Basic probability theory

22.10.2006
GE02: day 1 part 1

Yurii Aulchenko
Erasmus MC Rotterdam

Overview: probability theory definitions

- Experiment and experimental outcomes
- Events and Probability
- Mutually exclusive events & the rule of summation of probabilities
- Complementary events
- Independent events & the rule of multiplication of probabilities

Experiment

Any planned process of data collection

- Tossing a coin
- Measuring height is people
- Genotyping people
- Sampling pedigrees via proband
- ...

Composition of an experiment

Consists of a number of independent trials

- Tossing a coin 3 times
- Experiment with 3 trials

Each trial can result in some outcome

- Head or tail
- Trial with 2 outcomes

Many trials – many possible outcomes

- {**TTT**, **TTH**, **THT**, **HTT**, **THH**, **HTH**, **HHT**, **HHH**}
- Experiment with 8 outcomes

Importance of tossing coins

- Abstracts an experiment with binary outcome
- Probability theory
  - Binomial, Poisson & Normal distributions
  - Hypothesis testing
- Genetic epidemiology
  - Mendel’s lows
  - Hardy-Weinberg equilibrium
  - Genetic drift

Problem

- How many outcomes exist for experiment in which n coin tosses is made?

Number of experimental outcomes

After 1 trial

2 outcomes

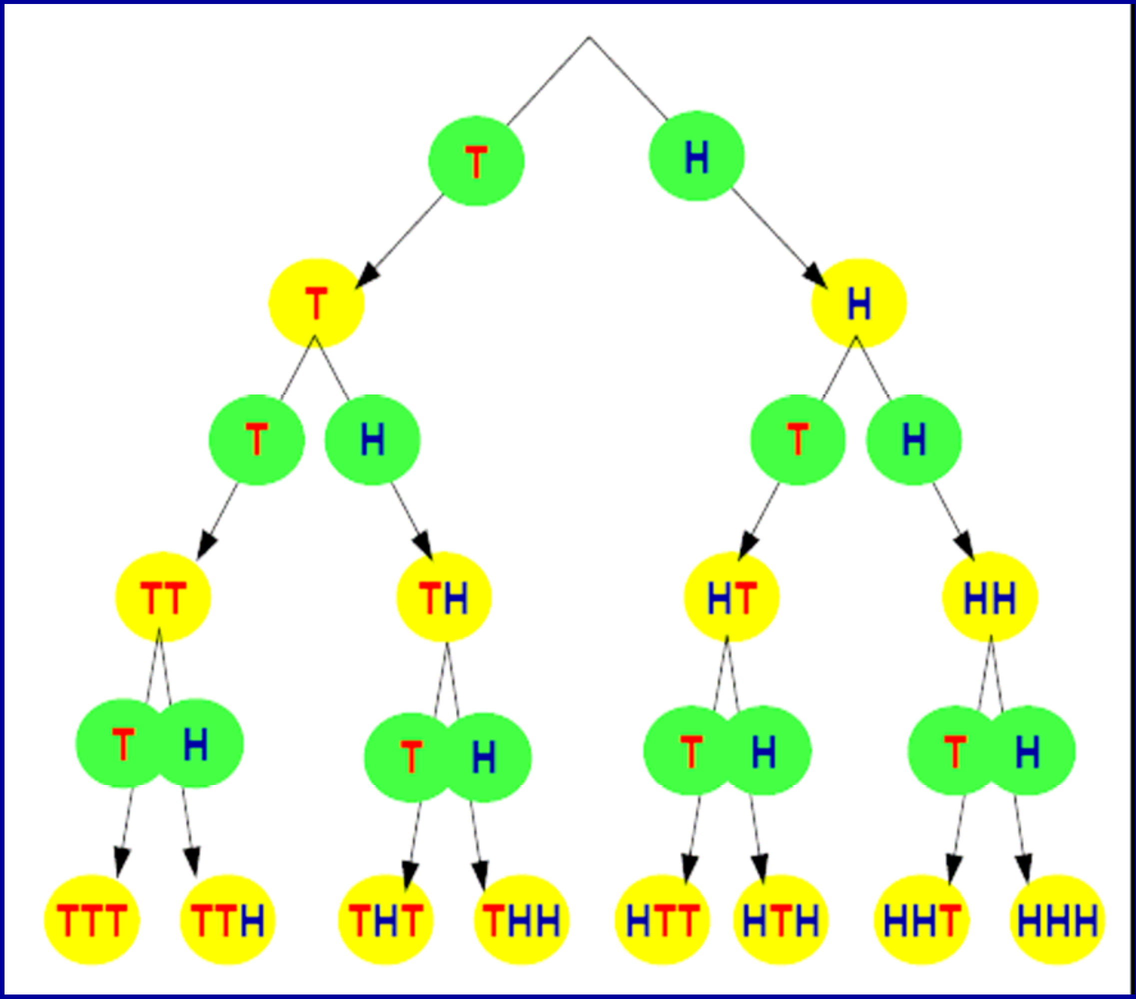
After 2 trials

2 x 2 = 4 outcomes

After 3 trials

4 x 2 = 8 outcomes

After *n* trials: 2^n outcomes



Event

Is defined as a single experimental outcome

- All three heads {HHH}
- Two heads and then 1 tail {HHT}

… or a set of outcomes

- A head in 2nd trial {THT, HHT, THH, HHH}
- Exactly 1 head {HTT, THT, TTH}
- More then 1 head {HHH, HHT, HTH, THH}

Probability

A function of event taking values between 0 and 1

- Frequently denoted as P(event)

Measures how likely is event

- if P(A) is close to 0 then A is unlikely
- if P(A) is close to 1 then A is very likely

Probabilities of all possible experimental outcomes sum to one

Probability of heads or tails

Tossing a coin once

- Outcome: either head (**H**) or tail (**T**), mutually exclusive
- If coin is fair, both are equally likely
- Therefore P(**H**) = P(**T**) = 0.5 (or 50%)

More formal

- Because of symmetry P(**H**) = P(**T**)
- H or T are all possible outcomes => P(**H**) + P(**T**) = 1
- Thus P(**H**) = P(**T**) = 0.5

Probability in *n* trials

- 2^n outcomes are possible, all are equivalent
- Then probability of a particular outcome is $1/2^n$
- For example, for 2 trials:
  - P(**HH**) = ¼
  - P(**HT**) = ¼
  - P(**TH**) = ¼
  - P(**TT**) = ¼

Mutually exclusive events

If the occurrence of one event precludes the occurrence of the other. In three trials

- Events **HHT** and **HTH** are mutually exclusive
- Events "more then 1 head" and "two heads" are not

If events A and B are mutually exclusive, then

- Pr(A or B) = Pr(A) + Pr(B)

Example

What is probability to have event A = ("one or more heads" in two trials)?

- P(**HH**) = ¼
- P(**HT**) = ¼
- P(**TH**) = ¼
- P(**TT**) = ¼

P("one or more heads") =

P(**HH** or **HT** or **TH**) =

P(**HH**) + P(**HT**) + P(**TH**) =

¼ + ¼ + ¼ = ¾

Genetic example

Consider a gene with two alleles

- N (Normal) and
- D (Disease)

Probability of observing a person with genotypes NN, DN and DD in a population are

- P(NN)=0.81,
- P(ND)=0.18 and
- P(DD)=0.01

Problem

- What is probability that a random person is a carrier of the D allele?

- A **carrier** of an allele is a person having at least one copy of this allele in the genotype

Solution

= P(carrier) =

= P(ND or DD) =

[mutually exclusive =>]

= P(ND) + P(DD) =

= 0.18 + 0.01 =

= 0.19

Complementary event

Define B = “not A”

B is called an “event, complementary to A”

“A or B” covers all possible outcomes

- P(A or B) = P(A or “NOT A”) = 1

B and A are mutually exclusive

- P(A or B) = P(A) + P(B) = 1

Therefore P(B) = 1 – P(A)

Examples

What is probability to have "one or more heads" in two trials?

P("more than 1 head") =

[complementary event = "no heads"]

1 – P("no heads") = 1 – P(TT) = 1 – ¼ = ¾

What is probability that a person is carrier? [P(NN)=0.81, P(ND)=0.18, P(DD)=0.01]

P("carrier") =

[complementary event = "not carrier"]

1 – P("not carrier") = 1 – P(NN) = 1 – 0.81 = 0.19

Independent events

Two events are **independent** if the outcome of one has no effect on the outcome of the second event

- Event "**having head in first toss**" and "**having head in second toss**" are independent
- Genotypes of two random people from a population are independent

Probability: independent events

- Two events A and B are independent when
  - Pr(A and B) = Pr(A) Pr(B)
- Sex of next offspring does not depend on the sex of the previous
  - P(boy) = P(girl) = ½

What is probability of three girls born?

- P("girl" and "girl" and "girl") = ½ ½ ½ = 1/8
- The same applies to having Heads three times

Problem

- In some elderly population, prevalence of hypertension (HT) is 42% in female and 57% in male. What is the probability that
  - Both spouses are HT?
  - Both spouses are NOT HT?
- Assume independence

Solution

- Both spouses are HT

  $P(husband=HT \,\&\, wife=HT) =$

  $P(male=HT)\ P(female=HT) = 0.57\ 0.42 = 0.24$

- Both spouses are NOT HT

  $P(husband \neq HT \,\&\, wife \neq HT) =$

  $P(husband \neq HT)\ P(wife \neq HT) =$

  $[1 - P(male=HT)]\ [1 - P(female=HT)] =$

  $0.43\ 0.58 = 0.25$

Non-independent events

Risk of disease in women whose partner had a disease compared to those whose partner did not

Disease	Odds Ratio (95% CI)
Hypertension	1.39 (1.14 to 1.70)
Depression	2.18 (1.78 to 2.67)
Asthma	1.68 (1.45 to 1.94)

Shared lifestyle risk factors?

Hippisley-Cox et al., BMJ 2002;325:636