Basics of bayesian statistics in R

Prasanna Bhogale

BerlinR meetup

12 December 2018

Contents

The model

- simplified system that approximates the real world
- hypothesis about how a particular dataset might be generated
- sometimes, mathematically or computationally tractable set of equations

Science often involves constructing models of the real world and testing them.

$$\text{Model} \longrightarrow \text{Predictions}$$

but, how does one construct a good model ?

Inference

$$\text{Data} \longrightarrow \text{Model} \longrightarrow \text{Predictions}$$

- Practitioner identifies a class of models with some parameters θ
- Inference consists of using data to estimate the parameter vector θ
- Bayesian inference : estimating the full probability distribution of θ given available data

Bayes theorem

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

- $P(\theta|D)$: Posterior probability
- $P(D|\theta) = l(\theta|D)$: Likelihood
- $P(\theta)$: Prior
- $P(D) = \int P(D|\theta)P(\theta)d\theta$: Evidence (normalization factor)

Inference example - water on earth

- Task : estimating the proportion of the globe covered by water by randomly dropping pins and checking if they are on water, or land
- So, we have a binomial process where probability of dropping the pin on water is p_w and $p_l = 1 - p_w$. Our parameters are $\theta = \{p_l\}$.
- Assume we know nothing, so we assume a uniform prior for p_l

  `see land_water_bayesian.Rmd`

as seen in the bayesian stats course by Richard Mcelreath

Why Bayesian inference is hard

- For high dimensional θ likelihood $P(D|\theta)$ can be hard to compute, since the volume of θ space where this is significant is small
- The Evidence $\int P(D|\theta)P(\theta)d\theta$ is similarly difficult
- Smart samplers (like Hamiltonian Monte Carlo) are needed
- Choosing informative priors is an art
- Bayesian inference is a LOT more computationally expensive than statistical inference

The GPU revolution

- High degree of parallelism
- Rapid development of algorithms and implementations leveraging parallelism
- Mature, easy to use bayesian tools built on top of gradient descent infrastructure (Edward, Pyro)
- in R, apart from JAGS, stan, BUGS which are well established and mostly, DON'T use the GPU there is Greta which does.

Thank you

Prasanna Bhogale

pbhogale@gmail.com