I. Conditional Expectation

The conditional expectation of Y given X is defined by

$$E[Y|X=x] = \sum_y yP[Y=y|X=x] = \sum_y y\frac{P(X=x,Y=y)}{P(X=x)}.$$

Prove E[E[Y|X]] = E[Y]

Prove E[h(X)Y|X] = h(X)E[Y|X]

II. LLSE

Linear Least Squares Estimate (LLSE)

$$L[Y|X] = E(Y) + \frac{cov(X,Y)}{var(X)}(X - E(X)).$$

Let X, Y be i.i.d. Uniform(-1,1). Calculate L(Y|Y+2X). [Var(X) = 2Var(U(0,1))]

Let X, Y, Z be i.i.d. N(0,1) and V= 2X + 3Y + 4Z, W= X+Y+Z. Find L[V|W].

III. Markov Chains

Definitions

P is a ***transition probability matrix*** if:

1. All of the entries are non negative
2. The sum of entries in each row is 1

A ***Markov chain*** is defined by four things: $(\mathscr{X}, \pi_0, P, \{X_n\}_{n=0}^{\infty})$

$\mathscr{X}$	Set of states
π_0	Initial probability distribution
P	Transition probability matrix
$\{X_n\}_{n=0}^{\infty}$	Sequence of random variables where:

$$Pr[X_0 = i] = \pi_0(i), i \in \mathscr{X}$$
$$Pr[X_{n+1} = j \mid X_n = i, X_{n-1}, \ldots, X_0] = P(i,j), \forall n \geq 0, \forall i, j \in \mathscr{X}$$

A Markov chain is ***irreducible*** if we can go from any state to any other state, possibly in multiple steps

Define value ***d(i)*** for each state i as:

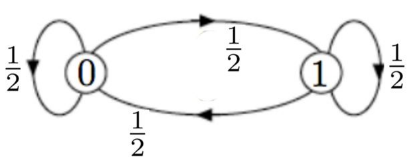
$$d(i) := g.c.d\{n > 0 \mid P^n(i,i) = Pr[X_n = i | X_0 = i] > 0\}, i \in \mathscr{X}$$

If d(i) = 1, then the Markov chain is ***aperiodic***. If d(i) ≠ 1, then the Markov chain is periodic and its ***period*** is d(i).

A distribution π is ***invariant*** if $\pi P = \pi$.

Theorem 24.3: A finite irreducible Markov chain has a unique invariant distribution
Theorem 24.4: All irreducible and aperiodic Markov chains converge to the unique invariant distribution. If a Markov chain is finite and reducible, the amount of time spent in each state approaches the invariant distribution as n grows large

Equations that model what will happen at the next step are called ***first step equations***



Denote β(i, j) as the expected amount of time it would take to move from i to j.

β(0, 1) = 1 + ½ β(0, 1)
β(1, 1) = 0

Life of Alex

Alex is enjoying college life. She spends a day either studying, partying, or looking for housing for the next year. If she is studying, the chances of her studying the next day are 30%, the chances of her partying the next day are 50%, and the chances of her looking for housing the next day are 20%. If she is partying, the chances of her partying the next day are 10%, the chances of her studying the next day are 60%, and the chances of her looking for housing the next day are 30%. If she is looking for housing, the chances of her looking for housing the next day are 50%, the chances of her partying the next day are 30% and the chances of her studying the next day are 20%.

a) Draw a Markov chain to visualize Alex's life.

b) Write out a matrix to represent this Markov chain

c) If Alex studies on Monday, what is the chance that she is partying on Friday? (Don't do the math, just write out the expression that you would use to find it.)

d) What percentage of her time should Alex expect to use looking for housing?

e) If Alex parties on Monday, what is the chance of Alex partying again before studying?

Prehistoric States

A prehistoric civilization survives by hunting game in the forests near their home. At the beginning of the hunting season, all the young men go out to the forest. After the first day, those who have a kill, which happens with probability 1/2, return home. Everyone who has been out for two days, even if without a kill, returns home for rest. And everyone who goes home goes back out the next day.

a. What are the states in this scenario? Draw a Markov chain.

b. What is the transition matrix? The initial vector?

c. Is this Markov chain reducible? Is it periodic?

d. What is the invariant vector?

e. What are the distributions after one week?

f. What is the expected length of hunting trip?

Stanford Cinema

You have a database of an infinite number of movies. Each movie has a rating that is uniformly distributed in $\{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$ independent of all other movies. You want to find two movies such that the sum of their ratings is greater than 7.5 (7.5 is not included).

a. A Stanford student chooses two movies each time and calculates the sum of their ratings. If is less than or equal to 7.5, the student throws away these two movies and chooses two other movies. The student stops when he/she finds two movies such that the sum of their ratings is greater than 7.5. What is the expected number of movies that this student needs to choose from the database?

b. A Berkeley student chooses movies from the database one by one and keeps the movie with the highest rating. The student stops when he/she finds the sum of the ratings of the last movie that he/she has chosen and the movie with the highest rating among all the previous movies is greater than 7.5. What is the expected number of movies that the student will have to choose?

Bet On It

Smith is in jail and has 3 dollars; he can get out on bail if he has 8 dollars. A guard agrees to make a series of bets with him. If Smith bets A dollars, he wins A dollars with probability 0.4 and loses A dollars with probability 0.6.

a. Find the probability that he wins 8 dollars before losing all of his money if he bets 1 dollar each time.

b. Find the probability that he wins 8 dollars before losing all of his money if he bets, each time, as much as possible but not more than necessary to bring his fortune up to 8 dollars

c. Which strategy gives Smith the better chance of getting out of jail?

Tossing Coins
A fair coin is tossed repeatedly and independently. Find the expected number of tosses till the pattern HTH appears.

Two Chains

Show that if P is the transition matrix of an irreducible chain with finitely many states, then $Q := (1/2)(I + P)$ is the transition matrix of an irreducible and aperiodic chain.

Show that P and $(1/2)(I + P)$ have the same stationary distributions. Discuss, physically, how the two chains are related.

IV. Continuous

Given the following density functions, identify if they are valid random variables. If yes, find the expectation and variance. If no, what rules does the variable violate?

1. f(x) = ¼ on (½, 9/2), = 0 elsewhere

2. f(x) = x - ½ on (0, $$\infty$$)

For a discrete random variable X we have Pr(X within [a, b]) that we can calculate directly by finding how many points in the probability space fall in the interval and how many total points are in the probability space. How do we find Pr(X within [a, b]) for a continuous random variable?

Are there any values of a, b for which we have a valid pdf? If not, why? If yes, what values?

f(x) = -1 $a<x<b$

f(x) = 0 $a<x<b$ (Are there any values of a,b for which we have a valid pdf?)

f(x) = 10000, $0<x<a$ (Are there any values of a for which we have a valid pdf?)

For what values of the parameters are the following functions probability density functions? What is the expectation and variance of the random variable that the function represents?

$f(x) = ax$, $0<x<1$, $f(x) = 0$ otherwise

$f(x) = -2x$, $a<x<b$, ($a=0$ OR $b=0$), $f(x) = 0$ otherwise

$f(x) = c$, $-30<x<-20$, $-5<x<5$, $60<x<70$, $f(x) = 0$ otherwise

Define a continuous random variable R as follows: we pick a random point on a disk of radius 1; the value of R is distance of this point from the center of the disk. We will find the probability density function of this random variable.

1. What is (should be) the probability that R is between 0 and ½? Why?

2. What is (should be) the probability that R is between *a* and *b*, for any $0 \le a \le b \le 1$?

3. What is a function f(*x*), for which $\int_a^b f(x)dx$ satisfies these same probabilities?

4. Define *g*(*x*), the probability density function for R.

V. Distributions

There are certain jellyfish that don't age called hydra. The chances of them dying is purely due to environmental factors, which we'll call *lambda*. On average, 2 hydras die within 1 day.
What is the probability you have to wait at least 5 days for a hydra dies?

(a) Let X and Y be two independent *discrete* random variables. Derive a formula for expressing the distribution of the sum $S = X + Y$ in terms of the distributions of X and of Y.

(b) Use your formula in part (a) to compute the distribution of $S = X + Y$ if X and Y are both discrete and uniformly distributed on $\{1,...,K\}$.

(c) Suppose now X and Y are *continuous* random variables with densities f and g respectively (X,Y still independent). Based on part (a) and your understanding of continuous random variables, give an educated guess for the formula of the density of $S = X + Y$ in terms of f and g.

(d) Use your formula in part (c) to compute the density of S if X and Y have both uniform densities on $[0, a]$.

(e) Show that if X and Y are independent normally distributed variables, then $X + Y$ is also a normally distributed variable.

VI. Joint Distributions

1. Suppose X and Y are independent and X has probability density function $g(x) = 6x(1-x)$ for $0 <= x <=1$, and that Y has probability density function $h(y) = 12y^2(1-y)$ for $0 <= y <=1$.
    a. Find the probability density function of (X,Y)

    b. Find $P(X+Y <= 1)$

2. Suppose that (X,Y) has probability $f(x,y) = 6x^2y$ for x and y in [0,1].
    a. Find the probability of density function of X.

    b. Find the pdf.

    c. Are X and Y independent.

3. Choose m real numbers at random between [0,1]. Let X be the largest one of these numbers. Find the pdf of f(x) and the E[X].

4. Intuitively, find the value that minimizes (E[(X-a)^2]), then prove your answer.