1. Assume you are using a biased coin to construct a Skip List. The probability of getting a head is 1/3 and probability of getting a tail is 2/3. When deciding the number of levels at which you will have a link for an element to be inserted you keep tossing your biased coin until you get a tail and the number of levels with links is equal to the number of heads you obtain before getting a tail.

(a) Estimate the probability that an element has links on levels 0 − i (and possibly also on higher levels).

(b) Assume such a Skip List has n elements; estimate the expected number of elements with links on level i.

(c) Estimate the probability to have at least one element on level $k \log_3 n$ where $k$ is an arbitrary integer $k \geq 1$.

(d) Estimate the expected value E of k such that k is the least integer so that the number of levels does not exceed $k \log_3 n$.

(e) Estimate the probability that an element with links up to level i also has a link at level $i + 1$.

(f) Estimate the expected number of elements between any two consecutive elements with a link on level $i + 1$ which have links on all levels from 0 to i but not on level $i + 1$.

(g) Estimate the search time in such a SkipList.

(h) So what do you think, is searching in such a SkipList built with such a biased coin asymptotically faster or slower than searching in the usual SkipList built with an unbiased coin?

(Each item is worth 3 points so in total the whole problem is worth 24 points.)
The solution follows what we have on the slides line by line, just by adjusting the probabilities accordingly.

(a) The probability of getting $i$ consecutive heads when flipping such a biased coin $i$ times is $(1/3)^i$. Thus, an element has links on levels $0 - i$ (and possibly also on higher levels) with probability $1/3^i$.

(b) If $n$ elements belong to a set with a probability $p$ each, then the expected size of that set is $np$. Thus, an $n$ element Skip List has on average $n/3^i$ elements with links on level $i$.

(c) Let $\#(i)$ denote the number of elements on level $i$. Since the expected number of elements having a link at level $i$ is $E[\#(i)] = n/3^i$, by the Markov inequality the probability of having at least one element at level $i$ satisfies

$$P(\#(i) \geq 1) \leq \frac{E[\#(i)]}{1} = \frac{n}{3^i}.$$ 

Thus, the probability to have an element on level $k \log_3 n$ is smaller than $n/3^{k \log_3 n} = n/3^{k \log_3 n} = n/n^k = 1/n^{k-1}$. Thus, the probability that level $k \log_3 n$ is nonempty is smaller than $1/n^{k-1}$.

(d) The above implies that the expected value $E$ of $k$ such that $k$ is the least integer so that the number of levels is $\leq k \log_3 n$ is bounded by

$$E \leq \sum_{k=1}^{\infty} \frac{k}{n^{k-1}} = \left( \frac{n}{n-1} \right)^2$$

(using one of the standard tricks covered in class to evaluate such a sum, for example for all $-1 < x < 1$,

$$\frac{d}{dx} \sum_{k=0}^{\infty} x^k = \frac{d}{dx} \left( \frac{1}{1-x} \right)$$

which implies

$$\sum_{k=0}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}$$

now just take $x = 1/n$.) Thus, the expected number of levels is barely larger than $\log_3 n$ and is certainly smaller than $2 \log_3 n$.

(e) If an element has a link at a level $i$ then with probability $1/3$ it also has a link at level $i + 1$. 

2
Thus, the expected number of elements which have links only on levels $0 - i$ and which are between two consecutive elements with a link on level $i + 1$ is smaller than (note: we go up with probability $1/3$ and do not go up with probability $2/3$)

$$0 \times \frac{1}{3} + 1 \times \frac{2}{3} \times \frac{1}{3} + 2 \times \left(\frac{2}{3}\right)^2 \times \frac{1}{3} + 3 \times \left(\frac{2}{3}\right)^3 \times \frac{1}{3} + \ldots$$

$$= \frac{1}{3} \times \frac{2}{3} \times \sum_{k=1}^{\infty} k \left(\frac{2}{3}\right)^{k-1} = \frac{1}{3} \times \frac{2}{3} \times \frac{1}{(1 - 2/3)^2} = 2$$

So once on level $i$, on average we will have to inspect only $2+1$ (1 for the starting element) i.e., three elements on that level before going to a lower level.

To summarise, on average, there will be fewer than $2 \log_3 n$ levels to go down, with visiting on average only three elements per each level. Consequently, the average search time is bounded by $2 \times 3 \times \log_3 n$.

This compares with $2 \times 2 \times \log_2 n$ for the unbiased coin. Since $\log_3 n = \Theta(\log_2 n)$ we conclude that asymptotically the two Skip Lists have the same run time. (Note that the ratio of these two estimates is

$$\frac{2 \times 2 \times \log_2 n}{2 \times 3 \times \log_3 n} = \frac{\ln 9}{\ln 8} \approx 1.06$$

so the biased coin Skip List might be negligibly faster.)

2. Assume that you have a device which produces lists of 10,000 uniformly distributed random integers in the range $0 - 10^{10}$. You need to sort 10,000 of these lists and have an option to use either the QuickSort or the Randomised Quick Sort. Which one would do the job faster? Explain why. (10 points)

Since the input numbers are all equally likely, there is no point in taking a random pivot because the standard, first element pivot is already random. However, producing random index value to take for the pivot in the Randomised Quick Sort takes some additional (but small) amount of time. So the randomised Quick Sort will be slightly slower than the standard, non-randomised Quick Sort.

3. You have an algorithm which on an input of length $n$ succeeds in certain task with a probability of $1/n$. You can run it $k(n) \times n$ many times, where $k(n)$ is
a function you have to choose. What should you choose for \( k(n) \) to succeed at least once with probability of at least \( 1 - 1/n \)? (16 points)

If you ran your algorithm \( k(n) \times n \) many times, the probability to fail in all attempts is equal to

\[
\left(1 - \frac{1}{n}\right)^{k(n) \times n} = \left(\left(1 - \frac{1}{n}\right)^n\right)^{k(n)} \approx \frac{1}{e^{k(n)}}
\]

using the fact that

\[
\left(1 - \frac{1}{n}\right)^n \approx \frac{1}{e}
\]

If you ran your algorithm \( k(n) \times n \) many times, the probability to fail in all attempts is at most \( \frac{1}{e^{kn}} \). Thus, if you pick \( k(n) = \ln n \) you get that the probability to fail in \( n \ln n \) attempts is at most \( \frac{1}{e^{n \ln n}} = \frac{1}{n} \) and consequently the probability of at least one success is at least \( 1 - 1/n \).

4. Consider a 4X4 board consisting of 16 squares with a spider walking on it; see the figure below. Every second the spider moves from the square where it is at the moment to any of the available adjacent squares vertically up or down or horizontally left or right (but not diagonally and not off the board) with equal probability in all available directions. Consider the Markov process whose states are the 16 squares where the spider can be.

(a) Consider the transition probabilities which corresponds to such a Markov chain with states enumerated as shown on the figure so that the transition probabilities satisfy the stated requirements. Specify the transition probabilities:

i. From C1 to all other states. (2 points)
ii. From C2 to all other states. (2 points)
iii. From C6 to all other states. (2 points)

(b) Is such a Markov Chain irreducible, i.e., is the underlying graph strongly connected? (8 points)

(c) Is it periodic? (this is a tiny bit tricky so think carefully if the length of every path starting and ending at the same square must be divisible by a fixed number) (16 points)

See the next page for a figure.
(a) 

\[ P(1 \rightarrow 2) = P(1 \rightarrow 5) = 1/2 \quad \text{all other probabilities from 1 are 0} \]

\[ P(2 \rightarrow 1) = P(2 \rightarrow 3) = P(2 \rightarrow 6) = 1/3 \quad \text{all other probabilities from 2 are 0} \]

\[ P(6 \rightarrow 2) = P(6 \rightarrow 5) = P(6 \rightarrow 7) = P(6 \rightarrow 10) = 1/4 \quad \text{all other probabilities from 6 are 0} \]

(b) Clearly, there is a path from every square to every other square always going to a directly adjacent square, so the Markov chain is irreducible.

(c) This Markov Chain is periodic, with the length of all loops starting and ending at any square divisible by 2. To see that, just notice that if you start from any cell the number of times you went one square to the left must be equal to the number of times you went one square to the right; similarly, the number of times you went one square up must be equal to the number of times you went one cell down, in order to return to the starting point. So the number of horizontal moves is even, being equal to the number of moves to the left plus equal number of moves to the right and similarly for the number of moves vertically.

5. Assume you have a ball of unit radius in a space of a very high dimension \( d \), say \( d \geq 1000 \). You draw a point from such a unit ball uniformly at random. Show that the probability that it will be at a distance at most \( 1/d \) from the surface
of such a ball is larger than $1 - 1/e$ where $e$ is the Euler constant $2.71\ldots$ (20 points)

Note that the point will be at the distance of less than $1/d$ just in case it belongs to the annulus between the ball of radius 1 and the ball of radius $1 - 1/d$. The probability that a point belongs to such an annulus is equal to

$$\frac{V(d) - (1 - 1/d)^dV(d)}{V(d)} = 1 - (1 - 1/d)^d$$

where $V(d)$ is the volume of a unit ball of dimension $d$. Here we used the fact that a ball of radius $r$ in $d$–dimensional space has volume $r^d V(d)$. We now use the fact that for large $d$,

$$(1 - 1/d)^d \approx 1/e$$