DISCRETE PROBABILITY

Ω is a finite or countable set – called the *Probability Space*

$\mathbf{P} : \Omega \to \mathbf{R}^+$.

$\sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1$.

If $\omega \in \Omega$ then $\mathbf{P}(\omega)$ is the *probability* of ω.

Fair Coin
$\Omega = \{H, T\}, \mathbf{P}(H) = \mathbf{P}(T) = 1/2$.

Dice
$\Omega = \{1, 2, \ldots, 6\}, \mathbf{P}(i) = 1/6,\ 1 \leq i \leq 6$.

Both are examples of a *uniform distribution*:

$$\mathbf{P}(\omega) = \frac{1}{|\Omega|} \qquad \forall \omega \in \Omega.$$

Geometric or number of Bernouilli trials until success

$\Omega = \{1, 2, \ldots, \}, \mathbf{P}(k) = (1-p)^{k-1}p, \quad k \in \Omega$.

Repeat "experiment" until success – k is the total number of trials.
p is the probability of success S and $1-p$ is the probability of failure F.

$\mathbf{P}(S) = p, \mathbf{P}(FS) = p(1-p),$
$\mathbf{P}(FFS) = (1-p)^2p, \mathbf{P}(FFFS) = (1-p)^3p \ldots,$.

Note that

$$\sum_{k=1}^{\infty}(1-p)^{k-1}p = \frac{p}{1-(1-p)} = 1.$$

Roll Two Dice

Probability Space 1:
$\Omega = [6]^2 = \{(x_1, x_2) : \ 1 \leq x_1, x_2 \leq 6\}$
$\mathbf{P}(x_1, x_2) = 1/36$ for all x_1, x_2.

Probability Space 2:
$\Omega = \{2, 3, 4, , \ldots, 12\}$
$\mathbf{P}(2) = 1/36$, $\mathbf{P}(3) = 2/36$, $\mathbf{P}(4) = 3/36, \ldots, \mathbf{P}(12) = 1/36$.

Events

$A \subseteq \Omega$ is called an *event*.

$$\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\omega).$$

(i) **Two Dice**
$A = \{x_1 + x_2 = 7\}$
where x_i is the value of dice i.
$A = \{(1,6),(2,5),\ldots,(6,1)\}$ and so

$$\mathbf{P}(A) = 1/6.$$

(ii) **Pennsylvania Lottery**

Choose 7 numbers I from$[80]$. Then the state randomly chooses $J \subseteq [80]$, $|J| = 11$.

$$WIN = \{J : J \supseteq I\}.$$

$\Omega = \{11$ element subsets of $[80]\}$ with uniform distribution.
$|WIN|$ = number of subsets which contain I – $\binom{73}{4}$.

$$\begin{aligned}\mathbf{P}(WIN) &= \frac{\binom{73}{4}}{\binom{80}{11}} = \frac{\binom{11}{7}}{\binom{80}{7}} \\ &= \frac{9}{86637720} \approx \frac{1}{9,626,413}.\end{aligned}$$

Poker

Choose 5 cards at random. $|\Omega| = \binom{52}{5}$, uniform distribution.

(i) **Triple** – 3 cards of same value e.g. Q,Q,Q,7,5.
$\mathbf{P}(\text{Triple}) = (13 \times 4 \times 48 \times 44/(2\binom{52}{5})) \approx .021$.

(ii) **Full House** – triple plus pair e.g. J,J,J,7,7.
$\mathbf{P}(\text{FullHouse}) = (13 \times 4 \times 12 \times 6)/\binom{52}{5} \approx .007$.

(iii) **Four of kind** – e.g. 9,9,9,9,J.
$\mathbf{P}(\text{Four of Kind}) = (13 \times 48)/\binom{52}{5} = 1/16660 \approx .00006$.

Birthday Paradox

$\Omega = [n]^k$ – uniform distribution, $|\Omega| = n^k$.
$D = \{\omega \in \Omega$; symbols are distinct}.

$$\mathbf{P}(D) = \frac{n(n-1)(n-2)\ldots(n-k+1)}{n^k}.$$

$n = 365, k = 26$ – birthdays of 26 randomly chosen people.

$\mathbf{P}(D) < .5$ i.e. probability 26 randomly chosen people have distinct birthdays is <.5. (Assumes people are born on random days).

Balls in Boxes

m distinguishable balls in n distinguishable boxes.
$\Omega = [n]^m = \{(b_1, b_2, \ldots, b_m)\}$ where b_i denotes the box containing ball i.
Uniform distribution.

$$E = \{\text{Box 1 is empty}\}.$$

$$\begin{aligned} \mathbf{P}(E) &= \frac{(n-1)^m}{n^m} \\ &= \left(1 - \frac{1}{n}\right)^m \\ &\to e^{-c} \qquad \text{as } n \to \infty \end{aligned}$$

if $m = cn$ where $c > 0$ is *constant*.

Explanation of limit

: $(1 - 1/n)^{cn} \to e^{-c}$.

$1 + x \leq e^x$ for all x;

1. $x \geq 0$: $1 + x \leq 1 + x + x^2/2! + x^3/3! + \cdots = e^x$.
2. $x < -1$: $1 + x < 0 \leq e^x$.
3. $x = -y, 0 \leq y \leq 1$:
   $1 - y \leq 1 - y + (y^2/2! - y^3/3!) + (y^4/4! - y^5/5!) + \cdots = e^{-y}$.
4. So $(1 - 1/n)^{cn} \leq (e^{-1/n})^{cn} = e^{-c}$.

$$e^{-x-x^2} \leq 1 - x \text{ if } 0 \leq x \leq 1/100. \tag{1}$$

$$\begin{aligned}
\log_e(1-x) &= -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots \\
&\geq -x - \frac{x^2}{2} - x^2\left(\frac{x}{3} + \frac{x^2}{3} - \cdots\right) \\
&= -x - \frac{x^2}{2} - \frac{x^3}{3(1-x)} \\
&\geq -x - x^2.
\end{aligned}$$

This proves (1). So, for large n,

$$\begin{aligned}
(1-1/n)^{cn} &\geq \exp\{-cn(1/n + 1/n^2)\} \\
&= \exp\{-c - c/n\} \\
&\to \epsilon^{-c}.
\end{aligned}$$

Random Walk

A particle starts at 0 on the real line and each second makes a random move left of size 1, (probability 1/2) or right of size 1 (probability 1/2).
Consider n moves. $\Omega = \{L, R\}^n$.
For example if $n = 4$ then $LLRL$ stands for move left, move left, move right, move left.
Each sequence ω is given an equal probability 2^{-n}.
Let $\mathbf{X}_n = \mathbf{X}_n(\omega)$ denote the position of the particle after n moves.
Suppose $n = 2m$. What is the probability $\mathbf{X}_n = 0$?

$$\frac{\binom{n}{m}}{2^n} \approx \sqrt{\frac{2}{\pi n}}.$$

Stirling's Formula: $n! \approx \sqrt{2\pi n}(n/e)^n$.

Boole's Inequality

$A, B \subseteq \Omega$.

$$\begin{aligned} \mathbf{P}(A \cup B) &= \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) \\ &\leq \mathbf{P}(A) + \mathbf{P}(B) \end{aligned} \tag{2}$$

If A, B are *disjoint* events i.e. $A \cap B = \emptyset$ then

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$$

.

Example: Two Dice. $A = \{x_1 \geq 3\}$ and $B = \{x_2 \geq 3\}$.
Then $\mathbf{P}(A) = \mathbf{P}(B) = 2/3$ and

$$\mathbf{P}(A \cup B) = 8/9 < \mathbf{P}(A) + \mathbf{P}(B).$$

More generally,

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbf{P}(A_i). \qquad (3)$$

Inductive proof

Base case: $n = 1$

Inductive step: assume (3) is true.

$$\begin{aligned}
\mathbf{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &\leq \mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbf{P}(A_{n+1}) \text{ by (2)} \\
&\leq \sum_{i=1}^{n} \mathbf{P}(A_i) + \mathbf{P}(A_{n+1}) \text{ by (3)}
\end{aligned}$$

Colouring Problem

Theorem Let $A_1, A_2, \ldots, A_n$ be subsets of A and $|A_i| = k$ for $1 \le i \le n$. If $n < 2^{k-1}$ then there exists a partition $A = R \cup B$ such that

$$A_i \cap R \neq \emptyset \text{ and } A_i \cap B \neq \emptyset \qquad 1 \le i \le n.$$

[R = Red elements and B= Blue elements.]

Proof Randomly colour A.

Ω = $\{R, B\}^A = \{f : A \to \{R, B\}\}$, uniform distribution.

$$BAD = \{\exists i : A_i \subseteq R \text{ or } A_i \subseteq B\}.$$

Claim: $\mathbf{P}(BAD) < 1$.

Thus $\Omega \setminus BAD \neq \emptyset$ and this proves the theorem.

$$BAD(i) = \{A_i \subseteq R \text{ or } A_i \subseteq B\}$$

$$BAD = \bigcup_{i=1}^{n} BAD(i).$$

$$\begin{aligned}
\mathbf{P}(BAD) &\leq \sum_{i=1}^{n} \mathbf{P}(BAD(i)) \\
&= \sum_{i=1}^{n} \left(\frac{1}{2}\right)^{k-1} \\
&= n/2^{k-1} \\
&< 1.
\end{aligned}$$

Example of system which is not 2-colorable.

Let $n = \binom{2k-1}{k}$ and $A = [2k-1]$ and

$$\{A_1, A_2, \ldots, A_n\} = \binom{[2k-1]}{k}.$$

Then in any 2-coloring of $A_1, A_2, \ldots, A_n$ there is a set A_i all of whose elements are of one color.

Suppose A is partitioned into 2 sets R, B. At least one of these two sets is of size at least k (since $(k-1)+(k-1) < 2k-1$). Suppose then that $R \geq k$ and let S be any k-subset of R. Then there exists i such that $A_i = S \subseteq R$.

Tournaments

n players in a tournament each play each other i.e. there are $\binom{n}{2}$ games.

Fix some k. Is it possible that for every set S of k players there is a person w_S who beats everyone in S?

Suppose that the results of the tournament are decided by a random coin toss.

Fix S, $|S| = k$ and let $\mathcal{E}_S$ be the event that nobody beats everyone in S.

The event

$$\mathcal{E} = \bigcup_S \mathcal{E}_S$$

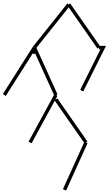
is that there is a set S for which w_S does not exist.

We only have to show that $\mathbf{Pr}(\mathcal{E}) < 1$.

$$\begin{aligned}
\mathbf{Pr}(\mathcal{E}) &\leq \sum_{|S|=k} \mathbf{Pr}(\mathcal{E}_S) \\
&= \binom{n}{k}(1-2^{-k})^{n-k} \\
&< n^k e^{-(n-k)2^{-k}} \\
&= \exp\{k \ln n - (n-k)2^{-k}\} \\
&\to 0
\end{aligned}$$

since we are assuming here that k is fixed independent of n.

Random Binary Search Trees



A binary tree consists of a set of *nodes*, one of which is the *root*. Each node is connected to 0,1 or 2 nodes below it and every node other than the root is connected to exactly one node above it. The root is the highest node.

The depth of a node is the number of edges in its path to the root.

The depth of a tree is the maximum over the depths of its nodes.

Starting with a tree T_0 consisting of a single root r, we grow a tree T_n as follows:

The n'th *particle* starts at r and flips a fair coin. It goes left (L) with probability 1/2 and right (R) with probability 1/2.

It tries to move along the tree in the chosen direction. If there is a node below it in this direction then it goes there and continues its random moves. Otherwise it creates a new node where it wanted to move and stops.

Let D_n be the depth of this tree.
Claim: for any $t \geq 0$,

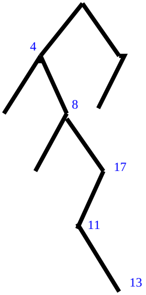
$$\mathbf{P}(D_n \geq t) \leq (n2^{-(t-1)/2})^t.$$

Proof The process requires at most n^2 coin flips and so we let $\Omega = \{L, R\}^{n^2}$ – most coin flips will not be needed most of the time.

$$DEEP = \{D_n \geq t\}.$$

For $P \in \{L, R\}^t$ and $S \subseteq [n], |S| = t$ let $DEEP(P, S)$ = {the particles $S = \{s_1, s_2, \ldots, s_t\}$ follow P in the tree i.e. the first i moves of s_i are along P, $1 \leq i \leq t$}.

$$DEEP = \bigcup_P \bigcup_S DEEP(P, S).$$



S={4,8,11,13,17}

t=5 and DEEP(P,S) occurs if
4 goes L...
8 goes LR...
17 goes LRR...
11 goes LRRL...
13 goes LRRLR...

$$
\begin{aligned}
\mathbf{P}(DEEP) &\le \sum_P \sum_S \mathbf{P}(DEEP(P,S)) \\
&= \sum_P \sum_S 2^{-(1+2+\cdots+t)} \\
&= \sum_P \sum_S 2^{-t(t+1)/2} \\
&= 2^t \binom{n}{t} 2^{-t(t+1)/2} \\
&\le 2^t n^t 2^{-t(t+1)/2} \\
&= (n2^{-(t-1)/2})^t.
\end{aligned}
$$

So if we put $t = A \log_2 n$ then

$$\mathbf{P}(D_n \ge A \log_2 n) \le (2n^{1-A/2})^{A \log_2 n}$$

which is very small, for $A > 2$.

Conditional Probability

Suppose $A \subseteq \Omega$. We define an *induced* probability $\mathbf{P}_A$ by

$$\mathbf{P}_A(\omega) = \frac{\mathbf{P}(\omega)}{\mathbf{P}(A)} \qquad \text{for } \omega \in A.$$

Usually write $\mathbf{P}(B \mid A)$ for $P_A(B)$.
If B is an arbitrary subset of Ω we write

$$\mathbf{P}(B \mid A) = \mathbf{P}_A(A \cap B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)}.$$

Two fair dice are thrown. Given that the first shows 3, what is the probability that the total exceeds 6? The answer is obviously $\frac{1}{2}$ since the second must show 4,5 or 6.
In detail: $\Omega = [6]^2$ with uniform mesaure. Let

$$\begin{aligned} A &= \{(i,j) \in [6]^2 : \; i = 3\}. \\ B &= \{(i,j) \in [6]^2 : \; i + j > 6\} \\ A \cap B &= \{(i,j) \in [6]^2 : \; i = 3, j > 3\}. \end{aligned}$$

Thus

$$\mathbf{P}(A) = \frac{1}{6}, \; \mathbf{P}(A \cap B) = \frac{3}{36} = \frac{1}{12}$$

and so

$$\mathbf{P}(A \mid B) = \frac{1/12}{1/6} = \frac{1}{2}.$$

Suppose that a family has two children and that each child is equally likely to be a Boy or a Girl. What is the probability that the family has two boys, given that it has at least one Boy.
Probability Space: {BB,BG,GB,GG}
with uniform measure, where *GB* means that the first child is a Girl and the second child is a Boy etc..
$\mathcal{A} = \{$At least one child is a Boy}=$\{BG, GB, BB\}$. So

$$\mathbf{P}(\mathcal{A}) = 3/4.$$

and

$$\mathbf{P}(\{BB\} \mid \mathcal{A}) = \frac{\mathbf{P}(\{BB\} \cap \mathcal{A})}{\mathbf{P}(\mathcal{A})} = \frac{\mathbf{P}(\{BB\})}{\mathbf{P}(\mathcal{A})} = \frac{1}{3}.$$

Monty Hall Paradox

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say Number a, and the host, who knows what's behind the other doors, opens another door b which has a goat. He then says to you, 'Do you want to pick door Number $c \neq a, b$?' Is it to your advantage to take the switch?
The door hiding the car has been chosen randomly, if door a hides the car then the host chooses b randomly and there is an implicit assumption that you prefer a car to a goat.

Incorrect Analysis:

Probability space is $\{1,2,3\}$ where i denotes that the car is behind door i and $\mathbf{P}(i) = \frac{1}{3}$ for $i = 1,2,3$.

Answer: Assume for example that $a = 1$. Then for $b = 2,3$,

$$\mathbf{P}(1 \mid \text{not } b) = \frac{\mathbf{P}(1)}{\mathbf{P}(\text{not } b)} = \frac{\mathbf{P}(1)}{\mathbf{P}(1) + \mathbf{P}(5-b)} = \frac{1}{2}.$$

So there is no advantage to be gained from switching.

Correct Analysis:

Assume that $a = 1$.
Probability space is $\{12, 13, 23, 32\}$ where ij denotes that the car is behind door i and the host opens door j.

$\mathbf{P}(12) = \mathbf{P}(13) = \frac{1}{6}$ and $\mathbf{P}(23) = \mathbf{P}(32) = \frac{1}{3}$. So,

$$\mathbf{P}(1) = \mathbf{P}(12) + \mathbf{P}(13) = \frac{1}{3}.$$

So there is an advantage to be gained from switching.
Look at it this way: The probability the car is not behind door a is 2/3 and switching causes you to win whenever this happens!

Binomial

n coin tosses.
$p = \mathbf{P}(Heads)$ for each toss.
$\Omega = \{H, T\}^n$.

$$\mathbf{P}(\omega) = p^k(1-p)^{n-k}$$

where k is the number of H's in ω.

E.g. $\mathbf{P}(HHTTHTHHTHHTHT) = p^8(1-p)^6$.

Fix k. $A = \{\omega : \ H$ appears k times$\}$
$\mathbf{P}(A) = \binom{n}{k}p^k(1-p)^{n-k}$. If $\omega \in A$ then

$$\mathbf{P}_A(\omega) = \frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{1}{\binom{n}{k}}$$

i.e. *conditional* on there being k heads, each sequence with k heads is equally likely.

Balls in boxes

m distinguishable balls in n distinguishable boxes.
Let

$$E_i = \{\text{Box } i \text{ is empty}\}.$$

$$\mathbf{Pr}(E_1) = \mathbf{Pr}(E_2) = \left(1 - \frac{1}{n}\right)^m$$

and

$$\mathbf{Pr}(E_1 \cap E_2) = \left(1 - \frac{2}{n}\right)^m < \mathbf{Pr}(E_1)\mathbf{Pr}(E_2).$$

So

$$\mathbf{Pr}(E_1 \mid E_2) < \mathbf{Pr}(E_1).$$

We say that the two events are *negatively correlated*.

Law of Total Probability

Let $B_1, B_2, \ldots, B_n$ be pairwise disjoint events which partition Ω. For any other event A,

$$\mathbf{P}(A) = \sum_{i=1}^{n} \mathbf{P}(A \mid B_i)\mathbf{P}(B_i).$$

Proof

$$\begin{aligned} \sum_{i=1}^{n} \mathbf{P}(A \mid B_i)\mathbf{P}(B_i) &= \sum_{i=1}^{n} \mathbf{P}(B_i \cap A) \\ &= \mathbf{P}\left(\bigcup_{i=1}^{n} (B_i \cap A)\right) \qquad (4) \\ &= \mathbf{P}(A). \end{aligned}$$

There is equality in (4) because the events $B_i \cap A$ are pairwise disjoint.

We are given two urns, each containing a collection of colored balls. Urn 1 contains 2 Red and 3 Blue balls. Urn 2 contains 3 Red and 4 Blue balls. A ball b_1 is drawn at random from Urn 1 and placed in Urn 2 and then a ball b_2 is chosen at random from Urn 2 and examined. What is the probability that b_2 is Blue? Let

$$\begin{aligned} A &= \{b_2 \text{ is Blue}\} \\ B_1 &= \{b_1 \text{ is Blue}\} \\ B_2 &= \{b_1 \text{ is Red}\} \end{aligned}$$

Then

$$\mathbf{P}(B_1) = \frac{3}{5},\ \mathbf{P}(B_2) = \frac{2}{5},\ \mathbf{P}(A \mid B_1) = \frac{5}{8},\ \mathbf{P}(A \mid B_2) = \frac{1}{2}.$$

So,

$$\mathbf{P}(A) = \frac{5}{8} \times \frac{3}{5} + \frac{1}{2} \times \frac{2}{5} = \frac{23}{40}.$$

Secretary Problem

There are n applicants for a secretarial position and CEO Pat will interview them in random order. The rule is that Pat must decide on the spot whether to hire the current applicant or interview the next one. Pat is an excellent judge of quality, but she does not know the set of applicants a priori. She wants to give herself a good chance of hiring the best.

Here is her strategy: She chooses a number $m < n$, interviews the first m and then hires the first person in $m+1, \ldots, n$ who is the best so far. (There is a chance that she will not hire anyone).

Let S be the event that Pat chooses the best person and let P_i be the event that the best person is the ith applicant. Then

$$\mathbf{Pr}(S) = \sum_{i=1}^{n} \mathbf{Pr}(S \mid P_i)\mathbf{Pr}(P_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Pr}(S \mid P_i).$$

Now Pat's strategy implies that $\mathbf{Pr}(S \mid P_i) = 0$ for $1 \le i \le m$. If P_i occurs for $i > m$ then Pat will succeed iff the best of the first $i - 1$ applicants (j say) is one of the first m, otherwise Pat will mistakenly hire j. Thus, for $i > m$, $\mathbf{Pr}(S \mid P_i) = \frac{m}{i-1}$ and hence

$$\mathbf{Pr}(S) = \frac{m}{n}\sum_{i=m+1}^{n} \frac{1}{i-1}.$$

Now assume that n is large and that $m = \alpha n$. Then

$$\mathbf{Pr}(S) \sim \alpha(\ln n - \ln \alpha n) = \alpha \ln 1/\alpha.$$

Pat will want to choose the value of α that maximises $f(\alpha) = \alpha \ln 1/\alpha$. But $f'(\alpha) = \ln 1/\alpha - 1$ and so the optimum choice for α is $1/e$. In which case,

$$\mathbf{Pr}(S) \sim e^{-1}.$$

2 sets $S, T \subseteq [n]$ are chosen (i) independently and (ii) uniformly at random from all possible sets. ($\Omega = \{0,1\}^{2n}$). Let

$$A = \{|S| = |T| \text{ and } S \cap T = \emptyset\}.$$

For each $X \subseteq [n]$ we let $B_X = \{S = X\} = \{(X, T) : \ T \subseteq [n]\}$. Thus for each X, $\mathbf{P}(B_X) = 2^{-n}$. So,

$$\begin{aligned} \mathbf{P}(A) &= \sum_X \mathbf{P}(A \mid B_X)\mathbf{P}(B_X) \\ &= 2^{-n} \sum_X \binom{n - |X|}{|X|} 2^{-n} \qquad (5) \\ &= 4^{-n} \sum_{k=0}^{n} \binom{n}{k} \binom{n-k}{k}. \end{aligned}$$

(5) follows from the fact that there are $\binom{n-|X|}{|X|}$ subsets of the same size as X which are disjoint from X.

Independence

Two events A, B are said to be *independent* if

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B),$$

or equivalently

$$\mathbf{P}(A \mid B) = \mathbf{P}(A).$$

(i) **Two Dice**
$A = \{\omega : \ x_1 \text{ is odd}\}$, $B = \{\omega : \ x_1 = x_2\}$.
$|A|$=18, $|B|$=6, $|A \cap B|$=3.
$\mathbf{P}(A) = 1/2$, $\mathbf{P}(B) = 1/6$, $\mathbf{P}(A \cap B) = 1/12$. A, B are independent.
(ii) $A = \{x_1 \geq 3\}$, $B = \{x_1 \geq x_2\}$.
$|A|$=24, $|B|$=21, $|A \cap B|$=18.
$\mathbf{P}(A) = 2/3$, $\mathbf{P}(B) = 7/12$, $\mathbf{P}(A \cap B) = 1/2$. A, B are not independent.

Random Bits

Suppose $\Omega = \{0,1\}^n = \{(x_1, x_2, \ldots, x_n) : \ x_j = 0/1\}$ with uniform distribution.

Suppose event A is determined by the values of x_i, $i \in \Delta_A$

e.g. if $A = \{x_1 = x_2 = \cdots = x_{10} = 0\}$ then $\Delta_A = \{1, 2, \ldots, 10\}$.

More Precisely: for $S \subseteq [n]$ and $x \in \Omega$ let $x_S \in \{0,1\}^S$ be defined by $(x_S)_i = x_i$, $i \in S$.

Ex. $n = 10$, $S = \{2, 5, 8\}$ and

$x = (0, 0, 1, 0, 0, 1, 1, 1, 1, 0\}$. $x_S = \{0, 0, 1\}$.

A is *determined* by Δ_A if $\exists S_A \subseteq \{0,1\}^{\Delta_A}$ such that $x \in A$ iff $x_{\Delta_A} \in S_A$. Furthermore, no subset of Δ_A has this property.

In our example above,

$S_A = \{(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)\}$ – ($|S_A| = 1$ here.)

Claim:

if events A, B are such that $\Delta_A \cap \Delta_B = \emptyset$ then A and B are independent.

$$\mathbf{P}(A) = \frac{|S_A|}{2^{|\Delta_A|}} \text{ and } \mathbf{P}(B) = \frac{|S_B|}{2^{|\Delta_B|}}.$$

$$\begin{aligned}
\mathbf{P}(A \cap B) &= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} 1_{\{x_{\Delta_A} \in S_A, x_{\Delta_B} \in S_B\}} \\
&= \frac{1}{2^n} |S_A| \; |S_B| 2^{n-|I_A|-|I_B|} \\
&= \mathbf{P}(A)\mathbf{P}(B).
\end{aligned}$$

Random Variables

A function $\zeta : \Omega \to \mathbf{R}$ is called a random variable.

Two Dice

$\zeta(x_1, x_2) = x_1 + x_2$.

$p_k = \mathbf{P}(\zeta = k) = \mathbf{P}(\{\omega : \ \zeta(\omega) = k\})$.

k	2	3	4	5	6	7	8	9	10	11	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Coloured Balls

$\Omega = \{m$ indistinguishable balls, n colours $\}$. Uniform distribution.
ζ = no. colours used.

$$p_k = \frac{\binom{n}{k}\binom{m-1}{k-1}}{\binom{n+m-1}{m}}.$$

If $m = 10, n = 5$ then
$p_1 = \frac{5}{1001}, p_2 = \frac{90}{1001}, p_3 = \frac{360}{1001}, p_4 = \frac{420}{1001},$
$p_5 = \frac{126}{1001}.$

Binomial Random Variable $B_{n,p}$.

n coin tosses. $p = \mathbf{P}(Heads)$ for each toss.
$\Omega = \{H, T\}^n$.

$$\mathbf{P}(\omega) = p^k(1-p)^{n-k}$$

where k is the number of H's in ω.
$B_{n,p}(\omega)$ = no. of occurrences of H in ω.

$$\mathbf{P}(B_{n,p} = k) = \binom{n}{k} p^k(1-p)^{n-k}.$$

If $n = 8$ and $p = 1/3$ then
$p_0 = \frac{2^8}{3^8}$, $p_1 = 8 \times \frac{2^7}{3^8}$, $p_2 = 28 \times \frac{2^6}{3^8}$,
$p_3 = 56 \times \frac{2^5}{3^8}$, $p_4 = 70 \times \frac{2^4}{3^8}$, $p_5 = 56 \times \frac{2^3}{3^8}$,
$p_6 = 28 \times \frac{2^2}{3^8}$, $p_7 = 8 \times \frac{2}{3^8}$, $p_8 = \frac{1}{3^8}$

Poisson Random Variable $Po(\lambda)$.

$\Omega = \{0, 1, 2, \ldots, \}$ and

$$\mathbf{P}(Po(\lambda) = k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \text{for all } k \geq 0.$$

This is a limiting case of $B_{n,\lambda/n}$ where $n \to \infty$.
$Po(\lambda)$ is the number of occurrences of an event which is individually rare, but has constant expectation in a large population.

Fix k, then

$$\begin{aligned}\lim_{n\to\infty} \mathbf{P}(B_{n,\lambda/n} = k) &= \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k e^{-\lambda}}{k!}\end{aligned}$$

Explanation of $\binom{n}{k} \approx n^k/k!$ for fixed k.

$$\begin{aligned}\frac{n^k}{k!} &\geq \binom{n}{k} \\ &= \frac{n^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \\ &\geq \frac{n^k}{k!} \left(1 - \frac{k(k-1)}{2n}\right)\end{aligned}$$

Expectation (Average)

$\mathbf{Z}$ is a random variable. Its *expected value* is given by

$$\begin{aligned}\mathbf{E}(\mathbf{Z}) &= \sum_{\omega\in\Omega} \zeta(\omega)\mathbf{P}(\omega)\\ &= \sum_{k} k\mathbf{P}(\zeta = k).\end{aligned}$$

Ex: **Two Dice**

$\zeta = x_1 + x_2$.

$$\mathbf{E}(\zeta) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + \cdots + 12 \times \frac{1}{36} = 7.$$

10 indistinguishable balls, 5 colours. Z is the number of colours actually used.

$$\mathbf{E}(Z) = \frac{5}{1001} + 2 \times \frac{90}{1001} + 3 \times \frac{360}{1001} + 4 \times \frac{420}{1001} + 5 \times \frac{126}{1001}.$$

In general: n colours, m balls.

$$\begin{aligned}
\mathbf{E}(\zeta) &= \sum_{k=1}^{n} \frac{k\binom{n}{k}\binom{m-1}{k-1}}{\binom{n+m-1}{m}} \\
&= n\sum_{k=1}^{n} \frac{\binom{n-1}{k-1}\binom{m-1}{k-1}}{\binom{n+m-1}{m}} \\
&= n\sum_{k-1=0}^{n-1} \frac{\binom{n-1}{k-1}\binom{m-1}{m-k}}{\binom{n+m-1}{m}} \\
&= \frac{n\binom{n+m-2}{m-1}}{\binom{n+m-1}{m}} \\
&= \frac{mn}{n+m-1}.
\end{aligned}$$

Geometric

$\Omega = \{1, 2, \ldots, \}$
$\mathbf{P}(k) = (1-p)^{k-1}p$, $\zeta(k) = k$.

$$\begin{aligned} \mathbf{E}(\zeta) &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\ &= \frac{p}{(1-(1-p))^2} \\ &= \frac{1}{p} \end{aligned}$$

= expected number of trials until success.

Binomial

$B_{n,p}$.

$$
\begin{aligned}
\mathbf{E}(B_{n,p}) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\
&= np(p+(1-p))^{n-1} \\
&= np.
\end{aligned}
$$

Poisson

$Po(\lambda)$.

$$\begin{aligned}
\mathbf{E}(Po(\lambda)) &= \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} \\
&= \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} \\
&= \lambda.
\end{aligned}$$

Suppose $\mathbf{X}, \mathbf{Y}$ are random variables on the same probability space Ω.

Claim: $\mathbf{E}(\mathbf{X}+\mathbf{Y}) = \mathbf{E}(\mathbf{X}) + \mathbf{E}(\mathbf{Y})$.

Proof:

$$\begin{aligned}
E(\mathbf{X}+\mathbf{Y}) &= \sum_{\alpha}\sum_{\beta}(\alpha+\beta)\mathbf{P}(\mathbf{X}=\alpha,\mathbf{Y}=\beta) \\
&= \sum_{\alpha}\sum_{\beta}\alpha\mathbf{P}(\mathbf{X}=\alpha,\mathbf{Y}=\beta) + \sum_{\alpha}\sum_{\beta}\beta\mathbf{P}(\mathbf{X}=\alpha,\mathbf{Y}=\beta) \\
&= \sum_{\alpha}\alpha\sum_{\beta}\mathbf{P}(\mathbf{X}=\alpha,\mathbf{Y}=\beta) + \sum_{\beta}\beta\sum_{\alpha}\mathbf{P}(\mathbf{X}=\alpha,\mathbf{Y}=\beta) \\
&= \sum_{\alpha}\alpha\mathbf{P}(\mathbf{X}=\alpha) + \sum_{\beta}\beta\mathbf{P}(\mathbf{Y}=\beta) \\
&= \mathbf{E}(\mathbf{X}) + \mathbf{E}(\mathbf{Y}).
\end{aligned}$$

In general if $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are random variables on Ω then

$$\mathbf{E}(\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n) = \mathbf{E}(\mathbf{X}_1) + \mathbf{E}(\mathbf{X}_2) + \cdots + \mathbf{E}(\mathbf{X}_n)$$

Binomial

Write $B_{n,p} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n$ where $\mathbf{X}_i = 1$ if the ith coin comes up heads.

$$\mathbf{E}(B_{n,p}) = \mathbf{E}(\mathbf{X}_1) + \mathbf{E}(\mathbf{X}_2) + \cdots + \mathbf{E}(\mathbf{X}_n) = np$$

since $\mathbf{E}(\mathbf{X}_i) = p \times 1 + (1-p) \times 0$.

Same probability space. $\zeta(\omega)$ denotes the number of occurrences of the sequence H, T, H in ω.
$\zeta = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_{n-2}$ where $\mathbf{X}_i = 1$ if coin tosses $i, i+1, i+2$ come up H, T, H respectively. So

$$\mathbf{E}(\zeta) = \mathbf{E}(\mathbf{X}_1) + \mathbf{E}(\mathbf{X}_2) + \cdots + \mathbf{E}(\mathbf{X}_{n-2}) = (n-2)p^2(1-p),$$

since $\mathbf{P}(x_i = 1) = p^2(1-p)$.

m **indistinguishable** balls, n colours. Z is the number of colours actually used.

$Z_i = 1 \leftrightarrow$ colour i is used.

$Z = Z_1 + \cdots + Z_n$= number of colours actually used.

$$\begin{aligned}
\mathbf{E}(Z) &= \mathbf{E}(Z_1) + \cdots + \mathbf{E}(Z_n) \\
&= n\mathbf{E}(Z_1) \\
&= n\mathbf{Pr}(Z_1 \neq 0) \\
&= n\left(1 - \frac{\binom{n+m-2}{m}}{\binom{n+m-1}{m}}\right). \\
&= n\left(1 - \frac{n-1}{n+m-1}\right) \\
&= \frac{mn}{n+m-1}.
\end{aligned}$$

m distinguishable balls, n boxes

$$\begin{aligned} \zeta &= \text{number of non-empty boxes.} \\ &= \zeta_1 + \zeta_2 + \cdots + \zeta_n \end{aligned}$$

where $\zeta_i = 1$ if box i is non-empty and = 0 otherwise. Hence,

$$\mathbf{E}(\zeta) = n\left(1 - \left(1 - \frac{1}{n}\right)^m\right),$$

since $\mathbf{E}(\zeta_i) = \mathbf{P}($ box i is non-empty$)$ = $\left(1 - (1 - \frac{1}{n})^m\right)$.

Why is this different from the previous frame?
The answer is that the indistinguishable balls space is obtained by partitioning the distinguishable balls space and then giving each set of the partition equal probability as opposed to a probability proportional to its size.

For example, if the balls are indistinguishable then the probability of exactly one non-empty box is $n \times \binom{m+n-1}{n-1}^{-1}$ whereas, if the balls are distinguishable, this probability becomes
$n \times n^{-m}$.

A problem with hats

There are n people standing a circle. They are blind-folded and someone places a hat on each person's head. The hat has been randomly colored Red or Blue.

They take off their blind-folds and everyone can see everyone else's hat. Each person then simultaneously declares (i) my hat is red or (ii) my hat is blue or (iii) or I pass.

They win a big prize if the people who opt for (i) or (ii) are all correct. They pay a big penalty if there is a person who incorrectly guesses the color of their hat.

Is there a strategy which means they will win with probability better than 1/2?

Suppose that we partition $Q_n = \{0,1\}^n$ into 2 sets W, L which have the property that L is a cover i.e. if $x = x_1x_2\cdots x_n \in W = Q_n \setminus L$ then there is $y_1y_2\cdots y_n \in L$ such that $h(x,y) = 1$ where

$$h(x,y) = |\{j : x_j \neq y_j\}|.$$

Hamming distance between x and y.

Assume that $0 \equiv Red$ and $1 \equiv Blue$. Person i knows x_j for $j \neq i$ (color of hat j) and if there is a unique value of x_i which places x in W then person i will declare that their hat has color i.

If indeed $x \in W$ then there is at least one person who will be in this situation and any such person will guess correctly.

Is there a small cover L?

Let $p = \frac{\ln n}{n}$. Choose L_1 randomly by placing $y \in Q_n$ into L_1 with probability p.

Then let L_2 be those $z \in Q_n$ which are not at Hamming distance ≤ 1 from some member of L_1.

Clearly $L = L_1 \cup L_2$ is a cover and
$\mathbf{E}(|L|) = 2^n p + 2^n (1-p)^{n+1} \leq 2^n (p + e^{-np}) \leq 2^n \frac{2 \ln n}{n}$.

So there must exist a cover of size at most $2^n \frac{2 \ln n}{n}$ and the players can win with probability at least $1 - \frac{2 \ln n}{n}$.

Conditional Expectation

Suppose $A \subseteq \Omega$ and Z is a a random variable on Ω. Then

$$\mathbf{E}(Z \mid A) = \sum_{\omega \in A} Z(\omega)\mathbf{P}(\omega \mid A) = \sum_{k} k\mathbf{P}(\zeta = k \mid A).$$

Ex: **Two Dice**

$\zeta = x_1 + x_2$ and $A = \{x_1 \geq x_2 + 4\}$.
$A = \{(5,1),(6,1),(6,2)\}$ and so $\mathbf{P}(A) = 1/12$.

$$\mathbf{E}(Z \mid A) = 6 \times \frac{1/36}{1/12} + 7 \times \frac{1/36}{1/12} + 8 \times \frac{1/36}{1/12} = 7.$$

Let $B_1, B_2, \ldots, B_n$ be pairwise disjoint events which partition Ω. Let Z be a random variable on Ω. Then

$$\mathbf{E}(Z) = \sum_{i=1}^{n} \mathbf{E}(Z \mid B_i)\mathbf{Pr}(B_i).$$

Proof

$$\begin{aligned}
\sum_{i=1}^{n} \mathbf{E}(Z \mid B_i)\mathbf{P}(B_i) &= \sum_{i=1}^{n} \sum_{\omega \in B_i} Z(\omega)\frac{\mathbf{P}(\omega)}{\mathbf{P}(B_i)}\mathbf{P}(B_i) \\
&= \sum_{i=1}^{n} \sum_{\omega \in B_i} Z(\omega)\mathbf{P}(\omega) \\
&= \sum_{\omega \in \Omega} Z(\omega)\mathbf{P}(\omega) \\
&= \mathbf{E}(Z).
\end{aligned}$$

First Moment Method

This is really Boole's inequality in disguise.
Let X be a random variable that takes values in $\{0, 1, 2, \ldots\}$.
Then

$$\mathbf{Pr}(X \geq 1) \leq \mathbf{E}(X)$$

Proof

$$\begin{aligned} \mathbf{E}(X) &= \mathbf{E}(X \mid X = 0)\mathbf{Pr}(X = 0) + \mathbf{E}(X \mid X \geq 1)\mathbf{Pr}(X \geq 1) \\ &\geq \mathbf{Pr}(X \geq 1). \end{aligned}$$

Union Distinct Families

Let $\mathcal{A}$ be a family of sub-sets of $[n]$. We say that $\mathcal{A}$ is *Union Distinct* if for distinct $A, B, C, D \in \mathcal{A}$ we have $A \cup B \neq C \cup D$. We use the probabilistic method to show the existence of a union distinct family of exponential size.

Suppose that $\mathcal{A}$ consists of p randomly and independently chosen sets $X_1, X_2, \ldots, X_p$. Let Z denote the number of 4-tples i, j, k, l such that $X_i \cup X_j = X_k \cup X_l$. Then

$$\mathbf{E}(Z) = p(p-1)(p-2)(p-3)\mathbf{Pr}(X_i \cup X_j = X_k \cup X_l) = \\ p(p-1)(p-2)(p-3)\left(\frac{5}{8}\right)^n.$$

(Observe that $\mathbf{Pr}(x \in (X_i \cup X_j) \setminus (X_k \cup X_l)) = 3/16.$)

So if $p \leq (8/5)^{n/4}$ then

$$\mathbf{Pr}(Z \geq 1) \leq \mathbf{E}(Z) < p^4 \left(\frac{5}{8}\right)^n \leq 1$$

implying that there exists a union free family of size p.

There is a small problem here in that we might have repetitions $X_i = X_j$ for $i \neq j$. Then our set will not be of size p.

But if Z_1 denotes the number of pairs i, j such that $X_i = X_j$ then

$$\mathbf{Pr}(Z_1 \neq 0) \leq \mathbf{E}(Z_1) = \binom{p}{2} 2^{-n}$$

and so we should really choose p so that
$\mathbf{Pr}(Z + Z_1 \neq 0) \leq \mathbf{E}(Z) + \mathbf{E}(Z_1) < p^4 \left(\frac{5}{8}\right)^n + p^2 \left(\frac{1}{2}\right)^n \leq 1.$

Average case of Quicksort

Quicksort is an algorithm for sorting numbers. Given distinct $x_1, x_2, \ldots, x_n$ we

1. Randomly choose an integer p between 1 and and n – the pivot.
2. Divide the remaining numbers into 2 sets L, R where $L = \{x_j : x_j < x_p\}$ and $R = \{x_j : x_j > x_p\}$.
3. Recursively sort L, R.

Let T_n be the expected number of comparisons taken by Quicksort.

We have $T_0 = 0$ and for $n \geq 1$

$$T_n =$$

$$\sum_{i=1}^{n} \mathbf{E}(No.\ comparisons \mid p\ is\ i'th\ largest)\mathbf{Pr}(p\ is\ i'th\ largest) =$$

$$\sum_{i=1}^{n}(n-1+T_{i-1}+T_{n-i}) \times \frac{1}{n}$$

$$= n - 1 + \frac{2}{n}\sum_{i=0}^{n-1} T_i$$

or

$$nT_n = n(n-1) + 2\sum_{i=0}^{n-1} T_i.$$

Let $T(x) = \sum_{n=0}^{\infty} T_n x^n$ be the generating function for T_n.

We note that

$$\sum_{n=1}^{\infty} nT_n x^n = xT'(x).$$

$$\sum_{n=1}^{\infty} n(n-1)x^n = \frac{2x^2}{(1-x)^3}.$$

$$\sum_{n=1}^{\infty} \left(\sum_{i=0}^{n-1} T_i \right) x^n = \frac{xT(x)}{1-x}.$$

Thus,

$$T'(x) = \frac{2x}{(1-x)^3} + \frac{2T(x)}{1-x}$$

or

$$(1-x)^2T'(x) - 2(1-x)T(x) = \frac{2x}{1-x}$$

or

$$\frac{d}{dx}((1-x)^2T(x)) = \frac{2x}{1-x}$$

and so

$$(1-x)^2T(x) = C - 2x - 2\ln(1-x).$$

$$(1-x)^2T(x) = C - 2x - 2\ln(1-x).$$

Now $T(0) = 0$ implies that $C = 0$ and so

$$\begin{aligned} T(x) &= -\frac{2x}{(1-x)^2} - \frac{2\ln(1-x)}{(1-x)^2} \\ &= -2\sum_{n=0}^{\infty} nx^n + 2\sum_{n=0}^{\infty}\left(\sum_{k=1}^{n}\frac{n-k+1}{k}\right)x^n \end{aligned}$$

So

$$\begin{aligned} T_n &= -4n + 2(n+1)\sum_{k=1}^{n}\frac{1}{k} \\ &\approx 2n\ln n. \end{aligned}$$

Hashing

Let $U = \{0, 1, \ldots, N - 1\}$ and $H = \{0, 1, \ldots, n - 1\}$ where n divides N and $N \gg n$. $f : U \to H$, $f(u) = u \mod n$.
(H is a hash table and U is the universe of objects from which a subset is to be stored in the table.)
Suppose $u_1, u_2, \ldots, u_m,\ m = \alpha n,$ are a random subset of U. A copy of u_i is stored in “cell” $f(u_i)$ and u_i’s that “hash” to the same cell are stored as a linked list.
Questions: u is chosen uniformly from U.
(i) What is the expected time T_1 to determine whether or not u is in the table?
(ii) If it is given that u is in the table, what is the expected time T_2 to find where it is placed?
Time = The number of comparisons between elements of U needed.

Let $M = N/n$, the average number of $u's$ that map to a cell. Let X_k denote the number of u_i for which $f(u_i) = k$. Then

$$\begin{aligned}
\mathbf{E}(T_1) &= \sum_{k=1}^{n} \mathbf{E}(T_1 \mid f(u) = k)\mathbf{P}(f(u) = k) \\
&= \frac{1}{n} \sum_{k=1}^{n} \mathbf{E}(T_1 \mid f(u) = k) \\
&\leq \frac{1}{n} \sum_{k=1}^{n} \mathbf{E}(X_k) \\
&= \frac{1}{n} \mathbf{E}\left(\sum_{k=1}^{n} X_k \right) \\
&= \alpha.
\end{aligned}$$

Let X denote $X_1, X_2, \ldots, X_n$ and let $\mathcal{X}$ denote the set of possible values for X. Then

$$\begin{aligned}
\mathbf{E}(T_2) &= \sum_{X \in \mathcal{X}} \mathbf{E}(T_2 \mid X)\mathbf{P}(X) \\
&= \sum_{X \in \mathcal{X}} \sum_{k=1}^{n} \mathbf{E}(T_2 \mid f(u) = k, X)\mathbf{P}(f(u) = k)\mathbf{P}(X) \\
&= \sum_{X \in \mathcal{X}} \sum_{k=1}^{n} \mathbf{E}(T_2 \mid f(u) = k, X)\frac{X_k}{m}\mathbf{P}(X) \\
&= \sum_{X \in \mathcal{X}} \sum_{k=1}^{n} \left(\frac{1 + X_k}{2}\right) \frac{X_k}{m}\mathbf{P}(X) \\
&= \frac{1}{2m} \sum_{X \in \mathcal{X}} \sum_{k=1}^{n} X_k(1 + X_k)\mathbf{P}(X)
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}(T_2) &= \frac{1}{2} + \frac{1}{2M}\mathbf{E}(X_1^2 + \cdots + X_n^2) \\
&= \frac{1}{2} + \frac{1}{2\alpha}\mathbf{E}(X_1^2) \\
&= \frac{1}{2} + \frac{1}{2\alpha}\sum_{t=1}^{m} t^2 \frac{\binom{M}{t}\binom{N-M}{m-t}}{\binom{N}{m}}.
\end{aligned}$$

If α is small and t is small then we can write

$$\begin{aligned} \frac{\binom{M}{t}\binom{N-M}{m-t}}{\binom{N}{m}} &\approx \frac{M^t}{t!}\frac{(N-M)^{m-t}}{(m-t)!}\frac{m!}{N^m} \\ &\approx \left(1-\frac{1}{n}\right)^m \frac{m^t}{t!n^t} \\ &\approx \frac{\alpha^t e^{-\alpha}}{t!}. \end{aligned}$$

Then we can further write

$$\mathbf{E}(T_2) \approx \frac{1}{2} + \frac{1}{2\alpha}\sum_{t=1}^{\infty} t^2 \frac{\alpha^t e^{-\alpha}}{t!} = 1 + \frac{\alpha}{2}$$

Random Walk

: Suppose we do n steps of previously described random walk. Let ζ_n denote the number of times the walk visits the origin. Then $\zeta_n = \mathbf{Y}_0 + \mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_n$ where $\mathbf{Y}_i = 1$ if $\mathbf{X}_i = 0$ – recall that $\mathbf{X}_i$ is the position of the particle after i moves.

But

$$\mathbf{E}(\mathbf{Y}_i) = \begin{cases} 0 & i \text{ odd} \\ \binom{i}{i/2} 2^{-i} & i \text{ even} \end{cases}$$

So

$$\begin{aligned} \mathbf{E}(\zeta_n) &= \sum_{\substack{0 \le m \le n \\ m \text{ even}}} \binom{m}{m/2} 2^{-m}. \\ &\approx \sum \sqrt{2/(\pi m)} \\ &\approx \frac{1}{2} \int_0^n \sqrt{2/(\pi x)} dx \\ &= \sqrt{2n/\pi} \end{aligned}$$

Finding Minimum

Consider the following program which computes the minimum of the n numbers $x_1, x_2, \ldots, x_n$.

```
begin
    $min := \infty$;
    for $i = 1$ to $n$ do
    begin
        if $x_i < min$ then $min := x_i$
    end
    output $min$
end
```

If the x_i are all different and in random order, what is the expected number of times that that the statement $min := x_i$ is executed?

Ω = {permutations of $1, 2, \ldots, n$} – uniform distribution. Let X be the number of executions of statement $min := x_i$. Let

$$X_i = \begin{cases} 1 & \text{statement executed at } i. \\ 0 & \text{otherwise} \end{cases}$$

Then $X_i = 1$ iff $x_i = \min\{x_1, x_2, \ldots, x_i\}$ and so

$$\mathbf{P}(X_i = 1) = \frac{(i-1)!}{i!} = \frac{1}{i}.$$

[The number of permutations of $\{x_1, x_2, \ldots, x_i\}$ in which x_i is the largest is $(i-1)!$.]

So

$$\begin{aligned}
\mathbf{E}(X) &= \mathbf{E}\left(\sum_{i=1}^{n} X_i\right) \\
&= \sum_{i=1}^{n} \mathbf{E}(X_i) \\
&= \sum_{i=1}^{n} \frac{1}{i} \qquad (= H_n) \\
&\approx \log_e n.
\end{aligned}$$

Independent Random Variables

Random variables $\mathbf{X}, \mathbf{Y}$ defined on the same probability space are called independent if for all α, β the events $\{\mathbf{X} = \alpha\}$ and $\{\mathbf{Y} = \beta\}$ are independent.
Example: if $\Omega = \{0, 1\}^n$ and the values of X, Y depend only on the values of the bits in disjoint sets Δ_X, Δ_Y then X, Y are independent.
E.g. if X = number of 1's in first m bits and Y = number of 1's in last $n - m$ bits.
The independence of X, Y follows directly from the disjointness of $\Delta_{\{X=\alpha\}}$ and $\Delta_{\{Y=\beta\}}$.

If X and Y are **independent** random variables then

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y).$$

$$\begin{aligned}
\mathbf{E}(XY) &= \sum_{\alpha}\sum_{\beta} \alpha\beta \mathbf{P}(X = \alpha, Y = \beta) \\
&= \sum_{\alpha}\sum_{\beta} \alpha\beta \mathbf{P}(X = \alpha)\mathbf{P}(Y = \beta) \\
&= \left[\sum_{\alpha} \alpha \mathbf{P}(X = \alpha)\right]\left[\sum_{\beta} \beta \mathbf{P}(Y = \beta)\right] \\
&= \mathbf{E}(X)\mathbf{E}(Y).
\end{aligned}$$

This is not true if X and Y are not independent. E.g. Two Dice:
$X = x_1 + x_2$ and $Y = x_1$.
$\mathbf{E}(X) = 7$, $\mathbf{E}(Y) = 7/2$ and
$\mathbf{E}(XY) = \mathbf{E}(x_1^2) + \mathbf{E}(x_1 x_2) = 91/6 + (7/2)^2$.

Inequalities

Markov Inequality: let $X : \Omega \to \{0, 1, 2, \ldots, \}$ be a random variable. For any $t \geq 1$

$$\mathbf{P}(X \geq t) \leq \frac{\mathbf{E}(X)}{t}.$$

Proof

$$\begin{aligned}
\mathbf{E}(X) &= \sum_{k=0}^{\infty} k\mathbf{P}(X = k) \\
&\geq \sum_{k=t}^{\infty} k\mathbf{P}(X = k) \\
&\geq \sum_{k=t}^{\infty} t\mathbf{P}(X = k) \\
&= t\mathbf{P}(X \geq t).
\end{aligned}$$

In particular, if $t = 1$ then

$$\mathbf{P}(X \neq 0) \leq \mathbf{E}(X)$$

m distinguishable balls, n boxes

Z = number of empty boxes.

$$m \geq (1+\epsilon)n \log_e n.$$

$$\begin{aligned}
\mathbf{E}(Z) &= n\left(1-\frac{1}{n}\right)^m \\
&\leq ne^{-m/n} \\
&\leq ne^{-(1+\epsilon)\log_e n} \\
&= n^{-\epsilon}.
\end{aligned}$$

So $\mathbf{P}(\exists \text{ an empty box}) \leq n^{-\epsilon}$.

Variance:

$Z : \Omega \to \boldsymbol{R}$ and $\mathbf{E}(Z) = \mu$.

$$\begin{aligned}
\mathbf{Var}(Z) &= \mathbf{E}((Z-\mu)^2) \\
&= \mathbf{E}(Z^2 - 2\mu Z + \mu^2) \\
&= \mathbf{E}(Z^2) - \mathbf{E}(2\mu Z) + \mathbf{E}(\mu^2) \\
&= \mathbf{E}(Z^2) - 2\mu\mathbf{E}(Z) + \mu^2 \\
&= \mathbf{E}(Z^2) - \mu^2.
\end{aligned}$$

Ex. Two Dice. $\zeta(x_1, x_2) = x_1 + x_2$.

$$\mathbf{Var}(\zeta) = \frac{2^2\times 1}{36} + \frac{3^2\times 2}{36} + \frac{4^2\times 3}{36} + \frac{5^2\times 4}{36} + \frac{6^2\times 5}{36} + \frac{7^2\times 6}{36} + \frac{8^2\times 5}{36} + \frac{9^2\times 4}{36} + \frac{10^2\times 3}{36} + \frac{11^2\times 2}{36} + \frac{12^2\times 1}{36} - 7^2 = \frac{35}{6}$$

Binomial: $Z = B_{n,p}$, $\mu = np$.

$$\begin{aligned}
\textbf{Var}(B_{n,p}) &= \sum_{k=1}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} - \mu^2 \\
&= \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} + \mu - \mu^2 \\
&= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} (1-p)^{n-k} \\
&= n(n-1)p^2 (p + (1-p))^{n-2} + \mu - \mu^2 \\
&= n(n-1)p^2 + \mu - \mu^2 \\
&= np(1-p).
\end{aligned}$$

Chebycheff Inequality

Now let $\sigma = \sqrt{\mathbf{Var}(Z)}$.

$$\begin{aligned} \mathbf{P}(|Z-\mu| \geq t\sigma) &= \mathbf{P}((Z-\mu)^2 \geq t^2\sigma^2) \\ &\leq \frac{\mathbf{E}((Z-\mu)^2)}{t^2\sigma^2} \qquad (6) \\ &= \frac{1}{t^2}. \end{aligned}$$

(6) comes from the Markov inequality applied to the random variable $(Z-\mu)^2$.

Back to Binomial: $\sigma = \sqrt{np(1-p)}$.

$$\mathbf{P}(|B_{n,p} - np| \geq t\sqrt{np(1-p)}) \leq \frac{1}{t^2}$$

which implies

$$\mathbf{P}(|B_{n,p} - np| \geq \epsilon np) \leq \frac{1}{\epsilon^2 np}$$

[Law of large numbers.]

Hoeffding's Inequality – Simple Case

Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values such that $\mathbf{Pr}(X_i = 1) = 1/2 = \mathbf{Pr}(X_i = -1)$ for $i = 1, 2, \ldots, n$. Let $X = X_1 + X_2 + \cdots + X_n$. Then for any $t \geq 0$

$$\mathbf{Pr}(|X| \geq t) < 2e^{-t^2/2n}.$$

Proof: For any $\lambda > 0$ we have

$$\begin{aligned} \mathbf{Pr}(X \geq t) &= \mathbf{Pr}(e^{\lambda X} \geq e^{\lambda t}) \\ &\leq e^{-\lambda t}\mathbf{E}(e^{\lambda X}). \end{aligned}$$

Now for $i = 1, 2, \ldots, n$ we have

$$\mathbf{E}(e^{\lambda X_i}) = \frac{e^{-\lambda} + e^{\lambda}}{2} = 1 + \frac{\lambda^2}{2!} + \frac{\lambda^4}{4!} + \cdots < e^{\lambda^2/2}.$$

So, by independence,

$$\mathbf{E}(e^{\lambda X}) = \mathbf{E}\left(\prod_{i=1}^{n} e^{\lambda X_i}\right) = \prod_{i=1}^{n} \mathbf{E}(e^{\lambda X_i}) \leq e^{\lambda^2 n/2}.$$

Hence,

$$\mathbf{Pr}(X \geq t) \leq e^{-\lambda t + \lambda^2 n/2}.$$

We choose $\lambda = t/n$ to minimise $-\lambda t + \lambda^2 n/2$. This yields

$$\mathbf{Pr}(X \geq t) \leq e^{-t^2/2n}.$$

Similarly,

$$\begin{aligned} \mathbf{Pr}(X \leq -t) &= \mathbf{Pr}(e^{-\lambda X} \geq e^{\lambda t}) \\ &\leq e^{-\lambda t}\mathbf{E}(e^{-\lambda X}) \\ &\leq e^{-\lambda t + \lambda^2 n/2}. \end{aligned}$$

Discrepancy

Suppose that $|X| = n$ and $\mathcal{F} \subseteq \mathcal{P}(X)$. If we color the elements of X with Red and Blue i.e. partition X in $R \cup B$ then the discrepancy $disc(\mathcal{F}, R, B)$ of this coloring is defined

$$disc(\mathcal{F}, R, B) = \max_{F \in \mathcal{F}} disc(F, R, B)$$

where $disc(F, R, B) = ||R \cap F| - |B \cap F||$ i.e. the absolute difference between the number of elements of F that are colored Red and the number that are colored Blue.

Claim:

If $|\mathcal{F}| = m$ then there exists a coloring R, B such that $disc(\mathcal{F}, R, B) \leq (2n \log_e(2m))^{1/2}$.

Proof Fix $F \in \mathcal{F}$ and let $s = |F|$. If we color X randomly and let $Z = |R \cap F| - |B \cap F|$ then Z is the sum of s independent ± 1 random variables.

So, by the Hoeffding inequality,

$$\mathbf{Pr}(|Z| \geq (2n \log_e(2m))^{1/2}) < 2e^{-n \log_e(2m)/s} \leq \frac{1}{m}.$$

Switching Game:

We are given an $n \times n$ matrix A where $A(i,j) = \pm 1$. We interpret $A(i,j) = 1$ as the light at i,j is on.

Now suppose that $x, y \in \{\pm 1\}^n$ are switches. The light at i,j is on if $A(i,j)x_i y_j = 1$ and off otherwise.

Let $\sigma(A) = \max_{x,y} \left| \sum_{i,j} A(i,j)x_i y_j \right|$ be the maximum absolute difference between the number of lights which are on and those that are off, obtaianble by switching.

Claim: There exists A such that $\sigma(A) \leq cn^{3/2}$ where $c = 2(\ln 2)^{1/2}$.

Fix $x, y \in \{\pm 1\}^n$ and let A be a random ± 1 matrix. Consider the random variable

$$Z_{x,y} = \sum_{i,j} A(i,j) x_i y_j.$$

This is the sum of n^2 independent random variables ($A(i,j)x_i y_j$) taking values in ± 1.

It follows from the Hoeffding inequality that

$$|Z_{x,y}| \geq cn^{3/2} < 2e^{-(cn^{3/2})^2/2n^2} = 2^{-2n}$$

So

$$\mathbf{Pr}(\max_{x,y} |Z_{x,y}| \geq cn^{3/2}) < 2^n \times 2^n \times = 2^{-2n} = 1.$$

Hence there exists A such that $\sigma(A) \leq cn^{3/2}$.