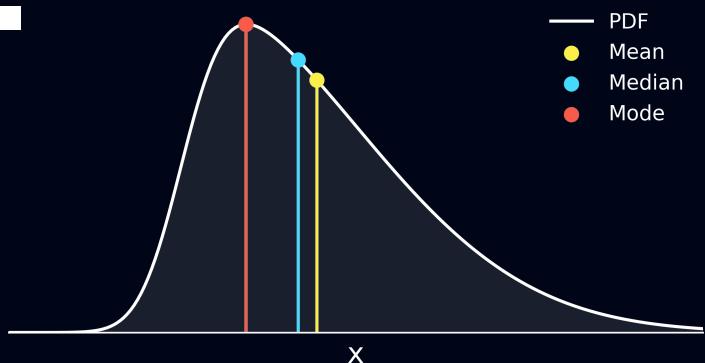


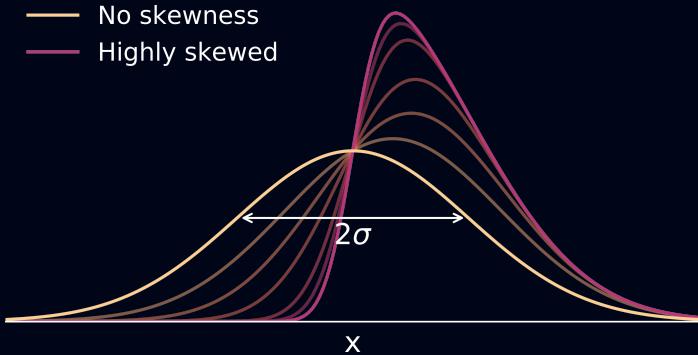
Summary Statistics

Central Values

1. Mean: Sum of all points over num points
2. Mode: The most common value
3. Median: Middle of the sorted list of



— No skewness
— Highly skewed



Width and Skewness

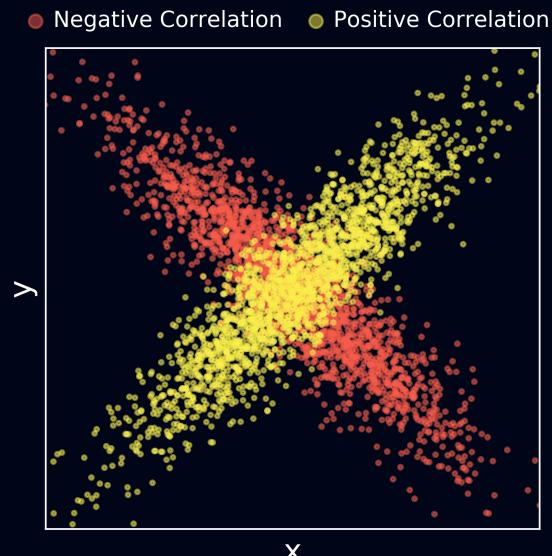
Standard Deviation: average distance from the mean.

Skewness: asymmetry of the distribution.

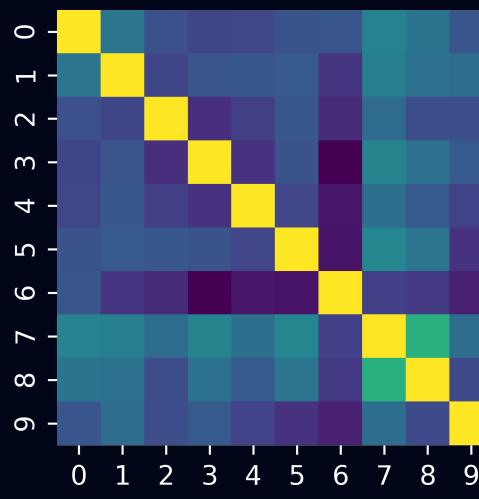
Correlation

Correlation

Positive correlation means if one variable increases the other usually does as well.



A correlation matrix allows a quick summary of all variable correlations.

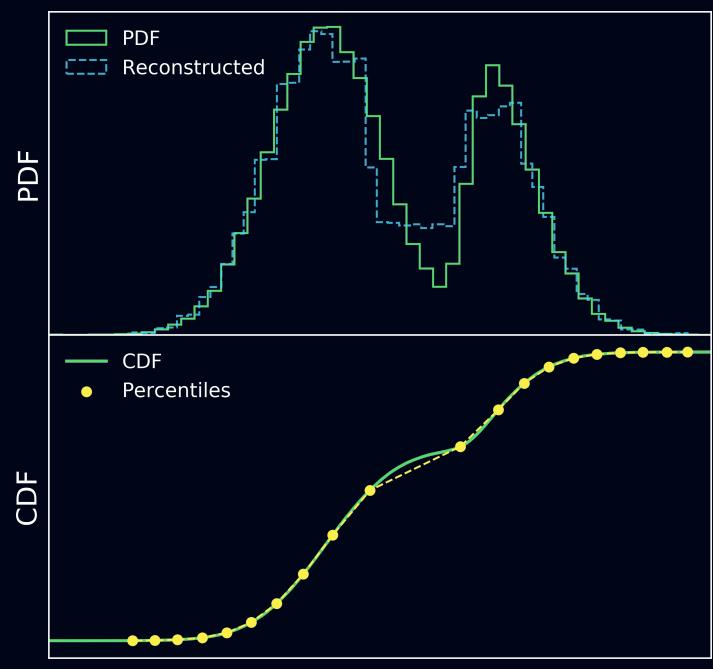


Percentiles

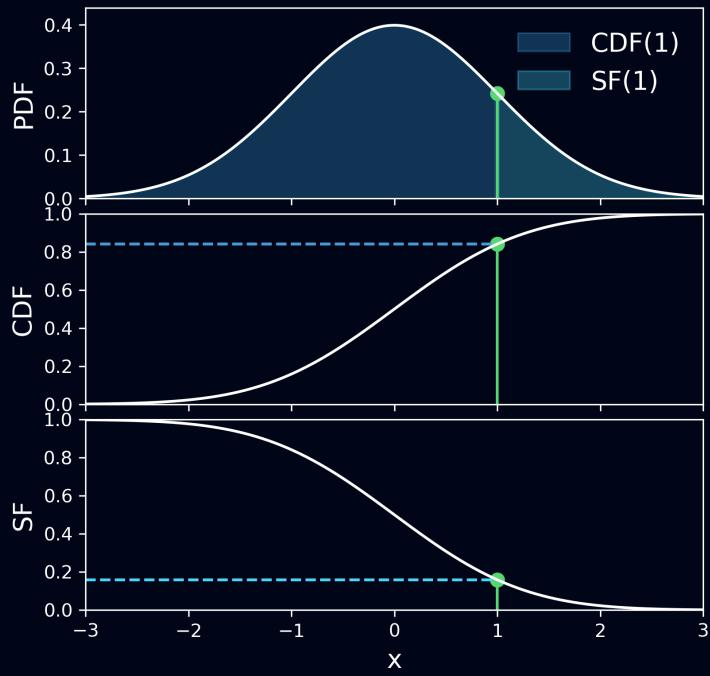
Percentiles

Percentiles allow an easy way to compress a 1D distribution down to an arbitrary number of percentile points.

An example with 20 percentiles (yellow) used to summarise the distribution (green), giving reconstructed results (blue). More than 20 points would give



Probability Basics



PDF and PMF

Probability density (continuous) or prob. mass (discrete). Probability of x falling between A and B is the area under the PDF from A to B .

Cumulative Distribution Function

Probability from negative infinity to x . Also known as the area under the PDF to the left.

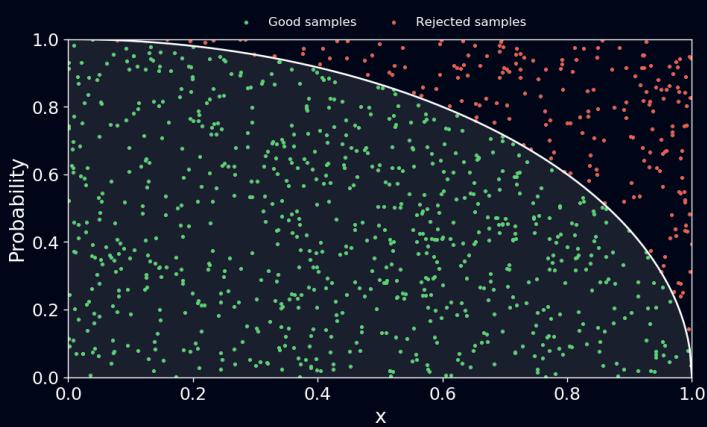
Survival Function

Area from x to positive infinity. $1 - \text{CDF}$.

Sampling Methods

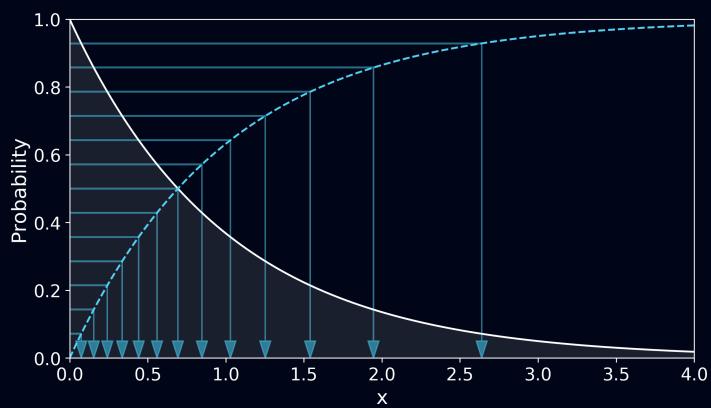
Rejection Sampling

Uniformly pick an x and y . Accept if $y < p(x)$, else pick another x and y . Repeat.

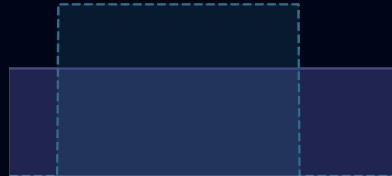


Inversion Sampling

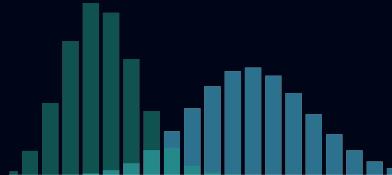
Integrate to get the CDF from the PDF, and then invert it (swap all xs and ys). Uniformly sample from the CDF and use the inverted function to get an x .



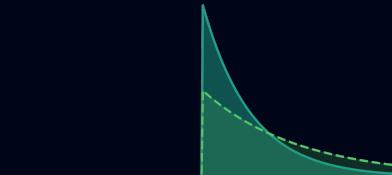
Common Distributions



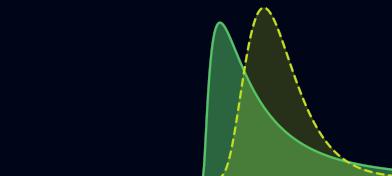
Uniform
Rolling a dice



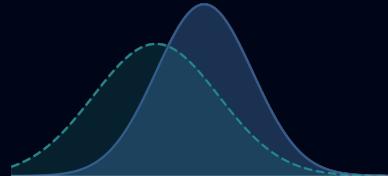
Binomial
Flipping a coin



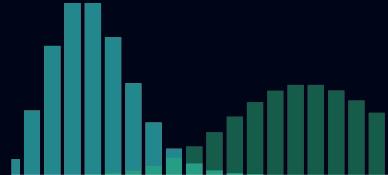
Exponential
Product failure



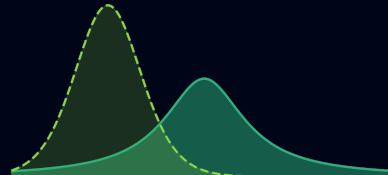
Lognormal
Mass, size,
positive values



Normal
Central Limit
Theorem



Poisson
Calls in a call
center



Student-t
Small-data
samples



Skewnormal
Generic
skewed dist