CUSP UCSL Summer 2016:
Probabilities and Probability distributions.

dr. federica bianco
fbianco@nyu.edu
@fedhere

Material related to this lecture can be found at
https://github.com/fedhere/UInotebooks/tree/master/UCSL2016

why do we bother with probabilities?

IF WE COULD SAMPLE THE ENTIRE *POPULATION* WE WOULD NOT NEED STATISTICS. Statistics allows us to say something about the entire population from what we know about a subsample.

- Outlier detection: e.g. is this event rare?

  (Hurrican Sandy, Meeting a Dinosaur in the streets of NYC)

- Prediction: based on current knowledge and current state, what is the probability that an event will occur?

  (can you predict a desease epidemic?)

- Average and typical: what is the average asthma in a certain area of the city? is it the same as in a difference area, where traffic is lower, or where there are fewer construction sites?

  (better allocation of resources, connectinc causes and effect)

Classical *Discrete* distibutions:

$$Bin(X = k \quad | \quad n, p) \qquad = \qquad \frac{n!}{(n-k)!k!}p^k(1-p)^{n-k}$$

$$Poi(X = k \quad | \quad \lambda) \qquad = \qquad e^{-\lambda}\frac{\lambda^k}{k!}$$

Classical *Discrete* distibutions

$$P(1 \text{ head in } 1 \text{ toss}) = Bin(1, 0.5) = 0.5$$

$$P(1 \text{ tail in } 1 \text{ tail}) = Bin(1, 0.5) = 0.5$$

Classical *Discrete* distibutions:

$$Bin(X = k \quad | \quad n, p) \qquad = \qquad \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

$$Poi(X = k \quad | \quad \lambda) \qquad = \qquad e^{-\lambda} \frac{\lambda^k}{k!}$$

Classical *Discrete* distibutions:

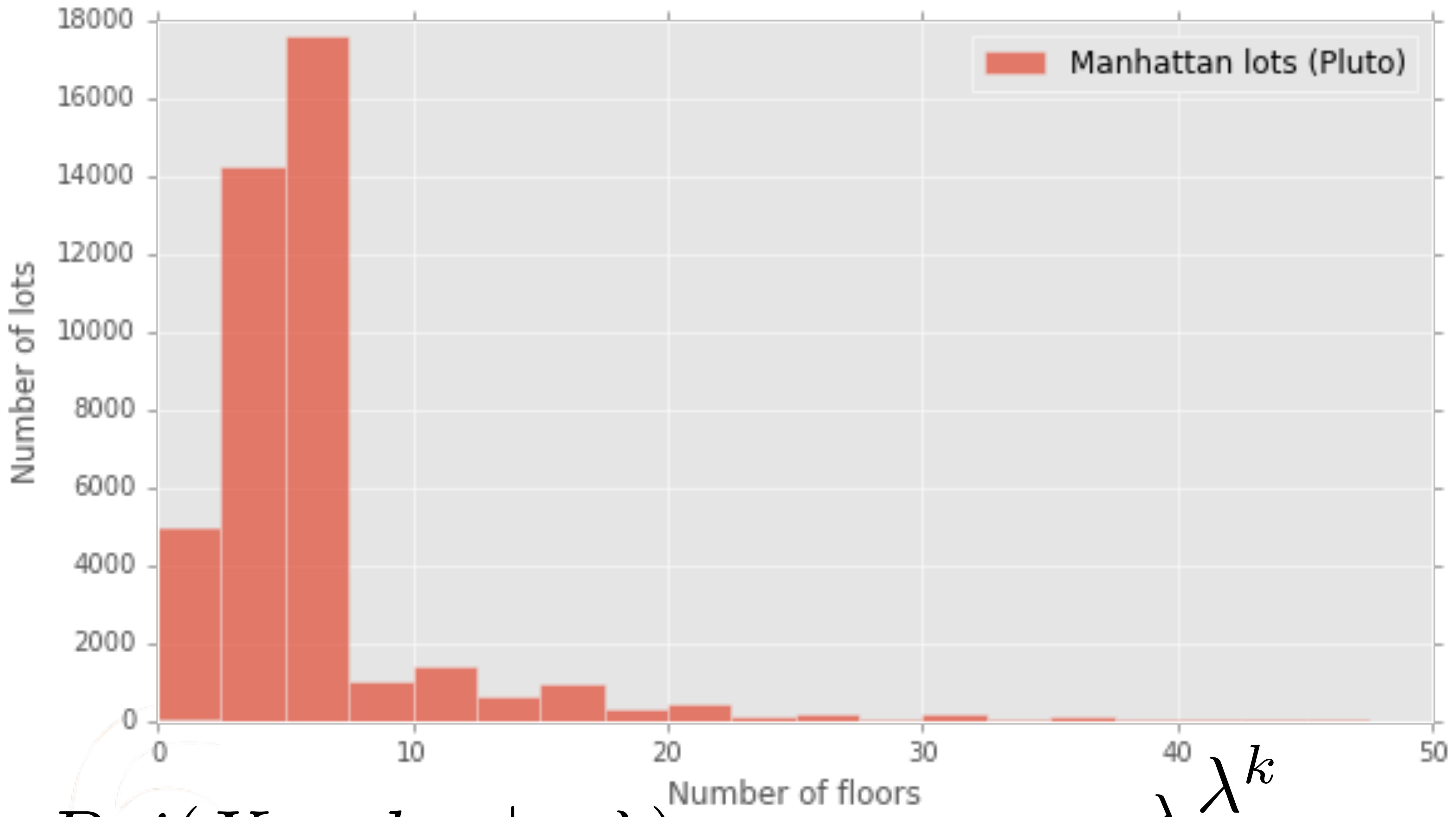
$$Bin(X = k \quad | \quad n, p) \qquad = \qquad \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

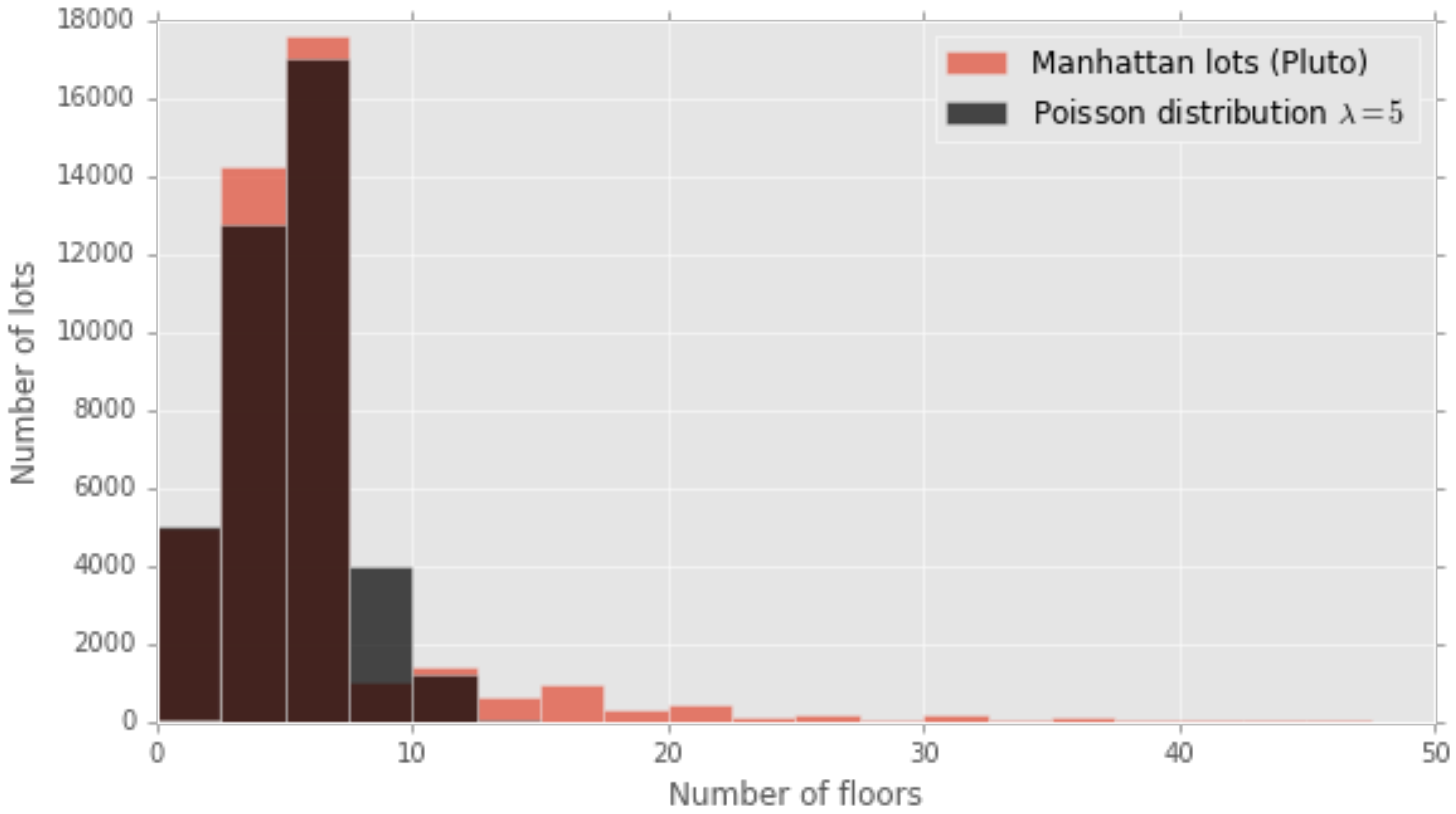
“Count statistics”:
the poisson distribution describes the probability of *number of occurrences*
non-negative integer values

$$Poi(X = k \quad | \quad \lambda) \qquad = \qquad e^{-\lambda} \frac{\lambda^k}{k!}$$

Classical *Discrete* distibutions

Classical *Discrete* distibutions:





$$Poi(X = k \quad | \quad \lambda) \quad = \quad e^{-\lambda}\frac{\lambda^k}{k!}$$

Classical *Continuous* distibutions:

Classical *Discrete* distibutions:

$$Bin(X = k \quad | \quad n, p) \qquad = \qquad \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

P(X=2 | 2) = 0.27,

$$Poi(X = k \quad | \quad \lambda) \qquad = \qquad e^{-\lambda} \frac{\lambda^k}{k!}$$

P(X=7 | 2) = 0.003

Classical *Continuous* distibutions:

Classical *Discrete* distibutions:

$$Bin(X = k \quad | \quad n, p) \qquad = \qquad \frac{n!}{(n-k)!k!}p^k(1-p)^{n-k}$$

$$Poi(X = k \quad | \quad \lambda) \qquad = \qquad e^{-\lambda}\frac{\lambda^k}{k!}$$

P(X=2 | 2) = 0.27,

P(X=7 | 2) = 0.003

P(~~X=0~~.2 | 2)

Classical *Discrete* distibutions:

$$Bin(X = k \quad | \quad n, p) \qquad = \qquad \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

$$Poi(X = k \quad | \quad \lambda) \qquad = \qquad e^{-\lambda} \frac{\lambda^k}{k!}$$

P(X=2 | 2) = 0.27,

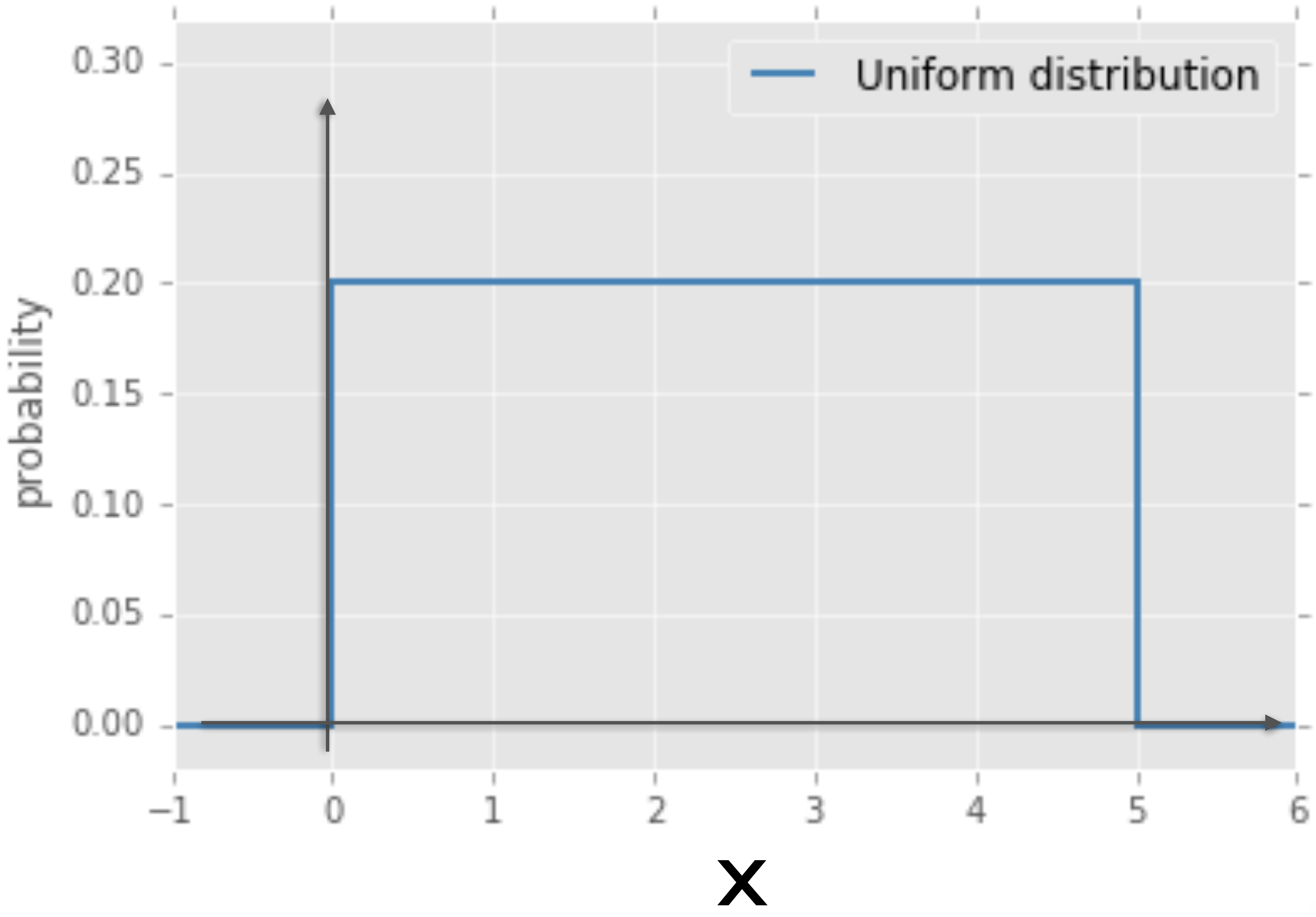
P(X=7 | 2) = 0.003

P(~~X=0.2~~ | 2)

5!=5 x 4 x 3 x 2 x 1=120

Classical *Continuous* distibutions:

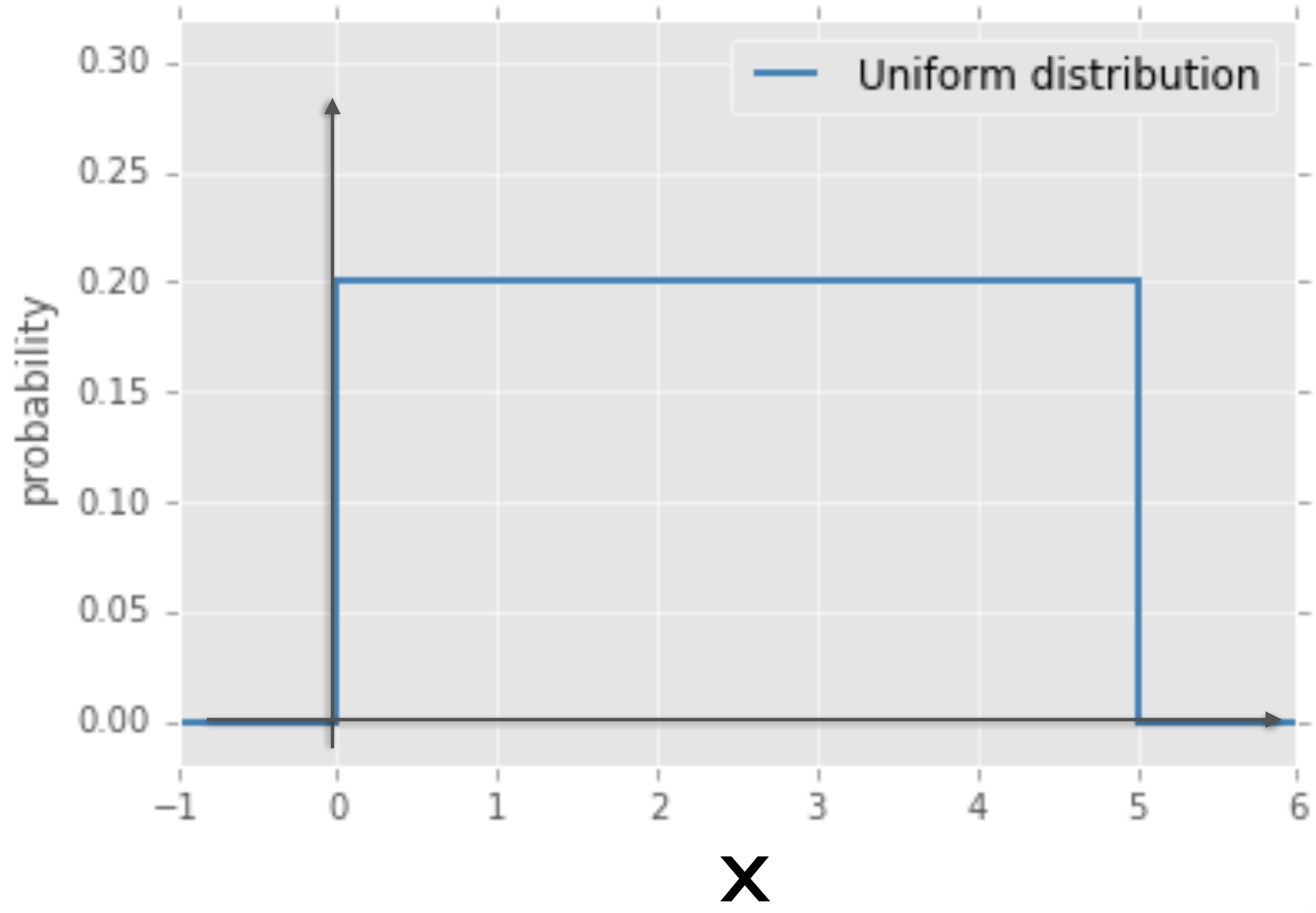
$$P(x) = \begin{cases} \mathsf{c} & 0 \le x \le 5 \\ 0 & \text{elsewhere} \end{cases}$$



Classical *Continuous* distibutions:

$$P(x) = \begin{cases} c & 0 \leq x \leq 5 \\ 0 & \text{elsewhere} \end{cases}$$

P(X = any value) = 1 by definition

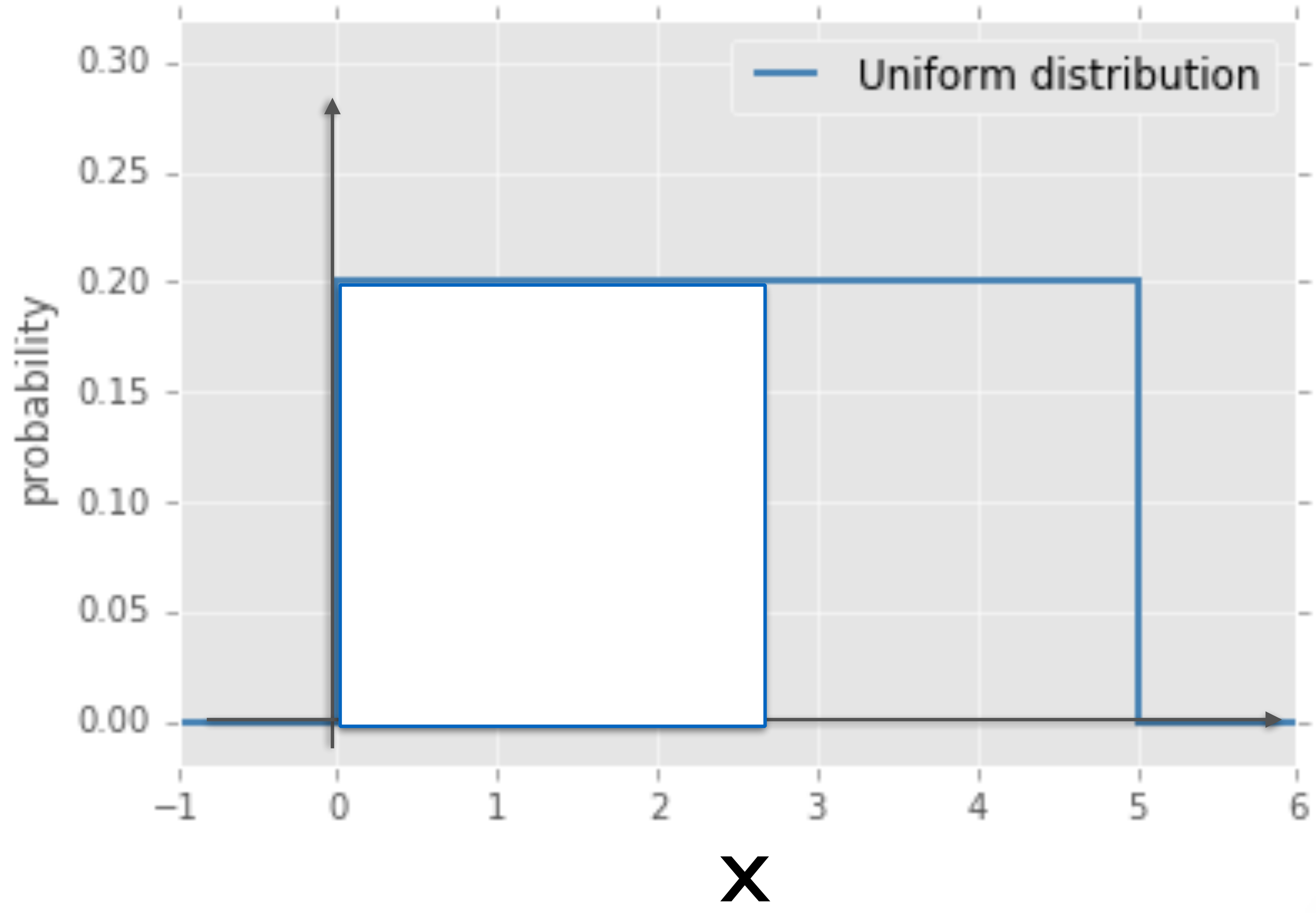


Classical *Continuous* distibutions:

$$P(x) = \begin{cases} c & 0 \leq x \leq 5 \\ 0 & \text{elsewhere} \end{cases}$$

P(X = any value) = 1 by definition

P(0< X = < 2.5) = 1/2 = 0.5



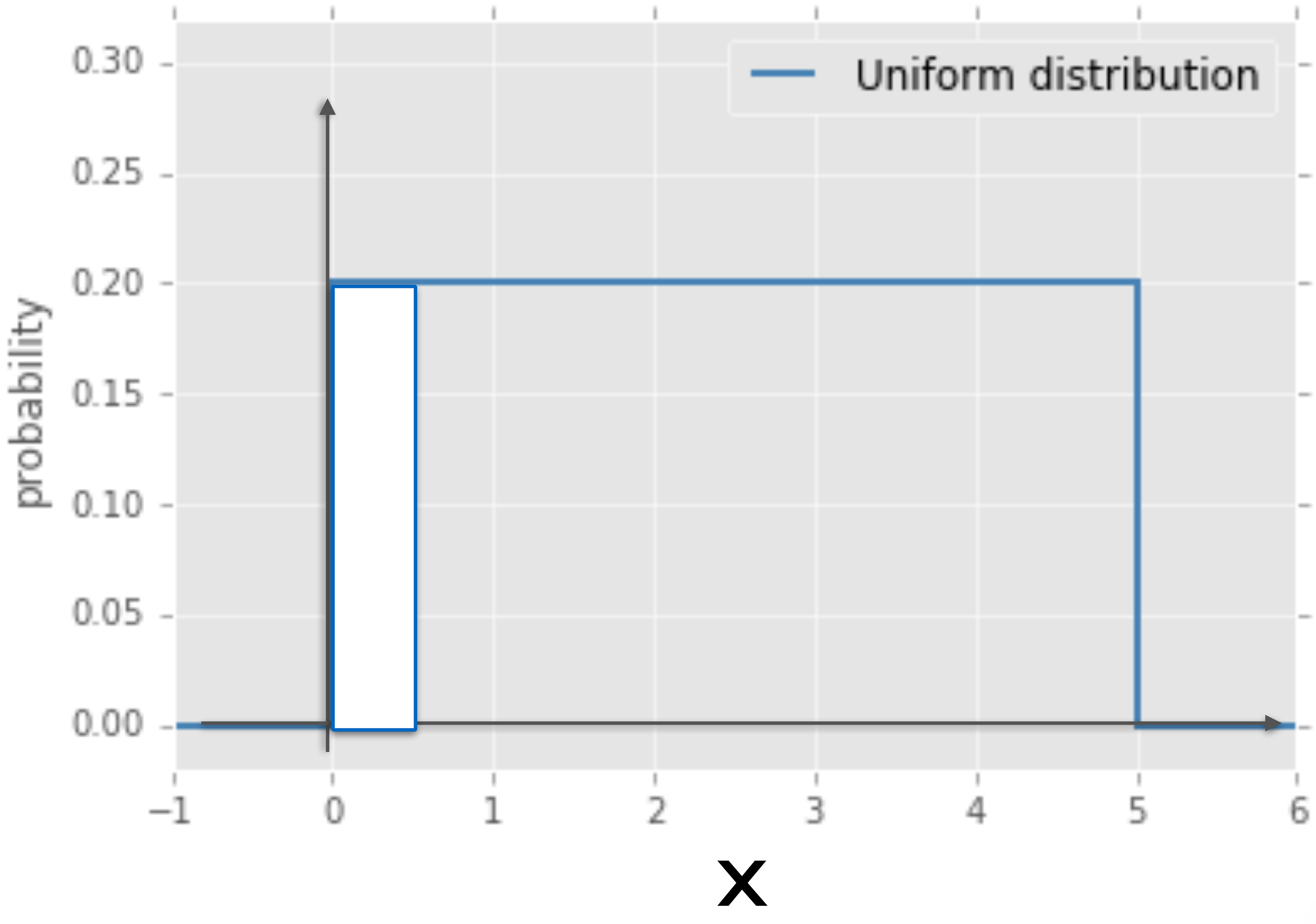
Classical *Continuous* distibutions:

$$P(x) = \begin{cases} c & 0 \leq x \leq 5 \\ 0 & \text{elsewhere} \end{cases}$$

P(X = any value) = 1 by definition

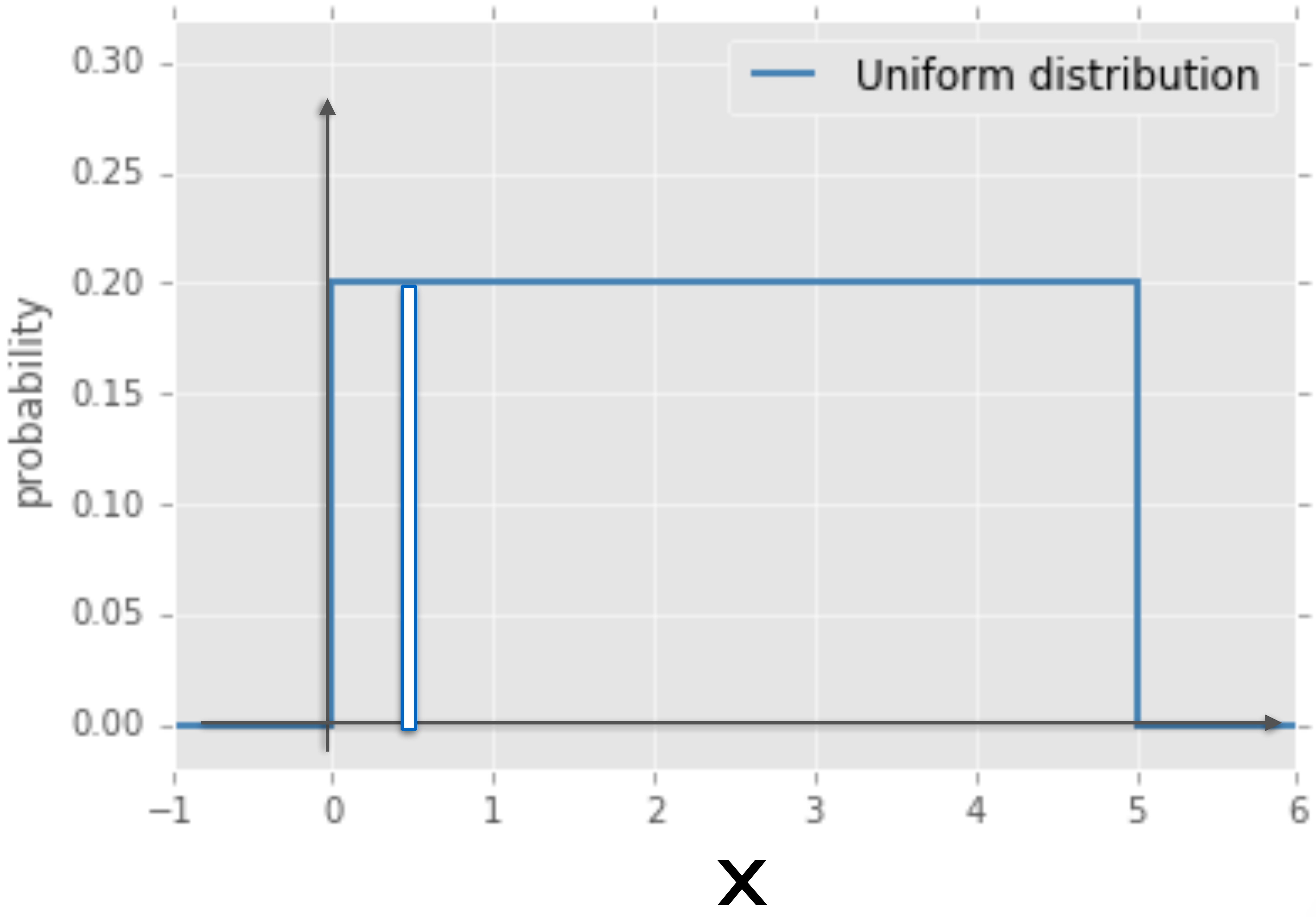
P(0< X = < 2.5) = 1/2 = 0.2x2.5 = 0.5

P(0< X = < 0.5) = 1/10 = 0.2x0.5 = 0.1



Classical *Continuous* distibutions:

$$P(x) = \begin{cases} c & 0 \leq x \leq 5 \\ 0 & \text{elsewhere} \end{cases}$$



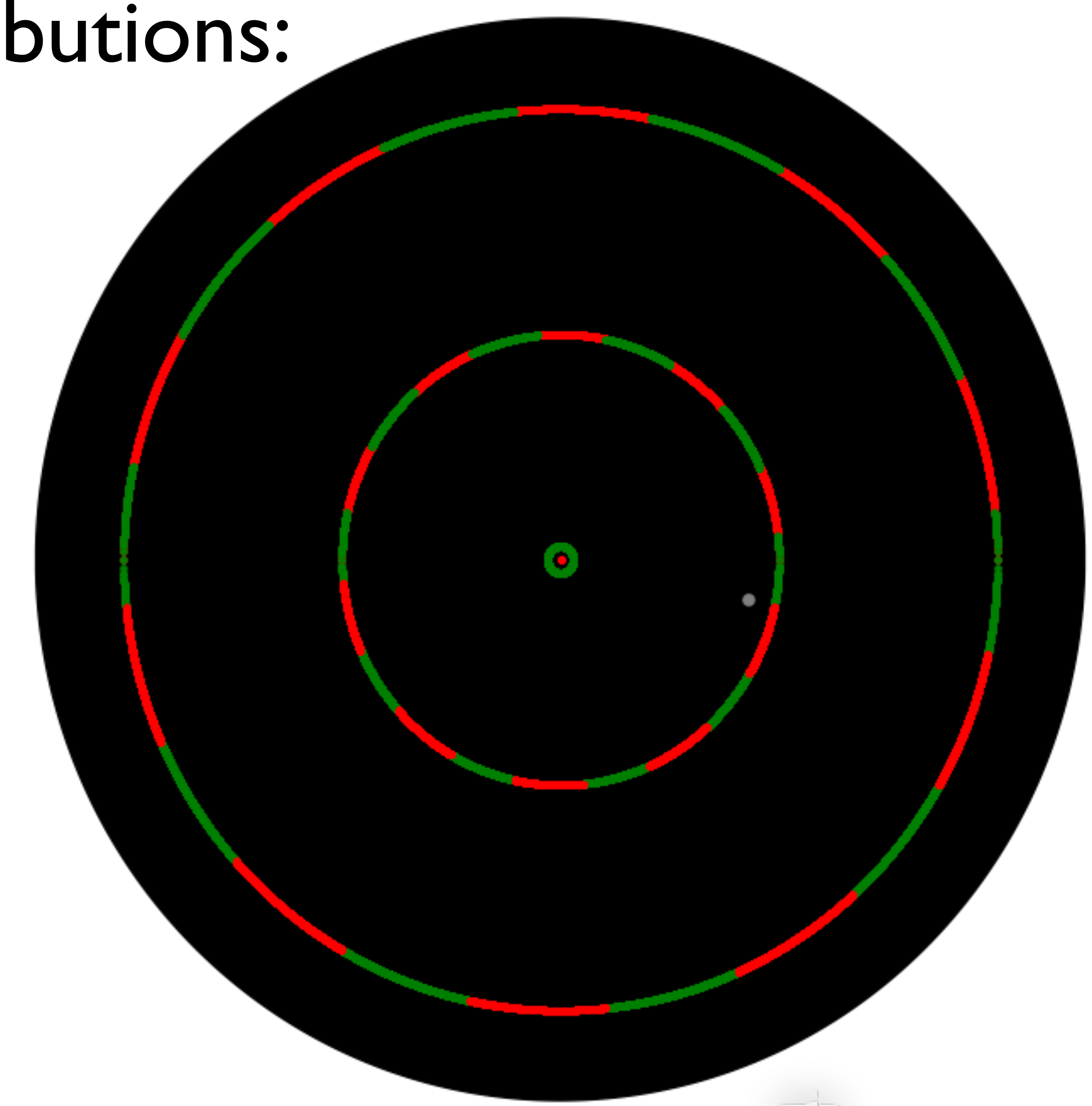
P(X = any value) = 1 by definition

P(0< X = < 2.5) = 1/2 = 0.2x2.5 = 0.5

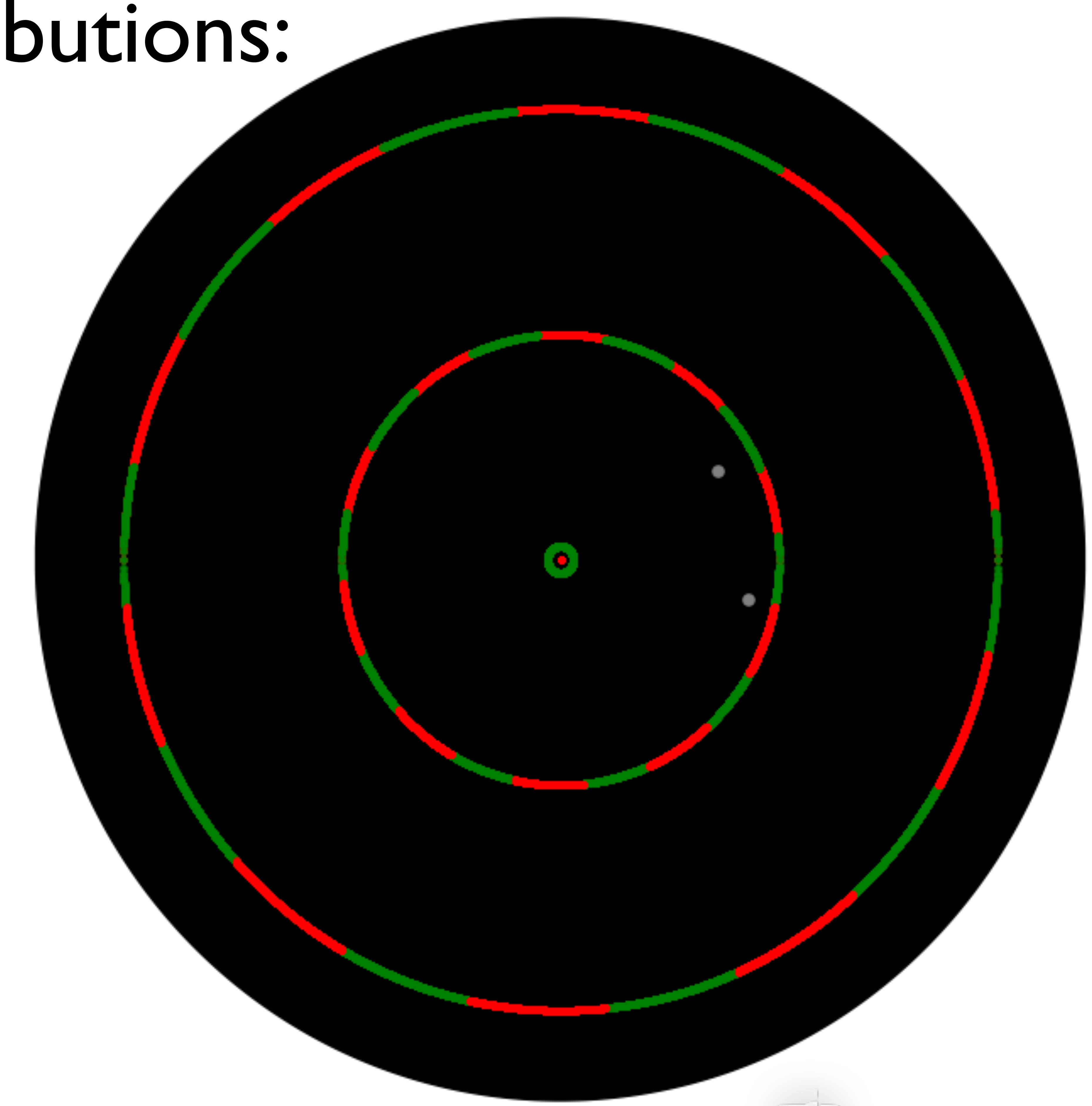
P(0< X = < 0.5) = 1/10 = 0.2x0.5 = 0.1

P(0.45< X = < 0.5) = 1/100 = 0.2x0.05 = 0.01

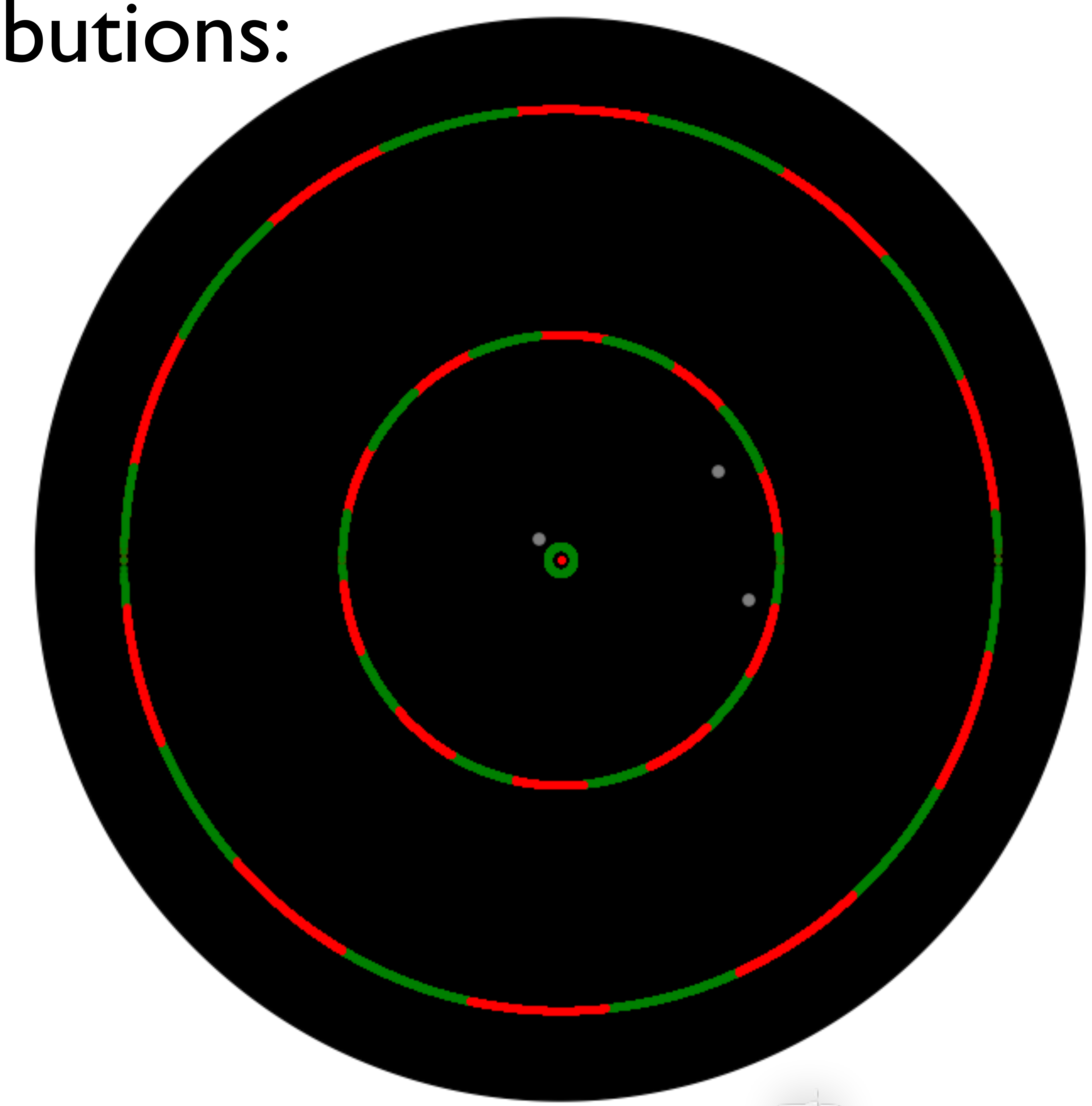
Classical *Continuous* distibutions:



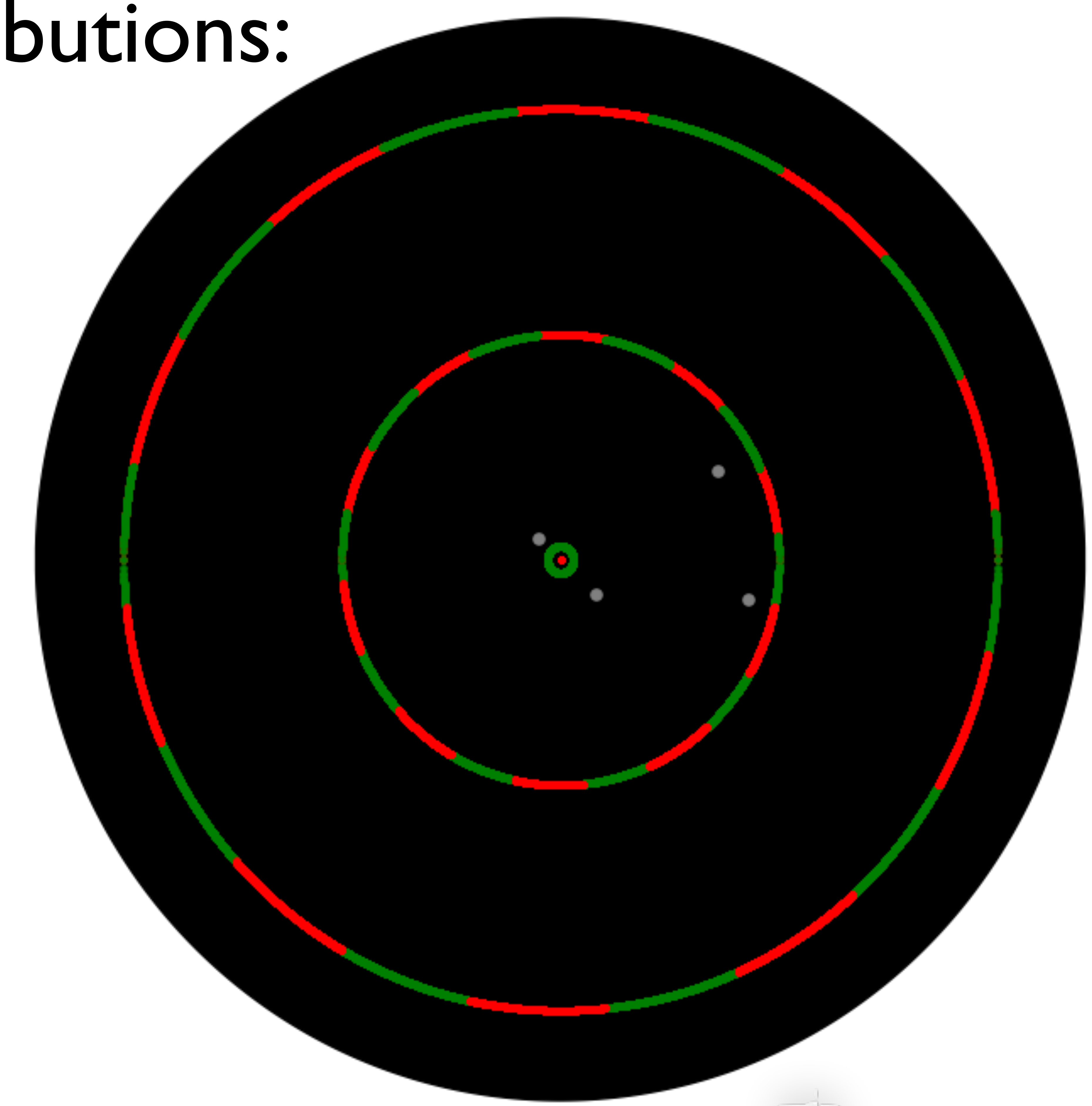
Classical *Continuous* distibutions:



Classical *Continuous* distibutions:

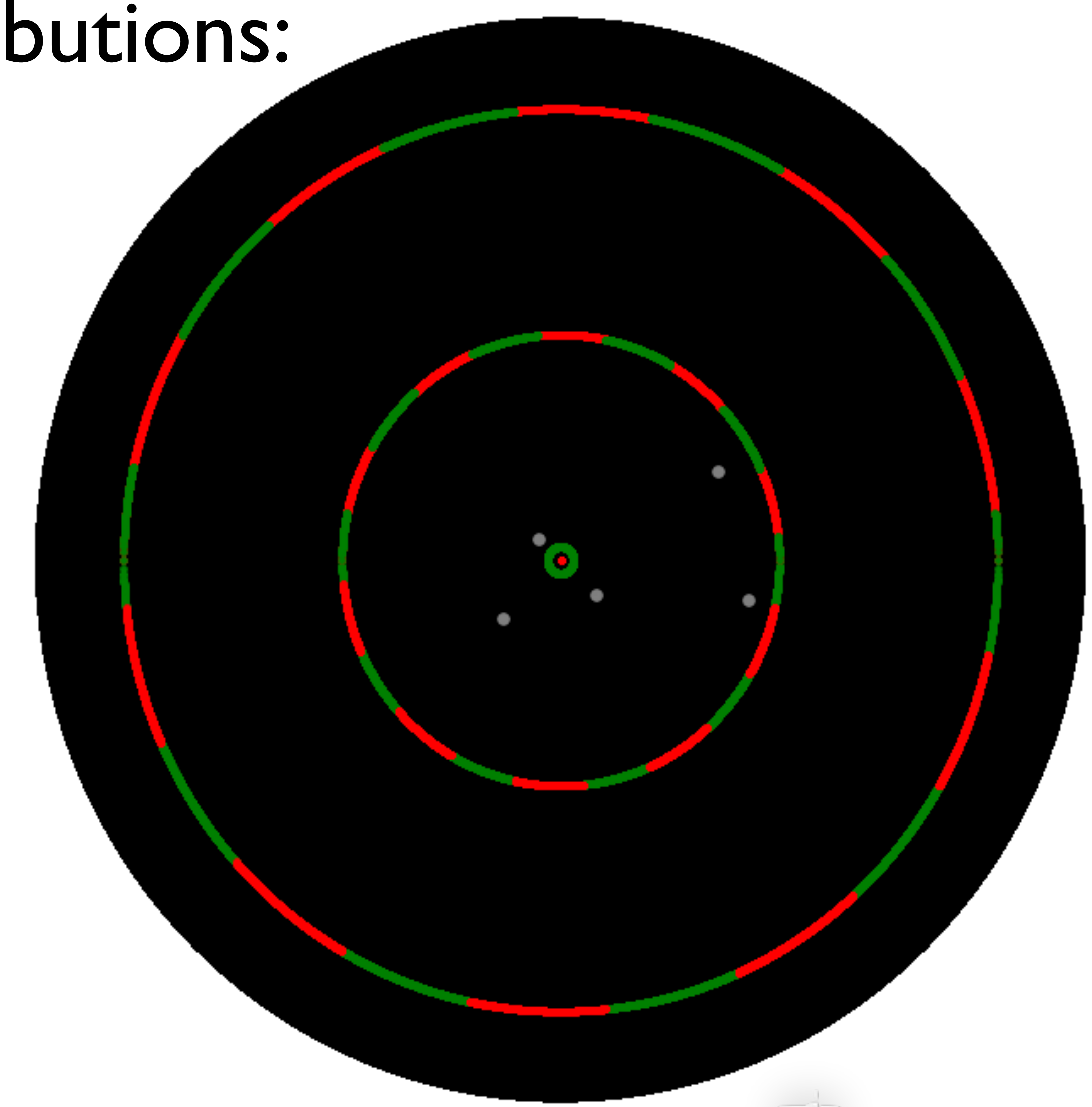


Classical *Continuous* distibutions:

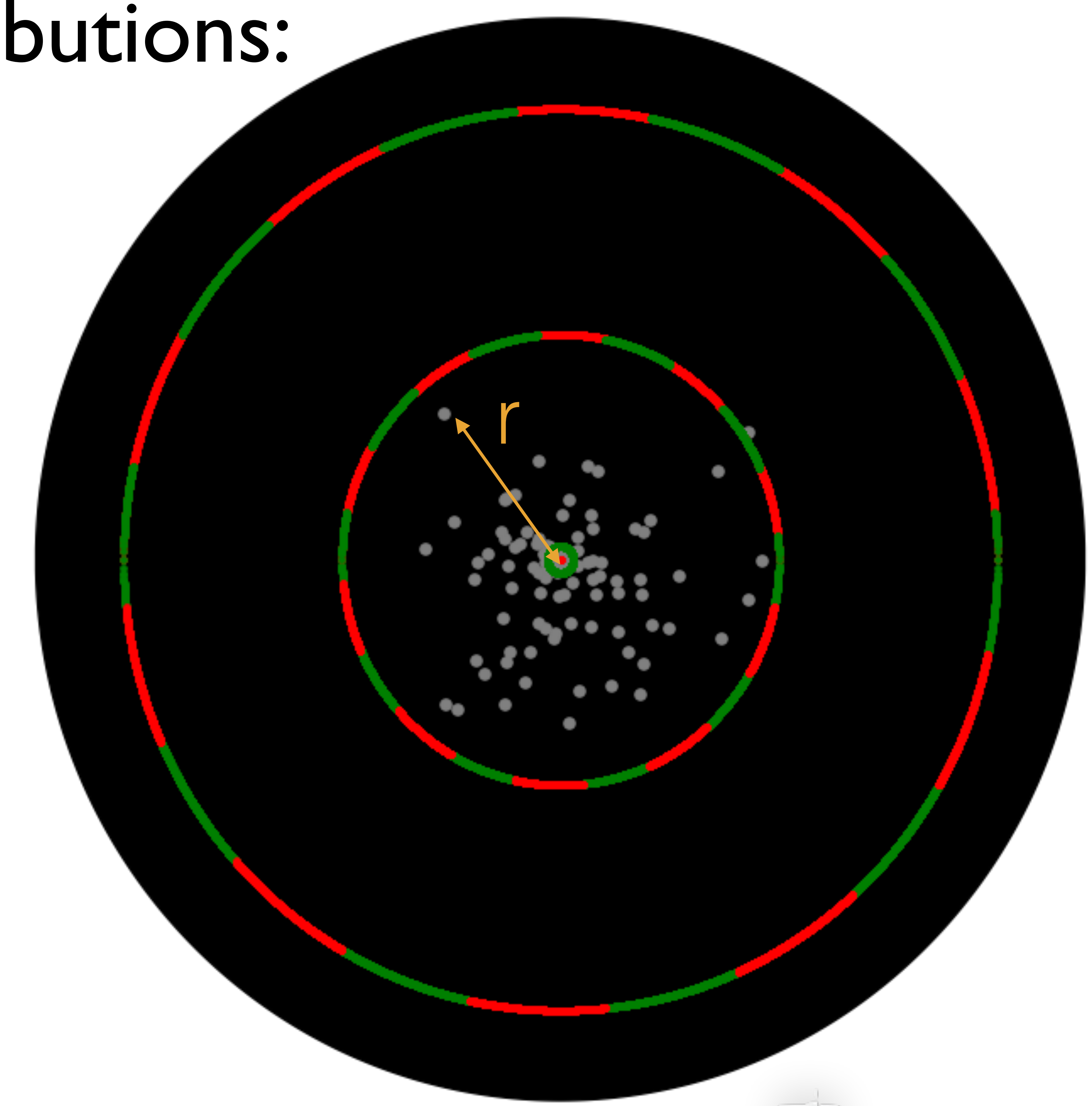


Classical *Continuous* distibutions

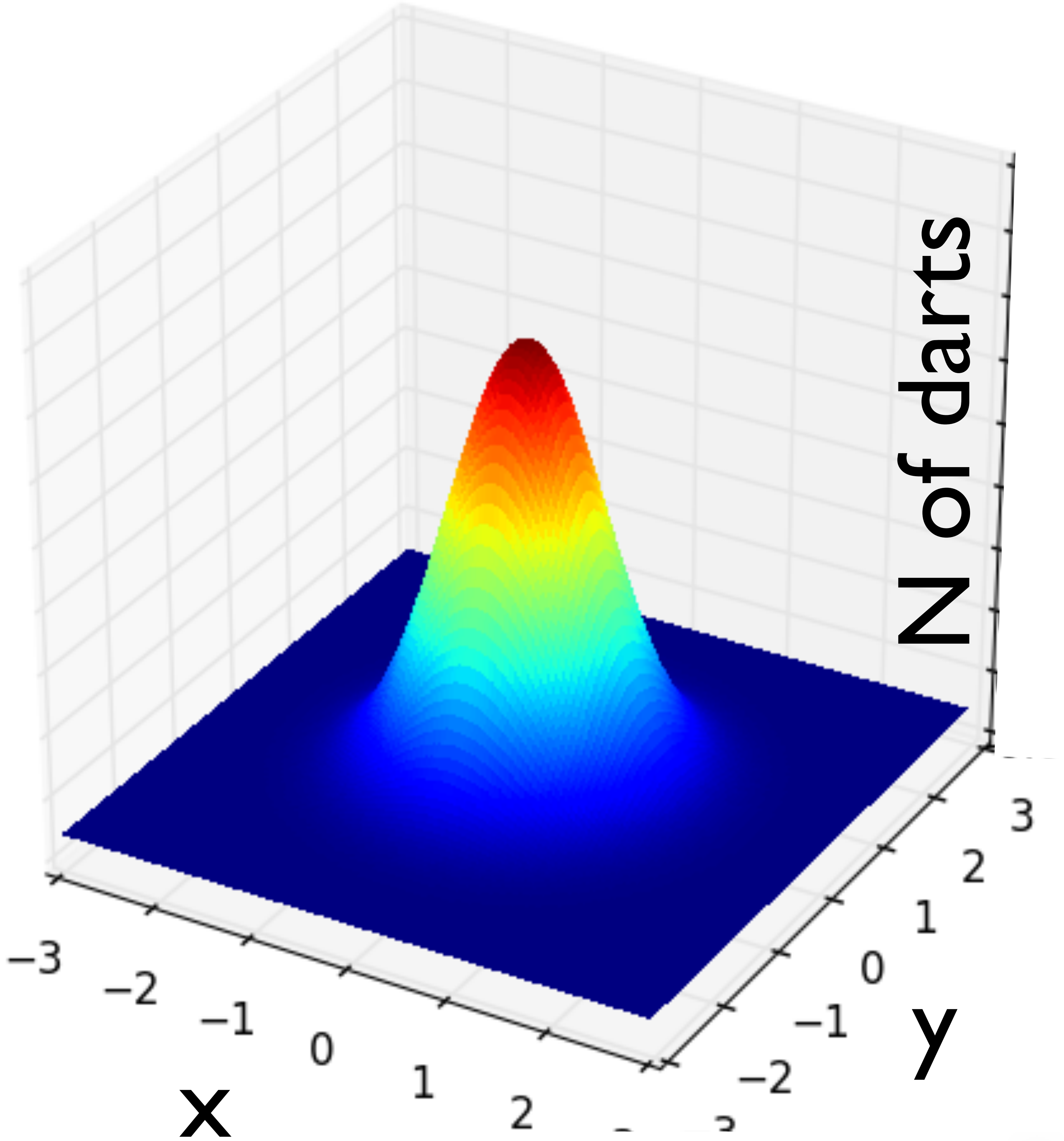
Classical *Continuous* distibutions:

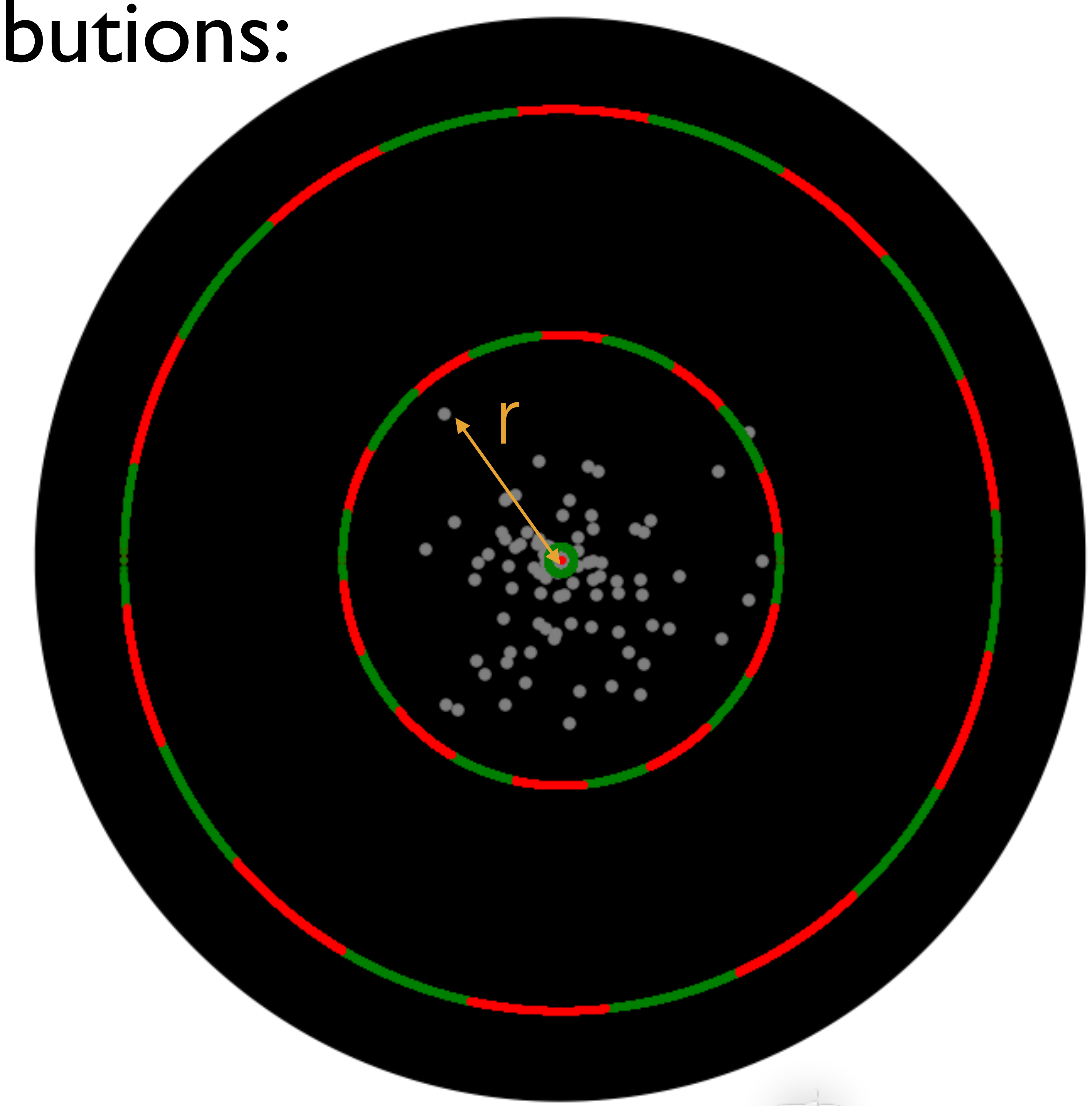


Classical *Continuous* distibutions:

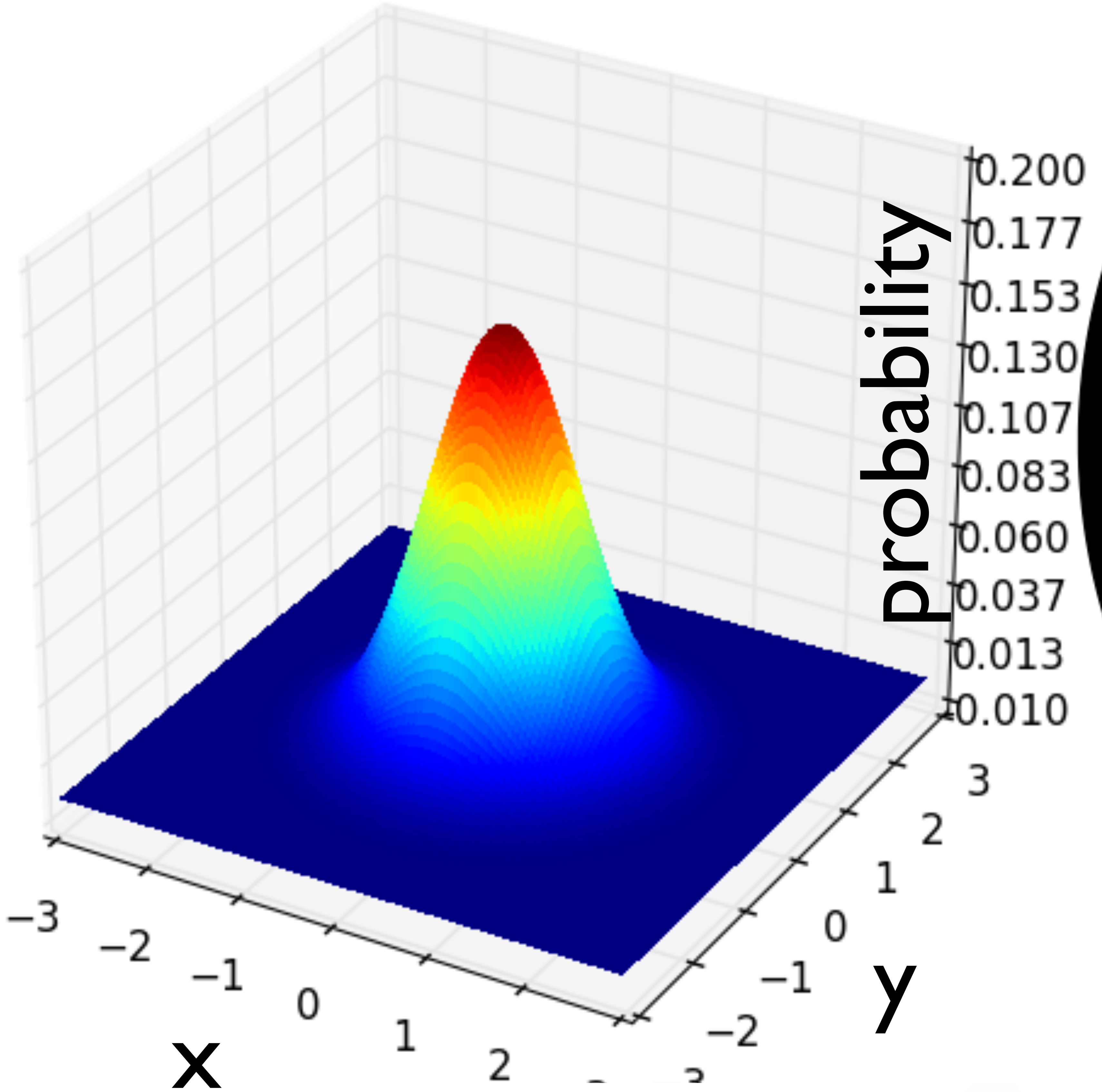


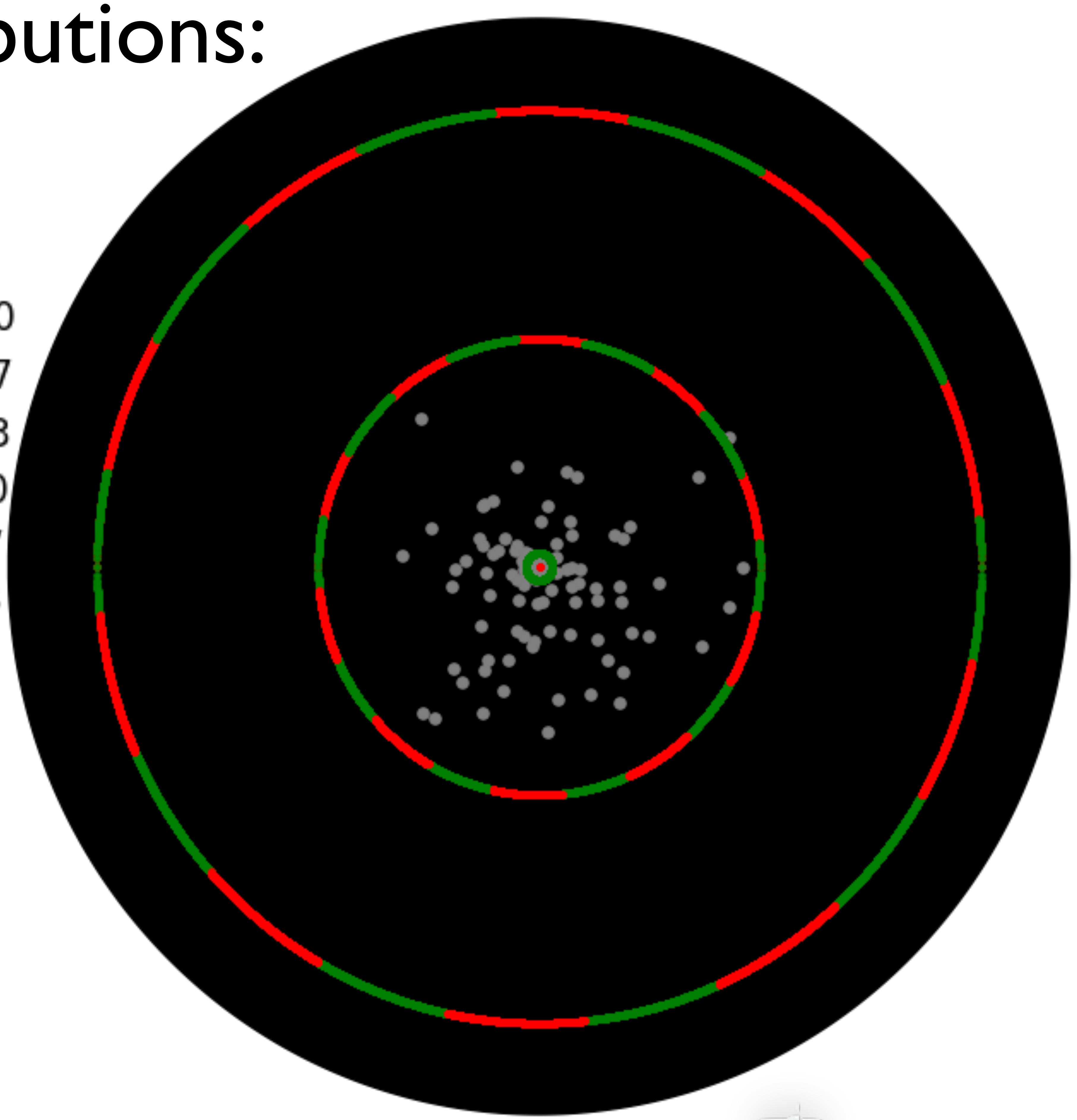
Classical *Continuous* distibutions:





Classical *Continuous* distibutions:





Normal Distribution

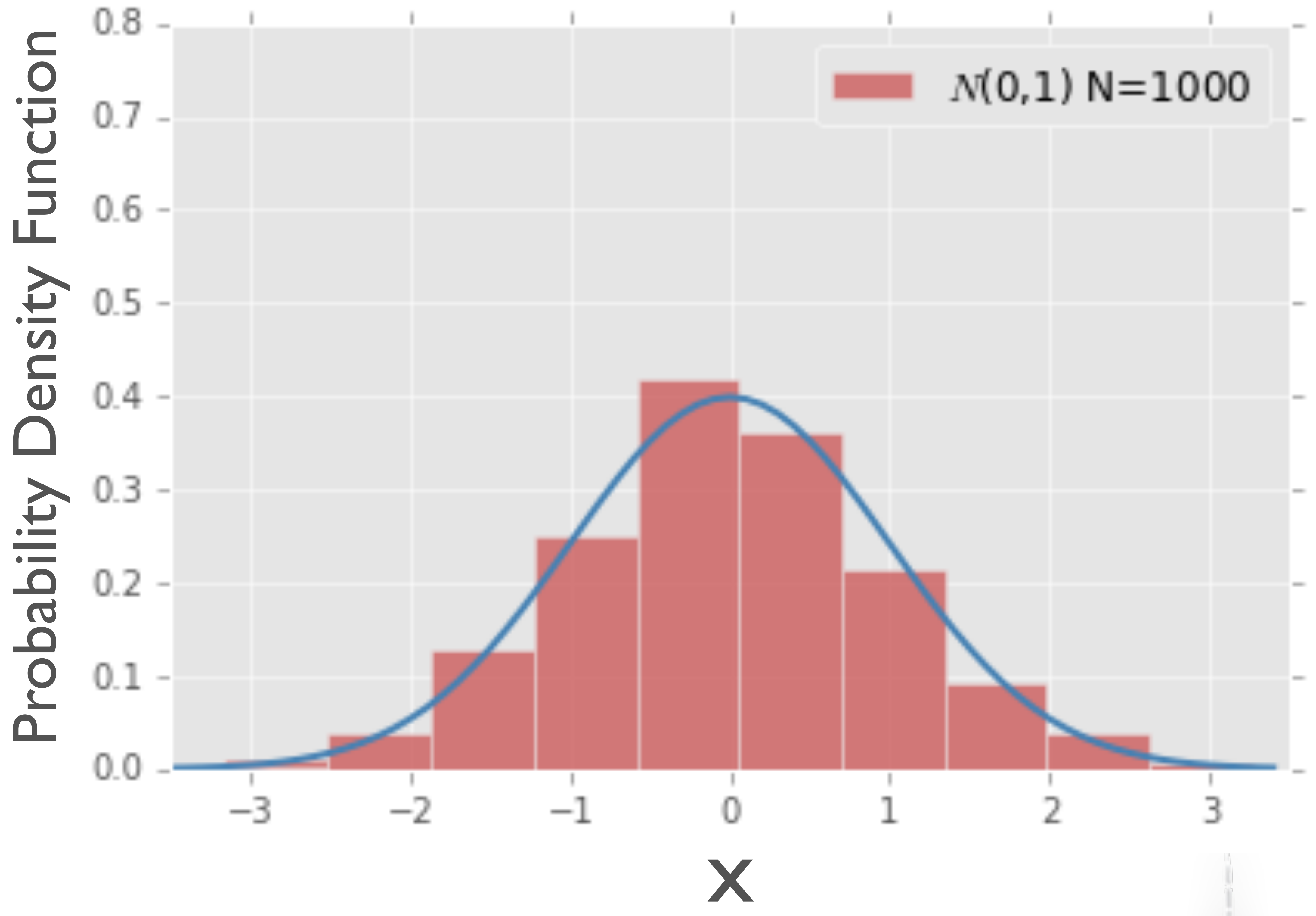
Classical *Continuous* distibutions:

Gaussian (or Normal) distribution

$$N(x \mid \mu, \sigma) \quad = \quad \frac{1.0}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

μ : mean

σ : standard deviation



Classical *Continuous* distibutions:

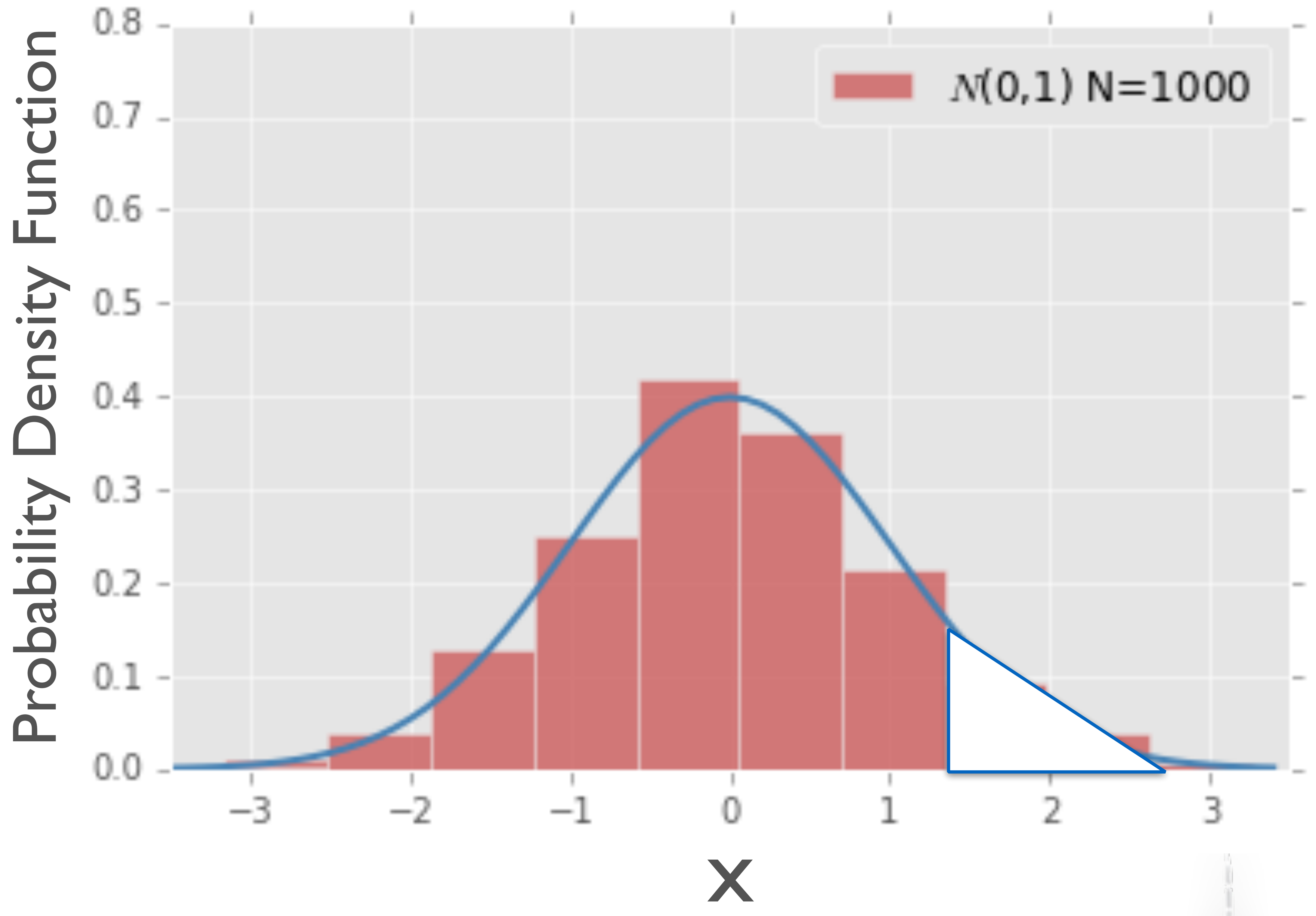
Gaussian (or Normal) distribution

$$N(x \mid \mu, \sigma) = \frac{1.0}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

P(X = any x) = 1 by definition

x ∈ [-∞, ∞]

P(X > x) = $\int_x^\infty N(x \mid \mu, \sigma)$



Classical *Continuous* distibutions:

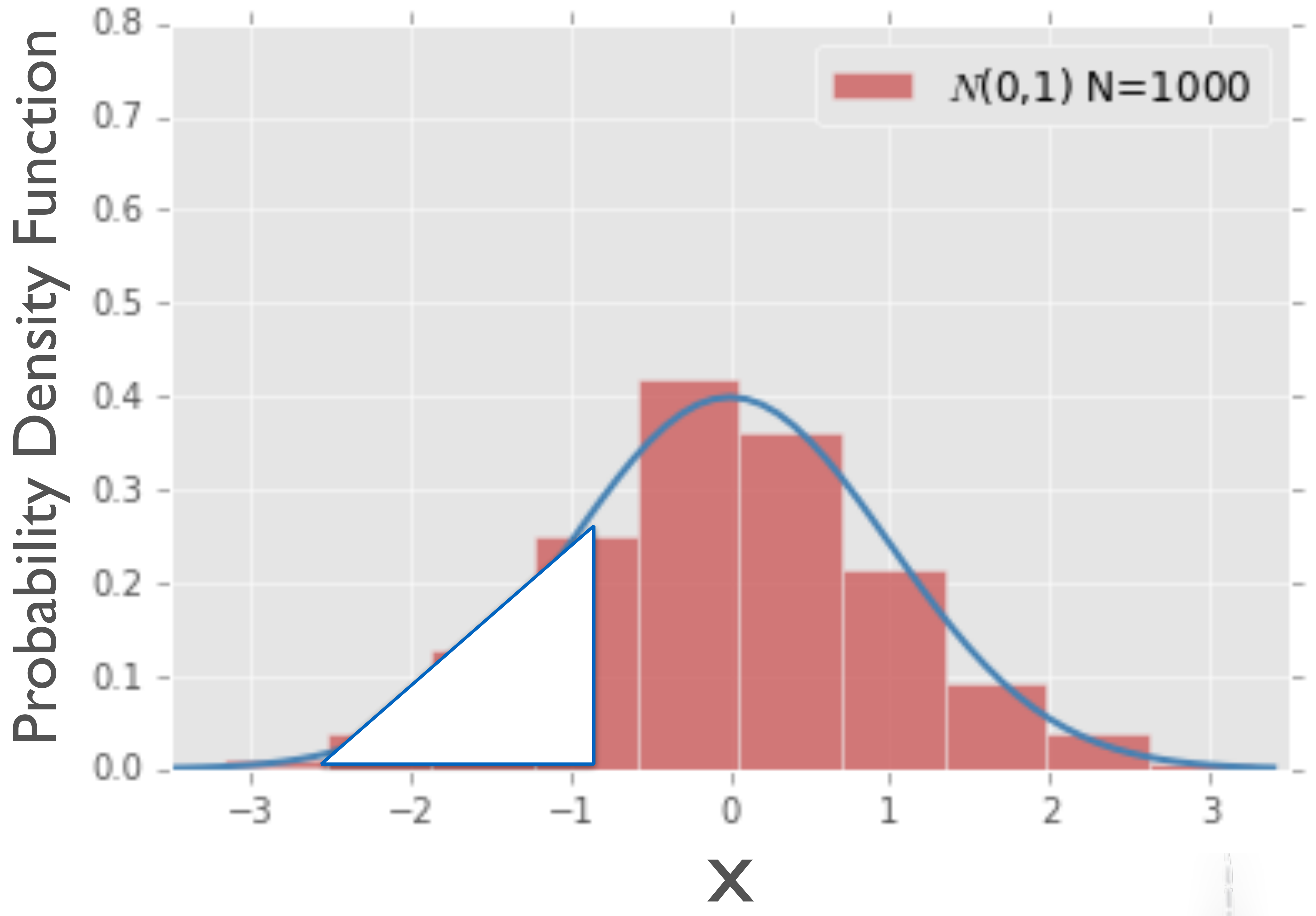
Gaussian (or Normal) distribution

$$N(x \mid \mu, \sigma) \quad = \quad \frac{1.0}{\sqrt{2\pi\sigma}}\, e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

P(X = any x) = 1 by definition

x ∈ [-∞, ∞]

P(X < some value)

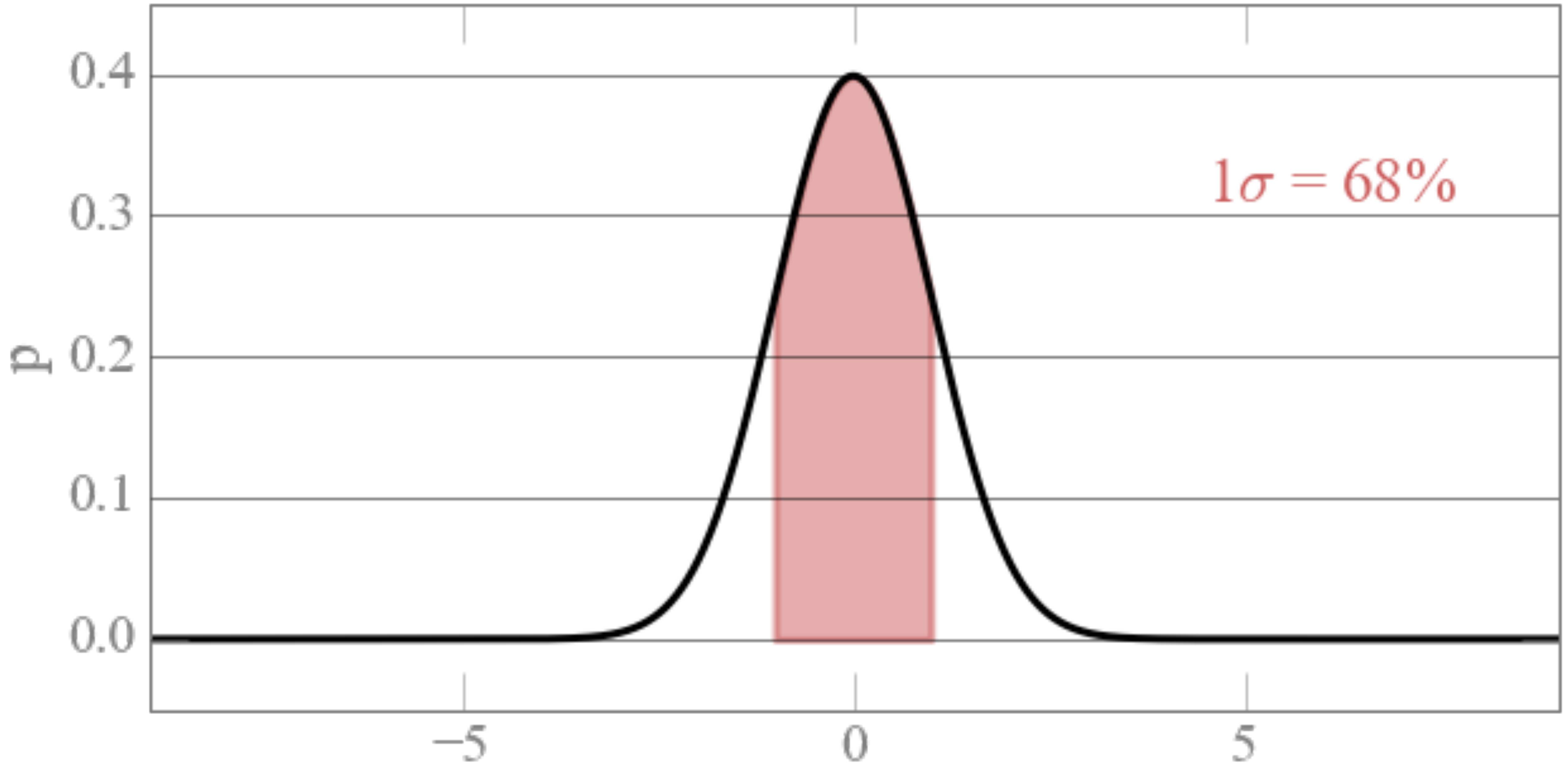


Normal Distribution

Classical *Continuous* distibutions:

Gaussian (or Normal) distribution

$$N(x \mid \mu, \sigma) = \frac{1.0}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

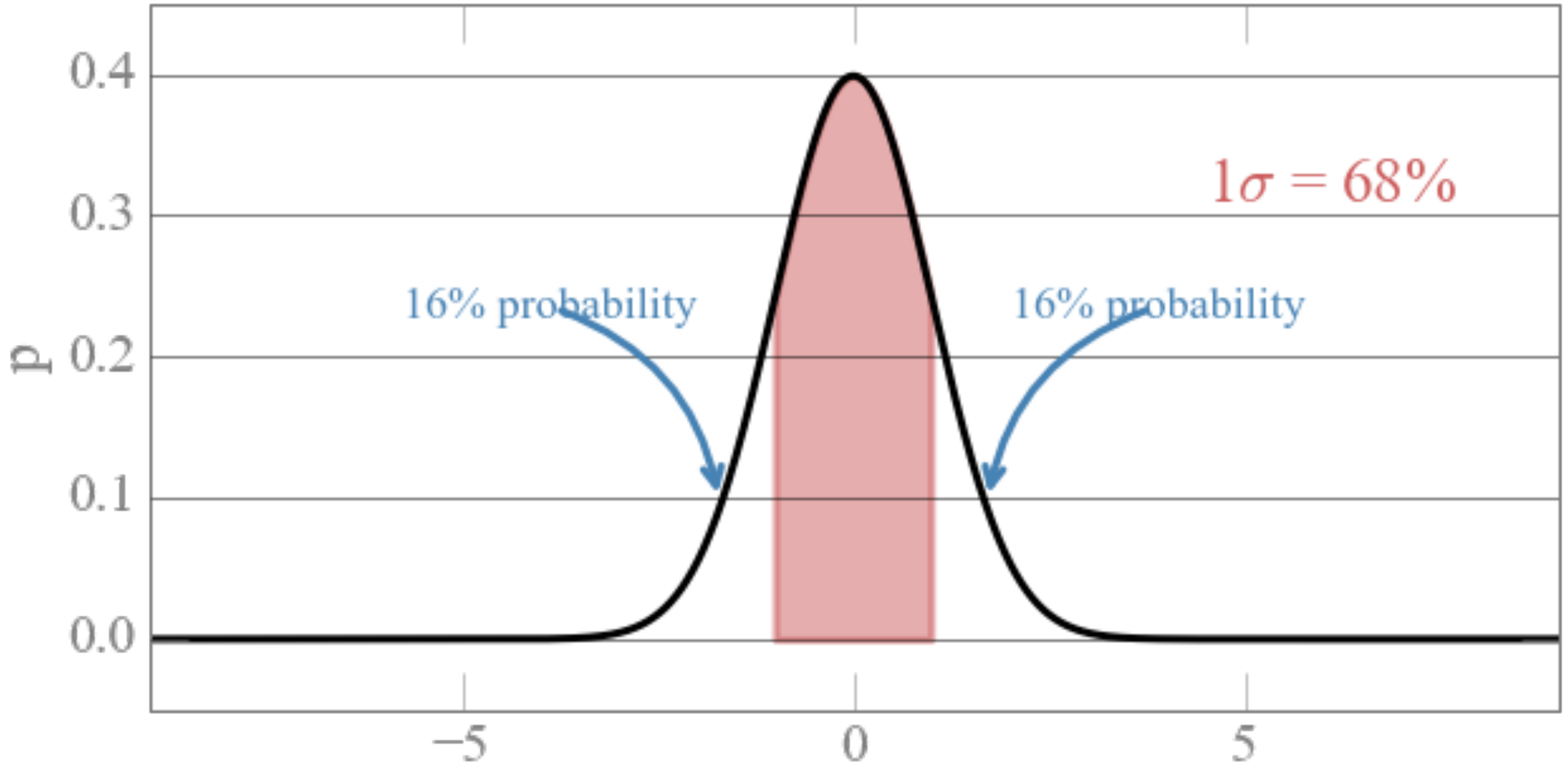


Normal Distribution

Classical *Continuous* distibutions:

Gaussian (or Normal) distribution

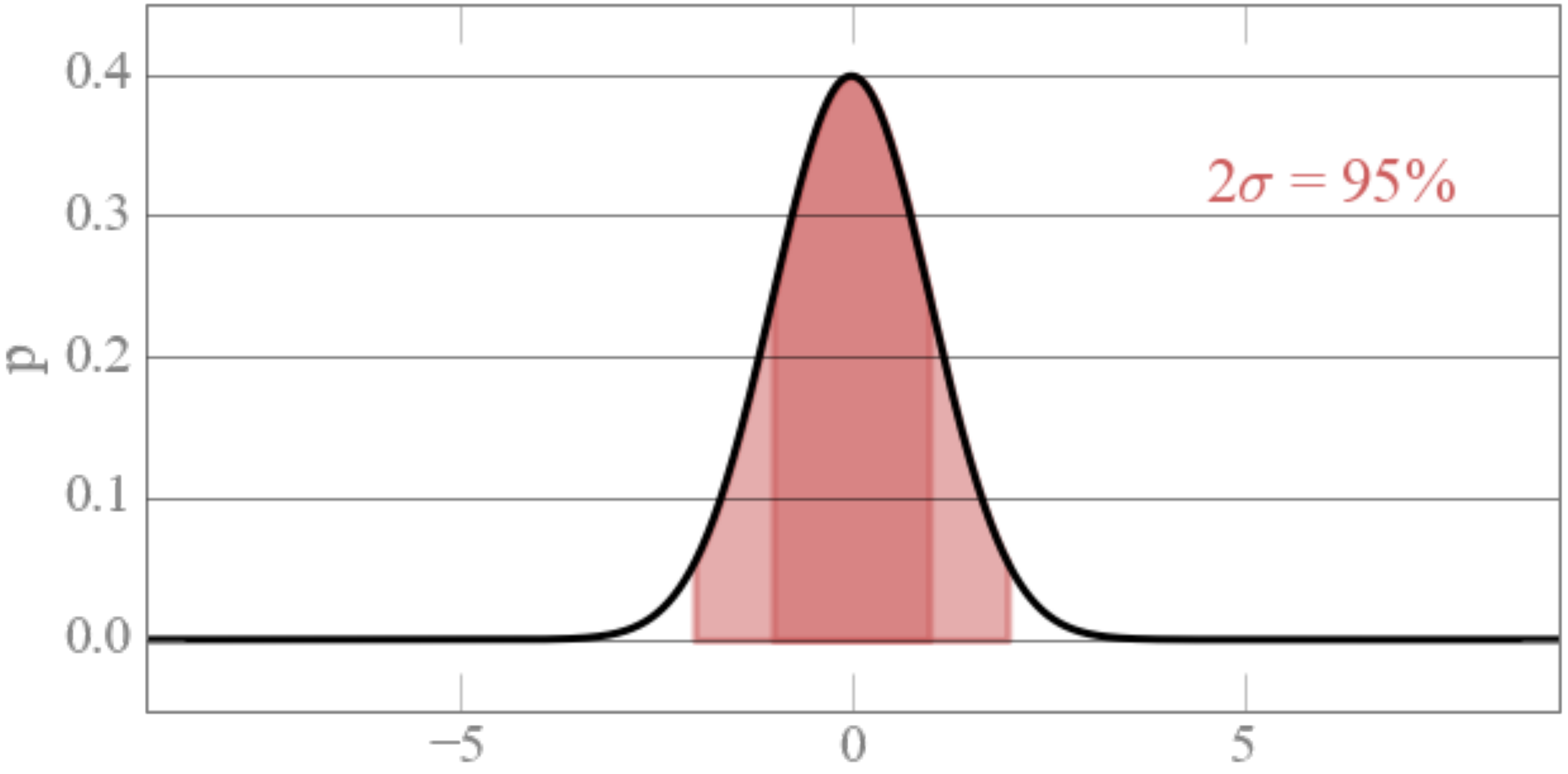
$$N(x \mid \mu, \sigma) \quad = \quad \frac{1.0}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{\sigma^2}}$$



Classical *Continuous* distibutions:

Gaussian (or Normal) distribution

$$N(x \mid \mu, \sigma) \quad = \quad \frac{1.0}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

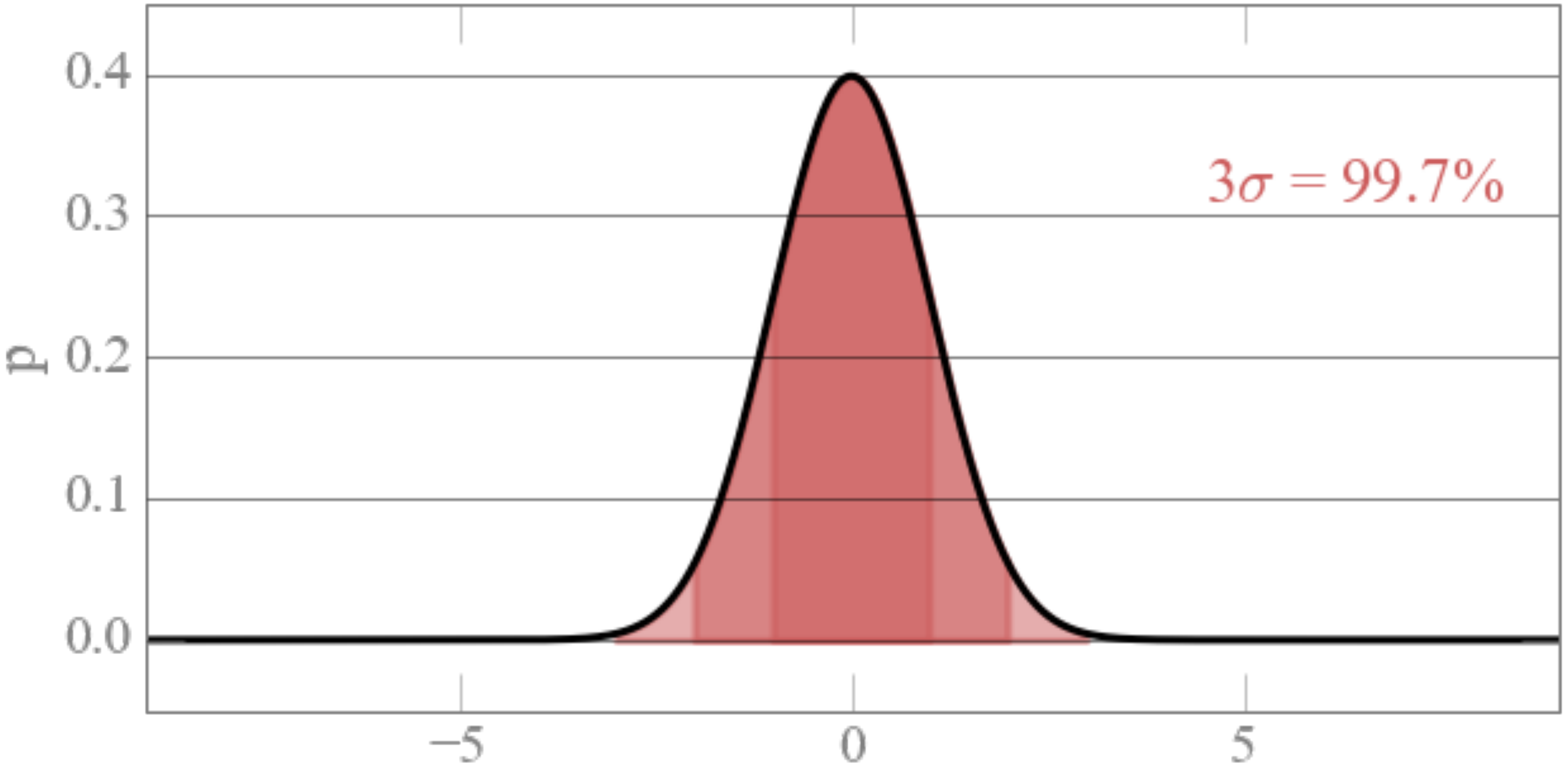


Normal Distribution

Classical *Continuous* distibutions:

Gaussian (or Normal) distribution

$$N(x \mid \mu, \sigma) \quad = \quad \frac{1.0}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{\sigma^2}}$$



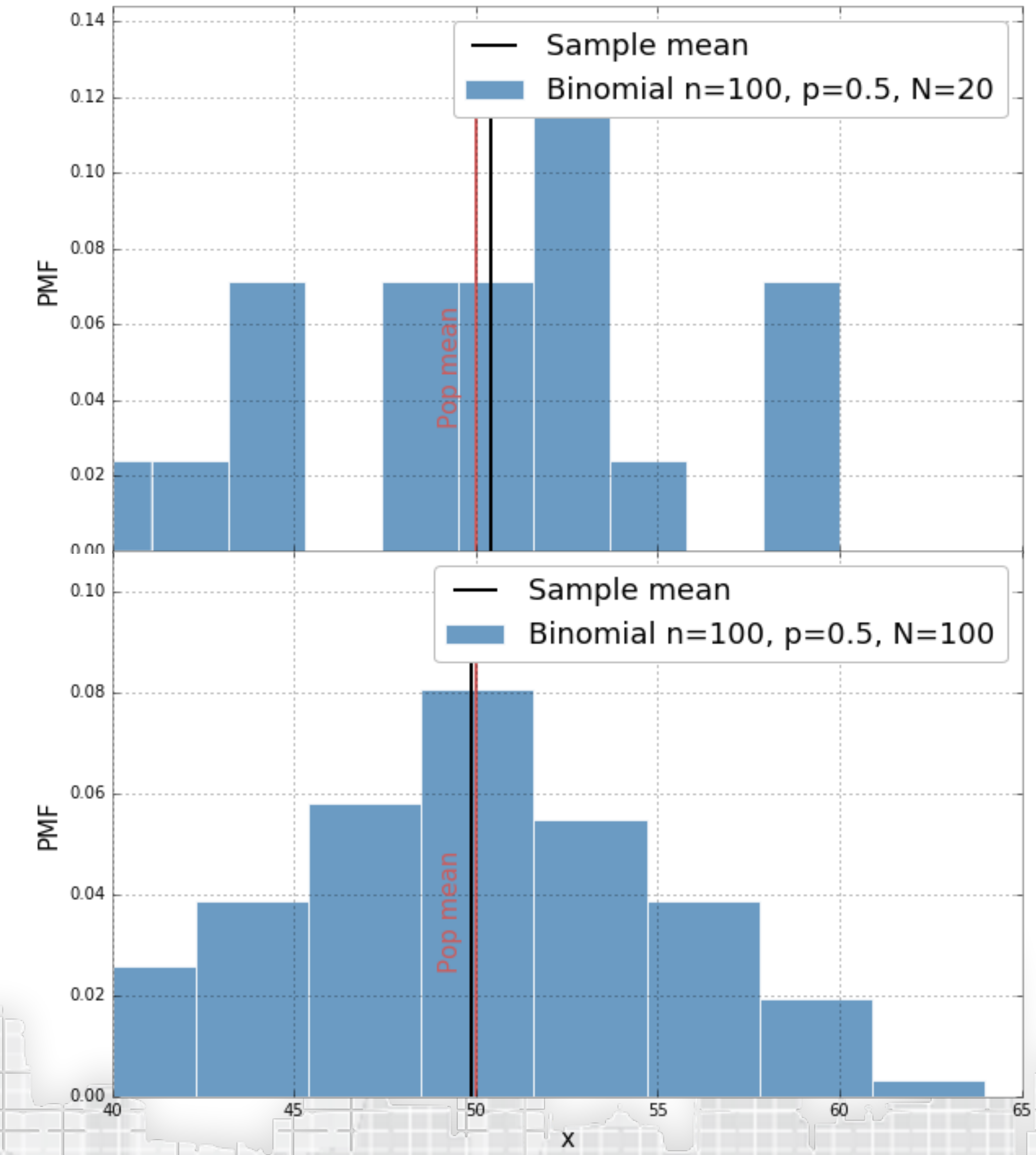
we will encounter this over and over again
when we talk about *p-values* and *hypothesis testing*

Central Limit Theorem

for ANY distribution, as the sample size increases, the *distribution of sample means* converges to the *population mean*

AND

the distribution of the means is Gaussian

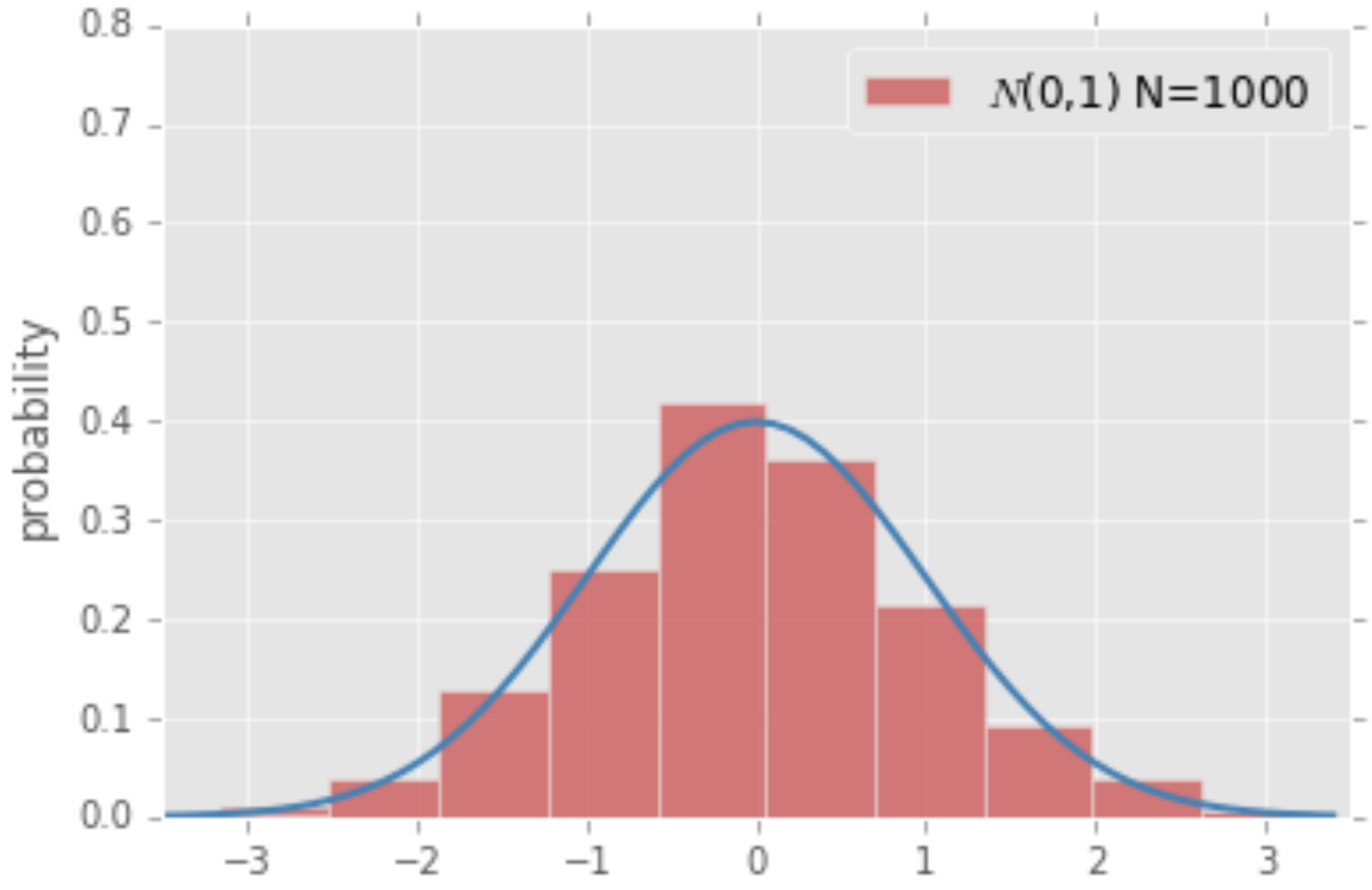


<u>*Central Limit Theorem:*</u>

for ANY distribution, as the sample size increases, the *distribution of sample means* converges to the *population mean*

AND

the distribution of the means is Gaussian



<u>*Pois*(λ) ⟶ *N*(λ, √λ) *for* λ ⟶ ∞:</u>

The Poisson *Pois*(λ) distribution converges to a *Gaussian* distribution *N*(λ, √λ) for large λ

https://github.com/fedhere/UInotebooks/blob/master/UCSL2016/UCSL2016%20continuous%20distributions%20demo.ipynb

Key Points

- Statistics allows us to make inference about the world, even if we cannot observe all of the world: inference on the population from a sample.

- Some statistical distributions describe any phenomena well: the Poisson distribution, which is a discrete distribution, and the Gaussian distribution, which is a continuous distribution.

- If the samples are large enough, the Poisson, Binomial, and a few other distribution start looking like a Gaussian. So in many cases we can make the assumption that the phoenomena we study are Gaussian in nature.

- If we have many samples (patients in many clinical trials for example) the means of the samples is itself going to be Gaussian-distributed (*Central limit theorem & Law of large numbers*).