Probability space

Sample space: S = set of all potential outcomes of experiment

E.g., flip two coins: S = {(H,H), (H,T), (T,H), (T,T)}

Events: $\mathbf{E} \subseteq \mathbf{S}$ is an arbitrary subset of the sample space

≥ 1 head in 2 flips: E = {(H,H), (H,T), (T,H)} S =

Probability:

A function from subsets of S to real numbers – Pr: $2^S \rightarrow [0,1]$

Probability Axioms:

Axiom 1 (Non-negativity): $0 \leq \Pr(E)$

Axiom 2 (Normalization): $\Pr(S) = 1$

Axiom 3 (Additivity): $EF = \emptyset \Rightarrow \Pr(E \cup F) = \Pr(E) + \Pr(F)$

1

equally likely outcomes

Simplest case: sample spaces with equally likely outcomes.

Coin flips: S = {Heads, Tails}

Flipping two coins: S = {(H,H),(H,T),(T,H),(T,T)}

Roll of 6-sided die: S = {1, 2, 3, 4, 5, 6}

Pr(each outcome) = $\dfrac{1}{|S|}$ uniform distribution

In that case,

$$\Pr(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S} = \frac{|E|}{|S|}$$

conditional probability & chain rule

General defn: $P(E \mid F) = \dfrac{P(EF)}{P(F)}$ where P(F) > 0

Implies: P(EF) = P(E|F) P(F) *("the chain rule")*

General definition of Chain Rule:

$$P(E_1 E_2 \cdots E_n) = P(E_1)P(E_2 \mid E_1)P(E_3 \mid E_1, E_2) \cdots P(E_n \mid E_1, E_2, \ldots, E_{n-1})$$

3

Law of total probability

$P(E) = P(EF) + P(EF^c)$
$= P(E|F)\,P(F) + P(E|F^c)\,P(F^c)$
$= P(E|F)\,P(F) + P(E|F^c)\,(1-P(F))$

weighted average, conditioned on event F happening or not.

More generally, if $F_1, F_2, \ldots, F_n$ partition S (mutually exclusive, $\bigcup_i F_i = S$, $P(F_i)>0$), then

$P(E) = \sum_i P(E|F_i)\,P(F_i)$

weighted average, conditioned on events F_i happening or not.

(Analogous to reasoning by cases; both are very handy.)

4

Bayes Theorem

Most common form:

$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E)}$$

Expanded form (using law of total probability):

$$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid F^c)P(F^c)}$$

Proof:

$$P(F \mid E) = \frac{P(EF)}{P(E)} = \frac{P(E \mid F)P(F)}{P(E)}$$

5

Independence

Independence of events

Intuition: E is independent of F if the chance of E occurring is not affected by whether F occurs.

Formally:

$Pr(E|F) = Pr(E)$ or $Pr(E \cap F) = Pr(E)Pr(F)$

These two definitions are equivalent.

7

Independence

Draw a card from a shuffled deck of 52 cards.

E: card is a spade

F: card is an Ace

Are E and F independent?

8

Independence

Toss a coin 3 times. Each of 8 outcomes equally likely.
Define

A = {at most one T} = {HHH, HHT, HTH, THH}

B = {both H and T occur}= $\{HHH, TTT\}^c$

Are A and B independent?

9

Independence as an assumption

It is often convenient to **assume** independence.

People often assume it without noticing.

Example: A sky diver has two chutes. Let

E = {main chute doesn't open} Pr (E) = 0.02

F = {backup doesn't open} Pr (F) = 0.1

What is the chance that at least one opens assuming independence?

10

Independence as an assumption

It is often convenient to **assume** independence.

People often assume it without noticing.

Example: A sky diver has two chutes. Let

E = {main chute doesn't open} Pr (E) = 0.02

F = {backup doesn't open} Pr (F) = 0.1

What is the chance that at least one opens assuming independence?

Note: Assuming independence doesn't justify the assumption! Both chutes could fail because of the same rare event, e.g. freezing rain.

11

Using independence to define a probabilistic model

We can **define** our probability model via independence.

Example: suppose a biased coin comes up heads with probability 2/3, independent of other flips.

Sample space: sequences of 3 coin tosses.

Pr (3 heads)=?

Pr (3 tails) = ?

Pr (2 heads) = ?

12

biased coin

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

P(n heads in n flips)
P(n tails in n flips)
P(HHTHTTT)
P(exactly k heads in n flips)

13

biased coin

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) = $(1-p)^n$

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

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.

14

biased coin

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

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

How does this compare to $p=1/2$ case?

15

biased coin

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

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

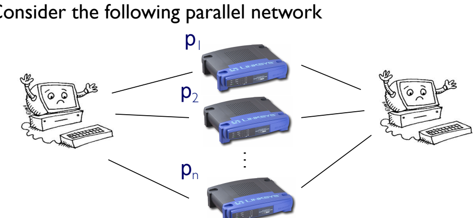
Note when $p=1/2$, this is the same result we would have gotten by considering *n* flips in the "equally likely outcomes" scenario. But $p\neq 1/2$ makes that inapplicable. Instead, the *independence* assumption allows us to conveniently assign a probability to each of the 2^n outcomes, e.g.:

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

16

network failure

Consider the following parallel network



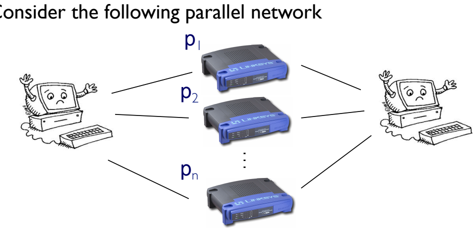
n routers, i^{th} has probability p_i of failing, independently

P(there is functional path) =

17

network failure

Consider the following parallel network

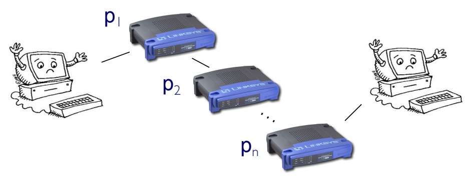


n routers, i^{th} has probability p_i of failing, independently
P(there is functional path) = 1 – P(all routers fail)

= $1 - p_1 p_2 \cdots p_n$

18

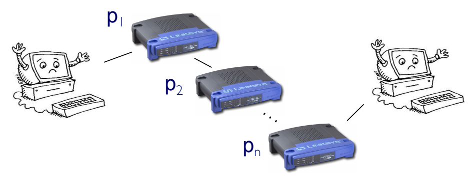
Contrast: a *series* network



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

P(there is functional path) =

Contrast: a *series* network



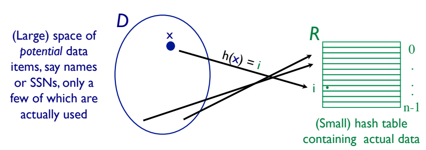
n routers, i^{th} has probability p_i of failing, independently

P(there is functional path) =

P(*no* routers fail)

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

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



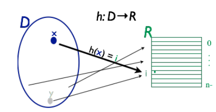
A solution: *hash function* h:D→{0,...,n-1} crunches/scrambles names from large space into small one.

E.g., if x is integer: h(x) = x mod n

Everything that hashes to same location stored in linked list.

Good hash functions *approximately* randomize placement.

Scenario: Hash m≤n keys from D into size n hash table.



How well does it work?

Worst case: *All* collide in one bucket. (Perhaps too pessimistic?)

Best case: *No* collisions. (Perhaps too optimistic?)

A middle ground: Probabilistic analysis.

Below, for simplicity, assume

- Keys drawn from D randomly, independently (with replacement)
- h maps equal numbers of domain points into each range bin, i.e., |D| = k|R| for some integer k, and $|h^{-1}(i)| = k$ for all $0 \le i \le n-1$

Many possible questions; a few analyzed below

m keys hashed (uniformly) into a hash table with n buckets

Each key hashed is an *independent* trial

E = at least one key hashed to first bucket

What is P(E) ?

Solution:

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

$P(F_i)$ =?

Event (F_1 F_2 … F_m) = no keys hashed to first bucket

P(E)= ?

m keys hashed (uniformly) into a hash table with n buckets

Each key hashed is an *independent* trial

E = at least one key hashed to first bucket

What is P(E) ?

Solution:

F_i = key 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 … F_m) = no keys hashed to first bucket

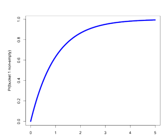
P(E)

$= 1 - P(F_1 F_2 \cdots F_m)$

$\overset{\text{indp}}{=} 1 - P(F_1)\, P(F_2) \cdots P(F_m)$

$= 1 - ((n-1)/n)^m$

$\approx 1\text{-}\exp(-m/n)$



hashing

m keys 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 first k buckets gets ≥ 1 key
What is P(E) ?
Solution:
F_i = at least one key hashed into i-th bucket
P(E) =

25

hashing

m keys 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 first k buckets gets ≥ 1 key
What is P(E) ?
Solution:
F_i = at least one key 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(no strings hashed to buckets 1 to k)
$= 1 - (1 - p_1 - p_2 - \cdots - p_k)^m$

26

Perfect hashing (i)

Let $|R| = n$, $D_0 \subseteq D$, $|D_0| = m$. A hash function h: $D \rightarrow R$ is *perfect* for D_0 if h: $D_0 \rightarrow R$ is injective (no collisions). How likely is that?

1) Fix h; pick m elements of D_0 *independently* at random $\in D$

P(h is perfect for D_0) =

27

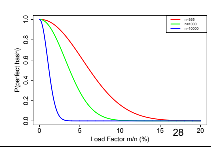
Perfect hashing (i)

Let $|R| = n$, $D_0 \subseteq D$, $|D_0| = m$. A hash function h: $D \rightarrow R$ is *perfect* for D_0 if h: $D_0 \rightarrow R$ is injective (no collisions). How likely is that?

1) Fix h; pick m elements of D_0 *independently* at random $\in D$
Again, suppose h maps $(1/n)^{th}$ of D to each element of R. This is like the birthday problem:

P(h is perfect for D_0) = $\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-m+1}{n}$

Except for very empty tables, a "perfect" hash is improbable



28

If E and F are independent,
then so are E and F^c
and so are E^c and F
and so are E^c and F^c

29

If E and F are independent,
then so are E and F^c
and so are E^c and F
and so are E^c and F^c

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

30

Independence of several events

Three events E, F, G are mutually independent if

$Pr(E \cap F) = Pr(E)Pr(F)$

$Pr(F \cap G) = Pr(F)Pr(G)$

$Pr(E \cap G) = Pr(E)Pr(G)$

$Pr(E \cap F \cap G) = Pr(E)Pr(F)Pr(G)$

31

Pairwise independent

E, F and G are pairwise independent if E is independent of F, F is independent of G, and E is independent of G.

Example: Toss a coin twice.

E = {HH, HT}

F = {TH, HH}

G = {HH, TT}

These are pairwise independent, but not mutually independent.

32

Independence of several events

Three events E, F, G are mutually independent if

$Pr(E \cap F) = Pr(E)Pr(F)$

$Pr(F \cap G) = Pr(F)Pr(G)$

$Pr(E \cap G) = Pr(E)Pr(G)$

$Pr(E \cap F \cap G) = Pr(E)Pr(F)Pr(G)$

If E, F and G are independent, then E will be independent of any event formed from F and G.

Example: Show that E is independent of F U G.

Pr (F U G | E) = Pr (F | E) + Pr (G | E) – Pr (FG | E)

= Pr (F) + Pr (G) - Pr (EFG)/Pr(E)

= Pr (F) + Pr (G) - Pr (FG)= Pr(F U G)

33

Summary - Independence

Events E & F are *independent* if

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

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

Dependent means correlated, associated, (partially) predictive

Independence can greatly simplify calculations

Independence can be used to define probability models.

34