CS5824: Advanced Machine Learning

Dawei Zhou

CS, Virginia Tech

Please keep your face covering on!

COMPUTER SCIENCE
VIRGINIA TECH

Basics on Probability

Coin Flips

- You flip a coin with one side labeled head (H) and the other labeled tail (T)
  - Experiment: the action of tossing coins
  - Outcome space: {H, T}
  - Event: any subset of the outcome space, e.g., {H}
  - P(Event): the probability of an event occurring
- You flip 100 coins
  - Head with probability 0.5
  - How many heads would you expect?

Coin Flips cont.

- You flip a coin
  - Head with probability p
  - Binary random variable
  - Bernoulli trial with success probability p
- You flip k coins
  - How many heads would you expect
  - Number of heads X: discrete random variable
  - Binomial distribution with parameters k and p

Discrete Random Variables

- Random variables (RVs) which may take on only a **countable** number of **distinct** values
  - E.g., the total number of heads X you get if you flip 100 coins

- X is a RV with arity *k* if it can take on exactly one value out of $\{x_1, \cdots, x_k\}$,
  - E.g., the possible values that X can take are 0, 1, 2,…, 100

Probability of Discrete RV

- Probability mass function (pmf): $\mathrm{P}\left(\mathrm{X}=x_i\right)$
- Easy facts about pmf
  - $\mathrm{P}\left(\mathrm{X}=x_i\right) \in [0,1]$
  - $\sum_i \mathrm{P}\left(\mathrm{X}=x_i\right)=1$
  - $\mathrm{P}\left(\mathrm{X}=x_i \cap \mathrm{X}=x_j\right)=0$ if $i \neq j$
  - $\mathrm{P}\left(\mathrm{X}=x_i \cup \mathrm{X}=x_j\right)=\mathrm{P}\left(\mathrm{X}=x_i\right)+\mathrm{P}\left(\mathrm{X}=x_j\right)$ if $i \neq j$

Common Distributions

- Uniform $X \sim U[1,\ldots,N]$
  - X takes values 1, 2, …, *N*
  - $P(X = i) = 1/N$
  - E.g., picking balls of different colors from a box
- Binomial $X \sim Bin(n, p)$
  - X takes values 0, 1, …, *n*
  - $P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$
  - E.g., coin flips

Coin Flips of Two Persons

- Your friend and you both flip coins
  - Head with probability 0.5
  - You flip 50 times; your friend flip 100 times
  - How many heads will both of you get

Joint Distribution

- Given two discrete RVs X and Y, their **joint distribution** is the distribution of X and Y together
  - E.g., P(You get 21 heads AND you friend get 70 heads)
- $\sum_x \sum_y \mathrm{P}\left(\mathrm{X}=x \cap \mathrm{Y}=y\right)=1$
  - E.g.,

$$\sum_{i=0}^{50} \sum_{j=0}^{100} \mathrm{P}\left(\text{You get } i \text{ heads AND your friend get } j \text{ heads}\right)=1$$

Conditional Probability

- $P\left(X = x \middle| Y = y\right)$ is the probability of $X = x$, given the occurrence of $Y = y$
  - E.g., you get 0 heads, given that your friend gets 61 heads
- $P\left(X = x \middle| Y = y\right) = \dfrac{P\left(X = x \cap Y = y\right)}{P\left(Y = y\right)}$

Law of Total Probability

- Given two discrete RVs X and Y, which take values in $\{x_1, \cdots, x_m\}$ and $\{y_1, \cdots, y_n\}$, we have

$$\begin{aligned}\mathrm{P}\left(\mathrm{X}=x_i\right) &= \sum_j \mathrm{P}\left(\mathrm{X}=x_i \cap \mathrm{Y}=y_j\right) \\ &= \sum_j \mathrm{P}\left(\mathrm{X}=x_i \middle| \mathrm{Y}=y_j\right)\mathrm{P}\left(\mathrm{Y}=y_j\right)\end{aligned}$$

Marginalization

Marginal Probability

Joint Probability

$$\mathrm{P}\left(\mathrm{X}=x_i\right)=\sum_j \mathrm{P}\left(\mathrm{X}=x_i \cap \mathrm{Y}=y_j\right)$$

$$=\sum_j \mathrm{P}\left(\mathrm{X}=x_i \middle| \mathrm{Y}=y_j\right)\mathrm{P}\left(\mathrm{Y}=y_j\right)$$

Conditional Probability

Marginal Probability

Bayes Rule

- X and Y are discrete RVs…

$$P\left(X = x \middle| Y = y\right) = \frac{P\left(X = x \cap Y = y\right)}{P\left(Y = y\right)}$$

$$P\left(X = x_i \middle| Y = y_j\right) = \frac{P\left(Y = y_j \middle| X = x_i\right) P\left(X = x_i\right)}{\sum_k P\left(Y = y_j \middle| X = x_k\right) P\left(X = x_k\right)}$$

Independent RVs

- Intuition: X and Y are independent means that $X = x$ **neither** makes it **more or less** probable that $Y = y$
- Definition: X and Y are independent iff

$$P\left(X = x \cap Y = y\right) = P\left(X = x\right)P\left(Y = y\right)$$

More on Independence

- $$\mathrm{P}\left(\mathrm{X}=x\cap\mathrm{Y}=y\right)=\mathrm{P}\left(\mathrm{X}=x\right)\mathrm{P}\left(\mathrm{Y}=y\right)$$

$$\mathrm{P}\left(\mathrm{X}=x\middle|\mathrm{Y}=y\right)=\mathrm{P}\left(\mathrm{X}=x\right) \qquad \mathrm{P}\left(\mathrm{Y}=y\middle|\mathrm{X}=x\right)=\mathrm{P}\left(\mathrm{Y}=y\right)$$

- E.g., no matter how many heads you get, your friend will not be affected, and vice versa

Conditionally Independent RVs

- Intuition: X and Y are conditionally independent given Z means that once Z is **known**, the value of X does not add any **additional** information about Y
- Definition: X and Y are conditionally independent given Z iff

$$\mathrm{P}\left(\mathrm{X}=x \cap \mathrm{Y}=y \middle| \mathrm{Z}=z\right)=\mathrm{P}\left(\mathrm{X}=x \middle| \mathrm{Z}=z\right)\mathrm{P}\left(\mathrm{Y}=y \middle| \mathrm{Z}=z\right)$$

More on Conditional Independence

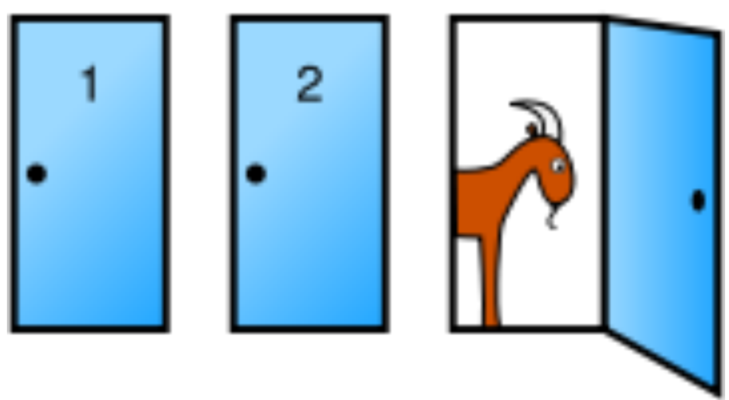
$$\mathrm{P}\left(\mathrm{X}=x \cap \mathrm{Y}=y \middle| \mathrm{Z}=z\right)=\mathrm{P}\left(\mathrm{X}=x \middle| \mathrm{Z}=z\right)\mathrm{P}\left(\mathrm{Y}=y \middle| \mathrm{Z}=z\right)$$

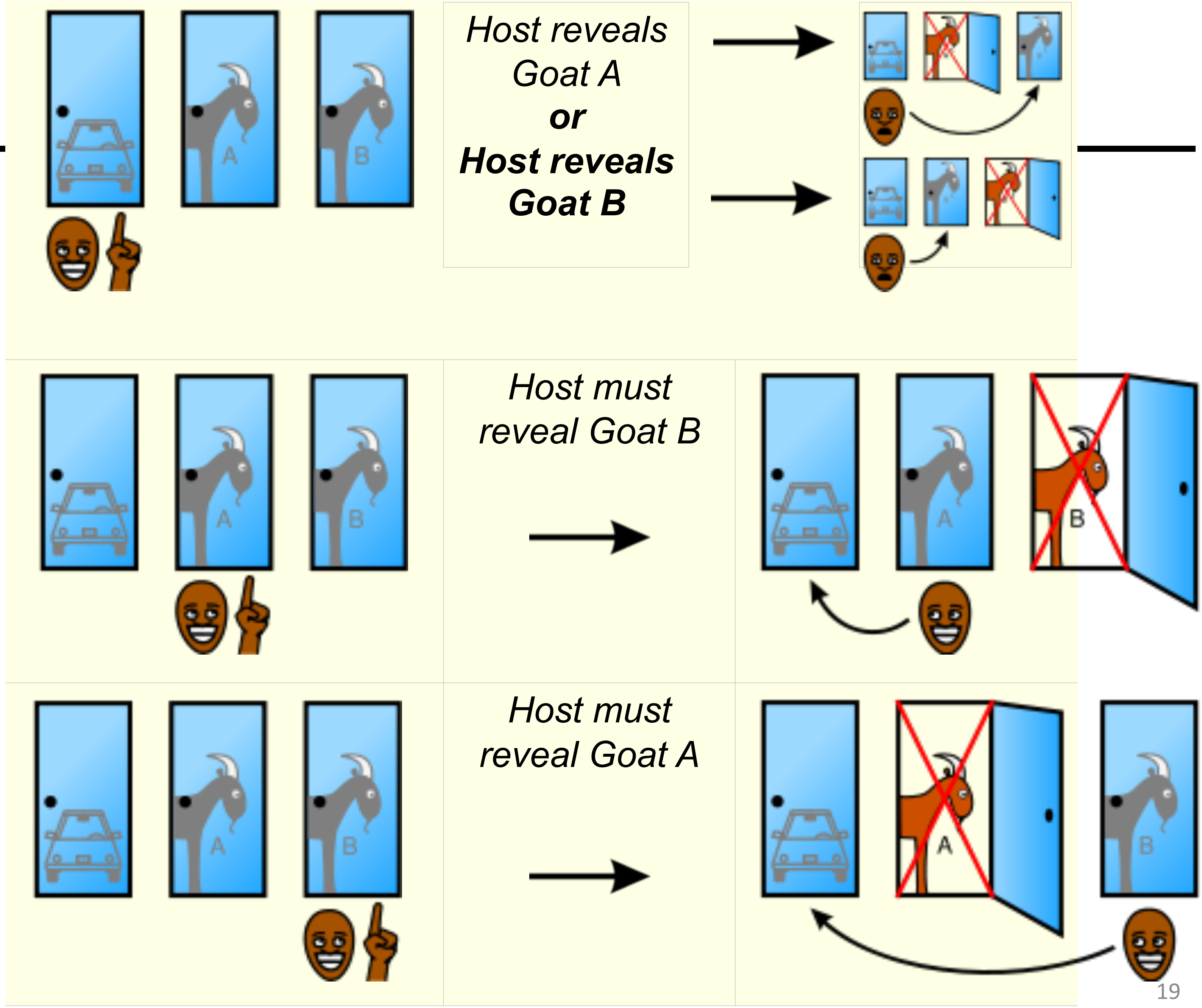
$$\mathrm{P}\left(\mathrm{X}=x \middle| \mathrm{Y}=y, \mathrm{Z}=z\right)=\mathrm{P}\left(\mathrm{X}=x \middle| \mathrm{Z}=z\right)$$

$$\mathrm{P}\left(\mathrm{Y}=y \middle| \mathrm{X}=x, \mathrm{Z}=z\right)=\mathrm{P}\left(\mathrm{Y}=y \middle| \mathrm{Z}=z\right)$$

Monty Hall Problem

- You're given the choice of three doors: Behind one door is a car; behind the others, goats.
- You pick a door, say No. 1
- The host, who knows what's behind the doors, opens another door, say No. 3, which has a goat.
- Do you want to pick door No. 2 instead?




Host reveals Goat A
or
Host reveals Goat B
A
B
Host must reveal Goat B
Host must reveal Goat A

Monty Hall Problem: Bayes Rule

- C_k : the car is behind door *k*, *k* = 1, 2, 3
- $P(C_k) = 1/3$
- H_{ij} : the host opens door *j* after you pick door *i*

- $P(H_{ij} | C_k) = \begin{cases} 0 & i = j \\ 0 & j = k \\ 1/2 & i = k \\ 1 & i \neq k, j \neq k \end{cases}$

Monty Hall Problem: Bayes Rule cont.

- WLOG, *i*=1 (your choice), *j*=3 (the host's choice)
- $P\left(C_1 \middle| H_{13}\right) = \dfrac{P\left(H_{13} \middle| C_1\right) P\left(C_1\right)}{P\left(H_{13}\right)}$
- $P\left(H_{13} \middle| C_1\right) P\left(C_1\right) = \dfrac{1}{2} \cdot \dfrac{1}{3} = \dfrac{1}{6}$

Monty Hall Problem: Bayes Rule cont.

- $P(H_{13}) = P(H_{13}, C_1) + P(H_{13}, C_2) + P(H_{13}, C_3)$

$$= P(H_{13} | C_1) P(C_1) + P(H_{13} | C_2) P(C_2)$$

$$= \frac{1}{6} + 1 \cdot \frac{1}{3}$$

$$= \frac{1}{2}$$

- $P(C_1 | H_{13}) = \dfrac{1/6}{1/2} = \dfrac{1}{3}$

Monty Hall Problem: Bayes Rule cont.

- $P\left(C_1 \middle| H_{13}\right) = \frac{1/6}{1/2} = \frac{1}{3}$
- $P\left(C_2 \middle| H_{13}\right) = 1 - \frac{1}{3} = \frac{2}{3} > P\left(C_1 \middle| H_{13}\right)$
- *You should switch!*