# Introduction to Probability

# Random Experiment

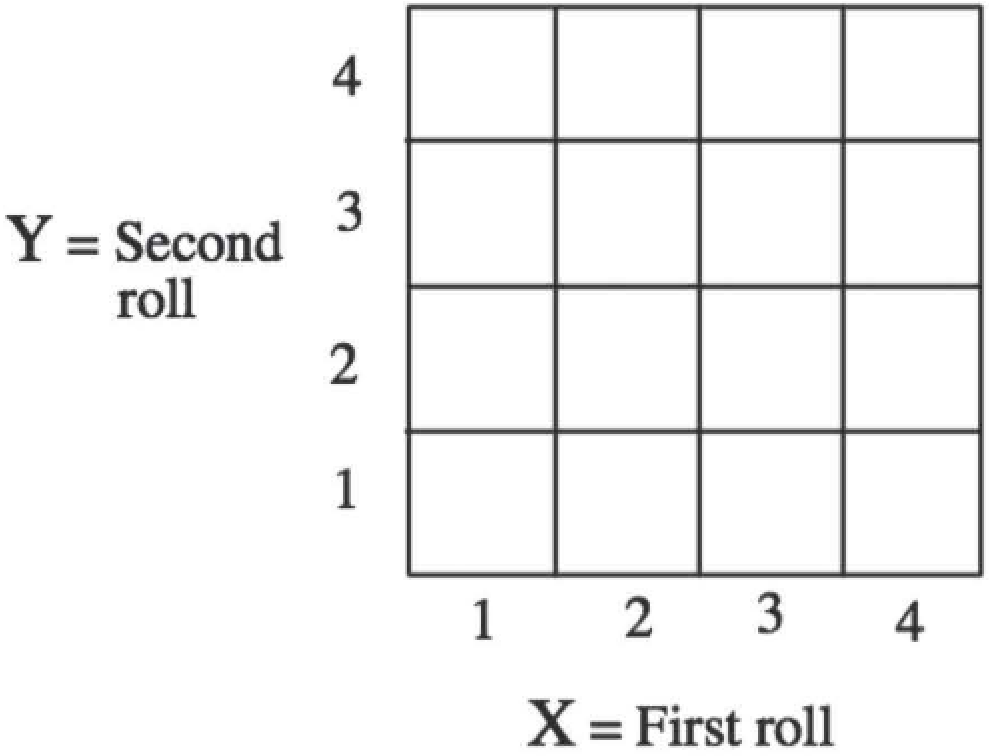
- Any well-defined procedure that produces an observable outcome that could not be perfectly predicted in advance
  - Tossing a coin
  - Rolling a dice twice
  - Goals scored in a soccer match
  - Rainfall today

- Outcome: result of a random experiment
- Random variable: outcome mapped to a numerical value

# Sample Space

- Set of all possible outcomes – $\Omega$
- Elements of the set must be
  - Mutually exclusive
  - Collectively exhaustive

# Sample Space: Two rolls of a 4 faced die



# Probability Events

- Event: a subset of the sample space
- Probability of an event: Value P(A) assigned to an event A
- Event space (Σ): collection of all possible events

- σ-algebra on a set X is a nonempty collection Σ of subsets of X closed under complement, countable unions, and countable intersections. The ordered pair ( X , Σ ) is called a measurable space.
- ( Ω , Σ , P ) is a called probability space, with sample space Ω, event space Σ and probability measure P s.t. satisfying the axioms

# Axioms of Probability

- Nonnegativity: $\mathbf{P}(A) \geq 0$
- Normalization: $\mathbf{P}(\Omega) = 1$
- (Finite) additivity: (to be strengthened later)
  If $A \cap B = \emptyset$, then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$

# Simple consequences (Theorems):

**Axioms**

$\mathbf{P}(A) \geq 0$

$\mathbf{P}(\Omega) = 1$

For disjoint events:

$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$

**Consequences**

$\mathbf{P}(A) \leq 1$

$\mathbf{P}(\emptyset) = 0$

$\mathbf{P}(A) + \mathbf{P}(A^c) = 1$

$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C)$

# More consequences (Theorems):

- If $A \subset B$, then $\mathbf{P}(A) \leq \mathbf{P}(B)$
- $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$
- $\mathbf{P}(A \cup B) \leq \mathbf{P}(A) + \mathbf{P}(B)$

# Probability Models

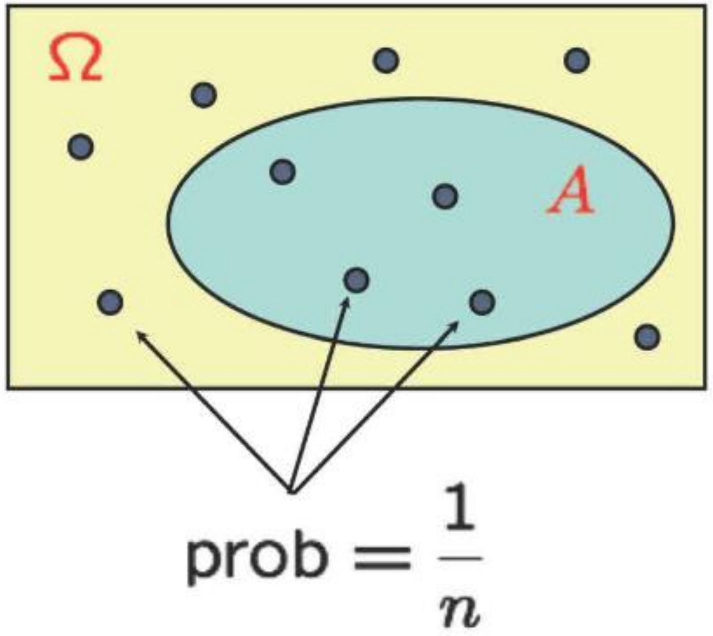
- Rules/mechanism for assigning probability values to events

- Should be faithful to real life phenomenon
- Should produce valid probability values

# Discrete Uniform Law

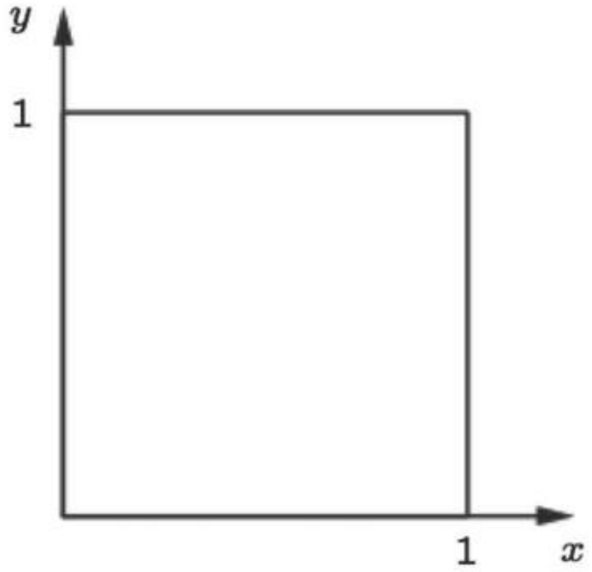
- Assume $\Omega$ consists of $n$ equally likely elements
- Assume $A$ consists of $k$ elements

$\mathbf{P}(A) =$



# Continuous Uniform Law

- $(x, y)$ such that $0 \leq x, y \leq 1$




- **Uniform** probability law: Probability $=$ Area

$$\mathbf{P}\Big(\{(x, y) \mid x + y \leq 1/2\}\Big) =$$

$$\mathbf{P}\Big(\{(0.5, 0.3)\}\Big) =$$

# Probability Calculation Steps

- Specify the sample space
- Specify a probability model
- Identify an event of interest
- Calculate ...

# Building Probability Models

- Depends on interpretation of probability values

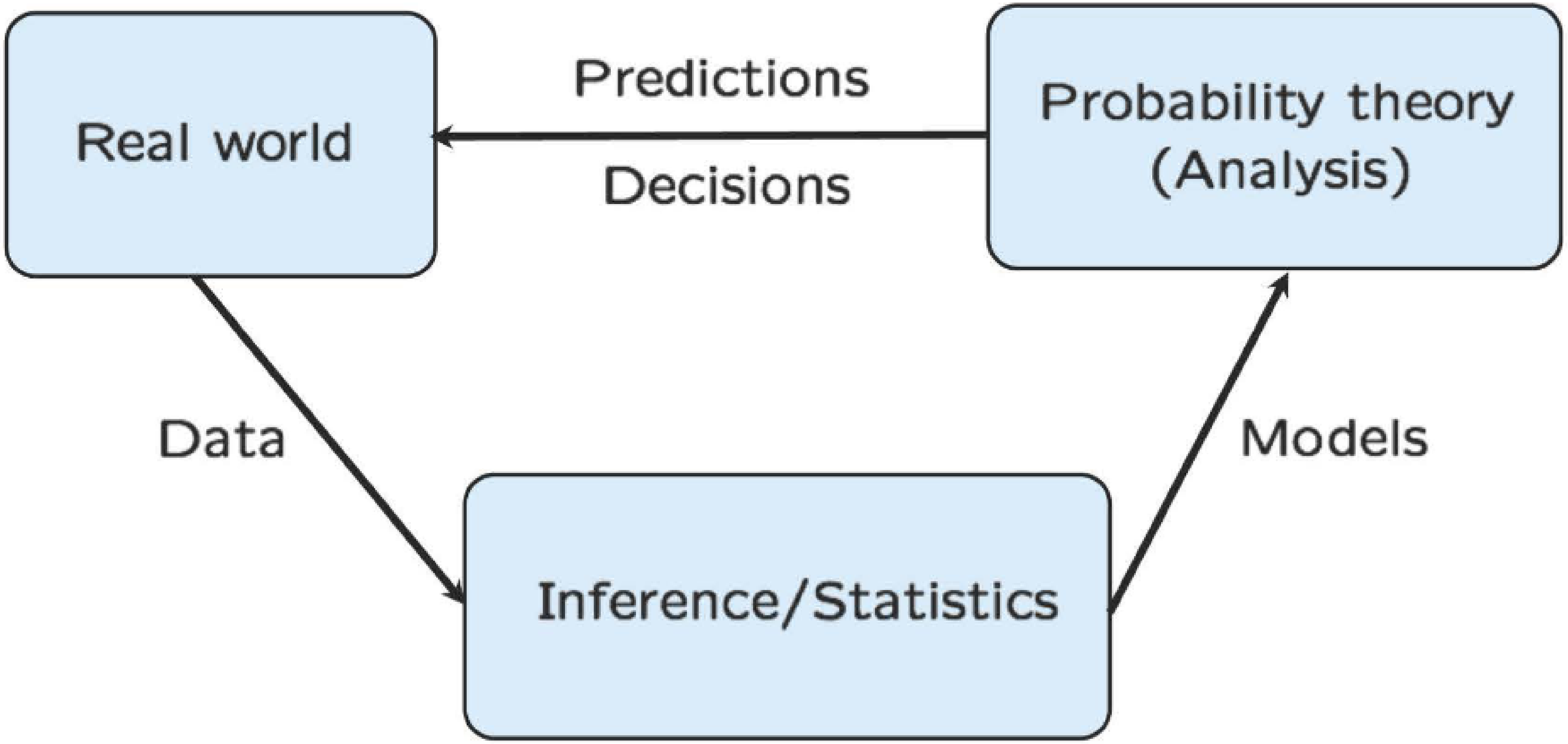
- Frequentist: (Theorem: “Frequency” of event A “is” P(A))
  - Are probabilities frequencies?

  \- P(coin toss yields heads) = 1/2

  \- P(the president of ... will be reelected) = 0.7

- Probabilities are often interpreted as:
  - Description of beliefs/knowledge

# Building Probability Model: Statistical Inference/Learning

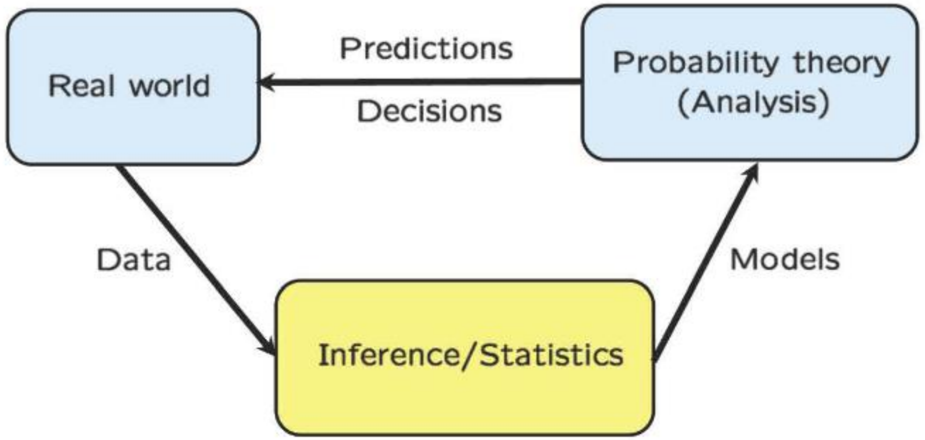


# Model Inference vs Estimating a Variable

$$X = aS + W$$

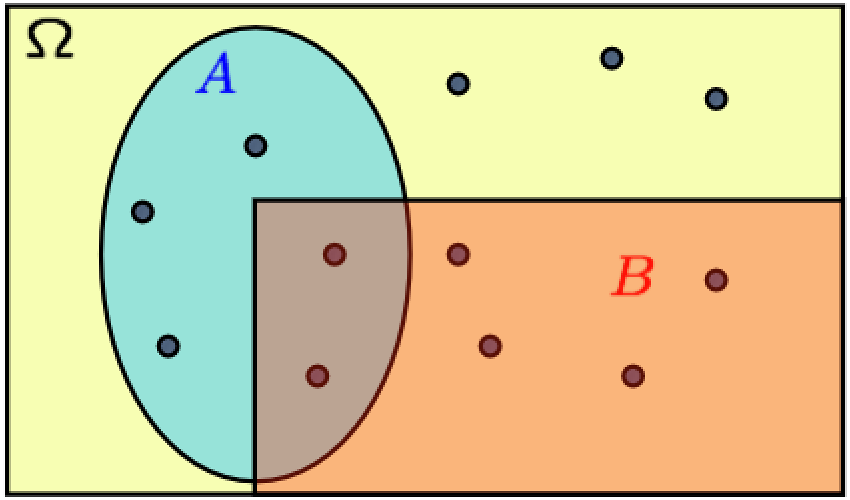
- Model building:
  - know "signal" $S$, observe $X$
  - infer $a$

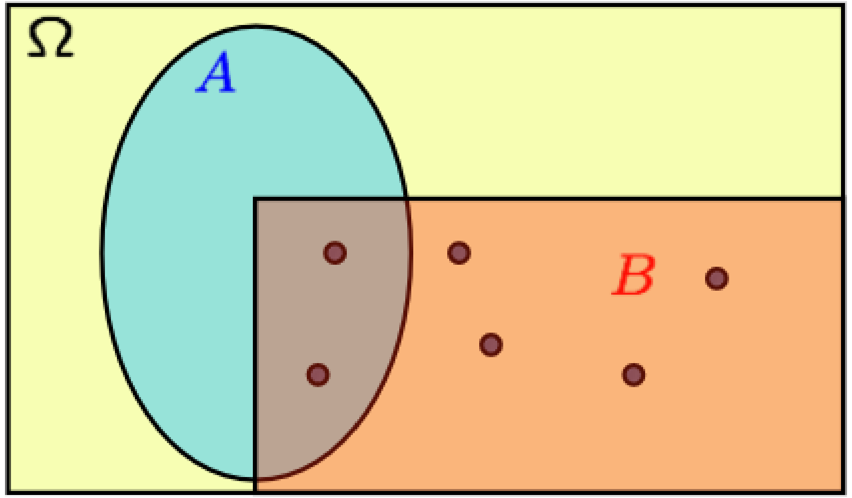
- Variable estimation:
  - know $a$, observe $X$
  - infer $S$



# Conditioning

- Use new information to revise a model (12 equally likely outcomes)

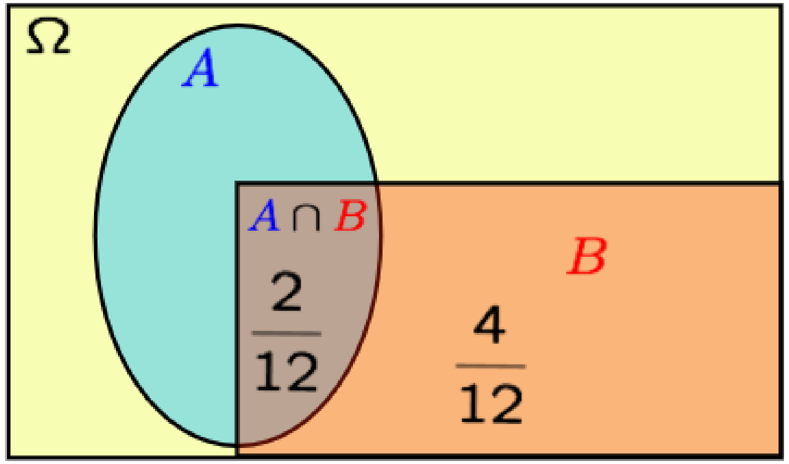



$\mathbf{P}(A) = \frac{5}{12}$ $\quad \mathbf{P}(B) = \frac{6}{12}$

$\mathbf{P}(A \mid B) =$ $\quad \mathbf{P}(B \mid B) =$

# Conditional Probability

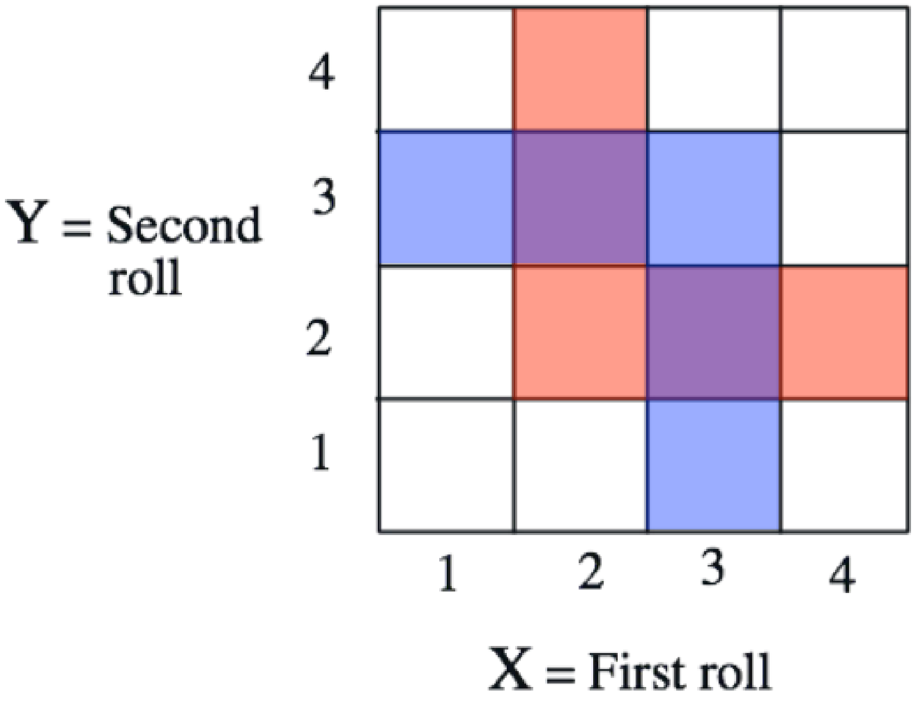



- $\mathbf{P}(A \mid B) =$ "probability of $A$, given that $B$ occurred"

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

defined only when $\mathbf{P}(B) > 0$

# Example: 2 rolls of a 4 sided die




- Let $B$ be the event: $\min(X, Y) = 2$

  Let $M = \max(X, Y)$

$\mathbf{P}(M = 1 \mid B) =$

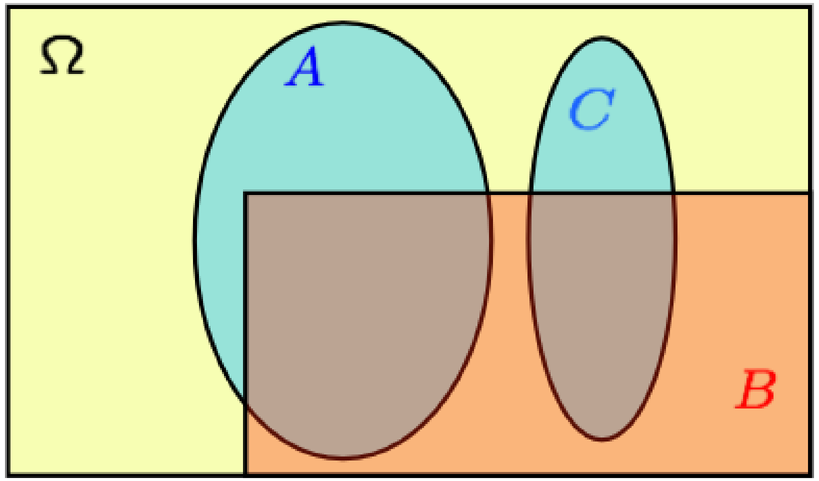
$\mathbf{P}(M = 3 \mid B) =$

# Conditional Probabilities Share Property of Original Probabilities

$\mathbf{P}(A \mid B) \geq 0$ assuming $\mathbf{P}(B) > 0$

$\mathbf{P}(\Omega \mid B) =$

$\mathbf{P}(B \mid B) =$




If $A \cap C = \emptyset$, then $\mathbf{P}(A \cup C \mid B) = \mathbf{P}(A \mid B) + \mathbf{P}(C \mid B)$

# Model Based on Conditional Probability

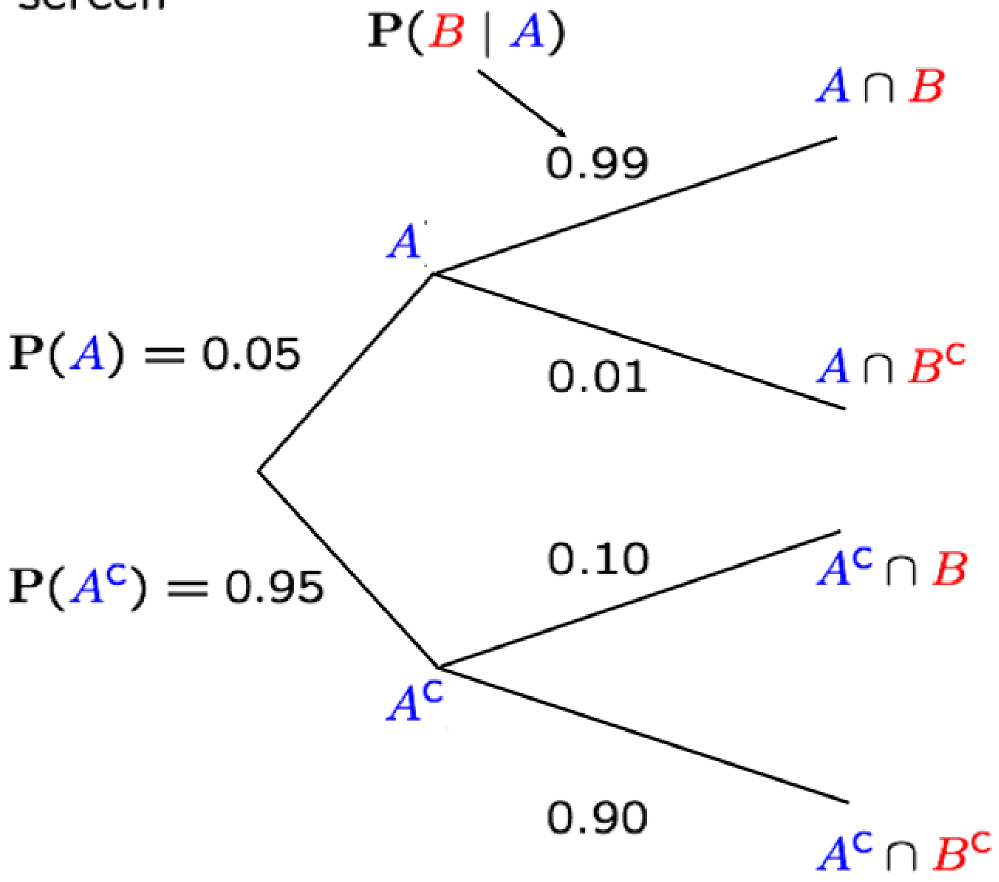
Event $A$: Airplane is flying above

Event $B$: Something registers on radar screen

- $\mathbf{P}(A \cap B) =$

- $\mathbf{P}(B) =$

- $\mathbf{P}(A \mid B) =$



# Multiplication Rule

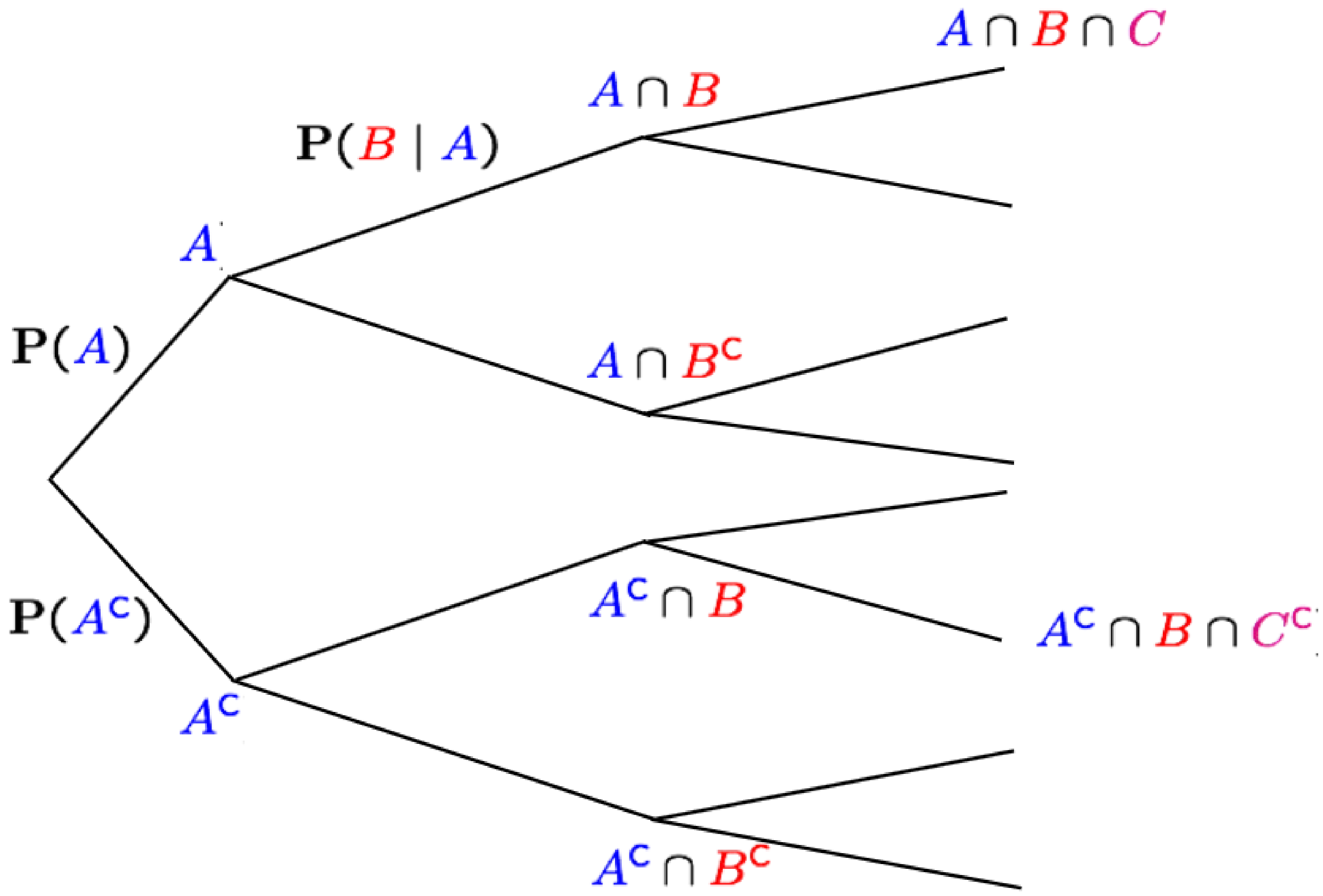
**The multiplication rule**

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

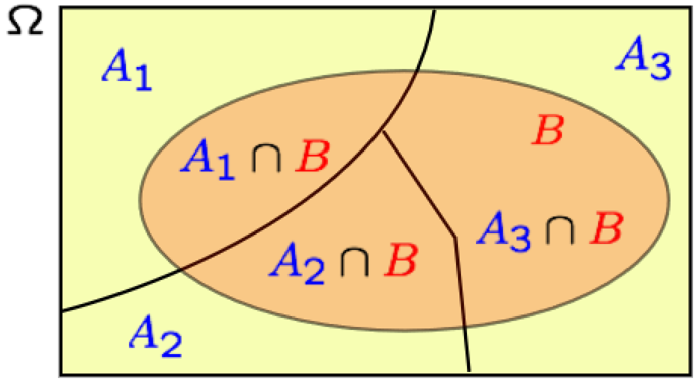
$$P(A \cap B) = P(B)\,P(A \mid B)$$

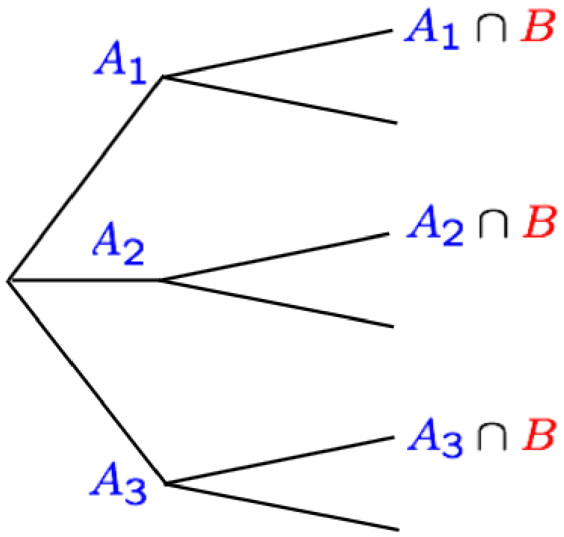
$$= P(A)\,P(B \mid A)$$

$$P(A^c \cap B \cap C^c) =$$



# Total Probability Theorem



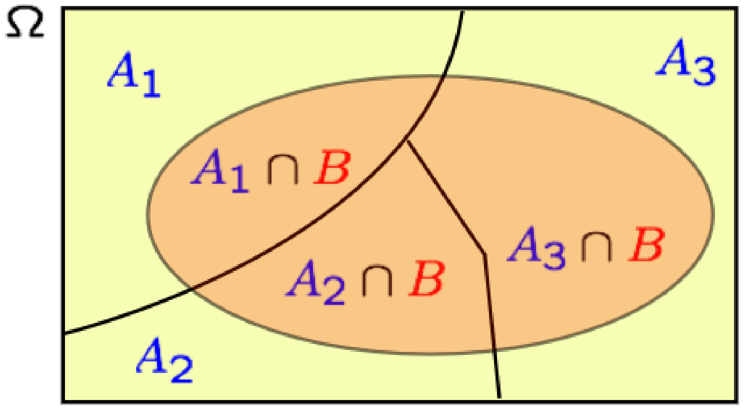



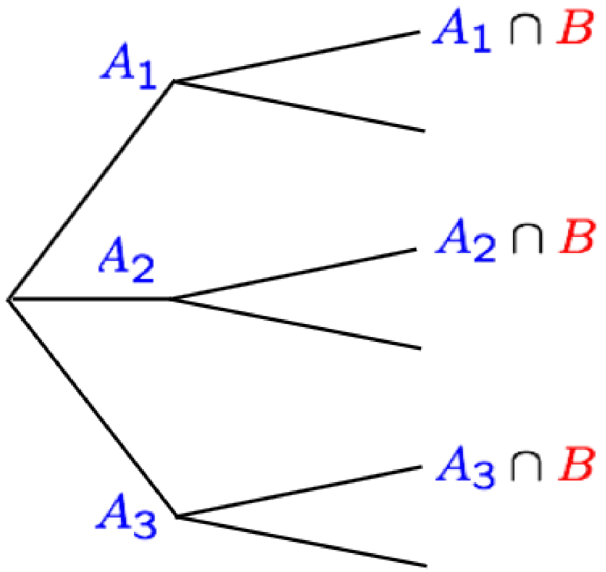

- Partition of sample space into $A_1, A_2, A_3$
- Have $\mathbf{P}(A_i)$, for every $i$
- Have $\mathbf{P}(B \mid A_i)$, for every $i$

$\mathbf{P}(B) =$

$$\boxed{\mathbf{P}(B) = \sum_i \mathbf{P}(A_i)\,\mathbf{P}(B \mid A_i)}$$

# Bayes Rule




- Partition of sample space into $A_1, A_2, A_3$
- Have $\mathbf{P}(A_i)$, for every $i$ initial "beliefs"
- Have $\mathbf{P}(B \mid A_i)$, for every $i$

revised "beliefs," given that $B$ occurred:

$$\mathbf{P}(A_i \mid B) =$$

$$\boxed{\mathbf{P}(A_i \mid B) = \frac{\mathbf{P}(A_i)\mathbf{P}(B \mid A_i)}{\sum_j \mathbf{P}(A_j)\mathbf{P}(B \mid A_j)}}$$

# Bayesian Inference

- initial beliefs $\mathbf{P}(A_i)$ on possible causes of an observed event $B$
- model of the world under each $A_i$: $\mathbf{P}(B \mid A_i)$

$$A_i \xrightarrow[\mathbf{P}(B \mid A_i)]{\text{model}} B$$

- draw conclusions about causes

$$B \xrightarrow[\mathbf{P}(A_i \mid B)]{\text{inference}} A_i$$

# Joint Probability

- For events A and B, **joint probability** Pr(AB) stands for the probability that both events happen.
- Example: A={HH}, B={HT, TH}, what is the joint probability Pr(AB)?

# Independence

- Two events ***A and B are independent*** in case

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

- A set of events $\{A_i\}$ is independent in case

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

# 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:
  - Continuous case:
  - Probability density function
  - Probability mass function

$$\Pr(X = x) = p_\theta(x)$$

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

# Random Variable: Example

- Let S be the set of all sequences of three rolls of a die. Let X be the sum of the number of dots on the three rolls.
- What are the possible values for X?
- $\Pr(X = 5) = ?$, $\Pr(X = 10) = ?$

# Expectation

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

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

  - In an empirical sample, x1, x2,…, xN,

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

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

- Expectation of sum of random variables

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

# Variance

- The variance of a random variable X is the expectation of $(X\text{-}E[x])^2$ :

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

# Bernoulli Distribution

- The outcome of an experiment can either be success (i.e., 1) and failure (i.e., 0).
- Pr(X=1) = p, Pr(X=0) = 1-p, or
- E[X] = p, Var(X) = p(1-p)

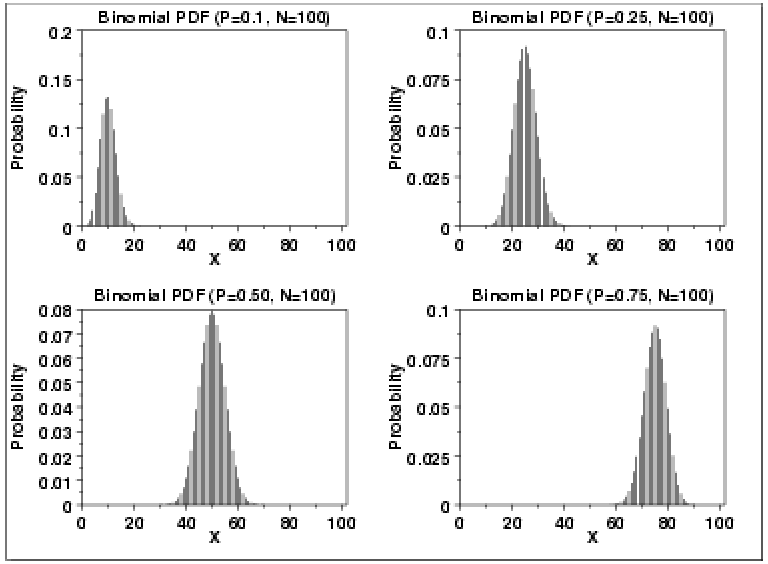
# Binomial Distribution

- n draws of a Bernoulli distribution
  - $X_i$~Bernoulli(p), X=$\sum_{i=1}^{n} X_i$, X~Bin(p, n)
- Random variable X stands for the number of times that experiments are successful.

$$\Pr(X = x) = p_\theta(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

- E[X] = np, Var(X) = np(1-p)

# Plots of Binomial Distribution



# Poisson Distribution

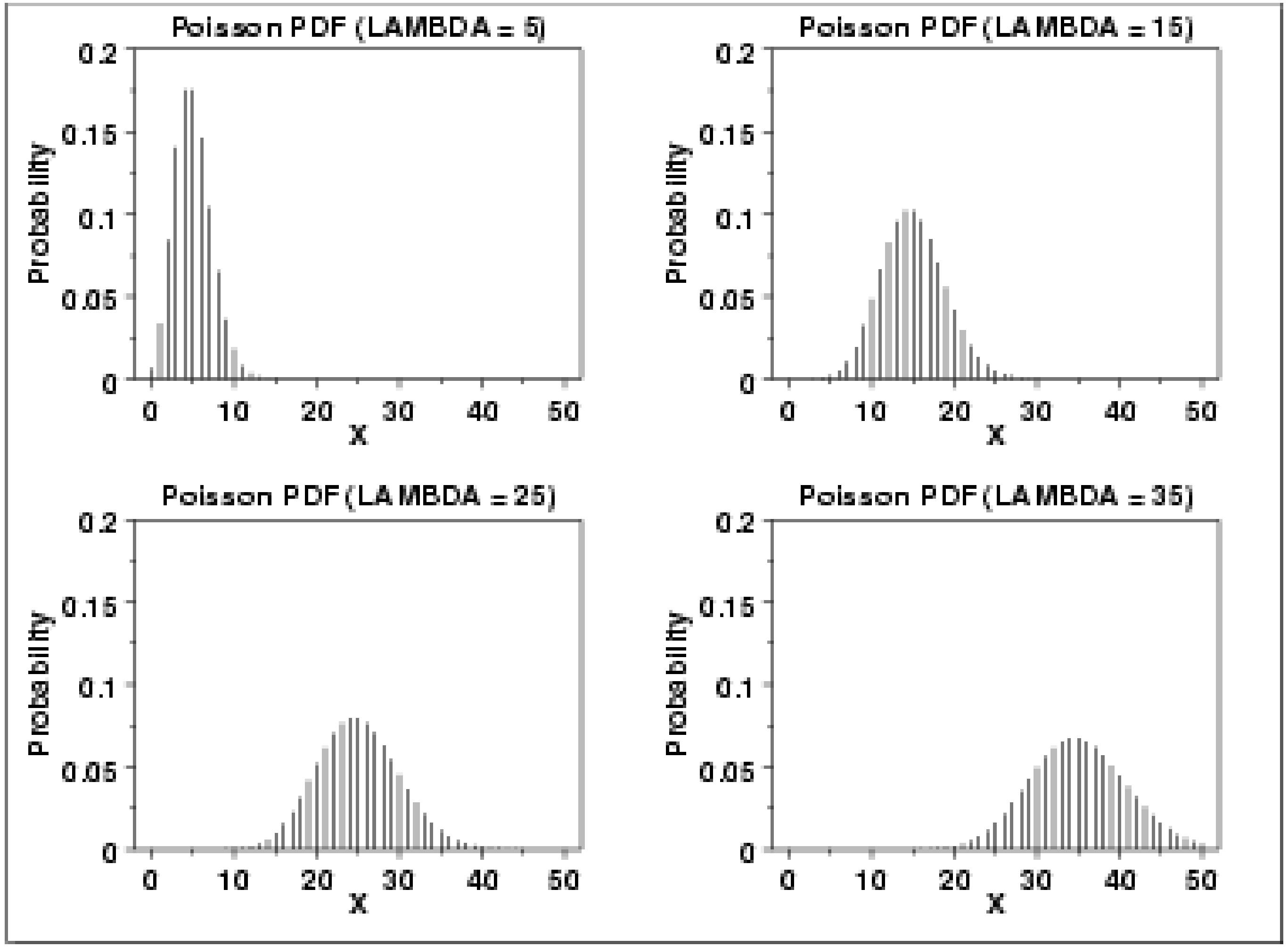
- Coming from Binomial distribution
  - Fix the expectation $\lambda$=np
  - Let the number of trials n$\rightarrow\infty$

  A Binomial distribution will become a Poisson distribution

$$\Pr(X = x) = p_\theta(x) = \begin{cases} \dfrac{\lambda^x}{x!} e^{-\lambda} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- E[X] = $\lambda$, Var(X) = $\lambda$

# Plots of Poisson Distribution



# Normal (Gaussian) Distribution

- $X \sim N(\mu, \sigma)$

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$$\Pr(a \le X \le b) = \int_a^b p_\theta(x)dx = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx$$

- $E[X] = \mu$, $Var(X) = \sigma^2$
- If $X_1 \sim N(\mu_1, \sigma_1)$ and $X_2 \sim N(\mu_2, \sigma_2)$, $X = X_1 + X_2$ ?