TA Lecture 13 - Markov Chain & Final Review

June 11, 13

School of Information Science and Technology,
ShanghaiTech University

上海科技大学
ShanghaiTech University

Outline

HW Problems

Final Review

Previous Final Exams

Problem 1

Given a coin with the probability p of landing heads. p is unknown and we need to estimate its value through data. In our data collection model, we have n independent tosses, the result of each toss is either Head or Tail. Let X denote the number of heads in the total n tosses. Now we conduct experiments to collect data and find $X = k$. Then we need to find $\hat{p}$, the estimation of p.

Problem 1 Continued

(a) Assume p is an unknown constant. Find $\hat{p}$ through the MLE (Maximum Likelihood Estimation) rule.

(b) Assume p is a random variable with a prior distribution $p \sim \text{Beta}(a, b)$, where a and b are known constants. Find $\hat{p}$ through the MAP (Maximum a Posterior Probability) rule.

(c) Assume p is a random variable with a prior distribution $p \sim \text{Beta}(a, b)$, where a and b are known constants. Find $\hat{p}$ through the MMSE (Minimal Mean Squared Error) rule.

Problem 1 Solution

Problem 2

Let X be the height of a randomly chosen adult man, and Y be his father's height, where X and Y have been standardized to have mean 0 and standard deviation 1. Suppose that (X, Y) is Bivariate Normal, with $X, Y \sim \mathcal{N}(0, 1)$ and $\text{Corr}(X, Y) = \rho$.

Problem 2 Continued

(a) Let $y = ax + b$ be the equation of the best line for predicting Y from X (in the sense of minimizing the mean squared error), e.g., if we were to observe $X = 1.3$ then we would predict that Y is $1.3a + b$. Now suppose that we want to use Y to predict X, rather than using X to predict Y. Give and explain an intuitive guess for what the slope is of the best line for predicting X from Y.

(b) Find a constant c (in terms of ρ) and an r.v. V such that $Y = cX + V$, with V independent of X. Hint: Start by finding c such that $\text{Cov}(X, Y - cX) = 0$.

(c) Find a constant d (in terms of ρ) and an r.v. W such that $X = dY + W$, with W independent of Y.

(d) Find $E(Y \mid X)$ and $E(X \mid Y)$.

(e) Reconcile (a) and (d), giving a clear and correct intuitive explanation.

Problem 2 Solution

Problem 3

Two chess players, Vishy and Magnus, play a series of games. Given p, the game results are i.i.d. with probability p of Vishy winning, and probability $q = 1 - p$ of Magnus winning (assume that each game ends in a win for one of the two players). But p is unknown, so we will treat it as an r.v. To reflect our uncertainty about p, we use the prior $p \sim \text{Beta}(a, b)$, where a and b are known positive integers and $a \geq 2$.

Problem 3 Continued

(a) Find the expected number of games needed in order for Vishy to win a game (including the win). Simplify fully; your final answer should not use factorials or Γ.

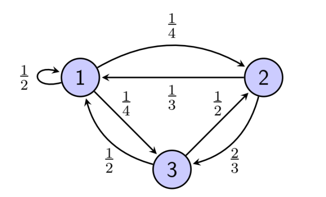
(b) Explain in terms of independence vs. conditional independence the direction of the inequality between the answer to (a) and $1 + E(G)$ for $G \sim \text{Geom}\left(\frac{a}{a+b}\right)$.

(c) Find the conditional distribution of p given that Vishy wins exactly 7 out of the first 10 games.

Problem 3 Solution

Problem 4

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



(a) Is this chain irreducible?

(b) Is this chain aperiodic?

(c) Find the stationary distribution of this chain.

(d) Is this chain reversible?

Problem 4 Solution

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 4 Solution

Definition

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.

Problem 4 Solution

Definition

A row vector $\mathbf{s} = (s_1, ..., s_M)$ 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

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

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 4 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 4 Solution

Theorem

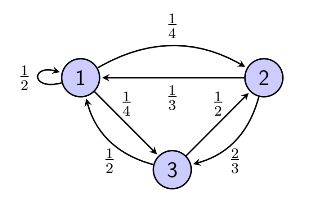
If for an irreducible Markov chain with transition matrix $Q = (q_{i,j})$, there exists a probability solution π to the detailed balance equations

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

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

Problem 5

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



(a) Find $P(X_3 = 3 \mid X_2 = 2)$ and $P(X_4 = 1 \mid X_3 = 2)$.

(b) If $P(X_0 = 2) = \frac{2}{5}$, find $P(X_0 = 2, X_1 = 3, X_2 = 1)$.

(c) Find $P(X_2 = 1 \mid X_0 = 2), P(X_2 = 2 \mid X_0 = 2)$, and $P(X_2 = 3 \mid X_0 = 2)$.

(d) Find $E(X_2 \mid X_0 = 2)$.

Problem 5 Solution

Problem 6

A fair coin is flipped repeatedly. We use H to denote the "Head appeared" and T to denote the "Tail appeared".

(a) What is the expected number of flips until the pattern HTHT is observed?

(b) What is the expected number of flips until the pattern THTT is observed?

(c) What is the probability that pattern HTHT is observed earlier than THTT?

Problem 6 Solution

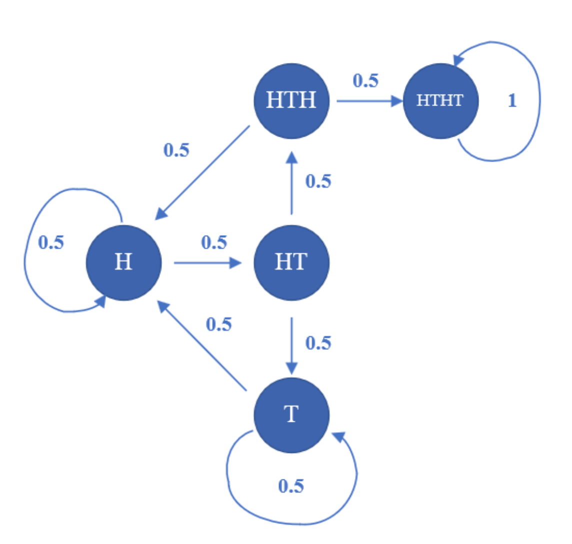


Figure 1: 6(1)

Problem 6 Solution

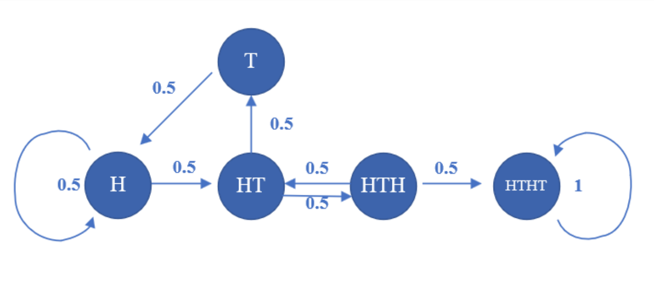


Figure 2: 6(1)

Problem 6 Solution

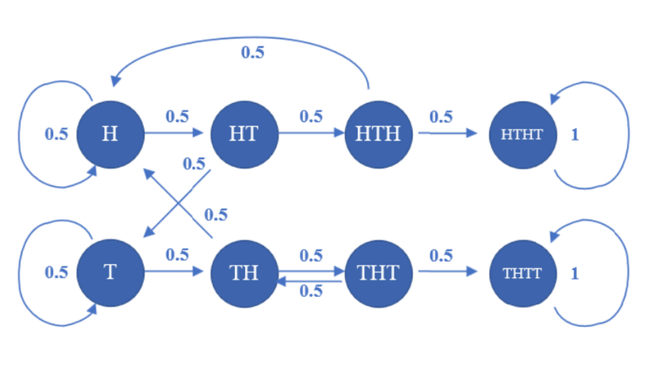


Figure 3: 6(1)

Outline

Outline of Topics

Part I

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

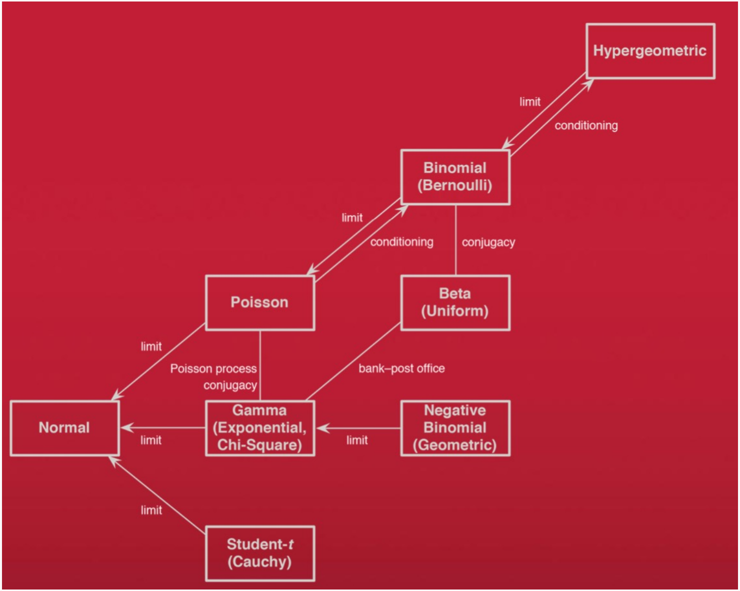
Part II

- 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 $M_{X,Y}(t)$
- $E(XY) = E(X)E(Y)$
- $\text{Corr}(X, Y) = \text{Cov}(X, Y) = 0$

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

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

Outline

HW Problems

Final Review

Previous Final Exams

Q1 2021

- Please describe the difference and connection between probability and statistics.
- Please describe the pros and cons of Bayesian statistical inference and classical statistical inference.

Q1 Solution

Q2 2021

Let X be a discrete r.v. whose distinct possible values are $x_0, x_1, \ldots,$ and let $p_k = P(X = x_k)$. The entropy of X is $H(X) = \sum_{k=0}^{\infty} p_k \log(1/p_k)$

- Find $H(X)$ for $X \sim Geom(p)$.
- Let X and Y be i.i.d. discrete r.v.s. Show that $P(X = Y) \geq 2^{-H(X)}$

Q2 Solution

Q3 2021

Let $X_1 \sim Expo(\lambda_1)$,$X_2 \sim Expo(\lambda_2)$ and$X_3 \sim Expo(\lambda_3)$ be independent.

- Find$E(X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3)$ in terms of $\lambda_1, \lambda_2, \lambda_3$.
- Find $P(X_1 = min(X_1, X_2, X_3))$

Q3 Solution

Q4 2021

Let $X \sim Gamma(a, \lambda)$, $Y \sim Gamma(b, \lambda)$. Assume X and Y are independent.

- Find the joint distribution of $T = X + Y$ and $W = \frac{X}{X+Y}$.
- Find the distribution of T and W respectively.
- Find $E(W)$

Q4 Solution

Q5 2021

Instead of predicting a single value for the parameter, we give an interval that is likely to contain the parameter: A $1-\delta$ confidence interval for a parameter p is an interval $[\hat{p}-\epsilon, \hat{p}+\epsilon]$ such that $\Pr(p \in [\hat{p}-\epsilon, \hat{p}+\epsilon]) \geq 1-\delta$. Now we toss a coin with probability p landing heads and probability $1-p$ landing tails. The parameter p is unknown and we need to estimate its value from experiments results. We toss such coin N times. Let $X_i = 1$ if the i th result is head, otherwise 0. We estimate p by using $\hat{p} = \frac{X_1 + \ldots + X_N}{N}$. Find the $1-\delta$ confidence interval for p, then discuss the impacts of δ and N.

Q5 2021 Continued

Hint: You can use the following Hoeffding bound: Let the random variables $X_1, X_2, \ldots, X_n$ be independent with $\mathrm{E}(X_i) = \mu, a \leq X_i \leq b$ for each $i = 1, \ldots, n$, where a, b are constants. Then for any $\epsilon \geq 0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq 2e^{-\frac{2n\epsilon^2}{(b-\alpha)^2}}$$

Q5 Solution

Q6 2021

Given a coin with the probability p of landing heads. p is unknown and we need to estimate its value through data. In our data collection model, we have n independent tosses, the result of each toss is either Head or Tail. Let X denote the number of heads in the total n tosses. Now we conduct experiments to collect data and find $X = k$. Then we need to find $\hat{p}$, the estimation of p.

Q6 2021 Continued

- Assume p is an unknown constant. Find $\hat{p}$ through the MLE (Maximum Likelihood Estimation) rule.
- Assume p is a random variable with a prior distribution $p \sim Beta(a, b)$, where a and b are known constants. Find $\hat{p}$ through the MAP (Maximum a Posterior Probability) rule.
- Assume p is a random variable with a prior distribution $p \sim Beta(a, b)$, where a and b are known constants. Find $\hat{p}$ through the MMSE (Minimal Mean Squared Error) rule.

Q6 Solution

Q7 2021

We know that the MMSE of X given Y is given by $g(Y) = \mathrm{E}[X \mid Y]$. We also know that the Linear Least Square Estimate (LLSE) of X given Y, denoted by $L[Y \mid X]$, is shown as follows:

$$L[Y \mid X] = \mathrm{E}(Y) + \frac{\mathsf{Cov}(X, Y)}{\mathsf{Var}(X)}(X - \mathrm{E}(X))$$

Q7 2021 Continued

Now we wish to estimate the probability of landing heads, denoted by θ, of a biased coin. We model θ as the value of a random variable Θ with a known prior PDF $f_\Theta \sim \text{Unif}(0, 1)$. We consider n independent tosses and let X be the number of heads observed.

- Show that $\mathrm{E}[(\Theta - \mathrm{E}[\Theta \mid X])h(X)] = 0$ for any real function $h(\cdot)$.
- Find the MMSE $\mathrm{E}[\Theta \mid X]$ and LLSE $L[\Theta \mid X]$. (Eve's law: $\text{Var}(Y) = \mathrm{E}(\text{Var}(Y \mid X)) + \text{Var}(\mathrm{E}(Y \mid X))$.)

Q7 Solution

Q8 2021

Show the following inequalities.

(a) Let $X \sim \text{Pois}(\lambda)$. If there exists a constant $a > \lambda$, then

$$\mathbb{P}(X \geq a) \leq \frac{e^{-\lambda}(e\lambda)^a}{a^a}.$$

(b) Let X be a random variable with finite variance σ^2. Then for any constant $a > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{2\sigma^2}{\sigma^2 + a^2}.$$

Q8 Solution

Q9 2021

Let $X \sim \mathcal{N}(0, 1), Y \sim \mathcal{N}(0, 1); X$ and Y are independent. Now let $Z_1 = \sin(X + Y), Z_2 = \cos(X + Y)$.

(a) Find $\mathrm{E}(Z_1)$ and $\mathrm{E}(Z_2)$.

(b) Find $\mathrm{Var}(Z_1)$ and $\mathrm{Var}(Z_2)$.

Q9 Solution

Q1 2022

- Please describe the difference and connection between probability and statistics.
- Please describe the pros and cons of Bayesian statistical inference and classical statistical inference.

Q2 2022

Let X and Y be two continuous random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} x + cy^2 & \text{if } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of constant c.

(b) Find the joint probability $P(0 \le X \le 1/2, 0 \le Y \le 1/2)$.

Q2 Solution

Q3 2022

Let X and Y be two continuous random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 6xy & \text{if } 0 \leq x \leq 1, 0 \leq y \leq \sqrt{x} \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the marginal distributions of X and Y. Are X and Y independent?
(b)) Find $E[X \mid Y = y]$ and $\text{Var}[X \mid Y = y]$ for $0 \leq y \leq 1$.

Q3 Solution

Q4 2022

Given a coin with the probability p of landing heads. p is unknown and we need to estimate its value through data. In our data collection model, we have n independent tosses, the result of each toss is either Head or Tail. Let X denote the number of heads in the total n tosses. Now we conduct experiments to collect data and find $X = k$. Then we need to find $\hat{p}$, the estimation of p.

Q4 2022 Continued

- Assume p is an unknown constant. Find $\hat{p}$ through the MLE (Maximum Likelihood Estimation) rule.
- Assume p is a random variable with a prior distribution $p \sim Beta(a, b)$, where a and b are known constants. Find $\hat{p}$ through the MAP (Maximum a Posterior Probability) rule.
- Assume p is a random variable with a prior distribution $p \sim Beta(a, b)$, where a and b are known constants. Find $\hat{p}$ through the MMSE (Minimal Mean Squared Error) rule.

Q5 2022

We know that the MMSE of X given Y is given by $g(Y) = \mathrm{E}[X \mid Y]$. We also know that the Linear Least Square Estimate (LLSE) of X given Y, denoted by $L[Y \mid X]$, is shown as follows:

$$L[Y \mid X] = \mathrm{E}(Y) + \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}(X - \mathrm{E}(X))$$

Now we wish to estimate the probability of landing heads, denoted by θ, of a biased coin. We model θ as the value of a random variable Θ with a known prior PDF $f_\Theta \sim \text{Unif}(0, 1)$. We consider n independent tosses and let X be the number of heads observed.

- Show that $\mathrm{E}[(\Theta - \mathrm{E}[\Theta \mid X])h(X)] = 0$ for any real function $h(\cdot)$.
- Find the MMSE $\mathrm{E}[\Theta \mid X]$ and LLSE $L[\Theta \mid X]$. (Eve's law: $\text{Var}(Y) = \mathrm{E}(\text{Var}(Y \mid X)) + \text{Var}(\mathrm{E}(Y \mid X))$.)

Q6 2022

Laplace's law of succession says that if $X_1, X_2, \ldots, X_{n+1}$ are conditionally independent Bern (p) r.v.s given p, but p is given a Unif$(0,1)$ prior to reflect ignorance about its value, then

$$P\left(X_{n+1} = 1 \mid X_1 + \cdots + X_n = k\right) = \frac{k+1}{n+2}.$$

As an example, Laplace discussed the problem of predicting whether the sun will rise tomorrow, given that the sun did rise every time for all n days of recorded history; the above formula then gives $(n+1)/(n+2)$ as the probability of the sun rising tomorrow (of course, assuming independent trials with p unchanging over time may be a very unreasonable model for the sunrise problem).

Q6 2022 Continued

(a) Find the posterior distribution of p given $X_1 = x_1, \ldots, X_n = x_n$, and show that it only depends on the sum of the $x_j, j \in \{1, \ldots, n\}$.
(b) Prove Laplace's law of succession, using a form of LOTP to find

$$P\left(X_{n+1} = 1 \mid X_1 + \cdots + X_n = k\right)$$

by conditioning on p.

Q6 Solution

Q7 2022

A handy rule of thumb in statistics and life is as follows: Conditioning often makes things better. This problem explores how the above rule of thumb applies to estimating unknown parameters. Let θ be an unknown parameter that we wish to estimate based on data $X_1, \ldots, X_n$ (these are r.v.s before being observed, and then after the experiment they "crystallize" into data). In this problem, θ is viewed as an unknown constant, and is not treated as an r.v. as in the Bayesian statistical inference. Let T_1 be an estimator for θ (this means that T_1 is a function of $X_1, \ldots, X_n$ which is being used to estimate θ). A strategy for improving T_1 (in some problems) is as follows.

Q7 2022 Continued

Suppose that we have an r.v. R such that $T_2 = E(T_1 \mid R)$ is a function of $X_1, \ldots, X_n$ (in general, $E(T_1 \mid R)$ might involve unknowns such as θ but then it couldn't be used as an estimator). Also suppose that $P(T_1 = T_2) < 1$, and that $E\left(T_1^2\right)$ is finite.
(a) Use Jensen's inequality and Adam's Law to show that T_2 is better than T_1 in the sense that the mean squared error is less, i.e.,

$$E\left[(T_2 - \theta)^2\right] < E\left[(T_1 - \theta)^2\right].$$

Q7 2022 Continued

(b) The bias of an estimator T for θ is defined to be $b(T) = E(T) - \theta$. An important identity in statistics, a form of the bias-variance trade-off, is that the mean squared error equals the variance plus the squared bias:

$$E\left[(T-\theta)^2\right] = \mathrm{Var}(T) + (b(T))^2.$$

Use this identity and Eve's law to give an alternative proof of the result from (a).

(c) Now suppose that $X_1, \ldots, X_n$ are i.i.d. with mean θ, and consider the special case $T_1 = X_1, R = \sum_{j=1}^{n} X_j$. Find T_2 in a simplified form, and check that it has a lower mean squared error than T_1 for $n \geq 2$. Also, explain what happens to T_1 and T_2 as $n \to \infty$.

Q7 Solution

Let X and Y be two independent random variables satisfying first success distribution $FS(p)$.
(a) (5 points) Define $Z_1 = X - Y$. Find the PMF of Z_1 and $E(Z_1)$.
(b) (5 points) Define $Z2 = \frac{X}{Y}$. Find the PMF of Z_2 and $E(Z_2)$.

Q8 Solution

Q9 2022

A scientist makes two measurements X, Y, considered to be i.i.d. random variables.
(a) If $X, Y \sim \mathcal{N}(0, 1)$. Find the correlation between the larger and smaller of the values, i.e., $\text{Corr}(\max(X, Y), \min(X, Y))$.
(b) If $X, Y \sim \text{Unif}(0, 1)$. Find the correlation between the larger and smaller of the values, i.e., $\text{Corr}(\max(X, Y), \min(X, Y))$.

Q9 Solution

Q1 2023

(5 points) Please describe the pros and cons of **Bayesian statistical inference** and **Classical statistical inference**. Then explain why conjugate priors are important for Bayesian statistical inference.

Q1 solution

Q2 2023

(10 points) Let X, Y be two continuous random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} cx^2y & \text{if } 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) (5 points) Find the value of constant c.

(b) (5 points) Find the conditional probability $P(Y \le \frac{X}{4} \big| Y \le \frac{X}{2})$.

Q2 solution

Q3 2023

(15 points) Let X and Y be two integer random variables with joint PMF

$$P_{X,Y}(x,y) = \begin{cases} \frac{1}{6 \cdot 2^{\min(x,y)}} & \text{if } x, y \geq 0, |x-y| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) (5 points) Find the marginal distributions of X and Y. Are X and Y independent?
(b) (5 points) Find $P(X = Y)$.
(c) (5 points) Find $E[X|Y=2]$ and $\text{Var}[X|Y=2]$.

Q3 solution

Q4 2023

(20 points) Let Z_1, Z_2 be two $i.i.d.$ random variables satisfying standard normal distributions, $i.e.$, Z_1, $Z_2 \sim \mathcal{N}(0,1)$. Define

$$X = \sigma_X Z_1 + \mu_X;$$
$$Y = \sigma_Y \left(\rho Z_1 + \sqrt{1-\rho^2} Z_2\right) + \mu_Y,$$

where $\sigma_X > 0, \sigma_Y > 0, -1 < \rho < 1$.

(a) (5 points) Show that X and Y are bivariate normal.
(b) (5 points) Find the correlation coefficient between X and Y, $i.e.$,Corr(X, Y).
(c) (5 points) Find the joint PDF of X and Y.
(d) (5 points) Find $E[Y|X]$ and Var$[Y|X]$.

Q4 solution

Q5 2023

(10 points) Let the random variable $X \sim \mathcal{N}(\mu, \tau^2)$. Given $X = x$, random variables $Y_1, Y_2, \ldots, Y_n$ are *i.i.d.* and have the same conditional distribution, *i.e.*, $Y_i | X = x \sim \mathcal{N}(x, \sigma^2)$. Define the sample mean $\bar{Y}$ as follows:

$$\bar{Y} = \frac{Y_1 + \cdots + Y_n}{n}.$$

(a) (4 points) Find the posterior PDF of X given $\bar{Y}$.
(b) (3 points) Find the MAP (Maximum a Posterior Probability) estimates of X given $\bar{Y}$.
(c) (3 points) Find the MMSE estimates of X given $\bar{Y}$. (We know that the MMSE of X given Y is given by $g(Y) = E[X|Y]$).

Q5 solution

Q6 2023

(10 points) Let $X_1 \sim$Expo$(\lambda_1), X_2 \sim$Expo(λ_2) and $X_3 \sim$Expo(λ_3) be independent.

(a) (5 points) Find $E(X_1 + X_2 + X_3 | X_1 > 1, X_2 > 2, X_3 > 3)$ in terms of $\lambda_1, \lambda_2, \lambda_3$.

(b) (5 points) Find $P(X_1 = \min(X_1, X_2, X_3))$.

Q6 solution

Q7 2023

(10 points) Let $X \sim$Expo(λ), $Y \sim$Expo(λ); X and Y are independent.

(a) (5 points) Find $E(X|X+Y)$.

(b) (5 points) Find $E(X^2|X+Y)$.

Q7 solution

Q8 2023

(10 points) Instead of predicting a single value for the parameter, we give an interval that is likely to contain the parameter: A $1-\delta$ confidence interval for a parameter p is an interval $[\hat{p}-\epsilon, \hat{p}+\epsilon]$ such that $P\left(p \in [\hat{p}-\epsilon, \hat{p}+\epsilon]\right) \geq 1-\delta$. Now we toss a coin with probability p landing heads and probability $1-p$ landing tails. The parameter p is unknown and we need to estimate its value from experiment results. We toss such coin N times. Let $X_i = 1$ if the ith result is head, otherwise 0. We estimate p by using $\hat{p} = \frac{X_1 + \cdots + X_N}{N}$. Find the $1-\delta$ confidence interval for p, then discuss the impacts of δ and N.

Q8 2023 Continued

Hint: You can use the following Hoeffding bound: Let the random variables $X_1, X_2, \ldots, X_n$ be independent with $E(X_i) = \mu, a \leq X_i \leq b$ for each $i = 1, \ldots, n$, where a, b are constants. Then for any $\epsilon \geq 0$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right).$$

Q8 solution

Q9 2023

(10 points) Show the following inequalities.
(a) (5 points) Let $X \sim$Pois(λ). If there exists a constant $a > \lambda$, then

$$P(X \geq a) \leq \frac{e^{-\lambda}(e\lambda)^a}{a^a}.$$

(b) (5 points) Let X be a random variable with finite variance σ^2. Then for any constant $a > 0$,

$$P(|X - \mathbb{E}[X]| \geq a) \leq \frac{2\sigma^2}{\sigma^2 + a^2}.$$

Q9 solution

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.