COMP 680
Statistics for Computing and Data Science
Week 10: Stochastic Processes

Su Chen, Assistant Teaching Professor,
Rice D2K Lab

Outline

Definition

- A stochastic process $\{X_t : t \in T\}$ is a collection of random variables indexed by t.

Definition

- A stochastic process $\{X_t : t \in T\}$ is a collection of random variables indexed by t.
  - X_t defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$
  - X_t takes values in the state space $\mathcal{X}$

Definition

- A stochastic process $\{X_t : t \in T\}$ is a collection of random variables indexed by t.
  - X_t defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$
  - X_t takes values in the state space $\mathcal{X}$
- $X_t = X(t, \omega) : (t, \omega) \in (T, \Omega) \to \mathcal{X} \subseteq \mathbb{R}$
  - the state space $\mathcal{X}$ can be discrete or continuous
  - the index set T can be discrete or continuous

Definition

- A stochastic process $\{X_t : t \in T\}$ is a collection of random variables indexed by t.
  - X_t defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$
  - X_t takes values in the state space $\mathcal{X}$
- $X_t = X(t, \omega) : (t, \omega) \in (T, \Omega) \to \mathcal{X} \subseteq \mathbb{R}$
  - the state space $\mathcal{X}$ can be discrete or continuous
  - the index set T can be discrete or continuous
- Can have 4 different types:
  - the discrete-time ones are more common

Examples

- iid samples is a trivial case with discrete index
    - X_t's in general are NOT independent!

Examples

- iid samples is a trivial case with discrete index
  - X_t's in general are NOT independent!
- Stock price
  - X_t would be the price of a particular stock at time t
  - continuous state

Examples

- iid samples is a trivial case with discrete index
  - X_t's in general are NOT independent!
- Stock price
  - X_t would be the price of a particular stock at time t
  - continuous state
- The weather
  - discrete state: sunny, rain, cloud
  - continuous state: temperature

Examples

- iid samples is a trivial case with discrete index
  - X_t's in general are NOT independent!
- Stock price
  - X_t would be the price of a particular stock at time t
  - continuous state
- The weather
  - discrete state: sunny, rain, cloud
  - continuous state: temperature
- Brownian motion
  - X_t would be the position of a particle at time t
  - continuous state

Outline

Markov Property

- A discrete-time stochastic process is Markovian (or has the Markov Property) if for all t and all $x \in \mathcal{X}$

$$\mathbb{P}(X_t = x | X_0, X_1, \cdots X_{t-1}) = \mathbb{P}(X_t = x | X_{t-1})$$

Markov Property

- A discrete-time stochastic process is Markovian (or has the Markov Property) if for all t and all $x \in \mathcal{X}$

$$\mathbb{P}(X_t = x | X_0, X_1, \cdots X_{t-1}) = \mathbb{P}(X_t = x | X_{t-1})$$

- also known as “memoryless” property:
  - the future only depends on present, but not the past
  - often as a model assumption

Markov Chains

- Discrete-time discrete-state stochastic processes that have the Markov property

Markov Chains

- Discrete-time discrete-state stochastic processes that have the Markov property
- For each t, X_t is a discrete random variable with a distribution

Markov Chains

- Discrete-time discrete-state stochastic processes that have the Markov property
- For each t, X_t is a discrete random variable with a distribution
- Rich and complex theory:
    - do the distributions converge?
    - if so, what is the limiting distribution?

Markov Chains

- Discrete-time discrete-state stochastic processes that have the Markov property
- For each t, X_t is a discrete random variable with a distribution
- Rich and complex theory:
  - do the distributions converge?
  - if so, what is the limiting distribution?
- Applications in:
  - finance
  - machine learning
  - engineering

Transition Probability

- Key quantities of a Markov chain that define the probabilities of jumping from one state to another
    - "state" is a set of values each X_t can take

Transition Probability

- Key quantities of a Markov chain that define the probabilities of jumping from one state to another
  - "state" is a set of values each X_t can take
- Define the probability of jumping from state i to state j:

$$p_{ij} = \mathbb{P}(X_{t+1} = j | X_t = i)$$

Transition Probability

- Key quantities of a Markov chain that define the probabilities of jumping from one state to another
  - "state" is a set of values each X_t can take
- Define the probability of jumping from state i to state j:

$$p_{ij} = \mathbb{P}(X_{t+1} = j | X_t = i)$$

- A Markov chain is homogeneous if $\mathbb{P}(X_{t+1} = j | X_t = i)$ does not depend on t.
  - $\mathbb{P}(X_{t+1} = j | X_t = i) = \mathbb{P}(X_1 = j | X_0 = i)$

Transition Matrix

Suppose a Markov chain has a total of n states, we can define a $n \times n$ transition matrix P whose (i, j) element is the transition probability from state i to state j:

$$p_{ij} = \mathbb{P}(X_{t+1} = j | X_t = i)$$

Transition Matrix

Suppose a Markov chain has a total of n states, we can define a $n \times n$ transition matrix P whose (i, j) element is the transition probability from state i to state j:

$$p_{ij} = \mathbb{P}(X_{t+1} = j | X_t = i)$$

- Transition matrix for n states:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

Transition Matrix

Suppose a Markov chain has a total of n states, we can define a $n \times n$ transition matrix P whose (i, j) element is the transition probability from state i to state j:

$$p_{ij} = \mathbb{P}(X_{t+1} = j | X_t = i)$$

- Transition matrix for n states:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

- row sum is always 1, why?

Weather Forecasting

- Example: states = {1=sunny, 2=cloudy, 3=rainy}

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

Weather Forecasting

- Example: states = {1=sunny, 2=cloudy, 3=rainy}

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

- Today is sunny, what is the probability that
  - tomorrow is rainy?

Weather Forecasting

- Example: states = $\{1{=}\text{sunny},\ 2{=}\text{cloudy},\ 3{=}\text{rainy}\}$

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

- Today is sunny, what is the probability that
  - tomorrow is rainy?
  - the day after tomorrow is rainy??

Weather Forecasting

- Example: states = $\{1{=}\text{sunny},\ 2{=}\text{cloudy},\ 3{=}\text{rainy}\}$

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

- Today is sunny, what is the probability that
  - tomorrow is rainy?
  - the day after tomorrow is rainy??
  - one week after today is rainy???

K-step Transition Probabilities

- If you want to predict the future, what we really need is the transition probability from state i to state j in arbitrary time steps k:

$$p_{ij}(k) = \mathbb{P}(X_{t+k} = j | X_t = i)$$

K-step Transition Probabilities

- If you want to predict the future, what we really need is the transition probability from state i to state j in arbitrary time steps k:

$$p_{ij}(k) = \mathbb{P}(X_{t+k} = j | X_t = i)$$

- It turns out $p_{ij}(k)$ is the (i, j) element in the matrix P^k, where:

$$P^k = \underbrace{P \times P \times \cdots \times P}_{\text{multiply the matrix k times}}$$

  - matrix multiplication!

Marginal Distribution

- For each t, X_t is a discrete random variable with a distribution
  - can we find the marginal PMF for X_t?
  - be able to predict for any future time t

Marginal Distribution

- For each t, X_t is a discrete random variable with a distribution
  - can we find the marginal PMF for X_t?
  - be able to predict for any future time t
- Yes, but we need:
  - initial distribution $\pi_0 = (p_1, p_2, \cdots p_n)$
  - transition matrix P

Marginal Distribution

- For each t, X_t is a discrete random variable with a distribution
  - can we find the marginal PMF for X_t?
  - be able to predict for any future time t
- Yes, but we need:
  - initial distribution $\pi_0 = (p_1, p_2, \cdots p_n)$
  - transition matrix P
- It turns out:

$$\pi_k = \pi_0 \times P^k$$

Marginal Distribution

- For each t, X_t is a discrete random variable with a distribution
  - can we find the marginal PMF for X_t?
  - be able to predict for any future time t
- Yes, but we need:
  - initial distribution $\pi_0 = (p_1, p_2, \cdots p_n)$
  - transition matrix P
- It turns out:

$$\pi_k = \pi_0 \times P^k$$

  - matrix multiplication!
  - $\pi_k = (p_1^{(k)}, p_2^{(k)}, \cdots p_n^{(k)})$ is the marginal PMF for X_k!

Weather Forecasting

- Example: states = {1=sunny, 2=cloudy, 3=rainy}

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

Weather Forecasting

- Example: states = {1=sunny, 2=cloudy, 3=rainy}

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

- Today is sunny, what is the initial distribution?

Weather Forecasting

- Example: states = {1=sunny, 2=cloudy, 3=rainy}

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

- Today is sunny, what is the initial distribution?
  - what is the weather "distribution" for tomorrow?

Weather Forecasting

- Example: states = {1=sunny, 2=cloudy, 3=rainy}

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.4 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

- Today is sunny, what is the initial distribution?
  - what is the weather "distribution" for tomorrow?
- What if I don't know today is sunny, but I believe it is 90% sunny, 10% cloudy?

Stationary Distribution

Interested in the distribution of long-term states:

$$\lim_{k\to\infty} \pi_k = \pi_0 \times P^k = \pi_{stat}?$$

Stationary Distribution

Interested in the distribution of long-term states:

$$\lim_{k\to\infty} \pi_k = \pi_0 \times P^k = \pi_{stat}?$$

- If such limit exists, then it must satisfy:

$$\pi_{stat} \times P = \pi_{stat}$$

  - proof:

Stationary Distribution

- Stationary Distribution is also known as steady-state distribution, invariant distribution or equilibrium.

Stationary Distribution

- Stationary Distribution is also known as steady-state distribution, invariant distribution or equilibrium.
- How to find π_{stat} in general? exist?? unique???

Stationary Distribution

- Stationary Distribution is also known as steady-state distribution, invariant distribution or equilibrium.
- How to find π_{stat} in general? exist?? unique???
  - irreducible
  - ergodic (aperiodic)

Stationary Distribution

- Stationary Distribution is also known as steady-state distribution, invariant distribution or equilibrium.
- How to find π_{stat} in general? exist?? unique???
  - irreducible
  - ergodic (aperiodic)
- Solve the linear system:

$$\pi_{stat} P = \pi_{stat} \Rightarrow \pi_{stat}(P - I) = \mathbf{0}$$

Example: Hardy-Weinberg

Suppose a gene can be type A or a, and let $(p, q, 1-p-q)$ denote the fraction of genotypes AA, Aa and aa in population.

Example: Hardy-Weinberg

Suppose a gene can be type A or a, and let $(p, q, 1-p-q)$ denote the fraction of genotypes AA, Aa and aa in population.

- It turns out the genotypes in future generations is a Markov chain with a stationary distribution!

Example: Hardy-Weinberg

Suppose a gene can be type A or a, and let $(p, q, 1-p-q)$ denote the fraction of genotypes AA, Aa and aa in population.

- It turns out the genotypes in future generations is a Markov chain with a stationary distribution!
- $P =?$
- $\pi_{stat} =?$

Inference for Markov Chain

- Unknown parameters: π_0 and P

Inference for Markov Chain

- Unknown parameters: π_0 and P
- Observed data: realization of $X_0, X_1, \cdots X_T$ from the Markov Chain
  - with only one data for X_0, can't estimate π_0

Inference for Markov Chain

- Unknown parameters: π_0 and P
- Observed data: realization of $X_0, X_1, \cdots X_T$ from the Markov Chain
  - with only one data for X_0, can't estimate π_0
- Maximum likelihood estimate of P:
  - total of n^2 parameters!
  - what is the likelihood function?

Inference for Markov Chain

- Unknown parameters: π_0 and P
- Observed data: realization of $X_0, X_1, \cdots X_T$ from the Markov Chain
  - with only one data for X_0, can't estimate π_0
- Maximum likelihood estimate of P:
  - total of n^2 parameters!
  - what is the likelihood function?
- Not surprisingly, we have:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i}$$

Inference for Markov Chain

- Unknown parameters: π_0 and P
- Observed data: realization of $X_0, X_1, \cdots X_T$ from the Markov Chain
    - with only one data for X_0, can't estimate π_0
- Maximum likelihood estimate of P:
    - total of n^2 parameters!
    - what is the likelihood function?
- Not surprisingly, we have:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i}$$

    - n_{ij}: observed number of transition from state i to j
    - $n_i = \sum_{k=1}^{n} n_{ik}$: observed number of state i

Outline

Poisson Process

- A counting process to model occurrence of events over time
  - traffic accidents, insurance claims, arrival of phone calls/emails...
  - single parameter λ: intensity rate

Poisson Process

- A counting process to model occurrence of events over time
  - traffic accidents, insurance claims, arrival of phone calls/emails...
  - single parameter λ: intensity rate
- Continuous-time, discrete-state $\mathcal{X} = \{0, 1, 2, \cdots\}$

Poisson Process

- A counting process to model occurrence of events over time
  - traffic accidents, insurance claims, arrival of phone calls/emails...
  - single parameter λ: intensity rate
- Continuous-time, discrete-state $\mathcal{X} = \{0, 1, 2, \cdots\}$
- X_t is a Poisson random variable for any given t:

$$X_t \sim \text{Poisson}(\lambda t)$$

Poisson Distribution

- Recall $X \sim \text{Poisson}(\lambda)$ with PMF:

$$\mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, \cdots$$

  - $\mathbb{E}[X] = \mathbb{V}(X) = \lambda$

Poisson Distribution

- Recall $X \sim \text{Poisson}(\lambda)$ with PMF:

$$\mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, \cdots$$

  - $\mathbb{E}[X] = \mathbb{V}(X) = \lambda$
- Poisson is the limiting distribution of Binomial(n, p)
  - as $n \to \infty, p \to 0, np \to \lambda$,
  - PMF converges:

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

Time Between Occurrence

- If X_t is a Poisson Process with intensity rate λ
- Then time between occurrence T has exponential distribution:

Time Between Occurrence

- If X_t is a Poisson Process with intensity rate λ
- Then time between occurrence T has exponential distribution:
  - $T \sim \text{Exponential}(\lambda)$
  - PDF: $f(t) = \lambda e^{-\lambda t}$ for $t \geq 0$
  - $\mathbb{E}[T] = 1/\lambda$, $\mathbb{V} = 1/\lambda^2$

Time Between Occurrence

- If X_t is a Poisson Process with intensity rate λ
- Then time between occurrence T has exponential distribution:
  - $T \sim \text{Exponential}(\lambda)$
  - PDF: $f(t) = \lambda e^{-\lambda t}$ for $t \geq 0$
  - $\mathbb{E}[T] = 1/\lambda$, $\mathbb{V} = 1/\lambda^2$
- Proof

$$\mathbb{P}(T > t) = \mathbb{P}(X_t = 0) = e^{-\lambda t}$$

Example

Mr. D drinks and drives once in 4 years, and every time he is caught by police. According to the laws of his state, the third time he is caught drinking and driving results in the loss of his driver's license.

- Assume Poisson process is the correct model for "rare events" such as drinking and driving

Example

Mr. D drinks and drives once in 4 years, and every time he is caught by police. According to the laws of his state, the third time he is caught drinking and driving results in the loss of his driver's license.

- Assume Poisson process is the correct model for "rare events" such as drinking and driving
- What is the probability that Mr. D will keep his license for at least 10 years?

Outline