Probability: The Study of Randomness
Randomness and Probability Models

IPS Chapters 4.1 and 4.2

Objectives (IPS Chapters 4.1 and 4.2)

Randomness and Probability models

- Probability and Randomness

- Sample spaces

- Probability rules

- Assigning probabilities: finite number of outcomes

- Assigning probabilities: equally likely outcomes

- Independence and multiplication rule

Randomness and probability

A phenomenon is **random** if individual outcomes are uncertain, but there is nonetheless a regular distribution of outcomes in a large number of repetitions.
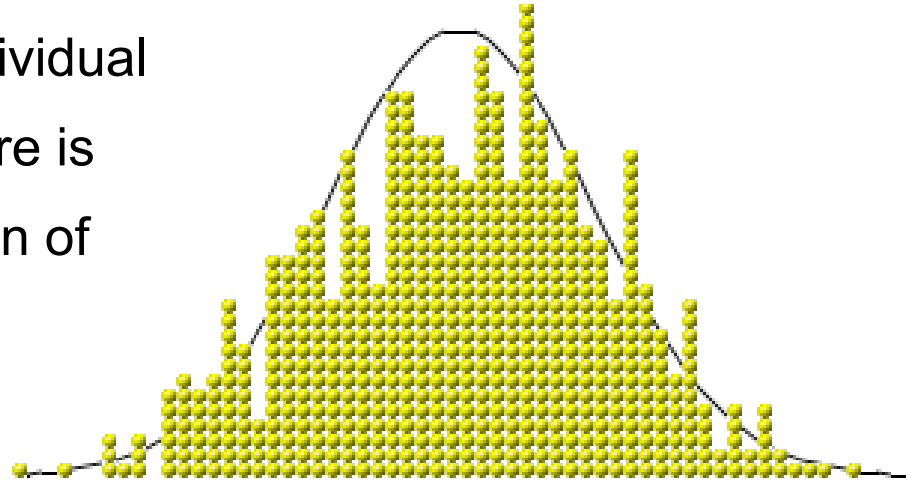
The **probability** of any outcome of a random phenomenon can be defined as the proportion of times the outcome would occur in a very long series of repetitions.

Coin toss

The result of any single coin toss is random. But the result over many tosses is predictable, as long as the trials are **independent** (i.e., the outcome of a new coin flip is not influenced by the result of the previous flip).
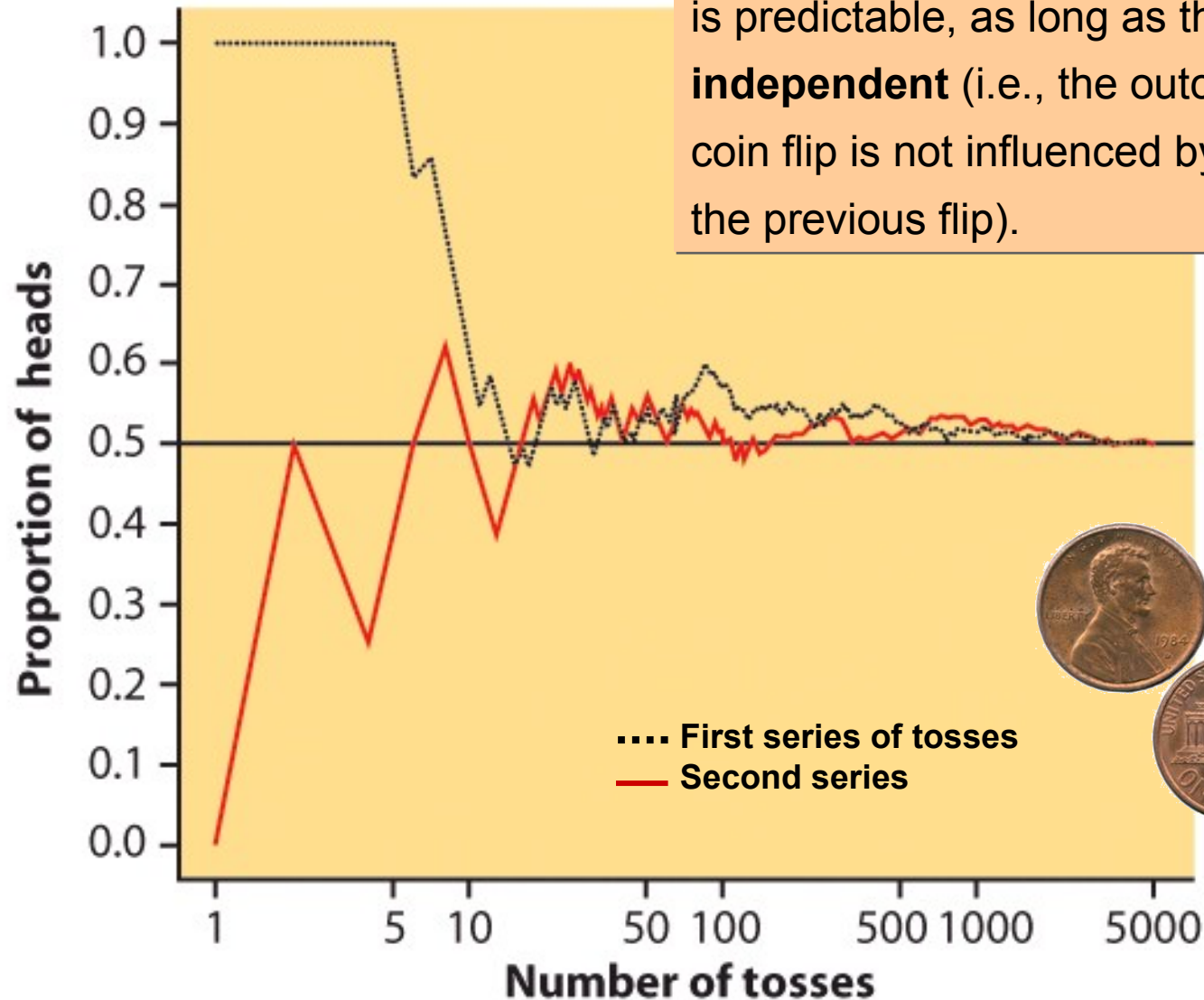
The probability of heads is 0.5 = the proportion of times you get heads in many repeated trials.

Proportion of heads (y-axis): 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.0

Number of tosses (x-axis): 1, 5, 10, 50, 100, 500, 1000, 5000

•••• **First series of tosses**
—— **Second series**

Two events are **independent** if the probability that one event occurs on any given trial of an experiment is not affected or changed by the occurrence of the other event.

When are trials not independent?

Imagine that these coins were spread out so that half were heads up and half were tails up. Close your eyes and pick one. The probability of it being heads is 0.5. However, if you don't put it back in the pile, the probability of picking up another coin that is heads up is now less than 0.5.

The trials are independent only when you put the coin back each time. It is called **sampling with replacement.**

Probability models

Probability models describe, mathematically, the outcome of random processes. They consist of two parts:

1) **S = Sample Space**: This is a set, or list, of <u>all</u> possible outcomes of a random process. An **event** is a subset of the sample space.

2) A **probability** for each possible event in the sample space S.

Example: Probability Model for a Coin Toss:
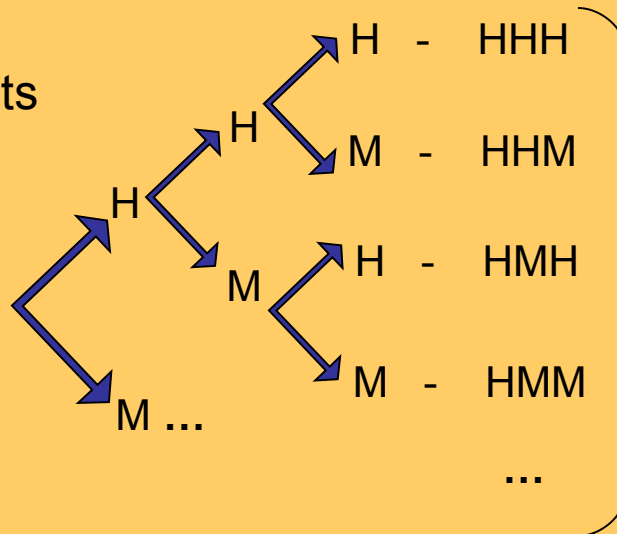
S = {Head, Tail}

Probability of heads = 0.5

Probability of tails   = 0.5

Sample spaces

It's the question that determines the sample space.

A. A basketball player shoots three free throws. What are the possible sequences of hits (H) and misses (M)?

H → H → H - HHH

H → H → M - HHM

H → M → H - HMH

H → M → M - HMM

H ...

M ...

$S = \{$ HHH, HHM, HMH, HMM, MHH, MHM, MMH, MMM $\}$

Note: 8 elements, 2^3

B. A basketball player shoots three free throws. What is the number of baskets made?

$S = \{ 0, 1, 2, 3 \}$

C. A nutrition researcher feeds a new diet to a young male white rat. What are the possible outcomes of weight gain (in grams)?

$S = [0, \infty[=$ (all numbers ≥ 0)

Probability rules

1) Probabilities range from 0 (*no chance of the event*) to 1 (*the event has to happen*).

For any event A, $0 \leqq P(A) \leqq 1$

Probability of getting a Head = 0.5
We write this as: P(Head) = 0.5

P(neither Head nor Tail) = 0
P(getting either a Head or a Tail) = 1

2) Because some outcome must occur on every trial, the sum of the probabilities for all possible outcomes (the sample space) must be exactly 1.

P(sample space) = 1

Coin toss: S = {Head, Tail}

P(head) + P(tail) = 0.5 + 0.5 =1
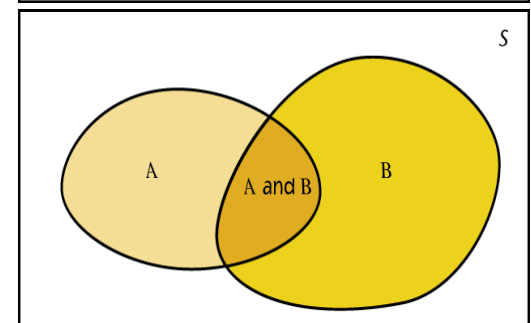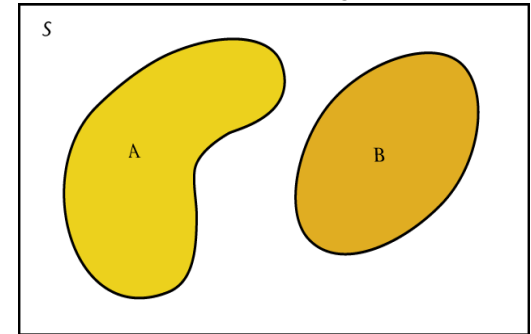→ P(sample space) = 1

Probability rules (*cont,d*)

3) Two events A and B are **disjoint** if they have no outcomes in common and can never happen together**.** The probability that A <u>or</u> B occurs is then the sum of their individual probabilities.

P(A or B) = "P(A U B)" = P(A) + P(B)
This is the **addition rule for disjoint events.**

A and B <u>not</u> disjoint

Example: If you flip two coins, and the first flip does not affect the second flip:

S = {HH, HT, TH, TT}. The probability of each of these events is 1/4, or 0.25.

The probability that you obtain "only heads <u>or</u> only tails" is:

$$P(\text{HH or TT}) = P(\text{HH}) + P(\text{TT}) = 0.25 + 0.25 = 0.50$$

Probability rules (*cont,d*)

Coin Toss Example:
S = {Head, Tail}
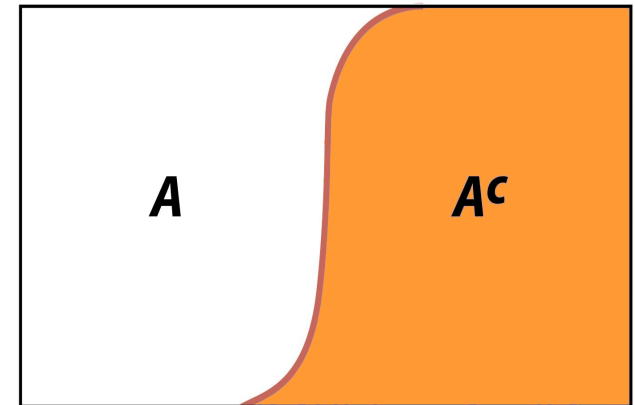Probability of heads = 0.5
Probability of tails = 0.5

4) The complement of any event A is the event that A does not occur, written as A^c.

The complement rule states that the probability of an event not occurring is 1 minus the probability that is does occur.

$P(\text{not A}) = P(A^c) = 1 - P(A)$

Tail^c = not Tail = Head

$P(\text{Tail}^c) = 1 - P(\text{Head}) = 0.5$

Venn diagram:
Sample space made up of an event A and its complementary A^c, i.e., everything that is not A.

Probabilities: finite number of outcomes

Finite sample spaces deal with **discrete data** — data that can only take on a limited number of values. These values are often integers or whole numbers.

Throwing a die: $S = \{ \boxed{•}, \boxed{••}, \boxed{•••}, \boxed{••••}, \boxed{•••••}, \boxed{••••••} \}$

$S = \{1, 2, 3, 4, 5, 6\}$

The individual outcomes of a random phenomenon are always disjoint.

➔ The probability of any event is the sum of the probabilities of the outcomes making up the event (addition rule).

M&M candies

If you draw an M&M candy at random from a bag, the candy will have one of six colors. The probability of drawing each color depends on the proportions manufactured, as described here:

Color	Brown	Red	Yellow	Green	Orange	Blue
Probability	0.3	0.2	0.2	0.1	0.1	?

What is the probability that an M&M chosen at random is blue?

S = {brown, red, yellow, green, orange, blue}

$P(S) = P(\text{brown}) + P(\text{red}) + P(\text{yellow}) + P(\text{green}) + P(\text{orange}) + P(\text{blue}) = 1$

$P(\text{blue}) = 1 - [P(\text{brown}) + P(\text{red}) + P(\text{yellow}) + P(\text{green}) + P(\text{orange})]$

$= 1 - [0.3 + 0.2 + 0.2 + 0.1 + 0.1] = 0.1$

What is the probability that a random M&M is either red, yellow, or orange?

$P(\text{red or yellow or orange}) = P(\text{red}) + P(\text{yellow}) + P(\text{orange})$

$= 0.2 + 0.2 + 0.1 = 0.5$

Probabilities: equally likely outcomes

We can assign probabilities either:

◻ **empirically** → from our knowledge of numerous similar past events

  ◻ Mendel discovered the probabilities of inheritance of a given trait from experiments on peas without knowing about genes or DNA.

◻ **or theoretically** → from our understanding of the phenomenon and symmetries in the problem

  ◻ A 6-sided fair die: each side has the same chance of turning up
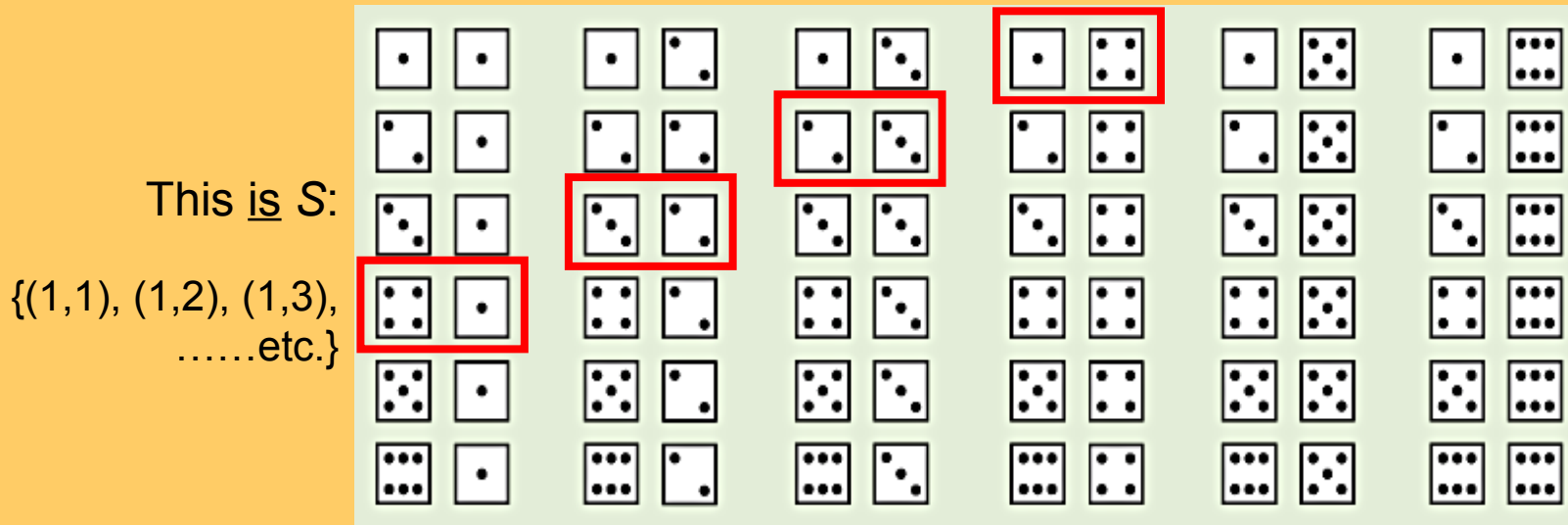 ◻ Genetic laws of inheritance based on meiosis process

If a random phenomenon has k equally likely possible outcomes, then each individual outcome has probability $1/k$.

And, for any event A:
$$P(\text{A}) = \frac{\text{count of outcomes in A}}{\text{count of outcomes in S}}$$

Dice

You toss two dice. What is the probability of the outcomes summing to 5?

This is <u>S</u>:

{(1,1), (1,2), (1,3), ……etc.}

There are 36 possible outcomes in *S*, all equally likely (given fair dice).

Thus, the probability of any one of them is 1/36.

P(the roll of two dice sums to 5) =

P(1,4) + *P*(2,3) + *P*(3,2) + *P*(4,1) = 4 / 36 = 0.111

The gambling industry relies on probability distributions to calculate the odds of winning. The rewards are then fixed precisely so that, **on average,** players lose and the house wins.

The industry is very tough on so called "cheaters" because their probability to win exceeds that of the house. Remember that it is a business, and therefore it has to be profitable.

Probability rules (*cont,d*)

Coin Toss Example:
S = {Head, Tail}
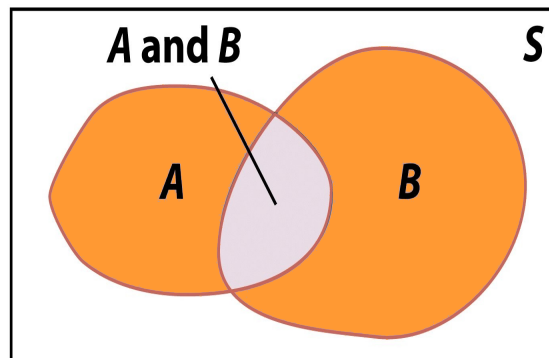Probability of heads = 0.5
Probability of tails = 0.5

5) Two events A and B are independent if knowing that one occurs does not change the probability that the other occurs.

If A and B are independent, P(A and B) = P(A)P(B)

This is the **multiplication rule for independent events**.

Two consecutive coin tosses:
P(first Tail and second Tail) = P(first Tail) * P(second Tail) = 0.5 * 0.5 = 0.25

Venn diagram:

Event A and event B. The intersection represents the event {A and B} and outcomes common to both A and B.

A couple wants three children. What are the arrangements of boys (B) and girls (G)?

Genetics tell us that the probability that a baby is a boy or a girl is the same, 0.5.

Sample space: {BBB, BBG, BGB, GBB, GGB, GBG, BGG, GGG}

➜ All eight outcomes in the sample space are **equally likely.**
  The probability of each is thus 1/8.

➜ Each birth is independent of the next, so we can use the **multiplication rule**.

  Example: $P(BBB) = P(B)* P(B)* P(B) = (1/2)*(1/2)*(1/2) = 1/8$

A couple wants three children. What are the numbers of girls (X) they could have?

The same genetic laws apply. We can use the probabilities above and the **addition rule for disjoint events** to calculate the probabilities for X.

  Sample space: {0, 1, 2, 3}
  ➜ $P(X = 0) = P(BBB) = 1/8$
  ➜ $P(X = 1) = P(BBG$ or BGB or $GBB) = P(BBG) + P(BGB) + P(GBB) = 3/8$
  …

Value of X	0	1	2	3
Probability	1/8	3/8	3/8	1/8

Probability: The Study of Randomness
Random Variables

IPS Chapters 4.3 and 4.4

Objectives (IPS Chapters 4.3 and 4.4)

Random variables

- Discrete random variables

- Continuous random variables

- Normal probability distributions

- Mean of a random variable

- Law of large numbers

- Variance of a random variable

- Rules for means and variances

Discrete random variables

A **random variable** is a variable whose value is a numerical outcome of a random phenomenon.

A basketball player shoots three free throws. We define the random variable *X* as the number of baskets successfully made.

A **discrete random variable** *X* has a <u>finite</u> number of possible values.

A basketball player shoots three free throws. The number of baskets successfully made is a discrete random variable (*X*). *X* can only take the values 0, 1, 2, or 3.
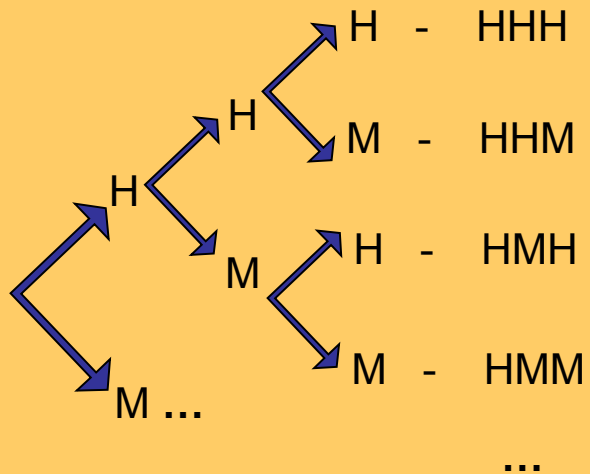
The probability distribution of a random variable X lists the values and their probabilities:

Value of X	x_1	x_2	x_3	$\cdots$	x_k
Probability	p_1	p_2	p_3	$\cdots$	p_k

The probabilities p_i must add up to 1.

A basketball player shoots three free throws. The random variable X is the number of baskets successfully made.

```
            H  -  HHH
      H
            M  -  HHM
  H
            H  -  HMH
      M
            M  -  HMM
  M ...
          ...
```

Value of X	0	1	2	3
Probability	1/8	3/8	3/8	1/8

		HMM	HHM	
		MHM	HMH	
MMM	MMH	MHH	HHH

The probability of any event is the sum of the probabilities p_i of the values of X that make up the event.

A basketball player shoots three free throws. The random variable X is the number of baskets successfully made.

What is the probability that the player successfully makes at least two baskets ("at least two" means "two or more")?

Value of X	0	1	2	3
Probability	1/8	3/8	3/8	1/8

```
                              HMM    HHM
                              MHM    HMH
              MMM      MMH     MHH    HHH
```

$P(X{\geq}2) = P(X=2) + P(X=3) = 3/8 + 1/8 = 1/2$

What is the probability that the player successfully makes fewer than three baskets?

$P(X<3) = P(X=0) + P(X=1) + P(X=2) = 1/8 + 3/8 + 3/8 = 7/8$ or

$P(X<3) = 1 – P(X=3) = 1 – 1/8 = 7/8$

Continuous random variables

A **continuous random variable** *X* takes all values in an **interval**.

Example: There is an infinity of numbers between 0 and 1 (e.g., 0.001, 0.4, 0.0063876).

How do we assign probabilities to events in an infinite sample space?

❑ We use **density curves** and compute probabilities for **intervals**.

❑ The probability of any event is the area under the density curve for the values of *X* that make up the event.

Area = 0.4

Height = 1

X 0   0.3   0.7   1
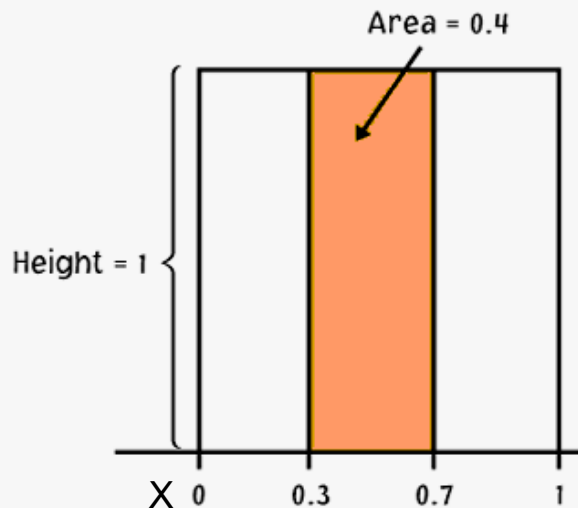
This is a uniform density curve for the variable *X*.

The probability that *X* falls between 0.3 and 0.7 is the area under the density curve for that interval:

$P(0.3 \leq X \leq 0.7) = (0.7 - 0.3)*1 = 0.4$

Intervals

The probability of a single event is meaningless for a continuous random variable. Only intervals can have a non-zero probability, represented by the area under the density curve for that interval.
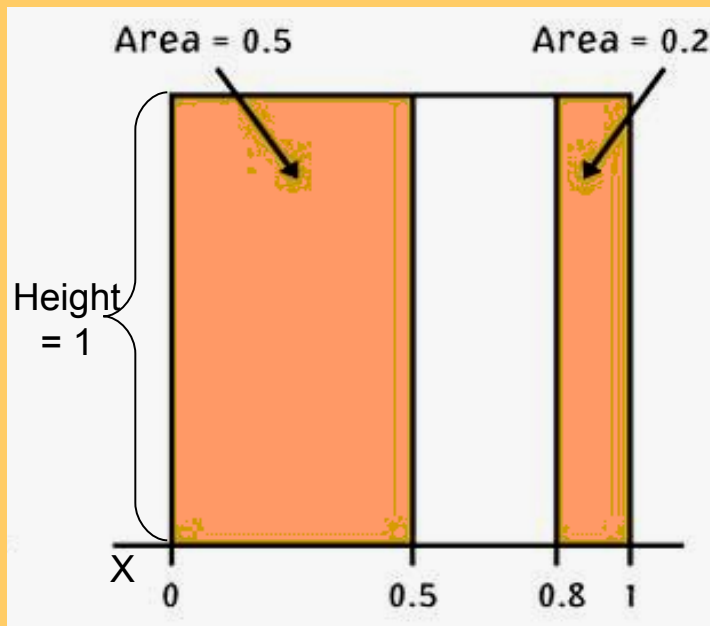
The probability of a single event is zero:

$$P(X=1) = (1 - 1)*1 = 0$$

The probability of an interval is the same whether boundary values are included or excluded:

$$P(0 \leq X \leq 0.5) = (0.5 - 0)*1 = 0.5$$
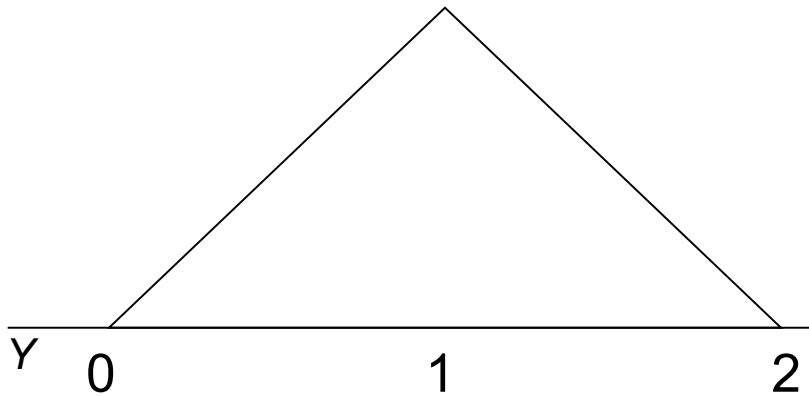$$P(0 < X < 0.5) = (0.5 - 0)*1 = 0.5$$
$$P(0 \leq X < 0.5) = (0.5 - 0)*1 = 0.5$$

$$P(X < 0.5 \text{ or } X > 0.8) = P(X < 0.5) + P(X > 0.8) = 1 - P(0.5 < X < 0.8) = 0.7$$

We generate two random numbers between 0 and 1 and take Y to be their sum. Y can take any value between 0 and 2. The density curve for Y is:
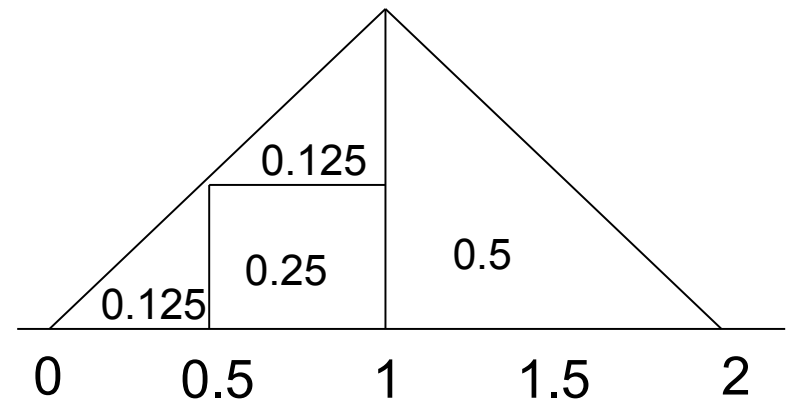

Height = 1. We know this because the base = 2, and the area under the curve has to equal 1 by definition.

The area of a triangle is ½ (base*height).

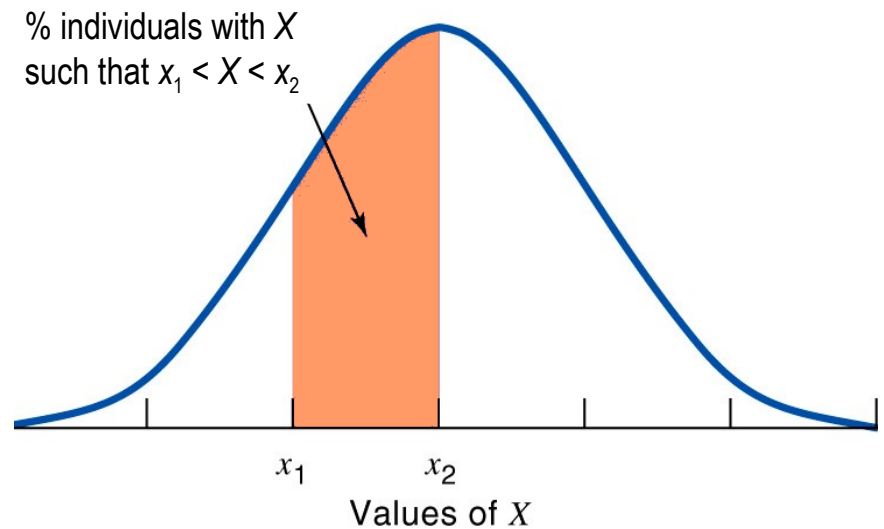What is the probability that Y is < 1?

What is the probability that Y < 0.5?

Continuous random variable and population distribution

The shaded area under a density curve shows the proportion, or %, of individuals in a population with values of X between x_1 and x_2.

% individuals with X
such that $x_1 < X < x_2$

x_1 x_2
Values of X

Because the <u>probability</u> of drawing **one individual at random** depends on the <u>frequency</u> of this type of individual in the population, the probability is also the shaded area under the curve.

$P(x_1 < X < x_2)$

x_1 x_2
Values of X
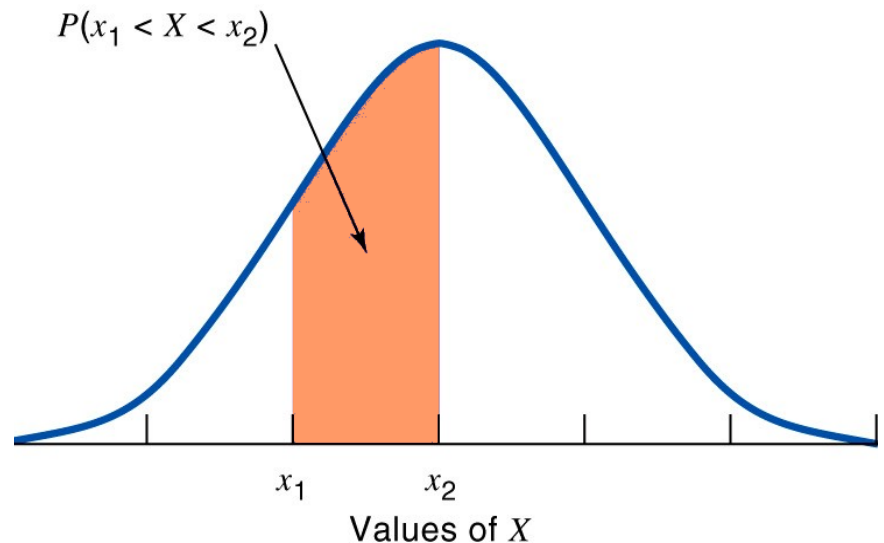
Normal probability distributions

The probability distribution of many random variables is a **normal distribution.** It shows what values the random variable can take and is used to assign probabilities to those values.
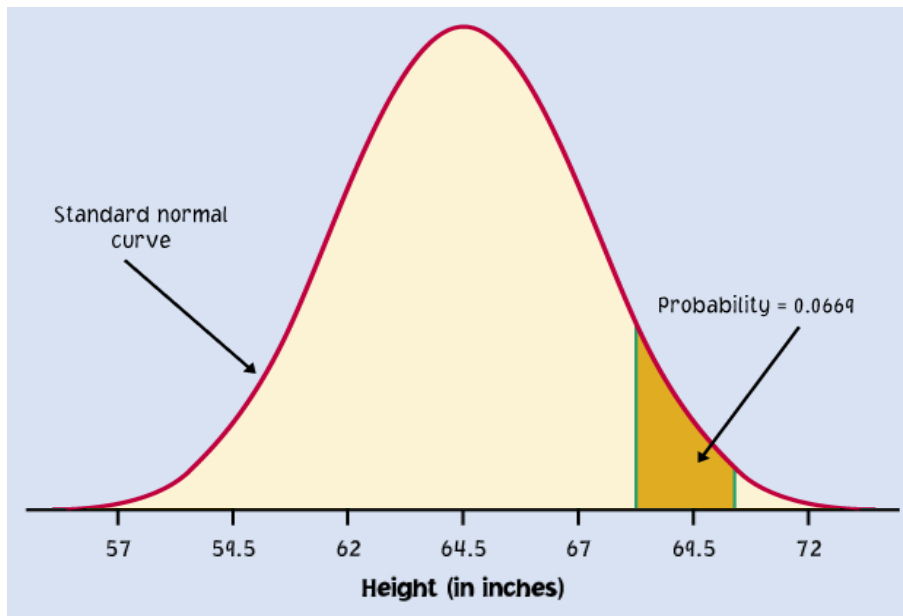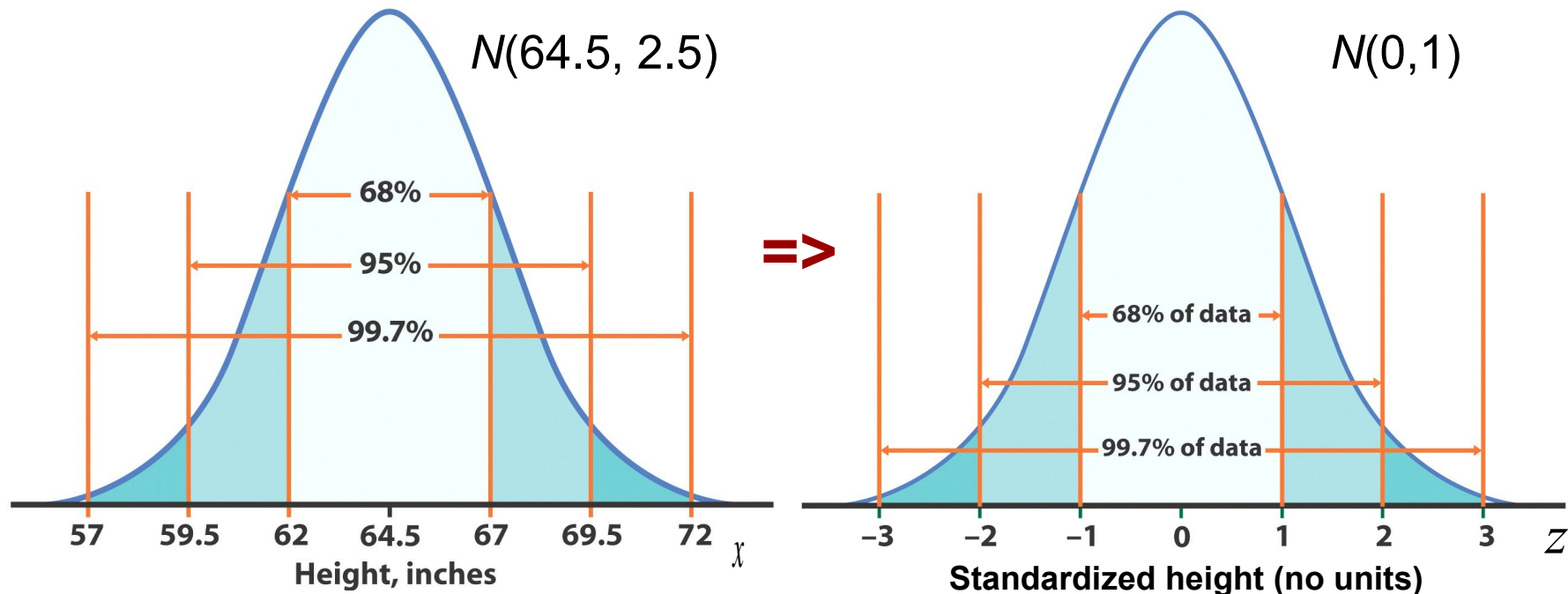

Example: Probability distribution of women's heights.

Here, since we chose a woman randomly, her height, *X*, is a random variable.

To calculate probabilities with the normal distribution, we will standardize the random variable (*z* score) and use Table A.

Reminder: standardizing $N(\mu,\sigma)$

We **standardize** normal data by calculating *z*-scores so that any Normal curve $N(\mu,\sigma)$ can be transformed into the standard Normal curve $N(0,1)$.

$N(64.5, 2.5)$

68%
95%
99.7%

57   59.5   62   64.5   67   69.5   72   *x*
Height, inches

=>

$N(0,1)$

68% of data
95% of data
99.7% of data

−3   −2   −1   0   1   2   3   *z*
Standardized height (no units)

$$z = \frac{(x - \mu)}{\sigma}$$

€

What is the probability, if we pick one woman at random, that her height will be some value X? For instance, between 68 and 70 inches $P(68 < X < 70)$?
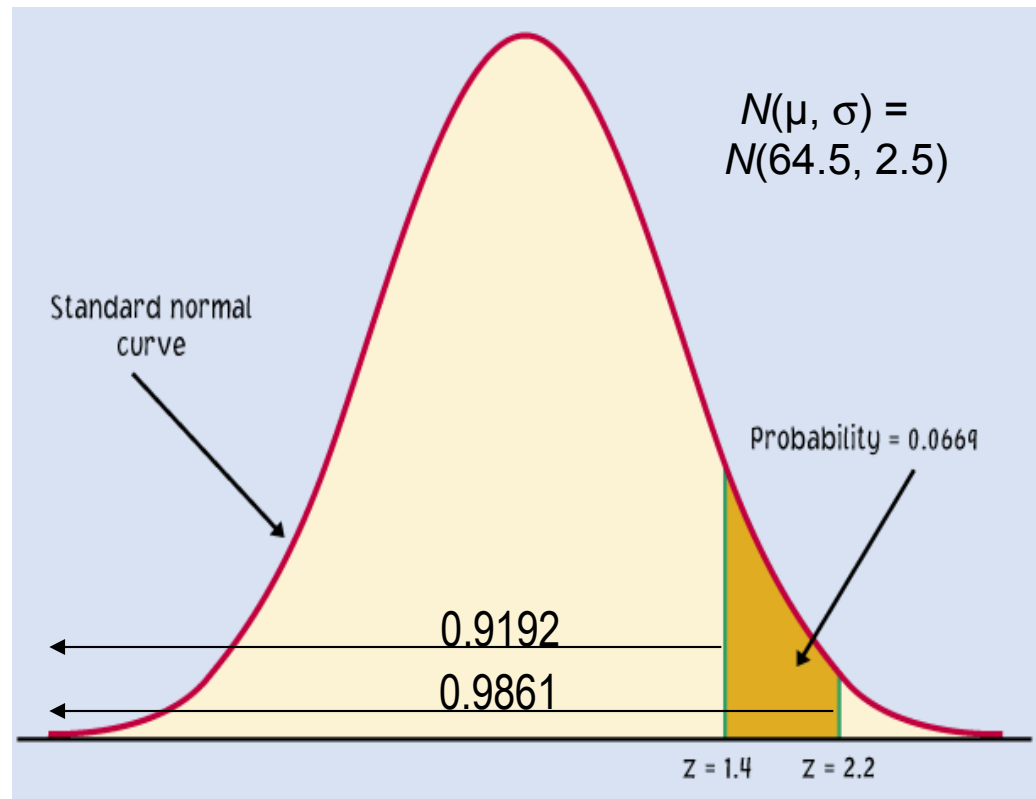
Because the woman is selected at random, X is a random variable.

$$z = \frac{(x - \mu)}{\sigma}$$

As before, we calculate the z-scores for 68 and 70.

€For x = 68", $z = \dfrac{(68 - 64.5)}{2.5} = 1.4$

For x = 70", $z = \dfrac{(70 - 64.5)}{2.5} = 2.2$

$N(\mu, \sigma) = N(64.5, 2.5)$

Standard normal curve

Probability = 0.0669

0.9192

0.9861

Z = 1.4   Z = 2.2

€

The area under the curve for the interval [68" to 70"] is 0.9861 − 0.9192 = 0.0669.

Thus, the probability that a randomly chosen woman falls into this range is 6.69%.

$P(68 < X < 70)$ = 6.69%

Inverse problem:

Your favorite chocolate bar is dark chocolate with whole hazelnuts.
The weight on the wrapping indicates 8 oz. Whole hazelnuts vary in weight, so
how can they guarantee you 8 oz. of your favorite treat? You are a bit skeptical...

To avoid customer complaints and
lawsuits, the manufacturer makes
sure that 98% of all chocolate bars
weigh 8 oz. or more.

The manufacturing process is
roughly normal and has a known
variability σ = 0.2 oz.

How should they calibrate the
machines to produce bars with a
mean μ such that $P(x < 8$ oz.$) =$
2%?
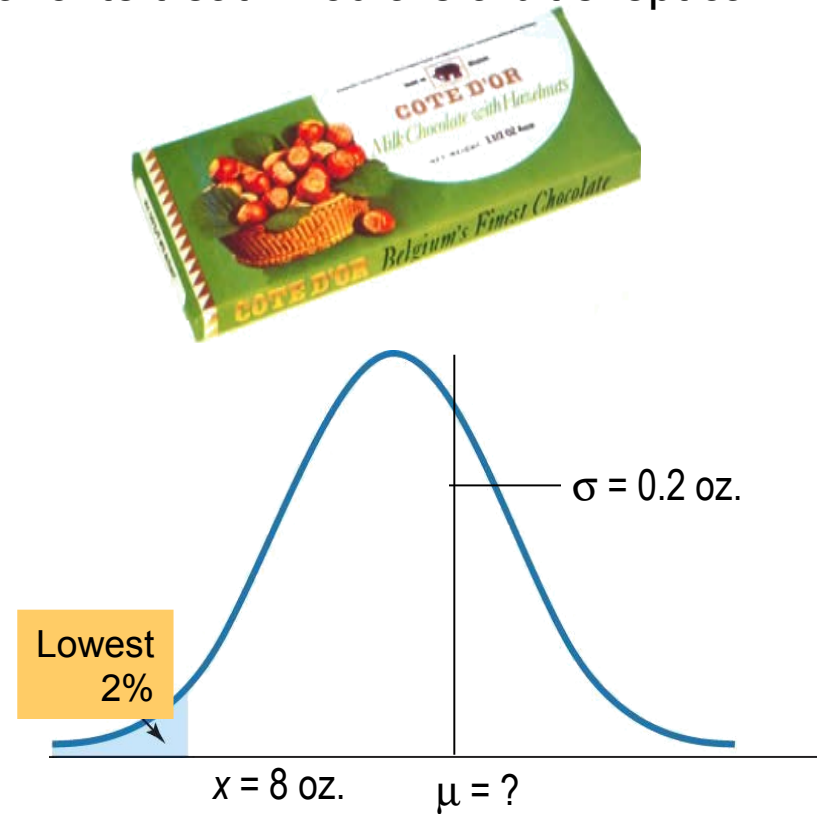
σ = 0.2 oz.

Lowest
2%

x = 8 oz.   μ = ?

How should they calibrate the machines to produce bars with a mean μ such that $P(x < 8 \text{ oz.}) = 2\%$?

σ = 0.2 oz.

Lowest 2%

x = 8 oz.   μ = ?
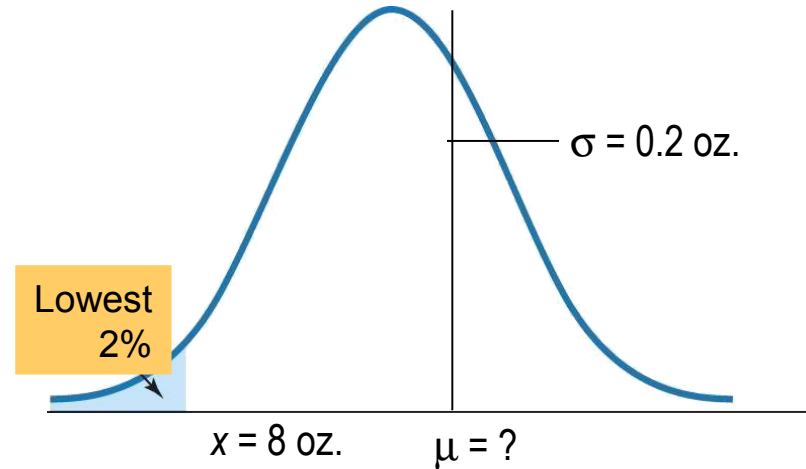
Here we know the area under the density curve (2% = 0.02) and we know *x* (8 oz.).

We want μ.

In table A we find that the *z* for a left area of 0.02 is roughly *z* = -2.05.

$$z = \frac{(x - \mu)}{\sigma} \iff \mu = x - (z * \sigma) \qquad \mu = 8 - (-2.05 * 0.2) = 8.41 \, oz.$$

Thus, your favorite chocolate bar weighs, on average, 8.41 oz. Excellent!!!

Mean of a random variable

The mean *x* bar of a set of observations is their arithmetic average.

The mean **μ** of a random variable *X* is a weighted average of the possible values of *X*, reflecting the fact that all outcomes might not be equally likely.

A basketball player shoots three free throws. The random variable *X* is the number of baskets successfully made ("H").

	HMM	HHM		
	MHM	HMH		
MMM	MMH	MHH	HHH	

Value of *X*	0	1	2	3
Probability	1/8	3/8	3/8	1/8

The mean of a random variable *X* is also called **expected value** of *X*.

Mean of a discrete random variable

For a discrete random variable X with probability distribution →

Value of X	x_1	x_2	x_3	$\cdots$	x_k
Probability	p_1	p_2	p_3	$\cdots$	p_k

the mean μ of X is found by multiplying each possible value of X by its probability, and then adding the products.

$$\mu_X = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k$$
$$= \sum x_i p_i$$

A basketball player shoots three free throws. The random variable X is the number of baskets successfully made.
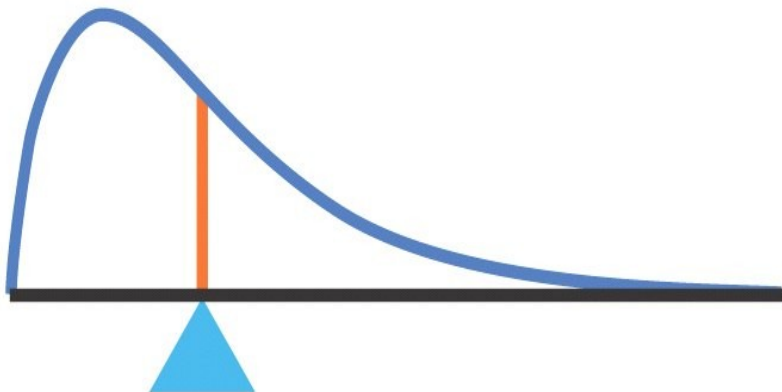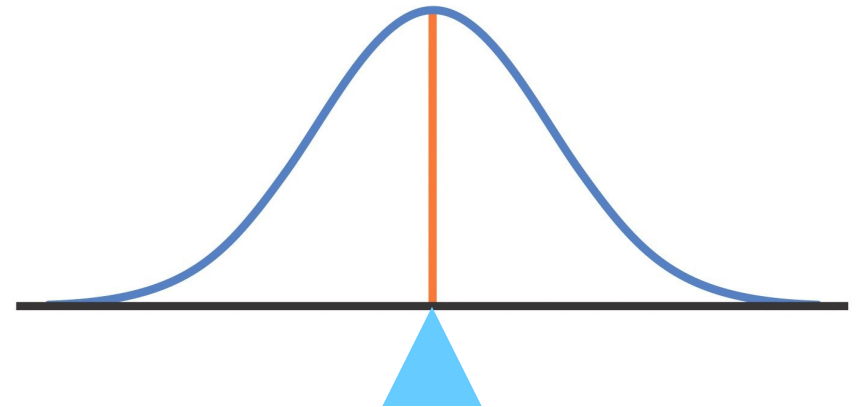
Value of X	0	1	2	3
Probability	1/8	3/8	3/8	1/8

The mean μ of X is

μ = (0*1/8) + (1*3/8) + (2*3/8) + (3*1/8)

= 12/8 = 3/2 = 1.5

Mean of a continuous random variable

The probability distribution of continuous random variables is
described by a density curve.

The mean lies at the center of
symmetric density curves
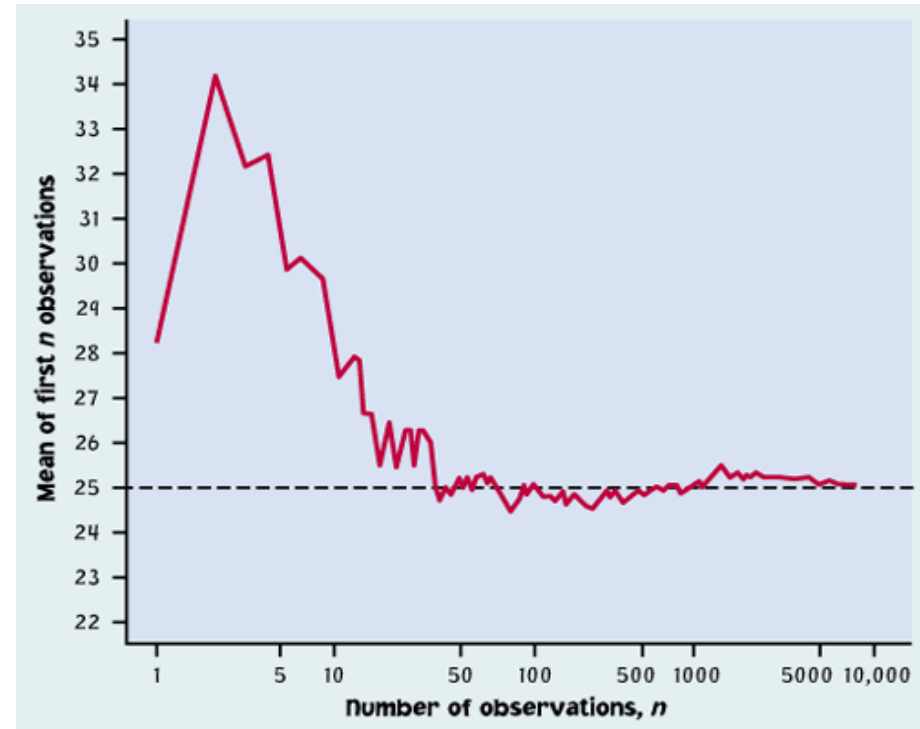such as the normal curves.

Exact calculations for the mean of
a distribution with a skewed
density curve are more complex.

Law of large numbers

As the number of randomly drawn observations (n) in a sample increases, the mean of the sample (x bar) gets closer and closer to the population mean μ.

This is the **law of large numbers.** It is valid for <u>any</u> population.

Note: We often intuitively expect predictability over a few random observations, but it is wrong. The law of large numbers only applies to <u>really</u> large numbers.

Variance of a random variable

The variance and the standard deviation are the measures of spread that accompany the choice of the mean to measure center.

The **variance σ^2_X** of a random variable is a weighted average of the squared deviations $(X - \mu_X)^2$ of the variable X from its mean μ_X. Each outcome is weighted by its probability in order to take into account outcomes that are not equally likely.

The larger the variance of X, the more scattered the values of X on average. The positive square root of the variance gives the **standard deviation σ** of X.

Variance of a discrete random variable

For a discrete random variable X with probability distribution →

Value of X	x_1	x_2	x_3	$\cdots$	x_k
Probability	p_1	p_2	p_3	$\cdots$	p_k

and mean μ_X, the variance σ^2 of X is found by multiplying each squared deviation of X by its probability and then adding all the products.

$$\sigma_X^2 = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \cdots + (x_k - \mu_X)^2 p_k$$
$$= \sum (x_i - \mu_X)^2 p_i$$

A basketball player shoots three free throws. The random variable X is the number of baskets successfully made.

μ_X = 1.5.

Value of X	0	1	2	3
Probability	1/8	3/8	3/8	1/8

The variance σ^2 of X is

$\sigma^2 = 1/8*(0-1.5)^2 + 3/8*(1-1.5)^2 + 3/8*(2-1.5)^2 + 1/8*(3-1.5)^2$

$= 2*(1/8*9/4) + 2*(3/8*1/4) = 24/32 = 3/4 = .75$

Rules for means and variances

If X is a random variable and a and b are fixed numbers, then

$$\mu_{a+bX} = a + b\mu_X$$

$$\sigma^2_{a+bX} = b^2\sigma^2_X$$

If X and Y are two independent random variables, then

$$\mu_{X+Y} = \mu_X + \mu_Y$$

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$

If X and Y are NOT independent but have correlation ρ, then

$$\mu_{X+Y} = \mu_X + \mu_Y$$

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$$

Investment

You invest 20% of your funds in Treasury bills and 80% in an "index fund" that represents all U.S. common stocks. Your rate of return over time is proportional to that of the T-bills (X) and of the index fund (Y), such that $R = 0.2X + 0.8Y$.

Based on annual returns between 1950 and 2003:

- Annual return on T-bills $\mu X = 5.0\%$ $\sigma X = 2.9\%$
- Annual return on stocks $\mu Y = 13.2\%$ $\sigma Y = 17.6\%$
- Correlation between X and $Y \rho = -0.11$

$\mu_R = 0.2\mu_X + 0.8\mu_Y = (0.2*5) + (0.8*13.2) = 11.56\%$

$\sigma^2_R = \sigma^2_{0.2X} + \sigma^2_{0.8Y} + 2\rho\sigma_{0.2X}\sigma_{0.8Y}$

$= 0.2*2\sigma^2_X + 0.8*2\sigma^2_Y + 2\rho*0.2*\sigma_X*0.8*\sigma_Y$

$= (0.2)2(2.9)2 + (0.8)2(17.6)2 + (2)(-0.11)(0.2*2.9)(0.8*17.6) = 196.786$

$\sigma_R = \sqrt{196.786} = 14.03\%$

The portfolio has a smaller mean return than an all-stock portfolio, but it is also less risky.

Probability: The Study of Randomness
General Probability Rules

IPS Chapter 4.5

Objectives (IPS Chapter 4.5)

General probability rules
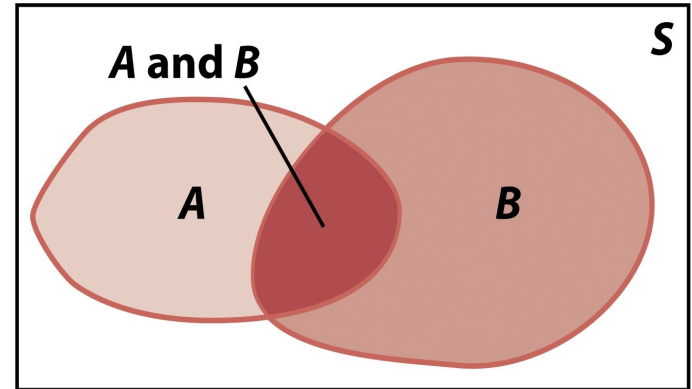
- ◻ General addition rules

- ◻ Conditional probability

- ◻ General multiplication rules

- ◻ Tree diagrams

- ◻ Bayes's rule

General addition rules

General addition rule for <u>any</u> two events A and B:

The probability that A occurs,

B occurs, or both events occur is:

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

What is the probability of randomly drawing either an ace or a heart from a deck of 52 playing cards? There are 4 aces in the pack and 13 hearts. However, 1 card is both an ace and a heart. Thus:

$P(\text{ace or heart}) = P(\text{ace}) + P(\text{heart}) - P(\text{ace and heart})$

$$= 4/52 + 13/52 - 1/52 = 16/52 \approx .3$$

Conditional probability

Conditional probabilities reflect how the probability of an event can change if we know that some other event has occurred/is occurring.

- ❑ Example: The probability that a cloudy day will result in rain is different if you live in Los Angeles than if you live in Seattle.

- ❑ Our brains effortlessly calculate conditional probabilities, updating our "degree of belief" with each new piece of evidence.

The conditional probability of event B given event A is: (provided that $P(A) \neq 0$)

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

General multiplication rules

▫ The probability that <u>any</u> two events, A and B, both occur is:

$$P(A \text{ and } B) = P(A)P(B|A)$$

This is the **general multiplication rule.**

▫ If A and B are independent, then $P(A \text{ and } B) = P(A)P(B)$

(A and B are independent when they have no influence on each other's occurrence.)

What is the probability of randomly drawing either an ace or heart from a deck of 52 playing cards? There are 4 aces in the pack and 13 hearts.
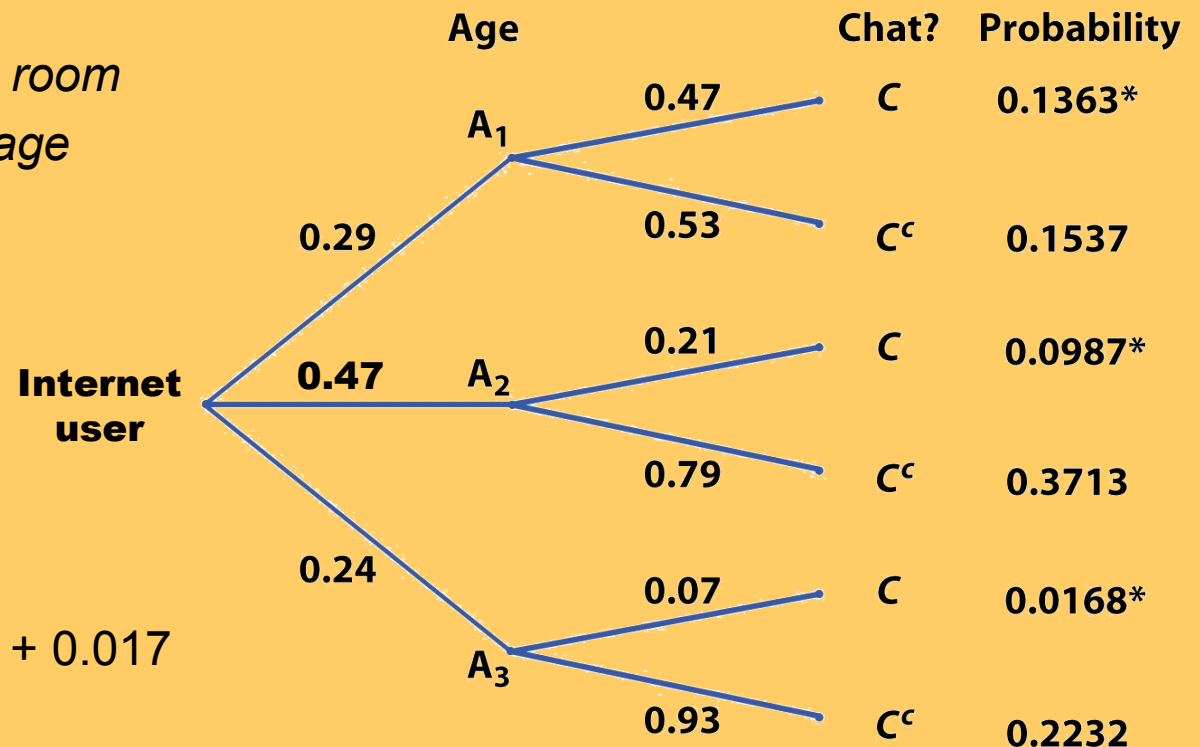
$P(\text{heart}|\text{ace}) = 1/4$ $P(\text{ace}) = 4/52$

➔$P(\text{ace and heart}) = P(\text{ace})* P(\text{heart}|\text{ace}) = (4/52)*(1/4) = 1/52$

Notice that heart and ace are independent events.

Probability trees

Conditional probabilities can get complex, and it is often a good strategy to build a **probability tree** that represents all possible outcomes graphically and assigns conditional probabilities to subsets of events.

*Tree diagram for chat room habits for three adult age groups*

|	Age	Chat?	Probability
	A_1 0.47	C	0.1363*
0.29	0.53	C^c	0.1537
Internet user 0.47	A_2 0.21	C	0.0987*
	0.79	C^c	0.3713
0.24	A_3 0.07	C	0.0168*
	0.93	C^c	0.2232

P(chatting) $= 0.136 + 0.099 + 0.017$
$= 0.252$

About 25% of all adult Internet users visit chat rooms.

Breast cancer screening

If a woman in her 20s gets screened for breast cancer and receives a positive test result, what is the probability that she does have breast cancer?
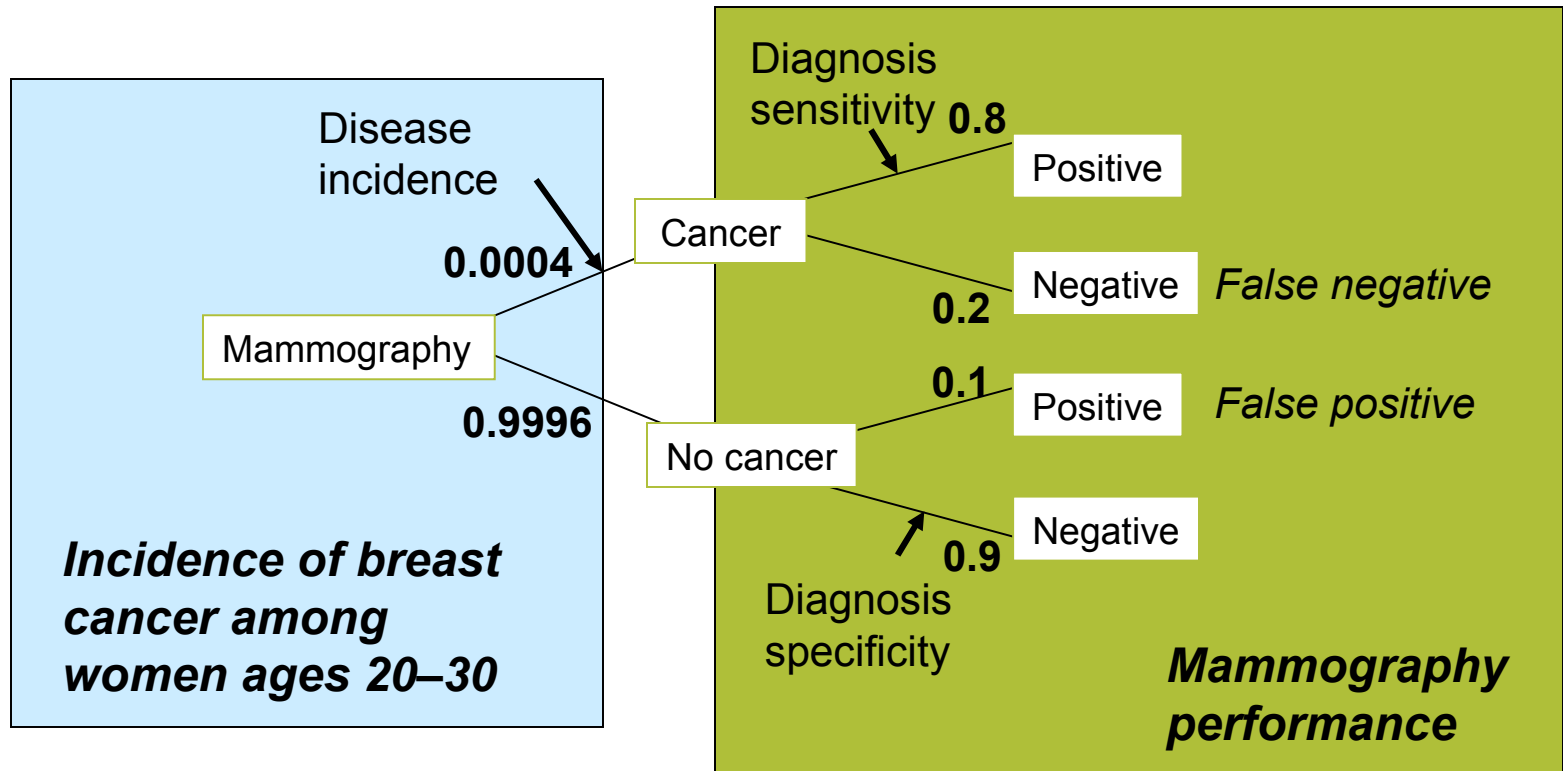
She could either have a positive test and have breast cancer or have a positive test but not have cancer (false positive).
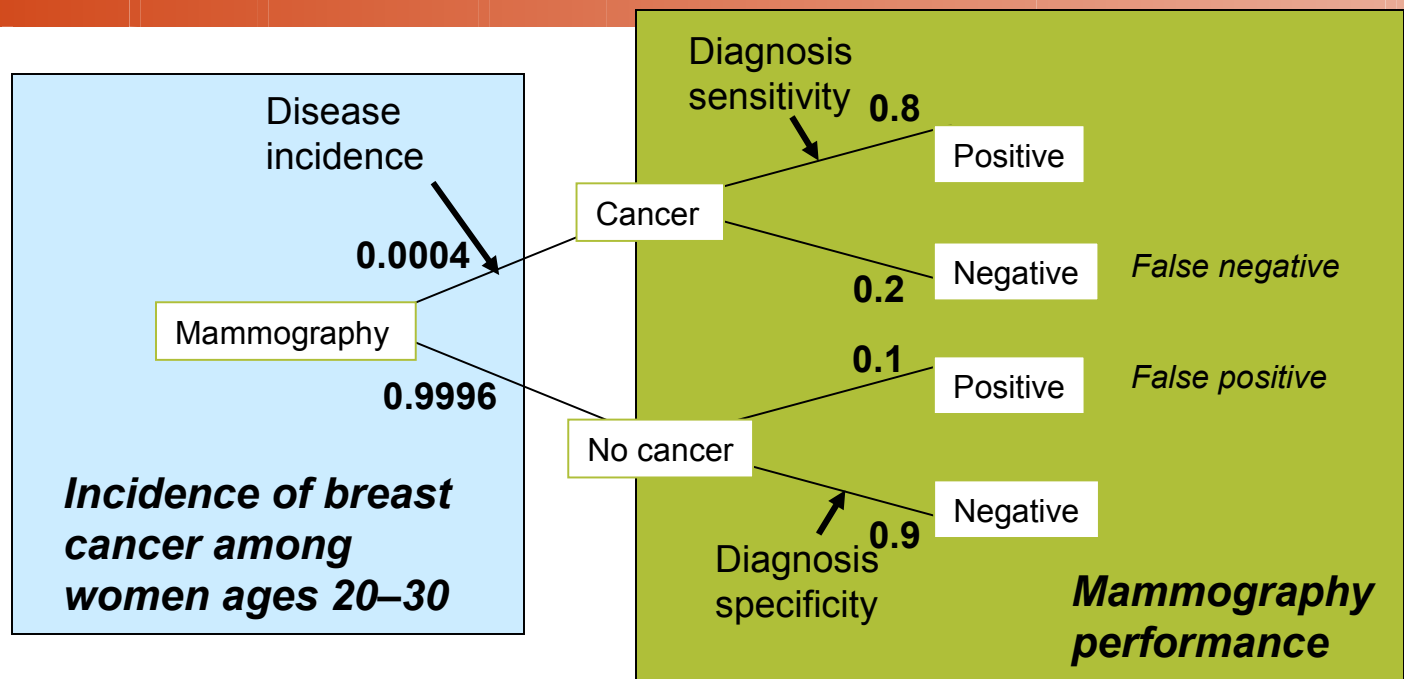
Possible outcomes given the positive diagnosis: positive test and breast cancer or positive test but no cancer (false positive).

$$P(cancer \mid pos) = \frac{P(cancer\ and\ pos)}{P(cancer\ and\ pos) + P(nocancer\ and\ pos)}$$

$$= \frac{0.0004 * 0.8}{0.0004 * 0.8 + 0.9996 * 0.1} \approx 0.3\%$$

This value is called the positive predictive value, or *PV+*. It is an important piece of information but, unfortunately, is rarely communicated to patients.

Bayes's rule

An important application of conditional probabilities is Bayes's rule. It is the foundation of many modern statistical applications beyond the scope of this textbook.

* If a sample space is decomposed in k disjoint events, A_1, A_2, … , A_k — none with a null probability but $P(A_1) + P(A_2) + … + P(A_k) = 1$,
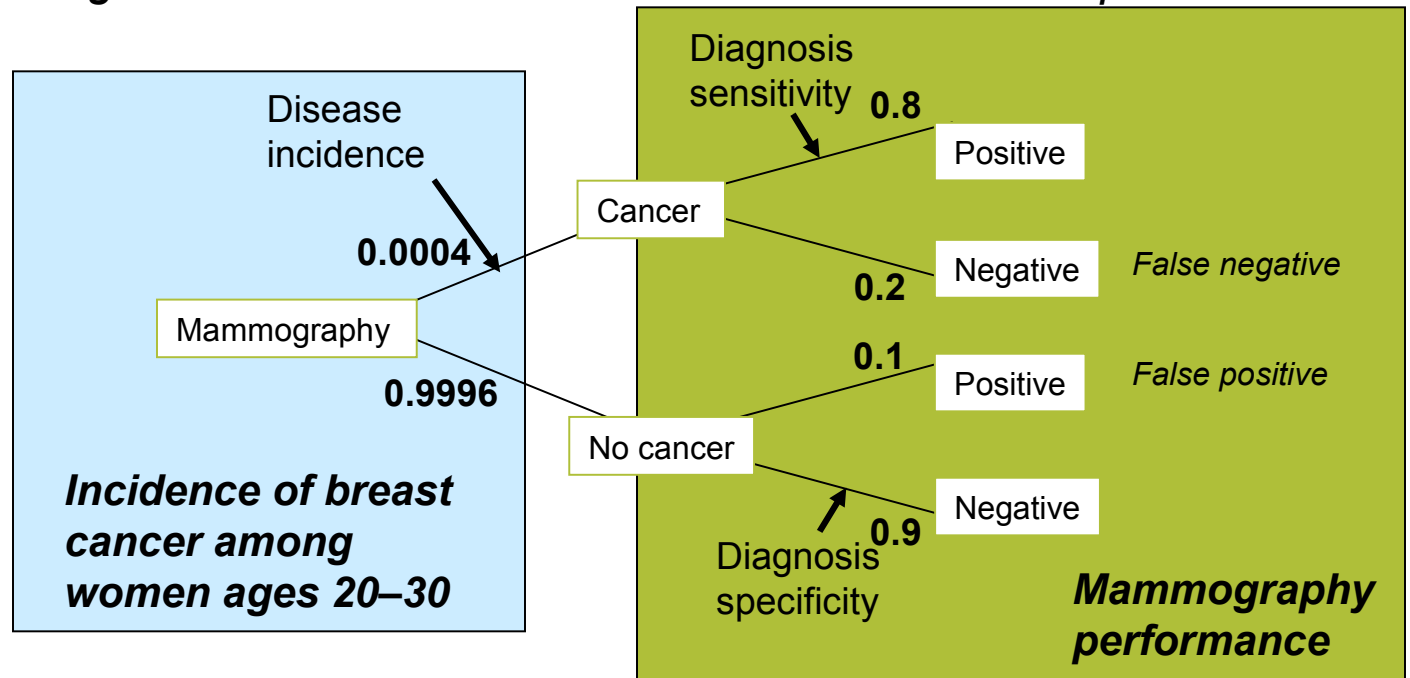
* And if C is any other event such that $P(C)$ is not 0 or 1, then:

$$P(A_i \mid C) = \frac{P(C \mid A_i)P(A_i)}{P(C \mid A_1)P(A_1) + P(C \mid A_2)P(A_2) + \cdots + P(A_k)P(C \mid A_k)}$$

However, it is often intuitively much easier to work out answers with a probability tree than with these lengthy formulas.

If a woman in her 20s gets screened for breast cancer and receives a positive test result, what is the probability that she does have breast cancer?

Disease incidence

Diagnosis sensitivity **0.8**

Positive

0.0004

Cancer

Negative   *False negative*

0.2

Mammography

0.1   Positive   *False positive*

0.9996

No cancer

Negative

0.9

Diagnosis specificity

Incidence of breast cancer among women ages 20–30

Mammography performance

This time, we use Bayes's rule:

$$P(A_i \mid C) = \frac{P(C \mid A_i)P(A_i)}{P(C \mid A_1)P(A_1) + P(C \mid A_2)P(A_2) + \cdots + P(A_k)P(C \mid A_k)}$$

A1 is cancer, A2 is no cancer, *C* is a positive test result.

$$P(cancer \mid pos) = \frac{P(pos \mid cancer)P(cancer)}{P(pos \mid cancer)P(cancer) + P(pos \mid nocancer)P(nocancer)}$$

$$= \frac{0.8 * 0.0004}{0.8 * 0.0004 + 0.1 * 0.9996} \approx 0.3\%$$