Randomized Algorithms

Balls and Bins: Some basics

Joseph Chuang-Chieh Lin

Dept. CSIE,
National Taiwan Ocean University

Outline

- The Birthday Paradox – Revisited
- Balls into Bins
- Poisson Distribution

Exponential function - revisited

- By definition of e^x:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots$$

$$\Rightarrow \quad e = \sum_{k=0}^{\infty} \frac{1}{k!} = 1 + x + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} \cdots$$

Exponential function - revisited

- By the Binomial Theorem:

$$\begin{aligned}\left(1+\frac{1}{n}\right)^n &= \sum_{k=0}^{n}\binom{n}{k}\left(\frac{1}{n}\right)^k \\ &= \sum_{k=0}^{n}\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\cdot\frac{1}{n^k} \\ &= \sum_{k=0}^{n}\left(\frac{n}{n}\right)\cdot\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)\cdot\frac{1}{k!}.\end{aligned}$$

Exponential function - revisited

- By the Binomial Theorem:

$$\begin{aligned}\left(1+\frac{1}{n}\right)^n &= \sum_{k=0}^{n}\binom{n}{k}\left(\frac{1}{n}\right)^k \\ &= \sum_{k=0}^{n}\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\cdot\frac{1}{n^k} \\ &= \sum_{k=0}^{n}\left(\frac{n}{n}\right)\cdot\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)\cdot\frac{1}{k!}.\end{aligned}$$

$$\therefore \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = \sum_{k=0}^{\infty}\frac{1}{k!}.$$

Exponential function - revisited

- Useful approaches:

$$\left(1+\frac{1}{n}\right)^{-m} = \left(\left(1+\frac{1}{n}\right)^{n}\right)^{-m/n} \approx e^{-m/n}.$$

$$\frac{k^k}{k!} < \sum_{i=0}^{\infty} \frac{k^i}{i!} = e^k.$$

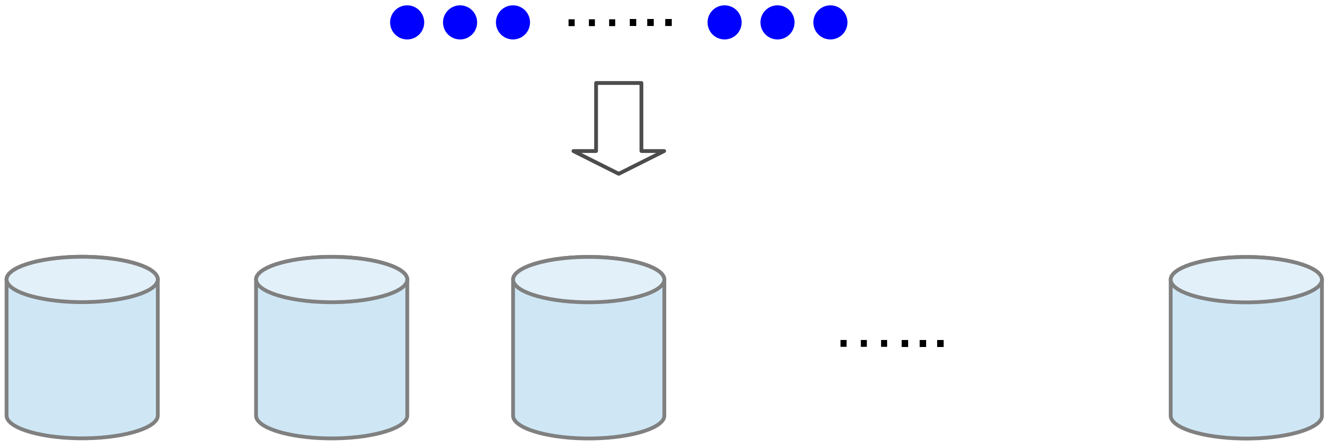
$$\therefore k! > \left(\frac{k}{e}\right)^k.$$

Exercise

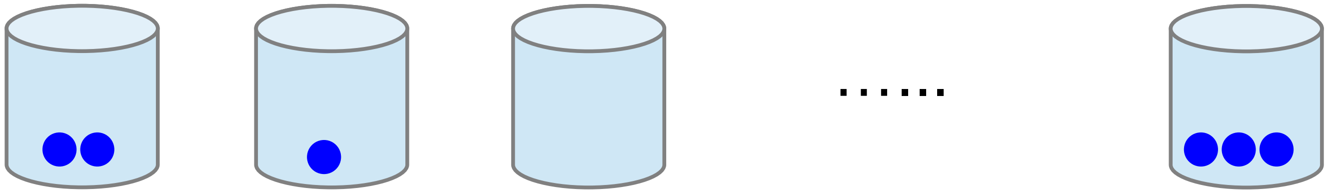
- Show that for $|x| \le 1$, $e^x(1-x^2) \le 1 + x \le e^x$.

- Let $\left(1-\left(1-\frac{1}{n}\right)^{km}\right)^k \approx (1-e^{\square})^k$.
  Find $\square = ?$

m balls into n bins

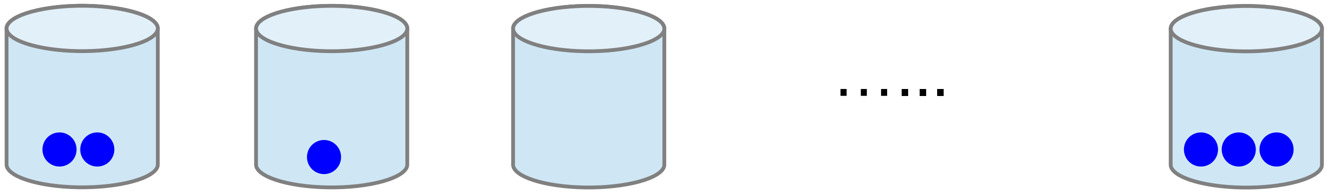


m balls into n bins



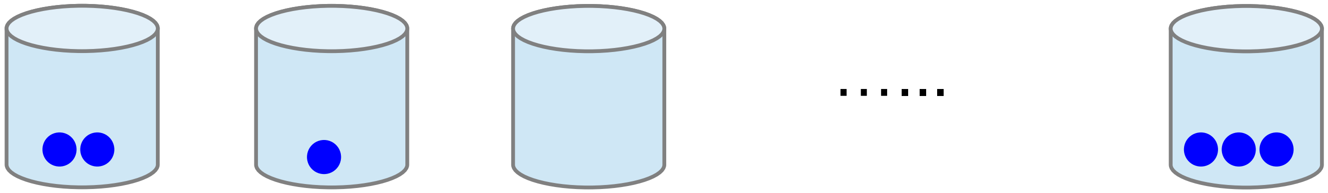
m balls into n bins

- Example. balls: people, bins: birthdays (mm/dd).



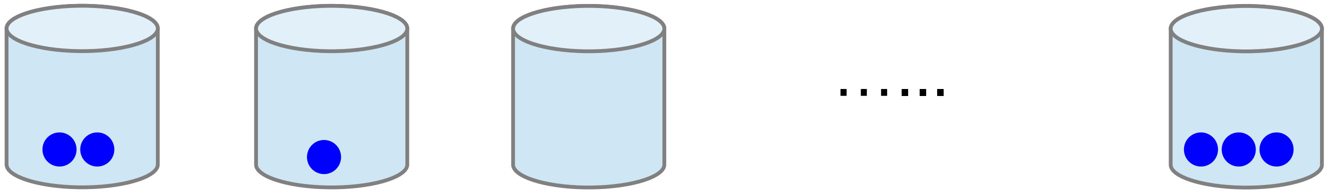
m balls into *n* bins

- Example. balls: people, bins: birthdays (mm/dd).
- Let's say there are 30 people in a room. What's the probability that no two people in the room share the same birthday?



m balls into n bins

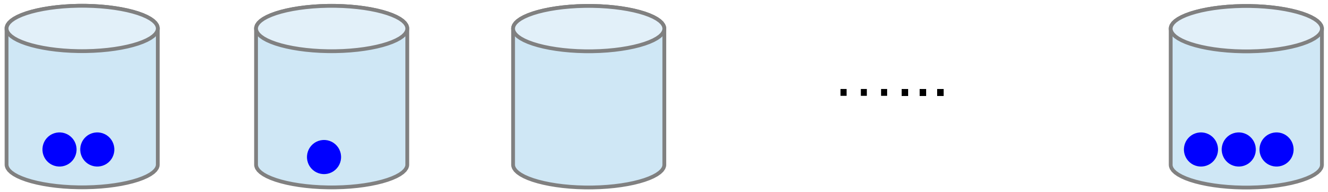
- Example. balls: people, bins: birthdays (mm/dd).
- Let's say there are 30 people in a room. What's the probability that no two people in the room share the same birthday?



$$\frac{\binom{365}{30}30!}{365^{30}}.$$

m balls into n bins

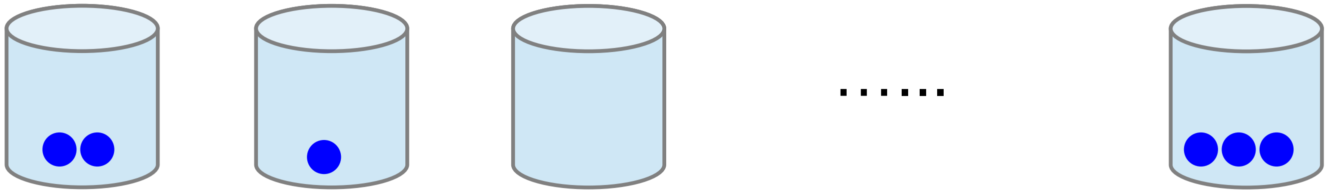
- Example. balls: people, bins: birthdays (mm/dd).
- Let's say there are 30 people in a room. What's the probability that no two people in the room share the same birthday?



$$\left(1-\frac{1}{365}\right)\cdot\left(1-\frac{2}{365}\right)\cdots\left(1-\frac{29}{365}\right).$$

m balls into *n* bins

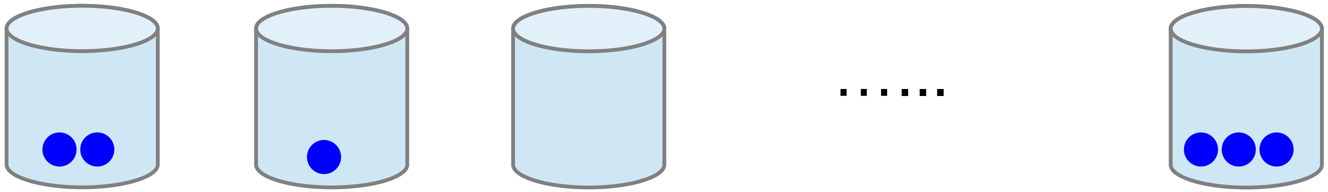
- <u>Example</u>. balls: people, bins: birthdays (mm/dd).
- In general, for ***m*** people in a room and ***n*** possible birthdays, what's the probability that no two people in the room share the same birthday?



$$\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{m-1}{n}\right)=\prod_{j=1}^{m-1}\left(1-\frac{j}{n}\right).$$

m balls into *n* bins

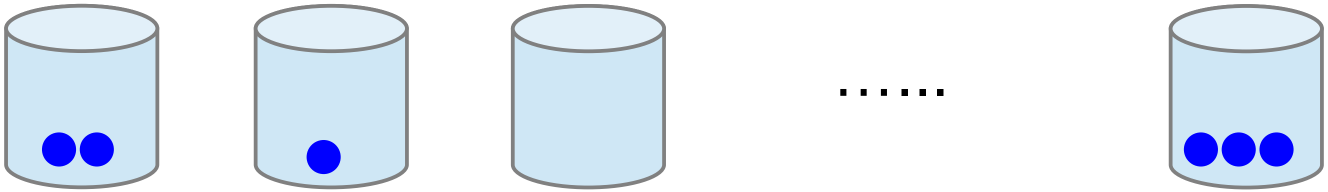
- Example. balls: people, bins: birthdays (mm/dd).
- In general, for ***m*** people in a room and ***n*** possible birthdays, what's the probability that no two people in the room share the same birthday?



$$\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{m-1}{n}\right)=\prod_{j=1}^{m-1}\left(1-\frac{j}{n}\right)\approx\prod_{j=1}^{m-1}e^{-j/n}=\exp\left\{-\sum_{j=1}^{m-1}\frac{j}{n}\right\}=e^{-m(m-1)/2n}\approx e^{-m^2/2n}.$$

m balls into *n* bins

- Example. balls: people, bins: birthdays (mm/dd).
- In general, for ***m*** people in a room and ***n*** possible birthdays, what's the probability that no two people in the room share the same birthday?

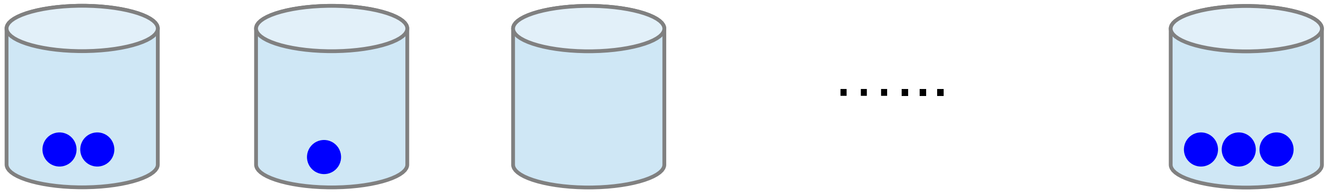


- Set the probability threshold to be ½:

$$\frac{m^2}{2n} = \ln 2 \Rightarrow m = \sqrt{2n \ln 2} \approx 22.49.$$

m balls into *n* bins

- <u>Example</u>. balls: people, bins: birthdays (mm/dd).
- In general, for ***m*** people in a room and ***n*** possible birthdays, what's the probability that no two people in the room share the same birthday?



- Set the probability threshold to be ½:

$$\frac{m^2}{2n} = \ln 2 \Rightarrow m = \sqrt{2n \ln 2} \approx \boxed{22.49.}$$

Matching your observation?

n balls into n bins: maximum load

- n balls are thrown independently and uniformly at random into n bins.
- What's the probability that the maximum "load" is more than L?
  - Maximum number of balls in one bin.

n balls into *n* bins: maximum load

- ***n*** balls are thrown independently and uniformly at random into *n* bins.
- What's the probability that the maximum "load" is more than L?
  - Maximum number of balls in one bin.
    - The average load is $n/n = 1$?!

n balls into *n* bins: maximum load

- ***n*** balls are thrown independently and uniformly at random into *n* bins.
- What's the probability that the maximum "load" is more than *L*?
  - Maximum number of balls in one bin.

- The probability that bin 1 receives at least *M* balls is at most

$$\binom{n}{M}\left(\frac{1}{n}\right)^M$$

n balls into *n* bins: maximum load

- ***n*** balls are thrown independently and uniformly at random into *n* bins.
- What's the probability that the maximum "load" is more than *L*?
  - Maximum number of balls in one bin.

- The probability that bin 1 receives at least *M* balls is at most

$$\binom{n}{M}\left(\frac{1}{n}\right)^{M}$$

union bound

n balls into *n* bins: maximum load

- ***n*** balls are thrown independently and uniformly at random into *n* bins.
- What's the probability that the maximum "load" is more than *L*?
  - Maximum number of balls in one bin.

- The probability that bin 1 receives at least *M* balls is at most

$$\binom{n}{M}\left(\frac{1}{n}\right)^M = \frac{n(n-1)\cdots(n-M+1)}{M!}\left(\frac{1}{n}\right)^M \leq \frac{1}{M!} \leq \left(\frac{e}{M}\right)^M .$$

n balls into *n* bins: maximum load

- The probability that bin 1 receives at least *M* balls is at most

$$\binom{n}{M}\left(\frac{1}{n}\right)^M = \frac{n(n-1)\cdots(n-M+1)}{M!}\left(\frac{1}{n}\right)^M \le \frac{1}{M!} \le \left(\frac{e}{M}\right)^M .$$

- Applying the union bound again:

n balls into n bins: maximum load

- The probability that bin 1 receives at least M balls is at most

$$\binom{n}{M}\left(\frac{1}{n}\right)^M = \frac{n(n-1)\cdots(n-M+1)}{M!}\left(\frac{1}{n}\right)^M \le \frac{1}{M!} \le \left(\frac{e}{M}\right)^M .$$

- Applying the union bound again:

$$n\left(\frac{e}{M}\right)^M \quad \le \quad n\left(\frac{e\ln\ln n}{3\ln n}\right)^{3\ln n/\ln\ln n}$$

n balls into n bins: maximum load

- The probability that bin 1 receives at least M balls is at most

$$\binom{n}{M}\left(\frac{1}{n}\right)^M = \frac{n(n-1)\cdots(n-M+1)}{M!}\left(\frac{1}{n}\right)^M \le \frac{1}{M!} \le \left(\frac{e}{M}\right)^M.$$

- Applying the union bound again:

$$\begin{aligned}
n\left(\frac{e}{M}\right)^M &\le n\left(\frac{e\ln\ln n}{3\ln n}\right)^{3\ln n/\ln\ln n} \\
&\le n\left(\frac{\ln\ln n}{\ln n}\right)^{3\ln n/\ln\ln n} = e^{\ln n}\left(\frac{e^{\ln\ln\ln n}}{e^{\ln\ln n}}\right)^{3\ln n/\ln\ln n} \\
&= e^{\ln n}(e^{\ln\ln\ln n-\ln\ln n})^{3\ln n/\ln\ln n} \\
&= e^{-2\ln n+3(\ln n)(\ln\ln n)/\ln\ln n} \\
&\le \frac{1}{n}.
\end{aligned}$$

Application: Bucket Sort

- A set of $n = 2^m$ numbers chosen uniformly at random in $[0, 2^k)$, $k \geq m$.

Bucket Sort:

- Stage 1: place the elements into n buckets.
  - j^{th} bucket: holds all elements whose first m binary digits corresponds to j.
    - e.g., $n = 2^{10}$, bucket 3 contains all elements whose first 10 digits are 0000000011.
- Stage 2: sort each bucket using any standard sorting algorithm.

Application: Bucket Sort

- A set of $n = 2^m$ numbers chosen uniformly at random in $[0, 2^k)$, $k \geq m$.

Bucket Sort: Expected $O(n)$ time?!

- Stage 1: place the elements into n buckets.
  - j^{th} bucket: holds all elements whose first m binary digits corresponds to j.
    - e.g., $n = 2^{10}$, bucket 3 contains all elements whose first 10 digits are 0000000011.
- Stage 2: sort each bucket using any standard sorting algorithm.

Application: Bucket Sort

- A set of $n = 2^m$ numbers chosen uniformly at random in $[0, 2^k)$, $k \geq m$.

Bucket Sort:

- Stage 1: place the elements into n buckets.
  - j^{th} bucket: holds all elements whose first m binary digits corresponds to j.
    - e.g., $n = 2^{10}$, bucket 3 contains all elements whose first 10 digits are 0000000011.
  - The number of elements landing in a specific bucket: Binomial(n, $1/n$).
- Stage 2: sort each bucket using any standard sorting algorithm.

Application: Bucket Sort

- X_j: the number of elements landing in bucket j.
- The time to sort bucket j: $c(X_j)^2$, for some constant c.
- The expected time for sorting in Stage 2:

$$\mathbf{E}\left[\sum_{j=1}^{n} c(X_j)^2\right] = c\sum_{j=1}^{n} \mathbf{E}[X_j^2] = cn\mathbf{E}[X_1^2].$$

Application: Bucket Sort

- X_j: the number of elements landing in bucket j.
- The time to sort bucket j: $c(X_j)^2$, for some constant c.
- The expected time for sorting in Stage 2:

The same for all buckets.

$$\mathbf{E}\left[\sum_{j=1}^{n} c(X_j)^2\right] = c\sum_{j=1}^{n}\boxed{\mathbf{E}[X_j^2]} = cn\mathbf{E}[X_1^2].$$

Application: Bucket Sort

- X_j: the number of elements landing in bucket j.
- The time to sort bucket j: $c(X_j)^2$, for some constant c.
- The expected time for sorting in Stage 2:

$$\mathbf{E}\left[\sum_{j=1}^{n} c(X_j)^2\right] = c\sum_{j=1}^{n}\mathbf{E}[X_j^2] = cn\mathbf{E}[X_1^2].$$

$$\begin{aligned}
\mathbf{E}[X_1^2] &= \mathbf{Var}[X_1] + (\mathbf{E}[X_1])^2 \\
&= n \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right) + \left(n \cdot \frac{1}{n}\right)^2 \\
&= 2 - \frac{1}{n} \\
&< 2.
\end{aligned}$$

Assignment 05

- Suppose that *n* balls are thrown independently and uniformly at random into *n* bins.
  - Find the conditional *probability* that bin 1 has one ball given that exactly one ball fell into the first three bins.
  - Find the conditional expectation of the number of balls in bin 1 under the condition that bin 2 received no balls.

Questions

- What is the probability that **a given bin is empty**?
- What is the expected number of **empty bins**?

- The probability that the *i*th bin remains empty is

$$\left(1 - \frac{1}{n}\right)^m \approx e^{-m/n}.$$

- The probability that the *i*th bin remains empty is

$$\left(1 - \frac{1}{n}\right)^m \approx e^{-m/n}.$$

- The expected number of empty bins:

$$n \cdot \left(1 - \frac{1}{n}\right)^m \approx ne^{-m/n}.$$

- The probability that the ith bin remains empty is

$$\left(1-\frac{1}{n}\right)^m \approx e^{-m/n}.$$

- The expected number of empty bins:

$$n \cdot \left(1-\frac{1}{n}\right)^m \approx n\boxed{e^{-m/n}}.$$

The expected “fraction”

- Generalization:

  The probability that a given bin has ***r*** balls is

- Generalization:

The probability that a given bin has ***r*** balls is

$$
\begin{aligned}
p_r &= \binom{m}{r}\left(\frac{1}{n}\right)^r\left(1-\frac{1}{n}\right)^{m-r} \\
&= \frac{1}{r!}\frac{m(m-1)\cdots(m-r+1)}{n^r}\left(1-\frac{1}{n}\right)^{m-r}. \\
&\approx \frac{e^{-m/n}(m/n)^r}{r!}.
\end{aligned}
$$

Expected number of bins with exactly *r* balls $\approx np_r$.

The Poisson Distribution

Siméon Poisson
(1781−1840)
Wikipedia

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

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

The Poisson Distribution

Siméon Poisson
(1781−1840)
Wikipedia

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

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

$$\Updownarrow$$

$$\frac{e^{-m/n}(m/n)^r}{r!}.$$

The Poisson Distribution

Siméon Poisson
(1781−1840)
Wikipedia

- Try to verify if it's proper:

$$\begin{aligned}\sum_{j=0}^{\infty}\Pr[X=j] &= \sum_{j=0}^{\infty}\frac{e^{-\mu}\mu^j}{j!}\\ &= e^{-\mu}\sum_{j=0}^{\infty}\frac{\mu^j}{j!}\\ &= 1.\end{aligned}$$

The Poisson Distribution

Siméon Poisson
(1781−1840)
Wikipedia

- The expectation:

$$
\begin{aligned}
\mathbf{E}[X] &= \sum_{j=0}^{\infty} j \Pr[X = j] = \sum_{j=1}^{\infty} j \frac{e^{-\mu}\mu^{j}}{j!} \\
&= \mu \cdot \sum_{j=1}^{\infty} \frac{e^{-\mu}\mu^{j-1}}{(j-1)!} \\
&= \mu \cdot \sum_{j=0}^{\infty} \frac{e^{-\mu}\mu^{j}}{j!} \\
&= \mu.
\end{aligned}
$$

The Poisson Distribution

Siméon Poisson
(1781−1840)
Wikipedia

- The sum of finite number of independent Poisson random variables is a Poisson random variable.

$$
\begin{aligned}
\Pr[X+Y=j] &= \sum_{k=0}^{j} \Pr[(X=k)\cap(Y=j-k)] \\
&= \sum_{k=0}^{j} \frac{e^{-\mu_1}\mu_1^k}{k!}\frac{e^{-\mu_2}\mu_2^{(j-k)}}{(j-k)!} \\
&= \frac{e^{-(\mu_1+\mu_2)}}{j!}\sum_{k=0}^{j}\frac{j!}{k!(j-k)!}\mu_1^k\mu_2^{(j-k)} \\
&= \frac{e^{-(\mu_1+\mu_2)}}{j!}\sum_{k=0}^{j}\binom{j}{k}\mu_1^k\mu_2^{(j-k)} \\
&= \frac{e^{-(\mu_1+\mu_2)}(\mu_1+\mu_2)^j)}{j!}.
\end{aligned}
$$

The Poisson Distribution

Siméon Poisson
(1781−1840)
Wikipedia

- The moment generating function of a Poisson random variable with parameter μ is

$$M_X(t) = e^{\mu(e^t - 1)}.$$

The Poisson Distribution

Siméon Poisson
(1781−1840)
Wikipedia

- The moment generating function of a Poisson random variable with parameter μ is

$$M_X(t) = e^{\mu(e^t-1)}.$$

For any t,

$$\mathbf{E}[e^{tX}] = \sum_{k=0}^{\infty} \frac{e^{-\mu}\mu^k}{k!} e^{tk} = e^{\mu(e^t-1)} \sum_{k=0}^{\infty} \frac{e^{-\mu e^t}(\mu e^t)^k}{k!}$$

$$= \quad e^{\mu(e^t-1)}.$$

The Poisson Distribution

Siméon Poisson
(1781−1840)
Wikipedia

- The moment generating function of a Poisson random variable with parameter μ is

$$M_X(t) = e^{\mu(e^t-1)}.$$

For any t,

$$\mathbf{E}[e^{tX}] = \sum_{k=0}^{\infty} \frac{e^{-\mu}\mu^k}{k!} e^{tk} = e^{\mu(e^t-1)} \sum_{k=0}^{\infty} \frac{e^{-\mu e^t}(\mu e^t)^k}{k!}$$

$$= \quad e^{\mu(e^t-1)}.$$

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{(\mu_1+\mu_2)(e^t-1)}.$$

Limit of the Binomial Distribution

- Theorem. Let $X_n \sim$ Binomial(n, p) be a binomial random variable
  - p: a function of n
  - $\lim_{n\to\infty} np = \lambda$ is a constant, independent of n.

  Then for any fixed k,

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

Scenario

- m balls into n bins.
  - $m = f(n)$;
  - $\lim_{m\to\infty} m/n = \lambda$;
  - X_m: the number of balls in a specific bin.
    - Binomial(m, $1/n$).

- From the theorem:

$$\lim_{m\to\infty} \Pr[X_m = r] = \frac{e^{-m/n}(m/n)^r}{r!}.$$