# I. Random Variables

## Axioms of Probability

1) $P(A) \geq 0$
2) $P(\Omega) = 1$ (whole set)
3) If $A \cap B = \varphi$, then $P(A \cup B) = P(A) + P(B)$ mutually exclusive (M.E.)

   * If not, $P(A \cup B) = P(A) + P(B) - P(AB)$

**Independence** $P(AB) = P(A) \cdot P(B) \rightarrow P(A|B) = \dfrac{P(AB)}{P(B)} = P(A)$

**Bayes' theorem**

H: Hypothesis
E: Event

$$\underbrace{P(H|E)}_{\text{posterior}} = \frac{\overbrace{P(E|H)}^{\text{likelihood}}\overbrace{P(H)}^{\text{prior}}}{\underbrace{P(E)}_{\text{evidence}}}$$

## Definitions of Random Variables

### Probability Distribution Function (P.D.F)

$P(X \leq x) = F_X(x) \geq 0$ $\begin{cases} \text{for continuous-type } X, & P\{X = x\} = 0 \\ \text{for discrete-type } X, & p_i = P\{X = x_i\} \end{cases}$

### Probability density Function (p.d.f)

$$\begin{cases} \text{for continuous-type } X, & f_X(x) = \dfrac{dF_X(x)}{dx} \\ \text{for discrete-type } X, & f_X(x) = \sum_i p_i \delta(x - x_i) \end{cases}$$

also known as probability mass function (p.m.f)

$$P\{x_1 \leq X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f_X(x)\,dx$$

# Some common distributions

## Continuous-type r.v.

| | |
|---|---|
| Normal (Gaussian) $X \sim N(\mu, \sigma^2)$ | $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ <br> $F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \triangleq G\left(\frac{x-\mu}{\sigma}\right)$ <br> mean <br> variance |
| Uniform $X \sim U(a,b)$ | $f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$ <br> mean: $\frac{a+b}{2}$ <br> variance: $\frac{(a-b)^2}{12}$ |
| Exponential $X \sim \exp(\lambda)$ | $f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$ <br> mean: $\theta = \frac{1}{\lambda}$ <br> variance: $\theta^2 = \frac{1}{\lambda^2}$ |

## Discrete-type r.v.

| | |
|---|---|
| Bernoulli | $p(X=1) = p,\ p(X=0) = 1-p$ <br> mean: $p$ <br> variance: $p(1-p)$ |
| Binomial | $X \sim B(n,p)$, if $p(X=k) = \binom{n}{k} p^k q^{n-k}$, $k = 0, 1, \cdots, n$ <br> mean: $np$, variance: $np(1-p)$ |
| Poisson | $X \sim P(\lambda)$, if $p(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, \cdots, \infty$ <br> mean = variance: $\lambda$ |

## How to characterize the r.v.

**Mean**

$$\eta_X = \bar{X} = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx \qquad \text{continuous}$$

$$= \int x \sum_i p_i \delta(x - x_i) dx = \sum_i x_i p_i \qquad \text{discrete}$$

if $Y = g(X)$, then $\mu_Y = E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy$

$$= E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx \qquad \text{continuous}$$

$$= \sum_i g(x_i) p_i \qquad \text{discrete}$$

**Variable** $\sigma_X^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx \quad > 0$

**Moments** $m_n = E(X^n), \quad n \geq 1$

$\mu_n = E[(X-\mu)^n] \longrightarrow$ central moments

$E[(X-a)^n] \longrightarrow$ generalized moments of $X$ about $a$

$E[|X|] \longrightarrow$ absolute moments of $X$

**Characteristic Function**

$$\Phi_X(w) = E(e^{jXw}) = \int_{-\infty}^{\infty} e^{jxw} f_X(x)\,dx \quad \leq \Phi_X(0) = 1$$

(frequency domain)

Uses: ① $\Phi_{X+Y}(w) = \Phi_X(w)\,\Phi_Y(w)$

can be used to calculate the p.d.f of $X+Y$

② Calculate the $k$-th moment of $X$

$$\Phi_X(w) = E(e^{jXw}) = E\left[\sum_{k=0}^{\infty} \frac{(jwX)^k}{k!}\right] \qquad \text{series}$$

$$= \sum_{k=0}^{\infty} j^k \frac{E(X^k)}{k!} w^k = 1 + jE(X)w - \frac{E(X^2)}{2}w^2 + \cdots + j^k \frac{E(X^k)}{k!} w^k + \cdots$$

Therefore, the $k$th moment of $X$ is

$$\boxed{E(X^k) = \frac{1}{j^k} \left.\frac{\partial^k \Phi_X(w)}{\partial w^k}\right|_{w=0}, \quad k \geq 1}$$

③ if $X \sim N(\mu, \sigma^2)$, $\Phi_X(w) = e^{(j\mu w - \sigma^2 w^2/2)}$

$X \sim N(0, \sigma^2)$, $\Phi_X(w) = e^{-\sigma^2 w^2/2}$

# 2. Joint Distributions

## Definition

$F_{XY}(x,y) = P(X \le x, Y \le y)$ $\qquad \geqslant 0$

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$

$$P((X,Y) \in D) = \iint_{(x,y)\in D} f_{XY}(x,y)\,dx\,dy$$

**Marginal statistics** $\quad F_X(x) = F_{XY}(x, +\infty), \quad F_Y(y) = F_{XY}(+\infty, y)$

$$f_X(x) = \int_0^{+\infty} f_{XY}(x,y)\,dy, \quad f_Y(y) = \int_0^{+\infty} f_{XY}(x,y)\,dx$$

if discrete, then
$$\begin{cases} P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij} \\ P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p_{ij} \end{cases}$$

joint p.d.f $\;\underset{\not\longleftarrow}{\longrightarrow}\;$ marginal p.d.f.

only when $X, Y$ are independent

**Independence** $\quad F_{XY}(x,y) = F_X(x)F_Y(y)$

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

**One special case** $\quad$ if $X, Y$ are independent, $Z = X + Y$

then $$f_Z(z) = \int_{-\infty}^{+\infty} \underbrace{f_X(z-y)f_Y(y)\,dy}_{\text{convolution of p.d.f.}}$$

## Functions of two R.V.s

① $Z = T_1(X,Y) \quad W = T_2(X,Y)$

$$f_{ZW}(z,w) = \sum_i \frac{1}{|J(x_i,y_i)|} f_{XY}(x_i,y_i) \qquad \text{where } J(x_i,y_i) = \det\begin{pmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{pmatrix}$$

① $X = T_3(Z,W), \; Y = T_4(Z,W)$

$$f_{ZW}(z,w) = \sum_i |J(z_i,w_i)| \, f_{XY}(x_i,y_i) \qquad \text{where } J(x_i,y_i) = \det\begin{pmatrix} \frac{\partial x}{\partial z} & \frac{\partial y}{\partial z} \\ \frac{\partial x}{\partial w} & \frac{\partial y}{\partial w} \end{pmatrix}$$

### Another special case

If $X, Y$ are zero mean independent Guassian r.v.s with common variance, then

$$\begin{cases} \sqrt{X^2+Y^2} & : \text{Rayleigh Distribution (magnitude)} \\ \tan^{-1}\left(\frac{Y}{X}\right) & : \text{Uniform Distribution (phase)} \end{cases}$$

## Properties

**Covariance** $\quad Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - \mu_X \mu_Y$

$$|Cov(X,Y)| \le \sqrt{Var(X)\,Var(Y)}$$

**Correlation Coefficient** $(-1 \le \rho_{XY} \le 1)$ $\quad \rho_{XY} = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \dfrac{Cov(X,Y)}{\sigma_X \sigma_Y}$

**Uncorrelated** $\quad \rho_{XY} = 0$, or $E(XY) = E(X)E(Y)$

**Orthogonality** $\quad E(XY) = 0$

Independence $\xrightarrow{\qquad}$ Uncorrelation $\Longleftrightarrow$ Orthogonality

Uncorrelation $\not\Rightarrow$ Independence: only when $X, Y$ are jointly Guassian

Uncorrelation $\Longleftrightarrow$ Orthogonality: if $X, Y$ has zero mean

**Joint Moments** of order $(k,m)$ $E[X^kY^m]=\int_{-\infty}^{+\infty}\int_{-\infty}^{\infty} x^k y^m f_{XY}(x,y)\,dx\,dy$

**Joint Characteristic Function**

$$\Phi_{XY}(u,v)=E(e^{j(Xu+Yv)})=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{j(xu+yv)} f_{XY}(x,y)\,dx\,dy \le \Phi_{XY}(0,0)=1$$

$$=E\left(1+j(Xu+Yv)+j^2\frac{(Xu+Yv)^2}{2!}+\cdots+j^k\frac{(Xu+Yv)^k}{k!}\right)$$

$$\Rightarrow E(XY)=\frac{1}{j^2}\frac{\partial^2\Phi_{XY}(u,v)}{\partial u\partial v}\bigg|_{u=0,v=0}$$

if $X,Y$ are independent $\begin{cases}\Phi_{XY}(u,v)=\Phi_X(u)\Phi_Y(v)\\ \Phi_{XY}(u,0)=\Phi_X(u),\ \Phi_{XY}(0,v)=\Phi_Y(v)\end{cases}$

**Conditional p.d.f.** $f_{X|Y}(x|y)=\frac{f_{XY}(x,y)}{f_Y(y)} \overset{\text{if dependent}}{=\!=} f_X(x)$

$$P(X=x_i|Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)} \quad \text{(discrete)}$$

**Extension to multiple r.v.'s**

- Joint PDF

$$F_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)\equiv P(X_1\le x_1,X_2\le x_2,\ldots X_n\le x_n)$$

- Joint pdf

$$f_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)\equiv\frac{\partial^n F_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)}{\partial x_1\partial x_2\ldots\partial x_n}$$

- **Independent**

$$F_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)=F_{X_1}(x_1)F_{X_2}(x_2)\ldots F_{X_n}(x_n)$$
$$f_{X_1X_2\ldots X_n}(x_1,x_2,\ldots x_n)=f_{X_1}(x_1)f_{X_2}(x_2)\ldots f_{X_n}(x_n)$$

- **i.i.d.** (independent, identically distributed)
  - The random variables are independent and have the same distribution.
  - Example: outcomes from repeatedly flipping a coin.

# 3. Sequences of Random Variables

## Concentration inequalities

**Markov Inequality:** non-negative X

$$P(X \geq a) \leq \frac{E(X)}{a} \quad (a > 0)$$

**Generalised Markov Inequality:** strictly increasing and non-negative g

$$P(X \geq a) = P(g(X) \geq g(a)) \leq \frac{E(g(X))}{g(a)}, \ a > 0$$

**Chebyshev Inequality:** when $g(X) = X^2$

$$P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}, \ a > 0$$

in particular $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

**Chernoff Bound:**

$$P(X > a) \leq \min_{\lambda > 0} e^{-\lambda a} E[e^{\lambda X}] = \min_{\lambda > 0} e^{-\lambda a} \Phi_X(\lambda), \ a > 0$$

(need to choose a suitable $\lambda$)

**Bernstein Inequality:** $X_1, X_2, \cdots, X_n$ : i.i.d. $\overline{X_n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$

$|X_i - \mu| < M$

$$P(|\overline{X_n} - \mu| \geq t) \leq 2\exp\left(-\frac{nt^2}{2(\sigma^2 + \frac{tM}{3})}\right), \ t > 0$$

**Hoeffding Inequality:** $X_1, X_2, \cdots, X_n$: independent, $\overline{X_n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$

$X_i \in [a_i, b_i]$

$$P(|\overline{X_n} - E(\overline{X_n})| \geq t) \leq 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right), \ t > 0$$

# The law of Large Numbers (L.L.N)

Suppose $X_1, X_2, \cdots, X_n$ are i.i.d. with finite $E(X_i) = \mu$.

Then $\bar{X}_n$ converges to $\mu$

$$\begin{cases} \text{in probability}: \quad \forall \varepsilon > 0, \ \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 & \text{(Weak)} \\ \text{almost surely}: \quad P(\lim_{n\to\infty} X_n = X) = 1 & \text{(Strong)} \end{cases}$$

non-examinable

# Central Limit Theorem

Suppose $X_1, X_2, \cdots, X_n$ are i.i.d with $\begin{cases} \text{zero mean} \\ \text{common variance } \sigma^2 \end{cases}$

When $n \to \infty$, the scaled sum $Y = \dfrac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} \to N(0, \sigma^2)$

(doesn't hold good for Cauchy distribution because of undefined variances)

# 4. Parameter Estimation

## Maximum Likelihood

* joint p.d.f of $X_1 \sim X_n$ depending on $\theta$
* a function of $\theta$ with respect to $X_1 \sim X_n$

**Likelihood function** $f_X(x_1, x_2, \cdots, x_n; \theta)$

**Likelihood equation** $\hat{\theta}_{ML} = \arg\max_{\theta} f_X(x_1, x_2, \cdots, x_n; \theta)$

**Log-likelihood function** $L(x_1, x_2, \cdots, x_n; \theta) \triangleq \log f_X(x_1, x_2, \cdots, x_n; \theta)$

$$\left.\frac{\partial \log f_X(x_1, x_2, \cdots, x_n; \theta)}{\partial \theta}\right|_{\theta = \hat{\theta}_{ML}} = 0$$

**Unbiased estimator** $E[\hat{\theta}_{ML}(x)] = \theta$

**Asymptotically unbiased estimator** $E[\hat{\theta}_{ML}(x)] = \theta$, as $n \to \infty$

**Consistent estimator** $Var(\hat{\theta}_{ML}) \to 0$, as $n \to \infty$

## Cramer-Rao Bound (Used to determine the best unbiased estimator)

known as efficient estimator

$$Var(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

Fisher information

where $I(\theta) = E\left[\left(\frac{\partial}{\partial \theta} \ln f_X(x_1, x_2, \cdots, x_n; \theta)\right)^2\right] = -E\left[\frac{\partial^2}{\partial \theta^2} \ln f_X(x_1, x_2, \cdots, x_n; \theta)\right]$

## Bayesian Estimation (MAP)

the likelihood function

the chosen a-priori p.d.f.

$$f_{\theta|X}(\theta | x_1, x_2, \cdots, x_n) = \frac{f_{X|\theta}(x_1, x_2, \cdots, x_n | \theta) f_\theta(\theta)}{f_X(x_1, x_2, \cdots, x_n)}$$

a-posteriori p.d.f constructed from observation

known

# 5. Stochastic Processes

## Definition

$X(t,\xi)$ For fixed $t$, $X_1 = X(t_1,\xi_i)$ is a r.v.

$X_1 = X(t_1)$ and $X_2 = X(t_2)$ are two different r.v.s

$$\begin{cases} F_X(x_1,x_2,t_1,t_2) = P\{X(t_1) \le x_1,\ X(t_2) \le x_2\} \\ f_X(x_1,x_2,t_1,t_2) \triangleq \dfrac{\partial^2 F_X(x_1,x_2,t_1,t_2)}{\partial x_1 \partial x_2} \end{cases}$$

**Mean** $\mu(t) = E\{X(t)\} = \int_{-\infty}^{\infty} x f_X(x,t)\,dx$

**Autocorrelation** $R_{XX}(t_1,t_2) = E\{X(t_1)X^*(t_2)\}$

$$= \iint x_1 x_2^* f_X(x_1,x_2,t_1,t_2)\,dx_1 dx_2$$

① $R_{XX}(t_1,t_2) = R_{XX}^*(t_2,t_1)$ ② $R_{XX}(t,t) = E\{|X(t)|^2\} \overset{\text{(power)}}{\ge} 0$

③ For any set of constants $\{a_i\}_{i=1}^n$, $\sum_{i=1}^n \sum_{j=1}^n a_i a_j^* R_{XX}(t_i,t_j) \ge 0$

**Autocovariance** $C_{XX}(t_1,t_2) = R_{XX}(t_1,t_2) - \mu_X(t_1)\mu_X^*(t_2)$

## Stationary Stochastic Processes

**Strict-Sense** $n^{th}$-order:

$$f_X(x_1,x_2,\cdots,x_n,t_1,t_2,\cdots,t_n) \equiv f_X(x_1,x_2,\cdots,x_n,t_1+c,t_2+c,\cdots,t_n+c)$$

**Wide-Sense** ① $E\{X(t)\} = \mu$

② $E\{X(t_1)X^*(t_2)\} = R_{XX}(t_2-t_1)$ only depends on difference $\tau$

Strict-Sense $\substack{\Longrightarrow \\ \not\Longleftarrow}$ Wide-Sense

only when Guassian

# Ergodic Processes

**Mean**

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} X(t)\,dt = E[X(t)] = \mu$$

**Autocorrelation**

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} X(t)X^*(t-\tau)\,dt = R_{XX}(\tau)$$

limited time $\longrightarrow$ whole behaviour

# Deterministic systems

Deterministic Systems $\rightarrow$ Memoryless Systems

Deterministic Systems $\rightarrow$ Systems with Memory

Strict-sense stationary input $\rightarrow$ Memoryless system $\rightarrow$ Strict-sense stationary output.

Wide-sense stationary input $\rightarrow$ Memoryless system $\rightarrow$ Not necessarily stationary in any sense.

$X(t)$ stationary Gaussian with $R_{XX}(\tau)$ $\rightarrow$ Memoryless system $\rightarrow$ $Y(t)$ stationary, but *not* necessarily Gaussian with $R_{XY}(\tau)=\eta R_{XX}(\tau)$.

Systems with Memory: { Time-Invariant Systems + Linear Systems } $\longrightarrow$ LTI Systems

## For LTI systems:

$h(t)$: impulse response of the system

$$Y(t) = X(t) * h(t)$$

① $X(t)$ wide-sense stationary process $\rightarrow$ LTI system $h(t)$ $\rightarrow$ $Y(t)$ wide-sense stationary process.

(a)

② $X(t)$ strict-sense stationary process $\rightarrow$ LTI system $h(t)$ $\rightarrow$ $Y(t)$ strict-sense stationary process (see Text for proof )

(b)

③ $X(t)$ Gaussian process (also stationary) $\rightarrow$ LTI system $h(t)$ $\rightarrow$ $Y(t)$ Gaussian process (also stationary)

(c)

④ white noise $\longrightarrow$ colored noise

## Output statistics

for W.S.S. input $X(t)$

$$\mu_Y(t) = \mu_X(t) * h(t) = \mu_X \int_{-\infty}^{+\infty} h(\tau)\,d\tau = \mu_X c, \text{ constant}$$

$$R_{XY}(t_1,t_2) = R_{XX}(t_1,t_2) * h^*(t_2) = R_{XY}(\tau) = R_{XX}(\tau) * h^*(-\tau)$$

$$R_{YY}(t_1,t_2) = R_{XX}(t_1,t_2) * h^*(t_2) * h(t_1) = R_{YY}(\tau) = R_{XX}(\tau) * h^*(-\tau) * h(\tau)$$

✽ { white: time/frequency structure of the process
gaussian: probability distribution of the samples at $t_0$ }

# White Noise Process

Initial definition $\begin{cases} ① \ E[W(t)] = 0 \\ ② \ R_{WW}(t_1,t_2) = E[W(t_1)W^*(t_2)] = 0 \text{ unless } t_1 = t_2 \end{cases}$

time | frequency

W.S.S $R_{WW}(t_1,t_2) = q\delta(t_1-t_2) = q\delta(\tau) \xleftrightarrow{F.T.} S_{WW}(w) = q$ power

White noise + LTI = Coloured noise $N(t) = h(t) * W(t)$

$$R_{nn}(\tau) = q\rho(\tau) = q\,h(\tau) * h^*(-\tau) \xleftrightarrow{F.T.} S_{nn}(w) = q|H(w)|^2$$

# Discrete Time stochastic processes

$$\mu_n = E\{X(nT)\}$$

$$R(n_1,n_2) = E\{X(n_1T)X^*(n_2T)\}$$

$$C(n_1,n_2) = R(n_1,n_2) - \mu_{n_1}\mu_{n_2}^*$$

$X(nT)$ is W.S.S if $\begin{cases} E[X(nT)] = \mu \\ E[X((k+n)T)X^*(kT)] = R(n) = r_n = r_{-n}^* \end{cases}$

# Auto Regressive Moving Average (ARMA) Process

$$\underbrace{X(n)}_{\text{output}} = -\sum_{k=1}^{p} a_k X(n-k) + \sum_{k=0}^{q} b_k \underbrace{W(n-k)}_{\text{input}}$$

ARMA(p,q)

Z-transform: $X(z)\sum_{k=0}^{p} a_k z^{-k} = W(z)\sum_{k=0}^{q} b_k z^{-k}$. $a_0 \equiv 1$

or: $H(z) = \dfrac{X(z)}{W(z)} = \dfrac{b_0 + b_1z^{-1} + b_2z^{-2} + \cdots + b_qz^{-q}}{1 + a_1z^{-1} + a_2z^{-2} + \cdots + b_pz^{-p}} = \sum_{k=0}^{\infty} \underbrace{h(k)}_{\text{impulse response}} z^{-k}$

$X(n) = W(n) * h(n) = \sum_{k=0}^{\infty} h(n-k)W(k)$

$\begin{cases} \text{If } q=0, \text{ it is a AR(p) process (all-pole process)} \\ \text{If } p=0, \text{ it is a MA(q) process (all-zero process)} \end{cases}$

# 6. Power Spectrum

## Fourier Transform ( $w = 2\pi f$ )

$$\begin{cases} X(w) = \int_{-\infty}^{\infty} x(t) e^{-jwt} dt \\ X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi f t} dt \end{cases} \quad \checkmark \qquad \begin{cases} x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(w) e^{jwt} dt \\ x(t) = \int_{-\infty}^{\infty} X(w) = e^{j2\pi f t} dt \end{cases} \quad \checkmark$$

For deterministic signals, $|X(w)|^2$ is energy spectrum

## Power Spectrum for Stochastic Processes

**Wiener-Khinchin Theorem** For w.s.s process, $R_{XX}(\tau) \xleftrightarrow{F.T} S_{XX}(w)$

$$\begin{cases} S_{XX}(w) = \int_{-\infty}^{+\infty} R_{XX}(\tau) e^{-jw\tau} d\tau \\ R_{XX}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{XX}(w) e^{jw\tau} dw \end{cases}$$

When $\tau = 0$, $R_{XX}(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \underbrace{S_{XX}(w)\,dw}_{\text{power in the band}} = E\{|X(t)|^2\} = P$ the total power

$$\begin{cases} R_{XX}(\tau) \text{ nonnegative-definite} \Leftrightarrow S_{XX}(w) \geq 0 \\ X(t) \text{ real} \Leftrightarrow R_{XX}(\tau) = R_{XX}(-\tau) \Leftrightarrow S_{XX}(w) = S_{XX}(-w) \geq 0 \end{cases}$$

## Power Spectra and Linear Systems

**w.s.s process**

$$S_{XY}(w) = F\{R_{XX}(\tau) * h^*(-\tau)\} = S_{XX}(w) H^*(w)$$

$$S_{YY}(w) = F\{R_{YY}(\tau)\} = S_{XY}(w) H(w) = S_{XX}(w) |H(w)|^2$$

**w.s.s white noise**

$$R_{WW}(\tau) = q\delta(\tau) \Leftrightarrow S_{WW}(w) = q$$

$$S_{YY}(w) = q|H(w)|^2$$

## Discrete-Time Process $X(nT)$ $\{r_k\}_{-\infty}^{+\infty}$

$$\begin{cases} R_{XX}(\tau) = \sum_{k=-\infty}^{+\infty} r_k \delta(\tau - kT) \\ S_{XX}(w) = \sum_{k=-\infty}^{+\infty} r_k e^{-jkwT}, \quad \text{periodic}, \ 2B = \frac{2\pi}{T} \end{cases}$$

$\Rightarrow$ inversely, $r_k = \frac{1}{2B}\int_{-B}^{B} S_{XX}(w) e^{jkwT} dw$

$$r_0 = \frac{1}{2B}\int_{-B}^{B} S_{XX}(w) dw = \underbrace{P}_{\text{total power}}$$

$$\begin{cases} S_{XY}(w) = S_{XX}(w) H^*(e^{jw}) \\ S_{YY}(w) = S_{XX}(w) |H(e^{jw})|^2 \end{cases}$$

## Matched Filter

$$r(t) = s(t) + w(t) \xrightarrow[\text{at } t=t_0]{h(t)} y(t) = y_s(t) + n(t)$$

Suppose $w(t)$ to be white noise with spectral density $N_0$,

then $SNR \le \frac{E_s}{N_0}$ and

equality is guaranteed only if $\begin{cases} H(w) = S^*(w) e^{-jwt_0} \\ \text{or } h(t) = s(t_0 - t) \end{cases}$

If the receiver is causal, it is $\begin{cases} h_{opt}(t) = s(t_0 - t) u(t) \\ (SNR)_{opt} = \frac{\int_0^{t_0} s^2(t) dt}{N_0} \end{cases}$

# 7. Mean-Square Estimation

## Definition

estimation $\hat{Y} = \varphi(X_1, X_2, \cdots, X_n) = \varphi(\underline{X})$

error $\varepsilon(\underline{X}) = Y - \hat{Y} = Y - \varphi(\underline{X})$

MMSE: find an estimator $\varphi$ to minimise $E\{|\varepsilon|^2\}$

Under MMSE criterion $\hat{Y} = E\{Y|\underline{X}\}$

$\sigma_\varepsilon^2 = E\{var(Y|\underline{X})\} \geq 0$

Explicit Calculation $f_{X,Y}(x,y) \to f_X(x) \to f_{Y|X}(y|x)$

(not necessarily linear)

$\to \hat{Y} = \int y f_{Y|X}(y|x)\,dy$

## Linear MMSE Estimator (Special Case)

Definition $\hat{Y}_l = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ where $a_i$ are to be determined

orthogonality principle

the best estimator $E\{\varepsilon X_k^*\} = 0$, $k = 1, 2, \cdots, n$

can be used to construct equations and solve $a_1 \sim a_n$

## Nonlinear Orthogonality Rule

Let $h(\underline{X})$ represent any functional form of the data

and $E\{Y|\underline{X}\}$ the best estimator for $Y$

With $\varepsilon = Y - E\{Y|\underline{X}\}$, there is $E\{\varepsilon h(\underline{X})\} = 0$

which implies $\varepsilon = Y - E\{Y|\underline{X}\} \perp h(\underline{X})$

**Theorem** If $X_1, X_2, \cdots, X_n$ and $Y$ are jointly Guassian zero-mean the best estimate is always linear

$$E[\varepsilon X_k^*] = 0, \quad E[\varepsilon] = 0, \quad E[\varepsilon | X] = 0$$

## Linear Prediction

**Problem Setup** $\hat{X}_{n+1} = \sum_{i=1}^{n} a_i X_i$ one-step ahead prediction

$$\begin{cases} \varepsilon_n = X_{n+1} - \hat{X}_{n+1} = X_{n+1} - \sum_{i=1}^{n} a_i X_i \\ E\{\varepsilon_n X_k^*\} = 0, \text{ for } k = 1, 2, \cdots, n \end{cases}$$

Suppose $X_i$ represents the sample of w.s.s. stochastic process

so that $E\{X_i X_k^*\} = r_{i-k} = r^*_{k-i}$

Hence $E\{\varepsilon_n X_k^*\} = E\{X_{n+1} X_k^*\} - \sum_{i=1}^{n} a_i E\{X_i X_k^*\} = 0$

$$\Rightarrow \sum_{i=1}^{n} a_i r_{i-k} = r_{n+1-k}$$

## Wiener-Hopf Equation

$$\begin{pmatrix} r_0 & r_1 & r_2 & \cdots & r_{n-1} \\ r_1^* & r_0 & r_1 & \cdots & r_{n-2} \\ r_2^* & r_1^* & r_0 & \cdots & r_{n-3} \\ & & \vdots & & \\ r_{n-1}^* & r_{n-2}^* & \cdots & r_1^* & r_0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} r_n \\ r_{n-1} \\ r_{n-2} \\ \vdots \\ r_1 \end{pmatrix}.$$

$$Ra = r$$
$$a = R^{-1} r$$
$$\sigma_n^2 = r_0 - r^* R^{-1} r$$

we want to solve

# 8. Markov Chains

**Definition** $X = \{X_n, n = 0, 1, \cdots\}$

provided that $P\{X_{n+1} = j \mid \underline{X_n = i}, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0\}$

$$= P\{X_{n+1} = j \mid \underline{X_n = i}\} = P_{ij}$$

only depends on the current state $X_n$

**Homogeneous** $P_{ij}$ is independent of time $(n)$

**transition matrix** $P = \{P_{ij}\}$ $\begin{cases} \text{for each } i, j, \ P_{ij} \geq 0 \\ \text{for each } i, \ \sum_j P_{ij} = 1 \ (\text{row}) \end{cases}$

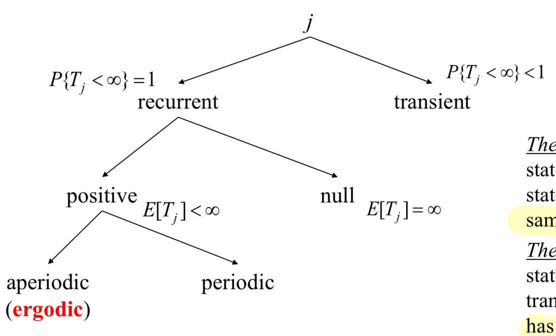
**Chapman-Kolmogorov Equations** $P_{ij}^{n+m} = \sum_{k \in E} P_{ik}^n P_{kj}^m$

$\pi^{(0)}$ is the probability distribution of time 0

then at time $n$, $\pi^{(n)} = \pi^{(0)} P^n$

## Classification of States

**Classification of States**




closed set

absorbing state

irreducible

*Theorem 2*. In an irreducible Markov chain, either all states are transient, all states are recurrent null, or all states are recurrent positive (i.e., all states are of the same type).

*Theorem 3*. In a finite-state Markov chain, all recurrent states are positive, and it is impossible that all states are transient. If the Markov chain is also irreducible, then it has no transient states.

## Limiting probabilities

In an irreducible aperiodic homogenous Markov chain,

$$\pi = \lim_{n\to\infty} \pi^{(n)} = \lim_{n\to\infty} \pi^{(0)} P^n$$

always exists and is independent of $\pi^{(0)}$

① all states are transient / recurrent null $\Rightarrow \pi_j = 0$ for all $j$.

② all states are recurrent positive

$\Rightarrow \pi_j > 0$ for all $j$ and $\pi_j = \sum_{i\in E} \pi_i P_{ij}$, $\sum_{j\in E} \pi_j = 1$

or written as: $\pi = \pi P$, $\pi 1 = 1$

Computing: ① get $c\pi$ ② normalise

Observation:
finite-state irreducible, aperiodic
$\Rightarrow$ the rows of $P^n \to \pi$

If $m_j$ is the expected time between two returns to $j$,
then $\pi_j = \frac{1}{m_j}$. When means the limiting probability of $j$ ($\pi_j$)
= the rate at which $j$ is visited
= the long-run proportion of time that the process is in state $j$

## Ergodic Theorem

If a Markov chain $X = \{X_n, n=0,1,2,\dots\}$ is ergodic,
then for any bounded $f$, $\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \to \underbrace{E[f(X)]}_{\text{with respect to } \pi}$

time average $\longrightarrow$ ensemble average

# 9. Continuous-Time Processes

$$p(X(t_n)=j \mid X(t_1)=i_1, \cdots, \underline{X(t_{n-1})=i_{n-1}}) = p(X=t_n)=j \mid \underline{X(t_{n-1})=i_{n-1}})$$

$\begin{cases} \text{Poisson Process with continuous time and discrete state space} \\ \text{Wiener Process with continuous time and continuous state space} \end{cases}$

## Poisson Process

**definition** $X(t)=n(0,t)$.

$\underline{\lambda: \text{intensity of}}$ $\underline{\text{the Poisson process}}$

features of $n(t_1,t_2)$:

① number of arrivals $n(t_1,t_2)$ in time $(t_1,t_2)$ is a poisson r.v.. $p\{n(t_1,t_2)=k\}=e^{-\lambda t}\frac{(\lambda t)^k}{k!}$

② if $(t_1,t_2)$ and $(t_3,t_4)$ are nonoverlapping, the r.v.s $n(t_1,t_2)$ and $n(t_3,t_4)$ are independent

$$E[X(t)]=E[n(0,t)]=\lambda t$$

$$E[X^2(t)]=E[n^2(0,t)]=\lambda t+\lambda^2 t^2$$

$$R_{XX}(t_1,t_2)=\lambda^2 t_1 t_2+\lambda \min(t_1,t_2) \longrightarrow \text{not a w.s.s process}$$

**properties**

① **Sum** $X_1(t)$ & $X_2(t)$ are two independent Poisson processes

$\Rightarrow X_1(t)+X_2(t)$ is Poisson process, $(\lambda_1+\lambda_2)t$

② **Random Selection**

For Bernoulli r.v. $N_i$. where $P(N_i=0)=p$. $p(N_i=1)=q=1-p$

Define $Y(t)=\sum_{i=1}^{X(t)} N_i$, $Z(t)=\sum_{i=1}^{X(t)}(1-N_i)=X(t)-Y(t)$

$\downarrow$ Poisson process $\lambda p t$

$\downarrow$ Poisson process $\lambda q t$

# inter-arrival distribution

① $\tau_i,\ i=0,1,2,\cdots$ : Inter-arrival duration between events

$$\begin{cases} F_{\tau_i}(t) = 1 - e^{-\lambda t} & \tau_i \overset{i.i.d}{\sim} \exp(\lambda) \\ f_{\tau_i}(t) = \lambda e^{-\lambda t} & E(\tau_i) = \frac{1}{\lambda} \end{cases}$$

② $t_n = \sum_{i}^{n} \tau_i$ : nth arrival

$$\begin{cases} F_{t_n} = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \\ f_{t_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t} \end{cases}$$

(Erlang-n distribution)

# Guassian Process

**Definition** each finite dimensional vector $\{X(t_1), X(t_2), \cdots, X(t_n)\}$ has the multivariate normal distribution $N(\mu(t), \Sigma(t))$

is stationery if $\begin{cases} ① E[X(t)] \text{ is constant for all } t \\ ② \Sigma(t) = \Sigma(t+c) \text{ for all } t \text{ and } c \end{cases}$

# Wiener Process

a special guassian process

a) $W$ has independent increments,

i.e. $W(s+t) - W(s)$ and $W(u+v) - W(u)$ are independent when $[s, s+t]$ and $(u, u+v]$ are nonoverlapping

b) $W(s+t) - W(s)$ is distributed as $N(0, \sigma^2 t)$ for all $s, t \geq 0$

c) sample paths of $W$ are continuous $\Rightarrow$ not stationery

$$E[W(t)] = 0,\quad Var[W(t)] = \sigma^2 t,\quad C(s,t) = \sigma^2 \min\{s,t\}$$

# 10. Martingales

**Definition** $E\{X_{n+1} | X_n, X_{n-1}, \cdots, X_1, X_0\} = X_n$

**Example 1: Gambler's fortune** 'doubling the stake' strategy

**Example 2: Markov chain** that satisfies $\sum_j j P_{ij} = i$

For finite chains of size $N$, $Px = x$, $x \in [1, 2, \cdots, N]^T$

* $P$ cannot be a primitive matrix, no stationery distribution
* Not irreducible, as least two close sets

**Example 3: DeMoivre's Martingale**

win: $p$. lose: $q = 1-p$. Current capital: $S_n$

then $Y_n = \left(\frac{q}{p}\right)^{S_n}$ generates a martingale

## Stopping Time (unpredictable)

$T$ is a stopping time $\Rightarrow$ $E\{X_T\} = E\{X_0\}$

this equation can be used to calculate the probability of stopping states

## Generalisation definition

When $\{X_n\}$ itself is not a martingale,

we can find $\{S_n = \phi(X_n) : n \geq 1\}$ which is a generalised martingale:

$$E[S_{n+1} | X_1, X_2, \cdots, X_n] = S_n$$

**Submartingale** $E[S_{n+1} | X_1, X_2, \cdots, X_n] \geq S_n$ $\uparrow$ less strong

**supermartingale** $E[S_{n+1} | X_1, X_2, \cdots, X_n] \leq S_n$ $\downarrow$ stronger

① martingale = both submartingale & supermartingale

② $\{S_n\}$ is submartingale $\rightarrow$ $\{-S_n\}$ is supermartingale

## Doob Decomposition

submartingale $S_n = \underbrace{M_n}_{\text{martingale}} + \underbrace{Y_n}_{\text{increasing predictable sequence}}$

## Martingale convergence theorem

Let $\{S_n\}$ be a submartingale with finite means.

there exists a random variable $S_\infty$ such that

$$S_n \rightarrow \infty \text{ almost surely as } n \rightarrow \infty$$