Probability Distributions

Joe

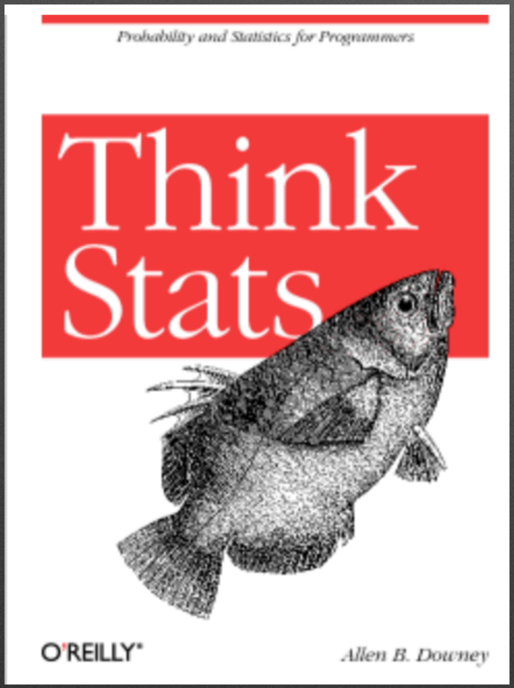
Introduction

Session Objective

1. Define the fundamental continuous and discrete probability distributions

2. Use matplotlib to visualize distributions of data and discuss data vis fundamentals

Resource



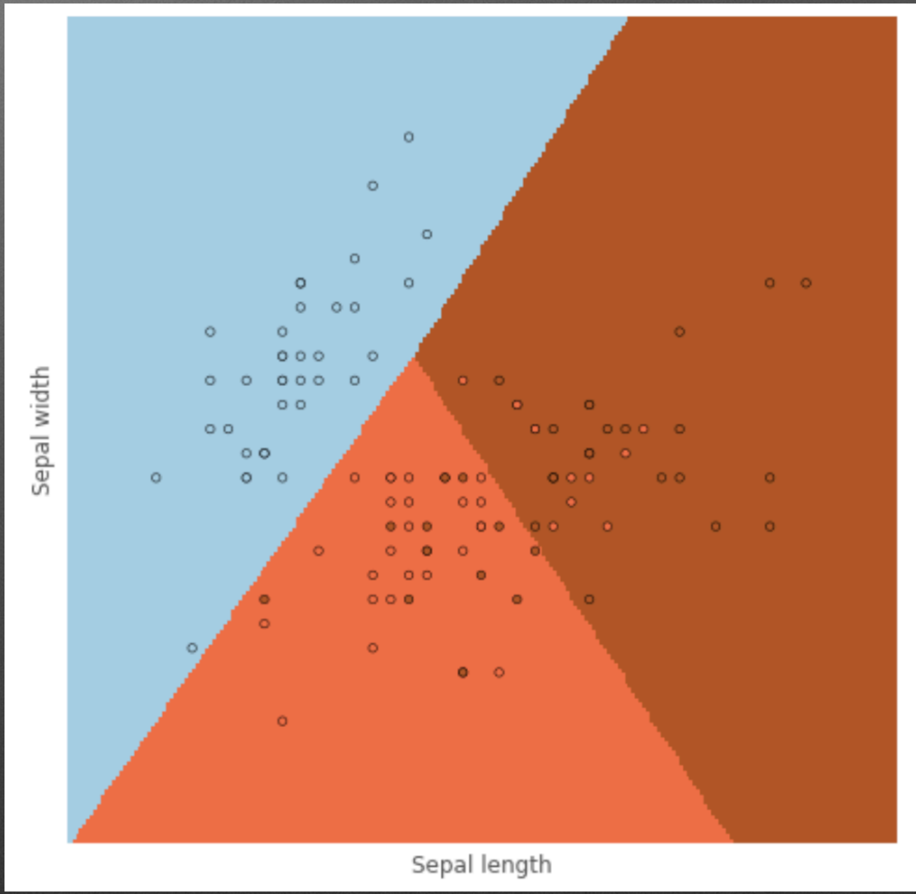
- “Think Stats” is available free online, or through amazon (I purchased a copy, I like authors)
- Great resource if any of the material today goes by too fast

Basics of Data Visualization

Data Visualization

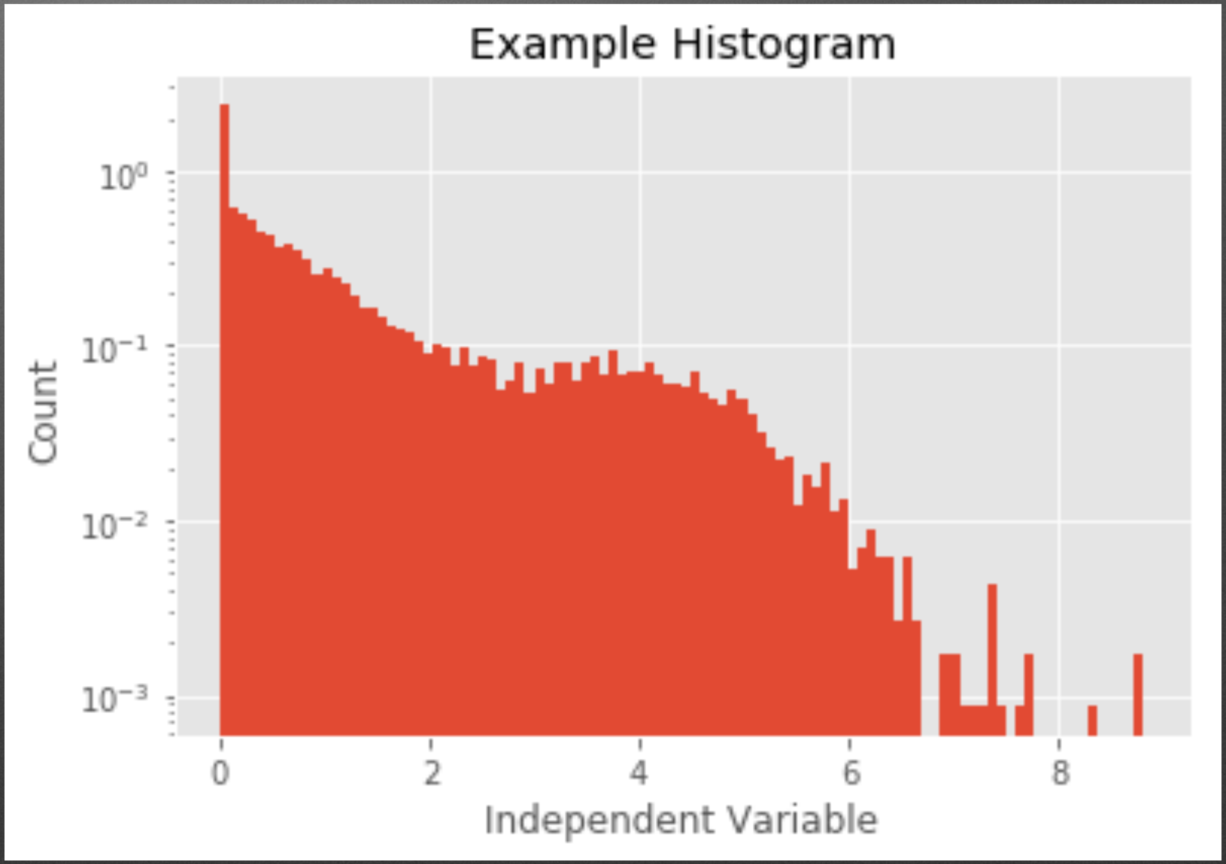
- Data visualization is a skill that you only get better at through <u>thoughtful</u> practice

- Depicted is the predictions of a logistic regression model on the iris dataset



Histograms

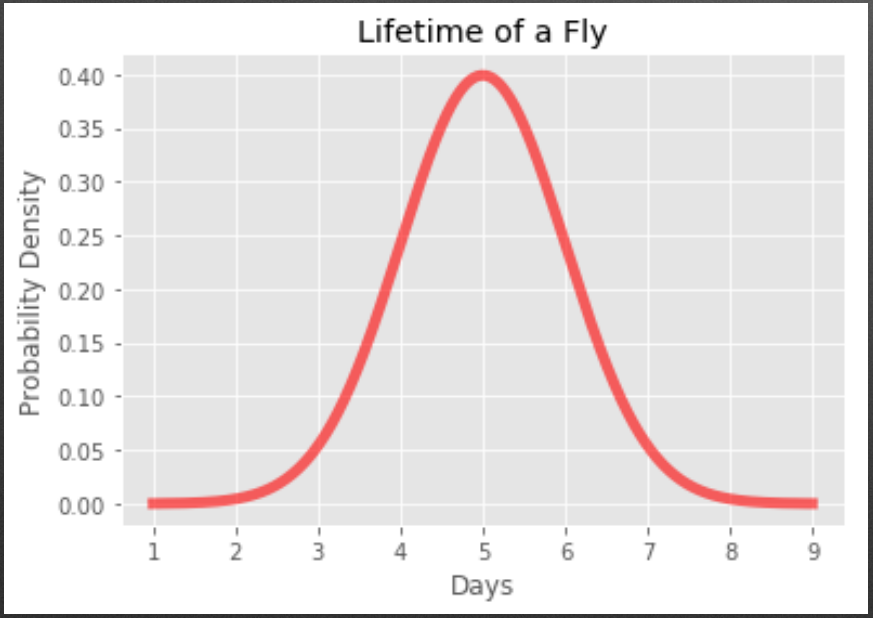
- The histogram is one of the more fundamental depictions of data.
- Histograms depict the how often a certain value occurs in a set of data.
- Let's hop into a notebook and go over some basic histogram creation.



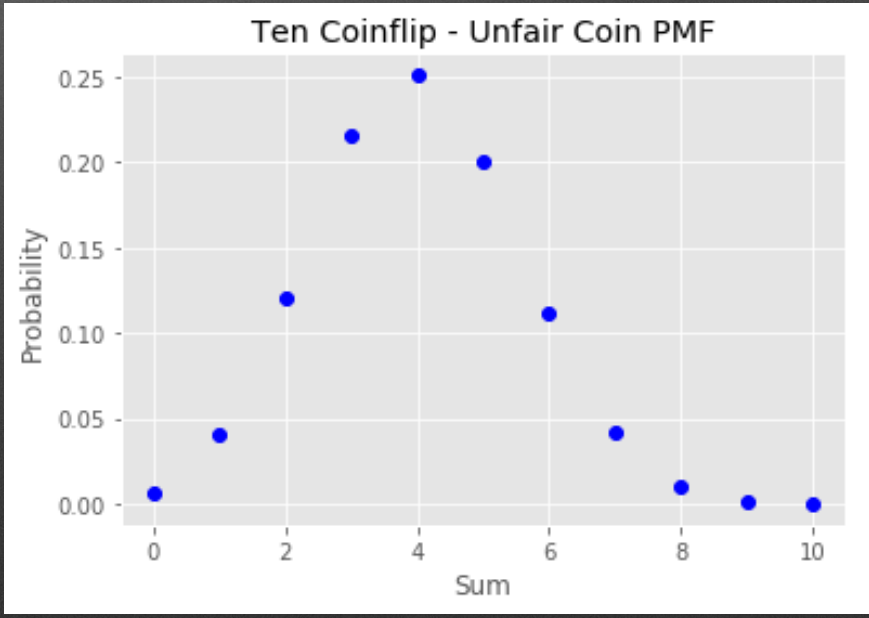
Fundamentals of Distributions

Continuous vs. Discrete

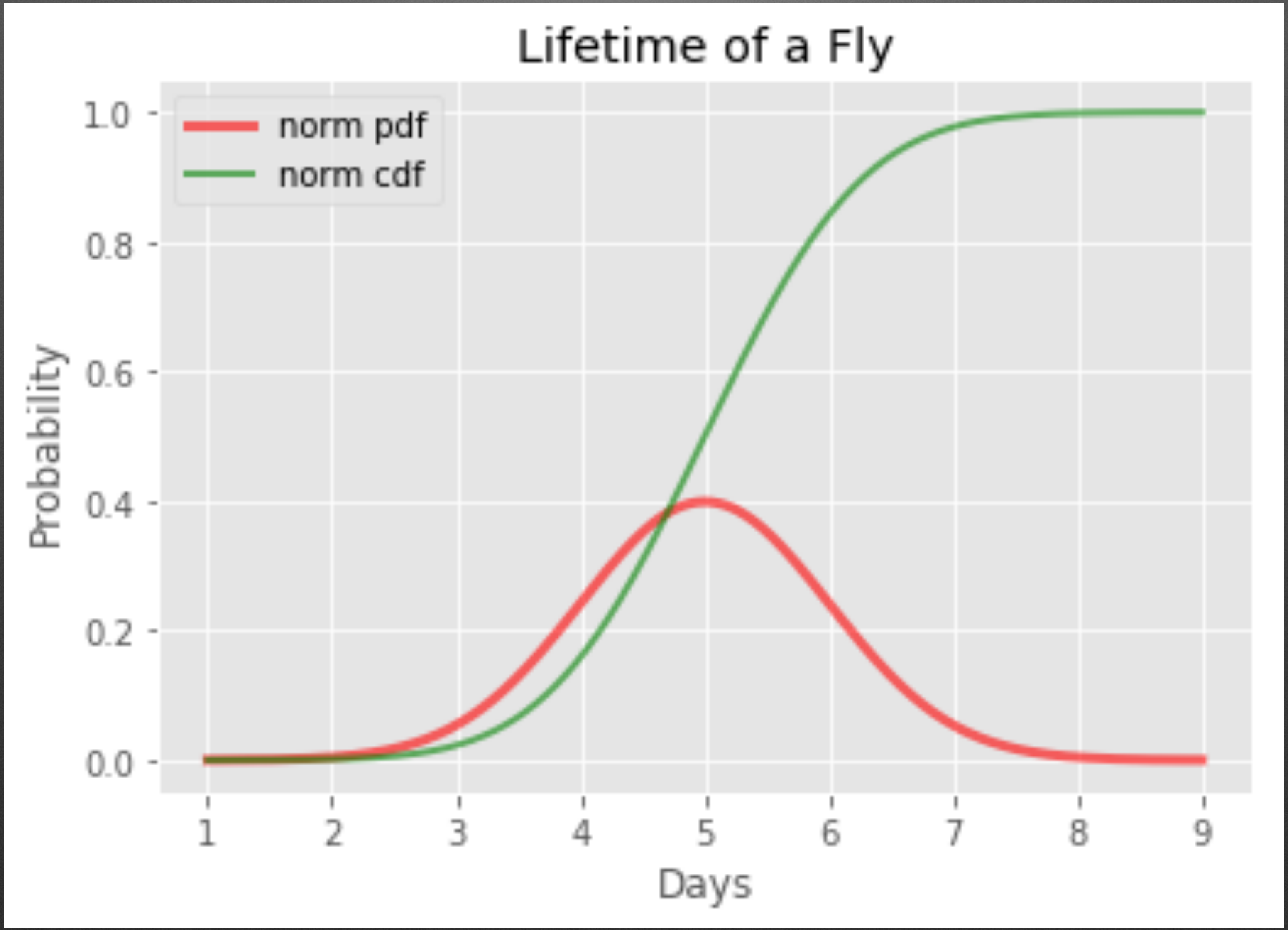
Continuous
- floating point numbers
- the life span of a fly
- prob. at a point is 0, only an integral has non-zero probability



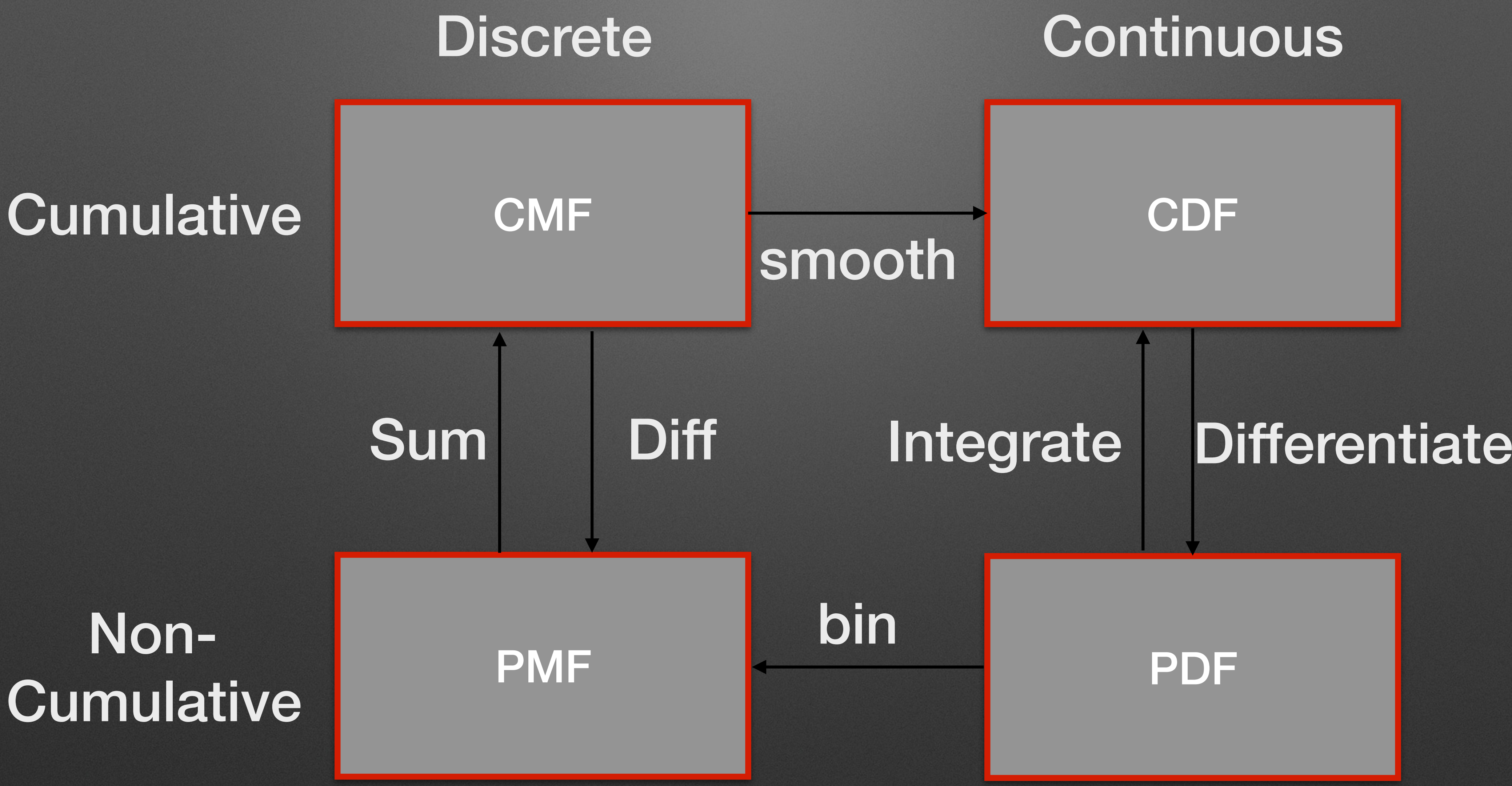
Discrete
- integers
- e.g. the number of heads when flipping coins



Cumulative Distribution Functions



Relationships



Expectation & Variance

Recall: Expectation and Variance

For **discrete** random variables (let P be the PMF of the r.v. X):

$$E(X) = \sum_{s \in S} s * P(X = s)$$

$$Var(X) = \sum_{s \in S} (s - E(X))^2 * P(X = s)$$

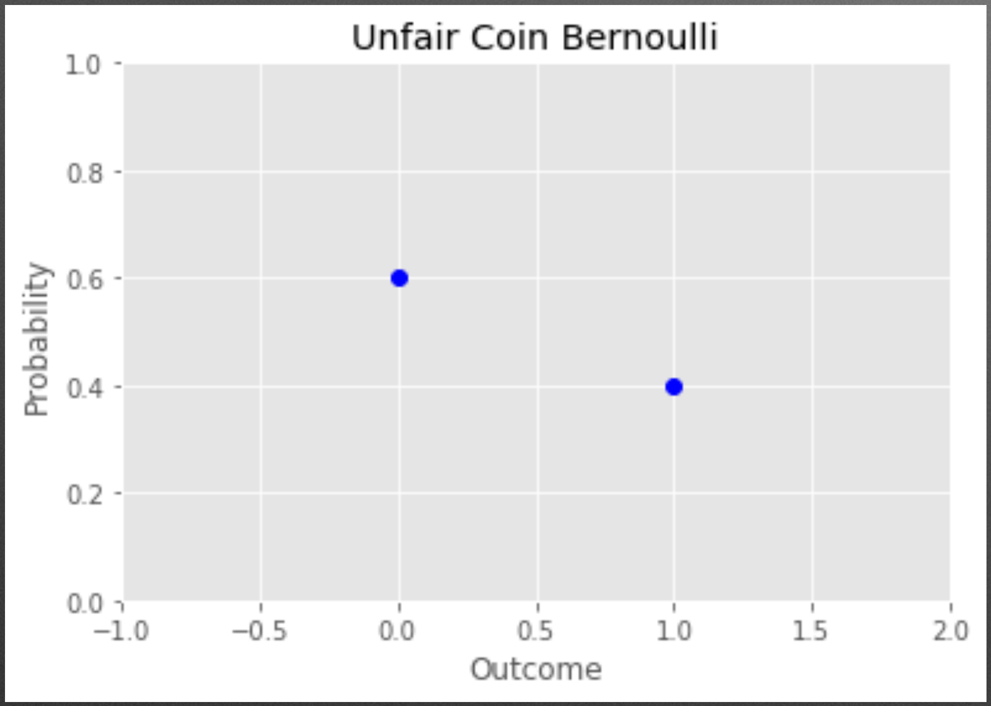
For **continuous** random variables (let f is the PDF of r.v. X):

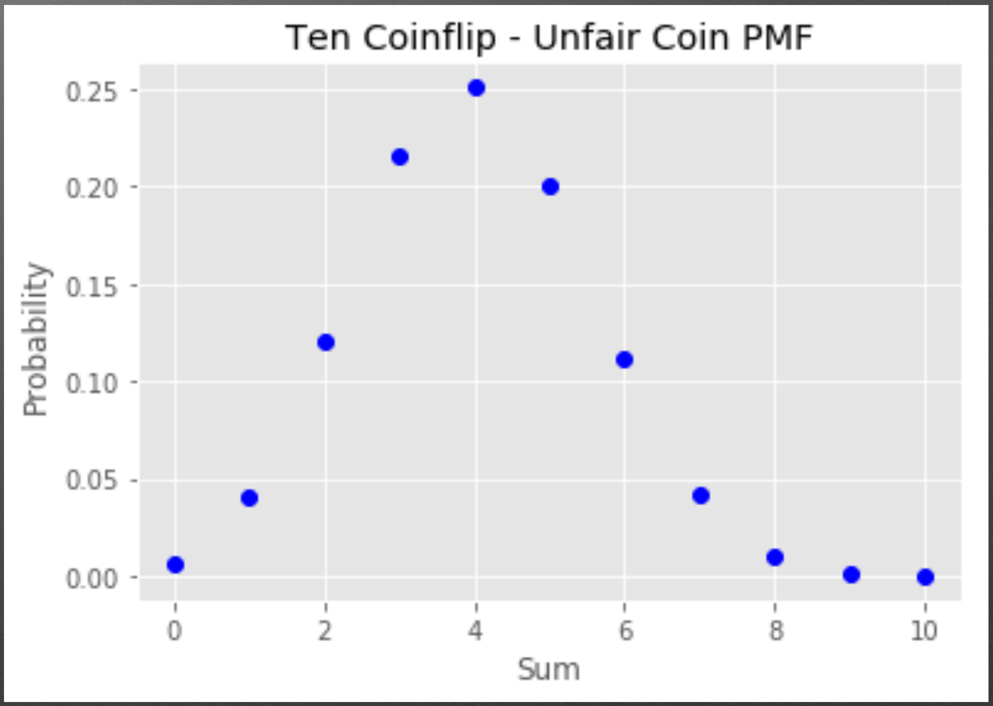
$$E(X) = \int_{x=-\infty}^{\infty} x * f(x) dx$$

$$Var(X) = \int_{x=-\infty}^{\infty} (x - E(X))^2 * f(x) dx$$

Important Discrete Distributions

Bernoulli & Binomial





PMF: $P[success] = p$, $P[failure] = 1 - p$

Support: $\{\text{success, failure}\}$

Mean: p

Variance: $p(1-p)$

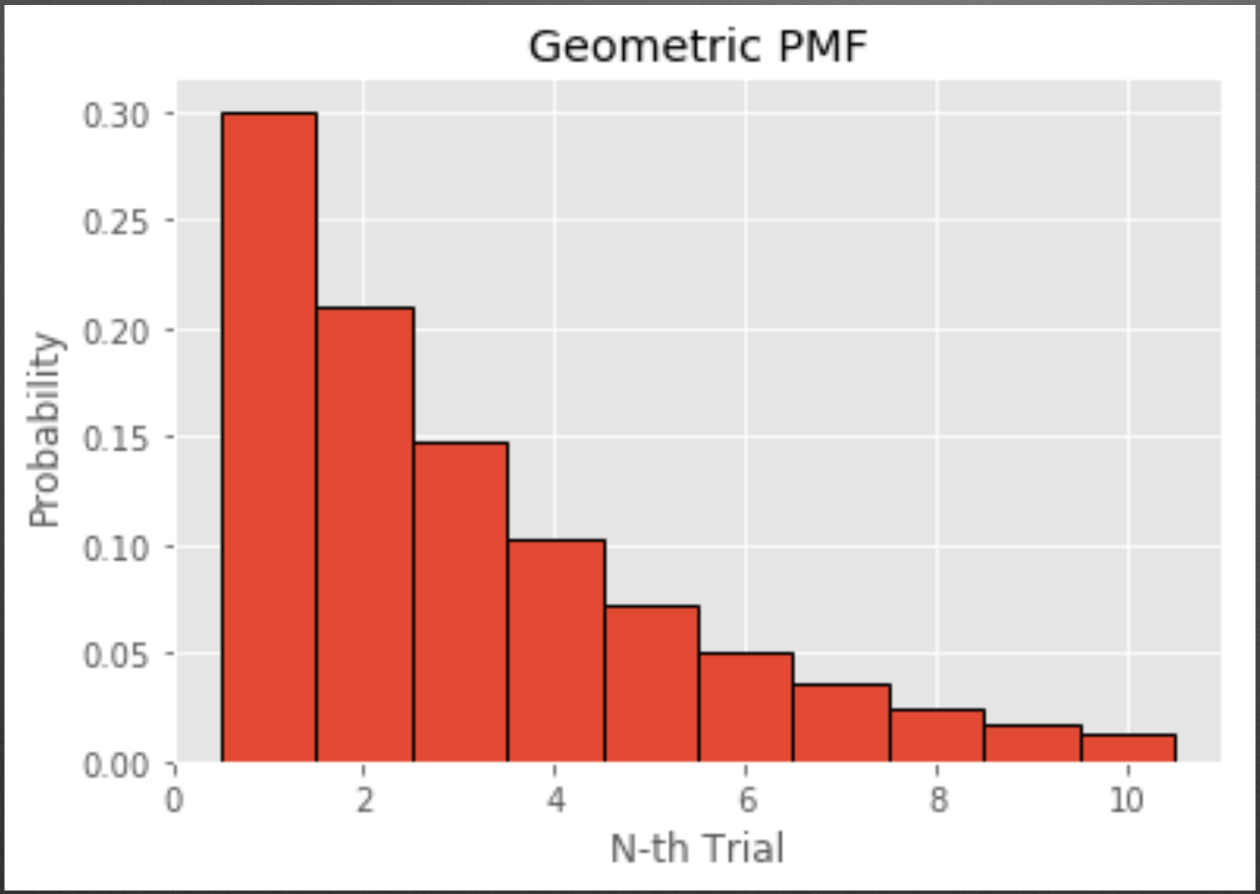
PMF: $P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$

Support: $k \in \{0, 1, \ldots, n\}$

Mean: np

Variance: $np(1-p)$

Geometric Distribution



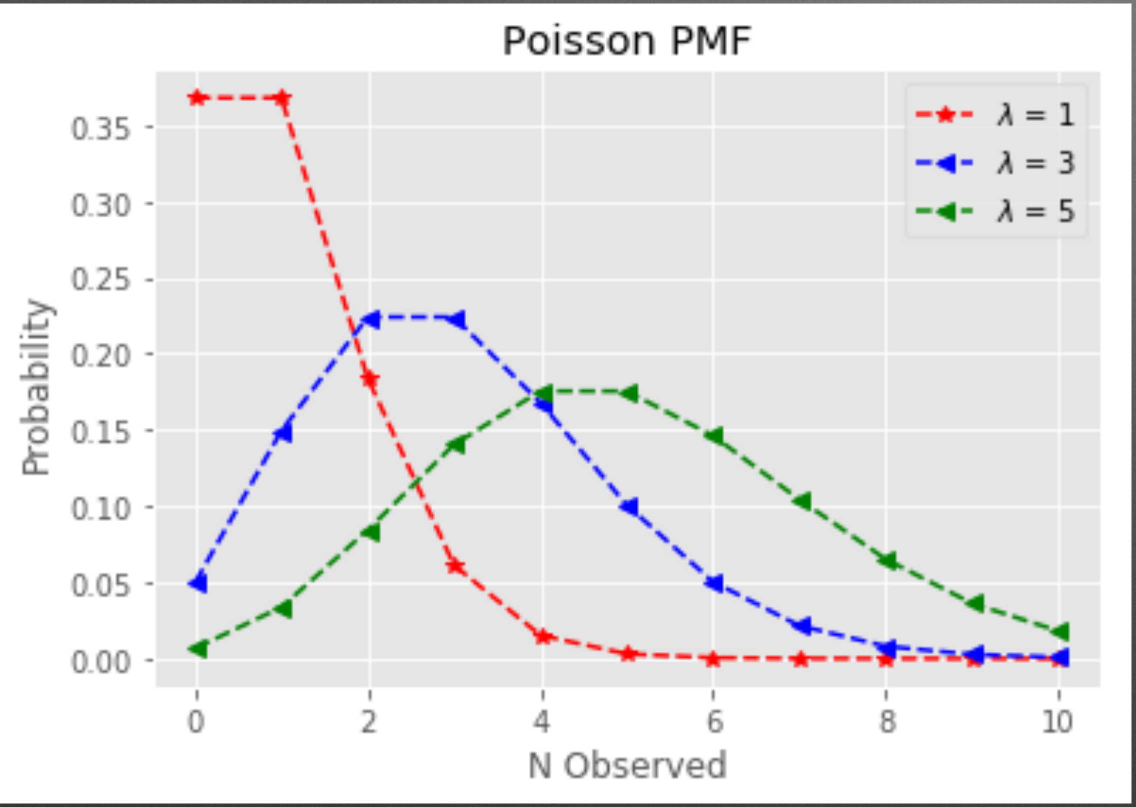
PMF: $P[X = k] = p(1-p)^{k-1}$

Support: $k \in \{0, 1, \dots\}$

Mean: $\frac{1}{p}$

Variance: $\frac{1-p}{p^2}$

Poisson Distribution



PMF: $P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$

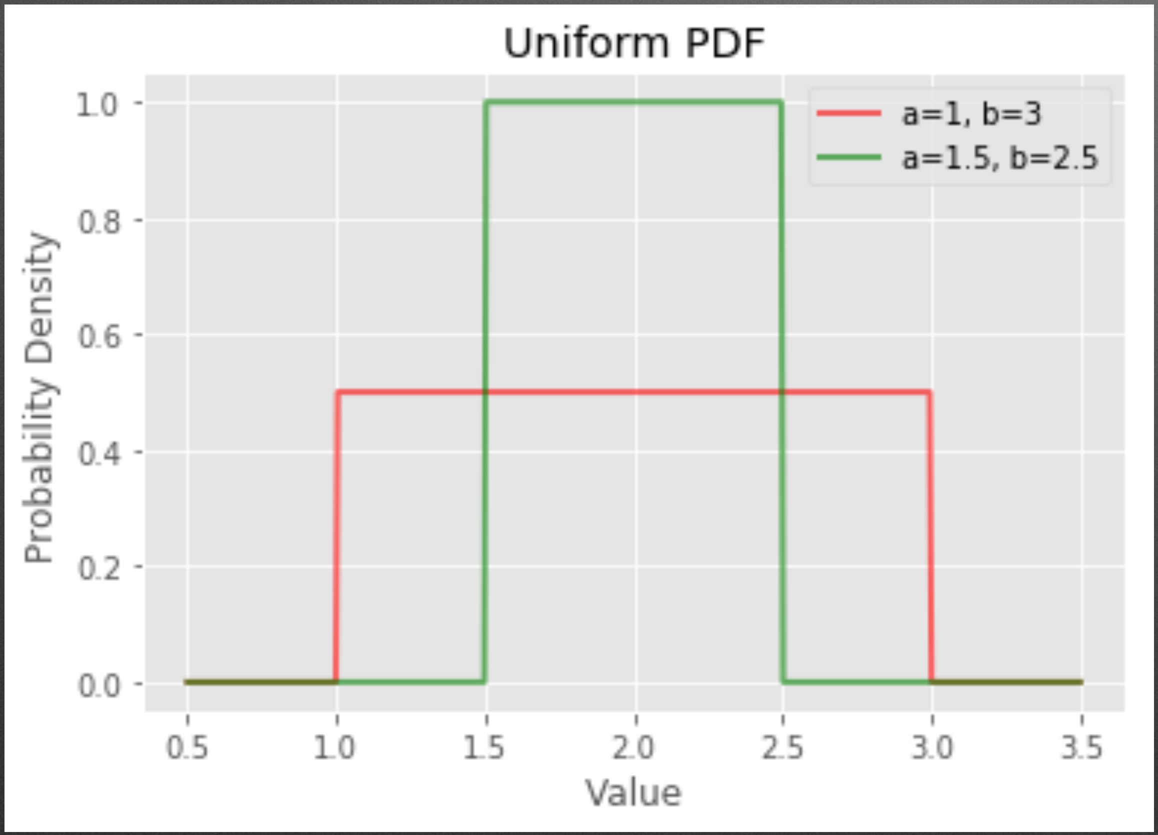
Support: $k \in \{0, 1, 2, \ldots\}$

Mean: λ

Variance: λ

Important Continuous Distributions

Uniform Distribution



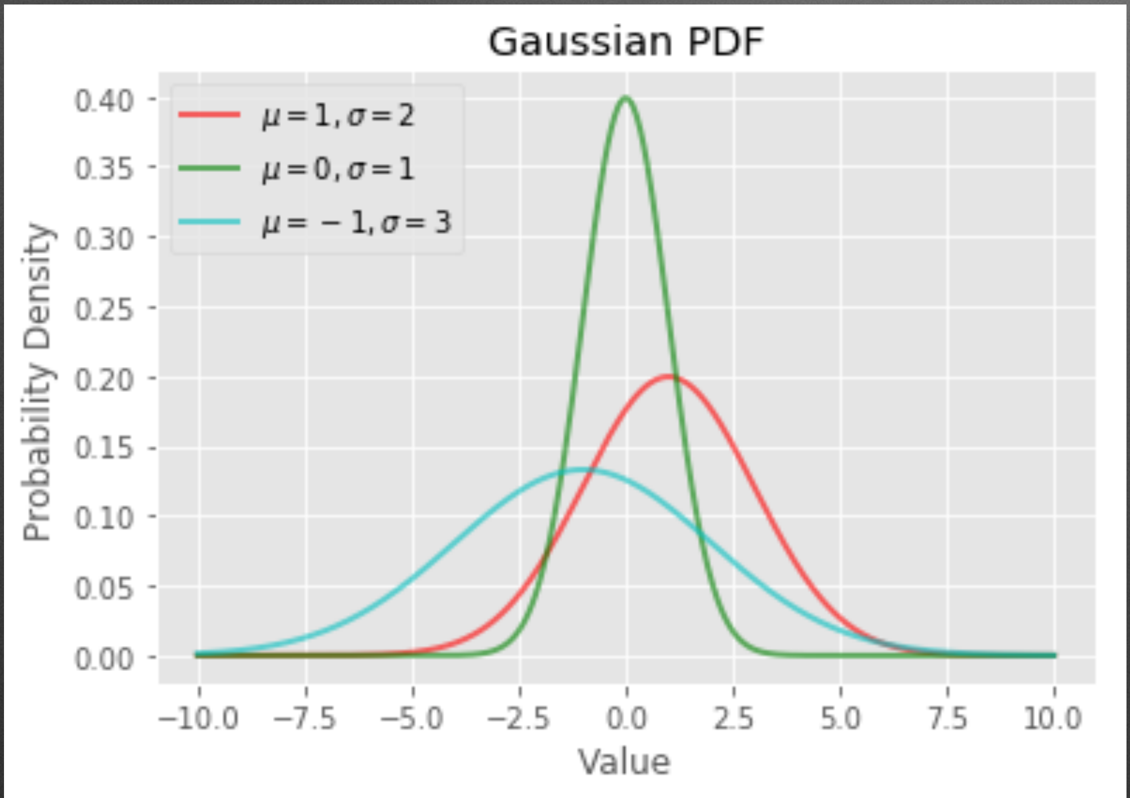
PDF: $f(x) = \frac{1}{b-a} * [\theta(a-x) - \theta(b+a-x)]$

Support: $x \in [a, b]$

Mean: $\frac{a+b}{2}$

Variance: $\frac{(b-a)^2}{2}$

Normal Distribution



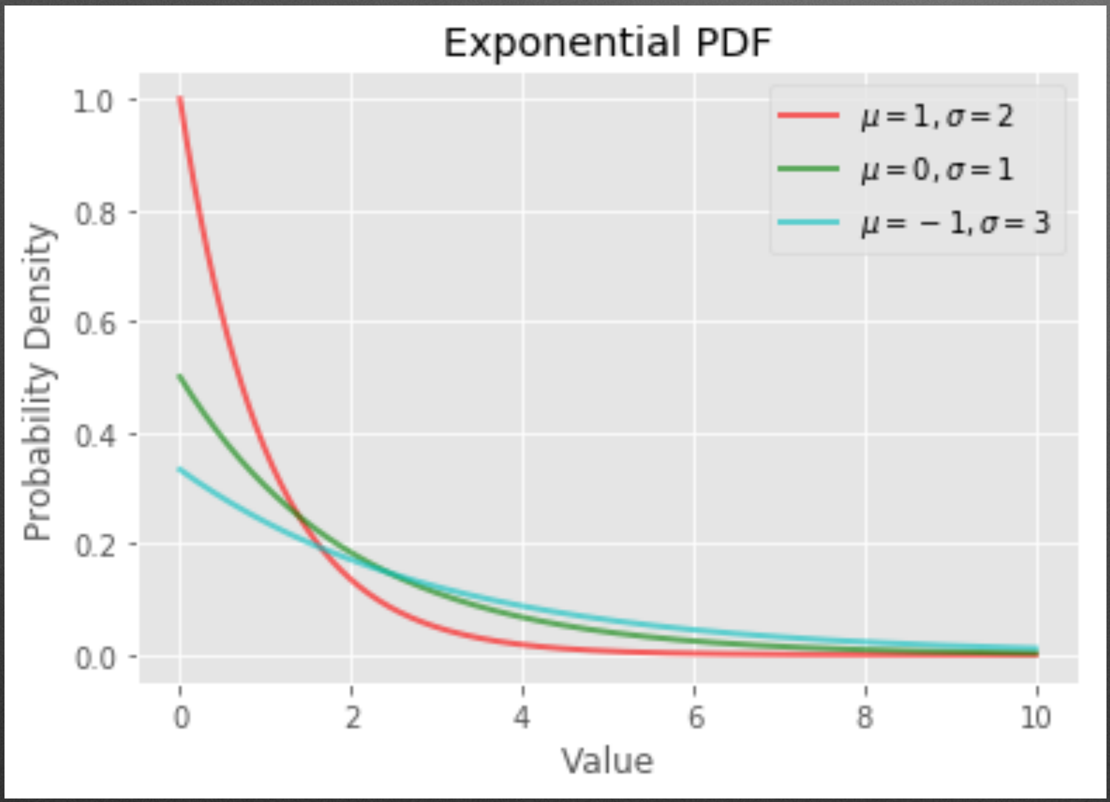
PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$

Support: $x \in (-\infty, \infty)$

Mean: μ

Variance: σ^2

Exponential Distribution



PDF: $f(x) = \lambda \exp(-\lambda x)$

Support: $x \in [0, \infty)$

Mean: $\frac{1}{\lambda}$

Variance: $\frac{1}{\lambda^2}$