Computational Stochastic Processes
Problem Sheet 1

U. Vaes
with problems from G. Pavliotis, A. Duncan and S. Gomes

February 2020

You are free to return a selection of your work to me for marking. This is entirely optional and the mark will not count for assessment.

Generating non-uniform random variables

Problem 1 (Generalized Bernoulli Distribution)**.** Suppose X is a discrete valued random variable taking values i with probability p_i for $i \in \{1, \ldots, k\}$ where $\sum_{i=1}^{k} p_i = 1$.

1. Write down the CDF $F(x)$ for the probability distribution of this r.v..

   Solution. *One can easily check that*

   $$F(x) = \mathbb{P}[X \leq x] = \sum_{i=1}^{\lfloor x \rfloor} p_i$$

2. Write down an expression for the generalised inverse of the CDF $F(x)$.

   Solution. *Let G be the generalised inverse of F, then for $y \in (0, 1]$:*

   $$G(y) = \inf \left\{ j \in \{1, \ldots, k\} \mid \sum_{i=1}^{j-1} p_i < y \leq \sum_{i=1}^{j} p_i \right\}$$

3. Use the inverse transform method to derive an algorithm to sample from this distribution.

   Solution. *(a) Sample $u \sim U(0, 1)$.*
   (b) Set $j = 1$.
   (c) While $u > \sum_{i=1}^{j} p_i$, set $j := j + 1$.
   (d) Output $X = j$.

4. Implement a sampler based on this scheme using a programming language of your choice.

5. For $k = 4$, $p_1, p_2, p_3, p_4 = 0.125, 0.125, 0.375, 0.375$ generate $N = 10^3$ samples and generate a normalized histogram from this sample to verify that each value is generated with the correct probability.

Solution. *See the Jupyter notebook. We can check that the right proportion of each of the integers, $j = 1, 2, 3, 4$ was generated.*

Problem 2 (Sample from Gamma(k, λ) distribution)**.** When $k \in \mathbb{N}$, it is known that

$$X_1 + \ldots + X_k \sim \text{Gamma}(k, \lambda),$$

where $X_1, \ldots, X_k$ are iid Exp(λ) distributed random variables.

1. Based on this observation, write a scheme to generate Gamma(k, λ) distributed samples, where $k \in \mathbb{N}$.

   Solution. *We use the inverse transform method to generate k, $Exp(\lambda)$ random variables, and add them, i.e.*

   (a) Let $u_1, \ldots, u_k \sim U(0, 1)$ iid.
   (b) Set $X = -\frac{1}{\lambda} \sum_{i=1}^{k} \log(u_i)$.

2. Suppose now that $k \in \mathbb{R}_{\geq 0}$. We wish to implement a rejection sampler for Gamma(k, λ) using proposal density of the form

   $$g(x) = \frac{\lambda_0^{k_0}}{\Gamma(k_0)} x^{k_0 - 1} e^{-\lambda_0 x},$$

   where $k_0 = \lfloor k \rfloor$, and $\lambda_0 > 0$. Calculate the upper bound $M = \sup_{x \geq 0} f(x)/g(x)$, and show that necessarily $\lambda < \lambda_0$ and

   $$M = \frac{\lambda^k}{\Gamma(k)} \frac{\Gamma(k_0)}{\lambda_0^{k_0}} \left(\frac{k - k_0}{\lambda - \lambda_0} \right)^{k - k_0} e^{-(k - k_0)}.$$

   Solution. *The pdf of X is given by*

   $$f(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}, \quad k > 1.$$

   The ratio $f(x)/g(x)$ is given by

   $$\frac{\lambda^k}{\Gamma(k)} \frac{\Gamma(k_0)}{\lambda_0^{k_0}} x^{k - k_0} e^{-(\lambda - \lambda_0)x}.$$

   We want a bound for f/g for all $x \geq 0$. In particular, it is automatically bounded at $x = 0$ since $k \geq k_0$. But to be bounded as $x \to +\infty$, we require that $\lambda - \lambda_0 > 0$. Under this assumption it is straightforward to check that the maximum is obtained at x^* given by

   $$x^* = \frac{k - k_0}{\lambda - \lambda_0}.$$

For this value of x, we have

$$\begin{aligned} M &= f(x^*)/g(x^*) \\ &= \frac{\lambda^k}{\Gamma(k)} \frac{\Gamma(k_0)}{\lambda_0^{k_0}} (x^*)^{k-k_0} e^{-(\lambda-\lambda_0)x^*} \\ &= \frac{\lambda^k}{\Gamma(k)} \frac{\Gamma(k_0)}{\lambda_0^{k_0}} (x^*/e)^{k-k_0}. \end{aligned}$$

3. Optimise over λ_0 to minimise M.

   Solution. *Take the derivative of M with respect to λ_0. After some simplification we see that the minimum occurs when*

   $$\lambda_0 = \frac{k_0}{k}\lambda,$$

   and since $\lambda_0 < \lambda$, this is consistent with out previous assumption.

4. Implement a rejection sampling scheme to sample from Gamma distribution. Implement also a naive rejection sampler based on a proposal distributed according to the standard Cauchy distribution, and compare the performance of both.

   Solution. *See the Jupyter notebook. I suggest you run it for a large N (say 100000), with fixed λ and vary k (for example, $k = 1.1, 4.5, 9.5$). You should notice that the rejection based sampler with Cauchy proposals does not scale well as $k \to \infty$, whereas our new approach has run-time almost independent of k.*

Monte Carlo simulation

Problem 3 (Monte Carlo Simulation)**.** We want to estimate the area inside the Batman curve via Monte Carlo simulation. To this end, let (X, Y) be a uniformly distributed random variable in the rectangle $[-L_x, L_x] \times [-L_y, L_y]$, with $L_x = 7.25$ and $L_y = 4$, and let $f(\cdot, \cdot)$ denote the indicator function of the surface inside the Batman curve. For the student's convenience, this function is implemented in *Python* below:

```
def batman_indicator(x, y):

    # We'll initialize at one and remove parts one by one
    result = np.ones(x.shape)

    # Ellipse
    ellipse = (x/7)**2 + (y/3)**2 - 1 >= 0
    result[np.where(ellipse)] = 0

    # Bottom curve on [-3, 3]
    bottom = (abs(x) < 4) * \
             (y <= abs(x/2) - ((3*np.sqrt(33)-7)/112)*x**2 - 3
              + np.sqrt(np.maximum(0, 1-(abs(abs(x)-2) - 1)**2)))
    result[np.where(bottom)] = 0

    # Top curve
    top = (abs(x) > .75) * (abs(x) < 1) * (y > 9 - 8*abs(x)) \
```

```
          + (abs(x) > .5) * (abs(x) < .75) * (y > 3*abs(x) + .75) \
          + (abs(x) < .5) * (y > 2.25) \
          + (abs(x) > 1) * (abs(x) < 3) * \
            (y > (6*np.sqrt(10)/7+(1.5-.5*abs(x))-(6*np.sqrt(10)/14)*\
                  np.sqrt(np.maximum(0, 4-(abs(x)-1)**2))))
    result[np.where(top)] = 0
    return result
```

The exact area of the Batman sign is given below:

```
# Exact area
I = (955/48) - (2/7) * (2*np.sqrt(33) + 7*np.pi + 3*np.sqrt(10) * (np.pi - 1)) \
    + 21 * (np.arccos(3/7) + np.arccos(4/7))
```

1. Using a sample of size $n = 1000$, generate a 95% confidence interval for the area of the Batman sign, $I := 4\,L_x\,L_y\,\mathbb{E}[f(X,Y)]$ based on

   (a) Chebychev's inequality;

   (b) The central limit theorem (CLT);

   (c) Bikelis' theorem.

   Solution. *We define*

   $$\hat{I}_n = (4\,L_x\,L_y)\,\frac{1}{n}\sum_{i=1}^{n} f(X_i,Y_i), \qquad (X_i,Y_i) \sim \mathcal{U}([-L_x,L_x]\times[-L_y,L_y])$$

   From the lecture notes, we know that the $(1-\alpha)$ *confidence intervals are given by*

   (a) Chebychev's inequality: $\left[\hat{I}_N - \frac{\sigma}{\sqrt{N\alpha}}, \hat{I}_N + \frac{\sigma}{\sqrt{N\alpha}}\right]$.

   (b) The CLT: $\left[\hat{I}_N - c_\alpha\frac{\sigma}{\sqrt{N}}, \hat{I}_N + c_\alpha\frac{\sigma}{\sqrt{N}}\right]$, *where* c_α *is such that* $2(1-\phi(c_\alpha)) = \alpha$ *and* $\phi(\cdot)$ *is the CDF of* $\mathcal{N}(0,1)$.

   (c) By Bikelis' theorem (see lecture notes),

   $$\mathbb{P}\left(\left|\frac{\hat{I}_n - I}{\sigma/\sqrt{n}}\right| \le a\right) \ge 2(1-\phi(a)) - \frac{2\,\xi}{\sqrt{n}\,\sigma^3\,(1+|a|)^3} =: \Psi_n(a).$$

   where $\sigma^2 = \mathbb{E}[Z^2]$ *and* $\xi = \mathbb{E}[Z^3]$, *where* $Z = 4\,L_x\,L_y\,f(X,Y) - I$. *Numerically, we can estimate* σ *and* ξ *based on the data from the Monte Carlo simulation, and construct a confidence interval by finding the smallest possible* a *such that* $\Psi_n(a) \ge 1-\alpha = .95$.

2. Perform this previous simulation 1000 times independently, and measure how many of the reported confidence intervals actually contained I.

   Solution. *Running this for* $\alpha = 0.05$ *and* $N = 1000$, *we see that approximately* 95% *of the confidence intervals obtained from the CLT contain* I, *while* 100% *of the confidence intervals obtained from Chebychev contain* I, *which suggests that the Chebychev confidence intervals are too loose!*

3. Use Hoeffding's inequality to derive $100(1-\alpha)$%-confidence intervals and compare with the ones obtained via Chebychev and the CLT.

Theorem 1. *Let $Z_1, \ldots, Z_n$ be iid random variables supported within the interval $[0,1]$ with mean μ. Then*

$$\mathbb{P}\left[|S_n - \mu| \geq a\right] \leq 2e^{-2na^2},$$

where $S_n = n^{-1}\sum_{i=1}^n Z_i$ and $a > 0$.

Solution. *We apply Hoeffding's inequality with the random variables $Z_i = f(X_i, Y_i)$ We need*

$$2e^{-2na^2} \leq \alpha,$$

or equivalently

$$a \geq \sqrt{\frac{-\log(\alpha/2)}{2n}},$$

for confidence $100(1-\alpha)\%$. For a satisfying this inequality,

$$\mathbb{P}\left[|\hat{I}_n - I| \geq 4\, L_x\, L_y a\right] \leq 2e^{-2na^2} \leq \alpha.$$

Repeating the computation above with the new confidence interval (the code is identical), we observe that the Hoeffding confidence intervals contain I with probability approximately 1. This implies that the bounds are somewhat coarse. This is not too surprising in this case: the Hoeffding inequality does not depend on the variance of the estimator! Without this information, it is natural that the confidence intervals would be quite conservative.

4. Find a control variate that enables a reduction of the variance of the estimator.

   Solution. *See the Jupyter notebook.*

Problem 4 (Density estimation using Histograms)**.** A very common problem in computational stochastic methods is the estimation of density, i.e. given a stream of iid samples of some rv X we wish to accurately recover the density $p(x)$ of X. The MC approach to density estimation is to express the density $p(x)$ as a limit of expectations of random variables, namely

$$p(x) = \lim_{h\to 0} \frac{\mathbb{E}\mathbf{1}\left[x < X < x + h\right]}{h}.$$

Thus, for small h, we use the following MC estimator

$$\hat{p}_n(x) = \frac{1}{hn}\sum_{i=1}^n \mathbf{1}\left[x < X_i < x + h\right]$$

1. Assume that $p \in C^1(\mathbb{R})$, use Taylor's theorem to show that $\hat{p}_n(x)$ is asymptotically biased (and thus biased). Show that the bias goes to zero as $h \to 0$.

   Solution. *Here is a simpler proof that doesn't need Taylor's theorem.*

$$\mathbb{E}[\hat{p}_n(x)] = \frac{1}{h}\mathbb{E}_{X\sim p}(\mathbf{1}[x < X < x + h]) = \frac{1}{h}\int_x^{x+h} p(y)\, dy = \frac{1}{h}(F(x+h) - F(x)),$$

where F is the CDF of p and so, if $F \in C^1(\mathbb{R})$

$$\lim_{h\to 0} \mathbb{E}[\hat{p}_n(x)] = p(x).$$

Indeed, we don't even need $F \in C^1$, only that p is integrable, using the Lebesgue differentiation theorem.

2. Compute the variance of the estimator $\hat{p}(x)$. Show that $\mathrm{Var}[\hat{p}_n(x)] \to \infty$ as $h \to 0$.

   Solution. *It is straightforward to check that*

$$n \,\mathrm{Var}[\hat{p}_n(x)] = \frac{1}{h^2}(F(x+h) - F(x))(1 - F(x+h) + F(x)).$$

   which is asymptotic to $p(x)/h$ as $h \to 0$. Therefore $\mathrm{Var}[\hat{p}_n(x)] \to \infty$ as $h \to 0$.

Taking $h \to 0$, the bias goes to zero, while the variance blows up. This suggests that a good choice of h must involve some kind of "trade-off" between variance and bias.

3. Assuming again that $p \in C^1(\mathbb{R})$. Write an expression for the MSE of $\hat{p}_n$ which is correct to $o(h)$.

   Solution. *Clearly the bias induced by a single sample is given by*

$$\mathbb{E}\left[\frac{1}{h}\mathbf{1}[x < X < x+h] - p(x)\right] = \frac{F(x+h) - F(x) - hF'(x)}{h}.$$

   which, by Taylor's theorem, equals $hF''(x)/2 + h^2F'''(x')/6 = h\,p'(x)/2 + O(h^2)$, if we assume that $p''(x)$ is bounded. Therefore the mean squared error is given by

$$MSE(\hat{p}_n) \approx \frac{p(x)}{hn} + \frac{h^2}{4}p'(x)^2. \tag{1}$$

4. Find the value of h which minimises the MSE. Conclude that MSE is minimized taking $h = O(N^{-1/3})$, in which case the MSE goes to zero with rate $N^{-2/3}$.

   Solution. *For a fixed N we should select h so that it roughly balances the terms on the RHS of (1). Writing the RHS of (1) as $a/h + bh^2$, we see that the minimum is attained when $h = (a/2b)^{1/3}$, so that*

$$h_* = \left(\frac{2p(x)}{p'(x)^2}\right)^{1/3} N^{-1/3},$$

   is the choice which minimises the MSE, at which value

$$MSE(\hat{p}_n) \approx KN^{-2/3}.$$

Variance reduction techniques

Problem 5 (Importance Sampling)**.** Consider the problem of estimating the moments of the distribution

$$p(x) = \frac{1}{2}e^{-|x|},$$

called the double exponential density. The CDF of this function is

$$F(x) = \frac{1}{2}e^x\mathbf{1}[x \le 0] + \frac{1}{2}(1 - e^{-x}/2)\mathbf{1}[x > 0],$$

which is a piecewise function and difficult to invert. Indeed, we cannot "easily" sample from this distribution. Suppose we wish to estimate the second moment of the distribution $\mathbb{E}[X^2]$.

1. Using importance sampling distribution $\mathcal{N}(0, 4)$ construct an importance sampler for computing $\mathbb{E}[X^2]$.

   Solution. *Our objective is to calculate the integral*

   $$I = \int_{-\infty}^{\infty} x^2 \frac{1}{2} e^{-|x|}\, dx.$$

   We can rewrite this as

   $$\int_{-\infty}^{\infty} x^2 \frac{\frac{1}{2}e^{-|x|}}{\frac{1}{\sqrt{8\pi}}e^{-x^2/8}} \frac{1}{\sqrt{8\pi}} e^{-x^2/8}\, dx.$$

   Note that $\frac{1}{\sqrt{8\pi}}e^{-x^2/8}$ *is the* $\mathcal{N}(0, 4)$ *density. We can construct the following importance sampler for* I*:*

   $$\hat{I}_n^{is} = \sum_{i=1}^{n} X_i^2 \frac{\frac{1}{2}e^{-|X_i|}}{\frac{1}{\sqrt{8\pi}}e^{-X_i^2/8}},$$

   where $X_1, X_2, \ldots$ *are iid* $\mathcal{N}(0, 4)$ *samples.*

2. Implement this sampler in a programming language of your choice, and generate 10^5 samples, and compute the mean. The true value of this expectation should be 2.

3. Can you use the expression for the variance of the importance sampler (or some other method) to find a better choice of σ^2 for a proposal distribution $\mathcal{N}(0, \sigma^2)$? Implement this scheme and compare the performance computationally.

   Solution. *See the Jupyter notebook.*

In machine learning one often needs to compute expectations with respect to the *Gumbel distribution*

$$p(x) = \exp(x - \exp(x)).$$

4. Show that $\mathbb{E}[\exp(X)] < \infty$, where $X \sim p$.

   Solution.

   $$\mathbb{E}[\exp(X)] = \int e^{2x - e^x}\, dx,$$

   but clearly $e^{2x-e^x} \leq e^{-|x|}$ *(check for positive and negative values of* x*). Therefore*

   $$\mathbb{E}[\exp(X)] \leq \int_{\mathbb{R}} e^{-|x|}\, dx = 2$$

5. Using a standard Gaussian importance distribution, implement a regular importance sampler approximating $\mathbb{E}[\exp(X)]$ in a programming language of your choice.

6. Similarly implement a self-normalized version of the importance sampler.

7. Compare the performance of both by computing the variance of both estimators, approximated over 10^3 independent runs.

   Solution. *See Jupyter notebook.*

Problem 6 (Gambler's ruin)**.** Here we consider again a problem that was discussed in the workbook on variance reduction techniques. Assume that $\{Z_i\}_{i=0}^{N-1}$ are independent $\mathcal{N}(0,\sigma^2)$ random variables and define

$$S_k = s_0 + \sum_{i=0}^{k-1} Z_i, \qquad k = 1, \dots, N.$$

we want to calculate the probability of ruin within the first N games, given by

$$I = \mathbb{P}\left(\min_{k\in\{1,\dots,N\}} S_k \leq 0 \right),$$

In order to better estimate I with a Monte Carlo method, we will use importance sampling with an important distribution given by the PDF of the $\mathbb{R}^N$-valued random variable V obtained by

$$V_k = s_0 + \sum_{i=0}^{k-1} b(V_i) + \sum_{i=0}^{k-1} Z_i, \qquad k = 1, \dots, N, \tag{2}$$

where $b(\cdot)$ is real-valued function.

1. Calculate the likelihood ratio $g(v)$ between the PDF of $S = (S_1, \dots, S_N)^T$ and that of $V = (V_1, \dots, V_N)^T$.

   Solution. *The calculation of the PDF of V is based on the fact that, if X and Y are two random variables with joint PDF $f_{X,Y}(x,y)$, then*

   $$f_{X,Y}(x,y) = f_X(x)\, f_{Y|X}(y|x)$$

   Here $f_X(x)$ is the marginal density for X and $f_{Y|X}(y|x)$ is the conditional density function of Y given that $X = x$. Applying this equality repeatedly

   $$\begin{aligned} f_{V_1,\dots,V_n}(v_1,\dots,v_n) &= f_{V_1,\dots,V_{n-1}}(v_1,\dots,v_{n-1})\, f_{V_n|V_1,\dots,V_{n-1}}(v_n|v_1,\dots,v_n) \\ &= f_{V_1,\dots,V_{n-2}}(v_1,\dots,v_{n-2})\, f_{V_{n-1}|V_1,\dots,V_{n-2}}\, f_{V_n|V_1,\dots,V_{n-1}}(v_n|v_1,\dots,v_{n-1}) \\ &= \dots \\ &= f_{V_1}(v_1)\, f_{V_2|V_1}(v_2|v_1)\, f_{V_3|V_1,V_2}(v_3|v_1,v_2)\dots \end{aligned}$$

   Since V_{k+1} depends only on V_k, we can simplify the previous equation:

   $$f_{V_1,\dots,V_n}(v_1,\dots,v_n) = f_{V_1}(v_1)\, f_{V_2|V_1}(v_2|v_1)\, f_{V_3|V_2}(v_3|v_2)\ \dots$$

   (Note that we do not write $f_{V_1|V_0}(v_1|v_0)$ because $V_0 = s_0$ was assumed to be deterministic.) Next, observe that the conditional densities are Gaussians,

   $$f_{V_{k+1}|V_k}(v_{k+1}|v_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \left| v_{k+1} - v_k - b(v_k) \right|^2 \right),$$

   which leads to

   $$f_{V_1,\dots,V_N}(v_1,\dots,v_N) = \left| \frac{1}{\sqrt{2\pi\sigma^2}} \right|^N \exp\left(-\frac{1}{2\sigma^2} \sum_{k=0}^{N-1} \left| v_{k+1} - v_k - b(v_k) \right|^2 \right).$$

   Let $\psi(v) := f_{V_1,\dots,V_N}(v_1,\dots,v_N)$. With a similar reasoning, we obtain that the probability

density function of $(S_1, \ldots, S_N)$ *is given by*

$$\pi(v) = \left| \frac{1}{\sqrt{2\pi\sigma^2}} \right|^N \exp\left(-\frac{1}{2\sigma^2} \sum_{k=0}^{N-1} |v_{k+1} - v_k|^2 \right).$$

The likelihood ratio, i.e. the ratio between the nominal and importance PDFs, is given by

$$g(v) := \frac{\pi(v)}{\psi(v)} = \exp\left(-\frac{1}{\sigma^2} \left(\sum_{k=0}^{N-1} (v_{k+1} - v_k)\, b(v_k) - \frac{1}{2} \sum_{k=0}^{N-1} |b(v_k)|^2 \right) \right). \tag{3}$$

2. Show that, if $V = (V_1, \ldots, V_N)^T$ is obtained from (2), then the likelihood ratio evaluated at V admits the following expression:

$$g(V) = \exp\left(-\frac{1}{\sigma^2} \left(\sum_{k=0}^{N-1} b(V_k)\, Z_k + \frac{1}{2} \sum_{k=0}^{N-1} |b(V_k)|^2 \right) \right), \tag{4}$$

where we used the notation $V_0 = s_0$.

Solution. *This follows directly from the fact that*

$$V_{k+1} - V_k = b(V_k) + Z_k, \qquad k = 0, \ldots, N-1,$$

with again $V_0 = s_0$. *Note that* (5) *is a pointwise equality of random variables: for any* ω *in the sample space* Ω,

$$g(V(\omega)) = \exp\left(-\frac{1}{\sigma^2} \left(\sum_{k=0}^{N-1} b(V_k(\omega))\, Z_k(\omega) + \frac{1}{2} \sum_{k=0}^{N-1} |b(V_k(\omega))|^2 \right) \right). \tag{5}$$

3. Calculate the expectation $\mathbb{E}[g(V)]$. Was the result expected?

Solution. *For* $k \in \{0, \ldots, N\}$, *let*

$$g_k(V) = \exp\left(-\frac{1}{\sigma^2} \left(\sum_{i=0}^{k-1} b(V_i)\, Z_i + \frac{1}{2} \sum_{i=0}^{k-1} |b(V_i)|^2 \right) \right),$$

with the convention that the empty sum is zero. Clearly $g_0(V) = 1$ *and we calculate that*

$$g_{k+1}(V) = g_k(V) \exp\left(-\frac{1}{\sigma^2} \left(b(V_k)\, Z_k + \frac{1}{2} |b(V_k)|^2 \right) \right).$$

By the "pulling out known factors" property of conditional expectation, we obtain that

$$\mathbb{E}[g_{k+1}(V) | Z_0, \ldots, Z_{k-1}] = g_k(V)\, \mathbb{E}\left[\exp\left(-\frac{1}{\sigma^2} \left(b(V_k)\, Z_k + \frac{1}{2} |b(V_k)|^2 \right) \right) | Z_0, \ldots, Z_{k-1} \right].$$

For $Z_0, \ldots, Z_{k-1}$ *given, the argument the exponential is a normal random variable with mean* $m = -\frac{1}{2\sigma^2} |b(V_k)|^2$ *and variance* $s^2 = \frac{1}{\sigma^2} |b(V_k)|^2$. *By the properties of the log-normal distribution, we deduce*

$$\mathbb{E}[g_{k+1}(V) | Z_0, \ldots, Z_{k-1}] = g_k(V)\, e^{m + \frac{s^2}{2}} = g_k(V).$$

It then follows from a simple recursion that $\mathbb{E}[g(V)] = 1$. *This was expected because*

$$\mathbb{E}[g(V)] = \int_{\mathbb{R}^N} \left(\frac{\pi(v)}{\psi(v)}\right) \psi(v)\, dv = 1.$$

4. For the parameters $N = 10$ and $\sigma = .2$, calculate I by using importance sampling using the modified dynamics (2). Can you find a choice of the function $b(\cdot)$ that produces better results than the constant function $b(\cdot) = -.1$?

   Solution. *See the Jupyter notebook.*

Problem 7 (Control Variates)**.** Let $X \sim p$ and suppose we want to evaluate

$$\mathbb{P}(X > a) = \int_a^\infty p(x)\, dx.$$

Suppose that p is symmetric around zero, so that $\mathbb{P}(X > 0) = \frac{1}{2}$. Form a control variate estimator

$$\hat{I}_n^c = \frac{1}{n}\sum_{i=1}^n \left[\mathbf{1}(X_i > a) + \alpha\left(\mathbf{1}(X_i > 0) - \frac{1}{2}\right)\right]$$

1. Compute the variance $\mathrm{Var}[\hat{I}_n^c]$.

   Solution. *Let* $X_i \sim p$, *then we can compute the required variances and covariance as follows:*

$$\mathrm{Var}[\mathbf{1}[X_i > a]] = \int \mathbf{1}[X_i > a]^2 p(x)\, dx - \left(\int \mathbf{1}[X_i > a] p(x)\, dx\right)^2 = \mathbb{P}[X_i > a](1 - \mathbb{P}[X_i > a]),$$

$$\mathrm{Var}[\mathbf{1}[X_i > 0]] = \int \mathbf{1}[X_i > 0]^2 p(x)\, dx - \left(\int \mathbf{1}[X_i > 0]^2 p(x)\, dx\right)^2 = \frac{1}{4},$$

   and

$$Cov[\mathbf{1}[X_i > 0], \mathbf{1}[X_i > a]] = \frac{1}{2}\mathbb{P}[X_i > a].$$

   Therefore,

$$\mathrm{Var}[\hat{I}_n^c] = \frac{1}{n}\left(\mathbb{P}[X_i > a](1 - \mathbb{P}[X_i > a]) + \alpha\mathbb{P}[X_i > a] + \frac{\alpha^2}{4}\right)$$

2. Find the optimal value of α for which there is maximum reduction in variance. Is it computable in practice? Find a range of α over which there will be some improvement in variance.

   Solution. *The variance is minimized (differentiate and equal to zero) when*

$$\alpha = -2\mathbb{P}[X_i > a],$$

   and the minimal variance will be

$$\mathrm{Var}[\hat{I}_n^c] = \frac{1}{n}\left(\mathbb{P}[X_i > a] - 2\mathbb{P}[X_i > a]^2\right).$$

   We would not be able to compute this α *in practice, however,* $\alpha\mathbb{P}[X_i > \alpha] + \alpha^2/4$ *will be negative for* $\alpha \in [-4\mathbb{P}[X_i > \alpha], 0]$. *Choosing* α *in this range we will see a decrease in variance. Of course, if* $X > \alpha$ *is a rare event, this is easier said than done, since* $\mathbb{P}[X_i > \alpha]$ *will be very small.*

3. Suppose now that p is the standard Gaussian distribution, and $a = 3$. Implement both a standard MC estimator $\hat{I}_n$ and appropriately tuned control variate estimator $\hat{I}_n^c$. Plot the 95% intervals for both as a function of n.

   Solution. *See the Jupyter notebook. Performing the numerical experiment, one will see insignificant difference between the control variate estimator and the standard one. This isn't surprising, the control variate doesn't provide that much additional information about the distribution.*

4. Employing the fact that there there is an explicit formula for the moments of p, construct a control variate that produces a better variance reduction. Is it worth doing?

   Solution. *In the Jupyter notebook, we construct a polynomial control variate that produces a modest reduction of the variance. For a fixed computational cost, however, the estimator is worse than the naive Monte Carlo estimator.*

Problem 8 (Importance sampling with Gaussian mixture)**.** Consider the function

$$f(x, y) = e^{-\beta V_1(x,y)} + e^{-\beta V_2(x,y)}$$

where $\beta > 0$ and

$$V_1(x, y) = (x - 0.5)^2 + \frac{1}{2}(y + 0.1)^4, \quad V_2(x, y) = 0.75(x + 0.4)^2 + (y - 0.5)^4.$$

1. Use a Monte-Carlo algorithm to estimate the integral

$$Z = \int_{-1}^{1} \int_{-1}^{1} f(x, y)\, dx\, dy,$$

   for $\beta = 100$.

2. Plot the variance of the estimator as a function of the number of random samples that you are generating. Use a deterministic numerical method to estimate Z and calculate the error as a function of the number of random samples.

   Solution. *See the Jupyter notebook. Note that when we sample from a uniform distribution in* $[-1, 1] \times [-1, 1]$ *to compute this integral, we are indeed computing*

$$\int_{-1}^{1} \int_{-1}^{1} f(x, y) \frac{1}{4}\, dx\, dy$$

   where $p(x, y) = \frac{1}{4}$ *is the pdf of* $\mathcal{U}([-1, 1] \times [-1, 1])$*. So the estimator has to come multiplied by* 4 *to overcome this.*

3. Choose an appropriate distribution $\psi(x, y)$ and estimate Z by using importance sampling. Justify the choice of $\psi(x, y)$ and plot the variance and the error of the estimator as a function of the number of samples.

   Solution. *By inspecting the function* f *(either by looking at the polynomials* V_1 *and* V_2 *or by looking at the plots), we realise that* f *has two areas with more "importance". These will have mean at* $(x, y) = (0.5, -0.1)$ *and* $(-0.4, 0.5)$*. So a good first guess for importance distribution*

ψ would be a sum of two Gaussians, one of them with mean $\mu_1 = (0.5, -0.1)$ and the other $\mu_2 = (-0.4, 0.5)$. For the variance, we look at the coefficients in V_1 and V_2. So we choose

$$\psi(x, y) = 0.5\mathcal{N}(\mu_1, \Sigma_1) + 0.5\mathcal{N}(\mu_2, \Sigma_2),$$

where

$$\Sigma_1 = \begin{bmatrix} \frac{1}{\beta} & 0 \\ 0 & \frac{1}{0.5\beta} \end{bmatrix}, \qquad \Sigma_2 = \begin{bmatrix} \frac{1}{0.75\beta} & 0 \\ 0 & \frac{1}{\beta} \end{bmatrix}.$$

We can vary the coefficients in front of each Gaussian after further inspection of the plot of f (feel free to experiment with this!), but for the purposes of this exercise it is sufficient to use 0.5. To sample from ψ we need to sample from a generalised Bernoulli function that takes the values 1 and 2 with probabilities 0.5 and 0.5 (this will be different if you change the coefficients in front of each Gaussian!). For the solution, see the Jupyter notebook.

Simulation of continuous-time Gaussian processes

Problem 9 (Simulation of Markovian Gaussian processes)**.** A very useful property of multivariate Gaussian random variables is that if we condition on part of the random vector, the resulting distribution remains Gaussian. To see this, suppose that

$$\mathbf{X} = (X_1, X_2)^\top \sim \mathcal{N}(\mathbf{m}, \Sigma), \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Then we know that

$$\mathbf{X}_1 \sim \mathcal{N}(\mathbf{m}_1, \Sigma_{11}), \quad \text{and} \quad \mathbf{X}_2 \sim \mathcal{N}(\mathbf{m}_2, \Sigma_{22}).$$

Furthermore, the conditional distribution of $\mathbf{X}_2$ conditional on $\mathbf{X}_1$ is a multivariate normal with

$$\mathbb{E}[\mathbf{X}_2 \,|\, \mathbf{X}_1] = \mathbf{m}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_1 - \mathbf{m}_1). \tag{6}$$

and

$$\mathrm{Var}(\mathbf{X}_2 \,|\, \mathbf{X}_1) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}. \tag{7}$$

Using these properties we can develop a more efficient scheme to simulate Gaussian processes, more specifically to interpolate between already simulated points of a Gaussian process.

1. Suppose we wish to generate X_{n+1} at time t_{n+1} given that we have already generated $X_0, \ldots, X_n$.

   (a) Specify the conditional distribution of X_{n+1} given $X_0, \ldots, X_n$.

   Solution. *Using equations (6) and (7), we know that the conditional distribution of X_{n+1} given $X_0, \ldots, X_n$ is a Normal distribution with mean*

   $$\mathbb{E}[X_{n+1} \,|\, X_0, \ldots, X_n] = \mathbb{E}(X_{n+1}) + \Sigma_{n+1,(1,\ldots,n)}\Sigma^{-1}_{(1,\ldots,n),(1,\ldots,n)}([X_0, \ldots, X_n] - [\mathbb{E}(X_0), \ldots, \mathbb{E}(X_n)]),$$

   and variance

   $$\mathrm{Var}(X_{n+1} \,|\, X_0, \ldots, X_n) = \mathrm{Var}(X_{n+1}) - \Sigma_{n+1,(1,\ldots,n)}\Sigma^{-1}_{(1,\ldots,n),(1,\ldots,n)}\Sigma_{(1,\ldots,n),n+1}.$$

(b) Use this to construct a numerical scheme to simulate a Gaussian stochastic process.

Solution. *Given $X_0, \ldots, X_n$ we generate X_{n+1} by generating a sample from a normal distribution with mean and variance described above.*

2. Suppose additionally that the Gaussian process $X(t)$ is Markovian, so that in particular, you only need to know the value of $X(t_n)$ to generate $X(t_{n+1})$. Construct a scheme to iteratively sample $X(t_i)$ over a sequence of points $t_0 < t_1 < t_2 < \ldots$.

Solution. *If X_t is Markovian, we only need X_n to generate X_{n+1}. The formulas above simplify to*

$$\mathbb{E}[X_{n+1} \,|\, X_0, \ldots, X_n] = \mathbb{E}[X_{n+1} \,|\, X_n] = \mathbb{E}(X_{n+1}) + \frac{Cov(X_n, X_{n+1})}{\mathrm{Var}(X_n)}(X_n - \mathbb{E}(X_n)),$$

and

$$\mathrm{Var}(X_{n+1} \,|\, X_0, \ldots, X_n) = \mathrm{Var}(X_{n+1} \,|\, X_n) = \mathrm{Var}(X_{n+1}) - \frac{Cov(X_n, X_{n+1})^2}{\mathrm{Var}(X_n)}.$$

So to generate X_{n+1} we only need to sample from a Normal distribution with this mean and variance.

(a) In the case of Brownian motion, show that the update formula can be written as:

$$X(t_{i+1}) = X(t_i) + \left(\sqrt{t_{i+1} - t_i}\right) Z,$$

where $Z \sim \mathcal{N}(0, 1)$.

Solution. *The Brownian motion has $\mathbb{E}(X_t) = 0$ and $Cov(X_t, X_s) = \min(t, s)$ (and in particular $\mathrm{Var}(X_t) = t$). So we have*

$$\mathbb{E}[X_{t_n+1} \,|\, X_{t_n}] = \frac{\min(t_n, t_{n+1})}{t_n} X_n = X_n,$$

and

$$\mathrm{Var}(X_{t_{n+1}} \,|\, X_{t_n}) = t_{n+1} - \frac{\min(t_n, t_{n+1})^2}{t_n} = t_{n+1} - t_n.$$

So the conditional distribution of $X_{t_{n+1}}$ given X_{t_n} is $\mathcal{N}(X_n, t_{n+1} - t_n)$ and we can write

$$X_{t_{n+1}} = X_{t_n} + \sqrt{t_{n+1} - t_n} Z, \qquad Z \sim \mathcal{N}(0, 1)$$

as required.

(b) Derive a similar update formula for the stationary Ornstein-Uhlenbeck process with mean 0 and covariance $C(s, t) = \exp(\alpha|t - s|/2)$.

Solution. *For the Ornstein-Uhlenbeck process we have $\mathbb{E}(X_t) = 0$, $C(s, t) = \exp(\alpha|t - s|/2)$ and $\mathrm{Var}(X_t) = 1$. Plugging this in the previous expressions, you obtain the desired update.*

Problem 10 (Karhunen–Loève expansion)**.** Consider the Gaussian random field $X(x)$ in $\mathbb{R}$ with covariance function

$$\gamma(x, y) = e^{-a|x-y|}$$

where $a > 0$.

1. Simulate this field: generate samples and calculate the first four moments.

   Solution. *Note that we have already done this when we simulated Gaussian processes (the OU process!). See Problem8.m for solution.*

2. Consider $X(x)$ for $x \in [-L, L]$. Calculate analytically the eigenvalues and eigenfunctions of the integral operator $\mathcal{K}$ with kernel $\gamma(x, y)$,

$$\mathcal{K}f(x) = \int_{-L}^{L} \gamma(x, y) f(y)\, dy.$$

   Use this in order to obtain the Karhunen-Loéve expansion for X. Plot the first five eigenfunctions when $a = 1$, $L = 0.5$. Investigate (either analytically or by means of numerical experiments) the accuracy of the KL expansion as a function of the number of modes kept.

   Solution. *We need to compute eigenvalues λ and eigenfunctions Φ such that*

$$\int_{-L}^{L} e^{-a|x-y|}\Phi(y)\ dy = \lambda\Phi(x).$$

   We note that we can write this as

$$\int_{-L}^{x} e^{-a(x-y)}\Phi(y)\ dy + \int_{x}^{L} e^{a(x-y)}\Phi(y)\ dy = \lambda\Phi(x). \tag{8}$$

   We differentiate (8) *with respect to x to obtain*

$$-a\int_{-L}^{x} e^{-a(x-y)}\Phi(y)\ dy + a\int_{x}^{L} e^{a(x-y)}\Phi(y)\ dy = \lambda\Phi'(x), \tag{9}$$

   where $'$ represents derivative with respect to x. Differentiating again, we obtain

$$(-2a + a^2\lambda)\Phi(x) = \lambda\Phi''(x). \tag{10}$$

   Defining $\omega^2 = \frac{-2a+\lambda a^2}{\lambda}$, we obtain the ODE

$$\Phi''(x) - \omega^2\Phi(x) = 0, \qquad -L \le x \le L.$$

   We need two boundary conditions. We obtain them by evaluating equations (8) *and* (9) *at $x = -L$ and $x = L$ and rearranging:*

$$a\Phi(L) + \Phi'(L) = 0, \tag{11}$$
$$a\Phi(-L) - \Phi'(-L) = 0. \tag{12}$$

   It can be shown that equation (10) *with boundary conditions* (11)–(12) *only has solutions when $\omega^2 \ge 0$ and that these are of the form*

$$\Phi(x) = c_1\cos(\omega x) + c_2\sin(\omega x).$$

In order to compute the eigenvalues, we use the boundary conditions and obtain

$$c_1(a - \omega\tan(\omega L)) + c_2(\omega + a\tan(\omega L)) = 0,$$
$$c_1(a - \omega\tan(\omega L)) - c_2(\omega + a\tan(\omega L)) = 0.$$

This system has nontrivial solutions only if its determinant is zero, which gives the following:

$$(a - \omega\tan(\omega L)) = 0,$$
$$(\omega + a\tan(\omega L)) = 0.$$

Denoting the solutions to these equations by ω, ω^*, we obtain the eigenvalues and eigenfunctions:

$$\Phi_n(x) = \frac{\cos(\omega_n x)}{\sqrt{L + \frac{\sin(2\omega_n L)}{2\omega_n}}}, \qquad \lambda_n = \frac{2a}{\omega_n^2 + a^2}, \qquad n \text{ even}, \tag{13}$$

$$\Phi_n^*(x) = \frac{\sin(\omega_n^* x)}{\sqrt{L - \frac{\sin(2\omega_n^* L)}{2\omega_n^*}}}, \qquad \lambda_n^* = \frac{2a}{{\omega_n^*}^2 + a^2}, \qquad n \text{ odd}. \tag{14}$$

$$\tag{15}$$

And we can write

$$X(t,\omega) = \sum_{n=1}^{\infty} \left[\xi_n \sqrt{\lambda_n}\Phi_n(x) + \xi_n^* \sqrt{\lambda_n^*}\Phi_n^*(x)\right],$$

where ξ_n, ξ_n^* are computed by taking the inner product of X with the eigenfunctions.

3. Develop a numerical method for calculating the first few eigenvalues/eigenfunctions of $\mathcal{K}$ with $a = 1$, $L = -0.5$. Use the numerically calculated eigenvalues and eigenfunctions to simulate $X(x)$ using the KL expansion. Compare with the analytical results and comment on the accuracy of the calculation of the eigenvalues and eigenfunctions and on the computational cost.

   Solution. *Not presented here. Would need to write the eigenfunctions in a chosen basis and obtain generalised eigenvalue problem. Then solve it numerically and compare with the solutions obtained analytically.*

Additional problems

Problem 11 (Sampling uniformly on spheres and balls)**.** A rv has uniform distribution on the d–dimensional ball $D = \{x \in R^d : |x|^2 \leq 1\}$ if the rv takes values almost surely in D and has distribution

$$p(x)\,dx = \frac{dx}{\int_D 1\,dx}.$$

Similarly, a rv has uniform distribution on the sphere $C = \{x \in \mathbb{R}^d : |x|^2 = 1\}$ if the rv takes values almost surely in C and has distribution,

$$q(dx) = \frac{1}{\lambda(C)}\lambda(dx)$$

where $\lambda(dx)$ is the spherical measure on C.

1. Given $U \sim U(0,1)$, show that $(\cos(2\pi U), \sin(2\pi U))$ is uniformly distributed on the 2D circle.

   Solution. *Given an arc A_{θ_1,θ_2} of the circle defined by angles $\theta_1 < \theta_2$, the probability*

   $$\mathbb{P}[(\cos(2\pi U), \sin(2\pi U)) \in A_{\theta_1,\theta_2}] = \mathbb{P}[2\pi U \in [\theta_1, \theta_2]] = \frac{1}{2\pi}(\theta_2 - \theta_1).$$

   It follows that the density of this random variable is $1/2\pi$, which implies that it is uniform on C.

2. Suppose we want to sample uniformly from the 3D sphere. We use the spherical coordinate transformation $\psi \in [0, \pi]$, and $\theta \in [0, 2\pi]$,

   $$(x(\theta,\psi), y(\theta,\psi), z(\theta,\psi)) = (\sin(\psi)\cos(\theta), \sin(\psi)\sin(\theta), \cos(\psi)).$$

   (a) Let $\mathbf{X}$ be a uniformly distributed random variable on the sphere. Writing the spherical measure in spherical coordinates, write down the marginal densities of the random variables ψ and θ, and write down the CDF of ψ.

   Solution. *Let $\mathbf{X}$ be a uniformly distributed random variable on the sphere. Given a region $B = \{x(\theta,\psi), y(\theta,\psi), z(\theta,\psi)) : \theta_1 \le \theta \le \theta_2, \psi_1 \le \psi \le \psi_2\}$.*

   $$\mathbb{P}[\mathbf{X} \in B] = \frac{1}{4\pi}\int_{\theta_1}^{\theta_2}\int_{\psi_1}^{\psi_2} sin(\psi) d\psi\, d\theta.$$

   Let $\theta(\mathbf{X}), \psi(\mathbf{X})$ be the corresponding spherical coordinates of the rv $\mathbf{X}$, then

   $$\mathbb{P}[\theta \in [\theta_1, \theta_2]] = \frac{1}{2\pi}(\theta_2 - \theta_1),$$

   i.e. θ is uniformly distributed on $[0, 2\pi]$. Similarly,

   $$\mathbb{P}[\psi(\mathbf{X}) \in [\psi_1, \psi_2]] = \frac{1}{2}\int_{\psi_1}^{\psi_2} \sin(\psi)\, d\psi,$$

   so that ψ has density $\frac{1}{2}\sin(\psi)\, d\psi$. The CDF is therefore,

   $$\mathbb{P}[\psi(\mathbf{X}) \le k] = \frac{1}{2}(1 - \cos(k)), \text{ for } k \in [0, \pi],$$

   which has inverse $G(y) = \arccos(1 - 2y)$.

   (b) Based on the previous step, generate samples ψ and θ using an appropriate method, and construct a sampler generate samples from the 2–sphere.

   Solution. *Similar to the Box-Muller transform, we sample θ uniformly, and ψ using an inverse function transform, and then transform from spherical to Euclidean coordinates.*

   i. Sample $u_1 \sim U[0,1]$ and let $\theta = 2\pi u_1$.
   ii. Sample $u_2 \sim U[0,1]$ and let $\psi = \arccos(1 - 2u_2)$.
   iii. Output $\mathbf{X} = (\sin(\psi)\cos(\theta), \sin(\psi)\sin(\theta), \cos(\psi))$.

3. Given $(X, Y) \sim U(D)$, show that

$$\left(\frac{X}{\sqrt{X^2+Y^2}}, \frac{Y}{\sqrt{X^2+Y^2}}\right) \sim U(C),$$

where C is the 1-sphere in $\mathbb{R}^2$. Construct a rejection based sampler to generate samples $U(C)$ using proposals with distribution $U([-1,1] \times [-1,1])$. This algorithm can be readily generalised sample from spheres in arbitrary dimensions. How do you expect the average performance to depend on dimension?

Solution. *Given $\theta_1 \leq \theta_2$ let A_{θ_1,θ_2} be the arc on C between angles θ_1 and θ_2 and let S_{θ_1,θ_2} be the corresponding sector in D. Then*

$$\mathbb{P}\left[\frac{(X,Y)}{\sqrt{X^2+Y^2}} \in A_{\theta_1,\theta_2}\right] = \mathbb{P}[(X,Y) \in S_{\theta_1,\theta_2}] = \frac{1}{\pi}\int_0^1 \int_{\theta_1}^{\theta_2} r\, dr\, d\theta = \frac{\theta_2 - \theta_1}{2\pi},$$

as required. We can propose the following sampling scheme

(a) Generate $U_1, U_2 \sim U[-1,+1]$.

(b) If $(U_1, U_2) \in D$, then output $(U_1, U_2)/\sqrt{U_1^2 + U_2^2}$.

(c) Otherwise return to step 1.

4. Suppose we can generate samples $X, Y \sim \mathcal{N}(0,1)$ iid. Show that

$$\left(\frac{X}{\sqrt{X^2+Y^2}}, \frac{Y}{\sqrt{X^2+Y^2}}\right) \sim U(C).$$

Solution. *Let X and Y be $\mathcal{N}(0,1)$ distributed. For $\theta_1 \leq \theta_2$,*

$$\begin{aligned}\mathbb{P}\left[(X,Y)/\sqrt{X^2+Y^2} \in A_{\theta_1,\theta_2}\right] &= \frac{1}{2\pi}\int_{\cup_{t\geq 0} tA_{\theta_1,\theta_2}} e^{-(x^2+y^2)/2}\, dx\, dy \\ &= \frac{1}{2\pi}\int_0^{\infty}\int_{\theta_1}^{\theta_2} re^{-r^2/2}\, dr\, d\theta = \frac{1}{2\pi}(\theta_2 - \theta_1),\end{aligned}$$

as required.

Problem 12 (Sampling Gaussian random variables)**.** First we consider the problem of generating Gaussians using rejection sampling

1. The standard Cauchy distribution is a continuous probability distribution having pdf:

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Use the inverse transform method to derive an algorithm to sample from this distribution.

2. Using the Cauchy distribution as proposal, use the rejection algorithm to generate samples from the standard Gaussian distribution $e^{-x^2/2}/\sqrt{2\pi}$. Implement a function in a programming language of your choice to sample Gaussian random variables using this scheme.

Solution. *Let*

$$p(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \text{ and } g(x) = \frac{1}{\pi(1+x^2)}.$$

It is straightforward to check that $\sup_x p(x)/g(x)$ *is attained at* $x = \pm 1$*, at which*

$$M = \sup_x [p(x)/g(x)] = e^{-1/2}\sqrt{\pi/2}.$$

We then implement a standard rejection sampler using this proposal.

3. Would it be possible to work the other way round, i.e., use rejection sampling to produce Cauchy distributed draws from using a Gaussian proposal distribution?

   Solution. *The reason why this is not possible is that there is no finite upper bound* M *in this case, indeed, in this case*

$$M = \sup_{x \in \mathbb{R}} \frac{e^{x^2/2}\sqrt{2\pi}}{\pi(1+x^2)} = \infty.$$

4. Write code to implement the Box-Muller sampling algorithm described in class. Frequently the BM-algorithm is cited as being slow due to the necessity to compute cosines and sines. Use the rejection based method described in the previous problem to obtain samples which have the same distribution as $(\cos(2\pi U_2), \sin(2\pi U_2))$. Implement code to sample Gaussian random variables using this scheme.

5. Using timing functions provided in your language (or use the shell `time` command), calculate the time of execution for all three methods, after generating 10^6 samples. Which is the fastest? Is this what you expected?

6. Use the Kolmogorov–Smirnov test as a check that the generated random numbers follow the standard normal distribution.

7. Suppose we wish to sample a pair of Gaussian random variables X_1 and X_2 having means μ_i, variances σ_i^2 and correlation ρ. By assuming that the Cholesky decomposition of the covariance matrix is of the form

$$C = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix},$$

   find expressions for a_{11}, a_{21} and a_{22}, and solve them to generate samples from X_1, X_2.

   Solution.

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = CC^\top = \begin{pmatrix} a_{11}^2 & a_{11}a_{21} \\ a_{11}a_{21} & a_{21}^2 + a_{22}^2 \end{pmatrix},$$

   so that $a_{11} = \sigma_1$*,* $a_{21} = \rho\sigma_2$*, and so* $a_{22} = \sigma_2\sqrt{(1-\rho^2)}$*. Therefore to sample from* $\mathcal{N}(\mu, \Sigma)$*, we generate* $Z_1, Z_2 \sim \mathcal{N}(0,1)$ *iid, and generate*

$$\begin{aligned} X_1 &= \mu_1 + \sigma_1 Z_1 \\ X_2 &= \mu_2 + \rho\sigma_2 Z_1 + \sigma_2\sqrt{1-\rho^2} Z_2. \end{aligned}$$