topic 12: statistical distributions

stuff to learn today:

1. E(X) & Var(X)
2. PMF, PDF, CDF
3. Bernoulli & Binomial Distributions
4. Normal Distributions
5. One Sample Z-Test

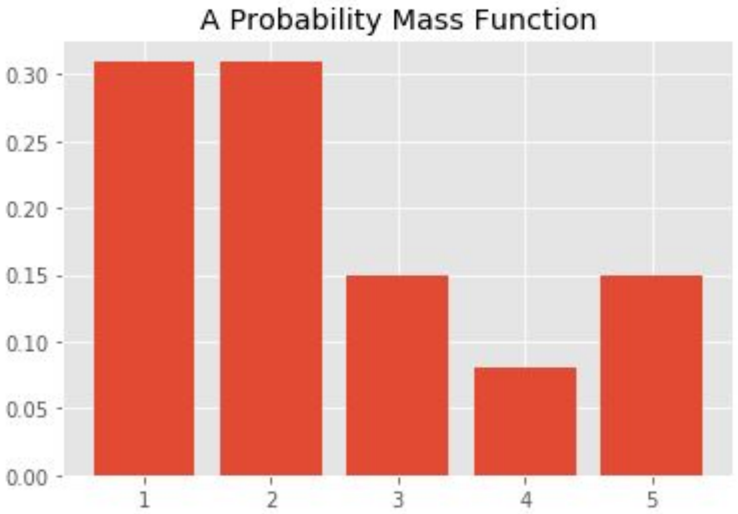
expected value & variance

- for a **random variable**:
  - expected value = $E(X) = \sum p(x_i) \cdot x_i$
  - variance = $Var(X) = \sum p(x_i) \cdot (x_i - E(X))^2$

- for specific named distributions, there are often formulas for the expected value and variance

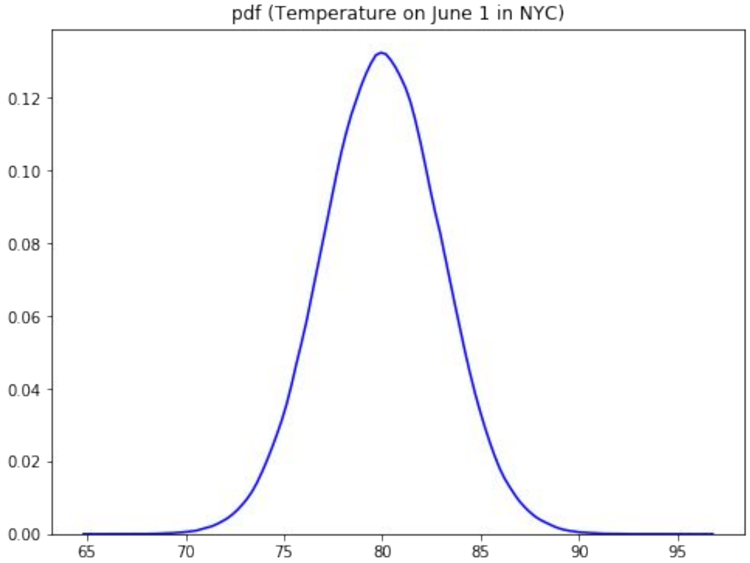
probability mass function

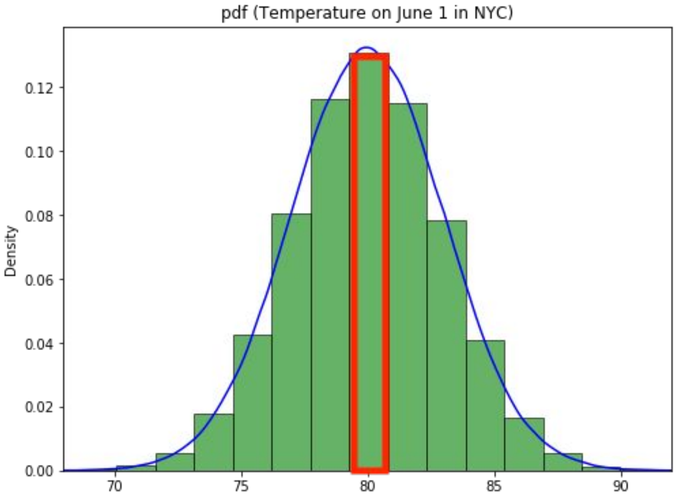
- associates probabilities with discrete random variables
- discrete = a known number of possible outcomes



probability density function

- associates probabilities with **continuous** random variables

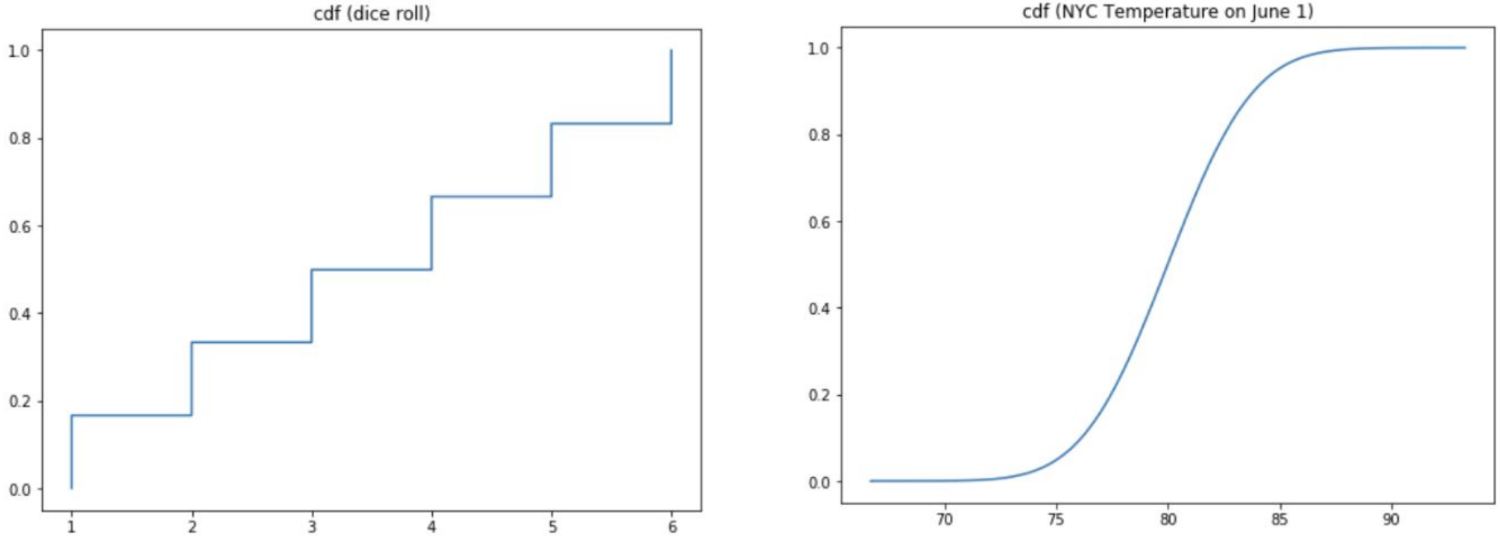




13% of the time you'll observe a temperature between 79.3 and 80.8 degrees

cumulative distribution function

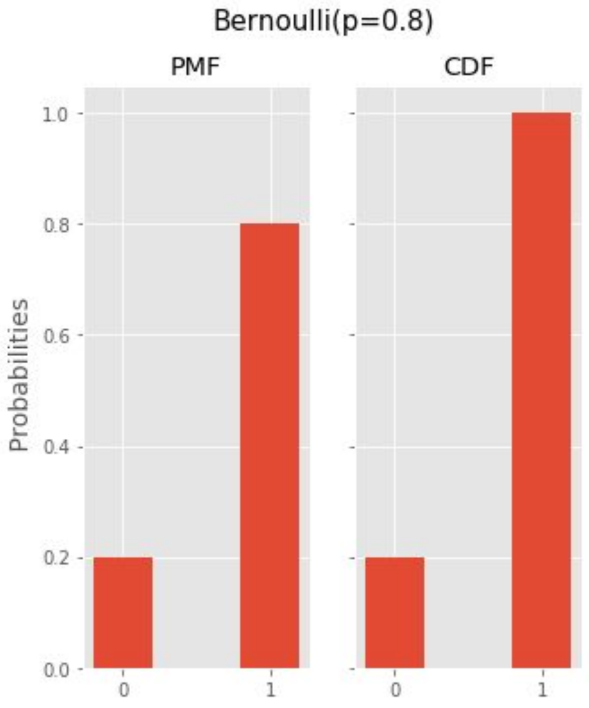
- shows $P(X \leq x)$ for any x within the sample space



the bernoulli distribution

- a one-trial, binary outcome experiment

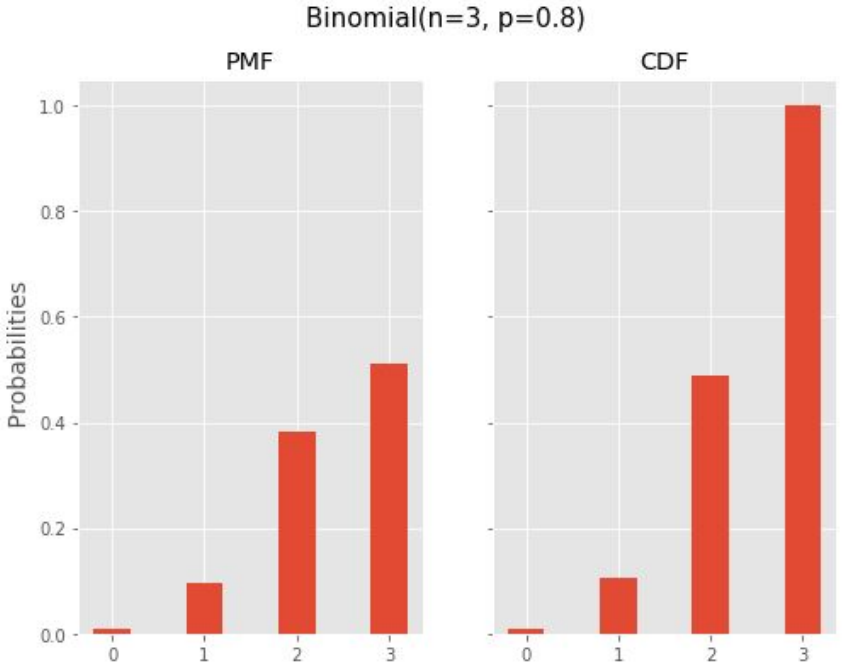
- $X \sim Ber(p=0.8)$
- $E(X) = 0.8$
- $Var(X) = 0.16$



the binomial distribution

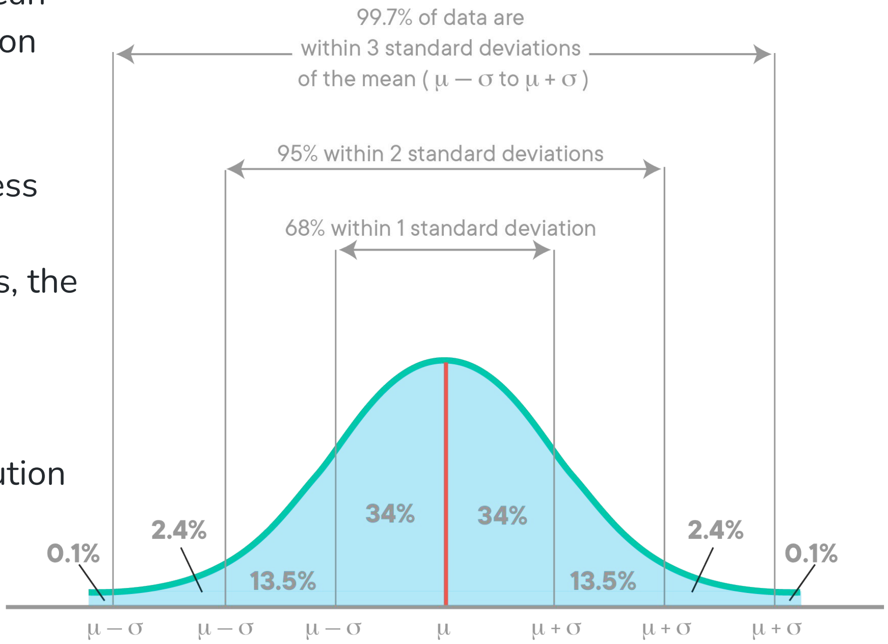
- a multi-trial, binary outcome experiment

- $X \sim Bi(n = 3, p=0.8)$
- $P(X=k) = {}^{n}C_{k}\,(p)^{k}\,(1-p)^{n-k}$
- $E(X) = 2.4$
- $Var(X) = 0.48$

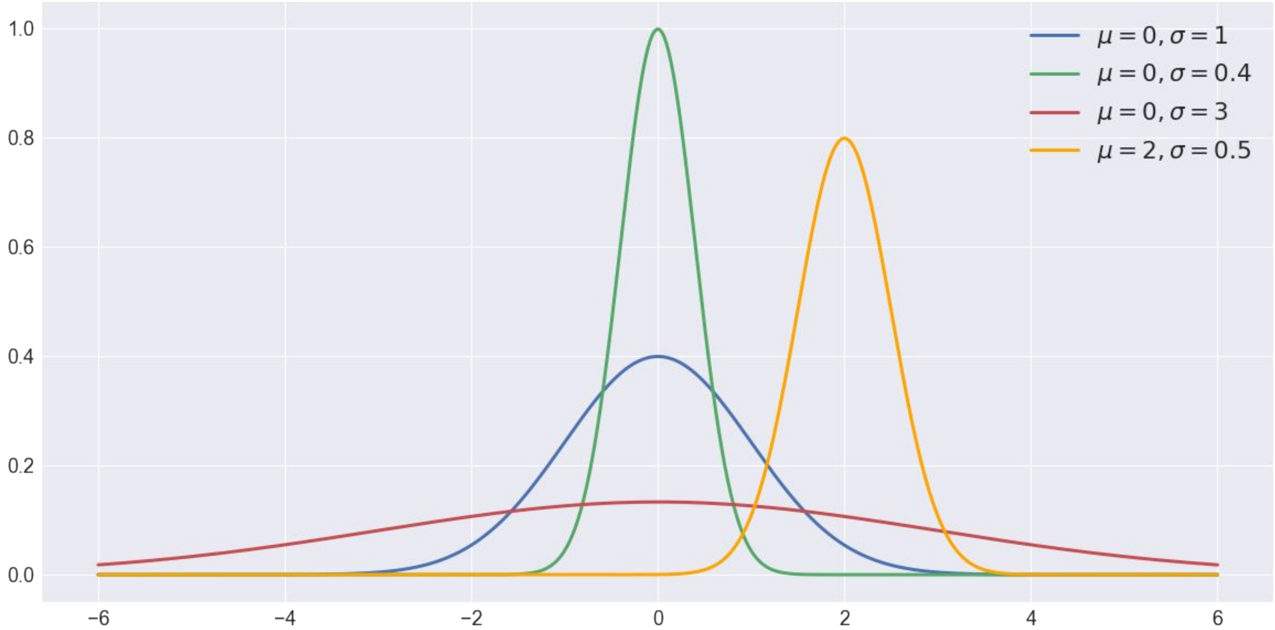


the Normal distribution

- also known as a Gaussian / bell curve
- Normal distributions are symmetric around their mean
- The mean, median, and mode of a normal distribution are equal
- The area under the bell curve is equal to 1.0
- Normal distributions are denser in the center and less dense in the tails
- Normal distributions are defined by two parameters, the mean and the standard deviation
- Around 68% of the area of a normal distribution is within *one standard deviation* of the mean
- Approximately 95% of the area of a normal distribution is within two standard deviations of the mean



expected value & variance



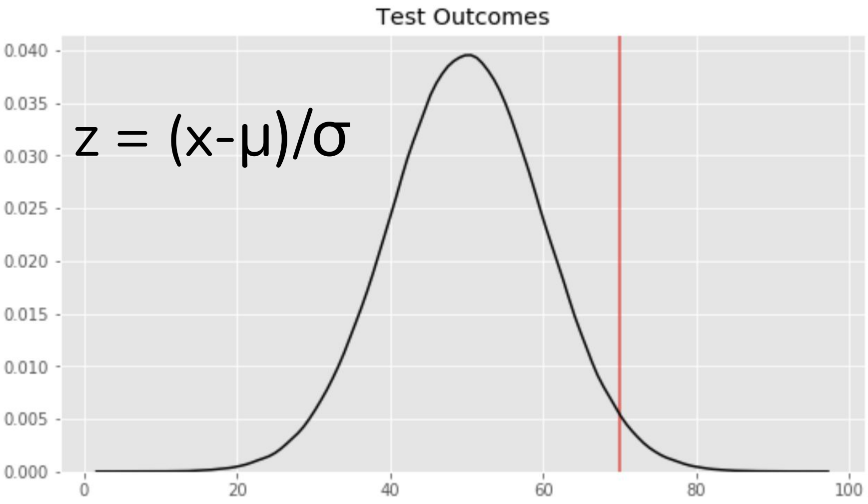
- $E(X) = \mu$
- $Var(X) = \sigma^2$

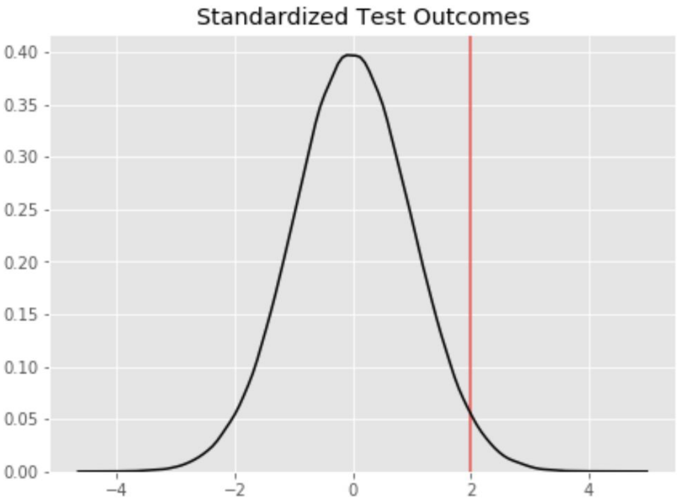
PDF	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
CDF	$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$

the Standard Normal distribution

- $E(X) = 0$, $Var(X) = 1$
- Allows us to compare different normal distributions
- z-scores!

$$z = (x-\mu)/\sigma$$





z-score

- a z-score tells us how many standard deviations away from the mean a point would be in a Standard Normal distribution
- z-scores are associated with cumulative probabilities, retrieved from the z-table
- $z = (x-\mu)/\sigma$ for a single point
- $z = (x-\mu)/(\sigma/\sqrt{n})$ for a sample