Bayesian Data Analysis

2020

Week 1: Probability theory – the logic of science

Jarno.Vanhatalo@helsinki.fi

Aims of the course

- Basic theory related to Bayesian modeling
  - how probabilities are used to summarize the uncertainty in our knowledge
  - how Bayes theorem can be used to update these uncertainties
  - Basic concepts related to probabilistic modeling and statistical inference
  - Hierarchical model
- Common models
  - Binomial, Gaussian, Poisson, (generalized) linear model…
  - Hierarchical models
- How to use R and Stan for calculations

Practical arrangements

- All information is provided in the moodle page of the course
  - Course requirements and grading
  - Lectures
  - Exercises
    - The emphasis is on hands-on practicing
    - Ask for help for exercises
  - Study material
    - A comment on course book: It is excellent ”hand book” to Bayesian statistics. However, the book is heavy reading at places and contains lengthy examples and ”hard to follow sentences”. Try to pick the essential, don’t get stuck with examples.

Aims of the week

To be familiar with…

- how probabilities are used to summarize uncertainty
- how Bayes theorem can be used to update knowledge
- how probabilistic predictions are made

Probability as a measure of uncertainty

- Bayesian theory provides a consistent tool to
  - Represent uncertainty by probability
  - To update knowledge when new evidence is obtained

What is uncertainty?

Few picks from earlier course:

- Measurement error
  - Inability to measure exactly, Randomness
  - … we try to solve the epistemic uncertainty through experiment, from which we obtain data that contain both the epistemic and aleatory uncertainty.
  - I only thought about epistemic uncertainty when I wrote about uncertainty before I had read the pre-course material. It was kind of strange because I deal with the random error of measurements every day in my studies.
- Process
  - Natural variation -> randomness
- How reliable estimate/statement is
  - Smaller uncertainty -> higher reliability
  - High uncertainty -> caution
- I somewhat disagree with … (O'Hagan). The author divides uncertainty into … (aleatory uncertainty) and … (epistemic uncertainty) …. there is nothing *fundamentally* unpredictable about a coin toss or a measurement error ... The reason why we treat them as purely random is because *in practice* I can't have enough knowledge about, … the mechanical causes of error … But *in principle* I could know, and there would be nothing *inherently* random about them. … If we want examples of true randomness (i.e. *real* aleatory uncertainty), …: the *fundamentally* random nature of certain quantum phenomena.
- …And it also appears that the Bayesian way of seeing data and what can be done with it is totally different to what I am used to. So I look forward to seeing which type of statistics I will be advocate after it.
- Lack of knowledge
  - Descriptor of belief about uncertain things
- Bayesian approach looks a bit more subjective at the beginning
- Is there such a thing as aleatory uncertainty
- …

Taxonomy of uncertainty

Uncertainty can be divided into two categories

- Aleatory (stochastic) uncertainty, which originates from randomness
  - We can not make observations that would help to reduce this uncertainty
- Epistemic (knowledge) uncertainty, which originates from the lack of knowledge
  - We can make observations to reduce this uncertainty
    - The epistemic uncertainty changes when information changes
  - Two observers may have different epistemic uncertainty

Example
- Coin tossing
- Bag of balls

Taxonomy of uncertainty

Example: draw one ball from a bag of balls

- The ratio of the black and white balls is known
  - There is aleatory uncertainty about the colour of the next ball drawn
- The ratio of the black and white balls is unknown
  - There is aleatory and epistemic uncertainty present (uncertainty about the ratio)
  - The epistemic uncertainty changes when balls are drawn
- Draw all the balls at once and count them
  - there would be no aleatory uncertainty about the outcome
  - only epistemic uncertainty about the contents of the bag

Taxonomy of uncertainty

- Strict division between aleatory and epistemic uncertainty is problematic
  - Practice:
    - (Part of) aleatory uncertainty can be reduced to epistemic if we can add more conditions
    - Uncertainty about uncertainty and unknown unknowns
  - Philosophy:
    - ”is there such thing as pure aleatory uncertainty?” (E.g. Heisenberg’s uncertainty principle, Quantum physics)
- More on this throughout the course

Probability as a measure of uncertainty

- Let
  - E be the event we are interested in
  - I be our background information
- $\Pr(E|I)$ is the probability of an event given (”under conditions”) I
  - $\Pr(E|I)$: number between 0 and 1
  - $\Pr(E|I) = 0$: E is impossible
  - $\Pr(E|I) = 1$: E happens for sure
  - $\Pr(E|I) = 0.64$: E might or might not happen
  - $\Pr(E1|I) > P(E2|I)$: $E1$ will happen more likely than $E2$

Mathematical treatment of probabilities

- Notation

$$\Pr(E1 \text{ and } E2) = \Pr(E1, E2)$$

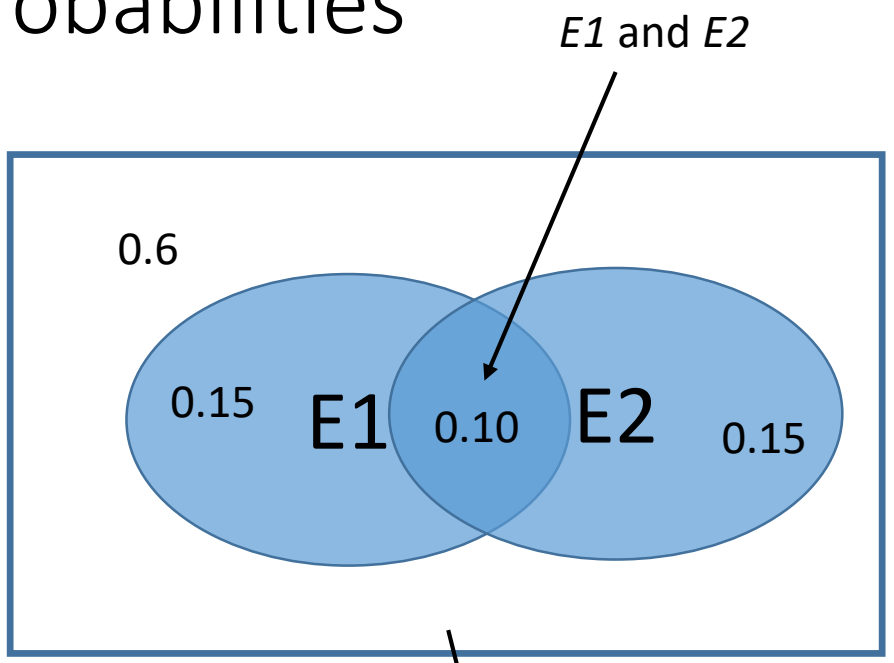
- Sum rule

Pr(E1 or E2)=Pr(E1)+Pr(E2)-Pr(E1,E2)

- Definition of conditional probability

$$\Pr(E1|E2) = \frac{\Pr(E1 \text{ and } E2)}{Pr(E2)} = \frac{\Pr(E1, E2)}{Pr(E2)}$$

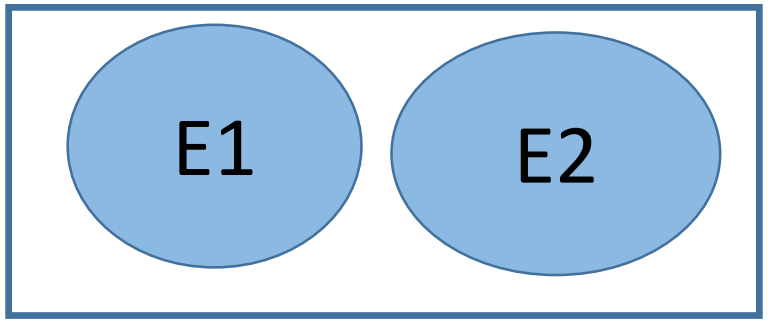
- Product rule $\Pr(E1, E2) = \Pr(E1|E2)\Pr(E2)$



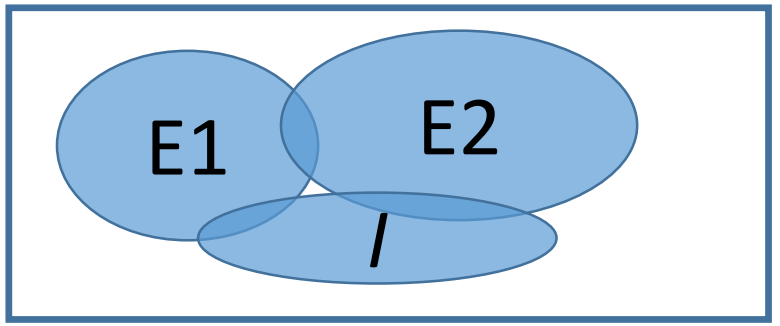
$$\Pr(E1|E2) = \frac{0.10}{0.10 + 0.15} = 0.4$$

Mathematical treatment of probabilities

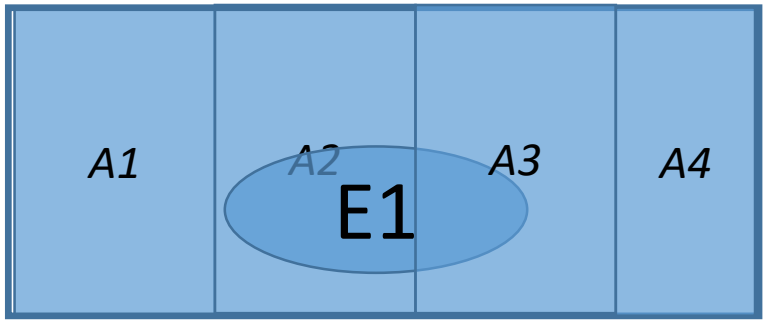
- Definition of statistical independence:
  $E1$ and $E2$ are statistically independent if
  $$\Pr(E1, E2) = \Pr(E1)\Pr(E2)$$



- Conditional independen:
  If E1 and E2 are independent given I, then
  $$\Pr(E1,E2|I)=\Pr(E1|I)\Pr(E2|I)$$



- Marginalization:
  If $A_1, A_2, \ldots, A_n$ partition the sample space, then
  $$\Pr(E|I)= \sum_{i=1}^{n} \Pr(A_i, E|I)$$



Mathematical model

- A description of a system/phenomenon using mathematical language
- A good model represents well the essential aspects of the phenomenon of interest
- A formal tool for logical inference under uncertainty

Hint. Check/refresh the probability axioms, e.g., from
https://en.wikipedia.org/wiki/Probability_axioms

Updating knowledge

Example: draw one ball from a bag of balls

- Let
  - $\theta \in [0,1]$: the proportion of black balls
  - n : the number of draws with replacement
  - y : number of black balls,
- Observation model (aleatory uncertainty)

$$p(y|\theta,n,I) = \text{Bin}(y|\theta,n) \propto \theta^y(1-\theta)^{n-y}$$

- Prior (epistemic uncertainty before observation)

$$p(\theta|I) = \begin{cases} 1, \text{if } 0 < \theta < 1 \\ 0, \text{otherwise} \end{cases}$$

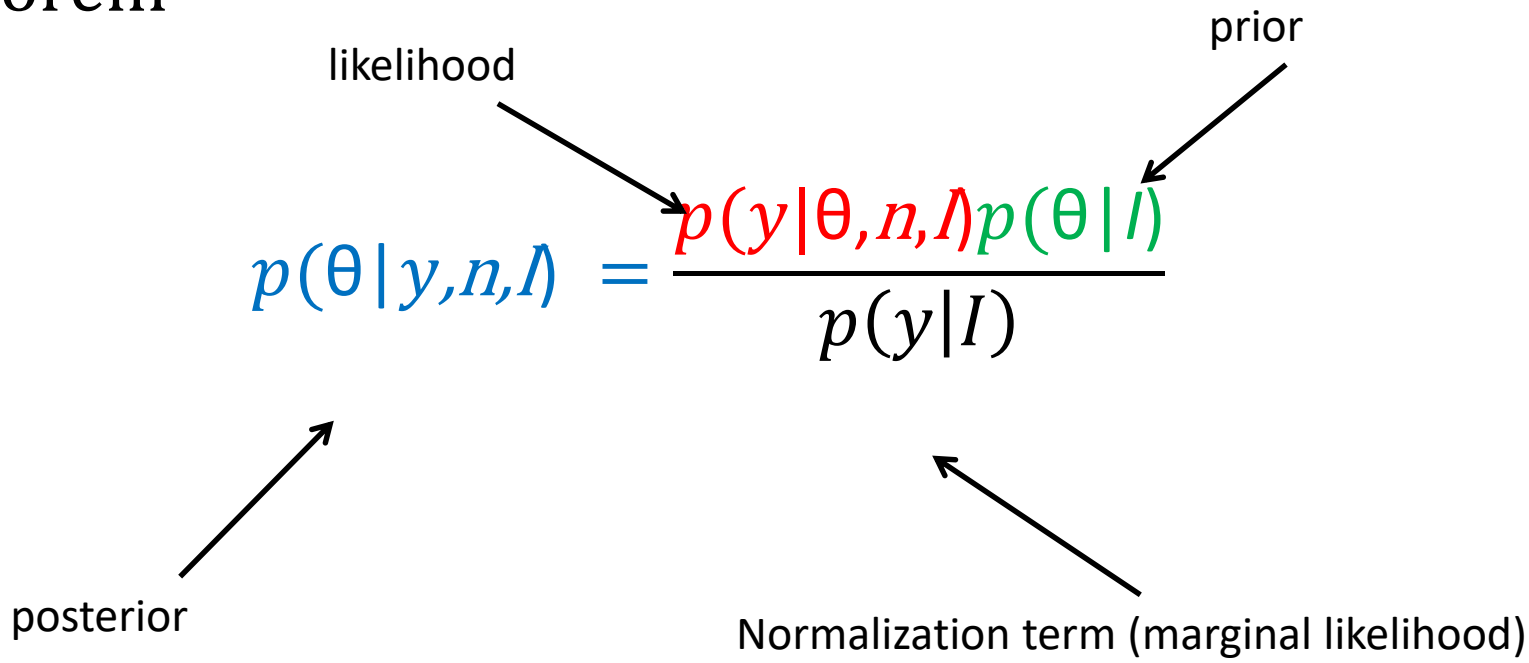
- Posterior (epistemic uncertainty after observation)

$$p(\theta|y,n,I) = ?$$

Example
- Illustration of Bag of balls example

Updating knowledge

- Bayes theorem



Structure of the Bayes theorem

$$p(\theta|y,n,I) = \frac{p(y|\theta,n,I)p(\theta|I)}{p(y|I)}$$

- Observation model, $p(y|\theta,n,I)$
  - Description for observations <u>given</u> variables of interest
  - tells the probability to observe y given particular value of θ
- Likelihood function, $p(y|\theta,n,I)$
  - when y has been observed and fixed, $p(y|\theta,n,I)$ is a function of θ
  - Sometimes denoted by $L(\theta,y)$ to distinguish from the observation model
- The two terms are often used inconsistently

Structure of the Bayes theorem

$$p(\theta|y,n,I) = \frac{p(y|\theta,n,I)p(\theta|I)}{p(y|I)}$$

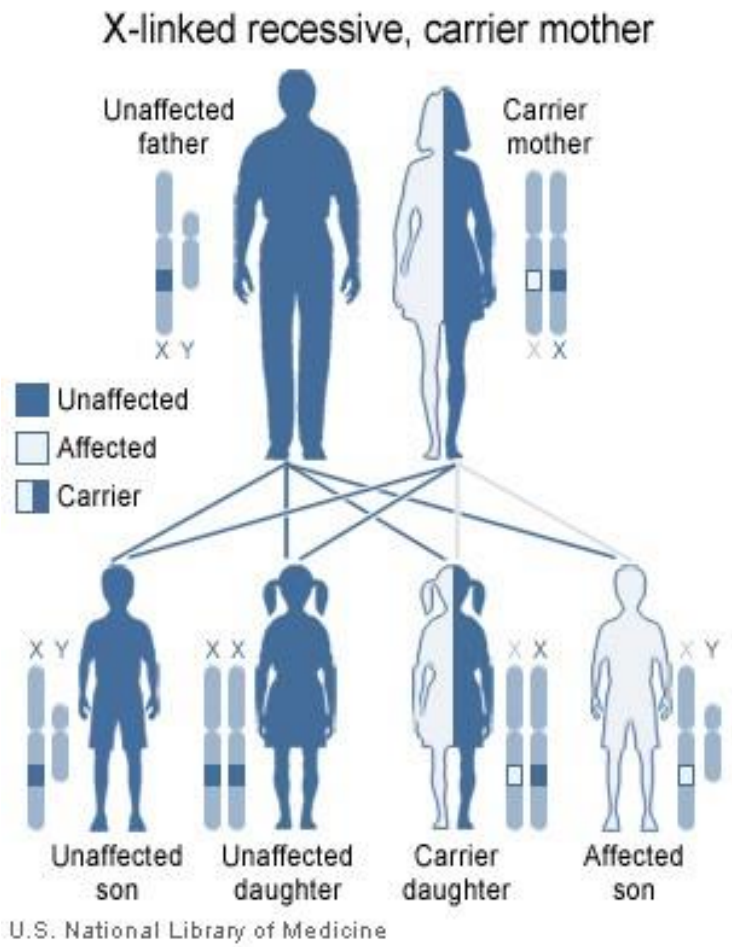
- Prior, $p(\theta|I)$
  - description of what is known about θ before observations
  - epistemic uncertainty before observations
  - the observation model and the prior are inseparable
- Background information, I
  - Present everywhere in the model
    - Likelihood $p(y|\theta,n,I)$
    - Prior $p(\theta|I)$
    - Posterior $p(\theta|y,n,I)$

Structure of the Bayes theorem

$$p(\theta|y,n,I) = \frac{p(y|\theta,n,I)p(\theta|I)}{p(y|I)}$$

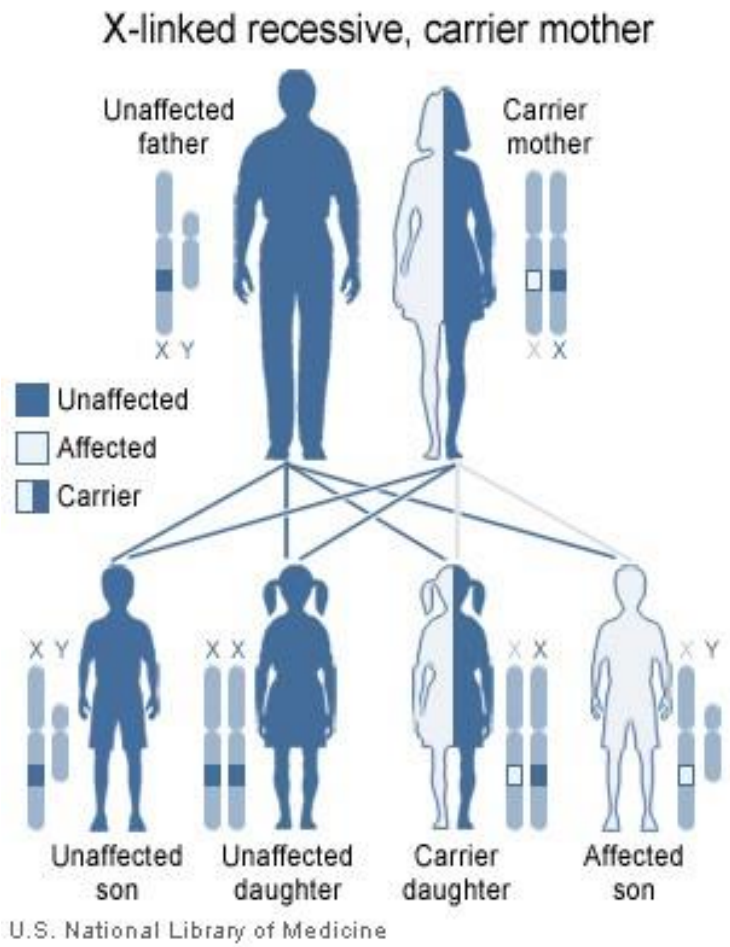
- posterior, $p(\theta|y,n,I)$
  - Combines the background information with information in data
  - Description of what is known about θ <u>after observing data</u>
- Normalization term $p(y|I)$
  - ensures that the overall probability is 1
  - $$p(y|I) = \begin{cases} \int p(y|\theta,n,I)p(\theta|I)\,d\theta, & \text{if } \theta \text{ is continuos} \\ \sum p(y|\theta_i,n,I)p(\theta_i|I), & \text{if } \theta \text{ is discrete} \end{cases}$$

Example: Hemophilia



- Prior
  - An inheritable X-chromosome-linked resessive disease
  - A woman's brother suffers from hemophilia.
  - Her mother and father are healthy
  - How probably the woman is carrier?
- data
  - She has 2 healthy sons
  - How probably the woman is carrier?

Example: Hemophilia



- What is the probability that third son is healthy?

Hint. Check/refresh the probability axioms, e.g., from https://en.wikipedia.org/wiki/Probability_axioms

Prediction

$$p(\theta|y,n,I) = \frac{p(y|\theta,n,I)p(\theta|I)}{p(y|I)}$$

- Let $y = \{y_1, \ldots, y_n\}$ represent observations (data)
- $\tilde{y}$ represent a new observation/measurement (data point) that is yet to be made
  - $\tilde{y}$ is an unknown whose exact value we are uncertain about
  - The prediction for $\tilde{y}$ is

$$p(\tilde{y}|I,y,n) = \sum p(\tilde{y}|\theta_i,I)p(\theta_i|y,I)$$

- Prediction includes both aleatory and epistemic uncertainty

Prediction

$$p(\theta|y,n,I) = \frac{p(y|\theta,n,I)p(\theta|I)}{p(y|I)}$$

- Compare
  - Posterior predictive distribution

$$p(\tilde{y}|I,y) = \sum p(\tilde{y}|\theta_i,I)p(\theta_i|y,I)$$

  - Prior predictive distibution

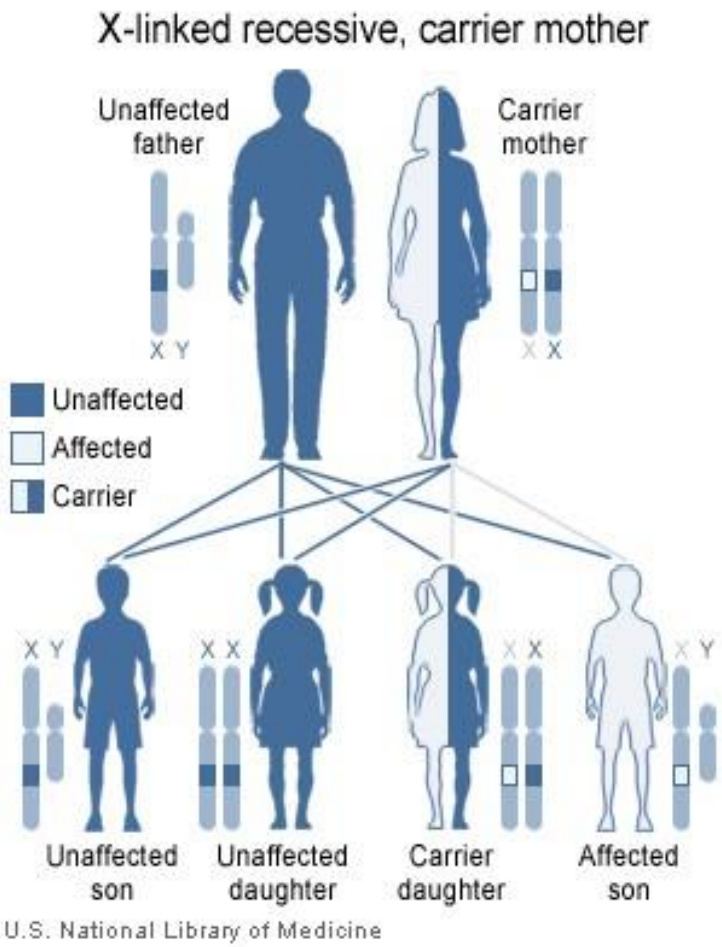
$$p(\tilde{y}|I) = \sum p(\tilde{y}|\theta_i,I)p(\theta_i|I)$$

Sequential use of Bayes formula

- The posterior of earlier analysis can be used as prior for new analysis
  - Old posterior $p(\theta|y,I)$
  - New data $\tilde{y}$
  - New posterior

$$p(\theta|\tilde{y},y,n,I) = \frac{p(\tilde{y}|\theta,I)p(\theta|y,I)}{p(\tilde{y}|y,I)} = \frac{p(\tilde{y},y|\theta,I)p(\theta|I)}{p(\tilde{y},y|,I)}$$

Example: Hemophilia



- The third son was healthy. How probably the woman is carrier?
- Calculate with two approaches
  - use your previous posterior as prior
  - Combine all data (3 healthy sons) and redo the first exercise

About uncertainty

- All information is uncertain (?)
- However, in order to act we have to treat some conditions as known
  - The background knowledge I
  - All models are conditional to I!
- Mathematical description is not perfect
  - ”All models are wrong but some are usefull” (George P. Box)
- Bayesian theory provides a tool to update the uncertain knowledge

About subjectivity

- Aleatory uncertainty is seemingly objective
  - The choice of model for the aleatory uncertainty is subjective
- Epistemic uncertainty is clearly subjective
  - Conditional to analyst's knowledge
  - Two analyst's may have different knowledge ("different I")
- So where is the objectivity of science?
  - All scientific inference is conditional to background assumptions
  - Bayesian analysis enforces one to be transparent about his/her assumptions
  - "objective" background assumptions are obtained through inter-subjectivity
    - Scientists agree on the assumptions ("a common I")

Binomial distribution

- Parameter of interest θ is the probability a of successes in a number of trials which can result in success or failure. Assume fixed number of n trials and y number of successes y$\sim$Bin(n,θ)

$$p(y|\theta)\binom{n}{y}\theta^y(1-\theta)^{n-y} \propto \theta^y(1-\theta)^{n-y}$$

- Used to model the number of successes in a sample of size *n* drawn with replacement from a population of size N.
- If the sampling is carried out without replacement the resulting distribution is a hypergeometric distribution (the draws are not independent). If the total population size, *N*, is much larger than the sample size, *n*, (that is *N≫n)*, the binomial distribution is a good approximation for hypergeometric distribution.

This week

- Uncertainty can be quantified with probabilities
- Bayes theorem is used to update the knowledge
  - posterior $\propto$ likelihood $\times$ prior

$$p(\theta|y,n,I) = \frac{p(y|\theta,n,I)p(\theta|I)}{p(y|I)}$$

Next week

- Technical necessities
  - Basic calculations in R
  - Visualising and calculating probability densities
- practical models
  - Binomial model