Mathematical Statistics II

MAT 5191 Lecture Notes

University of Ottawa

Hanan Ather

Winter 2022

Contents

These are my notes for MAT 5191: Mathematical Statistics II. These notes are *my* interpretation of the content covered in class. They are by no means comprehensive; my primary aim was to distill and summarize the important topics discussed in each lecture.

§1 January 11, 2022

§1.1 Introduction

In this course we will cover important topics of Mathematical Statistics. This course covers methods of hypothesis testing thoery and interval estimation in the context of **Parametric Statistics** and **Classical Non parametric Statistics**.

In a typical statistical problem our objective is to get information about the distribution **P** of a random variable X based on n independent observations $X_1, ..., X_n$ of $\boldsymbol{X}$.

Definition 1.1 (Random Sample)

Random variable is a function defined as

$$X(S, \mathbb{P}) \to \mathcal{X}$$

Where S is the sample space, and $\mathbb{P}$ is a probability measure. **Random sample** is defined as direct product of

$$S \times \cdots \times S \to \underbrace{(\mathcal{X} \times \cdots \times \mathcal{X})}_{\text{sample space}}$$

Sample space is just a set of all possible values of a random sample. In general for simplicity we assume that $(\mathcal{X} \times \cdots \times \mathcal{X}) = \mathbb{R}^n$.

A **hypothesis** is a statement regarding the parameter of the distribution or distribution itself. The two complementary hypothesis in a hypothesis testing problem are called the *null hypothesis* (H_0) and *alternative hypothesis* (H_1).

Example 1.2. We may consider testing the hypothesis of symmetry of a cdf $F(x)$ about zero:

$$H_0 : F_1(x) = F_2(x), \quad \text{for all } x \in \mathbb{R},$$

Based on two independent samples $X_1, .., X_n$ and $Y_1, ..., Y_m$ from continuous distributions with cdf's F_1 and F_2, respectively.

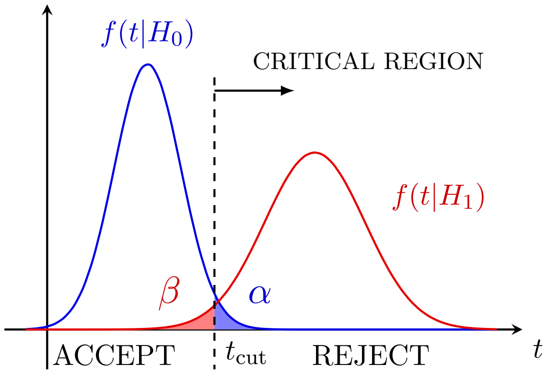
Definition 1.3 (Hypothesis test)

A hypothesis test is a statement regarding the parameter of the distribution or the distribution itself. A **nonrandomized test function** assigns to each value $\boldsymbol{x} = (x_1, ..., x_1)$ of $\boldsymbol{X} = (X_1, ..., X_1)$ of the the following decisions:

$$\phi(\boldsymbol{X}) = \begin{cases} 0 & \text{if } \boldsymbol{x} \in C \\ 1 & \text{if } \boldsymbol{x} \in A, \end{cases} \tag{1.1}$$

i For which sample values the decision is made to accept H_0 as true.

ii For which sample values H_0 is rejected and H_1 is accepted as true.



The subset of the sample space for which H_0 will be rejected is called the **critical region** (or rejection region). The complement of the rejection region is called **acceptance region**.

In hypothesis testing we can commit two types of errors: reject H_0 when H_0 is true (*type I error*) or to accept H_0 when H_0 is actually false (*type II error*).

To minimize the probability of both type I and type II error we impose the following asymmetry between both types of error: we select a small number $\alpha \in (0,1)$, called **level of significance**, and impose the condition that

$$\mathbb{P}(\text{type I error}) = \mathbb{P}_\theta(X \in C) \le \alpha, \quad \text{for all } \theta \in \Theta_0.$$

Subject to this condition, we minimize,

$$\mathbb{P}(\text{type II error}) = \mathbb{P}_\theta(X \in A), \quad \text{for all } \theta \in \Theta_1,$$

or equivalently we maximize

$$1 - \mathbb{P}(\text{type II error}) = \mathbb{P}_\theta(X \in A), \quad \text{for all } \theta \in \Theta_1. \tag{1.2}$$

The quantity,

$$\sup_{\theta \in \Theta_0} \mathbb{P}(\boldsymbol{X} \in C)$$

is called the *size* of the test with critical region C. With this approach the decision maker believes that the consequence of wrongly rejecting H_0 is more severe than the decision of wrongly accepting it, and therefore the size of the test is kept at a small level. The probability in 1.2 is called the **power** of the test against the alternative.

Definition 1.4 (Power Function)
Considered as a function of θ for all $\theta \in \Theta$, the probability

$$\beta(\theta) = \mathbb{P}_\theta(\boldsymbol{X} \in C), \quad \theta \in \Theta,$$

is called the **power function** of the test with critical region C.

We note that based on 1.3, for non-randomized test ϕ with critical region C, we have

$$\beta_\phi(\theta) = \mathbb{P}_\theta(\boldsymbol{X} \in C) = 1 \cdot \mathbb{P}_\theta(\boldsymbol{X} \in C) + 0 \cdot \mathbb{P}_\theta(\boldsymbol{X} \in A) = \mathbb{E}_\theta[\phi(\boldsymbol{X})]$$

Definition 1.5 (Confidence set)

Let $\boldsymbol{X} = (X_1, ..., X_n)$ be a random sample from a distribution $\boldsymbol{P} \in \mathcal{P} = \{\boldsymbol{P}_\theta : \theta \in \Theta \subseteq \mathbb{R}^k\}$. A random set $S(\boldsymbol{X})$ is said to constitute a **confidence set** for θ of level $(1-\alpha)$ if

$$\boldsymbol{P}_\theta(\theta \in S(\boldsymbol{X})) \geq 1 - \alpha, \quad \text{for all } \theta \in \Theta$$

§2 January 13, 2022

§2.1 Neyman-Pearson Fundamental Lemma

- typically the test ϕ that maximizes power against a particular alternative depends on the *alternative*.
- there is an important exception: when the alternative is simple, $\Theta = \{\theta_1\}$, the problem is completely specified by

$$\max_\phi \beta_\phi(\theta) = \max_\phi \mathbb{E}_\theta[\phi(\boldsymbol{X})]$$

  subject to the condition

$$\mathbb{E}_\theta[\phi(\boldsymbol{X})] \leq \alpha \quad \text{for all } \theta \in \Theta_0$$

- this maximization problem reduces to the mathematical problem of maximizing an integral (or sum) subject to some conditions. The solution to this problem is called the *most powerful (MP) test of level alpha*

Definition 2.1 (Uniformly most powerful test)

A level α test which maximizes power among all tests of level α is said to be **uniformly most powerful (UMP)** level α test. Thus, ϕ is UMP level α test if:

(i) $\sup_{\theta \in \Theta_0} \beta_\phi(\theta) = \alpha$

(ii) for any other test ϕ^* which satisfies (i) has $\beta_\phi(\theta) \geq \beta_{\phi^*}(\theta)\ \forall \theta \in \Theta_1$

Theorem 2.2 (The Neyman-Pearson Fundamental Lemma) Let $\boldsymbol{X} = (X_1, ..., X_n)$ be a random sample from a probability distribution $\boldsymbol{P_\theta}$ with pdf/pmf $f(x;\theta)$, $\theta \in \Theta = \{\theta_0, \theta_1\}$. Suppose that we are interested in testing two simple hypothesis $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ at level α.

(a) For testing H_0 versus H_1 there exists a test ϕ and a constant k such that

$$\mathbb{E}_{\theta_0}\phi(\boldsymbol{X}) = \alpha \tag{2.1}$$

and

$$\phi(\boldsymbol{X}) = \begin{cases} 1 & \text{if } f(\boldsymbol{x};\theta_1) > kf(\boldsymbol{x};\theta_0) \\ 0 & \text{if } f(\boldsymbol{x};\theta_1) < kf(\boldsymbol{x};\theta_0) \end{cases} \tag{2.2}$$

(b) If a test satisfies 2.1 and 2.2 for some k, then it is a UMP level α test.

(c) If ϕ is the most powerful at level $alpha$ for testing H_0 versus H_1, then for some k it satisfies 2.2. It also satisfies 2.1 unless there exists a test of size $< \alpha$ with power 1.

§2.2 Geometric Interpretation of Neyman-Pearson Lemma

If we consider the set

$$B = \left\{(\alpha, \beta) \in [0,1]^2 : \text{there exists a test } \phi \text{ such that } \alpha = \mathbb{E}_{\theta_0}[\phi(\boldsymbol{X})], \beta = \mathbb{E}_{\theta_1}[\phi(\boldsymbol{X})]\right\}$$

It can be shown that the set B is:

(a) convex;

(b) contains the points $(0,1)$ and $(1,1)$

(c) symmetric about the point $(1/2, 1/2)$ in the sense that if $(\alpha, \beta) \in B$ then the point $(1-\alpha, 1-\beta)$ also belongs to B

(d) closed.

§3 January 18, 2022

§3.1 Examples of Most Powerful tests

Example 3.1. Let $X_1, .., X_n$ be a random sample from normal $N(\mu, \sigma^2)$ distribution, where μ is unknown and σ^2 is known. If we test

$$H_0 : \mu = 0 \text{ vs. } H_1 : \mu = \mu_0$$

for some $\mu_0 > 0$. The likelihood ratio is equal to

$$\frac{f(\boldsymbol{x}; \mu_0)}{f(\boldsymbol{x}; 0} = \frac{exp(\frac{1}{-2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2)}{exp(\frac{1}{-2\sigma^2}\sum_{i=1}^{n} x_i^2)}$$

From this and Theorem 2.2, the critical region $\left\{\boldsymbol{x} : \frac{f(\boldsymbol{x};\mu_0)}{f(\boldsymbol{x};0} > k\right\}$ of the most powerful level α test is equivalent to the region $\{\boldsymbol{x} : \sum_{i=1}^{n} x_i > k'\}$, where the constant k' satisfies the level α constraint,

$$\mathbb{P}_{H_0}\left(\sum_{i=1}^{n} X_i \geq k' = \alpha\right)$$

We note that under H_0, $\sum_{i=1}^{n} X_i \sim N(0, n\sigma^2)$,

$$\mathbb{P}_{H_0}\left(N(0,1) \geq \frac{k'}{\sigma\sqrt{n}}\right) = \alpha$$

or,

$$1 - \Phi\left(\frac{k'}{\sigma\sqrt{n}}\right) = \alpha$$

or,

$$k' = \sigma\sqrt{n}\Phi^{-1}(1-\alpha).$$

Thus, for the observed value $\boldsymbol{x} = (x_1, ..., x_n)$ of $\boldsymbol{X} = (X_1, ..., X_n)$, the MP level α test rejects H_0 in favour of H_1 if

$$\sum_{i=1}^{n} x_i > \sigma\sqrt{n}\Phi^{-1}(1-\alpha) \iff \bar{x} > \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha)$$

§3.2 P-values

Definition 3.2 (p-value)
A p-value $p(\boldsymbol{X})$ is a test statistic satisfying $0 \leq p(\boldsymbol{x}) \leq 1$ for every sample point $\boldsymbol{x}$. The **p-value** or observed size is defined by

$$p = p(\boldsymbol{X}) = inf\{\alpha \in (0,1) : \boldsymbol{X} \in C_\alpha\}$$

where C_α is the critical region of level α test.

Lemma 3.3
Let $\boldsymbol{X} = (X_1, .., X_n)$ be a random sample of probability distribution $\mathbb{P}_\theta$, $\theta \in \Theta$. Consider $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta$

§4 January 20, 2022

§4.1 UMP tests: distributions with monotone likelihood ratio

- The cases that both null and alternative hypothesis are simple is mainly a theoretical situation. In most practical applications hypothesis are composite.

- Suppose that $\boldsymbol{X} = (X_1, ..., X_n)$ is a random sample from a probability distribution $\boldsymbol{P}_\theta$, where $\theta \in \Theta \subseteq \mathbb{R}$, that is θ is a *real-valued parameter* and suppose we wish to test

  $$H_0 : \theta \leq \theta_0 \ \ vs. \ \ H_1 : \theta > \theta_0,$$

  where θ_0 is a given number from the parameter space.

The MP test depends on the value of $\theta \in (\theta_0, \infty) =: \Theta_1$, and in general, is then not UMP.

Definition 4.1
A family of pdf's/pmf's $\{f(x;\theta) : \theta \in \Theta \subseteq \mathbb{R}\}$ is said to be **monotone likelihood ratio** (MLR) in the statistic $T(\boldsymbol{X})$ if there exists a function $T : \mathbb{R}^n \to \mathbb{R}$ such that whenever $\theta_1, \theta_2\Theta$ with $\theta_1 < \theta_2$, $\dfrac{f(\boldsymbol{x};\theta_2)}{f(\boldsymbol{x};\theta_1)}$ is a nondecreasing function of $T(\boldsymbol{X})$ on the set $\{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x};\theta_1) > 0 \text{ or } f(\boldsymbol{x};\theta_2) > 0\}$

Let $\boldsymbol{X} = (X_1, ..., X_n)$ be a random sample from a probability distribution $\boldsymbol{P}_{\boldsymbol{\theta}}$ with pdf/pmf $f(x;\theta), \theta \in \Theta \subseteq \mathbb{R}$, and let the family $\{f(x;\theta) : \theta \in \Theta\}$ have the MLR in $T(\boldsymbol{X})$.

(1) For testing $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$, there exists a UMP test level of α, which is given by

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{if } T(\boldsymbol{x}) > k \\ \gamma, & \text{if } T(\boldsymbol{x}) = k \\ 0 & \text{if } T(\boldsymbol{x}) < k, \end{cases} \tag{4.1}$$

where k and $\gamma \in (0, 1)$ are determined by

$$\mathbb{E}_{\theta_0}[\phi(\boldsymbol{X})] = \alpha \tag{4.2}$$

(2) The power function $\beta(\theta) = \mathbb{E}_{\theta_0}[\phi(\boldsymbol{X})]$, of this test is *strictly increasing* for all points θ for with $0 < \beta(\theta) < 1$.

(3) For all $\theta' \in \Theta$, the test determined by 4.1 and 4.2 is UMP for testing $H_0 : \theta \leq \theta'$ vs. $H_1 : \theta > \theta$ at level $\alpha = \beta(\theta')$.

(4) For any $\theta < \theta_0$, the test determined by 4.1 and 4.2 minimizes $\beta(\theta)$

By interchanging the inequalities we obtain a solution to the dual problem of testing $H_0 : \theta \geq \theta_0$ vs. $H_1 : \theta < \theta_0$ by the following level α test

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{if } T(\boldsymbol{x}) < k \\ \gamma, & \text{if } T(\boldsymbol{x}) = k \\ 0, & \text{if } T(\boldsymbol{x}) > k, \end{cases} \tag{4.3}$$

where k and γ are determined by

$$\mathbb{E}_{\theta_0}\phi(\boldsymbol{X}) = \alpha \tag{4.4}$$

§5 January 25, 2022

§6 January 27, 2022

§6.1 UMP tests for composite hypothesis

Previous lecture we proved that UMP exists for testing the hypothesis

$$H_0 : \theta \leq \theta_1 \text{ or } \theta \geq \theta_2 \text{ vs. } H_1 : \theta_1 \leq \theta \leq \theta_2 \tag{6.1}$$

if the family of pdf/pmf belong to a certain exponential family with a strictly monotone function $Q(\theta)$. We have also seen that no UMP exists for the hypothesis $H_0 : \theta = \theta_1$ vs $H_1 : \theta \neq \theta_1$. Similarly we run into the same problem if we interchange H_0 and H_1 in equation (5.1).

Definition 6.1 (Unbiased test)
Let $\boldsymbol{X} = (X_1, ..., X_n)$ be a random sample from probability distribution with pdf/pmf $f(x;\theta), \theta \in \Theta \in \mathbb{R}, k \geq 1$. Then for testing

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1$$

at level α a test ϕ based on $\boldsymbol{X}$ is said to be **unbiased** if

$$\mathbb{E}_\theta \phi(\boldsymbol{X}) \leq \alpha \text{ for all } \theta \in \Theta_0 \text{ and } \mathbb{E}_\theta \phi(\boldsymbol{X}) \geq \alpha \text{ for all } \theta \in \Theta_1$$

In other words, a test is **unbiased** if the probability of *type I error* is at most α and the *power* of the test is at least α.

Definition 6.2 (UMPU)
A test is **uniformly most powerful unbiased** (UMPU) if it is the UMP within the class of all unbiased tests.

The next condition provides us with conditions for when UMPU exists.

Theorem 6.3 Let $\boldsymbol{X} = (X_1, ..., X_n)$ be a random sample from probability distribution with pdf/pmf $f(x;\theta), \theta \in \Theta \in \mathbb{R}, k \geq 1$, where $f(x;\theta)$ is given by

$$f(x;\theta) = c(\theta)h(x)e^{Q(\theta)t(x)}, x \in \mathbb{R}, \theta \in \Theta$$

Then for testing,

$$H_0 : \theta_1 \leq \theta \leq \theta_2 \text{ vs. } H_1 : \theta \leq \theta_1 \text{ or } \theta \geq \theta_2$$

at level α, there exists a UMPU test which is given by

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } T(\boldsymbol{x}) \leq c_1 \text{ or } T(\boldsymbol{x}) \geq c_2 \\ \gamma & \text{if } T(\boldsymbol{x}) = c_i, i = 1, 2, c_1 < c_2 \\ 0 & \text{otherwise} \end{cases}$$

§7 Feb 1, 2022

§7.1 Bayesian tests

The Bayesian approach assumes that θ is a realization of a random variable θ with prior distribution π on Θ. The prior distribution reflects our opinion about the parameter θ. Let ϕ be a test for testing H_0 versus H_1, where C is the critical region, and A is the acceptence region.

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in C \\ 0 & \text{if } \boldsymbol{x} \in A \end{cases} \tag{7.1}$$

Consider the **loss function**

$$L(\theta,\phi)=\begin{cases}0 & \text{if } \theta\in\Theta_0 \text{ and } \phi=0 \text{ or } \theta\in\Theta_1 \text{ and } \phi=1\\ L_1 & \text{if } \theta\in\Theta_0 \text{ and } \phi=1\\ L_2 & \text{if } \theta\in\Theta_1 \text{ and } \phi=0,\end{cases}$$

where L_1 and L_2 are some positive constants. The **risk function** or the expected loss over all values of $\boldsymbol{X}$ is

$$R(\theta,\phi)=L(\theta,1)\mathbb{P}_\theta(\boldsymbol{X}\in C)+L(\theta,0)\mathbb{P}_\theta(\boldsymbol{X}\in A) \tag{7.2}$$

A test ϕ for which equation (7.2) is minimized is desirable. Consider the test

$$H_0:\theta=\theta_0 \;\; vs. \;\; H_1:\theta=\theta_1$$

where $\Theta_0=\{\theta_0\}$ and $\Theta_1=\{\theta_1\}$. Let $\alpha=\mathbb{P}_\theta(\boldsymbol{x}\in C)$ (*type I error*). And $\beta=\mathbb{P}_{\theta_1}(\boldsymbol{x}\in C)$. Now we can express the risk function in (7.2) in the following form:

$$R(\theta,\phi)=\begin{cases}L_1\alpha, & \text{if } \theta=\theta_0(\textit{type I error})\\ L_2(1-\beta), & \text{if } \theta=\theta_1(\textit{type II error})\end{cases} \tag{7.3}$$

For a prior distribution π on Θ,

$$p_0=\pi(\theta=\theta_0),\quad p_1=\pi(\theta=\theta_1)$$

The the **Bayes risk** of ϕ with respect to prior π is defined by

$$R_\pi(\phi)=p_0R(\theta_0,\phi)+p_1R(\theta,\phi)$$

where $R(\theta,\phi)$ is given by (7.3). A test ϕ for which the Bayes risk is minimal is called the **Bayes test** with respect to prior π. The following theorem is an analogue of the Neyman-Pearson Lemma.

Theorem 7.1 Let $\boldsymbol{X}=(X_1,...,X_n)$ be a random sample from probability distribution with pdf/pmf $f(x;\theta),\theta\in\Theta=\{\theta_0,\theta_1\}\subseteq\mathbb{R}^k,k\geq 1$. Let π be a prior distribution on Θ and let

$$p_0=\pi(\theta=\theta_0),\quad p_1=\pi(\theta=\theta_1)$$

For testing $H_0:\theta=\theta_0$ *vs.* $H_1:\theta=\theta_1$, there exists a *Bayes test* ϕ_π corresponding to the prior π which minimizes the *Bayes risk* $R_\pi(\phi)$. The test is given by

$$\phi_\pi(\boldsymbol{x})=\begin{cases}1, & \text{if } f(\boldsymbol{x};\theta_1)>\frac{p_0L_1}{p_1L_1}f(\boldsymbol{x},\theta_0)\\ 0, & \text{otherwise}\end{cases}$$

As we can see in 7.1, the Bayes test ϕ_π, is a likelihood ratio test and is the most powerful test for testing $H_0:\theta=\theta_0$ *vs.* $H_1:\theta=\theta_1$ at level $\mathbb{P}_{\theta_0}(\boldsymbol{X}\in C)$, where the‘ critical region C is given by

$$C=\left\{\boldsymbol{x}\in R^n:f(\boldsymbol{x};\theta_1)>\frac{p_0L_1}{p_1L_1}f(\boldsymbol{x},\theta_0)\right\},$$

which follows from the Neyman-Pearson Lemma.

§7.2 Minimax tests

Suppose $\boldsymbol{X} = (X_1, ..., X_n)$ be a random sample from probability distribution with pdf/pmf $f(x;\theta), \theta \in \Theta =\subseteq \mathbb{R}^k, k \geq 1$ and consider testing $H_0 : \theta \in \Theta_0$ *vs.* $H_1 : \theta \in \Theta_1$. And let $L(\theta, \phi)$ be the same loss function as before,

$$L(\theta, \phi) = \begin{cases} 0 & \text{if } \theta \in \Theta_0 \text{ and } \phi = 0 \text{ or } \theta \in \Theta_1 \text{ and } \phi = 1 \\ L_1 & \text{if } \theta \in \Theta_0 \text{ and } \phi = 1 \\ L_2 & \text{if } \theta \in \Theta_1 \text{ and } \phi = 0, \end{cases}$$

Consider the risk same risk function (7.2).

Definition 7.2 (Minimax test)
The test ϕ for testing $H_0 : \theta = \theta_0$ *vs.* $H_1 : \theta = \theta_1$ is called the **minimax test** if for any other test ϕ^* one has

$$\max(R(\theta_0, \phi), R(\theta_1, \phi)) \leq \max(R(\theta_0, \phi^*), R(\theta_1, \phi^*))$$

We have the following result regarding the existence of minimax test,

Theorem 7.3 (Existence of Minimax test) Let $\boldsymbol{X} = (X_1, ..., X_n)$ be a random sample from probability distribution with pdf/pmf $f(x;\theta), \theta \in \Theta = \{\theta_0, \theta_1\} \subseteq \mathbb{R}^k, k \geq 1$. Consider testing $H_0 : \theta = \theta_0$ *vs.* $H_1 : \theta = \theta_1$ at level α. Define the subset C of the sample space $\mathbb{R}^n$ as follows

$$C = \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}; \theta_1) > cf(\boldsymbol{x}, \theta_0)\}$$

and assume there is a determination of the constant c such that

$$R(\theta_0, \phi) = R(\theta_1, \phi) \quad \text{or equivalently,} \quad L_1 \mathbb{P}_{\theta_0}(\boldsymbol{X} \in C) = L_2 \mathbb{P}_{\theta_1}(\boldsymbol{X} \in A).$$

Then the test

$$\phi(\boldsymbol{x}) \begin{cases} 1 & \text{if } \boldsymbol{x} \text{ in } C \\ 0 & \text{otherwise} \end{cases}$$

is minimax test.