Introduction to Bayesian Inference

Contents

1 Introduction to Bayesian inference

Configuring R

Functions from these packages will be used throughout this document:

```
library(conflicted) # check for conflicting function definitions
# library(printr) # inserts help-file output into markdown output
library(rmarkdown) # Convert R Markdown documents into a variety of formats.
library(pander) # format tables for markdown
library(ggplot2) # graphics
library(ggfortify) # help with graphics
library(dplyr) # manipulate data
library(tibble) # `tibble`s extend `data.frame`s
library(magrittr) # `%>%` and other additional piping tools
library(haven) # import Stata files
library(knitr) # format R output for markdown
library(tidyr) # Tools to help to create tidy data
library(plotly) # interactive graphics
library(dobson) # datasets from Dobson and Barnett 2018
library(parameters) # format model output tables for markdown
library(haven) # import Stata files
library(latex2exp) # use LaTeX in R code (for figures and tables)
library(fs) # filesystem path manipulations
library(survival) # survival analysis
library(survminer) # survival analysis graphics
library(KMsurv) # datasets from Klein and Moeschberger
library(parameters) # format model output tables for
library(webshot2) # convert interactive content to static for pdf
library(forcats) # functions for categorical variables ("factors")
library(stringr) # functions for dealing with strings
library(lubridate) # functions for dealing with dates and times
```

Here are some R settings I use in this document:

```r
rm(list = ls()) # delete any data that's already loaded into R

conflicts_prefer(dplyr::filter)
ggplot2::theme_set(
  ggplot2::theme_bw() +
        # ggplot2::labs(col = "") +
    ggplot2::theme(
      legend.position = "bottom",
      text = ggplot2::element_text(size = 12, family = "serif")))

knitr::opts_chunk$set(message = FALSE)
options('digits' = 6)

panderOptions("big.mark", ",")
pander::panderOptions("table.emphasize.rownames", FALSE)
pander::panderOptions("table.split.table", Inf)
conflicts_prefer(dplyr::filter) # use the `filter()` function from dplyr() by default
legend_text_size = 9
run_graphs = TRUE
```

Suppose $X_1, ..., X_n \ \sim_{\text{iid}} \ N(M, 1)$

Suppose $M \sim N(0, 1)$.

Then:

$$\begin{aligned} p(M = \mu | X = x) &\propto p(M = \mu, X = x) \\ &= p(X = x | M = \mu) p(M = \mu) \\ &\propto \exp\left\{-\frac{1}{2} n\mu^2 - 2\mu n\bar{x}\right\} \exp\left\{-\frac{1}{2}\mu^2\right\} \\ &= \exp\left\{-\frac{1}{2}(n+1)\mu^2 - 2\mu n\bar{x}\right\} \\ &\propto \exp\left\{-\frac{1}{2}(n+1)(\mu - \frac{n}{n+1}\bar{x})^2\right\} \end{aligned}$$

So:

$$p(M = \mu | X = x) \sim N(\frac{n}{n+1}\bar{x}, (n+1)^{-1})$$

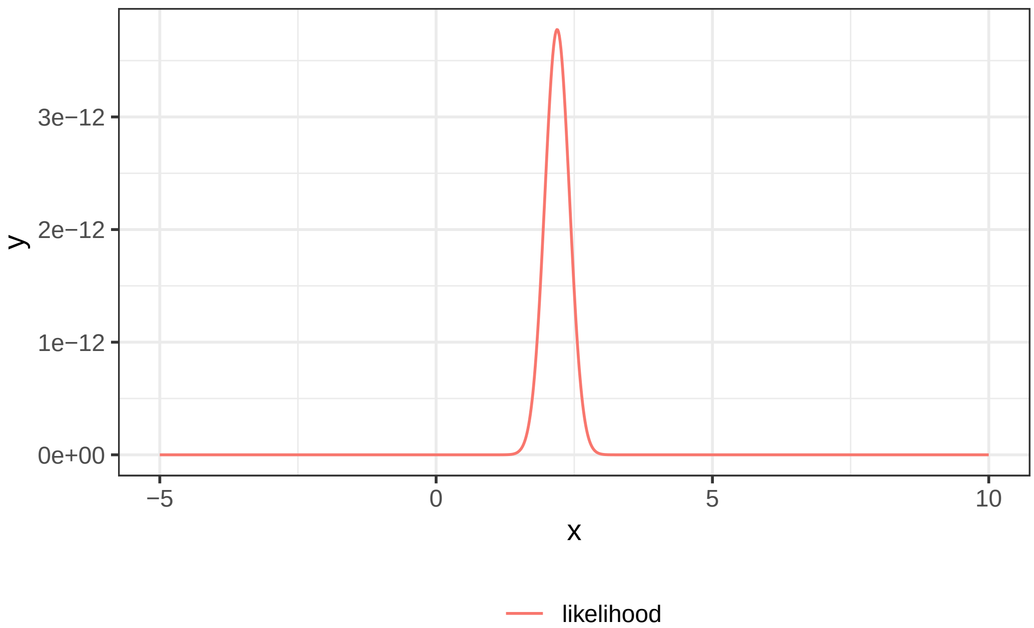
Let's put this in perspective.

Here's a frequentist CI:

```r
set.seed(1)
mu <- 2
sigma <- 1
n <- 20
x <- rnorm(n = n, mean = mu, sd = sigma)
xbar <- mean(x)
se <- sigma / sqrt(n)
CI_freq <- xbar + se * qnorm(c(.025, .975))
print(CI_freq)
#> [1] 1.75226 2.62879
```

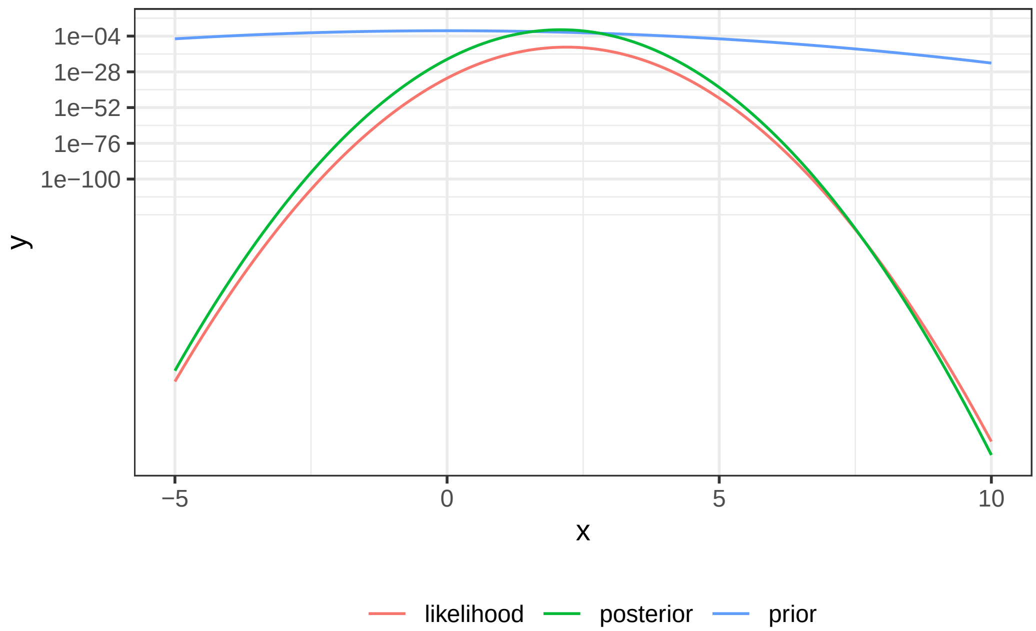
```r
lik0 <- function(mu) dnorm(x = x, mean = mu, sd = 1) |> prod()
lik <- function(mu) {
  (2 * pi * sigma^2)^(-n / 2) *
    exp(
      -1 / (2 * sigma^2) *
        (sum(x^2) - 2 * mu * sum(x) + n * (mu^2))
```

```r
    )
}
library(ggplot2)
ngraph <- 1001
plot1 <- ggplot() +
  geom_function(fun = lik, aes(col = "likelihood"), n = ngraph) +
  xlim(c(-5, 10)) +
  theme_bw() +
  labs(col = "") +
  theme(legend.position = "bottom")
print(plot1)
```



Here's a Bayesian CI:

```r
mu_prior_mean <- 0
mu_prior_sd <- 1
mu_post_mean <- n / (n + 1) * xbar
mu_post_var <- 1 / (n + 1)
mu_post_sd <- sqrt(mu_post_var)
CI_bayes <- qnorm(
  p = c(.025, .975),
  mean = mu_post_mean,
  sd = mu_post_sd
)
print(CI_bayes)
#> [1] 1.65851 2.51391
prior <- function(mu) dnorm(mu, mean = mu_prior_mean, sd = mu_prior_sd)
posterior <- function(mu) dnorm(mu, mean = mu_post_mean, sd = mu_post_sd)
plot2 <- plot1 +
  geom_function(fun = prior, aes(col = "prior"), n = ngraph) +
  geom_function(fun = posterior, aes(col = "posterior"), n = ngraph)
print(plot2 + scale_y_log10())
```



Here's $p(M \in (l(x), r(x)) | X = x)$:

```
pr_in_CI <- pnorm(
  CI_freq,
  mean = mu_post_mean,
  sd = mu_post_sd
) |> diff()
print(pr_in_CI)
#> [1] 0.930583
```

1.1 Example with JAGS

This example demonstrates Bayesian inference using JAGS (Just Another Gibbs Sampler) for a simple Bernoulli model from Dobson's text.

1.1.1 Load Required Packages

```
library(rjags)
library(runjags)
runjags::findJAGS()
#> [1] "/usr/bin/jags"
```

1.1.2 Run the JAGS Model

We'll use a simple Bernoulli model to estimate a probability parameter using Bayesian inference.

```
# JAGS chain initialization function
initsfunction <- function(chain) {
  stopifnot(chain %in% (1:4)) # max 4 chains allowed
  rng_seed <- (1:4)[chain]
  rng_name <- c(
    "base::Wichmann-Hill", "base::Marsaglia-Multicarry",
    "base::Super-Duper", "base::Mersenne-Twister"
  )[chain]
  return(list(".RNG.seed" = rng_seed, ".RNG.name" = rng_name))
}
```

```r
# Generate sample data
set.seed(1)
data1 <- rbinom(n = 91, size = 1, prob = .6)

# Run JAGS model
jags_post0 <- run.jags(
  n.chains = 2,
  inits = initsfunction,
  model = system.file("extdata/model.dobson.jags", package = "rme"),
  data = list(r = data1, N = length(data1)),
  monitor = "p"
)
#> Compiling rjags model...
#> Calling the simulation using the rjags method...
#> Note: the model did not require adaptation
#> Burning in the model for 4000 iterations...
#> Running the model for 10000 iterations...
#> Simulation complete
#> Calculating summary statistics...
#> Calculating the Gelman-Rubin statistic for 1 variables....
#> Finished running the simulation
```

1.1.3 Examine Results

```r
jags_post0$mcmc |> as.array() |> head()
#>       chain
#> iter       [,1]     [,2]
#>   5001 0.584687 0.550332
#>   5002 0.576621 0.562455
#>   5003 0.651983 0.628144
#>   5004 0.598258 0.604274
#>   5005 0.611863 0.583400
#>   5006 0.565635 0.606325
```

1.2 Other resources

UC Davis courses

- STA 015C[1]: "Introduction to Statistical Data Science III"
- STA 035C[2]: "Statistical Data Science III"
- STA 145[3]: "Bayesian Statistical Inference"
- ECL 234[4]: "Bayesian Models - A Statistical Primer"
- PLS 207[5]: "Applied Statistical Modeling for the Environmental Sciences"
- PSC 205H[6]: "Applied Bayesian Statistics for Social Scientists"
- POL 280[7]: "Bayesian Methods: for Social & Behavioral Sciences"
- BAX 442[8]: "Advanced Statistics"

Books

- Ross (2022) is a free online textbook
- "Population health thinking with Bayesian networks" (Aragon 2018) is on my to-read list

[1] https://catalog.ucdavis.edu/search/?q=STA+015C
[2] https://catalog.ucdavis.edu/search/?q=STA+035C
[3] https://catalog.ucdavis.edu/search/?q=STA+145
[4] https://catalog.ucdavis.edu/search/?q=ECL+234
[5] https://catalog.ucdavis.edu/search/?q=PLS+207
[6] https://catalog.ucdavis.edu/search/?q=PSC+205H
[7] https://catalog.ucdavis.edu/search/?q=POL+280
[8] https://catalog.ucdavis.edu/search/?q=BAX+442

- McElreath (2020)
  - very popular recently
  - author used to be a UCD professor
  - ECL 234[9] uses this book
  - videos: https://www.youtube.com/playlist?list=PLDcUM9US4XdPz-KxHM4XHt7uUVGWWVSus
  - course materials: https://github.com/rmcelreath/stat_rethinking_2024
- Korner-Nievergelt and Korner-Nievergelt (2015)
- Cowles (2013)
- Kéry, Schaub, and Beissinger (2012)
- Hobbs and Hooten (2015) has been used in PLS 207[10]

Aragon, Tomas J. 2018. "Population Health Thinking with Bayesian Networks." https://escholarship.org/uc/item/8000r5m5.

Cowles, Mary Kathryn. 2013. *Applied Bayesian Statistics: With R and OpenBUGS Examples*. Vol. 98. Springer Texts in Statistics. New York, NY: Springer Nature. https://doi.org/10.1007/978-1-4614-5696-4.

Hobbs, N. Thompson, and Mevin B Hooten. 2015. *Bayesian Models: A Statistical Primer for Ecologists*. STU - Student edition. Princeton: Princeton University Press.

Kéry, Marc., Michael. Schaub, and Steven R. Beissinger. 2012. *Bayesian Population Analysis Using WinBUGS : A Hierarchical Perspective*. 1st ed. Boston: Academic Press. https://shop.elsevier.com/books/bayesian-population-analysis-using-winbugs/kery/978-0-12-387020-9.

Korner-Nievergelt, Fränzi, and Fränzi Korner-Nievergelt. 2015. *Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and Stan*. 1st ed. Amsterdam, [Netherlands: Academic Press.

McElreath, Richard. 2020. *Statistical Rethinking : A Bayesian Course with Examples in R and Stan*. Second edition. Chapman & Hall/CRC Texts in Statistical Science Series. Boca Raton, FL: CRC Press.

Ross, Kevin. 2022. *An Introduction to Bayesian Reasoning and Methods*. Online. https://bookdown.org/kevin_davisross/bayesian-reasoning-and-methods/.

[9] https://catalog.ucdavis.edu/search/?q=ECL+234
[10] https://catalog.ucdavis.edu/search/?q=PLS+207