Topic V:
Hypothesis Testing

Wei You

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Fall, 2023

Introduction

What is a hypothesis? It is a scientific claim:

- Smoking damages health.
- The coin is not fair.

What is a statistical hypothesis? It is a mathematically precise hypothesis.

- The life expectancy for non-smoking people has distribution F and that for smoking people is G. Then $\mu_F - \mu_G > 0$.
- The coin has Bernoulli distribution with head probability p. Then $p \neq 0.5$.

Statistical Hypothesis

A statistical hypothesis is a hypothesis that is testable on the basis of observing a sample of data.

Null and Alternative Hypotheses

In hypothesis testing, the goal is to see if there is sufficient statistical evidence to reject a presumed null hypothesis in favor of a conjectured alternative hypothesis.

- We observe the random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$, i.e. the "evidence".
- Null hypothesis H_0: some common belief that I want to disprove.
  Example: smoking does not affect health.
- Alternative hypothesis H_1: something I suspect.
  Example: smoking affects health.

Consider a family of distribution parametrized by θ, the general format is

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1$$

where Θ_0 is some subset of the parameter space and $\Theta_0 \cap \Theta_1 = \emptyset$. Usually $\Theta_1 = \Theta_0^c$.

Types of Hypotheses

- **Simple hypothesis**: hypotheses that specify only one possible distribution

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1.$$

- **Composite hypothesis**: hypotheses that specify more than one possible distribution

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

Example: One-sided hypotheses:

$$H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0$$
$$H_0 : \theta \geq \theta_0 \quad \text{versus} \quad H_1 : \theta < \theta_0$$

Example: Two-sided hypotheses:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0$$

Hypothesis Testing

Because the sample $\boldsymbol{X}$ is the only available statistical evidence, our decision has to be based on $\boldsymbol{X}$.

Definition (Nonrandomized test)

A nonrandomized test of H_1 versus H_2 is specified by a critical region R with the convention that we reject H_0 when $X \in R$ and do not reject H_0 when $X \in R$.

In other words, a (nonrandomized) test φ is a statistic $\varphi(\boldsymbol{X})$ mapping to $\{0, 1\}$, with 1 indicates rejection. So that the rejection region or critical region R is

$$R = \{\boldsymbol{X} : \varphi(\boldsymbol{X}) = 1\}.$$

Evaluating a Test

Say we have a test now for the hypothesis H_0 and its alternative H_1.

$$\text{not reject } H_0 \text{ if } \boldsymbol{X} \notin R$$
$$\text{reject } H_0 \text{ if } \boldsymbol{X} \in R$$

- How can we say that this test is good, i.e., the critical region R is reasonable?
- We want it to make fewer false judgments:
  - Type I error: H_0 true, reject
  - Type II error: H_0 wrong, not reject

	Not reject H_0	Reject H_0
H_0 true	Correct	Type I error
H_0 wrong	Type II error	Correct

Probability of Errors

A test is assessed by the probabilities of making two types of error:

- If $H_0 : \theta \in \Theta_0$ is true, then

$$\alpha(\theta) = \mathbb{E}_\theta[\varphi(\boldsymbol{X})] = \mathbb{P}_\theta(\varphi(\boldsymbol{X}) = 1) = \mathbb{P}_\theta(\boldsymbol{X} \in R)$$

  is the probability of a type I error.
  - If H_0 is simple, then it is a single probability.
  - If H_0 is composite, then there is a set of probabilities.
- If $H_0 : \theta \in \Theta_1$ is true, then

$$1 - \beta(\theta) = 1 - \mathbb{E}_\theta[\varphi(\boldsymbol{X})] = \mathbb{P}_\theta(\varphi(\boldsymbol{X}) = 0) = \mathbb{P}_\theta(\boldsymbol{X} \notin R)$$

  is the probability of a type II error.

We want both errors to be small.

- Probability of type I error rate $\mathbb{P}_{\theta\in\Theta_0}(\boldsymbol{X} \in R)$, so smaller means a smaller R
- Probability of type II error rate $\mathbb{P}_{\theta\in\Theta_1}(\boldsymbol{X} \in R^c)$, so smaller means a larger R.
- So the two errors cannot be minimized simultaneously.
  Example: In an extreme case, we always reject H_0, so R is the whole sample space. No type II error but huge type I error.

Evaluating a Test

- In the mathematical formulation for hypothesis testing just presented, the hypotheses H_0 and H_1 have a symmetric role.
- But in applications H_0 generally represents the status quo, or what someone would believe about θ without compelling evidence to the contrary.
- In view of this, attention is often focused on tests that have a small chance of error when H_0 is correct (i.e., falsely reject H_0, type I error).

Significance level

The level of significance of the test is defined as

$$\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\boldsymbol{X} \in R).$$

In words, the level α is the worst chance of falsely rejecting H_0.

Power

Recall that the probability of type II error is

$$1 - \beta(\theta) = 1 - \mathbb{E}_\theta[\varphi(\boldsymbol{X})] = \mathbb{P}_\theta(\varphi(\boldsymbol{X}) = 0) = \mathbb{P}_\theta(\boldsymbol{X} \notin R).$$

Power

For $\theta \in \Theta_1$, the power is defined as

$$\beta(\theta) \equiv \mathbb{E}_\theta[\varphi(\boldsymbol{X})] = \mathbb{P}_\theta(\boldsymbol{X} \in R).$$

The power is the ability to reject when the alternative hypothesis is correct.

The power depends on the specific value of θ. If the alternative hypothesis is composite, then there is a set of probabilities.

A test φ_1 is uniformly more powerful than a test φ_2, if

$$\mathbb{E}_\theta[\varphi_1(\boldsymbol{X})] \geq \mathbb{E}_\theta[\varphi_2(\boldsymbol{X})]$$

for every $\theta \in \Theta_1$.

So we want a test to have small significance level and be powerful.

As we have explained, they cannot be achieved simultaneously.

We control α (say 0.05), and try to find the uniformly most powerful test at level α.

- The choice of α is usually somewhat subjective. In some applications, the tolerance for type-I error is much tighter, e.g. in criminal-justice systems (presumption of innocence).

Power Function

The significant level and the power can be conveniently summarized into a single function called the power function, defined as the chance of rejecting H_0 as a function of $\theta \in \Theta$.

Definition (Power Function)

$$\beta(\theta) \equiv \mathbb{E}_\theta[\varphi(\boldsymbol{X})] = \mathbb{P}_\theta(\mathbf{X} \in R) = \begin{cases} \text{probability of type I error} & \text{if } \theta \in \Theta_0, \\ \text{1-probability of type II error} & \text{if } \theta \in \Theta_0^c. \end{cases}$$

- $\beta(\theta) = \alpha(\theta)$ is the type I error when $\theta \in \Theta_0$.
- $1 - \beta(\theta)$ is the type II error when $\theta \in \Theta_0^c$.
- $\alpha = \max\{\beta(\theta) : \theta \in \Theta_0\}$ is the size of the test. The smaller the better.
- $\beta(\theta)$ is the power of the test when $\theta \in \Theta_0^c$. The larger the better.

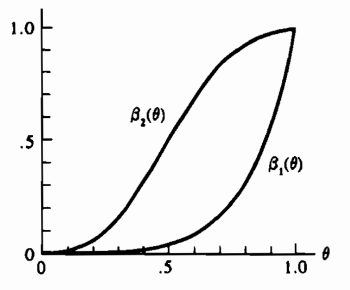
Example: Binomial power function. Let $X \sim$Binomial$(5, \theta)$.
Consider testing $H_0 : \theta \leq 0.5$ versus $H_1 : \theta > 0.5$.

Test 1 rejects H_0 if and only if $X = 5$ ($R = \{5\}$).

$$\beta_1(\theta) = \mathbb{P}_\theta(X \in R) = \mathbb{P}_\theta(X = 5) = \theta^5$$

Test 2 rejects H_0 if $X = 3, 4, 5$ ($R = \{3, 4, 5\}$).

$$\beta_2(\theta) = \mathbb{P}_\theta(X \in \{3, 4, 5\})$$
$$= \binom{5}{3}\theta^3(1-\theta)^2 + \binom{5}{4}\theta^4(1-\theta)^1 + \theta^5$$



p-value

- Consider a sequence of tests with rejection region R_α at significance level α.
- As α decreases, R_α shrinks and we will fail to reject $\mathbf{x}$ eventually.

Definition

For given tests R_α, the p-value of the observed value $\mathbf{x}$ of $\mathbf{X}$, denoted $p(\mathbf{x})$, is defined to be the smallest significance level for which H_0 is rejected, given $\mathbf{X} = \mathbf{x}$.

$$\mathbb{P}_0(p(\boldsymbol{X}) \leq \alpha) = \alpha.$$

- When $\alpha = p(\mathbf{x})$, $\mathbf{x}$ is at the boundary of R_α.
- Informally, $p(\mathbf{x})$ is interpreted as the probability of an outcome to be "more extreme" than $\mathbf{x}$, under the null hypothesis H_0.

Sufficiency and Tests

With a sufficient statistic T is available, we may restrict our attention to tests based on the sufficient statistic T.

Theorem

If $T(\boldsymbol{X})$ is a sufficient statistic for θ, then for any test φ, the test

$$\psi = \psi(T) = \mathbb{E}[\varphi(\boldsymbol{X})|T]$$

has the same power function as φ,

$$\mathbb{E}_\theta[\varphi(\boldsymbol{X})] = \mathbb{E}_\theta[\mathbb{E}_\theta[\varphi(\boldsymbol{X})|T]] = \mathbb{E}_\theta[\psi(T)].$$

Once again, sufficiency is for the validity of the test, not for the power of the test.

Randomized Tests

It is sometimes convenient to allow external randomization to "help" the researcher decide between the two hypotheses.

Test/critical function

Randomized tests are characterized by a test or critical function φ that maps a sample $\boldsymbol{x}$ to $[0,1]$. Given $\boldsymbol{X}=\boldsymbol{x}$, $\varphi(\boldsymbol{x})$ is the chance of rejecting H_0.

- The power function β still gives the chance of rejecting H_0

$$\beta(\theta)=\mathbb{P}_\theta(\text{reject } H_0)=\mathbb{E}_\theta[\mathbb{P}_\theta(\text{reject } H_0|\boldsymbol{X})]=\mathbb{E}_\theta[\varphi(\boldsymbol{X})].$$

- A nonrandomized test with critical region R can be viewed as a randomized test with $\varphi=\mathbb{1}_R$.
- Convex combinations of randomized tests are still random tests.

Most Powerful Tests

Let $\mathcal{C}$ be a class of tests for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$.

Definition (UMP)

A test in class $\mathcal{C}$, with power function $\beta(\theta)$, is a uniformly most powerful (UMP) class $\mathcal{C}$ test if

$$\beta(\cdot) \geq \beta'(\cdot) \text{ on } \Theta_0^c$$

for any power function $\beta'(\theta)$ of a test in class $\mathcal{C}$.

If $\mathcal{C}$ denote the class of level α test, it is called a UMP level α test.

- UMP tests are considered to be the best, but may not always exist.

Simple Versus Simple Testing

Recall simple hypotheses are specified by only one possible distribution

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1.$$

- The power function for a test φ has two values

$$\alpha = \mathbb{E}_{\theta_0}[\varphi(\boldsymbol{X})] \quad \text{and} \quad \beta = \mathbb{E}_{\theta_1}[\varphi(\boldsymbol{X})]$$

- We consider the constrained maximization problem of maximizing the power β among all tests φ with significance level at most $\alpha = \mathbb{E}_{\theta_0}[\varphi(\boldsymbol{X})]$.
- I.e., we aim to find the UMP test for simple hypotheses.
  - For simple alternative hypothesis, finding UMP is maximizing $\beta = \mathbb{E}_{\theta_1}[\varphi(\boldsymbol{X})]$.

Simple Versus Simple Testing

Lemma

Suppose $k \geq 0$ *and* φ^* *maximizes*

$$\mathbb{E}_{\theta_1}[\varphi(\boldsymbol{X})] - k\mathbb{E}_{\theta_0}[\varphi(\boldsymbol{X})]$$

among all critical functions, and $\mathbb{E}_{\theta_0}[\varphi(\boldsymbol{X})] = \alpha$*. Then* φ^* *maximizes* $\beta = \mathbb{E}_{\theta_1}[\varphi(\boldsymbol{X})]$ *over all* φ *with level at most* α*.*

Proof. Suppose $\mathbb{E}_{\theta_0}[\varphi] \leq \alpha$. Then

$$\begin{aligned}\mathbb{E}_{\theta_1}[\varphi] &\leq \mathbb{E}_{\theta_1}[\varphi] - k\mathbb{E}_{\theta_0}[\varphi] + k\alpha \\ &\leq \mathbb{E}_{\theta_1}[\varphi^*] - k\mathbb{E}_{\theta_0}[\varphi^*] + k\alpha \\ &= \mathbb{E}_{\theta_1}[\varphi^*]\end{aligned}$$

- Note that, we need only consider tests with level exactly α.

Simple Versus Simple Testing

To maximize the power, note that

$$\begin{aligned}\mathbb{E}_{\theta_1}[\varphi] - k\mathbb{E}_{\theta_0}[\varphi] &= \int [f(\boldsymbol{x}|\theta_1) - kf(\boldsymbol{x}|\theta_0)]\varphi(\boldsymbol{x})d\boldsymbol{x} \\ &= \int_{f(\boldsymbol{x}|\theta_1) > kf(\boldsymbol{x}|\theta_0)} |f(\boldsymbol{x}|\theta_1) - kf(\boldsymbol{x}|\theta_0)|\varphi(\boldsymbol{x})d\boldsymbol{x} \\ &\quad - \int_{f(\boldsymbol{x}|\theta_1) < kf(\boldsymbol{x}|\theta_0)} |f(\boldsymbol{x}|\theta_1) - kf(\boldsymbol{x}|\theta_0)|\varphi(\boldsymbol{x})d\boldsymbol{x}\end{aligned}$$

Clearly, any test φ^* maximizing this expression must have

$$\varphi^*(\boldsymbol{x}) = \begin{cases} 1, & \text{when } f(\boldsymbol{x}|\theta_1) > kf(\boldsymbol{x}|\theta_0), \\ 0, & \text{when } f(\boldsymbol{x}|\theta_1) < kf(\boldsymbol{x}|\theta_0). \end{cases}$$

- Proof for the discrete case follows exactly the same lines.

Likelihood Ratio Tests

The maximizer φ^* are based on the likelihood ratio

$$\lambda(\boldsymbol{x}) = L(\theta_1|\boldsymbol{x})/L(\theta_0|\boldsymbol{x}), \quad \text{where } L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta).$$

I.e., $\varphi^*(\boldsymbol{x}) = 1$ if $\lambda(\boldsymbol{x}) > k$ and $\varphi^*(\boldsymbol{x}) = 0$ if $\lambda(\boldsymbol{x}) < k$. (When $\lambda(\boldsymbol{x}) = k$, $\varphi^*(\boldsymbol{x})$ can take any value in $[0, 1]$.)

Likelihood ratio test

Any test that satisfies the following is called a likelihood ratio test

$$\varphi(\boldsymbol{x}) = \begin{cases} 1, & \text{when } \lambda(\boldsymbol{x}) > k, \\ \gamma, & \text{when } \lambda(\boldsymbol{x}) = k, \\ 0, & \text{when } \lambda(\boldsymbol{x}) < k. \end{cases}$$

Here, γ can take any value in $[0, 1]$.

Example: Exponential distribution. Suppose we take one observation $X \sim \text{Exp}(\theta)$, i.e., with density $f_\theta(x) = \theta e^{-\theta x} \mathbb{1}(x \geq 0)$, and that we would like to test $H_0 : \theta = 1$ versus $H_1 : \theta = \theta_1 > 1$.

A likelihood ratio test φ is equal to 1 if

$$\lambda(X) = \frac{\theta_1 e^{-\theta_1 X}}{e^{-X}} > k,$$

equivalently, if

$$X < \frac{\log(\theta_1/k)}{\theta_1 - 1} \equiv k'.$$

φ is equal to 0 if $X > k'$. (When $X = k'$ the test can take any value in $[0, 1]$, but the choice will not affect any power calculations since $\mathbb{P}_\theta(X = k') = 0$.)

The level of this likelihood ratio test is

$$\alpha = \mathbb{P}(X < k') = 1 - \alpha^{-k'},$$

which gives $k' = -\log(1 - \alpha)$.

Hence, if

$$\varphi_\alpha(x) = \begin{cases} 1, & \text{when } X < -\log(1-\alpha), \\ 0, & \text{when } X > -\log(1-\alpha), \end{cases}$$

then ϕ_α maximizes $\mathbb{E}_{\theta_1}[\varphi(X)]$ over all φ with level at most α.

- This test φ_α, which is optimal for testing test $H_0 : \theta = 1$ versus $H_1 : \theta = \theta_1 > 1$, does not depend on the value θ_1.
- Feature like this allows for using the same optimal test regardless of the alternative hypothesis, which is essential because the parameter is unknown.
- In such case, φ_α is a UMP test for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$.

Lemma

Suppose that there is a test φ of level α such that for every $\theta_1 \in \Theta_1$, φ is the UMP for testing H_0 versus $H_1 : \theta = \theta_1$. Then φ is UMP for testing H_0 versus $H_1 : \theta \in \Theta_1$.

Example: Binomial distribution. Suppose we take one observation $X \sim \text{Bin}(2, \theta)$ and that we would like to test $H_0 : \theta = 1/2$ versus $H_1 : \theta = \theta_1 = 3/4$, then

$$\lambda(X) = \frac{\binom{2}{X}(3/4)^X(1/4)^{2-X}}{\binom{2}{X}(1/2)^X(1/2)^{2-X}} = \frac{3^X}{4},$$

Under H_0,

$$\lambda(x) = \begin{cases} 1/4, & \text{with probability } 1/4, \\ 3/4, & \text{with probability } 1/2, \\ 9/4, & \text{with probability } 1/4, \end{cases}$$

Suppose the desired significance level is $\alpha = 5\%$.

- If $k < 9/4$, then $\varphi(2) = 1$. But then $\mathbb{E}_{\theta_0}[\varphi(X)] \geq \varphi(2)\mathbb{P}_{\theta_0}(X = 2) = 1/4 > \alpha$.
- If $k > 9/4$, then φ is identically 0. Hence, then $\mathbb{E}_{\theta_0}[\varphi(X)] = 0 < \alpha$.
- So $k = 9/4$, which implies $\varphi(0) = \varphi(1) = 0$. To achieve the desired α, we set $\varphi(2) = 1/5$. (A randomized test!)

Neyman-Pearson Lemma

Previously, when we derive that the likelihood ratio test is optimal, we did not specify $\mathbb{E}_{\theta_0}[\varphi] = \alpha$. As we see from the binomial example, randomized test needed to achieve level α in a typical situation.

To find a likelihood ratio test with any desired level $\alpha \in [0, 1]$

- First k is adjusted so that $\mathbb{P}_{\theta_0}(\lambda(\boldsymbol{X}) > k)$ and $\mathbb{P}_{\theta_0}(\lambda(\boldsymbol{X}) < k)$ bracket α.
- Then a value $\gamma \in [0, 1]$ is chosen for $\varphi(\boldsymbol{X})$ when $\lambda(\boldsymbol{X}) = k$ to achieve level α.

Theorem (Neyman-Pearson Lemma – Existence)

Given any level $\alpha \in [0, 1]$, there exists a likelihood ratio test φ_α with level α, and any likelihood ratio test with level α maximizes β among all tests with level at most α.

Neyman-Pearson Lemma

Theorem (Neyman-Pearson Lemma – Uniqueness)

Fix $\alpha \in [0,1]$, let φ_α be the UMP likelihood ratio test with critical value k. Define

$$B = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \neq k\}.$$

If another test φ^* maximizes β among all tests with level at most α, then φ^* and φ_α must agree on B almost surely.

- In other words, in dealing with simple versus simple hypothesis testing, the UMP test must be a likelihood ratio test.
- If the measure of B is 0 (usually the case for continuous distributions), then we have a unique nonrandomized test; otherwise, a randomized test is necessary to achieve the desired α.

Proof. Assume $k < \infty$ and let $B_1 = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) > k\}$ and $B_2 = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) < k\}$.

- Let φ_α be the UMP likelihood ratio test. Then, $\mathbb{E}_{\theta_1}[\varphi^*] = \mathbb{E}_{\theta_1}[\varphi_\alpha]$.
- φ_α maximizes $\mathbb{E}_{\theta_1}[\varphi] - k\mathbb{E}_{\theta_0}[\varphi]$, hence

$$k\mathbb{E}_{\theta_0}[\varphi_\alpha] = k\alpha \leq k\mathbb{E}_{\theta_0}[\varphi^*] \Rightarrow \mathbb{E}_{\theta_0}[\varphi^*] = \alpha.$$

- Hence, we have $\mathbb{E}_{\theta_1}[\varphi_\alpha] - k\mathbb{E}_{\theta_0}[\varphi_\alpha] = \mathbb{E}_{\theta_1}[\varphi^*] - k\mathbb{E}_{\theta_0}[\varphi^*]$.
- Recall

$$\mathbb{E}_{\theta_1}[\varphi_\alpha] - k\mathbb{E}_{\theta_0}[\varphi_\alpha] = \int \mathbb{1}_{B_1} |f(\boldsymbol{x}|\theta_1) - kf(\boldsymbol{x}|\theta_0)| d\boldsymbol{x}$$

- Then

$$\int \mathbb{1}_{B_1} |f(\boldsymbol{x}|\theta_1) - kf(\boldsymbol{x}|\theta_0)|(1 - \varphi^*(\boldsymbol{x}))d\boldsymbol{x} + \int \mathbb{1}_{B_2} |f(\boldsymbol{x}|\theta_1) - kf(\boldsymbol{x}|\theta_0)|\varphi^*(\boldsymbol{x})d\boldsymbol{x} = 0.$$

- Since integrands are nonnegative, they must be 0 almost surely.

Corollary of Neyman-Pearson Lemma

Corollary

Suppose $T(\mathbf{X})$ is a sufficient statistics for θ with pdf/pmf $g(t|\theta_i)$, $i = 0, 1$. Then any likelihood ratio test $\varphi_{T,\alpha}$ based on T, i.e.,

$$\varphi_{T,\alpha}(t) = \begin{cases} 1, & \text{when } \lambda(t) > k, \\ \gamma, & \text{when } \lambda(t) = k, \\ 0, & \text{when } \lambda(t) < k, \end{cases}$$

is a UMP test, where $\lambda(t) = g(t|\theta_1)/g(t|\theta_0)$.

Proof. Factorization theorem implies that

$$f(\mathbf{x}|\theta_i) = g(T(\mathbf{x})|\theta_i)h(\mathbf{x}) \quad i = 0, 1.$$

Composite Hypotheses

Previously, we studied the UMP tests for simple hypotheses.

- The case where H_0 and H_1 are both simple is mainly of theoretical interest.
- When a hypothesis is not simple, it is called composite.
- We saw an example with exponential distribution, where UMP likelihood ratio test for composite H_1 exists.

Unsurprisingly, we may still need the likelihood ratio tests for composite hypotheses.

Uniformly Most Powerful Test

Recall the definition of UMP level α test.

A test φ^* with level α is called uniformly most powerful if

$$\beta^*(\theta) = \mathbb{E}_\theta[\varphi^*] \geq \beta(\theta) = \mathbb{E}_\theta[\varphi], \quad \forall \theta \in \Theta_1,$$

for all φ with level at most α.

- Uniformly most powerful tests for composite hypotheses generally only arise when the parameter of interest is univariate, $\theta \in \Theta \subset \mathbb{R}$ and the hypotheses are of the form $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where θ_0 is a fixed constant.
- In addition, the family of densities needs to have an appropriate structure.

Monotone Likelihood Ratio

Recall the exponential example, where UMP exists for composite hypotheses. We now extend that result to a class of parametric problems in which the likelihood functions have a special property.

Definition (Monotone Likelihood Ratio)

A family of densities $f(\boldsymbol{x}|\theta), \theta \in \Theta \subset \mathbb{R}$ has monotone likelihood ratios (in $T(\boldsymbol{X})$) if there exists a statistic $T(\boldsymbol{X})$ such that whenever $\theta_1 < \theta_2$, the likelihood ratio $f(\boldsymbol{x}|\theta_2)/f(\boldsymbol{x}|\theta_1)$ is a nondecreasing function of T.

- We assume that the distributions are identifiable, $\mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$ whenever $\theta_1 \neq \theta_2$.
- Natural conventions concerning division by zero are used here, with the likelihood ratio interpreted as $+\infty$ when $f(x|\theta_2) > 0$ and $f(x|\theta_1) = 0$.
- On the null set where both densities are zero the likelihood ratio is not defined and monotonic dependence on T is not required.

Lemma

Suppose that $f(\boldsymbol{x}|\theta), \theta \in \Theta \subset \mathbb{R}$ has monotone likelihood ratios in $T(\boldsymbol{X})$. If ψ is a nondecreasing function of T, then $g(\theta) = E[\psi(T)]$ is a nondecreasing function of θ.

Proof. Let $\theta_1 < \theta_2$,

- Let $A = \{\boldsymbol{x} : f(\boldsymbol{x}|\theta_1) > f(\boldsymbol{x}|\theta_2)\}$ and $a = \sup_{\boldsymbol{x}\in A} \psi(T(\boldsymbol{x}))$.
- Let $B = \{\boldsymbol{x} : f(\boldsymbol{x}|\theta_1) < f(\boldsymbol{x}|\theta_2)\}$ and $b = \inf_{\boldsymbol{x}\in B} \psi(T(\boldsymbol{x}))$.
- Then $b \geq a$ because MLR and ψ is nondecreasing.

$$
\begin{aligned}
g(\theta_2) - g(\theta_1) &= \int \psi(T(\boldsymbol{x}))(f(\boldsymbol{x}|\theta_2) - f(\boldsymbol{x}|\theta_1))d\boldsymbol{x} \\
&\geq a \int_A (f(\boldsymbol{x}|\theta_2) - f(\boldsymbol{x}|\theta_1))d\boldsymbol{x} + b \int_B (f(\boldsymbol{x}|\theta_2) - f(\boldsymbol{x}|\theta_1))d\boldsymbol{x} \\
&= (b-a) \int_B (f(\boldsymbol{x}|\theta_2) - f(\boldsymbol{x}|\theta_1))dx \\
&\geq 0
\end{aligned}
$$

Example: Exponential family. If the densities $f(x|\theta)$ form a 1-parameter exponential family,

$$f(\boldsymbol{x}|\theta) = \left(\prod_i h(x_i)\right) c^n(\theta) \exp\left(\eta(\theta) \sum_i t(x_i)\right),$$

with $\eta(\theta)$ strictly increasing, then if $\theta_2 > \theta_1$,

$$\frac{f(\boldsymbol{x}|\theta_2)}{f(\boldsymbol{x}|\theta_1)} = \frac{c^n(\theta_2)}{c^n(\theta_1)} \exp\left((\eta(\theta_2) - \eta(\theta_1)) \sum_i t(x_i)\right),$$

which is increasing in $T = \sum_i t(x_i)$.

- Binomial, Poisson, negative binomial, normal, exponential[1], gamma, beta...

[1] Recall our exponential example!

Example: Uniform$(0, \theta)$. For $\theta_1 < \theta_2$,

$$\frac{f(\boldsymbol{x}|\theta_2)}{f(\boldsymbol{x}|\theta_1)} = \frac{\theta_1^n}{\theta_2^n} \frac{\mathbb{1}(x_{(n)} \leq \theta_2)}{\mathbb{1}(x_{(n)} \leq \theta_1)}$$

is a nondecreasing function of $x_{(n)}$ for $\boldsymbol{x}$'s at which at least one of $f(x|\theta_2)$ and $f(x|\theta_1)$ is positive, i.e., when $x_{(n)} < \theta_2$.

UMP for One-Sided Tests

Theorem

Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Suppose the family of densities has monotone likelihood ratios in $T(\boldsymbol{X})$. The test φ^* given by the following is UMP

$$\varphi^*(\boldsymbol{x}) = \begin{cases} 1, & \text{when } T(\boldsymbol{x}) > c, \\ \gamma, & \text{when } T(\boldsymbol{x}) = c, \\ 0, & \text{when } T(\boldsymbol{x}) < c, \end{cases}$$

where c and α are adjusted to achieve any desired level α.

Proof.

- Consider simple hypotheses $\theta = \theta_0$ versus $\theta = \theta_1$. Previously, we have the following being UMP

$$\varphi(\boldsymbol{x}) = \begin{cases} 1, & \text{when } \lambda(\boldsymbol{x}) > c, \\ \gamma, & \text{when } \lambda(\boldsymbol{x}) = c, \\ 0, & \text{when } \lambda(\boldsymbol{x}) < c. \end{cases}$$

- Since the family of densities has monotone likelihood ratios, this UMP test can be chosen to be the same as φ based on $T(\boldsymbol{x})$, with possibly different c and γ.
- Note that this test does not depend on θ_1 (because $T(\boldsymbol{X})$ is a statistic). Hence, it is UMP for $\theta = \theta_0$ versus $H_1 : \theta > \theta_0$.
- Note that φ^* is nondecreasing, hence a previous Lemma implies that $\beta(\theta) = \mathbb{E}_\theta[\varphi^*(T(\boldsymbol{X}))]$ is nondecreasing.
- Then φ^* is UMP for testing H_0 versus H_1, because $\beta(\theta) \leq \alpha = \beta(\theta_0)$ for all $\theta \leq \theta_0$.

UMP One-Sided Tests for Exponential Families

Corollary

Suppose X follows an exponential family with strictly monotone $\eta(\theta)$.

1. *If $\eta(\theta)$ is strictly increasing, then the UMP for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ is given by*

$$\varphi^*(\boldsymbol{x}) = \begin{cases} 1, & \textit{when } T(\boldsymbol{x}) > c, \\ \gamma, & \textit{when } T(\boldsymbol{x}) = c, \\ 0, & \textit{when } T(\boldsymbol{x}) < c, \end{cases}$$

2. *If $\eta(\theta)$ is strictly decreasing or $H_0 : \theta \geq \theta_0$ ($H_1 : \theta < \theta_0$), then the UMP is*

$$\varphi^*(\boldsymbol{x}) = \begin{cases} 1, & \textit{when } T(\boldsymbol{x}) < c, \\ \gamma, & \textit{when } T(\boldsymbol{x}) = c, \\ 0, & \textit{when } T(\boldsymbol{x}) > c, \end{cases}$$

Examples

Example: One-sided test for normal mean μ with known σ. $H_0 : \mu \leq \mu_0$. We have $T(\boldsymbol{X}) = \bar{X}$ with $\eta(\theta) = n\mu/\sigma^2$ strictly increasing, the UMP is given by

$$\varphi^*(\boldsymbol{x}) = \begin{cases} 1, & \text{when } \bar{X} > \sigma z_{1-\alpha}/\sqrt{n} + \mu_0, \\ 0, & \text{when } \bar{X} < \sigma z_{1-\alpha}/\sqrt{n} + \mu_0, \end{cases}$$

The value of γ does not matter. So we have a nonrandomized test.

Example: One-sided test for Bernoulli. $H_0 : p \leq p_0$. We have $T(\boldsymbol{X}) = \sum_i X$ with $\eta(p) = \log(p/(1-p))$ strictly increasing. In this case, we shall need a randomized test.

Example: One-sided test for Poisson. $H_0 : \theta \leq \theta_0$. We have $T(\boldsymbol{X}) = \sum_i X$ with $\eta(\theta) = \log(\theta)$ increasing. In this case, we shall need a randomized test.

- Usually, we need a randomized test when the distribution is discrete in order to achieve any desired level α.

Examples

Example: Uniform$(0, \theta)$. This is not an exponential family, but still has monotone likelihood ratio in $T(\boldsymbol{X}) = X_{(n)}$. The UMP is given by

$$\varphi^*(\boldsymbol{x}) = \begin{cases} 1, & \text{when } T(\boldsymbol{x}) > c, \\ \gamma, & \text{when } T(\boldsymbol{x}) = c, \\ 0, & \text{when } T(\boldsymbol{x}) < c. \end{cases}$$

To calculate c and γ, note that the density of $X_{(n)}$ is $n\theta^{-1}x^{n-1}\mathbb{1}(x \leq \theta)$. The UMP test is nonrandomized with

$$\alpha = \mathbb{E}_{\theta_0}[\varphi^*] = n\theta_0^{-1}\int_c^{\theta_0} x^{n-1}dx = 1 - \frac{c^n}{\theta_0^n}.$$

So $c = \theta_0(1-\alpha)^{1/n}$.

Two-Sided Tests

Now we consider two-sided hypothesis

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_0 : \theta \neq \theta_0$$

Definition (Two-sided test)

A test φ is called two-sided if there are finite constants $t_1 \leq t_2$ such that

$$\varphi(\boldsymbol{x}) = \begin{cases} 1, & \text{when } T(\boldsymbol{x}) < t_1 \text{ or } T(\boldsymbol{x}) > t_2, \\ 0, & \text{when } T(\boldsymbol{x}) \in (t_1, t_2), \end{cases}$$

In addition, the test should not be one-sided.

Nonexistence of UMP Two-Sided Test

Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ for a population with one-parameter exponential family, i.e., with density

$$c(\theta)h(x)e^{\eta(\theta)T(x)}, \quad \theta \in \Theta.$$

Assume that $\eta(\theta)$ is strictly increasing.

- Decompose into $H_0 : \theta = \theta_0$ vs $H_1 : \theta \geq \theta_0$ and $H_0 : \theta = \theta_0$ vs $H_1 : \theta \leq \theta_0$.
- There are two level α UMP tests for one-sided alternatives

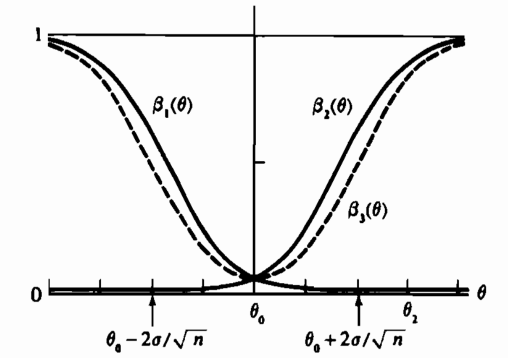
$$\varphi_+(\boldsymbol{x}) = \begin{cases} 1, & \text{when } T(\boldsymbol{x}) > c_+, \\ \gamma, & \text{when } T(\boldsymbol{x}) = c_+, \\ 0, & \text{when } T(\boldsymbol{x}) < c_+, \end{cases} \quad \text{and} \quad \varphi_-(\boldsymbol{x}) = \begin{cases} 1, & \text{when } T(\boldsymbol{x}) < c_-, \\ \gamma, & \text{when } T(\boldsymbol{x}) = c_-, \\ 0, & \text{when } T(\boldsymbol{x}) > c_-. \end{cases}$$

- If $\theta_- < \theta_0 < \theta_+$, then φ_+ has the maximal power at θ_+, and φ_- has the maximal power at θ_-.
- Decompose into $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_-$ or $H_1 : \theta = \theta_+$. By uniqueness in Neyman-Pearson Lemma, there cannot be a UMP level α two-sided test.

Nonexistence of UMP Two-Sided Test

Example:

- Test 1: $R_1 = (-\infty, \theta_0 - \sigma z(\alpha)/\sqrt{n})$ with power function β_1.
- Test 2: $R_2 = (\theta_0 + \sigma z(\alpha)/\sqrt{n}, \infty)$ with power function β_2.
- Test 3: $R_3 = R_1 \cup R_2$ with power function β_3.



Unbiasedness of Power Function

When a UMP test does not exist, we impose reasonable restrictions on the test to be considered, then find the best one among the restricted class of tests.

The following is a restriction says that the test is at least as good as a silly guess

- where we reject H_0 with probability α, independent of the observation.

Definition (Unbiased tests)

A level α test with power function $\beta(\theta)$ is unbiased if and only if $\beta(\theta') \geq \alpha$ for every $\theta' \in \Theta_0^c$ and $\alpha \geq \beta(\theta'')$ for every $\theta'' \in \Theta_0$.

- A UMP test among the class of unbiased tests is called a uniformly most powerful unbiased (UMPU) test.
- A UMP test is always unbiased. (Why?) Hence, the discussion of UMPU is only relevant when UMP does not exist.

UMPU for Exponential Family (Without Proof)

Theorem

Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ for a population with one-parameter exponential family, i.e., with density

$$c(\theta)h(x)e^{\eta(\theta)T(x)}, \quad \theta \in \Theta,$$

where $\eta(\theta)$ is differentiable and strictly increasing with $0 < \eta'(\theta_0) < \infty$. Then there exist a two-sided, level α test φ^* with $\beta'_{\varphi^*}(\theta_0) = 0$. Any such test is a uniformly most powerful unbiased test.

Remarks

Examples of UMPU in normal families (UMP do not exist)

- One-sample two-sided z-test
- One-sample two-sided t-test
- Two-sided Chi-squared test, but with certain unequal tail
- One-sided two-sample F-test
- Two-sided two-sample F-test, but with certain unequal tail
- One-sided and two-sided two-sample t-test

Likelihood Ratio Tests

Previously, we introduced the likelihood ratio tests from simple hypothesis. Now consider general hypothesis, possibly composite.

- We allow θ to be a vector $\boldsymbol{\theta}$. Let Θ be the entire parameter space.
- The likelihood ratio test statistic for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}.$$

To motivate this definition of likelihood ratio, note that

$$L(\theta_1|\boldsymbol{x})/L(\theta_0|\boldsymbol{x}) > c_0, \text{ for some } c_0 > 0$$

is equivalent to

$$L(\theta_0|\boldsymbol{x})/\max\{L(\theta_0|\boldsymbol{x}), L(\theta_1|\boldsymbol{x})\} < c,$$

for $c = (1/c_0) \wedge 1$ (excerise).

Intuition of Likelihood Ratio Tests

The likelihood ratio test is defined in a similar way, with $\lambda(\boldsymbol{x})$ replaced:

$$\varphi(\boldsymbol{x}) = \begin{cases} 1, & \text{when } \lambda(\boldsymbol{x}) < c, \\ 0, & \text{when } \lambda(\boldsymbol{x}) > c. \end{cases}$$

- $L(\theta|\mathbf{x})$: how likely $\mathbf{x}$ happens.
- $\sup_{\Theta_0} L(\theta|\mathbf{x})$: the likelihood of the most likely parameter in Θ_0.
- $\lambda(\mathbf{x})$ is small: for parameters in H_0, the sample is relatively unlikely to happen compared to parameters in H_1.
- fail to reject H_0 only if the ratio is large enough.
- Connection to MLE:
  - Denominator: the MLE of θ in Θ, the maximizer of $L(\theta|x)$ over Θ.
  - Numerator: the MLE of θ in Θ_0, the maximizer of $L(\theta|x)$ over Θ_0.

Normal LRT

Example: (Known variance) Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathcal{N}(\theta, 1)$. Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

$$\lambda(\mathbf{x}) = \frac{(\sqrt{2\pi})^{-n} \exp\left[\sum_{i=1}^{n}(x_i - \theta_0)^2/2\right]}{(\sqrt{2\pi})^{-n} \exp\left[\sum_{i=1}^{n}(x_i - \bar{x})^2/2\right]} = \ldots = \exp\left[-n(\bar{x} - \theta_0)^2/2\right]$$

The rejection region $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$ for some $c \in [0, 1]$

$$\{\mathbf{x} : |\bar{x} - \theta_0| \geq \sqrt{-2\log(c)/n}\}$$

- It has the same form as the classical one-sample z-test for normal mean.
- There is a one-to-one correspondence between c and α.
- Similarly, the one-sided tests are also LRTs.
- Note that the test depends on the sample only through the sufficient statistic $\bar{x}$!
- This is an UMPU.

Exponential LRT

Example: (Location family) Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x|\theta) = e^{-(x-\theta)}$, $x \geq \theta$. Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

The likelihood function is

$$L(\theta|\mathbf{x}) = \begin{cases} e^{-\sum x_i + n\theta} & \theta \leq x_{(1)} \\ 0 & \theta > x_{(1)} \end{cases} \qquad \Rightarrow \qquad \hat{\theta} = x_{(1)}$$

So

$$\lambda(\mathbf{x}) = \begin{cases} 1 & x_{(1)} \leq \theta_0 \\ e^{-n(x_{(1)}-\theta_0)} & x_{(1)} > \theta_0 \end{cases}$$

The rejection region is

$$\{\mathbf{x} : x_{(1)} \geq \theta_0 - \frac{\log(c)}{n}\}$$

The test depends on the sample only through the sufficient statistic $x_{(1)}$!

Likelihood Ratio Tests

If $T(X)$ is a sufficient statistics for θ, we can obtain its likelihood function (pdf/pmf) $L^*(\theta|t) = g(t|\theta)$, which yields LRT $\lambda^*(t)$.

Theorem

If $T(\mathbf{X})$ is a sufficient statistic for θ and $\lambda^*(t)$ and $\lambda(\mathbf{x})$ are the LRT statistic based on T and $\mathbf{X}$, respectively, then for every sample $\mathbf{x}$

$$\lambda^*(T(\mathbf{x})) = \lambda(\mathbf{x}).$$

Proof. Apply the factorization theorem,

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})} = \frac{\sup_{\Theta_0} f(\mathbf{x}|\theta)}{\sup_{\Theta} f(\mathbf{x}|\theta)} = \frac{\sup_{\Theta_0} g(T(\mathbf{x})|\theta)h(\mathbf{x})}{\sup_{\Theta} g(T(\mathbf{x})|\theta)h(\mathbf{x})} = \lambda^*(T(\mathbf{x})).$$

$\lambda^*(\cdot)$ depends on the sample only through $T(\cdot)$, as in the previous two examples.

Normal LRT

Example: (Unknown σ) Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathcal{N}(\mu, \sigma)$. Consider testing $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$.

$$\lambda(\mathbf{x}) = \frac{\sup_{\mu \leq \mu_0, \sigma^2 \geq 0} L(\mu, \sigma|\mathbf{x})}{\sup_{\mu \in \mathbb{R}, \sigma^2 \geq 0} L(\mu, \sigma|\mathbf{x})} = \begin{cases} 1 & \text{if } \hat{\mu} \leq \mu_0 \\ \frac{L(\mu_0, \hat{\sigma}_0^2|\mathbf{x})}{L(\hat{\mu}, \hat{\sigma}^2|\mathbf{x})} & \text{if } \hat{\mu} > \mu_0 \end{cases}$$

- Recall that $(\hat{\mu}, \hat{\sigma}^2) = (\bar{X}, \frac{n-1}{n} S^2)$ is the MLE of (μ, σ^2).
- For the constrained problem, using the method of Lagrange multipliers, MLE yields $(\mu_0, \hat{\sigma}_0^2)$ for $\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2$ when $\hat{\mu} > \mu_0$.
- The rejection region of the LRT is $R = \left\{ S^2 / \sum_{i=1}^n (X_i - \mu_0)^2 < c \right\}$.
- This has the same form as the classical one-sample t-tests $R = \left\{ \sqrt{n}(\bar{X} - \mu_0)/S > t_{n-1}(1-\alpha) \right\}$. (use the algebraic identity)

Sequential testing

In all previous tests, we have fixed the sample size n in advance. But in many applications (e.g., clinical trials), we often observe the data sequentially.

- Continue to gather samples until a confident conclusion can be made.
- This idea goes back to Wald (1945), and is referred as *sequential probability ratio test* (SPRT).

Sequential probability ratio test

We consider testing

$$H_0 : X_i \overset{i.i.d.}{\sim} f_0 \quad \text{versus} \quad H_1 : X_i \overset{i.i.d.}{\sim} f_1.$$

The likelihood ratio is

$$\lambda_n(\boldsymbol{X}_n) = \prod_{i=1}^{n} \frac{f_1(X_i)}{f_0(X_i)} = \frac{L_1(\boldsymbol{X})}{L_0(\boldsymbol{X})}$$

SPRT

SPRT requires two thresholds $0 < c_0 < c_1 < 1$:

- if $\lambda_n(\boldsymbol{X}_n) \geq c_1$, then reject H_0;
- if $\lambda_n(\boldsymbol{X}_n) \leq c_0$, then accept H_0;
- if $c_0 \leq \lambda_n(\boldsymbol{X}_n) \leq c_1$, we continue to collect sample.

Sequential probability ratio test – thresholds

- The thresholds c_0 and c_1 are chosen to control the type I ($\alpha = \mathbb{P}_0(\lambda_n(\boldsymbol{X}_n) \geq c_1)$) and type II ($1 - \beta = \mathbb{P}_1(\lambda_n(\boldsymbol{X}_n) \leq c_0)$) error.

Let's express the two types of error.

$$\beta \geq \mathbb{P}_1(\lambda_n(\boldsymbol{X}_n) \geq c_1) = \int_{\lambda_n(\boldsymbol{x}) \geq c_1} \frac{L_1(\boldsymbol{x})}{L_0(\boldsymbol{x})} L_0(\boldsymbol{x}) \mathrm{d}x = \int_{\lambda_n(\boldsymbol{x}) \geq c_1} \lambda_n(\boldsymbol{x}) L_0(\boldsymbol{x}) \mathrm{d}x$$

$$\geq c_1 \int_{\lambda_n(\boldsymbol{x}) \geq c_1} L_0(\boldsymbol{x}) \mathrm{d}x = c_1 \alpha.$$

Similarly, we can show that

$$1 - \alpha \geq \mathbb{P}_0(\lambda_n(\boldsymbol{X}_n) \leq c_0) = \int_{\lambda_n(\boldsymbol{x}) \leq c_0} \frac{L_0(\boldsymbol{x})}{L_1(\boldsymbol{x})} L_1(\boldsymbol{x}) \mathrm{d}x = \int_{\lambda_n(\boldsymbol{x}) \leq c_0} \lambda_n^{-1}(\boldsymbol{x}) L_0(\boldsymbol{x}) \mathrm{d}x$$

$$\geq c_0^{-1} \int_{\lambda_n(\boldsymbol{x}) \leq c_0} L_1(\boldsymbol{x}) \mathrm{d}x = c_0^{-1}(1 - \beta).$$

Sequential probability ratio test – thresholds

Now we have

$$\beta \geq c_1\alpha, \quad \text{and} \quad 1-\alpha \geq c_0^{-1}(1-\beta).$$

This implies that

$$c_1 \leq \frac{\beta}{\alpha}, \quad \text{and} \quad c_0 \geq \frac{1-\beta}{1-\alpha}.$$

So we set

$$c_1 = \frac{\beta}{\alpha}, \quad \text{and} \quad c_0 = \frac{1-\beta}{1-\alpha}.$$

to guarantee both type I and type II error are controlled at level α and β, respectively.

Sequential probability ratio test – expected stopping time

Let's consider the expectation of the log-likelihood ratio statistic at a fixed time n.

$$\mathbb{E}_1[\log \lambda_n] = \mathbb{E}_1\left[\sum_{i=1}^{n} \log \frac{f_1(X_i)}{f_0(X_i)}\right] = \sum_{i=1}^{n} \mathbb{E}_1\left[\log \frac{f_1(X_i)}{f_0(X_i)}\right] = k \cdot KL(f_1 \| f_0).$$

Where $KL(f_1\|f_0) = \mathbb{E}_1\left[\log \frac{f_1(X_i)}{f_0(X_i)}\right]$ is the Kullback-Leibler divergence between f_1 and f_0. Similarly,

$$\mathbb{E}_0[\log \lambda_n] = -k \cdot KL(f_0 \| f_1).$$

We have one complication: we do not know when the test will stop, i.e.,

$$\tau = \inf\{n \geq 0 : \lambda_n \geq c_1, \text{ or } \lambda_n \leq c_0\}$$

is a (random) stopping time. More importantly, τ is not independent of λ_n.

Sequential probability ratio test – expected stopping time

Luckily, we have Wald's identity:

Theorem

Let $Y_1, Y_2, \ldots$ be i.i.d. sample with mean μ. Let τ be a random variable such that $\mathbb{E}[\tau] < \infty$ and the event $\{\tau = t\}$ is determined by $Y_1, \ldots, Y_t$ and independent of $Y_i, i > t$. Then

$$\mathbb{E}\left[\sum_{i=1}^{\tau} Y_i\right] = \mu \mathbb{E}[\tau].$$

With Wald's identity, we have

$$\mathbb{E}_1[\log \lambda_\tau] = \mathbb{E}_1[\tau] \cdot KL(f_1 \| f_0) \Rightarrow \mathbb{E}_1[\tau] = \frac{\mathbb{E}_0[\log \lambda_\tau]}{KL(f_1 \| f_0)},$$

$$\mathbb{E}_0[\log \lambda_\tau] = -\mathbb{E}_0[\tau] \cdot KL(f_0 \| f_1) \Rightarrow \mathbb{E}_0[\tau] = \frac{\mathbb{E}_1[\log \lambda_\tau]}{-KL(f_0 \| f_1)}.$$

Sequential probability ratio test – expected stopping time

$$\mathbb{E}_0[\log \lambda_\tau] \approx \frac{\alpha \log(c_1) + (1-\alpha)\log(c_0)}{-KL(f_0 \| f_1)} = \frac{\alpha \log(\beta/\alpha) + (1-\alpha)\log((1-\beta)/(1-\alpha))}{-KL(f_0 \| f_1)}$$

$$\mathbb{E}_1[\log \lambda_\tau] \approx \frac{\beta \log(c_1) + (1-\beta)\log(c_0)}{KL(f_1 \| f_0)} = \frac{\beta \log(\beta/\alpha) + (1-\beta)\log((1-\beta)/(1-\alpha))}{KL(f_1 \| f_0)}$$

- Expected stopping time increases as the errors α or $1-\beta$ decreases, and as KL decreases.
- Wald's sequential test is optimal in the sense that it minimizes the expected stopping time among all tests with the same type I and type II error. (Proof omitted.)