# Two Choices Paradigm

**Theorem 1.** *Let* $n$ *balls be sequentially placed into* $n$ *bins in the following manner. For each ball,* $d \geq 2$ *bins are chosen independently and uniformly at random (with replacement). Each ball is placed in the least full of the* $d$ *bins at the time of placement, with ties broken randomly. After all the balls are placed, the maximum load of any bin is at most* $\ln \ln n / \ln d + O(1)$ *with probability* $1 - o(1/n)$.

# Proof Idea

Bound the number of bins with at least $i$ balls for all $i$.

We will find a sequence of values $\beta_i$ such that the number of bins with load at least $i$ is bounded above by $\beta_i$, whp.

**Height** of a ball $= 1 +$ the number of balls in the bin where it is placed.

No. of bins with with at least $i$ balls $<=$ No. of balls of height at least $i$.

$\beta_{i+1} \leq cn(\beta_i/n)^d$.

We will show, whp, $\frac{\beta_{i+1}}{n} \leq 2(\frac{\beta_i}{n})^d$.

# Notation

Time $t$ refers to the state immediately after the $t$th ball is placed.

$h(t)$ — height of the $t$th ball.

$\nu_i(t)$ — no. of bins with load at least $i$ at time $t$.

$\mu_i(t)$ — number of balls with height at least $i$ at time $t$.

$\nu_i(n)$ — $\nu_i$

$\mu_i(n)$ — $\mu_i$

$\nu_i(t) \leq \mu_i(t)$.

# Lemmas

**Lemma 1.** *Let $B(n,p)$ be a binomial r.v. with parameters $n$ and $p$. Then*

$$\Pr(B(n,p) \geq 2np) \leq e^{-np/3}$$

**Lemma 2.** *Let $X_1, X_2, \ldots, X_n$ be a sequence of r.v. in an arbitrary domain and let $Y_1, Y_2, \ldots, Y_n$ be a sequence of 0-1 r.v. with the property that $Y_i = Y_i(X_1, \ldots, X_i)$.*

*If $\Pr(Y_i = 1 | X_1, \ldots, X_{i-1}) \leq p$, then*

$$\Pr(\sum_{i=1}^{n} Y_i > k) \leq \Pr(B(n,p) > k)$$

.

# Proof of Theorem

We will construct $\beta_i$ such that, whp, $\nu_i(n) \leq \beta_i$ for all $i$.

Let $\beta_4 = n/4$ and let $\beta_{i+1} = 2\beta_i^d/n^{d-1}$ for $4 \leq i < i^*$, for some $i^*$ (chosen later).

Let $E_i$ denote the event that $v_i(n) \leq \beta_i$.

We show that, whp, if $E_i$ holds then $E_{i+1}$ holds, for $4 \leq i < i^*$.

Fix a value $i$ in the above range. Let $Y_t$ be a 0-1 r.v such that

$Y_t = 1$ iff $h(t) \geq i+1$ and $v_i(t-1) \leq \beta_i$.

Let $w_j$ represent the bins selected by the $j$th ball.

Then $\Pr(Y_t = 1 | w_1, \ldots, w_{t-1}) \leq (\beta_i/n)^d$.

Hence,

$$\Pr(\sum_{t=1}^{n} Y_t > k) \leq \Pr(B(n, (\beta_i/n)^d) > k)$$

.

Conditioned on $E_i$ we have $\sum_{t=1}^{n} Y_t = \mu_{i+1}$.

Conditioned on $E_i$ we have $\sum_{t=1}^{n} Y_t = \mu_{i+1}$.

Since $\nu_{i+1} \leq \mu_{i+1}$,

$$\Pr(v_{i+1} > k | E_i) \leq \Pr(\mu_{i+1} > k | E_i)$$

$$= \Pr(\sum_{t=1}^{n} Y_t > k | E_i) \leq \Pr(\sum_{t=1}^{n} Y_t > k) / \Pr(E_i)$$
$$\leq \Pr(B(n, (\beta_i/n)^d) > k) / \Pr(E_i)$$

.

$$\Pr(\nu_{i+1} > \beta_{i+1} | E_i)$$

$$\leq \Pr(B(n, (\beta_i/n)^d) > 2n(\beta_i/n)^d)) / \Pr(E_i)$$

$$\leq e^{-n((\beta_i/n)^d)/3}(1/\Pr(E_i))$$

$\leq \frac{1}{n^2 \Pr(E_i)}$ whenever $n(\beta_i/n)^d \geq 6 \ln n$.

Thus, $\Pr(\neg E_{i+1} | E_i) \leq \frac{1}{n^2 \Pr(E_i)}$ whenever $n(\beta_i/n)^d \geq 6 \ln n$.

Next we remove the conditioning.

$\Pr(\neg E_{i+1})$

$$= \Pr(\neg E_{i+1}|E_i)\Pr(E_i) + \Pr(\neg E_{i+1}|\neg E_i)\Pr(\neg E_i)$$

$$\leq \Pr(\neg E_{i+1}|E_i)\Pr(E_i) + \Pr(\neg E_i) \leq \Pr(\neg E_i) + 1/n^2.$$

Hence, whenever $n(\beta_i/n)^d) \geq 6\ln n$ and $E_i$ holds, then so does $E_{i+1}$.

We can show that when $i = i^* = \frac{\ln\ln n}{\ln d} + O(1)$, then $n(\beta_i/n)^d < 6\ln n$.

Use induction to show that $\beta_{i+4} \leq n/2^{d^i}$.

Thus, $\Pr(\neg E_{i^*}) \leq i^*/n^2$.

We now handle the case where $n(\beta_i/n)^d < 6\ln n$. We have

$$\Pr(\nu_{i^*+1} > 18\ln n | E_{i^*}) \le \Pr(\mu_{i^*+1} > 18\ln n | E_{i^*})$$

$$\le \Pr(B(n, 6\ln n/n) \ge 18\ln n)/\Pr(E_{i^*}) \le \frac{1}{n^2 \Pr(E_{i^*})}.$$

$$\Pr(\nu_{i^*+1} > 18\ln n) \le \Pr(\neg E_{i^*}) + 1/n^2 \le (i^* + 1)/n^2.$$

$$\Pr(\nu_{i^*+3} \geq 1) \leq \Pr(\mu_{i^*+3} \geq 1) \leq \Pr(\mu_{i^*+2} \geq 2).$$

$$\Pr(\mu_{i^*+2} \geq 2 | \nu_{i^*+1} \leq 18 \ln n) \leq \frac{\Pr(B(n, (18 \ln n/n)^d \geq 2)}{\Pr(v_{i^*+1} \leq 18 \ln n)}$$

$$\leq \frac{\binom{n}{2}(18 \ln n/n)^{2d}}{\Pr(v_{i^*+1} \leq 18 \ln n)}.$$

Thus,

$$\Pr(\nu_{i^*+3} \geq 1) \leq \Pr(\mu_{i^*+2} \geq 2)$$

$$\leq \Pr(\mu_{i^*+2} \geq 2 | \nu_{i^*+1} \leq 18 \ln n) \Pr(v_{i^*+1} \leq 18 \ln n)$$

$$+ \Pr(v_{i^*+1} > 18 \ln n)$$

$$\leq \frac{(18 \ln n)^{2d}}{n^{2d-2}} + (i^* + 1)/n^2 = o(1)$$

if $d \geq 2$.

Thus the probability that the maximum loaded bin is more than $i^* + 3 = \ln \ln n / \ln d + O(1)$ is $o(1/n)$.