Foundations of data science

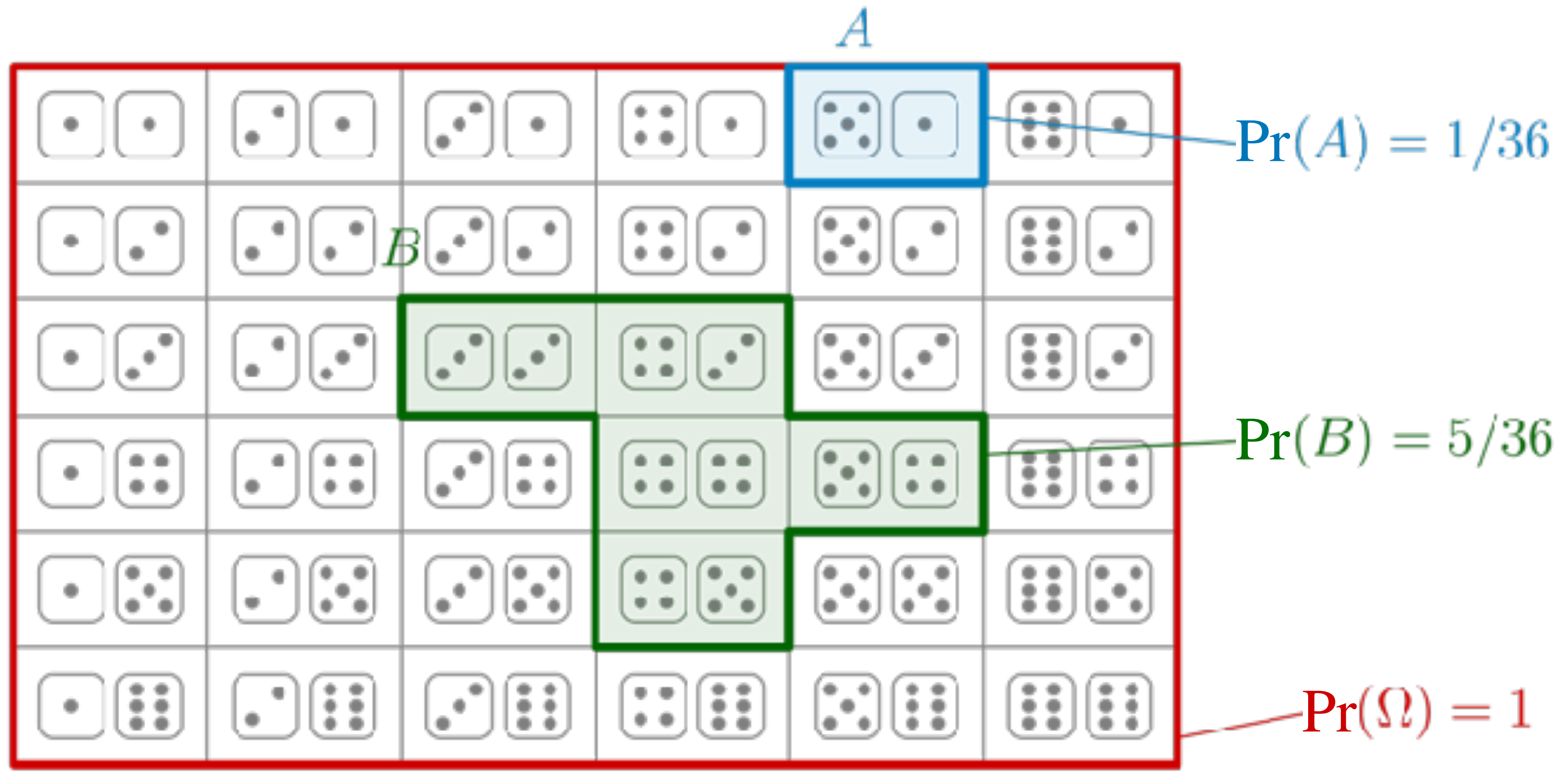
Probability Space

刘明谋 Nanjing University, Suzhou, 2025 Fall

Probability Space

Sample Space (样本空间)

- <u>Sample space</u> Ω: set of all possible outcomes of an experiment (<u>**samples**</u>).
  - **Example:** all sides of a dice; all outcomes of a sequence of coin tosses; …
- Each $\omega \in \Omega$ is called a <u>**sample**</u> (样本) or <u>**elementary event**</u> (基本事件).
- An <u>event</u> (事件) is a subset $A \subseteq \Omega$ of the sample space.



Discrete Probability Space

$(\Omega, \Pr)$

- Sample space Ω: set of all possible outcomes of an experiment (**samples**).
  - **Example:** all sides of a dice; all outcomes of a sequence of coin tosses; …
- Each $\omega \in \Omega$ is called a **sample** (样本) or **elementary event** (基本事件).
- For **discrete** probability space (where Ω is *finite* or *countably infinite*):
  - **probability mass function** (*pmf*) $p : \Omega \to [0,1]$ satisfies $\sum_{\omega \in \Omega} p(\omega) = 1$
  - the probability of event $A \subseteq \Omega$ is given by $\Pr(A) = \sum_{\omega \in A} p(\omega)$

Sample Space and Events

- <u>Sample space</u> Ω: set of all possible outcomes of an experiment (<u>samples</u>).
  - **Example**: all sides of a dice; all outcomes of a sequence of coin tosses; …
- A family $\Sigma \subseteq 2^{\Omega}$ of subsets of Ω, called <u>events</u> (事件), satisfies:
  - $\emptyset$ and Ω are events (the ***<u>impossible event</u>*** “不可能事件” and ***<u>certain event</u>*** “必然事件”);
  - if A is an event, then so is its complement $A^c = \Omega \backslash A$;
  - if (countably many) $A_1, A_2, \ldots$ are events, then so is $\bigcup_i A_i$ (and $\bigcap_i A_i$)

σ-Algebra (σ-代数)

- A family $\Sigma \subseteq 2^\Omega$ of subsets of Ω is called a **σ-algebra** or **σ-field**, if:
  - $\emptyset \in \Sigma$
  - $A \in \Sigma \implies A^c \in \Sigma$ (where $A^c = \Omega \backslash A$ denotes A's compliment in Ω)
  - $A_1, A_2, \ldots \in \Sigma \implies \bigcup_i A_i \in \Sigma$ (for countably many $A_1, A_2, \ldots \in \Sigma$)
- Examples:
  - $\Sigma = 2^\Omega$
  - $\Sigma = \{\emptyset, \Omega\}$
  - $\Sigma = \{\emptyset, A, A^c, \Omega\}$ for any $A \subseteq \Omega$

Sets as Events

Notation	Set interpretation	Event interpretation
$\omega \in \Omega$	Member of Ω	Elementary event
$A \subseteq \Omega$	Subset of Ω	Event A occurs
A^c	Complement of A	Event A does not occur
$A \cap B$	Intersection	Both A and B
$A \cup B$	Union	Either A or B or both
$A \backslash B$	Difference	A, but not B
$A \oplus B$	Symmetric difference	Either A or B, but not both
$\varnothing$	Empty set	Impossible event
Ω	Whole space	Certain event
$A \subseteq B$	Inclusion	A implies B
$A \cap B = \varnothing$	Set disjointness	A and B cannot both occur

Probability Space and Measure

Andrey Kolmogorov
Андре́й Колмого́ров
(1903-1987)

$(\Omega, \Sigma, \Pr)$

- Let $\Sigma \subseteq 2^{\Omega}$ be a <u>**σ-algebra**</u>.
- A <u>**probability measure**</u> (概率测度), also called <u>**probability law**</u> (概率律), is a function $\Pr : \Sigma \to [0,1]$ satisfying:
  - (*unitary/normalized*) $\Pr(\Omega) = 1$;
  - (*σ-additive*) for **disjoint (不相容)** $A_1, A_2, \ldots \in \Sigma$: $\Pr\left(\bigcup_i A_i\right) = \sum_i \Pr(A_i)$.
- The triple $(\Omega, \Sigma, \Pr)$ is called a <u>**probability space**</u>.

Classical Examples of Probability Space

- 古典概型 (classic probability): ***discrete uniform probability law***

  Finite sample space Ω, each outcome $\omega \in \Omega$ has equal probability.

  For every event $A \subseteq \Omega$: $\Pr(A) = \dfrac{|A|}{|\Omega|}$

- 几何概型 (geometric probability): continuous probability space such that

  For every event $A \in \Sigma$: $\Pr(A) = \dfrac{\text{Vol}(A)}{\text{Vol}(\Omega)}$

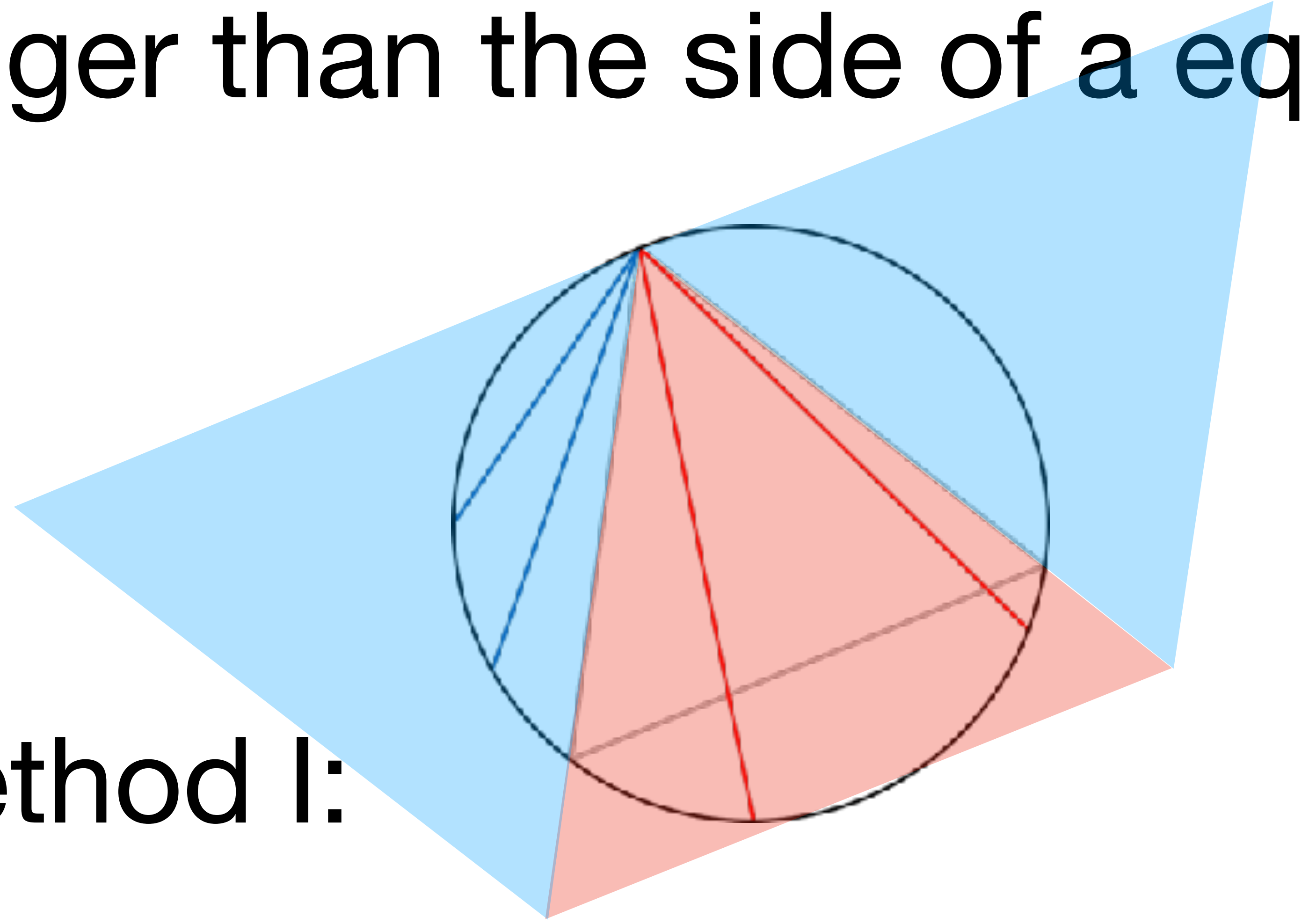
  - Bertrand's paradox
  - Buffon's needle problem

$\Pr \propto \angle$

Bertrand Paradox (贝特朗悖论)

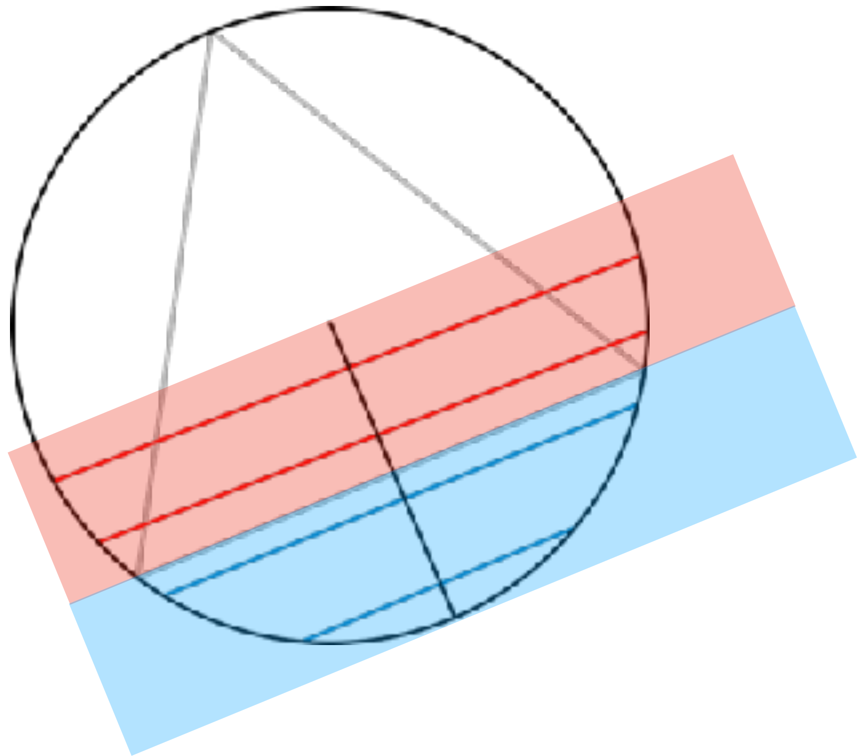
introduced in *Calcul des probabilités* (1889) by Joseph Bertrand

- What is the probability of the event A that a ***random chord of a circle*** is longer than the side of a equilateral triangle inscribed in a circle?
  - method I:



$$\Pr(A) = \frac{60^\circ}{180^\circ} = \frac{1}{3}$$

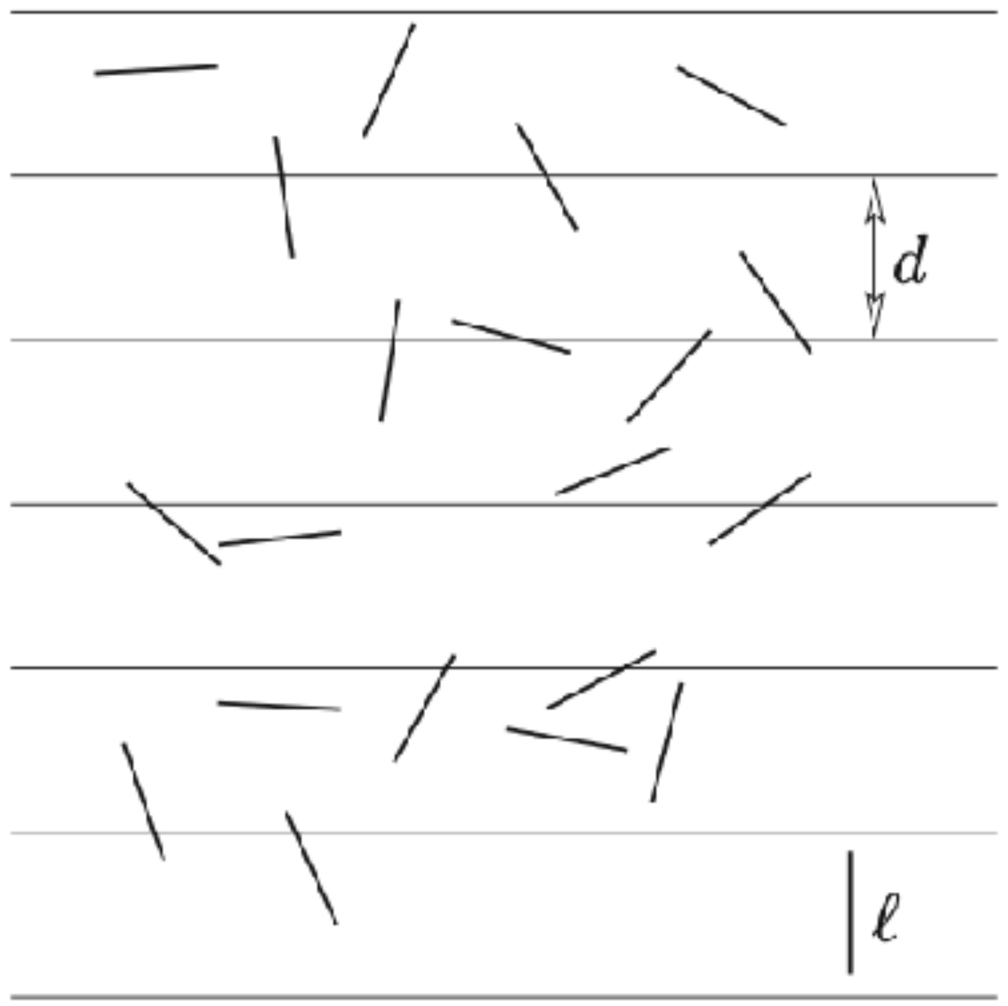
  - method II:



$$\Pr'(A) = \frac{\frac{r}{2}}{r} = \frac{1}{2}$$

Buffon's Needle Problem (蒲丰投针问题)
(Georges-Louis Leclerc de Buffon in 1733, and in 1777)

- Suppose that you drop a short needle of length ℓ on ruled paper, with distance d between parallel lines.
- What is the probability that the needle comes to lie in a position where it crosses one of the lines?
- For $\ell < d$, this probability is calculated as:



$$\Pr(A) = \frac{\text{Vol}(A)}{\text{Vol}(\Omega)} = \frac{2}{d\pi}\int_0^\pi \frac{\ell}{2}\sin(x)\,\mathrm{d}x = \frac{2\ell}{d\pi}$$

$x \in [0,\pi]$: angle between the needle and the parallel line below it

$y \in [0,d/2]$: distance from the center of the needle to the closest parallel line

Event $A = \left\{(x,y) \in [0,\pi] \times \left[0,\frac{d}{2}\right] \mid y \le \frac{\ell}{2}\sin(x)\right\}$

- A ***Monte Carlo method*** for estimating π

Basic Properties of Probability

All followings can be deduced from the **axioms** of probability space:

- $\Pr(A^c) = 1 - \Pr(A)$
- $\Pr(\varnothing) = 0$ $\qquad \Pr(A) > 0 \implies A \neq \varnothing$ **(the probabilistic method)**
- $\Pr(A \setminus B) = \Pr(A) - \Pr(A \cap B)$
- $A \subseteq B \implies \Pr(A) \leq \Pr(B)$
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- ***Not even wrong***: "自然数是偶数的概率为1/2"
  （然而 "[0,1]中均匀实数是有理数的概率为0" 却是正确的）

Union Bound

- **Union bound** (Boole's inequality): for events $A_1, A_2, \ldots A_n \in \Sigma$

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr(A_i)$$

- **Example:** A *randomized algorithm* has n types of errors, each occurring with prob. $\leq p$

Let A_i be the event that type-i error occurs.

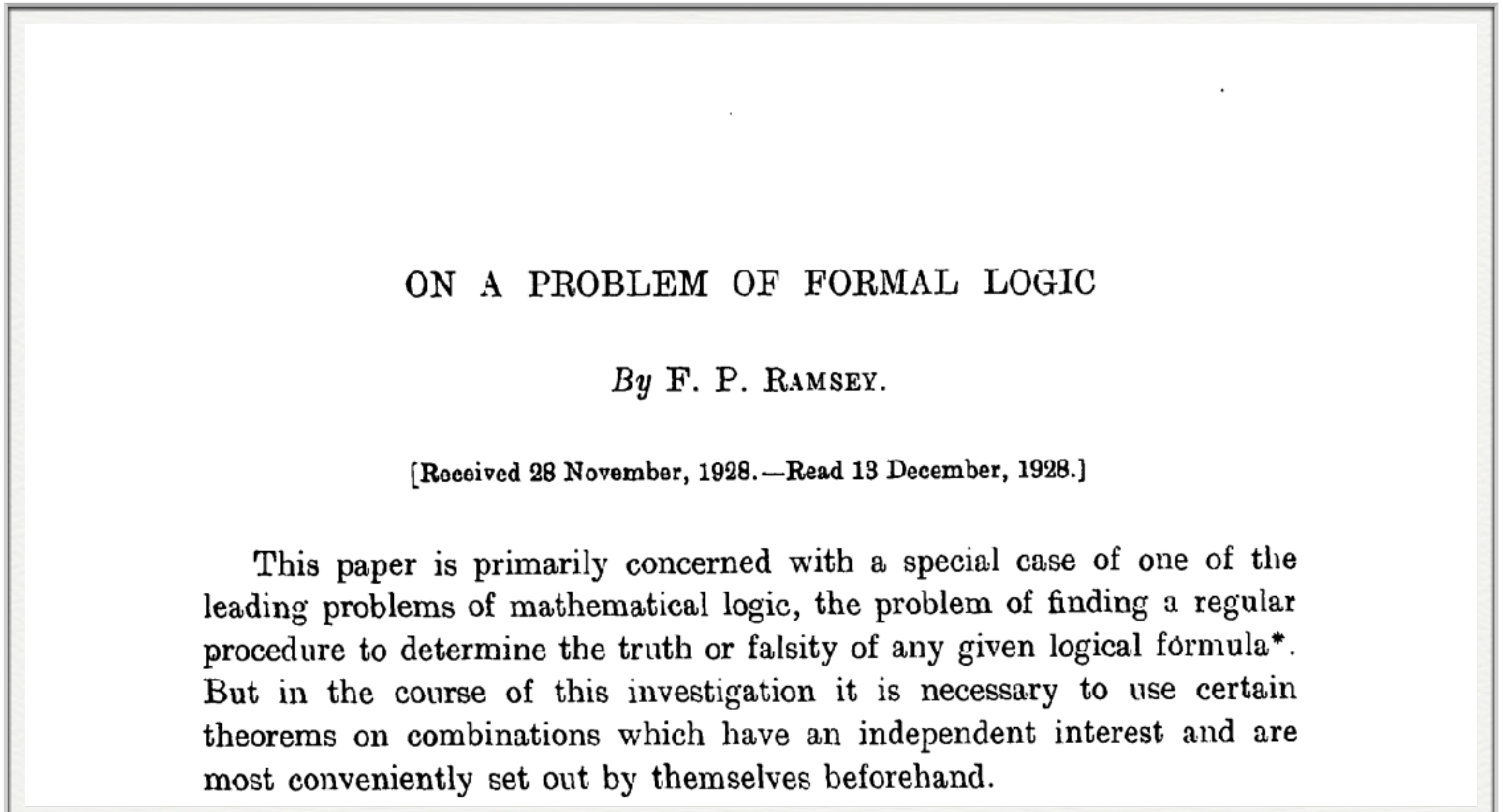
$$\Pr[\text{ no error occurs }] = \Pr\left(\bigcap_{i=1}^{n} A_i^c\right) = 1 - \Pr\left(\bigcup_{i=1}^{n} A_i\right) \geq 1 - np$$

Holds unconditionally.
(tight if all bad events are disjoint)

Ramsey Theory (Frank Ramsey, 1928)

“In any party of six people, either at least three of them are mutual strangers or at least three of them are mutual acquaintances.”

Any 2-coloring of K_6 must contain a *monochromatic* K_3.



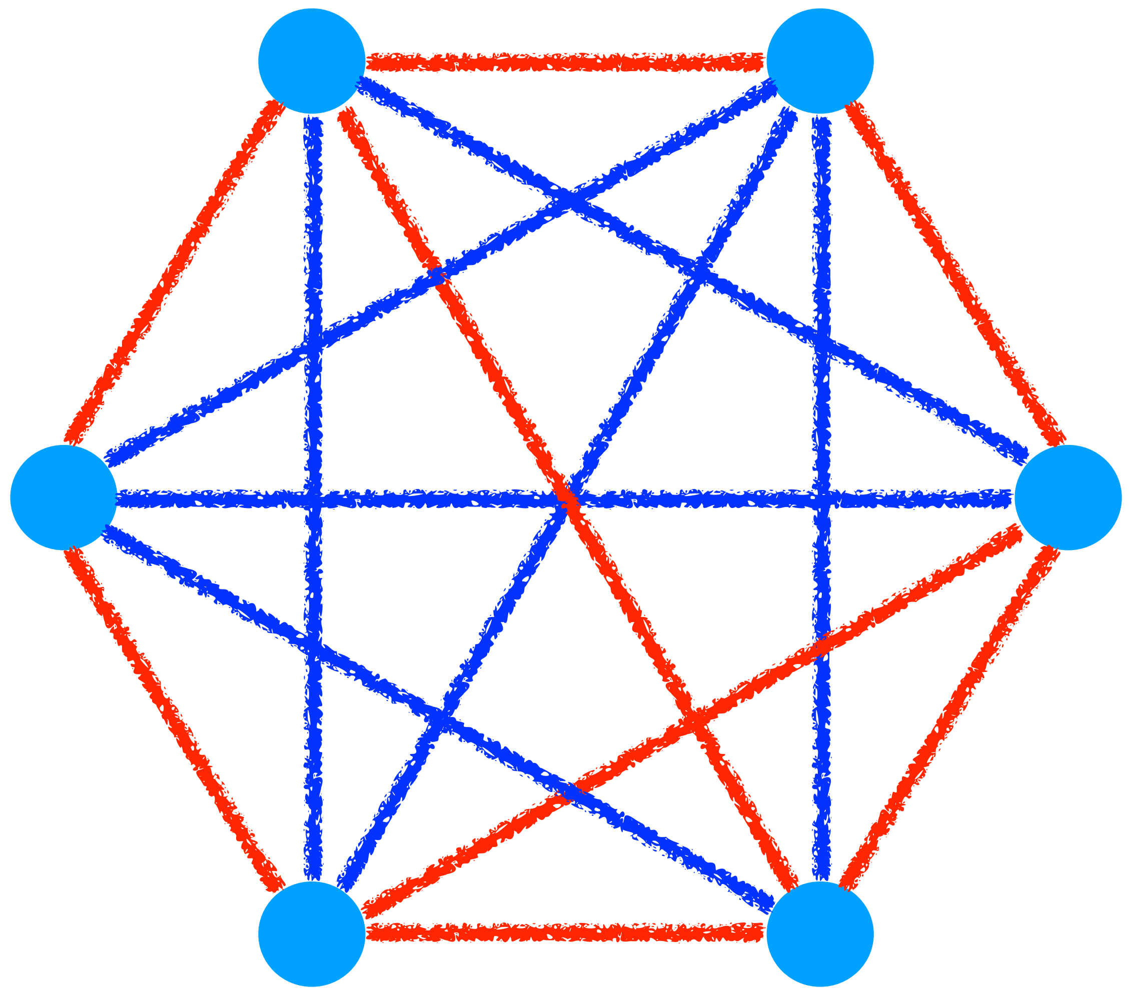
ON A PROBLEM OF FORMAL LOGIC

By F. P. RAMSEY.

[Received 28 November, 1928.—Read 13 December, 1928.]

This paper is primarily concerned with a special case of one of the leading problems of mathematical logic, the problem of finding a regular procedure to determine the truth or falsity of any given logical formula*. But in the course of this investigation it is necessary to use certain theorems on combinations which have an independent interest and are most conveniently set out by themselves beforehand.

Frank Ramsey
(1903-1930)



Ramsey Theory (Frank Ramsey, 1928)

"In any party of six people, either at least three of them are mutual strangers or at least three of them are mutual acquaintances."

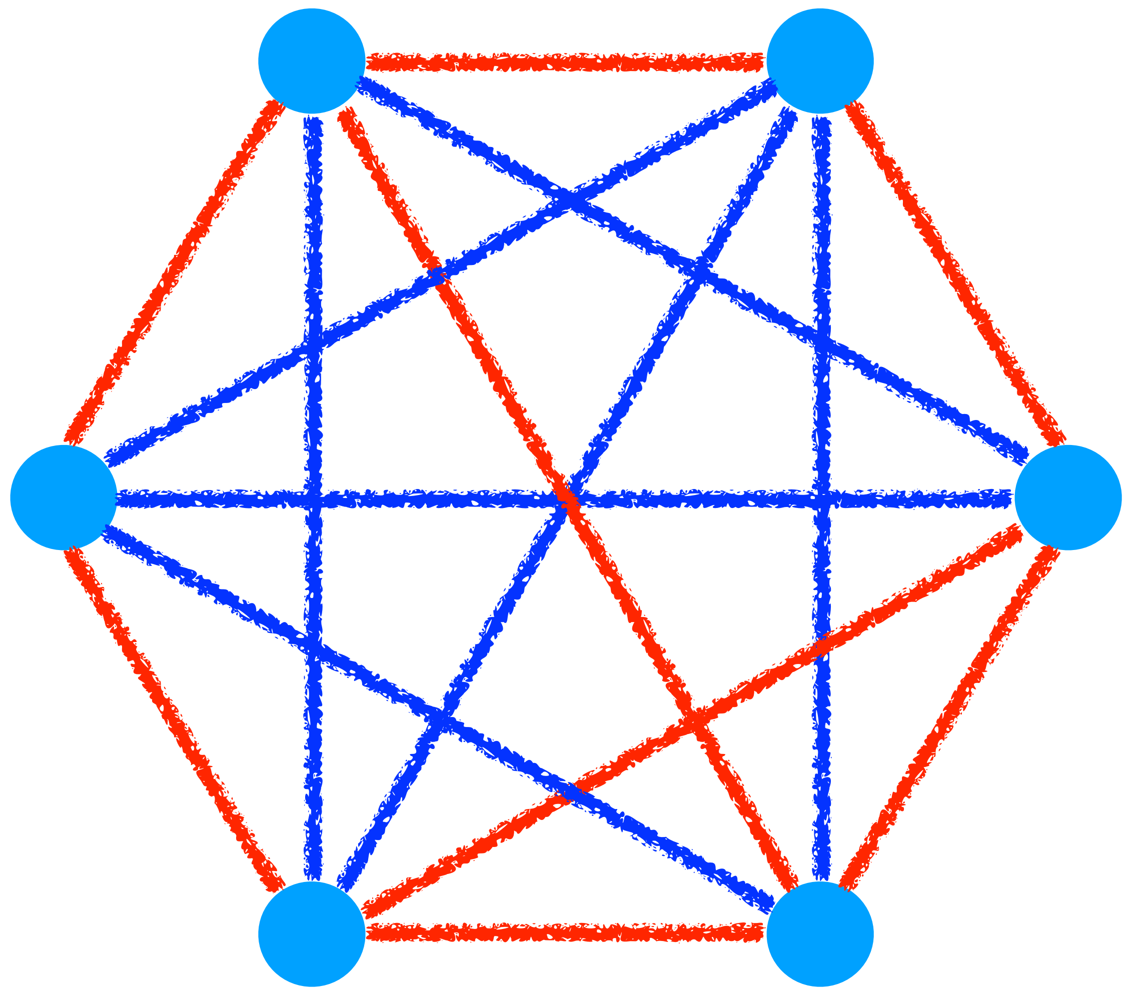
(★) Any 2-coloring of K_n must contain a ***monochromatic*** K_k.

- Ramsey number $R(k,k)$:= the smallest n satisfy (★)
- $R(3,3) = 6$

Ramsey Theorem (1928):

$R(k,k)$ is finite for all $k > 0$

Frank Ramsey
(1903-1930)



Ramsay Number

- The exact values of Ramsey numbers $R(k, k)$ are notoriously hard to compute.

Values / known bounding ranges for Ramsey numbers $R(s, t)$ (sequence A212954 in the OEIS)

t \ s	1	2	3	4	5	6	7	8	9	10
1	**1**	1	1	1	1	1	1	1	1	1
2		**2**	3	4	5	6	7	8	9	10
3			**6**	9	14	18	23	28	36	40–41[14]
4				**18**	25[9]	36–40	49–58	59[15]–79	73–105	92–135
5					**43–46**[11]	59[16]–85	80–133	101–193	133–282	149[15]–381
6						**102–160**	115[15]–270	134[15]–423	183–651	204–944
7							**205–492**	219–832	252–1368	292–2119
8								**282–1518**	329–2662	343–4402
9									**565–4956**	581–8675
10										**798–16064**

The Probabilistic Method
(for Ramsey number lower bound)

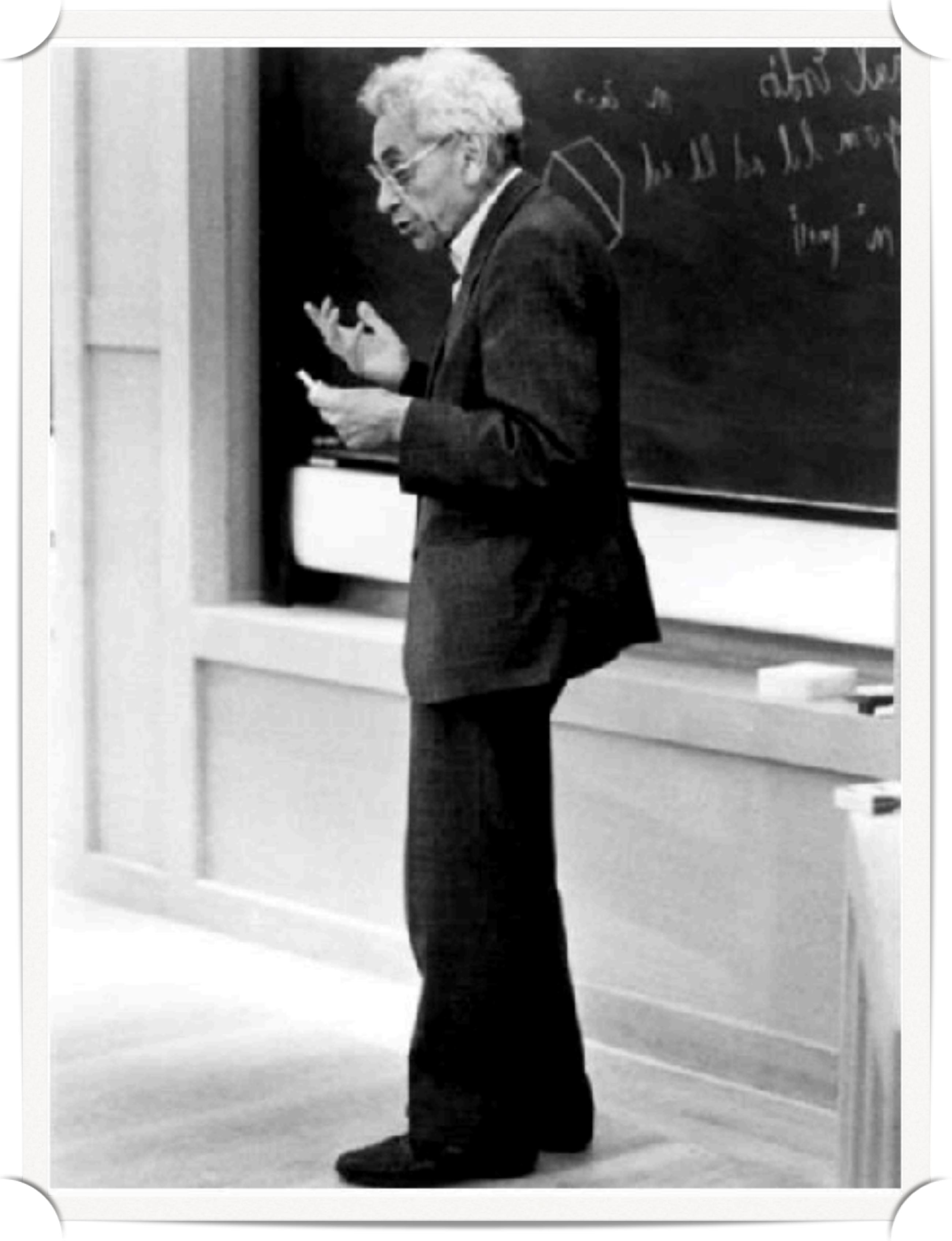
Theorem (Erdős 1947):

If $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, then ∃ 2-coloring of K_n with no monochromatic K_k subgraph.

- $\Longrightarrow R(k,k) >$ any n satisfying

$$\binom{n}{k}2^{1-\binom{k}{2}} < 1$$

- In particular, $R(k,k) > k2^{k/2-2}$



Paul Erdős
Erdős Pál
(1913-1996)

The Probabilistic Method
(for Ramsey number lower bound)

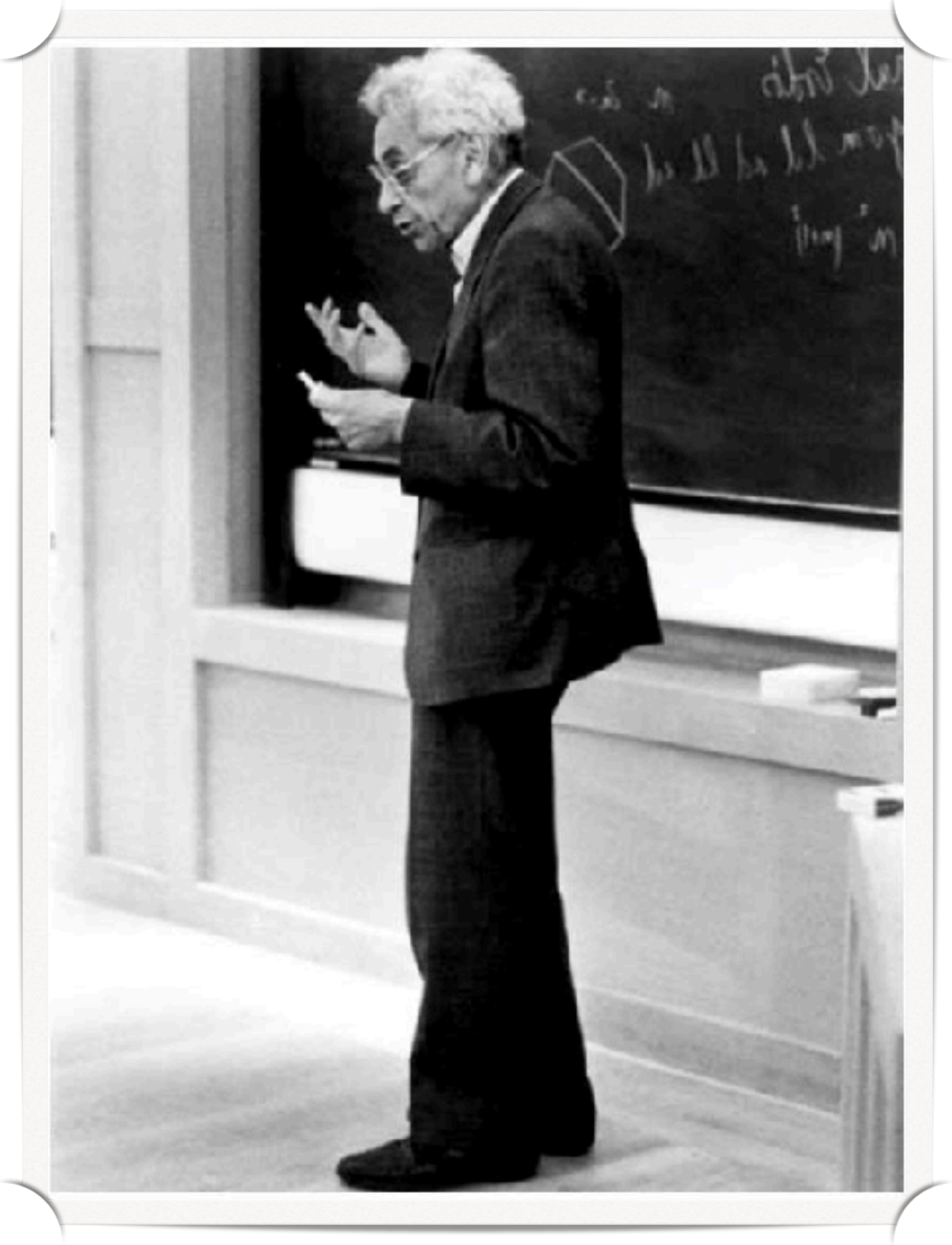
Theorem (Erdős 1947):

If $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, then $\exists$ 2-coloring of K_n with no monochromatic K_k subgraph.

Idea of the proof: Construct a probability law $\Pr$ on the sample space $\Omega = \{\textit{all 2-colorings of } K_n\}$

Show that $\Pr(A) > 0$ for the $A \subset \Omega$ defined as
$A = \{\textit{2-colorings of } K_n \textit{ with no monochromatic } K_k\}$

$\Longrightarrow A \neq \emptyset$, i.e. $\exists$ such 2-coloring of K_n w/o mono-K_k



Paul Erdős
Erdős Pál
(1913-1996)

The Probabilistic Method

(for Ramsey number lower bound)

Theorem (Erdős 1947):

If $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, then ∃ 2-coloring of K_n with no monochromatic K_k subgraph.

Proof: Color each edge of K_n red or blue uniformly at random.

- For each subset S of k vertices, define A_S={K_S is monochromatic}, then $\Pr(A_S) = 2^{1-\binom{k}{2}}$
- By **union bound**: $\Pr\left(\bigcup_S A_S\right) \leq \binom{n}{k}2^{1-\binom{k}{2}}$, which is < 1 by assumption.
- $\Pr[$no monochromatic $K_k] = 1 - \Pr[\exists$monochromatic $K_k] = 1 - \Pr\left(\bigcup_S A_S\right) > 0$

The Probabilistic Method: There exists such a non-Ramsey 2-coloring!

Principles of Inclusion-Exclusion

- **Principle of inclusion-exclusion**: for events $A_1, A_2, \ldots A_n \in \Sigma$,

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i) - \sum_{i<j} \Pr(A_i \cap A_j) + \sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k) - \cdots$$

$$= \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ S \neq \emptyset}} (-1)^{|S|-1} \Pr\left(\bigcap_{i \in S} A_i\right)$$

- **Boole-Bonferroni Inequality**: for events $A_1, A_2, \ldots A_n \in \Sigma$, for any $k \geq 0$

$$\sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ 1 \leq |S| \leq 2k}} (-1)^{|S|-1} \Pr\left(\bigcap_{i \in S} A_i\right) \leq \Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ 1 \leq |S| \leq 2k+1}} (-1)^{|S|-1} \Pr\left(\bigcap_{i \in S} A_i\right)$$

Derangement (错排)

(*le problème des rencontres, 1708*)

- The probability that a random permutation $\pi : [n] \xrightarrow[\text{onto}]{\text{1-1}} [n]$ has no fixed point (i.e. there is no $i \in [n]$ such that $\pi(i) = i$).
- Let A_i be the event that $\pi(i) = i$.

$$\Pr\left(\bigcap_{i \in S} A_i\right) = \frac{(n - |S|)!}{n!}$$

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} \sum_{S \in \binom{\{1,2,\ldots,n\}}{k}} (-1)^{k-1} \Pr\left(\bigcap_{i \in S} A_i\right) = \sum_{k=1}^{n} \binom{n}{k} (-1)^{k-1} \frac{(n-k)!}{n!} = -\sum_{k=1}^{n} \frac{(-1)^k}{k!}$$

$$\Pr[\,\pi \text{ has no fixed point}\,] = \Pr\left(\bigcap_{i=1}^{n} A_i^c\right) = 1 - \Pr\left(\bigcup_{i=1}^{n} A_i\right) = 1 + \sum_{k=1}^{n} \frac{(-1)^k}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \to \frac{1}{e} \text{ as } n \to \infty$$

Continuity of Probability Measures*

- Let $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ be an increasing sequence of events, and write A for their limit

$$A = \bigcup_{i=1}^{\infty} A_i = \lim_{i \to \infty} A_i \,.$$

Then $\Pr(A) = \lim_{i \to \infty} \Pr(A_i)$.

- **Proof**: Express A as a disjoint union $A = A_1 \uplus (A_2 \backslash A_1) \uplus (A_3 \backslash A_2) \uplus \cdots$. Then

$$\begin{aligned}
\Pr(A) &= \Pr(A_1) + \sum_{i=1}^{\infty} \Pr(A_{i+1} \backslash A_i) \\
&= \Pr(A_1) + \lim_{n \to \infty} \sum_{i=1}^{n-1} [\Pr(A_{i+1}) - \Pr(A_i)] \\
&= \lim_{n \to \infty} \Pr(A_n)
\end{aligned}$$

Continuity of Probability Measures*

- Let $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ be an increasing sequence of events, and write A for their limit

$$A = \bigcup_{i=1}^{\infty} A_i = \lim_{i \to \infty} A_i \,.$$

  Then $\Pr(A) = \lim_{i \to \infty} \Pr(A_i)$.

- Let $B_1 \supseteq B_2 \supseteq B_3 \supseteq \ldots$ be an decreasing sequence of events, and write B for their limit

$$B = \bigcap_{i=1}^{\infty} B_i = \lim_{i \to \infty} B_i \,.$$

  Then $\Pr(B) = \lim_{i \to \infty} \Pr(B_i)$.

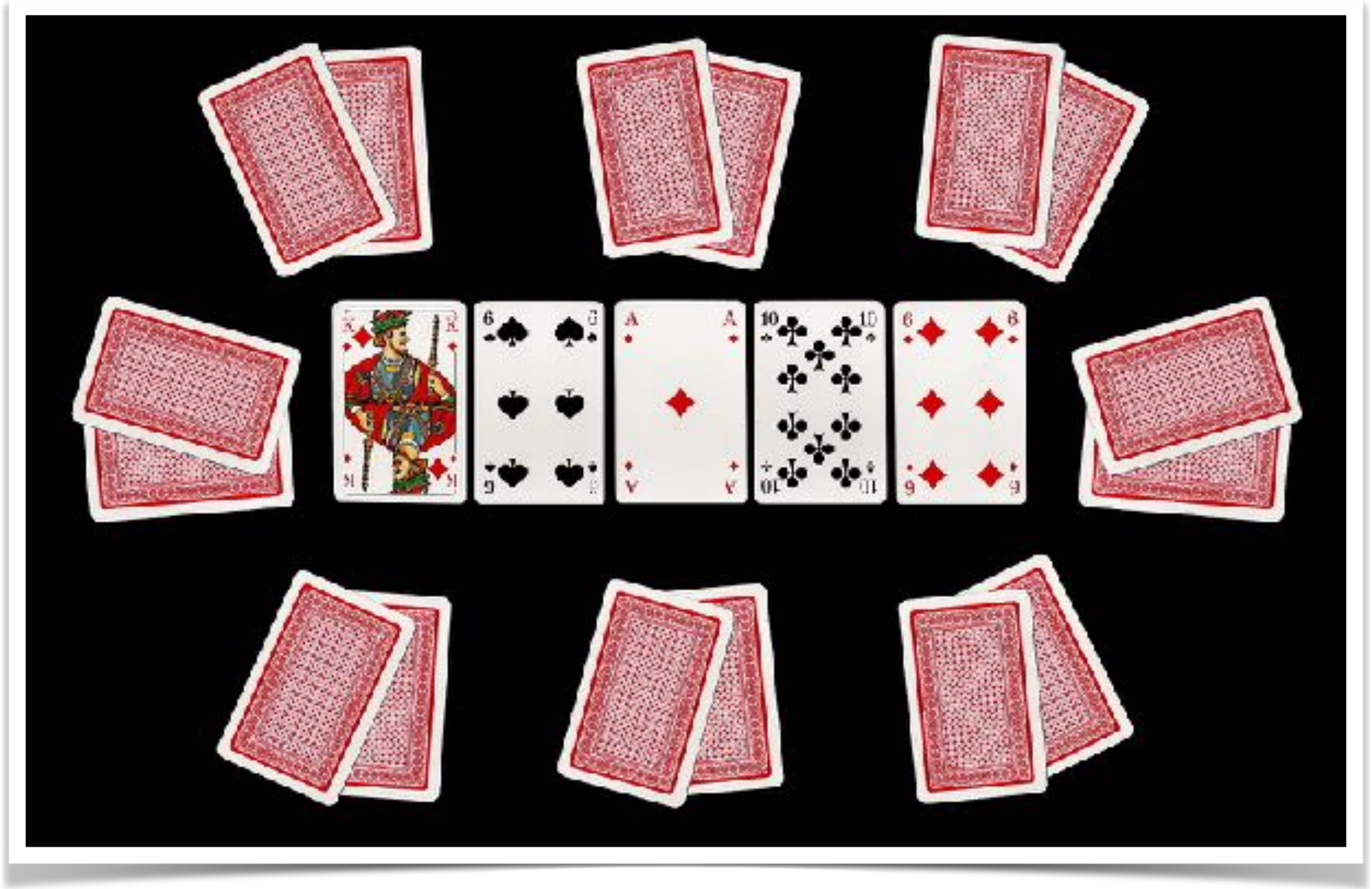
- **Proof**: Consider the complements $B_1^c \subseteq B_2^c \subseteq B_3^c \subseteq \ldots$ which is an increasing sequence.

Null and Almost Surely Events*

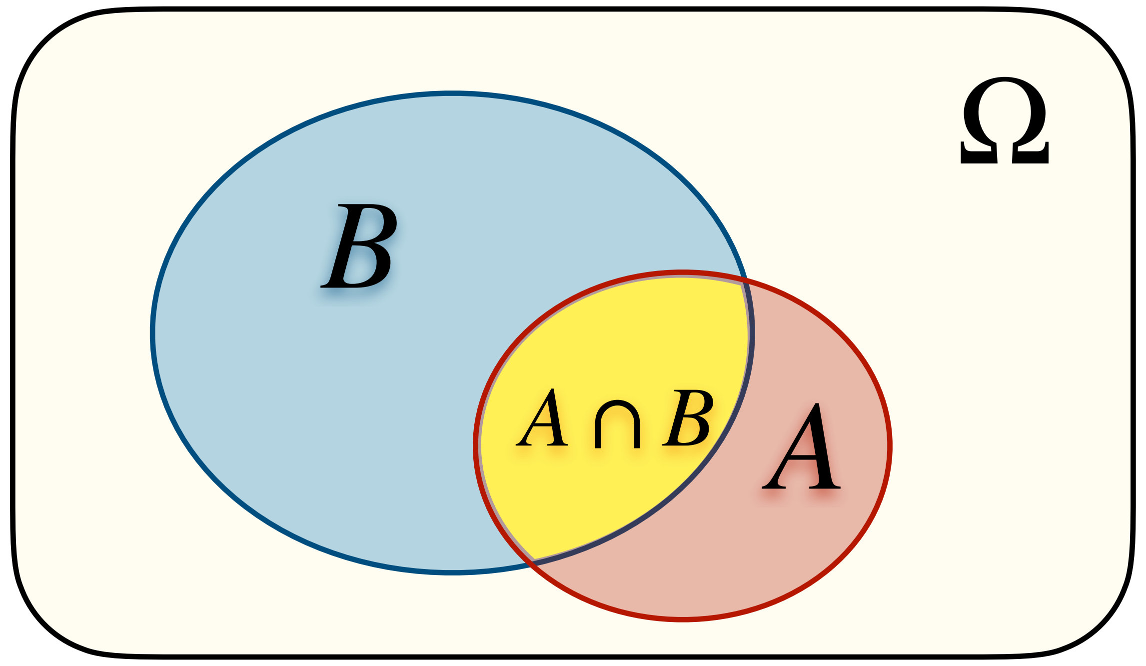
- An event $A \in \Sigma$ is called <u>**null**</u> if $\Pr(A) = 0$.
  - A null event is not necessarily the <u>impossible</u> event $\emptyset$.
- An event $A \in \Sigma$ occurs <u>**almost surely**</u> (<u>**a.s.**</u>) if $\Pr(A) = 1$.
  - An event that occurs a.s., is not necessarily the <u>certain</u> event Ω.
- A probability space is called <u>**complete**</u>, if all subsets of null events are events.
  - Without loss of generality: we only consider complete probability spaces

    (if we start with an incomplete one, we can complete it without changing the probabilities)

Conditional Probability



Conditional Probability



- Frequently, we need to make such statement:

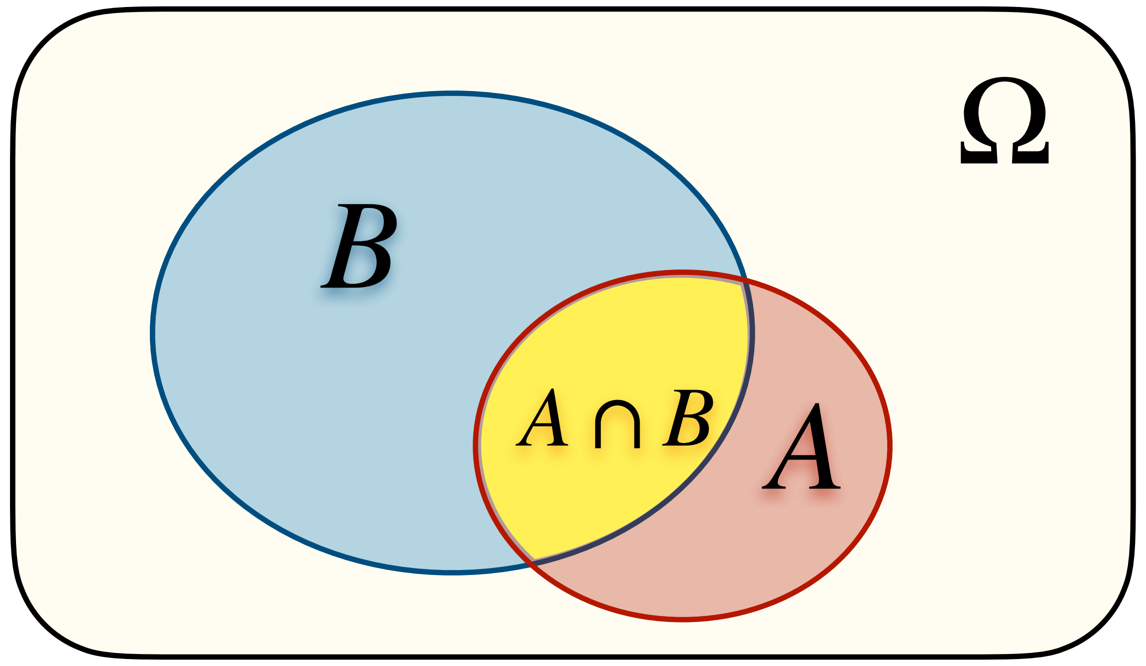
  "*The probability of A is p, given that B occurs.*"

- For discrete uniform law: $p = \dfrac{|A \cap B|}{|B|} = \dfrac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \dfrac{\Pr(A \cap B)}{\Pr(B)}$

- Let A be an event, and let B be an event that $\Pr(B) > 0$.
  The conditional probability that A **occurs given that** B **occurs** is defined to be

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Conditional Probability



- Let A be an event, and let B be an event that $\Pr(B) > 0.$
  The <u>conditional probability</u> that A **occurs given that** B **occurs** is defined to be

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

- $\Pr(\cdot \mid B)$ is a well-defined probability law:
  - sample space is B
  - $\Sigma^B = \{A \cap B \mid A \in \Sigma\}$ is a σ-algebra
  - the law $\Pr(\cdot \mid B)$ satisfies the probability axioms

Fair Coins out of a Biased One

(von Neumann's Bernoulli factory)

- John von Neumann (1951): "Suppose you are given a coin for which the probability of **HEADS**, say p, is **unknown**. How can you use this coin to generate unbiased (fair) coin-flips."

- **Protocol:** Repetitively flip the coin until a $\mathtt{HT}$ or $\mathtt{TH}$ is encountered, output $\mathtt{H}$ if $\mathtt{HT}$ is encountered, and output $\mathtt{T}$ if otherwise.

- Consider any two consecutive coin flips:

$$\Pr(\mathtt{HT} \mid \{\mathtt{HT}, \mathtt{TH}\}) = \Pr(\mathtt{TH} \mid \{\mathtt{HT}, \mathtt{TH}\}) = \frac{p(1-p)}{2p(1-p)} = \frac{1}{2}$$

The Two Child Problem
(boy or girl paradox)

- Martin Gardner (1959): "Knowing that I have two children and at least one of them is girl, what is the probability that both children are girls?"

- Consider a uniform law $\Pr$ over $\Omega = \{\mathtt{BB}, \mathtt{BG}, \mathtt{GB}, \mathtt{GG}\}$

$$\Pr(\{\mathtt{GG}\} \mid \{\mathtt{BG}, \mathtt{GB}, \mathtt{GG}\}) = \frac{\Pr(\{\mathtt{GG}\})}{\Pr(\{\mathtt{BG}, \mathtt{GB}, \mathtt{GG}\})}$$

$$= \frac{1/4}{3/4} = \frac{1}{3}$$

Laws for Conditional Probability

- **Chain rule**:

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \Pr\left(A_i \mid \bigcap_{j<i} A_j\right)$$

- **Law of total probability**: For partition $B_1, B_2, \ldots, B_n$ of Ω,

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \cap B_i) = \sum_{i=1}^{n} \Pr(A \mid B_i) \Pr(B_i)$$

- **Bayes' law**: For partition $B_1, B_2, \ldots, B_n$ of Ω,

$$\Pr(B_i \mid A) = \frac{\Pr(B_i) \Pr(A \mid B_i)}{\Pr(A)} = \frac{\Pr(B_i) \Pr(A \mid B_i)}{\Pr(A \mid B_1) \Pr(B_1) + \cdots + \Pr(A \mid B_n) \Pr(B_n)}$$

Chain Rule

(General Product Rule / Law of Successive Conditioning)

- Assuming that all the involved conditions have positive probabilities, we have

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \Pr\left(A_i \mid \bigcap_{j<i} A_j\right)$$

- **Proof:** Due to the telescopic product

$$\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \frac{\Pr\left(\bigcap_{i=1}^{n} A_i\right)}{\Pr\left(\bigcap_{i=1}^{n-1} A_i\right)} \cdot \frac{\Pr\left(\bigcap_{i=1}^{n-1} A_i\right)}{\Pr\left(\bigcap_{i=1}^{n-2} A_i\right)} \cdots \frac{\Pr\left(A_1 \cap A_2\right)}{\Pr\left(A_1\right)} \cdot \Pr(A_1)$$

Birthday "Paradox"

"一个班级若想要100%地保证有两个人同一天过生日，需要班上有超过366人；
但若仅想让这件事发生的可能性超过99%，则班上有超过57人就足够了。"

- Consider uniform random mapping $f : [n] \to [m]$

$$\Pr[\ f \text{ is 1-1 }] = \frac{m!/(m-n)!}{m^n} = \prod_{i=1}^{n}\left(1 - \frac{i-1}{m}\right)$$

- <u>Balls-into-bins</u> model: throwing n balls into m bins one-by-one at random

$$\Pr[\text{every ball is thrown to an empty bin}] \quad = \epsilon \text{ for } n \approx \sqrt{2m\ln(1/\epsilon)}$$

$$= \prod_{i=1}^{n} \Pr[\text{ball } i \text{ is in thrown into an empty bin} \mid \text{every ball } j < i \text{ is in an empty bin}] = \prod_{i=1}^{n}\left(1 - \frac{i-1}{m}\right)$$

$$\approx \exp\left(-\sum_{i=1}^{n}\frac{i-1}{m}\right) \approx \exp\left(-\frac{n^2}{2m}\right)$$

Law of Total Probability

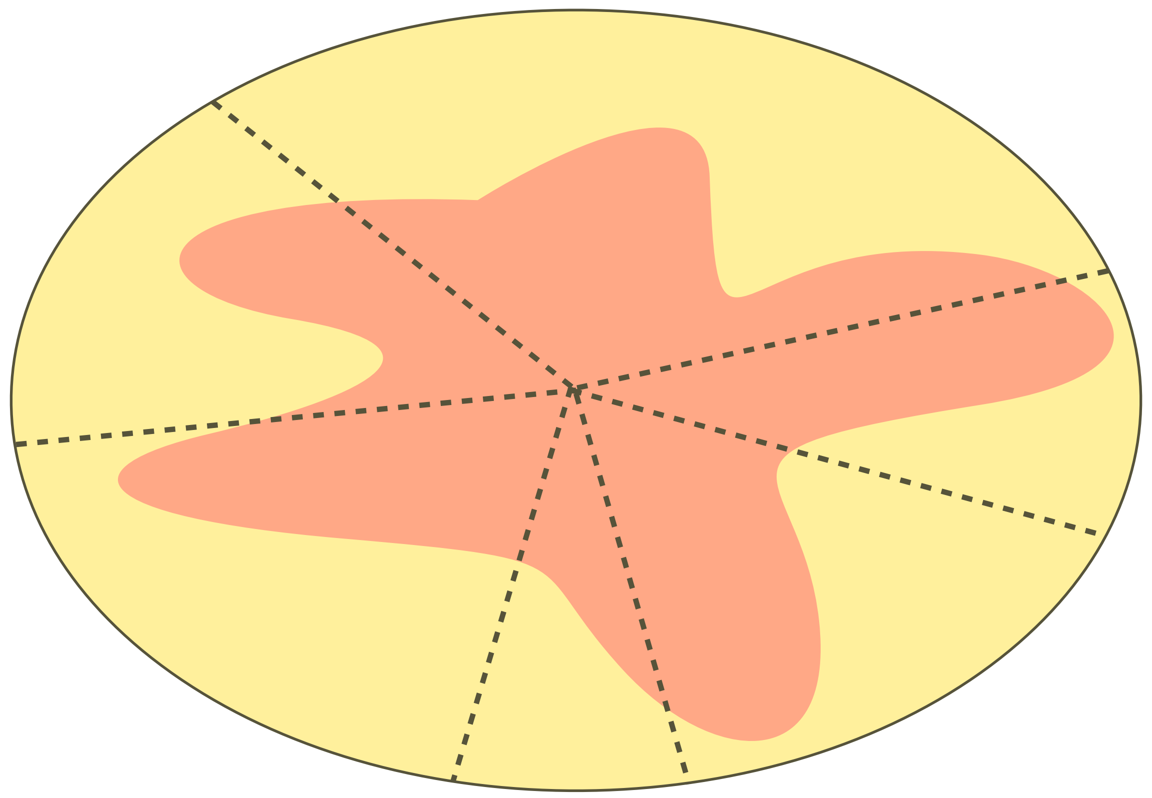
- Let events $B_1, B_2, \ldots, B_n$ be a partition of Ω such that $\Pr(B_i) > 0$ for all i. Then:

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \cap B_i) = \sum_{i=1}^{n} \Pr(A \mid B_i) \Pr(B_i)$$

- **Proof**: $A \cap B_1, A \cap B_2, \ldots, A \cap B_n$ are disjoint and $A = \bigcup_{i=1}^{n} (A \cap B_i)$

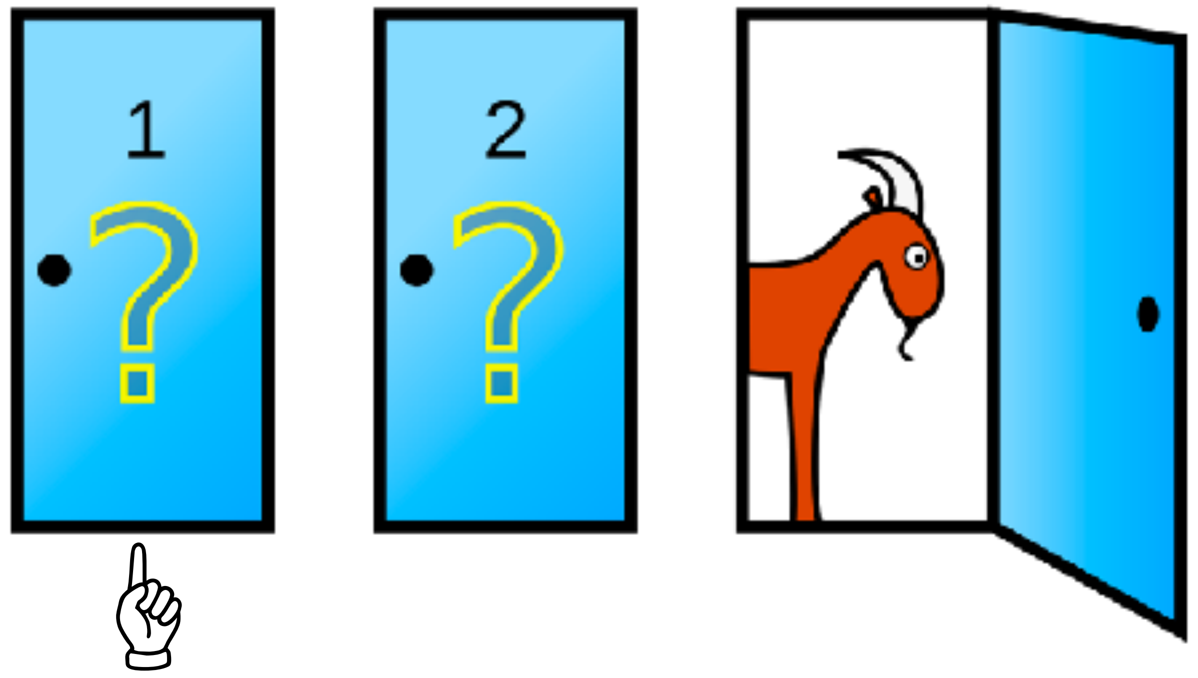
$$\implies \Pr(A) = \sum_{i=1}^{n} \Pr\left(A \cap B_i\right)$$

Moreover: $\Pr\left(A \cap B_i\right) = \Pr(A \mid B_i) \Pr(B_i)$.



Monty Hall Problem

(three doors problem)



- Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats.
- You pick a door, say No.1, and the host, who knows what's behind the doors, opens another door, say No.3, which has a goat. He then says to you, "Do you want to pick door No.2?" Is it to your advantage to switch your choice?

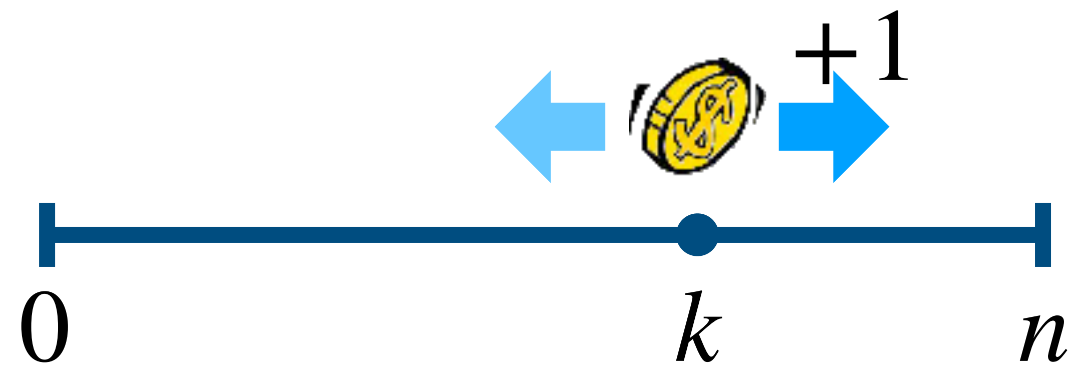
- Define event A : you win at last

  event B : you pick the car at first

$$\Pr(A) = \begin{cases} \Pr(B) = 1/3 & \text{if not switching} \\ \Pr(A \mid B)\Pr(B) + \Pr(A \mid B^c)\Pr(B^c) & \text{if switching} \\ = 0 + 1 \cdot 2/3 = 2/3 & \end{cases}$$

Gambler's Ruin

(Symmetric Random Walk in One-Dimension)



- A gambler plays a fair gambling game: At each step, he flips a fair coin, earns 1 point if it's HEADs, and loses 1 point if otherwise. He starts with k points, and will keep playing until either his points reaches 0 (**lose**) or $n > k$ (**win**).

- Define events A: the gambler loses; and B: the 1st coin flip returns HEADs

- Let $\Pr_k$ be the law that the gambler starts with k points.

$$\Pr_k(A) = \frac{1}{2}\Pr_k(A \mid B) + \frac{1}{2}\Pr_k(A \mid B^c) \ = \frac{1}{2}\Pr_{k+1}(A) + \frac{1}{2}\Pr_{k-1}(A)$$

$$\Pr_k(A) = \begin{cases} \frac{1}{2}(\Pr_{k+1}(A) + \Pr_{k-1}(A)) = 1 - \frac{k}{n} & \text{if } 0 < k < n \\ 1 & \text{if } k = 0 \\ 0 & \text{if } k = n \end{cases}$$

Bayes' Law

(Bayes' Theorem)

- For events A, B that $\Pr(A), \Pr(B) > 0$, we have

$$\Pr(B \mid A) = \frac{\Pr(B)\Pr(A \mid B)}{\Pr(A)}$$

- Let events $B_1, B_2, \ldots, B_n$ be a partition of Ω such that $\Pr(B_i) > 0$ for all i.
  If event A has $\Pr(A) > 0$, then

$$\Pr(B_i \mid A) = \frac{\Pr(B_i)\Pr(A \mid B_i)}{\Pr(A)} = \frac{\Pr(B_i)\Pr(A \mid B_i)}{\Pr(A \mid B_1)\Pr(B_1) + \cdots + \Pr(A \mid B_n)\Pr(B_n)}$$

Dominating False Positives

- A rare disease occurs with probability 0.001.
- 5% testing error:
  - A person with the disease tested $\begin{cases} + & 95\,\% \\ - & 5\,\% \end{cases}$; a person without the disease tested $\begin{cases} + & 5\,\% \\ - & 95\,\% \end{cases}$
- **If a person is tested "+", what is the probability that he/she is ill?**

$$\Pr(\texttt{ill} \mid +) = \frac{\Pr(\texttt{ill})\Pr(+ \mid \texttt{ill})}{\Pr(+)} = \frac{\Pr(\texttt{ill})\Pr(+ \mid \texttt{ill})}{\Pr(+ \mid \texttt{ill})\Pr(\texttt{ill}) + \Pr(+ \mid \neg\texttt{ill})\Pr(\neg\texttt{ill})}$$

$$= \frac{0.001 \times 95\,\%}{95\% \times 0.001 + 5\% \times 0.999} \approx 1.87\,\%$$

Simpson's Paradox

	Women		Men	
	Drug I	**Drug II**	**Drug I**	**Drug II**
Success	200	10	19	1000
Fail	1800	190	1	1000

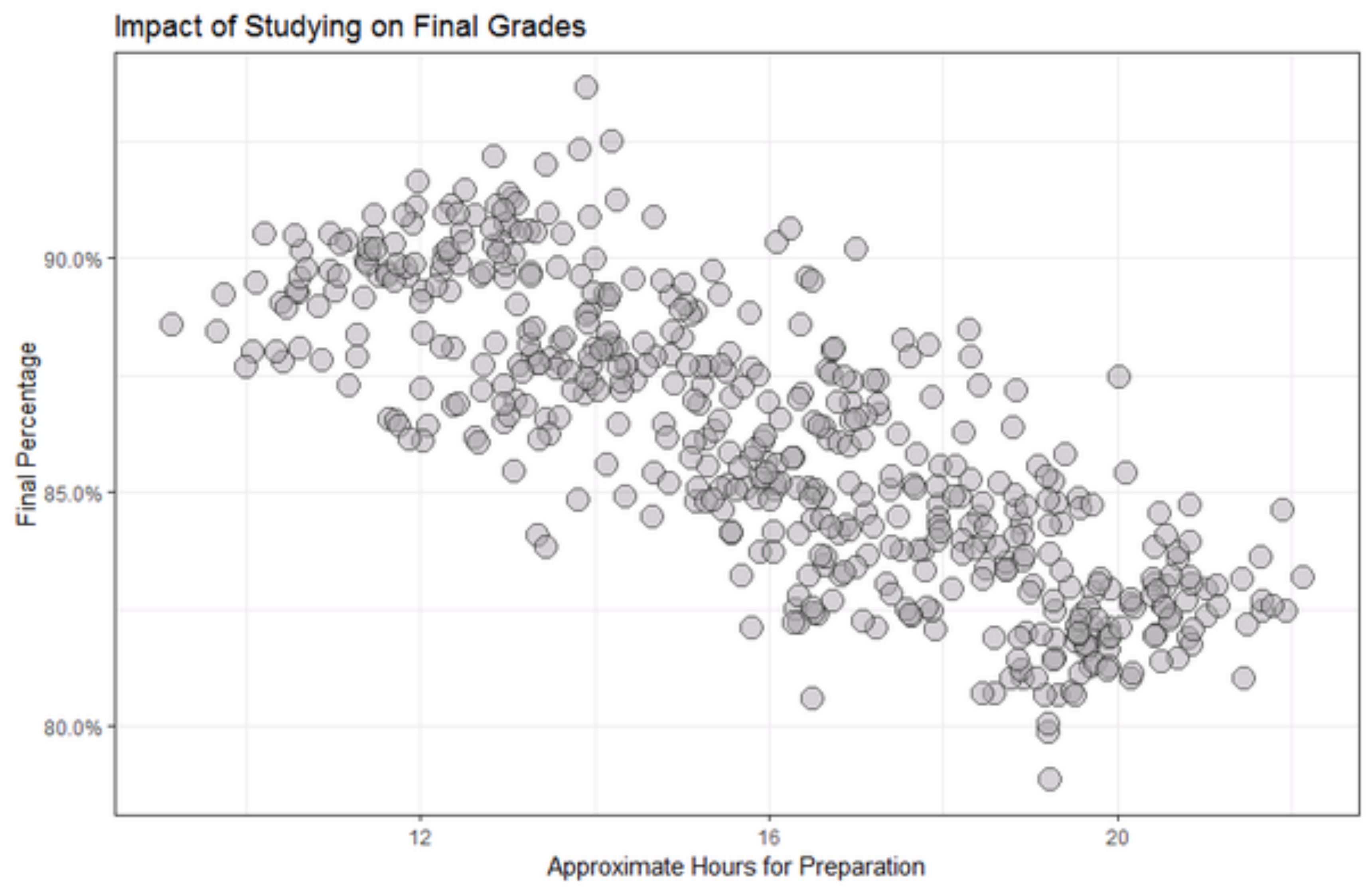
- Results of clinical trials for 2 drugs:
- Which drug is more effective?
  - **Drug-I is better**: for women 1/10 (I) $>$ 1/20 (II), for men 19/20 (I) $>$ 1/2 (II)
  - **Drug-II is better**: overall success rate 219/2020 (I) < 1010/2200 (II)
- **In *Probability***: It's possible that for events A, B and partition $C_1, \ldots, C_n$ of Ω
  - in case for each C_i, the occurrence of B has positive influence on A:
    $$\Pr(A \mid B \cap C_i) > \Pr(A \mid B^c \cap C_i) \text{ for all } i$$
  - but overall, the occurrence of B has negative influence on A:
    $$\Pr(A \mid B) < \Pr(A \mid B^c)$$

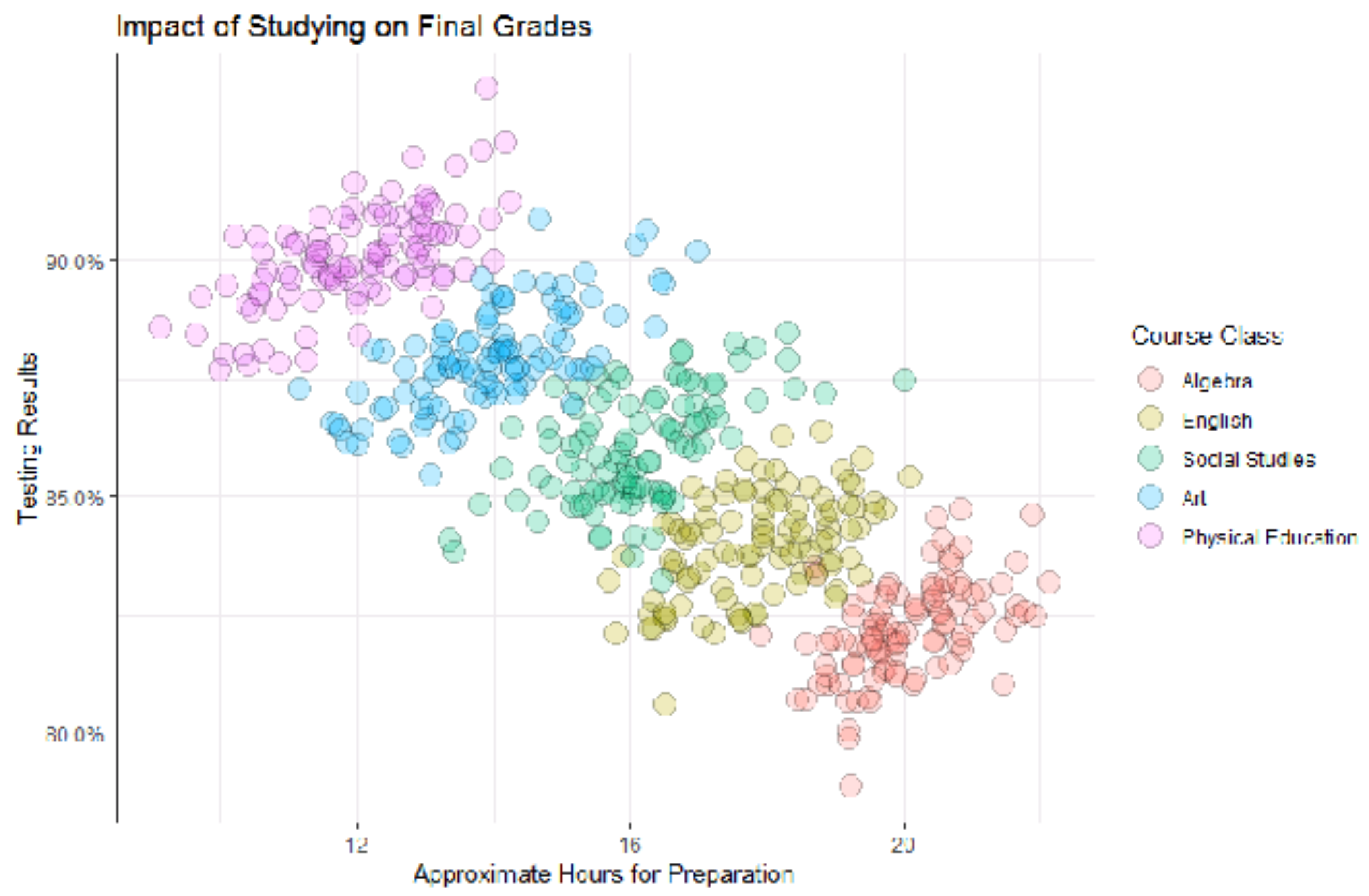
"no scientific discovery is named after its original discoverer."

Simpson's Paradox

(Edward H. Simpson in 1951; Karl Pearson in 1899; Udny Yule in 1903)

- **Example**: Correlation between hours for studying and grades.
  - Overall, it appears that lengths of studying have negative impact on grades. (*The longer the students study, the worse their grades are!*)
  - But truly the they are positively correlated in every course.





Independence

Independence of *Two* Events

- The occurrence of some event B changes the probability of another event A, from $\Pr(A)$ to $\Pr(A \mid B)$.

- If the occurrence of B has no influence on that of A, i.e. $\Pr(A \mid B) = \Pr(A)$, then A is said to be **independent** of B.

- The two events A and B are called **independent** if

$$\Pr(A \cap B) = \Pr(A)\Pr(B)$$

- **Propositions**: if $\Pr(B) > 0$: $\Pr(A \mid B) = \Pr(A) \iff \Pr(A \cap B) = \Pr(A)\Pr(B)$

$$\Pr(A \cap B) = \Pr(A)\Pr(B) \iff \Pr(A \cap B^c) = \Pr(A)\Pr(B^c)$$

Conditional independence

- Two events A and B are **conditionally independent** given C if $\Pr(C) > 0$ and

$$\Pr(A \cap B \mid C) = \Pr(A \mid C) \Pr(B \mid C)$$

- If $\Pr(B \cap C) > 0$: $\Pr(A \cap B \mid C) = \Pr(A \mid C) \Pr(B \mid C) \iff \Pr(A \mid B \cap C) = \Pr(A \mid C)$

- Example: any two events are independent but not conditionally independent given the third event

A: coin-1 is H; B: coin-2 is H; C: coin-1 $\neq$ coin-2;

- Example: A and B are not independent, but they are conditionally independent given C

A: X is tall; B: X knows a lot of math; C: X is 19 years old;

Suppose that X is a random person

Independence of *Several* Events

- A family $\{A_i \mid i \in I\}$ of events is called **(mutually) independent** if for all finite subsets $J \subseteq I$

$$\Pr\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \Pr(A_i)$$

- An event A is called **(mutually) independent** of a family $\{B_i \mid i \in I\}$ of events if for all disjoint finite subsets $J^+, J^- \subseteq I$

$$\Pr(A) = \Pr\left(A \mid \bigcap_{i \in J^+} B_i \cap \bigcap_{i \in J^-} B_i^c\right)$$

Product Probability Space

- Probability space constructed from a sequence of *independent experiments*.
- Consider *discrete* probability spaces $(\Omega_1, p_1), (\Omega_2, p_2), \ldots, (\Omega_n, p_n)$.
- The product probability space (Ω, p) is constructed as:
  - sample space $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$
  - $\forall \omega = (\omega_1, \ldots, \omega_n) \in \Omega$: *pmf* $p(\omega) = p_1(\omega_1) \cdots p_n(\omega_n)$
- For general probability spaces $(\Omega_1, \Sigma_1, \Pr_1), \ldots, (\Omega_n, \Sigma_n, \Pr_n)$, the product probability space $(\Omega, \Sigma, \Pr)$ can be constructed similarly, where Σ is the unique smallest σ-algebra that contains $\Sigma_1 \times \cdots \times \Sigma_n$, and the law $\Pr$ is a natural extension onto such Σ from the product probabilities:
$$\forall A = (A_1, \ldots, A_n) \in \Sigma_1 \times \cdots \times \Sigma_n, \Pr(A) = \Pr(A_1) \cdots \Pr(A_n)$$

Dependency Structure

- The followings are all possible:
  - $A_1, A_2, \ldots, A_n$are mutually independent and $B_1, B_2, \ldots, B_n$ are mutually independent, but A_i and B_i are not independent for every $1 \leq i \leq n$.
  - For every $1 \leq i \leq n$, A_i and B_i are independent, but for every $1 \leq i < j \leq n$, neither A_i and A_j, nor B_i and B_j, are independent.
  - For an arbitrary undirected graph $G(V, E)$ on vertices $V = \{A_1, \ldots, A_n\}$, each A_i is mutually independent of all A_j's that are not adjacent to A_i in G.

Limited Independence

- A family $\{A_i \mid i \in I\}$ of events is called ***pairwise* independent** if for all distinct $i, j \in I$

$$\Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j)$$

- Mutually independent events must be pairwise independent.
- Pairwise independent events are not necessarily mutually independent.
- **Example**: parities (XOR's) of random bits

A: coin-1 is H; B: coin-2 is H; C: coin-3 is H;

D: coin-1 $\neq$ coin-2; E: coin-2 $\neq$ coin-3; F: coin-3 $\neq$ coin-1;

G: # of H in coins-1,2,3 is odd;

Triply Independent but not pairwise

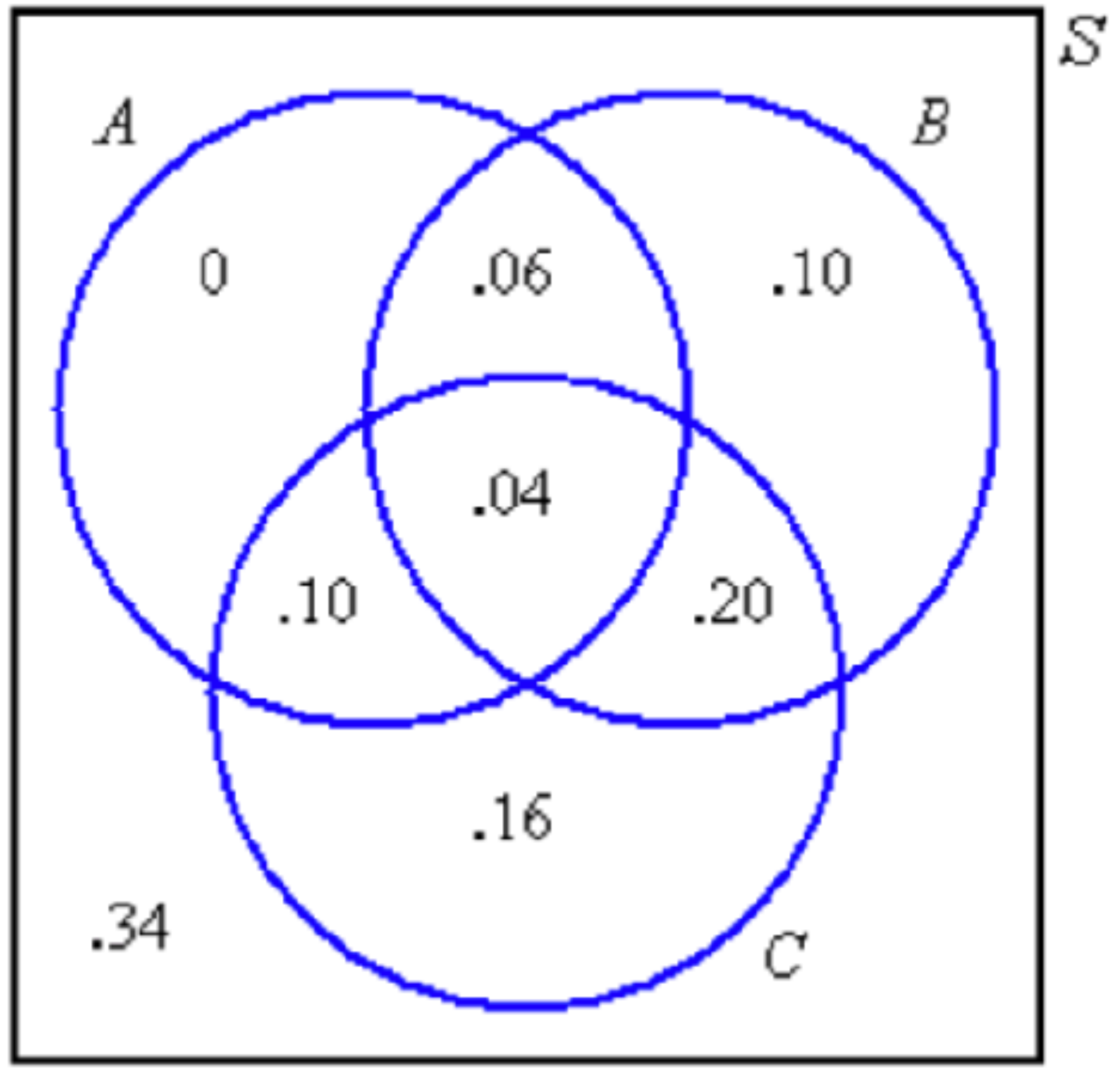


FIGURE 1

- $\Pr(A \cap B \cap C) = \Pr(A)\Pr(B)\Pr(C)$ but no pairwise independence
- Example and figure is from George, Glyn, "Testing for the independence of three events," Mathematical Gazette 88, November 2004, 568

Error Reduction (one-sided case)

- Decision problem $f: \{0,1\}^* \to \{0,1\}$.
- Monte Carlo randomized algorithm $\mathscr{A}$ with ***one-sided*** error:
  - $\forall x \in \{0,1\}^*$: $f(x) = 1 \implies \mathscr{A}(x) = 1$
  - $\forall x \in \{0,1\}^*$: $f(x) = 0 \implies \Pr[\mathscr{A}(x) = 0] \geq p$
- $\mathscr{A}^n$: independently run $\mathscr{A}$ for n times, return $\wedge$ of the n outputs

$$f(x) = 0 \implies \Pr[\mathscr{A}^n(x) = 1] \leq (1 - p)^n$$

- The one-sided error is reduced to ϵ by repeating $n \approx \frac{1}{p} \ln \frac{1}{\epsilon}$ times.

Binomial Probability

- Consider n *independent* tosses of a coin, in which each coin toss returns HEADs independently with probability p.
- We say that we have a sequence of **Bernoulli trials** (伯努利实验), in which each trial succeeds with probability p.
- **Binomial probability**: $p(k) = \Pr(k \text{ successes out of } n \text{ trials})$

$$= \sum_{S \in \binom{[n]}{k}} \Pr(\forall i \in S : i\text{th trial succeeds}) \Pr(\forall i \in [n] \setminus S : i\text{th trial fails})$$

$$= \sum_{S \in \binom{[n]}{k}} p^{|S|}(1-p)^{n-|S|} = \binom{n}{k} p^k (1-p)^{n-k}$$

$p(k)$ is a well-defined *pmf* on $\Omega = \{0,1,\ldots,n\}$

$$\sum_{k=0}^{n} p(k) = 1 \text{ (binomial Thm.)}$$

Controlling a Fair Voting

- In a society of n isolated (independent) and neutral (uniform) people, how many people are there enough to manipulate the result of a majority vote with 95% certainty.

- Consider n independent coin tosses of a fair coin.

$$\Pr[\,|\#\mathtt{HEADs} - \#\mathtt{TAILs}| \geq t] = \Pr[\#\mathtt{HEADs} \leq \tfrac{n}{2} - \tfrac{t}{2}] + \Pr[\#\mathtt{HEADs} \geq \tfrac{n}{2} + \tfrac{t}{2}]$$

$$= \sum_{k \leq (n-t)/2} \binom{n}{k} 2^{-n} + \sum_{k \geq (n+t)/2} \binom{n}{k} 2^{-n}$$

$$= 2^{1-n} \sum_{k \leq (n-t)/2} \binom{n}{k}$$

(entropy bound on the volume of a Hamming ball)

$$\leq 2^{1-n+nH\left(\frac{1}{2} - \frac{t}{2n}\right)}$$

where $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$

$$H(x) \approx 1 - \frac{2}{\ln 2}\left(x - \frac{1}{2}\right)^2 + O\left(\left(x - \frac{1}{2}\right)^3\right)$$

$$\approx 2\exp\left(-\frac{t^2}{2n}\right)$$

≤ 0.05 when $t \geq 2\sqrt{n}$

Error Reduction (two-sided case)

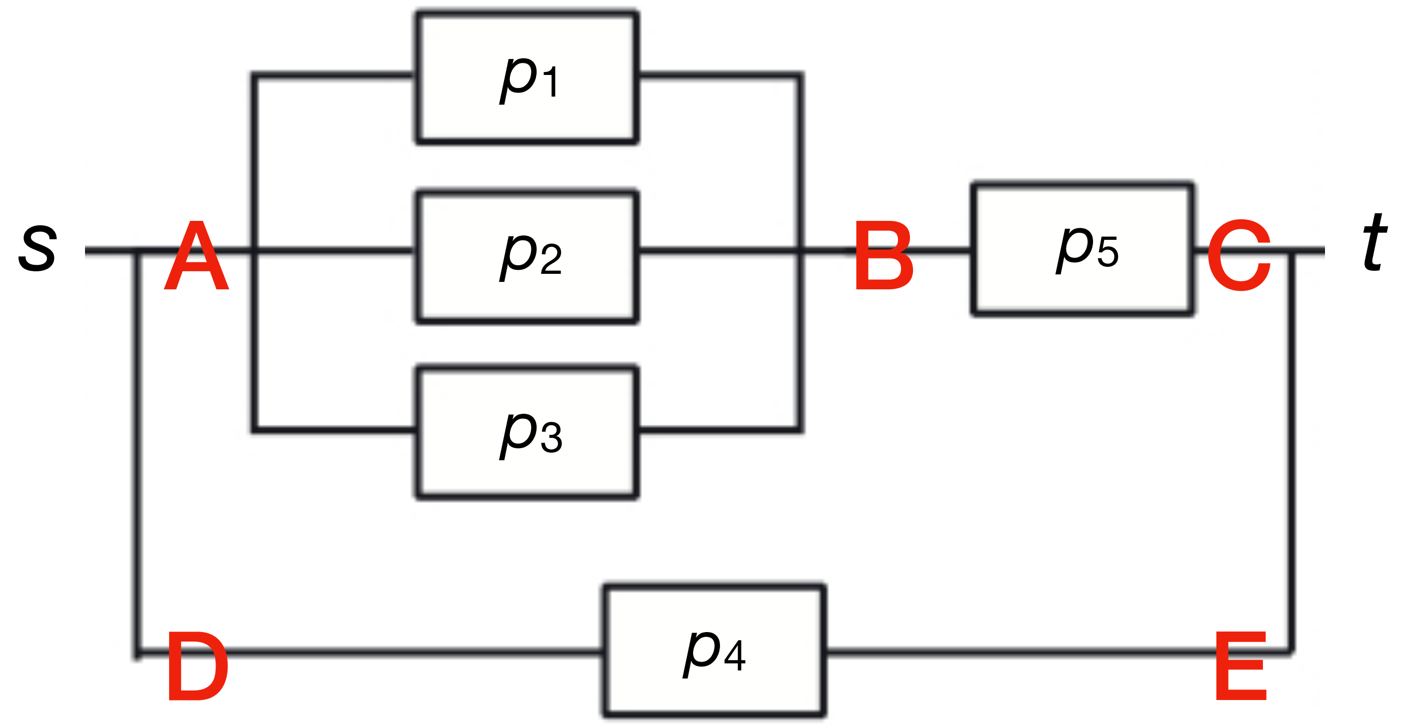
- Decision problem $f : \{0,1\}^* \to \{0,1\}$.
- Monte Carlo randomized algorithm $\mathscr{A}$ with ***two-sided*** error:
  - $\forall x \in \{0,1\}^*: \quad \Pr[\mathscr{A}(x) = f(x)] \geq \frac{1}{2} + p$
- $\mathscr{A}^n$: independently run $\mathscr{A}$ for n times, return majority of the n outputs

$$\Pr[\mathscr{A}^n(x) \neq f(x)] \leq \sum_{k<\frac{n}{2}} \binom{n}{k} \left(\frac{1}{2}+p\right)^k \left(\frac{1}{2}-p\right)^{n-k} \leq \exp(-p^2 n)$$

$$\leq \epsilon \text{ when } n \geq \frac{1}{p^2} \ln \frac{1}{\epsilon}$$

- How to calculate this? (concentration inequalities)

Network Reliability



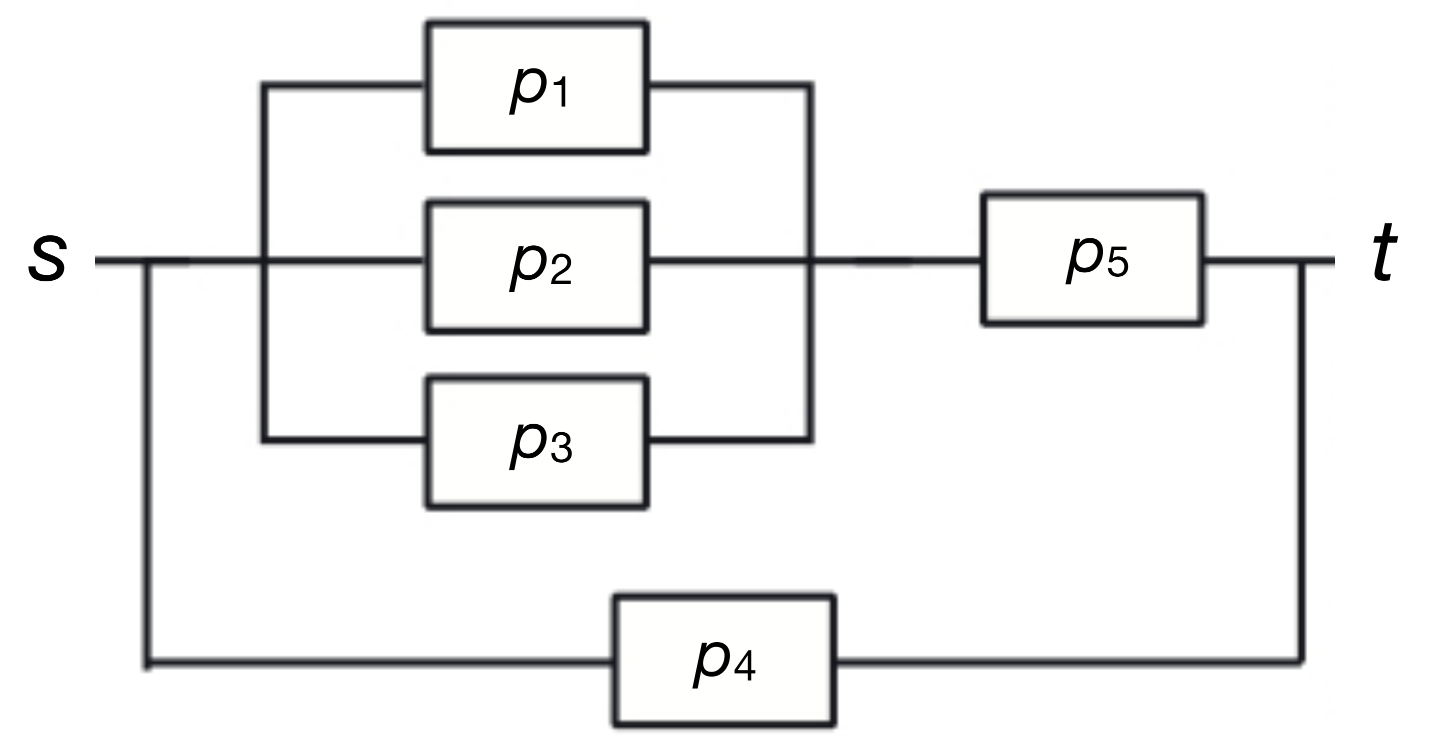
- A serial-parallel (串并联) network connects s to t.
- Suppose that each edge $e = uv$ connects uv independently with probability p_e.
- <u>s-t reliability</u> $P_{st} \triangleq \Pr[\, s \text{ and } t \text{ are connected} \,]$

$$= 1 - (1 - P_{AC})(1 - P_{DE}) \;= 1 - (1 - P_{AC})(1 - p_4)$$

$$P_{AC} = P_{AB}P_{BC} = P_{AB}p_5$$

$$P_{AB} = 1 - (1 - p_1)(1 - p_2)(1 - p_3)$$

Network Reliability



- A ~~serial-parallel (串并联)~~ network connects s to t.
- Suppose that each edge $e = uv$ connects uv independently with probability p_e.
- <u>s-t **reliability**</u> $P_{st} \triangleq \Pr[$ s and t are connected $]$
- <u>(*all-terminal*) **network reliability**</u>: $\triangleq \Pr[$ the resulting network is connected $]$
- For general networks:
  - s-t reliability is **#P-complete** (Leslie Valiant, 1979)
  - all-terminal network reliability is **#P-complete** (Mark Jerrum, 1981)