3. Exercise 8.23: In a tandem queue, customers arrive to an $M/M/1$ queue according to a Poisson process of rate $\lambda$ with service times independent and exponentially distributed with parameter $\mu_1$. After completing service at this first queue, the customers proceed immediately to a second queue, also being served by a single server, where service times are independent and exponentially distributed with parameter $\mu_2$. Find the stationary distribution of this system. (*Hint:* Try to generalize the form of the stationary distribution for a single queue.)

Let $i$ and $j$ be the number of customers in the first and second queue respectively at some moment. The next event may be a new arrival to the first queue, proceeding from the first queue to the second queue (if $i > 0$) or completing service at the second queue (if $j > 0$). Waiting times for these events are independent and exponentially distributed with parameters $\lambda$, $\mu_1$ and $\mu_2$ respectively. By lemma 8.5 we know that the time before the first of these happens is also exponentially distributed with parameter $\theta_{(i,j)}$, where

$$\begin{align*}
\theta_{(0,0)} &= \lambda \\
\theta_{(0,j)} &= \lambda + \mu_2 \\
\theta_{(i,0)} &= \lambda + \mu_1 \\
\theta_{(i,j)} &= \lambda + \mu_1 + \mu_2.
\end{align*}$$

Transition probabilities from state $(i, j)$ to state $(k, h)$, denoted by $p_{(i,j),(k,h)}$, are also given by lemma 8.5:

$$\begin{align*}
p_{(0,0),(1,0)} &= 1 \\
p_{(0,j),(1,j)} &= \frac{\lambda}{\lambda + \mu_2} \\
p_{(0,j),(0,j-1)} &= \frac{\mu_2}{\lambda + \mu_2} \\
p_{(i,0),(i+1,0)} &= \frac{\lambda}{\lambda + \mu_1} \\
p_{(i,0),(i-1,0)} &= \frac{\mu_1}{\lambda + \mu_1} \\
p_{(i,j),(i+1,j)} &= \frac{\lambda}{\lambda + \mu_1 + \mu_2} \\
p_{(i,j),(i-1,j+1)} &= \frac{\mu_1}{\lambda + \mu_1 + \mu_2} \\
p_{(i,j),(i,j-1)} &= \frac{\mu_2}{\lambda + \mu_1 + \mu_2}.
\end{align*}$$

For the stationary distribution, denoted by $\pi_{(i,j)}$, we have the rate equations (equation 8.4 in the book). By plugging the above values $\theta_{(i,j)}$ and $p_{(i,j),(k,h)}$ into the rate equations we get the following set of equations:

$$\begin{align*}
\pi_{(0,0)} \lambda &= \pi_{(0,1)} \mu_2 \\
\pi_{(0,j)} (\lambda + \mu_2) &= \pi_{(1,j-1)} \mu_1 + \pi_{(0,j+1)} \mu_2 \\
\pi_{(i,0)} (\lambda + \mu_1) &= \pi_{(i-1,0)} \lambda + \pi_{(i,1)} \mu_2 \\
\pi_{(i,j)} (\lambda + \mu_1 + \mu_2) &= \pi_{(i-1,j)} \lambda + \pi_{(i+1,j-1)} \mu_1 + \pi_{(i,j+1)} \mu_2
\end{align*}$$

We need to find out $\pi_{(i,j)}$ that satisfies these equations and the constraint $\sum_{i,j} \pi_{(i,j)} = 1$. First we note that clearly the second queue does not have any effect on the first queue and thus the marginal stationary distribution of the first queue is the same as for a single $M/M/1$ queue, that is, $\pi_i = \sum_j \pi_{(i,j)} = (1 - \lambda/\mu_1) (\lambda/\mu_1)^i$. Thus we have $\pi_{(i,j)} = \pi_{jj} \pi_i = \pi_{jj} (1 - \lambda/\mu_1) (\lambda/\mu_1)^i$, where $\pi_{jj}$ denotes the conditional stationary distribution of $j$ given that the state of the first queue is $i$. On the other hand, if we let $\mu_1$ increase towards infinity, then in the limit all customers advance immediately to the second queue once they get to the first queue and therefore their arrival to the second queue is a Poisson process with rate $\lambda$. Hence, in the limit also the
distribution of the second queue is the same as for a single $M/M/1$ queue. We make guess, that this is the case for all $\mu_1$, that is $\pi_{ij} = (1 - \lambda/\mu_2)(\lambda/\mu_2)^i$ and thus 

$$\pi(i,j) = \left(1 - \frac{\lambda}{\mu_1}\right) \left(\frac{\lambda}{\mu_1}\right)^i \left(1 - \frac{\lambda}{\mu_2}\right) \left(\frac{\lambda}{\mu_2}\right)^j.$$ 

It is easy to check that this satisfies all the equations above and the constraint $\sum_{i,j} \pi(i,j) = 1$, and therefore, is the stationary distribution of the tandem queue.

2. Exercise 10.3: Show that the following alternative definition is equivalent to the definition of an FPRAS given in the chapter: A fully polynomial randomized approximation scheme (FPRAS) for a problem is a randomized algorithm for which, given an input $x$ and any parameter $\varepsilon$ with $0 < \varepsilon < 1$, the algorithm outputs an $(\varepsilon, 1/4)$-approximation in time that is polynomial in $1/\varepsilon$ and the size of the input $x$. (Hint: To boost the probability of success from $3/4$ to $1 - \delta$, consider the median of several independent runs of the algorithm. Why is the median a better choice than the mean?)

We need to show that the following two definitions for FPRAS are equivalent:

1. FPRAS is a randomized algorithm which, given input $x$ and parameters $\varepsilon$ and $\delta$ (with $0 < \varepsilon, \delta < 1$), outputs an $(\varepsilon, \delta)$-approximation in time poly($1/\varepsilon, \ln \delta^{-1}, |x|$).

2. FPRAS is a randomized algorithm which, given input $x$ and a parameter $\varepsilon$ (with $0 < \varepsilon < 1$), outputs an $(\varepsilon, 1/4)$-approximation in time poly($1/\varepsilon, |x|$).

It is easy to see that if the algorithm follows definition 1, then it also follows definition 2 when we select that $\delta = 1/4$. What remains, is to show the other direction.

Let $A$ be an algorithm that follows definition 2. Then the output $X$ of $A$ approximates the correct value $V$ such that 

$$\Pr(|X - V| \leq \varepsilon V) \geq 3/4.$$ 

(1)

We construct a new algorithm $A'$, that repeats $A$ independently $n = O(\ln(1/\delta))$ times for its input and then outputs the median $X_{(n/2)}$ of the results $X_1, \ldots, X_n$. The running time of the algorithm $A'$ is poly($1/\varepsilon, \ln \delta^{-1}, |x|$), which is in line with definition 1.

Consider then the probability of failure. If $X_{(n/2)}$ is outside of the range $U = [(1 - \varepsilon)V, (1 + \varepsilon)V]$, then at least half of the all results $X_i$ returned by $A$ must be outside of this range. From (3) it follows that the probability that one result $X_i \not\in U$ is at most $1/4$. We can use independent Bernoulli trials to get an upper bound for the failure probability for multiple results. Let $Z_1, \ldots, Z_n$ be independent indicator variables with parameter $p = 1/4$ and let $Z = \sum Z_i$. Then $\mathbf{E}[Z] = n/4$ is an upper bound for the expected number of failures. By using the Chernoff bound (4.2) from the book (page 64) we get that

$$\Pr(X_{(n/2)} \not\in U) \leq \Pr(Z \geq n/2) = \Pr(Z \geq (1 + 1)n/4) \leq e^{-n/12}.$$ 

Therefore, if we choose $n > 12 \ln(1/\delta)$, the probability of failure is less than $\delta$ and $A'$ is a FPRAS according to definition 1.
3. Exercise 10.5:

a) Let $S_1, S_2, \ldots, S_m$ be subsets of a finite universe $U$. We know $|S_i|$ for $1 \leq i \leq m$. We wish to obtain an $(\epsilon, \delta)$-approximation to the size of the set

$$S = \bigcup_{i=1}^{m} S_i.$$  

We have available a procedure that can, in one step, choose an element uniformly at random from a set $S_j$. Also, given an element $x \in U$, we can determine the number of sets $S_j$ for which $x \in S_j$. We call this number $c(x)$.

Define $p_i$ to be

$$p_i = \frac{|S_i|}{\sum_{j=1}^{m} |S_j|}.$$  

The $j$th trial consists of the following steps. We choose a set $S_j$, where the probability of each set $S_i$ being chosen is $p_i$, and then we choose an element $x_j$ uniformly at random from $S_j$. In each trial the random choices are independent of all other trials. After $t$ trials, we estimate $|S|$ by

$$\left(\frac{1}{t} \sum_{j=1}^{t} \frac{1}{c(x_j)}\right) \left(\sum_{i=1}^{m} |S_i|\right).$$  

Determine – as a function of $m$, $\epsilon$ and $\delta$ – the number of trials needed to obtain an $(\epsilon, \delta)$-approximation to $|S|$.

Let $X_1, \ldots, X_t$ be our randomly selected elements, $Z_t = 1/c(X_i)$, $T = \sum_{j=1}^{m} |S_j|$ and $Y = \frac{t}{T} \sum_{j=1}^{m} Z_t$ such that $Y$ is our estimate for $|S|$. The probability, that an element $x$ gets selected as a sample, is

$$\text{Pr}(X_t = x) = \sum_{i : x \in S_i} p_i \frac{1}{|S_i|} = \sum_{i : x \in S_i} \frac{1}{|S|} = \frac{c(x)}{T}.$$  

Thus,

$$\mathbb{E}[Z_t] = \mathbb{E} \left[ \frac{1}{c(X_t)} \right] = \sum_{x \in S} \text{Pr}(X_t = x) \frac{1}{c(X_t)} = \frac{|S|}{T},$$

and furthermore

$$\mathbb{E}[Y] = \frac{T}{t} \sum_{j=1}^{t} \mathbb{E}[Z_j] = |S|,$$

which means that the expectation of our estimate is correct.

Now we would like to use Theorem 10.1 to determine the number of trials needed to obtain an $(\epsilon, \delta)$-approximation, but the theorem requires indicator random variables which our variables $Z_t$ are not. We modify the procedure such that once we obtain an element $X_t$, then instead of setting directly $Z_t = 1/c(X_t)$, we set $Z'_t = 1$ with probability $1/c(X_t)$ and $Z'_t = 0$ with probability $1 - 1/c(X_t)$. Now $\mathbb{E}[Z'_t | X_t] = 1/c(X_t)$ and therefore

$$\mathbb{E}[Z'_t] = \mathbb{E}[\mathbb{E}[Z'_t | X_t]] = \mathbb{E} \left[ \frac{1}{c(X_t)} \right] = \frac{|S|}{T},$$

that is, the expectation of $Z'_t$ remains the same as the expectation of $Z_t$. Thus, also the expectation of $Y' = \frac{T}{t} \sum_{j=1}^{m} Z'_j$ remains also the same as the expectation of $Y$. In addition, since instead of sampling values that belong to an interval $[0, 1]$ the modified version samples the end points, $Z'_t$ has larger deviation than $Z_t$ and thus $Y'$ is less peaked around its expectation than $Y$. As a consequence the tail probabilities increase for the modified procedure and the number of samples that is enough for the modified procedure will be enough also for the original procedure. (The actual proof for this is Exercise 4.19 on page 87.)

But now $Z'_t$s are i.i.d. 0-1 variables, which means that we can use Theorem 10.1. The number of samples that is enough for $(\epsilon, \delta)$-approximation is therefore

$$t \geq \frac{3\ln(2/\delta)}{\epsilon^2 \mathbb{E}[Z'_t]}.$$
Since $E[Z'] = |S|/T \leq 1/m$, the above condition is satisfied by selecting

$$t \geq \frac{3m\ln(2/\delta)}{\epsilon^2}.$$ 

b) Explain how to use your results from part (a) to obtain an alternative approximation algorithm for counting the number of solutions to a DNF problem.

Let $F = C_1 \lor C_2 \lor \cdots \lor C_l$ be our DNF formula. Assume again, that there is no a variable and its negation in any clause. Denote by $S_t$ the set of assignments that satisfy clause $C_t$. Then we know that $|S_t| = 2^{n-l}$, where $n$ is the number of all variables and $l_t$ is the number of literals in $C_t$. Now it is easy to choose a random assignment from $S_t$ by fixing the variable included in $C_t$ so that $C_t$ gets satisfied and randomly selecting values for the rest. In addition, for any assignment it is easy to test which of the clauses are satisfied by it, that is, which sets $S_t$ contain it. Since the set $S$ of all assignments that satisfy $F$ is the union over sets $S_t$, we can use the algorithm from part (a) to approximate the size of $S$.

4. Exercise 10.6: The problem of counting the number of solutions to a knapsack instance can be defined as follows: Given items with sizes $a_1, a_2, \ldots, a_n > 0$ and an integer $b > 0$, find the number of vectors $(x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$ such that $\sum_{i=1}^n a_i x_i \leq b$. The number $b$ can be thought of as the size of a knapsack, and the $x_i$ denote whether or not each item is put into the knapsack. Counting solutions correspond to counting the number of different sets of items that can be placed in the knapsack without exceeding its capacity.

Let $c(\bar{x}) = \sum_{i=1}^n a_i x_i$ for a vector $\bar{x} \in \{0, 1\}^n$ denoting a set of items. The set $\bar{x}$ can be placed in the knapsack of capacity $b$ if $c(\bar{x}) \leq b$.

a) A naive way of counting the number of solutions to this problem is to repeatedly choose $(x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$ uniformly at random, and return the $2^n$ times the fraction of samples that yield valid solutions. Argue why this is not a good strategy in general; in particular, argue that it will work poorly when each $a_i$ is 1 and $b = \sqrt{n}$.

In general, if $b$ is small, then there are only few vectors $\bar{x}$ that satisfy the condition $c(\bar{x}) \leq b$ and thus the probability of hitting such a vector is small when the samples are drawn uniformly.

For a concrete example, consider the case where $a_i = 1$ for all $i$ and $b = \sqrt{n}$. Denote by $X_i$ the random variable corresponding the bit $x_i$ in the vector and let $X = \sum X_i$. The solution is valid if $X \leq b$. Since $X_i$s are independent Poisson trians with probability $p = 1/2$, the bound (4.11) (page 71 in the book) says that

$$\Pr(X \leq (1 - \delta)n/2) \leq \exp\left(-\delta^2 n \frac{n}{2}\right).$$

By setting $(1 - \delta)n/2 = \sqrt{n}$, we get that $\delta = 1/2/\sqrt{n}$ and

$$\frac{n}{2} = \frac{1}{2} (\sqrt{n} - 2) \approx \frac{n}{2}.$$

This means, that the probability of finding a valid solution decreases exponentially with respect to $n$. Thus, to find at least one valid solution, the algorithm needs to sample (in expectation) an exponential number of random vectors.

b) Consider a Markov chain $X_0, X_1, \ldots$ on vectors $(x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$. Suppose $X_j$ is $(x_1, x_2, \ldots, x_n)$. At each step, the Markov chain chooses $i \in [1, n]$ uniformly at random. If $x_i = 1$ then $X_{j+1}$ is obtained from $X_j$ by setting $x_i$ to 0. If $x_i = 0$, then $X_{j+1}$ is assigned from $X_j$ by setting $x_i$ to 1 if $a_i > b$. Otherwise $X_{j+1} = X_j$.

Argue that this Markov chain has a uniform stationary distribution whenever $\sum_{i=1}^n a_i > b$. Be sure to argue that the chain is irreducible and aperiodic.

The chain is irreducible since from any valid state $\bar{x}$ we can get to any other valid state $\bar{y}$ by first setting all 1-bits in $\bar{x}$ to 0 one by one (thus moving to state $(0, 0, \ldots, 0)$) and then setting one by one those bits to 1 that are 1 in $\bar{y}$.
From the condition $\sum_{i=1}^n a_i > b$ it follows that there is at least one invalid state that exceeds the capacity of the knapsack (namely $(1,1,\ldots,1)$). It also means that there is at least one valid state, which does not contain all the elements but would exceed the capacity if any item was added to it. Therefore this state has a self loop and the chain is aperiodic.

Since the transition probabilities of the algorithm are

$$P_{x,y} = \begin{cases} 1/n & \text{if } \bar{x} \neq \bar{y} \text{ and } \bar{y} \in N(\bar{x}) \\ 0 & \text{if } \bar{x} \neq \bar{y} \text{ and } \bar{y} \notin N(\bar{x}) \\ 1 - |N(x)|/n & \text{if } \bar{x} = \bar{y} \end{cases}$$

the condition of Lemma 10.7 are fullfilled and the stationary distribution of the chain is uniform.

c) Argue that, if we have an FPAS for the knapsack problem, then we can derive an FPRAS for the problem. To set the problem up properly, assume without loss of generality that $a_1 \leq a_2 \leq \cdots \leq a_n$. Let $b_0 = 0$ and $b_i = \sum_{j=1}^i a_j$. Let $\Omega(b_i)$ be the set of vectors $(x_1,x_2,\ldots,x_n) \in \{0,1\}^n$ that satisfy $\sum_{i=1}^n a_i x_i \leq b_i$. Let $k$ be the smallest integer such that $b_k \geq b$. Consider the equation

$$|\Omega(b)| = \frac{|\Omega(b_1)|}{|\Omega(b_{k-1})|} \times \frac{|\Omega(b_2)|}{|\Omega(b_{k-2})|} \times \cdots \times \frac{|\Omega(b_k)|}{|\Omega(b_0)|}.$$

You will need to argue that $|\Omega(b_i)|/|\Omega(b_1)|$ is not too small. Specifically, argue that $|\Omega(b_i)| \leq (n+1)|\Omega(b_{i-1})|$. We will build a FPRAS for $1/|\Omega(b)|$. Then we can easily get a FPRAS for $|\Omega(b)|$ by approximating $1/|\Omega(b)|$ with precision $\epsilon/(1-\epsilon)$. Let $r_i = |\Omega(b_{i-1})|/|\Omega(b_i)|$ for $i < k$ and $r_k = |\Omega(b_{k-1})|/|\Omega(b_k)|$. Since $|\Omega(b_k)| = 1$, we then have $1/|\Omega(b)| = \prod_{i=1}^k r_i$. Let $\tilde{r}_i$ be $\epsilon/(2n), \delta/n$-approximation for $r_i$. Then by Lemma 10.3, we have

$$\Pr\left(\left|\prod_{i=1}^k \tilde{r}_i - 1\right| \leq \epsilon\right) \geq 1 - \delta$$

and thus $\prod_{i=1}^k \tilde{r}_i$ is an $(\epsilon, \delta)$-approximation for $1/|\Omega(b)|$.

To get each value $\tilde{r}_i$, we first use a FPAUS algorithm to produce $M$ samples from $(\epsilon/(6n^2))$-uniform distribution over $\Omega(b_i)$ (or $\Omega(b)$ for $i = k$) and then use the fraction of the samples belonging to $\Omega(b_{i-1})$ as the approximation $\tilde{r}_i$.

First we need to prove that $r_i$s are not too small. We show that $r_i \geq 1/(n+1)$ by using a similar approach as the proof of Lemma 10.4. Let $f : \Omega(b_1) \setminus \Omega(b_{i-1}) \rightarrow \Omega(b_{i-1})$ be a mapping such that $f(\bar{x})$ returns a set $\bar{y}$ where the heaviest element of $\bar{x}$ is removed, that is, the 1-bit with the largest index is set to 0. By definition, we have $b_{i-1} < c(\bar{x}) \leq b_i$, which is equivalent to $\sum_{j=1}^{i-1} a_j < c(\bar{x}) \leq \sum_{j=1}^{i} a_j$. On the other hand, the index of the largest element in $\bar{x}$ must be at least $i$, since otherwise $\bar{x}$ would belong to $\Omega(b_{i-1})$. Therefore removing the largest element decreases the total weight by at least $a_i$, which means that $c(\bar{y}) \leq b_{i-1}$ and thus $\bar{y}$ must belong to $\Omega(b_i)$.

As there are a total of $n$ bits that can be set to 0, each $\bar{y}$ has at most $n$ vectors $\bar{x}$ such that $\bar{y} = f(\bar{x})$. It then follows that

$$|\Omega(b_i) \setminus \Omega(b_{i-1})| \leq n |\Omega(b_{i-1})|$$

and thus

$$r_i = \frac{|\Omega(b_{i-1})|}{|\Omega(b_i)|} = \frac{|\Omega(b_{i-1})|}{|\Omega(b_{i-1})| + |\Omega(b_i) \setminus \Omega(b_{i-1})|} \geq \frac{|\Omega(b_{i-1})|}{|\Omega(b_{i-1})| + n |\Omega(b_{i-1})|} = \frac{1}{1 + n}$$

Following the proof of Lemma 10.4, the properties of our sampling algorithm imply that

$$|E[\hat{r}_i] - r_i| \leq \frac{\epsilon}{6n^2}.$$  \hspace{1cm} (2)

By combining this and the above observation, we get

$$E[\hat{r}_i] \geq r_i - \frac{\epsilon}{6n^2} \geq \frac{1}{n+1} - \frac{\epsilon}{6n^2} \geq \frac{1}{n+2} \geq \frac{1}{2n}.$$
Then by Theorem 10.1 we know that \( \hat{r}_i \) is an \((\varepsilon/(12n), \delta/n)\)-approximation for \( E[\hat{r}_i] \), if the number of samples \( M \) satisfies

\[
M \geq \frac{3 \ln(2n/\delta)}{((\varepsilon/(12n))^2 (1/(2n)))} = \frac{864 \ln(2n/\delta) n^3}{\varepsilon^2},
\]

which is polynomial in \( n, \ln(\delta^{-1}) \) and \( \varepsilon^{-1} \). Thus, with probability at least \( 1 - \delta/n \) the following holds:

\[
1 - \frac{\varepsilon}{12n} \leq \frac{\hat{r}_i}{E[\hat{r}_i]} \leq 1 + \frac{\varepsilon}{12n},
\]

(3)

On the other hand, from (2) and the fact that \( r_i \geq 1/(n + 1) \geq 1/(2n) \) we get that

\[
1 - \frac{\varepsilon}{3n} \leq \frac{E[\hat{r}_i]}{r_i} \leq 1 + \frac{\varepsilon}{3n},
\]

(4)

Combining (3) and (4) we see (like in the proof of Lemma 10.4) that with probability at least \( 1 - \delta/n \) we have

\[
1 - \frac{\varepsilon}{2n} \leq \frac{\hat{r}_i}{r_i} \leq 1 + \frac{\varepsilon}{2n},
\]

that is, \( \hat{r}_i \) is an \((\varepsilon/(2n), \delta/n)\)-approximations for \( r_i \). The claim thus follows.

5. **Exercise 10.12:** The following generalization of the Metropolis algorithm is due to Hastings. Suppose that we have a Markov chain on a state space \( \Omega \) given by the transition matrix \( Q \) and that we want to construct a Markov chain on this state space with a stationary distribution \( \pi_x = b(x)/B \), where for all \( x \in \Omega \), \( b(x) > 0 \) and \( B = \sum_{x \in \Omega} b(x) \) is finite. Define a new Markov chain as follows. When \( X_n = x \), generate a random variable \( Y \) with \( \Pr(Y = y) = Q_{xy} \). Notice that \( Y \) can be generated by simulating one step of the original Markov chain. Set \( X_{n+1} \) to \( Y \) with probability

\[
\min \left( \frac{\pi_y Q_{yx}}{\pi_x Q_{xy}}, 1 \right),
\]

and otherwise set \( X_{n+1} = X_n \). Argue that if this chain is aperiodic and irreducible, then it is also time reversible and has a stationary distribution given by the \( \pi_x \).

It is enough to show that the chain is time reversible, the claim then follows from Theorem 7.10. Let \( P \) be the transition matrix of the chain. Let \( x \neq y \). If \( Q_{xy} = 0 \) (or \( Q_{yx} = 0 \)), then also \( P_{xy} = 0 \) and \( P_{yx} = 0 \) and therefore \( \pi_x P_{yx} = \pi_y P_{xy} \). Assume then that \( Q_{xy} > 0 \) and \( Q_{yx} > 0 \). Then we have

\[
P_{xy} = Q_{xy} \min \left\{ \frac{\pi_y Q_{yx}}{\pi_x Q_{xy}}, 1 \right\}.
\]

Now if \( \frac{\pi_y Q_{yx}}{\pi_x Q_{xy}} \leq 1 \), then \( P_{xy} = Q_{xy} \) and consequently

\[
P_{yx} = Q_{yx} \frac{\pi_y Q_{xy}}{\pi_x Q_{xy}} = \frac{\pi_x}{\pi_y} P_{xy}.
\]

Thus, \( \pi_x P_{yx} = \pi_y P_{xy} \). By symmetry this also holds if \( \frac{\pi_y Q_{yx}}{\pi_x Q_{xy}} \geq 1 \). Therefore, the chain is time reversible and has \( \pi \) as its stationary distribution.

6. **Exercise 11.11:** Show that the Markov chain for sampling all independent sets of size exactly \( k \leq n/3(\Delta + 1) \) in a graph with \( n \) nodes and maximum degree \( \Delta \), as defined in Section 11.2.3, is ergodic and has a uniform stationary distribution.

To show that the chain is ergodic, we need to show that it is finite, aperiodic and irreducible. The chain is finite since the number of independent sets is finite. It is aperiodic since every state has a nonzero probability of staying at the same state.
The condition \( k \leq n/(3\Delta + 3) \) implies that \( (\Delta + 1)k \leq n/3 \). Thus any independent set of size \( k \) together its neighboring vertices covers at most \( n/3 \) nodes of the graph. Two independent sets \( x \) and \( y \) together with their neighbors therefore cover at most two thirds of the graph. We can move from \( x \) to \( y \) by using the remaining third as a temporary location for the vertices in independent set: Let \( z \) be any independent set contained in the remaining third. (Such a set exists since, as we saw above, third of the graph is always enough for an independent set of size \( k \) and its neighbors.) As the vertices in \( z \) do not have neighbors in \( x \) or \( y \), we can first exchange all vertices in \( x \) to the vertices in \( z \) one by one and then exchange these vertices to the vertices in \( y \) one by one. Therefore we can conclude that the chain is also irreducible and thus ergodic.

Then we need to show that the stationary distribution of the chain is uniform. Two states \( x \) and \( y \) are neighbors if they differ in one vertex. The maximum number of neighbors is \( N = \max_x |N(x)| \leq k(n-k) \leq kn \), since there are \( k \) possible differing vertices and for each at most \( n-k \) possible alternatives. On the other hand, the probability of moving to a specific neighboring state is \( 1/(kn) \) and the transition matrix is therefore

\[
P_{x,y} = \begin{cases} 
1/(kn) & \text{if } x \neq y \text{ and } y \in N(x) \\
0 & \text{if } x \neq y \text{ and } y \notin N(x) \\
1 - |N(x)|/(kn) & \text{if } x = y.
\end{cases}
\]

Since \( kn \geq N \) and the chain is ergodic, by Lemma 10.7 the stationary distribution is uniform.

7. **Exercise 11.14:** Consider the following variation on shuffling for a deck of \( n \) cards. At each step, two specific cards are chosen uniformly at random from the deck, and their positions are exchanged. (It is possible both choices give the same card, in which case no change occurs.)

a) Argue that the following is an equivalent process: at each step, a specific card is chosen uniformly at random from the deck, and a position from \( [1,n] \) is chosen uniformly at random; then the card at position \( i \) exchanges positions with the specific card chosen.

Both in the original process and in the modified process both cards are chosen independently and uniformly at random (choosing the second card by position does not change this fact). Therefore the processes are equivalent.

b) Consider the coupling where the two choices of card and position are the same for both copies of the chain. Let \( X_t \) be the number of cards whose positions are different in the two copies of the chain. Show that \( X_t \) is nonincreasing over time.

Let \( c \) be the chosen card and \( i \) be the chosen position. There are four cases:

(a) If the positions of \( c \) are same in both chains and the card in \( i \)th position is the same in both chains then the cards in both positions remain to be same in both chains after the exchange and thus \( X_t \) does not change.

(b) If the positions of \( c \) are different in the two chains and the card in \( i \)th position is the same in both chains then in the exchange the cards that were in \( i \)th position move to different position but the cards \( c \) move to the same position \( i \). Thus \( X_t \) remains the same.

(c) If the positions of \( c \) are same in both chains and the card in \( i \)th position is different in the two chains then after the exchange cards \( c \) still have the same position and the cards in the other position still are different. Thus \( X_t \) remains the same.

(d) If the positions of \( c \) are different in the two chains and the card in \( i \)th position is different in the two chains then in the exchange the cards \( c \) move into the same position and no pair of same cards in same position move in different position. Thus \( X_{t+1} \) decreases by at least one (in the best case it can decrease by three).

Therefore \( X_t \) never increases over time.

c) Show that

\[
\Pr(X_{t+1} \leq X_t - 1 \mid X_t > 0) \geq \left( \frac{X_t}{n} \right)^2.
\]
The event $X_{t+1} \leq X_t - 1$ corresponds to the case 4 above. Since there are $X_t$ of $n$ cards in different locations, the probability that the positions of the selected card are different in the two chains is $X_t/n$. Similarly, the probability that the card in the selected position is different in the two chains is $X_t/n$. Since the choices are independent, the probability of case 4 is

$$\Pr(X_{t+1} \leq X_t - 1 \mid X_t > 0) = \left( \frac{X_t}{n} \right)^2 \geq \left( \frac{X_t}{n} \right)^2.$$  

**d) Argue that the expected time until $X_t$ is 0 is $O(n^2)$, regardless of the starting state of the two chains.**

In the worst case $X_0 = n$ and $X_t$ always decreases by at most one in one step. If $X_t = k$, then the number $Y_k$ of exchanges before $X_t$ is decreased is geometrically distributed with parameter $\Pr(X_{t+1} \leq X_t - 1 \mid X_t = k) \geq (k/n)^2$. The expected number of exchanges before $X_t$ decreases is thus $E[Y_k] \leq (n/k)^2$ and the expected total time is therefore at most

$$E \left[ \sum_{k=1}^{n} Y_k \right] \leq \sum_{k=1}^{n} \frac{n}{k} \leq n^2 \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} = O(n^2).$$