Probability Review

Jeremy Irvin and
Daniel Spokoyny

Created from Maleki and Do's Probability Review for Stanford CS229

Consider flipping a coin twice.

Elements of Probability

- Sample Space (Ω): Set of all outcomes
- $\Omega = \{HH, HT, TH, TT\}$

Elements of Probability

- Event (E): A subset E of Ω, ie, a subset of outcomes
- E = {HH, HT}

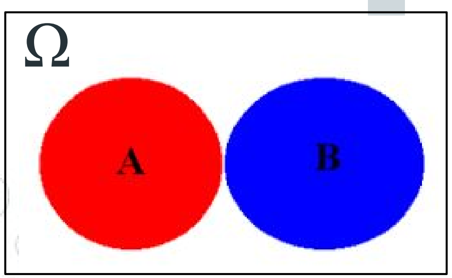
Elements of Probability

- Event Space (F): Set of all possible events, ie, set of all subsets of Ω
- F= {∅, {HH}, {TT}, {TH}, {TT}, {HH, HT} , …}

Elements of Probability

- Probability Measure (P): A function $P : \mathcal{F} \to \mathbb{R}$ satisfying:
  - (i) $P(A) \geq 0$, for all $A \in \mathcal{F}$

  - (ii) $P(\Omega) = 1$



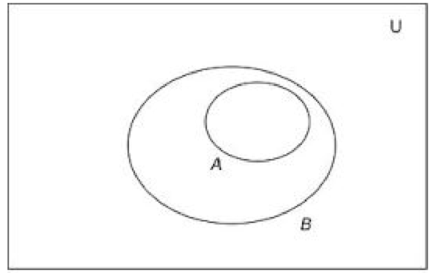
  - (iii) If $A_1, A_2, \ldots$ are disjoint events ($A_i \cap A_j = \emptyset$ when i≠j), then

$$P(\cup_i A_i) = \sum_i P(A_i)$$

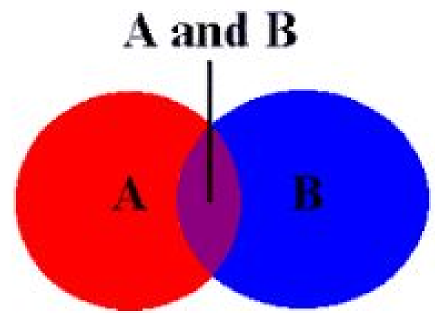
Simple Example

- P({HH}) = ¼ , P({HT}) = ¼, P({HH, HT}) = ½
- Notice {HH} and {HT} are disjoint events, and
    - P({HH, HT}) = P({HH}) + P({HT})

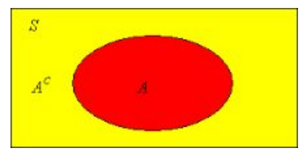
Properties



$A \subset B$



- If $A \subseteq B \implies P(A) \leq P(B)$.
- $P(A \cap B) \leq \min(P(A), P(B))$.
- (Union Bound) $P(A \cup B) \leq P(A) + P(B)$.
- $P(\Omega \setminus A) = 1 - P(A)$.

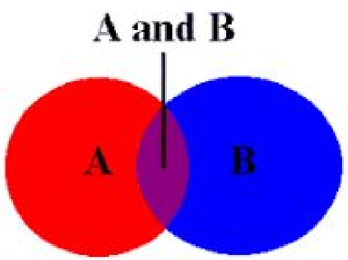


- If $A_1, \ldots, A_k$ are disjoint events with

$$\cup_{i=1}^{k} A_i = \Omega,$$

then $\sum_{i=1}^{k} P(A_k) = 1.$

Conditional Probability



- If B is an event with non-zero probability ($P(B) \neq 0$) then the <u>conditional probability</u> of A <u>given</u> B is

$$P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}$$

	23	2	25
	12	3	15
	35	5	40

- In other words, $P(A|B)$ is the probability of event A after observing event B.

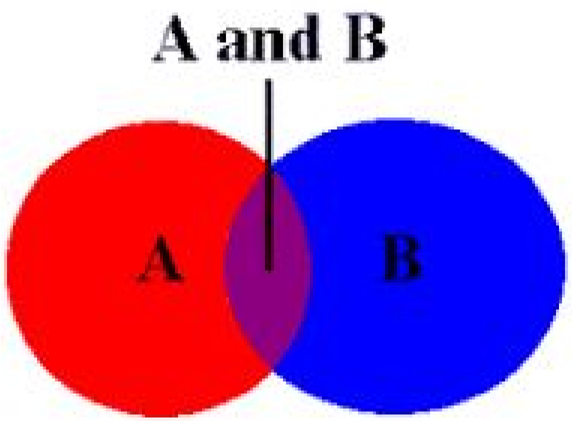
Independence

- A and B are independent if

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

or equivalently,

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



HH	HT
TH	TT

Bayes Theorem!!!

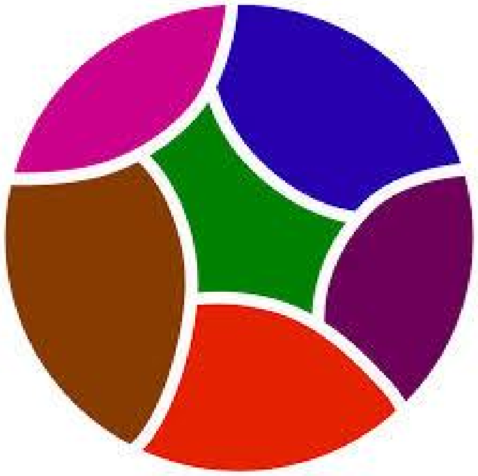
- If A and B are any two events, then

$$P(A|B) = \frac{P(A)\,P(B|A)}{P(B)}$$

- If {*A_j*} is a partition of the sample space, then

$$P(B) = \sum_j P(B \mid A_j) P(A_j),$$

$$\Rightarrow P(A_i \mid B) = \frac{P(B \mid A_i)\,P(A_i)}{\sum_j P(B \mid A_j)\,P(A_j)}.$$



Random Variables

- Suppose we flip 10 coins and want to know the number of coins which come up heads.

- Maybe we get the sequence:
  {HHTHTTTHTH}

Random Variables

- Real-valued functions of outcomes (such as the number of heads that appear among our 10 tosses) are known as <u>random variables.</u>

- More formally, a random variable X is a function $X : \Omega \longrightarrow \mathbb{R}$.

Random Variable Example

- Suppose we are flipping a coin 10 times.
- For any outcome $w \in \Omega$, let $X(w)$ be the number of heads which occur in w.
- X is discrete since it can only take on a countable amount of values {0,1…,10} (a random variable is continuous if it takes on an uncountable number of values) and

$$P(X = k) = P(\{w{:}X(w) = k\}) = 10Ck/2\text{^}10$$

Probability Mass Function (pmf)

- A <u>probability mass function (pmf)</u> corresponding to a discrete random variable X is a function $p_X : Val(X) \to [0, 1]$ where

$$p_X(x) := P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

- $$\sum_{x \in Val(X)} p_X(x) = 1$$

$$A \subseteq Val(X)$$

- $$\sum_{x \in A} p_X(x) = P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\})$$

Cumulative Distribution Function (cdf)

- A cumulative distribution function (cdf) corresponding to a random variable X is a function $F_X : \mathbb{R} \to [0, 1]$ which specifies a probability measure as

$$F_X(x) \triangleq P(X \leq x).$$

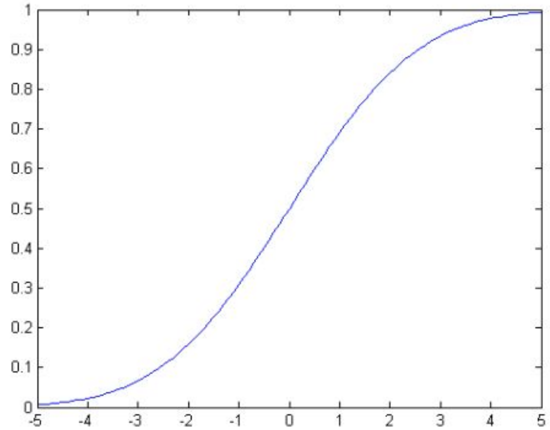


Figure 1: A cumulative distribution function (CDF).

cdf Properties

- $0 \le F_X(x) \le 1.$
- $\lim_{x\to-\infty} F_X(x) = 0.$
- $\lim_{x\to\infty} F_X(x) = 1.$
- $x \le y \Longrightarrow F_X(x) \le F_X(y).$

Probability Density Function (pdf)

- A <u>probability density function (pdf)</u> corresponding to a continuous random variable X with differentiable cdf F_X is a function $f_X : \Omega \to \mathbb{R}$ where

$$f_X(x) \triangleq \frac{dF_X(x)}{dx}.$$

- $f_X(x) \geq 0$.
- $\int_{-\infty}^{\infty} f_X(x) = 1.$
- $\int_{x \in A} f_X(x) dx = P(X \in A).$

Expected Value

- If X is a random variable with pmf $p_X(x)$ the expected value of X is defined as

$$\mathbb{E}[X] := \sum_x x P(X = x)$$

- Think of $\mathbb{E}[X]$ as a weighted average of the values x that X can take on with weights $p_X(x)$.

Expected Value

- $E[X]$ is called the <u>mean</u> of X

- $E[a] = a$ for any constant $a \in \mathbb{R}$

- $E[X + Y] = E[X] + E[Y]$ (linearity)

Variance

- The variance of a random variable X is a measure of how concentrated the distribution of X is around its mean $E[X]$

- Formally, the variance of X is defined

$$Var[X] \triangleq E[(X - E(X))^2] = E[X^2] - E[X]^2$$

- $Var[a] = 0$ for any constant $a \in \mathbb{R}$.

- $Var(aX) = a^2 Var(X)$

Common Discrete Distributions

- X~Bernoulli(p)

$$p(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

A single coin flip, with heads probability p.

- X~Binomial(n,p)

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

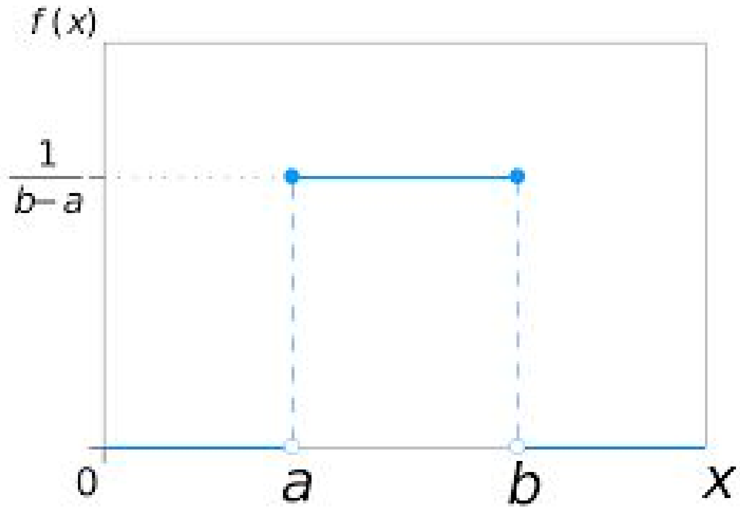
The number of heads in n independent flips of a coin with heads probability p.

Common Continuous Distributions

- X~Uniform(a,b)

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Any equal-sized interval occurs with equal probability:

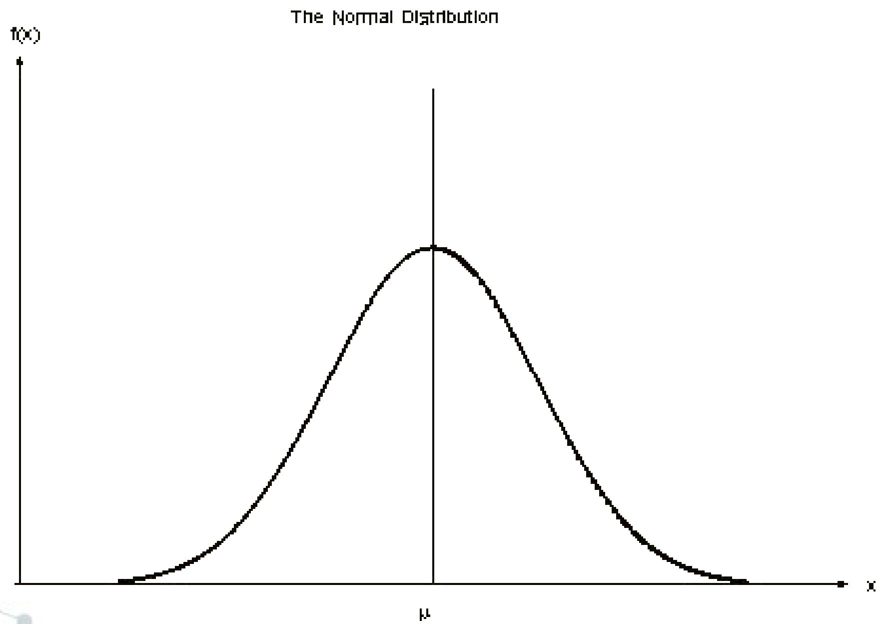


Common Continuous Distributions

- X~Normal(μ,σ²)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Also known as Gaussian, the typical 'bell-curve':



Expected Value and Variance Example

- Calculate the mean and the variance of the uniform random variable X with pdf $f_X(x) = 1$ for $x \in [0, 1]$ and $f_X(x) = 0$ elsewhere.
- Answer:
  E[X]=½ , Var(X) = 1/12

What Just Happened?

- Axioms of Probability
- Bayes Theorem
- Random Variables
- Common Distributions

Python demo

Gaussian Distribution Sampling

Done with probability!