The purpose of this tutorial sheet is to help you better understand the lecture material. Start early and do as many as you have time for. Even if you are unable to make much progress, you should still attend your tutorial.

Exercise 1. ***Inverse transform sampling***

The cumulative distribution function (cdf) $F_x(\alpha)$ of a (continuous or discrete) random variable x indicates the probability that x takes on values smaller or equal to α,

$$F_x(\alpha) = \mathbb{P}(x \leq \alpha). \tag{1}$$

For continuous random variables, the cdf is defined via the integral

$$F_x(\alpha) = \int_{-\infty}^{\alpha} p_x(u)\mathrm{d}u, \tag{2}$$

where p_x denotes the pdf of the random variable x (u is here a dummy variable). Note that F_x maps the domain of x to the interval $[0, 1]$. For simplicity, we here assume that F_x is invertible.

For a continuous random variable x with cdf F_x show that the random variable $y = F_x(x)$ is uniformly distributed on $[0, 1]$.
Hint: Determine the cdf of y.

Importantly, this implies that for a random variable y which is uniformly distributed on $[0, 1]$, the transformed random variable $F_x^{-1}(y)$ has cdf F_x. This gives rise to a method called "inverse transform sampling" to generate n iid samples of a random variable x with cdf F_x. Given a target cdf F_x, the method consists of:

- calculating the inverse F_x^{-1}
- sampling n iid random variables uniformly distributed on $[0, 1]$: $y^i \sim \mathcal{U}(0, 1)$, $i = 1, \ldots, n$.
- transforming each sample by F_x^{-1}: $x^i = F_x^{-1}(y^i)$, $i = 1, \ldots, n$.

By construction of the method, the x^i are n iid samples of x.

Exercise 2. ***Sampling from a Laplace random variable***

A Laplace random variable x of mean zero and variance one has the density $p(x)$

$$p(x) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right) \qquad x \in \mathbb{R}. \tag{3}$$

Use inverse transform sampling to generate n iid samples from x.

Exercise 3. ***Sampling from a restricted Boltzmann machine***

The restricted Boltzmann machine (RBM) is a model for binary variables $\mathbf{v} = (v_1, \ldots, v_n)^\top$ and $\mathbf{h} = (h_1, \ldots, h_m)^\top$ which asserts that the joint distribution of $(\mathbf{v}, \mathbf{h})$ can be described by the probability mass function

$$p(\mathbf{v}, \mathbf{h}) \propto \exp\left(\mathbf{v}^\top \mathbf{W}\mathbf{h} + \mathbf{a}^\top \mathbf{v} + \mathbf{b}^\top \mathbf{h}\right), \tag{4}$$

where $\mathbf{W}$ is a $n \times m$ matrix, and $\mathbf{a}$ and $\mathbf{b}$ vectors of size n and m, respectively. Both the v_i and h_i take values in $\{0, 1\}$. The v_i are called the "visibles" variables since they are assumed to be observed while the h_i are the hidden variables since it is assumed that we cannot measure them (see the additional practice material from tutorial 2).

Explain how to use Gibbs sampling to generate samples from the marginal $p(\mathbf{v})$,

$$p(\mathbf{v}) = \frac{\sum_{\mathbf{h}} \exp\left(\mathbf{v}^\top \mathbf{W}\mathbf{h} + \mathbf{a}^\top \mathbf{v} + \mathbf{b}^\top \mathbf{h}\right)}{\sum_{\mathbf{h},\mathbf{v}} \exp\left(\mathbf{v}^\top \mathbf{W}\mathbf{h} + \mathbf{a}^\top \mathbf{v} + \mathbf{b}^\top \mathbf{h}\right)}, \tag{5}$$

for any given values of $\mathbf{W}$, $\mathbf{a}$, and $\mathbf{b}$.

Hint: Use the results in the additional practice sheet of Tutorial 2.