Basic ideas behind all calculations

- Prior probability describes your uncertainty about something unknown, usually a model parameter, **BEFORE** taking into account what your data were.
- Posterior probability describes your uncertainty **AFTER** you have included your data.

Basic ideas behind all calculations

- Every probability is conditional to something, minimally at least to your ’*a priori*’ assumptions.
- P(X | Y) = ”probability that X is true, given that Y is true”.
- Statistical model of your data X is some **conditional distribution** $P(X|\theta)$, that is: conditional to the unknown parameter θ which we want to know better.

Basic ideas behind all calculations

- Probabilities can be for **discrete quantities**: e.g. X = ”the number of red balls in a bag of N balls”.
- So, before any data, you might select your **prior probability** as P(X)=1/(N+1) for all possible X. (0,1,2,…,N).
  - Here X is the unknown parameter, and your data will be the observed balls that will be drawn.

Basic ideas behind all calculations

- Probabilities can be about **continuous things** too: e.g. p = ”the proportion of black sheep among ’infinite’ number of sheep”.
- So, before any observations, you might select your **prior probability density** as $\pi(p)$=Uniform(0,1).
  - Here p is the unknown parameter, and your data will be the observed black sheep in a sample that will be collected.

Basic ideas behind all calculations

- All probabilities obey laws of probability calculus, that is: probabilities are ≥0 and ≤1, and add up to one.
  - Discrete: $\sum_X P(X) = 1$
  - Continuous: $\int_\Theta \pi(\theta) d\theta = 1$

  - Any function $\pi(\theta)$ for which the above integral (or sum) is finite, can be made a proper probability density (or point probability function) by dividing the original values by the integral (or sum).
  - (e.g. 3, 4, 7 → 3/14, 4/14, 7/14 which add up to 1)

Basic ideas behind all calculations

- Another important rule is the product rule for the probability of ”X and Y is true”
  - $P(X,Y) = P(X|Y)P(Y) = P(Y|X)P(X)$
  - here ’P’ can be probability density function, if you have continuous quantities.
- The product rule is generalized as
  - $P(X,Y,Z) = P(X|Y,Z)P(Y|Z)P(Z)$
  - Due to symmetry of $P(X,Y,Z)$ it can take several forms.

Basic ideas behind all calculations

- Think of the joint distribution of the unknown parameter θ, and the forthcoming data X: $P(X,\theta)$.
  - **This must obey the product rule too.**
  - **Hence,** from the symmetry of $P(X,\theta)=P(\theta,X)$, we also get Bayes formula: the **probability for the unknown θ, given the observed data X:**
  - $P(\theta \mid X) = P(X \mid \theta)P(\theta) / \text{constant}$

Basic ideas behind all calculations

- $P(\theta \mid X) = P(X \mid \theta)P(\theta)$ / constant
- This is the tool for all Bayesian inference: we aim to compute this conditional probability: the posterior probability.
- Since this probability is conditional to the data X, it gives an **updated probability**, compared to the prior probability $P(\theta)$.
- To calculate this, we need to be able to write what $P(\theta)$ and $P(X|\theta)$ are, **and then just calculate** $P(\theta|X)$ as shown. ('constant' is the normalizing constant).

Basic ideas behind all calculations

- If the prior P(θ) is chosen in a special way, so that it is 'conjugate' to P(X|θ), the posterior distribution comes out as a standard distribution, in the same family of distributions as P(θ).
- **Otherwise**, we need to approximate the posterior distribution as some normal density (is not always accurate) or compute posterior probabilities by numerical integration methods or Monte Carlo methods (→ WinBUGS/OpenBUGS).

Basic ideas behind all calculations

- In some contexts, the prior distribution is chosen to be ’as uninformative’ as possible, e.g. ’flat prior’.
- In some contexts, the prior is chosen to be informative, to represent real background knowledge that is thought to be relevant for the problem.
- The prior can also be constructed hierarchically, conditional to other parameters, thus making connection to other parts of data that provide information on common parameters (e.g. data from previous time interval, or neighbouring geographical areas). This prior cleverly makes use of the rest of the data.