Dr. Osonde Osoba

This week:

- Total Expectation
  - Conditional Expectations
  - Total Probability as Total Expectation : I_A
  - Substitution Law
- Probabilistic Inequalities
  - Markov and Chebyshev Inequalities
  - Jensen's Inequality : Convexity
  - Cauchy-Schwarz Inequality : σ_{xy} vs ρ_{xy}
    - Jointly Gaussian random variables

H/W

L-G : 7.15 – 7.17, 5.74, 5.75, 5.80, 6.50(a, b)

I// Conditional Expectation and Total Expectation:

Given two jointly distributed r.vs, (X,Y), the conditional pdf:

$$\boxed{f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}}$$

$f_{Y|X}$ is a typical pdf except that it expresses probabilities in terms of rv X or samples of $X=x$.

$$\boxed{P(Y\in A|X) = \int_A f_{Y|X}(y|x)\,dy = h(X)}$$

The conditional pdf gives a conditional expectation $\forall$ measurable function $g(\cdot)$ which are integrable wrt $f_{Y|X}$:

$$\boxed{E_{Y|X}[g(Y)|X] = \int g(y) f_{Y|X}(y|x)\,dy}$$

Think of $E[g(Y)|X]$ as the "optimal" estimate for $g(Y)$ given all information in X. $r(X) = E[g(Y)|X]$ is in fact the optimal $\mathcal{L}^2$-estimate for $g(Y)$ given observations of X.

$E_{Y|X}[\cdot]$ inherits all the properties of a typical integral:

i/ <u>Linearity / Homogeneity</u>:

$$E[a\cdot h(Y) + b\cdot g(Z) | X] = a\cdot E[h(Y)|X] + b\cdot E[g(Z)|X]$$

Conditioning also leads to some unique properties e.g.

ii/ [$E[\cdot|X]$ treats functions of X as constants]:

$$\boxed{E[h(X)\cdot g(Y)|X] = h(X)\cdot E[g(Y)|X]}$$

<u>Proof</u>: $\forall$

$$E[h(X)\cdot g(Y)|X=x] = \int h(x)\, g(y)\, f_{Y|X}(y|x)\, dy \quad \forall x$$
$$= h(x) \int g(y)\, f_{Y|X}(y|x)\, dy \quad \forall x$$
$$= h(x)\cdot E[g(Y)|X=x] \quad \forall x$$

$$\therefore\ E[h(X)\cdot g(Y)|X] = h(X)\cdot E[g(Y)|X] \ //$$

iii/ [Conditional Probabilities are Conditional Expectations]:

$$\boxed{\begin{aligned} P(Y\in A|X) &= \int I_A(y)\cdot f_{Y|X}(y|x)\, dy \\ &= E_{Y|X}[I_A(Y)|X] \end{aligned}}$$

It turns out that $E[\cdot|X]$ is the rigorous foundation for $P(Y\in A|X)$; $E[\cdot|X]$ is more fundamental than $P(\cdot|X)$.

$r(X) = E[g(X,Y)|X]$ is a function of X and thus a rv also. $r(x) = E[g(X,Y)|X=x]$ is a realization of $r(X)$. The substitution law links the two.

iv/ Substitution Law:

$$\boxed{E[g(X,Y)|X=x] = E[g(x,Y)|X=x]}$$

Proof:

recall $I_A(X) + I_{A^c}(X) = 1$; Let $A = \{x\}$

$$\Rightarrow g(X,Y) = g(X,Y)\left[I_{\{x\}}(X) + I_{\{x\}^c}(X)\right]$$

$$= g(X,Y)\cdot I_{\{x\}}(X) + g(X,Y)\cdot I_{\{x\}^c}(X)$$

$$\Rightarrow E[g(X,Y)|X] = E[\overbrace{g(x,Y)\cdot I_{\{x\}}(X)}^{=0\ \forall X\neq x}|X] + E[g(X,Y)\cdot I_{\{x\}^c}(X)|X]$$

by prop. ii/ above: $E[g(X,Y)I_A(X)|X] = I_A(X)\cdot E[g(X,Y)|X]$

$$\Rightarrow E[g(X,Y)|X] = I_{\{x\}}(X)\cdot E[g(x,Y)|X] + I_{\{x\}^c}(X)\cdot E[g(X,Y)|X]$$

$$\Rightarrow E[g(X,Y)|X=x] = \overbrace{I_{\{x\}}(x)}^{=1 \text{ since } X=x} E[g(x,Y)|X=x] + \overbrace{I_{\{x\}^c}(x)}^{=0 \text{ since } X=x}\cdot E[g(X,Y|X=x]$$

$$\therefore\quad E[g(X,Y)|X=x] = E[g(x,Y)|X=x]$$

iv. Total Expectation: $[T_E]$ (*)

$$\boxed{E_Y[g(Y)] = E_X\left[E_{Y|X}[g(Y)|X]\right]}$$

[aka. "Law of iterated Expectations"]

Proof:

$$\begin{aligned}
E_X E_{Y|X}[g(Y)|X] &= \int E_{Y|X}[g(Y)|X]\, f_X(x)\, dx \\
&= \iint g(y)\cdot f_{Y|X}(y|x)\cdot f_X(x)\, dx\, dy \\
&= \iint g(y)\cdot f_{XY}(x,y)\, dx\, dy \\
&= \int g(y)\left[\int f_{XY}(x,y)\, dx\right] dy \\
&= \int g(y)\, f_Y(y)\, dy
\end{aligned}$$

$$E_X E_{Y|X}[g(Y)|X] = E_Y[g(Y)]$$

[We used the facts that $f_{Y|X}\cdot f_X = f_{XY}$ and $\int f_{XY}\, dx = f_Y$]

Similarly:

$$E_{XY}[g(X,Y)] = E_X\left[E_{Y|X}[g(X,Y)]\right]$$

$$E_{XYZ}[g(X,Y,Z)] = E_{YZ}\left[E_{X|Y,Z}[g(X,Y,Z)\,|\,Y,Z]\right]$$

etc...

Total Probability is a special case of T_E when $g(\cdot)$ is an indicator function.

$$\boxed{P(Y\in A) = E_Y[I_A(Y)] = E_X\left[E_{Y|X}[I_A(Y)|X]\right]}$$

We can apply the "SIT" technique ~~for~~ to complicated expectations:

SIT for T_E

(1) T_E : Apply **T**otal **E**xpectation

(2) S : **S**ubstitute sample x for r.v X when $X=x$

(3) I : Apply **I**ndependence (if applicable!!!)

Extra practice in Gubner : 7.34 – 7.37

e.g.: $X \sim \exp(1)$, $Y|_{X=x} \sim N(0, x^2)$ | $E[Y^2 \cdot X^5] = ???$

$$E[Y^2X^5] \overset{T_E}{=} E_X\, E_{Y|X}[Y^2 \cdot X^5 \mid X]$$

$$= E_X[X^5 \cdot E_{Y|X}[Y^2|X]]$$

$$\overset{S}{=} \int_0^\infty x^5\, E_{Y|X}[Y^2|X=x]\, dx$$

$$E[Y^2|X=x] = \sigma^2_{Y|X} = x^2$$

$$\Rightarrow E[Y^2X^5] = \int_0^\infty x^5 \cdot x^2 \cdot f_X(x)\,dx = \int_0^\infty x^7 \cdot \exp(-x)\,dx$$

$$\therefore E[Y^2 \cdot X^5] = \Gamma(8) = 7!$$

ex: X, Z, U : independent

$X, Z \sim \exp(1)$

$U \sim U[-1/2, 1/2]$

$E[\exp(U(X+Z))] = ?$

Ans :

$$E[\exp(U(X+Z))] \overset{\text{TE}}{=} E_U\, E_{X,Z|U}[\cdots|U]$$

$$\overset{\text{S}}{=} \int_u f_U(u)\, E_{X,Z|U}[\exp(u(X+Z))\,|\,U=u]\, du$$

$$\overset{\text{Ind}}{=} \int_u f_U(u)\, E_{X,Z}[\exp(uX)\cdot\exp(uZ)]\, du$$

$$\overset{X \perp\!\!\!\perp Z}{=} \int_u f_U(u)\, E_X[\exp(uX)]\cdot E_Z(uZ)]\, du$$

but $[X \overset{d}{=} Z] \Rightarrow E_X[\exp(uX)] = E_Z[\exp(uZ)]$

$$\Rightarrow E[\exp(U(X+Z))] = \int_u (E_X[\exp(uX)])^2 \cdot f_U(u)\, du$$

$$= \int_{-1/2}^{1/2} \left(\int_0^{\infty} \exp(ux)\cdot\exp(-x)\, dx\right)^2 \cdot 1\, du$$

$$= \int_{-1/2}^{1/2} \left(\int_0^{\infty} \exp(x(u-1))\, dx\right)^2 du$$

$$= \int_{-1/2}^{1/2} \left(\frac{-1}{1-u}\right)^2 du$$

$$= \int_{1/2}^{3/2} x^{-2}\, dx \qquad \begin{bmatrix} x = 1-u \\ dx = -du \end{bmatrix}$$

$$= \left[-x^{-1}\right]_{0.5}^{1.5}$$

$$= -2/3 + 2$$

$$\therefore\ E[\exp(U(X+Z))] = 4/3$$

Ex: [Doubly Random Sum]:

Suppose $\{X_i\}_{i=1}^{\infty}$: iid with $\sigma_X^2 < \infty$

N : rv independent of all $\{X_i\}_{i\geq 1}$; $N \geq 1$

define $S_N = \sum_{i=1}^{N} X_i$

$E[S_N] = ???$

$$E[S_N] \overset{I_E}{=} E_N E_{S_N|N}[S_N | N]$$
$$\overset{S}{=} E_N E_{S_N|N}\left[\sum_{i=1}^{n} X_i \,\middle|\, N=n\right]$$
$$\overset{Ind}{=} E_N\left[E_{\vec{X}}\left[\sum_{i=1}^{n} X_i\right]\right]$$

$$= \sum_n P_N(n)\, E\left[\sum_{i=1}^{n} X_i\right]$$
$$\overset{iid}{=} \sum_n P_N(n) \cdot n \cdot \mu_X$$
$$= \mu_X \cdot \left(\sum_n n \cdot P_N(n)\right)$$
$$\boxed{E[S_N] = \mu_X \cdot \mu_N}$$

[Wald's Identity]

$Var[S_N] = ???$

The conditional variance:

$$\boxed{V[Y|X] = E\left[(Y - E[Y|X])^2 \,\middle|\, X\right]}$$

Claim: $V[Y|X] = E[Y^2|X] - E^2[Y|X]$

Proof:

$$
\begin{aligned}
V[Y|X] &= E\left[(Y - E[Y|X])^2 \,|\, X\right] \\
&= E[Y^2|X] + E^2[Y|X] - 2E\left[Y \cdot \overbrace{E[Y|X]}^{r(X)} \,\middle|\, X\right] \\
&= E[Y^2|X] + E^2[Y|X] - 2E[Y|X] \cdot E[Y|X] \\
&= E[Y^2|X] + E^2[Y|X] - 2E^2[Y|X]
\end{aligned}
$$

$$V[Y|X] = E[Y^2|X] - E^2[Y|X]$$

Total Variance $[T_V]$:

$$\boxed{V[Y] = \underbrace{E_X[V[Y|X]]}_{(a)} + \underbrace{V_X[E[Y|X]]}_{(b)}}$$

Proof:

(a)
$$
\begin{aligned}
E_X[V[Y|X]] &= E_X\left[E[Y^2|X] - E^2[Y|X]\right] \\
&= E_X E_{Y|X}[Y^2|X] - E_X\left[(E[Y|X])^2\right] \\
&\overset{T_E}{=} E_Y[Y^2] - E_X\left[(E[Y|X])^2\right]
\end{aligned}
$$

(b)
$$
\begin{aligned}
V_X[E[Y|X]] &= E_X\left[(E[Y|X])^2\right] - \left(E_X[E_{Y|X}[Y|X]]\right)^2 \\
&\overset{T_E}{=} E_X\left[(E[Y|X])^2\right] - \left(E_Y[Y]\right)^2
\end{aligned}
$$

$$
\begin{aligned}
\therefore V[Y] &\overset{(a)+(b)}{=} E[Y^2] - E_X\left[(E[Y|X])^2\right] + E_X\left[(E[Y|X])^2\right] - E_Y^2[Y] \\
&= E[Y^2] - \mu_Y^2
\end{aligned}
$$

So $Var[S_N] = E_N[Var[S_N|N]] + V_N[E[S_N|N]]$

$$E_N[Var[S_N|N]] \stackrel{S}{=} \sum_n P_N(n) \cdot Var[S_N|N=n]$$
$$\stackrel{Ind}{=} \sum_n P_N(n) \; Var[\sum_i^n X_i]$$
$$\stackrel{iid}{=} \sigma_X^2 \sum_n n P_N(n) = \sigma_X^2 \cdot \mu_N$$

$$V_N[E[S_N|N]] = E_N[(E[S_N|N])^2] - E_N^2[E[S_N|N]]$$
$$\stackrel{S, Ind}{=} \sum_n ((E[\sum^n X_i])^2 P_N(n)) - \left(\sum_n E[\sum^n X_i] P_N(n)\right)^2$$
$$= \sum_n n^2 \cdot \mu_X^2 \cdot P_N(n) - \left(\mu_X \cdot \sum_n n P_N(n)\right)^2$$

$$= \mu_X^2 \cdot \left[\left(\sum_n n^2 \cdot P_N(n)\right) - \left(\sum_n n \cdot P_N(n)\right)^2\right]$$

$$V_N[E[S_N|N]] = \mu_X^2 \cdot \sigma_N^2$$

$$\therefore \boxed{Var[S_N] = \sigma_X^2 \mu_N + \mu_X^2 \sigma_X^2}$$

More rigorously, if X is a positive r.v.

$$X : (\Omega, \mathcal{Q}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

from a probability space $(\Omega, \mathcal{Q}, P)$, conditional expectation $E[X|\mathcal{G}]$ are Radon-Nikodym derivatives of $\int_A X \, dP$ wrt P for event $A \in \mathcal{G} \subseteq \mathcal{Q}$. The familiar $E[X|Y]$ is really

sub-σ-algebra

$$E[X|Y] = E[X|\sigma(Y)]$$

i.e $\mathcal{G}$ is the sub-σ-algebra

$$\mathcal{G} = \sigma(Y) = \sigma(\{Y^{-1}(B) \mid B \in \mathcal{B}(\mathbb{R})\})$$

Thus $E[X|\mathcal{G}]$ is the theoretical foundation for $E[X|Y]$ and $P(A|B)$

Martingales

$\{S_n\}_{n\geq 1}$ is a martingale iff:

$$\boxed{E[S_{n+1} \mid S_1, \cdots, S_n] = S_n \quad \forall n}$$

[e.g. Paul Samuelson's work linking stock prices to martingales]

$\{S_n\}_{n\geq 1}$ is a martingale w.r.t. $\{X_n\}$ iff

$E[S_{n+1} \mid X_1, \ldots, X_n] = S_n$

$\{S_n\}_{n\geq 1}$ submartingale iff $E[S_{n+1} \mid S_1, \cdots, S_n] \geq S_n \quad \forall n$

$\{S_n\}_{n\geq 1}$ supermartingale iff $E[S_{n+1} \mid S_1, \cdots, S_n] \leq S_n \quad \forall n$

e.g. $\{X_i\}_{i\geq 1}$: iid with $E[X] = 0$

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

$$\Rightarrow E[S_{n+1} \mid S_1, \cdots, S_n] = E[X_{n+1} + S_n \mid \{S_i\}_{i=1}^{n}]$$

$$= S_n + E[X_{n+1} \mid \{S_i\}_{i=1}^{n}]$$

$\{X_i\}_{i=1}^{n+1}$ indep. $\Rightarrow \{X_{n+1}, g(X_1, \cdots, X_n)\}$ indep. $\forall g(\cdot)$

e.g. $g(X_1, \cdots, X_n) = \sum_{i=1}^{n} X_i \Rightarrow$

Thus X_{n+1} is indep. of all $\{S_i\}_{i=1}^{n}$

$$\Rightarrow E[X_{n+1} \mid \{S_i\}_{i=1}^{n}] = E[X_{n+1}] = 0 \quad \forall n$$

$$\therefore E[S_{n+1} \mid S_1, \cdots, S_n] = S_n$$

"Fair random walks are Martingales."

II/ Probabilistic Inequalities:

Goal: approximate $P(X \in A_\varepsilon)$ where

$A_\varepsilon = \{ |X - \mu_X| > \varepsilon \}$ [Tail probabilities]

Ideally we can do approximations using limited distribution info e.g μ_X and σ_X^2 only

(a) Markov Inequality: (><)

Suppose $X \geqslant 0$ and $E[X] < \infty$

then $\forall\ \varepsilon > 0$

$$\boxed{P(X > \varepsilon) < \frac{E[X]}{\varepsilon}} \qquad (M.I)$$

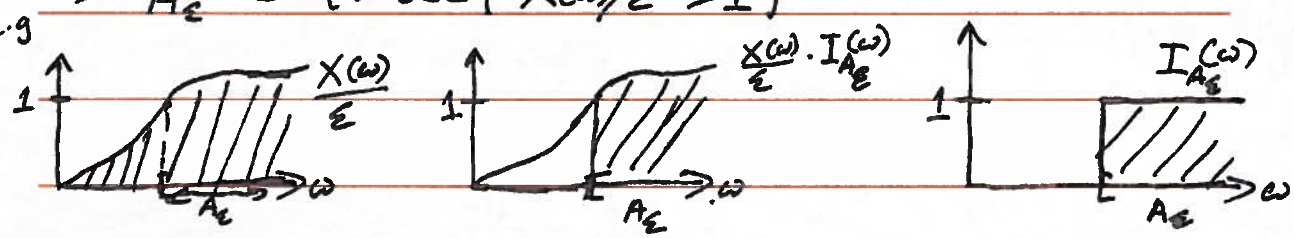
if $X \not\geqslant 0$ then

$$P(|X| > \varepsilon) < 1/\varepsilon \cdot E[|X|]$$

Proof: Define $A_\varepsilon = \{ \omega \in \Omega \mid X(\omega) > \varepsilon \}$

$\Rightarrow A_\varepsilon = \{ \omega \in \Omega \mid X(\omega)/\varepsilon > 1 \}$

e.g



i.e $X(\omega)/\varepsilon \ > \ X(\omega)/\varepsilon \cdot I_{A_\varepsilon}(\omega) \ > \ I_{A_\varepsilon}(\omega)$

$E[\cdot] \Rightarrow$

$$E[I_{A_\varepsilon}(\omega)] < E[X(\omega)/\varepsilon]$$

$$P(A_\varepsilon) < (1/\varepsilon) \cdot E[X]$$

$$\boxed{P(X > \varepsilon) < \frac{E[X]}{\varepsilon}} \quad \text{(Markov Ineq.)}$$

e.g. $X \sim \gamma(\alpha, \theta)$

$\Rightarrow P(X > 8) < 1/8 \cdot E[X] = \alpha\theta/8$

e.g. A DMV employee spends an avegge of 5 mins per customer. There are 11 people ahead of me. Find an upper bound for the probability that I waited > 1 hour.

Ans X_i : service time

$X_i \sim \exp(5)$ $[\theta = 5 \leftarrow$ avg. service time$]$

$Y = \sum_{i=1}^{11} X_i \Rightarrow Y \sim \gamma(11, 5)$

$$P(Y > 60) < \frac{E[Y]}{60} = 55/60 = 11/12$$

(b) Chebyshev's Inequality:

Let $A_\varepsilon = \{|X-\mu| > \varepsilon\}$ instead

Then $Y = |X-\mu|$ is a positive r.v and

$|E[Y]| < \infty$ if $\sigma_x^2 < \infty$.

Also $A_\varepsilon \equiv \{(X-\mu)^2 > \varepsilon^2\}$

$\Rightarrow$ By Markov inequality:

$$P(A_\varepsilon) < \frac{E[(X-\mu)^2]}{\varepsilon^2}$$

i.e $\boxed{P(|X-\mu|>\varepsilon) < \sigma_x^2/\varepsilon}$

e.g $\bar{X}_n = 1/n \sum_{i=1}^{n} X_i$ $\{X_i\}$ iid

$\Rightarrow E[\bar{X}_n] = \mu_x$ and $Var[\bar{X}_n] = \sigma_x^2/n$

$\therefore$ $P(|\bar{X}_n - \mu_x| > \varepsilon) < Var(\bar{X}_n)/\varepsilon^2 = \dfrac{\sigma_x^2}{n\varepsilon^2}$

$\Rightarrow \lim_{n\to\infty} P(|\bar{X}_n - \mu_x| > \varepsilon) < \lim_{n\to\infty} \left(\frac{\sigma_x^2}{n\varepsilon^2}\right) = 0$ [WLLN]

$\uparrow$ next week

Generalized Chebyshev-Markov Inequality:

$\forall\, \varepsilon > 0$

$$\boxed{P(|X-\alpha\mu| > \varepsilon) < \frac{E[|X-\alpha\mu|^n]}{\varepsilon^n}}$$

M.I $\Rightarrow (n=1, \alpha=0)$ CI $\Rightarrow (n=2, \alpha=1)$

(c) **Jensen's Inequality:**

$\phi(x)$ is a convex function (e.g $x^2, e^x, -\ln x, x^{2k}$)

$\iff \phi(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\phi(x_1) + (1-\lambda)\phi(x_2)$ $\left[\begin{matrix}\lambda \in [0,1] \\ \forall x_1, x_2\end{matrix}\right]$

$\iff \phi''(x) \geq 0 \quad \forall x$

$\iff$ epigraph(ϕ) is a convex set

$\iff$ if $\phi : \mathbb{R}^n \to \mathbb{R}$

the Hessian $\underline{H}_\phi = [[\partial_{x_i}\partial_{x_j}\phi(x)]]_{(i,j)}$ is a

positive semidefinite matrix

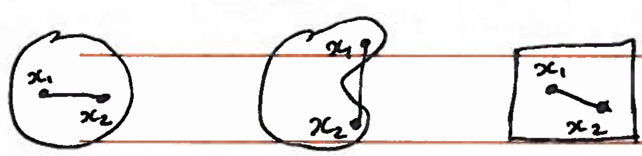
↖ ???

- [ϕ convex] $\iff$ [$-\phi$ concave]; [ϕ linear] $\iff$ [ϕ convex + concave]

- ϕ convex $\Rightarrow$ ϕ has a global minimum (**)
- The epigraph defn of convexity links functional convexity with set convexity.

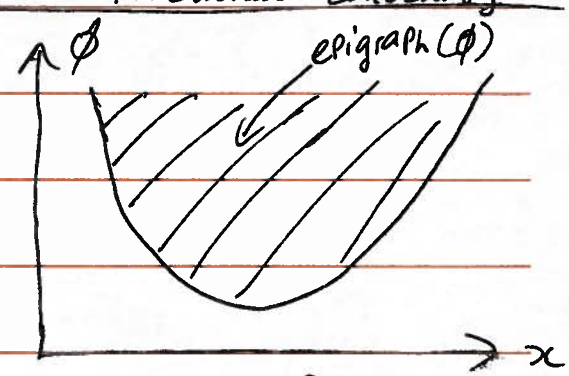
Set Convexity

Define $L(x_1, x_2)$ as the line between x_1, x_2. [S convex] $\iff$ [$L(x_1,x_2) \in S \;\; \forall x_1, x_2 \in S$] e.g



convex   not convex   convex

functional convexity



[ϕ convex] $\iff$ [epigraph(ϕ) is a convex set]

Jensen's inequality:

ϕ convex and X a random variable

$$\Rightarrow \quad \boxed{\phi\left(E[X]\right) \leq E[\phi(X)]} \quad \text{(Jensen's)}$$

e.g.: $\phi(x) = x^2$

$\Rightarrow \quad \phi(E[X]) = \mu_X^2 \quad ; \quad E[\phi(X)] = E[X^2]$

$\Rightarrow \mu_X^2 \leq E[X^2]$ which we already know

since $E[X^2] = \sigma_X^2 + \mu_X^2$ and $\sigma_X^2 \geq 0$

---

e.g. Geometric vs Arithmetic mean:

$\vec{x} \triangleq \{x_i\}_{i=1}^n$ (assume x_i distinct and >0)

$GM(\vec{x}) = \left(\prod_{i=1}^n x_i\right)^{1/n}$

$AM(\vec{x}) = 1/n \sum_{i=1}^n x_i \quad \leftarrow$ [a sample of the sample mean $\overline{X}_n$]

Claim: $AM(\vec{x}) \geq GM(\vec{x})$

Proof: [use Jensen's inequality for $\phi(x) = -\ln x$]

Define the PMF for X: $P_X(x) = \begin{cases} 1/n & \forall x \in \{x_i\}_{i=1}^n \\ 0 & \text{else} \end{cases}$

(verify)

$$\Rightarrow \quad E_X[X] = \sum_{i=1}^{n} x_i \cdot (1/n) = AM(\vec{x})$$

$\therefore$ By Jensen's :

$$\phi(AM(\vec{x})) \leq E_X[\phi(X)]$$

$$\Rightarrow \quad -\ln(AM(\vec{x})) \leq \sum_{i=1}^{n} \frac{1}{n}(-\ln x_i)$$

$$[x \in D] \Rightarrow \quad \ln(AM(\vec{x})) \geq 1/n \sum_{i}^{n} \ln x_i$$

$$[\exp(\cdot)] \Rightarrow \quad AM(\vec{x}) \geq \exp\left[\left(\sum \ln x_i\right)/n\right]$$

$$\exp\left[\frac{1}{n}\sum \ln(x_i)\right] = \left(\prod_i^n \exp(\ln(x_i))\right)^{1/n} = \left(\prod_i^n x_i\right)^{1/n} = GM(\vec{x})$$

$$\therefore \quad AM(\vec{x}) \geq GM(\vec{x})$$

in general : $HM \leq GM \leq AM$ where $HM = n\left(\sum_{i=1}^{n} 1/x_i\right)^{-1}$

Pythagorean means : $\{AM, GM, HM\}$.

(d) <u>Cauchy-Schwarz' Inequality</u> :

Recall :

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X \mu_Y$$

<u>Claim</u> :

$$\boxed{\sigma_{XY}^2 \leq \sigma_X^2 \sigma_Y^2} \qquad \text{(Cauchy-Schwarz)}$$

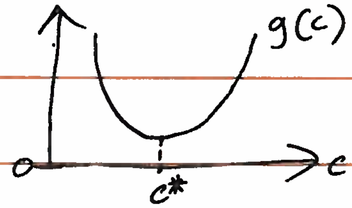
<u>Proof</u> : $\forall c \in \mathbb{R}$

$$g(c) = E\left[\left((X-\mu_X) - c(Y-\mu_Y)\right)^2\right] \geq 0$$

$$\Rightarrow \quad g(c) = E[(X-\mu_X)^2] + c^2 E[(Y-\mu_Y)^2] - 2cE[(X-\mu_X)(Y-\mu_Y)]$$

$$\Rightarrow \quad g(c) = \sigma_X^2 + c^2\sigma_Y^2 - 2c\,\sigma_{XY} \geq 0$$

i.e



find $c^* = \arg\min g(c)$

i.e $g'(c^*) = 0$

$\Rightarrow \boxed{c^* = \sigma_{xy}/\sigma_y^2}$

$g(c^*) \geq 0$ since $g(c) \geq 0 \quad \forall c$

$g(c^*) = \sigma_x^2 + \sigma_{xy}^2 \cdot \sigma_y^2/\sigma_y^4 - 2\sigma_{xy}^2/\sigma_y^2 \geq 0$

$\Rightarrow \quad \sigma_x^2 + \sigma_{xy}^2/\sigma_y^2 - 2\sigma_{xy}^2/\sigma_y^2 \geq 0$

$\Rightarrow \quad \sigma_x^2 - \sigma_{xy}^2/\sigma_y^2 \geq 0$

$\Rightarrow \quad \sigma_x^2 \geq \sigma_{xy}^2/\sigma_y^2$

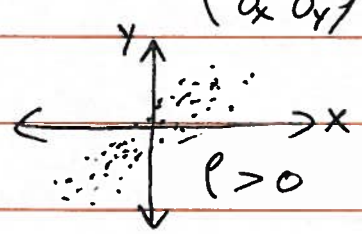
$\therefore \boxed{\sigma_{xy}^2 \leq \sigma_x^2 \sigma_y^2}$ //

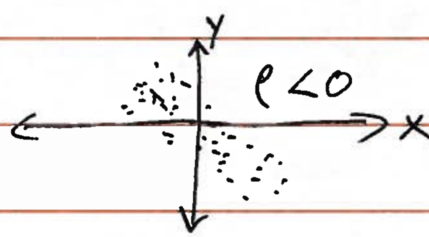
Thus if we define the population correlation coefficient

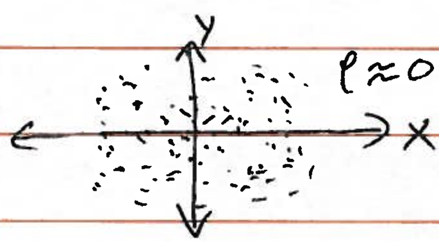
$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

C-S $\Rightarrow \left(\dfrac{\sigma_{xy}}{\sigma_x \sigma_y}\right)^2 = \rho_{xy}^2 \leq 1$ i.e $|\rho_{xy}| \leq 1$

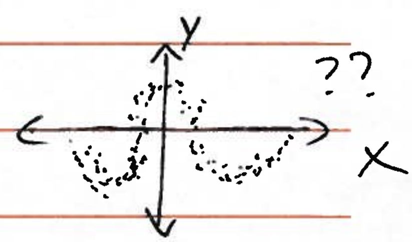
e.g.:







What about:



ρ_{XY} describes the dependence between X and Y

(X, Y) indep $\overset{\not\Leftarrow}{\Longrightarrow}$ $\rho_{XY} = 0$

In the case of <u>jointly-varying Gaussian</u> r.vs. (X,Y)

$\rho_{XY} = 0$ does imply that X and Y are independent.

This is because

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}}} \exp\left[\frac{-\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \left(\frac{y-\mu_y}{\sigma_y}\right)^2 + 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)}{2\left(1-\rho_{xy}^2\right)}\right]$$

factors into $f_{XY} = f_X \cdot f_Y$ when $\rho_{XY} = 0$.

<u>Generalizations of Cauchy-Schwarz</u>

The C-S inequality in probability is an instance of the more general C-S inequality of Hilbert spaces. A Hilbert space is just a "complete" space that has an inner product $\langle \cdot, \cdot \rangle$ defined. In probability

$$\langle X, Y \rangle = \int X(\omega) \cdot Y(\omega)\, dP = E[XY]$$

So we are working on the Hilbert space of square-integrable $(\sigma_x^2, \sigma_y^2 < \infty)$ functions with the above inner product. This (inner product) space of functions is called $\mathcal{L}^2$.

$\mathcal{L}^2$-space is the foundation of many areas of science and engineering. Each application $\mathcal{L}^2$ typically comes with its own Cauchy-Schwarz inequality.

e.g (a) [Signal-Processing]

for finite power signals i.e $\int |x(t)|^2 dt < \infty$

we also get finite power Fourier transforms $\int |X(\omega)|^2 d\omega < \infty$

C-S : $\mathrm{Var}[\tilde{x}(t)] \cdot \mathrm{Var}[\tilde{X}(\omega)] \geq 1/16\pi^2$ ← artifact of $\mathcal{F}\{\cdot\}$

where $\tilde{x}$ and $\tilde{X}$ are power normalized signals

(b) [Quantum-Mechanics]

Wave functions, ψ, are normalized complex $\mathcal{L}^2$-function, i.e.

$$\langle \psi | \psi \rangle \triangleq \int \psi^* \psi = 1$$

and for any "operator" $\hat{A}$, $\langle \psi | \hat{A} | \psi \rangle = \int \psi^* \hat{A} \psi \triangleq \langle \hat{A} \rangle$

Then for any 2 operators, $\hat{A}, \hat{B}$:

$$\sigma_A \sigma_B \geq \tfrac{1}{2} \left| \langle [\hat{A}, \hat{B}] \rangle \right| \quad \text{by Cauchy-Schwarz}$$

where $\sigma_A^2 = \langle \hat{A}^2 \rangle - (\langle A \rangle)^2$

and $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}A$

For $\hat{A} = \hat{x}$ (position operator), $\hat{B} = \hat{p}$ (momentum operator)

we get Heisenberg's Uncertainty Principle $\boxed{\sigma_x \sigma_p \geq \hbar/2}$

Other Probabilistic Inequalities :

i/ Hölder's Inequality

suppose $p > 1$ and $q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$

$$\Rightarrow \quad E[|X \cdot Y|] \leq E^{1/p}[|X|^p] \cdot E^{1/q}[|Y|^q]$$

$CS \Rightarrow$ Hölder w/ $p = 2 = q$

ii/ Lyapunov's Inequality

$$E^{1/\alpha}[|X|^\alpha] \leq E^{1/\beta}[|X|^\beta] \quad \text{for } 1 \leq \alpha < \beta < \infty$$

iii/ Minkowski's Inequality : $1 \leq p < \infty$

$$E_{XY}^{1/p}[|X+Y|^p] \leq E_X^{1/p}[|X|^p] + E_Y^{1/p}[|Y|^p]$$

$$\|X\|_p \triangleq \left[\int |X|^p \, dP\right]^{1/p}$$