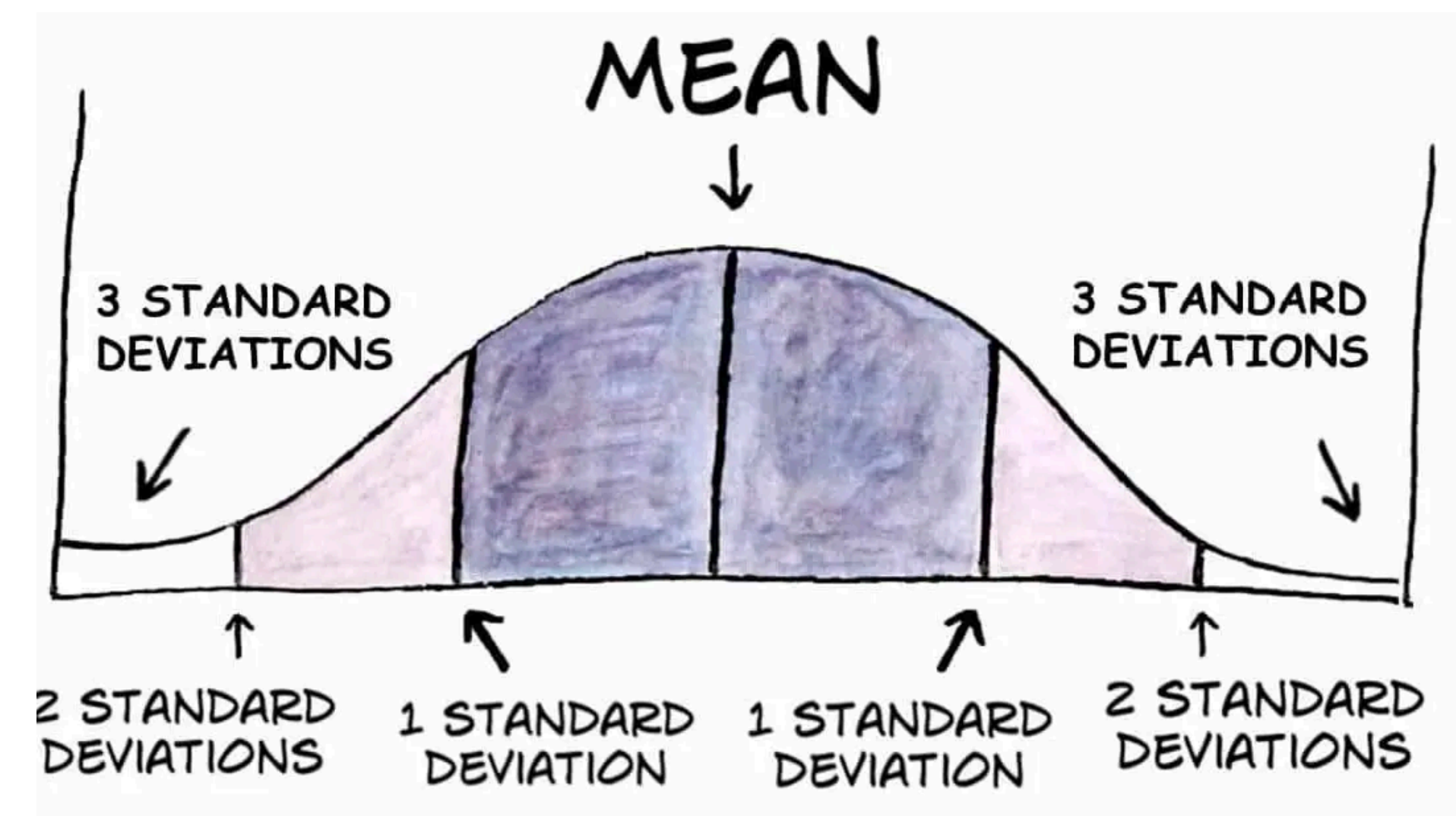


Probability Theory & Mathematical Statistics

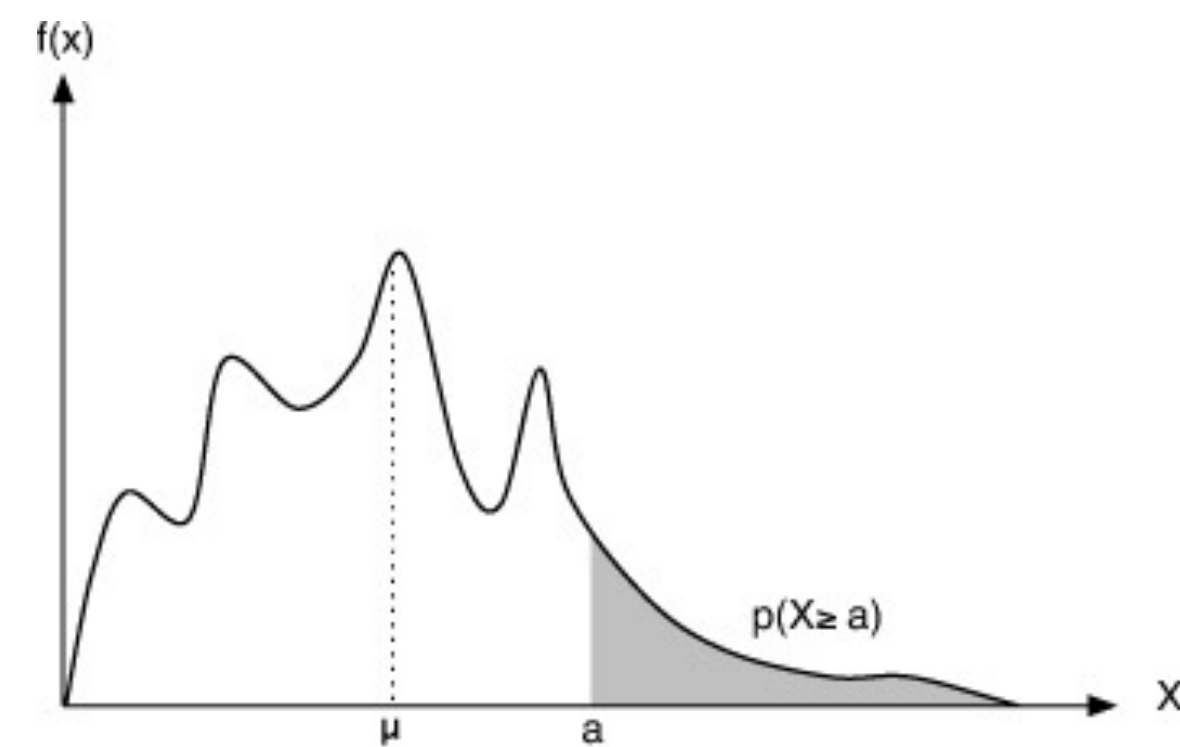
Moment and Deviation

Moments and Deviations



Markov's Inequality

(马尔可夫不等式)



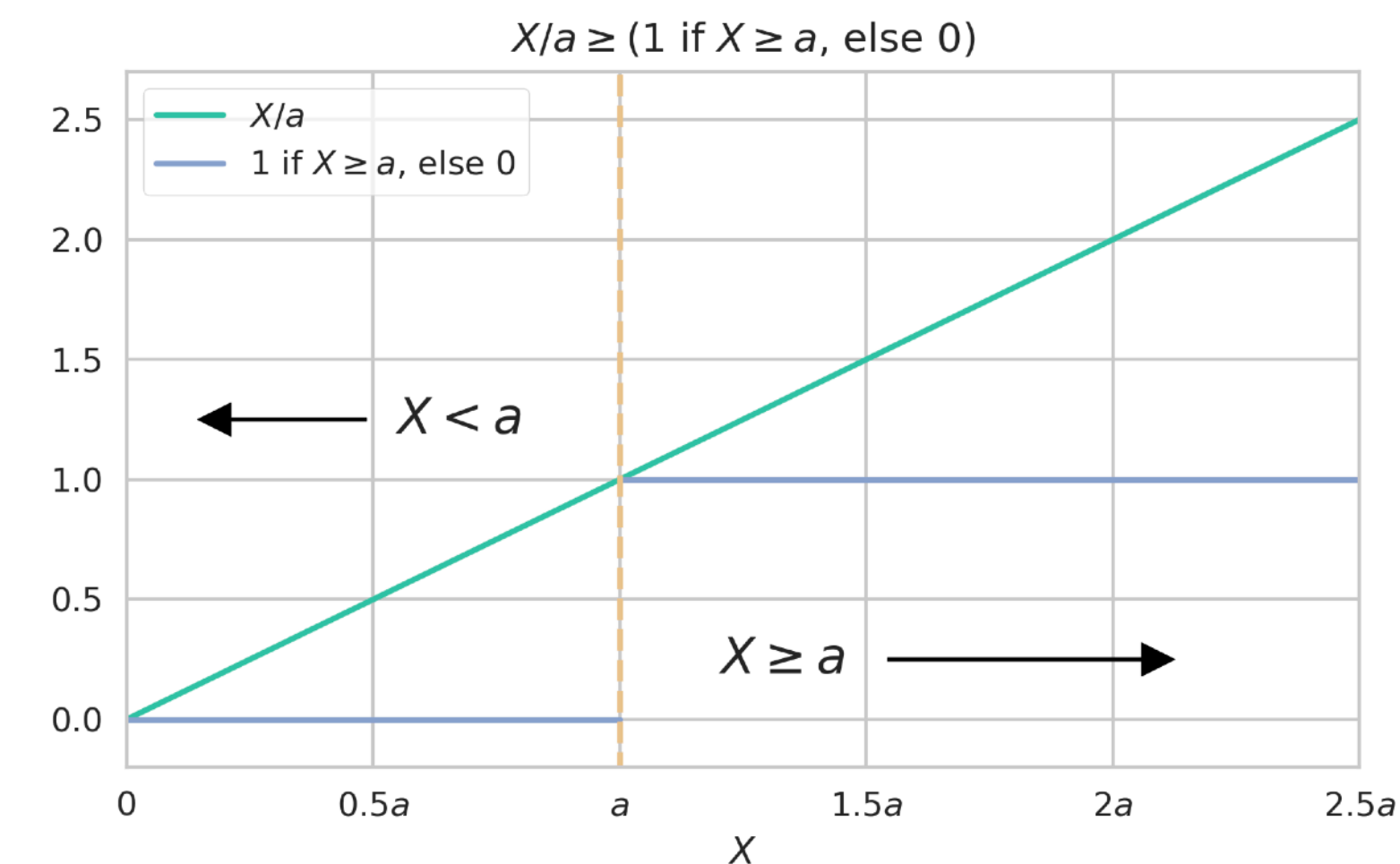
- Markov's inequality: Let X be a *nonnegative-valued* random variable. Then,

$$\text{for any } a > 0, \quad \Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

- **Proof** (by indicator): Let $I = I(X \geq a)$. Since $X \geq 0$ and $a > 0$, we have

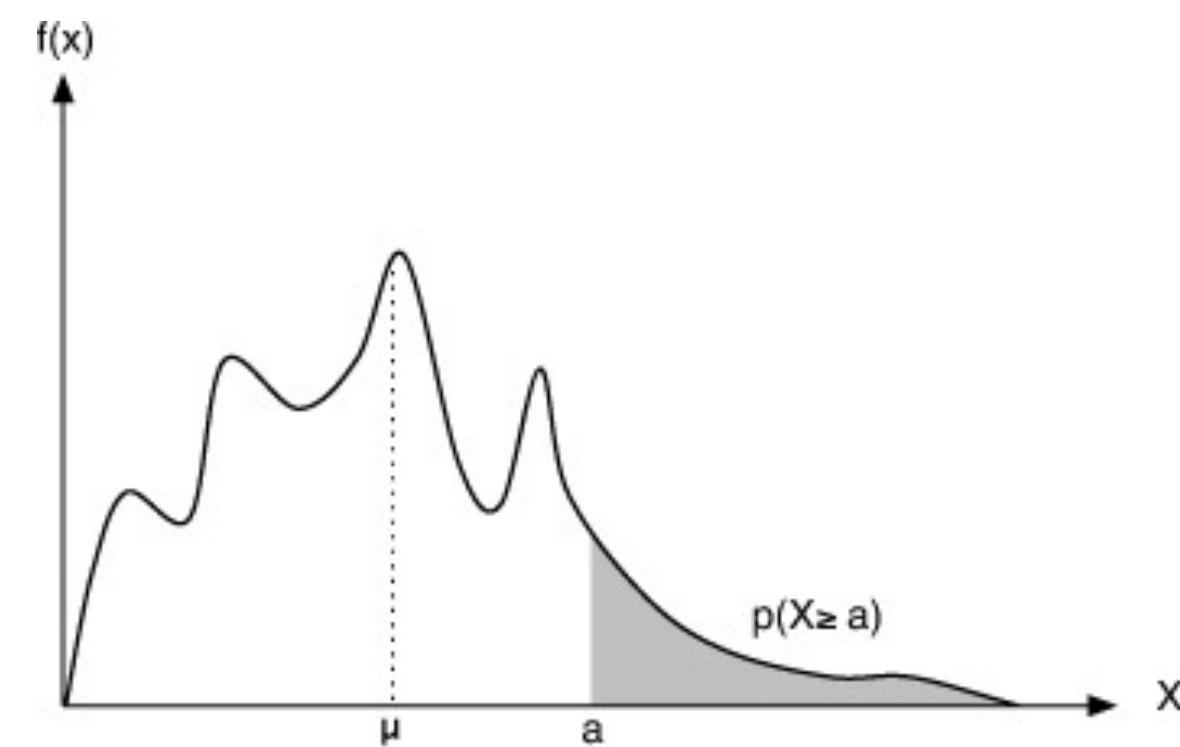
$$I = I(X \geq a) \leq \left\lfloor \frac{X}{a} \right\rfloor \leq \frac{X}{a}.$$

$$\text{Therefore, } \Pr(X \geq a) = \mathbb{E}[I] \leq \mathbb{E} \left[\frac{X}{a} \right] = \frac{\mathbb{E}[X]}{a}$$



Markov's Inequality

(马尔可夫不等式)



- Markov's inequality: Let X be a *nonnegative-valued* random variable. Then,

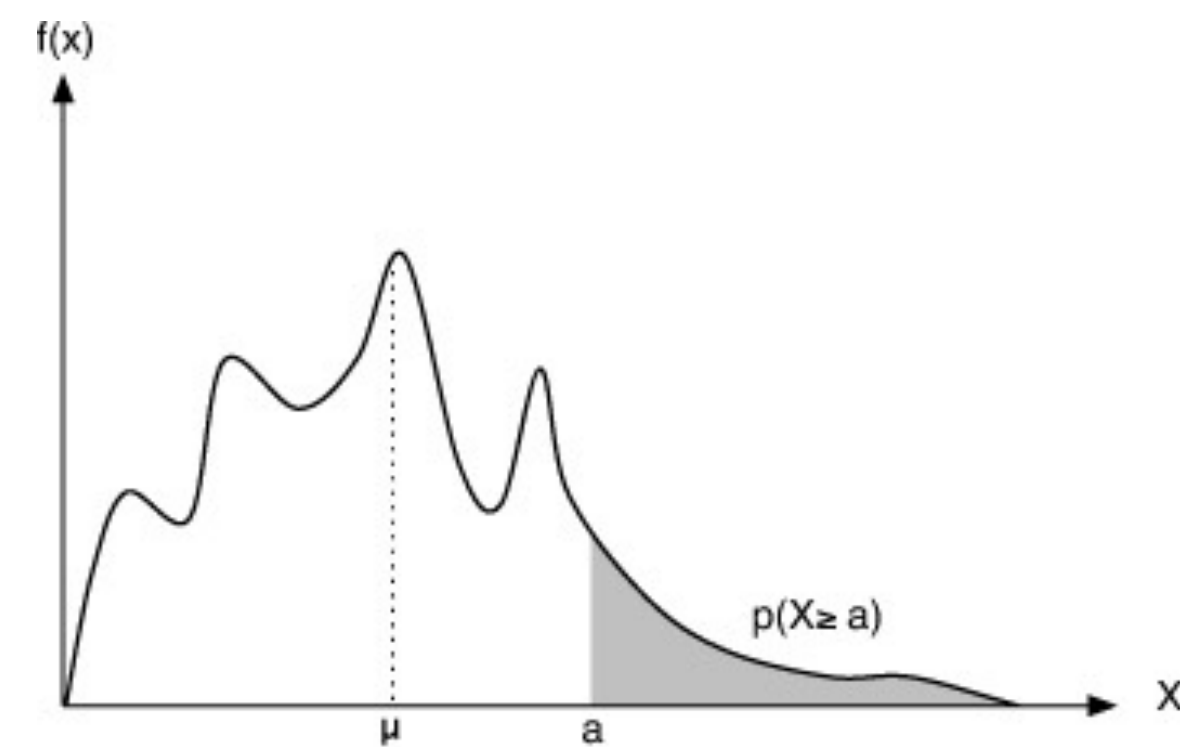
$$\text{for any } a > 0, \quad \Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

- **Proof** (by total expectation):

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X \mid X \geq a] \cdot \Pr(X \geq a) + \mathbb{E}[X \mid X < a] \cdot \Pr(X < a) \\ &\geq a \cdot \Pr(X \geq a) + 0 \cdot \Pr(X < a) = a \cdot \Pr(X \geq a) \\ \implies \Pr(X \geq a) &\leq \frac{\mathbb{E}[X]}{a} \end{aligned}$$

Markov's Inequality

(马尔可夫不等式)



- Markov's inequality: Let X be a *nonnegative-valued* random variable. Then,

$$\text{for any } a > 0, \quad \Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

- **Corollary**: for any $c > 1$, $\Pr(X \geq c\mathbb{E}[X]) \leq 1/c$
- **Tight in the worst case**: $\forall c > 1, \forall \mu \in \mathbb{R}, \exists$ nonnegative X with $\mathbb{E}[X] = \mu$, such that $\Pr(X \geq c\mu) = 1/c$
- **Lower tail variant** (sometimes called reverse Markov's inequality):
 $\Pr(X \leq a) \leq (u - \mathbb{E}[X]) / (u - a)$ requires X to have bounded range $X \leq u$

From Las Vegas to Monte Carlo



- Monte Carlo algorithm: randomized algorithms that are correct by chance



- Las Vegas algorithm: randomized algorithms that always give correct result upon termination (but may run for a random period of time before termination)
- If there is a Las Vegas algorithm \mathcal{A} with expected running time at most $t(n)$ for any input of size n (\mathcal{A} has worst-case expected time complexity $t(n)$):

Algorithm \mathcal{B} :

```
simulate algorithm  $\mathcal{A}$  up to  $\lceil t(n)/\epsilon \rceil$  steps;  
if algorithm  $\mathcal{A}$  terminates  
    return the output of  $\mathcal{A}$ ;  
else return an arbitrary answer;
```

- Algorithm \mathcal{B} is a Monte Carlo algorithm s.t.
 - \mathcal{B} has worst-case running time $\leq \lceil t(n)/\epsilon \rceil$
 - \mathcal{B} is correct with probability at least $1 - \epsilon$
(by Markov inequality)

Cliques in Random Graph

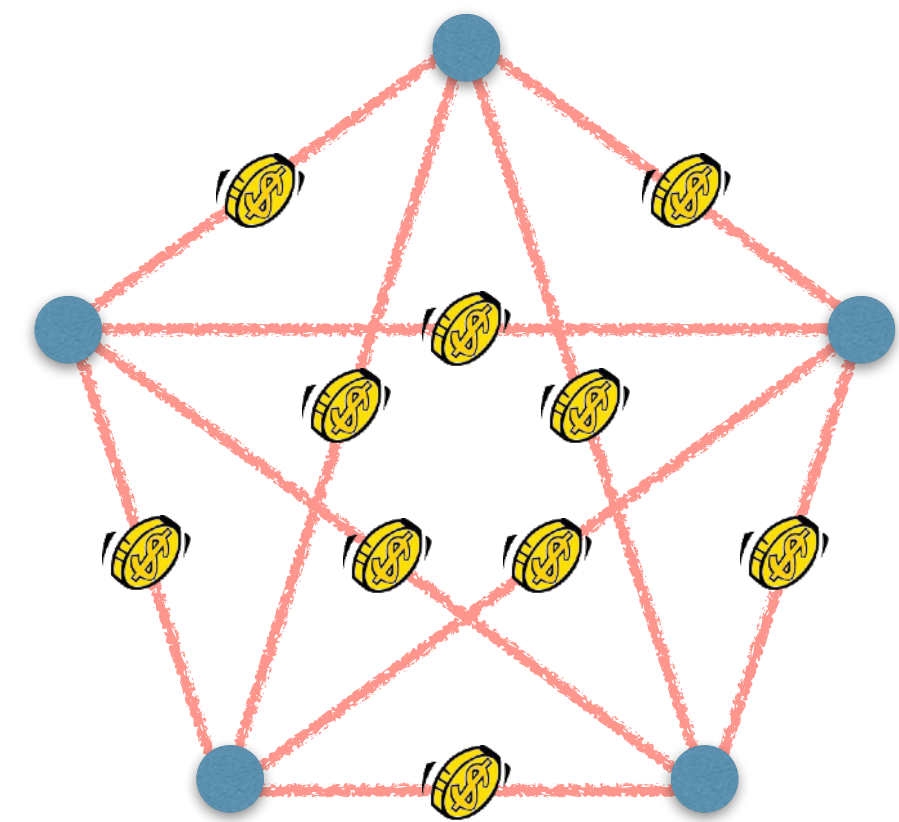
- $G(n, p)$: between every pair u, v among n vertices, an edge is added i.i.d. with prob. p
- Fix a constant integer $k \geq 3$. Let X be the number of k -cliques (K_k) in $G \sim G(n, p)$.
- For every distinct $S \subseteq \in [n]$ of size $|S| = k$, let $I_S = I(K_S \subseteq G)$. Then:

- $\mathbb{E}[I_S] = \Pr(K_S \subseteq G) = p^{\binom{k}{2}}$

- $X = \sum_{S \in \binom{[n]}{k}} I_S$

- **Linearity of expectation:** $\mathbb{E}[X] = \binom{n}{k} p^{\binom{k}{2}} \leq n^k p^{k(k-1)/2} = o(1)$ for $p = o(n^{-2/(k-1)})$

- **Markov's inequality:** $\Pr(X \geq 1) \leq \mathbb{E}[X] = o(1) \implies \Pr(X = 0) = 1 - o(1)$
 \implies If $p = o(n^{-2/(k-1)})$, then $G(n, p)$ is K_k -free **a.a.s.** (asymptotically almost surely)



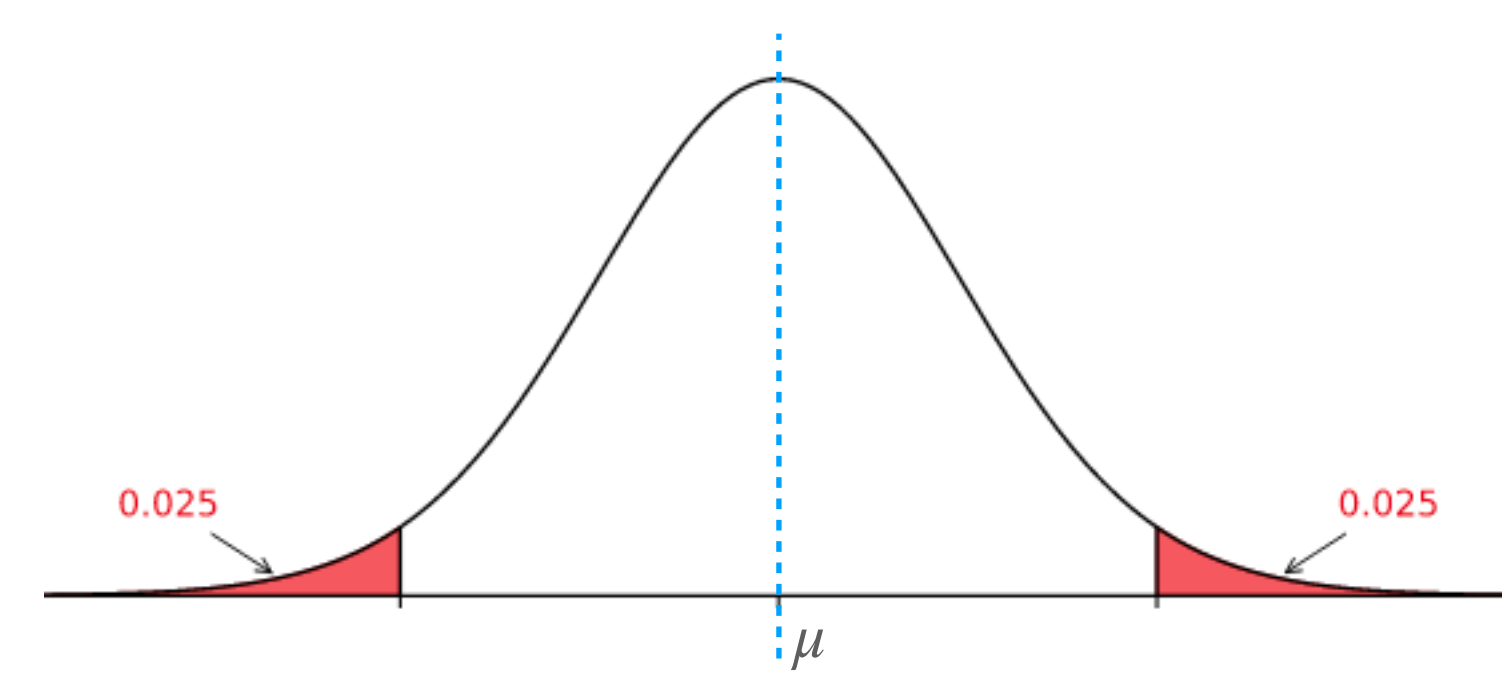
Generalized Markov's Inequality

- Let X be a random variable and $f: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ a nonnegative-valued function.

$$\text{For any } a > 0, \quad \Pr(f(X) \geq a) \leq \frac{\mathbb{E}[f(X)]}{a}$$

- **Proof:** Apply the Markov's inequality to the random variable $Y = f(X)$.
- **Applications:** useful if $f(X)$ can “*extract*” useful information about X
 - **Chebyshev's inequality, k th moment method:** $f(X)$ extracts the k th moment
 - **Chernoff-Hoeffding bounds, Bernstein inequalities:** $f(X)$ extracts all moments

Deviation Inequality



- Let X be a random variable with mean $\mu = \mathbb{E}[X]$. For $a > 0$

$$\Pr(|X - \mu| \geq a) \leq ?$$

- Applying Markov's inequality to $Y = |X - \mu|$ gives us

$$\Pr(|X - \mu| \geq a) \leq \frac{\mathbb{E}[|X - \mu|]}{a} \quad \text{difficult to calculate}$$

- Alternatively, we may apply Markov's inequality to $Y = (X - \mu)^2$

$$\Pr(|X - \mu| \geq a) = \Pr((X - \mu)^2 \geq a^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{a^2} \quad \text{Variance (2nd central moment)}$$

Variance (方差) and Moments (矩)

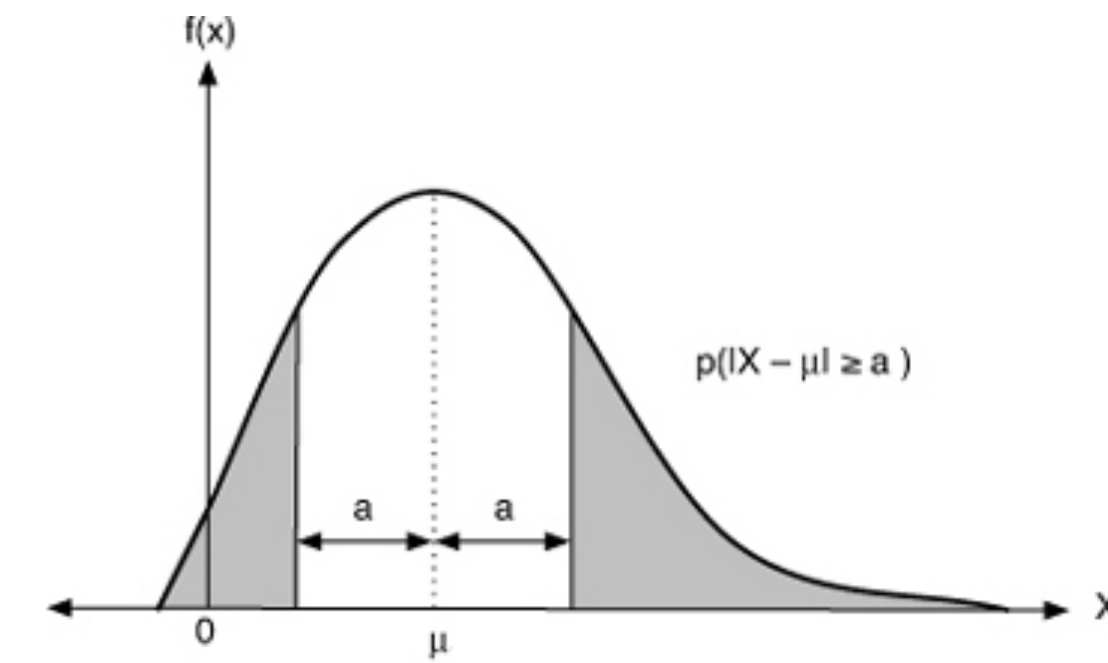
- For integer $k > 0$, the k th moment (k 阶矩) of a random variable X is $\mathbb{E}[X^k]$, and the k th central moment (k 阶中心矩) of X is $\mathbb{E}[(X - \mathbb{E}[X])^k]$.
- Sometimes, a random variable X is called **centralized** (中心化的) if $\mathbb{E}[X] = 0$. A random variable X can be centralized by $Y = X - \mathbb{E}[X]$.
- The variance (方差) of a random variable X is its 2nd central moment:

$$\mathbf{Var}[X] = \mathbb{E} [(X - \mathbb{E}[X])^2]$$

and the standard deviation (标准差) of X is $\sigma = \sigma[X] = \sqrt{\mathbf{Var}[X]}$

Chebyshev's Inequality

(切比雪夫不等式)



- Chebyshev's inequality: Let X be a random variable. For any $a > 0$,

$$\Pr(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbf{Var}[X]}{a^2}$$

- **Proof**: Apply Markov's inequality to $Y = (X - \mathbb{E}[X])^2$.

- **Corollary**: For standard deviation $\sigma = \sqrt{\mathbf{Var}[X]}$, for any $k \geq 1$,

$$\Pr(|X - \mathbb{E}[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

Median and Mean

- The median (中位数) of random variable X is defined to be any value m s.t.:

$$\Pr(X \leq m) \geq 1/2 \quad \text{and} \quad \Pr(X \geq m) \geq 1/2$$

- The expectation $\mu = \mathbb{E}[X]$ is the value that minimizes

$$\mathbb{E}[(X - \mu)^2]$$

- **Proof:** $f(x) = \mathbb{E}[(X - x)^2] = \mathbb{E}[X^2] - 2x\mathbb{E}[X] + x^2$ is convex and has $f'(\mu) = 0$

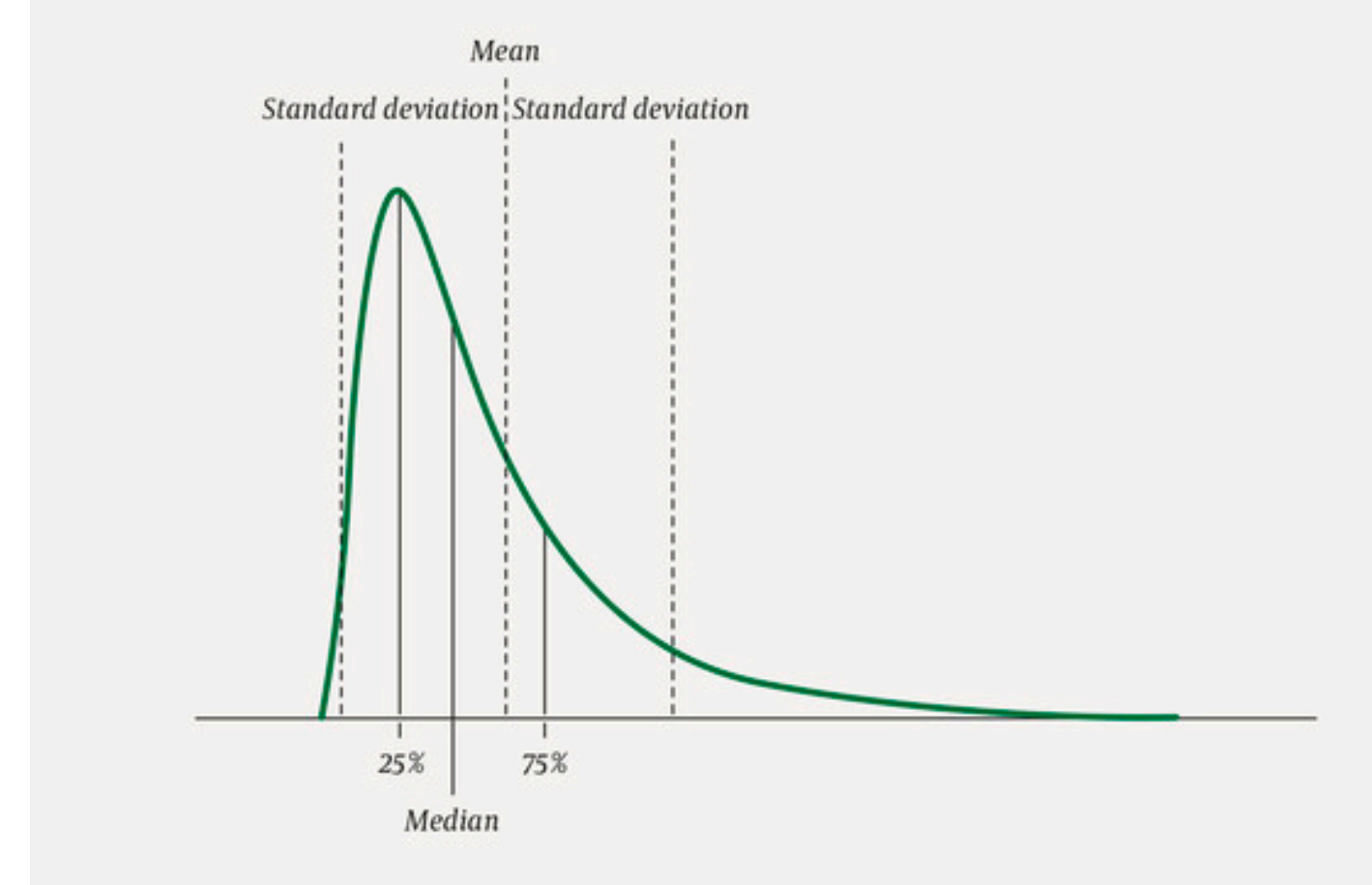
- The median m is the value that minimizes

$$\mathbb{E}[|X - m|]$$

- **Proof:** By symmetry, suppose non-median $y > m$ so that $\Pr(X \geq y) < 1/2$.

$$\begin{aligned} \mathbb{E}[|X - y| - |X - m|] &= (m - y) \Pr(X \geq y) + \sum_{m < x < y} (m + y - 2x) \Pr(X = x) + (y - m) \Pr(X \leq m) \\ &> (m - y)/2 + (y - m)/2 = 0 \end{aligned}$$

Median and Mean



- If X is a random variable with finite expectation μ , median m , and standard deviation σ , then

$$|\mu - m| \leq \sigma$$

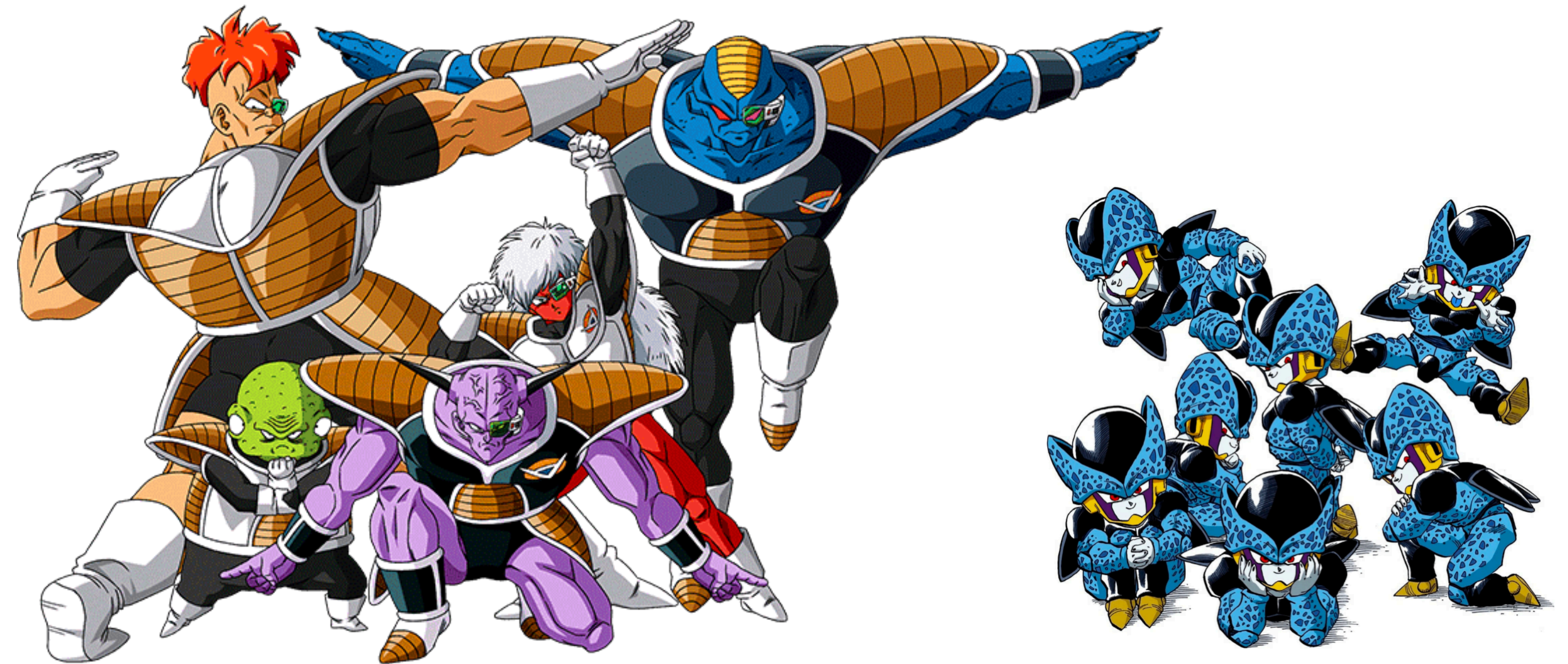
- **Proof:** $|\mu - m| = |\mathbb{E}[X] - m| = |\mathbb{E}[X - m]|$

$$\leq \mathbb{E}[|X - m|] \quad (\text{Jensen's inequality})$$

$$\leq \mathbb{E}[|X - \mu|] \quad (\text{the median } m \text{ minimizes } \mathbb{E}[|X - m|])$$

$$= \mathbb{E}\left[\sqrt{(X - \mu)^2}\right] \leq \sqrt{\mathbb{E}\left[(X - \mu)^2\right]} = \sigma \quad (\text{Jensen's inequality})$$

Variance



Calculation of Variance

$$\mathbf{Var}[X] = \mathbb{E} [(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

- **Proof:** $\mathbf{Var}[X] = \mathbb{E} [(X - \mathbb{E}[X])^2]$
 $= \mathbb{E} [X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2]$
 $= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2$
 $= \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- X is constant **a.s.** ($\Pr(X = \mathbb{E}[X]) = 1$) $\iff \mathbb{E}[X^2] = \mathbb{E}[X]^2 \iff \mathbf{Var}[X] = 0$

Variance of Linear Function

- For random variables X, Y and real number $a \in \mathbb{R}$:
 - $\mathbf{Var}[a] = 0$
 - $\mathbf{Var}[X + a] = \mathbf{Var}[X]$ (variance is a central moment)
 - $\mathbf{Var}[aX] = a^2 \mathbf{Var}[X]$ (variance is quadratic)
 - $\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y])$
- **Proof:** All can be verified through $\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

Covariance (协方差)

- The covariance (协方差) of two random variables X and Y is

$$\mathbf{Cov}(X, Y) = \mathbb{E} [(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- **Properties: $\mathbf{Var}[X] = \mathbf{Cov}(X, X)$**
 - *Symmetric:* $\mathbf{Cov}(X, Y) = \mathbf{Cov}(Y, X)$
 - *Distributive:* $\mathbf{Cov}(X + Y, Z) = \mathbf{Cov}(X, Z) + \mathbf{Cov}(Y, Z)$
 $\mathbf{Cov}(aX, Y) = a\mathbf{Cov}(X, Y)$
- If X and Y are independent then

$$\mathbf{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$$

Covariance of Independent Variables

- If random variables X and Y are independent, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

- If random variables X_1, X_2, \dots, X_n are mutually independent, then

$$\mathbb{E} \left[\prod_{i=1}^n X_i \right] = \mathbb{E} \left[\prod_{i=1}^{n-1} X_i \right] \cdot \mathbb{E}[X_n] = \prod_{i=1}^n \mathbb{E}[X_i]$$

Proof: By change of variable (*LOTUS*)

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

Expectation of Product

- For random variables X and Y :

if X and Y independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

- (Cauchy-Schwarz)

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

- (Hölder) for any $p, q > 0$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$

$$\mathbb{E}[XY] \leq \mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q}$$

Correlation (相关性)

- The covariance (协方差) of two random variables X and Y is

$$\mathbf{Cov}(X, Y) = \mathbb{E} [(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- The correlation coefficient (相关性系数) of X and Y is

$$\rho(X, Y) = \frac{\mathbf{Cov}(X, Y)}{\sqrt{\mathbf{Var}[X] \cdot \mathbf{Var}[Y]}}$$

- Two random variables X and Y are called uncorrelated if $\mathbf{Cov}(X, Y) = 0$
- X and Y are uncorrelated means:
 - $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$
 - $\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y]$

Variance of Sum

- For random variables X, Y :

$$\mathbf{Var}[X + Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X, Y)$$

- For random variables X_1, X_2, \dots, X_n :

$$\mathbf{Var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \mathbf{Var}[X_i] + \sum_{i \neq j} \mathbf{Cov}(X_i, X_j)$$

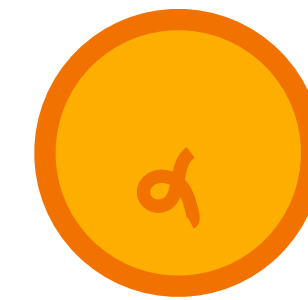
- For **pairwise** independent X_1, X_2, \dots, X_n :

$$\mathbf{Var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \mathbf{Var}[X_i]$$

Variance of Indicator



p



$1 - p$

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

$$X^2 = X \implies \mathbb{E}[X^2] = \mathbb{E}[X] = p$$

$$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1 - p)$$

- For the indicator random variable $X = I(A)$ of event A :

$$\mathbf{Var}[X] = \Pr(A)(1 - \Pr(A)) = \Pr(A) \Pr(A^c)$$

Variance of Discrete Uniform Distribution

- For integers $a \leq b$, let X be chosen from $[a, b] = \{a, a + 1, \dots, b\}$ **u.a.r.**

- $$\mathbb{E}[X] = \sum_{k=a}^b \frac{k}{b - a + 1} = \frac{a + b}{2}$$

- $$\mathbb{E}[X^2] = \sum_{k=a}^b \frac{k^2}{b - a + 1} = \frac{2b^2 + 2ab + 2a^2 + b - a}{6}$$

- $$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{(b - a)(b - a + 2)}{12}$$

Poisson Distribution

- For Poisson random variable $X \sim \text{Pois}(\lambda)$, recall $\mathbb{E}[X] = \lambda$, and

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{k \geq 0} k^2 \frac{e^{-\lambda} \lambda^k}{k!} = \sum_{k \geq 1} k \frac{e^{-\lambda} \lambda^k}{(k-1)!} \\ &= \sum_{k \geq 0} (k+1) \frac{e^{-\lambda} \lambda^{k+1}}{k!} = \lambda \sum_{k \geq 0} (k+1) \frac{e^{-\lambda} \lambda^k}{k!} \\ &= \lambda \mathbb{E}[X+1] = \lambda(\mathbb{E}[X] + 1) = \lambda(\lambda + 1)\end{aligned}$$

$$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda$$

Geometric Distribution (几何分布)



- For geometric random variable $X \sim \text{Geo}(p)$, recall $\mathbb{E}[X] = 1/p$, and

$$\mathbb{E}[X^2] = \sum_{k \geq 1} k^2 (1-p)^{k-1} p = (2-p)p^{-2}$$

$$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = (2-p)p^{-2} - p^{-2} = (1-p)/p^2$$

- **Total expectation:** $\mathbb{E}[X^2] = \mathbb{E}[X^2 | X > 1] \cdot (1-p) + \mathbb{E}[X^2 | X = 1] \cdot p$
 $= \mathbb{E}[((X-1)+1)^2 | X > 1] \cdot (1-p) + p$
(memoryless) $= \mathbb{E}[(X+1)^2] \cdot (1-p) + p$
 $= (1-p)\mathbb{E}[X^2] + 2(1-p)/p + 1$

$$\implies \mathbb{E}[X^2] = (2-p)/p^2 \implies \mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = (1-p)/p^2$$

Binomial Distribution (二项分布)

- For binomial random variable $X \sim \text{Bin}(n, p)$, recall $\mathbb{E}[X] = np$, and

$$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} - (np)^2$$

- **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
- For mutually independent X_1, \dots, X_n :

$$\mathbf{Var}[X] = \sum_{i=1}^n \mathbf{Var}[X_i] = np(1-p)$$

Negative Binomial Distribution (负二项分布)

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

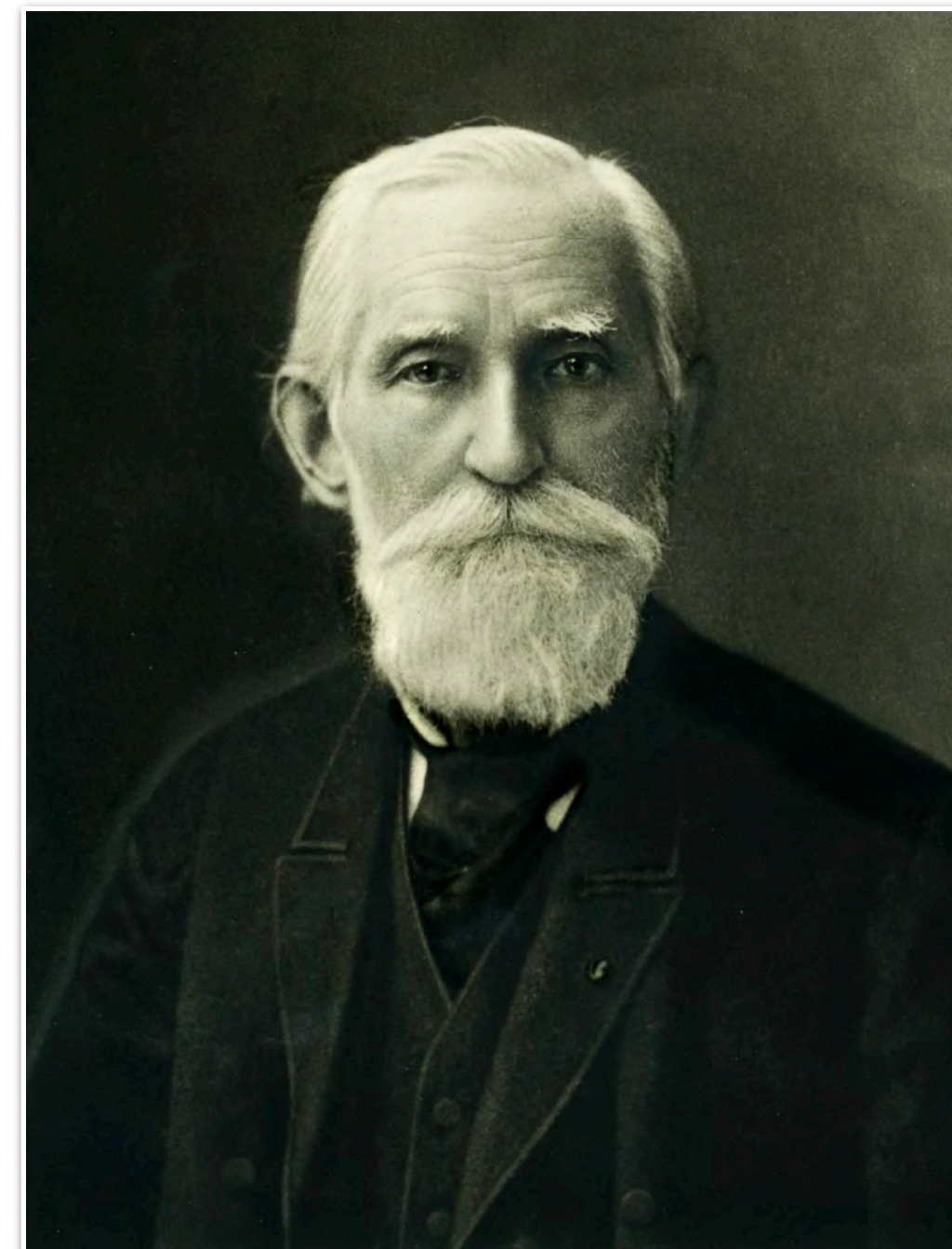
$$\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_{k \geq 1} k^2 \binom{k+r-1}{k} (1-p)^k p^{r-k} - r^2 (1-p)^2 / p^2$$

- **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

- For mutually independent X_1, \dots, X_r :

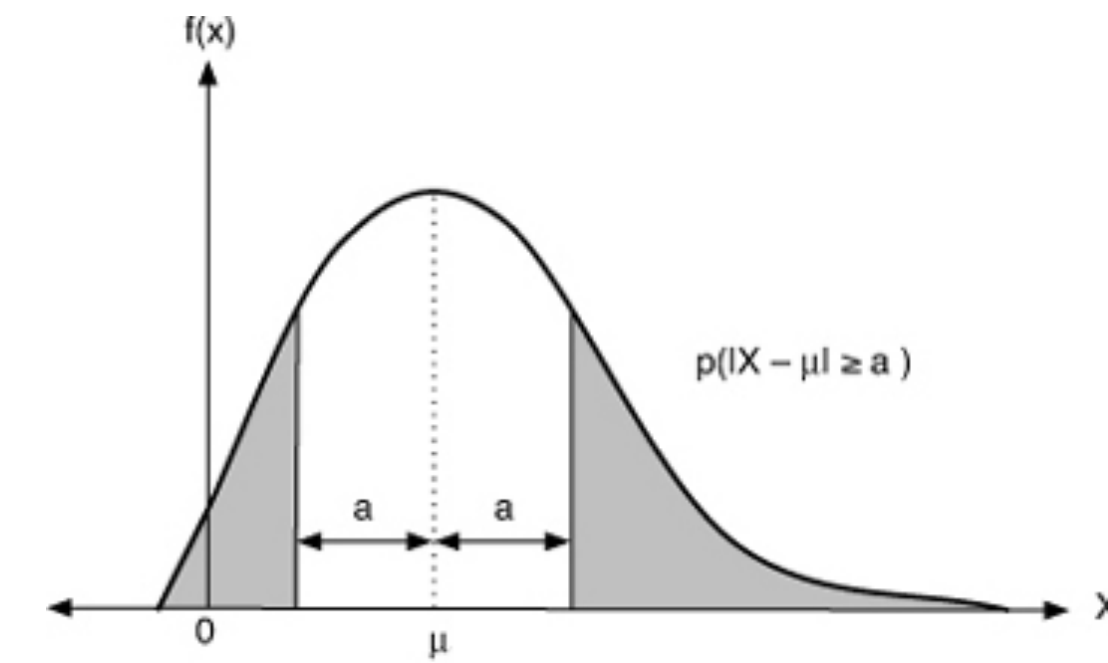
$$\mathbf{Var}[X] = \sum_{i=1}^r \mathbf{Var}[X_i - 1] = \sum_{i=1}^r \mathbf{Var}[X_i] = \frac{r(1-p)}{p^2}$$

Chebyshev (Чебышёв)'s Inequality



Chebyshev's Inequality

(切比雪夫不等式)



- Chebyshev's inequality: Let X be a random variable. For any $a > 0$,

$$\Pr(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbf{Var}[X]}{a^2}$$

- **Corollary**: For standard deviation $\sigma = \sqrt{\mathbf{Var}[X]}$, for any $k \geq 1$,

$$\Pr(|X - \mathbb{E}[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

- **Tight in the worst case**: $\forall k \geq 1$, $\forall \mu \in \mathbb{R}$ and $\forall \sigma > 0$, $\exists X$ with $\mathbb{E}[X] = \mu$ and $\mathbf{Var}[X] = \sigma^2$ such that $\Pr(|X - \mu| \geq k\sigma) = 1/k^2$

Unbiased Estimator

- Let X_1, \dots, X_n be *i.i.d.* random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbf{Var}[X_i] = \sigma^2$.

- Empirical mean: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

$$\mathbb{E}[\bar{X}] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mu \quad \text{and} \quad \mathbf{Var}[\bar{X}] = \frac{1}{n^2} \sum_{i=1}^n \mathbf{Var}[X_i] = \frac{\sigma^2}{n}$$

- Chebyshev's inequality:

$$\Pr(|\bar{X} - \mu| \geq \epsilon\mu) \leq \frac{\mathbf{Var}[\bar{X}]}{\epsilon^2\mu^2} = \frac{\sigma^2}{\epsilon^2\mu^2n} \leq \delta \quad \text{if } n \geq \frac{\sigma^2}{\epsilon^2\mu^2\delta}$$

(one-sided) Error Reduction

- Decision problem $f : \{0,1\}^* \rightarrow \{0,1\}$.
- Monte Carlo randomized algorithm \mathcal{A} with *one-sided* error:
for any input x and uniform *random seed* $r \in [p]$ for some **prime number** p
 - $f(x) = 1 \implies \Pr(\mathcal{A}(x, r) = 1) \geq \epsilon$
 - $f(x) = 0 \implies \mathcal{A}(x, r) = 0$ for all $r \in [p]$
- $\mathcal{A}^k(x, r_1, \dots, r_k) = \bigvee_{i=1}^k \mathcal{A}(x, r_i)$: for **mutually independent** $r_1, \dots, r_k \in [p]$
 - $f(x) = 1 \implies \Pr(\mathcal{A}^k(x, r_1, \dots, r_k) = 0) \leq (1 - \epsilon)^k$

Two-Point Sampling (2-Universal Hashing)

- Let $p > 1$ be a prime number and $[p] = \{0, 1, \dots, p - 1\} = \mathbb{Z}_p$.
- Pick $\mathbf{a}, \mathbf{b} \in [p]$ *u.a.r.* and let $r_i = (\mathbf{a} \cdot i + \mathbf{b}) \bmod p$ for $i = 1, 2, \dots, p$
 - $r_1, \dots, r_p \in [p]$ are pairwise independent
 - each r_i is uniformly distributed over $[p]$
- **Proof:** For any $i \neq j$, $\forall c, d \in [p]$, $\Pr(r_i = c \cap r_j = d) = 1/p^2$ because
$$\begin{cases} \mathbf{a} \cdot i + \mathbf{b} \equiv c \pmod{p} \\ \mathbf{a} \cdot j + \mathbf{b} \equiv d \pmod{p} \end{cases}$$
 has a unique solution $(\mathbf{a}, \mathbf{b}) \in [p]^2$
$$\Pr(r_i = c) = \Pr(\mathbf{a} \cdot i + \mathbf{b} \equiv c \pmod{p}) = \frac{1}{p} \sum_{\mathbf{a} \in [p]} \Pr(\mathbf{b} \equiv c - \mathbf{a}i \pmod{p}) = \frac{1}{p}$$

Derandomization with Two-Point Sampling

- \mathcal{A} : for any input x and uniform *random seed* $r \in [p]$ for **prime number** p
 - $f(x) = 1 \implies \Pr(\mathcal{A}(x, r) = 1) \geq \epsilon$
 - $f(x) = 0 \implies \mathcal{A}(x, r) = 0$ for all $r \in [p]$
- $\mathcal{A}^k(x, r_1, \dots, r_k) = \bigvee_{i=1}^k \mathcal{A}(x, r_i)$: $k \leq p$ for $r_i = (\mathbf{a} \cdot i + \mathbf{b}) \bmod p$ with uniform $\mathbf{a}, \mathbf{b} \in [p]$
 - If $f(x) = 0 \implies \mathcal{A}^k(x, r_1, \dots, r_k) = \bigvee_{i=1}^k \mathcal{A}(x, r_i) = 0$
 - If $f(x) = 1 \implies \Pr(\mathcal{A}(x, r_i) = 1) \geq \epsilon$ because each r_i is uniform over $[p]$
 - Let $X_i = \mathcal{A}(x, r_i)$ and let $X = \sum_{i=1}^k X_i$.
 - X_1, \dots, X_k are pairwise independent Bernoulli random variables with $\Pr(X_i = 1) \geq \epsilon$
 - $\Pr(\mathcal{A}^k(x, r_1, \dots, r_k) = 0) = \Pr(X = 0) \leq \Pr(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathbf{Var}[X]}{\mathbb{E}[X]^2}$
(Chebyshev's inequality)

Derandomization with Two-Point Sampling

- $\mathcal{A}^k(x, r_1, \dots, r_k) = \bigvee_{i=1}^k \mathcal{A}(x, r_i)$: $k \leq p$ and $r_i = (\mathbf{a} \cdot i + \mathbf{b}) \bmod p$ with uniform $\mathbf{a}, \mathbf{b} \in [p]$
- If $f(x) = 1 \implies \Pr(\mathcal{A}(x, r_i) = 1) \geq \epsilon$ because each r_i is uniform over $[p]$
- Let $X_i = \mathcal{A}(x, r_i)$ and let $X = \sum_{i=1}^k X_i$.
 - X_1, \dots, X_k are pairwise independent Bernoulli random variables with $\Pr(X_i = 1) \geq \epsilon$
 - $\Pr(\mathcal{A}^k(x, r_1, \dots, r_k) = 0) = \Pr(X = 0) \leq \Pr(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathbf{Var}[X]}{\mathbb{E}[X]^2} \leq \frac{1}{\epsilon k}$
 - Linearity of expectation: $\mathbb{E}[X] = \sum_{i=1}^k \mathbb{E}[X_i] \geq \epsilon k$
 - Pairwise independence: $\mathbf{Var}[X] = \sum_{i=1}^k \mathbf{Var}[X_i] \leq \sum_{i=1}^k \mathbb{E}[X_i^2] = \sum_{i=1}^k \mathbb{E}[X_i] = \mathbb{E}[X]$
- Reduce any 1-sided error $1 - \epsilon$ to $1/(\epsilon k)$ with $k \leq p$ runs of the algorithm using only **2 random seeds** in total.

Cliques in Random Graph (revisited)

- Fix a constant integer $k \geq 3$. Let X be the number of k -cliques (K_k) in $G \sim G(n, p)$.
- For every distinct $S \subseteq \in [n]$ of size $|S| = k$, let $I_S = I(K_S \subseteq G)$. Then:

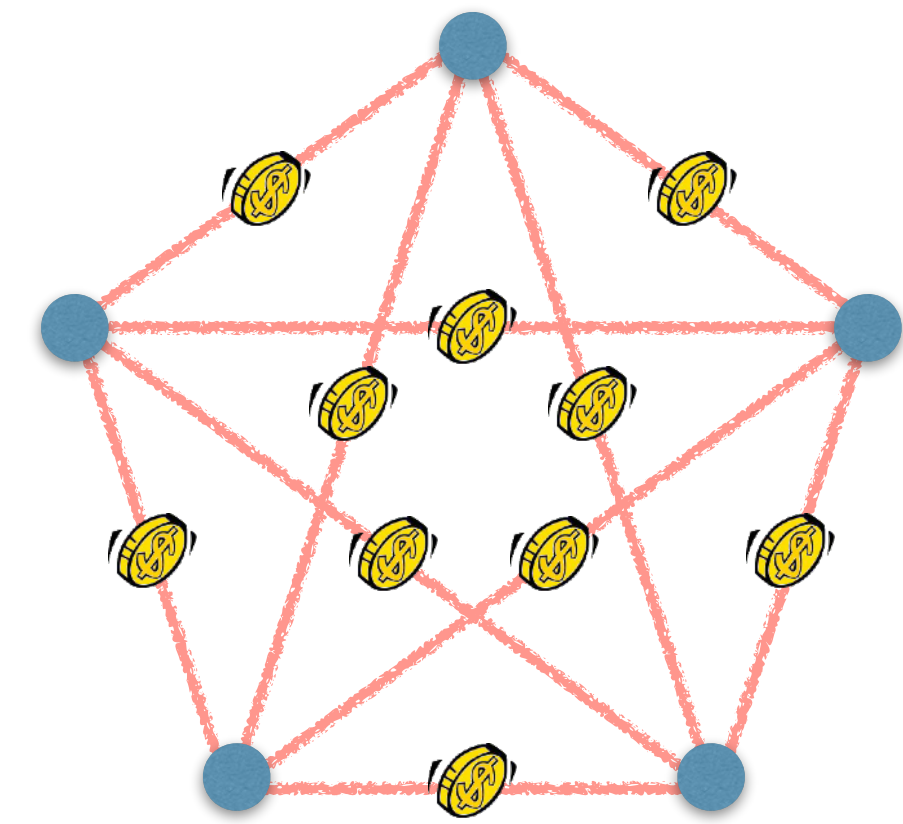
$$X = \sum_{S \in \binom{[n]}{k}} I_S \text{ and } \mathbb{E}[I_S] = \Pr(K_S \subseteq G) = p^{\binom{k}{2}}$$

- **Linearity of expectation:** $\mathbb{E}[X] = \binom{n}{k} p^{\binom{k}{2}} = \Theta \left(n^k p^{\binom{k}{2}} \right)$

$$\mathbb{E}[X] = \Theta \left(n^k p^{\binom{k}{2}} \right) = \begin{cases} o(1) & \text{if } p = o \left(n^{-2/(k-1)} \right) \\ \omega(1) & \text{if } p = \omega \left(n^{-2/(k-1)} \right) \end{cases}$$

$$\stackrel{\text{(Markov)}}{\implies} \Pr(X \geq 1) = o(1)$$

$$\stackrel{?}{\implies} \Pr(X \geq 1) = 1 - o(1)$$



Cliques in Random Graph (revisited)

- Fix a constant integer $k \geq 3$. Let X be the number of k -cliques (K_k) in $G \sim G(n, p)$.
- For every distinct $S \subseteq \in [n]$ of size $|S| = k$, let $I_S = I(K_S \subseteq G)$. Then:

$$X = \sum_{S \in \binom{[n]}{k}} I_S \text{ and } \mathbb{E}[X] = \Theta \left(n^k p^{\binom{k}{2}} \right) = \begin{cases} o(1) & \text{if } p = o \left(n^{-2/(k-1)} \right) \\ \omega(1) & \text{if } p = \omega \left(n^{-2/(k-1)} \right) \end{cases}$$

- Chebyshev: $\Pr(X = 0) \leq \Pr(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathbf{Var}[X]}{\mathbb{E}[X]^2} \leq \frac{1}{\mathbb{E}[X]} + \frac{\sum_{S \neq T} \mathbb{E}[I_S I_T]}{\mathbb{E}[X]^2}$

$$\begin{aligned} \mathbf{Var}[X] &= \sum_{S \in \binom{[n]}{k}} \mathbf{Var}[I_S] + \sum_{\substack{S \neq T \\ S, T \in \binom{[n]}{k}}} \mathbf{Cov}(I_S, I_T) = \sum_{S \in \binom{[n]}{k}} (\mathbb{E}[I_S^2] - \mathbb{E}[I_S]^2) + \sum_{\substack{S \neq T \\ S, T \in \binom{[n]}{k}}} (\mathbb{E}[I_S I_T] - \mathbb{E}[I_S] \mathbb{E}[I_T]) \\ &= \sum_S (\mathbb{E}[I_S] - \mathbb{E}[I_S]^2) + \sum_{S \neq T} (\mathbb{E}[I_S I_T] - \mathbb{E}[I_S] \mathbb{E}[I_T]) \\ &\leq \mathbb{E}[X] + \sum_{S \neq T} \mathbb{E}[I_S I_T] \end{aligned}$$

Cliques in Random Graph (revisited)

- Fix a constant integer $k \geq 3$. Let X be the number of k -cliques (K_k) in $G \sim G(n, p)$.
- For every distinct $S \subseteq \in [n]$ of size $|S| = k$, let $I_S = I(K_S \subseteq G)$. Then:

$$X = \sum_{S \in \binom{[n]}{k}} I_S \text{ and } \mathbb{E}[X] = \Theta \left(n^k p^{\binom{k}{2}} \right) = \begin{cases} o(1) & \text{if } p = o \left(n^{-2/(k-1)} \right) \\ \omega(1) & \text{if } p = \omega \left(n^{-2/(k-1)} \right) \end{cases}$$

- Chebyshev: $\Pr(X = 0) \leq \Pr(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathbf{Var}[X]}{\mathbb{E}[X]^2} \leq \frac{1}{\mathbb{E}[X]} + \frac{\sum_{S \neq T} \mathbb{E}[I_S I_T]}{\mathbb{E}[X]^2}$

$$\mathbb{E}[I_S I_T] = \Pr((K_S \cup K_T) \subseteq G) = p^{2\binom{k}{2} - \binom{|S \cap T|}{2}}$$

$$\sum_{\substack{S \neq T \\ S, T \in \binom{[n]}{k}}} \mathbb{E}[I_S I_T] = \sum_{\ell=2}^{k-1} \sum_{\substack{|S \cap T| = \ell \\ S, T \in \binom{[n]}{k}}} \mathbb{E}[I_S I_T] = \sum_{\ell=2}^{k-1} \binom{n}{2k-\ell} \cdot O(1) \cdot p^{2\binom{k}{2} - \binom{\ell}{2}} = O \left(n^{2k} p^{2\binom{k}{2}} \sum_{\ell=2}^{k-1} n^{-\ell} p^{-\binom{\ell}{2}} \right)$$

Cliques in Random Graph (revisited)

- Fix a constant integer $k \geq 3$. Let X be the number of k -cliques (K_k) in $G \sim G(n, p)$.

- For every distinct $S \subseteq \mathbb{E} [n]$ of size $|S| = k$, let $I_S = I(K_S \subseteq G)$. Then:

$$X = \sum_{S \in \binom{[n]}{k}} I_S \text{ and } \mathbb{E}[X] = \Theta \left(n^k p^{\binom{k}{2}} \right) = \begin{cases} o(1) & \text{if } p = o \left(n^{-2/(k-1)} \right) \\ \omega(1) & \text{if } p = \omega \left(n^{-2/(k-1)} \right) \end{cases}$$

- Chebyshev: $\Pr(X = 0) \leq \Pr(|X - \mathbb{E}[X]| \geq \mathbb{E}[X]) \leq \frac{\mathbf{Var}[X]}{\mathbb{E}[X]^2} \leq \frac{1}{\mathbb{E}[X]} + \frac{\sum_{S \neq T} \mathbb{E}[I_S I_T]}{\mathbb{E}[X]^2}$

$$= O \left(n^{-k} p^{-\binom{k}{2}} \right) + O \left(\sum_{\ell=2}^{k-1} n^{-\ell} p^{-\binom{\ell}{2}} \right) = O \left(\sum_{\ell=2}^k n^{-\ell} p^{-\binom{\ell}{2}} \right)$$

$$= o(1) \text{ if } p = \omega \left(n^{2/(1-k)} \right)$$

- $\implies \Pr(X \geq 1) \geq 1 - o(1)$

A “Threshold Behavior” in Random Graphs (Erdős–Rényi 1960)

- Fix a constant integer $k \geq 3$.
- Let $G \sim G(n, p)$, as $n \rightarrow \infty$:

$$\Pr(G \text{ contains a } K_k) = \begin{cases} o(1) & \text{if } p = o(n^{-2/(k-1)}) \\ 1 - o(1) & \text{if } p = \omega(n^{-2/(k-1)}) \end{cases}$$

- For $H(V, E)$ with $k = |V|$, $m = |E|$ s.t. every subgraph of H has density $\leq m/k$:

$$\Pr(G \text{ contains a subgraph } H) = \begin{cases} o(1) & \text{if } p = o(n^{-k/m}) \\ 1 - o(1) & \text{if } p = \omega(n^{-k/m}) \end{cases}$$

Weierstrass Approximation Theorem

(魏尔施特拉斯逼近定理)

- Weierstrass Approximation Theorem: Let $f : [0,1] \rightarrow [0,1]$ be a continuous function. For any $\epsilon > 0$, there exists a polynomial p such that

$$\sup_{x \in [0,1]} |p(x) - f(x)| \leq \epsilon$$

- **Proof**: Let integer n be sufficiently large (to be fixed later).

For $x \in [0,1]$, let $Y_x \sim \text{Bin}(n, x)$. Define polynomial p on $x \in [0,1]$ to be:

$$p(x) = \mathbb{E} \left[f \left(\frac{Y_x}{n} \right) \right] = \sum_{k=0}^n f \left(\frac{k}{n} \right) \binom{n}{k} x^k (1-x)^{n-k}$$

Let $f : [0,1] \rightarrow [0,1]$ be continuous. For $x \in [0,1]$, let $Y_x \sim \text{Bin}(n, x)$, and:

$$p(x) = \mathbb{E} \left[f \left(\frac{Y_x}{n} \right) \right] = \sum_{k=0}^n f \left(\frac{k}{n} \right) \binom{n}{k} x^k (1-x)^{n-k}$$

$$|p(x) - f(x)| = \left| \mathbb{E} [f(Y_x/n) - f(x)] \right| \leq \mathbb{E} \left[|f(Y_x/n) - f(x)| \right]$$

(f is continuous on $[0,1] \implies \exists \delta > 0$ s.t. $|f(x) - f(y)| \leq \epsilon/2$ for all $|x - y| \leq \delta$)

$$= \mathbb{E} \left[|f(Y_x/n) - f(x)| \mid |Y_x/n - x| \leq \delta \right] \cdot \Pr \left(|Y_x/n - x| \leq \delta \right) \\ + \mathbb{E} \left[|f(Y_x/n) - f(x)| \mid |Y_x/n - x| > \delta \right] \cdot \Pr \left(|Y_x/n - x| > \delta \right)$$

$$\leq \mathbb{E} [\epsilon/2] + |1 - 0| \cdot \Pr \left(|Y_x - nx| > n\delta \right) \leq \frac{\epsilon}{2} + \frac{x(1-x)}{n\delta^2} \quad (\text{Chebyshev})$$

$$\leq \frac{\epsilon}{2} + \frac{1}{4n\delta^2} \leq \epsilon \quad \text{if we choose } n \geq \frac{1}{2\epsilon\delta^2}$$

Weierstrass Approximation Theorem

(魏尔施特拉斯逼近定理)

- Weierstrass Approximation Theorem: Let $f : [0,1] \rightarrow [0,1]$ be a continuous function. For any $\epsilon > 0$, there exists a polynomial p such that

$$\sup_{x \in [0,1]} |p(x) - f(x)| \leq \epsilon$$

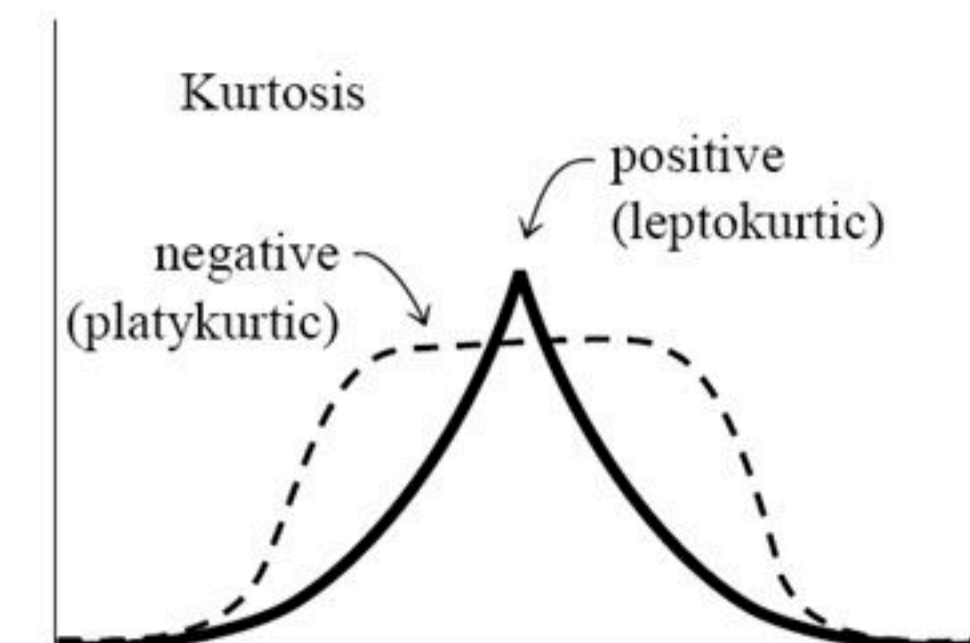
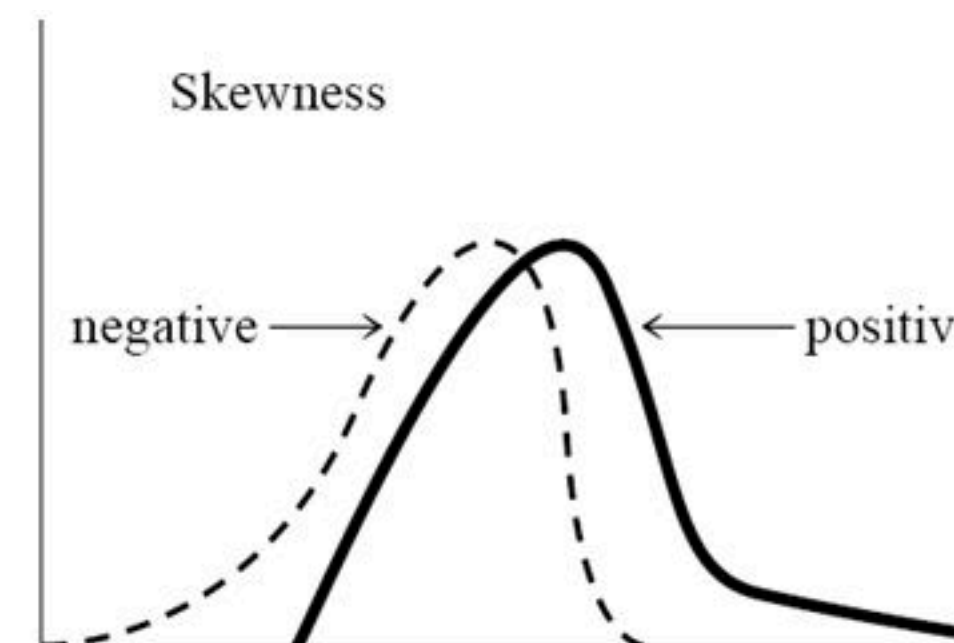
- **Proof:** By continuity, $\exists \delta > 0$ s.t. $|f(x) - f(y)| \leq \epsilon/2$ if $|x - y| \leq \delta$.

Let $n \geq 1/(2\epsilon\delta^2)$ be any integer. For $x \in [0,1]$, let $Y_x \sim \text{Bin}(n, x)$, and:

$$p(x) = \mathbb{E} \left[f\left(\frac{Y_x}{n}\right) \right] = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$

For any $x \in [0,1]$, it holds that $|p(x) - f(x)| \leq \epsilon$.

Higher Moments

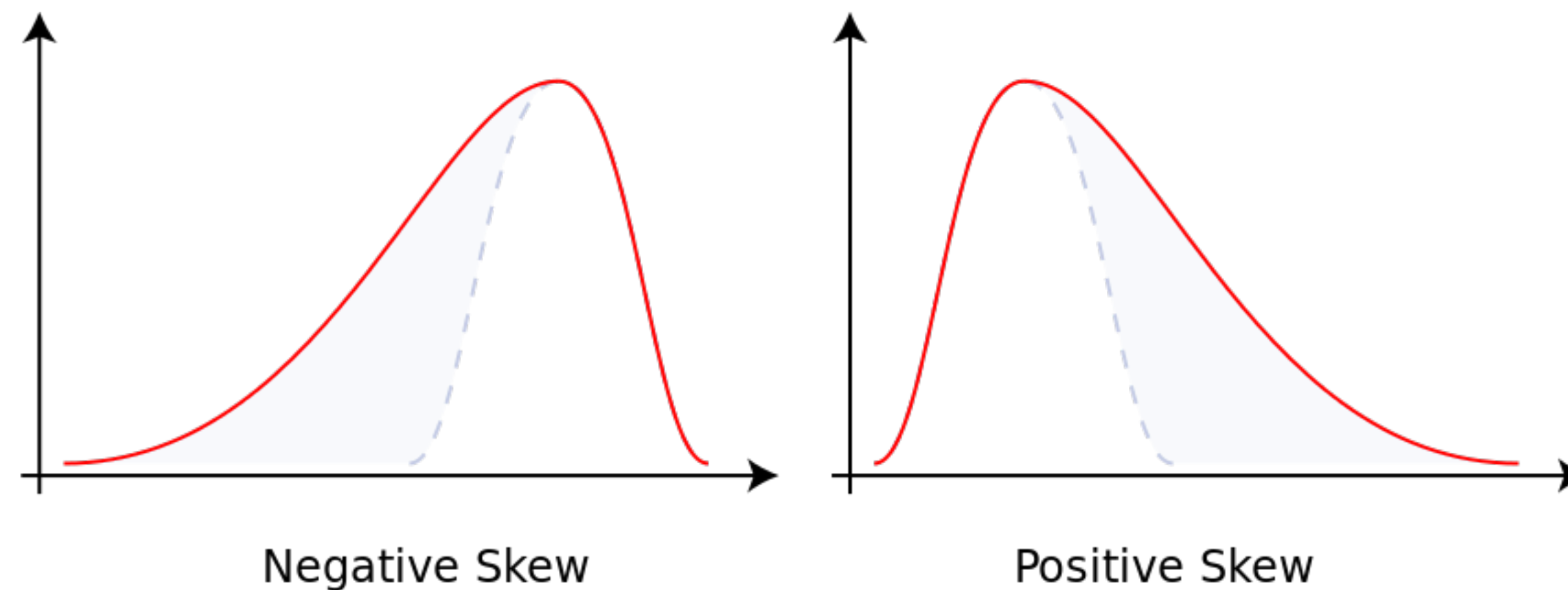


Skewness (偏度)

- The skewness (偏度) of a random variable X with expectation $\mu = \mathbb{E}[X]$ and standard deviation $\sigma = \sqrt{\mathbf{Var}[X]}$ is defined by

$$\text{Skew}[X] = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] = \frac{\mathbb{E}[(X - \mu)^3]}{\sigma^3}$$

standardized
moment
(of degree 3)

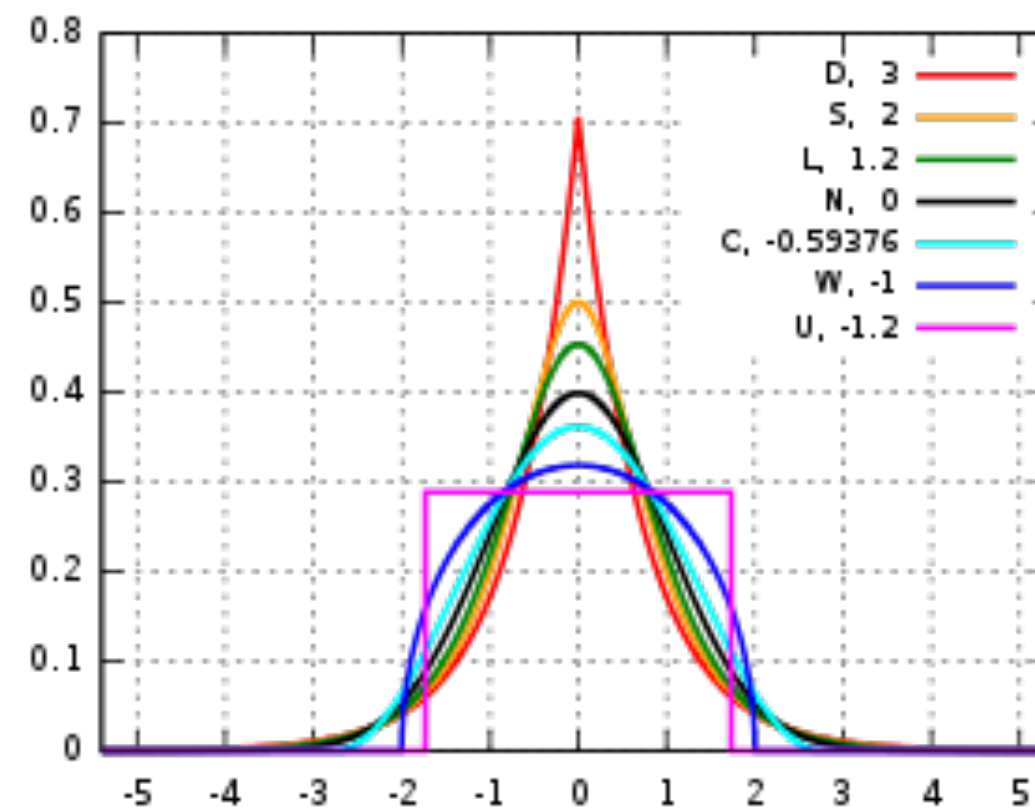


Kurtosis (峰度)

- The kurtosis (峰度) of a random variable X with expectation $\mu = \mathbb{E}[X]$ and standard deviation $\sigma = \sqrt{\mathbf{Var}[X]}$ is defined by

$$\text{Kurt}[X] = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] = \frac{\mathbb{E}[(X - \mu)^4]}{\sigma^4}$$

standardized
moment
(of degree 4)



The k th Moment Method

- Let X be a random variable with $\mathbb{E}[X] = \mu$. For any $C > 1$ and integer $k \geq 1$

$$\Pr \left(|X - \mu| \geq C \cdot \mathbb{E} \left[|X - \mu|^k \right]^{\frac{1}{k}} \right) \leq \frac{1}{C^k}$$

- **Proof:** Apply Markov's inequality to $Z = |X - \mu|^k$.

The Moment Problem

- Do moments $m_k = \mathbb{E}[X^k]$, $\forall k \geq 1$, uniquely identify the distribution of X ?
- If X takes values from a finite set $\{x_1, \dots, x_n\}$, then solving the Vandermonde system:

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^n & x_2^n & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}$$

can recover the *pmf* $p_i = p_X(x_i)$

The Moment Problem

- Do moments $m_k = \mathbb{E}[X^k]$, $\forall k \geq 1$, uniquely identify the distribution of X ?
 - If $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$ for all $k \geq 1$, are X and Y always identically distributed?
- If X and Y have the same moment generating function (MGF)

$$M_X(t) = \mathbb{E}[e^{tX}] = \sum_{k \geq 0} \frac{t^k \mathbb{E}[X^k]}{k!}$$

then X and Y are identically distributed.

- The MGF $M_X(t)$ is convergent if the sequence $\mathbb{E}[X^k]$ does not grow too fast.