Conditional Probability: **Probability of B given A:** $P(B|A)=\frac{P(A\cup B)}{P(A)}\rightarrow P(A\cup B)=P(A)\cdot P(B|A)=P(B)\cdot P(A|B)$ **Bayes Theorem:** $P(B|A)=\frac{P(A|B)\cdot P(B)}{P(A)}$; $P(A|B)=\frac{P(B|A)\cdot P(A)}{P(B)}$

Independence Events A,B are **statistically independent** IFF: $P(A\cup B)=P(A)\cdot P(B)$ **In general:** $\{A_i\}_{i=1}^n$ stat.indept IFF $\forall 2\le k\le n$: $P\left(\bigcap_{i=1}^k A_i\right)=\prod_{i=1}^k P(A_i)$

Random Variable: Random variable X is a transformation from the sampling space to real space: $X:\Omega\rightarrow R$

Discrete: X is DRV iff $F_X(x)$ is a steps function: $F_X(x)=\sum_i p_i\cdot U(x-x_i)$ **Mixed:** X is MRV iff $\exists$: $0<\alpha<1$, $F_{X_d}(\cdot)$ =CDF of a DRV, $F_{X_c}(\cdot)$ =CDF of a CRV, s.t. $F_X(x)$: $F_X(x)=\alpha F_{X_d}(x)+(1-\alpha)F_{X_c}(x)\ \forall x$

Cumulative Distribution Function (CDF): Every RV has a CDF, defined: $F_X(x)=P(X\le x)$

Properties of the CDF:		Calculating Probabilities using the CDF:	
$\forall x\in R,\ 0\le F(X)\le 1$	$\lim_{x\to\infty}F_X(x)=1$; $\lim_{x\to-\infty}F_X(x)=0$	$P(x<X\le y)=F_X(y)-F_X(x)$	$P(x\le X\le y)=F_X(y)-F_X(x)+P(X=x)$
$F_X(x)=F_X(x^+)$	Monotonically non-decreasing: if $x<y$ then $F_X(x)\le F_X(y)$	$P(X=x)=F_X(x)-F_X(x^-)$	$P(x\le X<y)=F_X(y)-F_X(x)+P(X=x)-P(X=y)$

Probability Distribution Function (PDF): If X is an RV, its PDF is defined as derivative of its CDF (when $\exists$): $f_X(x)=\frac{dF_X(x)}{dx}$

PDF Properties for continuous RV:

$\forall x;\ f_X(x)\ge 0$	$P(X=x)=0$	$P(a\le X\le b)=\int_a^b f_X(t)dt$
$P(X\le x)=F_X(x)=\int_{-\infty}^x f_X(t)dt$		

PDF Properties for Discrete RV:

$P(X\le x)=F_X(x)=\sum_{i=-\infty}^{x}f_X(x_i)$	$P(a\le X\le b)=\sum_{i=a^-}^{b^+}f_X(x_i)$	$P(a<X\le b)=\sum_{i=a^+}^{b^+}f_X(x_i)$
$P(a<x<b)=\sum_{i=a^+}^{b^-}f_X(x_i)$	$P(a\le X<b)=\sum_{i=a^-}^{b^-}f_X(x_i)$	$P(X=x)=\sum_{i=x^-}^{x^+}f_X(x_i)$

Function of Random Variable:

Let X be RV, and $g(\cdot)$ known deterministic function. Def: $Y=g(X)$. Find $F_Y(y)$ using $F_X(x)$: $F_Y(y)=P(Y\le y)=P(Y\in[-\infty,y])=P\left(X\in g^{-1}([-\infty,y])\right)$

Theorem: Let $Y=g(X)$, where X is an RV. Assume $Y=g(X)$ has a finite number of solutions denoted $\{x_i\}_{i=1}^n$. Assume that $\forall i\ g'(x)\neq 0$. So, PDF of Y is: $f_Y(y)=\sum_{i=1}^n\frac{f_X(x_i)}{|g'(x_i)|}$

Remark: number of solutions might be ∞. Demand it to be countable. If not - can't use this.

Joint Distribution Functions

Joint CDF: of n RV: $F_{X_1,X_2,\ldots X_n}(x_1,x_2,\ldots,x_n)=F_{\underline{X}}(\underline{x})=P(X_1\le x_1,X_2\le x_2,\ldots,X_n\le x_n)$

Joint PDF: Joint PDF of n random variables: $f_{X_1,X_2,\ldots X_n}(x_1,x_2,\ldots,x_n)=f_{\underline{X}}(\underline{x})=\frac{\partial^n F_{X_1,X_2\ldots X_n}(x_1,x_2,\ldots,X_n)}{\partial x_1\ldots\partial x_n}$

Characteristics

- $P(x_1\le X\le x_2,y_1\le Y\le y_2)=(x_2,y_2)-F_{XY}(x_1,y_2)-F_{XY}(x_2,y_1)+F_{XY}(x_1,y_1)$
- $F_{XY}(-\infty,y)=F_{XY}(x,-\infty)=0$
- $P(X\le x,y_1\le y_2)=F_{XY}(x,y_2)-F_{XY}(x,y_1)$
- $F_{XY}(x,\infty)=F_X(x)$
- $F_{XY}(\infty,\infty)=1$
- $P(x_1\le X\le x_2,Y\le y)=F_{XY}(x_2,y)-F_{XY}(x_1,y)$
- $F_{XY}(\infty,y)=F_Y(y)$

- $f_{X_1,X_2,\ldots X_n}(x_1,x_2,\ldots,x_n)\ge 0\ \forall x_1\ldots x_n$ and $\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}f_{X_1,X_2,\ldots X_n}(x_1,x_2,\ldots,x_n)dx_2dx_3\ldots dx_n=1$
- $F_{X_1,X_2,\ldots X_n}(x_1,x_2,\ldots,x_n)=\int_{-\infty}^{x_1}\ldots\int_{-\infty}^{x_n}f_{X_1,X_2,\ldots X_n}(\alpha_1,\alpha_2,\ldots,\alpha_n)d\alpha_n\ldots d\alpha_1$
- **Marginal PDF** (of RV X_1): $f_{X_1}(x_1)=\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}f_{X_1,X_2,\ldots X_n}(x_1,x_2,\ldots,x_n)dx_2dx_3\ldots dx_n$

Conditional PDF: Given discrete random variables $B_1,\ldots B_m$ and continuous RVs $X_1,\ldots X_k$, the following holds:

$f_{X_1,\ldots X_n|Y_1\ldots Y_k}(x_1,\ldots,x_n\mid y_1,\ldots,y_k)=\frac{f_{X_1\ldots X_n,Y_1\ldots Y_k}(x_1,\ldots x_n,y_1,\ldots y_k)}{f_{Y_1\ldots Y_k}(y_1,\ldots y_k)}$

$f_{X_1\ldots X_n,Y_1\ldots Y_k}(x_1,\ldots,x_n,y_1,\ldots,y_k)=f_{X_1\ldots X_n|Y_1\ldots Y_k}(x_1,\ldots,x_n|y_1,\ldots,y_k)f_{Y_1\ldots Y_k}(y_1,\ldots,y_k)=f_{Y_1\ldots Y_k|X_1\ldots X_n}(y_1,\ldots y_k|x_1,\ldots x_n)\cdot f_{X_1\ldots X_n}(x_1,\ldots,x_n)$

Law of Total Prob: $P(B_1=b_1,B_2=b_2,\ldots B_m=b_m)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}P(B_1=b_1,B_2=b_2,\ldots B_m=b_m\mid X_1=x_1,X_2=x_2,\ldots X_k=x_k)f_{X_1\ldots X_k}(x_1,x_2,\ldots x_k)dx_1dx_2\ldots dx_k$; $f_{X_1,\ldots X_k}(x_1,x_2,\ldots,x_k)=\sum_{b_1}\sum_{b_2}\ldots\sum_{b_m}f_{X_1\ldots X_k|B_1\ldots B_m}(x_1,\ldots,x_k|b_1,b_2,\ldots,b_m)P(B_1=b_1,B_2=b_2\ldots,B_m=b_m)$

Conditional Prob: $f_{X_1\ldots X_k|B_1\ldots B_m}(x_1,\ldots,x_k|b_1,b_2,\ldots,b_m)=\frac{P(B_1=b_1,B_2=b_2\ldots B_m=b_m\mid X_1=x_1,\ldots X_k=x_k)f_{X_1\ldots X_k}(x_1\ldots x_k)}{P(B_1=b_1,B_2=b_2,\ldots B_m=b_m)}$; $P(B_1=b_1,B_2=b_2,\ldots B_m=b_m\mid X_1=x_1,\ldots X_k=x_k)=\frac{f_{X_1\ldots X_k|B_1\ldots B_m}(x_1,\ldots,x_k\mid b_1,\ldots,b_m)P(B_1=b_1,B_2=b_2,\ldots,B_m=b_m)}{f_{X_1\ldots X_k}(x_1\ldots x_k)}$

Cond. Expectation: $E[X|Y=y]=\int_{-\infty}^{\infty}x\cdot f_{X|Y}(x|y)dx$; $[X|Y_1=y_1,Y_2=y_2,\ldots,Y_k=y_k]=\int_{-\infty}^{\infty}x\cdot f_{X|Y_1\ldots Y_k}(x|y_1,\ldots,y_k)dx$

Laplace Distr: $f_X(x)=\frac{\lambda}{2}e^{-\lambda|x|}$, $E[X]=0,E[X^2]=Var(X)=2/\lambda^2$

Smoothing Theorem (Law of Total Expectation):

If X is an RV with expected value $E[X]$, and Y is any RV on the same probability space, then: $E[X]=E\left[E[X|Y]\right]$, i.e. expected value of the conditional expected value of X given Y is same as the expected value of X. General formulation: $E[g(X,Y)]=E\left[E[g(X,Y)|Y]\right]$ and $E[X_1]=E\left[E[X_1|Y_1,\ldots,Y_k]\right]$. **Smoothing Find Conditional Expectation ("simple example"):** $E[X_1|Y_1,Y_2,\ldots Y_k]=E\left[E[X_1|Z_2,Z_2,\ldots,Z_m,Y_1,Y_2,\ldots Y_k]|Y_1,Y_2,\ldots Y_k\right]$

Joint Moments:

Covariance: $\sigma_{XY}=Cov(X,Y)=E[(X-E[X])\cdot(Y-E[Y])]=E[XY]-E[X]E[Y]=\mu_{11}$ **Correlation:** $m_{11}=E[XY]$ **Correlation Coeff.:** $\rho=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}$

Joint Moment: $m_{nk}=E[X^n\cdot Y^k]$ for $n+k=p$. Call m_{nk} the joint moment of order p **Joint Central Moments:** $\mu_{nk}=E[(X-E[X])^n\cdot(Y-E[Y])^k]$

Orthogonal: 2 RVs are orthogonal if correlation between them if 0: $E(X,Y)=0$. **Uncorrelated:** 2 RVs are uncorrelated if the covariance between tem is 0: $Cov(X,Y)=0$

Statistical Independence

Independence of a Pair of RVs: RVs X and Y are statistically independent IFF:
- $F_{XY}(x,y)=F_X(x)F_Y(y)$
- $F_{X|Y}(x|y)=F_X(x)$
- $f_{X|Y}(x|y)=f_X(x)$
- $f_{XY}(x,y)=f_X(x)\cdot f_Y(y)$

Independence Random Vectors:: $\underline{X}=[X_1,\ldots,X_n]^T$ $\underline{Y}=[Y_1,\ldots,Y_m]^T$, independent IFF:
- $F_{\underline{X},\underline{Y}}=F_{\underline{X}}(\underline{x})F_{\underline{Y}}(\underline{y})\ \ \forall\underline{x},\underline{y}$
- $F_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y})=F_{\underline{X}}(\underline{x})\ \ \forall\underline{x},\underline{y}$
- $f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y})=f_{\underline{X}}(\underline{x})\ \ \forall\underline{x},\underline{y}$
- $f_{\underline{X},\underline{Y}}(\underline{x},\underline{y})=f_{\underline{X}}(\underline{x})f_{\underline{Y}}(\underline{y})\ \ \forall\underline{x},\underline{y}$

Uncorrelated RVs: if $\rho_{XY}=0$

Orthogonality: RVs orthogonal if $E[XY]=0$

Logic Claims
- For 2 indepd random vectors $\underline{X},\underline{Y}$, $\forall g(\cdot)$ and $f(\cdot)$ deterministic scalar functs, this holds: $E[g(\underline{X})\cdot f(\underline{Y})]=E[g(\underline{X})]\cdot E[f(\underline{Y})]$
- Two uncorrelated RVs that are jGaus → statistical independence
- For two independent random variables X and Y, $\forall n,k$: $\mu_{1k}=\mu_{n1}=0$
- No correlation & one of the variables has expectation 0 → orthogonality
- Statistical independence → no correlation
- For two RVs X,Y: $Var(X+Y)=Var[X]+2\cdot Cov(X,Y)+Var(Y)$. Thus, variables are uncorrelated: $Var(X+Y)=Var(X)+Var(Y)$
- Joint Gaussian: 2 RVs that are JGauss with 0 covariance are stat. independent

Characteristic Function of Random Vectors/Variables

Characteristic Function of a Pair of Random Variables

For a RVs X,Y: $\phi_{XY}(\omega_1,\omega_2)=E\left[e^{j(\omega_1X+\omega_2Y)}\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{XY}(x,y)e^{j(\omega_1x+\omega_2y)}dxdy$. Moreover: $f_{XY}(x,y)=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi_{XY}(\omega_1,\omega_2)e^{j(\omega_1x+\omega_2y)}d\omega_1d\omega_2$. Moment order p: $m_{nk}=\frac{1}{j^p}\frac{\partial^p\phi_{XY}(\omega_1,\omega_2)}{\partial\omega_1^n\partial\omega_2^k}\Big]_{\omega_1=\omega_2=0}$, $n+k=p$

Characteristics: $\phi_{XY}(0,0)=1$ | $\phi_Y(\omega)=\phi_{XY}(0,\omega)$ | $\phi_X(\omega)=\phi_{XY}(\omega,0)$

Characteristic function of Random Vectors:

For a random vector $\underline{X}$ of length n, first characteristic function is defined as : (The vector $\underline{\omega}$ is of length n.) $\phi_X(\underline{\omega})=E\left[e^{j\underline{\omega}^T\underline{X}}\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}f_{\underline{X}}(\underline{x})e^{j\underline{\omega}^T\underline{x}}d\underline{x}$

$f_{\underline{X}}(\underline{x})=\frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\phi_{\underline{X}}(\underline{\omega})e^{j\underline{\omega}^T\underline{x}}d\underline{\omega}$

Moments can be made thru the 1st characteristic func similar to how we do it for RVs.

Law of Total Probability:

Given a set of disjoint events $\{A_i\}_{i=1}^{\infty}$ which meet $\bigcup_{i=1}^{\infty}A_i=\Omega$ (such sets are called partitions of probability space). Then for any event B we have: $(B)=P(B\cap\Omega)=P\left(B\cap\left(\bigcup_{i=1}^{\infty}A_i\right)\right)=P\left(\bigcup_{i=1}^{\infty}(B\cap A_i)\right)=\sum_{i=1}^{\infty}(B\cap A_i)=\sum_{i=1}^{\infty}P(B|A_i)P(A_i)$. The result is also correct for a finite partition ($\{A_i\}_{i=1}^n$). For that case, replace upper bounds with n in above expression.

Moments: Let X be an RV. We define for it:

Variance: $Var[X]=\int_{-\infty}^{\infty}(\alpha-E[X])^2\cdot f_X(\alpha)d\alpha=E[X^2]-E^2[X]$. Denoted σ^2,σ_X^2

Variance (Discrete): $Var[X]=\sum_i(x_i-E[X])^2\cdot P(X=x_i)=\sum_i x_i^2\cdot P(X=x_i)-E^2[X]$

Moment Order n: $m_n=E[X^n]=\int_{-\infty}^{\infty}\alpha^n\cdot f_X(\alpha)d\alpha$

Moment Order n (Dscr.): $m_n=E[X^n]=\sum_i x_i^nP(X=x_i)$

Notes: $E[X]=m_1$.| $Var[X]=\mu_2$. $Var[X]\ge 0\ \forall X$.| $Var[X]=0$ then $X=\eta_X$

Expected Value: $E[X]=\int_{-\infty}^{\infty}\alpha\cdot f_X(\alpha)d\alpha$. Denoted as η,η_X | **(Discrete):** $E[X]=\sum_i x_i\cdot P(X=x_i)$

Central Moment Order n: $\mu_n=E[(X-E[X])^2]$. For RV with $E[X]=0$ we have $\mu_n=m_n$

Variance of an RV=0 IFF its deterministic

Variance using Moments: $Var[X]=m_2-m_1^2$

Expectation is defined when the integral or sum converge absolutely. In cont.case: $\int_{-\infty}^{\infty}|\alpha|\cdot f_X(\alpha)d\alpha<\infty$

$Var(aX+bY)=a^2Var(X)+b^2Var(Y)+2Cov(X,Y)ab$

First Characteristic Function of Random Variables PDF and the Characteristic function are a Fourier Pair.

For RV X, its first characteristic function $\phi_X(\omega)$ is **defined as:** $\phi_X(\omega)=E\left[e^{j\omega X}\right]=\int_{-\infty}^{\infty}f_X(x)e^{j\omega x}dx$. Following holds: $f_X(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_X(\omega)e^{-j\omega x}d\omega$.

Moments from char func: $m_n=E[X^n]=\frac{1}{j^n}\cdot\frac{d^m\phi_X(\omega)}{d\omega^n}\Big]_{\omega=0}$

If X DRV: $\phi_X(\omega)\overset{\text{def}}{=}E\left[e^{j\omega X}\right]=\sum_i e^{j\omega x_i}P(X=x_i)$

If func Analytical: $\phi_X(\omega)=\sum_{n=0}^{\infty}m_n\cdot\frac{(j\omega)^n}{n!}$

Conclusion: If Characteristic Function analytical, the sequence of moments of RV defines it uniquely, so, defines PDF uniquely.

Tail Index Estimation

Markov's Ineq-ty: If $f_X(x)=0\ \forall x<0$, then $\forall\alpha>0$: $P(X\ge\alpha)\le\frac{\eta_X}{\alpha}$. **Proof:** $\eta=E[X]=_{X>0}\int_0^{\infty}xf_X(x)dx\ge\int_{\alpha}^{\infty}xf_X(dx)\ge\alpha\int_{\alpha}^{\infty}f_X(x)dx=\alpha\cdot P(X\ge\alpha)\rightarrow P(X\ge\alpha)\le\eta/\alpha$

Chernoff's Inequality: For an RV X: $P(X\ge\epsilon)\le e^{-a\epsilon}E[e^{aX}]$; $\forall a\ge 0$ and $P(X\le\epsilon)\le e^{-a\epsilon}E[e^{aX}]$; $\forall a\le 0$. **Alternatively:** $P(X\le\epsilon)\le e^{a\epsilon}E[e^{-aX}]$ $\forall a\ge 0$ **Proof:** $P(X\ge\epsilon)=_1 P(e^{aX}\ge e^{a\epsilon})\le_2 e^{-a\epsilon}E[e^{aX}]$. 1: function e^{ax} increases monotonically. 2: Markov's inequality.

Moment Generating Function: $E[e^{aX}]$ Is a func of determ-ic variable a. $M_X(a)=E[e^{aX}]$ then get: $\frac{d^n}{da^n}M_X(a)=E[X^ne^{aX}]\rightarrow m_n=E[X^n]=\frac{d^n}{da^n}M_X(a)\Big]_{a=0}$ If a is a +ve number, $M_X(\cdot)$ is 2-sided Laplace of $f_X(\cdot)$

Chebyshev's Inequality: For RV X with exect μ & var σ^2, $\forall\alpha>0$: $P(|X-\mu|\ge\alpha)\le\frac{\sigma^2}{\alpha^2}$

Partial Second Order Statistics:

Expectation of a random Vector:: $\underline{X}=\begin{bmatrix}X_1\\X_2\\\vdots\\X_n\end{bmatrix}$, $E[\underline{X}]=\underline{\eta}_X=\begin{bmatrix}E[X_1]\\E[X_2]\\\vdots\\E[X_n]\end{bmatrix}$

Correlation Matrix (deterministic): $R_{XX}=E[\underline{X}\underline{X}^T]=\begin{bmatrix}E[X_1X_1] & E[X_1X_2] & \ldots & E[X_1X_n]\\E[X_2X_1] & E[X_2X_2] & & E\{X_2X_n]\\\vdots & \vdots & \ddots & \vdots\\E[X_nX_1] & E[X_nX_2] & \ldots & E[X_nX_n]\end{bmatrix}$

Covariance Matrix (deterministic matrix): $C_{XX}=E\left[\left(\underline{X}-\underline{\eta}_X\right)\left(\underline{X}-\underline{\eta}_X\right)^T\right]=E[\underline{X}\underline{X}^T]-\underline{\eta}_X\underline{\eta}_X^T$

Cross-correlation matrix: $R_{XY}=E[\underline{X}\underline{Y}^T]=R_{YX}^T$

Cross-Covariance matrix: $C_{XY}=E\left[\left(\underline{X}-\underline{\eta}_X\right)\left(\underline{Y}-\underline{\eta}_Y\right)^T\right]=E[\underline{X}\underline{Y}^T]-\underline{\eta}_X\underline{\eta}_Y^T=C_{YX}^T$

Note auto-covariance and auto-correlation matrices are *positive, semi definite*, and *symmetrical*, thus they are *diagonalizable*., cross covariance mats are not necessarily positive semi-definite or symmetrical.

Linear Transformation of a random vector: Consider random vector $\underline{X}$ with covariance matrix C_{XX} and expectation vector $\underline{\eta}_X$: We define $\underline{Y}=\boldsymbol{A}\underline{X}+\underline{b}$, where A is a deterministic matrix and $\underline{b}$ is a deterministic vector. Then: $C_{YY}=\boldsymbol{A}C_{XX}\boldsymbol{A}^T$; $\underline{\eta}_Y=\boldsymbol{A}\underline{\eta}_X+\underline{b}$

Central Limit Theorem

For an i.i.d sequence of variable $\{X_i\}_{i=1}^{\infty}$ who distribute with finite expected value μ & finite variance $\sigma^2 < \infty$, define their normalized sum: $S_n = \frac{\sum_{i=1}^{n} X_n - n\mu}{\sqrt{n\sigma^2}}$. Then we have: $lim_{n\to\infty}\phi_{S_n}(\omega) = \lim_{n\to\infty} E[e^{i\omega S_n}]$ and $\lim_{n\to\infty} F_{S_n}(s) = \lim_{n\to\infty} P[S_n \le s] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$. I.e, distribution of S_n approaches (as $n \to \infty$) that of of a standard Gaussian Variable.

Gaussian Random Vector

Whitening: X is Gauss Vec, A is determin. Matrix, b is determin. Vec. Then $Y = AX + b$ is Gauss. If X is arbitrary RV, then $E(Y) = AE(X) + b, C_Y = AC_X A^T$. We want to whiten X to get $E(Y) = 0, C_Y = \boldsymbol{I}$

Definition: We say that $X \sim N(\mu, \sigma^2)$ when it has a PDF of: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Characteristic Function: $\phi_X(\omega) = E[e^{i\omega X}] = e^{i\omega\mu - \frac{1}{2}\sigma^2\omega^2}$. For $\sigma^2 = 0$ X is deterministic and $f_X(x) = \delta(x - \mu)$

Jointly Gaussian: The set of RVs $X_1, X_2, X_3 \dots X_n$ is Jointly Gaussian IFF every linear combination of them is Gaussian. Equivalently, $X_1, X_2, X_3 \dots X_n$ are jointly Gaussian IFF they can be represented as an affine transformation of n independent Gaussian Variables: $\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \underline{A}\begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} + \underline{b}$; $Z_i \sim N(0,1)$, $\{Z_i\}$ are IID, Where A is matrix and b is a vector

For Gauss RV: $E[X^4] = 3\sigma^4$, $E[X^{\text{odd}}] = 0$ (true if they have 0 mean??)

The Q-Function

Seen as how an analytical expression of the CDF of a Gaussian RV does not exist, we use the Q-function to deal with these cases. Let us denote $Z \sim N(0,1)$, then it holds that: $Q(x) = P(Z \ge y) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{\frac{z^2}{2}} dz$. $\forall$ Gaussian RV $Y \sim N(\mu, \sigma^2)$ s.t $\sigma^2 > 0$: $P(Y \ge y) = P\left(\frac{Y-\mu}{\sigma} \ge \frac{y-\mu}{\sigma}\right) = P\left(Z \ge \frac{y-\mu}{\sigma}\right) = Q\left(\frac{y-\mu}{\sigma}\right)$

Whitening Process: (1) Subtract mean vector, to obtain a Gaussian RVector w. zero mean. (2) Applying a linear transformation which makes the covariance matrix diagonal. (3) Applying normalization of both components, to obtain a variance of 1 for both components. **NOTE:** Given Gaussian RVec with invertible covariance matrix, by whitening we can get new Gaussian Rvec with any desired mean, correlation, variance

Random Gaussian Vectors: A random vector is Gaussian IFF every linear combination of its components is a Gaussian variable.

The **characteristic function of a Gaussian**: $\phi_{\underline{X}}(\underline{\omega}) = e^{j\underline{\eta}_X^T - \frac{1}{2}\underline{\omega}^T C_{XX}\underline{\omega}}$ where $\underline{\eta}_X$ is the expectation vector of $\underline{X}$ and C_{XX} is the covariance matrix of $\underline{X}$.

Gaussian PDF (if cov invertible): If covmatrix invertible, then $\underline{X}$ is a continuous Gaussian vector with PDF of $f_{\underline{X}}(\underline{x}) = \frac{1}{\sqrt{(2\pi)^n (\det C_{XX})}} e^{-\frac{1}{2}(\underline{x} - \underline{\eta}_X)^T C_{XX}^{-1}(\underline{x} - \underline{\eta}_X)}$.

Random vector with IID components → Gaussian (reverse not true)

Linear transformation of random Gaussian vector is a Gaussian Vector

Gauss RV PDF($\sigma^2 > 0$): $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $E[X] = \mu, E[X^2] = Var(X) = \sigma^2$

If $\underline{Y} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}$ is Gaussian then $\underline{X}_2 | \underline{X}_1$ is Gaussian

If $\underline{Y} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}$ is Gaussian and the Cross-covariance($\underline{X}_1, \underline{X}_2$) is 0, then X_1, X_2 are independent Gaussian Vectors

Gauss RV PDF($\sigma^2 = 0$): $f_X(x) = \delta(x - \mu)$

Fourth Moment of a Gaussian Random Vector (THERE IS A PROOF THAT WE OMMITED)

Jointly Gaussian: RVs $X_1, X_2, \dots, X_n$ JGaussian IFF each linear combination of them is Gaussian

For $\underline{X} = (X_1, X_2, X_3, X_4)$ a Gaussian random vector with expected value 0 (i.e $\underline{\eta}_X = 0$) it holds: $E[X_1X_2X_3X_4] = E[X_1X_2]E[X_3X_4] + E[X_1X_3]E[X_2X_4] + E[X_1X_4]E[X_2X_3]$

Estimation

Problem Definition: For given RV Y want to find an estimation of another RV X. The relationship between X and Y is some statistical connection which we can derive. For example, from the passing through some system: $\xrightarrow{X} \boxed{\text{Random System}} \xrightarrow{Y}$ where the system can implement addition of some kind of noise (an RV), a multiplication by another RV, etc. In essence, $\exists$ statistical model which connects between Y (the measurements) and X – the parameter we wish to estimate. Given Y we want to estimate X.

Method: The method of estimation depends on the desired quality of estimation. Namely, what is considered a good (optimal) estimator. For example an estimator that minimizes the error probability $P(\hat{X} \ne X)$ can be found, as well as an estimator minimizing the Mean Square Error $E\left[(X - \hat{X})^2\right]$. Overall we want to minimize the quantity: $D = E[d(e)]$; $e = X - \hat{X}$ Where $d(\cdot)$ is a non-negative distortion measure and e is the estimation error (an RV)

Estimator Function: Let us denote the estimator, which is a function of Y, as $\hat{X} = g(Y)$.

Argmax: Unlike the global max this returns the *inputs* at which the output is max.

$\boldsymbol{MSE} = E[e^2] = E[X^2] - E[\hat{X}^2] = Var(X) - Var(\hat{X}) = E[Var(X|Y)]$

Criterion of Minimum Probability of Error:

Distortion measure for the problem: $d(e) = \begin{cases} 1 & e \ne 0 \\ 0, & e = 0 \end{cases}$

Optimal Estimator: Solution of problem rising from Maximum A-posteriori Prob (MAP) estimation, its optimal est. is given by: $\hat{X}_{\text{MAP}} = g_{\text{MAP}}(y) = \arg\max_x P(X = x \,|Y = y)$

Criterion of Minimum Mean Square Error (MMSE):

Definitions:

Expected Value of Estimation Error: $E[e] = E[X - \hat{X}] = E[X] - E[\hat{X}] = \eta_X - E[\hat{X}]$. Our estimators have $E[e] = 0 \to E[\hat{X}] = \eta_X$. **Expected Value of MSE:** $E[e^2] = E\left[(X - \hat{X})^2\right]$. For estimator with expected value of estimation error=0: $Var(e) = E[e^2] - E^2[e] = E[e^2]$. **Distortion measure for the problem:** $d(e) = e^2$. **Solution:** Estimator that minimizes MSE (called MMSE estimator).

Optimal Estimator:

Optimal Estimator: brings MSE to a minimum, given by: $\hat{X}_{\text{opt}} = E[X|Y]$. Notice that $E[X|Y]$ is **only** a function of Y. | Optimal Estimator gives zero expectation error | **MSE in Opt. Estimator**: $Var(e) = E[e^2] = E\left[(X - \hat{X})^2\right] = E[E[X^2|Y] - E^2[X|Y]] = E[Var(X|Y)]$.Averaging of the conditional variance according to Y. | **Alt. expression for the MSE:** $Var(e) = \sigma_X^2 - \sigma_{\hat{X}_{\text{opt}}}^2$ | **Perpendicularity**: Estimator is MMSE IFF estimation error $\perp$ $\forall$ functions of the measurments: $\forall g(Y)$ $E[e \cdot g(Y)] = 0$

Optimal Linear Estimator: In form $aY + b$ (a,b constants), which brings the MSE to a minimum is given by: $\hat{X}_{\text{BLE}} = E[X] + \frac{Cov(X,Y)}{Var(Y)}(Y - E[Y])$

$\hat{X}_{MMSE} = E[X|Y = y]$

Perpendicularity Property (Opt Lin Est): An estimator is a BLE estimator IFF estimation error is $\perp$ $\forall$ linear functions of the measurements: $\forall a, b: E[e \cdot (aY + b)] = 0$

Optimal Lin Est-r gives **zero expectation error**

MSE for of optimal linear estimator: $Var(e) = E[e^2] = E\left[(X - \hat{X})^2\right] = \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2} = \sigma_X^2(1 - \rho_{XY}^2)$

Unbiased: $E[X] = E_Y[E[X|Y]] = E[\hat{X}]$

Pythagorean Property: $E[X^2] = E[\hat{X}^2] + E[e^2]$

Gaussian: $E[X|Y] = E[X] + \frac{Cov(X,Y)}{Var(Y)}(Y - E[Y])$

$\hat{X}_{MMSE} = \hat{X}_{LMMSE}$

$\hat{X}_{BLE} = E(X) + \frac{Cob(X,Y)}{Var(Y)}(Y - E(Y))$

Estimation of Random Vector from Random Vector

If all components of vectors X and Y are JGaussian, then the optimal estimator of X given Y by MMSE is the optimal linear.est

$\underline{X}$ **: estimated vec** | $\underline{Y}$ **: measurements vect** | $\underline{\hat{X}} = g(\underline{Y})$ **: estimator**
$\underline{e} = \underline{X} - \underline{\hat{X}}$ **: estimation error** | $d(\underline{X}, \underline{\hat{X}}) = \sum_{i=1}^{n} d(X_i, \hat{X}_i)$ **: divisible distortion measure**

Optimal Estimator Minimum Probability of Error Criterion: minimize $P(\hat{X}_i \ne X_i)$ $i = 1, \dots, n$
Optimal in MSE sense(MMSE Criterion): minimize $E[\|\underline{e}\|^2] = E[\sum_{i=1}^{N} e_i^2] = \sum_{i=1}^{N} E[e_i^2]$

Estimation of a random Vector from another Random Vector: Estimator $\underline{\hat{X}} = g(\underline{Y})$ is optimal in MMSE sense IFF the estimation error is orthogonal to any function of the measurements. The only solution is the conditional expectation estimator $\underline{\hat{X}} = E[\underline{X}|\underline{Y}]$. Notice that for all I the estimator of X_i is the conditional expectation of X_i given $\underline{Y}$: $\hat{X}_i = E[X_i|\underline{Y}]$. That is to say, each of the elements of $\underline{X}$ is estimated separately.

Covariance Matrix of Estimation Error:

$C_{\underline{e}} = E\left[\left(\underline{X} - \underline{\hat{X}}_{MMSE}\right)\left(\underline{X} - \underline{\hat{X}}_{MMSE}\right)^T\right] = C_{\underline{X}} - C_{\underline{\hat{X}}_{MMSE}}$. On diagonal of this matrix appear **error variances** of the estimators $\hat{X}_i$ (namely, MSEs are scalars)

Optimal Linear Estimation in the MMSE Sense: An estimator in the form $\underline{\hat{X}} = A\underline{Y} + \underline{b}$, which minimizes MSE. Estimator is optimal IFF error is $\perp$ $\forall$**linear** functions of the measurements. If measurements vec $\underline{Y}$ has invertible cov matrix, opt linear est: $\underline{\hat{X}}_{BLE} = \underline{\eta}_X + C_{XY}C_Y^{-1}(\underline{Y} - \underline{\eta}_Y)$CovMat of EstError: $C_{\underline{e}} = E\left[\left(\underline{X} - \underline{\hat{X}}_{BLE}\right)\left(\underline{X} - \underline{\hat{X}}_{BLE}\right)^T\right] = C_{\underline{X}} - C_{\underline{XY}}C_{\underline{Y}}^{-1}C_{\underline{YX}}$**Interesting example:** Pass a vector $\underline{X}$ through a noisy linear system, i.e. $\underline{Y} = H\underline{X} + \underline{N}$ where $\underline{X}$ and $\underline{N}$ are uncorrelated random vectors with expectation 0 and covariance matrices $C_{\underline{X}}$ and $C_{\underline{N}}$ respectively. Optimal linear estimator is: $\underline{\hat{X}}_{BLE} = \underline{\eta}_X + C_{XY}C_Y^{-1}(\underline{Y} - \underline{\eta}_Y) = C_{XY}C_Y^{-1}\underline{Y} = C_X H^T(HC_XH^T + C_N)^{-1}\underline{Y}$. **Covariance matrix of the estimator error** will be: $C_{ee}^{MMSE} = C_X - C_XH^T(HC_XH^T + C_N)^{-1}HC_X$,

First Order Marginal Distribution

When time parameter is constant ($t = t_0$), we get a unique RV $X(t_0)$. This RV has a PDF, $f_X(x\,; t_0)$. The expectation and variance can be found, generally be a function of time: $E[X(t)] = \eta_X(t)$; $Var(X(t)) = \sigma_X^2(t)$

Second Order Marginal Distribution:

Let us sample the process at two different times t_1 and t_2, so as to get two random variables $X(t_1).X(t_2)$ with some statistical dependence between them. For any two times, a joint PDF exists: $f_{X_1,X_2}(x_1, x_2\,; t_1, t_2)$. From this function, we may find the **second order statistics** (i.e all the second order moments) of the process. **Auto-Correlation Function:** $R_X(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2 f_{X_1X_2}(x_1, x_2; t_1, t_2)\, dx_1 dx_2$ (also written as $C_{XX}(\cdot), R_{XX}(\cdot)$)

Auto-covariance Function: $C_X(t_1, t_2) = E[(X(t_1) - \eta_X(t_1))(X(t_2) - \eta_X(t_2))] = R_X(t_1, t_2) - \eta_X(t_1)\eta_X(t_2)$

Auto-correlation coefficient funct: $\rho_X(t_1, t_2) = \frac{C_X(t_1,t_2)}{\sqrt{C_X(t_1,t_1)C_X(t_2,t_2)}}$

Marginal Distribution of Order n:

Let us look at the samples of the random process in n different times, $t_1, t_2, \dots, t_n$. These samples are random variables with some interrelationship $X(t_1), \dots, X(t_n)$ defined by the n −dimensioned PDF: $f_{X_1 \dots X_n}(x_1, \dots, x_n\,; t_1 \dots t_n)$. The complete statistical information of the process is defined by joint PDF of any n samples for any value of n.

Gaussian Random Process:

If $X[n]$is Gaussian Random Process then $X[-n]$ is a Gaussian Random Process

Definition: $X(\cdot)$ is a Gaussian random process IFF any vector of its samples is a Gaussian random vector.

Claim: The complete statistics of a Gaussian random process are characterized by the auto-correlation function and the expectation function.

Similarly (to explanation): $[C_{\underline{Y}}]_{l,m} = R_X(n_l, n_m) - \mu_X(n_l)\mu_X(n_m)$

Explanation: Assume $X[n]$ Gaussian random process in discrete Time. It has mean $\mu_X[n]$ and auto-correlation $R_X[l, m]$. Denote a vector of size k of its samples: $\underline{Y} = [X[n_1]\;\; X[n_2] \dots \;\; X[n_k]]^T$. From definition, this is a Gaussian Rand vector. So, to characterize its statistics requires expectation vector $\underline{\mu_Y} = E[\underline{Y}]$ & covariance matrix $C_{\underline{Y}} = E[\underline{YY}^T] - \underline{\mu_Y} \cdot \underline{\mu_Y}^T$. However: $\underline{\mu_Y} = E\left[[X[n_1]\;\; X[n_2] \dots \;\; X[n_k]]^T\right] = [[E[X(n_1)]]\;\; [E[X(n_2)]] \dots [E[X(n_k)]]^T = [\mu_X(n_1)\;\; \mu_X(n_2) \dots \mu_X(n_k)]^T$.

Jointly Gaussian Random Processes: Random processes $\{X_i(\cdot)\}_{i=1,2,\dots,n}$ are called Jointly Gaussian IFF any vector consisting of their samples is a Gaussian Random Vector.

Joint Gaussian (2): Vars made by a linear transform of a Gauss Rand Vec are JGauss

Gaussian Random Process Definition (2): A random process $X(t)$ is Gaussian if for any series of times $t_1 .. t_n$, collection of vars $[X(t_1), \dots, X(t_n)]$ is a Gaussian Rvec

Jointly Gaussian Processes: $X(t), Y(t)$ are JGauss processes if for any series of time t_1 , … , t_n the collection of vars $[X(t_1), \dots, X(t_n), Y(t_1), \dots, Y(t_n)]$ is Gaussian Rvec.

Stationarity

$X(t)$ SSS for any time series $t_1, \dots t_n$ and for any const time shift τ, the Rvec $[X(t_1), \dots, X(t_n)]$ and Rvec $X(t_1 + \tau), \dots, X(t_n + \tau)$ have the same distribution

Strict Sense Stationary (SSS): A random process is SSS if for all values of n, the PDF of order n of the process does not change due to a shifting in time, namely:

Continuous Time X(t) : $f_{X_1,\dots,X_n}(x_1, \dots, x_n\,; t_1, \dots, t_n) = f_{X_1,\dots,X_n}(x_1, \dots, x_n\,; t_1 + \tau, \dots, t_n + \tau)$ $\forall n \in \mathbb{N}$, $\forall t_i, \tau \in \mathbb{R}$, $i \in [1, n]$

Discrete time X[n]: $f_{X_1,\dots,X_n}(x_1, \dots, x_n\,; k_1, \dots, k_n) = f_{X_1,\dots,X_n}(x_1, \dots, x_n\,; k_1 + m, \dots, k_n + m)$ $\forall n \in \mathbb{N}, \forall k_i, m \in \mathbb{Z}$, $i \in [1, n]$

In other words, the PDF is dependent only on the time difference, and invariant to shifting in the time axis

Wide Sense Stationary: An RV is WSS if following 2 conditions are met: expected value does not depend on time: $\eta_X(t) = \eta_X$. the auto-correlation is dependent only on time difference: $R_X(t_1, t_2) = R_X(t_1 - t_2)$. **IID Implies SSS**

Continuous Time: $E[X(t)] = \eta_X$, $R_X(t_1, t_2) = R_X(t_1 - t_2)$

Discrete time: $E[X[k]] = \eta_X$, $R_X[k_1, k_2] = R_X[k_1 - k_2]$

SSS →WSS. WSS does not mean SSS (unless gaussian)

Properties of the auto-correlation function of a real WSS Random Process:

$R_X(0) = E[X^2(t)] \ge 0$ – positive | $R_X(0) \ge |R_X(\tau)|$ | $R_X(\tau) = R_X(-\tau)$ – Even Function | $R_X(\tau)$ – is a real function | $\rho_X(\tau) = \frac{C_X(\tau)}{C_X(0)}$ – correlation coefficient

$\sum_{i=1}^{M}\sum_{j=1}^{N} a_i a_j R_X(t_i, t_j) \ge 0$ – the correlation matrix is positive semi-definite. THIS DOES NOT MEAN THAT THE ELEMENTS OF THE MATRIX ARE NON-NEGATIVE.

Stationarity and Processes Independent of Time and Memoryless

We will prove the following: Let $X(t)$ be a random process and $g(\cdot)$ be arbitrary real function: $g: R \to R$. Define random process $Z(t)$ as: $Z(t) = g(X(t))$ $\forall t$. (It is said that the random process $Z(t)$ is obtained by passing random process X(t) through a time-invariant & memoryless system) **Then:** If $X(t)$ is SSS then necessarily $Z(t)$ is SSS, & If $X(t)$ is WSS then $Z(t)$ is not neccesarily WSS

Generalizations and Highlights: [1]: Generalize for system with memory but time-invariant: If $g: R^m \to R$ and $Z(t) = g(\{X(t-\tau_i\}_{i=1}^m), \forall t$, such that $\{\tau_i\}_{i=1}^m$ is a series of consts not dependent on the time $t, then\ if$ X(t) is SSS then $Z(t)$ is necessarily SSS. [**2**]: **If $g(\cdot)$ is a system independent of time but linear to its parameters, i.e.:** $Z(t) = g(\{X(t-\tau_i\}_{i=1}^m) = \sum_{i=1}^m a_i X(t-\tau_i) + b$, where $\{a_i\}_{i=1}^{i=m}$, b, are consts indep of time. Then if $X(t)$ is WSS then $Z(t)$ is necessarily WSS.

Auto-Regressive (A.R) Random Process:

An AR random process in discrete time $X[n]$ is **defined as**: $X[n] = g(X[n-1], W[n]) \quad \forall n > n_0 .\, g(\cdot)$ is deterministic function, and $W[n]$ is the innovations process, and can be $-\infty$. If n_0 is finite, the initial conditions must also be specified (i.e. the distribution of the sample $X[n_0]$) In order to characterize $X[n]$. $X[n_0]$ is independent of $\{W[n]\}_{n=0}^{\infty}$

Let us solve the following difference equation for later use: $z[n] = az[n-1] + n \;\; n \geq 1,\; a \neq 1$ Where $z[n]$ is some deterministic function with initial condition $z[0]$. Then: $z[n] = az[n-1] + n = a(az[n-2] + b) + b = \cdots = a^n z[0] + b\sum_{i=0}^{n-1} a^i = a^n z[0] + b\frac{1-a^n}{1-a} = a^n\left(z[0] - \frac{b}{1-a}\right) + \frac{b}{1-a}$ Therefore: $\boxed{z[n] = a^n\left(a[0] - \frac{b}{1-a}\right) + \frac{b}{1-a} \quad n \geq 0}$

Summary of Necessary and Sufficient Conditions for WSS

(1): $|a| < 1$ **(2):** $\eta_{X_0} = \frac{\eta_W}{1-a}$ (3): $\sigma_{X_0}^2 = \frac{\sigma_W^2}{1-a^2}$ If (1) holds then guaranteed: $\eta_X[n] \xrightarrow{n\to\infty} \frac{\eta_W}{1-a}$ and $C_X[n, n-k] \xrightarrow{n\to\infty} a^{|k|}\frac{\sigma_W^2}{1-a^2}$ regardless of other 2 conditions. We say the process is asymptotically WSS.

Power Spectral Density of a WSS Random Process: PSD Defined as $\boldsymbol{S_{XX}}(\omega) = \int_{-\infty}^{\infty} R_{XX}(\tau)\, e^{-j\omega\tau} d\tau$

Properties of Power Spectral Density:	Deterministic Function	Real Function by definition	Even function (for real processes)	$\forall\omega\ S_x(\omega) \geq 0$
	Inverse Fourier Transform: $R_{XX}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XX}(\omega)e^{j\omega t} d\omega$. Thus: $E[X^2(t)] = R_{XX}(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XX}(\omega)d\omega$			

If random process has expected value $\neq$ than zero, then PSD has a delta function in the origin, whose height is at least $2\pi(E[X(t)])^2$. PSD is a measure of second order statistics of the random process. Different random processes can have same PSD

Cross-Spectral Density: For X(t) , Y(t), two real JWSS random processes, the cross-spectral density $\boldsymbol{S_{XY}}(\omega)$ is defined as: $R_{XY}(\tau) \overset{FT}{\leftrightarrow} S_{XY}(\omega)$. $S_{XY}(\omega) = \int_{-\infty}^{\infty} R_{XY}(\tau)e^{-j\omega t} d\tau$

Properties of the Cross-Spectrum:	Deterministic Function	Not necessarily Real	**Inverse FT:** $R_{XY}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XY}(\omega)e^{j\omega t} d\omega$	**Property(by definition)**: $S_{YX}(\omega) = S_{XY}^*(\omega) = S_{XY}(-\omega)$
	Symmetric Conjugate Property: $S_{XY}(\omega) = S_{XY}^*(-\omega)$. Proof: $S_{XY}(\omega) = \int_{-\infty}^{\infty} R_{XY}(\tau)e^{-j\omega t}d\tau = \overline{\int_{-\infty}^{\infty}\overline{R_{XY}(\tau)e^{-j\omega t}d\tau}} = \overline{\int_{-\infty}^{\infty} R_{XY}(\tau)e^{-j(-\omega)t}d\tau} = S_{XY}^*(-\omega)$			

Discrete Time Definitions: The definitions for a random process in discrete time are very similar, where the continuous FT is exchanged by the DTFT. If X_n is real and WSS random process in discrete time, then: $R_{XX}[k] \overset{DTFT}{\longleftrightarrow} S_{XX}(e^{j\omega})$; $S_{XX}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} R_{XX}[k]e^{-j\omega k}$ $\quad R_{XX}[k] = \frac{1}{2\pi}\int_0^{2\pi} S_{XX}(e^{j\omega})e^{j\omega k}d\omega$. If X_n , Y_n are real JWSS random processes in discrete time, then: $R_{XY}[k] \overset{DTFT}{\longleftrightarrow} S_{XY}(e^{j\omega})$; $S_{XY}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} R_{XY}[k]e^{-j\omega k}$ $\quad R_{XY}[k] = \frac{1}{2\pi}\int_0^{2\pi} S_{XY}(e^{j\omega})e^{j\omega k}d\omega$

WSS Random Process through an LTI System:

Let $X(t)$ be a WSS process. $X(t)$ is set as the input of an LTI system with impulse response $h(t)$ whose Fourier Transform is $H(\omega)$. In the output of the system is the random process $Y(t)$, that is JWSS with $X(t)$. Furthermore: $\eta_Y = E[Y(t)] = E\left[\int_{-\infty}^{\infty} h(\alpha)X(t-\alpha)d\alpha\right] = \eta_X \int_{-\infty}^{\infty} h(\alpha)d\alpha = \eta_X H(\omega = 0)$

$R_{YX}(\tau) = E[Y(t)X(t-\tau)] = E\left[\left(h(t) \star X(t)\right)X(t-\tau)\right] = E\left[\left(\int_{-\infty}^{\infty} h(\alpha)X(t-\alpha)d\alpha\right) X(t-\tau)\right] = \int_{-\infty}^{\infty} h(\alpha)E[X(t-\alpha)X(t-\tau)]d\alpha = \int_{-\infty}^{\infty} h(\alpha)R_{XX}(\tau-\alpha)d\alpha = h(\tau) \star R_{XX}$

$R_{YY}(\tau) = E[Y(t)Y(t-\tau)] = E\left[\left(\int_{-\infty}^{\infty} h(\alpha)X(t-\alpha)d\alpha\right)Y(t-\tau)\right] = \int_{-\infty}^{\infty} h(\alpha)E[X(t-\alpha)Y(t-\tau)]d\alpha = \int_{-\infty}^{\infty} h(\alpha)R_{XY}(\tau-\alpha)d\alpha = h(\tau) \star R_{XY}(\tau) = h(\tau) \star R_{XX}(\tau)h(-\tau)$

$R_{XY}(\tau) = R_{YX}(-\tau) = \int_{-\infty}^{\infty} h(\alpha)R_{XX}(-\tau-\alpha)d\alpha = R_{XX}(-\tau) \star h(-\tau) = R_{XX}(\tau) \star h(-\tau) \mid\to S_{YZ}(\omega) = H(\omega)S_{XX}(\omega) \quad , S_{XY}(\omega) = S_{XX}(\omega)H^*(\omega),\ \ S_{YY}(\omega) = S_{XX}(\omega)|H(\omega)|^2$

For **discrete time** random processes and systems: $R_{YX}[k] = h[k] \star R_{XX}[k]$; $R_{XY}[k] = R_{YX}[-k] = R_{XX}[k] \star h[-k]$; $R_{YY}[k] = h[k] \star R_{XX}[k] \star h[-k]$
$S_{YX}(e^{j\omega}) = H(e^{j\omega})S_{XX}(e^{j\omega}) \quad , S_{XY}(e^{j\omega}) = S_{XX}(e^{j\omega})H^*(e^{j\omega}), \quad S_{YY}(e^{j\omega}) = S_{XX}(e^{j\omega})\left|H(e^{j\omega})\right|^2$

Y is WSS because $\eta_Y = const.$ And R_{YY} depends on τ. Also, $\boldsymbol{X}\&\boldsymbol{Y}$ **are JWSS**

Power Spectral Density of a Random Process

Consider random process $X(t)$, with known PSD $S_{XX}(\omega)$ which passes through following filter: $H_{\omega_0,\Delta}(\omega) = \begin{cases} 1 \quad \omega \in \left(-\omega_0 - 2\pi\frac{\Delta}{2}, -\omega_0 + 2\pi\frac{\Delta}{2}\right) \cup \left(w_0 - 2\pi\frac{\Delta}{2}, \omega_0 + 2\pi\frac{\Delta}{2}\right) \\ 0\,;\text{else} \end{cases}$ .We can find the power **spectral density at the output** of the filter $X_{\omega_0,\Delta}(t)$ by using: $S_{X_{\omega_0,\Delta}}(\omega) = S_{XX}(\omega)\left|H_{\omega_0,\Delta}(\omega)\right|^2$.The **power of the process at the output** of the filter is: $E\left[X_{\omega_0,\Delta}^2(t)\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{X_{\omega_0,\Delta}}(\omega)e^{j\omega\cdot 0}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XX}(\omega)\left|H_{\omega_0,\Delta}(\omega)\right|^2 d\omega = 2\frac{1}{2\pi}\int_{\omega_0-2\pi\frac{\Delta}{2}}^{\omega_0+2\pi\frac{\Delta}{2}} S_{XX}(\omega)\cdot 1\cdot d\omega$. Now we can perform **approximation** to an integral for $\Delta \to 0$: $\frac{1}{\pi}\int_{\omega_0-2\pi\frac{\Delta}{2}}^{\omega_0+2\pi\frac{\Delta}{2}} S_{XX}(\omega)d\omega \approx \frac{1}{\pi}2\pi\Delta S_{XX}(\omega_0) = 2\Delta S_{XX}(\omega_0)$. **Get the following connection**: $S_{XX}(\omega_0) = \lim_{\Delta\to 0}\frac{E\left[X_{\omega_0,\Delta}^2(t)\right]}{2\Delta}$

White Noise:

Continuous: Random process $X(t)$ is white noise in continuous time IFF it is WSS & satisfies $R_{XX}(\tau) = \sigma^2\delta(\tau)$ **Discrete:** Random process Y_n is wn IFF it is WSS & satisfies $R_{YY}[k] = \sigma^2\delta[k]$

Notice in both cases, every pair of different samples of the random process is uncorrelated. For Gaussian white noise, every pair of different samples of the process are statistically independent.

Ergodicity: In case of ergodic processes, the statistics can be extracted by looking at one sample function at different times.

Strong Ergodicity: $E\left[g\left(X(t_1), \ldots, X(t_n)\right)\right] = \lim_{T\to\infty}\frac{1}{T}\int_0^T g\left(X(t_1+\tau), \ldots, X(t_n+\tau)\right)d\tau\ w.p1\ \forall g(\cdot)$ where $E\left[g\left(X(t_1), \ldots, X(t_n)\right)\right] = \int \ldots \int g(x_1, \ldots, x_n)f_X(x_1, \ldots, x_n\,; t_1, \ldots, t_n)dx_1 \ldots dx_n$

Mean Ergodicity (1): $C_X(\tau) \xrightarrow{\tau\to\infty} = 0$

Slutzky Theorem: A WSS process is EWRTM IFF: $\frac{1}{T}\int_0^T R_{XX}(\tau)d\tau \xrightarrow{T\to\infty} 0$, so if $R_{XY}(\tau) \xrightarrow{\tau\to\infty} 0$ then process is EWRTM

Ergodicity wrt. Mean: $m_T = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} X(t)dt$ (mT =time average) $\quad m_T \xrightarrow{t\to\infty} E[X(t)]$ in MS sense

Mean Erg(2): $\frac{1}{T}\int_{-T}^{T}|C_X(\tau)|d\tau \xrightarrow{T\to\infty} 0$

Overall mean ergodic condition: $m_T \xrightarrow{\text{mean square converg}} E[X(t)] = \eta_X \leftrightarrow E[(m_T - \eta_X)^2] \xrightarrow{T\to\infty} 0$.,If true, process EWRTM

Ergodic wrt Mean if: $\lim_{T\to\infty} m_T = E[X(t)]$. $E[X(t)]$ must be indep time

Mean Ergodicity (3): $\int_0^{\infty}|C_X(\tau)|d\tau < \infty$

Joint Stationary Random Proceses

JSSS: Random processes $X(t), Y(t)$ are JSSS if for any time series $t_1, \ldots, t_n$ and for any constant time shift τ, the Rvec $[X(t_1, \ldots, X(t_n), Y(t_1), \ldots, Y(t_n)]$ and the RVec $[X(t_1+\tau), \ldots, X(t_n+\tau), Y(t_1+\tau), \ldots, Y(t_n+\tau)]$ have the same distribution.

JWSS: Random Processes $X(t), Y(t)$ are JWSS if holds: $X(t)$ WSS, $Y(t)$ WSS, Cross Correlation depends only on time difference: $R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] = R_{XY}(t_1 - t_2)$

If $X(t), Y(t)$ both stat.indep SSS then they are JSSS

If $X(t), Y(t)$ both stat. indep WSS then they are JWSS

If $X(t), Y(t)$ are JSSS then they are also JWSS

If $X(t), Y(t)$ are JGaussian JWSS then they are also JSSS

Optimal Linear Estimation (In the case of two JWSS Processes) - Wiener Filter

Given two JWSS Rprocesses $X(t), Y(t)$ w. expectations zero, calculate optimal linear MMSE estimator of $X(t)$from the samples $\{Y(s), s \in \mathcal{R}\}$: According to MSE criterion this is LTI: $\hat{X}(t) = h(t) \star Y(t)$. The error of the optimal estimator is orthogonal to any linear func of measurments: $e(t) = X(t) - \hat{X}(t)$, $E[e(t)Y(t-\tau)] = 0\ \forall\tau \to R_{eY}(\tau) = R_{XY}(\tau) - R_{\hat{X}Y}(\tau) = 0\ \forall\tau$. Applying Fourier transform obtain: $S_{eY}(\omega) = S_{XY}(\omega) - S_{\hat{X}Y}(\omega) = 0\ \forall\omega \to S_{XY}(\omega) = S_{\hat{X}Y}(\omega) \to H(\omega)S_Y(\omega) = S_{XY}(\omega)$. Conclusion: for any frequency s.t. $S_Y(\omega) > 0$ applies, the frequency response of the **optimal MSE filter** is: $H(\omega) = \frac{S_{XY}(\omega)}{S_Y(\omega)}$. *OLD VERSION OF THE FORMER*: $\hat{X}(t) = \int_{-\infty}^{\infty} h(t,s)Y(s)ds$. **solution**: $\hat{X}_{LMMSE}(t) = h_{Wiener}(t) \star Y(t)$, where $h_{Wiener}(t)$ **is given by:** $H_{WIener}(\omega) = \frac{S_{XY}(\omega)}{S_{YY}(\omega)}$. **Estimation error** process of optimal linear estimator $e(t) = X(t) - \hat{X}_{LMMSE}(t)$ is orthogonal to all linear functions of samples: $E[e(t')g(t) \star Y(t)] = 0\ \ \forall t, t'$ for any $g(t)$. **MSE:** $E[e^2] = \frac{1}{2\pi}\int \left[S_{XX}(\omega) - \frac{S_{XY}^2(\omega)}{S_{YY}(\omega)}\right]d\omega$

Autocorrelation funct of estimation error: $R_{ee}(\tau) = E[e(t+\tau)e(t)] = E\left[e(t+\tau)\left(X(t) - \hat{X}_{LMMSE}(t)\right)\right] =_1 E[e(t+\tau)X(t)] = E\left[\left(X(t+\tau) - \hat{X}_{LMMSE}(t+\tau)\right)X(t)\right] = R_{XX}(\tau) - E\left[\hat{X}_{LMMSE}(t+\tau)X(t)\right] = R_{XX}(\tau) - E\left[\hat{X}_{LMMSE}(t+\tau)\left(e(t) + \hat{X}_{LMMSE}(t)\right)\right] =_2 R_{XX}(\tau) - R_{XX}(\tau)$. (1), (2) follow since estimation error orthogonal to all linear functions of samples, hence also to $\hat{X}_{LMMSE}(t)$

Spectrum of the Estimation Error: $S_e(\omega) = S_X(\omega) - S_{\hat{X}}(\omega) = S_X(\omega) - |H_{Wiener}(\omega)|^2 S_Y(\omega) = \begin{cases} S_X(\omega) - \frac{|S_{XY}(\omega)|^2}{S_Y(\omega)}\,; S_Y(\omega) > 0 \\ S_X(\omega); S_Y(\omega) = 0 \end{cases}$

MSE: $E[e^2(t)] = R_e(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_e(\omega)d\omega$

where $X(t)$ and $N(t)$ are uncorrelated random processes having zero expectation, then the optimal

Special Case – Additive Orthogonal Noise: If $Y(t) = X(t) + N(t)$ filter is: $H(\omega) = \frac{S_{XX}(\omega)}{S_{XX}(\omega) + S_{NN}(\omega)}$. Similar results are true for discrete time.

Autocorrelation Ergodicity: Given SSS Rprocess $X(t)$ want to estimate its autocorrelation: $R_X(\tau) \triangleq E[X(t+\tau)X(t)]\forall\tau$. Notice that estimation of autocorrelation of $X(t)$ at point τ *is* equiv to estimation of the expectation of the group of Rprocesses $Z_\tau(t) \overset{\text{def}}{=} X(t+\tau)X(t)$. I.e, it holds that Autocorrelation ergodicity of $X(t) \leftrightarrow$Mean Ergodicity of $Z_{\tau(t)}\forall\tau$. NOTE: TO check Z mean-ergodic by Slutsky, necessary to require it to be WSS. $X(t)$ mus be WSS and also the 4th moment must depend on time difference only. To have autocorellation ergodicity of variable $Y(t) = \begin{cases} X_1(t); A = 1 \\ X_2(t); A = -1 \end{cases}$ we must satisfy $\frac{1}{2}[R_{X1}^2(\tau) + R_{X2}^2(\tau)] = \left[\frac{R_{X1}(\tau) + R_{X2}(\tau)}{2}\right]^2$. This holds IFF $\boldsymbol{R_{X1}(\tau) = R_{X2}(\tau)}$

Wiener Random Process

Increments of a Random Process: For some random process $X(t)$ define the increments process in the range $(t_1, t_2]$ as follows: $X(t_1, t_2) \overset{\text{def}}{=} X(t_2) - X(t_1)$

Random Process with Independent Increments: It is said that a random process $X(t)$ has independent increments IFF the increments vector $\left[\ X\left(t_1^{(s)}, t_1^{(f)}\right) \quad X\left(t_2^{(s)}, t_3^{(f)}\right) \quad \ldots \quad X\left(t_k^{(s)}, t_k^{(f)}\right)\ \right]$ is a random vector whose samples are independent for all $\left\{t_i^{(s)}, t_i^{(f)}\,|t_i^{(s)} \leq t_i^{(f)} \quad 1 \leq i \leq k\right\}$ such that $\left(t_i^{(s)}, t_i^{(f)}\right] \cap \left(t_j^{(s)}, t_j^{(f)}\right] = \phi \quad 1 \leq i < j \leq k$, and $\forall k > 1$

Random Walk:$X_n = X_{n-1} + W_n$; $X_0 = 0\ \ w.p\ 1$;$W_n = \begin{cases} 1\ w.p. 0.5 \\ -1\ w.p\ 0.5 \end{cases}$, iid and independent of X_0.$E[X_n] = 0$; $Var[X_n] = E[X_n^2] = n$; $C_X[n,m]$=$E[X_n \cdot X_m] =^{n>m} E[(X_m + \sum_{k=m+1}^n W_k)X_m] = E[X_m \cdot X_m] = m$. We assumed that $n > m$ and used the fact that X_m is independent of W_k for $k > m$. For general m, n we get $C_X[n,m] = \min\{n, m\}$

Defining Wiener Process (Limit of Random Walk): Process $X_{T,d}(t)$ is def as $X_{T,d}(t) = 0 \ \ t < 0$ (assumed $X_{T,d}(t) = 0 \ \ t < 0$) . d is size of the step, and $T > 0$ is time it takes per step. **Holds**: $E\left[X_{T,d}(t)\right] = 0$, $Var\left[X_{T,d}(t)\right] = d^2\left\lfloor\frac{t}{T}\right\rfloor \ \ t > 0$. Require that $d\char`^2 = \alpha T$ for some const $\alpha > 0$, & look at small T's for which $N = \left\lfloor\frac{t}{T}\right\rfloor \approx \frac{t}{T}$: $X_{T,d}(t) = \sqrt{\alpha T}\sum_{k=1}^N W_k = \sqrt{\alpha t}\underbrace{\frac{1}{\sqrt{N}}\sum_{k=1}^N W_k}_{\text{Normalized sum}}$.**Define** Wiener process $X(t)$ as: $X(t) = \lim_{T\to 0} X_{T,d}(t) \sim N(0, \alpha t)$ (by central limit theorem)

Properties:

Wiener is an independent increments process	**Expectation, Autocorrelation and Autocovariance:** $\eta_X(t) = 0$, $R_X(t_1, t_2) = C_X(t_1, t_2) = \alpha \cdot \min\{t_1, t_2\} \quad t_1, t_2 t > 0$
Wiener Process is a Gaussian Random Process.	**Distribution of the increments of the process:** $X(t_1, t_2) \sim N\left(0, \alpha(t_2 - t_1)\right) \quad 0 \leq t_1 < t_2$

If $X[n]$ is i.i.d process, then is no correlation between samples at diff. times. Thus $X[n]$ is a white noise w. $R_X[k] = \sigma_X^2\delta[k]$

Derivative Process of Wiener: $X'(t) = \frac{dX(t)}{dt} = \lim_{\epsilon\to 0} X_\epsilon(t)$, $\eta_X(t) = 0$, $C_X(t_1, t_2) = \alpha\delta(t_1 - t_2)$. approximation $\epsilon > 0$ to derivative: $X_e(t) = \frac{X(t+\epsilon) - X(t)}{\epsilon} = \frac{X(t, t+\epsilon)}{\epsilon}$, $\eta_{X_{\epsilon(t)}} = 0, C_{X_\epsilon}(t_1, t_2) = \begin{cases} 0\ \ if\ |t_1 - t_2| \geq \epsilon \\ \frac{\alpha}{\epsilon}\cdot\left(1 - \frac{|t_1 - t_2|}{\epsilon}\right)\ \ if\ |t_1 - t_2| < \epsilon \end{cases}$

Markov Chains

Markovian Process: A process in discrete time $\{X_n\}$ that satisfies: $P(X_{n+1} = j\,|X_n = i, X_{n-1} = i_1, \ldots X_{n-k} = i_k) = P(X_{n+1} = j \mid X_n = i)$. I.e, given current value, future doesn'n depend on past.

Homogenous Markovian Process: Homogenous if $\forall i, j, n$ it holds: $P(X(n+1) = j \mid X(n) = i) = P(X(n) = j \mid X(n-1) = i) \equiv p_{ij}$. I.e transition probability of the chain is independent of time.

Probabilities Vector: If set of all possible states is a finite group $\{1, 2, \ldots, J\}$ then we can define the probabilities vector of the chain at time n: $\pi^{(n)} = [P(X_n = 1), P(X_n = 2), \ldots, P(X_n = J)]$

Transition Property of the Chain: $P(X_{n+1}=j|X_n=i)$

Probability of Passing from state i to j: p_{ij}

Stochastic Matrix: transition matrix is stochastic, satisfies $\sum_{j=1}^{J} p_{ij}=1\ \forall i$

Transition Matrix: If chain is homogenous, then we define the transition matrix: $\boldsymbol{P}=\begin{bmatrix} p_{11} & \cdots & p_{1J} \\ \vdots & p_{ij} & \vdots \\ p_{J1} & \cdots & p_{JJ}\end{bmatrix}$. For a non-homog process the matrix is dependent on time. Each row i of the matrix represents prob of passing from state i to any of the other states.

The following discussion pertains only to Homogenous Markov Chains with a finite number of states, unless stated otherwise:

Based on the Law of Total Probability: $P(X_{n+1}=j)=\sum_{i=1}^{J}P(X_{n+1}=j\,|X_n=i)\cdot P(X_n=i)$. This is to say, the probabilities vector satisfies: $\underline{\pi}^{(n)}=\underline{\pi}^{(n-1)}\boldsymbol{P}=\underline{\pi}^{(n-2)}\boldsymbol{P}^2=\cdots=\underline{\pi}^{(n-m)}\boldsymbol{P}^m=\underline{\pi}^{(0)}\boldsymbol{P}^n$, where $\underline{\pi}^{(0)}$ is the initial probabilities vector and $\boldsymbol{P}$ is the transition matrix.

Chapman-Kolmogorov Formula: For all Markovian Processes with discrete samples space (not necessarily homogenous), for any three times $n_1<n_2<n_3$: $P\left(X_{n_3}=j\mid X_{n_1}=i\right)=\sum_k P\left(X_{n_3}=j\mid X_{n_2}=k\right)P\left(X_{n_2}=k\mid X_{n_1}=i\right)$.

Conclusion from Chapman-Kolmogorov: Element ij of matrix $\boldsymbol{P}^m$ is $=P(X_n=j\mid X_{n-m}=i)=\{\boldsymbol{P}^m\}_{ij}$

Characterization of states

Accessible: For $i\neq j$ state j is accessible from state i if $\exists$ a probability of getting from state i to state j. Symbolized as $i\to j$

Recurrent state: i recurrent if comm-s w. all states accessible by it (if starting in this state, it is guaranteed to return to it).

Transient state: A state that is not recurrent is called transient. In other words, state i is transient IFF $\exists\ i\neq j$ $s.t.\ i\to j$ but not $j\to i$. Namely a state from which we can get to another state but cannot return to it.

Period of a state: Period of a state is the greatest common divisor of length of all the possible paths that start and end in the same state

Communicating states: States i,j communicate if $i>j$ and $j\to i$. Symbolized as $i\leftrightarrow j$.

Classes

Class: Group of states that communicate with one another and do not communicate with other states.

Transient class: In a class that has transient state, it is guaranteed that all the states in it will be transient and it is called a transient class.

Irreducible Markov Chain: Chain with only one class. This class must be recurrent

Recurrent class: In class with a recurrent state, its guaranteed that all states in it will be recurrent. In all Markov chains with finite number of states, there is at least one rec. class.

Period of a Class: All the states in the same class have the same period. This period is called the period of the class. A class with period 1 is aperiodic.

Stationary Distribution: The distribution $\underline{\pi}_s$ is stationary is $\underline{\pi}_s=\underline{\pi}_s\cdot\boldsymbol{P}$. A chain can have a number of different stationary distributions. **Property 1:** If π is a stationary distribution and i is transient state, then $\pi_i=0$ **Property 2:** $\lim_{n\to\infty}\underline{\pi}^{(n)}$ does not necessarily exist, and it can be depend on initial condition $\underline{\pi}^{(0)}$. If limit exists for some $\underline{\pi}^{(0)}$, then it is equal to some stationary distribution $\underline{\pi}_s$.

Perron-Frobenius Theorem: the following claims are a special case of the Perron-Frobenius theorem ,and are only true for a Markov chain with a finite number of states:

$\forall$ Markov chains,$\exists$ one stationary distribution at least. I.e there always exists a "legal" solution (whose elements are non-negative and sum up to 1) to the equation $\underline{\pi}=\underline{\pi}\cdot\boldsymbol{P}$.

If chain has only 1 recurrent class, then there is only 1 solution to the equation above.

If chain has r different rec-nt classes, then $\exists\ r$ independent stationary distributions.

Inverting a Matrix: Write **A|I** I is identity matrix. Perform row operations on A until A is identity matrix. Repeat those same operations on **I**. At the end of this process, **I** will become the inverse matrix of **A**

If chain has only 1 recurrent class, which is aperiodic, then$\forall\ i,j$: $\lim_{n\to\infty}[P^n]_{ij}=\lim_{n\to\infty}\pi_j^{(n)}=\pi_j$, where π is stationary distribution. I.e, matrix P^n approaches the following: $P^n\xrightarrow{n\to\infty}\begin{bmatrix}\pi_1 & \cdots & \pi_J\\ \vdots & \ddots & \vdots\\ \pi_1 & \cdots & \pi_J\end{bmatrix}$. This is true for any initial condition $\pi^{(0)}$. I.e, probability of being in state j approaches π_j independently of initial state. In this case we say that the chain is **ergodic** and its **memory vanish** (with n)

If a chain contains r classes, all of which recurrent: After "rearranging" them (re-indexing the states, s.t states of each recurrent class are consecutive), $\boldsymbol{P}$ will be block diagonal: $\boldsymbol{P}=\begin{bmatrix}A_1 & & 0\\ & A_2 & \\ 0 & & \ddots\end{bmatrix}$. r independent stationary solutions, e.g: $\underline{\pi}=[\underline{\pi}_1 \vdots 0 \vdots\cdots\vdots 0],[0 \vdots \underline{\pi}_2 \vdots 0\vdots\cdots\vdots 0],[0\vdots 0\vdots \underline{\pi}_3\vdots\cdots\vdots 0],\ldots,[0\vdots\cdots\vdots 0\vdots \underline{\pi}_r]$. $\underline{\pi}_i$ is the only sol of $\underline{\pi}_i=\underline{\pi}_i\cdot A_i$. Any linear comb. of the sols is also a stationary sol, provided $\forall$ coefficients are >0 & their sum = 1 (since this is a probabilities vector)

If a chain contains r classes, all of which recurrent & aperiodic: Matrix P^n: $\boldsymbol{P}^\infty=\lim_{n\to\infty}\boldsymbol{P}^n=\begin{bmatrix}\Pi_1 & 0 & \ldots & 0\\ 0 & \Pi_2 & \ldots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \Pi_r\end{bmatrix}$, Π_i square matrices of form $\Pi_i=\begin{bmatrix}\underline{\pi}_i\\ \vdots\\ \underline{\pi}_i\end{bmatrix}$

If only one recurrent class exists, but it is periodic with a period $d>1$: Then $\boldsymbol{P}^n$ **does not converge, but** $P^{n\cdot d}=(P^d)^n$ does converge. There is convergence $\forall$ d−th moment: $\{0,d,2d,\ldots,nd,\ldots\},\{1,1+d,1+2d,\ldots,1+nd,..\}\ldots$

Inverse Markov Chain: stationary distro of inv.chain is same as of original chain. $\boldsymbol{P}_{ij}^{\text{inv}}=\frac{\boldsymbol{P}_{ji}\pi_j}{\pi_i}$ (only true if already converging to a stationary distro?)

Combined Stationary Distro: If you have π_1 ,π_2, then $\pi_{\text{general}}=\alpha\pi_1+(1-\alpha)\pi_2$

Jacobian Stuff

Problem: X,Y be RVs w. PDF: $f_{XY}(x,y)$. Def $V=h(X,Y),W=g(X,Y)$, find PDF of V & W? Solution (assume countable sols for eq-n set): For each set of vals (v,w) solve set of equations: $v=h(x,y),w=g(x,y)$. Denote sols as: $\{(x_i,y_i)\}_{i=1}^{n}$

Jacobian Transformation: $\frac{\partial(v,w)}{\partial(x,y)}=\det\begin{bmatrix}\frac{dv}{dx} & \frac{dv}{dy}\\ \frac{dw}{dx} & \frac{dw}{dy}\end{bmatrix}$, $\frac{\partial(x,y)}{\partial(v,w)}=\det\begin{bmatrix}\frac{dx}{dv} & \frac{dx}{dw}\\ \frac{dy}{dv} & \frac{dy}{dw}\end{bmatrix}$

Joint PDF of V and W: $f_{VW}(v,w)={}_1\sum_{i=1}^{n}f_{XY}(x_i,y_i)\left|\frac{\partial(x,y)}{\partial(v,w)}\right|_{x=x_i,y=y_i}={}_2\sum_{i=1}^{n}\frac{f_{XY}(x_i,y_i)}{\left|\frac{\partial(v,w)}{\partial(x,y)}\right|_{x=x_i,y=y_i}}$. In (1) Jacobian is a funct of v,w, and in (2) it's a func of x,y.

Autocovariance of a Rprocess: $C_X(t_1,t_2)=Cov\big(X(t_1),X(t_2)\big)=R_X(t_1,t_2)-\eta_X(t_1)\eta_X(t_2)$

Poisson Distribution

Sum of stat.indep Poisson RVs: Is Poisson as well. $P_r(N_n=m)=\frac{e^{-QT}(QT)^m}{m!}$ $Q=\sum_{k=1}^{n}\lambda_k$

If PoiP is w. stat. indep. Increments then distr of $N(t_3)-N(t_2)$ given $N(t_1),N(t_2)$ is Poisson w. param. $\lambda(t_3-t_2)$

Point Process: Discrete events in continuous time. In some tyupes the events tend to occur in clusters. $X(t)=\sum_j\delta(t-t_j)$

Counting Process; $N(t)=\sum_j n(t-t_j)$

We can denote $N(t)$ as a posson variable: $N\sim P,\lambda T$

Mean of RProcess: $\eta_X=E[X(t)]$

$P(n)=e^{-\lambda T}\frac{(\lambda T)^n}{n!}$, def $\lambda T=Q\to P(n)=e^{-!}\frac{Q^n}{n!}$, $E[n]=Q\ \&\ V(n)=Q$

$E[n^2]=e^{-Q}Q^2\sum_{n=2}^{\infty}\frac{Q^{n-2}}{(n-2)!}+Q$

Characteristic Function of Poisson Variable: $\phi_n(\omega)=e^{Q(e^{i\omega}-1)}$

Stirling's Formula: $n!\sim\sqrt{2\pi n}e^{-n}n^n$

Poisson Process: Events happen randomly and independently of each other. $P[n]=P[n\text{ events in }[0,t]]=\frac{m!}{n!(m-n)!}\left(\frac{Q}{m}\right)^n\left(1-\frac{Q}{m}\right)^{m-n}$. As $m\to\infty$: $P(n)=\frac{Q^n}{n!}e^{-Q}$

Rate Parameter: (events(avg)/time): $\lambda=\frac{Q}{t}$

$E[N(t_1)|N(t_2)]=\frac{t_1}{t_2}N(t_2)$

$P(n,t)=P_{N(t)}(n)=P\,[N(t)=n]=\frac{e^{-\lambda t}(\lambda t)^n}{n!}$. N(t) is the counting Poisson Process [count of events occurring in a pois-point process between times 0 and t]

Autocorrelation: $R_{XX}(t_1,t_2)=\lambda^2t_1t_2+\lambda\min(t_1,t_2)$

Autocorelation of Rprocess: $\boldsymbol{R_X}(t_1,t_2)=E[X(t_1)X(t_2)]$]

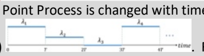
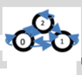
Exponential Distribution | $f_{\Delta t}(\tau)=\lambda e^{-\lambda\tau}u(\tau)$ | $E[\Delta t)=1/\lambda$ | $V(\Delta t)=1/\lambda^2$ | SNR = 1 | Exponential distribution is memoryless | **CDF:** $F(\tau_2)=1-e^{-\lambda\tau_2}$ | **PDF:** $f(\tau_2)=\lambda e^{-\lambda e_2}$

Erlang Distribution: | **Two parameters**: k −shape, λ −rate | **PDF:** $f_X=\frac{\lambda^kx^{k-1}e^{-\lambda x}}{(k-1)!}$ | **CDF:** $P(k,\lambda x)=\frac{\gamma(k,\lambda x)}{(k-1)!}=1-\sum_{n=0}^{k-1}\frac{1}{n}$ | **Mean:** $\frac{k}{\lambda}$ | **Variance:** $\frac{k}{\lambda^2}$

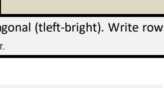
MATRIX TRANSPOSE: Reflect A over its main diagonal (tleft-bright). Write rows of A as colums of A_T. Write columns of A as rows of A_T.

$X_n=a\cdot X_{n-1}+W_n$ (autoregressive process) ; $z[n]=az[n-1]+b=a(az[n-2]+b)+b=\cdots=a^nz[0]+b\sum_{i=0}^{n-1}a^i=a^nz[0]+b\frac{1-a^n}{1-a}=a^n\left(z[0]-\frac{b}{1-a}\right)+\frac{b}{1-a}$, hence $z[n]=a^n\left(z[0]-\frac{b}{1-a}\right)+\frac{b}{1-a},n\geq 0$

P1: Measurement of variable X is obscured by multiplicative noise n s.t measured quantity is $Y=Xn$. $Var(X)=\sigma_X^2,E[X]=0,Var(n)=\sigma_n^2,E[n]=\eta_n\neq 0$. X,n stat.indep. **A)** Variable $\hat{X}_1=Y/\eta_n$ proposed as estimator for X. Is it biased? What is the MSE? **ANS:** $E\left[X-\widehat{X_1}\right]=E[X]-E\left[\frac{Y}{\eta_n}\right]=E[X]-\frac{1}{\eta_n}E[X]E[n]=0$, so unbiased. MSE: $MSE\left(\widehat{X_1}\right)=E\left[\left(X-\widehat{X_1}\right)^2\right]=E[X^2]-2E\left[X\widehat{X_1}\right]+E\left[\widehat{X_1}^2\right]$. $E\left[X\widehat{X_1}\right]=E\left[X\frac{Y}{\eta_n}\right]=E\left[X\frac{Xn}{\eta_n}\right]=\frac{1}{\eta_n}E[X^2]E[n]=E[X^2]=\sigma_X^2$. $E\left[\widehat{X_1}^2\right]=E\left[\frac{Y^2}{\eta_n^2}\right]=\frac{1}{\eta_n^2}E[X^2]E[n^2]=\frac{\sigma_X^2(\sigma_n^2+\eta_n^2)}{\eta_n^2}$. $MSE\left(\widehat{X_1}\right)=\frac{\sigma_X^2(\sigma_n^2+\eta_n^2)}{\eta_n^2}-\sigma_X^2=\sigma_X^2\left(\frac{\sigma_n^2+\eta_n^2}{\eta_n^2}-1\right)=\frac{\sigma_X^2\sigma_n^2}{\eta_n^2}$. **B)** Hoping to improve est, propose $\widehat{X_2}=\alpha\frac{Y}{\eta_n}$. Find optimal α w.r.t MSE, determine if est. is biased, is it better that $\widehat{X_1}$? **ANS:** $MSE\left(\widehat{X_2}\right)=E[X^2]-2\alpha E[X^2]+\alpha^2E[X^2]\frac{E[n^2]}{\eta_n^2}=\sigma_X^2\left(1-2\alpha+\alpha^2\frac{\sigma_n^2+\eta_n^2}{\eta_n^2}\right)$. This is quadratic in α. Has unique minimum. Find by taking derivative w.r.t α, solving =0. $\frac{\partial MSE(\widehat{X_2})}{\partial\alpha}=\sigma_X^2\left(-2+2\alpha\frac{\sigma_n^2+\eta_n^2}{\eta_n^2}\right)=0\to\alpha^*=\frac{\eta_n^2}{\sigma_n^2+\eta_n^2}$. Proving unbiased: $E\left[X-\hat{X}_2\right]=E[X]-E\left[\alpha\frac{Y}{\eta_n}\right]=E[X]-E\left[\alpha\frac{Xn}{\eta_n}\right]=E[X]-\frac{\alpha}{\eta_n}E[X]E[n]=E[X](1-\alpha)=0$, so unbiased $\forall\alpha$. $\widehat{X_2}$ better than $\widehat{X_1}$ or else α would have been 1.**C)** Now assume X,n are Gaussian. Is the linear estimator from (b) optimal? **ANS:** Lin comb. Of stat. indep normal distr. RVs is a normally distributed RV, so if measurement was $Y=X+n$, then would have been Gauss and $\widehat{X_2}=\hat{X}_{BLE}=\hat{X}_{MMSE}$. But, since $Y=Xn$, X,Y are not JointGauss, thus $\hat{X}_{BLE}\neq\hat{X}_{MMSE}$.**D)** Given that X is Gaussian, and n is a binary variable equal to 0,1 with equal probabilities. Find opt. est. of X given Y. **ANS:** If Y=0 there are 2 options: (1) $n=0$, i.e Y doesn't give any info about X. Hence $\hat{X}_{MMSE}=E[X|Y]=E[X]=0$. (2) $X=0$, thus $\hat{X}_{MMSE}=Y=X=0$. If $Y\neq 0$ then $n\neq 0$, i.e $n=1$, hence $Y=X$. Thus $\hat{X}_{MMSE}=Y=X$

P2: Rate param $\lambda(t)$ of Poisson Point Process is changed with time in a piecewise constant manner over intervals of duration T. This means that rate is equal to $\lambda(t)=\lambda_1$ for $t\in[0,T],\lambda(t)=t_2$ for $t\in[T,2T]$, and so on, s.t the n_{th} value of λ_n prevails for $t\in[(n-1)T,nT]]$ (see fig) . Reminder: ProbDist of Poisson Variable N with param Q is: $\begin{cases}\frac{e^{-Q}Q^m}{m!}\ ;Q>0\\ \delta_{m,0}\ ;Q=0\end{cases}$ where $\delta_{m,0}=\begin{cases}1\ ;m=0\\ 0\ ;\ m\neq 0\end{cases}$. **A)** Assume that $\lambda_k=k\ mod\ 3$, i.e $\lambda_1=1,\lambda_2=2,\lambda_3=0,\lambda_4=1\ldots$ Find distribution of N_n – number of events accumulated until time nT (namely, find probability $P_r(N_n=m)$) for all $n\geq 1$. **ANS:** from class, sum of stat.indep. Poisson RVs is Poisson distributed. Hence: $P_r(N_n=m)=\frac{e^{QT}(QT)^m}{m!}$, $Q=\sum_{k=1}^{n}\lambda_k$. Let us divide the answer for different values of n: For $n\leq 3$: $Q=\begin{cases}1\ ;n=1\\ 3\ ;n=2,3\end{cases}$. For $n>3$: $Q=\begin{cases}1+\frac{n-1}{3}+2\frac{n-1}{3}=n\ ;n\ mod\ 3=1\\ 1+2+\frac{n-2}{3}+2\frac{n-2}{3}=n+1\ ;n\ mod\ 3=2\\ \frac{n}{3}+2\frac{n}{3}=n\ ;n\ mod\ 3=0\end{cases}$. Overall we got: $Q=\begin{cases}n+1\ ;n\ mod\ 3=2\\ n\ ;else\end{cases}$. **B)** Assume that λ_k is a homogenous Markov chain whose state diagram is given as:  .Write into the diagram the transition probabilities that ensure that λ_k is an IID process, s.t $\lambda_k=0$ w.p 0.1, $\lambda_k=1$ w.p 0.7, and $\lambda_k=2$ w.p 0.2. **ANS:** λ_k would be IID if its prob.distr. does not change in time. Denote $\boldsymbol{P}$ as the transition matrix of the process, and $\underline{\pi_k}$ is the probability distr. Of λ_k at time k. In these terms we want $\underline{\pi_k}=[0.1\ \ 0.7\ \ 0.2]\ \forall k$. Hence we need $\underline{\pi_k}\boldsymbol{P}=\underline{\pi_k}$. From class, this is the case when $\boldsymbol{P}=\begin{bmatrix}\underline{\pi_k}\\ \underline{\pi_k}\\ \underline{\pi_k}\end{bmatrix}=\begin{bmatrix}0.1 & 0.7 & 0.2\\ 0.1 & 0.7 & 0.2\\ 0.1 & 0.7 & 0.2\end{bmatrix}$. **C)** Under the conditions of (b), find the distribution of N_1 and N_2. **ANS:** $P_r(N_1=m)=\sum_{k=0}^{2}P_r(N_1=m\,|\lambda_1=k)\,P_r(\lambda_1=k)=\begin{cases}0.7\frac{e^{-T}T^m}{m!}+0.2\frac{e^{-2T}(2T)^m}{m!}\ ;\ m>0\\ 0.1+0.7e^{-T}+0.2e^{-2T}\ ;m=0\end{cases}$. As in (a), sum of stat. indep. Poisson RVs is Poisson distr-d. Thus: $P_r(N_2=m)=\sum_{k,j=0}^{2}P_r(N_2=m\,|\lambda_1=k,\lambda_2=j)P_r(\lambda_1=k,\lambda_2=j)$. Divide into cases. For $m=0$: $P_r(N_2=0)=0.1^2+0.1\times 0.7e^{-T}+0.1\times 0.2e^{-2T}+0.7\times 0.1e^{-T}+0.7\times 0.7e^{-2T}+0.7\times 0.2e^{-3T}+0.2\times 0.1\,e^{-2T}+0.2\times 0.7e^{-3T}+0.2\times 0.2e^{-4T}$. For $m>0$: $P_r(N_2=m)=0.1\times 0.7\frac{e^{-T}T^m}{m!}+0.1\times 0.2\frac{e^{-2T}(2T)^m}{m!}+0.7\times 0.1\frac{e^{-T}T^m}{m!}+0.7\times 0.7\frac{e^{-2T}(2T)^m}{m!}+0.7\times 0.2\frac{e^{-3T}(3T)^m}{m!}+0.2\times 0.1\frac{e^{-2T}(2T)^m}{m!}+0.2\times 0.7\frac{e^{-3T}(3T)^m}{m!}+0.2\times 0.2\frac{e^{-4T}(4T)^m}{m!}$. **D)** In the following part, λ_n can receive one of two values, 0 or 1. It is given that $\lambda_1=1$, and that $\lambda_n=1$ when $N_{n-1}=N_{n-2}$ and $\lambda_n=1-\lambda_{n-1}$ when $N_{n-1}>N_{n-2}$. Prove that λ_n is a Markov process. Plot its transition diagram, determine if it has a limit distribution, and if so, find it. **ANS:** First, notice that $P_r(N_{n-1}=N_{n-2})=P_r(N_{n-1}-N_{n-2}=0)$. Second, since $N_{n-1}\geq N_{n-2}$ holds (N_n is a Poisson process), $P_r(N_{n-1}>N_{n-2})=1-P_r(N_{n-1}-N_{n-2}=0)$. We know that $P_r(\lambda_n=1)=P_r(N_{n-1}-N_{n-2}=0|\lambda_{n-1}=1)P_r(\lambda_{n-1}=1)+P_r(N_{n-1}-N_{n-2}\geq 0|\lambda_{n-1}=0)P_r(\lambda_{n-1}=0)=e^{-T}P_r(\lambda_{n-1}=1)+P_r(\lambda_{n-1}=0)$. This means that the value ofλ_n is a Markpv process. State diagram: . Transition Matrix: $\boldsymbol{P}=\begin{bmatrix}e^{-T} & 1-e^{-T}\\ 1 & 0\end{bmatrix}$. Clearly this process has 1 recurrent class which is a-periodic. Hence $\exists$a limit distr. Finding it: $\underline{\pi_s}\boldsymbol{P}=[\pi_1\ \ \pi_0]\begin{bmatrix}e^{-T} & 1-e^{-T}\\ 1 & 0\end{bmatrix}=[\pi_1\ \ \pi_0]=\underline{\pi_s}\ s.t.\pi_1+\pi_0=1$. Solving this yields: $\underline{\pi_s}=\left[\frac{1}{2-e^{-T}}\ \frac{1-e^{-T}}{2-e^{-T}}\right]$.

P3: A) The process $x(t)$ is to be linearly estimated from another process $y(t)$. Prove that $\hat{x}(t)$ is the best such estimator in the MSE sense if it is unbiased and satisfies the orthogonality condition $E\big[(x(t)-\hat{x}(t))y(t')\big]=0\ \ \forall t,t'$. Does this relation require $x(t)$ WSS? You can assume means of $x(t),y(t)$ are 0. **ANS:** Proof for orthogonality was given in class. Further if we assume that $x(t),y(t)$ are JWSS (which implies $x(t)$ WSS), then the LMMSE Estimator is the Wiener Filter, but orthogonality still holds. **B)** Assume the given setup, where $x(t)$ is a zero-mean WSS signal filtered simultaneously by two different filters w.transfer functions $H_1(\omega),H_2(\omega)$ as shown.It is given that $R_X(\tau)=\sigma_x^2e^{-B|\tau|}$ (B +ve const), $H_1(\omega)=\begin{cases}1\ ;\omega\in[0,\pi B]\\ 0\ ;else\end{cases}$, $H_2(\omega)=H_1(-\omega)$ and $R_{n_2}(\tau)=N_0\delta(\tau)$. FIND $\hat{x}_1(t)$, $\hat{x}_2(t)$, which are the best linear estimators of $x(t)$ obtained from the signals $y_1(t)$ and $y_2(t)$ respectively. Note that at each time t, the estimator $\hat{x}_j(t)$ makes use of its knowledge of $y_j(t)$ at all past and future times.



ANS: As seen in class, in this case $x(t),y_1(t)$ are JWS, and so are $x(t),y_2(t)$. Hence $\widehat{x_1}(t),\widehat{x_2}(t)$ are given by wiener filters. This case is the special case of orthogonal additive noise, so the Wieners are given by the closed form formula: $\widehat{x_1}(t)=g_1(t)\star y_1(t)$, $\widehat{x_2}(t)=g_2(t)\star y_2(t)$, with: $G_1(\omega)=\frac{S_{X_1}(\omega)}{S_{X_1}(\omega)+S_{N_1}(\omega)}$, $G_2(\omega)=\frac{S_{X_2}(\omega)}{S_{X_2}(\omega)+S_{N_2}(\omega)}$, where we define $x_1(t)=h_1(t)\star x(t)$,$X_2(t)=h_2(t)\star x(t)$. $S_{N_1}(\omega)=S)N_2\,(\omega)=N_0$. $S_{X_1}(\omega)=|H_1(\omega)|^2S_X(\omega)=\begin{cases}S_X(\omega)\ ;w\in[0,\pi B]\\ 0:else\end{cases}$. $S_{X_2}(\omega)=|H_2(\omega)|^2S_X(\omega)=\begin{cases}S_X(\omega)\ ;w\in[-\pi B,0]\\ 0:else\end{cases}$. $S_X(\omega)=\mathcal{FT}\{\sigma_x^2e^{-B|\tau|}\}=\sigma_x^2\int_{-\infty}^{\infty}e^{-B|\tau|}e^{-i\omega\tau}d\tau=\sigma_x^2\left[\int_{-\infty}^{0}e^{\tau(B-i\omega)}d\tau+\int_0^\infty e^{-\tau(B+i\omega)}d\tau\right]=\sigma_x^2\left[\int_0^\infty e^{-\tau(B-i\omega)}d\tau+\int_0^\infty e^{-\tau(B+i\omega)}d\tau\right]=\sigma_x^2\left[\frac{1}{B-i\omega}+\frac{1}{B+i\omega}\right]=\frac{2B\sigma_x^2}{B^2+\omega^2}$. In conclusion: $G_1(\omega)=\begin{cases}\frac{2B\sigma_x^2/(B^2+\omega^2)}{[2B\sigma_x^2/(B^2+\omega^2)]+N_0}\ ;\omega\in[0,\pi B]\\ 0\ ;else\end{cases}=\begin{cases}\frac{2B\sigma_x^2}{2B\sigma_x^2+N_0(B^2+\omega^2)}\ ;\omega\in[0,\pi B]\\ 0\ ;else\end{cases}$. $G_2(\omega)=\begin{cases}\frac{2B\sigma_x^2}{2B\sigma_x^2+N_0(B^2+\omega^2)}\ ;\omega\in[0,\pi B]\\ 0\ ;else\end{cases}$. **C)** It was suggested to look at linear est. in the form $\hat{x}(t)=w_1(t)\star y_1(t)+w_2(t)\star y_2(t)$. Express the optimal frequency responses $W_1(\omega)$ or $w_1(t)$ and $W_2(\omega)$ of $w_2(t)$ as a function of the power spectrum of $x(t)$, of N_0 , and of the functions $H_1(\omega),H_2(\omega)$. Is this the optimal linear estimator of $x(t)$ on the basis of $y_1(t)$ and $y_2(t)$? Explain. **ANS:** A key observation is to realize that $y_1(t),y_2(t)$ contain information regarding $x(t)$ in *disjoint frequency bands*. Hence, similarly to sol of Q3 of Rec11, the optimal linear estimator is: $\hat{x}(t)=\hat{x}_1(t)+\hat{x}_2(t)=g_1(t)\star y_1(t)+g_2(t)\star y_2(t)$. Hence $g_1(t)=w_1(t)$, $g_2(t)=w_2(t)$, $G_1(\omega)=W_1(\omega)$, $G_2(\omega)=W_2(\omega)$.