Conditional probability
Formula of total probability

23.10.2005

GE02 day 2 part 1

Yurii Auchenko

Erasmus MC Rotterdam

Colour blindness: experiment

- Experiment: drawing a random subject from a total population of N people
- In a subject, we can observe the following features
  - Sex = {M, F}
  - Colour-blindness = {D, U}
- ...We aim to predict the risk (the probability) that this random subject is colour-blind

Relations between events

M and F are mutually exclusive

P(M&F) = 0

D and U are mutually exclusive

P(D&U) = 0

Sex and colour blindness are not:

P(M&U) > 0

P(M&D) > 0

P(F&U) > 0

P(F&D) > 0

Numbers

Let

- number of affected is N_D
- number of unaffected is $N_U = N - N_D$
- number of males is N_M
- number of females is $N_F = N - N_M$

We also know

- number of affected males, $N_{D\&M}$
- number of affected females, $N_{D\&F}$

Probabilities

Then the probability that a random subject is colour-blind is

- N_D/N

But we know that frequency of colour-blindness in males is higher then in female!

- Or, to say it more formal, probability that a person is colour-blind, depends on sex

Using more information in risk prediction

- Our risk prediction may gain accuracy if we utilize the information on sex
- What is the probability that a random male is affected? Or, better to say, what is probability of being affected GIVEN the person is male?

  – $P(D|M) = N_{M\&D}/N_M = P(M\&D)/P(M)$

Conditional probability

Probability of being colour-blind given sex

- P(D|M)
- is an example of conditional probability

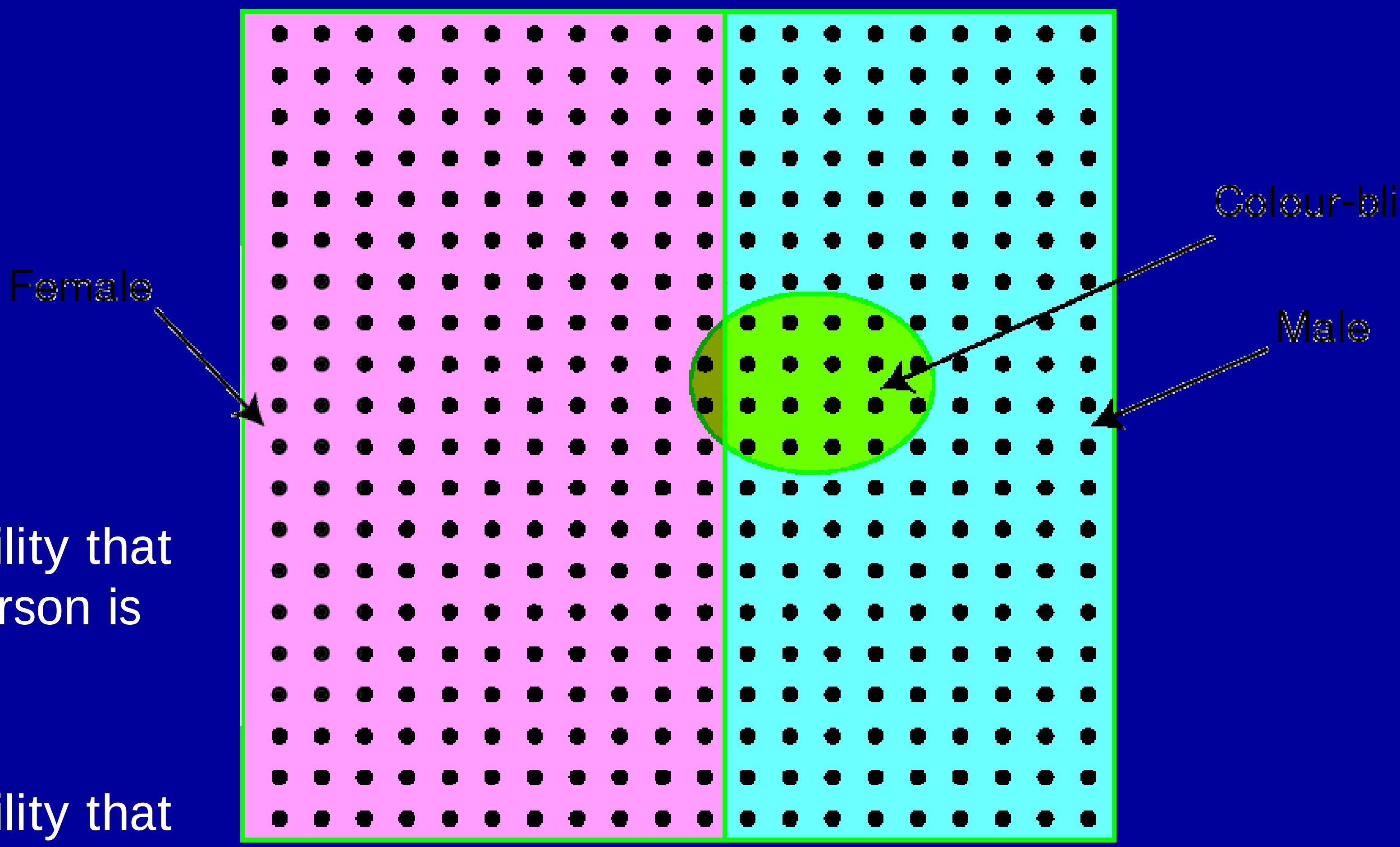
There are many genetic probabilities that are conditional

- transmission probabilities
- penetrances
- ...

Generally, $P(A|B) = P(A\&B)/P(B)$

Problem

- Compute
  - P(D)
  - P(M)
  - P(F)
  - P(D|M)
  - P(D|F)
- Compute probability that a colour-blind person is male,
  - P(M|D)
- Compute probability that a colour-blind person is female,
  - P(F|D)



Solution

- N = 400
- P(D) = 20/400 = 1/20 = 5%
- P(M) = 180/400 = 9/20
- P(F) = 220/400 = 11/20
- P(D|M) = 18/180= 1/10 = 10%
- P(D|F) = 2/220 = 1/110 = 0.9%

- P(M|D) = 18/20 = P(M&D)/P(D)
- P(F|D) = 2/20 = P(F&D)/P(D)

Problem

There are two bowls full of cookies. Bowl #1 has 10 chocolate chip cookies and 30 plain cookies, while bowl #2 has 20 of each

- What is probability to pick up a plain cookie from bowl #1?
- ... #2?
- What is probability to pick up a a bowl at random and then cookie at random and then to discover that it is a plain one?
- If you pick up a bowl at random and then a cookie at random and discover that it was a plain one, what is probability that you picked it up from the bowl #1?
- ... from bowl #2?

Solution

- Denote bowl as B and cookie as C
  - $P(C=plain|B=1) = N_{plain\ in\ \#1}/N_{\#1} = 30/40 = 3/4$
  - $P(C=plain|B=2) = N_{plain\ in\ \#2}/N_{\#2} = 20/40 = 1/2$
  - $P(C=plain) = N_{plain}/N = 50/80 = 5/8$

  - $P(B=\#1|C=plain) = N_{plain\ in\ \#1}/N_{plain} = 30/50 = 3/5$
  - $P(B=\#2|C=plain) = N_{plain\ in\ \#2}/N_{plain} = 20/50 = 2/5$

Problem

- Let in population there are 2 alleles, M and N
- Frequency of M, $P(M)=0.05$
- Penetrances (conditional probability of having disease given genotype) are
  - $P(D|MM)=1.0$
  - $P(D|MN)=0.7$
  - $P(D|NN)=0.03$
- Assuming HWE, what is the frequency of disease in the population?

Solution

- Frequency of M, P(M)=0.05. Thus, assuming HWE,
  - P(MM) = 0.0025, P(MN) = 0.095, P(NN) = 0.9025
  - Of MM, who make 0.0025 of the population, all are ill, thus, they contribute 0.0025 to the frequency of the diseas
  - Of MN, who make 9.5% of the population, 70% are ill, thus, they contribute 0.095*0.7 = 0.0665 to the frequency of the disease
  - Of NN, 3% are ill, they contribute 0.9025*0.03 = 0.0271 to the disease

Solution

- Thus, the frequency of disease is

0.0025 (these ill among MM) +

0.0665 (among MN) +

0.0271 (among NN) = 0.0961 =

9.61% of the population are ill

Formula of total probability

- We were following schema

P(M)	0,05		
g	P(g)	P(D|g)	P(g)*P(D|g)
MM	0,0025	1,0000	0,0025
MN	0,0950	0,7000	0,0665
NN	0,9025	0,0300	0,0271
		P(D)=	0,0961

And the computations were done using the formula

$$P(D) = \sum_{g=MM,MN,NN} P(D \mid g) P(g) =$$

$$P(D \mid MM) P(MM) + P(D \mid DM) P(DM) + P(D \mid DD) P(DD)$$

Problem

- Use the total probability formula to find out the chance to pick up a a bowl at random and then cookie at random and then to discover that it is a CHOCOLATE one

Solution

$$P(C = chocolate) = \sum_{bowl=1,2} P(C = choc \mid B = bowl) P(B = bowl) =$$

P(C=choc|bowl=1)P(bowl=1) +

P(C=choc|bowl=2)P(bowl=2) =

¼ ½ + ½ ½ = 3/8