Boston University

Department of Electrical and Computer Engineering

EC505 – STOCHASTIC PROCESSES

Information Sheet

Fall 2016

Professor:	Venkatesh Saligrama Rm 438, 8 St. Mary's Street 353-1040, **srv@bu.edu**
Office hours:	Monday, 2–3 pm; Friday, 4–6 PM
Class:	MW 12-2, PHO 202
Web Site:	http://learn.bu.edu/
Required text:	None
Notes:	*Course Notes on Stochastic Processes* by D. A. Castañon & W. C. Karl available from the class web site. In addition, other materials will be handed out throughout the term.
Prerequisites:	MA381 or EK500, Introduction to Probability SC401, Signals and Systems MA142, Linear Algebra In general the course assumes a fluency in linear systems as well as basic probability. A facility with linear algebra is strongly recommended and helpful. The subject material demands a high level of maturity, dedication, and commitment to understanding the concepts in depth.
Homework:	Homework will be assigned roughly weekly. They are for you to clear up your confusions with the material through extended thought, to develop proficiency through practice, and to learn the concepts. They must be handed in to me by the date they are due. No late homework will be accepted. Doing the homework will be essential to your understanding of the material. Do not wait till the last minute before doing the homework!
Exams:	There will be 2 exams during the semester and a final during the final exam period. Midterm 1 – October 10 PHO 202 Midterm 2 – November 14 PHO 202 Final – TBD
Grading Policy:	Homework: 20% Midterm 1: 25% Midterm 2: 25% Final: 30%

Course Policies

Academic Conduct

The student handbook defines Academic Misconduct as follows: “Academic misconduct occurs when a student intentionally misrepresents his or her academic accomplishments or impedes other students’ chances of being judged fairly for their academic work. Knowingly allowing others to represent your work as theirs is as serious an offense as submitting another’s work as your own.” This basic definition applies to SC505. If you are ever in doubt as to the legitimacy of an action, please talk to me immediately. The penalities for plagiarism at BU are severe.

Make-ups

There will be no make-up exams. If you have a legitimate excuse, such as illness as documented by a doctor’s note, then the scores of your other exams will be weighted more highly to compensate for the missed exam. If you do not have a legitimate excuse, you will be given a grade of zero for the exam.

Incompletes

Incompletes will not be given to students who wish to improve their grade by taking the course in a subsequent semester. An incomplete may be given for medical reasons where a doctor’s note is provided. The purpose of incompletes are to allow a student *who has essentially completed the course* and who has a legitimate interruption in the course, to complete the remaining material in another semester. Students will not be given an opportunity to improve their grade by doing ”extra work”.

Homework, Dates, Etc.

Homeworks are due by the end of the day on the date stated. Late homeworks will not be accepted. No homework scores will be dropped.

Students are responsible for being aware of the drop dates for the current semester. Drop forms will not be back-dated.

Syllabus
EC505 STOCHASTIC PROCESSES
Fall 2016

Topic	# Lectures	Reading Notes
I. Probability review	3	1, 2.1-2.2
probability space, axioms, definitions		
random variables, bernoulli, poisson, gaussian rvs		
random vectors, multivariate gaussian, properties		
II. Definition and characterization of random processes	6	2–4
distribution description		
moments, important classes of processes,		
IIP process, Brownian motion, markov chains,		
time averages, stationarity and ergodicity		
mean square calculus, power spectral density		
Exam review & exam	1	
III. System response with random signals	4	5-7
LTI system response, Shaping filters		
discrete time linear models and system identification		
modulation & sampling		
IV. Signal detection	5	8, 9
detection/classification of a random variable		
detection of vectors & disc. time signals		
spectral decomp., Karhunen-Loève expansions		
detection of continuous time signals		
Exam review & exam	2	
V. Estimation	7–9	10,11,12
estimation of a random variables, Bayesian estimation		
estimation of a nonrandom variables, Maximum-likelihood		
Kalman filtering, HMMs and Graphical Models		
Course review	1	
Final Exam		

Reference Texts

1. H. Stark and J. W. Woods, *Probability Random Processes and Estimation Theory for Engineers*, Prentice-Hall, 1986. Nice alternative to text for some topics in the course, especially early on. On reserve.

2. K. Sam Shanmugan, *Random Signals: Detection, Estimation, and Data Analysis*, Wiley, 1988. On reserve.

3. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, 1991. On reserve.

4. R. M. Gray and L. D. Davisson, *Random Processes: A Mathematical Approach for Engineers*, Prentice-Hall, 1986. Bridges the gap between formal mathematical texts and engineering texts on probability theory. On reserve.

5. A. Drake, *Fundamentals of Applied Probability Theory*, McGraw-Hill, 1967. Basic engineering text on probability theory.

6. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vols. I and II, Wiley, 1968. Valuable formal reference set on probability theory.

7. S. M. Kay, *Fundamentals of Statistical Signal Processing and Estimation Theory*, Prentice-Hall, 1993. Accessible and thorough treatment of estimation theory.

8. E. Lee and D. G. Messerschmitt, *Digital Communication*, Kluwer Academic, 1988. Advanced reading on applications in communication theory

9. M. Loeve, *Probability Theory I*, Springer-Verlag, fourth ed., 1977. Formal but reasonably readable treatment of probability theory. A classic.

10. A. V. Oppenheim and R. W. Schafer,*Discrete-Time Signal Processing*, Prentice-Hall, 1989. Standard text on discrete-time linear systems and signals.

11. A. V. Oppenheim and A.S. Willsky, *Signals and Systems*, Prentice-Hall, 1983. Basic undergraduate text on both continuous-time and discrete-time linear systems and signals.

12. E. Parzen,*Stochastic Processes*, Holden-Day, 1962. Classic, formal text on stochastic processes.

13. G. Strang, *Linear Algebra and its Applications*, Harcourt Brace Jovanovich, third ed., 1968. Standard reference text on linear algebra.

14. C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, 1992. Very accessible alternative to text for some topics in the course (all done in discrete-time).

15. H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I*, Wiley, 1968. Classic and valuable reference text on detection and estimation theory.

Boston University

Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Notes on Notation

Random Variables

- X: Random variable
- x: Value the vandom variable takes on
- $P_X(x)$: Probability distribution function (PDF) for a random variable. Also known as a cumulative distribution function (CDF).
- $p_X(x)$: Probability density function (pdf) for a continuous-valued random variable
- $p_X(x)$: Usually a probability mass function (pmf) for a discrete-valued random variable, but could be a pdf
- m_X or m_X or $E[X]$: Mean
- $\sigma_X^2 = E[X^2] - m_X^2 = \text{Var}(X)$: Variance of X
- $E_Y[Y] = E_X[g(X)]$: Expectation of function $Y = g(X)$. $E[Y] = \int y p_Y(y)\,dy = \int g(x)\,p_X(x)\,dx$
- Conditional expectation: $E_{X|Y}[X|Y]$
  - $E_{X|Y}[X|Y] = \int x\,p_{X|Y}(x|y)\,dx$
  - $E_{X|Y}[X|Y]$ is a random variable
  - $E_{X|Y}[X|Y = y]$ is a deterministic value
  - Note that: $E[X] = E_Y[E_{X|Y}[X|Y]]$
- Cross-moments:
  - $E[XY]$: Correlation
  - Orthogonal RVs: $E[XY] = 0$
  - $\text{Cov}(X, Y) = E[XY] - m_X m_Y$: Covariance
  - Uncorrelated RVs: $\text{Cov}(X, Y) = 0$ (also true for independent RVs)

Random Vectors

- $\underline{X}$ or sometimes $\mathbf{X}$: Random vector
- $\underline{x}$ or sometimes $\mathbf{x}$: Vector value a random vector takes on
- X_i: Element i of a random vector
- x_i: Value that element i of a random vector takes on
- $p_{\underline{X}}(\underline{x})$ or $p_{\mathbf{X}}(\mathbf{x})$: pdf for a continuous-valued random vector
- $m_{\underline{X}}$ or $m_{\underline{X}}$ or $E[\underline{X}]$: Mean
- $R_{XX} = R_X = E[\underline{X}\underline{X}^T]$: Autocorrelation matrix (diagonal for orthogonal elements)
- $\Sigma_X = \Lambda_X = C_X = K_X = R_{XX} - m_{\underline{X}} m_{\underline{X}}^T$: covariance matrix (diagonal for uncorrelated elements or independent elements).
- Cross-moments:
  - $R_{XY} = E[\underline{X}\underline{Y}^T]$: (Cross) Correlation matrix
  - Orthogonal Random vectors: $R_{XY} = 0$
  - $\Sigma_{XY} = \Lambda_{XY} = C_{XY} = K_{XY} = R_{XY} - m_{\underline{X}} m_{\underline{Y}}^T$: (Cross) Covariance matrix
  - Uncorrelated Random vectors: $\Sigma_{XY} = 0$ (also true for independent RVs)

Random Processes

- $\{X(t)\}$ or $\{X(n)\}$: Random process with t for continuous time and n for discrete time
- $X(t)$ or $X(n)$: Random variable corresponding to time t (or n) of the random process or perhaps the random process itself
- $x(t)$ or $x(n)$: Waveform "value" (i.e. sample path) that the random process takes on or the scalar value that the random variable at time t or n takes on
- $p_{X(t_1),\cdots,X(t_k)}(x_1,\cdots,x_k)$ or $p_{X(t_1),\cdots,X(t_k)}(\underline{x})$ or $p_X(x_1,\cdots,x_k;t_1,\cdots,t_k)$ or $p_X(\underline{x};\underline{t})$: k-th order marginal pdf for a random process. $p(\cdot)$ may be used for the pmf, i.e. discrete valued case
- $p_{X(t_1)|X(t_2)\cdots,X(t_k)}(x_1|x_2\cdots,x_k)$ or $p_X(x_1|x_2\cdots,x_k;t_1,\cdots,t_k)$: example of a conditional distribution. $p(\cdot)$ may be used for the pmf, i.e. discrete valued case
- $p_{X(t_1),\cdots,X(t_k)}(x_1,\cdots,x_k) = p_{X(t_1+\tau),\cdots,X(t_k+\tau)}(x_1,\cdots,x_k)$ or $p_X(x_1,\cdots,x_k;t_1,\cdots,t_k) = p_X(x_1,\cdots,x_k;t_1+\tau,\cdots,t_k+\tau)$ for all t_i: k-th order stationary
- Strictly stationary or strict sense stationary or SSS: k-th order stationary for all k, the complete distribution and any marginals do not change with a time shift
- $m_X(t)$ or $m_X(t)$ or $E[X(t)]$: Mean of the process. In general a function of time
- $R_{XX}(t,s) = R_X(t,s) = E[X(t)X(s)]$ ($R_{XX}(m,n)$ for discrete valued): Autocorrelation function. $R_{XX}(t,s) = A(t)\delta(t-s)$ for an orthogonal process
- $C_{XX}(t,x) = C_X(t,s) = K_{XX}(t,s) = K_X(t,s) \equiv R_{XX}(t,s) - m_X(t)m_X(s)$ ($C_{XX}(m,n)$ for discrete valued): Autocovariance function. $C_{XX}(t,s) = A(t)\delta(t-s)$ for an uncorrelated process
- Cross-moments
  - $R_{XY}(t,s) = E[X(t)Y(s)]$ ($R_{XY}(m,n)$ for discrete valued): Cross-correlation function.
  - $X(t)$ and $Y(t)$ orthogonal processes: $R_{XY}(t,s) = 0$
  - $C_{XY}(t,x) = K_{XY}(t,s) \equiv R_{XY}(t,s) - m_X(t)m_Y(s)$ ($C_{XY}(m,n)$ for discrete valued): Autocovariance function.
  - $X(t)$ and $Y(t)$ uncorrelated processes: $C_{XY}(t,s) = 0$ (also true for independent RVs)
- Wide sense stationary or weak sense stationary or weakly stationary or WSS: 1st and 2nd order moments do not change with time
  - $m_X(t) = m_X$: Mean is constant
  - $R_{XX}(t,t+\tau) = R_{XX}(\tau)$, $C_{XX}(t,t+\tau) = C_{XX}(\tau)$: 2nd order moments can be written in terms of time differences. Note: Many references take the other convention and define $R_{XX}(t+\tau,t) = R_{XX}(\tau)$, etc.
  - $S_{XX}(f)$ or $S_X(f)$ or $S_{XX}(\omega)$ or $S_X(\omega)$: $= \mathcal{F}[R_{XX}(\tau)]$ Power spectral density
  - $S_{XX}(s)$ or $S_X(s)$ Power spectral density represented with respect via the Laplace transform: $= \mathcal{L}[R_{XX}(\tau)]$
- $X(t)$ and $Y(t)$ jointly Wide sense stationary or jointly weak sense stationary or jointly weakly stationary or JWSS: Joint 1st and 2nd order moments do not change with time
  - $X(t)$ and $Y(t)$ are individually WSS
  - $R_{XY}(t,t+\tau) = R_{XY}(\tau)$, $C_{XY}(t,t+\tau) = C_{XY}(\tau)$: cross-moments can be written in terms of time differences. Note: Many references take the other convention and define $R_{XY}(t+\tau,t) = R_{XY}(\tau)$, etc.
  - $S_{XY}(f)$ or $S_{XY}(\omega)$: $= \mathcal{F}[R_{XY}(\tau)]$ Cross
  - $S_{XY}(s)$ Cross-power spectral density represented with respect via the Laplace transform: $= \mathcal{L}[R_{XY}(\tau)]$ power spectral density

Boston University
Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Notes on Probability

1. Probability Space
   - A triple $(\Omega, \mathcal{F}, P)$ that describes outcomes of a random experiment
     (a) Ω: Set of elementary outcomes, sample space.
     (b) $\mathcal{F}$: Set of events – Conditions we care about. Subsets of Ω closed under countable unions and intersections, and complementation
     (c) $P(\cdot)$: Probability measure – satisfies axioms:
        i. $P(\Omega) = 1$
        ii. $P(A) \geq 0,\ \forall\ A \in \mathcal{F}$.
        iii. $P(\cup A_i) = \sum P(A_i)$ if $A_i \cap A_j = \emptyset$, $i \neq j$.
   - Note: In continuous sample spaces can have events that can occur, but have zero probability and events that have probability 1, but may not occur (e.g. Experiment: choose real number in [0, 1], with uniform distribution, Event 1: Choose 1/2, Event 2: Choose any number but 1/2).

2. Random Variables
   - A function mapping outcomes to real numbers:

3. One Random Variable
   - Probability Distribution Function (PDF): $P_X(x) \equiv P(X \leq x)$
     – $P[x_1 < X \leq x_2] = P_X(x_2) - P_X(x_1)$
   - Probability Density Function (pdf): $p_X(x) = \dfrac{dP_X(x)}{dx}$
     – $P(A) = \displaystyle\int_A P_X(x)\,dx$
   - $E\left[g(X)\right] \equiv \displaystyle\int_{-\infty}^{\infty} g(x) p_X(x)\,dx$
   - Mean: $m_X = E[X]$
   - n-th moment: $E[X^n] = \int_{-\infty}^{\infty} x^n p_X(x)\,dx$
   - Variance: $\sigma_X^2 = E\left[(X - m_X)^2\right] = E[X^2] - (E[X])^2$

4. Two Random Variables X, Y
   - Joint Distribution Function: $P_{X,Y}(x, y) \equiv P[(X \leq x) \cap (Y \leq y)]$
   - Joint Density Function: $p_{X,Y}(x, y) = \dfrac{\partial^2 P_{X,Y}(x, y)}{\partial x \partial y}$
   - Marginal pdf: $p_X(x) = \displaystyle\int_{-\infty}^{\infty} p_{X,Y}(x, y)\,dy$

- Conditional Density: $p_{X|Y}(x \mid y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)} = \dfrac{p_{Y|X}(y|x)p_X(x)}{p_Y(y)}$
- Bayes' Rule: $p_{X|Y}(x \mid y) = \dfrac{p_{Y|X}(y \mid x)p_X(x)}{p_Y(y)}$
- X, Y <u>statistically independent</u> $\Leftrightarrow p_{X,Y}(x,y) = p_X(x)p_Y(y)$
- Expected value: $E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)p_{X,Y}(x,y)\,dx\,dy$
- Correlation: $E[XY]$
- Covariance: $\sigma_{XY} = E[(X-m_X)(Y-m_Y)] = E[XY] - m_X m_Y = \mathrm{Cov}(X,Y)$
- X, Y uncorrelated $\Leftrightarrow \sigma_{XY} = 0 \Leftrightarrow E[XY] = E[X]E[Y]$
- X, Y orthogonal $\Leftrightarrow E[XY] = 0$
- $(X,Y$ Independent$)$ $\Rightarrow$ $(X,Y$ Uncorrelated$)$, but $(X,Y$ Independent$)$ $\not\Leftarrow$ $(X,Y$ Uncorrelated$)$.
- Conditional Expectation (mean) of X given Y: $E[X \mid Y] = \int_{-\infty}^{\infty} x p_{X|Y}(x|y)\,dx$

5. Random Vectors

- $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$, $\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$
- Joint distribution function:
  $P_{\underline{X}}(\underline{x}) = P[(X_1 \le x_1), \ldots, (X_N \le x_N)]$ or $P_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) = P[(X_1 \le x_1), \ldots, (X_N \le x_N), (Y_1 \le y_1), \ldots, (Y_N \le y_N)]$
- Joint density: $p_{\underline{X}}(\underline{x}) = \dfrac{\partial^N P_{\underline{X}}(\underline{x})}{\partial x_1 \ldots \partial x_N}$ or $p_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) = \dfrac{\partial^{2N} P_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{\partial x_1 \ldots \partial x_N, \partial y_1 \ldots \partial y_N}$
- $\underline{X}$, $\underline{Y}$ independent if $p_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) = p_{\underline{X}}(\underline{x})p_{\underline{Y}}(\underline{y})$
- Conditional Density: $p_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \dfrac{p_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{p_{\underline{Y}}(\underline{y})} = \dfrac{p_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x})p_{\underline{X}}(\underline{x})}{p_{\underline{Y}}(\underline{y})}$
- $E[\underline{g}(\underline{x})] = \begin{pmatrix} E[g_1(\underline{x})] \\ \vdots \\ E[g_N(\underline{x})] \end{pmatrix} = \int_{-\infty}^{\infty} \underline{g}(\underline{x}) f(\underline{x})\,d\underline{x}$
- Mean Vector: $E[\underline{X}] = m_{\underline{X}} = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_N] \end{pmatrix}$
- Covariance Matrix:
  $\mathrm{Cov}(\underline{X},\underline{X}) = \Lambda_{\underline{X}} = \Sigma_{\underline{X}} = E[(\underline{X} - m_X)(\underline{X} - m_X)^T] = E[\underline{X}\,\underline{X}^T] - m_{\underline{X}} m_{\underline{X}}^T$
- Cross-Covariance Matrix:
  $\mathrm{Cov}(\underline{X},\underline{Y}) = \Lambda_{\underline{XY}} = \Sigma_{\underline{XY}} = E[(\underline{X} - m_{\underline{X}})(\underline{Y} - m_{\underline{Y}})^T] = E[\underline{X}\underline{Y}^T] - m_{\underline{X}} m_{\underline{Y}}^T$
- Conditional Mean: $E[\underline{X}|\underline{Y}] = \int_{-\infty}^{\infty} \underline{x}\, f(\underline{x}|\underline{y}) d\underline{x}$
- Conditional Covariance: $\Sigma_{X|Y} = \int_{-\infty}^{\infty} (\underline{x} - E[\underline{x}|\underline{y}])(\underline{x} - E[\underline{x}|\underline{y}])^T f(\underline{x}|\underline{y})\,d\underline{x}$

- Uncorrelated: $\Sigma_{\underline{XY}} = 0, \Rightarrow E[\underline{XY}^T] = E[\underline{X}]E[\underline{Y}]^T$.
- Orthogonal: $E[\underline{XY}^T] = 0$.

Boston University
Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Overall Class Summary

---

1. **Characterization and Manipulation of Random Processes**
   - Tools of stochastic processes
   - Joint pdfs
   - Marginal pdfs
   - Conditional pdfs
   - General Expectation $E[g(x)]$
   - Conditional Expectation $E[g(x)|y]$
   - Moments: Means, autocorrelations, autocovariances, power spectral density
   - Distribution-based properties: Stationarity, wide-sense stationarity, ergodicity, independence, IIP, Markov, etc
   - Moment-based properties: Uncorrelated, orthogonal, etc
   - Functions of random variables (Derived pdfs): Equivalent events have equal probability
   - Random vectors
   - Special random processes and their means and variances: e.g.
     - Gaussian
     - Poisson
     - Exponential
     - etc
2. **Random Signals and Systems**
   - I/O relationships for random processes through linear systems
   - I/O relationships for LTI systems and Stationary processes
   - Time, frequency, and Laplace/z domain expressions
   - Special case: Discrete time linear models and finding system parameters (AR, MA, ARMA)
   - LCCDE descriptions
   - Sampling
   - Noise modeling and special processes: Wiener process (Gaussian noise), Poisson process, White noise,...
   - Spectral Factorization for shaping filter design

3. **Detection Theory/Hypothesis Testing**

- $x = H_i$ Discrete-valued
- Given:
  - Observation Model $p_{Y|X}(y|x = H_i)$
  - Prior Model $p_X(x = H_i)$
  - Costs C_{ij} = Cost of deciding H_i and H_j true
- Find Optimal detection rule to minimize $E(\text{Cost})$.
- Special Case Rules: MAP Rule (MPE=$C_{ij} = 1 - \delta_{ij}$), ML Rule (MPE and $P_i = P_j$), Neyman-Person (max P_D for $P_F \leq \alpha$), etc
- Binary vs multi-valued/M-ary detection (For Binary: ROC, likelihood-ratio test, P_D, P_F, etc)
- Finding performance: $\Pr(\text{Err}) = P_e$ or $E(\text{Cost})$
- Detection of signals – role of matched filter for a signal in white noise
- KL expansions for signal detection, White vs Colored noise
- Special results: Min distance classifier, Gaussian processes, MPE rule, etc.

4. **Estimation Theory**

- General Bayes (Random Parameter) Estimation
  - Setup:
    (a) Parameter Model: $P_X(x)$, Probabilistic Prior Density
    (b) Observation Process: $P_{Y|X}(y \mid x)$, Conditional density
    (c) Costs: $J(\widehat{x}, x)$ = Cost of Estimating $\widehat{x}$ when x True.
  - Estimation Rule: Minimize Expected Cost $\Longrightarrow \widehat{x}(y) = \arg\min_x E[J(\widehat{x}, x)] = \arg\min_x E[J(\widehat{x}, x) \mid y]$
  - Performance Measures: Define error $e \equiv x - \widehat{x}(y)$
    * $E[\text{Cost}] = E[J(\widehat{x}, x)]$
    * Bias: $b \equiv E[e]$. Just a number for Bayes Estimation.
    * Error Covariance: $\Lambda_e \equiv E\left[(e-b)(e-b)\right] = E\left[ee^T\right] - bb^T$ Uncertainty in estimate
    * Mean Square Error: $\text{MSE} = E[e^Te] = \text{Tr}\left[[E[ee^T]\right] = \text{Tr}\left[\Lambda_e + bb^T\right]$
- Bayes Least Squares Estimation (BLSE)
  - Cost: $J(\widehat{x}, x) = ||\widehat{x} - x||^2 = ||e||^2 \Longrightarrow$ BLSE is Minimum Mean Square Error Estimate (MSEE)
  - Estimate: $\widehat{x}_B(y) = E[x \mid y]$. $\Longrightarrow$ BLSE Estimate is Conditional Mean
  - Bias: $b = E[x - \widehat{x}_B(y)] = E[x] - E\left[E[x \mid y]\right] = 0$. $\Longrightarrow$ BLSE estimates are unbiased
  - Error Covariance: $\Lambda_B = E\left[(e-0)(e-0)^T\right] = E\left[\Lambda_{x|y}(y)\right]$. Expected value of conditional covariance
  - $E[\text{Cost}] = \text{MSE} = E[e^Te] = \text{Tr}\left\{\Lambda_B\right\} = \text{Tr}\left\{E[\Lambda_{x|y}]\right\}$. Minimum value of MSE over all estimators (linear and nonlinear).
  - Alternate characterization of BLSE
    * $E\left[x - \widehat{x}_B(y)\right] = 0$. Unbiased
    * $E\left\{[x - \widehat{x}_B(y)]\, g(y)\right\} = 0, \forall g(\cdot)$. Error orthogonal to any function of the data

– Gaussian Vector Case:

$$\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \underline{m}_x \\ \underline{m}_y \end{bmatrix}, \begin{bmatrix} \Lambda_x & \Lambda_{xy} \\ \Lambda_{xy}^T & \Lambda_y \end{bmatrix}\right) \implies \begin{aligned} \widehat{\underline{x}}_B(y) &= \underline{m}_x + \Lambda_{xy}\Lambda_y^{-1}\left(\underline{y} - \underline{m}_y\right) \\ \Lambda_B &= \Lambda_{x|y} = \Lambda_x - \Lambda_{xy}\Lambda_y^{-1}\Lambda_{xy}^T \\ \text{Cost} &= \text{MSE} = \text{Tr}\left(\Lambda_B\right) \end{aligned}$$

Estimate is linear in Gaussian case and $\Lambda_{x|y}$ is not a function of $\underline{y}$

- Bayes Maximum A Posteriori Estimation (MAP)
  - Cost: $J(\widehat{x}, x) = \begin{cases} 1 & |\widehat{x} - x| > \Delta \\ 0 & |\widehat{x} - x| > \Delta \end{cases} \quad \Delta \to 0.$ Uniform Cost
  - Estimate: $\widehat{x}_{MAP}(y) = \arg\max_x p_{X|Y}(x \mid y) = \arg\max_x p_{Y|X}(y \mid x) p_X(x). \implies$ MAP Estimate is Conditional Mode
  - MAP Equation for Estimate: $\left.\dfrac{\partial \ln\left[p_{Y|X}(y \mid x)\right]}{\partial x} + \dfrac{\partial \ln\left[p_X(x)\right]}{\partial x}\right|_{x=\widehat{x}_{MAP}(y)} = 0$
  - Bias: $b = E[x - \widehat{x}_{MAP}(y)] \neq 0$ in general. $\implies$ MAP estimates can be biased
  - MAP Estimation requires knowledge of details of density
- Bayes Linear Least Squares Estimation (BLLSE)
  - BLSE with estimator constrained to have a linear form: $\widehat{x}_L(y) = C\underline{y} + \underline{d}$
  - Estimate: $\widehat{x}_L(y) = \underline{m}_x + \Lambda_{xy}\Lambda_y^{-1}\left(\underline{y} - \underline{m}_y\right)$
  - BLLSE Estimators only require second order properties
  - Bias: $b = E[x - \widehat{x}_L(y)] = 0. \implies$ LLSE estimates are unbiased
  - Error Covariance: $\Lambda_L = E\left[(e-0)(e-0)^T\right] = \Lambda_x - \Lambda_{xy}\Lambda_y^{-1}\Lambda_{xy}^T$.
  - $E[\text{Cost}] = \text{MSE} = E[e^T e] = \text{Tr}\left\{\Lambda_L\right\}$. Minimum value of MSE over all linear estimators.
  - Alternate characterization of BLLSE. Unique linear function of y such that:
    * $E\left[x - \widehat{x}_L(y)\right] = 0$. Unbiased
    * $E\left\{\left[x - \widehat{x}_L(y)\right]\underline{y}^T\right\} = 0, \forall g(\cdot)$. Error orthogonal to (linear functions of) the data
- General Nonrandom Parameter Estimation
  - Setup:
    (a) Parameter Model: x is an unknown deterministic parameter
    (b) Observation Process: $P_{Y|X}(y \mid x)$, Parameterized density (aka "likelihood function").
  - Estimation Rule: No general procedure as in Bayes case.
  - Performance Measures: Define error $e(x) \equiv x - \widehat{x}(y)$
    * All are a function of x and not just numbers.
    * Bias: $b(x) = E[e \mid X = x] \equiv E[x - \widehat{x}(y) \mid X = x] = \int[x - \widehat{x}(y)]p_{Y|X}(y \mid x)\, dy$.
    * Error Covariance: $\Lambda_e(x) \equiv E\left[(e - b(x))(e - b(x))^T \mid X = x\right]$
    * Mean Square Error: $\text{MSE}(x) = E[e^T e \mid X = x] = \text{Tr}\left[\left[E[ee^T \mid X = x\right]\right] = \text{Tr}\left[\Lambda_e(x) + b(x)b(x)^T\right]$.
    * Cramer-Rao Estimation Error Covariance Bound

· If $\widehat{x}(y)$ is any unbiased (nonrandom parameter) estimate of x and $\Lambda_e(x)$ its associated estimation error covariance:

$$\Lambda_e(x) \quad \geq \quad \frac{1}{I_Y(x)}, \qquad I_Y(x) = E\left\{ \left[\frac{\partial}{\partial x} \ln p_{Y|X}(y \mid x) \right]^2 \middle| X = x \right\} = -E\left\{ \frac{\partial^2}{\partial x^2} \ln p_{Y|X}(y \mid$$

· Any unbiased estimator that achieves the CRB is termed efficient.

- Maximum Likelihood Estimation (Nonrandom parameter)
  - Estimate: $\widehat{x}_{ML}(y) = \arg\max_x P_{Y|X}(y \mid x)$.
  - ML Equation for Estimate: $\left.\frac{\partial \ln \left[p_{Y|X}(y \mid x) \right]}{\partial x} \right|_{x=\widehat{x}_{ML}(y)} = 0 \qquad \Longrightarrow$ Limit of MAP as $\partial p_X(x)/\partial x \to 0$.
  - Performance:
    * If an efficient estimator does exist it is $\widehat{x}_{ML}(y)$ and in this case $\widehat{x}_{ML}(y)$ is the minimum variance, unbiased estimator.
    * If an efficient estimator does not exist, there may be unbiased estimators with lower variances.
  - ML Facts:
    * If $z = g(x) \Longrightarrow \widehat{z}_{ML}(y) = g\left(\widehat{x}_{ML}(y)\right)$
    * As number of observations $N \to \infty$ ML estimate is asymptotically unbiased, efficient, and consistant.

5. **LLSE Estimation of Random Processes based on Random Processes**

- $x(t)$, $y(\tau)$ assumed zero-mean. If not estimate $\tilde{x}(t) = (x(t) - m_x(t))$ based on $\tilde{y}(t) = (y(t) - m_y(t))$
- Linear estimator $\Longrightarrow$ Only need second order properties $K_{xx}(t,\tau)$, $K_{yx}(t,\tau)$, $K_{yy}(t,\tau)$.
- Form of estimator:

$$\text{CT}: \ \widehat{x}(t) = \int_{T_i}^{T_f} h(t,\sigma) y(\sigma)\, d\sigma$$

- Orthogonality conditions for optimal solution $\Longrightarrow$ Wiener-Hopf Equations
- General Wiener-Hopf Equations for optimal estimator:

$$\text{CT:} \quad K_{xy}(t,\tau) \quad = \quad \int_{T_i}^{T_f} h(t,\sigma)\, K_{yy}(\sigma,\tau)\, d\sigma \qquad \forall \tau \in [T_i,\, T_f]$$

- General Error Covariance:

$$\text{CT:} \quad \Lambda_{LSE}(t) \quad = \quad K_{xx}(t,t) - \int_{T_i}^{T_f} h(t,\sigma)\, K_{yx}(\sigma,t)\, d\sigma$$

- Discrete-time, finite-length: Same solution as for random vectors (i.e. normal equations) Estimate: $\Lambda_{xy} = h^T \Lambda_{yy}$. Error Variance: $\Lambda_{LSE} = \Lambda_x - h^T \Lambda_{xy^T}$
- Noncausal Wiener Filter:

- LLSE, $x(t)$, $y(t)$ zero mean
- $x(t)$, $y(t)$ Jointly wide-sense stationary
- Observation interval: $T_i = -\infty$, $T_f = +\infty$
- Optimal Estimate: WH equation just a convolution – use transform techniques

$$\text{CT:} \quad H_{nc}(s) \;=\; \frac{S_{yx}(s)}{S_{yy}(s)}$$

- Error-Covariance:

$$\text{CT:} \quad \Lambda_{nc} \;=\; K_{xx}(0) - \int_{-\infty}^{\infty} h(u)\, K_{yx}(u)\, du$$

or

$$\begin{aligned}
\text{CT:} \quad S_{ee}(s) &= S_{xx}(s) - \frac{S_{yx}(s)S_{yx}(-s)}{S_{yy}(s)} \\
S_{ee}(j\omega) &= S_{xx}(j\omega) - \frac{|S_{yx}(j\omega)|^2}{S_{yy}(j\omega)} \\
\Lambda_{nc} &= R_{ee}(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{ee}(j\omega)\, d\omega
\end{aligned}$$

- <u>Causal Wiener Filter:</u>
  - LLSE, $x(t)$, $y(t)$ zero mean
  - $x(t)$, $y(t)$ Jointly wide-sense stationary
  - Observation interval: $T_i = -\infty$, $T_f = t$
  - Optimal Estimate: Whiten data first via $W(s)$ then use CWF for white noise $G(s)$.

$$\text{CT:} \quad H_c(s) \;=\; W(s)\,G(s) = \underbrace{\frac{1}{S^{+}_{yy}(s)}}_{\text{Whitening Filter}} \; \underbrace{\left\{\frac{S_{yx}(s)}{S^{-}_{yy}(s)}\right\}_{+}}_{\text{CWF for Innovations}}$$

  - Error-Covariance:

$$\text{CT:} \quad \Lambda_c \;=\; K_{xx}(0) - \int_{0}^{\infty} h(\tau)\, K_{yx}(\tau)\, d\tau$$

or

$$\text{CT:} \quad \Lambda_c \;=\; K_{xx}(0) - \int_{0}^{\infty} K^2_{\nu x}(\tau)\, d\tau = K_{xx}(0) - \int_{0}^{\infty} g^2(\tau)\, d\tau$$

- Similar expressions for Discrete-time.
- $\Lambda_{nc} \leq \Lambda_c$
- Know important special cases: e.g. $y(t) = x(t) + v(t)$, $x(t) \perp v(t)$.

6. **Recursive Filtering and the Discrete-Time Kalman Filter**

- Basic concept of sequential estimation of a random variable from sequential observations.

- Key idea of recursive estimation is whitening – role of orthogonality relationships
- Kalman-Filtering: LLSE for problem satisfying particular assumptions:
  - Covariance structure for $x(t)$ specified implicitly via state space model:
$$\underline{x}(t+1) = A(t)\underline{x}(t) + B(t)\underline{u}(t) + G(t)\underline{w}(t)$$
  - Observation Model:
$$\underline{y}(t) = C(t)\underline{x}(t) + \underline{v}(t)$$
- Notation:
$$\begin{aligned} \underline{\widehat{x}}(t \mid s) &= \text{LLSE of } \underline{x}(t) \text{ given } y(\tau),\ \tau \leq s \\ \underline{e}(t \mid s) &= \underline{x}(t) - \underline{\widehat{x}}(t \mid s) \\ P(t \mid s) &= E\left[\underline{e}(t \mid s)\underline{e}(t \mid s)^T\right] \end{aligned}$$
- To solve:
  (a) Set up state space model, identify $A(t)$, $B(t)$, $G(t)$, $C(t)$, $u(t)$, $w(t)$, $v(t)$, and covariances $Q(t)$, $R(t)$.
  (b) Find initial conditions: $\underline{\widehat{x}}(t_0|t_0-1)$, $P(t_0|t_0-1)$
  (c) Iterate Kalman filtering equations
- Kalman Filtering Equations:

  Initialization:
$$\begin{aligned} \underline{\widehat{x}}(t_0|t_0-1) &= \underline{m}_x(t_0) \\ P(t_0|t_0-1) &= P_x(t_0) \end{aligned}$$
  Update Step:
$$\underline{\widehat{x}}(t|t) = \underline{\widehat{x}}(t|t-1) + P(t|t-1)C^T(t)\left[C(t)P(t|t-1)C^T(t) + R(t)\right]^{-1}\left[\underline{y}(t) - C(t)\underline{\widehat{x}}(t|t-1)\right]$$
$$P(t|t) = P(t|t-1) - P(t|t-1)C^T(t)\left[C(t)P(t|t-1)C^T(t) + R(t)\right]^{-1}C(t)P(t|t-1)$$
  Prediction Step:
$$\underline{\widehat{x}}(t+1|t) = A(t)\underline{\widehat{x}}(t|t) + B(t)\underline{u}(t)$$
$$P(t+1|t) = A(t)P(t|t)A^T(t) + G(t)Q(t)G^T(t)$$
- Kalman gain: $K = P(t|t-1)C^T(t)\left[C(t)P(t|t-1)C^T(t) + R(t)\right]^{-1}$
- For stationary processes and long observation intervals, Kalman filter $\Longrightarrow$ Causal Wiener Filter as $t \longrightarrow \infty$. Steady state analysis.

7. **Advice:**
   - Have basic results at your fingertips
   - Know the assumptions/conditions behind formulas that you use!
   - Perform sanity checks on answers – go back to basics if totally stuck (i.e. defining equation of expectation, variance etc)
   - Know Fourier/Laplace transforms, partial fraction expansions, and properties
   - Make sure you're clear on difference between e.g. $S_{yy}^{+}(s)$ and $\{S_{yy}\}_{+}$
   - Don't forget to deal with the mean! (e.g. $R_{xx}(t)$ vs $K_{xx}(t)$, etc.)

Boston University
Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Exam 2 Summary

1. **Linear Systems and Random Processes:** $Y(t) = \int h(t,\tau)X(\tau)\,d\tau$
   - Complete characterization difficult $\Longrightarrow$ Use second order relationships
   - General Relations between 2nd order statistics:
     - Mean: $m_Y(t) = \int_{-\infty}^{\infty} h(t,\tau)m_X(\tau)\,d\tau$
     - Cross-correlation: $R_{YX}(t,s) = \int_{-\infty}^{\infty} h(t,\tau)R_{XX}(\tau,s)\,d\tau$
     - Output-correlation: $R_{YY}(t,s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t,\tau)R_{XX}(\tau,\sigma)h(s,\sigma)\,d\tau d\sigma$
2. **LTI Systems and Wide-Sense Stationary Processes:**
   - If $h(t)$ LTI and $X(t)$ WSS $\Longrightarrow$ $X(t)$,$Y(t)$ are JWSS
   - LTI Time-domain Relations between 2nd order statistics:
     - Mean: $m_Y = m_X \int_{-\infty}^{\infty} h(\tau)\,d\tau = H(0)m_X$
     - Cross-correlation: $R_{YX}(t) = \int_{-\infty}^{\infty} h(-\tau)R_{XX}(t-\tau)\,d\tau = h(-t) * R_{XX}(t)$
     - Output-correlation: $R_{YY}(t) = h(t) * R_{XX}(t) * h(-t)$
   - LTI Frequency-domain Relations:
     - Cross-PSD: $S_{YX}(j\omega) = H(-j\omega)S_{XX}(j\omega)$ or $S_{YX}(s) = H(-s)S_{XX}(s)$
     - Output-PSD: $S_{YY}(j\omega) = |H(j\omega)|^2 S_{UU}(j\omega)$ or $S_{YY}(s) = H(s)H(-s)S_{UU}(s)$
   - Shaping Filter: LTI system $H(s)$ driven by white noise.
   - Properties of PSD of $S_{YY}(s)$: Quadrantal Symmetry of poles and zeros.
   - Spectral Factorization: Can always write $S_{YY}(s) = G(s)G(-s)$ with $G(s)$ stable and causal. Yields shaping filter for given $S_{YY}(s)$.
3. **DT Linear Models:**
   - Autoregressive (AR). All pole model, IIR
     - $x(n) = \sum_{i=1}^{P} a_i x(n-i) + w(n)$, $R_{WW}(n) = \sigma^2\delta(n)$
     - $R_{XX}(m) = \sum_{i=1}^{P} a_i R_{XX}(m-i) + \sigma^2\delta(m)$
     - Yule-Walker Equations. Linear equations for coefficients a_i:

$$\begin{bmatrix} R_{XX}(1) \\ R_{XX}(2) \\ \vdots \\ R_{XX}(P) \end{bmatrix} = \begin{bmatrix} R_{XX}(0) & R_{XX}(1) & \cdots & R_{XX}(P-1) \\ R_{XX}(1) & R_{XX}(0) & & R_{XX}(P-2) \\ \vdots & & \ddots & \vdots \\ R_{XX}(P-1) & R_{XX}(P-2) & \cdots & R_{XX}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

- Moving Average (MA). All zero model, FIR
  - $x(n) = \sum_{k=1}^{Q} b_k w(n-k) + w(n)$, $R_{WW}(n) = \sigma^2 \delta(n)$
  - Nonlinear equations for coefficients b_i:

$$R_{XX}(m) = \begin{cases} \sigma^2 \left(\sum_{k=m}^{Q} b_k b_{k-m} \right) = \sigma^2 b(m) * b(-m) & m \leq Q \\ 0 & m > Q \end{cases}$$

  - Correlation function is finite for MA
- Autoregressive Moving Average (ARMA). Both poles and zeros
  - $x(n) = \sum_{i=1}^{P} a_i x(n-i) + \sum_{k=1}^{Q} b_k w(n-k) + w(n)$, $R_{WW}(n) = \sigma^2 \delta(n)$
  - In general, equations for coefficients a_i, b_k are nonlinear:

$$R_{XX}(m) = \sum_{i=1}^{P} a_i R_{XX}(m-i) + \sum_{j=0}^{Q} b_j R_{XW}(m-j), \qquad b_0 = 1$$

  - For $m > Q$: $R_{XX}(m) = \sum_{i=1}^{P} R_{XX}(m-i)$.
  - Solution approach:
    (a) Solve linear equations for $m > Q$ for AR coefficients
    (b) Solve NL equations for MA coefficients: $S_{XX}(z)A(z)A(z^{-1}) = B(z)B(z^{-1})\sigma^2$

4. **Sampling of Random Processes:**

- Stochastic Nyquist Criterion Exists
- Result: If $S_{XX}(\omega) = 0$ for $|\omega| > W$ (i.e. is bandlimited) and $T_s < \pi/W$ then:

$$\lim_{N \to \infty} \left[2T_s \frac{W}{2\pi} \sum_{n=-N}^{N} X(nT_s) \frac{\sin W(t-nT_s)}{W(t-nT_s)} \right] \stackrel{mss}{=} X(t)$$

- Result: If $S_{XX}(\omega) = 0$ for $|\omega| > W$ and $T_s < \pi/W$ then:

$$\lim_{N \to \infty} 2T_s \frac{W}{2\pi} \sum_{n=-N}^{N} R_{XX}(nT_s) \frac{\sin W(\tau - nT_s)}{W(\tau - nT_s)} = R_{XX}(\tau)$$

5. **Detection/Hypothesis Testing:**

- Deterministic Decision Rule: Mapping of Observation space onto H_0, H_1.
- Bayes Risk Approach:
  - Priors $P_i = \Pr(H_i)$
  - Observation Model: $P(y|H_i)$
  - Costs: C_{ij} = Cost of deciding H_i when H_j true.
  - Choose decision rule to min E(cost)
  - Likelihood Ratio Test (LRT) minimizes E(cost):

$$\mathcal{L}(y) = \frac{P_{Y|H_1}(y|H_1)}{P_{Y|H_0}(y|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{(C_{10} - C_{00})P_0}{(C_{01} - C_{11})P_1} = \eta$$

- Special Cases:
  - MPE: $C_{ij} = 1 - \delta_{ij} \Longrightarrow$ <u>MAP decision rule</u>

$$P_{H_1|Y}(H_1|y) \underset{H_0}{\overset{H_1}{\gtrless}} P_{H_0|Y}(H_0|y)$$

  - MPE and $P_0 = P_1 = 1/2 \Longrightarrow$ <u>ML decision rule</u>

$$P_{Y|H_1}(y|H_1) \underset{H_0}{\overset{H_1}{\gtrless}} P_{Y|H_0}(y|H_0)$$

  - Gaussian Problems
- Randomized Tests: Given Two LRTs (LRT_1 and LRT_2) use LRT_1 with probability p and LRT_2 with probability $1-p$. Has performance on line connecting two P_D,P_F pairs.

- Discrete Random Variables – know how to handle
- <u>Sufficient Statistic</u>: Function of the data that contains all information needed for test
- <u>Performance</u>:
  - $E[\text{Cost}] = \underbrace{C_{00}P_0 + C_{01}P_1}_{\text{Fixed Cost}} + \underbrace{(C_{10} - C_{00})P_0 P_F + (C_{11} - C_{01})P_1 P_D}_{\text{Fn of threshold } \eta}$
  - $\Pr(\text{error}) = \Pr[\text{choose } H_0,\, H_1 \text{ true}] + \Pr[\text{choose } H_1,\, H_0 \text{ true}] = (1 - P_D)P_1 + P_F P_0$
  - Both only depend on:

$$\begin{aligned} P_D &= \Pr(\text{Choose } H_1|H_1) = \int_{\{y|\text{say } H_1\}} P(y|H_1)\,dy = \int_{\mathcal{L}>\eta} P(\mathcal{L}|H_1)\,d\mathcal{L} \\ P_F &= \Pr(\text{Choose } H_1|H_0) = \int_{\{y|\text{say } H_1\}} P(y|H_0)\,dy = \int_{\mathcal{L}>\eta} P(\mathcal{L}|H_0)\,d\mathcal{L} \end{aligned}$$

  - <u>Receiver Operating Characteristic:</u> (ROC): Plot of P_D vs P_F as threshold is varied.
    - Know properties – Concave, etc
    - Discrete random variables: ROC consists of points
    - Role of randomized tests
- <u>Minimax Tests</u>: Minimize the maximum $E[\text{Cost}]$ as P_1 is varied. Minimax test satisfies $P_D = \left(\frac{C_{01} - C_{00}}{C_{01} - C_{11}}\right) - \left(\frac{C_{10} - C_{00}}{C_{01} - C_{11}}\right) P_F$
- <u>Neyman-Pearson Tests</u>: Maximize P_D subject to $P_F \leq \alpha$. Solution is a LRT for some threshold
- <u>M-ary Bayes Hypothesis Tests:</u>
  - Solution is:

$$\text{Choose } H_k \text{ if } \quad \sum_{j=0}^{M-1} C_{kj} P(H_j|y) \leq \sum_{j=0}^{M-1} C_{ij} P(H_j|y) \qquad \forall i$$

  - Generates set of $M(M-1)/2$ unique comparisons defining decision regions:

$$\sum_{j=0}^{M-1} C_{kj} P_j P_{Y|H_j}(y|H_j) \underset{\text{Not } H_i}{\overset{\text{Not } H_k}{\gtrless}} \sum_{j=0}^{M-1} C_{ij} P_j P_{Y|H_j}(y|H_j) \qquad \forall\, i,k \text{ pairs}$$

- Define $L_j(y) = \frac{P_{Y|H_j}(y|H_j)}{P_{Y|H_0}(y|H_0)}$, then test is:

$$\sum_{j=0}^{M-1} C_{kj} P_j L_j(y) \underset{\text{Not } H_i}{\overset{\text{Not } H_k}{\gtrless}} \sum_{j=0}^{M-1} C_{ij} P_j L_j(y) \qquad \forall\, i,k \text{ pairs}$$

  Linear Decision Boundaries in L_i space

- Special Cases:
  - MPE Cost Assignment $C_{ij} = 1 - \delta_{ij} \implies$ MAP decision rule:

$$\text{Choose } H_k \text{ if } \quad P(H_k|y) \geq P(H_i|y) \quad \forall i$$

  - MPE and $P_i = 1/M \implies$ ML decision rule:

$$\text{Choose } H_k \text{ if } \quad P(y|H_k) \geq P(y|H_i) \quad \forall i$$

  - Minimum Distance Classifier
    - $P_{\underline{Y}|H_K}(\underline{y}|H_k) = N(\underline{y}; \underline{m}_k, I)$
    - ML Rule, $P_k = 1/M$
    - $\implies$ Minimum Distance Classifier

$$\text{Choose } H_k \text{ if } \quad ||\underline{y} - \underline{m}_k||^2 \leq ||\underline{y} - \underline{m}_i||^2 \quad \forall i$$

6. **Series Expansions, KLE, and Detection of Continuous Time Processes:**

- Series expansions of stochastic processes: $X(t) = \sum_{i=1}^{\infty} X_i \phi_i(t)$
- KLE:
  - Find good basis functions for stochastic processes
  - Want uncorrelated coefficients: $E[X_i X_j] = \lambda \delta_{ij}$
  - Basis given by solutions to KL equation:

$$\int_{T_0}^{T_1} R_{XX}(t,\tau)\phi_m(\tau)\, d\tau = \lambda_m \phi_m(t)$$

  - Eigendecomposition of $R_{XX}(t,\tau)$
  - Gives optimal approximation of $X(t)$.
  - White Noise: Every complete orthonormal basis (CON) is a KL basis
- Detection of CT waveforms
  - 1 Known signal in white noise:

$$\begin{aligned} H_0: &\quad y(t) = w(t), \qquad R_{WW}(\tau) = \sigma^2\delta(\tau) \\ H_1: &\quad y(t) = s(t) + w(t) \end{aligned}$$

    - Choose $\phi_1(t) = s(t)/\sqrt{E}$ and remaining ϕ_i to form CON set
    - y_1 is a sufficient statistic for problem
    - Matched Filter: $y_1 = \int y(s)s(t)\, dt \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$
    - Performance depends on signal energy, not structure.

– 2 Known signals in white noise:

$$\begin{array}{ll} H_0: & y(t) = s_0(t) + w(t), \quad R_{WW}(\tau) = \delta(\tau) \\ H_1: & y(t) = s_1(t) + w(t) \end{array}$$

* Approach 1) Let $y'(t) = y(t) - s_0(t)$ and apply previous results
* Approach 2) Let subset of basis functions span signal subspace

– M Known signals in white noise:

$$\begin{array}{ll} H_0: & y(t) = s_0(t) + w(t), \quad R_{WW}(\tau) = \delta(\tau) \\ H_1: & y(t) = s_1(t) + w(t) \\ H_2: & y(t) = s_2(t) + w(t) \\ & \vdots \\ H_M: & y(t) = s_M(t) + w(t) \end{array}$$

* Project onto signal subspace. Choose $\phi_1(t),\cdots,\phi_{M+1}(t)$ to span the space of the signals.

$$\begin{array}{llllllll} H_0: & y_1 = s_{01} + w_1 & H_1: & y_1 = s_{11} + w_1 & & H_M: & y_1 = s_{M1} + w_1 \\ & y_2 = s_{02} + w_2 & & y_2 = s_{12} + w_2 & & & y_2 = s_{M2} + w_2 \\ & \vdots & & \vdots & & & \vdots \\ & y_M = s_{0M} + w_M & & y_M = s_{1M} + w_M & \cdots & & y_M = s_{MM} + w_M \\ & y_{M+1} = w_{M+1} & & y_{M+1} = w_{M+1} & & & y_{M+1} = w_{M+1} \\ & \vdots & & \vdots & & & \vdots \end{array}$$

– Known signals in correlated noise:

$$\begin{array}{ll} H_0: & y(t) = s_0(t) + w(t), \quad R_{WW}(\tau) \neq \delta(\tau) \\ H_1: & y(t) = s_1(t) + w(t) \end{array}$$

* Choose $\phi_i(t)$ via KLE of noise $w(t)$. w_i uncorrelated, but need all coefficients in general.

$$\begin{array}{llll} H_0: & y_1 = w_1 & H_1: & y_1 = s_1 + w_1 \\ & y_2 = w_2 & & y_2 = s_2 + w_2 \\ & y_3 = w_3 & & y_3 = s_3 + w_3 \\ & \vdots & & \vdots \end{array}$$

* In practice, truncate after some number of terms

7. **Advice:**

- Have basic results at your fingertips
- Know the assumptions/conditions behind formulas that you use!
- Perform sanity checks on answers – go back to basics if totally stuck (i.e. defining equation of expectation, variance etc)

Boston University
Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Exam 1 Summary

---

1. **Probability and Random Variables:**

- Axiomatic definition of probability: Triple $(\Omega, \mathcal{F}, P)$, Ω = Set of outcomes, $\mathcal{F}$ = Set of events, and $P(\cdot)$ = Probability measure which satisfies axioms:
  (a) $P(\Omega) = 1$
  (b) $P(A) \geq 0, \forall A \in \mathcal{F}$.
  (c) $P(\cup A_i) = \sum P(A_i)$ if $A_i \cap A_j = \emptyset$, $i \neq j$.
- One Random Variable
  - Probability Distribution Function (PDF): $P_X(x) \equiv P(X \leq x)$
    - $P[x_1 < X \leq x_2] = P_X(x_2) - P_X(x_1)$
  - Probability Density Function (pdf): $p_X(x) = \dfrac{dP_X(x)}{dx}$
    - $P(A) = \int_A p_X(x)\,dx$
  - $E[g(X)] \equiv \int_{-\infty}^{\infty} g(x) p_X(x)\,dx$
  - Mean: $m_X = E[X]$
  - n-th moment: $E[X^n] = \int_{-\infty}^{\infty} x^n p_X(x)\,dx$
  - Variance: $\sigma_X^2 = E\left[(X - m_X)^2\right] = E[X^2] - (E[x])^2$
- Two Random Variables X, Y
  - Joint Distribution Function: $P_{X,Y}(x,y) \equiv P[(X \leq x) \cap (Y \leq y)]$
  - Joint Density Function: $p_{X,Y}(x,y) = \dfrac{\partial^2 P_{X,Y}(x,y)}{\partial x \partial y}$
  - Marginal pdf: $p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,dy$
  - Conditional Density: $p_{X|Y}(x \mid y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)} = \dfrac{p_{Y|X}(y|x) p_X(x)}{p_Y(y)}$
  - Bayes' Rule: $p_{X|Y}(x \mid y) = \dfrac{p_{Y|X}(y \mid x) p_X(x)}{p_Y(y)}$
  - X, Y <u>statistically independent</u> $\Leftrightarrow p_{X,Y}(x,y) = p_X(x) p_Y(y)$
  - Expected value: $E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) p_{X,Y}(x,y)\,dx\,dy$
  - Correlation: $E[XY]$
  - Covariance: $\sigma_{XY} = E[(X - m_X)(Y - m_Y)] = E[XY] - m_X m_Y = \mathrm{Cov}(X,Y)$
  - X, Y uncorrelated $\Leftrightarrow \sigma_{XY} = 0 \Leftrightarrow E[XY] = E[X]E[Y]$
  - X, Y orthogonal $\Leftrightarrow E[XY] = 0$
  - $(X, Y$ Independent$)$ $\underset{\not\Longleftarrow}{\Longrightarrow}$ $(X, Y$ Uncorrelated$)$,
  - Conditional Expectation (mean) of X given Y: $E[X \mid Y] = \int_{-\infty}^{\infty} x p_{X|Y}(x|y)\,dx$
- Random Vectors

- $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}, \underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$
- Joint distribution function:
  $P_{\underline{X}}(\underline{x}) = P[(X_1 \leq x_1), \ldots, (X_N \leq x_N)]$ or $P_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) = P[(X_1 \leq x_1), \ldots, (X_N \leq x_N), (Y_1 \leq y_1), \ldots, (Y_N \leq y_N)]$
- Joint density: $p_{\underline{X}}(\underline{x}) = \dfrac{\partial^N P_{\underline{X}}(\underline{x})}{\partial x_1 \ldots \partial x_N}$ or $p_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) = \dfrac{\partial^{2N} P_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{\partial x_1 \ldots \partial x_N, \partial y_1 \ldots \partial y_N}$
- $\underline{X}$, $\underline{Y}$ independent if $p_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) = p_{\underline{X}}(\underline{x}) p_{\underline{Y}}(\underline{y})$
- Conditional Density: $p_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \dfrac{p_{\underline{X},\underline{Y}}(\underline{x},\underline{y})}{p_{\underline{Y}}(\underline{y})} = \dfrac{p_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) p_{\underline{X}}(\underline{x})}{p_{\underline{Y}}(\underline{y})}$
- $E[\underline{g}(\underline{x})] = \begin{pmatrix} E[g_1(\underline{x})] \\ \vdots \\ E[g_N(\underline{x})] \end{pmatrix} = \int_{-\infty}^{\infty} \underline{g}(\underline{x}) f(\underline{x})\, d\underline{x}$
- Mean Vector: $E[\underline{X}] = m_{\underline{X}} = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_N] \end{pmatrix}$
- Covariance Matrix:
  $\mathrm{Cov}(\underline{X},\underline{X}) = \Lambda_{\underline{XX}} = \Sigma_{\underline{XX}} = E[(\underline{X} - m_X)(\underline{X} - m_X)^T] = E[\underline{XX}^T] - m_{\underline{X}} m_{\underline{X}}^T$
- Convariance matrix constraints: $\Sigma_{\underline{XX}}$ must be a symmetric, PSD matrix
- Cross-Covariance Matrix:
  $\mathrm{Cov}(\underline{X},\underline{Y}) = \Lambda_{\underline{XY}} = \Sigma_{\underline{XY}} = E[(\underline{X} - m_{\underline{X}})(\underline{Y} - m_{\underline{Y}})^T] = E[\underline{XY}^T] - m_{\underline{X}} m_{\underline{Y}}^T$
- Conditional Mean: $E[\underline{X}|\underline{Y}] = \int_{-\infty}^{\infty} \underline{x}\, f(\underline{x}|\underline{y}) d\underline{x}$
- Conditional Covariance: $\Sigma_{X|Y} = \int_{-\infty}^{\infty} (\underline{x} - E[\underline{x}|\underline{y}])(\underline{x} - E[\underline{x}|\underline{y}])^T f(\underline{x}|\underline{y})\, d\underline{x}$
- Uncorrelated: $\Sigma_{\underline{XY}} = 0, \Rightarrow E[\underline{XY}^T] = E[\underline{X}]E[\underline{Y}]^T$.
- Orthogonal: $E[\underline{XY}^T] = 0$.
- Gaussian Random Vectors: $\underline{a}^T \underline{X}$ is a Gaussian random variable for all $\underline{a}$.

2. **Characterization and Manipulation of Random Processes**

- Complete Characterization of Random Processes: In terms of N-th order probability distribution or density functions $p_{X(t_1),X(t_2),\cdots,X(t_N)}(x_1, x_2, \cdots, x_N)$ for all t_i, N.
- Joint pdfs
- Marginal pdfs
- Conditional pdfs
- Mean: $m_x(t) = E[X(t)]$
- Autocorrelation: $R_{XX}(t_1,t_2) = E[X(t_1)X(t_2)]$
- Autocovariance: $K_{XX}(t_1,t_2) = C_{XX}(t_1,t_2) = E[X(t_1)X(t_2)] - E[X(t_1)]E[X(t_2)]$
- Constraints on $R_{XX}(t_1,t_2)$, $K_{XX}(t_1,t_2)$: $\mathrm{Var}\left[\int a(t)X(t)\, dt\right] \geq 0$
- General Expectation $E[g(x)]$
- Conditional Expectation $E[g(x)|y]$
- Second order characterization: Partial characterization in terms of $m_x(t)$ and $K_{XX}(t_1,t_2)$.
- Special Types of Stochastic Processes

- Gaussian: $X(t)$ is Gaussian process $\Longleftrightarrow \sum_{i=1}^{N} a_i X(t_i)$ a Gaussian random variable for all a_i, t_i, N.
- Markov: $p_{X(t_N)|X(t_{N-1}),X(t_{N-2}),\cdots,X(t_1)}(X_N|X_{N-1},\cdots,X_1) = p_{X(t_N)|X(t_{N-1})}(X_N|X_{N-1})$, for all t_i with $t_i \geq t_{i-1}$
- IIP: $X(t_i) - X(t_{i-1})$ independent of $X(t_{i-1})$ for all $t_i \geq t_{i-1}$. IIP $\longrightarrow$ Markov. IIP $\longrightarrow K_{XX}(t,s) = \mathrm{Var}[\min(t,s)]$
- Strict Sense Stationary: $p_{X(t_1),X(t_2),\cdots,X(t_N)}(x_1,x_2,\cdots,x_N) = p_{X(t_1+\tau),X(t_2+\tau),\cdots,X(t_N+\tau)}(x_1,x_2,\cdots,x_N)$ for all τ, N.
- Wide Sense or weakly Stationary: $m_x(t) = m_x$, $K_{XX}(t_1,t_2) = K_{XX}(t_2 - t_1)$

- Special random processes and their means and variances: e.g.
  - Poisson Counting Process: $\Pr[N(t) = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$, $m_N(t) = \lambda t$, $K_{nn}(t,s) = \lambda \min(t,s)$. $N(t)$ is IIP.
  - Random Telegraph Wave
  - Random Walk
  - Wiener Process: $m_x(t) = 0$, $K_{XX}(t,s) = \alpha \min(t,s)$, IIP

3. **Convergence, Mean Square Calculus**

- Mean Square Convergence: $\lim_{n\to\infty} E\left[(X_n - X)^2\right] = 0$
- Cauchy Criterion for MSS Convergence: $\lim_{n\to\infty} E\left[(x_n - x_m)^2\right] \to 0$
- Mean Square Continuity: $\lim_{\varepsilon\to 0} X(t+\varepsilon) \overset{\text{mss}}{=} X(t)$. Mean Square Continuous if $R_{XX}(t_1,t_2)$ continuous.
- Mean Square Derivative: $\lim_{\varepsilon\to 0}\left(\dfrac{X(t+\varepsilon) - X(t)}{\varepsilon}\right) \overset{\text{mss}}{=} \dot{X}(t)$. Mean Square differentiable if $\frac{\partial^2}{\partial t_1 \partial t_2} R_{XX}(t_1,t_2)$ exists.
- Mean Square Integral: $\lim_{N\to\infty} \left.\sum_{i=1}^{N} X(s+i\Delta)\Delta\right) \overset{\text{mss}}{=} Y$. Mean square integrable if $\int_s^t \int_s^t R_{XX}(\sigma,\tau)\,d\sigma d\tau$ exists.

4. **Ergodicity**

- Idea: Time average = Ensemble average in MSS sense.
- Ergodic in the Mean: $\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} X(t)\,dt = \lim_{T\to\infty} \langle m_x\rangle_T \overset{\text{mss}}{=} m_x$
- Ergodic in Autocorrelation: $\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} X(t+\tau)X(t)\,dt = \lim_{T\to\infty} \langle R_{XX}(\tau)\rangle_T \overset{\text{mss}}{=} R_{XX}(\tau)$
- Completely Ergodic: $\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} g\{X(t)\}\,dt \overset{\text{mss}}{=} E[g\{X(t)\}]$

5. **Power Spectral Density**

- For $X(t)$ WSS
- Beware WSS Notation: $R_{XY}(t,t+\tau) \equiv R_{XY}(t+\tau-t) = R_{XY}(\tau)$ verses $R_{XY}(t,t+\tau) \equiv R_{XY}(t-t-\tau) = R_{XY}(-\tau)$.
- Power Spectral Density: $S_{XX}(\omega) = F[R_{XX}(\tau)] = \int_{-\infty}^{\infty} R_{XX}(\tau)e^{-j\omega\tau}d\tau$, $S_{XX}(f) = F[R_{XX}(\tau)] = \int_{-\infty}^{\infty} R_{XX}(\tau)e^{-j2\pi f\tau}d\tau$
- Inverse Power Spectral Density: $R_{XX}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XX}(\omega)e^{j\omega\tau}\,d\omega = \int_{-\infty}^{\infty} S_{XX}(f)e^{j2\pi f\tau}\,df$
- Cross-Spectral Density: $S_{XY}(\omega) = F[R_{XY}(\tau)] = \int_{-\infty}^{\infty} R_{XY}(\tau)e^{-j\omega\tau}d\tau$, $S_{XY}(f) = F[R_{XY}(\tau)] = \int_{-\infty}^{\infty} R_{XY}(\tau)e^{-j2\pi f\tau}d\tau$

- Inverse Cross-Power Spectral Density: $R_{XY}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XY}(\omega)e^{j\omega\tau}\,d\omega = \int_{-\infty}^{\infty} S_{XY}(f)e^{j2\pi f\tau}\,df$
- For sequences: $S_{XX}(f) = \sum_{n=-\infty}^{\infty} R_{XX}(n)e^{-j2\pi fn}$, $-1/2 < f < 1/2$. $S_{XX}(\omega) = \sum_{n=-\infty}^{\infty} R_{XX}(n)e^{-j\omega n}$, $-\pi < \omega < \pi$.
- $S_{XX}(f)$ average power at frequency f.
- Properties of $S_{XX}(f)$ and $S_{XY}(f)$:
  - $S_{XX}(\omega)$ real, non-negative.
  - Total average power in $X(t)$: $R_{XX}(0) = \int_{-\infty}^{\infty} S_{XX}(\omega)\,d\omega$
  - $X(t)$ Real $\longrightarrow S_{XX}(\omega) = S_{XX}(-\omega)$
  - $\alpha X(t) \longrightarrow \alpha^2 S_{XX}(\omega)$
  - $\frac{d}{dt}X(t) \longrightarrow \omega^2 S_{XX}(\omega)$
  - $X(t)e^{j\omega_0 t} \longrightarrow S_{XX}(\omega - \omega_0)$
  - $X(t) + b \longrightarrow S_{XX}(\omega) + 2\pi|b|^2\delta(\omega)$.

6. **Advice:**
   - Have basic results at your fingertips
   - Know the assumptions/conditions behind formulas that you use!
   - Perform sanity checks on answers – go back to basics if totally stuck (i.e. defining equation of expectation, variance etc)
   - Don't forget to deal with the mean! (e.g. $R_{XX}(t)$ vs $K_{XX}(t)$, etc.)

Boston University

Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Stochastic Process Examples

---

1. $w(n) =$ i.i.d. $\sim N(0, \sigma^2)$. So the discrete sequence is a series of independent Gaussians at each time. This process is known as "white Gaussian noise."

   (a) First Order (Marginal) Density: Since for each n the process is *defined* to be a Gaussian the first order density is trivially obtained as: $p_{X(t)}(x) = N(x; 0, \sigma^2)$.

   (b) Second Order (Marginal) Density: Note that $X(t_1)$ and $X(t_2)$ are independent random variables for all $t_1 \neq t_2$. Since they are individually Gaussian and independent they are thus jointly Gaussian. In particular, the second order density is given by the pdf of the vector $[X(t_1), X(t_2)]^T$, which is a Gaussian vector completely characterized by its mean and covariance:

$$\begin{bmatrix} X(t_1) \\ X(t_2) \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \right)$$

2. $X(t) = Zt$, where $Z \sim N(0, 1)$. The sample paths are just lines through the origin.

   (a) First Order (Marginal) Density: Since for each (fixed) t the process is just that (fixed) t times Z, a Gaussian random variable, the first order density is just a Gaussian and thus completely characterized by its mean and covariance. The mean and covariance are given by:

$$E[X(t)] = E[Zt] = tE[Z] = 0$$
$$E[X^2(t)] = E[t^2 Z^2] = t^2 E[Z^2] = t^2$$

   Thus the first order density for any time t is given by:

$$X(t) \sim N(0, t^2)$$

   (b) Second Order (Marginal) Density: Note that $X(t_1)$ and $X(t_2)$ are just different numbers (i.e. t_1 and t_2) times the *same* Gaussian random variable. Since arbitrary linear combinations of $X(t_1) = t_1 Z$ and $X(t_2) = t_2 Z$ will just be another constant times Z (i.e. $aX(t_1) + bX(t_2) = at_1 Z + bt_2 Z = (at_1 + bt_2)Z$), $X(t_1)$ and $X(t_2)$ are clearly jointly Gaussian random variables. Their joint density function is thus completely characterized by the mean vector and covariance matrix, which is given as:

$$\begin{aligned} E\begin{bmatrix} X(t_1) \\ X(t_2) \end{bmatrix} &= \begin{bmatrix} E(t_1 Z) \\ E(t_2 Z) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ E\left(\begin{bmatrix} X(t_1) \\ X(t_2) \end{bmatrix} \begin{bmatrix} X(t_1) & X(t_2) \end{bmatrix} \right) &= E\left(\begin{bmatrix} t_1^2 Z^2 & t_1 t_2 Z^2 \\ t_1 t_2 Z^2 & t_2^2 Z^2 \end{bmatrix} \right) = \begin{bmatrix} t_1^2 & t_1 t_2 \\ t_1 t_2 & t_2^2 \end{bmatrix} \end{aligned}$$

   Thus:

$$\begin{bmatrix} X(t_1) \\ X(t_2) \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} t_1^2 & t_1 t_2 \\ t_1 t_2 & t_2^2 \end{bmatrix} \right)$$

(c) First Order Conditional Density: From the above arguments we know that $X(t_1)$ and $X(t_2)$ jointly Gaussian, so the marginal density $p_{X(t_1)|X(t_2)}(x_1|x_2)$ must also be Gaussian. Being Gaussian, it is completely characterized by its mean and covariance. But we have already seen in class the expressions for the conditional means and variances of a jointly Gaussian random variables. In particular applying these earlier results for Gaussian random variables we have for the mean:

$$E[X(t_1)|X(t_2) = x_2] = \mu_{X(t_1)} + \frac{\sigma_{X(t_1)X(t_2)}}{\sigma_{X(t_2)X(t_2)}} \left(x_2 - \mu_{X(t_2)}\right) = \frac{t_1}{t_2} X(t_2)$$

But this is just the line through the origin through the point $X(t_2)$, which isn't surprising if we remember what the sample paths look like. Now for the variance:

$$\text{Var}(X(t_1)|X(t_2)) = \sigma_{X(t_1)X(t_1)} - \frac{\sigma^2_{X(t_1)X(t_2)}}{\sigma_{X(t_1)X(t_1)}} = t_1^2 - \frac{t_1^2 t_2^2}{t_2^2} = 0$$

Thus there is no variablity! Again this is not surprising if we keep in mind the sample paths. Together we have:

$$p_{X(t_1)|X(t_2)}(x_1|x_2) = N\left(\frac{t_1}{t_2} x_2, 0\right)$$

which is really deterministic.

3. $X(t) = A\sin(2\pi f_0 t + \Theta)$, where A is a random variable uniformly distributed on the interval $[-1,\, 1]$, Θ is a random variable independent of A, uniformly distributed on the interval $[0,\, 2\pi]$, and f_0 is a constant. Note that the sample paths of such a random process are just sinusoids with random phase and amplitude, as depicted in the figure below:

(a) Mean function: The key is that Θ and A are independent:

$$m_X(t) \;=\; E\left[A\sin(2\pi f_0 t + \Theta)\right] = E\left[A\right] E\left[\sin(2\pi f_0 t + \Theta)\right] = m_A \int_{-\pi}^{\pi} \frac{1}{2\pi} \sin(2\pi f_0 t + \Theta)\, d\Theta = 0$$

(b) Correlation function:

$$\begin{aligned}
R_X(t_1, t_2) &= E\left[X(t_1)X(t_2)\right] = E\left[A^2 \sin(2\pi f_0 t_1 + \Theta)\sin(2\pi f_0 t_2 + \Theta)\right] \\
&= E\left[A^2\right] E\left[\sin(2\pi f_0 t_1 + \Theta)\sin(2\pi f_0 t_2 + \Theta)\right] \\
&= \frac{1}{2} E\left[A^2\right] \left(E\left[\cos[2\pi f_0(t_1 - t_2)]\right] - E\left[\cos[2\pi f_0(t_1 + t_2) + 2\Theta]\right]\right) \\
&= \frac{1}{2} E\left[A^2\right] \cos[2\pi f_0(t_1 - t_2)] \\
&= \frac{1}{24} \cos[2\pi f_0(t_1 - t_2)]
\end{aligned}$$

where we have used the fact that $E\left[\cos[2\pi f_0(t_1 + t_2) + 2\Theta]\right] = 0$. Note that this correlation function only depends on time difference $t_2 - t_1$.

(c) Correlation coefficient: Simple calculation shows that:

$$r_{XX}(t_1, t_2) \quad = \quad \cos[2\pi f_0(t_1 - t_2)]$$

Note that while the sample functions look deterministic, if you know $X(t_1)$ there is still uncertainty in the value of $X(t_2)$ *except* when $(t_1 - t_2)2\pi f_0 = n\pi$.

4. $X(t) = A\cos(2\pi f_0 t + \Theta)$, where A and f_0 are constants and Θ is a random variable, uniformly distributed on the interval $[0,\, 2\pi]$. This might be a model for a signal in a communication system, where the frequency and amplitude are known but the phase is random. Note that the sample paths of such a random process are just sinusoids, as depicted in the figure below:

(a) First Order (Marginal) Density: In this example we can't take the Gaussian short cuts we did above. So we are forced to use the "methods of events" wherein we find the probability distribution function by assigning equal events equal probability then differentiating the resulting distribution function to get the density function. Thus first we want to find:

$$P_{X(t)}(x) = \Pr(X(t) \leq x)$$

Now we know that:

$$\Pr(X(t) \leq x) = \Pr(\{\theta | X(t) \leq x\})$$

where the set of θ defined in the second probability is the set of all θ for which the first inequality is satisfied. These are the equivalent events.
Now, lets find $\{\theta | X(t) \leq x_0\}$. From its definition we know that $X(t) = A\cos(2\pi f_0 t + \Theta)$. Consider the "phaser diagram" shown in the figure below for our situation for a fixed time t_0 and where $|x_0| \leq A$.

If we want $X(t_0) = A\cos(2\pi f_0 t_0 + \theta) \leq x_0$, then from the figure we can see that we want:

$$\cos^{-1}\left(\frac{x}{A}\right) \leq 2\pi f_0 t_0 + \theta \leq \left(2\pi - \cos^{-1}\left(\frac{x}{A}\right)\right)$$

Or, equivalently, we want:

$$\left(\cos^{-1}\left(\frac{x}{A}\right) - 2\pi f_0 t\right) \le \theta \le \left(2\pi - \cos^{-1}\left(\frac{x}{A}\right) - 2\pi f_0 t\right)$$

This is the set $\{\theta | X(t) \le x\}$. Thus we have

$$\begin{aligned}
\Pr(X(t) \le x) &= \Pr\left[\left(\cos^{-1}\left(\frac{x}{A}\right) - 2\pi f_0 t\right) \le \theta \le \left(2\pi - \cos^{-1}\left(\frac{x}{A}\right) - 2\pi f_0 t\right)\right] \\
&= \int_{\cos^{-1}\left(\frac{x}{A}\right)-2\pi f_0 t}^{2\pi-\cos^{-1}\left(\frac{x}{A}\right)-2\pi f_0 t} p_\Theta(\theta)\, d\theta = \frac{1}{2\pi}\left[2\pi - 2\cos^{-1}\left(\frac{x}{A}\right)\right] \\
&= 1 - \frac{1}{\pi}\cos^{-1}\left(\frac{x}{A}\right) = P_{X(t)}(x)
\end{aligned}$$

Now we can obtain the first order marginal density by taking the derivate with respect to x of the distribution function we have just found. This yields for $|x| \le A$:

$$\begin{aligned}
p_{X(t)}(x) &= \frac{\partial}{\partial x} P_{X(t)}(x) = \frac{\partial}{\partial x}\left[1 - \frac{1}{\pi}\cos^{-1}\left(\frac{x}{A}\right)\right] \\
&= \frac{-1}{\pi}\left(\frac{\partial}{\partial x}\cos^{-1}\left(\frac{x}{A}\right)\right) = \frac{-1}{\pi}\left(\frac{-1}{\sqrt{1-\left(\frac{x}{A}\right)^2}}\left(\frac{1}{A}\right)\right) \\
&= \begin{cases} \frac{1}{\pi\sqrt{A^2-x^2}} & |x| \le A \\ 0 & \text{else} \end{cases}
\end{aligned}$$

This density is as shown in the figure below:

Boston University

Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Notes on Second-Order Statistics and Linear Systems

This summary is a compendium of results on second-order statistics arising from random processes passing through linear systems. I have attempted to present the real, multivariable case and use $H(t)$ for the impulse response matrix and $\mathrm{H}(s)$ for transformed quantities. Further, for compactness I have only presented the radian form of the transforms. This handout follows the convention for stationary systems that e.g. $R_{XY}(\tau) = R_{XY}(t, t+\tau)$. This is the convention used in the notes, but is not universally used, so beware!

Continuous Time	**Discrete Time**
$\underline{u}(t) \rightarrow$ wide sense stationary	$\underline{u}(n) \rightarrow$ wide sense stationary
$H(t,\tau)$ stable, multivariable, real, $[H(t,\tau)]_{ij} = h_{ij}(t,\tau)$	$H(n,k)$ stable, multivariable, real, $[H(n,k)]_{ij} = h_{ij}(n,k)$
$\underline{y}(t) = \int_{-\infty}^{\infty} H(t,\tau)\underline{u}(\tau)\, d\tau$	$\underline{y}(n) = \sum_{k=-\infty}^{\infty} H(n,k)\underline{u}(k)$
$\underline{m}_Y(t) = \int_{-\infty}^{\infty} H(t,\tau)\underline{m}_U(\tau)\, d\tau$	$\underline{m}_Y(n) = \sum_{k=-\infty}^{\infty} H(n,k)\underline{m}_U(k)$
$R_{YZ}(t,s) = \int_{-\infty}^{\infty} H(t,\tau)R_{UZ}(\tau,s)\, d\tau$	$R_{YZ}(m,n) = \sum_{k=-\infty}^{\infty} H(m,k)R_{UZ}(k,n)$
$R_{ZY}(t,s) = R_{YZ}^T(t,s)$	$R_{ZY}(m,n) = R_{YZ}^T(m,n)$
$R_{YU}(t,s) = \int_{-\infty}^{\infty} H(t,\tau)R_{UU}(\tau,s)\, d\tau$	$R_{YU}(m,n) = \sum_{k=-\infty}^{\infty} H(m,k)R_{UU}(k,n)$
$R_{UY}(t,s) = R_{YU}^T(s,t)$	$R_{UY}(m,n) = R_{YU}^T(n,m)$
$R_{YY}(t,s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(t,\sigma)R_{UU}(\sigma,\tau)H^T(s,\tau)\, d\tau d\sigma$	$R_{YY}(m,n) = \sum_{k=-\infty}^{\infty}\sum_{\ell-\infty}^{\infty} H(m,k)R_{UU}(k,\ell)H^T(n,\ell)$
$R_{YY}(t,s) = R_{YY}^T(s,t)$	$R_{YY}(m,n) = R_{YY}^T(n,m)$

LTI, WSS, Continuous Time	**LSI, WSS, Discrete Time**
$H(t,\tau) \rightarrow H(t-\tau)$	$H(n,k) \rightarrow H(n-k)$
$R_{YU}(t,s) \rightarrow R_{YU}(s-t)$	$R_{YU}(m,n) \rightarrow R_{YU}(n-m)$
$\underline{y}(t) = \int_{-\infty}^{\infty} H(t-\tau)\underline{u}(\tau)\, d\tau = H(t) * \underline{u}(t)$	$\underline{y}(n) = \sum_{k=-\infty}^{\infty} H(n-k)\underline{u}(k) = H(n) * \underline{u}(n)$
$\underline{m}_Y = \underline{m}_U \int_{-\infty}^{\infty} H(t-\tau)\, d\tau = \mathrm{H}(0)\underline{m}_U$	$\underline{m}_Y = \underline{m}_U \sum_{k=-\infty}^{\infty} H(n-k) = \mathrm{H}(1)\underline{m}_U$
$R_{YZ}(\tau) = \int_{-\infty}^{\infty} H(-\sigma)R_{UZ}(\tau-\sigma)\, d\sigma$	$R_{YZ}(m) = \sum_{k-\infty}^{\infty} H(-k)R_{UZ}(m-k)$
$= H(-\tau) * R_{UZ}(\tau)$	$= H(-m) * R_{UZ}(m)$
$R_{ZY}(\tau) = R_{YZ}(-\tau)^T$	$R_{ZY}(m) = R_{YZ}(-m)^T$
$R_{YU}(\tau) = \int_{-\infty}^{\infty} H(-\sigma)R_{UU}(\tau-\sigma)\, d\sigma$	$R_{YU}(m) = \sum_{k=-\infty}^{\infty} H(-k)R_{UU}(m-k)$
$= H(-\tau) * R_{UU}(\tau)$	$= H(-m) * R_{UU}(m)$
$R_{UY}(\tau) = R_{YU}(-\tau)^T$	$R_{UY}(m) = R_{YU}(-m)^T$
$R_{YY}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(-\gamma)R_{UU}(\tau-\gamma-\sigma)H(\sigma)^T\, d\sigma d\gamma$	$R_{YY}(m) = \sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty} H(-k)R_{UU}(m-k-\ell)H(\ell)^T$
$= H(-\tau) * R_{UU}(\tau) * H(\tau)^T$	$= H(-m) * R_{UU}(m) * H(m)^T$
$R_{YY}(\tau) = R_{YY}(-\tau)^T$	$R_{YY}(m) = R_{YY}(-m)^T$

Power Spectral Density	**Power Spectral Density**
$\mathrm{H}^*(j\omega) = \mathrm{H}^T(-j\omega)$	$\mathrm{H}^*(e^{j\omega}) = \mathrm{H}^T(e^{-j\omega})$
$X(t)$, $Z(t)$ JWSS	$X(n)$, $Z(n)$ JWSS
$\Downarrow$	$\Downarrow$
$S_{XZ}(\omega) = \int_{-\infty}^{\infty} R_{XZ}(\tau)e^{-j\omega\tau}\, d\tau$	$S_{XZ}(\omega) = \sum_{k=-\infty}^{\infty} R_{XZ}(k)e^{-j\omega k}$
$S_{XZ}(\omega) = S_{ZX}^T(-\omega)$	$S_{XZ}(\omega) = S_{ZX}^T(-\omega)$
$S_{YZ}(\omega) = \mathrm{H}(-j\omega)S_{UZ}(\omega)$	$S_{YZ}(\omega) = \mathrm{H}(e^{-j\omega})S_{UZ}(\omega)$
$S_{ZY}(\omega) = S_{ZU}(\omega)\mathrm{H}^T(j\omega)$	$S_{ZY}(\omega) = S_{ZU}(e^{j\omega})\mathrm{H}^T(\omega)$
$S_{YU}(\omega) = \mathrm{H}(-j\omega)S_{UU}(\omega)$	$S_{YU}(\omega) = \mathrm{H}(e^{-j\omega})S_{UU}(\omega)$
$S_{UY}(\omega) = S_{UU}(\omega)\mathrm{H}^T(j\omega)$	$S_{UY}(\omega) = S_{UU}(e^{j\omega})\mathrm{H}^T(\omega)$
$S_{YY}(\omega) = \mathrm{H}(-j\omega)S_{UU}(\omega)\mathrm{H}^T(j\omega)$	$S_{YY}(\omega) = \mathrm{H}(e^{-j\omega})S_{UU}(\omega)\mathrm{H}^T(e^{j\omega})$
$= S_{YU}(\omega)\mathrm{H}^T(j\omega)$	$= S_{YU}(\omega)\mathrm{H}^T(e^{j\omega})$
$= \mathrm{H}(-j\omega)S_{UY}(\omega)$	$= \mathrm{H}(e^{-j\omega})S_{UY}(\omega)$

Power Spectral Density	**Power Spectral Density**
$X(s) = \int_{-\infty}^{\infty} x(t)e^{-st}\, dt$	$X(z) = \sum_{k=-\infty}^{\infty} x(k)z^{-k}$
$\mathrm{H}^*(s) = \mathrm{H}^T(-s)$	$\mathrm{H}^*(z) = \mathrm{H}^T(z^{-1})$
$X(t)$, $Z(t)$ JWSS	$X(n)$, $Z(n)$ JWSS
$\Downarrow$	$\Downarrow$
$S_{XZ}(s) = \int_{-\infty}^{\infty} R_{XZ}(\tau)e^{-s\tau}\, d\tau$	$S_{XZ}(z) = \sum_{k=-\infty}^{\infty} R_{XZ}(k)z^{-k}$
$S_{XZ}(s) = S_{ZX}^T(-s)$	$S_{XZ}(z) = S_{ZX}^T(z^{-1})$
$S_{YZ}(s) = \mathrm{H}(-s)S_{UZ}(s)$	$S_{YZ}(z) = \mathrm{H}(z^{-1})S_{UZ}(z)$
$S_{ZY}(s) = S_{ZU}(s)\mathrm{H}^T(s)$	$S_{ZY}(z) = S_{ZU}(z)\mathrm{H}^T(z)$
$S_{YU}(s) = \mathrm{H}(-s)S_{UU}(s)$	$S_{YU}(z) = \mathrm{H}(z^{-1})S_{UU}(z)$
$S_{UY}(s) = S_{UU}(s)\mathrm{H}^T(s)$	$S_{UY}(z) = S_{UU}(z)\mathrm{H}^T(z)$
$S_{YY}(s) = \mathrm{H}(-s)S_{UU}(s)\mathrm{H}^T(s)$	$S_{YY}(z) = \mathrm{H}(z^{-1})S_{UU}(z)\mathrm{H}^T(z)$
$= S_{YU}(s)\mathrm{H}^T(s)$	$= S_{YU}(z)\mathrm{H}^T(z)$
$= \mathrm{H}(-s)S_{UY}(s)$	$= \mathrm{H}(z^{-1})S_{UY}(z)$

Boston University
Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Notes on LTI Input/Output Relationships for Single-Input/Single-Output Systems

Deterministic Case

$x(t) \to \boxed{h(t)} \to y(t)$

	Continuous Time	Discrete Time
Signals	BLT: $X(s) = \int_{-\infty}^{\infty} x(t)e^{-st}\,dt$	BZT: $X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}$
	CTFT (Radians): $X(j\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\,dt$	DTFT (Radians): $X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}$
	CTFT (Hertz): $X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt$	DTFT (Hertz): $X(f) = \sum_{n=-\infty}^{\infty} x(n)e^{-j2\pi fn}$
	ICFT: $x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\omega)e^{j\omega t}\,d\omega$	IDFT: $x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega t}\,d\omega$
	$= \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}\,df$	$= \int_{-1}^{1} X(f)e^{j2\pi ft}\,df$
General Systems	$y(t) = \int_{-\infty}^{\infty} h(t,\tau)u(\tau)\,d\tau$	$y(n) = \sum_{n=-\infty}^{\infty} h(n,m)u(m)$
Linear Time Invariant Systems	$h(t-\tau)$	$h(n-m)$
	$y(t) = \int_{-\infty}^{\infty} h(t-\tau)x(\tau)\,d\tau = h(t) * x(t)$	$y(n) = \sum_{m=-\infty}^{\infty} h(n-m)x(m) = h(n) * x(n)$
	$Y(s) = H(s)U(s)$	$Y(z) = H(z)U(z)$
	$Y(f) = H(f)X(f)$	$Y(f) = H(f)X(f)$
	$Y(j\omega) = H(j\omega)X(j\omega)$	$Y(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega})$
Poles	LHP $\to$ Causal, Stable	Inside Unit Disk $\to$ Causal, Stable
	RHP $\to$ Acausal, Stable	Outside of Unit Disk $\to$ Acausal, Stable

Continuous-Time LTI Random Case. $h(t)$ Real

$\{X(t)\} \to \boxed{h(t)} \to \{Y(t)\}$

General Processes: Time Domain	Wide Sense Stationary Processes: Time Domain	Wide Sense Stationary Processes: Frequency Domain
$m_Y(t) = h(t) * m_X(t)$	$m_Y = m_X \int h(t)\,dt = m_X H(0)$	
$R_{YY}(t,s) = h(t) * R_{XX}(t,s) * h(s)$	$R_{YY}(\tau) = R_{XX}(\tau) * h(-\tau) * h(\tau)$	$S_{YY}(f) = S_{XX}(f)\|H(f)\|^2$
$= R_{XY}(t,s) * h(t)$	$= R_{XY}(\tau) * h(-\tau)$	$= S_{XY}(f)H(-f) = S_{XY}(f)H^*(f)$
$= R_{YX}(t,s) * h(s)$	$= R_{YX}(\tau) * h(\tau)$	$= S_{YX}(f)H^*(-f) = S_{YX}(f)H(f)$
$R_{XY}(t,s) = R_{XX}(t,s) * h(s)$	$R_{XY}(\tau) = R_{XX}(\tau) * h(\tau)$	$S_{XY}(f) = S_{XX}(f)H^*(-f) = S_{XX}(f)H(f)$
$R_{YX}(t,s) = R_{XX}(t,s) * h(t)$	$R_{YX}(\tau) = R_{XX}(\tau) * h(-\tau)$	$S_{YX}(f) = S_{XX}(f)H(-f) = S_{XX}(f)H^*(f)$
	$R_{XY}(\tau) = R_{YX}(-\tau)$	$S_{XY}(f) = S_{YX}(f)^*$

Stationarity

$\underset{\text{WSS}}{\{X(t)\}} \to \boxed{h(t)} \to \underset{\text{WSS}}{\{Y(t)\}}$

$\underset{\text{SSS}}{\{X(t)\}} \to \boxed{h(t)} \to \underset{\text{SSS}}{\{Y(t)\}}$

Existence

$$\left.\begin{array}{l} \bullet\ m_X,\ R_{XX}(\tau)\ \text{Exist} \\ \bullet\ \text{System BIBO Stable} \\ \quad (\text{i.e. } \int_{-\infty}^{\infty} |h(t)|dt < \infty \\ \quad \text{or } \sum_{n=-\infty}^{\infty} |h(t)|) \end{array}\right\} \Longrightarrow m_Y,\ R_{YY}(\tau)\ \text{Exist}$$

Boston University

Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Notes on Gaussian Variables, Vectors and Processes

Gaussian Random Variables

- Gaussian probability density function:

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(x-m_x)^2}{2\sigma_x^2}\right)$$

- pdf completely characterized by two parameters: Mean $= m_x$, Variance $= \sigma_x^2$.
- Notation: $X \sim N(m_x, \sigma_x^2)$, $p_X(x) = N(x; m_x, \sigma_x^2) = N(x - m_x; 0, \sigma_x^2) = N\left(\frac{x-m_x}{\sigma_x}; 0, 1\right)$
-
- Characteristic Function: $\Psi_X(\omega) = E\left[e^{j\omega x}\right] = e^{j\omega m_x - \omega^2\sigma_x^2/2}$
- If $X \sim N(m, \sigma^2)$ and $Z = aX + b$, a, b, constant, then $Z \sim N(am + b, \underline{a}^2\sigma^2)$.

Gaussian Random Vectors

- $\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}$ is a Gaussian random vector (equivalently $\{X_1, \cdots, X_N\}$ are jointly Gaussian random variables) if and only if $Y = \underline{a}^T\underline{X}$ is a Gaussian random variable for all $\underline{a} \neq \underline{0}$
- Joint pdf completely characterized by: Mean vector $= m_{\underline{x}}$, Covariance matrix $= \Sigma_x$.
- Notation: $\underline{X} \sim N(m_{\underline{x}}, \Sigma_x)$, $p_{\underline{X}}(\underline{x}) = N(\underline{x}; m_{\underline{x}}, \Sigma_x) = N(\underline{x} - m_{\underline{x}}; 0, \Sigma_x) = N\left(\Sigma_x^{-1/2}(\underline{x} - m_{\underline{x}}); 0, I\right)$
- Gaussian probability density function:

$$p_{\underline{X}}(\underline{x}) = \frac{1}{\sqrt{(2\pi)^N |\Sigma_x|}} \exp\left(-\frac{(\underline{x} - m_{\underline{x}})^T\Sigma_x^{-1}(\underline{x} - m_{\underline{x}})}{2}\right)$$

- All subsets of X_i are jointly Gaussian
- Each X_i is a Gaussian random variable: $X_i \sim N\left(\underline{m}_i, (\Sigma_x)_{ii}\right)$
- Linear combination of Gaussians are also Gaussian: If $\underline{Z} = A\underline{X} + \underline{b}$, $\underline{X} \sim N(m_{\underline{x}}, \Sigma)$ then $\underline{Z} \sim N(Am_{\underline{x}} + \underline{b}, A\Sigma_x A^T)$.
- Conditional density $p_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y})$ of jointly Gaussian random vectors $\underline{X}$, $\underline{Y}$ is also Gaussian with

$$m_{\underline{X}|\underline{Y}} = m_{\underline{X}} + \Sigma_{\underline{XY}}\Sigma_{\underline{YY}}^{-1}(\underline{Y} - m_{\underline{Y}})$$

$$\Sigma_{\underline{X}|\underline{Y}} = \Sigma_{\underline{X}} - \Sigma_{\underline{XY}}\Sigma_{\underline{YY}}^{-1}\Sigma_{\underline{YX}}$$

- For Gaussian random vectors (only): Independent $\iff$ Uncorrelated

Gaussian Random Processes

- $X(t)$ is a Gaussian random process if and only if all its finite dimensional distributions are Gaussian random vectors. i.e. if and only if $\underline{Z} = \begin{bmatrix} X(t_1) \\ \vdots \\ X(t_N) \end{bmatrix} \sim N(m_{\underline{z}}, \Sigma_z)$ for all N and all t_i.
- A Gaussian random process is completely characterized by the quantities: Mean function $= m_x(t)$, Covariance function $K_{xx}(t, x)$.

Boston University

Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Notes on Estimation

The elements of estimation theory are depicted in the above figure. We summarize:

- Decision rule is a mapping of observation y to an estimate $\widehat{x}(y)$.
- Three elements exist:
    1. Model of Nature/Parameter Space
    2. Model of the Observation Process: $P_{Y|X}(y \mid x)$
    3. Estimation Rule

Random Parameters

- General Bayes (Random Parameter) Estimation
    - Setup:
        1. Parameter Model: $P_X(x)$, Probabilistic Prior Density
        2. Observation Process: $P_{Y|X}(y \mid x)$, Conditional density
        3. Costs: $J(\widehat{x}, x) =$ Cost of Estimating $\widehat{x}$ when x True.
    - Estimation Rule: Minimize Expected Cost $\Longrightarrow \widehat{x}(y) = \arg\min_x E[J(\widehat{x}, x)] = \arg\min_x E[J(\widehat{x}, x) \mid y]$
    - Performance Measures: Define error $e \equiv x - \widehat{x}(y)$
        * $E[\text{Cost}] = E[J(\widehat{x}, x)]$
        * Bias: $b \equiv E[e]$. Just a number for Bayes Estimation.
        * Error Covariance: $\Lambda_e \equiv E\left[(e-b)(e-b)^T\right] = E\left[ee^T\right] - bb^T$ Uncertainty in estimate
        * Mean Square Error: $\text{MSE} = E[e^T e] = \text{Tr}\left[[E[ee^T]\right] = \text{Tr}\left[\Lambda_e + bb^T\right]$
- Bayes Least Squares Estimation (BLSE)
    - Cost: $J(\widehat{x}, x) = ||\widehat{x} - x||^2 = ||e||^2 \Longrightarrow$ BLSE is Minimum Mean Square Error Estimate (MSEE)
    - Estimate: $\widehat{x}_B(y) = E[x \mid y]$. $\Longrightarrow$ BLSE Estimate is Conditional Mean
    - Bias: $b = E[x - \widehat{x}_B(y)] = E[x] - E\left[E[x \mid y]\right] = 0$. $\Longrightarrow$ BLSE estimates are unbiased
    - Error Covariance: $\Lambda_B = E\left[(e-0)(e-0)^T\right] = E\left[\Lambda_{x|y}(y)\right]$. Expected value of conditional covariance
    - $E[\text{Cost}] = \text{MSE} = E[e^T e] = \text{Tr}\left\{\Lambda_B\right\} = \text{Tr}\left\{E[\Lambda_{x|y}]\right\}$. Minimum value of MSE over all estimators (linear and nonlinear).
    - Alternate characterization of BLSE
        * $E\left[x - \widehat{x}_B(y)\right] = 0$. Unbiased
        * $E\left\{\left[x - \widehat{x}_B(y)\right] g(y)\right\} = 0$, $\forall g(\cdot)$. Error orthogonal to any function of the data
    - Gaussian Vector Case:

$$\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \underline{m}_x \\ \underline{m}_y \end{bmatrix}, \begin{bmatrix} \Lambda_x & \Lambda_{xy} \\ \Lambda_{xy}^T & \Lambda_y \end{bmatrix}\right) \Longrightarrow \begin{array}{rcl} \widehat{\underline{x}}_B(y) &=& \underline{m}_x + \Lambda_{xy}\Lambda_y^{-1}\left(\underline{y} - \underline{m}_y\right) \\ \Lambda_B &=& \Lambda_{x|y} = \Lambda_x - \Lambda_{xy}\Lambda_y^{-1}\Lambda_{xy}^T \\ \text{Cost} &=& \text{MSE} = \text{Tr}\left(\Lambda_B\right) \end{array}$$

    Estimate is linear in Gaussian case and $\Lambda_{x|y}$ is not a function of $\underline{y}$

- <u>Bayes Maximum A Posteriori Estimation (MAP)</u>
  - Cost: $J(\widehat{x}, x) = \begin{cases} 1 & |\widehat{x} - x| > \Delta \\ 0 & |\widehat{x} - x| > \Delta \end{cases} \qquad \Delta \to 0. \qquad$ Uniform Cost
  - Estimate: $\widehat{x}_{MAP}(y) = \arg\max_x p_{X|Y}(x \mid y) = \arg\max_x p_{Y|X}(y \mid x) p_X(x)$. $\Longrightarrow$ <u>MAP Estimate is Conditional Mode</u>
  - MAP Equation for Estimate: $\left.\dfrac{\partial \ln \left[p_{Y|X}(y \mid x)\right]}{\partial x} + \dfrac{\partial \ln \left[p_X(x)\right]}{\partial x}\right|_{x=\widehat{x}_{MAP}(y)} = 0$
  - Bias: $b = E[x - \widehat{x}_{MAP}(y)] \neq 0$ in general. $\Longrightarrow$ <u>MAP estimates can be biased</u>
  - MAP Estimation requires knowledge of details of density
- <u>Bayes Linear Least Squares Estimation (BLLSE)</u>
  - BLSE with estimator constrained to have a <u>linear</u> form: $\widehat{x}_L(y) = C\underline{y} + \underline{d}$
  - Estimate: $\widehat{x}_L(y) = \underline{m}_x + \Lambda_{xy}\Lambda_y^{-1}\left(\underline{y} - \underline{m}_y\right)$
  - <u>BLLSE Estimators only require second order properties</u>
  - Bias: $b = E[x - \widehat{x}_L(y)] = 0$. $\Longrightarrow$ <u>LLSE estimates are unbiased</u>
  - Error Covariance: $\Lambda_L = E\left[(e-0)(e-0)^T\right] = \Lambda_x - \Lambda_{xy}\Lambda_y^{-1}\Lambda_{xy}^T$.
  - $E[\text{Cost}] = \text{MSE} = E[e^T e] = \text{Tr}\left\{\Lambda_L\right\}$. Minimum value of MSE over all <u>linear</u> estimators.
  - Alternate characterization of BLLSE. Unique linear function of y such that:
    * $E\left[x - \widehat{x}_L(y)\right] = 0$. <u>Unbiased</u>
    * $E\left\{\left[x - \widehat{x}_L(y)\right]\underline{y}^T\right\} = 0$, $\forall g(\cdot)$. <u>Error orthogonal to (linear functions of) the data</u>

Nonrandom Parameters

- <u>General Nonrandom Parameter Estimation</u>
  - Setup:
    1. Parameter Model: x is an <u>unknown deterministic</u> parameter
    2. Observation Process: $P_{Y|X}(y \mid x)$, <u>Parameterized</u> density (aka "likelihood function").
  - Estimation Rule: No general procedure as in Bayes case.
  - Performance Measures: Define error $e(x) \equiv x - \widehat{x}(y)$
    * All are a function of x and not just numbers.
    * Bias: $b(x) = E[e \mid X = x] \equiv E[x - \widehat{x}(y) \mid X = x] = \int [x - \widehat{x}(y)] p_{Y|X}(y \mid x)\, dy$.
    * Error Covariance: $\Lambda_e(x) \equiv E\left[(e - b(x))(e - b(x))^T \mid X = x\right]$
    * Mean Square Error: $\text{MSE}(x) = E[e^T e \mid X = x] = \text{Tr}\left[\left[E[ee^T \mid X = x]\right] = \text{Tr}\left[\Lambda_e(x) + b(x)b(x)^T\right]\right.$.
    * <u>Cramer-Rao Estimation Error Covariance Bound</u>
      · If $\widehat{x}(y)$ is <u>any unbiased</u> (nonrandom parameter) estimate of x and $\Lambda_e(x)$ its associated estimation error covariance:
      $$\Lambda_e(x) \quad \geq \quad \frac{1}{I_Y(x)}, \qquad I_Y(x) = E\left\{\left[\frac{\partial}{\partial x}\ln p_{Y|X}(y \mid x)\right]^2 \middle| X = x\right\} = -E\left\{\frac{\partial^2}{\partial x^2}\ln p_{Y|X}(y \mid x)\middle| X = x\right\}$$
      · Any <u>unbiased</u> estimator that achieves the CRB is termed <u>efficient</u>.
- <u>Maximum Likelihood Estimation (Nonrandom parameter)</u>
  - Estimate: $\widehat{x}_{ML}(y) = \arg\max_x P_{Y|X}(y \mid x)$.
  - ML Equation for Estimate: $\left.\dfrac{\partial \ln \left[p_{Y|X}(y \mid x)\right]}{\partial x}\right|_{x=\widehat{x}_{ML}(y)} = 0 \qquad \Longrightarrow$ Limit of MAP as $\partial p_X(x)/\partial x \to 0$.
  - Performance:
    * If an efficient estimator <u>does</u> exist it is $\widehat{x}_{ML}(y)$ and in this case $\widehat{x}_{ML}(y)$ is the minimum variance, unbiased estimator.
    * If an efficient estimator <u>does not</u> exist, there may be unbiased estimators with lower variances.
  - ML Facts:
    * If $z = g(x) \Longrightarrow \widehat{z}_{ML}(y) = g\left(\widehat{x}_{ML}(y)\right)$
    * As number of observations $N \to \infty$ ML estimate is asymptotically unbiased, efficient, and consistant.

Boston University
Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Notes on Ergodicity

- Basic Idea: Time Averages $\equiv$ Ensemble Averages
- Key required property: Process at widely separated enough points should look independent. Need process to decorrelate "fast enough".
- Ergodicity in the Mean
  - Def: WSS process $X(t)$ is ergodic in the mean if

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} X(t)\,dt = \lim_{T\to\infty} \langle x(t)\rangle_T \overset{\text{mss}}{=} m_x$$

  - Can think of as an "estimator:" $\widehat{m}_x = \langle x(t)\rangle_T \equiv \frac{1}{2T}\int_{-T}^{T} X(t)\,dt$
  - Must have:

Mean: $\lim_{T\to\infty} E\left[\frac{1}{2T}\int_{-T}^{T} X(t)\,dt\right] = E[X(t)] = m_x$ — Always true for WSS

Variance: $\lim_{T\to\infty} \mathrm{Var}\left[\frac{1}{2T}\int_{-T}^{T} X(t)\,dt\right] = \lim_{T\to\infty} E\left[\left(\frac{1}{2T}\int_{-T}^{T} X(t)\,dt - m_x\right)^2\right] = 0$ — This is MSS condition

  - First condition will always be true for WSS processes, thus focus on the second.
  - $\mathrm{Var}\left[\frac{1}{2T}\int_{-T}^{T} X(t)\,dt\right] = \frac{1}{2T}\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)K_{XX}(\tau)\,d\tau$
  - Results:

$X(t)$ is ergodic in mean $\iff \lim_{T\to\infty}\frac{1}{2T}\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)K_{XX}(\tau)\,d\tau = 0$ (Necessary and Sufficient)

$X(t)$ is ergodic in mean $\Longleftarrow \lim_{T\to\infty}\frac{1}{T}\int_{-T}^{T}|K_{XX}(\tau)|\,d\tau = 0$ (Sufficient)

$X(t)$ is ergodic in mean $\Longleftarrow \int_{-\infty}^{\infty}|K_{XX}(\tau)| < \infty$ (Sufficient)

$X(t)$ is ergodic in mean $\Longleftarrow \begin{cases} K_{XX}(0) < \infty \\ \lim_{\tau\to\infty} K_{XX}(\tau) = 0 \end{cases}$ (Sufficient)

- Ergodicity in Autocorrelation
  - Def: WSS process $X(t)$ is ergodic in autocorrelation if for any τ:

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} X(t+\tau)X(t)\,dt = \lim_{T\to\infty}\langle x(t+\tau)x(t)\rangle_T \overset{\text{mss}}{=} R_{XX}(\tau)$$

  - Need to show $\lim_{T\to\infty}\mathrm{Var}\left[\langle x(t+\tau)x(t)\rangle_T\right] = 0$

- Result: WSS $X(t)$ is ergodic in autocorrelation $\Longleftrightarrow$

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-2T}^{2T} \left(1 - \frac{|\tau|}{2T}\right) K_{Y_sY_s}(\tau)\, d\tau = 0 \qquad \forall s$$

  where $Y_s(t) \equiv X(t+s)X(t)$.

- Complete Ergodicity
  - Def: A WSS process $X(t)$ is completely ergodic if for any function $g(\cdot)$:

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g\left\{X(t)\right\}\, dt \stackrel{\text{mss}}{=} E\left[g\left\{X(t)\right\}\right]$$

  - Strongest type of ergodicity. Implies others.
  - Thm: A WSS Gaussian Process $X(t)$ is completely ergodic if

$$\int_{-\infty}^{\infty} |K_{XX}(\tau)| < \infty$$

Boston University

Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

<u>Notes on Detection</u>

The elements of detection theory or hypothesis testing are depicted in the above figure. Three components are needed:

1. A model of the underlying phenomenon
2. A model of the observation process
3. A decision rule – i.e. a mapping of the observation space to decisions.

Binary Hypothesis Testing

- Only two possibilities for X: H_0, H_1
- Decision rule is mapping of observation to one of H_0 or H_1.
- <u>Conditional</u> Probabilities:

$$\begin{aligned} P_F &= \Pr(\text{Choose } H_1|H_0) = \Pr \text{ of False Alarm} && \Longleftarrow \text{"Type I Error"} \\ P_D &= \Pr(\text{Choose } H_1|H_1) = \Pr \text{ of Detection} && \\ P_M &= \Pr(\text{Choose } H_0|H_1) = \Pr \text{ of Miss} = 1 - P_D && \Longleftarrow \text{"Type II Error"} \end{aligned}$$

- <u>Bayes Risk Formulation</u>:
  - Minimize "Bayes' Risk" $= E[\text{Cost}]$ given
    1. Apriori Probabilities: $P_i = \Pr[H_i]$
    2. Observation Model: $P_{Y|H_i}(y|H_i)$
    3. Costs: C_{ij} = Cost of deciding H_i when H_j is true
  - Solution is the <u>Likelihood ratio test</u> (LRT):

$$\mathcal{L} = \frac{P_{Y|H_1}(y|H_1)}{P_{Y|H_0}(y|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{(C_{10} - C_{00})P_0}{(C_{01} - C_{11})P_1} = \eta$$

  - $\mathcal{L}$ is a <u>sufficient statistic</u> $\Longrightarrow$ LRT is a <u>scalar</u> test regardless of the dimension of the data space!
  - Probability of Error:

$$\Pr(\varepsilon) = \Pr[\text{choose } H_0,\, H_1 \text{ true}] + \Pr[\text{choose } H_1,\, H_0 \text{ true}] = P_M P_1 + P_F P_0$$

  - Corresponding Bayes' Risk:

$$E[\text{Cost}] = \underbrace{C_{00}P_0 + C_{01}P_1}_{\text{Fixed Cost}} + \underbrace{(C_{10} - C_{00})P_0P_F + (C_{11} - C_{01})P_1P_D}_{\text{Fn of threshold } \eta}$$

- Special Cases:

- Minimum Probability of Error (MPE) Cost Assignment $C_{ij} = 1-\delta_{ij} \Longrightarrow$ MAP decision rule:

$$P_{H_1|Y}(H_1|y) \underset{H_0}{\overset{H_1}{\gtrless}} P_{H_0|Y}(H_0|y)$$

- MPE and $P_0 = P_1 = 1/2 \Longrightarrow$ ML decision rule:

$$P_{Y|H_1}(y|H_1) \underset{H_0}{\overset{H_1}{\gtrless}} P_{Y|H_0}(y|H_0)$$

- Minimax Tests
  * P_0, P_1 unknown $\Longrightarrow$ Choose η in LRT to minimize the maximum expected cost as function of P_0.
  * $P_D = \left(\dfrac{C_{01}-C_{00}}{C_{01}-C_{11}}\right) - \left(\dfrac{C_{10}-C_{00}}{C_{01}-C_{11}}\right) P_F$
- Neyman-Pearson Tests
  * P_0, P_1 unknown and C_{ij} unknown $\Longrightarrow$ Maximize P_D subject to $P_F \leq \alpha$.

• Performance
  - Can write both Bayes risk $E[\text{Cost}]$ and Pr(error) as a function of only P_D and P_F (assuming C_{ij}, P_i are fixed) $\Longrightarrow$ P_D, P_F contain the information necessary for performance analysis of *any* decision rule
  - Receiver Operating Characteristic: Plot of $P_D(\eta)$ vs $P_F(\eta)$ as threshold η in LRT is varied
  - $P_D = \displaystyle\int_{\{y|\text{Say } H_1\}} P(y \mid H_1)\, dy, \qquad P_F = \int_{\{y|\text{Say } H_1\}} P(y \mid H_0)\, dy$
  - Properties:
    1. $(P_D, P_F) = (0,0)$ and $(P_D, P_F) = (1,1)$ are always on the ROC
    2. ROC is boundary between what is achievable and what is not
    3. η is slope of ROC at point $(P_D(\eta), P_F(\eta))$
    4. ROC for LRT always has $P_D \geq P_F$
    5. ROC is concave downwards
    6. For discrete random variables ROC consists of points
  - d^2 Statistic: Measure of distance between hypotheses or difficulty of test.

    LRT of form: (Test statistic) $\underset{H_0}{\overset{H_1}{\gtrless}}$ (Threshold)

$$d^2 \equiv \frac{(E[\text{test statistic} \mid H_1] - E[\text{test statistic} \mid H_0])^2}{\sqrt{\text{Var}(\text{test statistic} \mid H_1)\,\text{Var}(\text{test statistic} \mid H_0)}}$$

• Randomized Tests
  - Given Two LRTs (LRT_1 and LRT_2) defined by two thresholds η_1, η_2 with corresponding performance (P_{D_1}, P_{F_1}), (P_{D_2}, P_{F_2}).
  - Randomized test given by using LRT_1 with probability p and LRT_2 with probability $1-p$.
  - Performance of randomized test is on line connecting (P_{D_1}, P_{F_1}) and (P_{D_2}, P_{F_2}):

$$P_D(p) = pP_{D_1} + (1-p)P_{D_2}, \qquad P_F(p) = pP_{F_1} + (1-p)P_{F_2}$$

- Important Case – General (Vector) Gaussian Detection: $P_{\underline{Y}|H_i}(\underline{y}|H_i) = N(\underline{y}; \underline{m}_i, \Lambda_i)$, $i = 0, 1$

$$\frac{1}{2}(\underline{y}-\underline{m}_0)^T \Lambda_0^{-1}(\underline{y}-\underline{m}_0) - \frac{1}{2}(\underline{y}-\underline{m}_1)^T \Lambda_1^{-1}(\underline{y}-\underline{m}_1) \underset{H_0}{\overset{H_1}{\gtrless}} \ln\left(\eta\sqrt{\frac{|\Lambda_1|}{|\Lambda_0|}}\right)$$

  - If $\Lambda_0 = \Lambda_1 = \Lambda$

$$\ell(\underline{y}) = (\underline{m}_1 - \underline{m}_0)^T \Lambda^{-1}\underline{y} \underset{H_0}{\overset{H_1}{\gtrless}} \ln(\eta) + \frac{1}{2}\left(\underline{m}_1^T \Lambda^{-1}\underline{m}_1 - \underline{m}_0^T \Lambda^{-1}\underline{m}_0\right) = \Gamma$$

    Note: $\ell(\underline{y})$ is a Gaussian random variable.
  - Scalar case: If $\Lambda_0 = \Lambda_1 = \Lambda = \sigma^2$, $m_1 > m_0$

$$y \underset{H_0}{\overset{H_1}{\gtrless}} \frac{m_1 + m_0}{2} + \frac{\sigma^2 \ln(\eta)}{(m_1 - m_0)}$$

  - If $\Lambda_0 = \Lambda_1 = \sigma^2 I$, and $\eta = 1 \Rightarrow$ Minimum Distance Receiver:

$$\left\|\underline{y} - \underline{m}_0\right\|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \left\|\underline{y} - \underline{m}_1\right\|^2$$

  - If $\underline{m}_0 = \underline{m}_1 = \underline{0}$

$$\underline{y}^T \left(\Lambda_0^{-1} - \Lambda_1^{-1}\right)\underline{y} \underset{H_0}{\overset{H_1}{\gtrless}} 2\ln\left(\eta\sqrt{\frac{|\Lambda_1|}{|\Lambda_0|}}\right) = \Gamma'$$

  - If $\underline{m}_0 = \underline{m}_1 = \underline{0}$, $\Lambda_0 = \sigma_0^2 I$, $\Lambda_1 = \sigma_1^2 I$

$$\ell'(\underline{y}) = \underline{y}^T\underline{y} \underset{H_0}{\overset{H_1}{\gtrless}} 2\left(\frac{\sigma_1^2\sigma_0^2}{\sigma_1^2 - \sigma_0^2}\right)\ln\left(\eta\frac{\sigma_1}{\sigma_0}\right)$$

    Note: $\ell'(\underline{y})$ is not a Gaussian random variable.

M-ary Hypothesis Testing

- M possibilities for X: H_i, $i = 0, \cdots, M-1$
- Decision rule is mapping of observation space to one of H_i.
- Bayes Risk Formulation:
  - Minimize "Bayes' Risk" $= E[\text{Cost}]$ given
    1. Apriori Probabilities: $P_i = \Pr[H_i]$
    2. Observation Model: $P_{Y|H_i}(y|H_i)$
    3. Costs: C_{ij} = Cost of deciding H_i when H_j is true
  - Solution is:

$$\text{Choose } H_k \text{ if} \quad \sum_{j=0}^{M-1} C_{kj} P(H_j|y) \le \sum_{j=0}^{M-1} C_{ij} P(H_j|y) \qquad \forall i$$

  - Generates set of $M(M-1)/2$ unique comparisons defining decision regions:

$$\sum_{j=0}^{M-1} C_{kj} P_j P_{Y|H_j}(y|H_j) \underset{\text{Not } H_i}{\overset{\text{Not } H_k}{\gtrless}} \sum_{j=0}^{M-1} C_{ij} P_j P_{Y|H_j}(y|H_j) \qquad \forall\, i, k \text{ pairs}$$

– Define $L_j(y) = \frac{P_{Y|H_j}(y|H_j)}{P_{Y|H_0}(y|H_0)}$, then test is:

$$\sum_{j=0}^{M-1} C_{kj} P_j L_j(y) \underset{\text{Not } H_i}{\overset{\text{Not } H_k}{\gtrless}} \sum_{j=0}^{M-1} C_{ij} P_j L_j(y) \qquad \forall\, i, k \text{ pairs}$$

<u>Linear Decision Boundaries in L_i space</u>

- Special Cases:
  - <u>MPE</u> Cost Assignment $C_{ij} = \delta_{ij} \implies$ <u>MAP decision rule</u>:

$$\text{Choose } H_k \text{ if} \quad P(H_k|y) \geq P(H_i|y) \quad \forall i$$

  - MPE and $P_i = 1/M \implies$ <u>ML decision rule</u>:

$$\text{Choose } H_k \text{ if} \quad P(y|H_k) \geq P(y|H_i) \quad \forall i$$

- Important Case: $P_{\underline{Y}|H_K}(\underline{y}|H_k) = N(\underline{y}; \underline{m}_k, I)$
  - ML Rule, $P_k = 1/M \implies$ <u>Minimum Distance Receiver</u>

$$\text{Choose } H_k \text{ if} \quad ||\underline{y} - \underline{m}_k||^2 \leq ||\underline{y} - \underline{m}_i||^2 \quad \forall i$$

Series Expansions, KLE, and Detection of Continuous Time Processes

- Series expansions of stochastic processes: $X(t) = \sum_{i=1}^{\infty} X_i \phi_i(t)$
- KLE:
  - Find good basis functions for stochastic processes
  - Want uncorrelated coefficients: $E[X_i X_j] = \lambda \delta_{ij}$
  - Basis given by solutions to KL equation:

$$\int_{T_0}^{T_1} R_{XX}(t, \tau) \phi_m(\tau)\, d\tau = \lambda_m \phi_m(t)$$

  - Eigendecomposition of $R_{XX}(t, \tau)$
  - Gives optimal <u>approximation</u> of $X(t)$.
  - White Noise: Every complete orthonormal basis (CON) is a KL basis
- Detection of CT waveforms
  - 1 Known signal in white noise:

$$\begin{aligned} H_0: &\quad y(t) = w(t), \quad R_{WW}(\tau) = \sigma^2 \delta(\tau) \\ H_1: &\quad y(t) = s(t) + w(t) \end{aligned}$$

    * Choose $\phi_1(t) = s(t)$ and the remaining ϕ_i to form CON set
    * y_1 is a sufficient statistic for problem
    * Matched Filter: $y_1 = \int y(s) s(t)/\sqrt{E}\, dt \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$
    * Performance depends on signal energy, not structure.

- 2 Known signals in white noise:

$$\begin{array}{ll} H_0: & y(t) = s_0(t) + w(t), \qquad R_{WW}(\tau) = \delta(\tau) \\ H_1: & y(t) = s_1(t) + w(t) \end{array}$$

  * Approach 1) Let $y'(t) = y(t) - s_0(t)$ and apply previous results
  * Approach 2) Let subset of basis functions span signal subspace

- M Known signals in white noise:

$$\begin{array}{ll} H_0: & y(t) = s_0(t) + w(t), \qquad R_{WW}(\tau) = \delta(\tau) \\ H_1: & y(t) = s_1(t) + w(t) \\ H_2: & y(t) = s_2(t) + w(t) \\ & \vdots \\ H_M: & y(t) = s_M(t) + w(t) \end{array}$$

  * Project onto signal subspace. Choose $\phi_1(t),\cdots,\phi_{M+1}(t)$ to span the space of the signals.

$$\begin{array}{llllll} H_0: & y_1 = s_{01} + w_1 & H_1: & y_1 = s_{11} + w_1 & & H_M: \; y_1 = s_{M1} + w_1 \\ & y_2 = s_{02} + w_2 & & y_2 = s_{12} + w_2 & & \quad\;\; y_2 = s_{M2} + w_2 \\ & \vdots & & \vdots & \cdots & \quad\;\; \vdots \\ & y_M = s_{0M} + w_M & & y_M = s_{1M} + w_M & & \quad\;\; y_M = s_{MM} + w_M \\ & y_{M+1} = w_{M+1} & & y_{M+1} = w_{M+1} & & \quad\;\; y_{M+1} = w_{M+1} \\ & \vdots & & \vdots & & \quad\;\; \vdots \end{array}$$

- Known signals in correlated noise:

$$\begin{array}{ll} H_0: & y(t) = s_0(t) + w(t), \qquad R_{WW}(\tau) \neq \delta(\tau) \\ H_1: & y(t) = s_1(t) + w(t) \end{array}$$

  * Choose $\phi_i(t)$ via KLE of noise $w(t)$. w_i uncorrelated, but need all coefficients in general.

$$\begin{array}{llll} H_0: & y_1 = w_1 & H_1: & y_1 = s_1 + w_1 \\ & y_2 = w_2 & & y_2 = s_2 + w_2 \\ & y_3 = w_3 & & y_3 = s_3 + w_3 \\ & \vdots & & \vdots \end{array}$$

  * In practice, truncate after some number of terms

Boston University
Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Notes on Finding Derived Distributions

Consider the random experiment with the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω is the top right quadrant of the unit circle centered at zero, $\mathcal{F}$ is a Borel field $\mathcal{B}(\Re^2)$, and $\mathcal{P}$ is given by a uniform distribution. This means that the probability of an event $E \in \mathcal{F}$ is equal to the area of E times $4/\pi$, or $p_{XY}(x, y) = 4/\pi$ for $(x, y) \in \Omega$. An outcome of this experiment will be denoted by $\omega = (x, y) \in \Re^2$ which give the x-axis and y-axis values.

Define a new 2-dimensional random vector (R, Θ), where $R(\omega)$ is the distance from the origin to ω and $\Theta(\omega)$ is the angle from the x-axis. We would like to find the new probability space $(\Omega_{R\Theta}, \mathcal{F}_{R\Theta}, \mathcal{P}_{R\Theta})$. The new sample space is $\Omega_{R\Theta} = [0, 1] \times [0, \pi/2]$ and the new event space is the Borel field corresponding to this sample space $\mathcal{F}_{R\Theta} = \mathcal{B}(\Omega_{R\Theta})$ since the new sample space is continuous. The new probability measure is given by a density function, which we can find using either of the two following methods.

Method I: Method of Equivalent Events (General Solution)
Find the cumulative distribution function using the fact that equivalent events have equal probability:

$$P_{R\Theta}(r, \theta) = \Pr\left(\{(\rho, \phi) : \rho \le r, \phi \le \theta\}\right) = \frac{\pi r^2 \theta}{2\pi} \cdot \frac{4}{\pi} = \frac{2}{\pi} r^2 \theta$$

for $\theta \in [0, \pi/2]$ and $r \in [0, 1]$. Now take the partial derivatives with respect to r and θ to obtain the density function

$$p_{R\Theta}(r, \theta) = \frac{\partial^2}{\partial r \partial \theta} P_{R\Theta}(r, \theta) = \frac{4r}{\pi}.$$

Method II: Jacobian Solution
Note that the functions $R = (X^2 + Y^2)^{1/2}$ and $\Theta = \arctan(\frac{Y}{X})$ are differentiable and there is one root $(x_1, y_1) : x_1 = r\cos\theta, y_1 = r\sin\theta$. The Jacobian is

$$|J(x, y)| = \begin{vmatrix} \frac{\partial}{\partial x} r & \frac{\partial}{\partial y} r \\ \frac{\partial}{\partial x}\theta & \frac{\partial}{\partial y}\theta \end{vmatrix} = \begin{vmatrix} x(x^2 + y^2)^{-1/2} & y(x^2 + y^2)^{-1/2} \\ -y(x^2 + y^2)^{-1} & x(x^2 + y^2)^{-1} \end{vmatrix} = (x^2 + y^2)^{-1/2}$$

and

$$|J(x_1, y_1)| = \frac{1}{r}.$$

Then

$$p_{R\Theta}(r, \theta) = \frac{p_{XY}(x_1, y_1)}{|J(x_1, y_1)|} = \frac{4/\pi}{1/r} = \frac{4r}{\pi}.$$

In this problem, where the original distribution is uniform, it is easier to use the "equivalent events" approach. For more complex distributions, however, it is often easier to use the Jacobian solution, when it is an option.

Certainly the variables X and Y are dependent, since you know that if $X = .95$ then $Y \neq .95$. However, it turns out that R and Θ are independent. To show this, first find the marginal distributions:

$$p_R(r) = \int_0^{\pi/2} f(r, \theta) d\theta = \frac{\pi}{2}\frac{4r}{\pi} = 2r \qquad 0 \le r \le 1$$

$$p_\Theta(\theta) = \int_0^1 f(r, \theta) dr = \frac{2}{\pi} \qquad 0 \le \theta \le \pi/2$$

We thus see that:

$$p_{R\Theta}(r, \theta) = \frac{4r}{\pi} = 2r\frac{2}{\pi} = p_R(r)p_\Theta(\theta)$$

for all r and θ.

Boston University
Department of Electrical and Computer Engineering

EC505 STOCHASTIC PROCESSES

Notes on Convergence, Derivatives, and Integrals

---

1. Important for studying integration, differentiating, series, limits.
2. Deterministic Convergence:
   - Definition for sequences: Sequence x_n converges to limit x if for an $\varepsilon > 0$ there exists an n_0 such that for all $n > n_0(\varepsilon)$, $|x_n - x| < \varepsilon$.
   - Cauchy Criterion: Sequence x_n converges to a limit if and only if $|x_n - x_m| \to 0$ as $n, m \to \infty$. No need for knowledge of limit to test convergence!
3. Mean Square Sense (MSS) Convergence of Sequences of Random Variables
   - Since random will include probabilistic stucture
   - Mean Square Sense Convergence:
   $$\lim_{n\to\infty} E\left[(X_n - X)^2\right] = 0$$
   - Notation:
   $$\begin{aligned} \lim_{n\to\infty} X_n &\overset{\text{mss}}{=} X \\ \underset{n\to\infty}{\text{l.i.m.}}\, X_n &= X \end{aligned}$$
   - Cauchy Criterion for MSS Convergence: $\lim_{n\to\infty} E\left[(X_n - X_m)^2\right] \to 0$. No need for knowledge of limit to test convergence!
4. Mean Square Continuity (MSC) of Random Processes
   - Def: A random process $X(t)$ is mean square sense continuous at t if
   $$\lim_{\varepsilon\to 0} X(t+\varepsilon) \overset{\text{mss}}{=} X(t)$$
   - Result: $X(t)$ is mean square continuous at t if $R_{XX}(t_1, t_2)$ is continuous at $t_1 = t_2 = t$
   - Result: A WSS process $X(t)$ is mean square continuous at $\forall t$ if $R_{XX}(\tau)$ is continuous at $\tau = 0$
5. Mean Square Differentiation of Random Processes
   - Def: A random process $X(t)$ has a mean square derivative at t if
   $$\lim_{\varepsilon\to 0}\left(\frac{X(t+\varepsilon) - X(t)}{\varepsilon}\right) \overset{\text{mss}}{=} \dot{X}(t) \text{ (i.e. some limit)}$$
   - Result: $X(t)$ is mean square differentiable at t if $\frac{\partial^2}{\partial t_1 \partial t_2} R_{XX}(t_1, t_2)$ exists at $t_1 = t_2 = t$.
   - Result: A WSS process $X(t)$ is mean square differentiable for all t if $\frac{d^2}{d\tau^2} R_{XX}(\tau)$ exists at $\tau = 0$.
   - Mean: $E\left[\dot{X}(t)\right] = \dot{m}_x(t)$

- Correlation Functions:

$$\begin{aligned} R_{x\dot{x}}(t_1,t_2) &= \frac{\partial}{\partial t_2} R_{XX}(t_1,t_2) \\ R_{y\dot{x}}(t_1,t_2) &= \frac{\partial}{\partial t_2} R_{yx}(t_1,t_2) \\ R_{\dot{x}\dot{x}}(t_1,t_2) &= \frac{\partial^2}{\partial t_1 \partial t_2} R_{XX}(t_1,t_2) \end{aligned}$$

- $X(t)$ WSS $\Longrightarrow R_{\dot{x}\dot{x}}(\tau) = -\dfrac{d^2}{d\tau^2} R_{XX}(\tau)$
- White Noise: $W(t) = \dot{X}(t)$, where $X(t)$ is a Wiener Process. $m_w(t) = 0$, $R_{ww}(t_1,t_2) = R_{\dot{x}\dot{x}}(t_1,t_2) = \delta(t_1 - t_2)$

6. Mean Square Integration of Random Processes

- Def: A random process $X(t)$ has a mean square integral over $[s, t]$ if:

$$\lim_{N\to\infty} \left. \sum_{i=1}^{N} X(s+i\Delta)\Delta \right) \stackrel{\text{mss}}{=} Y \quad \text{(i.e. some limit)}$$

  where $\Delta = (t-s)/N$
- Result: $X(t)$ is mean square integrable over $[s, t]$ if $\int_s^t \int_s^t R_{XX}(\sigma,\tau)\, d\sigma\, d\tau < \infty$.
- Result: A WSS process $X(t)$ is mean square integrable over $[s, t]$ if $\int_0^{(t-s)} |R_{XX}(\tau)|\, d\tau < \infty$.
- Mean: $E[Y(t)] = E\left[\int_s^t X(\tau)\, d\tau\right] = \int_s^t m_x(\tau)\, d\tau$
- Correlation Functions: $R_{yy}(t_1,t_2) = \int_s^t \int_s^t R_{XX}(\tau_1,\tau_2)\, d\tau_1\, d\tau_2$
- Wiener Process: $X(t) = \int_0^t W(s)\, ds$, where $W(t)$ is white noise, $m_x(t) = 0$, $R_{XX}(\tau) = \sigma^2\delta(\tau)$. $m_x(t) = \int_0^t m_w(s)\, ds = 0$, $R_{XX}(t_1,t_2) = \int_0^{t_2}\int_0^{t_1} R_{ww}(\tau_1,\tau_2)\, d\tau_1\, d\tau_2 = \sigma^2 \min(t_1,t_2)$