# Introduction to tests

Olga Klopp

## Introductory example

- We observe 10 coin flips and we obtain the following result (heads and tails):

$$(H, H, T, T, H, T, H, H, T, H)$$

  Question : Is the coin balanced ?

- Answer to this question implies to **make a decision**:

$$\varphi = \varphi(H, H, T, T, H, T, H, H, T, H)$$

$$= \begin{cases} \textbf{we accept} \text{ the hypothesis "the piece is balanced"} \\ \textbf{we reject} \text{ the hypothesis "the coin is balanced"} \end{cases}$$

- Modeling : These observations are modeled by the following statistical experience

$$\mathcal{E}^{10} = \big(\{0,1\}^{10}, \{\mathbb{P}_\theta, \theta \in [0,1]\}\big),$$

  where $(H = 1, T = 0)$

$$\mathbb{P}_\theta = \big(\theta\delta_1 + (1-\theta)\delta_0\big)$$

- and we "translate" the question into mathematical terms: solve the next test problem

$$\boxed{H_0 : \theta = \frac{1}{2} \quad \text{against} \quad H_1 : \theta \neq \frac{1}{2}}$$

**Definition 1.** *The hypothesis $H_0$ is called null hypothesis.*

*The hypothesis $H_1$ is called alternative hypothesis.*

## Definitions

**Definition 2.** *1. Let $Z$ be our observations.*

*2. Let $\Theta = \Theta_0 \cup \Theta_1$ be a partition of the parameter space.*

*3. We consider the following test problem*

$$\boxed{H_0 : \theta \in \Theta_0 \quad \textit{against} \quad H_1 : \theta \in \Theta_1}$$

*A test or a decision rule is a statistic of the form*

$$\varphi(Z) = I(Z \in \mathcal{R}) = \begin{cases} H_0 & \textit{"we accept"} \\ H_1 & \textit{"we reject"} \end{cases}$$

*$\mathcal{R}$ is called* **region of rejection** *or* **critical region**.

## Building a test

- $Z = (X_1, \dots, X_{10})$ : observation in the Bernoulli sampling model $\{\mathbb{P}_\theta : 0 < \theta < 1\}$ where $\mathbb{P}_\theta = \theta\delta_1 + (1-\theta)\delta_0$
- we look at the following test problem:

$$\boxed{H_0 : \theta = \frac{1}{2} \quad \text{against} \quad H_1 : \theta \neq \frac{1}{2}}$$

- we build a **test** using MLE $\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}} = \bar{X}_n = \bar{X}_{10}$ ($n = 10$) and a given threshold $t_0$:

$$\varphi(Z) = I(Z \in \mathcal{R}) = \begin{cases} H_0 & \text{when } \left|\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}} - \frac{1}{2}\right| \leq t_0 \\ H_1 & \text{if not} \end{cases}$$

- the **region of rejection** :

$$\mathcal{R} = \left\{z \in \{0,1\}^{10} : \left|\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}}(z) - \tfrac{1}{2}\right| > t_0\right\}$$

- In the previous test

$$\varphi(Z) = I(Z \in \mathcal{R}) = \begin{cases} H_0 & \text{when } \left|\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}} - \frac{1}{2}\right| \leq t_0 \\ H_1 & \text{if not} \end{cases}$$

  this test was built using the MLE estimator $\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}}$. This is a classic approach: here the MLE plays the role of the **test statistic**

- We can compute MLE using our data $\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}} \overset{example}{=} 0,6$ and so make a decision.

  Question : how to choose the threshold $t_0$?

- is there a better test choice? better test statistics?

## Two types of decision errors

**Definition 3.** *Let $\varphi$ be a test with the region of rejection $\mathcal{R}$ (that is $\varphi(z) = I(z \in \mathcal{R})$)*

- *Type 1 error (wrongly reject )*

$$\sup_{\theta \in \boldsymbol{\Theta_0}} \mathbb{P}_\theta[Z \in \mathcal{R}]$$

- *Type 2 error (wrongly accept)*

$$\theta \in \boldsymbol{\Theta_1} \mapsto \mathbb{P}_\theta[Z \notin \mathcal{R}] = 1 - \pi_\varphi(\theta)$$

  *where $\pi_\varphi(\theta) = \mathbb{P}_\theta[Z \in \mathcal{R}]$ is the* **function of power** *of the test*

**Note** : "*reject*" = reject $H_0$; "*accept*" = accept $H_0$
In our example:
The test $\varphi(z) = I(|\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}}(z) - 0.5| > t_0)$ can make two types of errors:

- wrongly reject :

$$\text{Reject} \quad (\varphi(Z) = H_1) \quad \text{while} \quad \theta = \tfrac{1}{2} \in \Theta_0 = \{\tfrac{1}{2}\}$$

  in this case, the **type 1 error** is

$$\mathbb{P}_{0.5}[|\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}}(Z) - 0.5| > t_0]$$

- wrongly accept :

$$\text{Accept} \quad (\varphi(Z) = H_0) \quad \text{while} \quad \theta \neq \tfrac{1}{2}$$

  in this case, the function of the **type 2 error** is

$$\theta \neq 1/2 \mapsto \mathbb{P}_\theta[|\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}}(Z) - 0.5| \leq t_0]$$

## Asymptotic level of a test

**Definition 4.** *In the sampling model we build a test sequence* $(\varphi_n)$ *(where* $n$ *is the number of observations). Let* $\alpha \in (0,1)$*. We say that* $(\varphi_n)$ *is of the* **asymptotic level** $\alpha$ *when*

$$\forall \theta \in \mathbf{\Theta_0}, \quad \limsup_{n\to\infty} \mathbb{P}_\theta[\varphi_n = H_1] \leq \alpha$$

1. we want to be sure that asymptotically the probability of rejecting wrongly is less than $\alpha$
2. ideally, we would like to be able to set a level for both types of errors (1st and 2nd) but this is not possible in general. We will therefore favor a control of the type 1 error by constructing tests of given asymptotic level: **introduction of an asymmetry**

In our example:

Under the null hypothesis $H_0$ :

$$\boxed{\sqrt{n}\left(\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}} - 0.5\right) \xrightarrow{d} \mathcal{N}(0, 1/4)}$$

in particular, for $g \sim \mathcal{N}(0,1)$,

$$\mathbb{P}_{0.5}[\left|\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}} - 0.5\right| > t_{n,\alpha}] \longrightarrow \mathbb{P}[|g| > q_{1-\alpha/2}] = \alpha$$

where, using the $1-\alpha/2$-quantile $q_{1-\alpha/2}$ of $g$ we put

$$\boxed{t_{n,\alpha} = \frac{q_{1-\alpha/2}}{2\sqrt{n}}}$$

Under $H_1$: for any $\theta \in \Theta_0^c = (0,1) - \{0.5\}$,

$$\boxed{\sqrt{n}\left|\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}} - 0.5\right| \overset{a.s.}{\to} +\infty}$$

In our example, the observations are:

$$z = (H, H, T, T, H, T, H, H, T, H)$$

then, the MLE is equal $\widehat{\theta}_{10}^{\mathtt{ml}}(z) = 0.6$ and, for a given $\alpha \in (0,1)$ the test is

$$\varphi(Z) = \begin{cases} H_0 & \text{when } \left|\widehat{\theta}_{10}^{\mathtt{ml}}(Z) - \frac{1}{2}\right| \leq t_{10,\alpha} = \frac{q_{1-\alpha/2}}{2\sqrt{10}} \\ H_1 & \text{if not} \end{cases}$$

For example:

1. for $\alpha = \mathbf{5}\%$, we have $\mathbf{q_{1-\alpha/2} \approx 1.96}$ and $t_{10,\alpha} \approx 0.31$ then as $\left|\widehat{\theta}_{10}^{\mathtt{ml}}(z) - \frac{1}{2}\right| = 0.1 \leq t_{10,5\%}$, we **accept**
2. for $\alpha = \mathbf{10}\%$, we have $\mathbf{q_{1-\alpha/2} \approx 1.64}$, alors $t_{10,\alpha} \approx 0.26$ then as $\left|\widehat{\theta}_{10}^{\mathtt{ml}}(z) - \frac{1}{2}\right| = 0.1 \leq t_{10,10\%}$, we **accept**

## Introducing the $p$ -value

The choice of the **significance level** $\alpha$ is arbitrary. In the previous example, we will

- accepter as long as $q_{1-\alpha/2} \geq 2\sqrt{10} * \left|\widehat{\theta}_{10}^{\mathtt{ml}} - \frac{1}{2}\right|$
- reject once $q_{1-\alpha/2} \leq 2\sqrt{10} * \left|\widehat{\theta}_{10}^{\mathtt{ml}} - \frac{1}{2}\right|$

The **limit value of $\alpha$ for which the decision toggles** that is the $\alpha$ such that

$$q_{1-\alpha/2} = 2\sqrt{10} * \left| \widehat{\theta}_{10}^{\mathtt{ml}} - \tfrac{1}{2} \right|$$

is called the **p-value**. Here the p-value is given by the equation (with respect of $\alpha$)

$$q_{1-\alpha/2} = 2\sqrt{10} * 0.1 \approx 0.63$$

that is $\alpha =$ p-value $\approx 0.525$.

**Definition 5.** *Let $\varphi_\alpha$ be a test of the asymptotic level $\alpha$ and the reject zone $\mathcal{R}_\alpha$. The following statistic is called the* **p-value** *of the test:*

$$p - value = \inf(\alpha \in (0,1) : Z \in \mathcal{R}_\alpha)$$

- It is the critical threshold where the decision switches:

$$\varphi_\alpha(Z) = \begin{cases} H_0 & \text{when } \alpha \leq \mathbf{p-value} \\ H_1 & \text{when } \alpha > \mathbf{p-value} \end{cases}$$

- the **p-value** quantifies how confident we are accepting $H_0$.

| **p-value** | **level of confidence accepting $H_0$** | decision |
|---|---|---|
| $p < 0.01$ | very weak | *reject* (with confidence) |
| $0.01 \leq p < 0.05$ | weak | *reject* |
| $0.05 \leq p < 0.1$ | strong | *accept* |
| $0.1 \leq p$ | very strong | *accept* (with confidence) |

- large p-value : the test does not allow to reject $H_0$
- small p-value: even if we take a very small confidence level, the test will reject $H_0$ (while we have a strong aversion for the type 1 risk, i.e. to reject wrongly)
- in our example, we have the p-value $\approx 0.525$, which implies that we will accept.

## Meaning of acceptance

**To accept $H_0$ does not mean that $H_0$ is true**

1. by default, we accept $H_0$ unless we bring a "proof" that $H_0$ is not acceptable
2. *Accept* only means that we have not been able to prove that $H_0$ is not acceptable: we prefer to say that the test does not allow to reject rather than "we accept".
3. A *proof* is the observation of a "rare" event under $H_0$: "under $H_0$, the test statistic takes a value that can be considered rare" is a proof of rejection
4. the "rarity" of an event is fixed by the (asymptotic) level $\alpha$
5. if in our example, the true $\theta = 0.5 + 10^{-10}$, it is very likely that we will not reject while $H_0$ is false

## Meaning of rejection

**Only the rejection is informative**

1. *reject* means that it has been proven that $H_0$ can not be accepted
2. given a confidence level $\alpha$, we reject when the value taken by the test statistic is rare being known its law under $H_0$ (what is a rare event depends on the confidence level $\alpha$)

3. in general, **under $H_0$**, we know the asymptotic law of the test statistic (ex. : $\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}} \sim \mathcal{N}(0.5, (4n)^{-1})$); if the value taken by this statistic is unlikely for its asymptotic law (ex. : $\widehat{\theta}_{\mathrm{n}}^{\mathtt{ml}} = 0.9$) then we will reject = we have a proof that $H_0$ is not acceptable

4. the $p$-value measures how rare is the observed value of the test statistic for its law under $H_0$.

**The choice of hypothesis is important**

1. For a partition $\Theta = A \cup B$, there is no equivalence between the two test problems

$$H_0 : \theta \in A \text{ against } H_1 : \theta \in B$$

and

$$H_0 : \theta \in B \text{ against } H_1 : \theta \in A$$

2. We choose assumptions based on interest in the problem : the hypothesis $H_0$ is privileged

3. the hypothesis $H_0$ is privileged because we decided to cover against the risk of the fist type before the risk of 2nd type: i.e, we want to avoid, first and foremost, rejecting wrongly and consequently, we tend to "accept too much"

4. it is easier to accept than to reject because rejection requires a "proof" that acceptance is not sustainable

# Methodology for asymptotic tests

a) find a **test statistic** (often an estimator) $\widehat{\theta}_n$

b) such that under $H_0$, we have an asymptotic normality

$$\sqrt{\frac{n}{v(\widehat{\theta}_n)}}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{d} V$$

(or $v(\theta_0)$ instead of $v(\widehat{\theta}_n)$)

c) and such that under $H_1$ :

$$\sqrt{\frac{n}{v(\widehat{\theta}_n)}}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{a.s.} +\infty$$

(with or without absolute values)

d) we use this statistic to build a test of asymptotic level $\alpha \in (0, 1)$ setting

$$\varphi(Z) = \begin{cases} H_0 & \text{when } \left(\widehat{\theta}_n - \theta_0\right) \leq \frac{q^V_{1-\alpha}\sqrt{v(\widehat{\theta}_n)}}{\sqrt{n}} := t_{n,\alpha} \\ H_1 & \text{if not} \end{cases}$$

The form of the test (here $\widehat{\theta}_n - \theta_0$ smaller than something) is given by the behavior of $\widehat{\theta}_n - \theta_0$ under $H_1$.

e) under $H_0$ we have

$$\mathbb{P}[reject] = \mathbb{P}\left[\widehat{\theta}_n - \theta_0 > t_{n,\alpha}\right] \longrightarrow \mathbb{P}[V > q^V_{1-\alpha}] = \alpha$$

It is therefore a test of the asymptotic level $\alpha$.

# Choice of tests: notion of optimality for tests

## Simple hypothesis versus simple alternative

This is the case when $\Theta = \{\theta_0, \theta_1\}$ with $\theta_0 \neq \theta_1$ and

$$\boxed{\Theta_0 = \{\theta_0\} \text{ against } \Theta_1 = \{\theta_1\}}$$

Question: Does it exist an optimal test $\varphi^\star$, in the sense that: $\forall$ test $\varphi$, we have simultaneously better control over both errors (1st and 2nd type)

$$\begin{cases} \mathbb{P}_{\theta_0}\left[\varphi^\star = H_1\right] \leq \mathbb{P}_{\theta_0}\left[\varphi = H_1\right] & \text{"probability of wrongly reject"} \\ \mathbb{P}_{\theta_1}\left[\varphi^\star = H_0\right] \leq \mathbb{P}_{\theta_1}\left[\varphi = H_0\right] & \text{"probability of wrongly accept"} \end{cases}$$

## Neyman - Pearson approach

- Assumptions $H_0$ and $H_1$ are "dis-symmetric" : $H_0$ is "more important" than $H_1$ in the following sense: we impose a type 1 error:

$$\underline{\text{we want to avoid, first and foremost, a false rejection}}$$

**Definition 6.** *For a given $\alpha \in (0,1)$, a hypothesis test $\varphi = \varphi_\alpha$ with the null hypothesis $H_0 : \theta \in \Theta_0$ against an alternative $H_1$ is of significance* **level** $\alpha$ *if*

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta\left[\varphi_\alpha = H_1\right] \leq \alpha$$

- The significance level *alpha* for a test does not say anything about the second type error (or power)

**Definition 7.** *Let $\varphi$ be a test with the rejection region $\mathcal{R}$. The* **power** *of $\varphi$ is the function $\pi_\varphi : \theta \in \Theta_1 \mapsto \mathbb{P}_\theta[Z \in \mathcal{R}]$ (the probability that it will reject a false null hypothesis)*

- The Neyman - Pearson approach : $\alpha \in (0,1)$, among all the tests of a given significance level $\alpha$, look for the one(s) that are the most powerful

**Definition 8.** *A Uniformly More Powerful (UMP) test is a hypothesis test which has the greatest power among all possible tests of a given size $\alpha$:*

1. $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta\left[\varphi^\star = H_1\right] \leq \alpha$
2. *and if $\varphi$ is such that $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta\left[\varphi = H_1\right] \leq \alpha$ then*

$$\forall \theta \in \Theta_1, \pi_{\varphi^\star}(\theta) \geq \pi_\varphi(\theta)$$

## Neyman-Pearson lemma

### Neyman-Pearson test (likelihood ratio test)

For the case of a simple null hypothesis (that is $\Theta_0 = \{\theta_0\}$) against a simple alternative hypothesis (that is $\Theta_1 = \{\theta_1\}$), an UMP test exists: it is the Neyman-Pearson test (or likelihood ratio test) whose construction is as follows:

- density $f(\theta, z)$, $\theta \in \{\theta_0, \theta_1\}$
- We choose a critical region of the form

$$\boxed{\mathcal{R}(c) = \left\{f(\theta_1, z) > c f(\theta_0, z)\right\}, \quad c > 0}$$

  and we tune $c = t_{n,\alpha}$ so that

$$\boxed{\mathbb{P}_{\theta_0}\left[Z \in \mathcal{R}(t_{n,\alpha})\right] = \alpha}$$

- This test (if this equation admits a solution) is of the level $\alpha$. We can shows that it's UMP.

**Proposition 1** (Neyman-Pearson Lemma)**.** *Let $\alpha \in [0,1]$. If the equation*

$$\boxed{\mathbb{P}_{\theta_0}\left[f(\theta_1, Z) > t_{n,\alpha} f(\theta_0, Z)\right] = \alpha}$$

*admits a solution $t_{n,\alpha}$ then the test with the critical region*

$$\mathcal{R}_\alpha = \left\{z : f(\theta_1, z) > t_{n,\alpha} f(\theta_0, z)\right\}$$

*is of the level $\alpha$ and UMP for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$.*

- In general, if $U = f(\theta_1, Z)/f(\theta_0, Z)$ is well defined, then $\mathbb{P}_{\theta_0}\left[U > t_{n,\alpha}\right] = \alpha$ admits a solution.

### Neyman-Pearson test example

We observe $Z = (X_1, \ldots, X_n)$ where $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$. For $\theta_0 < \theta_1$, we consider the test problem

$$\boxed{H_0 : \theta = \theta_0 \text{ against } H_1 : \theta = \theta_1}$$

The likelihood is

$$f(\theta, Z) = \tfrac{1}{(2\pi)^{n/2}} \exp\Big(-\frac{1}{2}\sum_{i=1}^{n} X_i^2 + n\theta\overline{X}_n - \frac{n\theta^2}{2}\Big)$$

The likelihood ratio is

$$\frac{f(\theta_1, Z)}{f(\theta_0, Z)} = \exp\Big(n(\theta_1 - \theta_0)\overline{X}_n - \frac{n}{2}(\theta_1^2 - \theta_0^2)\Big)$$

- rejection region of N-P. test:

$$\begin{aligned}\mathcal{R}(c) &= \big\{z \in \mathbb{R}^n : f(\theta_1, z) > c f(\theta_0, z)\big\}\\ &= \big\{(x_1, \ldots, x_n)^\top \in \mathbb{R}^n : n(\theta_1 - \theta_0)\overline{x}_n - \frac{n}{2}(\theta_1^2 - \theta_0^2) > \log c\big\}\\ &= \big\{(x_1, \ldots, x_n)^\top \in \mathbb{R}^n : \overline{x}_n > \frac{\theta_0 + \theta_1}{2} + \tfrac{\log c}{n(\theta_1 - \theta_0)}\big\}\end{aligned}$$

- The choice of $c$: we solve

$$\mathbb{P}_{\theta_0}\left[\overline{X}_n > \tfrac{1}{2}(\theta_0 + \theta_1) + \tfrac{\log c}{n(\theta_1 - \theta_0)}\right] = \alpha$$

  **Under $\mathbb{P}_{\theta_0}$** :

$$\overline{X}_n \sim \mathcal{N}\Big(\theta_0, \tfrac{1}{n}\Big)$$

Solving to find $c$ : for $g \sim \mathcal{N}(0,1)$,

$$\mathbb{P}\left[\theta_0 + \frac{1}{\sqrt{n}} g > \frac{1}{2}(\theta_0 + \theta_1) + \frac{\log c}{n(\theta_1 - \theta_0)}\right] = \alpha$$

that is $\mathbb{P}\left[g > \frac{\sqrt{n}}{2}(\theta_1 - \theta_0) + \frac{1}{\sqrt{n}}\frac{\log c}{\theta_1 - \theta_0}\right] = \alpha$, which implies

$$\frac{\sqrt{n}}{2}(\theta_1 - \theta_0) + \frac{1}{\sqrt{n}}\frac{\log c}{\theta_1 - \theta_0} = q_{1-\alpha},$$

where $q_{1-\alpha}$ is the $1 - \alpha$ order quantile of $\mathcal{N}(0,1)$

- Conclusion: the NP test of the level $\alpha$ has a rejection region $\mathcal{R}(c_\alpha)$ where

$$c_\alpha = \exp\left(\sqrt{n}(\theta_1 - \theta_0)q_{1-\alpha} - \frac{n(\theta_1-\theta_0)^2}{2}\right)$$

which can be written as

$$\mathcal{R}(c_\alpha) = \{(x_1,\dots,x_n)^\top \in \mathbb{R}^n : \overline{x}_n > \theta_0 + t_{n,\alpha}\} \text{ where } t_{n,\alpha} = \frac{q_{1-\alpha}}{\sqrt{n}}.$$

We see that the NP test has the following form:

$$\varphi(Z) = \begin{cases} H_0 & \text{when } \overline{X}_n \leq \theta_0 + t_{n,\alpha} \\ H_1 & \text{if not} \end{cases} \quad \text{where } t_{n,\alpha} = \frac{q_{1-\alpha}}{\sqrt{n}}$$

rem.: the value of $\theta_1$ does not intervene in the NP test.

- the **power** of this test:

$$\pi_\varphi(\theta_1) = \mathbb{P}_{\theta_1}[\overline{X}_n > \theta_0 + t_{n,\alpha}] = \mathbb{P}[g > \sqrt{n}(\theta_0-\theta_1) + q_{1-\alpha}]$$

as under $\mathbb{P}_{\theta_1}$, $\overline{X}_n \sim \mathcal{N}(\theta_1, 1/n)$.

rem. : Power increases when $n$ increases and when $|\theta_0 - \theta_1|$ increases. The alternative intervenes only in power.

# Classical tests in the Gaussian sampling model

## Gaussian tests

### Testing for mean: known variance

We observe $Z = (X_1,\dots,X_n) \sim \mathcal{N}(\mu, \sigma^2 \mathrm{Id}_n)$ where $\sigma$ is known. We consider the following test problem:

$$\boxed{H_0 : \mu \leq \mu_0 \quad \text{against} \quad H_1 : \mu > \mu_0}$$

Idea: we estimate $\mu$ and we reject $H_0$ if our estimator is "larger" than $\mu_0$. We consider tests of the following form:

$$\varphi_\alpha(Z) = \begin{cases} H_0 & \text{if } \overline{X}_n < \mu_0 + t_{n,\alpha} \\ H_1 & \text{if not} \end{cases}$$

We choose the **threshold** $t_{n,\alpha}$ such that

$$\sup_{\mu \leq \mu_0} \mathbb{P}_\mu\left[\varphi_\alpha(Z) = H_1\right] = \alpha$$

Upper bounding the type 1 error. Let $\mu \leq \mu_0$. Under $\mathbb{P}_\mu$, $\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$, then for $g \sim \mathcal{N}(0,1)$

$$\begin{aligned}\mathbb{P}_\mu\left[\overline{X}_n - \mu_0 \geq t_{n,\alpha}\right] &= \mathbb{P}\left[(\mu + \tfrac{\sigma}{\sqrt{n}}g) - \mu_0 \geq t_{n,\alpha}\right] \\ &= \mathbb{P}\left[\tfrac{\sigma}{\sqrt{n}}g \geq t_{n,\alpha} + (\mu_0 - \mu)\right] \\ &\leq \mathbb{P}\left[\tfrac{\sigma}{\sqrt{n}}g \geq t_{n,\alpha}\right] \overset{\text{we whant}}{=} \alpha\end{aligned}$$

We take

$$\boxed{t_{n,\alpha} = \frac{\sigma q_{1-\alpha}}{\sqrt{n}}}$$

In particular, we have:

$$\sup_{\mu \leq \mu_0} \mathbb{P}_\mu\left[\varphi_\alpha(Z) = H_1\right] = \mathbb{P}_{\mu_0}\left[\varphi_\alpha(Z) = H_1\right]$$

**Calculation of the power of the test:** Let $\mu > \mu_0$. Under $\mathbb{P}_\mu$, the distribution of $\overline{X}_n$ is $\mathcal{N}(\mu, \sigma^2/n)$ the the power function of the test is

$$\begin{aligned} \mu \in (\mu_0, +\infty) \mapsto \mathbb{P}_\mu \left[\overline{X}_n - \mu_0 \geq t_{\alpha,n}\right] \\ = \mathbb{P}\left[g \geq \frac{\sqrt{n}(\mu_0 - \mu)}{\sigma} + q_{1-\alpha}\right] \end{aligned}$$

Rem. :

- the power tends to 1 when $n$ tends to $+\infty$,
- it's a UMP test.

**Testing for mean: unknown variance**

- **Main ingredient**:

$$s_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{n}{n-1}\left(\widehat{\sigma}_n^2\right)^{\mathtt{mv}}$$

  then

$$(n-1)\frac{s_n^2}{\sigma^2} \sim \chi^2(n-1)$$

  and

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{s_n} \sim \text{Student}(n-1)$$

  and these variables are **pivotal**: their distribution does not depend on $\mu, \sigma^2$ under $\mathbb{P}_{\mu,\sigma^2}$.

- $\chi^2$ and Student (with $k$ degree of freedom) distributions are classic and study independently.

**Testing for mean: composite hypothesis**

- We test $H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$. A test of the level $\alpha$ is given by the following rejection region:

$$\mathcal{R}_\alpha = \left\{z \in \mathbb{R}^n : T(z) > q^{\mathfrak{T}}_{1-\alpha,n-1}\right\}$$

  where

$$T(Z) = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{s_n}$$

  and $q^{\mathfrak{T}}_{1-\alpha,n-1}$ = quantile of order $1-\alpha$ of Student distribution with $n-1$ degrees of freedom :

$$\mathbb{P}\left[\text{Student}_{n-1} > q^{\mathfrak{T}}_{1-\alpha,n-1}\right] = \alpha$$

- We test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. A test of the level $\alpha$ is given by $\mathcal{R}_\alpha = \left\{z \in \mathbb{R}^n : \left|T(z)\right| > q^{\mathfrak{T}}_{1-\alpha/2,n-1}\right\}$.

**Testing for variance**

- We test $H_0 : \sigma^2 \leq \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$. A test of the level $\alpha$ : is given by the rejection region

$$\mathcal{R}_\alpha = \left\{z \in \mathbb{R}^n : V(z) > q^{\chi^2}_{1-\alpha,n-1}\right\},$$

  where

$$V(Z) = \frac{1}{\sigma_0^2}\sum_{i=1}^n (X_i - \overline{X})^2$$

  and

$$\mathbb{P}\left[\text{Chi-deux}_{n-1} > q^{\chi^2}_{1-\alpha,n-1}\right] = \alpha.$$

Exercises: testing mean and variance.