## CS70: Lecture25.

Markov Chains 1.5

1. Review
2. Distribution
3. Irreducibility
4. Convergence

## Review

- Markov Chain:
- Finite set $\mathscr{X} ; \pi_{0} ; P=\{P(i, j), i, j \in \mathscr{X}\}$;
- $\operatorname{Pr}\left[X_{0}=i\right]=\pi_{0}(i), i \in \mathscr{X}$
- $\operatorname{Pr}\left[X_{n+1}=j \mid X_{0}, \ldots, X_{n}=i\right]=P(i, j), i, j \in \mathscr{X}, n \geq 0$
- Note:
$\operatorname{Pr}\left[X_{0}=i_{0}, X_{1}=i_{1}, \ldots, X_{n}=i_{n}\right]=\pi_{0}\left(i_{0}\right) P\left(i_{0}, i_{1}\right) \cdots P\left(i_{n-1}, i_{n}\right)$
- First Passage Time:
- $A \cap B=\emptyset ; \beta(i)=E\left[T_{A} \mid X_{0}=i\right] ; \alpha(i)=P\left[T_{A}<T_{B} \mid X_{0}=i\right]$
- $\beta(i)=1+\sum_{j} P(i, j) \beta(j)$;
- $\alpha(i)=\sum_{j} P(i, j) \alpha(j) . \alpha(A)=1, \alpha(B)=0$.

Stationary distributions: Example 2

$$
\begin{aligned}
& P=\left[\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right] \\
& \pi P=\pi \Leftrightarrow[\pi(1), \pi(2)]\left[\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right]=[\pi(1), \pi(2)] \Leftrightarrow \pi(1)=\pi(1) \text { and } \pi(2)=\pi(2) .
\end{aligned}
$$

Every distribution is invariant for this Markov chain. This is obvious, since $X_{n}=X_{0}$ for all $n$. Hence, $\operatorname{Pr}\left[X_{n}=i\right]=\operatorname{Pr}\left[X_{0}=i\right], \forall(i, n)$.

## Discussion.

We have seen a chain with one stationary,
and a chain with many
When is here just one?

Distribution of $X_{n}$


Recall $\pi_{n}$ is a distribution over states for $X_{n}$.
tationary distribution: $\pi=\pi P$
Distribution over states is the same before/after transition
probability entering $i: \sum_{i j} P(j, i) \pi(j)$.
probability leaving $i$ : $\pi_{i}$.
are Equal!
Distribution same after one step.
Questions? Does one exist? Is it unique?
If it exists and is unique. Then what?
Sometimes the distribution as $n \rightarrow \infty$

## Irreducibility.

Definition A Markov chain is irreducible if it can go from every state $i$ o every state $j$ (possibly in multiple steps).

## Examples


[A]

[B]

[C]
[A] is not irreducible. It cannot go from (2) to (1).
$[B]$ is not irreducible. It cannot go from (2) to (1).
[C] is irreducible. It can go from every $i$ to every $j$.
If you consider the graph with arrows when $P(i, j)>0$, irreducible means that there is a single connected component

Existence and uniqueness of Invariant Distribution

Theorem A finite irreducible Markov chain has one and only one invariant distribution.
That is, there is a unique positive vector $\pi=[\pi(1), \ldots, \pi(K)]$ such that $\pi P=\pi$ and $\sum_{k} \pi(k)=1$.
Ok. Now.
Only one stationary distribution if irreducible (or connected.)

## Long Term Fraction of Time in States

Theorem Let $X_{n}$ be an irreducible Markov chain with invariant distribution $\pi$. Then, for all $i, \frac{1}{n} \sum_{m=0}^{n-1} 1\left\{X_{m}=i\right\} \rightarrow \pi(i)$, as $n \rightarrow \infty$. Example 2:


## Long Term Fraction of Time in States

Theorem Let $X_{n}$ be an irreducible Markov chain with invariant distribution $\pi$.
Then, for all $i$,

$$
\frac{1}{n} \sum_{m=0}^{n-1} 1\left\{X_{m}=i\right\} \rightarrow \pi(i), \text { as } n \rightarrow \infty .
$$

The left-hand side is the fraction of time that $X_{m}=i$ during steps $0,1, \ldots, n-1$. Thus, this fraction of time approaches $\pi(i)$
Proof: Lecture note 24 gives a plausibility argument.

## Convergence to Invariant Distribution

Question: Assume that the MC is irreducible. Does $\pi_{n}$ approach the unique invariant distribution $\pi$ ?
Answer: Not necessarily. Here is an example:


Assume $X_{0}=1$. Then $X_{1}=2, X_{2}=1, X_{3}=2, \ldots$
Thus, if $\pi_{0}=[1,0], \pi_{1}=[0,1], \pi_{2}=[1,0], \pi_{3}=[0,1]$, etc.
Hence, $\pi_{n}$ does not converge to $\pi=[1 / 2,1 / 2]$.
Notice, all cycles or closed walks have even length.

## Long Term Fraction of Time in States

Theorem Let $X_{n}$ be an irreducible Markov chain with invariant distribution $\pi$. Then, for all $i, \frac{1}{n} \sum_{m=0}^{n-1} 1\left\{X_{m}=i\right\} \rightarrow \pi(i)$, as $n \rightarrow \infty$. Example 1:


$$
\begin{aligned}
& P=\left[\begin{array}{ll}
0 & 1 \\
1 & 0
\end{array}\right] \quad \pi P=\pi \Rightarrow \pi=[1 / 2,1 / 2 \\
& { }_{4}{ }_{n}
\end{aligned}
$$

The fraction of time in state 1 converges to $1 / 2$, which is $\pi(1)$.

## Periodicity

## Definition: Periodicity is gcd of the lengths of all closed walks Previous example: 2. <br> Definition If periodicity is $\mathbf{1}$, Markov chain is said to be aperiodic Otherwise, it is periodic.

Example


[B]
[A]: Closed walks of length 3 and length $4 \Longrightarrow$ periodicity $=1$. [B]: All closed walks multiple of $3 \Longrightarrow$ periodicity $=2$.

## Convergence of $\pi_{n}$

Theorem Let $X_{n}$ be an irreducible and aperiodic Markov chain with invariant distribution $\pi$. Then, for all $i \in \mathscr{X}$,

$$
\pi_{n}(i) \rightarrow \pi(i), \text { as } n \rightarrow \infty .
$$

## Example



## CS70: Continuous Probability.

## Continuous Probability 1

1. Examples
2. Events
3. Continuous Random Variables

## Convergence of $\pi_{n}$

Theorem Let $X_{n}$ be an irreducible and aperiodic Markov chain with invariant distribution $\pi$. Then, for all $i \in \mathscr{X}$,

$$
\pi_{n}(i) \rightarrow \pi(i), \text { as } n \rightarrow \infty
$$

## Example




## Uniformly at Random in $[0,1]$.

Choose a real number $X$, uniformly at random in $[0,1]$.
What is the probability that $X$ is exactly equal to $1 / 3$ ? Well, ..., 0 .

What is the probability that $X$ is exactly equal to 0.6? Again, 0 .
In fact, for any $x \in[0,1]$, one has $\operatorname{Pr}[X=x]=0$.
How should we then describe 'choosing uniformly at random in $[0,1]$ '? Here is the way to do it:

$$
\operatorname{Pr}[X \in[a, b]]=b-a, \forall 0 \leq a \leq b \leq 1 .
$$

Makes sense: $b-a$ is the fraction of $[0,1]$ that $[a, b]$ covers.

## Summary

## Markov Chains

- Markov Chain: $\operatorname{Pr}\left[X_{n+1}=j \mid X_{0}, \ldots, X_{n}=i\right]=P(i, j)$
- FSE: $\beta(i)=1+\sum_{j} P(i, j) \beta(j) ; \alpha(i)=\sum_{j} P(i, j) \alpha(j)$.
- $\pi_{n}=\pi_{0} P^{n}$
- $\pi$ is invariant iff $\pi P=\pi$
- Irreducible $\Rightarrow$ one and only one invariant distribution $\pi$
- Irreducible $\Rightarrow$ fraction of time in state $i$ approaches $\pi(i)$
- Irreducible + Aperiodic $\Rightarrow \pi_{n} \rightarrow \pi$.
- Calculating $\pi$ : One finds $\pi=[0,0 \ldots, 1] Q^{-1}$ where $Q=\cdots$.


## Uniformly at Random in $[0,1]$.

Let $[a, b]$ denote the event that the point $X$ is in the interval $[a, b]$.

$$
\operatorname{Pr}[[a, b]]=\frac{\text { length of }[a, b]}{\text { length of }[0,1]}=\frac{b-a}{1}=b-a .
$$

Intervals like $[a, b] \subseteq \Omega=[0,1]$ are events.
More generally, events in this space are unions of intervals. Example: the event $A$ - "within 0.2 of 0 or 1 " is $A=[0,0.2] \cup[0.8,1]$. Thus,

$$
\operatorname{Pr}[A]=\operatorname{Pr}[[0,0.2]]+\operatorname{Pr}[[0.8,1]]=0.4 .
$$

More generally, if $A_{n}$ are pairwise disjoint intervals in [ 0,1 ], then

$$
\operatorname{Pr}\left[\cup_{n} A_{n}\right]:=\sum_{n} \operatorname{Pr}\left[A_{n}\right] .
$$

Many subsets of $[0,1]$ are of this form. Thus, the probability of those sets is well defined. We call such sets events

Uniformly at Random in $[0,1]$.


Note: A radical change in approach.
Finite prob. space: $\Omega=\{1,2, \ldots, N\}$, with $\operatorname{Pr}[\omega]=p_{\omega}$.
$\Longrightarrow \operatorname{Pr}[A]=\sum_{\omega \in A} p_{\omega}$ for $A \subset \Omega$.
Continuous space: e.g., $\Omega=[0,1]$
$\operatorname{Pr}[\omega]$ is typically 0 .
Instead, start with $\operatorname{Pr}[A]$ for some events $A$.
Event $A=$ interval, or union of intervals.

## Uniformly at Random in $[0,1]$.



Think of $f(x)$ as describing how
one unit of probability is spread over $[0,1]$ : uniformly!
Then $\operatorname{Pr}[X \in A]$ is the probability mass over $A$.
Observe:

- This makes the probability automatically additive.
- We need $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x) d x=1$.


## Uniformly at Random in $[0,1]$.


$\operatorname{Pr}[X<x]=x$ for $x \in[0,1]$. Also, $\operatorname{Pr}[X<x]=0$ for $x<0$
$\operatorname{Pr}[X<x]=1$ for $.2 x>1$.
Define $F(x)=\operatorname{Pr}[X \leq x]$.


Then we have $\operatorname{Pr}[X \in(a, b]]=\operatorname{Pr}[X \leq b]-\operatorname{Pr}[X \leq a]=F(b)-F(a)$.
Thus, $F(\cdot)$ specifies the probability of all the events!

Uniformly at Random in $[0,1]$.


Discrete Approximation: Fix $N \gg 1$ and let $\varepsilon=1 / N$.
Define $Y=n \varepsilon$ if $(n-1) \varepsilon<X \leq n \varepsilon$ for $n=1, \ldots, N$.
Then $|X-Y| \leq \varepsilon$ and $Y$ is discrete: $Y \in\{\varepsilon, 2 \varepsilon, \ldots, N \varepsilon\}$.
Also, $\operatorname{Pr}[Y=n \varepsilon]=\frac{1}{N}$ for $n=1, \ldots, N$.
Thus, $X$ is 'almost discrete'

Uniformly at Random in $[0,1]$.

$\operatorname{Pr}[X \in(a, b]]=\operatorname{Pr}[X \leq b]-\operatorname{Pr}[X \leq a]=F(b)-F(a)$.
An alternative view is to define $f(x)=\frac{d}{d x} F(x)=1\{x \in[0,1]\}$. Then

$$
F(b)-F(a)=\int_{a}^{b} f(x) d x
$$

Thus, the probability of an event is the integral of $f(x)$ over the event:

$$
\operatorname{Pr}[X \in A]=\int_{A} f(x) d x
$$

Nonuniformly at Random in $[0,1]$.


This figure shows a different choice of $f(x) \geq 0$ with $\int_{-\infty}^{\infty} f(x) d x=1$ It defines another way of choosing $X$ at random in $[0,1]$.
Note that $X$ is more likely to be closer to 1 than to 0 .
One has $\operatorname{Pr}[X \leq x]=\int_{-\infty}^{x} f(u) d u=x^{2}$ for $x \in[0,1]$
Also, $\operatorname{Pr}[X \in(x, x+\varepsilon)]=\int_{x}^{x+\varepsilon} f(u) d u \approx f(x) \varepsilon$.

Another Nonuniform Choice at Random in [0, 1].


This figure shows yet a different choice of $f(x) \geq 0$ with $\int_{-\infty}^{\infty} f(x) d x=1$.
It defines another way of choosing $X$ at random in $[0,1]$.
Note that $X$ is more likely to be closer to $1 / 2$ than to 0 or 1 .
For instance, $\operatorname{Pr}[X \in[0,1 / 3]]=\int_{0}^{1 / 3} 4 x d x=2\left[x^{2}\right]_{0}^{1 / 3}=\frac{2}{9}$.
Thus, $\operatorname{Pr}[X \in[0,1 / 3]]=\operatorname{Pr}[X \in[2 / 3,1]]=\frac{2}{9}$ and
$\operatorname{Pr}[X \in[1 / 3,2 / 3]]=\frac{5}{9}$.

## Discrete Approximation

Fix $\varepsilon \ll 1$ and let $Y=n \varepsilon$ if $X \in(n \varepsilon,(n+1) \varepsilon]$
Thus, $\operatorname{Pr}[Y=n \varepsilon]=F_{X}((n+1) \varepsilon)-F_{X}(n \varepsilon)$.
Note that $|X-Y| \leq \varepsilon$ and $Y$ is a discrete random variable.
Also, if $f_{X}(x)=\frac{d}{d x} F_{X}(x)$, then $F_{X}(x+\varepsilon)-F_{X}(x) \approx f_{X}(x) \varepsilon$.
Hence, $\operatorname{Pr}[Y=n \varepsilon] \approx f_{X}(n \varepsilon) \varepsilon$
Thus, we can think of $X$ of being almost discrete with $\operatorname{Pr}[X=n \varepsilon] \approx f_{X}(n \varepsilon) \varepsilon$

## General Random Choice in $\Re$

Let $F(x)$ be a nondecreasing function with $F(-\infty)=0$ and $F(+\infty)=1$.
Define $X$ by $\operatorname{Pr}[X \in(a, b]]=F(b)-F(a)$ for $a<b$. Also, for
$a_{1}<b_{1}<a_{2}<b_{2}<\cdots<b_{n}$,

$$
\begin{aligned}
\operatorname{Pr}[X \in & \left.\left(a_{1}, b_{1}\right] \cup\left(a_{2}, b_{2}\right] \cup\left(a_{n}, b_{n}\right]\right] \\
& =\operatorname{Pr}\left[X \in\left(a_{1}, b_{1}\right]\right]+\cdots+\operatorname{Pr}\left[X \in\left(a_{n}, b_{n}\right]\right] \\
& =F\left(b_{1}\right)-F\left(a_{1}\right)+\cdots+F\left(b_{n}\right)-F\left(a_{n}\right) .
\end{aligned}
$$

Let $f(x)=\frac{d}{d x} F(x)$. Then,

$$
\operatorname{Pr}[X \in(x, x+\varepsilon]]=F(x+\varepsilon)-F(x) \approx f(x) \varepsilon .
$$

Here, $F(x)$ is called the cumulative distribution function (cdf) of $X$ and $f(x)$ is the probability density function (pdf) of $X$.
To indicate that $F$ and $f$ correspond to the $\operatorname{RV} X$, we will write them $F_{X}(x)$ and $f_{X}(x)$.

## Example: CDF

Example: hitting random location on gas tank.
Random location on circle.


Random Variable: $Y$ distance from center.
Probability within $y$ of center:

$$
\begin{aligned}
\operatorname{Pr}[Y \leq y] & =\frac{\text { area of small circle }}{\text { area of dartboard }} \\
& =\frac{\pi y^{2}}{\pi}=y^{2} .
\end{aligned}
$$

Hence

$$
F_{Y}(y)=\operatorname{Pr}[Y \leq y]=\left\{\begin{array}{lr}
0 & \text { for } y<0 \\
y^{2} & \text { for } 0 \leq y \leq 1 \\
1 & \text { for } y>1
\end{array}\right.
$$

## $\operatorname{Pr}[X \in(x, x+\varepsilon)]$

An illustration of $\operatorname{Pr}[X \in(x, x+\varepsilon)] \approx f_{X}(x) \varepsilon$ :


Thus, the pdf is the 'local probability by unit length. It is the 'probability density.'

Calculation of event with dartboard..

## Probability between .5 and .6 of center? Recall CDF.

$$
\begin{aligned}
& F_{Y}(y)=\operatorname{Pr}[Y \leq y]=\left\{\begin{array}{lr}
0 & \text { for } y<0 \\
y^{2} & \text { for } 0 \leq y \leq 1 \\
1 & \text { for } y>1
\end{array}\right. \\
& \begin{aligned}
\operatorname{Pr}[0.5<Y \leq 0.6] & =\operatorname{Pr}[Y \leq 0.6]-\operatorname{Pr}[Y \leq 0.5] \\
& =F_{Y}(0.6)-F_{Y}(0.5) \\
& =.36-.25 \\
& =.11
\end{aligned}
\end{aligned}
$$

PDF.

Example: "Dart" board.
Recall that

$$
\begin{gathered}
F_{Y}(y)=\operatorname{Pr}[Y \leq y]=\left\{\begin{array}{lr}
0 & \text { for } y<0 \\
y^{2} & \text { for } 0 \leq y \leq 1 \\
1 & \text { for } y>1
\end{array}\right. \\
f_{Y}(y)=F_{Y}^{\prime}(y)=\left\{\begin{array}{lr}
0 & \text { for } y<0 \\
2 y & \text { for } 0 \leq y \leq 1 \\
0 & \text { for } y>1
\end{array}\right.
\end{gathered}
$$

The cumulative distribution function (cdf) and probability distribution function (pdf) give full information
Use whichever is convenient.

## $\operatorname{Expo}(\lambda)$

The exponential distribution with parameter $\lambda>0$ is defined by

$$
f_{X}(x)=\lambda e^{-\lambda x} 1\{x \geq 0\}
$$

$F_{X}(x)= \begin{cases}0, & \text { if } x<0 \\ 1-e^{-\lambda x}, & \text { if } x \geq 0\end{cases}$



Note that $\operatorname{Pr}[X>t]=e^{-\lambda t}$ for $t>0$

## Target



## Continuous Random Variables

Continuous random variable $X$, specified by

1. $F_{X}(x)=\operatorname{Pr}[X \leq x]$ for all $x$.

Cumulative Distribution Function (cdf)
$\operatorname{Pr}[a<X \leq b]=F_{X}(b)-F_{X}(a)$
$1.10 \leq F_{X}(x) \leq 1$ for all $x \in \Re$. 1.2 $F_{X}(x) \leq F_{X}(y)$ if $x \leq y$
2. $\operatorname{Or} f_{X}(x)$, where $F_{X}(x)=\int_{-\infty}^{x} f_{X}(u) d u$ or $f_{X}(x)=\frac{d\left(F_{X}(x)\right)}{d x}$. Probability Density Function (pdf).
$\operatorname{Pr}[a<X \leq b]=\int_{a}^{b} f_{X}(x) d x=F_{X}(b)-F_{X}(a)$
$2.1 f_{X}(x) \geq 0$ for all $x \in \Re$
$2.2 \int_{-\infty}^{\infty} f_{X}(x) d x=1$.
Recall that $\operatorname{Pr}[X \in(x, x+\delta)] \approx f_{X}(x) \delta$
$X$ "takes" value $n \delta$, for $n \in Z$, with $\operatorname{Pr}[X=n \delta]=f_{X}(n \delta) \delta$
$U[a, b]$


A Picture


The pdf $f_{X}(x)$ is a nonnegative function that integrates to 1 .
The cdf $F_{X}(x)$ is the integral of $f_{X}$.

$$
\begin{aligned}
& \operatorname{Pr}[x<X<x+\delta] \approx f_{X}(x) \delta \\
& \operatorname{Pr}[X \leq x]=F_{X}(x)=\int_{-\infty}^{x} f_{X}(u) d u
\end{aligned}
$$

Multiple Continuous Random Variables

One defines a pair $(X, Y)$ of continuous RV s by specifying $f_{X, Y}(X, Y)$ for $x, y \in \mathfrak{R}$ where
$f_{X, Y}(x, y) d x d y=\operatorname{Pr}[X \in(x, x+d x), Y \in(y+d y)]$.
The function $f_{X, Y}(x, y)$ is called the joint pdf of $X$ and $Y$. Example: Choose a point $(X, Y)$ uniformly in the set $A \subset \mathfrak{R}^{2}$. Then

$$
f_{X, Y}(x, y)=\frac{1}{|A|} 1\{(x, y) \in A\}
$$

where $|A|$ is the area of $A$.
Interpretation. Think of $(X, Y)$ as being discrete on a grid with mesh size $\varepsilon$ and $\operatorname{Pr}[X=m \varepsilon, Y=n \varepsilon]=f_{X, Y}(m \varepsilon, n \varepsilon) \varepsilon^{2}$.
Extension: $\mathbf{X}=\left(X_{1}, \ldots, X_{n}\right)$ with $f_{\mathbf{x}}(\mathbf{x})$.

## Example of Continuous $(X, Y)$

Pick a point $(X, Y)$ uniformly in the unit circle.


Thus, $f_{X, Y}(x, y)=\frac{1}{\pi} 1\left\{x^{2}+y^{2} \leq 1\right\}$.
Consequently,

$$
\begin{aligned}
& \operatorname{Pr}[X>0, Y>0]=\frac{1}{4} \\
& \operatorname{Pr}[X<0, Y>0]=\frac{1}{4} \\
& \operatorname{Pr}\left[X^{2}+Y^{2} \leq r^{2}\right]=\frac{r^{2}}{\pi} \\
& \operatorname{Pr}[X>Y]=\frac{1}{\Omega} .
\end{aligned}
$$

## Summary

## Continuous Probability 1

1. pdf: $\operatorname{Pr}[X \in(x, x+\delta]]=f_{X}(x) \delta$.
2. CDF: $\operatorname{Pr}[X \leq x]=F_{X}(x)=\int_{-\infty}^{x} f_{X}(y) d y$.
3. $U[a, b]: f_{X}(x)=\frac{1}{b-a} 1\{a \leq x \leq b\} ; F_{X}(x)=\frac{x-a}{b-a}$ for $a \leq x \leq b$.
4. $\operatorname{Expo}(\lambda)$ :
$f_{X}(x)=\lambda \exp \{-\lambda x\} 1\{x \geq 0\} ; F_{X}(x)=1-\exp \{-\lambda x\}$ for $x \leq 0$
5. Target: $f_{X}(x)=2 x 1\{0 \leq x \leq 1\} ; F_{X}(x)=x^{2}$ for $0 \leq x \leq 1$.
6. Joints: Is this $4 / 20$ ?

Joint pdf: $\operatorname{Pr}[X \in(x, x+\delta), Y=(y, y+\delta))=f_{X, Y}(x, y) \delta^{2}$

