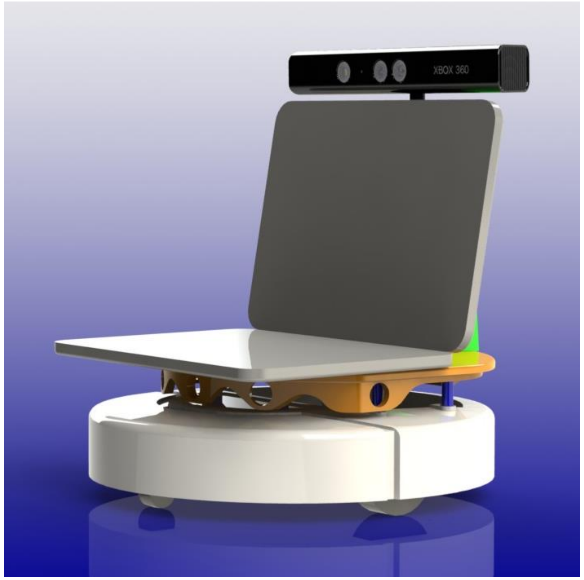
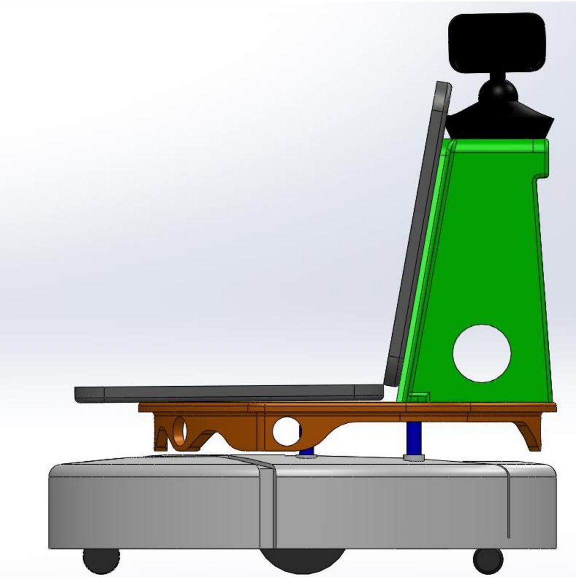
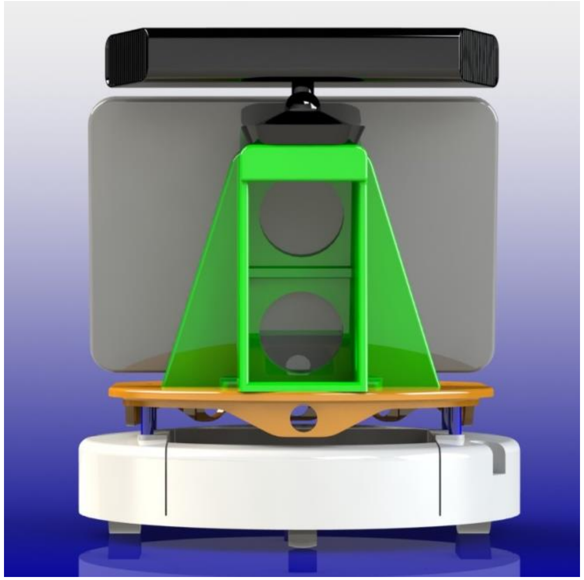
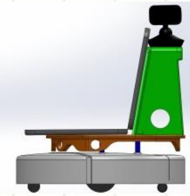
MAE 493G, CpE 493M, Mobile Robotics

7. Basic Statistics



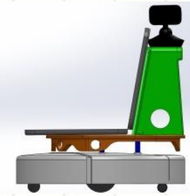



Instructor: Yu Gu, Fall 2013



Probability Vs. Statistics

- The best explanation I have found is here: http://www.cs.sunysb.edu/~skiena/jaialai/excerpts/node12.html
- "Probability deals with predicting the likelihood of future events, while statistics involves the analysis of the frequency of past events";
- "Probability theory enables us to find the consequences of a given ideal world, while statistical theory enables us to measure the extent to which our world is ideal";
- We really need both in robotics: statistics for data analysis and probability to make decisions.

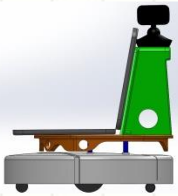


Random Variable

- **Random Variable**: a random variable or stochastic variable is a variable whose value is subject to variations due to chance. A random variable (*X*) is defined on a set of possible outcomes (*S*) and a probability distribution that associates each outcome with a probability. As opposed to other mathematical variables, a random variable conceptually *does not have a single, fixed value* (even if unknown); rather, it can take on a set of possible different values, each with an associated probability;

- **Discrete Random Variable** - it may assume any of a specified list of exact values (i.e. dice);

- **Continuous Random Variable** - it may assume any numerical value in an interval or collection of intervals (i.e. error in the distance measurement using a ruler).



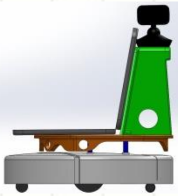
Basic Statistical Terms

- **Mode**: values that occur most frequently in a distribution (may have more than one mode in a data set);
- **Median:** value midway in the frequency distribution;
- **Mean:** Mean can have different meanings:

  1. arithmetic average: $x_m = \frac{1}{N}\sum_{i=1}^{N} x_i$
  2. the expected value E(*X*) of a random variable.
  3. the mean (μ) of a probability distribution.

- **Range**: measure of dispersion about mean (maximum minus minimum). when max and min are unusual values, range may be a misleading measure of dispersion;
- **Deviation**: the difference between the observed value and the mean:

$$d_i = x_i - \mu$$



More Statistical Terms

- **Standard Deviation**: standard deviation shows how much variation or "dispersion" exists from the average:

$$\sigma = \left[\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \right]^{\frac{1}{2}}$$

- **Variance**: variance is a measure of how far a set of numbers is spread out:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

- But we don’t really know the value of μ …For a series of N measurements, the most probable estimate of mean μ is the average. We refer this as the **sample mean** to distinguish it from the true mean μ:

$$\mu \approx x_m = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- Our best estimate of the standard deviation would be from:

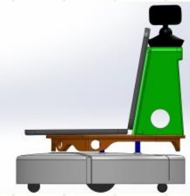
Sample Standard Deviation:

$$\sigma \approx \left[\frac{1}{N-1} \sum_{i=1}^{N} (x_i - x_m)^2 \right]^{\frac{1}{2}}$$

- Our best estimate of the variance would be from:

Sample Variance:

$$\sigma^2 \approx \frac{1}{N-1} \sum_{i=1}^{N} (x_i - x_m)^2$$

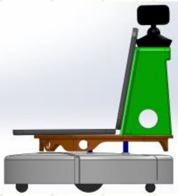


CEP and RMS

- The Root Mean Square (RMS) and Circular Error Probable (CEP) are commonly used to measure sensor errors (e.g. GPS error);
- CEP is defined as the radius of a circle centered on the true value that contains 50% of the actual measurements;
- So a GPS receiver with 1 meter CEP accuracy will have outputs within one meter of the true position 50% of the time;
- RMS is defined as the square root of the mean of the squares of the values:

$$x_{rms} = \sqrt{\frac{1}{n}(x_1^2 + x_2^2 + \ldots + x_n^2)}$$

- A GPS receiver with 1 meter RMS accuracy will be within one meter of the true measurement ~65% of the time;
- Variability of the measurement error as indicated by the standard deviation is *around the mean instead of zero*. Therefore, the RMS of the differences between measurements and the truth (if we know it) is a meaningful measure of the error.



Statistical Analysis of Data

- **Example #1**

Reading	X_i (m)	$d_i = x_i - x_m$	$(x_i - x_m)^2$
1	5.30	-0.313	0.098
2	5.73	0.117	0.014
3	6.77	1.157	1.339
4	5.26	-0.353	0.125
5	4.33	-1.283	1.646
6	5.45	-0.163	0.027
7	6.09	0.477	0.228
8	5.64	0.027	0.001
9	5.81	0.197	0.039
10	5.75	0.137	0.019

Sample mean value

$$x_m = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{10}\cdot 56.13 = 5.613m$$

Biased standard deviation

$$\sigma_b = \left[\frac{1}{N}\sum_{i=1}^{N}\left(x_i - x_m\right)^2\right]^{\frac{1}{2}} = \sqrt{\frac{1}{10}\cdot 3.533} = 0.594m$$

Sample standard deviation

$$\sigma = \left[\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - x_m\right)^2\right]^{\frac{1}{2}} = \sqrt{\frac{1}{10-1}\cdot 3.533} = 0.627m$$

Range

$$[4.33,\ 6.77],\ \text{or},\ 6.77 - 4.33 = 2.44m$$

- MATLAB Functions: ‘mean’, ‘median’, ‘mode’, ‘min’, ‘max’, ‘var’, ‘std’, ‘rms’,...

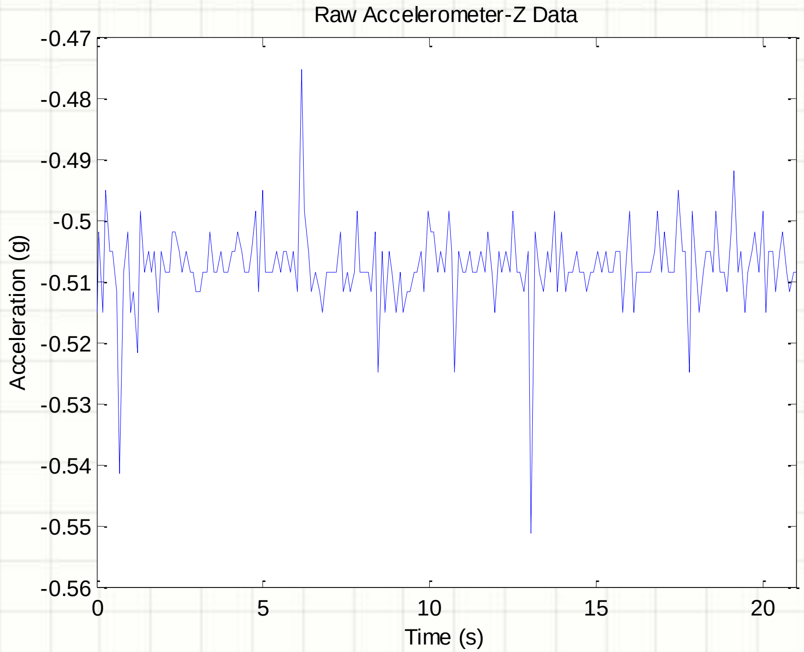
Statistical Analysis of Data (Cont.)

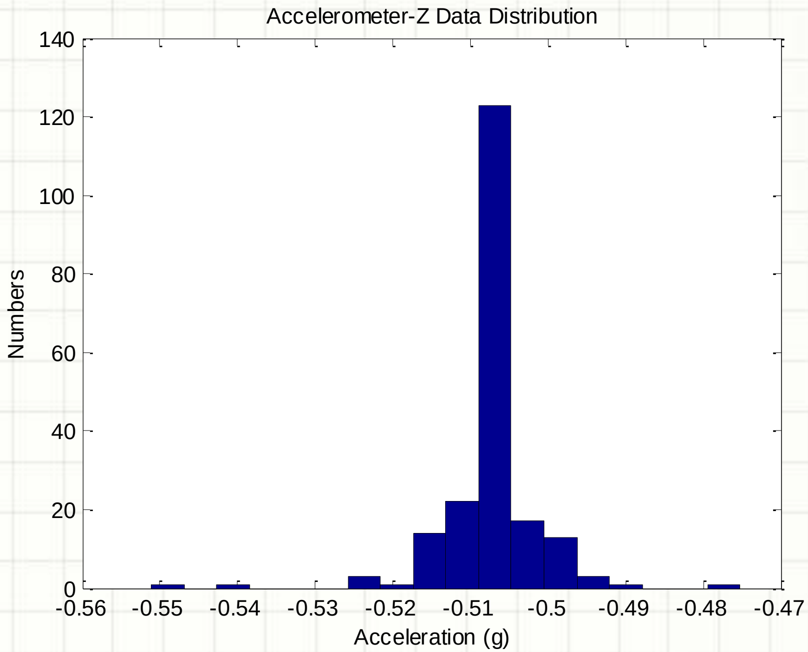
- **Example #2:** Two hundred steps of z-axis accelerometer data:

  Mean = -0.5073g; Median = Mode = -0.5082g;

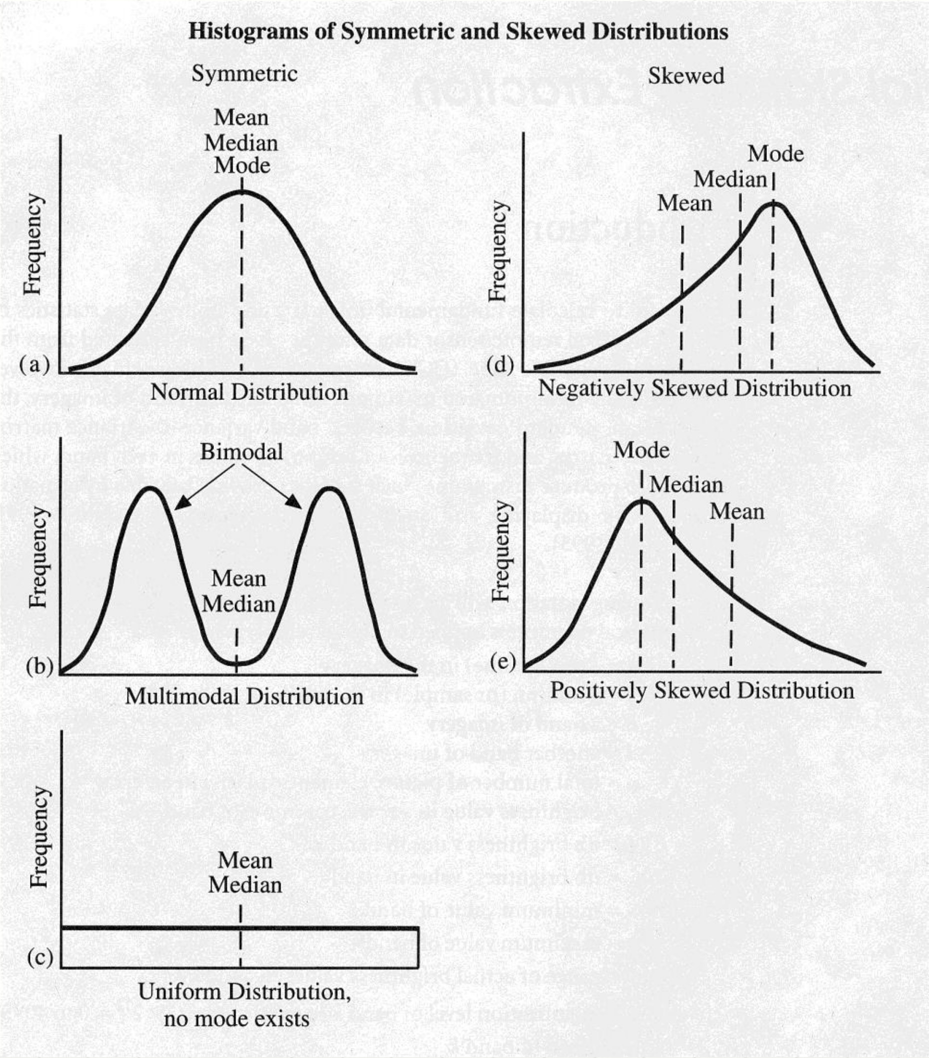
  Standard Deviation = 0.0067g; Variance = 4.5279e-05 g^2

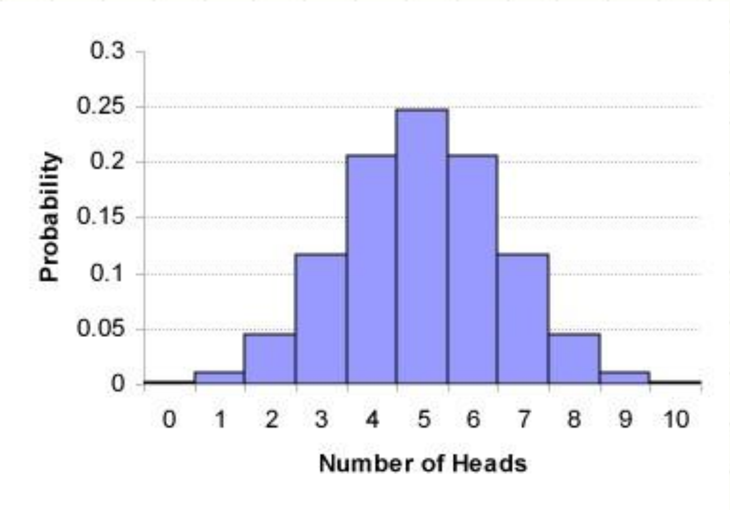
  Min = -0.5511 g; Max = - 0.4752g.





Probability Distributions





Flip a fair coin 10 times

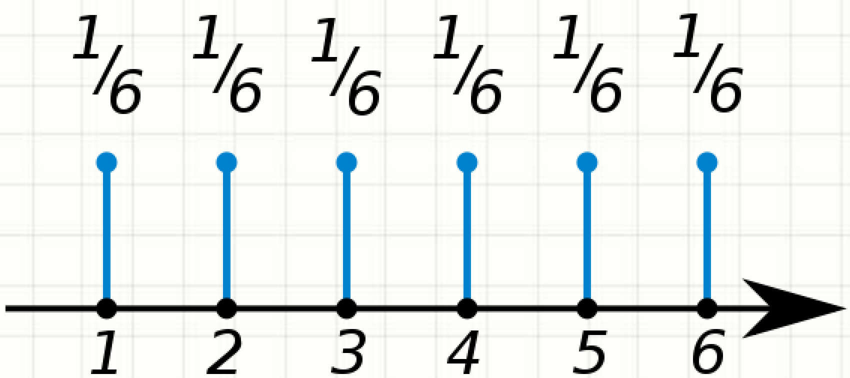
(Discrete Distribution)

from: Jensen, 1996

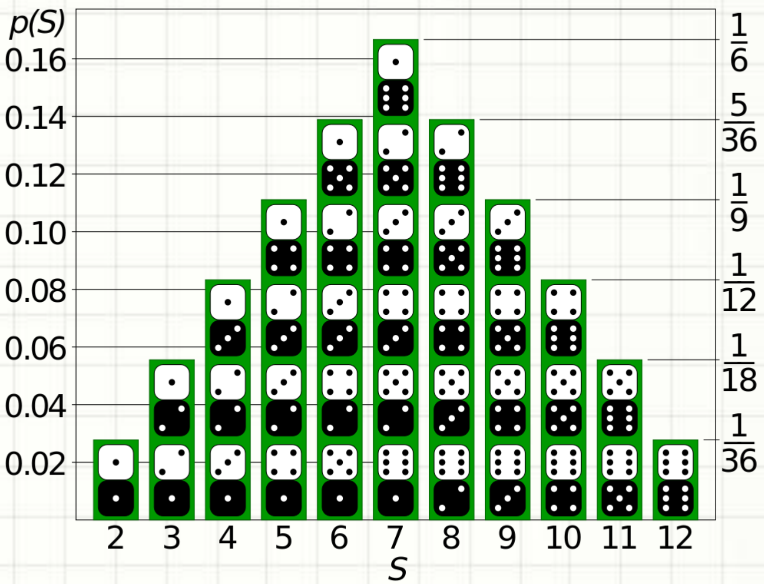
Probability Mass Function

- A discrete random variable X is defined as a random variable that assumes values from a countable set, that is $S = \{x_1, x_2, x_3, \ldots\}$
- The probability mass function (pmf) describes the relative chance for a random variable to take on a given value. It is defined as: $f_X(x) = \Pr(X = x)$
- The total probability for all hypothetical outcomes x: $\sum_{x \in S} f_X(x) = 1$
- For a fair dice:

$$f_X(x) = \begin{cases} 1/6, & x \in \{1,2,3,4,5,6\} \\ 0, & x \notin \{1,2,3,4,5,6\} \end{cases}$$



PMF for a Fair Dice



PMF for Two Dices

Probability Density Function

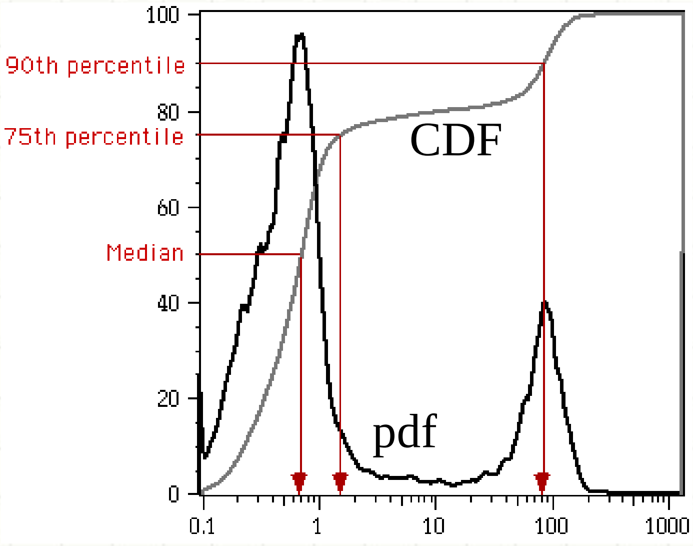
- A continues random variable may assume any numerical value in an interval or collection of intervals.
- The probability density function (pdf) of a continues random variable is the counterpart of pmf for a discrete random variable.
- A random variable X has density f_X, where f_X is a non-negative if:

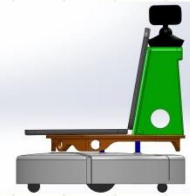
$$\Pr[a \le x \le b] = \int_a^b f_X(x)dx$$

- The cumulative distribution function (CDF) of X, F_X, is defined as:

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du, \quad f_X(x) = \frac{d}{dx} F_X(x)$$

- Intuitively, one can think of $f_X(x)dx$ as being the probability of X falling within the infinitesimal interval $[x, x + dx]$.



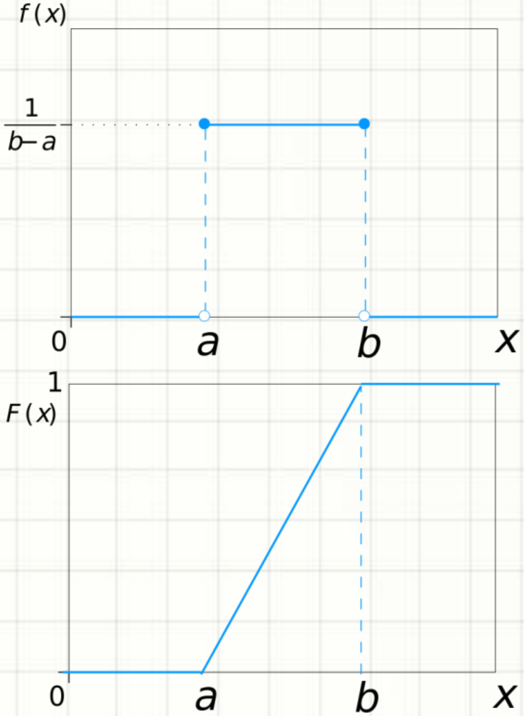


Uniform Distribution

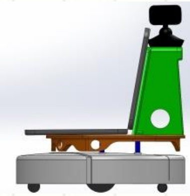
- Continuous uniform distribution is a family of symmetric probability distributions such that for each member of the family, all intervals of the same length on the distribution's support are equally probable;
- The support is defined by the two parameters, *a* and *b*, which are its minimum and maximum values.
- The pdf and CDF for a uniformly distributed random variable between [*a*, *b*] are:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \dfrac{x-a}{b-a} & \text{for } a \le x < b \\ 1 & \text{for } x \ge b \end{cases}$$



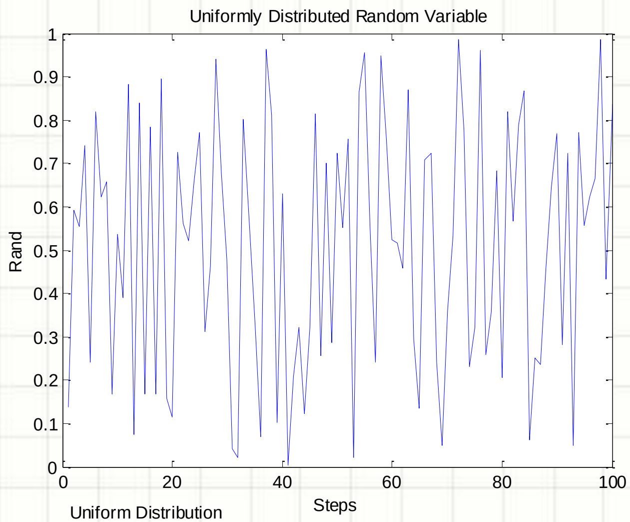
- The *quantization noise* can be approximated by the uniform distribution.

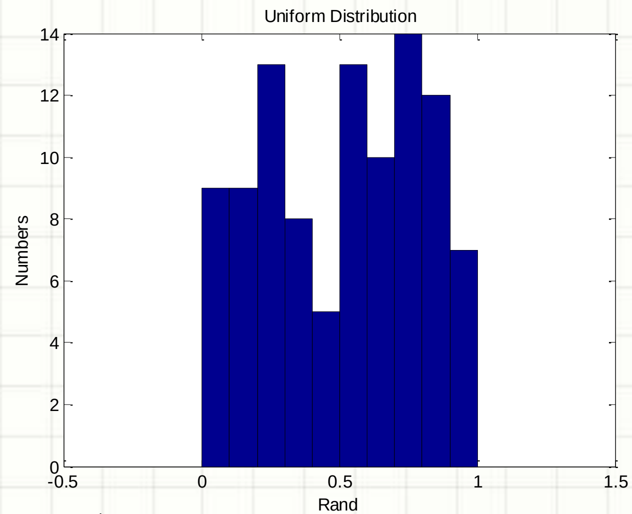


Uniform Distribution Example

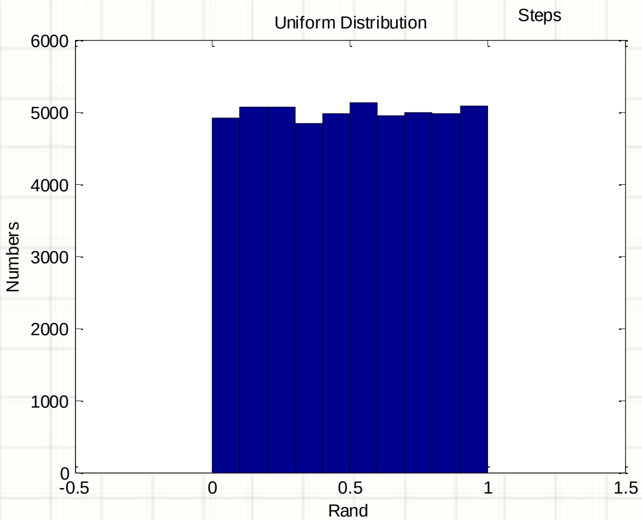
- The MATLAB ‘rand’ function generates uniformly distributed pseudorandom numbers:

1,000 Numbers:

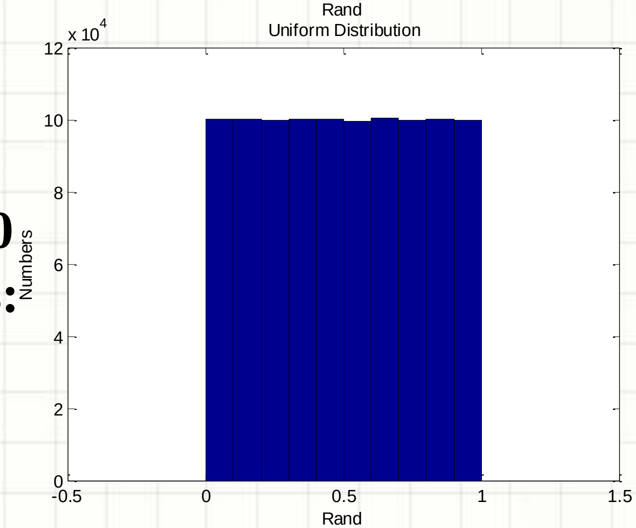


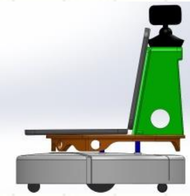


50,000 Numbers:



1,000,000 Numbers:

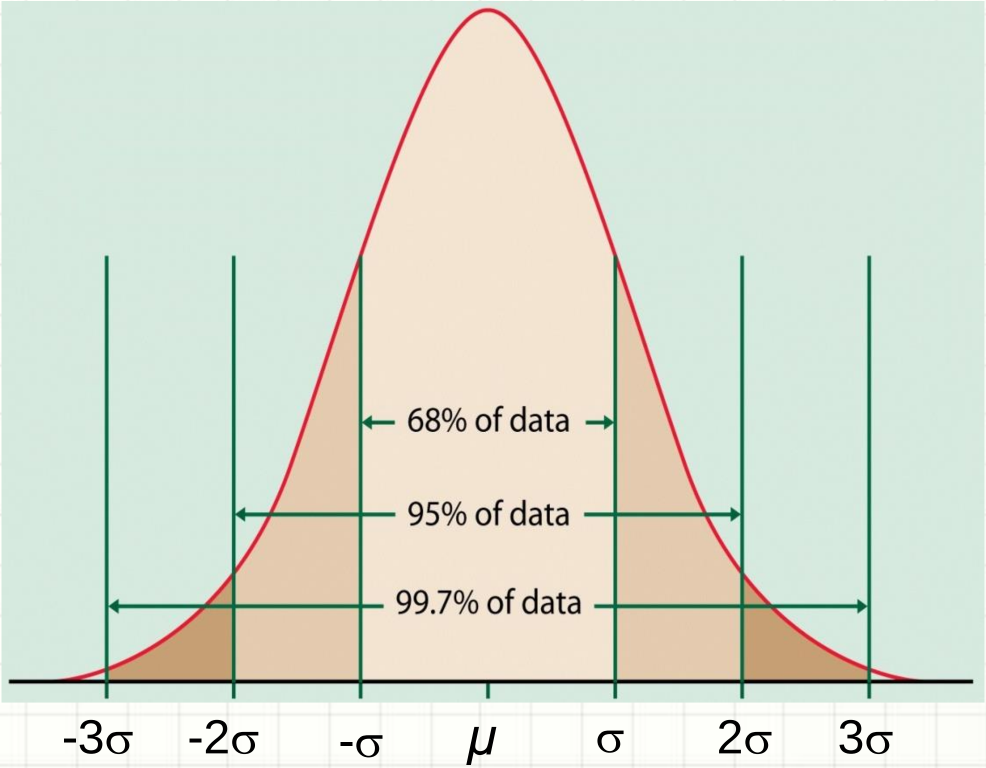




Normal/Gaussian Distribution

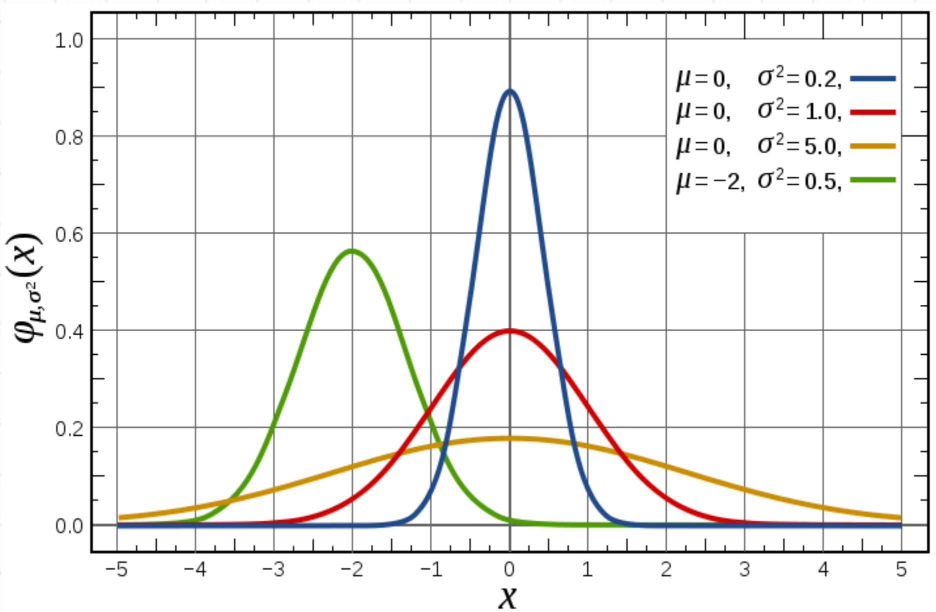
- The Gaussian distribution is sometimes informally called the bell curve.
- The pdf for the Gaussian distribution is: $f\left(x\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- If the mean μ = 0 and the standard deviation σ = 1, the distribution is called the standard normal distribution
or the unit normal distribution.
- MATLAB function:
R = normrnd(MU,SIGMA)

Deviation	Probability
0.6745σ	0.5000
σ	0.6827
2σ	0.9545
3σ	0.9973

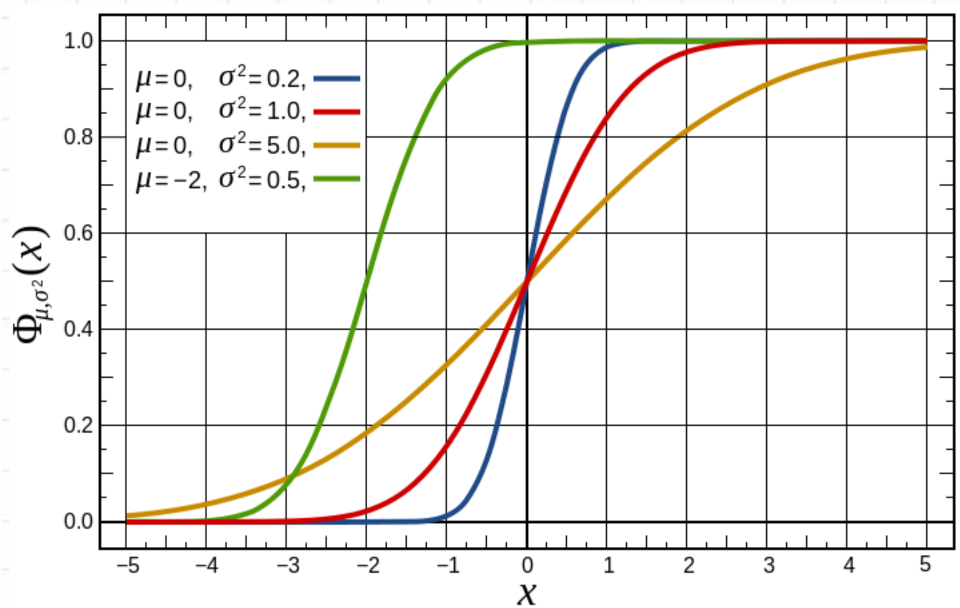


Gaussian Distribution (Cont.)

- Gaussian distribution is widely used in engineering, especially robotics;
- It is also the most important type of distribution in measurement;
- It is mathematically simple: the distribution is fully captured by two parameters – mean (the best guess) and variance (*confidence level* about the guess).



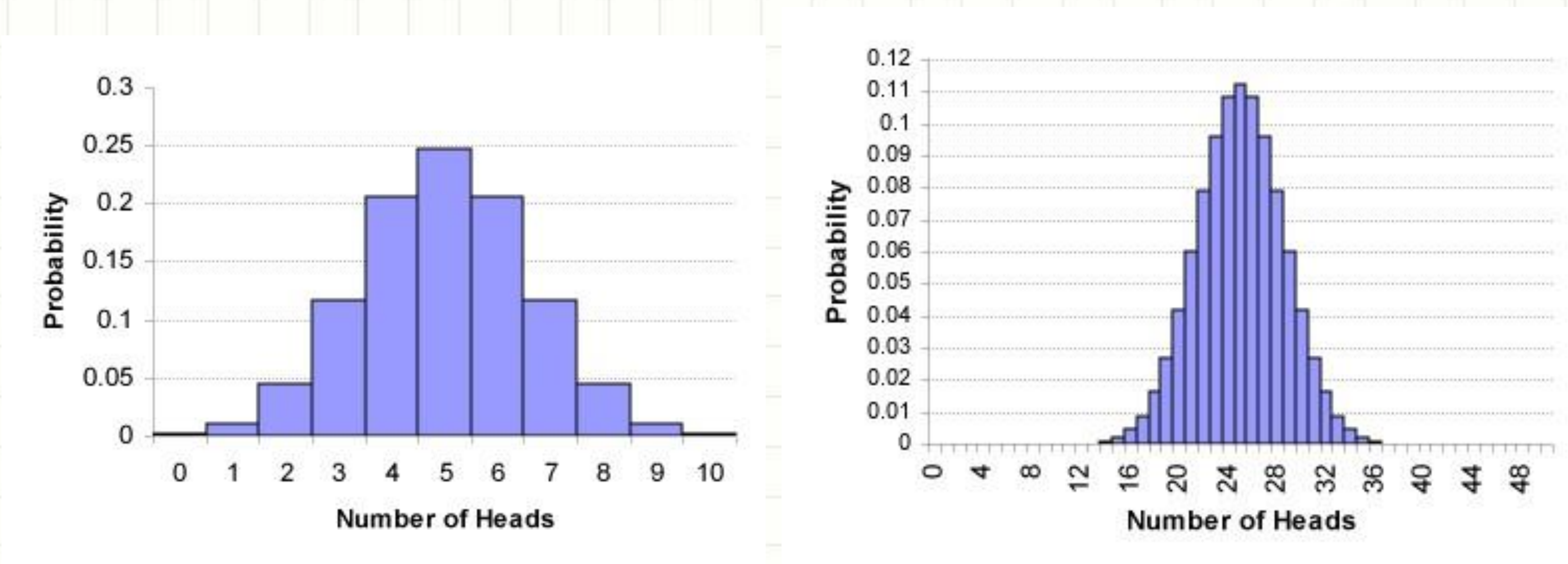
pdf



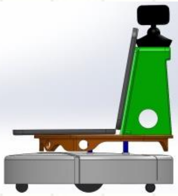
CDF

Central Limit Theorem

- In probability theory, the **Central Limit Theorem** (CLT) states that, given certain conditions, the mean of a sufficiently large number of independent random variables (no matter what's the distribution may look like), each with finite mean and variance, will be approximately **normally distributed**!
- This explains why a lot of things are normally distributed in nature (lifetime of light bulbs, years of education, weight of rocks, ...)

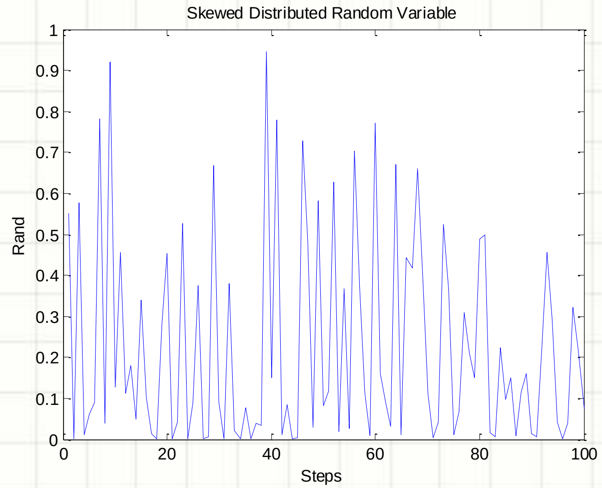


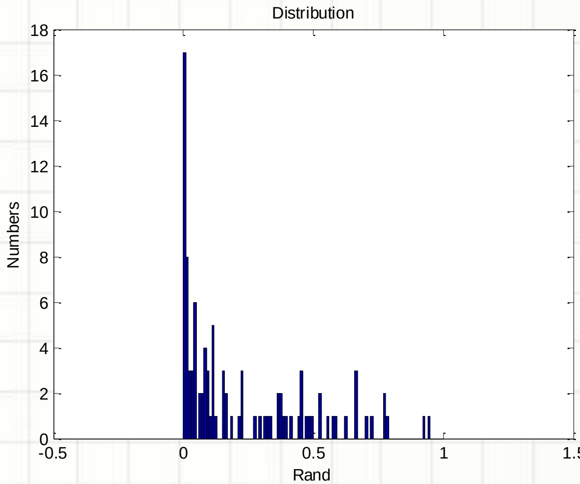
Flip a fair coin 10 and 50 times



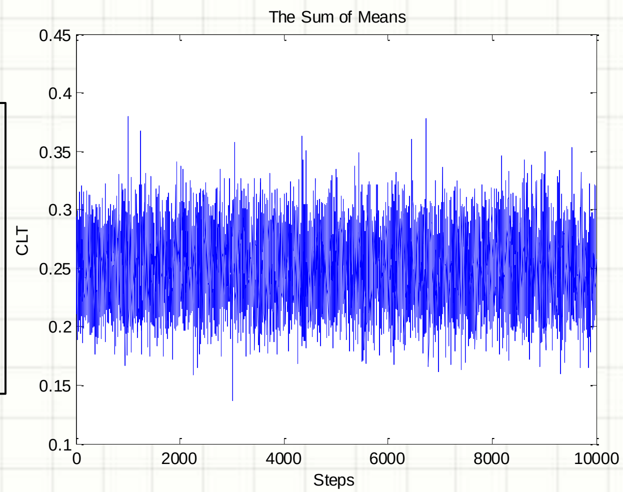
Central Limit Theorem Example

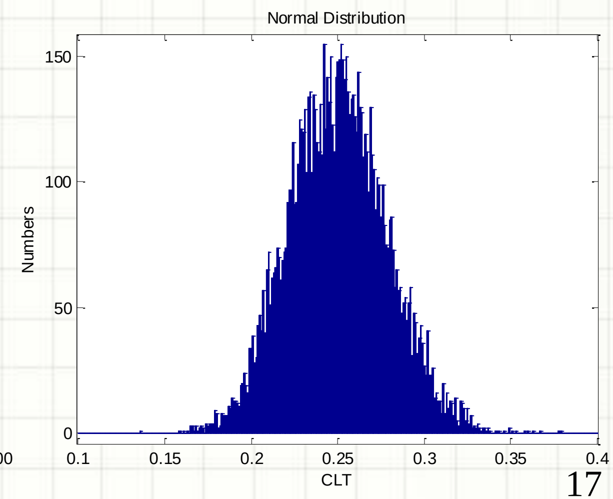
- Let's try to use rand^3 to create a random variable *X* with a skewed distribution. First plot the distribution with 100 values:
- Repeat the process 10,000 times and create a new random variable *Y* that is defined as the mean of *X* (every 100 values).
- What is the distribution of *Y*?

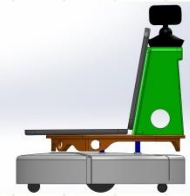





```
for j=1:10000
    for i=1:100
        X(i)=rand^3;
    end
    Y(j)=mean(X);
end
```

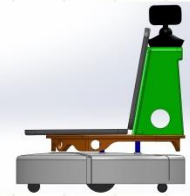






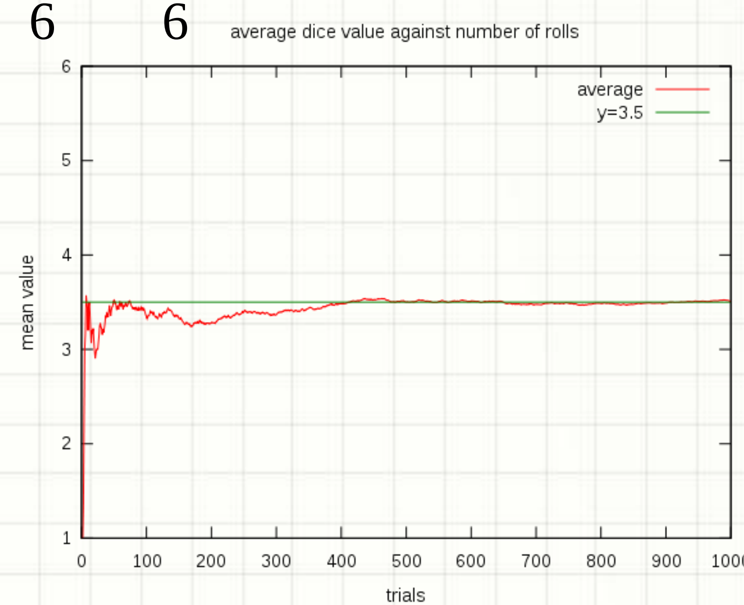
Measurement Noise

- Why measurement noises are mostly normally distributed?
  - Because the measurement noise is the average effect of a large number of random underlining contribution factors such as quantization, vibration, electromagnetic interference, etc...
  - This is true for most of the time but not always! A safer bet is to perform a statistical analysis of the data and see it yourself!
  - There are technics to handle those non-Gaussian noise distributions (but is beyond the scope of this class...).



Expected Value

- The expected value is a weighted average of all possible values;
- Suppose a discrete random variable *X* can take value x_1 with probability p_1, value x_2 with probability p_2, and so on, up to value x_k with probability p_k. Then the expectation of this random variable *X* is defined as: $$E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_k p_k = \sum_{i=1}^{k} x_i p_i$$
- For example, the expected value for throwing a fair dice is: $$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$
- If you roll the dice enough times, the average value will approach 3.5;
- The expected value helps us to find *not random* property of a random variable (think about casino!), thus can help us (and robots) to make decisions (not to gamble!).



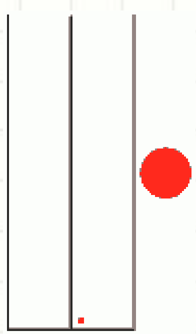
Expected Value (Cont.)

- For a continues random variable *X*, the expected value can be calculated using the pdf of the random variable:

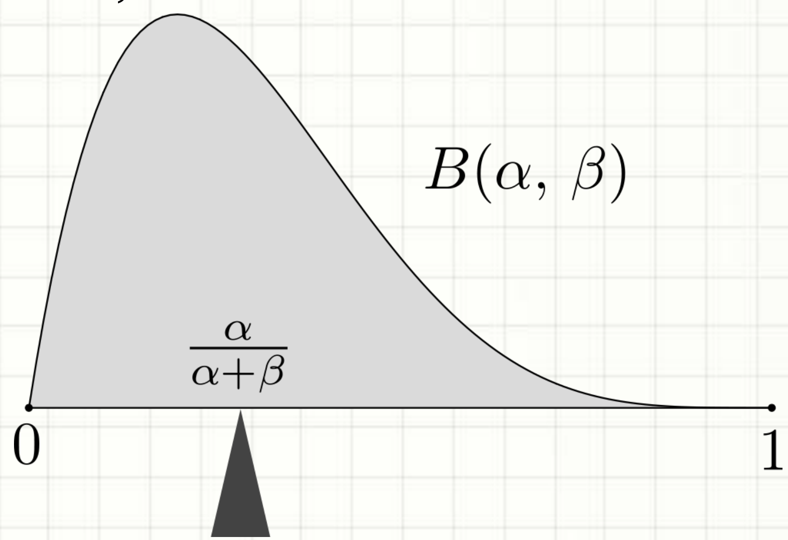
$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

- Law of large numbers: the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed;

- We often use the expected value for prediction in robotics.



Simulation of Coin Flips to Show the Law of Large Numbers



The mass of probability distribution is balanced at the expected value

Expected Value and Variance

- The expected value of a constant is equal to the constant itself, E[c] = c;
- The expected value operator is linear:

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

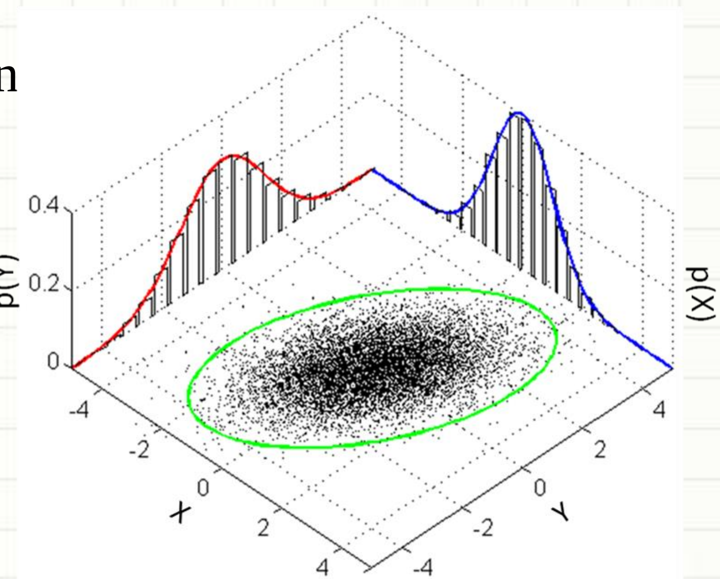
- For a discrete random variable with a probability mass function: $x_1 \rightarrow p_1, \ldots x_n \rightarrow p_n$, the variance: $Var(X) = \sum_{i=1}^{n} p_i (x_i - \mu)^2 = E\left[(X-\mu)^2\right]$ where $\mu = E[X]$
- For a continues random variable with a pdf *f*(*x*), the variance:

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = E\left[(X-\mu)^2\right]$$

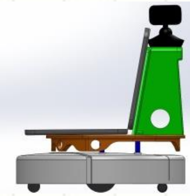
- Therefore: $Var(X) = E\left[(X-\mu)^2\right] = E\left[X^2 - 2XE[X] + E[X]^2\right] = E\left[X^2\right] - E[X]^2$
- In other words: “mean of square minus square of mean”.

Multivariate Random Variable

- Many times, we are dealing with events that involves multiple variables. For example, the 2D position of a robot can be described by two variables (X, Y) and each of them could be random;
- A multivariate random variable or random vector $\mathbf{X} = (X_1, \ldots, X_n)^T$ is a list of mathematical variables each of whose value is unknown, either because the value has not yet occurred or because there is imperfect knowledge of its value;
- Multivariate distribution, is a generalization of the one-dimensional distribution to higher dimensions;
- The expected value or mean of a random vector **X** is a fixed vector E(**X**) whose elements are the expected values of the respective random variables.
- But what about the variance?



A multivariate Gaussian distribution



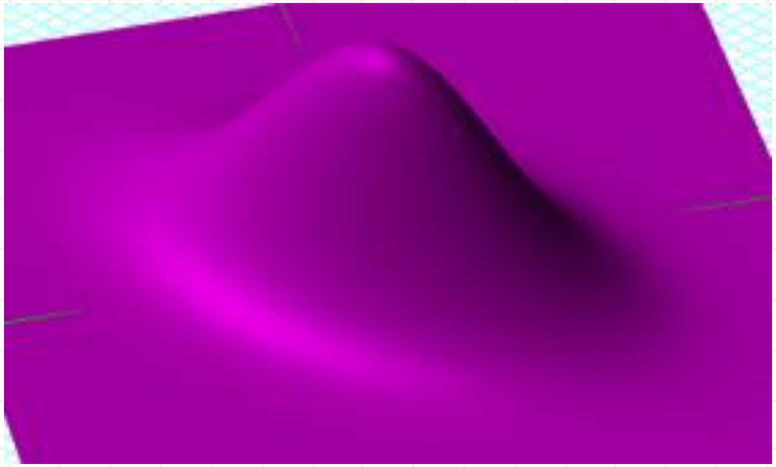
Covariance

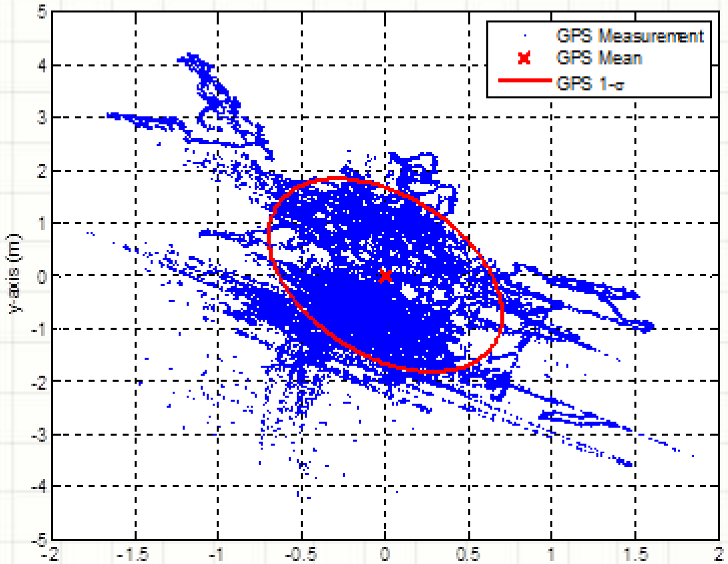
- Recall that the variance for a random variable is: $Var(X) = \sigma^2(X) = E\left[(X-\mu)^2\right]$
- Covariance is an generalization of the concept of variance to a random vector;
- Covariance is a measure of how much two random variables change together;
- The covariance between two jointly distributed random variables *X* and *Y* is defined as: $\sigma(X,Y) = E\left[(X - E[X])(Y - E[Y])\right]$
- Variance is a special case of the covariance when the two variables are identical: $\sigma(X,X) = \sigma^2(X)$
- The covariance matrix of an $n\times 1$ random vector $\mathbf{X}$ is an $n\times n$ matrix whose *i*, *j* element is the covariance between the i^{th} and the j^{th} random variables. The covariance matrix is the expected value, element by element, of the $n\times n$ matrix computed as:

$$\mathrm{cov}(\mathbf{X}) = E\left[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T\right]$$

Covariance Matrix

- For example, if we have a position measurement device that reports the x and y position of a robot. If the measurements are distributed with a covariance matrix of $\begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}$, it will have a distribution like in the figure below (left);
- This distribution is somewhat like the 5.5 hours of GPS measurement of the fire hydrant that we saw earlier in the class.



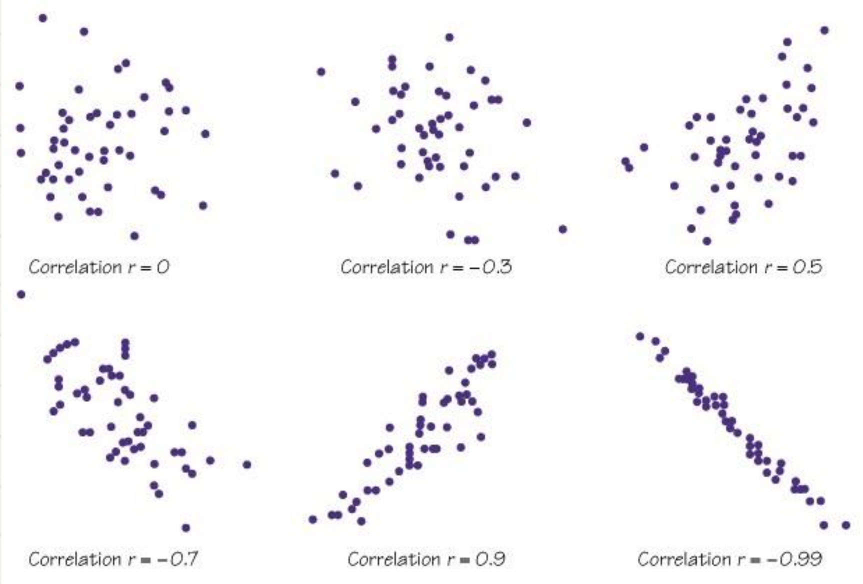


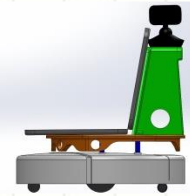
Correlation

- <u>Correlation</u> is related to the concept of <u>dependence</u> (or <u>independence</u>);
- <u>Pearson correlation coefficient</u> between two random variables *X* and *Y* is defined as:

$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sigma_X \sigma_Y}$$

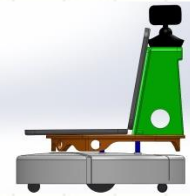
- It has a value of 1 for positive correlation, -1 for negative correlation, and 0 for no correlation (independent);
- If the none diagonal terms in the covariance matrix are all zero, then all the components of the random vector are mutually independent or uncorrelated;
- **Correlation does not imply causation!**





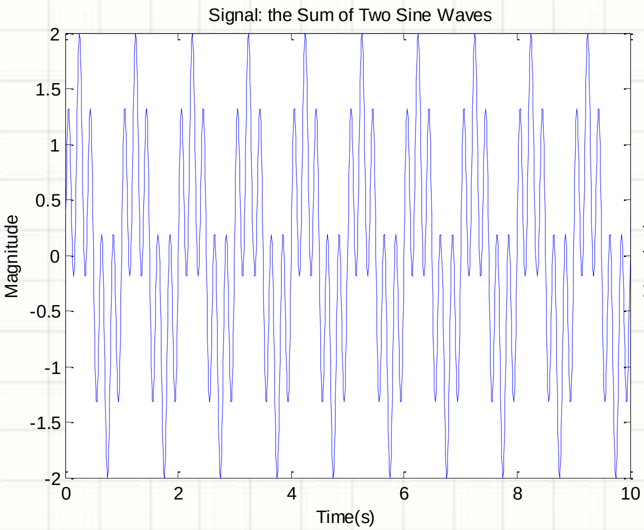
Autocorrelation and PSD

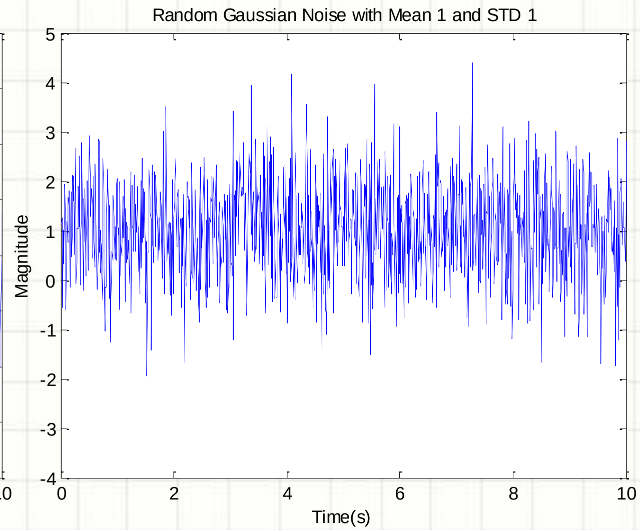
- Autocorrelation is the cross-correlation of a signal with itself.
- If X_t is a <u>stationary process</u> then the mean μ and the variance σ^2 are time-independent, the autocorrelation is: $$R(\tau) = \frac{E\left[(X_t - \mu)(X_{t+\tau} - \mu)\right]}{\sigma^2}$$
- It can be used to find frequency components in a signal;
- The Power Spectral Density (PSD) is given by the Fourier transform (FT) of its autocorrelation function (we only need to know the concept instead of mathematical details in this class);
- It describes how the power of a signal or time series is distributed over the different frequencies;
- The MATLAB functions for calculating the covariance, correlation, autocorrelation, and PSD are 'cov', 'corrcoef', 'autocorr', and 'pwelch' respectively.

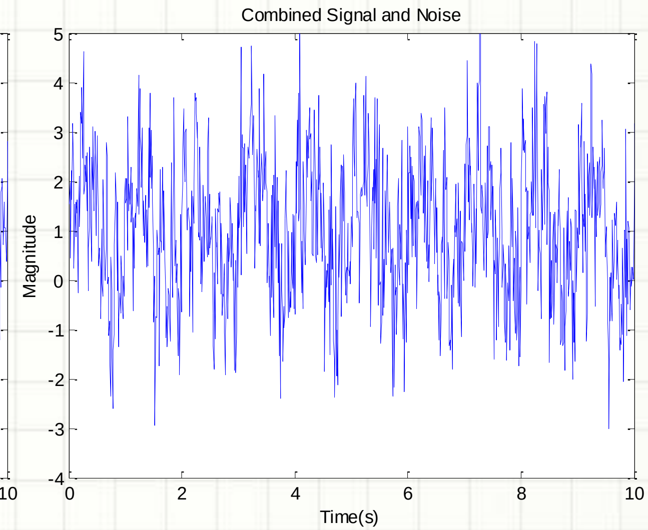


PSD Example

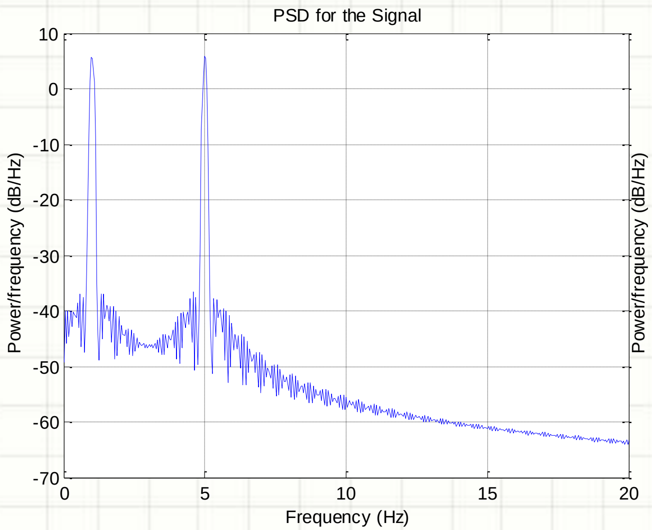
Signal: 1Hz and 5Hz Sine Wave + Gaussian Noise = Can you see the signal?

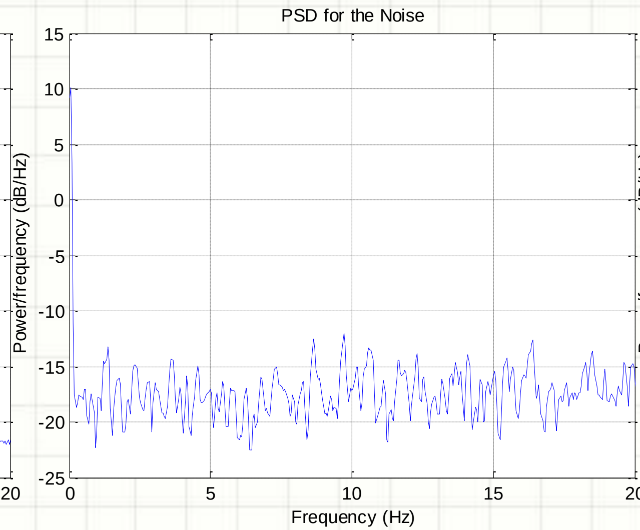


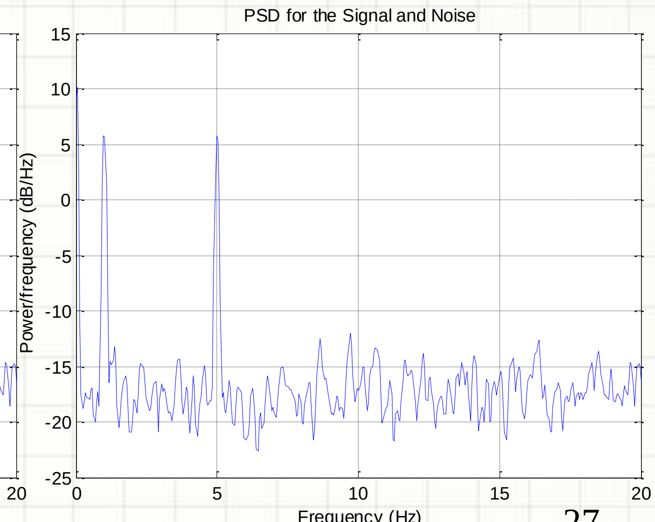




The PSDs:

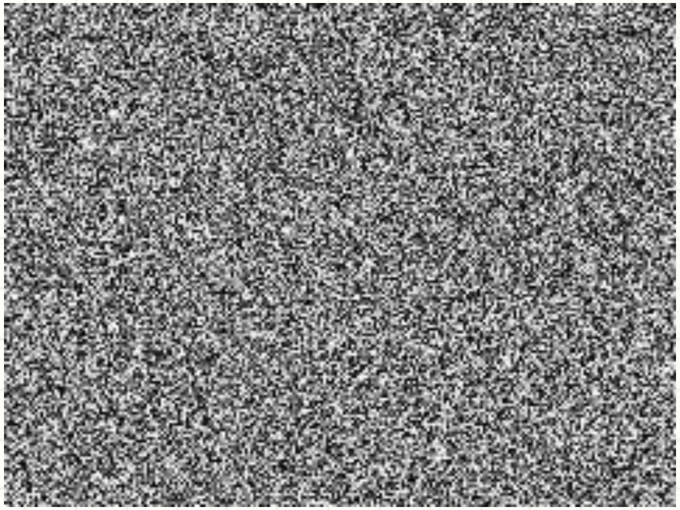


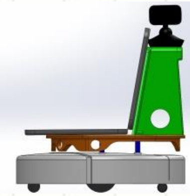




White Noise

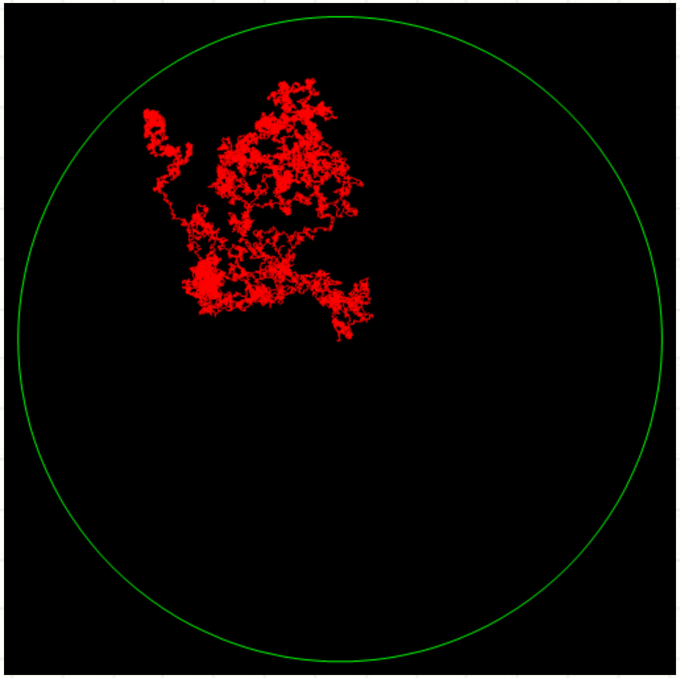
- White noise is a random signal with a flat (constant) PSD. In other words, a signal that contains equal power within any frequency band with a fixed width;
- White noise draws its name from white light, which is commonly (but incorrectly) assumed to have a flat spectral power density over the visible band;
- Gaussian white noise is a good approximation of many real-world situations and generates mathematically tractable models;
- However, Gaussian does not imply white, nor is the other way around;
- Being uncorrelated in time does not restrict the values a signal can take.



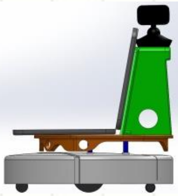


Random Walk

- A *random walk* is a mathematical formalization of a path that consists of a succession of random steps. For example, the path traced by a molecule as it travels in a liquid or a gas;
- Imagining that you are really bored (I hope you are not!) and decide to let two coins to direct you. You will walk to left if one of the coin is head and right if tail. You will walk forward if another coin is head and backward if tail. Repeat this process many times and you will be walking in a trajectory like the figure on the right;



- As silly as it might look, random walk is a fundamental process that captures the dynamics of many events such as the price of a stock.

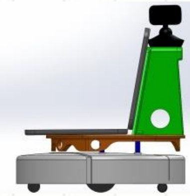


Random Walk in Navigation

- For us, random walk is directly related to inertial navigation;
- In inertial navigation, the position is estimated by integrating the accelerometer data twice;
- As we discussed earlier, the sensor error can mostly be modeled as *Gaussian white noise*. So what would be the velocity and position error look like if this assumption is true?

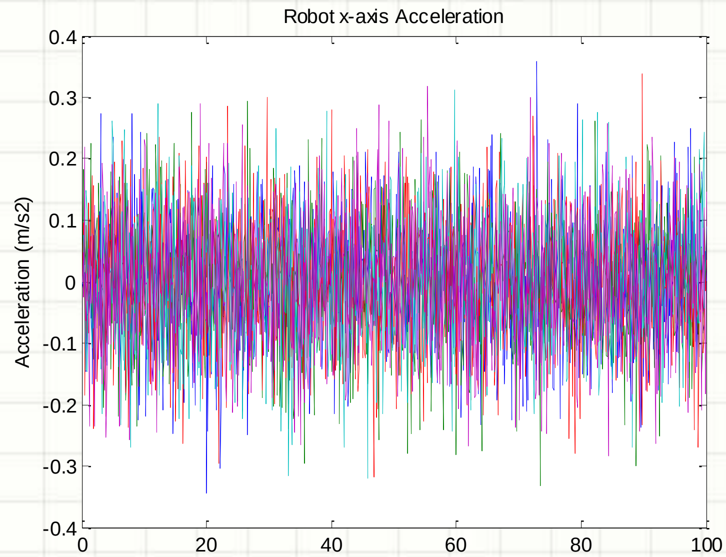
```
N=1000;      % Total number of data
Ts=0.1;     % The sampling time
K=5;         % The number of runs
Ax=zeros(K,N);
Vx=zeros(K,N);
Px=zeros(K,N);
T=Ts:Ts:(N*Ts); % The time
for k=1:K
    for i=2:N
        Ax(k,i)=Ax(k,i)+normrnd(0,0.1);
        Vx(k,i)=Vx(k,i-1)+Ax(k,i)*Ts;
        Px(k,i)=Px(k,i-1)+Vx(k,i)*Ts;
    end
end
```

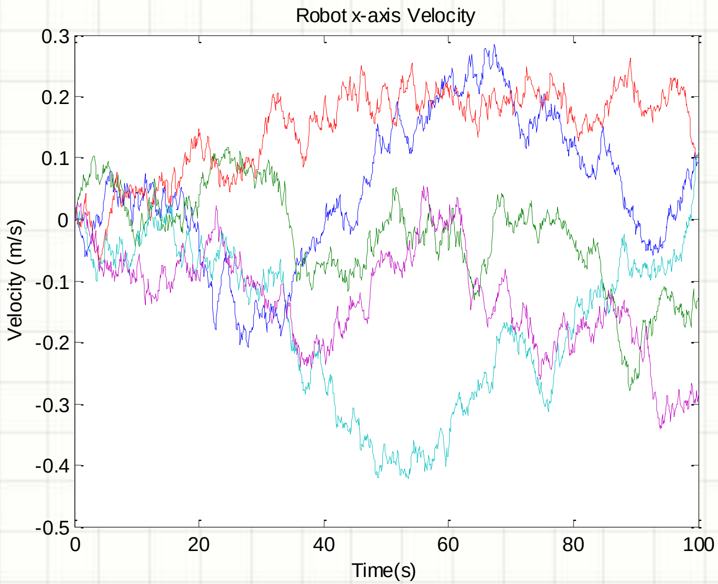
- We can find it out with a MATLAB simulation. We assume the accelerometer noise is zero mean and with a standard deviation of $0.1 m/s^2$;
- Sounds pretty harmless, right?

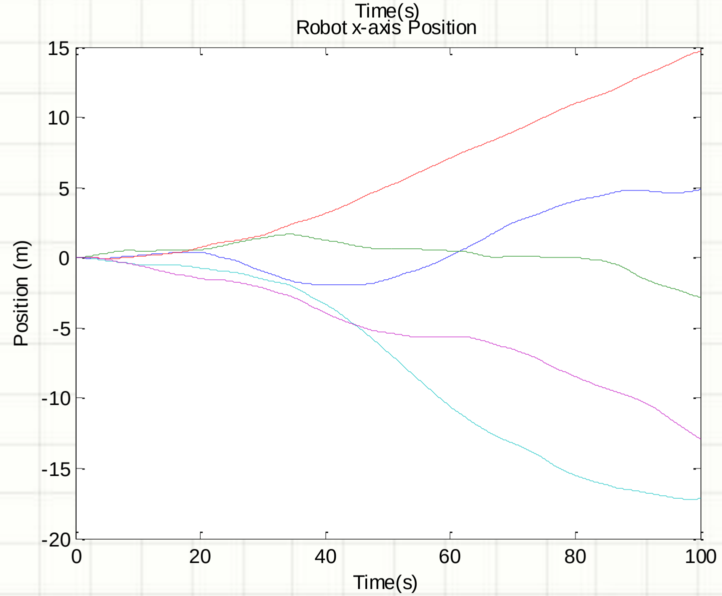


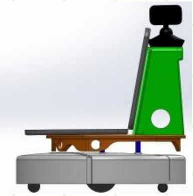
Navigation with Random Walk

- Let’s run it five times with the same assumptions and with no motion;
- The <u>velocity random walk</u> for accelerometers and <u>angular random walk</u> for rate gyroscopes are among the most important specs for inertial sensors.



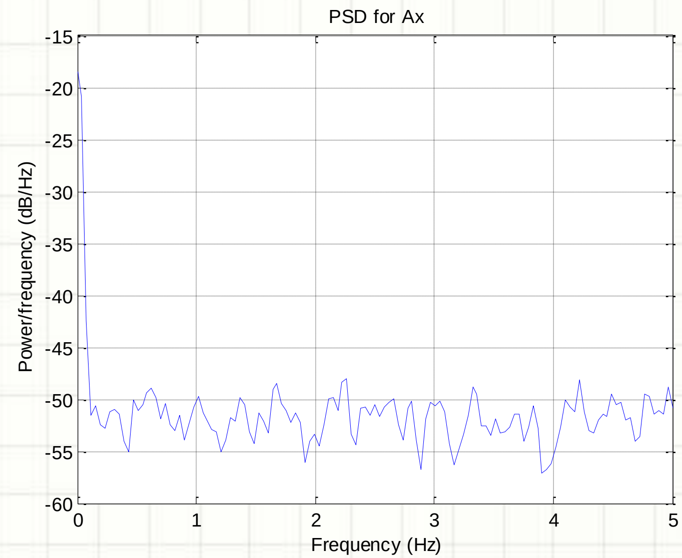


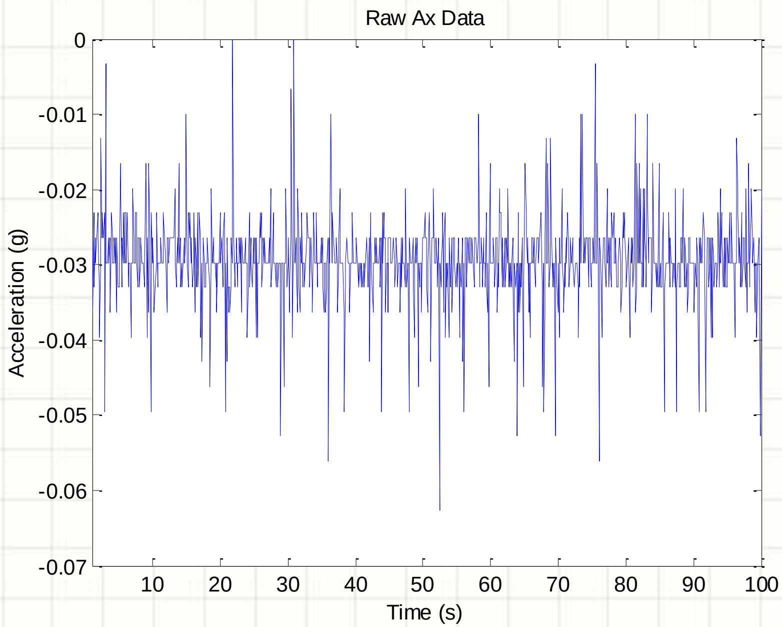


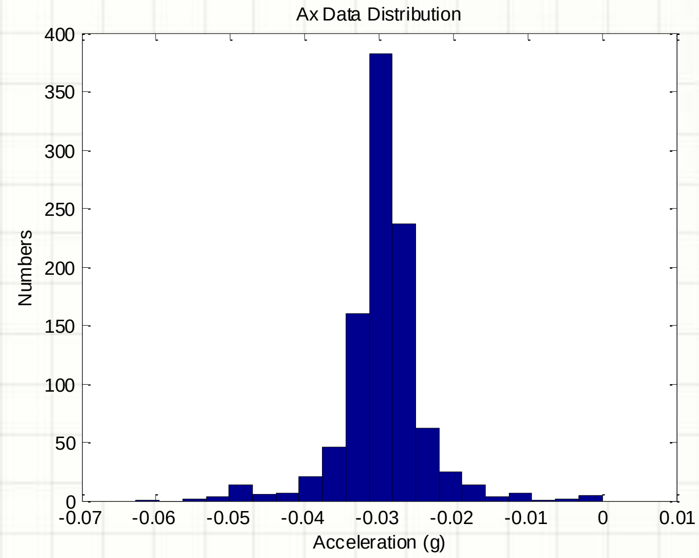


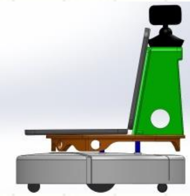
Analyzing of Sensor Noise

- 100 Seconds of accelerometer data at 10Hz under static conditions were collected and analyzed;
- Looks like the noise on each individual channel is pretty close to white and Gaussian!







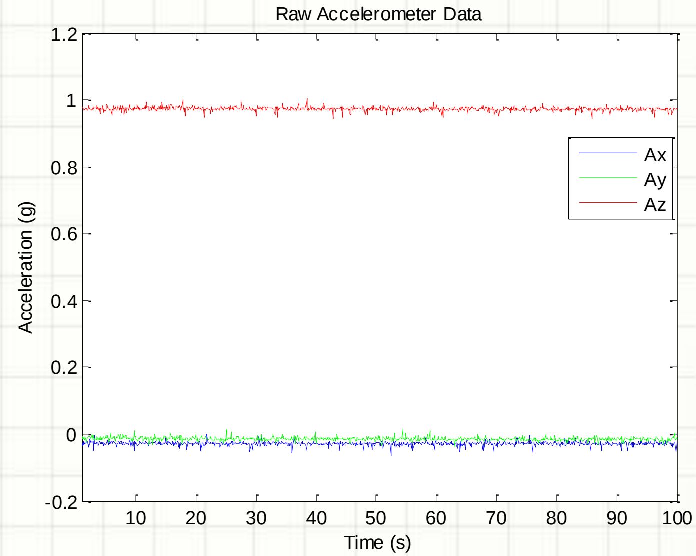


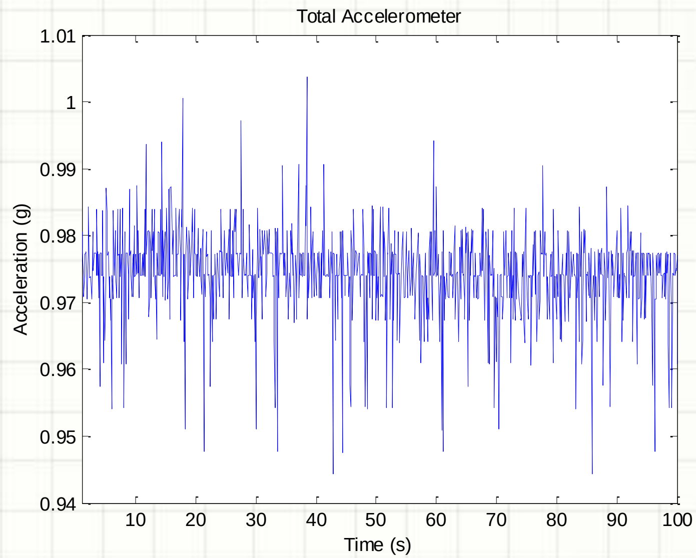
Looking at All Three Channels

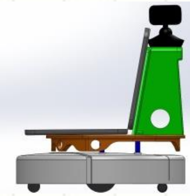
- Total acceleration looks a little less than 1g;
- The covariance matrix:

$$Cov(Ax, Ay, Az) = 1.0e-04 \times \begin{bmatrix} 0.3665 & -0.0048 & -0.0465 \\ -0.0048 & 0.3143 & 0.0181 \\ -0.0465 & 0.0181 & 0.4131 \end{bmatrix}$$

- The 3-axis measurements are reasonably independent from each other!
- The Az accelerometer is the worst!
- There is more coupling between Ax-Az than between Ax-Ay.







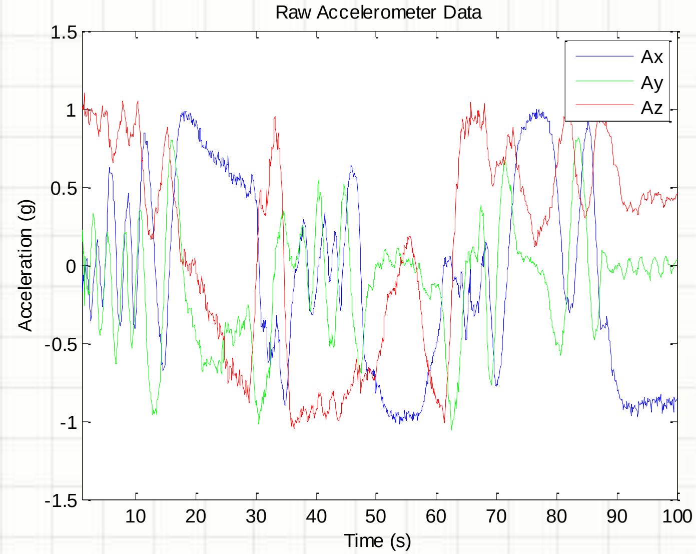
Accelerometers in Motion

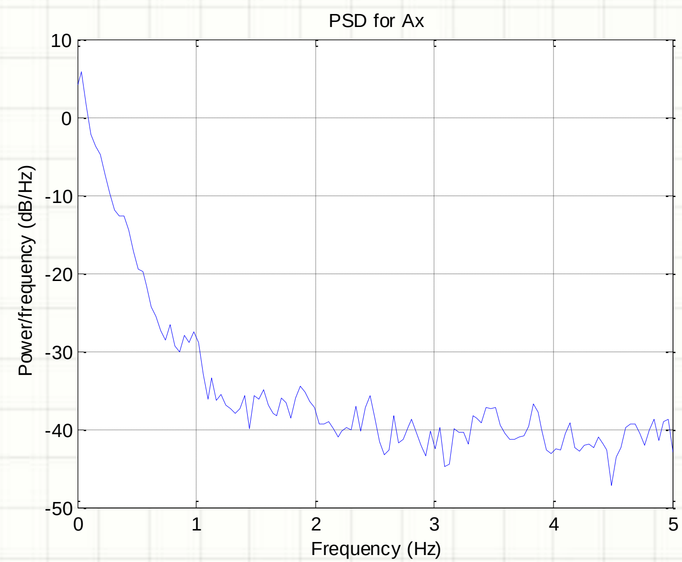
- The covariance matrix:

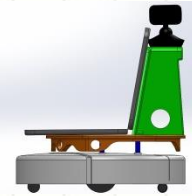
$$Cov(Ax, Ay, Az) = \begin{bmatrix} 0.3816 & -0.0378 & -0.0330 \\ -0.0378 & 0.1388 & 0.0433 \\ -0.0330 & 0.0433 & 0.4221 \end{bmatrix}$$

- Can you explain what happened?

(Since the signal and noise are mixed now, the covariance matrix no longer only contain information about the measurement error. To quantify the sensor performance, we need to know the truth – to some degree...)







Summary

- Let's conclude the section by asking a few questions:
  1. How can a random variable be described mathematically?
  2. What is the difference between mean and expected value?
  3. How do we know if two random variables are independent?
  4. How do we isolate low frequency signals from the noise?
  5. Why does the output of an Inertial Navigation System (INS) drift when all the measurement errors are zero mean?
  6. Why most measurement errors are Gaussian? And why do we consider ourselves lucky here?
  7. Why does it make sense to average multiple measurements?
  8. What if the measured parameter is changing itself?

Further Reading

- Probability Vs. Stastistics. http://www.cs.sunysb.edu/~skiena/jaialai/excerpts/node12.html
- Why are normal distributions normal? http://aidanlyon.com/sites/default/files/Lyon-normal_distributions.pdf
- Search Wikipedia for keywords: “Probability Density Function”, “Probability Mass Function”, “Gaussian Distribution”, “Expected Value”, “Multivariate Random Variable”, “Covariance”, “Correlation”, “Power Spectral Density”, “White Noise”, “Random Walk”, etc.