Chapter 16: Stochastic Processes

DSCC 462

Computational Introduction to Statistics

Anson Kahng

Fall 2022

Final Project Announcement

- Due next Friday: short 1-2 minute video describing your approach to the last question (#7) – mid-project check-in
  - Best videos will be asked to present in class (~5 minutes or so) on Tuesday, December 13
  - Extra credit opportunity!
- Official announcement will be posted after class today

Plan for Today

- Stochastic processes (how random variables evolve over time)
- Poisson processes (counting arrivals)
- Markov chains (any memoryless process)

Stochastic Processes

- A stochastic process is a process that changes over time in a random way
- Examples:
  - Random walk (when is not random, how is random)
  - Radioactive decay (when is random, how is not random)
  - Financial derivatives (when and how are both random)

Stochastic Processes

- Represented as an indexed collection of random variables, where the index usually represents time
- $\{X_t\}, t \in T$
- Discrete time process: sequence $X_1, X_2, \ldots$
- Continuous time process: process on a subset $t \in [0,\infty)$

Counting Processes

- A counting process is a stochastic process, $N(t)$, defined on $t \in [0,\infty)$ satisfying the following conditions
  - $N(0) = 0$
  - $N(t)$ is nondecreasing in t
  - $N(t)$ always increments by $+1$
- Think of $N(t)$ as an arrival process
  - At $N(0) = 0$, no one has arrived yet; one person arrives at a time; no one leaves
- For $t > s$, we have $N(t) - N(s)$ is the number of arrivals in the interval $(s, t]$

Poisson Processes

- A counting process is a Poisson process with rate λ if the following hold
  - Independent increments: $N(t_1) - N(s_1)$ and $N(t_2) - N(s_2)$ are independent whenever $s_1 < t_1 < s_2 < t_2$
  - Stationary increments: Distribution of $N(t) - N(s)$ only depends on $t - s$ for any $s < t$
  - Number of arrivals in any interval of length $s > 0$ has a $Pois(\lambda s)$ distribution

Poisson Processes Example

- The number of customers arriving at a department store can be modeled by a Poisson process with intensity $\lambda = 10$ customers / hour
- Find the probability that there are 2 customers between 11:00 and 11:20
- What about between 12:15 and 12:35?

Poisson Processes Example

- The number of customers arriving at a department store can be modeled by a Poisson process with intensity $\lambda = 10$ customers / hour
- Find the probability that there are 2 customers between 11:00 and 11:20
  - Interval is of length $1/3$ hour
  - $\Pr(X = 2) = \dfrac{e^{-\frac{10}{3}}\left(\frac{10}{3}\right)^2}{2!} \approx 0.2$
- What about between 12:15 and 12:35?
  - The same!

Poisson Processes

- Let's look at the time between Poisson process events
- Let X_i be the time between arrivals $i - 1$ and i
- $\Pr(X_2 > t + s \,|\, X_1 > t) = \Pr(N(t + s) - N(t) = 0) = e^{-\lambda s}$
  - If $X_2 > t + s$, then there were no arrivals between t and $t + s$
- Time between events is exponentially distributed with rate λ

Poisson Processes: Summary

- When to use them:
  - Counting / arrival processes (+1 increments at random times)
- Properties:
  - Number of arrivals in a time frame follows a Poisson distribution
  - Time between arrivals follows an exponential distribution

Markov Chains

- Markov chains are (discrete time) stochastic properties with a set of *states* and probabilistic *transitions* between the states
- Have *memoryless property* (Markovian property)
- Consider a discrete time stochastic process X_i
- X_i is a Markov process if:

$$\Pr(X_n = j \mid X_1, X_2, \ldots, X_{n-1}) = \Pr(X_n = j \mid X_{n-1})$$

- Intuitively: Probability of an event depends only on the previous state

Markov Chains

- In this lecture, we will consider Markov chains that are
  - Discrete time (time proceeds in jumps of one increment)
  - Time-homogeneous (probabilities do not change over time)
  - Finite state space
- There exist more complicated Markov chains, but they're out of scope for today

Properties of Markov Chains

- $\sum_j \Pr(X_n = j \mid X_{n-1} = i) = 1$ (outgoing edges from each state sum to 1)
- Can express Markov chains as a graph
    - Nodes represent states and directed edges represent transitions between states

Transition Matrices

- Can also represent a Markov chain by a transition matrix P:
- $$P = \begin{bmatrix} p_{00} & p_{01} & \cdots \\ p_{10} & p_{11} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$
- Entry $(i,j) = p_{ij} = \Pr(X_n = j \mid X_{n-1} = i)$ is the probability of going from state i (row coordinate) to state j (column coordinate)
- We assume that transition probabilities do not change with time (time homogeneous)

Properties of Markov Chains: Example

- Setup: Consider the process of repeatedly flipping an unfair coin with probability $\alpha > 1/2$ of being heads and $1 - \alpha < 1/2$ of being tails. This is expressible as a Markov chain.
- Question 1: How can we express this as a graph?
- Question 2: How can we express this as a transition matrix?

Absorbing States

- In some Markov chains, there are absorbing states, which are states that have no outgoing transitions
  - Expressed as a self-edge with probability 1
- Pictorially:

Example: Gambler's Ruin

- Consider the Gambler's Ruin setup: A gambler starts with s dollars. Every time step, he bets \$1 that a tossed coin will come up heads. The coin has probability $1/2$ of being heads and $1/2$ of being tails. The gambler keeps betting until either (1) he has no money left, or (2) he reaches some value $n > s$, at which point he collects his winnings and leaves.

- Express this as a graph:

Example: Gambler’s Ruin

- Question: What is the probability of winning? (Key idea: recurrences!)

Example: Gambler's Ruin

- Question: What is the probability of winning? (Key idea: recurrences!)

Let R_i = probability of winning starting at i
Recurrence: $R_i = (1/2)R_{i-1} + (1/2)R_{i+1}$, so $R_i - R_{i-1} = R_{i+1} - R_i$
Because $R_0 = 0$, we have $R_1 = R_2 - R_1 \implies R_2 = 2R_1$
Continuing up, we have that in general $R_k = kR_1$
Using the fact that $R_n = 1$, we see that $R_1 = 1/n$
In general, $R_s = s/n$

Properties of Markov Chains

- Suppose we want to know how likely it is for a Markov chain to be in state j after k transitions, given that it started in state i

$$\Pr(X_{n+k} = j \mid X_n = i) = P_{ij}^k$$

Here, P_{ij}^k is the $(i, j)^{th}$ entry in P^k

- Let's consider $k = 2$ transitions

- Get as a result of matrix multiplication!

Properties of Markov Chains

- Suppose we want to know how likely it is for a Markov chain to be in state j after k transitions, given that it started in state i

$$\Pr(X_{n+k} = j \mid X_n = i) = P_{ij}^k$$

Here, P_{ij}^k is the $(i, j)^{th}$ entry in P^k

- Let's consider $k = 2$ transitions

$$\begin{aligned}
P_{ij}^2 &= \Pr(X_{n+2} = j \mid X_n = i) \\
&= \sum_h \Pr(X_{n+2} = j \mid X_{n+1} = h, X_n = i) \Pr(X_{n+1} = h \mid X_n = i) \\
&= \sum_h \Pr(X_{n+2} = j \mid X_{n+1} = h) \Pr(X_{n+1} = h \mid X_n = i) \\
&= \sum_h p_{ih} p_{hj}
\end{aligned}$$

- Get as a result of matrix multiplication!

Properties of Markov Chains

- What about specific path over m transitions?
- Path: $i_0 \rightarrow i_1 \rightarrow \ldots \rightarrow i_{m-1} \rightarrow i_m$
- Because every transition is given, we can just use the product rule
- $\Pr(X_m = i_m, X_{m-1} = i_{m-1}, \ldots, X_1 = i_1 \,|\, X_0 = i_0) = p_{i_0,i_1} \cdot p_{i_1,i_2} \cdot \ldots \cdot p_{i_{m-1},i_m}$

Properties of Markov Chains

- We may be interested in the long-run behavior of a Markov chain
- In order to analyze this, we must introduce some new terminology
  - A state i has **period** $k \geq 1$ if any path starting at and returning to state i must take a number of steps divisible by k
  - If $k = 1$, the state is **aperiodic**; else, state is **periodic**
  - If there is a path between any two states, the chain is **irreducible**
  - A state is **recurrent** or **transient** if the process will eventually return to that state
  - A recurrent state is **positive recurrent** if it is expected to return within a finite number of steps
  - A state is **ergodic** if it is positive recurrent and aperiodic
  - A Markov chain is ergodic if all its states are ergodic ("best behaved chains")

Stationary Distributions

- Long-run behavior of a Markov chain is given by the **stationary distribution** $\pi = (\pi_1, \ldots, \pi_n)$ where π_i is the steady-state probability that the chain is in state i

- Let $N_i(k)$ be the number of transitions into state i after the k^{th} transition

$$\pi_i = \lim_{k \to \infty} \frac{N_i(k)}{k}$$

- The stationary distribution is *invariant* by the transition matrix:

$$\pi = \pi P$$

- Any ergodic Markov chain has a unique stationary distribution

Stationary Distributions: Example

- A population of 100 people either go to McDonald's or Burger King for fast food every day. Every day, 10% of people who went to McDonald's the previous day switch to Burger King, and 20% of people who went to Burger King the previous day switch to McDonald's. After a while, the proportion of the population that goes to each restaurant stabilizes. What is the stationary distribution?

Stationary Distributions: Example

- A population of 100 people either go to McDonald's or Burger King for fast food every day. Every day, 20% of people who went to McDonald's the previous day switch to Burger King, and 30% of people who went to Burger King the previous day switch to McDonald's. After a while, the proportion of the population that goes to each restaurant stabilizes. What is the stationary distribution?

  Let m = # McDonald's, b =# Burger King
  $m = 0.8m + 0.3b$
  $b = 0.7b + 0.2m$
  Solving yields $2m = 3b$, and because $m + b = 100$, we have $m = 60,\ b = 40$

  Check that this makes sense for the Markov chain via matrix multiplication (algebra omitted)

Markov Chain Example

- I decide to flip a fair coin until I get three heads in a row.
- Draw a Markov chain representing this process
- What is the expected number of flips until I succeed?

Markov Chain Example

- I decide to flip a fair coin until I get three heads in a row.
- Draw a Markov chain representing this process
- What is the expected number of flips until I succeed?

  $X_0 = 1 + (1/2)(X_0 + X_1)$
  $X_1 = 1 + (1/2)(X_0 + X_2)$
  $X_2 = 1 + (1/2)(X_0 + X_3)$
  $X_3 = 0$
  Solving yields $X_0 = X_1 + 2$, $X_1 = X_2 + 4$, $X_1 = 12$, so $X_0 = 14$

Continuous Time Markov Chains

- Continuous time here means that transitions are not discretized to happen at $t = 1,2,\ldots$, but rather over $t \in [0,\infty)$

- Often involve waiting processes or distributions over when the next transition occurs

- Formally: $X(t)$ is a continuous time Markov chain if $X(t) \in S$ for some discrete state space and the following also holds:

$$\Pr(X(t+s) = j \,|\, X(s) = i, X(u) = h, u \in [0,s)) = \Pr(X(t+s) = j \,|\, X(s) = i)$$

- Often applied to birth and death processes, with the requirement that state transitions can only occur between adjacent integers (or radioactive decay)

Stochastic Processes Summary

- Randomness is common
- These processes give us tools to reason about randomness
- Many extensions of these frameworks are foundational for advanced topics in machine learning, data science, physics, etc.