Statistics for Data Science -1

Lecture 7.1: Conditional Probability: Contingency tables

Usha Mohan

Indian Institute of Technology Madras

Learning objectives

1. Understand notion of conditional probability, i.e find the probability of an event given another event has occurred.

Learning objectives

1. Understand notion of conditional probability, i.e find the probability of an event given another event has occurred.
2. Distinguish between independent and dependent events.

Learning objectives

1. Understand notion of conditional probability, i.e find the probability of an event given another event has occurred.
2. Distinguish between independent and dependent events.
3. Solve applications of probability.

Contingency tables: Joint, Marginal, and Conditional probabilities

From tables to probability

- Recall the cell phone usage versus gender example when we discussed about association between categorical variables and the concept of relative frequencies.
- Percentages computed within rows or columns of a contingency table correspond to conditional probabilities
- Convert contingency tables into probabilities, we use the counts to define probabilities.

Relative frequency

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	10	34	**44**
Male	14	42	**56**
Column total	**24**	**76**	**100**

Relative frequency

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	10	34	**44**
Male	14	42	**56**
Column total	**24**	**76**	**100**

Divide each count by 100

Relative frequency

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	10	34	**44**
Male	14	42	**56**
Column total	**24**	**76**	**100**

Divide each count by 100

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	10/100	34/100	**44/100**
Male	14/100	42/100	**56/100**
Column total	**24/100**	**76/100**	**100**

Joint probabilities

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	0.10	0.34	**0.44**
Male	0.14	0.42	**0.56**
Column total	**0.24**	**0.76**	**100**

- Displayed in cells of a contingency table
- Represent the probability of an intersection of two or more events
- In the example: there are four joint probabilities; e.g.,
  - P(Female **and** Not owning a smartphone) $= 0.10$
  - P(Male and Owning a smartphone) $= 0.42$

Marginal probability

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	0.10	0.34	**0.44**
Male	0.14	0.42	**0.56**
Column total	**0.24**	**0.76**	**100**

- Displayed in the margins of a contingency table
- Is the probability of observing an outcome with a single attribute, regardless of its other attributes
- In the example: There are four marginal probabilities, e.g.,
  - $P(\text{Female}) = 0.10 + 0.34 = 0.44$
  - $P(\text{Owning a smartphone}) = 0.34 + 0.42 = 0.76$

Conditional probability

Conditional probability

- Find conditional probabilities to answer questions like
  - "among Female buyers, what is the chance a someone owns

Conditional probability

- Find conditional probabilities to answer questions like
  - "among Female buyers, what is the chance a someone owns a phone?"

Conditional probability

- Find conditional probabilities to answer questions like
  - "among Female buyers, what is the chance a someone owns a phone?"
  - "Among people who don't own a phone, how many are male?"

Conditional probability

- Find conditional probabilities to answer questions like
  - "among Female buyers, what is the chance a someone owns a phone?"
  - "Among people who don't own a phone, how many are male?"
- Recognize the answers

Conditional probability

- Find conditional probabilities to answer questions like
  - "among Female buyers, what is the chance a someone owns a phone?"
  - "Among people who don't own a phone, how many are male?"
- Recognize the answers
  - "among Female buyers, what is the chance a someone owns a phone?"- row relative frequency

Conditional probability

- Find conditional probabilities to answer questions like
  - "among Female buyers, what is the chance a someone owns a phone?"
  - "Among people who don't own a phone, how many are male?"
- Recognize the answers
  - "among Female buyers, what is the chance a someone owns a phone?"- row relative frequency
  - "Among people who don't own a phone, how many are male?"- column relative frequency

Conditional probability

We restrict the sample space to a row or column.

Conditional probability

We restrict the sample space to a row or column.

Conditional probability

We restrict the sample space to a row or column.

- "among Female buyers, what is the chance a someone owns a phone?"- Restrict sample space to only "Females"- First row

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	10/44	34/44	**44**
Male	14/56	42/56	**56**
Column total	**24/100**	**76/100**	**100**

Conditional probability

We restrict the sample space to a row or column.

- "among Female buyers, what is the chance a someone owns a phone?"- Restrict sample space to only "Females"- First row

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	10/44	34/44	**44**
Male	14/56	42/56	**56**
Column total	**24/100**	**76/100**	**100**

$$P(\text{Doesn't own a phone}|\text{Female}) = \frac{10}{44} = \frac{P(\text{Female} \cap \text{Doesn't own a phone})}{P(\text{Female})}$$

Conditional probability

Conditional probability

We restrict the sample space to a row or column.

Conditional probability

We restrict the sample space to a row or column.

- "Among people who don't own a phone, how many are male?"- Restricting sample space to only people who "don't own a phone"- First column

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	10/24	34/76	**44/100**
Male	14/24	42/76	**56/100**
Column total	**24**	**76**	**100**

Conditional probability

We restrict the sample space to a row or column.

- "Among people who don't own a phone, how many are male?"- Restricting sample space to only people who "don't own a phone"- First column

	Own a smartphone		
Gender	**No**	**Yes**	**Row total**
Female	10/24	34/76	**44/100**
Male	14/24	42/76	**56/100**
Column total	**24**	**76**	**100**

$$P(\text{Female}|\text{Doesn't own a phone}) = \frac{10}{24} = \frac{P(\text{Female} \cap \text{Doesn't own a phone})}{P(\text{Doesn't own a phone})}$$

Section summary

- Revisited contingency tables and introduced notions of
  1. Joint probability
  2. Marginal probability
  3. Conditional probability