Chapter 1: Introduction to Bayesian Statistics

ST4234: Bayesian Statistics

Semester 2, AY 2019/2020

Department of Statistics and Applied Probability

National University of Singapore

LI Cheng

stalic@nus.edu.sg

Outline

Module Information

Schedule

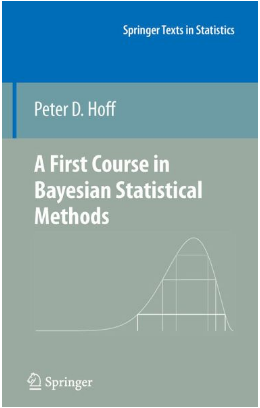
- Tuesday 6pm(+5min) to ~~10pm~~ 9pm
- 45min lecture + 5min break + 45min lecture + 10min break + 40min tutorial
- Tutorials start in the 3rd week
- Webcast available on LumiNUS
- Ask question during the break, after class, or through email
- Office consultation via appointment

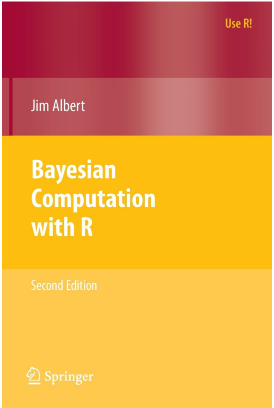
Website

- LumiNUS for everything (lecture notes, sample codes, tutorials, webcasts, announcements, surveys)

Module Information

Textbooks

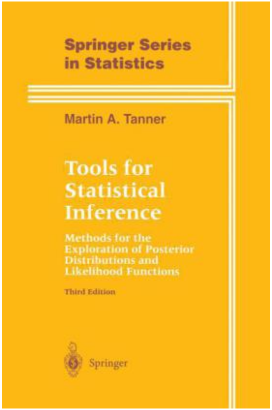


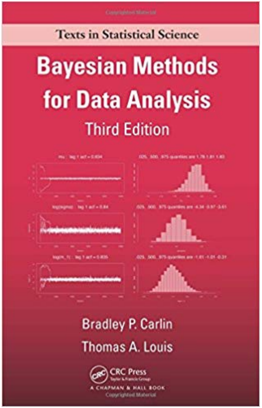


Ebooks from Springer can be downloaded from NUS library website.

Module Information

Reference books





Ebooks from Springer can be downloaded from NUS library website.

Module Information

Assessment

- **Tutorial Participation and Presentation (10%)**
  - About 10 tutorial sessions. Arrangement will be decided after the 2nd week, depending on the total enrollment.
  - Each group has 1 week to work on the tutorial problems. Submission through LumiNUS before the class on Tuesday.
  - Marked based on clarity and participation (not on correctness). **Group members are subject to mark reduction if they fail to contribute to the group discussion and preparation of solution files, or they are absent from the presentation.**
  - You are encouraged to work together!
  - You will get full marks for this part, as long as you show your effort and contribution to your teamwork.

Module Information

Assessment (cont'd)

- **1 Data Analysis Project (15%)**
  - Assigned after the midterm.
  - You will be asked to achieve some goals and answer some questions about some dataset(s).
  - Submit your **individual final report** electronically on LumiNUS before the reading week.
  - Requirements on format will be given later (restriction on the length of the report, typeset, etc.)
  - Handwritten or email submissions will not be accepted.
  - No late submission will be accepted.

Module Information

Assessment (cont'd)

- **Midterm Exam (20%)**
  - Tentatively scheduled on Tuesday, 3 March, 6:15pm - 7:45pm, venue TBA
- **Final Exam (55%)**
  - Saturday, April 25, 1pm - 3pm, venue TBA
  - **If you cannot make this date, please drop the module immediately.**
  - There will be no make-up exams.

Module Information

Prerequisites

- ST2132, or department approval.
- Familiar with calculus and basic concepts of mathematical statistics.
- Some prior knowledge in R programming is helpful.

Some Advice

You are strongly encouraged to

- Attend lectures and tutorials. Attempt tutorial problems.
- Ask questions when you do not understand.
- Read through the examples and try out the sample codes.

Module Information

Some Disclaimer

This is the first time I teach this module, so

- Please let me know if you spot any error in the lecture notes or the tutorials.
- I may update the pdf files of lecture notes on LumiNUS if a correction is necessary. All such corrections will be recorded chronically in a `.txt` file in each subfolder.
- I will refrain from spamming your emails with such corrections.

Acknowledgement

Many thanks to Linda Tan and David Nott for generously sharing their course materials.

Outline

Conditional probability

- Conditional probability: $\mathrm{P}(A|B) = \begin{cases} \dfrac{\mathrm{P}(A,B)}{P(B)} & \text{if } \mathrm{P}(B) > 0 \\ 0 & \text{otherwise.} \end{cases}$
- Multiplicative formula: $\mathrm{P}(A,B) = P(A|B)P(B) = \mathrm{P}(B|A)\mathrm{P}(A)$.
- **Bayes' Theorem:**

$$\begin{aligned} \mathrm{P}(A|B) &= \frac{\mathrm{P}(A,B)}{\mathrm{P}(B)} \\ &= \frac{\mathrm{P}(A,B)}{\mathrm{P}(B,A) + \mathrm{P}(B,A^c)} \\ &= \frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B|A)\mathrm{P}(A) + \mathrm{P}(B|A^c)\mathrm{P}(A^c)} \end{aligned}$$

Motivating example

- Women have a "1 in 8" (about 12%) risk of developing breast cancer at some time in their lives.
- If a patient has breast cancer, then the probability that a mammogram returns a positive result is 0.83.
- There is also a probability of 0.07 that a mammogram returns a positive result when the patient does not have breast cancer.
- What is the probability that a patient does not have breast cancer given that the mammogram returns a positive result?

$$\mathrm{P}(D^c|+) = \frac{\mathrm{P}(D^c, +)}{\mathrm{P}(+)} = \frac{\mathrm{P}(+|D^c)\mathrm{P}(D^c)}{\mathrm{P}(+|D)\mathrm{P}(D) + \mathrm{P}(+|D^c)\mathrm{P}(D^c)}$$
$$= \frac{0.07(1-0.12)}{0.83(0.12) + 0.07(1-0.12)} = 0.38.$$

Bayesian Inference

- Suppose we have some **unknown quantity** θ (possibly a vector) that we wish to learn about and we observe some **data** y relevant to their values.
- For example, θ could be parameters in a model or missing data.
- In Bayesian statistics, we need to specify
    - a **sampling model** or a probability distribution which describes how y depends on θ. This is expressed as a probability density function (pdf) $p(y|\theta)$, which we call the **likelihood function**.
    - a **prior distribution** $p(\theta)$ which expresses any prior knowledge or beliefs that we have about their values before observing the data.

Bayesian Inference

- Bayes' theorem (named after Thomas Bayes (1701–1761)) allows prior beliefs to be updated using observed data.

Bayes' Theorem

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

$$\Rightarrow p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

Bayesian Inference

- The prior and likelihood are used to compute the conditional distribution of θ given the data y using Bayes' Theorem. This conditional distribution $p(\theta|y)$ is called the **posterior distribution**.

  ### Bayes' Theorem

  $$p(\theta|y) = \frac{p(y,\theta)}{p(y)} = \frac{p(y,\theta)}{\int p(y,\theta')\mathrm{d}\theta'}$$

  $$= \frac{p(y|\theta)p(\theta)}{\int p(y|\theta')p(\theta')\,\mathrm{d}\theta'}.$$

- Here, we are assuming θ is continuous. If θ is discrete, the integral is just replaced by a sum over all possible values of θ.

Bayesian Inference

- As the denominator does not depend on θ (constant w.r.t. θ),

$$\underbrace{p(\theta|y)}_{\text{Posterior}} \propto \underbrace{p(y|\theta)}_{\text{Likelihood}} \underbrace{p(\theta)}_{\text{Prior}}.$$

- An easy way to remember Bayes' Theorem is

  "the posterior is proportional to the likelihood times the prior".

- Bayes' Theorem tells us how our prior beliefs about θ should be updated after observing new information.

- The posterior distribution summarizes the information in the data and prior and is used for inference.

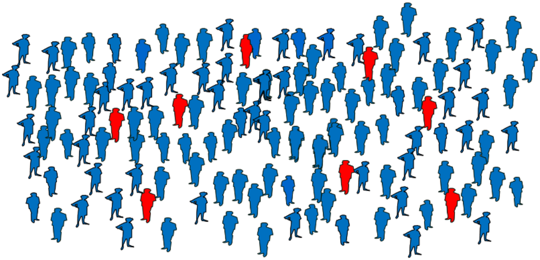
Bayesian Inference

- The denominator $p(y) = \int p(y|\theta')p(\theta')\,\mathrm{d}\theta'$ is called the **marginal distribution** of the data y.
- We can think of it as a **normalization constant** that ensures $\int p(\theta|y)\,\mathrm{d}\theta = 1$.
- In complex models, this integral is often intractable.
- Hence, finding the posterior distribution usually requires the use of advanced numerical methods, which we will learn later in this course.

Example: Estimating probability of a rare event

This example is from Section 1.2.1 of Peter Hoff's book.

- Suppose we are interested in estimating θ, the fraction of individuals infected with a certain disease in a city, $0 \leq \theta \leq 1$.
- A random sample of 20 individuals from the city are checked for infection. The data y records the total number of people in the sample who are infected; $y \in \{0, 1, 2, \ldots, 20\}$.



Example: Estimating probability of a rare event

Sampling model

- Let Y denote the number of infected individuals in the sample.
- A reasonable sampling model is $Y \sim \text{Binomial}(20, \theta)$.
- The likelihood is then

$$p(y|\theta) = \mathsf{P}(Y = y|\theta) = \binom{20}{y}\theta^y(1-\theta)^{20-y}.$$

- As the parameter θ is unknown, the conditioning of the likelihood on θ is expressed explicitly.

Example: Estimating probability of a rare event

Prior distribution

- Other studies from the country indicate that the infection rate in comparable cities ranges from 0.05 to 0.20, with an average of 0.10.
- This information suggests that we use a prior distribution $p(\theta)$ that assigns a substantial amount of probability to (0.05, 0.20) and has a mean close to 0.10.
- As there are many probability distributions that satisfy these conditions and our prior information is limited, we will use a prior that has these characteristics, but whose mathematical form is of computational convenience.

Example: Estimating probability of a rare event

Beta distribution as the prior

- We encode the prior information using a beta distribution defined on the interval $[0, 1]$.
- If $\theta \sim \text{Beta}(a, b)$ where $a > 0, b > 0$,

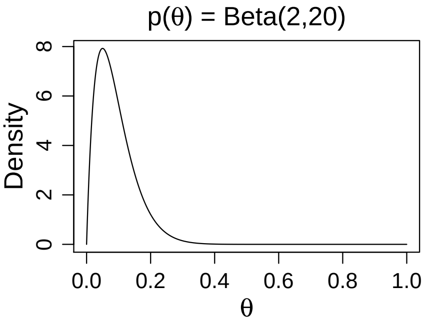
$$p(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{\mathrm{B}(a, b)}, \;\; 0 \le \theta \le 1,$$

  where $\mathrm{B}(a, b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is the beta function and $\Gamma(\cdot)$ denotes the gamma function.
- $\mathrm{E}(\theta) = \dfrac{a}{a+b}$, $\mathrm{Var}(\theta) = \dfrac{ab}{(a+b)^2(a+b+1)}$ and the mode of θ is $\dfrac{a-1}{a+b-2}$.

Example: Estimating probability of a rare event

Beta distribution as the prior

- Suppose we let $a = 2$, $b = 20$.
- This prior distribution is shown in the figure below.



- $E(\theta) = \dfrac{2}{2+20} = 0.09.$
- Mode $\theta = \dfrac{2-1}{2+20-2} = 0.05.$
- $P(\theta < 0.10) = 0.64.$
- $P(0.05 < \theta < 0.20) = 0.66$.

Example: Estimating probability of a rare event

Posterior distribution

- Suppose $y = 0$ is observed, i.e. no one in the sample is infected.
- Then the likelihood is $p(y|\theta) = \binom{20}{0}\theta^0(1-\theta)^{20-0} = (1-\theta)^{20}$.
- The posterior distribution can be computed as

$$\begin{aligned} p(\theta|y) &\propto p(y|\theta)p(\theta) \\ &\propto (1-\theta)^{20} \cdot \frac{\theta^{2-1}(1-\theta)^{20-1}}{\mathrm{B}(2,20)} \\ &\propto \theta^{2-1}(1-\theta)^{40-1}. \end{aligned}$$

- Thus $p(\theta|y) = \theta^{2-1}(1-\theta)^{40-1}/C$ where C is a normalizing constant. Since $\int p(\theta|y)\,\mathrm{d}\theta = 1$, $C = \mathrm{B}(2,40)$ and the posterior is a Beta(2, 40) distribution.

Example: Estimating probability of a rare event

Posterior distribution

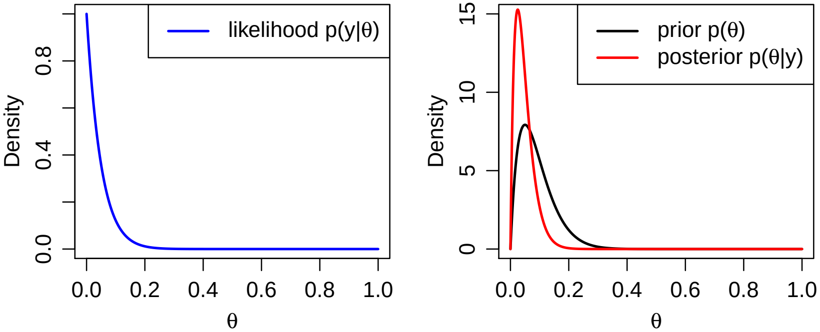
- Suppose $y = 0$ is observed, i.e. no one in the sample is infected.
- Then the likelihood is $p(y|\theta) = \binom{20}{0}\theta^0(1-\theta)^{20-0} = (1-\theta)^{20}$.
- The posterior distribution can be computed as

$$\begin{aligned} p(\theta|y) &\propto p(y|\theta)p(\theta) \\ &\propto (1-\theta)^{20} \cdot \frac{\theta^{2-1}(1-\theta)^{20-1}}{\mathrm{B}(2,20)} \\ &\propto \theta^{2-1}(1-\theta)^{40-1}. \end{aligned}$$

- Thus $p(\theta|y) = \theta^{2-1}(1-\theta)^{40-1}/C$ where C is a normalizing constant. Since $\int p(\theta|y)\,\mathrm{d}\theta = 1$, $C = \mathrm{B}(2,40)$ and the posterior is a Beta(2, 40) distribution.

Example: Estimating probability of a rare event

Compare prior and posterior



- The mass of the posterior density is concentrated further to the left than the prior distribution and is more peaked.
- The shift to the left of $p(\theta)$ is because the observation $y = 0$ provides evidence of a low value of θ.

Example: Estimating probability of a rare event

Compare prior and posterior

- The posterior is more peaked than the prior because it combines information from the data and the prior, and thus contains more information (and less uncertainty) than the prior.
- The peak of the posterior is at 0.025 and the posterior mean, $\mathrm{E}(\theta|y) = 0.048$. The posterior probability that $\theta < 0.10$, $\mathrm{P}(\theta < 0.10|y) = 0.93$.
- The posterior distribution $p(\theta|y)$ provides us with a model for learning about the city-wide infection rate θ.
- An individual whose prior beliefs about θ were represented by a Beta(2,20) distribution now has beliefs represented by a Beta(2,40) distribution.

Example: Estimating probability of a rare event

Sensitivity analysis

- Suppose we discuss the survey results with some health officials. A discussion among a diverse group of people might benefit from describing posterior beliefs corresponding to a variety of prior distributions.
- Generally, if we observe y infected individuals out of a sample of n subjects and consider a $\text{Beta}(a, b)$ prior,

$$\begin{aligned} p(\theta|y) &\propto p(y|\theta)p(\theta) \\ &\propto \theta^{y}(1-\theta)^{n-y} \times \theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta^{a+y-1}(1-\theta)^{b+n-y-1}. \end{aligned}$$

- The posterior distribution $p(\theta|y)$ is $\text{Beta}(a+y, b+n-y)$.

Example: Estimating probability of a rare event

Sensitivity analysis

- The posterior mean is

$$
\begin{aligned}
E(\theta|y) = \frac{a+y}{a+b+n} &= \frac{\cancel{a+b}}{a+b+n}\left(\frac{a}{\cancel{a+b}}\right) + \frac{\cancel{n}}{a+b+n}\left(\frac{y}{\cancel{n}}\right) \\
&= \frac{w}{w+n}\theta_0 + \frac{n}{w+n}\bar{y}
\end{aligned} \tag{1}
$$

  where $\theta_0 = \dfrac{a}{a+b}$ is the prior mean of θ and $w = a + b$.
- The posterior mean is a weighted average of the sample mean $\bar{y}$ and the prior mean θ_0.
- In terms of estimation, θ_0 represents our prior guess at the true value of θ and w represents our confidence in this guess, expressed on the same scale as the sample size.

Example: Estimating probability of a rare event

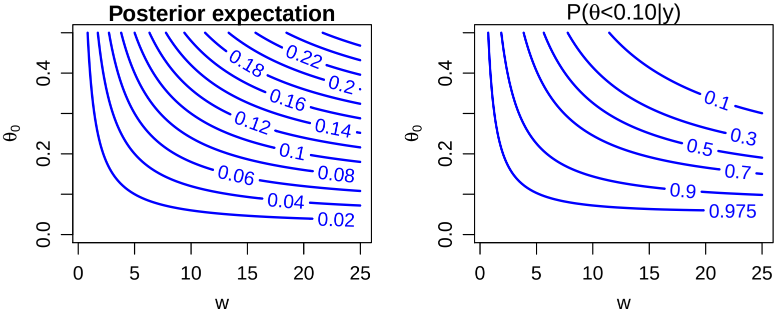
Sensitivity analysis

$$\begin{aligned} \theta_0 &= \frac{a}{a+b} \\ w &= a+b \end{aligned} \quad \Rightarrow \quad \begin{aligned} a &= \theta_0 w \\ b &= w(1-\theta_0) \end{aligned}$$

- Given a prior guess θ_0 with degree of confidence w, this prior belief can be approximated with a $\text{Beta}(w\theta_0, w(1-\theta_0))$ distribution.
- The posterior belief is then represented by a $\text{Beta}(w\theta_0 + y, w(1-\theta_0) + n - y)$ distribution.
- To explore how posterior information is affected by differences in prior opinion, let us compute the posterior expectation and $\text{P}(\theta < 0.10|y)$ for a range of θ_0 and w values.

Example: Estimating probability of a rare event

Sensitivity analysis



- Right plot is useful if city officials wish to recommend a vaccine to the public unless they are reasonably sure that the infection rate is less than 0.10.
- People with weak prior beliefs or low prior expectations are generally 90% or more certain that the infection rate is below 0.10.

Example: Estimating probability of a rare event

Comparison to non-Bayesian methods

- A standard estimate of a population proportion θ is the sample mean $\bar{y} = \frac{y}{n}$, the fraction of infected people in the sample.
- A 95% confidence interval (CI) for θ is

$$\bar{y} \pm 1.96\sqrt{\bar{y}(1-\bar{y})/n}.$$

  If n is large, this interval contains the true θ with probability 0.95.
- For our sample, $\bar{y} = 0$ and the CI is also just a single point 0. Note that our sample size $n = 20$ is small.
- Bayesian statistics provide a formal approach to combine information from different sources. When we do not have much information in the data (e.g. small sample), making use of external knowledge can be important.

Example: Estimating probability of a rare event

Comparison to non-Bayesian methods

- Let us consider the posterior mean $\mathrm{E}(\theta|y)$ as an estimate of θ. From (1),
$$\hat{\theta} = \frac{w}{w+n}\theta_0 + \frac{n}{w+n}\bar{y},$$
where θ_0 represents a prior guess at the true value of θ and w represents the degree of confidence in this guess.
- If the sample size n is large, $\bar{y}$ is a reliable estimate of θ. The estimator $\hat{\theta}$ takes advantage of this by placing more weight on $\bar{y}$ when n is large and less when n is small.
- Statistical properties of $\bar{y}$ and $\hat{\theta}$ are essentially the same for large n. However, for small n, the variability of $\bar{y}$ might be more than our uncertainty about θ_0 and $\hat{\theta}$ allows us to combine the data with prior information to stabilize our estimation of θ.

Example: Estimating probability of a rare event

Comparison to non-Bayesian methods

- $\hat{\theta}$ can be interpreted as a Bayesian estimator using a certain class of prior distributions.
- However, even if a particular prior distribution $p(\theta)$ does not exactly reflect our prior information, the corresponding posterior distribution $p(\theta|y)$ can still be a useful means of providing stable inference and estimation for situations where the sample size is low.

Outline

Module Information

Introduction to Bayesian Statistics

Why Bayes?

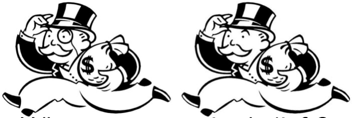
Introduction to R

Why Bayes?

- From the motivating example, we see that Bayesian methods provide
  - the ability to formally incorporate prior information into an analysis (this frees the analyst from ad hoc adjustments of results that "don't look right").
  - formal structure for statistical decision problems: it allows compromise estimates to emerge naturally (e.g. weighted average of sample mean and prior mean) and automatically gives increasing weight to the direct estimate as it becomes more reliable.
  - stable inference and estimation when the sample size is low.

Why Bayes?

- Bayesian methods require a sampling model $p(y|\theta)$ and a prior distribution $p(\theta)$ on all unknown quantities θ in the model.
- Cox (1946, 1961) and Savage (1954, 1972) prove that if $p(\theta)$ and $p(y|\theta)$ represent a person's beliefs, then Bayes' rule is an optimal method of updating these beliefs about θ given new information y.
- In practical data analysis, it can be hard to formulate what our prior beliefs are, and so $p(\theta)$ is often chosen for computational convenience.
- However, this does not mean that $p(\theta|y)$ is not useful. If $p(\theta)$ approximates our beliefs, then the fact that $p(\theta|y)$ is optimal under $p(\theta)$ means that it will also serve as a good approximation to what our posterior beliefs should be.



What are your prior beliefs?

Why Bayes?

- In other situations it may not be our beliefs that are of interest. Rather, we want to use Bayes' rule to explore how the data would update the beliefs of people with differing prior opinions.

- Of particular interest might be the posterior beliefs of someone with weak prior information. This has motivated the use of "diffuse" prior distributions, which assign probability more or less evenly over large regions of the parameter space.
- In many complicated statistical problems there are no obvious non-Bayesian methods of estimation or inference. In these situations, Bayes' rule can be used to generate estimation procedures, and the performance of these procedures can be evaluated using non-Bayesian criteria. In many cases it has been shown that Bayesian procedures work very well, even for non-Bayesian purposes.

Outline

What is R

- An open source programming language and software environment for statistical computing and graphics.
- Based on S language created by Chambers et al. at Bell Labs.
- Free Software. Developed by Ross Ihaka and Robert Gentleman in New Zealand.
- Reliable and comprehensive; v1.0 released February 2000.
- Current version is 3.6.2 (as of 8 January 2020).
- Now supported by a volunteer team of developers together with many contributors.

How to get it

To download and install R,

- Go to https://cran.r-project.org/;
- Select the link according to your system (Linux/Mac/Windows);
- Download and install according to the instructions.

RStudio is a free and open-source IDE of R. You can download it for free from https://www.rstudio.com/products/rstudio/download/.

RStudio is recommended.

R Tutorials

- *An introduction to R* is the official guide, by W. N. Venables, D. M. Smith, and the R Core Team.
- *Using R for Data Analysis and Graphics* by John Maindonald.
- Many free R tutorials are available online (find them in your search engine). They usually have well organized webpages that navigate you through the basics, e.g. `http://www.r-tutor.com/`
- The online help pages of R functions.

R Packages

- R packages are the main reason that R is popular for data analysis in the statistics community.
- To this day, the CRAN (comprehensive R archive network) has over 14,000 packages.
- The top downloaded packages are `Rcpp`, `ggplot2`, `digest`, etc. Most of the top downloaded packages are related to data structure and visualization.
- We will introduce some of the useful packages and functions as we proceed.

Next we illustrate the basic objects in R and how to perform simple data analysis in R.

R Markdown

- R Markdown is a very useful tool to streamline documentation and production related to R codes.
- You can access the online tutorial from the R markdown webpage.
- It allows the output format can be either docx or pdf files.

Next we illustrate the basic objects in R and how to perform simple data analysis in R.

Creating vectors, matrices and sequences

- Create a vector $a = [1, 2, 3, 4]$: `a <- c(1,2,3,4)`
- Create a matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$:
    - `A <- cbind(c(1,4,7), c(2,5,8), c(3,6,9))`
    - `A <- rbind(c(1,2,3), c(4,5,6), c(7,8,9))`
    - `A <- matrix(1:9, nrow=3, ncol=3, byrow=TRUE)`
- Create a sequence b from 0 to 1 with increment 0.1:
    - `b <- seq(from=0, to=1, by=0.1)`
- Create a sequence c from 0 to 1 of length 11:
    - `c <- seq(from=0, to=1, length.out=11)`

Basic operations and writing functions

- Operations such as `+, -, *, /, exp, log` are applied element-wise to vectors and matrices if the dimensions are correct. For example, `c(1,2,3)*c(4,5,6)` yields a vector $[4, 10, 18]$. To perform matrix multiplication, use `%*%`.

- To write a function for computing say the mean $\left(\dfrac{a}{a+b}\right)$ of a Beta(a, b) distribution, we can write:

```
beta_mean <- function(a,b) {a/(a+b)}
```

  "beta_mean" is the name of this function and its arguments are a and b. To apply this function to $a = 1$, $b = 2$, use `beta_mean(1,2)`. The order of the arguments is important unless we explicitly specify `beta_mean(b=2,a=1)`.

Reading datasets into R

- To read a dataset into R from a text document named "dataset" in our current working directory, use

```
df <- read.table(file="dataset.txt", header=TRUE)
```

  if the first line of the file contains the names of the variables. Otherwise set `header=FALSE`

- The dataset is then read into R as the object "df" which is a data frame (collection of variables). To refer to a variable named "var1" in `df`, we can type `df$var1`.

- To convert the dataframe into a matrix, we can type `as.matrix(df)`.

Plotting

- The following code produces the 1 by 4 figure in the following slide.
- We first create a sequence of values from 0 to 1 with increment 0.1 stored in the vector "theta".

```
theta <- seq(from=0, to=1, by=0.1)
```

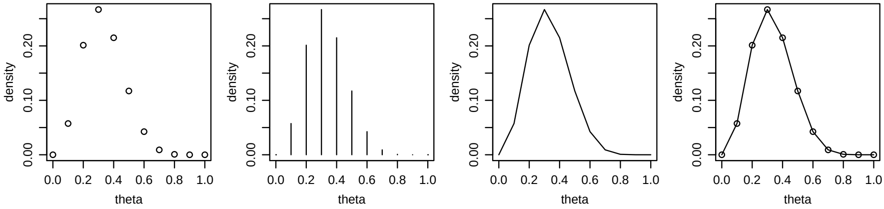
- Then we compute the Binomial$(10, \theta)$ density $\binom{10}{x}\theta^x(1-\theta)^{10-x}$ for $x = 3$ and $\theta \in \{0, 0.1, \ldots, 1\}$.

```
density <- dbinom(x=3, size=10, p=theta)
```

- Set up a 1 by 4 plot and do a scatterplot of "density" against "theta". The x-axis and y-axis labels can be specified using `xlab` and `ylab`.

```
par(mfrow=c(1,4))
plot(theta, density, xlab="theta", ylab="density")
```

Plotting



- The option `type="h"` produces histogram-like vertical lines, `type="l"` produces lines joining the points and `type="o"` plots the points and produces lines joining the points.

```
plot(theta,density,type="h",xlab="theta",ylab="density")
plot(theta,density,type="l",xlab="theta",ylab="density")
plot(theta,density,type="o",xlab="theta",ylab="density")
```

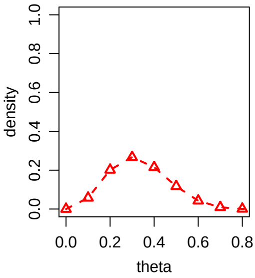
Plot parameters

- We can also specify other plot parameters. E.g.

```
plot(theta, density, type="o", xlab="theta", ylab="density",
  col="red", xlim=c(0,0.8), ylim=c(0,1), lwd=2, lty=2, pch=2)
```

Here,

- `col` specifies color of line and plot points,
- `xlim` and `ylim` give x and y coordinates ranges,
- `lwd` specifies line width,
- `lty` specifies line type (dotted, dashed etc),
- `pch` specifies type of plot symbols.



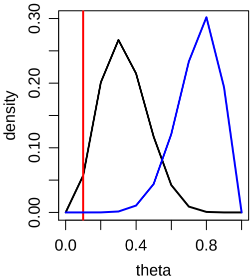
Use `help(par)` and `help(points)` for further details.

- To add a line to an existing plot, use `abline`. Specify `v`: *x*-value for a vertical line, `h`: *y*-value for a horizontal line or `a,b`: intercept and slope.
- To add points to an existing plot, use `points` (works like `plot`).
- Example:

```
plot(theta, density, type="l",
    xlab="theta", ylab="density",
    ylim=c(0,0.3), lwd=2)

abline(v=0.1, col="red", lwd=2)

points(theta,
    dbinom(x=8, size=10, p=theta),
    type="l", col="blue", lwd=2)
```



Contour plot

- In this example, we use `contour` to draw the contours of a bivariate normal density with zero mean and covariance matrix I_2.
- We require the package `mvtnorm` which provides the function `dmvnorm` for computing the density of a multivariate normal distribution.
- The arguments of `dmvnorm` are (1) the vector at which the density is to be computed, (2) the mean (zero vector by default) and (3) the covariance matrix (identity matrix by default) of the normal distribution.
- First we need to construct a grid of x and y values on which we compute the density. We construct sequences `x` and `y`, each containing 100 values distributed evenly between -2 and 2. Then we compute the density on this 100×100 grid, stored in the matrix `z`.

Contour plot

- The R-code below produces the contour plot.

```
require(mvtnorm)
N <- 100
x <- seq(from=-2, to=2, length.out=N)
y <- seq(from=-2, to=2, length.out=N)
z <- matrix(0,N,N)
for (i in 1:N){
  for (j in 1:N){
    z[i,j] <- dmvnorm(c(x[i],y[j]))
}}
contour(x, y, z, col="blue", lwd=2,
  labcex = 0.9, xlab = "x", ylab = "y")
```