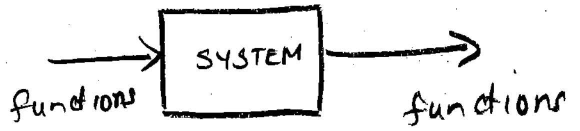
1

ITCSC- INC Winter School Day 2

Fourier Analysis in Communications & Probability Theory

System



$x(t)$ $y(t)$ $\underline{y}(t) = H[\underline{x}(t)]$

$t \to$ variable "signal)

Properties of Systems

a) Linearity : mapping defined by the system is a linear mapping

$$H[\alpha\, \underline{x}_1(t) + \beta\, \underline{x}_2(t)] = \alpha\, H[\underline{x}_1(t)] + \beta\, H[\underline{x}_2(t)]$$

b) Time Invariance:

$$H[x(t-t_0)] = \underline{y}(t-t_0) \quad \forall t_0$$

where $\underline{y}(t) = H[\underline{x}(t)]$

c) Causality

if $x_1(t) = x_2(t) \quad \forall t \le t_0$

then $y_1(t) = y_2(t) \quad \forall t \le t_0$

where $\underline{y}_1(t) = H[\underline{x}_1(t)]$

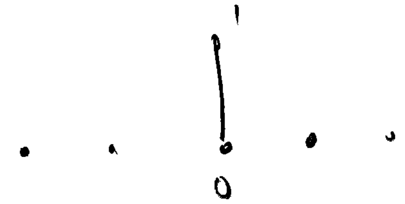
$\underline{y}_2(t) = H[\underline{x}_2(t)]$

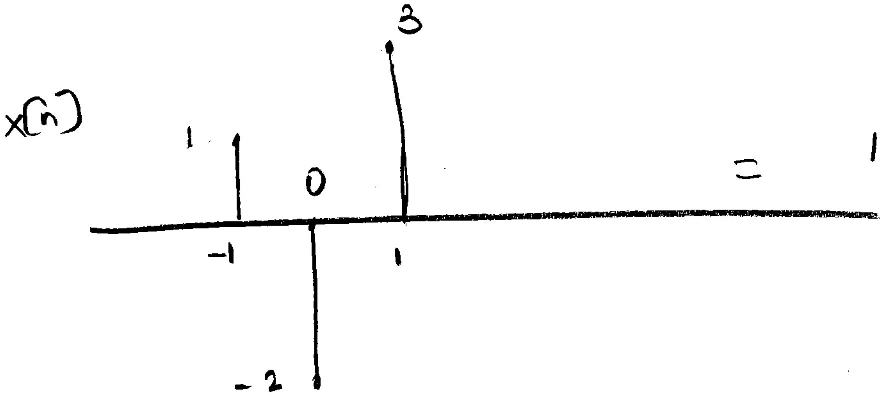
LTI Systems : Systems that are both linear & time-invariant

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = a\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$x[n] = \sum_{m=-\infty}^{\infty} x[m]\,\delta[n-m]$$

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$$





$$= 1 \times \delta[n+1] - 2\,\delta[n] + 3\,\delta[n-1]$$

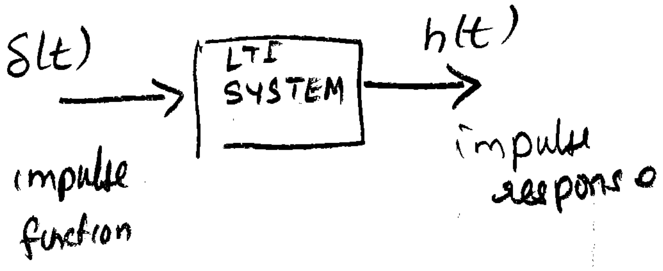
In continuous time

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\,\delta(t-\tau)\,d\tau$$

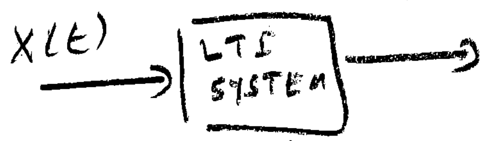
$\hookrightarrow$ Dirac delta distribution

$$\int_{I_\varepsilon} f(t)\,\delta(t)\,dt = \begin{cases} 0, & 0 \notin I \\ f(0), & 0 \in I \end{cases}$$

$$\int_{-\varepsilon} f(t)\,\delta(t)\,dt = f(0) \quad \forall \varepsilon > 0$$



then



$$\underline{x}(t) = \int_{-\infty}^{\infty} x(\tau)\, \underline{\delta}(t-\tau)\, d\tau$$

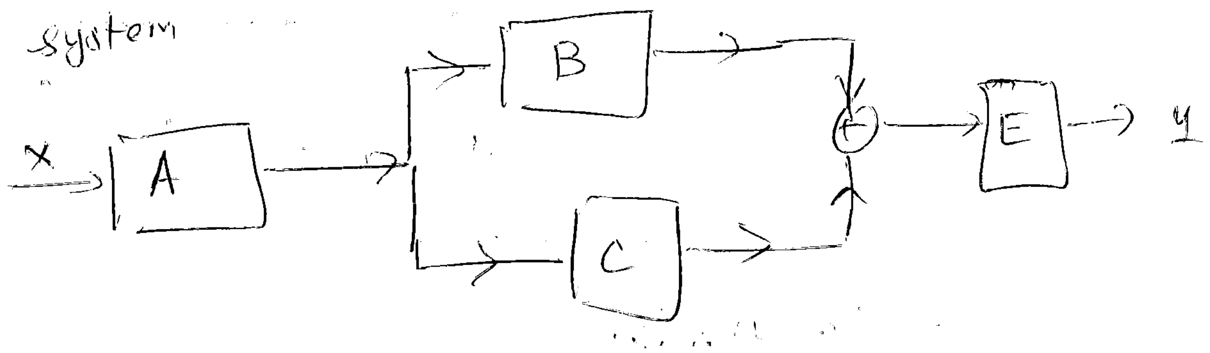
$$\therefore\ H[\underline{x}(t)] = \int_{-\infty}^{\infty} x(\tau)\, H[\underline{\delta}(t-\tau)]\, d\tau$$

$$\underline{y}(t) = \int_{-\infty}^{\infty} x(\tau)\, h(t-\tau)\, d\tau$$

convolution

---

system



$$\underline{y} = E(B+C)Ax$$

$x \in \mathbb{R}^n$

$A, B, C, E \in \mathbb{R}^{n\times n}$

symmetric matric

$$A = Q_A \Lambda_A Q_A^T$$
$$B = Q_B \Lambda_B Q_B^T$$
$$C = Q_C \Lambda_C Q_C^T$$
$$D = Q_D \Lambda_D Q_D^T$$

Eigen decomposition of real symmetric matrix

$$Ax = \lambda x$$

$$A x_1 = \lambda_1 x_1$$
$$A x_2 = \lambda_2 x_2$$

$$x_2^T A x_1 = \lambda_1 (x_2^T x_1)$$
$$= \lambda_2 (x_2^T x_1)$$
$$\cancel{(x_2 A)} (A x_2)^T x_1 = \lambda_2 x_2^T x_1$$

$$Q = x_1 \; . \; . \; x_n$$

$$A \underbrace{\begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}}_{Q} = \begin{bmatrix} \lambda_1 x_1 & \cdots & \lambda_n x_n \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_n \end{bmatrix}$$

$$AQ = Q\Lambda$$
$$AQQ^T = Q\Lambda Q^T \qquad QQ^T = I$$
$$A = Q\Lambda Q^T$$

$$y = Q\Lambda_E Q^T (Q\Lambda_C Q^T + Q\Lambda_B Q^T) Q\Lambda_A Q^T$$
$$= Q \Lambda_E (\Lambda_C + \Lambda_B) \Lambda_A Q^T x$$

LTI systems share a same set of eigenvectors

~~Fourie~~ Eigenfunction of a system

$e(t) \longrightarrow$ [system] $\xrightarrow{\lambda_e e(t)}$

$e^{j2\pi ft} \longrightarrow$ [LTI system, $h(t)$] $\longrightarrow$

$$\int_{-\infty}^{\infty} h(\tau) e^{j2\pi f(t-\tau)} d\tau$$
$$= e^{j2\pi ft} \cdot \underbrace{\int_{-\infty}^{\infty} h(\tau) e^{-j2\pi f\tau} d\tau}_{\hat{H}(f)}$$

$h(t)$ is a signal then

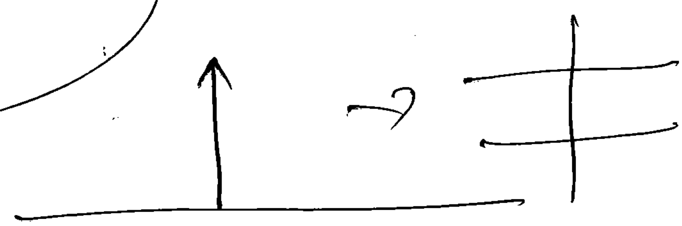
$$\hat{H}(f) = \int_{-\infty}^{\infty} h(t)\, e^{-j2\pi f t}\, dt$$

Fourier Inversion

$$h(t) = \int_{-\infty}^{\infty} \hat{H}(f)\, e^{j2\pi f t}\, df$$

Parseval's theorem

$$\int_{-\infty}^{\infty} |h(t)|^2 dt = \int_{-\infty}^{\infty} |\hat{H}(f)|^2 df$$



If $h(t) = \delta(t)$

$$\hat{H}(f) = \int_{-\infty}^{\infty} \delta(t)\, e^{-j2\pi f t}\, dt = 1$$

$$x(t) = e^{j2\pi f_0 t}$$

$\cancel{\hat{x}(f)}$

Duality principle

$$\text{F.T.}\left[\text{F.T.}\left[x(t)\right]\right] = x(-t)$$

$$\hat{x}(\sigma) = \int x(t)\, e^{-j2\pi f t}\, dt$$

$$\hat{\hat{x}}(\tau) = \int \hat{x}(\sigma)\, e^{-j2\pi\sigma\tau}\, d\sigma$$

$$\hat{\hat{x}}(-t) = \int \hat{x}(\sigma)\, e^{j2\pi\sigma t}\, d\sigma = x(t)$$

Guess $\boxed{\hat{X}(f) = \delta(f - f_0)}$

$$x(t) = \int_{-\infty}^{\infty} \hat{X}(f)\, e^{j2\pi f t}\, df = e^{j2\pi f_0 t}$$

$$h(t) = \int_{-\infty}^{\infty} \hat{H}(f)\, e^{j2\pi f t}\, df$$

Parseval's Theorem

$$\int_{-\infty}^{\infty} |h(t)|^2 dt = \int |\hat{H}(f)|^2 df$$

$h^* \rightarrow$ complex conjugate

Pf

$$\int_{-\infty}^{\infty} h(t)\, h^*(t)\, dt$$

$$= \int_{-\infty}^{\infty} h(t) \left[\int_{-\infty}^{\infty} \hat{H}(f)\, e^{j2\pi f t}\, df \right]^* dt$$

$$= \int_{-\infty}^{\infty} h(t) \left(\int_{-\infty}^{\infty} \hat{H}^*(f)\, e^{-j2\pi f t}\, df \right) dt$$

$$= \int_{-\infty}^{\infty} \hat{H}^*(f) \left[\int_{-\infty}^{\infty} h(t)\, e^{-j2\pi f t}\, dt \right] df$$

$$= \int_{-\infty}^{\infty} \hat{H}^*(f)\, \hat{H}(f)\, df$$

$$\int_A \left[\int_B f(x,y)\, dx \right] dy = \int_B \left[\int_A f(x,y)\, dy \right] dx$$

When can I do this (Fubini's theorem)

$f \geq 0$ or $\int_A \int_B |f(x,y)|\, dx\, dy < \infty$

$x^2(t)$ $\int_{-\infty}^{\infty} |x(t)|^2 dt = 1$

Heisenberg's inequality for Fourier Transform. $\int_{-\infty}^{\infty} |\hat{x}(f)|^2 df = 1$

$$\left(\int_{-\infty}^{\infty} t^2 |x^2(t)|^2 dt\right) \int \left(\int_{-\infty}^{\infty} f^2 |\hat{x}^2(f)|^2 df\right) \geq \frac{1}{16\pi^2}$$

$t\, x^2(t)\, dt \to 0$ as $|t| \to \infty$

Cauchy - Schwartz inequality

vector space

$$|\langle a, b\rangle|^2 \leq \langle a, a\rangle \langle b, b\rangle$$

$\langle a - \alpha b, a - \alpha b\rangle \geq 0$ Take $\alpha = \frac{\langle a, b\rangle}{\langle b, b\rangle}$

$$\left|\left(\int_{-\infty}^{\infty} f(t) g^*(t)\, dt\right)\right|^2 \leq \int_{-\infty}^{\infty} |f(t)|^2 dt \int_{-\infty}^{\infty} |g(t)|^2 dt$$

$$1 = \int_{-\infty}^{\infty} x^2(t)\, dt = -\int_{-\infty}^{\infty} 2t\, x(t)\, x'(t)\, dt$$

$$\because \int_{-\infty}^{\infty} \frac{d}{dt}\left(t\, x^2(t)\right) dt = 0$$

$$\Rightarrow 4 \int_{-\infty}^{\infty} t^2 |x(t)|^2 dt \int_{-\infty}^{\infty} |x'(t)|^2 dt \geq 1$$

$$x(t) = \int_{-\infty}^{\infty} \hat{x}(f)\, e^{j2\pi f t}\, df$$

$$x'(t) = \int_{-\infty}^{\infty} \left(j2\pi f\, \hat{x}(f)\right) e^{j2\pi f t}\, df$$

$$\Rightarrow 16\pi^2 \int_{-\infty}^{\infty} t^2 |x(t)|^2 dt \int f^2 |\hat{x}(f)|^2 df \geq 1$$

Discrete Fourier Transform

$$\left(x[0] \; \ldots \; x[N-1]\right) \xleftrightarrow{DFT} \left(\hat{x}(0) \; \ldots \; \hat{x}(N-1)\right)$$

$$\hat{x}(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n]\, e^{-j 2\pi \frac{nk}{N}} \qquad \text{(DFT)}$$

$$x(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{x}(k)\, e^{j 2\pi \frac{nk}{N}} \qquad \text{inverse}$$

Prove Heisenberg's inequality for DFT

support $(x[0], \ldots, x[N-1])$ = no. of non-zero elements

$$\text{supp}(\underline{x})\ \text{supp}(\underline{\hat{x}}) \geq N \qquad x \neq 0$$

~~Proof~~ Step 1: Let $N_t = |\text{supp}(x)|$

Claim: $\hat{x}(k)$ cannot have N_t consecutive zero's

If Claim holds then $N_f \cdot (N_t - 1) + N_f \geq N$

suppose non-zero elements of x are $x[n_1], \ldots, x[n_{N_t}]$

~~So~~ Assume $\hat{x}[k], \ldots, \hat{x}(k+N_t-1)$ are zero

$$\begin{bmatrix} \hat{x}[k] \\ \vdots \\ \hat{x}[k+N_t-1] \end{bmatrix} = \begin{bmatrix} e^{-j2\pi \frac{n_1 k}{N}} & \cdots & e^{-j2\pi \frac{n_{N_t} k}{N}} \\ & & \\ e^{-j2\pi \frac{n_1 (k+N_t-1)}{N}} & \cdots & \end{bmatrix} \begin{bmatrix} x[n_1] \\ \vdots \\ x[n_{N_t}] \end{bmatrix}$$

Van der Monde

$$\begin{bmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{bmatrix}$$

Fourier Series

here $x(t)$ is a periodic function with period T

Express (under certain mild conditions)

$$x(t) = \sum_{k=-\infty}^{\infty} a_k e^{-j\frac{2\pi kt}{T}}$$

$$\langle x_1(t), x_2(t)\rangle = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x_1(t)\, x_2^*(t)\,dt$$

$$\left\langle e^{j\frac{2\pi kt}{T}}, e^{j\frac{2\pi lt}{T}} \right\rangle = \begin{cases} 1, & k = l \\ 0, & \text{o.w.} \end{cases}$$

Parseval theorem says

$$\frac{1}{T}\int_{-\infty}^{\infty} |x(t)|^2 dt = \sum_{k=-\infty}^{\infty} |a_k|^2$$

Use this to prove

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \cdots = \frac{\pi^2}{6}$$

$$1 + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \cdots = \frac{\pi^4}{90}$$

$$1 + \frac{1}{2^6} + \frac{1}{3^6} + \cdots = \frac{\pi^6}{945}$$

$$1 + \frac{1}{2^3} + \frac{1}{3^3} + \cdots = \text{Apery's constant}$$

$$\eta(s) = \sum_{n=1} \frac{1}{n^s} \quad ; \ Re(s) > 1$$

Riemann Zeta function

All the non-trivial zero's of the zeta function lie on the line $Re(s) = \frac{1}{2}$

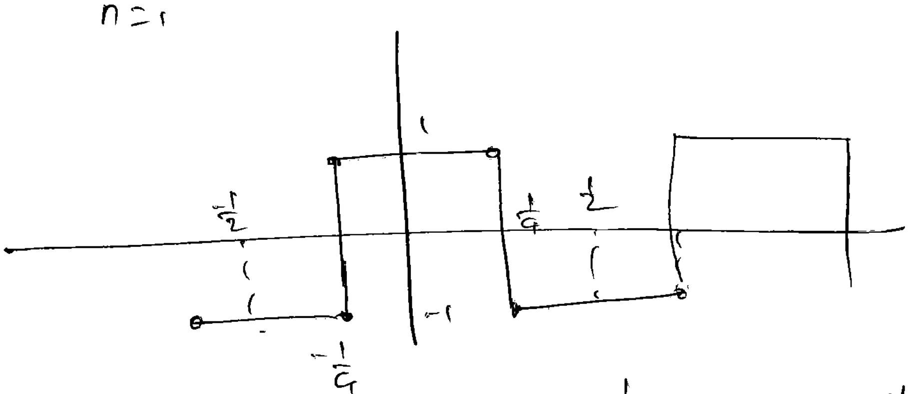
(Clay millenium problem)

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{\prod_{i=1}^{\infty}\left(1 - \frac{1}{p_i^s}\right)} \qquad Re(s)$$

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

Period = 1

$x(t)$



$$a_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} x(t)\, e^{-j2\pi kt}\, dt$$

$a_0 = 0$

$K \neq 0$

$$a_k = \int_{-\frac{1}{2}}^{-\frac{1}{4}} -e^{-j2\pi kt}\, dt + \int_{-\frac{1}{4}}^{\frac{1}{4}} e^{-j2\pi kt}\, dt - \int_{\frac{1}{4}}^{\frac{1}{2}} e^{-j2\pi kt}\, dt$$

$$= \left.\frac{e^{-j2\pi kt}}{j2\pi k}\right|_{-\frac{1}{2}}^{-\frac{1}{4}} + \left.\frac{e^{-j2\pi kt}}{-j2\pi k}\right|_{-\frac{1}{4}}^{\frac{1}{4}} + \left.\frac{e^{-j2\pi kt}}{j2\pi k}\right|_{\frac{1}{4}}^{\frac{1}{2}}$$

$$= \frac{e^{\frac{j\pi k}{2}} - e^{j\pi k}}{j2\pi k} + \frac{e^{\frac{j\pi k}{2}} - e^{-\frac{j\pi k}{2}}}{j2\pi k} + \frac{e^{-j\pi k} - e^{-\frac{j\pi k}{2}}}{j2\pi k}$$

$$= \frac{e^{\frac{j\pi k}{2}} - e^{-\frac{j\pi k}{2}}}{\ldots}$$

$a_k = 0 \quad , \quad k \text{ is even}$

$$a_k = \frac{e^{\frac{j\pi k}{2}} - e^{-\frac{j\pi k}{2}}}{j\pi k} = \frac{2\sin\left(\frac{\pi k}{2}\right)}{\pi k}$$

By Parseval's theorem

$$\frac{1}{1}\int_{-\frac{1}{2}}^{\frac{1}{2}} |x^2(t)|\,dt = \sum_{k=-\infty}^{\infty} |a_k|^2$$

$$1 = 2\sum_{\substack{k\geq 1\\ k:\text{odd}}} \frac{4}{\pi^2 k^2}$$

$$\Rightarrow \sum_{\substack{k\geq 1\\ k:\text{odd}}} \frac{1}{k^2} = \frac{\pi^2}{8}$$

$$\sum_{k:k\geq 1} \frac{1}{k^2} = \sum_{\substack{k\geq 1\\ k:\text{odd}}} \frac{1}{k^2} + \frac{1}{4}\sum_{k:k\geq 1} \frac{1}{k^2}$$

$$x = \frac{\pi^2}{8} + \frac{x}{4}$$

$$\Rightarrow \frac{3x}{4} = \frac{\pi^2}{8} \Rightarrow x = \frac{\pi^2}{6}$$

Analytic means

$f'(z)$ exists at all points

$z \in B(z_0; \delta)$

Parseval's theorem, Cauchy-Schwartz

Eigen decomposition

$$h(t) = \sum \hat{H}(f)\, e^{j2\pi f t}\, df$$

$$\underline{X} = a_1 \underline{e_1} + a_2 \underline{e_2} + \cdots + a_n \underline{e_n}$$

$\cancel{e}$ $x(t) = e^{t}$

$x(t) = e^{t^2}$

~~Ex~~ $\sin\left(\frac{1}{t}\right)$

<u>Probability theory</u>

Probability Theory

$(\Omega, \mathcal{F}) \rightarrow$ measurable space

$\mathcal{F}$; σ-algebra

collections of subsets of Ω which satisfy the following properties

$$\begin{cases} \text{(i)} & \phi \in \mathcal{F} \\ \text{(ii)} & A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F} \\ \text{(iii)} & A_i \in \mathcal{F} \Rightarrow \bigcup_i A_i \in \mathcal{F} \end{cases}$$

σ-algebra

$\mu \rightarrow$ countably additive probability measure on $(\Omega, \mathcal{F})$ if

(a) $\mu(\phi) = 0$, $\mu(\Omega) = 1$

(b) if A_i's are pairwise disjoint

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i)$$ (countable additivity)

$\Omega = [0, 1]$, $\mathcal{B}$ = Borel σ-algebra
smallest sigma algebra containing Borel sets

Exercise $\mathcal{A}$ is a collection of sets

$\sigma(\mathcal{A})$: smallest σ-algebra containing $\mathcal{A}$

show that if $\mathcal{A} = [r_1, r_2]$ $\quad 0 \le r_1 < r_2 \le 1$, $r_1, r_2 \in \mathbb{Q}$

then $\mathcal{B} = \sigma(\mathcal{A})$

---

random variable : "measurable mapping" from Ω to $\mathbb{R}$

$$X : (\Omega, \mathcal{F}) \longmapsto (\mathbb{R}, \mathbb{B}_{\mathbb{R}})$$

$\forall E \in \mathbb{B}_{\mathbb{R}}$

$\{\omega : X(\omega) \in E\} \in \mathcal{F}$

Expectations

$E[X]$

a) ~~indicatiory~~ indication function

$E \in \mathcal{F}$

$$1_E(\omega) = \begin{cases} 1, & \omega \in E \\ 0, & \omega \in E^c \end{cases}$$

$$E[1_E] = \mu(E)$$

b) simple function

$E_i \in \mathcal{F}$, E_i's are disjoint

$$X = \sum_{i=1}^{k} a_i 1_{E_i}$$

$$E[X] = \sum_{i=1}^{k} a_i \mu(E_i)$$

~~b)~~ c) $X(\omega) \geq 0$

$$E[X] = \sup_{\substack{X_n \leq X \\ X_n\text{-simple}}} E[X_n]$$

d) $X_+ = \max(X, 0)$, $X_- = \min(X, 0)$

$$E[X] = E[X_+] + E[X_-]$$

as long as $E[X_+] < \infty$ or $E[X_-] > -\infty$

Notions of Convergence

a) pointwise convergence

$X_n(\omega) \xrightarrow[n \to \infty]{} X(\omega)$ pointwise if $\forall \omega$

b) convergence almost surely

$E = \{\omega : X_n(\omega) \xrightarrow[n \to \infty]{} X(\omega)\}$, if $\mu(E) = 1$

c) convergence in norm, $\mathbb{L}_2$

$E[|X_n - X|^2] \to 0$ as $n \to \infty$

(5)

d) convergence in measure

$X_n \xrightarrow{\mu} X$

$\forall \varepsilon > 0$

& let $A_n = \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$

then $\mu(A_n) \underset{n\to\infty}{\longrightarrow} 0$ $\quad \forall \varepsilon > 0$

a.s. convergence $\Rightarrow$ convergence in measure

convergence in norm $\Rightarrow$

~~Weak convergence~~

For any random variable

$F_X(x) = P\{\omega : X(\omega) \le x\}$

$F_X(x)$: a) non-decreasing

b) $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$

c) $F_X(x)$ is right continuous

$\xleftarrow{1-1}$ countably additive probability measures on the real line

points of continuity

$A_n = \{x : F_X(x) - F_X(x^-) \ge \frac{1}{n}\}$

$\bigcup_n A_n$ is countable

Weak convergence (convergence of distributions or measure)

$X_n \xrightarrow{w} X$

if $F_{X_n}(x) \longrightarrow F_X(x)$ $\quad \forall x \in \mathcal{C}$: Points of continuity of $F_X(x)$

$X_n = 1 + \frac{1}{n}$ w.p. 1

$X = 1$ w.p. 1

$F_{X_n}(1) = 0 \quad \forall n$

$F_X(1) = 1$

~~Central lim~~ Levy - Cramer Continuity theorem

$\phi_X(t) = \mathbb{E}[e^{itX}]$ characteristic function

~~Le~~ . Levy's continuity theorem

The following two statements are equivalent

a) $X_n \xrightarrow{\omega} X$

b) $\phi_{X_n}(t) \to \phi_X(t) \quad \forall t$

$$F(b) - F(a) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \phi_X(t) \left(\frac{e^{-itb} - e^{-ita}}{-it} \right) dt$$ inversion

---

Central Limit Theorem

$X_1 \perp\!\!\!\perp X_2$ $\quad P\{\omega : X_1(\omega) \in E_1,\ X_2(\omega) \in E_2\}$
$= P\{\omega : X_1(\omega) \in E_1\}\ P\{\omega : X_2(\omega) \in E_2\}$

if $X_1, X_2, \ldots$ are mutually independent, identically distributed and

$E[X^2] = \sigma^2, \quad E(X) = 0$

~~then~~ then $\dfrac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{\omega} \mathcal{N}(0, \sigma^2)$

Pf

$$\phi_n(t) = E\left[e^{it\frac{(X_1 + \cdots + X_n)}{\sqrt{n}}}\right] = \phi_X\left(\frac{t}{\sqrt{n}}\right)^n$$

5

$$\left(\phi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n = E\left[e^{i\frac{t}{\sqrt{n}}X}\right]$$

$$= \left(1 - \frac{t^2\sigma^2}{2n} + o\left(\frac{1}{n}\right)\right)^n$$

$$\to e^{-\frac{t^2\sigma^2}{2}}$$

characteristic function of Gaussian

Lindeberg's theorem

X_i are ~~mutually~~ pairwise independent, zero-mean and

$$s_n^2 = \sum_{i=1}^n E[X_i^2], \quad s_n^2 \to \infty \text{ as } n \to \infty$$

then if

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^n \int_{|x|>\varepsilon s_n} x^2 \, d\alpha_n \to 0 \quad \forall \varepsilon > 0$$

then

$$\frac{X_1 + \dots + X_n}{s_n} \xrightarrow{w} N(0,1)$$

---

if they where identically distributed

$$s_n^2 = n\sigma^2$$

$$E\left[X^2 1_{X^2 \geq \varepsilon^2 s_n^2}\right] \to 0 \text{ as } n \to \infty$$

$$X^2 = \lim_{n\to\infty} X^2 1_{X^2 \leq \varepsilon^2 s_n^2}$$

$X_n \uparrow X$, $X_n, X \geq 0$
$E[X] < \infty$ then
$E[X_n] \to E[X]$

monotone convergence

Probability theory Thm 3.18 AIMS
(S.R.S. Varadhan) NYU

Information Theory

Information measures

$X \in \{1, \ldots, |X|\}$ finite set

$$H(X) = -\sum_x p(x) \log_2 p(x)$$

$p(1), \ldots, p(k)$

$$\binom{n}{n p(1) \ \ldots \ n p(k)} \doteq 2^{n H(p_1, \ldots, p_k)}$$

(Exercise: Stirling's formulae)

Suppose I toss the k-sided coin n-times and I want to convey the outcomes then I can give you $n H(X)$ bits and you will be able to tell the outcome with

---

$$H(X|Y) = \sum_y P(Y=y)\, H(X|Y=y)$$
$$= \sum_{x,y} - p(x,y) \log p(x|y)$$

Entropy is concave in $p(x)$

$f(\underline{x})$ is concave $\Rightarrow$

$$f(\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2) \geq \alpha f(\underline{x}_1) + (1-\alpha) f(\underline{x}_2) \quad \forall\, \alpha \in [0,1)$$

if f is concave

$$E[f(X)] \leq f(E(X))$$

Question coin with bias

$q_1, \ldots, q_K$

toss the coins (independent outcomes) and

observe

$np_1, \ldots, np_K$

$$\binom{n}{np_1 \ldots np_K} q_1^{np_1} \cdots q_K^{np_K}$$

$$\doteq 2^{n\left[-\sum_i p_i \log p_i + \sum_i p_i \log q_i\right]}$$

$$= 2^{-n\left[\sum_i p_i \log \frac{p_i}{q_i}\right]}$$

$$D(p\|q) = \sum_i p_i \log \frac{p_i}{q_i}$$

Claim $D(p\|q) \geq 0$

$$-D(p\|q) = \sum_i p_i \log \frac{q_i}{p_i} \leq \underbrace{\log\left(\sum_i p_i \frac{q_i}{p_i}\right)}_{=0}$$

$p(x)$ & $q(x)$ $\xmapsto{r(y|x)}$

$$\tilde{p}(y) = \sum_x p(x)\, r(y|x)$$

$$\tilde{q}(y) = \sum_y \cancel{p(x)}\, q(x)\, r(y|x)$$

Show that $D(\tilde{p}(y)\|\tilde{q}(y)) \leq D(p(x)\|q(x))$

(data processing inequality)

$$p(x)\, r(y|x) = \tilde{p}(y)\, \hat{r}(x|y)$$

$$q(x)\, r(y|x) = \tilde{q}(y)\, \bar{r}(x|y)$$

$$D_{KL}\left(p(x)\, r(y|x) \,\|\, q(x)\, r(y|x)\right)$$

$$= \sum_{x,y} p(x)\, r(y|x) \log \frac{p(x)\, \cancel{r(y|x)}}{q(x)\, \cancel{r(y|x)}}$$

$$= \sum_x p(x) \log \frac{p(x)}{q(x)} = D\left(p(x) \,\|\, q(x)\right)$$

$$= \sum_{x,y} \tilde{p}(y)\, \hat{r}(x|y) \log \left(\frac{\tilde{p}(y)\, \hat{r}(x|y)}{\tilde{q}(y)\, \bar{r}(x|y)} \right)$$

$$= \sum_y \tilde{p}(y) \log \frac{\tilde{p}(y)}{\tilde{q}(y)} + \sum_y \tilde{p}(y) \Big(\underbrace{\sum_x \hat{r}(x|y) \log \frac{\hat{r}(x|y)}{\bar{r}(x|y)}}_{\geq 0} \Big)$$

$$\therefore \sum_y \tilde{p}(y) \log \frac{\tilde{p}(y)}{\tilde{q}(y)} \leq D\left(p(x) \,\|\, q(x)\right)$$

---

mutual information

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

$$= D_{KL}\left(p(x y) \,\|\, p(x)p(y)\right) \geq 0$$

$$= H(X) - H(X|Y)$$

$$= H(Y) - H(Y|X)$$

<u>Suppose</u> [Shannon's channel coding theorem]

$$X \longrightarrow \boxed{p(y|x)} \longrightarrow Y$$

capacity = maximum bits per transmission receiver can reconstruct the messa.

$$C = \max_{p(x)} I(X;Y)$$

$$M \longrightarrow \boxed{\text{Encode}} \xrightarrow{X^n} \boxed{p(y|x)} \xrightarrow{Y^n} \boxed{\text{Decode}} \longrightarrow \hat{M}$$

$\in [1, \ldots, 2^{nR}]$

We say that a rate R bits / transmission is achievable if $\exists$ a sequence of encoding & decoders (n) such that

$$P_r\{\hat{M} \neq M\} \to 0 \quad \text{as} \quad n \to \infty$$

capacity is the maximum achievable rate

Theorem $\quad C = \max_{p(x)} I(X;Y)$

<u>channel</u> $\to$ memoryless $\quad p(y_i \mid x^{i-1}, y^{i-1}, x_i) = p(y_i|x_i)$

$$p(y^n/x^n) = \prod_i p(y_i|x_i)$$

Converse to channel coding theorem

<u>Fano's inequality</u> $\quad M \in \{1, \ldots, |M|\}$

$$P_r\{\hat{M} \neq M\} \leq \varepsilon$$

$$H(M|\hat{M}) \leq 1 + \varepsilon \log |M|$$

<u>Proof</u> $\quad E = \begin{cases} 1, & M \neq \hat{M} \\ 0, & \text{o.w.} \end{cases}$

$$\begin{aligned} H(M|\hat{M}) &\leq H(M, E|\hat{M}) \\ &= H(E|\hat{M}) + H(M|E, \hat{M}) \quad \text{ch} \\ &\leq H(E) + P(E=1)\, H(M|E=1, \hat{M}) \\ &\leq 1 + \varepsilon \log |M| \end{aligned}$$

(10)

Given a sequence of codebooks (encoding decoding scheme)
n 2^{nR}

$P(E) \xrightarrow[n\to\infty]{} 0$

$M \in \{1, \dots, 2^{nR}\}$

$X^n(m)$

$p(m, x^n, y^n)$ distribution induced by a codebook

$= p(m)\, p(x^n|m)\, p(y^n|x^n)$

$= p(m)\, p(x^n|m) \prod_i p(y_i|x_i)$

$M \to X^n \to Y^n \to \hat{M}$

$I(M;\hat{M}) \le I(X^n;Y^n)$

$nR \le H(M)$

$= H(M|\hat{M}) + I(M;\hat{M})$

$\le \left[1 + P(E)\log|2^{nR}|\right] + I(M;\hat{M})$

$= 1 + nP(E)R + I(M;\hat{M})$

$\le 1 + nRP(E) + I(X^n;Y^n)$

$I(X^n;Y^n) = H(Y^n) - H(Y^n|X^n)$

$= \sum_i \left[H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, X_i, X^{n\setminus i})\right]$

$= \sum_i H(Y_i|Y^{i-1}) - H(Y_i|X_i)$

$\le \sum_i H(Y_i) - H(Y_i|X_i)$

$= \sum_i I(X_i;Y_i)$

$\le n \max_{p(x)} I(X;Y)$

$nR \le 1 + nR\,P_n(E) + nC$

$R \le \frac{1}{n} + P_n(E)R + C$

Achieve

$p(x)$

$R = I(X;Y) - \varepsilon$

Random Coding argument

$X^n(m) \sim \prod_i p(x_i)$

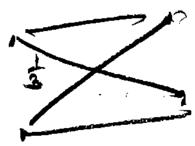
decoder

$\mathcal{A} = \{ m : (X^n(m), y^n) \in T_\varepsilon^{(n)}(X,Y) \}$

$y^n \quad y_1, \ldots \quad y_n$

$x_{m1}, \ldots \quad x_{mn}$

$\frac{1}{2}$ 0 1



$y^n =$ 0 0 0 1 1 0 1 1 0 1 1 0

$x^n(1)$ 1 1 1 0 0 1 0 0 1 0 0 1

$p(x,y)$

0	0	$\frac{1}{3}$
0	1	$\frac{1}{6}$
1	0	$\frac{1}{6}$
1	1	$\frac{1}{3}$

0	1	$\frac{1}{2}$
1	0	$\frac{1}{2}$

if $|\mathcal{A}| = 1$ then $\hat{m}$ = message in $\mathcal{A}$

otherw error

$M = 1$ is tx $\{(X^n(1), y^n) \in T_\varepsilon^n(X,Y)\}$ $\quad X^n(1) = \prod_i p(x_i)$

$x^n(1)$ $\quad 0 \ldots 0 \; 1 1 \ldots 1 \quad 0 \ldots 0 \; 1 \ldots 1$

$0 \ldots 0 \; 1 \ldots 1 \quad 0 \ldots 0 \; 1 \ldots 1$

(12)

$$E_m = \{ (X^n(m), Y^n) \in T_\varepsilon^{(n)}(X,Y) \}$$

$$P(X^n(1), Y^n, X^n(m)) = P(X^n(1))\, P_{Y|X}(Y^n|X^n(1))\, P(X^n(m)) \quad , m \neq 1$$

$$P(Y^n, X^n(m)) = P_{Y^n}(Y^n)\, P(X^n)$$

$$P(E_m) = 2^{-n D(P(X,Y) \| P(X)P(Y))}$$

$$= 2^{-n I(X;Y)}$$

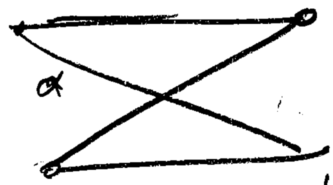
$$P\left(\bigcup_{m=2}^{2^{nR}} E_m\right) \leq \sum_m P(E_m) \leq 2^{nR}\, 2^{-n I(X;Y)}$$

$$R = I(X;Y) - \varepsilon$$

$\cancel{X}^n = \{-1, 1\}^n$

$Y^n = \{-1, 1\}^n$

$\cancel{p(y^n|x^n)}$



$p(Y=0|X=0) = 1-\alpha \qquad P(Y=1|X=1) = 1-\alpha$

$p(Y=1|X=0) = \alpha \qquad p(Y=0|X=1) = \alpha$

Conjecture

$$\max_{\substack{\text{all boolean functions} \\ f(x^n),\, g(y^n)}} I(f(X^n); g(Y^n)) = I(X_1; Y_1)$$

$$= 1 - H(\alpha)$$

$$\max_{f} I(f(X^n); f(Y^n)) = I(X_1; Y_1)$$
