Conditional probability $\boxed{P(E \mid F) = \frac{P(E \cap F)}{P(F)}}$ assuming that $P(F) \neq 0$

INDEPENDENT EVENTS.

We say that two events A and B are independent if

$$P(A \cap B) = P(A)\,P(B) \qquad (*)$$

It follows from this and the definition of conditional probability that if A and B are independent then

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

Alternatively, one can take one of the two equalities in (**) as definition of independence and derive both the other and (*). Thus (*) and (**) are equivalent.

Going back to our example where $S=\{a,b,c,d,e,f,g\}$, $A=\{a,c,d,e\}$, $B=\{a,e,g\}$, $A\cap B=\{a,e\}$ and

a	b	c	d	e	f	g
$\frac{1}{14}$	$\frac{2}{14}$	0	$\frac{1}{14}$	$\frac{6}{14}$	$\frac{1}{14}$	$\frac{3}{14}$

$P(A)=\frac{1}{14}+0+\frac{1}{14}+\frac{6}{14}=\frac{8}{14}$, $P(B)=\frac{1}{14}+\frac{6}{14}+\frac{3}{14}=\frac{10}{14}$, $P(A\cap B)=\frac{1}{14}+\frac{6}{14}=\frac{7}{14}$ $\quad P(A)P(B)=\frac{8}{14}\cdot\frac{10}{14}=\frac{80}{14^2}$

not independent

On the other hand, if $S=\{a,b,c,d,e,f,g,h,i\}$ and

a	b	c	d	e	f	g	h	i
$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	0	$\frac{2}{9}$	0	$\frac{1}{9}$	$\frac{1}{9}$

Then $E=\{a,b,c,e\}$ and $F=\{c,d,e,g\}$ are independent. In fact, $P(E)=\frac{1}{3}$, $P(F)=\frac{1}{3}$,

$E\cap F=\{c,e\}$, $P(E\cap F)=\frac{1}{9}$ and thus $\underbrace{P(E\cap F)}_{\frac{1}{9}}=\underbrace{P(E)P(F)}_{\frac{1}{3}\cdot\frac{1}{3}=\frac{1}{9}}$.

E and F are independent

Bayes' formula

Let E and F be two events such that $P(E) > 0$ and $P(F) > 0$. Then

Assuming $P(F) \neq 0$

$$\boxed{P(E \mid F) = \frac{P(E \cap F)}{P(F)}} \tag{1}$$

and

$$\boxed{P(F \mid E) = \frac{P(E \cap F)}{P(E)}} \tag{2}$$

note that $F \cap E = E \cap F$

so $P(F \cap E) = P(E \cap F)$

Assuming $P(E) \neq 0$

From (2) we get that

multiply both sides by $P(E)$

$$\underbrace{P(F \mid E)\, P(E) = P(E \cap F)}_{\text{put it in the numerator of (1)}} \tag{3}$$

Substituting (3) into (1) we get

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)} = \frac{P(F \mid E)\, P(E)}{P(F)}$$

Bayes' formula (4)

Example: D = you have a disease

$\neg D$ = you don't have the disease

$+$ = you test positive

$-$ = you test negative

Information: $P(D) = 5\%$ base rate $\Rightarrow 100\% - 5\% = 95\%$

sensitivity of the test : $P(+|D) = 92\%$

specificity of the test : $P(-|\neg D) = 88\%$

$\Rightarrow P(+|\neg D) = 100\% - 88\% = 12\%$

$$P(D|+) = \frac{P(+|D)\,P(D)}{P(+)} \quad \text{by Bayes' rule}$$

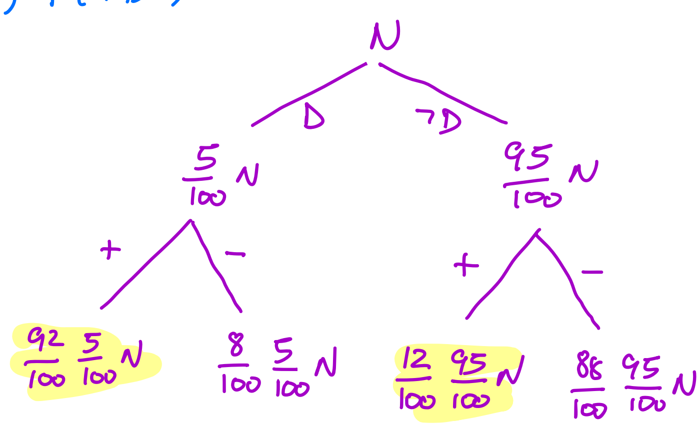
$$= \frac{\frac{92}{100}\cdot\frac{5}{100}}{P(+)} = \frac{\frac{92}{100}\cdot\frac{5}{100}}{\underbrace{P(+\cap D)} + \underbrace{P(+\cap\neg D)}}$$

$$+ = (+\cap D) \cup (+\cap\neg D) \quad \text{disjoint}$$

$$P(+) = P(+\cap D) + P(+\cap\neg D)$$

$$P(+\cap D) = P(+|D)\,P(D)$$

$$P(+\cap\neg D) = P(+|\neg D)\,P(\neg D)$$



$$P(D|+) = \frac{\frac{92}{100}\cdot\frac{5}{100}}{\frac{92}{100}\cdot\frac{5}{100} + \frac{12}{100}\cdot\frac{95}{100}}$$

$$P(D|+) = \frac{\frac{92}{100}\cdot\frac{5}{100}\,N}{\frac{92}{100}\,\frac{5}{100}\,N + \frac{12}{100}\cdot\frac{95}{100}\cdot N}$$

Bayes' theorem

Bayes' formula says that $\boxed{P(E \mid F) = \dfrac{P(F \mid E)\,P(E)}{P(F)}}$. From set theory we have that, given any two sets A and B,

$$A = (A \cap B) \cup (A \cap \neg B) \tag{5}$$

$(A \cap B) \cap (A \cap \neg B) = \phi$

and the two sets $(A \cap B)$ and $(A \cap \neg B)$ are disjoint. Thus $P(A) = P(A \cap B) + P(A \cap \neg B)$.

Hence in the denominator of Bayes' formula we can replace $P(F)$ with

Then, using conditional probability we get that $P(F \cap E) =$ and

$P(F \cap \neg E) =$

Thus $P(F) =$.

Replacing this in Bayes' formula we get

$$P(E \mid F) = \frac{P(F \mid E)\,P(E)}{P(F \mid E)\,P(E) + P(F \mid \neg E)\,P(\neg E)}$$

Bayes' theorem (6)

$$P(E \mid F) = \frac{P(F \mid E)\,P(E)}{P(F \mid E)P(E) + P(F \mid \neg E)P(\neg E)}$$

EXAMPLE.

Enrollment in a class is as follows: 60% econ majors (E), 40% other majors ($\neg E$). In the past, 80% of the econ majors passed and 65% of the other majors passed. A student tells you that she passed the class. What is the probability that she is an econ major? Let P stand for "Pass the class".

$$P(E) = \frac{60}{100} \qquad P(\neg E) = \frac{40}{100}$$

$$P(E \mid P_{ass}) = \frac{P(Pass \mid E) \cdot P(E)}{P(Pass \mid E) \cdot P(E) + P(Pass \mid \neg E)\, P(\neg E)}$$

$$\left.\begin{aligned} P(Pass \mid E) &= \frac{80}{100} \\ P(Pass \mid \neg E) &= \frac{65}{100} \end{aligned}\right\}$$

$$= \frac{\frac{80}{100} \cdot \frac{60}{100}}{\frac{80}{100} \cdot \frac{60}{100} + \frac{65}{100} \cdot \frac{40}{100}}$$

Back to previous examples

EXAMPLE 1. Testing for a disease

Base rate of a disease: percentage of the population that has the disease

Sensitivity of a test: percentage of those who have the disease that tests positive

Specificity of a test: percentage of those who **do not** have the disease that tests negative

Suppose:

Base rate = 6% $= P(D)$ $\quad P(\neg D) = 94\%$

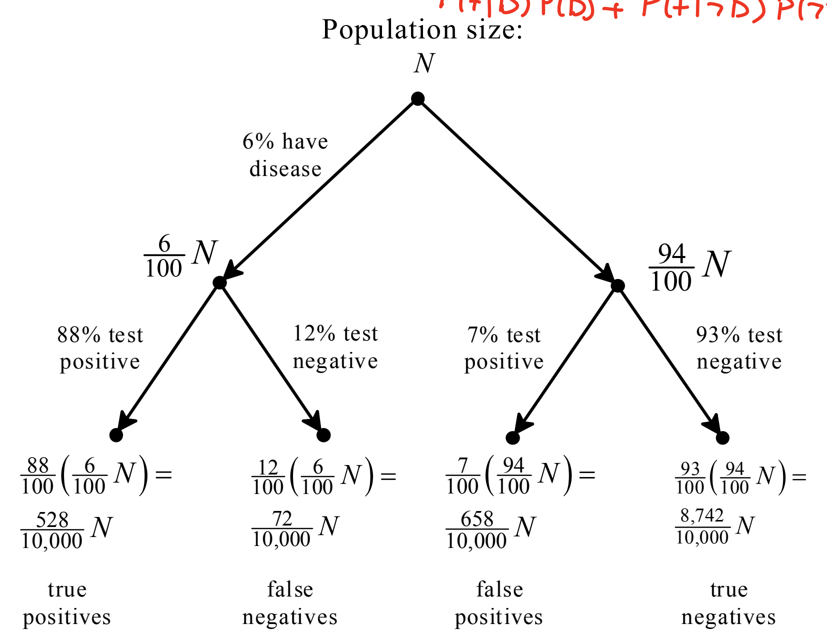
Sensitivity = 88% $= P(+|D)$ $\quad P(-|D) = 12\%$

Specificity = 93% $= P(-|\neg D)$ $\quad P(+|\neg D) = 7\%$

Suppose you test positive. What is the probability that you have the disease? $P(D|+)$

Previous analysis:

$$P(D|+) = \frac{P(+|D)\,P(D)}{P(+|D)\,P(D) + P(+|\neg D)\,P(\neg D)} = \frac{88 \cdot 6}{88 \cdot 6 + 7 \cdot 94}$$



The probability of having the disease, conditional on testing positive is:

$$\frac{528\dfrac{N}{10{,}000}}{528\dfrac{N}{10{,}000} + 658\dfrac{N}{10{,}000}} = \frac{528\dfrac{N}{10{,}000}}{1{,}186\dfrac{N}{10{,}000}} = \frac{528}{1{,}186} = 0.4452 = 44.52\%$$

EXAMPLE 2. More than two categories

Base rates of seniors who graduated within the past 6 months:

CS	*Econ*	*Hist*	*Math*
28%	40%	12%	20%

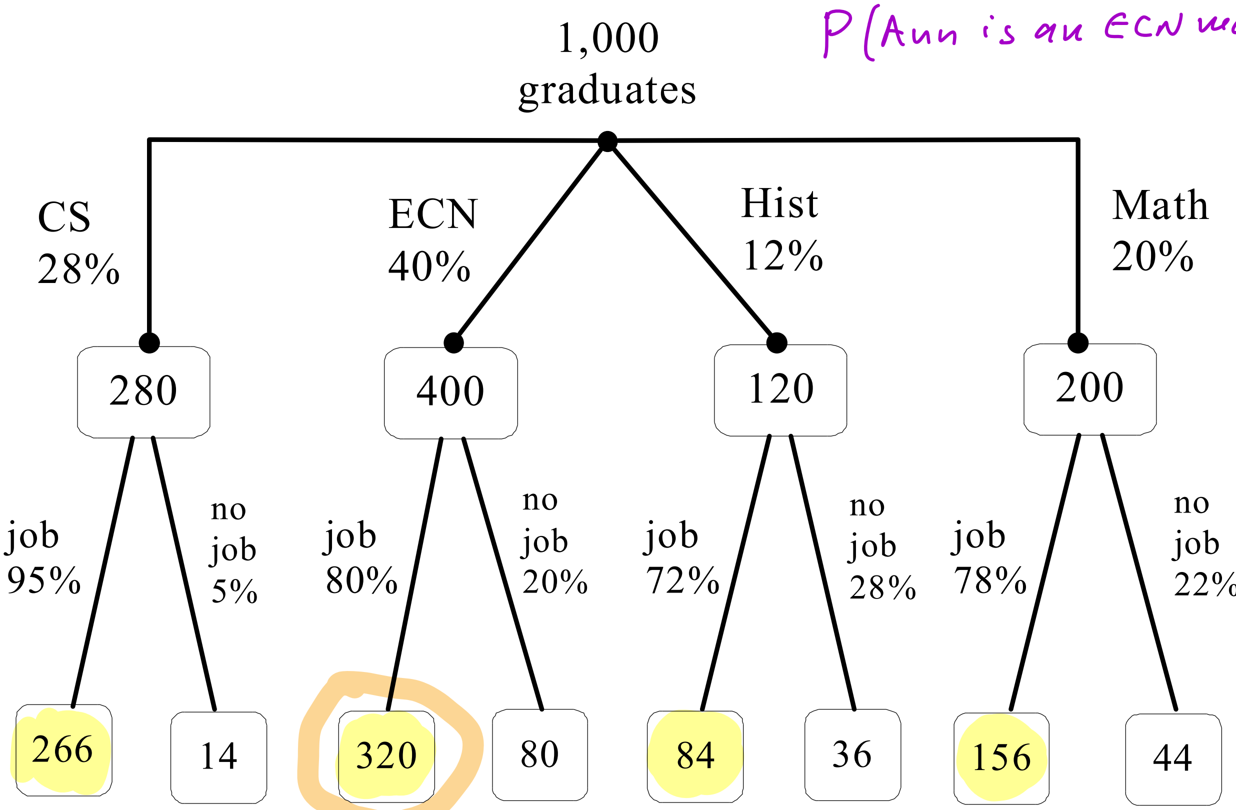
Percentages of those who found a job within 6 months of graduation by major:

Major:	*CS*	*Econ*	*Hist*	*Math*
% who found a job:	95%	80%	70%	78%

You learn that Ann graduated 6 months ago and has already found a job. What is the probability that Ann is an Econ major?

P(Ann is an ECN major | found a job)

base rates



The **probability of Ann being and Econ major, given that she found a job is thus**:

$$\frac{320}{266+320+84+156} = \frac{320}{826} = 0.3874 = 38.74\%$$

ASSUMING NO DOUBLE MAJORS

$E_1 \cap E_2 = \emptyset$

$E_1 \cap E_3 = \emptyset$

$E_1 \cap E_4 \neq \emptyset$

$E_2 \cap E_4 = \emptyset$

$E_3 \cap E_4 = \emptyset$

$E_1 \cup E_2 \cup E_3 \cup E_4 = S$

E_1	E_2	E_3	E_4
CS	*Econ*	*Hist*	*Math*
28%	40%	12%	20%

Major:	*CS*	*Econ*	*Hist*	*Math*
% who found a job:	95%	80%	70%	78%

Now we need a version of Bayes' rule that allows for more than two conditioning events.

Let S be the set of states and $\{E_1, E_2, ..., E_m\}$ be a partition of S, that is,

- $E_1 \cup E_2 \cup ... \cup E_m = S$
- For all $i, j \in \{1, 2, ...m\}$ with $i \neq j$, $E_i \cap E_j = \varnothing$

Let $F \subseteq S$ be an arbitrary event. Then

$F = (F \cap E_1) \cup (F \cap E_2) \cup ... \cup (F \cap E_m)$, all disjoint events. Thus

$$P(F) = \underbrace{P(F \cap E_1)}_{} + \underbrace{P(F \cap E_2)}_{} + ... + \underbrace{P(F \cap E_m)}_{}$$

for $i = 1, \ldots, n$

Hence, $P(E_i \mid F) = \dfrac{P(F \mid E_i)\,P(E_i)}{P(F \mid E_1)\,P(E_1) + P(F \mid E_2)\,P(E_2) + \cdots + P(F \mid E_n)\,P(E_n)}$

$$\begin{array}{cccc} CS & Econ & Hist & Math \\ 28\% & 40\% & 12\% & 20\% \end{array}$$

and, by hypothesis, CS, ECN, HIS, MAT is an exhaustive list of majors:

$$P(CS) = \frac{28}{100}, \quad P(ECN) = \frac{40}{100}, \quad P(HIS) = \frac{12}{100}, \quad P(MAT) = \frac{20}{100}$$

Major:	CS	$Econ$	$Hist$	$Math$
% who found a job:	95%	80%	70%	78%

$$P(J \mid CS) = \frac{95}{100}, \quad P(J \mid ECN) = \frac{80}{100}, \quad P(J \mid HIS) = \frac{70}{100}, \quad P(J \mid MAT) = \frac{78}{100}$$

$$P(\neg J \mid CS) = \frac{5}{100}, \quad P(\neg J \mid ECN) = \frac{20}{100}, \quad P(\neg J \mid HIS) = \frac{30}{100}, \quad P(\neg J \mid MAT) = \frac{12}{100}$$

Thus

$$P(ECN \mid J) = \frac{P(J|ECN)P(ECN)}{P(J|CS)P(CS)+P(J|ECN)P(ECN)+P(J|HIS)P(HIS)+P(J|MAT)P(MAT)}$$

$$= \frac{\frac{80}{100}\times\frac{40}{100}}{\frac{95}{100}\times\frac{28}{100}+\frac{80}{100}\times\frac{40}{100}+\frac{70}{100}\times\frac{12}{100}+\frac{78}{100}\times\frac{20}{100}} = 0.3874 = 38.74\%$$