Lecture 12 Continuous Probability Distribution

BIO210 Biostatistics

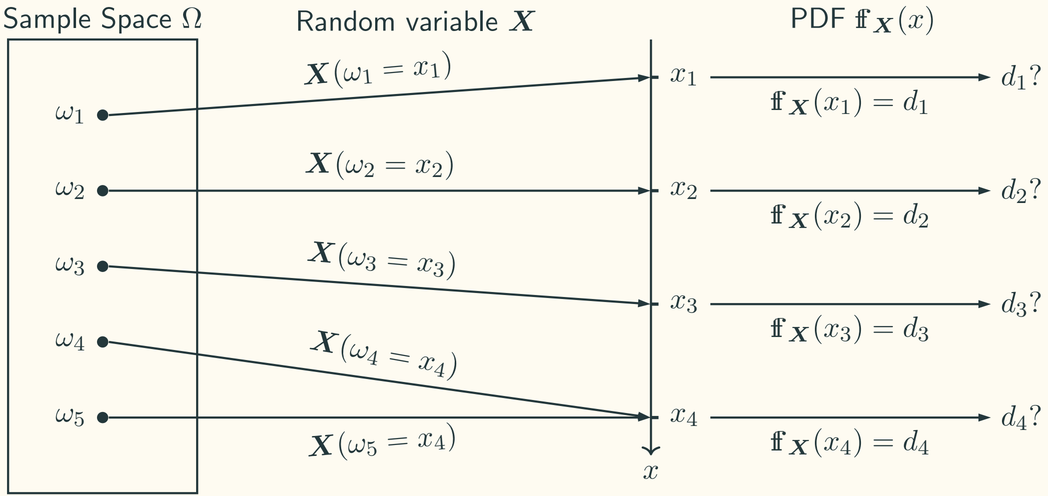
Xi Chen

Fall 2024

School of Life Sciences
Southern University of Science and Technology

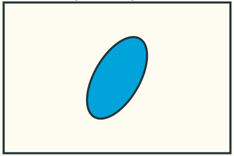
南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

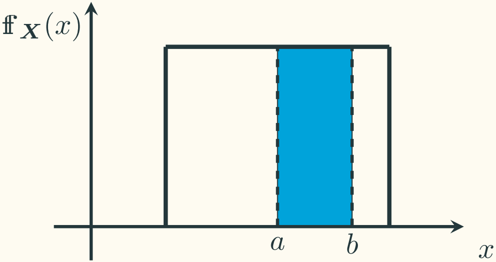
Probability Density Function (PDF)



Probability Density Function (PDF)

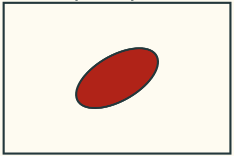
Sample space Ω

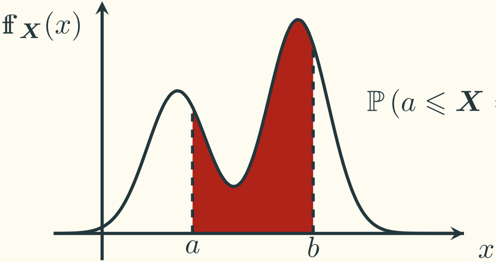




$$\mathbb{P}\,(a \leqslant \boldsymbol{X} \leqslant B)$$
$$= \mathbb{f}_{\boldsymbol{X}}(x) \cdot (b - a)$$

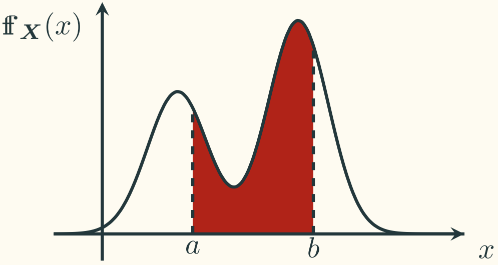
Sample space Ω





$$\mathbb{P}\,(a \leqslant \boldsymbol{X} \leqslant b) = \int_a^b \mathbb{f}_{\boldsymbol{X}}(x)\mathrm{d}x$$

Probability Density Function (PDF)



$$\mathbb{f}_{\boldsymbol{X}}(x) \geqslant 0, \int_{-\infty}^{+\infty} \mathbb{f}_{\boldsymbol{X}}(x)\,\mathrm{d}x = 1$$

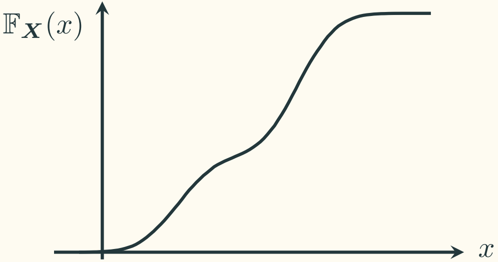
$$\mathbb{P}\left(\boldsymbol{X} = a\right) = ?$$

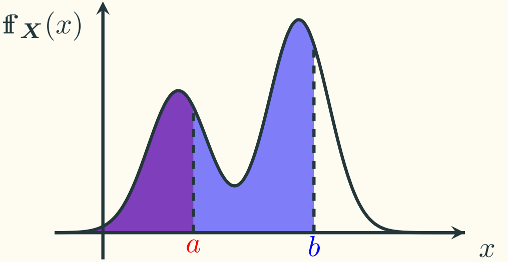
$$\mathbb{P}\left(x \leqslant \boldsymbol{X} \leqslant x + \delta\right)$$

$$= \int_{x}^{x+\delta} \mathbb{f}_{\boldsymbol{X}}(x)\,\mathrm{d}x = \mathbb{f}_{\boldsymbol{X}}(x) \cdot \delta$$

$$\mathbb{f}_{\boldsymbol{X}}(x) = \frac{\mathbb{P}\left(x \leqslant \boldsymbol{X} \leqslant X + \delta\right)}{\delta}$$

Cumulative Distribution Function



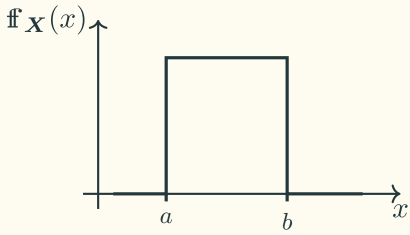


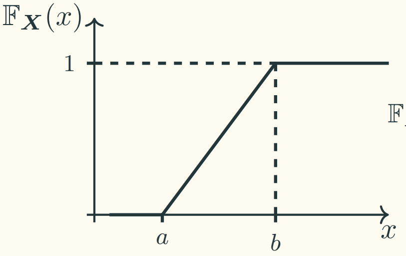
$$\mathbb{F}_{\boldsymbol{X}}(x) = \mathbb{P}(\boldsymbol{X} \leqslant x) = \int_{-\infty}^{x} \mathbb{f}_{\boldsymbol{X}}(t)\,\mathrm{d}t$$

$$\mathbb{F}_{\boldsymbol{X}}(a) = \mathbb{P}(\boldsymbol{X} \leqslant a) = \int_{-\infty}^{a} \mathbb{f}_{\boldsymbol{X}}(x)\,\mathrm{d}x$$

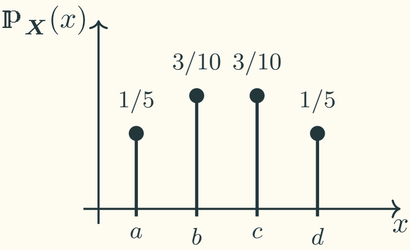
$$\mathbb{F}_{\boldsymbol{X}}(b) = \mathbb{P}(\boldsymbol{X} \leqslant b) = \int_{-\infty}^{b} \mathbb{f}_{\boldsymbol{X}}(x)\,\mathrm{d}x$$

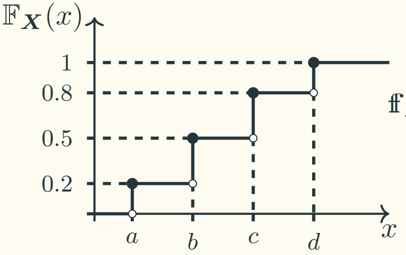
Cumulative Distribution Functions (CDFs)





$$\mathbb{F}_{\boldsymbol{X}}(x) = \mathbb{P}\left(\boldsymbol{X} \leqslant x\right)$$
$$= \int_{-\infty}^{x} \mathbb{f}_{\boldsymbol{X}}(t)\,\mathrm{d}t$$





$$\mathbb{f}_{\boldsymbol{X}}(x) = \mathbb{P}\left(\boldsymbol{X} \leqslant x\right)$$
$$= \sum_{k \leqslant x} \mathbb{P}_{\boldsymbol{X}}(k)$$

Expectation and Variance

The continuous case

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

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

$$\mathbb{V}\mathrm{ar}(X) = \sigma_X^2 = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^2\right]$$
$$= \int_{-\infty}^{+\infty} (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^2\, \mathbb{f}_{\boldsymbol{X}}(x)\, \mathrm{d}x$$
$$= \mathbb{E}\left[\boldsymbol{X}^2\right] - (\mathbb{E}[\boldsymbol{X}])^2$$

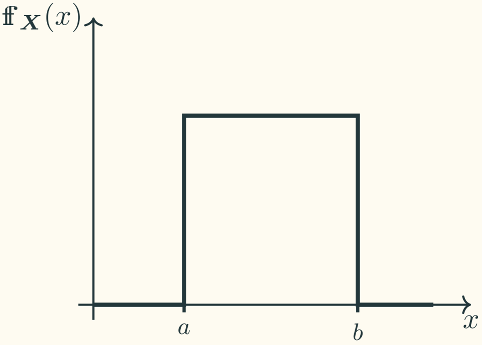
The discrete case

$$\mathbb{E}[\boldsymbol{X}] = \sum_x x\mathbb{P}_{\boldsymbol{X}}(x)$$

$$\mathbb{E}[g(\boldsymbol{X})] = \sum_x g(x)\mathbb{P}_{\boldsymbol{X}}(x)$$

$$\mathbb{V}\mathrm{ar}(\boldsymbol{X}) = \sigma_{\boldsymbol{X}}^2 = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^2\right]$$
$$= \sum_x (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^2\, \mathbb{P}_{\boldsymbol{X}}(x)$$
$$= \mathbb{E}\left[\boldsymbol{X}^2\right] - (\mathbb{E}[\boldsymbol{X}])^2$$

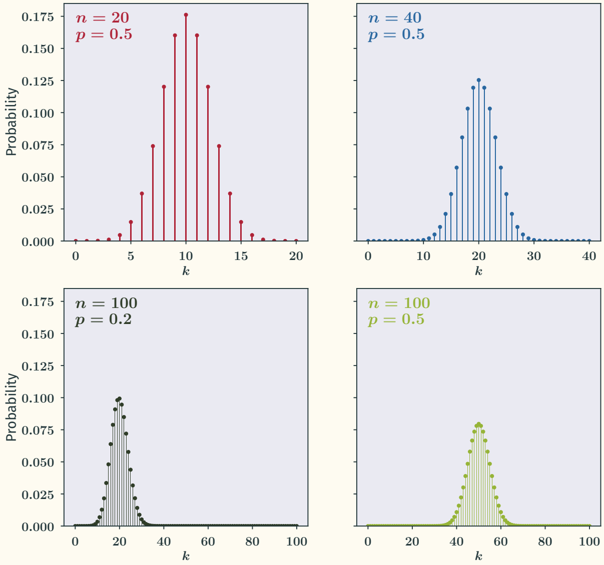
Continuous Uniform Distribution

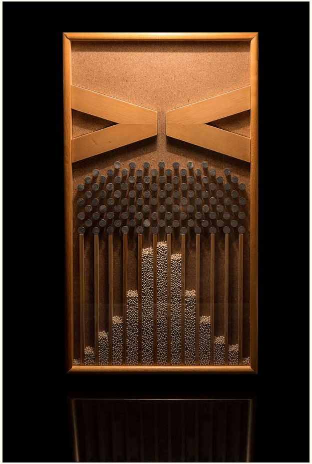


$$f_{X}(x) = \begin{cases} \frac{1}{b-a}, & a \leqslant x \leqslant b \\ 0, & x < a \text{ or } x > b \end{cases}$$

$$\mathbb{E}[X] = ? \qquad \mathbb{V}\text{ar}(X) = ?$$

The Idea of The Normal Distributions



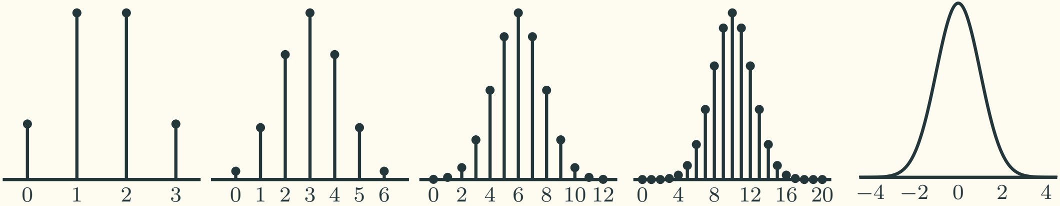


The Bean Machine
by Francis Galton

A Little History of The Normal Distribution - Binomial Approximation

Abraham de Moivre: The Doctrine of Chances, 1738

- The middle term of the binomial coefficient is approximately: $\dfrac{2}{\sqrt{2\pi n}}$
- Stirling's formula: $n! \simeq n^n e^{-n}\sqrt{2\pi n}$



The de Moivre–Laplace Theorem

When n becomes large, and np, nq are also large:

$$\mathbb{P}_{\boldsymbol{X}}(k) = \binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi}\sqrt{npq}} \cdot e^{-\dfrac{(k-np)^2}{2npq}}, where\ q = 1 - p$$

A Little History of the Normal Distribution - The Error Curve

Carl Friedrich Gauss: Theoria motus corporum coelestium in sectionibus conicis solem ambientium (Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections), 1809.

- least squares
- maximum likelihood
- normal distribution

Pierre Simon de Laplace

- In 1782: $\int_{-\infty}^{\infty} e^{-t^2} \mathrm{d}t = \sqrt{\pi}$
- In 1810: the central limit theorem

A Little History of the Normal Distribution - Beyond Errors

Adolphe Quetelet: the average man - Letters addressed to H.R.H. the grand duke of Saxe Coburg and Gotha, on the Theory of Probabilities as Applied to the Moral and Political Sciences, 1846.

TABLE 1: Chest measurement of Scottish soldiers

Girth	Frequency
33	3
34	18
35	81
36	185
37	420
38	749
39	1,073
40	1,079
41	934
42	658
43	370
44	92
45	50
46	21
47	4
48	1