Randomized Algorithms

The Poisson Approximation

Joseph Chuang-Chieh Lin

Department of Computer Science & Engineering
National Taiwan Ocean University

Spring 2026

Outline

The problem with balls and bins

- Throw m balls independently and uniformly into n bins.
- Let $X_i^{(m)}$ be the load of bin i.
- For each fixed bin,

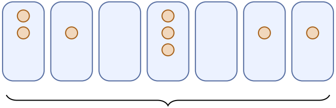
$$X_i^{(m)} \sim \mathrm{Binomial}(m, 1/n) \approx \mathrm{Poisson}(m/n).$$

The problem with balls and bins

- Throw m balls independently and uniformly into n bins.
- Let $X_i^{(m)}$ be the load of bin i.
- For each fixed bin,

$$X_i^{(m)} \sim \text{Binomial}(m, 1/n) \approx \text{Poisson}(m/n).$$

- But the vector $(X_1^{(m)}, \ldots, X_n^{(m)})$ is not independent.



$X_1^{(m)}, \ldots, X_n^{(m)}$ are dependent

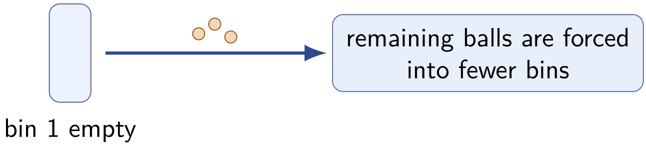
exact case: exactly m balls total

A small dependency example

Suppose m balls have been thrown into n bins.

Dependence

If bin 1 is empty, then all m balls must be distributed among the other $n-1$ bins. This changes the distribution of bin 2.

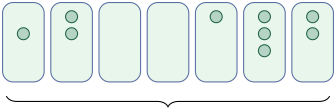


Why independence is valuable

- Independent loads allow product calculations:

$$\Pr[Y_1 < M, \dots, Y_n < M] = \prod_{i=1}^{n} \Pr[Y_i < M].$$

- Independent loads also permit Chernoff-style tools.
- The Poisson approximation replaces a dependent occupancy vector by an independent one.



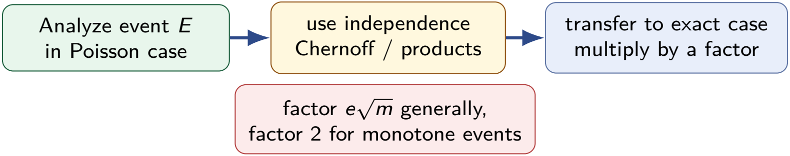
Poisson case: total is random, mean m

The central idea

Poissonization

Instead of fixing exactly m balls, let bin loads be independent Poisson random variables with mean m/n.

$$Y_1^{(m)}, \ldots, Y_n^{(m)} \quad \text{independent}, \qquad Y_i^{(m)} \sim \text{Poisson}(m/n).$$



Exact case versus Poisson case

Model	Load variables	Total load
Exact case	$X_1^{(m)}, \ldots, X_n^{(m)}$	exactly m
Poisson case	$Y_1^{(m)}, \ldots, Y_n^{(m)}$ independent	random, mean m

Key question

How much error do we introduce by replacing the **exact case** with the **Poisson case**?

Notation for the exact case

Say that we throw k balls independently and uniformly into n bins.

$$X_i^{(k)} = \#\{\text{balls in bin } i\}.$$

The joint distribution is multinomial:

$$\Pr\left[(X_1^{(k)}, \ldots, X_n^{(k)}) = (k_1, \ldots, k_n)\right] = \frac{\binom{k}{k_1, k_2, \ldots, k_n}}{n^k}.$$

whenever $\sum_{i=1}^{n} k_i = k$.

Notation for the Poisson case

Let

$$Y_1^{(m)}, Y_2^{(m)}, \ldots, Y_n^{(m)}$$

be independent Poisson random variables with common mean

$$\lambda = \frac{m}{n}.$$

Then

$$\Pr[Y_i^{(m)} = k_i] = e^{-m/n} \frac{(m/n)^{k_i}}{k_i!}.$$

Main benefit

The variables $Y_1^{(m)}, \ldots, Y_n^{(m)}$ are independent.

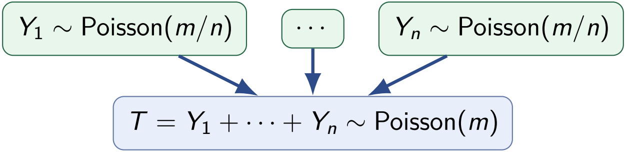
The total number of balls in the Poisson case

Because sums of independent Poisson variables are Poisson,

$$\sum_{i=1}^{n} Y_i^{(m)} \sim \text{Poisson}(m).$$

Thus the Poisson case has total number of balls

$$T = \sum_{i} Y_i^{(m)}, \qquad \mathbb{E}[T] = m.$$



Why conditioning should recover the exact case

The exact model fixes the total number of balls.

independent Poisson loads $Y_i \sim \text{Poisson}(m/n)$ → condition on $\sum_{i=1}^{n} Y_i = k$ → exact k balls $X_i^{(k)}$

conditioning restores the fixed-total model

Observation

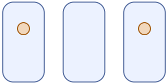
The only difference between the two models is whether the total number of balls is fixed or random.

A concrete mini-example

Take $n = 3$ bins and condition on total load $k = 2$.

$$(k_1, k_2, k_3) = (1, 0, 1).$$

- Exact model probability:
$$\frac{2!}{1!0!1!} \frac{1}{3^2} = \frac{2}{9}.$$
- Poisson model conditioned on total 2 gives the same value.



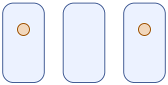
load vector $(1, 0, 1)$

A concrete mini-example

Take $n = 3$ bins and condition on total load $k = 2$.

$$(k_1, k_2, k_3) = (1, 0, 1).$$

- Exact model probability:
$$\frac{2!}{1!0!1!}\frac{1}{3^2} = \frac{2}{9}.$$
- Poisson model conditioned on total 2 gives the same value.



load vector $(1, 0, 1)$

$Y_1 + Y_2 + Y_3 \sim \text{Poisson}(2)$, so $\Pr[Y_1 + Y_2 + Y_3 = 2] = e^2(2^2/2!) = 2e^{-2}$.
$\Pr[(Y_1, Y_2, Y_3) = (1, 0, 1)] = e^{2/3}(2/3)^1/1!) \cdot e^{2/3}(2/3)^0/0!) \cdot e^{2/3}(2/3)^1/1!) = e^{-2}(2/3)^2$.

What is being approximated?

- We are not saying each bin is merely marginally Poisson.
- We want to approximate the joint load vector.
- The Poisson case gives independence, but loses the fixed-total constraint.

Conditioning Identity

Theorem 1

The distribution of

$$(Y_1^{(m)}, \ldots, Y_n^{(m)})$$

conditioned on

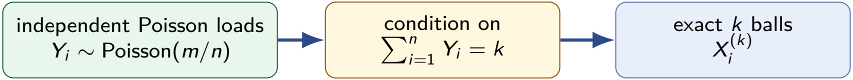
$$\sum_{i=1}^{n} Y_i^{(m)} = k$$

is the same as the distribution of

$$(X_1^{(k)}, \ldots, X_n^{(k)}),$$

regardless of the value of m.

What Theorem 1 says intuitively



conditioning restores the fixed-total model

Interpretation

If independent Poisson loads happen to sum to k, then their conditional allocation across the bins is exactly the same as throwing k balls uniformly into n bins.

Exact model probability

For nonnegative integers $k_1, \ldots, k_n$ satisfying

$$\sum_{i=1}^{n} k_i = k,$$

throwing k balls into n bins gives

$$\Pr\left[(X_1^{(k)}, \ldots, X_n^{(k)}) = (k_1, \ldots, k_n)\right] = \frac{k!}{k_1! \cdots k_n!} \frac{1}{n^k}.$$

- Multinomial coefficient chooses which balls go to each bin.
- Factor n^{-k} is the probability of a particular assignment of balls to bins.

Poisson numerator

In the Poisson case, independence gives

$$\Pr[Y_1 = k_1, \ldots, Y_n = k_n] = \prod_{i=1}^{n} e^{-m/n} \frac{(m/n)^{k_i}}{k_i!}.$$

Since $\sum_{i=1}^{n} k_i = k$,

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

Poisson denominator

The denominator for conditioning is

$$\Pr\left[\sum_{i=1}^{n} Y_i = k\right].$$

Since $\sum_{i=1}^{n} Y_i \sim \text{Poisson}(m)$,

$$\Pr\left[\sum_{i=1}^{n} Y_i = k\right] = e^{-m}\frac{m^k}{k!}.$$

Key cancellation

The terms involving m cancel out in the conditional probability.

The cancellation

$$\Pr\left[Y_1 = k_1, \ldots, Y_n = k_n \mid \sum_i Y_i = k\right]$$

$$= \frac{e^{-m}(m/n)^k/(k_1!\cdots k_n!)}{e^{-m}m^k/k!} = \frac{k!}{k_1!\cdots k_n!}\frac{1}{n^k}.$$

Why “regardless of m”?

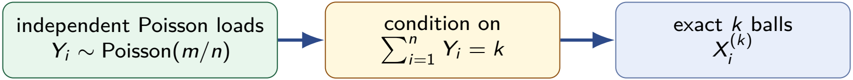
The conditional distribution does not depend on the original Poisson mean m/n.

$$\frac{e^{-m}(m/n)^k/(k_1!\cdots k_n!)}{e^{-m}m^k/k!} = \frac{k!}{k_1!\cdots k_n!}\frac{1}{n^k}.$$

Message

Once the total number of balls is fixed at k, the value of the Poisson mean no longer matters.

Theorem 1 as a bridge



conditioning restores the fixed-total model

- Exact model is hard because loads are dependent.
- Poisson model is easy because loads are independent.
- Theorem 1 says the exact model is a conditioned Poisson model.

Expectation transfer

Recall that

- $Y_1^{(m)}, \ldots, Y_n^{(m)}$: independent Poisson random variables with mean m/n;
- For m balls thrown into n bins independently and uniformly at random, $X_i^{(m)} :=$ the number of balls in the ith bin.

Theorem 2

Let $f(x_1, \ldots, x_n)$ be any **nonnegative** function. Then

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

Expectation transfer

Recall that

- $Y_1^{(m)}, \ldots, Y_n^{(m)}$: independent Poisson random variables with mean m/n;
- For m balls thrown into n bins independently and uniformly at random, $X_i^{(m)} :=$ the number of balls in the ith bin.

Theorem 2

Let $f(x_1, \ldots, x_n)$ be any **nonnegative** function. Then

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

A bound in the independent Poisson world transfers to the exact world, losing a factor $e\sqrt{m}$.

Why f must be nonnegative

Let $Y := (Y_1, Y_2, \ldots, Y_n)$. The proof starts from

$$\mathbb{E}[f(Y)] = \sum_{k=0}^{\infty} \mathbb{E}\left[f(Y) \mid \sum_{i=1}^{n} Y_i = k\right] \Pr\left[\sum_{i=1}^{n} Y_i = k\right].$$

If $f \geq 0$, then keeping only the $k = m$ term gives a valid lower bound:

$$\mathbb{E}[f(Y)] \geq \mathbb{E}\left[f(Y) \mid \sum_{i=1}^{n} Y_i = m\right] \Pr\left[\sum_{i=1}^{n} Y_i = m\right].$$

Nonnegativity

Dropping terms from a sum is safe only if the terms are nonnegative.

Apply Theorem 1 inside the proof

By Theorem 1,

$$\left\{ (Y_1^{(m)}, \ldots, Y_n^{(m)}) \;\middle|\; \sum_{i=1}^{n} Y_i = m \right\}$$

has the same distribution as

$$(X_1^{(m)}, \ldots, X_n^{(m)}).$$

Therefore

$$\mathbb{E}\left[f(Y) \;\middle|\; \sum_{i=1}^{n} Y_i = m \right] = \mathbb{E}[f(X)].$$

The remaining probability factor

Since $\sum_{i=1}^{n} Y_i \sim \text{Poisson}(m)$,

$$\Pr\left[\sum_{i=1}^{n} Y_i = m\right] = e^{-m}\frac{m^m}{m!}.$$

Thus

$$\mathbb{E}[f(Y)] \geq \mathbb{E}[f(X)]\, e^{-m}\frac{m^m}{m!}.$$

Rearranging gives

$$\mathbb{E}[f(X)] \leq \frac{m!e^m}{m^m}\,\mathbb{E}[f(Y)].$$

Where the factor $e\sqrt{m}$ comes from

Lemma 1 gives the loose factorial bound (by Stirling's approximation)

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

Plugging this into

$$\mathbb{E}[f(X)] \le \frac{m!e^m}{m^m}\,\mathbb{E}[f(Y)]$$

yields

$$\mathbb{E}[f(X)] \le e\sqrt{m}\,\mathbb{E}[f(Y)].$$

A Lemma from Stirling's approximation

Lemma 1

For $n \geq 1$,

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

- We have discussed that $n^n \geq n!$ and $n! \geq (n/e)^n$ in the classes.
- Hence the lemma provides a slightly better upper bound on $n!$.

Proof idea for Lemma 1

Use the fact that: $\ln(n!) = \sum_{i=1}^{n} \ln i$.

Proof idea for Lemma 1

Use the fact that: $\ln(n!) = \sum_{i=1}^{n} \ln i$. Since $\ln x$ is concave,

$$\int_{i-1}^{i} \ln x \, dx \geq \frac{\ln(i-1) + \ln i}{2} \quad (i \geq 2).$$

Proof idea for Lemma 1

Use the fact that: $\ln(n!) = \sum_{i=1}^{n} \ln i$. Since $\ln x$ is concave,

$$\int_{i-1}^{i} \ln x \, dx \geq \frac{\ln(i-1) + \ln i}{2} \quad (i \geq 2).$$

This yields

$$\int_{1}^{n} \ln x \, dx = n \ln n - n + 1 \geq \ln(n!) - \frac{\ln n}{2}.$$

Exponentiating gives the factorial bound.

How to use Theorem 2

1. Choose a **nonnegative** function f of the bin loads.
2. Analyze $\mathbb{E}[f(Y_1, \ldots, Y_n)]$ using independence.
3. Transfer the bound to the exact case:

$$\mathbb{E}[f(X)] \leq e\sqrt{m}\,\mathbb{E}[f(Y)].$$

Typical choices of f

An **indicator** of a rare event, the **number** of bad bins, or a function measuring overflow.

Why rare events are the target

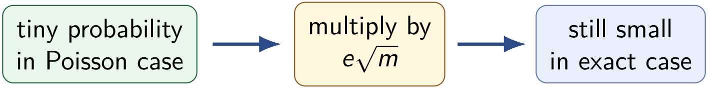
For an event E, Theorem 2 can imply

$$\Pr_{\mathrm{Exact}}[E] \leq e\sqrt{m}\,\Pr_{\mathrm{Poisson}}[E].$$

If

$$\Pr_{\mathrm{Poisson}}[E] \ll \frac{1}{e\sqrt{m}},$$

then the exact-case probability is still small.



The Event Version (Using indicators)

Corollary 1

Any event that takes place with probability p in the Poisson case takes place with probability at most

$$pe\sqrt{m}$$

in the exact case.

Why

Take f to be the indicator function of the event.

$$f(x_1, \ldots, x_n) = \mathbf{1}_E(x_1, \ldots, x_n).$$

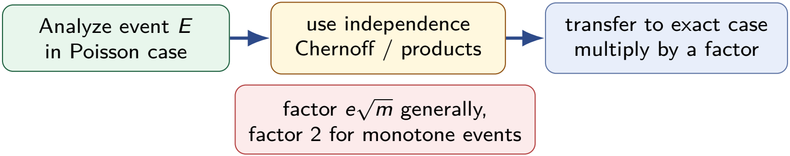
Applications using Indicator Functions

If $f = \mathbf{1}_E$, then

$$\mathbb{E}[f] = \Pr[E].$$

Therefore, Theorem 2 immediately gives

$$\Pr_{\mathrm{Exact}}[E] \le e\sqrt{m}\,\Pr_{\mathrm{Poisson}}[E].$$



When the $e\sqrt{m}$ factor is acceptable

Suppose the Poisson case gives

$$\Pr_{\text{Poisson}}[E] \le \frac{1}{e\sqrt{m}}\,\delta.$$

Then Corollary 1 gives

$$\Pr_{\text{Exact}}[E] \le e\sqrt{m} \cdot \frac{\delta}{e\sqrt{m}} = \delta.$$

Rule of thumb

The method is especially useful for rare bad events.

The monotonicity improvement

For many natural events, the probability is monotone in the number of balls.

- Increasing event: some bin has load at least M.
- Decreasing event: every bin has load at most $M - 1$.

Theorem 3 (Exercise)

If $f \geq 0$ and $\mathbb{E}[f(X^{(m)})]$ is monotone in m, then

$$\mathbb{E}[f(X^{(m)})] \leq 2\,\mathbb{E}[f(Y^{(m)})].$$

The Monotone Event Version

Corollary 2

Let E be an event whose probability is either monotonically increasing or monotonically decreasing in the number of balls.

If E has probability p in the Poisson case, then E has probability $\leq 2p$ in the exact case.

Improvement

For monotone events, the transfer factor improves from $e\sqrt{m}$ to 2.

Examples of monotone events

Event	Monotonicity in m
some bin has load $\geq M$	increasing
maximum load is $< M$	decreasing
all bins are nonempty	increasing
at least r empty bins	decreasing

Note

Monotonicity lets us compare a random total number of balls with a fixed total number of balls more sharply.

Choosing the right transfer theorem

Object	Assumption	Exact-case bound
nonnegative f	none	$e\sqrt{m}\,\mathbb{E}_{\text{Poisson}}[f]$
event E	none	$e\sqrt{m}\,\Pr_{\text{Poisson}}[E]$
monotone f	monotone expectation	$2\,\mathbb{E}_{\text{Poisson}}[f]$
monotone event E	monotone probability	$2\,\Pr_{\text{Poisson}}[E]$

Memory aid

General result: $e\sqrt{m}$; monotone result: 2.

Why the Poisson case is easier

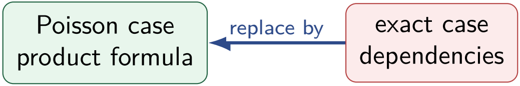
In the Poisson case, the loads are independent:

$$Y_i \sim \mathrm{Poisson}(m/n).$$

So, for example,

$$\Pr[\max_i Y_i < M] = \prod_{i=1}^{n} \Pr[Y_i < M].$$

In the exact case, the analogous probability does not factor.



Application: maximum load when $m = n$

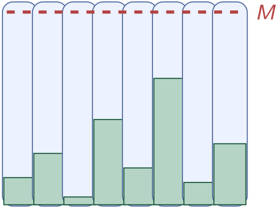
Now specialize to throwing n balls into n bins.

Goal

Show an almost-matching lower bound on the maximum load:

$$\max_i X_i^{(n)} \geq \frac{\ln n}{\ln \ln n}$$

with probability at least $1 - 1/n$ for sufficiently large n.



Set the threshold

$$\text{Let} \quad M = \frac{\ln n}{\ln \ln n}.$$

We want to show that some bin has load at least M with high probability.

upper bounds from union bound $O(\ln n / \ln \ln n)$ → Poisson approximation gives matching lower scale

In the Poisson case, loads are iid Poisson(1)

When $m = n$, the Poisson mean is

$$\lambda = \frac{m}{n} = 1.$$

Thus

$$Y_i \sim \text{Poisson}(1), \qquad \Pr[Y_i = M] = \frac{e^{-1}}{M!}.$$

Hence

$$\Pr[Y_i \geq M] \geq \frac{1}{eM!}.$$

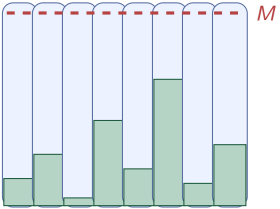
No heavy bin in the Poisson case

Because the Y_i are independent,

$$\Pr[\text{no bin has load at least } M] \le \left(1 - \frac{1}{eM!}\right)^n.$$

Using $1 - x \le e^{-x}$,

$$\left(1 - \frac{1}{eM!}\right)^n \le \exp\left(-\frac{n}{eM!}\right).$$



What remains to show

We want the Poisson failure probability to be at most n^{-2}:

$$\exp\left(-\frac{n}{eM!}\right) \leq n^{-2}.$$

What remains to show

We want the Poisson failure probability to be at most n^{-2}:

$$\exp\left(-\frac{n}{eM!}\right) \le n^{-2}.$$

It is enough to show

$$M! \le \frac{n}{2e\ln n}.$$

Why?

What remains to show

We want the Poisson failure probability to be at most n^{-2}:

$$\exp\left(-\frac{n}{eM!}\right) \le n^{-2}.$$

It is enough to show

$$M! \le \frac{n}{2e\ln n}.$$

Why?

$$\Rightarrow \quad \frac{n}{eM!} \ge 2\ln n,$$

so the exponential is at most $e^{-2\ln n} = n^{-2}$.

Bounding $M!$

Use Lemma 1 (a sharper upper bound on $M!$):

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

For sufficiently large M (also n), this implies the simpler estimate

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

With

$$M = \frac{\ln n}{\ln \ln n},$$

we have

$$\ln M = \ln \ln n - \ln \ln \ln n.$$

The key asymptotic calculation

$$\ln M! \leq M \ln M - M + \ln M \leq M \ln M \text{ (for sufficiently large } M\text{)}.$$

Substitute $M = \ln n / \ln\ln n$:

$$\begin{aligned} M \ln M &= \frac{\ln n}{\ln\ln n}(\ln\ln n - \ln\ln\ln n) \\ &= \ln n - \frac{\ln n \, \ln\ln\ln n}{\ln\ln n} \leq \ln n - \frac{\ln n}{\ln\ln n}. \end{aligned}$$

The key asymptotic calculation

$$\ln M! \leq M \ln M - M + \ln M \leq M \ln M \text{ (for sufficiently large } M\text{)}.$$

Substitute $M = \ln n / \ln \ln n$:

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

Then we have

$$\ln M! \leq \ln n - \frac{\ln n}{\ln \ln n} \leq \ln n - \ln \ln n - \ln(2e)$$

for sufficiently large n

The key asymptotic calculation

$$\ln M! \le M \ln M - M + \ln M \le M \ln M \text{ (for sufficiently large } M\text{)}.$$

Substitute $M = \ln n / \ln \ln n$:

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

Then we have

$$\ln M! \le \ln n - \frac{\ln n}{\ln \ln n} \le \ln n - \ln \ln n - \ln(2e)$$

for sufficiently large n (The last $\le$: using $\ln \ln n = o\left(\frac{\ln n}{\ln \ln n}\right)$).

Therefore the Poisson failure probability is tiny. From

$$M! \leq \frac{n}{2e \ln n},$$

we get

$$\Pr_{\text{Poisson}}\left[\max_{i \in \{1,\dots,n\}} Y_i < M\right] \leq \exp\left(-\frac{n}{eM!}\right) \leq n^{-2}.$$

Poisson conclusion

In the independent Poisson world, with probability at least $1 - n^{-2}$, some bin has load at least

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

Transfer back to the exact case

Let

$$E = \left\{ \max_{i \in \{1,\ldots,n\}} X_i^{(n)} < M \right\}.$$

This is the bad event: no bin reaches threshold M.

Using Corollary 1 with $m = n$,

$$\Pr_{\text{Exact}}[E] \le e\sqrt{n}\,\Pr_{\text{Poisson}}[E] \le \frac{e\sqrt{n}}{n^2} = \frac{e}{n^{3/2}} < \frac{1}{n}$$

for sufficiently large n.

A constant-detail remark

The proof arranges the Poisson failure probability to be at most n^{-2}; after transfer by $e\sqrt{n}$, this is still below $1/n$ for large n.

A constant-detail remark

The proof arranges the Poisson failure probability to be at most n^{-2}; after transfer by $e\sqrt{n}$, this is still below $1/n$ for large n.

- General transfer: multiply by $e\sqrt{n}$, so n^{-2} becomes $e/n^{3/2} < 1/n$.

A constant-detail remark

The proof arranges the Poisson failure probability to be at most n^{-2}; after transfer by $e\sqrt{n}$, this is still below $1/n$ for large n.

- General transfer: multiply by $e\sqrt{n}$, so n^{-2} becomes $e/n^{3/2} < 1/n$.
- Monotone transfer could also use a factor 2 for this monotone event.

A constant-detail remark

The proof arranges the Poisson failure probability to be at most n^{-2}; after transfer by $e\sqrt{n}$, this is still below $1/n$ for large n.

- General transfer: multiply by $e\sqrt{n}$, so n^{-2} becomes $e/n^{3/2} < 1/n$.
- Monotone transfer could also use a factor 2 for this monotone event.
- The asymptotic threshold is unaffected by this transfer choice.

Conclusion

For large n,

$$\Pr\left[\max_{i\in\{1,\ldots,n\}} X_i^{(n)} \geq \frac{\ln n}{\ln\ln n}\right] \geq 1 - \frac{1}{n}.$$

Alternative transfer: monotone events

The event

$$E = \left\{ \max_{i \in \{1,\dots,n\}} X_i^{(n)} < M \right\}$$

is monotone decreasing in the number of balls. Thus Corollary 2 gives

$$\Pr_{\text{Exact}}[E] \le 2 \Pr_{\text{Poisson}}[E].$$

If

$$\Pr_{\text{Poisson}}[E] \le n^{-2},$$

then Corollary 2 also gives

$$\Pr_{\text{Exact}}[E] \le 2n^{-2} < 1/n.$$

Discussions