

**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  $\alpha$ ,

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

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

$$F_x(\alpha) = \int_{-\infty}^{\alpha} p_x(u) du, \quad (2)$$

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  $[0, 1]$ :  $y^i \sim \mathcal{U}(0, 1)$ ,  $i = 1, \dots, n$ .
- transforming each sample by  $F_x^{-1}$ :  $x^i = F_x^{-1}(y^i)$ ,  $i = 1, \dots, n$ .

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

**Solution.** We start with the cumulative distribution function (cdf)  $F_y$  for  $y$ ,

$$F_y(\beta) = \mathbb{P}(y \leq \beta). \quad (\text{S.1})$$

Since  $F_x(x)$  maps  $x$  to  $[0, 1]$ ,  $F_y(\beta)$  is zero for  $\beta < 0$  and one for  $\beta > 1$ . We next consider  $\beta \in [0, 1]$ .

Let  $\alpha$  be the value of  $x$  that  $F_x$  maps to  $\beta$ , i.e.  $F_x(\alpha) = \beta$ , which means  $\alpha = F_x^{-1}(\beta)$ . Since  $F_x$  is a non-decreasing function, we have

$$F_y(\beta) = \mathbb{P}(y \leq \beta) = \mathbb{P}(F_x(x) \leq \beta) = \mathbb{P}(x \leq F_x^{-1}(\beta)) = \mathbb{P}(x \leq \alpha) = F_x(\alpha). \quad (\text{S.2})$$

Since  $\alpha = F_x^{-1}(\beta)$  we obtain

$$F_y(\beta) = F_x(F_x^{-1}(\beta)) = \beta \quad (\text{S.3})$$

The cdf  $F_y$  is thus given by

$$F_y(\beta) = \begin{cases} 0 & \text{if } \beta < 0 \\ \beta & \text{if } \beta \in [0, 1] \\ 1 & \text{if } \beta > 1 \end{cases} \quad (\text{S.4})$$

which is the cdf of a uniform random variable on  $[0, 1]$ . Hence  $y = F_x(x)$  is uniformly distributed on  $[0, 1]$ .

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

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

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

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

**Solution.** The main task is to compute the cumulative distribution function (cdf)  $F_x$  of  $x$  and its inverse. The cdf is by definition

$$F_x(\alpha) = \int_{-\infty}^{\alpha} \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|u|\right) du. \quad (S.5)$$

We first consider the case where  $\alpha \leq 0$ . Since  $-|u| = u$  for  $u \leq 0$ , we have

$$F_x(\alpha) = \int_{-\infty}^{\alpha} \frac{1}{\sqrt{2}} \exp\left(\sqrt{2}u\right) du \quad (S.6)$$

$$= \frac{1}{2} \exp\left(\sqrt{2}u\right) \Big|_{-\infty}^{\alpha} \quad (S.7)$$

$$= \frac{1}{2} \exp\left(\sqrt{2}\alpha\right). \quad (S.8)$$

For  $\alpha > 0$ , we have

$$F_x(\alpha) = \int_{-\infty}^{\alpha} \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|u|\right) du \quad (S.9)$$

$$= 1 - \int_{\alpha}^{\infty} \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|u|\right) du \quad (S.10)$$

where we have used the fact that the pdf has to integrate to one. For values of  $u > 0$ ,  $-|u| = -u$ , so that

$$F_x(\alpha) = 1 - \int_{\alpha}^{\infty} \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}u\right) du \quad (S.11)$$

$$= 1 + \frac{1}{2} \exp\left(-\sqrt{2}u\right) \Big|_{\alpha}^{\infty} \quad (S.12)$$

$$= 1 - \frac{1}{2} \exp\left(-\sqrt{2}\alpha\right). \quad (S.13)$$

In total, for  $\alpha \in \mathbb{R}$ , we thus have

$$F_x(\alpha) = \begin{cases} \frac{1}{2} \exp\left(\sqrt{2}\alpha\right) & \text{if } \alpha \leq 0 \\ 1 - \frac{1}{2} \exp\left(-\sqrt{2}\alpha\right) & \text{if } \alpha > 0 \end{cases} \quad (S.14)$$

Figure 1 visualises  $F_x(\alpha)$ .

As the figure suggests, there is a unique inverse to  $y = F_x(\alpha)$ . For  $y \leq 1/2$ , we have

$$y = \frac{1}{2} \exp\left(\sqrt{2}\alpha\right) \quad (S.15)$$

$$\log(2y) = \sqrt{2}\alpha \quad (S.16)$$

$$\alpha = \frac{1}{\sqrt{2}} \log(2y) \quad (S.17)$$

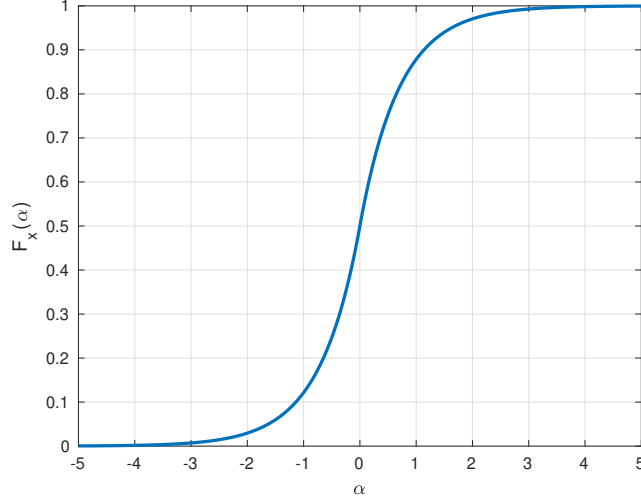


Figure 1: The cumulative distribution function  $F_x(\alpha)$  for a Laplace distributed random variable.

For  $y > 1/2$ , we have

$$y = 1 - \frac{1}{2} \exp(-\sqrt{2}\alpha) \quad (\text{S.18})$$

$$-y = -1 + \frac{1}{2} \exp(-\sqrt{2}\alpha) \quad (\text{S.19})$$

$$1 - y = \frac{1}{2} \exp(-\sqrt{2}\alpha) \quad (\text{S.20})$$

$$\log(2 - 2y) = -\sqrt{2}\alpha \quad (\text{S.21})$$

$$\alpha = -\frac{1}{\sqrt{2}} \log(2 - 2y) \quad (\text{S.22})$$

The function  $y \mapsto g(y)$  that occurs in the log

$$g(y) = \begin{cases} 2y & \text{if } y \leq \frac{1}{2} \\ 2 - 2y & \text{if } y > \frac{1}{2} \end{cases} \quad (\text{S.23})$$

is shown below and can be written as  $g(y) = 1 - 2|y - 1/2|$ .

We thus can write the inverse  $F_x^{-1}(y)$  of the cdf  $y = F_x(\alpha)$  as

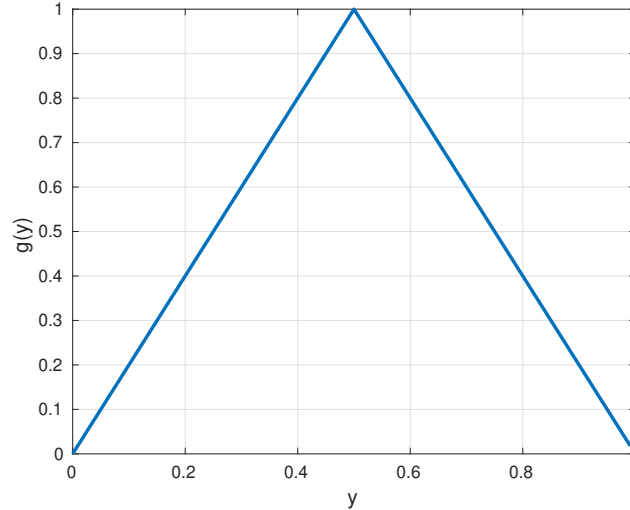
$$F_x^{-1}(y) = -\text{sign}\left(y - \frac{1}{2}\right) \frac{1}{\sqrt{2}} \log\left[1 - 2\left|y - \frac{1}{2}\right|\right]. \quad (\text{S.24})$$

To generate  $n$  iid samples from  $x$ , we first generate  $n$  iid samples  $y^i$  that are uniformly distributed on  $[0, 1]$ , and then compute for each  $F_x^{-1}(y^i)$ . The properties of inverse transform sampling guarantee that the  $x^i$ ,

$$x^i = F_x^{-1}(y^i) \quad (\text{S.25})$$

are independent and Laplace distributed.

Inverse transform sampling can be used to generate samples from many standard distributions. For example, it allows one to generate Gaussian random variables from uniformly distributed random variables. The method is called the Box-Muller transform, see e.g. <https://en>.



[wikipedia.org/wiki/Box-Muller\\_transform](https://wikipedia.org/wiki/Box-Muller_transform). How to generate the required samples from the uniform distribution is a research field on its own, see e.g. [https://en.wikipedia.org/wiki/Random\\_number\\_generation](https://en.wikipedia.org/wiki/Random_number_generation) and <http://statweb.stanford.edu/~owen/mc/Ch-unifrng.pdf>.

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

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

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

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  $p(\mathbf{v})$ ,

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

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

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

**Solution.** In order to generate samples  $\mathbf{v}^{(k)}$  from  $p(\mathbf{v})$  we generate samples  $(\mathbf{v}^{(k)}, \mathbf{h}^{(k)})$  from  $p(\mathbf{v}, \mathbf{h})$  and then ignore the  $\mathbf{h}^{(k)}$ .

Gibbs sampling is a MCMC method to produce a sequence of samples  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \dots$  that follow a pdf/pmf  $p(\mathbf{x})$  (if the chain is run long enough). Assuming that  $\mathbf{x}$  is  $d$ -dimensional, we generate the next sample  $\mathbf{x}^{(k+1)}$  in the sequence from the previous sample  $\mathbf{x}^{(k)}$  by:

1. picking (randomly) an index  $i \in \{1, \dots, d\}$
2. sampling  $x_i^{(k+1)}$  from  $p(x_i | \mathbf{x}_{\setminus i}^{(k)})$  where  $\mathbf{x}_{\setminus i}^{(k)}$  is vector  $\mathbf{x}$  with  $x_i$  removed, i.e.  $\mathbf{x}_{\setminus i}^{(k)} = (x_1^{(k)}, \dots, x_{i-1}^{(k)}, x_{i+1}^{(k)}, \dots, x_d^{(k)})$

3. setting  $\mathbf{x}^{(k+1)} = (x_1^{(k)}, \dots, x_{i-1}^{(k)}, x_i^{(k+1)}, x_{i+1}^{(k)}, \dots, x_d^{(k)})$ .

For the RBM, the tuple  $(\mathbf{h}, \mathbf{v})$  corresponds to  $\mathbf{x}$  so that a  $x_i$  in the above steps can either be a hidden variable or a visible. Hence

$$p(x_i | \mathbf{x}_{\setminus i}) = \begin{cases} p(h_i | \mathbf{h}_{\setminus i}, \mathbf{v}) & \text{if } x_i \text{ is a hidden variable } h_i \\ p(v_i | \mathbf{v}_{\setminus i}, \mathbf{h}) & \text{if } x_i \text{ is a visible variable } v_i \end{cases} \quad (\text{S.26})$$

( $\mathbf{h}_{\setminus i}$  denotes the vector  $\mathbf{h}$  with element  $h_i$  removed, and equivalently for  $\mathbf{v}_{\setminus i}$ )

To compute the conditionals on the right hand side, we use the following result from Tutorial 2 (additional practice):

$$p(\mathbf{h}|\mathbf{v}) = \prod_{i=1}^m p(h_i|\mathbf{v}), \quad p(h_i = 1|\mathbf{v}) = \frac{1}{1 + \exp\left(-\sum_j v_j W_{ji} - b_i\right)}, \quad (\text{S.27})$$

$$p(\mathbf{v}|\mathbf{h}) = \prod_{i=1}^n p(v_i|\mathbf{h}), \quad p(v_i = 1|\mathbf{h}) = \frac{1}{1 + \exp\left(-\sum_j W_{ij} h_j - a_i\right)}. \quad (\text{S.28})$$

Given the independencies between the hidden given the visible and vice versa, we have

$$p(h_i | \mathbf{h}_{\setminus i}, \mathbf{v}) = p(h_i | \mathbf{v}) \quad p(v_i | \mathbf{v}_{\setminus i}, \mathbf{h}) = p(v_i | \mathbf{h}) \quad (\text{S.29})$$

so that the expressions for  $p(h_i = 1|\mathbf{v})$  and  $p(v_i = 1|\mathbf{h})$  allow us to implement the Gibbs sampler.

Given the independencies, it makes further sense to sample the  $\mathbf{h}$  and  $\mathbf{v}$  variables in blocks: first we sample all the  $h_i$  given  $\mathbf{v}$ , and then all the  $v_i$  given the  $\mathbf{h}$  (or vice versa). This is also known as block Gibbs sampling.

In summary, given a sample  $(\mathbf{h}^{(k)}, \mathbf{v}^{(k)})$ , we thus generate the next sample  $(\mathbf{h}^{(k+1)}, \mathbf{v}^{(k+1)})$  in the sequence as follows:

- For all  $h_i$ ,  $i = 1, \dots, m$ :
  - compute  $p_i^h = p(h_i = 1|\mathbf{v}^{(k)})$
  - sample  $u_i$  from a uniform distribution on  $[0, 1]$  and set  $h_i^{(k+1)}$  to 1 if  $u_i \leq p_i^h$ .
- For all  $v_i$ ,  $i = 1, \dots, n$ :
  - compute  $p_i^v = p(v_i = 1|\mathbf{h}^{(k+1)})$
  - sample  $u_i$  from a uniform distribution on  $[0, 1]$  and set  $v_i^{(k+1)}$  to 1 if  $u_i \leq p_i^v$ .

As final step, after sampling  $S$  pairs  $(\mathbf{h}^{(k)}, \mathbf{v}^{(k)})$ ,  $k = 1, \dots, S$ , the set of visible  $\mathbf{v}^{(k)}$  form samples from the marginal  $p(\mathbf{v})$ .