239423
Eirik Sakariassen

Mandatory 2

Problem 1

a) As stated in the description, T is set to be the first time when two people in the room share the same birthday. People enters the room according to a Poisson process $(N_t)_{t>0}$ with rate $\lambda = 1$. We then have that the interarrival times X_1, X_2.. are independent and

$$X_i \sim exp(\lambda = 1), \quad \text{for i} = 1, 2, 3..$$

We obtain the total time spent after n arrivals by

$$T_n = \sum_{i=1}^{n} X_i$$

and hence we can obtain the time after K arrivals by

$$T = \sum_{i=1}^{K} X_i$$

The HPP is defined as a counting process, denoted by $\{N(t), t >= 0\}$, which represents the total number of events that have happened up to and including the time t. The expection of $N(t)_{t>=0}$ is

$$E[N(t)] = \lambda t$$

where $\lambda = 1$ in our case. We want to estimate the Poisson process at count K, which gives

$$E[K] = t$$

where t is the time when K happened $\Rightarrow$ t=T, giving

$$E[K] = T = \sum_{i=1}^{K} X_i$$

where $X_i \sim exp(\lambda = 1), \quad \text{for i} = 1, .., K$. Further we have that E(X) = β for an exponential distribution, where $\beta = 1/\lambda = 1$. This means, expected arrival time is

$$E[T] = E(\sum_{i=1}^{K} X_i) = \sum_{i=1}^{K} E(X) = \sum_{i=1}^{K} 1 = K$$

b) We want to estimate E(T) = θ followed by these properties:

$$\theta = \int_A g(t)dt$$

$$\Rightarrow \theta = \int_A h(t)f(t)dt = E(h(T))$$

where g(t) is decomposed into h(t)f(t) and f(t) is a probability density on the set A. T is then a rv with density f(t). We can use the crude Monte Carlo estimate

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} h(T_i)$$

In our case

$$\theta = \int_0^\infty (1 + \frac{t}{365})^{365} e^{-t} dt$$

We can set e^{-t} as f(t) since it is the pdf for $exp(\lambda = 1)$

$$\Rightarrow \hat{\theta} = E((1 + \frac{T}{365})^{365})$$

where T $\sim exp(\lambda = 1)$. I.e

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(1 + \frac{T_i}{365})^{365}$$

$T_1, .., T_n$ iid $\sim exp(\lambda = 1)$

c) For a non-homogeneous Poisson process, the intensity λ varies over time t. In this case,

$$\lambda(t) = t$$

which is a linear function where $\lambda(t) <= \lambda_{max}$ for all t in [0,t]. If we want to simulate NHPP by the thinning method over the interval [0,t], we can use this algorithm:

1. Simulate an HPP with intensity λ_{max}, which gives simulated arrival times $s_1, .., s_N$
2. Simulate N random variables $u_i, .., u_N$ from U[0,1] distribution.
3. For each arrival time s_i, accept it if $u_i <= \lambda(s_i)/\lambda_{max}$. If not, delete it.

The remaining arrival times may then be returned. Based on this, we can see that the algorithm will delete approximately half of the values. In cases where λ_{max} is high compared to most of the $\lambda(t)$ values, most arrival times will be rejected.

Since the integrated intensity in this case is easily inverted, and the thinning method rejects a lot of values, the transformation method will be a more efficient alternative. We have that

$$\wedge(t) = \int_0^t \lambda(u) du = \int_0^t u du = \frac{t^2}{2}$$

We can then easily find the inverted intensity of $\wedge(t)$ by

$$\wedge(t) = \frac{t^2}{2}$$

$$\Rightarrow y = \frac{t^2}{2}$$
$$\Rightarrow t = \sqrt{2y}$$

since $t > 0$. $\wedge^{-1}(t) = \sqrt{2t}$.

A simulation algorithm for the continuous-time birthday problem in the NHPP case based on the transformation method may look something like this:

1. Simulate the pdf for the birthday problem over the interval [0,366], giving the probabilities $p_1, .., p_{366}$
2. Then simulate an HPP with intensity 1, giving the arrival times $a_1, .., a_{366}$
3. Calculate $s_1 = \wedge^{-1}(a_1), .., s_{366} = \wedge^{-}(a_{366})$. This is the NHPP arrival times.
4. Simulate 366 random variables $u_i, .., u_{366}$ from the U[0,1] distribution
5. For each u_i, check if $u_i > p_i$. For the first occurrence, return i as the number of people in the room (K value) and s_i as the time (T value)
6. Run the simulation a large number of times and calculate the mean value of both K and T.

We want to estimate

$$\theta = \int_a^b g(x)dx = (b-a)E(g(X)),$$

where g(x) = $\wedge(t) = \int_0^t \lambda(s)ds$. We may use the basic (crude) Monte Carlo method in order to estimate it:

$$\hat{\theta} = \frac{b-a}{n}\sum_{i=1}^{n} g(U_i) = \frac{t}{n}\sum_{i=1}^{n} U_i \tag{1}$$

$U_i \sim U[a,b]$. From the lecture notes on antithetic variables, we know that if the negative correlation between U_i and V_i (a pair of antithetic variables) extends the correlation between $g(U_i)$ and $g(V_i)$, we could reduce the variance in equation 1 if antithetic variables replaces the iid uniform rvs.

The lecture notes further explains that if g(x) is monotonic on [a,b], then $cov(g(U_i), g(V_i)) < 0$ when $cov(U_i, V_i) < 0$. As $\wedge(t)$ is monotonic on [0,t], we may use the method of antithetic variables to estimate θ by:

$$\hat{\theta}_{AT} = \frac{b-a}{n}(\sum_{i=1}^{n/2} g(U_i) + \sum_{i=1}^{n/2} g(V_i)),$$

where $U_1, .., U_{n/2}$ are iid from U[a,b] and $V_i = a + b - U_i$ for $i = 1, 2, .., n/2$.

Problem 2

a) The processes described in 2a) can be uniquely described by their transition matrix P, which is

$$P_{x,y} = P(X_{n+1} = y | X_n = x), \quad x, y \in S$$

Each entry in this squared matrix represents the probability of a transition from x to y. Generally, if a transition matrix T has all positive entries for some power of T, it is said to be **regular**. The markov Chain represented by T is a **regular Markov chain**.

For the random walk model on the complete graph, each next step has an 1/11 probability (including "going back" to itself). The transition matrix P is then a 11x11 matrix where all entries are 1/11, hence $P^1 > 0$. Based on this, we may also derive that

$$\pi_{complete} = \frac{1}{11}$$

which means that each vertex has equal probability also after a long time.

The transition probabilities for the random walk on a discrete circle is:

$$P(Y_{n+1} = i | Y_n = j) = \begin{cases} \frac{1}{2} & if \quad i \equiv j \pm 1 \quad mod \quad 11 \\ 0 & otherwise \end{cases}$$

for all $i, j \in \{1, .., 11\}$. The transition matrix P follows a pattern where $P_{i,j} = P_{i+1(mod11), j+1(mod11)}$, having two positive values per row. We have that $P_{1,2} = 1/2$ and $P_{1,11} = 1/2$, the rest of the values at row 1 equals to zero. Then for row 2, $P_{2,3} = 1/2$ and $P_{2,1} = 1/2$ etc. For this transition matrix P, we have that $P^{11} > 0$, e.g. it is a **regular Markov chain**.

For a Markov chain $(X_n)_{n>=0}$ with state space S and edges E that jumps to a neighboring vertex with probability proportional to the number of its neighbors, we have

$$P(X_{n+1} = j | X_n = i) = \begin{cases} \frac{1}{deg(i)} & if \quad (i,j) \in E \\ 0 & otherwise \end{cases}$$

This is true for (Y_n) (having to go left or right with 50/50 chance), therefore we can calculate π_j based on this formula from the lecture notes:

$$\pi_j = \frac{deg(j)}{\sum_{i>=1} deg(i)} = \frac{deg(j)}{2|E|}, \forall j \in S$$

which for the random walk (Y_n) on the discrete circle gives:

$$\pi_{circle} = \frac{2}{2 \cdot 11} = \frac{1}{11}$$

d)

As the probability of choosing head increases, the distribution approaches the case in c). This makes a lot of sense since we sample the same way as in c) whenever head is tossed. (Proven in the R file). As p approaches 0, visits becomes more random and the probabilities for all websites will approach 1/(number of websites).

Problem 3

a) In order to estimate the profit for strategy A, we first have to find the probability that you sell within the time frame [0,1]. We assume that bids arrive at times of a Poisson process with rate $\lambda = 1$. These bids are modelled as i.i.d. random variables from the unif[0,1] distribution. The bigger the price θ is, the smaller is the chance of accepting a bid from $U \sim U[0,1]$, and hence we may model the acceptance of bids as a Poisson process with a rate parameter $1-\theta$. The probability of of selling within [0,1] is then

$$P(S(1) > 1) = 1 - P(S(1) = 0) = 1 - \frac{((1-\theta)t)^0}{0!}\theta)e^{-(1-\theta)t} = 1 - e^{(\theta-1)}$$

Further, we know that the expectation for a random variable from a uniform continuous distribution is

$$E(X) = \frac{1}{2}(a+b)$$

having a as the lower bound and b as the upper bound. In our case, the profit must be between the price itself (θ) and 1, which gives expected profit $E(X) = \frac{1}{2}(1+\theta)$. We then have to consider this expected profit based on the probability of selling within the time window, and end up with the final expected value

$$E(profit) = \frac{1}{2}(a+b)\left(1 - e^{(\theta-1)}\right)$$

d) We have that

$$E(t) = 1 - e^{-1/2}\int_0^1 e^{\frac{t^2}{2}}dt,$$

and we want to estimate the integral based on the hit and miss method. This method says that if we consider $\theta = \int_a^b g(x)dx$ and assume that $0 <= g(x) <= c$ for any $x \in [a,b]$,

$$\theta = c(b-a)E(I(Y <= g(X))) \tag{2}$$

A proposed algorithm for 2 is:

1. Find a c satisfying $0 <= g(x) <= c$
2. Generate Y $\sim$ U[0,c]

3. Generate X $\sim$ U[a,b]
4. Find $Z = Y <= g(X)$
5. Estimate the integral by multiplying the area of the box (c(b-a)) with the proportion of (X,Y) pairs below g (Take the mean of Z)
6. Set the estimated integral into the formula and return the answer

In order to calculate the number of simulations needed to have approximately 95% probability that the estimate is at most ε=0.02 from the true value, we can use

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \varepsilon \quad \Rightarrow n > \frac{z_{\alpha/2}^2\sigma^2}{\varepsilon^2} \tag{3}$$

where $\sigma^2 = SD(\overline{X})^2 n$ (from the lecture notes on uncertainty in estimates).
For the hit or miss estimator in 2, we have that

$$\widehat{SD}(\hat{\theta}_{HM}) = c(b-a)\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \tag{4}$$

where $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} I(Y_i <= g(X_i))$. By combining (3) and (4), we obtain

$$\begin{aligned} n &> \frac{z_{\alpha/2}^2\widehat{SD}(\hat{\theta}_{HM})^2 n}{\varepsilon^2} \\ \Rightarrow n &> \frac{z_{\alpha/2}^2}{\varepsilon^2}(\frac{c(b-a)\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}) \\ \iff n &> \left(\frac{z_{\alpha/2}c(b-a)}{\varepsilon}\right)^2\hat{p}(1-\hat{p}) \\ \Rightarrow n &> \left(\frac{1.96e^{1/2}(1-0)}{0.02}\right)^2 0.71(1-0.71) \approx 5375 \end{aligned} \tag{5}$$

where $\hat{p}$ was estimated to 0.71 in 3e). Based on this, we have to run our simulation 5375 times.
The general Monte Carlo approach for approximation of integrals is described in 1b). This can be written alternatively as

$$\theta = \int_A g(x)dx = \int_A \frac{g(x}{f(x)}f(x)dx = E\left(\frac{g(X)}{f(X)}\right) \tag{6}$$

We want to choose the density f(x) smart in order to minimize the variance of the estimate.

$$\hat{\theta}_{IS} = \frac{1}{n}\sum_{i=1}^{n}\frac{g(X_i)}{f(X_i)} \tag{7}$$

For our integral $\int_0^1 e^{\frac{t^2}{2}}$dt, I propose a function f(x) from the triangular distribution with a=0, b=1 and c=1, giving

$$f(x) = \frac{2(x-0)}{(1-0)(1-0)} = 2x \tag{8}$$

We then finally have that

$$\hat{\theta}_{IS} = \frac{1}{n}\sum_{i=1}^{n} \frac{e^{\frac{Xi^2}{2}}}{2X_i} \tag{9}$$

The simulation algorithm is then as follows:

1. Generate $X_1, ., X_n$ from the triangular distribution with a=0, b=1, c=1 (2x).
2. Return the mean of $\frac{e^{\frac{Xi^2}{2}}}{2X_i}$