Randomness and Computation
or, “Randomized Algorithms”

Mary Cryan

School of Informatics
University of Edinburgh
Balls in Bins

- \( m \) balls, \( n \) bins, and balls thrown \textit{uniformly at random} into bins (usually one at a time).

- Magic bins with no upper limit on capacity.

- Common model of random allocations and their affect on overall load and load balance, typical distribution in the system.

- “Classic” question - what does the distribution look like for \( m = n \)? Max load? (\textit{with high probability} results are what we want).

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

We will show an \( \Omega\left(\frac{\ln(n)}{\ln \ln(n)}\right) \) bound today.
Some preliminary observations, definitions

The probability is of a specific bin (bin $i$, say) being empty:

$$(1 - \frac{1}{n})^m \sim e^{-m/n}.$$ 

Expected number of empty bins: $\sim ne^{-m/n}$

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

$$p_r \sim e^{-m/n} \frac{m^r}{r!} \frac{1}{n}.$$ 

Definition (5.1)

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

$$\Pr[X = j] = e^{-\mu} \frac{\mu^j}{j!}.$$
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 $\lim_{n \to \infty} np = \lambda$ is a constant (independent of $n$), then for any fixed $k \in \mathbb{N}_0$

$$
\lim_{n \to \infty} \Pr[X_n = k] = \frac{e^{-\lambda} \lambda^k}{k!}.
$$
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 $(X_1^{(m)}, \ldots, X_n^{(m)})$ for the joint distribution (note the various $X_i^{(m)}$s are not independent).

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

We write $(Y_1^{(m)}, \ldots, Y_n^{(m)})$ to denote a joint distribution of independent Poisson r.v.s which are all 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 \( (X_1^{(m)}, \ldots, X_n^{(m)}) \) for the joint distribution (note the various \( X_i^{(m)} \)'s are not independent).

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

We write \( (Y_1^{(m)}, \ldots, Y_n^{(m)}) \) to denote a joint distribution of independent Poisson r.vs which are all 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 $(X_1^{(m)}, \ldots, X_n^{(m)})$ for the joint distribution (note the various $X_i^{(m)}$s are not independent).

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

We write $(Y_1^{(m)}, \ldots, Y_n^{(m)})$ to denote a joint distribution of independent Poisson r.v.s which are all independent.
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 \( (X_1^{(m)}, \ldots, X_n^{(m)}) \) for the joint distribution (note the various \( X_i^{(m)} \)s are not independent).

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

We write \( (Y_1^{(m)}, \ldots, Y_n^{(m)}) \) to denote a joint distribution of independent Poisson r.v.s which are all 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 $(X_{1}^{(m)}, \ldots, X_{n}^{(m)})$ for the joint distribution (note the various $X_{i}^{(m)}$s are not independent).

For the “Poisson approximation” we take $\lambda = \frac{m}{n}$, and write $Y_{i}^{(m)}$ to denote a Poisson r.v with parameter $\lambda = m/n$.

We write $(Y_{1}^{(m)}, \ldots, Y_{n}^{(m)})$ to denote a joint distribution of independent Poisson r.vs which are all independent.
Some preliminaries

Theorem (5.7)
Let $f(x_1, \ldots, x_n)$ be a non-negative function. Then

$$E[f(X^{(m)}_1, \ldots, X^{(m)}_n)] \leq e\sqrt{m} \cdot E[f(Y^{(m)}_1, \ldots, Y^{(m)}_n)].$$

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 in bins”

Lemma

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 $\lambda = \frac{n}{n} = 1$. Let $M = \lceil \frac{\ln(n)}{\ln \ln(n)} \rceil$. For any bin $i$ (say),

$$\Pr_{\text{Poiss}}[\text{bin } i \text{ has load } \geq M] \geq \Pr_{\text{Poiss}}[\text{bin } i \text{ has load } = M] = \frac{1^M e^{-1}}{M!} = \frac{1}{eM!}$$

In our Poisson model, the bins are independent, so the probability no bin has load $\geq M$ (our bad event) is at most

$$\left(1 - \frac{1}{eM!}\right)^n \leq e^{-n/(eM!)}. $$
Lower bound for $n$ “balls in bins”

**Lemma**

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 \( \lambda = \frac{n}{n} = 1 \). Let \( M = \left\lceil \frac{\ln(n)}{\ln \ln(n)} \right\rceil \). For any bin \( i \) (say),

\[
\Pr_{\text{Poiss}}[\text{bin } i \text{ has load } \geq M] \\
\geq \Pr_{\text{Poiss}}[\text{bin } i \text{ has load } = M] \\
= \frac{1^M e^{-1}}{M!} = \frac{1}{eM!}
\]

In our Poisson model, the bins are independent, so the probability *no bin has load* \( \geq M \) (our bad event) is at most

\[
\left( 1 - \frac{1}{eM!} \right)^n \leq e^{-n/(eM!)}.\]
Lower bound for $n$ “balls in bins”

Lemma

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 $\lambda = \frac{n}{n} = 1$. Let $M = \lceil \frac{\ln(n)}{\ln \ln(n)} \rceil$. For any bin $i$ (say),

$$\Pr_{\text{Poiss}}[\text{bin } i \text{ has load } \geq M] \geq \Pr_{\text{Poiss}}[\text{bin } i \text{ has load } = M] = \frac{1^M e^{-1}}{M!} = \frac{1}{eM!}.$$

In our Poisson model, the bins are independent, so the probability no bin has load $\geq M$ (our bad event) is at most

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

RC (2018/19) – Lecture 10 – slide 7
**Lower bound for n “balls in bins”**

**Lemma**

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

**Proof.**

For the Poisson variables, we have \( \lambda = \frac{n}{n} = 1 \). Let \( M = \lceil \frac{\ln(n)}{\ln \ln(n)} \rceil \). For any bin \( i \) (say),

\[
\Pr_{\text{Poiss}}[\text{bin } i \text{ has load } \geq M] \\
\geq \Pr_{\text{Poiss}}[\text{bin } i \text{ has load } = M] \\
= \frac{1^M e^{-1}}{M!} = \frac{1}{eM!}
\]

In our Poisson model, the bins are independent, so the probability *no bin has load \( \geq M \) (our bad event)* is at most

\[
\left( 1 - \frac{1}{eM!} \right)^n \leq e^{-n/(eM!)}. 
\]
Lower bound for $n$ “balls in bins”

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

Proof.
For the Poisson variables, we have $\lambda = \frac{n}{n} = 1$. Let $M = \lceil \frac{\ln(n)}{\ln \ln(n)} \rceil$. For any bin $i$ (say),

$$
\Pr_{\text{Poiss}}[\text{bin } i \text{ has load } \geq M] \\
\geq \Pr_{\text{Poiss}}[\text{bin } i \text{ has load } = M] \\
= \frac{1^M e^{-1}}{M!} = \frac{1}{eM!}
$$

In our Poisson model, the bins are independent, so the probability no bin has load $\geq M$ (our bad event) is at most

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

RC (2018/19) – Lecture 10 – slide 7
Lower bound for $n$ “balls in bins”

Lemma

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

Proof.

For the Poisson variables, we have $\lambda = \frac{n}{n} = 1$. Let $M = \lceil \frac{\ln(n)}{\ln\ln(n)} \rceil$. For any bin $i$ (say),

$$\mathbb{P}_{\text{Poiss}}[\text{bin } i \text{ has load } \geq M] \geq \mathbb{P}_{\text{Poiss}}[\text{bin } i \text{ has load } = M] = \frac{1^M e^{-1}}{M!} = \frac{1}{eM!}.$$

In our Poisson model, the bins are independent, so the probability no bin has load $\geq M$ (our bad event) is at most

$$\left(1 - \frac{1}{eM!}\right)^n \leq e^{-n/(eM!)}. $$
Lower bound for $n$ “balls in bins”

Proof of Lemma 5.1 cont’d.

We now relate $\Pr_{\text{Poiss}}[\text{bin } i \text{ has load } \geq M]$ to the probability of the same event in the balls-in-bins model.

Corollary 5.9 tells us that when we consider the exact balls-in-bins distribution $(X_1^{(n)}, \ldots, X_n^{(n)})$, that the probability of the event “no bin has $\geq M$ balls” is at most

$$e^{\sqrt{n}} \cdot e^{-n/(eM!)}.$$

We want this less than $n^{-1}$, i.e., we want $e^{1-n/(eM!)} \leq n^{-3/2}$. Taking $\ln(\cdot)$ of both sides, this happens if

$$\left(1 - \frac{n}{eM!}\right) \leq \frac{3}{2} \ln(n) \iff 1 + \frac{3}{2} \ln(n) \leq \frac{n}{eM!}.$$

Now $M! \leq e^{\sqrt{M}} \left(\frac{M}{e}\right)^M \leq M \left(\frac{M}{e}\right)^M$ (Lemma 5.8), hence $\frac{n}{eM!} \geq \frac{ne^M}{e^{M+1}}$. 

RC (2018/19) – Lecture 10 – slide 8
Lower bound for \( n \) “balls in bins”

**Proof of Lemma 5.1 cont’d.**

We now relate \( \Pr_{Poiss}[\text{bin } i \text{ has load } \geq M] \) to the probability of the same event in the balls-in-bins model.

Corollary 5.9 tells us that when we consider the exact balls-in-bins distribution \((X_1^{(n)}, \ldots, X_n^{(n)})\), that the probability of the event “no bin has \( \geq M \) balls” is at most

\[
e^{\sqrt{n}} \cdot e^{-n/(eM!)}.\]

We want this less than \( n^{-1} \), i.e., we want \( e^{1-n/(eM!)} \leq n^{-3/2} \). Taking \( \ln(\cdot) \) of both sides, this happens if

\[
\left(1 - \frac{n}{eM!}\right) \leq -\frac{3}{2} \ln(n) \iff 1 + \frac{3}{2} \ln(n) \leq \frac{n}{eM!}.
\]

Now \( M! \leq e^{\sqrt{M}} (\frac{M}{e})^M \leq M (\frac{M}{e})^M \) (Lemma 5.8), hence \( \frac{n}{eM!} \geq \frac{ne^M}{e^{M+1}} \).

\[\square\]
Proof of Lemma 5.1 cont’d.
We now relate $\Pr_{\text{Poiss}}[\text{bin } i \text{ has load } \geq M]$ to the probability of the same event in the balls-in-bins model.

Corollary 5.9 tells us that when we consider the exact balls-in-bins distribution $(X_1^{(n)}, \ldots, X_n^{(n)})$, that the probability of the event “no bin has $\geq M$ balls” is at most

$$e^{\sqrt{n}} \cdot e^{-n/(eM!)}.\]$$

We want this less than $n^{-1}$, ie we want $e^{1-n/(eM!)} \leq n^{-3/2}$. Taking $\ln(\cdot)$ of both sides, this happens if

$$\left(1 - \frac{n}{eM!}\right) \leq -\frac{3}{2} \ln(n) \Leftrightarrow 1 + \frac{3}{2} \ln(n) \geq \frac{n}{eM!}.\]

Now $M! \leq e^{\sqrt{M}}\left(\frac{M}{e}\right)^M \leq M\left(\frac{M}{e}\right)^M$ (Lemma 5.8), hence $\frac{n}{eM!} \geq \frac{ne^M}{e^{M+1}}$. \qed
Lower bound for \( n \) “balls in bins”

Proof of Lemma 5.1 cont’d.

Therefore it will suffice to show that \( 1 + \frac{3}{2} \ln(n) \leq \frac{ne^M}{e^{M+1}} \), or (for sufficiently large \( n \)), that

\[
2 \ln(n) \leq \frac{ne^M}{e^{M+1}}.
\]

Taking the \( \ln \) of both sides, this happens (using \( M \sim \frac{\ln(n)}{\ln \ln(n)} \)) when

\[
\ln(2) + \ln \ln(n) \leq \left( \ln(n) + \frac{\ln(n)}{\ln \ln(n)} \right) - \left( 1 + \left( \frac{\ln(n)}{\ln \ln(n)} + 1 \right) \left( \ln \ln(n) - \ln \ln \ln(n) \right) \right),
\]

ie, exactly when

\[
1 + \ln(2) + \ln \ln(n) \leq \left( \ln(n) + \frac{\ln(n)}{\ln \ln(n)} \right) - \ln(n) - \ln \ln(n) + \frac{\ln(n) \ln \ln \ln(n)}{\ln \ln(n)} + \ln \ln \ln(n),
\]

ie, exactly when

\[
1 + \ln(2) + 2 \ln \ln(n) \leq \frac{\ln(n)}{\ln \ln(n)} + \frac{\ln(n) \ln \ln \ln(n)}{\ln \ln(n)} + \ln \ln \ln(n).
\]
Proof of Lemma 5.1 cont’d.

Therefore it will suffice to show that \(1 + \frac{3}{2} \ln(n) \leq \frac{ne^M}{e^{M+1}}\), or (for sufficiently large \(n\)), that

\[
2 \ln(n) \leq \frac{ne^M}{e^{M+1}}.
\]

Taking the \(\ln\) of both sides, this happens (using \(M \sim \frac{\ln(n)}{\ln \ln(n)}\)) when

\[
\ln(2) + \ln \ln(n) \leq \left(\ln(n) + \frac{\ln(n)}{\ln \ln(n)}\right) - \left(1 + \left(\frac{\ln(n)}{\ln \ln(n)} + 1\right)(\ln \ln(n) - \ln \ln \ln(n))\right),
\]

ie, exactly when

\[
1 + \ln(2) + \ln \ln(n) \leq \left(\ln(n) + \frac{\ln(n)}{\ln \ln(n)}\right) - \ln(n) - \ln \ln(n) + \frac{\ln(n) \ln \ln \ln(n)}{\ln \ln(n)} + \ln \ln \ln(n),
\]

ie, exactly when

\[
1 + \ln(2) + 2 \ln \ln(n) \leq \frac{\ln(n)}{\ln \ln(n)} + \frac{\ln(n) \ln \ln \ln(n)}{\ln \ln(n)} + \ln \ln \ln(n).
\]
Lower bound for $n$ “balls in bins”

Proof of Lemma 5.1 cont’d.

Therefore it will suffice to show that $1 + \frac{3}{2} \ln(n) \leq \frac{ne^M}{e^{M+1}}$, or (for sufficiently large $n$), that

$$2\ln(n) \leq \frac{ne^M}{e^{M+1}}.$$ 

Taking the $\ln$ of both sides, this happens (using $M \sim \frac{\ln(n)}{\ln\ln(n)}$) when

$$\ln(2) + \ln\ln(n) \leq \left(\ln(n) + \frac{\ln(n)}{\ln\ln(n)}\right) - \left(1 + \left(\frac{\ln(n)}{\ln\ln(n)} + 1\right)(\ln\ln(n) - \ln\ln\ln(n))\right),$$

ie, exactly when

$$1 + \ln(2) + \ln\ln(n) \leq \left(\ln(n) + \frac{\ln(n)}{\ln\ln(n)}\right) - \ln(n) - \ln\ln(n) + \frac{\ln(n)\ln\ln\ln(n)}{\ln\ln(n)} + \ln\ln\ln(n),$$

ie, exactly when

$$1 + \ln(2) + 2\ln\ln(n) \leq \frac{\ln(n)}{\ln\ln(n)} + \frac{\ln(n)\ln\ln\ln(n)}{\ln\ln(n)} + \ln\ln\ln(n).$$
Lower bound for $n$ “balls in bins”

Proof of Lemma 5.1 cont’d.

Therefore it will suffice to show that $1 + \frac{3}{2} \ln(n) \leq \frac{ne^M}{e^{M^2+1}}$, or (for sufficiently large $n$), that

$$2 \ln(n) \leq \frac{ne^M}{e^{M^2+1}}.$$ 

Taking the $\ln$ of both sides, this happens (using $M \sim \frac{\ln(n)}{\ln \ln(n)}$) when

$$\ln(2) + \ln \ln(n) \leq \left( \ln(n) + \frac{\ln(n)}{\ln \ln(n)} \right) - \left( 1 + \left( \frac{\ln(n)}{\ln \ln(n)} + 1 \right) \ln \ln(n) - \ln \ln \ln(n) \right),$$

ie, exactly when

$$1 + \ln(2) + \ln \ln(n) \leq \left( \ln(n) + \frac{\ln(n)}{\ln \ln(n)} \right) - \ln(n) - \ln \ln(n) + \frac{\ln(n) \ln \ln \ln(n)}{\ln \ln(n)} + \ln \ln \ln(n),$$

ie, exactly when

$$1 + \ln(2) + 2 \ln \ln(n) \leq \frac{\ln(n)}{\ln \ln(n)} + \frac{\ln(n) \ln \ln \ln(n)}{\ln \ln(n)} + \ln \ln \ln(n).$$
Proof of Lemma 5.1 cont’d.
To show that

\[ 1 + \ln(2) + 2 \ln \ln(n) \leq \frac{\ln(n)}{\ln \ln(n)} + \frac{\ln(n) \ln \ln \ln(n)}{\ln \ln(n)} + \ln \ln \ln(n) \]

we will multiply across by \( \ln \ln(n) \), to verify the equivalent inequality

\[ (1+\ln(2)) \ln \ln(n) + 2(\ln \ln(n))^2 \leq \ln(n) + \ln(n) \ln \ln \ln(n) + \ln \ln \ln(n) \ln \ln(n). \]

At this point we notice that we have two terms on the right (\( \ln(n) \) and \( \ln(n) \ln \ln \ln(n) \)) which are exponentially larger than the two terms on the lhs - both lhs terms only grow wrt \( \ln \ln(n) \).

We do not need to check the numbers - as \( n \) grows the rhs will certainly be greater than the lhs.

Hence our claim holds.
Proof of Lemma 5.1 cont’d.
To show that

$$1 + \ln(2) + 2 \ln \ln(n) \leq \frac{\ln(n)}{\ln \ln(n)} + \frac{\ln(n) \ln \ln \ln(n)}{\ln \ln(n)} + \ln \ln \ln(n)$$

we will multiply across by $\ln \ln(n)$, to verify the equivalent inequality

$$(1 + \ln(2)) \ln \ln(n) + 2(\ln \ln(n))^2 \leq \ln(n) + \ln(n) \ln \ln \ln(n) + \ln \ln \ln(n) \ln \ln(n).$$

At this point we notice that we have two terms on the right ($\ln(n)$ and $\ln(n) \ln \ln \ln(n)$) which are exponentially larger than the two terms on the lhs - both lhs terms only grow wrt $\ln \ln(n)$.

We do not need to check the numbers - as $n$ grows the rhs will certainly be greater than the lhs.

Hence our claim holds.
Proof of Lemma 5.1 cont’d.

To show that

\[ 1 + \ln(2) + 2 \ln \ln(n) \leq \frac{\ln(n)}{\ln \ln(n)} + \frac{\ln(n) \ln \ln \ln(n)}{\ln \ln(n)} + \ln \ln \ln(n) \]

we will multiply across by \( \ln \ln(n) \), to verify the equivalent inequality

\[ (1+\ln(2)) \ln \ln(n) + 2(\ln \ln(n))^2 \leq \ln(n) + \ln(n) \ln \ln(n) + \ln \ln \ln(n) \ln \ln(n). \]

At this point we notice that we have two terms on the right (\( \ln(n) \) and \( \ln(n) \ln \ln \ln(n) \)) which are exponentially larger than the two terms on the lhs - both lhs terms only grow wrt \( \ln \ln(n) \).

We do not need to check the numbers - as \( n \) grows the rhs will certainly be greater than the lhs.

Hence our claim holds.
References and Exercises

▶ Sections 5.1, 5.2 of “Probability and Computing”.

▶ Sections 5.3 and 5.4 have all precise details of our $\Omega \left( \frac{\ln(n)}{\ln \ln(n)} \right)$ result.

▶ Section 5.5 on Hashing is worth a read and has none of the Poisson stuff (I’m skipping it because of time limitations).

Exercises
I will release a tutorial sheet.