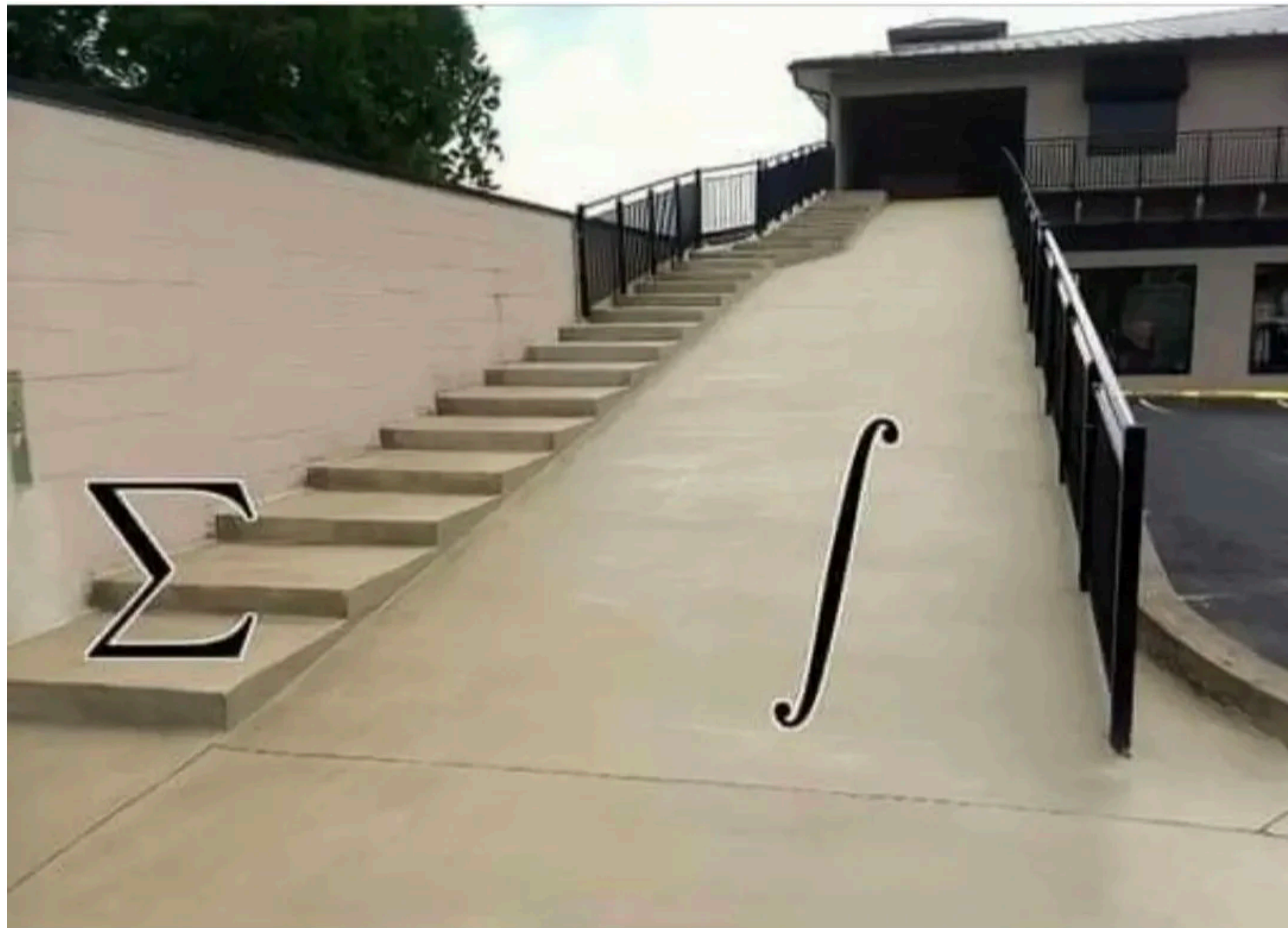
Probability Theory & Mathematical Statistics

Continuous Random Variable

尹一通 Nanjing University, 2024 Spring

Continuous Random Variable



Random Variable

- Given $(\Omega, \Sigma, \Pr)$, a random variable is a function $X : \Omega \to \mathbb{R}$
  - satisfying that $\forall x \in \mathbb{R}, \{\omega \in \Omega \mid X(\omega) \le x\} \in \Sigma$ (i.e. X is Σ-measurable)
- Events: $X \in B$, where $B \subseteq \mathbb{R}$ is countable ∩,∪,\ of intervals $(a, b]$
- The cumulative distribution function (CDF) (累积分布函数) of a random variable X is the $F_X : \mathbb{R} \to [0,1]$ given by

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

- **CDF** $F_X : \mathbb{R} \to [0,1]$ is monotonically nondecreasing with

$$\lim_{x\to-\infty} F_X(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F_X(x) = 1$$

Continuous Random Variable

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

$$F_X(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u$$

for some integrable function $f : \mathbb{R} \to [0,\infty)$

- f is called probability density function (*pdf*) (概率密度函数) of X.

Probability Density Function

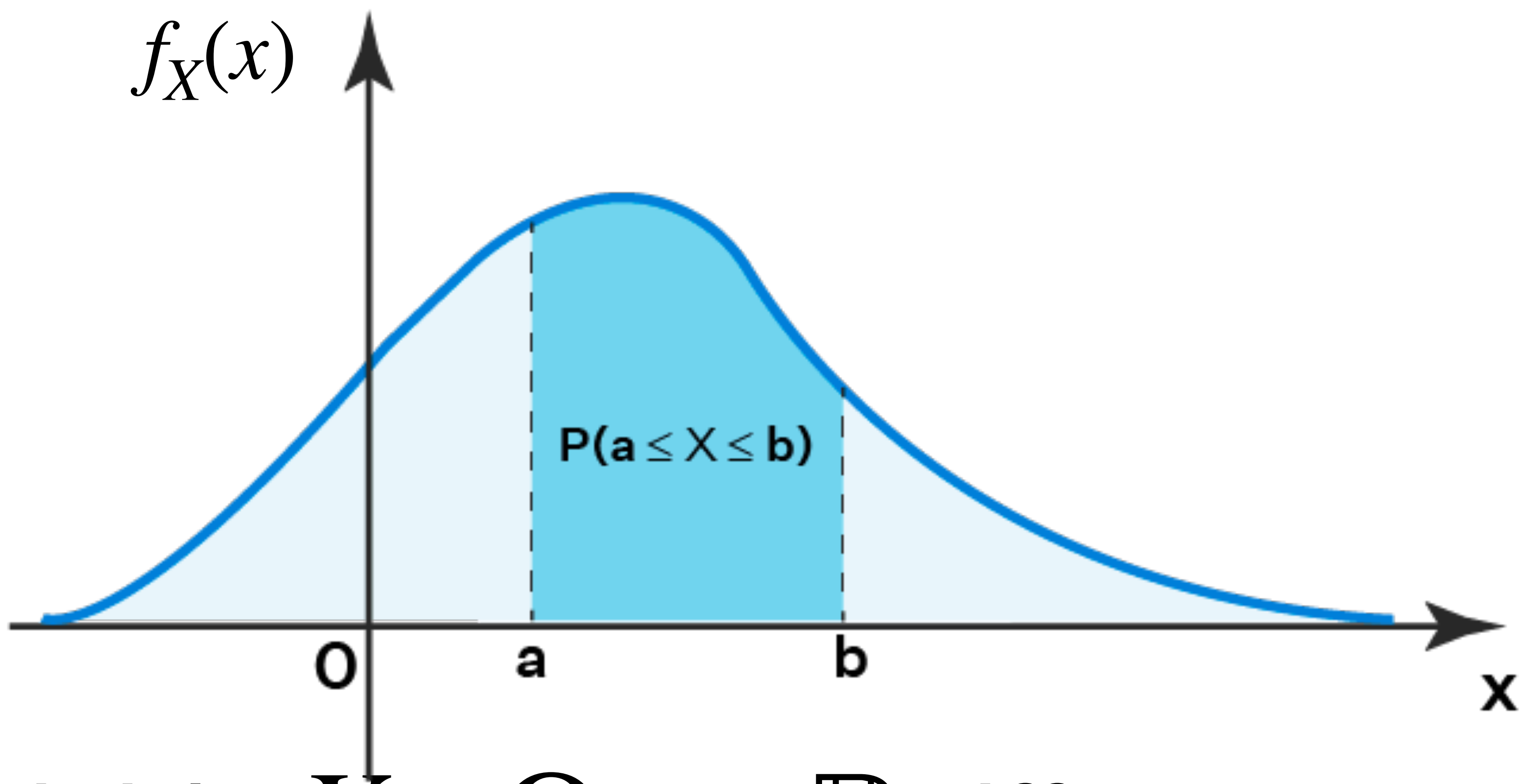
- The density function f is *NOT* uniquely determined by $F_X(x) = \int_{-\infty}^{x} f(u)\, \mathrm{d}u$
  - If F_X is differentiable then we normally set $f_X(x) = F'_X(x)$
- For continuous random variable X, it holds that

$$\Pr(X = x) = 0 \text{ for all } x \in \mathbb{R}$$

- The density value $f_X(x) \geq 0$ is *NOT* a probability, but a proportion ("*density*")

$$\Pr(x < X \leq x + \Delta x) = F_X(x + \Delta x) - F_X(x) \approx f_X(x)\Delta x$$

Probability Density Function



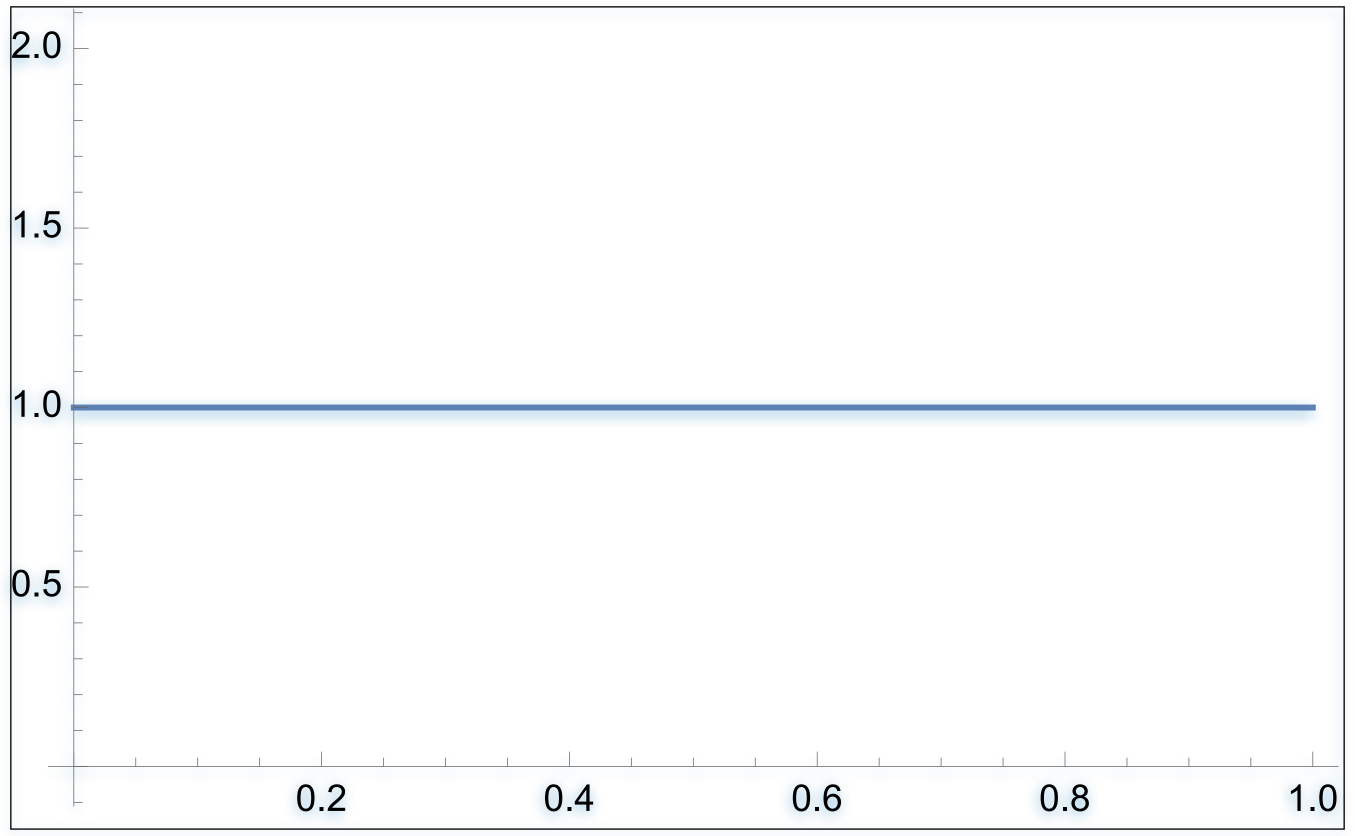
- f_X is the density function of a continuous random variable $X : \Omega \to \mathbb{R}$, iff
  - $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = \Pr(-\infty < X < \infty) = 1$
  - $f_X(x) \geq 0$ for all $x \in \mathbb{R}$

$\Longrightarrow$

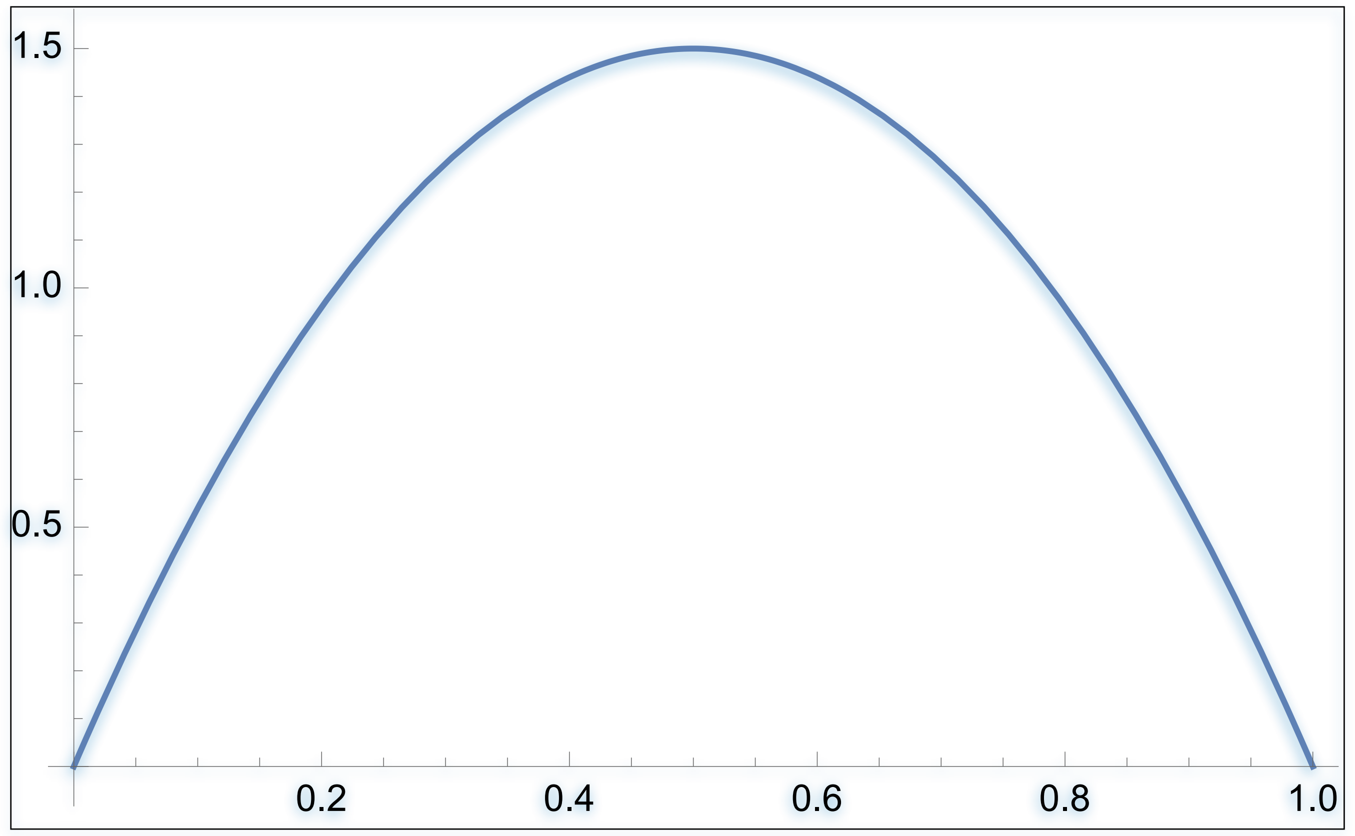
- $\Pr(a \leq X \leq b) = \Pr(a \leq X < b) = \Pr(a < X \leq b) = \Pr(a < X < b) = \int_a^b f_X(x)\,\mathrm{d}x$
- $\Pr(X = x) = 0$ for all $x \in \mathbb{R}$

Examples

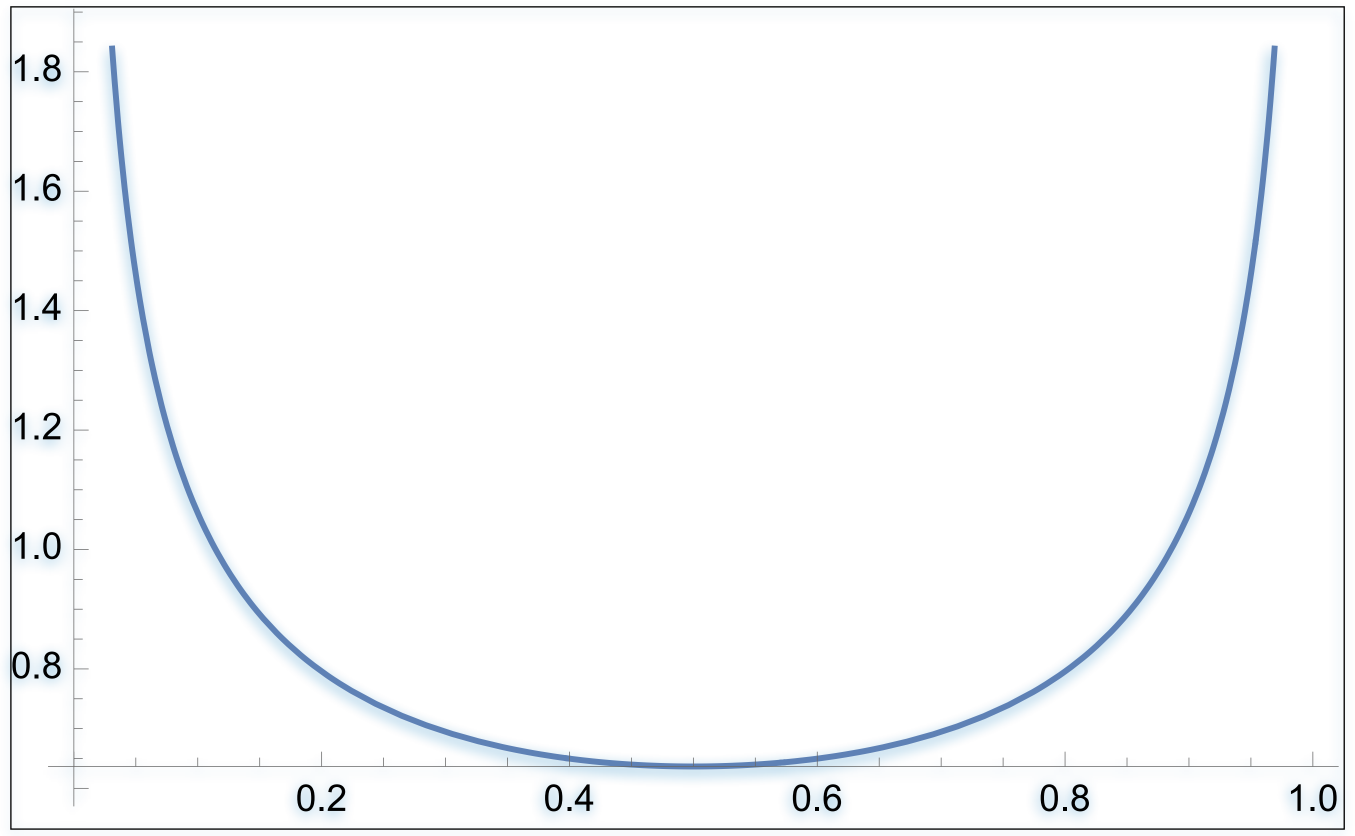
$f_X(x) = I_{[0,1]}(x)$

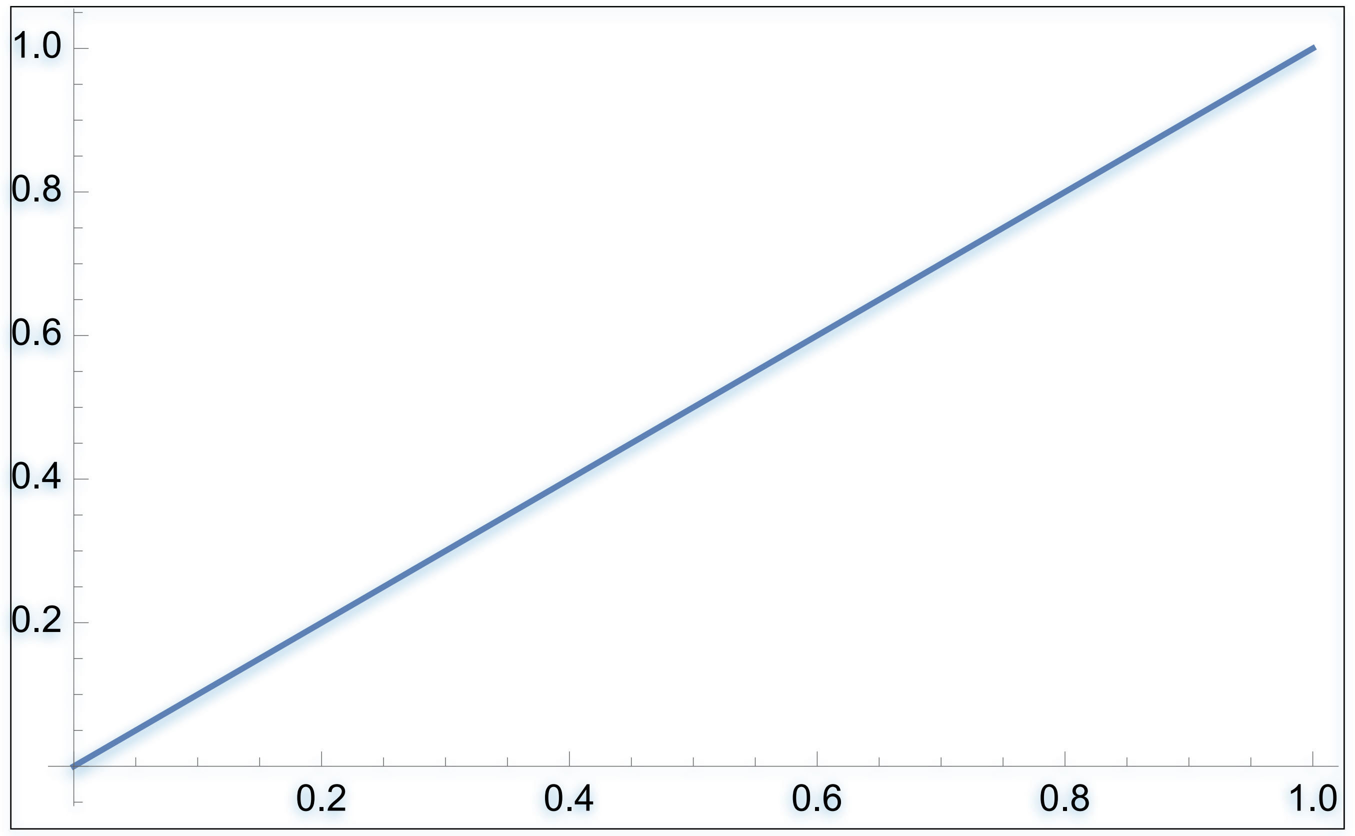


$f_Y(y) = 6y(1-y)I_{[0,1]}(y)$

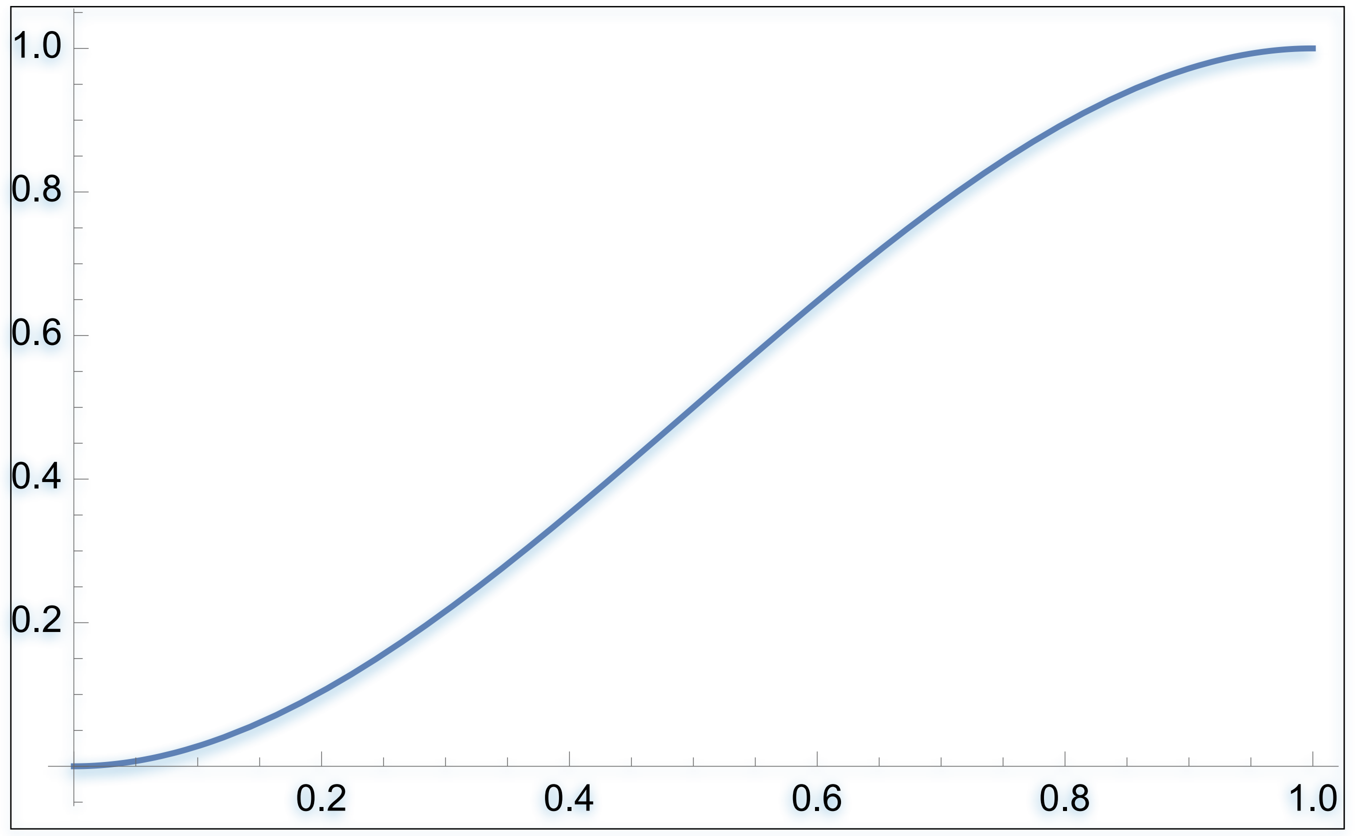


$$f_Z(z) = \frac{1}{\pi\sqrt{z(1-z)}} I_{[0,1]}(z)$$

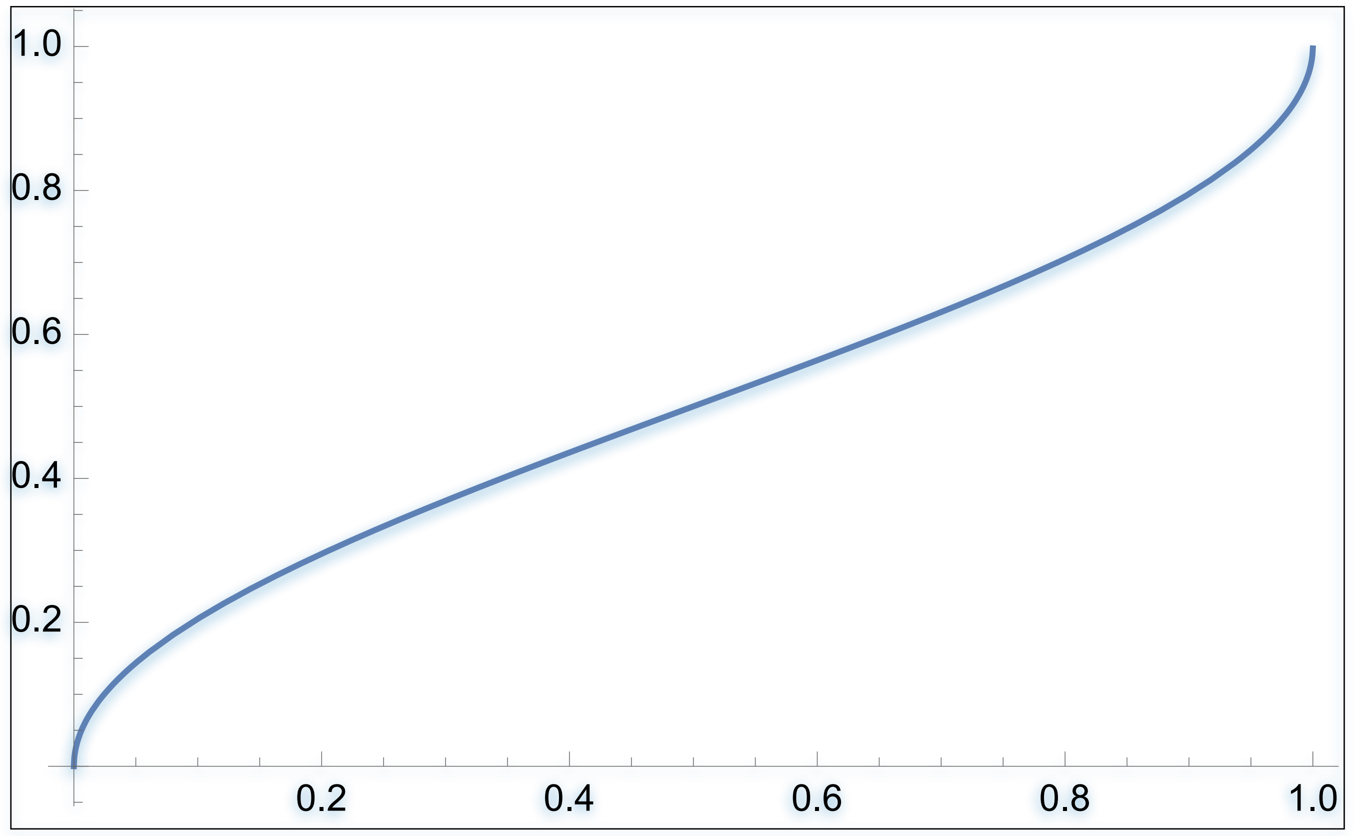




CDF $F_X(x)$



CDF $F_Y(y)$



CDF $F_Z(z)$

Continuous Probability Space*

- Let $\mathscr{I}$ be the collection of all open intervals in $\mathbb{R}$.
  The **Borel σ-field** $\mathscr{B}$ is the smallest σ-algebra that contains $\mathscr{I}$.
  - Each $B \in \mathscr{B}$, called a **Borel set**, is a countable $\cap, \cup, \setminus$ of open intervals.
- For Borel set $B \in \mathscr{B}$, the **Lebesgue integral**:

$$\mu(B) := \Pr(X \in B) = \int_B f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} I_B(x)\,\mathrm{d}F_X(x)$$

- $(\mathbb{R}, \mathscr{B}, \mu)$ is a well-defined probability space.
- If $g : \mathbb{R} \to \mathbb{R}$ is **Borel-measurable** ($\forall y \in \mathbb{R}, \{x \in \mathbb{R} \mid g(x) \leq y\} \in \mathscr{B}$), then $g(X)$ is also a random variable

Lebesgue Integration*

- Let $(\mathbb{R}, \mathscr{B}, \mu)$ be a **probability** (or measure) **space**.

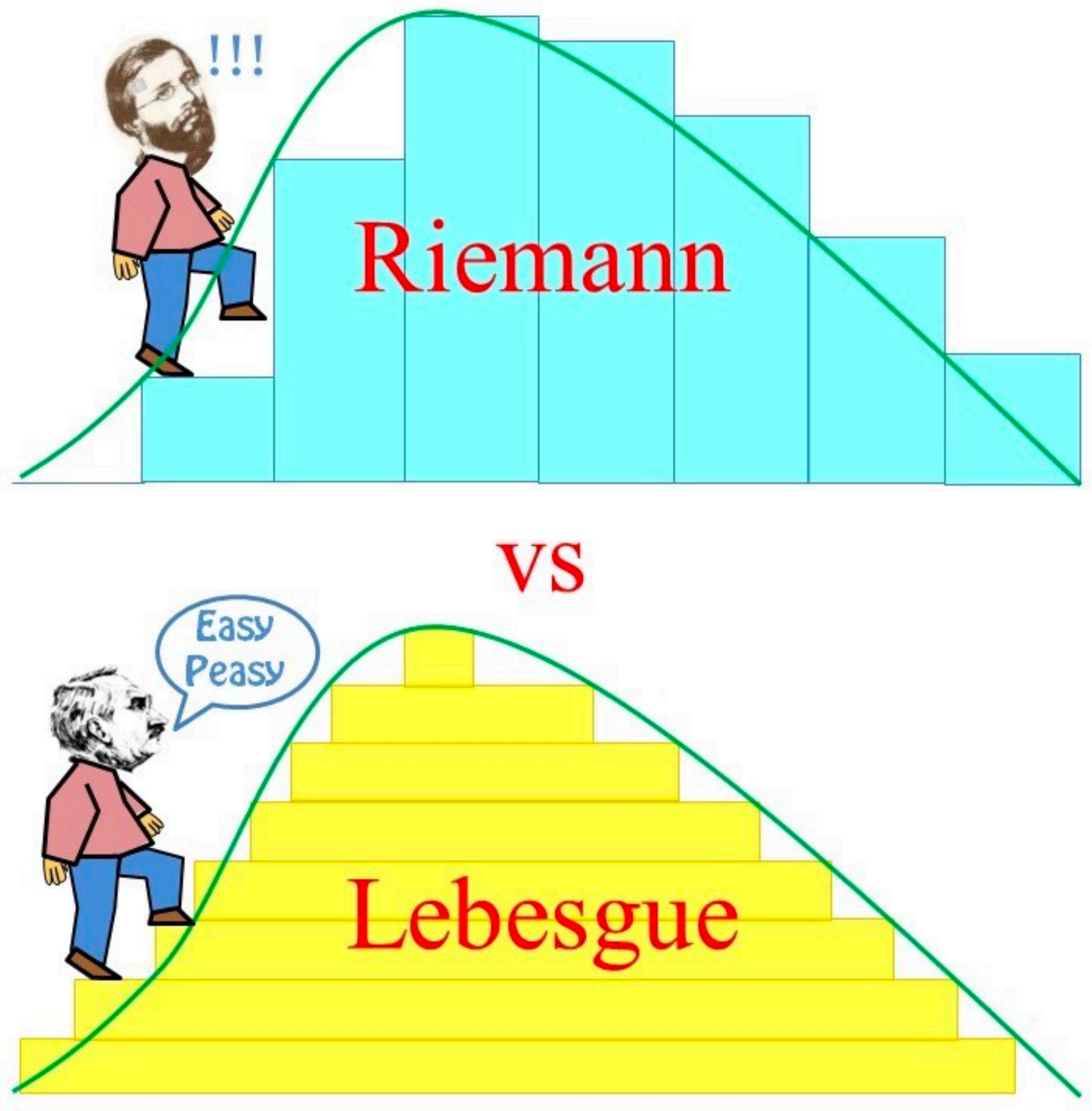
- Assume that f is (Borel-)**measurable** and **nonnegative**. For $B \in \mathscr{B}$, define:

$$f^*(t) := \mu(\{x \in B \mid f(x) > t\})$$

- The Lebesgue integral is defined by:

$$\int_B f \, \mathrm{d}\mu := \int_0^\infty f^*(t) \, \mathrm{d}t$$

- For general f, let $f = f^+ - f^-$ for nonnegative f^+, f^-



Pathological Examples*

- Non-measurable set:
  - Vitali set: $V \subseteq [0,1]$ that contains a single point from each coset of $\mathbb{Q}$ in $\mathbb{R}$
- Functions that are Lebesgue-integrable but not Riemann-integrable:
  - Dirichlet function (indicator of *rational* number) $f : \mathbb{R} \to \{0,1\}$
  - Characteristic function of a measurable set but is discontinuous everywhere
- Uncountable subset of $[0,1]$ with 0 measure:
  - Cantor set

Joint Distribution

- The <u>joint distribution function</u> of X and Y is the function $F_{X,Y} : \mathbb{R} \to [0,1]$ given by

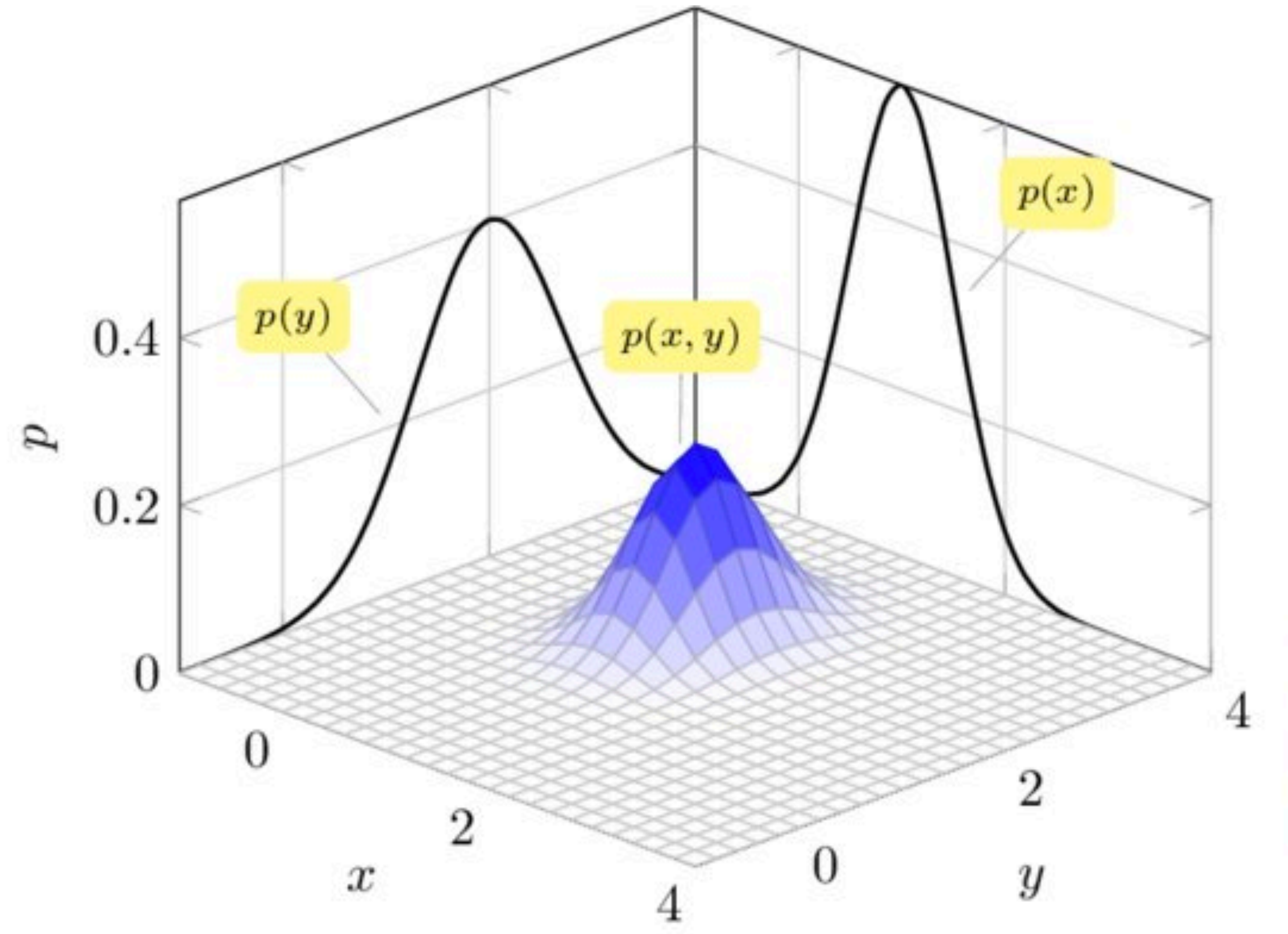
$$F_{X,Y}(x, y) = \Pr(X \leq x \cap Y \leq y)$$

- The random variables X and Y are **(jointly) continuous** with <u>**joint (probability) density function**</u> $f_{X,Y} : \mathbb{R}^2 \to [0,\infty)$ if for all $x, y \in \mathbb{R}$

$$F_{X,Y}(x, y) = \int_{v=-\infty}^{y} \int_{u=-\infty}^{x} f_{X,Y}(u, v)\, \mathrm{d}u\, \mathrm{d}v$$

- $f_{X,Y}(x, y) = \dfrac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$ (assuming $F_{X,Y}$ being sufficiently differentiable)

Marginal Distribution



- The marginal distribution functions of X and Y are:

$$F_X(x) = \Pr(X \leq x) = F_{X,Y}(x, \infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{u,Y}(u, y)\, \mathrm{d}y, \mathrm{d}u$$

$$F_Y(y) = \Pr(Y \leq y) = F_{X,Y}(\infty, y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_{X,Y}(x, v)\, \mathrm{d}x\, \mathrm{d}v$$

- The marginal density functions of X and Y are:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, \mathrm{d}y \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, \mathrm{d}x$$

Independence

- Random variables X and Y are called <u>independent</u> if:

$$X \leq x \text{ and } Y \leq y \text{ are independent events for all } x, y \in \mathbb{R}$$

which is equivalent to

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad \text{for all } x, y \in \mathbb{R}$$

which, for **continuous** random variables X and Y, is equivalent to

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{for all } x, y \in \mathbb{R}$$

- For Borel-measurable $g, h : \mathbb{R} \to \mathbb{R}$ (which means $g(X)$,$h(Y)$ are random variables),

$$X \text{ and } Y \text{ are independent} \implies g(X) \text{ and } h(Y) \text{ are independent}$$

(because $g(X)$ is Σ-measurable for Σ-measurable X and Borel-measurable g)

Conditional Distribution

- Let X be a continuous random variable. Let A be an event with $\Pr(A) > 0$. The <u>conditional distribution function</u> of X given the occurrence of A is:

$$F_{X|A}(x) = \Pr(X \leq x \mid A) = \int_{-\infty}^{x} f_{X|A}(u)\,\mathrm{d}u$$

where the density function $f_{X|A}(x) = \dfrac{\mathrm{d}F_{X|A}(x)}{\mathrm{d}x}$

- **Law of total probability**: For partition $B_1, \ldots, B_n$ of Ω with $\Pr(B_i) > 0$ for all i,

$$f_X(x) = \sum_{i=1}^{n} \Pr(B_i) f_{X|B_i}(x)$$

Proof: By taking derivative w.r.t. x on both sides of $\Pr(X \leq x) = \sum_{i=1}^{n} \Pr(B_i)\Pr(X \leq x \mid B_i)$

Conditional Distribution

- For (jointly) continuous random variables X and Y, the <u>conditional distribution function</u> of X given $Y = y$ is the function $F_{X|Y}(\cdot \mid y)$ given by

$$F_{X|Y}(x \mid y) = \Pr(X \leq x \mid Y = y) = \int_{-\infty}^{x} \frac{f_{X,Y}(u, y)}{f_Y(y)} \mathrm{d}u$$

- This definition makes sense because:

$$\Pr(X \leq x \mid y \leq Y \leq y + \mathrm{d}y) = \frac{\Pr((X \leq x) \cap (y \leq Y \leq y + \mathrm{d}y))}{\Pr(y \leq Y \leq y + \mathrm{d}y)}$$

$$= \frac{\int_{u=-\infty}^{x} f_{X,Y}(u, y)\mathrm{d}y\, \mathrm{d}u}{f_Y(y)\mathrm{d}y} = \int_{u=-\infty}^{x} \frac{f_{X,Y}(u, y)}{f_Y(y)} \mathrm{d}u$$

Conditional Distribution

- For (jointly) continuous random variables X and Y, the <u>conditional distribution function</u> of X given $Y = y$ is the function $F_{X|Y}(\,\cdot \mid y)$ given by

$$F_{X|Y}(x \mid y) = \Pr(X \leq x \mid Y = y) = \int_{-\infty}^{x} \frac{f_{X,Y}(u, y)}{f_Y(y)} \, \mathrm{d}u$$

- The <u>conditional density function</u> of $F_{X|Y}$ is given by

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \text{ for any } y \text{ such that } f_Y(y) > 0$$

Conditional Distribution

- For (jointly) continuous random variables X and Y, the <u>**conditional distribution function**</u> of X given $Y = y$ is the function $F_{X|Y}(\,\cdot \mid y)$ given by

$$F_{X|Y}(x \mid y) = \Pr(X \leq x \mid Y = y) = \int_{-\infty}^{x} \frac{f_{X,Y}(u, y)}{f_Y(y)}\,\mathrm{d}u$$

- **Law of total probability:** Let $B \subseteq \mathbb{R}$ be an event (a Borel set). For jointly continuous X and Y where Y has positive density over $\Omega_Y \subseteq \mathbb{R}$

$$\Pr(X \in B) = \int_{\Omega_Y} \Pr(X \in B \mid Y = y) \cdot f_Y(y)\,\mathrm{d}y = \int_{\Omega_Y} f_Y(y) \int_B \frac{f_{X,Y}(x, y)}{f_Y(y)}\,\mathrm{d}x\,\mathrm{d}y$$

Expectation

- The expectation (or mean) of a continuous random variable X with *pdf* f_X (and *CDF* F_X) is given by

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x\,\mathrm{d}F_X(x)$$

- The kth moment of X is accordingly defined by

$$\mathbb{E}[X^k] = \int_{-\infty}^{\infty} x^k f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x^k\,\mathrm{d}F_X(x)$$

- These quantities are well-defined when the integrals exist.

Double Counting

- If continuous random variable X takes on only **nonnegative** values (i.e. the **density** value $f_X(x) = 0$ if $x < 0$)

$$\mathbb{E}[X] = \int_0^\infty [1 - F_X(x)]\, \mathrm{d}x = \int_0^\infty \Pr(X > x)\, \mathrm{d}x$$

- **Proof**:

$$\int_0^\infty [1 - F_X(x)]\, \mathrm{d}x = \int_0^\infty \Pr(X > x)\, \mathrm{d}x = \int_0^\infty \left(\int_x^\infty f_X(u)\, \mathrm{d}u \right) \mathrm{d}x$$

$$= \int_{u=0}^\infty f_X(u) \int_{x=0}^u 1\, \mathrm{d}x\, \mathrm{d}u = \int_0^\infty u f_X(u)\, \mathrm{d}u = \mathbb{E}[X]$$

Change of Variables, *LOTUS*

- If X is a continuous random variable and $g(X)$ is a random variable, then

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$$

- **Proof**: First, assume $g \geq 0$. Denote $B_y = \{x \mid g(x) > y\}$.

$$\mathbb{E}[g(X)] = \int_0^{\infty} \Pr(g(X) > y)\,\mathrm{d}y = \int_0^{\infty} \int_{B_y} f_X(x)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty} f_X(x) \int_0^{g(x)} 1\,\mathrm{d}y\,\mathrm{d}x = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$$

For general $g : \mathbb{R} \to \mathbb{R}$, write $g = g^+ - g^-$ for $g^+, g^- \geq 0$. Then

$$\mathbb{E}[g(X)] = \mathbb{E}[g^+(X)] - \mathbb{E}[g^-(X)] = \int_{-\infty}^{\infty} g^+(x) f_X(x)\,\mathrm{d}x - \int_{-\infty}^{\infty} g^-(x) f_X(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$$

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: $\mathbb{E}[aX + b] = \int_{-\infty}^{\infty} (ax + b) f_X(x)\,\mathrm{d}x = a \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x + b \int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = a\mathbb{E}[X] + b$

$$\mathbb{E}[X + Y] = \iint_{\mathbb{R}^2} (x + y) f_{(X,Y)}(x, y)\,\mathrm{d}x\,\mathrm{d}y = \iint_{\mathbb{R}^2} x f_{(X,Y)}(x, y)\,\mathrm{d}x\,\mathrm{d}y + \iint_{\mathbb{R}^2} y f_{(X,Y)}(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} x f_{(X)}(x)\,\mathrm{d}x + \int_{-\infty}^{\infty} y f_{(X,Y)}(y)\,\mathrm{d}y = \mathbb{E}[X] + \mathbb{E}[Y]$$

Monotonicity of Expectation

- Let X and Y be random variables.
  - If $X \geq 0$, then $\mathbb{E}[X] \geq 0$
  - If $X \geq Y$, then $\mathbb{E}[X] \geq \mathbb{E}[Y]$

Proof: If $X \geq 0$, then $\mathbb{E}[X] = \int_0^\infty [1 - F_X(x)]\,\mathrm{d}x$ is clearly nonnegative

If $X \geq Y$, then let $Z = X - Y \geq 0$, then by linearity of expectation

$$0 \leq \mathbb{E}[Z] = \mathbb{E}[X] - \mathbb{E}[Y] \implies \mathbb{E}[X] \geq \mathbb{E}[Y]$$

Total Expectation

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

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

- $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$, thus $\Pr(B) = \int_{-\infty}^{\infty} \Pr(B \mid X = x) f_X(x) \, \mathrm{d}x$ by $\mathbb{E}[\mathbb{E}[I_B \mid X]]$

Proof: $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) \, \mathrm{d}x = \int_{-\infty}^{\infty} x \sum_{i=1}^{n} \Pr(B_i) f_{X|B_i}(x) \, \mathrm{d}x = \sum_{i=1}^{n} \mathbb{E}[X \mid B_i] \Pr(B_i)$

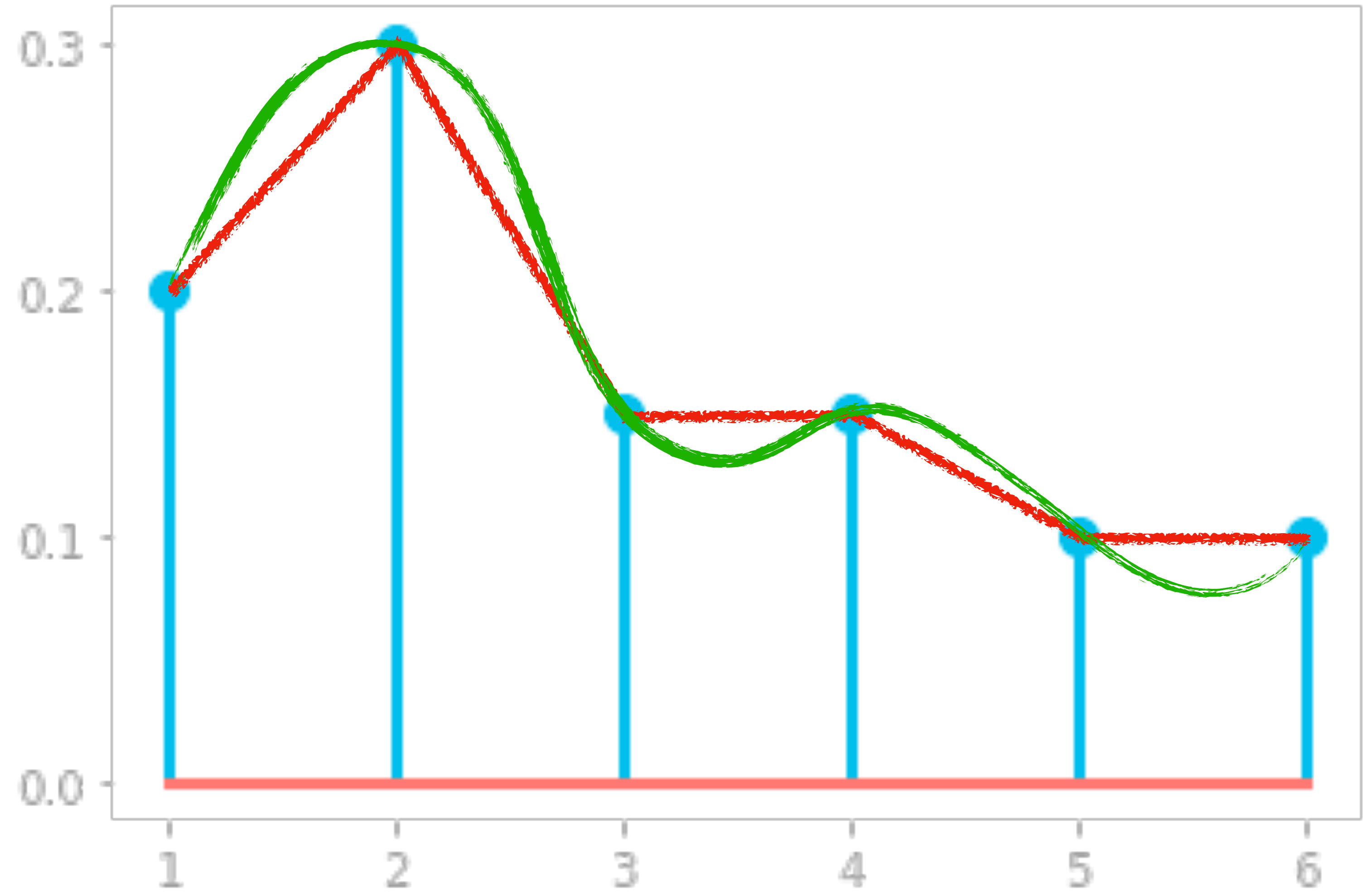
$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y] f_Y(y) \, \mathrm{d}y = \int_{-\infty}^{\infty} f_Y(y) \int_{-\infty}^{\infty} x \frac{f_{X,Y}(x, y)}{f_Y(y)} \, \mathrm{d}x \, \mathrm{d}y = \int_{-\infty}^{\infty} x f_X(x) \, \mathrm{d}x = \mathbb{E}[X]$$

Expectation of Product

- If random variables X and Y are **independent**, then
$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$
and hence $\mathbf{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$
- Consequently, $\mathbf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{Var}[X_i]$ for pairwise independent $X_1, \ldots, X_n$

Proof: By change of variable
$$\mathbb{E}[XY] = \iint_{\mathbb{R}^2} xyf_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y = \iint_{\mathbb{R}^2} xyf_X(x)f_Y(y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{-\infty}^{\infty} xf_X(x)\,\mathrm{d}x \int_{-\infty}^{\infty} yf_Y(y)\,\mathrm{d}y = \mathbb{E}[X]\mathbb{E}[Y]$$

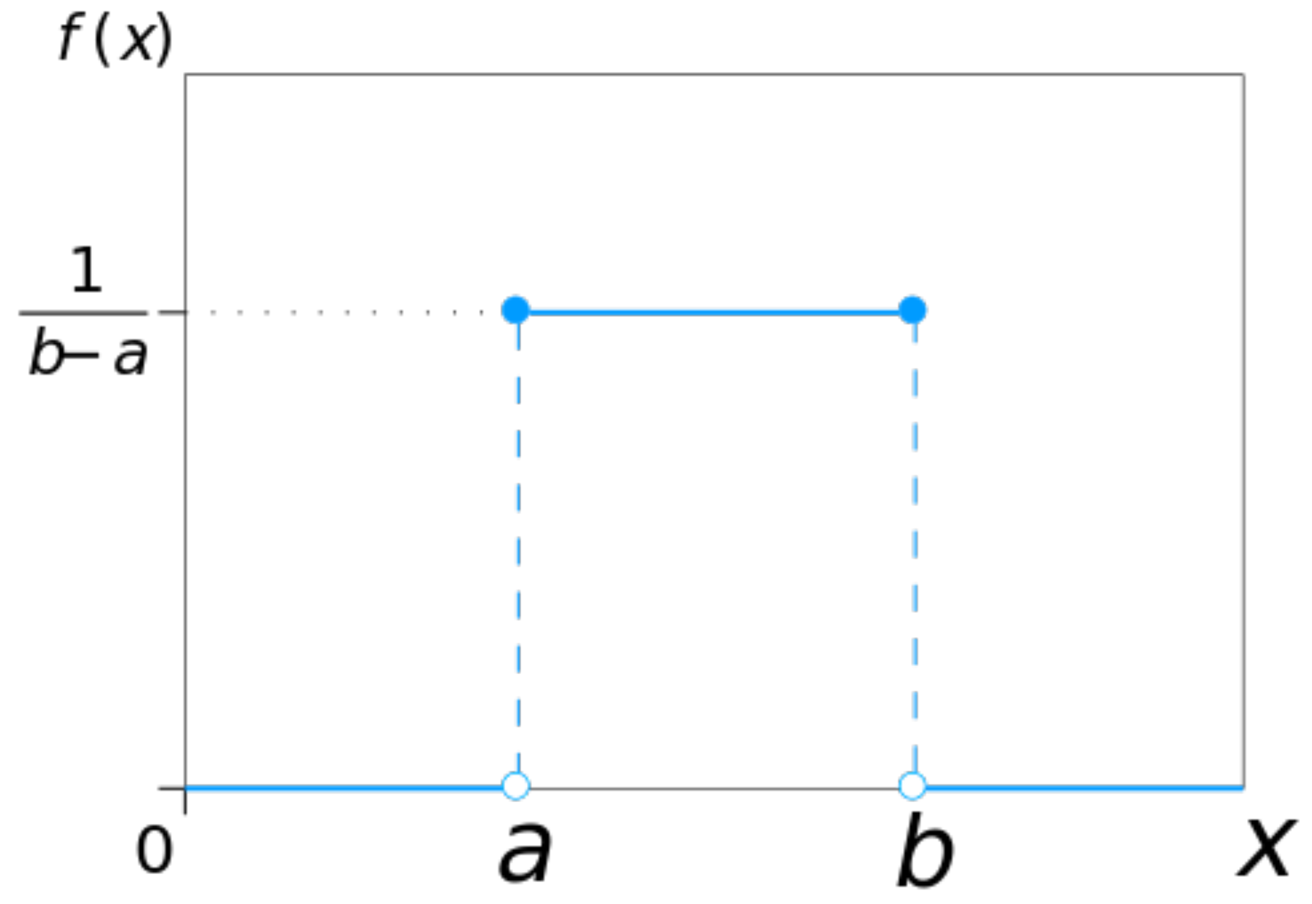
Continuous
Probability Distributions

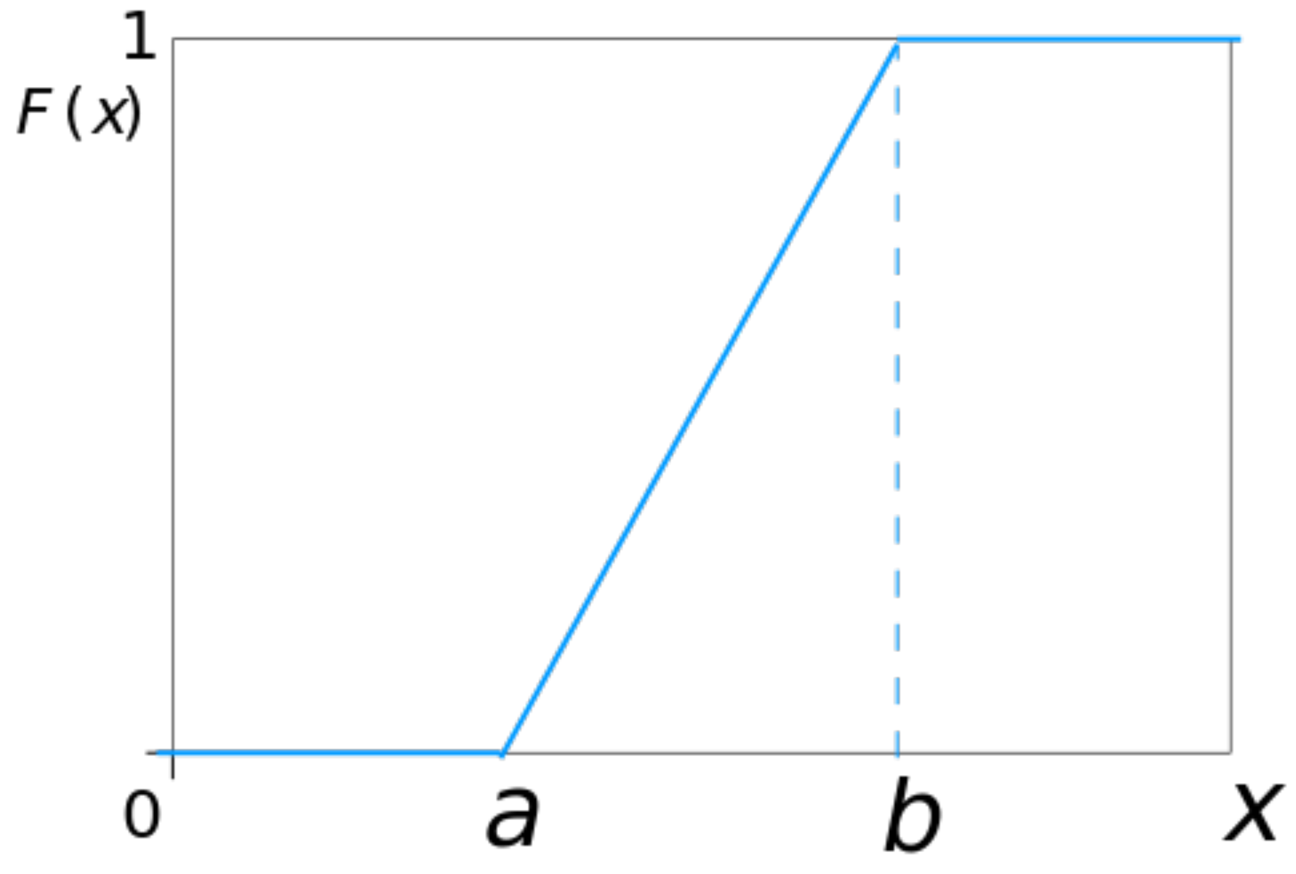


Continuous Uniform Distribution

- The random variable X is <u>uniform</u> on $[a, b]$ (or $[a, b)$, $(a, b]$, (a, b)), if it has

$$\text{pdf } f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \text{CDF } F(x) = \begin{cases} 0 & \text{if } x \le a \\ \frac{x-a}{b-a} & \text{if } a < x \le b \\ 1 & \text{if } x > b \end{cases}$$





- There is no uniform distribution on $[a, \infty)$, or on $(-\infty, b]$, or on $\mathbb{R}$

Continuous Uniform Distribution

- The random variable X is **uniform** on $[a, b]$ (or $[a, b)$, $(a, b]$, (a, b)), if it has

$$\text{pdf } f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \text{CDF } F(x) = \begin{cases} 0 & \text{if } x \le a \\ \frac{x-a}{b-a} & \text{if } a < x \le b \\ 1 & \text{if } x > b \end{cases}$$

- $\mathbb{E}[X] = \int_a^b \frac{x}{b-a} \mathrm{d}x = \frac{a+b}{2}$ (same as the discrete uniform distribution on $[a, b]$)
- $\mathbb{E}[X^2] = \int_a^b \frac{x^2}{b-a} \mathrm{d}x = \frac{a^2+b^2+ab}{3} \implies \mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{(b-a)^2}{12}$

(different from the discrete uniform distribution on $[a, b]$)

Rejection Sampling

- Let X be a uniform random variable on $[a, b]$. Then for any $[c, d] \subseteq [a, b]$,

$$\Pr(X \in [c, d]) = \frac{d - c}{b - a}$$

  and given that $X \in [c, d]$, the **conditional distribution** of X is uniform on $[c, d]$:

$$\Pr(X \leq x \mid X \in [c, d]) = \begin{cases} 0 & \text{if } x \leq c \\ \frac{x - c}{d - c} & \text{if } c < x \leq d \\ 1 & \text{if } x > d \end{cases}$$

- **Proof**: $\Pr(X \leq x \mid X \in [c, d]) = \dfrac{\Pr(X \in [a, x] \cap [c, d])}{\Pr(X \in [c, d])}$

Function of Random Variable

(Induced probability distribution)

- What is the *pdf* f_Y of $Y = g(X)$ for a continuous random variable X with *pdf* f_X and (Borel measurable) function $g : \mathbb{R} \to \mathbb{R}$? (Assume g is monotonically increasing)

- CDF of Y is: $F_Y(y) = \Pr(Y \leq y) = \Pr(g(X) \leq y) = \int_{\{x | g(x) \leq y\}} f_X(x)\,\mathrm{d}x$

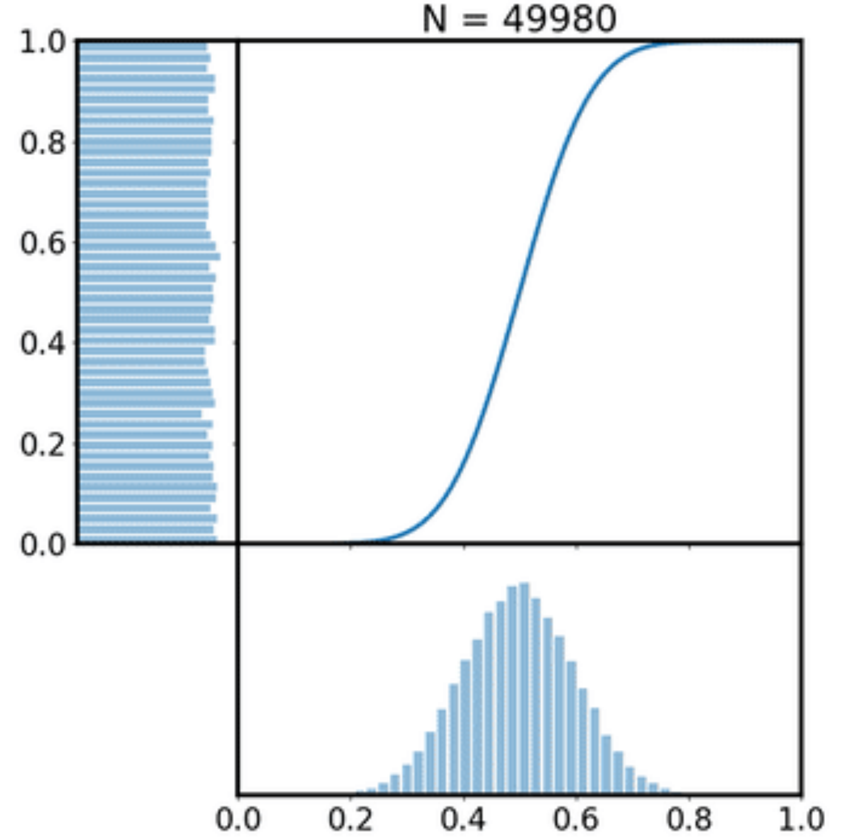
$$= \int_{-\infty}^{g^{-1}(y)} f_X(x)\,\mathrm{d}x = F_X\left(g^{-1}(y)\right)$$

- Taking derivative: $f_Y(y) = \dfrac{\mathrm{d}F_Y(y)}{\mathrm{d}y}$

$$= \frac{\mathrm{d}F_X\left(g^{-1}(y)\right)}{\mathrm{d}y} = f_X\left(g^{-1}(y)\right)\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} = f_X\left(g^{-1}(y)\right)\frac{1}{g'(y)}$$

Simulation of General Distribution

(Inverse transform sampling)



- Let random variable U be uniform on $[0,1]$. Let $F : \mathbb{R} \to [0,1]$ be a CDF.
  - If F is continuous, then the random variable $X = F^{-1}(U)$ has CDF F.
  - If F is the CDF of a integer-valued discrete random variable, then the discrete random variable $X = k$ iff $F(k-1) < U \leq F(k)$ has CDF F.
- **Idea**: The inverse function F^{-1} gives the **quantile** of the random variable X.

Proof: $\Pr(X \leq x) = \Pr(F^{-1}(U) \leq x) = \Pr(U \leq F(x)) = \dfrac{F(x) - 0}{1 - 0} = F(x)$

$$\Pr(X = k) = \Pr(F(k-1) < U \leq F(k)) = F(k) - F(k-1)$$

Stochastic Domination and Coupling

- If random variables X and Y satisfy $F_X(u) \leq F_Y(u)$ for all $u \in \mathbb{R}$, then we say that <u>X **dominates** Y **stochastically**</u>, and write $X \succeq_{\text{st}} Y$
- $X \succeq_{\text{st}} Y$ iff there is a <u>coupling</u> (X', Y') of X and Y with marginal distributions $F_{X'} = F_X$ and $F_{Y'} = F_Y$ respectively, such that $\Pr(X' \geq Y') = 1$

Proof: Let random variable U be uniform on $[0,1]$.

Let $X' = F_X^{-1}(U)$ and $Y' = F_Y^{-1}(U)$ (or if X, Y are discrete, $\begin{array}{l} X' = k \text{ iff } F_X(k-1) < U \leq F_X(k) \\ Y' = k \text{ iff } F_Y(k-1) < U \leq F_Y(k) \end{array}$).

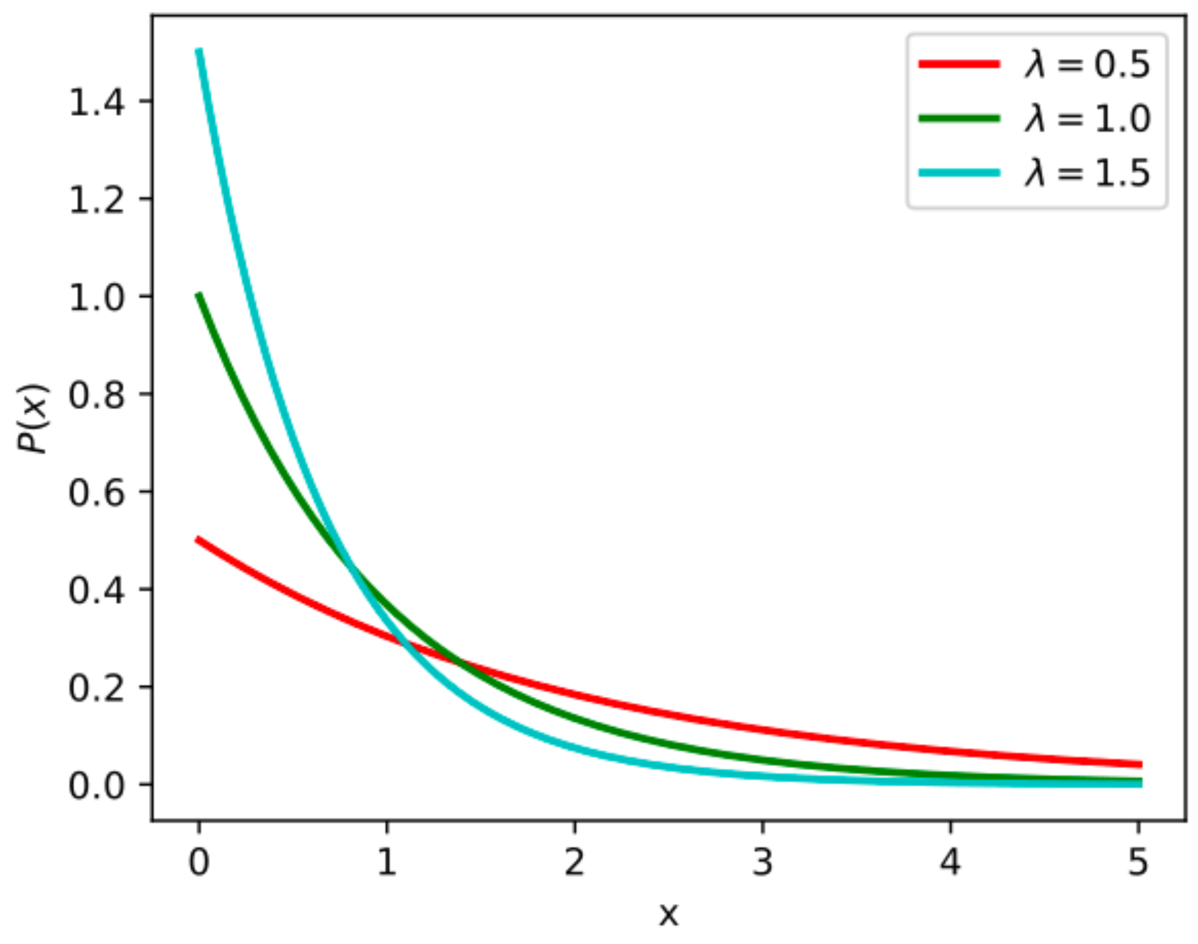
By inverse transform sampling, marginal distributions $F_{X'} = F_X$ and $F_{Y'} = F_Y$

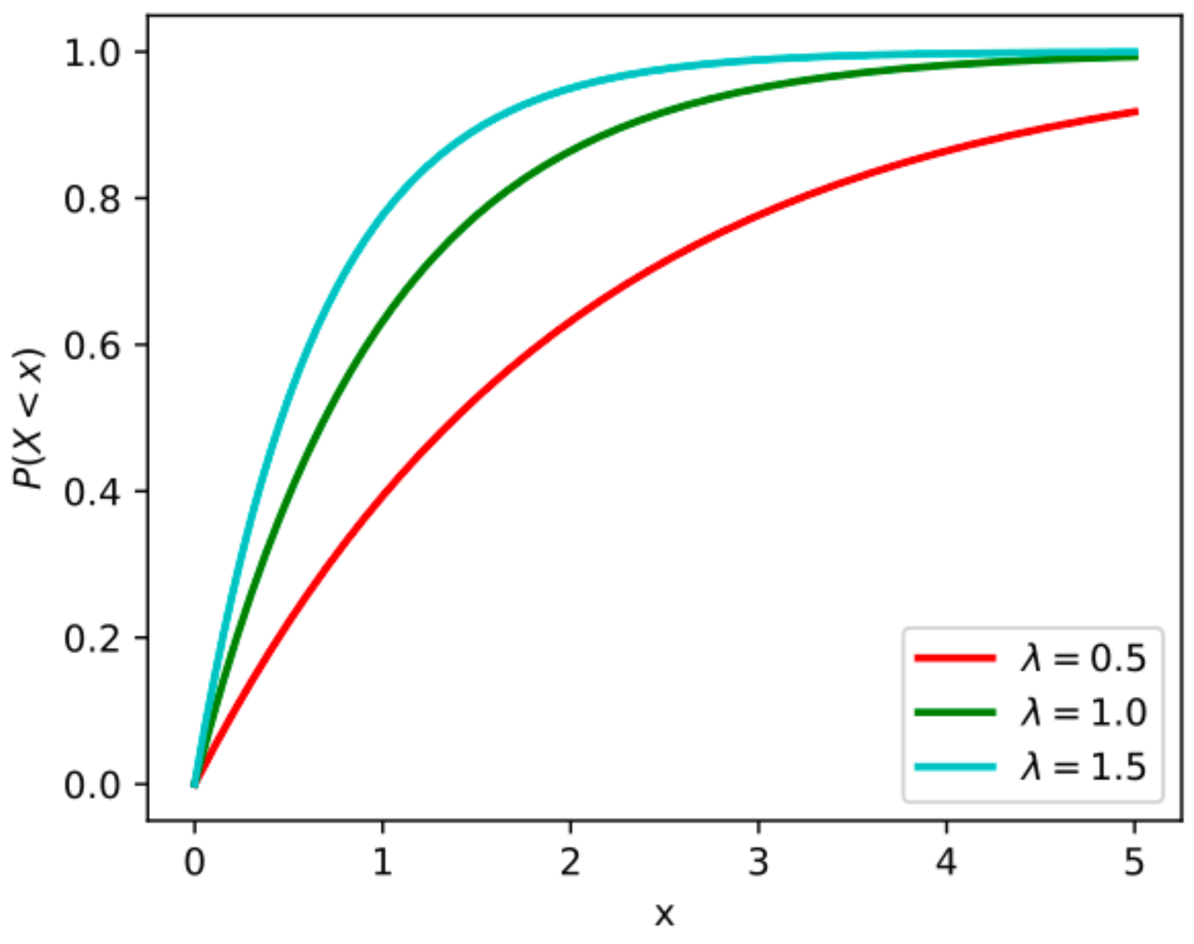
And $X' = F_X^{-1}(U) \geq F_Y^{-1}(U) = Y'$ since $F_X(u) \leq F_Y(u)$ for all $u \in \mathbb{R}$

Exponential Distribution

- The random variable X is exponential with parameter $\lambda > 0$, if it has

$$\text{pdf } f(x) = \lambda e^{-\lambda x} \quad \text{and} \quad \text{CDF } F(x) = 1 - e^{-\lambda x}, \quad \text{for } x \geq 0$$





- **Continuous limit of geometric distribution:** an i.i.d. Bernoulli trial (with $p = \lambda\delta$) being performed after every δ time elapse, let X be the time of the first success

$$\Pr(X > x) = (1 - p)^{x/\delta} = (1 - \lambda\delta)^{x/\delta} \quad \to e^{-\lambda x} \quad \text{as } \delta \downarrow 0$$

Exponential Distribution

- The random variable X is exponential with parameter $\lambda > 0$, if it has

$$\textbf{pdf}\ \ f(x) = \lambda \mathrm{e}^{-\lambda x} \quad \text{and} \quad \textbf{CDF}\ \ F(x) = 1 - \mathrm{e}^{-\lambda x}, \quad \text{for } x \geq 0$$

- $\mathbb{E}[X] = \int_0^\infty x\lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x = -\int_0^\infty x\,\mathrm{d}\mathrm{e}^{-\lambda x} = \left(-x\mathrm{e}^{-\lambda x}\right)\Big|_0^\infty + \int_0^\infty \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda}$
- alternatively, $\mathbb{E}[X] = \int_0^\infty (1 - F(x))\,\mathrm{d}x = \int_0^\infty \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda}$
- $\mathbb{E}[X^2] = \int_0^\infty x^2\lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{2}{\lambda^2} \implies \mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}$

Exponential Distribution

- **Exponential** random variable X is memoryless: for $s, t \geq 0$

$$\Pr(X > s + t \mid X > t) = \Pr(X > s)$$

It's the only memoryless continuous random variable

Proof: $\Pr(X > s + t \mid X > t) = \dfrac{\Pr(X > s + t)}{\Pr(X > t)} = \dfrac{\mathrm{e}^{-\lambda(s+t)}}{\mathrm{e}^{-\lambda t}} = \mathrm{e}^{-\lambda s} = \Pr(X > s)$

- If $X_1, \ldots, X_n$ are **independent exponential** random variables with parameters $\lambda_1, \ldots, \lambda_n$, respectively, then $\min_{1 \leq i \leq 0} X_i$ is exponential with parameter $\sum_{i=1}^{n} \lambda_i$.

Proof: $\Pr\left(\min_{1 \leq i \leq n} X_i > x\right) = \Pr\left(\bigcap_{1 \leq i \leq n} (X_i > x)\right) = \prod_{i=1}^{n} \Pr(X_i > x) = \prod_{i=1}^{n} \mathrm{e}^{-x\lambda_i} = \mathrm{e}^{-x \sum_{i=1}^{n} \lambda_i}$

Poisson Point Process

(Stochastic counting process with exponential interarrival)

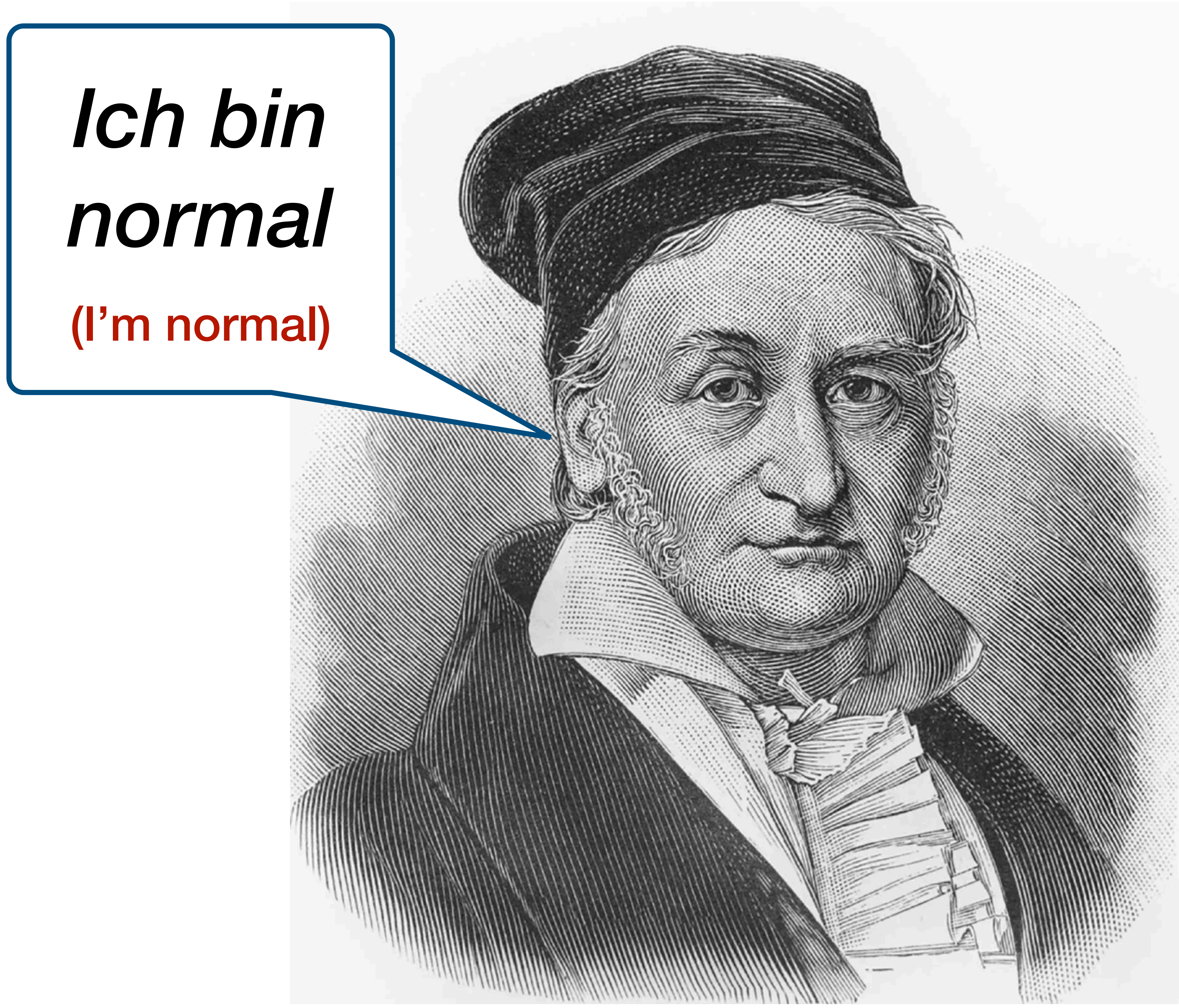
- The **Poisson process** $\{N(t) \mid t \geq 0\}$ with rate $\lambda > 0$ is a continuous time process defined as follows —— imagine we have such a clock:
  - $N(t)$ counts the number of times the clock rings up to time t, initially $N(0) = 0$;
  - The time elapse (**interarrival time**) between any two consecutive ringings (including the time elapse before 1st ringing) is independent exponential with parameter λ

- Due to **memoryless** and **minimum:** The process defined by k independent clocks with the same rate λ is equivalent to the 1-clock process with rate $k\lambda$

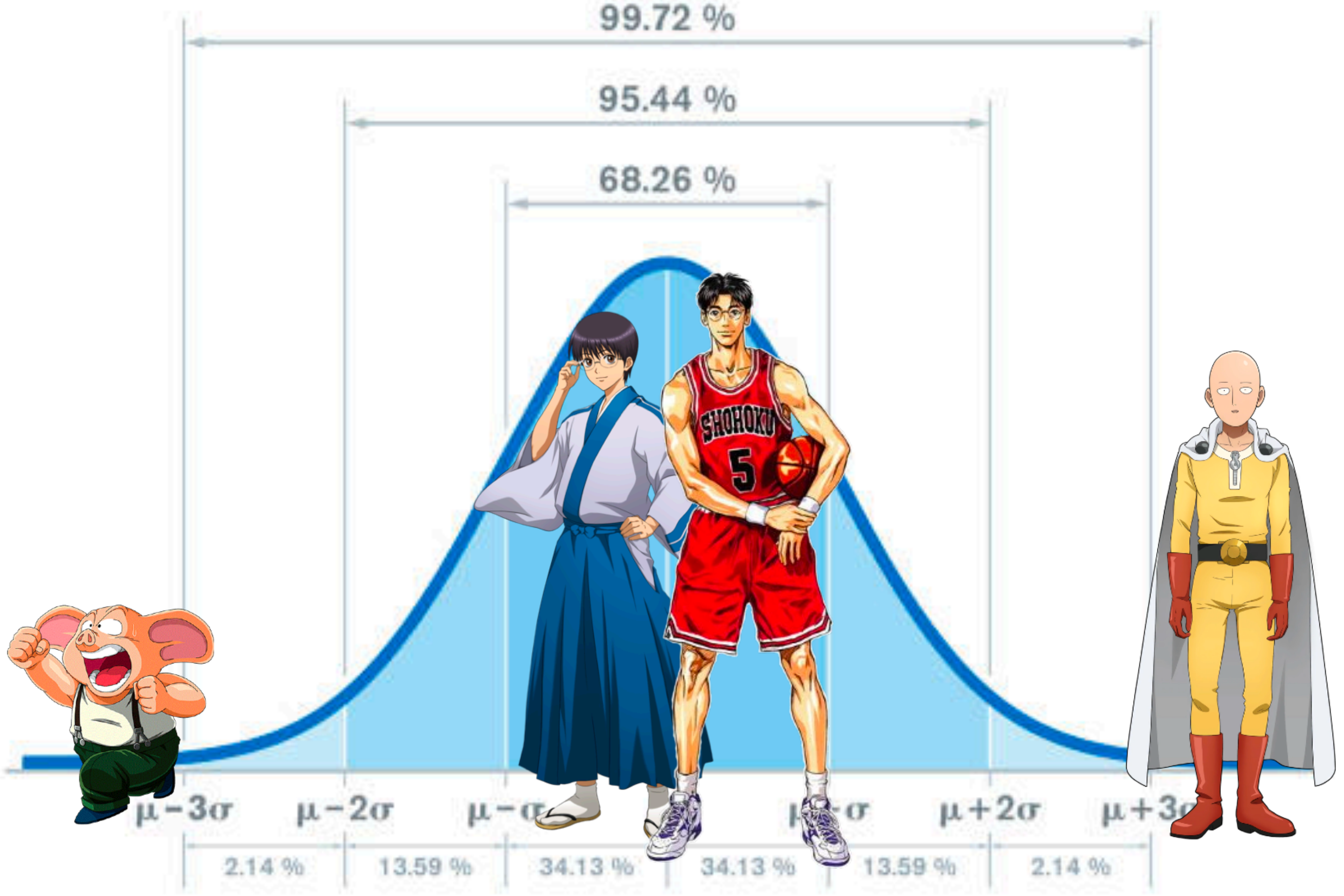
- (Poisson distribution) For any $t, s \geq 0$ and any integer $n \geq 0$,

$$\Pr(N(t+s) - N(s) = n) = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

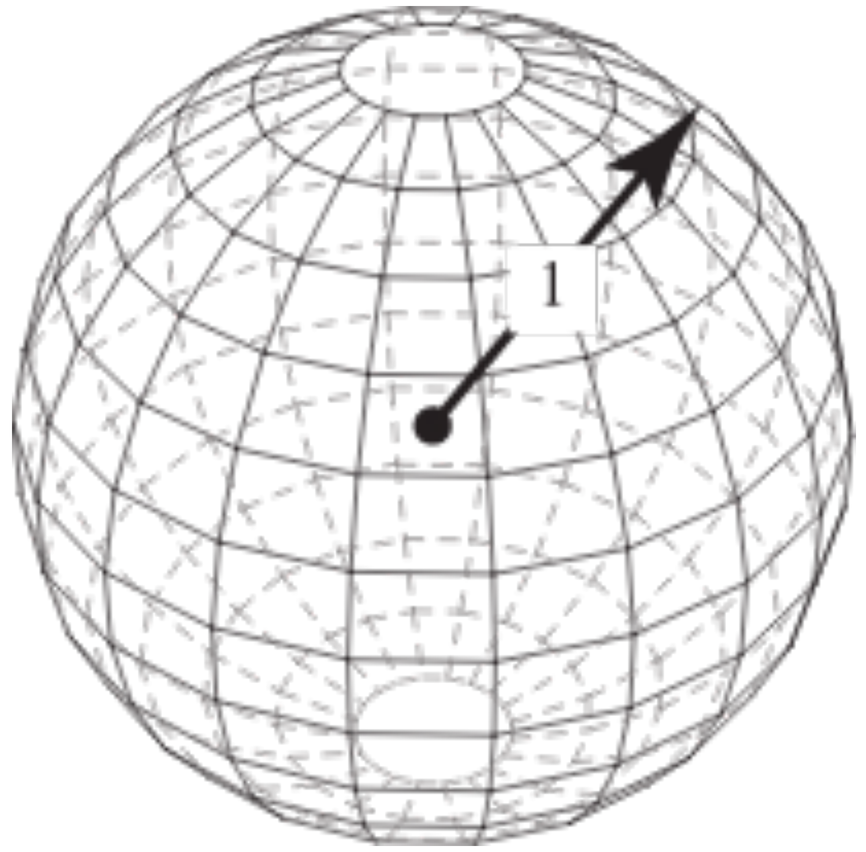
Gaussion Distribution



Normal Distribution



Generating a Uniform Unit Vector



- **Goal**: Draw a **uniform** random **unit** vector $U \in \mathbb{R}^n$ with $\|U\|_2 = 1$

- A standard approach is to draw *i.i.d.* $X_1, \ldots, X_n \in \mathbb{R}$ and then normalize:

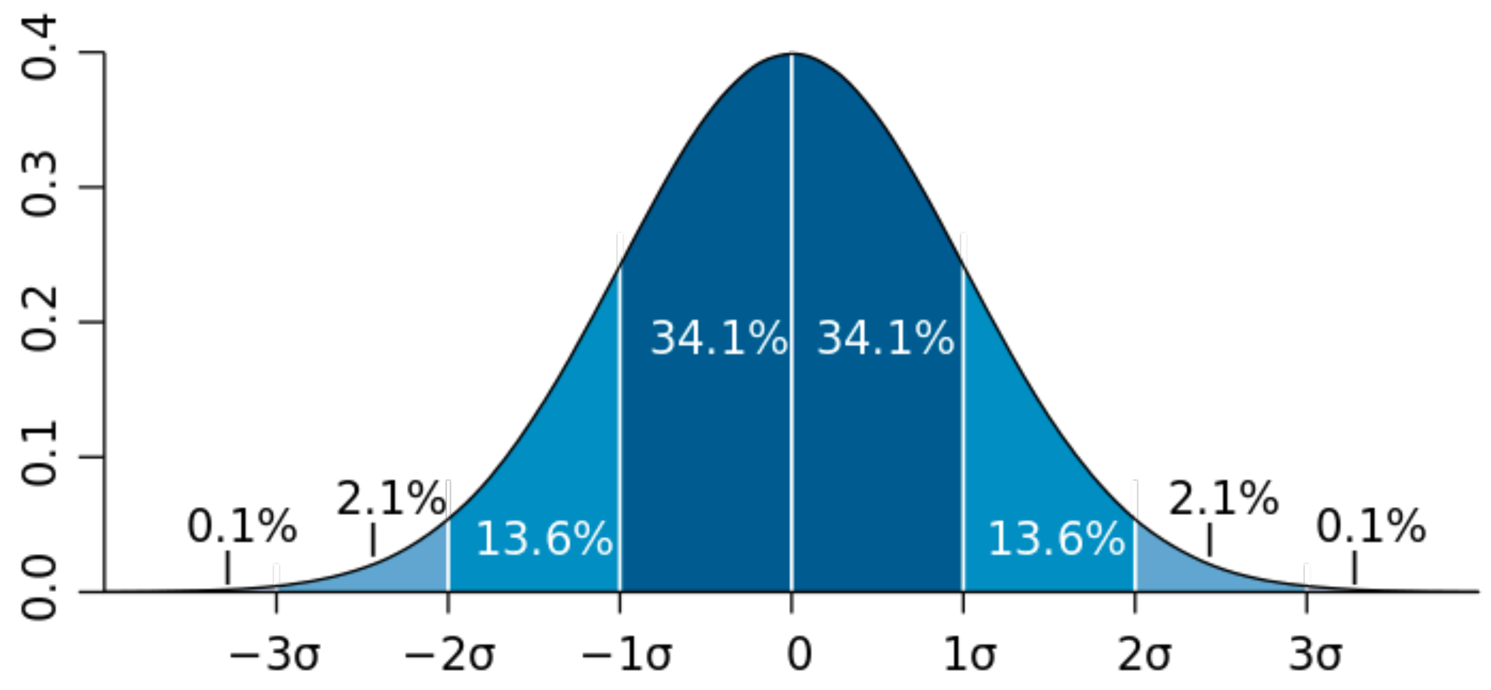
$$U = \frac{(X_1, \ldots, X_n)}{\|(X_1, \ldots, X_n)\|_2} \text{ so that } \|U\|_2 = 1$$

- Such U is uniform on the **unit sphere**, if the joint density of $(X_1, \ldots, X_n)$ is **spherically symmetric:**

$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} f_{X_i}(y_i) = f_{(X_1,\ldots,X_n)}(y_1, \ldots, y_n), \text{ given } \|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2$$

- This suggests random variable with density $f(x) \propto \exp(-c \cdot x^2)$

Normal Distribution (正态分布)
(the most important continuous distribution)



- A continuous random variable X is a **normal** or **Gaussian** random variable with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, if it has the density function:

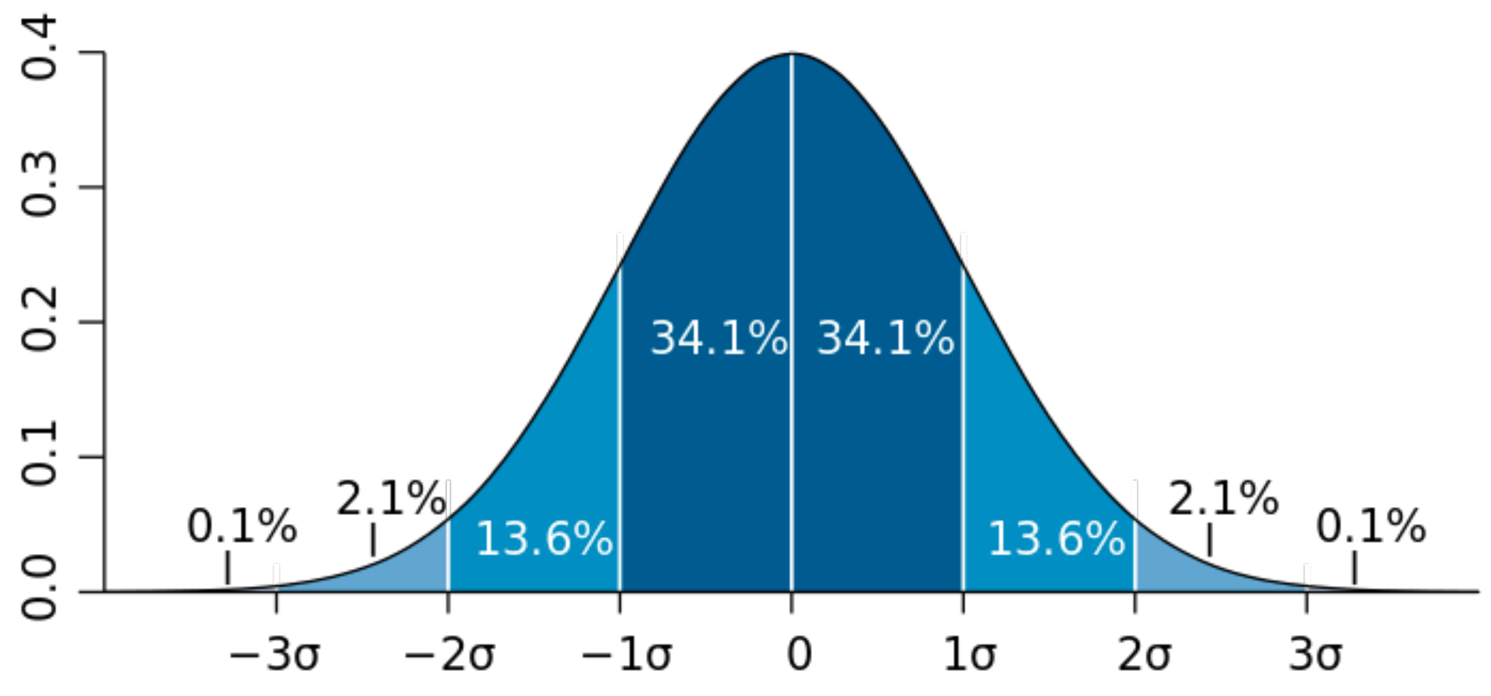
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad -\infty < x < \infty$$

- The distribution is denoted by $N(\mu, \sigma^2)$. If $\mu = 0$ and $\sigma = 1$, then

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty$$

is the density of the **standard normal distribution** (标准正态分布) $N(0,1)$

Normal Distribution (正态分布)
(the most important continuous distribution)

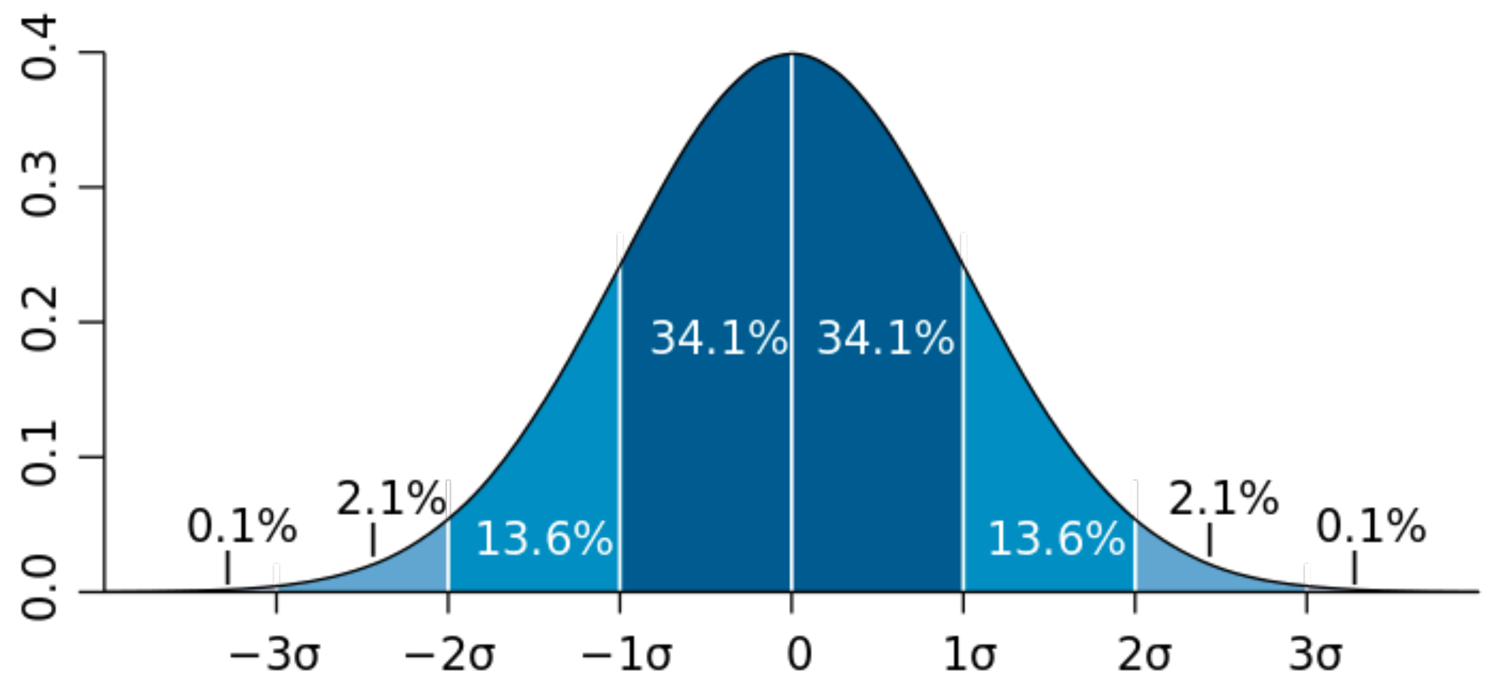


- It defines a probability distribution because the **Gaussian integral**

$$\int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}$$

- The normal distribution is important because it naturally arrises in many ways:
  - (*de Moivre–Laplace theorem*) The normal distribution is a continuous limit of the binomial distribution $B(n, p)$ as $n \to \infty$
  - (*central limit theorem*, *CLT*) The sum of large number of independent random variables is approximately normally distributed.

Normal Distribution (正态分布)



- If X is a **normal** random variable with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, then

$$\mathbb{E}[X] = \mu \quad \text{and} \quad \mathbf{Var}[X] = \sigma^2$$

Proof: $\mathbb{E}[X] = \mu$ because the density $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is symmetric w.r.t. $x = \mu$

$$\mathbf{Var}[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \,\mathrm{d}x = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 \mathrm{e}^{-y^2/2} \,\mathrm{d}y$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left(-y\mathrm{e}^{-y^2/2}\right)\Big|_{-\infty}^{\infty} + \sigma^2 \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-y^2/2} \,\mathrm{d}y}_{=1} = \sigma^2$$

Linear Transformation of Normal Distribution

- If $X \sim N(\mu, \sigma^2)$, then for any constants $a \neq 0$ and b, the random variable

$$Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$$

Proof: Suppose $a > 0$. $F_Y(y) = \Pr(Y \leq y) = \Pr(X \leq (y-b)/a) = F_X\left(\frac{y-b}{a}\right)$

By chain rule for derivative: $f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$

Suppose $a < 0$. $F_Y(y) = 1 - F_X\left(\frac{y-b}{a}\right)$ and $f_Y(y) = \frac{1}{-a} f_X\left(\frac{y-b}{a}\right)$

Thus, $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi}\,|a|\,\sigma} \mathrm{e}^{-\frac{(y-a\mu-b)^2}{2a^2\sigma^2}}$

Linear Transformation of Normal Distribution

- If $X \sim N(\mu, \sigma^2)$, then for any constants $a \neq 0$ and b, the random variable

$$Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$$

$\Longrightarrow$

- If $X \sim N(\mu, \sigma^2)$, then $\dfrac{X - \mu}{\sigma} \sim N(0,1)$
- If $X \sim N(0,1)$, then $\sigma X + \mu \sim N(\mu, \sigma)$

Linear Transformation of Normal Distribution

- If $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ are independent, then

$$X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2)$$

Convolution (卷积)

- The <u>convolution</u> of density functions f_X and f_Y is a function $f_X * f_Y$ given by

$$f_X * f_Y(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\,\mathrm{d}y$$

- If continuous random variables X and Y are **independent**, then

$$f_{X+Y} = f_X * f_Y$$

- **Proof**: $\Pr(X+Y \le z) = \iint_{u+v\le z} f_X(u) f_Y(v)\,\mathrm{d}u\,\mathrm{d}v = \int_{u=-\infty}^{\infty}\int_{v=-\infty}^{z-u} f_X(u) f_Y(v)\,\mathrm{d}v\,\mathrm{d}u$

$$= \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{z} f_X(x) f_Y(y-x)\,\mathrm{d}y\,\mathrm{d}x = \int_{y=-\infty}^{z}\int_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x\,\mathrm{d}y$$

Linear Transformation of Normal Distribution

- If $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ are independent, then

$$X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2)$$

Proof: By convolution

$$f_{X+Y}(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x = \frac{1}{2\pi\sigma\tau}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}-\frac{(z-x-\nu)^2}{2\tau^2}}\,\mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}}\mathrm{e}^{-\frac{(z-\mu-\nu)^2}{2(\delta^2+\tau^2)}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}}}\exp\left[-\frac{\left(x-\frac{\sigma^2(z-\nu)+\tau^2\mu}{\sigma^2+\tau^2}\right)^2}{2\cdot\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}}\right]\mathrm{d}x$$

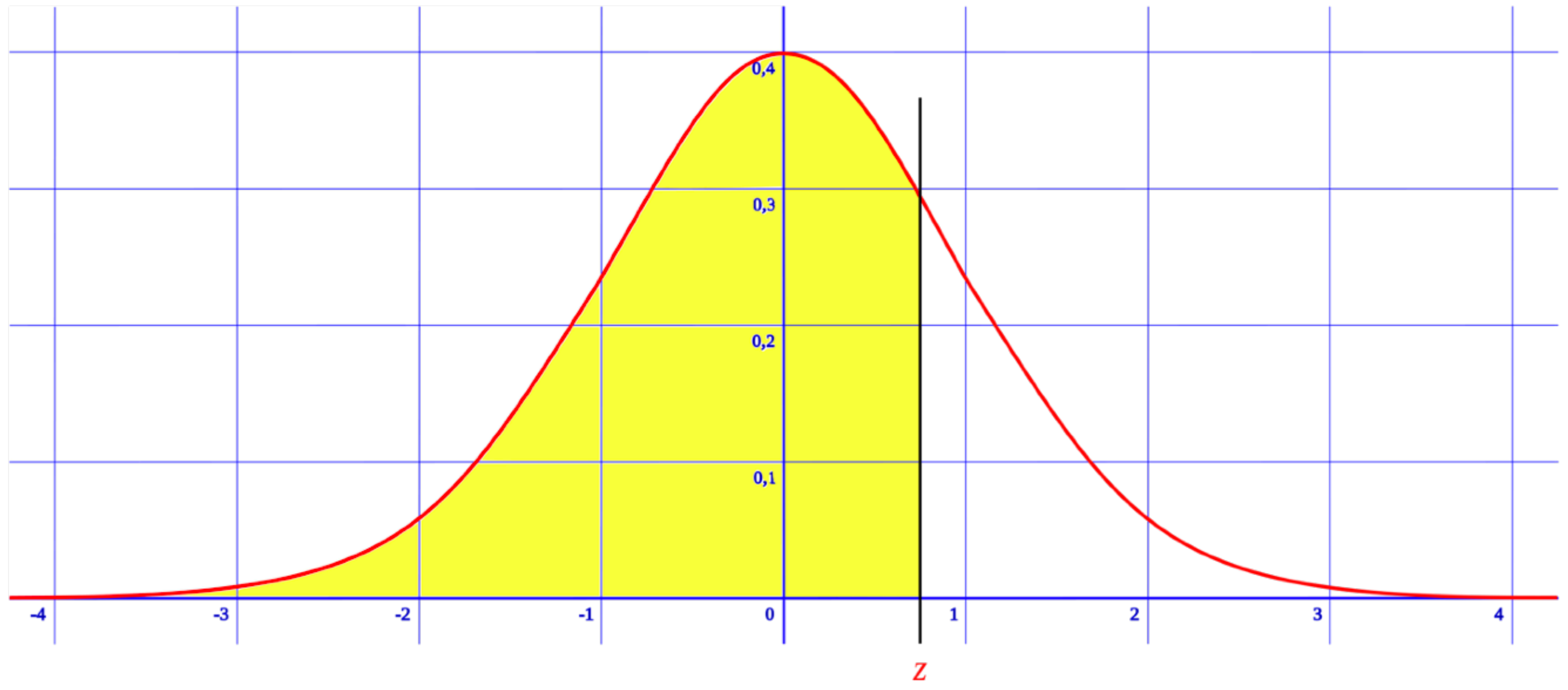
=1

Standard Normal Distribution

- The **PDF** and **CDF** of standard normal random variable $X \sim N(0,1)$ are given respectively by:
  - $\phi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$
  - $\Phi(z) = \Pr(X \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \, \mathrm{d}x$
- The CDF $\Phi(z)$ does not have a closed-form expression, but is usually computed numerically and given by a **standard normal table**

Standard Normal Table

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$



- $\Phi(-z) = 1 - \Phi(z)$ by symmetry
- For $X \sim N(\mu, \sigma^2)$ and any x:

$$\Pr(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

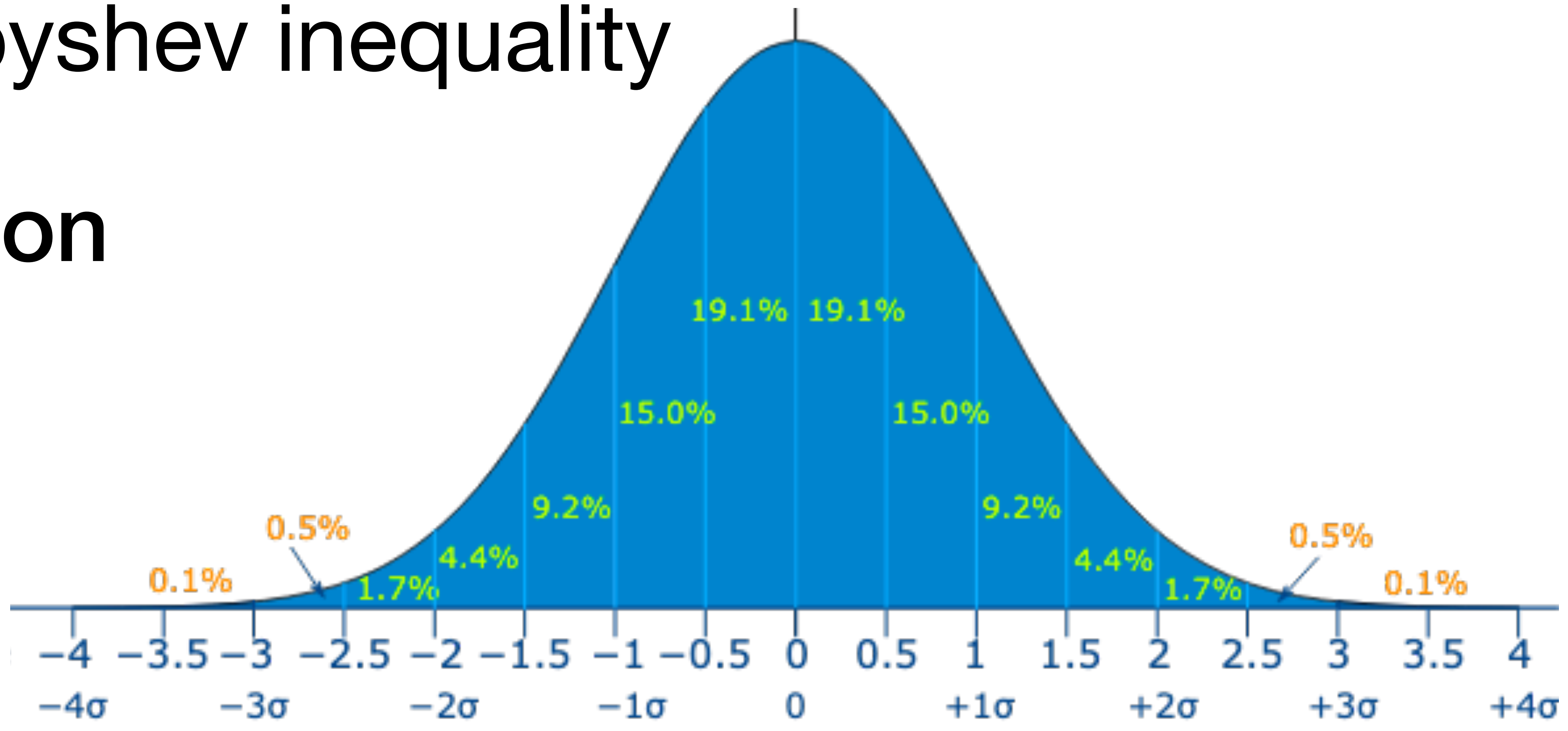
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

Large Deviation (Concentration) Bound

- If $X \sim N(\mu, \sigma^2)$, then for any $a > 0$,

$$\Pr(|X - \mu| \geq a\sigma) \leq 2\mathrm{e}^{-a^2/2}$$

- Still numerically weaker than the **“68-95-99.7” empirical rule**
- Much sharper than the $1/a^2$ bound by Chebyshev inequality
- Proved using the **moment generating function**



68–95–99.7 rule

Moment Generating Function

- The **moment generating function** (**MGF**) of a random variable X is

$$M_X(t) = \mathbb{E}[\mathrm{e}^{tX}]$$

- Expand e^{tX} by Maclaurin series: $M_X(t) = \mathbb{E}[\mathrm{e}^{tX}] = \sum_{k \geq 0} \frac{t^k \mathbb{E}[X^k]}{k!}$

- In fact, the kth moment is given by $\mathbb{E}[X^k] = M_X^{(k)}(0)$ where $M_X^{(k)}(0)$ is the kth derivative of $M_X(t)$ evaluated at $t = 0$

- If $M_X(t) = M_Y(t)$ for all $t \in [-\delta, \delta]$ for some $\delta > 0$, then X, Y are identically distributed.

MGF of Normal Distribution

- The **moment generating function** of **standard normal** $X \sim N(0,1)$ is

$$M_X(t) = \mathbb{E}[\mathrm{e}^{tX}] = \mathrm{e}^{t^2/2}$$

Proof: $M_X(t) = \mathbb{E}[\mathrm{e}^{tX}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{tx}\mathrm{e}^{-x^2/2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2+tx}\,\mathrm{d}x$

$$= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{t^2/2} \int_{-\infty}^{\infty} \mathrm{e}^{(x-t)^2/2}\,\mathrm{d}x = \mathrm{e}^{t^2/2}$$ (because $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{(x-t)^2/2}\,\mathrm{d}x = 1$)

Large Deviation (Concentration) Bound

- If $X \sim N(\mu, \sigma^2)$, then for any $a > 0$,

$$\Pr(|X - \mu| \geq a\sigma) \leq 2e^{-a^2/2}$$

Proof: Consider the standardized $Z = (X - \mu)/\sigma \sim N(0,1)$.

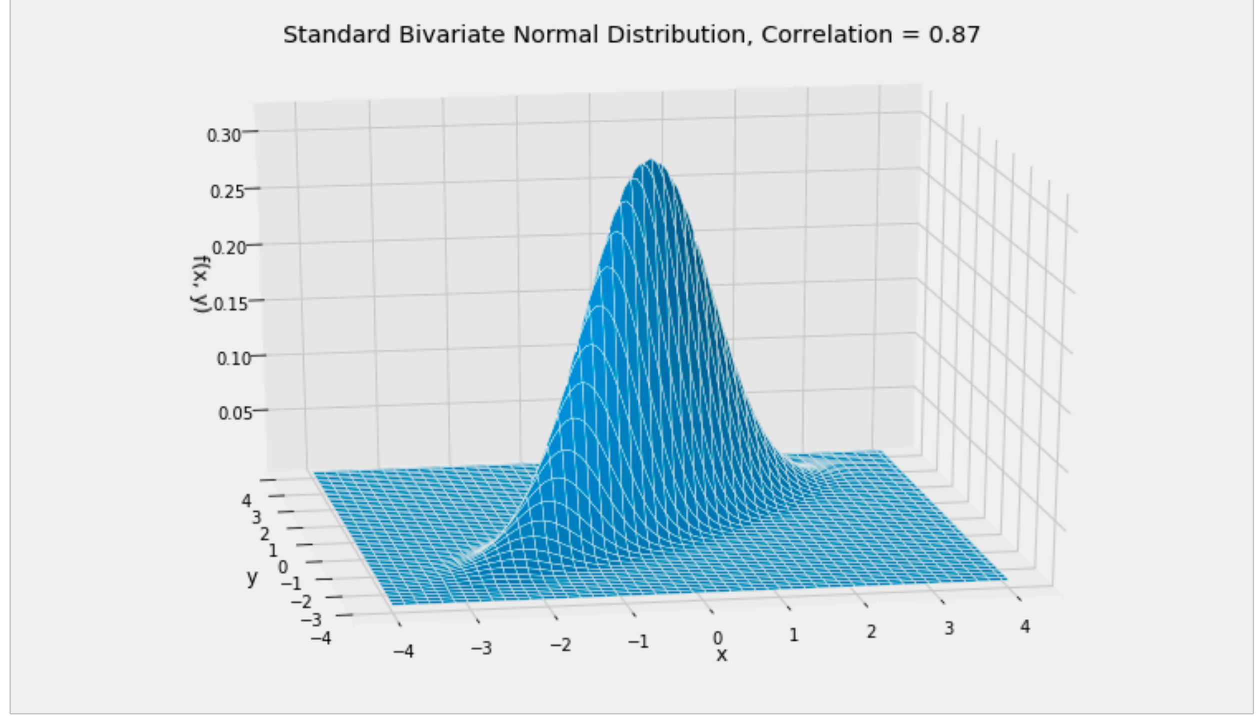
The upper tail: $\Pr(X - \mu \geq a\sigma) = \Pr(Z \geq a) = \Pr(e^{tZ} \geq e^{ta})$ (for any $t > 0$)

(Markov inequality) $\leq \mathbb{E}[e^{tZ}]/e^{ta} = e^{t^2/2 - ta}$

$\leq e^{-a^2/2}$ (choosing $t = a$ that minimizes $e^{t^2/2 - ta}$)

The lower tail: $\Pr(X - \mu \leq -a\sigma) = \Pr(Z \leq -a)$ is symmetric.

Bivariate Normal Distribution



- The joint density of <u>**standard bivariate normal**</u> random variables (X, Y) with parameter $-1 < \rho < 1$, is given by

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right)$$

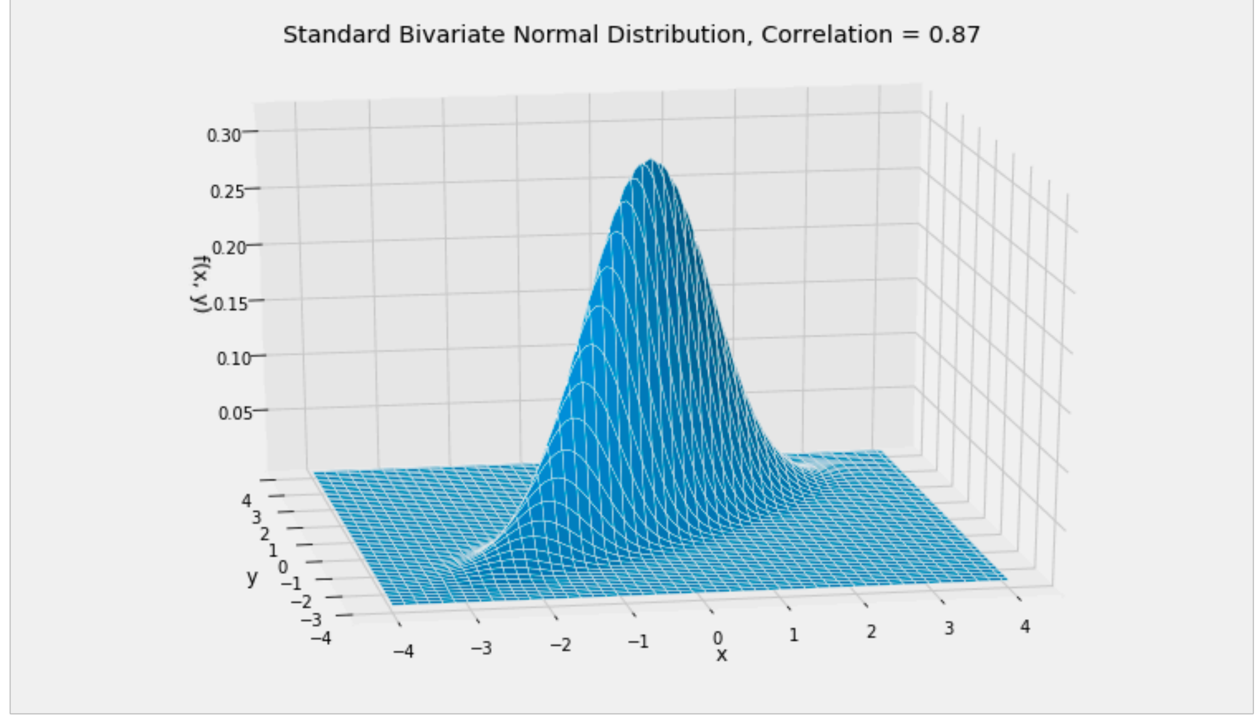
- The **marginal distributions** of X and Y are $N(0,1)$. And

$$\mathbf{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y = \rho$$

- $\rho = 0 \implies f_{X,Y}(x, y) = \phi(x)\phi(y)$

  standard bivariate normal variables are independent $\iff$ they are uncorrelated

Bivariate Normal Distribution



- The joint density of general <u>**bivariate normal**</u> random variables (X, Y) with means μ_1, μ_2, variances σ_1^2, σ_2^2, and correlation $-1 < \rho < 1$, is given by

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x,y)}$$

where $Q(x, y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$

$$= (x-\mu_1, y-\mu_2)\begin{bmatrix}\sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2\end{bmatrix}^{-1}(x-\mu_1, y-\mu_2)^T$$

- Marginally $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, and $\mathbf{Cov}(X, Y) = \sigma_1\sigma_2\rho$

Multivariate Normal Distribution*

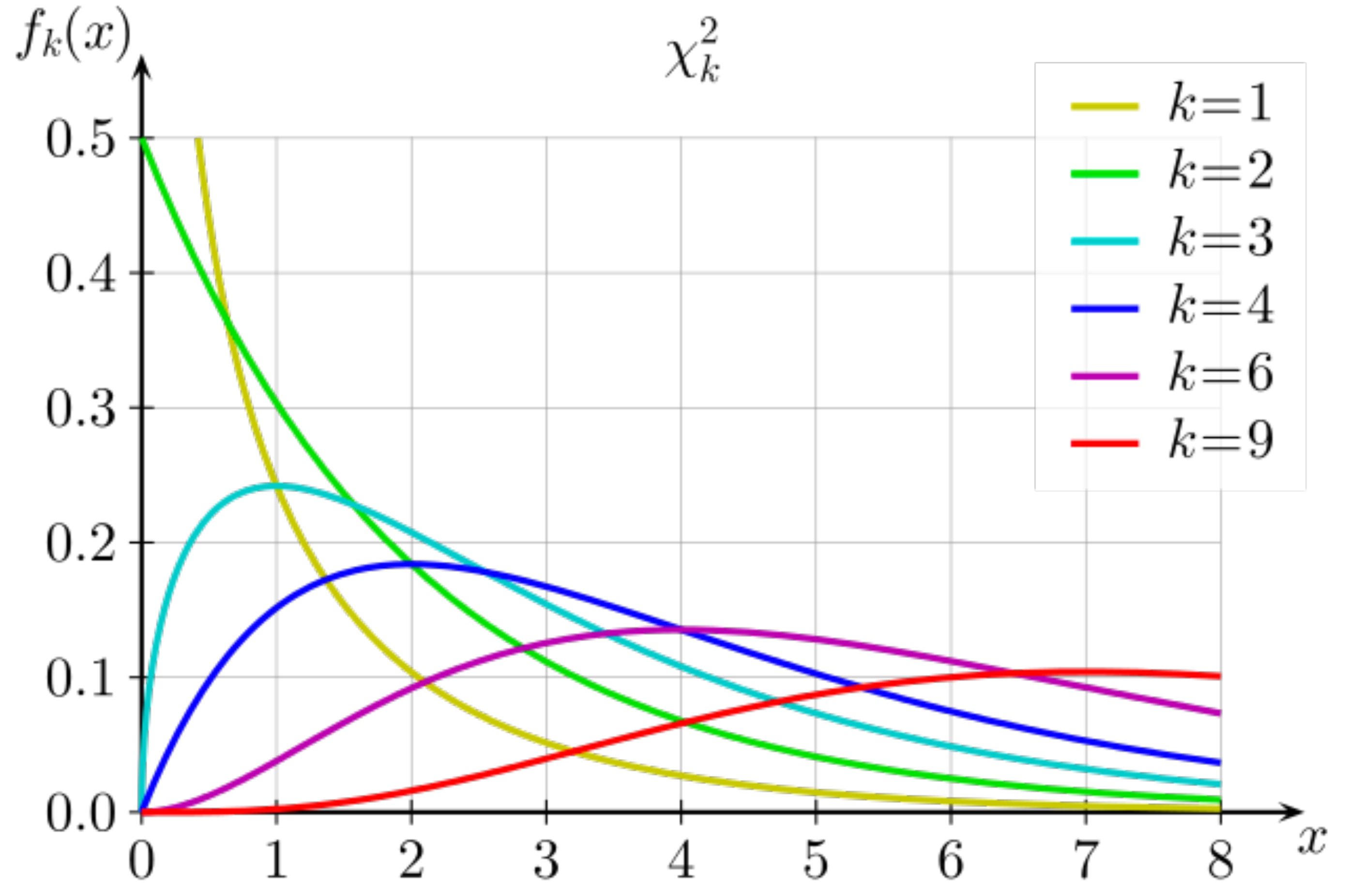
- A random vector $Y = (Y_1, \ldots, Y_n)$ has a multivariate normal distribution, iff there is an $\boldsymbol{A} \in \mathbb{R}^{n \times k}$, a vector $X = (X_1, \ldots, X_k)$ of k independent standard normal random variables, and a vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n$, such that
$$Y^T = AX^T + \boldsymbol{\mu}^T$$

- If further $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^T = \mathbb{E}[(Y - \boldsymbol{\mu})(Y - \boldsymbol{\mu})^T]$ has full rank, then the density of Y is
$$f(\boldsymbol{y}) = f(y_1, \ldots, y_n) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})^T}$$

- Denote $Y \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Marginally, each $Y_i \sim N(\mu_i, \Sigma_{ii})$, and $\mathbf{Cov}(Y_i, Y_j) = \Sigma_{ij}$

- **Equivalent characterization:** $\forall \boldsymbol{a} \in \mathbb{R}^n, \langle \boldsymbol{a}, Y \rangle = a_1 Y_1 + \cdots + a_n Y_n$ is normal

Chi-squared (χ^2) distribution



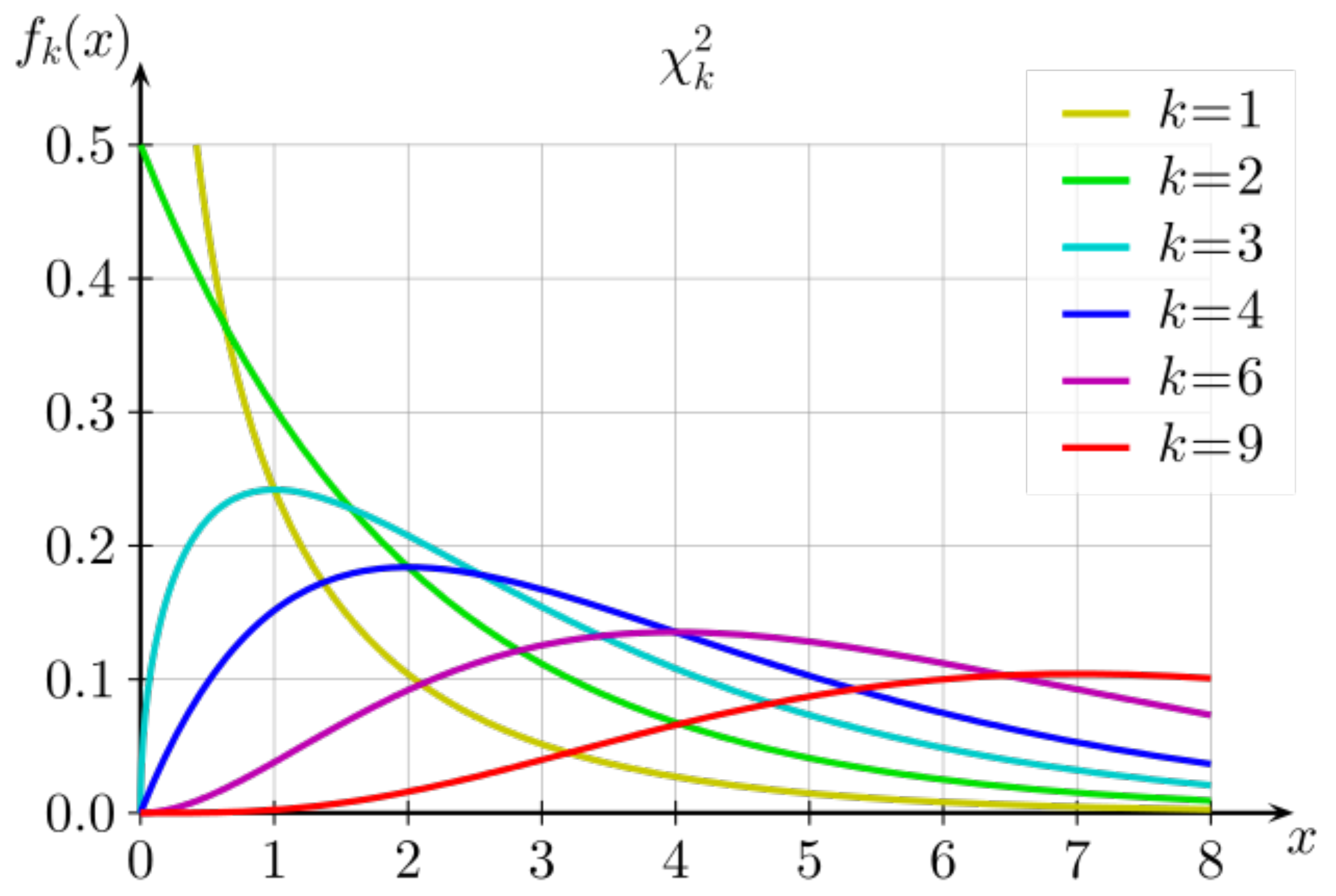
- If $Z_1, \ldots, Z_k$ are **independent standard normal** random variables, then

$$Q = \sum_{i=1}^{k} Z_i^2$$

follows the **chi-squared** (卡方) distribution with k **degrees of freedom**, denoted as $Q \sim \chi^2(k)$

- $\mathbb{E}[Z_i^2] = \mathbf{Var}[Z_i] = 1$ since $Z_i \sim N(0,1) \implies \mathbb{E}[Q] = k$
- sum of independent $\chi^2(k)$ and $\chi^2(l)$ random variables follows $\chi^2(k + l)$

Chi-squared (χ^2) distribution



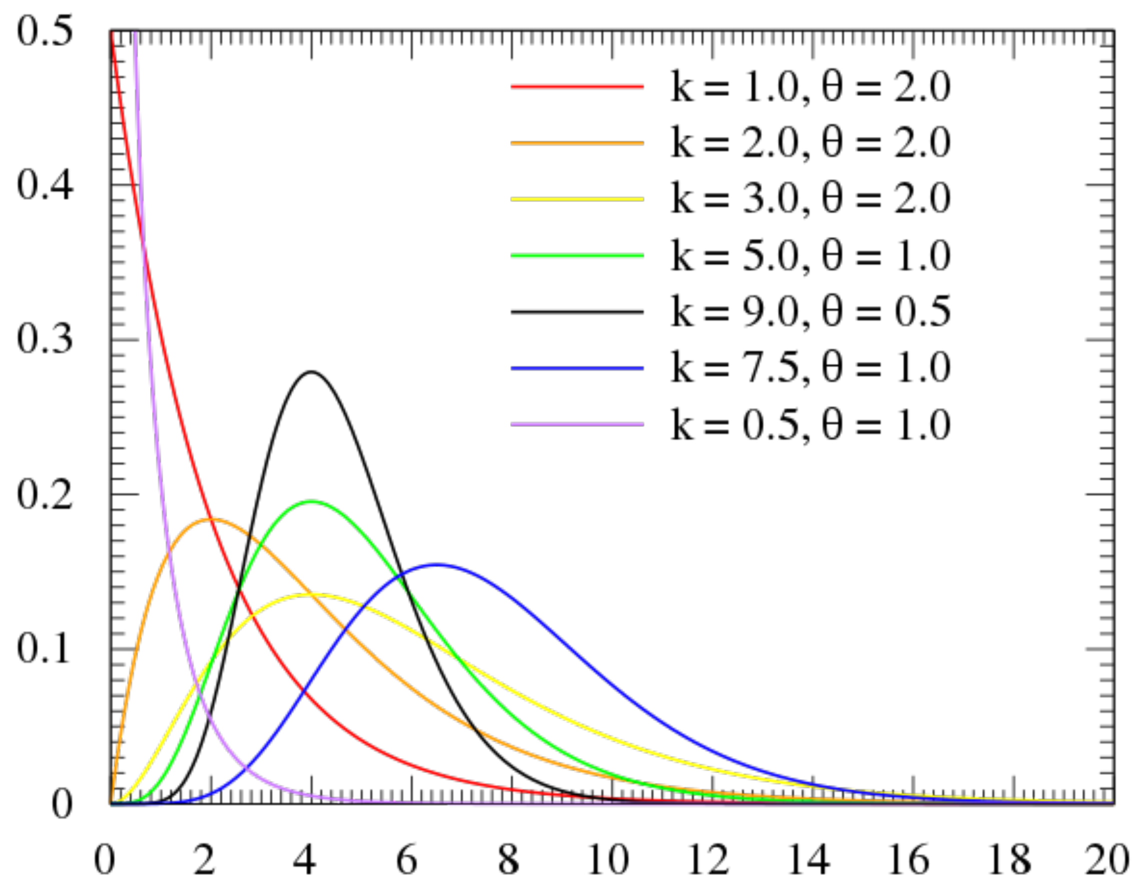
- Let $Z \sim N(0,1)$ and $Y = Z^2$. Then for any $y \geq 0$,

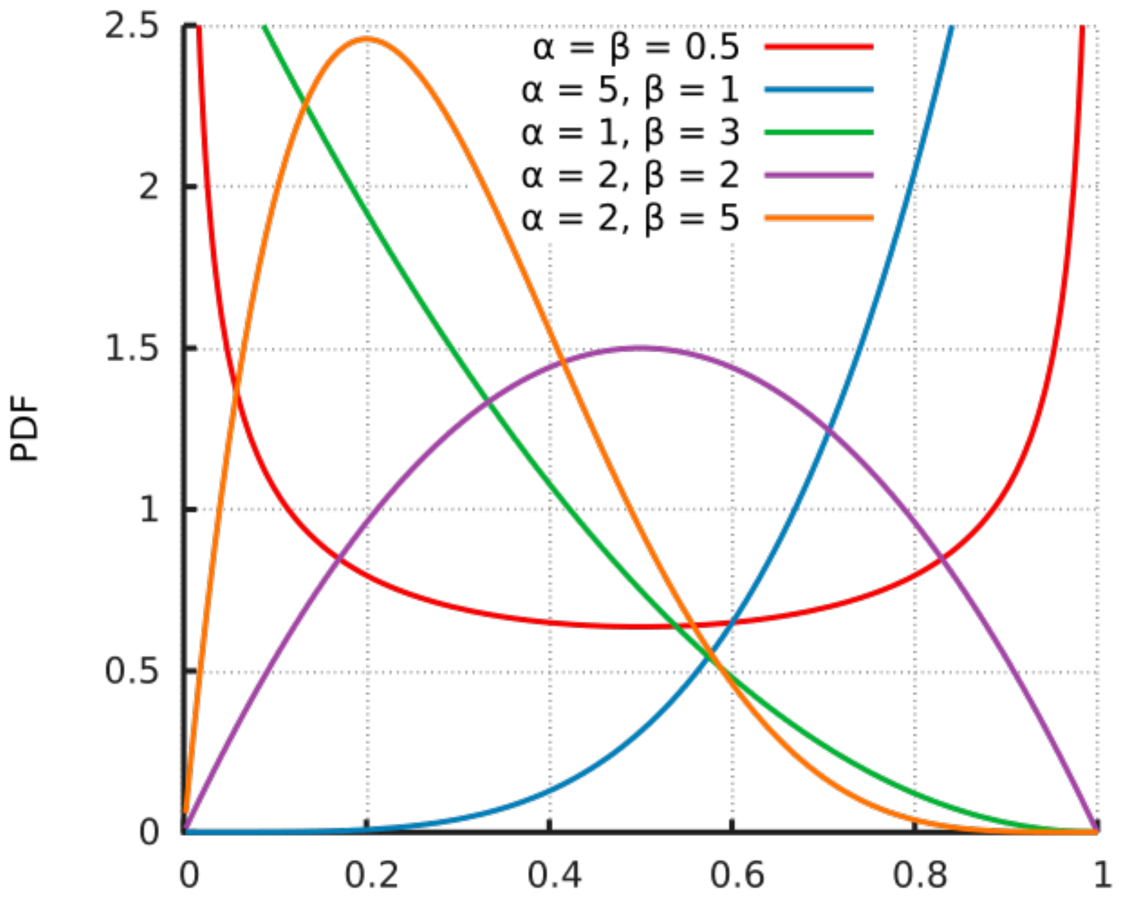
$$F_Y(y) = \Pr(Y \leq y) = \Pr(Z^2 \leq y) = \Pr(-\sqrt{y} \leq Z \leq \sqrt{y})$$

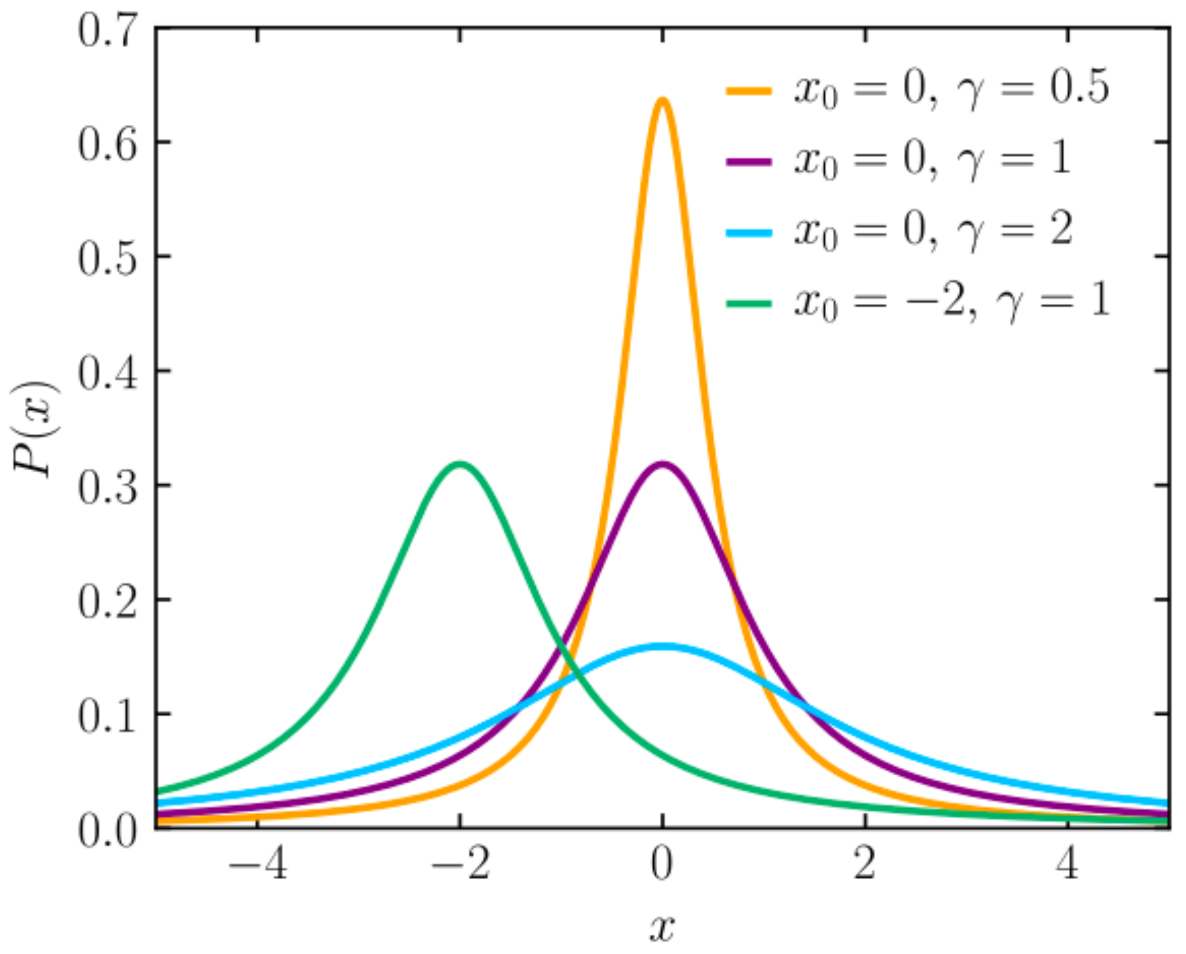
$$= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1$$

- By chain rule: $f_Y(y) = \dfrac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = 2\dfrac{\mathrm{d}\Phi(\sqrt{y})}{\mathrm{d}y} = \dfrac{1}{\sqrt{y}}\phi(\sqrt{y}) = \dfrac{1}{\sqrt{2\pi y}}\mathrm{e}^{-y/2}$
- $\chi^2(1)$ has pdf $f(x) = \dfrac{1}{\sqrt{2\pi x}}\mathrm{e}^{-x/2}$.
- For general integer $k \geq 1$: $\chi^2(k)$ has pdf $f(x) = \dfrac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}\mathrm{e}^{-x/2}$

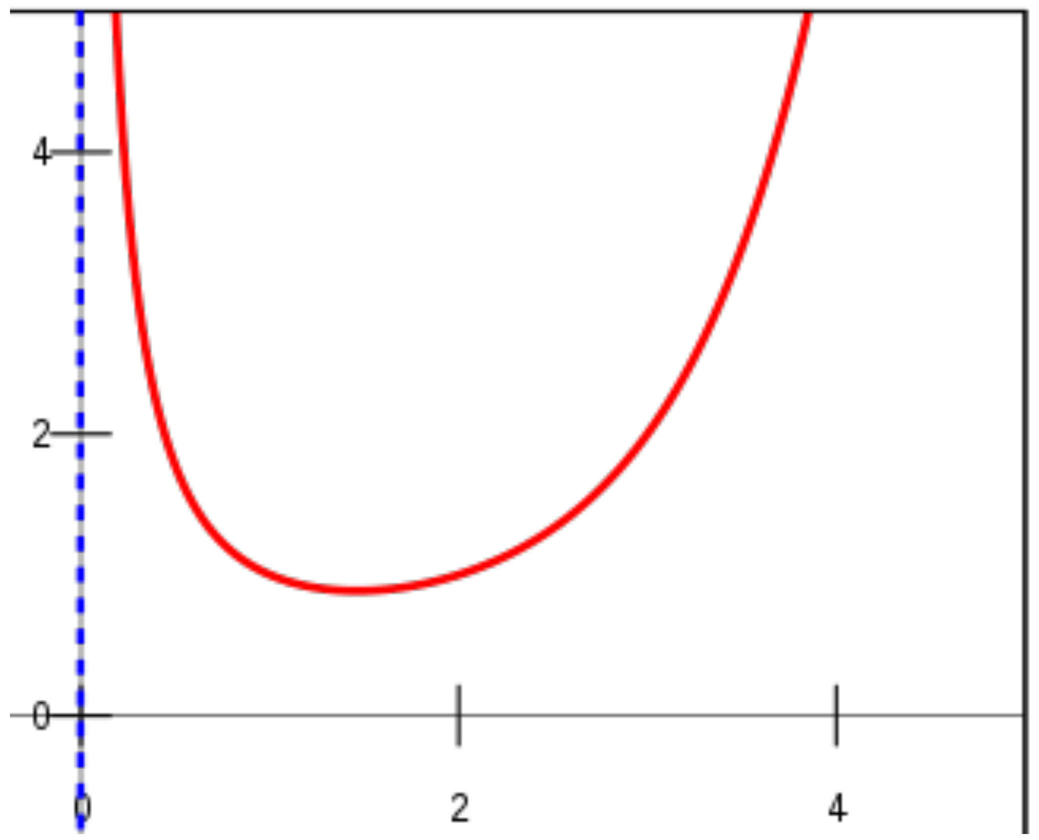
Other Continuous Probability Distributions







Gamma Function*

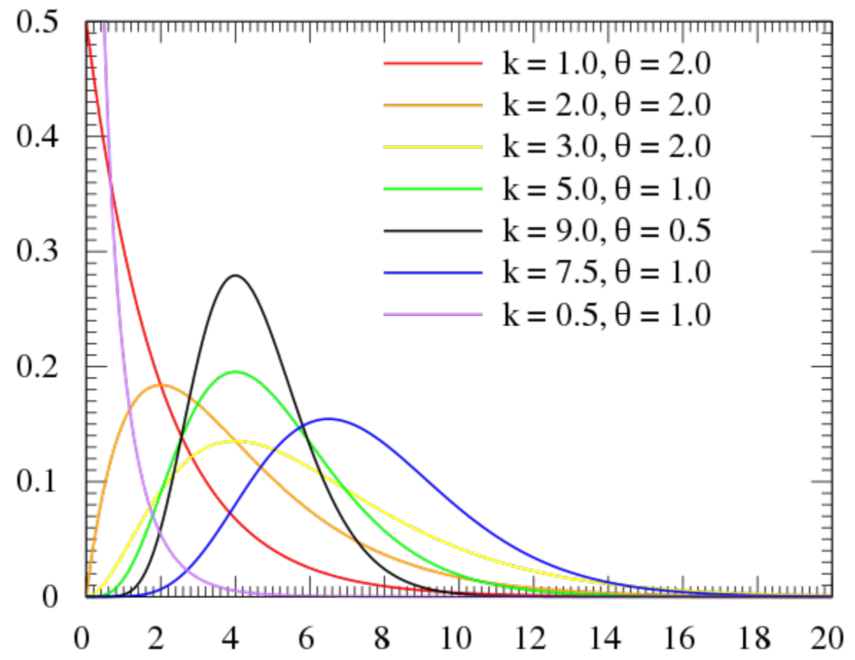


- Gamma function $\Gamma(x)$: analytic extension of factorial $\Gamma(n) = (n-1)!$

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \, dt, \quad \text{for } x > 0$$

- $\Gamma(1) = \int_0^\infty e^{-t} \, dt = 1$ gives exponential distribution with $\lambda = 1$
- $\Gamma(k) = \int_0^\infty (\lambda t)^{k-1} \lambda e^{-\lambda t} \, dt = \mathbb{E}[(\lambda X)^{k-1}]$ for exponential X with parameter 1

Gamma Distribution*



- The random variable X has the <u>**gamma distribution**</u> with parameters $k, \lambda > 0$, denoted $\Gamma(k, \lambda)$ or $\text{Gamma}(k, \lambda)$, if it has the density

$$f_X(x) = \frac{1}{\Gamma(k)}(\lambda x)^{k-1} \cdot \lambda e^{-\lambda x}, \quad \text{for } x \geq 0, \quad \text{where } \Gamma(k) = \int_0^\infty t^{k-1} e^{-t} \, dt$$

- $\Gamma(1,\lambda)$ is the exponential distribution with parameter λ
- If $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$ are independent, then $X + Y \sim \Gamma(\alpha + \beta, \lambda)$
- $\sum_{i=1}^{k} X_i \sim \Gamma(k, \lambda)$ if $X_1, \ldots, X_k$ are i.i.d. exponential random variables with parameter λ
- $\Gamma(k/2, 1/2)$, for integer $k \geq 1$, is the $\chi^2(k)$ distribution

Gamma Distribution*

- The **moment generating function** of **gamma** random variable $X \sim \Gamma(k, \lambda)$ is

$$M_X(t) = \mathbb{E}[\mathrm{e}^{tX}] = \left(1 - \frac{t}{\lambda}\right)^{-k} \quad \text{for } t < \lambda$$

Proof: $M_X(t) = \mathbb{E}[\mathrm{e}^{tX}] = \int_0^\infty \mathrm{e}^{tx} \cdot f_X(x)\,\mathrm{d}x = \frac{\lambda^k}{\Gamma(k)} \int_0^\infty x^{k-1}\mathrm{e}^{-(\lambda-t)x}\,\mathrm{d}x$

(substituting $(\lambda - t)x = u$) $= \frac{\lambda^k}{\Gamma(k)(\lambda-t)^k} \int_0^\infty u^{k-1}\mathrm{e}^{-u}\,\mathrm{d}u = \frac{\lambda^k \Gamma(k)}{\Gamma(k)(\lambda-t)^k} = \left(1 - \frac{t}{\lambda}\right)^{-k}$

Gamma Distribution*

- The moment generating function of gamma random variable $X \sim \Gamma(k, \lambda)$ is

$$M_X(t) = \mathbb{E}[\mathrm{e}^{tX}] = \left(1 - \frac{t}{\lambda}\right)^{-k} \quad \text{for } t < \lambda$$

- If $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$ are independent, then for $t < \lambda$

$$M_{X+Y}(t) = \mathbb{E}[\mathrm{e}^{t(X+Y)}] = \mathbb{E}[\mathrm{e}^{tX} \cdot \mathrm{e}^{tY}] = \mathbb{E}[\mathrm{e}^{tX}] \cdot \mathbb{E}[\mathrm{e}^{tY}] = \left(1 - \frac{t}{\lambda}\right)^{-(\alpha+\beta)}$$

$\implies X + Y \sim \Gamma(\alpha + \beta, \lambda)$ (since MGF uniquely identifies distribution)

Poisson Point Process*

- Let $\{N(t) \mid t \geq 0\}$ be a **Poisson process** with rate $\lambda > 0$. For any $t, s \geq 0$ and any integer $n \geq 0$,

$$\Pr(N(t+s) - N(s) = n) = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

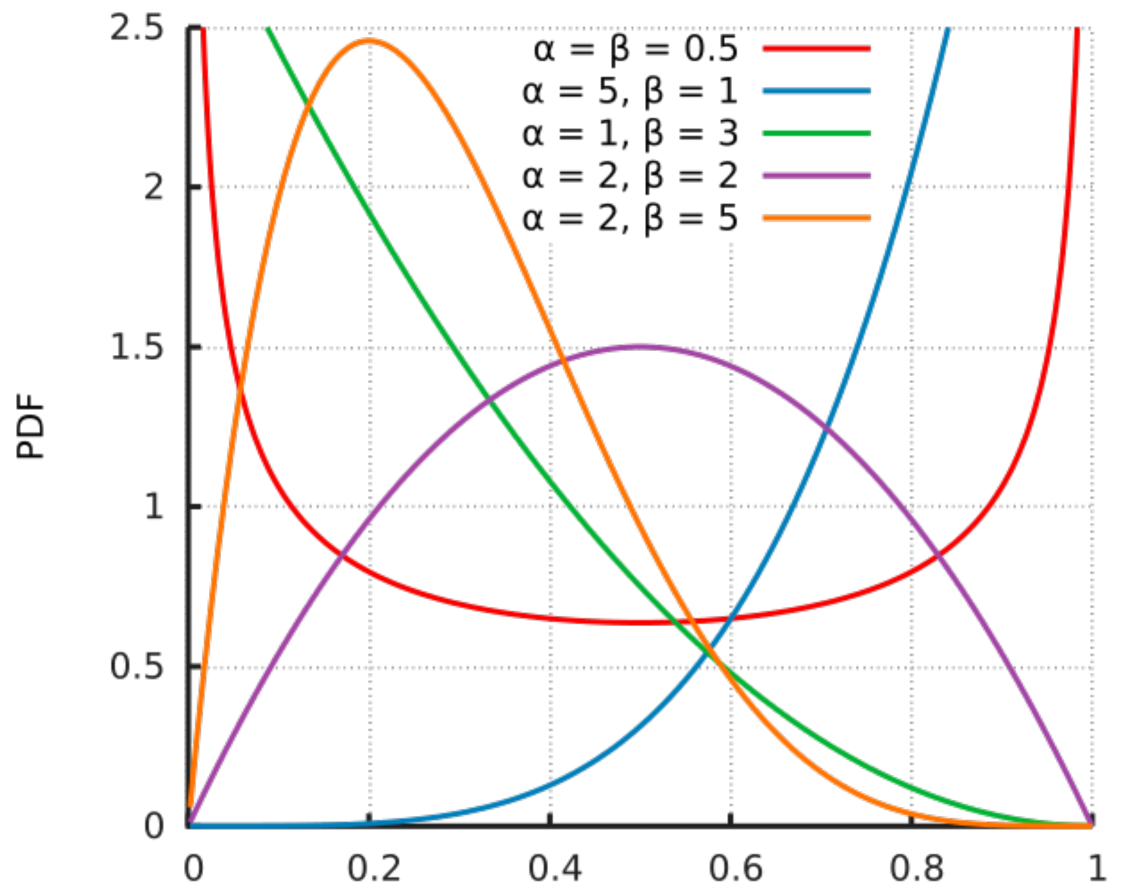
Proof: By memoryless, it suffices to prove $\Pr(N(t) = n) = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}$

For i.i.d. exponential random variables $X_1, X_2, \ldots$ with parameter λ

$$\Pr(N(t) = n) = \Pr\left(\sum_{i=1}^{n} X_i \leq t \cap \sum_{i=1}^{n+1} X_i > t\right) = \int_0^t f_{\sum_1^n X_i}(x) \cdot \Pr\left(X_{n+1} > t - x\right)\, \mathrm{d}x$$

$$= \int_0^t \frac{(\lambda x)^{n-1}\lambda \mathrm{e}^{-\lambda x}}{\Gamma(n)} \mathrm{e}^{-\lambda(t-x)}\, \mathrm{d}x = \frac{\lambda^n \mathrm{e}^{-\lambda t}}{\Gamma(n)}\int_0^t x^{n-1}\, \mathrm{d}x = \frac{\lambda^n \mathrm{e}^{-\lambda t} t^n}{n\Gamma(n)} = \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

Beta Distribution*



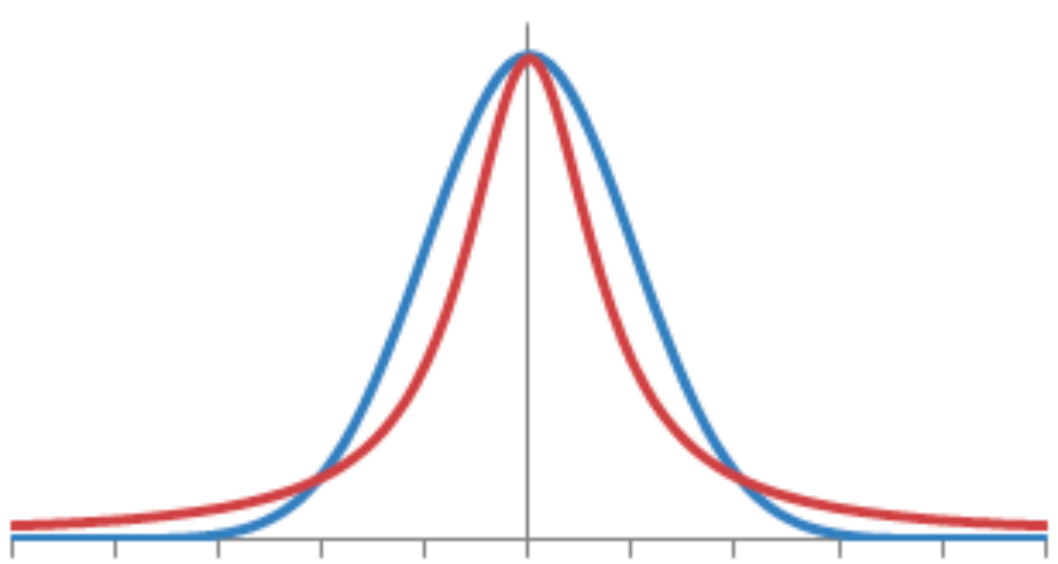
- The random variable X is **beta** with parameters $a, b > 0$, if it has the density

$$f_X(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \qquad \text{for } 0 \leq x \leq 1,$$

where $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x$ is the **beta function**

- $\text{Beta}(1,1) \sim$ uniform on $[0,1]$, and $\text{Beta}(1,n) \sim \min_{1\leq i\leq n} X_i$ for i.i.d. X_i uniform on $[0,1]$
- If $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$ are independent, then $\frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$

Cauchy Distribution*



- The random variable X has the **Cauchy distribution**, if it has the density

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty$$

- The Cauchy random variable X ***does not have any moment***:

$$\mathbb{E}[X^k] = \infty \text{ for all } k \geq 1$$

- The moment generating function $M_X(t) = \mathbb{E}[e^{tX}]$ does not exist