Monty Hall problem from a game show

- There are three doors, say A, B, C, behind one of them is a prize
- Player chooses one door
- Monty opens one of the doors that the player did not choose and does not have a prize, and offers the player to switch from her choice to the other remaining unrevealed door

- Should she switch or not?

- Our goal is to familiarize with the notations
- We will focus on discrete random variables, for now as rigorous treatment of continuous random variables require sigma-algebra and Borel sets, which is outside the scope of this course
Sample space, events, probability

- **sample space** \( \Omega \)
  - a set of all outcomes
  - an outcome is \( \omega \in \Omega \)
  - for example, if the player’s strategy is to switch
    \( \omega = (\text{prize A, player B, Monty C, player wins}) \)

- a **probability space** is a **sample space** \( \Omega \) and a **probability measure** \( \mu \) such that
  - \( 0 \leq \mu(\omega) \), and
  - \( \sum_{\omega \in \Omega} \mu(\omega) = 1 \)

- an **event** \( A \) is any subset of \( \Omega \), and we define
  - \( \mu(A) = \sum_{\omega \in A} \mu(\omega) \)
Random variables

- **a random variable** is a mapping from sample space $\Omega$ to some set $\mathcal{X}$ and we denote a random variable by $x \in \mathcal{X}$ or $x_i \in \mathcal{X}_i$ if there are many random variables in the problem
- for example, $x_1 \in \{A, B, C\}$ can represent the location of the prize
- $x_2 \in \{A, B, C\}$ can represent the player’s initial choice
- $x_3 \in \{A, B, C\}$ can represent the door Monty opens
- $x_4 \in \{0,1\}$ can represent whether the player wins or not

- The event $\omega = \text{(prize A, player B, Monty C, player does not switch)}$ is then mapped to a **random vector** $x = (x_1, x_2, x_3, x_4) = (A, B, C, 0)$

- $\mu$ in the definition of the original probability space induces a **probability distribution** of the random variable/vector, denoted by $\mu(x)$
- one should visualize $\mu(x)$ as a table with $3 \times 3 \times 3 \times 2$ non-negative entries that sum to one
Probability distribution

- consider the prize being place randomly, then the induced probability distribution is
  \[ \mu(x_1 = A) = \mu(x_1 = B) = \mu(x_1 = C) = 1/3 \]
- consider the player choosing randomly,
  \[ \mu(x_2 = A) = \mu(x_2 = B) = \mu(x_2 = C) = 1/3 \]
- for the random vector \( x = (x_1, x_2, x_3, x_4) \), the joint probability distribution is denoted by \( \mu(x) = \mu(x_1, x_2, x_3, x_4) \)
  - for example, if the player’s strategy is not to switch, then
    \[ \mu(A, B, C, 0) = 1/9 \]
    \[ \mu(A, B, C, 1) = 0 \]
    \[ \mu(A, A, B, 1) = 1/18 \]
    \[ \mu(A, A, C, 1) = 1/18 \]
Probability distribution

- given a joint distribution $\mu(x_1, \ldots, x_n)$, a marginal distribution of one random variable $x_1$ is defined as
  \[
  \mu(x_1) = \sum_{x_2, x_3, \ldots, x_n} \mu(x_1, \ldots, x_n)
  \]
  and the process of computing such marginal is referred to as marginalizing out $(x_2, \ldots, x_n)$

- similarly,
  \[
  \mu(x_1, x_2) = \sum_{x_3, \ldots, x_n} \mu(x_1, \ldots, x_n)
  \]

- for example, to decide whether we should switch or not, we need to compute the marginal distribution of $\mu(x_4 = 1)$ for the player who switches, and for the player who does not switch
Conditional probability distribution

• for the player who does not switch,
  • one way to compute the marginal is to enumerate all possible outcomes:  
    \[ \mu(x_4 = 1) = \mu(A, A, A, 1) + \mu(A, A, B, 1) + \cdots + \mu(C, C, C, 1) \]
  • a simpler approach is to notice that
    \[ \mu(x_4 = 1) = \mu(x_1 = x_2) = \mu_{12}(A, A) + \mu_{12}(B, B) + \mu_{12}(C, C) \]
    where \( \mu_{12} \) denotes the second order marginal distribution of \( \mu(x_1, x_2) \)
    and this is \( 1/9+1/9+1/9=1/3 \)

• for the player who does switch,
  • again, one could enumerate
  • a simpler approach is to notice that
    \[ \mu(x_4 = 1) = \mu(x_1 \neq x_2) \]
    and as \( \mu(x_1 \neq x_2) = 1 - \mu(x_1 = x_2) \), we know that it is \( 2/3 \)
Independence

• Two random variables are independent
  • if $\mu(x_1, x_2) = \mu(x_1)\mu(x_2)$ for all $x_1, x_2$
  • for example, the location of the prize $x_1$ and the initial choice of the player $x_2$ are independent, as neither knows the other
  • one can easily check that $\mu(x_1, x_2) = \mu(x_1)\mu(x_2)$

• We denote independence by $x_1 \perp x_2$

• For the example, if $x_1, x_2, x_3, x_4$ are mutually independent, then we only need $2 + 2 + 2 + 1$ entries to store $\mu(x)$, as it decomposes (or factorizes) as $\mu(x) = \mu(x_1)\mu(x_2)\mu(x_3)\mu(x_4)$, and each function can be stored with 2,2,2,1 values respectively (because probability sum to one, we don’t need to store the last entry in each of the functions)
• Independence give efficiency (in storing and also in computing, as we will see)
• in practice, independent random variables are rare, but conditional independence is abundant
Conditional probability distribution

- conditional probability of $x_2$ given $x_1$ is defined as
  \[ \mu(x_2 \mid x_1) = \frac{\mu(x_1, x_2)}{\mu(x_1)} \]

- for example, $\mu(x_2 \mid x_1) = \mu(x_2) = 1/3$ as $x_1 \perp x_2$

\[ \begin{align*}
  \mu(x_1 = A, x_2 = A \mid x_3 = C) &= \frac{\mu(x_1 = A, x_2 = A, x_3 = C)}{\mu(x_3 = C)} = \frac{1/18}{1/3} = 1/6 \\
  \mu(x_1 = B, x_2 = A \mid x_3 = C) &= \frac{\mu(x_1 = B, x_2 = A, x_3 = C)}{\mu(x_3 = C)} = \frac{1/9}{1/3} = 1/3 \\
  \mu(x_2 = A \mid x_1 = B, x_3 = B) &= \frac{\mu(x_2 = A, x_1 = B, x_3 = B)}{\mu(x_1 = B, x_3 = B)} = \frac{0}{0} = 0
\end{align*} \]

- note that $\mu(x_1, x_2) = \mu(x_1)\mu(x_2 \mid x_1)$

- which gives the chain rule:
  \[
  \mu(x_1, x_2, \ldots, x_n) = \mu(x_1, \ldots, x_{n-1})\mu(x_n \mid x_1, \ldots, x_{n-1}) \\
  = \mu(x_1, \ldots, x_{n-1})\mu(x_{n-1} \mid x_{n-2})\mu(x_n \mid x_{n-1}) \\
  = \prod_{i=1}^{n} \mu(x_i \mid x_{i-1})
  \]

where $x_i^j = (x_i, x_{i+1}, \ldots, x_j)$ and $x_1^0 = x_1$

One interpretation of the conditional probability is that the probability that the prize is not behind my chosen door is larger.
Conditional independence

Two random variables are **conditionally independent**

- if \( \mu(x_1, x_2 \mid x_3) = \mu(x_1 \mid x_3)\mu(x_2 \mid x_3) \) for all \( x_1, x_2, x_3 \)
- this notion captures many real-life scenarios
- We denote it by \( x_1 \perp x_2 \mid x_3 \)

For example, consider 4 random variables

\( x_1 = \) weather, \( x_2 = \) cavity, \( x_3 = \) toothache, \( x_4 = \) catch

- Weather in \{sunny, rain, cloudy, snow\}
- Cavity in \{0,1\}, Toothache in \{0,1\}, Catch in \{0,1\}
- It is clear that weather is independent of any other variables
- The joint distribution \( \mu(x_2, x_3, x_4) \) can be represented by a table as

<table>
<thead>
<tr>
<th>toothache</th>
<th>¬ toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>catch</td>
<td>¬ catch</td>
</tr>
<tr>
<td>cavity</td>
<td>.108</td>
</tr>
<tr>
<td>¬ cavity</td>
<td>.016</td>
</tr>
</tbody>
</table>

- Note that it requires \( 2 \times 2 \times 2 - 1 \) numbers to store this table
- But we know that catch is independent of toothache, conditioned on cavity
- This can be confirmed via

\[ \mu(\text{toothache}, \text{catch} | \text{cavity}) = \mu(\text{toothache} | \text{cavity})\mu(\text{catch} | \text{cavity}) \]

- One implication of such conditional independence structure is that by the chain rule,

\[ \mu(x_1, x_2, x_3) = \mu(x_1)\mu(x_2, x_3 | x_1) \]

Which requires \(1 + 2 \times 3\) numbers to store, but by the conditional independence, we have

\[ \mu(x_1, x_2, x_3) = \mu(x_1)\mu(x_2 | x_1)\mu(x_3 | x_1) \]

This only requires \(1+2+2\) numbers to store

There can be significant efficiency gain in using the conditional independence structure