Prob. RC Note

Chia-Cheng, Hao
Yan-Wei, Su

October 30, 2022

Abstract

This is a note for recitation in 2022 Fall undergraduate Probability Theory in the Applied Mathematics department at NYCU. The instructor of the course is Professor Yuan-Chung Sheu (website), and the two TA are Chia-Cheng, Hao and Yan-Wei, Su. The Github site for this note is at https://github.com/18Allen/PT_RC_material/blob/main/Notes/master.pdf

Contents

Chapter 0

RC 0

Recitation 0

0.1 Basic rules

22 Sep. 18:30

- Recitation: Thursday 12:30 to 13:30, 18:30 to 19:30
- My office hour is right after recitation.
- Grade distribution: (quiz correction + attend rc): 1×4; mid correction + attend rc: 2; Attend one rc: 4

0.2 Review of some high school comminatory and probability tricks

Early chapters are about high school counting things over again. We will swiftly go through the concept.

0.2.1 Permutation and Combination

- Permutation
  #ways to form an ordering of m out of n different things.

$$P(n,m)$$

- Combination
  #ways to form a group of m with n different things.

$$C(n,m), \quad \binom{n}{m}$$

- Multinomial coefficents (Gra21)
  #ways to divide a set of n elements into r (distinguishable) subsets of $n_1, n_2, \dots, n_r$ elements.

$$\frac{n!}{n_1!n_2!\dots n_r!}$$

0.2.2 Set Operation

- De Morgan's Law

0.3 Axioms of Probability

0.3.1 Probability Space

The **probability space** is a triple $\Omega, \mathcal{F}, P$ that contains

- The **sample space** Ω contains all possible outcome.
- The σ-algebra $\mathcal{F}$ is the **event space**. It is a subset of the power set of Ω we are interested in.
- The **probability measure** P is a function $P : \mathcal{F} \to [0, 1]$ that satisfies the three axioms.
  1. $P(\Omega) = 1$
  2. Non-negative
  3. Countable additivity for disjoint sets in $\mathcal{F}$.

0.3.2 Standard process

A standard process of solving these problems (HW1,2) is to find the size of possible outcome first. Then, finding the size of desired event, and the ratio of the two is the prob.

Example. (Gra21) Example 3.11
You have 10 pairs of socks in the closet. Pick 8 socks at random. For every i, compute the probability that you get i complete pairs of socks.

- # outcome:
- # desirable outcome:
- the probability is :

Example. (Gra21) Problem 3.2 (HW2 problem 2)
Three married couples take seats around a table at random. Compute $P(\text{no wife sits next to her husband})$.
Use Inclusion-Exclusion principle to compute the probability of its complement event.

0.3.3 Why do we need to set σ-algebra: Vitali set

Have you ever wonder: Why would I need $\mathcal{F}$ if I have Ω already? As the textbook said, you won't have any problem with this notion. However, things get messy when we encounter set of infinity size. The idea of "length" will not be clear then. We use **Vitali set** V as an example on $\mathbb{R}$ to show that we can't have a measure on V. This problem is one of the reason that we only put probability measure on $\mathcal{F}$.

For more information: How the Axiom of Choice Gives Sizeless Sets | Infinite Series

0.4 Homework Help

TBD

Chapter 1

RC 1

Recitation 1

1.1 Review

29 Sep. 18:30

1.1.1 Axioms of Probability

A **probability space** is a triple $\Omega, \mathcal{F}, P$ that contains

- The **sample space** Ω contains all possible outcome.
- The σ-algebra $\mathcal{F}$ is the **event space**. It is a subset of the power set of Ω we are interested in.
- The **probability measure** P

1.1.2 Probability (measure)

- The **probability measure** P is a function $P : \mathcal{F} \to [0, 1]$ that satisfies the three axioms.
  1. $P(\Omega) = 1$
  2. Non-negative
  3. Countable additivity for disjoint sets in $\mathcal{F}$.

1.1.3 σ-algebra

$\mathcal{F}$ is call a σ-algebra on a set Ω If

1. $\emptyset \in \mathcal{F}$
2. If $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. If $A_1, A_2, \cdots \in \mathcal{F}$, then $\cup_{i=1} A_i \in \mathcal{F}$

> **Example.** Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, find a minimal* σ-algebra that contains the sets $\{1, 2, 3\}, \{1\}$

Answer: [1]

1.1.4 Conditional Probability

For the general definition, take events $A, B,$ and assume that $P(B) > 0$. The *conditional probability* of the event A given B equals

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

[1] $\mathcal{F} = \{\emptyset, \Omega, \{1\}, \{1, 2, 3\}, \{4, 5, 6\}, \{2, 3, 4, 5, 6\}, \{2, 3\}, \{1, 4, 5, 6\}\}$

1.1.5 Independence

Events $A_1, \ldots, A_n$ are independent if

$$P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$$

Please note that it means you can not just check $P(A_i \cap A_j) = P(A_i)P(A_j), i \neq j$. One example is teh following

Example (Pairwise Independence but not independent variables). ((Pan19) Exercise 1.4.2) Consider a regular tetrahedron die painted blue, red and green on three sides and painted in all three colours on the fourth side. If the die is equally likely to land on any side, show that the appearances of these colours on the side it lands on are pairwise-independent but not independent.

1.2 Problems

Exercise. Put r distinguishable balls into n different boxes. What is the probability of all n boxes are occupied?

Probabilistic approach. $A_i = \{\text{Box } i \text{ is occupied}\}$ for all $i = 1, \cdots, n$.

$$\mathbb{P}(\text{No empty boxes}) = \mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big). \tag{1.1}$$

(Inclusion-exclusion principle!)

Combinatorial approach. Define $A(r, n)$ as the number of distributions such that all n boxes are non-empty when you put r balls into them. Then the probability is $A(r, n)/n^r$. Knowing $A(r, n + 1)$ can lead us to $A(r, n)$:

$$A(r, n) = \sum_{k=1}^{r-1} \binom{r}{k} A(r - k, n - 1). \tag{1.2}$$

Knowing this, we can prove that:

$$A(r, n) = \sum_{\nu=0}^{n} (-1)^{\nu} \binom{n}{\nu} (n - \nu)^r. \tag{1.3}$$

(Use Induction!)

(Think about it!) What is the probability of exactly m boxes are occupied?

$$\text{number of distributions } = \binom{n}{m} \times A(r, m). \tag{1.4}$$

$A(r, m) = \mathbb{P}(r \text{ balls into } m \text{ boxes and all of the boxes are occupied}) \times m^r$.

Chapter 2

RC 2

Recitation 2

2.1 Review

13 Oct. 18:30

2.1.1 Random Variables

(([Cha]) chapter 2) It is a 'function' mapping Ω to some number, i.e

$$X : \Omega \to \mathbb{R}$$

Notation

- $X \in A, \quad A \subseteq \mathbb{R}$: The set containing $\omega \in \Omega$ such that $X(\omega) \in B$. $X \in B = \{\omega \in \Omega | X(\omega) \in B\}$

2.1.2 Discrete Random Variables

The **range** of X is finite or countable.

- **probability mass function**: A measure on x_i given by the measure in Ω

$$f(x_i) = P(X = x_i)$$

- Because $X \in x_i$ for $i = 1, 2, \ldots$ from a disjoint partition (Why?) of Ω, we have

$$\sum_{i=1}^{\infty} f(x_i) = P(\Omega) = 1$$

  TBD(Panchenko assume that Ω is countable)

- Independence of random variables: (We won't go through the detail in class, **strongly recommend** you to check out ([Cha]) Proposition 3.) It comes from the Independence of events(1.1.5). This is a theorem showing you the equivalent way of defining independence. One using subcollection, one using collection of all.

  Additional note: An infinite sequence of r.v $X_1, X_2, \ldots$ is called Independence if **for any n**, $X_1, X_2, \ldots, X_n$ are independent.

2.1.3 Some Discrete r.v

The most important thing to remember about a r.v (before a test) are 1. PMF; 2. expectation; 3. Varaiance. Examples about the r.v or the relation with other r.v's are helpful, too. You can those info in ([Gra21]) Chapter 5.

- **Bernoulli r.v** Bernoulli(p)

- **Binomial r.v** binomial(n, p)
- **Poisson r.v** Poisson(λ)
- **Geometric r.v** Geometric(p)

2.1.4 Joint Probability Mass Function

Let $X_1, \ldots, X_n$ be discrete random variables defined on the same sample space. The function

$$f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

is called the joint probability mass function (joint p.m.f) of the r.v $X_1, \ldots, X_n$.

marginal p.m.f: sum out the rest variables then the one(s) you are interested in, i.e

$$f_i(x) = \sum_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n} f(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

Theorem. Let $X_1, \ldots, X_n$ be discrete random variables with joint p.m.f f. Suppose that

$$f(x_1, \ldots, x_n) = h_1(x_1) \cdots h_n(x_n)$$

for some p.m.f $h_1, \ldots, h_n$. Then $X_1, \ldots, X_n$ are independent. and $f_i = h_i$, for $i = 1, \ldots, n$.

Example (Joint p.m.f)**.** .

- (In the lecture note) Compute the p.m.f of Geometric r.v from Bernoulli's
- Compute the p.m.f of Binomial r.v from Bernoulli
- (Joint p.m.f when not independent) Consider two independent r.v X, Y follows the p.m.f
$$f(-1) = f(0) = f(1) = \frac{1}{3}$$
Then, we consider $U = XY, V = Y$. Compute the respective p.m.f of U and V. Also, compute the joint p.m.f $f_{U,V}(u, v)$. What do you observe?

Exercise (stability of Poisson random variables)**.** Suppose $X_n \sim$ Pois(λ) are i.i.d. Then Define $S_n = X_1 + \cdots + X_n$. Show that $S_n \sim$ Pois$(n\lambda)$.

2.2 Preview: Expectation

Let X be be a discrete random variable. The expected value or expectation or mean of X is defined as

$$E(X) = \sum_x x P(X = x)$$

Some simple example is like 'The expected number of head of fair coin toss', or 'average outcome of rolling a six-sided dice' etc.

How do we extend the idea to multiple random variables?

Theorem. (Proposition 6. in (Cha)). Let $X_1, \ldots, X_n$ be discrete random variables and $Y =$

$f(X_1, \ldots, X_n)$ for some function f. Then,

$$E(Y) = \sum_{x_1,\ldots,x_n} f(x_1, \ldots, x_n) P(X_1 = x_1, \ldots, X_n = x_n).$$

With this notion we can use the linearity of expectation.

Exercise (Linearity of expectation). .

- Acquire the expectation of Binomial r.v by direct computation and linearity
- Define two independent r.v X_1 Poisson(λ_1) and X_2 Poisson(λ_2) Calculate the expectation of $Y = X_1 + X_2$ by the linearity of expectation and by the stability of Poisson random variables.

Note. The Linearity of expectation does not require the sequence of r.v being independent. (Next time we will show that this is not true for **variance**)

2.3 Extra note

2.3.1 From Bernoulli to Binomial

Exercise (Peak of Binomial distribution). (Feller p.59 q.5) The probability p_k that a given cell contains exactly k balls (with a total r balls and n boxes) is given by the binomial distribution. Show that the most probable number v satisfies

$$\frac{r+n-1}{n} \le v \le \frac{r+1}{n}$$

(In other words, it is asserted that $p_0 < p_1 < \cdots < p_{v-1} \le p_v > p_{v+1} \cdots > p_r$)

(Hint: Don't use derivative to find local extremum).

2.3.2 From Binomial to Poisson

Exercise (Limiting form). (Feller p.59 q.6) If $n \to \infty$ and $r \to \infty$ so that the average number $\lambda = \frac{r}{n}$ of balls per cell remains constant. Then,

$$p_k \to e^{-\lambda} \lambda^k / k!$$

Chapter 3

RC 3

Recitation 3

3.1 Review

20 Oct. 18:30

Independence of random variables:

Exercise. Let X and Y be independent discrete random variables and let $g, h : \mathbb{R} \to \mathbb{R}$. Show that $g(X)$ and and $h(Y)$ are independent.

3.1.1 Expectation

Theorem. (Proposition 6. in (Cha)). Let $X_1, \ldots, X_n$ be discrete random variables and $Y = f(X_1, \ldots, X_n)$ for some function f. Then,

$$E(Y) = \sum_{x_1,\ldots,x_n} f(x_1, \ldots, x_n) P(X_1 = x_1, \ldots, X_n = x_n).$$

With this notion we can use the linearity of expectation.

3.2 Applications

3.2.1 Geometric r.v

Example (Coupon collector problem). ((Gra21) Example 8.6) Sample from n cards, with replacement, indefinitely. Let N be the number of cards you need to sample for a complete collection, i.e., to get all different cards represented. What is $E[N]$?

3.2.2 Piosson Approximation to binomial

Example (HW5,Q4). ((Pan19) Exercise 1.3.6) When you bet on black in Roulette, your chances are 18/38. Suppose also that bets over \$250 are not
allowed by the casino. You decide to play the following strat- egy: you start with a \$1 bet and double the bet until either you
win (the same amount as the bet) or the bet exceeds \$250; then you start again with a \$1 bet and repeat. We will call this sequence of bets in the strategy until restart with a \$1 bet 'one round'. If you play 1000 rounds of this strategy, what is the probability that your total winnings/losses are 0? Compute the exact formula and then compare it with Poisson approximation.

Solution. Suppose in the i-th round, $1 \leq i \leq 1000$, the money you'll get if end up winning at the n-th bet is

$$-1 - 2 - \cdots - 2^{n-2} + 2^{n-1} = 2^{n-1} - (2^{n-1} - 1) = 1,$$

which is independent to n. Otherwise, if you end up lose all the bets, then the money you'll earn is

$$-1 - 2 \cdots - 2^7 = -(2^8 - 1).$$

Note that the maximum number of bets is 8 times since $2^8 > 250$. Clearly, you'll earn this amount with probability $q = (1-p)^8$. Therefore, $\mathbb{P}(\text{ends up getting } 1 \text{ dollar in a round}) = 1 - q$.

Now define

$$\begin{aligned}
N &= \text{total earning in } 1000 \text{ rounds} \\
W &= \text{total number of rounds that ends up getting } 1 \text{ dollar} \\
L &= \text{total number of rounds that ends up getting } -255 \text{ dollars}
\end{aligned}$$

Clearly $W + L = 1000$, and

$$N = 1 \cdot W + (-255) \cdot L.$$

Moreover, one can see

$$\mathbb{P}(L = k) = \binom{1000}{k} q^k (1-q)^{1000-k},$$

which is a $Bin(1000, q)$ random variable. Then

$$\mathbb{P}(N \geq 0) = \mathbb{P}(W - 255L \geq 0) = \mathbb{P}(1000 - 256L \geq 0) = \mathbb{P}(L < 4).$$

Using Poisson approximation to L, one can get

$$\left|\mathbb{P}(L < 4) - \mathbb{P}(X < 4)\right| \leq \frac{\lambda^2}{1000},$$

if $\lambda = 1000q$ and $X \sim Pois(\lambda)$.

Example. ((Pan19) Exercise 1.3.5.) The Kicker for the Dallas Cowboys scores an average of 2 Field goals per game. Over this players 150 game career, what is the probability that in at least one game he scored exactly 6 Field goals?

3.3 Variance

Variance of a random variable X is the second moment of the demeaned $X - E[X]$, i.e

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

Example. Calculate the variance of Poisson random variable

Note: Second moments for some random variable may not exists (L^2 will be save), and in those cases, the variance will not be defined. See Cauchy distribution.

3.4 Extra note

Exercise. Let X and Y be independent geometric random variables with respective parameters α and β. Show that

$$P(X + Y = z) = \frac{\alpha\beta}{\alpha - \beta}\{(1-\beta)^{z-1} - (1-\alpha)\}$$

Chapter 4

RC 4

Recitation 4

4.0.1 Variance

27 Oct. 18:30

Variance of a random variable X is the second moment of the demeaned $X - E[X]$, i.e

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

4.0.2 Covariance

The definition of covariance is like that for variance, but for two random variable.

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Take Y and X the same then you get the definition for variance. Below are two properties for the use of covariance

- (Proposition 8)Bilinear:
  Let $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$ be random variables and $a_1, \ldots, a_m, b_1, \ldots, b_n$ be real numbers. Let $\mathcal{U} = a_1 X_1 + \ldots, a_m X_m$ and $V = b_1 Y_1, \ldots, b_n Y_n$. Then

  $$\text{Cov}(\mathcal{U}, V) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j \text{Cov}(X_i, Y_j).$$

  The immediate use of this proposition is Corollary 4 and 6. This is useful when you consider the variance of something like $S_n = \sum_{i=1}^{n} X_i$, where X_i's are i.i.d random variable. Then, $\text{Var}(S_n) = n \cdot \text{Var}(X_1)$. You can verify this by thinking the variance of Bernoulli(p) and Bin(n,p), one is $p(1-p)$ and the other is $np(1-p)$.

- (Corrollary 5)Independent implies covariance zero
  If two random variables are independent, then there covariance is zero (**NOT the other way around!**) The exercise below show that the converse is false.

> **Exercise.** Find a counterexample of two discrete random variable X, Y such that $\text{Cov}(X, Y) = 0$, but they are not independent.

Answer: [1]

[1] Y with $f_y(1) = f_y(-1) = f(0) = \frac{1}{3}$ (uniform), and $X = |Y|$

4.1 The method of indication

((Cha) p.33)" The method of indicators is a technique for evaluating the expected value/variance of a random variable by finding a way to write it as a sum of indicator function."

The **indicator** of A is a random variable, denoted by 1_A, (you can understand it as taking A or not), defined as follows

$$1_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}$$

If X can be represented as $X = \sum_{i=1}^{n} 1_{A_i}$, we can use **linearity of expectation** to write

$$E[X] = \sum_{i=1}^{n} E[1_{A_i}]$$

(Remember here we don't have to worry about independence of 1_{A_i} .)

4.1.1 For expectation

Example (Coin Run). .

- A biased coin is tossed n times, and heads shows with probability p on each toss. A run is a sequence of throws which result in the same outcome, so that, for example, the sequence

  $$HHTHTTH$$

  contains five runs. Show that the expected number of runs is $1 + 2(n - l)p(1 - p)$. Find the variance of the number of runs.

- As a review, what is the expected length of a run.

- A head run is simply a continuous sequence of heads. Consider the first problem but with head run.

Exercise. Of the $2n$ people in a given collection of n couples, exactly m die. Assuming that the m have been picked at random, find the mean number of surviving couples. This problem was formulated by Daniel Bernoulli in 1768.

Hint: Find indicator on the survival for each couple.

Wiki Best of luck with your exams.

Bibliography

(Cha) Sourav Chatterjee. *Lecture Notes For MATH151*. URL: https://souravchatterjee.su.domains//math151notes.pdf.

(Gra21) Janko Gravner. *Lecture Notes for Introductory Probability Introduction to Probability*. sbd, 2021. URL: https://www.math.ucdavis.edu/~gravner/MAT135A/resources/lecturenotes.pdf.

(Pan19) Dmitry Panchenko. *Introduction to Probability Theory*. 2019. URL: https://drive.google.com/file/d/1Rpkr-NCEyqygvypR65RznaZkb36KHB29/view.