EE 278 Lecture Notes 10
Course Summary

- Basic Probability and Random Variables
- First and Second Moments, Correlation
- Gaussians
- Convergence and Limit Theorems
- Classes of Random Processes
- LTIS with WSS Process Input
- Linear Estimation

Basic Probability and Random Variables

- Sample space Ω, outcome ω, events $\subset 2^{\Omega}$, probability measure $\mathrm{P}(\cdot)$
- Law of total probability:
  - events: $\mathrm{P}(B) = \sum_i \mathrm{P}(A_i \cap B)$ if A_i's partition Ω
  - pmf: $p_X(x) = \sum_y p(x,y)$
  - pdf: $f_X(x) = \int f_{X,Y}(x,y)\,dy$
  - mixed: $f_Y(y) = \sum_\theta p_\Theta(\theta) f_{Y|\Theta}(y|\theta)$

    $p_\Theta(\theta) = \int f_Y(y) p_{\Theta|Y}(\theta|y)\,dy$

- Bayes rule
  - events: $\mathrm{P}(A_j \mid B) = \dfrac{\mathrm{P}(B \mid A_j)}{\sum_i \mathrm{P}(B \mid A_i)\mathrm{P}(A_i)} \mathrm{P}(A_j)$ if A_i's partition Ω
  - pmf: $p_{X|Y}(x|y) = \dfrac{p_{Y|X}(y|x)}{\sum_{x'} p_{Y|X}(y|x')p_X(x')} p_X(x)$
  - pdf: $f_{X|Y}(x|y) = \dfrac{f_{Y|X}(y|x)}{\int f_X(x')f_{Y|X}(y|x')\,dx'} f_X(x)$
  - mixed: $p_{\Theta|Y}(\theta|y) = \dfrac{f_{Y|\Theta}(y|\theta)}{\sum_{\theta'} p_\Theta(\theta')f_{Y|\Theta}(y|\theta')} p_\Theta(\theta)$

$$f_{Y|\Theta}(y|\theta) = \frac{p_{\Theta|Y}(\theta|y)}{\int f_Y(y')p_{\Theta|Y}(\theta|y')dy'} f_Y(y)$$

- Signal detection: MAP, ML, and minimum distance decoding rules

- Independence
  - events: for all subsets of the set of events $A_1, A_2, \ldots, A_n$,

$$\mathrm{P}(A_{i_1}, A_{i_2}, \ldots, A_{i_k}) = \prod_{j=1}^k \mathrm{P}(A_{i_j})$$

  - cdf: $F_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n F_{X_i}(x_i)$ for all $\mathbf{x}$
  - pmf: $p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n p_{X_i}(x_i)$ for all $\mathbf{x}$
  - pdf: $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n f_{X_i}(x_i)$ for all $\mathbf{x}$
- Conditional independence: for all x_1, x_2, x_3,

$$f_{X_1,X_3|X_2}(x_1, x_3|x_2) = f_{X_1|X_2}(x_1|x_2)f_{X_3|X_2}(x_3|x_2)$$

- Uncorrelation: $\mathrm{Cov}(X, Y) = \mathrm{E}\big((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\big) = 0$
  - independence $\Rightarrow$ uncorrelation, but not conversely

- Functions of random variables
- Application: generation of random variables ($Y = F_X(X)$ is $\mathrm{U}[0,1]$)
- Conditional expectation
  - conditional expectation is a random variable $\mathrm{E}(g(X,Y) \,|\, Y)$ that takes the values $\mathrm{E}(g(X,Y) \,|\, Y = y) = \int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x|y)\, dx$
  - iterated expectation: $\mathrm{E}(g(X,Y)) = \mathrm{E}_Y\left(\mathrm{E}_X(g(X,Y)|Y)\right)$
  - $\mathrm{E}(X \,|\, Y)$ is the best MSE estimate of X given Y; its MSE is $\mathrm{E}(\mathrm{Var}(X \,|\, Y))$
- Bounds
  - Union of events: $\mathrm{P}(\cup_{i=1}^{n} A_i) \le \sum_{i=1}^{n} \mathrm{P}(A_i)$
  - Cauchy-Schwarz: $(\mathrm{E}(XY))^2 \le \mathrm{E}(X^2)\,\mathrm{E}(Y^2)$; equality iff $X = aY$
  - Jensen: if $g(X)$ is convex, then $\mathrm{E}(g(X)) \ge g(\mathrm{E}(X))$
  - Markov: if $X \ge 0$ and $a > 1$ then $\mathrm{P}\{X \ge a\,\mathrm{E}(X)\} \le \dfrac{1}{a}$
  - Chebychev: $\mathrm{P}\{|X - \mathrm{E}(X)| \ge a\sigma_X\} \le \dfrac{1}{a^2}$

First and Second Moments, Correlation

- Two random variables
  - mean: $\mathrm{E}(X)$
  - second moment: $\mathrm{E}(X^2)$, also called average power
  - variance: $\mathrm{Var}(X) = \mathrm{E}\left((X - \mathrm{E}(X))^2\right) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$
  - correlation: $\mathrm{E}(XY)$
  - covariance: $\mathrm{Cov}(X,Y) = \mathrm{E}\left((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$
- Random vectors
  - mean vector: $\mathrm{E}(\mathbf{X}) = \boldsymbol{\mu}_{\mathbf{X}}$
  - correlation matrix: $\mathrm{E}(\mathbf{X}\mathbf{X}^T)$
  - covariance matrix: $\Sigma_{\mathbf{X}} = \mathrm{E}\left((\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\right) = \mathrm{E}(\mathbf{X}\mathbf{X}^T) - \mathrm{E}(\mathbf{X})\,\mathrm{E}(\mathbf{X}^T)$
    A matrix can be a covariance (correlation) matrix iff it is real, even, nonnegative definite
  - crosscovariance matrix: $\Sigma_{\mathbf{XY}} = \mathrm{E}(\mathbf{X}\mathbf{Y}^T) - \mathrm{E}(\mathbf{X})\,\mathrm{E}(\mathbf{Y}^T)$

- Random processes
  - mean function: $\mathrm{E}(X(t))$
  - autocorrelation function: $R_X(t_1, t_2) = \mathrm{E}(X(t_1)X(t_2))$
  - crosscorrelation function: $R_{XY}(t_1, t_2) = \mathrm{E}(X(t_1)Y(t_2))$
- WSS random processes
  - $\mathrm{E}(X(t))$ is constant and $R_X(t_1, t_2) = R_X(t_1 - t_2) = R_X(\tau)$
  - $R(\tau)$ is an autocorrelation function iff it is real, even, nonnegative definite (equivalent to $S_X(f) = \mathcal{F}[R(\tau)] \geq 0$ for all f)

    $|R_X(\tau)| \leq R_X(0) = \mathrm{E}(X^2(t))$ (average power)
    If $R_X(T) = R_X(0)$ then $X(t)$ and $R_X(\tau)$ are periodic with period T
  - Power spectral density: $S_X(f) = \mathcal{F}[R_X(\tau)]$ real, even and ≥ 0
  - crosscorrelation function: $R_{XY}(t_1 - t_2) = R_{XY}(\tau)$
    cross power spectral density: $S_{XY}(f) = \mathcal{F}[R_{XY}(\tau)]$

Gaussians

- Gaussian random variable: $X \sim \mathcal{N}(\mu, \sigma^2)$ has pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Gaussian random vector: $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ has joint pdf

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

  where $|\Sigma| = \det \Sigma$.
- Properties of Gaussian random vectors
  - uncorrelation implies independence
  - linear transformation of GRV yields GRV; if A is $m \times n$ full rank matrix, where $m \leq n$,

$$\mathbf{Y} = A\mathbf{X} \sim \mathcal{N}(A\boldsymbol{\mu}_{\mathbf{X}},\, A\Sigma_{\mathbf{X}}A^T)$$

  - marginals of GRV are Gaussian

◦ Conditionals of GRV are Gaussian; if

$$\begin{bmatrix} \mathbf{Y} \\ -- \\ \mathbf{X} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ -- \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{Y}} & \mid & \Sigma_{\mathbf{YX}} \\ -- & \mid & -- \\ \Sigma_{\mathbf{XY}} & \mid & \Sigma_{\mathbf{X}} \end{bmatrix} \right),$$

then

$$\mathbf{X} \mid \{\mathbf{Y} = \mathbf{y}\} \sim \mathcal{N}(\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{Y}}^{-1}\mathbf{y},\ \Sigma_{\mathbf{X}} - \Sigma_{\mathbf{XY}}\Sigma_{\mathbf{Y}}^{-1}\Sigma_{\mathbf{YX}})$$

◦ best MSE estimate of GRV given GRV is linear

- Gaussian random process: all finite order pdfs (joint pdfs of a finite set of sample) are Gaussian
  - ◦ Examples: discrete-time WGN, discrete-time Wiener process, Gauss-Markov, bandlimited WGN
  - ◦ WSS ⇒ SSS

Convergence and Limit Theorems

- w.p.1: $\mathrm{P}\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\} = 1$ (no problems in final on this)
- in m.s.: $\lim_{n\to\infty} \mathrm{E}\left((X_n - X)^2\right) = 0$
- in probability: $\lim_{n\to\infty} \mathrm{P}\{|X_n - X| > \epsilon\} = 0$ for every $\epsilon > 0$,
  - ◦ Weak law of large numbers: if $X_1, X_2, X_3, \ldots$ are i.i.d. r.v.s with finite mean and variance, then $S_n \to \mathrm{E}(X)$ in probability
- in distribution: $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$
- Central limit theorem: if $X_1, X_2, X_3, \ldots$ are i.i.d. r.v.s with finite mean $\mathrm{E}(X)$ and variance σ_X^2 then

$$\sum_{i=1}^{n} \frac{(X_i - \mathrm{E}(X))}{\sigma_X \sqrt{n}} \to \mathcal{N}(0,1) \text{ in distribution}$$

- CLT also holds for i.i.d sequence of random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ with finite mean $\boldsymbol{\mu}$ and nonsingular covariance matrix Σ :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{X}_i - \boldsymbol{\mu}) \rightarrow \mathcal{N}(\mathbf{0}, \Sigma) \text{ in distribution}$$

- Relationships

with probability 1 $\Longrightarrow$ in probability $\Longrightarrow$ in distribution

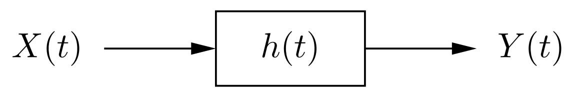
in mean square $\Longrightarrow$ in probability

Random Processes

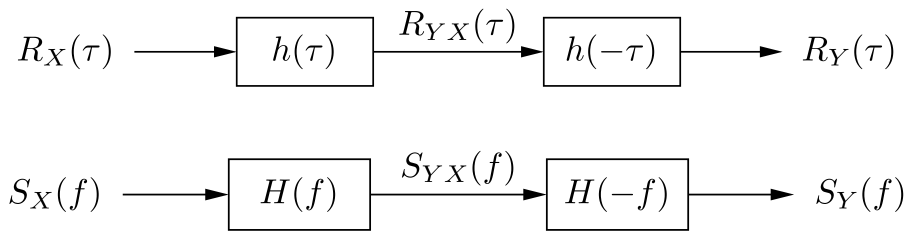
- IID: Bernoulli, discrete time WGN
- Markov: IID, random walk, Gauss-Markov
- Independent increment: random walk, discrete-time Wiener
- Gaussian: WGN, discrete-time Wiener, Gauss-Markov
- SSS: all n-th order distributions are time invariant, e.g., periodic signal with random phase, IID, WSS Gaussian
- WSS: mean and autocorrelation functions are time invariant
- Discrete-time white noise: WSS with zero mean and $S_X(f) = N$, $|f| < \frac{1}{2}$
- Continuous-time bandlimited white noise: zero mean and $S_X(f) = \frac{N}{2}$, $|f| < B$
- White noise: WSS with zero mean and $S_X(f) = \frac{N}{2}$ for all f

Response of LTI System To WSS Process Input

- $X(t)$ WSS process, $-\infty < t < \infty$



- $X(t)$ and $Y(t)$ are jointly WSS with



Linear MSE Estimation

- Scalar case: Signal X and observation Y are zero mean

$$\hat{X} = \frac{\sigma_{XY}}{\sigma_Y^2} Y = \sigma_X \rho_{X,Y} \frac{Y}{\sigma_Y}$$

$$\text{MSE} = \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2} = (1 - \rho_{X,Y}^2)\sigma_X^2$$

- Vector case: Signal X and observation $\mathbf{Y}$ are zero mean

$$\hat{X} = \Sigma_{X\mathbf{Y}} \Sigma_{\mathbf{Y}}^{-1} \mathbf{Y}$$

$$\text{MSE} = \sigma_X^2 - \Sigma_{X\mathbf{Y}} \Sigma_{\mathbf{Y}}^{-1} \Sigma_{\mathbf{Y}X}$$

- WSS process infinite smoothing case:

  Signal $X(t)$ and observation $Y(\tau)$ for $-\infty < \tau < \infty$ are zero mean jointly WSS

  The best linear MSE estimate is of the form

$$\hat{X}(t) = h(t) * Y(t) = \int_{-\infty}^{\infty} Y(\tau)h(t-\tau)\, d\tau\,,$$

  where the transfer function and the MSE of the best linear MSE estimate are

$$H(f) = \frac{S_{XY}(f)}{S_Y(f)}$$

$$\text{MSE} = \int_{-\infty}^{\infty} S_X(f)\, df - \int_{-\infty}^{\infty} \frac{|S_{XY}(f)|^2}{S_Y(f)}\, df$$

$$= \mathrm{E}\left((X(t))^2\right) - \left(R_{XY}(\tau) * h(-\tau)\right)\Big|_{\tau=0}$$

Where do you go from here?

- EE courses that use EE 278:
  - Communications: EE 276, 279, 359, 374, 376A,B, 379A,B,C
  - Signal and image processing: EE378, 355, 363, 368, 372, 398A,B
- More on probability and random processes:
  - Stat 217, 218
  - Stat 310A,B,C
  - Stat 317