Assignment 2

Exercise 1

1. Simulate $n = 100$ draws from a Normal, $\mathrm{N}(5, 9)$, distribution (using `rnorm`). Estimate the mean with the first $1, \dots, n$ (very n) draws. Discuss and visualise convergence of the estimates.
2. Simulate $n = 10000$ draws from a Cauchy distribution with scale one by drawing from $\frac{\mathrm{N}(0,1)}{\mathrm{N}(0,1)}$. Estimate the mean with the first $1, \dots, n$ draws. Discuss and visualise convergence of the estimates.

Exercise 2

You have observations on daily *Alles Gurgelt* tests in your office — $\mathbf{y} = (y_1, \dots, y_n)$ — and want to learn about the prevalence. There are 20 colleagues who test everyday. Assume that the data is independent and identically distributed.[1]

1. What is the class of conjugate priors for this problem? Derive the posterior distribution $p(\theta \mid \mathbf{y})$.
2. Assume you have observations for thirty days ($n = 30$) with a total of ten positive test ($\sum_i^n y_i = 10$). Determine and briefly explain several point estimators of θ.
3. Discuss sources of prior information for this problem and compare the impact of different priors on your point estimates.
4. Discuss the assumption of independent and identically distributed data. How could you (conceptually) improve the model with this in mind?

Exercise 3

Write an **R** function to simulate n observations from the model $\mathbf{y} = \alpha + \mathbf{X}\beta + \mathbf{e}$.[2] Draw the k independent variables from distributions of your choice, and the error from a Normal with mean zero and standard deviation σ. The function should have arguments to set n, k, α, β, and σ; it should return a list with the simulated data, $\mathbf{y}$ and $\mathbf{X}$.

1. Simulate data with $k = 1$ and $\sigma = 1$. Plot the regressor $\mathbf{x}$ and regressand $\mathbf{y}$ in a scatterplot; add a LS regression line. Repeat this 1,000 times and store β_{LS} every time. Then create a histogram of the LS estimates — what do you see?
2. Assume you know that $\sigma^2 = 1$. What are the latent values of the model?
3. Come up with a potentially interesting regression you want to run. Explain and draw ways you expect a single coefficient of interest, β_j, to look like a priori.
4. Simulate data with $k = 1$ and $\sigma = 1$ — you can assume you know σ. Set a Normal prior, $\mathrm{N}(\mu_0, \sigma_0)$, for β — decide on parameters μ_0 and σ_0 for this pior. Compute and plot the posterior density for simulated data with increasing n (e.g. $n \in \{50, 100, 200\}$).[3]

Exercise 4

1. Suppose you have data $\mathbf{y} \sim \mathrm{N}(\mu, 1)$, and want to estimate μ. Specify a Normal prior $\mu \sim \mathrm{N}(\mu_0, \sigma_0^2)$. Derive the posterior $p(\mu \mid \mathbf{y})$ by applying Bayes' theorem. Create histograms of two priors of your choice.
2. Suppose you have data $\mathbf{y} \sim \mathrm{N}(5, \sigma^2)$, and want to estimate σ^2. Work with the precision, σ^{-2}, and specify a Gamma prior $\sigma^{-2} \sim \mathrm{G}(0.5, \eta)$ with single parameter η. Derive the posterior $p(\sigma^2 \mid \mathbf{y})$ by applying Bayes' theorem. Visualise the prior density for $\eta \in \{0.01, 1, 100\}$.
3. Suppose you are uncomfortable with choosing a value for η, and want to include this parameter in your model. Discuss a suitable prior distribution for η, and visualise the prior $\sigma^2|\eta$ by first simulating draws from η, and then σ^2 repeatedly.

[1] The *Binomial distribution*, $Y \sim \mathrm{Binom}(20, \theta)$ with unknown parameter θ is relevant to this.
[2] We generally group α and β together and just use $\bar{\mathbf{X}} = [\mathbf{1}\,\mathbf{X}]$ for derivations.
[3] The posterior density is given by $\mathrm{N}(\mu_n, \sigma_n^2)$, where $\sigma_n = (\sigma_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}$ and $\mu_n = \sigma_n \left[\sigma_0^{-1}\mu_0 + \mathbf{X}'\mathbf{y}\right]$.