# Math Tools Lab

## Summary Statistics and Probability Distributions

30 Oct 2020

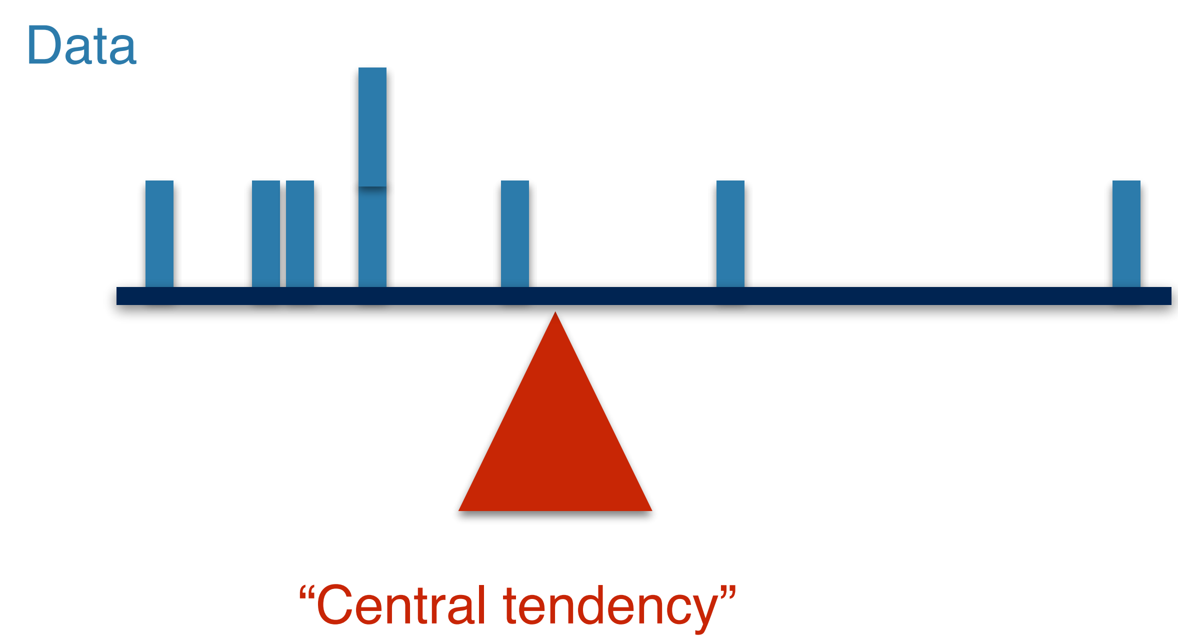
# Summary Statistics

- If you had to choose one number to explain your data, what number would you choose?
- Generally speaking you may want some point which is close to your cloud all of your data points
  - What do we mean by close? How do we measure that?

$$\arg\min_c \left[ \frac{1}{N} \sum_{n=1}^{N} |x_n - c|^p \right]^{1/p}$$

- Norms:

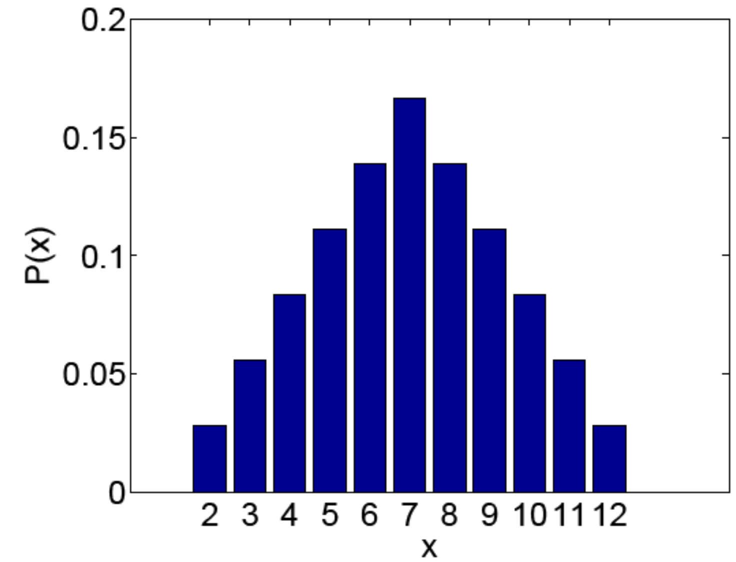
| | |
|---|---|
| $L_0$ | Mode |
| $L_1$ | Median |
| $L_2$ | Mean |



# Exercises: Summary Stats

# Discrete Probability Distributions: PMF

A discrete random variable is one that can equal a **finite number of values**. We can describe the distribution of a discrete r.v. using a **probability mass function (pmf)**, which must be non-negative for all inputs and sum to 1. A pmf describes the probability that a discrete r.v. $X$ takes on a particular value, $P(X = x)$.




What is the pmf of (2)?

Examples of discrete r.v.'s

(1) $X$ = fair die

(2) $X$ = sum of two rolled dice

(3) $X$ = maximum of two rolled dice

(4) $X$ = # of coin flips until the 1st heads

(5) $X$ = number of kids in a family

| x | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P(X=x) | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 |

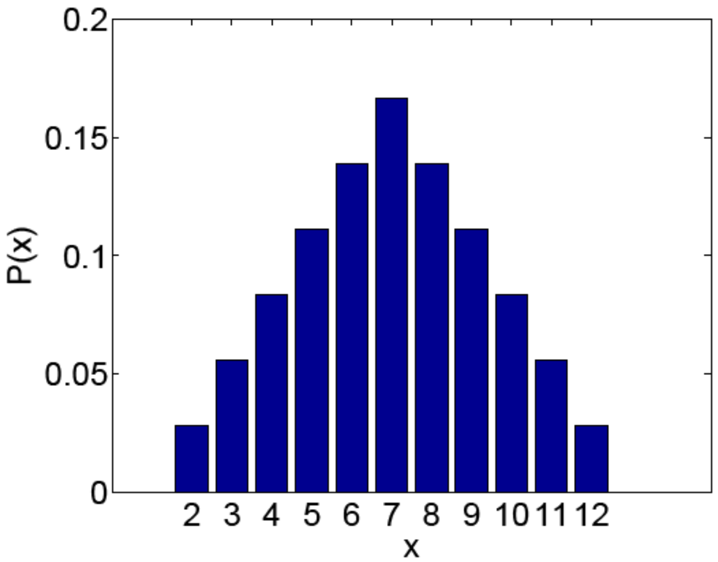
Now, what is the probability that the sum of two rolls is at least 5? What about no more than 8?

$$P(X \geq 5) = \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{25}{36}$$

$$P(X \leq 8) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36} = \frac{26}{36}$$

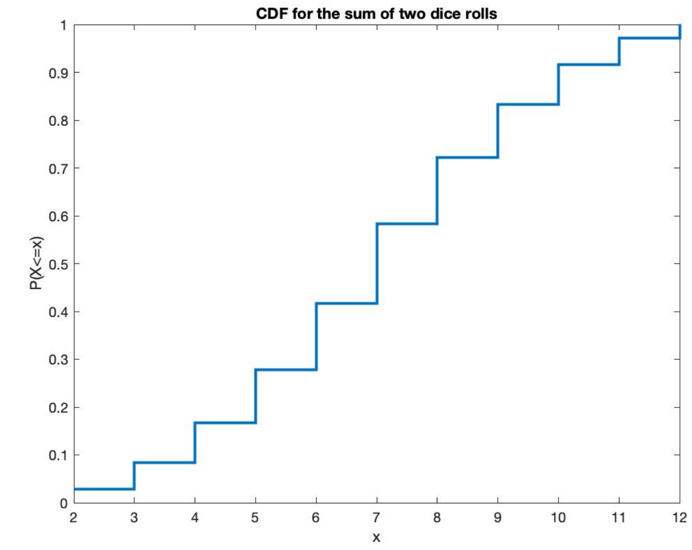
# Discrete Probability Distributions: CDF

Another way to describe a r.v.'s distribution is with a **cumulative distribution function (cdf)**, which hdescribes the probability that a discrete r.v. $X$ is less than or equal to a particular value $k$, $P(X \leq k)$.




Back to our example: what is the cdf of the sum of two dice rolls?

| x | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| P(X=x) | 1/36 | 2/36 | 3/36 | 4/36 | 5/36 | 6/36 | 5/36 | 4/36 | 3/36 | 2/36 | 1/36 |
| P(X≤ x) | 1/36 | 3/36 | 6/36 | 10/36 | 15/36 | 21/36 | 26/36 | 30/36 | 33/36 | 35/36 | 36/36 |




We can use the cumulative distribution function to generate data/draw samples from a distribution (will go through in exercise)

# Expected Value

The **expected value** of a r.v. $X$ is a **weighted average of all the possible values** of $X$.

Discrete r.v.: $E(X)$ is a sum

$$\mathbb{E}(x) = \sum_{n=1}^{N} x_n \, p(x_n)$$

Linearity of expectation: If $Y = aX + b$, then $E(Y) = E(aX + b) = aE(X) + b$

Example 1: Suppose you roll a die, and are paid \$1 for odd rolls and \$2 for even rolls. What is the expected value for one roll?

| x | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| P(X=x) | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |
| Amount (\$) | 1 | 2 | 1 | 2 | 1 | 2 |

$$E(X) = 1(\frac{1}{6}) + 2(\frac{1}{6}) + 1(\frac{1}{6}) + 2(\frac{1}{6}) + 1(\frac{1}{6}) + 2(\frac{1}{6}) = \frac{9}{6} = 1.5$$

If you were given data on how much you money you made on each roll, you could get the sample mean of the money made on one roll

$$\frac{1}{N}\sum_{n=1}^{N} x_n = \frac{1}{N}\overrightarrow{1}^T\overrightarrow{x}$$

# Variance and Standard Deviation

For any r.v., we want to summarize its central tendency but also its spread. The variance of a r.v. $X$ is given by finding the **average squared deviation** of X from its mean, $E(X)$.

$$Var(X) = \sigma_X^2 = E[(X - E(X))^2] = E(X^2) - [E(X)]^2$$

$$s_X^2 = \frac{1}{N}\sum_{n=1}^{N} x_n^2 - \bar{x}^2 = \frac{1}{N}\|\overrightarrow{x}\|^2 - \bar{x}^2$$

Since variance averages **squared deviations**, its units are the square of the units of the original r.v. So we define **standard deviation.**

$$\sigma_X = \sqrt{Var(X)}$$

$$s_X = \sqrt{\frac{1}{N}\|\overrightarrow{x}\|^2 - \bar{x}^2}$$

Example: Let $X$ = the number of bases a baseball player earns per at-bat. Given the probability function below, find the expected value, variance, and standard deviation of $X$.

| k | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| P(X=k) | 0.65 | 0.25 | 0.06 | 0.01 | 0.03 |

$$E(X) = 0(0.65) + 1(0.25) + 2(0.06) + 3(0.01) + 4(0.03) = 0.52$$

$$E(X^2) = \sum_{k=0}^{4} k^2 P(X = k) = 1.06$$

$$Var(X) = E(X^2) - E(X)^2 = 1.06 - 0.52^2 = 0.7896$$

$$sd(X) = \sqrt{Var(X)} = 0.8886$$

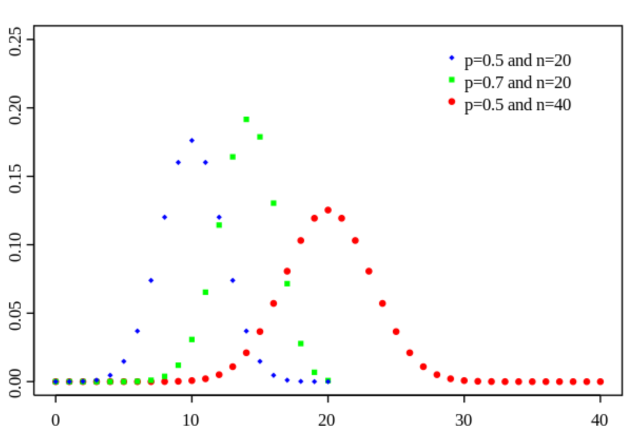
# Discrete Probability Distributions: Examples

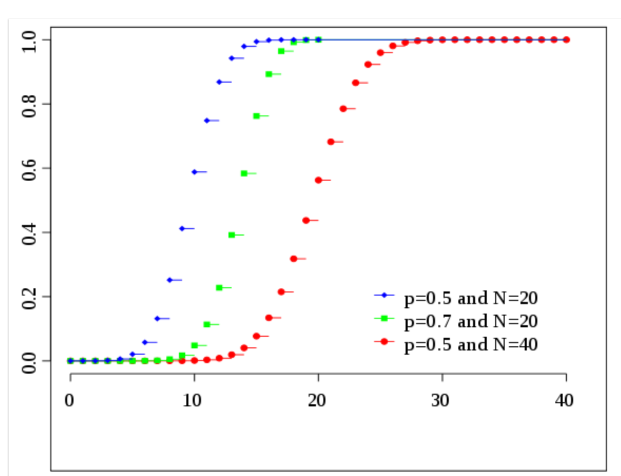
## Binomial Distribution

The distribution for the number of successes in a sequence of $n$ independent experiments, each with 2 possible outcomes where $p$ is the probability of success. Example: flipping a coin.

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

Mean = $np$; variance = $np(1-p)$



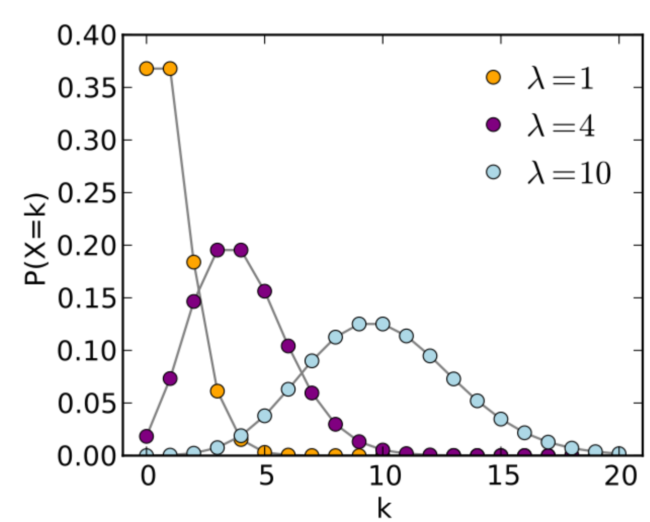



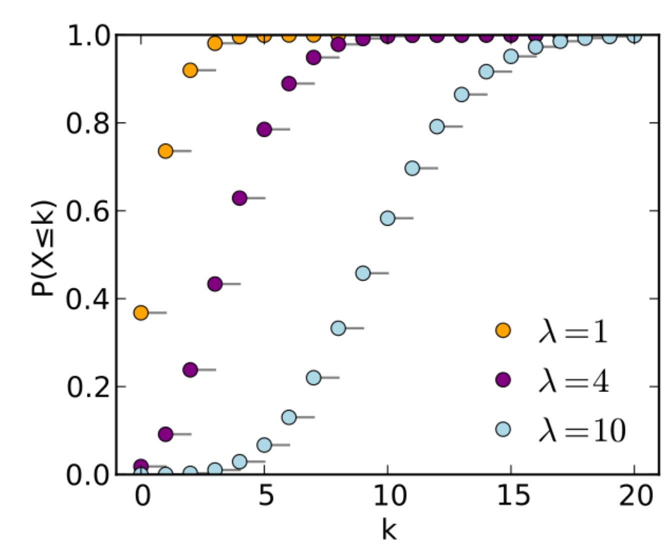

## Poisson Distribution

The distribution that expresses the probability that a given number of events will occur in a fixed interval of time $\lambda$. The event must occur with a constant rate and independently of the last event. Example: neural spike counts.

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Mean = $\lambda$; variance = $\lambda$



# Exercises: Probability Distributions

# Covariance and the covariance matrix

Covariance measures the **joint variability of two r.v.'s**. If both $X$ and $Y$ tend to be "big" at the same time, then $Cov(X, Y) > 0$. If one tends to be "big" when the other is "small", then $Cov(X, Y) < 0$.

$$Cov(X, Y) = Cov(Y, X) \\ = E(XY) - E(X)E(Y)$$

$$Cov(X, X) = Var(X)$$

The covariance matrix is a matrix whose element in the $i, j$ position represents **the covariance between the $i^{th}$ and $j^{th}$ elements** of a random vector (a r.v. with multiple dimensions). This generalizes the notion of covariance to multiple dimensions. For example, the variation of a collection of two-dimensional points cannot be fully characterized by a single number or just the variances in the $x$ and $y$ directions: a $2x2$ matrix is needed.

eg. We have random variables X and Y, then the covariance matrix is

$$\begin{bmatrix} Cov(X, X) & Cov(X, Y) \\ Cov(Y, X) & Cov(Y, Y) \end{bmatrix} = \begin{bmatrix} Var(X) & Cov(X, Y) \\ Cov(Y, X) & Var(Y) \end{bmatrix}$$

To get the sample covariance matrix from data $\vec{d_n} = \begin{bmatrix} x_1 & x_2, & \dots & x_n \\ y_1 & y_2 & \dots & y_n \end{bmatrix}$

$$\frac{1}{N}\begin{bmatrix} ||\vec{x}||^2 & \vec{x}^T\vec{y} \\ \vec{y}^T\vec{x} & ||\vec{y}||^2 \end{bmatrix} - \begin{bmatrix} \bar{x}^2 & \bar{x}\bar{y} \\ \bar{y}\bar{x} & \bar{y}^2 \end{bmatrix} = \begin{bmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{bmatrix}$$

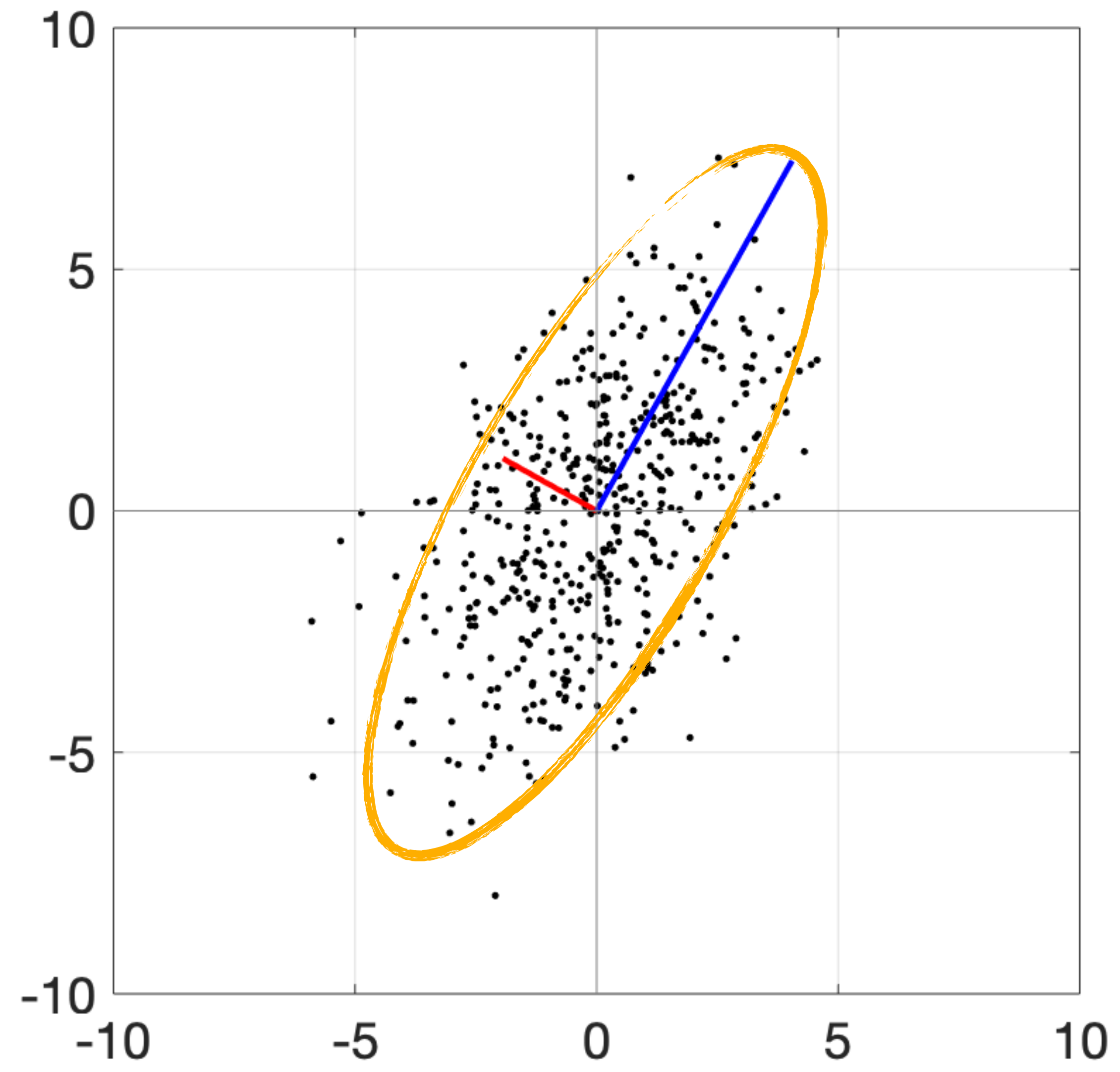
$$= \frac{1}{N}(\vec{d_n} - \bar{d})'(\vec{d_n} - \bar{d})$$

# Covariance and the covariance matrix

How can we use the covariance matrix?

Recall when we used it during the PCA Lab: we wanted to find the directions that captured the most spread in the data

One of the ways we did this was by getting the eigenvalue decomposition of C = X'X which we now know is the covariance matrix.




With this in mind you'll go through two exercises which help us understand how a linear transformation of the data affects the covariance matrix (and the data), as well as how to get rid of covariances in the data (PCA whitening)

# Exercises: Covariance Matrix & Transformations