Compiled by Tian Xiao

ST2131 Probability

AY2020/21 Semester 2

1. Combinatorial Analysis

1.1 Basic Principle of Counting

Suppose two experiments are to be performed: Experiment 1 has m outcomes, Experiment 2 has n outcomes, then together there are mn outcomes.

1.2 Permutation

Suppose there are n distinct objects, then total number of permutations is $n!$.

Suppose there are n objects and n_a of them are **alike**, then total number of permutations is $\frac{n!}{n_a!}$.

Suppose there are n people sitting **in a circle**, then total number of permutations is $(n-1)!$.

1.3 Combination

Suppose there are n distinct objects, from which we choose r as a group, then total number of combinations $\binom{n}{r} = \frac{n!}{r!(n-r)!}$.

For $1 \leq r \leq n$, $\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$.

Binomial Theorem: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

$\sum_{k=0}^{n} \binom{n}{k} = 2^n$. // $x = y = 1$.

$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$. // $x = -1$, $y = 1$.

$\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots$.

Suppose there are n distinct objects and we are to divide them into r groups of size $n_1, n_2, \ldots, n_r$, then total number of combinations $\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \ldots n_r!}$.

Multinomial Theorem: $(x_1 + x_2 + \cdots + x_r)^n = \sum_{n_1+n_2+\cdots+n_r=n} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \ldots x_r^{n_r}$.

Suppose $x_1 + x_2 + \cdots + x_r = n$, then total number of different **positive** integer-valued vectors $(x_1, x_2, \ldots, x_n)$ is $\binom{n-1}{r-1}$.

Suppose $x_1 + x_2 + \cdots + x_r = n$, then total number of different **non-negative** integer-valued vectors $(x_1, x_2, \ldots, x_n)$ is $\binom{n+r-1}{r-1}$.

2. Axioms of Probability

2,1 Sample Spaced and Events

The sample space, S, is the set of all possible outcomes of an experiment.

Any subset of S is an event.

2.2 Axioms of Probability

The probability, P, is a function satisfying:
(i) For any event E, $0 \leq P(E) \leq 1$;
(ii) $P(S) = 1$;
(iii) For any sequence of mutually exclusive events $E_1, E_2, \ldots$, $P(\bigcup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} P(E_k)$.

Proposition 2.1: $P(\emptyset) = 0$.

Proposition 2.2: For any finite sequence of mutually exclusive events $E_1, E_2, \ldots, E_n$, $P(\bigcup_{k=1}^{n} E_k) = \sum_{k=1}^{n} P(E_k)$.

Proposition 2.3: $P(E^c) = 1 - P(E)$.

Proposition 2.4: If $A \subset B$, then $P(A) \leq P(B)$.

Inclusion/Exclusion Principle: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. // what if there are n events?

2.3 Sample Spaces Having Equally Likely Outcomes

2.4 Probability as a Continuous Set Function

A sequence of events is **increasing** if $E_1 \subset E_2 \subset \cdots$

Proposition 2.6: $P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n)$.

3. Conditional Probability & Independence

3.1 Conditional Probability

The conditional probability of A given B, $P(A|B) = \frac{P(AB)}{P(B)}$.

Multiplication Rule: $P(AB) = P(A)P(B|A)$.

3.2 Bayes' Formulas

$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c)$.

Bayes' First Formula: Suppose $A_1, A_2, \dots, A_n$ partition S, then $P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k)$.

Bayes' Second Formula: Suppose $A_1, A_2, \dots, A_n$ partition S, then $P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^{n} P(A_k)P(B|A_k)}$.

The **odds** of an event A is $\frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A)}$.

3.3 Independent Events

A and B are **independent** if $P(AB) = P(A)P(B)$.

3,4 De Méré-Pascal Problem

3.5 Gambler's Ruin Problem

3.6 Algebra of Conditional Probability

Proposition 3.4: Let A be an event such that $P(A) > 0$, then the following three conditions hold:
(i) $0 \le P(B|A) \le 1$;
(ii) $P(S|A) = 1$;
(iii) For any sequence of mutually exclusive events $B_1, B_2, \dots$, $P(\bigcup_{k=1}^{\infty} B_k \,|A) = \sum_{k=1}^{\infty} P(B_k|A)$.

4. Discrete Random Variable

4.1 Random Variable

A random variable, X, is a mapping from the sample space to real numbers $X: S \to \mathbb{R}$.

4.2 Discrete Random Variable

A random variable is **discrete** if the range of X is either finite or countably infinite.

The **probability mass function**, p_X, is defined as

$$p_X(x) = \begin{cases} P(X = x) & \text{if } x = x_1, x_2, \dots; \\ 0 & \text{otherwise.} \end{cases}$$

$\sum_{k=1}^{\infty} p_X(x_k) = 1$.

The **cumulative distribution function**, F_X, is defined as $F_X(x) = P(X \le x)$.

4.3 Expected Value

$E(X) = \sum_x x\, p_X(x)$.

4.4 Expected Value of a Function of a Random Variable

$E\big(g(X)\big) = \sum_x g(x) p_X(x)$.

Corollary 4.2: $E(aX + b) = aE(X) + b$.

4.5 Variance and Standard Deviation

$Var(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$.

$\sigma(X) = \sqrt{Var(X)}$.

$Var(aX + b) = a^2 Var(X)$.

$\sigma(aX + b) = |a|\sigma(X)$.

4.6-4.8 Distributions of Discrete Random Variable

$E(X)$	$Var(X)$
Bernoulli Distribution, $Be(p)$: success or failure. $P(X = 1) = p$; $P(X = 0) = 1 - p$.	
p	$p(1-p)$
Binomial Distribution, $Bin(n, p)$: number of successes in n trials. $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.	
np	$np(1-p)$
Geometric Distribution, $Geom(p)$: number of trials required to obtain the first success. $P(X = k) = p(1-p)^{k-1}$.	
$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial Distribution, $NB(r, p)$: number of trails required to obtain r successes. $P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$.	

$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Poisson Distribution, $Po(\lambda)$: number of events occurring in a fixed interval if the events occur independently with a constant mean rate. $P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$.	
λ	λ
Hypergeometric Distribution, $H(n,N,m)$: number of red balls if we choose n balls from a set of m red balls and $N-m$ blue balls. $P(X=k)=\frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$.	
$\frac{nm}{N}$	$\frac{nm}{N}[\frac{(n-1)(m-1)}{N-1}+1-\frac{nm}{N}]$

4.9 Distribution Functions and Probability Mass Functions

Properties of Distribution Function:

(i) If $a<b$, then $F_X(a)\le F_X(b)$.

(ii) $\lim_{b\to\infty} F_X(b)=1$; $\lim_{b\to-\infty} F_X(b)=0$.

(iii) $\lim_{x\to b^-} F_X(x)$ always exists.

(iv) $\lim_{x\to b^+} F_X(x)=F_X(b)$.

5. Continuous Random Variable

5.1 Continous Random Variable

$p(a<X\le b)=\int_a^b f_X(x)dx$, where f_X is called the **probability density function** (p.d.f.) of X.

The **distribution function** of X, $F_X(x)=P(X\le x)=\int_{-\infty}^{x} f_X(t)dt$.

5.2 Expectation and Variance

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

$Var(X)=\int_{-\infty}^{\infty}\left(x-E(X)\right)^2 f_X(x)dx$.

Proposition 5.1:

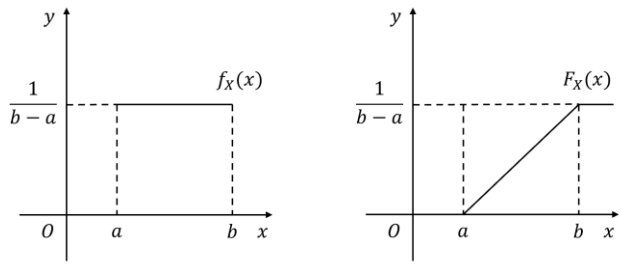
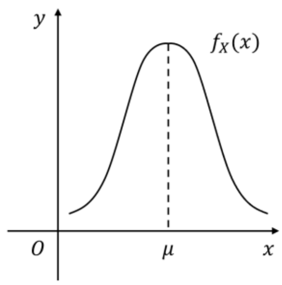
(i) $E[g(x)]=\int_{-\infty}^{\infty} g(x)f_X(x)dx$.

(ii) $E(aX+b)=aE(X)+b$.

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

Tail Sum Formula: Suppose X is a **non-negative** continuous random variable, then $E(X)=\int_0^{\infty} P(X>x)dx$.

5.3-5.6 Distributions of Continuous Random Variable

$E(X)$	$Var(X)$
Uniform Distribution, $U(a,b)$: 	
$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal Distribution, $N(\mu,\sigma^2)$: $f_X(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.  **3-σ Rule**: If an observation is taken from a normal population with mean μ and variance σ^2, then it is very likely (99.74%) that it lies within 3 standard deviation of the mean.	
μ	σ^2
Standard Normal Distribution, $Z\sim N(0,1)$: $f_Z(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$. **Normalisation**: If $X\sim N(\mu,\sigma^2)$, then $\frac{X-\mu}{\sigma}\sim N(0,1)$.	
0	1
Exponential Distribution, $Exp(\lambda)$: $f_X(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$. $F_X(x)=\begin{cases}0, & x\le 0\\ 1-e^{-\lambda x}, & x>0\end{cases}$. **Memoryless Property of Exponential Distribution**: $P(X>s+t\mid X>t)=P(X>s)$.	
$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Gamma Distribution, $Gamma(\alpha, \lambda)$: If events are occurring independently with a constant mean rate, then the amount of time one has to wait until a total of n events has occurred is a random variable which follows $Gamma(n, \lambda)$.

$$f_X(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & x < 0 \end{cases},$$

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

// $Gamma(1, \lambda) = Exp(\lambda)$.

// if $X \sim Gamma(\frac{n}{2}, \frac{1}{2})$, then $X \sim \chi^2(n)$.

$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$

Weibull Distribution, $W(v, \alpha, \beta)$:

$$f_X(x) = \begin{cases} \frac{\beta}{\alpha}\left(\frac{x-v}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-v}{\alpha}\right)^\beta}, & x > v \\ 0, & x \leq v \end{cases}.$$

// $W(0,1,\lambda) = Exp(\lambda)$.

$\alpha\Gamma(1 + \frac{1}{\beta})$	$\alpha^2[\Gamma\left(1+\frac{2}{\beta}\right) - \left(\Gamma\left(1+\frac{1}{\beta}\right)\right)^2]$

Cauchy Distribution parametrised with θ and positive α:

$$f_X(x) = \frac{1}{\pi\alpha[1+\left(\frac{x-\theta}{\alpha}\right)^2]}$$

not exist	not exist

Beta Distribution, $Beta(a, b)$:

$$f_X(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, & 0 < x < 1 \\ 0, & otherwise \end{cases},$$

where $B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$.

// $Beta(1,1) = U(0,1)$.

$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$

5.7 Approximations of Binomial Random Variables

Normal Approximation: $Bin(n, p) \approx N(np, np(1-p))$.

// good if $np(1-p) \geq 10$.

Continuity Correction: Suppose $X \sim Bin(n, p)$ and is approximated as $X \sim N(np, np(1-p))$, then:

$P(X = k) = P(k - \frac{1}{2} < X < k + \frac{1}{2})$;

$P(X \geq k) = P(X \geq k - \frac{1}{2})$;

$P(X \leq k) = P(X \leq k + \frac{1}{2})$.

Poisson Approximation: $Bin(n, p) \approx Po(np)$.

// working rule: if $p < 0.1$, put $\lambda = np$; if $p > 0.9$, put $\lambda = n(1-p)$ and work in terms of "failure".

5.8 Distribution of a Function of a Random Variable

Suppose $g(x)$ is a strictly monotonic, differentiable function of X. Then the probability density function of $Y = g(X)$ is given by $f_Y(y) =$

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

6. Jointly Distributed Random Variables

6.1 Joint Distribution Function

The **joint distribution function** of X and Y, $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$.

The **marginal distribution function** of X, $F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y)$.

$P(X > a, Y > b) = 1 - F_X(a) - F_Y(b) + F_{X,Y}(a, b)$.

$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_1, b_1) + F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1)$.

The **jointly probability mass function** of X and Y, $p_{X,Y}(x, y) = P(X = x, Y = y)$.

The **marginal probability mass function** of X, $p_X(x) = P(X = x) = \sum_{y \in \mathbb{R}} p_{X,Y}(x, y)$.

The **joint probability density function** of X and Y is $f_{X,Y}(x, y)$, where $P\left((X, Y) \in C\right) = \iint_{(x,y) \in C} f_{X,Y}(x, y) dx dy$.

The **marginal probability density function** of X and Y, $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$.

6.2 Independent Random Variables

Proposition 6.1: The following statements are equivalent for discrete random variables:

(i) Random variables X and Y are independent.

(ii) For all x and y, $p_{X,Y}(x, y) = p_X(x) p_Y(y)$.

(iii) For all x and y, $F_{X,Y}(x, y) = F_X(x) F_Y(y)$.

Proposition 6.2: The following statements are equivalent for continuous random variables:

(i) Random variables X and Y are independent.

(ii) For all x and y, $f_{X,Y}(x, y) = f_X(x) f_Y(y)$.

(iii) For all x and y, $F_{X,Y}(x, y) = F_X(x) F_Y(y)$.

Proposition 6.3: Random variables X and Y are independent if and only if there exist functions $g, h: \mathbb{R} \to \mathbb{R}$ such that for all $x, y \in \mathbb{R}$ we have $f_{X,Y}(x,y) = h(x)g(y)$.

<u>6.3 Sum of Independent Random Variables</u>

$F_{X+Y}(a) = P(X + Y \leq a) = \int_{-\infty}^{\infty} F_X(a-y) f_Y(y) dy = \int_{-\infty}^{\infty} F_Y(a-x) f_X(x) dx$. Here, F_{X+Y} is called the **convolution** of F_X and F_Y.

<u>6.4 Conditional Distributions (Discrete)</u>

The **conditional probability mass function** of X given that $Y = y$ is given by $p_{(X|Y)}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$.

The **conditional distribution function** of X given that $Y = y$ is given by $F_{(X|Y)}(x|y) = P(X \leq x | Y = y) = \sum_{a \leq x} p_{(X|Y)}(a|y)$.

Proposition 6.6: If X is independent of Y, then $p_{(X|Y)}(x,y) = p_x(X)$.

<u>6.5 Conditional Distributions (Continuous)</u>

The **conditional probability density function** of X given that $Y = y$ is given by $f_{(X|Y)}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$.

The **conditional distribution function** of X given that $Y = y$ is given by $F_{(X|Y)}(x|y) = P(X \leq x | Y = y) = \int_{-\infty}^{x} f_{(X|Y)}(t|y) dt$.

Proposition 6.7: If X is independent of Y, then $f_{(X|Y)}(x,y) = f_x(X)$.

<u>6.6 Joint Probability Distribution Function of Functions of Random Variables</u>

Proposition 6.8: Assume that the following conditions are satisfied:
(i) Let X and Y be jointly continuously distributed random variables with known joint probability density function;
(ii) Let U and V be functions of X and Y in the form $U = g(X,Y), V = h(X,Y)$, and we can uniquely solve X and Y in terms of U and V, say $x = a(u,v)$ and $y = b(u,v)$;
(iii) The functions g and h have continuous partial derivatives at all points (x,y) and $J(x,y) = \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix} \neq 0$ at all points (x,y);
Then the **joint probability density function** of U and V is given by $f_{U,V}(u,v) = f_{X,Y}(x,y)|J(x,y)|^{-1}$, where $x = a(u,v)$ and $y = b(u,v)$.
// what about there are more variables?

<u>6.7 Jointly Distributed Random Variables: $n \geq 3$</u>

Rather same.

7. Properties of Expectation

<u>7.1 Expectation of Sums of Random Variables</u>

$E[g(X,Y)] = \sum_y \sum_x g(x,y) p_{X,Y}(x,y)$.

$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$.

Monotone Property: If $X < Y$, then $E(X) < E(Y)$.

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

<u>7.2 Covariance, Variance of Sums, and Correlation</u>

The **covariance** of X and Y, $Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y)$.

If $Cov(X,Y) = 0$, then X and Y are uncorrelated.

Proposition 7.2: If X and Y are independent, then for any functions g, h, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.

Corollary 7.3: If X and Y are independent, then $Cov(X,Y) = 0$.

$Var(X) = Cov(X,X)$.

$Cov(X,X) = Cov(Y,X)$.

Variance of a Sum: $Var(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} Var(X_k) + 2\sum_{1 \leq i < j \leq n} Cov(X_i, X_j)$.

Under independence, variance of sum = sum of variances.

The **correlation (coefficient)** of X and Y, $\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(y)}}$.

$-1 \leq \rho(X,Y) \leq 1$.

$|\rho(X,Y)|$ near 1 implied linearity.

If X and Y are independent, then $\rho(X,Y) = 0$.

7.3 Conditional Expectation

$E[X|Y=y] = \sum_x x\, p_{(X|Y)}(x|y)$.

$E[X|Y=y] = \int_{-\infty}^{\infty} x\, f_{(X|Y)}(x|y)dx$.

$E[g(X)|Y=y] = \sum_x g(x) p_{(X|Y)}(x|y)$.

$E[g(X)|Y=y] = \int_{-\infty}^{\infty} g(x)\, f_{(X|Y)}(x|y)dx$.

$E[\sum_{k=1}^{n} X_k | Y=y] = \sum_{k=1}^{n} E[X_k|Y=y]$.

Proposition 7.4: $E[X] = E[E[X|Y]]$.

7.4 Moment Generating Functions

The **moment generating function** of X, $M_X(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p_X(x) & \text{(discrete)} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)dx & \text{(continuous)} \end{cases}$.

$E(X^n) = M_x^n(0)$, where $M_x^n(0) = \frac{d^n}{dt^n} M_x(t)|_{t=0}$.

Multiplicative Property: $M_{X+Y}(t) = M_x(t) M_Y(t)$.

Uniqueness Property: If $M_X(t) \equiv M_Y(t)$, then $X \equiv Y$.

Moment generating of function of various distributions:

Distribution	M.G.F.
$X \sim Be(p)$	$M_X(t) = 1 - p + pe^t$
$X \sim Bin(n,p)$	$M_X(t) = (1 - p + pe^t)^n$
$X \sim Geom(p)$	$M_X(t) = \frac{pe^t}{1-(1-p)e^t}$
$X \sim Po(\lambda)$	$M_X(t) = e^{\lambda(e^t - 1)}$
$X \sim U(\alpha,\beta)$	$M_X(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta - \alpha)t}$
$X \sim Exp(\lambda)$	$M_X(t) = \frac{\lambda}{\lambda - t}$
$X \sim N(\mu,\sigma^2)$	$M_x(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

7.5 Joint Moment Generating Functions

The **joint moment generating function** $M_{X_1,X_2,\ldots,X_n}(t_1,t_2,\ldots,t_n) = E[e^{t_1X_1 + t_2X_2 + \cdots + t_nX_n}]$.

It also has **uniqueness property**.

$X_1, X_2, \ldots, X_n$ are independent if and only if $M_{X_1,X_2,\ldots,X_n}(t_1,t_2,\ldots,t_n) = M_{X_1}(t_1) M_{X_2}(t_2) \ldots M_{X_n}(t_n)$.

Proposition 7.8: Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed normal random variables with mean μ and variance σ^2, then the sample mean $\bar{X}$ and sample variance S^2 are independent. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ and $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

8. Limit Theorems

8.1 Introduction

8.2 Chebyshev's Inequality and the Weak Law of Large Numbers

Markov's Inequality: Let X be a non-negative random variable. For $a > 0$, we have $P[X \geq a] \leq \frac{E[X]}{a}$.

Chebyshev's Inequality: Let X be a random variable with mean μ, then for $a > 0$, $P[|X - \mu| \geq a] \leq \frac{Var(X)}{a^2}$.

If $Var(X) = 0$, then X is constant.

The Weak Law of Large Numbers: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, with common mean μ. Then for any $\epsilon > 0$, $P\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0$ as $n \to \infty$.

8.3 Central Limit Theorem

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

Normal Approximation: Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables, each having mean μ and variance σ^2. Then, for large n, the distribution of $\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$ is approximately standard normal.

8.4 The Strong Law of Large Numbers

The Strong Law of Large Numbers: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu =$

$E(X_i)$. Then **with probability 1**, $\frac{X_1+X_2+\cdots+X_n}{n} \to \mu$ as $n \to \infty$.

8.5 Other Inequalities

One-sided Chebyshev's Inequality: X is a random variable with mean 0 and finite variance σ^2, then for any $a > 0$, $P(X \geq a) \leq \frac{\sigma^2}{\sigma^2+a^2}$.

Jensen's Inequality: If $g(x)$ is a convex function, then $E[g(X)] \geq g(E[X])$ provided that the expectations exist and are finite.

References

AY2020/21 Semester 2 ST2131 Lecture Notes by Prof. Chan Yiu Man.

Annex I: Cumulative Probability for Standard Normal Distribution, $P(Z < z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999
3.1	0.99903	0.99906	0.9991	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.9994	0.99942	0.99944	0.99946	0.99948	0.9995
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.9996	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.9997	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.9998	0.99981	0.99981	0.99982	0.99983	0.99983