Conditional probability

Relapse

Researchers randomly assigned 72 chronic users of cocaine into three groups: desipramine (antidepressant), lithium (standard treatment for cocaine) and placebo. Results of the study are summarized below.

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

http://www.oswego.edu/~srp/stats/2_way_tbl_1.htm

Marginal probability

What is the probability that a patient relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

Marginal probability

What is the probability that a patient relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	*48*	24	*72*

P(relapsed) = $\frac{48}{72} \approx 0.67$

Joint probability

What is the probability that a patient received the antidepressant (desipramine) <u>and</u> relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

Joint probability

What is the probability that a patient received the antidepressant (desipramine) and relapsed?

	relapse	no relapse	total
desipramine	*10*	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	*72*

P(relapsed and desipramine) = $\frac{10}{72} \approx 0.14$

Conditional probability

Conditional probability

The conditional probability of the outcome of interest B given condition A is calculated as

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

Conditional probability

Conditional probability

The conditional probability of the outcome of interest B given condition A is calculated as

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

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

$$P(relapse|desipramine) = \frac{P(relapse\ and\ desipramine)}{P(desipramine)}$$

Conditional probability

Conditional probability

The conditional probability of the outcome of interest B given condition A is calculated as

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

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

$$P(relapse|desipramine)$$
$$= \frac{P(relapse\ and\ desipramine)}{P(desipramine)}$$
$$= \frac{10/72}{24/72}$$

Conditional probability

Conditional probability

The conditional probability of the outcome of interest B given condition A is calculated as

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

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

$$\begin{aligned} & P(relapse|desipramine) \\ &= \frac{P(relapse \text{ and } desipramine)}{P(desipramine)} \\ &= \frac{10/72}{24/72} \\ &= \frac{10}{24} \end{aligned}$$

Conditional probability

Conditional probability

The conditional probability of the outcome of interest B given condition A is calculated as

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

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

$$P(relapse|desipramine) = \frac{P(relapse\ and\ desipramine)}{P(desipramine)} = \frac{10/72}{24/72} = \frac{10}{24} = 0.42$$

Conditional probability (cont.)

If we know that a patient received the antidepressant (desipramine), what is the probability that they relapsed?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

Conditional probability (cont.)

If we know that a patient received the antidepressant (desipramine), what is the probability that they relapsed?

	relapse	no relapse	total
desipramine	*10*	14	*24*
lithium	18	6	24
placebo	20	4	24
total	48	24	72

P(relapse | desipramine) = $\frac{10}{24} \approx 0.42$

Conditional probability (cont.)

If we know that a patient received the antidepressant (desipramine), what is the probability that they relapsed?

	relapse	no relapse	total
desipramine	*10*	14	*24*
lithium	18	6	24
placebo	20	4	24
total	48	24	72

P(relapse | desipramine) = $\frac{10}{24} \approx 0.42$

P(relapse | lithium) = $\frac{18}{24} \approx 0.75$
P(relapse | placebo) = $\frac{20}{24} \approx 0.83$

Conditional probability (cont.)

If we know that a patient relapsed, what is the probability that they received the antidepressant (desipramine)?

	relapse	no relapse	total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

Conditional probability (cont.)

If we know that a patient relapsed, what is the probability that they received the antidepressant (desipramine)?

	relapse	no relapse	total
desipramine	*10*	14	24
lithium	18	6	24
placebo	20	4	24
total	*48*	24	72

P(desipramine | relapse) = $\frac{10}{48} \approx 0.21$

Conditional probability (cont.)

If we know that a patient relapsed, what is the probability that they received the antidepressant (desipramine)?

	relapse	no relapse	total
desipramine	*10*	14	24
lithium	18	6	24
placebo	20	4	24
total	*48*	24	72

P(desipramine | relapse) = $\frac{10}{48} \approx 0.21$

P(lithium | relapse) = $\frac{18}{48} \approx 0.375$

P(placebo | relapse) = $\frac{20}{48} \approx 0.42$

General multiplication rule

- Earlier we saw that if two events are independent, their joint probability is simply the product of their probabilities. If the events are not believed to be independent, the joint probability is calculated slightly differently.

General multiplication rule

- Earlier we saw that if two events are independent, their joint probability is simply the product of their probabilities. If the events are not believed to be independent, the joint probability is calculated slightly differently.
- If A and B represent two outcomes or events, then

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

Note that this formula is simply the conditional probability formula, rearranged.

General multiplication rule

- Earlier we saw that if two events are independent, their joint probability is simply the product of their probabilities. If the events are not believed to be independent, the joint probability is calculated slightly differently.
- If A and B represent two outcomes or events, then

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

  Note that this formula is simply the conditional probability formula, rearranged.
- It is useful to think of B as the outcome of interest and A as the condition.

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is $\frac{60}{100} = 0.6$.

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is $\frac{60}{100} = 0.6$.
- The probability that a randomly selected student is a social science major given that they are female is

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is $\frac{60}{100} = 0.6$.
- The probability that a randomly selected student is a social science major given that they are female is $\frac{30}{50} = 0.6$.

Independence and conditional probabilities

Consider the following (hypothetical) distribution of gender and major of students in an introductory statistics class:

	social science	non-social science	total
female	30	20	50
male	30	20	50
total	60	40	100

- The probability that a randomly selected student is a social science major is $\frac{60}{100} = 0.6$.
- The probability that a randomly selected student is a social science major given that they are female is $\frac{30}{50} = 0.6$.
- Since $P(SS|M)$ also equals 0.6, major of students in this class does not depend on their gender: P(SS | F) = P(SS).

Independence and conditional probabilities (cont.)

Generically, if $P(A|B) = P(A)$ then the events A and B are said to be independent.

Independence and conditional probabilities (cont.)

Generically, if $P(A|B) = P(A)$ then the events A and B are said to be independent.

- Conceptually: Giving B doesn't tell us anything about A.

Independence and conditional probabilities (cont.)

Generically, if $P(A|B) = P(A)$ then the events A and B are said to be independent.

- Conceptually: Giving B doesn't tell us anything about A.
- Mathematically: We know that if events A and B are independent, $P(A \text{ AND } B) = P(A) \times P(B)$. Then,

$$P(A|B) = \frac{P(A \text{ AND } B)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A)$$

Bayes' Theorem

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

Example of using Bayes' Theorem

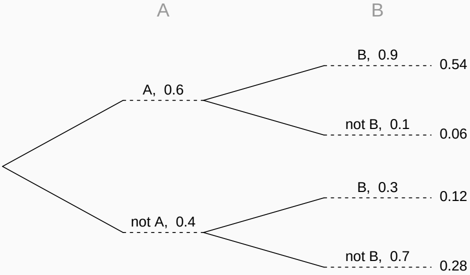
- You are told $P(A) = 0.6$, $P(B|A) = 0.9$, and $P(B|A^c) = 0.3$.

Example of using Bayes' Theorem

- You are told $P(A) = 0.6$, $P(B|A) = 0.9$, and $P(B|A^c) = 0.3$.
- Determine $P(A|B)$.

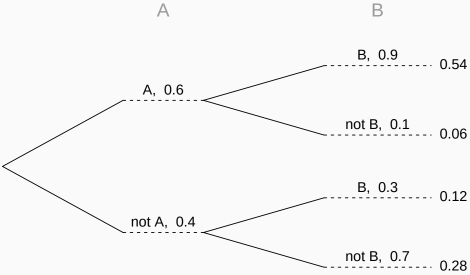
Example of using Bayes' Theorem

- You are told $P(A) = 0.6$, $P(B|A) = 0.9$, and $P(B|A^c) = 0.3$.
- Determine $P(A|B)$.



Example of using Bayes' Theorem

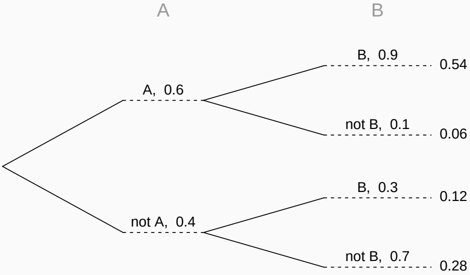
- You are told $P(A) = 0.6$, $P(B|A) = 0.9$, and $P(B|A^c) = 0.3$.
- Determine $P(A|B)$.



	A	A^c	*Total*
B	*0.54*	*0.12*	*0.66*
B^c	*0.06*	*0.28*	*0.34*
Total	*0.6*	*0.4*	*1*

Example of using Bayes' Theorem

- You are told $P(A) = 0.6$, $P(B|A) = 0.9$, and $P(B|A^c) = 0.3$.
- Determine $P(A|B)$.

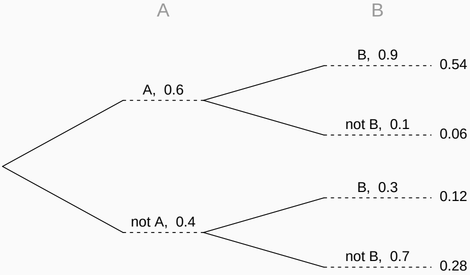


	A	A^c	*Total*
B	*0.54*	*0.12*	*0.66*
B^c	*0.06*	*0.28*	*0.34*
Total	*0.6*	*0.4*	*1*

$P(A|B) =$

Example of using Bayes' Theorem

- You are told $P(A) = 0.6$, $P(B|A) = 0.9$, and $P(B|A^c) = 0.3$.
- Determine $P(A|B)$.



	A	A^c	*Total*
B	*0.54*	*0.12*	*0.66*
B^c	*0.06*	*0.28*	*0.34*
Total	*0.6*	*0.4*	*1*

$$P(A|B) = \frac{0.54}{0.54+0.12} = \frac{0.54}{0.66}$$
$$\approx 0.818$$

Breast cancer screening

- American Cancer Society estimates that about 1.7% of women have breast cancer.

  http://www.cancer.org/cancer/cancerbasics/cancer-prevalence
- Susan G. Komen For The Cure Foundation states that mammography correctly identifies about 78% of women who truly have breast cancer.

  http://ww5.komen.org/BreastCancer/AccuracyofMammograms.html
- An article published in 2003 suggests that up to 10% of all mammograms result in false positives for patients who do not have cancer.

  http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1360940

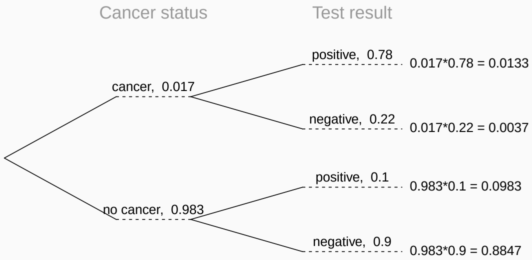
Note: These percentages are approximate, and very difficult to estimate.

Inverting probabilities

When a patient goes through breast cancer screening there are two competing claims: patient **has cancer** and patient **doesn't have cancer**. If a mammogram yields a positive result, what is the probability that patient actually has cancer?

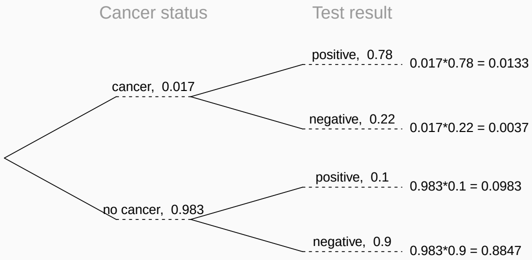
Inverting probabilities

When a patient goes through breast cancer screening there are two competing claims: patient **has cancer** and patient **doesn't have cancer**. If a mammogram yields a positive result, what is the probability that patient actually has cancer?



Inverting probabilities

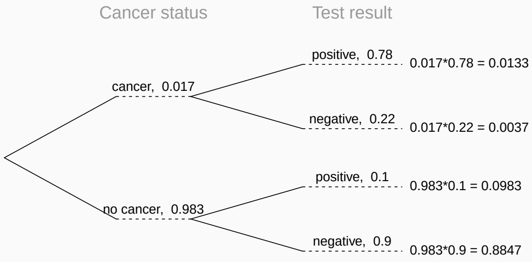
When a patient goes through breast cancer screening there are two competing claims: patient **has cancer** and patient **doesn't have cancer**. If a mammogram yields a positive result, what is the probability that patient actually has cancer?



$$P(C|+) \quad =$$

Inverting probabilities

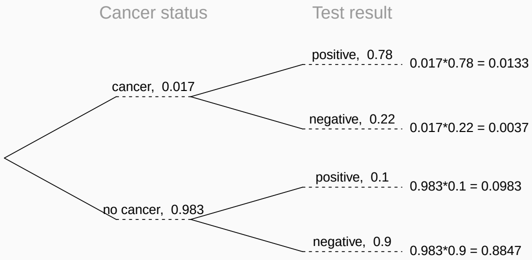
When a patient goes through breast cancer screening there are two competing claims: patient **has cancer** and patient **doesn't have cancer**. If a mammogram yields a positive result, what is the probability that patient actually has cancer?



$$P(C|+) = \frac{P(C \text{ AND } +)}{P(+)}$$

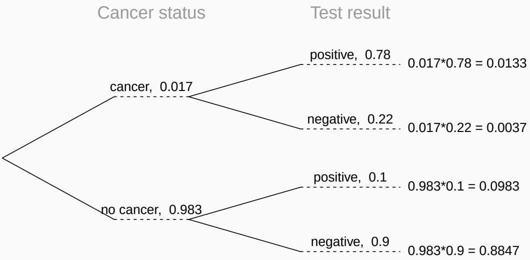
Inverting probabilities

When a patient goes through breast cancer screening there are two competing claims: patient **has cancer** and patient **doesn't have cancer**. If a mammogram yields a positive result, what is the probability that patient actually has cancer?



$$P(C|+) = \frac{P(C \text{ AND } +)}{P(+)} = \frac{0.0133}{0.0133 + 0.0983}$$

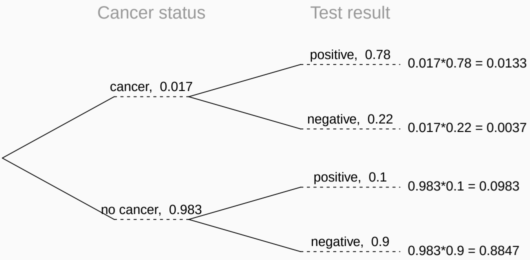
When a patient goes through breast cancer screening there are two competing claims: patient **has cancer** and patient **doesn't have cancer**. If a mammogram yields a positive result, what is the probability that patient actually has cancer?



$$
\begin{aligned}
P(C|+) &= \frac{P(C \text{ AND } +)}{P(+)} \\
&= \frac{0.0133}{0.0133 + 0.0983} \\
&= 0.12
\end{aligned}
$$

Inverting probabilities

When a patient goes through breast cancer screening there are two competing claims: patient **has cancer** and patient **doesn't have cancer**. If a mammogram yields a positive result, what is the probability that patient actually has cancer?



$$
\begin{aligned}
P(C|+) &= \frac{P(C \text{ AND } +)}{P(+)} \\
&= \frac{0.0133}{0.0133 + 0.0983} \\
&= 0.12
\end{aligned}
$$

Note: Tree diagrams are useful for inverting probabilities: we are given $P(+|C)$ and asked for $P(C|+)$.

Practice

Suppose a woman who gets tested once and obtains a positive result wants to get tested again. In the second test, what should we assume to be the probability of this specific woman having cancer?

(a) 0.017

(b) 0.12

(c) 0.0133

(d) 0.88

Practice

Suppose a woman who gets tested once and obtains a positive result wants to get tested again. In the second test, what should we assume to be the probability of this specific woman having cancer?

(a) 0.017

(b) *0.12*

(c) 0.0133

(d) 0.88

Practice

What is the probability that this woman has cancer if this second mammogram also yielded a positive result?

(a) 0.0936

(b) 0.088

(c) 0.48

(d) 0.52

Practice

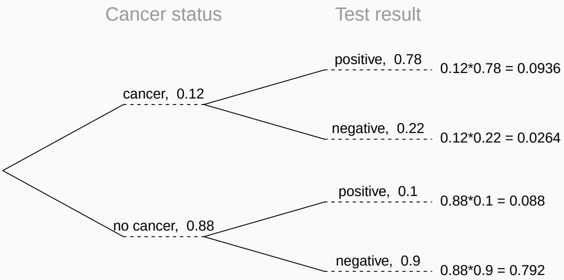
What is the probability that this woman has cancer if this second mammogram also yielded a positive result?

(a) 0.0936

(b) 0.088

(c) 0.48

(d) *0.52*



Practice

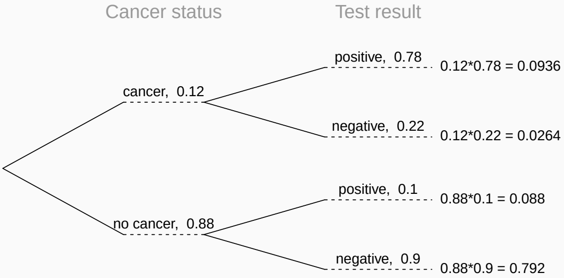
What is the probability that this woman has cancer if this second mammogram also yielded a positive result?

(a) 0.0936

(b) 0.088

(c) 0.48

(d) *0.52*



$$P(C|+) = \frac{P(C \text{ } and \text{ } +)}{P(+)} = \frac{0.0936}{0.0936 + 0.088} = 0.52$$

Bayes' Theorem

- The conditional probability formula we have seen so far is a special case of the Bayes' Theorem, which is applicable even when events have more than just two outcomes.

Bayes' Theorem

- The conditional probability formula we have seen so far is a special case of the Bayes' Theorem, which is applicable even when events have more than just two outcomes.
- *Bayes' Theorem:*

$$P(outcome\ A_1\ of\ variable\ 1 \mid outcome\ B\ of\ variable\ 2)$$
$$= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)}$$

where $A_2, \cdots, A_k$ represent all other possible outcomes of variable 1.

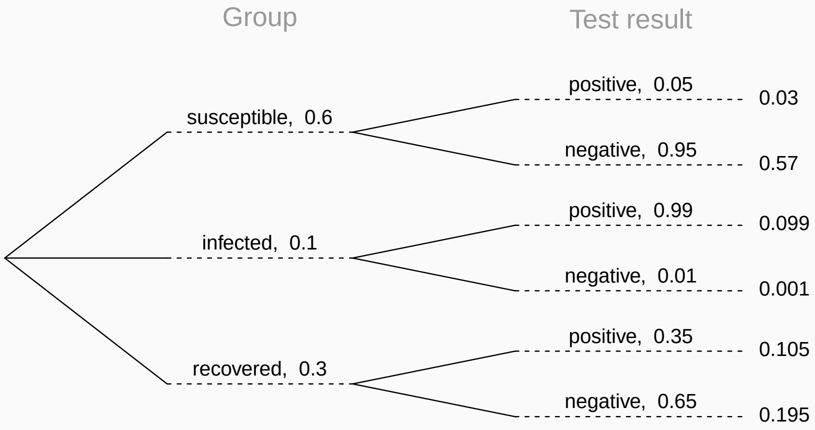
Application activity: Inverting probabilities

A common epidemiological model for the spread of diseases is the SIR model, where the population is partitioned into three groups: Susceptible, Infected, and Recovered. This is a reasonable model for diseases like chickenpox where a single infection usually provides immunity to subsequent infections. Sometimes these diseases can also be difficult to detect.

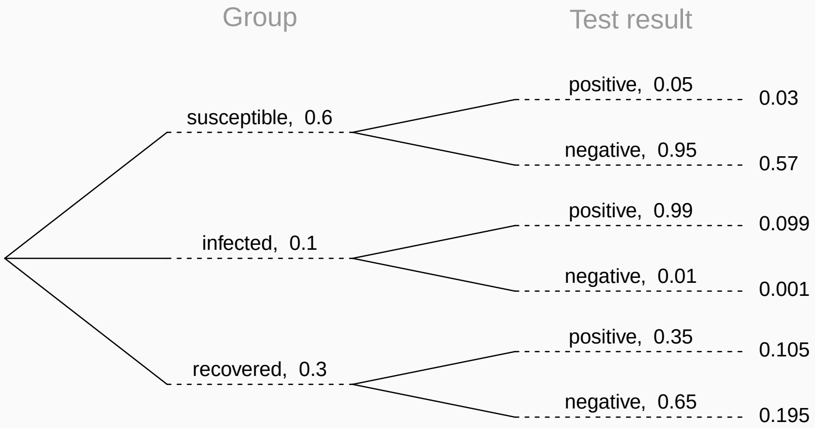
Imagine a population in the midst of an epidemic where 60% of the population is considered susceptible, 10% is infected, and 30% is recovered. The only test for the disease is accurate 95% of the time for susceptible individuals, 99% for infected individuals, but 65% for recovered individuals. (Note: In this case accurate means returning a negative result for susceptible and recovered individuals and a positive result for infected individuals).

Draw a probability tree to reflect the information given above. If the individual has tested positive, what is the probability that they are actually infected?

Application activity: Inverting probabilities (cont.)



Application activity: Inverting probabilities (cont.)



$$P(inf|+) = \frac{P(inf\ and\ +)}{P(+)} = \frac{0.099}{0.03 + 0.099 + 0.105} \approx 0.423$$