Intro to Econometrics: Recitation 3

Gustavo Pereira

September 23, 2019

Outline

- Review:
  - Statistical model
    - Definition
    - Examples
    - Identification, sufficiency
  - Statistical decision problem
    - Definition
    - Examples

Statistical model

Definition

- Idea: formalize statements such as
    1. Let $\{h_1, \ldots, h_{10}\}$ denote the outcome of 10 independent coin flips with probability p of landing heads

Statistical model

Definition

- Idea: formalize statements such as
  1. Let $\{h_1, \dots, h_{10}\}$ denote the outcome of 10 independent coin flips with probability p of landing heads
  2. "Let X_1, X_2, X_3 be iid uniform in $[0, \theta]$ where θ is an unknown positive real number"

Statistical model

Definition

- Idea: formalize statements such as
  1. Let $\{h_1, \ldots, h_{10}\}$ denote the outcome of 10 independent coin flips with probability p of landing heads
  2. "Let X_1, X_2, X_3 be iid uniform in $[0, \theta]$ where θ is an unknown positive real number"
  3. "Let $\{Y_t\}_{t \in 1,2,\ldots,T}$ be an AR(1) process with gaussian innovations"

Statistical model

Definition

- Idea: formalize statements such as
  1. Let $\{h_1, \ldots, h_{10}\}$ denote the outcome of 10 independent coin flips with probability p of landing heads
  2. "Let X_1, X_2, X_3 be iid uniform in $[0, \theta]$ where θ is an unknown positive real number"
  3. "Let $\{Y_t\}_{t \in 1,2,\ldots,T}$ be an AR(1) process with gaussian innovations"
- Claim. All statements equivalent to: "let $\mathbf{X}$ be a draw from some cdf $F : \mathbf{R}^S \to [0,1]$ where F is taken from some restricted set of CDFs,

$$F \in \mathfrak{F} \text{"}$$

Statistical model

Definition

- It's common to write

$$\mathfrak{F} = \{F_\theta\}_{\theta \in \Theta}$$

Statistical model

Definition

- It's common to write

$$\mathfrak{F} = \{F_\theta\}_{\theta \in \Theta}$$

- For example:

$$\mathfrak{F} = \{F : \mathbf{R} \to \mathbf{R} | F \text{ is the cdf of } U[a, b] \text{ for some } a \leq b\}$$

  - Does this represent a statistical model?

Statistical model

Definition

- It's common to write

$$\mathfrak{F} = \{F_\theta\}_{\theta \in \Theta}$$

- For example:

$$\mathfrak{F} = \{F : \mathbf{R} \to \mathbf{R} | F \text{ is the cdf of } U[a, b] \text{ for some } a \leq b\}$$

  - Does this represent a statistical model?

- We can define for $\theta = (a, b)$,

$$F_\theta = \frac{t - a}{b - a} \mathbf{1}_{[a,b]}(t)$$

Statistical model

Definition

- It's common to write

$$\mathfrak{F} = \{F_\theta\}_{\theta \in \Theta}$$

- For example:

$$\mathfrak{F} = \{F : \mathbf{R} \to \mathbf{R} | F \text{ is the cdf of } U[a, b] \text{ for some } a \leq b\}$$

  - Does this represent a statistical model?

- We can define for $\theta = (a, b)$,

$$F_\theta = \frac{t - a}{b - a} \mathbf{1}_{[a,b]}(t)$$

- With that indexing,

$$\mathfrak{F} = \{F_\theta\}_{\theta \in \Theta}$$

where $\Theta = \{(x, y) \in \mathbf{R}^2 : x \leq y\}$

Statistical model

Comment

- Why do we specify models with CDFs?

Statistical model

Comment

- Why do we specify models with CDFs?
- Reason: in Euclidean spaces, distribution of random variables is fully characterized by CDF

Statistical model

Comment

- Why do we specify models with CDFs?
- Reason: in Euclidean spaces, distribution of random variables is fully characterized by CDF
- However, if all CDFs in your model are absolutely continuous, it's equivalent to specify a family of PDFs

Statistical model

Comment

- Why do we specify models with CDFs?
- Reason: in Euclidean spaces, distribution of random variables is fully characterized by CDF
- However, if all CDFs in your model are absolutely continuous, it's equivalent to specify a family of PDFs
- In the course, we will do this interchangeably; if a model is specified in terms of PDFs, it's understood that we're considering only absolutely continuous distributions

Statistical model

Comment

- Why do we specify models with CDFs?
- Reason: in Euclidean spaces, distribution of random variables is fully characterized by CDF
- However, if all CDFs in your model are absolutely continuous, it's equivalent to specify a family of PDFs
- In the course, we will do this interchangeably; if a model is specified in terms of PDFs, it's understood that we're considering only absolutely continuous distributions
- We can also specify the model with more general probability distributions:

$$\{P_\theta : \mathcal{B}(\mathfrak{X}) \to [0,1]\}_{\theta \in \Theta}$$

  where $\mathfrak{X}$ a possibly more general space (e.g., a space of bounded continuous functions)

Statistical model

Example 1: ten coin flips

- Single coin flip:

$$F_p^1(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } x \in [0, 1) \\ 1 & \text{otherwise} \end{cases}$$

Statistical model

Example 1: ten coin flips

- Single coin flip:

$$F_p^1(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } x \in [0, 1) \\ 1 & \text{otherwise} \end{cases}$$

- Then the joint is $F_p(h_1, h_2, \ldots, h_{10}) = F_p^1(h_1) \cdots F_p^1(h_{10})$

Statistical model

Example 1: ten coin flips

- Single coin flip:

$$F_p^1(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } x \in [0, 1) \\ 1 & \text{otherwise} \end{cases}$$

- Then the joint is $F_p(h_1, h_2, \ldots, h_{10}) = F_p^1(h_1) \cdots F_p^1(h_{10})$
- Model:

$$\{F_p\}_{p \in [0,1]}$$

  - What is Θ ?

Statistical model

Example 2: Uniform $[0, \theta]$

- Three independent uniform $[0, \theta]$. We know that for a given θ

$$F_\theta^2(t) = \frac{t}{\theta}\mathbf{1}_{[0,\theta]}(t)$$

is the cdf of $U[0, \theta]$ for non-negative θ.

Statistical model

Example 2: Uniform $[0, \theta]$

- Three independent uniform $[0, \theta]$. We know that for a given θ

$$F_\theta^2(t) = \frac{t}{\theta}\mathbf{1}_{[0,\theta]}(t)$$

is the cdf of $U[0, \theta]$ for non-negative θ.

- Thus joint is

$$F_\theta(x_1, x_2, x_3) = F_\theta^2(x_1)F_\theta^2(x_2)F_\theta^2(x_3)$$

and statistical model is

$$\{F_\theta\}_{\theta>0}$$

Statistical model

Example 3: AR(1) with Gaussian innovations

- An "AR(1) with Gaussian innovations" means that

$$Y_t - \mu = \rho(Y_{t-1} - \mu) + \epsilon_t$$

  where ϵ_t are drawn iid $N(0, \sigma^2)$.

  - Note: need to make assumption about Y_0. Assume fixed.

Statistical model

Example 3: AR(1) with Gaussian innovations

- An "AR(1) with Gaussian innovations" means that

$$Y_t - \mu = \rho(Y_{t-1} - \mu) + \epsilon_t$$

  where ϵ_t are drawn iid $N(0, \sigma^2)$.

  - Note: need to make assumption about Y_0. Assume fixed.

- Equivalently,

$$Y_t | Y_{t-1}, \ldots, Y_1 \sim N(\mu + \rho(Y_{t-1} - \mu), \sigma^2)$$

Statistical model

Example 3: AR(1) with Gaussian innovations

- An "AR(1) with Gaussian innovations" means that

$$Y_t - \mu = \rho(Y_{t-1} - \mu) + \epsilon_t$$

where ϵ_t are drawn iid $N(0, \sigma^2)$.

  - Note: need to make assumption about Y_0. Assume fixed.

- Equivalently,

$$Y_t | Y_{t-1}, \ldots, Y_1 \sim N(\mu + \rho(Y_{t-1} - \mu), \sigma^2)$$

- How do you write the joint CDF? By what parameters will it be indexed?

Statistical model

Identification & sufficiency

- Summary of previous discussion: a statistical model is a family of distributions, $\{F_\theta : \mathbf{R}^S \to [0,1]\}_{\theta \in \Theta}$.

Statistical model

Identification & sufficiency

- Summary of previous discussion: a statistical model is a family of distributions, $\{F_\theta : \mathbf{R}^S \to [0,1]\}_{\theta \in \Theta}$.
- If each $\theta \in \Theta$ induces a unique distribution, the model is called identified.

Statistical model

Identification & sufficiency

- Summary of previous discussion: a statistical model is a family of distributions, $\{F_\theta : \mathbf{R}^S \to [0,1]\}_{\theta \in \Theta}$.
- If each $\theta \in \Theta$ induces a unique distribution, the model is called identified.
  - Mathematically: the model is identified iff for every $\theta \neq \theta'$, there exists $x \in \mathbf{R}^S$ such that $F_\theta(x) = F_{\theta'}(x)$

Statistical model

Identification & sufficiency

- Summary of previous discussion: a statistical model is a family of distributions, $\{F_\theta : \mathbf{R}^S \to [0,1]\}_{\theta \in \Theta}$.
- If each $\theta \in \Theta$ induces a unique distribution, the model is called identified.
  - Mathematically: the model is identified iff for every $\theta \neq \theta'$, there exists $x \in \mathbf{R}^S$ such that $F_\theta(x) = F_{\theta'}(x)$
  - What if the model was specified in terms of PDFs? What about general probability distributions?

Statistical model

Identification & sufficiency

- Summary of previous discussion: a statistical model is a family of distributions, $\{F_\theta : \mathbf{R}^S \to [0,1]\}_{\theta \in \Theta}$.
- If each $\theta \in \Theta$ induces a unique distribution, the model is called identified.
  - Mathematically: the model is identified iff for every $\theta \neq \theta'$, there exists $x \in \mathbf{R}^S$ such that $F_\theta(x) = F_{\theta'}(x)$
  - What if the model was specified in terms of PDFs? What about general probability distributions?
- A *statistic* is any function $T : \mathbf{R}^S \to \mathbf{R}^K$. We say that T is sufficient if

$$\mathbf{P}_\theta(\cdot | T(\cdot))$$

does not depend on θ. Intuitively, if you condition on $T(X)$, the full data become uninformative about θ.

Statistical model

Identification & sufficiency

- Example: let X_1 and X_2 be iid $N(\mu, 1)$.

Statistical model

Identification & sufficiency

- Example: let X_1 and X_2 be iid $N(\mu, 1)$.
    - Model here is $\{F_\mu\}_{\mu \in \mathbf{R}}$ where F_μ is cdf of independent joint normal with mean (μ, μ) and identity variance matrix
- Then $T(X_1, X_2) = X_1 + X_2$ is sufficient.

Statistical model

Identification & sufficiency

- Example: let X_1 and X_2 be iid $N(\mu, 1)$.
  - Model here is $\{F_\mu\}_{\mu \in \mathbf{R}}$ where F_μ is cdf of independent joint normal with mean (μ, μ) and identity variance matrix
- Then $T(X_1, X_2) = X_1 + X_2$ is sufficient.
- Before proof: note that crucially the data is 2 dimensional, but the sufficient statistic is 1d

Statistical model

Identification & sufficiency

- Example: let X_1 and X_2 be iid $N(\mu, 1)$.
  - Model here is $\{F_\mu\}_{\mu \in \mathbf{R}}$ where F_μ is cdf of independent joint normal with mean (μ, μ) and identity variance matrix
- Then $T(X_1, X_2) = X_1 + X_2$ is sufficient.
- Before proof: note that crucially the data is 2 dimensional, but the sufficient statistic is 1d
- Now:

$$\begin{bmatrix} X_1 \\ X_2 \\ T(X_1, X_2) \end{bmatrix} \sim \mathcal{N}_3\left(\begin{bmatrix} \mu \\ \mu \\ 2\mu \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}\right)$$

Statistical model

Identification & sufficiency

- Example: let X_1 and X_2 be iid $N(\mu, 1)$.
  - Model here is $\{F_\mu\}_{\mu \in \mathbf{R}}$ where F_μ is cdf of independent joint normal with mean (μ, μ) and identity variance matrix
- Then $T(X_1, X_2) = X_1 + X_2$ is sufficient.
- Before proof: note that crucially the data is 2 dimensional, but the sufficient statistic is 1d
- Now:

$$\begin{bmatrix} X_1 \\ X_2 \\ T(X_1, X_2) \end{bmatrix} \sim \mathcal{N}_3 \left(\begin{bmatrix} \mu \\ \mu \\ 2\mu \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix} \right)$$

- To find conditional distribution of X_1 and X_2 given $T(X_1, X_2)$, use the BLP trick.

Statistical model

Identification & sufficiency

- Math:

$$E[X_1|X_1+X_2] = E[X_2|X_1+X_2] = \frac{X_1+X_2}{2}$$

  moreover, conditional variance also doesn't depend on μ

  - Statistical decision problem
    - Definition
    - Examples

Statistical decision problem
Definition

- Definition: statistical decision problem is

$$(\Theta, A, \mathcal{L}, \{F_\theta\}_{\theta \in \Theta})$$

where

Statistical decision problem

Definition

- Definition: statistical decision problem is

$$(\Theta, A, \mathcal{L}, \{F_\theta\}_{\theta \in \Theta})$$

where

1. Θ is a parameter space

Statistical decision problem

Definition

- Definition: statistical decision problem is

$$(\Theta, A, \mathcal{L}, \{F_\theta\}_{\theta \in \Theta})$$

where

1. Θ is a parameter space
2. A is a space of actions

Statistical decision problem

Definition

- Definition: statistical decision problem is

$$(\Theta, A, \mathcal{L}, \{F_\theta\}_{\theta \in \Theta})$$

where

1. Θ is a parameter space
2. A is a space of actions
3. $\mathcal{L}$ is a utility/loss function

Statistical decision problem

Definition

- Definition: statistical decision problem is

$$(\Theta, A, \mathcal{L}, \{F_\theta\}_{\theta\in\Theta})$$

where

1. Θ is a parameter space
2. A is a space of actions
3. $\mathcal{L}$ is a utility/loss function
4. $\{F_\theta\}$ is a statistical model
    - ⋆ Remember: this can be alternatively specified as $\{P_\theta\}_{\theta\in\Theta}$ or $\{f_\theta\}_{\theta\in\Theta}$

Statistical decision problem

Interpretation

- Statistician is supposed to decide something. Examples:

Statistical decision problem
Interpretation

- Statistician is supposed to decide something. Examples:
  1. Pick the θ that she thinks generated the data

$$A = \Theta$$

Statistical decision problem

Interpretation

- Statistician is supposed to decide something. Examples:
  1. Pick the θ that she thinks generated the data
  $$A = \Theta$$
  2. Given a split $\Theta = \Theta_0 \sqcup \Theta_1$, pick which of Θ_0 or Θ_1 is more likely to contain the parameter that generated data
  $$A = \{0, 1\}$$

Statistical decision problem

Interpretation

- Statistician is supposed to decide something. Examples:
  1. Pick the θ that she thinks generated the data
  $$A = \Theta$$
  2. Given a split $\Theta = \Theta_0 \sqcup \Theta_1$, pick which of Θ_0 or Θ_1 is more likely to contain the parameter that generated data
  $$A = \{0, 1\}$$
  3. Pick a subset $C \subseteq \Theta$ where she thinks the true θ falls in
  $$A = \text{reasonable subsets of } \Theta$$

Statistical decision problem

Interpretation

- Model this as a sequential game.

Statistical decision problem

Interpretation

- Model this as a sequential game.
  - *First stage:* Nature picks $\theta \in \Theta$. This is not observable by statistician

Statistical decision problem

Interpretation

- Model this as a sequential game.
  - *First stage:* Nature picks $\theta \in \Theta$. This is not observable by statistician
  - *Stage* $1\frac{1}{2}$*:* Nature randomly draws $X \sim F_\theta$

Statistical decision problem

Interpretation

- Model this as a sequential game.
  - *First stage:* Nature picks $\theta \in \Theta$. This is not observable by statistician
  - *Stage* $1\frac{1}{2}$*:* Nature randomly draws $X \sim F_\theta$
  - *Second stage:* Statistician chooses action a

Statistical decision problem

Interpretation

- Model this as a sequential game.
  - *First stage:* Nature picks $\theta \in \Theta$. This is not observable by statistician
  - *Stage* $1\frac{1}{2}$*:* Nature randomly draws $X \sim F_\theta$
  - *Second stage:* Statistician chooses action a
- At the terminal nodes, statistician gets the loss $\mathcal{L}(a, \theta)$

Statistical decision problem

Interpretation

- Model this as a sequential game.
  - *First stage:* Nature picks $\theta \in \Theta$. This is not observable by statistician
  - *Stage* $1\frac{1}{2}$*:* Nature randomly draws $X \sim F_\theta$
  - *Second stage:* Statistician chooses action a
- At the terminal nodes, statistician gets the loss $\mathcal{L}(a, \theta)$
- Let $\mathfrak{X} \subset \mathbf{R}^S$ denote the (common) support of F_θ. A strategy for the statistician in this game is a function

$$d : \mathfrak{X} \to A$$

I.e. a specification of an action for every possible decision node she faces

This strategy is called a decision rule in the mathematical statistical jargon

Statistical decision problem

Risk function

- What sort of criterion should we use to rank decision rules?

Statistical decision problem

Risk function

- What sort of criterion should we use to rank decision rules?
- We use the expected utility paradigm. For fixed θ, we postulate that

$$d_1(\cdot) \precsim_\theta d_2(\cdot) \iff \mathbf{E}_\theta \left[\mathcal{L}(d_1(X), \theta)\right] \geq \mathbf{E}_\theta \left[\mathcal{L}(d_2(X), \theta)\right]$$

  - With respect to what are we taking the expectation?

Statistical decision problem

Risk function

- What sort of criterion should we use to rank decision rules?
- We use the expected utility paradigm. For fixed θ, we postulate that

$$d_1(\cdot) \precsim_\theta d_2(\cdot) \iff \mathbf{E}_\theta\left[\mathcal{L}(d_1(X), \theta)\right] \geq \mathbf{E}_\theta\left[\mathcal{L}(d_2(X), \theta)\right]$$

  - With respect to what are we taking the expectation?
- This expectation is called *risk*. Notation:

$$R(d, \theta) := \mathbf{E}_\theta\left[\mathcal{L}(d(X, \theta)\right] = \int_{\mathbf{R}^S} d(x, \theta) dF_\theta(x)$$

Statistical decision problem

Risk function

- What sort of criterion should we use to rank decision rules?
- We use the expected utility paradigm. For fixed θ, we postulate that

$$d_1(\cdot) \precsim_\theta d_2(\cdot) \iff \mathbf{E}_\theta\left[\mathcal{L}(d_1(X),\theta)\right] \geq \mathbf{E}_\theta\left[\mathcal{L}(d_2(X),\theta)\right]$$

  - With respect to what are we taking the expectation?
- This expectation is called *risk*. Notation:

$$R(d,\theta) := \mathbf{E}_\theta\left[\mathcal{L}(d(X,\theta)\right] = \int_{\mathbf{R}^S} d(x,\theta) dF_\theta(x)$$

- Analogy with game theory: *dominated* strategies
  - A decision rule that is not weakly dominated is called admissible