Statistic

Zambelli Lorenzo
BSc Applied Mathematics

September 2021-October 2021

1 Introduction

Definition 1 ***Parameter*** *A parameter is a constant that defines the population pmf/pdf* $f(x)$

Definition 2 ***Statistic*** *A statistic is an observable function* $T : \mathbb{R}^n \to \mathbb{R}$*of a random sample (of a collection of random variables) such that* T *does not depend on any unknown parameters*

Definition 3

$$\text{sample mean} : \quad \overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$$
$$\text{sample variance} : \quad S_n^2 = \frac{1}{n-1} \sum_{i \le n} (X_i - \overline{X})^2$$

Lemma 4

$$S_n^2 = \frac{1}{n-1} \sum_{i \le n} X_i^2 - \frac{n}{n-1} \overline{X}_n^2$$

Theorem 5 ***Unbiasedness of sample mean variance*** *Let* $X_1, ..., X_n$ *be independent and identically distributed with mean* μ *and variance* σ^2*. Then,*

1. $\mathbb{E}(\overline{X}_n) = \mu$
2. $\mathbb{E}(S_n^2) = \sigma^2$

Remark:

- We write $X_1, ..., X_n \sim \mathcal{F}_\theta$ to indicate that $X_1, ..., X_n$ is a random sample of size n from a distribution $\mathcal{F}_\theta$ that depends on the parameter(s) θ
- For a given random sample $X_1, ..., X_n \sim \mathcal{F}_\theta$ we use for the joint density the notation $f_\theta(x_1, ..., x_n)$ instead of $f_{X_1,...,X_n}(x_1, ..., x_n)$ and for the density of X_3 at x_3 as $f_\theta(x_3)$ instead of $f_{X_3}(x_3)$

Definition 6 ***Random Sample*** *A random sample of size* n *is a sequence* $X_1, ..., X_n$ *of independent random variables all with the same pdf/pmf, say say* $f(x)$*. We thus have*

$$f_\theta(x_1, ..., x_n) = \prod_{1 \le i \le n} f_\theta(x_i)$$

we say that f *is the* ***population pdf/pmf***

Terminology

- We use small letters for the realizations of random variables.
- Given realizations $x_1, ..., x_n$, we define: $\overline{x_n} := \frac{1}{n} \sum_{1 \le i \le n} x_i$

1.1 Useful knowledge form probability theory

Definition 7 (Quatiles) *Consider a random variable with distribution $\mathcal{F}_\theta$. The α-quatile q_α of the distribution $\mathcal{F}_\theta$ is defined as*

$$\mathbb{P}_\theta(X \leq q_\alpha) = \alpha \Leftrightarrow F_\theta(q_\alpha) = \alpha$$

where F_θ is the cumulative distribution function.

Remark: For symmetric distributions (with $f_\theta(x) = f_\theta(-x)$) we have that $q_\alpha = -q_{1-\alpha}$

Definition 8 (Some properties of the normal distributions) *Consider a Gaussian distribution random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$:*

- $(X + b) \sim \mathcal{N}(\mu + b, \sigma^2)$
- $a \cdot X \sim \mathcal{N}(a \cdot \mu, a^2 \cdot \sigma^2)$
- $a(X + b) \sim \mathcal{N}(a \cdot (\mu + b), a^2 \cdot \sigma^2)$

Now consider a random sample $X_i \sim \mathcal{N}(\mu, \sigma^2$ for all $i = 1, ..., n$, then

- $\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right)$
- $(\overline{X}_n - \mu) \sim \mathcal{N}\left(0, \frac{1}{n}\sigma^2\right)$
- $\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu)\mathcal{N}(0, 1)$

Definition 9 *A sequence of $X_1, X_2, ...$ of random variables converges in probability to a constant $c \in \mathbb{R}$ if $\forall \epsilon > 0$:*

$$\mathbb{P}(|X_n - c| > \epsilon) \to 0$$

which can be read as: "as n gets larger, it becomes very unlikely that Xn is far from c". We write $X_n \xrightarrow[n\to\infty]{\mathbb{P}} c$. Instead of "converges in probability" we sometimes also say converges weakly.

Theorem 10 ***Weak Law of Large Numbers*** *Let $X_1, X_2, ...$ independent and identically distributed with $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$ then*

$$\overline{X}_n \underset{\mathbb{P}}{\to} \mu \qquad \lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > \epsilon) = 0$$

Theorem 11 (Law of Large Numbers) *Consider a random sample from a distribution $\mathcal{F}_\theta$*

$$X_1, ..., X_n \sim \mathcal{F}_\theta \qquad \text{or short: } X \sim \mathcal{F}_\theta$$

then for $n \to \infty$: $\overline{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}(X_1)$ i.e convergence in probability.

More Generally, we have for any $k \in \mathbb{N}$:

$$\text{for } n \to \infty: \quad \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{\mathbb{P}} \mathbb{E}[X_1^k]$$

Definition 12 ***converges in distribution*** *A sequence of random variables $X_1, X_2, ...$ converges in distribution to a random variable X if*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for every $x \in \mathbb{R}$ at which $F_X(x)$ is continuous. We denote this by

$$X_n \xrightarrow[d]{n\to\infty} X$$

Lemma 13 *If X is continuous and $X_n \xrightarrow[d]{n\to\infty} X$, then*

$$\mathbb{P}(X_n = x) \xrightarrow{n\to\infty} 0$$

for all $x \in \mathbb{R}$

Proposition 14 *If X is continuous and $X_n \xrightarrow[d]{n\to\infty} X$, then for every interval $I \subset \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}(X_n \in I) = \mathbb{P}(X \in I)$$

Theorem 15 ***Central Limit Theorem*** *Let $X_1, X_2, ...$ be independent and identically distributed with mean μ and variance σ^2 (both finite). Then,*

$$\sqrt{n} \cdot \frac{\overline{X} - \mu}{\sigma} \xrightarrow[d]{n\to\infty} Z, \qquad \textit{where } Z \sim \mathcal{N}(0,1)$$

Remarks:

$$\sqrt{n} \cdot \frac{\overline{X} - \mu}{\sigma} = \frac{\sum_{i=1}^{n} X_i - \mu n}{\sigma\sqrt{n}}$$

Theorem 16 ***Normal approximation to binomial*** *When n is large and p is not too close to 0 or 1, we have the approximation*

$$X \sim \mathrm{Bin}(n,p) \approx Y \sim \mathcal{N}(np, np(1-p))$$

where

$$\mathbb{P}(X \le b) \approx \int_{-\infty}^{b+\frac{1}{2}} f_Y(y)\,dy = F_Y\left(b + \frac{1}{2}\right), \quad \mathbb{P}(X \ge a) \approx \int_{a-\frac{1}{2}}^{\infty} f_Y(y)\,dy = 1 - F_Y\left(a - \frac{1}{2}\right)$$

this approximation holds if $n \ge 15$, $np \ge 5$ and $n(1-p) \ge 5$.

Theorem 17 ***Chebyshev Inequality*** *Let X an random variable,*

$$\mathbb{P}(|X - \mathbb{E}(X)| > x) \le \frac{\mathrm{Var}(X)}{x^2}, \quad x > 0$$

Theorem 18 (Markov Inequality) *For a single random variable $X \sim \mathcal{F}_\theta$ with sample space $S_X \subseteq \mathbb{R}_0^+$, we have for all $r > 0$ the Markov inequality:*

$$\mathbb{P}_\theta(X \ge r) \le \frac{E[X]}{r}$$

Definition 19 (Chi-Square distribution) *Consider a sample from a standard Gaussian distribution, $X \sim \mathcal{N}(0,1)$. Then the random variable:*

$$S = \sum_{i=1}^{n} X_i^2$$

is Chi-squared distributed with n degree of freedom, symbolically: $S \sim \chi_n^2$. And we have $\mathbb{E}(S) = n$ and $\text{Var}(S) = 2n$

Remark: for any gaussian sample $X_n \sim \mathcal{N}(\mu, \sigma^2)$

$$S = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2$$

Definition 20 (t-distribution) *Consider a standard Gaussian distributed random variable X and a Chi-squared distributed random variable S with n degree of freedom. If X and S are statistically independent, then the random variable*

$$T = \frac{X}{\sqrt{\frac{1}{n} S}}$$

is t-distributed with n degree of freedom, symbolically $T \sim t_n$ where $\mathbb{E}(T) = 0$ and fro $n > 2$ $\text{Var}(T) = \frac{n}{n-2}$.

Remark: For $n \to \infty$ $t_n \xrightarrow{D} \mathcal{N}(0,1)$

Definition 21 (F-distribution) *Consider two Chi-squared distributed random variable S_1 and S_2 with n_1 and n_2 degree of freedom. If S_1 and S_2 are statistically independent, then the random variable*

$$F = \frac{\frac{1}{n_1} S_1}{\frac{1}{n_2} S_2}$$

is F-distributed with parameters n_1 and n_2, symbolically $F \sim F_{n_1,n_2}$, where for $n_2 > 2$ $\mathbb{E}(F) = \frac{n_2}{n_2-2}$

Theorem 22 (Cauchy Schwartz- Inequality) *for two random variable Y, Z we have*

$$|\text{Cov}(Y,Z)| \leq \sqrt{\text{Var}(Y)\,\text{Var}(Z)}$$

Theorem 23 (Jensen's inequality) *Let $X \sim \mathcal{F}_\theta$ be a random variable on the possibly infinite interval (a,b) and let the function $g()$ be differentiable and convex on (a,b). If $\mathbb{E}(X)$ and $\mathbb{E}(g(X)$ both exist, then*

$$\mathbb{E}(g(X) \geq g(\mathbb{E}(X))$$

Definition 24 (Information inequality) *Let $X \sim \mathcal{F}_\theta$ be a random variable with $\theta \in \Theta$ and density $f_\theta()$. Moreover, let θ_0 be the true parameter. Then:*

$$\mathbb{E}_{\theta_0}(\log(f_{\theta_0}(X))) \geq \mathbb{E}_{\theta_0}(\log(f_\theta(X)))$$

Theorem 25 (continuous mapping theorem) *Given $\{X_n\}_{n\in\mathbb{N}}$ and a continuous function $g()$, we have:*

1. $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$
2. $X_n \xrightarrow{D} X \Rightarrow g(X_n) \xrightarrow{D} g(X)$

Theorem 26 (Slutsky's theorem) *For two sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ with*

$$X_n \xrightarrow{D} X \quad Y_n \xrightarrow{P} c$$

where X is a random variable and $c \in \mathbb{R}$ is a constant, we have

1. $X_n + Y_n \xrightarrow{D} X + c$
2. $X_n \cdot Y_n \xrightarrow{D} c \cdot X$
3. $\frac{X_n}{Y_n} \xrightarrow{D} \frac{1}{c}X$ *if* $c \neq 0$

1.2 Sufficiently of a Statistic

Definition 27 *A statistic T is called sufficient for θ if the conditional density of X given $T(X)$, $f_\theta(x|t(x))$ does not depend on θ. That is, if we have: $f_\theta(x|t(x)) = f(x|t(x))$*

Hence, a statistic T is called sufficient for θ if we do not lose any information about θ when 'summarizing'

The sufficiency principle:
Consider two random samples X and Y of size n from the same distribution $\mathcal{F}_\theta$ and a statistic T that is sufficient for θ. Given two realizations $X = x$ and $Y = y$ with $T(X) = T(Y)$, the inference about θ should be the same in both cases.

Theorem 28 (Factorization theorem) *Given a random sample $X \sim \mathcal{F}_\theta$, then T is a sufficient statistic for θ if and only if the joint density $f_\theta(x)$ of X can be factorized into:*

$$f_\theta(x) = g(t(x);\theta) \cdot h(x) \quad \text{for all } x = (x_1, ..., x_n) \in S_X$$

Definition 29 (Exponential family) *A distribution $\mathcal{F}_\theta$ with θ containing d parameters ($|\theta| = d$) belongs to the exponential family if the density f_θ of $\mathcal{F}_\theta$ can be decomposed into:*

$$f_\theta(x) = h(x) \cdot \exp\left\{ \sum_{d \leq j \leq 1} \mu_j(\theta) T_j(x) - A(\theta) \right\}$$

2 Estimators

The idea is how large should n be such that $\overline{X}_n$ approximates μ well?

Definition 30 *Let $X \sim \mathcal{F}_\theta$ be a random sample, then an* ***estimator*** *is a statistic $T(X)$ that is used to estimate the unknown parameter θ.*

Remark: If the purpose of the statistic is to estimate the parameter θ, the statistic is usually denoted $\hat{\theta}(X)$ or short $\hat{\theta}$.

2.1 Method of Moments (MM) Estimators

Consider a distribution $\mathcal{F}_\theta$, where θ covers d unknown parameters ($|\theta| = d$) and a random sample from this distribution $X_1, ..., X_n \sim \mathcal{F}_\theta$.

LLN implies for $k = 1, ..., d$: $\frac{1}{n}\sum_{1\leq i\leq n} X_i^k \xrightarrow[\mathbb{P}]{n\to\infty} \mathbb{E}[X_1^k]$

We then try to solve the system of d equations that follows from the LLN.

2.2 Likelihood and Maximum Likelihood

Let Θ denote the parameter space, i.e the space of all possible parameters θ

Definition 31 (Likelihood) *The likelihood (function) is defined as* $L : \Theta \to \mathbb{R}_0^+$ *with* $L(\theta) := f_\theta(x_1, ..., x_n)$

Remark:

- For any θ the likelihood tells us 'how likely' the realizations $x_1, ..., x_n$ are if θ is the true parameter.
- If the sample is from a discrete distribution, $L(\theta)$ is the probability of the realizations $x_1, .., x_n$
- If the sample is from a continuous distributions, then $f_\theta(x_1, ..., x_n)$ and $L(\theta)$ are no probabilities.

Definition 32 (Maximum Likelihood (ML) Estimator) *Given a random sample* $X_1, ..., X_n \sim \mathcal{F}_\theta$ *the Maximum Likelihood (ML) Estimator of* $\theta \in \Theta$ *is defined as:*

$$\hat{\theta}_{ML} := \text{argmax}_{\theta\in\Theta}\{L(\theta)\}$$

where $L(\theta) = f_\theta(x_1, ..., x_n)$ *is the likelihood*

Important Trick: It is computationally much easier to maximize the log-likelihood $\log(L(\theta))$. Since the logarithm is a monotone trasformation, we have:

$$\hat{\theta}_{ML} := \text{argmax}_{\theta\in\Theta}\{L(\theta)\} = \text{argmax}_{\theta\in\Theta}\{l(\theta)\}$$

where $l(\theta) = \log(L(\theta))$ is the log-likelihood

Definition 33 (Consistency of the ML estimator) *Consider a random sample* $X_n \sim \mathcal{F}_\theta$ *with* $\theta \in \Theta$ *and densinty* $f_\theta()$. *Let* θ_0 *denote the true parameter. Under regulatory conditions, the* ML *estimator is consistent for* θ_0

$$\hat{\theta}_{ML,n} \xrightarrow{P} \theta_0$$

Required conditions:

1. *The sample space* S_X *does not depend on* θ
2. θ_0 *is an interior point of* Θ

3. *The log-likelihood $l_X(\theta)$ is differentiable in θ*

4. *θ_0 is the unique solution of $l'_X(\theta) = 0$*

Definition 34 (Asymptotic Efficiency of the ML) *Given a random sample $X_n \sim \mathcal{F}_\theta$ with parameter space Θ. The ML estimator $\hat{\theta}_{ML,n}$ of θ is an efficient estimator if:*

$$\sqrt{n} \cdot (\hat{\theta}_{ML,n} - \theta) \xrightarrow{D} \mathcal{N}(0, \frac{1}{I(\theta)}$$

where $I(\theta)$ is the expected Fisher information, under the following regulatory condition

1. *The parameter space $\Theta \subset \mathbb{R}$ must be open*

2. *The density $f_\theta()$ must be 3-times differentiable w.r.t θ*

3. *The sample space S_X is not allowed to depend on θ*

2.3 Study the estimators

Definition 35 *The bias of the estimator $\hat{\theta}_n$ is defined as*

$$B(\hat{\theta}_n) = \mathbb{E}(\hat{\theta}_n) - \theta$$

Definition 36 *The estimator $\hat{\theta}_n$ is an unbiased estimator of θ if for all $n \in \mathbb{N} : \mathbb{E}(\hat{\theta}_n) = \theta$*

Definition 37 *The estimator $\hat{\theta}_n$ is an asymptotically unbiased estimator of θ if for $n \to \infty$: $\mathbb{E}(\hat{\theta}_n) \to \theta$*

Definition 38 (Mean Squared Error (MSE)) *The Mean Squared Error of $\hat{\theta}_n$ is defined as:*

$$\text{MSE}(\hat{\theta}_n) = \mathbb{E}\left[(\hat{\theta}_n - \theta)^2\right]$$

Remark: Note that $\text{MSE}(\hat{\theta}_n) = \text{Var}(\hat{\theta}_n) + B(\hat{\theta}_n)^2$

Definition 39 *Let $X \sim \mathcal{F}_\theta$ be a random sample, and $g : \Theta \to \mathbb{R}$ be a function. The statistic $T(X)$ is called an unbiased estimator of $g(\theta)$ if*

$$\mathbb{E}(T(X)) = g(\theta)$$

Theorem 40 (The Cramer-Rao Theorem) *Consider a sample of size n $X \sim \mathcal{F}_\theta$, and an unbiased estimator $\hat{\theta}$ of θ. Then (under certain regulatory condition)*

$$\text{Var}(\hat{\theta}) \geq \frac{1}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta} l_X(\theta)\right)^2\right]}$$

where $l_X(\theta)$ is the log-likelihood.

Remark:

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta} l_X(\theta)\right)^2\right] = n \cdot \mathbb{E}\left[\left(\frac{\partial}{\partial\theta} l_{X_1}(\theta)\right)^2\right] = -\mathbb{E}\left[\left(\frac{\partial^2}{\partial\theta^2} l_X(\theta)\right)^2\right]$$

Definition 41 (Expected Fisher information (of a sample of size $n = 1$)) *Given a random sample $X_n \sim \mathcal{F}_\theta$ we define the expected Fisher information (of a sample of size $n = 1$) as*

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta} l_{X_1}(\theta)\right)^2\right]$$

Definition 42 (Observerd Fisher information) *Slutsky's theorem allows us to replace the expected Fisher information $I(\theta)$ by the observer Fisher information $I(\hat{\theta}_{ML,n})$, because*

$$\hat{\theta}_{ML,n} \xrightarrow{D} \theta \Rightarrow I(\hat{\theta}_{ML,n}) \xrightarrow{D} I(\theta)$$

Theorem 43 (Rao-Blackwell Theorem) *Consider a random sample $X \sim \mathcal{F}_\theta$ and a function $g : \Theta \to \mathbb{R}$. If we have*

1. *The statistic $W = W(X)$ is unbiased estimator of $g(\theta)$*
2. *The statistic $T = T(X)$ is sufficient for θ*

we can define a new estimator

$$\phi(T) = \mathbb{E}(W|T)$$

with

1. $\mathbb{E}(\phi(T)) = g(\theta)$*, i.e $\phi(T)$ is an unbiased estimator of $g(\theta)$*
2. $\mathrm{Var}(\phi(T)) \leq \mathrm{Var}(W)$*, i.e the variance of $\phi(T)$ is potentially smaller than the variance of W*

2.3.1 Asymptotic Statistic

Definition 44 (Sequence of estimators) *Consider a random sample $X_n \sim \mathcal{F}_\theta$ with increasing sample size. Then, $\hat{\theta}_n$ is a estimator for θ in $X_n \sim \mathcal{F}_\theta$. We define a sequence of estimators of θ $\{\hat{\theta}\}_{n\in\mathbb{N}}$*

Definition 45 *Let $X_1, ..., X_n$ be a random sample of pmf/pdf with parameter θ. We say that $\hat{\theta}_n$ is **consistent estimator** of θ if*

$$\forall\theta \in \Theta : \hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}} \theta$$

Proposition 46 *Given a random sample $X_n \sim \mathcal{F}_\theta$ and an estimator $\hat{\theta}_n$ of θ if we have*

1. $\mathbb{E}(\hat{\theta}_n) \xrightarrow{n\to\infty} \theta \Leftrightarrow B(\hat{\theta}_n) \xrightarrow{n\to\infty} 0$
2. $\mathrm{Var}(\hat{\theta}_n) \xrightarrow{n\to\infty} 0$

then it follows that $\hat{\theta}_n$ is a consistent estimator

Definition 47 (Asymptotic Efficiency) *Given a random sample $X_n \sim \mathcal{F}_\theta$ with parameter space Θ. An estimator $\hat{\theta}_n$ of θ is an efficient estimator if for all $\theta \in \Theta$:*

$$\sqrt{n} \cdot (\hat{\theta}_n - \theta) \xrightarrow{D} \mathcal{N}(0, \frac{1}{I(\theta)}$$

where $I(\theta)$ is the expected Fisher information

3 Statistical test

Definition 48 (Statistical Hypothesis) *Consider a random sample $X \sim \mathcal{F}_\theta$ with parameter space Θ. We consider a partition of Θ:*

$$\Theta = \Theta_0 \cup \Theta_1 \quad \text{with } \Theta_0 \cap \Theta_1 = \emptyset$$

A (statistical) hypothesis H is a statement about θ, i.e

- *Null hypothesis $H_0 : \theta \in \Theta_0$*
- *Alternative hypotheses $H_1 : \theta \in \Theta_1$*

Definition 49 (Statistical Hypothesis Test) *Consider a random sample of size n, in short $X \sim \mathcal{F}_\theta$ with sample space S_X and parameter space Θ with partition*

$$\Theta = \Theta_0 \cup \Theta_1 \quad \text{with: } \Theta_0 \cap \Theta_1 = \emptyset$$

Given the two hypothesis $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_1$, a statistical hypothesis test is a decision rule D that selects one of the two hypothesis based on realizations of X:

$$D : S_X \to \{H_0, H_1\}$$

We note that $D(X)$ is a statistic.

Definition 50 (Test statistic) *The test decision rule is based on a test statistic $W = W(X)$ with $W : S_X \to \mathbb{R}$, where $\mathbb{R} = R \cup R^c$ with R be the rejection region. Then, the decision rule is define as follows*

$$D(x) = \begin{cases} H_0 & W(x) \in R^c \\ H_1 & W(x) \in R \end{cases}$$

Given a realization $X = x$.

Remark: A good statistical test should fulfill:

1. $\mathbb{P}_{\theta\in\Theta_0}(W(X) \in R)$ is closed to 0
2. $\mathbb{P}_{\theta\in\Theta_1}(W(X) \in R)$ is closed to 1

Definition 51 (Power Function) *The power function of a statistical test is defined as*

$$\beta : \Theta \to [0, 1]$$

with

$$\beta(\theta) = \mathbb{P}_\theta(W(X) \in R)$$

where $\theta \in \Theta$ is the true parameter.

Remark:

1. For $\theta \in \Theta_0$ the power function should be low
2. For $\theta \in \Theta_1$ the power function should be high

3. A good test statistic has a high power $\beta(\theta)$ for $\theta \in \Theta_1$.

Definition 52 (Test level) *A statistical test is called a test to the level $\alpha \in [0, 1]$ if*

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$$

That is, if under H_0 the probability to commit a type 1 error is bounded by α.

A statistical test can have two outcomes:

- You reject H_0 and you claim that H_1 is right.
- You do not reject H_0, but you do not confirm H_0 either. You don't claim anything. (you do not have enough informatio to confirm H_0.

In principle, you could make two mistakes:

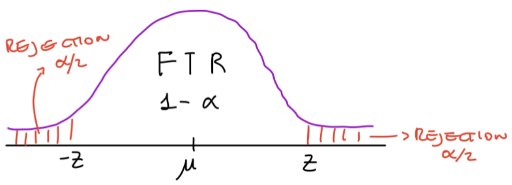
- H_0 is right, but you claim H_1 is right. ('type 1 error')
- H_1 is right, but you claim H_0 is right. ('type 2 error')

Tests are constructed such that the probability for making an 'error of type 1' is lower than or equal to α. A widely used (conventional) 'test level' is $\alpha = 0.05$.

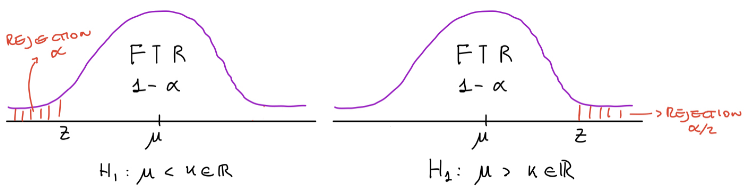
If the tests rejects the null hypothesis, statisticians say: 'The test was significant to the level α'

There exists two type of test, the two sided test problem and the one side test problem.

Definition 53 (Two sided test problem) *A two sided test problem is a problem where we have $H_0 : \mu = k \in \mathbb{R}$ and $H_1 : \mu \notin k \in \mathbb{R}$. Let $W(X) \sim \mathcal{F}_\mu$, and this be a test level to α. In the picture we have $W(X)$ distribution (we assumed for sake of simplicity that is a symmetric distribution) with z be the critical value.*



Definition 54 (One sided test problem) *A two sided test problem is a problem where we have $H_0 : \mu > k$ or $\mu < k$ and $H_1 : \mu < k$ or $\mu > k$, where $k \in \mathbb{R}$. Let $W(X) \sim \mathcal{F}_\mu$, and this be a test level to α. In the picture we have $W(X)$ distribution (we assumed for sake of simplicity that is a symmetric distribution) with z be the critical value.*



Definition 55 (Likelihood ratio (RT) test statistic) *Consider a random sample* $X \sim \mathcal{F}_\theta$ *with* $\theta \in \Theta$ *and a partition* $\Theta = \Theta_0 \cup \Theta_1$*, and the test problem*

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

The likelihood ratio test statistic is defined as:

$$\lambda(X) = \frac{\sup_{\theta \in \Theta_0}\{L_X(\theta)\}}{\sup_{\theta \in \Theta_0 \cup \Theta_1}\{L_X(\theta)\}}$$

where $L_X(\cdot)$ *is the likelihood.*

Remark:low values of $\lambda(X)$ suggest that θ is more likely to be in Θ_1.

Definition 56 (Likelihood ratio test) *A likelihood ratio test LRT makes use of the likelihood ratio test statistic. The LRT is based on the decision rule:*

$$D_\lambda(X) = \begin{cases} H_0 & \lambda(X) > c \\ H_1 & \lambda(X) \leq c \end{cases}$$

where $c \in [0, 1]$*. The test level* α *depends on the value of* c*.*

Definition 57 (Uniform most powerful test (UMP)) *a test* $D(X)$ *is the uniform most powerful test if all other test* $\tilde{D}(X)$ *to the same level* α *have less power on* Θ_1*. That is, if we have*

$$\mathbb{P}_\theta(D(X) = H_1) \geq \mathbb{P}_\theta(\tilde{D}(X) = H_1)$$

for all $\theta \in \Theta_1$ *and any level* α *test* $\tilde{D}$

Lemma 58 (Neyman Person Lemma) *Consider a random sample* $X \sim \mathcal{F}_\theta$ *and a simple test problem*

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

A test that employs as test statistic the density ratio

$$W(X) = \frac{f_{\theta_0}(X)}{f_{\theta_1}(X)}$$

and uses the rejection region $R = \{x \in S_X : W(X) < k\}$*, so that the decision rule is*

$$D(X) = \begin{cases} H_1 & W(X) < k \\ H_0 & W(X) \geq k \end{cases}$$

is the UMP test of level $\alpha = \mathbb{P}_{\theta_0}(W(X) < k)$

Lemma 59 *Consider a random sample* $X \sim \mathcal{F}_\theta$ *with sufficient statistic* $T(X)$ *and a simple test problem*

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

A test that employs as test statistic the sufficient statistic density ratio

$$W(X) = \frac{f_{T,\theta_0}(T(X))}{f_{T,\theta_1}(T(X))}$$

and uses the rejection region $R = \{t \in S_T : W(t) < k\}$*, so that the decision rule is*

$$D(T(X)) = \begin{cases} H_1 & W(T(X)) < k \\ H_0 & W(T(X)) \geq k \end{cases}$$

is the UMP test of level $\alpha = \mathbb{P}_{\theta_0}(W(T(X)) < k)$

Definition 60 (Monotone Likelihood Ratio) *Consider a random sample* $X \sim \mathcal{F}_\theta$ *with sufficient statistic* $T(X)$*.* $T(X)$ *has a monotone likelihood ratio if*

$$W(t) = \frac{f_{T,\theta_0}(t)}{f_{T,\theta_1}(t)}$$

is a monotone function of $t \in S_T$*. For every* $k > 0$ $(W(X) < k)$ *there is a* $t_0 \in \mathbb{R}$ *with*

1. $t > t_0$ *if monotonically decreasing*
2. $t < t_0$ *if monotonically increasing*

Theorem 61 (Karlin-Ruben Theorem) *Consider a random sample* $X \sim \mathcal{F}_\theta$ *with sufficient statistic* $T(X)$ *having a monotone likelihood ratio, and the composite test problem*

$$H_0 : \theta \leq \theta_0 \quad H_1 : \theta > \theta_0$$

1. *If* $T(X)$ *has a monotonically decreasing likelihood ration, then the test that reject* H_0 *if* $T > t_0$ *is UMP of the level* $\alpha = \mathbb{P}_{\theta_0}(T(X) > t_0)$
2. *If* $T(X)$ *has a monotonically increasing likelihood ration, then the test that reject* H_0 *if* $T < t_0$ *is UMP of the level* $\alpha = \mathbb{P}_{\theta_0}(T(X) < t_0)$

Definition 62 (Asymptotic LR test) *Consider a random sample* $X \sim \mathcal{F}_\theta$ *with parameter space* Θ *and the test problem*

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

where $\Theta = \Theta_0 \cup \Theta_1$ *is a partition, and the likelihood ratio statistic*

$$\lambda_n(X) = \frac{\sup_{\theta \in \Theta_0}\{L_X(\theta)\}}{\sup_{\theta \in \Theta_0 \cup \Theta_1}\{L_X(\theta)\}}$$

under the following regulatory condition

1. $\Theta \subset \mathbb{R}$ *must be an open set*
2. *The sample space* S_X *is not allowed to depend on* θ
3. *The density* $f_\theta(x)$ *must be 3-times differentiable w.r.t* θ

we have under H_0

$$-2\log(\lambda_n(X)) \xrightarrow{D} \chi^2_1$$

Definition 63 (P-value) *The p-value is the lowest test level* α *to which* H_0 *could have been rejected.*

Definition 64 (One sample t-test (two sided)) *Consider a sample from a Gaussian distribution $X_n \sim \mathcal{N}(\mu, \sigma^2)$ with two unknown parameters μ and σ^2, and the test problem*

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

Under the null-hypothesis, we have

$$T(X) = \frac{\sqrt{n} \cdot (\overline{X}_n - \mu_0)}{\sqrt{S_n^2}} \sim t_{n-1}$$

A two-sided one sample t-test to the level α employs the decision rule:

$$D(X) = \begin{cases} H_0 & T(X) \in [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}] \\ H_1 & \text{otherwise} \end{cases}$$

where $q_{\frac{\alpha}{2}}$ and $q_{1-\frac{\alpha}{2}}$ are the quantiles of the t_{n-1} distribution

Definition 65 (One sample t-test (one sided) version 1) *Consider a sample from a Gaussian distribution $X_n \sim \mathcal{N}(\mu, \sigma^2)$ with two unknown parameters μ and σ^2, and the test problem*

$$H_0 : \mu \leq \mu_0 \quad H_1 : \mu \not> \mu_0$$

Under the null-hypothesis, we have

$$T(X) = \frac{\sqrt{n} \cdot (\overline{X}_n - \mu_0)}{\sqrt{S_n^2}} \sim t_{n-1}$$

A one-sided one sample t-test to the level α employs the decision rule:

$$D(X) = \begin{cases} H_0 & T(X) \leq q_{1-\alpha} \\ H_1 & T(X) > q_{1-\alpha} \end{cases}$$

where $q_{1-\alpha}$ is the quantiles of the t_{n-1} distribution. Note that here the likelihood ratio is monotonically increasing.

Definition 66 (One sample t-test (one sided) version 2) *Consider a sample from a Gaussian distribution $X_n \sim \mathcal{N}(\mu, \sigma^2)$ with two unknown parameters μ and σ^2, and the test problem*

$$H_0 : \mu \geq \mu_0 \quad H_1 : \mu \not< \mu_0$$

Under the null-hypothesis, we have

$$T(X) = \frac{\sqrt{n} \cdot (\overline{X}_n - \mu_0)}{\sqrt{S_n^2}} \sim t_{n-1}$$

A one-sided one sample t-test to the level α employs the decision rule:

$$D(X) = \begin{cases} H_0 & T(X) \geq q_{\alpha} \\ H_1 & T(X) < q_{\alpha} \end{cases}$$

where q_{α} is the quantiles of the t_{n-1} distribution. Note that here the likelihood ratio is monotonically increasing

Definition 67 (Two sample t-test (unpaired, two-sided)) *Consider two idependent Guassian samples $X_n \sim \mathcal{N}(\mu_x, \sigma^2)$ and $Y_m \sim \mathcal{N}(\mu_y, \sigma^2)$, where μ_x, μ_y, σ^2 are unknown parameters, and the test problem*

$$H_0 : \mu_x - \mu_y = \mu^\star \quad H_1 : \mu_x - \mu_y \neq \mu^\star$$

Under H_0 we have

$$T(X,Y) = \frac{\overline{X}_n - \overline{Y}_m - \mu^\star}{\sqrt{S^2_{n,m}\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t_{n+m-2}$$

where

$$S^2_{n,m} = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2 + \sum_{i=1}^m (Y_i - \overline{Y}_m)^2}{n+m-2}$$

An unpaired two sample t-test to the level α employs the decision rule

$$D(X) = \begin{cases} H_0 & T(X,Y) \in [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}] \\ H_1 & \textit{otherwise} \end{cases}$$

where $q_{\frac{\alpha}{2}}$ and $q_{1-\frac{\alpha}{2}}$ are the quantiles of the t_{n+m-2} distribution

Remark about statistical test structure:

- There is a test problem H_0 vs H_1
- There is a statistical test that can be computed from the observed data
- Under H_0 the test statistic has a well-known distribution
- the user specifies the test level $\alpha \in [0, 1]$
- A rejection region is specified such that under the null-hypothesis the probability that the test statistic takes values in the rejection region is bounded by α (erro of type 1, i.e rejecting H_0, thought is true).
- if the test statistic takes value in the rejection region, the alternative hypothesis is confirmed

3.1 Confidence interval

Definition 68 (Confidence interval (CI)) *Consider a random sample $X \sim \mathcal{F}_\theta$ with $\theta \in \Theta$. An interval $[L(X), U(X)]$ that contains the unknown parameter θ with probability $1 - \alpha$ is called a $1 - \alpha$ confidence interval for θ we have*

$$\forall \theta \in \Theta : \mathbb{P}_\theta(L(X) \leq \theta \leq U(X)) \geq 1 - \alpha \quad \Leftrightarrow \inf_{\theta \in \Theta} \{p_\theta(L(X) \leq \theta \leq U(X))\} \geq 1 - \alpha$$

where $U(X), L(X)$ are statistic.

Definition 69 (Connection between tests and CI) *Consider a random sample $X \sim \mathcal{F}_\theta$ with $\theta \in \Theta$. For every $\theta_0 \in \Theta$ we can formulate the test problem*

$$H_0 : \theta = \theta_0 \quad H_1 : \theta \neq \theta_0$$

Assume we can for each $\theta_0 \in \Theta$ perform a statistical level α test with test statistic $W(X)$ and rejection region R_{θ_0}. Then a $1 - \alpha$ confidence interval for θ_0 is given by

$$CI(X) = \{\theta : W(X) \notin R_{\theta_0}\}$$

Remark: Note that the true parameter θ_0 is in $CI(X)$ with probability $1-\alpha$.

Definition 70 (Wald test and confidence intervals) *Recall the definition of asymtotically efficient for the maximul likelihood estimator of a random sample $X \in \mathcal{F}_\theta$. Then, the asymptotic $1-\alpha$ confidence interval for θ is*

$$\hat{\theta}_{ML,n} \pm q_{1-\frac{\alpha}{2}} \frac{1}{\sqrt{n \cdot I(\hat{\theta}_{ML,n}}}$$