Probability and Statistics

Robert Stengel
Robotics and Intelligent Systems MAE 345
Princeton University, 2017

Learning Objectives

- Concepts and reality
  - Interpretations of probability
  - Measures of probability
- Scalar uniform, Gaussian, and non-Gaussian distributions
  - Probability density and mass functions
  - Expected values
- Bayes's Law
- Central Limit Theorem
- Propagation of the state's probability distribution

Copyright 2017 by Robert Stengel. All rights reserved. For educational use only.
http://www.princeton.edu/~stengel/MAE345.html

Probability

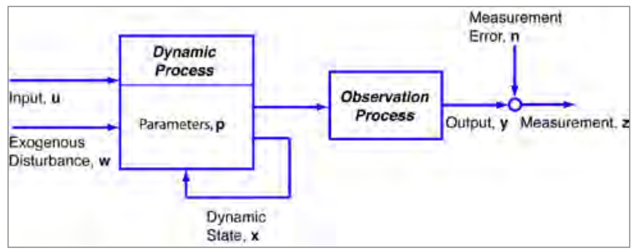
- ... a way of expressing knowledge or belief that an event will occur or has occurred

Statistics

- The science of making effective use of numerical data relating to groups of individuals or experiments

How Do Probability and Statistics Relate to Robotics and Intelligent Systems?

- **Decision-making under uncertainty**
- **Controlling random dynamic processes**



Concepts and Reality

(Papoulis)

- **Theory may be exact**
  - **Deals with averages of phenomena with many possible outcomes**
  - **Based on models of behavior**
- **Application can be only approximate**
  - **Measure of our state of knowledge or belief that something may or may not be true**
  - **Subjective assessment**

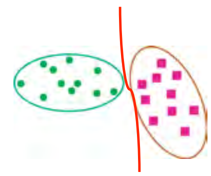
$$A : event$$
$$P(A) : probability\ of\ event$$
$$n_A : number\ of\ times\ A\ occurs\ experimentally$$
$$N : total\ number\ of\ trials$$
$$P(A) \approx \frac{n_A}{N}$$

Interpretations of Probability

(Papoulis)

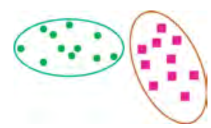
- **Axiomatic Definition (Theoretical interpretation)**
  - **Probability space, abstract objects (outcomes), and sets (events)**
  - **<u>Axiom 1</u>: Pr(A_i) ≥ 0**
  - **<u>Axiom 2</u>: Pr("certain event") = 1 = Pr [all events in probability space (or universe)]**
  - **<u>Axiom 3</u>: Independent events,**

$$\boxed{\Pr\left(A_i \text{ and } A_j\right) = \Pr\left(A_i \cap A_j\right) = \Pr\left(A_i\right)\Pr\left(A_j\right)}$$



  - **<u>Axiom 4</u>: Mutually exclusive events,**

$$\boxed{\Pr\left(A_i \text{ or } A_j\right) = \Pr\left(A_i \cup A_j\right) = \Pr\left(A_i\right) + \Pr\left(A_j\right)}$$



  - **<u>Axiom 5</u>: Non-mutually exclusive events,**

$$\boxed{\Pr\left(A_i \text{ or } A_j\right) = \Pr\left(A_i\right) + \Pr\left(A_j\right) - \Pr\left(A_i\right)\Pr\left(A_j\right)}$$

Interpretations of Probability

(Papoulis)

- **Relative Frequency (Empirical interpretation)**

$$\boxed{\Pr\left(A_i\right) = \lim_{N\to\infty}\left(\frac{n_{A_i}}{N}\right)}$$

N = number of trials (total)
n_{Ai} = number of trials with attribute A_i

- **Classical ("Favorable outcomes" interpretation)**

$$\boxed{\Pr\left(A_i\right) = \frac{n_{A_i}}{N}}$$

N is finite
n_{Ai} = number of outcomes "favorable to" A_i

- **Measure of belief (Subjective interpretation)**
  - **Pr(A_i) = <u>measure of belief</u> that A_i is true (similar to fuzzy sets)**
  - **Informal induction precedes deduction**
  - **Principle of insufficient reason (i.e., total prior ignorance):**
    - **e.g., if there are 5 event sets, A_i, i = 1 to 5, Pr(A_i) = 1/5 = 0.2**

Favorable Outcomes Example: Probability of Rolling a "7" with Two Dice

(Papoulis)

- **Proposition 1: 11 possible sums, one of which is 7**

$$\Pr\left(A_i\right) = \frac{n_{A_i}}{N} = \frac{1}{11}$$

- **Proposition 2: 21 possible pairs, not distinguishing between dice**
  - **3 pairs: 1-6, 2-5, 3-4**

$$\Pr\left(A_i\right) = \frac{n_{A_i}}{N} = \frac{3}{21}$$

- **Proposition 3: 36 possible outcomes, distinguishing between the two dice**
  - **6 pairs: 1-6, 2-5, 3-4, 6-1, 5-2, 4-3**

$$\Pr\left(A_i\right) = \frac{n_{A_i}}{N} = \frac{6}{36}$$

Propositions are knowable and precise; outcome of rolling the dice is not.

Steps in a Probabilistic Investigation

(Papoulis)

1) **Physical *(Observation)*: Determine probabilities, Pr(A_i), of various events, A_i, by experiment**
   - **Experiments cannot be exact**

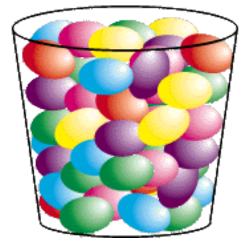
2) **Conceptual *(Induction)*: Assume that Pr(A_i) satisfies certain axioms and theorems, allowing deductions about other events, B_i, based on Pr(B_i)**
   - **Build a model**

3) **Physical *(Deduction)*: Make predictions of B_i based on Pr(B_i)**

Empirical (or Relative) Frequency of Discrete, Mutually Exclusive Events in Sample Space

$$\Pr(x_i) = \frac{n_i}{N} \quad \text{in } [0,1]; \quad i = 1 \text{ to } I$$

- ***N*** **= total number of events**
- $\boldsymbol{n_i}$ **= number of events with value** $\boldsymbol{x_i}$
- ***I*** **= number of different values**
- $\boldsymbol{x_i}$ **= ordered set of hypotheses or values**



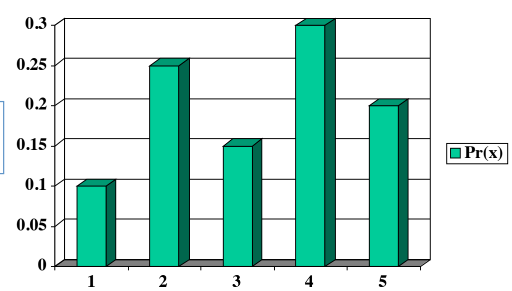
x **is a random variable**

Empirical (or Relative) Frequency of Discrete, Mutually Exclusive Events in Sample Space

- ***x*** **is a random variable**
- **Equivalent sets**

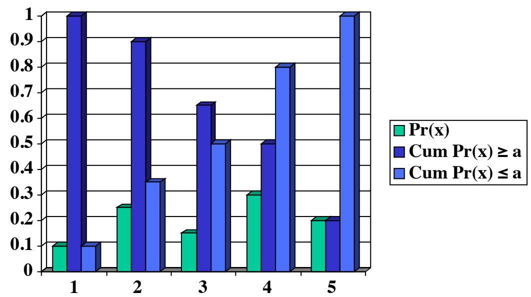
$$A_i = \{x \in U \mid x = x_i\} \quad ; \quad i = 1 \text{ to } I$$

- **Cumulative probability over all sets**



$$\sum_{i=1}^{I} \Pr(A_i) = \sum_{i=1}^{I} \Pr(x_i) = \frac{1}{N} \sum_{i=1}^{I} n_i = 1$$

Cumulative Probability, Pr(x ≥/≤ a), and Discrete Measurements of a Continuous Variable

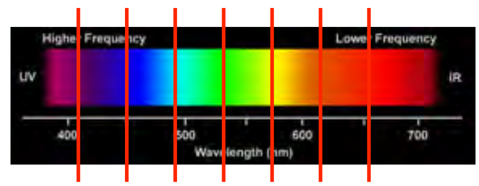


Suppose x represents a continuum of colors

x_i is the center of a band in x

$$\Pr\left(x_i \pm \Delta x / 2\right) = n_i / N$$

$$\sum_{i=1}^{I} \Pr\left(x_i \pm \Delta x / 2\right) = 1$$



Probability Density Function, pr(x) Cumulative Distribution Function, Pr(x < X)

Probability density function

$$\mathrm{pr}\left(x_i\right) = \frac{\Pr\left(x_i \pm \Delta x / 2\right)}{\Delta x}$$

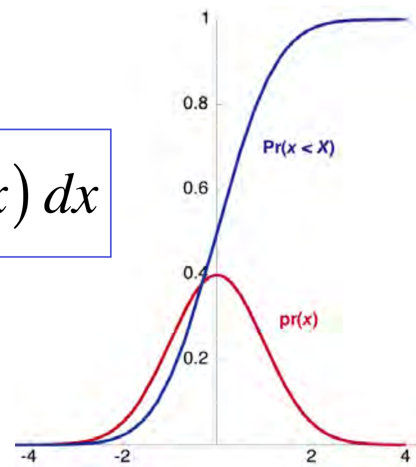
$$\sum_{i=1}^{I} \Pr\left(x_i \pm \Delta x / 2\right) = \sum_{i=1}^{I} \mathrm{pr}\left(x_i\right) \Delta x \xrightarrow[I \to \infty]{\Delta x \to 0} \int_{-\infty}^{\infty} \mathrm{pr}(x)\, dx = 1$$

Cumulative distribution function

$$\Pr\left(x < X\right) = \int_{-\infty}^{X} \mathrm{pr}(x)\, dx$$

Probability Density Function, pr(x) Cumulative Distribution Function, Pr(x <X)

$$\Pr\left(x < X\right) = \int_{-\infty}^{X} \mathrm{pr}\left(x\right)\,dx$$



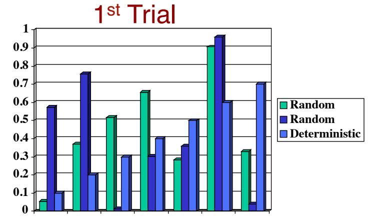
Random Number Example

Statistical -- not deterministic -- properties prior to actual event

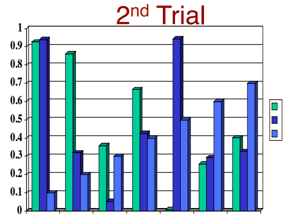
- **Excel spreadsheet: 2 random rows and one deterministic row**
  - **[RAND()] generates a uniform random number on each call**

=RAND()	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()
=RAND()	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()	=RAND()
0.1	0.2	0.3	0.4	0.5	0.6	0.7

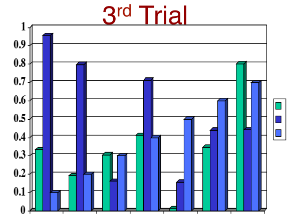
1st Trial



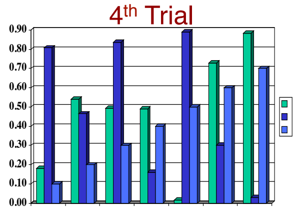
2nd Trial



3rd Trial



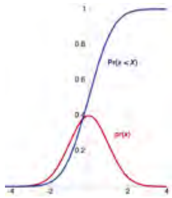
4th Trial



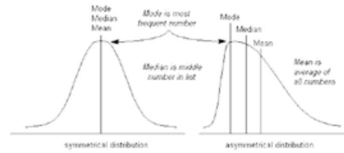
Output for 4th trial

0.18	0.54	0.49	0.49	0.02	0.73	0.88
0.81	0.46	0.84	0.16	0.89	0.30	0.03
0.10	0.20	0.30	0.40	0.50	0.60	0.70

Once the experiment is over, the results are determined



Properties of Random Variables

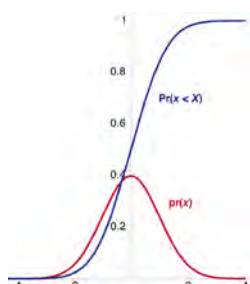


- **Mode**
  - **Value of *x* for which pr(*x*) is maximum**
- **Median**
  - **Value of *x* corresponding to 50th percentile**
  - **Pr(*x* < median) = Pr(*x* ≥ median) = 0.5**
- **Mean**
  - **Value of *x* corresponding to statistical average**
- **First moment of *x* = Expected value of *x***

"Force"

$$\bar{x} = E(x) = \int_{-\infty}^{\infty} x \,\mathrm{pr}\left(x\right) dx$$

"Moment arm"



Expected Values

- **Mean Value is the first moment of x**

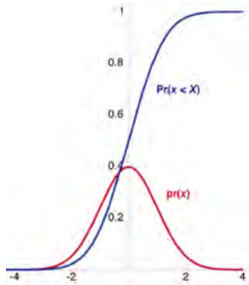
$$\boxed{\bar{x} = E(x) = \int_{-\infty}^{\infty} x \,\mathrm{pr}\left(x\right) dx}$$

- **Second central moment of *x* = Variance**
  - **Variance from the mean value rather than from zero**
  - **Smaller value indicates less uncertainty in the value of *x***

$$\boxed{E\left[\left(x-\bar{x}\right)^2\right] = \sigma_x^2 = \int_{-\infty}^{\infty} \left(x-\bar{x}\right)^2 \mathrm{pr}\left(x\right) dx}$$

- **Expected value of a function of x**

$$\boxed{E\left[f(x)\right] = \int_{-\infty}^{\infty} f(x) \,\mathrm{pr}\left(x\right) dx}$$



Expected Value is a Linear Operation

Expected value of sum of random variables

$$
\begin{aligned}
E[x_1 + x_2] &= \int_{-\infty}^{\infty} (x_1 + x_2)\,\mathrm{pr}(x)\,dx \\
&= \int_{-\infty}^{\infty} x_1\,\mathrm{pr}(x)\,dx + \int_{-\infty}^{\infty} x_2\,\mathrm{pr}(x)\,dx = E[x_1] + E[x_2]
\end{aligned}
$$

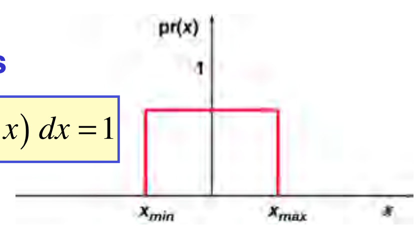
Expected value of constant times random variable

$$
E[k\,x] = \int_{-\infty}^{\infty} k\,x\,\mathrm{pr}(x)\,dx = k\int_{-\infty}^{\infty} x\,\mathrm{pr}(x)\,dx = k\,E[x]
$$

Mean Value of a Uniform Random Distribution

- Used in most random number generators (e.g., RAND)
- Bounded distribution
- Example is symmetric about the mean

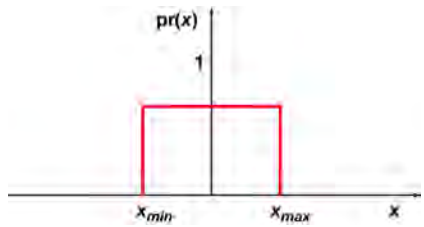
$$
\int_{-\infty}^{\infty} \mathrm{pr}(x)\,dx = 1
$$



$$
pr(x) = \begin{cases} 0 & ; \quad x < x_{\min} \\ \dfrac{1}{x_{\max} - x_{\min}} & ; \quad x_{\min} < x < x_{\max} \\ 0 & ; \quad x > x_{\max} \end{cases}
$$

$$
\begin{aligned}
\bar{x} = E(x) &= \int_{-\infty}^{\infty} x\,\mathrm{pr}(x)\,dx = \int_{x_{\min}}^{x_{\max}} \frac{x}{x_{\max} - x_{\min}}\,dx \\
&= \frac{1}{2}\frac{x_{\max}{}^2 - x_{\min}{}^2}{x_{\max} - x_{\min}} = \frac{1}{2}(x_{\max} + x_{\min})
\end{aligned}
$$

Variance and Standard Deviation of a Uniform Random Distribution



Variance

$$x_{\min} = -x_{\max} \triangleq a$$

$$E\left[\left(x-\bar{x}\right)^2\right] = \sigma_x{}^2 = \frac{1}{2a}\int_{-a}^{a} x^2\,dx = \left.\frac{x^3}{6a}\right|_{-a}^{a} = \frac{a^2}{3}$$

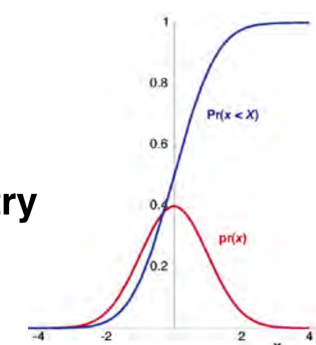
Standard deviation

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{a^2}{3}} = \frac{a}{\sqrt{3}}$$

Gaussian (Normal) Random Distribution

- **Used in some random number generators (e.g., RANDN)**
- **Unbounded, symmetric distribution**
- **Defined entirely by its mean and standard deviation**

$$\mathrm{pr}(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}}$$



Mean value; from symmetry

$$E(x) = \int_{-\infty}^{\infty} x\,\mathrm{pr}(x)\,dx = \bar{x}$$

Variance

$$E\left[\left(x-\bar{x}\right)^2\right] = \int_{-\infty}^{\infty} (x-\bar{x})^2\,\mathrm{pr}(x)\,dx = \sigma_x^2$$

Units of x and σ_x are the same

Probability of Being Close to the Mean
(Gaussian Distribution)

$$\mathrm{pr}(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}}$$

- **Probability of being within** $\pm 1\sigma_x$

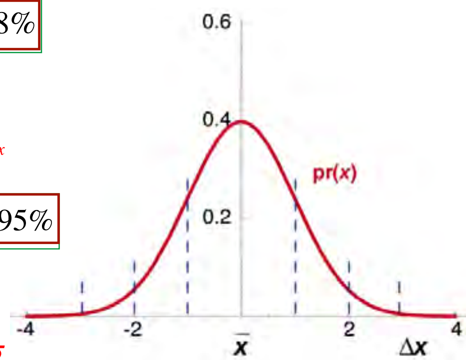
$$\mathrm{Pr}\left[x < (\bar{x} + \sigma_x)\right] - \mathrm{Pr}\left[x < (\bar{x} - \sigma_x)\right] \approx \boxed{68\%}$$

- **Probability of being within** $\pm 2\sigma_x$

$$\mathrm{Pr}\left[x < (\bar{x} + 2\sigma_x)\right] - \mathrm{Pr}\left[x < (\bar{x} - 2\sigma_x)\right] \approx \boxed{95\%}$$

- **Probability of being within** $\pm 3\sigma_x$

$$\mathrm{Pr}\left[x < (\bar{x} + 3\sigma_x)\right] - \mathrm{Pr}\left[x < (\bar{x} - 3\sigma_x)\right] \approx \boxed{99\%}$$



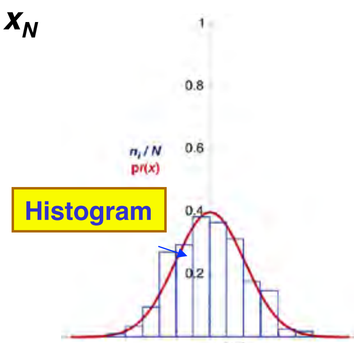
Experimental Determination of Mean and Variance

- **Sample mean for *N* data points, $x_1, x_2, \ldots, x_N$**

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

- **Sample variance for same data set**

$$\sigma_x^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{(N-1)}$$



- **Divisor is (*N* – 1) rather than *N* to produce an unbiased estimate**
  - **Only (*N* – 1) terms are independent**
  - **If N is large, the difference is inconsequential**
- **Distribution is not necessarily Gaussian**
  - **Prior knowledge: fit histogram to known distribution**
  - **Hypothesis test: determine best fit (e.g., Rayleigh, binomial, Poisson, ...)**

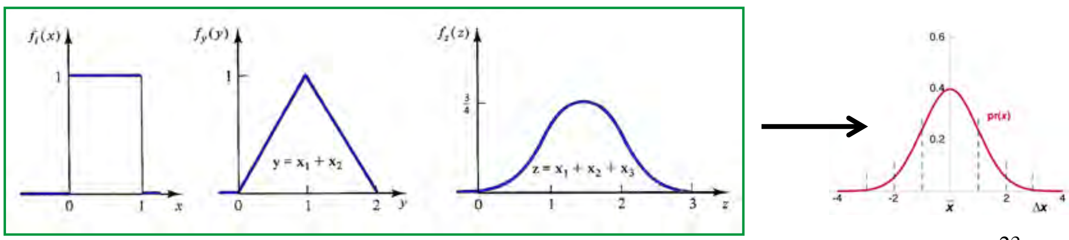
Central Limit Theorem

Probability density function of the sum of 2 random variables the convolution of their probability density functions (Papoulis, 1990)

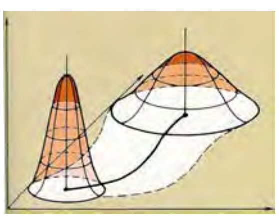
$$y = x_1 + x_2$$

$$pr(y) = \int_{-\infty}^{+\infty} pr\left[x_1\left(x_2\right)\right] pr(x_2)\,dx_2 = \int_{-\infty}^{+\infty} pr(y - x_2)\,pr(x_2)\,dx_2$$

The probability distribution of the sum of variables with any distributions approaches a normal distribution as the number of variables approaches infinity



CLTsim2017.m



Joint Probability *(n = 2)*

Suppose *x* can take *I* values and *y* can take *J* values; then,

$$\sum_{i=1}^{I} \Pr(x_i) = 1 \quad ; \quad \sum_{j=1}^{J} \Pr(y_j) = 1$$

If *x* and *y* are independent,

$$\Pr\left(x_i, y_j\right) = \Pr\left(x_i \wedge y_j\right) = \Pr\left(x_i\right)\Pr\left(y_j\right)$$

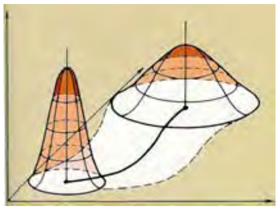
and

$$\sum_{i=1}^{I}\sum_{j=1}^{J} \Pr\left(x_i, y_j\right) = 1$$

Pr(y_j)

Pr(x_i)

	0.5	**0.3**	**0.2**	
0.6	0.3	0.18	0.12	0.6
0.4	0.2	0.12	0.08	0.4
	0.5	0.3	0.2	1



Conditional Probability
(n = 2)

If x and y are *not independent*, probabilities are related

Probability that x takes i^{th} value when y takes j^{th} value

Similarly

$$\Pr\left(x_i \mid y_j\right) = \frac{\Pr\left(x_i, y_j\right)}{\Pr\left(y_j\right)}$$

$$\Pr\left(y_j \mid x_i\right) = \frac{\Pr\left(x_i, y_j\right)}{\Pr\left(x_i\right)}$$

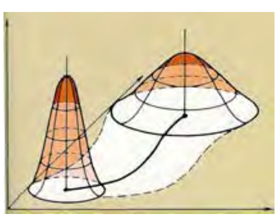
$$\Pr\left(x_i \mid y_j\right) = \Pr\left(x_i\right)$$

iff x and y are independent of each other

$$\Pr\left(y_j \mid x_i\right) = \Pr\left(y_j\right)$$

iff x and y are independent of each other

Conditional probability does not address causality



Applications of Conditional Probability
(n = 2)

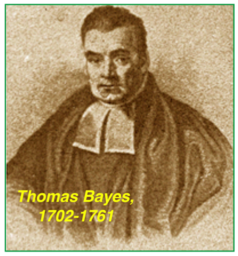
Joint probability can be expressed in two ways

$$\Pr\left(x_i, y_j\right) = \Pr\left(y_j \mid x_i\right)\Pr\left(x_i\right) = \Pr\left(x_i \mid y_j\right)\Pr\left(y_j\right)$$

Unconditional probability of each variable is expressed by a sum of terms

$$\Pr\left(x_i\right) = \sum_{j=1}^{J}\left[\Pr\left(x_i \mid y_j\right)\Pr\left(y_j\right)\right]$$

$$\Pr\left(y_j\right) = \sum_{i=1}^{I}\left[\Pr\left(y_j \mid x_i\right)\Pr\left(x_i\right)\right]$$



Bayes' s Rule

Bayes' s Rule proceeds from the previous results

Probability of x taking the value x_i conditioned on y taking its j^{th} value

$$\Pr\left(x_i \mid y_j\right) = \frac{\Pr\left(y_j \mid x_i\right)\Pr\left(x_i\right)}{\Pr\left(y_j\right)} = \frac{\Pr\left(y_j \mid x_i\right)\Pr\left(x_i\right)}{\sum_{i=1}^{I}\Pr\left(y_j \mid x_i\right)\Pr\left(x_i\right)}$$

... and the converse

$$\Pr\left(y_j \mid x_i\right) = \frac{\Pr\left(x_i \mid y_j\right)\Pr\left(y_j\right)}{\Pr\left(x_i\right)} = \frac{\Pr\left(x_i \mid y_j\right)\Pr\left(y_j\right)}{\sum_{j=1}^{J}\Pr\left(x_i \mid y_j\right)\Pr\left(y_j\right)}$$

Multivariate Statistics and Propagation of Uncertainty

Inner and Outer Products of Vectors

$$\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}; \quad \mathbf{y} = \begin{bmatrix} c \\ d \end{bmatrix}$$

Inner Product

$$\mathbf{x}^T\mathbf{y} = \mathbf{x} \bullet \mathbf{y} = \begin{bmatrix} a & b \end{bmatrix}\begin{bmatrix} c \\ d \end{bmatrix} = ac + bd$$

Outer Product

$$\mathbf{x}\mathbf{y}^T = \mathbf{x} \otimes \mathbf{y} = \begin{bmatrix} a \\ b \end{bmatrix}\begin{bmatrix} c & d \end{bmatrix} = \begin{bmatrix} ac & ad \\ bc & bd \end{bmatrix}$$

Multivariate Expected Values: Mean Value Vector and Covariance Matrix

Mean value vector of the dynamic state

$$\bar{\mathbf{x}} = E(\mathbf{x}) = \int_{-\infty}^{\infty} \mathbf{x}\,\mathrm{pr}(\mathbf{x})\,d\mathbf{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \ldots \\ \bar{x}_n \end{bmatrix}$$

$$\dim(\mathbf{x}) = n \times 1$$

Covariance matrix of the state

$$\mathbf{P} \triangleq E\left[(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^T\right] = \int_{-\infty}^{\infty} (\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^T\,\mathrm{pr}(\mathbf{x})\,d\mathbf{x}$$

If the state variation is Gaussian, its probability distribution is

$$\mathrm{pr}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\,|\mathbf{P}|^{1/2}}\,e^{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T \mathbf{P}^{-1}(\mathbf{x}-\bar{\mathbf{x}})}$$

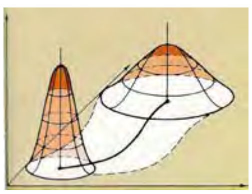
State Covariance Matrix is the Expected Value of the Outer Product of the Variations from the Mean

$$\mathbf{P} = E\left[(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^T\right]$$
$$= \begin{bmatrix} \sigma_{x_1}{}^2 & \rho_{12}\sigma_{x_1}\sigma_{x_2} & \dots & \rho_{1n}\sigma_{x_1}\sigma_{x_n} \\ \rho_{21}\sigma_{x_2}\sigma_{x_1} & \sigma_{x_2}{}^2 & \dots & \rho_{2n}\sigma_{x_2}\sigma_{x_n} \\ \dots & \dots & \dots & \dots \\ \rho_{n1}\sigma_{x_n}\sigma_{x_1} & \rho_{n2}\sigma_{x_n}\sigma_{x_2} & \dots & \sigma_{x_n}{}^2 \end{bmatrix}$$

$$\sigma_{x_1}{}^2 = \textit{Variance of } x_1$$

$$\rho_{12} = \textit{Correlation coefficient for } x_1 \textit{ and } x_2$$
$$-1 < \rho_{ij} < 1$$
$$\rho_{12}\sigma_{x_1}\sigma_{x_2} = \textit{Covariance of } x_1 \textit{ and } x_2$$

Gaussian probability distribution is totally described by its mean value and covariance matrix



$$\mathrm{pr}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\,|\mathbf{P}|^{1/2}}\, e^{-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T \mathbf{P}^{-1}(\mathbf{x}-\bar{\mathbf{x}})}$$

Stochastic Model for Propagating Mean Values and Covariances of Variables

LTI discrete-time model with known coefficients

$$\mathbf{x}_{k+1} = \mathbf{\Phi}\mathbf{x}_k + \mathbf{\Gamma}\mathbf{u}_k + \mathbf{\Lambda}\mathbf{w}_k, \quad \mathbf{x}_0 \quad \textit{given}$$

Mean and covariance of the state

$$\bar{\mathbf{x}}_0 = E[\mathbf{x}_0]; \quad \mathbf{P}_0 = E\left\{[\mathbf{x}_0 - \bar{\mathbf{x}}_0][\mathbf{x}_0 - \bar{\mathbf{x}}_0]^T\right\}$$

$$\bar{\mathbf{x}}_k = E[\mathbf{x}_k]; \quad \mathbf{P}_k = E\left\{[\mathbf{x}_k - \bar{\mathbf{x}}_k][\mathbf{x}_k - \bar{\mathbf{x}}_k]^T\right\}$$

Covariance of the disturbance with zero mean value

$$\bar{\mathbf{w}}_k = \mathbf{0}; \quad \mathbf{Q}_k = E\left\{[\mathbf{w}_k][\mathbf{w}_k]^T\right\}$$

Mean of perfectly known control vector

$$\mathbf{u}_k = \bar{\mathbf{u}}_k = E[\mathbf{u}_k]; \quad \mathbf{U}_k = \mathbf{0}$$

Mean Value and Covariance of the Disturbance

$$\bar{\mathbf{w}} = E(\mathbf{w}) = \int_{-\infty}^{\infty} \mathbf{w} \, \mathrm{pr}(\mathbf{w}) \, d\mathbf{w} = \begin{bmatrix} \bar{\mathbf{w}}_1 \\ \bar{\mathbf{w}}_2 \\ \dots \\ \bar{\mathbf{w}}_n \end{bmatrix}$$

$\dim(\mathbf{w}) = s \times 1$

$$\mathbf{Q} \triangleq E\left[(\mathbf{w} - \bar{\mathbf{w}})(\mathbf{w} - \bar{\mathbf{w}})^T\right] = \int_{-\infty}^{\infty} (\mathbf{w} - \bar{\mathbf{w}})(\mathbf{w} - \bar{\mathbf{w}})^T \, \mathrm{pr}(\mathbf{w}) \, d\mathbf{w}$$

If the disturbance is Gaussian, its probability distribution is

$$\mathrm{pr}(\mathbf{w}) = \frac{1}{(2\pi)^{s/2} \, |\mathbf{Q}|^{1/2}} e^{-\frac{1}{2}(\mathbf{w}-\bar{\mathbf{w}})\mathbf{Q}^{-1}(\mathbf{w}-\bar{\mathbf{w}})}$$

Dynamic Model to Propagate the Mean Value of the State

$$E(\mathbf{x}_{k+1}) = E(\mathbf{\Phi}\mathbf{x}_k + \mathbf{\Gamma}\mathbf{u}_k + \mathbf{\Lambda}\mathbf{w}_k)$$

If disturbance mean value is zero

$$\bar{\mathbf{x}}_{k+1} = \mathbf{\Phi}\bar{\mathbf{x}}_k + \mathbf{\Gamma}\bar{\mathbf{u}}_k + 0, \quad \bar{\mathbf{x}}_0 \text{ given}$$

Dynamic Model to Propagate the Covariance of the State

$$\mathbf{P}_{k+1} = E\left\{\left[\mathbf{x}_{k+1} - \overline{\mathbf{x}}_{k+1}\right]\left[\mathbf{x}_{k+1} - \overline{\mathbf{x}}_{k+1}\right]^T\right\}$$
$$= E\left[\left(\mathbf{\Phi}\left[\mathbf{x}_k - \overline{\mathbf{x}}_k\right] + \mathbf{\Gamma}\mathbf{u}_k + \mathbf{\Lambda}\mathbf{w}_k\right)\left(\mathbf{\Phi}\left[\mathbf{x}_k - \overline{\mathbf{x}}_k\right] + \mathbf{\Gamma}\mathbf{u}_k + \mathbf{\Lambda}\mathbf{w}_k\right)^T\right]$$

Expected values of cross terms are zero

$$\mathbf{P}_{k+1} = E\left\{\mathbf{\Phi}\left[\mathbf{x}_k - \overline{\mathbf{x}}_k\right]\left[\mathbf{x}_k - \overline{\mathbf{x}}_k\right]^T{}_k \mathbf{\Phi}^T + 0 + \mathbf{\Lambda}\mathbf{w}_k\mathbf{w}^T{}_k\mathbf{\Lambda}^T{}_k\right\}$$
$$= \mathbf{\Phi}E\left\{\left[\mathbf{x}_k - \overline{\mathbf{x}}_k\right]\left[\mathbf{x}_k - \overline{\mathbf{x}}_k\right]^T{}_k\right\}\mathbf{\Phi}^T + \mathbf{\Lambda}E\left(\mathbf{w}_k\mathbf{w}^T{}_k\right)\mathbf{\Lambda}^T$$
$$= \mathbf{\Phi}\mathbf{P}_k\mathbf{\Phi}^T + \mathbf{\Lambda}\mathbf{Q}_k\mathbf{\Lambda}^T, \quad \mathbf{P}_0 \text{ given}$$

LTI System Propagation of the Mean and Covariance

Propagation of the Mean Value

$$\overline{\mathbf{x}}_{k+1} = \mathbf{\Phi}\overline{\mathbf{x}}_k + \mathbf{\Gamma}\overline{\mathbf{u}}_k, \quad \overline{\mathbf{x}}_0 \text{ given}$$

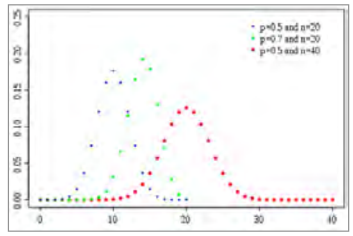
Propagation of the Covariance

$$\mathbf{P}_{k+1} = \mathbf{\Phi}\mathbf{P}_k\mathbf{\Phi}^T + \mathbf{\Lambda}\mathbf{Q}_k\mathbf{\Lambda}^T, \quad \mathbf{P}_0 \text{ given}$$

Both propagation equations are *linear*

Some Non-Gaussian Distributions

- **Binomial Distribution**
  - **Random variable, *x***
  - **Probability of *k* successes in *n* trials**
  - **Discrete probability distribution described by a probability mass function, pr(*x*)**



$$\mathrm{pr}(x) = \frac{n!}{k!(n-k)!} p(x)^k \left[1 - p(x)\right]^{n-k} \triangleq \begin{pmatrix} n \\ k \end{pmatrix} p(x)^k \left[1 - p(x)\right]^{n-k}$$

$$= \text{ probability of exactly } k \text{ successes in } n \text{ trials, in } (0,1)$$

$$\sim \text{ normal distribution for large } n$$

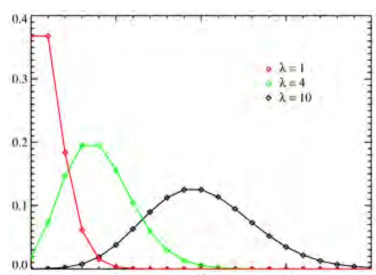
Parameters of the distribution

$p(x)$: probability of occurrence, in $(0,1)$

n : number of trials

Some Non-Gaussian Distributions

- **Poisson Distribution**
  - **Probability of a number of events occurring in a fixed period of time**
  - **Discrete probability distribution described by a probability mass function**



$$\mathrm{pr}(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$\lambda =$ Average rate of occurrence of event (per unit time)

$k =$ # of occurrences of the event

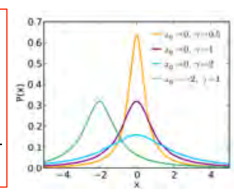
$\mathrm{pr}(k) =$ probability of k occurrences (per unit time)

$\sim$ normal distribution for large λ

- **Cauchy-Lorentz Distribution**
  - **Mean and variance are undefined**
  - **"Fat tails": extreme values more likely than normal distribution**
  - **Central limit theorem fails**

$$\mathrm{pr}(x) = \frac{\gamma}{\pi\left[\gamma^2 + (x - x_0)^2\right]}$$

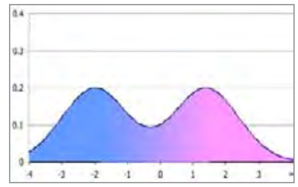
$$\mathrm{Pr}(x) = \frac{1}{\pi}\tan^{-1}\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}$$



Some Non-Gaussian Distributions

Bimodal Distributions

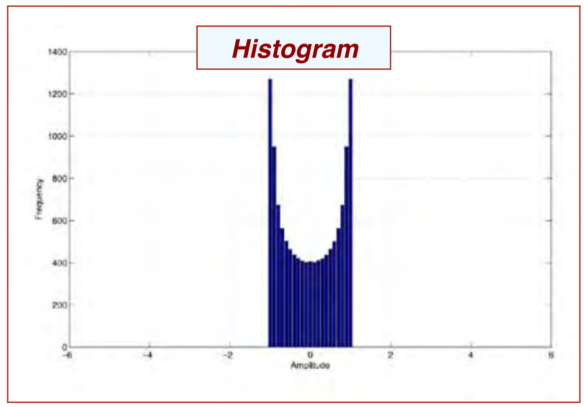
- **Bimodal Distribution**
  - Two Peaks
  - e.g., concatenation of 2 normal distributions with different means



- **Random Sine Wave**

$$x = A\sin\left[\omega t + \text{ random phase angle}\right]$$

$$pr(x) = \begin{cases} \dfrac{1}{\pi\sqrt{A^2 - x^2}} & |x| \le A \\ 0 & |x| > A \end{cases}$$



Next Time:
Machine Learning:
Classification of Data Sets

Supplemental Material

Correlation and Independence

- **Probability density functions of two random variables, *x* and *y***

$pr(x)$ and $pr(y)$ given for all x and y in $(-\infty,\infty)$

$pr(x,y)$: Joint probability density function of x and y

$$\int_{-\infty}^{\infty} pr(x)\,dx = 1; \quad \int_{-\infty}^{\infty} pr(y)\,dy = 1; \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} pr(x,y)\,dx\,dy = 1;$$

- **Expected values of *x* and *y***
  - Mean values
  - Covariance

$$E(x) = \int_{-\infty}^{\infty} x\,pr(x)\,dx = \bar{x}$$

$$E(y) = \int_{-\infty}^{\infty} y\,pr(y)\,dy = \bar{y}$$

$$E(xy) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\,pr(x,y)\,dx\,dy$$

Independence *(probability)* and Correlation *(expected value)*

x **and** ***y*** **are independent if**

$$pr(x,y) = pr(x)\,pr(y) \text{ at every } x \text{ and } y \text{ in } (-\infty,\infty)$$
$$pr(x\,|\,y) = pr(x); \quad pr(y\,|\,x) = pr(y)$$

Dependence

$$pr(x,y) \neq pr(x)\,pr(y) \text{ for some } x \text{ and } y \text{ in } (-\infty,\infty)$$

x **and** ***y*** **are uncorrelated if**

$$E(xy) = E(x)E(y) = \bar{x}\,\bar{y}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\; pr(x,y)\,dx\,dy = \int_{-\infty}^{\infty} x\; pr(x)\,dx \int_{-\infty}^{\infty} y\; pr(y)\,dy$$

Correlation

$$E(xy) \neq E(x)E(y)$$

Which Combinations are Possible?

Independence and lack of correlation

$$pr(x,y) = pr(x)\,pr(y) \text{ at every } x \text{ and } y \text{ in } (-\infty,\infty)$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\; pr(x,y)\,dx\,dy = \int_{-\infty}^{\infty} x\; pr(x)\,dx \int_{-\infty}^{\infty} y\; pr(y)\,dy = \bar{x}\,\bar{y}$$

Independence and correlation

$$pr(x,y) = pr(x)\,pr(y) \text{ at every } x \text{ and } y \text{ in } (-\infty,\infty)$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\; pr(x,y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\; pr(x)\,pr(y)\,dx\,dy \neq \int_{-\infty}^{\infty} x\; pr(x)\,dx \int_{-\infty}^{\infty} y\; pr(y)\,dy = \bar{x}\,\bar{y}$$

Dependence and lack of correlation

$$pr(x,y) \neq pr(x)\,pr(y) \text{ for some } x \text{ and } y \text{ in } (-\infty,\infty)$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\; pr(x,y)\,dx\,dy = \int_{-\infty}^{\infty} x\; pr(x)\,dx \int_{-\infty}^{\infty} y\; pr(y)\,dy = \bar{x}\,\bar{y}$$

Dependence and correlation

$$pr(x,y) \neq pr(x)\,pr(y) \text{ for some } x \text{ and } y \text{ in } (-\infty,\infty)$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\; pr(x,y)\,dx\,dy \neq \int_{-\infty}^{\infty} x\; pr(x)\,dx \int_{-\infty}^{\infty} y\; pr(y)\,dy = \bar{x}\,\bar{y}$$

Correlation, Orthogonality, and Dependence of Two Random Variables

If two variables are *uncorrelated*

$$E(xy) = E(x)E(y)$$

Two variables are *orthogonal* if

$$E(xy) = 0$$

Two variables are *independent* if

$$\mathrm{pr}(x, y) = \mathrm{pr}(x)\mathrm{pr}(y)$$

Given independent x and y

$$E\left[g(x)h(y)\right] = E\left[g(x)\right]E\left[h(y)\right]$$

Still no notion of causality

Example

$$\mathbf{x}_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} x_{1_k} \\ x_{2_k} \end{bmatrix}$$

2nd-order LTI system

$$\mathbf{x}_{k+1} = \mathbf{\Phi}\mathbf{x}_k + \mathbf{\Lambda}\mathbf{w}_k, \quad \mathbf{x}_0 = \mathbf{0}$$

Gaussian disturbance, $\mathbf{w}_k$, with independent, uncorrelated components

$$\bar{\mathbf{w}} = \begin{bmatrix} \bar{w}_1 \\ \bar{w}_2 \end{bmatrix}; \quad \mathbf{Q} = \begin{bmatrix} \sigma_{w_1}^2 & 0 \\ 0 & \sigma_{w_2}^2 \end{bmatrix}$$

Propagation of state mean and covariance

$$\bar{\mathbf{x}}_{k+1} = \mathbf{\Phi}\bar{\mathbf{x}}_k + \mathbf{\Lambda}\bar{\mathbf{w}}$$

$$\mathbf{P}_{k+1} = \mathbf{\Phi}\mathbf{P}_k\mathbf{\Phi}^T + \mathbf{\Lambda}\mathbf{Q}\mathbf{\Lambda}^T, \quad \mathbf{P}_0 = 0$$

Off-diagonal elements of **P** and **Q** express correlation

$$\boxed{\bar{\mathbf{x}}_{k+1} = \mathbf{\Phi}\bar{\mathbf{x}}_k + \mathbf{\Lambda}\bar{\mathbf{w}}}$$

$$\boxed{\mathbf{P}_{k+1} = \mathbf{\Phi}\mathbf{P}_k\mathbf{\Phi}^T + \mathbf{\Lambda}\mathbf{Q}\mathbf{\Lambda}^T, \quad \mathbf{P}_0 = 0}$$

Example

Independence and lack of correlation in state

$$\mathbf{\Phi} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}; \quad \mathbf{\Lambda} = \begin{bmatrix} c & 0 \\ 0 & d \end{bmatrix}$$

Independent dynamics and correlation in state

$$\mathbf{\Phi} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}; \quad \bar{\mathbf{x}}_0 \neq \mathbf{0}; \quad \bar{\mathbf{w}} = \bar{w}; \quad \mathbf{\Lambda} = \begin{bmatrix} c \\ c \end{bmatrix}$$

Dependence and lack of correlation in nonlinear output

$$\mathbf{\Phi} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}; \quad \mathbf{\Lambda} = \begin{bmatrix} e & 0 \\ 0 & f \end{bmatrix}$$

$$\mathbf{z} = \begin{bmatrix} x_1 \\ x_2^2 \end{bmatrix}; \quad \text{Conjecture (t.b.d.)}$$

Dependence and correlation in state

$$\mathbf{\Phi} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}; \quad \mathbf{\Lambda} = \begin{bmatrix} e & 0 \\ 0 & f \end{bmatrix}$$

2nd-Order Example of Uncertainty Propagation

Position and Velocity

LTI Dynamic System with Random Disturbance

$$\begin{bmatrix} x_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} x_k \\ v_k \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} u_k + \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} w_k$$

Propagation of the Mean Value

$$\begin{bmatrix} \bar{x}_{k+1} \\ \bar{v}_{k+1} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \bar{x}_k \\ \bar{v}_k \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \bar{u}_k$$

Propagation of the Covariance

$$\begin{bmatrix} p_{xx_{k+1}} & p_{xv_{k+1}} \\ p_{vx_{k+1}} & p_{vv_{k+1}} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} p_{xx_k} & p_{xv_k} \\ p_{vx_k} & p_{vv_k} \end{bmatrix} \begin{bmatrix} \phi_{11} & \phi_{21} \\ \phi_{12} & \phi_{22} \end{bmatrix} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \sigma_{w_k}^2 \begin{bmatrix} \lambda_1 & \lambda_2 \end{bmatrix}$$