Notion of probability:

$$P_k = \lim_{n\to\infty} f_k(n) = \lim_{n\to\infty} \frac{N_k(n)}{n}$$

Number of times

Outcome number

Axioms of probability:

$P(A) \geq 0$ for any A

$P(S) = 1$

$P(A \cup B) = P(A \text{ or } B) = P(A) + P(B)$

Conditional Probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Independent events:

$$P(A,B) = P(A|B)P(B) = P(A)P(B)$$

Cum. Distribution Function (CDF):

$F_X(x) = P[X \leq x]$ for **ALL** x

Prob. Mass Function (PMF):

- $P_X(X=a)$

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Prob. Density Function(PDF):

$$P[a < X < b] = \int_a^b f_X(x)dx$$

Uniform:

$$f_X(x) = \begin{cases} 1/(b-a) & a \leq x < b \\ 0 & \text{otherwise} \end{cases}$$

Gaussian:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Expected Value:

$$E[X] = \mu_X = \sum_{\text{all } x} x P_X(x)$$

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

Variance:

$$Var[X] = \sigma_X^2 = E\left[(X - E[X])^2\right]$$

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

$$Var[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)dx$$

Distributions:

Joint probability mass function

Marginal probability mass functions

Bernoulli:

$$P_X(x) = \begin{cases} 1-p & x = 0 \\ p & x = 1 \\ 0 & \text{otherwise} \end{cases}$$

Pairs of variables:

$$\mathbf{X} = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} |\text{difference between dice}| \\ \text{first die} \end{pmatrix}$$

joint PMF:

$$P(X = x, Y = y) = P_{X,Y}(x,y)$$

marginal PMF:

$$P_X(x) = P_{X,Y}(x, \text{any } y) = \sum_{\text{all } y} P_{X,Y}(x,y)$$

joint CDF:

$$F_{X,Y}(x_1, y_1) = P(X \leq x_1, Y \leq y_1)$$

properties:

- $F_{X,Y}(-\infty, y_1) = P[X \leq -\infty, Y \leq y_1] = 0$
- $F_{X,Y}(\infty, y_1) = P[X \leq \infty, Y \leq y_1] = P[Y \leq y_1] = F_Y(y_1)$

joint PDF:

$$f_{X,Y}(x,y) = \frac{d^2 F_{X,Y}(x,y)}{dx\,dy} = \frac{d^2 P(X \leq x, Y \leq y)}{dx\,dy}$$

properties:

- $f_{X,Y}(x,y) \geq 0$
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx\,dy = 1$

Conditional PMF/PDF:

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)} \qquad f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$f_{X|Y=0}(x) = \frac{f_{X,Y}(x, y=0)}{f_Y(y=0)}$$

marginal PDF:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

j-PDF ⇔ j-CDF

- 1-D Case

$$F_X(a) = \int_{-\infty}^{a} f_X(x)dx$$

- 2-D Case

$$F_{X,Y}(a,b) = \underbrace{\int_{-\infty}^{a}\int_{-\infty}^{b} f_{X,Y}(x,y)dydx}_{P(X \leq a, Y \leq b)}$$

Independent random Variables:

X, Y independent $\Leftrightarrow P_{X,Y}(x,y) = P_X(x)P_Y(y)$

X, Y independent $\Leftrightarrow f_{X,Y}(x,y) = f_X(x)f_Y(y)$

Multivariate Gaussian

- When we have a **pair** of Gaussian random variables, and we write the pair $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi \det(C_{\mathbf{X}})^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_{\mathbf{X}})' C_{\mathbf{X}}^{-1} (\mathbf{x} - \mu_{\mathbf{X}})\right)$$

- Here, the mean vector $\mu_{\mathbf{X}} = \begin{bmatrix} E[X_1] \\ E[X_2] \end{bmatrix}$
- and the **covariance matrix**:

$$C_{\mathbf{X}} = \begin{bmatrix} Var[X_1] & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var[X_2] \end{bmatrix}$$

*Expectation: E[**X**] = E[(X, Y)] = [E[X], E[Y]] ⇒*

$$\begin{bmatrix} \int_{-\infty}^{\infty} x f_X(x)dx \\ \int_{-\infty}^{\infty} y f_Y(y)dy \end{bmatrix}$$

Covariance: $Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$

Independent → Uncorrelated: (Uncorrelated   Independent)

- Correlation $E[XY] = E[X]E[Y]$
- Covariance $Cov(X,Y) = E[XY] - E[X]E[Y] = 0$

- $Cov(X,Y) = 0 \quad \Leftrightarrow$ X and Y are uncorrelated
- $E[XY] = 0 \quad \Leftrightarrow$ X and Y are orthogonal

Expected value X_t:

$$\mu(t_k) = E[X(t_k)] = \int_{-\infty}^{\infty} x f_{X(t_k)}(x)dx$$

RANDOM PROCESSES

Random sinusoidal process: $X_n = A\sin(2\pi f n + \phi)$

$X(t_k) \to f_{X_{t_k}}(x_{t_k}) = f_{X_k}(x_k)$

$X(t_1)...X(t_k)... \to f_{X_1, X_2, ..., X_k, ...}(x_1, x_2, ..., x_k, ...)$

IID Random Processes: →

- all $X(t_k)$ are mutually independent random variables for all t_k
- all $X(t_k)$ have the **same** pdf for all t_k

- *Independent*
- *Identically*
- *Distributed*

$$f_{X_1, X_2, ...}(x_1, x_2, ...) = f_{X_1}(x_1)f_{X_2}(x_2)...$$
$$= f_X(x_1)f_X(x_2)...$$

Autocovariance:

$$C_X(t,\tau) = Cov[X(t), X(t+\tau)]$$
$$= E\left[(X(t) - \mu_X(t))(X(t+\tau) - \mu_X(t+\tau))\right]$$
$$= E\left[X(t)X(t+\tau)\right] - E[X(t)]E[X(t+\tau)]$$
$$C_X(t, \tau = 0) = E[X(t)X(t)] - E[X(t)]E[X(t)]$$
$$= Var[X(t)]$$

Bernoulli:

$$P_{X_k}(x_k) = \begin{cases} p^{x_k}(1-p)^{1-x_k} & x = 0,1 \\ 0 & \text{otherwise} \end{cases}$$

$$P_{X_1, X_2}(x_1, x_2) = P_{X_1}(x_1)P_{X_2}(x_2)$$
$$= p^{x_1}(1-p)^{1-x_1}p^{x_2}(1-p)^{1-x_2}$$
$$= p^{x_1+x_2}(1-p)^{2-x_1-x_2}$$

$$P_{X_1...X_n}(x_1, ..., x_n) = \begin{cases} p^{x_1+...+x_n}(1-p)^{n-(x_1+...+x_n)} & x_i = 0,1 \\ 0 & \text{otherwise} \end{cases}$$

Uncorrelated Processes:

- *∀ X(t), X(t + τ) are uncorrelated →* $C_X(t,\tau) = \begin{cases} var(t) & \text{for all } t \text{ and } \tau = 0 \\ 0 & \text{for all } t \text{ and } \tau \neq 0 \end{cases}$
- *∀ X(t), X(t + τ) are orthogonal →* $R_X(t,\tau) = \begin{cases} E[X^2(t)] & \text{for all } t \text{ and } \tau = 0 \\ 0 & \text{for all } t \text{ and } \tau \neq 0 \end{cases}$

Counting: counting nº of Bernoulli vars X_i = 1 → $P_{N(k)}(j) = \binom{k}{j} p^j (1-p)^{k-j}$

Poisson: Nº of events in [0, t] with rate λ

$$P[N(t) = n] = P_{N(t)}(n) = \frac{(\lambda t)^n}{n!}\exp(-\lambda t)$$

Stationary Processes:

- *Every joint-pdf shift invariant: →* $f_{X(t_1), X(t_2), ..., X(t_k)}(x_1, x_2, .., x_k) =$
$f_{X(t_1+\Delta t), X(t_2+\Delta t), ..., X(t_k+\Delta t)}(x_1, x_2, .., x_k)$

Consequences:

- *marginal pdfs independent of t:* $f_{X(t)}(x) = f_{X(t+\Delta t)}(x) = f_X(x)$
- *expected value and variance independent of t:* $\mu_X(t) = E[X(t)] = \mu_X$ | $Var_X(t) = Var[X(t)] = Var[X] = \sigma_X^2$
- *2d joint-pdf shift invariant: → only 'distance' t_1 t_2 matters:*

$$f_{X(t_1), X(t_2)}(x_1, x_2) = f_{X(t_1+\Delta t), X(t_2+\Delta t)}(x_1, x_2) \qquad R_X(t,\tau) = R_X(\tau)$$
$$= f_{X(0), X(t_2 - t_1)}(x_1, x_2) \qquad C_X(t,\tau) = C_X(\tau) = R_X(\tau) - \mu_X^2$$

Wide-Sense stationary Processes:

- *Show processes are stationary is 'impossible'*
- *A process is wide-sense stationary ←→ :*

$\mu_X(t) = \mu$ for all t

$R_X(t,\tau) = R_X(\tau)$ for all t

Stationary Process
reflects properties of the joint-pdf

Wide-Sense Stationary process
reflects properties (only) of
•expected value
•autocorrelation function

$\mu_X(n) = \mu$ for all n

$R_X(n,k) = R_X(k)$ for all n

AutoCorrelation Function:

$R_X(0) \geq 0$

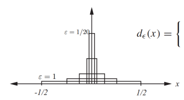
$R_X(k) = R_X(-k)$

$|R_X(k)| \leq R_X(0)$

$\lim_{k\to\infty} R_X(k) = \mu_X^2$

- With random time signals, we often (sometimes even implicitly) assume $E[X(k)] = 0$

Delta Function: →



$$d_\epsilon(x) = \begin{cases} 1/\epsilon & -\epsilon/2 \leq x \leq \epsilon/2, \\ 0 & \textit{otherwise.} \end{cases}$$

$$\delta(x) = \lim_{\epsilon\to 0} d_\epsilon(x).$$

- The delta function is used to simulate a discrete probability mass with a continuous function
- Nice property: $\int_{-\infty}^{\infty} g(x)\delta(x - x_0)\,dx = g(x_0).$

Estimation of the mean (using sample):

Use: $f_X(x) = \frac{1}{n}\sum_{i=1}^{n}\delta(x - x_i)$

Then: $E[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{n}\sum_{i=1}^{n}\delta(x - x_i)dx = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$M_n(k) = \frac{1}{n}\sum_{i=1}^{n} X_i(k)$$

- *Many Realizations:* =constant for all k!
- *Over time:* $\tilde{M}_n(k) = \frac{1}{n}\sum_{i=1}^{n} X(i)$ *→ (only for WSS **ergodic** processes)*

Estimated Autocorrelation function

- Instead of autocorrelation function based on ensembles

$$R_X(m)=\frac{1}{n}\sum_{i=1}^{n}X_i(k)X_i(k+m)$$

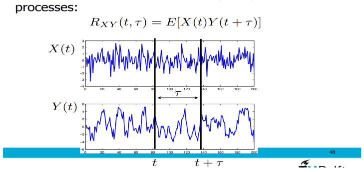
the autocorrelation function is estimated on the basis of a single realization

$$\tilde{R}_X(m)=\frac{1}{n}\sum_{i=1}^{n}X(i)X(i+m)$$

Cross-correlation function

- Cross-correlation, defined for two jointly WSS stochastic processes:

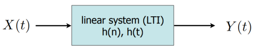
$R_{XY}(t,\tau)=E[X(t)Y(t+\tau)]$



$$\begin{aligned}R_{XY}(\tau)&=E[X(t)Y(t+\tau)]\\&=E[Y(t+\tau)X(t)]\\&=E[Y(t)X(t-\tau)]\\&=R_{YX}(-\tau)\end{aligned}$$

Random Signal Processing

Linear Time Invariant systems:



$$y(n)=\sum_{k=-\infty}^{\infty}h(k)x(n-k)=h(n)*x(n)$$ $$y(n)=\sum_{k=-\infty}^{\infty}h(k)x(n-k)=h(n)*x(n)$$

E[] of filtered WSS process:

$$E[Y(n)]=E[h(n)*X(n)]=E\left[\sum_{k=-\infty}^{\infty}h(k)X(n-k)\right]$$

$$=\sum_{k=-\infty}^{\infty}h(k)E[X(n-k)]=E[X(n)]\cdot\sum_{k=-\infty}^{\infty}h(k)=E[X(n)]H_0$$

$$\mu_Y=E[X(n)]H_0=\mu_X H_0$$

$$\mu_Y=E[Y(t)]=E[X(t)]H_0=\mu_X\int_{-\infty}^{\infty}h(t)dt$$

R(k) for filtered WSS process:

$$\begin{aligned}R_Y(k)&=E[Y(n)Y(n+k)]\\&=E[(h(n)*X(n))(h(n)*X(n+k)]\\R_Y(k)&=E[\sum_m h(m)X(n-m)\sum_p h(p)X(n+k-p)]\\&=\sum_m h(m)\sum_p h(p)E[X(n-m)X(n+k-p)]\\&=\sum_m h(m)\sum_p h(p)R_X(k-p+m)\\&=h(-k)*h(k)*R_X(k)\end{aligned}$$

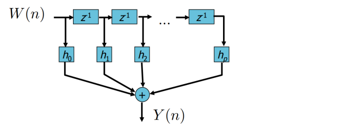
White Gaussian noise:

- W(t) White Gaussian noise iff:

1. W(t) is stationary
2. Expected value $\mu_W=0$
3. Autocorrelation function $R_W(\tau)=\eta_0\delta(\tau)$

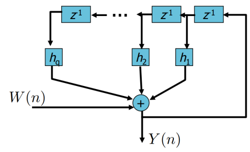
- Note: mathematically nice, but physically impossible!
- Because **average power**: $E[W^2(t)]=R_W(0)=\infty$

- Tapped delay line or Finite Impulse Response Filter (FIR) system



$$Y(n)=h_0W(n)+h_1W(n-1)+...+h_pW(n-p)$$

- Infinite Impulse Response (IIR) system



$$Y(n)=h_1Y(n-1)+h_2Y(n-2)+...+h_qY(n-q)+W(n)$$

If we input **white noise** into these LTI systems, then (ARMA)

- output of FIR filter is called a **moving average** (MA) process
- output of IIR filter is called an **autoregressive** (AR) process

- First order AR process $Y(n)=h_1Y(n-1)+W(n)$
- Combined

$$R_Y(k)=\begin{cases}h_1R_Y(k-1) & k>0\\ h_1R_Y(k+1) & k<0\\ \frac{1}{1-h_1^2}\sigma_W^2 & k=0\end{cases}\longrightarrow R_Y(k)=\frac{\sigma_W^2}{1-h_1^2}h_1^{|k|}$$

The Optimal Predictor:

- Predictor: $\hat{x}(n)=\sum_{k=1}^{N}h_kx(n-k)$ Find these weights h_k
- Minimize: $\sigma_{\Delta x}^2=E[(x(n)-\hat{x}(n))^2]$
  - note that (here) the prediction takes place on original data
- Solution for N linear prediction coefficients h_k

$$\frac{\partial\sigma_{\Delta x}^2}{\partial h_i}=0\Rightarrow R_x(i)=\sum_{k=1}^{N}h_kR_x(i-k)\qquad i=1,2,\dots,N$$

- Since

$$R_Y(\tau)=h(\tau)*h(-\tau)*R_X(\tau)$$

Fourier transform

- then

$$\begin{aligned}S_Y(f)&=H(f)H^*(f)S_X(f)\\&=|H(f)|^2S_X(f)\end{aligned}$$

Usage of Fourier transforms. The power of Spectral Density (PSD)

***used to:** describe det. signals and linear systems | analyze & design signals and lin. sys*

- **Power Spectral Density (PSD)**

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

$$S_X(f)=F\{R_X(k)\}=\sum_{k=-\infty}^{\infty}R_X(k)\exp(-j2\pi fk)$$

PSD only exists for WSS random processes!

- Continuous time:

$$S_X(f)=F\{R_X(\tau)\}=\int_{-\infty}^{\infty}R_X(\tau)\exp(-j2\pi f\tau)d\tau$$
$$R_X(\tau)=\int_{-\infty}^{\infty}S_X(f)\exp(j2\pi f\tau)df$$

- Discrete time:

$$S_X(f)=F\{R_X(k)\}=\sum_{k=-\infty}^{\infty}R_X(k)\exp(-j2\pi fk)$$
$$R_X(k)=\int_{-1/2}^{1/2}S_X(f)\exp(j2\pi fk)df$$

***properties:** $S_X(-f)=S_X(f)$ | always >= 0 | integral over PSD is av. pow*

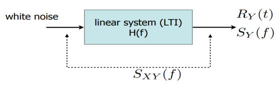
$$\int_{-\infty}^{\infty}S_X(f)df=E[X(t)^2]=R_X(0)$$

Cross relation function:

$$R_{XY}(\tau)=h(\tau)*R_X(\tau)$$
$$S_{XY}(f)=H(f)S_X(f)$$

cross power spectral density

Application: system identification



- When feeding white noise into a linear system, we obtain

$$S_{XY}(f)=H(f)S_X(f)=\frac{N_0}{2}H(f)$$

PREDICTION

Mean squared error: $e=E\left[(X-\hat{x}_B)^2\right]$ → minimize → set derivative 0: $\hat{x}_B=E[X]$

$$\begin{aligned}\hat{x}_M&=E[X|Y=y]\\&=\int_{-\infty}^{\infty}xf_{X|Y}(x|y)dx\\&=\int_{-\infty}^{\infty}x\frac{f_{X,Y}(x,y)}{f_Y(y)}dx\end{aligned}$$

$$C(X,\hat{x}_U)=\begin{cases}0 & |X-\hat{x}_U|<\varepsilon\\ 1 & \text{otherwise}\end{cases}$$ (very small value)

- Minimise:

$$e=E[C(X,\hat{x}_U)]=\arg\max_{\hat{x}_U}f_{X|Y}(\hat{x}_U|y)$$

Sample mean:

$$M_n=\frac{1}{n}\sum_{i=1}^{n}X_i\qquad E[M_n]=\frac{1}{n}(nE[X])=\mu_X$$

Linear Estimation of X given Y:

$\hat{x}_L(y)=ay+b$ → min: $e_L=E\left[(X-\hat{x}_L(Y))^2\right]$

Solve:

$$a^*=\frac{Cov[X,Y]}{Var[Y]}\qquad b^*=E[X]-a^*E[Y]$$

Markov & Chebyshev Inequalities:

$$P[X\geq c^2]\leq\frac{E[X]}{c^2}\qquad P[|Y-\mu_Y|\geq c]\leq\frac{Var[Y]}{c^2}$$

Variance of Sample Mean:

$$Var[M_n]=\frac{1}{n}Var[X]$$

Weak Law of large Numbers:

$$\lim_{n\to\infty}P[|M_n-\mu_X|>\varepsilon]=0$$

Probability Density Function of M_n:

$$f_{W_n}(w)=f_{X_1}(w)*f_{X_2}(w)*...*f_{X_n}(w)$$

Central Limit Theorem:

- When X_i are iid random variables
  - with expected value μ_X
  - and variance σ^2
- then for $W_n=X_1+X_2+...+X_n$

$$f_{W_n}(w)=f_{X_1}(w)*f_{X_2}(w)*...*f_{X_n}(w)$$

- can be approximated by a Gaussian pdf
  - with expected value $n\mu_X$
  - and variance $n\sigma_X^2$

MARKOV CHAINS

Markov property: *Conditional PMF of X_{n+1} only depends on X_n*

$$P_{ij}=P(X_{n+1}=j|X_n=i)\qquad\sum_{j=0}^{\infty}P_{ij}=1$$

time new state old state

State Transition Mtx:

$$\mathbf{P}=\begin{bmatrix}P_{00}&P_{01}&P_{02}\\P_{10}&P_{11}&P_{12}\\P_{20}&P_{21}&P_{22}\end{bmatrix}=\begin{bmatrix}0&1.0&0\\0&0.7&0.3\\0.5&0.3&0.2\end{bmatrix}$$

State Probabilities:

$$\mathbf{p}(n)=\mathbf{p}(n-1)\mathbf{P}$$
$$\mathbf{p}(n)=\mathbf{p}(0)\mathbf{P}^n$$

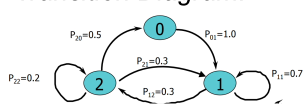
Limiting State Probabilities:

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

$\mathbf{p}(n)=\mathbf{p}(n-1)\mathbf{P}$ ⇨ $\pi=\pi\,\mathbf{P}$

$$\pi_0+\pi_1+\pi_2=1$$

Transition Diagram:



State classification:

- **Accessibility**: State j is accessible from state i if $P_{ij}(n)>0$ for some n>0. (i→j)
- **Communicating states**: States i and j communicate if i→j and j→i. (i ↔ j)
- **Communicating Class**: A nonempty subset of states in which all states communicate
- **Periodic**: State i has period d, if d is the largest integer that divides the length of all paths to state i. (gcd)
- **Aperiodic**: State i is aperiodic if d=1.
- All states in a communicating class have the same period.