EC505 CHEAT SHEET

PROBABILITY REVIEW

- Probability space: a triplet $(\Omega, \mathcal{F}, P)$
  - Ω a nonempty set, the sample space, with $\omega \in \Omega$ an outcome
  - $\mathcal{F}$ a set of subsets of Ω called events; F satisfies the axioms:
    - A.1 $\Omega \in F$
    - A.2 If $A \in F$ then $A^c \in F$
    - A.3 If $A, B \in F$ then $A \cup B \in F$, $\bigcup_{i=1}^{\infty} A_i \in F$
  - P a probability measure satisfying:
    - P.1 $P(A) \geq 0 \ \forall A \in F$
    - P.2 If $A, B \in F$ and A, B mutually exclusive, $P(A \cup B) = P(A) + P(B)$ and if $A_1, A_2 \ldots$ a sequence of mutually exclusive events then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
    - P.3 $P(\Omega) = 1$
- Properties of probability spaces:
  - If $A \subset B$ then $P(A) \leq P(B)$
  - $P(A) + P(A)^c = 1$
  - $P(A \cup B) = P(A) + P(B) - P(AB)$
  - $P(\emptyset) = 0$
  - $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC)$
- Continuity of probability lemma: Suppose $B_1, B_2, \ldots$ is a sequence of events
  - a) If $B_1 \subset B_2 \subset \ldots$, then $\lim_{j \to \infty} P(B_j) = P(\bigcup_{i=1}^{\infty} B_i)$
  - b) If $B_1 \supset B_2 \supset \ldots$, then $\lim_{j \to \infty} P(B_j) = P(\bigcap_{i=1}^{\infty} B_i)$
- independence: $P(A_{i_1} A_{i_2} A_{i_3} \cdots A_{i_j}) = P(A_{i_1}) P(A_{i_2}) P(A_{i_3}) \ldots P(A_{i_j})$
- pairwise independence: Any two events are independent
- conditional probability: $P(A|B) = \frac{P(AB)}{P(B)}$
- partition: Events $E_1 \ldots E_k$ mutually exclusive and $\Omega = E_1 \cup E_2 \cup \ldots \cup E_k$
- Bayes Theorem: $P(E_i|A) = \frac{P(AE_i)}{P(A)} = \frac{P(A|E_i)P(E_i)}{P(A)}$

$$= \frac{P(A|E_i)P(E_i)}{P(A|E_1)P(E_1) + \ldots + P(A|E_k)P(E_k)}$$

- infinitely often: The event $\{A_n$ infinitely often$\}$ is the set of $\omega \in \Omega$ such that $\omega \in A_n$ for infinitely many values of n.
- Borel-Cantelli Lemma: Let $(A_n : n \geq 1)$ be a sequence of events and let $p_n = P(A_n)$
  - a) If $\sum_{n=1}^{\infty} p_n < \infty$, then $P\{A_n$ infinitely often$\} = 0$
  - b) If $\sum_{n=1}^{\infty} p_n = \infty$, and $A_1, A_2 \ldots$ are mutually independent, then $P\{A_n$ infinitely often$\} = 1$
- Random Variable: given a probability space $(\Omega, \mathcal{F}, P)$, a random variable is a function X from Ω to $\mathbb{R}$ that is $\mathcal{F}$ measureable, meaning for any # c, $\{\omega : X(\omega) \leq c\} \in \mathcal{F}$
  - CDF: cumulative distribution function $F_X(c) = P\{\omega : X(\omega) \leq c\} = P\{X \leq c\}$
    - PROPERTIES:
    - F.1 F is nondecreasing
    - F.2 $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$
    - F.3 F is right continuous
  - DISCRETE RV:
    - probability mass function $p_X(x) = P\{X = x\}$
    - CDF $F_X(x) = \sum_{y: y \leq x} p_X(y)$
  - Continuous RV:
    - CDF: $F_X(x) = \int_{-\infty}^{x} f_X(y) dy$
    - $F^{-1}(u) = \min\{x : F(x) \geq u\}$
    - PDF: $f_X(x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} P\{x \leq X \leq x + \epsilon\}$
    - $P\{X \in A\} = \int_A f_X(x) dx$
- EXPECTATION: $E[X] = \sum_{i=1}^{\infty} x_i P\{X = x_i\}$ $E[X] = \sum x p_X(x)$ $= \int_{-\infty}^{\infty} x f_X(x) dx$
  - PROPERTIES:
  - E.1 (Linearity) $E[cX] = cE[X]$, $E[X+Y] = E[X] + E[Y]$
  - E.2 (Preservation of order) If $P\{X \geq Y\} = 1$ and $E[Y] > -\infty$, then $E[X] \geq E[Y]$
  - E.3 (Law of the unconscious statistician): $E[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$
  - E.4 (Integration by Parts): $E[X] = \int_0^{\infty} (1 - F(x)) dx - \int_{-\infty}^{0} F_X(x) dx$
- Variance: $Var(X) = E[(X - E(X))^2] = E[X^2] - E[X]^2$
- Markov Inequality: For Y nonnegative, $c > 0$, $P\{Y \geq c\} \leq \frac{E[Y]}{c}$
- Chebyshev Inequality: For X w/ finite mean μ, variance σ^2, then for $d > 0$ $P\{|X - \mu| \geq d\} \leq \frac{\sigma^2}{d^2}$
- characteristic function of a random variable: $\Phi_X(u) = \int_{-\infty}^{\infty} \exp(jux) f_X(x) dx$
  - If $E[X^k]$ exists & is finite for an integer $k \geq 1$, the derivatives of Φ_X up to order k exist and are continuous and $\Phi_X^{(k)}(0) = j^k E[X^k]$
  - Z transform: $\Psi_X(z) = E[z^X] = \sum_{k=0}^{\infty} z^k p_X(k)$
    - $\Psi_X^{(k)}(1) = E[X(X-1) \cdots (X-k+1)]$

FREQUENTLY USED DISTRIBUTIONS

- Bernoulli: $Be(p)$, $0 \leq p \leq 1$
  - pmf: $p(i) = \begin{cases} p & i=1 \\ 1-p & i=0 \\ 0 & \text{else} \end{cases}$
  - z-transform: $1 - p + pz$ (# of heads in one coin flips)
  - mean: p
  - variance: $p \cdot (1-p)$
- Binomial: $Bi(n,p)$, $n \geq 1$, $0 \leq p \leq 1$
  - pmf: $p(i) = \binom{n}{i} p^i (1-p)^{n-i}$ $0 \leq i \leq n$
  - z-transform: $(1 - p + pz)^n$ (# of heads in n independent coin flips)
  - mean: np, variance: $np(1-p)$
- Poisson:
  - pmf: $p(i) = \frac{\lambda^i e^{-\lambda}}{i!}$, $i \geq 0$
  - mean: λ
  - variance: λ (# of phone calls placed during a 10 second interval in a large city)
  - z transform: $\exp(\lambda(z-1))$ ★Limit of binomial as $n \to +\infty$, $p \to 0$ s.t. $np \to \lambda$
- Geometric: $Geo(p)$, $0 < p \leq 1$
  - pmf: $p(i) = (1-p)^{i-1} p$ $i \geq 1$
  - mean: $\frac{1}{p}$ (# of independent coin flips until heads first appears)
  - z-transform: $\frac{pz}{1 - z + pz}$
  - variance $\frac{1-p}{p^2}$
  - SIGNIFICANT PROPERTY (MEMORYLESS): If X has geometric distribution $P\{X > i\} = (1-p)^i$ for integers $i \geq 1$. So X has the memoryless property: $P(X > i + j \mid X > i) = P(X > j)$ for $i, j \geq 1$.
- Gaussian: $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma \geq 0$
  - pdf (if $\sigma^2 > 0$): $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$
  - characteristic function: $\exp\left(j u \mu - \frac{u^2 \sigma^2}{2}\right)$
  - pmf: (if $\sigma^2 = 0$): $p(x) = \begin{cases} 1 & x = \mu \\ 0 & \text{else} \end{cases}$
  - mean: μ variance: σ^2
  - CDF: $\Phi(c)$
    - $Q(c) = 1 - \Phi(c) = \int_c^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$
  - CENTRAL LIMIT THEOREM: If $X_1, X_2, \ldots$ are independent and identically distributed w/ mean μ and variance σ^2, then for any c,

$$\lim_{n \to \infty} P\left\{\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}} \leq c\right\} = \Phi(c)$$

- Exponential: $Exp(\lambda)$, $\lambda > 0$
  - pdf: $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$
  - mean: $1/\lambda$, variance $1/\lambda^2$ (e.g. time elapsed b/w 12pm and the first telephone call placed in a large city)
  - char. function: $\frac{\lambda}{\lambda - ju}$
  - SIGNIFICANCE (memoryless property): If $X \sim Exp(\lambda)$, $P\{X \geq t\} = e^{-\lambda t}$ for $t \geq 0$, so X has the memoryless property: $P\{X \geq s + t \mid X \geq s\} = P\{X \geq t\}$ $s, t \geq 0$
- Uniform: $U(a,b)$ $-\infty < a < b < \infty$
  - pdf: $f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{else} \end{cases}$
  - char function: $\frac{e^{jub} - e^{jua}}{ju(b-a)}$
  - mean: $\frac{a+b}{2}$
  - var: $\frac{(b-a)^2}{12}$
- Gamma: $Gamma(n, \alpha)$
  - $n, \alpha > 0$ (n real valued)
  - pdf: $f(x) = \frac{\alpha^n x^{n-1} e^{-\alpha x}}{\Gamma(n)}$, $x \geq 0$, where $\Gamma(n) = \int_0^{\infty} s^{n-1} e^{-s} ds$
  - char funct: $\left(\frac{\alpha}{\alpha - ju}\right)^n$
  - mean $= n/\alpha$ var: n/α^2
  - SIGNIFICANCE: If n is a positive integer then $\Gamma(n) = (n-1)!$ and a Gamma random variable has the same distribution as the sum of n independent, $Exp(\alpha)$ distributed RVs
- Rayleigh(σ^2):
  - pdf: $f(r) = \frac{r}{\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right)$ $r > 0$
  - mean $\sigma\sqrt{\frac{\pi}{2}}$, variance $\sigma^2(2 - \pi/2)$
  - CDF: $1 - \exp\left(\frac{r^2}{2\sigma^2}\right)$
  - SIGNIFICANCE: If X & Y independent, $N(0, \sigma^2)$, then $(X^2 + Y^2)^{1/2}$ has the Rayleigh(σ^2) distribution
- Failure rate functions: $h(t) \triangleq \lim_{\epsilon \to 0} \frac{P(t < T \leq t + \epsilon \mid T > t)}{\epsilon}$, given the item is working at time t, the probability it fails in the next ϵ time units is $h(t)\epsilon + o(\epsilon)$
  - $h(t) = \frac{f_T(t)}{1 - F_T(t)}$, $F(t) = 1 - e^{-\int_0^t h(s) ds}$
- Sylvester's Criterion for positive definiteness: an nxn matrix is positive definite iff all of its leading principal minors are positive (upper left 1x1, 2x2, 3x3, & M itself)
- correlation coefficient: $\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$

- joint CDF: of $X_1, X_2, \ldots X_m$ RVs is $F_{X_1, \ldots X_m}(x_1, \ldots x_m) = P(X_1 \leq x_1, \ldots X_m \leq x_m)$
  - Note the marginal pdf/pmf is given by:

$$F_1(x) = F_{X_1}F_{X_2}(x_1, +\infty) = \int_{-\infty}^{\infty} f_{X_1 X_2}(x_1, x_2) dx_2$$

$$p_{X_1}(x_1) = \sum_{x_2} p_{X_1 X_2}(x_1, x_2)$$

- joint characteristic function: $\Phi_{X_1 \ldots X_m}(x_1, \ldots x_m) = E[e^{j(X_1 x_1 + \ldots + X_m x_m)}]$
- If random variables are independent, then $F_{X_1 X_2 \ldots X_m}(x_1, \ldots x_m) = F_{X_1}(x_1) \ldots F_{X_m}(x_m)$
- conditional pdf: $f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$, $-\infty < x < +\infty$
- conditional expectation: $E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$

CORRELATION & COVARIANCE

- correlation: $E[XY]$
- covariance: $Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$
- correlation coefficient: $\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$ → $|Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)}$
- Schwarz Inequality: $|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}$
- Triangle inequality: $E[(X+Y)^2]^{1/2} \leq E[X^2]^{1/2} + E[Y^2]^{1/2}$
- Linearity of Covariance: $Cov(X+Y, U+V) = Cov(X,U) + Cov(X,V) + Cov(Y,U) + Cov(Y,V)$
  - $Cov(aX + b, cY + d) = ac\,Cov(X,Y)$

RANDOM VECTORS

- a random vector X of dimension m has the form $X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}$, where $X_1, \ldots X_m$ are random variables. We can consider the joint distribution of $X_1 \ldots X_m$ as the distribution of X

$$\left[f_Y(y) = \frac{f_X(x)}{\left|\frac{\partial y}{\partial x}(x)\right|} = f_X(x)\left|\frac{\partial x}{\partial y}(y)\right|\right]$$

MATRICES

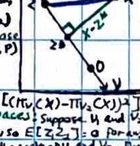
- An mxn matrix over $\mathbb{R}$ has the form $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & \cdots & & a_{mn} \end{bmatrix}$, $a_{ij} \in \mathbb{R} \ \forall i,j$
- transpose of an mxn matrix $A = (a_{ij})$ is the nxm matrix $A^T = (a_{ji})$
- A matrix is symmetric if $A = A^T$
- A matrix's diagonal comprises the entries of the form a_{ii}
- A square matrix is diagonal if all entries off the diagonal are 0
- Given $A_{m \times k}$, $B_{k \times n}$, the product AB is the mxn matrix w/ ij'th element $\sum_{l=1}^{k} a_{il} b_{lj}$
- A vector x is an mx1 matrix
- The set of all m-dimensional vectors over $\mathbb{R}^m$ is the m-dimensional Euclidean space $\mathbb{R}^m$
- inner product of two vectors is $x^T y$, scalar, $\sum_{i=1}^{m} x_i y_i$
- Two vectors x, y are orthogonal if $x^T y = 0$
- norm/Euclidean length of a vector is $\|x\| = (x^T x)^{1/2}$
- A set of vectors $e_1, \ldots e_n$ is orthonormal if they are all orthogonal to each other and $\|e_i\| = 1 \ \forall i$
- A set of vectors $v_1, \ldots v_n$ in $\mathbb{R}^m$ span $\mathbb{R}^m$ if any vector in $\mathbb{R}^m$ can be expressed as a linear combo $\alpha_1 v_1 + \alpha_2 v_2 + \ldots \alpha_n v_n$ for some $\alpha_1, \ldots \alpha_n \in \mathbb{R}$
- An orthonormal set of vectors in $\mathbb{R}^m$ spans $\mathbb{R}^m$ iff $n = m$
  - ↳ $e_1, \ldots e_n$
- An orthonormal basis for $\mathbb{R}^m$ is an orthonormal set of m vectors in $\mathbb{R}^m$ & corresponds to a coordinate system for $\mathbb{R}^m$ w/ coordinates for v given by $\alpha_i = e_i^T v$, and $v = \alpha_1 e_1 + \ldots + \alpha_m e_m$
- A square matrix is orthonormal if any of the following are satisfied
  - 1. $U^T U = I$ 2. $UU^T = I$ 3. The columns form an orthonormal basis
- Given an mxm orthonormal matrix U and $v \in \mathbb{R}^m$, the coordinates of v relative to U are given by $U^T v$
- Given a square matrix A, a vector e is an eigenvector of A and λ is an eigenvalue of A iff $Ae = \lambda e$
- Determinant of a square matrix A given by $det(A) = \sum_{\pi} (-1)^{\pi} \prod_{i=1}^{m} a_{i\pi(i)}$
- Properties of determinants
  - 1. If B is obtained from A by multiplication of a row or column of A by a scalar constant c then $det(B) = c\,det(A)$
  - 2. Given U subset of $\mathbb{R}^m$, $V = \{Ax : x \in U\}$, then $volume(U) = |A| \times volume(V)$
  - 3. $det(AB) = det(A)det(B)$
  - 4. $det(A) = det(A^T)$
  - 5. $|U| = 1$ if U is orthonormal
  - 6. Columns of A span $\mathbb{R}^n$ iff $det(A) \neq 0$
  - 7. Equation $p(\lambda) = det(\lambda I - A)$ defines characteristic polynomial of A
  - 8. The zeros of the characteristic polynomial are the eigenvalues of A and $det(A) = \prod_{i=1}^{n} \lambda_i$
- Let K be an mxm symmetric matrix w/ eigenvalues $\lambda_1, \ldots \lambda_m$ with an orthonormal basis consisting of the corresponding eigenvectors $e_1, \ldots e_m$. Let U be the orthonormal matrix w/ columns $e_1, \ldots e_m$ and let Λ be the diagonal matrix w/ diagonal entries given by eigenvalues. We have $KU = U\Lambda$, so $K = U\Lambda U^T$ and $\Lambda = U^T K U$
- A symmetric mxm matrix A is positive semidefinite if $\alpha^T A \alpha \geq 0 \ \forall \alpha \in \mathbb{R}^m$
- A symmetric matrix is positive semidefinite iff eigenvalues are nonnegative

RANDOM VECTORS AND MINIMUM MEAN SQUARED ERROR ESTIMATION

- A random vector X of dimension m has the form $X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}$ and the expectation $E[X]$ defined by $E[X] \triangleq \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_m] \end{pmatrix}$
- Suppose Y another RV on the same probability space as X w/ dimension n. The cross-correlation matrix of X and Y is the mxn matrix $E[XY^T]$, which has ij'th entry $E[X_i Y_j]$
- The cross covariance matrix of X and Y, denoted by $Cov(X,Y)$, is the matrix of ij'th entry $Cov(X_i, Y_j)$
- Correlation matrix of X is $E[XX^T]$ and covariance matrix $Cov(X,X) = Cov(X)$
- Elementary properties of expectation, correlation, and covariance:
  - 1. $E[AX + b] = AE[X] + b$
  - 2. $Cov(X,Y) = E[X(Y - E[Y])^T] = E[(X - E[X])Y^T] = E[XY^T] - E[X]E[Y]^T$
  - 3. $E[(AX)(CY)^T] = AE[XY^T]C^T$
  - 4. $Cov(AX + b, CY + d) = ACov(X,Y)C^T$
  - 5. $Cov(AX + b) = ACov(X)A^T$
  - 6. $Cov(W + X, Y + Z) = Cov(W,Y) + Cov(W,Z) + Cov(X,Y) + Cov(X,Z)$
  - ★If $E[X] = 0$ or $E[Y] = 0$, then cross correlation and cross covariance matrix are equal!★
- PROPOSITION: Correlation & covariance matrices are symmetric positive semidefinite matrices AND if K is a symmetric positive semidefinite matrix, then K is the covariance and correlation matrix for some mean 0 random vector X.
- Characteristic function of random vector: $\Phi_X(u) = E[\exp(ju^T X)]$
- Orthogonality principle for MMSE estimation
  - Let X be an RV w/ some known distribution; we want to estimate X.
  - If we estimate w/ a constant b, the error will be $X - b$
  - The mean square error (MSE) is $E[(X-b)^2]$

$$E[(X-b)^2] = E[((X - E[X]) + (E[X] - b))^2]$$
$$= E[(X - E[X])^2 + 2\cancel{(X - E[X])(E[X] - b)} + (E[X] - b)^2]$$
$$= Var(X) + (E[X] - b)^2$$

- clearly MSE minimized w.r.t. b when $b = E[X]$, and the minimum possible value is $Var(X)$
- Random variables are orthogonal $(X \perp Y)$ if $E[XY] = 0$
- The Orthogonality Principle: Let V be a closed linear subspace of $L^2(\Omega, \mathcal{F}, P)$ and let $X \in L^2(\Omega, \mathcal{F}, P)$, for some probability space.
  - a) (Existence & Uniqueness) There exists a unique element Z^* $(\Pi_V(X))$ in V s.t. $E[(X - Z^*)^2] \leq E[(X - Z)^2] \ \forall Z \in V$
  - b) (Characterization) Let W be a random variable. Then $W = Z^*$ iff
    - i) $W \in V$
    - ii) $(X - W) \perp Z$ for all Z in V
  - c) (Error expression) The minimum mean square error (MMSE) is given by $E[(X - Z^*)^2] = E[X^2] - E[(Z^*)^2]$ (by orthogonality!)
- Linearity of projection: Suppose V a closed linear subspace of $L^2(\Omega, \mathcal{F}, P)$, X_1 and X_2 are in $L^2(\Omega, \mathcal{F}, P)$ and a_1, a_2 constants. Then $\Pi_V(a_1 X_1 + a_2 X_2) = a_1 \Pi_V(X_1) + a_2 \Pi_V(X_2)$



- Projections onto nested subspaces: Suppose V_1, V_2 are closed linear subspaces of $L^2(\Omega, F, P)$ s.t. $V_2 \subset V_1$. Then for any $X \in L^2(\Omega, F, P)$, $\Pi_{V_2}(X) = \Pi_{V_2}(\Pi_{V_1}(X))$ and $E[(X - \Pi_2(V))^2] = E[(X - \Pi_{V_1}(X))^2] + E[(\Pi_{V_1}(X) - \Pi_{V_2}(X))^2]$
- Projection onto the span of orthogonal subspaces: Suppose V_1 and V_2 are closed linear subspaces of $L^2(\Omega, F, P)$ s.t. $V_1 \perp V_2$, so $E[Z_1 Z_2] = 0$ for any $Z_1 \in V_1$, $Z_2 \in V_2$. Let $V = V_1 \oplus V_2 = \{Z_1 + Z_2 : Z_i \in V_i\}$ denote the span of V_1 and V_2. Then for any $X \in L^2(\Omega, F, P)$, $\Pi_V(X) = \Pi_{V_1}(X) + \Pi_{V_2}(X)$, and the MMSE satisfies $E[(X - \Pi_V(X))^2] = E[X^2] - E[(\Pi_{V_1}(X))^2] - E[(\Pi_{V_2}(X))^2]$
- Variance of linear combo of RVs:
  - $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X,Y)$
  - $Var(aX - bY) = a^2 Var(X) + b^2 Var(Y) - 2ab\,Cov(X,Y)$

CONDITIONAL EXPECTATION AND LINEAR ESTIMATORS

- Suppose a random variable X is to be estimated using an observed RV Y of dimension m. Suppose $E[X^2] < +\infty$ and consider the most general class of estimators based on Y by setting $V = \{g(Y) : g : \mathbb{R}^m \to \mathbb{R}, E[g(Y)^2] < +\infty\}$
- We have the implicit conditions that g is Borel measureable and g(Y) is a random var.
- The projection of X onto this class V is the unconstrained MMSE estimator of X given Y

$$g^*(y) = E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$$

OPTIMAL ESTIMATOR: $g^*(Y) = E[X|Y]$

- For a random vector X, we similarly have $g^*(Y) = E[X|Y] = \begin{bmatrix} E[X_1|Y] \\ \vdots \\ E[X_m|Y] \end{bmatrix}$
- $Cov(e) = Cov(X) - Cov(E[X|Y])$ $e = X - E[X|Y]$
- Linear Estimators: Let X, Y be random vectors, $E[\|X\|^2] < +\infty$ and $E[\|Y\|^2] < +\infty$. We want estimators of the form $AY + b$ to minimize the MSE. We call these linear estimators b/c each coordinate of $AY + b$ is a linear combo of $Y_1, \ldots Y_m$ and 1.
  - To identify the optimal linear estimator, we apply the orthogonality principle for each coordinate of X w/ $V = \{c_0 + c_1 Y_1 + c_2 Y_2 + \ldots c_n Y_n\}$
  - e is the estimation error $e = X - (AY + b)$, and we must choose A, b s.t. $e_i \perp Z$ for all $Z \in V$, so $e_i \perp 1 \ \forall i$, $e_i \perp Y_j \ \forall i,j$
    - ↳ implies $E e_i = 0 \Rightarrow E[e_i Y_j] = Cov(e_i, Y_i)$
- So, $\hat{E}[X|Y] = E[X] + Cov(X,Y)Cov(Y,Y)^{-1}(Y - E[Y])$
  - $Cov(e) = Cov(X) - Cov(\hat{E}[X|Y])$
  - $= Cov(X) - Cov(X,Y)Cov(Y,Y)^{-1}Cov(X,Y)$ → $E[e^T e]$, MSE
- ORDERING OF MSEs: $E[(X - E[X|Y])^2] \leq E[(X - \hat{E}[X|Y])^2] \leq Var(X)$ (by projections onto nested subspaces)
- Tower property of conditional expectation: $E[E[X|Y_1, Y_2]|Y_1] = E[X|Y_1]$
  - $E[E[X|Y]] = E[X]$

JOINT GAUSSIAN DISTRIBUTION & GAUSSIAN RVS

- LEMMA: Suppose $X_1, X_2, \ldots X_n$ are independent Gaussian RVs. Then any linear combo $a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$ is a Gaussian RV
- joint Gaussian distribution: A collection $(X_i : i \in I)$ of random variables has a joint Gaussian distribution (and the X_i's are said to be jointly Gaussian) if every finite linear combination of $(X_i : i \in I)$ is a Gaussian random variable. A random vector X is called a Gaussian random vector if its coordinate RVs are jointly Gaussian. A collection of random vectors has a joint Gaussian distribution if all coordinate RVs of all vectors are jointly Gaussian.
- We write X is a $N(\mu, K)$ random vector if X a Gaussian random vector w/ mean vector μ, covariance matrix K
- PROPOSITION
  - a) If $(X_i : i \in I)$ has a joint Gaussian distribution, then each of the random variables itself is Gaussian
  - b) If the random variables $X_i : i \in I$ are each Gaussian and independent (so $X_{i_1}, X_{i_2}, \ldots X_{i_n}$ are independent for any finite number of indices $i_1, i_2, \ldots i_n$), then $(X_i : i \in I)$ has a joint Gaussian distribution
  - c) (Preservation of joint Gaussian property under linear combinations and limits) Suppose $(X_i : i \in I)$ has a joint Gaussian. Let $(Y_j : j \in J)$ denote a collection of RVs s.t. each Y_j is a finite linear combo of $(X_i : i \in I)$ and let $(Z_k : k \in K)$ be a set of RVs s.t. each Z_k is a limit in probability of a sequence from $(Y_j : j \in J)$. Then $(Y_j : j \in J)$ and $(Z_k : k \in K)$ each have a joint Gaussian distribution
  - d) The characteristic function of an $N(\mu, K)$ random vector is given by $\Phi_X(u) = E[e^{ju^T X}] = e^{ju^T \mu - \frac{1}{2} u^T K u}$
    - For joint Gaussian: $Cov(0) \Leftrightarrow$ uncorrelated
  - e) If X an $N(\mu, K)$ random vector and K a diagonal matrix ($Cov(X_i, X_j) = 0$ for $j \neq i$), then the coordinates $X_1, \ldots X_m$ are independent
  - f) An $N(\mu, K)$ random vector X s.t. K is nonsingular has a pdf:

$$f(x) = \frac{1}{(2\pi)^{m/2} |K|^{1/2}} \exp\left(\frac{-(x-\mu)^T K^{-1} (x-\mu)}{2}\right)$$

    (Any random vector X s.t. Cov(X) is singular does not have a pdf)
  - g) If X and Y are jointly Gaussian vectors, then they are independent iff $Cov(X,Y) = 0$
- Proposition: Let X, Y be jointly Gaussian vectors and $y \in \mathbb{R}$. The conditional distribution of X given $Y = y$ is $N(\hat{E}[X|Y=y], Cov(e))$. In particular, $E[X|Y=y] = \hat{E}[X|Y=y]$ and $E[X|Y] = \hat{E}[X|Y]$

LINEAR INNOVATIONS SEQUENCES

- Let $X, Y_1, \ldots Y_n$ be random vectors w/ finite second moments on the same prob. space
- If $E[Y_i] = 0 \ \forall i$, $E[Y_i Y_j^T] = 0$ for $i \neq j$, then

★$\hat{E}[X|Y_1, \ldots Y_n] = \bar{X} + \sum_{i=1}^{n} \hat{E}[X - \bar{X}|Y_i]$ (by orthog.) $(\bar{X} = E[X])$

- Let e be the estimation error:

$$e = X - \bar{X} - \sum_{i=1}^{n} \hat{E}[X - \bar{X}|Y_i]$$

- Must show that $E[e(Y_1^T c_1 + \ldots + Y_n^T c_n + b)] = 0 \ \forall c_1, \ldots c_n$ and b.
- It's enough to show $E[e] = 0$, $E[eY_j^T] = 0 \ \forall j$
- $\hat{E}[X - \bar{X}|Y_i]$ has form $B_i Y_i$ since $X - \bar{X}$ and Y_i have mean 0, so $E[e] = 0$
- Also $E[eY_j^T] = E[(X - \hat{E}[X|Y_j])Y_j^T] - \sum_{i: i \neq j} E[B_i Y_i Y_j^T]$
- Note in order to use ★, 1, $Y_1, \ldots Y_n$ must be orthogonal
- We can orthogonalize in the following way:
  - $\tilde{Y}_1 = Y_1 - E[Y_1]$, and for $k \geq 2$ $\tilde{Y}_k = Y_k - \hat{E}[Y_k|Y_1, \ldots Y_{k-1}]$
- It is possible to prove that the set of all RVs obtained by linear transformation of $1, Y_1, \ldots Y_k$ is equal to the set of all RVs obtained by linear transformation of $1, \tilde{Y}_1, \ldots \tilde{Y}_k$.
- So, for any random vector X, we have

$$E[X|Y_1, \ldots Y_n] = \hat{E}[X|\tilde{Y}_1, \ldots \tilde{Y}_n] = \bar{X} + \sum_{i=1}^{n} \hat{E}[X - \bar{X}|\tilde{Y}_i]$$
$$= \bar{X} + \sum_{i=1}^{n} Cov(X, \tilde{Y}_i)Cov(\tilde{Y}_i)^{-1}\tilde{Y}_i$$

$$\tilde{Y}_1 = Y_1 - E[Y_1] \quad \tilde{Y}_k = Y_k - E[Y_k] - \sum_{i=1}^{k-1} Cov(X, \tilde{Y}_i)Cov(\tilde{Y}_i)^{-1}\tilde{Y}_i$$

DISCRETE-TIME KALMAN FILTERING

- GOAL: Estimate $x_0, x_1, \ldots$ from observed sequence $y_0, y_1, \ldots$
  - State: $x_{k+1} = F_k x_k + w_k$
  - Observation: $y_k = H_k^T x_k + v_k$
- ASSUMPTIONS:
  - $x_0, v_0, v_1, \ldots w_0, w_1, \ldots$ are pairwise uncorrelated
  - $E[x_0] = \bar{x}_0$, $Cov(x_0) = P_0$, $E[w_k] = 0$, $Cov(w_k) = Q_k$, $E[v_k] = 0$, $Cov(v_k) = R_k$
  - F_k, H_k, Q_k, R_k, for $k \geq 0$; P_0 are known matrices
  - $\bar{x}_0$ a known vector
- Let $\bar{x}_k = E[x_k]$ and $P_k = Cov(x_k)$. These quantities are recursively determined:
  - $\bar{x}_{k+1} = F_k \bar{x}_k$ and $P_{k+1} = F_k P_k F_k^T + Q_k$
- Let $y^k = (y_0, y_1, \ldots y_k)$ represent the observations up to time k and define

$$\hat{x}_{i|j} = \hat{E}[x_i|y^j], \quad \Sigma_{i|j} = Cov(x_i - \hat{x}_{i|j})$$

- We want to compute $\hat{x}_{k+1|k}$
- KALMAN FILTER EQUATIONS:

$$\hat{x}_{k+1|k} = [F_k - K_k H_k^T]\hat{x}_{k|k-1} + K_k y_k$$
$$= F_k \hat{x}_{k|k-1} + K_k[y_k - H_k^T \hat{x}_{k|k-1}]$$

Where $\hat{x}_{0|-1} = \bar{x}_0$, $\Sigma_{0|-1} = P_0$

$$K_k = F_k \Sigma_{k|k-1} H_k [H_k^T \Sigma_{k|k-1} H_k + R_k]^{-1}$$

$$\Sigma_{k+1|k} = F_k[\Sigma_{k|k-1} - \Sigma_{k|k-1} H_k (H_k^T \Sigma_{k|k-1} H_k + R_k)^{-1} H_k^T \Sigma_{k|k-1}]F_k^T + Q_k$$

$E[XY] = \iint xy f_{XY}(x,y)$

CONDITIONAL DISTRIBUTION: $f(y|x) = P(Y=y|X=x)$

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$

HW 3.4 Consider $X \sim N(0, \sigma_x^2)$, $V \sim N(0, \sigma_v^2)$, H Bernoulli w/ param p
$Y = HX + V$, H, X, V independent

a) Find $\hat{E}[X|Y]$, the LMMSE estimate of X based on Y, variance of estimates and MSE.

$\mu_x = E[X] = 0$ $\mu_y = E[Y] = E[HX + V] = E[X]E[H] + E[V] = 0$

$Cov(X) = E[X^2] = \sigma_x^2$ $Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)]$
$= E[XY] = E[X(HX+V)]$
$= E[H]E[X^2] + E[X]E[V]$
$= \sigma_x^2 p$

$Cov(Y) = E[Y^2] = E[(HX+V)^2]$
$= E[H\cancel{X}^2] + E[V^2] + 2E[HVX]$
$= E[H^2]E[X^2] + \sigma_v^2 + 0$
$= p\sigma_x^2 + \sigma_v^2$

$$\hat{E}[X|Y=y] = \mu_x + Cov(X,Y)(Cov(Y))^{-1}(y - \mu_y)$$
$$= \left[\frac{p\sigma_x^2}{p\sigma_x^2 + \sigma_v^2}\right] y$$

$$Var(\hat{E}[X|Y]) = E\left[\left(\left(\frac{p\sigma_x^2}{p\sigma_x^2+\sigma_v^2}\right)Y - 0\right)^2\right]$$
$$= \frac{p^2\sigma_x^4}{(p\sigma_x^2+\sigma_v^2)^2} E[Y^2] = \frac{p^2\sigma_x^4}{p\sigma_x^2+\sigma_v^2}$$

$$MSE(\hat{E}[X|Y]) = Cov(X) - Cov(X,Y)Cov(Y)^{-1}Cov(Y,X)$$
$$= \sigma_x^2 - \frac{(p\sigma_x^2)^2}{p\sigma_x^2+\sigma_v^2} = \frac{\sigma_x^4 p(1-p) + \sigma_x^2\sigma_v^2}{p\sigma_x^2 + \sigma_v^2}$$

b) Show $E[X|Y=y] = \Sigma_{i=0}^{1} P(H=i|Y=y)E[X|Y=y, H=i]$.

$E[X|Y=y] = E[E[X|H=i, Y=y]|Y=y]$
$= \sum_{i=0}^{1} P(H=i|Y=y)E[X|H=i, Y=y]$

c) Find $E[X|Y]$ the MMSE (BLS) estimate of X given Y.

We need to compute $E[X|Y=y]$

From b) this is $\Sigma_{i=0}^{1} P(H=i|Y=y)E[X|H=i, Y=y]$

When $H=0$,
$E[X|Y=y, H=0] = E[X|V=y] = E[X] = 0$

When $H=0$,
$Y = X+V$, so $E[X|Y=y, H=1]$ is the BLS estimate for X given $Y = X+V$ where X, V are independent Gaussians, so the estimator is linear!

$E[Y=y, H=1] = \hat{E}[X|Y=y, H=1]$
$= E[X|H=1] + Cov(X,Y|H=1)Cov(Y|H=1)^{-1}(y - E[Y|H=1])$
$= 0 + \frac{\sigma_x^2}{\sigma_x^2+\sigma_v^2}(y - 0) = \frac{\sigma_x^2 y}{\sigma_x^2+\sigma_v^2}$

Now we just need $P(H=1|Y=y)$

$$P(H=1|Y=y) = \frac{f_{Y|H}(y|H=1)P(H=1)}{f_Y(y)} = \frac{f_{Y|H}(y|H=1)P(H=1)}{\Sigma_{i=0}^{1} f_{Y|H}(y|H=i)P(H=i)}$$
$$= \frac{pN(0, \sigma_x^2+\sigma_v^2)_y}{pN(0, \sigma_x^2+\sigma_v^2)_y + (1-p)N(0, \sigma_v^2)_y}$$

So, $E[X|Y=y] = P(H=0|Y=y)E[X|H=0, Y=y] + P(H=1|Y=y) \cdot E[X|H=1, Y=y]$

$$= \frac{pN(0, \sigma_x^2+\sigma_v^2)_y}{pN(0,\sigma_x^2+\sigma_v^2)_y + (1-p)N(0,\sigma_v^2)_y} \cdot \frac{\sigma_x^2}{\sigma_x^2+\sigma_v^2} y$$

3.8 Let Z be a 2D Gaussian RV with mean μ_Z, cov mat Λ_Z

$$Z = \begin{bmatrix} W \\ V \end{bmatrix}, \mu_Z = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Lambda_Z = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$$

Let X be Gaussian RV w/ mean $\mu_x = 2$, variance $\lambda_x = 8$, Z, X indep.
We define Y as: $Y = (2+W)X + V$

a) Find $\hat{x}_{LLS}(y)$, the LLSE of X based on $Y=y$.

Y is not Gaussian as it involves the product of two Gaussians.
Compute needed stats:

$E[Y] = E[2+W]E[X] + E[V] = 2E[X] = 4$

$E[Y^2] = E[(2+W)^2X^2 + 2(2+W)XV + V^2]$
$= E[(2+W)^2]E[X^2] + E[2(2+W)V]E[X] + E[V^2]$
$= (4+2)(4+8) + (2)(2) + 4 = 80$

$Var(Y) = 80 - 4^2 = 64$

$E[XY] = E[(2+W)X^2 + XV] = E[2+W]E[X^2] + E[XV]$
$= 2(12) + 0 = 24$

$Cov[XY] = 24 - E[X]E[Y] = 24 - 2 \cdot 4 = 16$

So, $\hat{x}_{LLS}(y) = 2 + \frac{16}{64}(y-4) = 2 + \frac{y-4}{4}$

b) Find λ_{LLS}, mean squared estimation error
$\lambda_{LLS} = 8 - \frac{16^2}{64} = 4$

Common Integrals

$\int \frac{1}{x} dx = \ln|x|$ $\int a^u du = \frac{a^u}{\ln a}$ $\int ue^u du = (u-1)e^u$

$\int x^n dx = \frac{1}{n+1} x^{n+1}$ $\int \ln u \, du = u\ln(u) - u$

$\int \frac{1}{ax+b} dx = \frac{1}{a}\ln|ax+b|$ IBP: $\int u\,dv = uv - \int v\,du$

$Y = ZX_1 + (1-Z)X_2$ NOT Gauss if Z bernoulli

$E[X\cos(Y)|Y] = E[X|Y]\cos(Y)$

Kalman: $E[e_{t|t}e_{t|t}^T] \le E[e_{t|t-1}e_{t|t-1}^T]$
↳ TRUE b/c you get more measurements

HW 3.5 Let X, Y, Z be RVs w/ finite 2nd moments and we want to estimate X. Answer the following T/F questions.

a) $E[(X - E[X|Y])^2] \le E[(X - \hat{E}[X|Y, Y^2])^2]$

$\boxed{\text{TRUE}}$ $\hat{E}[X|Y, Y^2]$ is the best linear estimator of X based on 1, Y, Y²; $E[X|Y]$ is the best estimator of X based on any measurable function of Y, so the MSE can only be smaller

b) $E[(X - E[X|Y])^2] = E[(X - \hat{E}[X|Y, Y^2])^2]$ if X, Y jointly Gaussian.

$\boxed{\text{TRUE}}$ $E[(X - E[X|Y])^2] \le E[(X - \hat{E}[X|Y,Y^2])^2] \le E[(X - \hat{E}[X|Y])^2]$
$= E[(X - E[X|Y])^2]$

MMSE = LMMSE for jointly Gaussian X, Y!

c) $E[(X - E[E[X|Z]|Y])^2] \le E[(X - E[X|Y])^2]$

$\boxed{\text{FALSE}}$ Counterexample: let $\mu_x = 0$, $\sigma_x^2 \ne 0 = 1$
Suppose $Y = X$, Z RV independent of X. Then $E[X|Z] = E[X] = 0$
Then left side: $E[(X-0)^2] = 1$, right: 0 b/c $E[X|Y] = X$

d) If $E[(X - E[X|Y])^2] = Var(X)$, then X, Y independent

$\boxed{\text{FALSE}}$ Claims $E[(X - E[X|Y])^2] = E[(X - E[X])^2] \Rightarrow$ independence
It is possible for $E[X|Y] = E[X]$ w/o X and Y being independent
COUNTEREXAMPLE: $X = WY$, W RV independent of Y w/ 0 mean and unit variance.
So given Y, conditional distribution of X has mean 0 variance Y^2
$E[X|Y] = 0$, $E[(X - E[X|Y])^2] = Var(X)$, but X, Y not independent.

3.6 Suppose Y is Gaussian, w/ $\mu = 0$, $\sigma^2 = 1$, and $X = |Y|$. Find the estimator for X of the form $\hat{X} = a + bY + cY^2$ which minimizes the MSE.

$$\begin{bmatrix} E[|Y|] = 0.8 \\ E[|Y|^4] = 3 \\ E[|Y|^3] = 1.6 \end{bmatrix}$$

a) Use orthogonality to derive equations for a, b, c
At the optimal estimator, error is orthogonal to Y^c for $c = 0, 1, 2$.

$E[(|Y| - (a + bY + cY^2))] = 0 \Rightarrow 0.8 - a - c = 0$
$E[(|Y| - (a + bY + cY^2))Y] = 0 \Rightarrow -b = 0$
$E[(|Y| - (a + bY + cY^2))Y^2] = 0 \Rightarrow 1.6 - a - 3c = 0$

b) Find $\hat{X}$:
We get $0.8 - 4c = 0 \Rightarrow c = 0.4$, $a = 0.4$, $\hat{X}(Y) = 0.4 + 0.4Y^2$

c) Find MSE:
$MSE = E[X^2] - \hat{E}[X^2]$ (by orthogonality)
$= E[Y^2] - E[(0.4 + 0.4Y^2)^2]$
$= 1 - 0.16 - 0.32E[Y^2] - 0.16E[Y^4] = 0.04$

3.7 Let $X = \frac{1}{1+U}$, U uniformly distributed over [0,1]. X a deterministic function of U, so if we allowed non linear estimates, $E[X|U] = X$ w/ no mean square error.

a) Evaluate $\hat{x}_{LLS}(U)$, optimal linear least squares estimate

$E[X] = E\left[\frac{1}{1+u}\right] = \int_0^1 \frac{1}{1+u} du = \ln(2)$

$E[X^2] = \int_0^1 \left(\frac{1}{1+u}\right)^2 du = 1 - \frac{1}{2} = \frac{1}{2}$

$Var[X] = \frac{1}{2} - (\ln(2))^2 = 0.0195$

$E[U] = \frac{1}{2}$

$Var[U] = \frac{1}{12}$

$E[XU] = \int_0^1 \frac{u}{1+u} du = 1 - \ln(2)$

$Cov[X,U] = E[XU] - E[X]E[U] = 1 - (3/2)\ln(2) = -0.04$

$\hat{x}_{LLS} = E[X] + \frac{Cov[X,U]}{Var(U)}(U - \frac{1}{2}) = \ln(2) - 0.48(U - 0.5)$

b) Calculate MSE $E[(X - \hat{x}_{LLS}(U))^2]$

$E[(X - \hat{x}_{LLS}(U))^2] = Var[X] - \frac{Cov(X,U)^2}{Var(U)} = 0.5 - (\ln 2)^2 - 12(1 - \frac{3}{2}\ln(2))^2$

3.9 Let X and Y be random variables s.t. X exponential and conditioned on X, Y is exponential w/ parameter X.

$p_X(x) = \frac{1}{a} e^{-x/a} u(x)$
$p_{Y|X}(y|x) = xe^{-xy}u(y)$

a) Determine $\hat{x}_{BLS}(y) \triangleq E[X|Y=y]$, $\Lambda_{X|Y}(y) = E[(X - \hat{x}_{BLS}(Y))^2 | Y=y]$ and $\Lambda_{BLS} = E[(X - \hat{x}_{BLS}(Y))^2]$

We want $\hat{x}_{BLS}(y) = E[X|Y]$, so we need $f_{X|Y}(x|y)$.
Find $f_Y(y)$ and then use Bayes rule.

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)\,dx$$
$$= \int_0^\infty xe^{-xy}u(y)\frac{1}{a}e^{-x/a}dx = u(y)\frac{1}{a}\int_0^\infty xe^{-x(y+1/a)}dx$$
$$= \frac{u(y)}{a}\frac{1}{(y+\frac{1}{a})^2}$$

So, $f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = (y + \frac{1}{a})^2 xe^{-x(y+1/a)}u(x)$

Then $\hat{x}_{BLS} = E[X|Y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx$, etc.

b) Compute $\hat{x}_{map}(y)$

$$\left[\hat{x}_{MAP} = \arg\max_x f_{X|Y}(x|y) = \arg\max_x (y+\frac{1}{a})^2 xe^{-x(y+1/a)}u(x) = \arg\max_x xe^{-x(y+1/a)}, x \ge 0\right]$$

Take derivative and set to 0

PRACTICE EXAM #2

$Y \sim N(0, \sigma^2)$, Bayes least square estimate of X given Y is $\hat{x}_{BLS}(y) = y^2$

a) Could X & Y be jointly Gaussian?
No, $E[X|Y]$ would be linear in that case

b) Find $\hat{x}_{LLS}(y)$, so $\hat{x}_{LLS}(y) = \alpha y + \beta$

$\hat{x}_{LLS} = \mu_X + \frac{Cov(X,Y)}{Cov(Y)}(y - \cancel{\mu_Y})$

$= \mu_X + \frac{\sigma_{XY}}{\sigma_Y^2} y$ $\sigma_{XY} = E[XY] - E[X]\cancel{E[Y]}$ 0
$= E[XY] = E[YE[X|Y]]$
$= E[Y \cdot Y^2]$
$= E[Y^3]$
$= 0$

So, $\hat{x}_{LLS} = \mu_X$
$= E[E[X|Y^2]]$
$= E[Y^2] = \sigma^2$

c) If error variance of $\hat{x}_{LLS}(y)$ is $\lambda_{LLS} = 3\sigma^4$, find BLS:

$\lambda_{LLS} = \sigma_X^2 - \frac{\sigma_{XY}^2}{\sigma_Y^2} \Rightarrow \lambda_{LLS} = \sigma_X^2 = 3\sigma^4$

MMSE $= E[X^2] - E[(\hat{x}(y))^2]$
$= 3\sigma^4 - E[E[X|Y]^2]$
$= 3\sigma^4 - E[Y^4] = 3\sigma^4 - 3\sigma^4 = 0$

3.11 Let $X_0, V_0, W_0, W_1, V_1, V_2$ be a sequence of zero mean pairwise uncorrelated RVs with $Var(X_0) = 4$, $Var(W_t) = 1$, for $t = 0, 1$, $Var(V_t) = 1$ for $t = 0, 1, 2$. For $t = 1, 2$, let
$X_t = \rho X_{t-1} + W_t$
And for $t = 0, 1, 2$, let $Y_t = X_t + V_t$

a) Find the covariance matrix of $(X_0, X_1, X_2)^T$

$E[X_t] = 0 \; \forall t \ge 0$.
$Var(X_t) = E[X_t^2] = \rho^2 E[X_{t-1}^2] + E[W_t^2] = \rho^2 Var(X_{t-1}) + 1$, $Var(X_0) = 4$
So, $Var(X_1) = 4\rho^2 + 1$, $Var(X_2) = 4\rho^4 + \rho^2 + 1$
B/C X_0, W_0 uncorrelated, we have
$E[X_0X_1] = E[X_0(\rho X_0 + W_0)] = \rho E[X_0^2] = 4\rho$
$E[X_1X_2] = \rho E[X_1^2] = 4\rho^2 + \rho$
$E[X_0X_2] = E[X_0(\rho^2 X_0 + \rho W_0 + W_1)] = 4\rho^2$

$$\text{So, } Cov\begin{bmatrix} X_0 \\ X_1 \\ X_2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 4 & 4\rho & 4\rho^2 \\ 4\rho & 4\rho^2+1 & 4\rho^3+\rho \\ 4\rho^2 & 4\rho^3+\rho & 4\rho^4+\rho^2+1 \end{bmatrix}$$

b) Find the LMMSE estimate of X_2 based on Y_0, Y_1, Y_2 and corresponding MSE

KALMAN FILTER PROBLEM!
Matrix equivalences: $A = \rho$, $C = 1$, $Var(W_t) = Q = 1$, $Var(V_t) = R = 1$
Initialize: $\hat{X}_{0|-1} = 0$; $\Sigma_{0|-1} = 4$

Update for $k = 0, 1, 2$:
$$\hat{X}_{k|k} = \hat{X}_{k|k-1} + \Sigma_{k|k-1}C^T(C\Sigma_{k|k-1}C^T + R)^{-1}(Y_k - E[Y_k])$$
$$= \hat{X}_{k|k-1} + \frac{\Sigma_{k|k-1}}{\Sigma_{k|k-1}+1}(Y_k - \hat{X}_{k|k-1})$$

$$\Sigma_{k|k} = \Sigma_{k|k-1} - \Sigma_{k|k-1}C^T(C\Sigma_{k|k-1}C^T + R)^{-1}C\Sigma_{k|k-1}$$
$$= \Sigma_{k|k-1} - \frac{\Sigma_{k|k-1}^2}{\Sigma_{k|k-1}+1} = \frac{\Sigma_{k|k-1}}{\Sigma_{k|k-1}+1}$$

PREDICT: For $k = 0, 1$, $\hat{X}_{k+1|k} = \rho\hat{X}_{k|k}$, $\Sigma_{k+1|k} = \rho^2\Sigma_{k|k} + 1$

Now compute: $\Sigma_{0|0} = \frac{4}{4+1} = 4/5$

$\Sigma_{1|0} = (4/5)\rho^2 + 1$

$\Sigma_{1|1} = \frac{4/5\rho^2 + 1}{4/5\rho^2 + 2}$

$\Sigma_{2|1} = \rho^2\left(\frac{4/5\rho^2+1}{4/5\rho^2+2}\right) + 1$

MSE! LMMSE will be $\hat{X}_{2|2}$
$$\Sigma_{2|2} = \frac{\rho^2\left(\frac{4/5\rho^2+1}{4/5\rho^2+2}\right)+1}{\rho^2\left(\frac{4/5\rho^2+1}{4/5\rho^2+2}\right)+2}$$

3.10 INNOVATION SEQUENCE

Let $(Y_1, Y_2, Y_3, X)^T$ be a zero-mean RV w/ $\Sigma = \begin{pmatrix} 2 & 1 & 1 & 2 \\ 1 & 2 & 1 & 2 \\ 1 & 1 & 2 & 2 \\ 2 & 2 & 2 & 4 \end{pmatrix}$

a) Let $\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3$ denote the innovations sequence (unnormalized).
Find A s.t. $\begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \tilde{Y}_3 \end{bmatrix} = A\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$

$\tilde{Y}_1 = Y_1$, $Var(\tilde{Y}_1) = 2$
$\tilde{Y}_2 = Y_2 - \hat{E}[Y_2|\tilde{Y}_1] = Y_2 - \frac{1}{2}Y_1$, $Var(\tilde{Y}_2) = 2 - \frac{1}{2} = 3/2$
$\tilde{Y}_3 = Y_3 - \hat{E}[Y_3|\tilde{Y}_1, \tilde{Y}_2] = Y_3 - \hat{E}(Y_3|\tilde{Y}_1) - E[Y_3|\tilde{Y}_2]$
$= Y_3 - \frac{\tilde{Y}_1}{2} - Cov(Y_3, \tilde{Y}_2)\frac{2}{3}\tilde{Y}_2 = Y_3 - \frac{\tilde{Y}_1}{2} - Cov(Y_3, Y_2 - \frac{Y_1}{2})(\frac{2}{3})(Y_2 - \frac{Y_1}{2})$
$= Y_3 - \frac{Y_1}{2} - \frac{1}{3}(Y_2 - \frac{Y_1}{2}) = Y_3 - \frac{1}{3}Y_2 - \frac{1}{3}Y_1$
$Cov(\tilde{Y}_3) = 2 + 4/9 + 2/9 - 4/3 = 4/3$

$$\text{So, } A = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ -1/3 & -1/3 & 1 \end{bmatrix}$$

b) Find the correlation matrix of $(\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3)^T$ and the cross covariance matrix $Cov(X, (\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3)^T)$.

$$Cov\left(\begin{bmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \\ \tilde{Y}_3 \end{bmatrix}\right) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3/2 & 0 \\ 0 & 0 & 4/3 \end{bmatrix}$$

Cross covariance given by
$Cov(X, (\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3)^T) = E[(X\tilde{Y}_1, X\tilde{Y}_2, X\tilde{Y}_3)]$
$= (2, 1, 2/3)$

c) Find constants a, b, c, minimizing $E[(X - a\tilde{Y}_1 - b\tilde{Y}_2 - c\tilde{Y}_3)^2]$
This is the LLSE estimator of X and we can compute it directly

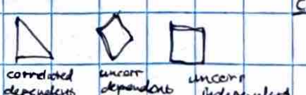
$$\hat{E}[X|\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3] = \hat{E}[X|\tilde{Y}_1] + \hat{E}[X|\tilde{Y}_2] + \hat{E}[X|\tilde{Y}_3]$$
$$= \tilde{Y}_1 + \frac{2}{3}\tilde{Y}_2 + \frac{1}{2}\tilde{Y}_3$$

↳ for some reason $a = \frac{E[X\tilde{Y}_1]}{Var(\tilde{Y}_1)} = \frac{2}{2} = 1$

$b = \frac{E[X\tilde{Y}_2]}{Var(\tilde{Y}_2)} = \frac{1}{3/2} = 2/3$

$c = \frac{E[X\tilde{Y}_3]}{Var(\tilde{Y}_3)} = \frac{2/3}{4/3} = 1/2$

$$\hat{E}[X|Y_1, \ldots, Y_n] = \hat{E}[X|\tilde{Y}_1, \ldots, \tilde{Y}_n]$$
$$= \bar{X} + \sum_{i=1}^{n} \hat{E}[X - \bar{X}|\tilde{Y}_i]$$
$$= \bar{X} + \sum_{i=1}^{n} Cov(X, \tilde{Y}_i)Cov(\tilde{Y}_i)^{-1}\tilde{Y}_i$$



CDF RIGHT-CONTINUOUS

ECS05 Final Review

- **random process**: an indexed collection of random variables $X = (X_t : t \in \mathbb{T})$ all on the same probability space $(\Omega, \mathcal{F}, P)$
  - Three ways to view:
    1. For each fixed t, X_t a function on Ω
    2. X a function on $\mathbb{T} \times \Omega$ w/ value $X_t(\omega)$ for $t \in \mathbb{T}$ and $\omega \in \Omega$
    3. For each fixed ω, $X_\omega(t)$ a function of t called the sample path corresponding to ω.
  - Notation: $\mu_X(t) = E[X_t]$ $\quad R_X(s,t) = E[X_s X_t]$

$$C_X(s,t) = Cov(X_s, X_t) = E[X_s X_t] - E[X_s]E[X_t]$$

$$F_{X,n}(x_1,t_1; \dots; x_n,t_n) = P\{X_{t_1} \le x_1, \dots, X_{t_n} \le x_n\}$$

- **second order random process**: a random process $(X_t : t \in \mathbb{T})$ s.t. $E[X_t^2] < +\infty$ $\forall t \in \mathbb{T}$
  - mean, correlation, and covar functions for 2nd order RP are all well-defined & finite
- **Gaussian random process**: random variables X_t making up X are jointly Gaussian.
  - 2 or more RPs are said to be jointly Gaussian if all the RVs comprising them are jointly Gaussian.
- **increment** of a random process $X = (X_t : t \in \mathbb{T})$ over an interval $[a,b]$ is the RV $X_b - X_a$.
- A random process has **independent increments** if for any positive integer n, any $t_0 < t_1 < \dots < t_n$ in $\mathbb{T}$. The increments $X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ are mutually independent.
- A random process is **martingale** if $E[X_t]$ finite for all t and for any $n \in \mathbb{Z}^+$ and $t_1 < t_2 < \dots < t_n < t_{n+1}$,

$$E[X_{t_{n+1}} | X_{t_1}, \dots X_{t_n}] = X_{t_n}, \text{ i.e. } E[(X_{t_{n+1}} - X_{t_n}) | X_{t_1}, \dots X_{t_n}] = 0$$

- **Brownian motion**: (aka a Wiener process) w/ parameter $\sigma^2 > 0$ is a random process $W = (W_t : t \ge 0)$ s.t.
  - B.0 $P\{W_0 = 0\} = 1$
  - B.1 W has independent increments
  - B.2 $W_t - W_s$ has the $N(0, \sigma^2(t-s))$ dist. for $t \ge s$
  - B.3 $P\{W_t$ is a continuous function of $t\} = 1$, or W is sample path continuous w/ probability 1.

STATISTICS: $\mu_W(t) = E[W_t] = E[W_t - W_0] = 0$

for $s \le t$: $R_W(s,t) = E[W_s W_t]$ ~~$= E[W_s \dots]$~~

$$= E[(W_s - W_0)(W_s - W_0 + W_t - W_s)]$$

$$= E[(W_s - W_0)^2] = \sigma^2 s$$

$$C_W(s,t) = R_W(s,t) = \sigma^2 (s \wedge t) \leftarrow \min(s,t)$$

- Brownian motion is a Martingale, Gaussian
- **counting function**: a function f on $\mathbb{R}_+$ that is nondecreasing, right-continuous, & integer-valued.
- $f(t)$ = # of counts observed during the interval $(0,t]$
- an increment $f(b) - f(a)$ is the number of counts in the interval $(a,b]$
- If t_i denotes the time of the ith count for $i \ge 1$, then f can be described by the sequence (t_i). (count times)
- If $u_1 = t_1$, $u_i = t_i - t_{i-1}$ for $i \ge 2$, f can be described by the sequence (u_i) (intercount times)

$$\boxed{f(t) = \sum_{n=1}^{\infty} I_{\{t \ge t_n\}} \quad t_n = \min\{t : f(t) \ge n\} \quad t_n = u_1 + \dots + u_n}$$

- **counting process**: a random process whose sample paths are counting functions w/ probability 1.
- **poisson process**: Let $\lambda \ge 0$. A Poisson process w/ rate λ is a random process $N = (N_t : t \ge 0)$ s.t.
  - N.1 N is a counting process
  - N.2 N has independent increments
  - N.3 $N(t) - N(s)$ has $Poi(\lambda(t-s))$ distribution for $t \ge s$
  - PROPERTIES:
    - intercount times $U_1, U_2, \dots$ are mutually independent, $Exp(\lambda)$ RVs.
    - for each $\tau > 0$, N_τ is a Poisson random variable with param $\lambda\tau$ and given $\{N_\tau = n\}$, the times of the n counts during $[0,\tau]$ are the same as n independent, Unif$[0,\tau]$ RVs reordered to be nondecreasing. I.e. for any $n \ge 1$, the conditional density of the first n count times $(T_1, \dots, T_n)$ given $\{N_\tau = n\}$ is:

$$f(t_1, \dots, t_n | N_\tau = n) = \begin{cases} \frac{n!}{\tau^n} & 0 < t_1 < \dots t_n < \tau \\ 0 & \text{else} \end{cases}$$

- **stationary**: random process $X = (X_t : t \in \mathbb{T})$ s.t. $\mathbb{T} = \mathbb{Z}$ or $\mathbb{R}$. Then X is stationary if for any $t_1, \dots, t_n$ and s in $\mathbb{T}$, the random vectors $(X_{t_1}, \dots X_{t_n})$ and $(X_{t_1+s}, \dots X_{t_n+s})$ have the same distribution (JOINT STATS UNAFFECTED BY TIME SHIFTS)

$$F_{X,n}(x_1, t_1; \dots x_n, t_n) = F_{X,n}(x_1, t_1+s; \dots; x_n, t_n+s)$$

  - $\mu_X(t)$, $E[X_t^2]$ do not depend on t
  - $E[X_{t_1} X_{t_2}] = E[X_{t_1+s} X_{t_2+s}]$ $\forall s$, so $R_X(t_1+s, t_2+s)$ does not depend on s.
- **wide sense stationary**: 2nd order random process s.t. $\mu_X(t) = \mu_X(s+t)$, $R_X(t_1,t_2) = R_X(t_1+s, t_2+s)$
  - For WSS processes, we can use the notation:

$$E[X_t] = \mu_X, \; E[X_{t_1} X_{t_2}] = R_X(t_1 - t_2)$$

$$C_X(t_1,t_2) = C_X(t_1 - t_2), \; C_X(\tau) = R_X(\tau) - \mu_X^2$$

X Gauss & WSS $\Longrightarrow$ X stationary

- **conditional independence of random vectors**: Let X, Y, Z be random vectors. Conditional independence of X and Z given Y is denoted $X - Y - Z$. If X, Y, Z discrete, $X - Y - Z$ holds if

$$P(X = i, Z = k | Y = j) = P(X = i | Y = j) P(Z = k | Y = j) \; \forall i,j,k$$

with $P\{Y = j\} > 0$. Equivalently $X - Y - Z$ if

$$P\{X = i, Y = j, Z = k\} P\{Y = j\} = P\{X = i, Y = j\} P\{Z = k, Y = j\}$$

$\forall i,j,k$ (intuitively Y serves as state)

Equivalently (again), $X - Y - Z$ if

$$P(Z = k | X = i, Y = j) = P(Z = k | Y = j) \; \forall i,j,k \text{ w/ } P\{X = i, Y = j\} > 0$$

- symmetric in X and Z ($X - Y - Z$ equiv to $Z - Y - X$)
- If X, Y, Z have a joint pdf, we can define $X - Y - Z$ as follows:

$$f_{XZ|Y}(x, z | y) = f_{X|Y}(x|y) f_{Z|Y}(z|y) \quad (f_Y(y) > 0)$$

OR

$$f_{Z|XY}(z | x,y) = f_{Z|Y}(z|y) \quad (f_{XY}(x,y) > 0)$$

- **Markov Process**: A random process $X = (X_t : t \in \mathbb{T})$ is a markov process if for any $t_1, \dots t_{n+1} \in \mathbb{T}$ w/ $t_1 < \dots < t_{n+1}$, the following conditional independence condition holds:

$$(X_{t_1}, \dots X_{t_n}) - X_{t_n} - X_{t_{n+1}}$$

random walk, Brownian motion, Poisson processes, recursive sequences

equivalently

$$(X_{t_1}, \dots X_{t_n}) - X_{t_n} - (X_{t_n}, \dots X_{t_{n+m}})$$

convergence of sequences of random variables

- **almost surely (a.s.)**: all RVs defined on the same probability space and $P[\lim_{n\to\infty} X_n = X] = 1$. Denoted $X_n \xrightarrow{a.s.} X$.
- **in probability**: all RVs defined on same prob space & for any $\epsilon > 0$, $\lim_{n\to\infty}(P[|X - X_n| \ge \epsilon]) = 0$. Denoted $X_n \xrightarrow{p} X$
- **mean square sense (m.s.)**: all RVs defined on same prob space, $E[X_n^2] < \infty \; \forall n$, and $\lim_{n\to\infty} E[(X_n - X)^2] = 0$. Denoted $X_n \xrightarrow{m.s.} X$
- **In distribution**: $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all continuity points of F_X. Denoted $X_n \xrightarrow{d} X$.

$$\begin{tikzcd} a.s. \arrow[dr] & & \\ & p \arrow[r] & d \\ m.s. \arrow[ur] & & \end{tikzcd}$$

if sequence dominated by single RV w/ finite 2nd moment

- **limits for cts time random processes**: Let $X = (X_t : t \in \mathbb{T})$ be a random process s.t. $\mathbb{T}$ is all of or an interval of $\mathbb{R}$. Fix $t_0 \in \mathbb{T}$. The process X_t has a limit Y at t_0:
  - a) in the m.s. sense if for any $\epsilon > 0$, $\exists \delta > 0$ s.t. $E[(X_s - Y)^2] < \epsilon$ w/ $s \in \mathbb{T}$, $|s - t_0| < \delta$. Equiv condition: $X_{s_n} \xrightarrow{m.s.} Y$ as $n \to \infty$ when $s_n \to t_0$
  - b) in probability if given $\epsilon > 0$, $\exists \delta > 0$ s.t. $P[|X_s - Y| \ge \epsilon] < \epsilon$ when $s \in \mathbb{T}$, $|s - t_0| < \delta$. Equiv condition: $X_{s_n} \xrightarrow{p} Y$ as $n \to \infty$ when $s_n \to t_0$
  - c) in distribution if for any continuity point c of F_Y and any $\epsilon > 0$ $\exists \delta > 0$ s.t. $|F_{X,1}(c,s) - F_Y(c)| < \epsilon$ whenever $s \in \mathbb{T}$ and $|s - t_0| < \delta$. Equivalent condition: $X_{s_n} \xrightarrow{d} Y$ as $n \to \infty$ whenever $s_n \to t_0$

$$[F_{X,1}(c,s) = P\{X_s \le c\}].$$

  - d) almost surely if $\exists$ an event F_{t_0} having prob. 1 s.t. $F_{t_0} \subset \{\omega : \lim_{s \to t_0} X_s(\omega) = Y(\omega)\}$.
- **continuity at a point**: for each $t_0 \in \mathbb{T}$ fixed, $X = (X_t : t \in \mathbb{T})$ is cts at t_0 in any one of the four senses (m.s, p, d, a.s) if $\lim_{s \to t_0} X_s = X_{t_0}$ in the corresponding sense
- **continuity for a whole random process**: A random process X_t is said to be
  - ms cts if it is ms cts at each t
  - cts in p if it is cts in p at each t
  - cts in d if it is cts in d. at each t
  - a.s cts at each t if a.s cts at each t
  - a.s sample-path cts if $F \subset \{X_t$ cts in $t\}$ for some event F w/ P(F)=1

a.s. sample path cts $\Longrightarrow$ a.s. cts at each t

- **m.s. continuity**: suppose $(X_t : t \in \mathbb{T})$ is a 2nd order process. The following are equiv:
  - i) R_X cts at all points of the form (t,t)
  - ii) R_X is cts over $\mathbb{T} \times \mathbb{T}$
  - iii) X m.s. cts

If X is m.s. cts, then the mean function, $\mu_X(t)$ is cts. If X WSS, the following are equivalent:

i) $R_X(\tau)$ cts at $\tau = 0$ ii) X is ms cts iii) $R_X(\tau)$ cts over all of $\mathbb{R}$

- **m.s. differentiable**: For each t fixed, the random process $X = (X_t : t \in \mathbb{T})$ is ms diff'able at t if the following limit exists:

$$\lim_{s \to t} \frac{X_s - X_t}{s - t} \text{ m.s.}$$

If this limit exists, it is the m.s. deriv. of X at t denoted X'_t. The whole RP X is said to be ms diff'able if it is ms diff'able at each t. It is said to be ms cts diff'able if it is ms diff'able & X' is ms cts.

- **ms differentiability properties**
  - a) If X m.s. diff'able, then μ_X is diff'able and $\mu'_X(t) = \mu_{X'}(t)$
  - b) If X m.s. diff'able, $R_{X'X} = \partial_1 R_X$ and $R_{XX'} = \partial_2 R_X$ and the autocorr of X' is given by $R_{X'} = \partial_1 \partial_2 R_X = \partial_2 \partial_1 R_X$
  - c) X is m.s. diff'able at t iff the following limit exists & is finite:

$$\lim_{s,s' \to t} \frac{R_X(s,s') - R_X(s,t) - R_X(t,s') + R_X(t,t)}{(s-t)(s'-t)}$$

  - d) X is m.s. cts'ly diff'able iff R_X, $\partial_2 R_X$, $\partial_1 \partial_2 R_X$ exist & are cts
  - e) If X WSS, X is m.s. cts'ly diff'able iff $R_X(\tau)$, $R'_X(\tau)$, and $R''_X(\tau)$ exist and are cts functions of τ. If X ms cts diff'able then X, X' are jointly WSS, X' has mean 0, and $R_{X'}(\tau) = -R''_X(\tau)$,

$$R_{X'X}(\tau) = R'_X(\tau), \; R_{XX'}(\tau) = R'_X(\tau)$$

  - f) If X WSS and ms diff'able, $R'_X(0)$ exists and $= 0$
  - g) If X ms diff'able Gaussian, then X and its derivative process X' are jointly Gaussian
- **product rule**: If X an m.s. diff'able RP and f a diff'able function, then product $Xf = (X(t)f(t) : t \in \mathbb{R})$ is ms diff'able and $(Xf)' = X'f + Xf'$
- **Riemann sum for RP**: given partition of $(a,b]$ $((t_0,t_1], \dots (t_{n-1}, t_n])$, sample points $v_k \in (t_{k-1}, t_k]$ for $1 \le k \le n$

$$\sum_{k=1}^{n} X_{v_k} h(v_k)(t_k - t_{k-1})$$

norm of partition: $\max_k |t_k - t_{k-1}|$

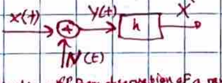
- **m.s. Riemann integral**: $\int_a^b X_t h(t) dt$ is said to exist and its value is the random variable I if the following is true. Given $\epsilon > 0$, $\exists \delta > 0$ s.t. $E[(\sum_{k=1}^{n} X_{v_k} h(v_k)(t_k - t_{k-1}) - I)^2] \le \epsilon$ when the norm of the partition is $\le \delta$.

i.e. sequence of Riemann sums m.s. converge to I as $m \to \infty$

If $X_t h(t)$ defined over whole real line, Riemann integrable over every bounded interval $[a,b]$, then

$$\int_{-\infty}^{\infty} X_t h(t) dt = \lim_{a,b \to \infty} \int_{a}^{b} X_t h(t) dt \text{ m.s.}$$

- **m.s. Riemann integration properties**: (assume $X_t h(t)$, $Y_t k(t)$ m.s. integrable over $[a,b]$)
  - a) The integral $\int_a^b X_t h(t) dt$ exists in the ms Riemann sense iff $\int_a^b \int_a^b R_X(s,t) h(s) h(t) ds dt$ exists as a 2D Riemann integral w/ aligned sampling. Exists if X m.s. piecewise cts over $[a,b]$ and h piecewise cts over $[a,b]$
  - b) $E\left[\int_a^b X_t h(t) dt\right] = \int_a^b \mu_X(t) h(t) dt$
  - c) $E\left[\left(\int_a^b X_t h(t) dt\right)^2\right] = \int_a^b \int_a^b R_X(s,t) h(s) h(t) ds dt$
  - d) $Var\left(\int_a^b X_t h(t) dt\right) = \int_a^b \int_a^b C_X(s,t) h(s) h(t) ds dt$
  - e) $E\left[\left(\int_a^b X_s h(s) ds\right)\left(\int_a^b Y_t k(t) dt\right)\right] = \int_a^b \int_a^b R_{XY}(s,t) h(s) k(t) ds dt$
  - f) $Cov\left(\int_a^b X_s h(s) ds, \int_a^b Y_t k(t) dt\right) = \int_a^b \int_a^b C_{XY}(s,t) h(s) k(t) ds dt$
  - g) $\int_a^b X_t h(t) + Y_t k(t) dt = \int_a^b X_t h(t) dt + \int_a^b Y_t k(t) dt$
- **Fundamental theorem of ms calculus**: Let X be a m.s. cts'ly diff'able random process. For $a < b$, $X_b - X_a = \int_a^b X'_t dt$ (ms Riemann integral)
- If X a Gaussian RP, X, all of its ms derivatives, and all its ms Riemann integrals of X of the form $I(a,b) = \int_a^b X_t h(t) dt$ that exist are jointly Gaussian.



E.g. Consider Estimating a RP from observation of a random process + noise. Assume X, N jointly WSS, mean 0, w/ known auto corr. functions. $R_{XN} = 0$. Observed process given by $Y = X + N$. $S_{XY} = S_X$, $S_Y = S_X + S_N$

So optimal filter: $\boxed{H(\omega) = \frac{S_{XY}(\omega)}{S_Y(\omega)}} = \frac{S_X(\omega)}{S_X(\omega) + S_N(\omega)}$

Associated MMSE: $\int_{-\infty}^{\infty} \left(S_X(\omega) - \frac{S_X(\omega)^2}{S_X(\omega) + S_N(\omega)}\right) \frac{d\omega}{2\pi} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_X(\omega) S_N(\omega)}{S_X(\omega) + S_N(\omega)} d\omega$

$$E[|X_t - \hat{X}_t|^2] = E[|X_t|^2] - E[|\hat{X}_t|^2] = \int_{-\infty}^{\infty} \left(S_X(\omega) - \frac{|S_{XY}(\omega)|^2}{S_Y(\omega)}\right) \frac{d\omega}{2\pi}$$

IF $S_X(\omega) = \frac{1}{1+\omega^2} \Longleftrightarrow R_X(\tau) = \frac{e^{-|\tau|}}{2}$, $S_N(\omega) = \frac{4}{4+\omega^2} \Longleftrightarrow R_N(\tau) = e^{-2|\tau|}$, $R_{XN} = 1$

$$H(\omega) = \frac{S_X(\omega)}{S_X(\omega) + S_N(\omega)} = \frac{\frac{1}{1+\omega^2}}{\frac{1}{1+\omega^2} + \frac{4}{4+\omega^2}} = \frac{4+\omega^2}{8+5\omega^2} = \frac{1}{5} + \frac{\frac{12}{5}}{8+5\omega^2}$$

$\frac{1}{5}\delta(t) \leftrightarrow \frac{1}{5}$ $\quad \forall \alpha$, $e^{-\alpha|t|} \leftrightarrow \frac{2\alpha}{\omega^2 + \alpha^2}$

Random Processes & Linear Systems and Spectral analysis

- The output $(Y_s : s \in \mathbb{R})$ of a linear system w/ impulse response function $h(s,t)$ and a random process input $(X_t : t \in \mathbb{R})$ is defined by

$$Y_s = \int_{-\infty}^{\infty} h(s,t) X_t dt$$

- The integral defining the output Y_s for fixed s exists iff the following Riemann integral exists & is finite: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h^*(s,\tau) h(s,t) R_X(t,\tau) dt d\tau$ (*)
- **Sufficient condition for Y_s to be well-defined**: R_X a bounded cts function, $h(s,t)$ cts in t w/ $\int_{-\infty}^{\infty} |h(s,t)| dt < \infty$
- **Mean function of the output**: $\mu_Y(s) = E\left[\int_{-\infty}^{\infty} h(s,t) X_t dt\right] = \int_{-\infty}^{\infty} h(s,t) \mu_X(t) dt$
- **cross corr b/w input & output**: $R_{YX}(s,\tau) = E\left[\int_{-\infty}^{\infty} h(s,t) X_t dt X_\tau^*\right]$

$$= \int_{-\infty}^{\infty} h(s,t) R_X(t,\tau) dt$$

- **corr of output**: $R_Y(s,u) = E\left[Y_s \left(\int_{-\infty}^{\infty} h(u,\tau) X_\tau d\tau\right)^*\right]$ (complex conj.)

$$= \int_{-\infty}^{\infty} h^*(u,\tau) R_{YX}(s,\tau) d\tau$$

$$= \int\int_{-\infty}^{\infty} h^*(u,\tau) h(s,t) R_X(t,\tau) dt d\tau$$

- A linear system is **bounded input bounded output (bibo) stable** if the output is bounded whenever the input is bounded. If system time invariant, bibo stability equivalent to condition $\int_{-\infty}^{\infty} |h(t)| dt < \infty$ (**)
- A linear system is **time invariant** if $h(s,t)$ depends on s,t only through $s-t$. We write $h(s-t)$ instead of $h(s,t)$ & defining relation becomes a convolution $Y = h * X$.
- If (**) holds and input X satisfies $|X_s| < L$ $\forall s$, then $Y = X * h$ satisfies

$$|y(t)| \le L \int_{-\infty}^{\infty} |h(t)| dt \quad \forall t$$

- If X WSS, then R_X bounded by $R_X(0)$, so if X WSS, ms cts and the linear system is LTI and bibo stable, (*) exists & is bounded by

$$R_X(0) \left(\int_{-\infty}^{\infty} |h(\tau)| d\tau\right)^2 < \infty$$

- Thus, the output of a LTI bibo stable system is well-defined in ms sense if the input is a stationary (WSS), ms cts process.
- **CONVOLUTIONS**: given f, g functions over $\mathbb{R}$, the convolution is the function $f * g$ which can be defined in the following ways:

$$f * g(t) = \int_{-\infty}^{\infty} f(s) g(t-s) ds = \int_{-\infty}^{\infty} f(t-s) g(s) ds$$

Equivalently for any real a,b:

$$f * g(a+b) = \int_{-\infty}^{\infty} f(a+s) g(b-s) ds$$

- SALIENT FEATURE: the integral of the product of f and g w/ the args of f and g ranging over $\mathbb{R}$ in such a way that the sum of the two args is held constant & = the value we are evaluating the conv. at
- Convolution is **commutative** and **associative**
- If X WSS and the system is LTI, X, Y jointly WSS with

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(t) dt, \; R_{YX} = h * R_X, \; R_Y = h * \tilde{h} * R_X \quad \tilde{h} = h^*(-t)$$

$$h * \tilde{h}(t) = \int_{-\infty}^{\infty} h(s) \tilde{h}(t-s) ds$$

$$= \int_{-\infty}^{\infty} h(s) h^*(s-t) ds$$

- The **Fourier transform** of a function g mapping $\mathbb{R}$ to $\mathbb{C}$ is formally defined by $\hat{g}(\omega) = \int_{-\infty}^{\infty} e^{-j\omega t} g(t) dt$
- **PROPERTIES OF FOURIER TRANSFORMS:**
  1. Linearity: $\widehat{ag + bh} = a\hat{g} + b\hat{h}$
  2. Inversion: $g(t) = \int_{-\infty}^{\infty} e^{j\omega t} \hat{g}(\omega) \frac{1}{2\pi} d\omega$
  3. Convolution to multiplication: $\widehat{g * h} = \hat{g}\hat{h}$, $\hat{g} * \hat{h} = 2\pi \widehat{gh}$
  4. Parseval's Identity: $\int_{-\infty}^{\infty} g(t) h^*(t) dt = \int_{-\infty}^{\infty} \hat{g}(\omega) \hat{h}^*(\omega) \frac{d\omega}{2\pi}$
  5. Transform of time reversal: $\hat{\tilde{h}} = \hat{h}^*$ (where $\tilde{h}(t) = h^*(-t)$)
  6. Differentiation to multiplication by $j\omega$: $\widehat{\frac{dg}{dt}}(\omega) = j\omega \hat{g}(\omega)$
  7. Pure sinusoid to delta function: For ω_0 fixed $\widehat{e^{j\omega_0 t}}(\omega) = 2\pi\delta(\omega - \omega_0)$
  8. Delta function to pure sinusoid: For t_0 fixed: $\widehat{\delta(t - t_0)}(\omega) = e^{-j\omega t_0}$
- Given an LTI system w/ impulse response h, its Fourier transform $H(\omega)$ is known as the **transfer function**
- $y = h * x$ becomes $\hat{y}(\omega) = H(\omega)\hat{x}(\omega)$ in frequency domain
- Given a WSS process $X = (X_t : t \in \mathbb{R})$, the FT of its correlation function R_X is called the **power spectral density** (S_X)
- If X and Y are jointly WSS, the FT of R_{XY} is denoted S_{XY} and called the **cross power spectral density function**

$$\left[S_{YX}(\omega) = H(\omega) S_X(\omega), \; S_Y(\omega) = |H(\omega)|^2 S_X(\omega)\right]$$

- S_X well defined if $\int_{-\infty}^{\infty} |R_X(t)| dt < \infty$
- $S_{YX} = S_{XY}^*$, taking $Y = X$, $R_X = \tilde{R}_X$, $S_X = S_X^*$
- $E[|X_t|^2] = R_X(0) = \int_{-\infty}^{\infty} S_X(\omega) \frac{d\omega}{2\pi}$ (power/total power of X)
- **Power of X in frequency interval $[a,b]$**: $E[|Y_t|^2] = \int_{-\infty}^{\infty} S_Y(\omega) \frac{d\omega}{2\pi}$

$$= \int_a^b S_X(\omega) \frac{d\omega}{2\pi}$$

$E[Y^2(t)] = R_Y(0)$

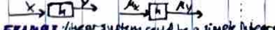
Mean ergodic: $\lim_{t\to\infty} \frac{1}{t} \int_0^t X_u du = \mu_X$ m.s. $\sim$ $\frac{1}{2T} \int_{-T}^{T} |C_{XX}(\tau)| d\tau \to 0$ suff

- $X_t(\omega)$ if bounded can always be expanded as a FS

$$\frac{1}{2T} \int_{-T}^{T} Y_t dt = E[Y_t] = E[X_{t+\tau} X_t] = R_{XX}(\tau)$$

$\to R_{XX}(\tau)$

Ergodic in autocorr w/ delay τ



EXAMPLE: linear systems could be a simple integrator from time 0 defined by

$$Y_s = \begin{cases} \int_0^s X_t dt & s \ge 0 \\ 0 & \text{else} \end{cases} \Longrightarrow h(s,t) = \begin{cases} 1 & s \ge t \ge 0 \\ 0 & \text{else} \end{cases}$$

Given Q, $p_{ij}^J = \begin{cases} -q_{ij}/q_{ii} & (\ne j) \\ 0, & i = j \end{cases}$

$$E[H_i^3] = \frac{-1}{q_{ii}} + \sum_j p_{ij} E[H_j^3]$$

$$= \frac{-1}{q_{ii}} + \sum_j \left(\frac{q_{ij}}{q_{ii}}\right) E[H_j^3]$$

$$\left[j\omega = s = 2\pi j f\right]$$

Practice Exam Solutions

Problem #1 (T/F)

a) For $X(t)$ a zero mean WSS process w/ autocorrelation function $R_X(\tau) = \max(1-|\tau|, 0)$, it is mean-squared continuous. TRUE Autocorrelation function is continuous at $\tau=0$

b) Consider WSS process $X(t)$ w/ mean 2, autocovar $K_X(\tau) = 8e^{-4|\tau|}$, that is input to a linear system with transfer function $H(s) = \frac{1}{(s+2)(s+2)}$. The output $Y(t)$ has mean $m_Y(t) = 4$ False. H is 1, the gain at 0 frequency times the input mean.

c) Assume a WSS process $X(t)$ has autocovariance $K_X(\tau) = 5e^{-|\tau|}$. Then for such a process

$$E[X(t)] = m_X = \lim_{T\to\infty} \int_{-T}^{T} X(s)\,ds$$ where the limit is interpreted in the m.s. sense.

FALSE: You must average over time which the right hand side fails along.

d) Consider a WSS process $V(t)$ that is input to a causal linear system w/ transfer function $H(s) = \frac{1}{s-1}$ to obtain output $Y(t)$. Then $Y(t)$ is also WSS.
FALSE. $H(s) = 1/s-1$ is unstable, as it was specified that this was a causal system ⟹ output will not be WSS

e) Suppose we have zero mean, mean-square continuous stochastic process $X(t)$ on the interval $[0,1]$ with autocovariance $K_X(t_1, t_2) = e^{-|t_1-t_2|}$. Let $f(t), g(t)$ be functions s.t. $f(t)=1$, $g(t) = (\sqrt{3})(1-2t)$ for $t\in[0,1]$. Note that

$$\int_0^1 f^2(t)dt = 1 \quad \int_0^1 g^2(t)dt = 1 \quad \int_0^1 f(t)g(t)dt = 0 \text{ (i.e. } f,g \text{ orthonormal)}$$

Define the random coefficients

$$C_1 = \int_0^1 f(t)X(t)dt \quad C_2 = \int_0^1 g(t)X(t)dt$$

Then the coefficients C_1, C_2 satisfy $E[C_1C_2] = 0$

TRUE. If $f(t), g(t)$ were eigenfunctions of the autocovariance, this would be simply true. We can easily verify $f(t)$ not an eigenfunction:

$$\int_0^1 e^{-|t_1-t_2|}dt_1 = \int_0^{t} e^{t_1-t}dt_1 + \int_t^1 e^{t-t_1}dt_1 = 1 - e^{-t} + 1 - e^{t-1}$$

The right hand side is not a constant (multiple of $f(t)$) as it would have to be for eigenfunctions. We can similarly check $g(t)$ is not an eigenfunction.
However, if we directly compute $E[C_1C_2]$ we get orthogonality

$$E[C_1C_2] = \int_0^1\int_0^1 3(1-2t_2)e^{-|t_1-t_2|}dt_1dt_2$$
$$= \int_0^1 3(1-2t_2)(2-e^{-t_2}-e^{t_2-1})dt_2$$
$$= 3[2-(1-e^{-1})-(1-e^{-1}) - 2 + 2(1-2e^{-1}) + 2e^{-1}] = 0 \checkmark$$

Problem #2: Constructing a process $M(t)$ that resembles a random walk except jumps will be different from ±1. Will be a discrete time process w/ $M(0)=0$ and

$$M(n+1) = M(n) + (n+1)J(n+1)$$

where $J(n)$ are independent, identically distributed Gaussian RVs w/ mean 0, var 1.
Closed form: $M(n) = \sum_{k=1}^{n} kJ(k)$

$M(t)$ will change values at each discrete time, but the jumps will be getting bigger on average, linearly w/ the time index. We can view this as a random walk with random, time-dependent jumps.

a) Compute $m_M(n) = E[M(n)]$ for a given integer n. $m_M(k) = E[\sum_{j=1}^{k} jJ(j)] = \sum_{j=1}^{k} jE[J(j)] = 0$

b) Compute $E[M(n)^2]$ (can leave answer as a sum.
$E[M(k)^2] = Var(M(k)) = \sum_{j=1}^{k} Var(jJ(j))$ (b/c $J(j)$ are uncorrelated.
$E[M(k)^2] = \sum_{j=1}^{n} j^2 = \frac{k(k+1)(2k+1)}{6}$

c) From the recursive equation for $M(n)$ in (1) above, compute a recursive equation for $m_M(n) = E[M(n)]$ and a recursive equation for its variance $\Sigma(n) = Var(M(n))$
Taking expectations of both sides, we get
$E[M(n+1)] = E[M(n)] + E[(n+1)J(n+1)] \Rightarrow m_M(n+1) = m_M(n)$
$\Sigma(n+1) = \Sigma(n) + (n+1)^2$ (b/c the two terms $M(n)$ and $(n+1)J(n+1)$ independent.

d) Is $M(n)$ IIP? YES. Let $m>n$, $M(m)-M(n)$ independent of $M(n)$ b/c
$M(m) - M(n) = \sum_{k=n+1}^{m} kJ(k)$

e) Compute the autocorrelation function $R_M(m,n) = E[M(m)M(n)]$. You can leave the answer as a sum.
Because of the IIP, if $m>n$, $E[M(n)M(m)] = E[M(n)(M(m)-M(n)+M(n))] = E[M(n)^2]$
If $m<n$, we get $R(n,m) = \sum^{\min(m,n)} k^2 = \frac{(2\min(m,n)+1)(\min(n,m)+1)\min(n,m)}{6}$

f) Is $M(n)$ a Gaussian process? Yes, Σ of Gaussian processes $J(k)$

Problem #3: A cts-time digital comm. system. Either nothing is transmitted (bit = 0, hypothesis H_0) or the signal $A_s(t)$ is transmitted (bit = 1, hypothesis H_1) where $s(t)$ is shown in the figure. Receiver observes $Y(t)$, $t\in[0,2]$, composed of the transmitted signal corrupted by additive Gaussian noise $W(t)$ w/ power spectral density $S_W(\omega) = 1$. The two possible signals are equally likely a priori:

$$Y(t) = \begin{cases} W(t) & \text{under hypothesis } H_0 \\ As(t) + W(t) & \text{under hypothesis } H_1 \end{cases}$$

a) Suppose $A=1$. What is a good orthonormal basis $\{\phi_i(t)\}$ to use in analyzing this problem?

b) Design a detector to achieve the minimum probability of error in detecting the correct hypothesis. Specify the required processing of the data $Y(t)$ and associated decision regions.

c) What is the probability of error for your detector in b) in terms of $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-x^2/2}dx$
Boundary of region for deciding H_1 in terms of the coefficient c_1 is $\sqrt{2}/2$ standard deviations away from the mean under $H(0)$ (0) so $P_{err} = Q(\sqrt{2}/2)$

d) Suppose A a Gaussian RV independent of time and $W(t)$, $A\sim N(0,1)$. What is the optimal MPE ~~detector~~ receiver? Specify required processing of $Y(t)$, decision regions.
Temporal dependence hasn't changed so can use the same temporal basis function. Reduce problem to that of detecting the signal based on coefficient c_1

$$c_1 = \int_0^2 Y(t)\phi_1(t)dt$$

Under H_1: $c_1 = \int_0^2 As(t)\phi_1(t)dt + \int_0^2 W(t)\phi_1(t)dt = A\sqrt{2}+V$, $V\sim N(0,1)$, so $c_1\sim N(0,3)$

Under H_0, $c_1 \sim N(0,1)$, So optimal detector: $\frac{\frac{1}{\sqrt{6\pi}}e^{-c_1^2/6}}{\frac{1}{\sqrt{2\pi}}e^{-c_1^2/2}} \gtrless_{H_0}^{H_1} 1$

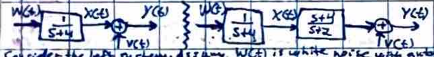
So, larger magnitudes of c_1 support H_1, which has larger covar.

$$\frac{c_1^2}{2} - \frac{c_1^2}{6} + \ln\left(\frac{1}{\sqrt{3}}\right) = \frac{c_1^2}{3} - \frac{1}{2}\ln(3) \gtrless_{H_0}^{H_1} 0$$

$$c_1^2 \gtrless_{H_0}^{H_1} \frac{1}{2}\ln(3)$$

e) For your detector in part d) write an expression for $P_r(err)$.
Let $z = \sqrt{1/2 \ln(3)}$ be the threshold on the magnitude of c_1 to decide b/w H_0, H_1. Under H_0, $c_1\sim N(0,1)$, so $P_F = 2Q(z)$
Under H_1, $c_1\sim N(0,3)$, which has standard deviation $\sqrt{3}$, so threshold is $z/\sqrt{3}$ standard deviations away from the mean 0, and $P_D = 2Q(z/\sqrt{3})$. $P_r(err)$ is the average of these w/ prior prob 1/2 for each hypothesis: $P_{error} = Q(z/\sqrt{3}) + Q(z)$

Problem #4



Consider the left system. Assume $W(t)$ is white noise with autocovariance $K_W(t) = \delta(t)$. The input process $V(t)$ is also white noise independent of $W(t)$ w/ autocov $K_V(\tau) = \frac{1}{9}\delta(\tau)$.

a) Compute the power spectral densities for $S_X(\omega)$, $S_Y(\omega)$.
$S_X(\omega) = \frac{1}{\omega^2+16}$ $S_Y(\omega) = S_X(\omega) + S_V(\omega) = \frac{1}{\omega^2+16} + 1/9 = \frac{(\omega^2+25)}{9(\omega^2+16)}$

b) Compute the cross spectral power density for $S_{YX}(\omega j)$: $S_{YX}(j\omega) = S_X(\omega) = \frac{1}{\omega^2+16}$

c) Find the noncausal linear filter $H_{nc}(j\omega)$ that minimizes the MSE in estimating $X(t)$ based on $Y(t)$. $H_{nc}(j\omega) = \frac{S_X(j\omega)}{S_Y(j\omega)} = \frac{9}{25+\omega^2}$

d) Find the variance of the error in the noncausal estimate of $X(t)$ based on observing $Y(t)$
$S_E(\omega) = S_X(\omega) - |S_{YX}(\omega)|^2\frac{1}{S_Y(\omega)} = \frac{1}{16+\omega^2} - \frac{1}{(16+\omega^2)^2}\frac{9(\omega^2+16)}{(\omega^2+25)}$
$= \frac{\omega^2+25-9}{(16+\omega^2)(\omega^2+25)} = \frac{1}{\omega^2+25}$
$K_E(\tau) = \mathcal{F}^{-1}\left(\frac{1}{25+\omega^2}\right) = 0.1e^{-5|\tau|}$ $E[\varepsilon^2(t)] = K_E(0) = 0.1$

e) Consider the right system, assume $V(t)$'s autocovar is $K_V(\tau) = \frac{1}{5}\delta(t)$. Find the non-causal linear filter $H_{nc}(j\omega)$ that minimizes the MSE in estimating $X(t)$ based on the process $Y(t)$.
$S_X(\omega) = \frac{1}{\omega^2+16}$, $S_Y(\omega) = \frac{1}{(\omega^2+4)} + \frac{1}{5} = \frac{\omega^2+9}{5(\omega^2+4)}$
$S_{YX}(j\omega) = \frac{4-j\omega}{2-j\omega}S_X(\omega) = \frac{4-j\omega}{2-j\omega}\frac{1}{16+\omega^2} = \frac{1}{(2-j\omega)(4+j\omega)}$
$H_{nc}(j\omega) = \frac{\frac{1}{(2-j\omega)(4+j\omega)}}{\frac{9+\omega^2}{5(4+\omega^2)}} = \frac{1}{(2-j\omega)(4+j\omega)}\frac{5(4+\omega^2)}{9+\omega^2} = \frac{10+5j\omega}{(4+j\omega)(9+\omega^2)}$

f) Compute variance of error of the non-causal estimate of $X(t)$ from e)
$S_E(\omega) = S_X(\omega) - |S_{YX}(\omega)|^2\frac{1}{S_Y(\omega)}$
$= \frac{1}{\omega^2+16} - \frac{1}{(4+\omega^2)(16+\omega^2)}\frac{5(\omega^2+4)}{\omega^2+9}$
$= \frac{1}{\omega^2+16}\left[1 - \frac{5}{\omega^2+9}\right]$
$= \frac{\omega^2+4}{(\omega^2+16)(\omega^2+9)} = \frac{A}{\omega^2+16} + \frac{B}{\omega^2+9}$ $A+B=1$, $9A+16B=4$, $B=-5/7$, $A=12/7$
$K_E(\tau) = \frac{12}{56}e^{-4|\tau|} - \frac{5}{42}e^{-3|\tau|}$
$E[\varepsilon^2(t)] = K_E(0) = \frac{12}{56} - \frac{5}{42} = \frac{2}{21}$

HW #8 Solutions

8.1 Suppose X a stationary Markov process w/ mean 0, state space $\{-1,1\}$, and transition rate matrix $Q = \begin{pmatrix} -\alpha & \alpha \\ \alpha & -\alpha \end{pmatrix}$, $\alpha\geq 0$. $\alpha = 0$ a possible case.

a) Find $R_X(\tau)$.
$E[X_t] = 0\ \forall t$. X_t takes values $\{-1,1\}$, distrib. of X_t for any t must be $\pi_i = 0.5$ for $i\in\{-1,1\}$. Solving Kolmogorov eqs. yields the distribution of the process at a time $t\geq 0$ for any given initial distribution:

$$\frac{\partial\pi_j(t)}{\partial t} = \sum_{\substack{i\in S\\ i\neq j}}\pi_i(t)q_{ij} - \sum_{\substack{i\in S\\ i\neq j}}\pi_j(t)q_{ji} \Rightarrow \frac{\partial\pi(t)}{\partial t} = \pi(t)Q$$

$$\frac{\partial\pi_1(t)}{\partial t} = -\alpha\pi_1(t) + \alpha\pi_2(t) \quad \pi_1(t) = 1 - \pi_2(t)$$
$$= -\alpha\pi_1(t) + \alpha(1-\pi_1(t))$$
$$= \alpha - 2\alpha\pi_1(t)$$

$H(t)$ 2x2
Top row: $\pi(t)$ for IC $\pi(0) = (1,0)$
Bottom: $\pi(t)$ for IC $\pi(0) = (0,1)$

$$\pi_1(t) = \pi_1(0)e^{-2\alpha t} + \int_0^t e^{-2\alpha(t-s)}\alpha\,ds$$
$$= \pi_1(0)e^{-2\alpha t} + \alpha\left(\tfrac{1}{2\alpha}\right)(1-e^{-2\alpha t})$$

So
$$\pi(t) = \pi(0)e^{-2\alpha t} + \left(\tfrac{1}{2}, \tfrac{1}{2}\right)(1-e^{-2\alpha t})$$
Transition probs: $p_{ij}(\tau) = \begin{cases} 0.5(1+e^{-2\alpha\tau}) & i=j \\ 0.5(1-e^{-2\alpha\tau}) & i\neq j\end{cases}$
So, for $\tau\geq 0$

$$R_X(\tau) = E[X(\tau)X(0)] = \sum_{i\in\{-1,1\}}\sum_{j\in\{-1,1\}} P[X(0)=i, X(\tau)=j]\,ij$$
$$= \sum\sum \pi_i p_{ij}(\tau)ij$$
$$= \frac{1}{4}\left[(1+e^{-2\alpha\tau}) + (1+e^{-2\alpha\tau}) - (1-e^{-2\alpha\tau}) - (1-e^{-2\alpha\tau})\right]$$
$$= e^{-2\alpha\tau}$$
$\forall t$, $R_X(\tau) = e^{-2\alpha|\tau|}$

b) For what values of $\alpha\geq 0$ is X m.s. cts? Diffable? Ergodic? Mean (m.s)
Continuous $\forall\alpha$ b/c R_X cts
Diffable for $\alpha = 0$, b/c R_X is twice ctsly diffable in a neighborhood of 0 iff $\alpha = 0$
Ergodic for $\alpha>0$, since $\lim_{\tau\to\infty}R_X(\tau) = 0$ for $\alpha>0$, $\lim_{\tau\to\infty}R_X(\tau)\neq 0$ if $\alpha=0$

8.2 Suppose $Z = (Z_t: 0\leq t\leq T)$ has the form $Z_t = \sum_{n=1}^{N}X_n\xi_n(t)$ s.t. the functions $\xi_1,\ldots,\xi_N$ are orthonormal over $[0,T]$, vector $X = (X_1,\ldots,X_N)^T$ has a correlation matrix K ($\det(K)\neq 0$). We say Z has rank N. Suppose K not diagonal. Describe the Karhunen-Loeve expansion of Z (describe an orthonormal basis $(\phi_n, n\geq 1)$, and eigenvals for the KL expansion of X.
K positive def, so can be factorized as $K = U\Lambda U^T$, where Λ diagonal w/ nonneg. diagonal entries $\lambda_1,\ldots,\lambda_N$, U orthonormal, $U^{-1} = U^T$. By given assumption,

$$R_Z(s,t) = \sum_{i=1}^{N}\sum_{j=1}^{N} K_{ij}\xi_i(s)\xi_j(t) = [\xi_1(s)|\cdots|\xi_N(s)]\,U\Lambda U^T\begin{bmatrix}\xi_1(t)\\ \xi_N(t)\end{bmatrix}$$

Let $[\phi_1|\cdots|\phi_N] = [\xi_1|\cdots|\xi_N]U$, or equivalently, $\phi_j(t) = \sum_{i=1}^{N}\xi_i(t)u_{ij}$, we have

$$R_Z(s,t) = [\phi_1(s)|\cdots\phi_N(s)]\Lambda\begin{bmatrix}\phi_1(t)\\ \phi_N(t)\end{bmatrix} = \sum_{i=1}^{N}\lambda_i\phi_i(s)\phi_i(t).$$

Also $\int_0^T \phi_i(t)\phi_j(t)dt$
$$= \int_0^T\sum_{a=1}^{N}\sum_{b=1}^{N}\xi_a(t)u_{ai}\xi_b(t) = \sum_{a=1}^{N}\sum_{b=1}^{N}I_{\{a=b\}}u_{ai}u_{bj} = \sum_{a=1}^{N}u_{ai}u_{aj}I_{\{i=j\}}$$

So, $\phi_1,\ldots,\phi_N$ are orthonormal eigenfunctions for R_Z, $\lambda_1,\ldots$ the eigenvalues. The jth coordinate in the K-L expansion of Z can be expressed in terms of X by
$\langle Z,\phi_j\rangle = \int_0^T Z_t\phi_j(t)dt = \sum_{i=1}^{N}\int_0^T Z_t\xi_i(t)u_{ij}dt = \sum_{i=1}^{N}X_iu_{ij}$. Equivalently, the vector of coordinates for the K-L expansion of Z is given by U^TX. In the case of a finite rank process, finding the K-L expansion is equivalent to diagonalizing the correlation matrix K of the random vector of coefficient variables.

8.3 Let $X = (X_t: 0\leq t\leq T)$ be a random process given by $X(t) = AB\sin\left(\frac{\pi t}{T}\right)$, where A, T pos. constants, $B\sim N(0,1)$.

a) Expected total energy of X: $E\left[\int_0^T X_t^2dt\right] = A^2E[B^2]\int_0^T\sin^2\left(\frac{\pi t}{T}\right)dt = A^2\int_0^T\frac{1-\cos\left(\frac{2\pi t}{T}\right)}{2}dt$
$= A^2T/2$

b) Mean & covar functions of X?
$E[X_t] = AE[B]\sin\left(\frac{\pi t}{T}\right) = 0$ $C_X(s,t) = A^2\sin\left(\frac{\pi s}{T}\right)\sin\left(\frac{\pi t}{T}\right)$, $0\leq s,t\leq T$

c) Describe KL expansion of X. Suppose φ an eigenfunction of C_X for some $\lambda\neq 0$. Then
$\lambda\varphi(s) = \int_0^T A^2\sin(\pi s/T)\sin(\pi t/T)\varphi(t)dt = \sin(\pi s/T)\underbrace{\int_0^T A^2\sin(\pi t/T)\varphi(t)dt}_{\text{constant}}$
So φ proportional to $\sin(\pi s/T)$. Thus
$\varphi_1(t) = \sqrt{\frac{2}{T}}\sin\left(\frac{\pi t}{T}\right)$ $0\leq t\leq T$, $\lambda_1 = \frac{A^2T}{2}$
KL expansion of X is $X_t = \langle X,\varphi_1\rangle\varphi_1(t)$ where
$\langle X,\varphi_1\rangle = \int_0^T AB\sin\left(\frac{\pi t}{T}\right)\sqrt{\frac{2}{T}}\sin\left(\frac{\pi t}{T}\right)dt = AB\sqrt{\frac{T}{2}}$
Coordinates of X relative to $\varphi_1, \varphi_2, \ldots$ given by $X \leftrightarrow \left(AB\sqrt{\frac{T}{2}}, 0, 0, \ldots\right)$

d) Let $N = (X_t: 0\leq t\leq T)$ be a Gauss white noise process independent of X w/ $R_N(t) = \sigma^2\delta(t)$, $Y = X+N$. Same basis functions for X can be used for KL expansions of N, Y. $N_1 = \langle N,\varphi_1\rangle$, $Y_1 = \langle Y,\varphi_1\rangle$. $Y_1 = X_1 + N_1$. Find $E[B|Y_1]$ & resulting MSE
Coordinates of N given by $N\leftrightarrow(N_1, N_2, \ldots)$ where N_i are $N(0,\sigma^2)$, coordinates of Y given by $Y\leftrightarrow\left(N_1 + AB\sqrt{\frac{T}{2}}, N_2, N_3, \ldots\right)$
Random vars B, Y_1 are mean 0 so $E[B|Y_1] = \frac{Cov(B,Y_1)}{Var(Y_1)}Y_1 = \frac{(A\sqrt{T/2})Y_1}{\sigma^2 + \frac{A^2T}{2}}$
$= \frac{A}{\sigma^2 + \left(\frac{A^2T}{2}\right)}\int_0^T\sin\left(\frac{\pi t}{T}\right)Y_t\,dt$

$MMSE = Var(B) - \frac{Cov(B,Y_1)^2}{Var(Y_1)} = 1 - \frac{\frac{A^2T}{2}}{\sigma^2 + \frac{A^2T}{2}}$

8.4 Let $X(t)$ be a Gaussian random process, WSS, w/ mean $E[X(t)] = 1$ and power spectral density function $S_X(\omega)$ in radians given by $S_X(\omega) = \frac{3}{1-\omega^2} + 2\pi\delta(\omega)$

a) Compute the autocovar $K_X(t)$
To find $K_{XX}(t) = R_{XX}(t) - \mu_X^2$ we need $R_{XX}(\tau)$, $\mathcal{F}^{-1}(S_{XX}(\omega))$.
Tables tell us $R_X(\tau) = e^{-|\tau|} + 1$, $K_X(\tau) = R_X(\tau) - 1 = e^{-|\tau|}$

b) Is $X(t)$ ergodic in autocorrelation?
$\int_{-\infty}^{\infty}|K_X(\tau)|d\tau = \int_{-\infty}^{\infty}|e^{-|\tau|}|d\tau = 2\int_0^\infty e^{-\tau}d\tau = 2 < \infty$
Gaussian + absolutely integrable autocovar ⟹ totally ergodic

c) Is X mean square continuous? Diffable?
MSC: show $\lim_{t\to t_0}E[(X(t)-X(t_0))^2] = 0$ $E[(X(t)-X(t_0))^2] = 2R_{XX}(0) - 2R_{XX}(t-t_0) = 2-2e^{-|t-t_0|}$
The limit is 0, so cts! (also $R_X(\tau)$ cts at $\tau=0$) ⟹ MSC
MSD: NO b/c $R_X(\tau)$ is not twice diffable at $\tau = 0$

d) Define $Y(t) = \frac{d}{dt}X(t)$. Using generalized functions if necessary, find mean and autocorr of $Y(t)$. Is $Y(t)$ WSS? $E[Y(t)] = E[\frac{d}{dt}X(t)] = \frac{d}{dt}E[X(t)] = \frac{d}{dt}1 = 0$
$K_Y(t) = \frac{d^2}{dt^2}K_X(t) = -\frac{d^2}{dt^2}e^{-|t|}$ $\frac{d}{dt}e^{-|t|} = \begin{cases} -e^{-|t|} & t>0 \\ e^{-|t|} & t<0\end{cases}$

$\frac{d}{dt}e^{-|t|} = e^{-|t|}[1-2u(t)]$, where $u(t)$ is the unit step function.
$\frac{d^2}{dt^2}e^{-|t|} = e^{-|t|}[1-2u(t)]^2 - 2\delta(t)e^{-|t|} = e^{-|t|}[1-2\delta(t)]$
Subbing: $K_Y(t) = -\frac{d^2}{dt^2}K_X(t) = e^{-|t|}[2\delta(t)-1] = 2\delta(t) - e^{-|t|}$
Yes WSS (constant mean, autocorr dependent on time diff only)

8.5 Consider LTI system w/ input $u(t)$, output $x(t)$. The diff. equation of the system: $\frac{d}{dt}x(t) = -x(t) + 3u(t)$, transfer function $H(s) = \frac{3}{s+1}$

a) Assume $u(t)$ is a stochastic process, WSS, mean $\mu_u = 5$, $K_u(\tau) = 5\delta(t)$. Find expression for the mean of the output $x(t)$.
$\mu_x = H(0)\mu_u = 15$

b) Find expression for power spectral density of $x(t)$, $S_x(\omega)$
$R_u(\tau) = \mu_u^2 + K_u(t) = 25 + 5\delta(t)$
$S_x(f) = H(2\pi jf)H(-2\pi jf)S_u(f) = \frac{9}{1+(2\pi f)^2}(25+5\delta(f))$

c) Define $y(t) = x(t) + v(t)$, where $v(t)$ a random process independent of $u(t)$ w/ 0 mean, $S_v(\omega) = 6$. Find $S_y(\omega)$.
$[S_y(f) = S_x(f) + S_v(f) = S_x(f) + 3 = 3 + \frac{9}{1+(2\pi f)^2}(25+5\delta(f))]$

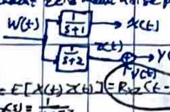
d) Suppose $z(t) = y(t) + x(t)$. Compute $S_z(\omega)$.
$z(t) = y(t) + x(t) = 2x(t) + v(t)$
$S_z(f) = 4S_x(f) + S_v(f) = 3 + \frac{36}{1+(2\pi f)^2}(25+5\delta(f))$

8.6 Assume input to an LTI system w/ impulse response $h(t) = e^{-t}u(t)$ is a zero mean Gauss random process w/ $S_X(\omega) = \eta/2$, ω in rad/sec.

a) Find $S_Y(\omega)$
$H(s) = \frac{1}{s+1}$, $S_Y(f) = |H(2\pi f)|^2S_X(f) = \frac{\eta}{2}\frac{1}{4\pi^2f^2+1} = \frac{\eta}{2}\frac{1}{\omega^2+1}$

b) Find $E[Y^2(t)]$ (avg power of output)
transform back to autocor: $K_Y(\tau) = \frac{\eta}{4}e^{-|\tau|}$
$E[Y^2(t)] = K_Y(0) = \frac{\eta}{4}$

8.8 Consider $W(t)$, $V(t)$ independent zero mean noise processes w/ $E[W(t)W(\tau)] = \delta(t-\tau)$, $E[V(t)V(\tau)] = \frac{1}{5}\delta(t-\tau)$



a) Compute $R_{XY}(t)$
$R_{XY}(\tau) = E[X(t)Y(t+\tau)]$
$Y(t) = Z(t) + V(t)$
$R_{XY}(t-s) = E[X(t)Y(s)] = E[X(t)Z(t)] = R_{XZ}(t-s)$
Note $Z(s) = \frac{1}{s+2}W(s)$, so $S_Z(s) = \frac{1}{4-s^2}$
$X(s) = \frac{1}{s+1}W(s)$, $S_X(s) = \frac{1}{1-s^2}$
→ $Z(s) = \frac{s+1}{s+2}W(s)$
$S_{ZX}(s) = S_{YX}(s) = \frac{-s+1}{-s+2}S_X(s) = \frac{-s+1}{-s+2}\frac{1}{1-s^2} = \frac{1}{(1+s)(2-s)}$
$S_{ZX}(s) = \frac{1/3}{2-s} + \frac{1/3}{1+s}$
$R_{ZX}(\tau) = R_{YX}(\tau) = e^{-\tau}u(\tau) + e^{2\tau}u(-\tau)$

b) Compute $S_{XY}(\omega)$
$\tau = s-t$ $S_{XY}(s) = S_X(s)\frac{s+1}{s+2} = \frac{1}{(1-s)(s+2)}$

c) Determine NC Wiener filter for estimating $X(t)$ given $Y(t)$.
$S_Y(s) = S_Z(s) + S_V(s)$
$= \left(\frac{1}{2+s}\right)\left(\frac{1}{2-s}\right)S_W(s) + S_V(s) = \left(\frac{1}{2+s}\right)\left(\frac{1}{2-s}\right)\cdot 1 + \frac{1}{5} = \frac{1}{(4-s^2)} + \frac{1}{5}$
$= \frac{9-s^2}{5(4-s^2)}$
Then $H(s) = \frac{S_{YX}(s)}{S_Y(s)} = \frac{1}{(2-s)(s+1)}\frac{5(2-s)(2+s)}{(9-s^2)} = \frac{5(s+2)}{(s+1)(9-s^2)}$

d) Compute MSE error for NC Wiener filter.
$S_E(s) = S_X(s) - \frac{S_{YX}(s)S_{YX}(-s)}{S_Y(s)}$
$= \left(\frac{1}{1+s}\right)\left(\frac{1}{1-s}\right)S_{WW}(s) - \left(\frac{1}{(2-s)(1-s)}\right)\left(\frac{1}{(2+s)(1-s)}\right)\frac{5(s-2)(2+s)}{(9-s^2)}$
$= \frac{1}{1-s^2} - \frac{5}{(1-s^2)(9-s^2)} = \frac{4-s^2}{(1-s^2)(9-s^2)}$
$S_e(j\omega) = \frac{4+\omega^2}{(1+\omega^2)(9+\omega^2)} = \frac{A}{1+\omega^2} + \frac{B}{9+\omega^2}$ $9A+B=4$, $A+B=1$ ⟹ $A = 3/8$, $B = 5/8$
$R_e(\tau) = \frac{3}{16}e^{-|\tau|} + \frac{5}{48}e^{-3|\tau|}$ $R_e(0) = 7/24$

8.11 Consider 4 equally likely signals $s_i(t)$: $s_1(t) = r(t)$, $s_2(t) = \frac{1}{3}r(t)$, $s_3(t) = -\frac{1}{3}r(t)$, $s_4(t) = -r(t)$, where $r(t) = \sqrt{2/T}\sin(2\pi t/T)$. Signal under H_i consists of $Y(t) = s_i(t) + W(t)$, $R_W(t) = 2\delta(t)$. $\int_0^T r^2(t)dt = 1$.

a) What is the MPE decision rule?
Use a basis expansion where $f_1(t) = r(t)$, $f_i(t)$, $i>1$ orthonormal functions ⊥ $r(t)$. Coeff expansions:

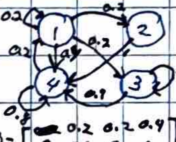
$c_1 = \int_0^T Y(t)f_1(t)dt = \begin{cases} N(1,2) & H_1 \\ N(1/3,2) & H_2 \\ N(-1/3,2) & H_3 \\ N(-1,2) & H_4\end{cases}$

$c_n = \int_0^T Y(t)f_n(t)dt \sim N(0,2)\ \forall H_i$, $n\geq 2$
Opt. decision rule: obtain c_1 & make decision as scalar

$$u(c_1) = \begin{cases} H_1 & c_1\geq 2/3 \\ H_2 & 0\leq c_1 < 2/3 \\ H_3 & -2/3\leq c_1 < 0 \\ H_4 & c_1 < -2/3\end{cases}$$

b) What is $P_r(err)$ in terms of $Q(x)$
H_1: threshold $(1/3\sqrt{2})$ s.d.s away from mean.
$P(e|H_1) = Q\left(\frac{1}{3\sqrt{2}}\right)$
$P_e = P(e|H_i)P(H_i) = \frac{3}{2}Q(1/3\sqrt{2})$
λ: 2 decreases, error decreases!

HW #7 Consider a cts time homogeneous Markov chain w/ states $\{0,1,2,3\}$. Arrival rate 3 customers/minute, service rate 2 cust/min.



$$Q = \begin{bmatrix} [redacted] & 0.2 & 0.2 & 0.4 \\ 0 & -1 & 0 & 1 \\ & & & 0.9 \end{bmatrix}$$

a) Assume store starts empty, what is the expected time until store is full?

$Q_{23}\pi_2 = Q_{32}\pi_3$

For cts Markov chains $Q^T\pi = 0$

HW #4

An ECS05 student is trying to decide whether to undergo an operation and consults 3 experts, who each give him a binary recommendation. Expert 1 is conditionally correct w/ a probability of p_1, expert 2 is conditionally correct w/ a probability of p_2, expert 3 is conditionally correct w/ probability p_3. The student wants to find an optimal way of combining the answers of the experts to maximize the probability of making a correct overall decision. The a priori probability of being a good candidate is $Pr(yes) = 1/2$. Experts are indep.

a) Formulate the detection problem by identifying the observation space and the probability associated with each observation conditioned on the hypotheses $H_0 = no$, $H_1 = yes$.

$[y_1, y_2, y_3]$	$Pr(Y=y\mid H_1)$	$Pr(Y=y\mid H_0)$
0 0 0	$(1-p_1)(1-p_2)(1-p_3)$	$p_1p_2p_3$
0 0 1	$(1-p_1)(1-p_2)p_3$	$p_1p_2(1-p_3)$
0 1 0	$(1-p_1)p_2(1-p_3)$	$p_1(1-p_2)p_3$
0 1 1	$(1-p_1)p_2p_3$	$p_1(1-p_2)(1-p_3)$
1 0 0	$p_1(1-p_2)(1-p_3)$	$(1-p_1)p_2p_3$
1 0 1	$p_1(1-p_2)p_3$	$(1-p_1)p_2(1-p_3)$
1 1 0	$p_1p_2(1-p_3)$	$(1-p_1)(1-p_2)p_3$
1 1 1	$p_1p_2p_3$	$(1-p_1)(1-p_2)(1-p_3)$

$$\mathcal{L}(y) = \frac{P(y\mid H_1)}{P(y\mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P(H_0)}{P(H_1)}\frac{(C_{10}-C_{00})}{(C_{01}-C_{11})}$$

b) Find the optimal decision rule that minimizes the probability of making a mistake and the associated prob. of error for the following two cases

i) $p_1=0.8$, $p_2=0.6$, $p_3=0.7$

$$\mathcal{L}(y) = \frac{P(y\mid H_1)}{P(y\mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1 \Rightarrow \mathcal{L}(y) = \prod_{i=1}^{3}\left(\frac{p_i}{1-p_i}\right)^{(2y_i-1)} \underset{H_0}{\overset{H_1}{\gtrless}} 1$$

$$Pr(err) = Pr(yes\mid H_0)P(H_0) + Pr(no\mid H_1)P(H_1) = \tfrac{1}{2}[Pr(yes\mid H_0) + Pr(no\mid H_1)]$$

$[y_1, y_2, y_3]$	$Pr(Y=y\mid H_1)$	$Pr(Y=y\mid H_0)$	Decision
0 0 0	0.024	0.336	NO
0 0 1	0.056	0.144	NO
0 1 0	0.036	0.224	NO
0 1 1	0.084	0.096	NO
1 0 0	0.096	0.084	YES
1 0 1	0.224	0.036	YES
1 1 0	0.144	0.056	YES
1 1 1	0.336	0.024	YES

$\boxed{\text{Rule: Use Expert \#1}}$

$\boxed{Pr(err) = 0.2 = 1 - p_1}$

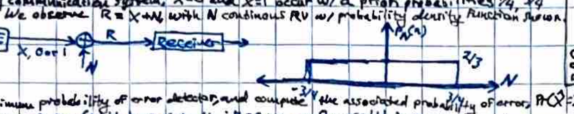
ii) when $p_1 = 0.75$, $p_2 = 0.6$, $p_3 = 0.7$

$[y_1, y_2, y_3]$	$Pr(Y=y\mid H_1)$	$Pr(Y=y\mid H_0)$	Decision
0 0 0	0.03	0.315	NO
0 0 1	0.07	0.135	NO
0 1 0	0.045	0.21	NO
0 1 1	0.105	0.07	YES
1 0 0	0.07	0.105	NO
1 0 1	0.21	0.045	YES
1 1 0	0.135	0.07	YES
1 1 1	0.315	0.03	YES

$\boxed{\text{Rule: go with majority}}$

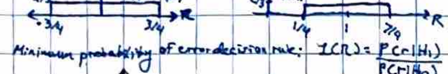
$$Pr(err) = \tfrac{1}{2}[(0.07+0.045+0.07+0.03) + (0.03+0.07+0.045+0.07)] = 0.215$$

4.2 In the binary communication system, $X=0$ and $X=1$ occur w/ a priori probabilities 1/4, 3/4 respectively. We observe $R = X + N$, with N continuous RV w/ probability density function shown.



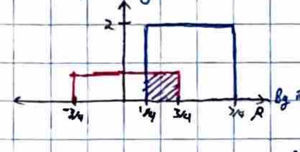
a) Find the minimum probability of error detector, and compute the associated probability of error, $Pr(\hat{X} \neq X)$

$H_0: R = N \rightarrow P_{R|H}(r\mid H_0) = P_N(r)$ $H_1: R = N+1 \rightarrow P_{R|H}(r\mid H_1) = P_N(r-1)$



Minimum probability of error decision rule: $\mathcal{L}(R) = \frac{P(r\mid H_1)}{P(r\mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{P_1} = \frac{1}{3}$

$$3P(r\mid H_1) \underset{H_0}{\overset{H_1}{\gtrless}} P(r\mid H_0)$$



By inspection, it is clear that MPE rule: $r \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{4}$

$$P_r(\varepsilon) = P_0P_F + P_1P_M = \frac{1}{4}P_F + \frac{3}{4}P_M = \frac{1}{4}\int_{1/4}^{\infty} P(r\mid H_0)\,dr + \frac{3}{4}\int_{-\infty}^{1/4} P(r\mid H_1)\,dr = \frac{1}{4}\left(\frac{1}{3}\right) + \frac{3}{4}(0) = \frac{1}{12}$$

b) Suppose the receiver does not know the true a priori probabilities, so it decides an ML detector. Find the ML detector and the associated probability of error if the rule is applied when the true prior probabilities are indeed different. Is the ML detector unique?

ML rule given by ~~[illegible]~~ $P(r\mid H_1) \underset{H_0}{\overset{H_1}{\gtrless}} P(r\mid H_0)$

Notice: $P(r\mid H_1) < P(r\mid H_0)$, $r < 1/4$

$=$ $1/4 \le r < 3/4$

$>$ $r \ge 3/4$

So, any decision rule of the form $r \underset{H_0}{\overset{H_1}{\gtrless}} \alpha$ $\alpha \in [1/4, 3/4)$ is a valid ML rule.

So, the rule is not unique. $P_r(\varepsilon) = P_0P_F + P_1P_M = P_0\int_{\alpha}^{\infty} P(r\mid H_0)\,dr + P_1\int_{-\infty}^{\alpha} P(r\mid H_1)\,dr$

$$= P_0\int_{\alpha}^{3/4}\frac{2}{3}\,dr + P_1\int_{1/4}^{\alpha}\frac{2}{3}\,dr = \frac{2}{3}\left[P_0\left(\frac{3}{4}-\alpha\right) + P_1\left(\alpha - \frac{1}{4}\right)\right]$$

$$= \frac{2}{3}\left[\frac{3P_0 - P_1}{4} + \alpha(P_1 - P_0)\right] \rightarrow \text{function of } \alpha \text{ unless } P_1 = P_0$$

4.3 There is suspicion ground water has been polluted by a chemical company. H_1 = polluted, H_0 = drinkable. $P_{X|H_1}(x\mid H_1) = xe^{-x}u(x)$, $P_{X|H_0}(x\mid H_0) = e^{-x}u(x)$. Design a hypothesis testing rule s.t. P_D, prob. that you decide to sue given water is polluted, is maximized s.t. P_F (prob. false detection) = 0.1.

NEYMAN-PEARSON TEST

$$\frac{P(x\mid H_1)}{P(x\mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \Rightarrow \frac{xe^{-x}}{e^{-x}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \Rightarrow x \underset{H_0}{\overset{H_1}{\gtrless}} \eta \ge 0$$

To calculate P_F for any value of η: $P_F = P_r(\text{Choose } H_1\mid H_0) = P_r(x > \eta\mid H_0)$

$$= \int_{\eta}^{\infty} e^{-x}u(x)\,dx = e^{-\eta},\ \eta \ge 0 \Rightarrow 0.1 = e^{-\eta} \Rightarrow \eta = -\ln(0.1)$$

The corresponding value of P_D:

$$P_D = P_r(\text{Choose } H_1\mid H_1) = P_r(x > \eta\mid H_1) = \int_{\eta}^{\infty} xe^{-x}u(x)\,dx = -xe^{-x}\Big|_{\eta}^{\infty} - e^{-x}\Big|_{\eta}^{\infty} = \eta e^{-\eta} + e^{-\eta}$$

4.4 A decision about the presence/absence of a target is made on the basis of observation y. If the target is present, $Y = A + X$, where $A > 0$ is a known constant. If target not present, $Y = X$. $X \sim N(0, N_0)$. A priori prob. that the target is not present is 0.999.

a) Find the MAP rule and P(error)

$P(y\mid H_0) = N(y; 0, N_0)$, $P(y\mid H_1) = N(y; A, N_0)$, $P_0 = 0.999$, $P_1 = 0.001$

$$\text{MAP: } P(H_1\mid y) \underset{H_0}{\overset{H_1}{\gtrless}} P(H_0\mid y) \Rightarrow P(y\mid H_1)P_1 \underset{H_0}{\overset{H_1}{\gtrless}} P(y\mid H_0)P_0 \Rightarrow y \underset{H_0}{\overset{H_1}{\gtrless}} \frac{2N_0\ln\left(\frac{P_0}{P_1}\right) + A^2}{2A} = \Gamma_{MAP}$$

$$P_r(error) = P_FP_0 + P_MP_1 = P_r(y > \Gamma_{MAP}\mid H_0)P_0 + P_r(y < \Gamma_{MAP}\mid H_1)P_1$$

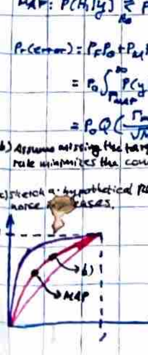
$$= P_0\int_{\Gamma_{MAP}}^{\infty} P(y\mid H_0)\,dy + P_1\int_{-\infty}^{\Gamma_{MAP}} P(y\mid H_1)\,dy = P_0\int_{\Gamma_{MAP}}^{\infty} N(y; 0, N_0)\,dy + P_1\int_{-\infty}^{\Gamma_{MAP}} N(y; A, N_0)\,dy$$

$$= P_0Q\left(\frac{\Gamma_{MAP}}{\sqrt{N_0}}\right) + P_1Q\left(\frac{-(\Gamma_{MAP} - A)}{\sqrt{N_0}}\right)$$

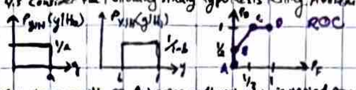
b) Assume missing the target is 10x worse than falsely detecting ($C_{00} = C_{11} = 0$, $C_{01} = 10$, $C_{10} = 1$). What decision rule minimizes the conditional risk for y?

$$y \underset{H_0}{\overset{H_1}{\gtrless}} \frac{2N_0\ln\left(\frac{P_0}{10P_1}\right) + A^2}{2A} \quad (\text{increases } P_D!)$$

c) Sketch a hypothetical ROC curve for the two decision rules. Sketch it if you have 5 indep. observations and if noise increases.



4.5 Consider the following binary hypothesis testing problem:



ROC shows possible (P_F, P_D) pairs as threshold γ is varied from $-\infty$ to ∞ for the decision rule $y \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$

a) What point on the ROC corresponds to $\gamma = b$? $\gamma = a$?

Shape of ROC tells us $b < a$, $a, b > 0$, $a, b < 1$. b/c as γ varies, the ROC has a region where P_F decreases where $P_D = 1$. The densities must overlap since otherwise we could achieve $P_F = 0$, $P_D = 1$.

So, $\gamma = b \rightarrow$ point C, $\gamma = a \rightarrow$ point B

b) Find values of a and b

At C: $\gamma = b \Rightarrow P_F = 1/3 = \frac{a-b}{a} \Rightarrow b = \frac{2}{3}a$

At B: $\gamma = a \Rightarrow P_D = 1/2 = \frac{1-a}{1-b} \Rightarrow 1 - \frac{2}{3}a = 2 - 2a$

$\boxed{a = 3/4 \quad b = 1/2}$

c) If $P_1 = P_0 = 0.5$, what (P_D, P_F) pair on the ROC corresponds to the γ w/ min $P(\varepsilon)$?

What is $P(\varepsilon)$? $P_r(\varepsilon) = P_0P_F + P_1(1 - P_D) = \frac{1}{2}(P_F + 1 - P_D)$

On the line connecting B and C, $P_D = \frac{1}{2} + \frac{3}{2}P_F$, so $P(err) = \frac{1}{2}\left(\frac{3}{2} - \frac{1}{3}P_D\right)$

So, we want P_D as large as possible, so we choose point C. $P(err) = 1/6$

d) Consider LRT $\mathcal{L}(y) = \frac{P_{Y|H}(y\mid H_1)}{P_{Y|H}(y\mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta$. Choose η s.t. $P_D = 1$, $P_F = 1/3$. Is η unique?

That pair is point C. The H_1 decision region must include $[b, 1]$ and exclude $[0, b]$.

$$P_{Y|H}(y\mid H_1) \underset{H_0}{\overset{H_1}{\gtrless}} \eta P_{Y|H}(y\mid H_0) \Rightarrow \frac{1}{1-b} > \eta\frac{1}{a},\ \eta > 0$$

So, $0 < \eta < \frac{a}{1-b} = 3/2$ η not unique, many choices work b/c densities are flat.

4.6 When a RAM chip is selected from a batch there is a probability α that it is defective. So the # of defective chips in N randomly selected is a binomial RV:

$$P_r[K = k] = \frac{N!\,\alpha^k(1-\alpha)^{N-k}}{k!(N-k)!}$$

We receive a shipment of N chips from batch 0 (prob. defect α_0) or batch 1 (prob. defect α_1), w/ prior probs P_0 and $P_1 = 1 - P_0$. Assume $0 < \alpha_0 < \alpha_1 < 1$. Want to determine which production batch our chips came from by testing all N units. Let K = # of defective RAM chips found out of N.

a) Find the MPE rule for deciding which batch the chips came from.

MPE means $C_{01} = C_{10} = 1$ (symmetric costs). So, threshold for likelihood test is $\tau = P_0/1-P_1$ $P[K = k\mid H_i] = \frac{N!\,\alpha_i^{\,k}(1-\alpha_i)^{N-k}}{k!(N-k)!}$

$$LR: \Lambda(k) = \frac{P[K=k\mid H_1]}{P[K=k\mid H_0]} = \frac{\alpha_1^{\,k}(1-\alpha_1)^{N-k}}{\alpha_0^{\,k}(1-\alpha_0)^{N-k}}$$

$$\text{So, MPE rule: } \frac{\alpha_1^{\,k}(1-\alpha_1)^{N-k}}{\alpha_0^{\,k}(1-\alpha_0)^{N-k}} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{1-P_1} = \tau$$

b) Sketch ROC for N = 2.

Possible values of measurements are k = 0, 1, 2

	0	1	2
$P(k\mid H_1)$	$(1-\alpha_1)^2$	$2\alpha_1(1-\alpha_1)$	α_1^2
$P(k\mid H_0)$	$(1-\alpha_0)^2$	$2\alpha_0(1-\alpha_0)$	α_0^2

, we know $\alpha_1 > 0$ and we have the ordering on the ratios:

$$\frac{(1-\alpha_1)^2}{(1-\alpha_0)^2} \le \frac{(1-\alpha_1)\alpha_1}{(1-\alpha_0)\alpha_0} \le \frac{\alpha_1^2}{\alpha_0^2}$$

ROC will have 4 points:

- (1, 1) corresponds with threshold 0
- when threshold is higher than likelihood for k=0 but lower than the LR for k=1 will have prob (detection) $2(\alpha_1)(1-\alpha_1) + \alpha_1^2$, $P_F = 2(\alpha_0)(1-\alpha_0) + \alpha_0^2$
- 3rd point: threshold higher than LR for k=1, lower than LR for k=2, $P_D = \alpha_1^2$, $P_F = \alpha_0^2$.
- last point (0, 0) corresponds to thresholds higher than the LR at k=2

Consider a 4-hypothesis ~~detection~~ problem where we send a vector signal into a communications channel: $H_0: S = (0,0)$, $H_1: S = (0,1)$, $H_2: S = (1,1)$, $H_3: S = (1,0)$. We observe $Y = S^T + V$, where V is zero mean 2D Gaussian RV w/ $\Sigma = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/4 \end{bmatrix}$

a) Find the optimal decision boundaries.

Assuming each hypothesis equally likely, we look for ML rule. Likelihoods: $f_Y(Y\mid H_0) \sim N(\begin{bmatrix}0\\0\end{bmatrix}, \Sigma)$, $f_Y(Y\mid H_1) \sim N(\begin{bmatrix}0\\1\end{bmatrix}, \Sigma)$, $f_Y(Y\mid H_2) \sim N(\begin{bmatrix}1\\1\end{bmatrix}, \Sigma)$, $f_Y(Y\mid H_3) \sim N(\begin{bmatrix}1\\0\end{bmatrix}, \Sigma)$. Optimal ML rule is to find the largest likelihood for each Y.

$$\ln(f_Y(Y\mid H_0)) + \ln(0.5\pi) = \frac{-y_1^2}{2} - \frac{y_2^2}{2}$$
$$\ln(f_Y(Y\mid H_1)) + \ln(0.5\pi) = -\frac{y_1^2}{2} - \frac{(y_2-1)^2}{2}$$
$$\ln(f_Y(Y\mid H_2)) + \ln(0.5\pi) = -\frac{(y_1-1)^2}{2} - \frac{(y_2-1)^2}{2}$$
$$\ln(f_Y(Y\mid H_3)) + \ln(0.5\pi) = -\frac{(y_1-1)^2}{2} - \frac{y_2^2}{2}$$

The decision ~~rule~~ simplifies to:

$y_1 < 0.5, y_2 < 0.5 \rightarrow H_0$ $y_1 < 0.5, y_2 > 0.5 \rightarrow H_1$

$y_1 > 0.5, y_2 > 0.5 \rightarrow H_2$ $y_1 > 0.5, y_2 < 0.5 \rightarrow H_3$

b) Compute P_D for each H_i

By symmetry, P_D will be the same for each H_i

For H_0: $P_D = \int_{-\infty}^{0.5}\int_{-\infty}^{0.5} \frac{2}{\pi}e^{-2(y_1)^2 - 2(y_2)^2}\,dy_1\,dy_2 = \Phi(1)^2$, where

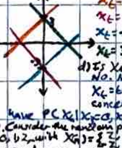
$\Phi(c)$ is the char. function for the standard Gaussian

(See this by using $z_1 = 2(y_1)$, $z_2 = 2(y_2)$)

$$P_D = \left[\int_{-\infty}^{1}\frac{1}{\sqrt{2\pi}}e^{-y_1^2/2}\,dy_1\right]\left[\int_{-\infty}^{1}\frac{1}{\sqrt{2\pi}}e^{-y_2^2/2}\,dy_2\right] = \Phi(1)\Phi(1)$$

5.1 Define $X_t = 3A + Bt$, where A, B RVs w/ $P\{A=1\} = P\{A=-1\} = P\{B=1\} = P\{B=-1\} = 0.5$

a) Sketch sample functions



$X_t = 3 + t$
$X_t = -3 - t$
$X_t = -3 + t$
$X_t = 3 - t$

b) Find $P\{X_t \ge 0\}$ ~~[illegible]~~ for each $t \ge 0$

$P\{X_t \ge 0\} = 0.5\ \forall t > 0$ except $t = 3$ where $P\{X_t \ge 0\} = 3/4$

c) Find $P\{X_t \ge 0\}$ ~~[illegible]~~ for EVERY $t \ge 0$

1 of the four sample paths stays positive $\forall t \ge 0$ so 0.25 is the prob.

d) Is X_t a Markov process?

No. Note $P(X_t\mid X_3 = 0)$ has the prob dist divided equally b/w $X_t = t - 3$ and $X_t = 3 - t$, but $P(X_t\mid X_3 = 0, X_0 = 3)$ has all of its concentrated on $X_t = 3 - t$. If it were Markov, we would have $P(X_t\mid X_3 = 0, X_0 = 3) = P(X_t\mid X_3 = 0)$ which it is not.

5.2 Consider the random process defined over the index of natural numbers 0, 1, 2, ... with $X(n) = \begin{cases} Z, & n \text{ even} \\ -Z, & n \text{ odd} \end{cases}$, where $Z \sim N(m, 1)$

a) Find mean and autocovariance of $X(n)$ $\mu(n) = \begin{cases} m & n \text{ even} \\ -m & n \text{ odd} \end{cases}$

$$K_X(n,m) = Cov[X(n), X(m)] = \begin{cases} Cov[Z, Z] = 1, & |n-m| \text{ even} \\ Cov[Z, -Z] = -1, & |n-m| \text{ odd} \end{cases}$$

b) Is $X(n)$ wide-sense stationary? Yes if m = 0, otherwise no ★

c) Is $X(n)$ Markov? YES $P(X_n\mid X_{n-1}, \ldots, X_0) = P(X\mid X_{n-1}) = \delta(X_n + X_{n-1})$

d) Is $X(n)$ strict-sense stationary? Not if m ≠ 0, yes if m = 0

e) Is $X(n)$ Gaussian? YES (w/ a caveat) If you take a collection of K sampling times, joint prob. density of the vector of sampled process values has a very degenerate distribution

f) Is $X(n)$ an independent increments process? No. Consider $X(2) - X(1) = 2Z$. That is not independent of the increment $X(1) - X(0) = -2Z$

5.3 Let $\{W_t, t \ge 0\}$ be standard Brownian motion

a) Express $P[W_3 \ge \frac{W_2 + W_4}{2} + 1]$ in terms of the Q function

$Z = W_3 - 1 - \frac{W_2 + W_4}{2}$ is a Gaussian RV. $E[Z] = -1$

$R_W(t,s) = \min(t,s)$, so

$$Var(Z) = E[(W_3 - \tfrac{W_2 + W_4}{2})^2]$$
$$= E[W_3^2] + 0.25E[W_2^2] + 0.25E[W_4^2] - E[W_2W_3] - E[W_3W_4] + 0.5E[W_2W_4]$$
$$= R_W(3,3) + 0.25R_W(2,2) + 0.25R_W(4,4) - R_W(2,3) - R_W(3,4) + 0.5R_W(2,4)$$
$$= 3 + 0.5 + 1 - 2 - 3 + 1 = 0.5$$

So, $P(Z) = Q(1/\sqrt{2}) = Q(\sqrt{2})$

b) Compute $E[W_1\mid W_3]$ $E[W_1\mid W_3] = \frac{Cov(W_1, W_3)}{Var(W_3)}W_3 = \frac{1}{3}W_3$

c) Compute $E[W_2\mid W_1, W_3]$

$$E[W_2\mid W_1, W_3] = Cov(W_2, \begin{bmatrix}W_1\\W_3\end{bmatrix})\,Cov(\begin{bmatrix}W_1\\W_3\end{bmatrix})^{-1}\begin{bmatrix}W_1\\W_3\end{bmatrix}$$
$$= [1\ 2]\begin{bmatrix}3/2 & -1/2\\-1/2 & 1/2\end{bmatrix}\begin{bmatrix}W_1\\W_3\end{bmatrix} = 0.5(W_1 + W_3)$$

5.4 Let $\{N_t, t \ge 0\}$ be a Poisson process with rate λ

a) Give a simple expression for $P(N_1 \ge 1\mid N_2 = 2)$ in terms of λ

Prob (no arrivals in interval of length 1) given from Poisson distribution w/ parameter λ, as is the probability there is only one arrival

$$P(N_1 \ge 1, N_2 = 2) = P(N_1 = 1, N_2 = 2) + P(N_1 = 2, N_2 = 2)$$
$$= P(N_1 = 1, N_2 - N_1 = 1) + P(N_1 = 2, N_2 - N_1 = 0)$$
$$= (\lambda e^{-\lambda})^2 + \left(\frac{1}{2}\lambda^2e^{-\lambda}\right)e^{-\lambda} = \frac{3}{2}\lambda^2e^{-2\lambda}$$

$$P(N_2 = 2) = \frac{(2\lambda)^2}{2}e^{-2\lambda} = 2\lambda^2e^{-2\lambda}$$

$$\text{So, } P(N_1 \ge 1\mid N_2 = 2) = \frac{P(N_1 \ge 1, N_2 = 2)}{P(N_2 = 2)} = \frac{\frac{3}{2}\lambda^2e^{-2\lambda}}{2\lambda^2e^{-2\lambda}} = 3/4$$

indep. of λ!

b) Give an expression for $P(N_2 = 2\mid N_1 \ge 1)$

~~Note~~ $P(N_2 = 2\mid N_1 > 2) = 0$ (no departures)

Want to compute $P(N_2 = 2\mid N_1 \in \{1,2\})$

$P(N_1 \in \{1,2\}, N_2 = 2) = \frac{3}{2}\lambda^2e^{-2\lambda}$, $P(N_1 \in \{1,2\}) = \lambda e^{-\lambda} + \frac{\lambda^2}{2}e^{-\lambda}$

$$P(N_2 = 2\mid N_1 \in \{1,2\}) = \frac{3/2\,\lambda^2e^{-2\lambda}}{\lambda e^{-\lambda} + \frac{\lambda^2}{2}e^{-\lambda}} = \frac{3}{2}\frac{\lambda e^{-\lambda}}{1 + \lambda/2}$$

c) Let $X_t = N_t - \lambda t$. Compute $E[X_t]$ $E[X_t] = E[N_t] - \lambda t = 0$

d) Given X_t as above, let $\tau \ge t$. Is $E[X_\tau\mid X_t] = X_t$?

$$E[X_\tau\mid X_t] = E[X_t + (X_\tau - X_t)\mid X_t] = E[X_t + (N_\tau - N_t) - \lambda(\tau - t)\mid X_t]$$
$$= X_t + E[N_\tau - N_t\mid X_t] - \lambda(\tau - t)$$
$$= X_t + E[N_\tau - N_t] - \lambda(\tau - t) = X_t \quad (\text{from IIP})$$

5.5 Let $X = \{X_n : n \in \mathbb{Z}\}$, where $\mathbb{Z}$ is the integers, consist of IID RVs that are Poisson distributed w/ mean $\beta > 0$. Define $Y_n = X_n + X_{n-1}$

a) Show Y_n is Poisson w/ mean 2β for each n

$$P(Y_n = m) = \sum_{k=0}^{m} P(X_n = k)P(X_{n-1} = m-k) = \sum_{k=0}^{m}\frac{\beta^k}{k!}e^{-\beta}\frac{\beta^{m-k}}{(m-k)!}e^{-\beta}$$
$$= e^{-2\beta}\frac{\beta^m}{m!}\sum_{k=0}^{m}\frac{m!}{k!(m-k)!} = e^{-2\beta}\frac{\beta^m}{m!}2^m = \frac{e^{-2\beta}(2\beta)^m}{m!}$$

(Sum is expansion of $(1+1)^m$)

b) Show X is a completely stationary random process. Since X_n are IID, the joint probability dist. function of any finite collection of times $n_1, \ldots, n_k$ will be the product of the individual probs. at each time & this will be time-invariant

c) Is Y completely stationary? YES. Y is formed by a time-invariant averaging of the process X. The joint prob dist. of Y at any finite collection of times $n_1, \ldots, n_k$ will be the same as the joint prob dist of the shifted vals for $n_i + s$.

d) Compute $\mu_Y(n) = E[Y_n]$ 2β (from a))

e) Compute $K_Y(n,m) = E[Y_nY_m] - \mu_Y(n)\mu(m)$

If $|n-m| > 1$, $K_Y(n,m) = 0$ b/c Y_n, Y_m independent

We consider n = m:

Y_n = sum of Poisson RVs, so cov is sum of [illegible] cov = 2β

n = m+1:

$Cov(X_{m+1} + X_m, X_m + X_{m-1}) = Cov(X_m, X_m) = \beta$

So $K_Y(n,m) = \begin{cases} 2\beta & n = m \\ \beta & |n-m| = 1 \\ 0 & \text{else} \end{cases}$

5.6 Let $X = \{X_n : n \in \mathbb{Z}\}$, and $Z = \{Z_n : n \in \mathbb{Z}\}$ be random processes s.t. $Y_n = X_n^2$, $Z_n = X_n^3$ $\forall n$

a) If X is Markov Y is Markov.

FALSE. Reflected random walk on -1, 0, 1, 2, moving R/L w/ prob = 1/2

Suppose $Y_2 = 1$. $P(Y_3 = 0\mid Y_2 = 1, Y_1 = 4) = 1/2 \neq 3/4 = P(Y_3 = 0\mid Y_2 = 1, Y_1 = 0)$

b) If X is Markov then Z is Markov

TRUE b/c the function is 1-1 and invertible

$P(Z_n = z\mid Z_{n-1}, Z_{n-2}, \ldots) = P(X_n = z^{1/3}\mid X_{n-1}, X_{n-2}, \ldots) = P(X_n = z^{1/3}\mid X_{n-1})$
$= P(Z_n = z\mid Z_{n-1})$

c) If X SSS then Y SSS TRUE Given sampling times $n_1, \ldots, n_m$, the joint finite distribution function of Y at these times depends on $Y = X^2$. So for any shift k, $F(X_{n_1}, \ldots X_{n_m}) = F(X_{n_1+k}, \ldots X_{n_m+k})$, the transformation $Y = X^2$ will give $F(Y_{n_1}, \ldots Y_{n_m}) = F(Y_{n_1+k}, \ldots, Y_{n_m+k})$

d) If X WSS, Y WSS. FALSE:

$x(t) = \cos(t + \Theta)$, Θ discrete & uniform in $\{0, \pi/2, \pi, 3\pi/2\}$

$E[x(t)] = 0$ $x(t)x(s) = \cos(t+\Theta)\cos(s+\Theta)$
$= \frac{1}{2}(\cos(s+t+2\Theta) + \cos(t-s))$

$$E[x(t)x(s)] = \frac{1}{2}\cos(t-s) + \frac{1}{8}(\cos(s+t) + \cos(s+t+\pi/2) + \cos(s+t+\pi) + \cos(s+t+3\pi/2))$$
$$= 0.5\cos(t-s) \rightarrow WSS$$

$y(t) = x^2(t) = \cos(t+\Theta)\cos(t+\Theta) = \frac{1}{2} + 1/2\cos(2t + 2\Theta)$

$E[y(t)] = 1/2$, $z(t) = 2(y(t) - 0.5) = \cos(2t + 2\Theta)$

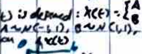
$z^2(t) = 0.5 + 0.5\cos(4t + 4\Theta)$

$$E[z^2] = 0.5 + 1/8(\cos(4t) + \cos(4t + 2\pi) + \cos(4t + 4\pi) + \cos(4t + 6\pi))$$
$$= 0.5 + 0.5\cos(4t) \rightarrow \text{DEPENDS ON } t$$

e) If X IIP then Y IIP. FALSE. Increment of $Y = X_n^2 - X_m^2 = (X_n - X_m)(X_n + X_m)$. Even if $X_n - X_m$ independent of X_m, $X_n + X_m$ will not be, so increment won't be indep. of Y_m.

5.7 A random process X(t) is defined: $X(t) = \begin{cases} A & t < \Theta \\ B & t \ge \Theta \end{cases}$, where A, B, Θ are indep. Gaussian RVs $A \sim N(-1,1)$, $B \sim N(1,1)$, $\Theta \sim N(0,1)$

a) Sketch a sample function



b) Find $P_{X(t)}(x)$, $P_{X(t_2)|X(t_1)}(x_2\mid x_1)$ for $t_2 > t_1$. Is X(t) a Gaussian rand. pr.?

Find joint density for $X(t)$, Θ & sum out Θ

$$P_{X(t_0)}(x) = \int P_{X(t_0),\Theta}(x,\theta)\,d\theta = \int P_{X(t_0)|\Theta}(x\mid\theta)P_\Theta\,d\theta$$

$\Theta \le t_0 \rightarrow P_{X(t_0)|\Theta}(x\mid\theta) \sim N(+1,1)$, $\Theta > t_0 \rightarrow P_{X(t_0)|\Theta}(x\mid\theta) \sim N(-1,1)$

$$P_X(t_0) = \int_{\theta \le t_0} N(x; 1,1)p_\Theta(\theta)\,d\theta + \int_{\theta > t_0} N(x; -1,1)p_\Theta(\theta)\,d\theta$$
$$= N(x; +1,1)\int_{\theta \le t_0} p_\Theta(\theta)\,d\theta + N(x; -1,1)\int_{\theta > t_0} p_\Theta(\theta)\,d\theta$$
$$= P_{X(t_0)|\Theta \le t_0}(x\mid\Theta \le t_0)P_r(\Theta \le t_0) + P_{X(t_0)|\Theta > t_0}(x\mid\Theta > t_0)P_r(\Theta > t_0)$$
$$= N(x; 1,1)[1 - Q(t_0)] + N(x; -1,1)Q(t_0)$$

Now, note $X(t_1)$, $X(t_2)$ are independent Gauss RVs if $t_1 < \Theta < t_2$, and $X(t_1) = X(t_2)$ otherwise

So, $P_{X(t_2)|X(t_1)}(x_2\mid x_1) = \delta(x_2 - x_1)P_r[\Theta \le t_1 \text{ or } t_2 < \Theta]$
$+ N(x_2; +1,1)P_r[t_1 < \Theta \le t_2]$
$= \delta(x_2 - x_1)[1 - Q(t_1) + Q(t_2)]$
$+ N(x_2; 1,1)[Q(t_1) - Q(t_2)]$
$= \delta(x_2 - x_1)[Q(-t_1) + Q(t_2)] +$
$N(x_2; 1,1)[Q(t_1) - Q(t_2)]$

$\boxed{1 - Q(t_1) = Q(-t_1)}$

NOT GAUSSIAN

c) Is X_t SSS? NO. First order density not stationary

d) Is X_t Markov? NO. Markovianity requires

$P_r[X(t_N) = x_N\mid X(t_{N-1}) = x_{N-1}, X(t_{N-2}) = x_{N-2}, \ldots X(t_1) = x_1]$
$= P_r[X(t_N) = x_N\mid X(t_{N-1}) = x_{N-1}]$

Suppose $x_{N-1} \neq x_{N-2}$. Then $P_r(X(t_N) = x_{N-1}) = 1$.

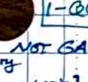
So $P_r[X(t_N) = X(t_{N-1}) = x_{N-1}, X(t_{N-2}) = x_{N-2}, x_{N-1} \neq x_{N-2}] = 1$

but $P_r[X(t_N) = x_{N-1}\mid X(t_{N-1}) = x_{N-1}] < 1$, so not Markov

e) Is X(t) IIP? NO, b/c all IIP processes are Markov

5.8 Let $W = (W_t : t \ge 0)$ be a standard Brownian motion w/ $\sigma^2 = 1$. Let $B_t = W_t - tW_1$ for $0 \le t \le 1$. $B = (B_t : 0 \le t \le 1)$ is a mean zero Gaussian random process called a Brownian bridge.

a) Sketch a sample path



b) Find autocorr. of B

For $s, t \in [0,1]$, if $s \le t$, $R_B(s,t) = E[(W_s - sW_1)(W_t - tW_1)]$
$= s - st - st + st = s - st = s(1-t)$

So, $R_B(s,t) = \begin{cases} s(1-t), & s \le t \\ t(1-s) & t \le s \end{cases}$

c) Is B Markov? $\rho(s,t)$ = corr. coeff. b/w X_s, X_t:

$$\rho(s,t) = \frac{Cov(B_s, B_t)}{\sqrt{Var(B_s)Var(B_t)}} = \frac{s(1-t)}{\sqrt{s(1-s)t(1-t)}} = \sqrt{\frac{s}{1-s}}\sqrt{\frac{1-t}{t}}$$

So, for $0 \le r < s < t \le 1$, $\rho(r,s)\rho(s,t) = \rho(r,t)$, so YES, MARKOV!

d) Show B is independent of W_1.

Show $Cov(B_t, W_1) = 0\ \forall\ 0 \le t \le 1$ (uncorr. joint Gauss vecs. indep.)

$Cov(B, W_1) = Cov(W_t - tW_1, W_1) = Cov(W_t, W_1) - tCov(W_1, W_1) = t - t = 0$ ✓

e) Let $X_t = (1-t)W_{\frac{t}{1-t}}$ for $0 \le t < 1$ and let $X_1 = 0$. Let X be the random process $X = (X_t : 0 \le t \le 1)$. X a zero mean Gaussian RP. Find autocorr of X

For $0 \le s \le t \le 1$, $\frac{s}{1-s} \le \frac{t}{1-t}$, so

$$R_X(s,t) = Cov((1-s)W_{s/1-s}, (1-t)W_{t/1-t})$$
$$= (1-s)(1-t)Cov(W_{s/1-s}, W_{t/1-t}) = (1-s)(1-t)\frac{s}{1-s} = s(1-t)$$

X also a Brownian bridge!

6.3 ~~Let $X = (X_k : k \in \mathbb{Z})$ be a random process s.t. X_k are independent Poisson~~ ~~[illegible]~~ Customers enter system according to a Poisson point process on $\mathbb{R}$ of rate λ, meaning the # of arrivals, N(a, b], in an interval (a, b] has the Poisson dist. w/ mean $\lambda(b-a)$ and the #s of arrivals in disjoint intervals are independent. Each customer stays for 1 unit of time, indep. of other customers. # of customers in the system at a given time given by $X_t = N(t-1, t]$

a) Find mean and autocov of X Since $N(t-1, t]$ Poisson w/ mean λ, $E[X_t] = \lambda$

If $|s-t| \ge 1$, X_s indep of X_t, else they have an overlap length of $1 - |s-t|$.

If $0 \le t - s < 1$, $(s-1, s] = (s-1, t-1] \cup (t-1, s]$, $(t-1, t] = (t-1, s] \cup (s, t]$.

So, $X_s = N(s-1, t-1] + N(t-1, s]$, $X_t = N(t-1, s] + N(s, t]$.

$$C_X(s,t) = Cov(N(s-1, t-1] + N(t-1, s], N(t-1, s] + N(s, t])$$
$$= Var(N(t-1, s]) = \lambda(s - t + 1) = \lambda(1 - |t-s|)$$

b) Is X stationary, WSS? YES and YES!

Let $n \ge 1$, $t_1, \ldots, t_n$ fixed. $u_0 < u_1 \ldots < u_m$ s.t. $t_i - 1$ and t_i are in $\{u_0, \ldots u_m\}$ for $1 \le i \le n$. ∃ m×n matrix G w/ each entry equal to 0 or 1 s.t.

$(X_{t_1+s}, \ldots X_{t_n+s}) = (N(u_0+s, u_1+s], \ldots N(u_{m-1}+s, u_m+s])\,G$

for all s, since random vec on right consists of indep Poisson RVs, not dependent on s. So X stationary.

c) Is X Markov?

NO. The past tells us more about the fu[illegible] than the present alone

d) Find a simple expression for $P(X_t = 0$ for $t \in [0,1])$ [illegible] in terms of λ.

The system empty during [0, 1] iff there a[illegible] no arrivals during (-1, 1]. So $P\{X_t = 0 \text{ for } t \in [0,1]\} = P\{N(-1,1] = 0\} = \exp(-2\lambda)$

e) Find simple expression for $P(X_t > 0$ for $t \in [0,1]\}$, in terms of λ.

Equivalent to there being at least one customer in the queue at time 0. T = ~~[illegible]~~ time of first arrival from time zero, S be time of last arrival before time zero. S, T indep, exp. dist. random vars w/ parameter λ

$$P\{X_t > 0 \text{ for } t \in [0,1]\} = P\{S + T \le 1\}$$
$$= \int_0^1\int_0^{1-t}\lambda^2e^{\lambda(s+t)}\,ds\,dt = 1 - (1 - \lambda)e^{-\lambda}$$

Binary Hypothesis Testing

- X is either H_0 or H_1
- Decision rule is mapping of observation to one of H_0 or H_1
- Conditional probabilities
  - $P_F = P_r(\text{Choose } H_1 | H_0)$ = Pr False alarm ← Type I error
  - $P_D = P_r(\text{Choose } H_1 | H_1)$ = Pr of Detection
  - $P_M = P_r(\text{Choose } H_0 | H_1)$ = Pr of Miss = $1 - P_D$ ← Type II error
- **Bayes Risk Formulation**
  - Minimize "Bayes Risk" = E[Cost] given
    1. A priori probabilities $P_i = P_r[H_i]$
    2. Observation model: $P_{Y|H_i}(y|H_i)$
    3. Costs C_{ij} = Cost of decide H_i when H_j is true
- Solution is the likelihood ratio test (LRT):

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

} scalar test regardless of data dim

- Probability of error: $P_r(\varepsilon) = P_r(\text{Choose } H_0, H_1) + P_r[\text{Choose } H_1, H_0]$

$$= P_M P_1 + P_F P_0$$

- 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 $C_{ij} = 1 - \delta_{ij}$

$$\Rightarrow \text{MAP decision rule } P_{H_1|Y}(H_1|y) \underset{H_0}{\overset{H_1}{\gtrless}} P_{H_0|Y}(H_0|y)$$

- MPE and $P_0 = P_1 = 1/2 \Rightarrow$ 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 ⇒ choose η in LRT to minimize the max expected cost as a function of P_0.
- $P_D = \left(\frac{C_{01}-C_{00}}{C_{01}-C_{11}}\right) - \left(\frac{C_{10}-C_{00}}{C_{01}-C_{11}}\right)P_F$
- Neyman-Pearson tests: P_0, P_1 unknown and C_{ij} unknown: maximize P_D subject to $P_F \leq \alpha$.
- Receiver Operating Characteristic: Plot of $P_D(\eta)$ vs $P_F(\eta)$ as threshold η in LRT is varied
  - $P_D = \int_{\{Y | \text{say } H_1\}} p(y|H_1)\,dy \quad P_F = \int_{\{Y | \text{say } H_1\}} p(y|H_0)\,dy$
- Properties:
  1. $(P_D, P_F) = (0,0)$ and $(P_D, P_F) = (1,1)$ are always on the ROC
  2. ROC is boundary b/w what is achievable & what is not
  3. η is the slope of the 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 RVs, ROC consists of points

Prediction, interpolation, extrapolation: Gaussian Process

$$E[X(t_1)|X(t_2)] = \mu_X(t) + \frac{K_{XX}(t_1,t_2)}{K_{XX}(t_2,t_2)}(x(t_2) - \mu_X(t_2))$$

$$\Sigma_{X(t_1)|X(t_2)} = K_{XX}(t_1,t_1) - \frac{K^2_{XX}(t_1,t_2)}{K_{XX}(t_2,t_2)}$$

$$f_{X(t_1)|X(t_2)}(x_1|x_2) = N(x_1; E[X(t_1)|X(t_2) = x_2], \Sigma_{X(t_1)|X(t_2)})$$

random process: indexed collection of random variables all on the same probability space $(\Omega, \mathcal{F}, P)$. If $\mathbb{T} = \mathbb{Z}$, X a discrete time random process. If $\mathbb{T} = \mathbb{R}$, X a cts time random process

- For each t fixed X_t is a function on Ω
- X a function on $\mathbb{T} \times \Omega$ w/ value $X_t(\omega)$ for given $t \in \mathbb{T}, \omega \in \Omega$
- For each ω fixed w/ $\omega \in \Omega$, $X_t(\omega)$ is a function of t called the sample path corresponding to ω
- NOTATION: $\mu_X(t) = E[X_t]$, $R_X(s,t) = E[X_s, X_t]$, $K_X(s,t) = \text{Cov}(X_s, X_t)$ $= R_{XX}(s,t) - \mu_X(s)\mu_X(t)$

$$F_{X,n}(x_1,t_1; \ldots x_n,t_n) = P\{X_{t_1} \leq x_1, \ldots X_{t_n} \leq x_n\}$$

$$P_{X,n}(x_1,t_1; \ldots x_n,t_n) = P\{X_{t_1} = x_1, \ldots X_{t_n} = x_n\}$$

- A random process is Gaussian if the RVs X_t, $t \in \mathbb{T}$ are jointly Gaussian
- A random process has independent increments if for any positive integer n, any $t_0 < t_1 < \ldots < t_n$ in $\mathbb{T}$, the increments $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are mutually independent
- **Brownian motion (Wiener process)**: w/ parameter $\sigma^2 > 0$ is a random process $W = (W_t; t \geq 0)$ s.t.

$\mu_{W}(t) = 0$, $R_W(s,t) = \sigma^2 s$, $s \leq t$

B.0 $P\{W_0 = 0\} = 1$

B.1 W has independent increments

B.2 $W_t - W_s$ has the $N(0, \sigma^2(t-s))$ dist. for $t \geq s$

B.3 $P\{W_t$ is a continuous function of $t\} = 1$ (W is sample path cts w/ prob 1)

counting function: a function f on $\mathbb{R}_+$ where $f(0) = 0$, f nondecreasing, right cts, and integer valued

- $f(t)$ = # of counts observed during $(0, t]$.
- $f(b) - f(a)$ = # of counts in $(a, b]$
- Can describe w/ sequence t_i, where t_i = time of ith count — count times

OR sequence u_i, where $u_1 = t_1$, $u_i = t_i - t_{i-1}$ — intercount times

Poisson process: Let $\lambda \geq 0$. A poisson process w/ rate λ is a random process $N = (N_t : t \geq 0)$ s.t.

N.1 N is a counting process N.2 N is IIP

N.3 $N(t) - N(s)$ has $Poi(\lambda(t-s))$ distribution for $t \geq s$

- Let N be a counting process, $\lambda > 0$. The following are equivalent
  - a) N is a Poisson process w/ rate λ
  - b) The intercount times are independent, Exp(λ) RVs.
  - c) For each $\tau > 0$, N_τ is a Poisson RV w/ parameter $\lambda\tau$ and given $\{N_\tau = n\}$, the times of the n counts during $[0, \tau]$ are the same as n independent, unif $[0, \tau]$ RVs reordered to be nondecreasing. So, for any $n \geq 1$, the conditional density of the first n count times $(T_1, \ldots T_n)$ given $\{N_T = n\}$ is $\left[f(t_1, \ldots, t_n | N_\tau = n) = \begin{cases} \frac{n!}{\tau^n} & 0 < t_1 < \ldots < t_n \leq \tau \\ 0 & \text{else} \end{cases} \right]$

- stationary: n-fold pdfs invariant to time shifts → IID process. IIP random walk → IIP $\not\Rightarrow$ Stationary!!
- **WSS** $X = (X_t, t \in \mathbb{R})$ or $X = (X_n, n \in \mathbb{Z})$
  - 2nd order, e.g. $E[X_t^2] < \infty\ \forall t$
  - 2nd order moments invariant to time shift

$$\mu_X(t) = \mu_X(t - \tau) = \mu_X(0)$$

$$R_{XX}(t,s) = R_{XX}(t-\tau, s-\tau) = R_{XX}(t-s, 0) = R_{XX}(t-s)$$

NOT WSS: Random walk, Brownian motion, PCP

Poisson: pmf $= p(i) = \frac{\lambda^i e^{-\lambda}}{i!}$, $i \geq 0$ mean λ, variance λ

Exponential: Exp(λ), $\lambda > 0$

pdf: $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$ mean $1/\lambda$, var $1/\lambda^2$

Gaussian: $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma \geq 0$

pdf: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$ mean μ, var σ^2

cdf $\Phi(c)$, $Q(c) = 1 - \Phi(c) = \int_c^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$

IIP Process: for any $t_1 \leq t_2 \leq t_3 \leq \ldots t_n$, the increments $X_{t_i} - X_{t_{i-1}}$ are all independent and in particular $X_{t_i} - X_{t_{i-1}}$ is independent of X_{i-1}

RANDOM WALK 2ND ORDER PROPS: $\mu_X(n) = E[X_n] = n(2p-1)$

$$R_{XX}(n,m) = E[X_n - X_m]E[X_m] + E[X_m^2]$$

$$\hookrightarrow E[(X_n - X_m + X_m)X_m]$$

MARKOV PROCESS: X, Y, Z random vectors if X - Y - Z forms a chain; conditioned on Y, X and Z are independent

$$f_{XZ|Y}(x,z|y) = f_{X|Y}(x|y)f_{Z|Y}(z|y)$$

$$f_{XYZ}(x,y,z) = f_{X|Y}(x,y)f_{Z|Y}(z,y)f_Y(y)$$

(e.g. random walk, any IID IIP process) recursive eq, stochastic diff/table equation

Markov $\not\Rightarrow$ IIP (e.g. $X_n = \frac{1}{2}X_{n-1} + W_n$), W_n IID

autocovariance/correlation of a Poisson process