Markov Chain & Final Review

Jan

School of Information Science and Technology,
ShanghaiTech University

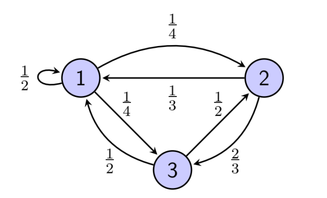
Outline

HW Problems

Final Review

Problem 1

Given a Markov chain with state-transition diagram shown as follows:



(a) Is this chain irreducible?

(b) Is this chain aperiodic? yes

(c) Find the stationary distribution of this chain.

(d) Is this chain reversible?

Problem 1 Solution

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$Q^{(2)}$ $Q^{(3)}$

Q_{12}, Q_{21}, Q_{13}, Q_{31}, Q_{23}, Q_{32} non-zero

Definition

A Markov chain with transition matrix Q is *irreducible* if for any two states i and j, it is possible to go from i to j in a finite number of steps (with positive probability). That is, for any states i, j there is some positive integer n such that the (i, j) entry of Q^n is positive. A Markov chain that is not irreducible is called *reducible*.

Problem 1 Solution

Definition

$d(1)=1$. $(2,3) \rightarrow$ state $2,3$. $\Rightarrow d(2)=d(3)=1$

For a Markov chain with transition matrix Q, the *period* of state i, denoted $d(i)$, is the greatest common divisor of the set of possible return times to i. That is,

$$d(i) = gcd\{n > 0 : Q^n_{i,i} > 0\}.$$

If $d(i) = 1$, state i is said to be *aperiodic*. If the set of return times is empty, set $d(i) = +\infty$.

Definition

A Markov chain is called *periodic* if it is irreducible and all states have period greater than 1.

A Markov chain is called *aperiodic* if it is irreducible and all states have period equal to 1.

Definition

A row vector $\mathbf{s} = (s_1, \ldots, s_M)$ such that $s_i \geq 0$ and $\sum_i s_i = 1$ is a *stationary distribution* for a Markov chain with transition matrix Q if

$$\sum_i s_i q_{i,j} = s_j.$$

(1) $\vec{s}Q = \vec{s}$

$\vec{s}(Q - I) = 0$

$\sum_i s_i = 1$

$\Rightarrow \vec{s} = \left(\frac{16}{35}, \frac{9}{35}, \frac{2}{7}\right)$

for all j, or equivalently,

$$\mathbf{s}Q = \mathbf{s}.$$

Theorem

Given a Markov chain with finite state space.

- *If such Markov chain is irreducible, then it has a unique stationary distribution. In this distribution, every state has positive probability.*
- *If such Markov chain is irreducible and aperiodic, with stationary distribution* $\mathbf{s}$ *and transition matrix* Q*, then* $P(X_n = i)$ *converges to* s_i *as* $n \to \infty$*. In terms of the transition matrix,* Q^n *converges to a matrix in which each row is* $\mathbf{s}$*.*

Problem 1 Solution

Definition

Let $Q = (q_{i,j})$ be the transition matrix of a Markov chain. Suppose there is $\mathbf{s} = (s_1, \ldots, s_M)$ with $s_i \geq 0$, $\sum_i s_i = 1$, such that

$$s_i q_{i,j} = s_j q_{j,i}$$

for all states i and j. This equation is called the *reversibility* or *detailed balance* condition, and we say that the chain is *reversible* with respect to $\mathbf{s}$ if it holds.

Theorem

Suppose that $Q = (q_{i,j})$ is a transition matrix of a Markov chain that is reversible with respect to a nonnegative vector $\mathbf{s} = (s_1, \ldots, s_M)$ whose components sum to 1. Then $\mathbf{s}$ is a stationary distribution of the chain.

Problem 1 Solution

Theorem

If for an irreducible Markov chain with transition matrix $Q = (q_{i,j})$, there exists a probability solution π to the detailed balance equations

$$\pi_i q_{i,j} = \pi_j q_{j,i}$$

for all pairs of states i, j, then this Markov chain is reversible and the solution π is the unique stationary distribution.

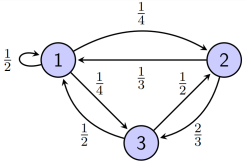
(d). NO.

If chain is reversible $\leftarrow$ $\begin{cases} \pi_1 \cdot q_{12} = \pi_1 \cdot \frac{1}{4} = \pi_2 \cdot q_{21} = \pi_2 \cdot \frac{1}{3} \\ \pi_1 \cdot \frac{1}{4} = \pi_3 \cdot \frac{1}{2} \\ \pi_2 \cdot \frac{2}{3} = \pi_3 \cdot \frac{1}{2} \end{cases}$ $\Rightarrow \pi_1 = \pi_2 = \pi_3 = 0$

$\sum_i \pi_i = 1 \Rightarrow$ irreversible

Problem 2

Given a Markov chain with state-transition diagram shown as follows:



(a) Find $P(X_9 = 3 \mid X_8 = 1)$ and $P(X_8 = 2 \mid X_7 = 3)$.

(b) If $P(X_0 = 3) = \frac{1}{2}$, find $P(X_0 = 3, X_1 = 1, X_2 = 2, X_4 = 3)$.

(c) Find $E(X_8 \mid X_6 = 2)$.

(d) Find $\mathrm{Var}(X_7 \mid X_5 = 3)$.

Problem 2 Solution

$$Q = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

(a) $P(X_9=3 \mid X_8=1) = \frac{1}{4}$. $\quad P(X_8=2 \mid X_7=3) = \frac{1}{2}$

(b). $P(\underline{X_0=3}, X_1=1, X_2=2, X_4=3)$

$= P(\underline{X_1=1}, X_2=2, X_4=3 \mid X_0=3) \cdot P(X_0=3)$

$= P(X_2=2, X_4=3 \mid X_1=1) \cdot P(X_1=1 \mid X_0=3) \cdot P(X_0=3)$

$= \underline{P(X_4=3 \mid X_2=2)}\ \underline{P(X_2=2 \mid X_1=1)}\ \underline{P(X_1=1 \mid X_0=3)} \cdot \underline{P(X_0=3)} = \frac{1}{192}$

$P(X_4=3 \mid X_2=2) = P(X_4=3, \underline{X_3=1} \mid X_2=2) + P(X_4=3, X_3=2 \mid X_2=2) + P(X_4=3, X_3=3 \mid X_2=2)$

$= P(X_4=3 \mid X_3=1) \cdot P(X_3=1 \mid X_2=2) + \cdots = \frac{1}{12}$

Problem 2 Solution

(c). $E(X_8 \mid X_6=2) = \sum_{i=1}^{3} i\, \underbrace{P(X_8=i \mid X_6=2)} = \dfrac{19}{12}.$

$$Q^{(2)} = Q^2 = \begin{bmatrix} \frac{11}{24} & \frac{1}{4} & \frac{7}{24} \\ \frac{1}{2} & \frac{5}{12} & \frac{1}{12} \\ \frac{5}{12} & \frac{1}{8} & \frac{11}{24} \end{bmatrix}$$

Problem 2 Solution

(d) $E(X_7 \mid X_5=3) = ?$ $\quad E(X_7^2 \mid X_5=3) = ?$

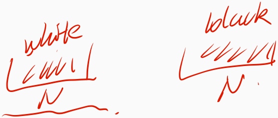
$$E(X_7 \mid X_5=3) = \sum_{i=1}^{3} i\, P(X_7=i \mid X_5=3) = \frac{49}{24}.$$

$$E(X_7^2 \mid X_5=3) = \sum_{i=1}^{3} i^2 P(X_7=i \mid X_5=3) = \frac{121}{24}.$$

$$\Rightarrow \quad Var(X_7 \mid X_5=3) = \frac{503}{576}$$

Problem 2 Solution

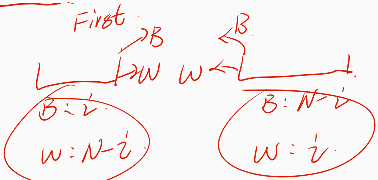
Problem 3



There are two urns with a total of $2N$ distinguishable balls. Initially, the first urn has N white balls and the second urn has N black balls. At each stage, we pick a ball at random from each urn and interchange them. Let X_n be the number of black balls in the first urn at time n. This is a Markov chain on the state space $\{0, 1, \ldots, N\}$.

(a) Find the transition probabilities of the chain.

(b) Find the stationary distribution of the chain.



Problem 3 Solution

(a). We can know:

$$P(X_{n+1}=1 \mid X_n=0)=1, \quad P(X_{n+1}=N-1 \mid X_n=N)=1.$$

If $X_n=i$, $i\in\{1,\dots,N-1\}$, $X_{n+1}\in\{i-1,i,i+1\}$.

① $X_{n+1}=i-1$. First: black. Second: white.

$$P(X_{n+1}=i-1 \mid X_n=i)=\frac{i}{N}\cdot\frac{i}{N}=\frac{i^2}{N^2}$$

② $X_{n+1}=i+1$. First: white. Second: Black. $P(X_{n+1}=i+1 \mid X_n=i)=\frac{(N-i)^2}{N^2}$.

③ $X_{n+1}=i$. ------. $P(X_{n+1}=i \mid X_n=i)=\frac{2i(N-i)}{N^2}$.

Problem 3 Solution

(b) Note two important observations:

- The Markov chain is irreducible.
- The Markov chain is a step-by-step analogy to the story of the Hypergeometric distribution.

These two observations lead to the guess of the stationary distribution as $\mathbf{s} = [s_0, \ldots, s_i, \ldots, s_N]$ with the PMF of the Hypergeometric distribution, i.e.,

$$s_i = \frac{\binom{N}{i}\binom{N}{N-i}}{\binom{2N}{N}}.$$

Due to irreducibility, we justify the proposed distribution by checking the detailed balance equation:

$$s_i q_{ij} = s_j q_{ji}, \quad \forall i, j \in \{0, 1, \ldots, N\}.$$

For state $i = 0$, the only non-trivial case we need to check is state $j = 1$ since there is no direct transition to other states. Therefore, we have that

$$s_0 q_{01} = s_1 q_{10},$$

which simplifies as:

$$\frac{\binom{N}{0}\binom{N}{N}}{\binom{2N}{N}} \cdot 1 = \frac{\binom{N}{1}\binom{N}{N-1}}{\binom{2N}{N}} \cdot \frac{1^2}{N^2}.$$

Problem 3 Solution

We use the fact that $\binom{N}{1} = N$ and $\binom{N}{N-1} = N$ to simplify:

$$\frac{1}{\binom{2N}{N}} = \frac{N \cdot N}{\binom{2N}{N} \cdot N^2}.$$

Similarly, for $i = N$, the only non-trivial case is for $j = N - 1$, which is true using the same calculations. For $i = 1, \ldots, N - 1$, non-trivial cases happen for $j = i - 1$ and $j = i + 1$. We are going to show that the equation holds for $1 < i \leq N - 1$ and for $j = i - 1$ (all other calculations are similar or we have already shown). We have that

$$s_i q_{i,i-1} = s_{i-1} q_{i-1,i}.$$

Substituting the values:

$$\frac{\binom{N}{i}\binom{N}{N-i}}{\binom{2N}{N}} \cdot \frac{i^2}{N^2} = \frac{\binom{N}{i-1}\binom{N}{N-i+1}}{\binom{2N}{N}} \cdot \frac{(N-i+1)^2}{N^2}.$$

Expanding the binomial coefficients:

$$\frac{\binom{N}{i} \cdot \binom{N}{N-i} \cdot i^2}{N!} = \frac{\binom{N}{i-1} \cdot \binom{N}{N-i+1} \cdot (N-i+1)^2}{N!}.$$

Rewriting in factorial form:

$$\frac{\frac{N!}{i!(N-i)!} \cdot \frac{N!}{(N-i)!i!} \cdot i^2}{N!} = \frac{\frac{N!}{(i-1)!(N-i+1)!} \cdot \frac{N!}{(N-i+1)!(i-1)!} \cdot (N-i+1)^2}{N!}.$$

Simplifying:

$$\frac{N!}{(i-1)!(N-i)!} \cdot \frac{1}{(N-i)!} \cdot \frac{i^2}{i} = \frac{N!}{(i-1)!(N-i)!} \cdot \frac{1}{(i-1)!} \cdot (N-i+1)^2.$$

Therefore:

$$s_i q_{i,i-1} = s_{i-1} q_{i-1,i}.$$

Hence, we have shown that the chain is reversible, and s is the stationary distribution.

Method 2:

$\sum_i s_i = 1$;

$s_i q_{i,i+1} = s_{i+1} q_{i+1,i}$

$s_i \cdot \frac{(N-i)^2}{N^2} = s_{i+1} \cdot \frac{(i+1)^2}{N^2}$

$\Rightarrow s_i(s_0)$

$\Rightarrow s_0 = \boxed{}$

$\Rightarrow s = \text{- - - -}$

Problem 3 Solution

Outline

HW Problems

Final Review

Outline of Topics

Part I

- Probability & Counting
- Conditional Probability
- Random Variables
- Expectations
- Continuous Random Variables

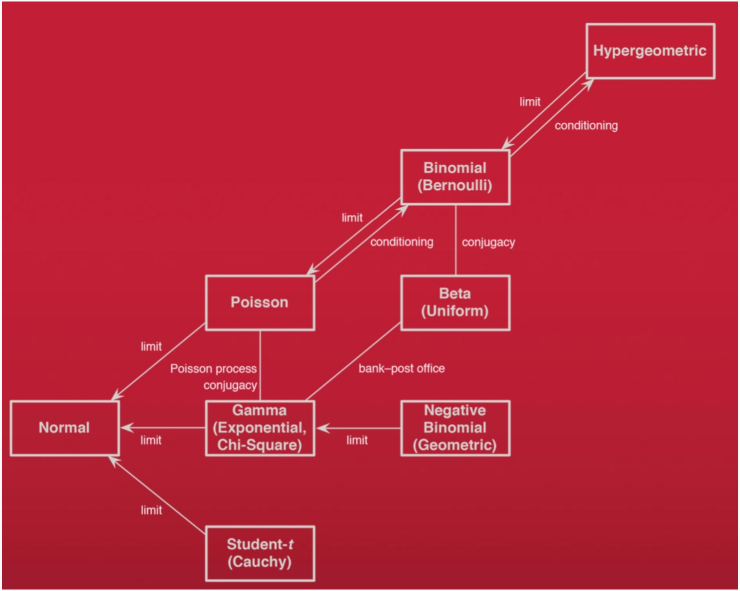
Part II

- Joint Distributions
- Transformations
- Monto Carlo Methods & Concentration Inequalities
- Statistical Inference
- Markov Chains

Random Variables

- First success & Geometric
- Exponential & Poisson & Gamma
- Bernoullis
- Uniform
- Normal & MVN
- Beta & Binomial

Random Variables



Random Variables

- Ordering
  - Order statistics
  - Max & Min operators
- Jointness
  - Independence: pairwise & conditional
  - Correlation & Covariance
- Transformation
  - Change of variables & convolution
  - Inverse transform method
- Relationship
  - Discrete to continuous with δ-step method
  - Conjugacy: Beta-Binomial, Normal-Normal, Gamma-Pois
  - Connection: Beta-Gamma, Uniform-Beta, Binomial-Gamma

Independence

- $P(X, Y) = P(X)P(Y)$ ⟺
- $P(X|Y) = P(X)$ with $P(Y) \neq 0$
- Factorization of PDF $f_{X,Y}(x, y)$ and MGF $M_{X,Y}(t)$
- $E(XY) = E(X)E(Y)$
- $\mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y) = 0$ $\not\Rightarrow$ independence

Tools

- Bayes' rule, LOTP & LOTE, LOTUS: 1D & 2D
- Indicator & linearity of expectation
- First-step analysis & recursive equations
- Conditional expectation: Adam's & Eve's law
- Generating functions: PGF & MGF
- Symmetry
    - $X+Y, XY, |X-Y|, \frac{X}{Y}, \frac{X}{X+Y}, \frac{Y}{X+Y}$
    - The property of i.i.d. continuous random variables
    - Normal distributions

Lecture 6: Joint Distributions Summary

1. **Discrete Multivariate R.V.s**
   - **Joint CDF:** $F_{X,Y}(x, y) = P(X \leq x, Y < y)$
   - **Joint PMF:** $P_{X,Y}(x, y) = P(X = x, Y = y)$
   - **Marginal PMF:** $P_X(x) = \sum_y P_{X,Y}(x, y)$
   - **Conditional PMF:** $P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)}$
   - **Independence:** $P_{X,Y}(x, y) = P_X(x)P_Y(y)$
2. **Continuous Multivariate R.V.s**
   - **Joint CDF:** $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$
   - **Joint PDF:** $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$
   - **Marginal PDF:** $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$
   - **Conditional PDF:** $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$
   - **Independence:** $f_{X,Y}(x, y) = f_X(x)f_Y(y)$

3. Covariance and Correlation

$Cov(X+Y, Z) = Cov(X, Z) + Cov(Y,$

- **Covariance:** $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$
- **Correlation:** $\rho(X, Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$
- **Properties:** Covariance properties, correlation bounds, independence implies uncorrelated.

4. Multinomial Distribution

- **Story:** Allocation of n objects into k categories with probabilities $p_1, p_2, \ldots, p_k$.
- **Joint PMF:** $P(X_1 = n_1, \ldots, X_k = n_k) = \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k}$
- **Marginals:** $X_i \sim \text{Binomial}(n, p_i)$
- **Covariance:** $\text{Cov}(X_i, X_j) = -np_ip_j$

5. Multivariate Normal

- **Definition:** Linear combinations of components are Normal.
- **Parameters:** Mean vector μ, covariance matrix Σ.
- **Properties:** Uncorrelated implies independent, subvectors are Normal.

6. Change of Variables

- **1D Transformation:** $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right|$
- **nD Transformation:** Using Jacobian matrix $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\partial x}{\partial y} \right|$
- **Example:** Box-Muller transformation for generating Bivariate Normal.

7. **Convolutions**

- **Discrete:** $P(T = t) = \sum_x P(X = x)P(Y = t - x)$
- **Continuous:** $f_T(t) = \int_{-\infty}^{\infty} f_X(x)f_Y(t - x)dx$

Lecture 7: Monte Carlo Methods Summary

1. **Sampling: Random Variable Generation**
   - **Inverse Transform Method**:
     - For continuous distributions: $X = F^{-1}(U)$ where $U \sim \text{Unif}(0, 1)$.
     - For discrete distributions: Find the smallest k such that $U \leq F(x_k)$.
   - **Acceptance-Rejection Method**:
     - Generate $Y \sim g$ and $U \sim \text{Unif}(0, 1)$.
     - Accept Y if $U \leq \frac{f(Y)}{cg(Y)}$.
2. **Sampling: Random Vector Generation**
   - **Change of Variables**:
     - Use Jacobian determinant to transform variables.
     - Example: Generating uniform distribution over an ellipse or sphere.

3. **Monte Carlo Integration**
   - **Estimation of Integrals**:
     - Use sample mean to approximate integrals: $\int_a^b g(x)dx \approx \frac{b-a}{n}\sum_{i=1}^n g(X_i)$.
   - **Importance Sampling**:
     - Re-weight samples from a proposal distribution g to estimate integrals with respect to f.
4. **Asymptotic Analysis: Law of Large Numbers**
   - **Weak Law of Large Numbers (WLLN)**:
     - $\frac{1}{n}\sum_{i=1}^n X_i$ converges in probability to μ.
   - **Strong Law of Large Numbers (SLLN)**:
     - $\frac{1}{n}\sum_{i=1}^n X_i$ converges almost surely to μ.

5. Non-asymptotic Analysis: Inequalities

- **Jensen's Inequality**:
  - For convex functions: $E[g(X)] \geq g(E[X])$.
- **Markov's Inequality**:
  - $P(|X| \geq a) \leq \frac{E[|X|]}{a}$.
- **Chebyshev's Inequality**:
  - $P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$.
- **Chernoff's Bound**:
  - $P(X \geq a) \leq \inf_{t>0} \frac{E[e^{tX}]}{e^{ta}}$.
- **Hoeffding's Inequality**:
  - For bounded independent variables:
    $P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq 2e^{-\frac{2n\epsilon^2}{(b-a)^2}}$.

1. Conditional Expectation Given an Event

- **Definition:** The expected value of a random variable given that a specific event has occurred.
- **Formulas:**
  - For discrete random variables:

$$E(Y|A) = \sum yP(Y = y|A)$$

  - For continuous random variables:

$$E(Y|A) = \int yf_{Y|A}(y)dy$$

- **Law of Total Expectation (LOTE):**

$$E(Y) = \sum E(Y|A_i)P(A_i) \quad \text{for a partition } \{A_i\}.$$

2. Conditional Expectation Given a Random Variable

- **Definition:** $E(Y|X)$ is a random variable that is a function of X, denoted $g(X)$, where $g(x) = E(Y|X = x)$.

$X = x$

- **Properties:**
  - **Linearity:**

$$E(Y_1 + Y_2|X) = E(Y_1|X) + E(Y_2|X)$$

  - **Taking Out What's Known:**

$$E(h(X)Y|X) = h(X)E(Y|X)$$

- **Adam's Law:**

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

- **Eve's Law:**

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

3. Prediction & Estimation

- **Minimum Mean Square Error (MMSE) Estimator:** $E(Y|X)$ is the best predictor of Y given X in terms of minimizing the mean squared error.
- **Linear Least Squares Estimation (LLSE):**

$$\hat{Y} = E(Y) + \frac{Cov(X, Y)}{Var(X)}(X - E(X))$$

- **Geometric Interpretation:** $E(Y|X)$ is the orthogonal projection of Y onto the space of functions of X.

4. Application Case: Kalman Filter

- **Purpose:** Optimal recursive data processing algorithm for estimating the state of a dynamic system from a series of noisy measurements.
- **Steps:**
  - **Prediction Step:** Predict the state at the next time step based on previous state estimates.
  - **Update Step:** Update the prediction with the actual measurement to get a new state estimate.
- **Optimality:** Kalman filter is optimal for linear systems with Gaussian noise, providing the MMSE estimate.

Lecture 9: Statistical Inference Summary

1. **Overview of Statistical Inference:**
   - **Statistical Inference:**
     - Process of extracting information from data.
     - Core parts: Point Estimation, Interval Estimation, Hypothesis Testing.
   - **Bayesian vs. Frequentist Approaches:**
     - Bayesian: Parameters are random variables with prior distributions.
     - Frequentist: Parameters are fixed unknown constants.
2. **Our Focus: Bayesian Statistical Inference:**
   - **Bayesian Inference:**
     - Use of prior distributions, likelihood models, and posterior distributions.
     - Maximum A Posteriori (MAP) estimation.

3. **Beta and Gamma Distributions:**
   - **Beta Distribution:**
     - PDF, properties, and relationship with the binomial distribution.
   - **Gamma Distribution:**
     - PDF, properties, and relationship with the exponential and Poisson distributions.
   - **Beta-Gamma Connection:**
     - Sum of Gamma variables and their ratios.
4. **Conjugate Prior: A Weapon of Bayesian:**
   - **Conjugate Priors:**
     - Simplify posterior calculations.
     - Beta-Binomial and Dirichlet-Multinomial conjugacy.

Stochastic Processes

- **Definition**: A stochastic process is a collection of random variables $\{X_t, t \in I\}$ defined on a common state space S. If I is discrete, it's a discrete-time stochastic process; if I is continuous, it's a continuous-time stochastic process.

Markov Model

- **Model Selection**: Discrete time stochastic process balances complexity and simplicity.
- **Motivation**: Introduced by Andrey Markov in 1906, between independent and completely dependent random variables.
- **Components**: Sequence of random variables, state space, and Markov property (future depends only on present).
- **Classification**: Continuous Markov Process, Discrete-Time Markov Chain and so on.

Markov Property and Transition Matrix

- **Markov Property**: $P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = P(X_{n+1} = j | X_n = i)$.
  For time-homogeneous Markov chains,
  $P(X_{n+1} = j | X_n = i) = q_{i,j}$ (constant independent of n).
- **Transition Matrix**: For a Markov chain with state space $\{1, 2, \cdots, M\}$, the $M \times M$ matrix $Q = (q_{i,j})$ is the transition matrix.
- **Examples**: Rainy-Sunny Markov Chain, Markov chain in Russian literature study, Gambler's Ruin, Coupon Collector, and Random Walk on a Graph.

Basic Computations

- **n-step Transition Probability**: $q_{i,j}^{(n)} = P(X_n = j | X_0 = i)$ is the (i, j) entry of Q^n. Chapman-Kolmogorov relationship: $q_{i,j}^{(m+n)} = \sum_k q_{i,k}^{(m)} q_{k,j}^{(n)}$. $Q^{(2)}$
- **Distribution of X_n**: If initial distribution is $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_M)$ with $\alpha_i = P(X_0 = i)$, then $P(X_n = j) = (\alpha Q^n)_j$.
- **Examples**: Calculations like $P(X_3 = \frac{1}{\varepsilon} | X_2 = 1)$, $P(X_4 = 3 | X_3 = 2)$, $P(X_0 = 1, X_1 = 2, X_2 = 3)$, $P(X_2 = j | X_0 = 1)$ for $j = 1, 2, 3$, and $E(X_2 | X_0 = 1)$.

Classification of States

- **Recurrent and Transient States**: State i is recurrent if $\sum_{n=1}^{\infty} p_{i,i}^{(n)} = \infty$; transient if $\sum_{n=1}^{\infty} p_{i,i}^{(n)} < \infty$.
- **Irreducible and Reducible Chain**: Irreducible if for any two states i and j, can go from i to j in finite steps. In an irreducible Markov chain with finite state space, all states are recurrent.
- **Period**: The period of state i, $d(i)$, is the greatest common divisor of possible return times to i. If $d(i) = 1$, state i is aperiodic.
- **Examples**: Classification in Gambler's Ruin and Coupon Collector Markov chains.

Stationary Distribution

- **Definition**: A row vector $s = (s_1, \cdots, s_M)$ is a stationary distribution for a Markov chain with transition matrix Q if $sQ = s$.
- **Examples**: For double stochastic matrix, uniform distribution is stationary. For two-state Markov chain, stationary distribution can be calculated. In an irreducible Markov chain with finite state space, there is a unique stationary distribution. If aperiodic, $P(X_n = i)$ converges to stationary distribution as $n \to \infty$.

Reversibility

- **Definition**: A Markov chain is reversible with respect to s if $s_i q_{i,j} = s_j q_{j,i}$ for all states i and j.
- **Reversibility Implies Stationary**: If reversible with respect to s, then s is a stationary distribution.
- **Examples**: If transition matrix is symmetric, uniform distribution is the unique stationary distribution. For random walk on undirected graph, stationary distribution can be calculated using detailed balance equation.

Model-Based Problems

- Birthday problem: static & dynamic
- Sequence of coin tosses: biased or not
- Gambler's ruin & Random walk
- Coupon collector: given total number or not
- Pattern matching: coin or dice
- Chicken-egg problem & Poisson process
- Bank-post office

Model-Free Problems

- Computation via definitions
  - PMF, PDF, CDF, Joint distribution
  - Expectation, PGF, MGF
  - Markov chains
- Approximation
  - CLT & Law of Large Number
  - Poisson approximation & Law of Small Number
  - Non-asymptotic inequalities
- Estimation
  - MLE & MAP
  - Confidence interval
  - MMSE & LLSE

A Final Note

Please, Check Out the Previous Example Papers.

Please, Check Out the Slides.

Please, Check Out the HWs.

Please, Check Out the Textbooks.