EECS 310: Discrete Math	**Lecture 13**
Probability	

Reading: MIT OpenCourseWare 6.042 Chapter 16

Independence

[[*Intuitively, flip two coins in different cities, outcome of one does not change other. Independence formalizes this.*]]

Def: Events A, B *independent* if

- $\Pr[B] = 0$ or
- $\Pr[A|B] = \Pr[A]$

[[*Does not mean A, B disjoint, in fact disjoint events are NOT independent; knowing B happens means A did not happen if A, B disjoint.*]]

Claim: A, B independent iff $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$.

Proof:

1. $\Pr[B] = 0$:
   - indep. by defn
   - both sides zero
2. $\Pr > 0$:
   - $\Pr[A \cap B] = \Pr[A|B] \cdot \Pr[B]$
   - and $\Pr[A|B] = \Pr[A]$ iff A, B independent

Example: Coin toss:

- n coin tosses, each prob. 1/2 heads [[*indep by assumption*]]
- M_{ij} = event i'th toss = j'th toss
  1. sample space: 3-bit sequences
  2. event M_{ij}: sequences where i'th bit = j'th bit
  3. outcome probability: $(1/2)^n$, uniform
  4. event probability: $2^{n-1}/2^n = 1/2$

Question: $M_{ij}, M_{i'j'}$ indep. if

- $i = i'$, $j = j'$, $i \neq j$?
  $$\Pr[M_{ij} \cap M_{ij}] = \Pr[M_{ij}] = 1/2$$
  but
  $$\Pr[M_{ij}] \cdot \Pr[M_{ij}] = (1/2)^2 = 1/4$$
- none equal?
  $$\Pr[M_{ij} \cap M_{i'j'}] = 2^{n-2}/2^n = 1/4$$
  and
  $$\Pr[M_{ij}] \cdot \Pr[M_{i'j'}] = (1/2)^2 = 1/4$$
- $i = i'$, $j \neq j'$, $i \neq j$, $i \neq j'$?
  $$\Pr[M_{ij} \cap M_{ij'}] = 2^{n-2}/2^n = 1/4$$
  and
  $$\Pr[M_{ij}] \cdot \Pr[M_{i'j'}] = (1/2)^2 = 1/4$$

Def: events $E_1, \ldots, E_n$ mutually independent if $\forall i, \forall S \subseteq \{1, \ldots, n\} - \{i\}$,

$$\Pr[\cap_{j \in S} E_j] = 0$$

or

$$\Pr[E_i | \cap_{j \in S} E_j] = \Pr[E_i].$$

Claim: mutually independent iff $\forall S \subseteq \{1, \ldots, n\}$,

$$\Pr[\cap_{i \in S} E_i] = \prod_{i \in S} \Pr[E_i].$$

Example: Coin toss:

Question: Are $\{M_{ij}\}$ mutually indep.?

$$\Pr[M_{12} \cap M_{23} \cap M_{31}] = 2^{n-2}/2^n = 1/4$$

but

$$\Pr[M_{12}] \cdot \Pr[M_{23}] \cdot \Pr[M_{31}] = 1/8$$

Def: $\{M_{ij}\}$ k-wise indep. iff every subset of k is mutually indep.

Example: Coin toss: $\{M_{ij}\}$ are 2-wise (or pairwise) indep.

Example: Birthday paradox:

Question: Probability two of us have same birthday?

Variables: m people, N days

Assumptions:

- for each person, all bdays equally likely

  [[*actually more likely to be born on a week day; most common birthday Oct. 5th; least common May 22nd.*]]

- bdays mutually indep

  [[*not if there are twins, for example*]]

[[*Assumptions valid for CS applications, will see later.*]]

Four-step method:

1. sample space: map people i to bdays b_i

$$S = \{(b_1, \ldots, b_m) | b_i \in \{1, \ldots, N\}\}$$

2. events:

   A = event ≥ 2 people have same bday [[*hard to evaluate, use complement instead*]]

   A^c = event no two people have same bday

$$A^c = \{(b_1, \ldots, b_m) | \forall i \neq j, b_i \neq b_j\}$$

   Recall: $\Pr[A] = 1 - \Pr[A^c]$

3. outcome prob.:

   - $\Pr[b_i = k] = 1/N$ [[*by 1'st assumption*]]
   - $\Pr[(b_1, \ldots, b_m) = (k_1, \ldots, k_m)] = \prod_i \Pr[b_i = k_i] = (1/N)^m$ [[*by 2'nd assumption*]]

   so uniform

4. event prob.:

   - $|A^c| = N(N-1)\ldots(N-m+1) = N!/(N-m)!$
   - $|S| = N^m$

   so $\frac{N!}{N^m (N-m)!}$

Claim: Stirling approx: $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$

a lot of math... $\Pr[\text{2 of 23 share bday}] > 1/2$.

- 0.4 for $n = 20$
- 0.7 for $n = 30$

- 0.999998876 for $n = 100$

[[*Poll class to see how many people have same bday.*]]

Alternatively:

1. sample space: $S = \{(b_1, \ldots, b_m)\}$
2. events:
   $B_i =$ event $b_i \notin \{b_1, \ldots, b_{i-1}\}$

$$\begin{aligned} \Pr[A^c] &= \Pr[B_1 \cap \ldots \cap B_m] \\ &= \Pr[B_1] \ldots \Pr[B_n | B_1, \ldots, B_{n-1}] \end{aligned}$$

3. outcome prob.: uniform
4. event prob.:
   $\Pr[B_i | \cap_{j<i} B_j] = 1 - \Pr[B_i^c | \cap_{j<i} B_j] = 1 - (i-1)/N$

Claim: $(1-x) \leq e^{-x}$ for all x (good approx. if x close to 0)

$$\begin{aligned} \Pr[A^c] &= \Pr[B_1] \ldots \Pr[B_m | B_1, \ldots, B_{m-1}] \\ &= (1 - 0/d) ... (1 - (n-1)/d) \\ &\leq e^{-0/d} e^{-1/d} ... e^{-(m-1)/N} \\ &= e^{-m^2/2N} \end{aligned}$$

Note: Constant prob. of collision for $m \geq \sqrt{2N}$.

Note: Hashing:

- want to store m records
- using N keys
- function h maps record to key
- *if* h maps randomly, need $N = m^2$-sized array to avoid collisions