Applied Statistics for Data Science

LECTURE 1: REVIEW OF PROBABILITY: AXIOMS, RULES, AND APPROXIMATION

1

الاحصاء التطبيقي : هو الاساس لعلم البيانات من خلال طرق احصائيه
وطرق تحليليه و التركيز على البيانات في الحياه اليوميه بهدف لايجاد
الحلول لمشاكل

Overview:

Applied statistics is a foundation upon which data science has been built. Through statistical methods, analysis, and an emphasis on real-world data, applied statisticians seek concrete solutions to tangible problems.

الاحتمالية

❖ What is Probability?

مقياس لامكانيه حدوث الحدث

- ***Probability*** is a measure of the likelihood of an event to occur. Many events cannot be predicted with total certainty.

ولايمكن التنبؤ بدقه لاحتماليه حدوث حدث

- Using probability, one can predict only the chance of an event to occur, i.e., how likely they are going to happen.

P

P : تحسب فرصه حدوث الحدث

- ***For example,*** when a coin is tossed, there is a probability to get heads or tails.

H T

- ***Probability can be reduced to three axioms.***

مسلمات
بديهيات

2

Overview:

مسلمة

❖ What do we mean by an axiom?

شيء مثبت ذاتيا (لا يحتاج اثبات)

- An axiom is typically something that is mathematically self-evident.
- *More precisely,* an axiom is a statement which we have assumed to be true. That is, there is no proving an axiom.

جمله تعتبر صحيحه وليس لها اثبات

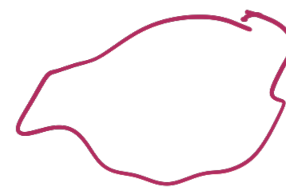
- **In order to understand the axioms for probability,** Lets define the following :

مجموعة النواتج — فضاء العينة — (H) (T) — S

- A set of outcomes called the sample space S.

مجموعة جزئية من S — الحدث

- The sample space is comprised of subsets called events $E_1, E_2, \ldots, E_n$.
- Assume that there is a way of assigning a probability to any event E.
- The probability of the event E is denoted by P(E).



$P(E_1)$ احتماليه حدوث الحادث E_1

3

Overview:

مسلمات الاحتمال

P(blue)=0

❖ Axioms of Probability:

Axiom 1:

دائماً الاحتمال موجب

- For any event E , P(E) ≥ 0
- → probability cannot be negative.

$P(E)=0 \Longleftarrow$ حادث مستحيل

- → The smallest value for P(E) is zero if P(E)=0 , then, the event E may considered impossible

Axiom 2:

احتمال حدوث فضاء العينة = ١

- Probability of the sample space S is P(S)=1 = 100%
- → i.e., 100 percent
- → S contains all possible outcoms , thus, the outcome of each trial always belongs to S which means the event S always occurs P(S)=1.

اي محاولة في التجربه سوف تكون النواتج ضمن S

- *For example:* when rolling the dice S={1,2,3,4,5,6} , and since the outcome is always among the numbers 1 through 6 , P(S)=1.

$P(1) = \frac{1}{6}$ $\quad P(4) = \frac{1}{6}$

4

$P(2) = \frac{1}{6}$ $\quad P(5) = \frac{1}{6}$ $\quad \frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}$

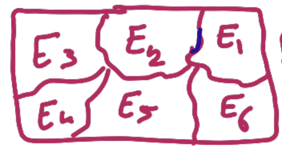
$P(3) = \frac{1}{6}$ $\quad P(6) = \frac{1}{6}$ $\quad P(S) = \frac{6}{6} = 1$

disjoint = mutually exclusive : الحوادث المتنافيه
حوادث لا يمكن ان تحدث سويا
حدوثهما معاً $= A \cap B = 0$

Overview:

Axiom 3:

$P(E_1)+P(E_2)+P(E_3)\cdots = P(E_1 \cup E_2 \cup E_3 \cdots)$



$P(S)=1$

$E_1 \cup E_2$

- If E_1, E_2, E_3, … are disjoint events , then
- → $P(E_1 \cup E_2 \cup E_3 \cup \cdots) = P(E_1)+P(E_2)+P(E_3)+\ldots$
- → The basic idea of axiom 3 if some events are disjoint (i.e., there is no overlap between them), then the probability of their union must be the summations of their probabilities.
- *For example:* In a presidential election, there are four candidates. Lets call them A, B, C, and D. Based on the polling analysis, you estimate that A has a 20 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that A **or** B win the election?

20% 40%

مجموع الحوادث المتنافيه = اتحاد جميع هذه الحوادث U

Solution:

- The events {A wins} , {B wins} , {C wins} , and {D wins} are disjoint events.
- From axiom 3 : $P(A\ wins\ or\ B\ wins) = P(\{A\ wins\} \cup \{B\ wins\})$
- $= P(\{A\ wins\}) \cup P(\{B\ wins\})$
- $= 0.2+0.4$
- $=0.6$

or U او
+

5

$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$

$0.2 + 0.4 = 0.6$

Overview:

❖ **Rules of Probability:** قواعد الاحتمال

Definitions:

•Two events are **mutually exclusive** or **disjoint** if they cannot occur at the same time.

•The probability that Event A occurs, given that Event B has occurred, is called a **conditional probability**. The conditional probability of Event A, given Event B, is denoted by the symbol P(A|B).

•The **complement** of an event is the event not occurring. The probability that Event A will not occur is denoted by P(A').

•The probability that Events A and B *both* occur is the probability of the **intersection** of A and B. The probability of the intersection of Events A and B is denoted by P(A ∩ B). If Events A and B are mutually exclusive, P(A ∩ B) = 0.

•The probability that Events A or B occur is the probability of the **union** of A and B. The probability of the union of Events A and B is denoted by P(A ∪ B) .

•If the occurrence of Event A changes the probability of Event B, then Events A and B are **dependent**. On the other hand, if the occurrence of Event A does not change the probability of Event B, then Events A and B are **independent**.

6

Overview:

Axiom 3:

- If $E_1, E_2, E_3, \ldots$ are disjoint events , then
- $\rightarrow$ P($E_1 \cup E_2 \cup E_3 \cup \cdots$)=P($E_1$)+P($E_2$)+P($E_3$)+…
- $\rightarrow$ The basic idea of axiom 3 if some events are disjoint (i.e., there is no overlap between them), then the probability of their union must be the summations of their probabilities.
- *For example:* In a presidential election, there are four candidates. Lets call them **A, B, C, and D**. Based on the polling analysis, you estimate that A has a 20 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that A **or** B win the election?

Solution:

- The events {A wins} , {B wins} , {C wins} , and {D wins} are disjoint events.
- From axiom 3 : $P(A\ wins\ or\ B\ wins) = P(\{A\ wins\} \cup \{B\ wins\})$
- $= P(\{A\ wins\}) \cup P(\{B\ wins\})$
- = 0.2+0.4
- =0.6

5

Overview:

قواعد الاحتمال

❖Rules of Probability:

Definitions: تعريفات

الحوادث المتنافية

•Two events are **mutually exclusive** or **disjoint** if they cannot occur at the same time. الحوادث المتنافية: هي الحوادث التي لا يمكن ان تحدث سويًا

•The probability that Event A occurs, $\boxed{\text{given}}$ that Event B has occurred, is called a **conditional probability**. The conditional probability of Event A, given Event B, is denoted by the symbol P(A|B).

الاحتمال الشرطي

P(A/B الاحتمال الشرطي :
احتمال حدوث A بشرط B (given)

المتممة

•The **complement** of an event is the event not occurring. The probability that Event A will not occur is denoted by P(A'). المتممة : احتمال عدم حدوث الحادث

لن يحدث

P(A')

6

$P(\bar{A})$
$P(A^c)$

Overview:

احتمال حدوث الحادثين سوياً (سميا تقاطع)

❖Rules of Probability:

(Both) (and) $P(A \cap B)$

Definitions:

في حالة كان الحادثين متنافيين $P(A \cap B) = 0$

•The probability that Events A and B *both* occur is the probability of the **intersection** (تقاطع) of A and B. The probability of the intersection of Events A and B is denoted by $P(A \cap B)$. If Events A and B are mutually exclusive, $P(A \cap B) = 0$.

•The probability that Events A or B occur is the probability of the **union** (اتحاد) of A and B. The probability of the union of Events A and B is denoted by $P(A \cup B)$. احتمال حدوث الحادث A أو B هو احتمال اتحاد الحادثين

$P(A \cup B)$

7

Overview:

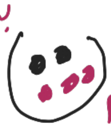
❖Rules of Probability:

Definitions:

•If the occurrence of Event A changes the probability of Event B, then Events A and B are **dependent** (غير مستقل). On the other hand, if the occurrence of Event A does not change the probability of Event B, then Events A and B are **independent** (مستقله).

اذا كان حدوث الحادث A يؤثر في حدوث الحادث B تسمى الحوادث (غير مستقله)

بدون ارجاع



$P(red) = \frac{3}{5}$ $P(Black) = \frac{2}{4}$

اذا كان حدوث الحادث A لا يؤثر على حدوث B تسمى الحوادث مستقله (مع الارجاع)

$P(red) = \frac{3}{5}$ $P(Black) = \frac{2}{5}$

8

Overview:

40%   60%
A' { A
1

❖ Rules of Probability:

- **Rule of Subtraction:** قاعده الطرح

The probability that event A will occur is equal to 1 minus the probability that event A will not occur.

$$P(A) = 1 - P(A')$$

Example: احتماليه حدوث الحادث A هو (1 - احتماليه عدم حدوثه)

Suppose, for example, the probability that Bill will graduate from college is 0.80. What is the probability that Bill will not graduate from college?

0.8 ≈ 80%   20%

Based on the rule of subtraction,

- The probability that Bill will graduate is P(A)
- The probability that Bill will not graduate is P(A')
- → 0.80 = 1- P(A') → P(A') = 1- 0.80 = 0.20 .

$P(A') = 1 - P(A)$
$= 1 - 0.80$
$= 0.20 = 20\%$

9

سحب مع الارجاع :. ما احتمال سحب اول كره حمراء والثانيه سوداء $= P(r,B) = P(r) \cdot P(B) = \frac{3}{5} \cdot \frac{2}{5} = \frac{6}{25}$

سحب بدون ارجاع :. احتمال سحب الكره الاولى • احتمال سحب الثانيه سوداء بشرط الاولى حمراء $\frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20}$

Continue ... قاعده الضرب

- **Rule of Multiplication:**

احتماليه حدوث الحادثين سويا = ساوي ضربهم وناخذ بعين الاعتبار الحادث الاول على انه

The probability that Events A and B both occur is equal to the probability that Event A occurs times the probability that Event B occurs, given that A has occurred.

$$P(A \cap B) = P(A)\, P(B|A)$$

P(B)

مستقله
$\boxed{P(A \cap B) = P(A) \cdot P(B)}$

غير مستقله
$\boxed{P(A \cap B) = P(A) \cdot P(B/A)}$

Example:

An urn contains 6 red marbles and 4 black marbles. Two marbles are drawn without replacement from the urn. What is the probability that both of the marbles are black? بدون ارجاع

مستقله $P(B/A) = P(B)$

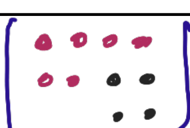
Let's state that : A = the event that the first marble is black;

and let B = the event that the second marble is black.

A سحب الاولى اسود
B سحب الثانيه اسود

10

اذا كان السؤال مع الارجاع (حوادث مستقله)



$P(A \cap B) = P(1Black) \cdot P(2Black) = \frac{4}{10} \cdot \frac{4}{10} = \frac{16}{100}$

اذا كان السؤال بدون ارجاع اذاً الحوادث غير مستقلة

$P(A \cap B) = P(A) \cdot P(B/A) = \frac{4}{10} \cdot \frac{3}{9} = \frac{12}{90} = 0.133$

13.3%

Continue …

- **Rule of Multiplication:**

Solution Continue:

From the question we know the following:

•In the beginning, there are 10 marbles in the urn, 4 of which are black. **Therefore, P(A) = 4/10.**

•After the first selection, there are 9 marbles in the urn, 3 of which are black. **Therefore, P(B|A) = 3/9.**

Therefore, based on the rule of multiplication:

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

$$P(A \cap B) = (4/10) * (3/9) = 12/90 = 2/15 = 0.133$$

Continue …

قاعدة الجمع

- **Rule of Addition:**

∪ +

The probability that Event A or Event B occurs is equal to the probability that Event A occurs plus the probability that Event B occurs minus the probability that both Events A and B occur.

X

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

Note: Invoking the fact that P(A ∩ B) = P(A)P(B | A), the Addition Rule can also be expressed as:

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A)P(B | A)}$$

احتمالية حدوث الحادث A أو B ← جمع

اذا كانت الحوادث متنافية (A ∪ B) فقط الجمع

اذا كانت الحوادث غير متنافية $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B) = 20 + 30 - 5$

20 غير متنافية CS M 30 5

Continue …

- **Rule of Addition:**

Example:

A student goes to the library. The probability that she checks out (a) a work of fiction is 0.40, (b) a work of non-fiction is 0.30, and (c) both fiction and non-fiction is 0.20. What is the probability that the student checks out a work of fiction, non-fiction, or both?

$P(F) = 0.4 \qquad P(N) = 0.3$

$$P(F \cup N) = P(F) + P(N) - P(F \cap N)$$
$$= 0.4 + 0.3 - 0.2 = 0.5$$

13

Continue …

- **Rule of Addition:**

Solution:

- **Let F** = the event that the student checks out fiction;
- **and let N =** the event that the student checks out non-fiction.
- *Then, based on the rule of addition:*

$$P(F \cup N) = P(F) + P(N) - P(F \cap N)$$

$$P(F \cup N) = 0.40 + 0.30 - 0.20 = 0.50$$

14