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

$$(1-1/n)^{cn} \leq (e^{-1/n})^{cn} = e^{-c}.$$

$$e^{-x-x^2} \le 1 - x \text{ if } 0 \le x \le 1/100. \qquad (1)$$

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

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

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

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 $X_n = X_n(\omega)$ denote the position of the particle after n moves.

Suppose $n = 2m$. What is the probability $X_n = 0$?

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

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

Boole's Inequality

$A, B \subseteq \Omega$.

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

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 $1 \leq i \leq n$. If $n < 2^{k-1}$ then there exists a partition $A = R \cup B$ such that

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

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

Proof Randomly colour A.
$\Omega = \{R, B\}^A = \{f : A \to \{R, B\}\}$, 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.

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

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

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

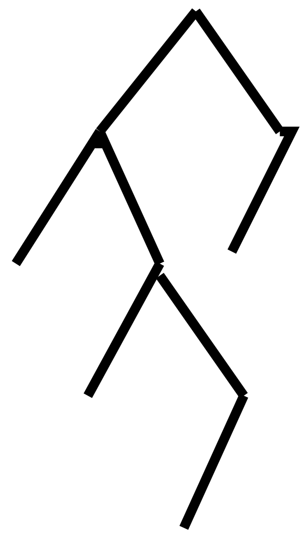
Explanation:
For any set $X \subseteq A$ and any $x \in \{R, B\}^X$ we have

$$\mathbf{P}(f(X) = x) = 2^{-|X|}.$$

1. The number of ω such that $f(X) = x$ is $2^{|A|-|X|}$.

2. $f(X) = x$ just depends on the random colours assigned to X and so is *independent* of colours not in X.

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 $t \geq 0$,

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

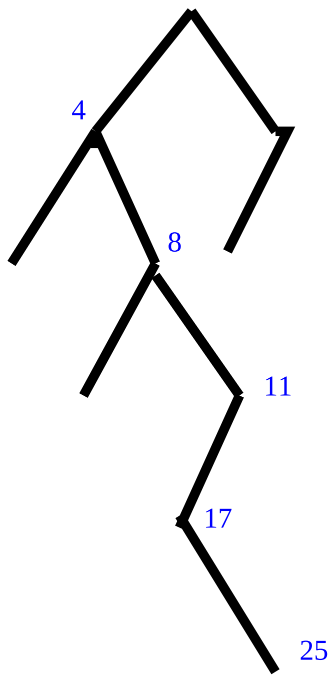
Proof The process requires at most n^2 coin flips and so we let $\Omega = \{L,R\}^{n^2}$ – most coin flips will not be needed most of the time.

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

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={4,8,11,17,25}

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

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

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 $A > 2$.