ECON 7710

Econometrics I

Lecture notes 4.

Statistical decision problem:

- Parameter inference ($\theta \in \Theta \subset \mathbb{R}^p$) is input to statistical decisions
- Statistical decision: function $d = \delta(x)$ (x is observation)
- Consequence of decision is a loss $L(\theta, d) \in \mathbb{R}_+$
- Focus on long-term average loss $E[L(\theta, \delta(X))]$
- **Example:**
  - $X = (X_1, \ldots, X_n)$, $X_i \sim N(\xi, \sigma^2)$
  - Decision: select interval $[\underline{L}, \overline{L}]$
  - Loss function: 0 if decision is correct and depends on the distance of the true ξ from the interval otherwise
  - Loss minimization problem: minimize loss w.r.t. the lendth of the interval
- Introduce risk function
$$R(\theta, \delta) = E_\theta[L(\theta, \delta(X))]$$
- Choosing optimal decision is impossible since θ is unknown
- One approach to optimality is to choose decisions that provide uniformly smaller risk

Testing, basic definitions:

- Define the problem as accepting or rejecting a certain hypothesis regarding P_θ
- Decision problem if called *a test* of the hypothesis

- Parameter space generates a class of distributions $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$
- Then $\mathcal{P} = H \cup K$ (where hypothesis is true for H and not true for K) with $\Theta = \Theta_H \cup \Theta_K$
- The hypothesis mathematically is a statement: $P_\theta \in H$
- K is the class of alternatives
- d_0 is the decision to not reject H and d_1 is the decision to reject H
- If $X \in S$ is the r.v. corresponding to the sample, then $S = S_0 \cup S_1$, s.t. H is not rejected if $x \in S_0$ and rejected if $x \in S_1$ (x is the realization of X)
- S_0 is called acceptance region and S_1 is called the rejection region (or *critical region*)
- Two types of errors:
  1. Type I: Rejecting H when it is true
  2. Type II: Not rejecting H when it is false
- Desirable property: both errors are small, but this is impossible
- **Definition:** Level of significance is a number $0 < \alpha < 1$ such that

$$P_\theta(\delta(X) = d_1) = P_\theta(X \in S_1) \leq \alpha, \quad \forall\, \theta \in \Theta_H$$

- Test is constructed by maximizing $P_\theta(\delta(X) = d_1) = P_\theta(X \in S_1)$ (the power of the test) over $\theta \in \Theta_K$ setting the significance level to fixed α
- **Definition:** Quantity

$$\sup_{\theta \in \Theta_H} P_\theta(X \in S_1)$$

  is called the size of the test.
- **Definition:** Probability $\beta(\theta) = P_\theta(\delta(X) = d_1)$ considered as a function of θ is called the power function of the test
- Randomized test: $\phi(x)$ is the probability of rejection

- $\phi(x)$ is the critical function; for deterministic tests it is indicator for critical region
- Construction of test: select $\phi(\cdot)$ to maximize

$$\beta_\phi(\theta) = E_\theta[\phi(X)], \ \forall \theta \in \Theta_K,$$

subject to

$$E_\theta[\phi(X)] \le \alpha, \ \forall \theta \in \Theta_H.$$

- Test that maximized power against a certain alternative may depend on this alternative
- Finding tests that maintain power against all alternatives will be very important

Testing, fundamental results:

- **Defnition:** Class of distributions is called *simple* if it contains a single distribution, otherwise it is called *composite*.
- Consider simple hypothesis K, and P_0 are distributions under K and P_1 are distributions under H
  - Test design is based on solving

$$\max_S \sum_{x \in S} P_1(x)$$

  subject to

$$\sum_{x \in S} P_0(x) \le \alpha.$$

  - Solution chooses x with the highest ratio $r(x) = \frac{P_1(x)}{P_0(x)}$
  - Formally S is the set $r(x) > c$ where

$$P_0(X \in S) = \sum_{r(x) > c} P_0(x) = \alpha.$$

- **Theorem (Neyman-Pearson fundamental lemma):** Let P_1 and P_0 be probability distributions with densities p_1 and p_0.

(i) *Existence:* For testing $H : p_0$ against the alternative $K : p_1$ there exists a test $\phi(\cdot)$ and a constant k, s.t.

$$E_0[\phi(X)] = \alpha \qquad (*)$$

and

$$\phi(x) = \begin{cases} 1, \text{ when } p_1(x) > k\, p_0(x), \\ 0, \text{ when } p_1(x) < k\, p_0(x). \end{cases} \qquad (**)$$

(ii) *Sufficient condition for most powerful test.* If the test satisfies (*) and (**) for some k then it is most powerful for testing p_0 against p_1 at level α.

(iii) *Necessary condition for most powerful test.* If ϕ is most powerful at level α for testing p_0 against p_1, then for some k it satisfies (**) almost everywhere. It also satisfies (*) unless there exists a test of size $< \alpha$ with power 1.

- **Corollary:** Let β denote the power of the most powerful level α test for testing P_0 against P_1. Then $\alpha < \beta$ unless $P_0 = P - 1$

- **Example:** $X \sim N(\mu, 1)$. Test the null $\mu = 0$ against the alternative $\mu = \mu' > 0$.
  - Likelihood ratio
  $$\frac{p_1(x)}{p_0(x)} = \frac{e^{-\frac{(x-\mu')^2}{2}}}{e^{-\frac{x^2}{2}}} = e^{x\mu' - \frac{1}{2}(\mu')^2}.$$
  - $p_1(x)/p_0(x)$ is monotone increasing in x
  - $p_1(x)/p_0(x) > k$ corresponds to $x > c$ for some c
  - Set $P_0(X > c) = \alpha$, meaning that $c = z_{1-\alpha}$ ($1 - \alpha$ quantile of standard normal distribution).

p-values:

- Suppose that $p_1(x)/p_0(x)$ is continuous, then the most powerful level α test is non-randomized and rejects if $p_1(x)/p_0(x) > k(\alpha)$

- Varying α generates nested critical regions with $S_\alpha \subset S_{\alpha'}$ for $\alpha < \alpha'$

- **Definition:** The smallest significance level at which the hypothesis would be rejected given observation

$$\widehat{p} = \widehat{p}(x) = \inf\{\alpha : x \in S_\alpha\}$$

is called *p-value*.

- **Example (continued):** The rejection region

$$S_\alpha = \{x : x > z_{1-\alpha}\} = \{x : 1 - \Phi(x) < \alpha\}.$$

  - $\widehat{p} = 1 - \Phi(x)$.

- **Lemma:** Suppose that X has distribution P_θ for some $\theta \in \Theta$ and H specifies $\theta \in \Theta_H$. Suppose that for rejection region $S_\alpha \subset S_{\alpha'}$ if $\alpha < \alpha'$.

  (i) If for $\theta \in \Theta_H$

  $$P_\theta(X \in S_\alpha) \leq \alpha, \quad 0 < \alpha < 1,$$

  then the distribution of $\widehat{p}$ under $\theta \in \Theta_H$ satisfies

  $$P_\theta(\widehat{p} \leq u) \leq u \ \ \forall\, 0 \leq u \leq 1.$$

  (ii) If for $\theta \in \Theta_H$

  $$P_\theta(X \in S_\alpha) = \alpha, \quad 0 < \alpha < 1,$$

  then the distribution of $\widehat{p}$ under $\theta \in \Theta_H$ satisfies

  $$P_\theta(\widehat{p} \leq u) = u \ \ \forall\, 0 \leq u \leq 1,$$

  i.e. $\widehat{p}$ is uniformly distributed over $[0, 1]$.

Uniformly most powerful tests:

- **Definition:** The real-parameter distribution family $\{p_\theta(x), \quad \theta \in \Theta\}$ is said to have monotone likelihood ratio if there exists a real-valued function $T(x)$ such that for any $\theta < \theta'$ the distributions P_θ and $P_{\theta'}$ are distinct and the ratio $p_{\theta'}(x)/p_\theta(x)$ is a non-decreasing function of $T(x)$.

- **Theorem:** Let θ be a real parameter and let random variable X have probability density $p_\theta(x)$ with monotone likelihood ratio in $T(x)$.

  (i) For testing $H : \theta \le \theta_0$ against $K : \theta > \theta_0$, there exists a UMP (uniformly most powerful) test, which is given by

  $$\phi(x) = \begin{cases} 1, & \text{when } T(x) > C, \\ \gamma, & \text{when } T(x) = C, \\ 0, & \text{when } T(x) < C, \end{cases}$$

  where C and γ are determined by

  $$E_{\theta_0}[\phi(X)] = \alpha.$$

  (ii) The power function

  $$\beta(\theta) = E_\theta[\phi(X)]$$

  of this test is strictly increasing for all points θ for which $0 < \beta(\theta) < 1$.

  (iii) For all θ' the test determined in (i) is UMP for testing $H' : \theta \le \theta'$ against $K' : \theta > \theta'$ at level $\alpha = \beta(\theta')$.

  (iv) For any $\theta < \theta_0$ the test minimizes $\beta(\theta)$ among all tests satisfying $E_{\theta_0}[\phi(X)] = \alpha$.

- **Example:** Hypergeometric distribution

  $$P_D(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$$

  - Note that $\max\{0, n+D-N\} \le x \le \min\{n, D\}$
  - Likelihood ratio is
    $$\frac{P_{D+1}(x)}{P_D(x)} = \frac{D+1}{N-D}\,\frac{N-D-n+x}{D+1-x}$$
    if $n + D + 1 - N \le x \le D$ (and 0 or $+\infty$ at the end points)
  - $T(x) = x$ and there exists UMP for $H : D \le D_0$ against $K : D > D_0$ that rejects H when x is too large.

Confidence bounds:

- An object of interest in many cases is the upper and lower bound for a given parameter (or sets of parameters)
- wlog we consider the lowe bound
- $\underline{\theta}(x)$ is a function of observations
- We choose the confidence level $1-\alpha$ such that

$$P_\theta(\underline{\theta}(X) \leq \theta) \geq 1-\alpha, \ \ \forall\theta.$$

- **Definition:** Function $\underline{\theta}(x)$ is called a lower confidence bound for θ at confidence level $1-\alpha$
- The construction of $\underline{\theta}(\cdot)$ should minimize the probability $P_\theta(\underline{\theta}(X) \leq \theta')$ for any $\theta' < \theta$ (which will give the uniformly most accurate lower confidence bound for θ).
- **Definition:** A family of subsets $S(x) \subset \Theta$ is said to constitute a family of confidence sets at confidence level $1-\alpha$ if

$$P_\theta(\theta \in S(X)) \geq 1-\alpha, \ \ \forall\, \theta \in \Theta$$

- A lower confidence bound is a special case of the confidence set $S(x) = \{\theta \,:\, \underline{\theta}(x) \leq \theta < \infty\}$
- **Theorem:**
  (i) For each $\theta_0 \in \Theta$ let $A(\theta_0)$ be the acceptance region of the α-level test for testing $H(\theta_0) \,:\, \theta = \theta_0$ and for each sample point x let $S(x)$ denote the set of parameter values

$$S(x) = \{\theta \,:\, x \in A(\theta),\, \theta \in \Theta\}.$$

  Then $S(x)$ is a family of confidence sets for θ at confidence level $1-\alpha$.

  (ii) If for all θ_0, $A(\theta_0)$ is UMP for testing $H(\theta_0)$ at level α against the alternative $K(\theta_0)$, then for each $\theta_0 \in \Theta$, $S(x)$ minimizs probability

$$P_{\theta_0}(\theta_0 \in S(X)) \ \ \forall\, \theta \in K(\theta_0)$$

  among all level $1-\alpha$ families of confidence sets for θ.

- **Corollary:** Let the family of densities $p_\theta(x)$ have mootone likelihood ratio in $T(x)$ and suppose that the cumulative distribution function $F_\theta(t)$ of $T = T(X)$ is a continuous function in each of the variables t and θ when the other is fixed.

  (i) There exists a uniformly most accurate confidence bound $\underline{\theta}$ for θ at each confidence level $1 - \alpha$

  (ii) If x denotes the observed values of X and t the observed values of $T(X)$ and if the equation

  $$F_\theta(t) = 1 - \alpha$$

  has a solution $\theta = \widehat{\theta}$ in Θ then this solution is unique and $\underline{\theta}(x) = \widehat{\theta}$.