Bayes' Theorem (2.4)

Prof. Tesler

Math 186
January 14, 2008

First scenario: rain and cancelling classes

Example

- There's a 30% chance of rain tomorrow.
- **If it rains,** there's a 10% chance classes will be cancelled.
- **If it doesn't rain,** there's only a 2% chance classes will be cancelled.

Questions

1. What's the probability it will rain and classes will be cancelled?
2. What's the total probability (whether or not it rains) that classes will be cancelled?
3. If classes are indeed cancelled tomorrow, what is the probability that it was due to rain?

Probability it will rain and classes will be cancelled?

- There's a 30% chance of rain tomorrow.
- **If it rains,** there's a 10% chance classes will be cancelled.
- **If it doesn't rain,** there's only a 2% chance classes will be cancelled.

Express the data using event notation

- Event A = rains tomorrow

  $$P(A) = .30 \qquad P(A^c) = .70$$

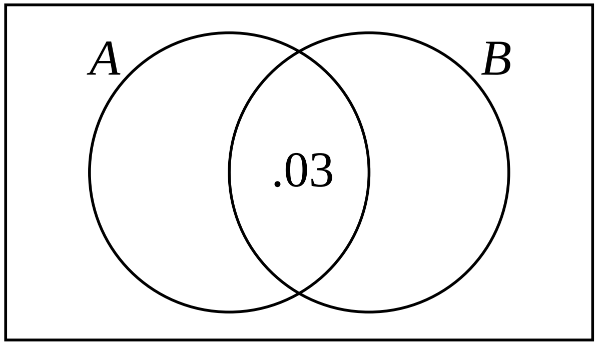
- Event B = classes are cancelled tomorrow

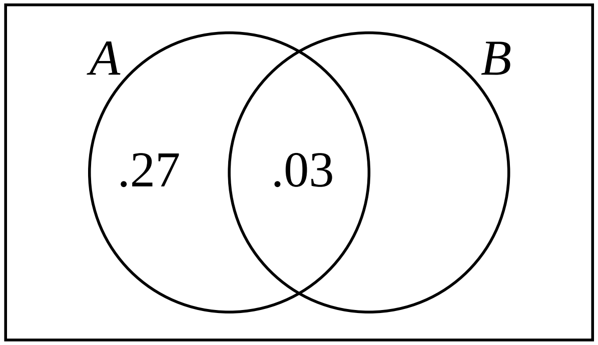
  $$P(B|A) = .10 \qquad P(B|A^c) = .02$$

Probability it will rain and classes will be cancelled?

- Events: $A =$ rains tomorrow, $B =$ classes cancelled tomorrow.
- There's a 30% chance of rain tomorrow. $P(A) = .30, \quad P(A^c) = .70$
- **If it rains,** there's a 10% chance classes will be cancelled.
- **If it doesn't rain,** there's a 2% chance classes will be cancelled. $P(B|A) = .10, \quad P(B|A^c) = .02$

- Express the question using event notation: $P(A \cap B) = ?$
- Note $P(A|B) = \frac{P(A \cap B)}{P(B)}$ gives $P(A \cap B) = P(A|B)P(B)$.
  Similarly, $P(A \cap B) = P(B|A)P(A)$.
- So $P(A \cap B) = P(B|A)P(A) = (.10)(.30) = \boxed{\mathbf{.03}} = \boxed{\mathbf{3\%}}$.

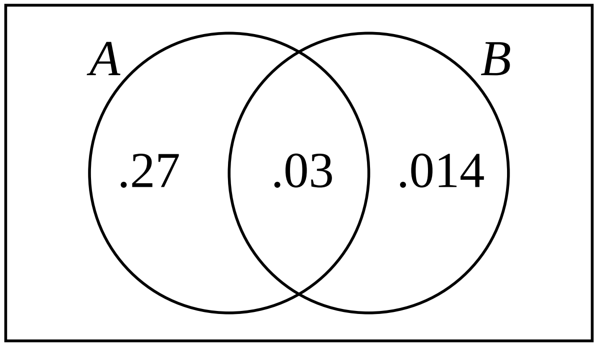


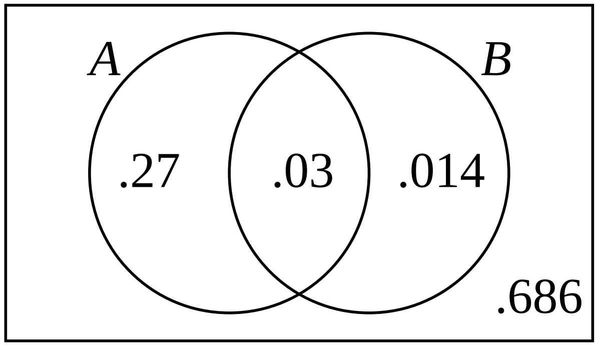


Total probability classes will be cancelled?

- Events: $A =$ rains tomorrow, $B =$ classes cancelled tomorrow.
- There's a 30% chance of rain tomorrow. $P(A) = .30, \quad P(A^c) = .70$
- **If it rains,** there's a 10% chance classes will be cancelled.
- **If it doesn't rain,** there's 2% chance classes will be cancelled.
  $P(B|A) = .10, \quad P(B|A^c) = .02$

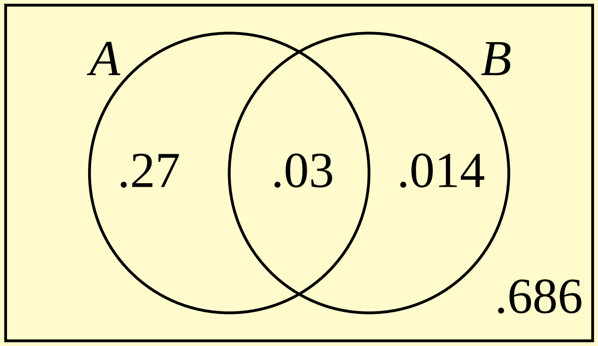
- Express the question using event notation: $P(B) = ?$
- $$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap A^c) \\ &= P(B|A)P(A) + P(B|A^c)P(A^c) \\ &= (.10)(.30) + (.02)(.70) = .03 + .014 = \boxed{\mathbf{.044}} = \boxed{\mathbf{4.4\%}} \end{aligned}$$





If classes are cancelled, what's the probability it was due to rain?

- Events: $A =$ rains tomorrow, $B =$ classes cancelled tomorrow.
- There's a 30% chance of rain tomorrow. $P(A) = .30, \quad P(A^c) = .70$
- **If it rains,** there's a 10% chance classes will be cancelled.
- **If it doesn't rain,** there's 2% chance classes will be cancelled.
  $P(B|A) = .10, \quad P(B|A^c) = .02$



- Express the question using event notation: $P(A|B) = ?$
- $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(B|A)P(A)}{P(B)} = \dfrac{(.10)(.30)}{.044}$ $\boxed{= \mathbf{.68\overline{18} \approx 68.2\%}}$

Bayes' Theorem (simple version)

Theorem (Bayes' Theorem)

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

This lets us express the probability of A given B, in terms of the probability of B given A.

Alternate formulation of Bayes' Theorem

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

where we used

$$P(B) = P(B \cap A) + P(B \cap A^c) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Second scenario: location of genes on chromosomes

Length of chromosomes & fraction in genes

Chromosome #	Length (nt)	Prob. gene @ each nt
1	1,000,000	0.05
2	2,000,000	0.06
3	3,000,000	0.07
4	4,000,000	0.08
Total	**10,000,000**	

Sample space, events, and probabilities

- S = sample space = all positions
  $N(S)$ = 10,000,000
- A_i = positions on chromosome i
  $P(A_1) = 1{,}000{,}000/10{,}000{,}000 = .1$
  $P(A_2) = 2/10 = .2 \quad P(A_3) = 3/10 = .3 \quad P(A_4) = 4/10 = .4$
- B = positions in genes
  $P(B|A_1) = .05 \quad P(B|A_2) = .06 \quad P(B|A_3) = .07 \quad P(B|A_4) = .08$

Breaking down the probabilities of events

Sample space, events, and probabilities

- A_i = positions on chromosome i
  $P(A_1) = .1 \quad P(A_2) = .2 \quad P(A_3) = .3 \quad P(A_4) = .4$
- B = positions in genes
  $P(B|A_1) = .05 \quad P(B|A_2) = .06 \quad P(B|A_3) = .07 \quad P(B|A_4) = .08$

Venn Diagram

	A_1	A_2	A_3	A_4
B	$B \cap A_1$	$B \cap A_2$	$B \cap A_3$	$B \cap A_4$
B^c	$B^c \cap A_1$	$B^c \cap A_2$	$B^c \cap A_3$	$B^c \cap A_4$

Breaking down the probabilities of events

Sample space, events, and probabilities

- A_i = positions on chromosome i
  $P(A_1) = .1 \quad P(A_2) = .2 \quad P(A_3) = .3 \quad P(A_4) = .4$
- B = positions in genes
  $P(B|A_1) = .05 \quad P(B|A_2) = .06 \quad P(B|A_3) = .07 \quad P(B|A_4) = .08$

Venn diagram with probabilities

	A_1	A_2	A_3	A_4	Total
B	$P(B \cap A_1)$	$P(B \cap A_2)$	$P(B \cap A_3)$	$P(B \cap A_4)$	$P(B)$
B^c	$P(B^c \cap A_1)$	$P(B^c \cap A_2)$	$P(B^c \cap A_3)$	$P(B^c \cap A_4)$	$P(B^c)$
Total	$P(A_1)$	$P(A_2)$	$P(A_3)$	$P(A_4)$	1

Breaking down the probabilities of events

Sample space, events, and probabilities

- A_i = positions on chromosome i
  $P(A_1) = .1 \quad P(A_2) = .2 \quad P(A_3) = .3 \quad P(A_4) = .4$
- B = positions in genes
  $P(B|A_1) = .05 \quad P(B|A_2) = .06 \quad P(B|A_3) = .07 \quad P(B|A_4) = .08$

Venn diagram with probabilities

Fill in top row with $P(B \cap A_i) = P(B|A_i)P(A_i)$,
and fill in column totals $P(A_i)$.

	A_1	A_2	A_3	A_4	Total
B	$(.05)(.1) = .005$	$(.06)(.2) = .012$	$(.07)(.3) = .021$	$(.08)(.4) = .032$	$P(B)$
B^c	$P(B^c \cap A_1)$	$P(B^c \cap A_2)$	$P(B^c \cap A_3)$	$P(B^c \cap A_4)$	$P(B^c)$
Total	.1	.2	.3	.4	1

Breaking down the probabilities of events

Sample space, events, and probabilities

- A_i = positions on chromosome i
  $P(A_1) = .1 \quad P(A_2) = .2 \quad P(A_3) = .3 \quad P(A_4) = .4$
- B = positions in genes
  $P(B|A_1) = .05 \quad P(B|A_2) = .06 \quad P(B|A_3) = .07 \quad P(B|A_4) = .08$

Venn diagram with probabilities

Fill in rest of table to complete the row and column totals.

	A_1	A_2	A_3	A_4	Total
B	.005	.012	.021	.032	.070
B^c	.095	.188	.279	.368	.930
Total	.1	.2	.3	.4	1

$$P(B) = P(B \cap A_1) + \cdots + P(B \cap A_4)$$
$$= P(B|A_1)P(A_1) + \cdots + P(B|A_4)P(A_4)$$

Questions

Events and probabilities

$P(A_1) = .1 \quad P(B|A_1) = .05$
$P(A_2) = .2 \quad P(B|A_2) = .06$
$P(A_3) = .3 \quad P(B|A_3) = .07$
$P(A_4) = .4 \quad P(B|A_4) = .08$

	A_1	A_2	A_3	A_4	Total
B	.005	.012	.021	.032	.070
B^c	.095	.188	.279	.368	.930
Total	.1	.2	.3	.4	1

- Over the whole genome, what fraction of positions are in genes?
  $P(B) = \boxed{\mathbf{.070}}$
- How many positions are in genes? $(10{,}000{,}000)(.070) = \boxed{\mathbf{700{,}000}}$
- If site x is in a gene, what's the probability x is on chromosome i?
  - $P(A_1|B) = \dfrac{P(B|A_1)P(A_1)}{P(B)} = \dfrac{(.10)(.05)}{.070} \approx \boxed{\mathbf{.0714}}$
  - $P(A_2|B) = \dfrac{(.20)(.06)}{.070} \approx \boxed{\mathbf{.1714}}$ $\quad P(A_4|B) = \dfrac{(.40)(.08)}{.070} \approx \boxed{\mathbf{.4571}}$

    $P(A_3|B) = \dfrac{(.30)(.07)}{.070} \approx \boxed{\mathbf{.3000}}$

Full version of Bayes' Theorem

Definition (Partition of S)

Events $A_1, \ldots, A_n$ *partition* the sample space S when

- $S = A_1 \cup \cdots \cup A_n$.
- $A_i \cap A_j = \emptyset$ for $i \neq j$. *(pairwise mutually exclusive)*
- $P(A_i) > 0$ for all i.

In other words, $A_1, \ldots, A_n$ are all nonempty with positive probability, and every element of the sample space is in exactly one of them.

Theorem (Bayes' Theorem)

Let $A_1, \ldots, A_n$ be mutually exclusive events that partition sample space S, and B be any event on S. Then

- $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$
- *If $P(B) > 0$ then for each $j = 1, \ldots, n$,*

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$