The general framework

The statistical/econometric procedures to be examined will involve constructions based on statistical models (collections of probability distributions) and the sample (a collection of random elements related to the aforementioned distributions).

⊖ those random elements will in most cases be random vectors or random matrices (eg. in the usual linear model

the sample Z_n is the random matrix

regressors

$$(y_n, X_n) = \begin{pmatrix} y_1 & x_{11} & x_{12} & \cdots & x_{1p} \\ y_2 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_n & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

dependent variable

)

$n \times (p+1)$

random variables

⊖ We will not get into the measure/probability theory details that formulate the notion of a random element (random variable, random vector, random matrix, random function, etc). However the reminder of their functionality could be useful:

> A random element with values in S is a function $X: \Omega \to S$, where:

* Ω is a (usually latent) non empty set equipped with a collection of subsets at which probabilities can be assigned in a "consistent" way, and (in our setting) a probability distribution that assigns probabilities to the above, say $\mathbb{P}$.

* S is a (usually non-latent) set that is usually equipped with a mathematical structure that discerns the collection of subsets that can be consistently assigned with probabilities.

* Then $X : \Omega \to S$ is an S-valued random element iff it additionally creates a correspondence between the subsets of Ω and the subsets of S at which probabilities can be consistently assigned.

* But the above then implies that X pushouts $\mathbb{P}$ from Ω to S, thus creating a probability distribution on S, say $\mathbb{P}^*$, with $\mathbb{P}^*(\cdot) = \mathbb{P}(X^{-1}(\cdot))$.

(i.e. the probability that $\mathbb{P}^*$ assigns to any appropriate subset of S, say A, equals the probability that $\mathbb{P}$ assigns to the subset of Ω formed by the elements of Ω that are mapped in A by X).

⊖ But what is S? (what is this extra structure that it carries?) In our investigations S will typically $\boxed{\text{have the structure of a metric space.}}$

Definition. A metric space (S, d) is a pair of a non-empty set S along with a metric (or distance function) d, which is any function $d: \underbrace{S \times S}_{\text{evaluated on pairs of elements of } S} \to \underbrace{\mathbb{R}}_{\text{assigns a real value that is interpretable as the distance between them}}$

evaluated on pairs of elements of S

assigns a real value that is interpretable as the distance between them that has the properties:

α. $\forall s, s' \in S$, $d(s,s') \geq 0$ (positive definite)

β. $d(s,s') = 0 \iff s = s'$ (separation)

γ. $\forall s, s' \in S$, $d(s,s') = d(s',s)$ (symmetry)

δ. $\forall s_1, s_2, s_3 \in S$, $d(s_1,s_2) \leq d(s_1,s_3) + d(s_3,s_2)$ (triangle inequality)

- d provides S (among others) with the means of exending notions encountered in real analysis (e.g. convergence, continuity, etc) there. Studying suchlike concepts in detail (Analysis on Metric Spaces) is obviously out of the scope of the course, but the examples and the notions below will be useful in what follows:

⊖ Some examples:

① $S = \mathbb{R}$, $d(s,s^*) = |s - s^*|$, $s, s^* \in \mathbb{R}$. → $\mathbb{R}$ with the usual metric

② $S = \mathbb{R}^p$, $d(s,s') = \left(\sum_{i=1}^{p} (s_i - s_i^*)^2\right)^{1/2}$, $s, s^* \in \mathbb{R}^p$ → Euclidean metric

$\left(s = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_p \end{pmatrix}, \ s_i \in \mathbb{R}, \ i = 1, \dots, p\right)$

Euclidean p-space

* Notice that $d(s,s') = (s - s^*)'(s - s^*)$

③ $S = M_{q \times p}$

↦ the set of $q \times p$ matrices with real entries

if $A, B \in M_{q \times p}$, $d(A,B) = \left(\sum_{i=1}^{q} \sum_{j=1}^{p} (A_{ij} - B_{ij})^2\right)^{1/2}$

where A_{ij} is the $(i,j)^{th}$ element of A. – Frobenius

(Can this be represented as a Euclidean-type metric?)

4 Suppose that Θ is a non empty set, and $f: \Theta \to \mathbb{R}$ a real function on Θ. f is bounded iff $\exists M > 0 : |f(\theta)| \leq M \ \forall \theta \in \Theta$.

(Why suchlike always exist?)

Let $B(\Theta, \mathbb{R})$ denote the set of bounded real functions on Θ. Then:

$$S = B(\Theta, \mathbb{R}) \text{ and if } f, g \in B(\Theta, \mathbb{R}),$$

$$d(f,g) = \sup_{\theta \in \Theta} |f(\theta) - g(\theta)|.$$

↳ Uniform metric

⊖ A quick overview of some useful elements of metric space analysis: (S,d) is thus a metric space

✱ If $s \in S$ and $\varepsilon > 0$, the open ball centered at s with radius ε, $B_\varepsilon(s) = \{s' \in S : d(s,s') < \varepsilon\}$.

The analogous closed ball $\bar{B}_\varepsilon(s) = \{s' \in S : d(s,s') \leq \varepsilon\}$.

It is easy to show that $B_\varepsilon(s) \neq \emptyset \ \forall s, \varepsilon$,

and $B_\varepsilon(s) \subseteq \bar{B}_\varepsilon(s) \ \forall \varepsilon, s$.

* A subset A of S is (d-) bounded iff $\exists s \in S, \varepsilon > 0$: $A \subseteq B_\varepsilon(s)$ (eq. $A \subseteq \bar{B}_\varepsilon(s)$).

- Easy example: balls are bounded.

* A subset A of S is (d-) open iff $\forall s \in A, \exists \varepsilon > 0$ (depends on s) : $B_\varepsilon(s) \subseteq A$.

- Easy example: open balls are open!

* A subset A of S is (d-) closed iff A^c is d-open

- Easy example: closed balls are closed!

[$\emptyset$, S are simultaneously (d-) open and closed]

* A sequence of elements of S, $(s_n)_{n \in \mathbb{N}}$ is a "vector" of the form $(s_0, s_1, \dots, s_n, \dots)$, $s_n \in S \ \forall n \in \mathbb{N}$

(equivalently a function $\mathbb{N} \to S$)

$(s_n)_{n\in\mathbb{N}}$ (d-) converges to $s \in S$ ($d\text{-}\lim s_n = s$)

iff $\forall \varepsilon > 0, \exists n^*(\varepsilon): s_n \in B_\varepsilon(s) \ \forall n \geq n^*(\varepsilon)$

iff $\boxed{\lim_{n\to+\infty} d(s_n, s) = 0}$

↳ a translation of the notion using the language of real analysis.

$*$ (S_1, d_1), (S_2, d_2) are metric spaces, and $f: S_1 \to S_2$. If $s_1 \in S_1$, f is (d_2/d_1)-continuous at s_1 iff $\forall (s_{1n})_{n\in\mathbb{N}}$, $s_{1n} \in S_1 \ \forall n \in \mathbb{N}$, with $(d_1\text{-})\lim s_{1n} = s_1$, $d_2\text{-}\lim f(s_{1n}) = f(s_1)$.

f is (d_2/d_1)-continuous iff it is (d_2/d_1)-continuous at s_1, $\forall s_1 \in S_1$.

↳ Again a translation using the language of real analysis.

In $\mathbb{R}^p$ a sequence of p vectors

$$\left(\begin{pmatrix} s_{11}\\ s_{21}\\ \vdots \\ s_{p1}\end{pmatrix}, \begin{pmatrix} s_{12}\\ s_{22}\\ \vdots\\ s_{p2}\end{pmatrix}, \dots, \begin{pmatrix} s_{1n}\\ s_{2n}\\ \vdots\\ s_{pn}\end{pmatrix}, \dots \right)$$

can be perceived as a "finite ordered list" of p sequences of real numbers $\left((s_{in})_{n\in\mathbb{N}}\right)_{i=1,\dots,p}$.
(real sequences)

Using this and the properties of the Euclidean metric (in connection to the usual metric in $\mathbb{R}$) it can be proven that convergence w.r.t. the Euclidean metric is equivalent to "pointwise" convergence of each of the real sequences involved and the limiting vector is the vector of the limits of the members of the ordered list.

The same is true in $M_{q\times p}$ and the Frobenius metric. (why?)

* Examples and further notions will be locally introduced when needed.

⊖ When S is equipped with a metric, its then produced open and closed subsets, in turn produce its subsets that are consistently assignable probabilities.

⊖ Thus in our considerations S will be equipped with a metric d, thus being a metric space; if $X: \Omega \to S$ then, an S-valued random element:

① X is a random variable when $S = \mathbb{R}$ (d is the usual metric)

② X is a random vector (p-dimensional) when $S = \mathbb{R}^p$ (d is the Euclidean metric)

③ X is a random $(q \times p)$ matrix when $S = M_{q \times p}$ (d is the Frobenius metric)

④ X is a random (bounded) real function (on Θ), when $S = B(\Theta, \mathbb{R})$ (d is the uniform metric).

⊖ It can be easily proven that a random p-vector is a p-vector of random variables, and a random $q \times p$ matrix, is a $q \times p$ matrix of random variables (also when $\Theta \subseteq \mathbb{R}^k$, then a random bounded real function on Θ, is a collection of random variables, one for each $\theta \in \Theta$.)

- During the course the sample will be mostly a random vector or a random matrix. Random functions will constitute statistical criteria (e.g. sum of squares, likelihood functions, etc.) → (relevant to optimization)
- Derivatives of such like function (when exist) will constitute random functions with values in $\mathbb{R}^p$ or in $M_{q \times p}$ (eg. Hessians).
- The relevant framework (e.g. elements of linear algebra, the $\mathcal{L}^2$ (least squares) geometry, elements of probability theory (e.g. probability distributions on $\mathbb{R}^p$, integration and moments, conditional expectations, etc)) are considered inherited from ECON850.