Discrete Space Markov Chains

Matthew Wascher

January 15, 2025

Conditional Probability

Conditional probability is a key idea in the analysis of Markov chains, so we will briefly review. Let A and B be events. Then we define the probability of A given B as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

provided that $P(B) > 0$.

Conditional Probability

Let X and Y be two discrete random variables with probability mass functions (pmfs) $p(x)$ and $p(y)$ and let $p(x, y)$ denote their joint pmf. Then we can define the random variable $X|Y$ by the pmf

$$p(x|y) = \frac{p(x, y)}{p(y)}$$

for values of y such that $p(y) > 0$.

The following also holds:

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x)$$

Conditional Probability

Bayes Theorem is a useful tool for relating different conditional probabilities. It states that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

provided all the conditional probabilities are well defined. Why is this true?

$$P(A|B)\,P(B) = P(B|A)\,P(A)$$

$$\frac{P(A\cap B)}{\cancel{P(B)}}\,\cancel{P(B)} = \frac{P(B\cap A)}{\cancel{P(A)}}\,\cancel{P(A)}$$

Conditional Probability

Suppose that during nesting season a certain turtle lays X eggs where $X \sim \mathrm{Poisson}(\lambda)$. Each egg hatches independently with probability p. Call the number of eggs that hatch Y.

1. What is the conditional distribution of $Y|X$?
2. What is the marginal distribution of Y?
3. On average, how many hatched eggs will result from this turtle's nesting season? What is the variance?
4. What is the conditional distribution of $X|Y$?

Conditional Probability

What is the conditional distribution of $Y|X$?

$$Y \sim Bin(X, p)$$

$$p(y) = \binom{x}{y} p^y (1-p)^{x-y}$$

Conditional Probability

What is the marginal distribution of Y?

$$p(y) = \sum_{x=0}^{\infty} p(y|x)\,p(x) = \sum_{x=0}^{\infty} \binom{x}{y} p^{y} (1-p)^{x-y} \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= p^{y} \sum_{\substack{x=0 \\ x\geq y}}^{\infty} \frac{\cancel{x!}}{y!(x-y)!} (1-p)^{x-y} \frac{e^{-\lambda}\lambda^{x}}{\cancel{x!}}$$

$$= \frac{p^{y}}{y!} \sum_{\substack{x=0 \\ x\geq y}}^{\infty} \lambda^{y} e^{-\lambda p} \frac{e^{-\lambda(1-p)}\left[\lambda(1-p)\right]^{x-y}}{(x-y)!}$$

$$= \frac{(\lambda p)^{y} e^{-\lambda p}}{y!} \underbrace{\sum_{\substack{x=0 \\ x\geq y}}^{\infty} \frac{e^{-\lambda(1-p)}\left[\lambda(1-p)\right]^{x-y}}{(x-y)!}}_{=1}$$

$$p(y) = \frac{(\lambda p)^{y}\, e^{-\lambda p}}{y!} \Rightarrow \quad Y \sim \text{Poisson}(\lambda p)$$

Conditional Probability

On average, how many hatched eggs will result from this turtle's nesting season? What is the variance?

$Y \sim \text{Poisson}(\lambda p)$

$E(Y) = \lambda p$

$V(Y) = \lambda p$

Conditional Probability

What is the conditional distribution of $X|Y$?

$$p(x|y) = \frac{p(y|x)\,p(x)}{p(y)}$$

$$= \frac{\binom{x}{y} p^y (1-p)^{x-y} \frac{e^{-\lambda} \lambda^x}{x!}}{\frac{e^{-\lambda p} (\lambda p)^y}{y!}}$$

Conditional Probability

We can also find the mean without needing to find the marginal distribution of Y using the **law of total expectation**, which states that

$$E_Y(Y) = E_X(E_{Y|X}(Y|X))$$

Why is this true?

$$E[Y] = \sum_y y\,p(y) = \sum_y y \sum_x p(x,y)$$

$$= \sum_y \sum_x y\,p(y|x)\,p(x)$$

$$= \sum_x \sum_y y\,p(y|x)\,p(x)$$

$$= \sum_x p(x) \underbrace{\sum_y y\,p(y|x)}_{E[Y|X]}$$

$$= E_x[E(Y|X)]$$

Conditional Probability

We can now apply the law of total expectation to our example and compute $E(Y)$ without finding $p(y)$.

$Y \mid X \sim Bin(X, p)$

$X \sim Poisson(\lambda)$

$$E[Y] = E(E[Y \mid X)]$$

$$= E[Xp] = pE[X]$$

$$= \lambda p$$

Conditional Probability

We can similarly compute the variance without finding $p(y)$ using the **law of total variance** which states

$$V(Y) = E(V(Y|X)) + V(E(Y|X))$$

The law of total variance can be proved by applying the law of total expectation to the identity

$$V(Y) = E(Y^2) - E(Y)^2$$

Conditional Probability

We can now apply the law of total variance to our example and compute $V(Y)$ without finding $p(y)$.

$$V(Y) = E[V(Y|X)] + V(E[Y|X])$$

$$E[Xp(1-p)] + V(Xp)$$

$$(p - p^2)E[X] + p^2 V(X)$$

$$(p - \cancel{p^2})\lambda + \cancel{p^2}\lambda = \lambda p$$

Stochastic Processes

A **stochastic process** is an indexed family of random variables. In this course we will interpret the index as time (though in other contexts it could be something else, e.g. space).

For now we will focus on discrete time, so our stochastic processeses take the form

$$\{X_t\}_{t=0}^{\infty}$$

The set of possible values that can be taken on by $\{X_t\}_{t=0}^{\infty}$ is called the **state space**.

In general the X_t are not independent, but we will assume the underlying process that generates $\{X_t\}_{t=0}^{\infty}$ obeys the time ordering. That is, knowing $X_0, X_1, \dots, X_{t-1}$ must be sufficient information to fully characterize the distribution of X_t.

Stochastic Processes

A few example of stochastic processes include the following

1. Let $X_0 = 0$. Define $X_{t+1} = X_t + 1$ with probability .5 and $X_{t+1} = X_t - 1$ with probability .5 (Simple random walk)
2. Let $X_0 = 0$. Define $X_{t+1} = X_t + 1$ with probability $p := \frac{\exp\left(-\sum_{s=0}^{t} X_s\right)}{1+\exp\left(-\sum_{s=0}^{t} X_s\right)}$ and $X_{t+1} = X_t - 1$ with probability $1 - p$
3. Let $X_0 = 0$. Define $X_{t+1} = .5X_t + Z_{t+1}$ where $\{Z_t\}_{t=1}^{\infty}$ is a collection of iid N(0,1) random variables.
4. A signaling device sends out a signal of either 0 or 1. If the last two signals X_t and X_{t-1} were different, the device sends out the signal $X_{t+1} = |1 - X_t|$, otherwise it sends out the signal $X_{t+1} = X_t$.

The Markov Property

A stochastic process $\{X_t\}_{t=0}^{\infty}$ is said to have the **Markov property** if for every $t \in \mathbf{N}$ and every set A that can be assigned a probability

$$P(X_t \in A | X_{t-1}, \ldots, X_0) = P(X_t \in A | X_{t-1})$$

In the case where the state space of $\{X_t\}_{t=0}^{\infty}$ is discrete we can write this as

$$P(X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_0 = x_0) = P(X_t = x_t | X_{t-1} = x_{t-1})$$

Intuitively, a stochastis process is Markov if given the present, the distribution of the next step does not depend on the past. The Markov property is sometimes also called **memorylessness**.

The Markov Property

Which of our example stochastic processes have the Markov property?

1. Yes
2. No
3. Yes
4. No

Markov Processes

A stochastic process that satisfies the Markov property is called a **Markov process**. Well-known examples of Markov processes include

1. Random walks
2. Poisson processes
3. Renewal processes
4. Brownian motion

A Markov process on a finite or countably infinite state space is called a **Markov chain**.

Markov Chains

We will focus on the case where the state space is discrete and consider some examples:

Gambler's Ruin - Suppose a gambler has $\$N_0$. He plays a game where each round he wins \$1 with probability p and loses \$1 with probability $1 - p$. He stops when goes bankrupt or reaches $\$N$.

Does this satisfy the Markov property? Yes

Markov Chains

We will focus on the case where the state space is finite and consider some examples:

Ehrenfest chain - Suppose N balls are distributed between two urns, call them left and right. Let X_n be the number of balls in the left urn at step n. For step $n+1$, move a ball from the left to the right urn with probability X_n/N and move a ball from the right urn to the left urn with probability $(N - X_n)/N$

Does this satisfy the Markov property? Yes

Transition Matrix

Because the outcome of the next step in a Markov chain only depends on the present, we can summarize the chain's one-step behavior with a **transition matrix**. Given a k state Markov chain define

$$P := \{p(i,j)\}_{i,j \in \{1,\dots,k\}}$$

where $p(i,j) = P(X_{n+1} = j | X_n = i)$

Transition Matrix

Consider the Gambler's Ruin with $N = 5$ and $p = .4$. What is the transition matrix?

$$\begin{array}{cc|cccccc} & & & & j & & & \\ & & 0 & 1 & 2 & 3 & 4 & 5 \\ \hline i & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ & 1 & .6 & 0 & .4 & 0 & 0 & 0 \\ & 2 & 0 & .6 & 0 & .4 & 0 & 0 \\ & 3 & 0 & 0 & .6 & 0 & .4 & 0 \\ & 4 & 0 & 0 & 0 & .6 & 0 & .4 \\ & 5 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}$$

Transition Matrix

Consider the Ehrenfest Chain with $N = 4$. What is the transition matrix?

$$
\begin{array}{cc|ccccc}
 & & & & j & & \\
 & & 0 & 1 & 2 & 3 & 4 \\
\hline
i & 0 & 0 & 1 & 0 & 0 & 0 \\
 & 1 & 1/4 & 0 & 3/4 & 0 & 0 \\
 & 2 & 0 & 1/2 & 0 & 1/2 & 0 \\
 & 3 & 0 & 0 & 3/4 & 0 & 1/4 \\
 & 4 & 0 & 0 & 0 & 1 & 0
\end{array}
$$

Transition Matrix

The transition matrix has the following properties

1. $p(i,j) \geq 0$ for all i,j (because these are probabilities)
2. $\sum_j p(i,j) = 1$ (because given $X_n = i$ X_{n+1} must attain some state j)

While we can describe a Markov chain in words (as in our previous examples) we can also define a Markov chain by writing down a transition matrix. Any matrix satisfying 1. and 2. above gives rise to a Markov chain.

Transition Matrix

Here is an example where we define a Markov chain by a transition matrix. Suppose the stock market is either a bull market 1 or a bear market 2. Let $p(i, j)$ be

$$\begin{array}{ccc} & \mathbf{1} & \mathbf{2} \\ \mathbf{1} & .3 & .7 \\ \mathbf{2} & .2 & .8 \end{array}$$

How would one describe this Markov chain in words?

Examples: Branching process

We will now consider some widely applied examples of Markov chains.

A **branching process** models the evolution of a population over generations. Let X_n be the number of individuals in generation n. Each individual i has Z_i offspring independently according to a distribution where $P(Z_i = k) = p_k$ for $k = 0, 1, \ldots$. So given $X_n > 0$

$$X_{n+1} = \sum_{i=1}^{X_n} Z_i$$

Once $X_n = 0$ for some n, then $X_m = 0$ for all $m > n$.

Examples: Branching process

A natural question one might ask is whether the population is guaranteed to eventually go extinct. It turns out that this depends on the offspring distribution, in particular its mean.

1. If $E(Z_i) > 1$ then there is positive probability the population never goes extinct.
2. If $E(Z_i) < 1$ then the population eventually goes extinct with probability 1.
3. If $E(Z_i) = 1$ then the population eventually goes extinct with probability 1 unless $P(Z_i = 1) = 1$.

Examples: Branching process

While the proof of the survival regime is difficult, we can gain some insight into why the population goes extinct when $E(Z_i) < 1$ by using Markov's inequality, which says for a non-negative random variable X

$E(Z_i) = m$

$$P(X \geq a) \leq \frac{E(X)}{a}$$

$X_0 = 1$

To use this result, first we need to study $E(X_n)$

Examples: Branching process

$E[Z_i] = m \quad m < 1$

Let $X_0 = 1$. What is $E(X_n)$?

$E[E[X_n | X_{n-1}]] = E\left[\sum_{i=0}^{X_{n-1}} Z_i\right] = \sum_{i=0}^{X_{n-1}} E[Z_i] = \sum_{i=0}^{X_{n-1}} m = mX_{n-1}$

$E[mX_{n-1}] = mE[E[X_{n-1} | X_{n-2}]] = m^2 X_{n-2}$

$\ldots = m^n X_0 = m^n$

What happens when we apply Markov's inequality to $P(X_n \geq 1)$?

$P(X_n \geq 1) \leq \dfrac{E[X_n]}{1} = m^n$

$m^n \to 0 \quad \text{if} \quad 0 < m < 1$

Examples: Wright-Fisher Model

The Wright-Fisher model is a model of genetic drift in a population. Consider a population of N genes where each gene is either type A or type a. To generate the next generation, choose N individuals uniformly at random with replacement from the current generation and copy them. If we let X_n be the number of type A individuals in generation n, what is the transition matrix?

$$A \quad A \quad a \quad a \longrightarrow A \quad A \quad a \quad A$$

$$p_n = \frac{X_n}{N} \qquad X_{n+1} \sim \text{Binomial}(N, p_n)$$

Examples: Wright-Fisher Model

Note that if we ever see $X_n = N$ or $X_n = 0$ then the chain stays in that state forever. States with this property, $p(x, x) = 1$, are called **absorbing states**. (Another example is $X_n = 0$ in the branching process).

A natural question we can ask about the Wright-Fisher model is, given an initial condition, what is the probability that X_n reaches n and all individuals in the population are type A?

Another natural question is how long it takes until the chain reaches any absorbing state.

Examples: Wright-Fisher Model

Not all Markov chains have absorbing states. We can modify the Wright-Fisher model to have no absorbing states by allowing for mutation. Suppose there exist $u, v \in (0,1)$ such that an individual with a type A parent is type a with probability u and an individual with a type a parent is type A with probability v.

The chain no longer fixes at $X_n = N$ and $X_n = 0$ because even if all individuals are of one type, some offspring me be other the opposite type.

A natural question in this case is whether the proportion of type A individuals eventually converges to some equilibrium distribution

Example: Two stage Markov chain

Suppose that a basketball player's success shooting depends on the outcome of their previous two shots

- They have a $1/2$ probability of making the next shot if they missed the previous two
- They have a $2/3$ probability of making the next if they split the previous two
- They have a $3/4$ probability of making the next if they made the previous two

If we let X_n be the result of a single shot, then this is not a Markov chain because X_n depends on X_{n-1} and X_{n-2}.

Example: Two stage Markov chain

However, we can construct a Markov chain by modifying the state space and considering $Y_n = (X_n, X_{n-1})$. Let's write down the transition matrix for Y_n

$$
\begin{array}{cc|cccc}
 & & \multicolumn{4}{c}{(X_n, X_{n+1})} \\
 & & HH & HM & MH & MM \\
\hline
 & HH & 3/4 & 1/4 & 0 & 0 \\
(X_{n-1}, X_n) & HM & 0 & 0 & 2/3 & 1/3 \\
 & MH & 2/3 & 1/3 & 0 & 0 \\
 & MM & 0 & 0 & 1/2 & 1/2
\end{array}
$$