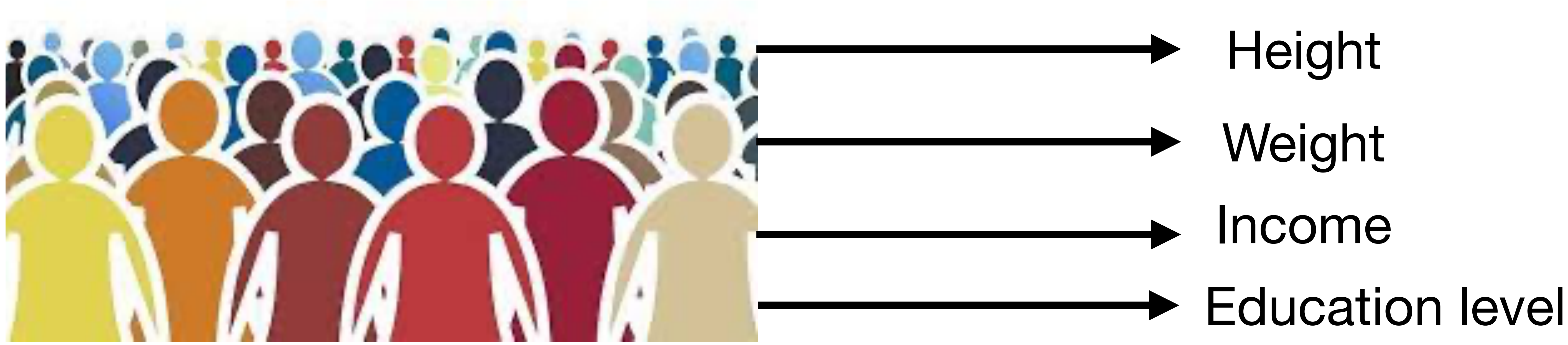
Probability and Statistics

Random Variable

谢润烁 Nanjing University, 2023 Fall

Why Random Variable?

- The Ω might be too concrete
- We care about mathematical properties



$$\Omega \xrightarrow{X/Y/Z/\ldots} \mathbb{R}/\mathbb{N}$$

What about Random Variable?

- How to describe a random variable?
  - $\Pr(X = k) = f(k)$?
- How to compute/approximate the probability of a certain event?
  - $\{X = 1\} \cap \{X = 3\}$
  - $\{a \le X \le b\}$
- How to reason about the overall behavior of a random variable?
  - Moment and Deviation $\rightarrow$ Concentration of Measure

Basics on Random Variable

Random Variable

- $X : \Omega \to \mathbb{R}$
  - Is every function $f : \Omega \to \mathbb{R}$ a legal random variable?
- We write $\{a \le X \le b\}$
  - instead of $\{\omega : a \le X(\omega) \le b\}$
  - We don't care too much about the concrete Ω when we're studying X

Function of Random Variable

- We can operate on numbers $\mathbb{N}$
  - 1 + 2
- We can also operate on function
  - $f \circ g$
- Random variable as well
  - $Z = f(X, Y)$

How to describe a random variable?

- In discrete case, it is okay to use $\Pr(X = k)$ to describe a random variable
- When things become continuous:
  - Uniform distribution on $[0,1]$
  - Uniform distribution on $[0,1] \times [0,1]$

Discrete and Continuous Variable

- Discrete: Probability Mass Function (pmf)

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

$$\Pr(X \leq x) = \sum_{k=-\infty}^{x} f(k), x \in \mathbb{N}$$

- Continuous: Probability Density Function (pdf)

$$\Pr(X \leq x) = \int_{-\infty}^{x} f(u)du, x \in \mathbb{R}$$

Distribution Function

Unify the discrete and continuous cases

- Distribution Function
  - a.k.a Cumulative Distribution Function (CDF)
- $F(x) = \Pr(X \leq x)$, for $\forall x \in \mathbb{R}$
- Answer: Is every function $f : \Omega \rightarrow \mathbb{R}$ a legal random variable?
  - Recall that $\Pr(A)$ is well-defined if and only if $A \in \Sigma$
  - X is a random variable if and only if $\forall x \in \mathbb{R} . \{X \leq x\} \in \Sigma$

Distribution Function

Unify the discrete and continuous cases

- aka Cumulative Distribution Function (CDF)
- $F(x) = \Pr(X \leq x)$, for $\forall x \in \mathbb{R}$
- Property
  - $F(-\infty) = 0$
  - $F(+\infty) = 1$
  - $F(x)$ is right-continuous

Expectation

Aggregate the data

- Discrete case

$$\mathbb{E}(X) = \sum_{x} x \cdot \Pr(X = x)$$

- Continuous case

$$\mathbb{E}(X) = \int_{x} xf(x)dx$$

Expectation

Linearity of Expectations

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
- No need for independence
- Generally,

$$\mathbb{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbb{E}(X_i)$$

$$\mathbb{E}(cX) = c\mathbb{E}(X)$$

Expectation

Averaging principle

- It is trivial that:
  - $\exists x \,.\, x \geq \mathbb{E}(X)$ and $\exists x \,.\, x \leq \mathbb{E}(X)$
  - Consider a class with average height (expectation) 175cm
    - There must be people with height ≤175cm
    - There must be people with height ≥175cm
- In probabilistic method
  - we use $\mathbb{E}(X)$ to prove some upper bounds and lower bounds of X
- Conditional Expectation can even be used to design algorithms

Variance

How data deviates from the expectation?

- $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}(X^2) - \mathbb{E}X^2$
- Standard Deviation: $\sqrt{\mathrm{Var}(X)}$
- If X and Y are independent

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

- They will be useful later, but not now

Elementary Random Variables

Road Map

Let’s explore their distribution, expectation and variance

Discrete

- Bernoulli
- Binomial
- Geometric
- Poisson

Continuous

- Uniform Distribution
- Exponential
- Normal

There will be tons of maths here

Road Map

Let’s explore their distribution, expectation and variance

Discrete

- **Bernoulli**
- Binomial
- Geometric
- Poisson

Continuous

- Uniform Distribution
- Exponential
- Normal

Bernoulli Distribution

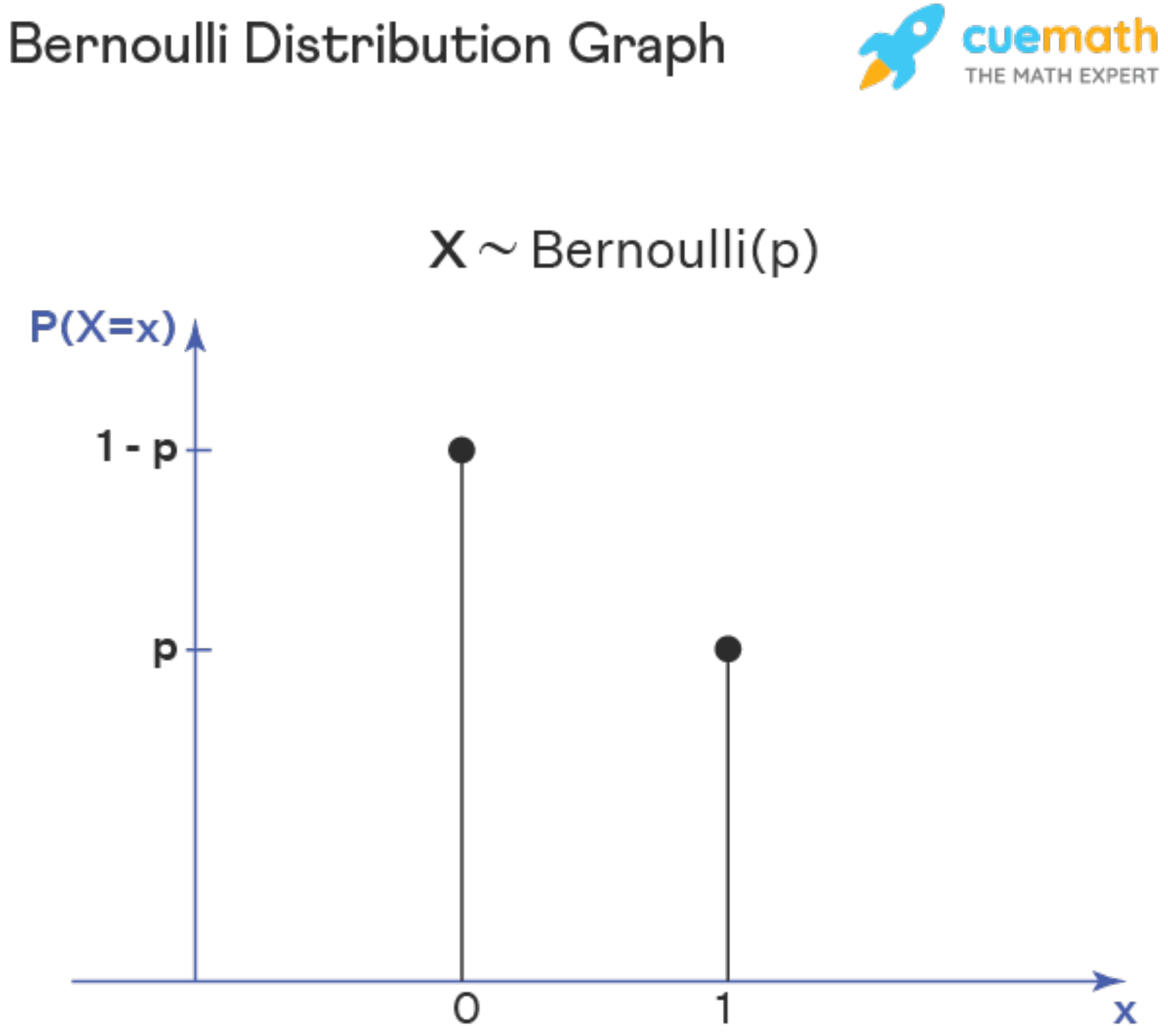
a.k.a Bernoulli Trial

- $X = \begin{cases} 1 & \text{trial succeeds} \\ 0 & \text{otherwise} \end{cases}$
- $\Omega = \{H, T\}$

Bernoulli Distribution

a.k.a Bernoulli Trial

- $X = \begin{cases} 1 & \text{trial succeeds} \\ 0 & \text{otherwise} \end{cases}$
    - $\Pr(X = 1) = p$
    - $\Pr(X = 0) = 1 - p$
- $\mathbb{E}(X) = p$
- $\mathrm{Var}(X) = p(1 - p)$



Road Map

Let’s explore their distribution, expectation and variance

Discrete

- Bernoulli
- **Binomial**
- Geometric
- Poisson

Continuous

- Uniform Distribution
- Exponential
- Normal

Binomial Distribution

Sum of Independent Bernoulli Distribution

- We don't throw one time; we throw n times
- For example, when $n = 3$

$$\Omega = \{HHH, HHT, HTH, HTT, TTT, TTH, THT, TTH\}$$

- How many heads can we get?

Binomial Distribution

Sum of Independent Bernoulli Distribution

- $X \sim B(n, p)$
  - $X = X_1 + X_2 + \ldots + X_n$
  - X_i: Bernoulli Distribution
  - $X_1, X_2, \ldots, X_n$ are i.i.d (independent identical distribution)

Binomial Distribution

Sum of Independent Bernoulli Distribution

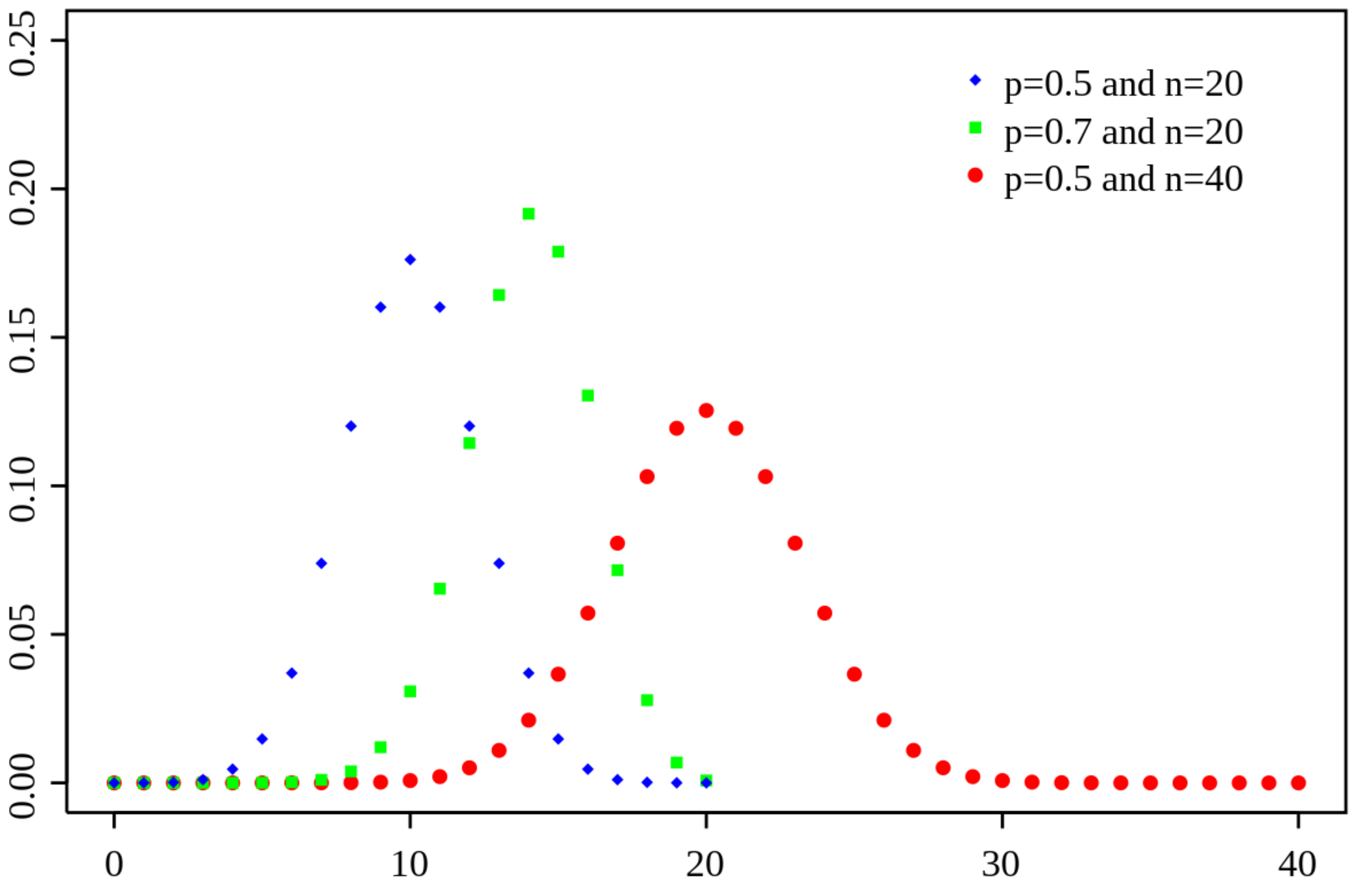
- $X \sim B(n, p)$

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

- $\mathbb{E}(X) = np$
  - Proved by definition
  - Or by linearity of Expectation (Much simpler)
- $\mathrm{Var}(X) = np(1-p)$

Binomial Distribution

Sum of Independent Bernoulli Distribution



Road Map

Let’s explore their distribution, expectation and variance

Discrete

- Bernoulli
- Binomial
- **Geometric**
- Poisson

Continuous

- Uniform Distribution
- Exponential
- Normal

Geometric Distribution

Do not stop until I hit it

- You can throw as many times as possible
- But if you get a head, you must stop
- $\Omega = \{H, TH, TTH, TTTH, \dots\}$

Geometric Distribution

Do not stop until I hit it

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

- $\mathbb{E}(X) = \dfrac{1}{p}$
- $\mathrm{Var}(X) = \dfrac{(1 - p)}{p^2}$

Road Map

Let’s explore their distribution, expectation and variance

Discrete

- Bernoulli
- Binomial
- Geometric
- **Poisson**

Continuous

- Uniform Distribution
- Exponential
- Normal

Poisson Distribution

What the hell is this?

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

- Consider a Binomial Distribution with $n = 37, p = \frac{1}{37}$
  - First 3 terms
  - $(1 - \frac{1}{37})^{37}, \binom{37}{1}(1 - \frac{1}{37})^{36}\frac{1}{37}, \binom{37}{2}(1 - \frac{1}{37})^{35}(\frac{1}{37})^2,$

Poisson Distribution

What the hell is this?

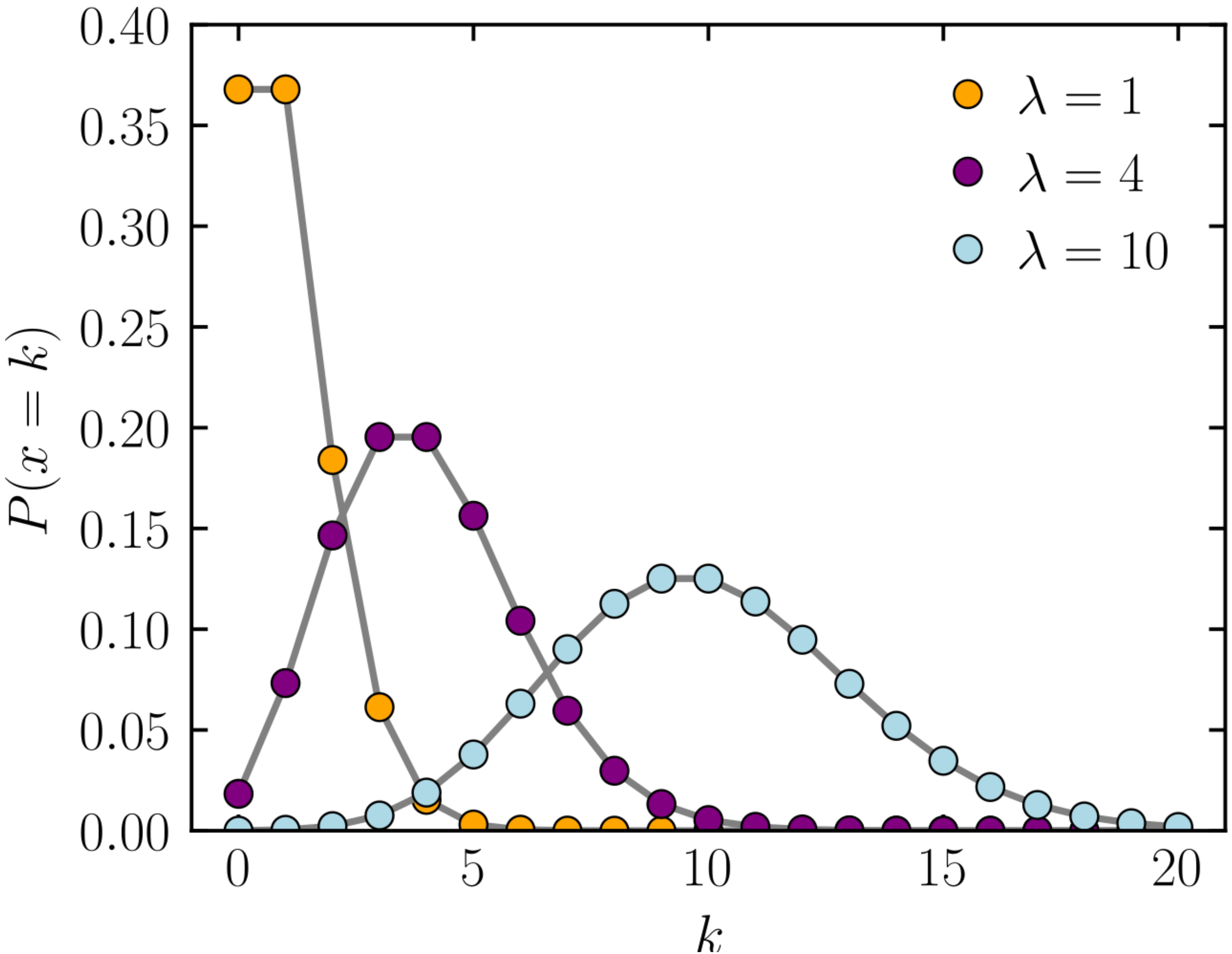
- $(1-\frac{1}{37})^{37}, \binom{37}{1}(1-\frac{1}{37})^{36}\frac{1}{37}, \binom{37}{2}(1-\frac{1}{37})^{35}(\frac{1}{37})^{2},$
- $c=(1-\frac{1}{37})^{37}\approx 0.363$
- $c, \frac{36}{37}c, \frac{36}{37}\times\frac{1}{2}c$

Poisson Distribution

What the hell is this?

- $c = (1 - \frac{1}{37})^{37} \approx 0.363$
- $c, \frac{36}{37}c, \frac{36}{37} \times \frac{1}{2}c$
- Approximation: $[e^{-1} \approx 0.368 \ / \ c]$, $[1 \ / \ \frac{36}{37}]$
- $e^{-1}, e^{-1}, \frac{1}{2}e^{-1}$

Poisson Distribution



Poisson Distribution

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

- $\mathbb{E}(X) = \lambda$
- $\text{Var}(X) = \lambda$
- "the Poisson law was at one time under the name of '*the law of small numbers*'."
- "Well-kept statistical data such as the number of Prussian cavalry men killed each year by a kick from a horse, or the number of child suicides in Prussia, were cited as typical examples of this remarkable distribution (see [Keynes])"

Road Map

Let's explore their distribution, expectation and variance

Discrete

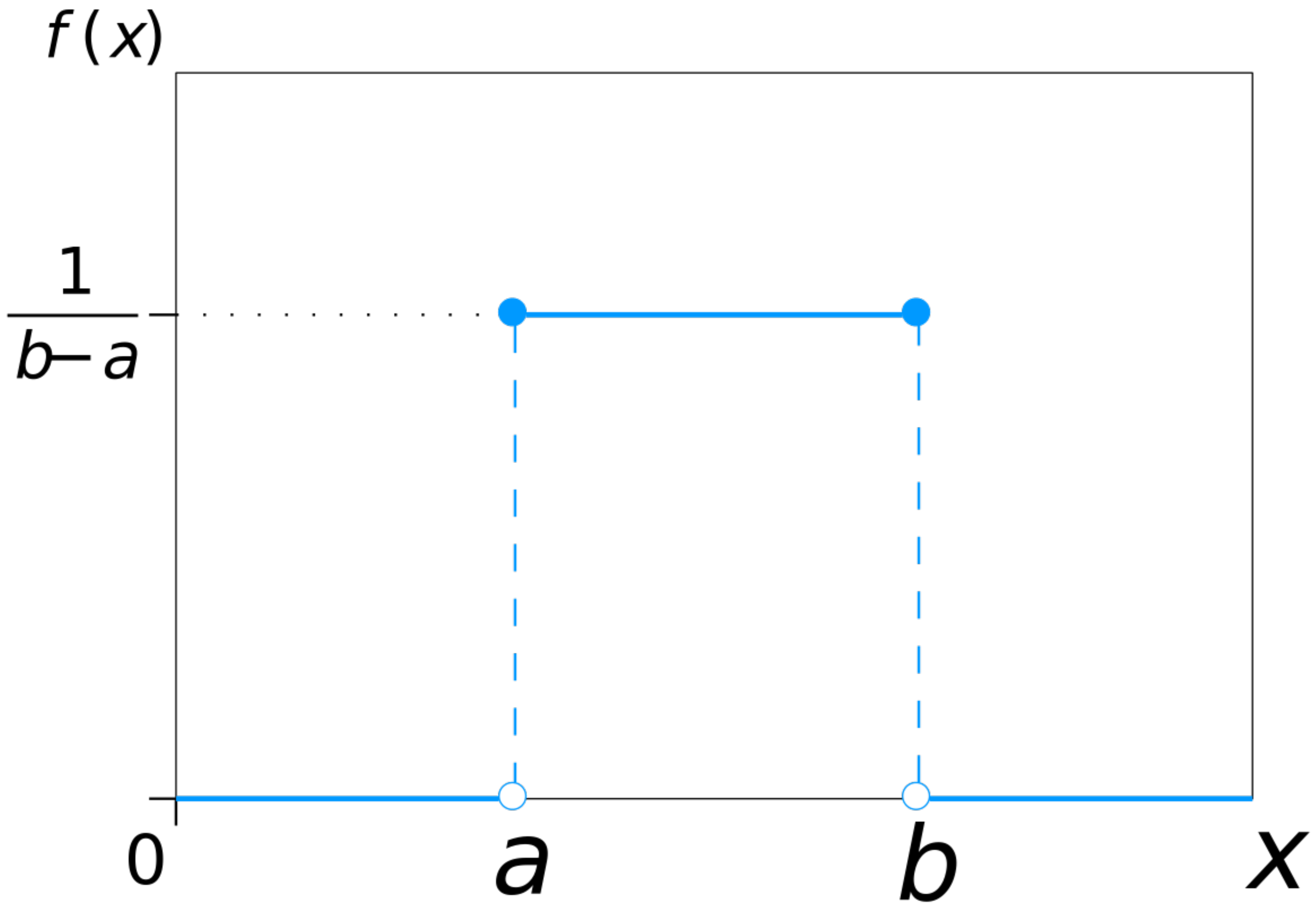
- Bernoulli
- Binomial
- Geometric
- Poisson

Continuous

- **Uniform Distribution**
- Exponential
- Normal

Uniform Distribution

- $f(x) = \begin{cases} \dfrac{1}{b-a} & , a \leq x \leq b \\ 0 & \text{,otherwise} \end{cases}$
- $\mathbb{E}(X) = (a+b)/2$
- $\text{Var}(X) = \dfrac{(b-a)^2}{12}$



Road Map

Let's explore their distribution, expectation and variance

Discrete

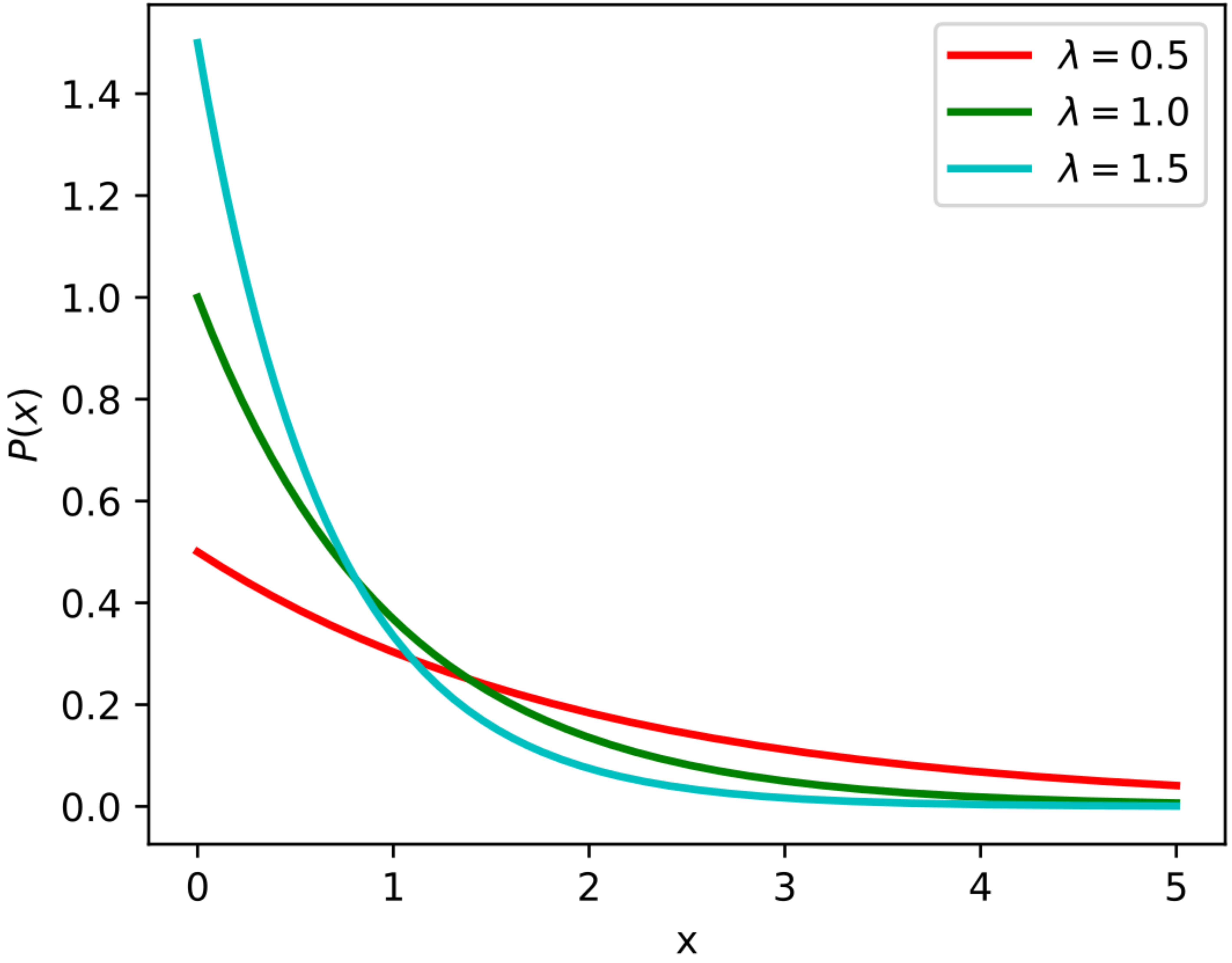
- Bernoulli
- Binomial
- Geometric
- Poisson

Continuous

- Uniform Distribution
- **Exponential**
- Normal

Exponential Distribution

- $f(x) = \lambda e^{-\lambda x}$, for $x \geq 0$
- $\mathbb{E}(X) = \dfrac{1}{\lambda}$
- $\mathrm{Var}(X) = \dfrac{1}{\lambda^2}$



Exponential Distribution

- Memoryless

$$\Pr(X > s + t \,|\, X > s) = \Pr(X > t)$$

- Useful model for various types of waiting time problems
  - telephone calls
  - service times
  - splitting of radioactive particles
  - ……

Road Map

Let’s explore their distribution, expectation and variance

Discrete

- Bernoulli
- Binomial
- Geometric
- Poisson

Continuous

- Uniform Distribution
- Exponential
- **Normal**

Normal Distribution

a.k.a Gaussian Distribution

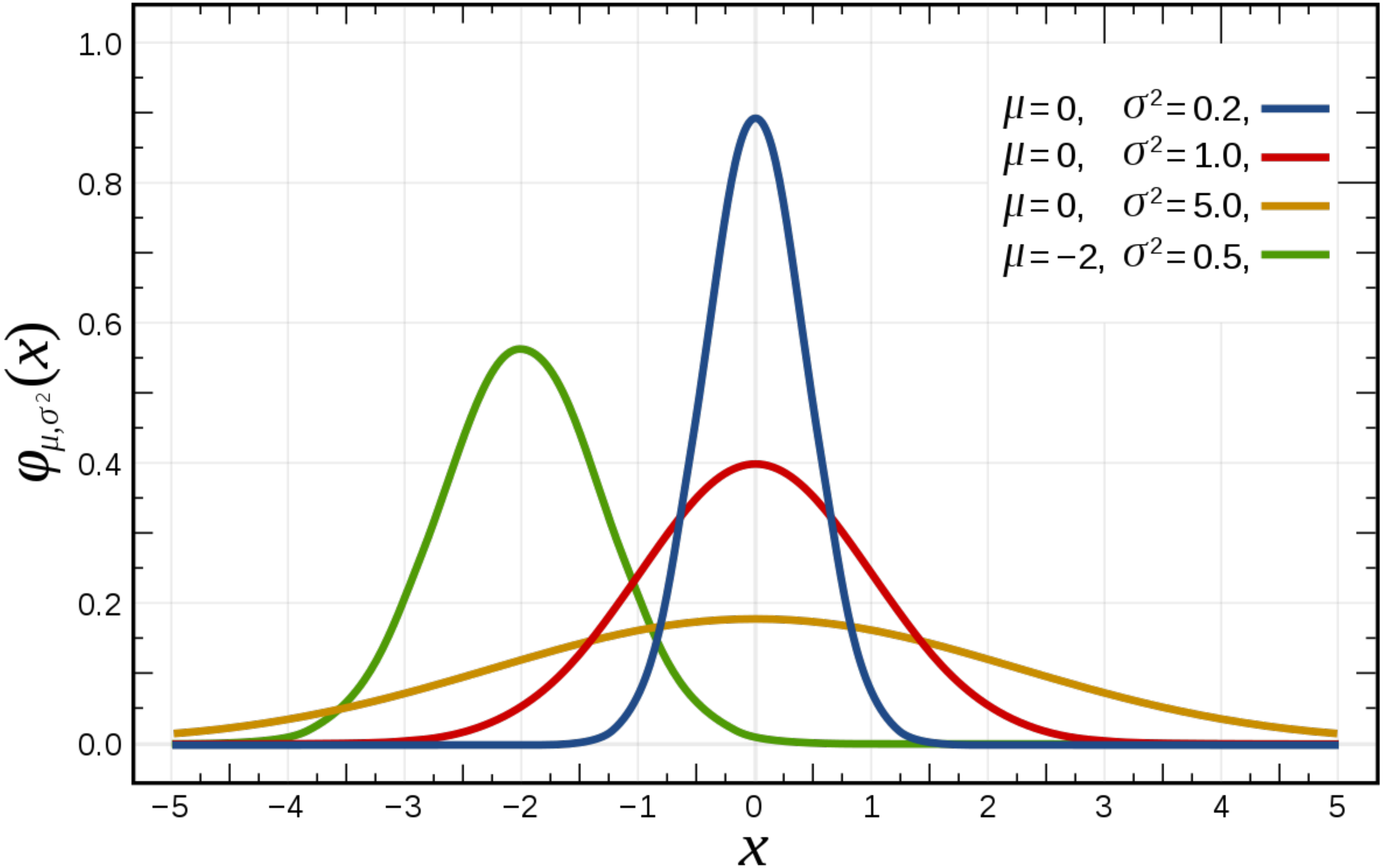
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, -\infty < x < \infty$$

- $\mathbb{E}(X) = \mu$
- $\text{Var}(X) = \sigma^2$
- standard/unit Normal Distribution: $\mu = 0, \sigma^2 = 1$

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

Normal Distribution

a.k.a Gaussian Distribution



Normal Distribution

a.k.a Gaussian Distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2}, -\infty < x < \infty$$

- $\mathbb{E}(X) = \mu$
- $\text{Var}(X) = \sigma^2$
- standard/unit Normal Distribution: $\mu = 0, \sigma^2 = 1$

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