Bayes assignment 1

Sean van der Merwe

2022/03/25

# Currently marking student 2012345678

# Instructions

The goal of this assignment is to gain insight for yourself into the impact of prior distributions on modelling. Specifically, we will use the $Normal\left(μ,σ\right)$ distribution that everyone is familiar with as an example, since we have ‘correct’ answers already derived for this problem. We will denote the standard deviation as $σ$ and work directly with that notation, which is not the most common notation but becoming more popular.

## Part A

1. Begin by simulating 5 samples from the standard Gaussian distribution. The samples must have size 20, 30, 40, 50, and 60.
2. Create a single summary table (6 columns) showing descriptive statistics of your 5 samples, including at least: n, mean, sd, median, quartiles, min, max. [20]

## Part B

1. Find, give, and reference at least one commonly used objective prior for $μ$ and $σ$. [10]
2. Why and how is the Empirical Bayes framework different to the more classical framework? [10]
3. Give, in formula/equation form, the sampling distribution of the mean and standard deviation of each sample. Then derive the implied distributions of $μ$ and $σ$ (given $n$). [5+5=10]

Lastly, for the questions below, also consider the following flawed priors: $μ∼N\left(1,1\right)$ and $σ∼χ\_{n−1}^{2}$.

## Part C

1. For every combination of sample and prior, simulate 10000 samples from the joint posterior of $μ$ and $σ$.
2. Summarise the simulations in the form of parameter estimates, via both the posterior mean and posterior median approaches. [2priors x 5samples x 2estimators = 20 marks]
3. For fun, discuss any superficial patterns that you see in your parameter estimates, especially in relation to the true values that you know. [5]

[25 marks for generally following directions and neat presentation. All code must be shown and explained, even though mark allocation is only at items with distinct outputs. Negative marking will be implemented for hard coding and poor presentation.]

# Memorandum

## Part A

seq(20, 60, 10) -> n
n |> lapply(rnorm) -> smples

smples |> sapply(\(x) { c(
 n = length(x),
 Average = mean(x),
 StdDev = sd(x),
 Median = median(x),
 LQ = quantile(x, 0.25),
 UQ = quantile(x, 0.75),
 Min = min(x),
 Max = max(x)
)}) |> kable(digits = 3)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| n | 20.000 | 30.000 | 40.000 | 50.000 | 60.000 |
| Average | 0.023 | -0.232 | -0.158 | 0.080 | 0.147 |
| StdDev | 0.804 | 0.866 | 1.148 | 0.787 | 1.026 |
| Median | -0.151 | -0.114 | -0.139 | -0.033 | 0.143 |
| LQ.25% | -0.487 | -0.776 | -0.759 | -0.454 | -0.535 |
| UQ.75% | 0.387 | 0.422 | 0.463 | 0.549 | 0.922 |
| Min | -1.399 | -1.989 | -2.565 | -1.452 | -2.132 |
| Max | 1.661 | 1.564 | 2.158 | 1.678 | 2.092 |