Today: Power of 2 choices
   - useful for hashing schemes

Balls into bins - simple scheme

Setting: n balls & n bins

Random assignment:
   For each ball, independently assign to a uniformly at random bin.

Let \( L(i) \) = load of bin \( i \) = # of balls assigned to bin \( i \)

Max load = \( \max_i L(i) \)

How large is the max load?

\( O(\log n) \) with high probability (whp)

Can show it's \( (1+o(1))\frac{\log n}{\log \log n} \) whp
\[ \Pr(\ell(i) > 2\log n) \leq \left( \frac{2\log n}{(2\log n)!} \right)^2 \left( \frac{1}{n} \right)^{2\log n} \]

\leq \left( \frac{e^{2\log n}}{(2\log n)!} \right)^2 \left( \frac{1}{n} \right)^{2\log n} \quad \text{(since } k! \geq \left( \frac{k}{e} \right)^k \text{)}

\leq \left( \frac{1}{2} \right)^{2\log n} \quad \text{for } n \text{ sufficiently large so that } \log n > e.

\leq \frac{1}{n^2}

Hence by a union bound,

with prob. \geq 1 - \frac{1}{n}, \ \text{max load} \leq 2\log n.

Can we improve this constant?

Yes, \[ \Pr(\ell(i) > \log n + \log \log n) \leq \left( \frac{1}{n} \right) \left( \frac{1}{\log n} \right) \]

Hence, max load \leq \log n + \log \log n \ \text{w. prob.} \geq 1 - \frac{1}{\log n}

this is \((1+o(1))\log n\)
Better approach:

Assign balls sequentially into bins

For ball $i = 1 \rightarrow n$:

- Choose 2 random bins $j$ & $k$.
- Check $L(j)$ & $L(k)$ for their current loads.
- If $L(j) < L(k)$ then:
  - assign ball $i$ to bin $j$.
- If $L(k) \leq L(j)$ then:
  - assign ball $i$ to bin $k$.

(In other words, assign ball $i$ to the least loaded of 2 randomly chosen bins.)
Theorem [Azar, Broder, Karlin, Upfal '94]

Max load is $O(\log \log n)$ with high probability. More generally, with $q \geq 2$ choices, it's $O\left(\frac{\log \log n}{\log q}\right)$.

Proof high-level idea:

Let $B_i$ = # of bins with load $\geq i$ at the end of the assignment.

Suppose we could prove $B_i \leq \beta_i$ w.h.p.

Then,

\[ P_e( \text{ball i is assigned to a bin with load} \geq i ) \leq \left( \frac{\beta_i}{n} \right)^2 \]

Since need $L(j), L(k) \geq i$ for it to occur.

Thus,

\[ B_{i+1} \leq Bin\left(n, \left(\frac{\beta_i}{n}\right)^2\right) \]

the mean of is $\frac{\beta_i^2}{n}$

& it should be close by Chernoff bound.
Thus, \( \beta_{i+1} = n \left( \frac{\beta_i}{n} \right)^2 = \frac{\beta_i^2}{n} \)

Note, \( \beta_2 = \frac{n}{2} \) holds since \( \leq \frac{n}{2} \) bins have \( \geq 2 \) balls.

Then, the recurrence solves to: \( \beta_{i+2} = \frac{1}{2^i} \)

hence for \( i^* = \log \log n \) we get \( \beta_{i+2} < 1 \)

So no \( \beta_{i+2} \) bins have load \( \geq i^* = \log \log n \).

Now let's formalize the proof.
Proof:

Base case: \( P_0 = \frac{n}{2e} \)

Note, \( \leq \frac{n}{2} \) bins have load \( \geq 6 \).

Since \( \frac{n}{6} < \frac{n}{2e} \) we know \( B_0 \leq \beta_0 \).

For \( i > 0 \), let

\[ \beta_{it} = \frac{e^{\beta_i}}{n} \]

Let the event \( \mathcal{A}_i = \sum B_i \leq \beta_i \gamma \)

& \( \mathcal{B}_i = \mathcal{A}_i = \sum B_i > \beta_i \gamma \)

Note,

\[
\Pr(\mathcal{B}_{it+1} | \mathcal{A}_i) = \Pr(\mathcal{B}_{it+1} | \mathcal{B}_i) \leq \frac{\Pr(\text{Bin}(n, \frac{\beta_i \gamma}{n}) > \beta_{i+1})}{\Pr(\mathcal{A}_i)}
\]
By a Chernoff bound, \( \Pr(X \geq eu) \leq e^{-\frac{e}{2}} \).

Thus, \( \Pr(B_{i+1} | A_i) \leq \frac{e^{\frac{e}{2n}}}{\Pr(A_i)} \leq \frac{\ln^2}{\Pr(A_i)} \)

assuming \( \frac{\beta_i^2}{n} \geq 2 \ln n \).

Now let's bound \( \Pr(A_i) \).

Claim: \( \Pr(B_i) \leq \frac{1}{n^2} \) assuming \( \frac{\beta_i^2}{n} \geq 2 \ln n \).

Using the claim, let \( i^* \) be the min \( i \) where \( \beta_i^2 < 2 \ln n \).

Since \( \beta_{i+1} = \frac{e \beta_i^2}{n} \) then \( i^* = \frac{\ln \ln n}{\ln 2} \).

Thus, for \( i^* = \ln \ln n \) we have \( \beta_{i^*} \leq \sqrt{2n \ln n} \) and we conclude that:

\( \leq \sqrt{2n \ln n} \) bins have load \( \geq \ln \ln n \) with high probability, \( \geq 1 - \frac{i^*}{n} \geq 1 - \frac{1}{n} \).
Proof of claim:

Base case: we know \( \Pr(B_0) = 0 \)

In general, recall \( \Pr(B_{i+1} | H_i) \leq \frac{X_i}{\Pr(H_i)} \)

Thus: \( \Pr(B_{i+1}) \leq \Pr(B_{i+1} | H_i) \Pr(H_i) + \Pr(B_{i+1} | B_i) \Pr(B_i) \)

\[ \leq \frac{X_i}{\Pr(H_i)} \Pr(H_i) + \frac{\Pr(B_{i+1} | B_i)}{\Pr(B_i)} \Pr(B_i) \]

\[ \leq \frac{1}{n^2} + \Pr(B_i) \]

\[ \leq \frac{1}{n^2} + \frac{i}{n^2} \quad \text{by induction} \]

\[ \leq \frac{(i+1)}{n^2} \]
Once again we know have that:

$$Pr(B_i) \leq \frac{i}{n^2} \leq \frac{1}{n} \text{ for all } i \text{ where }$$

$$\frac{\beta_i^2}{n} \geq 2 \ln n$$

Let $i^*$ be the min $i$ where $\beta_i^2 < 2 \ln n$.

Solving the recurrence $\frac{\beta_{i+1}^2}{n} = e \frac{\beta_i^2}{n}$

we have $i^* = \frac{\ln \ln n}{\ln 2}$ & since $\beta_i^2 \leq 2 \ln n$

so $\beta_{i^*} \leq \sqrt{2 \ln n}$.

Therefore, $\sqrt{2 \ln n}$ bins have load $\geq \ln \ln n$

with prob. $\geq 1 - \frac{1}{n}$. 
To finish off the proof:

Claim: \( \Pr(B_{i+2} \geq 1) = O(\frac{\log n}{n}) \)

Proof:

Let \( \mathcal{A}_{i+1} = \{ B_{i+1} \leq 6 \ln n \} \)

\[
\Pr(B_{i+1}) \leq \Pr(B_{i+1} \geq 6 \ln n | \mathcal{A}_{i+1}) \Pr(\mathcal{A}_{i+1}) + \Pr(B_{i+1})
\]

\[
\leq \Pr(B_{\text{Bin}}(n, \frac{2\ln n}{n}) \geq 6 \ln n) + \frac{1}{n^k}
\]

\[
\leq \frac{1}{n^2} + \frac{1}{n} \quad \text{by Chernoff bound}
\]

\[
= O(\frac{1}{n})
\]
Now for $i^*+2$:

$$
\Pr(B_{i^*+2} \geq 1) \leq \Pr(B_{i^*+2} \geq 1 \mid S_{i^*+1}) \Pr(S_{i^*+1}) + \Pr(B_{i^*+1})
$$

$$
\leq \Pr(\text{Bin}(n, (\frac{6\ln n}{n})^2) \geq 1) + O(\frac{1}{n})
$$

$$
\leq n \left( \frac{(\frac{6\ln n}{n})^2}{n} \right) + O(\frac{1}{n})
$$

$$
= O\left( \frac{1}{n} \left( \frac{1}{n} \right)^2 \right) = o(1).
$$