2 Week 2

2.1 Binomial model with "censored" observation

R-template `ex_binom_censored_observations_template.Rmd`.

(This exercise is modified from the exercise 2.1 in BDA3.)

Suppose you have a $Beta(1, 1)$ prior distribution on the probability θ that a coin will yield a head when spun in a specified manner. The coin is independently spun ten times, and 'heads' appear fewer than 3 times. You are not told how many heads were seen, only that the number is less than 3.

1) Write down the equation for the posterior density for θ.
2) Discretize interval $[0, 1]$ into 100 equally spaced intervals, calculate the posterior density at each of the discrete bins and draw the posterior density.
3) Using the discretized values of θ and their corresponding posterior density values a) draw the posterior cumulative density function of θ and b) calculate the posterior probability that $\theta < 0.3$.
4) Draw the posterior density for θ in case where you are told that the exact number of heads is 1. What is the apparent differences between these two posterior densities?

2.2 Probability of female birth given placenta previa

Read Sections 2.1 - 2.4 from BDA 3 and see `example_femaleBirthRate.Rmd` for additional help.

In this exercise you need the following special result. Assume we have a Binomial observation model

$$p(y|\theta, n) = \mathrm{Bin}(y|n, \theta) \propto \theta^y(1-\theta)^{n-y}. \tag{1}$$

The number of trials, n, is considered as fixed and the parameter θ is given a Beta prior

$$p(\theta) = \mathrm{Beta}(\theta|\alpha, \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}. \tag{2}$$

Comparing the observation model and prior gives insight that the Beta prior corresponds to $\alpha - 1$ prior successes and $\beta - 1$ prior failures. A special case is $\alpha = \beta = 1$ when the prior is uniform on the interval $[0, 1]$. The posterior distribution of θ is now also Beta distribution (see BDA 3, Chapter 2 for more details)

$$\theta|N, y \sim \mathrm{Beta}(\alpha + y, \beta + n - y). \tag{3}$$

It is suggested that a maternal condition called *placenta previa* may be more common in case of female baby than in case of male baby. An early study concerning the sex ratio in placenta previa births in Germany found that out of a total of $n = 980$ births $y = 437$ were female.

a) Given the above data, analyze the ratio of females in the *superpopulation* of all placenta previa births with the following steps:

1. Visualize the posterior density for the proportion of female births θ given n, y
2. Visualize cumulative density function
3. Calculate posterior median and central 90% posterior interval for θ.
4. Calculate the posterior probability, $p(\theta < 0.485|y, n)$ and $p(\theta > 0.485|y, n)$
5. Do all the above steps with i) uniform prior for θ and ii) three informative prior distributions for θ where the prior mean is always 0.5 (see Appendix A of BDA3, available at course we page)

b) Continue the analyses by using the uniform prior for θ and Sample 10 000 random draws from the posterior distribution and use them to

1. visualize the distribution with histogram
2. calculate the posterior median, standard deviation and coefficient of variation of θ
3. visualize the female-to-male sex ratio $\theta/(1 - \theta|y, n)$ with histogram
4. calculate $p(\theta/(1 - \theta) < 0.9|y, n)$

Note! Check carefully that you have the quantiles, parameter limits etc. correct. They are not necessarily the same as in the `example_femaleBirthRate.Rmd`.

2.3 Populations and population parameters

A very typical situation for statistician or data scientist is the following. Someone has collected data on a phenomenon and wants you to analyze it to answer his/her question. In this type of situation you have not had control on what type of data has been collected. Hence, it is important to realize whether the data can be used to answer the question at hand and, if yes, how.

Consider the following five questions and reply to a) what would be the superpopulation in that example, and b) could all or part of the data be considered as a random sample from that superpopulation *in practical terms*. Give a short justification.

1) Question: what is the average height of 5 year old Finns? Data: Heights and ages of all day care children at randomly chosen day cares throughout the Finland.

2) Question: what is the average height of 5 year old Finns? Data: Heights of all day care children at randomly chosen day cares throughout the Finland.

3) Question: In what proportion of sandy shores of Gulf of Bothnia can white fish spawn? Data: Data on type of shore and whether white fish has spawned there from randomly chosen locations in Gulf of Bothnia shore line.

4) Question: How big proportion of people in Puumala would be willing to use an electronic scooter? Data: User statistics of electronic scooters from Helsinki and population demography information from Helsinki and Puumala.

5) Question: How big proportion of people in Tampere would be willing to use an electronic scooter? Data: User statistics of electronic scooters from Helsinki and population demography information from Helsinki and Tampere.