Introduction to probability and statistics

Master in Cognitive Science 2025-2026

Lecture 3

October, 2025

Objectives

- Introduction to Bayesian statistics
- First example: Inference on a proportion "by hand"
- Conjugate priors
- Markov Chain Monte Carlo (MCMC)
- Inference using Jags

Rreadings: Lee and Wagenmakers, Bayesian Cognitive Modeling A practical Course. Chapter 1: The basics of Bayesian analysis

Introduction to Bayesian Statistics

Imagine we want to *estimate* a parameter θ, for example, the interspike time interval. However, we do not observe the parameter θ directly. The uncertainty or "degree of belief" about θ is quantified using probability, by assuming that θ is a random variable following some distribution $p(\theta)$ (the **prior**). Once data D are observed (for instance, from many trials of a neuron's spike train), the prior information or beliefs are updated to form the **posterior** distribution $p(\theta \mid D)$.

Bayes' rule

Bayes' rule tells us how to combine the information given by the data (the **likelihood** $p(D|\theta)$) with the prior knowledge to get the posterior distribution

Bayes' rule

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}. \qquad (1)$$

Notation: Since $P(D)$ does not involve on the parameter of interest $P(D)$, we write equation (1) as

$$p(\theta|D) \propto p(D|\theta)p(\theta). \qquad (2)$$

The symbol $\propto$ is to say that "it is proportiona tol" or they are equal up to normalizing constant.

Inference of a proportion

Bayesian model for Binomial data with uniform prior.
Suppose you pass an exam with $n = 100$ questions, and you get
$y = 70$ right questions. We want to model your score
$(\frac{\#\text{correct answers}}{\#\text{total questions}})$ θ with a discrete uniform prior over the set of
values $\{\frac{0}{100}, \frac{1}{100}, \frac{2}{100}, \cdots, \frac{100}{100}\}$.
What is the posterior probability of your score?

```
n = 50
y = 35
theta = seq(0, 100)/100
prior = rep(1/101, length = 101)
like = choose(n, y) * (theta^y) * (1 - theta)^(n - y)
post = prior * like / sum(like * prior)

plot(theta, post, type = "h", col = 2)
lines(theta, prior, type = "h", col = 4)
legend("topright", c("prior", "posterior"))
```
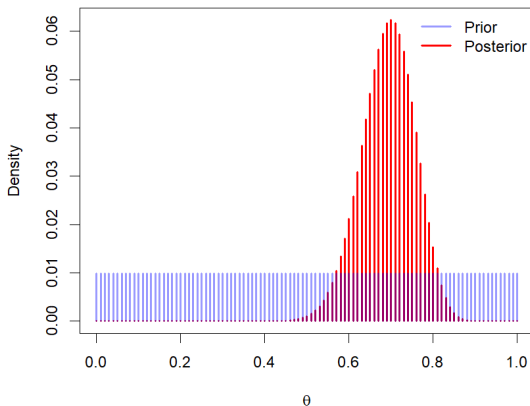
Prior and Posterior distributions

Figure 1: Prior and Posterior distribution of exam score

Prior and Posterior probability

```
interval = (findInterval(theta,c(0.6,0.8)) == 1)
cat("prior probability of theta in [0.6;0.8]= ",
  sum(prior*interval))

## prior probability of theta in [0.6;0.8]=  0.1980198

interval = (findInterval(theta,c(0.6,0.8)) == 1)
cat("prior probability of theta in [0.6;0.8]= ",
  sum(post*interval))

## prior probability of theta in [0.6;0.8]=  0.8865295
```

Conjugate Prior

> Given a likelihood distribution $p(x \mid \theta)$, if the posterior
> distribution belongs to the same family as the prior
> distribution, then the prior and posterior are said to be
> **conjugate** with respect to the chosen likelihood function.

Example: Suppose you didn't pass the official exam yet, and
we know that you prepared 4 out of the 5 chapters you had to
read. When you passed the practice exam you got y correct
answers out of n.

For the prior, we use a Beta distribution Beta(4, 1), which represents **4 "successes" and 1 "failure"** in our prior knowledge.

After observing data y out of n trials, the posterior is also a Beta distribution (due to conjugacy) with updated parameters:

$$\theta \mid y \sim \text{Beta}(y + 4, \, n - y + 1),$$

where the prior counts are added to the observed counts, reflecting how the data updates our prior beliefs.

$$\text{Likelihood} \begin{cases} y|\theta \sim Bin(n, \theta) \\ p(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}. \end{cases}$$

$$\text{Prior} \begin{cases} \theta \sim Be(4, 1) \\ \pi(\theta) = 4x^3 \cdot 1_{\{\theta \in (0,1)\}}. \end{cases}$$

$$\text{Posterior} \begin{cases} \theta|y \sim Be(y + 4, n - y + 1) \\ \pi(\theta|y) \propto \theta^{y+3}(1-\theta)^{n-y}. \end{cases}$$

Some Examples of Conjugate Priors

Likelihood	Conjugate Prior (and posterior)	Prior Hyperparam.	Posterior Hyperparam.
Bernoulli(p)	Beta(α, β)	α, β	$\alpha + \sum x_i, \beta + n - \sum x_i$
Binomial(n, p)	Beta(α, β)	α, β	$\alpha + \sum x_i, \beta + \sum(n - x_i)$
Poisson(λ)	Gamma(α, β)	α, β	$\alpha + \sum x_i, \beta + n$
Exponential(λ)	Gamma(α, β)	α, β	$\alpha + n, \beta + \sum x_i$

Can we always compute the distribution of the posterior?

In most real-world case studies, it is often difficult, and sometimes impossible, to obtain a closed-form analytic expression for the posterior distribution for many practically interesting choices of likelihood and prior. This limitation restricted the use of Bayesian inference for many years. The introduction of MCMC methods dramatically improved this situation by providing a technique to *sample* from the posterior distribution without requiring its analytic form.

Markov Chain Monte Carlo (MCMC)

- MCMC constructs a **Markov chain** whose *target distribution is the posterior distribution* of the parameters.
- Under suitable regularity conditions, the Markov chain converges to this target (equilibrium) distribution.
- Once converged, MCMC generates dependent samples from the posterior distribution.
- Inference is then performed using the empirical distribution of these samples.

Monte Carlo approximation of posterior characteristics

We are typically interested in quantities of the form

$$\mathbb{E}_{\theta|x}(g(\theta)) = \int_{\Theta} g(\theta)\pi(\theta|x)d\theta.$$

Since $\pi(\theta|x)$ is a density, we can use a sample $\theta_1, \ldots, \theta_M$ generated from the posterior density $\pi(\theta|x)$ and approximate by an empirical average

$$\bar{g}_M := \frac{1}{M} \sum_{m=1}^{M} g(\theta_m).$$

By the SLLN, $\bar{g}_M$ converges a.s. to $\mathbb{E}_{\theta|x}(g(\theta))$.

JAGS (Just Another Gibbs Sampler)

Download the latest version of JAGS at:
https://sourceforge.net/projects/mcmc-jags/files/
latest/download

Install the library `Rjags` in R:

```
install.packages("Rjags")
```

Setting Up the Bayesian Model

We want to model your score on an exam with 10 questions.
You passed last years exam and got 7 correct answers out of 10.

Model specification

```
model_string <- "
model {
  # Likelihood
  k ~ dbin(theta, n)

  # Prior
  theta ~ dunif(0,1)
}
"
```

The observed data in this case is:

Observed data

```
n <- 10 # Number of questions
k <- 7  # Number of correct answers
data_list <- list(k = k, n = n)
```

Then we initialize the JAGS sampler:

Model Initialization

```
model <- jags.model(
  textConnection(model_string),
  data_list,
  n.chains = 3,
  n.adapt = 1000
)
```

Posterior Sampling

```
# Burn-in
update(jags_model, 1000)
mcmc_samples <- coda.samples(model,
    variable.names = c("theta"), n.iter = 5000)
```

The variable `mcmc_samples` contains draws from the posterior distribution.

Exercise

1. Show that $\mathrm{Unif}(0, 1) = \mathrm{Beta}(1, 1)$.

2. Compare the empirical cumulative distribution function of the generated samples with the theoretical CDF of $\mathrm{Beta}(8, 4)$.

3. What are your conclusions?