Random Vectors

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

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

Notice Symmetry

$$\mathbb{E}[\underbrace{\underline{X}}_{m\times 1}\underbrace{\underline{X}^T}_{1\times m}] = \begin{bmatrix} \mathbb{E}[X_1^2] & \mathbb{E}[X_1X_2] & \cdots & \mathbb{E}[X_1X_m] \\ \mathbb{E}[X_2X_1] & \mathbb{E}[X_2^2] & \cdots & \\ \vdots & & \ddots & \\ \mathbb{E}[X_mX_1] & \mathbb{E}[X_mX_2] & \cdots & \mathbb{E}[X_m^2] \end{bmatrix}_{m\times m}$$

Variance of $\underline{X}_{m\times 1}$

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

K Covariance matrix

Cross-Correlation (b/t two random vectors)

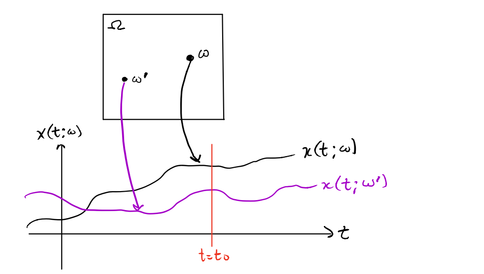
$$\mathbb{E}[\underbrace{\underline{X}}_{m\times 1}\underbrace{\underline{Y}^T}_{1\times n}]_{m\times n} = \{\mathbb{E}[X_iY_j]\}_{\substack{1\le i\le m \\ 1\le j\le n}}$$

i rows
j columns

$$Cov(\underline{X},\underline{Y}) = \{Cov(X_i,Y_j)\}_{\substack{1\le i\le m \\ 1\le j\le n}}$$

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

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

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

Specifically,

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

$\forall n \quad \{t_i\}_{i=1}^n \quad \Delta$

WIDE Sense Stationary: A random process is WSS if

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

(ii) $R_X(t_1,t_2) \triangleq \mathbb{E}[X_1(t)X_2(t)]$
$= R(\tau)$

where $\tau = t_2 - t_1$

Independence Review

Independent Events: $A \perp\!\!\!\perp B$ if $Pr[A\cap B] = Pr[A]Pr[B]$

or equivalently: $Pr[A|B] = Pr[A]$

Independent Random Variables:

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

or equivalently,

$$f_{X|Y}(x|y) = f_X(x)$$

Conditional Independent Events

$A \perp\!\!\!\perp B$ conditioned on C if

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

or equivalently,

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

Conditional Independent Random Variables:

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

or equivalently,

$$f_{X|Y,Z}(x|y,z) = f_{X|Z}(x|z)$$

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

Independent Increments: $\{X(t)\}$ has independent increments $\boxed{\text{if } \forall n}$, and for all $t_0 < t_1 < \dots < t_n$ the increments $X(t_1-t_0),\dots,X(t_n-t_{n-1})$ 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

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

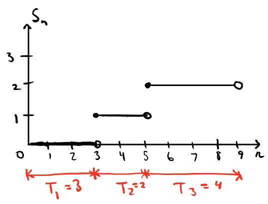
Binomial Counting Process

$X_i \sim$ 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 $\{T_i\}_{i=1}^\infty$

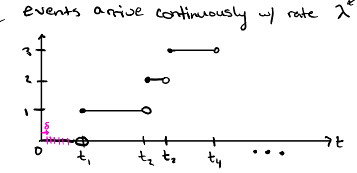
In above example,

$$T_1 \sim \text{Geometric}(p)$$

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

What about continuous counting processes?

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



Partition time into δ-length intervals.

Assumptions

For $\delta\to 0$

① The probability of having more than one arrivals in δ-interval is negligible (i.e. $\to 0$)

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

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

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

$\delta\to 0$

$p=\lambda\delta\to 0$

$n=\frac{t}{\delta}\to\infty$

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

Gaussian Random Processes

Definition:

$X(t)$ is Gaussian if $\forall n$, and $\forall \{t_i\}_{i=1}^n$ 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 $\mu_X(t)$ and autocorrelation function $R_X(t)$/ autocovariance function $C_X(t)$.

Example

"READ RIGHT VALUES!"

$$f_{X(t_1),X(t_2)}(x_1,x_2) \sim \mathcal{N}\left(\begin{bmatrix}\mu_X(t_1)\\ \mu_X(t_2)\end{bmatrix}, K=\begin{bmatrix}C_X(t_1) & C_X(t_1,t_2)\\ C_X(t_2,t_1) & C_X(t_2)\end{bmatrix}\right)$$

② WSS ⇒ S

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

① $X(0)=0$
② $\{X(t)\}_{t\ge 0}$ has independent increments
③ $X(t_2)-X(t_1) \sim \mathcal{N}(0,\sigma^2(t_2-t_1))$
④ Every sample path is continuous

Properties of Brownian Motion

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

② $R_X(t_1,t_2) = C_X(t_1,t_2) \quad t_2>t_1$
$= \mathbb{E}[X(t_1)X(t_2)]$
$= \mathbb{E}[X(t_1)(X(t_1)+X(t_2)-X(t_1))]$
$= \mathbb{E}[X^2(t_1)] + \mathbb{E}[X(t_1)]\mathbb{E}[X(t_2)-X(t_1)]$
just illustrates increments ⟶ $= \mathbb{E}[(X(t_1)-X(0))^2] + \mathbb{E}[X(t_1)-X(0)]\mathbb{E}[X(t_2)-X(t_1)]$
$= \sigma^2 t_1$

In general,

$C_X(t_1,t_2) = R_X(t_1,t_2) = \min(t_1,t_2)\sigma^2$

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

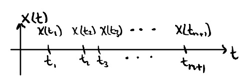
Also have $X(t_1), X(t_2),\dots,X(t_n)$ J.G $\forall n$.

i.e

$$\begin{bmatrix} X(t_1) \\ \vdots \\ X(t_n) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 \cdots & 0 \\ 1 & 1 & 1\; 0 \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}\begin{bmatrix} X(t_1)-X(0) \\ X(t_2)-X(t_1) \\ \vdots \\ X(t_n)-X(t_{n-1}) \end{bmatrix}$$

Markov Process

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



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

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 $\perp\!\!\!\perp$ past states

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

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

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

Jointly Gaussian

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

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

$Var(X)=\sigma^2$

X and Y are jointly Gaussian if

$Z=aX+bY$ is Gaussian $\forall a,b\ne 0\in\mathbb{R}$

Properties

① If JG, Uncorrelated ⇒ independence NOT generally true

Uncorrelated $= Cov(X,Y)=0$

independence ← 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 X,Y jointly Gaussian then $X|Y$ is Gaussian USEFUL

with mean $\mathbb{E}[X|Y] = \hat{X}_{LMMSE} = \mathbb{E}[X] + \frac{Cov(X,Y)}{Var(Y)}(Y-\mathbb{E}[Y])$

Variance MMSE $Var(X|Y) = Var(X) - \frac{Cov^2(X,Y)}{Var(Y)}$

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

in general, a rv w.r.t variable being conditioned on

For every realization of Y, have the same result

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

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{|K|}}e^{-\frac{1}{2}\left(\begin{pmatrix}x-\mu_X\\ y-\mu_Y\end{pmatrix}^T K^{-1}\begin{pmatrix}x-\mu_X\\ y-\mu_Y\end{pmatrix}\right)}$$

where

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

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

Discrete-Time Markov Chain

$\{X_n\}_{n\geq 0}$ is a random process

$k, \{n_i\}_{i=1}^{k}$

$X_{n_1}, \dots, X_{n_k} \to$ need joint pmf to describe it

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

$$= \Pr[X_0 = x_0]\Pr[X_1 = x_1 | X_0 = x_0]\Pr[X_2 = x_2 | X_1 = x_1, x_0 = 0] \cdots \Pr[X_k = x_k | X_{k-1} = x_{k-1}, \dots, X_0 = x_0]$$

$$= \Pr[X_0 = x_0]\underbrace{\Pr[X_1 = x_1 | X_0 = x_0]}_{\text{one-step transition probability}}\Pr[X_2 = x_2 | X_1 = x_1] \cdots \Pr[X_k = x_k | X_{k-1} = x_{k-1}]$$

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

i.e. Need $p_X(x_0)$ and $\{\Pr[X_{n+1} = j | X_n = i]\}_{i,j \in \mathcal{X}}$

Homogenous Markov Chain $\Rightarrow$ Time INVARIANT

$$\Pr[X_{n+1} = j | X_n = i] \triangleq p_{i,j} \quad \forall n$$

Transition Matrix:

$$\mathbb{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,M} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,M} \\ \vdots & \vdots & & \vdots \\ p_{M,1} & p_{M,2} & \cdots & p_{M,M} \end{bmatrix}$$

$\mathcal{X} = \{1, 2, \dots, M\}$

$0 \leq p_{i,j} \leq 1$

$\sum_j p_{i,j} = 1 \quad \forall i$

n-Step Transition Matrix

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

Chapman-Korogorov Equation

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

$i \to k \quad k \to j$

$\swarrow i \to j$

$$\mathbb{P}^{(m+n)} = \mathbb{P}^{(m)}\mathbb{P}^{(n)}$$

$$\mathbb{P}^{(n)} = \mathbb{P}\mathbb{P}^{(n-1)}$$
$$= \mathbb{P}\mathbb{P}\mathbb{P}^{(n-2)}$$
$$\vdots$$
$$= \mathbb{P}^n$$

State Probability

This is the distribution at time n.

$$p(n) = [\Pr[X_n = 1] \ \cdots \ \Pr[X_n = m]] \leftarrow \text{vector}$$

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

$$p(1) = p(0)\mathbb{P}$$
$$\vdots$$
$$p(n) = p(0)\mathbb{P}^n$$

$\{X_n\}_{n\geq 0}$: $\vec{p}(0) = \Pr[X_0 = 1, \dots, X_0 = m]$, $\left\{ \begin{array}{l} \text{pmf of } X_0 \\ \mathbb{P} \end{array} \right.$

row vector by definition: $\vec{p}(n) \triangleq \Pr[X_n = 1, \dots, X_n = m]$

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

Gaussian Random Vectors

$\mathbb{X}$ is a Gaussian random vector if its coordinates are JOINTLY GAUSSIAN

$\mathbb{X} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}$ a Gaussian random vector $\Leftrightarrow a_1 X_1 + \cdots + a_m X_m$ is Gaussian $\forall a_i \in \mathbb{R}, 1 \leq i \leq m$.

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

If $\mathbb{X} \sim \mathcal{N}(\mu, K)$, then

1. Any subvector of $\mathbb{X}$ is a Gaussian random vector
   - $X_i \sim \mathcal{N}(\mu_i, K_{i,i})$
   - $\begin{bmatrix} X_i \\ X_j \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_i \\ \mu_j \end{bmatrix}, \begin{bmatrix} K_{i,i} & K_{i,j} \\ K_{j,i} & K_{j,j} \end{bmatrix} \right)$
2. $\forall \mathbb{A}, \mathbb{b}$

$$\underset{n\times 1}{\mathbb{Y}} = \underset{n\times m}{\mathbb{A}}\underset{m\times 1}{\mathbb{X}} + \underset{n\times 1}{\mathbb{b}}$$

is a Gaussian random vector

$$\mathbb{Y} \sim \mathcal{N}(\mathbb{A}\mu + \mathbb{b}, \underbrace{\mathbb{A}K\mathbb{A}^T}_{\mathbb{E}[(\mathbb{Y} - \mathbb{E}[\mathbb{Y}])(\mathbb{Y} - \mathbb{E}[\mathbb{Y}])^T]})$$

3. If K is diagonal, every rv in $\mathbb{X}$ is independent of every other rv in $\mathbb{X}$
4. If $\mathbb{X}_{m\times 1}$ and $\mathbb{Y}_{n\times 1}$ are jointly Gaussian then they are independent iff $Cov(\mathbb{X}, \mathbb{Y}) = 0$

MMSE Estimate of $\mathbb{X}$ using $\mathbb{Y}$

$$\text{MSE} \quad \mathbb{E}[\|\mathbb{X} - \hat{\mathbb{X}}\|^2] = \mathbb{E}[\underset{1\times m}{(\mathbb{X} - \hat{\mathbb{X}})^T}\underset{m\times 1}{(\mathbb{X} - \hat{\mathbb{X}})}]$$

$$= \sum_{i=1}^{m} \mathbb{E}[X_i - \hat{X}_i]$$

Need to "design"

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

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

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

Our best MMSE estimator for X_i is

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

Need conditional distribution

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

LMMSE Estimator of $\mathbb{X}$

Similar to rv case,

$$\underset{m\times 1}{\hat{\mathbb{X}}} = \underset{m\times n}{\mathbb{A}}\underset{n\times 1}{\mathbb{Y}} + \underset{m\times 1}{\mathbb{b}}$$
$$= \underset{m\times 1}{\mathbb{E}[\mathbb{X}]} + \underset{m\times n}{Cov(\mathbb{X}, \mathbb{Y})}\underset{n\times n}{Cov^{-1}(\mathbb{Y})}\underset{n\times 1}{(\mathbb{Y} - \mathbb{E}[\mathbb{Y}])}$$

If $\mathbb{X}, \mathbb{Y}$ are jointly Gaussian, (similar to rv case)

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

Poisson Process:

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

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

Look at T_1

$$\Pr[T_1 > t] = (1 - \lambda\delta)^{\frac{t}{\delta}}$$

$$\lim_{\delta \to 0} (1 - \lambda\delta)^{t/\delta} = e^{-\lambda t}$$

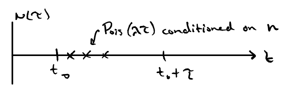
i.e $T_1 \sim \exp(\lambda)$ $\qquad \mathbb{E}[T_1] = \frac{1}{\lambda}$

Leads to ANOTHER definition

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

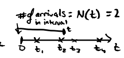
Definition: $\{N(t)\}_{t\geq 0}$ is a counting process w/ rate λ if it is a counting process such that $\forall \tau > 0$

any interval τ: $N(\tau) \sim \text{Pois}(\lambda\tau)$, and given $N(\tau) = n$ (n arrival times in interval τ), are iid $U \sim [0, \tau]$



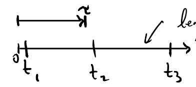
Facts about Poisson Process

$\{N(t)\}_{t\geq 0}$ w/ rate λ. $\lambda \to$ arrival rate/unit time



1. Process has independent increments

$$N(t_2) - N(t_1) \perp\!\!\!\perp N(t_3) - N(t_2)$$



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

2. # arrivals in interval $[t_1, t_2] = \mathcal{N}(t_2) - \mathcal{N}(t_1)$

where

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

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

where

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

4. Given $\mathcal{N}(\tau) = k$, these k arrivals are iid $\mathcal{U}[0, \tau)$

i.e $t_i \sim \mathcal{U}[0, k]$

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

$Var(N(t)) = \lambda t$ (Poisson property)

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

$$= \mathbb{E}[N(s)(N(s) + N(t) - N(s))]$$
$$= \mathbb{E}[(N(s))^2] + \mathbb{E}[N(s)(N(t) - N(s))] \quad \text{independ. increments}$$
$$= \lambda s + (\lambda s)^2 + \mathbb{E}[N(s)]\mathbb{E}[N(t) - N(s)] \quad \to \mathbb{E}[N(t)] - \mathbb{E}[N(s)]$$
$$= \lambda s + (\lambda s)^2 + \lambda s(\lambda(t - s))$$

$C_N(s,t) = \lambda \min(s,t) \leftarrow$ applies $\underline{\forall s, t}$ (note above we assumed $s < t$)

$$\hat{X}_{MMSE} = \mathbb{E}[X|Y]$$
$$MSE(\hat{X}_{MMSE}) = \mathbb{E}[Var(X|Y)]$$

Cauchy-Schwartz Inequality

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

Correlation Coefficient

normalization comes from Cauchy-Schwarz inequality

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

If $X_1, \dots, X_m$ are $\underline{\text{pairwise uncorrelated}}$, then

$$Var(X_1 + X_2 + \cdots + X_m) = Var(X_1) + Var(X_2) + \cdots + Var(X_m)$$

$$Var\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} Var(X_i)$$

Conditional Expectation

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

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

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

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