Report submissions are accepted in the following formats, one text document for the solutions, **name_lab1**. *Also submit an email with your R-files, with a file named* **name_lab1.R** *that can be used to run your analysis, remember to submit* **all** *of the files you have created, and also that the codes are possible to run. Email the files to* **jonas.wallin@stat.lu.se**.

Discussion between groups is permitted (and encouraged), as long as your answers and code reflects your own work.

Deadline: Friday 10/11 at 23.59

Bullets, indicates mandatory exercises, whereas star indicates voluntarily.

Prior

If you unaccustomed to R, on Rstudio. Before Starting, go through the tutorials on the homepage.

Basic R

In this section you are supposed to learn how to use basic R function in order to learn to calculate statistical properties of distribution.

- Assume that the random variable X has normal distribution with mean, μ, 2 and variance, σ^2, 3. Use R to calculate:

  1. $\mathbb{P}(X > 2)$
  2. $\mathbb{P}(1 < X \leq 3)$
  3. The value x_p such that $\mathbb{P}(X > x_p) = 0.82$.

  (1p)

- Assume that the random variable, X, has a Beta distribution,

  $$B(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},$$

  where $\alpha > 0$,$\beta > 0$ and $x \in [0,1]$. For $\alpha = 0.6$ and $\beta = 0.4$, use R to:

  1. Calculate $\mathbb{P}(X < 0.5)$.
  2. Find the value x_p such that $\mathbb{P}(X < x_p) = 0.8$.
  3. Plot the density function over $[0,1]$.
  4. Generate 100 random variables and visualize the result in a histogram.

  (1p)

* Write an R function that takes a vector and returns the sum of the values that are even. (2p)
  Hint: To check if the value **val** is even you can use **val %% 2==0**

Sampling statistics

In this section, the goal is to understand that sampling can be used instead of analytical integration ($\mathbb{P}(X \leq x)$ is an integral of the density of X.). This is fundamental in Bayesian inference, since almost no interesting Bayesian model can be done solely on analytical integration.

- Create a piece of code that, approximately, computes the probability:

  $$\mathbb{P}(|X|^3 - 3 > 10),$$

  where $X \sim N(0, 2)$. (1p)

* The command rgamma(n, scale = 2, shape=3) generates Gamma random variables, Γ. Use this compute, approximately, the Probability that $\mathbb{P}(\Gamma > 2)$, $\mathbb{E}[\Gamma]$ and $\mathbb{V}[\Gamma]$. (1p)

Bayesian statistics

In this section, the goal is get a basic understanding of Bayesian statistics.

* Solve 2H1 in the book. (0.5p)
* Solve 2H2 in the book. (0.5p)
* Solve 2H3 in the book. (0.5p)
* Solve 2H4 in the book. (0.5p)

- In this example we study the marble bag and try to discover how many blue marbles it is in the bag, we know that it is five marbles in the bag. In the file marble.txt is a vector of ten observations of marbles, where 0 is white marble and 1 is blue marble. We assume a uniform prior (all possible marble bags have the same probability.)
  - Visualize the posterior distribution of number of blue marbles in the bag. (1p)
  - What is the posterior mean, and posterior variance of the number blue marbles? (1p)

- In a German study of 980 births where the mother had placenta previa, 437 of the babies were female. This imposes a likelihood of the data

  $$n_f \sim Bin(n, p)$$

  here n_f is number of girls born, n is total number of babies and p the probability of a baby being a girl. We will use two prior:
  - $\pi_1(p)$ is a prior which gives uniform density to all values of p above 0.5, and zero to all p below 0.5.
  - $\pi_2(p)$ gives uniform values to all p (between zero and one).

1. Use the grid approximation approach in the book to generate posterior distribution. Comment on the posterior from both π_1 and π_2 do they look reasonable? (2p)
2. Use the sampling method to compute the posterior mean of p for both priors. (1p)
3. Use the sampling method to compute the posterior probability that the probability of having a girl is greater then having a boy. (2p)