Let us throw $n$ balls into random bins.

Pr that all $n$ balls land up in different bins.

Let $E_i$, $i = 2 \ldots n$ denote the event that ball $i$ ends up in a bin not containing any of the first $i-1$ balls.

$$\Pr\left(\bigcap_{i=2}^{n} E_i^c\right) = \Pr\left(E_2\right) \Pr\left(E_2^c | E_3\right) \Pr\left(E_3^c | E_2^c \cap E_4\right) \cdots \Pr\left(E_n^c | E_2^c \cap E_3^c \cap \ldots \cap E_{n-1}\right)$$

What is $\Pr\left(E_i^c | E_2^c \cap E_3^c \cap \ldots \cap E_{i-1}\right) = 1 - \frac{i-1}{n}$.

Given that the first $i-1$ balls have landed in distinct bins, what is the probability that the $i^{th}$ ball lands in a distinct bin.

Thus

$$\Pr\left(\bigcap_{i=2}^{n} E_i^c\right) = \prod_{i=2}^{n} \left(1 - \frac{i-1}{m}\right) \leq \prod_{i=2}^{n} e^{-\frac{i-1}{m}}.$$  

Since $1 - x \leq e^{-x}$

$$= e^{-\frac{1}{m} \sum_{i=2}^{n} (i-1)} = e^{-\frac{n(n-1)}{2m}}$$
Thm: for \( n = \sqrt{2m + 1} \), \( pr \) that all \( n \) balls lands up in \( m \) distinct bins is at most \( \frac{1}{e} \) and if \( n \) increases, the probability decreases.

Birthday problem:
\( m = 365 = \# \text{days} \)
\( n = \# \text{ of people} \)

When \( n = \sqrt{2 \times 365 + 1} = 28 \)
then two people will have birthday on same date \( \text{w} \) \( pr > \frac{1}{2} \)

This is also related to collisions in hashing.
Markov's Inequality

Let \( X \) be a r.v. that only takes positive values

\[ X: \Omega \to \mathbb{R}^+ \]

Claim: \( \forall t > 0: \Pr[X \geq t \cdot E(X)] \leq \frac{1}{t} \)

Proof (using Definition)

\[ E(X) = \sum_{x \geq 0} x \Pr(X = x) \]

\[ = \sum_{0 \leq x < s} x \Pr(X = x) + \sum_{x \geq s} x \Pr(X = x) \]

\[ \geq \sum_{x \geq s} x \Pr(X = x) \]

\[ \geq \sum_{x \geq s} s \cdot \Pr(X = x) \]

\[ = s \sum_{x \geq s} \Pr(X = x) \]

\[ = s \cdot \Pr(X \geq s) \]

Therefore

\[ \Pr(X \geq s) \leq \frac{1}{s} E(X). \]

Set \( s = t \cdot E(X) \) to obtain the claim. \( \square \)
Example: Recall Quicksort

\[ X = \# \text{ Comparisons of } qsort. \]

We showed that \( E(X) \leq 2n \ln n \)

What about

\[ \Pr(X \geq 2 \cdot E(X)) = ? \]

\[ \leq \frac{1}{2} \quad (\text{Markov’s Inequality}) \]

\[ \Rightarrow \Pr(X \geq 4n \ln n) \leq \frac{1}{2} \]

Here is an alternative quicksort algorithm

- Run \( qsort \)
  
  Success - If it finishes in \( \leq 4n \ln n \) steps then \text{DONE}.  
  Failure - Else stop it after \( 4n \ln n \) steps and rerun it.

What is \( \Pr(\text{New } qsort \geq k \cdot 4n \ln n \text{ comparisons}) = ? \)

\[ \Rightarrow \Pr(\text{we run } qsort \geq k \text{ times}) \]

\[ \Rightarrow \Pr(\text{k times in a row } qsort \text{ does not complete within } 4n \ln n \text{ steps}) \]

\[ \leq \left( \frac{1}{2} \right)^k \sim \text{ probability of } k\text{-failures in a row} \]
Let $T$ be the running time of new $q$sort.

$E(T) = \ ?$

Let $N = \# \text{times we run } q\text{sort}$

$T \leq N \cdot 4n \ln n$

$E(T) \leq 4n \ln n \cdot E(N)$

What is $E(N) = \ ?$

Let $p$ be the probability that $q\text{sort}$ finishes in $4n \ln n$ steps

Note that $\frac{1}{2} \leq p \leq 1$

$E(N) = ? = \sum_{k=1}^{\infty} k \cdot \Pr(N=k) = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \cdot p = \frac{1}{p} N$

Recall geometric r.v

Thus $E(T) \leq \frac{1}{p} \cdot 4n \ln n$

$\leq 8n \ln n$

$= O(n \ln n)$