Aug 23, 2023

Prof Dennis

First Econometrics Class

1. Use Teams
2. This semester Topics:-
   - Probability
   - Real Analysis / Measure Theory
     - ↳ Used across Micro/Macro/Trix.
   - Estimators (extreme value theories)

Model as a lens to see the data

   - Testing

⇒ We are not required to memorize anything. If you are memorizing anything, talk to Dennis → [Open book tests.]

⇒ Problem Sets → extension of classes.

= Programming → Jupyter // Python (Ask ChatGPT)

Probability Theory.

- Origin → $11^{th}/12^{th}$ century. → which is when people were thinking of gambling.
- Formalized in 1950s ⟶ relatively new // to shake off the gambling stuff.

  Calculus comes from 1700s.
  Geometry is 2^{nd} century AD

The phenomenon is abstract. → Can be anything.

- x → single element
- A → a set is a collection of elements.

  In python, set is going to be lists.

- $B \subset A$: all elements in B are in A but not vice versa.

  and $\exists\, x \in A$, s.t. $x \notin B$

- $B \subseteq A$:
- ϕ : Empty set. (Collection with no items/elements in it)
- Ω : sample space [• All the sets are going to be subsets of this sample space.]

  • Always need to define sample space.

- $A \cup B = \{x :$ s.t. $x \in A$ or $x \in B\}$
- $A \cap B = \{x :$ s.t. $x \in A$ & $x \in B\}$
- $A^c = \{x ; x \in \Omega, x \notin A\}$
- $A \cap A^c = \phi$, $A \cup A^c = \Omega$
- $A \setminus B = \{x :$ s.t. $x \in A, x \notin B\}$

  $A^c = \Omega \setminus A$

A^c

Properties:

1. • $A \cup B = B \cup A$
   • $A \cap B = B \cap A$
2. • $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$
3. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
   $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

[Can see these pictorially but can't use that as a proof.]

Try these properties with complements. (De Morgan Laws).

Algebra: is a set system $\mathcal{A}$ s.t.

1. If $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
2. If a finite collection of sets $A_1, A_2, \ldots, A_n \in \mathcal{A}$ then $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$
3. $\emptyset \in \mathcal{A}$

Very algorithmic approach to algebra.

Aug 28, 2023 : Class 2

Algebra

2' Infinite Union

If $A_1, A_2, \ldots A_n, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$] Countably infinite

then $\mathcal{A}$ is σ- algebra

2 prime

Example:

Ω = sample space = $\{H, T\}$

Measurable states = $(\Omega, \mathcal{A})$ → space → structure

We arrange the space to infuse it with probability.

$\Omega = \{H, T\}$

Trivial Algebra:

from step 1 $\mathcal{A} = \{\Omega, \phi\}$

Another example $\mathcal{A} = \{H, T, H \cup T, \phi\}$ { This is from the three steps }

$H \in \mathcal{A}$, $\cancel{H \in \mathcal{A}}$

$H \cup T \in \mathcal{A}$

$\Rightarrow \phi \in \mathcal{A}$ because $(H \cup T)^c = \phi$

- Function $f : \mathcal{A} \longmapsto \mathcal{F}$ is measurable if

$\forall F \in \mathcal{F}, \; f^{-1}(F) = A \in \mathcal{A}$

{ From now on in the course, we are only looking at σ algebra

$f(A) = \{f(x) : x \in A\}$ → evaluating at each x.

$\mathcal{B}$ is a σ- algebra of $[-\infty, x], \forall x \in \mathbb{R}$. It is called
↳ all xs.
Borel's σ- algebra.

(*) Is this surjective function?

(*) A measurable function $X : \mathcal{F} \to \mathcal{B}$ ($\mathcal{F}$ is σ algebra on Ω) is called a random variable.

(just a representation of the original space)

Example mapping:
$H \Rightarrow 1, 2$
$T \Rightarrow 0, 2$
$\phi \to \phi$

ω can be either H or T, so it is not measurable.

So we see it's not random or a variable, it's just a mapping.

(*) Probability Measure is $\boxed{\text{set function}}$ of the elements of $\mathcal{F}$ s.t.

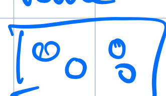
1. $\forall A \in \mathcal{F} \Rightarrow P(A) \geq 0$
2. $P(\Omega) = 1$
3. $\{A_i\}_{i=1}^{\infty}$ is s.t $A_i \in \mathcal{F}$, $A_i \cap A_j = \phi$ if $i \neq j$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

(σ- additivity of probability measure)

Doesn't have to add up to 1 (wow)

Can have something like a swiss cheese



Set function → Takes sets as inputs & outputs number

Probability space : tuple of $(\Omega, \mathcal{F}, P)$

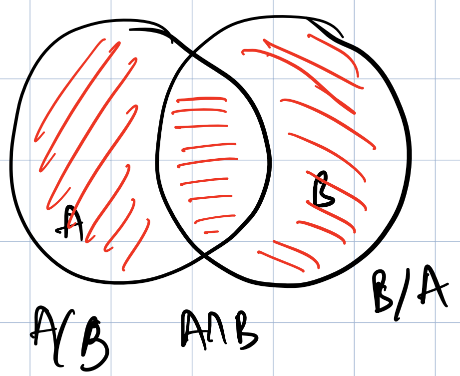
Properties of Probability space :—

1. A, A^c, $P(A \cup A^c) = P(\Omega) = 1 = P(A) + P(A^c)$

2. $P(\phi) = 0$, (when A is considered Ω, we'll have this implication.)

$P(A^c) = 1 - P(A)$

3. $P(A) = 1 - P(A^c) \leq 1$



$A = (A \setminus B) \cup (A \cap B)$

$B = (B \setminus A) \cup (A \cap B)$

$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A)$

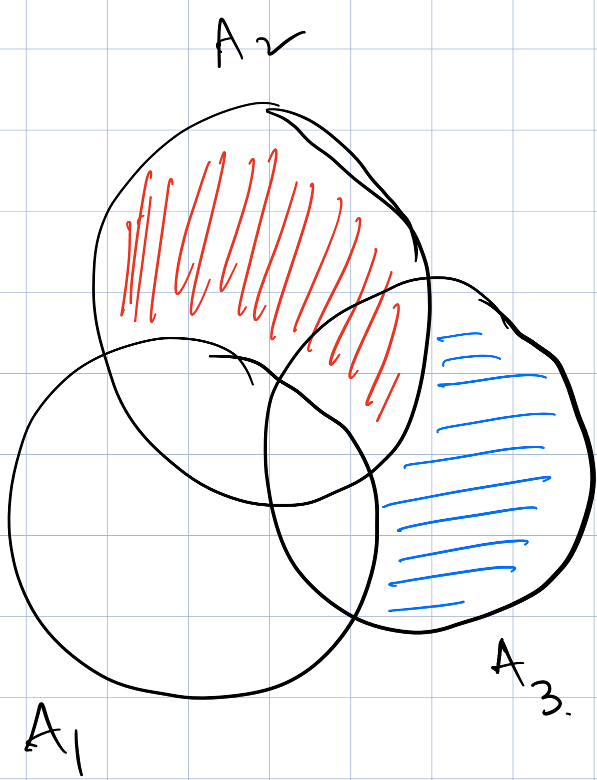
$$P(A) = P(A \setminus B) + P(A \cap B)$$
$$P(B) = P(B \setminus A) + P(A \cap B)$$
$$P(A \cup B) = P(A \setminus B) + P(A \cap B) + P(B \setminus A)$$
$$\Rightarrow P(A \cup B) - P(A) - P(B) = -P(A \cap B)$$
$$\Rightarrow P(A \cup B) \leq P(A) + P(B)$$

30 Aug 2023 | Class 3

$\{A_i\}_{i=1}^{\infty} \in \mathcal{F}$, $\underbrace{A_i \cap A_j = \phi}_{\text{disjoint}}$ if $i \neq j \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$P\left(\bigcup_{i=1}^{\infty} A_i\right)$ where $A_i \cap A_j \neq \phi$



Set $B_1 = A_1$

$B_2 = A_2 \setminus A_1$

$B_3 = A_3 \setminus (A_1 \cup A_2)$

$B_k = A_k \setminus \left(\bigcup_{i<k} A_i\right)$

$\Rightarrow \bigcup_{i=1}^{k} B_k = \bigcup_{i=1}^{k} A_k$

$B_{k+1} = A_{k+1} \setminus \left(\bigcup_{i=1}^{k} A_i\right)$

The B sets are disjoint.

$$B_k \subseteq A_k \Rightarrow P(B_k) \leq P(A_k)$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} P(B_i) \leq \sum_{i=1}^{n} P(A_i)$$

$$\Longrightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A)$$

$$(a+b)^n = (a+b)(a+b) \dots (a+b)$$

The concept of urns came from mathematicians looking at governance in Athens, where people put black/white marble in an urn as a vote! - WOW

Permutations :

n items $\quad n(n-1)(n-2) \dots 1 = n!$

- A collection with replacement.

Collection = k ~~items~~ marbles

$k \leq n$ (total number of marbles)

ordered $\quad$ unordered selections

	Ordered	Unor. collection
W/O replacement	$\frac{n!}{(n-k)!}$	$\frac{n!}{(n-k)!\,k!} = \binom{n}{k}$
with replacement	n^K	$\frac{(n+k-1)!}{k!\,(n-1)!} = \binom{n+k-1}{k}$

n	n	n	n		...	n
1	2	3	4	.	. .	k

$\Rightarrow n^k$

n	n-1	n-2	.	...	.	n-k+1
1	2	3	4	. .	. . .	k

$= \frac{n!}{(n-k)!}$

Unordered, w/o replacement:

In the above example, we can have an ordered collection, so we are overcounting. So we devide by $k!$

$$= \frac{n!}{(n-k)!\,k!}$$

$$(a+b)(a+b)\ldots(a+b) = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Unordered, with replacement.

0 0 | | | 0 | | | 0 0

$(n-1) \rightarrow$ lines

$k \rightarrow 0$

$$\frac{(n+k-1)!}{k!\,(n-1)!}$$

permutations

To account for shuffling

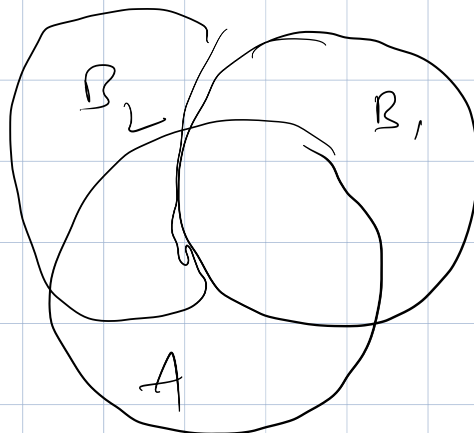
Events A & B are independent
if $P(A \cap B) = P(A)\,P(B)$ $= A \perp B$

Conditional Probability

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

To normalize by the probability of B.

Sept 4



$B_i \cap B_j = \phi$, if $i \neq j$

$\bigcup_{j=1}^{n} B_j \supseteq A$

As long as A is completely covered by Bs.

$$P(A) = \sum_{i=1}^{n} P(A|B_j)\,P(B_j)$$

weighting by the proportion of individuals in B_j.

$$P(A) = P(A \cap B_j)$$

This is correct when the following holds $\bigcup_{j=1}^{n} B_j \supseteq A \Rightarrow \bigcup_{j=1}^{n} (A \cap B_j) = A$.

The above tells us how to recover A from the Bs. (the disjoint sets).

$B_i \cap B_j = \phi$

$\bigcup_{j=1}^{n} B_j \supseteq A$

$$P(B_i \mid A) = \frac{P(A / B_i) \cdot P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)\, P(B_j)}$$

Posterior

Prior

Bayes Law

Likelihood function (from the model)

We'll do this in more detail later.

Random Variables

Measurable f^n mapping from sigma algebra to real domain.

Translate abstract set to a borel set / real number. → rainbows / unicorns etc.

Measurable space $(\mathbb{R}, \mathcal{B})$

Side Note.

$P_X(B) = P(X^{-1}(B))$

$X(A) \equiv \{X(\omega), \omega \in A\}$

Distribution of random variable X

$$P\Big(\underbrace{\{\omega : X(\omega) \in B\}}_{A}\Big) = P_X(B)$$

$$= P(X(\omega) \in B)$$

Probability space: $(\mathbb{R}, \mathcal{B}, P_X)$

- Mapped everything on real line
- Borel Algebra works on real line
- Standardized probability

After the mapping, RV converted everything to reals, so we are only concerned about real lines. We are no longer concerned about Ω (elements of the abstract space). But should remember that this is a contraction & one may lose some info in the process.

To make things simpler, we need not look at the entire distribution of R.V.

$$F_X(x) = P_X((-\infty, x]) \, , \; x \in \mathbb{R}$$

is the cumulative dist^n function.

<u>Properties of cdf :</u>

1) If $x_1 \leq x_2$, then $F_X(x_1) \leq F_X(x_2)$

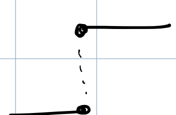
2) (cadlag) : $\lim_{x \downarrow x_0} F_X(x) = F(x_0)$

(Approaching x_0 from right)

$\lim_{x \uparrow x_0} F_X(x)$ exists

(Approaching x_0 from left)

$\Rightarrow$ the function can exhibit weird jumps.



jumps are only one way $\Rightarrow$ post jump, the function has to behave properly.

3) $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to +\infty} F_X(x) = 1$.

Kolmogorov

Theorem: Suppose that F satisfy properties 1–3,

?

then $\exists$ probability space $(\Omega, \mathcal{F}, P)$ and r.v. X on it s.t. $F(x) = F_X(x)$

Kolmogorov $\to$ One of the most accessible real analysis textbooks. // Another is Rudin.

Definition: R.v. X is called absolutely continuous if $\exists$ $f_X(\,)$, s.t. $P_X(B) = \int_B f_X(x) \cdot dx$

Examples:-

1. Bernoulli r.v. (p)

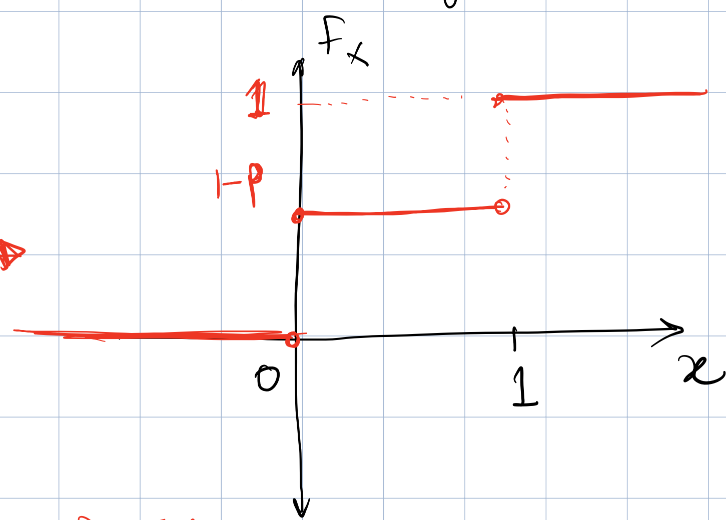
Coin tosses

rv X

$$X = \begin{cases} 1, & \text{with prob } p \\ 0, & \text{with prob } 1-p \end{cases}$$

$P_X((-\infty, x])$

$P_X((-\infty, 1])$



This is not an absolutely continuous R.V.

2. Binomial R.V. (p, n) number of coin tosses.

(n independent Bernoulli experiments)

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

probability of 1s

probability of 0s

Our summation of the combinatorics:
$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1.$$

(Newton's Binomial.)

$\binom{n}{k}$ we don't care about the orders.

3. Poisson Random Variable.

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$k = 0, 1, \ldots, +\infty$

$$\sum_{k=0}^{+\infty} P(X=k) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

6 Sep

4. Uniform Random Variable

$U[a,b]$ → uniform on segment b/w a & b

$$P_X(B) = \int_{B \cap [a,b]} \frac{1}{b-a} \, dx$$

5. Exponential Distribution (α)

$$P_X(B) = \int_{B \cap (0,+\infty)} \alpha e^{-\alpha x} \, dx$$

6. Normal Distribution (μ, δ^2)

$$P_X(B) = \int_B \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{(x-\mu)^2}{2\delta^2}} \, dx$$

Functional transformation of random variables:-

$$Y = g(X)$$

$$F_Y(y) = P(g(X) \leq y) = P(X \in g^{-1}((-\infty, y]))$$

If $g()$ is invertible (& increasing w.l.g), then

$$F_Y(y) = P(g(x) \leq y) = P(X \leq g^{-1}(y))$$

$$= F_X(g^{-1}(y))$$

$$\frac{d}{dt} \int_{a(t)}^{b(t)} f(x,t)\, dx = f(b(t),t)\, b'(t) - f(a(t),t) \cdot a'(t)$$

$$+ \int_{a(t)}^{b(t)} \frac{\partial f(x,t)}{\partial t} \cdot \partial x$$

Example $X \sim U[0,1]$, $Y = -\log(X)$

$\downarrow$

x is uniform on 0 to 1

CDF of R.V. Y :

$$F_Y(y) = P(-\log X \leq y)$$

$$= P(\log X \geq -y)$$

$$= P(X \geq e^{-y})$$

$$= 1 - P(X \leq e^{-y})$$

$$P_X(B) = \int_{B \cap [0,1]} dx = \int_0^{e^{-y}} dx = e^{-y}$$

$$= 1 - e^{-y}$$

Example

$\left\{ \begin{array}{c} \text{Chi square} \\ \text{RV with 1} \\ \underline{\text{dof}} \end{array} \right\}$

$X \sim N(0,1) \quad , \quad Y = X^2$

$$F_Y(y) = P(X^2 \leq y)$$

$$= P(-\sqrt{y} \leq x \leq \sqrt{y})$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \cdot dx$$

We need to apply

If the density exists (which it does because the RHS is differentiable:

$$f_Y(y) = F_Y'(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$$

Multiple Dimensions

Multidimensional extension of the analysis above

$$(\Omega, F, P) \rightarrow \text{in } \overset{RV}{\cancel{2d}} \text{ single dimension}$$

$$\mathbf{X}(\omega) = \begin{pmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_n(\omega) \end{pmatrix}$$

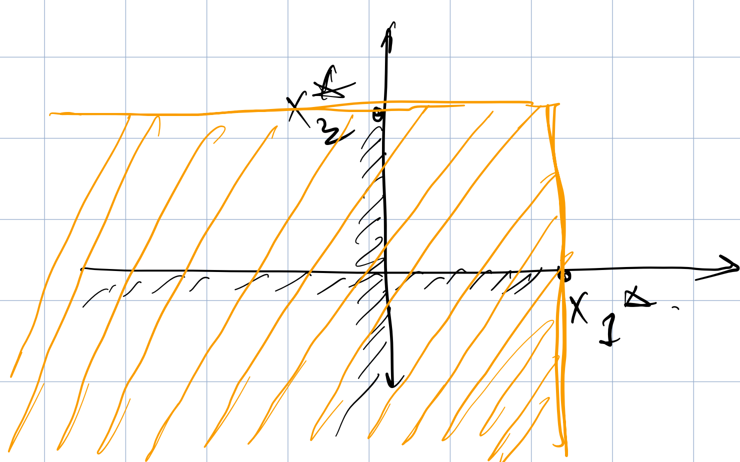
$\downarrow$

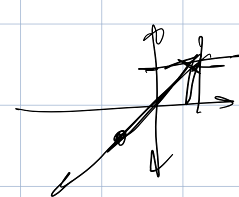
vector

$\mathbf{X} : \Omega \rightarrow \mathbb{R}^n$ is a random vector.

Cartesian Product Across Dimensions:-

$$[-\infty, x_1] \times [-\infty, x_2] \times \ldots \times [-\infty, x_n]$$





$$F_X(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

$$f_{\tilde{X}}(x_1, \ldots, x_{n-1}) = \lim_{x_n \to +\infty} F_X(x_1, \ldots, x_n)$$

$$\lim_{x_n \to -\infty} (x_1, \ldots, x_n) = 0$$

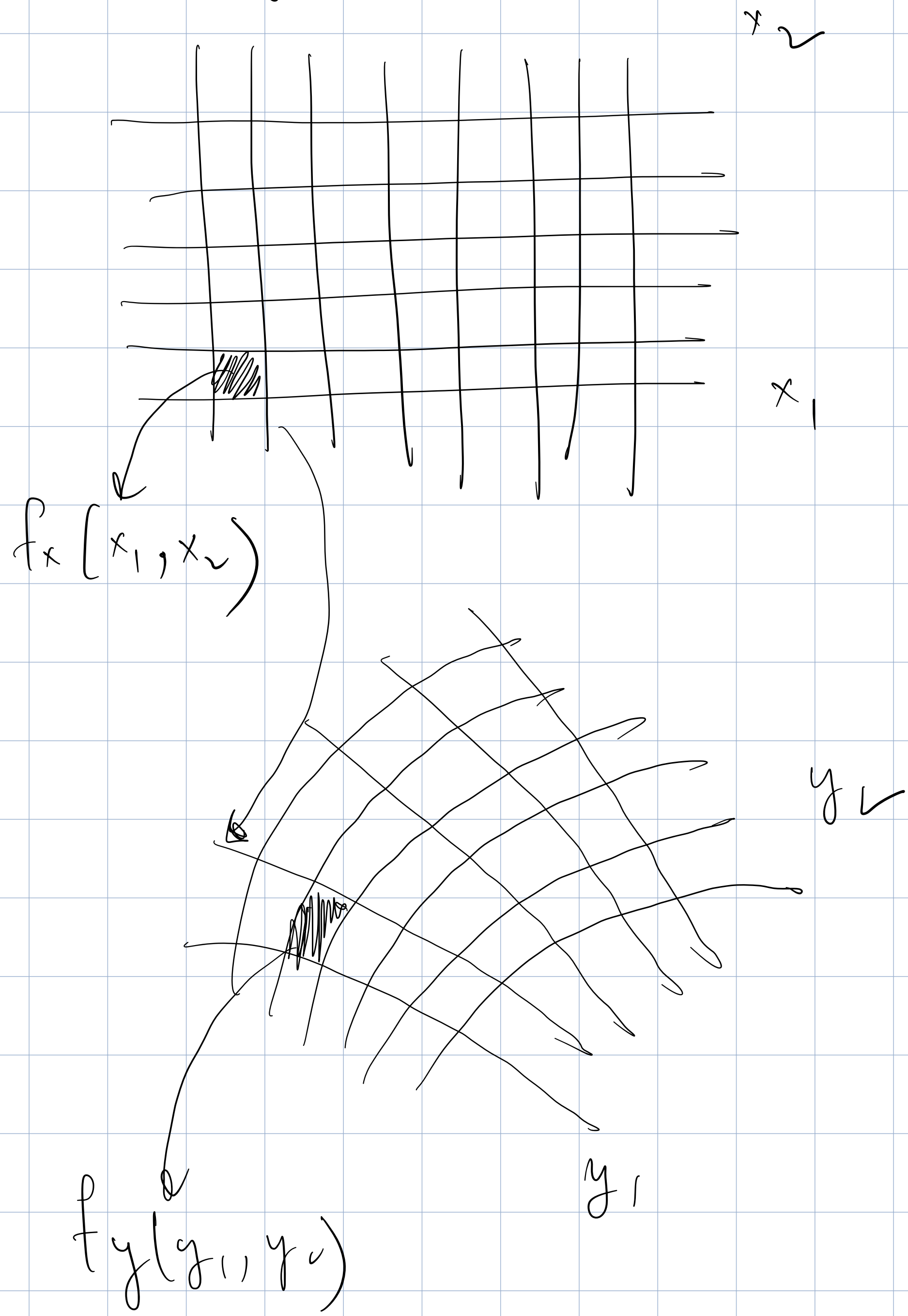
(*) If a f^n is not differentiable at some point → it should be a concern. for problems etc. (Need to look into how to find that)

If X is abs. continuous

$$\lim_{x_n \to \infty} \underbrace{F_X(x_1, \ldots, x_n)}_{\boxed{\text{CDF}}} = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(t_1, \ldots, t_n)\, dt_1 \ldots dt_n$$

$$f_{\tilde{X}}(x_1, \ldots, x_{n-1}) = \int_{-\infty}^{+\infty} f_X(x_1, \ldots, x_{n-1}, t_n)\, dt_n$$

→ density



On transforming f_X to f_Y the volumes of the density would no longer be preserved.
Jacobian is used to adjust for the change.

$X_1, \ldots, X_n$ $\qquad f_{X_1, \ldots X_n}(x_1 \ldots x_n)$

$$Y_1 = g_1(X_1, \ldots X_n)$$
$$\vdots$$
$$Y_n = g_n(X_1, \ldots X_n)$$
$$f_{X_1, X_n}(X_1, \ldots X_n)$$

$\Rightarrow$

$$X_1 = r_1(Y_1, \ldots, Y_n)$$
$$\vdots$$
$$X_n = r_n(Y_1, \ldots Y_n)$$

Computing density at particular points.

$$J = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial g_n}{\partial x_n} & & \frac{\partial g_n}{\partial x_n} \end{pmatrix}$$

$$\underbrace{f_{x_1,\dots x_n}\left(r_1(y_1\dots y_n),\dots,r_n(y_1,\dots y_n)\right)}_{\text{density of } f_{y_1\dots y_n}(y_1,\dots,y_n)}\underbrace{\left|\det\left(J^{-1}\right)\right|}_{}$$

density of $f_{y_1,\dots y_n}(y_1,\dots,y_n)$

(transformed variable.

To correct for the change in the volume.

$\boxed{\text{Sep 11}}$

- $P(A \cap B) = P(A)P(B)$ $\quad A \perp B$
  when events are independent.

- Independence of random variables

Random variables $X_1, \dots X_n$ are independent iff

$$F_{X_1 \dots n}(x_1 \dots x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n)$$

We don't need all the distⁿ of the random variables. We can simplify this further & work with statistics

Expectation

$$E[X] = \int_\Omega X(\omega)\, P(d\omega)$$

Lebesgue-Stieltjes Integral.

this does not matter except ~~to~~ knowing that when the space is discrete, then its all around more spaces (* ?)

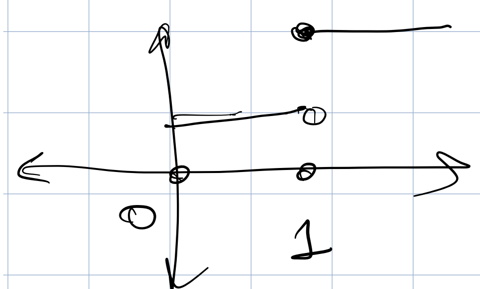
$\downarrow$

then we need to take the sum

$$= \int_{-\infty}^{\infty} x \, dF_X(x)$$

Interpret it as a switching device.

For Bernoulli random variable, we can't integrate but have to take sum



$$E(x) = 0(1-p) + 1 \cdot p = p.$$

When dealing with absolutely continuous r.v.; the above integral would be fully defined.

For absolutely continuous r.v.

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

Continuity here is about the probability density f^n.

When expectation is not defined!

Cauchy's distribution: Absolutely continuous rv → density is defined as:

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

$$E = \int_{-c}^{c} x \, f_X(x) \cdot dx = \int_{0}^{c} \frac{1}{\pi(1+x^2)} \cdot x \cdot dx$$

$$= \int_0^c \frac{1}{2\pi}\left(\log(1+x^2)\right) \cdot dx$$

$$= \frac{1}{2\pi} \log(1+c^2)$$

The expectation does not exist. The integral does not converge.

But if $f_X(x) = \frac{c}{1+x^{2+\delta}} \rightarrow$ (integral converges.)

Properties of expectations :—

1) $E(a+bX) = a + bE(X)$

2) For any X_1 & X_2 (r.v. where expectations exist), s.t. $|E[X_1]|$, $|E[X_2]| < \infty$

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

3) If $a \leq X \leq b$, then

$$a \leq E[X] \leq b$$

$$|E[X]| \leq E[|X|]$$

Q what does it mean to not have an expectation?

When we do multiple sampling, we won't have the same/similar expectation in different samples from the pop^n.

- Bootstrapping fails too

4) If $X \geq 0$ & $E[X] = 0 \Rightarrow X = 0$

$E[g(X)]$

$g(X)$ is also a R.V.

$$\mathbb{1}\{x \in A\} = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases}$$

$$E\left[\mathbb{1}\{X \in A\}\right] = \int_A P(dw) = P(A)$$

Examples :-

1. Bernoulli (p) $E[X] = p$

2. Binomial R.V. (n, p)

$E[X_1 + X_2 + \ldots\ldots + X_n]$ (Bernoulli R.V.)

$= np$

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} k \leftarrow$$ Can also do like this, just more tedious

$= np$

3. Poisson (λ)

$$E[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda}$$

Let $k-1=i$ $\Rightarrow \sum_{i=0}^{\infty} \frac{\lambda^{i+1}}{i!} e^{-\lambda}$

$$= \lambda e^{-\lambda} \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^i}{i!}}_{e^{\lambda}}$$

$$= \lambda$$

4. Normal R.V. :- (μ, δ^2)

$$f(x) = \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{(x-\mu)^2}{2\delta^2}}$$

$$E(X) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{(x-\mu)^2}{2\delta^2}} \cdot (x-\mu+\mu) \cdot dx$$

$$= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{(x-\mu)^2}{2\delta^2}} (x-\mu) dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{t^2}{2\delta^2}} t \cdot dt + \mu \int_{-\infty}^{\infty} \frac{1}{2\pi\delta^2} e^{-\frac{(x-\mu)^2}{2\delta^2}} \cdot dx$$

$= \quad 0 \quad + \quad 1$

$= 1$

Conditional Expectations :-

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

$$\Rightarrow \quad P(Y \in A \mid X \in B)$$

$$E[Y / X \in B] = \int_{\Omega} y(\omega) \, P(dy \mid X \in B)$$

Example (Survival Distribution)

$$F(X \leq x) = 1 - e^{-\alpha x}$$

$$F(X \leq x \mid X \geq a)$$

$$f_X(x) = E[\mathbb{1}\{X \leq x\}]$$

$$A = \mathbb{1}\{X \leq x\}$$
$$B = \mathbb{1}\{X \geq a$$

$$P(B) = E[\mathbb{1}\{X \geq a\}] = e^{-\alpha a}$$

$$A \cap B = \begin{cases} \phi, & \text{if } x < a \\ [a, x], & \text{if } x \geq a \end{cases}$$

$$P(A \cap B) = E[\mathbb{1}\{X \in [a, x]\}]$$
$$= e^{-\alpha a} - e^{-\alpha x}$$

$$E(X \leq x \mid X \geq a) = 1 - e^{-\alpha(x-a)}$$

September 13 : Machine learning // Making predictions

Minimization problem:-

Loss function $X - a$

What we are predicting modelled as r.v.

Our prediction

Risk function : $E[(X-a)^2]$

Expectations is linear operator (inherited from the integral we did).

$$E = [X^2 - 2aX + a^2]$$

$$= \underbrace{E[X^2] - 2aE[X] + a^2}_{\text{minimize wrt } a.}$$

a is a constant

$$a^* = E[X]$$

◦ Including a confounding variable:—

$$E[(X-a)^2 \mid Z] \Rightarrow a^*(Z) = E[X \mid Z]$$

Regression function (year 2)

A correlator which helps us to estimate better.

OLS is trying to solve the following problem:-

$$E\left[(X - a(Z))^2\right] = \int_{-\infty}^{\infty} (x - a(z))^2 f(x|z)\, dx,$$

Here the problem is to find an unknown function. Under some conditions, it would be conditional expectation.

We find a for each z.

$$a = \mathbb{E}[X \mid Z = z]$$

This is solved point-wise

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

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

Non-central

k^{th} moment of a r.v. : $E[X^k]$

- Central moments : $X - E[X]$

- k^{th} central moments : $E[(X-E[X])^k]$

Knowing moments of a r.v. can help us to find the whole distribution.

Examples :

1) Bernoulli R.V.

$$\mathrm{Var}(X) = p - p^2$$
$$= p(1-p).$$

{ Using this def: $E[X^2]-E[X]^2$ }

Properties of variance (non-exhaustive):—

1. $Var(X) \geq 0$, if $Var(X) = 0 \Rightarrow X = c$
a.e. (almost everywhere) / with probability = 1.

$\left(\min_a E\left[(X-a)^2 \right] = 0 \right)$ solving this

2. $Var(a + bX) = b^2 \, Var(X)$

Covariance

$$cov(X, Y) = E\left[(X - E[X])(Y - E[Y]) \right]$$

3. $Var(X+Y) = Var(X) + Var(Y) + 2cov(X, Y)$

If $X \perp Y$ (X & Y are independent)

Independence of joint distribution $\Rightarrow$ multiplication of marginal density distribution.

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

$$= E[X-E(X)]\,E[Y-E(Y)] = 0$$

<u>Examples</u> :

1) Bernoulli R.V.

$$\mathrm{Var}(X) = p - p^2$$
$$= p(1-p).$$

{ Using this def: $E[X^2] - E[X]^2$ }

2) Binomial Distribution

Binomial (n, p) $\quad X_1 + X_2 + \dots + X_n$

$$\mathrm{Var}(X_1 + X_2 + \dots + X_n)$$

All r.v. have bernoulli distn.

$$= n\,\underbrace{\mathrm{Var}(X_1)}$$

Var of bernoulli

$= np(1-p)$

3) Poisson (λ)

$$E[X^2] = \sum_{k=0}^{+\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= \sum_{k=0}^{\infty} k \frac{\lambda^k}{(k-1)!} e^{-\lambda}$$

We know

$$(\lambda^k)' = k \lambda^{k-1}$$

$$= e^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{k \overbrace{\lambda^{k-1}}^{(\lambda^k)'}}{(k-1)!} \quad ** $$

$$\sum_{k=0}^{\infty} \frac{k \lambda^{k-1}}{(k-1)!} = \sum_{k=1}^{\infty} \frac{(\lambda^k)'}{(k-1)!} = \left(\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \right)' \quad *$$

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda e^{\lambda}$$

Put in *

$$(\lambda e^{\lambda})' = e^{\lambda} + \lambda e^{\lambda}$$

Plugging in **

$\lambda + \lambda^2$

$Var(X) = \lambda$

18 September

Inequalities

- Cauchy-Schwarz Inequality

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

Take

$$(|a| - |b|)^2 \geq 0$$

$$\Rightarrow \quad 2|ab| \leq a^2 + b^2$$

$$a = \frac{X}{\sqrt{E[X^2]}} \qquad b = \frac{Y}{\sqrt{E[Y^2]}}$$

$$2 \cdot \frac{|XY|}{\sqrt{E[X^2]\,E[Y^2]}} \leq \frac{X^2}{E[X^2]} + \frac{Y^2}{E[Y^2]}$$

The inequality holds point wise, so it would also hold for expectation.

Taking expectation both side :-

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

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

- <u>Correlation Coefficient</u>

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Properties :-

1. If $X \perp Y \Rightarrow \rho_{XY} = 0$

2. By Cauchy Schwarz inequality:

$$E\left[\left|(X - E(X))(Y - E[Y])\right|\right] \leq \sqrt{Var(X)\,Var(Y)}$$

As, $\left|E[g(X)]\right| \leq E\left[|g(X)|\right]$, (property of expectation)

$$\Rightarrow \underbrace{\left|E[(X - E[X])(Y - E[Y])]\right|}_{|Cov(X,Y)|} \leq E\left[\left|(X - E(X))(Y - E[Y])\right|\right] \leq \sqrt{Var(X)\,Var(Y)}$$

$$\Rightarrow |\rho_{XY}| \leq 1 \Rightarrow -1 \leq \rho_{XY} \leq 1$$

3. If $|\rho_{XY}| = 1$, then with probability 1,
$Y = a + bX$ (linear dependence)

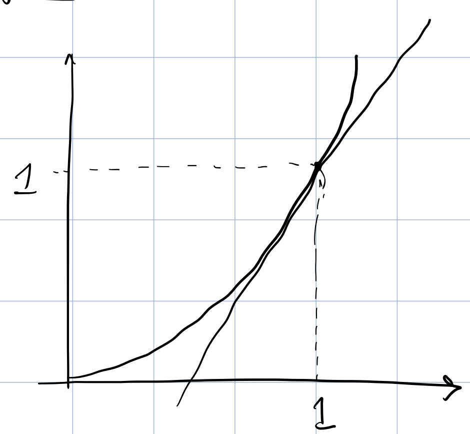
[In applied work, when we look at ρ_{XY}, imp to remember its only telling us about the linear relationship, there could be another r/p too.]

⊛ Hölder's Inequality :

Suppose $r > 1$, then take $s > 0$ s.t.
$\frac{1}{r} + \frac{1}{s} = 1$

$$E[|XY|] \leq E[|X|^r]^{1/r} E[|Y|^s]^{1/s}$$

Proof :—



Inequality b/w convex f^n & tangent line

Convex f^n always lie above the tangent line.

$$x^r - 1 \geq \underbrace{r(x-1)}_{\text{tangent line}} \text{ ⊛}$$

⊛ why do we take 1 here

how is this tangent line? Derivative?

Set $a = \frac{|X|^r}{E[|X|^r]}$, $b = \frac{[Y]^s}{E[|Y|^s]}$

Set $x = \left(\frac{a}{b}\right)^{1/r}$

Revisiting *, $r\left(\frac{a^{1/r}}{b^{1/r}} - 1\right) \leq \frac{a}{b} - 1$

Multiplying both sides by b

$$r\left(a^{1/r} \cdot b^{1/s} - b\right) \leq a - b$$

Plug the value of a & b.

$$r\left(\frac{|X|}{(E[|X|^r])^{1/r}} \frac{|Y|}{(E[|Y|^s])^{1/s}} - \frac{|Y|^s}{E[|Y|^s]}\right) \leq \frac{|X|^r}{E[|X|^r]} - \frac{|Y|^s}{E[|Y|^s]}$$

$$= r\left(\frac{|XY|}{(E[|X|^r])^{1/r}(E[|Y|^r])^{1/s}} - \frac{|Y|^s}{E[|Y|^s]}\right) \leq \frac{|X|^r}{E[|X|^r]} - \frac{|Y|^s}{E[|Y|^s]}$$

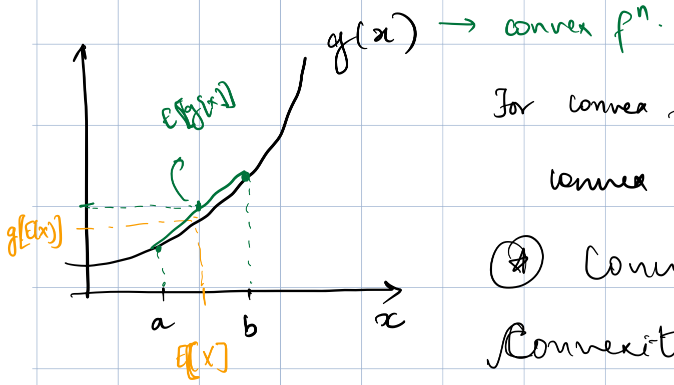
Taking expectations both sides

$$= r\left(\frac{E[|XY|]}{(E[|X|^r])^{1/r}(E[|Y|^r])^{1/s}} - \overbrace{\frac{E[|Y|^s]}{E[|Y|^s]}}^{=1}\right) \leq \overbrace{\frac{E[|X|^r]}{E[|X|^r]}}^{=1} - \overbrace{\frac{E[|Y|^s]}{E[|Y|^s]}}^{=1}$$

$$= \frac{E[|XY|]}{E[|X|^r]^{1/r} E[|Y|^r]^{1/s}} \leq 1$$

* Cauchy-Schwarz is a special case of Hölder's inequality

- Jensen's Inequality (pronounced as Yensen) → Danish



For convex f^n : $E[g[x]] \geq g(E[x])$ for convex f^n $g(\cdot)$

☆ Convexity does not need differentiability

Convexity without differentiability leads to kinks.

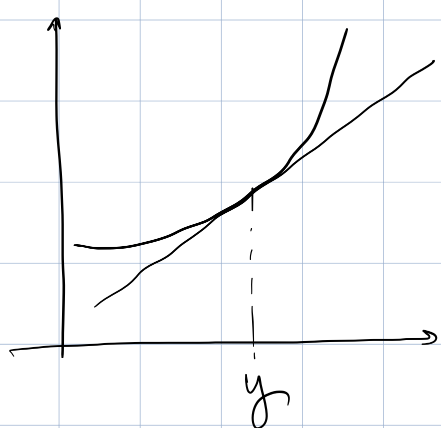
<u>Proof</u> :

20 September

$$\int_S dx = 0 \longrightarrow \text{measure.}$$

Real Analysis book: Kolmogorov, Fomin "Intro real analysis"

Jensen's Inequality



Tangent line

$$g(x) \geq g(y) + \nabla g(y)(x-y)$$

Let $y = E[X]$

$x = X$

$$g(X) \geq g(E[X]) + \nabla g(E[X])(X - E[X])$$

$$E[g(X)] \geq g(E[X]) + \nabla g(E[X])(E[X] - E[X])$$

For concave f^n; the inequality is flipped.

Markov's inequality

$$P(|X| \geq \varepsilon) \leq \frac{E[|X|]}{\varepsilon}$$

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

$$= \int_{|x| \geq \varepsilon} |x| f_X(x)\, dx + \int_{|x| < \varepsilon} |x| f_X(x)\, dx$$

$$\geq \int_{|x| \geq \varepsilon} |x| f_X(x)\, dx$$

$$\geq \int_{|x| \geq \varepsilon} \varepsilon f_X(x) \cdot dx \geq \varepsilon \int_{|x| \geq \varepsilon} f_X(x) \cdot dx$$

$$\geq \varepsilon P(|X| \geq \varepsilon)$$

$$\Rightarrow P(|X| \geq \varepsilon) \leq \frac{E[|X|]}{\varepsilon}$$

(how do you visualize this?)

Markov's inequality generalizes to arbitrary moments of r.v. if they exist! —

$$P(|X| \geq \varepsilon) \leq \frac{E[|X|^k]}{\varepsilon^k}$$

⑥ Existence of moments very deeply connected to the tail behaviour.

Chebyshev Inequality

$$P(|Y - E[Y]| \geq \varepsilon) \leq \frac{E[(Y - E(Y))^2]}{\varepsilon^2}$$

$$\leq \frac{Var(Y)}{\varepsilon^2}$$

$$\underbrace{|E[X]|}_{\text{this would exist}} \leq \underbrace{E[|X|]}_{\text{if this exists}}$$

but not the other way around.

COMPLEX VARIABLES.

- Roots of polynomials. → solving for engineering reasons.
- k^{th} degree polynomial should have k roots

$a_k x^k + a_{k-1} x^{k-1} + \dots + a_0 = 0$

Shortly held belief but shortly found wrong.

$x^2 = -1$

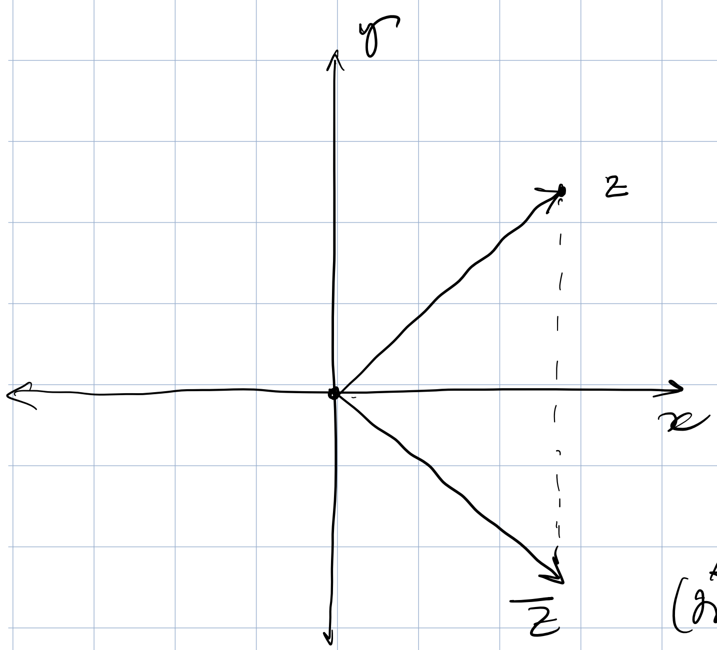
Imaginary number: $i = \sqrt{-1}$

$x = \pm i$

$x^2 = -1$; similarly &

~~$x = \sqrt{-1}$~~

Take real numbers and expand them to the complex domain.



complex numbers

$Z \in \mathbb{C}$

$Z = x + iy, \quad x, y \in \mathbb{R}$

$x = \mathrm{Re}(Z)$

$y = \mathrm{Im}(z)$

$|z| = \sqrt{x^2 + y^2}$

(~~In~~ As euclidian world)

$z^* = \bar{z} = x - iy.$

complex conjugate of complex no. z.

$$z\bar{z} = |z|^2 = (x+iy)(x-iy)$$
$$= x^2 + ixy - ixy - (i)^2 y^2$$
$$= x^2 + y^2$$

(*) Try to plot <u>sin</u>(z)

25 sep

Gen. form of complex no.
$x = a + bi$

$x^4 = -1$

$\Rightarrow (x^2)^2 = -1$

$\Rightarrow x^2 = \pm i$

$x = a + bi$

$x^2 = a^2 + b^2(i)^2 + 2abi = i$

$= a^2 - b^2 + 2abi = i$

$a^2 - b^2 = 0$

$2ab = 1 \Rightarrow ab = \frac{1}{2}$

$|a| = |b|$

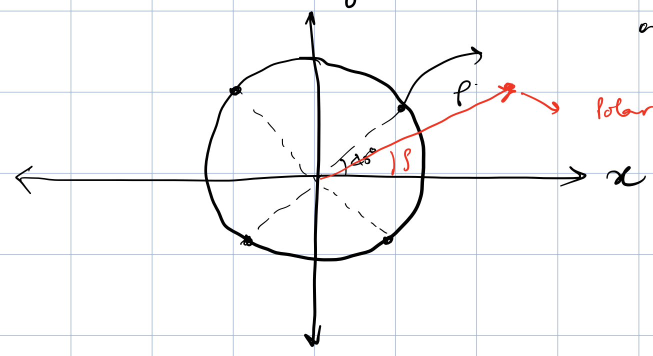
$a = b = \frac{1}{\sqrt{2}}$, $a = b = -\frac{1}{\sqrt{2}}$

Roots :-

$\frac{1+i}{\sqrt{2}}$ $\quad$ $-\frac{(1+i)}{\sqrt{2}}$

$\frac{1-i}{\sqrt{2}}$ $\quad$ $\frac{-1+i}{\sqrt{2}}$

Complex numbers can be expressed as points on the number line:



Polar coordinates:

complex no.

modulus

angle.

$z = \rho\cos\varphi + i\rho\sin\varphi = \rho(\cos\varphi + i\sin\varphi)$ *

Euler's Formula: $z = \rho e^{i\varphi}$ (Using Taylor series expansion)

$z^4 = -1 \Rightarrow \rho = 1$ Modulus $\to 1. \to \rho^4 e^{4i\varphi} = -1$

take mod on both side.

as it is distance, it has to be positive.

$e^{4i\varphi} = -1 = e^{i\pi}$

General expression $e^{i(\pi + 2\pi k)}$

Use the graph to see where -1 lie. can also confirm using *

$\varphi = \frac{\pi}{4} + \frac{\pi k}{2}$

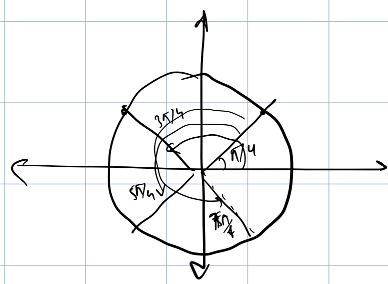
when $k=0$; $\varphi = \frac{\pi}{4}$ $\to$ first $45°$ degree pt.

$k=1$; $\varphi = \frac{\pi}{4} + \frac{\pi}{2} = \frac{3\pi}{4}$

$e^{i\pi/4} = \cos\left(\frac{\pi}{4}\right) + i\sin\left(\frac{\pi}{4}\right)$
$= \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}$

$k=2$, $\varphi = \frac{\pi}{4} + \pi = \frac{5\pi}{4}$

$k=3$; $\varphi = \frac{\pi}{4} + \frac{3\pi}{2} = \frac{\pi + 6\pi}{4} = \frac{7\pi}{4}$



$e^{i\pi/4} = \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}$

$e^{i3\pi/4} = -\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}$

$e^{i5\pi/4} = -\frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}$

$z^6 = -1$ (exact same logic)

* Characteristic Function of RV X

$\phi_X(t) = E[e^{itX}]$ (complex valued R.V.), $t \in \mathbb{R}$

If X is abs cont., $\phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(x).dx$

* Knowing the characteristic f^n is equivalent to knowing the dist^n function.

Properties

always on the unit circle.

Real positive number.

1. $\phi(0) = 1$, $|\phi_X(t)| \leq 1$, $\left|\int_{-\infty}^{+\infty} e^{itx} f_X(x).dx\right| \leq$

$\int_{-\infty}^{+\infty} f_X(x)dx = 1.$

~~Why?~~

density from $-\infty$ to $+\infty$.

2. $\phi_{a+bX}(t) = E\left[e^{it(a+bX)}\right] = E\left[e^{ita} \cdot e^{itbX}\right]$

$$= e^{ita} \cdot E\left[e^{itbX}\right]$$

$$= e^{ita}\,\phi_X(bt)$$

3. If $X_1, \dots, X_n$ are independent,

$$\phi_{X_1+X_2+\dots+X_n}(t) = \phi_{X_1}(t) \dots \phi_{X_n}(t)$$

This is equal to the Product of expectation.

$$E\left[e^{it(X_1+X_2+\dots+X_n)}\right] = E\left[e^{itX_1} e^{itX_2} \dots e^{itX_n}\right] = \phi_{X_1}(t)\,\phi_{X_2}(t) \dots \phi_{X_n}(t)$$

Examples

RV(1)

1. Bernoulli (p)

RV(0)

$$\phi_X(t) = p e^{it} + 1 - p = p(e^{it} - 1) + 1$$

2. Binomial (n, p) $\quad Y = X_1 + X_2 + \dots + X_n$, X_i is Bernoulli (p)

$$\phi_Y(t) = \left(p(e^{it} - 1) + 1\right)^n$$

3. Poisson (λ)

$$\phi_X(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{it}\lambda)^k}{k!}$$ (Taylor expansion of the exponent)

$$= e^{(e^{it}-1)\lambda}$$

Back to properties :—

4. $\phi_X(t)$ is uniformly continuous.

[CB define $m_X(t) = E[e^{tX}]$ instead of defining the complex numbers. That is bad because this requires the existence of expectation of the exponent of rv X which may not exist for many rvs.]

5. If k-th moment r.v. X exists, then

$$\phi_X^{(k)}(0) = (i)^k E[X^k]$$

$$E[e^{itX}]' = E[iXe^{itX}]$$

$$t=0 \Rightarrow iE[X]$$

$$E[e^{itX}]'' = E[i^2X^2e^{itX}]$$

$$t=0 \Rightarrow (i)^2[X^2]$$

$$\vdots \qquad \vdots$$

(Assuming can take differential inside the expectation.)

Sep 27

- $X = a$, w.p. $= 1$
  $\phi_X(t) = e^{iat}$

Theorem: function $\phi(t)$, s.t. $\phi(0) = 1$ is a characteristic function iff it is positive semi definite.

Def: Function $f(t)$ is PSD if $\forall\, t_1, \ldots, t_n \in \mathbb{R}$ matrix F with $F_{ij} = f(t_i - t_j)$ is PSD.

Theorem: If rv X is absolutely continous, then $\lim_{t \to \infty} \phi(t) = 0$. Eg $\rightarrow$ [bernoulli is not abs continous, so this property does not hold for it.]

Theorem: For r.v. X and Y, $F_X(z) \equiv F_Y(z)\ \forall\, z \in \mathbb{R}$ iff $\phi_X(t) = \phi_Y(t)\ \forall\, t \in \mathbb{R}$

X is uniform on $[a,b]$ $\longrightarrow$ continous R.V.

$$\phi_X(t) = \int_a^b e^{itx} \frac{1}{b-a}\, dx$$

Integrate over the density

$$= \frac{1}{b-a} \cdot \frac{1}{it} e^{itx} \Big|_a^b$$

$$= \frac{e^{itb} - e^{ita}}{itb - ita}$$

$$\frac{1}{i} = a + bi$$

$$\Rightarrow 1 = ai - b \qquad (i^2 = -1)$$

$$\Rightarrow a = 0 \; ; \; b = -1$$

$$\Rightarrow \frac{1}{i} = -i$$

Moment Generating Function. (Mgf)

$$m_X(t) = E\left[e^{tX}\right]$$

$$m_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) \cdot dx$$

(Not sure what this is called.)

More on this, in the book. Dennis does not like this because it may not always exist as already ~~listed~~ noted in the notes before.

$$e^{tX} = \sum_{k=0}^{\infty} \frac{(tX)^k}{k!}$$

$$m_X(t) = E\left[\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right]$$

$$= \sum_{k=0}^{\infty} \frac{t^k}{k!} E\left[X^k\right]$$

{We may not always be able to swap Σ & E. because of the summation to the "infinity"}

* [Stochastic Convergence]

$X_n \ (\Omega, \mathcal{F}, P), \quad n = 1, 2, \ldots\ldots$

$L_2: \ \|X_n\|^2_{L_2} = \int_{-\infty}^{+\infty} X_n^2(\omega)\, F_{X_n}(d\omega) = E[X_n^2]$

Did'nt understand why usual convergence doesnt work, but it doesn't so we have to define it differently now:

Need not be a constant!

X_n converges in probability to r.v. X
(notation: $X_n \xrightarrow{P} X$) if $\forall \varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) \longrightarrow 0, \text{ as } n \longrightarrow \infty$$

[The correlation b/w X_n & X as $n \to \infty$, becomes 1.]
(Question on case)

Convergence almost surely (a.s)

X_n converges to r.v. X almost surely (notation $X_n \xrightarrow{a.s.} X$) iff:

$$P\left(\left\{\omega : X_n(\omega) \not\longrightarrow X(\omega)\right\}\right) = 0$$

Almost because it could converge with Probability 0.

Eg:-

$$X_n = \begin{cases} 1, & \text{w.p. } \frac{1}{n} \\ 0, & \text{wp } 1 - \frac{1}{n} \end{cases}$$ (Bernoulli)

$$P(X_n = 1) \equiv P(|X_n - 0| > \varepsilon) = \frac{1}{n} \longrightarrow 0$$

deviating from 0 by more than ε for any ε. R.V. only takes 2 values (0, 1).

$$X_n \xrightarrow{P} 0$$

This will not converge almost surely. (To be done formally in the next class).

Q) When are they equivalent?

Theorem: Suppose that X_n is monotone (wlog $X_n \geq X_{n+1}$)

then $X_n \xrightarrow{P} X \iff X_n \xrightarrow{a.s.} X$

September 18

Mid Term : Around Oct 11

Properties of profit & supply functions

$\pi(p) = \max_{y \in Y} p \cdot y$

$y^*(p) = \arg\max_{y \in Y} p \cdot y$

$\pi : \mathbb{R}^n \longrightarrow \mathbb{R}$ (function)

$y^* : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ (Correspondence)

Note: Results do not rely on regularity assumptions (convexity, differentiability etc.) unless stated differently.

① $\pi(p)$ is non increasing in input prices and non-decreasing in output prices.

Proof: Take 2 price vectors p, p' and choose $y \in y^*(p)$ and $y' \in y^*(p')$

let $p_i' \geq p_i$ for all outputs, $p_i' \leq p_i$ for all inputs.

By WAPM; $p' \cdot y' \geq p' \cdot y$

consider $(p'-p) \cdot y = (p_1'-p_1) y_1 + \dots + (p_n'-p_n) y_n$

$y_i \geq 0$: $p_i' \geq p_i$ $\quad$ $y_i \leq 0$: $p_i' \leq p_i$

$$(p'-p) \cdot y \geq 0 \Rightarrow p' \cdot y \geq p \cdot y$$

$$\pi(p') = p' \cdot y' \geq p' \cdot y \geq p \cdot y = \pi(p)$$

$$\pi(p') \geq \pi(p)$$

② $\pi(p)$ is homogeneous of degree 1.

WTS: $\pi(tp) = t\pi(p)$

Proof: $\pi(tp) = \max_{y \in Y} (tp) \cdot y$

$$= \max_{y \in Y} t(p \cdot y)$$

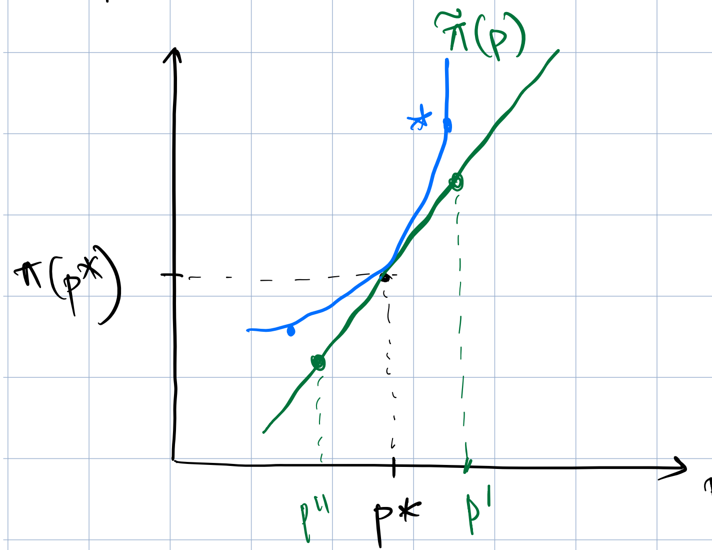
$$= t \max_{y \in Y} p \cdot y$$

$$= t\pi(p)$$

(3) $\pi(p)$ is convex in p

i.e. $\pi(\alpha p + (1-\alpha)p') \leq \alpha \pi(p) + (1-\alpha)\pi(p')$

Graphical intuition.



Single Output Case

At p^*, let optimal prod^n be $(-x^*, y^*)$

$$\pi(p^*) = p^* y^* - w^* \cdot x^*$$

$$\tilde{\pi}(p) = p y^* - w^* \cdot x^*$$

(*) Re-optimized points. When the price changes, either the profit is on the tangent line or above it [

If we do assume π is differentiable:

$$H(p) = D^2\pi(p) = \begin{bmatrix} \frac{\partial^2 \pi}{\partial p_1^2} & \cdots & \frac{\partial^2 \pi}{\partial p_1 \partial p_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \pi}{\partial p_n \partial p_1} & \cdots & \frac{\partial^2 \pi}{\partial p_n \partial p_n} \end{bmatrix}$$

$D^2_{\pi}(p)$ is positive semi-definite.

<u>Properties of $y(p)$:—</u>

① $y(p)$ is homogenous of degree 0.

Proof :—

$$\max_{y \in Y^*} (\alpha p) \cdot y$$

$$= \alpha \max_{y \in Y} p \cdot y$$

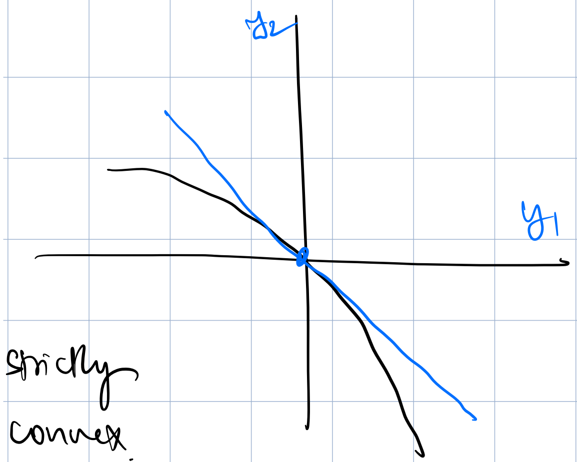
* includes both inputs & outputs whenever you see cap Y.

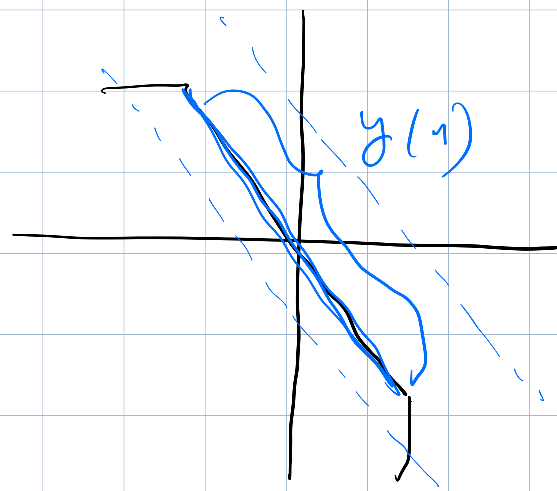
(Scaling all the prices should'nt change your optimal choices)

~~②~~ (a) If Y is convex, then $y(p)$ is convex set.

(b) If Y is strictly convex, then $y(p)$ is a singleton —(?).

Proof by graph: -





⊛ Envelope Theorem → v. imp.

Q: How does the optimal value of a maximization problem changes as a parameter changes?

$$\max_x f(x, \theta)$$

x: choice variable

θ: parameter eg. prices.

Define the value function → f^n of the parameters

$$V(\theta) = \max_x f(x, \theta)$$

[what is $\frac{dV}{d\theta}$?]

One way:

$$V(\theta) = f(x^*(\theta), \theta)$$

A change in θ has 2 effects :-

- Direct Effect : θ changes f directly through the second argument.
- Indirect Effect : θ changes x^*, which changes f.

$$\frac{dV}{d\theta} = \underbrace{\left.\frac{\partial f}{\partial x}\right|_{x = x^*(\theta)}}_{=0 \text{ as optimal pt. (FOC)}} \cdot \frac{dx^*}{d\theta} + \underbrace{\left.\frac{\partial f}{\partial \theta}\right|_{x = x^*(\theta)}}_{\text{evaluating at optimal pt.}}$$

$$\Rightarrow \frac{dV}{d\theta} = \left.\frac{\partial f}{\partial \theta}\right|_{x = x^*(\theta)}$$

- Found all over economics! -
  - Hotelling's lemma
  - Shepard's Lemma.
  - Myerson's Lemma.
  - Benveniste - Scheinkman formula (Macro)

When there are constraints:

$$\max_x f(x, \theta)$$

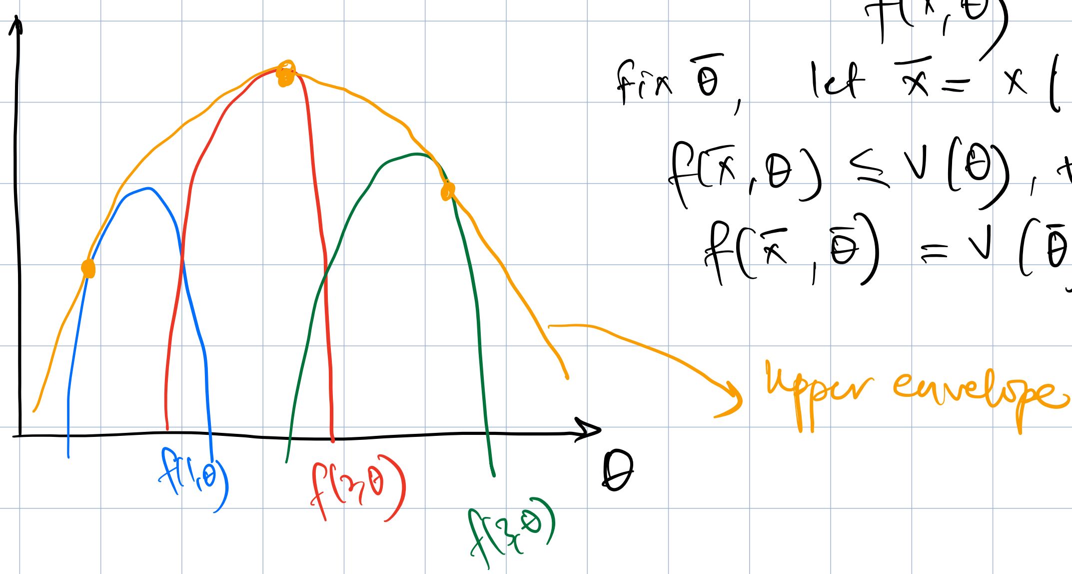
constraints

$$\text{s.t } g_r(x, \theta) \leq c_r \quad \text{for } r = 1, \ldots, k$$

$$\frac{\partial V}{\partial \theta} = \frac{\partial f}{\partial \theta} - \lambda_1 \frac{\partial g_1}{\partial \theta} - \cdots - \lambda_k \frac{\partial g_k}{\partial \theta}$$

To remember: $\mathcal{L} = f(x, \theta) + \lambda_1 (c_1 - g_1(x, \theta)) + \ldots$
$\ldots + \lambda_k (c_k - g_k(x, \theta))$

$$\frac{dV}{d\theta} = \frac{\partial \mathcal{L}}{\partial \theta}$$



$f(x,\theta)$

fix $\bar{\theta}$, let $\bar{x} = x(\bar{\theta})$

$f(\bar{x},\theta) \leq V(\theta)$, $\forall\, \theta$

$f(\bar{x},\bar{\theta}) = V(\bar{\theta})$

Hotelling's Lemma :-

If $\pi(p)$ is differentiable at $\bar{p}$, $\bar{p}_i > 0$, then

$$\frac{\partial \pi(\bar{p})}{\partial p_i} = y_i(\bar{p})$$

Proof: Restrict to single output case. (assume interior solⁿ)

$$\pi(p,w) = \max_{x \geq 0} \underbrace{pf(x) - w\cdot x}_{h(x)}$$

By the envelope theorem

$$\frac{d\pi}{dp}(p,w) = \frac{\partial h}{\partial p}\Big|_{x=x^*(p,w)}$$

$$= f(x^*(p,w)) = y^*(p,w)$$

$$\frac{d\pi}{dw_i} = \frac{\partial h}{\partial w_i} = -x_i\Big|_{x=x^*(p,w)} = -x_i^*(p,w)$$

Example:

$$f(x_1, x_2) = 30 x_1^{2/5} x_2^{2/5}$$

$$\max \; 30 p x_1^{2/5} x_2^{2/5} - w_1 x_1 - w_2 x_2$$

FOCs x_1 :

$$12 p x_1^{-3/5} x_2^{2/5} = w_1$$

$$12 p x_1^{2/5} x_2^{-3/5} = w_2$$

$$x_1^*(p,w) = \frac{12^5 p^5}{w_1^3 w_2^2} \quad (\ast\ast) \qquad x_2^*(p,w) = \frac{12^5 p^5}{w_1^2 w_2^3}$$

$$\pi(p,w) = 30p \cdot \frac{12^4 p^4}{(w_1 w_2)^2} - \frac{12^5 p^5}{(w_1 w_2)^2} - \frac{12^5 p^5}{(w_1 w_2)^2}$$

$$= \frac{1}{2}\left[\frac{12^5 p^5}{w_1^2 w_2^2}\right]$$

To find x_1^*

$$\frac{\partial \pi}{\partial w_1} = -\frac{2}{2}\left[\frac{12^5 p^5}{w_1^3 w_2^2}\right]$$

This is same as $(\ast\ast)$

So envelope theorem gives you a way to recover x_1^*.

(Recover supply from profit fn)

⊛ Integral form of Hotelling's lemma

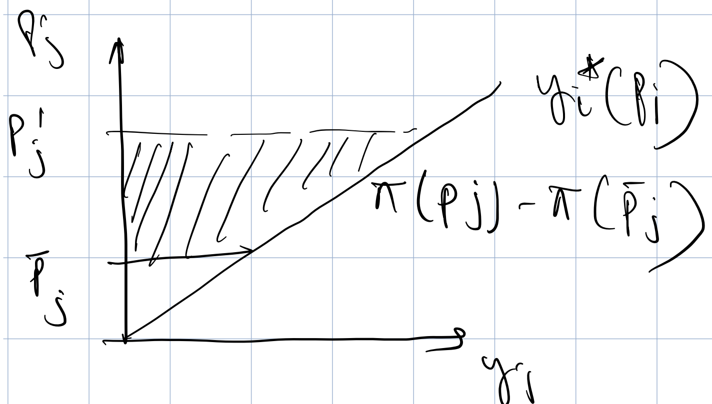
Hold all prices other than p_j fixed at $\bar{p}_{-j}$.

Then, $\pi(p_j, \bar{p}_{-j}) - \pi(\bar{p}_j, \bar{p}_{-j}) = \int_{\bar{p}_j}^{p_j} y_i(s, \bar{p}_{-j})\, ds$

By original Hotelling:

$$\int_{\bar{p}_j}^{p_j} y_j(s, \bar{p}_{-j})\, ds = \int_{\bar{p}_j}^{p_j} \frac{d\pi(s, \bar{p}_{-j})}{d p_j} \cdot ds$$

$$= \pi(p_j, \bar{p}_{-j}) - \pi(\bar{p}_j, \bar{p}_{-j})$$



September 20

Hotelling's Lemma:

$$\frac{\partial \pi}{\partial p_i} = y_i(p)$$

$$Dy(p) = \begin{bmatrix} \frac{\partial y_1}{\partial p_1} & \cdots & \frac{\partial y_n}{\partial p_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial p_n} & \cdots & \frac{\partial y_n}{\partial p_n} \end{bmatrix}$$

[Jacobian / Substitution matrix]

$\boxed{\text{In econ}}$

$$\frac{\partial y_i(p)}{\partial p_j} = \frac{\partial^2 \pi}{\partial p_j \partial p_i}$$

$$Dy(p) = \begin{bmatrix} \frac{\partial^2 \pi}{\partial p_1^2} & \cdots & \frac{\partial^2 \pi}{\partial p_1 \partial p_n} \\ \vdots & & \vdots \\ \frac{\partial^2 \pi}{\partial p_n \partial p_1} & \cdots & \frac{\partial^2 \pi}{\partial p_n^2} \end{bmatrix}$$

This matrix is positive semi-definite.

This implies:—

① $$\frac{\partial y_i}{\partial p_i} = \frac{\partial^2 \pi}{\partial p_i^2} \geq 0$$ (Law of supply)

② $$\frac{\partial y_i}{\partial p_j} = \frac{\partial^2 \pi}{\partial p_j \partial p_i} = \frac{\partial^2 \pi}{\partial p_i \partial p_j} = \frac{\partial y_j}{\partial p_i}$$

Substitution matrix is symmetric. (Not intuitive ⇒ math shows this.)

<u>LeChatelier Principle</u>

"Firms should respond more to price changes in the long run than the short run."

long run ↳ All the inputs are variable.

short run ↳ Some inputs are fixed (factory, contracts etc.)

Suppose, (y^*, x^*) (Single input model)

$= (y(p^*, w^*), x(p^*, w^*))$

at price (p^*, w^*)

Lets fix one input z_i^* in the short run, variable in the long run.

z_i^* is optimal at (p^*, w^*)

Q: How does x_j depend on w_j in the short vs. the long run?

$$\pi(p, w) = \max \ pf(x) - w \cdot x \quad \text{s.t. } x_j \geq 0$$

$$\left. \frac{\partial x_j(p,w)}{\partial w_j} \right|_{(p,w) = (p^*, w^*)}$$

$$\pi^s(p,w) = \max\ pf(x) - w.x$$

s.t. $x_j \geqslant 0$

$x_i = z_i^*$

$$\left.\frac{\partial x_j^s(p,w)}{\partial w_j}\right|_{(p,w)=(p^*,w^*)}$$

Let $h(p,w) = \pi(p,w) - \pi^s(p,w)$

Profits in the LR always have to be greater than in SR as more constraint can only reduce

By def., $h(p,w) \geqslant 0$ & $h(p^*,w^*) = 0$

(p^*,w^*) is a local minima.

$$\Rightarrow \frac{\partial^2 h(p^*,w^*)}{\partial w_j^2} \geqslant 0$$

Diagonal element of Hessian matrix has to be nonnegative for locally convex.

$$\frac{\partial^2 \pi(p^*,w^*)}{\partial w_j^2} - \frac{\partial^2 \pi^s(p^*,w^*)}{\partial w_j^2} \geqslant 0$$

Hotelling: $\frac{\partial \pi (p,w)}{\partial w_j} = -x_j(p,w)$

$$-\frac{\partial x_j(p^*,w^*)}{\partial w_j} + \frac{\partial x_j^c(p^*,w^*)}{\partial w_j} \geq 0$$

These are negative so when we take absolute values we flip the signs.

$$\left| \frac{\partial x_j^c(p^*,w^*)}{\partial w_j} \right| \leq \left| \frac{\partial w_j(p^*,w^*)}{\partial w_j} \right|$$

$\boxed{\text{Cost Minimization}}$

- Single output good firm.
- Pricing power in output market, but not input markets.
- Previous analysis does not work but we can do the following:—

cost minimization problem.

(1) Find the cheapest way to make <u>any</u> target output Y.

(2) Use the "cost function" $c(y)$ to choose optimal p/y combination.

CMP (Cost Minimization Problem) :-

- Fix a target level of output y.

$$\underbrace{c(w, y)}_{\text{parameters}} = \min_x \; w \cdot x$$

s.t.

$x_i \geq 0$

$f(x) \geq y$

Ignore non-negativity for now :-

$$\mathcal{L} = w \cdot x - \lambda\left(f(x) - y\right)$$

FOCs:

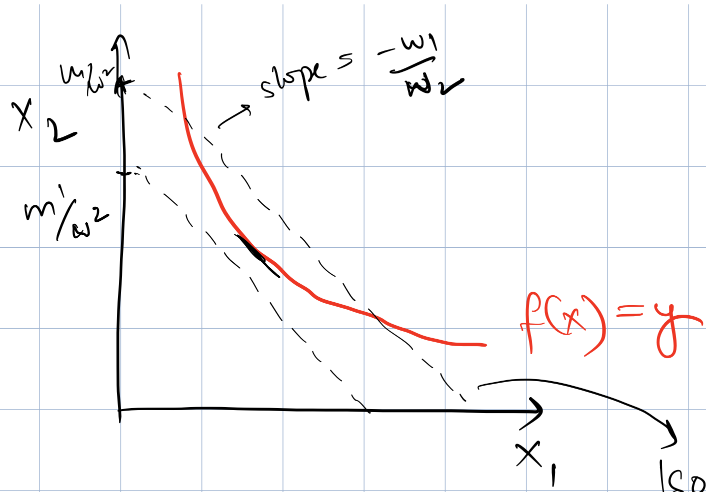
$$w_i - \lambda \frac{\partial f}{\partial x_i} = 0$$

or

$$w_i = \lambda \frac{\partial f}{\partial x_i} \quad \forall i$$

Divide i by j

$$\underbrace{\frac{w_i}{w_j}}_{\text{Economic rate of substitution}} = \underbrace{\frac{\partial f / \partial x_i}{\partial f / \partial x_j}}_{\boxed{\text{MRTS}}} \qquad (\text{for } x_i^*, x_j^* > 0)$$



Isocost line

Constraint: Isoquant

Objective: Isocost

$$w_1 x_1 + w_2 x_2 = m$$

$$x_2 = \frac{m}{w_2} - \frac{w_1}{w_2} x_1$$

The slope remains the same

Looking for tangency as the optimal point.

$x^*(w, y)$ conditional factor demand correspondence

$$c(w,y) = w \cdot x^*(w,y) = \min_{\substack{\text{s.t.} \\ x_i \geq 0 \\ f(x) \geq y}} w \cdot x$$

Envelope Theorem $\quad \frac{dV}{d\theta} = \frac{\partial \mathcal{L}}{\partial \theta}$

Here set $V = C$, $\theta = y$

$$\frac{\partial c}{\partial y} = \frac{\partial \mathcal{L}}{\partial y}$$

$$\boxed{\frac{\partial c}{\partial y} = \lambda}$$

$\Rightarrow$ λ is the marginal cost of increasing prodn by one unit.

"Shadow Prices" $\rightarrow$ Cost of tightening the constraint.
$[f(x) \geq y]$

Relationship with Profit Maximization Problem (PMP)

$$\max_{y \geq 0} \; py - c(w, y)$$

FOC: $p = \dfrac{\partial c}{\partial y} = $ (λ from CMP)

Things to keep in mind :-

- We have assumed a differentiable f, interior solutions.
- General FOCs :

$$\lambda \frac{\partial f}{\partial x_i}(x^*) - w_i \leq 0 \quad \forall i,$$

with equality if $x_i^* > 0$.

- KT conditions are necessary but not sufficient in general. They will be sufficient if $f(x)$ $\xrightarrow{\text{prodn } f^n}$ is concave.
- Existence / uniqueness issues.

Weak Axiom of Cost Minimization (WACM)

$\{(w^1, x^1, y^1), \ldots\ldots, (w^T, x^T, y^T)\}$

WACM: → what you did.

$\underbrace{w^t . x^t}_{\text{the cost you paid in a month}} \leq w^t . x^s$ → could have done for all $\underbrace{y^s \geq y^t}$.

Produced what you needed to produce in that month.

- Any firm that violates WACM is "irrational" / Are not minimizing costs.

[It should be less than any other cost plan for that month.]

If WACM does not hold ⇒ You are not cost minimizing.

Implication of WACM: — downward sloping demand.

Take $y^s = y^t$, WACM gives 2 inequalities

1. $w^t . x^t \leq w^t . x^s$
2. $w^s . x^s \leq w^s . x^t$

Add 1 & 2 $\quad w^t x^t + w^s x^s \leq w^t x^s + w^s x^t$

$\Rightarrow w^t x^t + w^s x^s - w^t x^s - w^s x^t \leq 0$

$\Rightarrow (w^t - w^s) . (x^t - x^s) \leq 0$

$\Delta w . \Delta x \leq 0$ (Talks about vector of price changes & not just a price change)

$$c(w,y) = \min_{x} \ w \cdot x$$
$$\text{s.t. } f(x) \geq y$$

① $c(w,y)$ is nondecreasing in w.

Proof: Take $w' \geq w$ & $x \in x^*(w,y)$, $x' \in x^*(w',y)$

$$c(w,y) = w \cdot x \underset{\text{WACM}}{\leq} w \cdot x' \underset{w' > w}{\leq} w' \cdot x' = c(w',y)$$

$$\Rightarrow \quad c(w,y) < c(w',y)$$

② $c(w,y)$ is nondecreasing in y

Let $y'' \geq y'$, $x' \in X^*(w,y')$, $x'' \in x^*(w,y'')$

$$c(w,y') = \min_{\substack{\text{s.t.} \\ f(x) \geq y'}} w \cdot x = w \cdot x' \underset{\text{WACM}}{\leq} w \cdot x''$$

$$= \min_{\substack{\text{s.t.} \\ f(x) \geq y''}} w \cdot x = c(w, y'')$$

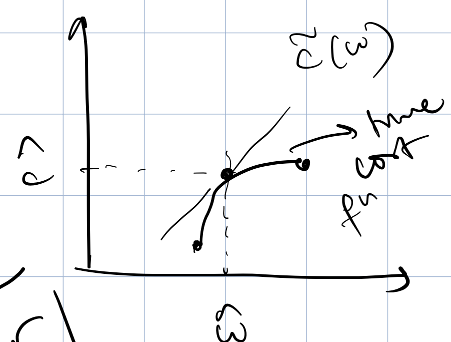
Inequality works because $f(x'') \geq y'' \geq y'$.

③ $c(w,y)$ is hd-1 in w.

$\iff c(tw,y) = tc(w,y)$

④ $c(w,y)$ is concave in w. NOT true for y.

(Proof is same as that of the π maximization being convex.)



When price increases, costs increase linearly & vice versa.

↓

don't let the firm respond.

When the firm responds, its cost would be equal or lower than the $\tilde{c}(w)$.

$\boxed{\text{H.W. for formal proof}}$

(5) If $v(y)$ is convex, then $x^*(w, y)$ is convex valued.
If $v(y)$ is strictly convex, then $x^*(w, y)$ is single-valued.

(6) $x^*(w, y)$ is hd-0 in w.

$$x^*(tw, y) = x^*(w, y)$$

→ Relative prices of the 2 goods are the same.

Implication :- (Looking for one good)!

$$x_i^*(t, w, y) = x_i^*(w, y)$$

Differentiate wrt t, eval @ $t = 1$.

$$\nabla_w x_i^*(tw, y) \cdot w = 0$$

$$\sum_{j=1}^{n} \frac{\partial x_i^*(tw, y)}{\partial w_j} \cdot w_j = 0$$

We know : $\dfrac{\partial x_i^*}{\partial w_i} \leq 0$

Say $n=2$

$$\frac{\partial x_1^*}{\partial w_1} w_1 + \frac{\partial x_2^*}{\partial w_2} w_2 = 0$$

negative

this is positive

$$\frac{\partial x_1^*}{\partial w_2} \geq 0 \quad \therefore \text{ goods 1 \& 2 are substitutes.}$$

If $n > 2$, then $\frac{\partial x_i^*}{\partial w_j} \geq 0$ for some good j.

Shephard's Lemma

If $c(w,y)$ is differentiable, and $\bar{w} > 0$, then:

$$x_i(\bar{w},y) = \frac{\partial c(\bar{w},y)}{\partial w_i}$$

Proof 1: Envelope Theorem

$$\mathcal{L} = w.x - \lambda(f(x)-y)$$

$$\frac{\partial c(w,y)}{\partial w_i} = \frac{\partial \mathcal{L}}{\partial w_i}\Bigg|_{x=x^*(w,y)}$$

$$\boxed{\frac{\partial c}{\partial w_i} = x_i^*(w,y)}$$

Proof 2: $\quad C(w, y) = w \cdot x^*(w, y)$

$$(*) \quad \frac{\partial c}{\partial w_i} = x_i^*(w, y) + \sum_{j=1}^{n} w_j \frac{\partial x_j^*(w, y)}{\partial w_i}$$

Recall the FOCs of CMP :—

$$(1) \quad \lambda \frac{\partial f}{\partial x_i} = w_i \qquad\qquad (2) \quad f(x^*(w, y)) = y$$

lets differentiate (2) wrt w_i :

$$\sum_{j=1}^{n} \frac{\partial f}{\partial x_j} \frac{\partial x_j^*}{\partial w_i} = 0$$

Let's plug (1) into (*)

$$\frac{\partial c}{\partial w_i} = x_i^*(w, y) + \sum_{j=1}^{n} \lambda \frac{\partial f}{\partial x_j} \frac{\partial x_j^*}{\partial w_i} = 0$$

$$\boxed{\frac{\partial c}{\partial w_i} = x_i^*(w, y)}$$

Substitution Matrix:

$$\begin{bmatrix} \frac{\partial x_i^*}{\partial w_i} & \cdots & \frac{\partial x_i^*}{\partial w_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n^*}{\partial w_1} & \cdots & \frac{\partial x_n^*}{\partial w_n} \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial^2 c}{\partial w_1^2} & \cdots & \frac{\partial^2 c}{\partial w_n \partial w_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 c}{\partial w_1 \partial w_n} & \cdots & \frac{\partial^2 c}{\partial w_n^2} \end{bmatrix}}_{\text{Hessian of } c \text{ which is concave.}}$$

* Law of Demand: $\frac{\partial x_i^*}{\partial w_i} \leq 0 \quad \forall i$ (Diagonal elements of Hessian matrix of concave f^n is negative!)

* Symmetric cross-price effects

$$\frac{\partial x_i^*}{\partial w_j} = \frac{\partial x_j^*}{\partial w_i}$$

Theorem: If f is CRS, then $c(w,y) = y \underbrace{c(w,1)}_{\text{unit cost of } f^n}$

Proof :—

Assume f is differentiable, interior solution

FOCs at $y=1$.

$$\frac{w_i}{w_j} = \frac{F_i(x^*)}{F_j(x^*)} \quad \text{and} \quad F(x^*) = 1$$

For general $\hat{y}$:

$$\frac{w_i}{w_j} = \frac{f_i(\hat{x})}{f_j(\hat{x})} \qquad f(\hat{x}) = \hat{y}$$

Consider inputs $\hat{x} = \hat{y}\, x^*$

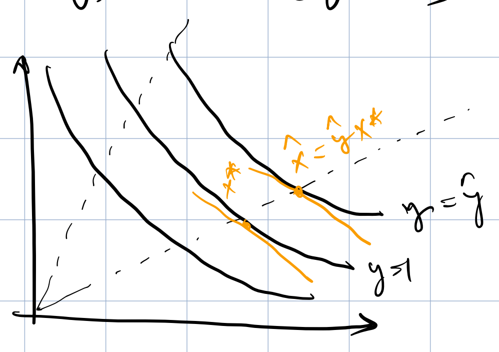
claim: $\hat{x}$ solves the FOCs at $\hat{y}$.

$$\frac{f_i(\hat{x})}{f_j(\hat{x})} = \frac{f_i(\hat{y}\, x^*)}{f_j(\hat{y}\, x^*)} \underset{\text{CRS}}{=} \frac{f_i(x^*)}{f_j(x^*)} = \frac{w_i}{w_j}$$

$$f(\hat{y}\, x^*) \underset{\text{by CRS}}{=} \hat{y}\, f(x^*) = \hat{y}$$

So, $\hat{x} = \hat{y}\, x^*$ solves the problem at target output $\hat{y}$.

$$c(w, \hat{y}) = w \cdot (\hat{y}\, x^*) = \hat{y}\,(w \cdot x^*) = \hat{y}\, c(w, 1)$$



The slope of the tangent lines are the same.