UNIT 1.2 PROBABILITY

Probability theory (PnG p. 3)

- *Probability theory resides within an axiomatic system: … its practicality comes from knowing how to use the theory to yield useful approximations.*
- We use probability theory to model a natural underlying process of a phenomenon (e.g. annual mortality rates of teenagers in the U.S., polls in election, times between computer crashes, etc.)

1. OPERATIONS ON EVENTS AND PROBABILITY

Define event and probability

- **Event:** Any **set** of possible outcomes of a random phenomenon (an experiment or a trial).
  - A, B, C, …

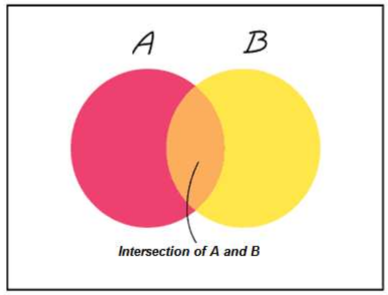
Example:

- A = an event that a 30-year-old woman is alive at age 70.
- B = an event her 30-year old husband is alive at age 70.

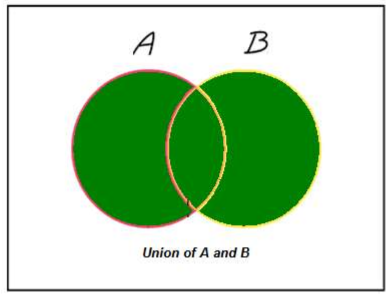
Events - examples

5

- Intersection of A and B, A∩B, is an event both A and B happen.
  - A∩B is an event that both of them are alive at age 70.



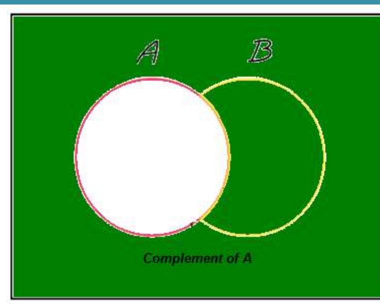
- The *union* of A and B, AUB, is an event either A or B or both happen.
  - AUB is an event that either the 30-year-old woman or her 30-year-old husband lives to age 70, or that they both live to be 70 years of age.



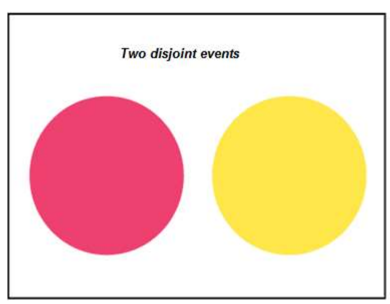
Events - examples

6

- The complement of A, A^c, is the event "not A".
  - A^c is the event that the 30-year-old woman dies before she reaches the age of 70.



- Two events A and B are *disjoint*, if they cannot occur simultaneously.
  - E.g. If A is the event that a newborn's birth weight is under 2000 grams and b is the event that it is between 2000 and 2499 grams.



Define probability of an event

7

- **Probability of an event:** the limit of relative frequency of the event if infinite number of identical and independent trials are conducted.
  (This is called the "Frequentist definition".)
  - P (A) denotes the probability of event A occurring.

$$
\begin{aligned}
P(A) &= P(\{a_1, a_2, \cdots, a_m\}) \\
&= \text{Probability that } \textbf{at least} \text{ one of the outcomes in event A occur} \\
&= \textbf{Probability of event A}
\end{aligned}
$$

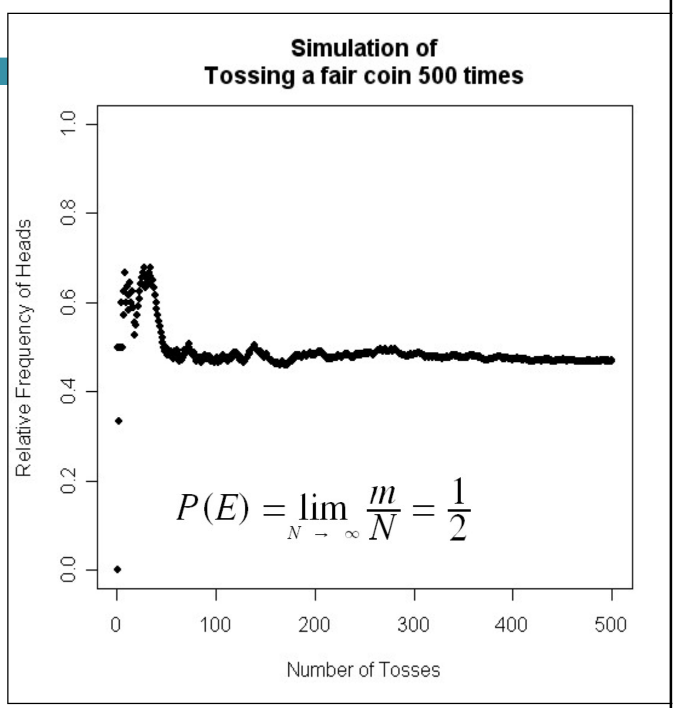
Frequentist definition of probability

Consider a trial to toss a fair coin once. Event E is observing head on one trial.

The probability of E, $P(E)$ is defined as following:

Run the trial N times and record the number of Heads ($= m$).

The probability is the limit of relative frequency (m/N).



Example – probability of an event

9

- National Health Interview Survey 1980-1981 [2]

Employment Status	Population	Impairments
Currently employed	98917	552
Currently unemployed	7462	27
Not in labor force	56778	368
Total	**163,157**	**947**

PnG

- P(a person has hearing impairment) =
- We assume 163,157 is large enough to satisfy the frequentist probability definition.

Example – probability of an event

10

- National Health Interview Survey 1980-1981 [2]

Employment Status	Population	Impairments
Currently employed	98917	552
Currently unemployed	7462	27
Not in labor force	56778	368
Total	**163,157**	**947**

- $P(H^c)$ = P(a person does not have hearing impairment) =
- $P(H \cap H^c)$ = P(a person has hearing impairment and the person does not have hearing impairment) = 0
- $P(H \cup H^c)$ = P(a person has hearing impairment or the person does not have hearing impairment) = 1

Example – probability of an event

11

- National Health Interview Survey 1980-1981 [2]

Employment Status	Population	Impairments
Currently employed	98917	552
Currently unemployed	7462	27
Not in labor force	56778	368
Total	**163,157**	**947**

  - $P(E_1)$ =P(a person is employed) =
  - $P(E_1^c)$ =P(a person is in not currently employed) = 1- $P(E_1)$ =
  - $P(E_1 \cup E_2) = P(E_1)+P(E_2) =$ (Why? Disjoint events)

Addition Rule: union of events in a trial

12

- Formal Addition Rule
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
    where $P(A \cap B)$ denotes the probability that A and B both occur at the same time as an outcome in a trial.
- When A and B are disjoint events
  - $P(A \cup B) = P(A) + P(B)$

Rules of Complementary Events

- A and A^c are disjoint
  - That is, it is impossible for an event and its complement to occur at the same time.

- $$P(A) + P(A^c) = 1$$
  $$P(A^c) = 1 - P(A)$$
  $$P(A) = 1 - P(A^c)$$

2. CONDITIONAL PROBABILITY

Definition - Conditional Probability

15

P(B | A) represents the probability for the second event *B* to occur assuming that the first event *A* has already occurred.

- "Conditional probability of B given A"

Example – conditional probability

16

- National Health Interview Survey 1980-1981 [2]

Employment Status	Population	Impairments
Currently employed	98917	552
Currently unemployed	7462	27
Not in labor force	56778	368
Total	**163,157**	**947**

- Let E_1 be the event that an individual is currently employed.
- Let H be the event that an individual has a hearing impairment.
- $P(E_1)$ = P(an individual is currently employed) = 0.6063 (why?)
- $P(H \cap E_1)$ = P(an individual is currently employed and has hearing impairment) = 0.00338 (why?)
- $P(H|E_1)$ = P(an individual has hearing impairment | the individual is currently employed) = 0.0056 (why?)
- Note that $P(H \cap E_1)/P(E_1)$ is also 0.0056.

Multiplication Rule

17

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

  or $P(A \cap B) = P(A) \cdot P(B|A)$

Definition – Independent events

18

- Independent Events

  Two events A and B are independent if the occurrence of one does not affect the probability of the occurrence of the other. (Several events are similarly independent if the occurrence of any does not affect the probabilities of occurrence of the others.)

Multiplication rule when independent

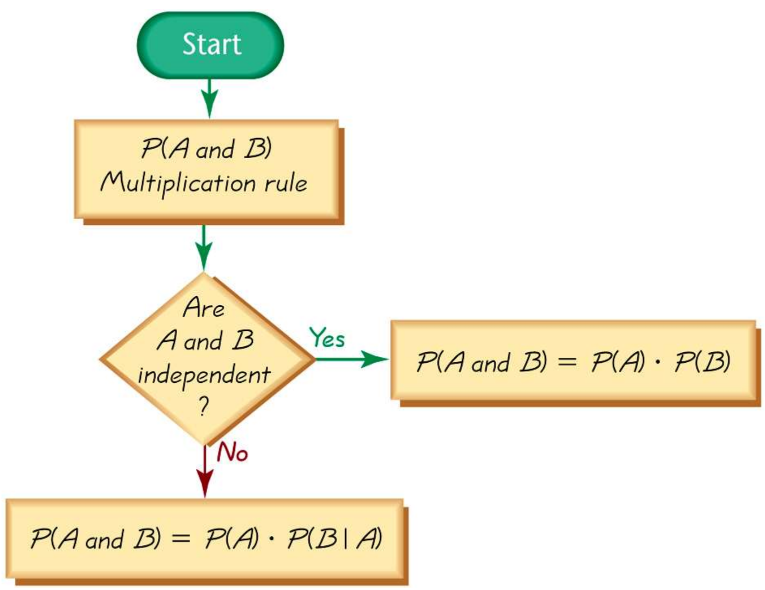
19

- Two events are independent if and only if
  $P(A \cap B) = P(A)P(B)$

19

Applying the Multiplication Rule

20



Elementary Statistics, Triola

20

3. BAYES' THEOREM

Example – probability of an event

- National Health Interview Survey 1980-1981 [2]

Employment Status	Population	Impairments
Currently employed	98917	552
Currently unemployed	7462	27
Not in labor force	56778	368
Total	**163,157**	**947**

- P$(E_1|H)$ = 0.583
- $\frac{P(H|E_1)P(E_1)}{P(H|E_1)P(E_1)+P(H|E_1^C)P(E_1^C)} = 0.583$ (!)
  - P(E_1) = 0.6063, P(E_1^C) = 0.3937
  - P$(H|E_1)$ = 0.0056, P$(H|E_1^C)$ = 0.0061 (why?)

Example – probability of an event

- National Health Interview Survey 1980-1981 [2]

Employment Status	Population	Impairments
Currently employed	98917	552
Currently unemployed	7462	27
Not in labor force	56778	368
Total	**163,157**	**947**

- $P(E_1|H) = \frac{P(H|E_1)P(E_1)}{P(H|E_1)P(E_1)+P(H|E_1^C)P(E_1^C)}$
  - We can directly compute $P(E_1|H) = 0.583$.
  - **This is useful when $P(E_1|H)$ is not directly computable.**

Bayes' theorem

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

4. DIAGNOSTIC TESTS

Trivia question

- A person is tested positive for a disease that affects 1 in 5,000. The test is known to accurately 99% of times showing positive results for people with disease and showing negative results for people without the disease. What is the chance that the person really does have this illness?

Application of Bayes' theorem: Diagnostic testing or screening

- Screening:

  Application of a test to who have not yet exhibited any clinical symptoms (p.135)

  - Test for steroid use of athletes
  - HIV screening by blood test
  - Pap smear for cervical cancer (p.136)
  - X-ray screening for tuberculosis

Application of Bayes' theorem: Diagnostic testing or screening

- Events
  - D = an individual has a particular disease
  - D^c = an individual does not have the disease
  - T^+ = positive screening test result happens
  - T^- = negative screening test result happens
- Sensitivity: probability of positive test result given that the individual actually has the disease
  - Sensitivity = $P(T^+|D)$

Application of Bayes' theorem: Diagnostic testing or screening

- Specificity: probability of negative test result given that the individual actually does not have the disease
  - Specificity = $P(T^- | D^c)$
- Prevalence: probability of an individual having the disease or proportion of people having the disease
  - Prevalence = P(D)

Diagnostic test

[*Yerushelmy, 1970*]	No Tuberculosis	Tuberculosis	Total
X-ray negative	*1739*	*9*	*1747*
positive	*51*	*22*	*73*
Total	*1790*	*30*	1820

- This was a case-control study.
  - Sensitivity = $P(T^+ | D) = 22/30 = 0.7333$
  - Specificity = $P(T^- | D^c) = 1739/1790 = 0.9715$
- What is $P(D|T^+)$, the positive predictive value?

Diagnostic test

[*Yerushelmy, 1970*]	No Tuberculosis	Tuberculosis	Total
X-ray negative	*1739*	*9*	*1747*
positive	*51*	*22*	*73*
Total	*1790*	*30*	1820

- Since this was a case-control study, we cannot directly estimate $P(D|T^+)$ "positive predictive value".
  - It's not 22/73.

Application of Bayes Theorem

Prevalence

Sensitivity

1-specificity

$$P(D\mid T^+) = \frac{P(D)P(T^+\mid D)}{P(D)P(T^+\mid D) + P(D^c)P(T^+\mid D^c)}$$

- P(D) = Prevalence: not obtainable from this data. <u>Because the data are not chosen from the population at random.</u>
- *Morbidity and Mortality Weekly Report, 38:16, 1987*
  - 9.3 cases of tuberculosis per 100,000 : P(D) = 0.000093

Application of Bayes Theorem

$$P(D \mid T^{+}) = \frac{0.000093 \times 0.7333}{0.000093 \times 0.7333 + (1 - 0.000093)(0.0285)} = 0.00239$$

- For every 10,000 positive x-rays, only 239 signal true cses of tuberculosis.

- Reference
  - Principles of Biostatistics (Pagano and Gauvreau)
  - *Elementary Statistics* by Triola, 10^{th} edition.
  - *Applied Statistics for Engineers and Scientists* by Petruccelli et al.

- Acknowledgements
  - Prof. Jayson Wilbur, WPI
  - Prof. Balgobin Nandram, WPI
  - Prof. Lee Jaeyong, Seoul National University
  - Some slides provided by Pearson Education, Inc Publishing as Pearson Addison-Weslely