# Methods 3, Week 4:

Bayesian Thinking: Experimental and Implementational Aspects

Maxime Sainte-Marie

## Bayes... again

$Pr(A|B) = Pr(A)\frac{Pr(B|A)}{Pr(B)}$

$$
\begin{aligned}
Pr(B|A) &= \text{Conditional probability to be reversed (Likelihood)} && (1)\\
Pr(A) &= \text{Marginal probability of unknown event (Prior)} && (2)\\
Pr(B) &= \text{Marginal probability of known event} && (3)
\end{aligned}
$$

- This reversing of conditional probabilities is a trivial consequence of the definition of conditional probability
  - It wasn't so in the time of Bayes

### From probability points to probability distributions

- Exact probabilities are rare, multiple factors at play
  - Prostate cancer: Family history, genetics, obesity, height, pesticides, ethnicity, hormones...
  - Representativity of samples, relative frequencies
  - Population quantities are not known or fixed
- Distributions allow to represent uncertainty regarding the true value of specific parameters
  - Assigns probabilities to the different possibles outcomes of a random variable
    * Graphically, a curve where the probability of an outcome is proportional to the height of the curve

### Distributional Bayes

$f(\theta|data) = f(\theta)\frac{f(data|\theta)}{f(data)}$

$$
\begin{aligned}
f(\theta|data) &= \text{posterior distribution of the parameter } \theta && (4)\\
f(data|\theta) &= \text{sampling density for the data} && (5)\\
&\quad \text{(not a proper density function,} && (6)\\
&\quad \text{but proportional to likelihood function)} && (7)\\
f(\theta) &= \text{Prior distribution for the parameter} && (8)\\
f(data) &= \text{Marginal probability of the data} && (9)\\
&\quad \text{(aka marginal likelihood, acts as normalizing constant)} && (10)
\end{aligned}
$$

## $f(\theta)$ = Marginal probability of unknown event (Prior)

- Something only Bayesians have

- Specifies
  - The possible values of the parameter
  - The relative prior plausibility of each
- Informative priors
  - Bigger effect on posterior distribution
  - Effect larger if the data is limited or highly variable

### $f(data|\theta)$ = Likelihood

- Measure of the extent to which available evidence provides support for particular values of a given parameter
- Bridges prior and posterior beliefs by quantifying how well a specific set of parameters fit the observed data
- Bayes Factor = relative likelihood ratio
  - Quantifies the strength of evidence for one hypothesis or model over another, given observed data

  $BF = Pr(data|H_1)/Pr(data|H_2)$

### $f(\theta|data)$ = Posterior

- Reduce uncertainty in parameters by balancing information from our data with our prior knowledge
  - Prior influence = The more informative the prior, the greater our prior certainty, the more influence it has over the posterior
  - Data influence = The more data we have, the more influence the data has over the posterior. Thus, if they have ample data two researchers with different priors will have similar posteriors
- Sequential Bayesian analysis
  - Incremental update of posterior distributions
    - Data order invariant
  - With increasing data, models converge

### $f(data)$ Marginal likelihood

- Normalizing constant making posterior density proper
- When unknown, its calculation rests on the law of total probability

$f(data) = \int f(data|\theta)f(\theta)d\theta$

- Simply scales the posterior density
  - Not taking into account does not alter the relative frequencies of different values of a random variable

$posterior = \frac{prior \cdot likelihood}{normalizing\ constant} \propto prior \cdot likelihood$

# Implementation

- Obtaining the analytical form of posterior distributions is generally intractable
  - Often no closed form expression for the product of the likelihood function and the prior distribution
    - A closed-form expression is a mathematical expression that can be evaluated in a finite number of expressions
      - Involves some kind of limit/infinity
    - Data distribution cannot be written down on paper
    - Exception: Conjugate priors and prior mixtures

## Approxating posterior distributions

- Posterior distributions most often approximated

  - According to the **Law of Large Numbers**, if we draw enough samples from any probability distribution, then the **Central Limit Theorem** says that even if we don't know what it should be then we can compute a Monte Carlo estimate to arrive at a good approximation of it

### MCMC Posterior Distribution Approximation

- Computer–driven sampling method allowing to characterize a distribution without knowing all of the distribution's mathematical properties by randomly sampling values out of the distribution.
- Monte-Carlo = Practice of estimating distribution properties through random samples examination
  - As the number of samples increases, it converges to an accurate representation of the posterior
- Markov Chain = Special sequential process, where each random sample is used as a stepping stone to generate the next random sample (hence the chain)
  - Similar to making pancakes!

## Basic steps to Bayesian Analysis (Kruschke, 2015: 25)

1. Identify the data relevant to the research questions (predictor & predicted variables).
2. Define a descriptive model for the relevant data.
3. Specify a prior distribution on the parameters.
4. Use Bayesian inference to re-allocate credibility across parameter values.
5. Interpret the posterior distribution with respect to theoretically meaningful issues (assuming that the model is a reasonable description of the data; see next step).
6. Check that the posterior predictions mimic the data with reasonable accuracy (posterior predictive check)