Kun Zhou UID: 204688165

P1.

i) I think step 3 is more efficient than step 2, because with $\sigma_0 = 1$, the $p(x, y)$ is very small when $\pi(x, y)$ is large. So It is hard for step 2 to get sample with more useful information.

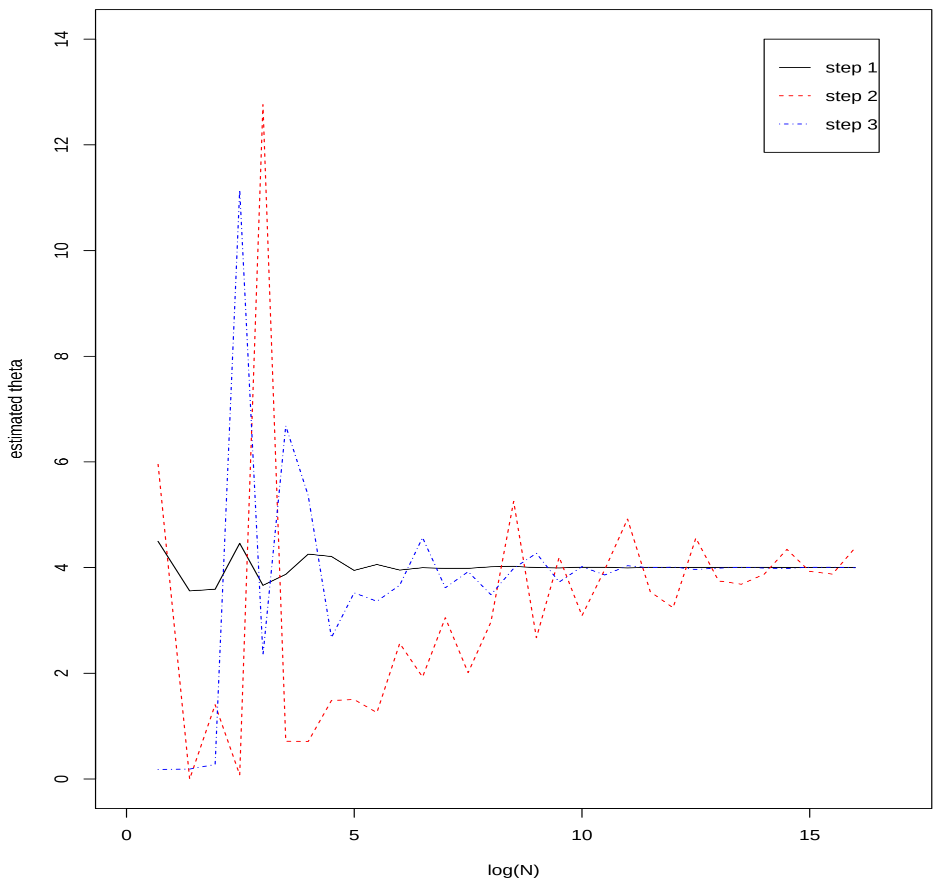


Figure 1: $\hat{\theta}$ against n

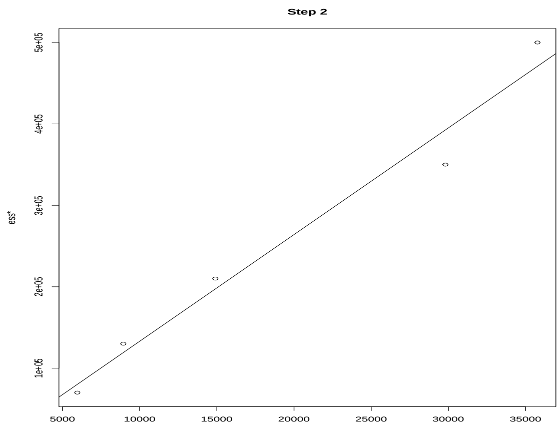
From Figure 1 , we can see $\hat{\theta}_1$ of step 1 quickly converges to 4 after sampling amount increase to 10^5. For step 2, $\hat{\theta}_2$ are prone to converge to 4, but the estiamted values are not stable even when $N = 10^{15}$. For step 3, $\hat{\theta}_3$ converges to 4 after the amount increases to 10^{10}. So step 1 is the best method and step 3 is better than step 2.

ii)

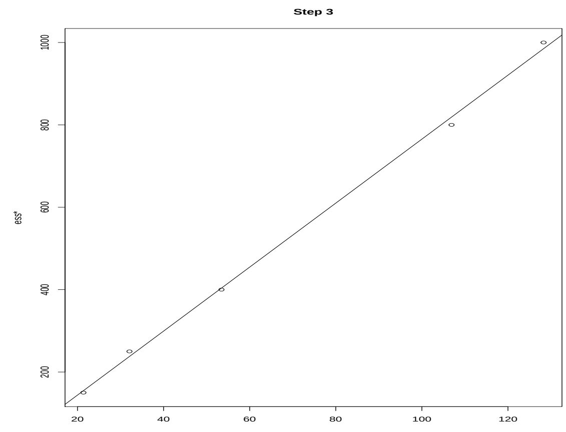
First calculate the $ess(n)$ for step 2 and step 3.

$$\omega = \frac{\pi(x,y)}{p(x,y)}$$
$$E[\omega] = E[\frac{\pi(x,y)}{p(x,y)}] = \int \frac{\pi(x,y)}{p(x,y)}p(x,y)dxdy = 1$$
$$E[\omega^2] = E[\frac{\pi(x,y)^2}{p(x,y)^2}] = \int \frac{\pi(x,y)^2}{p(x,y)}dxdy = \int \frac{\sigma_0^2}{2\pi}e^{-\left[(x-2)^2+(y-2)^2-\frac{1}{2\sigma_0^2}(x^2+y^2)\right]}dxdy$$
$$= \int \frac{\sigma_0^2}{2\pi}e^{-\left[(1-\frac{1}{2\sigma_0^2})x^2-4x+(1-\frac{1}{2\sigma_0^2})y^2-4y+8\right]}dxdy$$
$$= \int \frac{\sigma_0^2}{2\pi}e^{-(1-\frac{1}{2\sigma_0^2})\left[x-\frac{4\sigma_0^2}{2\sigma_0^2-1}\right]^2+\frac{8\sigma_0^2}{2\sigma_0^2-1}}e^{-(1-\frac{1}{2\sigma_0^2})\left[y-\frac{4\sigma_0^2}{2\sigma_0^2-1}\right]^2+\frac{8\sigma_0^2}{2\sigma_0^2-1}}e^{-8}dxdy$$
$$= \frac{\sigma_0^2}{2\pi}(\sqrt{\frac{\pi}{\frac{2\sigma_0^2-1}{2\sigma_0^2}}})^2e^{\frac{16\sigma_0^2}{2\sigma_0^2-1}}e^{-8}$$
$$= \frac{\sigma_0^4}{2\sigma_0^2-1}e^{\frac{8}{2\sigma_0^2-1}}$$
$$\Rightarrow ess(n) = \frac{n}{1+Var[\omega]} = \frac{n}{E[\omega^2]} = \frac{2\sigma_0^2-1}{\sigma_0^4}e^{\frac{-8}{2\sigma_0^2-1}}n$$

For step 2, $\sigma_0 = 1 \Rightarrow ess(n) = e^{-8}n = 0.000335n$. For step 3, $\sigma_0 = 4 \Rightarrow ess(n) = 0.09355n$.
For ess^*, calculating one $\hat{\theta}$ is not enough to estimate errors, so calculate $\hat{\theta}$ at least 500 times for each step and then calculate the standard deviation of each estimate. If 2 steps have a similar standard deviation, the estimated errors could be regarded as same. For example, if $n_1 = 5$, the standard deviation of step 1 is 0.6282757. For step 2, $n_2 = 210000$, the standard deviation of step 2 is 0.5757558. So we can think $ess^*(n_1) = ess^*(n_2)$ and theoretical value $n_2 = 5 * e^8 = 14904$.



(a) step 2



(b) step 3

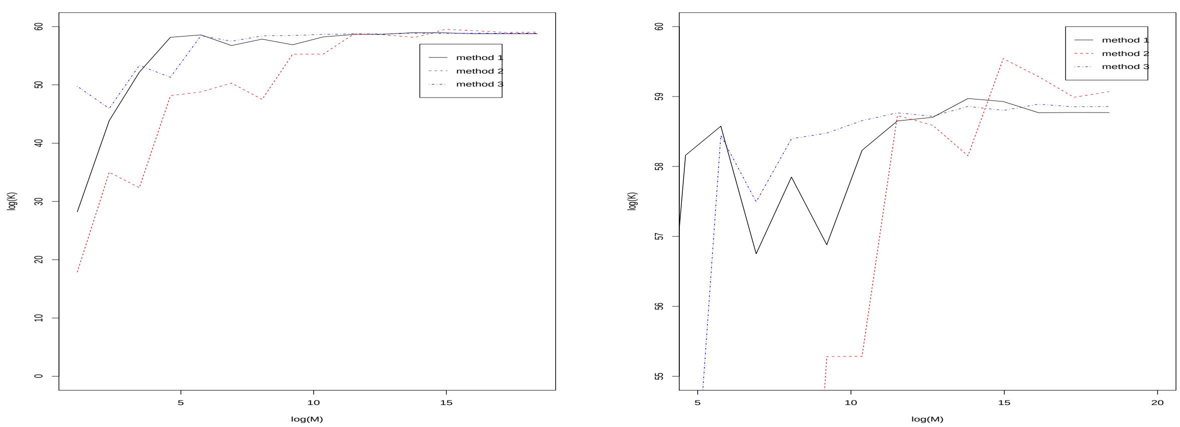
Figure 2

We can see ess and ess^* are proportional for both step 2 and step 3. From Figure 2a and Figure 2b, if we want step 2 and step 3 have the same effecive sample size, step 2 requires much more

samples.
The following is R code.

P2.

Method 1: I use probability $g_1(x) = \prod_{j=1}^{m} \frac{1}{k_j}$, where m is the total length of the path, and k_j is the number of possible choices at j-th step.
Method 2: I introduce an early termination rate $\epsilon = 0.05$ at each step (while in textbook, $\epsilon = 0.1$). So $g_2(x) = \prod_{j=1}^{m} \frac{1}{k_j * 0.95}$.
Method 3: For any walk that longer than 50, $u = 5$ more children based on it are generated and are reweighed by $\omega_0 = \frac{\omega}{u}$.
i)



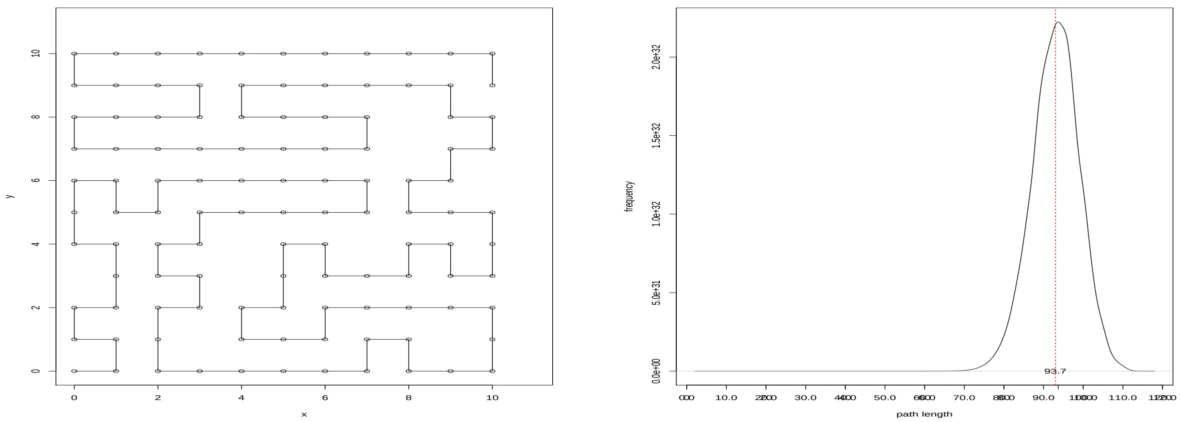
(a) Global log-log plot of K against M

(b) Local log-log plot of K against M

Figure 3

From Figure 3a , we can see that SIS processes of all 3 methods are converging with M increasing. Figure 3b demonstrates that Method 3 converges the fastest and Method 2 converges the most slowly. The limit values of Method 2 and Method 3 are higher than Method 1, which means the bias cannot be neglected. The following are the estimated value of Ks when $M = 10^8$ for the 3 methods respectively: $3.316744 * 10^{25}$, $4.512596 * 10^{25}$, $3.649155 * 10^{25}$.

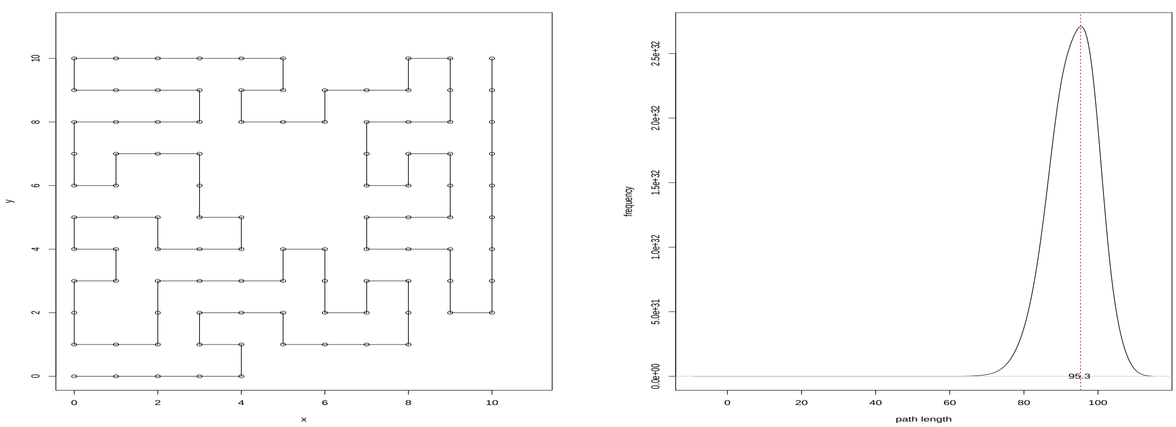
ii)
I applied the method mentioned in textbook.
$1.501552 * 10^{24}$
iii)



(a) Longest path (b) Density function of paths

Figure 4: Method 1

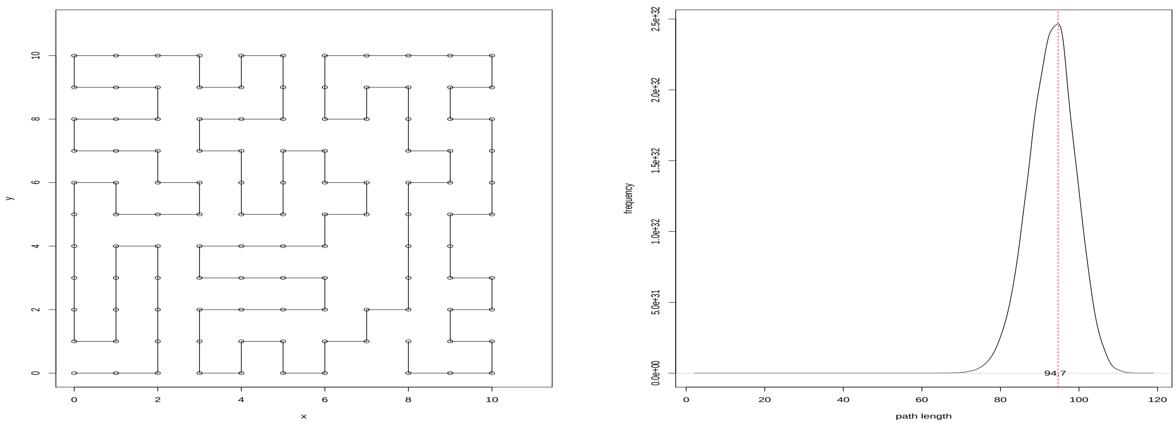
For Method 1, Figure 4a has length 111.



(a) Longest path (b) Density function of paths

Figure 5: Method 2

For Method 2, Figure 5a has length 100. Figure 5b is against intuition. The most dense point is 95.3 while the length of the longest path for Method 2 is 100. But it's reasonable, because $\frac{1}{\epsilon}^m$ makes longer paths' weight much more larger.



(a) Longest path

(b) Density function of paths

Figure 6: Method 3

For Method 3, Figure 6a has length 115. The most dense point is 94.7 which is a little higher than the one of Method 1, becuase there are more paths, having length more than 50 and some of which are adapted to new weights.
The following are C++ codes for all questions.