Independence

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.

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?

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

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

B = {at most two Heads}= $\{HHH\}^c$

Are A and B independent?

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?

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.

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) = ?

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(HHTHTTT) $= p^3(1-p)^4$

$p\cdot p\cdot(1-p)\cdot p\cdot(1-p)^3$

P(exactly k heads in n flips)

$$= \sum_{\substack{\text{outcomes} \\ \text{w/ exactly} \\ k \text{ heads}}} p^k(1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\Pr\left(\begin{matrix} TH\ldots H\ HT\ldots T \\ \underbrace{HHH\ldots H}_{k}\ \underbrace{TTT\ldots T}_{n-k} \end{matrix}\right) = p^k(1-p)^{n-k}$$

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) $= (1-p)^n$

$\Omega = \{H,T\}^n$

Pr(HHTHTTT)

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

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

HHTTTTHHH

Aside: note that the probability of *some* number of heads = $\boxed{\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}} = (p + (1-p))^n = 1$

as it should, by the binomial theorem.

$\{H,T\}^n$

$= \sum (\Pr(\omega))$

$\omega \in \Omega$ s.t. ω has k Heads

0 H's | 1 H | 2 H's | n H's

$\sum_{k=0}^{n} Pr(k\ H's)$

biased coin

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

$\left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k}$

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

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

all outcomes equally likely $\quad Pr(\omega) = \frac{1}{2^n} \quad \{H,T\}^n$

$$Pr(\text{exactly } k\ H's) = \frac{|\text{outcomes w/ } k\ H's|}{2^n} = \frac{\binom{n}{k}}{2^n}$$

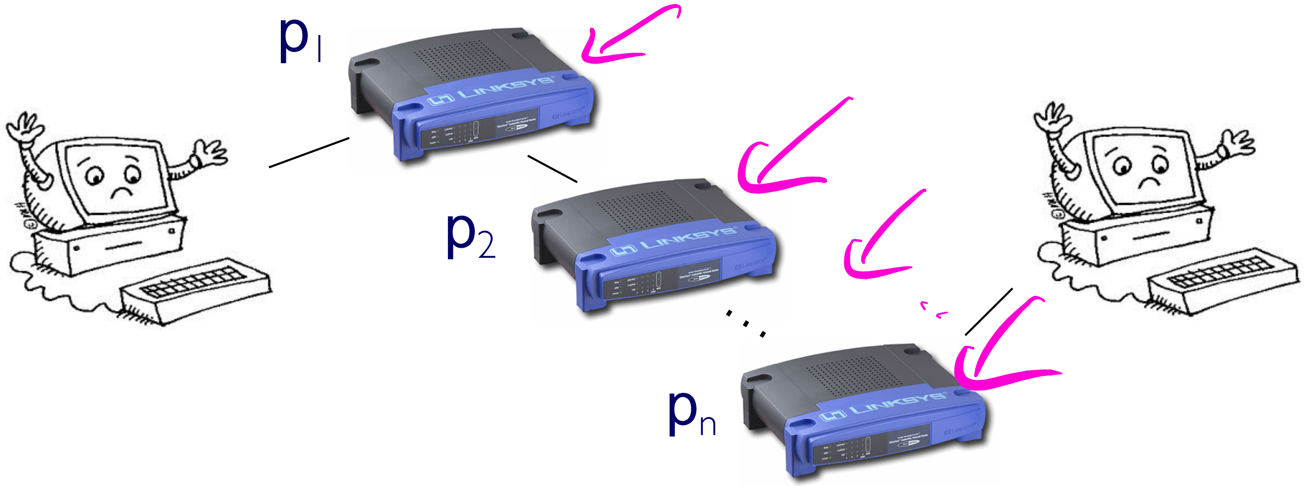
Suppose a biased coin comes up heads with probability p, *independent* of other flips

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 different from ½ 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) = $p^2(1-p)p(1-p)^3 = p^{\#H}(1-p)^{\#T}$

Contrast: a *series* network



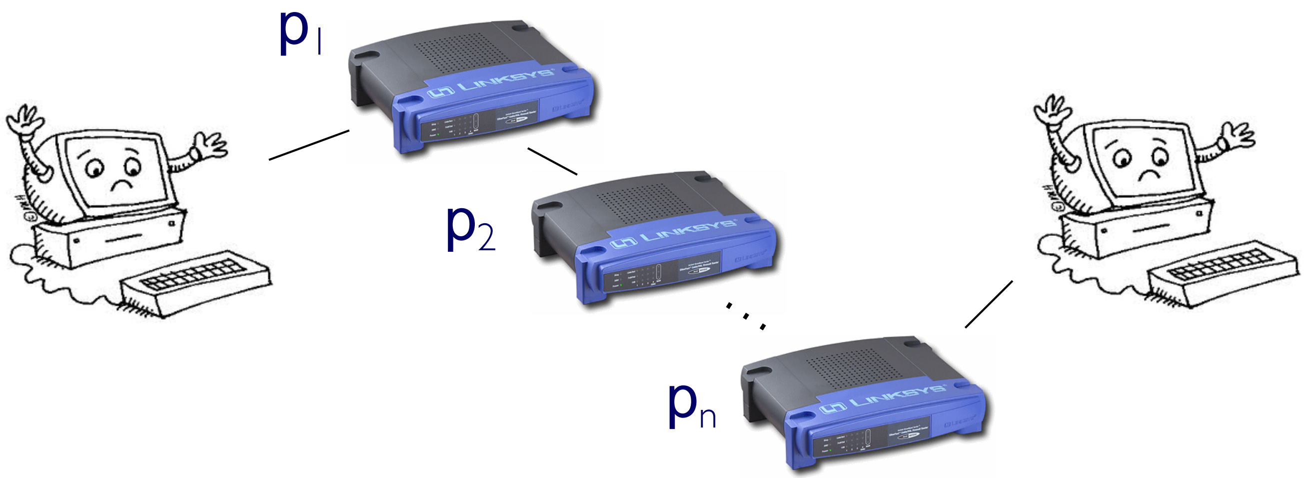
n routers, i^{th} has probability p_i of failing, independently

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

P(there is functional path) = $\prod_{i=1}^{n} (1-p_i)$

network failure

Contrast: a *series* network



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

P(there is functional path) =

P(*no* routers fail)

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

choosing random key

$Pr(x \in D \text{ selected}) = \frac{1}{|D|}$

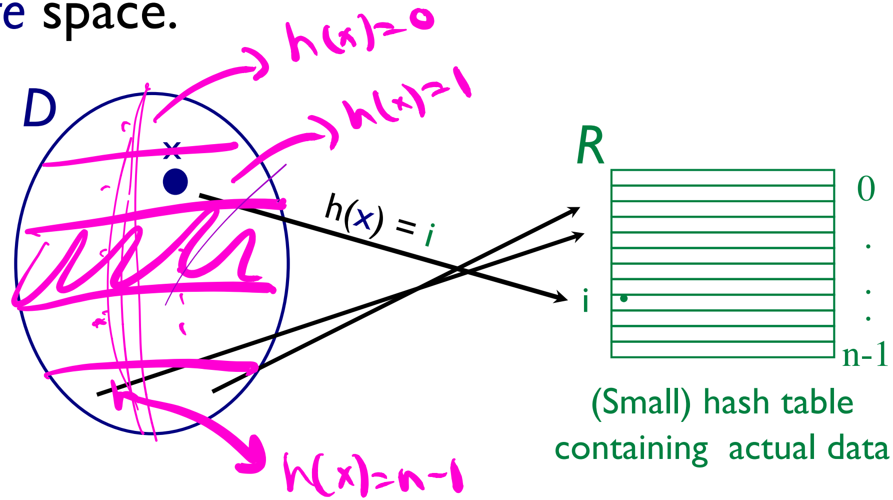
$Pr(\text{selected } x \text{ has } h(x) = i) = \frac{\#x\text{'s s.t. } h(x)=i}{|D|} = \frac{|D|/n}{|D|} = \frac{1}{n}$

random exp: selecting random $x \in D$

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

2^{128}

(Large) space of *potential* data items, say names or SSNs or IP addresses, only a few of which are actually used



(Small) hash table containing actual data

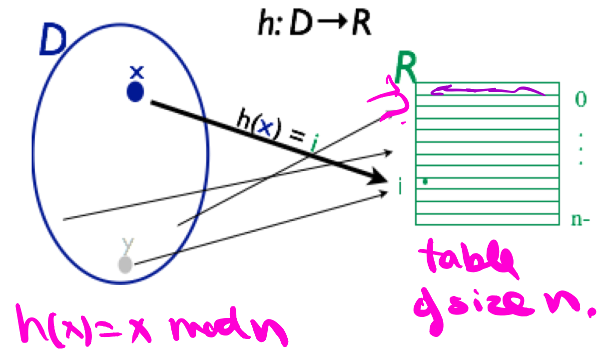
A solution: *hash function* h:D→{0,...,n-1} crunches/scrambles names from large space into small one.

E.g., if x is integer: $h(x) = x \bmod n$

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 $0 \leq i \leq n-1$

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 $\longrightarrow \forall i\ Pr(\text{goes into bin } i) = \frac{1}{n}$

E = at least one key hashed to first bucket

What is P(E) ?

$$Pr(\text{at least one key} \to 1^{st} \text{ bucket}) = 1 - \underbrace{Pr(\text{no keys} \to \text{first bucket})}$$

$$= 1 - \prod_{j=1}^{m} \underbrace{Pr(j^{th} \text{ key selected does not go to first bucket})}_{1-\frac{1}{n}}$$

$$= \underbrace{1 - \left(1 - \frac{1}{n}\right)}_{1+\left(-\frac{1}{n}\right)}^{m} \leq 1 - \left(e^{-\frac{1}{n}}\right)^{m} = 1 - e^{-\frac{m}{n}}$$

$$e^{x} = 1 + x + \frac{x^2}{2} + \dots$$

$e^x \geq 1+x$

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 $(F_1 F_2 \dots F_m)$ = no keys hashed to first bucket

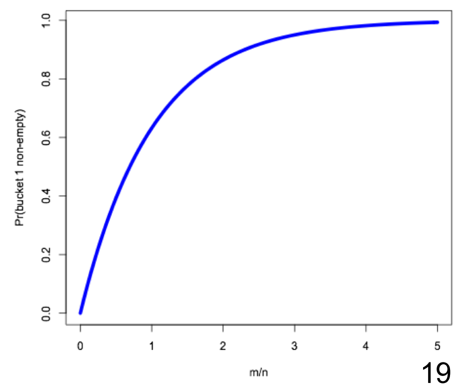
P(E)

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

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

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

$$\approx 1 - \exp(-m/n)$$



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

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

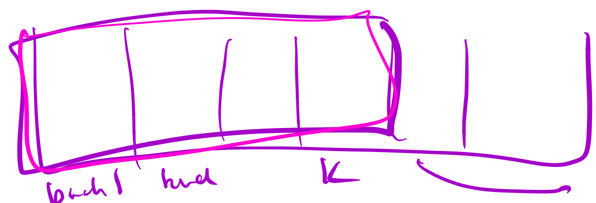
E = At least 1 of first k buckets gets ≥ 1 key

What is P(E) ?

$\sum p_i = 1$

$\Pr(\text{at least 1 of first k buckets gets} \geq 1 \text{ key})$

$= 1 - \Pr(\text{none of first k buckets gets} \geq 1 \text{ key})$



$= 1 - \prod_{j=1}^{m} \Pr(j^{th} \text{ key does not go into one of first k buckets})$

$= 1 - \left(\sum^{n} p_\ell \right)^m$

$\Pr(\text{j}^{\text{th}}\text{ key goes into one of first k buckets}) = p_1 + p_2 + \dots + p_k$

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

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

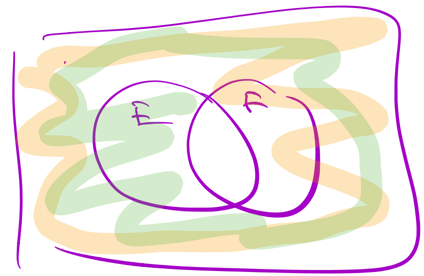
$F^c = \overline{F}$

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



$$
\begin{aligned}
Pr(E^c \cap F^c) &= 1 - Pr(E \cup F) \\
&= 1 - [Pr(E) + Pr(F) - Pr(E \cap F)] \\
&= 1 - Pr(E) - Pr(F) + Pr(E)Pr(F) \\
&= (1 - Pr(E))(1 - Pr(F)) \\
&= Pr(E^c)Pr(F^c)
\end{aligned}
$$

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: $P(EF^c) = P(E) - P(EF)$

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

$= P(E)\,(1-P(F))$

$= P(E)\,P(F^c)$

Independence of several events

Three events E, F, G are mutually independent if

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

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

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

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

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}

$Pr(E \cap F) = Pr(HH) = \frac{1}{4}$

$Pr(E)Pr(F) = \frac{1}{2} \cdot \frac{1}{2}$

$\underset{\frac{1}{4}}{Pr(E \cap F \cap G)} \neq \underset{(\frac{1}{2})^3}{Pr(E)Pr(F)Pr(G)}$

These are pairwise independent, but not mutually independent.

Independence of several events

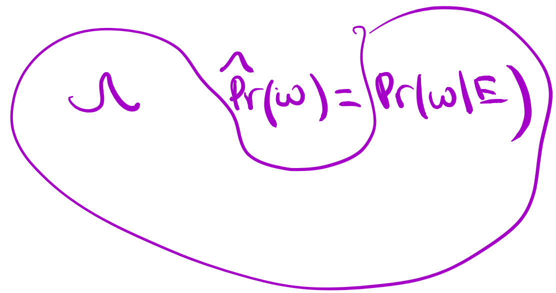
Three events E, F, G are mutually independent if

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

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

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

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



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

Example: 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)

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

$$P(EF|G) = P(E|G) \cdot P(F|G)\ ?$$

In general, *no*.

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

$E = \{ D_1 = 1 \}$ $\Pr(E) = \frac{1}{6}$

$F = \{ D_2 = 6 \}$ $\Pr(F) = \frac{1}{6}$

$G = \{ D_1 + D_2 = 7 \}$ $\Pr(G) = \frac{1}{6} = \Pr(\{(1,6)(2,5)(3,4)\cdots(6,1)\})$

E and F are independent

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

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

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

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

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!

$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$ $\Pr(A|B)\Pr(B) = \Pr(A \cap B)$

Definition:

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

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

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

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

$$\Pr(E|FG) = \frac{\Pr(E \cap F \cap G)}{\Pr(F \cap G)} = \frac{\Pr(E \cap F|G)\Pr(G)}{\Pr(F \cap G)}$$

$$= \frac{\Pr(E|G)\Pr(F|G)\Pr(G)}{\Pr(F \cap G)} = \Pr(E|G)$$

conditioning can also break DEPENDENCE

Randomly choose a day of the week

A = { It is not a Monday } $Pr(A) = 6/7$

B = { It is a Saturday } $Pr(B) = 1/7$

C = { It is the weekend } $Pr(C) = 2/7$

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:

$Pr(A|C) = 1$ $Pr(B|C) = \frac{1}{2}$ $Pr(A \cap B|C) = \frac{1}{2}$

conditioning can also break DEPENDENCE

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:

$P(A|C) = 1$, $P(B|C) = \frac{1}{2}$, $P(AB|C) = \frac{1}{2}$

$P(AB|C) = P(A|C)\,P(B|C) \Rightarrow A|C$ and $B|C$ independent

Dependent events can become independent by conditioning on additional information!

Another reason why conditioning is so useful

- 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
- Dependent means correlated, associated, (partially) predictive
- Independence can be used to define probability models.
- 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

Problem

In a group of N people 15% are left-handed.

Suppose that 100 times you pick a random person (each person is picked each time with probability 1/N) and ask that person if they are left-handed or not.

What is the probability that among the 100 queries, 55 people are left-handed?

$$\binom{100}{55}(0.15)^{55}(0.85)^{45}$$

You have 50 pairs of socks. No two have the same color and pattern. You reach in to your drawer and grab 5 random socks. What is the probability that there is a pair among the 5?

Case 1: the left and right sock from each pair are distinguishable. (i.e., all 100 socks are distinguishable).

Problem

You have 50 pairs of socks. No two have the same color and pattern. You reach in to your drawer and grab 5 random socks one at a time. What is the probability that there is a pair among the 5?

Case 1: the left and right sock from each pair are distinguishable.

$$1 - \frac{2^5 \binom{50}{5}}{\binom{100}{5}}$$

You have 50 pairs of socks. No two have the same color and pattern. You reach in to your drawer and grab 5 random socks one at at time. What is the probability that there is **no** pair among the k?

Case 1: the left and right sock from each pair are not distinguishable.

You have 10 pairs of socks. No two have the same color and pattern. You reach in to your drawer and grab 5 random socks one at a time. What is the probability that there is **no** pair among the k?

Case 1: the left and right sock from each pair are not distinguishable.

Pr (no pair in 1st) Pr (no pair in 1st and 2nd | no pair in 1st) Pr (3rd diff | 1st and 2nd diff) Pr (4th diff | 1st – 3rd diff) Pr (5th diff | 1st - 4th diff)

$$= \frac{1}{1} \cdot \frac{18}{19} \cdot \frac{16}{18} \cdot \frac{14}{17} \cdot \frac{12}{16}$$

Problem

Toss a red die and a green die. What is the probability that the sum mod 6 is 4 given that the green die shows a 5?

$$Pr((R+G) \bmod 6 = 4)|G=5) =$$

$$\frac{Pr(G=5 \text{ and } (R+G) \bmod 6 = 4)}{Pr(G=5)} = \frac{1}{6}$$

$$Pr(G=5 \text{ and } R = -1 \bmod 6) = Pr(G=5 \text{ and } R=5) = \frac{1}{36}$$