ECO239

Week 7

Probability (part 2)

Conditional probability

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

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

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

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

Conditional probability

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

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

	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

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

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

	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

P(relapse | desipramine) = 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

P(relapse | desipramine) = 10 / 24 ~ 0.42

P(relapse | lithium) = 18 / 24 ~ 0.75

P(relapse | placebo) = 20 / 24 ~ 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) = 10 / 48 ~ 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) = 10 / 48 ~ 0.21

P(lithium | relapse) = 18 / 48 ~ 0.38

P(placebo | relapse) = 20 / 48 ~ 0.42

Practice

- Of the PCs in the computer lab,
- 40% is touch screen (A)
- 70% has camera (B)
- 20% has both.

Q1: What’s the probability that a PC is touch screen given that it has camera?

Q2. What’s the probability that the PC has camera given that it is touch screen?

Q3. What’s the probability that a PC does not have either?

Q1: What's the probability that a PC is touch screen (A) given that it has camera (B)?

$P(A|B)= P(A \cap B)/P(B)=0.2/0.7 = 0.286$

Q2. What's the probability that the PC has camera (B)given that it is touch screen(A)?

$P(B|A)= P(B \cap A)/P(A)=0.2/0.4 = 0.5$

Q3. What's the probability that a PC does not have either?

$1-[P(A)+P(B)-P(A \cap B)]=1-[0.4+0.7-0.2]=0.1$

Multiplication Rule

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

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

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

$P(A \cap B)$: joint probability
$P(A|B)$: conditional probability
$P(A)$: marginal probability

Practice: $P(A \cap B)=P(A|B)P(B)=P(B|A)P(A)$

- Confirm the rule by using previous example
- 40% touch screen (A)
- 70% camera (B)
- 20% both

$P(A \cap B)=P(A|B)P(B)=0.286*0.7=0.2$

$P(A \cap B)=P(B|A)P(A)=0.5*0.4=0.2$

Independence

Two processes are *independent* if knowing the outcome of one provides no useful information about the outcome of the other.

$$\boxed{P(A \cap B) = P(A)P(B)}$$

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

Independence

Two processes are *independent* if knowing the outcome of one provides no useful information about the outcome of the other.

- Knowing that the coin landed on a head on the first toss does not provide any useful information for determining what the coin will land on in the second toss.
  >> Outcomes of two tosses of a coin are independent.

Independence

Two processes are *independent* if knowing the outcome of one provides no useful information about the outcome of the other.

- Knowing that the coin landed on a head on the first toss does not provide any useful information for determining what the coin will land on in the second toss.
  >> Outcomes of two tosses of a coin are independent.
- Knowing that the first card drawn from a deck is an ace does provide useful information for determining the probability of drawing an ace in the second draw.
  >> Outcomes of two draws from a deck of cards (without replacement) are dependent.

Practice

- Of the PCs in the computer lab,
- 40% is touch screen. (A)
- 70% has camera. (B)
- 20% has both.

Are event A and B statistically independent?

$P(A \cap B) = P(A)P(B)$???
$P(A) = 0.4$, $P(B) = 0.7$, $P(A)P(B)=0.28$
$\neq 0.2 = P(A \cap B)$ => Not independent.

Practice

Between January 9-12, 2013, SurveyUSA interviewed a random sample of 500 NC residents asking them whether they think widespread gun ownership protects law abiding citizens from crime, or makes society more dangerous. 58% of all respondents said it protects citizens. 67% of White respondents, 28% of Black respondents, and 64% of Hispanic respondents shared this view (Consider these are conditional probabilities (P(protect|Race)). Which of the below is true?

Opinion on gun ownership and race ethnicity are most likely
(a) complementary
(b) mutually exclusive
(c) independent
(d) dependent

http://www.surveyusa.com/client/PollReport.aspx?g=a5f460ef-bba9-484b-8579-1101ea26421b

Hint

- P(protect) = 0.58
- P(protect|White) = 0.67
- P(protect|Black) = 0.28
- P(protect|Hispanic)=0.64

P(protect|White) = P(protect) ???
P(protect|Black) = P(protect)???
P(protect|Hispanic)=P(protect)???

Practice

Between January 9-12, 2013, SurveyUSA interviewed a random sample of 500 NC residents asking them whether they think widespread gun ownership protects law abiding citizens from crime, or makes society more dangerous. 58% of all respondents said it protects citizens. 67% of White respondents, 28% of Black respondents, and 64% of Hispanic respondents shared this view. Which of the below is true?

Opinion on gun ownership and race ethnicity are most likely
(a) complementary
(b) mutually exclusive
(c) independent
(d) dependent

http://www.surveyusa.com/client/PollReport.aspx?g=a5f460ef-bba9-484b-8579-1101ea26421b

Product rule for independent events

P(A and B) = P(A) x P(B)

Or more generally, $P(A_1$, and, … and $A_k) = P(A_1) \times \ldots \times P(A_k)$

Product rule for independent events

P(A and B) = P(A) x P(B)

Or more generally, $P(A_1$, and, … and $A_k) = P(A_1) \times \ldots \times P(A_k)$

You toss a coin twice, what is the probability of getting two tails in a row?

Product rule for independent events

P(A and B) = P(A) x P(B)

Or more generally, $P(A_1$, and, … and $A_k) = P(A_1)$ x … x $P(A_k)$

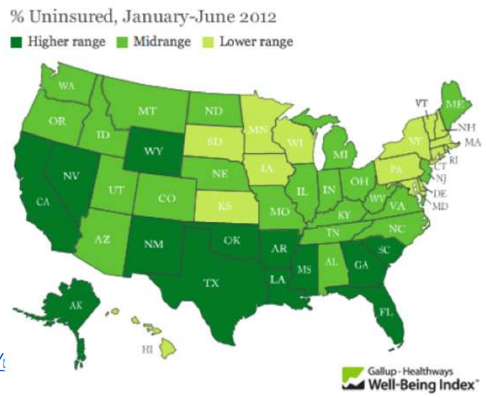
You toss a coin twice, what is the probability of getting two tails in a row?

P(T on the first toss) x P(T on the second toss) = ½ x ½ =¼

Practice

A recent Gallup poll suggests that 25.5% of Texans do not have health insurance as of June 2012. Assuming that the uninsured rate stayed constant, what is the probability that two randomly selected Texans are both uninsured?

(a) 25.5^2
(b) 0.255^2
(c) 0.255 x 2
(d) $(1 - 0.255)^2$

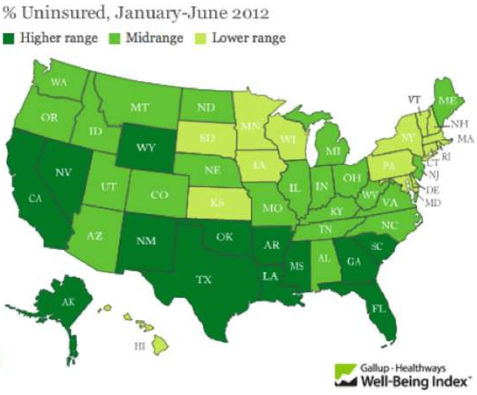


http://www.gallup.com/poll/156851/

Practice

A recent Gallup poll suggests that 25.5% of Texans do not have health insurance as of June 2012. Assuming that the uninsured rate stayed constant, what is the probability that two randomly selected Texans are both uninsured?

(a) 25.5^2
(b) 0.255^2
(c) 0.255×2
(d) $(1 - 0.255)^2$



http://www.gallup.com/poll/156851/

Practice

Roughly 20% of undergraduates at a university are vegetarian or vegan. What is the probability that, among a random sample of 3 undergraduates, at least one is vegetarian or vegan?

(a) $1 - 0.2 \times 3$
(b) $1 - 0.2^3$
(c) 0.8^3
(d) $1 - 0.8 \times 3$
(e) $1 - 0.8^3$

Practice

Roughly 20% of undergraduates at a university are vegetarian or vegan. What is the probability that, among a random sample of 3 undergraduates, at least one is vegetarian or vegan?

(a) $1 - 0.2 \times 3$
(b) $1 - 0.2^3$
(c) 0.8^3
(d) $1 - 0.8 \times 3$
(e) $1 - 0.8^3$

P(at least 1 from veg)
= 1 - P(none veg)
= $1 - 0.8^3$
= 1 - 0.512 = 0.488

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 and B) = P(A | B) x P(B)

  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 and B) = P(A | B) x P(B)

  Note that this formula is simply the conditional probability formula, rearranged.
- It is useful to think of A as the outcome of interest and B 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 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 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 60 / 100 = 0.6.
- The probability that a randomly selected student is a social science major given that they are female is 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 60 / 100 = 0.6.
- The probability that a randomly selected student is a social science major given that they are female is 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 $P(A \mid B) = P(A)$ then the events A and B are said to be independent.

Independence and conditional probabilities (cont.)

Generically, if $P(A \mid B) = P(A)$ 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 and B) = P(A) x P(B)*. Then,

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

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*

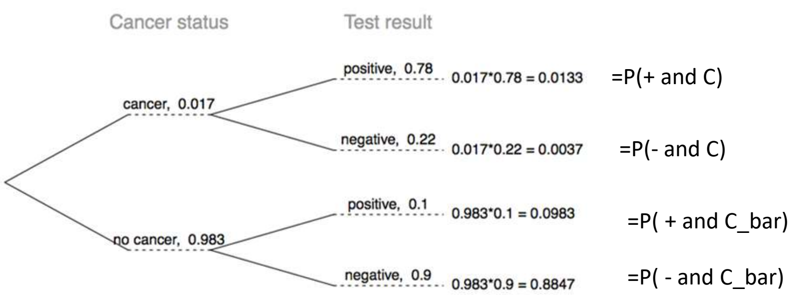
Note: These percentages are approximate, and very difficult to estimate.

Inverting probabilities

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

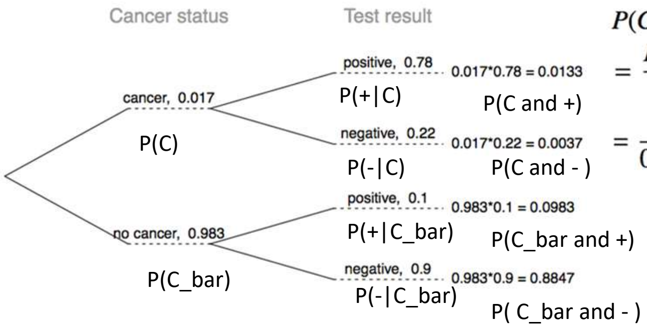
Inverting probabilities

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



$$P(C|+) = \frac{P(C \text{ and } +)}{P(+)} = \frac{0.0133}{0.0133 + 0.0983} = 0.12$$

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

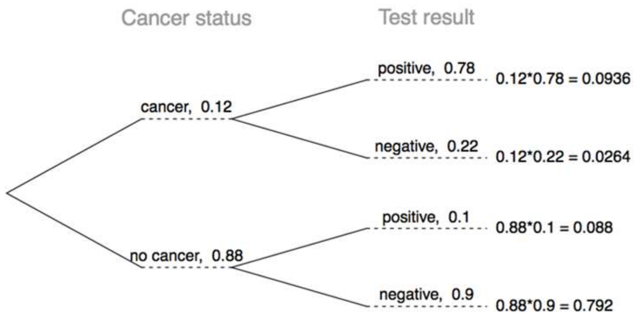
What is the probability that this woman has cancer if this second mammogram also yielded a positive result?

(a) 0.0936

(b) 0.088

(c) 0.48

(d) 0.52



Practice

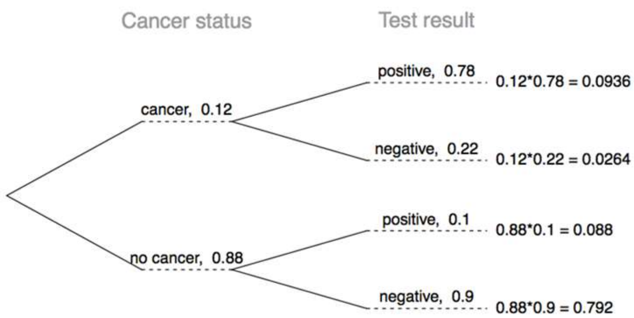
What is the probability that this woman has cancer if this second mammogram also yielded a positive result?

(a) 0.0936

(b) 0.088

(c) 0.48

(d) 0.52



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

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

$$P(\text{outcome } A \text{ of variable } 1 \mid \text{outcome } B \text{ of variable } 2)$$

$$= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)}$$

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

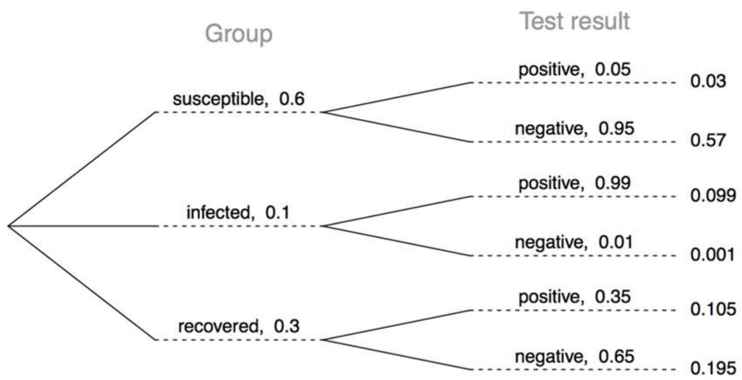
Application activity: inverting probabilities

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.

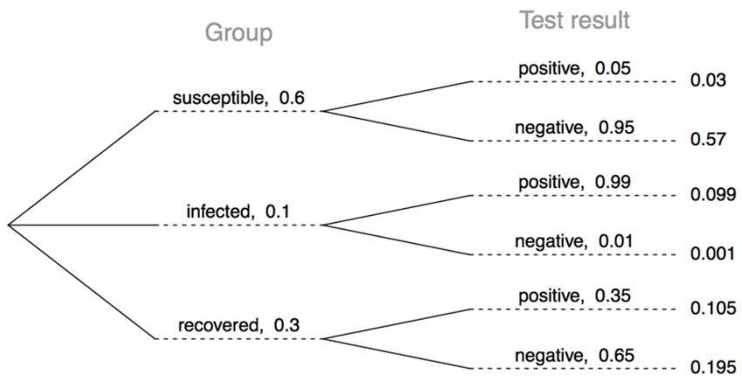
Imagine a population in the midst of an epidemic where 60% of the population is considered susceptible, 10% is infected, and 30% is recovered. The only test for the disease is accurate 95% of the time for susceptible individuals, 99% for infected individuals, but 65% for recovered individuals. (Note: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



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