrate the probability density function (pdf). This is similar to the integral formula that connects mass m, density ρ and volume V in physics (now you understand why the term "probability mass").

$$m = \int_V \rho(x)\,dx$$

Probability $\mathrm{P}[\mathrm{a} \le \mathrm{x} \le \mathrm{b}]$ can be interpreted as the area under the pdf f(x), which is the "probability mass" (shaded area below). If we integrate over the whole space where pdf f(x) is defined, then we get 1 (=total probability).

mass



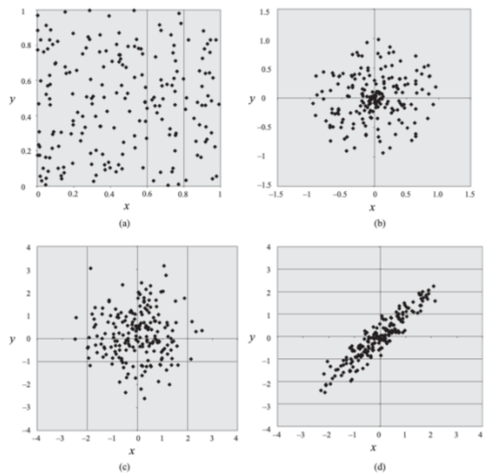
For a pdf extending between $-\infty$ and ∞, we have the normalization condition: $\mathrm{P}[-\infty \leq \mathrm{x} \leq \infty] = \int_{-\infty}^{\infty} f(x)\, dx = 1.$

Two random variables

In this section we will deal with two dimensional random numbers, i.e., for each outcome from the sample space S, there is a pair of random numbers (X,Y). The simplest example is throwing two dice, each given a random number.

Going back to one dimension, we can show the random numbers sampled from different distribution functions on scattergram.

If we have two random variables X and Y, we can show the random number set (X,Y) in 2D scattergrams. Examples:



Interpretation of the scattergrams:

Figure (a) shows random number pairs (X,Y) where X and Y seems to be independent of each other.

Figure (b) and (c) suggest that the random number pairs (X,Y) are sampled from a 2D distribution that is symmetric in the polar angle direction (something like a symmetric hill). (dependent)

Figure (d) shows a distribution where X and Y are correlated. For example, (-2,3) is not likely here. (dependent)

Pairs of Discrete Random Variables

If we're dealing with discrete (X,Y) pairs, then the **joint probability mass function** (jpmf) specifies the probabilities {X=x}∩{Y=y}.

$$p_{X,Y}(x,y) = P[X = x, Y = y]$$

Normalization condition (total probability is 1) is given by

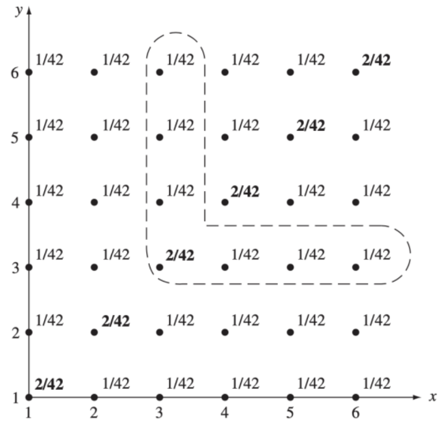
$$\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} p_{X,Y}(x_j, y_k) = 1.$$

For example, if X and Y are coin flip random numbers, there are four terms:
$p_{X,Y}(H,H)$, $p_{X,Y}(H,T)$, $p_{X,Y}(T,H)$ and $p_{X,Y}(T,T)$.

Note: P(X,Y)=P(X∩Y).

Example:

Random experiment of tossing two "loaded" dice and noting the pair of numbers (X,Y) facing up. The joint pmf $p_{X,Y}(j,k)$ for j=1,…,6 and k=1,…,6 is given by the two dimensional table shown in the figure below. Each entry in the table contains the value of the joint pmf $p_{X,Y}(j,k)$. (Note that each element is not 1/36; that's for fair dice.) Also, these two dice are affected from each other *magnetically*, so they tend to get the same outcome like (1,1), (2,2) etc. with slightly greater probability (2/42).



What is the probability of getting the minimum of both dice as 3, i.e., P[min(X,Y)=3]?

$$\begin{aligned} P[\min(X,Y) = 3] &= p_{X,Y}(6,3) + p_{X,Y}(5,3) + p_{X,Y}(4,3) \\ &\quad + p_{X,Y}(3,3) + p_{X,Y}(3,4) + p_{X,Y}(3,5) + p_{X,Y}(3,6) \\ &= 6\left(\frac{1}{42}\right) + \frac{2}{42} = \frac{8}{42}. \end{aligned}$$

Marginal Probability Mass Function

Consider now two loaded dice with the pmf given in the table below. What is the joint probability of the **die X** showing 3 and **die Y** showing anything, i.e., P(X=3,Y)=P(X=3∩Y)?

This corresponds to the pairs (3,1), (3,2), (3,3), (3,4), (3,5) and (3,6). The probability will be summation of the individual probabilities for these cases, which is 3/200+3/200+3/200+1/25+1/25+7/200=**4/25**.

Let's remake the table now by writing the summed probabilities on the margins. Each number in the *margin* gives the summation of all six number in that row or column.

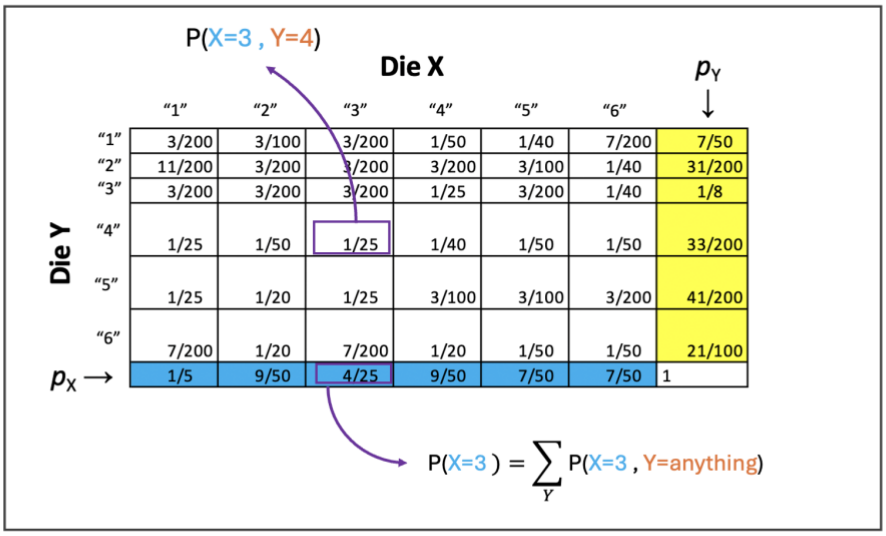
Die X

Die Y	"1"	"2"	"3"	"4"	"5"	"6"	p_Y ↓
"1"	3/200	3/100	3/200	1/50	1/40	7/200	7/50
"2"	11/200	3/200	3/200	3/200	3/100	1/40	31/200
"3"	3/200	3/200	3/200	1/25	3/200	1/40	1/8
"4"	1/25	1/50	1/25	1/40	1/50	1/50	33/200
"5"	1/25	1/20	1/25	3/100	3/100	3/200	41/200
"6"	7/200	1/20	7/200	1/20	1/50	1/50	21/100
p_X →	1/5	9/50	4/25	9/50	7/50	7/50	1

Notice that 4/25 we calculated above is in the blue margin; so calculated it by summing all individual probabilities for die Y when die X=3. 4/25 gives us X=3 and Y is anything. You can see why this probability is called "marginal" because it's written on the margin of the table.

Let's recap:

- Each white box shows the joint probability P(X,Y).
- The margins shows the marginals P(X) and P(Y):
  - P(X=a): probability of die X=a regardless of what Y is when two dice are thrown.
  - P(Y=a): probability of die Y=a regardless of what X is when two dice are thrown.
- Conditional probability cannot be easily read off from the table (it'll come later).



Another example, according to the table, what is the probability of getting Y=6 and X any number? It is 21/100, given in the yellow margin. Here p_X and p_Y show the **marginal probability mass functions**.

We can formally define what we did above. For a joint probability mass function (jpmf) $p_{X,Y}(x_i, y_j)$, the **marginal probability mass function** is found by summing over the one of the random variable:

$$p_X(\mathrm{x_i}) = \sum_j p_{X,Y}(\mathrm{x_i}, y_j) \quad \text{(Sum over all } y_j\text{)}$$

$$p_Y(\mathrm{y}_j) = \sum_i p_{X,Y}(\mathrm{x_i}, y_j) \quad \text{(Sum over all } \mathrm{x_i}\text{)}$$

So, to repeat question above: "What is the joint probability of the **die X** showing 3 and **die Y** showing anything?" Let's do it this time formally:

$$p_X(3) = \sum_j p_{X,Y}(3, y_j) = p_{X,Y}(3,1) + p_{X,Y}(3,2) + p_{X,Y}(3,3) + p_{X,Y}(3,4) + p_{X,Y}(3,5) + p_{X,Y}(3,6)$$

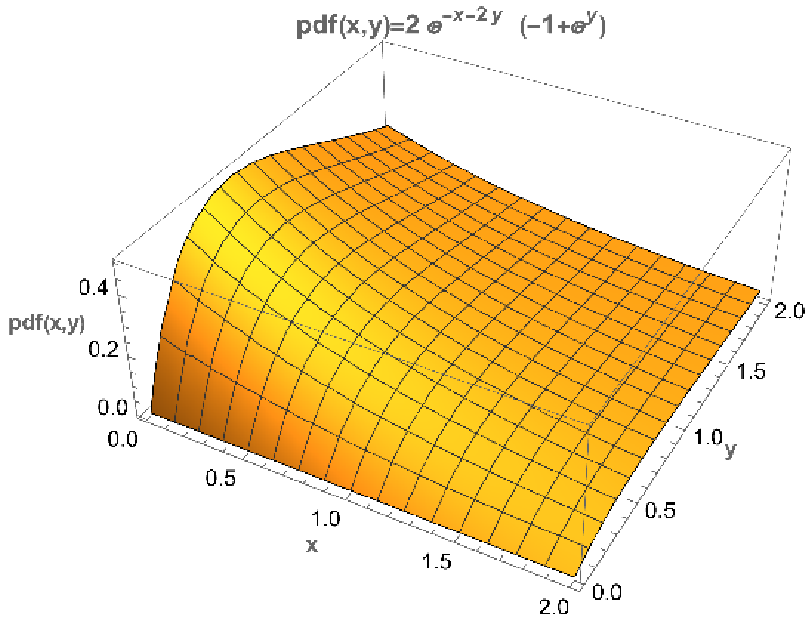
=3/200+3/200+3/200+1/25+1/25+7/200=4/25.

The Joint PDF of two continuous random variables

The joint PDF $f_{X,Y}$ should be normalized:

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x', y')\, dx'\, dy'.$$

Example:



Conditional Probability

Conditional probability $p_X(x|y)$ gives us the conditional pdf of X when we know the event Y=y happens. (The vertical symbol "|" appears in conditional probability distributions.) This is different from the joint pdf $p_{X,Y}(x, y)$; joint pdf gives us the probability of getting X=x and Y=y without one depending other with any condition.

In the discrete case, the joint pmf $p_{X,Y}(x_k, y_j)$, the conditional pmf $p_Y(y_j|x_k)$ (or $p_X(x_i|y_j)$) and the marginal pmf $p_X(x_k)$ (or $p_Y(y_j)$) are related to each other with these conditional pmf rules:

$$p_{X,Y}(x_k, y_j) = p_Y(y_j|x_k)p_X(x_k)$$

$$p_{X,Y}(x_k, y_j) = p_X(x_k|y_j)p_Y(y_j)$$

Remember, we can read off from the table the joint probabilities $p_{X,Y}(x_k, y_j)$ (white boxes) and marginal probabilities $p_X(x_k)$ or $p_Y(y_j)$ (blue and yellow boxes in the margins), but not conditional probabilities. We can <u>calculate</u> the conditional probabilities by using the formula above.

If you equate the right-hand sides of the two equations above, you get Bayes' theorem

$$p_X(x_k|y_j) = \frac{p_Y(y_j|x_k)p_X(x_k)}{p_Y(y_j)}$$

If we want to find the marginal pmfs, we sum $p_{X,Y}(x_k, y_j)$ over x_k or y_j.

$$p_Y(y_j) = \sum_k p_{X,Y}(x_k, y_j) = \sum_k p_Y(y_j|x_k)p_X(x_k)$$

$$p_X(x_k) = \sum_j p_{X,Y}(x_k, y_j) = \sum_j p_X(x_k|y_j)p_Y(y_j)$$

Notice that $p_X(x_k|y_j)$ is the pmf of the random variable X, and $p_Y(y_j|x_k)$ is the pmf of the random variable Y. Realize how the subscripts and the first variables match in $p_X(x_k|y_j)$ and $p_Y(y_j|x_k)$.

In the continuous case, the equations above are the same except one must replace discrete sums with integrals.

Example: Loaded dice

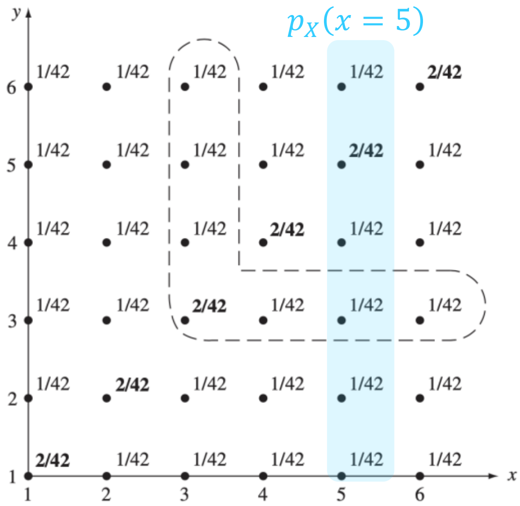
We have the pmf of two loaded dice shown in the table below. Remember, these two dice are affected from each other *magnetically* (so they are dependent) and they tend to show the same number up with (probability 2/42). Also remember, the table shows joint pmfs, not conditional pmfs! All joint pmfs in the table (36 values) add up to 1.

Find $p_Y(y|5)$, which means "the probability of die Y if we already got die X=5".

$p_Y(y|5)$ can be rewritten in terms of the joint pdf $p_{X,Y}(x,y)$ by using the formula introduced before:

$$p_Y(y|x=5) = \frac{p_{X,Y}(x=5,y)}{p_X(x=5)}$$

The marginal $p_X(x=5)$ can be calculated from the pmf table on the right:



$p_X(x=5) = 1/42 + 2/42 + 1/42 + 1/42 + 1/42 + 1/42 = 7/42 = 1/6$.

So, the conditional pmf is

$p_Y(y|x=5) = \frac{p_{X,Y}(x=5,y)}{p_X(x=5)} = \frac{p_{X,Y}(x=5,y)}{1/6}$.

For example, $p_Y(y=3|x=5) = \frac{p_{X,Y}(x=5,y=3)}{1/6} = \frac{1/42}{1/6} = \frac{1}{7}$.

Remember, $p_Y(y|x=5)$ is quantifying "what face *die Y* shows with what probability if *die X* is known to show face 5 up".

To list all possibilities:

$p_Y(y=1|x=5) = p_Y(y=2|x=5) = p_Y(y=3|x=5) = p_Y(y=4|x=5) = p_Y(y=6|x=5) = 1/7$.

$p_Y(y=5|x=5) = 2/7$.

So, all possibilities for $p_Y(y|x=5)$ is adding to 1.

Let's compare the joint and conditional pmfs for better understanding:

<u>Joint pmf (has "," or "∩" sign):</u>

$p_{X,Y}(y=1, x=5) = 1/42$ (found from the joint pmf table given above)

It is the probability of getting X=5 AND Y=1 out of all 36 cases when we roll both die. Sample space has 36 possibilities (X,Y)=(1,1), (1,2) … (6,5), (6,6), and the probability for each case is given in the joint pmf table given above.

Conditional pmf (has "|" sign):

$p_Y(y=1|x=5)=1/7$ (calculated above by using the formula $p_Y(y|x=5)=\frac{p_{X,Y}(x=5,y)}{p_X(x=5)}$)

The probability of getting die Y=1 when we *know* die X=5. When we know X=5, the sample space gets smaller, which becomes (X,Y)=(5,1), (5,2), (5,3), (5,4), (5,5) and (5,6); so six possibilities in total if X=5.

PAIRS OF JOINTLY CONTINUOUS VARIABLES

(*Here we're going to be specific and use Gaussian random variables since they appear frequently in science and engineering and also the math of Gaussian distributions is easy and instructive.*)

Consider the amount of rain that falls in Fatih, İstanbul and Seattle, Washington are random variables X and Y, and assume the amount of rain is a Gaussian random variable with mean and standard deviation (for example, μ=72 mm and σ=20 mm). These two Gaussian distributions will be independent since these two towns are thousands of miles apart and we wouldn't expect correlation between the amount of rainfall. So, the joint pdf factors, meaning it can be written as multiplication of two independent Gaussians, $f_{X,Y}(x,y)=f_X(x)f_Y(y)$.

[*Caution: That they are independent does not mean P(X,Y)=P(X∩Y)=0!*

P(X,Y)=0 would mean two events (rain in Fatih and rain in Seattle) are mutually exlusive, i.e., it rains in Fatih when it doesn't rain in Seattle and vice versa, where such a mutually exclusive scenario is not realistic.]

Now consider two adjacent towns in İstanbul, Zeytinburnu and Fatih. Since they're next to each other, the rainfall on one will be correlated by the rainfall in other. How correlated? There are a few possibilities. Maybe they're positively correlated, which means when it rains more in Zeytinburnu, and it rains relatively more in Fatih. The rainfall amounts in two cities may be negatively correlated, which means when it rains more in Zeytinburnu, and it rains relatively less in Fatih. When there is positive or negative correlation, then the distributions of both towns are not independent, so we cannot write $f_{X,Y}(x,y)$ as $f_X(x)f_Y(y)$, i.e., joint pdf does not factor).

Below, we provide the expression for the joint pdf $f_{X,Y}(x,y)$ of two Gaussian distributions. (Here we're interested in Gaussian joint pdfs because they appear frequent in engineering problems, but keep in mind that in general there may be joint distribution in any form)

The random variables X and Y are said to be **jointly Gaussian** if their joint pdf has the form

$$f_{X,Y}(x,y)=\frac{\exp\left\{\frac{-1}{2(1-\rho_{X,Y}^2)}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2-2\rho_{X,Y}\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{y-m_2}{\sigma_2}\right)+\left(\frac{y-m_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{X,Y}^2}}$$

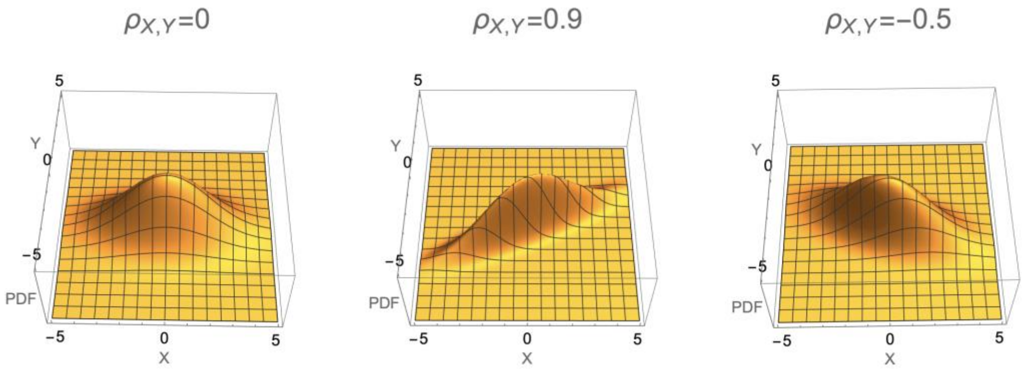
Here $\rho_{X,Y}$ is the correlation coefficient that quantifies the correlation between random variables X and Y. Notice that here $f_{X,Y}(x,y)$ is not equal to multiplication of two *independent* Gaussian distributions when $\rho_{X,Y}\neq 0$, i.e., joint pdf does not factor when events are correlated (dependent).

The random variables X and Y are said to be **jointly Gaussian** if their joint pdf has the form

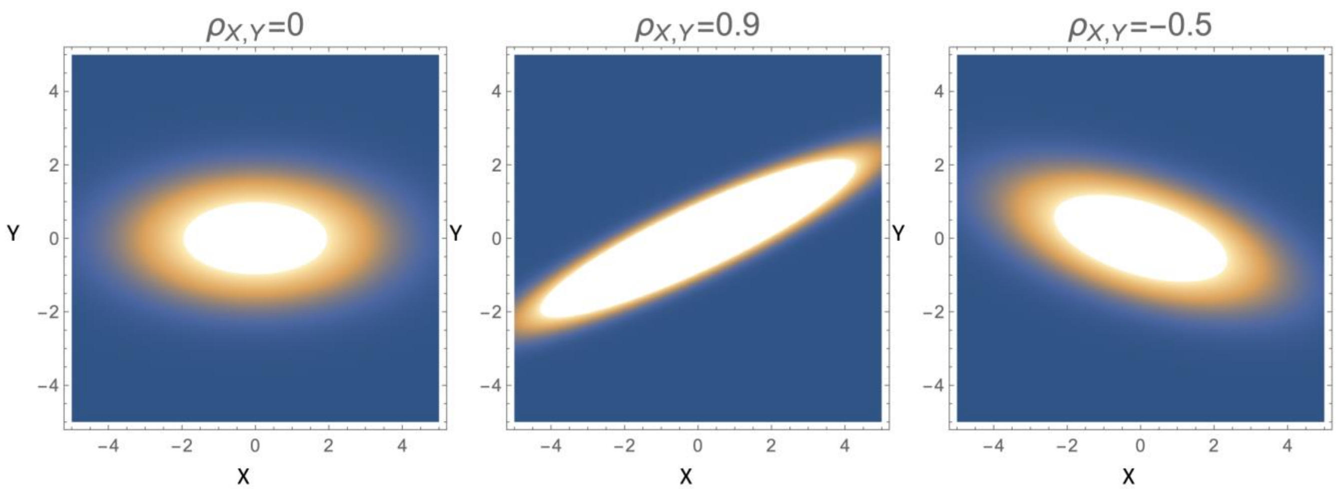
$$f_{X,Y}(x, y) = \frac{\exp\left\{\frac{-1}{2(1-\rho_{X,Y}^2)}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{y-m_2}{\sigma_2}\right) + \left(\frac{y-m_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{X,Y}^2}}$$

If $\rho_{X,Y} = 0$, then the cross term $\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{y-m_2}{\sigma_2}\right)$ disappears and $f_{X,Y}(x,y)$ becomes multiplication of two independent Gaussian distributions, i.e., $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

What does it mean for X and Y to be correlated? It means Y is increasing or decreasing while X is increasing. Below we plot the joint Gaussian pdf $f_{X,Y}(x,y)$ for different values of $\rho_{X,Y}$ for the means $m_1 = m_2 = 0$ (namely, the mean of the 2D distribution is at (0,0) point), and $\sigma_1 = 2$ and $\sigma_2 = 1$ (Here these parameters are chosen arbitrarily just to give you an example).



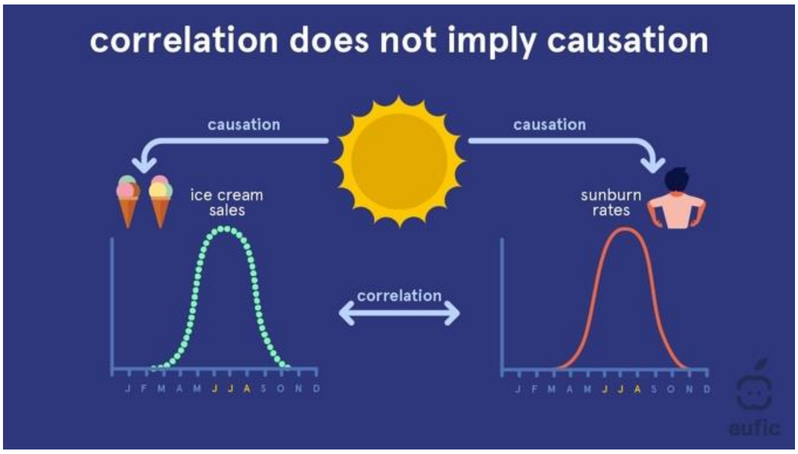
Or we can show it as a density plot:



Notice that we chose different standard deviations for these two distributions, that's why we get ellipse shape rather than a perfectly round disk.

Finally, here's a quick clarification about the concept "correlation":

In the example about rainfall in two cities we gave above, correlation doesn't mean the rainfall in one town affects the rainfall in the other town.



Learn the famous adage, "correlation does not imply causation".

Correlation basically means co-occurring.

There may be several factors that are causes of rainfall amounts in both towns.