Goals for the lecture

you should understand the following concepts
  • definition of probability
  • random variables
  • joint distributions
  • conditional distributions
  • chain rule
  • independence
  • union rule
  • Bayes theorem
  • expected values
  • multinomial distribution
  • probability density function
  • normal distribution
Definition of probability

• *frequentist* interpretation: the probability of an event from a random experiment is the proportion of the time events of 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 event

• 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 variable representing the outcome of an event

- example
  - $X$ represents the outcome of my flight to Chicago
  - we write the probability of my flight being on time as $P(X = \text{on-time})$
  - or when it’s clear which variable we’re referring to, we may use the shorthand $P(\text{on-time})$

Notation

- uppercase letters and capitalized words denote random variables
- lowercase letters and uncapsulalted words denote values
- we’ll denote a particular value for a variable as follows
  \[ P(X = x) \quad P(\text{Fever} = \text{true}) \]
- we’ll also use the shorthand form
  \[ P(x) \quad \text{for} \quad P(X = x) \]
- for Boolean random variables, we’ll use the shorthand
  \[ P(\text{fever}) \quad \text{for} \quad P(\text{Fever} = \text{true}) \]
  \[ P(\neg\text{fever}) \quad \text{for} \quad P(\text{Fever} = \text{false}) \]
Probability distributions

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

- **requirements:**
  \[ P(x) \geq 0 \quad \text{for every } x \]
  \[ \sum_x P(x) = 1 \]

Joint distributions

- **joint probability distribution:** the function given by \( P(X = x, Y = y) \)
- read “\( X \) equals \( x \) and \( Y \) equals \( y \)”
- **example**

<table>
<thead>
<tr>
<th>( x, y )</th>
<th>( P(X = x, Y = y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun, on-time</td>
<td>0.20</td>
</tr>
<tr>
<td>rain, on-time</td>
<td>0.20</td>
</tr>
<tr>
<td>snow, on-time</td>
<td>0.05</td>
</tr>
<tr>
<td>sun, late</td>
<td>0.10</td>
</tr>
<tr>
<td>rain, late</td>
<td>0.30</td>
</tr>
<tr>
<td>snow, late</td>
<td>0.15</td>
</tr>
</tbody>
</table>

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

• the marginal distribution of $X$ is defined by

\[ P(x) = \sum_y P(x,y) \]

"the distribution of $X$ ignoring other variables"

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

\[ P(x) = \sum_y \sum_z P(x,y,z) \]

Marginal distribution example

<table>
<thead>
<tr>
<th>joint distribution</th>
<th>marginal distribution for $X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x, y$</td>
<td>$P(X = x, Y = y)$</td>
</tr>
<tr>
<td>sun, on-time</td>
<td>0.20</td>
</tr>
<tr>
<td>rain, on-time</td>
<td>0.20</td>
</tr>
<tr>
<td>snow, on-time</td>
<td>0.05</td>
</tr>
<tr>
<td>sun, late</td>
<td>0.10</td>
</tr>
<tr>
<td>rain, late</td>
<td>0.30</td>
</tr>
<tr>
<td>snow, late</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Conditional distributions

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

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

"the distribution of $X$ given that we know the value of $Y"$

### Conditional distribution example

<table>
<thead>
<tr>
<th>joint distribution</th>
<th>conditional distribution for $X$ given $Y=$on-time</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x, y$</td>
<td>$P(X = x, Y = y)$</td>
</tr>
<tr>
<td>sun, on-time</td>
<td>0.20</td>
</tr>
<tr>
<td>rain, on-time</td>
<td>0.20</td>
</tr>
<tr>
<td>snow, on-time</td>
<td>0.05</td>
</tr>
<tr>
<td>sun, late</td>
<td>0.10</td>
</tr>
<tr>
<td>rain, late</td>
<td>0.30</td>
</tr>
<tr>
<td>snow, late</td>
<td>0.15</td>
</tr>
</tbody>
</table>
The product rule

• rearranging the definition of the conditional distribution

\[ P(x \mid y) = \frac{P(x, y)}{P(y)} \]

• leads to the product rule

\[ P(x, y) = P(x \mid y)P(y) \]

---

The chain rule

• by repeated application of the product rule, a joint distribution can be expressed as

\[ P(x_1, x_2, \ldots, x_n) = P(x_1) \prod_{i=2}^{n} P(x_i \mid x_1, \ldots, x_{i-1}) \]

• permits the calculation of the joint distribution of a set of random variables using only conditional probabilities

• important idea for Bayesian networks
Independence

- two random variables, $X$ and $Y$, are independent if

\[ P(x,y) = P(x) \times P(y) \quad \text{for all } x \text{ and } y \]

- equivalently

\[
\begin{align*}
P(X \mid Y) &= P(X) \\
P(Y \mid X) &= P(Y)
\end{align*}
\]

Conditional independence

- two random variables, $X$ and $Y$, are conditionally independent given $Z$ if

\[ P(x,y \mid z) = P(x \mid z) \times P(y \mid z) \quad \text{for all } x, y \text{ and } z \]
### Independence example #1

<table>
<thead>
<tr>
<th>$x, y$</th>
<th>$P(X = x, Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun, on-time</td>
<td>0.20</td>
</tr>
<tr>
<td>rain, on-time</td>
<td>0.20</td>
</tr>
<tr>
<td>snow, on-time</td>
<td>0.05</td>
</tr>
<tr>
<td>sun, late</td>
<td>0.10</td>
</tr>
<tr>
<td>rain, late</td>
<td>0.30</td>
</tr>
<tr>
<td>snow, late</td>
<td>0.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(X = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.3</td>
</tr>
<tr>
<td>rain</td>
<td>0.5</td>
</tr>
<tr>
<td>snow</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$y$</th>
<th>$P(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>on-time</td>
<td>0.45</td>
</tr>
<tr>
<td>late</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Are $X$ and $Y$ independent here? **NO.**

### Independence example #2

<table>
<thead>
<tr>
<th>$x, y$</th>
<th>$P(X = x, Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun, fly-United</td>
<td>0.27</td>
</tr>
<tr>
<td>rain, fly-United</td>
<td>0.45</td>
</tr>
<tr>
<td>snow, fly-United</td>
<td>0.18</td>
</tr>
<tr>
<td>sun, fly-Delta</td>
<td>0.03</td>
</tr>
<tr>
<td>rain, fly-Delta</td>
<td>0.05</td>
</tr>
<tr>
<td>snow, fly-Delta</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(X = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sun</td>
<td>0.3</td>
</tr>
<tr>
<td>rain</td>
<td>0.5</td>
</tr>
<tr>
<td>snow</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$y$</th>
<th>$P(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>fly-United</td>
<td>0.9</td>
</tr>
<tr>
<td>fly-Delta</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Are $X$ and $Y$ independent here? **YES.**
Probability of union of events

- the probability of the union of two events is given by:

\[ P(x \lor y) = P(x) + P(y) - P(x, y) \]

Bayes' rule (or theorem)

recall the product rule

\[ P(x, y) = P(x | y)P(y) = P(y | x)P(x) \]

dividing the expressions on the right by \( P(y) \)

\[ P(x | y) = \frac{P(y | x)P(x)}{P(y)} = \frac{\sum_{x'} P(y | x')P(x')}{P(y)} \]
Bayes rule example

- \( P(\text{stiff-neck}|\text{meningitis}) = 0.5 \)
- \( P(\text{meningitis}) = \frac{1}{50,000} \)
- \( P(\text{stiff-neck}) = \frac{1}{20} \)

\[
P(\text{meningitis}|\text{stiff-neck}) = \frac{P(\text{stiff-neck}|\text{meningitis})P(\text{meningitis})}{P(\text{stiff-neck})} = \frac{0.5 \times \frac{1}{50,000}}{\frac{1}{20}} = 0.0002
\]

Why use Bayes rule?

- Causal knowledge such as \( P(\text{stiff-neck}|\text{meningitis}) \) is often more reliably estimated than diagnostic knowledge such as \( P(\text{meningitis}|\text{stiff-neck}) \)
- Bayes’ rule lets us use causal knowledge to make diagnostic inferences
Expected values

• the *expected value* of a random variable that takes on numerical values is defined as:

\[ E[X] = \sum_{x} x \times P(x) \]

this is the same thing as the *mean*

• we can also talk about the expected value of a function of a random variable

\[ E[g(X)] = \sum_{x} g(x) \times P(x) \]

Expected value examples

\[ E[\text{Shoesize}] = \]

\[ 5 \times P(\text{Shoesize} = 5) + \ldots + 14 \times P(\text{Shoesize} = 14) \]

• Suppose each lottery ticket costs $1 and the winning ticket pays out $100. The probability that a particular ticket is the winning ticket is 0.001.

\[ E[\text{gain(Lottery)}] = \]

\[ \text{gain(winning)}P(\text{winning}) + \text{gain(losing)}P(\text{losing}) = \]

\[ ($100 - $1) \times 0.001 - $1 \times 0.999 = \]

\[ - $0.90 \]
The binomial distribution

- Distribution over the number of successes in a fixed number $n$ of independent trials (with same probability of success $p$ in each)

$$P(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

- E.g. the probability of $x$ heads in $n$ coin flips

![Binomial distribution w/ p=0.5, n=10](image)

The geometric distribution

- Distribution over the number of trials before the first failure (with same probability of success $p$ in each)

$$P(x) = (1 - p) p^x$$

- E.g. the probability of $x$ heads before the first tail

![Geometric distribution w/ p=0.5, n=10](image)
The multinomial distribution

- \(k\) possible outcomes on each trial
- probability \(p_i\) for outcome \(x_i\) in each trial
- distribution over the number of occurrences \(x_i\) for each outcome in a fixed number \(n\) of independent trials

\[ P(x) = \frac{n!}{\prod_i (x_i!)} \prod_i p_i^{x_i} \]

- e.g. with \(k=6\) (a six-sided die) and \(n=30\)

\[ P([7,3,0,8,10,2]) = \frac{30!}{7!\times3!\times0!\times8!\times10!\times2!} \left( p_1^7 p_2^3 p_3^0 p_4^8 p_5^{10} p_6^2 \right) \]

Continuous random variables

- up to now, we’ve considered only discrete random variables, but we can have RVs describing continuous variables too (weight, temperature, etc.)

- a continuous random variable is described by a probability density function (p.d.f.)
Probability density functions

- a continuous random variable is described by a
  probability density function \( f(x) \)

\[ \forall x \quad f(x) \geq 0 \]

\[ P[a \leq X \leq b] = \int_a^b f(x) \, dx \]

\[ \int_a^b f(x) \, dx = 1 \]

The normal (Gaussian) distribution

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
Some p.d.f.'s

- normal (Gaussian)
- uniform
- Gamma
- student's t