Probabilistic Modeling and Statistical Computing Fall 2015

September 29, 2015

Markov Chain as Bayesian Network

Consider Bayesian network

$$X_0 \to X_1 \to X_2 \to X_3 \to X_4 \to$$

where all the X_i have the same range ("state space") and conditional distribution of X_i depends only on values of X_{i-1}.

Markov Property:

$$\mathcal{P}(X_i = x | X_{i-1}, X_{i-2}, X_{i-3} \dots) = \mathcal{P}(X_i = x | X_{i-1})$$

Think of i as time. To assess distribution of X_i, only need X_{i-1}, not the more distant past.

Example: Simple Random Walk

- Each X_i has range $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$
- Start at 0: $\mathcal{P}(X_0 = 1) = 1$
- Take steps to left or right with equal probability $\frac{1}{2}$

$$\mathcal{P}(X_{i+1} = x + 1 | X_i = x) = 1/2$$
$$\mathcal{P}(X_{i+1} = x - 1 | X_i = x) = 1/2$$

We can also write this as

$$X_1 = X_0 + Y_1, \quad X_2 = X_1 + Y_2, \ldots$$

where the Y_i are independent, $\mathcal{P}(Y_i = \pm 1) = 1/2$.

Example: Roulette

Roulette: Always bet on Red

- Start with $X_0 = \$200$. Always bet \$1 on Red. X_i = amount after i bets.
- Win \$1 with $\mathcal{P} = 18/38$, lose \$1 with $\mathcal{P} = 20/38$.

$$\mathcal{P}(X_{i+1} = x + 1 | X_i = x) = 18/38$$
$$\mathcal{P}(X_{i+1} = x - 1 | X_i = x) = 20/38$$

Stop when you have lost all your money or when you have reached \$1,000.

Simulation and Analysis

Simulate this.
We can write

$$X_1 = X_0 + Y_1, \quad X_2 = X_1 + Y_2, \ldots$$

where the Y_i are independent,
$\mathcal{P}(Y_i = 1) = 18/38, \mathcal{P}(Y_i = -1) = 20/38.$
Thus

$$\mathcal{E}(Y_i) = 18/38 \cdot 1 + 20/38 \cdot (-1) = -1/19$$
$$\mathcal{E}(X_i) = \mathcal{E}(X_{i-1}) - 1/19$$

What does this mean?

St. Petersburg System

- Start with $X_0 = \$200$. Always bet on Red, starting with \$1.
- After a loss, double your last bet.
- After a win, bet \$1.

Stop when you have lost all your money or when you have reached \$1,000.

What is the rationale?

What is the right state space? How to simulate this?

Time-Homogeneous Markov Chain

Suppose the state space = range of each X_i is finite and the same, $\mathcal{R} = \{1, 2, \ldots, n\}$.
Need to specify all

$$\mathcal{P}(X_{i+1} = k | X_i = j), 1 \leq j, k \leq n$$

For each i, this is a $n \times n$ matrix. **Now assume this is independent of i.**
Transition matrix, probability of $j \rightsquigarrow k$

$$P_{jk} = \mathcal{P}(X_{i+1} = k | X_i = j), 1 \leq j, k \leq n$$

Example: Joe's Diet

Joe only eats pizza, chocolate ice cream, and kimchi. He eats once a day.

- On days after eating kimchi, he always eats ice cream.
- On days after ice cream, he eats pizza or ice cream with equal probability.
- On days after pizza, he eats kimchi 20% of the time and another pizza 80% of the time.

Set up the transition matrix!

Diet Transition Matrix

State space:

$$\mathcal{R} = \{\textit{kimchi} \simeq 1,\ \textit{ice cream} \simeq 2,\ \textit{pizza} \simeq 3\}$$

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & .5 & .5 \\ .2 & 0 & .8 \end{pmatrix}$$

Today Joe had pizza. What is the probability that he will eat ice cream in three days?

$$\mathcal{P}(X_3 = 2 | X_0 = 3)$$

Transition

Suppose P is the transition matrix of a time homogeneous Markov chain. Then for each i, j, k

$$\boxed{\mathcal{P}(X_{i+2} = k | X_i = j) = \left(P^2\right)_{jk}} \, .$$

More generally

$$\boxed{\mathcal{P}(X_{i+r} = k | X_i = j) = \left(P^r\right)_{jk}} \, .$$

for $r = 1, 2, 3, \ldots$.

Proof for $r = 2$

Suppose $X_i = j$ (fixed). Then

$$\mathcal{P}(X_{i+1} = \ell) = \mathcal{P}(X_{i+1} = \ell | X_i = j) = P_{j\ell}$$

$$\mathcal{P}(X_{i+2} = k | X_{i+1} = \ell) = P_{\ell k}$$

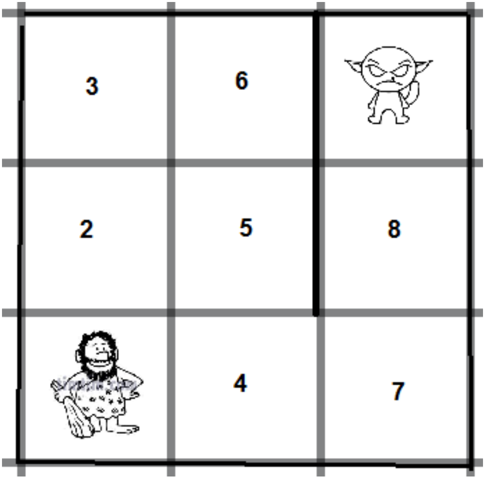
$$\mathcal{P}(X_{i+2} = k | X_i = j) = P_{j1}P_{1k} + P_{j2}P_{2k} + P_{j3}P_{3k} + \cdots$$

The right hand side is the (j, k) entry of P^2 by definition of the matrix product.

Joe's diet matrix:

$$P^3 = \begin{pmatrix} .1 & .25 & .65 \\ .13 & .225 & .645 \\ .128 & .260 & .612 \end{pmatrix}$$

Another random walk



Zog the caveman explores a new cave, starting in room 1 (lower left).
Every minute, he either stays in the current room (probability 1/2) or moves into one of the directly adjacent rooms with equal probability.
He cannot walk through walls.
There is a monster in room 9 that will eat him if he enters the room.

Set up the transition matrix and find the probability that he is still alive after half an hour.

Web surfing as random walk

- State space = set of all Internet websites. *How many states are there?*
- State of a random web surfer = the site she is looking at.
- Transitions: with large probability $1 - D$, click on one of the links on the page, all equally likely. With remaining probability D, jump to any other site on the web (i.e., get bored).

PageRank (US patent 6285999 B1) = Google's method for estimating long-term probabilities of visits to sites. *How can this be monetized?*

Example 1: Random Meeting

You and your friend Zoe have agreed to meet somewhere between 1 PM and 2 PM. Each of you will arrive at random (uniform distribution) and independently. Find the distribution of the time that you have to wait for your friend.

Model

Use two independent random variables $X_{me}, X_{Zoe} \sim U(1, 2)$ (arrival times).
We want the distribution of

$$\boxed{W = \max(X_{Zoe} - X_{me}, 0)}\ .$$

Example 2: Life of Crime

Brianna is a serial burglar who breaks into random houses to steal antique dolls. She risks being caught with probability p (small). If she is caught, she goes to prison for life. Otherwise she finds a doll with probability q. Find the distribution of the number of dolls that she steals.

Model

Use a geometric random variable N, success rate p, for the number of break-ins until she is caught. We want the distribution of $X \sim B(N, q)$. *Note that N is random!*

Example 3: Mattress Store

Customers arrive at a mattress store and buy one of k mattresses, with probability p_i for mattress i. They might also not buy any mattress, probability p_0.
Suppose n customers arrive independently. Let X_i be the number of mattresses i that is sold, and X_0 be the number of customers who don't buy anything. Find the distribution of $(X_0, X_1, \ldots, X_k)$.

Model

A new model is needed!

Discrete Multivariate Distribution

Assume that $(X_1, X_2, \ldots, X_n)$ is an n-dim random vector with finitely or countably many possible values. The joint probability mass function (pmf) is the function

$$f(x_1, \ldots, x_n) = \mathcal{P}(X_1 = x_1, \ldots, X_n = x_n)$$

defined for all possible vectors of values $(x_1, x_2, \ldots, x_n)$.

Mattress Store Example

Suppose there are $k = 2$ types of mattresses and therefore three probabilities p_0, p_1, p_2 that add up to 1.
In the case of one customer, the joint pmf for (X_0, X_1, X_2) is

$$\mathcal{P}((1,0,0)) = p_0$$
$$\mathcal{P}((0,1,0)) = p_1$$
$$\mathcal{P}((0,0,1)) = p_2$$

Mattress Store Example

Suppose there are $k = 2$ types of mattresses, probabilities p_0, p_1, p_2.
Some values of the joint pmf for (X_0, X_1, X_2) in the case of $N = 3$ customers:

$$\mathcal{P}((3,0,0)) = 1 \cdot p_0^3$$
$$\mathcal{P}((1,2,0)) = 3 \cdot p_0 p_1^2$$
$$\mathcal{P}((1,1,1)) = 6 \cdot p_0 p_1 p_2$$

The factors 1, 3, 6 count the number of ways in which these events can occur.

Burglar Example

The number N of successful break-ins has a geometric distribution, $\mathcal{P}(N = n) = p(1-p)^n$. The number of stolen dolls given N has a binomial distribution,

$$\mathcal{P}(X = k | N = n) = \binom{n}{k} q^k (1-q)^{n-k}$$

The joint pmf is the product:

$$\mathcal{P}(X = k, N = n) = \binom{n}{k} q^k (1-q)^{n-k} \cdot p \cdot (1-p)^n$$

This is understood to be zero if $k > n$.

Marginal Distributions

Assume that $(X_1, X_2, \ldots, X_n)$ is an n-dim random vector with joint pmf $f(x_1, \ldots, x_n)$. The marginal distribution of X_k is

$$\mathcal{P}(X_k = x_k) = \sum_{x_i,\, i \neq k} f(x_1, x_2, \ldots, x_n)\,.$$

*Sum over all possible values of all the **other** components of the random vector.*

Partial observation X_k, not complete observation $(X_1, \ldots, X_n)$.

Burglar Example

Joint pmf:

$$\mathcal{P}(X = k, N = n) = \binom{n}{k} q^k (1-q)^{n-k} \cdot p \cdot (1-p)^n$$

Marginal pmf of N:

$$\mathcal{P}(N = n) = \sum_{k=0}^{n} \binom{n}{k} q^k (1-q)^{n-k} \cdot p \cdot (1-p)^n$$
$$= p \cdot (1-p)^n$$

Binomial distribution. We already knew that.

Burglar Example

Marginal pmf of X:

$$\mathcal{P}(X = k) = \sum_{n=0}^{\infty} \binom{n}{k} q^k (1-q)^{n-k} \cdot p \cdot (1-p)^n$$
$$= \frac{pq^k(1-p)^k}{(p+q-pq)^{k+1}}$$

Geometric distribution, success rate $p/(p+q-pq)$. Then $\mathcal{E}(X) = \frac{1-p}{p} \cdot q$.

Mattress Store Example:

Marginal pmf of X_1 for $k = 3$ customers:

$$\begin{aligned}\mathcal{P}(X_1 = 0) &= \mathcal{P}(3,0,0) + \mathcal{P}(2,0,1) \\ &\quad + \mathcal{P}(1,0,2) + \mathcal{P}(0,0,3 \\ &= 1 - 3p_1 + 3p_1^2 - p_1^3 \\ &= (1 - p_1)^3\end{aligned}$$

The marginal distribution of X_1 is $B(3, p_1)$.
What are the marginal distributions of X_0 & X_2?

Conditional Distribution

... for the case of two random variables X, Y with joint pmf $f_{XY}(x, y)$. Suppose the marginal distribution of X_1 has pmf $f_X(x)$.

Given x such that $\mathcal{P}(X = x) = f_X(x) > 0$, define the conditional pmf of $Y|X = x$ as

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}.$$

What does this mean in terms of actual probabilities?

Burglar Example

Find the pmf of the random variable X conditioned on $N = n$.

Recall that N tells us the number of break-ins without getting caught. If $N = n$, then X has a $B(n, q)$ distribution. So that is also the marginal distribution.

Verify this using the definition!

Burglar Example

Find the pmf of the random variable N conditioned on $X = k$.

The police find k stolen objects in Brianna's house. How many houses did she break into?

$$f_{NX}(n,k) = \binom{n}{k} q^k (1-q)^{n-k} \cdot p \cdot (1-p)^n$$

$$f_X(k) = \frac{pq^k(1-p)^k}{(p+q-pq)^{k+1}}$$

Then $f_{N|X}(n|k)$ is the ratio of these two terms.

Burglar Example

With some messy algebra, we obtain the conditional distribution of N, given X:

$$f_{N|X}(n|k) = \binom{n}{k} \frac{((1-q)(1-p))^n (p+q-pq)}{\left(\frac{(1-q)(1-p)}{p+q-pq}\right)^k}$$