Beating the Birthday Paradox in Dining Cryptographer Networks

Pablo García[1] **Jeroen van de Graaf**[2] Alejandro Hevia[3] Alfredo Viola[4]

Universidad Nacional de San Luis, Argentina

Depto. de Ciência da Computação, Universidade Federal de Minas Gerais, Brazil

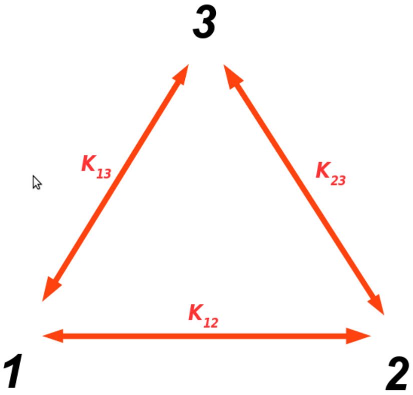
Dept. of Computer Science, University of Chile, Chile

Instituto de Computación, Universidad de la República, Uruguay

Project CEV: Challenges on Electronic Voting
Kick-off Meeting
May 6, 2015

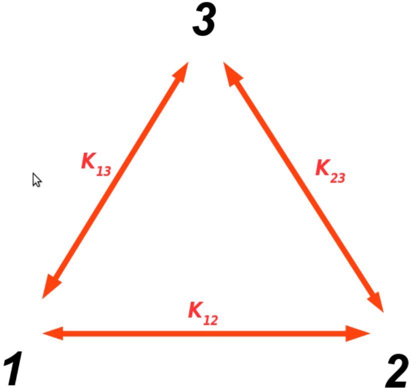
The Dining Cryptographers protocol

Preliminary phase: parties exchange random keypads



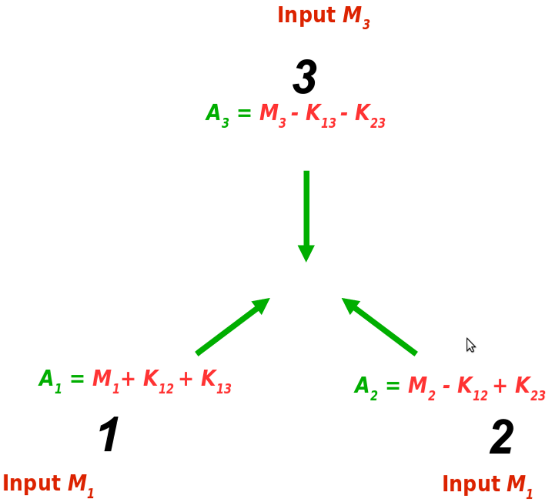
The Dining Cryptographers protocol

Preliminary phase: parties exchange random keypads



The Dining Cryptographers protocol

Broadcast phase: parties compute and broadcast their contribution



Input M_3

3

$A_3 = M_3 - K_{13} - K_{23}$

$A_1 = M_1 + K_{12} + K_{13}$

1

Input M_1

$A_2 = M_2 - K_{12} + K_{23}$

2

Input M_1

The Dining Cryptographers protocol

Consolidation phase: parties compute the message

$$
\begin{aligned}
A_1 + A_2 + A_3 = {} & M_1 + K_{12} + K_{13} \\
& + M_2 - K_{12} + K_{23} \\
& + M_3 - K_{13} - K_{23} \\
& = M_1 + M_2 + M_3
\end{aligned}
$$

Towards a DC Net

- Use bitwise XOR of strings, or addition/multiplication in some cyclic group.
- Repeat the protocol sequentially; we call these slots.
- Party i randomly choose a slot send their message M_i.
- Parties send the zero message in all the other slots.
- Problem can be modelled as a **balls-in-bins** problem.
- Amenable to be analyzed with techniques based on *Analytic Combinatorics*.

–	–	–	**2**	–	–	–	–	–	–	–	–
–	–	–	**8**	–	–	**1**	–	–	–	–	–
7	–	5	**6**	–	–	**4**	3	–	–	–	–

Problems:

1. Even if everybody is honest, collisions may occur
2. Dishonest parties can deviate from the protocol and disrupt communication

Why study DC Nets?

Anonymous Message Broadcast is usually implemented through Mix Nets:

- Based on public-key primitives → computational privacy only
- Many voting protocols use Mix nets for universal verifiability
- In voting, unconditional privacy is not sufficient

Q: What alternatives do we have?

Why study DC nets?

Anonymous Message Broadcast can be implemented through DC Nets:

- assumes private channels between participants
  - Admittedly a potential bottle-neck
  - Alternative setting with many users/voters and a small number of authorities mitigates the problem
  - Alternative Key Distribution mechanisms (post-quantum?) become now an option.
- assumes a broadcast channel
- The original DC Net provides **unconditional privacy**

Q: Can we substitute the Mix Net used in Tor using DC Nets?

- Practical prototypes based on DC Nets exist: Herbivore (Cornell), Consent + Verdict (Yale)
- Use DHKE with pairing and PRNG to create the keypads → **computational privacy**
- Some use Mix Nets to do slot reservation

Q: Can we make unconditionally private DC Nets practical?

2 Random Allocations

Collisions are barely avoidable, distrib. is irregular.

m = # Urns, n = # Balls

Theorem 0.

(i) Collisions occur early = *Birthday Paradox*

$$Ex\{\textit{First collision } /m \textit{ cells}\} \sim \sqrt{\frac{\pi m}{2}}$$

(ii) Probability of no collisions in a full table is

$$\Pr\{\textit{No collision } n = m\} = \frac{n!}{n^n} \sim \frac{e^{-n}}{\sqrt{2\pi n}}.$$

(iii) Empty cells disappear late = *Coupon Collector*

$$Ex\{\textit{All } m \textit{ cells non-empty}\} \sim m \cdot \log m.$$

(iv) Even in an α-sparse allocation, $\alpha = \frac{n}{m}$

$$Ex\{\textit{Max bucket occupancy}\} \sim \frac{\log n}{\log\log n}$$

Collision management is a necessity

The Poisson Law governs balls-in-urns models

Thow n balls into m buckets. Let

$$\boxed{\alpha = \frac{n}{m}}$$

be fixed, $0 < \alpha < \infty$. Then, asymptotically $(m, n \to +\infty)$

- the proportion of empty urns is $e^{-\alpha}$, 36% for $\alpha = 1$;
- the proportion of k-urns is a Poisson law of param. α

$$\text{Poisson}(\alpha, k) := e^{-\alpha}\frac{\alpha^k}{k\,!}.$$

Proof.

$$\begin{array}{lll} \text{empty urns} & m \times \left(\frac{m-1}{m}\right)^n & \sim m \times e^{-\alpha}. \\ k\text{–urns:} & m \times \binom{n}{k} \cdot \frac{(m-1)^{n-k}}{m^n} & \sim m \times e^{-\alpha}\frac{\alpha^k}{k!} \end{array}$$

$\Longrightarrow$ Analysis of separate chaining:

$$Ex\{C^-(m,n)\} \sim 1 + \sum_{k=1}^{\infty} k \cdot e^{-\alpha}\frac{\alpha^k}{k!} \equiv \boxed{1+\alpha}.$$

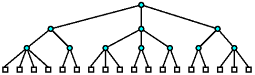
$$Ex\{C^+(m,n)\} \sim \boxed{1 + \tfrac{\alpha}{2}}$$

4 Analytic Combinatorics

Two basic principles ↦ "dictionaries"

SYMBOLIC METHODS

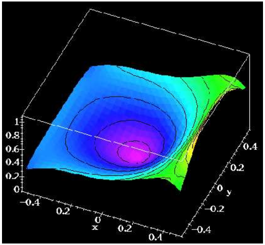
Generating functions



$\mapsto \quad \mathbf{z}^{11}$

$$\mathbf{z} + \mathbf{z}^2 + \mathbf{z}^3 + 2\,\mathbf{z}^4 + 2\,\mathbf{z}^5 + 4\,\mathbf{z}^6 + 5\,\mathbf{z}^7 + 9\,\mathbf{z}^8 + \cdots$$

Analytic functions and singularities



CONSTRUCTIONS

Dictionary (I)

$$\mathcal{F} \quad \mapsto \quad \{f_n\} \quad \mapsto \quad f(z) = \sum_n f_n \frac{z^n}{n!}.$$

$$\frac{1}{1-f} = 1 + f + f^2 + f^3 + \cdots$$

$$\exp(f) = 1 + f + \frac{1}{2!} f^2 + \frac{1}{3!} f^3 + \cdots$$

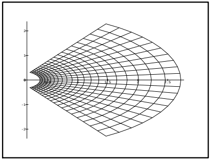
$$\begin{array}{lcl}
\mathbf{A} \cup \mathbf{B} & \mapsto & A(z) + B(z) \\
\mathbf{A} \times \mathbf{B} & \mapsto & A(z) \times B(z) \\
\mathsf{Seq}\ \mathbf{A} & \mapsto & \dfrac{1}{1 - A(z)} \\
\mathsf{Set}\ \mathbf{A} & \mapsto & \exp(A(z)) \\
\mathsf{Cycle}\ \mathbf{A} & \mapsto & \log \dfrac{1}{1 - A(z)}
\end{array}$$

COMPLEX ASYMPTOTICS

Dictionary (II)

Point of regularity. $f(z) \approx f(z_0) + f'(z_0)(z - z_0)$

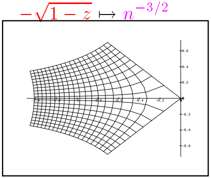
$\exp(z)$



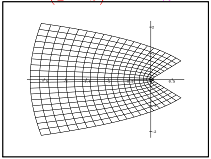
Point of singularity. $\neg\exists \; \frac{d}{dz} f(z)\Big|_{z_0}$

$$(1-z)^{\beta} \;\mapsto\; \frac{n^{\beta-1}}{\Gamma(\beta)}$$

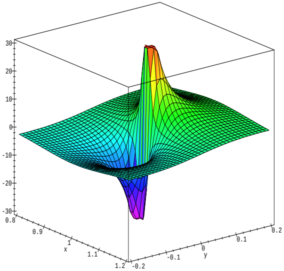
$-\sqrt{1-z} \mapsto n^{-3/2}$



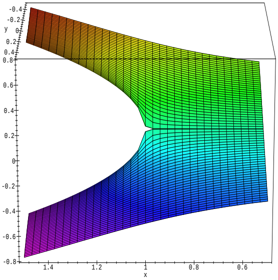
$-(1-z)^{3/2} \mapsto n^{-5/2}$



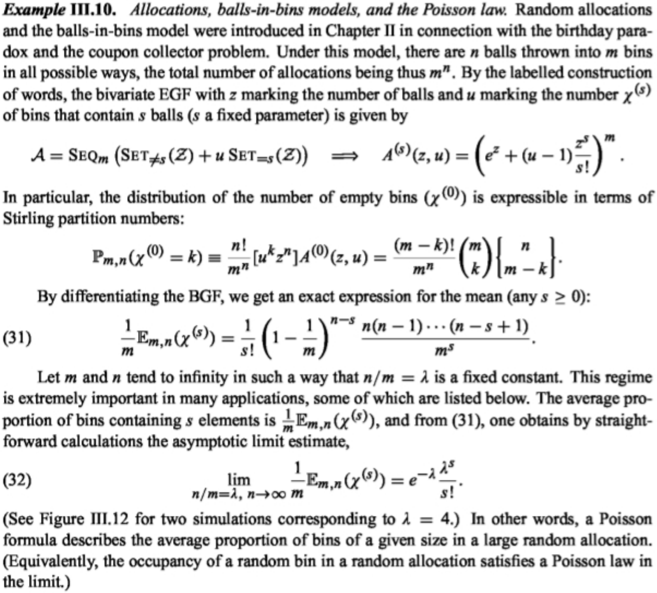
Permutations: $(1-z)^{-1}$



Trees: $1-\sqrt{1-z}$



Basics on balls-in-bins



***Example* III.10.** *Allocations, balls-in-bins models, and the Poisson law.* Random allocations and the balls-in-bins model were introduced in Chapter II in connection with the birthday paradox and the coupon collector problem. Under this model, there are n balls thrown into m bins in all possible ways, the total number of allocations being thus m^n. By the labelled construction of words, the bivariate EGF with z marking the number of balls and u marking the number $\chi^{(s)}$ of bins that contain s balls (s a fixed parameter) is given by

$$\mathcal{A} = \textsc{Seq}_m\left(\textsc{Set}_{\neq s}(\mathcal{Z}) + u\,\textsc{Set}_{=s}(\mathcal{Z})\right) \quad\Longrightarrow\quad A^{(s)}(z,u) = \left(e^z + (u-1)\frac{z^s}{s!}\right)^m .$$

In particular, the distribution of the number of empty bins ($\chi^{(0)}$) is expressible in terms of Stirling partition numbers:

$$\mathbb{P}_{m,n}(\chi^{(0)} = k) \equiv \frac{n!}{m^n}[u^k z^n]A^{(0)}(z,u) = \frac{(m-k)!}{m^n}\binom{m}{k}\left\{ {n \atop m-k} \right\}.$$

By differentiating the BGF, we get an exact expression for the mean (any $s \geq 0$):

$$(31) \qquad \frac{1}{m}\mathbb{E}_{m,n}(\chi^{(s)}) = \frac{1}{s!}\left(1 - \frac{1}{m}\right)^{n-s}\frac{n(n-1)\cdots(n-s+1)}{m^s}.$$

Let m and n tend to infinity in such a way that $n/m = \lambda$ is a fixed constant. This regime is extremely important in many applications, some of which are listed below. The average proportion of bins containing s elements is $\frac{1}{m}\mathbb{E}_{m,n}(\chi^{(s)})$, and from (31), one obtains by straightforward calculations the asymptotic limit estimate,

$$(32) \qquad \lim_{n/m=\lambda,\ n\to\infty} \frac{1}{m}\mathbb{E}_{m,n}(\chi^{(s)}) = e^{-\lambda}\frac{\lambda^s}{s!}.$$

(See Figure III.12 for two simulations corresponding to $\lambda = 4$.) In other words, a Poisson formula describes the average proportion of bins of a given size in a large random allocation. (Equivalently, the occupancy of a random bin in a random allocation satisfies a Poisson law in the limit.)

An example

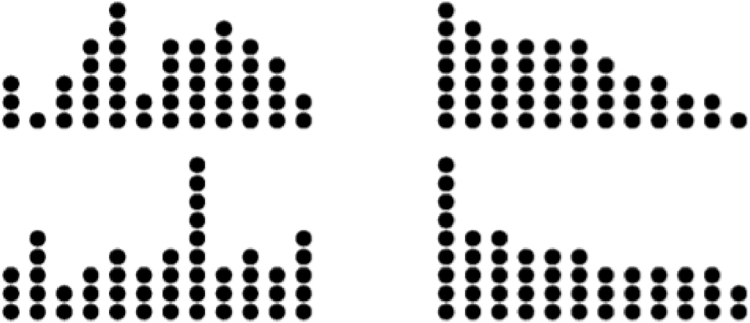


Figure III.12. Two random allocations with $m = 12$, $n = 48$, corresponding to $\lambda \equiv n/m = 4$ (left). The right-most diagrams display the bins sorted by decreasing order of occupancy.

First solution: sequential repetition

Suppose everybody behaves honestly. How can we deal with collisions?

–	–	–	**2**	–	–	–	–	–	–	–	–
–	–	–	**8**	–	–	**1**	–	–	–	–	–
7	–	5	**6**	–	–	**4**	3	–	–	–	–

Obvious strategy:

- Parties who see their message got through start sending the null message
- Repeat the process until every party has succeeded

Protocol for sequential repetition

#messages $n = 8$; #slots $m = 12$;

Round 1:

–	–	–	**2**	–	–	–	–	–	–	–	–
–	–	–	**8**	–	–	**1**	–	–	–	–	–
7	–	5	**6**	–	–	**4**	3	–	–	–	–

Round 2:

–	–	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	**2**	–
8	6	–	–	–	–	–	4	–	–	**1**	–

Protocol for sequential repetition

#messages $n = 8$; #slots $m = 12$;

Round 1:

–	–	–	**2**	–	–	–	–	–	–	–	–
–	–	–	**8**	–	–	**1**	–	–	–	–	–
7	–	5	**6**	–	–	**4**	3	–	–	–	–

Round 2:

–	–	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	**2**	–
8	6	–	–	–	–	–	4	–	–	**1**	–

Channel 3:

–	–	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	–	–
–	–	–	–	–	–	–	2	–	1	–	–

An exact analysis for the sequential repetition

- Let $G_m(u,z) = (e^{z/m} + (u-1)z/m)^m$.
- The coefficient $n![u^k z^n]G_m(u,z)$ is the probability that when throwing n balls in m bins we have k bins with one ball.
- These bins are filled by the messages that go through, and as a consequence do not have to be sent again in the next round.

Theorem

Let

$$\hat{G}_m(u,z,) = G_m(u/z,z)^r. \tag{1}$$

Then $n![u^n, z^{r-1}n]\hat{G}_m(u,z)$ is the probability generating function that all messages go through in less than or equal r rounds when n messages are sent in m slots, using the sequential algorithm.

Sketch of the proof.

The normalization u/z "marks" the number of balls that go through in each round. At the end n bins have been filled, and since n balls were originally thrown, we have to consider the coefficient z^{r-1}.

□.

Result for sequential repetition

- Even though this is an exact approach, it is far from trivial to extract coefficients.
- By the Poisson Law that governs the asymptotic behaviour of the problem in the case $n = \lambda m$ with $\lambda > 0$ a constant, the *expected* number of bins with one ball when $m, n \longrightarrow \infty$ and $n = \lambda m$ is $m\lambda e^{-\lambda}$.
- It can be seen that this value is concentrated around its mean.
- Let $n_0 = n$ be the initial number of messages
- Let $\lambda_0 = \frac{n_0}{m}$ (the "density" of message over the slots)
- E[number of **slots** with 1 message in round 1] $= n_0 e^{-\lambda_0}$.
- E[number of **messages** that fail in round 1] $= n_0 - n_0 e^{-\lambda_0}$.
- So $n_1 = n_0 - n_0 e^{-\lambda_0}$.

Result for sequential repetition

- Even though this is an exact approach, it is far from trivial to extract coefficients.
- By the Poisson Law that governs the asymptotic behaviour of the problem in the case $n = \lambda m$ with $\lambda > 0$ a constant, the *expected* number of bins with one ball when $m, n \to \infty$ and $n = \lambda m$ is $m\lambda e^{-\lambda}$.
- It can be seen that this value is concentrated around its mean.
- Let $n_0 = n$ be the initial number of messages
- Let $\lambda_0 = \frac{n_0}{m}$ (the "density" of message over the slots)
- E[number of **slots** with 1 message in round 1] $= n_0 e^{-\lambda_0}$.
- E[number of **messages** that fail in round 1] $= n_0 - n_0 e^{-\lambda_0}$.
- So $n_1 = n_0 - n_0 e^{-\lambda_0}$.
- In general: $n_{i+1} = \lambda_{i+1} m$ with $\lambda_{i+1} = \lambda_i(1 - e^{-\lambda_i})$.
- Let f denote the number of iterations needed. Then the recursion ends when $n_f < 1$, that is, when $\lambda_f < 1/m = \lambda_0/n$.
- We can show that, with overwhelming probability, f lies in an interval $[low(\delta), high(\delta)]$ where δ is some auxiliary parameter from the Chernoff bound.

Result for sequential repetition

Theorem

Let $0 < \lambda_0 = n/m < 1$ *and* $0 < \delta < 1$ *such that* $(1+\delta)\lambda_0 < 1$*, and let*

$$low(n, \lambda_0, \delta) := \lfloor \lg\left(\log\left(n/\lambda_0\right)\right) - \lg\left(-\log(((1-\delta)(1-\lambda_0/2)\lambda_0)\right)\rfloor$$

and

$$high(n, \lambda_0, \delta) := \lceil \lg\left(\log\left(n/\lambda_0\right)\right) - \lg\left(-\log((1+\delta)\lambda_0)\right)\rceil .$$

Then,

$$Pr[low(n, \lambda_0, \delta) \ \leq \ f \ \leq \ high(n, \lambda_0, \delta)] \ > \ 1 \ - e^{\ -\frac{\delta^2(1-e^{-\lambda_0})}{3}n}.$$

where lg *denotes the logarithm in base 2.*

Result for sequential repetition

Theorem

Let $0 < \lambda_0 = n/m < 1$ *and* $0 < \delta < 1$ *such that* $(1+\delta)\lambda_0 < 1$, *and let*

$$low(n, \lambda_0, \delta) := \lfloor \lg\left(\log\left(n/\lambda_0\right)\right) - \lg\left(-\log(((1-\delta)(1-\lambda_0/2)\lambda_0)\right)\rfloor$$

and

$$high(n, \lambda_0, \delta) := \lceil \lg\left(\log\left(n/\lambda_0\right)\right) - \lg\left(-\log((1+\delta)\lambda_0)\right)\rceil .$$

Then,

$$Pr[low(n, \lambda_0, \delta) \ \leq \ f \ \leq \ high(n, \lambda_0, \delta)] \ > \ 1 \ - e^{-\frac{\delta^2(1-e^{-\lambda_0})}{3}n}.$$

where lg *denotes the logarithm in base 2.*

Interpretation:

- This leads to $r = O(\log\log(n))$ rounds.
- In terms of space, if $m = n$, we have a $O(n \log\log(n))$ space algorithm.

Comparison with experimental results

n	λ_0	Prob	*low*	$\left\lceil \lg\left(\log\left(\frac{n}{\lambda_0}\right)\right)\right\rceil$	*high*	#*low*	#*low*+1	#*low*+2
10^4	0.3	0.9996	2	4	4	0	76	24
10^4	0.5	0.9999	3	4	4	0	97	*3*
10^4	0.7	0.9999	3	4	5	0	95	5
10^4	0.9	0.9999	3	4	5	0	97	3
10^5	0.3	1	3	4	4	10	90	0
10^5	0.5	1	3	4	4	0	86	*14*
10^5	0.7	1	3	4	5	0	88	12
10^5	0.9	1	3	4	5	0	85	15
10^6	0.3	1	3	4	4	0	100	0
10^6	0.5	1	3	4	5	0	29	71
10^6	0.7	1	3	4	5	0	28	72
10^6	0.9	1	4	4	6	24	76	0

Protocol for parallel repetition

$\#$messages $n = 8$; $\#$slots $m = 12$; $\#$channels $k = 3$; total $\#$slots $S = 36$

Channel 1:

–	–	–	**2**	–	–	–	–	–	–	–	–
–	–	–	**8**	–	–	**1**	–	–	–	–	–
7	–	5	**6**	–	–	**4**	3	–	–	–	–

Channel 2:

–	–	–	–	–	–	–	–	–	–	–	–
–	–	**7**	–	–	–	–	**1**	–	–	–	–
5	–	**2**	–	3	–	–	**8**	–	4	–	6

Channel 3:

–	–	–	–	–	–	–	–	–	–	–	–
–	**6**	–	–	–	–	**4**	–	–	–	–	–
–	**5**	8	1	–	–	**7**	2	–	–	–	3

Result for parallel repetition

Let n be the initial number of messages and m the number of slots per channel

Suppose we have k parallel channels

Denote $S = km$ as the total number of slots.

Result for parallel repetition

Let n be the initial number of messages and m the number of slots per channel

Suppose we have k parallel channels

Denote $S = km$ as the total number of slots.

Let $P_{m,n,k}$ be the probability that all balls are successful if k parallel channels are used.

- Given n and S, we want to find k that maximizes $P_{m,n,k}$.
- Given n and a very small error probability $\varepsilon = 1 - P_{m,n,k} = 1 - 2^b$, we want to find the minimum of total slots S needed.

A previous known Fact

- Here $\left\{ {a \atop b} \right\}$ denotes the Stirling numbers of the second kind (number of ways of placing a elements in b non-empty sets).
- The falling factorials are defined as $m^{\underline{j}} = m(m-1)(m-2)\ldots(m-j+1)$.

Fact

Let $C_{n,m}$ be the number of ways of placing n balls in m bins in such a way that no bin has one ball, and $T_{m,n,r} = \sum_{j\geq 0} m^{\underline{j+r}} C_{n-r,j}$. Morover, let $N_{m,n,r}$ be the number of ways of throwing n balls in m bins in such a way that r of the bins contain 1 ball. Then

1. $C_{n,m} = \sum_{i\geq 0} (-1)^i \binom{n}{i} \left\{ {n-i \atop m-i} \right\}$.
2. $N_{m,n,r} = \binom{n}{r} T_{m,n,r}$.
3. $P_{m,n,1} = \frac{N_{m,n,n}}{m^n} = \frac{m^{\underline{n}}}{m^n}$.

Sketch of the proof.

Using symbolic techniques

$$\begin{aligned} N_{m,n,r} &= n![u^r z^n](e^z + (u-1)z)^m \\ C_{N,M} &= N![z^N](e^z - 1 - z)^M. \end{aligned}$$

A more direct approach will be useful for the generalization.

- Choose r balls to go to the r bins with one ball ($\binom{n}{r}$).
- Choose r bins and place the remaining $n-r$ balls in the other $m-r$ bins in such a way that no bin has 1 ball ($T_{m,n,r}$). This can be achieved as follows
  1. Choose r bins to place 1 ball at each one ($m^{\underline{r}}$).
  2. Chose j bins from the remaining $m-r$ bins to place the $n-r$ balls that give collisions ($(m-r)^{\underline{j}}$). Together with the previous selection, this gives a factor $m^{\underline{r+j}}$.
  3. Place the $(n-r)$ balls in the j bins in such a way that none of these j bins are empy or have one ball ($C_{n-r,j}$).

□

This Fact clearly divides the calculation of $N_{m,n,r}$ in two parts.

- The selection of successful balls ($\binom{n}{r}$).
- The selection of the bins and placement of the unsuccessful balls ($T_{m,n,r}$).

This division is key to derive a general results for all k.

Result for parallel repetition

Theorem

Let $I_k = \{0 \leq j_1 = s_1 \leq n - t\} \cup \{0 \leq j_i \leq s_i \leq n-t,\ 2 \leq i \leq k\}$ *be a set of indices,* $\delta_i = s_i - j_i$ *and* $J_k = \sum_{i=1}^{k} j_i$*. Let* $Q^{(k)}_{m,n,t}$ *be the probability of having* t *failures when throwing* n *balls in* k *parallel channels with sets of* m *bins. Then*

1. $Q^{(k)}_{m,n,t} = \frac{R^{(k)}_{m,n,t}}{n^{km}}$, *with* $R^{(k)}_{m,n,t} = \binom{n}{t} \sum_{I_k} c^{(k)}_{n,t} \prod_{i=1}^{k} N_{m,n,s_i}$ *and*

$$\begin{aligned} c^{(1)}_{n,t} &= 1 \text{ which equals} \binom{J_0}{n-t-s_1},\ \text{when } s_1 = n-t, \\ c^{(k)}_{n,t} &= \frac{c^{(k-1)}_{n,t}}{\binom{J_{k-2}}{n-t-s_{k-1}}} \binom{n-t-J_{k-2}}{j_{k-1}} \binom{J_{k-2}}{\delta_{k-1}} \binom{J_{k-1}}{n-t-s_k}, \quad 2 \leq k. \end{aligned}$$

2. $P_{m,n,k} = Q^{(k)}_{m,n,0}$.

Sketch of the proof.

- $\binom{n}{t}$ gives the number of ways to choose the t balls that fail in all the channels.
- Then $n - t$ balls should fall in a bin with one ball in at least one of the channels.
- s_i is the number successful balls in channel i, j_i is the number of balls whose *first* successful try is channel i and J_i is the *total* number of successful balls up to channel i, with $1 \leq i \leq k$.
- Then $s_1 = j_1$ since this is the first successful channel for these balls, and $j_k = n - t - J_{k-1}$ since at the end $n - t$ balls are successful ($J_k = n - t$).
- Moreover, for $1 \leq i < k$, the j_i balls whose first successful try is channel i should be taken from the ones that have failed in all previous channels ($n - t - J_{i-i}$ of them), giving the factor $\binom{n-t-J_{i-1}}{j_i}$.
- Furthermore, the other $\delta_i = s_i - j_i$ balls should be chosen among the one already successful, giving the factor $\binom{J_{i-1}}{\delta_i}$.
- Since s_k are successful in the last channel, we have to choose the other $n - t - s_k$ unsucessful balls (but *successful* ones in the k channels) among the already successful J_{k-1} balls (giving the factor $\binom{J_{k-1}}{n-t-s_k}$).

Sketch of the proof(cont).

- The recursive definition of $c_{n,t}^{(k)}$ then follows by induction on k (the number of channels).
- Notice that when $k = 1$, $c_{n,t}^{(1)} = \binom{J_0}{n-t-s_1}$.
- Since $J_0 = 0$ this coefficient is 0, unless $s_1 = n - t$, when it takes the value of 1. This is actually the case, since all the $n - t$ balls should be successful in the first try.
- The proof is then completed by noticing that $\prod_{i=1}^{k} N_{m,n,s_i}$ counts the number of ways to place all the $n - s_i$ balls in $m - s_i$ bins in such a way that no bin has one ball.

□

Result for parallel repetition

The exact expression is very difficult to handle with large numbers, and asymptotic results are difficult to achieve.

Therefore we derived an alternative, approximate approach to find an analytic expression, based on Poisson approximations

Let $p = (1 - 1/m)^{n-1}$ and $S = km$.

Let $\hat{P}_{S,n,k}$ be the probability (in this model) that when n balls are thrown in k parallel channels each with m bins, all the balls are successful.

Result for parallel repetition

Theorem

1. $$\hat{P}_{S,n,k} = \left(1 - \left(1 - \left(1 - \frac{k}{S}\right)^{n-1}\right)^{k}\right)^{n}.$$

2. *When $S, n \to \infty$ and $k = o(S)$, then $p \approx e^{\frac{-nk}{S}}$, and then*

$$\hat{P}_{S,n,k} \approx \left(1 - \left(1 - e^{\frac{-nk}{S}}\right)^{k}\right)^{n}.$$

3. *For fixed S and n, the optimal number of channels is $k^* = \lfloor \frac{log(2)S}{n} \rfloor$. The corresponding optimal probability is*

$$\hat{P}^*_{S,n} \approx \left(1 - \frac{1}{2^{\frac{log(2)S}{n}}}\right)^{n}.$$

4. *For given n, the minimal value of S to achieve an exponentially small error $1/2^b$ is*

$$S_{min} = \frac{n}{log(2)}(b + lg(n)),$$

Sketch of the proof

- If n balls are thrown in m bins, then the expected number of bins with 1 ball is

$$E[\#\ bins\ with\ 1\ ball] = n\left(1-\frac{1}{m}\right)^{n-1}. \tag{2}$$

  If $n, m \to \infty$ this value can be approached by $ne^{-n/m}$.

- Its variance is

$$\begin{aligned} Var[\#\ bins\ with\ 1\ ball] &= n(n-1)\left(1-\frac{2}{m}\right)^{n-2}\left(1-\frac{1}{m}\right) \\ &+ n\left(1-\frac{1}{m}\right)^{n-1} - n^2\left(1-\frac{1}{m}\right)^{2n-2}. \end{aligned} \tag{3}$$

  Moreover, since when $n, m \to \infty$ this variance is approached by $ne^{-n/m}(1-e^{-n/m})$, the number of bins with one ball is concentrated around its mean.

- In the approximate model, let $p = (1-1/m)^{n-1}$. So the expected number of bins with one ball is np. Then, p can be interpreted as the probability that a given ball is successful when n balls are thrown in m bins. Furthermore $q = 1 - p$ can be interpreted as the probability that a given ball fails, since it falls in an bin with two or more balls.

Sketch of the proof(cont).

- We consider a Bernoulli process where the probability that a given ball fails in all the k channels is q^k. Hence, with probability $(1 - q^k)$ a given ball succeeds in at least one of the channels. As a consequence, the probability that all the n balls succeed in at least one channel is $(1 - q^k)^n$.
- Parts 1 and 2 are straightforward.
- For part 3, to find the optimal value of k, lets call $x = kn/S$ and so

$$\frac{\partial}{\partial k}\hat{P}_{S,n,k} \approx -\frac{\hat{P}_{S,n,k}\ n(1-e^{-x})^k}{1-(1-e^{-x})^k}\left(log(1-e^{-x}) + \frac{xe^{-x}}{1-e^{-x}}\right).$$

The maximum is at the value x_0 such that $\left(log(1-e^{-x_0}) + \frac{x_0 e^{-x_0}}{1-e^{-x_0}}\right) = 0$, giving $x_0 = log(2)$. As a consequence, the optimal number of channels is $k = \lfloor \frac{Sx_0}{n} \rfloor$, and

$$\hat{P}^*_{S,n} \approx \left(1 - \frac{1}{2^{\frac{Sx_0}{n}}}\right)^n. \tag{4}$$

- Moreover, for part 4, given n and a constant b (like b=80), if we want an exponentially small error in b ($\hat{P}^*_{S,n} = 1 - 1/2^b$, with $b > 0$), then from (4) and the approximation $(1 - \frac{1}{2^b})^{\frac{1}{n}} = 1 - \frac{1}{n2^b} + O(\frac{1}{n^2})$ then S should verify:

$$S = \frac{n}{x_0}(b + lg(n)), \tag{5}$$

where lg is the base 2 logarithm. □

Comparison with experimental results

Experimental results for the probability of success for a total of $S = 360$ and $k = 23$ messages, slightly adjusted values from the birthday paradox.

S	n	k	m	exact	approx1	approx2	experimental
360	23	1	360	0.4877	-	-	0.4861
360	23	2	180	0.7566	0.7348	0.7165	0.7587
360	23	3	120	0.8992	0.8961	0.8848	0.8967
360	23	4	90	0.9500	0.9493	0.9421	0.9497
360	23	5	72	-	0.9704	0.9654	0.9717
360	23	6	60	-	0.9801	0.9763	0.9802
360	23	8	48	-	0.9877	0.9849	0.9878
360	23	9	40	-	0.9891	0.9866	0.9891
360	23	10	36	-	**0.9898**	**0.9874**	**0.9901**
360	23	12	30	-	0.9898	0.9873	0.9897
360	23	15	24	-	0.9869	0.9838	0.9865
360	23	18	20	-	0.9799	0.9759	0.9803
360	23	20	18	-	0.9718	0.9670	0.9717
360	23	24	15	-	0.9410	0.9348	0.9406
360	23	30	12	-	0.8246	0.8226	0.8248

Comparison with experimental results

Experimental results for the number of slots needed to guarantee a probability of failure less than 2^{-b}. Ideally the last column should be 1.

n	b	S_{opt}	k_{opt}	experimental$/(1-2^{-b})$
23	12	549	16	1.000134
23	15	648	19	1.000015
30	12	732	16	1.000119
30	15	862	19	1.000005
40	12	1000	17	1.000034
40	15	1173	20	0.999995
50	12	1273	17	1.000024
50	15	1490	20	1.000010

Conclusion

Sequential repetition

- Always leads to success
- We need a $O(\log\log(n))$ rounds

Parallel repetition

- Leads to success with overwhelming probability
- Total space: given n, the minimum number of total slots to achieve error $\varepsilon < 1/2^b$ is

$$S_{min} = \frac{n}{log(2)}(b + lg(n)),$$

Birthday paradox

From 48 % of success to almost 99% of success using the same overall number of slots, $k = 10$ parallel channels