Discrete Probability II

COMPSCI 230 — Discrete Math

April 7, 2016
Outline

1. Joint and Marginal Probability Functions
2. Independent Outcomes and Events
3. Conditional Probabilities and Independence
Two Biased Coins

\[ \mathcal{S} = \{H, T\} \times \{H, T\} = \{(H, H), (H, T), (T, H), (T, T)\} \]

\[ \mathcal{E} = \{\emptyset, \{(H, H)\}, \{(H, T)\}, \{(T, H)\}, \{(T, T)\}, \ldots\} \]

\[ |\mathcal{S}| = 4 \Rightarrow |\mathcal{E}| = 2^4 = 16 \]

\( \binom{4}{k} \) events with \( k \) outcomes

- **Most general** probability function:
  
  \[ P((H, H)) = p_{HH} \quad P((H, T)) = p_{HT} \]
  
  \[ P((T, H)) = p_{TH} \quad P((T, T)) = p_{TT} \]

  with \( p_i \geq 0 \) and \( p_{HH} + p_{HT} + p_{TH} + p_{TT} = 1 \)
Joint Probability Function

\[
\begin{array}{c|ccc|c}
 & H_2 & T_2 & \\
\hline
H_1 & p_{HH} & p_{HT} & p_{H_1} \\
T_1 & p_{TH} & p_{TT} & p_{T_1} \\
\hline
p_{H_2} & p_{T_2} & p_S
\end{array}
\begin{array}{c|ccc|c}
 & H_2 & T_2 & \\
\hline
H_1 & 0.2 & 0.1 & 0.3 \\
T_1 & 0.4 & 0.3 & 0.7 \\
\hline
0.6 & 0.4 & 1
\end{array}
\]

- \(p_{HH} = P((H, H))\)
- Events \(\{(H, H)\}\) and \(\{(H, T)\}\) are mutually exclusive
- So \(p_{H_1} = \text{prob}(\{(H, \ast)\}) = \text{prob}(\{(H, H)\} \cup \{(H, T)\}) = \text{prob}(\{(H, H), (H, T)\}) = p_{HH} + p_{HT}\)
- Marginal distribution for coin 1 on right margin
- Marginal distribution for coin 2 on bottom margin
- Adding marginals yields 1 either way
Marginal Probability Functions

- $p_{H_1} = \text{prob}(\{(H, \ast)\}) = \text{prob}(\{(H, H)\} \cup \{(H, T)\}) = \text{prob}(\{(H, H), (H, T)\}) = p_{HH} + p_{HT}$
- More generally, a compound experiment produces outcome $(X, Y)$ in sample space $\mathcal{S} = \mathcal{S}_X \times \mathcal{S}_Y$
- $P(X) = \sum_{Y \in \mathcal{S}_Y} P(X, Y)$
- $P(Y) = \sum_{X \in \mathcal{S}_X} P(X, Y)$ (marginalization)
- Sum all rows (or all columns) of a joint probability function to obtain the marginal probability function
- $P(X), P(Y)$ defined on $\mathcal{S}_X$ and $\mathcal{S}_Y$
- Notice the overloading of the symbol $P$
- This abuse of notation is common in probability theory because of its convenience
Independent Outcomes

• Two independent, biased coins

\[ P((H, H)) = p_1 p_2 \quad P((H, T)) = p_1 q_2 \]
\[ P((T, H)) = q_1 p_2 \quad P((T, T)) = q_1 q_2 \]

• “Independent outcomes” = probabilities factor:

\[ P((H, H)) = P(H)P(H) \quad P((H, T)) = P(H)P(T) \]
\[ P((T, H)) = P(T)P(H) \quad P((T, T)) = P(T)P(T) \]
Joint Probability Function for **Independent** Outcomes

<table>
<thead>
<tr>
<th></th>
<th>$H_2$</th>
<th>$T_2$</th>
<th>$H_1$</th>
<th>$T_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_2$</td>
<td>$p_{H_1}p_{H_2}$</td>
<td>$p_{H_1}p_{T_2}$</td>
<td>$p_{H_1}$</td>
<td>$p_{T_1}$</td>
</tr>
<tr>
<td>$T_2$</td>
<td>$p_{T_1}p_{H_2}$</td>
<td>$p_{T_1}p_{T_2}$</td>
<td>$p_{T_1}$</td>
<td>$p_S$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$H_2$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1$</td>
<td>0.18</td>
<td>0.12</td>
</tr>
<tr>
<td>$T_1$</td>
<td>0.42</td>
<td>0.28</td>
</tr>
</tbody>
</table>

- Fill marginals first
- Joint probabilities are products of the marginals
- More generally, two events (as opposed to outcomes) $A$ and $B$ are independent iff

\[
\text{prob}(A \cap B) = \text{prob}(A) \cdot \text{prob}(B)
\]
Independent Outcomes and Events

Exclusivity and Independence

• $A, B \in \mathbb{E}$ mutually exclusive: $A \cap B = \emptyset$
  (A set-theoretical statement)
• $A, B$ independent: $\text{prob}(A \cap B) = \text{prob}(A) \text{prob}(B)$
  (A probabilistic statement)
• Two mutually exclusive events are independent only if at least one of them has zero probability:

  $$A \cap B = \emptyset \Rightarrow \text{prob}(A \cap B) = 0$$

  ... and if $A$ and $B$ are independent

  $$0 = \text{prob}(A \cap B) = \text{prob}(A) \text{prob}(B)$$

  ... so that either $\text{prob}(A) = 0$ or $\text{prob}(B) = 0$

• Two mutually exclusive events with nonzero probabilities are always dependent. Prove by contradiction:

• Assume $\text{prob}(A) > 0$, $\text{prob}(B) > 0$ and $A \cap B = \emptyset$ and $\text{prob}(A \cap B) = \text{prob}(A) \text{prob}(B)$
  Then, $\text{prob}(A \cap B) > 0$, but $\text{prob}(A \cap B) = \text{prob}(\emptyset) = 0$
Independent Events

• Two events $A$ and $B$ are independent iff

$$\text{prob}(A \cap B) = \text{prob}(A) \text{prob}(B)$$

• Example: Rolling a fair die, $p(O_i) = 1/6$ for $i = 1, \ldots, 6$
  • $\text{prob}(A) = \text{prob}({1, 3, 5}) = 1/2$, $\text{prob}(B) = \text{prob}({2, 5}) = 1/3$
  • $\text{prob}(A \cap B) = \text{prob}({5}) = 1/6 = \text{prob}(A) \text{prob}(B)$
  • $A, B$ are independent
• $\text{prob}(C) = \text{prob}({2, 3, 5}) = 1/2$
• $\text{prob}(A \cap C) = \text{prob}({3, 5}) = 1/3 \neq \text{prob}(A) \text{prob}(C) = 1/4$
• $A, C$ are dependent
• Easy condition to verify
• However, the definition seems arbitrary
• Seems to have to do uniquely with numbers
• To understand what it means, we introduce the notion of conditional probability
Conditional Probabilities

- Let $A, B \subseteq \mathcal{S}$ with $\text{prob}(B) > 0$

\[ \text{prob}(A|B) \triangleq \frac{\text{prob}(A \cap B)}{\text{prob}(B)} \]

- $\text{prob}(A)$ is the fraction of the probability mass of $A$ relative to that of $\mathcal{S}$

- $\text{prob}(A|B)$ is the fraction of the probability mass of $A$ relative to that of $B$

- $\text{prob}(A|B)$ is the probability of $A$ if the sample space shrinks from $\mathcal{S}$ to $B$
Conditional Probabilities

- $A, B \subseteq \mathbb{S}$ with $\text{prob}(B) > 0$

\[
\text{prob}(A|B) \triangleq \frac{\text{prob}(A \cap B)}{\text{prob}(B)}
\]

- Think of $\text{prob}(A)$ as a shorthand for $\text{prob}(A|\mathbb{S})$

\[
\text{prob}(A|\mathbb{S}) = \frac{\text{prob}(A \cap \mathbb{S})}{\text{prob}(\mathbb{S})} = \frac{\text{prob}(A)}{1} = \text{prob}(A)
\]
Conditioning and Independence

- $A$ and $B$ independent: $\text{prob}(A \cap B) = \text{prob}(A) \text{prob}(B)$
- Assume $\text{prob}(A) > 0$ and $\text{prob}(B) > 0$
  (otherwise $A$, $B$ are trivially independent)

\[
\text{prob}(A|B) = \frac{\text{prob}(A \cap B)}{\text{prob}(B)} = \frac{\text{prob}(A)\text{prob}(B)}{\text{prob}(B)} = \text{prob}(A)
\]

\[
\text{prob}(B|A) = \frac{\text{prob}(B \cap A)}{\text{prob}(A)} = \frac{\text{prob}(B)\text{prob}(A)}{\text{prob}(A)} = \text{prob}(B)
\]

- So if $\text{prob}(A), \text{prob}(B) > 0$ then

\[
\text{prob}(A \cap B) = \text{prob}(A) \text{prob}(B)
\]

$\iff \text{prob}(A|B) = \text{prob}(A)$

$\iff \text{prob}(B|A) = \text{prob}(B)$

- $A$ and $B$ are independent iff knowledge of one does not change the probability of the other
Example: Roll of a Die Case I

- Kate is about to roll a fair die in a separate room
- You bet on an odd outcome, $A = \{1, 3, 5\}$
- Your uncertainty is distributed over the entire sample space
- $\text{prob}(A) = 1/2$
- After the roll, I tell you that the outcome was either 2 or 5
- Your sample space has now shrunk to $B = \{2, 5\}$
- The only outcomes of $B$ that let you win are in $A \cap B = \{5\}$
- You need to compare the probability of $A \cap B$ with the total mass in $B$
- So your probability of winning after what I told you is

\[
\text{prob}(A|B) = \frac{\text{prob}(A \cap B)}{\text{prob}(B)} = \frac{\text{prob}\{\{5\}\}}{\text{prob}\{\{2, 5\}\}} = \frac{1/6}{2/6} = \frac{1}{2}
\]

- $\text{prob}(A|\Sigma)$ and $\text{prob}(A|B)$ are the same
- Knowing $B$ does not change the probability of $A$
- $A, B$ are independent
Example: Roll of a Die Case II

• Same situation, you still bet on an odd outcome, \( A = \{1, 3, 5\} \)
• Initially, \( \text{prob}(A) = 1/2 \)
• After the roll, I tell you that the outcome was either 2 or 3 or 5
• **Your sample space has now shrunk to** \( C = \{2, 3, 5\} \)
• The only outcomes of \( C \) that let you win are in \( A \cap C = \{3, 5\} \)
• You need to compare the probability of \( A \cap C \) with the total mass in \( C \)
• So your probability of winning after what I told you is

\[
\text{prob}(A|C) = \frac{\text{prob}(A \cap C)}{\text{prob}(C)} = \frac{\text{prob}(\{3, 5\})}{\text{prob}(\{2, 3, 5\})} = \frac{2/6}{3/6} = \frac{2}{3}
\]

• \( \text{prob}(A|S) \) and \( \text{prob}(A|C) \) are not the same
• Knowing \( C \) changes the probability of \( A \) (up or down)
• \( A, C \) are dependent
Interpretation of Conditional (In)dependence

- \( \text{prob}(A|\mathcal{S}) = 1/2 \) but \( \text{prob}(A|C) = 2/3 \)
- Knowledge of \( C \) has restricted the sample space from \( \mathcal{S} \) to \( C \)
- This restriction changed the probability of \( A \) occurring
- \( A \) depends on \( C \)
- \( A \) did not depend on \( B \)
- If \( A \) depends on \( C \) then \( C \) depends on \( A \)
- Proof: \( \text{prob}(A \cap C) = \text{prob}(A) \text{prob}(C) \) is symmetric in \( A \) and \( C \). Done.
Two Loaded Tetrahedral Dice

Roll two loaded tetrahedral dice $n = 1000$ times.

<table>
<thead>
<tr>
<th>Contingency Table</th>
<th>Joint Probability Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{XY}$</td>
<td>$n_Y$</td>
</tr>
<tr>
<td>1 2 3 4</td>
<td>272</td>
</tr>
<tr>
<td>1 68 63 72 69</td>
<td>253</td>
</tr>
<tr>
<td>2 67 53 72 61</td>
<td>225</td>
</tr>
<tr>
<td>3 55 55 50 65</td>
<td>250</td>
</tr>
<tr>
<td>4 69 60 66 55</td>
<td>1000</td>
</tr>
</tbody>
</table>

Probabilities are only coarse estimates because $n = 1000$. Both dice seem loaded. From marginals, $X$ die has low $p_X(2)$ and $Y$ die has low $p_Y(3)$. 
Two Loaded Tetrahedral Dice

<table>
<thead>
<tr>
<th>Joint Probability Estimate</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>(p_Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_{XY})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.068</td>
<td>.063</td>
<td>.072</td>
<td>.069</td>
<td>.272</td>
</tr>
<tr>
<td>2</td>
<td>.067</td>
<td>.053</td>
<td>.072</td>
<td>.061</td>
<td>.253</td>
</tr>
<tr>
<td>3</td>
<td>.055</td>
<td>.055</td>
<td>.050</td>
<td>.065</td>
<td>.225</td>
</tr>
<tr>
<td>4</td>
<td>.069</td>
<td>.060</td>
<td>.066</td>
<td>.055</td>
<td>.250</td>
</tr>
<tr>
<td>(p_X)</td>
<td>.259</td>
<td>.231</td>
<td>.260</td>
<td>.250</td>
<td>1</td>
</tr>
</tbody>
</table>

Are the dice independent? Multiply the marginals.

<table>
<thead>
<tr>
<th>Joint Probability Estimate</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>(p_Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_{XY})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.0704</td>
<td>.0628</td>
<td>.0707</td>
<td>.0680</td>
<td>.272</td>
</tr>
<tr>
<td>2</td>
<td>.0655</td>
<td>.0584</td>
<td>.0658</td>
<td>.0633</td>
<td>.253</td>
</tr>
<tr>
<td>3</td>
<td>.0583</td>
<td>.0520</td>
<td>.0585</td>
<td>.0563</td>
<td>.225</td>
</tr>
<tr>
<td>4</td>
<td>.0648</td>
<td>.0577</td>
<td>.0650</td>
<td>.0625</td>
<td>.250</td>
</tr>
<tr>
<td>(p_X)</td>
<td>.259</td>
<td>.231</td>
<td>.260</td>
<td>.250</td>
<td>1</td>
</tr>
</tbody>
</table>

Close, but not quite.
Outcomes from two dice seem to be slightly tied to each other. Magnets? Biased roll? Chance?