Chapter 9 Additional Topics in Probability

9.1 The Poisson process

$N(t)$: The number of events that occur in the time interval $[0, t]$.

The collection of random variables $\{N(t),\ t \geq 0\}$ is said to be a *Poisson process* having rate $\lambda,\ \lambda > 0$ if

(i) $N(0) = 0$

(ii) The number of events that occur in disjoint time intervals are independent.

(iii) The distribution of the number of events that occur in a given interval depends only on the length of that interval and not on its location.

(iv) $P\{N(h) = 1\} = \lambda h + o(h)$

(v) $P\{N(h) \geq 2\} = o(h)$

Lemma 1.1: For a Poisson process with rate λ,

$$P\{N(t) = 0\} = e^{-\lambda t}$$

$\{T_n, n = 1, 2, \ldots\}$: Sequence of interarrival times.

Proposition 1.1: $T_1, T_2, \ldots$ are independent exponential random variables each with mean $1/\lambda$.

- $S_n = \sum_{i=1}^{n} T_i$: Waiting time until the nth event.
- Probability density of S_n:

$$f_{S_n}(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} \quad x \geq 0$$

Theorem 1.1: For a Poisson process with rate λ,

$$P\{N(t) = n\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

- $N(t) \geq n \iff S_n \leq t$

$$\begin{aligned} P\{N(t) = n\} &= P\{N(t) \geq n\} - P\{N(t) \geq n+1\} \\ &= P\{S_n \leq t\} - P\{S_{n+1} \leq t\} \\ &= \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} dx - \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^n}{n!} dx \\ &= \frac{e^{-\lambda t}(\lambda t)^n}{n!} \end{aligned}$$

9.2 Markov chains

- Consider a sequence of random variables $X_1, X_2, \ldots, X_n$ and suppose that the set of possible values of these random variables is $\{0, 1, \ldots, M\}$.
- The sequence of random variables is said to form a Markov chain if each time the system is in the state i there is some fixed probability–call it P_{ij}–that it will next be in state j.

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = P_{ij}$$

- Transition probabilities of the Markov chain:

$$P_{ij} \geq 0 \qquad \sum_{j=0}^{M} P_{ij} = 1 \quad i = 1, 2, \ldots, M$$

- Transition probability matrix:

$$\begin{pmatrix} P_{00} & P_{01} & \cdots & P_{0M} \\ P_{10} & P_{11} & \cdots & P_{1M} \\ \vdots & & & \\ P_{M0} & P_{M1} & \cdots & P_{MM} \end{pmatrix}$$

Example 9.2a. Suppose that whether or not it rains tomorrow depends on previous weather conditions only through whether or not it is raining today. Suppose further that if it is raining today, then it will rain tomorrow with probability α, and, if it is not raining today, then it will rain tomorrow with probability β.

- If we say that the system is in state 0 when it rains and state 1 when it does not, then the system above is a two-side Markov chain having transition probability matrix

$$\begin{vmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{vmatrix}$$

- $P_{00} = \alpha = 1 - P_{01}, P_{10} = \beta = 1 - P_{11}$.

Example 9.2b. Consider a gambler who at each play of the game either wins 1 unit with probability p or loses 1 unit with probability $1 - p$.

- If we suppose that the gambler will quit playing when his fortune hits either 0 or M, then the gambler's sequence of fortunes is a Markov chain having transition probabilities

$$P_{i,i+1} = p = 1 - P_{i,i-1} \quad i = 1, \ldots, M - 1$$
$$P_{00} = P_{MM} = 1$$

Example 9.2c. The physicists P. and T. Ehrenfest considered a conceptual model for the movement of molecules in which M molecules are distributed among 2 urns. At each time point one of the molecules is chosen at random and is removed from its urn and placed in the other one.

- If we let X_n denote the number of molecules in the first urn immediately after the nth exchange, then $\{X_0, X_1, \ldots\}$ is a Markov chain with transition probabilities

$$\begin{aligned} P_{i,i+1} &= \frac{M-i}{M} \quad 0 \le i \le M \\ P_{i,i-1} &= \frac{i}{M} \quad 0 \le i \le M \\ P_{ij} &= 0 \quad \text{if } |j-i| > 1 \end{aligned}$$

Proposition 2.1
TheChapman-Kolmogorov equations:

$$P_{ij}^{(n)} = \sum_{k=0}^{M} P_{ik}^{(r)} P_{kj}^{(n-r)} \quad \text{for all } 0 < r < n$$

Example 9.2d. *A random walk.* An example of a Markov chain having a countably infinite state space is the *random walk*, tracks a particle as it moves along a one-dimensional axis.

- Suppose that at each point in time the particle will move either one step to the right or one step to the left with respective probabilities p and $1 - p$.

- That is, suppose the particle's path follows a Markov chain with transition probabilities

$$P_{i,i+1} = p = 1 - P_{i,i-1} \quad i = 0, \pm 1, \ldots$$

- If the particle is at state i, then the probability it will be at state j after n transitions is the probability that $(n-i+j)/2$ of these steps are to the right and $n-[(n-i+j)/2] = (n+i-j)/2$ are to the left.

- As each step will be to the, independently of the other steps, with probability p, it follows that the above is just the binomial probability

$$P_{ij}^{n} = \binom{n}{(n-i+j)/2} p^{(n-i+j)/2} (1-p)^{(n+i-j)/2}$$

where $\binom{n}{x}$ is taken to equal 0 when x is not a nonnegative integer less than or equal to n.

- The above can be rewritten as

$$P^{2n}_{i,i+2k} = \binom{2n}{n+k} p^{n+k}(1-p)^{n-k} \quad k = 0, \pm 1, \ldots, \pm n$$

$$P^{2n+1}_{i,i+2k+1} = \binom{2n+1}{n+k+1} p^{n+k+1}(1-p)^{n-k}$$
$$k = 0, \pm 1, \ldots, \pm n, -(n+1)$$

Theorem 2.1: For an ergodic Markov chain

$$\Pi_j = \lim_{n\to\infty} P^{(n)}_{ij}$$

exits, and the Π_j, $0 \leq j \leq M$, are the unique nonnegative solutions of

$$\Pi_j = \sum_{k=0}^{M} \Pi_k P_{kj}$$

$$\sum_{j=0}^{M} \Pi_j = 1$$

Example 9.2e. Consider Example 9.2a, in which we assume that if it rains today, then it will rain tomorrow with probability α; and, if it does not rain today, then it will rain tomorrow with probability β.

- From Theorem 2.1 it follows that the limiting probabilities of rain and of no rain, π_0 and π_1, are given by

$$\begin{aligned}\Pi_0 &= \alpha\Pi_0 + \beta\Pi_1 \\ \Pi_1 &= (1-\alpha)\Pi_0(1-\beta)\Pi_1 \\ \Pi_0 + \Pi_1 &= 1\end{aligned}$$

which yields

$$\Pi_0 = \frac{\beta}{1+\beta-\alpha} \quad \Pi_1 = \frac{1-\alpha}{1+\beta-\alpha}$$

- For instance, if $\alpha = .6$, $\beta = .3$, then the limiting probabilities of rain on the nth day is $\Pi_0 = \frac{3}{7}$.

Example 9.2f. Suppose in Example 9.2c that we are interested in the proportion of time there are j molecules in urn 1, $i = 0, \ldots, M$.

- By Theorem 2.1 these quantities will be the unique solution of

$$\Pi_0 = \Pi_1 \times \frac{1}{M}$$

$$\begin{aligned}
\Pi_j &= \Pi_{j-1} \times \frac{M-j+1}{M} + \Pi_{j+1} \times \frac{j+1}{M} \quad j = 1, \ldots, M \\
\Pi_M &= \Pi_{M-1} \times \frac{1}{M} \\
\sum_{j=0}^{M} &= 1
\end{aligned}$$

- However, as it is easily checked that

$$\Pi_j = \binom{M}{j}(\frac{1}{2})^M \qquad j = 0, \ldots, M$$

satisfy the equations above, it follows that these are the long-run proportions of time that the Markov chain is in each of the states. (See Problem 11 for an explanation of how one might have guessed at the foregoing solution.)

9.3 Surprise, uncertainty, and entropy

- Consider an event E that can occur when an experiment is performed.
- How surprised would we be to hear that E does, in fact, occur?

- $P(E) = 1/36$ is more surprised than $P(E) = 1/2$.
- $S(p)$: The surprise evoked by the occurrence of an event having probability p.

Axiom 1:

$$S(1) = 0$$

Axiom 2: $S(p)$ is a strictly decreasing function of p; that is, if $p < q$, then $S(p) > S(q)$.

Axiom 3: $S(p)$ is a continuous function of p.

Axiom 4:

$$S(pq) = S(p) + S(q) \quad 0 < p, q \leq 1$$

Theorem 3.1: If $S(\cdot)$ satisfies Axioms 1 through 4, then

$$S(p) = -C \log_2 p$$

where C is an arbitrary positive integer.

- $S(p^2) = S(p) + S(p) = 2S(p)$
- $S(p^m) = mS(p)$
- $S(p^{1/n}) = \frac{1}{n}S(p)$
- $S(p^{m/n}) = \frac{m}{n}S(p)$
- $S(p^x) = xS(p)$
- Let $x = -\log_2 p$, then $S(p) = xS(1/2) = -C\log_2 p$

Proposition 3.1:

$$H(X, Y) = H(Y) + H_Y(X)$$

Lemma 3.1:

$$\ln x \leq x - 1 \quad x > 0$$

with equality only at $x = 1$.

Theorem 3.2:

$$H_Y(X) \leq H(X)$$

with equality if and only if X and Y are independent.

9.4 Coding Theory and Entropy

Lemma 4.1: Let X take on the possible values $x_1, \ldots, x_N$. Then, in order for it to be possible to encode the values of X in binary sequences (none of which is an extension of another) of respective lengths $n_1, \ldots, n_N$,it is necessary and sufficient and sufficient that

$$\sum_{i=1}^{N} (\frac{1}{2})^{n_j} \leq 1$$

Theorem 4.1 The noiseless coding theorem: Let X take on the values $x_1, \ldots, x_N$with respective probabilities $p(x_1), \ldots, p(x_N)$.Then, for any coding of X that assigns n_i bits to x_i,

$$\sum_{i=1}^{N} n_i p(x_i) \geq H(X) = -\sum_{i=1}^{N} p(x_i) \log p(x_i)$$

Example 9.4a. Consider a random variable X with probability mass function

$$p(x_1) = \frac{1}{2} \quad p(x_2) = \frac{1}{4} \quad p(x_3) = p(x_4) = \frac{1}{8}$$

Since

$$\begin{aligned} H(X) &= -[\frac{1}{2} \log \frac{1}{2} + \frac{1}{4} \log \frac{1}{4} + \frac{1}{4} \log \frac{1}{8}] \\ &= \frac{1}{2} + \frac{2}{4} + \frac{3}{4} \\ &= 1.75 \end{aligned}$$

it follows from Theorem 4.1 that there is on more efficient coding scheme than

$$\begin{aligned} x_1 &\leftrightarrow 0 \\ x_2 &\leftrightarrow 10 \end{aligned}$$

$$x_3 \leftrightarrow 110$$
$$x_4 \leftrightarrow 111$$

Example 9.4b. Suppose that 10 independent tosses of a coin, having probability p of coming up heads, are made at location A and the result is to be transmitted to location B. The outcome of this experiment is a random vector $X = (X_1, \ldots, X_10)$, where X_i is 1 or 0 according to whether or not the outcome of the ith toss is heads. By the results of this section it follows that L, the average number of bits transmitted by any code, satisfies

$$H(X) \leq L$$

with

$$L \leq H(X) + 1$$

for at least one code. Now, since the X_i are independent, it follows from Proposition 3.1 and Theorem 3.2 that

$$\begin{aligned} H(X) = H(X_1, \ldots, X_n) &= \sum_{i=1}^{N} H(X_i) \\ &= -10[p \log p + (1-p)\log(1-p)] \end{aligned}$$

If $p = \frac{1}{2}$, then $H(X) = 10$, and it follows that we can do no better than just encoding X by its actual value. That is, for example, if the first 5 tosses come up heads and the last 5 tails, then the message 1111100000 is transmitted to location B.

However, if $p \neq \frac{1}{2}$, we can often do better by using a different coding scheme. For instance, if $p = \frac{1}{4}$, then

$$H(X) = -10(\frac{1}{4}\log\frac{1}{4} + \frac{3}{4}\log\frac{3}{4}) = 8.11$$

and thus there is an encoding for which the average length of th encoding message is no greater than 9.11.

One simple coding that is more efficient, in this case, than the identity code id to break up $(X_1, \ldots, X_10)$ into 5 pairs of 2 random variables each and then code each of the pairs as follows:

$$X_i = 0, X_{i+1} = 0 \leftrightarrow 0$$
$$X_i = 0, X_{i+1} = 1 \leftrightarrow 10$$

$$X_i = 1, X_{i+1} = 0 \leftrightarrow 110$$
$$X_i = 1, X_{i+1} = 1 \leftrightarrow 111$$

for $i = 1, 3, 5, 7, 9$. The total message then transmitted is the successive encodings of the above pairs.
For instance, if the outcome $TTTHHTTTTH$ is observed, then the message 010110010 is sent.
The average number of bits needed to transmit the message using this code is

$$5[1(\frac{3}{4})^2 + 2(\frac{1}{4})(\frac{3}{4}) + 3(\frac{1}{4})(\frac{3}{4}) + 3(\frac{1}{4})^2] = \frac{135}{16} \approx 8.44$$

Theorem 4.2 The noisy coding theorem: There is number C such that for any value R which is less than C, and any $\varepsilon > 0$, there exist a coding-decoding scheme that transmits at an average rate of R bits sent per signal and with an error (per bit) probability of less than ε. The largest such value of C, call it C^*, † is called the channel capacity, and for the binary symmetric channel,

$$C^* = 1 + p \log p + (1 - p) \log(1 - p)$$

† For an entropy interpretation of C^*, see Problem 18.