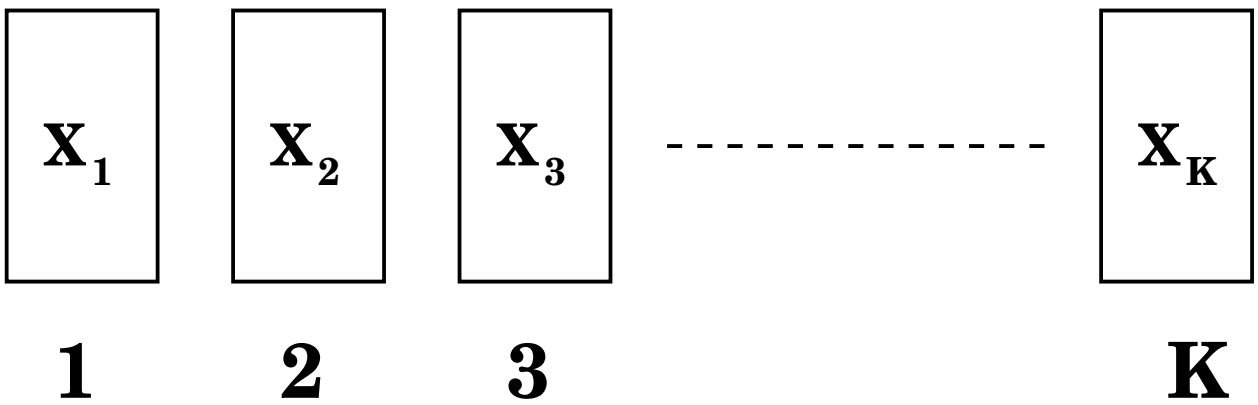
Lecture 5

The birthday problem

Gamblers ruin

Polya's urn

Problem # 3: The birthday problem



$X_1, X_2, \ldots, X_K$ are chosen <u>at random</u> from the numbers $1, 2, \ldots, N$

- What is the probability that <u>at least 2</u> of the K cards have the same (random) number on them?
- If $N = 365$, and cards are people, this is the "birthday problem".

A counting solution

$$\begin{aligned}\Omega &= [1, 2, \ldots, N]^K \\ &= \{\omega = (\omega_1, \ldots, \omega_K),\ \omega_i \in [1, \ldots, N]\}\end{aligned}$$

- $\#(\Omega) = N^K$
- $A = \{\text{at least 2 of } K \text{ cards have same number on them}\}$
- $A^c = \{\text{All } K \text{ cards have different numbers on them}\}$
- $\#(A^c) = N(N-1)(N-2)\ldots(N-K+1)$

$$P\{A\} = 1 - P\{A^c\} = 1 - \frac{N(N-1)(N-2)\ldots(N-K+1)}{N^K}$$

- Note:

$$P\{A^c\} = \left(\frac{N-1}{N}\right)\left(\frac{N-2}{N}\right)\ldots\left(\frac{N-K+1}{N}\right)$$

Multiplication rule for n events:

$$P\{A_1 A_2 \ldots A_n\}$$

$$= P\{A_1\}P\{A_2|A_1\}P\{A_3|A_1A_2\} \times \cdots P\{A_n|A_1A_2\ldots A_{n-1}\}$$

A simpler solution to the birthday problem

- $A = \{$at least 2 of K cards have same number on them$\}$
  - $A_1 =$ Second person has a different birthday to the first.
  - $A_2 =$ Third person has a different birthday to the first 2.
  - $A_{K-1} =$ Last (K-th) has a different b'day to the first $K-1$.
- $A^c = \{$All numbers different$\} = A_1 \cap A_2 \cap \ldots \cap A_{K-1}$.

$$
\begin{aligned}
P\{A\} &= 1 - P\{A^c\} \\
&= 1 - P\{A_1 A_2 \ldots A_{K-1}\} \\
&= 1 - P\{A_1\} P\{A_2|A_1\} P\{A_3|A_1 A_2\} \times \\
&\qquad \times \cdots P\{A_{K-1}|A_1 A_2 \ldots A_{K-2}\} \\
&= 1 - \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \ldots \left(1 - \frac{K-1}{N}\right)
\end{aligned}
$$

Calculus

$$(1) \qquad \frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots$$

$$(2) \qquad \ln(1-x) = \int \frac{-1}{1-x} dx = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} + \ldots$$

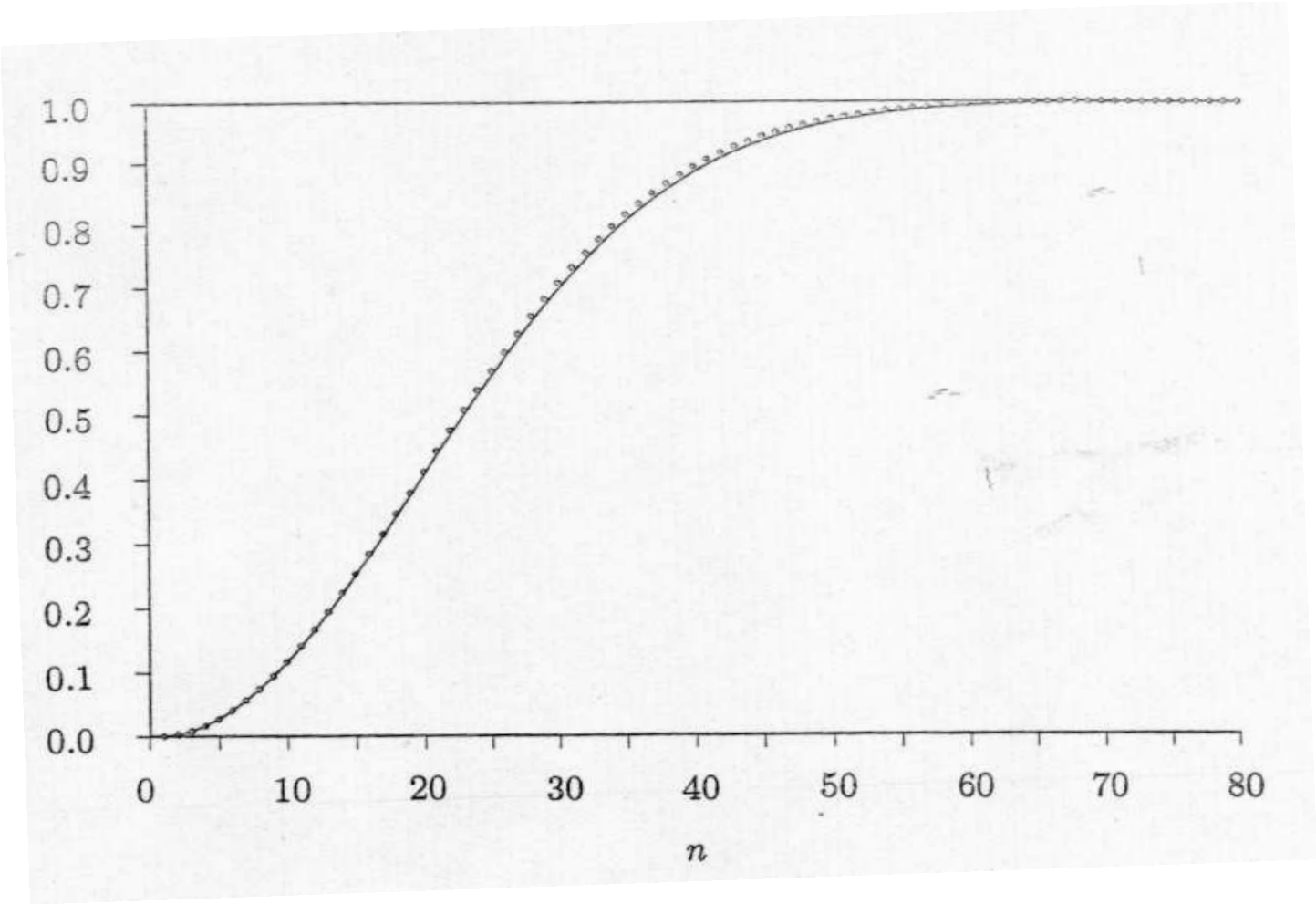
$$\boxed{a_n \equiv (1 - x/n)^n \to e^{-x}}$$

$$\begin{aligned} \ln a_n = -n \ln(1 - x/n) &= -n \left(\frac{x}{n} + \frac{x^2}{2n^2} + \frac{x^3}{3n^3} + \ldots \right) \\ &\to -x \\ \Rightarrow a_n &\to e^{-x}. \end{aligned}$$

- Similarly, if N is fixed, and $K \to \infty$, (exercise)

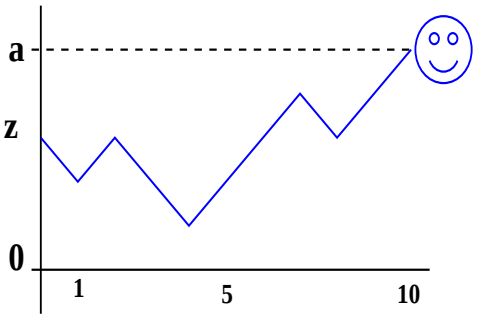
$$\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \ldots \left(1 - \frac{K-1}{N}\right) \to \exp\{-K(K-1)/2N\}$$

Birthday probabilities



Gambler's ruin

- Gambler starts with stake of z
- She goes home if she reaches a or 0
- Tosses a biased coin,
  $P\{H\} = p,\ P\{T\} = q = 1 - p$
  $H \Rightarrow$ increase by 1, $T \Rightarrow$ decrease by 1.



- $p_z = P\{\text{Reaches } a \text{ before } 0\},\ \ q_z = P\{\text{Reaches } 0 \text{ before } a\}$
- Will <u>prove</u> that $q_z = 1 - p_z$

$$
\begin{aligned}
q_z &= P\{\text{Reaching } 0 \text{ before } a\} \\
&= P\{0 \text{ before } a \big| H \text{ first}\} P\{H \text{ first}\} \\
&\qquad\qquad + P\{0 \text{ before } a \big| T \text{ first}\} P\{T \text{ first}\} \\
&= p q_{z+1} \ + \ q q_{z-1}
\end{aligned}
$$

$$\boxed{q_z \;=\; pq_{z+1} \;+\; qq_{z-1}}$$

- $q_0 = 1, \quad q_a = 0, \quad p_0 = 0, \quad p_a = 1$
- Solutions

$$p \neq q \quad \Rightarrow \quad q_z \;=\; \frac{(q/p)^a \;-\; (q/p)^z}{(q/p)^a \;-1}$$

$$p = q \quad \Rightarrow \quad q_z \;=\; 1 - \frac{z}{a}$$

- To check that $p_z = 1 - q_z$, solve the corresponding equation for p_z and sum
- See Feller for more details, and also gaming strategies

TABLE 1

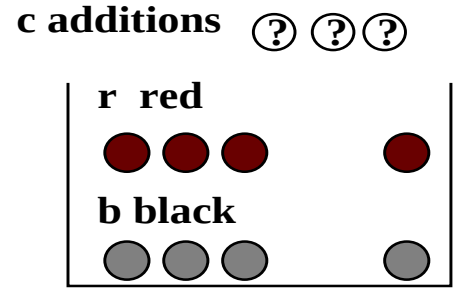
ILLUSTRATING THE CLASSICAL RUIN PROBLEM

p	q	z	a	Probability of		Expectation of	
				Ruin	Success	Gain	Duration
0.5	0.5	9	10	0.1	0.9	0	9
.5	.5	90	100	.1	.9	0	900
.5	.5	900	1,000	.1	.9	0	90,000
.5	.5	950	1,000	.05	.95	0	47,500
.5	.5	8,000	10,000	.2	.8	0	16,000,000
.45	.55	9	10	.210	.790	−1.1	11
.45	.55	90	100	.866	.134	−76.6	765.6
.45	.55	99	100	.182	.818	−17.2	171.8
.4	.6	90	100	.983	.017	−88.3	441.3
.4	.6	99	100	.333	.667	−32.3	161.7

The initial capital is z. The game terminates with ruin (loss z) or capital a (gain $a - z$).

Polya's urn

- b black and r red balls in urn
- One is chosen at random
- It is returned, along with
  c extra balls of the same colour
- Keep doing this: C_k = Colour of k-th ball drawn

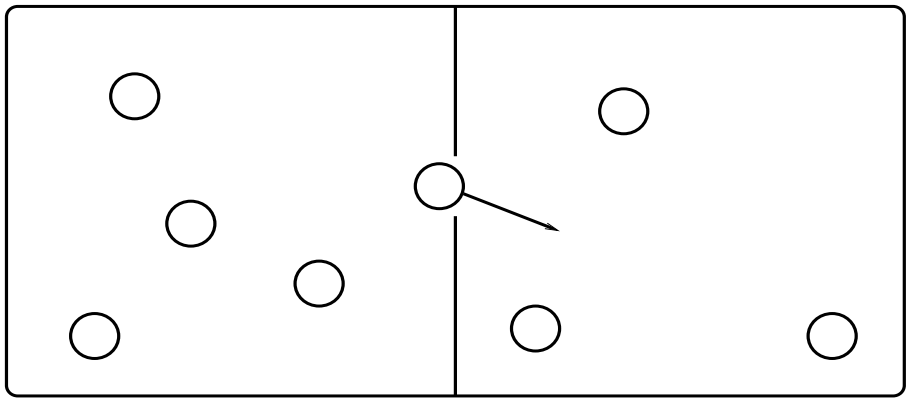


$$
\begin{aligned}
P\{C_1 = C_2 = B\} &= P\{C_2 = B \big| C_1 = B\} \times P\{C_1 = B\} \\
&= \frac{b}{b+r} \; \frac{b+c}{b+c+r}
\end{aligned}
$$

$$
\begin{aligned}
P\{C_2 = B\} &= P\{C_2 = B \big| C_1 = B\} \times P\{C_1 = B\} \\
&\qquad + P\{C_2 = B \big| C_1 = R\} \times P\{C_1 = R\} \\
&= \frac{b+c}{b+c+r} \; \frac{b}{b+r} + \frac{b}{b+c+r} \; \frac{r}{b+r} = \frac{b}{b+r}
\end{aligned}
$$

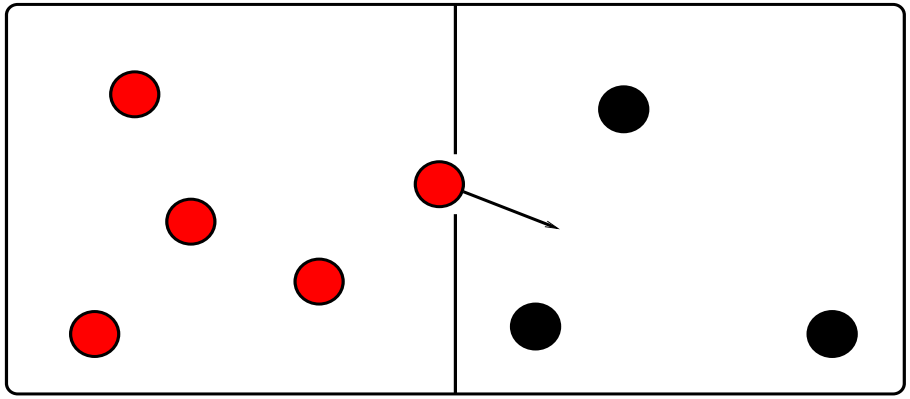
Heat exchange – Ehrenfest

- Heat generating particles move from one side of a container to another "at random":



Heat exchange – Ehrenfest

- Heat generating particles move from one side of a container to another "at random":



Further applications

- $c = 0$ is "sampling with replacement".
- $c = -1$ is "sampling without replacement".
- Many colours.
- Let X_n be the number of red balls chosen on the n-th draw. Then

$$\widehat{X}_n = \frac{X_n}{nc + r + b}$$

  is a martingale.

  i.e. Expectations (averages) don't change in time.
- Next problem: Find $P\{X_n = k\}$

Polya's urn after k steps By induction (conditioning on C_1 and using exactly the same argument) it is immediate that

$$P\{C_k = B\} \;=\; P\{C_1 = B\} \;=\; \frac{b}{b+r}$$

since

$$\begin{aligned} P\{C_k = B\} &= P\{C_k = B \big| C_1 = B\} \times P\{C_1 = B\} \\ &\qquad + P\{C_2 = B \big| C_1 = R\} \times P\{C_1 = R\} \\ &= \frac{b+c}{b+c+r}\,\frac{b}{b+r} + \frac{b}{b+c+r}\,\frac{r}{b+r} \\ &= \frac{b}{b+r} \end{aligned}$$

where the induction, in the second line, is for initial urns with $b+r+c$ balls, of which either $b+c$ or b are black and we look at stage $k-1$.

Another interesting property of the Polya urn:

$$P\{C_1 = \ldots = C_n = B,\ C_{n+1} = \ldots = C_{n+m} = R\}$$

$$= P\{C_1 = B\}P\{C_2 = B|C_1 = B\}P\{C_3 = B|C_1 = C_2 = B\}\ldots$$

$$\times P\{C_{n+m} = B|C_1 = \ldots = C_n = B, C_{n+1} = \ldots = C_{n+m-1} = R\}$$

$$= \frac{b(b+c)(b+2c)\cdots(b+nc-c)\cdot r(r+c)\cdots(r+mc-c)}{(b+r)(b+r+c)\cdots(b+r+(n+m)c)}$$

$$= \frac{r(r+c)\cdots(r+mc-c)\cdot b(b+c)(b+2c)\cdots(b+nc-c)}{(b+r)(b+r+c)\cdots(b+r+(n+m)c)}$$

$$= P\{C_1 = \ldots = C_m = R,\ C_{m+1} = \ldots = C_{n+m} = B\}$$

$$= P\{\text{ANY specific ordering of } n \text{ blacks and } m \text{ reds}\}$$