Lecture 12 Continuous Probability Distribution

BIO210 Biostatistics

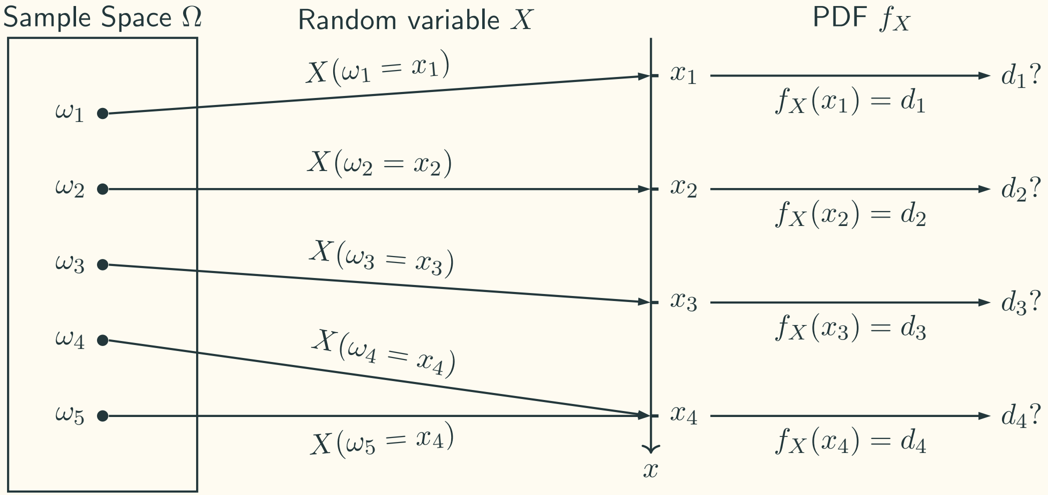
Xi Chen

Spring, 2022

School of Life Sciences
Southern University of Science and Technology

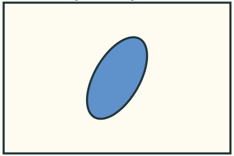
南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

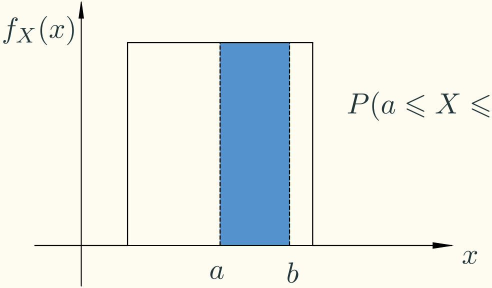
Probability Density Function (PDF)



Probability Density Function (PDF)

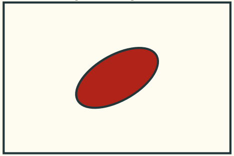
Sample space Ω

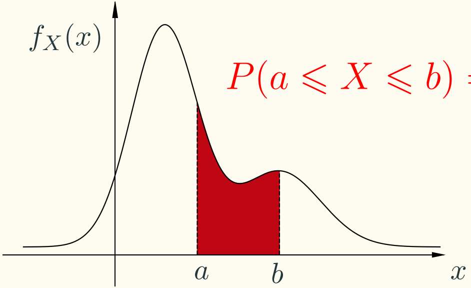




$$P(a \leqslant X \leqslant b) = f_X(x) \cdot (b - a)$$

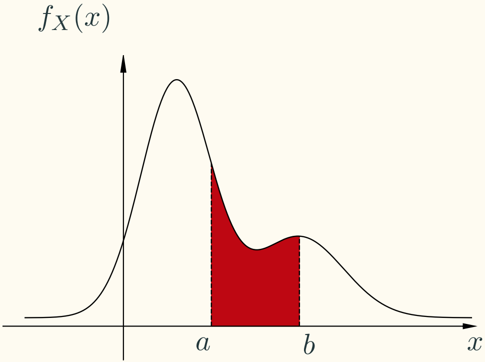
Sample space Ω





$$P(a \leqslant X \leqslant b) = \int_a^b f_X(x)\, dx$$

Probability Density Function (PDF)



$$f_X(x) \geqslant 0, \int_{-\infty}^{+\infty} f_X(x)\, dx = 1$$

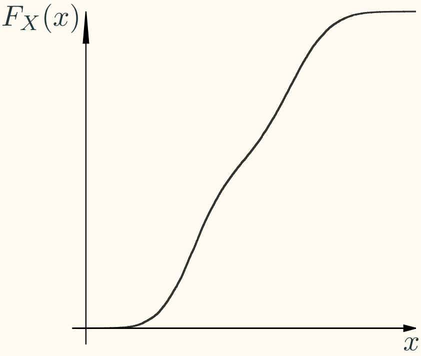
$$P(X = a) = ?$$

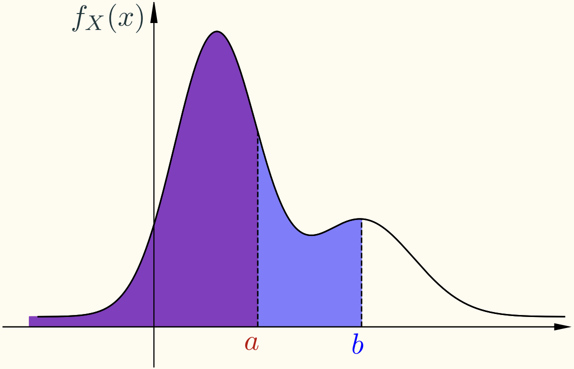
$$P(x \leqslant X \leqslant x + \delta)$$

$$= \int_{x}^{x+\delta} f_X(x)\, dx = f_X(x) \cdot \delta$$

$$f_X(x) = \frac{P(x \leqslant X \leqslant X + \delta)}{\delta}$$

Cumulative Distribution Function



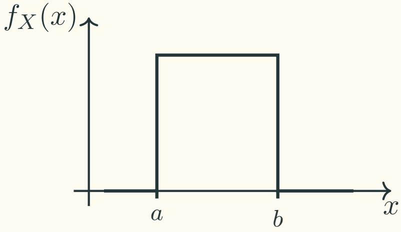


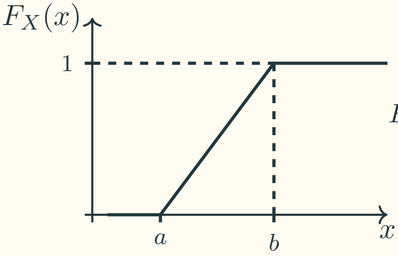
$$F_X(x) = P(X \leqslant x) = \int_{-\infty}^{x} f_X(t)\,dt$$

$$F_X(a) = P(X \leqslant a) = \int_{-\infty}^{a} f_X(x)\,dx$$

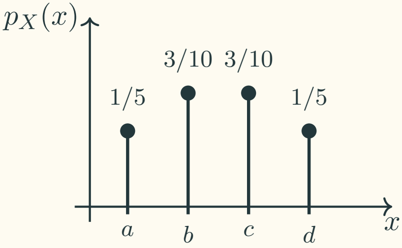
$$F_X(b) = P(X \leqslant b) = \int_{-\infty}^{b} f_X(x)\,dx$$

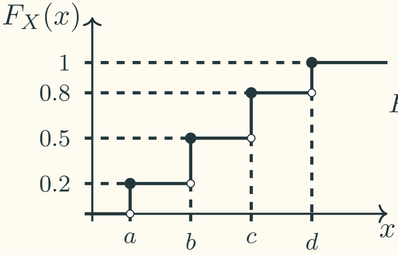
Cumulative Distribution Function (CDF)





$$F_X(x) = P(X \leqslant x) = \int_{-\infty}^{x} f_X(t)\, dt$$





$$F_X(x) = P(X \leqslant x) = \sum_{k \leqslant x} p_X(k)$$

Expectation and Variance

The continuous case

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\, dx$$

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

$$var(X) = \sigma_X^2 = E\left[(X - E[X])^2\right]$$

$$= \int_{-\infty}^{+\infty} (X - E[X])^2 f_X(x)\, dx$$

$$= E[X^2] - (E[X])^2$$

The discrete case

$$E[X] = \sum_x x p_X(x)$$

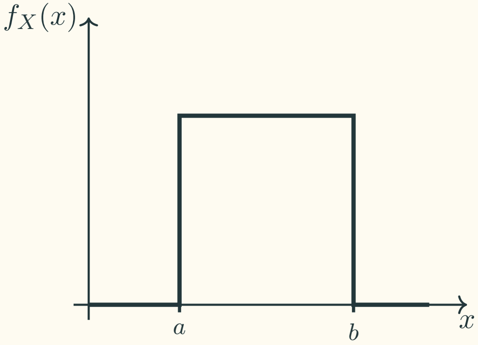
$$E[g(X)] = \sum_x g(x) p_X(x)$$

$$var(X) = \sigma_X^2 = E\left[(X - E[X])^2\right]$$

$$= \sum_x (X - E[X])^2 p_X(x)$$

$$= E[X^2] - (E[X])^2$$

Continuous Uniform Distribution

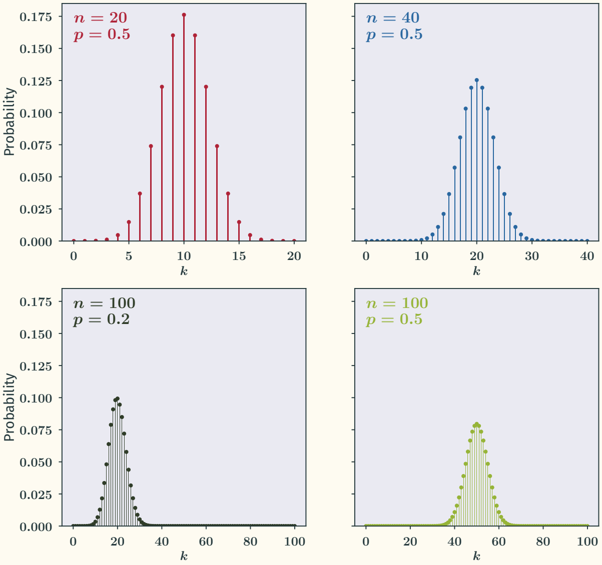


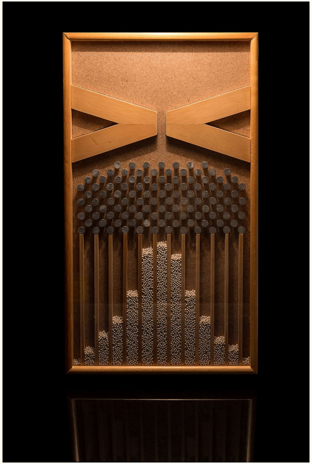
$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \leqslant x \leqslant b \\ 0\,, & x < a \ or \ x > b \end{cases}$$

$$E[X] = ?$$

$$var(X) = ?$$

The idea of the normal distribution



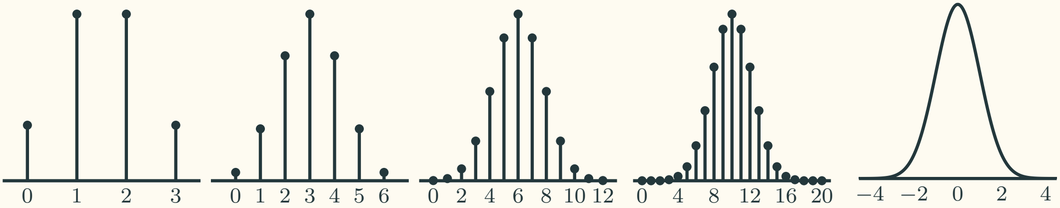


The Bean Machine
by Francis Galton

A Little History of the Normal Distribution

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: $n! \simeq n^n e^{-n}\sqrt{2\pi n}$
- In 1730: $\int_0^\infty e^{-x^2}dx = \dfrac{\sqrt{\pi}}{2}$



A Little History of the Normal Distribution

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 1810: the central limit theorem

Derivation of the Normal PDF equation

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

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