Definition of Independence

- $P(A|B) = P(A)$
- $P(A \cap B) = P(A)P(B)$
- Recall $A_1 \to A_n$ independent if
  - $P(\cap_j A_{ij}) = \prod_j P(A_{ij})$
  - For any indices $i_1 \to i_m, m \in [1, n]$

Claim: If $A_1 \to A_n$ are independent, then so are $A_1^c \to A_n^c$

Inclusion Exclusion Principle

- Fact: Let $B_1 \to B_n$ be any events (not necessarily independent or disjoint)
- Then: $P(\cap B_i) = \sum P(B_i) - \sum_{i<j} P(B_i \cap B_j) + \sum_{i<j<k} P(B_i \cap B_j \cap B_k) - ... + (-1)^{n+1} P(B_1 \cap B_2 \cap ... \cap B_n)$
  - [Similar to a Taylor's series]

Let's show $P(\cap A_i^c) = \prod P(A_i^c)$

- Demorgan:
  $P((\cup A_i)^c) = 1 - P(\cap A_i)$
  $= 1 - \sum P(A_i) + \sum_{i<j} P(A_i \cap A_j) - ... + (-1)^n P(A_1 \cap ... \cap A_n)$
- Fact: If $x_1 \to x_n$ are any real numbers, then:
  $(1 - x_1)(1 - x_2)...(1 - x_n)$
  $= 1 - \sum x_i + \sum_{i<j} x_i x_j - ... + (-1)^n * P(A_1)...P(A_n)$
  $= P(A_1^c)...P(A_n^c)$

Example: Circuit Problem

x|-1-|x

-|-2-|-

x|. . . |x

x|-n-|x

- Assume $A_1 \to A_n$ are independent
  $\to P(A_1) = P(A_2) = ... = P(A_n)$
- $W = \cup A_i$
- $P(W) = 1 - P(W^c)$
  $= 1 - P((\cup A_i)^c)$
  $= 1 - P(\cap A_i^c)$
  $= 1 - \prod P(A_i^c) = 1 - (1 - p)^n$

Example: Another circuit problem

x|-1-|-4-|x

-|xxx3xxx|-

x|-2-|-5-|x

- $A_i = \{i\text{-th component works}\}$
- Same assumptions
- $W = \{\text{circuit works}\}$
- $P(W) = P(W|A_3)P(A_3) + P(W|A_3^c)P(A_3^c)$
  $= P(W|A_3)p + P(W|A_3^c)(1 - p)$

Solve $P(W|A_3)$

- $P(W|A_3) = P((A_1 \cup A_2) \cap (A_4 \cup A_5)|A_3)$
  $= P((A_1 \cup A_2) \cap (A_4 \cup A_5))$ because independence
  $= P(A_1 \cup A_2)P(A_4 \cup A_5)$
  $= [P(A_1) + P(A_2) - P(A_1 \cap A_2)][...]$
  $= (2p - p^2)^2$

Solve $P(W|A_3^c)$

- $P(W|A_3^c) = P((A_1 \cap A_4) \cup (A_2 \cap A_5)|A_3^c)$
  $= P((A_1 \cap A_4) \cup (A_2 \cap A_5))$ because independence
  $= P(A) + P(B) - P(A \cap B)$
  $= 2p^2 - p^4$

Altogether

- $P(W) = (2p - p^2)^2 p + (2p^2 - p^4)(1 - p)$

Random Variables

- Informally, these are just numbers attached to the outcomes of experiments
- Formally, a random variable, say x, is a function $x : \Omega \to \mathbb{R}$

Example: You toss a coin three times (all outcomes are equally likely)

- $\Omega = \{$ HHH, HHT, ..., TTT $\}$
- $|\Omega| = 8$
- Consider the number of heads among the three tosses (call this X)
  - $X(HHH) = 3$
  - $X(HTH) = 2$
  - $X(TTH) = 1$
  - $X(TTT) = 0$
  - X takes on the set {0, 1, 2, 3}
- Consider the number of heads - number of tails, call this Y
  - $Y(HHH) = 3 - 0 = 3$
  - $Y(HTH) = 2 - 1 = 1$
  - $Y(TTH) = 1 - 2 = -1$
  - $Y(TTT) = 0 - 3 = -3$
  - Y takes on the set {-3, -1, 1, 3}

There's a much more efficient way of thinking about random variables

- Let X be a random variable taking possible values $x_1, x_2, ...$
- Define the probability mass function (pmf) of X as $p_X(x_j) = P(\{w : X(w) = x_j\}) = P(X = x_j)$
  - Note: X is random and x_j is fixed
- Pmf is also called a frequency function

Example: X = number of heads among 3 tosses

- $p_X(0) = \frac{1}{8}$
- $p_X(1) = \frac{3}{8}$
- $p_X(2) = \frac{3}{8}$

- $p_X(3) = \frac{1}{8}$

Note: A pmf captures all the statistical information there is to know about a random variable

Cumulative distribution function (cdf) is another function we can attach to a random variable

- $F_X(x) = P(X \leq x)$
- Compare this to $p_X(x) = P(X = x)$
- Fact: A cdf also captures all the statistical information about a random variable