Department of EECS - University of California at Berkeley
EECS126 - Probability and Random Processes - Fall 2003
REVIEW

These notes go through a number of problems to help you review the material. We classify the problems by topic but we do not restrict ourselves to a strict ordering of the concepts. These notes complement the review for midterm 2.

---

1 Probability Space

We have learned that a probability space is $\{\Omega, \mathcal{F}, P\}$ where Ω is a nonempty set, $\mathcal{F}$ is a σ-field of Ω, i.e., a collection of subsets of Ω that is closed under countable set operations, and $P : \mathcal{F} \to [0, 1]$ is a σ additive set function such that $P(\Omega) = 1$.

The idea is to specify the likelihood of various outcomes (elements of Ω). If one can specify the probability of individual outcomes (e.g., when Ω is countable), then one can choose $\mathcal{F} = 2^{\Omega}$, so that all sets of outcomes are events. However, this is generally not possible as the example of the uniform distribution on $[0, 1]$ shows.

Recall also the definition of conditional probability $P[A \mid B] = P(A \cap B)/P(B)$ and that of independence of two events and of mutual independence of a collection of events.

Example 1 *Pick three balls from an urn with fifteen balls that are identical except that ten are red and five are blue. Specify the probability space.*

One possibility is to specify the color of the three balls in the order they are picked. Then

$$\Omega = \{R, B\}^3, \mathcal{F} = 2^{\Omega}, P(\{RRR\}) = \frac{10}{15}\frac{9}{14}\frac{8}{13}, \ldots, P(\{BBB\}) = \frac{5}{15}\frac{4}{14}\frac{3}{13}.$$

Example 2 *We deal five cards from a perfectly shuffled 52-card deck. Specify the probability space.*

One possibility is to choose Ω to be the set of all the permutations of the numbers 1 to 52. Here, ω represents the shuffled deck. Then $\mathcal{F} = 2^{\Omega}$ and $P(\{\omega\}) = 1/(52!)$ for $\omega \in \Omega$.

Note that this probability space is bigger than necessary since it specifies more than the five cards we deal. However, it is easy to specify.

Example 3 *You flip a fair coin until you get three consecutive 'heads'. Specify the probability space.*

One possible choice is $\Omega = \{H, T\}^*$, the set of finite sequences of H and T. That is,

$$\{H, T\}^* = \cup_{n=1}^{\infty} \{H, T\}^n.$$

This set Ω is countable, so we can choose $\mathcal{F} = 2^{\Omega}$. Here,

$$P(\{\omega\}) = 2^{-n} \text{where } n := \text{length of } \omega.$$

This is another example of a probability space that is bigger than necessary, but easier to specify than the smallest probability space we need.

Example 4 *Let $\Omega = \{0, 1, 2, \ldots\}$. Let $\mathcal{F}$ be the collection of subsets of Ω that are either finite or whose complement is finite. Is $\mathcal{F}$ a σ-field?*

No, $\mathcal{F}$ is not closed under countable set operations. For instance, $\{2n\} \in \mathcal{F}$ for each $n \geq 0$ because $\{2n\}$ is finite. However,

$$A := \cup_{n=0}^{\infty}\{2n\}$$

is not in $\mathcal{F}$ because both A and A^c are infinite.

Example 5 *Choose two numbers uniformly but without replacement in $\{0, 1, \ldots, 10\}$. What is the probability that the sum is less than or equal to 10 given that the smallest is less than or equal to 5?*

Draw a picture of

$$\Omega = \{0, 1, \ldots, 10\}^2 \setminus \{(i, i) \mid i = 0, 1, \ldots, 10\}.$$

The outcomes in Ω all have the same probability. Let also

$$A = \{\omega \mid \omega_1 \neq \omega_2 \text{ and } \omega_1 + \omega_2 \leq 10\}, B = \{\omega \mid \omega_1 \neq \omega_2 \text{ and } \min\{\omega_1, \omega_2\} \leq 5\}.$$

The probability we are looking for is

$$\frac{|A \cap B|}{|B|} = \frac{|A|}{|B|}.$$

Your picture shows that $|A| = 10 + 9 + 8 + \cdots + 1 = 55$ and that $|B| = 10 \times 5 + 4 \times 5 = 70$. Hence, the answer is $55/70 = 11/14$.

Example 6 *In a class with 24 students, what is the probability that no two students have the same birthday?*

Let $N = 365$ and $n = 24$. The probability is

$$\alpha := \frac{N}{N} \times \frac{N-1}{N} \times \frac{N-2}{N} \times \cdots \times \frac{N-n+1}{N}.$$

To estimate this quantity we proceed as follows. Note that

$$\begin{aligned} ln(\alpha) &= \sum_{k=1}^{n} ln(\frac{N-n+k}{N}) \approx \int_1^n ln(\frac{N-n+x}{N})dx \\ &= N\int_a^1 ln(y)dy = N[yln(y) - y]_a^1 \\ &= -(N-n+1)ln(\frac{N-n+1}{N}) - (n-1). \end{aligned}$$

(In this derivation we defined $a = (N - n + 1)/N$.) With $n = 24$ and $N = 365$ we find that $\alpha \approx 0.48$.

2 Random Variables

A random variable is a function $X : \Omega \to \Re$ such that $X^{-1}((-\infty, x]) \in \mathcal{F}$ for all $x \in \Re$. We then define the c.d.f. $f_X(.)$ of X as

$$f_X(x) = P(X \leq x) := P(X^{-1}((-\infty, x])), x \in \Re.$$

The random variable X is continuous if

$$f_X(x) = \int_{-\infty}^{x} f_X(u)du, x \in \Re.$$

In that case, f_X is the p.d.f. of X. We know that

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)dF_X(x).$$

If the random variable takes countably many values, then the list of these values together with their probabilities is called the p.m.f.

The same story extends to multiple random variables and you know about j.p.m.f., j.c.d.f., j.p.d.f. You also know how to calculate $E(h(\mathbf{X}))$. In particular, you recall the variance, covariance, k-th moment. The vector notation has few secrets for you. For instance, you know that

$$cov(A\mathbf{X}, B\mathbf{Y}) = Acov(\mathbf{X}, \mathbf{Y})B^T.$$

You also understand the following derivation:

$$E(\mathbf{X}^T\mathbf{Y}) = E(tr(\mathbf{X}^T\mathbf{Y})) = E(tr(\mathbf{Y}\mathbf{X}^T)) = trE(\mathbf{Y}\mathbf{X}^T).$$

You should remember the simple p.m.f. and p.d.f. Here is a list of what I assume you know:

$$B(p), B(n,p), G(p), P(\lambda), U[a,b], Exp(\lambda), N(\mu, \sigma^2), N(\mu, \Sigma).$$

You also know the definition (and meaning) of independence and mutual independence and you know that the mean value of a product of independent random variables is the product of their mean values. You can also prove that functions of independent random variables are independent.

Example 7 *Give an example of a probability space and a real-valued function on Ω that is not a random variable.*

Let $\Omega = \{0, 1, 2\}, \mathcal{F} = \{\emptyset, \{0\}, \{1, 2\}, \Omega\}, P(\{0\}) = 1/2 = P(\{1, 2\})$, and $X(\omega) = \omega$ for $\omega \in \Omega$. The function X is not a random variable since

$$X^{-1}((-\infty, 1]) = \{0, 1\} \notin \mathcal{F}.$$

The meaning of all this is that the probability space is not rich enough to specify $P(X \leq 1)$.

Example 8 *Let X, Y be two points picked independently and uniformly on the circumference of the unit circle. Define $Z = ||X - Y||^2$. Find $f_Z(.)$.*

By symmetry we can assume that the point X has coordinates $(1,0)$. The point Y then has coordinates $(\cos(\theta), \sin(\theta))$ where θ is uniformly distributed in $[0, 2\pi]$. Consequently, $X - Y$ has coordinates $(1-\cos(\theta), -\sin(\theta))$ and $Z = (1-\cos(\theta))^2 + \sin^2(\theta) = 2(1-\cos(\theta)) =: g(\theta)$.

We now use the basic results on the density of a function of a random variable. To review how this works, note that if $\theta \in (\theta_0, \theta_0 + \epsilon)$, then

$$g(\theta) \in (g(\theta_0), g(\theta_0 + \epsilon)) = (g(\theta_0), g(\theta_0) + g\prime(\theta_0)\epsilon).$$

Accordingly,

$$g(\theta) \in (z, z + \delta)$$

if and only if

$$\theta \in ((\theta_n, \theta_n + \frac{\delta}{g\prime(\theta_n)})$$

for some θ_n such that $g(\theta_n) = z$. It follows that, if $Z = g(\theta)$, then

$$f_Z(z) = \sum_n \frac{1}{|g\prime(\theta_n)|} f_\theta(\theta_n).$$

In this expression, the sum is over all the θ_n such that $g(\theta_n) = z$.

Coming back to our example, $g(\theta) = z$ if $2(1 - \cos(\theta)) = z$. In that case, $|g\prime(\theta)| = 2|\sin(\theta)| = 2\sqrt{1 - (1 - \frac{z}{2})^2}$. Note that there are two values of θ such that $g(\theta) = z$ whenever $z \in (0, 4)$. Accordingly,

$$f_Z(z) = 2 \times \frac{1}{2\sqrt{1 - (1 - \frac{z}{2})^2}} \times \frac{1}{2\pi} = \frac{1}{2\pi\sqrt{z - \frac{z^2}{4}}}, \text{ for } z \in (0, 4).$$

Example 9 *Let $X_1, X_2, \ldots, X_n$ be i.i.d. $U[0,1]$ and $Y = max\{X_1, \ldots, X_n\}$. Calculate $E[X_1 \mid Y]$.*

Intuition suggests, and it is not too hard to justify, that if $Y = y$, then $X_1 = y$ with probability $1/n$, and with probability $(n-1)/n$ the random variable X_1 is uniformly distributed in $[0, y]$. Hence,

$$E[X_1 \mid Y] = \frac{1}{n}Y + \frac{n-1}{n}\frac{Y}{2} = \frac{n+1}{2n}Y.$$

Example 10 *Let X, Y be a pair of random variables. Find the value of a that minimizes the variance of $X - aY$.*

If we were allowed to choose (a, b) to minimize $E((X - aY - b)^2)$ we would pick $a = \text{cov}(X,Y)/\text{var}(Y)$. But

$$E((X - aY - b)^2) = \text{var}(X - aY - b) + (E(X - aY - b))^2 = \text{var}(X - aY) + (E(X - aY - b))^2.$$

Thus, (a, b) would be picked so that $E(X - aY - b) = 0$ and $\text{var}(X - aY)$ is minimized. Therefore, we know that the answer is $a = \text{cov}(X,Y)/\text{var}(Y)$.

Example 11 *Let $\{X_n, n \geq 1\}$ be i.i.d. $B(p)$. Assume that $g, h : \Re^n \to \Re$ have the property that if $\mathbf{x}, \mathbf{y} \in \Re^n$ are such that $x_i \leq y_i$ for $i = 1, \ldots, n$, then $g(\mathbf{x}) \leq g(\mathbf{y})$ and $h(\mathbf{x}) \leq h(\mathbf{y})$. Show, by induction on n, that $cov(g(\mathbf{X}), h(\mathbf{X})) \geq 0$ where $\mathbf{X} = (X_1, \ldots, X_n)$.*

The intuition is that $g(\mathbf{X})$ and $h(\mathbf{X})$ are large together and small together.

For $n = 1$ this is easy. We must show that $\text{cov}(g(X_1)h(X_1)) \geq 0$. By redefining $\tilde{g}(x) = g(x) - g(0)$ and $\tilde{h}x) = h(x) - h(0)$, we see that it is equivalent to show that $\text{cov}(\tilde{g}(X_1), \tilde{h}(X_1)) \geq 0$. In other words, we can assume without loss of generality that $g(0) = h(0) = 0$. If we do that, we need only show that

$$E(g(X_1)h(X_1)) = pg(1)h(1) \geq E(g(X_1))E(h(X_1)) = pg(1)ph(1)$$

which is seen to be satisfied since $g(1)$ and $h(1)$ are nonnegative and $p \leq 1$.

Assume that the result is true for n. Let $\mathbf{X} = (X_1, \ldots, X_n)$ and $V = X_{n+1}$. We must show that

$$E(g(\mathbf{X}, V)h(\mathbf{X}, V)) \geq E(g(\mathbf{X}, V)E(h(\mathbf{X}, V)).$$

We know, by the induction hypothesis, that

$$E(g(\mathbf{X}, i)h(\mathbf{X}, i)) \geq E(g(\mathbf{X}, i)E(h(\mathbf{X}, i)), \text{ for } i = 0, 1.$$

Assume, without loss of generality, that

$$E(g(\mathbf{X}, 0)) = 0.$$

Then we know that

$$E(g(\mathbf{X}, 1)) \geq 0 \text{ and } E(g(\mathbf{X}, 0)h(\mathbf{X}, 0)) \geq 0$$

and

$$E(h(\mathbf{X}, V)) \leq E(h(\mathbf{X}, 1)),$$

so that

$$\begin{aligned} E(g(\mathbf{X}, V))E(h(\mathbf{X}, V)) &= pE(g(\mathbf{X}, 1))E(h(\mathbf{X}, V)) \leq pE(g(\mathbf{X}, 1))E(h(\mathbf{X}, 1)) \\ &\leq pE(g(\mathbf{X}, 1)h(\mathbf{X}, 1)) \leq pE(g(\mathbf{X}, 1)h(\mathbf{X}, 1)) + (1-p)E(g(\mathbf{X}, 0)h(\mathbf{X}, 0)) \\ &= E(g(\mathbf{X}, V)h(\mathbf{X}, V)), \end{aligned}$$

which completes the proof.

3 Discrete Time Markov Chains

Recall that a sequence of random variables $\mathbf{X} = \{X_n, n \geq 0\}$ taking values in a countable set $\mathbf{S}$ is a Markov chain if

$$P[X_{n+1} = j \mid X_n = i, X_m, m \leq n-1] = P(i, j), \forall i, j \in \mathbf{S}, n \geq 0.$$

The key point of this definition is that, given the present value of X_n, the future $\{X_m, m \geq n+1\}$ and the past $\{X_m, m \leq n-1\}$ are independent. That is, the evolution of $\mathbf{X}$ starts afresh from X_n. In other words, the *state* X_n contains all the information that is useful for predicting the future evolution.

Example 12 *Let $\{X_n, n \geq 0\}$ be a Markov chain on $\mathbf{S}$ with transition probability matrix P and initial distribution π. Specify the probability space.*

The simplest choice is the *canonical* probability space defined as follows. $\Omega = \mathbf{S}^\infty$; $\mathcal{F}$ is the smallest σ-field that contains all the events of the form

$$\{\omega \mid \omega_0 = i_0, \ldots \omega_n = i_n\};$$

P is the σ-additive set function on $\mathcal{F}$ such that

$$P(\{\omega \mid \omega_0 = i_0, \ldots \omega_n = i_n\}) = \pi(i_0)P(i_0, i_1) \times \cdots \times P(i_{n-1}, i_n).$$

3.1 Recognizing a Markov Chain

It is essential to be able to determine whether a random sequence is Markov.

Example 13 *We flip a biased coin forever. Let $X_1 = 0$ and, for $n \geq 2$, let $X_n = 1$ if the outcomes of the n-th and $(n-1)$-st coin flips are identical and $X_n = 0$ otherwise. Is $\mathbf{X} = \{X_n, n \geq 1\}$ a Markov chain?*

Designate by Y_n the outcome of the n-th coin flip. Let $P(H) = p = 1 - q$. If $\mathbf{X}$ is a Markov chain, then

$$P[X_4 = 1 \mid X_3 = 1, X_2 = 1] = P[X_4 = 1 \mid X_3 = 1, X_2 = 0].$$

The left-hand side is

$$\begin{aligned} &P[(Y_1, Y_2, Y_3, Y_4) \in \{HHHH, TTTT\} \mid (Y_1, Y_2, Y_3) \in \{HHH, TTT\}] \\ &= \frac{P((Y_1, Y_2, Y_3, Y_4) \in \{HHHH, TTTT\})}{P((Y_1, Y_2, Y_3) \in \{HHH, TTT\})} = \frac{p^4 + q^4}{p^3 + q^3}. \end{aligned}$$

Similarly, the right-hand side is

$$\begin{aligned} &P[(Y_1, Y_2, Y_3, Y_4) \in \{THHH, HTTT\} \mid (Y_1, Y_2, Y_3) \in \{THH, HTT\}] \\ &= \frac{P((Y_1, Y_2, Y_3, Y_4) \in \{THHH, HTTT\})}{P((Y_1, Y_2, Y_3) \in \{THH, HTT\})} = \frac{qp^3 + pq^3}{qp^2 + pq^2} = p^2 + q^2. \end{aligned}$$

Algebra shows that the expressions are equal if and only if $p = 0.5$. Thus, if $\mathbf{X}$ is a Markov chain, $p = 0.5$. Conversely, if $p = 0.5$, then we see that the random variables $\{X_n, n \geq 2\}$ are i.i.d. $B(0.5)$ and $\mathbf{X}$ is therefore a Markov chain.

Example 14 *Show that a function of a Markov chain need not be a Markov chain.*

Here is a simple example. Let $X_n = (X_0 + n) mod 3$ where X_0 is uniformly distributed in $\{0, 1, 2\}$. That is, if $X_0 = 0$, then $(X_n, n \geq 0) = (0, 1, 2, 0, 1, 2, \ldots)$ whereas if $X_0 = 1$, then $(X_n, n \geq 0) = (1, 2, 0, 1, 2, \ldots)$, and similarly if $X_0 = 2$. Let $g(0) = g(1) = 5$ and $g(2) = 6$. Then $\{Y_n = g(X_n), n \geq 0\}$ is not a Markov chain. Indeed,

$$P[Y_2 = 6 \mid Y_1 = 5, Y_0 = 5] = 1 \neq P[Y_2 = 6 \mid Y_1 = 5] = \frac{1}{2}.$$

3.2 FSE

The *First Step Equations* are difference equations about some statistics of a Markov chain $\{X_n, n \geq 0\}$ that are derived by considering the different possible values of the first step, i.e., for X_1. In the notes we looked at the FSE for the probability of hitting 0 before B starting from A and for the average time until X_n hits either 0 or B. We look at some related examples.

Example 15 *A clumsy man tries to go up a ladder. At each step, he manages to go up one rung with probability p, otherwise he falls back to the ground. What is the average time he takes to go up to the n-th rung.*

Let $\beta(m)$ be the average time to reach the n-th rung, starting from the m-th one, for $m \in \{0, 1, 2, \ldots, n\}$. The FSE are

$$\begin{aligned}\beta(m) &= 1 + p\beta(m+1) + (1-p)\beta(0), \text{ for } m \in \{0, 1, \ldots, n-1\} \\ \beta(n) &= 0\end{aligned}$$

The first equation is of the form $\beta(m+1) = a\beta(m) + b$ with $a = 1/p$ and $b = -1/p - (1 - p)\beta(0)/p$. The solution is

$$\beta(m) = a^m \beta(0) + \frac{1-a^m}{1-a} b, m = 0, 1, \ldots, n.$$

Since $\beta(n) = 0$, we find that

$$a^n \beta(0) + \frac{1-a^n}{1-a} b = 0.$$

Substituting the values of a and b, we find

$$\beta(0) = \frac{1-p^n}{p^n - p^{n+1}}.$$

For instance, with $p = 0.8$ and $n = 10$, one finds $\beta(0) = 41.5$.

Example 16 *Consider a small deck of three cards 1, 2, 3. At each step, you take the middle card and you place it first with probability 1/2 or last with probability 1/2. What is the average time until the cards are in the reversed order 3, 2, 1?*

The possible states are the six permutations $\{123, 132, 312, 321, 231, 213\}$. The state transition diagram consists of these six states placed around a circle (in the order indicated), with a probability 1/2 of transition of one step clockwise or counterclockwise. Relabelling the states 1, 2, ..., 6 for simplicity, with 1 = 321 and 4 = 123, we write the FSE for the average time $\beta(i)$ from state i to state 1 as follows:

$$\begin{aligned}\beta(i) &= 1 + 0.5\beta(i+1) + 0.5\beta(i-1), i \neq 1 \\ \beta(1) &= 0.\end{aligned}$$

In these equations, the conventions are that 6 + 1 = 1 and 1 - 1 = 6. Solving the equations gives $\beta(1) = 0, \beta(2) = \beta(6) = 5, \beta(3) = \beta(5) = 8, \beta(4) = 9$. Accordingly, the answer to our problem is that it takes an average of 9 steps to reverse the order of the cards.

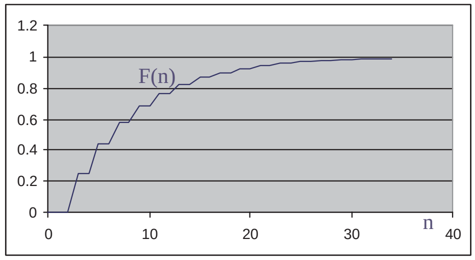

Figure 1: Graph of $F(n)$ in example 17

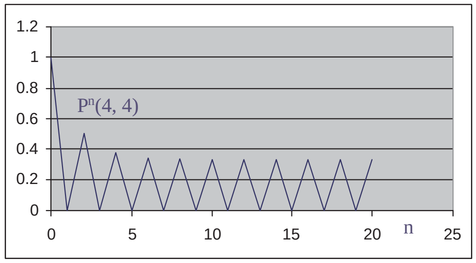

Figure 2: Graph of $P^n(4,4)$ in example 18

Example 17 *For the same Markov chain as in the previous example, what it the probability $F(n)$ that it takes at most n steps to reverse the order of the cards?*

Let $F(n;i)$ be the probability that it takes at most n steps to reach state 1 from state i, for $i \in \{1, 2, \ldots, 6\}$. The FSE for $F(n;i)$ are

$$\begin{aligned} F(n;i) &= 0.5F(n-1;i+1) + 0.5F(n-1;i-1), i \neq 1, n \geq 1 \\ F(n;1) &= 1, n \geq 0 \\ F(0;i) &= 1\{i = 1\}. \end{aligned}$$

Again we adopt the conventions that 6 + 1 = 1 and 1 - 1 = 6. We can solve the equations numerically and plot the values of $F(n) = F(4;n)$. The graph is shown in Figure 1.

Example 18 *Is the previous Markov chain periodic?*

Yes, it takes 2, 4, 6, ... steps to go from state i to itself. Thus, the Markov chain is periodic with period 2. Recall that this implies that the probability of being in state i does not converge to the invariant distribution $(1/6, 1/6, \ldots, 1/6)$. The graph in Figure 2 shows the probability of being in state 4 at time n given that $X_0 = 4$. This is derived by calculating $P(4,4)^n$. Since $P^{n+1}(4,j) = \sum_i P^n(4,i)P(i,j) = 0.5P^n(4,j-1) + 0.5P^n(4,j+1)$, one can compute recursively by iterating a vector with 6 elements instead of a matrix with 36.

Example 19 *We flip a fair coin repeatedly until we get either the pattern HHH or HTH. What is the average number of coin flips?*

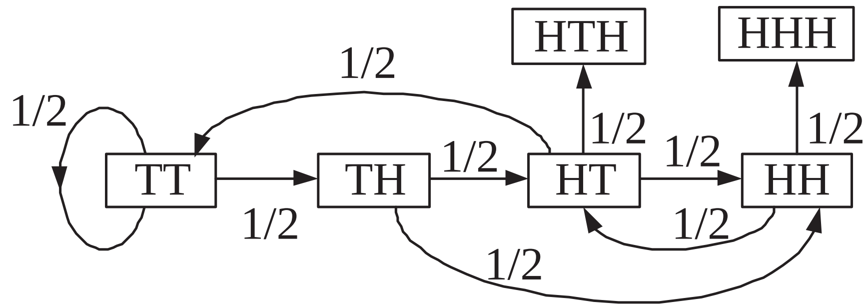

Figure 3: Transition diagram of X_n in example 19

Let X_n be the last two outcomes. After two flips, we start with X_0 that is equally likely to be any of the four pairs in $\{H, T\}^2$. Look at the transition diagram of Figure 3.

The FSE for the average time to hit one of the two states HTH or HHH from the other states are as follows:

$$\begin{aligned}\beta(TT) &= 1 + 0.5\beta(TT) + 0.5\beta(TH) \\ \beta(TH) &= 1 + 0.5\beta(HH) + 0.5\beta(HT) \\ \beta(HT) &= 1 + 0.5\beta(TT) \\ \beta(HH) &= 1 + 0.5\beta(HT)\end{aligned}$$

Solving these equations we find

$$(\beta(TT), \beta(HT), \beta(HH), \beta(TH)) = \frac{1}{5}(34, 22, 16, 24)$$

and the answer to our problem is then

$$2 + \frac{1}{4}(\beta(TT) + \beta(TH) + \beta(HT) + \beta(HH)) = \frac{34}{5} = 6.8.$$

4 Continuous Time Markov Chains

Recall that the random process $\mathbf{X} = \{X_t, t \geq 0\}$ taking values in the countable set $\mathbf{S}$ is a Markov chain with rate matrix $\mathbf{Q}$ and initial distribution π if

$$\begin{aligned}&P[X_{t+\epsilon} = j \mid X_t = i; X_s, s \leq t] = 1\{i = j\} + q(i,j)\epsilon + o(\epsilon), i, j \in \mathbf{S} \\ &P(X_0 = i) = \pi(i), i \in \mathbf{S}.\end{aligned}$$

In this definition, $q(i,j) \geq 0$ for $i \neq j$ and

$$q(i) := -q(i,i) = \sum_{j \neq i} q(i,j) < \infty, i \in \mathbf{S}.$$

Also, $\pi(i) \geq 0, i \in \mathbf{S}$ and $\sum_i \pi(i) = 1$.

The definition specifies the Markov property that given X_t the past and the future are independent. Recall that the Markov chain stays in state i for an exponentially distributed time with rate $q(i)$, then jumps to state j with probability $q(i,j)/q(i)$ for $j \neq i$, and the evolution continues in that way.

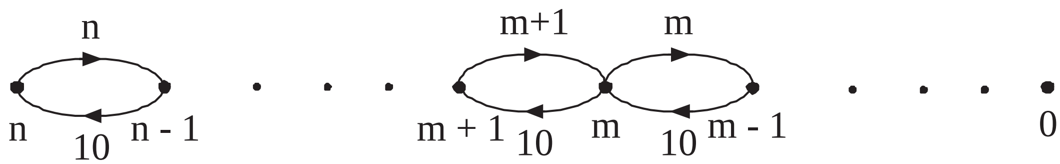

Figure 4: State transitions diagram in example 21

Example 20 *Consider n light bulbs that have independent lifetimes exponentially distributed with mean 1. What is the average time until the last bulb dies?*

Let X_t be the number of bulbs still alive at time $t \geq 0$. Because of the memoryless property of the exponential distribution, $\{X_t, t \geq 0\}$ is a Markov chain. Also, the rate matrix is seen to be such that

$$q(m) = q(m, m-1) = m, m \in \{1, 2, \ldots, n\}.$$

The average time in state m is $1/m$ and the Markov chain goes from state n to $n-1$ to $n-2$, and so on until it reaches 0. The average time to hit 0 is then

$$\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{3} + \frac{1}{2} + 1.$$

To fix ideas, one finds the average time to be about 3.6 when $n = 20$.

Example 21 *In the previous example, assume that the janitor replaces a burned out bulb after an exponentially distributed time with mean 0.1. What is the average time until all the bulbs are out?*

The rate matrix now corresponds to the state diagram shown in Figure 4. Defining $\beta(m)$ as the average time from state m to state 0, for $m \in \{0, 1, \ldots, n\}$, we can write the FSE as

$$\begin{aligned}
\beta(m) &= \frac{1}{m+10} + \frac{m}{m+10}\beta(m-1) + \frac{10}{m+10}\beta(m+1), \text{ for } m \in \{1, 2, \ldots, n-1\} \\
\beta(n) &= \frac{1}{n} + \beta(n-1) \\
\beta(0) &= 0.
\end{aligned}$$

If we knew $\beta(n-1)$, we could solve recursively for all values of $\beta(m)$. We could then check that $\beta(0) = 0$. Choosing $n = 20$ and adjusting $\beta(19)$ so that $\beta(0) = 0$, we find numerically that $\beta(20) \approx 2,488$.