

Random Vectors

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}, \quad \mathbb{E}[\mathbf{X}] = \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_m] \end{bmatrix}$$

Correlation (b/t a random vector w/ itself)

$$\text{Notice Symmetry} \quad \mathbb{E}[\mathbf{X} \mathbf{X}^T] = \begin{bmatrix} \mathbb{E}[X_1^2] & \mathbb{E}[X_1 X_2] & \cdots & \mathbb{E}[X_1 X_m] \\ \mathbb{E}[X_2 X_1] & \mathbb{E}[X_2^2] & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[X_m X_1] & \mathbb{E}[X_m X_2] & \cdots & \mathbb{E}[X_m^2] \end{bmatrix}$$

Variance of \mathbf{X}_{max}

$$\mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}]) (\mathbf{X} - \mathbb{E}[\mathbf{X}])^T] = \begin{bmatrix} \text{Var}(X_1) \text{Cov}(X_1, X_2) \cdots \text{Cov}(X_1, X_m) \\ \text{Cov}(X_2, X_1) \text{Var}(X_2) \cdots \vdots \\ \vdots \\ \text{Cov}(X_m, X_1) \text{Cov}(X_m, X_2) \cdots \text{Var}(X_m) \end{bmatrix}_{m \times m}$$

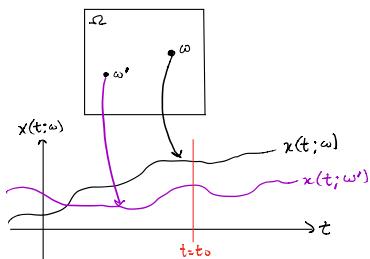
Cross-Correlation (b/t two random vectors)

$$\mathbb{E}[\mathbf{X} \mathbf{Y}^T] = \left\{ \mathbb{E}[X_i Y_j] \right\}_{1 \leq i \leq m, 1 \leq j \leq n}$$

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = \left\{ \text{Cov}(X_i, Y_j) \right\}_{1 \leq i \leq m, 1 \leq j \leq n}$$

Random Process

Each sample $\omega \in \Omega$ in the sample space is mapped to a time function $X(t, \omega)$



Sampling a random process at a particular time t_0 , you get a random variable $X(t_0, \omega)$, $X(t_0, \omega)$ random variables

$X(t; \omega)$ is a deterministic function of t , called a realization or a sample path of this random process

$X(t) \rightarrow$ r.p. $x(t) \rightarrow$ sample path

Stationarity: A random process is stationary if its statistical characteristics do not change over time.

Specifically,

$$\Delta \quad F_{X(t_1, \dots, t_n)}(x_1, \dots, x_n) = F_{X(t_1+\Delta, \dots, t_n+\Delta)}(x_1+\Delta, \dots, x_n+\Delta)$$

WIDE Sense Stationary: A random process is WSS if

$$(i) \mathbb{E}[X(t)] = \mu$$

$$(ii) R_{X(t)}(t_1, t_2) = \mathbb{E}[X(t_1)X(t_2)]$$

$$\text{where } \tau = t_2 - t_1$$

Independence Review

Independent Events: $A \perp B$ if $P[A \cap B] = P[A]P[B]$
or equivalently: $P[A|B] = P[A]$

Independent Random Variables:

$$X \perp\!\!\!\perp Y \text{ if } f_{XY}(x, y) = f_X(x)f_Y(y)$$

or equivalently,

$$f_{XY}(\omega_1, \omega_2) = f_X(\omega_1)f_Y(\omega_2)$$

Conditional Independent Events

$$A \perp B \text{ conditioned on } C \text{ if } P[A \cap B | C] = P[A|C]P[B|C]$$

or equivalently,

$$P[A | B \cap C] = P[A | C]$$

Conditional Independent Random Variables:

$$X \perp\!\!\!\perp Y \text{ if } f_{X|Z}(x|z)f_{Y|Z}(y|z)$$

or equivalently,

$$f_{XY|Z}(\omega_1, \omega_2 | z) = f_X(\omega_1 | z)f_Y(\omega_2 | z)$$

Increment: The increment of a r.p. $\{X(t)\}$ over an interval $[a, b]$ is the r.v. $X(b) - X(a)$ another r.v.

Independent Increments: $\{X(t)\}$ has independent increments if $\forall t_1, \dots, t_n$, and for all $t_1 < t_2 < \dots < t_n$ the increments $X(t_1, t_2), \dots, X(t_{n-1}, t_n)$ are independent

Stationary Increments

$\{X(t)\}$ has stationary increments if the distribution of $X(t + \tau) - X(t)$ depends only on τ , not t .

Key Idea: When characterizing processes w/ independent increments over non-overlapping intervals

Counting Function

$f(t) \quad (t \geq 0)$ is a counting function if $f(0) = 0$, $f(t)$ takes non-negative integers, is non-decreasing, and is right continuous.

Binomial Counting Process

$$X_i \sim \text{Bernoulli}(p)$$

$$S_n = \sum_{i=1}^n X_i$$

$$S_n \sim \text{Binomial}(n, p)$$

$$\Pr[S_n = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

T_1, T_2, T_3 are a sequence of inter-arrival times

Interarrival Time $\{T_i\}_{i=1}^\infty$

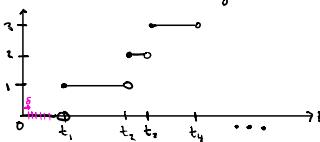
In above example,

$$T_i \sim \text{Geometric}(p)$$

$$\Pr[T_1 = k] = (1-p)^{k-1} p$$

What about continuous counting processes?

Say events arrive continuously w/ rate λ avg arrival/unit time



Partition time into δ -length intervals.

Assumptions:

For $\delta \rightarrow 0$

(1) The probability of having more than one arrival in δ -interval is negligible ($\delta \rightarrow 0$)

(2) Whether there is an arrival within a δ -interval is independent of arrivals in other δ -intervals

$$\{p = \lambda \delta\}$$

These assumptions make this similar to the binomial discrete case w/ $p = \lambda \delta$

$$\delta \rightarrow 0$$

$$p = \lambda \delta \rightarrow 0$$

$$n = \frac{\lambda}{\delta} \rightarrow \infty$$

$$S_n \sim \text{Binomial}\left(\frac{\lambda}{\delta}, \delta\right) \rightarrow \text{Poisson}(\lambda t)$$

Properties

(1) If JG, Uncorrelated \Rightarrow independence NOT generally true

$\text{Cov}(X, Y) = 0$ joint distribution factors into product of marginal distributions

(2) If you have two rvs which are marginally Gaussian AND they're independent, can conclude they are jointly Gaussian

(3) If XY jointly Gaussian then XM is Gaussian USEFUL

$$\text{mean: } E[X|Y] = X_{\text{mean}} + E[Y] \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}$$

$$\text{variance: } \text{Var}(X|Y) = \text{Var}(X) - \text{Cov}^2(X, Y) / \text{Var}(Y)$$

$$= \text{Var}(X)(1 - \rho_{XY}^2)$$

$$= \text{Var}(X)(1 - \rho_{XY}^2)$$

For every realization y , we get the same result!

a CONSTANT

Gaussian Random Processes

Definition:

$X(t)$ is Gaussian if $\forall n$, and $\forall t_1, \dots, t_n$ r.v.s

$X(t_1), \dots, X(t_n)$ are JOINTLY Gaussian

"Completely specified by first two moments"

Properties

(1) completely specified by its mean function $\mu(t)$ and autocorrelation function $R_{XX}(t)$ / $\text{autocovariance function } C_{XX}(t)$.

Example $f_{X(t_1), X(t_2)}(t_1, t_2) \sim N\left(\begin{bmatrix} \mu(t_1) & \mu(t_2) \\ \mu(t_2) & \mu(t_1) \end{bmatrix}, \begin{bmatrix} C_{XX}(t_1, t_1) & C_{XX}(t_1, t_2) \\ C_{XX}(t_2, t_1) & C_{XX}(t_2, t_2) \end{bmatrix}\right)$

(2) WSS \Rightarrow S

Definition: A Brownian Motion (also called a Wiener Process) with parameter σ^2 is a random process $\{X(t)\}_{t \geq 0}$ such that

$$(1) X(0) = 0$$

(2) $\{X(t)\}_{t \geq 0}$ has independent increments

$$(3) X(t_2) - X(t_1) \sim N(0, \sigma^2(t_2 - t_1))$$

(4) Every sample path is continuous

Properties of Brownian Motion

$$(1) \mu(t) = \mathbb{E}[X(t)] = \mathbb{E}[X(t) - X(0)] = 0$$

$$(2) R_{XX}(t) = \mathbb{E}[X(t)X(t)] \quad t \geq 0$$

$$= \mathbb{E}[X(t)(X(t_1) + X(t_2) - X(t_1))]$$

$$= \mathbb{E}[X(t_1)] + \mathbb{E}[X(t_1)]\mathbb{E}[X(t_2) - X(t_1)]$$

just illustrates $\rightarrow = \mathbb{E}[(X(t_1) - X(0))^2] + \mathbb{E}[X(t_1) - X(0)]\mathbb{E}[X(t_2) - X(t_1)] = \sigma^2 t$

In general,

$$C_{XX}(t_1, t_2) = R_{XX}(t_1, t_2) = \min(t_1, t_2) \sigma^2$$

$$(3) X(t) \sim N(0, \sigma^2 t)$$

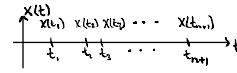
Also have $X(t_1), X(t_2), \dots, X(t_n) \quad t \in \mathbb{R}$

$$\begin{bmatrix} X(t_1) \\ X(t_2) \\ \vdots \\ X(t_n) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} X(0) \\ X(t_1) - X(0) \\ X(t_2) - X(0) \\ \vdots \\ X(t_n) - X(0) \end{bmatrix}$$

Markov Process

$\forall n$,

A random process $\{X(t)\}_{t \geq 0}$ is a Markov process if and $t, t_1, t_2, \dots, t_{n-1}$



$$\Pr[X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1] = \Pr[X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n]$$

i.e. conditioned on where you are now, the future is independent of the past.

We call the value we see the state.

Once you know the current state,

future state \perp past states

State: $X(t)$ is called the state at time t .

All the values $X(t)$ on time is called the state-space \mathcal{X}

If \mathcal{X} is discrete, then we call this a Markov chain

Jointly Gaussian

$$X \sim \mathcal{N}(\mu, \sigma^2) \text{ if } f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mathbb{E}[X] = \mu$$

$$\text{Var}(X) = \sigma^2$$

X and Y are jointly Gaussian if

$Z = aX + bY$ is Gaussian $\forall a, b \in \mathbb{R}$

For jointly Gaussian X and Y w/ $|D_{XY}| < 1$

$$f_{XY}(x, y) = \frac{1}{2\pi\sqrt{|D_{XY}|}} e^{-\frac{1}{2} \frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{1}{2} \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{1}{2} \frac{(xy - \mu_{XY})(x\mu_Y - y\mu_X)}{\sigma_X \sigma_Y \rho_{XY}}}$$

where

$$K = \begin{bmatrix} \text{Cov}(X, X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Cov}(Y, Y) \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho_{XY} \sigma_X \sigma_Y \\ \rho_{XY} \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix}$$

$$K^{-1} = \frac{1}{1 - \rho_{XY}^2} \begin{bmatrix} \sigma_X^2 & -\rho_{XY} \sigma_X \sigma_Y \\ -\rho_{XY} \sigma_X \sigma_Y & \sigma_Y^2 \end{bmatrix}$$

Discrete-Time Markov Chain

$\{X_n\}_{n=0}^{\infty}$ is a random process

K , $\{x_i\}_{i=0}^k$

$x_0, \dots, x_k \rightarrow$ need joint pmf to describe it

$$P_r[X_0=x_0, X_1=x_1, \dots, X_k=x_k]$$

$$= P_r[X_0=x_0] P_r[X_1=x_1 | X_0=x_0] P_r[X_2=x_2 | X_0=x_0, X_1=x_1] \cdots P_r[X_k=x_k | X_0=x_0, \dots, X_{k-1}=x_{k-1}]$$

one-step transition probability

i.e. for Markov process only need the marginal distribution of the initial state and the transition probabilities from one state to the next.

Need $P(x_0)$ and $\{P[x_{n+1} | X_n=i]\}_{i,j \in \mathcal{X}}$

Homogeneous Markov Chain \rightarrow Time INVARIANT

$$P_r[X_{n+1}=j | X_n=i] = P_{i,j} \quad \forall n$$

Transition Matrix:

$$P = \begin{bmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,n} \\ P_{1,0} & P_{1,1} & \cdots & P_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n,0} & P_{n,1} & \cdots & P_{n,n} \end{bmatrix} \quad \sum_j P_{i,j} = 1 \quad \forall i$$

n -Step Transition Matrix

$$P^{(n)} = \left\{ P_{i,j} \triangleq P_r[X_n=j | X_0=i] \right\}$$

Chapman-Korogorov Equation

$$P_{i,j}^{(m+n)} = \sum_{k \in \mathcal{X}} P_{i,k}^{(m)} P_{k,j}^{(n)}$$

$$\begin{aligned} P^{(m+n)} &= P^{(m)} P^{(n)} \\ P^{(n)} &= P P^{(n-1)} \\ &= P P P^{(n-2)} \\ &\vdots \\ &= P^n \end{aligned}$$

State Probability

This is the distribution at time n .

$$p(n) = [Pr[X_0=1] \ \cdots \ Pr[X_0=m]]^T \text{ vector}$$

$$p^{(0)} = [Pr[X_0=0] \ \cdots \ Pr[X_0=m]]$$

$$p^{(1)} = p^{(0)} P$$

$$\vdots$$

$$p^{(n)} = p^{(0)} P^n$$

$$\{X_n\}_{n=0}^{\infty} : \tilde{p}(0) = \underbrace{\Pr[X_0=1, \dots, X_0=m]}_{\substack{\text{pmf of } X_0 \\ P}} \quad \{X_n\}_{n=0}^{\infty} : \tilde{p}(n) \triangleq \Pr[X_0=1, \dots, X_n=m]$$

$$\Pr[X_n=j] = \sum_{i=1}^m \Pr[X_0=i] \Pr[X_0=j | X_0=i] = \tilde{p}(0) P^{(n)}$$

Cauchy-Schwartz Inequality

$$|E[XY]| \leq \sqrt{|E[X^2]| |E[Y^2]|}$$

Correlation Coefficient from Cauchy-Schwartz inequality

$$P_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \quad (|P_{X,Y}| \leq 1)$$

If X_1, \dots, X_m are pairwise uncorrelated, then

$$\text{Var}(X_1 + X_2 + \dots + X_m) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_m)$$

$$\text{Var}\left(\frac{1}{m} \sum_i X_i\right) = \frac{1}{m} \sum_i \text{Var}(X_i)$$

Gaussian Random Vectors

X is a Gaussian random vector if its coordinates are JOINTLY GAUSSIAN

$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}$ a Gaussian random vector \Leftrightarrow $a_i X_i + \dots + a_m X_m$ is Gaussian $\forall a_i \in \mathbb{R}, i \in \mathbb{N}$.

Use $X \sim N(\mu, K)$ to denote Gaussian random vector

If $X \sim N(\mu, K)$, then

① Any subvector of X is a Gaussian random vector

$$- X_i \sim N(\mu_i, K_{ii})$$

$$- \begin{bmatrix} X_i \\ X_j \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_i \\ \mu_j \end{bmatrix}, \begin{bmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{bmatrix}\right)$$

② $\forall A, b$

$$Y = AX + b$$

is a Gaussian random vector

$$Y \sim N(A\mu + b, AKA^T)$$

$$\mathbb{E}[Y] = A\mu + b$$

③ If K is diagonal, every $r.v.$ in \mathcal{X} is independent of every other $r.v.$ in \mathcal{X}

④ If X_{mn} and Y_{mn} are jointly Gaussian then they are independent iff $\text{Cov}(X, Y) = 0$

NMSE Estimate of X using Y

$$\text{NMSE} = \mathbb{E}\left[\|X - \hat{X}\|^2\right] = \mathbb{E}\left[\left(X - \hat{X}\right)^T \left(X - \hat{X}\right)\right] = \sum_{i=1}^m \mathbb{E}[X_i - \hat{X}_i]^2$$

Need to "design"

$$\hat{X} = g(Y) = \begin{bmatrix} g_1(Y) \\ g_2(Y) \\ \vdots \\ g_m(Y) \end{bmatrix}$$

Problem decoupled into m independent NMSE estimation problems, one for each X_i .

$$\hat{X}_i = g_i(Y)$$

Our best MMSE estimator for X_i is

$$\mathbb{E}[X_i | Y]$$

Need conditional distribution

$$f_{X_i | Y}(x_i | y) = \frac{f_{X_i, Y}(x_i, y)}{f_Y(y_1, y_2, \dots, y_m)}$$

LMMSE Estimator of X

Similar to rv case,

$$\begin{aligned} \hat{X} &= A\hat{Y} + b \\ &= \mathbb{E}[X] + \text{Cov}(X, Y) \text{Cov}^{-1}(Y)(Y - \mathbb{E}[Y]) \end{aligned}$$

If X, Y are jointly Gaussian, (similar to rv case)

$$\mathbb{E}[X | Y] = \mathbb{E}[X] + \text{Cov}(X, Y) \text{Cov}^{-1}(Y)(Y - \mathbb{E}[Y])$$

Poisson Process:

Definition: $\{N(t)\}_{t \geq 0}$ is Poisson process w/ rate λ if it is a counting process w/ independent increments and $N(t) - N(s) \sim \text{Pois}(\lambda(t-s)) \quad \forall t > s$.

Interarrival Time (note: T_i iid $\forall i$)

Look at T_1

$$\Pr[T_1 > t] = (1-\lambda)^t$$

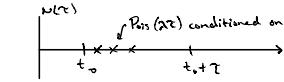
$$\lim_{t \rightarrow \infty} (1-\lambda)^t = e^{-\lambda t}$$

$$\therefore T_1 \sim \exp(\lambda) \quad \mathbb{E}[T_1] = \frac{1}{\lambda}$$

leads to ANOTHER definition

Definition: $\{N(t)\}_{t \geq 0}$ is a counting process w/ rate λ if it is a counting process w/ interarrival times iid exponentially distributed

Definition: $\{N(t)\}_{t \geq 0}$ is a counting process w/ rate λ if it is a counting process such that $\forall T > 0$, $N(T) \sim \text{Pois}(\lambda T)$, and given $N(t)=n$ (n arrival times in interval T), are iid $\text{Un}[0, T]$



Facts about Poisson Process #arrivals = N(t) = 2

① Process has independent increments

$$N(t_1) - N(t_0) \perp \!\!\! \perp N(t_2) - N(t_1)$$

length doesn't matter on interval, they just need NOT overlap

② #arrivals in interval $[t_1, t_2] = N(t_2) - N(t_1)$

where $N(t_2) - N(t_1) \sim \text{Pois}(\lambda(t_2 - t_1))$

③ The interarrival times $\{T_i\}_{i=1}^{\infty}$ are iid

where

$$T_i \sim \exp(\lambda)$$

④ Given $N(T) = k$, these k arrivals are iid $\text{Un}[0, T]$

i.e. $t_i \sim \text{Un}[0, k]$

⑤ $\mu_n(t) = \mathbb{E}[N(t)] = \lambda t$ i.e. arrival rate * time

\text{Var}(N(t)) = \lambda t \quad (\text{Poisson property})

$R_n(s, t) \triangleq \mathbb{E}[N(s)N(t)]$ assume $s \leq t$ if $s < t$ have var.

$$= \mathbb{E}[N(s)(N(s) + N(t-s) - N(s))] = \mathbb{E}[N(s)(N(t-s))] + \mathbb{E}[N(t-s)(N(t-s))] - \mathbb{E}[N(s)]\mathbb{E}[N(t-s)]$$

$$= \lambda s + (\lambda s)^2 + \mathbb{E}[N(s)]\mathbb{E}[N(t-s)] - \lambda s \cdot \lambda s = \lambda s + (\lambda s)^2 + \lambda s(\lambda(t-s))$$

$$= \lambda s + (\lambda s)^2 + \lambda s(\lambda(t-s)) \quad \text{if } t-s < \text{var}$$

$Cov(s, t) = \lambda \min(s, t) \quad \leftarrow \text{applies } \forall s, t \text{ (note above is assumed var.)}$

$$\begin{cases} \hat{X}_{\text{MLE}} = \mathbb{E}[X | Y] \\ \text{NSE}(\hat{X}_{\text{MLE}}) = \mathbb{E}[\text{Var}(X | Y)] \end{cases}$$

Conditional Expectation

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

$\mathbb{E}[X | Y]$ is a random variable which takes on value $\mathbb{E}[X | Y=y]$ w/ density $f_Y(y)$.

Since $\mathbb{E}[X | Y]$ is r.v. can take its expectation.

$$\mathbb{E}_Y[\mathbb{E}_{\text{nr}}[X | Y]] = \mathbb{E}[X]$$

