Suppose you are a content delivery network.

In a five-minute period, you get a certain number m of requests. Each needs to be served from one of your n servers.

How to distribute requests to balance the load?

- round robin/find the server with lowest load/etc.: require state, communication, resulting in delays.
- assign each job to a random server: how effective is this policy?
  - How many servers have no job?
  - How many servers have at least k jobs?
  - What is the maximum number of jobs in any server?
  - ...

Balls and Bins - Occupancy Problems

Assume that m balls are placed randomly is n boxes.

- How many boxes are empty?
- How many boxes have at least k balls?
- What is the maximum number of balls in any box?
- ...

Balls and Bins - Examples

Models many situations:

- Load balancing: balls = jobs, bins = servers;
- Data storage: balls = files, bins = disks;
- Hashing: balls = data keys, bins = hash table slots;
- Coupon Collector: balls = purchased coupons; bins = coupon types;
- ...

The Birthday Paradox

Having thirty people in the room, is it more likely or not that some two people in the room share the same birthday?

Assume birthdays uniformly distributed in $[1, \ldots, 365]$.

Count the configurations where no two people share a birthday. The probability that all birthdays are distinct is:

$$\frac{\binom{365}{30} 30!}{365^{30}} \approx 0.29. \tag{1}$$

We can also calculate this probability by considering one person at a time:

$$\left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdot \left(1 - \frac{3}{365}\right) \cdot \ldots \cdot \left(1 - \frac{29}{365}\right)$$

More generally, if there are m people and n possible birthdays, the probability that all m have different birthdays is

$$\left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdot\left(1-\frac{3}{n}\right)\cdot\ldots\cdot\left(1-\frac{m-1}{n}\right)$$

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

Using $1-\frac{k}{n}\approx \mathrm{e}^{-k/n}$ when $k << n$,

$$\begin{aligned}\prod_{j=1}^{m-1}\left(1-\frac{j}{n}\right) &\approx \prod_{j=1}^{m-1}\mathrm{e}^{-j/n}\\ &= \mathrm{e}^{-\sum_{j=1}^{m-1} j/n}\\ &= \mathrm{e}^{-m(m-1)/2n}\\ &\approx \mathrm{e}^{-m^2/2n}.\end{aligned}$$

We place m balls randomly into n bins, how many bins remain empty?
The probability that a given bin is missed by all m balls is

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

Let $X_j = 1$ if the j-th bin is empty else $X_j = 0$.

$E[X_j] = (1-\frac{1}{n})^m$.

$X = \sum_{j=1}^n X_j$ (number of empty bins)

$$E[X] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = n\left(1-\frac{1}{n}\right)^m \approx n\mathrm{e}^{-m/n}.$$

$$\Pr(X=0) = ?$$

How many bins have r balls?
The probability that a given bin has r balls is

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

For $m, n >> r$, (using that $(1-x) < e^{-x}$ for small x)

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

The Poisson distribution

Definition

A discrete Poisson random variable X with parameter μ (denoted by $P(\mu)$) is given by the following probability distribution on $j = 0, 1, 2, \ldots$.

$$\Pr(X = j) = \frac{e^{-\mu}\mu^j}{j!}.$$

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

Expectation

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

Lemma

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

Proof.

Consider two independent Poisson random variables X and Y with means μ_1 and μ_2. Now

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

□

Another Proof

$$E[\mathrm{e}^{tX}] = \sum_{k=0}^{\infty} \frac{\mathrm{e}^{-\mu}\mu^k}{k!}\mathrm{e}^{tk}$$

$$= \mathrm{e}^{\mu(\mathrm{e}^t-1)} \sum_{k=0}^{\infty} \frac{\mathrm{e}^{-\mu \mathrm{e}^t}\left(\mu \mathrm{e}^t\right)^k}{k!} = \mathrm{e}^{\mu(\mathrm{e}^t-1)}.$$

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

$$= \mathrm{e}^{(\mu_1+\mu_2)(\mathrm{e}^t-1)}$$

which is the moment generating function of a Poisson distribution with expectation $\mu_1 + \mu_2$.

Chernoff bound

Theorem

Let X be a Poisson random variable with parameter μ.

1. *If $x > \mu$, then $\Pr(X \geq x) \leq \frac{\mathrm{e}^{-\mu}(\mathrm{e}\mu)^x}{x^x}$;*
2. *If $x < \mu$, then $\Pr(X \leq x) \leq \frac{\mathrm{e}^{-\mu}(\mathrm{e}\mu)^x}{x^x}$.*

Proof.

For any $t > 0$ and $x > \mu$,

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

Hence

$$\Pr(X \geq x) \leq \mathrm{e}^{\mu(\mathrm{e}^t - 1) - tx}.$$

Choosing $t = \ln(x/\mu) > 0$ gives

$$\begin{aligned} \Pr(X \geq x) &\leq \mathrm{e}^{x - \mu - x \ln(x/\mu)} \\ &= \frac{\mathrm{e}^{-\mu}(\mathrm{e}\mu)^x}{x^x}. \end{aligned}$$

$\square$

Proof.

For any $t < 0$ and $x < \mu$,

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

Hence

$$\Pr(X \leq x) \leq \mathrm{e}^{\mu(\mathrm{e}^t - 1) - xt}.$$

Choosing $t = \ln(x/\mu) < 0$, gives

$$\begin{aligned} \Pr(X \leq x) &\leq \mathrm{e}^{x - \mu - x\ln(x/\mu)} \\ &= \frac{\mathrm{e}^{-\mu}(\mathrm{e}\mu)^x}{x^x}. \end{aligned}$$

$\square$

Limit of Binomial Distribution

Theorem

Let X_n be a binomial random variable with parameters n and p, where p is a function of n and $\lim_{n\to\infty} np = \lambda$ is a constant independent of n. Then for any fixed k,

$$\lim_{n\to\infty} \Pr(X_n = k) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}.$$

"Law of Rare Events"

“Law of Rare Events”

Some events which empirically have a Poisson distribution (according to sources on the Internet):

- Typos per page in printed books.
- Number of bomb hits per $.25\text{km}^2$ in South London during World War II.
- The number of soldiers killed by horse-kicks each year in each corps in the Prussian cavalry in the (late) 19th century.
- The number of goals in sports involving two competing teams.
- The number of yeast cells used when brewing Guinness beer.

Maximum per bin

Lemma

When n balls are thrown independently and uniformly at random into n bins, the probability that the maximum load is more than $3 \ln n / \ln \ln n$ is at most $1/n$ for n sufficiently large.

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

$$\binom{n}{M}\left(\frac{1}{n}\right)^M \leq \frac{1}{M!} \leq \left(\frac{\mathrm{e}}{M}\right)^M .$$

We use:

$$\frac{k^k}{k!} < \sum_{i=0}^{\infty} \frac{k^i}{i!} = \mathrm{e}^k \quad \Rightarrow \quad k! > \left(\frac{k}{\mathrm{e}}\right)^k .$$

The probability that any bin receives at least $M \geq 3 \ln n / \ln \ln n$ balls is bounded above by

$$
\begin{aligned}
n \left(\frac{e}{M} \right)^{M} &\leq n \left(\frac{e \ln \ln n}{3 \ln n} \right)^{3 \ln n / \ln \ln n} \\
&\leq n \left(\frac{\ln \ln n}{\ln n} \right)^{3 \ln n / \ln \ln n} \\
&= e^{\ln n} \left(e^{\ln \ln \ln n - \ln \ln n} \right)^{\frac{3 \ln n}{\ln \ln n}} \\
&= e^{-2 \ln n + \frac{3 (\ln n)(\ln \ln \ln n)}{\ln \ln n}} \\
&\leq \frac{1}{n} \qquad \text{(when } n \text{ is large enough)}
\end{aligned}
$$

Example: Number of Empty Bins

Theorem

Assume that m balls are placed randomly in n boxes. Assume that $m, n \to \infty$, such that the quantity $\lambda = ne^{-m/n}$ is bounded. Let $P_r(n, m)$ be the probability that exactly r boxes are empty. For each fixed r,

$$\lim_{n,m\to\infty} P_r(n, m) = e^{-\lambda}\frac{\lambda^r}{r!}.$$

X has a Poisson distribution with parameter λ:

$$Pr(X = i) \;=\; e^{-\lambda}\frac{\lambda^i}{i!}.$$

Inclusion-Exclusion

Let $E_1, ..., E_n$ be arbitrary events:

$$\begin{aligned} Pr(\cup_{i=1}^{n} E_i) &= \sum_{i=1}^{n} Pr(E_i) - \sum_{i<j} Pr(E_i \cap E_j) \\ &+ \sum_{i<j<k} Pr(E_i \cap E_j \cap E_k) \\ &- \cdots + (-1)^{\ell+1} \sum_{i_1<i_2<\cdots<i_\ell} Pr(\cap_{r=1}^{\ell} E_r) + \cdots \end{aligned}$$

Proof of the Theorem

We first compute $P_0(m, n)$.
Let E_1 be the event "box i is empty".

$$1 - P_0(m, n) \;=\; Pr(\cup_{i=1}^{n} E_i)$$

$$Pr(E_i) = \left(1 - \frac{1}{n}\right)^m$$

$$Pr(\cap_{i=1}^{k} E_i) = \left(1 - \frac{k}{n}\right)^m$$

$$\sum_{i_1 < i_2 < \cdots < i_k} Pr(\cap_{j=1}^{k} E_{i_j}) = \binom{n}{k}\left(1 - \frac{k}{n}\right)^m$$

Lemma

For any fixed $k > 0$,

$$\lim_{n,m\to\infty} \binom{n}{k} \left(1 - \frac{k}{n}\right)^m = \frac{\lambda^k}{k!}$$

Proof.

$$\frac{n^k}{k!} \left(1 - \frac{k}{n}\right)^k \leq \binom{n}{k} \leq \frac{n^k}{k!}$$

$$e^{-km/n} \left(1 - \frac{k^2}{n^2}\right) \leq \left(1 - \frac{k}{n}\right)^m \leq e^{-km/n}$$

$\square$

$$P_0(m,n) \; = \; 1 - \lambda + \frac{\lambda^2}{2} - \frac{\lambda^3}{3!} + ...$$

$$P_0(m,n) \; = \; \sum_{i=0}^{n} (-1)^i \frac{\lambda^i}{i!}$$

$$P_0(m,n) \to e^{-\lambda}$$

$$P_r(m,n) = \binom{n}{r} \left(1 - \frac{r}{n}\right)^m P_0(m, n-r)$$

$$P_r(m,n) \to \frac{\lambda^r}{r!} e^{-\lambda}$$