Lecture 12 Continuous Probability Distribution

BIO210 Biostatistics

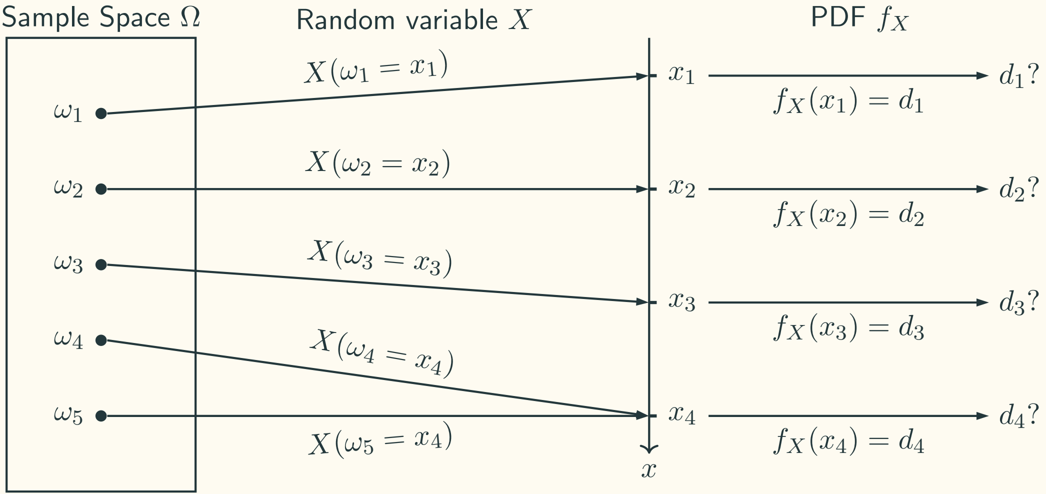
Xi Chen

Spring, 2023

School of Life Sciences
Southern University of Science and Technology

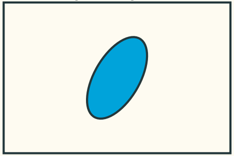
南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

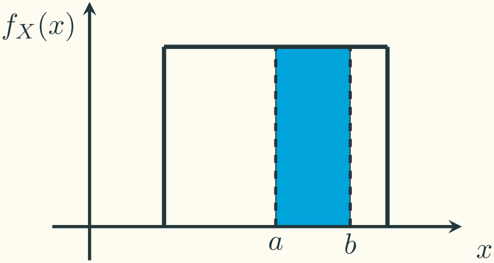
Probability Density Function (PDF)



Probability Density Function (PDF)

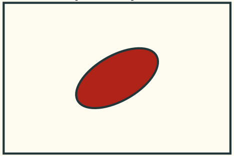
Sample space Ω

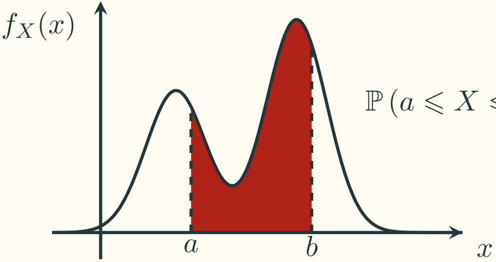




$$\mathbb{P}\left(a \leqslant X \leqslant B\right)$$
$$= f_X(x) \cdot (b-a)$$

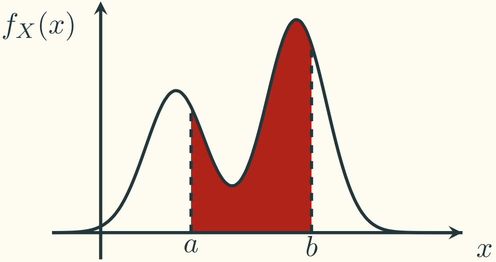
Sample space Ω





$$\mathbb{P}\left(a \leqslant X \leqslant b\right) = \int_a^b f_X(x)\mathrm{d}x$$

Probability Density Functions (PDFs)



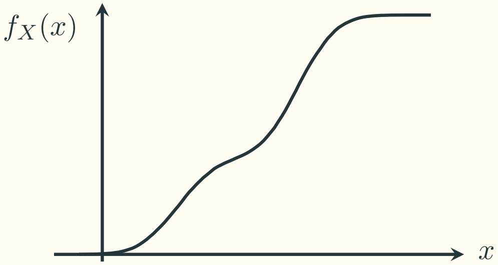
$$f_X(x) \geqslant 0, \int_{-\infty}^{+\infty} f_X(x)\,\mathrm{d}x = 1$$

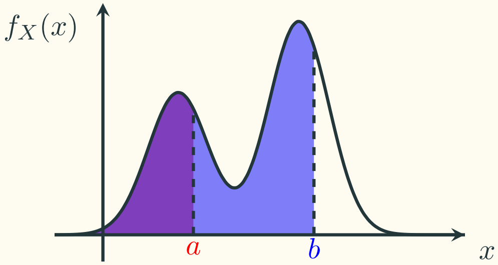
$$\mathbb{P}\left(X = a\right) = ?$$

$$\mathbb{P}\left(x \leqslant X \leqslant x + \delta\right)$$
$$= \int_{x}^{x+\delta} f_X(x)\,\mathrm{d}x = f_X(x) \cdot \delta$$

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

Cumulative Distribution Function



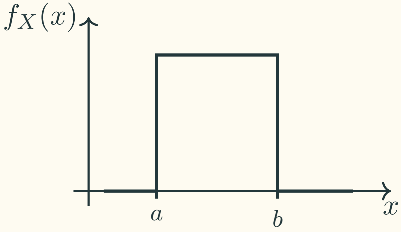


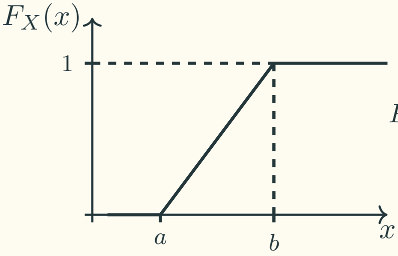
$$F_X(x) = \mathbb{P}\left(X \leqslant x\right) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t$$

$$F_X(a) = \mathbb{P}\left(X \leqslant a\right) = \int_{-\infty}^{a} f_X(x)\,\mathrm{d}x$$

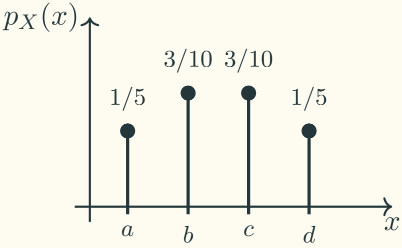
$$F_X(b) = \mathbb{P}\left(X \leqslant b\right) = \int_{-\infty}^{b} f_X(x)\,\mathrm{d}x$$

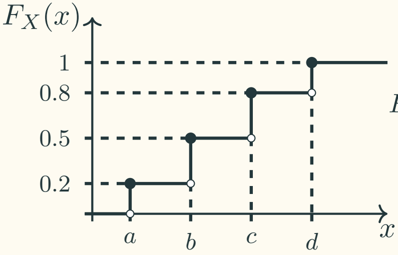
Cumulative Distribution Functions (CDFs)





$$
\begin{aligned}
F_X(x) &= \mathbb{P}(X \leqslant x) \\
&= \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t
\end{aligned}
$$





$$
\begin{aligned}
F_X(x) &= \mathbb{P}(X \leqslant x) \\
&= \sum_{k \leqslant x} p_X(k)
\end{aligned}
$$

Expectation and Variance

The continuous case

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

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

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

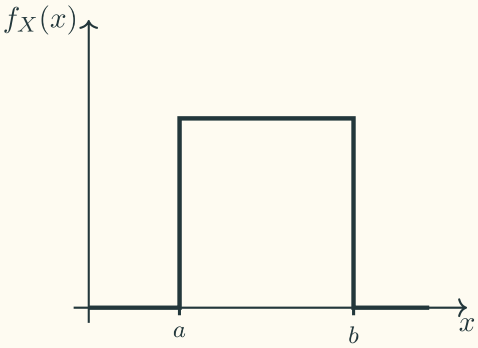
The discrete case

$$\mathbb{E}\left[X\right] = \sum_x x p_X(x)$$

$$\mathbb{E}\left[g(X)\right] = \sum_x g(x) p_X(x)$$

$$\mathbb{V}\mathrm{ar}\left(X\right) = \sigma_X^2 = E\left[(X - \mathbb{E}\left[X\right])^2\right]$$
$$= \sum_x (X - \mathbb{E}\left[X\right])^2 p_X(x)$$
$$= \mathbb{E}\left[X^2\right] - (\mathbb{E}\left[X\right])^2$$

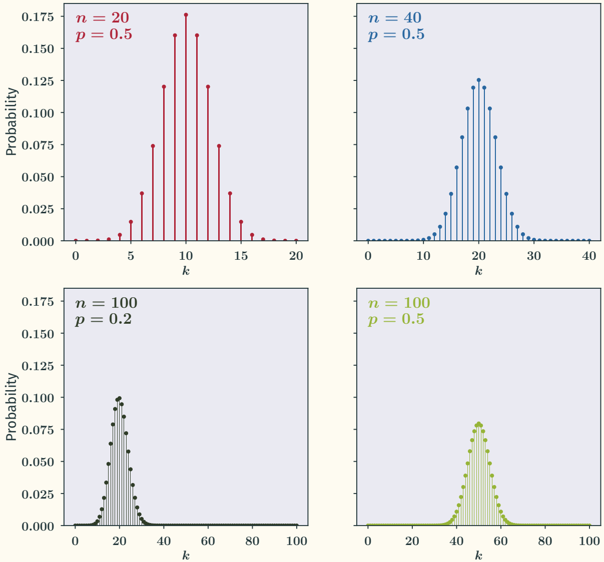
Continuous Uniform Distribution

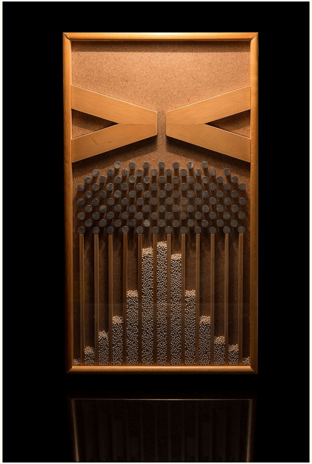


$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \leqslant x \leqslant b \\ 0\,, & x < a \text{ or } x > b \end{cases}$$

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

The Idea of The Normal Distributions



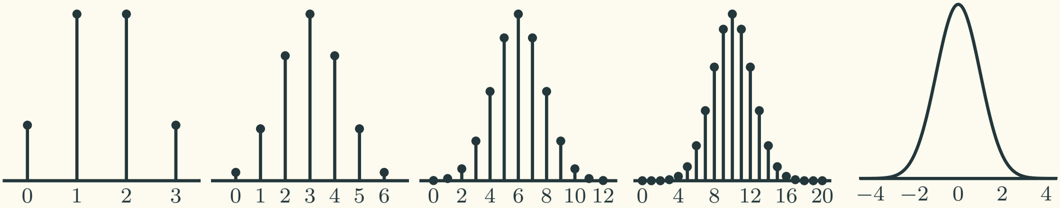


The Bean Machine
by Francis Galton

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: $n! \simeq n^n e^{-n} \sqrt{2\pi n}$



The de Moivre–Laplace Theorem

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

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

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