

# Bayesian Statistics Primer

## Rigorous Foundations with Practical Applications

Sarah Urbut, MD PhD

MGH and Broad Institute, Univ. of Chicago, DFCI

March 26, 2025

# Outline

- 1 Likelihood Ratio and Bayesian Inference
- 2 Mixture Models
- 3 Markov Chain Monte Carlo
- 4 Distributions and Multivariate Normal
- 5 Applications in Medical Research

# Bayes' Theorem: The Foundation

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \quad (1)$$

Where:

- $P(H|D)$  is the posterior probability of hypothesis  $H$  given data  $D$
- $P(D|H)$  is the likelihood of observing data  $D$  under hypothesis  $H$
- $P(H)$  is the prior probability of hypothesis  $H$
- $P(D)$  is the marginal likelihood (evidence)

# Likelihood Ratio for Model Comparison

$$LR = \frac{P(D|H_1)}{P(D|H_0)} \quad (2)$$

- Quantifies evidence in favor of  $H_1$  over  $H_0$
- Interpretation:
  - $LR = 1$ : Data equally likely under both hypotheses
  - $LR > 1$ : Data more likely under  $H_1$
  - $LR < 1$ : Data more likely under  $H_0$
- Jeffreys' scale for interpreting evidence strength

## R Example: Likelihood Ratio

```
# H0: data comes from N(0, 1)
# H1: data comes from N(2, 1)

# Generate data (true mean between hypotheses)
set.seed(123)
true_mean <- 1.5
data <- rnorm(30, mean = true_mean, sd = 1)

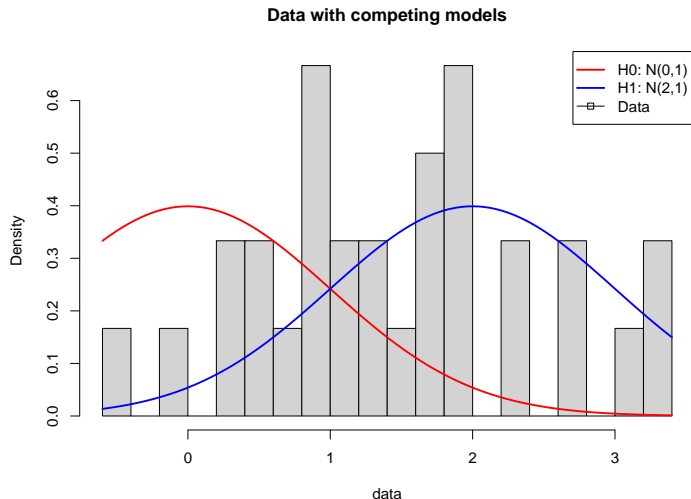
# Calculate likelihoods
likelihood_H0 <- prod(dnorm(data, mean = 0, sd = 1))
likelihood_H1 <- prod(dnorm(data, mean = 2, sd = 1))

# Calculate likelihood ratio
LR <- likelihood_H1 / likelihood_H0

# Log likelihood ratio (more stable)
log_LR <- sum(dnorm(data, mean = 2, sd = 1, log = TRUE)) -
  sum(dnorm(data, mean = 0, sd = 1, log = TRUE))

cat("Log likelihood ratio:", log_LR, "\n")
```

# Likelihood Ratio Visualization



data with competing models

Histogram of

# Introduction to Mixture Models

$$p(x) = \sum_{k=1}^K \pi_k f_k(x|\theta_k) \quad (3)$$

Where:

- $p(x)$  is the probability density of the mixture
- $\pi_k$  are the mixing weights ( $\sum_{k=1}^K \pi_k = 1$ )
- $f_k(x|\theta_k)$  are the component densities
- $\theta_k$  are the parameters of each component

# EM Algorithm for Mixture Models

- 1 **Initialize:** Choose starting values for  $\pi_k$  and  $\theta_k$
- 2 **E-step:** Calculate responsibilities

$$\gamma_{ik} = \frac{\pi_k f_k(x_i | \theta_k)}{\sum_{j=1}^K \pi_j f_j(x_i | \theta_j)} \quad (4)$$

- 3 **M-step:** Update parameters

$$\pi_k^{new} = \frac{1}{n} \sum_{i=1}^n \gamma_{ik} \quad (5)$$

$$\theta_k^{new} = \arg \max_{\theta_k} \sum_{i=1}^n \gamma_{ik} \log f_k(x_i | \theta_k) \quad (6)$$

- 4 Repeat E and M steps until convergence



## R Example: Gaussian Mixture Model

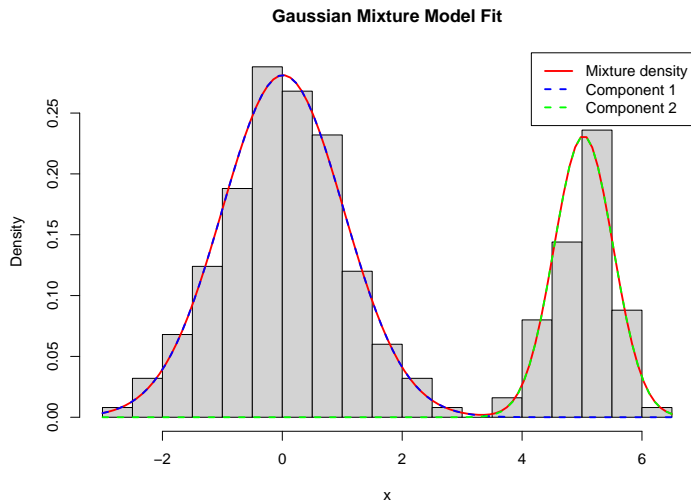
```
# Fitting a Gaussian mixture model using EM
library(mixtools)

# Generate data from mixture of two normals
set.seed(123)
n <- 500
true_pi <- c(0.7, 0.3) # Mixing proportions
true_means <- c(0, 5) # Component means
true_sds <- c(1, 0.5) # Component standard deviations

# Generate samples
z <- sample(1:2, size = n, replace = TRUE, prob = true_pi)
x <- rnorm(n, mean = true_means[z], sd = true_sds[z])

# Fit mixture model using EM algorithm
fit <- normalmixEM(x, k = 2)
```

# Gaussian Mixture Model Results



Fitted

Gaussian mixture model

# Metropolis-Hastings Algorithm

Goal: Sample from a target distribution  $p(x)$

- 1 Initialize  $x^{(0)}$
- 2 For  $t = 0, 1, 2, \dots$ :
  - 1 Propose  $x' \sim q(x'|x^{(t)})$  from proposal distribution
  - 2 Calculate acceptance ratio:

$$\alpha = \min \left( 1, \frac{p(x')q(x^{(t)}|x')}{p(x^{(t)})q(x'|x^{(t)})} \right) \quad (7)$$

- 
- 
- 3 Accept proposal with probability  $\alpha$ :

$$x^{(t+1)} = \begin{cases} x' & \text{with probability } \alpha \\ x^{(t)} & \text{with probability } 1 - \alpha \end{cases} \quad (8)$$

# Gibbs Sampling

Special case of Metropolis-Hastings where proposals are always accepted

- ① Initialize  $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_d^{(0)})$
- ② For  $t = 0, 1, 2, \dots$ :
  - ① Sample  $x_1^{(t+1)} \sim p(x_1 | x_2^{(t)}, x_3^{(t)}, \dots, x_d^{(t)})$
  - ② Sample  $x_2^{(t+1)} \sim p(x_2 | x_1^{(t+1)}, x_3^{(t)}, \dots, x_d^{(t)})$
  - ③  $\vdots$
  - ④ Sample  $x_d^{(t+1)} \sim p(x_d | x_1^{(t+1)}, x_2^{(t+1)}, \dots, x_{d-1}^{(t+1)})$

## R Example: Metropolis-Hastings

```
# Metropolis-Hastings for bimodal distribution
# Define target distribution (unnormalized)
target <- function(x) {
  0.5 * dnorm(x, -3, 1) + 0.5 * dnorm(x, 3, 1)
}

# Metropolis-Hastings algorithm
metropolis_hastings <- function(n_samples, proposal_sd) {
  samples <- numeric(n_samples)
  x_current <- 0 # Starting point

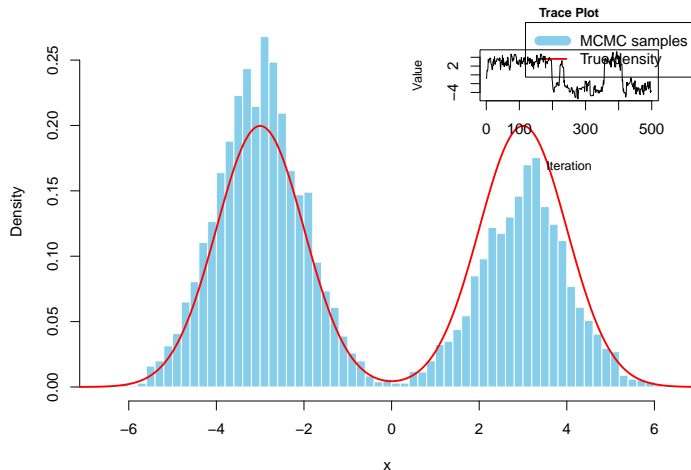
  for (i in 1:n_samples) {
    # Propose new value
    x_proposed <- rnorm(1, mean = x_current, sd = proposal_
      sd)

    # Calculate acceptance ratio
    ratio <- target(x_proposed) / target(x_current)

    # Accept or reject
    if (runif(1) < ratio) {
```

# MCMC Results

Metropolis-Hastings Sampling of Bimodal Distribution



Metropolis-Hastings sampling of bimodal distribution

# Beta and Dirichlet Distributions

## Beta Distribution:

$$p(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (9)$$

Where  $B(\alpha, \beta)$  is the Beta function.

## Dirichlet Distribution (multivariate generalization):

$$p(\mathbf{x}|\alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^K x_i^{\alpha_i-1} \quad (10)$$

Where  $\mathbf{x} = (x_1, \dots, x_K)$  with  $\sum_{i=1}^K x_i = 1$  and  $B(\alpha)$  is the multivariate Beta function.

# Multivariate Normal Distribution

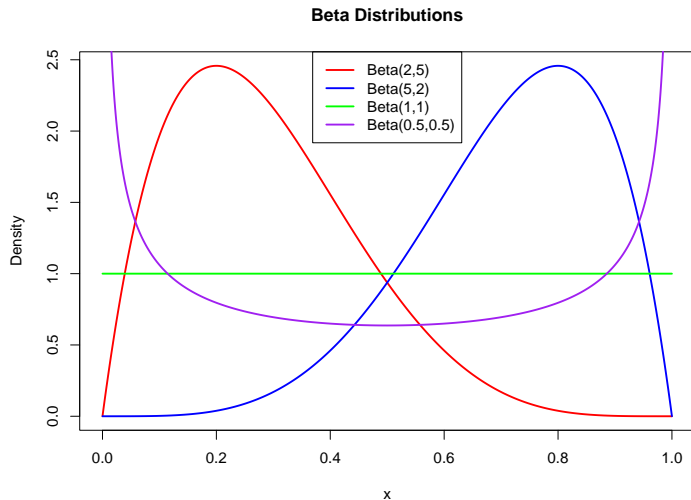
$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (11)$$

## Key Properties:

- Linear transformations: If  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ , then  $\mathbf{y} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$
- Marginal distributions are normal
- Conditional distributions are normal



# Beta Distribution Examples

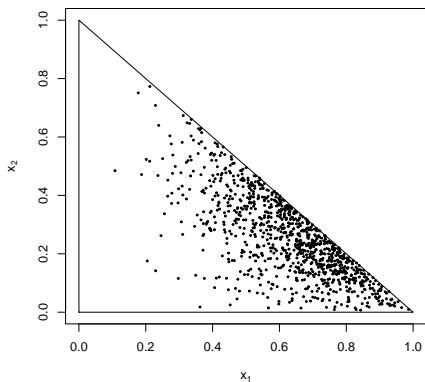


Various Beta

distributions

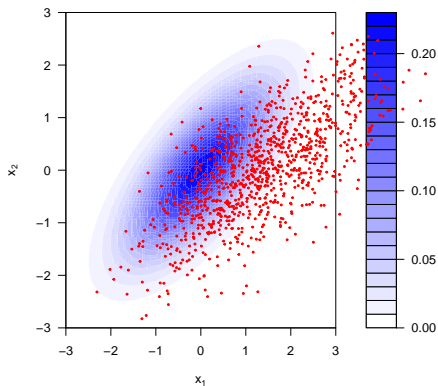
# Dirichlet and Multivariate Normal

Dirichlet(5,2,1) Samples on 2D Simplex



Dirichlet samples on simplex

Multivariate Normal Density



Multivariate normal density

# Bayesian Analysis of Clinical Trial Data

- **Traditional approach:** Fixed significance level (e.g.,  $p < 0.05$ )
- **Bayesian approach:** Calculate posterior probability of effect

**Example:** Treatment effect on cardiac outcomes

$$P(\theta > 0 | \text{data}) = \text{Probability treatment is beneficial} \quad (12)$$

$$P(|\theta| > \delta | \text{data}) = \text{Probability of clinically significant effect} \quad (13)$$

## Advantages:

- Incorporates prior knowledge from previous studies
- Provides direct probability statements about parameters
- Allows for sequential updating as data accumulates

# Factor Analysis for Disease Subtypes

## Model:

$$\mathbf{x}_i = \mathbf{\Lambda} \mathbf{f}_i + \epsilon_i \quad (14)$$

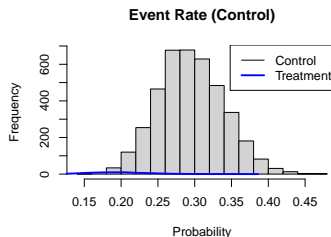
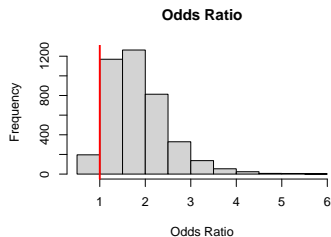
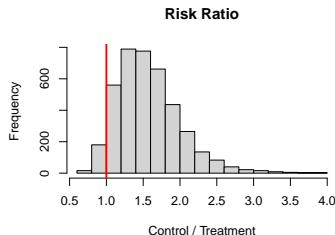
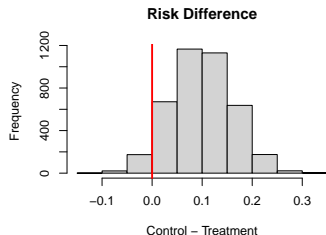
Where:

- $\mathbf{x}_i$  is the vector of observed variables for patient  $i$
- $\mathbf{\Lambda}$  is the factor loading matrix
- $\mathbf{f}_i$  are the latent factors
- $\epsilon_i$  is the error term

## Bayesian approach:

- Place priors on  $\mathbf{\Lambda}$  and error variances
- Infer posterior distributions of factors
- Identify disease subtypes based on factor patterns

# Clinical Trial Results

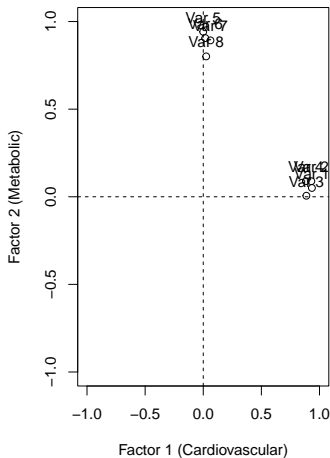


Posterior

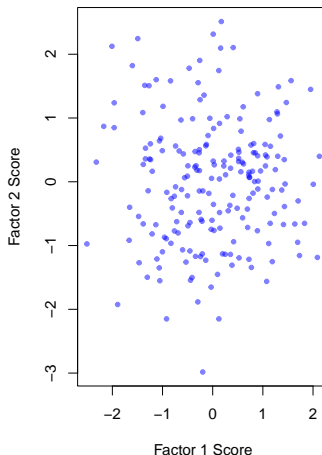
distributions of treatment effects

# Factor Analysis Results

Factor Loadings



Patient Distribution by Factors



Patient

distribution by disease factors

# Summary and Resources

## Key Concepts Covered:

- Likelihood ratio and Bayesian inference
- Mixture models and EM algorithm
- MCMC methods (Metropolis-Hastings and Gibbs sampling)
- Important distributions (Beta, Dirichlet, Multivariate Normal)
- Applications in medical research

## Resources for Further Learning:

- Five Minute Statistics:  
<http://stephens999.github.io/fiveMinuteStats/>
- Bayesian Data Analysis (Gelman et al.)
- Statistical Rethinking (McElreath)
- R packages: rstan, rstanarm, brms, bayesplot

# Thank you!

- 1 Giovanni Parmigiani, PhD and Sasha Gusev, PhD
- 2 Matthew Stephens, PhD
- 3 Pradeep Natarajan, MD MMSc