TA Lecture 05 - Midterm Review

April 10-11

School of Information Science and Technology,
ShanghaiTech University

上海科技大学
ShanghaiTech University

Outline

Midterm Review

Selected Previous HW Problems

HW 5

Summary of Counting

Choose k objects out of n objects, the number of possible ways:

	Order Matters	Order Not Matter
with replacement	n^k	$\binom{n+k-1}{k}$
without replacement	$n(n-1)\cdots(n-k+1)$	$\binom{n}{k}$

Property of Probability

Probability has the following properties, for any events A and B:

1. $P(A^c) = 1 - P(A)$.
2. If $A \subseteq B$, then $P(A) \leq P(B)$.
3. $P(A \bigcup B) = P(A) + P(B) - P(A \cap B)$.

For any events $A_1, \ldots, A_n$:

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k)$$
$$+ \ldots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n).$$

Conditional Probability

Definition

If A and B are events with $P(B) > 0$, then the *conditional probability* of A given B, denoted by $P(A|B)$, is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

- $P(A)$: *prior probability* of A.
- $P(A|B)$: *posterior probability* of A.

Theorem

For any events $A_1, \ldots, A_n$ with positive probabilities,

$$P(A_1, \ldots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \cdots P(A_n|A_1, \ldots, A_{n-1}).$$

Conditional Probability is Also Probability

- Condition on an event E, we update our beliefs to be consistent with this knowledge.
- $P(\cdot|E)$ is also a probability function with sample space S:
  - $0 \leq P(\cdot|E) \leq 1$ with $P(S|E) = 1$ and $P(\emptyset|E) = 0$.
  - If events $A_1, A_2, \ldots$ are disjoint, then $P(\cup_{j=1}^{\infty} A_j|E) = \sum_{j=1}^{\infty} P(A_j|E)$.
  - $P(A^c|E) = 1 - P(A|E)$.
  - Inclusion-exclusion: $P(A \cup B|E) = P(A|E) + P(B|E) - P(A \cap B|E)$.

Key Tool: Bayes Rule and LOTP

Theorem

For any events A and B with positive probabilities,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Theorem

Let $A_1, \ldots, A_n$ be a partition of the sample space S (i.e., the A_i are disjoint events and their union is S), with $P(A_i) > 0$ for all i. Then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

Key Tool: Independence

Definition

Events A and B are *independent* if

$$P(A \cap B) = P(A)P(B).$$

If $P(A) > 0$ and $P(B) > 0$, then this is equivalent to

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

“Independence”

Definition

Events A, B and C are *independent* if all of the following equations hold:

$$P(A \cap B) = P(A)P(B)$$
$$P(A \cap C) = P(A)P(C)$$
$$P(B \cap C) = P(B)P(C)$$
$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Definition

Events A and B are said to be *conditionally independent* given E if:

$$P(A \cap B|E) = P(A|E)P(B|E).$$

PMF: Watch the Support

Theorem

Let X be a discrete r.v. with support x_1, x_2,... (assume these values are distinct and, for notational simplicity, that the support is countably infinite; the analogous results hold if the support is finite). The PMF p_X of X must satisfy the following two criteria:

- *Nonnegative:* $p_X(x) > 0$ *if* $x = x_j$ *for some* j, *and* $p_X(x) = 0$ *otherwise;*
- *Sums to* 1*:* $\sum_{j=1}^{\infty} p_X(x_j) = 1$.

PMF: Watch the Support

Theorem

Let X be a discrete r.v. and $g : \mathbb{R} \to \mathbb{R}$. Then the support of $g(X)$ is the set of all y such that $g(x) = y$ for at least one x in the support of X, and the PMF of $g(X)$ is

$$P(g(X) = y) = \sum_{x:g(x)=y} P(X = x)$$

for all y in the support of $g(X)$.

Random Variables

Name	Param.	PMF or PDF	Mean	Variance
Bernoulli	p	$P(X=1)=p, P(X=0)=q$	p	pq
Binomial	n, p	$\binom{n}{k}p^k q^{n-k}$, for $k \in \{0,1,\dots,n\}$	np	npq
FS	p	pq^{k-1}, for $k \in \{1,2,\dots\}$	$1/p$	q/p^2
Geom	p	pq^k, for $k \in \{0,1,2,\dots\}$	q/p	q/p^2
NBinom	r, p	$\binom{r+n-1}{r-1}p^r q^n, n \in \{0,1,2,\dots\}$	rq/p	rq/p^2
HGeom	w, b, n	$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$, for $k \in \{0,1,\dots,n\}$	$\mu = \frac{nw}{w+b}$	$(\frac{w+b-n}{w+b-1})n\frac{\mu}{n}(1-\frac{\mu}{n})$
Poisson	λ	$\frac{e^{-\lambda}\lambda^k}{k!}$, for $k \in \{0,1,2,\dots\}$	λ	λ
Uniform	$a < b$	$\frac{1}{b-a}$, for $x \in (a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	μ, σ^2	$\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2
Log-Normal	μ, σ^2	$\frac{1}{x\sigma\sqrt{2\pi}}e^{-(\log x-\mu)^2/(2\sigma^2)}, x > 0$	$\theta = e^{\mu+\sigma^2/2}$	$\theta^2(e^{\sigma^2}-1)$
Expo	λ	$\lambda e^{-\lambda x}$, for $x > 0$	$1/\lambda$	$1/\lambda^2$
Gamma	a, λ	$\Gamma(a)^{-1}(\lambda x)^a e^{-\lambda x}x^{-1}$, for $x > 0$	a/λ	a/λ^2
Beta	a, b	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$, for $0 < x < 1$	$\mu = \frac{a}{a+b}$	$\frac{\mu(1-\mu)}{a+b+1}$

Memoryless Property

Theorem

Suppose for any positive integer n, discrete random variable X satisfies

$$P(X \geq n + k | X \geq k) = P(X \geq n)$$

for $k = 0, 1, 2, \ldots$, then $X \sim \mathrm{Geom}(p)$.

Expectation

Definition

The *expected value* (also called the *expectation* or *mean*) of a discrete r.v. X whose distinct possible values are $x_1, x_2, \cdots$ is defined by

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j)$$

If the support is finite, then this is replaced by a finite sum. We can also write

$$E(X) = \sum_{x} \underbrace{x}_{\text{value}} \underbrace{P(X = x)}_{\text{PMF at } x}$$

where the sum is over the support of X.

Expectation

The expected value of a sum of r.v.s is the sum of the individual expected values.

Theorem

For any r.v.s X, Y and any constant c,

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

$$E(cX) = cE(X)$$

Theorem

If X is a discrete r.v. and g is a function from $\mathbb{R}$ to $\mathbb{R}$, then

$$E\left(g\left(X\right)\right) = \sum_{x} g\left(x\right) P\left(X = x\right)$$

where the sum is taken over all possible values of X.

Indicator: Fundamental Bridge

Theorem

There is a one-to-one correspondence between events and indicator r.v.s, and the probability of an event A is the expected value of its indicator r.v. I_A:

$$P(A) = E(I_A).$$

Key Tool: Indicator

Let A and B be events. Then the following properties hold.

1. $(I_A)^k = I_A$ for any positive integer k.
2. $I_{A^c} = 1 - I_A$.
3. $I_{A\cap B} = I_A I_B$.
4. $I_{A\cup B} = I_A + I_B - I_A I_B$.

- Given n events $A_1, \ldots, A_n$ and indicators $I_j, j = 1, \ldots, n$.
- $X = \sum_{j=1}^{n} I_j$: the number of events that occur
- $\binom{X}{2} = \sum_{i<j} I_i I_j$: the number of pairs of distinct events that occur
- $E(\binom{X}{2}) = \sum_{i<j} P(A_i \cap A_j)$
  - $E(X^2) = 2\sum_{i<j} P(A_i \cap A_j) + E(X)$.
  - $Var(X) = 2\sum_{i<j} P(A_i \cap A_j) + E(X) - (E(X))^2$.

Variance

- For any r.v. X, $\mathrm{Var}(X) = E(X^2) - (EX)^2$.
- $Var(X + c) = Var(X)$ for any constant c.
- $Var(cX) = c^2 Var(X)$ for any constant c.
- If X and Y are independent, then $Var(X + Y) = Var(X) + Var(Y)$.
- $Var(X) \geq 0$ with equality if and only if $P(X = a) = 1$ for some constant a.

Key Tool: PGF

Definition

The *probability generating function* (PGF) of a nonnegative integer-valued r.v. X with PMF $p_k = P(X = k)$ is the generating function of the PMF. By LOTUS, this is

$$E\left(t^X\right) = \sum_{k=0}^{\infty} p_k t^k.$$

The PGF converges to a value in $[-1, 1]$ for all t in $[-1, 1]$ since $\sum_{k=0}^{\infty} p_k = 1$ and $\left|p_k t^k\right| \leq p_k$ for $|t| \leq 1$.

Let X be a nonnegative integer-valued r.v. with PMF $p_k = P(X = k)$, and the PGF of X is $g(t) = \sum_{k=0}^{\infty} p_k t^k$, we have

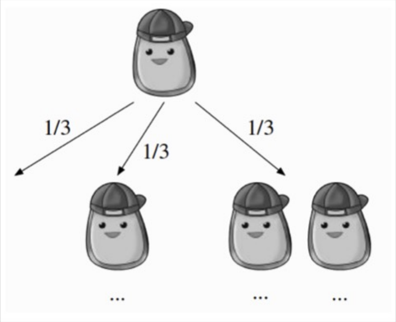
- $E(X) = g'(t)|_{t=1}$
- $E(X(X-1)) = g''(t)|_{t=1}$

Probabilistic Model: Birthday Problem

There are k people in a room. Assume each person's birthday is equally likely to be any of the 365 days of the year (we exclude February 29), and that people's birthdays are independent (we assume there are no twins in the room). What is the probability that two or more people in the group have the same birthday?

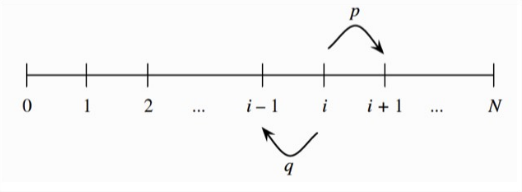
Probabilistic Model: Branching Process

A single amoeba, Bobo, lives in a pond. After one minute Bobo will either die, split into two amoebas, or stay the same, with equal probability, and in subsequent minutes all living amoebas will behave the same way, independently. What is the probability that the amoeba population will eventually die out?

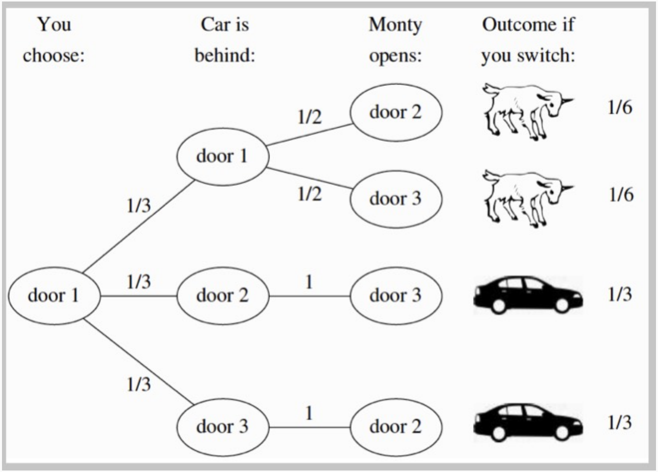


Probabilistic Model: Gambler's Ruin

Two gamblers, A and B, make a sequence of dollar 1 bets. In each bet, gambler A has probability p of winning, and gambler B has probability $q = 1 - p$ of winning. Gambler A starts with i dollars and gambler B starts with $N - i$ dollars; the total wealth between the two remains constant since every time A loses a dollar, the dollar goes to B, and vice versa. The game ends when either A or B is ruined (run out of money). What is the probability that A wins the game (walking away with all the money)?



Probabilistic Model: Monty Hall



Probabilistic Model: Coupon Collector

Suppose there are n types of toys, which you are collecting one by one, with the goal of getting a complete set. When collecting toys, the toy types are random (as is sometimes the case, for example, with toys included in cereal boxes or included with kids' meals from a fast food restaurant). Assume that each time you collect a toy, it is equally likely to be any of the n types. Let N denote the number of toys needed until you have a complete set. Find $E(N)$ and $Var(N)$.

Probabilistic Model: Pattern Matching

Suppose a coin with probability p for heads is tossed repeatedly, and we obtain a sequence of H and T (H denotes Head and T denotes Tail). Let N denote the number of toss to observe the first occurrence of the pattern "HH". Find $E(N)$ and $Var(N)$.

Outline

Midterm Review

Selected Previous HW Problems

HW 5

HW1 Problem 6

If each box of a brand of crispy instant noodle contains a coupon, and there are 108 different types of coupons. Given $n \geq 200$, what is the probability that buying n boxes can collect all 108 types of coupons?

HW2 Problem 4

In Monty Hall problem, now suppose the car is not placed randomly with equal probability behind the three doors. Instead, the car is behind door one with probability p_1, behind door two with probability p_2, and behind door three with probability p_3. Here $p_1 + p_2 + p_3 = 1$ and $p_1 \geq p_2 \geq p_3 > 0$. You are to choose one of the three doors, after which Monty will open a door he knows to conceal a goat. Monty always chooses randomly with equal probability among his options in those cases where your initial choice is correct. What strategy should you follow?

HW3 Problem 5

For X and Y binary digits (0 or 1), let $X \oplus Y$ be 0 if $X = Y$ and 1 if $x \neq y$ (this operation is called exclusive).

(a) Let $X \sim \text{Bern}(p)$ and $Y \sim \text{Bern}(1/2)$, independently. What is the distribution of $X \oplus Y$.

(b) Is $X \oplus Y$ independent of X? Is $X \oplus Y$ independent of Y? Be sure to consider both the case $p = 1/2$ and the case $p \neq 1/2$.

(c) Let $X_1, \ldots, X_n$ be i.i.d. Bern(1/2) R.V.s. For each nonempty subset J of $\{1, 2, \ldots, n\}$, let

$$Y_J = \oplus_{Y \in J} X_J.$$

Show that $Y_J \sim \text{Bern}(1/2)$ and that these $2^n - 1$ R.V.s are pairwise independent, but not independent.

HW4 Problem 5

People are arriving at a party one at a time. While waiting for more people to arrive they entertain themselves by comparing their birthdays. Let X be the number of people needed to obtain a birthday match.

(a) Show that $X = I_1 + I_2 + \cdots + I_{366}$, where I_j is the indicator r.v. for the event $X \geq j$. Then find $E(X)$ in terms of p_j's defined by $p_1 = p_2 = 1$ and for $3 \leq j \leq 366$,

$$p_j = (1 - \frac{1}{365})(1 - \frac{2}{365}) \cdots (1 - \frac{j-2}{365})$$

(b) Find the variance of X, both in terms of the p_j's and numerically. Hint:What is I_i^2, and what is $I_i I_j$ for $i < j$? Use this to simplify the expansion

$$X^2 = I_1^2 + \cdots + I_{366}^2 + 2 \sum_{j=2}^{366} \sum_{i=1}^{j-1} I_i I_j .$$

Outline

Midterm Review

Selected Previous HW Problems

HW 5

Problem 1

Nick and Penny are independently performing independent Bernoulli trials. For concreteness, assume that Nick is flipping a nickel with probability p_1 of Heads and Penny is flipping a penny with probability p_2 of Heads. Let $X_1, X_2, \ldots$ be Nick's results and $Y_1, Y_2, \ldots$be Penny's results, with $X_i \sim Bern(p_1)$ and $Y_j \sim Bern(p_2)$

Problem 1 Continued

(a) Find the distribution and expected value of the first time at which they are simultaneously successful, i.e., the smallest n such that $X_n = Y_n = 1$.
(b) Find the expected time until at least one has a success (including the success).
(c) For p1 = p2, find the probability that their first successes are simultaneous, and use this to find the probability that Nick's first success precedes Penny's.

Problem 1 Solution

Problem 2

A building has n floors, labeled $1, 2, \ldots, n$. At the first floor, k people enter the elevator, which is going up and is empty before they enter. Independently, each decides which of floors $2, 3, \ldots, n$ to go to and presses that button (unless someone has already pressed it)

(a) Assume for this part only that the probabilities for floors $2, 3, \ldots, n$ are equal. Find the expected number of stops the elevator makes on floors $2, 3, \ldots, n$.

(b) Generalize (a) to the case that floors $2, 3, \ldots, n$ have probabilities $p_2, \ldots, p_n$ (respectively); you can leave your answer as a finite sum

Problem 2 Solution

Problem 3

Given a random variable $X \sim Pois(\lambda)$ where $\lambda > 0$, show that for any non-negative integer k, we have the following identity:

$$E[\tbinom{X}{k}] = \frac{\lambda^k}{k!}$$

Problem 3 Solution

Problem 4

(a) Use LOTUS to show that for $X \sim Pois(\lambda)$ and any function g, $E(Xg(X)) = \lambda E(g(X+1))$. This is called the Stein-Chen identity for the Poisson.

(b) Find the forth moment $E(X^4)$ for $X \sim Pois(\lambda)$ by using the identity from (a) and a bit of algebra to reduce the calculation with the fact that X has mean λ and variance λ.

Problem 4 Solution

Problem 5

Suppose a fair coin is tossed repeatedly, and we obtain a sequence of H and T (H denotes Head and T denotes Tail). Let N denote the number of tosses to observe the first occurrence of the pattern "HTHT". Find $E(N)$ and $Var(N)$.

Problem 5 Solution

Problem 6 (Optional Challenging Problem)

An Erdos-Renyi random graph is formed on n vertices. Each unordered pair (edge) (i, j) of vertices is connected with probability p, independently of all the other pairs.

(a) A wedge (or path of length 2) is a tuple (i, j, k) where i, j, k are distinct and each of the edges (i, j) and (i, k) is connected. Let W denote the number of wedges contained in the random graph. Find appropriate condition under which W is approximately Poisson distributed.

(b) A triangle is a set of three vertices i, j, k such that each of the three edges (i, j), (j, k) and (i, k) is connected. Let T denote the number of triangles contained in the random graph. Find appropriate condition under which T is approximately Poisson distributed.

Problem 6 Solution