Probability Theory

Probability and Statistics for Data Science
CSE594 - Spring 2016

What is Probability?

What is Probability?

Examples

- outcome of flipping a coin (seminal example)
- amount of snowfall
- mentioning a word
- mentioning a word “a lot”

What is Probability?

The chance that something will happen.

Given infinite observations of an event, the proportion of observations where a given outcome happens.

Strength of belief that something is true.

“Mathematical language for quantifying uncertainty” - Wasserman

Probability (review)

Ω : Sample Space, set of all outcomes of a random experiment

A : Event (***A*** $\subseteq$ **Ω**), collection of possible outcomes of an experiment

P(*A*): Probability of event ***A*, P** is a function: events$\rightarrow\mathbb{R}$

Probability (review)

Ω : Sample Space, set of all outcomes of a random experiment

A : Event (***A*** ⊆ **Ω**), collection of possible outcomes of an experiment

P(*A*): Probability of event ***A*, P** is a function: events→ℝ

- **P(Ω)** = 1
- **P(*A*)** ≥ 0 **,** for all ***A***
- If A_1, A_2, … are disjoint events then: $$\mathrm{P}(\bigcup_i^\infty A_i) = \sum_i^\infty \mathrm{P}(A_i)$$

Probability (review)

Ω : Sample Space, set of all outcomes of a random experiment

A : Event (***A*** ⊆ **Ω**), collection of possible outcomes of an experiment

P(*A*): Probability of event ***A*, P** is a function: events→ℝ

P is a *probability measure*, if and only if

- **P(Ω)** = 1
- **P(*A*)** ≥ 0 **,** for all ***A***
- If A_1, A_2, … are disjoint events then: $$\mathrm{P}(\bigcup_{i}^{\infty} A_i) = \sum_{i}^{\infty} \mathrm{P}(A_i)$$

Probability

Examples

- outcome of flipping a coin (seminal example)
- amount of snowfall
- mentioning a word
- mentioning a word “a lot”

Probability (review)

Some Properties:

If $B \subseteq A$ then $P(A) \geq P(B)$

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

$P(A \cap B) \leq \min(P(A), P(B))$

$P(\neg A) = P(\Omega / A) = 1 - P(A)$

/ is set difference
$P(A \cap B)$ will be notated as $P(A, B)$

Probability (Review)

Independence

Two Events: A and B

Does knowing something about A tell us whether B happens (and vice versa)?

Probability (Review)

Independence

Two Events: *A* and *B*

Does knowing something about *A* tell us whether *B* happens (and vice versa)?

- A: first flip of a fair coin; B: second flip of the same fair coin
- A: mention or not of the word “happy”
  B: mention or not of the word “birthday”

Probability (Review)

Independence

Two Events: *A* and *B*

Does knowing something about *A* tell us whether *B* happens (and vice versa)?

- A: first flip of a fair coin; B: second flip of the same fair coin
- A: mention or not of the word “happy”
  B: mention or not of the word “birthday”

Two events, A and B, are *independent* iff **P(*A*, *B*) = P(*A*)P(*B*)**

Probability (Review)

Conditional Probability

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

Probability (Review)

Conditional Probability

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

H: mention “happy” in message, m

B: mention “birthday” in message, m

P(H) = .01 P(B) =.001 P(H, B) = .0005

P(H|B) = ??

Probability (Review)

Conditional Probability

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

H: mention "happy" in message, m

B: mention "birthday" in message, m

P(H) = .01   P(B) =.001   P(H, B) = .0005

P(H|B) = .50

H1: first flip of a fair coin is heads

H2: second flip of the same coin is heads

P(H2) = **0.5**   P(H1) = 0.5   P(H2, H1) = 0.25

P(H2|H1) = **0.5**

Probability (Review)

Conditional Probability

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

H1: first flip of a fair coin is heads

H2: second flip of the same coin is heads

P(H2) = **0.5** P(H1) = 0.5 P(H2, H1) = 0.25

P(H2|H1) = **0.5**

Two events, A and B, are *independent* iff **P(*A*, *B*) = P(*A*)P(*B*)**

P(*A*, *B*) = P(*A*)P(*B*) iff **P(*A*|*B*) = P(A)**

Probability (Review)

Conditional Probability

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

H1: first flip of a fair coin is heads

H2: second flip of the same coin is heads

P(H2) = **0.5** P(H1) = 0.5 P(H2, H1) = 0.25

P(H2|H1) = **0.5**

Two events, A and B, are *independent* iff **P(*A*, *B*) = P(*A*)P(*B*)**

P(*A*, *B*) = P(*A*)P(*B*) iff **P(*A*|*B*) = P(A)**

Interpretation of Independence:

Observing B has no effect on probability of A.

Why Probability?

Why Probability?

A formality to make sense of the world.

- To quantify uncertainty
  Should we believe something or not? Is it a meaningful difference?

- To be able to generalize from one situation or point in time to another.
  Can we rely on some information? What is the chance Y happens?

- To organize data into meaningful groups or “dimensions”
  Where does X belong? What words are similar to X?

Random Variables

X: A mapping from $\boldsymbol{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

Random Variables

X: A mapping from $\boldsymbol{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Example: $\boldsymbol{\Omega}$ = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

We may just care about how many tails? Thus,

X(<HHHHH>) = 0

X(<HHHTH>) = 1

X(<TTTHT>) = 4

X(<HTTTT>) = 4

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

We may just care about how many tails? Thus,

X(<HHHHH>) = 0

X(<HHHTH>) = 1

X(<TTTHT>) = 4

X(<HTTTT>) = 4

X only has 6 possible values: 0, 1, 2, 3, 4, 5

Random Variables

X: A mapping from $\boldsymbol{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Example: $\boldsymbol{\Omega}$ = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

We may just care about how many tails? Thus,

$X(<HHHHH>) = 0$

$X(<HHHTH>) = 1$

$X(<TTTHT>) = 4$

$X(<HTTTT>) = 4$

X only has 6 possible values: 0, 1, 2, 3, 4, 5

What is the probability that we end up with $k = 4$ tails?

$\mathbf{P}(X(\omega) = k)$ where $\omega \in \boldsymbol{\Omega}$

Random Variables

X: A mapping from $\boldsymbol{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Example: $\boldsymbol{\Omega}$ = 5 coin tosses = $\{\langle HHHHH \rangle, \langle HHHHT \rangle, \langle HHHTH \rangle, \langle HHHTH \rangle \ldots\}$

We may just care about how many tails? Thus,

$X(\langle HHHHH \rangle) = 0$

$X(\langle HHHTH \rangle) = 1$

$X(\langle TTTHT \rangle) = 4$

$X(\langle HTTTT \rangle) = 4$

X only has 6 possible values: 0, 1, 2, 3, 4, 5

What is the probability that we end up with $k = 4$ tails?

$$\mathbf{P}(X = k) := \mathbf{P}(\ \{\omega : X(\omega) = k\}\ ) \qquad \text{where } \omega \in \boldsymbol{\Omega}$$

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}
We may just care about how many tails? Thus,

X(<HHHHH>) = 0

X(<HHHTH>) = 1

X(<TTTHT>) = 4

X(<HTTTT>) = 4

X only has 6 possible values: 0, 1, 2, 3, 4, 5
What is the probability that we end up with $k = 4$ tails?

$\mathbf{P}(X = k) := \mathbf{P}(\ \{\omega : X(\omega) = k\}\ )$ where $\omega \in \mathbf{\Omega}$

$X(\omega) = 4$ for 5 out of 32 sets in **Ω**. Thus, assuming a fair coin, $\mathbf{P}(X = 4) = 5/32$

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

We may just care about how many tails? Thus,

$X(<HHHHH>) = 0$

$X(<HHHTH>) = 1$

$X(<TTTHT>) = 4$

$X(<HTTTT>) = 4$

X only has 6 possible values: 0, 1, 2, 3, 4, 5

What is the probability that we end up with $k = 4$ tails?

$\mathbf{P}(X = k) := \mathbf{P}(\ \{\omega : X(\omega) = k\}\ )$ where $\omega \in \mathbf{\Omega}$

$X(\omega) = 4$ for 5 out of 32 sets in **Ω**. Thus, assuming a fair coin, $\mathbf{P}(X = 4) = 5/32$

(Not a variable, but a function that we end up notating a lot like a variable)

Random Variables

X: A mapping from $\boldsymbol{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Example: $\boldsymbol{\Omega}$ = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

We may just care about how many tails? Thus,

X(<HHHHH>) = 0

X(<HHHTH>) = 1

X(<TTTHT>) = 4

X(<HTTTT>) = 4

X is a *discrete random variable* if it takes only a countable number of values.

X only has 6 possible values: 0, 1, 2, 3, 4, 5

What is the probability that we end up with k = 4 tails?

$\mathbf{P}(X = k) := \mathbf{P}(\ \{\omega : X(\omega) = k\}\ )$ where $\omega \in \boldsymbol{\Omega}$

$X(\omega) = 4$ for 5 out of 32 sets in $\boldsymbol{\Omega}$. Thus, assuming a fair coin, $\mathbf{P}(X = 4) = 5/32$

(Not a variable, but a function that we end up notating a lot like a variable)

Random Variables

$\mathbf{X}$: A mapping from Ω to $\mathbb{R}$ that describes the question we care about in practice.

$\mathbf{X}$ is a *continuous random variable* if it can take on an infinite number of values between any two given values.

$\mathbf{X}$ is a *discrete random variable* if it takes only a countable number of values.

Random Variables

$\mathbf{X}$: A mapping from $\mathbf{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Example: $\mathbf{\Omega}$ = inches of snowfall = $[0, \infty) \subseteq \mathbb{R}$

$\mathbf{X}$ is a *continuous random variable* if it can take on an infinite number of values between any two given values.

Random Variables

$\mathbf{X}$: A mapping from $\boldsymbol{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Example: $\boldsymbol{\Omega}$ = inches of snowfall = $[0, \infty) \subseteq \mathbb{R}$

$\mathbf{X}$ is a *continuous random variable* if it can take on an infinite number of values between any two given values.

X amount of inches in a snowstorm

$\mathbf{X}(\omega) = \omega$

Random Variables

$\mathbf{X}$: A mapping from $\boldsymbol{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Example: $\boldsymbol{\Omega}$ = inches of snowfall = $[0, \infty) \subseteq \mathbb{R}$

$\mathbf{X}$ is a *continuous random variable* if it can take on an infinite number of values between any two given values.

X amount of inches in a snowstorm

$\mathbf{X}(\omega) = \omega$

What is the probability we receive (at least) a *inches?*

$\mathbf{P}(X \geq a) := \mathbf{P}(\ \{\omega : X(\omega) \geq a\}\ )$

What is the probability we receive between a *and* b *inches?*

$\mathbf{P}(a \leq X \leq b) := \mathbf{P}(\ \{\omega : a \leq X(\omega) \geq b\}\ )$

Random Variables

X: A mapping from Ω to $\mathbb{R}$ that describes the question we care about in practice.

Example: Ω = inches of snowfall = $[0, \infty) \subseteq \mathbb{R}$

X is a *continuous random variable* if it can take on an infinite number of values between any two given values.

X amount of inches in a snowstorm

$\mathbf{X}(\omega) = \omega$

$\mathbf{P}(X = i) := 0$, for all $i \in \mathbf{\Omega}$

(probability of receiving <u>exactly</u> i inches of snowfall is zero)

What is the probability we receive (at least) a *inches?*

$\mathbf{P}(X \geq a) := \mathbf{P}(\ \{\omega : X(\omega) \geq a\}\ )$

What is the probability we receive between a *and* b *inches?*

$\mathbf{P}(a \leq X \leq b) := \mathbf{P}(\ \{\omega : a \leq X(\omega) \geq b\}\ )$

Probability Review

- what constitutes a probability measure?
- independence
- conditional probability
- random variables
  - discrete
  - continuous