Lec 2 - Probability Theory Review

Mustafa Elattar

Definition of Probability

- ***Experiment***: toss a coin twice

- ***Sample space***: possible outcomes of an experiment
  - S = {HH, HT, TH, TT}

- ***Event***: a subset of possible outcomes
  - A={HH}, B={HT, TH}

- ***Probability of an event***: a number assigned to an event Pr(A)
  - Axiom 1: Pr(A) $\geq$ 0
  - Axiom 2: Pr(S) = 1
  - Axiom 3: For every sequence of disjoint events

$$\Pr(\bigcup_i A_i) = \sum_i \Pr(A_i)$$

  - Example: Pr(A) = n(A)/N: frequentist statistics
    (If we repeat an experiment N times, and denote by n(A) the number of times we observe A, then Pr(A) = n(A)/N)

Joint Probability

- For events A and B, **joint probability** Pr(AB) (or Pr(A, B)) stands for the probability that both events happen

- Example: What is the probability that the first toss is H and the second toss is H?

  Pr(1st is H and 2nd is H)=Pr(1st is H)Pr(2nd is H)=0.5x0.5=0.25

- Example: A={HH}, B={HT, TH}, what is the joint probability Pr(AB)?

  Answer: 0

Independence

- Two events ***A and B are independent*** if

$$\Pr(AB) = \Pr(A)\Pr(B)$$

- A set of events $\{A_i\}$ are independent if

$$\Pr\left(\bigcap_i A_i\right) = \prod_i \Pr(A_i)$$

- Example: Drug test

	Women	Men
Success	200	1800
Failure	1800	200

A = {Patient is a Woman}

B = {Drug fails}

Is event A independent of event B?

- Pr(AB)=1800/4000, Pr(A)=2000/4000, Pr(B)=2000/4000

$\because$ Pr(AB) ≠ Pr(A)Pr(B) → A and B are dependent

Conditioning

- If A and B are events with Pr(A) > 0, the ***conditional probability of B given A*** is

$$\Pr(B \mid A) = \frac{\Pr(AB)}{\Pr(A)}$$

- Example: Drug test

	Women	Men
Success	200	1800
Failure	1800	200

A = {Patient is a Woman}

B = {Drug fails}

Pr(B|A) = ?

Pr(A|B) = ?

- Pr(B|A)=Pr(AB)/Pr(A)=(1800/4000)/(2000/4000)=0.9
- Pr(A|B)=Pr(AB)/Pr(B)=(1800/4000)/(2000/4000)=0.9

- Given A is independent from B, what is the relationship between Pr(A|B) and Pr(A)?

  Pr(A|B) = Pr(A)

Bayes' Rule

- Given two events A and B and suppose that Pr(A) > 0, then

$$\Pr(B \mid A) = \frac{\Pr(AB)}{\Pr(A)} = \frac{\Pr(A \mid B)\Pr(B)}{\Pr(A)}$$

- Example:

$\Pr(R) = 0.8$

$\Pr(W \mid R)$	R	$\neg R$
W	0.7	0.4
$\neg W$	0.3	0.6

R: It is a rainy day

W: The grass is wet

$\Pr(R \mid W)$ = ?

$$\Pr(R \mid W) = \frac{\Pr(W \mid R)\Pr(R)}{\Pr(W)} = \frac{0.7 \times 0.8}{\Pr(W)}$$

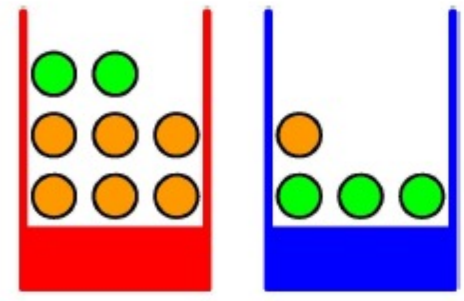
$$\Pr(W) = \Pr(WR) + \Pr(W \neg R) = \Pr(W \mid R)\Pr(R) + \Pr(W \mid \neg R)\Pr(\neg R)$$

$$= 0.7 \times 0.8 + 0.4 \times 0.2 = 0.64$$

Nile University

Bayes' Rule

- Example:
  Two boxes: one red and one blue
  Two kinds of fruit: apples and oranges
  Task: randomly pick one of the boxes
  and then select a fruit



  Let B represent the box (B = r or B = b)
  Let F represent the fruit (F = a or F = o)

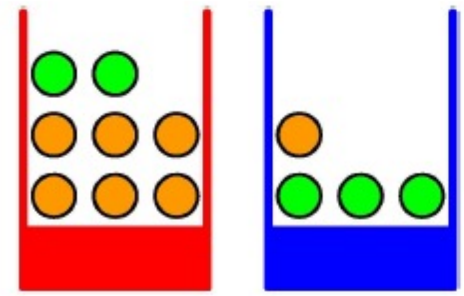
  Let the probability of picking the red box be 40% and the blue box be 60%

$$
\begin{aligned}
p(B = r) &= 4/10 \\
p(B = b) &= 6/10
\end{aligned}
$$

Bayes' Rule

- Conditional Probabilities

$$
\begin{aligned}
p(F=a|B=r) &= 1/4 \\
p(F=o|B=r) &= 3/4 \\
p(F=a|B=b) &= 3/4 \\
p(F=o|B=b) &= 1/4.
\end{aligned}
$$



- What is the probability of picking an apple? ($p(F = a)$)

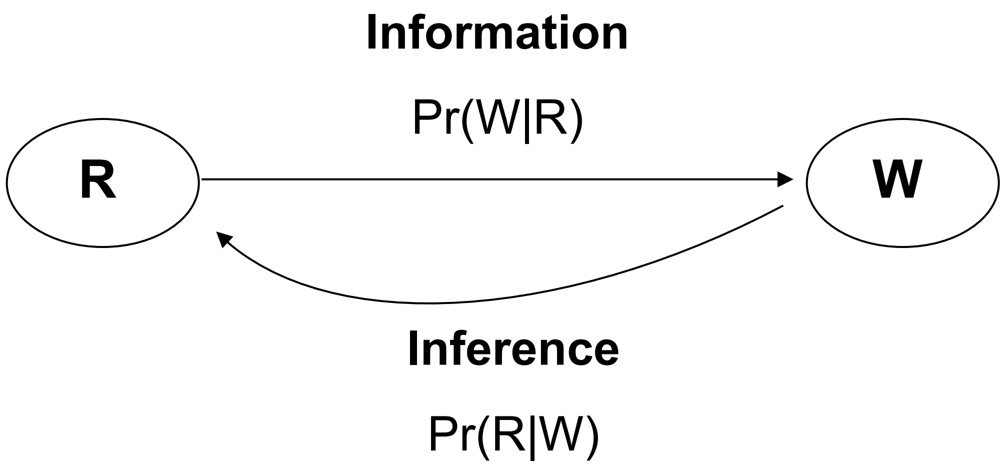
$$
\begin{aligned}
p(F=a) &= p(F=a|B=r)p(B=r)+p(F=a|B=b)p(B=b) \\
&= \frac{1}{4}\times\frac{4}{10}+\frac{3}{4}\times\frac{6}{10}=\frac{11}{20}
\end{aligned}
$$

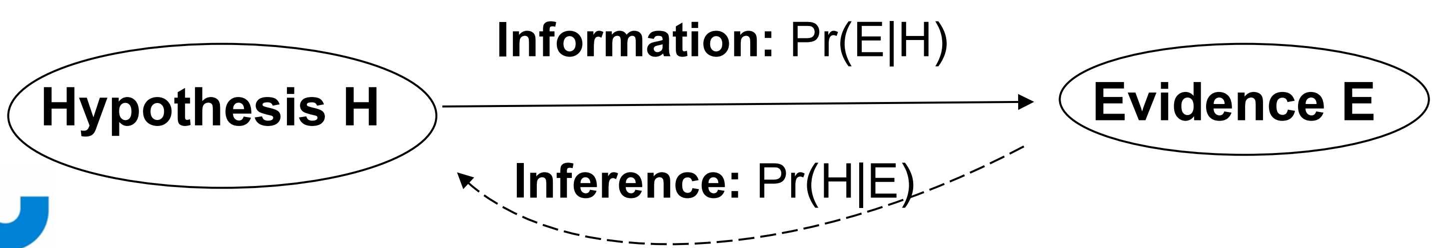
Bayes' Rule

	R	¬R
W	0.7	0.4
¬W	0.3	0.6

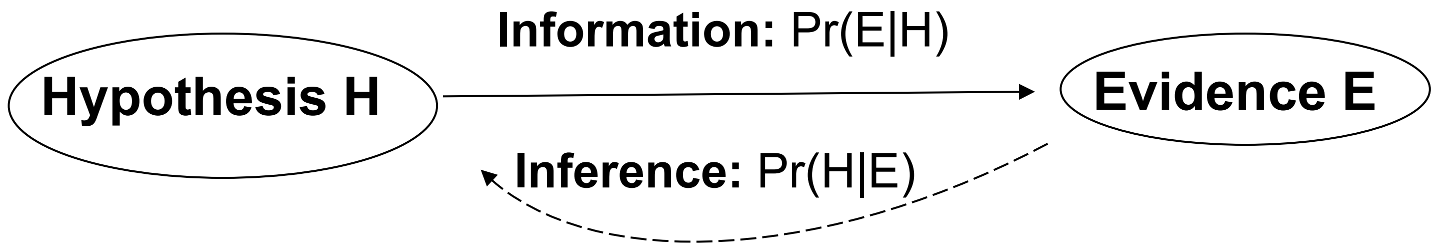
R: It rains

W: The grass is wet





Bayes' Rule

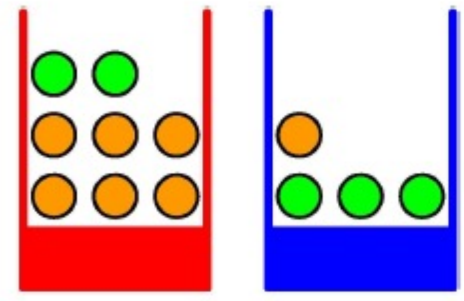


$$\underbrace{\Pr(H \mid E)}_{\text{Posterior}} = \frac{\overbrace{\Pr(E \mid H)}^{\text{Likelihood}}\ \overbrace{\Pr(H)}^{\text{Prior}}}{\Pr(E)}$$

- Prior Probability: Probability available ***before*** observing the evidence
- Posterior Probability: Probability obtained ***after*** observing the evidence

Bayes' Rule

- $p(B)$: Prior probability: Probability available before observing the identity of the fruit



- $p(B|F)$: Posterior probability:
  Probability obtained after observing the picked fruit

- Since $p(B = r)$ is 0.4, we are more likely to pick the blue box

- However, once we observe that the picked fruit is an orange we find that it's more likely that we picked from the red box (posterior probability)

$$p(B = r|F = o) = \frac{p(F = o|B = r)p(B = r)}{p(F = o)} = \frac{3}{4} \times \frac{4}{10} \times \frac{20}{9} = \frac{2}{3}$$

Bayes' Rule: Extended form

- Suppose that $B_1, B_2, \ldots B_k$ form a partition of S:

$$B_i \bigcap B_j = \varnothing; \quad \bigcup_i B_i = S$$

Suppose that $\Pr(B_i) > 0$ and $\Pr(A) > 0$. Then

$$\begin{aligned}
\Pr\left(B_i \middle| A\right) &= \frac{\Pr\left(A \middle| B_i\right)\Pr\left(B_i\right)}{\Pr\left(A\right)} \\
&= \frac{\Pr\left(A \middle| B_i\right)\Pr\left(B_i\right)}{\sum_{j=1}^{k} \Pr\left(AB_j\right)} \\
&= \frac{\Pr\left(A \middle| B_i\right)\Pr\left(B_i\right)}{\sum_{j=1}^{k} \Pr\left(B_j\right)\Pr\left(A \middle| B_j\right)}
\end{aligned}$$

Conditional Independence

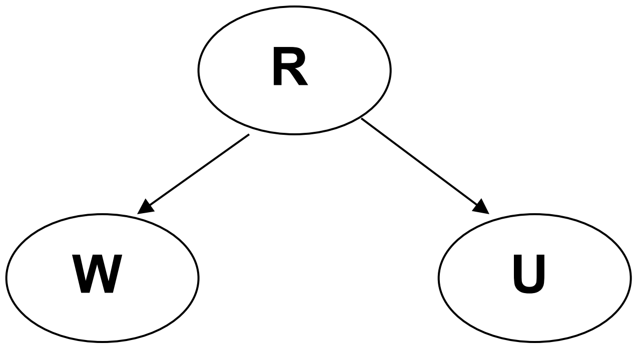
- Event A and B are ***conditionally independent given C*** if

$$Pr(AB|C)=Pr(A|C)Pr(B|C)$$

- A set of events $\{A_i\}$ are conditionally independent given C if

$$\Pr(\bigcap A_i \mid C) = \prod_i \Pr(A_i \mid C)$$

A More Complicated Example



R It rains

W The grass is wet

U People bring umbrella

U and W are conditionally independent given R →

$$\Pr(UW|R)=\Pr(U|R)\Pr(W|R)$$
$$\Pr(UW|\neg R)=\Pr(U|\neg R)\Pr(W|\neg R)$$

Pr(R) = 0.8

Pr(W|R)	R	¬R
W	0.7	0.4
¬W	0.3	0.6

Pr(U|R)	R	¬R
U	0.9	0.2
¬U	0.1	0.8

$$\Pr(U|W)=\frac{\Pr(UW)}{\Pr(W)}$$

$$\Pr(U|W)=?$$

$$\Pr(W)=\Pr(W|R)\Pr(R)+\Pr(W|\neg R)\Pr(\neg R)=0.7\times 0.8+0.4\times 0.2=0.64$$

$$\begin{aligned}\Pr(UW)&=\Pr(UW|R)\Pr(R)+\Pr(UW|\neg R)\Pr(\neg R)\\&=\Pr(U|R)\Pr(W|R)\Pr(R)+\Pr(U|\neg R)\Pr(W|\neg R)\Pr(\neg R)\\&=0.9\times 0.7\times 0.8+0.2\times 0.4\times 0.2=0.52\end{aligned}$$

$$\Pr(U|W)=\frac{0.52}{0.64}=0.81$$

Break

Random Variable and Distribution

- A ***random variable*** X is a numerical outcome of a random experiment

- The ***distribution*** of a random variable is the collection of possible outcomes along with their probabilities:

  – Discrete case: $\Pr(X = x) = p_\theta(x)$

  – Continuous case: $\Pr(a \le X \le b) = \int_a^b p_\theta(x)dx$

  θ represents the parameter(s) of the distribution

Random Variable: Example and Distribution

- Let S be the set of all sequences of three rolls of a dice. Let X be the sum of the number of dots on the three rolls

- What are the possible values of X?

  Answer: 3, 4, 5, 6, …, 18

- $\Pr(X = 5) = ?$

  To get $X = 5$: (1,1,3), (1,3,1), (3,1,1),

  (1,2,2), (2,1,2), (2,2,1)

  $\Pr(X = 5) = 6/6^3$

Expectation

- A random variable X given $\Pr(X = x)$. Then, its expectation is

$$E[X] = \sum_x x \Pr(X = x)$$

  - In an empirical sample, $x_1, x_2, \ldots, x_N$, the sample mean is

$$E[X] = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- Continuous case: $$E[X] = \int_{-\infty}^{\infty} x p_\theta(x) dx$$

- Expectation of sum of random variables

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

- Expectation of product of random variables

$$E[X_1 X_2] = \sum_{x_1} \sum_{x_2} x_1 x_2 \Pr(X_1 = x_1, X_2 = x_2)$$

If X_1 and X_2 are independent, $\Pr(X_1 = x_1, X_2 = x_2) = \Pr(X_1 = x_1)\Pr(X_2 = x_2)$

$$\therefore E[X_1 X_2] = \sum_{x_1} x_1 \Pr(X_1 = x_1) \sum_{x_2} x_2 \Pr(X_2 = x_2) = E[X_1]E[X_2]$$

Expectation: Example

- Let S be the set of all sequences of three rolls of a dice. Let X be the sum of the number of dots on the three rolls.

- What is $E[X]$?

  Answer: $X=X_1+X_2+X_3$

  $$E[X]= E[X_1] + E[X_2] +E[X_3]$$

  $$E[X_i]=\sum_{x_i} x_i \Pr(X_i = x_i)$$

  $$=1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=\frac{21}{6}$$

  $$\therefore E[X]=\frac{21}{6}+\frac{21}{6}+\frac{21}{6}=10.5$$

Expectation: Example

- Let S be the set of all sequences of three rolls of a die. Let X be the product of the number of dots on the three rolls

- What is $E[X]$?

  Answer: $X=X_1X_2X_3$

  Since the three rolls are independent, then

$$E[X]=E[X_1]E[X_2]E[X_3]$$

$$= \left(\frac{21}{6}\right)^3$$

Variance

- The variance of a random variable measures how much variability there is in the random variable

$$
\begin{aligned}
Var(X) &= E((X - E[X])^2) \\
&= E(X^2 + E[X]^2 - 2XE[X]) \\
&= E(X^2 - E[X]^2) \\
&= E[X^2] - E[X]^2
\end{aligned}
$$

- Population variance: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$, where μ is the sample mean

- Example

  Let X represent randomly sampled integer numbers in the range (1, 100):

  - {2, 50, 9, 4, 23, 65, 99} → $\sigma^2 = \frac{1}{7}\sum_{i=1}^{7}(x_i - 36)^2 = 1154.85$

  - {33, 34, 35, 36, 37, 38, 39} → $\sigma^2 = \frac{1}{7}\sum_{i=1}^{7}(x_i - 36)^2 = 4$

Covariance

- The covariance of two random variables measures the extent to which the two variables vary together

$$
\begin{aligned}
Cov[X,Y] &= E[(X-E[X])(Y-E[Y])] \\
&= E[XY - XE[Y] - YE[X] + E[X]E[Y]] \\
&= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\
&= E[XY] - E[X]E[Y]
\end{aligned}
$$

- Sample covariance $\quad C_{XY} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_X)(y_i - \mu_Y)$
- Example