

# Bayesian data analysis – reading instructions 3

Aki Vehtari

## Chapter 3

Outline of the chapter 3

- 3.1 Marginalisation
- 3.2 Normal distribution with a noninformative prior (very important)
- 3.3 Normal distribution with a conjugate prior (very important)
- 3.4 Multinomial model (can be skipped)
- 3.5 Multivariate normal with known variance (needed later)
- 3.6 Multivariate normal with unknown variance (glance through)
- 3.7 Bioassay example (very important, related to one of the exercises)
- 3.8 Summary (summary)

Normal model is used a lot as a building block of the models in the later chapters, so it is important to learn it now. Bioassay example is good example used to illustrate many important concepts and it is used in several exercises over the course.

R and Python demos at [https://avehtari.github.io/BDA\\_course\\_Aalto/demos.html](https://avehtari.github.io/BDA_course_Aalto/demos.html)

- demo3\_1: visualise joint density and marginal densities of posterior of normal distribution with unknown mean and variance
- demo3\_2: visualise factored sampling and corresponding marginal and conditional density
- demo3\_3: visualise marginal distribution of  $\mu$  as a mixture of normals
- demo3\_4: visualise sampling from the posterior predictive distribution
- demo3\_5: visualise Newcomb's data
- demo3\_6: visualise posterior in bioassay example

Find all the terms and symbols listed below. When reading the chapter, write down questions related to things unclear for you or things you think might be unclear for others. See also the additional comments below.

- marginal distribution/density
- conditional distribution/density
- joint distribution/density
- nuisance parameter
- mixture
- normal distribution with a noninformative prior
- normal distribution with a conjugate prior
- sample variance
- sufficient statistics
- $\mu, \sigma^2, \bar{y}, s^2$
- a simple normal integral

- Inv- $\chi^2$
- factored density
- $t_{n-1}$
- degrees of freedom
- posterior predictive distribution
- to draw
- N-Inv- $\chi^2$
- variance matrix  $\Sigma$
- nonconjugate model
- generalized linear model
- exchangeable
- binomial model
- logistic transformation
- density at a grid

### Conjugate prior for Gaussian distribution

BDA p. 67 (BDA3) mentions that the conjugate prior for Gaussian distribution has to have a product form  $p(\sigma^2)p(\mu|\sigma^2)$ . The book refers to (3.2) and the following discussion. As additional hint is useful to think the relation of terms  $(n-1)s^2$  and  $n(\bar{y}-\mu)^2$  in 3.2 to equations 3.3 and 3.4.

### Trace of square matrix

Trace of square matrix, trace,  $\text{tr } A$ ,  $\text{trace}(A)$ ,  $\text{tr}(A)$ , is the sum of diagonal elements. To derive equation 3.11 the following property has been used  $\text{tr}(ABC) = \text{tr}(CAB) = \text{tr}(BCA)$ .

### History and naming of distributions

See *Earliest Known Uses of Some of the Words of Mathematics* <http://jeff560.tripod.com/mathword.html>.

### Using Monte Carlo to obtain draws from the posterior of generated quantities

Chapter 3 discusses closed form posteriors for binomial and normal models given conjugate priors. These are also used as part of the assignment. The assignment also requires forming a posterior for derived quantities, and these posterior don't have closed form (so no need to try derive them). As we know how to sample from the posterior of binomial and normal models, we can use these posterior draws to get draws from the posterior of derived quantity.

For example, given posteriors  $p(\theta_1|y_1)$  and  $p(\theta_2|y_2)$  we want to find the posterior for the difference  $p(\theta_1 - \theta_2|y_1, y_2)$ .

1. Sample  $\theta_1^s$  from  $p(\theta_1|y_1)$  and  $\theta_2^s$  from  $p(\theta_2|y_2)$ , we can compute posterior draws for the derived quantity as  $\delta^s = \theta_1^s - \theta_2^s$  ( $s = 1, \dots, S$ ).
2.  $\delta^s$  are then draws from  $p(\delta^s|y_1, y_2)$ , and they can be used to illustrate the posterior  $p(\delta^s|y_1, y_2)$  with histogram, and compute posterior mean, sd, and quantiles.

This is one reason why Monte Carlo approaches are so commonly used.

## The number of required Monte Carlo draws

This will be discussed in chapter 10. Meanwhile, e.g., 1000 draws is sufficient.

## Bioassay

Bioassay example is an example of very common statistical inference task typical, for example, medicine, pharmacology, health care, cognitive science, genomics, industrial processes etc.

The example is from Racine et al (1986) (see ref in the end of the BDA3). Swiss company makes classification of chemicals to different toxicity categories defined by authorities (like EU). Toxicity classification is based on lethal dose 50% (LD50) which tells what amount of chemical kills 50% of the subjects. Smaller the LD50 more lethal the chemical is. The original paper mentions "1983 Swiss poison Regulation" which defines following categories for chemicals orally given to rats (mg/ml)

Class	LD50
1	<5
2	5-50
3	50-500
4	500-2000
5	2000-5000

To reduce the number of rats needed in the experiments, the company started to use Bayesian methods. The paper mentions that in those days use of just 20 rats to define the classification was very little. Book gives LD50 in log(g/ml). When the result from demo3\_6 is transformed to scale mg/ml, we see that the mean LD50 is about 900 and  $p(500 < LD50 < 2000) \approx 0.99$ . Thus, the tested chemical can be classified as category 4 toxic.

Note that the chemical testing is moving away from using rats and other animals to using, for example, human cells grown in chips, tissue models and human blood cells. The human-cell based approaches are also more accurate to predict the effect for humans.

logit transformation can be justified information theoretically when binomial likelihood is used.

Example codes in demo3\_6 can be helpful in exercises related to Bioassay example.

## Bayesian vs. frequentist statements in two group comparisons

When asking to compare groups, some students get confused as the frequentist testing is quite different. The frequentist testing is often focusing on a) differently named tests for different models and b) null hypothesis testing. In Bayesian inference a) the same Bayes rule and investigation of posterior is used for all models, b) null hypothesis testing is less common. We come later to decision making given posterior and utility/ cost function (Lecture 10.1) and more about null hypothesis testing (Lecture 12.1). Now it is assumed you will report the posterior (e.g. histogram), possible summaries, and report what you can infer from that. Specifically as in this assignment the group comparisons are based on continuous model parameter, the probability of 0 difference is 0 (later lecture 12.1 covers null hypothesis testing). Instead of forcing dichotomous answer (yes/no) about whether there is difference, report the whole posterior that tells also how big that difference might be. What big means depends on the application, which brings us back to the fact of importance of domain expertise. You are not experts on the application examples used in the assignment, but you can think how would you report what you have learned to a domain expert.

Frank Harrell's recommendations how to state results in two group comparisons are excellent <http://www.fharrell.com/2017/10/bayesian-vs-frequentist-statements.html>.