Summary

The following notes contain the Land of Oz example of a Markov chain, and a sample reduction, or a proof that HALF 3-CNF is NP-COMPLETE.

Markov Chains

We will consider the canonical example of a regular Markov chain: the weather in the Land of Oz from *Introduction to Probability* by Grinstead and Snell. The Land of Oz is quite nice but it has lousy weather. In fact, as you will see from the transition matrix given below, the weather in Oz is *never* nice two days in a row.

$$P = \begin{array}{c|ccc} & R & N & S \\ \hline R & 1/2 & 1/4 & 1/4 \\ N & 1/2 & 0 & 1/2 \\ S & 1/4 & 1/4 & 1/2 \end{array}$$

The entry in cell p_{ij} is the probability of transitioning from state i to state j. For example, cell p_{11} is the probability of a rainy day today and a rainy day tomorrow. Or, in more formal language, cell p_{ij} signifies the probability of rain tomorrow *given that it is raining today.* Thus, we can rewrite the matrix in the following manner:

$$P = \begin{array}{c|ccc} & R & N & S \\ \hline R & P(R|R) & P(N|R) & P(S|R) \\ N & P(R|N) & P(N|N) & P(S|N) \\ S & P(R|S) & P(N|S) & P(S|S) \end{array}$$

Now we want to consider the question of transitioning from state to state over time. For example, given that it is raining today, what is the probability that it is snowing two days from now? The event that it is snowing two days from now is the disjoint union of the following three events:

1. It is rainy tomorrow and snowy two days from now.
2. It is nice tomorrow and snowy two days from now.
3. It is snowy tomorrow and snowy two days from now.

Let us consider the first event. The probability that it is raining tomorrow is just the entry is $p_{11} = P(R|R)$. The probability of snow the day after that is $P(S|R)$ (Note that we are assuming the Land of Oz has a *forgetful* distribution, or that the distribution remains the same regardless of what events have occured in the past). Thus, the probability of snow after rain after rain is given by $P(S|R) \cdot P(R|R)$. Thus, we see the following equation:

$$P(\text{snow day after tomorrow} \mid \text{rain today}) = P(S|R) \cdot P(R|R) + P(S|N) \cdot P(N|R) + P(S|S) \cdot P(S|R) \quad (1)$$

$$p_{13}^{(2)} = p_{13} \cdot p_{11} + p_{23} \cdot p_{12} + p_{33} \cdot p_{13} \quad (2)$$

But this notation is extremely familar. This is just the entry $p_{13} \in P^2$ or the dot product of the matrix with itself. Thus we can introduce the following theorem:

- **Theorem:** Let $\mathbf{P}$ be the transition matrix of a Markov chain. The ij-th entry $p_{ij}^{(n)}$ of the matrix $\mathbf{P}^n$ gives the probability that the Markov chain, starting in the state s_i will be in state s_j after n steps.

Now we will notice a very interesting property of the probability distribution of the weather in the Land of Oz. Notice what happens when we continually raise the matrix to a power:

$$P^1 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 1/2 & 1/4 & 1/4 \\ N & 1/2 & 0 & 1/2 \\ S & 1/4 & 1/4 & 1/2 \end{array}, \quad P^2 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 0.4375 & 0.1875 & 0.3750 \\ N & 0.3750 & 0.2500 & 0.3750 \\ S & 0.3750 & 0.1875 & 0.4375 \end{array}, \quad P^3 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 0.4063 & 0.2031 & 0.3906 \\ N & 0.4063 & 0.1875 & 0.4063 \\ S & 0.3906 & 0.2031 & 0.4063 \end{array}$$

$$P^4 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 0.4023 & 0.1992 & 0.3984 \\ N & 0.3984 & 0.2031 & 0.3984 \\ S & 0.3984 & 0.1992 & 0.4023 \end{array}, \quad P^5 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 0.4004 & 0.2002 & 0.3994 \\ N & 0.4004 & 0.1992 & 0.4004 \\ S & 0.3994 & 0.2002 & 0.4004 \end{array}, \quad P^6 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 0.4001 & 0.2000 & 0.3999 \\ N & 0.3999 & 0.2002 & 0.3999 \\ S & 0.3999 & 0.2000 & 0.4001 \end{array}$$

$$P^7 = \begin{array}{c|ccc} & R & N & S \\ \hline R & 0.4000 & 0.2000 & 0.4000 \\ N & 0.4000 & 0.2000 & 0.4000 \\ S & 0.4000 & 0.2000 & 0.4000 \end{array}$$

Thus we can see an extraordinary property of this matrix: The long-range prediction of the weather in Oz is the same regardless of what the weather is today. When a matrix has this property, when the long-range prediction of the distribution is independent of the starting state, this is an example of a *regular* Markov chain. Now we consider the question of determining the weather tomorrow, given a particular distribution of the weather today. Consider the following theorem:

- **Theorem:** Let $\mathbf{P}$ be the transition matrix of a Markov chain, and let $\mathbf{u}$ be the probability vector which represents the starting distribution. Then the probability that the chain is in state s_i after n steps is the i-th entry of the vector:

$$\mathbf{u}^{(n)} = \mathbf{u}\mathbf{P}^n \tag{3}$$

- **Proof Sketch:** Consider the following explanation of the above theorem.

$$\begin{aligned} P(\text{rain tomorrow}) &= P(R|R) \cdot P(R) + P(R|N) \cdot P(N) + P(R|S) \cdot P(S) && (4) \\ &= p_{11} \cdot u_1 + p_{21} \cdot u_2 + p_{31} \cdot u_3 && (5) \end{aligned}$$

  But that equation is also very familiar to us: it is simply the product of the vector with the matrix $\mathbf{P}$. Proving that $\mathbf{u}^{(n)}$ is the probability distribution after n days is a straightforward inductive proof from here.

Using the above theorem, we will consider two different probability distributions on the inital state, and watch how they converge over time. Let $\mathbf{u} = \{.9, .05, .05\}$ and $\mathbf{v} = \{.05, .05, .9\}$, and consider the distribution after 3 days, and after 7 days:

$$\begin{aligned} \mathbf{u}^{(3)} = \mathbf{u}\mathbf{P}^3 &= (.9, .05, .05) \begin{pmatrix} 0.4063 & 0.2031 & 0.3906 \\ 0.4063 & 0.1875 & 0.4063 \\ 0.3906 & 0.2031 & 0.4063 \end{pmatrix} && (6) \\ &= (0.4055, 0.2023, 0.3922) && (7) \\ \mathbf{u}^{(7)} = \mathbf{u}\mathbf{P}^7 &= (.9, .05, .05) \begin{pmatrix} 0.4000 & 0.2000 & 0.4000 \\ 0.4000 & 0.2000 & 0.4000 \\ 0.4000 & 0.2000 & 0.4000 \end{pmatrix} && (8) \\ &= (0.4000, 0.2000, 0.4000) && (9) \\ \mathbf{v}^{(3)} = \mathbf{v}\mathbf{P}^3 &= (.05, .05, .9) \begin{pmatrix} 0.4063 & 0.2031 & 0.3906 \\ 0.4063 & 0.1875 & 0.4063 \\ 0.3906 & 0.2031 & 0.4063 \end{pmatrix} && (10) \\ &= (0.3922, 0.2023, 0.4055) && (11) \\ \mathbf{u}^{(7)} = \mathbf{u}\mathbf{P}^7 &= (.05, .05, .9) \begin{pmatrix} 0.4000 & 0.2000 & 0.4000 \\ 0.4000 & 0.2000 & 0.4000 \\ 0.4000 & 0.2000 & 0.4000 \end{pmatrix} && (12) \\ &= (0.4000, 0.2000, 0.4000) && (13) \\ & && (14) \end{aligned}$$

Thus, we see exactly what we expect to see based on the fact that the Land of Oz Markov chain is regular: the probability distribution converges independent of the starting point.

Half 3-CNF Satisfiability: CLRS 34.5-8

In the **Half 3-CNF Satisfiability** problem, we are given a 3-CNF (3 conjunctive normal form) formula ϕ with n variables and m clauses where m is even. We wish to determine whether there exists a truth assignment to the variables of ϕ such that exactly half the clauses evaluate to true and exactly half the clauses evaluate to false. Prove that the HALF 3-CNF is NP-COMPLETE.

Solution:
In order to prove that the HALF 3-CNF problem is NP-COMPLETE, we need to prove two things:

1. HALF 3-CNF $\in NP$
2. $\forall L \in \text{NP}, L \leq_p$ HALF 3-CNF, or HALF 3-CNF $\in$ NP-HARD

In order to prove that HALF 3-CNF is in NP, we need to show that a solution to a particular instance of the problem can be verified in polynomial time. Consider the following "witness" algorithm which takes an instance of HALF 3-CNF and a "certificate" as parameters. Our algorithm takes a boolean formula in conjunctive normal form, and the "certificate" or "proposed solution" is a list of the boolean variables and their proposed truth assignments. The algorithm walks the formula, evaluating each clause with the proposed truth assignments and returns true if exactly half the clauses are true, and false if not. This is clearly polynomial time because the boolean formula is only walked once.

In order to prove that HALF 3-CNF is NP-HARD, we can consider a reduction *from* a known NP-COMPLETE problem *to* HALF 3-CNF (or prove it directly, which we won't even consider). We choose to reduce from *from* 3-CNF-SAT *to* HALF 3-CNF. Because CIRCUIT-SAT $\leq_p$ SAT $\leq_p$ 3-CNF-SAT, if we are able to show 3-CNF-SAT $\leq_p$ Half 3-CNF, then this implies that Circuit-SAT $\leq_p$ Half 3-CNF. And since $L \leq_p$ Circuit-SAT $\forall L \in$ NP, then this implies that $L \leq_p$ Half 3-CNF $\forall L \in$ NP or that Half 3-CNF $\in$ NP-hard.

Consider the following reduction. First, we need to convert an instance of 3-CNF-SAT to HALF 3-CNF. Our approach is create a ϕ' which contains 4 times as many clauses as ϕ. Suppose ϕ contains m clauses. When creating ϕ', first we take all of the clauses from ϕ. Next, we create m clauses of the form:

$$(p \vee \neg p \vee q) \tag{15}$$

Clearly, these clauses are always true, regardless of the individual truth assignments for p and q. Next, we create $2m$ clauses of the form:

$$(p \vee q \vee r) \tag{16}$$

These clauses are always true or always false. Therefore, we have created a boolean formula ϕ' which contains all of the clauses ϕ and m clauses which are always true and $2m$ clauses which are either all true or all false. Clearly, this conversion takes polynomial time because we have only added 3 variables, and $3m$ clauses.

We also need to show that there exists a "yes" instance of 3-CNF-SAT if and only if there exists a "yes" instance of HALF 3-CNF.

- $\Longrightarrow$ Assume that there exists a truth assignment which causes ϕ to be true. Then, the m clauses which correspond to ϕ in ϕ' are true and there are m clauses which are always true. Thus, simply let p and r be false, and there exists a truth assignment which satisfies HALF 3-CNF, where half the clauses are true and half are false.

- $\Longleftarrow$ Assume that there exists a truth assignment which causes Half 3-CNF to be satisfied, or a truth assignment that causes half the clauses in ϕ' to be true and half false. But m clauses in ϕ' are always true, which means that the $2m$ clauses cannot be true if Half 3-CNF is satisfied (because then $3m$ clauses would be true which is more than half). Thus, the $2m$ clauses must be false, which means that ϕ is true, which means that 3-CNF-SAT is also satisfied.

Thus, we have shown that a "yes" instance of 3-CNF-SAT produces a "yes" instance of Half 3-CNF, and vice versa, which concludes our proof. □