STAT1201 – Summer Semester 2022

Lecture 3 - Randomness and Probability Theory

Dr. Wasanthi Thenuwara

15/03/2022

1

Lecture 3 – Randomness and Probability Theory

In this lecture, you will practice

- Difference between parameters and statistics
- Randomness and Probability
- Conditional probability
- Difference between discrete and continuous random variables
- Discrete and continuous probability distributions
- Expected value and standard deviations of discrete probability distributions

2

Population Parameters and Sample Statistics

- In lectures 1 and 2, we mainly focused on descriptive statistics.
- In inferential statistics, we draw conclusions about a population by examining a representative sample, that is taken from the respective population.
- A population is a complete set of individuals or objects that we want information about. For example, the Australia Census 2021 collected information from all the people living in Australia.
- A sample is a subset of population. For example, information of people living in Brisbane city council.
- Samples should be selected so that it is representative of the population, and it is not biased in any way.
- In STAT1201, we mainly focus on selecting random samples for scientific experiments.

3

Population Parameters and Sample Statistics

- A parameter is a numerical (summary) measure that describes a population characteristic.
- A statistic is a numerical (summary) measure that describes a sample characteristic.

Population Parameters	*Sample Statistics*
Population Size - N	Sample Size - n
Population Mean - μ	Sample Mean - $\overline{x}$
Population Variance - σ^2	Sample Variance - s^2
Population SD - σ	Sample SD - s
Population Proportion - p	Sample Proportion - $\hat{p}$

4

Population Parameters and Sample Statistics

Poll Question 1

The mean age of STAT1201 – Summer 2022 students is 18 years. The mean age of 20 randomly selected STAT1201 – Summer 2022 students - 18.3 years. The values of population mean (μ) and sample mean ($\overline{x}$) are

1. 18 and 18.3 respectively
2. 18.3 and 18 respectively
3. 18 and 18 respectively
4. 18.3 and 18.3 respectively

5

Population Parameters and Sample Statistics

Why μ = 18 years is different from $\bar{x} = 18.3$ years?

Sampling error.

- sampling error is an unavoidable consequence of being able to observe only a subset of the elements in the population.
- sampling errors can be reduced by increasing the sample size, and sometimes by using a different sampling selection approach.

6

Randomness and Probability

What is randomness in Statistics?

Describes a phenomenon in which the outcome of a single repetition is unpredictable in advance. However, there is a predictable long-term pattern that can be described by the distribution of the outcome of a large number of repetitions.

For example, consider tossing a coin. From the outcome of a previous toss, can you predict the outcome of the next tossing with certainty?

Randomness in samples of data

Random sampling is a sampling technique that does give every item in the population an equal chance of being selected.

Example - Suppose that you need to select 20 students for an experiment from STAT1201 – Summer 2022 student population. How do you do this?

7

Randomness and Probability

Poll Question 2

Suppose I select 20 students for an experiment from STAT1201 – Summer 2022 student population. Assuming 160 students have been enrolled for the course, what is the chance of you being selected for the sample?

1. 0
2. 1
3. 0.125
4. 20

8

Randomness and Probability

What is probability?

- Probability is how likely that a particular event will happen.
- Probabilities to outcomes can be assigned in three ways.
  - Subjective probability (reflects an individual's belief)
  - Calculated or theoretical probability (based on prior knowledge. e.g. if a six-sided dice is rolled, the chance of getting a 4 is 1/6)
  - Empirical probability (outcome is based on observed data).
- Probability must be between 0 and 1 (i.e. $0 \leq p \leq 1$)
- The probabilities of all possible outcomes associated with a particular random phenomenon must add up to 1 (i.e. $\sum_{i=1}^{n} p_i = 1$).

9

Randomness and Probability

Poll Question 3

Which of the following is a valid probability model for tossing a coin?

1. P(H)=0.25, P(T)=0.25
2. P(H)=0.5, P(T)=0.5
3. P(H)=0.25, P(T)=0.85
4. P(H)=0.5, P(T)=0.25

10

Key Probability concepts

Sample space (Ω) - set of all possible outcomes that might be observed in a random process.

Event (A) - A subset of sample space. An event occurs one of the outcomes in it occurs.

Example: Suppose you toss a coin three times and define A as the event of seen only two heads. What is Ω? What is A?

Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
A = {HHT, HTH, THH}
The probability of an event A = P(A)
P(A) = $\frac{N(A)}{N(\Omega)}$

In the previous example, P(A) = 3/8

P(Ω) = 1

11

Key Probability concepts

Poll Question 4

Suppose you toss a coin three times. What is the probability of seen three heads?

1. 0.5
2. 0.25
3. 0.75
4. 0.125

12

Key Probability concepts

The complement $(\bar{A})$ of an event A is the set of all outcomes in Ω not in A.

$P(\bar{A}) = 1 - P(A)$

The union of two events A and B $(A \cup B)$ is the set of all outcomes in A, or in B, or in both.

$P(A \cup B) = P(A)+P(B)-P(A \cap B)$

The intersection of two events A and B $(A \cap B)$ is the set of outcomes in both A and B.

If the two events, A and B are disjoint, then

$P(A \cup B) = P(A) + P(B)$

13

Conditional Probability - P(A|B)

Probability of event A occurring if B has already occurred.

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

Example: Consider the following contingency table for the survey data.

	Arcadia	Colmar	Hofn	Sum
Female	9	9	8	26
Male	9	16	9	34
Sum	18	25	17	60

What is the probability of a randomly selected female is living in Arcadia?

14

Conditional Probability - P(A|B)

	Arcadia	Colmar	Hofn	Sum
Female	9	9	8	26
Male	9	16	9	34
Sum	18	25	17	60

To use the conditional probability formula, define events as follows.

A = Living in Arcadia

B = Being a female

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

P(A|B) = $\frac{9/60}{26/60}$

P(A|B) = $\frac{9}{26}$

P(A|B) = 0.3462

15

Independent Events and Conditional Probability

- The two events are independent, if one event occurs, it does NOT affect the probability of a different event occurring.
- Only if A and B are independent events the probability of A occurring, given B has already occurred, be the same as just the probability of A.

P(A|B) = P(A)

Similarly, P(B|A) = P(B)

- if A and B are independent events;

  P($A \cap B$) = P(A)*P(B)

16

Independent Events and Conditional Probability

Poll Question 5

A six sided dice is rolled. What is the probability that the number rolled is a 3, if a head is tossed on a coin.

1. 1/6
2. 1/12
3. 4/6
4. 3/6

17

Conditional Probability Exercise

Discuss with your friends and find the answer. Share your answers in Ed.

A certain type of disease is present in 10% of the population. A test for the diagnosis of this disease is not perfect. The previous diagnostic test results revealed that a 2% rate of false positive outcomes and 4% rate of false negative outcomes. If a randomly selected person in the population tests positive, what is the probability she has the disease?

18

Random Variables

A random variable is a random process with numerical outcomes.

Examples:

- Number of text messages students receive during this lecture hour. The possible outcomes are 0, 1, 2, …, n.
- Time to complete the STAT1201 exam. This can take any hours between 0 and 2 (0.25hrs, 1.38hrs, 1.95hrs, …)

We will focus on discrete random variables and continuous random variables.

19

Random Variables

Discrete Random Variable

A random variable that has a countable number of possible values.

Examples: Number of children in a family; Number of left-handed students in STAT1201 class; Outcome of rolling a 6-sided dice.

Discrete random variables are usually generated from experiments in which things are 'counted', not 'measured'.

Continuous Random Variable

A random variable where the data can take infinitely many values.

Examples: Height of the STAT1201 students; Blood haemoglobin level.
Continuous random variables are usually generated from experiments in which things are 'measured', not 'counted'.

20

Discrete and Continuous Probability Distributions

Discrete Probability Distribution

The listing of all possible values of a discrete random variable X along with their associated probabilities.

Example: Define X = Number shown by rolling a six-sided dice (X = 1, 2, 3, 4, 5, 6). Then the probability distribution of X can be written as follows.

X	P(X=x)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

21

Discrete and Continuous Probability Distributions

Discrete Probability Distribution

Example: The following table shows the probability distribution of the number of children (X) in a family and the associated probabilities from a random sample of families living in Brisbane.

X	P(X=x)
0	0.21
1	0.45
2	0.23
3	0.11

What is the probability that no more than two children in a family?

P($X \leq 2$) = P(X=0) + P(X=1) + P(X=2)

P($X \leq 2$) = 0.89

22

Discrete Probability Distribution

X	P(X=x)
0	0.21
1	0.45
2	0.23
3	0.11

Poll Question 6

What is the probability that at least one child in a family?

1. 0.21
2. 0.45
3. 0.79
4. 0.11

23

Discrete and Continuous Probability Distributions

The most popular discrete probability distributions are: Binomial Distribution and Poisson Distribution

The most popular continuous probability distributions are:

Uniform Distribution; Normal Distribution; Exponential Distribution; t-Distribution; Chi-square Distribution and F-Distribution.

24

Expected value (Mean) and Variance of a Discrete Probability Distribution

Expected value or Mean (E(X) or μ)
Long run average of a random variable. If we repeat taking random samples of families living in Brisbane, the mean or expected number of children can be found as follows.

E(X) = μ = $\sum x.P(X = x)$

Using the children's distribution

E(X) = μ = 0x0.21 + 1x0.45 + 2x0.23 + 3x0.11 = 1.24

Variance (Var(X))
We can quantify the variability of a discrete random variable using squared deviations about the mean as we did for a sample of data.
Var(X) = $\sum P(X = x)(x - \mu)^2$
SD(X) = $\sqrt{Var(X)}$

25

Discrete Probability Distributions - Expected Value

Poll Question 7

Suppose that you are playing a game with your friend. The probability of winning 5 dollars is 0.4 and loosing 5 dollars is 0.6. If you play the game 10 times, how much you would expect to win?

1. 2 dollars
2. 20 dollars
3. -1 dollars
4. -10 dollars

26

Discrete and Continuous Probability Distributions

Continuous Probability Distribution f(x)

Continuous probability distribution functions cannot be presented in a table or histogram like we did for discrete probability distributions as there are uncountable number of possible outcomes.
The probability of any individual outcome is zero (0).
P(X=x) = 0

We always calculate the probability for a range of the continuous random variable, X.
P(X > a) ; P($a \leq X \leq b$) ; P($X \leq b$)

We can use the concept of integral to calculate these probabilities. We will discuss probability calculations for continuous probability distributions in Lectures 4 and 5.
E(X) = μ = $\int_{-\infty}^{\infty} f\ (x)x\ dx$
Var(X) = $\int_{-\infty}^{\infty} f\ (x)(x-\mu)^2\ dx$

27

Expected value and Variance of combined variables

Some important rules to know

Suppose X is a random variable.
Let Y = aX where a is a constant.
Then E(Y) = aE(X) and Var(Y) = a^2Var(X)

Let Y = aX + b where a and b are constants
Then E(Y) = aE(X) + b ; Var(Y) = a^2Var(X) ; SD(Y) = aSD(X)

Suppose X_1 and X_2 are two independent random variables.
Let Y = X_1 + X_2
Then E(Y) = E(X_1) + E(X_2) and Var(Y) = Var(X_1) + Var(X_2)

Let Y = X_1 - X_2
Then E(Y) = E(X_1) - E(X_2) and Var(Y) = Var(X_1) + Var(X_2)

28

Expected value and Variance of combined variables

Example

The length of lizards living in one island in Australia has a expected length of 50cm and a standard deviation of 8cm. Suppose a random sample of 9 lizards lengths are taken. What is the expected value and the standard deviation of total length of 9 lizards?

29

Next …

Reminders

Quizzes 2 and 3 are now open.

Lecture 4 – The Probability Distributions and Sampling Distributions

Thursday, 8 Dec 2022 at 12:00 via Zoom (818 1453 7986)

30