DA-STATS

Topic 05: Univariate Sampling

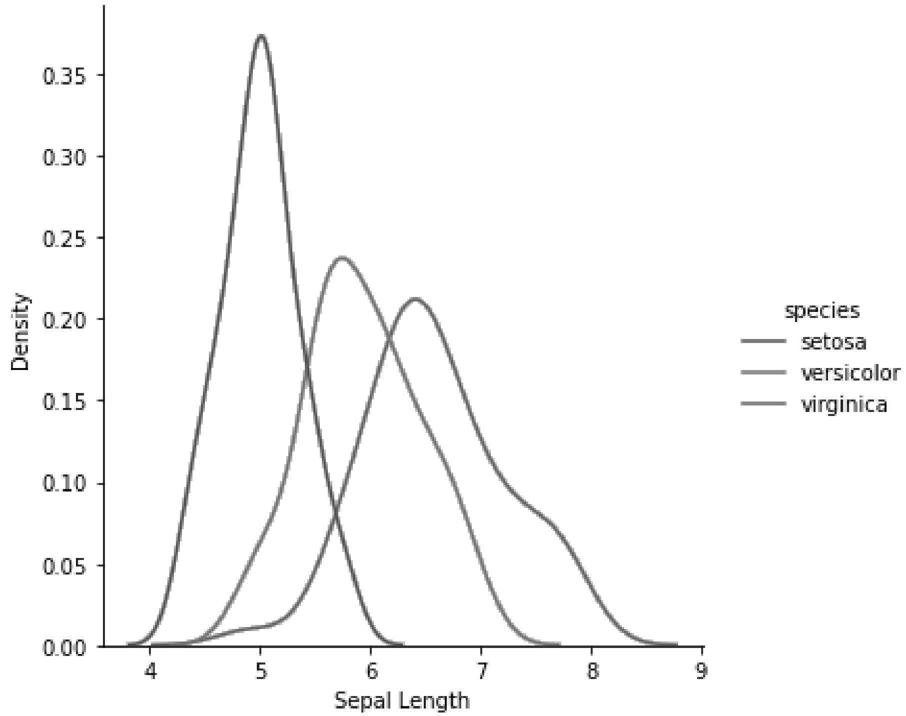
Shan E Ahmed Raza
Department of Computer Science
University of Warwick
shan.raza@warwick.ac.uk

Outline

- Distributions
- Probability Density Function (PDF)
- Cumulative Distribution Function (CDF)
- Sampling from Univariate Distribution
    - ➙Inverse CDF Sampling
    - ➙Reject Sampling
    - ➙Monte-Carlo Simulation

Sampling from univariate distributions

- How do we draw random numbers from a given distribution?
- What are distributions?



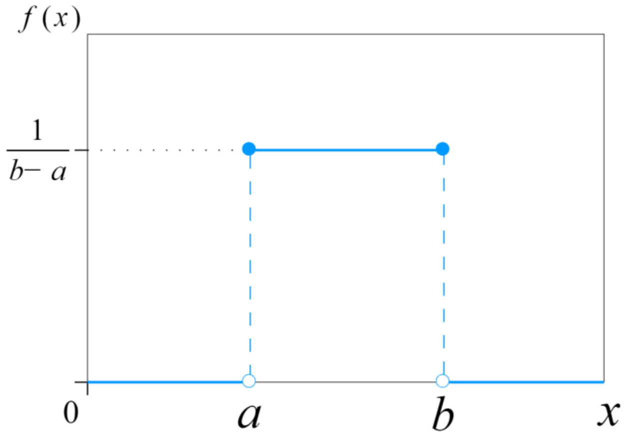
Distributions

- Mathematical function that gives the probabilities of occurrence of different possible outcomes for a random phenomenon
- Examples
  - Coin Toss
  - Dice
  - Someone getting sick from viral infection

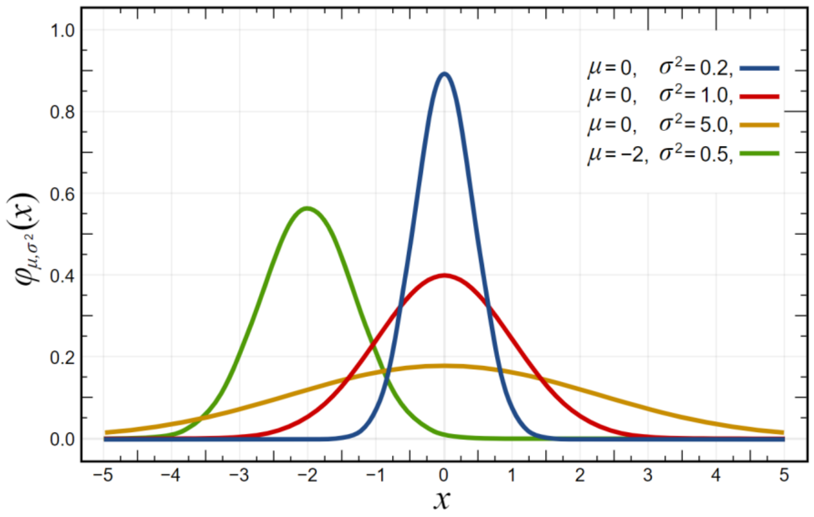
Distribution Examples

- Uniform Distribution

$$-p(x) = \frac{1}{b-a} \; for \; x \; \epsilon [a, b] \; else \; 0.0$$

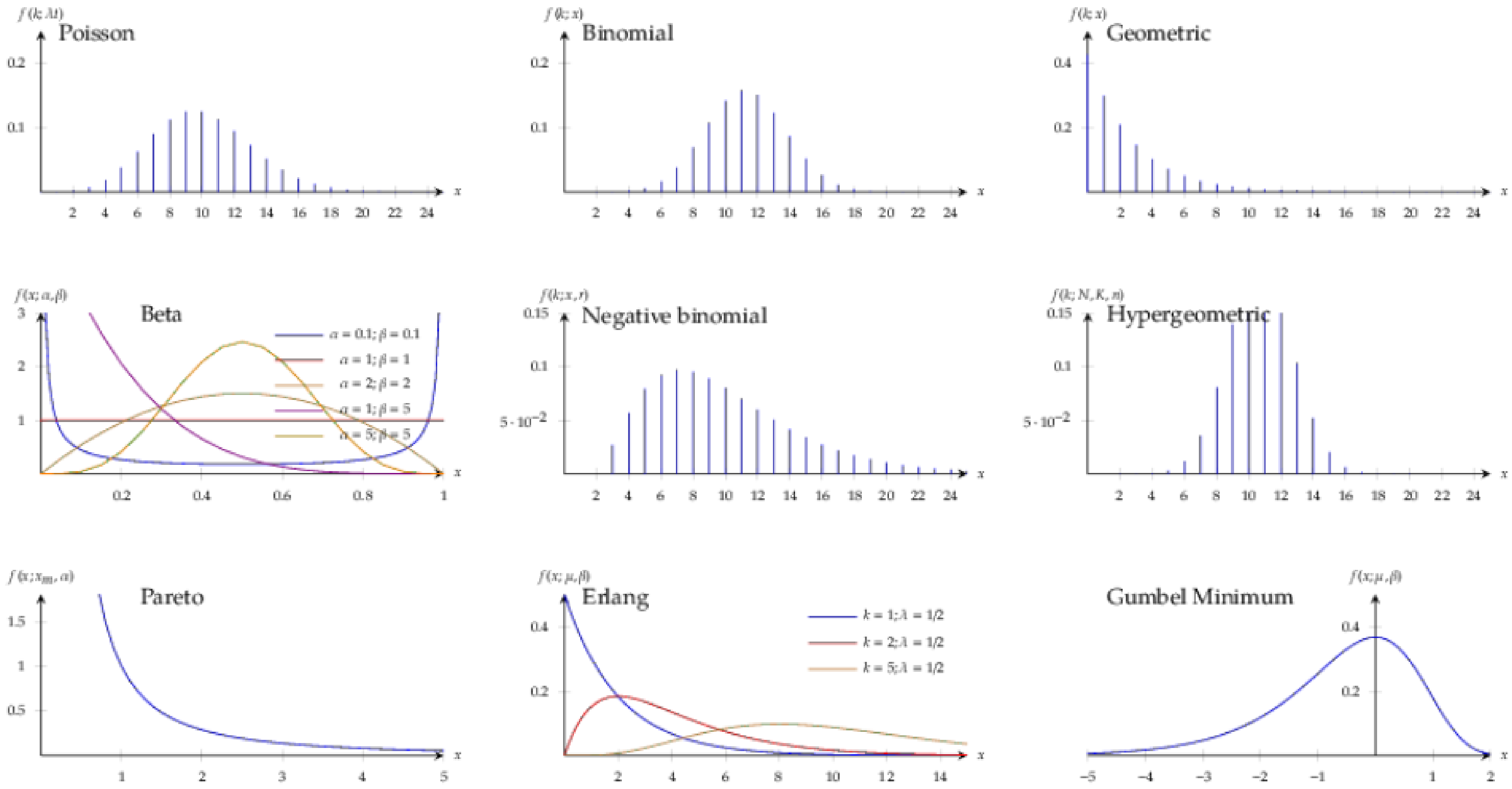


- Normal Distribution
- Gaussian Distribution
- Characterization of distributions

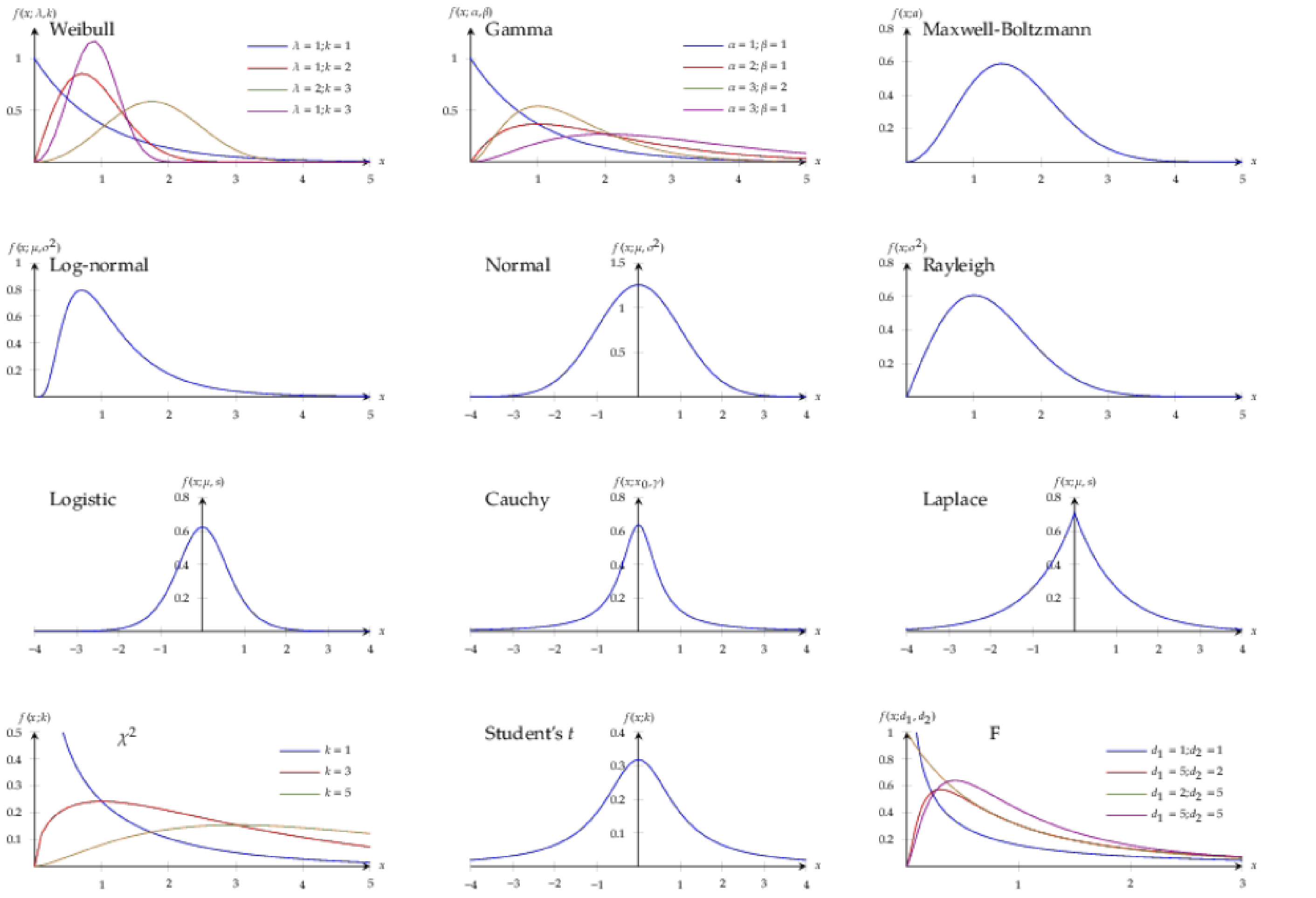


$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(x-\frac{\mu}{\sigma}\right)^2}$$

Distribution Examples



Distribution Examples



Central Limit Theorem

- Given a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement
- The **distribution of the sample means** will be approximately normally distributed.
- Holds true regardless of whether the source population is normal or skewed, provided the sample size is sufficiently large.
- If the population is normal, then the theorem holds true even for samples smaller than 30.

https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/BS704_Probability12.html

Sampling from a univariate distribution

- How to generate random numbers that follow a distribution?

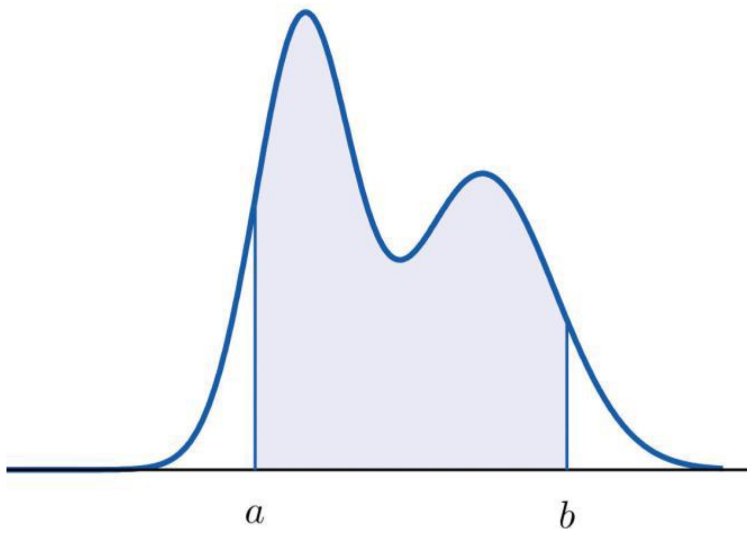
- Implementation
  - ➞ np.random.unform(0, 1, N)
  - ➞ np.random.gaussian

  - ➞ More in lab exercises

https://docs.scipy.org/doc/numpy 1.15.0/reference/routines.random.html

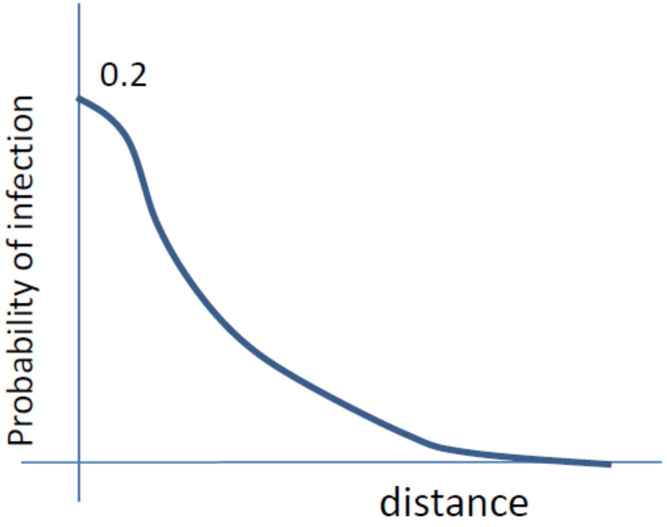
How are these Random Numbers Generated?

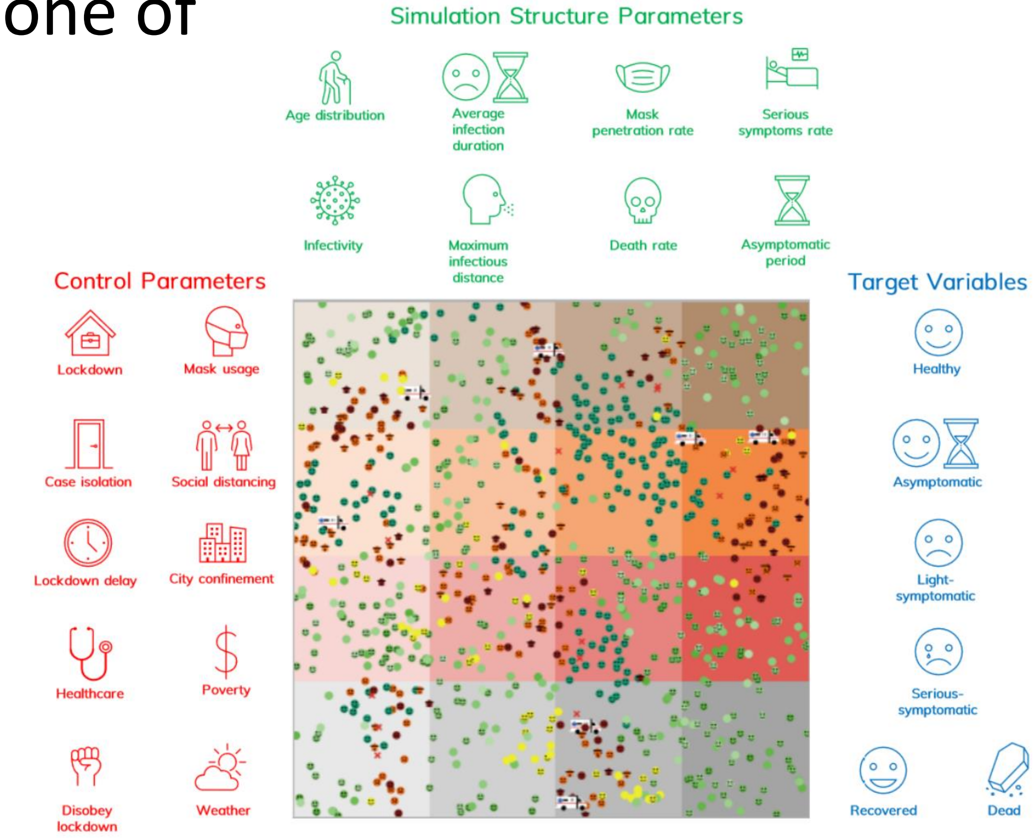
- How to sample from an arbitrary distribution?
- Let's say, you are required to generate random samples from a distribution like this on the right.
- Why would someone ask you to generate random samples that follow a distribution?



Application Example

- Let’s say, we want to simulate how a virus transmits between two people given that one of them is infected.





http://arxiv.org/abs/2011.11381

Inverse Transform Sampling

- Also known as: Inverse CDF Sampling

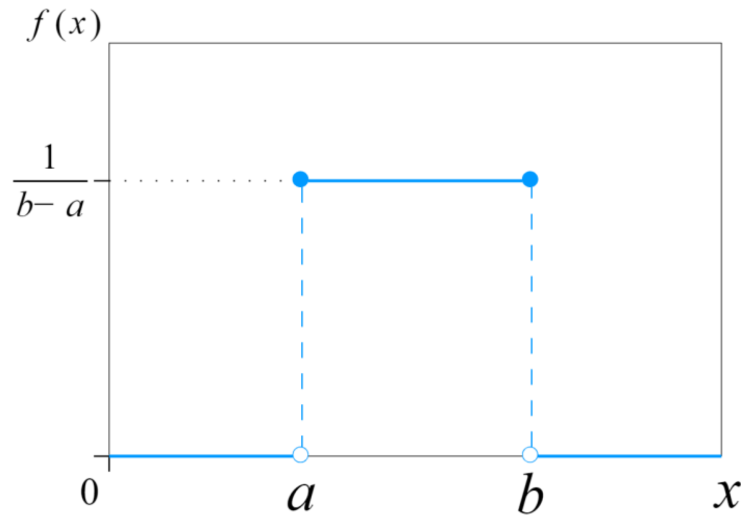
- What is a CDF?

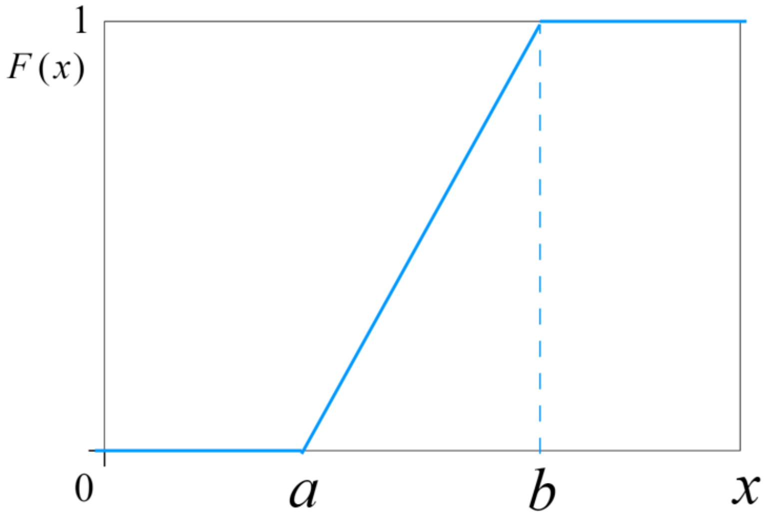
  - PDF (or PMF) Formally, probability density (mass) function
    - $p_X(x)$: probability of observing the random variable X with value x

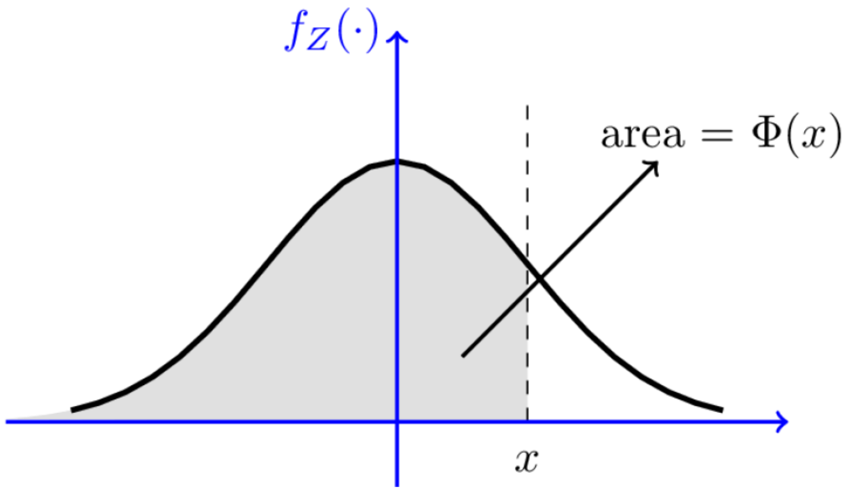
  - CDF (Cumulative Distribution Function)
    - The cumulative probability

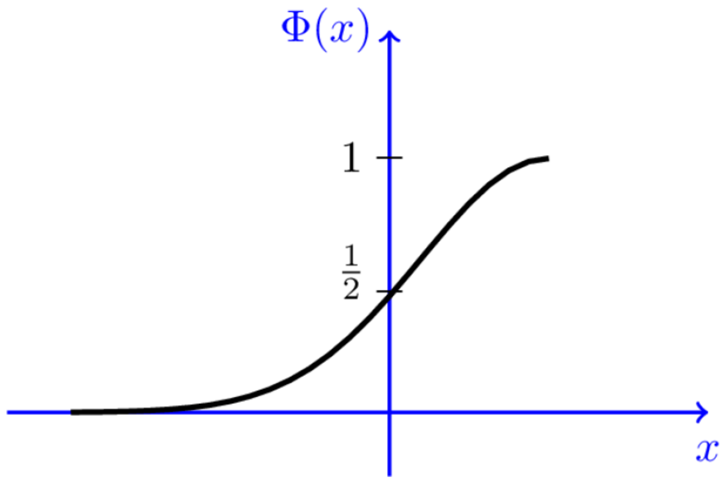
$$F_X(x) = p_X(X \leq x) = \int_{-\infty}^{x} p_X(x)$$

CDF Example





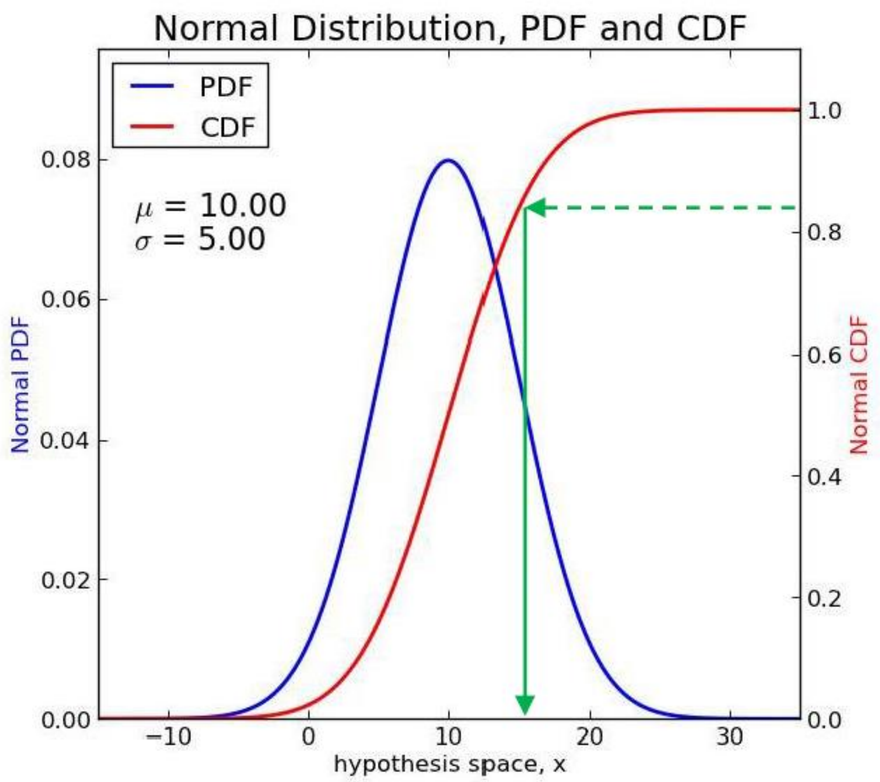




https://demonstrations.wolfram.com/ConnectingTheCDFAndThePDF/

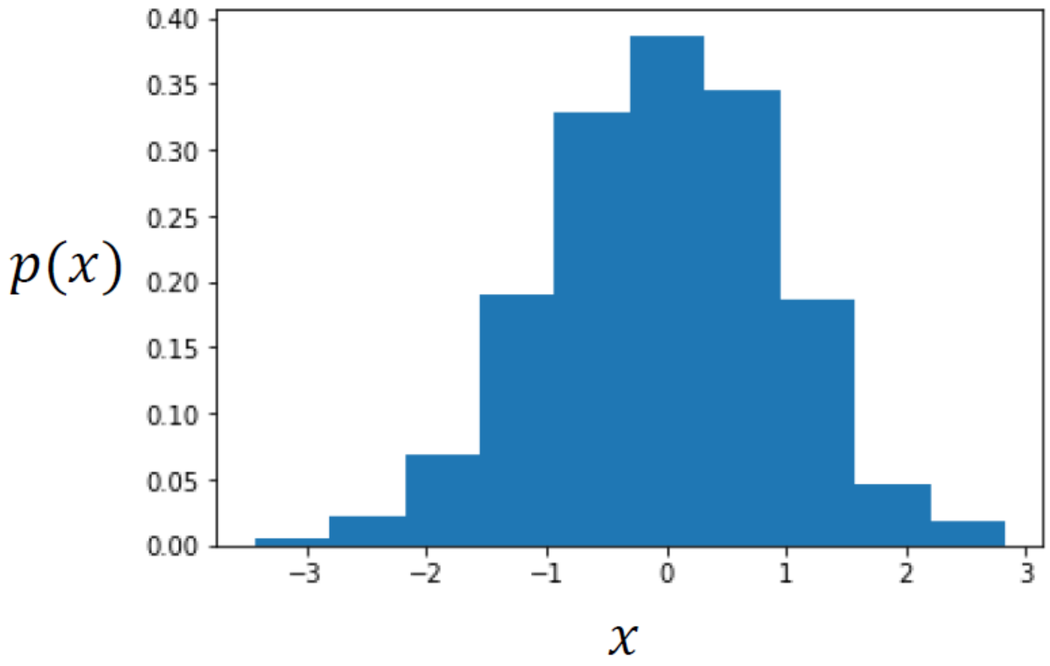
Inverse CDF Sampling

- Generate a random number u from the standard uniform distribution in the interval $[0, 1]$, e.g., from $U \sim Unif\,[0,1]$.
- Find the inverse of the desired CDF, e.g., $F_X^{-1}(x)$.
- Compute $X = F_X^{-1}(u)$. The computed random variable X has distribution $F_X(x)$.

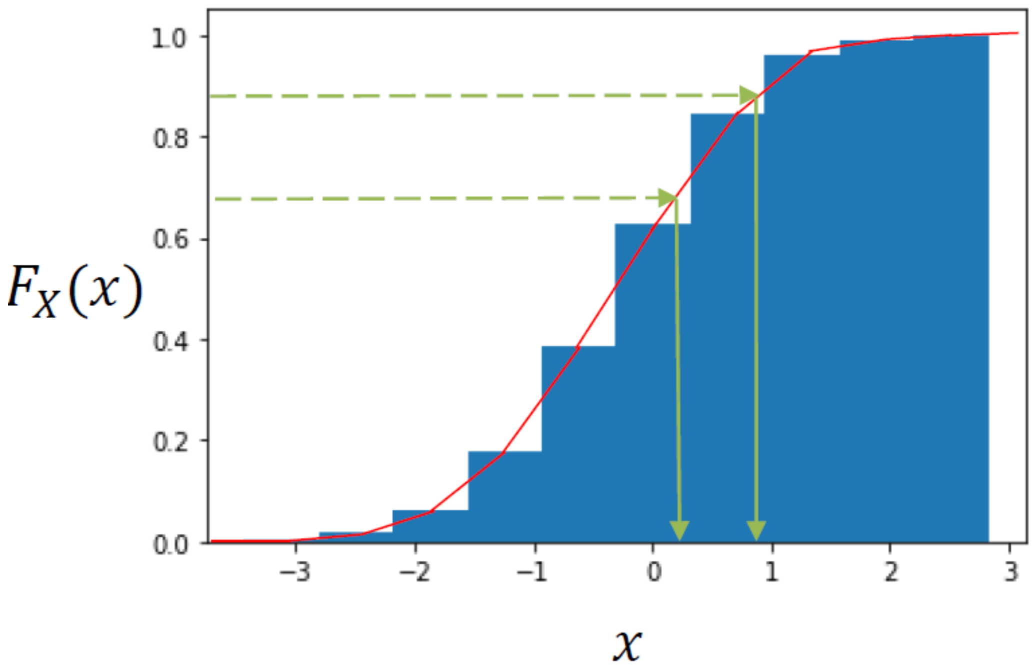


https://en.wikipedia.org/wiki/Inverse_transform_sampling

Exercise



0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$p(x)$
−3
−2
−1
0
1
2
3
x



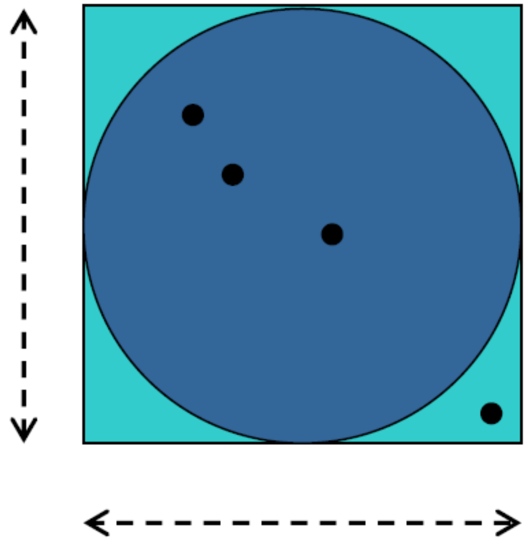
1.0
0.8
0.6
0.4
0.2
0.0
$F_X(x)$
−3
−2
−1
0
1
2
3
x

Sampling: Reject Sampling

- Keep samples that follow the distribution and reject others

Finding the value of π

- $r = 1.0 \Rightarrow A_{square} = (2r)^2 = 4.0,\ A_{circle} = \pi r^2 = \pi$
- $\frac{A_{circle}}{A_{square}} = \frac{\pi}{4.0} \Rightarrow \ \pi = 4\left(\frac{A_{circle}}{A_{square}}\right)$



- If we throw darts at random at an object, the percentage of darts that land inside it is proportional to its area relative.
- Throw darts at random within the square and calculate the percentage of darts that land inside the circle. This gives an approximate value of the ratio

$$\frac{A_{circle}}{A_{square}} = \frac{N_{circle}}{N_{square}}$$

- Use the above equation to calculate π

Throwing darts

- Generate coordinates using “random” module
  - ➞ x= random.random
  - ➞ y= random.random
- How would you check if (x, y) is inside the circle?
  - ➞ If $x^2 + y^2 < r^2$
- Throw a large number of darts and test what proportion lands inside. You can use it to estimate the value of π

- Further details in the lab session

Monte Carlo Simulation

- What you did is a monte carlo simulation of estimating the value of π

- It can be used for arbitrary simulation of systems

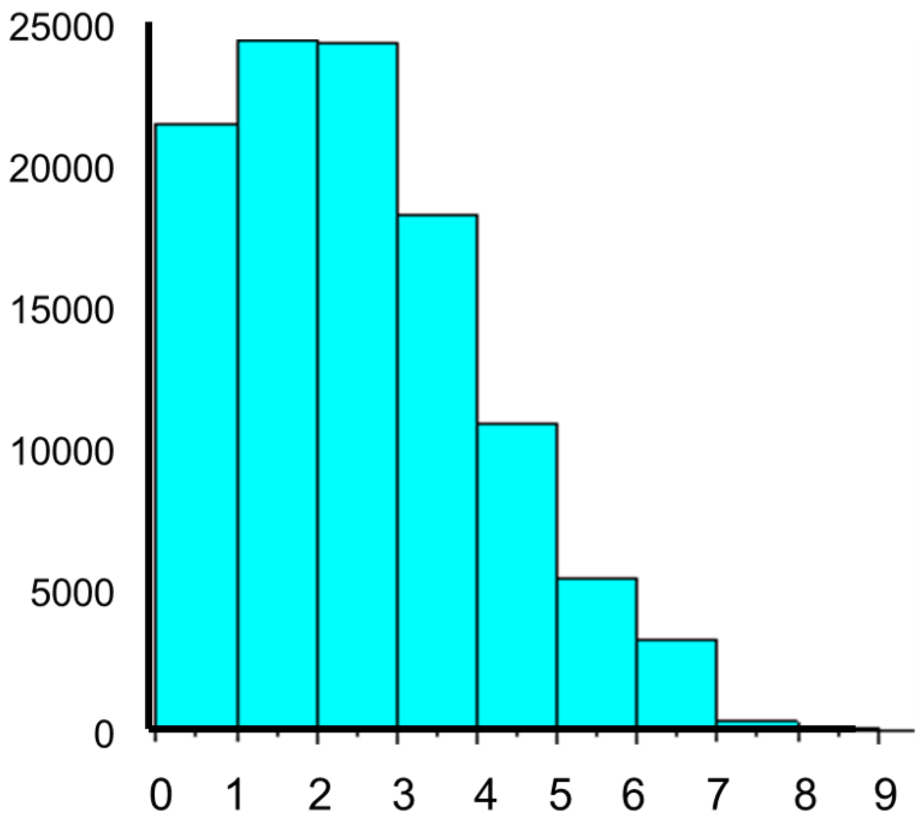
- Can be used for integrating arbitrary functions

- Finding risks, robustness, etc.

https://en.wikipedia.org/wiki/Monte_Carlo_method

Central Limit Theorem

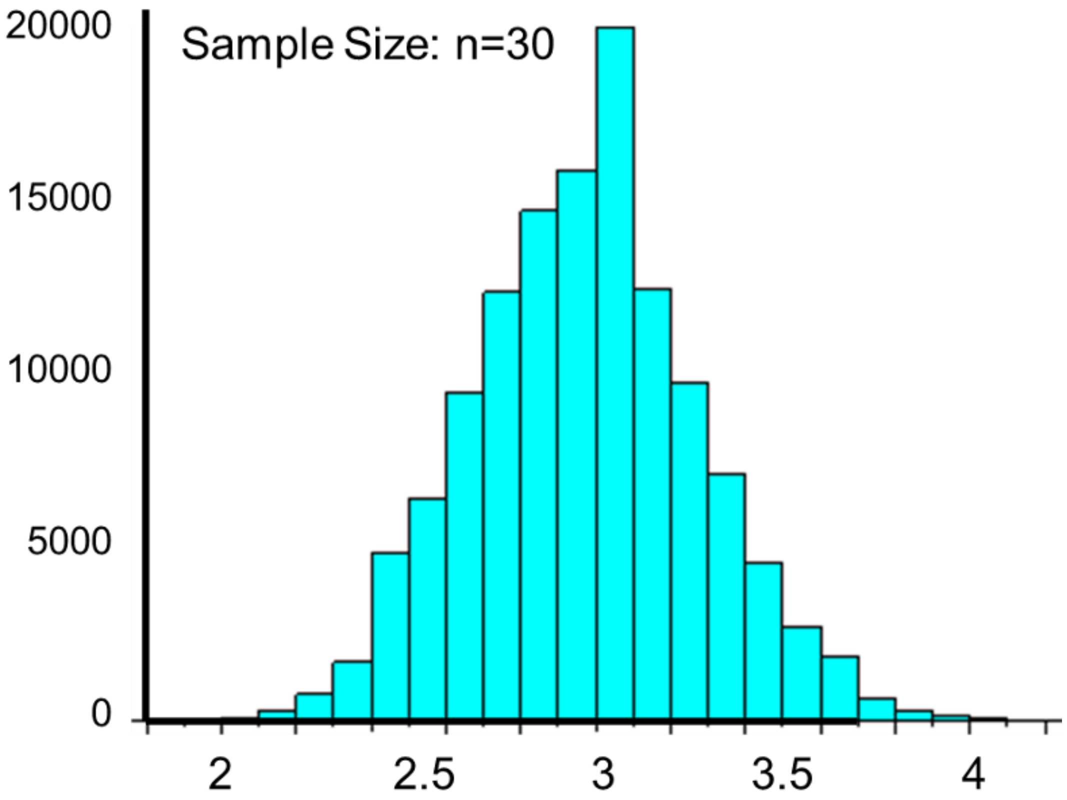
- Consider a Poisson distribution with $\mu = 3$ and $\sigma = 1.73$



https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/BS704_Probability12.html

Central Limit Theorem

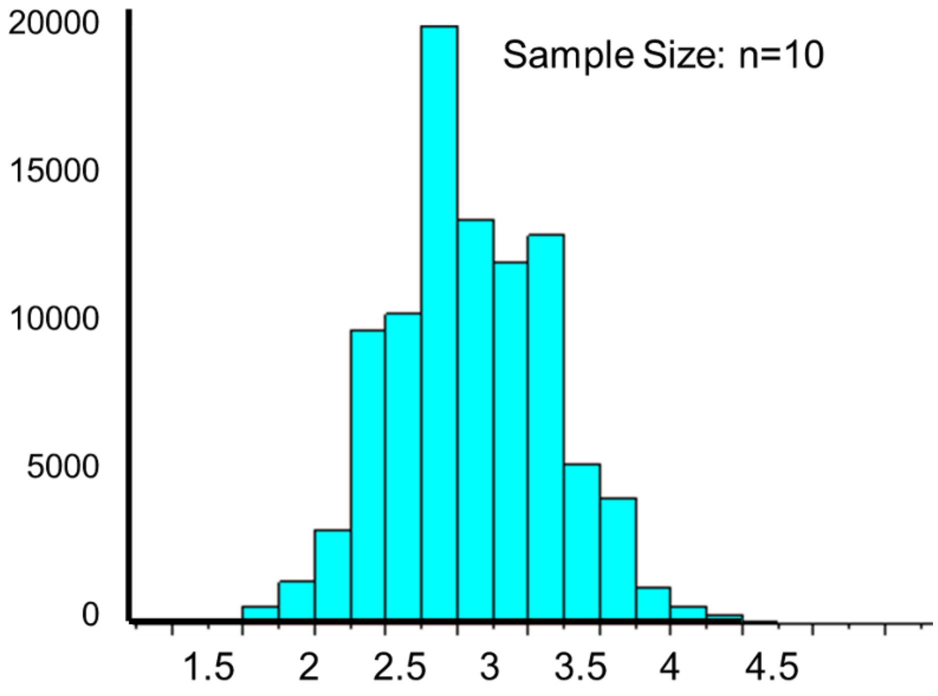
- This population is not normally distributed, but the Central Limit Theorem will apply if $n > 30$.
- If we take samples of size $n = 30$, we obtain samples distributed as shown in the graph with a $\mu = 3$ and $\sigma = 0.32$



https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/BS704_Probability12.html

Central Limit Theorem

- In contrast, with small samples of $n = 10$, we obtain samples distributed as shown in the graph.



https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/BS704_Probability12.html