NOTEBOOK FOR MA215 PROBABILITY

Lecturer: Prof.Hong

Author
Hongli Ye
Southern University of Science and Technology
2024 Fall

Contents

Preface

This is a self-made lecture note for Prof. Hong's MA215 Probability course at SUSTech in the Fall 2024 semester

Before reading, it is important to note that: these notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

If you find any typos or mistakes in this notes, please contact with `12311501@mail.sustech.edu.cn`, your suggestion will be very helpful!

1 Lecture 1 Basic of Probability 2024.09.12

Theorem 1.1. Basic principle of counting
Suppose there are two experiments. Experiment 1 has n results and experiment 2 has m results.
Then together there are $m \times n$ possible outcomes.

This basic theorem could be extended to many finite experiments by induction.

Definition 1.1. Permutation
Permutation means the different ordered arrangement of objects.

Theorem 1.2. Suppose we have n objects. Then there are $n! = \Pi_{i=1}^{n}(i) = 1 \times 2 \times \cdots \times n$ possible permutations.

Theorem 1.3. There are n objects, of which n_1 are alike, n_2 are alike,$\ldots$,n_r are alike.
Then there are $\frac{n!}{n_1! \times n_2! \times \ldots n_r!}$ possible outcomes.

Definition 1.2. Combination
Combination refers to selecting items from a set where order does not matter.

Theorem 1.4. If we choose r objects from a total of n differents objects at a time, then the # possible combinations of $\binom{n}{r}$

Theorem 1.5. Binomal Theorem
For any positive integer $n \geq 1$

$$(x+y)^k = \Sigma_{k=0}^{n}(\binom{n}{k}x^k y^{n-k})$$

Definition 1.3. Induction
Mathematical Induction is a proof method for natural numbers, consisting of a base case and an inductive step to show a statement holds for all natural numbers.

Mathematical Induction's basic step:

1. Basic step: The case holds when $n = 1$
2. Inductive step: Assume $n = k$ holds for some $k \geq 1$. Then $n = k + 1$ holds.

Quiz 1.1. From 8 women and 6 men, a committee of 3 men and 3 women is to be formed. How many different committees?

1. 2 of the men refuse to serve together?
2. 2 of the women refuse to serve together?
3. 1 man and 1 woman refuse to serve together?

2 Lecture 2 Probability Space 2024.09.19

Probability Space includes Sample Space, Events and Probability Measure.
Probability Space is a special case of measure theory.

Definition 2.1. Sample Space
The sample space S is the set of all possible outcomes of an experiment.

Definition 2.2. Event
An event is a subset of the sample space S, denoted $E \subset S$

Definition 2.3. Set Operation
Let E, F be two events and S is the sample space.

1. **Union**: $E \cup F = \{x|x \in E \text{ or } x \in F\}$
2. **Intersection**: $E \cap F = \{x|x \in E \text{ and } x \in F\}$
3. **Complement**: $E^c = \{x|x \notin E \text{ and } x \in S\}$
4. **Different**: $E - F = \{x|x \in E \text{ or } x \notin F\}$

Definition 2.4. Extension: $\sigma -$ algebra
Let $\mathcal{X}$ be a non-empty set. $\mathcal{F}$ is said to be a σ-algebra if:

1. $\mathbb{X} \in \mathbb{F}$
2. If $A \in \mathcal{F}, A^c \in \mathcal{F}$
3. If $A_1, A_2 \cdots \in \mathcal{F},$ then $\bigcup_{i=1}^{\infty}(A_i) \in \mathcal{F}$

Theorem 2.1. De Morgan's Law
For each $n \geq 1$, we have

$$(\bigcup_{i=1}^{n}(E_i))^c = \bigcap_{i=1}^{n}(E_i^c)$$

$$(\bigcap_{i=1}^{n}(E_i))^c = \bigcup_{i=1}^{n}(E_i^c)$$

Axiom 2.1. Axiom of Probability Let S be a sample space. For each event E, the probability $P(E)$ satisfies:

1. $0 \leq P(E) \leq 1$
2. $P(S) = 1$
3. For any sequence of mutually exclusive events $E_1, E_2 \ldots$, we have:

$$\Sigma_{i=1}^{\infty} P(E_i) = 1$$

Theorem 2.2. Basic corollaries:

1. $P(E) = 1 - P(E^c)$
2. If $E \subset F$, then $P(E) \leq P(F)$
3. $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
4. Inclusion-Exclusion Identity:(Extension of the line above)

$$P(\bigcup_{i=1}^{n}) = \Sigma_{i=1}^{n} P(E_i) - \Sigma_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \cdots + (-1)^{n+1} P(\bigcap_{i=1}^{n}(E_i))$$

Quiz 2.1. There are N cards numbered as $1, 2, \ldots, N$. Pick 1 card uniformly at random. Write down the number and return the card¿ Repeat for n times ($n > N, n = N, n < N$), we get a sequence $(x_1, x_2 \ldots, x_n)$.

1. P(the sequence is strictly increasing)
2. P(the sequence is non-decreasing)

3 Lecture 3 Conditional Probability and Independence 2024.09.26

Definition 3.1. Conditional Probability
For 2 events E,F such that $P(E) > 0$. The conditional probability F occurs given that E has occurred is denoted by:

$$P(F|E) = \frac{P(F \cap E)}{P(E)}$$

This definition give a new perspective into the conditional probability:

Theorem 3.1. If each outcome of a finite sample space is equally likely, then we may compute the conditional probability of the form $P(F|E)$ by using E as the reduced sample space.

Theorem 3.2. Multiplication Law
For events E, F, we have:

$$P(E \cap F) = P(E) \times P(F|E)$$

More generally:

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) \times P(E_2|E_1) \times P(E_3|(E_1 \cap E_2)) \dots P(E_n| \bigcap_{i=1}^{n-1}(E_i))$$

Theorem 3.2 is also called **chain rule**, which can be used in induction or some other methods.

Definition 3.2. Independence
For two events E, F. We say E and F are independent if:

$$P(E \cap F) = P(E) \times P(F) \text{ or } P(F|E) = P(F)$$

Theorem 3.3. Total Probability Formula
Let $A_1, A_2, \dots, A_n$ be mutually exclusive with $S = \bigcup_{k=1}^{n}(A_k)$.
Then $\forall$ event B:

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

Theorem 3.4. Bayes's Theorem
Let $A_1, A_2, \dots A_n$ be mutually exclusive so that $S = \bigcup_{k=1}^{n}(A_k)$.
Then $\forall$ event B:

$$P(A_j|B) = \frac{P(B|A_j) \times P(A_j)}{\Sigma_{i=1}^{n} P(B|A_i)P(A_i)}$$

Quiz 3.1. A gambler has a fair coin and a two-headed coin in his pocket.

1. He selects one of the coins at random; when he flips it, it shows heads. What is the probability that it is the fair coin?
2. Suppose that he flips the same coin a second time and, again, it shows heads. Now what is the probability that it is fair coin?
3. Suppose that he flips the same coin a third time and it shows tails. Now what is the probability that it is the fair coin?

4 Lecture 4 Discrete Random Variable 2024.10.10

Definition 4.1. Discrete Random Variable
A Random Variable $X : S \longrightarrow \mathbb{R}$.
If we take on at most a countable number of possible values is called discrete random R.V.

For example: 80 students, for which 70 are male. Choose 1 uniformly at random. Do this for 4 times. Let $X = \#$ of male students chosen.

Then X is a discrete R.V. taking values of {0,1,2,3,4}

Moreover:

$$\forall\ k \in \{0, 1, 2, 3, 4\}.P(X = k) = \binom{4}{k}(\frac{7}{8})^k(\frac{1}{8})^{1-k}$$

This is the probability mass functor of X.

Definition 4.2. Probability Mass Functor
For a discrete random variable X, we can define the probability mass functor(p.m.f), where $p(m)$ of X by

$$p(m) = P(X = m)$$

Definition 4.3. Special Random Variable

1. A random variable is said to be a **Bernoulli** random variable with parameter $p \in [0, 1]$ if:
   $$P(X = 0) = 1 - p \text{ and } P(X = 1) = p$$
   We say $X \sim \text{Bernoulli}(p)$
2. If we toss a coin independently for n times and let $X = \#$ of heads coming up, then X is said to be a **Binomial** random variable with parameter $p \in [0, 1]$
   Denoted by $X \sim \text{Bin}(n, p)$

The possible mass functor of $\text{Bin}(n, p)$ is:

$$P(X = k) = \binom{n}{k}p^k(1 - p)^{n-k}$$

For example: Alice is in a class of 80 students, after 100 independent trials. We count X as the # of times where Alice is picked. Then $X \sim \text{Bin}(100, \frac{1}{80})$

Remark: Binomial R.V. equals n times the addition of Bernoulli R.V.

Definition 4.4. Poisson Random variable

Let $X = \text{Bin}(n, \frac{\lambda}{n})$ for some $\lambda > 0$.

Then let $n \to \infty$, we can get a new p.m.f, which is the p.m.f of Poission R.V. :

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \ \forall k \geq 0$$

Denoted by $X \sim \text{Poission}(\lambda)$

Definition 4.5. Geometry Random Variable

There is a coin having probability $p \in (0, 1)$ of coming up heads. Toss the coin util it shows up head. Let $X = \#$ of tosses needed.

Then $X \sim \text{Geometric}(p)$, then p.m.f. of which is :

$$P(X = k) = (1 - p)^{k-1}p \ \forall k \geq 1$$

Denoted by $X \sim \text{Geometric}(p)$

The definition seems to be different from the Geometry Random Variable in Statistics. But they are actually the same.

Coupon Collector Problem:

Pick one card uniformly at random, record the number and then return the card. Repeat until we collect all the n numbers.

What is the average number of trials needed?

Definition 4.6. Expectation

For a discrete random variable, the expectation of X is defined by:

$$E(X) = \Sigma_{k=1}^{n} kP(X = k)$$

Quiz 4.1. Jim is conducting random walk on the real line starting from 0. For each time, independently of anything else, he moves one steps to the right with probability p, and to the left with probability $1 - p$. Let X_n be the position of Jim at time n. Find $P(X_n = k)$ for each $-n \leq k \leq n$

5 Lecture 5 Continuous Random Variable 2024.10.12

Definition 5.1. Probability Density Function
A non-negative function $f : (-\infty, \infty) \longrightarrow [0, \infty]$ is called a probability density function(p.d.f.), if:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Definition 5.2. Continuous Random Variable
A random variable X is called a continuous random variable if exists a p.d.f. f such that:

$$\forall a, b \in \mathbb{R},\ P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Remark: Let $b = a$ to get:

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x)\,dx = 0$$

Definition 5.3. Uniform Random Variable
A random variable $X \sim Uniform(\alpha, \beta)$, if the p.d.f. of X is:

$$f(x) = \frac{1}{\beta - \alpha} 1_{(\alpha,\beta)}(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{, otherwise} \end{cases}$$

Indicator function:

$$1_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

It is also called as characteristic function.
For example: Let $X \sim Unif(1, 5)$, find $P(X > 3.5)$
Solution:

$$P(X > 3.5) = P(X \geq 3.5)$$

$$P(X \geq 3.5) = lim_{b\to\infty} P(3.5 \leq X \leq b)$$

$$P(X \geq 3.5) = \int_{3.5}^{\infty} f(x)\,dx = \frac{3}{8}$$

Definition 5.4. Exponential Randon Variable
We say a X is an exponential random variable with parameter $\lambda > 0$ if the p.d.f. is:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Definition 5.5. Memoryless
We say a random variable is memoryless if:

$$P(X > t + s | X > t) = P(X > s) \; \forall t, s > 0$$

It is easy to prove that all exponential random variables are memoryless.

Theorem 5.1. If X is memoryless, then $X \sim Exp(\lambda)$ for some $\lambda > 0$.
Moreover, we proved that memoryless random variable equals to exponetial random variable.

For example: Let X be a continuous random variable with p.d.f.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x = 0 \end{cases}$$

Calculate $P(50 \leq X \leq 150)$
Solution:

1. Use $\int_{\infty}^{\infty} f(x)\,dx = 1$, we can get $\lambda = \frac{1}{100}$
2. $P(50 \leq X \leq 150) = P(X \geq 50) - P(X > 150) = e^{-\frac{50}{100}} - e^{-\frac{150}{100}}$

Definition 5.6. Gamma function:

$$\Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1}\,dy$$

Moreover:

$$\text{If } \alpha = n \in \mathbb{N}, \text{ then } \Gamma(n) = (n-1)!$$

Definition 5.7. Gamma Random Variable
Let X be a Gamma Random Variable, denoted by $X \sim Gamma(n, \lambda)$,then its p.d.f. is:

$$f(x) = \frac{x^{n-1} e^{-x/\lambda}}{\lambda^n \Gamma(n)}, \quad x > 0$$

In fact, if $X_1, X_2, \ldots, X_n$ are independent $Exp(\lambda)$, then

$$X_1 + X_2 + \cdots + X_n \sim \text{Gamma}(n, \lambda)$$

We can understand the Gamma random variable in both two ways.

Definition 5.8. Normal Random Variable
We say a $X \sim N(\mu, \sigma^2)$ is a normal random variable if the density is:

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

Definition 5.9. Expectation
For a continuous random variable X, the expectation of X is defined by

$$EX = E(X) = E[X] = \int_{\infty}^{\infty} x f(x)\, dx$$

For any function g, we have:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

Note: Expectation is actually a integration of a measurement.

Theorem 5.2. Properties of expectation
Let X,Y be two random variables:

1. $\forall c \in \mathbb{R}, E(c) = c$
2. If $X \geq 0$, then $EX \geq 0$
3. If $c \in \mathbb{R}, E(cX) = cEX$
4. $E[X+Y] = E[X] + E[Y]$

By properties 3 and 4, we know that expectation is **linear**.

For example: $X \sim Unif(0,1)$

$$EX = \int_{\infty}^{\infty} x f(x)\, dx = \frac{1}{2}$$

Definition 5.10. Variance
The variance of X is given by:

$$Var(X) = E[(X - EX)^2]$$

Moreover, we could also calculate by:

$$Var(X) = E(X^2) - (EX)^2$$

Quiz 5.1. Prove

$$\mu = \int_{-\infty}^{\infty} xf(x)\,dx \text{ and } \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

where:

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

6 Lecture 6 Expectation and Variance of special random variable

Theorem 6.1. Expectation and Variance of C.R.V

1. For $X \sim Exp(\lambda)$, we have:

$$E(X) = \frac{1}{\lambda} \text{ and } Var(X) = \frac{n}{\lambda}$$

2. For $X \sim Gamma(\alpha, \lambda)$, we have:

$$E(X) = \frac{1}{\lambda}\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \text{ and } Var(X) = \frac{1}{\lambda^2}\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} - (\frac{1}{\lambda}\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)})^2$$

3. For $X \sim N(\mu, \sigma^2)$, we have:

$$E(X) = \mu \text{ and } Var(X) = \sigma^2$$

Theorem 6.2. Expectation and Variance of D.R.V

For $X \sim Bernoulli(p)$, we have:

$$E(X) = p \text{ and } Var(X) = p(1-p)$$

For $X \sim Bin(n,p)$, we have:

$$E(X) = np \text{ and } Var(X) = np(1-p)$$

For $X \sim Poission(\lambda)$, we have:

$$E(X) = \lambda \text{ and } Var(X) = \lambda$$

For $X \sim Geo(p)$, we have:

$$E(X) = \frac{1-p}{p} \text{ and } Var(X) = \frac{1-p}{p^2}$$

Definition 6.1. Cumulative Distribution Function
For a random variable X, the cumulative distribution function(c.d.f.) of X is:

$$F_X(b) = P(X \leq b)$$

We notice that:

1. For discrete random variable:

$$F(b) = \Sigma_{m=-\infty}^{[b]} P(X = m)$$

2. For continuous random variable:

$$F'(b) = f(b)$$

But random variable have forms instead of these two kinds. See the quiz below:

Quiz 6.1. The cumulative distribution function of X is

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \leq x < 1 \\ \frac{2}{3} & 1 \leq x < 2 \\ \frac{11}{12} & 2 \leq x < 3 \end{cases}$$

(i) $P(x < 3)$

(ii) $P(x = 1)$

(iii) $P(x > \frac{1}{2})$

Theorem 6.3. 1. If $A_n \subset A_{n+1}, \forall n \geq 1$, then:

$$P(\bigcup_{n=1}^{\infty}) = lim_{n \to \infty} P(A_n)$$

2. If $B_{n+1} \subset B_n, \forall n \geq 1$, then:

$$P(\bigcap_{n=1}^{\infty}) = lim_{n \to \infty} P(B_n)$$

Theorem 6.4. Properties of Cumulative Distribution Function:
Let F be a cumulative distribution function.

1. F is a non-decreasing function, i.e.:
$$\forall\ a < b, F(a) \leq F(b)$$
2. $lim_{b\to-\infty}F(b) = 0, lim_{b\to\infty}F(b) = 1$
3. F is right continuous, i.e.:
$$\forall\ b \in \mathbb{R}, \forall lim_{n\to\infty}b_n = b,\ \text{ we have } lim_{n\to\infty}F(b_n) = F(b)$$
4. F has left limits, i.e.:
$$\forall b \in \mathbb{R}, \forall lim_{n\to\infty}(a_n) = a,\ \text{ we have } lim_{n\to\infty}F(a_n) = F(a^-) = F(x < a)$$

Use the **theorem6.3**, we could easily prove.
For example, we take $X \sim Bernoulli(p)$, then:

$$F_X(b) = \begin{cases} 1 & b \geq 1 \\ 1-p & 0 \leq b < 1 \\ 0, & b < 0 \end{cases}$$

It is a very traditional step function.

7 Lecture 7 Function of Random Varibale 2024.10.17

Theorem 7.1. If $X \sim N(\mu, \sigma^2)$, then:

$$Y = aX + b \sim N(a\mu + b, a^2\sigma^2), a, b \in \mathbb{R}$$

Quiz 7.1. If the pdf of X is:

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty$$

Show that $Y = \frac{1}{X}$ has the same pdf.

Theorem 7.2. Let X be a continuous random variable with pdf $f_X(x)$. Suppose $g(x)$ is a strictly monotonic(increasingly or decreasing), differentiable function. Then $Y = g(X)$ has a pdf:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)| & \text{if } y = g(x) \text{ for some } x. \\ 0 & if y \neq g(x), \forall\ x \end{cases}$$

Proof: $\forall y \in \mathbb{R}, F_Y(y) = P(Y \leq y) = P(g(x) \leq y)$. Assume g is increasing. Then $g(X) \leq y \leftrightarrow X \leq g^{-1}(y)$. So, $F_Y(y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$.

Theorem7.2 isn't useful since it has too many restrictions.

Now we do a summary on how to find a probability density function of $Y = g(X)$

1. Find the cdf of $Y = g(X)$, which means do some simple calculation.
2. Differentiate to find the density.
3. Specify in what region the result holds.

Theorem 7.3. Let $F(x)$ be the cdf of any random variable. Define for each $x \in (0, 1)$:

$$F^{-1}(x) = sup\{y \in \mathbb{R} : F(y) < x\}$$

8 Lecture 8 Multi-variables 2024.10.24

Definition 8.1. Joint cumulative distribution function:
For any random variables X, Y the joint cumulative distribution function of X and Y is defined by:

$$F(a, b) = P(X \leq a, Y \leq b)$$

Obviously, by the axiom of probability, we should have:

$$\lim_{a,b\to\infty}(F(a, b)) = 1$$

Notice that:

$$P(X \leq a) = lim_{b\to\infty}(P(X \leq a, Y \leq b))$$

Denote as $P(X \leq a) = F(a, \infty)$, and this is defined as **Marginal Probability Density Funcition**.

For discrete multi-random variables, we can define:

Definition 8.2. Joint probability mass function: When X, Y are both discrete random variables with p.m.f is given by p_X, p_Y.
The joint probability mass function:

$$p(X, Y) = P(X = x, Y = y)$$

Similar to the definition above, we have **Marginal Probability Mass Function:**

$$p_X(x) = \Sigma_y P(X = x, Y = y).$$

Similar to a single variable, we can also define independence in multi-variables.

Definition 8.3. Independent random variables
We say X, Y are independent if $\forall A, B \in \mathbb{R}$,

$$P(X \in A, Y \in B) = p(X \in A)P(Y \in B)$$

From the two definitions above, we could induce that:

Theorem 8.1. Two discrete random variables X, Y are independent if and only if $\forall x, y \in \mathbb{R}$,

$$p(x, y) = P_X(x)P_Y(y)$$

Definition 8.4. Jointly continuous
We say X and Y are jointly continuous if there exist a function $f(x, y0$ such that

$$\forall\ C \subset \mathbb{R}^2, P((X, Y) \in C) = \int_{(x,y)\in C} f(x, y)\, dxdy$$

The function **f(x,y)** is called the **joint probability distribution funcition** of X and Y.

Definition 8.5. joint cumulative distribution function:
The joint c.d.f. is then given by:

$$f_X(x) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x, y)\, dxdy$$

The definition of independence is still the same as before.

Definition 8.6. Expectation
For any joint p.m.f $p(x, y)$ or joint p.d.f $f(x, y)$, we have a $\forall$ function $g : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$,

$$E(g(X, Y)) = \Sigma_m \Sigma_n (g(m, n)p(m, n))$$

For example:

$$g(X, Y) = 1_{X\in A} 1_{Y \in B}$$

$$E[g(X, Y)] = E[1_{X\in A, Y\in B}] = P(X \in A, Y \in B) = \Sigma_m \Sigma_n p(m, n)$$

Or, in continuous situations:

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(X,Y)f(X,Y)\,dxdy$$

We need to notice that $1_{X\in A}$ and $1_{Y\in B}$ are beneficial **Characteristic functions**.

Another Example: A man and a woman promised to meet at 12 : 30P.M.. Assume the time they arrive are X and Y independently and satisfy:

$$X \sim \text{Unif}(12:15, 12:45)$$

$$Y \sim \text{Unif}(12:00, 1:00)$$

1. Calculate $P(\text{the man arrive first})$

$$X \sim \text{Unif}(-0.5, 0.5)Y \sim \text{Unif}(-1, 1)$$

$$\begin{aligned} P(X<Y) &= \int 1_{(X<Y)} \cdot f(x,y)\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \mathrm{d}x \int_{x}^{1} \mathrm{d}y \\ &= \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} (1-x)\ \mathrm{d}x \\ &= \frac{1}{2} \end{aligned}$$

2. Find the probability that the first to arrive waits no longer than 5 minutes.

$$\begin{aligned} P(|X-Y| < \frac{5}{30}) &= \iint 1_{(|X-Y|<\frac{1}{6})} \cdot f(x,y)\ \mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \mathrm{d}x \int_{x-\frac{1}{6}}^{x+\frac{1}{6}} \mathrm{d}y \\ &= \frac{1}{6} \end{aligned}$$

Quiz 8.1. The joint probability distribution function: of X, Y is :

$$f(x,y) = \begin{cases} c & \text{if } x^2+y^2 \le R^2 \\ 0 & \text{otherwise} \end{cases}$$

a) Find c

b) Find the marginal probability distribution functions of $f_X(x)$ and $f_Y(y)$.

This is the uniform distribution in circle plates.

Definition 8.7. Bivariate Normal Distribution
The joint probability of bivariate normal distribution is given by:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times exp\{-\frac{1}{2(1-\rho^2)^2}[(\frac{x-\mu_x}{\sigma_x})^2 + (\frac{y-\mu_y}{\sigma_y})^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\}$$

We denote it as $(X,Y) \sim N(\mu_x, \mu_y; \sigma_x^2, \sigma_y^2; \rho)$

Definition 8.8. Covariance
The covariance of X, Y is:

$$Cov(X,Y) := E[(X-EX)(Y-EY)]$$

By simple calculation, we know that:

$$Cov(X,X) = E[(X-EX)^2] = Var(X)$$

Now that we have expanded a single variable into bivariable, how can we get higher dimensions?

We could use **Matrix Form** to gain a beautiful expression of any finite dimension normal distribution.
If we let:

$$\begin{aligned} \vec{x} &= (x,y) \\ \vec{\mu} &= (\mu_x, \mu_y) \\ \Sigma &= \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \\ \Rightarrow \quad \det(\Sigma) &= \sigma_x^2\sigma_y^2(1-\rho^2) \\ \Sigma^{-1} &= \frac{1}{1-\rho^2} \cdot \begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix} \end{aligned}$$

Then we have:

$$f(x,y) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \cdot \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})^T\right\}$$

This is a very beautiful structure with general form.

Theorem 8.2. In Bivariate Normal Distribution, the X, Y independent are equivalent with:

1. $\rho = 0$
2. $Cov(X, Y) = 0$

9 Lecture 9 Sum of Independent Random Variables 2024.10.29

To understand the structure below better, we introduce the convolution of two functions.

Definition 9.1. Convolution: Let f and g be two functions

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)\,d\tau$$

Theorem 9.1. Sum of independent random variables
Let X, Y be independent continuous random variables. And $Z = X + Y$, then we have:

$$F_Z(z) = \int_{-\infty}^{\infty} F_X(z-y)f_Y(y)dy$$

Differentiate to obtain:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy \\ &= (f_X * f_Y)(z) \end{aligned}$$

Now we compute an example:
$X, Y \sim Exp(\lambda)$, and they are independent.
Compute the p.d.f. of $X + Y$:
Solution:

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} f_x(z-y)f_Y(y)\,dy \\ &= \int_0^z \lambda e^{\lambda(z-y)} \times \lambda e^{-\lambda y}\,dy \\ &= \lambda^2 e^{-\lambda z} z \end{aligned}$$

By observation, we can easily notice that $X + Y \sim Gamma(2, \lambda)$

Quiz 9.1. If the joint p.d.f. of (X, Y) is :

$$f(x,y) = \begin{cases} \frac{1}{2}(x+y)e^{-(x+y)}, & 0 < x, y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Find the p.d.f. of $Z = X + Y$.

10 Lecture 10 Conditional Distribution 2024.11.07

We begin with some easy conclusions:

Theorem 10.1. Here we discuss the sum of independent discrete random variables.

1. **Poisson:** If $X \sim Poisson(\lambda_1)$ and $Y \sim Poisson(\lambda_2)$ are two independent discrete random variables, then $X + Y \sim Poisson(\lambda_1 + \lambda_2)$
2. **Binomial:** If $X \sim Bin(n, p)$ and $Y \sim Bin(m, p)$ are two independent discrete random variables, then $X + Y \sim Bin(n + m, p)$

Proof:

1. For Poisson:

$$\begin{aligned} P(X+Y=n) &\overset{?}{=\!=} \frac{(\lambda_1+\lambda_2)^n}{n!} \cdot e^{-(\lambda_1+\lambda_2)} \\ \text{[Total Probability]} &= \sum_{k=0}^{n} P(X+Y=n|\, X=k) \cdot P(X=k) \\ &= \sum_{k=0}^{n} P(Y=n-k) \cdot P(X=k) \\ &= \sum_{k=0}^{n} \frac{\lambda_2^{n-k}}{(n-k)!} \cdot e^{-\lambda_2} \cdot \frac{\lambda_1^{n-k}}{k!} \cdot e^{-\lambda_1} \\ &= e^{-(\lambda_1+\lambda_2)} \cdot \frac{1}{n!} \cdot \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} \\ &= \frac{(\lambda_1+\lambda_2)^n}{n!} \cdot e^{-(\lambda_1+\lambda_2)} \end{aligned}$$

2. For Binomial:

$$\begin{aligned}
&\forall\, 0 \le N \le n+m \\
&P(X+Y=N) = \sum_{k=0}^{N\wedge n} P(Y=N-k|X=k)P(X=k) \\
&\qquad = \sum_{k=0}^{N\wedge n} \binom{m}{N-k} p^{N-k}(1-p)^{m-(N-k)} \binom{n}{k} p^k (1-p)^{n-k} \\
&\qquad = \sum_{k=0}^{N\wedge n} \binom{m}{N-k}\binom{n}{k} p^N (1-p)^{m+n-N}
\end{aligned}$$

Since we don't know the specific value of N, we need to do a classification discussion.

(a) $0 \le N \le n \wedge m$

$$\sum_{k=0}^{N} \binom{m}{N-k}\binom{n}{k} p^N (1-p)^{m+n-N} = \binom{m+n}{N} p^N (1-p)^{m+n-N}$$

(b) $n \vee m \le N \le n+m$

$$\sum_{k=0}^{n} \binom{m}{N-k}\binom{n}{k} p^N (1-p)^{m+n-N} = \binom{m+n}{N} p^N (1-p)^{m+n-N}$$

(c) $n \wedge m \le N \le n \vee m$ Similar hence omitted.

Definition 10.1. Conditional probability mass function:
If X, Y are two discrete random variables, we define the conditional probability mass function of X given $Y = y$:

$$\begin{aligned}
P_{X|Y}(x|y) &= P(X=x|Y=y) \\
&= \frac{p(x,y)}{P_Y(y)}
\end{aligned}$$

For example:
Suppose the $p(x,y)$ of (X,Y) is:

$$\begin{aligned}
p(0,0) &= 0.4 \quad p(0,1) = 0.2 \\
p(1,0) &= 0.1 \quad p(1,1) = 0.3
\end{aligned}$$

Find the conditional distribution of X given $Y = 1$.

proof:

$$\begin{aligned}
p(Y=1) &= p(0,1) + p(1,1) = 0.5 \\
P(X=0|\,Y=1) &= \tfrac{P(X=0,Y=1)}{P(Y=1)} = \tfrac{p(0,1)}{0.5} = \tfrac{2}{5} \\
P(X=1|\,Y=1) &= \tfrac{P(X=1,Y=1)}{P(Y=1)} = \tfrac{p(1,1)}{0.5} = \tfrac{3}{5}
\end{aligned}$$

Similarly, we can also define:

Definition 10.2. Conditional cumulative distribution function: If X, Y are two discrete random variables, we define the conditional cumulative distribution function of X given $Y = y$:

$$\begin{aligned} F_{X|Y}(x|y) &= P(X \leq x|Y = y) \\ &= \Sigma_{m \leq x} P(X = m|Y = y) \\ &= \Sigma_{m \leq x} \frac{P(m, y)}{P(Y = y)} \end{aligned}$$

For example:

If X and Y are independent R.V.s with $X \sim Poisson(\lambda_1)$, $Y \sim Poisson(\lambda_2)$, calculate the conditional distribution of X given $X + Y = n$.

proof: $\forall\, 0 \leq k \leq n$:

$$\begin{aligned} P(X = k|\, X + Y = n) &= \frac{P(X = k, X + Y = n)}{P(X + Y = n)} \\ &= \frac{e^{-\lambda_1} \cdot \frac{\lambda_1^k}{k!} \cdot e^{-\lambda_2} \cdot \frac{\lambda_2^{n-k}}{(n-k)!}}{e^{-(\lambda_1 + \lambda_2)} \cdot \frac{(\lambda_1 + \lambda_2)^n}{n!}} \\ &= \frac{n!}{k!(n-k)!} \cdot \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \cdot \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k} \\ &= \binom{n}{k} \cdot \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \cdot \left(1 - \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{n-k} \\ &\sim Bin(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}) \end{aligned}$$

Quiz 10.1. Joint probability density function of (X, Y) is:

$$f(x, y) = \begin{cases} xe^{-x(y+1)} & 0 < x, y < \infty \\ 0 & \text{otherwise} \end{cases}$$

1. Find $f_{X|Y}(x|y)$
2. Find $f_{Y|X}(y|x)$

Now we can see the **Total Probability Formula** from a new perspective.

Definition 10.3. Total Probability Formula:
Let $A_1, \ldots, A_n$ be mutually exclusive and $S = \bigcup_{k=1}^n A_k$. Then:

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

The continuous version:

$$P(B) = \int_{-\infty}^{\infty} P(B|Y=y) f_Y(y)\, dy$$

11 Lecture 11 Multiple Substitution 2024.11.12

Definition 11.1. Continuous version of total probability formula:

$$P(B) = \int_{-\infty}^{\infty} P(B|Y=y) \cdot f_Y(y)\, dy$$

$$\left(f_Y(y) = \frac{P(B \cap \{Y=y\})}{P(Y=y)} \right)$$

$P(B|Y)$ is a P.V. of Y.

$$P(B) = E[P(B|Y)] = \int_{-\infty}^{\infty} P(B|Y=y) \cdot f_Y(y)\, dy$$

Now recall the Bivariate Normal Distribution:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right] \right\}$$

Let:

$$\begin{cases} X \sim \mathcal{N}((\mu_x, \sigma_x^2)) \\ Y \sim \mathcal{N}(\mu_y, \sigma_y^2) \end{cases}$$

Given $Y = y$, the conditional pdf of X is:

$$\begin{aligned} f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} &= \frac{f(x,y)}{\frac{1}{\sqrt{2\pi\sigma_y^2}} \cdot e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}} \\ &\sim N(\mu_x + \rho \cdot \frac{\sigma_x}{\sigma_y}(y-\mu_y), \sigma_x^2(1-\rho^2)) \end{aligned}$$

If $\rho = 0$, then $X \sim N((\mu_x, \sigma_x^2)),\ Y \sim N(\mu_y, \sigma_y^2)$. i.e. They are independent.

Quiz 11.1. The joint pdf of (X, Y) is:

$$f(x,y) = \begin{cases} 2xe^{x^2-y} & 0 < x < 1, y > x^2 \\ 0 & \text{otherwise} \end{cases}$$

Find $f_{Y|X}(y|x)$ and $P(Y \geq \frac{1}{4}|X = x)$

proof:

1. $$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = e^{x^2-y}$$

2. $$P(Y \geq \frac{1}{4}|X = x) = \int_{x\vee\frac{1}{4}}^{\infty} f_{Y|X}(y|x)dy = \begin{cases} e^{x^2-\frac{1}{4}} & , 0 < x < \frac{1}{4} \\ 1 & , \frac{1}{2} < x < 1 \end{cases}$$

Now we introduce a important method in solving integral problems.

Theorem 11.1. Multiple Substitution:
Let $U = g(X, Y)$ and $V = h(X, Y)$.
The joint pdf of $(U, V) = (g(X, Y), h(X, Y))$ is

$$f_{UV}(u,v) = |J(u,v)| \cdot f_{XY}(q(u,v), r(u,v))$$

For example:
Suppose (X, Y) are independent $N(0, 1)$.

$$f(x,y) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}}$$

Let

$$\begin{cases} X = R \cdot cos\theta \\ Y = R \cdot sin\theta \end{cases} \quad , (R, \theta) \in \mathbb{R}^+ \times [0, 2\pi)$$

Consider (X, Y) jointly continuous R.V. with joint pdf $f(x, y)$.
Define:

$$\begin{cases} U = g(X, Y) \\ V = h(X, Y) \end{cases}$$

Let $K = \{(x, y) \in \mathbb{R}^2,\ f(x, y) > 0\}$
Set $G = \{(g(x, y), h(x, y)) \in \mathbb{R}^2, (x, y) \in K\} = \{(U, V)\}$

$$K = \mathbb{R}^2,\ G = [0, \infty) \times [0, 2\pi)$$

The map from K to G is bijective.

Solve

$$\begin{cases} u = g(x,y) \\ v = h(x,y) \end{cases}$$

To get

$$\begin{cases} x = q(u,v) \\ y = r(u,v) \end{cases}$$

Let

$$J(u,v) = \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} = \begin{vmatrix} \frac{\partial q(u,v)}{\partial u} & \frac{\partial q(u,v)}{\partial v} \\ \frac{\partial r(u,v)}{\partial u} & \frac{\partial r(u,v)}{\partial v} \end{vmatrix}$$

Theorem: The joint pdf of $(U,V) = (g(X,Y), h(X,Y))$ is

$$f_{UV}(u,v) = |J(u,v)| \cdot f_{XY}(q(u,v), r(u,v))$$

Hence we have:

$$f_R(R) = R \cdot e^{-\frac{R^2}{2}}$$
$$f_\Theta(\Theta) = \frac{1}{2\pi}$$

$\Rightarrow \Theta \sim$Unif(0,2π) Independent.

Definition 11.2. Box-Muller Algorithm:
Let U_1, U_2 be independent Unif$(0,1)$. Note that:

$$F_R(r) = \int_0^r R \cdot e^{-\frac{R^2}{2}} dR = 1 - e^{-\frac{r^2}{2}}$$

Then $F_R^{-1}(u) = \sqrt{-2\ln(1-u)}$
By letting

$$V_1 = \sqrt{-2\ln(1-U_1)} \left(or = \sqrt{-2\ln(U_1)} \right)$$

we get: $V_1 \overset{d}{=} R$

$$V_2 = 2\pi U_2 \sim \text{Unif}(0, 2\pi) \Rightarrow V_2 \overset{d}{=} \Theta$$

Then:

$$(V_1, V_2) \sim (R, \Theta)$$

So:

$$\begin{cases} X = V_1 cos \cdot V_2 \\ Y = V_1 sin \cdot V_2 \end{cases}$$

Independent $\quad X, Y \sim \mathcal{N}(0,1)$

This algorithm comes from the idea of solving Gauss Integral and give a clear way of how to transform uniform random variables into normal distributed random variables.

Definition 11.3. General Form of Expectation:
For a discrete R.V. X, the expectation of X is:

$$EX = \sum_{m=-\infty}^{\infty} m \cdot P(X = m)$$

given $E|X| < \infty$.

$$E|X| = \sum_{m=-\infty}^{\infty} |m| \cdot P(X = m) < \infty$$

For a continuous R.V. X, the expectation is:

$$EX = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Given:

$$E|X| = \int_{-\infty}^{\infty} |x| \cdot f(x)\, dx < \infty$$

$$Var(X) = E[(X - EX)^2] = \sum_{m=-\infty}^{\infty} (m - EX)^2 \cdot P(X = m)$$

12 Lecture 12 Order Statistic 2024.11.14

Definition 12.1. Order Statistic:
Let $X_1, X_2 \ldots, X_n$ be independent identically distributed random variables with a common probability density function f and commutative density function F
Then we define an ordered sequence of $X_1, X_2, \ldots, X_n$

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}$$

Theorem 12.1. The joint probability density function of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is:

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots, x_n) = n! \times f(x_1)f(x_2)\ldots f(x_n) \times 1_{x_1<x_2<\cdots<x_n}$$

Proof:
Using Infitesimal Method: $\forall \epsilon > 0$ small, note

$$\begin{aligned}
&P(X_{(1)} \in (x_1 - \frac{\epsilon}{2}, x_1 + \frac{\epsilon}{2}, \cdots, X_{(n)} \in (x_n - \frac{\epsilon}{2}, x_n + \frac{\epsilon}{2}))) \quad \forall x_1 < x_2 < \cdots < x_n \\
&= \int_{t_1 \in (x_1 - \frac{\epsilon}{2}, x_1 + \frac{\epsilon}{2})} dt_1 \cdot \int_{t_2 \in (x_2 - \frac{\epsilon}{2}, x_2 + \frac{\epsilon}{2})} dt_2 \cdot \int \cdots \int dt_n \\
&\approx f(x_1, \cdots, x_n) \cdot \epsilon^n + o(\epsilon^n)
\end{aligned}$$

On the other hand,

$$
\begin{aligned}
LHS &= n!P(X_{(1)} \in (x_1 - \frac{\epsilon}{2}, x_1 + \frac{\epsilon}{2}, \cdots, X_{(n)} \in (x_n - \frac{\epsilon}{2}, x_n + \frac{\epsilon}{2}))) \\
&= n! \prod_{k=1}^{n} P(X_k \in (x_k - \frac{\epsilon}{2}, x_k \in \frac{\epsilon}{2})) \\
&\approx n! \prod_{k=1}^{n} [f(x_k)\epsilon]
\end{aligned}
$$

Hence:

$$
f(x_1, x_2, \cdots x_n) = n! \prod_{k=1}^{n} f(x_k)
$$

For example:
If $X_1, \cdots, X_n$ are i.i.d. Unif(0,1), then:

$$
f_{X_{(1)}, X_{(2)}, \cdots, X_{(n)}}(x_1, x_2, \cdots, x_n) = n! \quad 0 < x_1 < x_2 < \cdots < x_n < 1
$$

For any $1 \leq j \leq n$, the marginal pdf of $X_{(j)}$ is:

$$
\begin{aligned}
f_{X(j)}(x) &= \frac{n!}{(n-j)!(j-1)!} \cdot P(X_1, \cdots, X_{j-1} < x, X_j = x, X_{j+1}, \cdots, X_n > x) \\
&= \frac{n!}{(n-j)!(j-1)!} \cdot x^{j-1}(1-x)^{n-j}
\end{aligned}
$$

$$
P(X_1 < x) = x \quad P(X_1 > x) = 1 - x \quad P(X_1 = x) \approx 1
$$

13 Lecture 13 2024.11.21

For example:

$X \sim \mathcal{N}(0,1)$, $E|X|^\alpha$

For example: If $X, Y \sim \mathcal{N}(0,1)$, independent, find $E(|X^2 + Y^2|^\alpha)$ for $\alpha \in \mathbb{R}$

Solution:

$$
\begin{aligned}
E(|X^2 + Y^2|^\alpha) &= \iint (x^2 + y^2)^\alpha \cdot f_{XY}(x, y) dxdy \\
&= \iint (x^2 + y^2)^\alpha \cdot \frac{1}{2\pi} \cdot e^{-\frac{1}{2}(x^2+y^2)} dxdy \\
&= \int_0^\infty r dr \cdot \int_0^{2\pi} d\theta \cdot (r^2)^\alpha \cdot \frac{1}{2\pi} \cdot e^{-\frac{1}{2}r^2} \\
&= \int_0^\infty r^{2\alpha+1} \cdot e^{-\frac{1}{2}r^2} dr \\
&\overset{t=\frac{1}{2}r^2}{=\!=\!=} \int_0^\infty e^{-t} \cdot (\sqrt{2\pi})^{2\alpha+1} \cdot \frac{1}{\sqrt{2\pi}} \\
&= \begin{cases} 2^\alpha \cdot \Gamma(\alpha + 1) & \forall \alpha > -1 \\ \infty & \forall \alpha \leq -1 \end{cases}
\end{aligned}
$$

Theorem 13.1. Properties of Expectation:
Let $X_1, X_2, \cdots, X_n$ be R.V.s such that $E|X_i| < \infty, \forall i$.

1. If $c_0, c_1, \cdots, c_n \in \mathbb{R}$, then:

$$E[c_0 + c_1X_1 + \cdots c_nX_n] = c_0 + c_1EX_1 + \cdots c_nEX_n$$

2. If $X_1, X_2, \cdots X_n$ are independent, then $\forall g_1, g_2, \cdots g_n$, we have:

$$E[\prod_{k=1}^{n} g_k(X_k)] = \prod_{k=1}^{\infty} E[g_k(X_k)]$$

Remark: $g_1(X_1), \cdots, g_n(X_n)$ are also independent.

For example:
A group of n men and n women is lined up at random.

1. Find the expectation number of men who have a women next to them.

Solution:

Let:

$$X_i = \begin{cases} 1 & \text{, If man i has a woman next to him.} \\ 0 & \text{, Otherwise.} \end{cases}$$

$$EX_{total} = E[X_1 + \cdots + X_n] = \sum_{k=1}^{n} EX_k = nEX_1$$

$$EX_1 = P(X_1 = 1) = \frac{1}{2n} \cdot \frac{n}{2n-1} + \frac{1}{2n} \cdot \frac{n}{2n-1} + (2n-2) \cdot \frac{1}{2n} \cdot (1 - \frac{(n-1)(n-2)}{(2n-1)(2n-2)})$$

$$\Rightarrow EX_1 = \frac{3n-1}{4n-2}$$

$$\Rightarrow EX_{total} = \frac{n(3n-1)}{4n-2}$$

2. Repeat part 1°, but assuming that the group is randomly seated at a round table.

$$EX_{table} = n \times 2n \times \frac{1}{2n} \cdot (1 - \frac{(n-1)(n-2)}{(2n-1)(2n-2)}) = \frac{3n^2}{4n-2}$$

For example:
20 individuals, 10 married couples, 5 tables.

1. If the seating is at random, find the expected number of couples that are seated at the same table.

$$EX_{total} = 10EX_1 = 10 \times \frac{3}{19} = \frac{30}{19}$$

2. If 2 men and 2 women are randomly chosen to be seated at each table, repeat 1°

$$EX_{total} = 10EX_1 = 10 \times \frac{2}{10} = 2$$

Quiz 13.1. Suppose there are n people coming to a party, they take turns to sit, for each table it has p probability not sitting it, if there is no table for him to sit, he will start a new table.

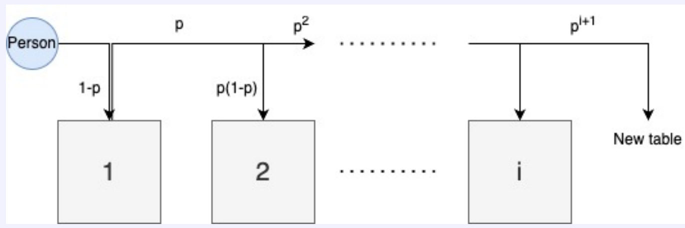

Figure 1: Quiz 13.1

What is the expectation of the numbers of table when all n people have taken their seats.

For example, this question has been discussed in lecture 4.
(Coupon - Collection problem) Suppose there are N different types of coupons, and each time it is equally likely to be any of the N types. Find expected number of coupons needed before obtaining a complete set of all the N types.

Solution:

Define X_i, $0 \le i \le N-1$ to be the number of additional coupons that needed to obtain after i distinct types have been collected.

$$X_{total\ number} = X_0 + X_1 + \cdots + X_{N-1}$$

$$P(X_i = k) = \left(\frac{i}{N}\right)^{k-1} \cdot \frac{N-i}{N} \ \Rightarrow X_i \sim Geometric\left(\frac{N-i}{N}\right)$$

$$EX_{total\ number} = \sum_{i=1}^{N-1} EX_i = \sum_{i=0}^{N-1} E\left(Geometric\left(\frac{N-i}{N}\right)\right) = N \cdot \sum_{k=1}^{N} \frac{1}{k} \approx N \cdot lnN$$

14 Lecture 14 Application of expectation 2024.11.26

Theorem 14.1.

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} P(\bigcap_{j=1}^{k} E_{i_j})$$

Let:

$$X_i = \begin{cases} 1 & \text{if } E_i \text{ occurs} \\ 0 & \text{if not} \end{cases}$$

Then it is easy to prove.

The variance of X is given by:

$$Var(X) = E[(X - EX)^2] = \begin{cases} \sum_{m=-\infty}^{n} (m-\mu)P(X=m) \\ \int_{-\infty}^{\infty} (x-\mu)f(x)\,dx \end{cases}$$

Now we introduce some properties of variance:

Theorem 14.2. Properties of variance:

1. If $EX = \mu$, $\mathrm{Var}(X) < \infty$, then:

   (a) $\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$, $\forall\, a, b \in \mathbb{R}$

   (b) $\mathrm{Var}(X) = \mathrm{E}((x-\mu)^2) < \mathrm{E}((x-c)^2)$, $\forall\, c \in \mathbb{R} \neq \mu$

2. Let $X_1, X_2, \ldots, X_n$ be random variables with $Var(X_i) < \infty$, then:

   (a) $\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n}\sum_{j=1}^{n}[E(X_iX_j) - \mu_i\mu_j]$

   (b) If $X_1, X_2, \ldots, X_n$ are independent.

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$

Proof:

1.a Easy to prove.

1.b

$$\begin{aligned} E[(x-c)^2] &= E[(x-\mu+\mu-c)^2] \\ &= E[(x-\mu)^2] + 2E[(x-\mu)(\mu-c)] + (\mu-c)^2 \\ &= E[(x-\mu)^2] + (\mu-c)^2 \\ &> E[(x-\mu)^2] \end{aligned}$$

2.a

$$\begin{aligned} LHS &= E[(\sum_{i=1}^{n} X_i - E(\sum_{i=1}^{n} X_i))^2] \\ &= E[(\sum_{i=1}^{n}(X_i - \mu_i))^2] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} E(X_iX_j) - \mu_i EX_j - \mu_j EX_i + \mu_i\mu_j \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}[E(X_iX_j) - \mu_i\mu_j] \end{aligned}$$

2.b Easy to prove by definition.

Definition 14.1. Corelation:
Let X, Y be two random variables.

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

To prove it, we introduce a well-known theory.

Theorem 14.3. Cauchy-Schwartz Theorem:
For all random variables X, Y:

$$|E(XY)| \leq E(|XY|) \leq \sqrt{EX^2 \times EY^2}$$

Proof:
Let X, Y be two random variables. Consider: $E[aX+Y]$ as a function of a

$$\begin{aligned} E[(aX+Y)^2] &= E[a^2X^2 + 2aXY + Y^2] \\ &= E[X^2]a^2 + 2E[XY]a + E[Y^2] \\ &\geq 0 \forall a \in \mathbb{R} \\ \text{So:} & \\ \Delta &= (2E[XY])^2 - 4EX^EY^2 \geq 0 \end{aligned}$$

Now we can prove $\rho \in [-1, 1]$.

Definition 14.2. Uncorrelated:
If $rho = 0$, or $Cov(X,Y) = 0$, we say X and Y are uncorrelated.

It is obvious that if X, Y are independent, then they must be uncorrelated.

Quiz 14.1. Joint probability density function of X, Y is:

$$f(x,y) = \frac{1}{y}e^{-y-\frac{x}{y}}, x > 0, y > 0$$

Find $EX, EY, Cov(X,Y)$

Recall the conditional probability density function of X given $Y = y$ is:

$$f_{X|Y} = \frac{f(x,y)}{f_Y(y)}$$

Now we can define:

Definition 14.3. Conditional expectation:
Let X and Y be two random variables:

$$E(X|Y=y) = \int x f_{X|Y}(x|y)\,dx = \int x \frac{f(x,y)}{f_Y(y)}\,dx$$

Theorem 14.4. Total formula of probability:
If X, Y be two continuous random variables, then:

$$E(X) = \int E(X|Y=y) f_Y(y)\,dy$$

For example:

A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 3 hours of travel. The second door leads to a tunnel that will return him to the mine after 5 hours of travel. The third door leads to a tunnel that will return him to the mine after 7 hours. If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

Solution:

Let X be the total time until safety and Y be the first choice of the door.

$$X = \begin{cases} 3 & \frac{1}{3} \\ 5 + X' & \frac{1}{3} \\ 7 + X' & \frac{1}{3} \end{cases}$$

Since the miner is at all times equally likely to choose any one of the doors. So we know

$$\begin{cases} EX & = \frac{1}{3}E(X|Y=1) + \frac{1}{3}E(X|Y=2) + \frac{1}{3}E(X|Y=3) \\ EX & = EX' \end{cases}$$

So:

$$\begin{aligned} EX &= \frac{1}{3}(3+5+7) + \frac{2}{3}EX \\ EX &= 15 \end{aligned}$$

15 Lecture 15 Probability generating function 2024.11.28

Definition 15.1. Probability generating function:
For a non-negative discrete random variable X, define :

$$\forall s \in \mathbb{R}, G_X(s) = E[s^X] = \sum_{n=0}^{\infty} s^n P(X = n)$$

is a function of s

The series is absolutely convergent at least for $|s| \leq 1$.
For example:

E.g.1 If $X \sim Bernoulli(p)$, then:

$$G_X(s) = 1 - p + sp$$

E.g.2 If $X \sim Poisson(\lambda)$, then:

$$\begin{aligned} G_X(s) &= \sum_{n=0}^{\infty} s^n e^{-\lambda} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(s\lambda)^n}{n!} \\ &= e^{s\lambda - \lambda} \end{aligned}$$

E.g.3 If $X \sim Binomial(n, p)$, then:

$$G_X(s) = (1 - p + sp)^n$$

Theorem 15.1. Properties of probability generating function:

1. $$\forall k \geq 0, P(X = k) = \frac{1}{k!} G^{(k)}(0).$$

2. (a) $G_X(1) = 1$
   (b) $G^{(1)}(1) = EX$
   (c) $G^{(2)}(1) = E[X(X-1)]$
   (d) By the property above:

$$Var(X) = EX^2 - (EX)^2 = G^{(2)}(1) + G^{(1)}(1) - (G^{(1)}(1))^2$$

3. If $X_1, X_2, \ldots, X_n$ are independent, then for $Y = \sum_{i=1}^{n} X_i$, we have:

$$G_Y(s) = E[s^Y] = E[s^{X_1} \ldots s^{X_n}]$$

All are easy to check by definition.

Definition 15.2. Moment generating function:

$$M_X t = E(e^{tX}) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E[X^n]$$

Remark: It is possible that $E(X^n) < \infty, \forall$ but $M_X(t) = \infty$

Quiz 15.1. The probability density function of X is;

$$f(x) = \frac{1}{2} e^{-|X|}, \forall\, x \in \mathbb{R}$$

1. Compute $M_X(t)$.
2. Compute $E[X^n]$ by $M_X(t)$.

Definition 15.3. Characteristic function:
The characteristic function of X is defined by :

$$\Phi_X(t) = E[e^{itX}] = E[cos(tX)] + iE[sin(tX)], \forall\, t \in \mathbb{R}$$

The cumulative density function of X is uniquely determined by $\Phi(t)$.

$$\Phi(t) = \int_{-\infty}^{\infty} e^{itX} dF(x)$$

For example:

$$\text{If} X \sim N(\mu, \sigma^2), \text{ then } E[e^{itX}] = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$$

Inversion Formula:

$\forall\, a < b \in \mathbb{R}$.

$$\frac{1}{2\pi} \lim_{T \to \infty} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \Phi(t)\, dt = \frac{1}{2}(F(b) + F(b^-)) - \frac{1}{2}(F(a) + F(a^-))$$

Equals to $F(b) - F(a)$ if left continuous.

16 Lecture 16 2024.12.05

I am absent from this lecture due to some reasons.

Definition 16.1. central moment:
Let X be a continuous random variable, with mean value μ. Then we denote:

$$\mu_p' = E[|X - \mu|^p]$$

We call μ_p' the pth central moment of X.

17 Lecture 17 Inequality and convergence 2024.12.10

Theorem 17.1. Markov's Inequality:
For any random variable X with mean μ, we have:

$$P(|X| \geq t) \leq \frac{\mu}{t} \forall t \geq 0.$$

Theorem 17.2. Chebyshev's Inequality:
For any random variable X with mean μ, variance σ^2, we have:

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}, \;\; \forall t < 0$$

In the real world application, Chebyshev's Inequality is **more precise than** Marcov's Inequality since it needs more information of X

By applying these two inequalities, we can get a important theorem.

Theorem 17.3. Weak Law of Large Numbers:
Let $X_1, X_2, \dots, X_n$ are i.i.d. random variables, with mean μ and variance σ^2, then:

$$\forall \epsilon > 0, \lim_{n \to \infty} P(|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu| > \epsilon) = 0.$$

Proof:
By Chebyshev Inequality:

$$P(|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu| > \epsilon) \leq \frac{Var(\frac{X_1 + X_2 + \cdots + X_n}{n})}{\epsilon^2}$$

By independence of $\{X_i\}$, we get:

$$\begin{aligned} Var(\frac{X_1 + X_2 + \cdots + X_n}{n}) &= \frac{\sum_{i=1}^n Var(X_i)}{n^2} \\ &= \frac{n\sigma^2}{n^2} \\ &= \frac{\sigma^2}{n} \end{aligned}$$

So:

$$P(|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n} = \frac{1}{n} \times (\frac{\sigma^2}{\epsilon^2})$$

Quiz 17.1. Suppose $X \geq 0$ has $EX = 10$.

1. Find a upper bound of $P(X > 15)$.
2. Suppose $Var(X) = 3$. Find a better bound for (a).

Definition 17.1. Four types of convergence:

1. **convergence in probability:**(p)
   Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables. We say X_n converges to X in probability in probability if:
   $$\forall\, \epsilon > 0,\, P(|X_n - X| > \epsilon) \to 0(n \to \infty)$$
2. **almost surely convergence:**(a.s.)
   $$P(\lim_{n\to\infty} X_n = X) = 1$$
3. **Convergence in distribution:**(d)
   For any $x \in \mathbb{R}$ such that x is a continuity point of $F_X(b) = P(X \leq b)$
   $$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x)$$
4. L^p **convergence for** $p > 0(L^p)$
   $$\lim_{p\to\infty} E|X_n - X|^p = 0$$

From strong convergence to weak convergence:

$$L^p \to p \to d$$
$$a.s. \to p \to d$$

There is direct relation between l^p and a.s convergence.

18 Lecture 18 2024.12.19

Theorem 18.1. Central Limit Theorem:
Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d random variables, with mean μ and variance σ^2. Then:
$$\lim_{n\to\infty} P(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \leq x) = P(Z \leq x), \forall\, x \in \mathbb{R}$$
Where $Z \sim N(0,1)$. That is:
$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} Z\ (n \to \infty)$$

This is actually describing how does $\sum_{i=1}^n X_i$ converges to $n\mu$

Theorem 18.2. Continuity Theorem:
Let $\xi_1, \xi_2, \dots, \xi_n$ be a sequence of random variables with characteristic function $\Phi_n(t) = E[e^{it\xi_n}]$. If $\Phi_n(t) \to \Phi(t), \forall\, t \in \mathbb{R}$ and $\Phi(t) = E[e^{itX}]$ for some random variables X and $\Phi(t)$ is continuous at 0. Then:

$$\xi_n \xrightarrow{d} X$$

Using **characteristic function sequence's** convergence, we can get convergence in distribution.

Quiz 18.1. Let $X_1, X_2, \dots, X_{10}$ be i.i.d. $\sim Unif(0,1)$.Let $S_{10} = \sum_{i=1}^{n} X_i$. Find $P(S_{10} > 6)$ using:

1. Markov's Inequality
2. Chebyshev's Inequality
3. Central limit theorem

From the result of the quiz, we can observe which one is the most accurate.

Theorem 18.3. Chernoff Bound:
$\forall\, a > 0$.

- $P(X \geq a) \leq \min_{t>0}\{E[\frac{e^{tX}}{e^{ta}}]\} \quad \forall\, t > 0$
- $P(X \leq a) \leq \min_{t<0}\{E[\frac{e^{tX}}{e^{ta}}]\} \quad \forall\, t < 0$

Application of this Chernoff bound:
For $Z \sim N(0,1)$,

$$P(Z \geq a) \leq \frac{E[e^{tZ}]}{e^{ta}} = e^{\frac{1}{2}t^2 - at}$$

Since this is correct for all $t > 0$.

$$P(Z \geq a) \leq e^{-\frac{1}{2}a^2} \quad \forall\, a > 0$$

This is a pretty good estimation for the normal distribution.

Theorem 18.4. Jensen's Inequality:
Take a convex function $f(x)$, then we have:

$$f(E[X]) \leq E[f(x)] \quad \forall\, x \in \mathbb{R}$$

19 Lecture 19 2024.12.24

Theorem 19.1. Strong Law of large numbers:
Let $X_1, X_2, \ldots, X_n$ be a sequence of o i.i.d.

Theorem 19.2. Borel-Cantelli Lemma:
For a sequence of events $\{A_n\}$ if $\sum_{k=1}^{\infty} P(A_k) < \infty$ then :

$$P(A_n \text{ infinitely often }) = 0$$

Proof:

$$\begin{aligned} E[\sum_{n=1}^{\infty} 1_{A_n}] &= \sum_{n=1}^{\infty} E[1_{A_n}] \\ &= \sum_{n=1}^{\infty} P(A_n) < \infty \end{aligned}$$

Quiz 19.1. If $X_n \xrightarrow{p} X$ and $P(| Y |< \infty) = 1$, then prove that:

$$X_n Y \xrightarrow{p} XY$$

Theorem 19.3. Dominated Convergence:
If $X_n \xrightarrow{a.s.} X$ and $|X_n|^p \leq Y, \forall n \geq 1$ and $EY < \infty$, then:

$$E[| X_n - X |^p] \to 0$$

So in some given situations, $a.s.$ can induce the L^p convergence.

A Answer for Quizes

1. **Quiz 1**
   (a) 896
   (b) 1000
   (c) 910
2. **Quiz 2**
   (a) $\frac{\binom{N}{n}}{N^n}$
   (b) $\frac{\binom{N+n-1}{n}}{N^n}$
3. **Quiz 3**
   (a) $\frac{1}{3}$
   (b) $\frac{1}{5}$
   (c) 1
4. **Quiz 4**
$$P(X_n = k) = \begin{cases} \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} (1-p)^{\frac{n-k}{2}} & \text{if } n+k \text{ odd} \\ 0 & \text{if } n+k \text{ even} \end{cases}$$
5. **Quiz 5**
6. **Quiz 6**
   (i) $\frac{11}{12}$
   (ii) $\frac{1}{6}$
   (iii) $\frac{3}{4}$
7. **Quiz 7**
8. **Quiz 8**
   (a) $c\ \frac{1}{\pi R^2}$
   (b) $f_X(x) = \frac{2}{\pi R^2}\sqrt{R^2 - x^2}$; $f_Y(y) = \frac{2}{\pi R^2}\sqrt{R^2 - y^2}$
9. **Quiz 9** $f_Z(z) = \frac{1}{2} z^2 e^{-z}$
10. **Quiz 10**
   (a) $f_{X|Y}(x|y) = (y+1)^2 x e^{-x(y+1)}$
   (b) $f_{Y|X}(y|x) = x e^{-xy}$
11. **Quiz 11**
12. **Quiz 12**
13. **Quiz 13**

14. **Quiz 14**

    (a) $EX = 1$

    (b) $EY = 1$

    (c) $Cov(X, Y) = 1$

15. **Quiz 15**

16. **Quiz 16**

17. **Quiz 17**

    (a) $\frac{2}{3}$

    (b) $\frac{3}{25}$

18. **Quiz 18**

    (a) $\frac{5}{6}$

    (b) $\frac{5}{12}$

    (c) $P(Z > \sqrt{1.2})$

B Extension Problem

References