ECO239

Week 7
Probability (2)

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 $P(A_k)$

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?

Product rule for independent events

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

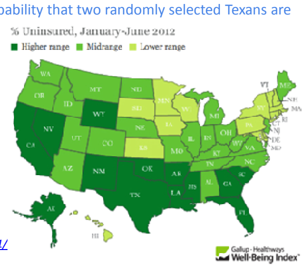
Or more generally, $P(A_1$, and, … and $A_k) = P(A_1)$ x … x $P(A_k)$

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?

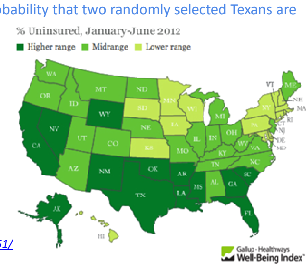


(a) 25.5^2
(b) 0.255^2
(c) 0.255 x 2
(d) $(1 - 0.255)^2$

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 x 2
(d) $(1 - 0.255)^2$

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

Practice

- If we were to randomly select 5 Texans, what is the probability that at least one is uninsured?

(25.5% of Texans do not have health insurance)

Putting everything together...

If we were to randomly select 5 Texans, what is the probability that at least one is uninsured?}

- If we were to randomly select 5 Texans, the sample space for the number of Texans who are uninsured would be:
  S = {0, 1, 2, 3, 4, 5}
- We are interested in instances where at least one person is uninsured:
  S = {0, 1, 2, 3, 4, 5}
- So we can divide up the sample space into two categories:
  S = {0, at least one}

Putting everything together...

Since the probability of the sample space must add up to 1:
P(at least 1 uninsured)
= 1 - *P(none insured)*

Putting everything together...

Since the probability of the sample space must add up to 1:
P(at least 1 uninsured)
= 1 - *P(none insured)*

$= 1 - (1 - 0.255)^5$

Putting everything together...

Since the probability of the sample space must add up to 1:
P(at least 1 uninsured)
= 1 - *P(none insured)*

$= 1 - (1 - 0.255)^5$
$= 1 - 0.745^5$

Putting everything together...

Since the probability of the sample space must add up to 1:
P(at least 1 uninsured)
= 1 - *P(none insured)*

$= 1 - (1 - 0.255)^5$
$= 1 - 0.745^5$
$= 1 - 0.23$

Putting everything together...

Since the probability of the sample space must add up to 1:

P(at least 1 uninsured)

= 1 - *P(none insured)*

= 1 - (1 - 0.255)^5

= 1 - 0.745^5

= 1 - 0.23

= 0.77

Putting everything together...

Since the probability of the sample space must add up to 1:

P(at least 1 uninsured)

= 1 - *P(none insured)*

= 1 - (1 - 0.255)^5

= 1 - 0.745^5

= 1 - 0.23

= 0.77

At least 1:

P(at least one) = 1 - *P(none)*

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 x 3
(b) 1 - 0.2^3
(c) 0.8^3
(d) 1 - 0.8 x 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 x 3
(b) 1 - 0.2^3
(c) 0.8^3
(d) 1 - 0.8 x 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 | 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.

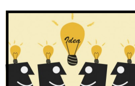
- Conceptually: Giving *B* doesn't tell us anything about *A*.
- Mathematically: We know that if events *A* and *B* are independent, *P(A and B) = P(A) x P(B)*. Then,

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

Joint and Marginal Probability

- Joint Probability: P(A∩B)=
(#outcomes satisfying A and B)/(total $ of elementary outcomes)

- Marginal Probability:
P(A) = P(A ∩B1)+P(A ∩B2)+…+P(A ∩Bk)
P(B) = P(A1 ∩B)+P(A2 ∩B)+…+P(Ah ∩B)
Where Ai and Bj are mutually exclusive and collectively exhaustive.



Practice

Event A: Frequency of watching "Big Bang Theory" {Regularly, Occasionally, Never}
Event B: education status {Low, Middle, High}

	High	Middle	Low	
Regularly	P(H∩R) =0.04	P(M∩R) =0.13	P(L∩R) =0.04	P(R) =
Occasionally	P(H∩O) =0.10	P(M∩O) =0.11	P(L∩O) =0.06	P(O) =
Never	P(H∩N) =0.13	P(M∩N) =0.17	P(L∩N) =0.22	P(N) =
	P(H) =	P(M) =	P(L) =	1

	High	Middle	Low	
Regularly	P(H∩R) =0.04	P(M∩R) =0.13	P(L∩R) =0.04	P(R) =0.21
Occasionally	P(H∩O) =0.10	P(M∩O) =0.11	P(L∩O) =0.06	P(O) =0.27
Never	P(H∩N) =0.13	P(M∩N) =0.17	P(L∩N) =0.22	P(N) =0.52
	P(H) =0.27	P(M) =0.41	P(L) =0.32	1

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.

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

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

=> Marginal probability

- P(B)= P(B∩A1)+P(B ∩A2)…+P(B ∩Ak)

=>P(B)=P(B|A1)P(A1)+P(B|A2)P(A2)…+P(B|Ak)P(Ak)

=> P(A|B)

$$= \frac{P(B|A)P(A)}{P(B|A1)P(A1) + P(B|A2)P(A2) + \cdots + P(B|Ak)P(Ak)}$$

Bayes' Theorem

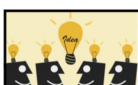
The conditional probability formula we have seen so far is a special case of the Bayes' Theorem, which is applicable even when events have more than just two outcomes.

Bayes' Theorem

$$P(outcome\ A\ of\ variable\ 1\ |\ outcome\ B\ of\ variable\ 2)$$

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

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



Practice

- Suppose 1% of women aged 40-50 have breast cancer. 80% of mammograms detect breast cancer given the cancer exists. 9% chance of false positive (tested positive given no cancer).

Q1. What is the probability a woman has breast cancer given that she just had a positive result?

Q2: What is the probability of the false test results?

HINT:
event A: a positive test result,
event B: the women has breast cancer.

More HINT

- We are looking for… P(B|A) or P(cancer|positive)
- What we are given are:
- P(B)=0.01
- P(A|B)=0.8
- P(A|B_bar)=0.09

P(B|A)= {P(A|B)P(B)}/P(A) = 0.8*0.01/P(A)=0.008/0.0971=0.0824

P(A)=P(A|B)P(B)+P(A|B_bar)P(B_bar)
=0.8*0.01+0.09*0.99=0.0971

Q2

- False Test Results

Case1: Given negative result, there is cancer.
Case2: Given positive result, there is no cancer.

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

$$= \frac{0.2 * 0.01}{1 - 0.0971} + \frac{0.09 * 0.99}{0.0971}$$

$$= 0.0022 + 0.9176 = 0.9198$$

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

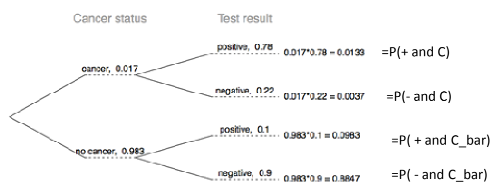
Note: These percentages are approximate, and very difficult to estimate.

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?

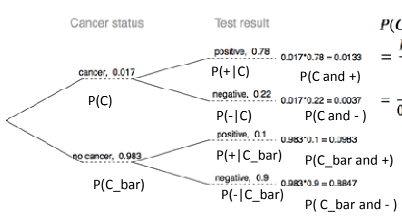
Inverting probabilities

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

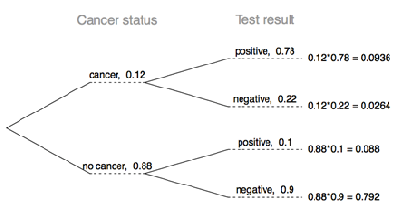
What is the probability that this woman has cancer if this second mammogram also yielded a positive result?

(a) 0.0936

(b) 0.088

(c) 0.48

(d) 0.52



Practice

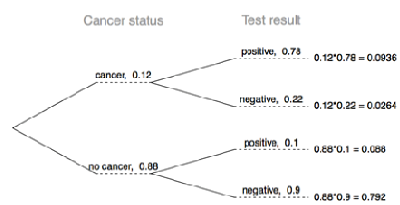
What is the probability that this woman has cancer if this second mammogram also yielded a positive result?

(a) 0.0936

(b) 0.088

(c) 0.48

(d) 0.52



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

Quiz 5 (Nov.15,2016)

Event A: 30% of students subscribing "Economists".
Event B: 45% of students subscribing "Newsweek".
10% of students subscribing both magazines.

Q1: What is the probability that a randomly chosen student subscribes "Economists" given that she also subscribes "Newsweek"?

Q2: What is the probability that a student does not subscribe either?

Q3. Are event A and event B statistically independent? Explain by using conditional probability.

Quiz 5_2 (Nov.15,2016)

- Consider a standard deck of 52 cards with four suits.
- Event A: the card is with figures (J,Q,K)
- Event B: the card is a black suit

Q1: Construct the probability table and calculate each probability.

Q2: Find the probability that the card is with a figure given that the card is a red suit.