Faculty of Life Sciences

Probabilities

Ib Skovgaard & Claus Ekstrøm
E-mail: ims@life.ku.dk

Program

- Probabilities
  - Sample space, events and probabilities
  - Rules for probability calculations
- Independence
- Conditional probability

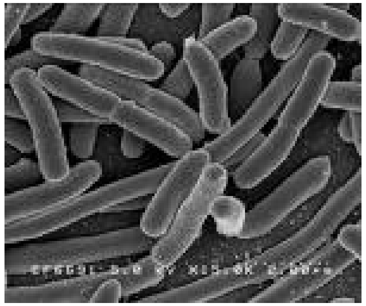
Probabilities

- Rules for probabilities are for statistics what arithmetic is for mathematics.
- Any probability statement, like a p-value or a confidence interval, is the result of a probability calculation.
- You will only see a tiny bit of it, but enough for computing precision of counts.

Example: diagnostic tests

Known from earlier experiments that,

- For meat with E. coli O157 the test is positive in 90% of the cases true positives.
- For meat without E. coli O157 the test is negative in 95% of the cases E.em E. coli 0157 is in 0.01% of the meat samples (the prevalence of E. coli 0157).



Problem: the test is not perfect!

What is the probability for a sample to be infected by the E. coli, if the test is positive?

Example: inverse sampling

Estimation of the fraction of leaves with aphids.

- Method 1: direct sampling. Inspect 60 leaves for aphids.

  Result 12 leaves with aphids (for example). Estimated fraction $12/60 = 0.20$. Uncertainty?
- Method 2: inverse sampling. Count the number of leaves inspected until one with aphids is found. Do that 10 times, say.

  Results (for example): Aphids found on leaf number

  2 7 2 9 1 3 7 1 3 2

Estimate for fraction of leaves aphids? Uncertainty?

Sample space and events

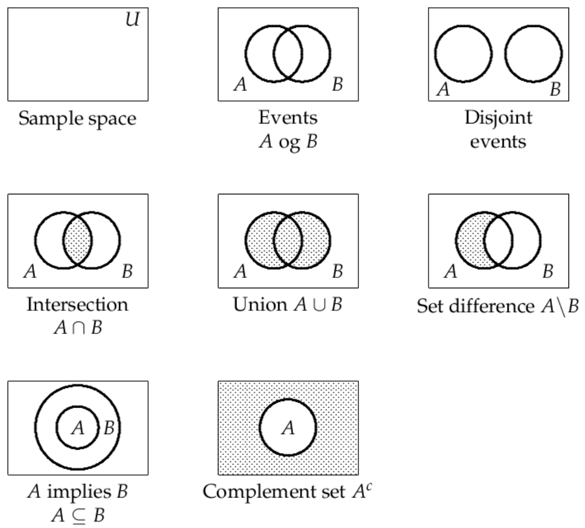
The sample space, U, is the set of all possible results. When throwing a die, the sample space is the set $U = \{1,2,3,4,5,6\}$.

An event, A, is a subset of $U, A \subseteq U$.

Example: the event *"an odd number"* when throwing a die,

$$A = \{1,3,5\}$$

Relations between events/sets



Definition of probability

Let U be the sample space for some trial. A probability distribution on U is a function P that assigns a number, $P(A)$, to any event, A say, and which satisfies the conditions,

1. $0 \leq P(A) \leq 1$ for any event, A.
2. $P(U) = 1$
3. $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$.

Three important probability rules

- Addition rule
  - P(Either A or B occurs) $= P(A) + P(B)$,
    if A and B exclude each other
- Multiplication rule

  P(A and B both occur) $= P(A) \cdot P(B)$,
  if A and B are independent
- Bayes' rule

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

A formula for "switching"a conditional probability

UNIVERSITY OF COPENHAGEN FACULTY OF LIFE SCIENCES

More computation rules

Let A and B be events in the sample space U. Then,

4. $P(B \setminus A) = P(B) - P(A \cap B)$.
5. $P(A) \leq P(B)$ If $A \subseteq B$.
6. $P(A^c) = 1 - P(A)$.
7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
8. $P(\emptyset) = 0$
9. $P(A_1 \cup \cdots \cup A_k) = P(A_1) + \cdots + P(A_k)$ if $A_1, \ldots, A_k$ is pairwise disjoint events, $A_i \cap A_j = \emptyset$ for all $i \neq j$.

UNIVERSITY OF COPENHAGEN FACULTY OF LIFE SCIENCES

Addition rule: use it with care!

If the probability of finding a certain DNA-sequence in a DNA-sample is 0.03, then what is the probability of finding the sequence in at least one out of five DNA-samples?

UNIVERSITY OF COPENHAGEN FACULTY OF LIFE SCIENCES

Conditional probabilities

Let A and B be events with $P(B) > 0$. The conditional probability for A given B is defined as

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

Interpretation: the probability of A among those cases where B occurs.

If $P(A|B) = P(A)$, A is said to be independent of B, (see next slide).

Interpretation: $P(A)$ is unaffected of whether B occurs.

Independence

Two events, A and B, are said to be independent, if

$$P(A \cap B) = P(A) \cdot P(B).$$

This is, in fact, the multiplication rule. Thus, the multiplication rule just repeats the definition of independent events.

Independent replications of a trial (or an experiment) means a series of trials for which events from the different trials are independent. (The trials do not affect each other).

Exercise

Two parents both have genotype Aa for eye color, that is, one allele (gene) for blue eyes (a) and one for brown eyes (A). What is the probability for their child to have brown eyes (genotype AA or Aa, but not aa)?

NOT ALLOWED to add, subtract, multiply or divide without mentioning which rule you use to allow it!

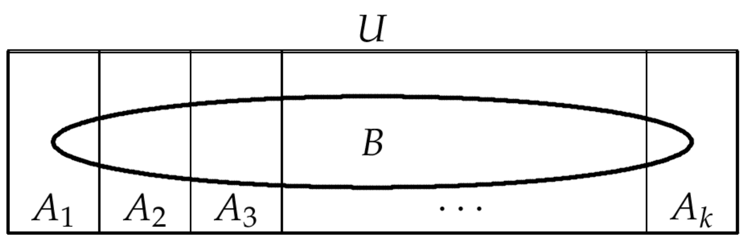
Bayes' rule

For events A and B with $P(A) > 0$ and $P(B) > 0$, Bayes' rule states that

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

Useful for "switching" conditional probabilities. (This essentially just repeats the definition of a conditional probability.)

Bayes' rule II



$$P(B) = P(B|A_1)P(A_1) + \ldots + P(B|A_k)P(A_k)$$

provided that the sets $A_1, \ldots, A_k$ partition the sample space (as shown in the figure), and $P(A_i) > 0$.

Salmonella Manhattan and smoked ham

Case-control investigation of a series of *Salmonella Manhattan* infections. Infected persons are compared with a control group.

Have eaten smoked ham from slaughterhouse S	Infected? Yes	No	Total
Yes	16	80	96
No	2	2500	2502
Total	18	2580	2598

We want: P(inf|ham). Hmmm...
We have: P(ham|inf).
We need: P(inf): from hospitals: P(inf)= 0.002.
We need also P(ham): from sales statistics: P(ham)= 0.033.

Actually, it suffices to know either P(inf) or P(ham).

UNIVERSITY OF COPENHAGEN FACULTY OF LIFE SCIENCES

Lecture summary: main points

- Probabilities and rules of computation
- Conditional probabilities
- Bayes' rule