Probability and Statistics review (part 1)

Francisco Pereira

Filipe Rodrigues

DTU Management Engineering
Department of Management Engineering

Outline

- Introduction
- Random variable, atom, and event
- Joint distribution
- Conditional probability
- Bayes theorem
- Independence
- Expectation

(Based on David MacKay, David Blei,
https://www.cs.princeton.edu/courses/archive/spring12/cos424/pdf/lecture02.pdf)

Introduction

Consider the “card problem"

- There are three cards:
  - Red/Red
  - Red/Blue
  - Blue/Blue
- I go through the following process
  1. Close my eyes and pick a card
  2. Pick a side at random
  3. Show you that side
- I show you Red. What’s the probability the other side is Red too?

Random variable, atom, and event

- In Algebra a variable, x, is an unknown value

  E.g. $2x = 4$
  It can take at most one value at a time.

- A random variable represents simultaneously a set of values
- Necessary in contexts where we *cannot* determine a unique value
  - Of course, theoretically, it also corresponds to one value...
  - But we can only determine its distribution
    $P(5 < X < 10) = 0.5$
- It can be a scalar (x), a vector ($\mathbf{x}$), a matrix ($\mathbf{X}$)...

Random variable, atom, and event

- Random variables take on values in a sample space. They can be discrete or continuous.
- For example:
  - Coin flip: {H, T}
  - Height: Positive values $(0, \infty)$
  - Temperature: real values $(-\infty, \infty)$
  - Number of words in a document: Positive integers $\{1, 2, ..., \infty\}$
- We call the values of random variables *atoms*.

Random variable, atom, and event

- A *discrete probability distribution* assigns probability to every atom in the sample space.
- For example, if X is an (unfair) coin, then

  $$P(X == H) = 0.7$$
  $$P(X == T) = 0.3$$

- The sum of probabilities of *any* distribution is 1

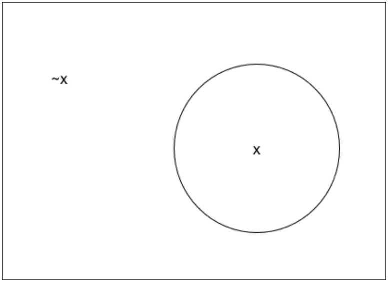
$$\sum_x P(X == x) = 1$$

- And all probabilities have to be greater or equal to 0
- Probabilities of disjunctions are sums over part of the space. E.g., the probability that a die is bigger than 3:

$$P(X > 3) = P(X == 4) + P(X == 5) + P(X == 6)$$

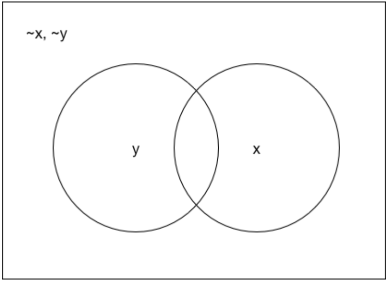
Random variable, atom, and event

- The figure below is helpful to understand these concepts well



- An *atom* is a point in the box. All atoms together form the *sample space*
- An *event* is a subset of atoms. Two events in the picture are x and $\sim x$
- The probability of an event is the sum of the probabilities of its atoms

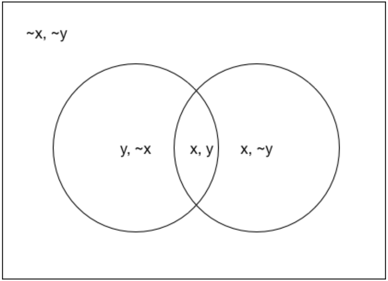
- In practice, we need to consider many variables at the same time
- An event would then combine atoms from multiple variables



- The joint distribution is a distribution over the configuration of all the random variables in the ensemble.

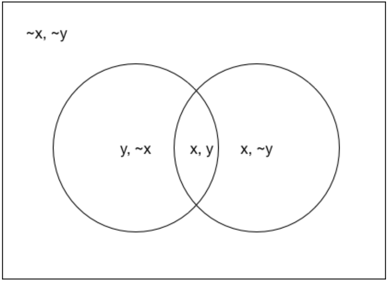
  For the figure, the function $P(X, Y)$ gives the probability of all possible combinations of X and Y
  Notice that $X \in \{x, \sim x\}$ and $Y \in \{y, \sim y\}$
  - Therefore $X, Y \in \{(x, y), (x, \sim y), (\sim x, y), (\sim x, \sim y)\}$

Joint distribution



- Some useful properties:
  - Union: $P(X \cup Y) = P(X) + P(Y) - P(X, Y)$
  - Marginalization: $P(X) = \sum_Y P(X, Y)$

Joint distribution



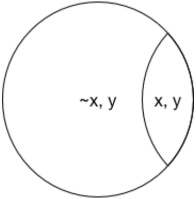
- Some useful properties:
  - Union: $P(X \cup Y) = P(X) + P(Y) - P(X, Y)$
  - **Marginalization**: $P(X) = \sum_Y P(X, Y)$

Joint distribution

- We can make a joint distribution of *two consecutive events*
- With the cards example:
  - X=first draw $\in$ {Red/Blue, Red/Red, Blue/Blue}
  - Y=second draw $\in$ {Red/Blue, Red/Red, Blue/Blue}
- X, Y $\in$ {(Red/Blue, Red/Blue), (Red/Blue, Red/Red), (Red/Blue, Blue/Blue), (Red/Blue, Red/Red), (Red/Red, Red/Red), (Red/Red, Blue/Blue), (Red/Blue, Blue/Blue), (Blue/Blue, Blue/Blue)}
- How to calculate the join distribution, $P(X, Y)$?

Conditional probability

- What about when we have observed one event, but want to know the probability of another one?
- The conditional probability of X given Y is the probability of event X when event Y is known



- So, we only concentrate on the subset of events where the specific value of Y occurs.
- In the above figure, we focus on when $Y = y$

$$p(X|Y = y) = \frac{P(X, Y = y)}{P(Y = y)}$$

Conditional probability

- We can now solve the card problem
- Let's have two events, X_1 for observed side of the card, and X_2 for the side we want to guess
- We need to calculate $P(X_2 = Red | X_1 = Red)$

$P(X_1 = Red) = \frac{1}{2}$
$P(X_1 = Red, X_2 = Red) = \frac{1}{3}$
therefore

$$P(X_2 = Red | X_1 = Red) = \frac{P(X_2 = Red, X_1 = Red)}{P(X_1 = Red)} = \frac{1/3}{1/2} = \frac{2}{3}$$

The chain rule

- Consider the conditional probability rule

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$

- It allows us to derive the chain rule, which defines the joint distribution as a product of conditionals:

$$P(X,Y) = P(X,Y)\frac{P(Y)}{P(Y)}$$

$$= P(X|Y)P(Y)$$

- In general, for any set of variables

$$P(X_1, X_2, ..., X_N) = \prod_{n=1}^{N} P(X_n|X_1, X_2, ..., X_{n-1})$$

- For example:

$$P(X,Y,Z) = P(X)P(Y|X)P(Z|Y,X)$$

Bayes theorem

- Using the chain rule, we can trivially say:

$$P(X|Y)P(Y) = P(Y|X)P(X)$$

  which means that [*Bayes theorem*]:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

- The Bayes theorem is an important foundation for Bayesian statistics, and particularly for Probabilistic Graphical Models!

Playtime!

- Open "1.Probability_Review.ipynb" in Jupyter
- Do Part 1, estimated duration 20 min

Independence

- Random variables are *independent* if knowing about X tells us nothing about Y.

$$P(Y|X) = P(Y)$$

- This means that their joint distribution is

$$P(X, Y) = P(X)P(Y)$$

  Why?

- A couple of examples:
  - Two persons, A, and B, start their trip in different parts of town. The transport mode for A is X and for B, it is Y. Are these two choices independence?
  - It's a rainy day. Two accidents happen on different roads of the city, far from each other. Are these two, independent events?

Independence

- Are these independent?
  - the speeds in adjacent road sections
  - the flow of pedestrians and the flow of cars in the same road
  - whether it is raining and the number of taxi cabs
  - whether it is raining and the amount of time it takes me to hail a cab
  - the departure time and arrival time of a trip

Independence

- Example: two coins, C_1, C_2 with $P(H|C_1) = 0.5, P(H|C_2) = 0.3$

  1. Suppose that I randomly choose a number $Z \in \{1, 2\}$, and take coin C_Z.
  2. I flip it twice, with results (X_1, X_2)

  Are X_1 and X_2 independent? What about if I know Z?

Conditional independence

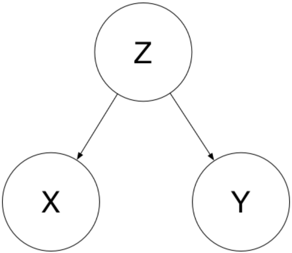
- X and Y are *conditionally independent* given Z

$$P(X|Y,Z) = P(X|Z)$$

- So, we can say that

$$X \perp\!\!\!\perp Y|Z \implies P(X,Y|Z) = P(X|Z)P(Y|Z)$$

- Graphical model notation:



Conditional independence

- X and Y are *conditionally independent* given Z

$$P(X|Y,Z) = P(X|Z)$$

- So, we can say that
- *If we know Z, then knowing about Y tells us nothing about X*

Playtime!

- Open "1.Probability_Review.ipynb" in Jupyter
- Do Part 2, estimated duration 30 min

Expectation

- The *expected value* of a random variable is the probability-weighted average of all possible values.
- In other words, it is the *mean* of the distribution of this random variable

$$E(X) = \sum_x x P(X = x)$$

- More generically (remember the f(x) can be itself a random variable)

$$E(f(X)) = \sum_x f(x) P(X = x)$$

Playtime!

- Open "1.Probability_Review.ipynb" in Jupyter
- Do Part 3, estimated duration 10 min