Basic probability
Mendel's lows

24.10.2005
GE02: day 1 part 1

Yurii Aulchenko
Erasmus MC Rotterdam

Experiment

- Any planned process of data collection
  - Tossing a coin
  - Measuring height is people
  - Genotyping people
  - Sampling pedigrees via proband
  - ...

Composition of an experiment

- It consists of a number of independent trials (or replications)
  - Tossing a coin 3 times
- Each trial can result in some outcome
  - Head or tail
- Many trials – many possible outcomes
  - {TTT, TTH, THT, HTT, THH, HTH, HHT, HHH}

Event

- A single experimental outcomes or as a set of outcomes, e.g.
  - All three heads (HHH)
  - Two heads and then 1 tail (HHT)
  - Exactly 1 head (HTT, HTH, HHT)
  - More then 1 head (HHH, HHT, HTH, THH)

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

Task

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

Number of experimental outcomes

- 1 toss : 2 outcomes
  - {**H**, **T**}
- 2 tosses : 4 outcomes
  - {**HH**, **HT**, **TH**, **TT**}
- 3 tosses : 8 outcomes
  - {**HHH**, **HHT**, **HTH**, **HTT**, **THH**, **THT**, **TTH**, **TTT**}

- *n* tosses : 2^n outcomes

Probability

- A function of event which takes 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
  - 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)

Task

- Experiment consist of tossing a coin three times
- What is the probability to have “exactly one head” or “exactly one tail”?

Solution

= Pr(exactly one head OR exactly one tail)

[are mutually exclusive =>]

= Pr(exactly one head) + Pr(exactly one tail) =

= Pr(HTT or THT or TTH) + Pr(THH or HTH or HHT)

[are mutually exclusive =>]

= Pr(HTT) + Pr(THT) + … + Pr(HHT) =

[each of 8 possible outcomes is equally likely =>]

= 6 / 8

More genetic example

- Consider a gene with two alleles
  - N (Normal) and
  - D (Disease)
- The 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

Task

- What is probability that a random person is a carrier of the disease allele?
- P.S. 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

Independent events

- Two events are **independent** if the outcome of one has no effect on the outcome of the second event
  - Tossing coins two times
    - 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)
- For example, the sex of next offspring does not depend on the sex of the previous
  - P(boy) = P(girl) = ½
- What is chance that all three children are girls?
  - P(all 3 girls) = ½ ½ ½ = 1/8
  - The same applies to having Heads three times

Task

- In some 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≠HT & wife≠HT) =

  P(husband≠HT) P(wife≠HT) =

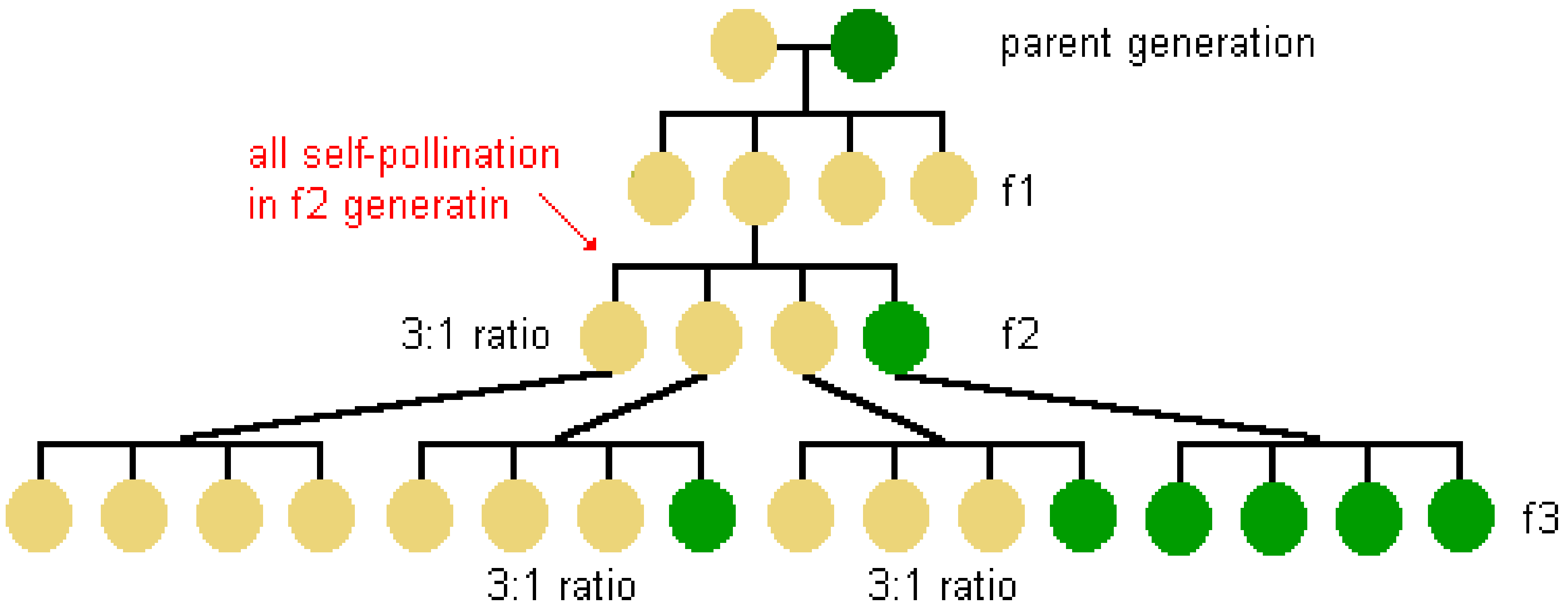
  [1 – P(male=HT)] [1 – P(female=HT)] =

  0.43 0.58 = 0.25

Mendel’s lows

- Pre-requisite: two parental forms, qualitatively different for a trait for a number of generations => parental forms are homozygous for different variants of the gene

Uniformity of F_1 and segregation of F_2



Mendel's low may be reduced to one assumption

- Three of concepts

Alleles: **Y**, **G**

Genotype	**Phenotype**
YY	**Yellow**
YG or GY	**Yellow**
GG	**Green**

- One assumption
  - The alleles are transmitted to the next generation in random manner

Uniformity of F_1

- **Yellow** parental form has genotype **YY**
- **Green** parental form has genotype **GG**
- Then, in F_1 all plants have genotype **YG** (**Y** from **Yellow** parent and **G** from **Green** parent)
- All F_1 will be **Yellow**.

Segregation of F_2

- According to random transmission assumption **YG** plants will produce 50% **Y** and 50% **G** gametes. These will randomly aggregate to give F_2:
  - P(**Y** & **Y**) = P(**Y**) P(**Y**) = ½ ½ = ¼ (**Yellow**)
  - P(**Y** & **G**) = P(**Y**) P(**G**) = ½ ½ = ¼ (**Yellow**)
  - P(**G** & **Y**) = P(**G**) P(**Y**) = ½ ½ = ¼ (**Yellow**)
  - P(**G** & **G**) = P(**G**) P(**G**) = ½ ½ = ¼ (**Green**)
- Thus ¾ will be **Yellow** and ¼ will be **Green**
- The famous 3:1 is established!

Task

- Consider two independent traits
  - Color, as in previous example and
  - Seed’s shape (wrinkled or smooth), which is controlled by alleles W and S, with S being dominant
- What is expected trait distribution in F_1 and F_2?

Free combination low: 9:3:3:1