Parametric Statistics

Hypothesis Testing

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

December 18, 2023

Lecture Summary

Statistical Inference

Given a probability model $f(x \mid \theta)$ (and possibly a prior $p(\theta)$) we may be interested in

- Parameter estimation
- Making decisions - Hypothesis testing, Chapter 9
  - e.g., If the disease affects 2% or more of the population, the state will launch a costly public health campaign.
  - Do we have evidence that θ is higher than 2% ?
- Other things like, prediction, experimental design, etc.

Hypothesis Testing

Should you get the coin?

- Your friend tells you that they will give you a fair coin for coin flipping.
- You are not sure the coin is fair.
- Your friend tells you you can test it.
- You toss it 100 times, you get 99 heads.
- Do you think the coin is fair?

Hypothesis Testing

Should you get the coin?

- Your friend tells you that they will give you a fair coin for coin flipping.
- You are not sure the coin is fair.
- Your friend tells you you can test it.
- You toss it 100 times, you get 99 heads.
- Do you think the coin is fair?

$$P(X = 99|\theta = 0.5) \approx 0$$

Probability of observing the data given that our friend is telling the truth is almost zero.

Steps to testing hypotheses: Neyman - Pearson

1. Make a claim for $\theta \in \Omega$.
2. Pick a null hypothesis H_0 and an alternative hypothesis H_1.
   - $H_0 : \theta = 0.5 \quad H_1 : \theta \neq 0.5$
3. Choose a significance level α (usually $\alpha = 0.05$ or 0.01).
4. Collect data.
   - Toss the coin 10 times.
5. Compute a p-value,
   $p = P$ (observing data at least as extreme as ours $\mid H_0$ is true).
6. State your conclusion.
   - If $p < \alpha$, "reject" the null hypothesis H_0 in favor of the alternative H_A. We say our result is statistically significant in this case!
   - Otherwise, "fail to reject" the null hypothesis H_0.

Types of Hypotheses

In general, consider a problem in which we wish to test the following hypotheses:

$$H_0 : \theta \in \Omega_0, \quad \text{and} \quad H_1 : \theta \in \Omega_1.$$

Simple and Composite Hypotheses.

If Ω_i contains just a single value of θ, then H_i is a simple hypothesis. If the set Ω_i contains more than one value of θ, then H_i is a composite hypothesis.

One-sided and Two-sided Hypotheses.

Let θ be a one-dimensional parameter. One-sided null hypotheses are of the form $H_0 : \theta \leq \theta_0$ or $H_0 : \theta \geq \theta_0$, with the corresponding one-sided alternative hypotheses being $H_1 : \theta > \theta_0$ or $H_1 : \theta < \theta_0$. When the null hypothesis is simple, such, the alternative hypothesis is usually two-sided, $H_1 : \theta \neq \theta_0$.

Critical region and the test statistic.

In general, consider a problem in which we wish to test the following hypotheses:

$$H_0 : \theta \in \Omega_0, \quad \text{and} \quad H_1 : \theta \in \Omega_1.$$

Suppose that we can observe a random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ drawn from a distribution that involves the unknown parameter θ. Let S denote the sample space of the n-dimensional random vector $\boldsymbol{X}$. In other words, S is the set of all possible values of the random sample.
The test procedure specifies a partitioning of the sample space S into two subsets.

- S_1 contains the values of $\boldsymbol{X}$ for which she will reject H_0.
- S_0 contains the values of $\boldsymbol{X}$ for which she will not reject H_0.
- S_1 is called the **critical region** of the test.

Critical Region

In our coin example,

$$H_0 : \theta \in \Omega_0 = \{0.5\}, \quad \text{and} \quad H_1 : \theta \in \Omega_1 = [0, 0.5) \cup (0.5, 1].$$

Let's say our data is 10 coin tosses.
Let's say we decide to reject the null hypothesis if we have less than 3 or more than 7 heads.
Then S_1 : All the random samples where $\sum_{i=1}^{N} X_i < 3$ or $\sum_{i=1}^{N} X_i > 7$.

Test Statistic / Rejection Region

In most hypothesis-testing problems, the critical region is defined in terms of a statistic, $T = r(\boldsymbol{X})$.

Rejection Region

Let $\boldsymbol{X}$ be a random sample from a distribution that depends on a parameter θ. Let $T = r(\boldsymbol{X})$ be a statistic, and let R be a subset of the real line. Suppose that a test procedure for our hypotheses is of the form "reject H_0 if $T \in R$."

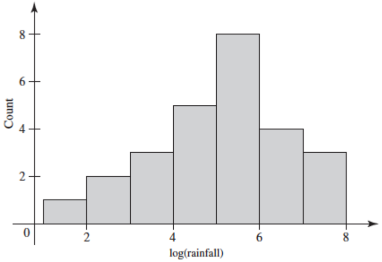
- T is a test statistic.
- R is the rejection region of the test.

In the coin example, $\sum_{i=1}^{N} X_i$ is a test statistic.

Example

Rain from Seeded Clouds

- Without seeding: $\mu = 4$.
- With seeding: $\mu = 5.136$.
- $\sum_{i=1}^{26} (X_i - \bar{X}_n)^2 = 40$.
- n=26.
- We want to answer the question: Is $\mu > 4$?



Power function

- $H_0 : \mu \leq 4 \quad H_1 : \mu > 4$.
- Let's say we decide to reject the null if $\bar{X}_n > 4 + c$.
- Rejection region: $(4 + c, \infty)$

Power function

- $H_0 : \mu \leq 4 \quad H_1 : \mu > 4.$
- Let's say we decide to reject the null if $\bar{X}_n > 4 + c$.
- Rejection region: $(4 + c, \infty)$

Power Function.

Let δ be a test procedure. The function $\pi(\theta \mid \delta)$ is called the power function of the test δ. If S_1 denotes the critical region of δ, then the power function $\pi(\theta \mid \delta)$ is determined by the relation

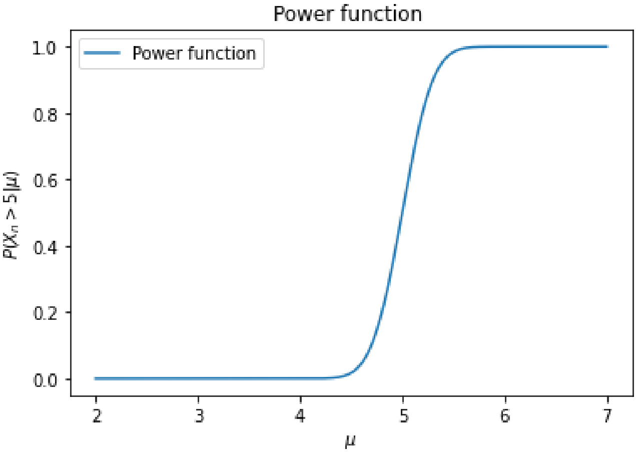
$$\pi(\theta \mid \delta) = \Pr\left(\boldsymbol{X} \in S_1 \mid \theta\right) \text{ for } \theta \in \Omega.$$

If δ is described in terms of a test statistic T and rejection region R, the power function is

$$\pi(\theta \mid \delta) = \Pr(T \in R \mid \theta) \quad \text{for} \quad \theta \in \Omega.$$

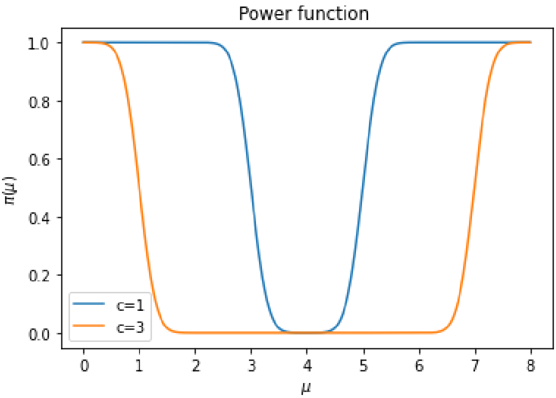
Power function

- $H_0 : \mu \leq 4 \quad H_1 : \mu > 4.$
- Let's say we decide to reject the null if $\bar{X}_n > 4 + 1$.
- Rejection region: $(5, \infty)$



Power function

- How about $H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$.
- Let's say we decide to reject the null if $|\bar{X}_n - \mu_0| > c$.
- Rejection region:$(-\infty, c - \mu_0) \cup (c + \mu_0, \infty)$



Types of Errors

- Type I error: Wrongly reject the null hypothesis.
- $\bar{X}_n \in (-\infty, c - \mu_0) \cup (c + \mu_0, \infty)$, but $\mu = \mu_0$.

- Type II error: Wrongly decide to not reject the null hypothesis.
- $\bar{X}_n \in [c - \mu_0, c + \mu_0]$, but $\mu \neq \mu_0$.

Relation to power function:

- If $\theta \in \Omega_0 : \pi(\theta \mid \delta) =$ probability of type I error
- If $\theta \in \Omega_1 : 1 - \pi(\theta \mid \delta) =$ probability of type II error

Size and Level of Tests

Want probability of both types of errors to be small

- Want $\pi(\theta \mid \delta)$ to be small for $\theta \in \Omega_0$ and large for $\theta \in \Omega_1$.
- Generally there is a trade-off between these probabilities.
- A popular method: Choose a number α_0 and pick δ such that

$$\pi(\theta \mid \delta) \leq \alpha_0 \quad \text{for } \theta \in \Omega_0$$

That is, we put an upper bound on the probability of type I error.

- The test is then called level α_0 test or we say that the test has significance level α_0
- The size $\alpha(\delta)$ of a test is defined as

$$\alpha(\delta) = \sup_{\theta \in \Omega_0} \pi(\theta \mid \delta)$$

- A test δ is a level α_0 test if and only if $\alpha(\delta) \leq \alpha_0$
- When the null hypothesis is simple $(H_0 : \theta = \theta_0)$ then $\alpha(\delta) = \pi\left(\theta_0 \mid \delta\right)$