Submit your solutions in a single pdf via email to `erik.strumbelj@fri.uni-lj.si`. Deadline is anytime before taking the exam.

The theoretical parts can be typeset (LaTeX recommended) or handwritten, as long as they are clear enough to read. Show **all** your work! Answers without supporting work will not be given credit. All computational results should be reproducible and code should be in a repository. The repository may be private, but you will have to share the content for the exam. Use Python and/or R.

(A) Let $X \sim N(0,1)$ and $f(x) = 10\exp(-5(x-3)^4)$. We are interested in computing the expectation of f with respect to X.

1. Approximate the integral using a quadrature based method (you may use a third-party library).
2. Approximate the integral using Monte Carlo (100 samples).
3. Repeat (2) 1000 times, compute a 90% CI using the normal approximation, and observe whether or not the CI contains the 'true' value (use (1)). Does it contain the true value 90% of the time? Discuss.
4. Approximate the integral using importance sampling (100 samples) with a sensible choice of surrogate distribution. Compare the variance of your importance sampling estimator with the variance of the Monte Carlo estimator from (2) and discuss.
5. Approximate the integral using Monte Carlo (100 samples) but instead of using the build-in generator for the standard normal distribution, implement rejection sampling using a logistic distribution-based envelope. Compare the variance of this estimator with the variance of estimator from (2) and discuss.
6. Approximate the integral using Monte Carlo (100 samples) but instead of using the build-in generator for the standard normal distribution, implement Metropolis-Hastings with a $U(x_i - \delta, x_i + \delta), \delta > 0$ proposal distribution. How do the variance of the estimator and the rejection rate change with δ? Find the optimal δ and compare the optimal variance with (2). What is the effective sample size (ESS)? You may use a library for MCMC variance estimation.

(B) For each of the Markov chains defined below do the following:

- find all communicating classes and show whether or not the chain is irreducible,
- for each state determine its period and whether it is transient, positive recurrent or null recurrent, and
- find all stationary distributions and show that there are no other stationary distributions.

1.

$$K = \begin{bmatrix} 0 & \frac{4}{5} & 0 & \frac{1}{5} & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{10} & \frac{2}{5} \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}.$$

2.

$$K = \begin{bmatrix} \frac{1}{10} & \frac{2}{10} & \frac{3}{10} & \frac{4}{10} \\ 0 & \frac{5}{10} & \frac{2}{10} & \frac{3}{10} \\ 0 & 0 & \frac{3}{10} & \frac{7}{10} \\ 0 & 0 & \frac{1}{10} & \frac{9}{10} \end{bmatrix}.$$

3.
$$K = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}.$$

4. $S = \mathbb{Z}$ (integers). If we are in state z, we have $\frac{1}{2}$ probability to move to $z-1$ and $\frac{1}{2}$ probability to move to $z+1$

5. $S = \mathbb{Z}$. If we are in state $z > 0$, we have $\frac{4}{5}$ probability to move to $z-1$ and $\frac{1}{5}$ probability to move to $z+1$. If we are in state $z < 0$, we have $\frac{4}{5}$ probability to move to $z+1$ and $\frac{1}{5}$ probability to move to $z-1$. If we are in $z = 0$, we have equal probability to move to $+1$ or -1.

(C) Plot a contour plot of this function (it should look like a banana):

```
fn <- function(x) {
  exp(-(x[1]^2)/200- 0.5 * (x[2]+ 0.05 * x[1]^2 - 100*0.05)^2 )
}
```

The goal is to sample from the distribution whose density is proportional to the above function.

Implement a Metropolis-Hastings sampler and for each of the proposal distributions below do the following: draw 1000 samples from 3 different starting points (= 3 different chains), for each chain plot the path of first 100 steps (in the contour plot; jitter rejected proposals so that they are visible), plot the traceplot and compute the ESS (for each of the two variables; you may use a library for MCMC variance estimation) and compute the rejection rate.

1. Sample uniformly from a square with side of length 1 and centered on the current state.
2. Sample uniformly from a square with side of length 20 and centered on the current state.
3. Suggest a proposal distribution that will be more efficient. Why is it difficult to come up with an efficient proposal distribution for this density?

(D) The goal of this problem is to implement a Metropolis-Hastings sampler for Bayesian logistic regression.

1. Generate a toy dataset with 100 observations, each with 5 real-valued predictors $x_1, ..., x_5$ and a binary dependent variable Y. All predictors are standard normal and $Y_i \sim Bernoulli(\frac{1}{1+e^{-\mu_i}})$, where $\mu_i = 3x_1 - 2x_2 + \frac{1}{10}x_3 + 1$. That is, we have a classification problem where only the first three predictors are relevant.
2. Implement the logistic regression likelihood with an arbitrary number of coefficients (same as above, but $\mu_i = \beta^T x_i$, where x_i is now a vector) and combine it with a weakly-informative standard normal prior $\beta_i \sim N(0, 100)$ to get a function that is proportional to the posterior distribution of the logistic regression model.
3. Use your own implementation of Metropolis-Hastings with a standard normal proposal to infer the parameters on the toy dataset. Generate 4 independent chains from the posterior with 10000 samples each. Inspect traceplots, rejection rate and lag-k covariances and compute effective sample size (ESS) for each parameter. You may use a library for MCMC variance estimation. Discuss if there is some reason for concern that the Markov chain is problematic.
4. Do your best to improve the efficiency of the sampler by changing the covariance matrix of the proposal distribution. Report the rejection rate and ESS for the most efficient M-H sampler you get. Compare with those obtained in (3) and discuss.
5. Estimate posterior means of β_i using the most efficient sampler. Compare with the ground truth. Estimate the posterior probability $P(|\beta_3| > \frac{1}{10}|\text{data})$.

The End.