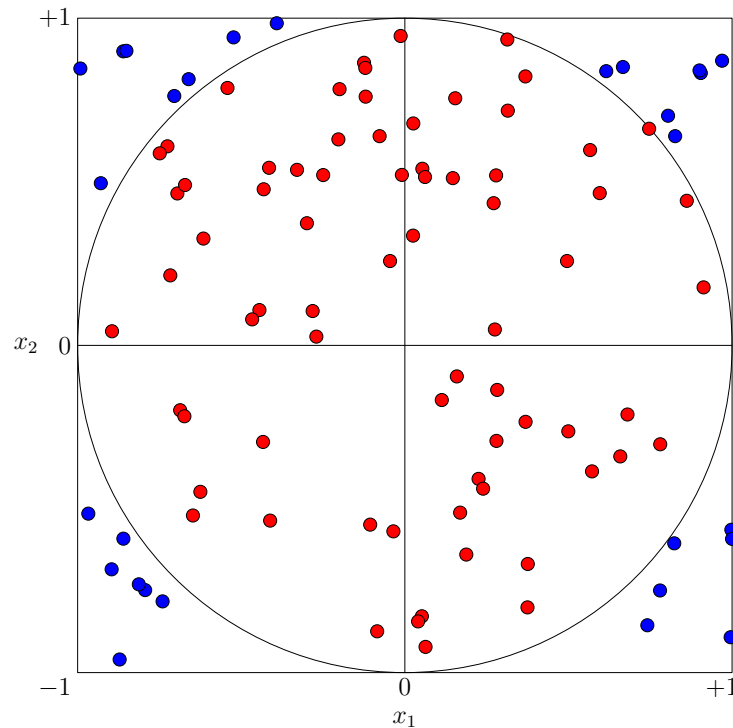


EE365 Homework 2

1. *Monte Carlo integration.* Consider a unit circle inscribed in a square, as shown below.



Each of the small circles drawn on this figure represents a random point that was generated in the square; the red and blue circles represent points inside and outside the unit circle, respectively. If we choose a point uniformly at random within the square, then the probability that the point lies in the unit circle is

$$p = \frac{\text{area of the unit circle}}{\text{area of the circumscribed square}}.$$

The Monte Carlo approximation of this probability is

$$\hat{p} = \frac{\# \text{ of sample points inside the unit circle}}{\# \text{ of sample points}}.$$

We know that the area of the circumscribed square is 4; if we knew p , then we could compute the area of the unit circle:

$$\text{area of the unit circle} = p(\text{area of the circumscribed square}) = 4p.$$

Replacing p with \hat{p} gives us a Monte Carlo estimate of the area of the unit circle. This method can allow us to obtain good estimates for the volumes of complicated or high-dimensional objects.

- (a) The n -dimensional unit ball is

$$B_n = \{x \in \mathbf{R}^n : \|x\|_2 \leq 1\},$$

where $\|x\|_2$ is the usual Euclidean distance. Use the Monte Carlo method to approximate the volume of B_n for $n = 1, \dots, 15$. For each n , use 1,000,000 samples to construct your Monte Carlo estimate. Plot the volume of B_n versus n .

- (b) Define the ε -skin of B_n as

$$S_n^\varepsilon = \{x \in \mathbf{R}^n : 1 - \varepsilon \leq \|x\|_2 \leq 1\},$$

The fraction of B_n occupied by S_n^ε is

$$F_n^\varepsilon = \frac{\text{vol}(S_n^\varepsilon)}{\text{vol}(B_n)}.$$

Use the Monte Carlo method to approximate $\text{vol}(S_n^{0.1})$ and $F_n^{0.1}$ for $n = 1, \dots, 15$. Plot $F_n^{0.1}$ versus n . Do you find this plot surprising? The phenomenon you observe is called sphere hardening, and has some important implications in coding theory.

2. *A quick Gaussian approximation.* We consider the Monte Carlo example from the Markov chain lecture. The random variables b_1, \dots, b_{50} are IID with

$$\mathbf{Prob}(b_i = 1) = 0.4, \quad \mathbf{Prob}(b_i = 0) = 0.6.$$

Our goal is to find or estimate

$$p = \mathbf{Prob} \left(\sum_{i=1}^{25} b_i \geq 0.6 \sum_{i=1}^{50} b_i \right).$$

The Monte Carlo simulation in the lecture notes suggests that p is around 0.14 or 0.15. In this exercise, you will compute an estimate of p using a simple Gaussian approximation. Carry out the following steps.

- (a) Explain why $p = \mathbf{Prob}(z \leq 0)$, where

$$z = - \sum_{i=1}^{25} b_i + (3/2) \sum_{i=26}^{50} b_i.$$

- (b) Find the mean μ and variance σ^2 of z , and form the approximation

$$p \approx \hat{p} = \mathbf{Prob}(\hat{z} \leq 0),$$

where $\hat{z} \sim \mathcal{N}(\mu, \sigma^2)$. Give the numerical value of \hat{p} , and check that it is consistent with the Monte Carlo estimate of p given in the lecture.

Hint. In matlab, $(1+\text{erf}(-\mu/(\text{sqrt}(2)*\text{sigma}))) / 2$ gives the probability that an $\mathcal{N}(\mu, \sigma^2)$ random variable is less than zero.

Remark. The Monte Carlo estimate \hat{p}_{mc} of p converges to p as N increases, where N is the number of samples used in the Monte Carlo estimate. The Gaussian approximation described in this exercise is a simple approximation that cannot be improved; its error depends on how non-Gaussian the random variable z is.

3. *Epidemic model simulation.* We consider an epidemic model. A simpler version of this model is simulated in the lecture notes, available on the class website. These lecture notes will not be covered in class, but are there for you to read.

In this model, each individual can be in one of four states: *susceptible* (S), *infected* (I), *deceased* (D), or *protected* (or immune) (P). An individual can transition from S to I (they become infected), from I to D (they die), from I to S (they recover, but without immunity), or from I to P (they recover, and now have immunity). The states D and P are absorbing: once an individual is in either of these states, they never leave. The transition probabilities for each individual are, with infection states ordered (S, I, D, P) ,

$$\begin{aligned} & \begin{bmatrix} 0.6 & 0.4 & 0 & 0 \\ 0.2 & 0.6 & 0.1 & 0.1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} && \text{with an infected neighbor,} \\ & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.2 & 0.6 & 0.1 & 0.1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} && \text{with no infected neighbors.} \end{aligned}$$

A typical individual has four neighbors: one above, one below, one to the left and one to the right; however, individuals on the boundary may be missing one or more of these neighbors. Use Monte Carlo simulation to estimate the expected number of individuals who are dead at $T = 50$, for a population of 100 individuals on a 10×10 grid, starting from each of the following initial distributions:

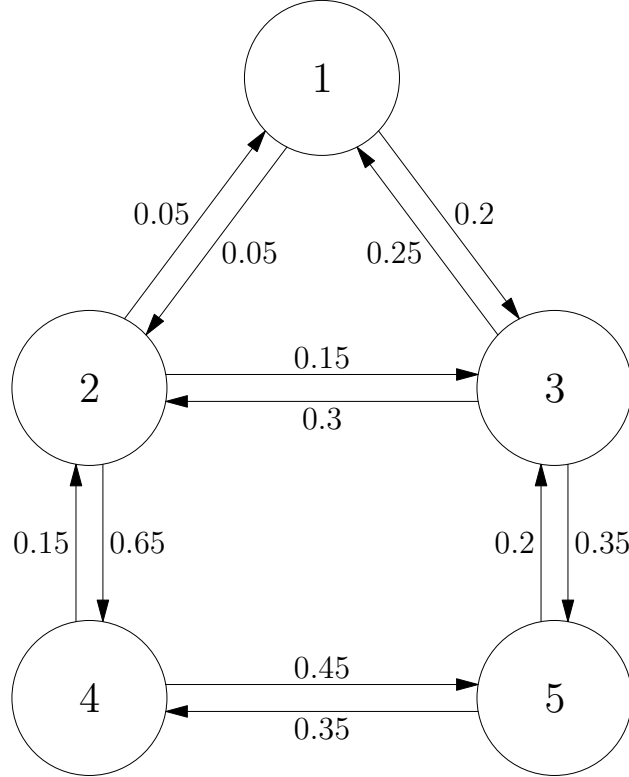
- Two individuals are infected, one at the (2, 2) location, and the other at the (9, 9) location. The rest of the population is susceptible.
- Two individuals are infected as above, but there are also eight protected individuals, at the following locations:

$$(2, 3), (3, 2), (3, 7), (4, 6), (6, 4), (7, 3), (8, 9), (9, 8).$$

(These individuals could be immunized, for example.) All others are susceptible.

An example script showing how to produce plots similar to those in the lecture notes is provided on the course website. You do not need to submit any plots, but you may find such plots interesting or useful for debugging your code.

4. *The Markov property.* Let $x_t : \Omega \rightarrow \mathcal{X}$ be a Markov chain, $i \in \mathcal{X}$ and $A, B \subseteq \mathcal{X}$. Determine whether each of the following statements is true or false. If a statement is true, give a proof; if a statement is false, give a counterexample.
- (a) $\mathbf{Prob}(x_2 \in B \mid x_1 \in A, x_0 = i) = \mathbf{Prob}(x_2 \in B \mid x_1 \in A)$
- (b) $\mathbf{Prob}(x_2 \in B \mid x_1 = i, x_0 \in A) = \mathbf{Prob}(x_2 \in B \mid x_1 = i)$
5. *Distribution propagation.* Consider a Markov chain specified by the following probability transition graph.



(We omit the self loops; the probabilities associated with these transitions can be inferred from the probabilities of the other transitions.) The initial distribution of the chain is

$$\pi_0 = [0.3 \quad 0.1 \quad 0.2 \quad 0.3 \quad 0.1].$$

We want to compute the quantity

$$p_T = \frac{1}{T+1} \sum_{t=0}^T \mathbf{Prob}(x_t = 1)$$

for $T = 50$.

- (a) Use distribution propagation to compute p_{50} exactly.
- (b) Use the Monte Carlo method to approximate p_{50} . Generate a plot of your estimate of p_{50} versus the number of sample trajectories used in the Monte Carlo method. Report the values of your approximation with 10, 100 and 1000 samples.

6. *Markov-chain representation of a moving-average process.* Let k be a positive integer, and let $w_{-k}, w_{-k+1}, w_{-k+2}, \dots$ be independent random variables. The k -step moving average of (w_t) is the sequence (x_t) , where

$$x_t = \frac{1}{k} \sum_{\tau=t-k}^{t-1} w_{\tau}.$$

- (a) Is (x_t) a Markov process?
- (b) Consider the sequence (z_t) , where

$$z_t = (x_t, w_{t-1}, \dots, w_{t-k+1}) \in \mathbb{R}^k.$$

- i. Show that $z_0, w_0, w_1, w_2, \dots$ are independent.
- ii. Find a function $f : \mathbb{R}^k \times \mathbb{R} \rightarrow \mathbb{R}^k$ such that $z_{t+1} = f(z_t, w_t)$.

We showed in lecture that these two facts imply that (z_t) is a Markov chain.