Exercise 1. ***Monte Carlo integration and importance sampling***

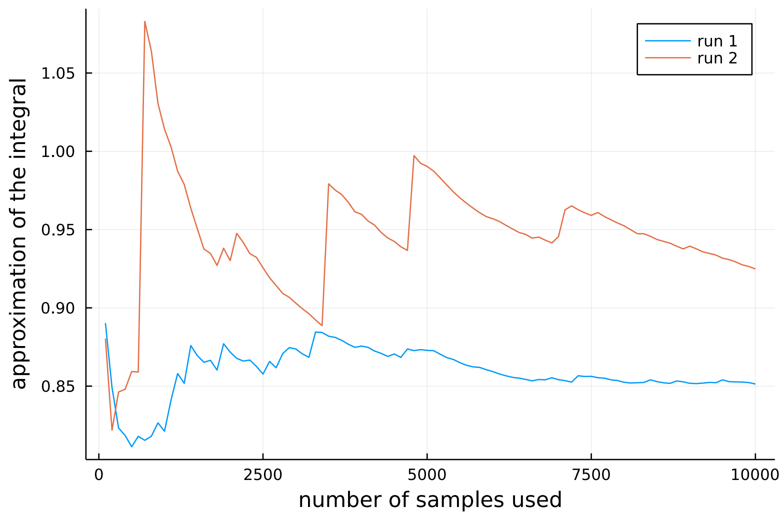
A standard Cauchy distribution has the density function (pdf)

$$p(x) = \frac{1}{\pi}\frac{1}{1+x^2} \tag{1}$$

with $x \in \mathbb{R}$. A friend would like to verify that $\int p(x)dx = 1$ but doesn't quite know how to solve the integral analytically. They thus use importance sampling and approximate the integral as

$$\int p(x)dx \approx \frac{1}{n}\sum_{i=1}^{n}\frac{p(x_i)}{q(x_i)} \qquad x_i \sim q \tag{2}$$

where q is the density of the auxiliary/importance distribution. Your friend chooses a standard normal density for q and produces the following figure:



The figure shows two independent runs. In each run, your friend computes the approximation with different sample sizes by subsequently including more and more x_i in the approximation, so that, for example, the approximation with $n = 2000$ shares the first 1000 samples with the approximation that uses $n = 1000$.

Your friend is puzzled that the two runs give rather different results (which are not equal to one), and also that within each run, the estimate very much depends on the sample size. Explain these findings, both mathematically and intuitively.

Exercise 2. ***Inverse transform sampling for*** $\mathcal{B}(x; 2, 1)$

We here use inverse transform sampling to sample from $\mathcal{B}(x; 2, 1)$.

(a) What is the density and cumulative distribution function (cdf) of $\mathcal{B}(x; 2, 1)$?

(b) Derive an explicit formula to generate samples $x \sim \mathcal{B}(x; 2, 1)$ from samples $u \sim \mathcal{U}(0, 1)$.

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

$$p(\mathbf{v}, \mathbf{h}) \propto \exp\left(\mathbf{v}^\top \mathbf{W} \mathbf{h} + \mathbf{a}^\top \mathbf{v} + \mathbf{b}^\top \mathbf{h}\right), \tag{3}$$

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.

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

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

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

Hint: You may use that

$$p(\mathbf{h}|\mathbf{v}) = \prod_{i=1}^{m} p(h_i|\mathbf{v}), \qquad p(h_i = 1|\mathbf{v}) = \frac{1}{1 + \exp\left(-\sum_j v_j W_{ji} - b_i\right)}, \tag{5}$$

$$p(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^{n} p(v_i|\mathbf{h}), \qquad p(v_i = 1|\mathbf{h}) = \frac{1}{1 + \exp\left(-\sum_j W_{ij} h_j - a_i\right)}. \tag{6}$$