Inference 2, 2023, lecture 10

Rolf Larsson

November 29, 2023

Today

Chap. 5. Testing hypotheses

- Test problems
- P value
- Decision rules
- The Neyman-Pearson test and lemma

Test problems

- Sample $\mathbf{X} = (X_1, ..., X_n)$, statistical model $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$.
- Let $\Theta = \Theta_0 \cup \Theta_1$ with $\Theta_0 \cap \Theta_1 = \varnothing$.
- Test H_0: $\theta \in \Theta_0$ vs H_1: $\theta \in \Theta_1$.
- Two-sided test: H_0: $\theta = \theta_0$, H_1: $\theta \neq \theta_0$.
- One-sided tests for $\Theta \subseteq \mathcal{R}$:
  - Test H_0: $\theta \geq \theta_0$ vs H_1: $\theta < \theta_0$
  - Test H_0: $\theta \leq \theta_0$ vs H_1: $\theta > \theta_0$
- Alternatively:
  - For $\Theta = (-\infty, \theta_0]$, test H_0: $\theta = \theta_0$ vs H_1: $\theta < \theta_0$.
  - For $\Theta = [\theta_0, \infty)$, test H_0: $\theta = \theta_0$ vs H_1: $\theta > \theta_0$.

P value

- Base the hypothesis test on a suitable **test statistic** $T(\mathbf{X})$.
- Suppose that H_0 is true.
- If the observed value $T(\mathbf{x}) = t_{\text{obs}}$ is "improbable" ("extreme"), then we have evidence against H_0.

Definition (5.1)

The **p value** corresponding to an observed $T(\mathbf{x}) = t_{\text{obs}}$ is the probability of $T(\mathbf{X})$ lying at or beyond t_{obs} in the "direction of the more extreme values" of the alternative, computed from the null distribution.

P value

Example 1:

- Let $\mathbf{X} = (X_1, ..., X_n)$ where the X_i are independent uniform on $[0, \theta]$.
- We have $n = 4$ and observations $\mathbf{x} = (0.74, 1.99, 0.42, 1.08)$.
- Let $\Theta = [2, \infty)$.
- Test H_0: $\theta = 2$ vs H_1: $\theta > 2$.
- Take $T(\mathbf{X}) = \max_i X_i$.
- Compute the p value.

P value

- For one-sided tests, i.e.
  - test H_0: $\theta \geq \theta_0$ vs H_1: $\theta < \theta_0$,
  - or test H_0: $\theta \leq \theta_0$ vs H_1: $\theta > \theta_0$,

  the P value is given by either $\mathrm{P}_0^T(T \leq t_{\mathrm{obs}})$ or $\mathrm{P}_0^T(T \geq t_{\mathrm{obs}})$, depending on the direction of the alternative and the test statistic T.
- For the two-sided test: H_0: $\theta = \theta_0$, H_1: $\theta \neq \theta_0$, the p value is (most often) given by

$$2 \min\{\mathrm{P}_0^T(T \leq t_{\mathrm{obs}}),\ \mathrm{P}_0^T(T \geq t_{\mathrm{obs}})\}.$$

P value

- Suppose that the test statistic T has a continuous distribution.
- Suppose the p value is given by $\mathrm{P}_0(T \leq t_{\mathrm{obs}})$, where P_0 is the probability calculated under H_0.
- The p value may be regarded as a random variable which is uniformly distributed on $[0,1]$. (Why?)

Decision rules

We now shift focus from the parameter space Θ to the sample space $\mathcal{X}$.

Definition (5.2)

A (nonrandomized) **test** is a statistic $\varphi : \mathcal{X} \to \{0, 1\}$ such that

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad \mathbf{x} \in C_1 \quad \text{(reject } H_0\text{)} \\ 0 & \text{if} \quad \mathbf{x} \in C_0 \quad \text{(do not reject } H_0\text{)} \end{cases}$$

where $\mathcal{X} = C_1 \cup C_0$, with $C_1 \cap C_0 = \varnothing$.
C_1 is called the **critical region**.

Consider example 1. If we want to reject with probability 0.05 under H_0, which is the critical region?

Decision rules

Example 2:

- Let $\mathbf{X} = (X_1, ..., X_n)$ where the X_i are independent discrete uniform on $[1, 2, ..., \theta]$.
- We have $n = 4$ and observations $\mathbf{x} = (52, 99, 35, 12)$.
- Let $\Theta = [100, \infty)$.
- Consider the test H_0: $\theta = 100$ vs H_1: $\theta > 100$.

1. Calculate the p value.
2. If we want to reject with probability *at most* 0.05 under H_0, which is the critical region?

Decision rules

Definition (5.3)

A **randomized test** φ is a step function on C_1, $C_=$ and C_0 to $[0, 1]$, where $\mathcal{X} = C_1 \cup C_= \cup C_0$ and C_1, $C_=$ and C_0 are mutually disjunct:

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad \mathbf{x} \in C_1 \quad (\text{reject } H_0) \\ \gamma & \text{if} \quad \mathbf{x} \in C_= \quad (\text{reject } H_0 \text{ with probablility } \gamma) \\ 0 & \text{if} \quad \mathbf{x} \in C_0 \quad (\text{do not reject } H_0) \end{cases}$$

Observe: Randomized tests are only of theoretical interest! They are never performed in practice. What would be the main problem if they were?

Example 2':
What should $C_=$ and γ be in example 2 if we want to reject with probability 0.05 under H_0?

Decision rules

Definition (5.4)

For the (randomized) test problem H_0: $\theta \in \Theta_0$ vs H_1: $\theta \in \Theta_1$ we define

1. The **error of type I**: reject H_0, but $\theta \in \Theta_0$.
2. The **error of type II**: do not reject H_0, but $\theta \in \Theta_1$.
3. The function defined on Θ by

$$\pi(\theta) = \mathrm{E}_\theta\{\varphi(\mathbf{X})\} = \mathrm{P}_\theta(C_1) + \gamma \mathrm{P}_\theta(C_=)$$

is called the **power function** of the test φ.

Note:

- The power function gives the probability of rejecting H_0 when the parameter equals θ.
- If $\theta \in \Theta_1$, this is one minus the probability of a type II error for this particular θ.

Decision rules

Calculate the power for

1. $\theta = 3$ in example 1.
2. $\theta = 150$ in example 2.

The Neyman-Pearson test

- Consider testing a simple H_0: $\theta = \theta_0$ vs a simple H_1: $\theta = \theta_1$.
- In other words, we test H_0: P_0 vs H_1: P_1.

Definition (5.5)

For testing a simple H_0 vs a simple H_1,
the **size** of the type I error is $\alpha = \mathrm{P}_0(\textit{reject } H_0)$.
The size of the type II error is $\beta = \mathrm{P}_1(\textit{do not reject } H_0)$.

In example 1 with H_0: $\theta = 2$, H_1: $\theta = 3$, calculate α and β when the critical regions are

1. $C_1 = \{t_{\text{obs}} > 1.975\}$
2. $C_1 = \{t_{\text{obs}} > 1.9\}$

The Neyman-Pearson test

The set of all possible $\{\alpha(\varphi), \beta(\varphi)\}$, where φ is a test, is called the set of $\boldsymbol{\alpha\beta}$**-representations**.

Theorem (5.1)

The set of $\alpha\beta$-representations

1. *is convex,*
2. *is included in the closed unit square,*
3. *includes the points $(0, 1)$ and $(1, 0)$.*

The Neyman-Pearson test

We want to find a test φ which is optimal in the sense that, given a pre-assigned **significance level** $\alpha = \mathrm{E}_0(\varphi)$, the power $1 - \beta = \mathrm{E}_1(\varphi)$ is as large as possible.

Definition (5.6)

A test φ^* is called **most powerful** (MP) of size α if.f. $\mathrm{E}_0\{\varphi^*(\mathbf{X})\} = \alpha$ and $\mathrm{E}_1\{\varphi^*(\mathbf{X})\} \geq \mathrm{E}_1\{\varphi(\mathbf{X})\}$ for all tests φ with $\mathrm{E}_0\{\varphi(\mathbf{X})\} \leq \alpha$.

The Neyman-Pearson test

Definition (5.7)

For real numbers $k \geq 0$ and $\gamma \in [0, 1]$, tests of the form

$$\varphi(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad p_0(\mathbf{x}) < kp_1(\mathbf{x}), \\ \gamma & \text{if} \quad p_0(\mathbf{x}) = kp_1(\mathbf{x}), \\ 0 & \text{if} \quad p_0(\mathbf{x}) > kp_1(\mathbf{x}) \end{cases}$$

are called **Neyman-Pearson (NP) tests**.

Note: For non randomized tests, this corresponds to rejecting if.f. **the likelihood ratio**

$$\Lambda^*(\mathbf{x}) = \frac{p_0(\mathbf{x})}{p_1(\mathbf{x})} < k.$$

The NP test is also called the likelihood ratio (LR) test.

The Neyman-Pearson test

The Neyman-Pearson lemma tells us that NP tests, given the significance level (size), are optimal in terms of power.

Theorem (5.2)

The Neyman-Pearson lemma.
For all $\alpha \in [0,1]$ there exist $\gamma = \gamma(\alpha)$ and $k = k(\alpha)$ such that the NP test with $\gamma(\alpha)$, $k(\alpha)$ has size α and is most powerful of size α.
Especially, for $\alpha > 0$,

1. *If there exists a $k = k(\alpha)$ such that $\mathrm{P}_0\{p_0(\mathbf{X}) < kp_1(\mathbf{X})\} = \alpha$, then $\gamma = \gamma(\alpha) = 0$. (The non randomized case.)*
2. *Otherwise take $k = k(\alpha)$ with $\mathrm{P}_0\{p_0(\mathbf{X}) < kp_1(\mathbf{X})\} < \alpha < \mathrm{P}_0\{p_0(\mathbf{X}) \leq kp_1(\mathbf{X})\}$ and*

$$\gamma = \gamma(\alpha) = \frac{\alpha - \mathrm{P}_0\{p_0(\mathbf{X}) < kp_1(\mathbf{X})\}}{\mathrm{P}_0\{p_0(\mathbf{X})) = kp_1(\mathbf{X})\}}.$$

News of today

- Tests, one-sided and two-sided
- P value
- In the continuous case, the P value can be seen as a uniformly distributed random variable.
- Test (function)
  - non randomized (continuous case)
  - randomized (discrete case)
- Critical region
- Errors of type I and II
- Power function
- Testing a one point H_0 vs a one point H_1:
  - The set of $\alpha\beta$ representations.
  - Most powerful test
  - Neyman-Pearson (likelihood ratio) tests
  - ...are optimal in terms of power **(Neyman-Pearson lemma)**