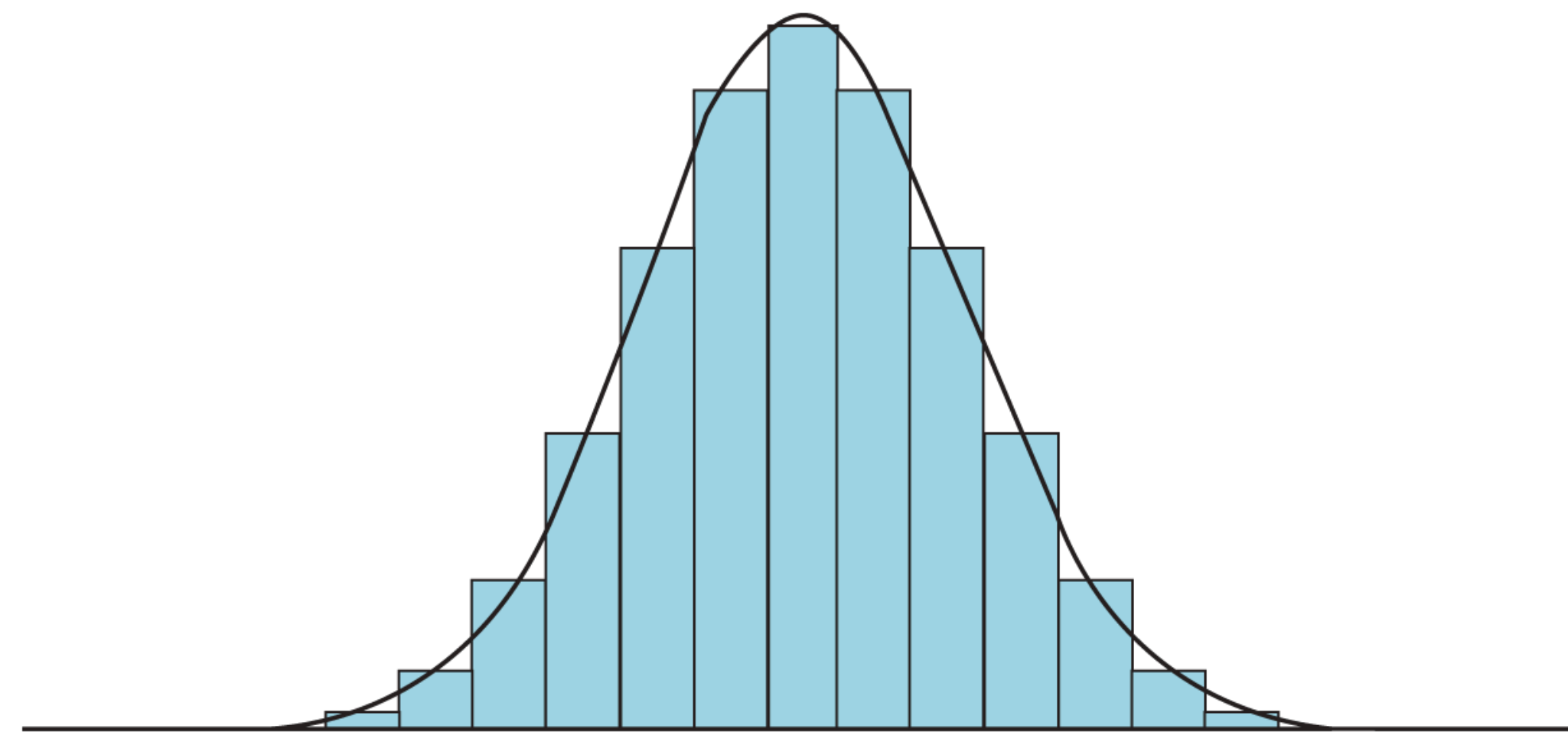


Probability Theory & Mathematical Statistics

Random Variable

Random Variable









“Variables” that are Random

- 令 X 和 Y 分别为两次掷骰子的结果：
 - 考虑 X^2 和 XY ——它们是相同的随机量吗？
 - $2X$ 和 $X + Y$ 呢？或者任意凸组合 $\lambda X + (1 - \lambda)Y$ 之间呢？
- 设正面朝上概率为 p ：令 X 表示连续抛硬币直至正面朝上为止的抛硬币次数；
令 Y 表示连抛 n 次硬币，其中正面朝上的次数；
- 令 X 表示从一个装有 M 个足球、 $N - M$ 个篮球的筐中（有/无放回地）取出 n 个球中足球的个数；
- n 个顶点，任意两点间独立以概率 p 连一条边，产生随机图 G ，令 $X = \chi(G)$ 为最小染色数；
- 令 X 为 $[0,1]$ 中均匀分布的随机实数；令 Y 为 $[0,\infty)$ 上满足 $\Pr(Y \geq y) = e^{-y}$ 的随机实数。


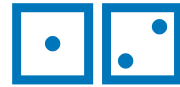



















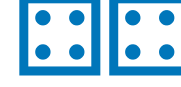














Random Variable

- Roll a , let X be the outcome of the roll, let $Y \in \{0,1\}$ indicate its oddness.

samples in Ω	values of X	values of Y
	1	1
	2	0
	3	1
	4	0
	5	1
	6	0

Random Variable

- Let X be the sum of two independent 🎲 rolls.

	2		3		4		5		6		7
	3		4		5		6		7		8
	4		5		6		7		8		9
	5		6		7		8		9		10
	6		7		8		9		10		11
	7		8		9		10		11		12

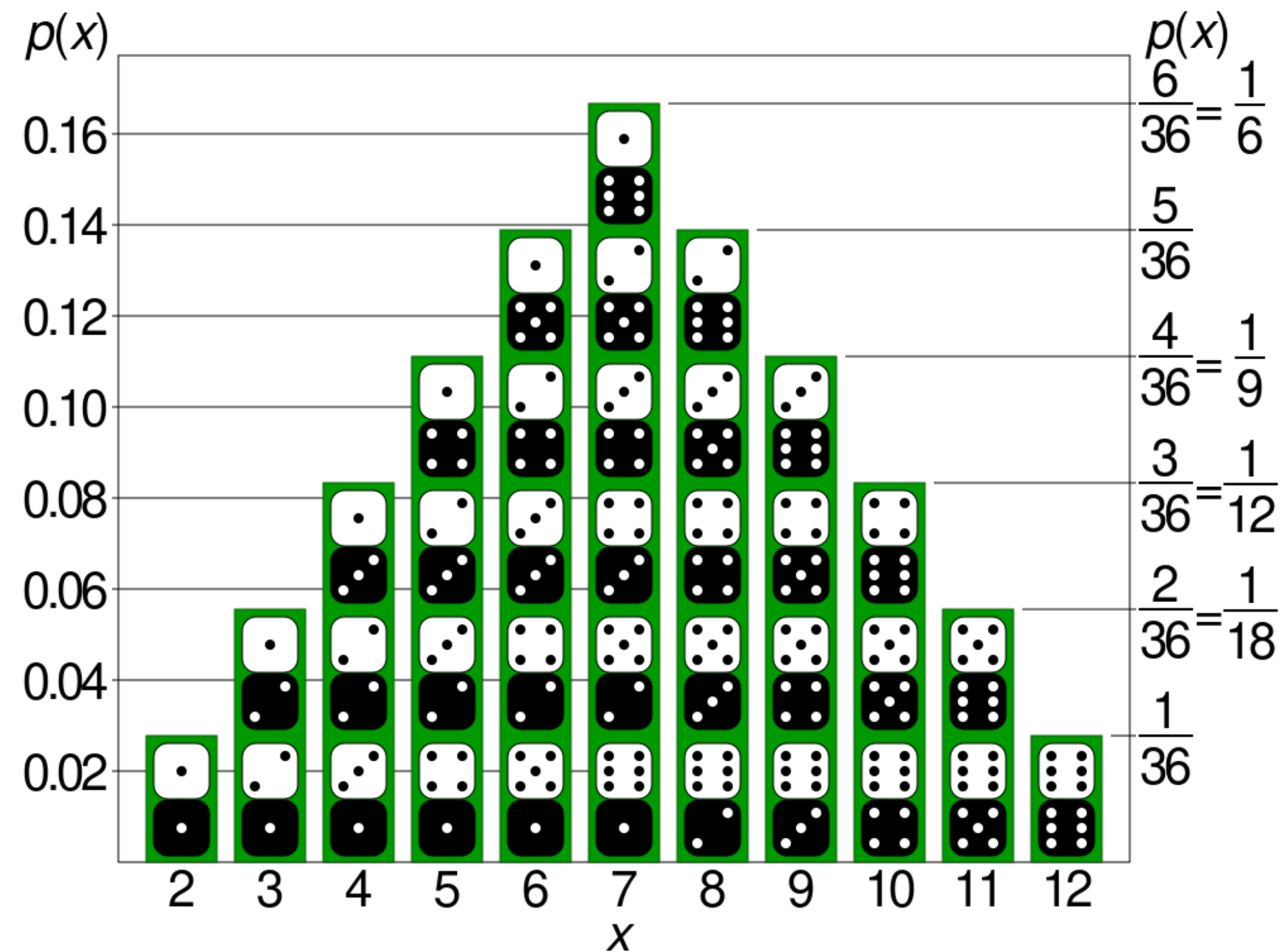
Random Variable (随机变量)

- Given $(\Omega, \Sigma, \text{Pr})$, a **random variable** is a function $X : \Omega \rightarrow \mathbb{R}$
 - satisfying that $\forall x \in \mathbb{R}, \{\omega \in \Omega \mid X(\omega) \leq x\} \in \Sigma$ (i.e. X is Σ -measurable)
- $X \leq x$ (where $x \in \mathbb{R}$) denotes the event $\{\omega \in \Omega \mid X(\omega) \leq x\}$
- $X > x$ (where $x \in \mathbb{R}$) denotes the event $\{\omega \in \Omega \mid X(\omega) > x\}$
- $X \in S$ (where $S \subseteq \mathbb{R}$ is countable \cap, \cup of intervals $(y, x]$) denotes the event $\{\omega \in \Omega \mid X(\omega) \in S\}$
- For discrete random variable $X : \Omega \rightarrow \mathbb{Z}$, this includes all subsets $S \subseteq \mathbb{Z}$

$$\text{Pr}(X \in S)$$

Distribution of Random Variable

- Let X be the sum of two **independent** 🎲 rolls.



Distribution (分布)

- The cumulative distribution function (**CDF**) (累积分布函数) or just distribution function (分布函数) of a random variable X is the $F_X : \mathbb{R} \rightarrow [0,1]$ given by

$$F_X(x) = \Pr(X \leq x)$$

- All probabilities regarding X can be deduced from F_X . (Prob. space is no longer needed.)
- Two random variables X and Y are identically distributed if $F_X = F_Y$
- Monotone: $\forall x, y \in \mathbb{R}$, if $x \leq y$ then $F_X(x) \leq F_X(y)$
- Bounded: $\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow \infty} F_X(x) = 1$

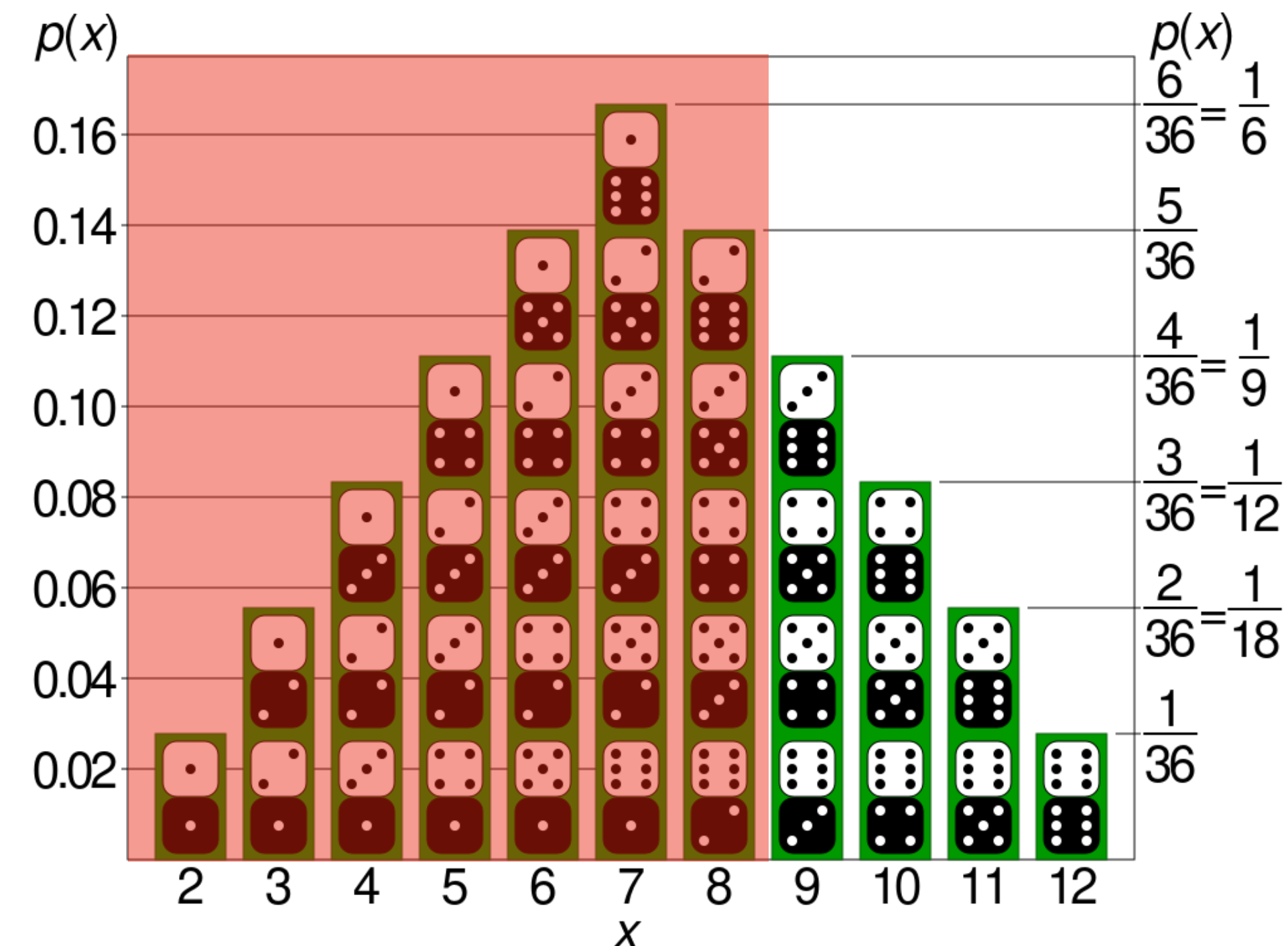
Discrete Random Variable

- A random variable $X : \Omega \rightarrow \mathbb{R}$ is called **discrete** if $X(\Omega)$ is countable.
- For a discrete random variable X , its **probability mass function (pmf)** (概率质量函数) $p_X : \mathbb{R} \rightarrow [0,1]$ is given by

$$p_X(x) = \Pr(X = x)$$

- The CDF F_X satisfies

$$F_X(y) = \sum_{x \leq y} p_X(x)$$



Continuous Random Variable

- A random variable $X : \Omega \rightarrow \mathbb{R}$ is called **continuous**, if its CDF can be expressed as

$$F_X(y) = \Pr(X \leq y) = \int_{-\infty}^y f_X(x) dx$$

for some integrable **probability density function (pdf)** (概率密度函数) f_X .

- Never mind what type of integral for now. (Riemann integral? Lebesgue integral?)
- There are random variables that are neither discrete nor continuous.

Independence

- Two *discrete* random variables X and Y are independent if $X = x$ and $Y = y$ are independent events for all x and y .
- *Discrete* random variables X_1, \dots, X_n are (mutually) independent if $X_1 = x_1, \dots, X_n = x_n$ are mutually independent events for all x_1, \dots, x_n
$$P_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \Pr(X_1 = x_1 \cap \dots \cap X_n = x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$$
- The pairwise (and k -wise) independence are defined in the same way.
 - **Example:** The construction of $2^n - 1$ pairwise independent random bits out of n mutually independent random bits by XOR.
- For *general* random variables, the events $X_i = x_i$ are replaced by $X_i \leq x_i$.

Random Vector (随机向量)

- Given (Ω, Σ, \Pr) , a random vector is an $X = (X_1, \dots, X_n)$ where each X_i is a random variable defined on the probability space (Ω, Σ, \Pr) .

- The joint CDF (联合累积分布函数) $F_X : \mathbb{R}^n \rightarrow [0,1]$ is given by

$$F_X(x_1, \dots, x_n) = \Pr(X_1 \leq x_1 \cap \dots \cap X_n \leq x_n)$$

- For *discrete* random vector, the joint mass function (联合质量函数) is given by

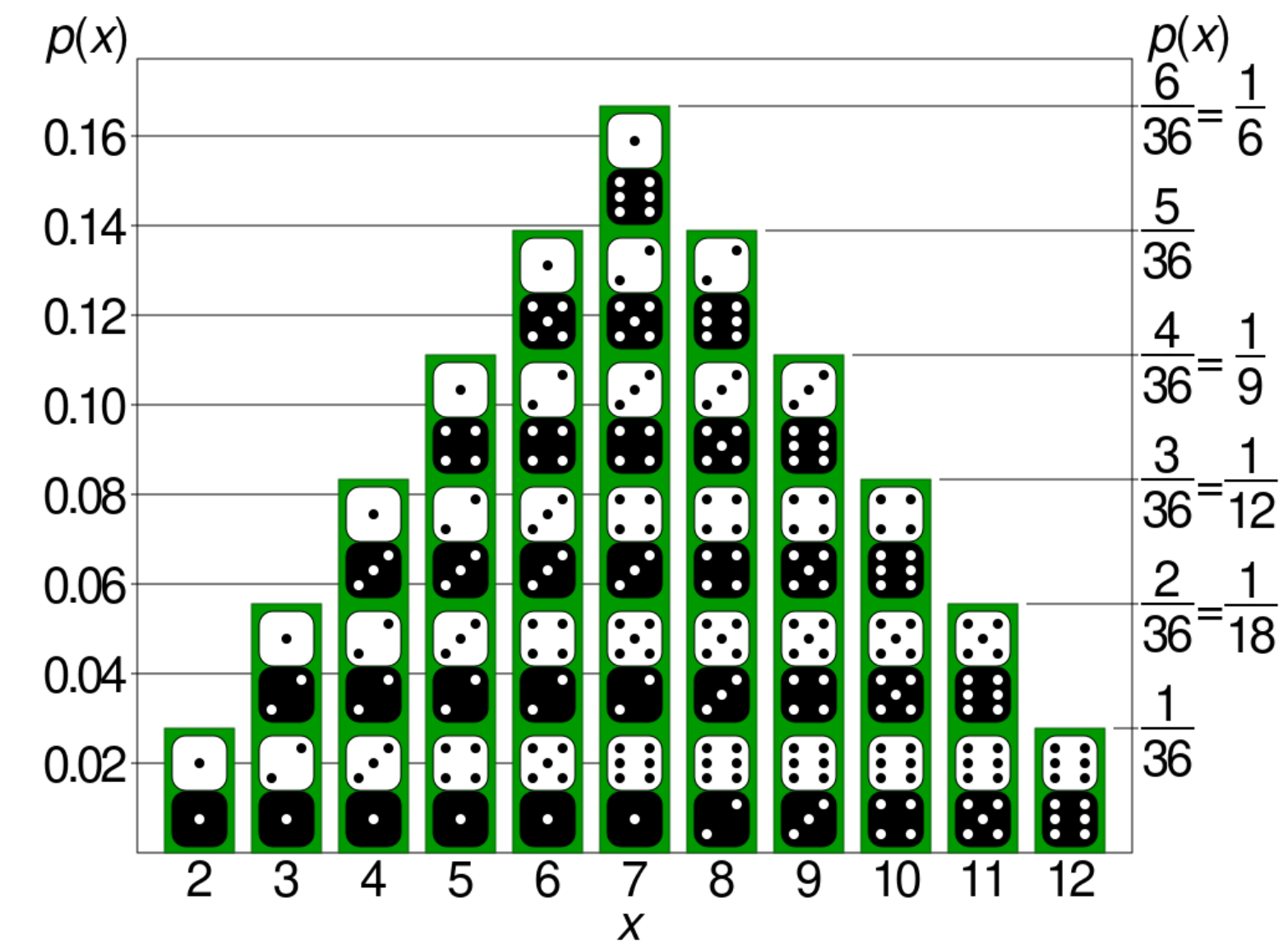
$$p_X(x_1, \dots, x_n) = \Pr(X_1 = x_1 \cap \dots \cap X_n = x_n)$$

- The marginal distribution of X_i in (X_1, \dots, X_n) is given by

$$p_{X_i}(x_i) = \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} p_{(X_1, \dots, X_n)}(x_1, \dots, x_n)$$

$Y \backslash X$	x_1	x_2	x_3	x_4	$p_Y(y) \downarrow$
y_1	$\frac{4}{32}$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{8}{32}$
y_2	$\frac{3}{32}$	$\frac{6}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{15}{32}$
y_3	$\frac{9}{32}$	0	0	0	$\frac{9}{32}$
$p_X(x) \rightarrow$	$\frac{16}{32}$	$\frac{8}{32}$	$\frac{4}{32}$	$\frac{4}{32}$	$\frac{32}{32}$

Discrete Random Variable



Probability Mass Function (概率质量函数)

- Consider *integer-valued* discrete random variable $X : \Omega \rightarrow \mathbb{Z}$
- Its probability mass function (*pmf*) $p_X : \mathbb{Z} \rightarrow [0,1]$ is given by

$$p_X(k) = \Pr(X = k)$$

- As histogram: p_X gives the “histogram” of the probability distribution
- As vector: p_X can be seen as a vector $p_X \in [0,1]^R$ such that $\|p_X(x)\|_1 = 1$, where $R = X(\Omega)$ is the range of values of X

- Its function $Y = f(X)$ is also a discrete random variable, where $p_Y(y) = \sum_{x:f(x)=y} p_X(x)$

Discrete Random Variables

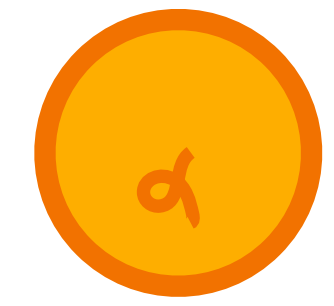
- Basic discrete probability distributions:
 - discrete uniform distribution (古典概型)
 - **Bernoulli trial** (coin flip)
 - **binomial distribution** (# of successes in n trials)
 - **geometric distribution** (# of trials to get a success)
 - negative binomial distribution
 - hypergeometric distribution
 - **Poisson distribution** (idealized binomial distribution)
 -
- Probability distributions of discrete objects:
 - multinomial distribution (balls into bins)
 - Erdős–Rényi model (random graph)
 - Galton-Watson process (random tree)
 -

Bernoulli Trial (伯努利试验)

(A coin flip)



p



$1 - p$

- A Bernoulli trial is an experiment with two possible outcomes.
- A Bernoulli random variable X takes values in $\{0,1\}$, its *pmf* is

$$p_X(k) = \Pr(X = k) = \begin{cases} p & \text{if } k = 1 \\ 1 - p & \text{if } k = 0 \end{cases}$$

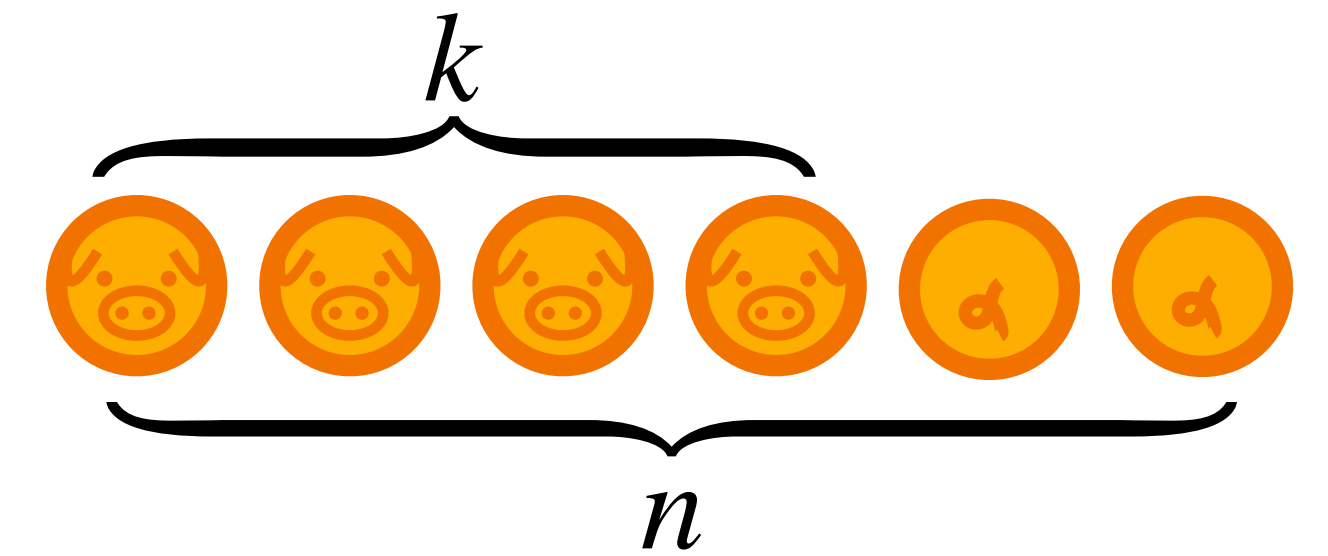
where $p \in [0,1]$ is a parameter.

- Indicator: For event A , the indicator X of A is a random variable defined by

$$X = I(A) = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}, \quad \text{a Bernoulli R.V. with parameter } \Pr(A)$$

Binomial Distribution (二项分布)

(Number of HEADS in n coin flips)



- X : number of successes in n i.i.d. (*independent and identically distributed*) Bernoulli trials with parameter p
- A binomial random variable X takes values in $\{0, 1, \dots, n\}$, and

$$p_X(k) = \Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n$$

- We say that X follows the binomial distribution with parameters n and p
denoted $X \sim \text{Bin}(n, p)$ or $B(n, p)$

Geometric Distribution (几何分布)

(Number of coin flips to get a HEADs)



- X : number of i.i.d. Bernoulli trials needed to get one success

- A geometric random variable X takes values in $\{1, 2, \dots\}$, and

$$p_X(k) = \Pr(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

- We say that X follows the geometric distribution with parameter $p \in [0, 1]$

denoted $X \sim \text{Geo}(p)$ or $\text{Geometric}(p)$

Geometric Distribution (几何分布)

(Number of coin flips to get a HEADs)



- Geometric random variable $X \sim \text{Geo}(p)$ is **memoryless**: for $k \geq 1, n \geq 0$

$$\Pr(X = k + n \mid X > n) = \Pr(X = k)$$

Proof:
$$\Pr(X = k + n \mid X > n) = \frac{\Pr(X = k + n)}{\Pr(X > n)} = \frac{(1 - p)^{n+k-1} p}{\sum_{k=n}^{\infty} (1 - p)^k p}$$
$$= \frac{(1 - p)^{k-1} p}{\sum_{k=0}^{\infty} (1 - p)^k p} = (1 - p)^{k-1} p$$

- Geometric distribution is the **only** discrete memoryless distribution (with the range of values $\{1, 2, \dots\}$).

Two Ways of Constructing Random Variables

- As a function of random variables $Y = f(X_1, X_2, \dots, X_n)$
 - Binomial Y : function f is sum, and (X_1, \dots, X_n) are i.i.d. Bernoulli trials
 - independent $Y_1 \sim \text{Bin}(n_1, p)$, $Y_2 \sim \text{Bin}(n_2, p) \implies Y_1 + Y_2 \sim \text{Bin}(n_1 + n_2, p)$
- As a stopping time T of a sequence X_1, X_2, \dots, X_T
 - A random variable T is a stopping time with respect to X_1, X_2, \dots if for all $t \geq 1$ the occurrence of $T = t$ is determined by the values of X_1, X_2, \dots, X_t
 - Geometric T : time for the first success in i.i.d. Bernoulli trials X_1, X_2, \dots

Sum of Independent Random Variables

- If discrete random variables X and Y are independent, then:

$$p_{X+Y}(z) = \Pr(X + Y = z) = \sum_x \Pr(X = x \cap Y = z - x) \quad \text{(total probability)}$$
$$\text{(independence)} \quad = \sum_x p_X(x)p_Y(z - x) = \sum_y p_X(z - y)p_Y(y)$$

- This defines a convolution (卷积) between mass functions:

$$p_{X+Y} = p_X * p_Y$$

Sum of Independent Random Variables

- If discrete random variables X and Y are independent, then:

$$p_{X+Y}(z) = \sum_x p_X(x)p_Y(z-x) = \sum_y p_X(z-y)p_Y(y)$$

- For *i.i.d.* Bernoulli random variables $X_1, \dots, X_n \in \{0,1\}$:

$$\begin{aligned} p_{X_1+\dots+X_n}(k) &= p \cdot p_{X_1+\dots+X_{n-1}}(k-1) + (1-p) \cdot p_{X_1+\dots+X_{n-1}}(k) \\ &= \binom{n-1}{k-1} p^k (1-p)^{n-k} + \binom{n-1}{k} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

Negative Binomial Distribution (负二项分布)

(“multiple successes” generalization of geometric distribution)

- X : number of failures in a sequence of i.i.d. Bernoulli trials before r successes
- A negative binomial random variable X takes values in $\{0, 1, 2, \dots\}$, and

$$p_X(k) = \Pr(X = k) = \binom{k + r - 1}{k} (1 - p)^k p^r = (-1)^k \binom{-r}{k} (1 - p)^k p^r$$

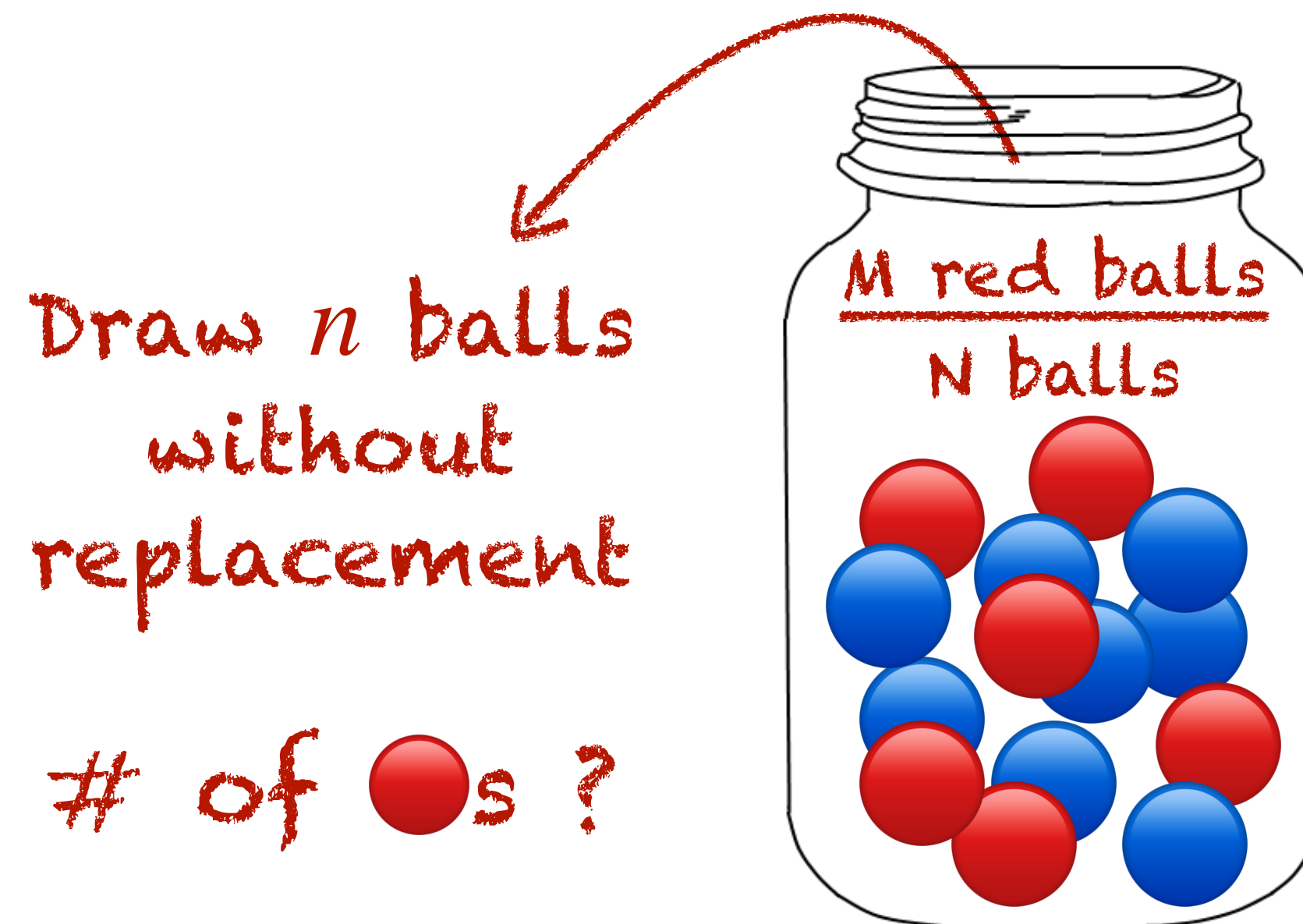
for $k = 0, 1, 2, \dots$

- We say that X follows the negative binomial distribution with parameters r, p
- $X = (X_1 - 1) + (X_2 - 1) + \dots + (X_r - 1)$ for i.i.d. $X_i \sim \text{Geo}(p)$

Hypergeometric Distribution (超几何分布)

(“without replacement” variant of binomial distribution)

- X : number of successes in n draws, without replacement (无放回), from a *finite population* of N objects, including exactly M ones, drawings of whom are counted as successes



Hypergeometric Distribution (超几何分布)

(“without replacement” variant of binomial distribution)

- X : number of successes in n draws, without replacement (无放回), from a *finite population* of N objects, including exactly M ones, drawings of whom are counted as successes
- A hypergeometric random variable X takes values in $\{0, 1, \dots, n\}$, and

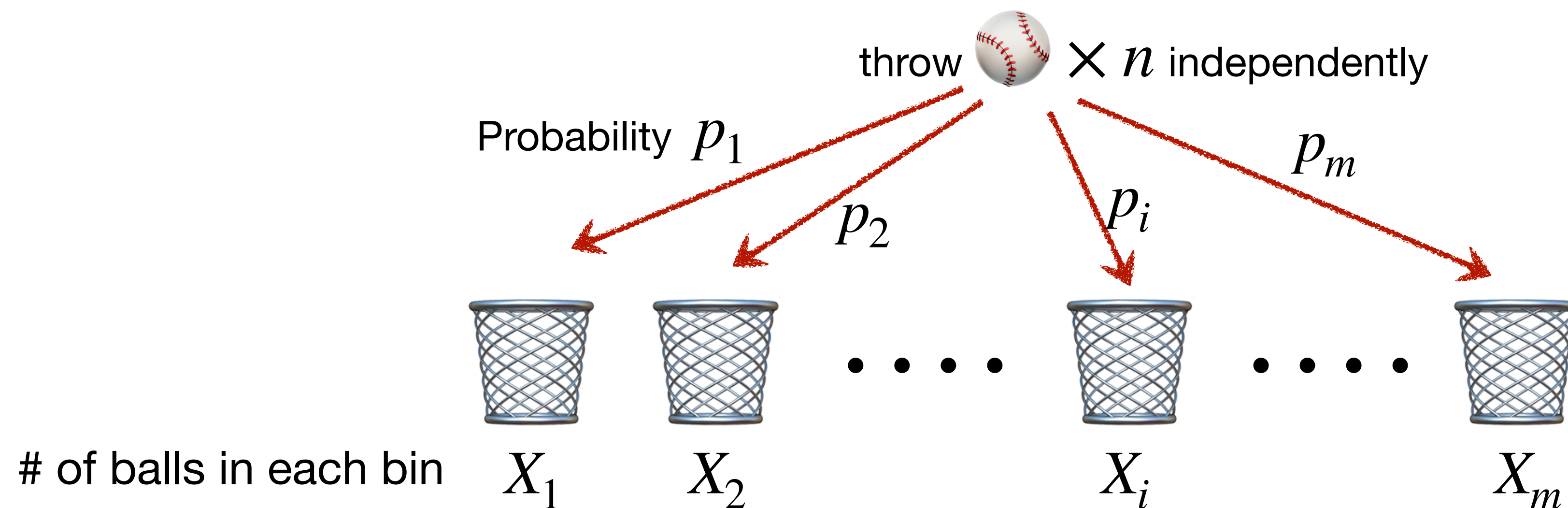
$$p_X(k) = \Pr(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \dots, n$$

- We say that X follows the hypergeometric distribution with parameters N, M, n , where $N \geq 0$, $0 \leq M \leq N$, and $0 \leq n \leq N$ are integers.

Multinomial Distribution (多项式分布)

(“multi-dimensional” generalization of binomial distribution)

- **Trials with multiple outcomes:** There are n *i.i.d.* trials, each having m possible outcomes, where the probability of the i th outcome is p_i . Let X_i be the # of i th outcomes.
- **Balls-into-bins model:** Throw n balls into m bins. Each ball is thrown independently such that the i th bin receives the ball with probability p_i . Let X_i be the # of balls in the i th bin.



Multinomial Distribution (多项式分布)

(“multi-dimensional” generalization of binomial distribution)

- Suppose that n balls are thrown into m bins, where each ball is thrown independently such that the i th bin receives the ball with probability p_i , where $p_1 + \dots + p_m = 1$ is given.

- (X_1, X_2, \dots, X_m) : the i th bin receives exactly X_i balls

- (X_1, \dots, X_m) takes values $(k_1, \dots, k_m) \in \{0, 1, \dots, n\}^m$ that $k_1 + \dots + k_m = n$, and

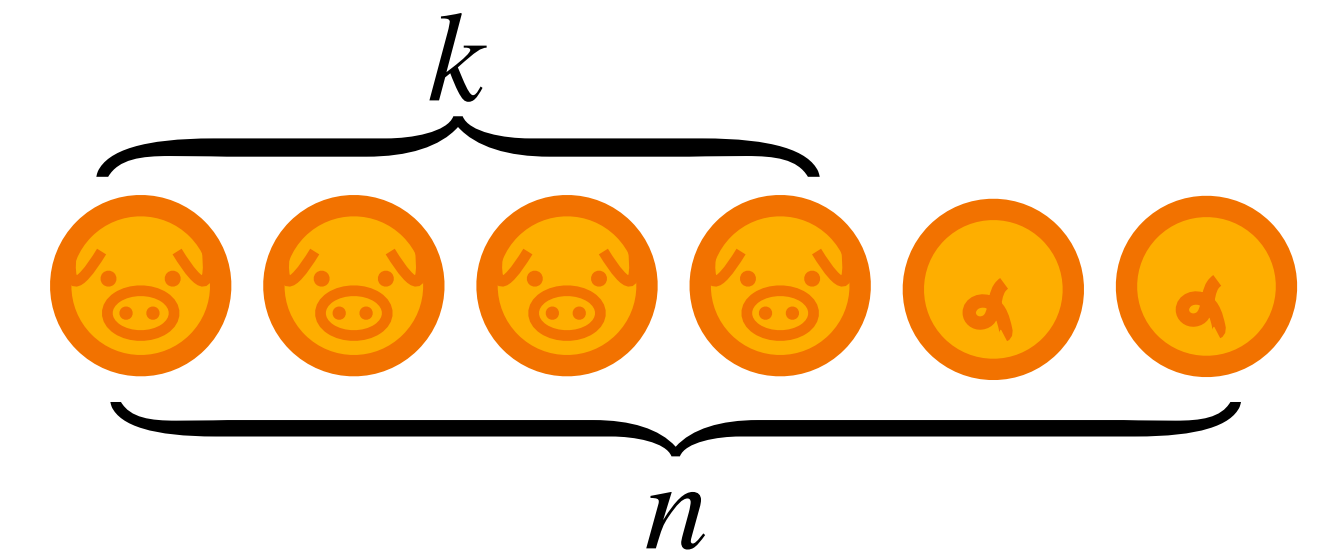
$$P_{(X_1, \dots, X_m)}(k_1, \dots, k_m) = \Pr \left(\bigcap_{i=1}^m (X_i = k_i) \right) = \frac{n!}{k_1! k_2! \dots k_m!} p_1^{k_1} p_2^{k_2} \dots p_m^{k_m}$$

- We say that (X_1, X_2, \dots, X_m) follows the multinomial distribution with parameters m , n , and $p = (p_1, \dots, p_m) \in [0, 1]^m$ such that $p_1 + \dots + p_m = 1$.

- $X_i \sim \text{Bin}(n, p_i)$ for each individual $1 \leq i \leq m$. (The marginal distribution of X_i is $\text{Bin}(n, p_i)$)

Binomial Distribution (二项分布)

(Number of HEADs in n coin flips)



- X : number of successes in n *i.i.d.* Bernoulli trials with parameter p
- A **binomial random variable** X takes values in $\{0, 1, \dots, n\}$, and

$$p_X(k) = \Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n$$

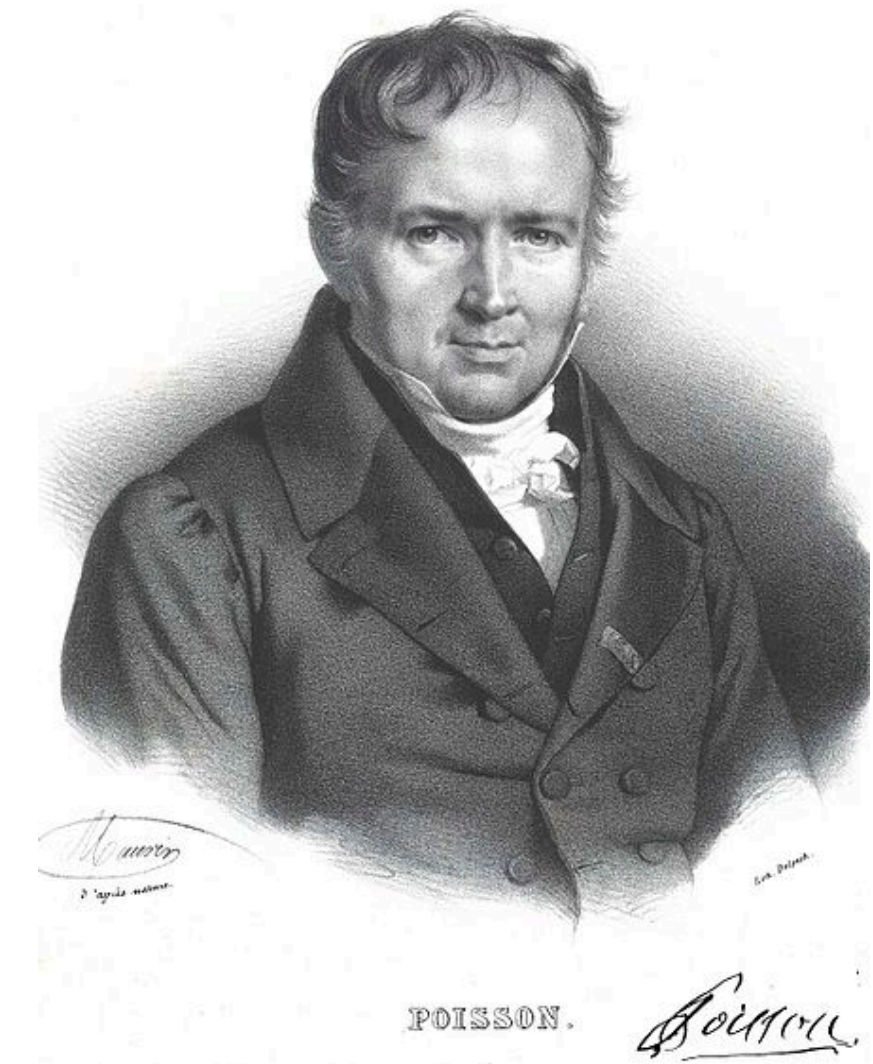
- Typical in real life: large unknown population size $n \rightarrow \infty$ with known $np = \lambda$

$$p_{\text{Bin}(n, \lambda/n)}(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n}{n} \frac{n-1}{n} \dots \frac{n-k+1}{n} \cdot \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

A “universal” distribution for all sufficiently large n , knowing the mean $\lambda = np$?

Poisson Distribution (泊松分布)

(Idealized binomial distribution when $n \rightarrow \infty$)



- A Poisson random variable X takes values in $\{0, 1, 2, \dots\}$, and

$$p_X(k) = \Pr(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

- It is a well-defined probability distribution over $\{0, 1, 2, \dots\}$: $\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1$
- We say that X follows the Poisson distribution with parameter $\lambda > 0$

denoted $X \sim \text{Pois}(\lambda)$

Sum of Poisson Variables

- Independent $X \sim \text{Bin}(n_1, p)$, $Y \sim \text{Bin}(n_2, p) \implies X + Y \sim \text{Bin}(n_1 + n_2, p)$
- By the *heuristics* $\text{Bin}(n, p) \approx \text{Pois}(np)$, it seems that the following should hold:

- independent $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2) \implies X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$

- **Proof:**
$$p_{X+Y}(k) = \Pr(X + Y = k) = \sum_{i=0}^k \Pr(X = i \cap Y = k - i) = \sum_{i=0}^k p_X(i)p_Y(k - i)$$
$$= \sum_{i=0}^k \frac{e^{-\lambda_1} \lambda_1^i}{i!} \frac{e^{-\lambda_2} \lambda_2^{k-i}}{(k-i)!} = \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \sum_{i=0}^k \binom{k}{i} \lambda_1^i \lambda_2^{k-i} = \frac{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^k}{k!}$$

Poisson Approximation

- (X_1, \dots, X_m) follows the multinomial distribution with parameters $m, n, p_1 + \dots + p_m = 1$
 - n balls are thrown into m bins independently according to the distribution (p_1, \dots, p_m)
 - after all n balls are thrown, the i th bin receives X_i balls
- (Y_1, \dots, Y_m) : each $Y_i \sim \text{Pois}(\lambda_i)$ independently, where $\lambda_i = np_i$

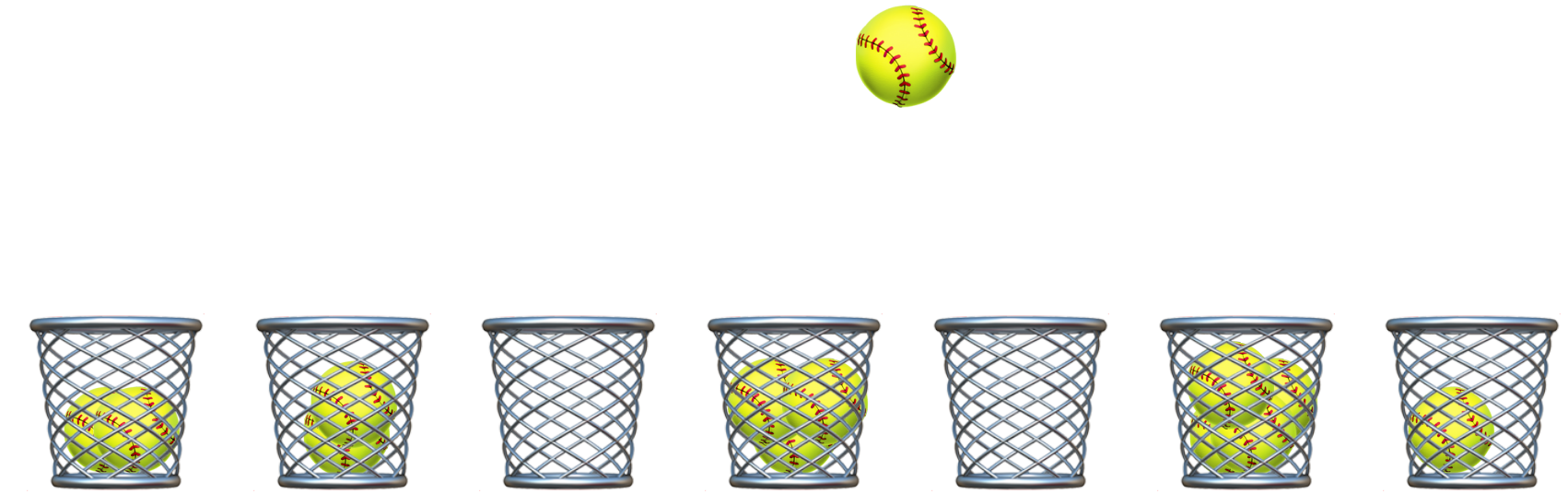
Proposition: (X_1, \dots, X_m) is identically distributed as (Y_1, \dots, Y_m) given that $\sum_{i=1}^m Y_i = n$

Proof: Observe that $Y_1 + \dots + Y_m \sim \text{Pois}(n)$. For any $k_1, \dots, k_m \geq 0$ that $k_1 + \dots + k_m = n$:

$$\begin{aligned} \Pr[(Y_1, \dots, Y_m) = (k_1, \dots, k_m) \mid Y_1 + \dots + Y_m = n] &= \left(\prod_{i=1}^m \frac{e^{-np_i} (np_i)^{k_i}}{k_i!} \right) / \left(\frac{e^{-n} n^n}{n!} \right) \\ &= \frac{n!}{k_1! \dots k_m!} p_1^{k_1} \dots p_m^{k_m} = \Pr[(X_1, \dots, X_m) = (k_1, \dots, k_m)] \end{aligned}$$

Balls into Bins

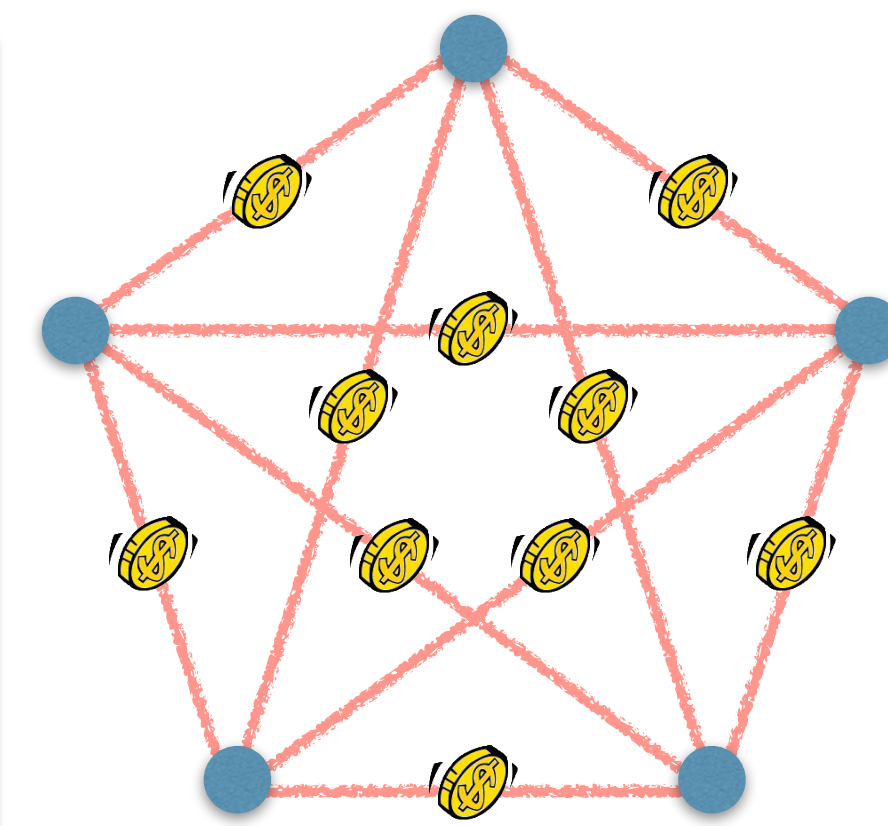
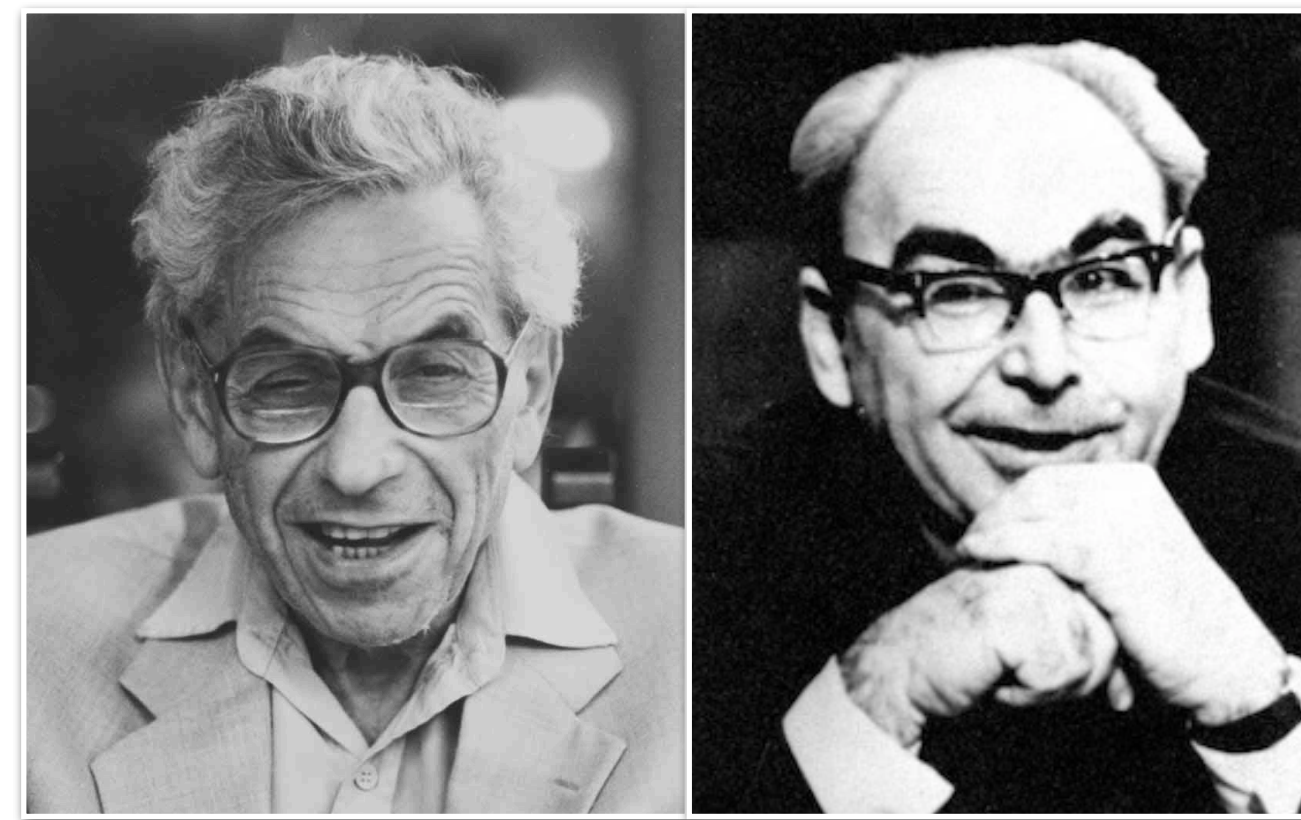
(Random mapping)



- Throw n balls into m bins uniformly at random (*u.a.r.*).
- Uniform random $f : [n] \rightarrow [m]$.
- The numbers of balls received in each bins (X_1, \dots, X_m) follow the multinomial distribution with parameters m, n and $(1/m, \dots, 1/m)$.
 - Birthday problem: the property of being injective (1-1)
 - Coupon collector problem: the property of being surjective (onto)
 - Occupancy (load balancing) problem: the *maximum load* $\max_i X_i$

Random Graph

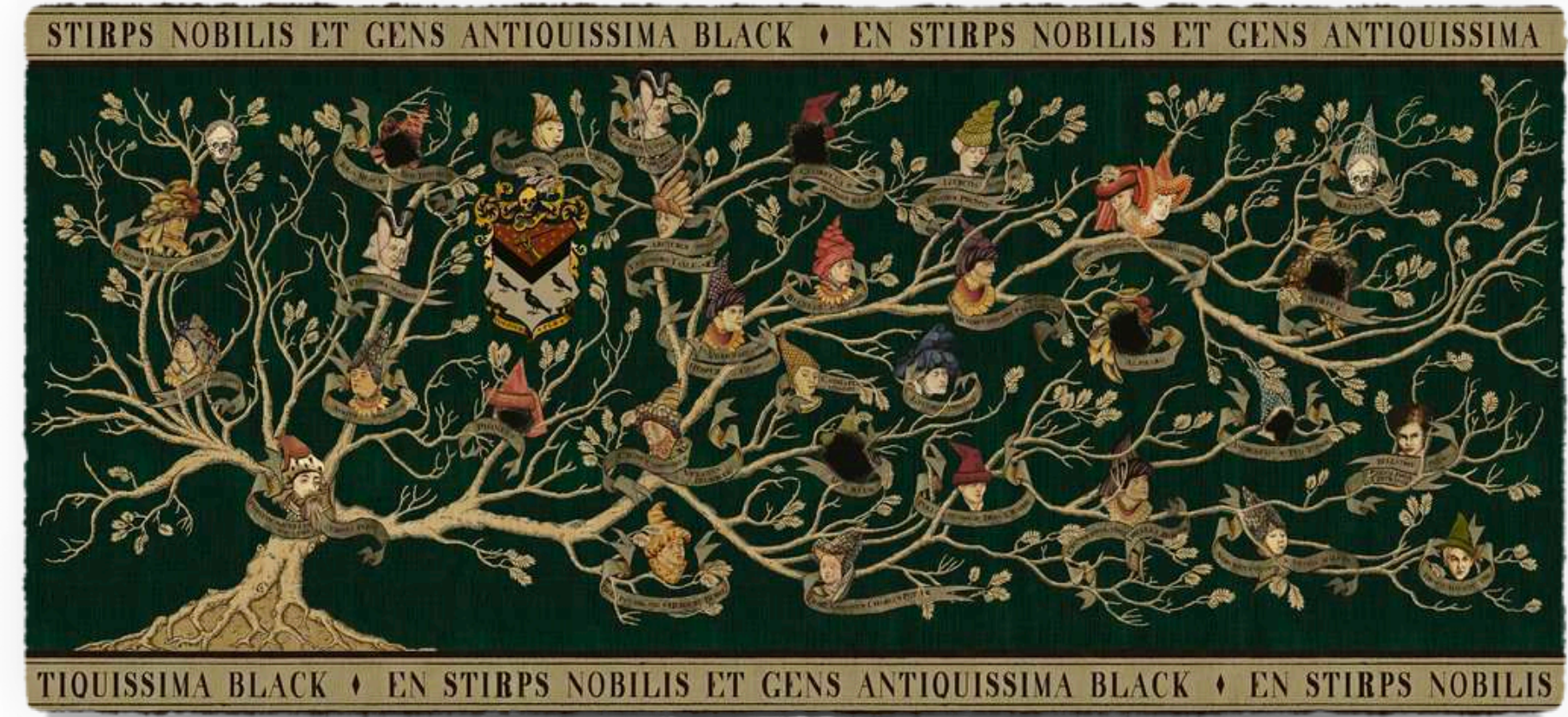
(Erdős–Rényi random graph model)



- $G \sim G(n, p)$: There are n vertices. For each pair u, v of vertices, an *i.i.d.* Bernoulli trial with parameter p is conducted, and an edge $\{u, v\}$ is added if the trial succeeds.
- $G(n, 1/2)$ gives the uniformly distributed random graph on n vertices.
- The number of edges in $G \sim G(n, p)$ follows the binomial distribution $\text{Bin} \left(\binom{n}{2}, p \right)$.
(Therefore, $G(n, p)$ is sometimes also called the *binomial random graph*)
- Random variables defined by $G \sim G(n, p)$: *chromatic number* $\chi(G)$, *independence number* $\alpha(G)$, *clique number* $\omega(G)$, *diameter* $\text{diam}(G)$, *connectivity*, *max-degree* $\Delta(G)$, *number of triangles*, *number of hamiltonian cycles*, ...

Random Tree

(Galton–Watson branching process)



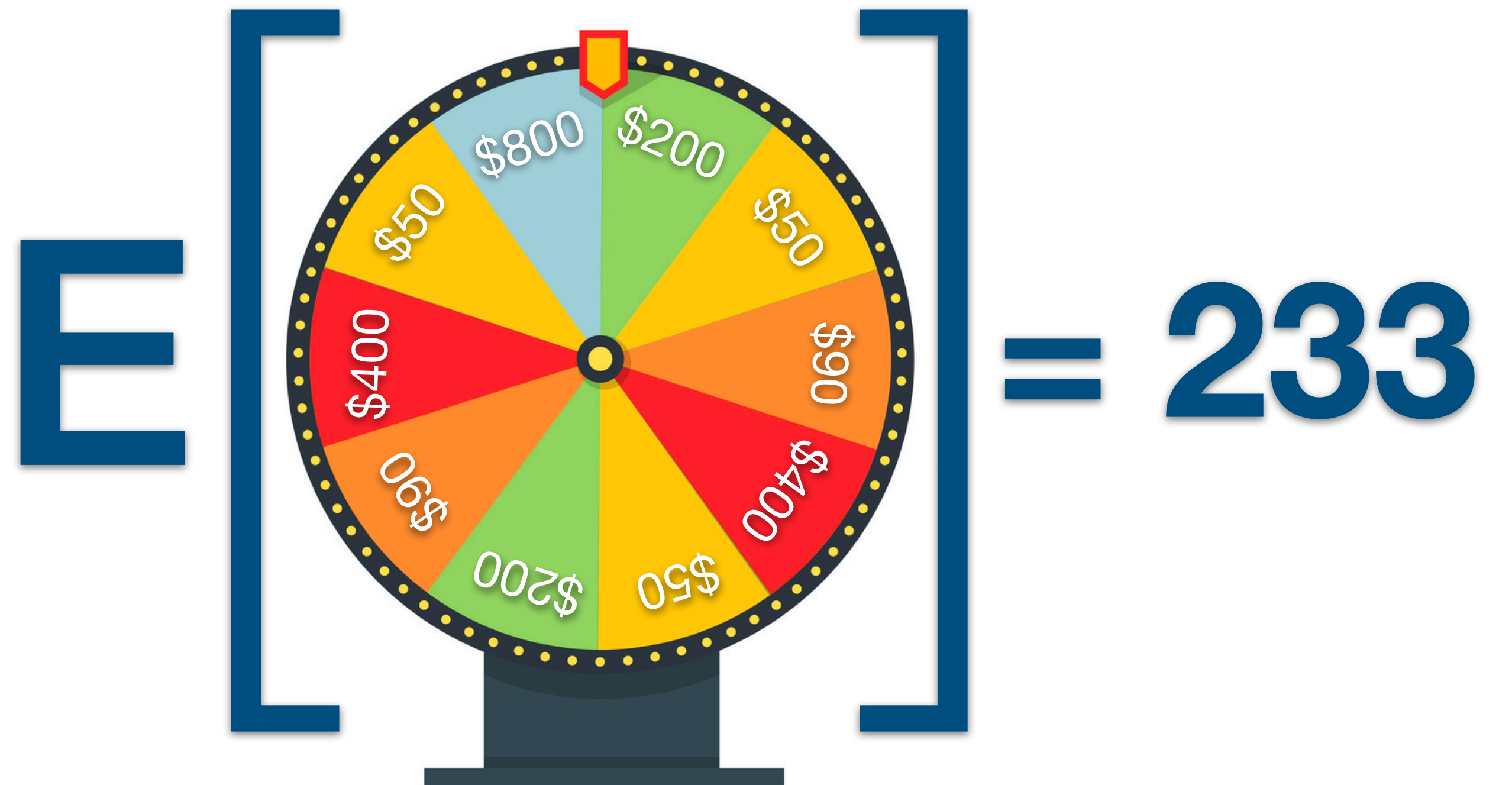
- A sequence of random variables X_0, X_1, X_2, \dots recursively defined by

$$X_0 = 1 \text{ and } X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)}$$

where $\{\xi_j^{(n)} \mid n, j \geq 0\}$ are *i.i.d.* non-negative integer-valued random variables (e.g. Poisson random variables)

- Random family tree: the j th family member in the n th generation has $\xi_j^{(n)}$ offsprings
- X_n : number of family members in the n th generation

Expectation



Expectation (数学期望)

- The expectation (or mean) of a discrete random variable X is defined to be

$$\mathbb{E}[X] = \sum_x xp_X(x)$$

where p_X denotes the *pmf* of X and the sum is taken over all x that $p_X(x) > 0$

- $\mathbb{E}[X]$ may be ∞ (we assume *absolute convergence* for $\mathbb{E}[X] < \infty$)

- **Example I:** $p_X(2^k) = 2^{-k}$ for $k = 1, 2, \dots$ (the St. Petersburg paradox)

- **Example II:** $X \in \mathbb{Z} \setminus \{0\}$ and $p_X(k) = \frac{1}{ak^2}$ where $a = \sum_{k \neq 0} k^{-2} = \frac{\pi^2}{3}$

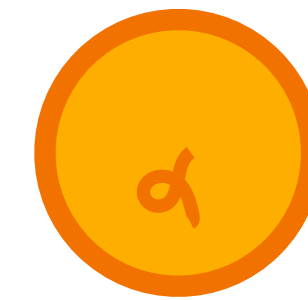
Perspectives of Expectation

- Computation of expectation:
 - straightforward computation (by definition)
 - **linearity of expectation** (by linearity)
 - **law of total expectation** (by case)
- Upper/lower bounds of expectation:
 - **Jensen's inequality** (by convexity)
 - monotonicity (by coupling)
- Implications of expectation:
 - **averaging principle** (the probabilistic method)
 - **tail inequalities** (the moment method)

Expectation of Indicator



p



$1 - p$

- For Bernoulli random variable $X \in \{0,1\}$ with parameter p

$$\mathbb{E}[X] = 0 \cdot (1 - p) + 1 \cdot p = p$$

- For the indicator random variable $X = I(A)$ of event A , where $X = 1$ if A occurs and $X = 0$ if otherwise (i.e. $\forall \omega \in \Omega, X(\omega) = 1$ if $\omega \in A$ and $X(\omega) = 0$ if $\omega \notin A$)

$$\mathbb{E}[X] = 0 \cdot \Pr(A^c) + 1 \cdot \Pr(A) = \Pr(A)$$

Poisson Distribution (泊松分布)

- Expectation of Poisson random variable $X \sim \text{Pois}(\lambda)$

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k \geq 0} k \frac{e^{-\lambda} \lambda^k}{k!} \\ &= \sum_{k \geq 1} \frac{e^{-\lambda} \lambda^k}{(k-1)!} \\ &= \sum_{k \geq 0} \frac{e^{-\lambda} \lambda^{k+1}}{k!} = \lambda \sum_{k \geq 0} \frac{e^{-\lambda} \lambda^k}{k!} \\ &= \lambda\end{aligned}$$

Change of Variables

(Law Of The Unconscious Statistician, *LOTUS*)

- For $f : \mathbb{R} \rightarrow \mathbb{R}$, for discrete X and $\mathbf{X} = (X_1, \dots, X_n)$:
 - $\mathbb{E}[f(X)] = \sum_x f(x)p_X(x)$
 - $\mathbb{E}[f(X_1, \dots, X_n)] = \sum_{(x_1, \dots, x_n)} f(x_1, \dots, x_n)p_X(x_1, \dots, x_n)$

Proof: Let $Y = f(X_1, \dots, X_n)$. Then

$$\begin{aligned}\mathbb{E}[f(X_1, \dots, X_n)] &= \sum_y y \Pr(Y = y) = \sum_y y \sum_{(x_1, \dots, x_n) \in f^{-1}(y)} \Pr((X_1, \dots, X_n) = (x_1, \dots, x_n)) \\ &= \sum_{(x_1, \dots, x_n)} f(x_1, \dots, x_n) \Pr((X_1, \dots, X_n) = (x_1, \dots, x_n)) \\ &= \sum_{(x_1, \dots, x_n)} f(x_1, \dots, x_n) p_X(x_1, \dots, x_n)\end{aligned}$$

Linearity of Expectation

- For $a, b \in \mathbb{R}$ and random variables X and Y :
 - $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
 - $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

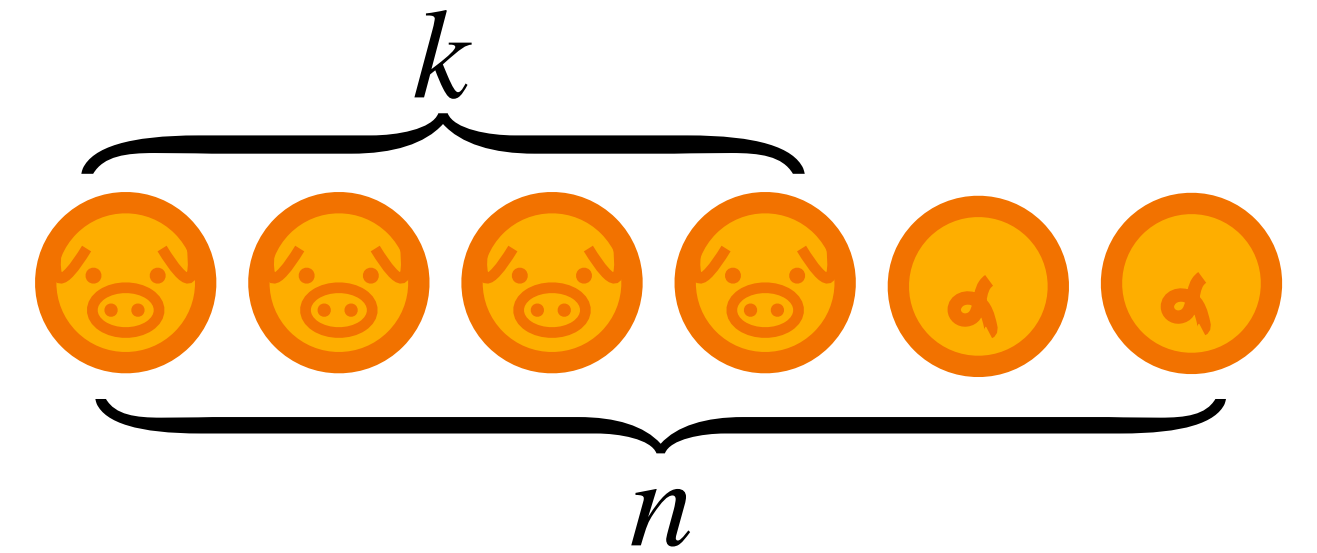
Proof:
$$\mathbb{E}[aX + b] = \sum_x (ax + b)p_X(x) = a \sum_x xp_X(x) + b \sum_x p_X(x) = a\mathbb{E}[X] + b$$

$$\begin{aligned}\mathbb{E}[X + Y] &= \sum_{x,y} (x + y) \Pr((X, Y) = (x, y)) \\ &= \sum_x x \sum_y \Pr((X, Y) = (x, y)) + \sum_y y \sum_x \Pr((X, Y) = (x, y)) \\ &= \sum_x x \Pr(X = x) + \sum_y y \Pr(Y = y) = \mathbb{E}[X] + \mathbb{E}[Y]\end{aligned}$$

Linearity of Expectation

- For $a, b \in \mathbb{R}$ and random variables X and Y :
 - $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
 - $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- For linear (affine) function f on random variables X_1, \dots, X_n
$$\mathbb{E}[f(X_1, \dots, X_n)] = f(\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$$
- It holds for *arbitrarily dependent* X_1, \dots, X_n

Binomial Distribution (二项分布)



- For binomial random variable $X \sim \text{Bin}(n, p)$

$$\mathbb{E}[X] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}$$

- **Observation:** $X \sim \text{Bin}(n, p)$ can be expressed as $X = X_1 + \dots + X_n$, where X_1, \dots, X_n are i.i.d. Bernoulli random variables with parameter p
- **Linearity of expectation:**

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_n] = np$$

Geometric Distribution (几何分布)



$I_k : 1 \quad 1 \quad 1 \quad 1 \quad 0 \dots$

- For geometric random variable $X \sim \text{Geo}(p)$

$$\mathbb{E}[X] = \sum_{k \geq 1} k(1-p)^{k-1}p$$

- **Observation:** $X \sim \text{Geo}(p)$ can be calculated by $X = \sum_{k \geq 1} I_k$,
where $I_k \in \{0,1\}$ *indicates whether all of the first $(k-1)$ trials fail*

- **Linearity of expectation:**

$$\mathbb{E}[X] = \sum_{k \geq 1} \mathbb{E}[I_k] = \sum_{k \geq 1} (1-p)^{k-1} = \frac{1}{p}$$

Negative Binomial Distribution (负二项分布)

- For negative binomial random variable X with parameters r, p

$$\mathbb{E}[X] = \sum_{k \geq 1} k \binom{k+r-1}{k} (1-p)^k p^r$$

- **Observation:** X can be expressed as $X = (X_1 - 1) + \dots + (X_r - 1)$, where X_1, \dots, X_r are i.i.d. geometric random variables with parameter p
- **Linearity of expectation:**

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_r] - r = r(1-p)/p$$

Hypergeometric Distribution (超几何分布)

- For hypergeometric random variable X with parameters N, M, n

$$\mathbb{E}[X] = \sum_{k=0}^n k \binom{M}{k} \binom{N-M}{n-k} / \binom{N}{n}$$

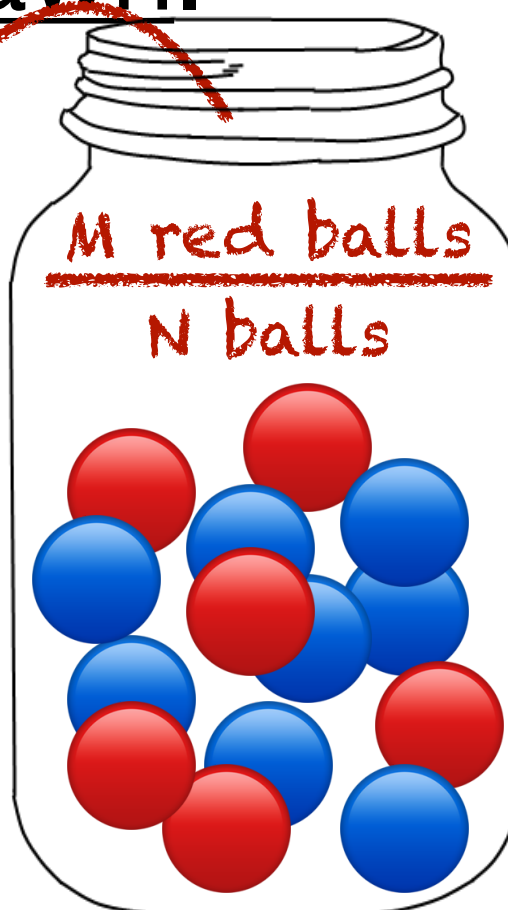
- **Observation:** each red ball (success) is drawn with probability $\binom{N-1}{n-1} / \binom{N}{n} = \frac{n}{N}$.

Then $X = X_1 + \dots + X_M$, where $X_i \in \{0,1\}$ indicates whether the i th red ball is drawn.

- **Linearity of expectation:**

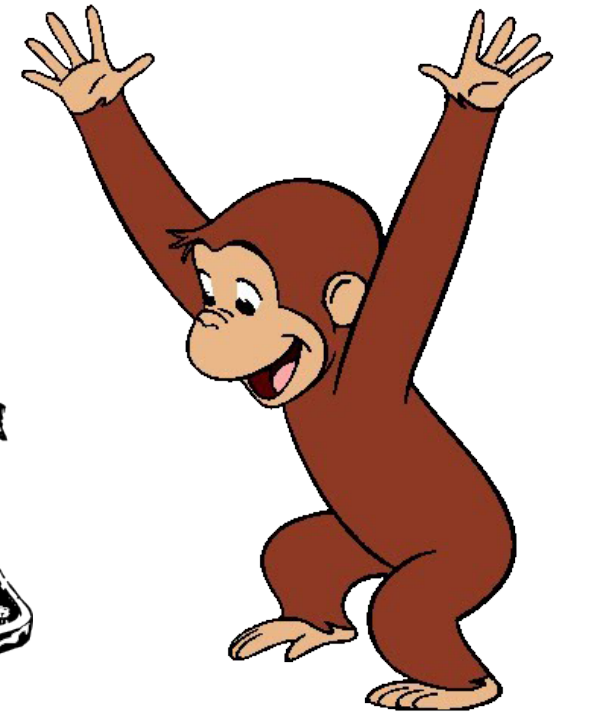
$$\mathbb{E}[X] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_M] = \frac{nM}{N}$$

Draw n balls
without
replacement



Pattern Matching

Hamlet



- $s = (s_1, \dots, s_n) \in Q^n$: uniform random string of n letters from alphabet Q with $|Q| = q$
- For pattern $\pi \in Q^k$, let X be the number of appearances of π in s as substring

• Let $I_i \in \{0, 1\}$ indicate that $\pi = (s_i, s_{i+1}, \dots, s_{i+k-1})$. Then $X = \sum_{i=1}^{n-k+1} I_i$

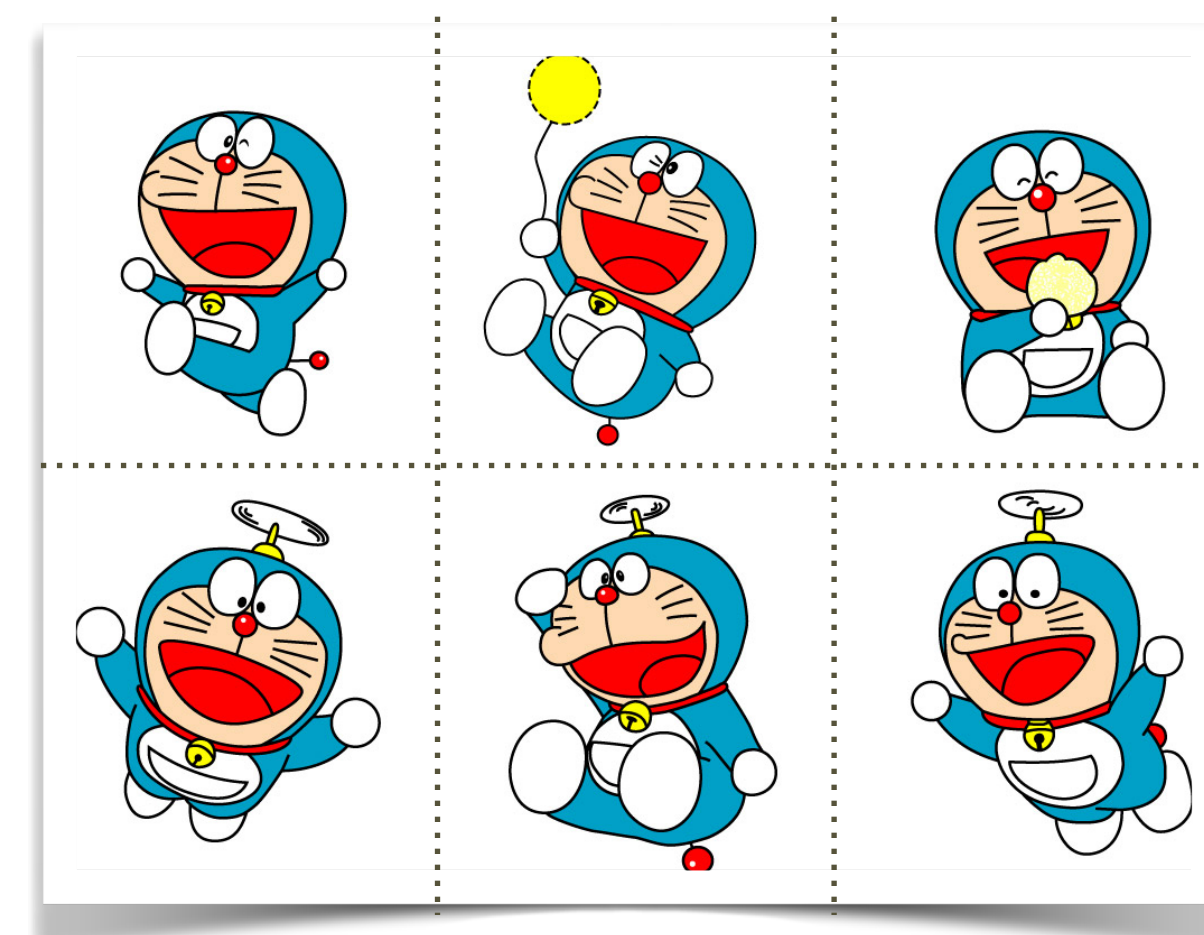
- **Linearity of expectation:**

$$\mathbb{E}[X] = \sum_{i=1}^{n-k+1} \mathbb{E}[I_i] = (n - k + 1)q^{-k}$$

- Expected time (position) for the first appearance? It may depend on the pattern π .

Optional Stopping Theorem (OST)

Coupon Collector



- Each cookie box comes with a uniform random coupon.
 - Number of cookie boxes opened to collect all n types of coupons
- **Balls-into-bins model:** throw balls one-by-one *u.a.r.* to occupy all n bins
 - X : total number of balls thrown to make all n bins nonempty
 - X_i : number of balls thrown while there are exactly $(i - 1)$ nonempty bins

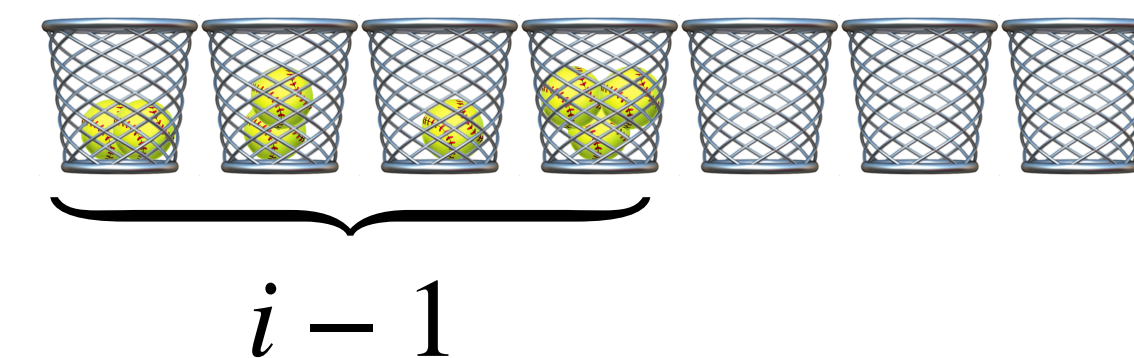
• X_i is geometric with parameter $p_i = 1 - \frac{i-1}{n}$ and $X = \sum_{i=1}^n X_i$



- **Linearity of expectation:**

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n \frac{n}{n-i+1} = n \sum_{i=1}^n \frac{1}{i} = nH(n) \approx n \ln n$$

(Harmonic number)



Double Counting

- For nonnegative random variable X that takes values in $\{0,1,2,\dots\}$

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \Pr[X > k]$$

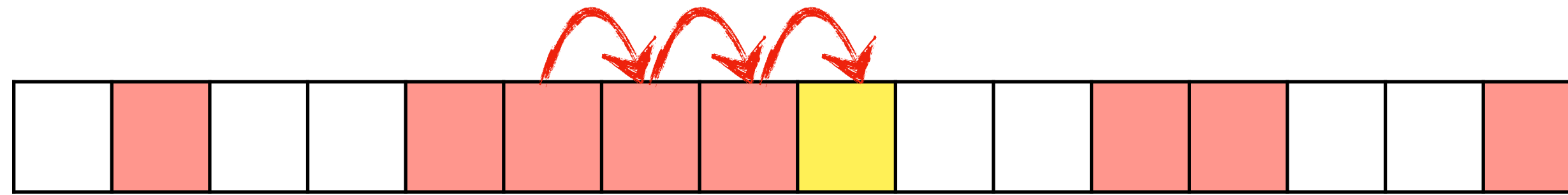
- **Proof I (Double Counting):**

$$\mathbb{E}[X] = \sum_{x \geq 0} x \Pr[X = x] = \sum_{x \geq 0} \sum_{k=0}^{x-1} \Pr[X = x] = \sum_{k \geq 0} \sum_{x > k} \Pr[X = x] = \sum_{k \geq 0} \Pr[X > k]$$

- **Proof II (Linearity of Expectation):** Let $I_k \in \{0,1\}$ indicate whether $X > k$.

Then $X = \sum_{k \geq 0} I_k$. By linearity, $\mathbb{E}[X] = \sum_{k \geq 0} \mathbb{E}[I_k] = \sum_{k \geq 0} \Pr[X > k]$

Open Addressing with Uniform Hashing



- Hash table: n keys from a universe U are mapped to m slots by hash function

$$h : U \rightarrow [m]$$

- Open addressing (开放寻址): hash collision is resolved by a probing strategy
 - when searching for a key $x \in U$, the i th probed slot is given by $h(x, i)$
- Linear probing: $h(x, i) = h(x) + i \pmod{m}$
- Quadratic probing: $h(x, i) = h(x) + c_1 i + c_2 i^2 \pmod{m}$
- Double hashing: $h(x, i) = h_1(x) + i \cdot h_2(x) \pmod{m}$
- Uniform hashing: $h(x, i) = \pi(i)$ where π is a uniform random permutation of $[m]$

Open Addressing with Uniform Hashing

- In a hash table with load factor $\alpha = n/m$, assuming uniform hashing, the expected number of probes in an unsuccessful search is at most $1/(1 - \alpha)$.
- **Proof:** Let X be the number of probes in an unsuccessful search.

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{\infty} \Pr(X > k) = 1 + \sum_{k=1}^{\infty} \Pr(X > k) \\ &= 1 + \sum_{k=1}^{\infty} \Pr\left(\bigcap_{i=1}^k A_i\right) \quad (\text{where } A_i \text{ is the event that the } i\text{th probed slot is occupied}) \\ &= 1 + \sum_{k=1}^{\infty} \prod_{i=1}^k \Pr\left(A_i \mid \bigcap_{j<i} A_j\right) \quad (\text{by chain rule}) \\ &= 1 + \sum_{k=1}^{\infty} \prod_{i=1}^k \frac{n - i + 1}{m - i + 1} \leq 1 + \sum_{k=1}^{\infty} \prod_{i=1}^k \frac{n}{m} = 1 + \sum_{k=1}^{\infty} \alpha^k = \sum_{k=0}^{\infty} \alpha^k = \frac{1}{1 - \alpha}\end{aligned}$$

Principle of Inclusion-Exclusion

- Let $I(A) \in \{0,1\}$ be the indicator random variable of event A . It's easy to verify:

★ $I(A^c) = 1 - I(A)$

♣ $I(A \cap B) = I(A) \cdot I(B)$

- For events A_1, A_2, \dots, A_n :

$$I\left(\bigcup_{i=1}^n A_i\right) \stackrel{(\star)}{=} 1 - I\left(\left(\bigcup_{i=1}^n A_i\right)^c\right) \stackrel{\text{(De Morgan's law)}}{=} 1 - I\left(\bigcap_{i=1}^n A_i^c\right) \stackrel{(\clubsuit)}{=} 1 - \prod_{i=1}^n I(A_i^c) \stackrel{(\star)}{=} 1 - \prod_{i=1}^n (1 - I(A_i))$$

$$\stackrel{\text{(binomial theorem)}}{=} 1 - \sum_{S \subseteq \{1, \dots, n\}} (-1)^{|S|} \prod_{i \in S} I(A_i) \stackrel{(\clubsuit)}{=} \sum_{\emptyset \neq S \subseteq \{1, \dots, n\}} (-1)^{|S|-1} I\left(\bigcap_{i \in S} A_i\right)$$

Principle of Inclusion-Exclusion

- Let $I(A) \in \{0,1\}$ be the indicator random variable of event A .
- For events A_1, A_2, \dots, A_n :

$$I\left(\bigcup_{i=1}^n A_i\right) = \sum_{\emptyset \neq S \subseteq \{1, \dots, n\}} (-1)^{|S|-1} I\left(\bigcap_{i \in S} A_i\right)$$

- By linearity of expectation:

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

Boole-Bonferroni Inequality

- For events A_1, A_2, \dots, A_n :

$$I\left(\bigcup_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n (1 - I(A_i)) = \sum_{k=1}^n (-1)^{k-1} \sum_{S \in \binom{\{1, \dots, n\}}{k}} I\left(\bigcap_{i \in S} A_i\right)$$

- **Observation:** $X_k \triangleq \binom{\sum_{i=1}^n I(A_i)}{k} = \sum_{S \in \binom{\{1, \dots, n\}}{k}} \prod_{i \in S} I(A_i) = \sum_{S \in \binom{\{1, \dots, n\}}{k}} I\left(\bigcap_{i \in S} A_i\right)$

and X_k as a binomial coefficient is *unimodal* in k

- For unimodal sequence X_k : $\sum_{k \leq 2t} (-1)^{k-1} X_k \leq \sum_{k=1}^n (-1)^{k-1} X_k \leq \sum_{k \leq 2t+1} (-1)^{k-1} X_k$

- Take expectation. By linearity of expectation \implies Bonferroni inequality

Limitation of Linearity

- Infinite sum: X_1, X_2, \dots

$$\mathbb{E} \left[\sum_{i=1}^{\infty} X_i \right] = \sum_{i=1}^{\infty} \mathbb{E}[X_i] \text{ if the } \textit{absolute convergence} \sum_{i=1}^{\infty} \mathbb{E}[|X_i|] < \infty \text{ holds}$$

This is possible: $\mathbb{E} \left[\sum_{i=1}^{\infty} X_i \right] < \infty$ and $\sum_{i=1}^{\infty} \mathbb{E}[X_i] < \infty$ but $\mathbb{E} \left[\sum_{i=1}^{\infty} X_i \right] \neq \sum_{i=1}^{\infty} \mathbb{E}[X_i]$

Counterexample: the **martingale** betting strategy in a fair gambling game

- A random number of random variables: X_1, X_2, \dots, X_N for random N

$$\mathbb{E} \left[\sum_{i=1}^N X_i \right] = \mathbb{E}[N] \mathbb{E}[X_1] \text{ ?}$$

Conditional Expectation (条件期望)

- The conditional expectation of a discrete random variable X given that event A occurs, is defined by

$$\mathbb{E}[X | A] = \sum_x x \Pr(X = x | A)$$

where the sum is taken over all x that $\Pr(X = x | A) > 0$

- To be well-defined, assume:
 - $\Pr(A) > 0$
 - the sum $\sum_x x \Pr(X = x | A)$ converges absolutely

Conditional Distribution (条件分布)

- The probability mass function $p_{X|A} : \mathbb{Z} \rightarrow [0,1]$ of a discrete random variable X given that event A occurs, is given by

$$p_{X|A}(x) = \Pr(X = x | A)$$

- $(X | A)$ can now be seen as a well-defined discrete random variable, whose distribution is described by the *pmf* $p_{X|A}$

- $\mathbb{E}[X | A] = \sum_x x \Pr(X = x | A)$ is just the expectation of $(X | A)$

- $\mathbb{E}[X | A]$ satisfies the properties of expectation, e.g. **linearity of expectation**

Law of Total Expectation

- Let X be a discrete random variable with finite $\mathbb{E}[X]$. Let events B_1, B_2, \dots, B_n be a partition of Ω such that $\Pr(B_i) > 0$ for all i .

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X \mid B_i] \Pr(B_i)$$

- The law of total probability is now a special case with $X = I(A)$

Proof: $\mathbb{E}[X] = \sum_x x \Pr(X = x) = \sum_x x \sum_{i=1}^n \Pr(X = x \mid B_i) \Pr(B_i)$ (law of total prob.)

$$= \sum_{i=1}^n \Pr(B_i) \sum_x x \Pr(X = x \mid B_i) = \sum_{i=1}^n \mathbb{E}[X \mid B_i] \Pr(B_i)$$

Analysis of QuickSort

- A *comparison-based* sorting algorithm
 - worst-case complexity: $O(n^2)$
 - average-case complexity: ? $t(n) = O(n \ln n)$ **verified by induction**

- Let $t(n) = \mathbb{E}[X_n]$, where X_n is the number of comparisons used in QSort(A) on a uniform random permutation A of n distinct numbers.
- **Law of total expectation:** Let B_i be the event that $A[1]$ is the i th smallest in A .

$$t(n) = \mathbb{E}[X_n] = \sum_{i=1}^n \mathbb{E}[X_n \mid B_i] \Pr(B_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[n - 1 + X_{i-1} + X_{n-i}] = n - 1 + \frac{2}{n} \sum_{i=0}^{n-1} t(i)$$

$$t(0) = t(1) = 0$$

QSort(A): an array A of n distinct entries

If $n > 1$ then do:

choose a pivot $x = A[1]$;

partition A into L with all entries $< x$,
and R with all entries $> x$;

QSort(L) and QSort(R);

Analysis of QuickSort

QSort(A): an array A of n distinct entries

If $n > 1$ then do:

choose a pivot $x = A[1]$;

partition A into L with all entries $< x$,
and R with all entries $> x$;

QSort(L) and QSort(R);

- Uniform random input:

- A is a uniform random permutation of $a_1 < \dots < a_n$

- Let $X_{ij} \in \{0,1\}$ indicate whether a_i and a_j are compared within QSort(A).

- **Observation I:** each pair of a_i, a_j are compared at most once.

$$\implies \text{total number of comparisons is } X = \sum_{i < j} X_{ij}$$

- **Observation II:** if a_i, a_j are still in the same array, then so are all a_k for $i < k < j$.

a_i, a_j are compared iff one of them is chosen as pivot when they are in the same array.

$$\implies \mathbb{E}[X_{ij}] = \Pr(a_i, a_j \text{ are compared}) = \Pr(\{a_i, a_j\} \mid \{a_i, a_{i+1}, \dots, a_j\}) = \frac{2}{j - i + 1}$$

- **Linearity of expectation:**

$$\mathbb{E}[X] = \sum_{i < j} \mathbb{E}[X_{ij}] = \sum_{i < j} \frac{2}{j - i + 1} = \sum_{i=1}^n \sum_{k=2}^{n-i+1} \frac{2}{k} \leq 2 \sum_{i=1}^n \sum_{k=1}^n \frac{1}{k} = 2nH(n) = 2n \ln n + O(n)$$

Conditional Expectation (条件期望)

$Y \backslash X$	x_1	x_2	x_3	x_4	$p_Y(y) \downarrow$
y_1	$\frac{4}{32}$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{8}{32}$
y_2	$\frac{3}{32}$	$\frac{6}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{15}{32}$
y_3	$\frac{9}{32}$	0	0	0	$\frac{9}{32}$
$p_X(x) \rightarrow$	$\frac{16}{32}$	$\frac{8}{32}$	$\frac{4}{32}$	$\frac{4}{32}$	$\frac{32}{32}$

- For random variables X, Y , the conditional expectation:

$$\mathbb{E}[X | Y]$$

is a random variable $f(Y)$ whose value is $f(y) = \mathbb{E}[X | Y = y]$ when $Y = y$

- Naturally generalized to $\mathbb{E}[X | Y, Z]$ for random variables X, Y, Z
- **Examples:**
 - $\mathbb{E}[X | Y]$: average height of the country of a random person on earth
 - $\mathbb{E}[X | Y, Z]$: average height of the gender of the country of a random person

Conditional Expectation (条件期望)

$Y \backslash X$	x_1	x_2	x_3	x_4	$p_Y(y) \downarrow$
y_1	$\frac{4}{32}$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{8}{32}$
y_2	$\frac{3}{32}$	$\frac{6}{32}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{15}{32}$
y_3	$\frac{9}{32}$	0	0	0	$\frac{9}{32}$
$p_X(x) \rightarrow$	$\frac{16}{32}$	$\frac{8}{32}$	$\frac{4}{32}$	$\frac{4}{32}$	$\frac{32}{32}$

- For random variables X, Y , the conditional expectation:

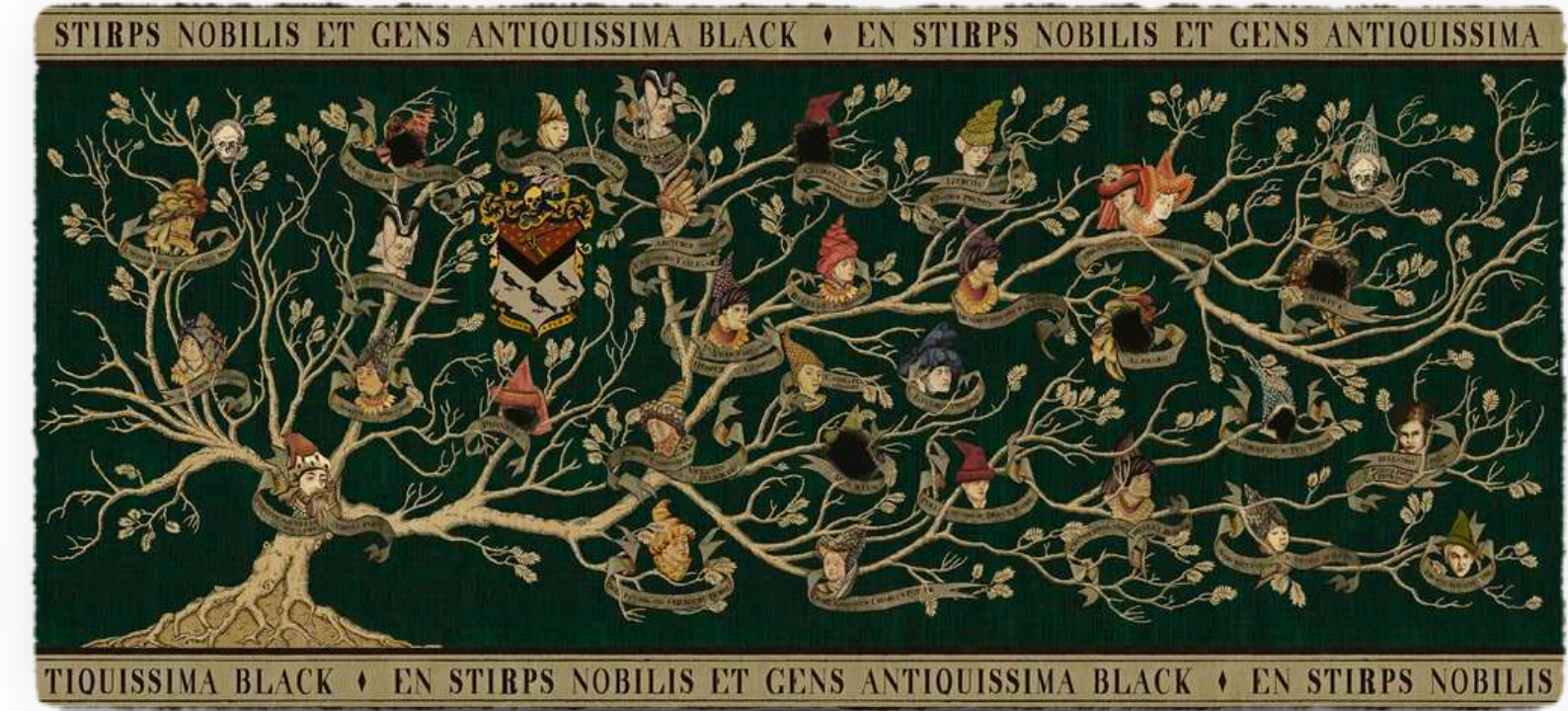
$$\mathbb{E}[X | Y]$$

is a random variable $f(Y)$ whose value is $f(y) = \mathbb{E}[X | Y = y]$ when $Y = y$

- Law of Total Expectation:** $\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X]$

- Proof:**
$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | Y]] &= \sum_y \mathbb{E}[X | Y = y] \Pr(Y = y) && \text{(by definition)} \\ &= \mathbb{E}[X] && \text{(law of total expectation)} \end{aligned}$$

Random Family Tree



- X_0, X_1, X_2, \dots is defined by $X_0 = 1$ and $X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)}$
where $\xi_j^{(n)} \in \mathbb{Z}_{\geq 0}$ are *i.i.d.* random variables with mean value $\mu = \mathbb{E}[\xi_j^{(n)}]$
- $X_0 = 1$ and $\mathbb{E}[X_1] = \mathbb{E}[\xi_1^{(0)}] = \mu$
- $\mathbb{E}[X_n | X_{n-1} = k] = \mathbb{E}\left[\sum_{j=1}^k \xi_j^{(n-1)} \mid X_{n-1} = k\right] = k\mu \implies \mathbb{E}[X_n | X_{n-1}] = X_{n-1}\mu$
- $\mathbb{E}[X_n] = \mathbb{E}[\mathbb{E}[X_n | X_{n-1}]] = \mathbb{E}[X_{n-1}\mu] = \mathbb{E}[X_{n-1}] \cdot \mu = \mu^n$
 $\implies \mathbb{E}\left[\sum_{n \geq 0} X_n\right] = \sum_{n \geq 0} \mathbb{E}[X_n] = \sum_{n \geq 0} \mu^n = \begin{cases} \frac{1}{1-\mu} & \text{if } 0 < \mu < 1 \\ \infty & \text{if } \mu \geq 1 \end{cases}$

Jensen's Inequality

- For general (non-linear) function $f(X)$ of random variable X

we don't have $\mathbb{E}[f(X)] = f(\mathbb{E}[X])$

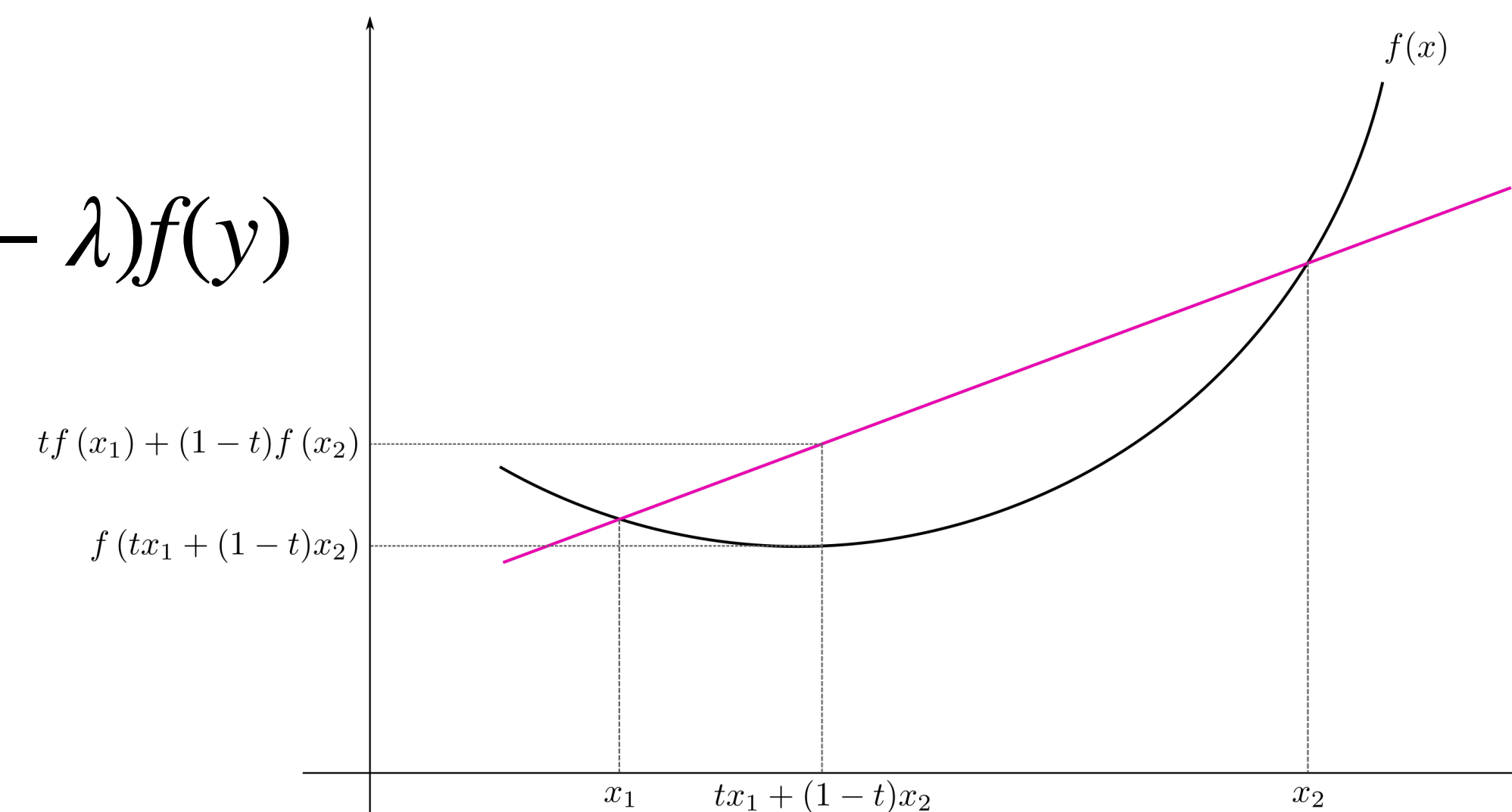
- But if the convexity of f is known, then the **Jensen's inequality** applies:

- f is **convex** $\iff f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$

$$\implies \mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$$

- f is **concave** $\iff f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$

$$\implies \mathbb{E}[f(X)] \leq f(\mathbb{E}[X])$$



Monotonicity of Expectation

- For random variables X and Y , for $c \in \mathbb{R}$:
(Y stochastically dominates X)
 - If $X \leq Y$ *a.s.* (almost surely, i.e. $\Pr(X \leq Y) = 1$), then $\mathbb{E}[X] \leq \mathbb{E}[Y]$
 - If $X \leq c$ ($X \geq c$) *a.s.*, then $\mathbb{E}[X] \leq c$ ($\mathbb{E}[X] \geq c$)
 - $\mathbb{E}[|X|] \geq |\mathbb{E}[X]| \geq 0$

Proof:

$$\begin{aligned}\mathbb{E}[X] &= \sum_x x \Pr(X = x) = \sum_x x \sum_y \Pr((X, Y) = (x, y)) \\ &= \sum_x x \sum_{y \geq x} \Pr((X, Y) = (x, y)) = \sum_y \sum_{x \leq y} x \Pr((X, Y) = (x, y)) \\ &\leq \sum_y \sum_{x \leq y} y \Pr((X, Y) = (x, y)) \leq \sum_y y \Pr(Y = y) = \mathbb{E}[Y]\end{aligned}$$

Averaging Principle

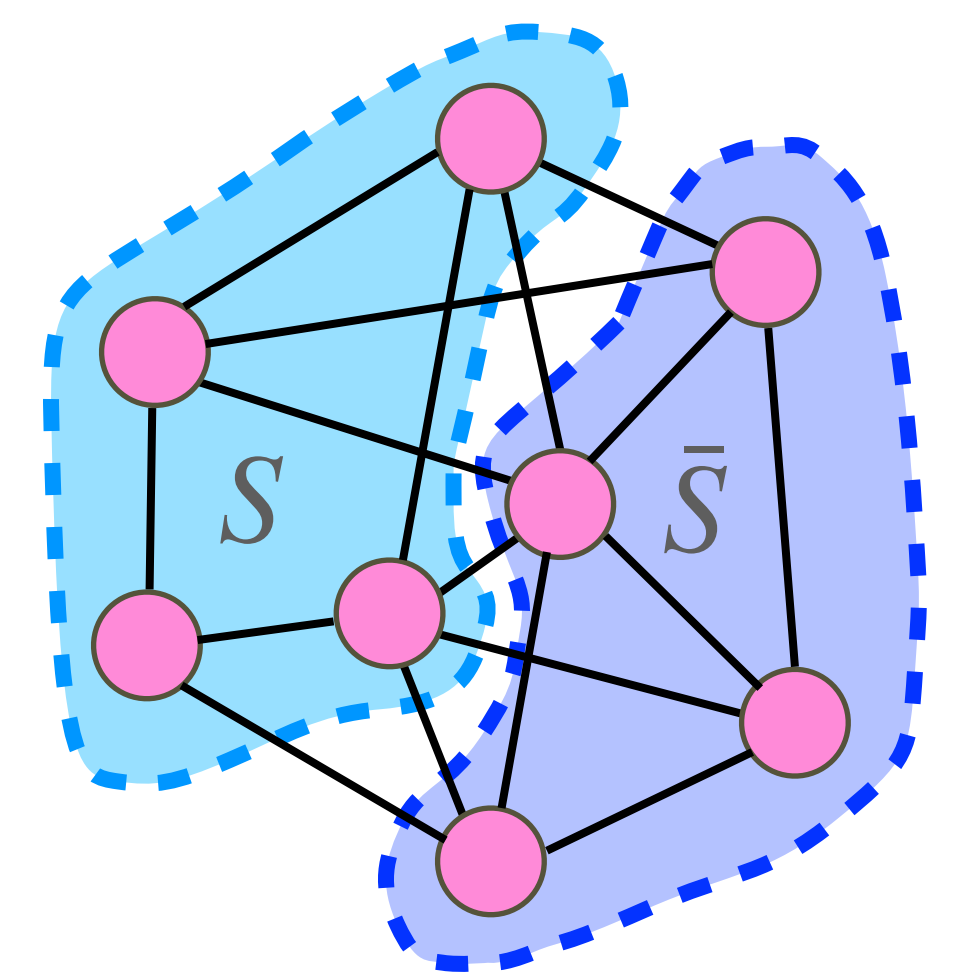
- $\Pr(X \geq \mathbb{E}[X]) > 0 \iff \text{if } \Pr(X < c) = 1 \text{ then } \mathbb{E}[X] < c$
- $\Pr(X \leq \mathbb{E}[X]) > 0 \iff \text{if } \Pr(X > c) = 1 \text{ then } \mathbb{E}[X] > c$
- By the Probabilistic Method:

$\exists \omega \in \Omega$ such that $X(\omega) \geq \mathbb{E}[X]$

$\exists \omega \in \Omega$ such that $X(\omega) \leq \mathbb{E}[X]$



Maximum Cut



- For an undirected graph $G(V, E)$:
 - Find an $S \subseteq V$ with largest **cut** $\delta S \triangleq \{\{u, v\} \in E \mid u \in S \wedge v \notin S\}$
- **NP-hard** problem (very unlikely to have efficient algorithms)

The average cut generated by pairwise independent bits is $\geq |E|/2$.

Proposition: There always exists a large enough cut of size $|\delta S| \geq |E|/2$.

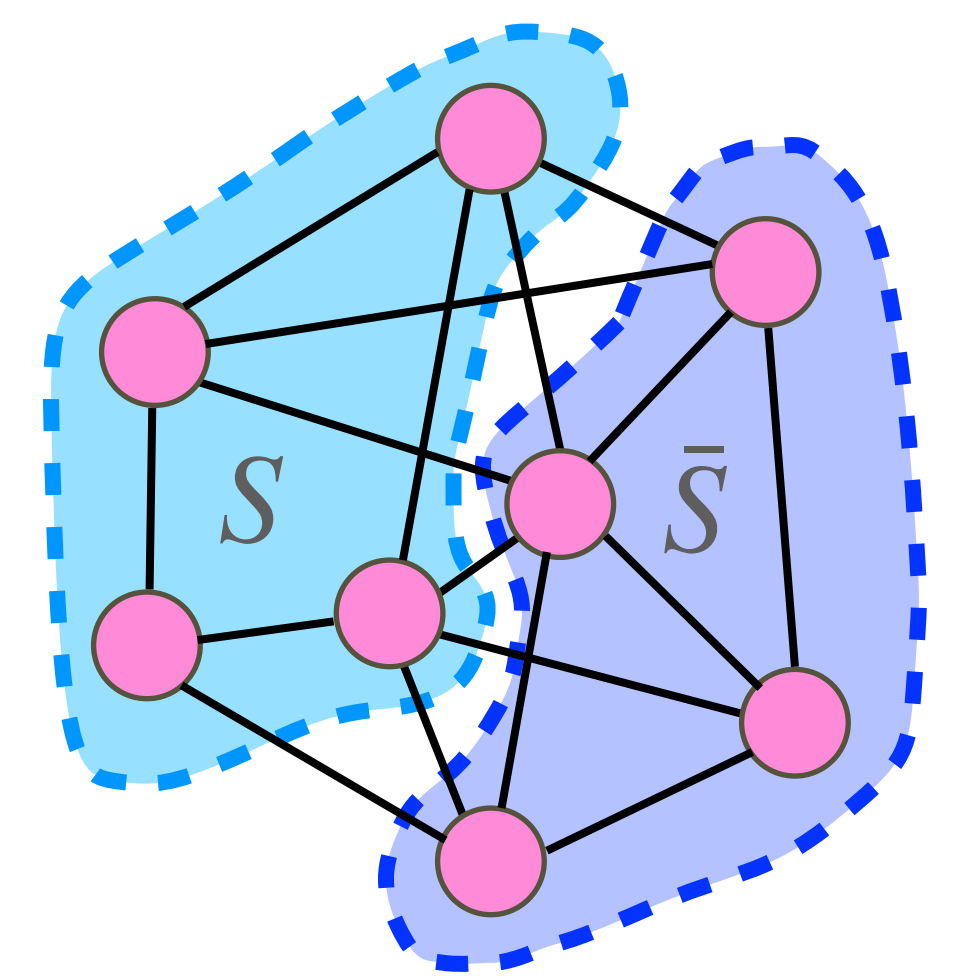
Proof: Let $Y_v \in \{0, 1\}$, for $v \in V$, be pairwise mutually independent uniform random bits.

Each $v \in V$ joins S iff $Y_v = 1$. Then it holds that $|\delta S| = \sum_{\{u, v\} \in E} I(Y_u \neq Y_v)$.

By linearity of expectation: $\mathbb{E}[|\delta S|] = \sum_{\{u, v\} \in E} \Pr(Y_u \neq Y_v) = |E|/2$.

Due to the probabilistic method: There exists such $S \subseteq V$ with $|\delta S| \geq |E|/2$.

Maximum Cut



- For an undirected graph $G(V, E)$:
 - Find an $S \subseteq V$ with largest cut $\delta S \triangleq \{ \{u, v\} \in E \mid u \in S \wedge v \notin S \}$
- **NP-hard** problem (very unlikely to have efficient algorithms)

Parity Search:

for all $\mathbf{b} \in \{0, 1\}^{\lceil \log_2(n+1) \rceil}$:

initialize $S_{\mathbf{b}} = \emptyset$;

for $i = 1, 2, \dots, n$:

if $\bigoplus_{j: \lfloor i/2^j \rfloor \bmod 2 = 1} b_j = 1$ then v_i joins $S_{\mathbf{b}}$;

return the $S_{\mathbf{b}}$ with the largest cut $\delta S_{\mathbf{b}}$;

Guarantees to return an $S \subseteq V$ with $|\delta S| \geq |E|/2$.