Review of Probability

Index

- Axioms of probability theory
- Conditional Probability.
- Independence, Mutual Exclusion and Exhaustive sets of events
- random variables- discrete and continuous
- Joint and Conditional Probabilities
- Law of Total Probability
- Marginalization Rules.

Simple Probability

Simple Probability

If there are n elementary events associated with a random experiment and m of n of them are favorable to an event A, then the probability of happening or occurrence of A is

$$P(A) = \frac{m}{n}$$

Axioms of Probability Theory

Pr*(A)* denotes probability that proposition *A* is true. (A is also called event, or random variable).

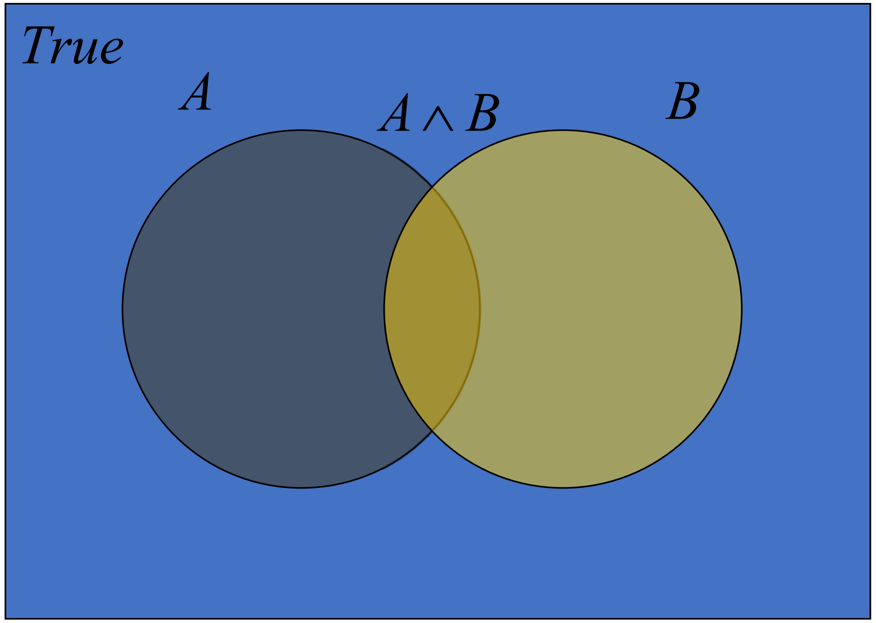
1. $$0 \le \Pr(A) \le 1$$

2. $$\Pr(True) = 1 \qquad\qquad \Pr(False) = 0$$

3. $$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \wedge B)$$

A Closer Look at Axiom 3

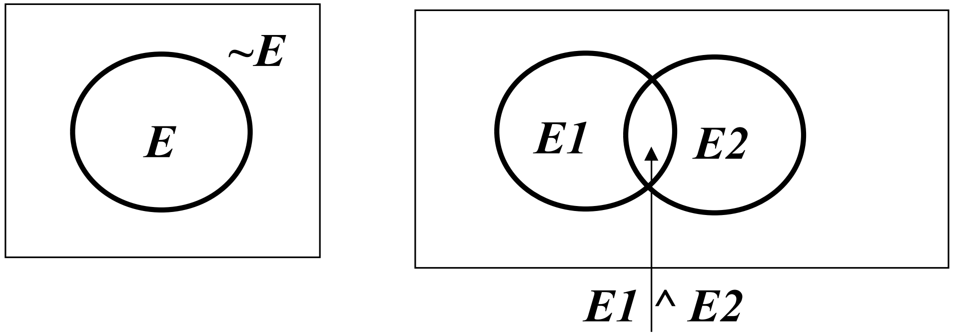
$$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \wedge B)$$



Probability of Events

- Sample space and events
  - Sample space *S*: (e.g., all people in an area)
  - Events $E1 \subseteq S$: (e.g., all people having cough)
    $E2 \subseteq S$: (e.g., all people having cold)
- Boolean operators between events (to form compound events)
  - Conjunctive (intersection): E1 ^ E2 ($E1 \cap E2$)
  - Disjunctive (union): E1 v E2 ($E1 \cup E2$)
  - Negation (complement): ~E ($E^C = S - E$)

- **Probabilities of compound events**
  - P(~E) = 1 – P(E) because P(~E) + P(E) =1
  - P(E1 v E2) = P(E1) + P(E2) – P(E1 ^ E2)
  - But how to compute the *joint probability* P(E1 ^ E2)?



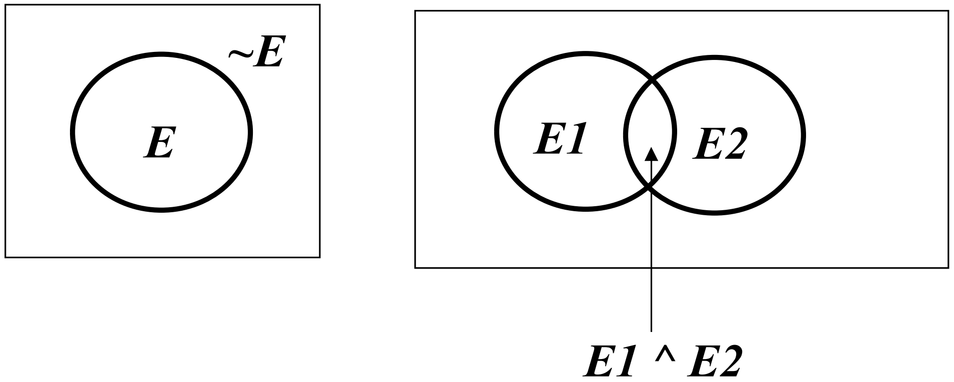
- **Conditional probability** (of E1, given E2)
  - How likely E1 occurs in the subspace of E2

$$P(E1 | E2) = \frac{|E1 \wedge E2|}{|E2|} = \frac{|E1 \wedge E2| / |S|}{|E2| / |S|} = \frac{P(E1 \wedge E2)}{P(E2)}$$

$$P(E1 \wedge E2) = P(E1 | E2) P(E2)$$

Using Venn diagrams and decision trees is very useful in proofs and reasonings

The main thing to remember for Bayes



$$P(E1 \mid E2) = \frac{|E1 \wedge E2|}{|E2|} = \frac{|E1 \wedge E2| / |S|}{|E2| / |S|} = \frac{P(E1 \wedge E2)}{P(E2)}$$

$$P(E1 \wedge E2) = P(E1 \mid E2)P(E2)$$

Independence, Mutual Exclusion and Exhaustive sets of events

- Independence assumption
  - Two events E1 and E2 are said to be independent of each other if

$P(E1 \mid E2) = P(E1)$ (given E2 does not change the likelihood of E1)

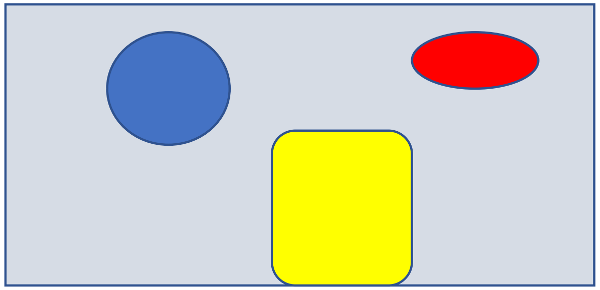
  - It can simplify the computation

$$P(E1 \wedge E2) = P(E1 \mid E2)P(E2) = P(E1)P(E2)$$

$$\begin{aligned} P(E1 \vee E2) &= P(E1) + P(E2) - P(E1 \wedge E2) \\ &= P(E1) + P(E2) - P(E1)P(E2) \\ &= 1 - (1 - P(E1)(1 - P(E2)) \end{aligned}$$

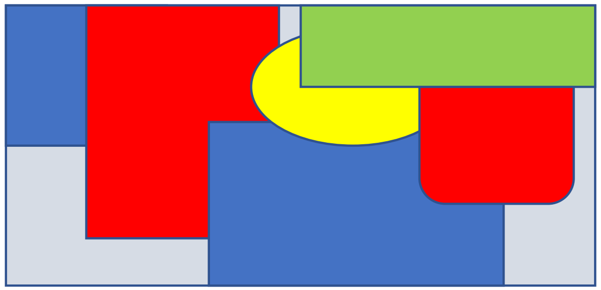
- Mutually exclusive (ME) and exhaustive (EXH) set of events
  - ME: $E_i \wedge E_j = \varnothing \;\; (P(E_i \wedge E_j) = 0), i, j = 1,..,n, i \neq j$
  - EXH: $E_1 \vee \ldots \vee E_n = S \;\; (P(E_1 \vee \ldots \vee E_n) = 1)$

Mutual Exclusive set of events



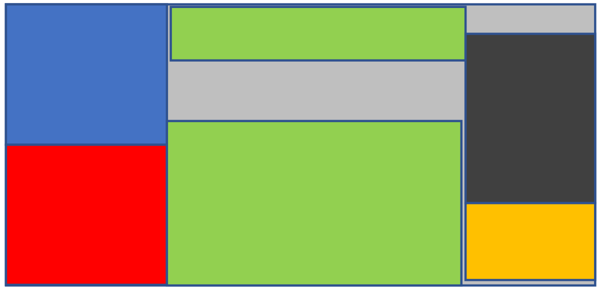
$$E_i \wedge E_j = \varnothing \ \ (P(E_i \wedge E_j) = 0), i, j = 1,..,n, i \neq j$$

Exhaustive sets of events



$$E_1 \vee \ldots \vee E_n = S \ \ (P(E_1 \vee \ldots \vee E_n) = 1)$$

Mutual Exclusive and Exhaustive set of events



No overlap

$$\boxed{E_i \wedge E_j = \varnothing \ \ (P(E_i \wedge E_j) = 0), i, j = 1,..,n, i \neq j}$$

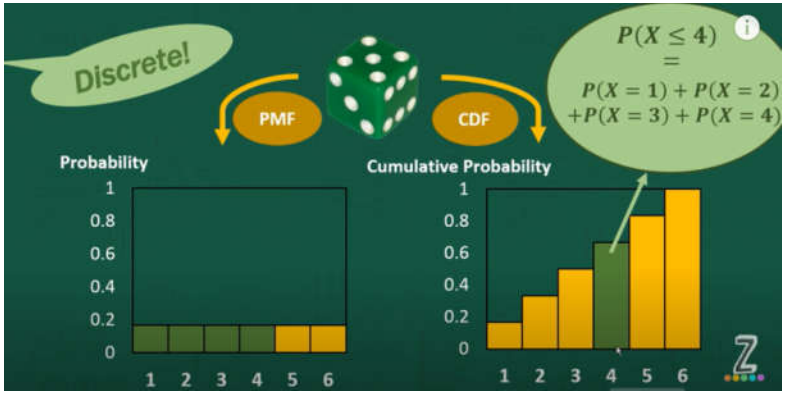
AND

$$\boxed{E_1 \vee \ldots \vee E_n = S \ \ (P(E_1 \vee \ldots \vee E_n) = 1)}$$

Random Variables

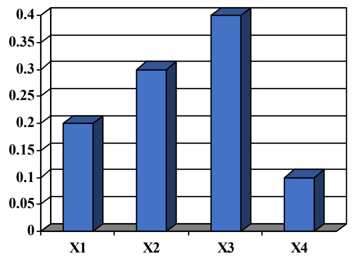
Discrete Random Variables: visualization

- Finite set of possible outcomes



$$X \in \{x_1, x_2, x_3, \ldots, x_n\}$$

$$P(x_i) \geq 0$$

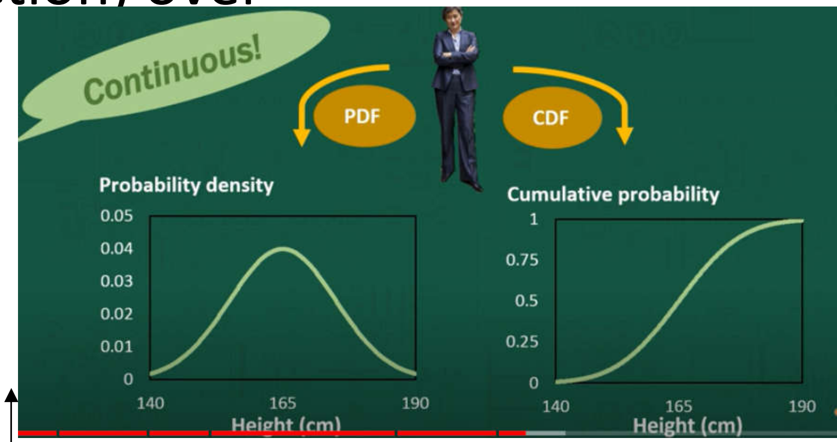


$$\sum_{i=1}^{n} P(x_i) = 1$$

X binary: $P(x) + P(\overline{x}) = 1$

Continuous Random Variable

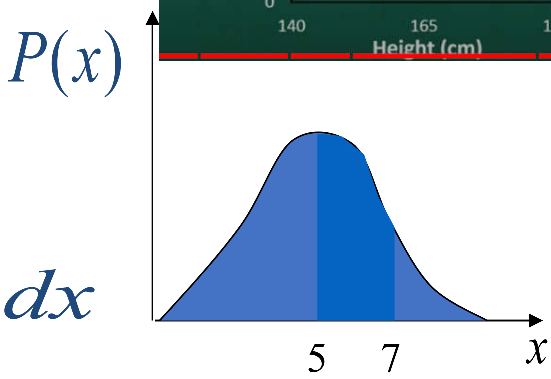
- Probability distribution (density function) over continuous values



$$X \in [0,10] \qquad P(x) \geq 0$$

$$\int_{0}^{10} P(x)dx = 1$$

$$P(5 \leq x \leq 7) = \int_{5}^{7} P(x)dx$$

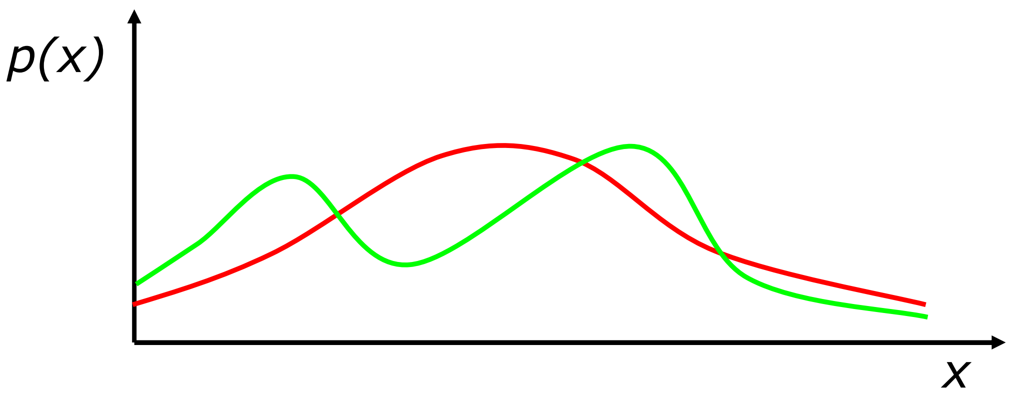


Continuous Random Variables

- X takes on values in the continuum.
- $p(X=x)$, or $p(x)$, is a probability density function (PDF).

$$\Pr(x \in [a,b]) = \int_a^b p(x)dx$$

- E.g.



Probability Distribution

- Probability distribution $P(X|\xi)$
  - X is a random variable
    - Discrete
    - Continuous
  - ξ is background state of information

Joint and Conditional Probabilities

Joint and Conditional Probabilities

- Joint Probabilities

$$P(x, y) \equiv P(X = x \wedge Y = y)$$

  - Probability that both *X*=*x* and Y=y

- Conditional Probabilities

$$P(x \mid y) \equiv P(X = x \mid Y = y)$$

  - Probability that *X*=*x* given we know that *Y*=*y*

Joint and Conditional Probability

- $P(X{=}x \text{ and } Y{=}y) = P(x,y)$

- If X and Y are independent then

$$P(x,y) = P(x)\,P(y)$$

- $P(x \mid y)$ is the probability of *x given y*

$$P(x \mid y) = P(x,y) \,/\, P(y)$$

divided

$$P(x,y) \;= P(x \mid y)\,P(y)$$

- If X and Y are independent then

$$P(x \mid y) = P(x)$$

Law of Total Probability

Discrete case

$$\sum_x P(x) = 1$$

$$P(x) = \sum_y P(x, y)$$

$$P(x) = \sum_y P(x \mid y) P(y)$$

Continuous case

$$\int p(x)\, dx = 1$$

$$p(x) = \int p(x, y)\, dy$$

$$p(x) = \int p(x \mid y)\, p(y)\, dy$$

Rules of Probability: Marginalization

- Product Rule

$$P(X,Y) = P(X \mid Y)P(Y) = P(Y \mid X)P(X)$$

- Marginalization

$$P(Y) = \sum_{i=1}^{n} P(Y, x_i)$$

X binary: $P(Y) = P(Y, x) + P(Y, \overline{x})$

Bayes Rule

$$P(H,E) = P(H \mid E)P(E) = P(E \mid H)P(H)$$

$$\boxed{P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}}$$

$$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \wedge B)$$

$$P(H,E) = P(H \mid E)P(E) = P(E \mid H)P(H)$$

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}$$

10/40 = probability that you smoke if you have cancer = P(smoke/cancer)

10/100 = probability that you have cancer if you smoke

1000-100 = 900 people who do not smoke

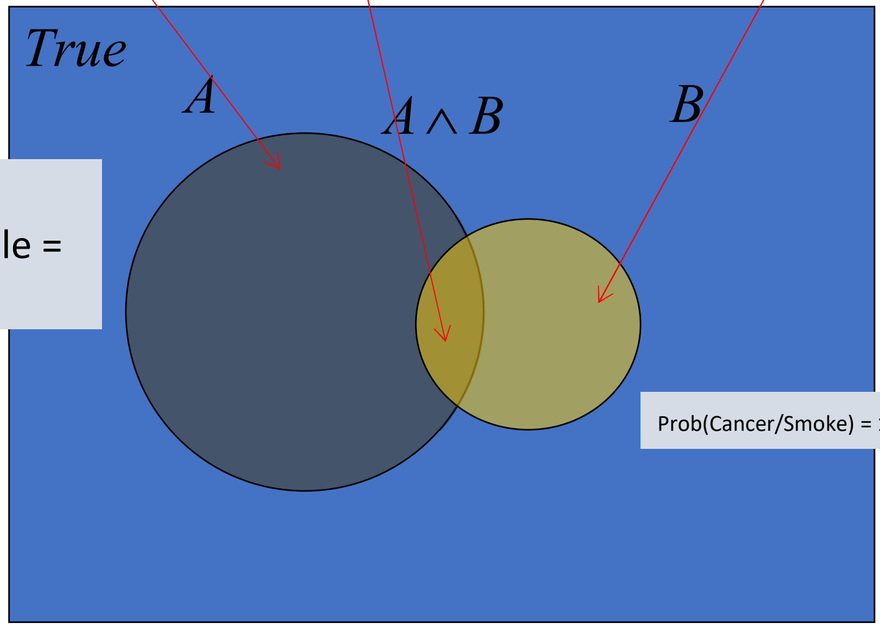
1000-40 = 960 people who do not have cancer

100 People who smoke

40 People who have cancer

10 People who smoke and have cancer

All people = 1000



E = smoke, H = cancer

Prob(Cancer/Smoke) =

P (smoke/Cancer) * P (Cancer) / P(smoke)

P(smoke) = 100/1000
P(cancer) = 40/1000
P(smoke/Cancer) = 10/40 = 25%

Prob(Cancer/Smoke) = 10/40 * 40/1000/ 100 = 10/1000 / 100 = 10/10,000 =/1000 = 0.1%

10/40 = probability that you smoke if you have cancer = P(smoke/cancer)

10/100 = probability that you have cancer if you smoke

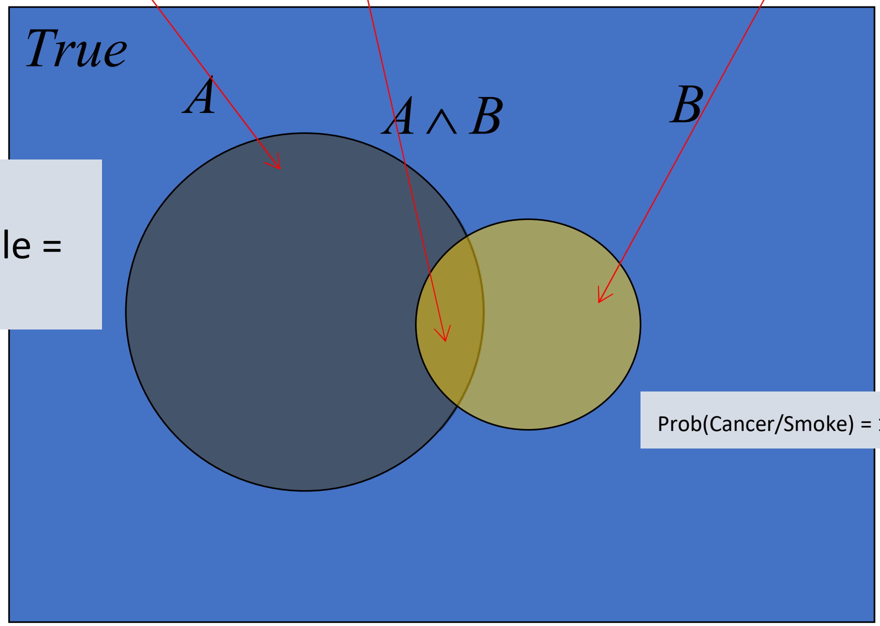
1000-100 = 900 people who do not smoke

1000-40 = 960 people who do not have cancer

100 People who smoke

10 People who smoke and have cancer

40 People who have cancer



All people = 1000

E = smoke, H = cancer

Prob(Cancer/Smoke) =

P (smoke/Cancer) * P (Cancer) / P(smoke)

P(smoke) = 100/1000
P(cancer) = 40/1000
P(smoke/Cancer) = 10/40 = 25%

Prob(Cancer/Smoke) = 10/40 * 40/1000/ 100 = 10/1000 / 100 = 10/10,000 = 1/1000 = 0.1%

E = smoke, H = cancer

Prob(Cancer/Not Smoke) =

P (Not smoke/Cancer) * P (Cancer) / P(Not smoke)

Prob(Cancer/Not smoke) = 30/40 * 40/100 / 900 =
30/100*900 = 30 / 90,000 = 1/3,000 = 0.03 %

Summary

- prior, conditional and joint probability for random variables
  - Prior probability: $P(X)$
  - Conditional probability: $P(X_1 | X_2), P(X_2 | X_1)$
  - Joint probability: $\mathbf{X} = (X_1, X_2), P(\mathbf{X}) = P(X_1, X_2)$
  - Relationship: $P(X_1, X_2) = P(X_2 | X_1)P(X_1) = P(X_1 | X_2)P(X_2)$
  - Independence: $P(X_2 | X_1) = P(X_2), P(X_1 | X_2) = P(X_1), P(X_1, X_2) = P(X_1)P(X_2)$
- Bayesian Rule

$$P(C | \mathbf{X}) = \frac{P(\mathbf{X} | C)P(C)}{P(\mathbf{X})} \Rightarrow Posterior = \frac{Likelihood \times Prior}{Evidence}$$