Probability and Stochastic Processes I I
Lecture 3a

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/stac62/staC632024.html

2024

III. Markov Chains

- consider probability model $(\Omega, \mathcal{A}, P)$ and $X : (\Omega, \mathcal{A}) \to (\mathcal{S}, \mathcal{C})$ as a *random object* so $X(\omega) \in \mathcal{S}, \mathcal{C}$ is a σ-algebra on $\mathcal{S}$ and $X^{-1}C \in \mathcal{A}$ for every $C \in \mathcal{C}$

- then $(\mathcal{S}, \mathcal{C}, P_X)$ is a probability model where $P_X(C) = P(X^{-1}C)$ (see PSPI notes)

- e.g., when $(\mathcal{S}, \mathcal{C}) = (\mathbb{R}^k, \mathcal{B}^k)$ then we call X a random vector (variable when $k = 1$)

- basically all the results we discussed for random vectors/variables also apply to random objects except those that depend on $\mathcal{S}$ being Euclidean

Example III.1

- $\mathcal{S} =$ set of all 2-dimensional similar triangles

- these "could" be as represented by the 3 angles as a (constrained and largest to smallest) point in $\mathbb{R}^3$ but what is the significance of the cdf in that case? ■

- assume all random objects X are defined on the same underlying probability space $(\Omega, \mathcal{A}, P)$ and refer to $\mathcal{S}$ as the *state space* (in PSPI we used the notation $\mathcal{X}$ for $\mathcal{S}$ but we'll follow the book)

- consider a stochastic process $\{X_t : t \in T\}$ where the set T can be totally ordered so that for $t_1, t_2 \in T$ we have one of $t_1 < t_2, t_1 = t_2$ or $t_1 > t_2$ being true

- for example, $T \subset \mathbb{R}^1$ and t corresponds to "time" but we can be more general (e.g., distance, vertices in a graph by number of connecting edges)

- suppose we observe the values of X_t for $t \leq t_0$ where we think of X_{t_0} as being the present "state" of the process

- then "roughly speaking" we call the process $\{X_t : t \in T\}$ a *Markov process (MP)* if for every $s_1 \geq 0, \ldots, s_n \geq 0$ and n, and $C \in \mathcal{C}^k$

$$P((X_{t_0+s_1}, \ldots, X_{t_0+s_n}) \in C \,|\, X_t : t \leq t_0) = P((X_{t_0+s_1}, \ldots, X_{t_0+s_n}) \in C \,|\, X_{t_0})$$

- in other words the probability distribution of the future given the present and past only depends on the present

- note - if in addition $P((X_{t_0+s_1}, \ldots, X_{t_0+s_n}) \in C \mid X_{t_0})$ only depends on the time differences from t_0, namely $s_1, \ldots, s_n$, then this is a *time homogeneous* (TH) Markov process, namely,

$$P((X_{t_0+s_1}, \ldots, X_{t_0+s_n}) \in C \mid X_{t_0})$$

is the same for every t_0, $C \in \mathcal{C}^n$

- there are Gaussian processes that are time homogeneous Markov processes (Brownian motion) but we restrict attention for now to situations where $\mathcal{S}$ and T are required to be countable (so various probability measures are discrete)

III.1 Definition and Basic Computations

Definition III.1 A stochastic process $\{X_t : t \in T\}$ is a discrete time, discrete space, time homogeneous *Markov chain* if:
(i) $\mathcal{S}$ is countable (typically represented as $\mathcal{S} \subset \mathbb{Z}, \mathcal{C} =$ power set),
(ii) T is countable (typically represented as $T = \mathbb{N}_0$),
(iii) there is an *initial probability measure* $P(X_0 = i) = \upsilon_i$ for $i \in \mathcal{S}$,
(iv) *Markov property (MP)* $P_{X_{n+1}}(\cdot \mid X_0, \ldots, X_n) = P_{X_{n+1}}(\cdot \mid X_n)$ with *transition probabilities* $p_{ij} = P(X_{n+1} = j \mid X_n = i)$ for every n.

Proposition III.1 For a MC the υ_i, p_{ij} determine all the finite dimensional distributions.

Proof:

$$
\begin{aligned}
& P(X_0 = i_0, \ldots, X_n = i_n) \\
= \; & P(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) P(X_n = i_n \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) \\
= \; & P(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) p_{i_{n-1} i_n} = \ldots = \upsilon_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}. \blacksquare
\end{aligned}
$$

- **note** - when we use the terminology "Markov chain" hereafter we will mean a s.p. that satisfies Definition III.1

- **note** - this gives a valid definition of a stochastic process (via Kolmogorov consistency and use $\sum_i \upsilon_i = 1$ and $\sum_j p_{ij} = 1$ for each i)

- **note** - when $\mathcal{S} \subset \mathbb{Z}, T = \mathbb{N}_0$ the initial probabilty distribution can be represented as a (possibly infinitely dimensional) row vector υ and the transition probabilities can be represented as a (possibly infinitely dimensional) matrix with (i, j)-th element p_{ij}, say $P = (p_{ij})$

- then

$$P(X_1 = j) \overset{TTP}{=} \sum_{i \in \mathcal{S}} P(X_0 = i) P(X_1 = j \mid X_0 = i) = \sum_{i \in \mathcal{S}} \upsilon_i p_{ij}$$

which is the product of the vector υ with the j-th column of P so

$$(\ldots, P(X_1 = j-1), P(X_1 = j), P(X_1 = j+1), \ldots) = \upsilon P$$

is the probability distribution of X_1

- also

$$
\begin{aligned}
&P(X_2 = j) \\
&\overset{TTP}{=} \sum_{i \in \mathcal{S}} P(X_2 = j \,|\, X_1 = i) P(X_1 = i) \\
&\overset{TH}{=} \sum_{i \in \mathcal{S}} p_{ij} P(X_1 = i) = \sum_{i \in \mathcal{S}} p_{ij} \sum_{k \in \mathcal{S}} \upsilon_k p_{ki} \\
&\overset{Fubini}{=} \sum_{k \in \mathcal{S}} \upsilon_k \sum_{i \in \mathcal{S}} p_{ki} p_{ij} = \sum_{k \in \mathcal{S}} \upsilon_k p_{kj}^{(2)}
\end{aligned}
$$

where $p_{kj}^{(2)} = \sum_{i \in \mathcal{S}} p_{ki} p_{ij} =$ the k-th row of P times the j-th column of P so $(p_{ij}^{(2)}) = PP = P^2$ and so

$$(\ldots P(X_2 = j-1), P(X_2 = j), P(X_2 = j+1), \ldots) = \upsilon P^2$$

is the probability distribution of X_2

- also

$$
\begin{aligned}
p_{kj}^{(2)} &= \sum_{i\in\mathcal{S}} p_{ki}p_{ij} = \sum_{i\in\mathcal{S}} P(X_1 = i \mid X_0 = k)P(X_1 = j \mid X_0 = i) \\
&\overset{TH}{=} \sum_{i\in\mathcal{S}} P(X_1 = i \mid X_0 = k)P(X_2 = j \mid X_1 = i) \\
&\overset{MP}{=} \sum_{i\in\mathcal{S}} P(X_1 = i \mid X_0 = k)P(X_2 = j \mid X_1 = i, X_0 = k) \\
&\overset{\text{below}}{=} \sum_{i\in\mathcal{S}} P(X_1 = i, X_2 = j \mid X_0 = k) = P(X_2 = j \mid X_0 = k)
\end{aligned}
$$

which is the probability of going from state k to state j in two time steps and so is a *2-step transition probability*

- note the last step uses (multiplication rule for conditional prob.)

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

- this can be generalized to give the distribution of X_n and the n-step transition probabilities

$$p_{ij}^{(n)} = P(X_n = j \mid X_0 = i)$$

Proposition III.2 For Markov chain with transition probability matrix P and initial prob. dist. υ, then $(p_{ij}^{(n)}) = P^n$ and

$$(\ldots P(X_n = j-1), P(X_n = j), P(X_n = j+1), \ldots) = \upsilon P^n.$$

Proof: (by induction) True for $n = 1$ and the above argument gives the inductive step. ■

Example III.2 - suppose

$$\begin{aligned} \mathcal{S} &= \{1,2,3,4\}, \\ \upsilon &= (1/4, 1/2, 1/8, 1/8), \\ P &= \begin{pmatrix} 0 & 1/3 & 1/2 & 1/6 \\ 1/3 & 0 & 1/2 & 1/6 \\ 1/6 & 1/6 & 0 & 2/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix} \end{aligned}$$

- a nonnegative matrix where each row sums to 1 is called a *stochastic matrix* and can serve as a transition matrix for a MC

- then for $n = 10$ the probability distribution of X_{10} is given by

$$\upsilon P^{10} = (0.2121419, 0.2121462, 0.3030114, 0.2727005)$$

and P^{10} is given by

0.2121699	0.2121530	0.3029984	0.2726787
0.2121530	0.2121699	0.3029984	0.2726787
0.2120860	0.2120860	0.3031143	0.2727137
0.2120977	0.2120977	0.3029867	0.2728179

- obtained via R commands

```
#computing probabilities
r1=c(0,1/3,1/2,1/6)
r2=c(1/3,0,1/2,1/6)
r3=c(1/6,1/6,0,2/3)
r4=c(1/3,1/3,1/3,0)
P=rbind(r1,r2,r3,r4)
P
P10=P%*%P%*%P%*%P%*%P%*%P%*%P%*%P%*%P%*%P
P10
nu=t(c(1/4,1/2,1/8,1/8))
nu%*%P10
```

note - if you want to start the chain in state i and then compute the probabilities for its evolution then simply put $\upsilon = e_i$ where e_i is the row vector with a 1 in the i-th place and 0's elsewhere

- so putting $\upsilon = e_1$ then

$$e_1 P^{10} = (0.2121699, 0.212153, 0.3029984, 0.2726787)$$

- how about simulating from the chain? below is 100 steps, including starting state, generated from this chain

```
[1] 2 3 4 3 4 2 1 2 4 1 3 4 1 3 4 2 4 2 1 2 3 2 1 2 1 2 4 1 3 4 1 2 3 4 3 4 2
[38] 1 3 1 4 1 3 4 2 1 2 3 4 2 1 2 3 4 2 3 4 1 3 2 4 1 2 3 4 3 4 1 2 1 4 1 3 4
[75] 1 3 4 1 2 1 3 2 1 3 4 1 2 1 3 4 2 1 4 1 3 1 3 4 1 3 4
```

computed using the R code

```
#simulating n steps from the chain
n=100
state=rep(0,101)
#starting state
U=runif(1,0,1)
if(U < 1/4) {i=1}
if(U >= 1/4 & U < 3/4) {i=2}
if(U >= 3/4 & U < 7/8) {i=3}
if(U >= 7/8) {i=4}
state[1]=i
for (k in 1:100){
# generate next step
U=runif(1,0,1)
if(i==1){
if(U < 1/3) {j=2}
if(U >= 1/3 & U < 5/6) {j=3}
if(U >= 5/6) {j=4}
}
```

```
if(i==2){
if(U < 1/3) {j=1}
if(U >= 1/3 & U < 5/6) {j=3}
if(U >= 5/6) {j=4}
}
if(i==3){
if(U < 1/6) {j=1}
if(U >= 1/6 & U < 1/3) {j=2}
if(U >= 1/4) {j=4}
}
if(i==4){
if(U < 1/3) {j=1}
if(U >= 1/3 & U < 2/3) {j=2}
if(U >= 2/3) {j=3}
}
state[k+1]=j
i=j
}
state
```
■

- suppose we want the distribution of

$$\bar{X}_{n:m} = \frac{1}{m} \sum_{i=0}^{m-1} X_{n+i}$$

for some n for which a closed form is difficult (or impossible) to obtain

- can we approximate this by simulation say computing $P\left(\bar{X}_{n:m} \leq x\right)$ for various x?

- if $X_1, X_2, \ldots$ were $i.i.d.$ then we would repeatedly generate the vector $(X_1, \ldots, X_m)$, compute $\bar{X}_{1:m}$ and determine the proportion of times $\bar{X}_{1:m} \leq x$ as the estimate

- but $X_1, X_2, \ldots$ are **not** $i.i.d.$ for a Markov chain and so we would have to repeatedly generate $X_0, X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m-1}$ compute $\bar{X}_{n:m}$ and determine the proportion of times $\bar{X}_{n:m} \leq x$ as the estimate

- if we can generate from the probability distribution of X_{n-1}, then can instead generate $X_{n-1}, X_n \,|\, X_{n-1}, X_{n+1} \,|\, X_n, \ldots, X_{n+m-1} \,|\, X_{n+m-2}$

- under some circumstances this can be made more effiicient when n is large (to be discussed) and this is connected with issues of convergence of the chain

Example III.3 *Random walks on* $\mathbb{Z}$

- suppose $Z_1, Z_2, \ldots$ are $i.i.d.$ with distribution given by a probability measure on $\mathbb{Z}$, stat. ind. of Z_0 also distributed on $\mathbb{Z}$ (the initial dist.)

- then the s.p. $\{X_n : n \in \mathbb{N}_0\}$ given by $X_n = \sum_{i=0}^{n} Z_i$ is a random walk with state space $\mathcal{S} = \mathbb{Z}$ and

$$\begin{aligned} & P(X_n = j \,|\, X_{n-1} = i, X_{n-2} = s_{n-2} \ldots, X_0 = s_0) \\ = \; & P(Z_n = j - i | \, X_{n-1} = i, X_{n-2} = s_{n-2}, \ldots, X_0 = s_0) \\ & \text{and since } Z_n \text{ is stat. ind. of } X_0, \ldots, X_{n-1} \\ = \; & P(Z_n = j - i) = P(Z_1 = j - i) = P(X_1 = j \,|\, X_0 = i) \end{aligned}$$

and so this is a time homogeneous Markov chain with transition matrix given by

$$(p_{ij}) = (P(Z_1 = j - i)) \in \mathbb{R}^{\mathbb{Z} \times \mathbb{Z}}$$

(all the rows of P are the same)

- when $Z_i \sim 2W_i - 1$ where $W_i \sim \text{Bernoulli}(p)$ this is a *simple random walk* (moves by increments/decrements of 1 at each step) and when $p = 1/2$ it is called a *symmetric simple random walk*, then

$$p_{ij} = \begin{cases} 0 & |i-j| \neq 1 \\ p & j = i+1 \\ 1-p & j = i-1 \end{cases}$$

$$P = \begin{pmatrix} & \ddots & \ddots & & \\ & \ddots & p & 0 & \\ \ddots & 1-p & 0 & p & \ddots \\ & 0 & 1-p & 0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

$$p_{ij}^{(n)} = P\left(\sum_{k=1}^{n} Z_k = j - i\right)$$

$$= \begin{cases} 0 & \text{if } n+j-i \text{ not even} \\ P\left(\sum_{k=1}^{n} W_k = \frac{n+j-i}{2}\right) & \text{if } n+j-i \text{ even} \end{cases}$$

$$= \begin{cases} 0 & \text{if } n+j-i \text{ not even} \\ \binom{n}{\frac{n+j-i}{2}} p^{\frac{n+j-i}{2}} (1-p)^{n-\frac{n+j-i}{2}} & \text{if } n+j-i \in \{0, 2, \ldots, 2n\} \end{cases}$$

since $\sum_{k=1}^{n} W_k \sim \text{binomial}(n, p)$ ■

Proposition III.3 For a Markov chain suppose $0 \le l_1 \le \cdots \le l_m, 0 \le k_1 \le \cdots \le k_n$, then

$$P_{(X_{k_n+l_1},\dots,X_{k_n+l_m})}(\cdot \,|\, X_{k_1},\dots,X_{k_n}) = P_{(X_{k_n+l_1},\dots,X_{k_n+l_m})}(\cdot \,|\, X_{k_n})$$

so the probability distribution of the future is determined by the present state.

Proof: We have that

$$\begin{aligned} & P(X_{k_n+l_1} = j_1, X_{k_n+l_2} = j_2, \dots, X_{k_n+l_m} = j_{l_m} \,|\, X_{k_1} = i_{k_1}, \dots, X_{k_n} = i_{k_n}) \\ = & \frac{P(X_{k_1} = i_{k_1}, \dots, X_{k_n} = i_{k_n}, X_{k_n+1} = j_1, X_{k_n+2} = j_2, \dots, X_{k_n+l_m} = j_{l_m})}{P(X_{k_1} = i_{k_1}, \dots, X_{k_n} = i_{k_n})}. \end{aligned}$$

Summing over

$$A = \left\{ \begin{array}{c} (s_0, s_1, \dots, s_{k_n+l_m}) : s_i \in \mathcal{S} \text{ with } s_i \text{ fixed when} \\ i \in \{k_1, \dots, k_n, k_n + l_1, \dots, k_n + l_m\} \end{array} \right\}$$

the numerator equals

$$\sum_A P\begin{pmatrix} X_0 = s_0, X_1 = s_1, \ldots, X_{k_n} = s_{k_n}, \\ X_{k_n+1} = s_{k_n+1}, X_{k_n+2} = s_{k_n+2}, \ldots, X_{k_n+l_m} = s_{k_n+l_m} \end{pmatrix}$$

$$\overset{MP}{=} \sum_A P\left(X_0 = s_0, X_1 = s_1, \ldots, X_{k_n} = s_{k_n},\right) \times$$

$$P(X_{k_n+1} = s_{k_n+1}, X_{k_n+2} = s_{k_n+2}, \ldots, X_{k_n+l_m} = s_{k_n+l_m} \mid X_{k_n} = i_{k_n})$$

$$= P(X_{k_1} = i_{k_1}, \ldots, X_{k_n} = i_{k_n}) \times$$

$$P(X_{k_n+l_1} = j_1, X_{k_n+l_2} = j_2, \ldots, X_{k_n+l_m} = j_{l_m} \mid X_{k_n} = i_{k_n})$$

and this gives the result. ■

Proposition III.4 *(Chapman-Kolmogorov equations) For a Markov chain and* $m, n, s \in \mathbb{N}_0$, then

$$\text{(i) } p_{ij}^{(m+n)} = \sum_k p_{ik}^{(m)} p_{kj}^{(n)}$$

$$\text{(ii) } p_{ij}^{(m+s+n)} = \sum_k \sum_l p_{ik}^{(m)} p_{kl}^{(s)} p_{lj}^{(n)}.$$

Proof: For (ii) we have

$$\begin{aligned} & p_{ij}^{(m+s+n)} = P(X_{m+s+n} = j \mid X_0 = i) \text{ and applying TTP} \\ = & \sum_k \sum_l P(X_{m+s+n} = j \mid X_0 = i, X_m = k, X_{m+s} = l) \times \\ & P(X_m = k, X_{m+s} = l \mid X_0 = i) \\ = & \sum_k \sum_l P(X_{m+s+n} = j \mid X_0 = i, X_m = k, X_{m+s} = l) \times \\ & P(X_{m+s} = l \mid X_0 = i, X_m = k) P(X_m = k \mid X_0 = i) \end{aligned}$$

$$
\begin{aligned}
&\quad \text{and by Prop. III.3} \\
&= \sum_k \sum_l P(X_{m+s+n} = j \mid X_{m+s} = l) \times \\
&\quad P(X_{m+s} = l \mid X_m = k) P(X_m = k \mid X_0 = i) \\
&\quad \text{and by TH} \\
&= \sum_k \sum_l p_{ik}^{(m)} p_{kl}^{(s)} p_{lj}^{(n)}
\end{aligned}
$$

and (i) follows similarly. ■

Corollary III.5 *(Chapman-Kolmogorov inequality)* $p_{ij}^{(m+n)} \geq p_{ik}^{(m)} p_{kj}^{(n)}$ for any k.

Exercises

III.1.1 Text 1.3.3

III.1.2 Text 1.3.97

III.1.3 Text 1.4.7

III.1.4 Prove part (i) of Prop. III.4.