Probability and Statistics

Robert Stengel
Optimal Control and Estimation MAE 546
Princeton University, 2015

- Concepts and reality
- Probability distributions
- Bayes' s Law
- Stationarity and ergodicity
- Correlation functions and power spectra
- Propagation of a probability distribution

http://www.princeton.edu/~stengel/MAE546.html
http://www.princeton.edu/~stengel/OptConEst.html

Concepts and Reality of Probability

(Papoulis, 1990)

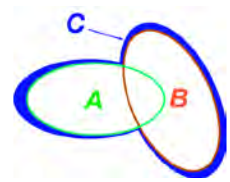
- Theory may be exact
  - Deals with averages of phenomena with many possible outcomes
  - Based on models of behavior
- Application can only be approximate
  - Measure of our state of knowledge or belief that something may or may not be true
  - Subjective assessment

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

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>: With no common elements,**



$$\Pr\left(A_i \cup A_j\right) = \Pr\left(A_i\right) + \Pr\left(A_j\right)$$

- **Relative Frequency (Empirical interpretation)**

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

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

Interpretations of Probability

(Papoulis)

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

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

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

- **Measure of belief (Subjective interpretation)**
  - **Pr(A_i) = measure of belief 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**

Probability

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

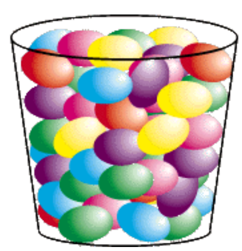
Statistics

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

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

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

- N = total number of events
- n_i = number of events with value x_i
- I = number of different values
- x_i = ordered set of hypotheses or values

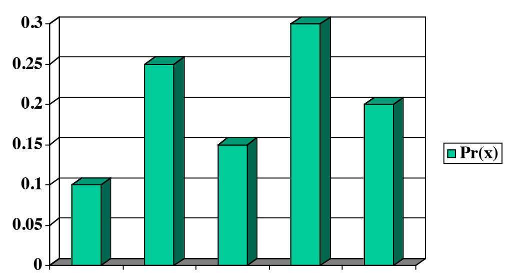


- x is a random variable
- Equivalent sets

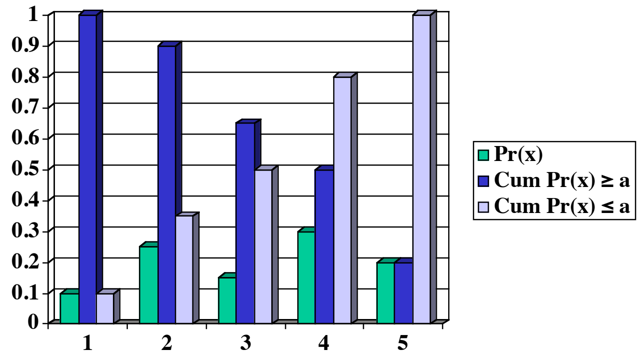
$$A_i = \left\{ x \in U \middle| \, x = x_i \right\} \quad ; \quad i = 1 \text{ to } I$$

- Cumulative probability over all sets

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



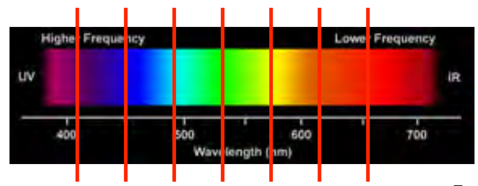
Cumulative Probability, Pr(x ≥/≤ a), and Discrete Measurements of a Continuous Variable



- 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), and Cumulative Distribution Function, Pr(x <X)

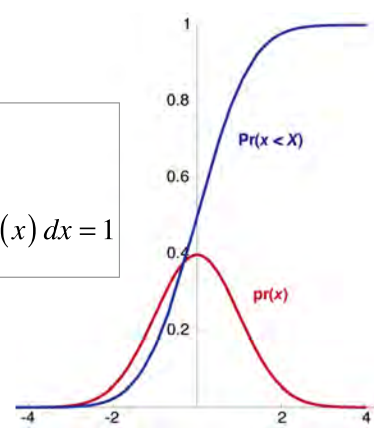
Probability density function

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

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

Cumulative distribution function

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

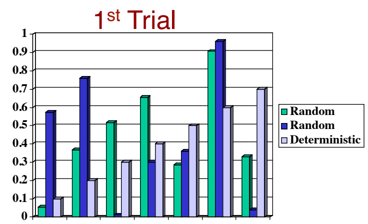


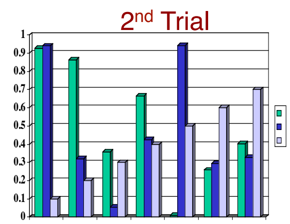
Gaussian probability density and cumulative distribution functions

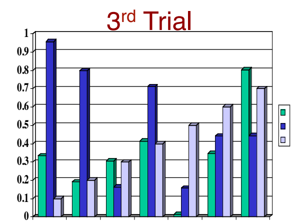
Random Number Example

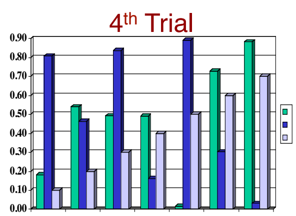
- **Statistical properties prior to actual event**
- **Excel spreadsheet: 2 random rows and one deterministic row**
  - **[RAND()] generates a uniform random number on each call**

=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





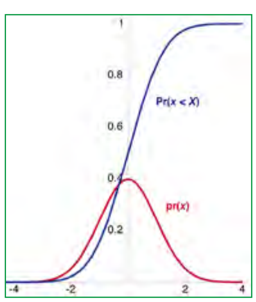




- **Output for 4th trial**

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



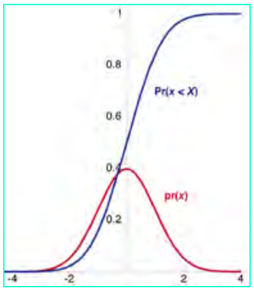
Properties of Random Variables

- **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”

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

“Moment arm”



Expected Values

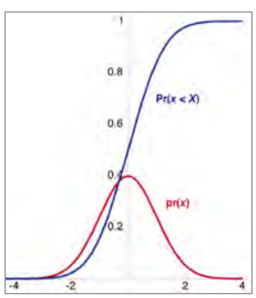
Mean Value $\bar{x} = E(x) = \int_{-\infty}^{\infty} x \, \mathrm{pr}(x) \, dx$

- **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**

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

- **Expected value of any function of x, *f(x)***

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



Expected Value is a Linear Operation

Expected value of sum of random variables

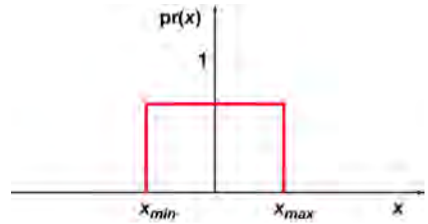
x_1 and x_2 need not be statistically independent

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

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

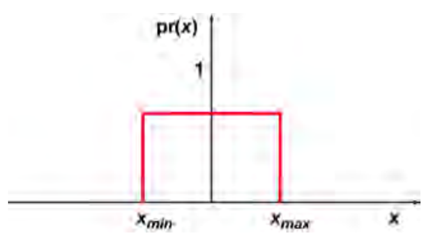
$$\mathrm{pr}(x) = \begin{cases} 0 & ; \quad x < x_{\min} \\ \dfrac{1}{x_{\max} - x_{\min}} & ; \quad x_{\min} < x < x_{\max} \\ 0 & ; \quad x > x_{\max} \end{cases}$$

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

Mean value

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

Variance and Standard Deviation of a Uniform Random Distribution



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

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

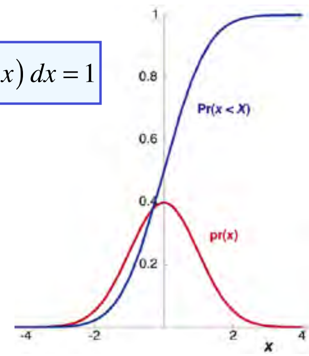
$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{a^2}{3}} = \frac{a}{\sqrt{3}}$$

Gaussian (Normal) Random Distribution

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

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

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



Mean value: First moment, μ_1

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

Variance: Second central moment, μ_2

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

Units of x and σ_x are the same

Probability of Being Close to the Mean
(Gaussian Distribution)

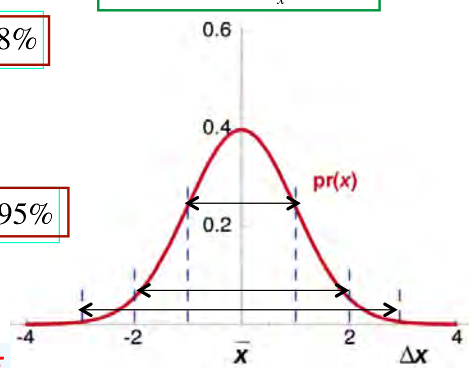
$$\boxed{\mathrm{pr}(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}}}$$

Probability of being within $\pm 1\sigma$

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

Probability of being within $\pm 2\sigma$

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



Probability of being within $\pm 3\sigma$

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

Experimental Determination of Mean and Variance

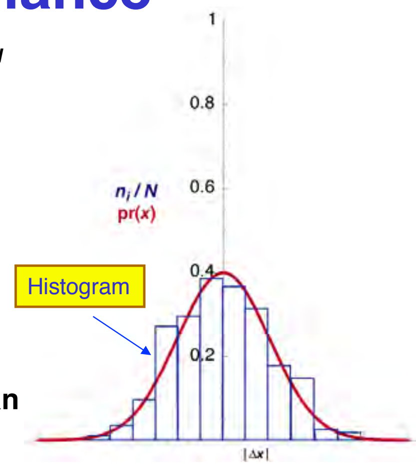
- **Sample mean** for **N data points**, $x_1, x_2, \ldots, x_N$

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

- **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
  - $(N-1)$ terms are independent
  - Inconsequential for large N

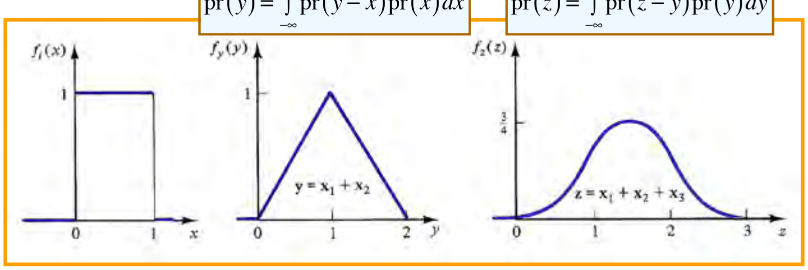


- **Distribution is not necessarily Gaussian**
  - **Prior knowledge**: fit histogram to known distribution
  - **Hypothesis test**: determine best fit (e.g., Rayleigh, binomial, Poisson, ...)

Central Limit Theorem

- **Probability distribution of the <u>sum of independent, identically distributed *(i.i.d.)* variables</u>**
  - Approaches **normal distribution** as number of variables approaches infinity
  - Summation of continuous random variables produces a **convolution** of probability density functions (Papoulis, 1990)
  - See *Supplemental Material* for sufficient conditions

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



Multiple Probability Densities and Expected Values

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

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

$\mathrm{pr}(x,y)$: **Joint probability density function** of x and y

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

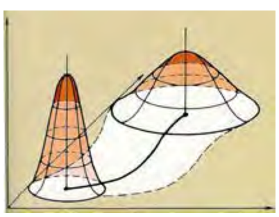
Expected values of *x* and *y*

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

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

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

Autocovariance $E(x^2) = \int_{-\infty}^{\infty} x^2\; pr(x)\,dx$



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 uncorrelated,

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

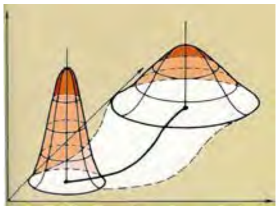
and

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

$\Pr(y_j)$

$\Pr(x_i)$

	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

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

- **Similarly**

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

$\Pr(x_i \mid y_j) = \Pr(x_i)$ iff x and y are independent of each other

$\Pr(y_j \mid x_i) = \Pr(y_j)$ iff x and y are independent of each other

Causality is not addressed by conditional probability

Dependence and Correlation

x and *y* are independent if

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

$$\mathrm{pr}(x \mid y) = \mathrm{pr}(x); \quad \mathrm{pr}(y \mid x) = \mathrm{pr}(y)$$

Dependence

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

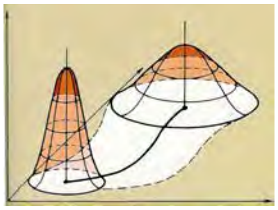
x and *y* are uncorrelated if

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

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

Correlation

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



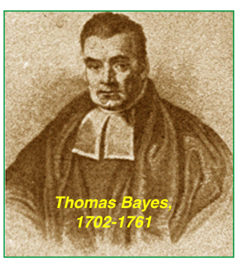
Applications of Conditional Probability (*n* = 2)

Joint probability can be expressed in two ways

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

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

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



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

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

... and the converse

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

Random Processes

Nonstationary Processes

Non-Stationary Process: Ensemble statistics (e.g., joint probability distribution and expected value) depend on t_1 and t_2

$$\boxed{\begin{aligned}&\mathrm{pr}_{ensemble}\left[x(t_1),x(t_2)\right]\\&E_{ensemble}\left[x(t_1)x(t_2)\right]\end{aligned}}$$

Stationary Processes

Stationary Process: Ensemble statistics (e.g., joint probability distribution and expected value) depend on Δt

$$\boxed{\begin{aligned}\mathrm{pr}_{ensemble}\left[x(t_1),x(t_2)\right] &= \mathrm{pr}_{ensemble}\left[x(t_1),x(t_1+\Delta t)\right] \\ &= \mathrm{pr}_{ensemble}\left[x(t),x(t+\Delta t)\right]\end{aligned}}$$

$$\boxed{E_{ensemble}\left[x(t_1)x(t_2)\right] = E_{ensemble}\left[x(t)x(t+\Delta t)\right]}$$

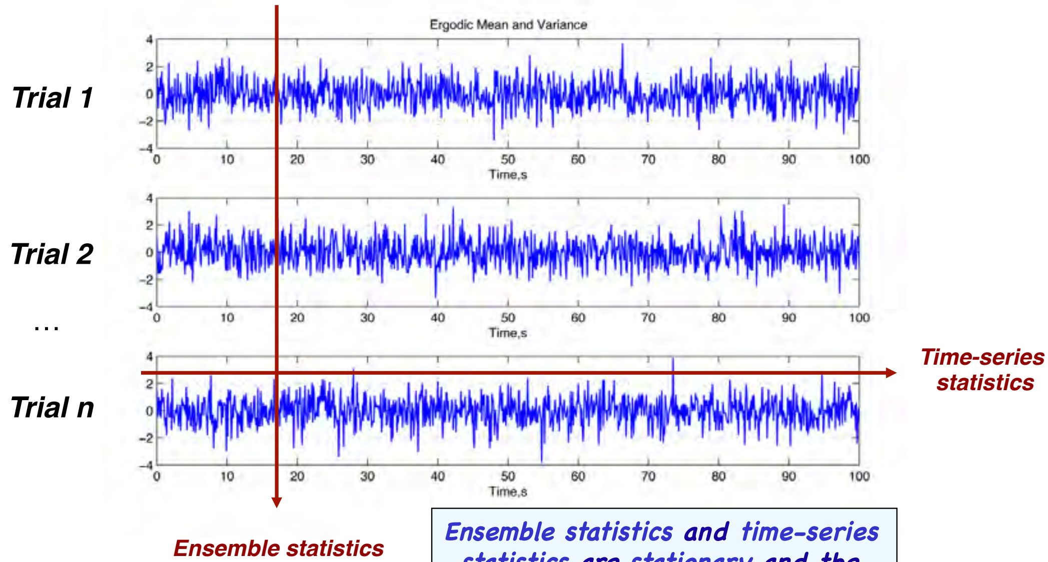
Ergodic Processes

Ergodic Process: Ensemble statistics and time-series statistics are stationary and the same

$$\boxed{\mathrm{pr}_{ensemble}\left[x(t),x(t+\Delta t)\right] = \mathrm{pr}_{time}\left[x(t),x(t+\Delta t)\right]}$$

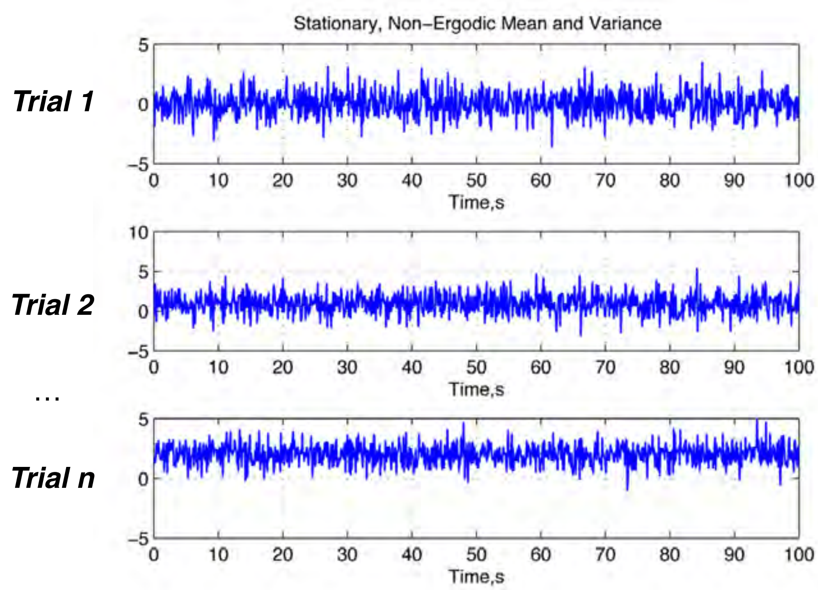
$$\boxed{E_{ensemble}\left[x(t)x(t+\Delta t)\right] = E_{time}\left[x(t)x(t+\Delta t)\right]}$$

Stationary, Ergodic Process



Ensemble statistics and time-series statistics are stationary and the same for each trial

Stationary, Non-Ergodic Process



*Ensemble statistics and time-series statistics are stationary and the **not** the same for each trial*

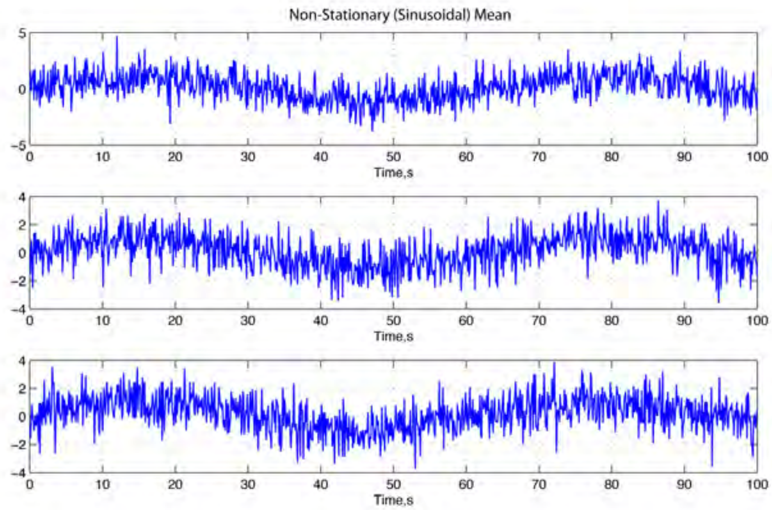
Non-Stationary Process: Sinusoidal Mean

Trial 1

Trial 2

…

Trial n



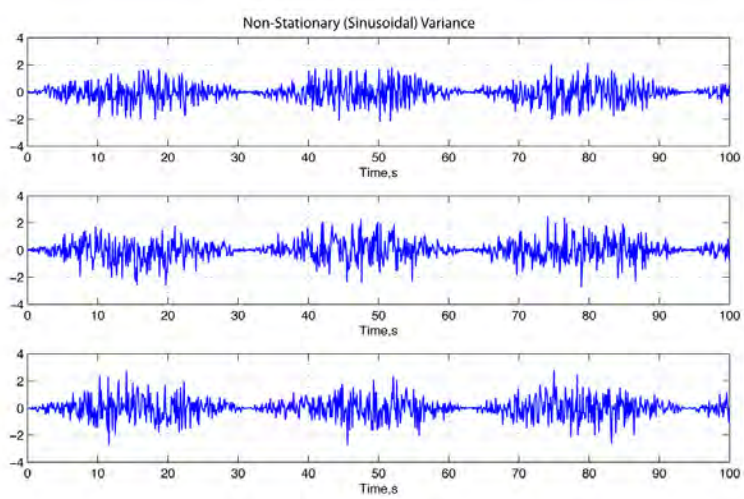
Non-Stationary Process: Sinusoidal Variance

Trial 1

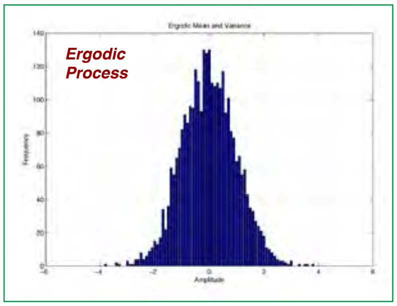
Trial 2

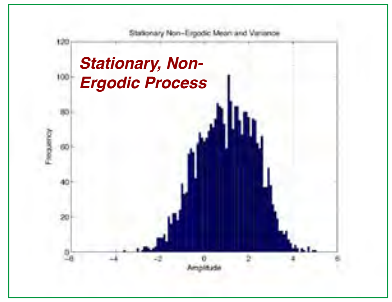
…

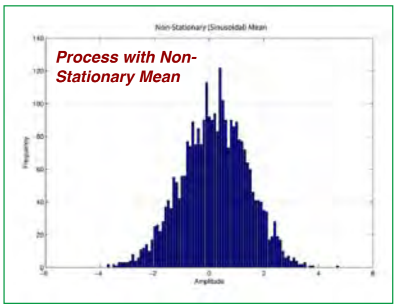
Trial n

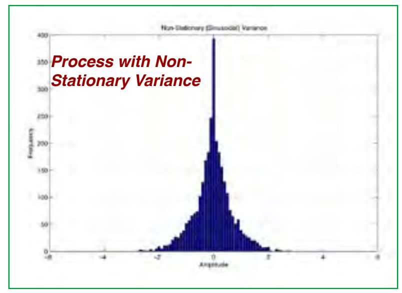


Histograms of Random Processes









Correlation and Covariance Functions

Autocorrelation Functions of Random Processes

Autocorrelation Function for Non-Stationary Variance

$$\boxed{E\left[x(t_1),x(t_2)\right]}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(t_1)x(t_2)\mathrm{pr}_{ensemble}\left[x(t_1),x(t_2)\right]dx(t_1)dx(t_2)$$

$$\triangleq \psi\left[x(t_1),x(t_2)\right]$$

Autocovariance Functions of Random Processes

Mean values subtracted from variables

$$\boxed{E\left\{\left[x(t_1)-\bar{x}(t_1)\right]\left[x(t_2)-\bar{x}(t_2)\right]\right\}}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[x(t_1)-\bar{x}(t_1)\right]\left[x(t_2)-\bar{x}(t_2)\right]\mathrm{pr}_{ensemble}\left[x(t_1),x(t_2)\right]dx(t_1)dx(t_2)$$

$$\triangleq \phi\left[\tilde{x}(t_1)\tilde{x}(t_2)\right], \quad where\ (\tilde{x}) \triangleq x-\bar{x}$$

Autocovariance Functions of Random Processes

With $t_1 = t_2$, autocovariance function = Variance

$$\boxed{\begin{aligned}E\left\{\left[x(t_1)-\bar{x}(t_1)\right]\left[x(t_2)-\bar{x}(t_2)\right]\right\} &= E\left\{\left[x(t_1)-\bar{x}(t_1)\right]^2\right\} \\ &= E\left[\tilde{x}^2(t_1)\right]=\sigma_x^2\end{aligned}}$$

$$\boxed{\tilde{x}(t) \triangleq \left[x(t)-\bar{x}(t)\right]}$$

Cross-Covariance Functions

Cross-covariance Function for Non-Stationary Variance

$$\boxed{E\left\{\left[x(t_1)-\bar{x}(t_1)\right]\left[y(t_2)-\bar{y}(t_2)\right]\right\}}$$

$$\boxed{\begin{aligned}&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[x(t_1)-\bar{x}(t_1)\right]\left[y(t_2)-\bar{y}(t_2)\right]\mathrm{pr}_{ensemble}\left[x(t_1),y(t_2)\right]dx(t_1)\,dy(t_2) \\ &= \phi\left[\tilde{x}(t_1)\tilde{y}(t_2)\right]\end{aligned}}$$

With $t_1 = t_2$, cross-covariance function

$$\boxed{E\left\{\left[x(t_1)-\bar{x}(t_1)\right]\left[y(t_1)-\bar{y}(t_1)\right]\right\}=\sigma_{xy}}$$

Covariance Functions for Stationary Processes

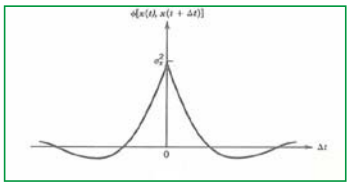
For stationary processes, statistics are independent of time

Autocovariance

$$\phi\left[\tilde{x}(t_1)\tilde{x}(t_2)\right] = \phi\left[\tilde{x}(t_1)\tilde{x}(t_1+\Delta t)\right] \triangleq \phi_{xx}(\Delta t)$$

Cross-covariance

$$\phi\left[\tilde{x}(t_1)\tilde{y}(t_2)\right] = \phi\left[\tilde{x}(t_1)\tilde{y}(t_1+\Delta t)\right] \triangleq \phi_{xy}(\Delta t)$$



$$\phi_{xx}(\Delta t) = \phi_{xx}(-\Delta t)$$

$\Delta t =$ Lag time

Ordering in product is immaterial; therefore,

Autocovariance function is symmetric

Cross-covariance function is not symmetric unless *y* = *x*

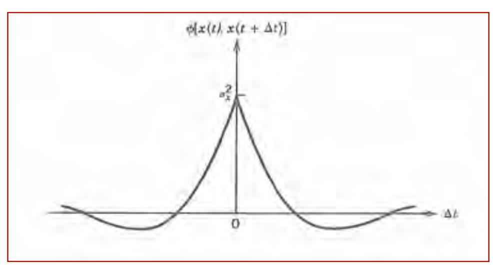
Properties of Continuous-Time Covariance Functions for Ergodic Processes

Correlation depends on Δt only

Convolution integrals of the dependent variables

$$\phi_{xx}(\Delta t) = E\left(x(t)x(t+\Delta t)\right) = \lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)x(t+\Delta t)\,dt$$

$$\phi_{yy}(\Delta t) = E\left(y(t)y(t+\Delta t)\right) = \lim_{T\to\infty}\frac{1}{T}\int_0^T y(t)y(t+\Delta t)\,dt$$



Cross-covariance function

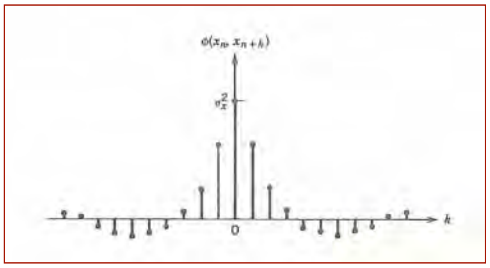
$$\phi_{xy}(\Delta t) = E\left(x(t)y(t+\Delta t)\right) = \lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)y(t+\Delta t)\,dt$$

Properties of Discrete-Time Covariance Functions for Ergodic Processes

k = Number of (±) lags

Convolution sums of the dependent variables

Discrete case approaches continuous case as *k* -> 0



$$\phi_{xx}(k) = E(x_n x_{n+k}) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}(x_n x_{n+k})$$

$$\phi_{yy}(k) = E(y_n y_{n+k}) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}(y_n y_{n+k})$$

Cross-covariance function

$$\phi_{xy}(k) = E(x_n y_{n+k}) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}(x_n y_{n+k})$$

Additional Properties of Covariance Functions for Stationary Processes

"Self" correlation ≥ lagged correlation

$$\phi_{xx}(0) \geq \phi_{xx}(\Delta t)$$

$$\phi_{yy}(0) \geq \phi_{yy}(\Delta t)$$

$$\phi_{xx}(0)\phi_{yy}(0) \geq \left[\phi_{xy}(\Delta t)\right]^2$$

Additional Properties of Correlation Functions for Stationary Processes

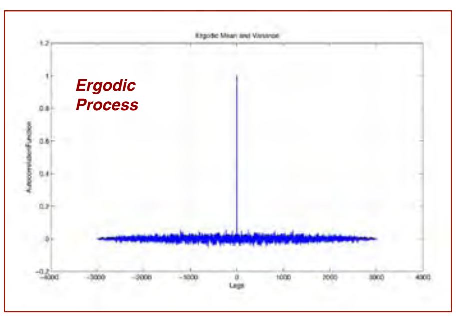
$$\phi_{xx}(k) = E\left(x_n x_{n+k}\right) = \lim_{N \to \infty} \sum_{n=1}^{N} \left(x_n x_{n+k}\right)$$

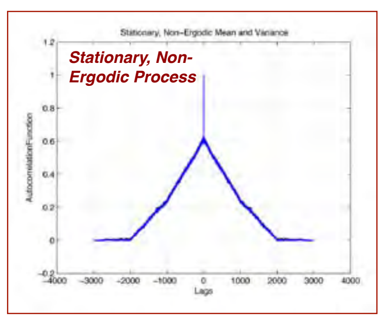
For finite length, number of products in correlation function estimate <u>decreases</u> as number of lags <u>increases</u>

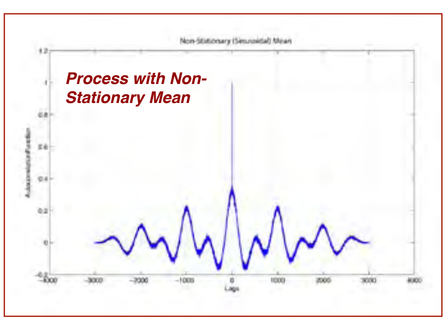
Time, sec	Phi(k)	# of Products		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1				
n				0	1	2	3	4	5	6	7	8	9	10				
x(n)	-			0.000	0.050	0.100	0.149	0.199	0.247	0.296	0.343	0.389	0.435	0.479	-		-	
x(n)	0.911	11		0.000	0.050	0.100	0.149	0.199	0.247	0.296	0.343	0.389	0.435	0.479	-		-	
x(n+1)	0.785	10	-		0.000	0.050	0.100	0.149	0.199	0.247	0.296	0.343	0.389	0.435		0.479	-	
x(n+2)	0.659	9	-	-		0.000	0.050	0.100	0.149	0.199	0.247	0.296	0.343	0.389		0.435		0.479

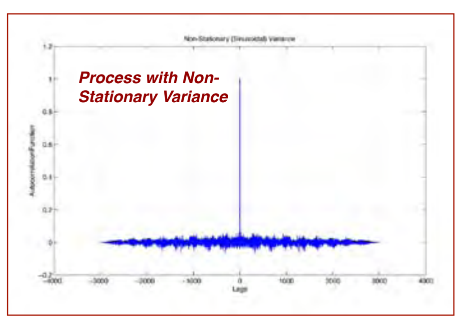
$$\text{Approximate correction: multiply } \phi \text{ by } \frac{n+k}{n}$$

AutoCovariance Functions for Previous Examples

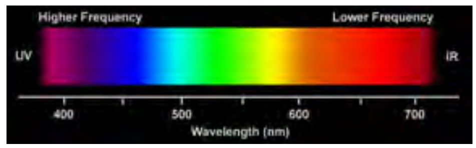






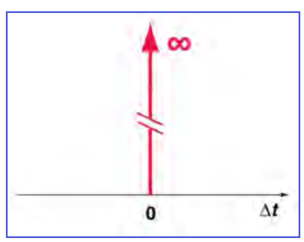


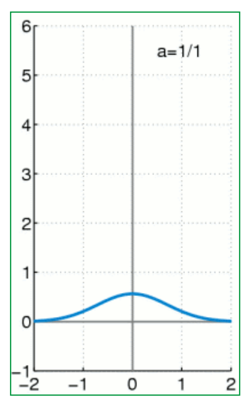
The Color of Noise



Dirac Delta Function

$$\delta(\Delta t) = \begin{cases} \infty, & \Delta t = 0 \\ 0, & \Delta t \neq 0 \end{cases}$$



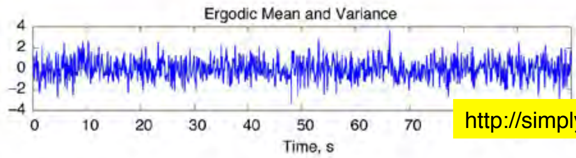


$$\int_{-\infty}^{\infty} \delta(t)\,dt = \lim_{\Delta t_1 \to 0} \int_{-\Delta t_1}^{\Delta t_1} \delta(\Delta t)\,d(\Delta t) = 1$$

Representation of Dirac delta function as Gaussian distribution with vanishing standard deviation

$$\delta(\Delta t) = \lim_{a \to 0} \frac{1}{a\sqrt{\pi}} e^{-\left(\frac{\Delta t}{a}\right)^2}$$

White Noise



http://simplynoise.com

1) Autocovariance function at zero lag

$$\phi_{xx}(0) = \sigma_x^2$$

2) Autocovariance function at non-zero lag = 0

3) Model using Dirac delta function

$$\phi_{xx}(\Delta t) = \phi_{xx}(0)\delta(\Delta t) = \sigma_x^2 \delta(\Delta t)$$

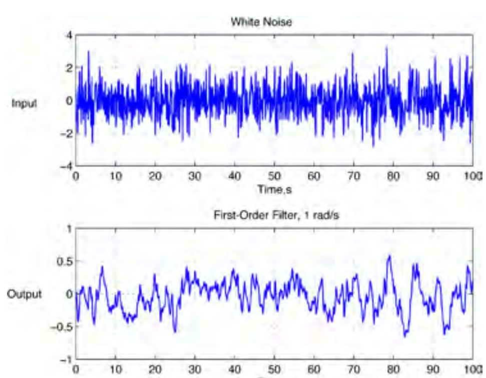
Conditional probability distribution = Unconditional probability distribution

$$\text{pr}\left[x(t) \middle| x(t+\Delta t)\right] = \text{pr}\left[x(t)\right]$$

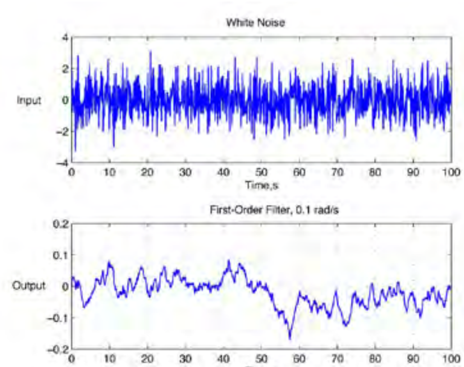
Colored Noise*: First-Order Filter Effects

$$\dot{x} = ax - au, \quad a < 0$$

***a* = –1 rad/s**



***a* = –0.1 rad/s**



**** = non-white noise***

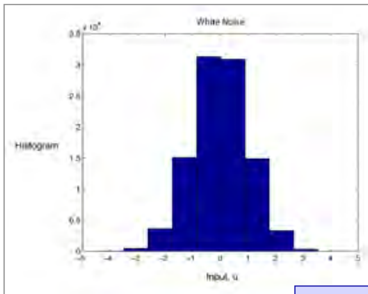
http://en.wikipedia.org/wiki/Colors_of_noise

http://yusynth.net/Modular/EN/NOISE/

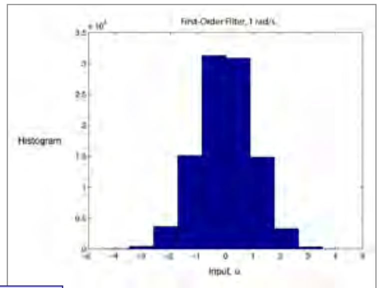
Colored Noise: Probability Density Functions

$$\dot{x} = -x + u$$

Input Histogram Function



Output Histogram Function



t_f **= 10,000 sec**

Probability density functions are virtually identical

Markov Sequences and Processes

- **Markov Sequence (Discrete Time)**
  - **Probability distribution of dynamic process at time $t_{k+1} > t_k > 0$, conditioned on the past history depends only on the state, x, at time t_k**

$$\Pr\left[x_{k+1} \mid \left(x_k, x_{k-1}, x_{k-2}, \cdots, 0\right)\right] = \Pr\left[x_{k+1} \mid x_k\right]$$

- **Markov Process (Continuous Time)**
  - **Probability distribution of dynamic process at time $s > t > 0$, conditioned on the past history depends only on the state, x, at time t**

Stochastic Steady State

Scalar LTI system with zero-mean Gaussian random input

$$x_{i+1} = ax_i + (1-a)u_i, \quad 0 < a < 1$$

Expected value

$$E(x_{i+1}) = E\left[ax_i + (1-a)u_i\right], \quad 0 < a < 1$$
$$= aE(x_i) + (1-a)E(u_i)$$

Expected value at stochastic steady state

$$E(x_{i+1}) = E(x_i)$$
$$(1-a)E(x_i) = (1-a)E(u_i)$$
$$E(x_i) = E(u_i) = 0$$

Autocovariance of a Markov Sequence

Scalar LTI system with zero-mean Gaussian random input

$$\phi_{xx}(1) = E(x_i x_{i+1}) = E\left[x_i a(x_i - u_i)\right]$$
$$= a\left[E(x_i^2) - E(x_i u_i)\right]$$

If input is white noise, $E(x_i u_i) = 0$

$$\phi_{xx}(1) = aE(x_i^2) = a\sigma_x^2$$

$$\phi_{xx}(k) = a^{|k|}\sigma_x^2, \quad 0 < a < 1$$

Autocovariance of a Markov Process

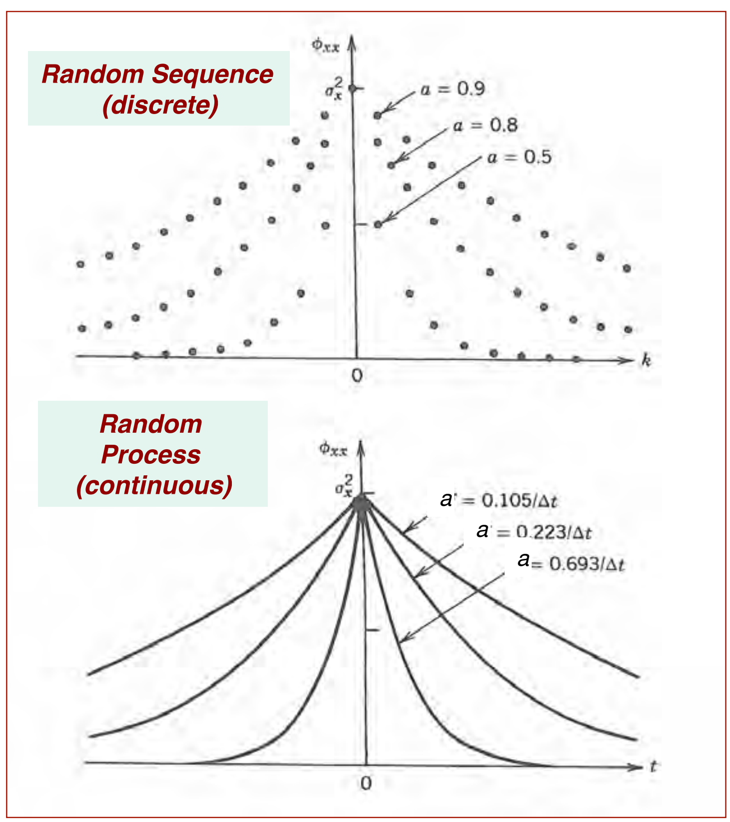
Expected value at stochastic steady state

$$\dot{x} = ax - au = a(x-u), \quad a<0$$
$$E(\dot{x}) = a\left[E(x) - E(u)\right] \xrightarrow[t\to\infty]{} 0$$
$$E(x) \xrightarrow[t\to\infty]{} E(u)$$

Autocovariance function of Markov Process

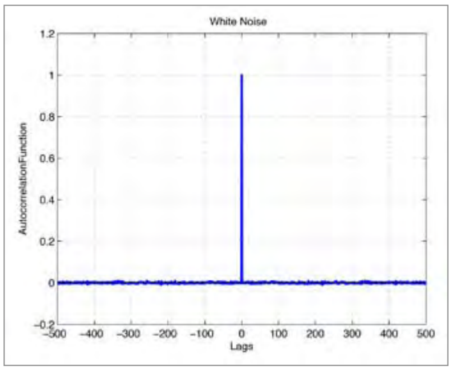
$$\phi_{xx}(\Delta t) = E\left[x(t)x(t+\Delta t)\right] = E\left[x(t)e^{a|\Delta t|}x(t)\right]$$
$$= e^{a|\Delta t|}E\left(x^2\right) = e^{a|\Delta t|}\sigma_x^2, \quad a<0$$

AutoCovariance Functions of Markov Sequences and Processes

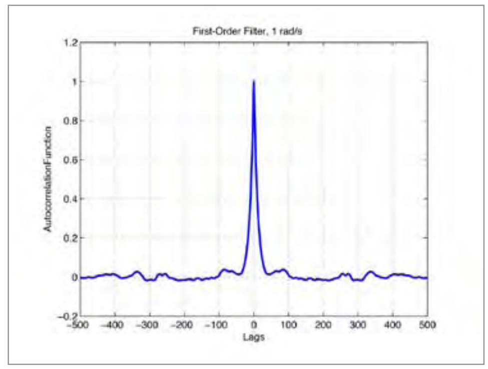


Colored Noise: First-Order Filter
a = –1 rad/s

Input AutoCovariance Function



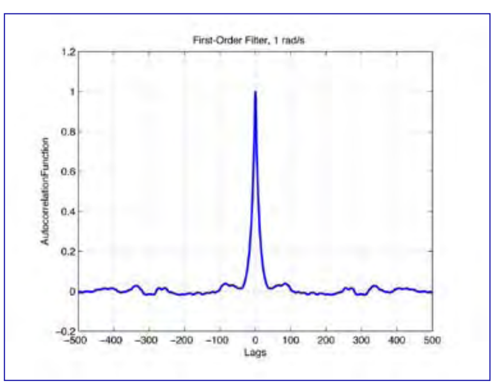
Output AutoCovariance Function



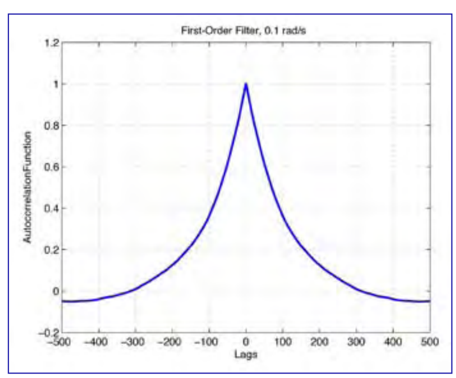
t_f = 10,000 sec

Colored Noise: Comparison of AutoCovariance Functions for 1st-Order Filters

a = –1 rad/s



a = –0.1 rad/s



t_f = 10,000 sec

Fourier Transform and Its Inverse

Fourier transform of *x(t)*

$$\boxed{F\left[x(t)\right] = X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt, \quad \omega = frequency, rad\,/\,s}$$

Inverse Fourier transform

$$\boxed{x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}\,d\omega}$$

$x(t)$ and $X(j\omega)$ are a Fourier transform pair

Power Spectral Density Function of a Random Process

Frequency distribution of the complex amplitude in *x*(*t*)

$$\boxed{X(\omega) = X_{\mathrm{Re}}(\omega) + jX_{\mathrm{Im}}(\omega)}$$

Complex conjugate

$$\boxed{X^*(\omega) = X_{\mathrm{Re}}(\omega) - jX_{\mathrm{Im}}(\omega)}$$

Power spectral density function

$$\begin{aligned} \left|X(\omega)\right|^2 &= X(\omega)X^*(\omega) \\ &= \left[X_{\mathrm{Re}}(\omega) + jX_{\mathrm{Im}}(\omega)\right]\left[X_{\mathrm{Re}}(\omega) - jX_{\mathrm{Im}}(\omega)\right] \\ &= X_{\mathrm{Re}}^{\;2}(\omega) + X_{\mathrm{Im}}^{\;2}(\omega) \triangleq \Psi_{xx}(\omega) \end{aligned}$$

Power Spectral Density and Autocorrelation Functions Form a Fourier Transform Pair

For a stationary process,

Power spectral density function

$$\Psi_{xx}(\omega) = \int_{-\infty}^{\infty} \psi_{xx}(\tau) e^{-j\omega\tau}\, d\tau$$

Autocorrelation function

$$\psi_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi_{xx}(\omega) e^{j\omega\tau}\, d\omega$$

Power Spectral Density and Autocovariance Functions for Variation from the Mean

for

$$\tilde{x}(t) = x(t) - \bar{x}$$

Power spectral density function

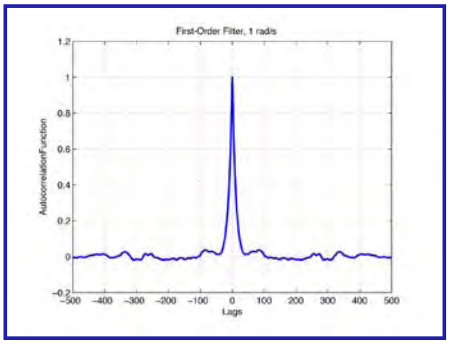
$$\Phi_{xx}(\omega) = \int_{-\infty}^{\infty} \phi_{xx}(\tau) e^{-j\omega\tau}\, d\tau$$

Autocovariance function

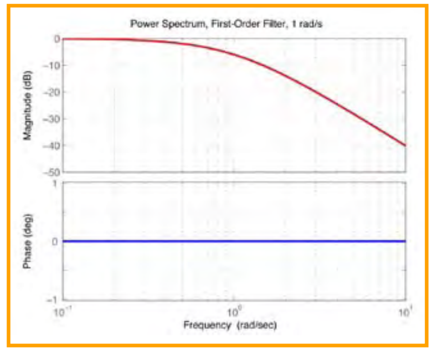
$$\phi_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_{xx}(\omega) e^{j\omega\tau}\, d\omega$$

Power Spectral Density and Autocovariance Functions for 1st-Order Filter Output, *a* = –1 rad/s, White-Noise Input

Autocovariance Function



Power Spectral Density Function



Power spectral density function has magnitude but not phase angle

Spectral Characteristics of White Noise

A

$$\phi_{xx}(0)=\sigma_x^2=\frac{1}{\pi}\int_0^\infty \Phi_{xx}(\omega)e^{j\omega((0)}\,d\omega=\frac{1}{\pi}\int_0^\infty \Phi_{xx}(\omega)\,d\omega$$

Which is $(1/\pi)\times area$ under the power spectral density curve

For white noise,

B

$$\Phi_{xx}(\omega)=\phi_{xx}(0)\int_{-\infty}^{\infty}\delta(\tau)e^{-j\omega\tau}\,d\tau=\phi_{xx}(0)=\ \textbf{constant}$$

However, **A** is infinite except when the constant integrand is zero

White Noise and Band-Limiting

- **Heuristic arguments for use of "white noise":**
  - **Real systems are band-limited by dynamics**
  - **Discrete-time systems are band-limited by sampling frequency**

Assume that power spectral density has a sharp cutoff at ω_B

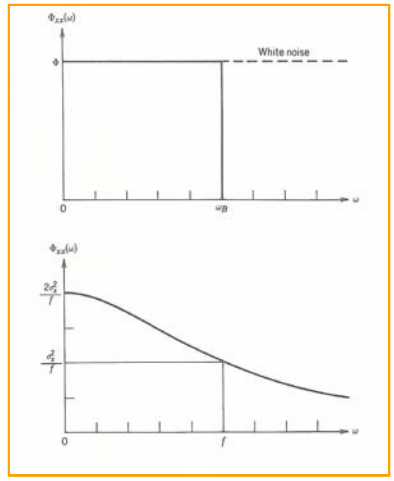
$$\Phi_{xx}(\omega) = \begin{cases} \Phi, & |\omega| \le \omega_B \\ 0, & |\omega| > \omega_B \end{cases}$$

$$\sigma_x{}^2 = \Phi\omega_B / 2\pi$$

$$\phi_{xx}(\tau) = \Phi \sin \omega_B \tau / 2\pi\tau$$

- **If $x(t)$ is a Markov process**

$$\Phi_{xx}(\omega) = \frac{2 f \sigma_x^2}{\omega^2 + f^2}$$



When power spectral density is plotted against frequency rather than log(frequency), rolloff does not seem as abrupt

Multivariate Statistics and Propagation of Uncertainty

Multivariate Expected Values:
Mean Value Vector and Covariance Matrix

Mean value vector of measurement error

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

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

Covariance matrix of measurement error

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

If the error is Gaussian, its probability distribution is

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

Covariance Matrix is Expected Value of the Outer Product

$$\mathbf{R} = E\left[(\mathbf{n}-\bar{\mathbf{n}})(\mathbf{n}-\bar{\mathbf{n}})^T\right]$$

$$= \begin{bmatrix} \sigma_{n_1}{}^2 & \rho_{12}\sigma_{n_1}\sigma_{n_2} & \dots & \rho_{1r}\sigma_{n_1}\sigma_{n_r} \\ \rho_{12}\sigma_{n_1}\sigma_{n_2} & \sigma_{n_2}{}^2 & \dots & \rho_{2r}\sigma_{n_2}\sigma_{n_r} \\ \dots & \dots & \dots & \dots \\ \rho_{1r}\sigma_{n_1}\sigma_{n_r} & \rho_{2r}\sigma_{n_2}\sigma_{n_r} & \dots & \sigma_{n_r}{}^2 \end{bmatrix}$$

$\sigma_{n_1}{}^2 = \textit{Variance of } n_1$

$\rho_{12} = \textit{Correlation coefficient for } n_1 \textit{ and } n_2$

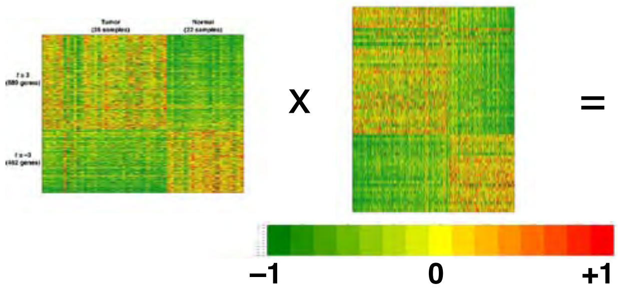
$\rho_{12}\sigma_{n_1}\sigma_{n_2} = \textit{Covariance of } n_1 \textit{ and } n_2$

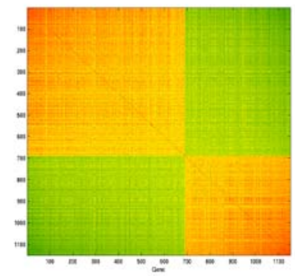
Gaussian probability distribution is completely described by its mean value and covariance matrix

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

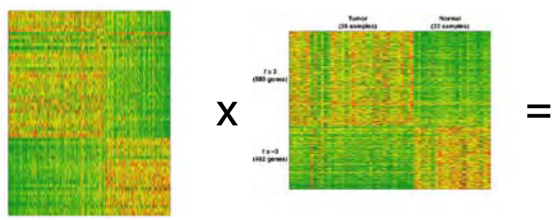
Covariance Matrix is Expected Value of the Outer Product

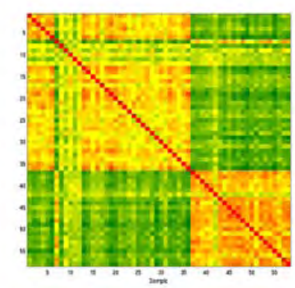
- Transcript correlation
  - Identifies possible components of biological circuits





- Sample tissue correlation
  - Confirms or questions the classification of samples





Mean Values and Covariances of State and Disturbance Vectors

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

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

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

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

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

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

If the disturbance is Gaussian, its probability distribution is

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

Uncertain Linear, Time-Invariant Dynamic Model

Linear, time-invariant, discrete-time model with known coefficients

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

Gaussian uncertainty models for initial condition and disturbance

$$\begin{array}{c}
\bar{\mathbf{x}}_0 = E[\mathbf{x}_0]; \quad \mathbf{P}_0 = E\left\{[\mathbf{x}_0 - \bar{\mathbf{x}}_0][\mathbf{x}_0 - \bar{\mathbf{x}}_0]^T\right\} \\
\bar{\mathbf{x}}_k = E[\mathbf{x}_k]; \quad \mathbf{P}_k = E\left\{[\mathbf{x}_k - \bar{\mathbf{x}}_k][\mathbf{x}_k - \bar{\mathbf{x}}_k]^T\right\} \\
\bar{\mathbf{w}}_k = \mathbf{0}; \quad \mathbf{Q}_k = E\left\{[\mathbf{w}_k][\mathbf{w}_k]^T\right\}
\end{array}$$

Control is assumed to be known

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

Propagation of the Ensemble Mean Value Estimate

Mean value of the state

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

Extrapolation

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

Propagation of the Ensemble State Covariance Estimate

State covariance matrix

Expected values of cross terms are zero

$$E\left\{\left[\mathbf{x}_{k+1}-\overline{\mathbf{x}}_{k+1}\right]\left[\mathbf{x}_{k+1}-\overline{\mathbf{x}}_{k+1}\right]^T\right\}=$$

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

$$\begin{aligned}\mathbf{P}_{k+1}&=E\left\{\boldsymbol{\Phi}\left[\mathbf{x}_k-\overline{\mathbf{x}}_k\right]\left[\mathbf{x}_k-\overline{\mathbf{x}}_k\right]^T{}_k\,\boldsymbol{\Phi}^T+\mathbf{0}+\boldsymbol{\Lambda}\mathbf{w}_k\mathbf{w}^T{}_k\boldsymbol{\Lambda}^T{}_k\right\}\\&=\boldsymbol{\Phi}E\left\{\left[\mathbf{x}_k-\overline{\mathbf{x}}_k\right]\left[\mathbf{x}_k-\overline{\mathbf{x}}_k\right]^T{}_k\right\}\boldsymbol{\Phi}^T+\boldsymbol{\Lambda}E\left(\mathbf{w}_k\mathbf{w}^T{}_k\right)\boldsymbol{\Lambda}^T\end{aligned}$$

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

Probability Density Function of the State Estimate

Uncertainty propagation model

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

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

Probability density function of the state is Gaussian

$$\mathrm{pr}(\mathbf{x}_k)=\frac{1}{(2\pi)^{n/2}\left|\mathbf{P}_k\right|^{1/2}}e^{-\frac{1}{2}(\mathbf{x}_k-\overline{\mathbf{x}}_k)^T\mathbf{P}_k^{-1}(\mathbf{x}_k-\overline{\mathbf{x}}_k)}$$

Propagating the state mean and covariance is equivalent to propagating the entire probability density function of the state

Example: Propagating a Scalar Probability Density Function

Scalar LTI system with zero-mean Gaussian random input

$$x_{k+1} = bx_k + \sqrt{1-b^2}\,u_k + \sqrt{1-b^2}\,w_k\,, \quad x_0 \ given$$

Propagation of the mean value

$$\bar{x}_{i+1} = b\bar{x}_i + \sqrt{1-b^2}\,\bar{u}_i\,, \quad \bar{x}_0 \ given$$

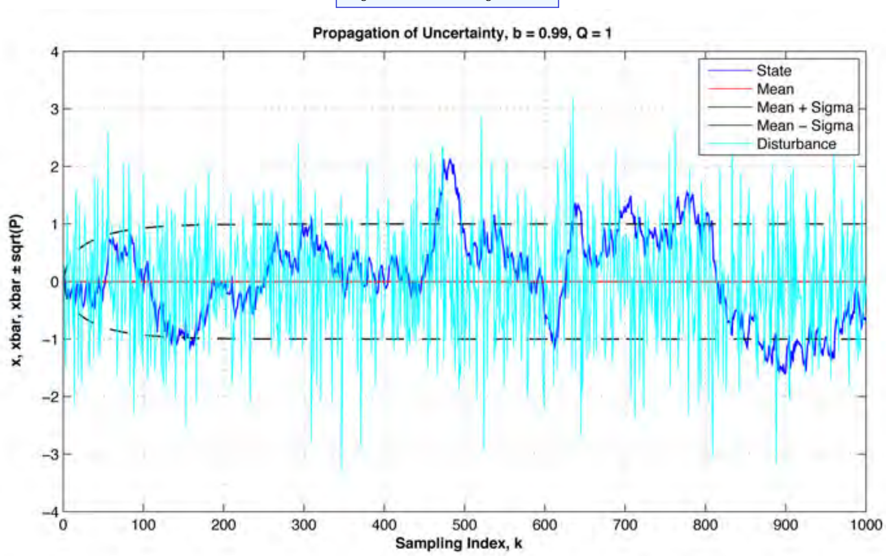
Propagation of the variance

$$P_{k+1} = b^2 P_k + \left(1-b^2\right)Q_k\,, \quad P_0 \ given$$

$$Q_k = E\left(w_k^2\right)$$

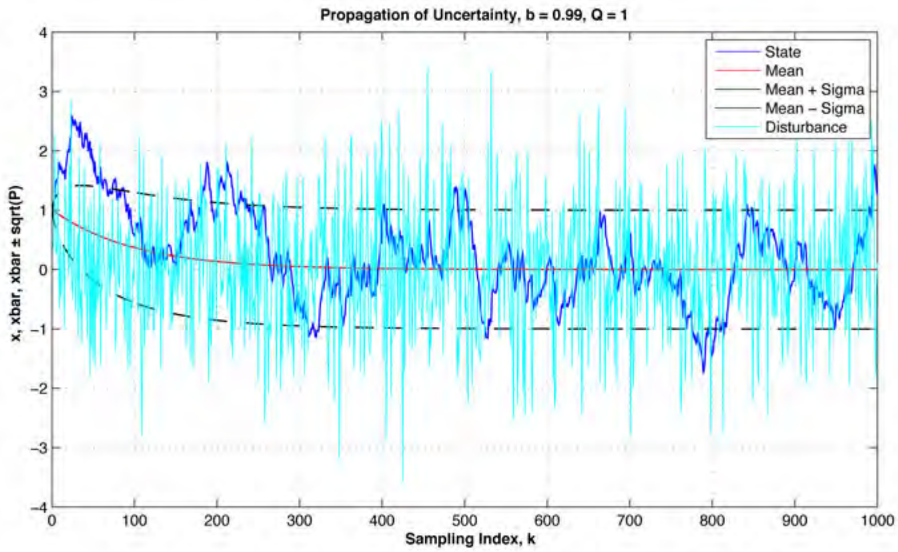
Example: Propagating a Scalar Probability Density Function

$$x_0 = 0; \quad P_0 = 0$$



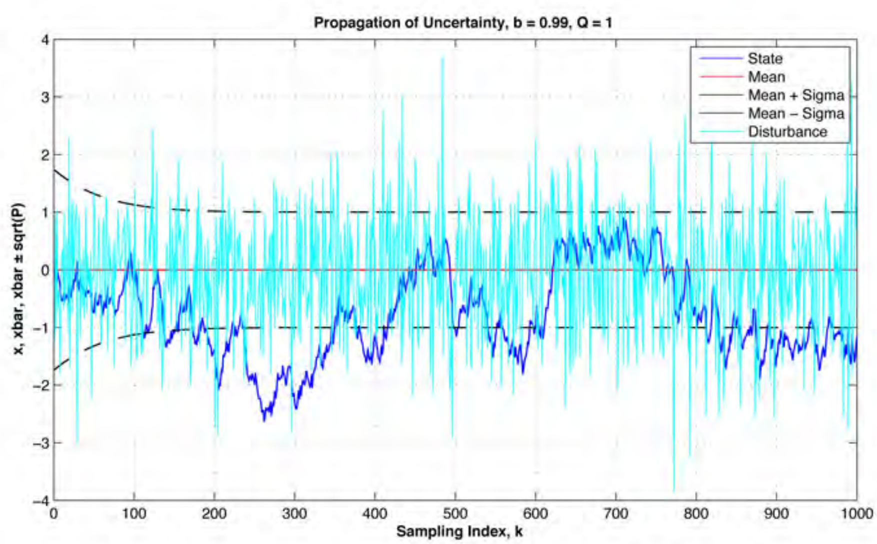
Example: Propagating a Scalar Probability Density Function

$$x_0 = 1; \quad P_0 = 0$$



Example: Propagating a Scalar Probability Density Function

$$x_0 = 0; \quad P_0 = 3$$



Next Time:
Least-Squares Estimation

Supplemental Material

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

(Papoulis)

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

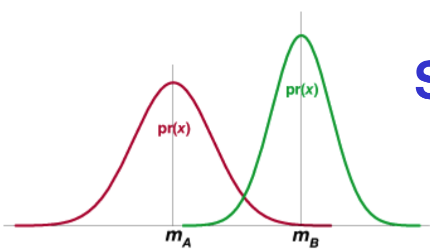
$$\Pr(A_i) = \frac{n_{A_i}}{N} = \frac{1}{11}$$

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

$$\Pr(A_i) = \frac{n_{A_i}}{N} = \frac{3}{21} = \frac{1}{7}$$

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

$$\Pr(A_i) = \frac{n_{A_i}}{N} = \frac{6}{36} = \frac{1}{6}$$



Simple Hypothesis Test: *t* Test

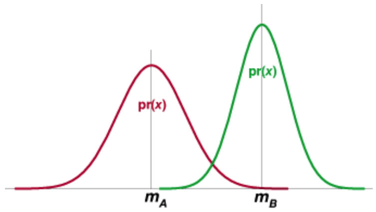
Is *A* different from *B*?

- *t* compares mean value difference of two data sets, normalized by standard deviations
  - $|t|$ is reduced by uncertainty in the data sets (s)
  - $|t|$ is increased by number of points in the data sets (n)

$$t = \frac{(m_A - m_B)}{\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}}$$

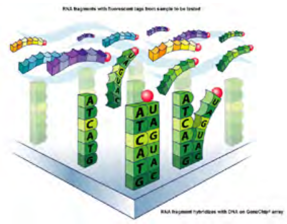
m :	Mean value of the data set
σ :	Standard deviation of data set
n :	Number of points in data set

- $|t| > 3$, $m_A \neq m_B$ with $\geq 99.7\%$ confidence (error probability ≤ 0.003 for Gaussian distributions) [$n > 25$]

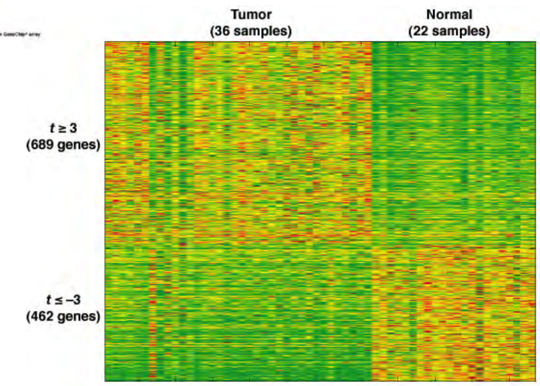


Application of *t* Test to DNA Microarray Data

(Data from Alon *et al*, 1999)



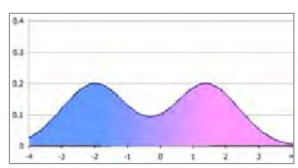
$$t = \left(m_T - m_N\right) \Big/ \sqrt{\frac{\sigma_T^2}{36} + \frac{\sigma_N^2}{22}}$$



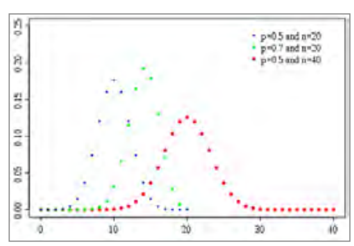
- 58 RNA samples representing tumor and normal tissue
- 1,151 transcripts are over/under-expressed in tumor/normal comparison ($p \leq 0.003$)
- Genetically dissimilar samples are apparent

Some Non-Gaussian Distributions

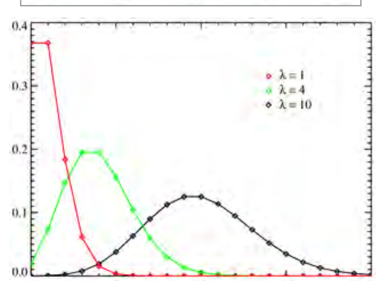
- **Bimodal Distribution**
  - Two Peaks
  - Often the sum of two unimodal distributions



- **Binomial Distribution**
  - Probability of *k* successes in *n* trials
  - Discrete probability distribution described by a probability mass function

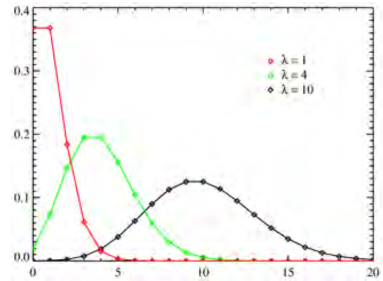


- **Poisson Distribution**
  - Probability of a number of events occurring in a fixed period of time
  - Discrete probability distribution described by a probability mass function



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

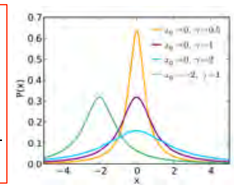
$\mathrm{pr}(k) =$ probability of k occurrences (per unit time)

$\sim$ 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

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

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

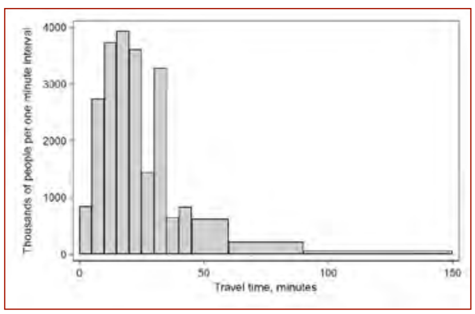


Histograms (Wikipedia Example)

Scaled Numbers

Area = Sum (+)

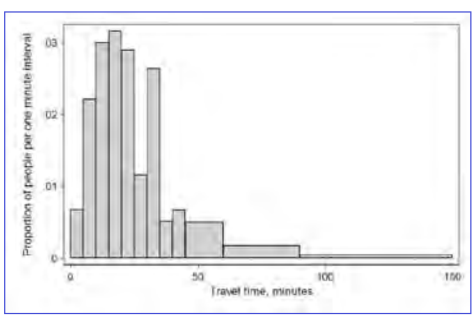
Interval	Width	Quantity	Quantity/width
0	5	4180	836
5	5	13687	2737
10	5	18618	3723
15	5	19634	3926
20	5	17981	3596
25	5	7190	1438
30	5	16369	3273
35	5	3212	642
40	5	4122	824
45	15	9200	613
60	30	6461	215
90	60	3435	57



Normalized Numbers

Area = 1

Interval	Width	Quantity (Q)	Q/total/width
0	5	4180	0.0067
5	5	13687	0.0221
10	5	18618	0.0300
15	5	19634	0.0316
20	5	17981	0.0290
25	5	7190	0.0116
30	5	16369	0.0264
35	5	3212	0.0052
40	5	4122	0.0066
45	15	9200	0.0049
60	30	6461	0.0017
90	60	3435	0.0005



Sufficient Conditions for Central Limit Theorem

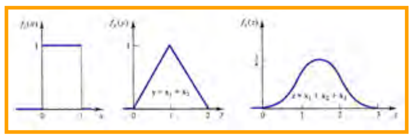
(Papoulis, 1990)

- The probability distribution of the <u>sum of independent, identically distributed *(i.i.d.)* variables</u> approaches a normal distribution as the number of variables approaches infinity

$$\boxed{\text{Random variables, } x_i\text{, are independent, identically distributed } (i.i.d.) \text{ and } E\left(x_i^3\right) \text{ is finite}}$$

$$\boxed{\text{Random variables, } x_i\text{, are bounded, } |x_i| < A < \infty, \text{ and } \sigma_i > a > 0}$$

$$\boxed{E\left(x_i^3\right) < B < \infty, \text{ and } \sum \sigma_n^2 \xrightarrow[n\to\infty]{} \infty}$$



Which Combinations are Possible?

Independence and lack of correlation

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

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

Independence and correlation

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

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

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

Dependence and lack of correlation

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

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

Dependence and correlation

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

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

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

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

Propagation of state mean and covariance

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

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

Off-diagonal elements of P and Q express correlation

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

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

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

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

Dependence and lack of correlation in nonlinear output

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

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

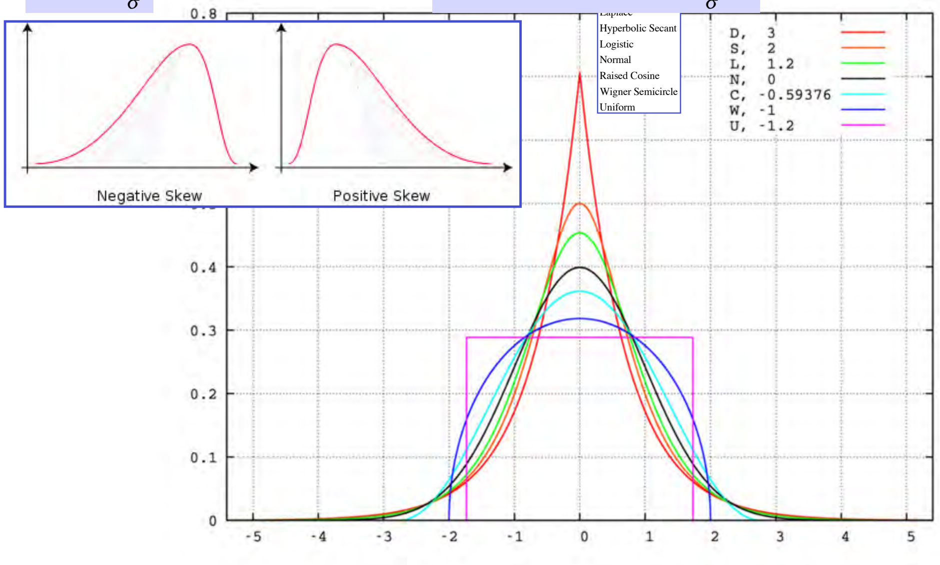
Dependence and correlation in state

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

Skewness and Kurtosis of Probability Distributions
(3^{rd} and 4^{th} Central Moments)

$\text{Skewness} = \dfrac{\mu_3}{\sigma^3}$

$\text{Excess Kurtosis} = \text{Kurtosis} - 3 = \dfrac{\mu_4}{\sigma^4} - 3$



Two Trials to Show Mean, Standard Deviation, Kurtosis, and Skewness

Stationary, Ergodic Mean and Variance
Mean = 0.0068123
Standard Deviation = 1.0006
Kurtosis = 3.0291
Skewness = -0.0093799

Stationary, Non-Ergodic Mean and Variance
Mean = 0.99847
Standard Deviation = 1.3156
Kurtosis = 2.5933
Skewness = -0.12198

Non-Stationary (Sinusoidal) Mean
Mean = 0.17438
Standard Deviation = 1.2058
Kurtosis = 2.904
Skewness = -0.042752

Non-Stationary (Sinusoidal) Variance
Mean = -0.0043818
Standard Deviation = 0.6804
Kurtosis = 4.6055
Skewness = 0.063442

Sine Wave
Mean = 6.6818e-06
Standard Deviation = 0.707
Kurtosis = 1.5006
Skewness = -2.8338e-05

Stationary, Ergodic Mean and Variance
Mean = 0.0050475
Standard Deviation = 0.99947
Kurtosis = 3.0485
Skewness = 0.0147

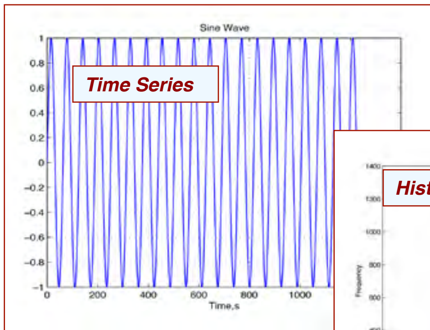
Stationary, Non-Ergodic Mean and Variance
Mean = 0.99337
Standard Deviation = 1.3262
Kurtosis = 2.6395
Skewness = -0.18208

Non-Stationary (Sinusoidal) Mean
Mean = 0.18366
Standard Deviation = 1.1833
Kurtosis = 2.83
Skewness = -0.10543

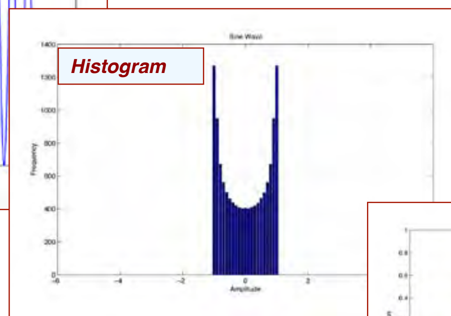
Non-Stationary (Sinusoidal) Variance
Mean = 0.0072196
Standard Deviation = 0.71168
Kurtosis = 5.0192
Skewness = -0.05492

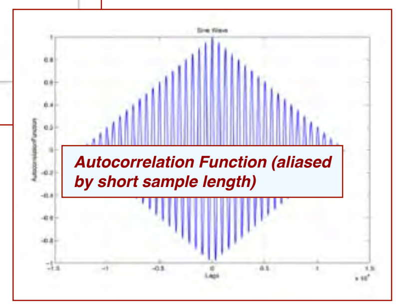
Sine Wave
Mean = 6.6818e-06
Standard Deviation = 0.707
Kurtosis = 1.5006
Skewness = -2.8338e-05

Sampled Sine Wave Statistics



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

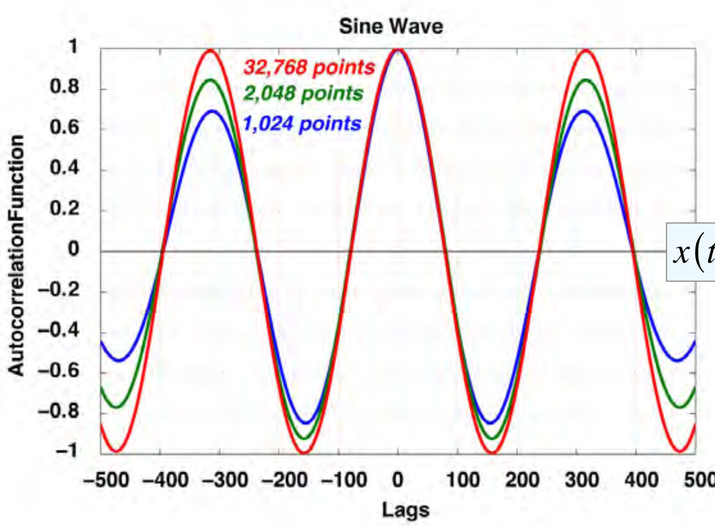




$$x(t) = \sin(0.1t)$$
$$\Delta t = 0.1\ s, \quad \#\,\text{points} = 1258$$

Bi-modal distribution

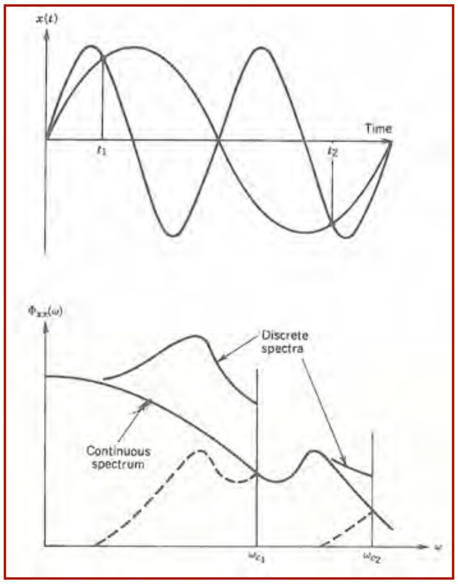
Sample-Length Effect on Estimate of Sine Wave Autocorrelation Function



$$x(t) = \sin(t/16\pi)$$

Autocorrelation function estimate becomes inaccurate when number of lags > ~10% of the number of sample points

Frequency Folding and Aliasing



τ : Sampling interval, sec

$\omega_D = \omega_C \tau$, rad/sec

$\omega_C = \dfrac{\pi}{\tau} =$ **Folding frequency**

- **Discrete Fourier transform is periodic in frequency increments of $\Delta\omega_D = 2\pi$**
- **Relationship of discrete Fourier transform to continuous Fourier transform**

$$X_D(\omega_C) = \frac{1}{\tau}\sum_{n=-\infty}^{\infty} X_C\left(\omega + \frac{2\pi n}{\tau}\right)$$

- **Distortion of signal and spectrum is called aliasing**

Propagation Program

```
%   Scalar Propagation of Mean and Variance
%   Rob Stengel
%   4/5/2011

    clear
    '=========================='
    'Propagation of Uncertainty'
    date
    b       =   0.99; b2      =   b*b;bb      =   1 - b^2;sqrb    =   sqrt(bb)
    w       =   []; x       =   []; xbar    =   [];
    x(1)    =   0;xbar(1) =   x(1)
    P(1)    =   3
    Q       =   1
    u       =   0
    for  k  =   1:999
        w(k)      =   randn(1);
        x(k+1)    =   b*x(k) + sqrb*u + sqrb*w(k);
        xbar(k+1) =   b*xbar(k) + sqrb*u;
        P(k+1)    =   b2*P(k) + bb*Q;
    end
    k       =   [1:1000];
    w(1000) =    randn(1);
    figure
    plot(k,x,'b',k,xbar,'r',k,(xbar+sqrt(P)),'k',k,(xbar-sqrt(P)),'k',k,w,'c'),grid
    title('Propagation of Uncertainty, b = 0.99, Q = 1'), xlabel('Sampling Index,
k'), ylabel('x, xbar, xbar ? sqrt(P)')
    legend('State','Mean','Mean + Sigma','Mean - Sigma','Disturbance')
```