

Statistics 301: Probability and Statistics

Probability

Module 3

Updated 2019

Definitions

probability: likelihood of an event in an experiment occurring; like the chance of precipitation tomorrow. There are personal probabilities (subjective) and empirical probabilities (objective)

outcome: result of experiment

event: any combination of outcomes, can be a collection of one or more outcomes; most often mathematically represented with capital (English) letters, like A , B , etc.

probability of event A : denoted as $P(A)$, meaning the chance that event A occurs

sample space (S): the entire collection of events from an experiment; the set of all possible outcomes of an experiment

trial: every instance of an experiment; the total number of trials = n

Probability Rules/Formulas

To start...

- (1) $0 \leq P(A) \leq 1$, probabilities must be between 0 and 1 (0% and 100%)
- (2) $\sum P(A_i) = 1 = S$, the sum of the probabilities for an experiment must sum to 1 (the sample space)
- (3) $P(A') = 1 - P(A)$, the complement of the probability of event A is 1 minus the probability of A ; it is whatever is not in A .
 - The symbol is “A-prime”, is stated as “A-complement” or “A-not” (complement, a number or quantity of something required to make a group complete, not to be confused with compliment, to give praise)

More on the way...

Intersections and Unions

union: the union of two events, A and B , is all of the outcomes from A or B . The symbol is a \cup and the key word to watch for is *or*; the probability of the union of two events A and B is $P(A \cup B)$, or $P(A \text{ or } B)$

intersection: the intersection of two events, A and B , is the set of events that A and B have in common (where they overlap). The symbol is a \cap and the key word to watch for is *and*; the probability of the intersection of two events A and B is $P(A \cap B)$, or $P(A \text{ and } B)$

Example

Let $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $A = \{1, 2, 3, 4, 5\}$, and $B = \{4, 5, 6, 7, 8\}$.

union:

$$A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

with $\{4,5\}$ not listed twice (no duplicates)

intersection:

$$A \cap B = \{4, 5\}$$

Complements

$$P(A) = 0.5, P(B) = 0.5, P(A \cap B) = 0.2$$

$$P(A') = 1 - P(A) = 1 - 0.5 = 0.5 \text{ or } 50\%$$

$$P(B') = 1 - P(B) = 1 - 0.5 = 0.5 \text{ or } 50\%$$

Events logistics II

mutually exclusive/disjoint: when there is no intersection between events, $P(A \cap B) = \emptyset$ (the intersection is an empty set, i.e. it does not exist). Events A and B cannot happen at the same time if they are disjoint (mutually exclusive). It is like being in two places at the same time (it cannot happen)

independent: If two events, A and B are independent, then the outcome of one event has no impact on the outcome of the other event; independence of events *cannot be assumed at your convenience*

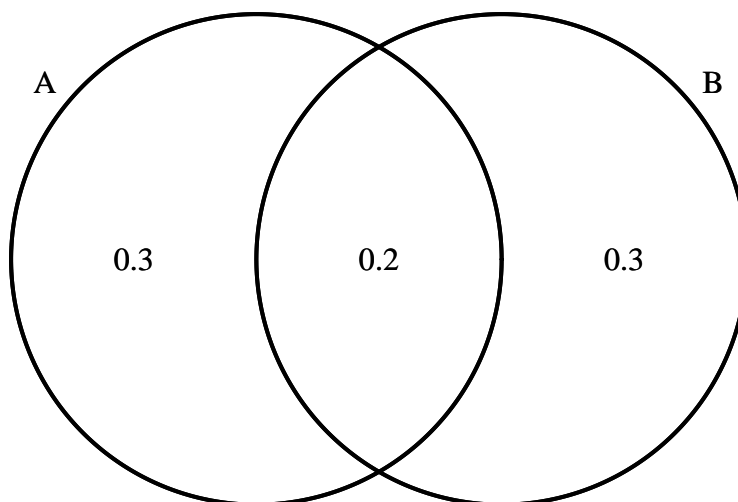
Disjoint events \neq independent events

Venn diagram I

A Venn diagram is a graph (picture) that represents the events of an experiment. It usually is a box that represents the sample space S and has circles that represent the events (A , B , etc.)

This one shows two events, A and B with an intersection.

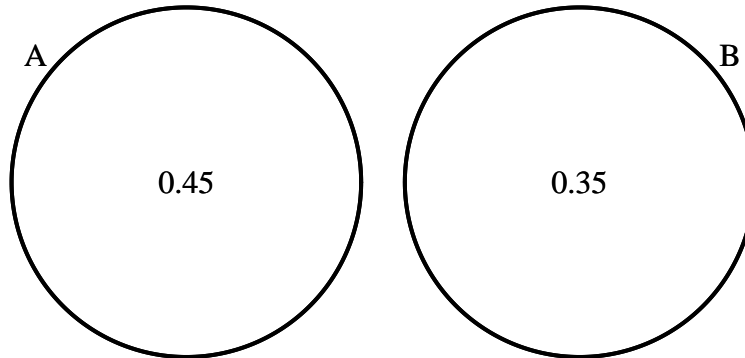
$$P(A) = 0.5, P(B) = 0.5, P(A \cap B) = 0.2$$



Venn diagram II

This one shows two events, A and B for disjoint/mutually exclusive events.

$$P(A) = 0.45, P(B) = 0.35, P(A \cap B) = \emptyset$$



Addition rule

(4) Addition rule

The probability of the *union* of A and B is:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

This formula can be modified if the union of one or both complements is required; if the union is given, then the intersection can be solved for. If two events are mutually exclusive, then the intersection is an empty set (it does not exist), the intersection is the sum of the probabilities of the two events. $P(A \cup B) = P(A) + P(B)$

Multiplication Rule

(5) Multiplication rule (**FOR INDEPENDENT EVENTS ONLY!!!**)

If two events, A and B are independent, the *intersection* of A and B can be calculated as:

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

This equation can also be used to prove or disprove independence

Again, it cannot be assumed, you have to either be told they are independent or prove it mathematically with the above formula.

Intersections and unions

Use addition rule for intersections (modification of the formulas is necessary to solve for values concerning complements)

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

When intersections are needed and you cannot assume independence, then use of the addition rule can be used and solved for the intersection.

$$P(A) = 0.5, P(B) = 0.5 \quad P(A \cap B) = 0.2$$

$$P(A \cup B) = 0.5 + 0.5 - 0.2 = 0.8$$

$$\text{Note that } = 1 - P(A \cap B) = 1 - 0.2 = 0.8$$

Conditional Probability

(6) Conditional probability

The probability of event A , given that event B has already occurred, is stated as $P(A \text{ given } B)$

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

This formula can be modified if the conditional probability of one or both complements is required; can also be used to prove independence. If A and B are independent, then $P(A|B) = P(A)$

Rules recap

- (1) $0 \leq P(A) \leq 1$
- (2) $\sum P(A_i) = 1 = S$
- (3) $P(A') = 1 - P(A)$
- (4) Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- (5) Multiplication rule: $P(A \cap B) = P(A)P(B)$ (independent events only)
- (6) Conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Pay attention to the formulas. They can be modified for finding complements, as well as solving for unknown values with a bit of algebra

Confusion Matrix

One way to help calculating probabilities of intersections, many time for use in other calculations

- (1) Create a 2X2 table, including row and column headers and totals
- (2) Label the rows with $P(A), P(A')$ and the columns with $P(B), P(B')$
- (3) The row totals are the values of $P(A), P(A')$ and the columns totals are $P(B), P(B')$
- (4) The grand total (bottom right corner cell) is always 1 ($\sum P(x_i) = 1$)
- (5) The upper left cell of the matrix (table) is $P(A \cap B)$; it has to be given or calculated
- (6) Solve for the other 3 cells of the matrix by way of row and column totals

Matrix setup

Table 4: Matrix template

	$P(B)$	$P(B')$	
$P(A)$	$P(A \cap B)$	$P(A \cap B')$	$p(a)$
$P(A')$	$P(A' \cap B)$	$P(A' \cap B')$	$p(a')$
	$p(b)$	$p(b')$	1

Matrix example

Suppose that $P(A) = 0.5$, $P(B) = 0.3$, $P(A \cap B) = 0.2$

Table 3: confusion matrix

	$P(B)$	$P(B')$	
$P(A)$	0.2	0.3	0.5
$P(A')$	0.1	0.4	0.5
	0.3	0.7	1

Matrix calculations

$$P(A \cap B') = 0.5 - 0.2 = 0.3$$

$$P(A' \cap B) = 0.3 - 0.2 = 0.1$$

$$P(A' \cap B') = 0.5 - 0.1 = 0.4 \text{ or } = 0.7 - 0.3 = 0.4$$

More matrix

$$P(A|B') = \frac{P(A \cap B')}{P(B')} = \frac{0.3}{0.7} \approx 0.429$$

Are A and B independent? We can use either $P(A \cap B) = P(A)P(B)$ or $P(A|B) = P(A)$ to prove it.

$$P(A \cap B) \stackrel{?}{=} P(A)P(B) \Rightarrow 0.2 \neq (0.5)(0.5)$$

$$P(A|B) \stackrel{?}{=} P(A) \Rightarrow 0.429 \neq 0.5$$

With either method, the results are the same. Since the statements were false, events A and B are not independent (they are dependent).

Example of flipping two fair coins

When you flip a fair coin (one that is not weighted so that one side is more likely to come up than the other), the chance that it is a head or a tail is equally likely.

List out all possible combinations

2 heads: (H,H); 1 head: (H,T), (T,H); 0 heads: (T,T)

$P(H) = 0.5, P(T) = 0.5$: flipping coins are independent events

2 heads: $P(H, H) = P(H1^{st} \cap H2^{nd}) = P(H1)P(H2) = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$

1 head: $P(H, T) = P[(H1^{st} \cap T2^{nd}) \cup (T1^{st} \cap H2^{nd})] = P(H1)P(T2) + P(T1)P(H2) = (\frac{1}{2})^2 + (\frac{1}{2})^2 = (\frac{1}{2})$

0 heads: $P(T, T) = P(T1^{st} \cap T2^{nd}) = P(T1)P(T2) = (\frac{1}{2})^2 = \frac{1}{4}$

Probability distribution of number of heads

Table 6: Coin...coin!

X=# heads	0	1	2
$P(x_i)$	0.25	0.5	0.25

Example of the sum of two 6-sided dice

When you roll a fair 6-sided die (one that is not weighted so that one side is more likely to come up than the others), the chance that it is a 1 is one-sixth, 2 is one-sixth, etc.; all sides are equally likely.

List out all possible combinations

sum=2: (1,1)

sum=3: (1,2), (2,1)

sum=4: (1,3), (3,1), (2,2)

sum=5: (1,4), (4,1), (2,3), (3,2)

sum=6: (1,5), (5,1), (2,4), (4,2), (3,3)

sum=7: (1,6), (6,1), (2,5), (5,2), (3,4), (4,3)

sum=8: (2,6), (6,2), (3,5), (5,3), (4,4)

sum=9: (3,6), (6,3), (4,5), (5,4)

sum=10: (4,6), (6,4), (5,5)

sum=11: (5,6), (6,5)

sum=12: (6,6)

Probability distribution of sum of two 6-sided dice

Table 2: Probability distribution of the sum of two 6-sided dice

	2	3	4	5	6	7	8	9	10	11	12
$P(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Contingency tables

Table displays sample values in relation to two different variables that may be dependent (contingent) on one another. You can display with counts or relative frequencies (probabilities)

Cell counts or probabilities are intersections, row and column totals are the individual counts or probabilities

Refer to the confusion matrix, it is the same thing. The inside cells of the table are intersections between rows and columns, with the row and column totals being the single events (A , B , etc.)

Table of counts

The following table shows a random sample of 100 hikers and the areas of hiking they prefer.

Table 1: Hiking locations

	Coastline	Near lakes/ streams	Mountain peaks	Total
male	16	25	14	55
female	18	16	11	45
Total	34	41	25	100

Counts table example

$$P(\text{male}) = \frac{55}{100} = 0.55, \quad P(\text{female}) = \frac{45}{100} = 0.45$$

$$P(\text{lakes}') = \frac{34 + 25}{100} = 0.59$$

$$P(\text{male} \cap \text{coastline}) = \frac{16}{100} = 0.16$$

$$P(\text{female} | \text{mtn}') = \frac{P(\text{female} \cap \text{mtn}')}{P(\text{mtn}')} = \frac{\frac{18+16}{100}}{\frac{34+41}{100}} = \frac{34}{75} \approx 0.453$$

Table of probabilities

Muddy Mouse lives in a cage with three doors. If Muddy goes out the first door, the probability that he gets caught by Alissa the cat is $1/5$ and the probability he is not caught is $4/5$. If he goes out the second door, the probability he gets caught by Alissa is $1/4$ and the probability he is not caught is $3/4$. The probability Alissa catches Muddy from the third door is $1/2$ and the probability he is not caught is $1/2$. It is equally likely that Muddy will choose any of the three doors so the probability of choosing each door is $1/3$. (book p. 194)

Mouse examples I

“If Muddy goes out the first door, the probability that he gets caught by Alissa the cat is $1/5$ and the probability he is not caught is $4/5$.” These are conditional probabilities with $P(\text{caught} | \text{door1}) = \frac{1}{5}$ and $P(\text{caught}' | \text{door1}) = \frac{4}{5}$

“If he goes out the second door, the probability he gets caught by Alissa is 1/4 and the probability he is not caught is 3/4.” $P(\text{caught}|\text{door}2) = \frac{1}{4}$ and $P(\text{caught}'|\text{door}2) = \frac{3}{4}$

“If he goes out the third door, the probability that he gets caught by Alissa the cat is 1/2 and the probability he is not caught is 1/2.” $P(\text{caught}|\text{door}3) = \frac{1}{2}$ and $P(\text{caught}'|\text{door}3) = \frac{1}{2}$

“It is equally likely that Muddy will choose any of the three doors so the probability of choosing each door is 1/3.” This means that his choice of any door is 1/3 so $P(\text{door}1) = P(\text{door}2) = P(\text{door}3) = \frac{1}{3}$

Mouse examples II

Use the formula for conditional probability to solve for the intersections

$$P(\text{caught} \cap \text{door}1) = P(\text{caught}|\text{door}1)P(\text{door}1) = \left(\frac{1}{5}\right) \left(\frac{1}{3}\right) = \frac{1}{15}$$

$$P(\text{caught}' \cap \text{door}1) = P(\text{caught}'|\text{door}1)P(\text{door}1) = \left(\frac{4}{5}\right) \left(\frac{1}{3}\right) = \frac{4}{15}$$

and so on...

Situations like Muddy the Mouse are challenging so make sure you read very carefully and reorganize the information so that you make sure as to what you really have

Muddy table

Table 5: Not with a mouse, not in a house...

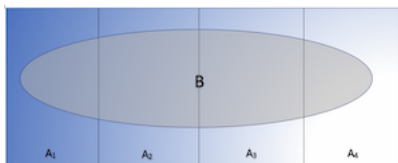
	Door #1	Door #2	Door #3	Total
caught	$\frac{1}{15}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{19}{60}$
not caught	$\frac{4}{15}$	$\frac{3}{12}$	$\frac{1}{6}$	$\frac{41}{60}$
Total	$\frac{5}{15}$	$\frac{4}{12}$	$\frac{2}{6}$	1

Law of Total Probability I

Reviewing Rule (6), conditional probability: the probability of event A , given that event B has already occurred:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$$

Many times the probability of interest is the denominator $P(B)$, but it is not directly known. When B arises in connection with events A_1, A_2, \dots, A_k , which constitute a *partition* of the sample space (i.e., they are mutually exclusive (disjoint) and make up the entire sample space); see figure.



Law of Total Probability II

If the probability of each A_i and the conditional probability of B given each A_i are all known, the Law of Total Probability expresses the probability of B as, and holds for any event B :

$$P(B) = \sum_{j=1}^k P(B|A_j)P(A_j) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)$$

The events of A_1, A_2, \dots, A_k can also be thought of as a stratification of the population.

Bayes' Theorem is to probability as Pythagorem's Theorem is to geometry¹

Bayes' Theorem

(7) Bayes' theorem: Let A_1, A_2, \dots, A_k be a collection of k mutually exclusive (disjoint) and exhaustive events with *prior* probabilities $P(A_i)$ ($i = 1, 2, \dots, k$). Then for any other event B for which $P(B) > 0$, the *posterior* probability of A_j given that B has occurred is

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum P(B|A_j)P(A_j)}$$

Another way to think about it: There is a partition A_1, A_2, \dots, A_k of the sample space and an event B (as shown in previous figure). The probabilities of the events A_i are given, and so are the conditional probabilities of B given that an A_i has occurred. Bayes' theorem answers the question: Given that B has occurred, what is the probability that a particular A_j has occurred?

(For just 2 events A and B : $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A')P(A')}$)

Bayes' theorem example I

Suppose that a test for using a particular drug has a 99% sensitivity rate and a 99% specificity rate. That is, the test will produce 99% true positive results for drug users and 99% true negative results for non-drug users. Suppose that 0.5% of people are users of the drug. What is the probability that a randomly selected individual with a positive test is actually a drug user?

$P(\text{user}) = 0.005$, $P(\text{user}') = 1 - P(\text{user}) = 1 - 0.005$, $P(+|\text{user}) = 0.99$, $P(-|\text{user}') = 0.99$, $P(+|\text{user}') = 1 - P(-|\text{user}') = 1 - 0.99 = 0.01$, and $P(-|\text{user}) = 1 - P(+|\text{user}) = 1 - 0.99 = 0.01$

$$P(\text{user}|+) = \frac{P(\text{user} \cap +)}{P(+)} = \frac{P(+|\text{user})P(\text{user})}{P(+|\text{user})P(\text{user})+P(+|\text{user}')P(\text{user}')} = \frac{(0.99)(0.005)}{(0.99)(0.005)+(0.01)(0.995)} = 0.3322$$

Interpretation of drug use and tests

Even if an individual tests positive, it is more likely that they do not use the drug than that they do. This is because the number of non-users is large compared to the number of users. The number of false positives outweighs the number of true positives. For example, if 1000 individuals are tested, there are expected to be 995 non-users and 5 users. From the 995 non-users, $0.01(995) \approx 10$ false positives are expected. From the 5 users, $(0.99)(5) \approx 5$ true positives are expected. Out of 15 positive results, only 5 are genuine. Yikes!

¹Jeffreys, Harold (1973). *Scientific Inference* (3rd ed.). Cambridge University Press. p. 31. ISBN 978-0-521-18078-8.

Bayes' theorem example II

The entire output of a factory is produced on three machines. The three machines account for 20%, 30%, and 50% of the factory output. The fraction of defective items produced is 5% for the first machine, 3% for the second machine, and 1% for the third machine. If an item is chosen at random from the total output and is found to be defective, what is the probability that it was produced by the third machine?

Here, the answer can be reached without recourse to the formula by applying the conditions to any hypothetical number of cases. For example, if 100,000 items are produced by the factory, 20,000 will be produced by Machine A, 30,000 by Machine B, and 50,000 by Machine C. Machine A will produce 1000 defective items, Machine B 900, and Machine C 500. Of the total 2400 defective items, only 500, or $5/24$ were produced by Machine C.

Bayes' example II solution

A solution is as follows. Let X_i denote the event that a randomly chosen item was made by the i^{th} machine (for $i = A, B, C$). Let Y denote the event that a randomly chosen item is defective. Then, we are given the following information:

$$P(X_A) = 0.2, P(X_B) = 0.3, P(X_C) = 0.5, P(Y|X_A) = 0.05, P(Y|X_B) = 0.03, P(Y|X_C) = 0.01$$

$$P(X_C|Y) = \frac{P(X_C \cap Y)}{P(Y)}. \text{ First find } P(Y) \text{ using the Law of Total Probability.}$$

$$P(Y) = P(Y|X_A)P(X_A) + P(Y|X_B)P(X_B) + P(Y|X_C)P(X_C) = (0.05)(0.2) + (0.03)(0.3) + (0.01)(0.5) = 0.024$$

$$\text{Now finish the problem: } P(X_C|Y) = \frac{P(X_C \cap Y)}{P(Y)} = \frac{P(Y|X_C)P(X_C)}{P(Y)} = \frac{(0.01)(0.5)}{0.024} = 0.2083 = \frac{5}{24}$$

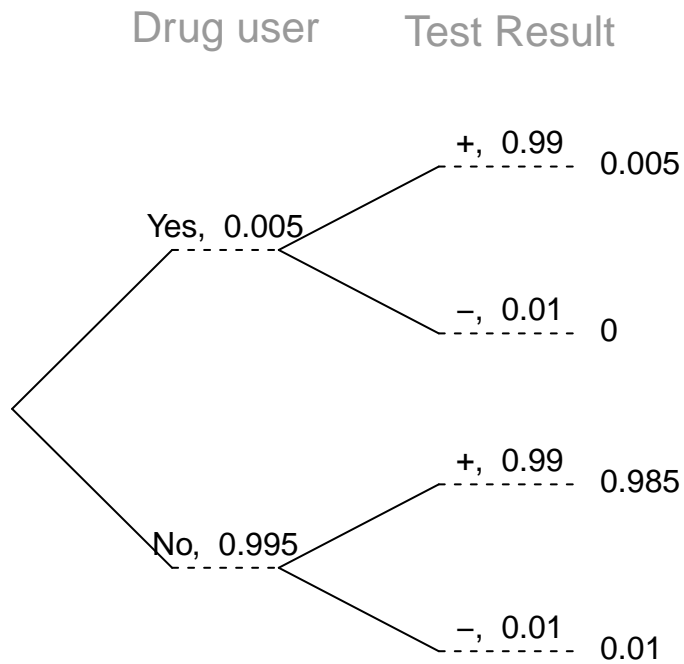
Interpretation of machines and defective products

Given that the item is defective, the probability that it was made by the third machine is only $5/24$. Although machine C produces half of the total output, it produces a much smaller fraction of the defective items. Hence the knowledge that the item selected was defective enables us to replace the prior probability $P(X_C) = 0.5$ by the smaller posterior probability $P(X_C|Y) = 0.2083$.

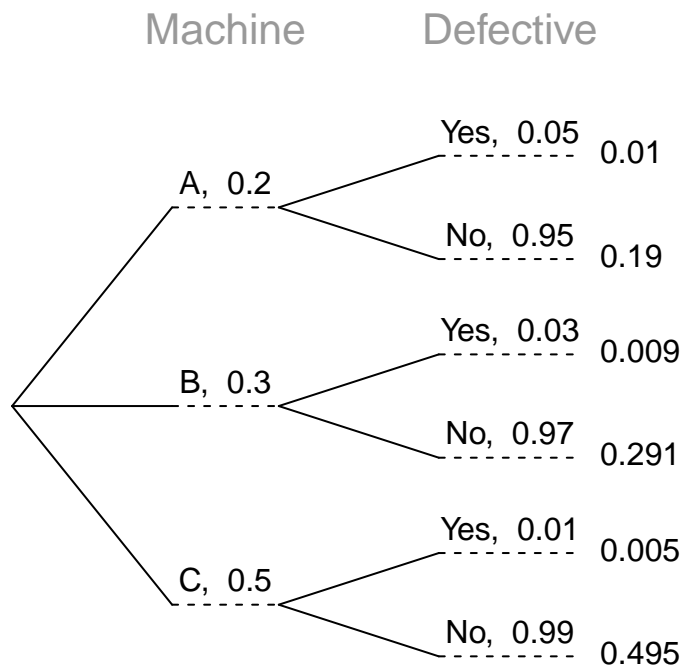
Tree diagrams

Another way to describe and visualize Bayes' theorem is to use a Tree diagram. Start with the initial condition branches, then the next branches are the conditional probabilities of the next event given the initial conditions. From there, most often the outside edge of the tree has the intersections computed by the product of the initial conditions and conditional probabilities.

Tree for drug use and tests



Tree for defects and machines



Rules recap, redux

(1) $0 \leq P(A) \leq 1$

(2) $\sum P(A_i) = 1 = S$

(3) $P(A') = 1 - P(A)$

(4) Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(5) Multiplication rule: $P(A \cap B) = P(A)P(B)$ (independent events only)

(6) Conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

(7) Bayes' Theorem: $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum P(B|A_j)P(A_j)}$