independence

Defn: Two events E and F are *independent* if

$P(EF) = P(E)\,P(F)$

If $P(F)>0$, this is equivalent to: $P(E|F) = P(E)$

Otherwise, they are called *dependent*

Roll two dice, yielding values D_1 and D_2

1) $E = \{ D_1 = 1 \}$

$F = \{ D_2 = 1 \}$

$P(E) = 1/6$, $P(F) = 1/6$, $P(EF) = 1/36$

$P(EF) = P(E) \cdot P(F) \Rightarrow$ E and F *independent*

Intuitive; the two dice are not physically coupled

2) $G = \{D_1 + D_2 = 5\} = \{(1,4),(2,3),(3,2),(4,1)\}$

$P(E) = 1/6$, $P(G) = 4/36 = 1/9$, $P(EG) = 1/36$

not independent!

E, G are *dependent* events

The dice are still not physically coupled, but "$D_1 + D_2 = 5$" couples them <u>mathematically</u>: info about D_1 constrains D_2. (But dependence/independence not always intuitively obvious; "use the definition, Luke".)

Two events E and F are *independent* if

$P(EF) = P(E)\,P(F)$

If $P(F)>0$, this is equivalent to: $P(E|F) = P(E)$

Otherwise, they are called *dependent*

Three events E, F, G are independent if

$P(EF) = P(E)\,P(F)$

$P(EG) = P(E)\,P(G)$ *and* $P(EFG) = P(E)\,P(F)\,P(G)$

$P(FG) = P(F)\,P(G)$

Example: Let X, Y be each {-1,1} with equal prob

$E = \{X = 1\}$, $F = \{Y = 1\}$, $G = \{XY = 1\}$

$P(EF) = P(E)P(F)$, $P(EG) = P(E)P(G)$, $P(FG) = P(F)P(G)$

but *P(EFG) = 1/4* !!! (because $P(G|EF) = 1$)

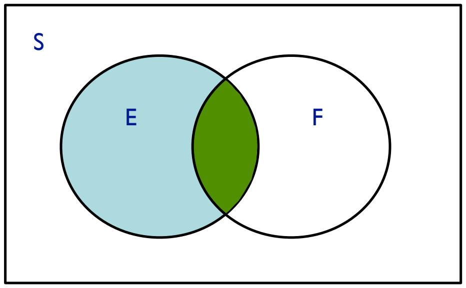
In general, events E_1, E_2, …, E_n are independent if for *every subset* S of {1,2,…, n}, we have

$$P\left(\bigcap_{i \in S} E_i\right) = \prod_{i \in S} P(E_i)$$

(Sometimes this property holds only for small subsets S. E.g., E, F, G on the previous slide are *pairwise* independent, but not fully independent.)

Theorem: E, F independent $\Rightarrow$ E, F^c independent

E = EF ∪ EF^c



Proof: $P(EF^c) = P(E) - P(EF)$
$= P(E) - P(E)\,P(F)$
$= P(E)\,(1-P(F))$
$= P(E)\,P(F^c)$

Theorem: if $P(E)>0$, $P(F)>0$, then
E, F independent $\Leftrightarrow P(E|F)=P(E) \Leftrightarrow P(F|E) = P(F)$

Proof: Note $P(EF) = P(E|F)\,P(F)$, regardless of in/dep.
Assume independent. Then

$P(E)P(F) = P(EF) = P(E|F)\,P(F) \Rightarrow P(E|F)=P(E)$ (÷ by P(F))

Conversely, $P(E|F)=P(E) \Rightarrow P(E)P(F) = P(EF)$ (× by P(F))

Suppose a biased coin comes up heads with probability p, *independent* of other flips

P(n heads in n flips) = p^n

P(n tails in n flips) = $(1-p)^n$

P(exactly k heads in n flips) $= \binom{n}{k} p^k (1-p)^{n-k}$

Aside: note that the probability of *some* number of heads = $\sum_k \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1$ as it should, by the binomial theorem.

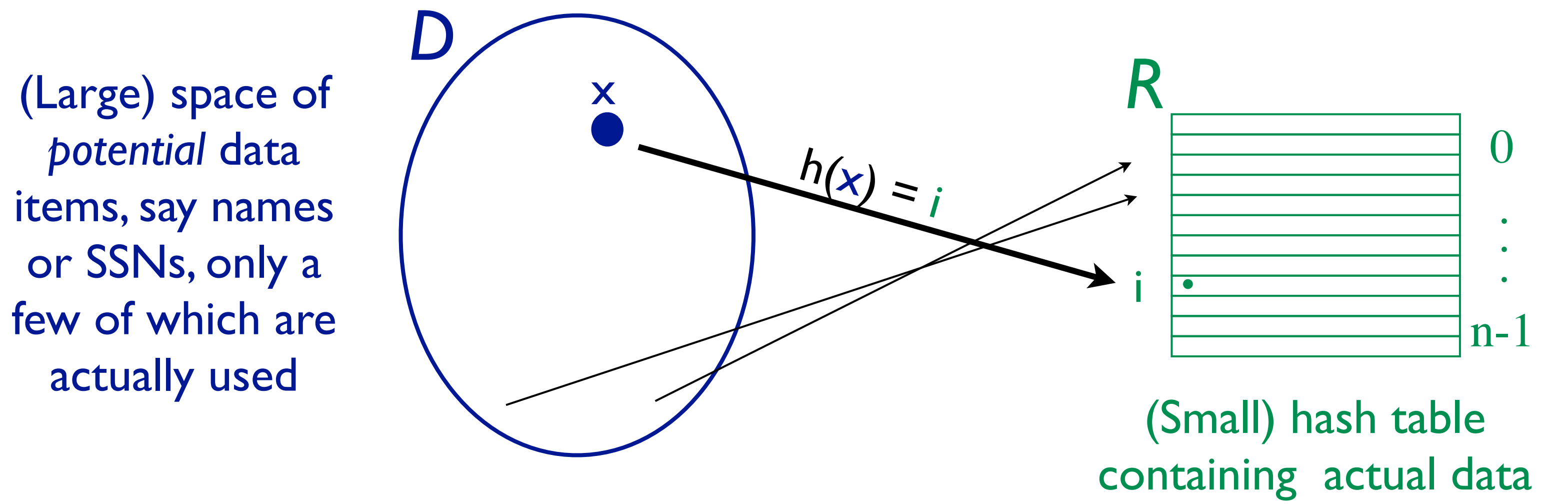
Suppose a biased coin comes up heads with probability p, *independent* of other flips

$$\text{P(exactly k heads in n flips)} = \binom{n}{k} p^k (1-p)^{n-k}$$

Note when p=1/2, this is the same result we would have gotten by considering n flips in the "equally likely outcomes" scenario. But $p \neq 1/2$ makes that inapplicable. Instead, the *independence* assumption allows us to conveniently assign a probability to each of the 2^n outcomes, e.g.:

$$\Pr(\text{HHTHTTT}) = p^2(1-p)p(1-p)^3 = p^{\#H}(1-p)^{\#T}$$

A data structure problem: *fast* access to *small* subset of data drawn from a *large* space.



A solution: *hash function* h:D→{0,...,n-1} crunches/scrambles names from large space into small one. E.g., if x is integer:

h(x) = x mod n

Good hash functions *approximately* randomize placement.

m strings hashed (uniformly) into a table with n buckets

Each string hashed is an *independent* trial

E = at least one string hashed to first bucket

What is P(E) ?

Solution:

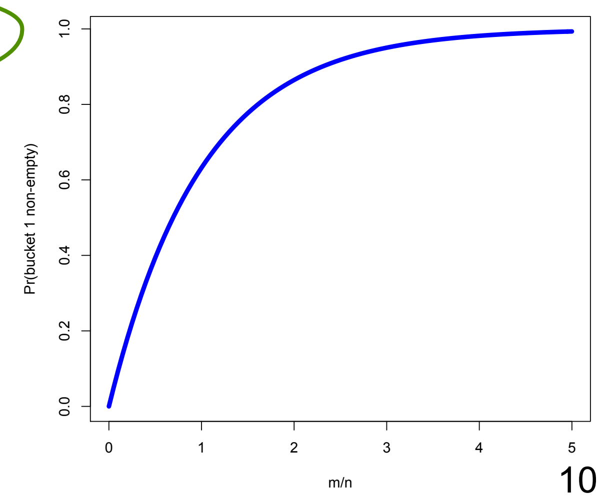
F_i = string i *not* hashed into first bucket (i=1,2,…,m)

$P(F_i) = 1 - 1/n = (n-1)/n$ for all i=1,2,…,m

Event $(F_1\ F_2\ \ldots\ F_m)$ = no strings hashed to first bucket

$$P(E) = 1 - P(F_1\ F_2 \cdots F_m)$$
$$= 1 - P(F_1)\ P(F_2) \cdots P(F_m)$$
$$= 1 - ((n-1)/n)^m$$
$$\approx 1-\exp(-m/n)$$

indp



m strings hashed (non-uniformly) to table w/ n buckets

Each string hashed is an *independent* trial, with probability p_i of getting hashed to bucket i

E = At least 1 of buckets 1 to k gets ≥ 1 string

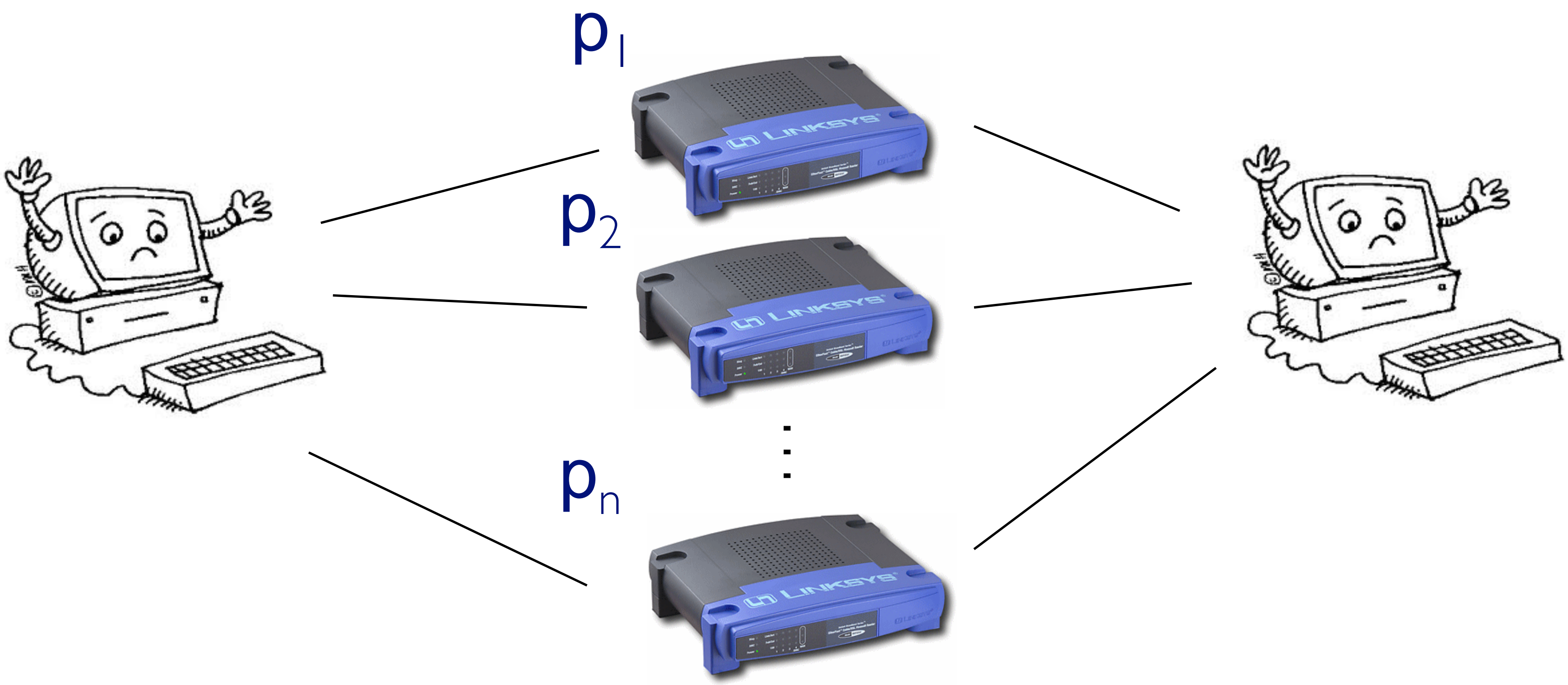
What is P(E) ?

Solution:

F_i = at least one string hashed into i-th bucket

$$P(E) = P(F_1 \cup \cdots \cup F_k) = 1 - P((F_1 \cup \cdots \cup F_k)^c)$$
$$= 1 - P(F_1^c F_2^c \ldots F_k^c)$$
$$= 1 - P(\text{no strings hashed to buckets 1 to k})$$
$$= 1 - (1 - p_1 - p_2 - \cdots - p_k)^m$$

Consider the following parallel network

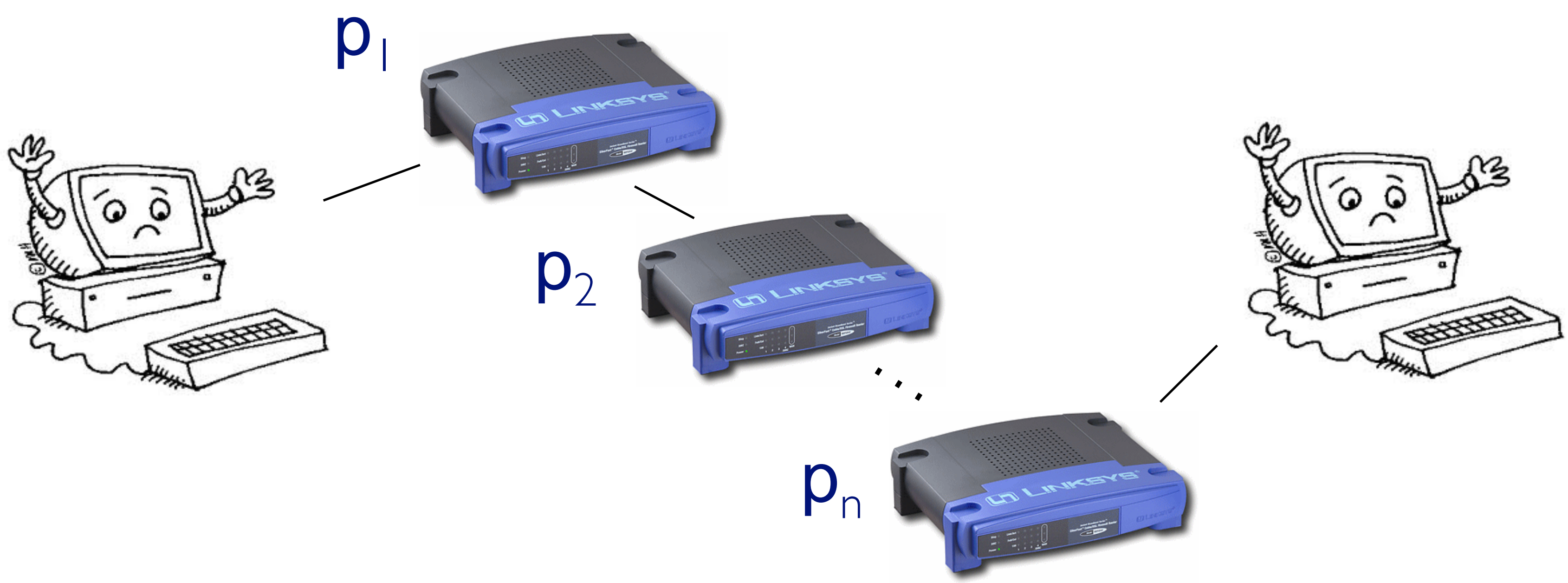


n routers, i[th] has probability p_i of failing, independently

P(there is functional path) = 1 – P(all routers fail)

$$= 1 - p_1 p_2 \cdots p_n$$

Contrast: a *series* network



n routers, i^{th} has probability p_i of failing, independently

P(there is functional path) =

$$P(\textit{no} \text{ routers fail}) = (1 - p_1)(1 - p_2) \cdots (1 - p_n)$$

Recall: Two events E and F are independent if

$$P(EF) = P(E)\,P(F)$$

If E & F are independent, does that tell us anything about

$P(EF|G)$, $P(E|G)$, $P(F|G)$,

when G is an arbitrary event? In particular, is

$$P(EF|G) = P(E|G)\,P(F|G) \text{ ?}$$

In general, *no*.

Roll two 6-sided dice, yielding values D_1 and D_2

$E = \{ D_1 = 1 \}$

$F = \{ D_2 = 6 \}$

$G = \{ D_1 + D_2 = 7 \}$

E and F are independent

$P(E|G) = 1/6$

$P(F|G) = 1/6$, but

$P(EF|G) = 1/6$, *not* $1/36$

so E|G and F|G are not independent!

Two events E and F are called *conditionally independent given G*, if

$P(EF|G) = P(E|G)\,P(F|G)$

Or, equivalently (assuming P(F)>0, P(G)>0),

$P(E|FG) = P(E|G)$

Say you are in a dorm with 100 students

- 10 are CS majors: P(C) = 0.1
- 30 get straight A's: P(A) = 0.3
- 3 are CS majors who get straight A's
- P(CA) = 0.03
- P(CA) = P(C) P(A), so C and A *independent*

At faculty night, only CS majors and A students show up

- So 37 students arrive
- Of 37 students, 10 are CS ⇒
  - $P(C \mid C \text{ or } A) = 10/37 = 0.27 < .3 = P(A)$
- Seems CS major *lowers* your chance of straight A's ☹
- Weren't they supposed to be independent?

In fact, CS and A are *conditionally dependent* at fac night

Randomly choose a day of the week

A = { It is not a Monday }

B = { It is a Saturday }

C = { It is the weekend }

A and B are dependent events

P(A) = 6/7, P(B) = 1/7, P(AB) = 1/7.

Now condition both A and B on C:

P(A|C) = 1, P(B|C) = ½, P(AB|C) = ½

P(AB|C) = P(A|C) P(B|C) ⇒ A|C and B|C independent

Dependent events can become independent by conditioning on additional information!

Events E & F are *independent* if

P(EF) = P(E) P(F), or, equivalently P(E|F) = P(E) (if p(E)>0)

More than 2 events are indp if, for *all subsets*, joint probability = product of separate event probabilities

Independence can greatly simplify calculations

For fixed G, conditioning on G gives a probability measure, P(E|G)

But "conditioning" and "independence" are orthogonal:

Events E & F that are (unconditionally) independent may become dependent when conditioned on G

Events that are (unconditionally) dependent may become independent when conditioned on G