Midterms Revision Guide

30.003 Probability and Statistics, Term 4 2019

Wei Min Cher

05 Jan 2020

Contents

1 W1: Probability and Statistics

1.1 Definitions

- Population: well defined collection of objects
- Sample: subset of population selected in certain manner
- Variable: any characteristic whose value may change from one object to another in population

- Probability: properties of populations known, question regarding sample taken from population are investigated **(deductive reasoning)**
- Statistics: characteristics of sample known from experiments, conclusions regarding population are made **(inductive reasoning)**

$$\begin{tikzcd} \text{Population} \arrow[r, bend left, "\text{Probability}"] & \text{Sample} \arrow[l, bend left, "\text{Statistics}"] \end{tikzcd}$$

1.2 Frequency

- Frequency: number of times value occurs in data set
- Relative frequency: fraction or proportion of times the value occurs

1.3 Range and mean

- Range: difference between largest and smallest sample values
- Mean: average of all values

- Population mean is denoted by μ
- Sample mean is denoted by $\bar{x}$, where

$$\bar{x} = \frac{\sum x_i}{n}, \text{ and } n \text{ denoting the number of data points}$$

1.4 Variance and standard deviation

- Variance: measures variability of data set
- Population variance is denoted by σ^2, where

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2, \text{ and } N \text{ denoting the size of the population}$$

- Sample variance is denoted by s^2, where

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\ , \text{ and } n \text{ denoting the size of the sample}$$

- Standard deviation is denoted as σ for population variance and s for sample variance, and is calculated either by:

$$\sigma = \sqrt{\sigma^2},\ \text{ or } s = \sqrt{s^2}$$

where σ^2 is the population variance and s^2 is the sample variance

  - Shortcut to calculate population variance:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu^2$$

1.5 Median

- Median: the middle value in a data set

1.6 Percentile

- value below which a given percentage of observations falls
- data set is ordered as $x'_1 \leq x'_2 \leq \cdots \leq x'_n$,
  where x'_1 and x'_n are the smallest and largest data values respectively
- x'_i corresponds to the $\frac{100(i-0.5)}{n}$th percentile

1.7 Sample space and events

- Sample space: the set of all possible outcomes of an experiment
  1. Collectively exhaustive
  2. Mutually exclusive
- Event: collection of outcomes contained in sample space Ω
  1. Simple event: exactly one outcome
  2. Compound event: > 1 outcome

1.8 Sample Space vs Population

- Sample space: contains mutually exclusive events
- Population: events can repeat many times

1.9 Set Theory

- Null event, $\varnothing$: event with no outcome
- Events A and B are mutually exclusive/disjoint if $A \cap B = \varnothing$

1.10 De Morgan's Laws

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$
$$A \cup B = A + B - A \cap B$$

1.11 Axiom of Probability

1. For any event A, P(A) $\leq$ 0.
2. P(Ω) = 1
3. Any infinite collection of mutually exclusive/disjoint events $A_1, A_2, A_3, \ldots, A_n$ satisfies

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = \sum_{i=1}^{\infty} P(A_i)$$

1.12 Properties of Probability

- For any event A, $P(A) + P(A^c) = 1$.
- $P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$
  $\because$ A and A^c are disjoint
- For any event A, $P(A) \leq 1$.
- For a null event $\varnothing$, $P(\varnothing) = 0$
  - Does **NOT** suggest A = $\varnothing$
- Similarly, P(A) = 1 does **NOT** suggest A = Ω

1.13 Equally likely outcomes

$$P(\text{equally likely event}) = \frac{1}{n}, \text{ where n is the number of equally likely events}$$

2 W1: Counting Technique

2.1 Finding probability

- Computing probability → counting

$$P(A) = \frac{N(A)}{N}$$

  - where $N(A)$ is the number of outcomes for event A,
    and N is the number of outcomes in the sample space

2.2 Tuple

- Group of k elements: k-tuple
- The 1$^{\text{st}}$ element is selected in n_1 ways; the 2$^{\text{nd}}$ element is selected in n_2 ways; the k$^{\text{th}}$ element is selected in n_k ways; such that *the elements are selected independently.*

2.3 Permutation

- Ordered subset
- Number of permutations of size k formed from n objects:

$$P_{k,n} = \frac{n!}{(n-k)!}$$

2.4 Combination

- Unordered subset of a group
- Number of combinations of size k formed from n objects:

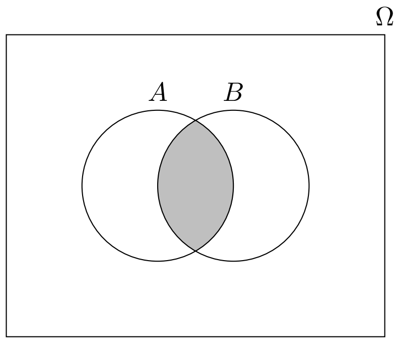
$$\binom{n}{k} \text{ or } C_{k,n} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!}$$

- Disregards the different outcomes due to order

3 W2: Conditional Probability

- Probability of event A given that event B has occurred: P(A|B)

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

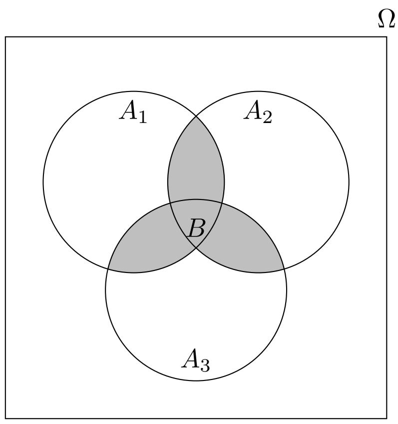


3.1 Law of Total Probability

- Let $A_1, A_2, \ldots, A_k$ be mutually exclusive and exhaustive events.

  For any other event B,

$$P(B) = \sum_{i=1}^{k} P(B \mid A_i)P(A_i)$$



3.2 Bayes' Theorem

$$\begin{aligned} P(A_j \mid B) &= \frac{P(A_j \cap B)}{P(B)} \\ &= \frac{P(B \cap A_j)}{P(B)} \\ &= \frac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^{k} P(B \mid A_i)P(A_i)} \end{aligned}$$

3.3 Independence of Random Variables

- Independence: occurrence/non-occurrence of one event has no bearing on the chance that the other will occur
  - $P(A \mid B) = P(A)$: A and B are independent
  - $P(A \mid B) \neq P(A)$: A and B are not independent
- Independence of A and B also implies $P(B \mid A) = P(B)$ if $P(A) > 0$

3.3.1 Multiplication Rule

- A and B are independent iff. $P(A \cap B) = P(A)P(B)$

3.3.2 Independence of several events

$$P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{ik}) = P(A_{i1})P(A_{i2}) \ldots P(A_{ik})$$

4 W2: Discrete Random Variable

4.1 Probability Mass Function (PMF) for Discrete RV

- For any pmf, $p(x) \leq 0$ and $\sum_{\text{all possible x}} p(x) = 1$

4.2 Bernoulli RV

- pmf of any Bernoulli RV:

$$p(x;\alpha) = \begin{cases} 1-\alpha, & \text{if } x = 0 \\ \alpha, & \text{if } x = 1 \\ 0, & \text{otherwise} \end{cases}$$

- α is a parameter, where $0 < \alpha < 1$

4.3 Bernoulli process

- A process with repeated independent trials
- 2 outcomes: 1 (success), 0 (failure)
- Success rate of trials is the same

4.4 Binomial distribution

- pmf of binomial RV:

$$p(x;n,p) = \begin{cases} C_{x,n} p^x (1-p)^{n-x}, & x = 0, 1, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

  - where n is the number of trials, and p is the success rate of each trial

4.5 Geometric distribution

- Probability distribution of number of Bernoulli trials X needed to get 1 success
- If $X = x$, $x - 1$ failures followed by success
- pmf of geometric RV:

$$p(x) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

  - where p is the success rate of each trial

4.6 Poisson distribution

- Used to model the number of occurrences of events in a time interval, where the average occurrence is λ
- pmf of Poisson RV:

$$p(x;\lambda) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, \ldots \\ 0, & \text{otherwise} \end{cases}$$

  - where λ is the parameter of Poisson distribution

4.7 Cumulative Distribution Function (CDF)

- CDF $F(x)$ of discrete RV X with pmf $p(x)$:

$$F(x) = P(X \leq x) = \sum_{y: y \leq x} p(y)$$

- $F(x)$ is the probability that the observed value is at most x
- Graph of $F(x)$ for discrete RV X is the linear combination of step functions, such that

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) = 1$$

5 W3: Expectation

5.1 Expected Value

- Expected value $E(X)$

$$E(X) = \mu_x = \sum_{x \in D} x \cdot p(x), \text{ provided that } \sum_{x \in D} |x| \cdot p(x) < \infty$$

- Expected value of a function $E[h(X)]$

$$E[h(X)] = \mu_{h(x)} = \sum_{x \in D} h(x) \cdot p(x)$$

- Expected value of a linear function $aX + b$

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

5.2 Variance

- Variance $V(X)$

$$V(X) = \textstyle\sum_{x \in D} (x - \mu)^2 p(x) = E[(X - \mu)^2], \text{ provided that the expectation exists}$$

OR

$$\text{Population variance, } \sigma^2 = V(X) = E(X^2) - [E(X)]^2$$

- Variance of a function $V[h(X)]$

$$V[h(X)] = \sum_{x \in D} \{h(x) - [E(X)]\}^2 \cdot p(x)$$

- Variance of a linear function $aX + b$

$$V(aX + b) = a^2 V(X)$$

$$\sigma_{aX+b} = |a| \sigma_x$$

5.3 Expected Value and Variance of Discrete PMFs

5.3.1 Bernoulli RV

- Expected value $E(X) = p$
- Variance $V(X) = p(1 - p)$

5.3.2 Binomial RV

- Expected value $E(X) = np$
- Variance $V(X) = np(1 - p)$

5.3.3 Geometric RV

- Expected value $E(X) = \dfrac{1}{p}$
- Variance $V(X) = \dfrac{1-p}{p^2}$

5.3.4 Poisson RV

- Expected value $E(X) = \lambda$
- Variance $V(X) = \lambda$

6 W3: Continuous Random Variable

6.1 Probability Density Function (PDF) for Continuous RV

- Probability described by the probability density function (pdf), measured between an interval

$$P(a \le X \le b) = \int_a^b f(x)dx$$

6.2 Uniform Distribution

$$\text{pdf } f(x; a, b) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

6.3 Exponential Distribution

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

6.4 Normal/Gaussian Distribution

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

6.5 Cumulative Distribution Function (CDF)

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)du$$

- Capital F means CDF, while small F means PDF
- For any a: $P(x > a) = 1 - F(a)$
- Between a and b: $P(a \le X \le b) = F(b) - F(a)$

6.5.1 Obtaining PDF from CDF

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

- The PDF is the derivative of the CDF.

6.6 Expected Value

- Expected value $E(X)$

$$E(X) = \mu_x = \sum_{x \in D} x \cdot p(x), \text{ provided that } \int_{\infty}^{\infty} |x| \cdot p(x) < \infty$$

- Expected value of a function $E[h(X)]$

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

- Expected value of a linear function $aX + b$

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

6.7 Variance

- Variance $V(X)$

$$\begin{aligned} V(X) = \mu_X^2 &= E[(X - \mu)^2] \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

- Variance of a linear function $aX + b$

$$V(aX + b) = a^2 V(X)$$

$$\sigma_{aX+b} = |a| \sigma_x$$

6.8 Expected Value and Variance of Continuous PDFs

6.8.1 Uniform RV

- Expected value $E(X) = \frac{1}{2}(a + b)$
- Variance $V(X) = \frac{1}{12}(b - a)^2$

6.8.2 Exponential RV

- Expected value $E(X) = \frac{1}{\lambda_E}$
- Variance $V(X) = \frac{1}{\lambda^2}$

7 W4: Useful Distributions

7.1 Poisson Approximation of Binomial Distributions

For any binomial distribution where n is large and p is small, such that $np > 0$,

$$b(x; n, p) \approx p(x; \lambda), \text{ where } \lambda = np$$

- Approximation can be safely applied if $n > 50$ and $np < 5$

7.2 Relationship between Poisson and Exponential Distributions

- Poisson distribution: Often used to model the number of occurrence of events in a time interval
- Exponential distribution: Often used to model the elapsed time between two successive events

Let $X_1, X_2, \ldots$ be the time when the 1st, 2nd, ... event occur.
The probability of waiting not more than t for the first event is $P(X_1 \leq t)$.

Deriving via Poisson Distribution

$$\begin{aligned} P(X_1 \leq t) &= 1 - P(X_1 > t) \\ &= 1 - P(\text{no event in [0, t]}) \\ &= 1 - \frac{\lambda^0 e^{-\lambda}}{0!} \\ &= 1 - e^{-\lambda} \\ &= 1 - e^{-\alpha t}, \text{where } \lambda = \alpha t \end{aligned}$$

Deriving via Exponential Distribution

$$\begin{aligned} P(X_1 \leq t) &= 1 - P(X_1 > t) \\ &= 1 - \int_t^\infty \alpha e^{-\alpha x} dx \\ &= 1 - \left[\frac{\alpha}{-\alpha} e^{-\alpha x}\right]_t^\infty \\ &= 1 - e^{-\alpha t} \end{aligned}$$

The rate of occurrence α in the Poisson distribution is the parameter of the exponential distribution.

7.3 Memoryless Property of Exponential Distribution

- Distribution of waiting time until a certain event does not depend on how much time has elapsed
- e.g. P(bulb can last for 600 h) = P(bulb can last for 900 h | bulb can last for 300 h)

7.4 Standard Normal Distribution

- Parameters: mean $\mu = 0$, variance $\sigma^2 = 1$
- Abbreviated $Z \sim N(0, 1)$
- pdf of Z:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty$$

- cdf of Z:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} f(u)\ du$$

  - Result can be found using standard normal table

7.4.1 z_α Notation

- Denotes value on the z axis for which α of the area under the z curve lies to the **RIGHT** of z_α
- $100(1 - \alpha)$th percentile of the standard normal distribution

7.5 Standardizing A Normal Distribution

- Normal RV: $X \sim N(\mu, \sigma^2)$
- Standard Normal RV: $Z = \dfrac{X - \mu}{\sigma}$
- Similarly,

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$

8 W4: Joint Probability Distribution

8.1 Joint Probability Mass Function

The joint probability mass function p(x, y) is defined for each pair of numbers (x, y) by

$$p(x, y) = P(X = x \text{ and } Y = y)$$

It must satisfy the following conditions:

1. $p(x, y) \geq 0$
2. $\sum_x \sum_y p(x, y) = 1$

The probability P[(X, Y) ∈ A] is obtained by summing the joint pmf over pairs in A:

$$P[(X, Y) \in A] = \sum_{(x,y)} \sum_{\in A} p(x, y)$$

8.2 Marginal Probability Mass Function

The marginal probability mass function of x, $p_X(x)$ is given by

$$p_X(x) = \sum_{y:p(x,y)>0} p(x, y) \text{ for each possible value of } x.$$

Similarly, the marginal probability mass function of y, $p_X(x)$ is given by

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x, y) \text{ for each possible value of } y.$$

- The word "marginal" indicates that the pmf is obtained from the joint probability distribution.
- We can obtain the marginal pmf from the joint pmf, however the reverse is not always true.

8.3 Joint Probability Density Function

The joint probability density function f(x, y) for two different RV is satisfies two conditions:

1. $f(x, y) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\ dx\ dy = 1$

For any two dimensional set A, where $a \leq x \leq b,\ c \leq y \leq d$,

$$\begin{aligned} P[(X, Y) \in A] &= \iint_A f(x, y)\ dx\ dy \\ &= \int_a^b \int_c^d f(x, y)\ dx\ dy \end{aligned}$$

- $P[(X, Y] \in A]$ is the volume beneath the surface above the region A

8.4 Marginal Probability Density Function

The marginal probability density function of X and Y, denoted by $f_X(x)$ and $f_Y(y)$ respectively, are

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\ dy, \quad -\infty < x < \infty$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\ dx, \quad -\infty < y < \infty$$

- Marginal pdf of X is the pdf of X
- The word "marginal" indicates that the pdf is obtained from the joint probability distribution.
- We can obtain the marginal pdf from the joint pdf, however the reverse is not always true.

8.5 Multiple Random Variables

If $X_1, X_2, \ldots, X_n$ are all discrete RVs, the joint pmf of the variables is

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

If $X_1, X_2, \ldots, X_n$ are all continuous RVs, the joint pdf of the variables with intervals $[a_1, b_1], \ldots, [a_n, b_n]$ is

$$P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \ldots, a_n \leq X_n \leq b_n)$$
$$= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(x_1, x_2, \ldots, x_n)\ dx_n \cdots dx_2\ dx_1$$

8.6 Independence of Random Variables

Two RVs X and Y are said to be independent if for **every pair** of x and y values:

$$p(x, y) = p_X(x) \cdot p_Y(y) \quad \text{for discrete RV}$$
$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \text{for continuous RV}$$

If the above is not satisfied for all (x, y), then X and Y are dependent.

9 W5: Conditional Distribution

9.1 Conditional Probability Mass Function

Let X and Y be two discrete RVs with pmf $p(x, y)$.

For any value x for which $p(x) > 0$, the conditional probability mass function of Y given that $X = x$ is

$$p_{Y|X}(y \mid x) = \frac{p(x, y)}{p_X(x)}$$

where $p_X(x)$ is the marginal pmf of X.

9.2 Conditional Probability Density Function

Let X and Y be two continuous RVs with pdf $f(x, y)$. For any value x for which $f(x) > 0$, the conditional probability density function of Y given that $X = x$ is

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

where $f_X(x)$ is the marginal pdf of X.

9.3 Conditional Distribution

- The summation of the conditional pmf or pdf over the entire sample space is 1.

$$\sum_y p_{Y|X}(y \mid x) = 1 \quad \text{for discrete RVs X and Y}$$
$$\int_{-\infty}^{\infty} f_{Y|X}(y \mid x) dy = 1 \quad \text{for continuous RVs X and Y}$$

9.4 Conditional Expectation

Let X and Y be jointly distributed RVs with pmf $p(x, y)$ or pdf $f(x, y)$. The expected value of a function $h(X, Y)$, denoted by $E[h(X, Y)]$ or $\mu_{h(X,Y)}$ is given by

$$E[(h(X, Y)] = \begin{cases} \displaystyle\sum_x \sum_y h(x, y)p(x, y) & \text{for discrete RVs X and Y} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)f(x, y) & \text{for continuous RVs X and Y} \end{cases}$$

9.5 Conditional Mean

Let X and Y be jointly distributed RVs with pmf $p(x, y)$ or pdf $f(x, y)$. The conditional mean of Y, given that $X = x$, denoted by $\mu_{Y|x}$ is given by

$$\mu_{Y|x} = E(Y \mid x) = \begin{cases} \displaystyle\sum_y yp(y \mid x) & \text{for discrete RVs X and Y} \\ \displaystyle\sum_y h(y)f(y \mid x)dy & \text{for continuous RVs X and Y} \end{cases}$$

9.6 Conditional Variance

Let X and Y be jointly distributed RVs with pmf $p(x, y)$ or pdf $f(x, y)$. The conditional mean of Y, given that $X = x$. denoted by $\sigma^2_{Y|x}$ is given by

$$\begin{aligned}
\sigma^2_{Y|x} &= E\{[Y - E(Y \mid x)]^2\} \\
&= E(Y^2 \mid x) - [E(Y \mid x)]^2
\end{aligned}$$

9.7 Law of Total Expectation

If X is a RV, and Y is a RV in the same probability space, then

$$E\left[E(X \mid Y)\right] = E(X)$$

i.e. expected value of the conditional expected value of X given Y is the = expected value of X

9.8 Covariance

The covariance between two variables X and Y, denoted by $\sigma_{X,Y}$ is given by

$$\begin{aligned}
\sigma_{X,Y} &= K(X, Y) = E[(X - \mu_x)(Y - \mu_y)] \\
&= \begin{cases} \displaystyle\sum_x \sum_y (x - \mu_x)(y - \mu_y)\ p(x, y) & \text{for discrete RVs X and Y} \\ \displaystyle\int_x \int_y (x - \mu_x)(y - \mu_y)\ f(x, y)\ dx\ dy & \text{for continuous RVs X and Y} \end{cases}
\end{aligned}$$

- Shortcut formula: $K(X, Y) = E(XY) - E(X)E(Y)$
- Value of covariance:
  - Positive $\sigma_{X,Y}$: positive linear relationship between X and Y
  - Near-zero $\sigma_{X,Y}$: no linear relationship between X and Y
  - Negative $\sigma_{X,Y}$: negative linear relationship between X and Y

9.9 Correlation

- Correlation coefficient $\rho_{X,Y}$: measure of degree of linear relationship between two RVs X and Y

$$\rho_{X,Y} = \widetilde{K}(X,Y) = \frac{K(X,Y)}{\sigma_X \sigma_Y}$$

  - It is always true that $-1 \leq \rho_{X,Y} \leq 1$

- If X and Y, then $\rho_{X,Y} = 0$
  - **BUT** $\rho_{X,Y}$ does not imply independence between X and Y

- Measure of linear relationship:
  - $|\rho| = 1$: Strong linear relationship between X and Y
  - $|\rho| \neq 1$: Not completely linear relationship between X and Y; could be strong non-linear relationship
  - $\rho = 0$: X and Y are uncorrelated

10 W5: Central Limit Theorem

10.1 Linear Combination of One RV

For a linear combination of one RV X, denoted by $aX + b$, the mean and variance are as follows:

- Mean, $E(aX + b) = aE(X) + b$
- Variance, $V(aX + b) = a^2 E(X)$

10.2 Linear Combination of Two RVs

For a linear combination of two RVs X and Y, where $W = aX + bY$,
the mean and variance are as follows:

	X, Y independent	**X, Y dependent**
Mean, $E(W)$	$aE(X) + bE(Y)$	
Variance, $V(W)$	$a^2V(X) + b^2V(Y)$	$a^2V(X) + b^2V(Y) + 2abK(X,Y)$

10.3 Linear Combination of Multiple RVs

For a linear combination of multiple RVs $X_1, X_2, \ldots, X_n$, where $W = \sum_{i=1}^{n} a_i x_i$, the mean and variance are as follows:

	RVs independent	**RVs dependent**
Mean, $E(W)$	$\sum_{i=1}^{n} a_i E(X_i)$	
Variance, $V(W)$	$\sum_{i=1}^{n} a_i^2 V(X_i)$	$\sum_{i=1}^{n} a_i^2 V(X_i) + 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} a_i a_j K(X_i, X_j)$

10.4 Linear Combination of Independent and Identically Distributed RVs

For a linear combination of independent and identically distributed (iid) RVs $X_1, X_2, \ldots, X_n$ where $W = \sum_{i=1}^{n} X_i$ with mean μ and variance σ^2, the mean and variance are as follows:

- Mean, $E(W) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu = n\mu$
- Variance, $V(W) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2$

10.5 Linear Combination of Normal RVs

For two normal RVs X and Y, where $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, the linear combination $W = X + Y$ is also a normal RV with mean $\mu_X + \mu_Y$ and variance $\sigma_X^2 + \sigma_Y^2$, i.e.

$$W \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

10.6 Sample Mean

Let $X_1, X2, \ldots, X_n$ be iid RVs with mean μ and variance σ^2.
The sample mean $\overline{X}$ can be calculated using the formula $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

The mean and variance of $\overline{X}$ is as follows:

- Mean, $E(\overline{X}) = \mu$
- Variance, $V(\overline{X}) = \frac{\sigma^2}{n}$

10.7 Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. The sample mean $\overline{X}$ can be calculated using the formula $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.

For a sufficiently large n, i.e. $\mathbf{n \leq 30}$, $\overline{X}$ has approximately a normal distribution with mean $E(\overline{X})$ and variance $V(\overline{X})$ as follows:

- Mean, $E(\overline{X}) = \mu$
- Variance, $V(\overline{X}) = \dfrac{\sigma^2}{n}$

If the distribution is close to a normal pdf, a small n yields a good approximation to a normal distribution.