Randomness and Computation

or, “Randomized Algorithms”

Heng Guo

(Based on slides by M. Cryan)

Balls into Bins

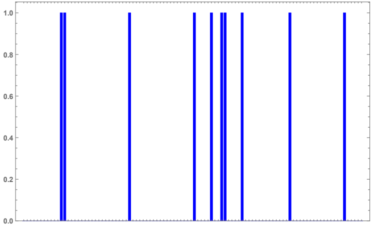
- m balls, n bins, and balls thrown uniformly at random and independently into bins (usually one at a time).
- Magic bins with no upper limit on capacity.
- Can be viewed as a random function $[m] \rightarrow [n]$.
- Common model of random allocations and their effects on overall *load* and *load balance* etc.

Many related questions:

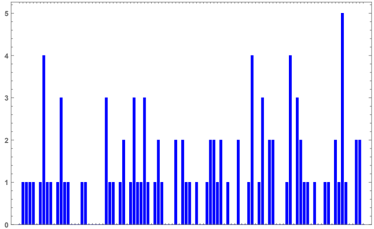
- How many balls do we need to cover all bins?
  (Coupon collector, *surjective mapping*)
- How many balls will lead to a collision?
  (Birthday paradox, *injective mapping*)
- What is the maximum load of each bin?
  (Load balancing)

Balls into Bins — maximum load

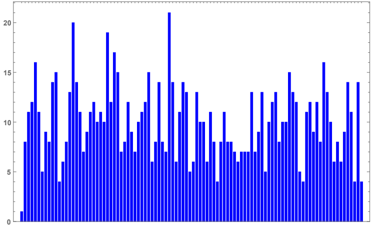
Depending on m/n, there are a few different scenarios. Fix $n = 100$, vary m.



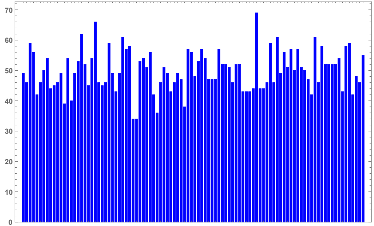
$m = 10$



$m = 100$



$m = 1000$



$m = 5000$

Balls into Bins — maximum load

$m = \Omega(n \log n)$ Maximum load is $\Theta\left(\frac{m}{n}\right)$, namely of the same order as the average load.

$m = n$ Maximum load is $\frac{\ln(n)}{\ln\ln(n)} + O(1)$ (same number of balls as bins).

We have already shown that when $m = n$ and n is sufficiently large, the maximum load is $\leq \frac{3\ln(n)}{\ln\ln(n)}$ with probability at least $1 - \frac{1}{n}$.

Today: a matching $\Omega(\frac{\ln(n)}{\ln\ln(n)})$ lower bound.

Poisson random variable

Probability p_r of a specific bin having r balls:

$$p_r = \binom{m}{r} \left(\frac{1}{n}\right)^r \left(1 - \frac{1}{n}\right)^{m-r}.$$

Note

$$p_r \sim \frac{e^{-m/n}}{r!} \left(\frac{m}{n}\right)^r,$$

where we consider r as a constant, m/n is a fixed constant, and $n \to \infty$.

Poisson random variable

Probability p_r of a specific bin having r balls:

$$p_r \;=\; \binom{m}{r}\left(\frac{1}{n}\right)^{r}\left(1-\frac{1}{n}\right)^{m-r}.$$

Note

$$p_r \sim \frac{e^{-m/n}}{r!}\left(\frac{m}{n}\right)^{r},$$

where we consider r as a constant, m/n is a fixed constant, and $n \to \infty$.

Definition (5.1)

A discrete Poisson random variable X with parameter μ is given by the following probability distribution on $j = 0, 1, 2, \ldots$:

$$\Pr[X = j] \;=\; \frac{e^{-\mu}\mu^{j}}{j!}.$$

Poisson as the limit of the Binomial Distribution

Theorem (5.5)

If X_n is a binomial random variable with parameters n and $p = p(n)$ such that $\lim_{n\to\infty} np = \mu$ is a constant (independent of n), then for any fixed $k \in \mathbb{N}_0$

$$\lim_{n\to\infty} \Pr[X_n = k] \;=\; \frac{e^{-\mu}\mu^k}{k!}.$$

Properties of Poisson

- It is well-defined:

$$\sum_{j=0}^{\infty} \frac{e^{-\mu}\mu^j}{j!} = e^{-\mu} \sum_{j=0}^{\infty} \frac{\mu^j}{j!} = 1.$$

- $\mathrm{E}[X] = \mu$
- $\mathrm{Var}[X] = \mu$
- The sum of two Poisson with parameters μ_1 and μ_2 is a Poisson with $\mu_1 + \mu_2$.

Concentration of Poisson r.v.

Theorem (5.4)

Let X be a Poisson random variable with parameter μ.

- *If $x = (1+\delta)\mu$ for $\delta > 0$, then*

$$\Pr[X \geq x] \leq \frac{e^{-\mu}(e\mu)^x}{x^x} = \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu};$$

- *If $x = (1-\delta)\mu$ for $0 < \delta < 1$, then*

$$\Pr[X \leq x] \leq \frac{e^{-\mu}(e\mu)^x}{x^x} = \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu};$$

Proof for Poisson concentration

Same argument as before:

$$\Pr[X \geq x] = \Pr[e^{tX} \geq e^{tx}] \leq \frac{\mathrm{E}[e^{tX}]}{e^{tx}}.$$

The claim follows from setting $t = \ln(x/\mu)$ and the following:

$$\mathrm{E}[e^{tX}] = e^{\mu(e^t - 1)}.$$

(It was $\leq$ in the sum of independent Bernoulli case.)

Poisson moment generating function

$$
\begin{aligned}
\mathrm{E}[e^{tX}] &= \sum_{i=0}^{\infty} e^{ti} \cdot \frac{e^{-\mu}\mu^i}{i!} \\
&= e^{-\mu} \sum_{i=0}^{\infty} \frac{(\mu e^t)^i}{i!} \\
&= e^{-\mu} e^{\mu e^t} \\
&= e^{\mu(e^t - 1)}
\end{aligned}
$$

Poisson modelling of balls-in-bins

Our balls in bins model has n bins, m (for variable m) balls, and the balls are thrown into bins independently and uniformly at random.

Each bin $X_i^{(m)}$ behaves like a binomial r.v. $B(m, \frac{1}{n})$.

Write $\mathbf{X}^{(m)} = (X_1^{(m)}, \ldots, X_n^{(m)})$ for the joint distribution. These $X_i^{(m)}$s are not independent.

Poisson modelling of balls-in-bins

Our balls in bins model has n bins, m (for variable m) balls, and the balls are thrown into bins independently and uniformly at random.

Each bin $X_i^{(m)}$ behaves like a binomial r.v. $B(m, \frac{1}{n})$.

Write $\mathbf{X}^{(m)} = (X_1^{(m)}, \ldots, X_n^{(m)})$ for the joint distribution. These $X_i^{(m)}$s are not independent.

For the "Poisson approximation" we take $\mu = \frac{m}{n}$, and write $Y_i^{(m)}$ to denote a Poisson r.v. with parameter $\mu = m/n$.

Write $\mathbf{Y}^{(m)} = (Y_1^{(m)}, \ldots, Y_n^{(m)})$ to denote a joint distribution of Poisson r.v.s which are all independent.

Notice that the sum $\sum_{i=1}^{n} Y_n^{(m)}$ is a Poisson r.v. with parameter $n \cdot \frac{m}{n} = m$.

Poisson approximation

Theorem (5.7)

Let $f(x_1, \ldots, x_n)$ be a non-negative function. Then

$$E[f(X_1^{(m)}, \ldots, X_n^{(m)})] \ \leq e\sqrt{m} \cdot E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})].$$

Poisson approximation

Theorem (5.7)

Let $f(x_1, \ldots, x_n)$ be a non-negative function. Then

$$E[f(X_1^{(m)}, \ldots, X_n^{(m)})] \ \le e\sqrt{m} \cdot E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})].$$

To bound the probability of some event, take f as its indicator function.

Corollary (5.9)

Any event that takes place with probability p in the "Poisson case" takes place with probability at most $pe\sqrt{m}$ in the exact balls-in-bins case.

Thus the Poisson approximation is good enough if p is sufficiently small.

Justify the Poisson approximation

$$\begin{aligned}
\mathrm{E}[f(\mathbf{Y}^{(m)})] &= \sum_{k=0}^{\infty} \mathrm{E}\left[f(\mathbf{Y}^{(m)}) \mid \sum_{i=1}^{n} Y_i^{(m)} = k\right] \Pr\left[\sum_{i=1}^{n} Y_i^{(m)} = k\right] \\
&\geq \mathrm{E}\left[f(\mathbf{Y}^{(m)}) \mid \sum_{i=1}^{n} Y_i^{(m)} = m\right] \Pr\left[\sum_{i=1}^{n} Y_i^{(m)} = m\right]
\end{aligned}$$

The theorem follow from two facts:

1. $\mathbf{Y}^{(m)}$ conditional on $\sum_{i=1}^{n} Y_i^{(m)} = k$ is the same as $\mathbf{X}^{(k)}$;
2. $\Pr[\sum_{i=1}^{n} Y_i^{(m)} = m] \geq \frac{1}{e\sqrt{m}}$.

Fact 1

The sum of $\mathbf{Y}^{(m)}$, denoted by Y, is a Poisson r.v. with parameter $n \cdot \frac{m}{n} = m$. Conditional on the sum, the probability that $\mathbf{Y}^{(m)}$ taking values $k_1, \ldots, k_n$ (where $\sum_{i=1}^{n} k_i = k$) is

$$\frac{\prod_{i=1}^{n} e^{-m/n}(m/n)^{k_i}/k_i!}{e^{-m}m^k/k!} = \frac{k!}{n^k \prod_{i=1}^{n} k_i!} = \frac{\binom{k}{k_1,\ldots,k_n}}{n^k},$$

which is exactly the probability that $\mathbf{X}^{(k)}$ taking values $k_1, \ldots, k_n$.

Fact 2

Stirling approximation:

$$m! \sim \sqrt{2\pi m}\left(\frac{m}{e}\right)^m.$$

The sum of $\mathbf{Y}^{(m)}$, denoted by Y, is a Poisson r.v. with parameter $n \cdot \frac{m}{n} = m$.

$$\Pr[Y = m] = e^{-m}m^m/m! \sim \frac{1}{\sqrt{2\pi m}}.$$

To make things rigorous (Lemma 5.8),

$$m! \leq e\sqrt{m}\left(\frac{m}{e}\right)^m.$$

Poisson approximation

Theorem (5.7)

Let $f(x_1, \ldots, x_n)$ be a non-negative function. Then

$$E[f(X_1^{(m)}, \ldots, X_n^{(m)})] \le e\sqrt{m} \cdot E[f(Y_1^{(m)}, \ldots, Y_n^{(m)})].$$

Corollary (5.9)

Any event that takes place with probability p in the "Poisson case" takes place with probability at most $pe\sqrt{m}$ in the exact balls-in-bins case.

Lower bound for n "balls into bins"

Lemma (5.12)

Let n balls be thrown independently and uniformly at random into n bins. Then (for n sufficiently large) the maximum load is at least $\frac{\ln(n)}{\ln\ln(n)}$ with probability at least $1 - \frac{1}{n}$.

Proof.

For the Poisson variables, we have $\mu = \frac{n}{n} = 1$. Let $M := \frac{\ln(n)}{\ln\ln(n)}$. For bin i,

$$\begin{aligned}\Pr[Y_i^{(m)} \geq M] &\geq \Pr\left[Y_i^{(m)} = M\right] \\ &= \frac{1^M e^{-1}}{M!} = \frac{1}{eM!}.\end{aligned}$$

(Chernoff bounds give an upper bound here.)

Lower bound for n "balls into bins"

Lemma (5.12)

Let n balls be thrown independently and uniformly at random into n bins. Then (for n sufficiently large) the maximum load is at least $\frac{\ln(n)}{\ln\ln(n)}$ with probability at least $1 - \frac{1}{n}$.

Proof.

For the Poisson variables, we have $\mu = \frac{n}{n} = 1$. Let $M := \frac{\ln(n)}{\ln\ln(n)}$. For bin i,

$$\begin{aligned} \Pr[Y_i^{(m)} \geq M] &\geq \Pr\left[Y_i^{(m)} = M\right] \\ &= \frac{1^M e^{-1}}{M!} = \frac{1}{eM!}. \end{aligned}$$

(Chernoff bounds give an upper bound here.)
In our Poisson model, the bins are independent, so the probability *no bin has load* $\geq M$ (our bad event) is at most

$$\left(1 - \frac{1}{eM!}\right)^n \leq e^{-n/(eM!)}.$$

Proof of Lemma 5.1 cont'd.

To tolerate the $e\sqrt{n}$ loss of the Poisson approximation, we want to show that

$$
\begin{aligned}
e^{-n/(eM!)} \le \frac{1}{n^2} \quad &\Leftrightarrow \quad 2\ln n \le \frac{n}{eM!} \\
&\Leftrightarrow \quad \ln\ln n + \ln 2 \le \ln n - \ln M! - 1 \\
&\Leftrightarrow \quad \ln M! \le \ln n - \ln\ln n - C,
\end{aligned}
$$

where $C = 1 + \ln 2$ is a constant.

Proof of Lemma 5.1 cont'd.

To tolerate the $e\sqrt{n}$ loss of the Poisson approximation, we want to show that

$$\begin{aligned} e^{-n/(eM!)} \le \frac{1}{n^2} \quad &\Leftrightarrow \quad 2\ln n \le \frac{n}{eM!} \\ &\Leftrightarrow \quad \ln\ln n + \ln 2 \le \ln n - \ln M! - 1 \\ &\Leftrightarrow \quad \ln M! \le \ln n - \ln\ln n - C, \end{aligned}$$

where $C = 1 + \ln 2$ is a constant.

Recall Lemma 5.8, $M! \le e\sqrt{M}\left(\frac{M}{e}\right)^M$.

$$\begin{aligned} \ln M! &\le M\ln M - M + \ln M \\ &= \frac{\ln n}{\ln\ln n}(\ln\ln n - \ln\ln\ln n) - \frac{\ln n}{\ln\ln n} + \ln M \\ &= \ln n - \frac{\ln n}{\ln\ln n} - (M\ln\ln\ln n - \ln M) \\ &\le \ln n - \frac{\ln n}{\ln\ln n}. \end{aligned}$$

Proof of Lemma 5.1 cont'd.

We want to show

$$\ln M! \leq \ln n - \ln \ln n - C,$$

where $C = 1 + \ln 2$ is a constant.

We have

$$\begin{aligned} \ln M! &\leq \ln n - \frac{\ln n}{\ln \ln n} \\ &\leq \ln n - \ln \ln n - C, \end{aligned}$$

since $\ln \ln n = o(\frac{\ln n}{\ln \ln n})$.

Proof of Lemma 5.1 cont'd.

We want to show

$$\ln M! \leq \ln n - \ln \ln n - C,$$

where $C = 1 + \ln 2$ is a constant.

We have

$$\begin{aligned} \ln M! &\leq \ln n - \frac{\ln n}{\ln \ln n} \\ &\leq \ln n - \ln \ln n - C, \end{aligned}$$

since $\ln \ln n = o(\frac{\ln n}{\ln \ln n})$.

Proof of Lemma 5.1 cont'd.

We have shown that in the Poisson case,

$$\begin{aligned}\Pr\left[\forall i \in [n],\ Y_i^{(m)} < M\right] &= \prod_{i=1}^{n} \Pr\left[Y_i^{(m)} < M\right] \\ &\le \left(1 - \frac{1}{eM!}\right)^n \le \frac{1}{n^2},\end{aligned}$$

where $M = \frac{\ln n}{\ln \ln n}$.

Proof of Lemma 5.1 cont'd.

We have shown that in the Poisson case,

$$\begin{aligned}\Pr\left[\forall i \in [n],\ Y_i^{(m)} < M\right] &= \prod_{i=1}^{n} \Pr\left[Y_i^{(m)} < M\right] \\ &\le \left(1 - \frac{1}{eM!}\right)^n \le \frac{1}{n^2},\end{aligned}$$

where $M = \frac{\ln n}{\ln \ln n}$.
Due to Poisson approximation,

$$\begin{aligned}\Pr\left[\forall i \in [n],\ X_i^{(m)} < M\right] &\le e\sqrt{n} \Pr\left[\forall i \in [n],\ Y_i^{(m)} < M\right] \\ &\le \frac{e\sqrt{n}}{n^2} < \frac{1}{n}.\end{aligned}$$

$\square$

References

- Sections 5.1 and 5.2 of "Probability and Computing".
- Sections 5.3 and 5.4 have all precise details of our $\Omega(\frac{\ln(n)}{\ln\ln(n)})$ result.
- We will skip the rest of Chapter 5.