EE 505 Part Review.

1. Pdf of linear transformation

$$V = aX + bY \qquad W = cX + eY \qquad \text{or} \quad \begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} a & b \\ c & e \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = A \begin{bmatrix} X \\ Y \end{bmatrix}$$

where $|A| = |ae - bc| \neq 0$

$$\Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} v \\ w \end{bmatrix}$$

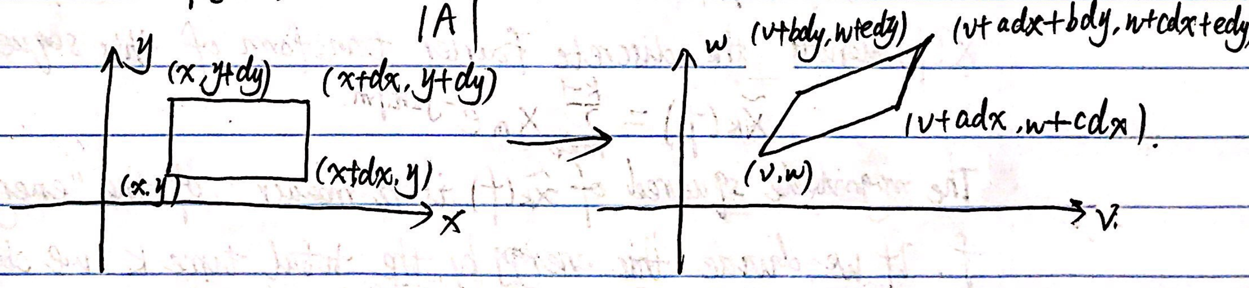
Equal probabilities. $f_{X,Y}(x,y)\, dx\, dy = f_{V,W}(v,w)\, dP$.

$$\Rightarrow f_{V,W}(v,w) = \frac{f_{X,Y}(x,y)}{\left| \frac{dP}{dx\,dy} \right|}$$

$$\left| \frac{dP}{dx\,dy} \right| = \frac{|ae-bc|(dx\,dy)}{(dx\,dy)} = |ae - bc| = |A|$$

$\bar{Z} = A\bar{X}$ where A is an $n \times n$ invertible matrix.

$$f_Z(\bar{z}) = \frac{f_X(A^{-1}\bar{z})}{|A|}$$



2. Pairs of jointly Gaussian Random variables

$$f_{X,Y}(x,y) = \frac{\exp\left\{ \frac{-1}{2(1-\rho_{X,Y}^2)} \left[\left(\frac{x-m_1}{\delta_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\delta_1}\right)\left(\frac{y-m_2}{\delta_2}\right) + \left(\frac{y-m_2}{\delta_2}\right)^2 \right] \right\}}{2\pi \delta_1 \delta_2 \sqrt{1-\rho_{X,Y}^2}}$$

The pdf is constant for x and y for which the argument of exponent is constant:

$$\left[\left(\frac{x-m_1}{\delta_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\delta_1}\right)\left(\frac{y-m_2}{\delta_2}\right) + \left(\frac{y-m_2}{\delta_2}\right)^2 \right] = \text{constant}.$$

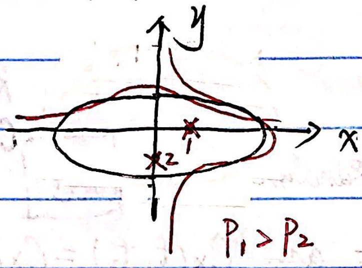
when $\rho_{X,Y} \neq 0$: $\theta = \frac{1}{2}\arctan\left(\frac{2\rho_{X,Y}\delta_1\delta_2}{\delta_1^2-\delta_2^2}\right)$

$\therefore$ $\delta_1 > \delta_2$: $0 < \theta < \frac{\pi}{4}$

$\delta_1 = \delta_2$: $\theta = \frac{\pi}{4}$

$\delta_1 < \delta_2$: $\frac{\pi}{4} < \theta < \frac{\pi}{2}$

Summary: ① when $\rho_{X,Y} = 0$, the contour of equal value of pdf:



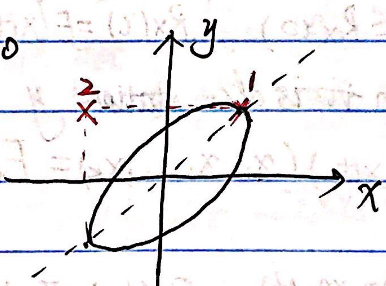
X and Y are independent.

① $\delta_1 \neq \delta_2$, it is an ellipse,

② $\delta_1 = \delta_2$, it is a circle.

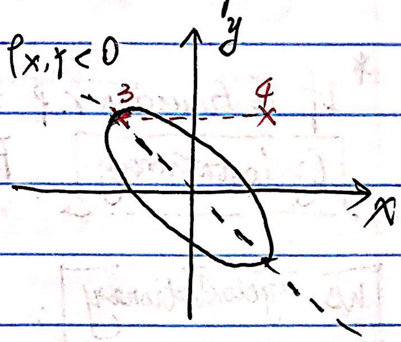
②. when $\rho_{X,Y} \neq 0$, the contour of equal value of pdf:

$\rho_{X,Y} > 0$



or

$\rho_{X,Y} < 0$



It must be a declining ellipse no matter if δ_1 is equal to δ_2

$$\left[\left(\frac{x-m_1}{\delta_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\delta_1}\right)\left(\frac{y-m_2}{\delta_2}\right) + \left(\frac{y-m_2}{\delta_2}\right)^2\right] = \text{constant}.$$

If $\rho_{X,Y} > 0$, constant 1 < constant 2, so point 2 need to be close to the original point.

If $\rho_{X,Y} < 0$, constant 3 < constant 4, So point 4 need to be close to the original point.

3. $\boxed{\text{Pdf of general transformation}}$

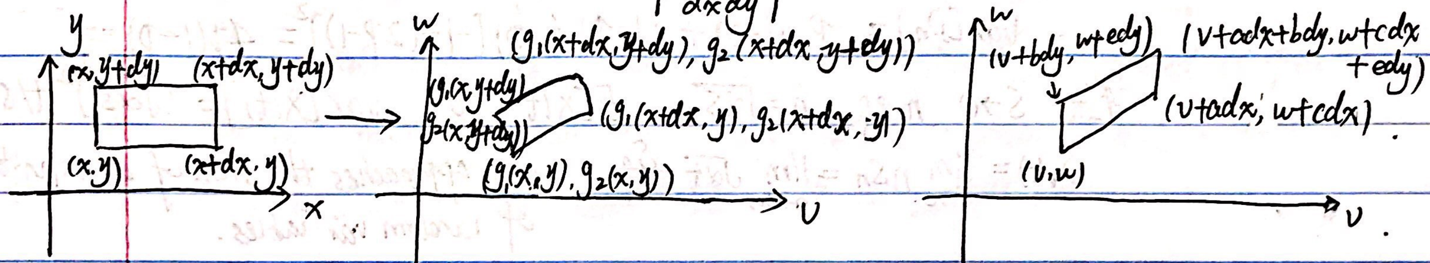
$V = g_1(X, Y)$ and $x = h_1(v, w)$

$W = g_2(X, Y)$ $\quad y = h_2(v, w)$

$\boxed{g_k(x+dx, y) \simeq g_k(x, y) + \frac{\partial}{\partial x} g_k(x, y)\, dx}$. $k = 1, 2$.

$$\Rightarrow f_{X,Y}(x, y)\, dx\, dy = f_{V,W}(v, w)\, dP$$

$$\Rightarrow f_{V,W}(v, w) = \frac{f_{X,Y}(h_1(v,w), h_2(v,w))}{\left|\frac{dP}{dx\,dy}\right|}$$



"stretch factor" at (v, w): $J(x,y) = \det \begin{bmatrix} \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{bmatrix}$ ... Jacobian.

$J(v,w) = \det \begin{bmatrix} \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \end{bmatrix}$ $\qquad |J(v,w)| = \frac{1}{|J(x,y)|}$

$$\Rightarrow f_{V,W}(v,w) = \frac{f_{X,Y}(h_1(v,w), h_2(v,w))}{|J(x,y)|} = f_{X,Y}(h_1(v,w), h_2(v,w))\,|J(v,w)|$$

$$\Rightarrow f_{\bar Z}(\bar z) = \left.\frac{f_{\bar X}(\bar x)}{|\det A|}\right|_{\bar x = A^{-1}\bar z} = \frac{f_{\bar X}(A^{-1}\bar z)}{|\det A|}.$$

4. $\boxed{\text{Vector random variables}}$

① For $\bar X = (X_1, X_2, \dots, X_n)^T$. ($X$ is a $1\times n$ matrix)

correlation matrix: $R_X = E[\bar X \bar X^T]$.

covariance matrix: $K_X = E[(X - m_X)(X - m_X)^T]$

$= E[XX^T] - m_X E[X^T] - E[X] m_X^T + m_X m_X^T$

$= R_X - m_X m_X^T$

For $\bar{Y} = A\bar{X}$:

$k_Y = E[(Y-m_Y)(Y-m_Y)^T] = E[(AX-Am_x)(AX-Am_x)^T]$.

$= E[A(X-m_x)(X-m_x)^T A^T] = A\,E[(X-m_x)(X-m_x)^T]A^T$

$= A k_x A^T$

$K_{XY} = E[(X-m_x)(Y-m_Y)^T] = E[(X-m_x)(X-m_x)^T A^T] = K_x A^T$.

If $\bar{x}$ are $\boxed{\text{uncorrelated}}$ and have unit variance. $k_x = I$

$\Rightarrow k_Y = A k_x A^T = A I A^T = AA^T$. ($Y$ are correlated)

② Diagonalization of covariance matrix. (we are interested in finding an $n\times n$ matrix A such that $Y = AX$ has a covariance matrix diagonal):

$P^T K_x P = \Lambda$ and $P^T P = I$. (Let $A = P^T$)

$K_x e_i = \lambda_i e_i$. where $e_i^T e_j = \delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i\neq j \end{cases}$

$\therefore P = [\bar{e}_1, \bar{e}_2, \dots \bar{e}_n]$ $\Lambda = \mathrm{diag}[\lambda]$. P is an orthonormal matrix.

$\Rightarrow K_x P = P\Lambda$.

$\Rightarrow P^T K_x P = P^T P \Lambda = \Lambda$

If we let $A = P^T$ and $Y = AX = P^T X$.

$\Rightarrow k_Y = P^T K_x P = \Lambda$.

In summary, any covariance matrix K_x can be diagonalized by a linear transformation.

$\det k_Y = \det P^T \det K_x \det P = \det \Lambda = \lambda_1 \lambda_2 \dots \lambda_n$.

k_Y is not invertible if one λ of K_x is zero.

$$X = PP^T X = PY = [\bar{e}_1, \bar{e}_2, \dots \bar{e}_n] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{k=1}^{n} Y_k \bar{e}_k$$

X can be expressed as a weighted sum of eigenvectors of K_x, where coefficient are uncorrelated r.v Y_K.

③ Generating r.v.s with specified covariance matrix (we wish to generate a r.v Y with an arbitrary valid covariance matrix k_Y).

X is a vector r.v uncorrelated, zero mean, unit variance.

$$Y = AX.$$

$$k_Y = A k_X A^T = AA^T$$

$$\because \quad P^T k_Y P = P^T P \Lambda = \Lambda.$$

$$\therefore \quad P \Lambda P^T = P P^T k_Y P P^T = k_Y = P \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} P^T = P \Lambda^{\frac{1}{2}} I (P \Lambda^{\frac{1}{2}})^T$$

where $A = P\Lambda^{\frac{1}{2}}$, $K_X = I$, P and Λ are eigenvector and eigenvalue of K_Y

$$\therefore \quad Y = AX = P\Lambda^{\frac{1}{2}} X.$$

Summary:

$$\begin{array}{ccc} K_Y & \underset{A=P}{\overset{A=P^T}{\rightleftarrows}} & K_X(\Lambda) \\ & \searrow^{A=P\Lambda^{\frac{1}{2}}} & \\ & & K_X(I) \end{array}$$

where P and Λ are eigenvector and eigenvalue of Y.

5. Jointly Gaussian Random vectors.

① $$f_{\bar{X}}(\bar{x}) \triangleq f_{X_1, X_2, \ldots, X_n}(x_1, \ldots, x_n) = \frac{\exp\{-\frac{1}{2}(\bar{x}-\bar{m})^T K^{-1} (\bar{x}-\bar{m})\}}{(2\pi)^{\frac{n}{2}} |K|^{\frac{1}{2}}}$$

where $\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ $\quad \bar{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}$

$$K = \begin{bmatrix} var(X_1) & cov(X_1, X_2) & \cdots & cov(X_1, X_n) \\ cov(X_1, X_2) & var(X_2) & \cdots & cov(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ cov(X_n, X_1) & \cdots & & var(X_n) \end{bmatrix}$$

Fact: The marginal pdf for x and z is also Gaussian and has the same set of means, variances and covariances.

② Linear transformation of Gaussian Random variable

Let $\bar{X} = (X_1, \dots, X_n)^T$ be jointly Gaussian with covariance matrix K_x, mean vector $\bar{m}_x$, $\bar{Y} = (Y_1, \dots Y_n)$ is defined by.

$$\bar{Y} = A\bar{X}$$

$$f_{\bar{Y}}(\bar{y}) = \frac{f_{\bar{X}}(A^{-1}\bar{y})}{|A|}$$

$$= \frac{\exp\{-\frac{1}{2}(A^{-1}\bar{y} - \bar{m}_x)^T K_x^{-1}(A^{-1}\bar{y} - m_x)\}}{(2\pi)^{\frac{n}{2}} |A| |K_x|^{\frac{1}{2}}}$$

$\because (A^{-1}y - \bar{m}_x) = A^{-1}(\bar{y} - A\bar{m}_x)$ and $(A^{-1}\bar{y} - \bar{m}_x)^T = (\bar{y} - A\bar{m}_x)^T A^{-1T}$

$$= (\bar{y} - A\bar{m}_x)^T A^{T-1}.$$

$\therefore (\bar{y} - A\bar{m}_x)^T A^{T-1} K_x^{-1} A^{-1}(\bar{y} - A\bar{m}_x) = (\bar{y} - A\bar{m}_x)^T (AK_xA^T)^{-1}(\bar{y} - A\bar{m}_x)$

$$= (\bar{y} - \bar{m}_Y)^T K_Y^{-1}(\bar{y} - \bar{m}_Y)$$

$\because \det(K_Y) = \det(AK_xA^T) = \det(A)\det(K_x)\det(A^T) = \det(A)^2\det(K_x)$.

$$\therefore \quad f_{\bar{Y}}(\bar{y}) = \frac{\exp\{-\frac{1}{2}(\bar{y} - \bar{m}_Y)^T K_Y^{-1}(\bar{y} - \bar{m}_Y)\}}{(2\pi)^{\frac{n}{2}} |K_Y|^{\frac{1}{2}}}$$

where $\bar{m}_Y = A\bar{m}_x$ and $K_Y = AK_xA^T$

If we wish to transform $\bar{X}$ to a vector $\bar{Y}$ of independent Gaussian R.V.

(Find a matrix A such that $AK_xA^T = \Lambda$)

$$\therefore f_{\bar{Y}}(\bar{y}) = \frac{\exp\{-\frac{1}{2}(\bar{y} - \bar{n})^T \Lambda^{-1}(\bar{y} - \bar{n})\}}{(2\pi)^{\frac{n}{2}} |\Lambda|^{\frac{1}{2}}}$$

$$= \frac{\exp\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - n_i)^2/\lambda_i\}}{[(2\pi\lambda_1)(2\pi\lambda_2)\dots(2\pi\lambda_n)]^{\frac{1}{2}}}$$

where λ_i are the diagonal components of Λ, $Y_1 \dots Y_n$ are independent r.vs with means n_i and variance λ_i

6. Estimation of random variables.

① MAP and ML estimators. (we are interested in finding the most probable input given the observation $Y=y$)

MAP (maximum a posterior) estimator:

$$\max_x P[X=x \mid Y=y] = \frac{P[Y=y \mid X=x]\,P[X=x]}{P[Y=y]}.$$

In some situations we know $P[Y=y \mid X=x]$ but we don't know a prior probability $P[X=x]$. We select the estimator value x as the value that maximize the likelihood of $Y=y$ (ML (maximum likelihood)) estimator.

$$\max_x P[Y=y \mid X=x].$$

Summary: MAP estimator for X given the observation Y:

$$\max_x f_X(X=x \mid Y=y)$$

ML estimator for X given the observation Y:

$$\max_x f_Y(Y=y \mid X=x).$$

② Minimum MSE Linear estimator. (The estimate for X is given by a function of the observation $\hat{X}=g(Y)$)

Mean square error: $e=E[(X-g(Y))^2]$.

$$\min_a E[(X-a)^2]=E[X^2]-2aE[X]+a^2 \Rightarrow a^*=E[X]$$

$$\Rightarrow E[(X-a^*)^2]=var(X).$$

$$\min_{a,b} E[(X-aY-b)^2] \Rightarrow b^*=E[X-aY]=E[X]-aE[Y]$$

$$\Rightarrow \min_a E[\{(X-E[X])-a(Y-E[Y])\}^2]$$

$$0=\frac{d}{da}E[(X-E[X])-a(Y-E[Y])^2].$$

$$= -2E[\{(X-E[X]) - a(Y-E[Y])\}(Y-E[Y])]$$

$$= -2(cov(X,Y) - a\,Var(Y))$$

$$\Rightarrow a^* = \frac{cov(X,Y)}{Var(Y)} = \rho_{X,Y}\frac{\delta_X}{\delta_Y}$$

∴ Minimum mean square error (mmse) linear estimator:

$$\hat{X} = a^* Y + b^*$$

$$= \rho_{X,Y}\delta_X \frac{Y-E[Y]}{\delta_Y} + E[X]$$

The mean square error of the best linear estimator is:

$$e_L^* = E[((X-E[X]) - a^*(Y-E[Y]))^2]$$

$$= E[((X-E[X]) - a^*(Y-E[Y]))(X-E[X])] - a^* E[((X-E[X]) - a^*(Y-E[Y]))(Y-E[Y])]$$

$$= E[((X-E[X]) - a^*(Y-E[Y]))(X-E[X])]$$

$$= var(X) - a^* cov(X,Y)$$

$$= var(X)(1-\rho_{X,Y}^2).$$

7. Random process.

A. Possion Process: (Counting process $N(t)$ can be approximated by the binomial counting process)

$$\lambda t = np \Rightarrow p = \frac{\lambda t}{n}.$$

$N(t)$ in the interval $[0,t]$ has a Possion distribution with mean λt

$$P[N(t)=k] = \binom{n}{k} p^k (1-p)^{n-k} = \lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1-\frac{\lambda t}{n}\right)^{n-k}$$

$$= \frac{(\lambda t)^k}{k!} \lim_{n\to\infty} \underbrace{\frac{n!}{(n-k)!}\left(\frac{1}{n^k}\right)}_{=1} \underbrace{\left(1-\frac{\lambda t}{n}\right)^n}_{= e^{-\lambda t}} \underbrace{\left(1-\frac{\lambda t}{n}\right)^{-k}}_{=1}$$

$$= \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad \text{for } k = 0, 1, \ldots$$

$E[N(t)=k]=\lambda t$, $Var[N(t)]=\lambda t$.

$$P[T\leq t] = 1-P[T>t]$$
$$=1-(1-p)^n$$
$$=1-(1-\frac{\lambda t}{n})^n$$
$$=1-e^{-\lambda t} \quad \text{as } n\to\infty.$$

$\Rightarrow$ T is an exponential

$\Rightarrow$ Interevent times in a possion process form an iid sequence of exponential r.v with mean $1/\lambda$

$\Rightarrow$ If the number of arrivals in interval $[0,t]$ is k, then the individual arrival times are distributed independently and uniformly in the interval

Exp: two independent possion process of rate λ_1, λ_2.

① Prob that a message arrives first on line 2.

$$P\{T_2<T_1\}=\int_0^\infty P\{T_1>t \mid T_2=t\} f_{T_2}(t)dt = \int_0^\infty e^{-\lambda_1 t}\cdot\lambda_2 e^{-\lambda_2 t}dt$$
$$=\frac{\lambda_2}{\lambda_1+\lambda_2}$$

② Pdf for the time until a message arrive on either line

$$F_T(t)=1-P\{T_1>t \text{ and } T_2>t\}=1-P\{T_1>t\}P\{T_2>t\}=1-e^{-(\lambda_1+\lambda_2)t}$$

③ Pmf for $N(t)$: $P[N(t)=k]=\frac{\{(\lambda_1+\lambda_2)t\}^k}{k!}e^{-(\lambda_1+\lambda_2)t}$.

B. Gaussian random process

$$f_{X_1,X_2,\dots X_k}(x_1,x_2,\dots,x_k)=\frac{\exp\{-\frac{1}{2}(\bar{x}-\bar{m})^T K^{-1}(\bar{x}-\bar{m})\}}{(2\pi)^{\frac{k}{2}}|K|^{\frac{1}{2}}}$$

where $\bar{m}=\begin{bmatrix} m_x(t_1) \\ \vdots \\ m_x(t_k)\end{bmatrix}$ $\quad K=\begin{bmatrix} C_x(t_1,t_1) & C_x(t_1,t_2) & \cdots & C_x(t_1,t_k) \\ C_x(t_2,t_1) & C_x(t_2,t_2) & \cdots & C_x(t_2,t_k) \\ \vdots & & & \\ C_x(t_k,t_1) & \cdots & & C_x(t_k,t_k)\end{bmatrix}$

Note: Gaussian r.p. have special property that their joint pdf's are completely specified by the mean function of the process $m_x(t)$ and by the variance function $C_x(t_1,t_2)$.

C. Wiener process (symmetric random walk process which takes steps of magnitude $\pm h$ every δ seconds) δ ($p=\frac{1}{2}$)

At time t, the process will have taken $n=[t/\delta]$ jumps,

$$X_\delta(t) = h(D_1 + D_2 + \cdots + D_{[t/\delta]}) = hS_n.$$

where $X_\delta(t)$ is the accumulated sum of the random step process up to time t.

$\Rightarrow E[X_\delta(t)] = hE[S_n] = 0$, $var[X_\delta(t)] = h^2 n\, var[D_n] = h^2 n$.

$\because var[D_n] = p\cdot[1-(2p-1)]^2 + (1-p)[-1-(2p-1)]^2 = 4p(1-p) = 1$

Let $\delta \to 0$, $h \to 0$, $h = \sqrt{\alpha\delta}$: $E[X(t)] = 0$. $var[X(t)] = (\sqrt{\alpha\delta})^2 (t/\delta) = \alpha t$.

$X(t) = \lim_{\delta\to 0} hS_n = \lim_{n\to\infty} \sqrt{\alpha t}\,\frac{S_n}{\sqrt{n}}$. ($X(t)$ approaches the sum of an infinite number of random variables.)

By central limit theorem: $f_{X(t)}(x) = \dfrac{1}{\sqrt{2\pi\alpha t}}\, e^{-\frac{x^2}{2\alpha t}}$

$$C_X(t_1, t_2) = \alpha \min(t_1, t_2) = \alpha t_1 \text{ for } t_1 < t_2$$

Note: Suppose S_n is the sum of n i.i.d radom variables X.

$m_S(n) = E[S_n] = nE[X] = nm$. $Var[S_n] = n\,var[X] = n\delta^2$

Suppose $n \le k$, so $n = \min(n,k)$. Then

$C_S(n,k) = E[(S_n - nm)(S_k - km)]$

$= E[(S_n - nm)\{(S_n - nm) + (S_n - km) - (S_n - nm)\}]$

$= E[(S_n - nm)^2] + E[(S_n - nm)(S_k - S_n - (k-n)m)]$.

Independent increments. $= E[(S_n - nm)^2] + E[(S_n - nm)]\,E[(S_k - S_n - (k-n)m)]$

$= E[(S_n - nm)^2] = Var[S_n] = n\delta^2$

$\therefore C_S(n,k) = \min(n,k)\,\delta^2$.

D. stationary random process.

$\boxed{\text{stationary}}$: $F_{X(t_1),\dots,X(t_k)}(x_1,\dots,x_k) = F_{X(t_1+\tau),\dots,X(t_k+\tau)}(x_1,\dots,x_k)$

doesn't depend on the placement of the time origin.

$\Downarrow \not\Uparrow$ $\qquad\qquad$ $\Downarrow \not\Uparrow$

$\boxed{\text{first order stationary}}$ $\quad \not\rightleftharpoons \quad$ $\boxed{\text{second order stationary}}$

$F_{X(t)}(x) = F_{X(t+\tau)}(x) = F_X(x)$ all t, τ

$\Downarrow$

mean stationary

$F_{X(t_1),X(t_2)}(x_1,x_2) = F_{X(0),X(t_2-t_1)}(x_1,x_2)$ for all t_1, t_2

$\Downarrow$

$R_X(t_1,t_2) = R_X(t_2-t_1)$ for all t_1, t_2

$C_X(t_1,t_2) = C_X(t_2-t_1)$ for all t_1, t_2.

Wide sense stationary Random process $\boxed{\text{WSS}}$

$m_X(t) = m$ for all t and $C_X(t_1,t_2) = C_X(t_1-t_2)$

For WSS process: Average power. $R_X(0) = E[X(t)^2]$ for all t.

$R_X(\tau) = E[X(t+\tau)X(t)] = E[X(t)X(t+\tau)] = R_X(-\tau)$

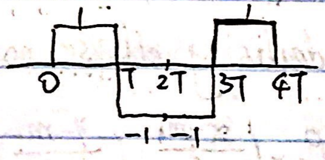
$|R_X(\tau)| \le R_X(0)$ [According to. $E[XY]^2 \le E[X^2]E[Y^2]$].

* If a Gaussian r.p is WSS, then it is also stationary.

$\boxed{\text{Cyclostationary}}$: $F_{X(t_1)X(t_2)\dots X(t_k)}(x_1,x_2,\dots x_k) = F_{X(t_1+mT),X(t_2+mT),\dots X(t_k+mT)}(x_1,x_2,\dots x_k)$

$\boxed{\text{WS cyclostationary}}$: $m_X(t+mT) = m_X(t)$, $C_X(t_1+mT, t_2+mT) = C_X(t_1,t_2)$

Exp: waveform:



$m_X(t) = 0$

$C_X(t_1,t_2) = \begin{cases} E[X(t_1)^2] = 1 & nT < t_1, t_2 < (n+1)T \\ E[X(t_1)]E[X(t_2)] = 0 & \text{otherwise} \end{cases}$

8. $\boxed{\text{Analysis and processing of random signals}}$ (WSS r.p.)

$E \triangleq \lim_{T\to\infty} \int_{-T}^{T} |x(t)|^2 dt = \int_{-\infty}^{\infty} |x(t)|^2 dt$. $\qquad P \triangleq \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} |x(t)|^2 dt$.

Note: ① $x(t)$ is an energy signal if and only if $0 < E < \infty$ so that $P = 0$

② $x(t)$ is a power signal if and only if $0 < P < \infty$, thus implying $E = \infty$.

Energy spectral Density.

$E \triangleq \int_{-\infty}^{\infty} |x(t)|^2 dt$

$= \int_{-\infty}^{\infty} x^*(t) \left[\int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df\right] dt$

$= \int_{-\infty}^{\infty} X(f) \left[\int_{-\infty}^{\infty} x^*(t) e^{j2\pi ft} dt\right] df$

$= \int_{-\infty}^{\infty} X(f) X^*(f) df = \int_{-\infty}^{\infty} |X(f)|^2 df$

$G(f) = |X(f)|^2.$

Power spectral Density.

$S_X(f) = \lim_{T\to\infty} E[\tilde{p}_T(f)] = \lim_{T\to\infty} \frac{1}{T} E[|\tilde{X}(f)|^2]$

$S_X(f) = \mathcal{F}\{R_X(\tau)\} = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f\tau} d\tau.$

Property: ① real-valued and even funct of f

② nonnegative

Average power.

$E[X^2(t)] = R_X(0) = \int_{-\infty}^{\infty} S_X(f) df.$

$S_X(f) = \mathcal{F}\{C_X(\tau)\} + m_X^2 \delta(f).$

↳ "dc" component of $X(t)$.

Cross-power spectral density $S_{XY}(f)$

$S_{X,Y}(f) = \mathcal{F}\{R_{X,Y}(\tau)\}.$

Einstein-Wiener-khinchin theorem:

Let $X_0, \dots X_{k-1}$ be k observations from discrete time, WSS process X_n.

$\tilde{X}(f)$ denote the discrete Fourier transform of this sequence.

$$\tilde{X}_k(f) = \sum_{m=0}^{k-1} X_m e^{-j2\pi fm}.$$

The magnitude squared of $\tilde{X}_k(f)$ is a measure of the "energy" at the frequency f. If we divide this energy by the total time k, we obtain an estimate for the "power" at the frequency f.

$$\tilde{p}_k(f) = \frac{1}{k} |\tilde{X}_k(f)|^2 \cdot \text{(periodogram estimate for psd).}$$

$\therefore E[\tilde{p}_k(f)] = \frac{1}{k} E[\tilde{X}_k(f) \tilde{X}_k^*(f)]$

$= \frac{1}{k} E\left[\sum_{m=0}^{k-1} X_m e^{-j2\pi fm} \sum_{i=0}^{k-1} X_i e^{j2\pi fi}\right]$

$= \frac{1}{k} \sum_{m=0}^{k-1} \sum_{i=0}^{k-1} E[X_m X_i] e^{-j2\pi f(m-i)}$

$= \frac{1}{k} \sum_{m=0}^{k-1} \sum_{i=0}^{k-1} R_X(m-i) e^{-j2\pi f(m-i)}.$

$m' = m - i \Rightarrow \quad = \frac{1}{k} \sum_{m'=-(k-1)}^{k-1} \{k - |m'|\} R_X(m') e^{-j2\pi fm'}$

$$= \sum_{m'=-(k-1)}^{k-1} \left\{1 - \frac{|m'|}{k}\right\} R_X(m') e^{-j2\pi f m'}$$

$E[\tilde{p}_k(f)] \to S_X(f)$ as $k \to \infty$.

For continous: $E[\tilde{p}_T(f)] = \int_{-T}^{T} \left\{1 - \frac{|\tau|}{T}\right\} R_X(\tau) e^{-j2\pi f \tau} d\tau$.

$E[\tilde{p}_T(f)] \to S_X(f)$ as $T \to \infty$.

Continuous-Time systems.

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

impulse response $h(t)$: $h(t) = T[\delta(t)]$

output: $y(t) = h(t) * x(t) = \int_{-\infty}^{\infty} h(s) x(t-s) ds = \int_{-\infty}^{\infty} h(t-s) x(s) ds$.

$E[Y(t)] = E\left[\int_{-\infty}^{\infty} h(s) X(t-s) ds\right] = \int_{-\infty}^{\infty} h(s) E[X(t-s)] ds = m_X H(0)$.

$R_Y(\tau) = E[Y(t) Y(t+\tau)] = E\left[\int_{-\infty}^{\infty} h(s) X(t-s) ds \int_{-\infty}^{\infty} h(r) X(t+\tau-r) dr\right]$.

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(s) h(r) E[X(t-s) X(t+\tau-r)] ds\, dr$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(s) h(r) R_X(\tau+s-r) ds\, dr.$$

$S_Y(f) = \int_{-\infty}^{\infty} R_Y(\tau) e^{-j2\pi f \tau} d\tau$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(s) h(r) R_X(\tau+s-r) e^{-j2\pi f \tau} ds\, dr\, d\tau.$$

let $u = \tau + s - r$ $$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(s) h(r) R_X(u) e^{-j2\pi f (u-s+r)} ds\, dr\, du.$$

$$= \int_{-\infty}^{\infty} h(s) e^{j2\pi f s} ds \int_{-\infty}^{\infty} h(r) e^{-j2\pi f r} dr \int_{-\infty}^{\infty} R_X(u) e^{-j2\pi f u} du$$

$$= H^*(f) H(f) S_X(f) = |H(f)|^2 S_X(f)$$

$R_{Y,X}(\tau) = E[Y(t+\tau) X(t)]$ $\quad (R_{X,Y}(\tau) = E[X(t+\tau) Y(t)]$.

$$= E\left[X(t) \int_{-\infty}^{\infty} X(t+\tau-r) h(r) dr\right]$$

$$= \int_{-\infty}^{\infty} E[X(t) X(t+\tau-r)] h(r) dr$$

$$= \int_{-\infty}^{\infty} R_X(\tau-r) h(r) dr.$$

$$= R_X(\tau) * h(\tau).$$

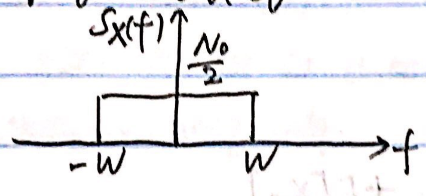
$\therefore S_{Y,X}(f) = H(f) S_X(f)$.

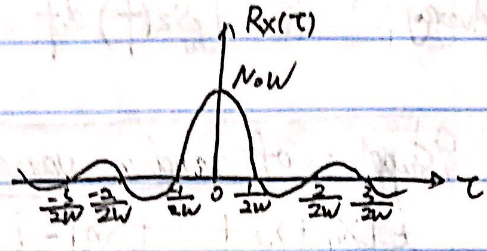
$\because R_{X,Y}(\tau) = R_{Y,X}(-\tau)$

$\therefore S_{X,Y}(f) = S^*_{Y,X}(f) = H^*(f) S_X(f)$

In general, $S_{X,Y}(f)$ is a complex function of f even if $X(t)$ and $Y(t)$ are both real-valued.

$\boxed{\text{White noise}}$ ("white" in analogy to white light, which contains all frequencies in equal amount).





$E[X^2(t)] = \int_{-W}^{W} \frac{N_0}{2} df = N_0 W$ $\quad R_X(\tau) = \frac{1}{2} N_0 \int_{-W}^{W} e^{j2\pi f\tau} df$

$$= \frac{1}{2} N_0 \frac{e^{-j2\pi W\tau} - e^{j2\pi W\tau}}{-j2\pi\tau}$$

$$= \frac{N_0 \sin(2\pi W\tau)}{2\pi\tau}$$

$X(t)$ and $X(t+\tau)$ are uncorrelated at $\tau = \frac{\pm k}{2W}$.

The term white noise refers to a random process $W(t)$ whose power spectral density is $N_0/2$ for all frequencies.

$$S_W(f) = \frac{N_0}{2} \quad \text{for all } f.$$

$\therefore W = \infty$ shows such a process must have infinite average power.

Here $R_W(\tau) = \frac{N_0}{2}\delta(\tau)$ (approximation).

which means $R_W(\tau) = \begin{cases} \sigma_W^2 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases}$ where $\sigma_W^2 = \frac{N_0}{2}$.

$\therefore S_W(f) = \mathcal{F}\{R_W(\tau)\} = \int_{-\infty}^{\infty} R_W(\tau) e^{-j2\pi f\tau} d\tau = \sigma_W^2$.

Physical mean of cross power spectral density. To compute the power (psd) of a process which is sum of two processes.

Exp: Psd of $Z(t) = X(t) + Y(t)$, where $X(t)$ and $Y(t)$ are jointly wss process.

$\therefore R_Z(\tau) = E[Z(t+\tau)Z(t)] = E[(X(t+\tau)+Y(t+\tau))(X(t)+Y(t))]$

$= R_X(\tau) + R_{YX}(\tau) + R_{XY}(\tau) + R_Y(\tau)$

Psd: $S_Z(f) = \mathcal{F}\{R_X(\tau) + R_{YX}(\tau) + R_{XY}(\tau) + R_Y(\tau)\}$

$= S_X(f) + S_{YX}(f) + S_{XY}(f) + S_Y(f)$.

Average power: $P_Z = \int_{-\infty}^{\infty} S_Z(f)\,df$

9. $\boxed{\text{Sum of random variables.}}$

$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$

$var(X_1 + X_2 + \cdots + X_n) = E\{\sum_{j=1}^{n}(X_j - E[X_j]) \sum_{k=1}^{n}(X_k - E[X_k])\}$

$= \sum_{k=1}^{n} var(X_k) + \sum_{j=1}^{n}\sum_{\substack{k=1 \\ j\neq k}}^{n} cov(X_j, X_k)$

$\boxed{\text{Sum of a random number } N \text{ of iid random variables.}}$

$S = \sum_{k=1}^{N} X_k$. (N is independent of X_k).

$E[S] = E[E[S|N]]$

$= E[N E[X]]$

$= E[N]E[X]$

$\because E[S^2|N] = E[\sum_{i=1}^{N} X_i \sum_{j=1}^{N} X_j] = \sum_{i=1}^{N}\sum_{j=1}^{N} E[X_i X_j] = N E[X^2] + (N^2 - N)E[X]^2$.

$\therefore E[S^2] = E[E[S^2|N]] = E[N]E[X^2] + (E[N^2] - E[N])E[X]^2$

$\therefore var[S] = E[S^2] - E[S]^2$

$= E[N]E[X^2] + E[N^2]E[X]^2 - E[N]E[X]^2 - E[N]^2 E[X]^2$

$= E[N]var(X) + var(N)E[X]^2$.

10. $\boxed{\text{Sample mean and LLN}}$

$M_n = S_n/n = (X_1+X_2+\cdots+X_n)/n$.

$E[M_n] = E[\frac{1}{n}\sum_{j=1}^{n} X_j] = \frac{1}{n}\sum_{j=1}^{n} E[X_j] = u$.

$Var[M_n] = \frac{1}{n^2} var[S_n] = \frac{n\delta^2}{n^2} = \frac{\delta^2}{n}$

$\Rightarrow$ ① variance of the sample mean approaches zero as the number of sample is increased.

② The probability that sample mean is close to the true mean approaches one as n becomes very large.

chebyshev inequality. $P[|M_n - E[M_n]| \geq \varepsilon] \leq \frac{var[M_n]}{\varepsilon^2} = \frac{\delta^2}{n\varepsilon^2}$.

$$\Rightarrow P[|M_n - u| < \varepsilon] \geq 1 - \frac{\delta^2}{n\varepsilon^2}.$$

we can select the number of samples n so that M_n is within ε of the true mean with probability $1-\delta$ or greater.

(sequence of sample means: $M_1, M_2, \ldots, M_j, \ldots$

A. Weak Law of Large Numbers: $\lim_{n\to\infty} P[|M_n - u| < \varepsilon] = 1$.

B. Strong Law of Large Numbers: $P[\lim_{n\to\infty} M_n = u] = 1$

Note: ① The weak law of large numbers doesn't address the question about what happens to the sample mean as a function of n. It states that for a large enough fixed value of n, the sample mean using n samples will be close to the true mean with high probability.

②. The strong law of large number states that with probability 1, every sequence of sample mean calculations will eventually approach and stay close to $E[X] = u$.

C. Central Limit Theorem: Let S_n be the sum of n iid random variables with finite mean $E[X]=u$ and finite variance δ^2, and

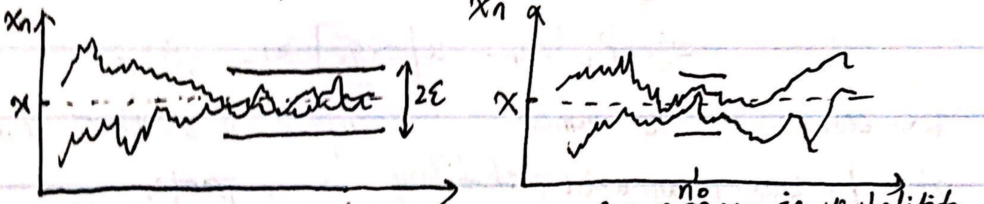
Let Z_n be the zero-mean, unit-variance random variable defined by

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad \text{or} \quad Z_n = \sqrt{n}\,\frac{M_n - \mu}{\sigma} \text{ (sample mean)}.$$

Note: CLT is the summands X_j can have any distribution as long as they have a finite mean and finite variance.

11. Brief summary of convergence of sequence of random varibbles.

A sequence of random variables X is a function that assigns a countably infinite number of real values to each outcome ξ from some sample spaces.



Almost-sure convergence　　convergence in probability

$\boxed{\text{Almost sure convergence:}}$ $P[\xi: X_n(\xi) \to X(\xi) \text{ as } n \to \infty] = 1$: The sample sequences convergence to the same value. Almost all sequences must eventually enter and remain inside a 2ε corridor. In almost-sure convergence some of the sample sequences may not converge, but these must all belong to ξ's that are in a set that has probability zero.

$\boxed{\text{Convergence in probability}}$: $P[|X_n(\xi) - X(\xi)| > \varepsilon] \to 0$ as $n \to \infty$: At specified time n_0 most sample sequences must be within ε of x. However, the sequences are not required to remain inside a 2ε corridor.

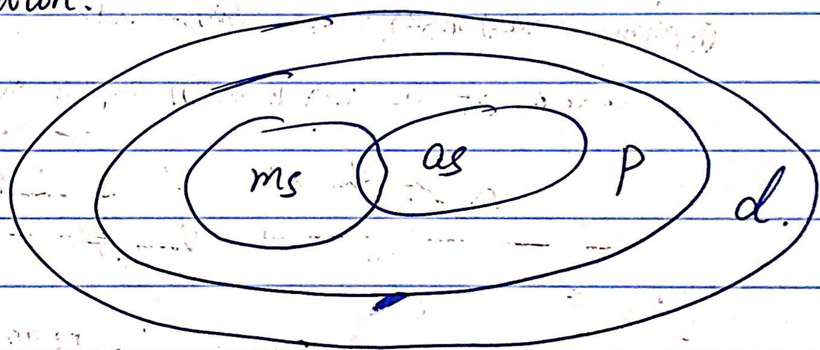
$\boxed{\text{Mean square Convergence:}}$ $E[(X_n(\xi) - X(\xi))^2] \to 0$ as $n \to \infty$: We may require that at particular time n_0, most sample sequences X_{n_0} be close to X in the sense that $E[(X_{n_0} - X)^2]$ is small. It doesn't address the behavior of entire sample sequences. (unlike almost-sure convergence).

$\boxed{\text{Covergence in distribution}}$ $F_n(x) \to F(x)$ as $n \to \infty$ for all x at which $F(x)$ is continuous.

Note. ① The strong law of large numbers is an example of almost sure convergence.

② The weak law of large numbers is an example of covergence in probability.

③ The central limit theorem is an example of covergence in distribution.



MLE与MAP的联系

在极大似然估计(MLE)中，我们求参数θ，通过使$p(x|\theta)$最大，此时θ为一个待估参数，其本身确定，即使目前未知。MLE是求怎样的参数θ可以让事件集发生的概率最大。通过不断改变固定的参数θ去寻找一个极大值

$$\arg\max_{\theta} \prod_{k=1}^{N} p(x_k|\theta).$$

在最大后验估计(MAP)中，将参数θ看成一个随机变量。MAP考量的是事件集X已经发生了，那在事件集发生的情况下，哪个θ发生的概率最大。与MLE正好相反

$$p(\theta|X) = \frac{p(\theta)\, p(X|\theta)}{p(X)}$$

Exp：抛一枚不均匀的硬币，抛了10次，9次正1次反，抛出正面的概率是？

MLE：假设抛出正面的概率为p，该事件概率为$p^9(1-p)$，求该事件发生概率的极大值所对应的p，此时p仅为一个函数的参数，而非随机变量。

MAP：通过求解最大的$p(\theta|X)$所对应的θ，X为"抛了10次，9次正1次反"这个事件，求在X已经发生的情况下，使得$p(\theta|X)$最大的θ。此时θ为一个随机变量。

MAP (Maximum A Posterior)

MLE是θ的函数，其求解过程是找到使得最大似然函数最大的参数θ。MAP将θ看成一个随机变量，并在已知样本集$\{x_1, x_2, \dots x_N\}$的条件下，估计θ。（Note：在MLE中，θ是一个定值，只是这个值未知，MLE是θ的函数，θ没有概率意义的。MAP中，θ是有概率意义的，θ有自己的分布，而这个分布函数需要通过已有样本集合X得到，即

MAP需要计算的是 $p(\theta|X)$

$\because$ $p(\theta|X) = \frac{p(\theta)\,p(X|\theta)}{p(X)}$

$\therefore$ $\underset{\theta}{\operatorname{argmax}}\, p(\theta|X) = \underset{\theta}{\operatorname{argmax}}\, p(\theta)\,p(X|\theta)$.

为了得到 θ，对 $p(\theta|X)$ 求导：$\frac{\partial p(\theta|X)}{\partial\theta} = \frac{\partial p(\theta)p(X|\theta)}{\partial\theta} = 0$.

Note. 这里 $p(X|\theta)$ 与 MLE中 $p(X|\theta)$ 是一样的。MAP与MLE区别：MAP是在MLE基础上加上 $p(\theta)$。从公式上看 MAP $= p(\theta) *$ MLE，但两种算法有本质区别，MLE将 θ 视为确定的未知值，MAP将 θ 视为一个随机变量。在MAP中，$p(\theta)$ 称为 θ 的先验，假设其服从均匀分布，即对于所有 θ 取值，$p(\theta)$ 都是同一变量，MAP和MLE会得到相同的结果。如果 $p(\theta)$ 方差非常小，$p(\theta)$ 近似均匀分布，MAP与MLE结果也非常接近。

Exp：袋1：cherry 100%。袋2：75% cherry + 25% Lemon。袋3：50% cherry + 50% lemon
袋4：25% cherry + 75% lemon 袋5：100% lemon.

从同一个袋子连续拿到2个lemon，最有可能是哪一个袋子？

MLE：$p(2个\text{lemon}|袋子) = p^2$。（$p$ 为从袋中拿到 lemon 的概率）.
$\Rightarrow$ 袋5.

上述MLE没有考虑模型本身的概率

假设拿到袋1或5的概率为0.1，袋2或4的概率为0.2，拿到3的概率为0.4.

MAP：$\underset{\theta}{\operatorname{argmax}}\, f(X|\theta)\,g(\theta) = p^2 g$. where $g(0.1, 0.2, 0.4, 0.2, 0.1)$

袋1：0
袋2：0.0125
袋3：0.125
袋4：0.28125
袋5：0.1

$\Rightarrow$ 袋4.

Note：MAP与MLE最大区别是MAP中加入了模型本身的概率分布，或者说，MLE中认为模型本身的概率是均匀的，即该概率为一个固定值。一个合理的先验概率假设很重要，否则会极大影响后验概率的结果。