TECHNISCHE UNIVERSITÄT WIEN

Introduction to Statistics

Tests of Hypotheses

LV Nr. 105.692

Summer Semester 2021

Theory of Hypothesis Testing

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

1. The parameter spaces Θ_0 and Θ_1 are any two disjoint subsets of the parameter space.
2. When Θ_0 is a singleton set (contains exactly one point), the null hypothesis is said to be *simple*.
   - For example, $\mu = 0$
3. The alternative hypothesis only motivates the choice of test statistic: e.g.,

$$Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$$

Decision theory and α levels

- The theory of statistical decisions is a large subject. When applied to hypothesis tests, it gives a different view.
- The point of a hypothesis test is to decide in favor of H_0 or H_1. The result is one of two decisions:
  1. accept H_0 or reject H_1 (both mean the same)
  2. reject H_0 or accept H_1 (both mean the same)
- In the decision-theoretic mode, the result of a test is just reported in these terms. No p-value is reported, hence no indication of the strength of evidence.

Type I and II errors

In a test of hypotheses, the sample space is partitioned into two disjoint regions: the rejection and the acceptance region

- Ω_1: the rejection region are the values of the test statistic T for which we reject the null at level α
- Ω_0: the acceptance region are the values of the test statistic for which we *cannot reject* the null at level α
- Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ is a vector of n iid random variables and $T(\mathbf{X})$ is a test statistic. The decision rule is

$$\text{if } \mathbf{X} \in \Omega_1 \quad \text{then reject } H_0$$

Type I and II errors

- We define two types of error with associated probabilities:

$$\begin{aligned}\alpha &= \mathbb{P}_{\theta\in\Theta_0}\left(\mathbf{X} \in \text{rejection region}\right)\\ &= \mathbb{P}\left(\mathbf{X} \in \Omega_1 \mid H_0 \text{ is true}\right)\\ &= \mathbb{P}(\text{Type I error})\\ \beta &= \mathbb{P}_{\theta\in\Theta_1}\left(\mathbf{X} \in \text{acceptance region}\right)\\ &= \mathbb{P}\left(\mathbf{X} \in \Omega_0 \mid H_0 \text{ is false}\right)\\ &= \mathbb{P}(\text{Type II error})\end{aligned}$$

- The power is now:

$$\begin{aligned}1-\beta &= \mathbb{P}_{\theta\in\Theta_1}\left(\mathbf{X} \in \Omega_1\right)\\ &= \mathbb{P}\left(\mathbf{X} \in \Omega_1 \mid H_0 \text{ is false}\right)\\ &= \textit{power}\end{aligned}$$

	Truth	
Decision	H_0	H_1
Accept H_0	correct decision $1-\alpha$	Type II error β
Reject H_0	Type I error α	correct decision $1-\beta$

Decision theory and α levels

If no p-value is reported, how is the test done?

A level of significance α is chosen

- If p-value $< \alpha$, then the test decides reject H_0.
- If p-value $\geqslant \alpha$, then the test decides accept H_0.

The decision theoretic view provides *less information*. Instead of giving the actual p-value, it is only reported whether the p-value is above or below α.

Type I error and p-value

- In the decision theoretic approach, α is the **smallest** p-value at which we can reject the null:
  - select an α (small) and **reject** H_0 if p-value $\leqslant \alpha$
- One can either calculate a *critical value* with respect to a chosen α and compare it with the observed test statistic, or calculate the p-value and compare it with α: the decision rule is exactly the same

Type I error and p-value

E.g. for testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$, when n is large,

1. Fix α and reject the null if

$$T(\mathbf{X}) = \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} > z_\alpha \iff \bar{x}_n > \mu_0 + z_\alpha \frac{z_n}{\sqrt{n}}$$

2. or, equivalently, compute

$$p - value = \mathbb{P}\left(Z > \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}\right)$$

and reject if p-value $\leqslant \alpha$.

Type I error and p-value

We toss a coin 10 times to test the null hypothesis that the coin is fair against the alternative that $\mathbb{P}(tails) > 0.5$. That is,

$$H_0 : p = 0.5 \quad vs \quad H_1 : p > 0.5$$

Fix $\alpha = 0.1$. We need to find the corresponding critical value C for the number of tails S so that

$$0.1 = \mathbb{P}_{p=0.5}\left(S \geqslant C\right) = 0.5^{10} \sum_{j=C}^{10} \binom{10}{j}$$

This has no exact integer solution (for any α). So, we find the minimal C such that the RHS does not exceed 0.1.

Type I error and p-value

$$0.1 = \mathbb{P}_{p=0.5}\left(S \geqslant C\right) = 0.5^{10} \sum_{j=C}^{10} \binom{10}{j}$$

1. No closed form solution: we start from the maximal $C = 10$ and reduce it until the probability exceeds 0.1. The previous value of C is the critical value.
   1. For $C = 8$, the probability is 0.0547 and for $C = 7$, 0.1719.
   2. Thus, the null is rejected if the number of tails is at least 8 and the actual level of this test is 0.0547.

Type I error and p-value

1. Now, suppose we observed 6 tails in 10 tosses.
2. The corresponding p-value is

$$p - value = \mathbb{P}_{p=0.5}(S \geqslant 6) = 0.5^{10} \sum_{j=6}^{10} \binom{10}{j} = 0.172$$

which is large enough to indicate that such a result is not so extreme under the null hypothesis and we will not reject it even for a "liberal" $\alpha = 0.1$.

Decision theory and α levels

- Ideally, the significance level α should be chosen carefully and reflect the costs and probabilities of false positive and false negative decisions.
- Since the decision-theoretic mode provides less information and isn't usually done properly
  - In practice, $\alpha = 0.05$ is usually thoughtlessly chosen

  many recent textbooks say it should not be used: always report the p-value, never report only a decision.

Combining Decision Theory and evidence based tests of hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1 \quad \Theta_0 \cap \Theta_1 = \emptyset, \Theta_0 \cup \Theta_1 = \Theta$$

- We have defined the power of a test as

$$\text{power of a test} = \mathbb{P}(\text{it will accept } H_1 \mid H_1 \text{ is true})$$

- We generalize it to a function of the parameter as follows:

Combining Decision Theory and evidence based tests of hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1 \quad \Theta_0 \cap \Theta_1 = \emptyset, \Theta_0 \cup \Theta_1 = \Theta$$

1. Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ is a vector of n iid random variables and $T(\mathbf{X})$ is a test statistic.
2. Let Ω_0 denote the acceptance and Ω_1 the rejection region of H_0. That is,

$$\text{if } \mathbf{X} \in \Omega_1 \quad \text{reject } H_0$$

3. The power of this test is the probability of rejecting H_0 as a function of θ:

$$\pi(\theta) = \mathbb{P}_\theta(\text{reject } H_0) = \mathbb{P}_\theta(\mathbf{X} \in \Omega_1) \tag{1}$$

Power

If α denotes the significance level, i.e., the probability value such that

$$p - value < \alpha \iff \text{reject } H_0$$

then

$$\alpha = \mathbb{P}_{\theta \in \Theta_0}(\mathbf{X} \in \Omega_1) = \pi(\theta),\ \theta \in \Theta_0$$

and

$$1 - \beta = \mathbb{P}_{\theta \in \Theta_1}(\mathbf{X} \in \Omega_1) = \pi(\theta),\ \theta \in \Theta_1$$

Maximal Power Tests

- So far, a test statistic $T(\mathbf{X})$ had to be given in advance.
- The core question in hypothesis testing is: what is the optimal test for a hypothesis?
- Ideally one would like to take the decision that has minimum Type I and II errors. But,

$$\alpha + \beta \neq 1$$

  so we cannot minimize them at the same time.
- Convention: Control the probability of Type I Error at a certain fixed level (α, significance) and find a test with minimal Type II error or maximal power.

Likelihood Ratio Test and Neyman-Pearson

- Do such tests exist?
- The answer comes from the *Neyman-Pearson* Lemma:
  - Assume $\mathbf{X} \sim f_\theta(\mathbf{x})$ and consider two simple hypotheses:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1$$

  - Denote the **likelihood ratio** as

$$\lambda(\mathbf{x}) = \frac{L(\theta_1; \mathbf{x})}{L(\theta_0; \mathbf{x})} = \frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})}$$

  - Then, the likelihood ration test with rejection region

$$\Omega_1 = \left\{ \mathbf{x} : \lambda(\mathbf{x}) = \frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} \geqslant C \right\} \tag{2}$$

    is the most powerful (MP) test among all tests at significance levels not larger than α, where

$$\alpha = \mathbb{P}_{\theta_0}(\lambda(\mathbf{X}) \geqslant C).$$

- The Neyman-Pearson Lemma was formulated for a fixed *critical value* C rather than α.
- For a continuous distribution $f_\theta(\mathbf{x})$, this is the same since α is a one-to-one function of C.
- We should be careful for discrete distributions, where $\lambda(\mathbf{x})$ can only take discrete values so that a critical value is the minimal possible C such that

$$\mathbb{P}_{\theta_0}(\lambda(\mathbf{X}) \geqslant C) \leqslant \alpha$$

- In this case, the resulting test will be over-conservative with true significance level $\alpha' < \alpha$.

Proof of Neyman-Pearson Lemma

We consider the case of continuous $f_\theta(\mathbf{x})$. For discrete, we replace integrals with sums.

Let π be the power of the LRT (2), that is,

$$\pi = \mathbb{P}_{\theta_1}(\lambda(\mathbf{X}) \geqslant C) = \mathbb{P}_{\theta_1}(\mathbf{X} \in \Omega_1) = \int_{\Omega_1} f_{\theta_1}(\mathbf{x}) d\mathbf{x}$$

Consider any other test at level $\alpha' \leqslant \alpha$ and let Ω_1' and π' be its rejection region and power, resp.

$$\alpha' = \mathbb{P}_{\theta_0}(\mathbf{X} \in \Omega_1') = \int_{\Omega_1'} f_{\theta_0}(\mathbf{x}) d\mathbf{x}, \quad \pi' = \mathbb{P}_{\theta_1}(\mathbf{X} \in \Omega_1') = \int_{\Omega_1'} f_{\theta_1}(\mathbf{x}) d\mathbf{x}$$

Proof of Neyman-Pearson Lemma (ctd)

We want to show $\pi \geqslant \pi'$:

$$\begin{aligned}
\pi - \pi' &= \int_{\Omega_1} f_{\theta_1}(\mathbf{x})d\mathbf{x} - \int_{\Omega_1'} f_{\theta_1}(\mathbf{x})d\mathbf{x} \\
&= \int_{\Omega_1 \cap \Omega_0'} f_{\theta_1}(\mathbf{x})d\mathbf{x} + \int_{\Omega_1 \cap \Omega_1'} f_{\theta_1}(\mathbf{x})d\mathbf{x} \\
&\quad - \int_{\Omega_1' \cap \Omega_1} f_{\theta_1}(\mathbf{x})d\mathbf{x} - \int_{\Omega_1' \cap \Omega_0} f_{\theta_1}(\mathbf{x})d\mathbf{x} \\
&= \int_{\Omega_1 \cap \Omega_0'} f_{\theta_1}(\mathbf{x})d\mathbf{x} - \int_{\Omega_1' \cap \Omega_0} f_{\theta_1}(\mathbf{x})d\mathbf{x}
\end{aligned}$$

(2) yields

$$f_{\theta_1}(\mathbf{x}) \geqslant C f_{\theta_0}(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega_1$$

and

$$f_{\theta_1}(\mathbf{x}) < C f_{\theta_0}(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega_0$$

Proof of Neyman-Pearson Lemma (ctd)

Therefore,

$$\begin{aligned}
\pi - \pi' &\geqslant C\left(\int_{\Omega_1\cap\Omega_0'} f_{\theta_0}(\mathbf{x})d\mathbf{x} - \int_{\Omega_1'\cap\Omega_0} f_{\theta_0}(\mathbf{x})d\mathbf{x}\right)\\
&= C\Bigg(\int_{\Omega_1\cap\Omega_0'} f_{\theta_0}(\mathbf{x})d\mathbf{x} + \int_{\Omega_1\cap\Omega_1'} f_{\theta_0}(\mathbf{x})d\mathbf{x}\\
&\quad - \int_{\Omega_1'\cap\Omega_1} f_{\theta_0}(\mathbf{x})d\mathbf{x} - \int_{\Omega_1'\cap\Omega_0} f_{\theta_0}(\mathbf{x})d\mathbf{x}\Bigg)\\
&= C\left(\int_{\Omega_1} f_{\theta_0}(\mathbf{x})d\mathbf{x} - \int_{\Omega_1'} f_{\theta_0}(\mathbf{x})d\mathbf{x}\right)\\
&= \alpha - \alpha' \geqslant 0 \quad \square
\end{aligned}$$

MP test for the normal mean

1. $X_1, \dots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$, σ is known.
2. We want to test $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_1$, $\mu_1 > \mu_0$.
3. The LR is

$$
\begin{aligned}
\lambda(\mathbf{x}) &= \frac{L(\mu_1; \mathbf{x})}{L(\mu_0; \mathbf{x})} \\
&= \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \mu_1)^2 - \sum_{i=1}^{n}(x_i - \mu_0)^2\right)\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\mu_0 - \mu_1)(2x_i - \mu_0 - \mu_1)\right) \\
&= \exp\left(\frac{1}{\sigma^2}(\mu_1 - \mu_0)n\left(\bar{x}_n - \frac{\mu_1 + \mu_0}{2}\right)\right)
\end{aligned}
$$

MP test for the normal mean

4. By NP Lemma, the MP test at level α rejects H_0 if

$$\lambda(\mathbf{X}) \geqslant C$$

with C satisfying

$$\mathbb{P}_{\mu_0}(\lambda(\mathbf{X}) \geqslant C) = \alpha$$

5. To solve this and find C we have to find the distribution of the LR under the null: typically difficult (or at least tedious).

MP test for the normal mean

6. Alternatively, we observe $\lambda(\mathbf{x})$ is an increasing function of $\bar{x}_n$ for $\mu_1 > \mu_0$ and that
$$\lambda(\mathbf{x}) \geqslant C \quad \iff \quad \bar{x}_n \geqslant C^\star$$
where
$$C = \exp\left(-\frac{1}{\sigma^2}(\mu_1 - \mu_0)n\left(C^\star - \frac{\mu_1 + \mu_0}{2}\right)\right)$$
7. The MP test at level α can then be re-written in terms of $\bar{X}_n$: reject H_0 if
$$\bar{X}_n \geqslant C^\star$$
with C satisfying
$$\mathbb{P}_{\mu_0}(\bar{X}_n \geqslant C^\star) = \mathbb{P}_{\mu_0}(\lambda(\mathbf{X}) \geqslant C) = \alpha$$

MP test for the proportion

Let $X_1, \ldots, X_n$ be iid Bernoulli(p) and suppose we want to test

$$H_0 : p = p_0 \quad vs \quad H_1 : p = p_1 > p_0$$

The likelihood ratio is

$$\lambda(\mathbf{x}) = \frac{L(p_1; \mathbf{x})}{L(p_0; \mathbf{x})} = \left(\frac{p_1}{p_0}\right)^{\sum_{i=1}^n x_i} \left(\frac{1-p_1}{1-p_0}\right)^{n-\sum_{i=1}^n x_i}$$

which is an increasing function of $\sum_i x_i$.

Therefore the LRT statistic is $\sum_{i=1}^n X_i$ and the corresponding MP test is

$$\text{reject } H_0 \quad \text{if} \quad \sum_{i=1}^n X_i \geqslant C$$

MP test for the proportion

The LRT is the same as the one we have already used.

Under the null,

$$\sum_{i=1}^{n} X_i \sim_{H_0} Bin(n, p_0)$$

so to find the critical value we have to proceed as for discrete distributions.

For large sample size n, we can use the normal approximation, in which case the LRT rejects H_0 if

$$\sum_{i=1}^{n} X_i \geqslant C = np_0 + z_\alpha \sqrt{np_0(1 - p_0)}$$

Uniformly Most Powerful tests

Suppose we consider the composite hypotheses:

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

Definition

A test is called a *uniformly most powerful* (UMP) among all tests at level α if its power satisfies

1. $\sup_{\theta \in \Theta_0} \pi(\theta) = \alpha$
2. $\pi(\theta) \geqslant \pi_1(\theta)$ for all $\theta \in \Theta_1$, where $\pi_1(\theta)$ is a power function of any other test at a level not larger than α.

Uniformly Most Powerful tests

Do UMP tests exist?

Example: Suppose $X_1, \dots, X_n$ are iide $\mathcal{N}(\mu, \sigma^2)$ with known σ and we want to test

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu > \mu_0$$

- Fix an arbitrary $\mu_1 > \mu_0$ and test the hypotheses

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu = \mu_1$$

  at level α.

- We already know the MP (LRT) test at level α rejects H_0 if

$$\bar{X}_n \geqslant \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

- This test does not depend on the *specific* μ_1 so that it is the MP for any $\mu_1 > \mu_0$. Hence, it is the UMP test.

Uniformly Most Powerful tests

- UMP tests most often do not exist.
- For example, for multivariate parameters, UMP tests do not exist except in singular cases
- **Example:** Consider again a normal sample but this time the variance is unknown and we want to test

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu = \mu_1$$

at level α, for $\mu_1 > \mu_0$

- In this case, the hypotheses are no longer simple:

$$\Theta_0 = \{(\mu_0, \sigma^2) : \sigma \geqslant 0\}, \quad \Theta_1 = \{(\mu_0, \sigma^2) : \sigma \geqslant 0\}$$

Uniformly Most Powerful tests

- The UMP test, if it exists, should be the most powerful test uniformly for all $\sigma \geqslant 0.$ most often do not exist.
- However, for any given σ, the corresponding MP test rejects H_0 if

$$\bar{X}_n \geqslant \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

  and depends on σ.
- Thus, even though we are only interested in μ and σ is a nuisance parameter, it is accounted for in the MP test, which cannot be uniformly MP across all σ values.

Monotone Likelihood Ratio and UMPs

- There is a class of hypothesis testing problems for a one parameter θ where a UMP test exists.
- A family of distributions $\{f_\theta, \theta \in \Theta\}$ with a one-dimensional parameter θ has a monotone likelihood ratio in a statistic $T(\mathbf{X})$ if for any $\theta_1 < \theta_2$, the likelihood ratio $f_{\theta_2}(\mathbf{x})/f_{\theta_1}(\mathbf{x})$ is a non-decreasing function of $T(\mathbf{x})$.
- (**HW**) Show that the one-parameter exponential family has a monotone likelihood ratio in a sufficient statistic $T(\mathbf{X})$ if the natural parameter $w(\theta)$ is a non-decreasing function in θ.

Monotone Likelihood Ratio and UMPs

Theorem

Let $\mathbf{X} \sim f_\theta(\mathbf{x})$, where f_θ belongs to a family of distributions with monotone likelihood ratio in a statistic $T(\mathbf{X})$. Then, there exists a UMP test at level α for testing the one-sided hypothesis $H_0 : \theta \leqslant \theta_0$ vs. the one-sided hypothesis $H_1 : \theta > \theta_0$, where H_0 is rejected if

$$T(\mathbf{X}) \geqslant C \quad \textit{and} \quad \mathbb{P}_{\theta_0}(T(\mathbf{X}) \geqslant C) = \alpha$$

(with the obvious modifications for discrete distributions).

This can be easily modified for testing $H_0 : \theta \geqslant \theta_0$ vs $H_1 : \theta < \theta_0$, in which case the UMP test rejects H_0 if

$$T(\mathbf{X}) \leqslant C, \quad \text{where} \quad \alpha = \mathbb{P}_{\theta_0}(T(\mathbf{X}) \leqslant C)$$

Monotone Likelihood Ratio and UMPs

This theorem says that if the family $\{f_\theta, \theta \in \Theta\}$ has a monotone likelihood ratio in a statistic $T(\mathbf{X})$ and the tested hypotheses are one-sided, then

- a UMP test exists
- $T(\mathbf{X})$ can be used as a test statistic
- To calculate the corresponding critical value one should use the distribution of $T(\mathbf{X})$ fro $\theta = \theta_0$.

Example: normal data with known variance

- A normal random sample from $\mathcal{N}(\mu, \sigma^2)$ with known σ^2 has a monotone ratio in $\bar{X}$
- Hence, the UMP test at level α for testing $H_0 : \mu \geqslant \mu_0$ vs $H_1 : \mu < \mu_0$ is to reject the null if
$$\bar{X} \leqslant \mu_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$
- Same result as before via a different route.

Example: exponential data

- Suppose $X_1, \ldots, X_n$ iid $\exp(\theta), f_\theta(x) = \theta e^{-\theta x}$.
- It has monotone likelihood ratio in $-\sum_{i=1}^n X_i$
- Hence, the UMP test at level α for testing $H_0 : \theta \leqslant \theta_0$ vs $H_1 : \theta > \theta_0$ is to reject the null if

  $$-\sum_{i=1}^n X_i \geqslant -C \iff \sum_{i=1}^n X_i < C$$

  where

  $$\alpha = \mathbb{P}_{\theta_0}(T(\mathbf{X}) \leqslant C)$$

- $\theta X_i \sim \exp(1) = \chi^2(2)/2$, therefore

  $$2\theta_0 \sum_{i=1}^n X_i \sim_{H_0} \chi^2(2n)$$

  with critical value

  $$C = \frac{1}{2\theta_0} \chi^2_{1-\alpha}(2n)$$

Generalized Likelihood Ratio Tests

- Let $\mathbf{X} = (X_1, \ldots, X_n) \sim f_\theta(\mathbf{x})$, $\theta \in \Theta$ and test

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_1$$

  with $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0 \cup \Theta_1 = \Theta$.

- A generalization of the likelihood ratio for composite hypotheses would be

$$\lambda^\star(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_1} L(\theta, \mathbf{x})}{\sup_{\theta \in \Theta_0} L(\theta, \mathbf{x})}$$

- The larger the value of $\lambda^\star(\mathbf{x})$ is the stronger the evidence against H_0, so it is a reasonable test statistic

Generalized Likelihood Ratio Tests

- It is more convenient to use the equivalent statistic

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta} L(\theta, \mathbf{x})}{\sup_{\theta \in \Theta_0} L(\theta, \mathbf{x})} = \frac{\sup_{\theta \in \Theta} f_\theta(\mathbf{x})}{\sup_{\theta \in \Theta_0} f_\theta(\mathbf{x})}$$

  where

$$\lambda(\mathbf{x}) = \max(\lambda^\star(\mathbf{x}), 1)$$

  is a nondecreasing function of $\lambda^\star(\mathbf{x})$.

- $\lambda(\mathbf{x})$ is called a generalized likelihood ratio (GLR).

- The corresponding generalized likelihood ratio test (GLRT) at level α rejects the null if

$$\lambda(\mathbf{x}) \geqslant C$$

  where C satisfies

$$\sup_{\theta \in \Theta_0} \mathbb{P}(\lambda(\mathbf{x}) \geqslant C) = \alpha$$

Calculating the GLRT

To calculate $\lambda(\mathbf{x})$ and the GLRT:

1. Find the MLE $\hat{\theta}$ of θ to calculate the numerator $\sup_{\theta\in\Theta} L(\theta, \mathbf{x}) = L(\hat{\theta}, \mathbf{x})$
2. Find the MLE $\hat{\theta}_0$ of θ_0 under the restriction $\theta \in \Theta_0$ to calculate the denominator $\sup_{\theta\in\Theta_0} L(\theta, \mathbf{x}) = L(\hat{\theta}_0, \mathbf{x})$
3. Form the generalized likelihood ratio

$$\lambda(\mathbf{x}) = \frac{L(\hat{\theta}, \mathbf{x})}{L(\hat{\theta}_0, \mathbf{x})}$$

   and find an equivalent simpler test statistic $T(\mathbf{x})$ if possible such that $\lambda(\mathbf{x})$ is its increasing function

4. Find the corresponding critical value for $T(\mathbf{x})$ solving

$$\alpha = sup_{\theta\in\Theta_0} \mathbb{P}(T(\mathbf{X}) \geqslant C)$$

Well known GLRTs

1. one and two-sample t-tests for normal means
2. χ^2 test for the normal variance
3. F-test for normal variances
4. F-test for comparing nested models in regression
5. Pearson's χ^2-test for goodness of fit

Example: one-sample t-test

1. $X_1, \ldots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$, σ unknown.
2. We want to test

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$$

3. Normal likelihood:

$$L(\mu, \sigma; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} \tag{3}$$

4. The MLEs are

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2$$

Example: one-sample t-test

5. Plug them in (3) to get

$$L(\hat{\mu}, \hat{\sigma}; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \right)^n e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\hat{\sigma}^2}} = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \right)^n e^{-\frac{n}{2}}$$

6. Under the null hypothesis, $\hat{\mu} = \mu_0$, so $\hat{\sigma}^2 = \sum_{i=1}^2 (x_i - \mu_0)^2 / n$ and,

$$L(\hat{\mu}_0, \hat{\sigma}_0; \mathbf{x}) = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}_0^2}} \right)^n e^{-\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\hat{\sigma}_0^2}} = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}_0^2}} \right)^n e^{-\frac{n}{2}}$$

Example: one-sample t-test

7. The GLR is

$$
\begin{aligned}
\lambda(\mathbf{x}) &= \frac{L(\hat{\mu}, \hat{\sigma}; \mathbf{x})}{L(\hat{\mu}_0, \hat{\sigma}_0; \mathbf{x})} \\
&= \left(\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^{n/2} \\
&= \left(\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^{n/2} \\
&= \left(1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)^{n/2} \\
&= \left(1 + \frac{1}{n-1} \left(\frac{\bar{x} - \mu_0}{s_0/\sqrt{n}} \right)^2 \right)^{n/2}
\end{aligned}
$$

where $s_0^2 = \sum_{i=1}^n (x_i - \mu_0)^2/(n-1)$.

Example: one-sample t-test

8. The GLR is an increasing function of $|T(\mathbf{x})|$, where

$$T(\mathbf{x}) = \frac{\bar{x} - \mu_0}{s_0/\sqrt{n}}$$

9. We reject $H_0 : \mu = \mu_0$ if

$$|T(\mathbf{x})| \geqslant C, \quad \text{where} \quad \alpha = \mathbb{P}_{\mu_0}(|T(\mathbf{X})| \geqslant C)$$

10. For $\mu = \mu_0$, $T(\mathbf{Y}) \sim t(n-1)$. Therefore, the GLRT rejects H_0 if

$$|T(\mathbf{x})| = \frac{\bar{x} - \mu_0}{s_0/\sqrt{n}} \geqslant t_{\alpha/2}(n-1)$$

which is the well-known one-sample t-test.

HW

1. Derive the GLRT for the normal variance: Assume $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, where both μ and σ are unknown. We want to test

$$H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 \neq \sigma_0^2$$

2. Let $X_1, \ldots, X_n$ be iid Uniform$(0, \theta)$
    1. Derive the MP test at level α for testing

    $$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1, \theta_1 > \theta_0$$

    2. Calculate the power of the MP test.

HW

1. Let $X_1, \dots, X_n$ be iid from a distribution with density

$$f_\theta(x) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}},\ y \geqslant 0, \theta > 0$$

   1. Derive the MP test at level α for testing two simple hypoheses

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1, \theta_1 > \theta_0$$

   2. Is there a UMP test at level α for testing the one-sided composite hypothesis

$$H_0 : \theta \leqslant \theta_0 \quad vs \quad H_1 : \theta > \theta_0$$

   What is its power function? (Hint: Show $X_i^2 \sim \exp(1/2\theta)$, so that $\sum_i X_i^2 \sim \theta\chi^2(2n)$).

HW

Let $X_1, \dots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$.

1. Assume first that μ is known.
    1. Find an MP test at level α for testing two simple hypoheses
    $$H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 = \sigma_1^2,\ \sigma_1 > \sigma_0$$
    2. Show that the MP test is a UMP test for testing
    $$H_0 : \sigma^2 \leqslant \sigma_0^2 \quad vs \quad H_1 : \sigma^2 > \sigma_0^2$$
    (Hint: $\sum_i (X_i - \mu)^2 \sim \sigma^2 \chi^2(n)$)
2. Now assume μ is unknown.
    1. Is there an MP test at level α for testing?
    $$H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 = \sigma_1^2,\ \sigma_1 > \sigma_0$$
    If not, find the corresponding GLRT.
    2. Is the above GLR test also a GLRT for testing the one-sided hypothesis?
    $$H_0 : \sigma^2 \leqslant \sigma_0^2 \quad vs \quad H_1 : \sigma^2 > \sigma_0^2$$
    3. Find the GLRT at level α for testing
    $$H_0 : \sigma^2 \geqslant \sigma_0^2 \quad vs \quad H_1 : \sigma^2 < \sigma_0^2$$