Probability review

CS B553
Spring 2013

Announcements

• Readings and lecture notes online on OnCourse
  – Under the “Wiki” tab

• Assignment 1 coming soon!

The Birthday Problem

• In a group of x people, what’s the probability that at least two share the same birthday?

Uses Stirling’s approximation:

\[
\text{# of people}
\]

Bayes’ Law

• For two events A and B,

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B)}
\]

• Useful when you want to know something about A, but all you can directly observe is B
  – This process is called Bayesian inference

Bayes’ Law Example #1

• I have two coins, one fair and one with heads on both sides. I choose a coin at random, flip it, and get heads. What is the probability that it is the fair coin?

Bayes’ Law Example #2

• A certain medical condition afflicts 0.1% of the population. There exists a test that is about 95% accurate. If the test says you have the condition, what’s the probability that you actually have the condition?
Random variables

- A random variable is like a multi-valued “attribute” that can be used to specify sets of outcomes
  - Formally, it’s a function that associates an attribute with each outcome, typically $f : S \rightarrow \mathbb{N}$ or $f : S \rightarrow \mathbb{R}$
  - E.g. Roll a die 5 times. Let random variable $X$ be the number of 1’s rolled.
    - Now $P(X=2)$ denotes probability of rolling two 1’s, i.e. $P(X = 2) = P(\{a \in S | f(a) = 2\})$, where $f(a)$ counts the number of 1’s in a given outcome, i.e. $f(11111)=5$, $f(11114)=4$, ...

Marginal distributions

- A probability distribution over the values of a random variable is called a marginal distribution
  - E.g. Roll a die 5 times. Let random variable $X$ be the number of 1’s rolled.
    - $P(X)$ is the marginal distribution over $X$

Joint distributions

- A probability distribution over multiple random variables is called a joint distribution
  - E.g. Roll a die 5 times. Let random variable $X$ be the number of 1’s rolled, and $Y$ be the sum of the rolls.
    - $P(X,Y)$ is the joint distribution, i.e. $P : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{R}$
    - E.g. $P(X=3, Y=6)$ refers to $P((X = 3) \cap (Y = 6))$

An example

- Given set of students, define random variables $X$ for Intelligence and $Y$ for grade in class
  - Joint distribution $P(X,Y)$ given by:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.07 0.18</td>
</tr>
<tr>
<td>B</td>
<td>0.28 0.09</td>
</tr>
<tr>
<td>C</td>
<td>0.35 0.03</td>
</tr>
</tbody>
</table>

- What are the marginal distributions $P(X)$ and $P(Y)$?

Conditional probabilities

- Conditional probabilities can also be written using random variables
  - Denoted $P(X|Y)$ for random variables $X$ and $Y$
  - For any value of $Y$, this is a distribution over $X$
  - Example: What’s $P(X=\text{low} \mid Y=\text{B})$?

<table>
<thead>
<tr>
<th></th>
<th>Intelligence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>low  high</td>
</tr>
<tr>
<td>A</td>
<td>0.07 0.18</td>
</tr>
<tr>
<td>B</td>
<td>0.28 0.09</td>
</tr>
<tr>
<td>C</td>
<td>0.35 0.03</td>
</tr>
</tbody>
</table>

Independence of events

- Two events are independent if $P(A \mid B) = P(A)$
  - Or, equivalently, if $P(B \mid A) = P(B)$
  - Independence denoted $A \perp B$
- The joint probability of independent events $A$ and $B$ both occurring is then simply:
  $P(A \cap B) = P(A)P(B)$
  - This idea of factoring a distribution into a product of two simpler distributions will be a recurring course theme!
### Conditional independence

- Sometimes events are both conditioned on the same event, but otherwise are independent
  - Suppose Mary and Bob are two independent people
  - E.g. Let A denote the event that it is raining, B denote event that Mary is wet, and C denote event that Bob is wet
  - Events B and C are not independent
  - But B and C are conditionally independent given A,
    
    \[
    P(B|A, C) = P(B|A)
    \]
    
    \[
    P(C|A, B) = P(C|A)
    \]
  - Denoted $B \perp C | A$

### Using distributions to answer questions

- Several ways we’ll use probability distributions
- **Probability queries**
  - Calculate distribution over some random variables given observed values for others
  - E.g. calculate $P(Y | X = x)$, where $X$ and $Y$ are sets of random variables
  - There might be a set of other random variables $Z$ that are neither the query variables nor the observations; these can be "marginalized out":
    
    \[
    P(Y | X = x) = \sum_Z P(Y, Z | X = x)
    \]

### Using distributions to answer questions

- Maximum A Priori (MAP) queries
  - Calculate the most likely values for some random variables, given observed values for others,
    
    \[
    \arg \max_Y P(Y = y | E = e)
    \]

### Expectation

- The expected value of a (discrete) random variable $X$,
  
  \[
  E[X] = \sum_{x \in \text{Val}(X)} x P(x)
  \]
  - E.g. Let $X$ be the amount of money you win in the $600 million Powerball Jackpot (Nov 2012). The probability of winning is about 1 in 200 million. What is $E[X]$?

### Properties of expectations

- Expectation is linear in the random variable:
  
  \[
  E[aX + b] = aE[X] + b
  \]
- Linearity of expectations:
  
  \[
  E[X + Y] = E[X] + E[Y]
  \]
- Products of random variables (if rv’s are independent):
  
  \[
  E[XY] = E[X]E[Y]
  \]

### Variance

- The variance of a random variable $X$ is,
  
  \[
  V[X] = E[(X - E[X])^2]
  \]
  - Or, equivalently,
    
    \[
    V[X] = E[X^2] - E[X]^2
    \]
- Sums of variances (only if $X$ and $Y$ are independent):
  
  \[
  V[aX + b] = a^2V[X]
  \]
  
  \[
  V[X + Y] = V[X] + V[Y]
  \]