Mathematical Foundations

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

SLIDES ADAPTED FROM DAVE BLEI AND LAUREN HANNAH

Random variable

- Probability is about *random variables.*
- A random variable is any “probabilistic” outcome.
- For example,
  - The flip of a coin
  - The height of someone chosen randomly from a population
- We’ll see that it’s sometimes useful to think of quantities that are not strictly probabilistic as random variables.
  - The temperature on 11/12/2013
  - The temperature on 03/04/1905
  - The number of times “streetlight” appears in a document

Random variable

- Random variables take on values in a *sample space*.
- They can be *discrete* or *continuous*:
  - Coin flip: $\{H, T\}$
  - Height: positive real values $(0, \infty)$
  - Temperature: real values $(-\infty, \infty)$
  - Number of words in a document: Positive integers $\{1, 2, \dots\}$
- We call the outcomes *events*.
- Denote the random variable with a capital letter; denote a realization of the random variable with a lower case letter.
- E.g., X is a coin flip, x is the value (H or T) of that coin flip.

Discrete distribution

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

$$\begin{aligned} P(X=H) &= 0.7 \\ P(X=T) &= 0.3 \end{aligned}$$

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

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

- The probabilities over the entire space must sum to one

Discrete distribution

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

$$\begin{aligned} P(X=H) &= 0.7 \\ P(X=T) &= 0.3 \end{aligned}$$

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

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

- The probabilities over the entire space must sum to one

Discrete distribution

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

$$\begin{aligned} P(X=H) &= 0.7 \\ P(X=T) &= 0.3 \end{aligned}$$

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

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

- The probabilities over the entire space must sum to one

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

Discrete distribution

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

$$\begin{aligned} P(X=H) &= 0.7 \\ P(X=T) &= 0.3 \end{aligned}$$

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

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

- The probabilities over the entire space must sum to one

$$\sum_{x} P(X=x)=1$$

Events

An *event* is a set of outcomes to which a probability is assigned

- drawing a black card from a deck of cards
- drawing a King of Hearts

Intersections and unions:

- Intersection: drawing a red and a King

$$P(A \cap B) \tag{1}$$

- Union: drawing a spade or a King

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2}$$

Events

An *event* is a set of outcomes to which a probability is assigned

- drawing a black card from a deck of cards
- drawing a King of Hearts

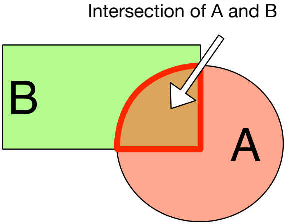
Intersections and unions:

- Intersection: drawing a red and a King

$$P(A \cap B) \tag{1}$$

- Union: drawing a spade or a King

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2}$$



Events

An *event* is a set of outcomes to which a probability is assigned

- drawing a black card from a deck of cards
- drawing a King of Hearts

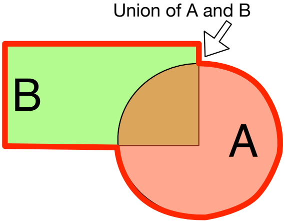
Intersections and unions:

- Intersection: drawing a red and a King

$$P(A \cap B) \tag{1}$$

- Union: drawing a spade or a King

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2}$$



Joint distribution

- Typically, we consider collections of random variables.
- The joint distribution is a distribution over the configuration of all the random variables in the ensemble.
- For example, imagine flipping 4 coins. The joint distribution is over the space of all possible outcomes of the four coins.

$$
\begin{aligned}
P(HHHH) &= 0.0625 \\
P(HHHT) &= 0.0625 \\
P(HHTH) &= 0.0625 \\
&\dots
\end{aligned}
$$

- You can think of it as a single random variable with 16 values.

Visualizing a joint distribution