Lecture 2, September 25th, 2025

1. Bayesian inference: general formulation + examples
2. Monte Carlo algorithms

Bayesian inference (generic setup)

Inputs:

- $\underline{y}$ — vector of data (could be a matrix or a tensor)
- $\underline{\theta}$ — vector of model parameters (e.g., θ - proportion)
- $p(\underline{y} \mid \underline{\theta})$ — density/prob defined by the data generated model (e.g., $Pr(y \mid \theta)$ - Binomial)
- $p(\underline{\theta})$ — density of the prior distribution that encodes our prior beliefs about $\underline{\theta}$ (e.g., $p(\theta)$ - Beta density)

Output:

$$p(\underline{\theta} \mid \underline{y}) = \frac{p(\underline{y} \mid \underline{\theta})\, p(\underline{\theta})}{\int p(\underline{y} \mid \underline{\theta})\, p(\underline{\theta})\, d\underline{\theta}}$$

— posterior density of model parameters

In practice, we use the following summaries of the posterior distribution:

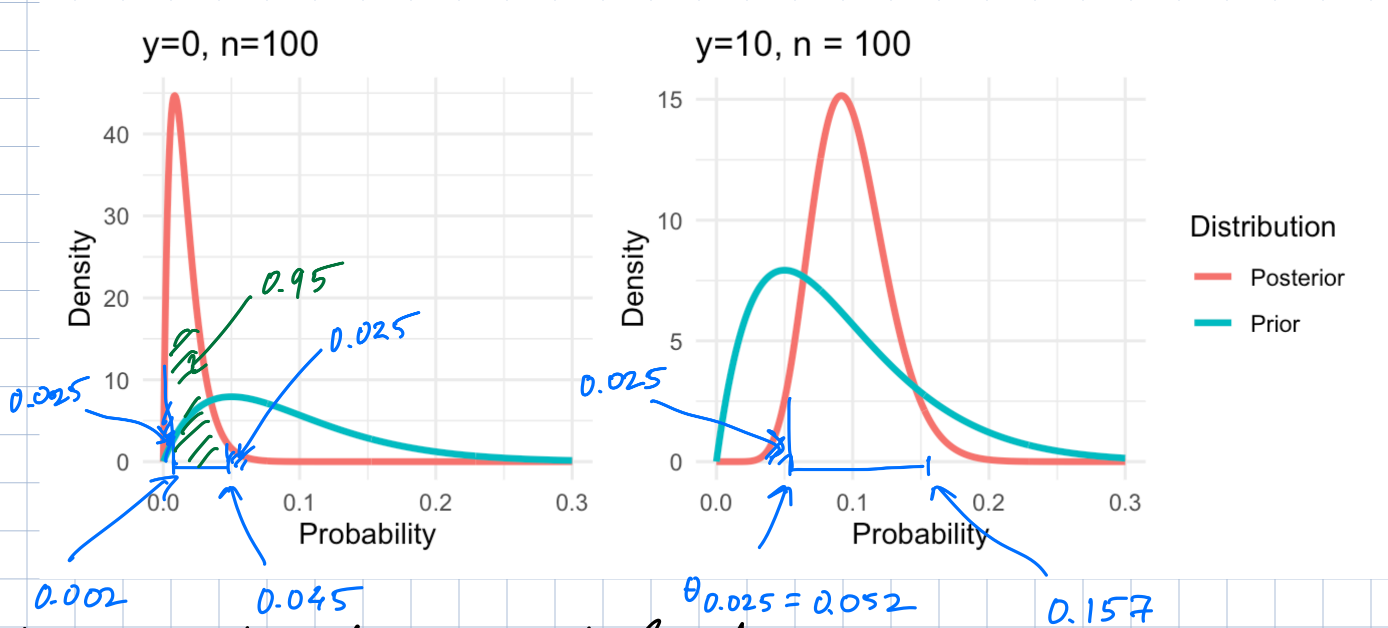
1) point estimation: median, mode, or mean of $p(\underline{\theta} \mid \underline{y})$. For example,

$$\hat{\underline{\theta}} = \underbrace{E(\underline{\theta} \mid \underline{y})}_{\text{posterior mean}} = \int \underline{\theta}\, p(\underline{\theta} \mid \underline{y})\, d\underline{\theta}$$

More concretely, in our proportion estimation example

$\theta \mid y \sim \text{Beta}(\alpha+y,\ n-y+\beta) \Rightarrow E(\theta \mid y) = \frac{\alpha+y}{\alpha+\cancel{y}+n-\cancel{y}+\beta} = \frac{\alpha+y}{\underbrace{\alpha+\beta}_{\text{prior contribution}}+n}$

$E(\theta \mid y) = \frac{2+0}{22+100} = \frac{2}{122} \approx 0.016 \qquad E(\theta \mid y) = \frac{2+10}{22+100} = \frac{12}{122} \approx 0.098$



2) uncertainty quantification:

Bayesian credible intervals for univariate parameters: $(\theta_{i,q/2},\ \theta_{i,1-q/2})$, where

$Pr(\theta_i < \theta_{i,q/2} \mid y) = q/2$ and $Pr(\theta_i < \theta_{i,1-q/2}) = 1-\frac{q}{2}$

$Pr(\theta_i > \theta_{i,1-q/2}) = \frac{q}{2}$

posterior quantiles

In our proportion estimation example:

$y=0,\ n=100,\ \alpha=2,\ \beta=20,\ q=0.05$

this gives us 95% credible interval

$(\theta_{0.025},\ \theta_{0.975}) = (0.002,\ 0.045)$

$y=10,\ n=100,\ \alpha=2,\ \beta=2,\ q=0.05$

$(\theta_{0.025},\ \theta_{0.975}) = (0.052,\ 0.157)$

There is no formula for quantiles, so these calculations must be done numerically (qbeta(·) in R)

Example : multiple linear regression

data: $\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ $\quad X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ & & \boxed{\underline{x}_i^T} & \\ \vdots & & & \\ 1 & x_{n1} & & x_{np} \end{pmatrix}$ ← we condition on covariates/predictors, so we do not model them

only y vector is random

n: sample size

p: number of covariates/predictors

Data generating model: $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$,

where $\varepsilon_1, \dots, \varepsilon_n \overset{iid}{\sim} N(0, \sigma^2)$ or

in vector form $y_i = \underline{x}_i^T \underline{\beta} + \varepsilon_i$

parameters: $\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \sigma^2$

Likelihood: $p(\underline{y} \mid \underline{\beta}, \sigma^2) =$ [independence assumption] =

$$= \prod_{i=1}^{n} p(y_i \mid \underline{\beta}, \sigma^2), \text{ where}$$

$$p(y_i \mid \underline{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \underline{x}_i^T \underline{\beta})^2}{2\sigma^2}}$$

priors: many choices, but let's use something simple:

$\beta_0, \beta_1, \dots, \beta_p \overset{iid}{\sim} N(0, s^2)$ ← constant of our choice

$\sigma^2 \sim \text{Gamma}(\cdot, \cdot)$

posterior distribution / density:

$$p(\underline{\beta}, \sigma^2 \mid \underline{y}) = \frac{p(\underline{y} \mid \underline{\beta}, \sigma^2)\, p(\underline{\beta})\, p(\sigma^2)}{\int \text{—} \shortparallel \text{—} \, d\underline{\beta}\, d\sigma^2}$$

Bad news: Bayesian inference requires calculating high dimensional integrals

Good news: We have algorithms for this task

Monte Carlo Integration

First, let's review the concept of mathematical expectation:

1) for discrete random variable X

$$E[g(X)] = \sum_{k=1}^{n} g(x_k)\, P(X = x_k)$$

1^{st} moment: $g(x) = x$

2^{nd} moment $g(x) = x^2$

2) for (absolutely) continuous random variable X:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

<u>Example</u>: exponential random variable

$$f(x) = \lambda e^{-\lambda x} \cdot 1_{\{x \geq 0\}}, \text{ where } \lambda > 0 \text{ - rate parameter}$$

$X \sim Exp(\lambda)$ $\quad E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{0}^{\infty} x \cdot \lambda \cdot e^{-\lambda x} = [\text{integration by parts}] = \ldots = \frac{1}{\lambda}$

Strong Law of Large Numbers (SLLN)

Let $X_1, X_2, \ldots$ be independent and identically distributed (iid) random variables with

$\mu = E(X_1) < \infty$ Then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu$$

SLLN says that empirical averages of iid random variables converge to the theoretical average/expectation

Monte Carlo Integration

<u>Objective</u>: $E[h(X)] = \int h(x) f(x)\, dx$, where X is a random variable with probability density function $f(x)$ or

$E(h(X)) = \sum_{k=1}^{\infty} h(x_k)\, p_k$, where X is a discrete random variable with prob. mass function $p_1, p_2, \ldots$

If $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ and $E(h(X_1)) < \infty$, then

$$\text{SLLN} \Rightarrow \underbrace{\frac{1}{n} \sum_{i=1}^{n} h(X_i)}_{\overline{h}_n} \approx E(h(X_i))$$

Example: second moment of the Beta distribution

Recall that in our estimating proportions example, one setting had:

$$\theta \sim \text{Beta}(2, 20)$$

$$y = 10, \quad n = 100$$

$$\theta \mid y \sim Beta(12, 110)$$

Objective : $E(\theta^2 \mid y)$ - second moment

In this case we can derive a formula :

$$X \sim Beta(\alpha, \beta) \Rightarrow E(X^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}$$

This result tells us that

$$E(\theta^2 \mid y) = \frac{12(12+1)}{(\quad)(\quad)} = 0.01$$

Let's see if we can get to the same number using Monte Carlo integration.

Our game plan :

1) generate $\theta_1, \ldots, \theta_m \sim Beta(12, 110)$

2) return $\sum_{i=1}^{m} \theta_i^2 \approx E(\theta^2 \mid y)$

computer demo here