Probability and Statistics

Part-6

MA2103 - 2023

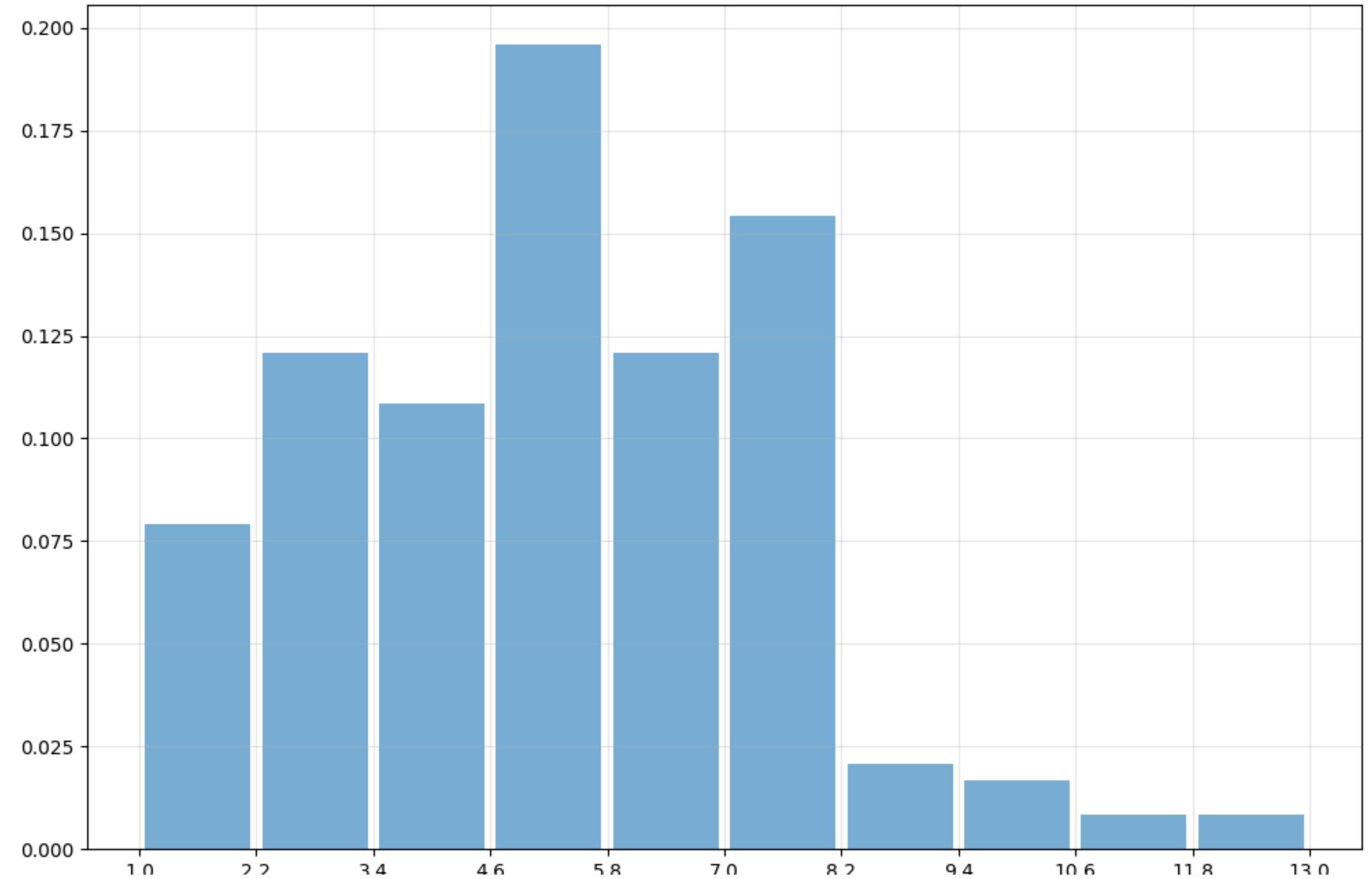
Rajesh Kumble Nayak

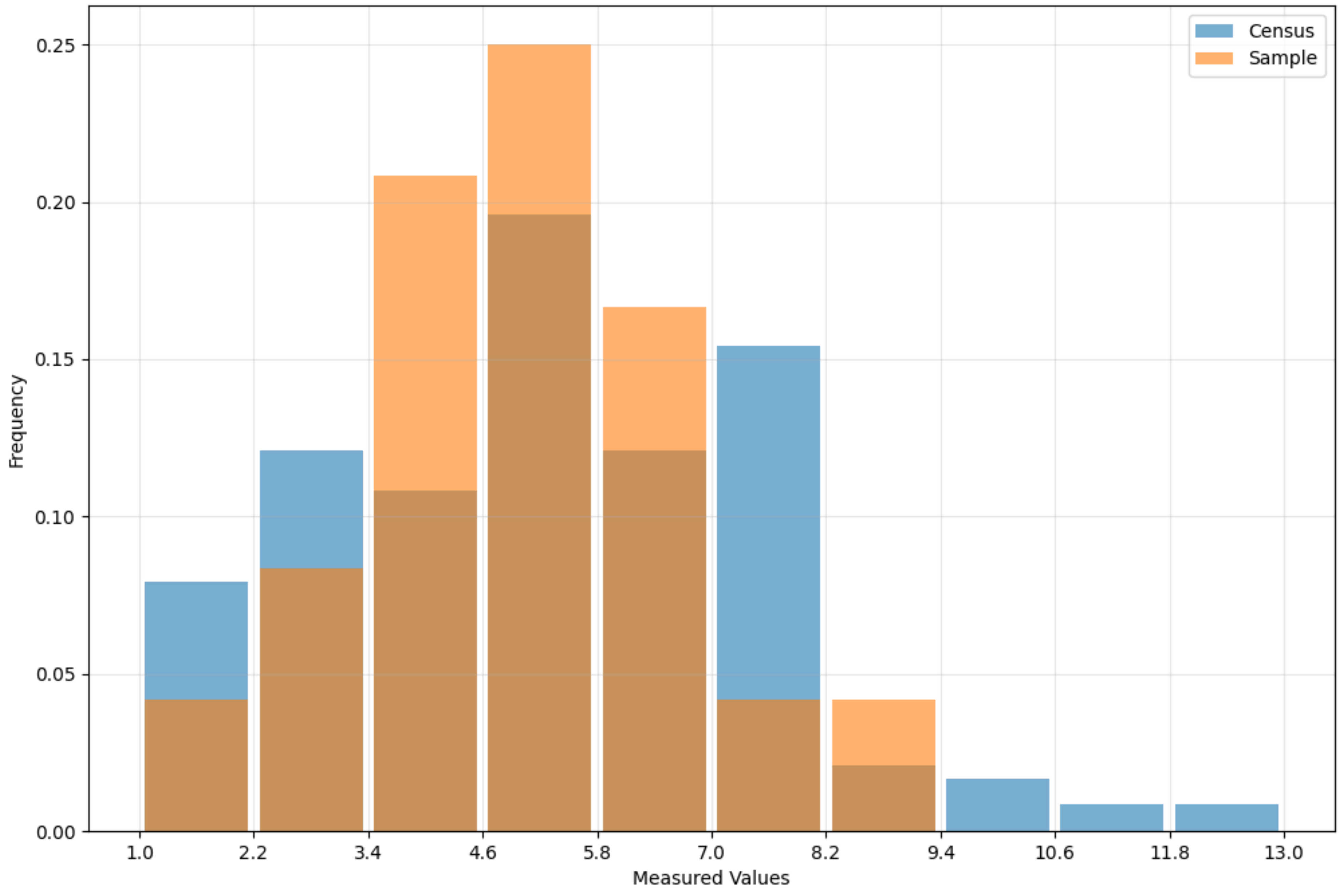
More on distribution function

As we have seen histogram provides chance of occurrence at specific value and given interval

However, it is not suitable for comparing two distributions.

We can normalise the histogram with total area under the curve and we get the probability density function or PDF.





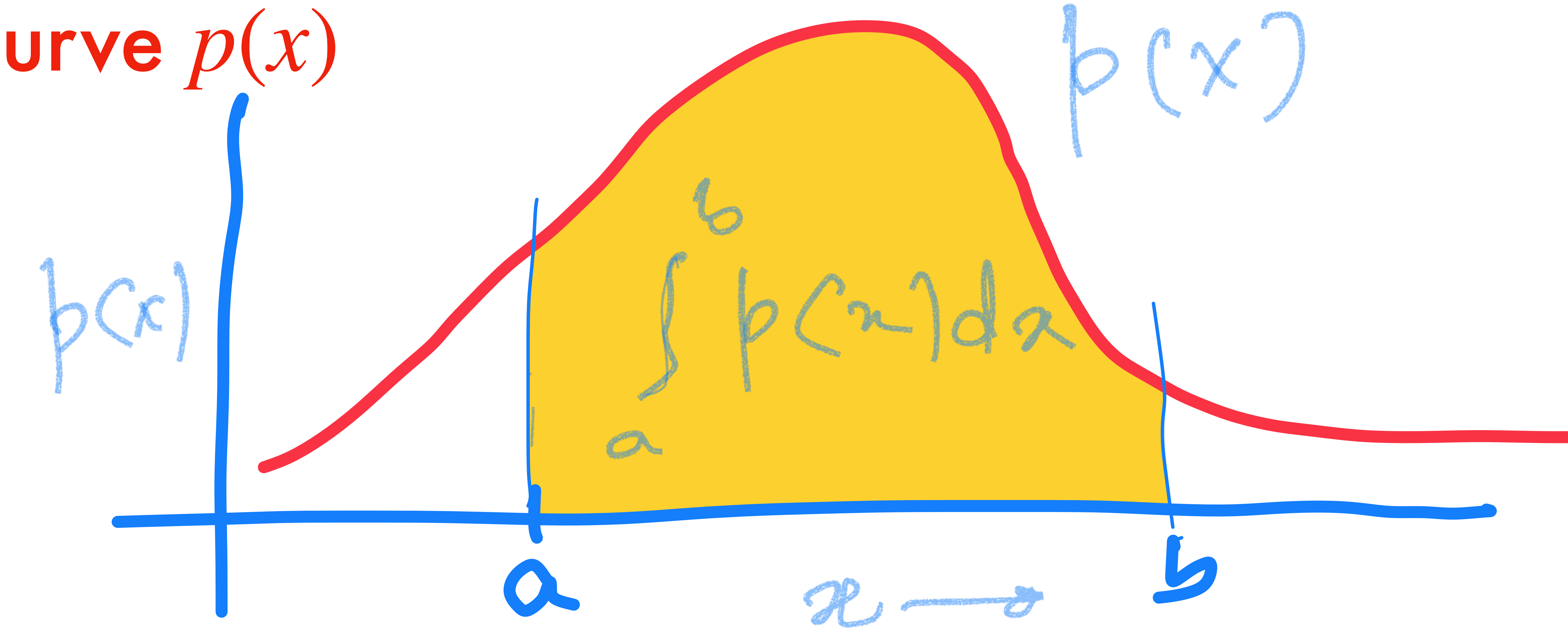
Probability Density Function (PDF)

Probability Density Function or Probability distribution $p(x)$ gives the probability of occurrence of event at x and $x + dx$ and is given by $p(x)dx$

One often normalises the PDF, i.e $\int_{-\infty}^{+\infty} p(x)dx = 1$

The probability of occurrence of event in the interval $[a, b]$ is given by:

$P([a, b]) = \int_{a}^{b} p(x)dx$, the area under the curve $p(x)$



Examples

Uniform probability density

Uniform probability density function is given by

$$p(x) = \begin{cases} \dfrac{1}{(b-a)}, & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Here, mean μ is given by $\mu = \dfrac{1}{2}(a+b)$ and $\sigma^2 = \dfrac{1}{12}(b-a)^2$

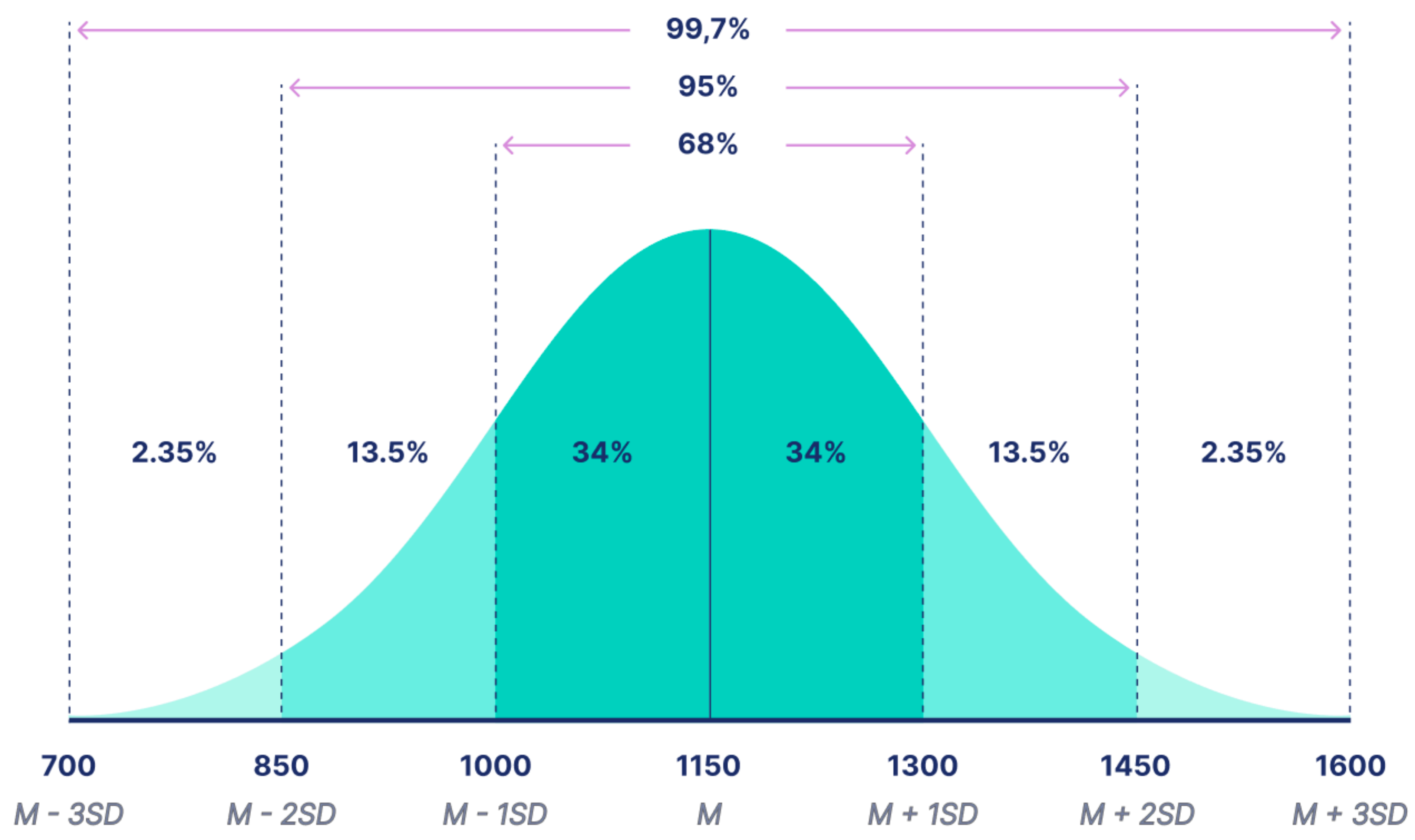
Here every point x is equally likely in the interval $[a, b]$

Ex: Tossing coin, throwing dice etc... 🎲 🎲 🪙 🃏

Gaussian or normal distribution:

Normal probability density function is given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



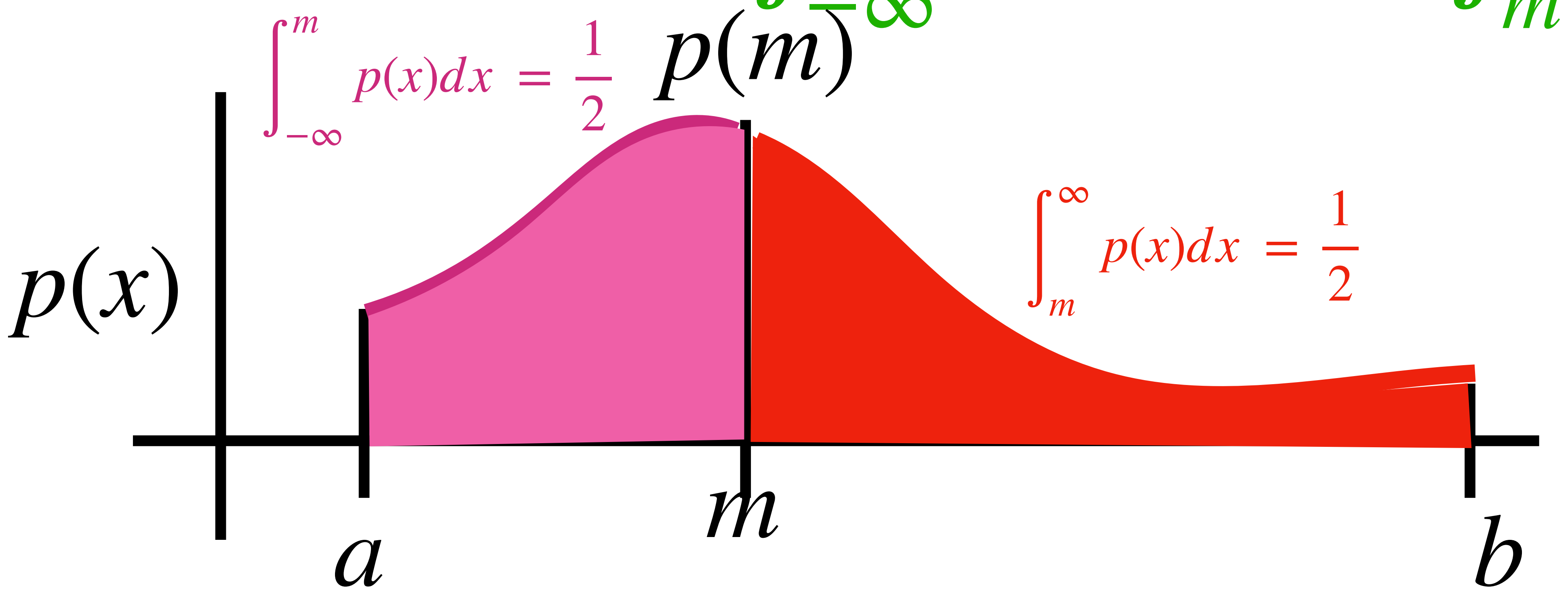
Here, mean μ is represented by μ and variance is given by σ^2

These are the most popular distributions

Statistical measures from the PDF

PDF $p(x)$ should contain all the statistical information information

- ➡ Normalisation: PDF, $p(x)$ is normalised i.e $\int_{-\infty}^{\infty} p(x)dx = 1$
- ➡ First moment or mean of a distribution is given by $\mu = \int_{-\infty}^{\infty} x\, p(x)dx$
- ➡ Variance of a distribution is given by $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2\; p(x)\, dx$
- ➡ Median of, `m' of a distribution is such that $\int_{-\infty}^{m} p(x)dx = \int_{m}^{\infty} p(x)dx = \frac{1}{2}$



Examples: ?Find the mean and median and variance of uniform distribution

PDF for the uniform distribution is given by,

$$p(x) = \begin{cases} \frac{1}{(b-a)}, & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Let's see if the $p(x)$ is normalised! $\int_{-\infty}^{\infty} p(x)\,dx = \int_{a}^{b} \frac{1}{(b-a)}\,dx = 1$

Let's find the mean, $\mu = \int_{-\infty}^{\infty} x\,p(x)dx = \frac{1}{2}(a+b)$

$$\mu = \int_{-\infty}^{\infty} x\,p(x)\,dx = \frac{1}{(b-a)} \int_{a}^{b} x\,dx = \frac{1}{2(b-a)} b^2 - a^2$$

Median of, `m' of a distribution is such that $\int_{-\infty}^{m} p(x)dx = \int_{m}^{\infty} p(x)dx = \frac{1}{2}$

Intuitively mid point between a and b should be the median

$$\int_a^m \frac{1}{b-a}\,dx + \int_m^b \frac{1}{(b-a)}\,dx = \frac{1}{2}$$

$$\Rightarrow \frac{a-m}{b-a} = \frac{m-b}{b-a} = \frac{1}{2}$$

$$a-m = \frac{1}{2}(b-a), \quad m-b = \frac{1}{2}(b-a)$$

$$\Rightarrow 2m = (a+b) \Rightarrow m = \frac{1}{2}(a+b)$$

Let's compute the variance, σ^2

$$\sigma^2 = \int_{-\infty}^{+\infty} \left(x-\mu\right)^2 p(x)\,dx$$

$$\sigma^2 = \int_a^b (x-\mu)^2 \cdot \frac{1}{(b-a)}\,dx = \frac{1}{(b-a)} \frac{(x-\mu)^3}{3}\Bigg|_a^b$$

$$= \frac{1}{3(b-a)}\left[(b-\mu)^3 - (a-\mu)^3\right]$$

$$= \frac{1}{3(b-a)}\left[\left(b-\frac{a}{2}-\frac{b}{2}\right)^3 - \left(a-\frac{a}{2}-\frac{b}{2}\right)^3\right]$$

$$= \frac{1}{3(b-a)}\left[\left(\frac{b}{2}-\frac{a}{2}\right)^3 + \left(\frac{b}{2}-\frac{a}{2}\right)^3\right]$$

$$\sigma^2 = \frac{1}{12}(b-a)^2$$

Examples: For Gaussian distribution given by, $p(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Show that μ is the mean and σ^2

We have $\displaystyle\int_{-\infty}^{+\infty} p(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$

Mean is given by, $\displaystyle\int_{-\infty}^{+\infty} xp(x)\, dx$

$$I = \int x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot dx$$

$$I = \int_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Let $y = x - \mu$ $\quad x = y + \mu$

$$I = \int_{-\infty}^{\infty} \frac{y+\mu}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2} \cdot dy$$

$$= \int_{-\infty}^{\infty} \frac{y}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2} \cdot d\cancel{x}y$$

$$+ \mu \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2} \cdot dy}_{1}$$

is zero because odd function

$$I = \mu$$

Let's compute the variance, σ^2

$$I = \int_{-\infty}^{+\infty} (x-\mu)^2 \; p(x)\, dx$$

$$I = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Substitute $y = \left(\frac{x-\mu}{\sigma}\right)$

and $dy = \frac{dx}{\sigma}$

With this I becomes

$$I = \sigma^2 \int_{-\infty}^{+\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

Because $\int_{-\infty}^{+\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = 1$

$$I = \sigma^2$$