Prob. RC Note

Chia-Cheng, Hao
Yan-Wei, Su

December 1, 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 $n_1, n_2, \ldots, n_r$ elements.

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

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(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 $i = 1, 2, \dots$ 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 $X_1, X_2, \dots$ is called Independence if **for any n**, $X_1, X_2, \dots, X_n$ 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

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

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

(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, $1 \le i \le 1000$, 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

N = total earning in 1000 rounds
W = total number of rounds that ends up getting 1 dollar
L = total number of rounds that ends up getting -255 dollars

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

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

Using Poisson approximation to L, one can get

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

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 $\mathcal{U} = a_1X_1 + \ldots, a_mX_m$ and $V = b_1Y_1, \ldots, b_nY_n$. Then

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

  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.

Chapter 5

RC 5

Recitation 5

5.1 Laws of large number

3 Nov. 18:30

Congratulation! You have made it to this chapter. The tools we are going to learn in this chapter is more structured compare to the previous ones. We start with Markov inequality and end with the weak law of large number. To give rough structure

$$\text{Markov inequality} \Rightarrow \text{Chebyshev inequality} \Rightarrow \text{WLLN}$$

5.1.1 Some inequalities

Although we don't use it directly, union bound is quite useful.

> **Note** (Union bound). For any events $A_1, \ldots, A_n$,
>
> $$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

Markov inequality

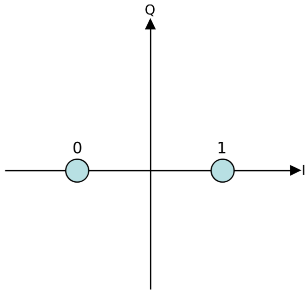
Markov inequality give the **tail bound** of the nonnegative random variable.

> **Note** (Markov inequality). Let X be a **nonnegative** r.v. Then, for any $t > 0$
>
> $$P(X \geq t) \leq \frac{E[X]}{t}$$

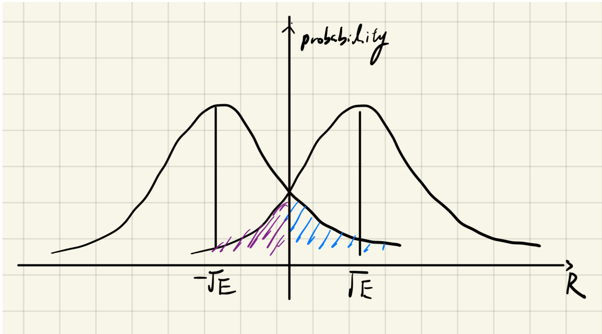
There are many approach to get Markov inequality, it is worthwhile to check them out. Before you do so, I would like to introduce a application of Markov inequality in communication technology.

BPSK

(Skipping this won't hurt your understanding of this chapter and will not be on any test.)
Abstraction
Binary in BPSK means that in each round of transmission, **transmitter** either send $-\sqrt{E_0}$ or $\sqrt{E_0}$, where E_0 is the transmitting energy. However, there will be some noise real world, and we usually take the noise as a Gaussian random variable w follows N(0, $\frac{N_0}{2}$) (It is okay we will use a graph to show this). If we assume that the receiving $x = \hat{b} + w$ greater than zero is 1 and 0 otherwise, the question will **how do we calculate the blue/purple area?** The blue area



Picture in the wiki page. BPSK symbol.



receiving signal under the effect of noise.

represent an error where 0 is recognized as 1. Qualitatively, we may skip a load of definition and conclude that the area of blue/purple is calculated as

$$\frac{1}{2}\int_{\sqrt{\frac{E_0}{N_0}}}^{\infty}\frac{2}{\sqrt{\pi}}e^{-u^2}du$$

As we know that this integral is hard to compute, so we want an estimation. This will become a easier problem if we consider X with p.d.f $f(x) = \frac{2}{\sqrt{pi}}e^{-x^2}, x \geq 0$, and thus $E[X] = \frac{e^{-x^2}}{\sqrt{\pi}}$ Hence, we may apply Markov inequality to get

$$P(X \geq c) \leq \frac{E[X]}{c} = \frac{e^{-c^2}}{c\sqrt{\pi}},$$

where $c = \sqrt{\frac{E_0}{N_0}}$.

Chebyshev inequality

The last inequality we will introduce here is Chebyshev inequality.

> **Note** (Chebyshev inequality)**.** Let X be **any** r.v. Then for **any** $t > 0$,
>
> $$P(|X - E[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$$

It is a direct application of Markov inequality as $P(|X - E[X]| \geq t) = P(|X - E[X]|^2 \geq t^2)$, and $E[(X - E[X])^2] = Var(X)$

5.1.2 Converge in probability

If we say $X_n \to c$ **in probability**, we means the sequence of random varaibles $(X_i)_{i=1}^{\infty}$ will eventually be really close to c in a way that those ω that does not map close to c will eventually has measure zero.

$$\lim_{n\to\infty} P(|X_n - c| \geq \epsilon) = 0$$

for every $\epsilon > 0$.

5.1.3 The weak law of large number (WLLN)

To give an intuition of WLLN, you may understand it as an justification of "Taking average of many sample will be really close to expectation value".

Theorem 5.1.1 (Weak Law of Large Numbers)**.** (Thm. 4 in (Cha))
Let $X_1, X_2, \ldots$ be an **i.i.d.** sequence of random variables with expected value μ and finite variance. For each n, let

$$\bar{X_n} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

be the average of the first n of these variables. Then as $n \to \infty$, $\bar{X_n} \to \mu$ **in probability**.

Direct application of Chebyshev inequality. If we consider **i.i.d** $(X_i)_{i=1}^{\infty}$, then the mean $\bar{X}_n$ will be really close to the expectation of X_1, $E[X_1]$. In formal expression we write $\bar{X}_n \to E[X_1]$ in probability.

Tips when using WLLN:
The key to this type of problem is to show that the variance will have **limit 0** as n grows.

Chapter 6

RC 6

Recitation 6

17 Nov. 18:30

Remark. Errata: page 5 of the "Continuous random variables" handout:

- $\cancelto{\mu\in\mathbf{R}}{\mu\mathbf{R}}$
- $X \sim$ Normal$(\mu, \cancelto{\sigma^2}{\sigma})$
- $N(\mu, \cancelto{\mu^2}{\mu})$
- $\cancelto{\frac{1}{\sqrt{2\pi\sigma^2}}}{\frac{1}{\sqrt{2\pi\sigma}}}$

6.1 Review: Laws of large number

Markov inequality $\Rightarrow$ Chebyshev inequality $\Rightarrow$ WLLN

6.1.1 Converge in probability

If we say $X_n \to c$ **in probability**, we means the sequence of random varaibles $(X_i)_{i=1}^{\infty}$ will eventually be really close to c in a way that those ω that does not map close to c will eventually has measure zero.

$$\lim_{n\to\infty} P(|X_n - c| \geq \epsilon) = 0$$

for every $\epsilon > 0$.

6.1.2 The weak law of large number (WLLN)

To give an intuition of WLLN, you may understand it as an justification of "Taking average of many sample will be really close to expectation value".

Theorem 6.1.1 (Weak Law of Large Numbers)**.** (Thm. 4 in (Cha))
Let $X_1, X_2, \ldots$ be an **i.i.d.** sequence of random variables with expected value μ and finite variance. For each n, let

$$\bar{X_n} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

be the average of the first n of these variables. Then as $n \to \infty$, $\bar{X_n} \to \mu$ **in probability**.

Direct application of Chebyshev inequality.

In many cases, you can't expect condition like i.i.d., but as we can see in the following example, you actually don't need it. WLLN only requires **second moment information**, such as covariance and variance.

> **Example** (You Don't need i.i.d. for WLLN). ((Pan19) Exercise 2.2.1)
> Suppose that random variables $X_1, \dots, X_n$ are uncorrelated, $\mu = E[\bar{X_n}]$ and Var(X_i) $\leq \sigma^2$ for all $i \leq n$. Then show that WLLN still holds.

Notice that the key is to show that Var(X_n) grows with the rate $\theta(n^2)$ as $n \to \infty$.

Combine with the use of indicator function, we can study random graph.

> **Example** (Expected number of triangles on Erdős-Rényi random graph). (Adapted from Prof. Chatterjee's STAT310 note Exercise 8.4.5) Define an undirected random graph on n vertices by putting an edge between any two vertices with probability p and excluding the edge with probability $1-p$, all edges independent. This is known as the Erdős-Rényi $G(n,p)$ random graph.
>
> - If $T_{n,p}$ is the number of triangles in this random graph, use the method of indicator to represent $T_{n,p}$.
> - Before we made calculation, can you guess why $T_{n,p}/n^3 \to p^3/6$?
>   Hint: Make reasons out of the number of indicator function, probability of one triangles being connected.
> - Show that $T_{n,p}/n^3 \to p^3/6$ in probability with WLLN. Hint: Break down the $\sum$Cov to cases depend on the relation of two triangles (share one edges, no share edges, same triangles)

6.2 Continuous Random Variables

6.2.1 Some Continuous Random Variables

These are some continuous r.v. that you should keep in mind.

- **Uniform r.v.** $X \sim$ **Unif**$([a,b])$:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } x \in [a,b]; \\ 0, & \text{otherwise .} \end{cases}$$

- **Exponential r.v.** $X \sim$ **Exp**(λ):

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0; \\ 0, & \text{otherwise .} \end{cases}$$

- **Normal(Gaussian) r.v.** $X \sim N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

  The $N(0,1)$ distribution is called **standard normal distribution**.

Before we introduce a more formal definition concerning continuous random variables, let's try to extend our understanding from discrete random ones.

> **Example.** ((Pan19) Exercise 4.1.2.) Compute the expectation of a uniform and exponential random variable.

We basically just turn summation to integration.

Exercise. ([Pan19] Exercise 4.1.4.) Compute the expectation of $E|X|$, where X follows $N(0, \sigma^2)$.

6.2.2 Probability Density Function

We may ignore some intricate definition of "niceness" of p.d.f. for now. Just remember that if a **non-negative** function f is given as p.d.f of random variable X, the probability of X taking value in A is

$$P(X \in A) = P(A) = \int_A f(x)dx$$

This means that we care about the value **"after" integration**, not density per se. .

- **p.d.f. can take on value larger than 1**
  For example: $f(x) = 21_{|X| \le \frac{1}{2}}$. $f(0) = 2$ but $\int_{-1/2}^{1/2} f(x)dx = 1$.
- **P($\mathbb{R}$)**
  By the definition above, $P(\mathbb{R}) = \int_{-\infty}^{\infty} f(x)dx = 1$

6.2.3 Expectation and Variance

. The expectation of a continuous random variable X with p.d.f. f is defined as

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

Providing that $xf(x)$ is absolutely integrable. (The reason is that we will sometimes change the order of integration).

Given the condition that $x^2 f(x)$ is absolutely integrable, we may define the **variance** of X as

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

6.2.4 Cumulative Distribution Function

Recall the first problem in quiz two that we ask you to draw a function $F_X(x) = P(X \le x)$. The cumulative distribution function (c.d.f.) of a continuous random variables X is defined just like the one for discrete r.v. .

$$F_X(x) = F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

In normal cases where your p.d.f. is continuous, the fundamental theorem of calculus implies that

$$f(x) = F'(x)$$

This property can be helpful when you use c.d.f. to compute p.d.f. after transformation. Check out the example in [Cha] page 55 about $Y \sim X^2$.

6.2.5 Change of Variable Formula

Let X be a continuous r.v. and $u : \mathbb{R} \to \mathbb{R}$ be a **strictly increasing function**. We define a new r.v. $Y = u(X)$.
What is the p.d.f. of Y?

$$P(Y \le y) = P(u(X) \le y) = P(X \le u^{-1}(y)).$$

Using the c.d.f. of X at $u^{-1}(y)$, we can compute the p.d.f. of y by taking the derivative

$$\frac{d}{dy} F(u^{-1}(y)) = F'(u^{-1}(y))(u^{-1})'(y)$$

> **Remark.** The case for **strictly decreasing** u is to compute the derivative of
> $$P(u(X) \leq y) = P(X \geq u^{-1}(y)) = 1 - F(u^{-1}(y))$$
> Hence, there will be an extra "-" in $-f(u^{-1}(y))u^{-1}(y)$.

This idea of changing the measure will extend to the change-of-variable for multivariable case. That will be the topic of next week.

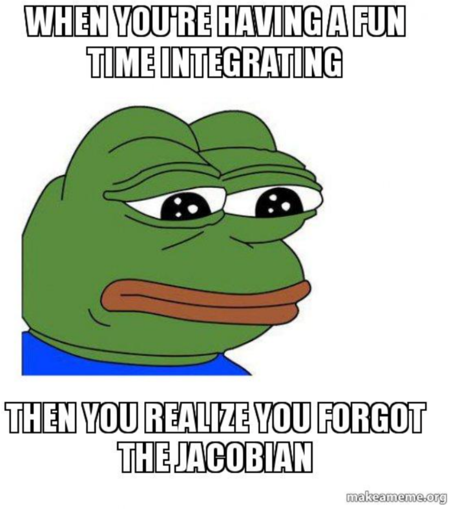


Figure 6.1: source

> **Example.** ((Pan19) Exercise 4.1.3.) What is the density of the random variable $X = g^3$, where g follows $N(0,1)$.

6.3 Extra

"Derive" Gaussian distribution

Have you ever wonder why Gaussian distribution look like that?
Normal distribution's probability density function derived in 5min

"Real" issues

- Not every set in $\mathbb{R}$ has probability defined.
- If we change value of a p.d.f on a measure-zero set, the result will still be a p.d.f.

Chapter 7

RC 7

Recitation 7

7.1 Continuous Random Variables

24 Nov. 18:30

Joint p.d.f

The concept for one variable r.v. still applies here. For continuous r.v., we won't use "probability at point". Instead, we will only use p.d.f in the context of "**measuring probability of event**", i.e. (Cha) Suppose that $X_1, \ldots, X_n$ are continuous r.v. defined on the same probability space. The n-tuple (means order things of length n) $X = (X_1, \ldots, X_n)$ is called a **random vector**. Note that

- $X : \Omega \to \mathbb{R}^n$
- non-negative $f : \mathbb{R}^n \to \mathbb{R}$ is the **joint p.d.f.** of $X_1, \ldots, X_n$ is such that for any "good" (i.e. measurable) set $A \subseteq \mathbb{R}^n$, we measure the probability of $X \in A$ by

$$P(X \in A) = \int_A f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

> **Exercise.** Given a joint p.d.f. $f(x, y)$ defined by
>
> $$f(x,y) = \begin{cases} 1, & \text{if } 0 \le x \le 1, 0 \le y \le 1; \\ 0, & \text{otherwise .} \end{cases}$$
>
> What is the probability of $A = [\frac{1}{4}, \frac{1}{2}] \times [\frac{1}{3}, \frac{2}{3}]$

- Independence
As the concept in previous chapters on discrete r.v., independent for r.v.'s means that we can calculate the probability of joint event "variable by variable", i.e.

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

The RHS equals (The p.d.f. of X_i is $f_i(x_i)$)

$$\int_{A_1} \cdots \int_{A_n} \prod_{i=1}^{n} f_i(x_i) dx_1 \ldots dx_n$$

, and the LHS is

$$\int_{A_1 \times \cdots \times A_n} f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

Hence, it is reasonable to define Independence by p.d.f., i.e.

> **Remark.** $X_1, \ldots, X_n$ are independent if
> $$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i)$$

7.1.1 Multivariable change of variables

Just remember what you learn in multivariable calculus, or implicit/explicit function theorem in advanced calculus. The intuition is:

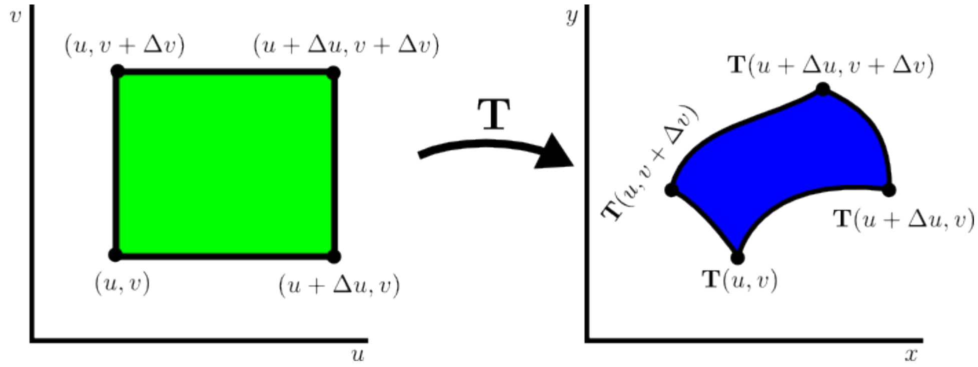


Figure 7.1: Transformation T stretch the small rectangle, so we need to give a "different unit" $|\det J_T|$ or $\left|\frac{\partial(v,w)}{\partial(x,y)}\right|$. source.

Usually you will be given transformation $(x, y) = T(v, w)$.
Then, $dxdy = \left|\frac{\partial(v,w)}{\partial(x,y)}\right| dvdw$. The new p.d.f. $g(v, w)$

$$g(v, w) = f(T(w, v))\,|J_T|$$

> **Exercise.** Exercise 3,6,11

Extra

- Counterexample to Fubini theorem

$$[f(i,j)] = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & \cdots \\ \vdots & \vdots & 0 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & 0 & \cdots \end{bmatrix}$$

  source

hi

Chapter 8

RC 8

Recitation 8

Application of

1 Dec. 18:30

Things that you can migrate from discrete to cont.

For 1D cont. r.v., apart from the definition of expectation, variance, and covariance, any other things depend on them still work, including

- Expectation of some function defined on random variables.
- Variance of a sum of r.v. is the sum of covariances.
- Inequalities(Markov's and Chebyshev...)
- Weak law of large numbers.

Mean vector and covariance matrix

((Cha)) Let $X = (X_1, \ldots, X_n)$ be an n-dimensional random vector (discrete or continuous).

- **Mean vector** $E[X]$
  The **mean vector** of X, denoted by $E[X]$ is the
- **Covariance matrix** Cov(X)
  The **covariance matrix** of X, denoted by Cov(X) is the nxn matrix Σ with

$$(\Sigma)_{ij} = \sigma_{ij} = \text{Cov}(X_i, X_j)$$

> **Example.** $X_1, \ldots, X_n$ are i.i.d. standard normal r.v.. Define a random vector $X = (X_1, \ldots, X_n)$ (standard normal vector).
> Then, Cov$(X) = I$

> **Remark.** The covaraince matrix Σ has some properies
>
> – Symmetric
> $$(\Sigma)_{ij} = \text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = (\Sigma)_{ji}$$
> Therefore, Σ is a **diagonalizable matrix**
>
> – positive semi definite In other word, given any $u \in \mathbb{R}^n$
> $$u^T \Sigma u \geq 0$$

- Linear transformation $Y = AX$
  $E[Y]$
  $$E[Y] = E[AX] = AE[X]$$
  Σ
  $$\sigma_{ij} = \text{Cov}((AX)_i, (AX)_j) = \sum_{1\le p,q\le n}^{n} a_{ip}a_{jq}\text{Cov}(X_p, X_q)$$
  Rearrange it you get
  $$\Sigma_Y = A\Sigma_X A^T$$

One useful fact to remember is

Theorem 8.0.1. Given a standard normal random vector X, and let $Z = \mu + AX$, where $\mu \in \mathbb{R}^n$ and A is invertible.
Then, the joint p.d.f. of Z is

$$h(z) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(\Sigma))^{\frac{1}{2}}} e^{-\frac{1}{2}(z-\mu)^T\Sigma^{-1}(z-\mu)}$$

where $\Sigma = AA^T$ is Σ_{AX}.

To give a simplified idea on page 74,75 of (Cha) and proposition 13.

Example (diagonalization). Although it is impractical to compute diagonalization of a matrix by hand, but it give you some idea.
If you are given a mulivariate normal distribution $Z \sim N(0, \Sigma)$, where

$$\Sigma = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$$

Find a matrix A such that each component of AZ is independent random variables.

Sol. We have

$$\Sigma \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and

$$\Sigma \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Let $v_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \end{bmatrix}^T, v_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \end{bmatrix}^T$, we define $V = [v_1, v_2]$
Then,

$$\Sigma = V \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} V^T$$

By checking the form of joint p.d.f. of Z, we have

$$\begin{aligned} h(z) &= \frac{1}{(2\pi)(\det(\Sigma))^{\frac{1}{2}}} e^{-\frac{1}{2}(z)^T\Sigma^{-1}(z)} \\ &= \frac{1}{(2\pi)(2\cdot 4)^{\frac{1}{2}}} e^{-\frac{1}{2}(z)^T\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}V^T(z)} \end{aligned} \tag{8.1}$$

Hence, the new gaussian normal vector $Y = V^TZ$ will have diagonal covaraince matrix, i.e. diagonal.

Exercise. HW9-6

Gaussian Stability

Remember in RC2 we introduced the stability of Poission random variables. Here we will show that

Theorem 8.0.2 (Gaussian stability). If $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, and g is a n-dim standard normal vector, then the random variable

$$X = \sum_{i=1}^{n} a_i g_i$$

has the Gaussian distribution N(0, $\sum_{i=1}^{n} a_i^2$)

Proof. This can be shown by considering the sum of two and then using induction. ((Cha) page 63, proposition 11).
Here we give a proof with more of linear algebra fasion. ((Pan19) Theorem 4.1)
Main idea:

- Find the complement space to $q_1 = \frac{a}{|a|}$, and thus the transformation matrix Q is unitary.
- For $Y = QX$, calculate the marginal p.d.f. of y_1. It will be easy because y_i are independent.

■

Since $a_i g_i \sim N(0, a_i^2)$, this means that the sum of independent Gaussian random variables is also Gaussian with the variance equal to the sum of variances.

Exercise. HW9-5

Hint: HW9-4

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.