Introduction to Probability

Random variables and independence

Outline

- Probability and sample spaces
- Random variables
- Distributions
- Independence

Definition of Probability

- *frequentist* interpretation: the probability of an event from a random experiment is proportion of the time events of the same kind will occur in the long run, when the experiment is repeated

- examples
  - the probability my flight to Chicago will be on time
  - the probability this ticket will win the lottery
  - the probability it will rain tomorrow

- always a number in the interval [0,1]
  0 means "never occurs"
  1 means "always occurs"

Sample Spaces

- *sample space*: a set of possible outcomes for some experiment

- examples
  - flight to Chicago: {on time, late}
  - lottery: {ticket 1 wins, ticket 2 wins,…,ticket *n* wins}
  - weather tomorrow:

    {rain, not rain} or

    {sun, rain, snow} or

    {sun, clouds, rain, snow, sleet} or…

Random Variables

- *random variable*: a function that maps the outcome of an experiment to a label (often a numerical value)
- example
  - X represents the outcome of my flight to Chicago
  - we write the probability of my flight being on time as $\Pr(X = \text{on-time})$
  - or when it's clear which variable we're referring to, we may use the shorthand $\Pr(\text{on-time})$

Notation

- UPPERCASE letters and Capitalized words denote random variables
- lowercase letters and uncapitalized words denote values
- we'll denote a particular value for a variable as follows

$$\Pr(X = x) \qquad \Pr(Fever = true)$$

- we'll also use the shorthand form

$$\Pr(x) \quad \text{for} \quad \Pr(X = x)$$

- for Boolean random variables, we'll use the shorthand

$$\Pr(fever) \;\; \text{for} \;\; \Pr(Fever = true)$$

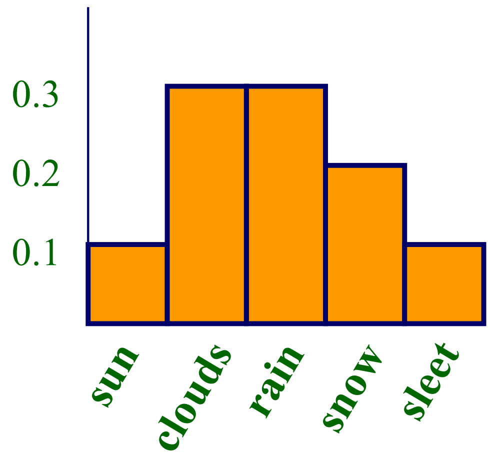
$$\Pr(\neg fever) \;\; \text{for} \;\; \Pr(Fever = false)$$

Probability Distributions

- if X is a random variable, the function given by $\Pr(X = x)$ for each x is the *probability distribution* of X
- requirements:

$$\Pr(x) \geq 0 \quad \text{for every } x$$

$$\sum_x \Pr(x) = 1$$



Joint Distributions

- *joint probability distribution*: the function given by $\Pr(X = x, Y = y)$
- read "X equals x <u>and</u> Y equals y"
- example

x, y	$\Pr(X = x, Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

probability that it's sunny and my flight is on time

Marginal Distributions

- the *marginal distribution* of X is defined by

$$\Pr(x) = \sum_y \Pr(x, y)$$

"the distribution of X ignoring other variables"

- this definition generalizes to more than two variables, e.g.

$$\Pr(x) = \sum_y \sum_z \Pr(x, y, z)$$

Marginal Distribution Example

joint distribution

x, y	$\Pr(X = x, Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

marginal distribution for X

x	$\Pr(X = x)$
sun	0.3
rain	0.5
snow	0.2

Conditional Distributions

- the *conditional distribution* of X given Y is defined as:

$$\Pr(X = x \mid Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}$$

"the distribution of X given that we know Y"

Conditional Distribution Example

joint distribution

x, y	$\Pr(X = x, Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

conditional distribution for X given Y=on-time

x	$\Pr(X = x \mid Y = \textit{on-time})$
sun	0.20/0.45 = 0.444
rain	0.20/0.45 = 0.444
snow	0.05/0.45 = 0.111

Independence

- two random variables, *X* and *Y*, are *independent* if and only if

$$\Pr(x, y) = \Pr(x) \times \Pr(y) \quad \text{for all } x \text{ and } y$$

- this means that

$$\Pr(x|y) = \Pr(x)$$

and

$$\Pr(y|x) = \Pr(y)$$

Independence Example #1

joint distribution

x, y	$\Pr(X = x, Y = y)$
sun, on-time	0.20
rain, on-time	0.20
snow, on-time	0.05
sun, late	0.10
rain, late	0.30
snow, late	0.15

marginal distributions

x	$\Pr(X = x)$
sun	0.3
rain	0.5
snow	0.2

y	$\Pr(Y = y)$
on-time	0.45
late	0.55

Are X and Y independent here? NO.

Independence Example #2

joint distribution

x, y	$\Pr(X = x, Y = y)$
sun, fly-United	0.27
rain, fly-United	0.45
snow, fly-United	0.18
sun, fly-Delta	0.03
rain, fly-Delta	0.05
snow, fly-Delta	0.02

marginal distributions

x	$\Pr(X = x)$
sun	0.3
rain	0.5
snow	0.2

y	$\Pr(Y = y)$
fly-United	0.9
fly-Delta	0.1

Are X and Y independent here? YES.

Conditional Independence

- two random variables X and Y are *conditionally independent* given Z if and only if

$$\Pr(X \mid Y, Z) = \Pr(X \mid Z)$$

"once you know the value of Z, knowing Y doesn't tell you anything about X"

- alternatively

$$\Pr(x, y \mid z) = \Pr(x \mid z) \times \Pr(y \mid z) \quad \text{for all } x, y, z$$

Conditional Independence Example

Flu	Fever	Vomit	Pr
true	true	true	0.04
true	true	false	0.04
true	false	true	0.01
true	false	false	0.01
false	true	true	0.009
false	true	false	0.081
false	false	true	0.081
false	false	false	0.729

Fever and Vomit are not independent: e.g. $\Pr(fever, vomit) \neq \Pr(fever) \times \Pr(vomit)$

Fever and Vomit are conditionally independent given Flu:

$$\Pr(fever, vomit \mid flu) = \Pr(fever \mid flu) \times \Pr(vomit \mid flu)$$

$$\Pr(fever, vomit \mid \neg flu) = \Pr(fever \mid \neg flu) \times \Pr(vomit \mid \neg flu)$$

etc.

Summary

- Probability and sample spaces
- Random variables
- Distributions
- Independence