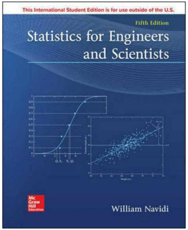
Introduction aux statistiques pour l'Ingénieur
Test Statistique

Stéphane Canu

asi.insa-rouen.fr/enseignants/~scanu

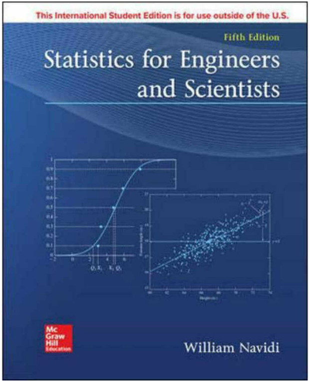
scanu@insa-rouen.fr



ITI 3, INSA Rouen Normandie, mai 2022

Lecture road map

1. La question du test
2. Formalisation des test d'hypothèse
3. Le cas de l'espérance d'une loi norma



https://moodle.insa-rouen.fr/course/view.php?id=93

Un exemple simple

Vous être l'ARS : comment décider ?

1. Un médicament soulage 1 personne sur 2 qui ont la maladie xxx
2. Un laboratoire pharmaceutique prétend que son nouveau médicament est meilleur (des pré études suggère que ce nouveau médicament est un candidat plausible)
3. On peut tester le nouveau médicament sur 50 volontaires qui ont la maladie xxx

Un exemple simple

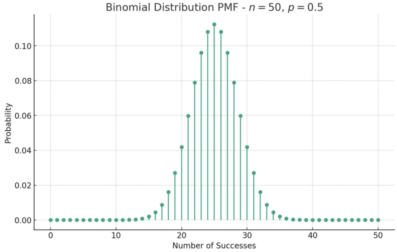
Vous être l'ARS : comment décider ?

1. Un médicament soulage 1 personne sur 2 qui ont la maladie xxx
$$H_0 : \ p = 0,5$$
2. Un laboratoire pharmaceutique prétend que son nouveau médicament est meilleur (des pré études suggère que ce nouveau médicament est un candidat plausible)
$$H_1 : \ p > 0,5$$
3. On peut tester le nouveau médicament sur 50 volontaires qui ont la maladie xxx
$$\text{on rejette } H_0 \text{ si } X > \text{seuil}$$

Le cas du médicament actuel (le monde sous H_0)

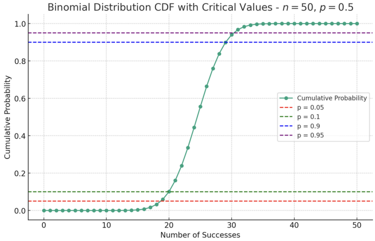
Comment déterminer le seuil ?
On veut être convaincu que le résultat du nouveau médicament n'est pas dû au hasard

ChatGPT



Here's the plot of the probability distribution for a binomial distribution with $n = 50$ tries and a probability $p = 0.5$ of success on each try. The x-axis represents the number of successes out of 50 tries, and the y-axis shows the probability of achieving that number of successes. This plot shows the symmetrical nature of the distribution when $p = 0.5$, centering around 25 successes.

ChatGPT



Here's the enhanced plot of the cumulative probability distribution for a binomial distribution with $n = 50$ tries and a probability $p = 0.5$. I've added dashed lines at cumulative probabilities of 0.05, 0.1, 0.9, and 0.95 to illustrate critical points on the distribution curve. Each line corresponds to a different color, making it easy to see where these probabilities fall in relation to the number of successes. [>_]

La question, hypothèse et modèle statistique

- La question : le nouveau médicament soulage -t-il plus de gens ?
- hypothèse : On veut être convaincu que le résultat du nouveau médicament n'est pas dû au hasard : H_0
- modèle statistique : le nombre de succès suit un e loi binomiale

si H_0 est vraie que peut-on s'attendre à observer

La question du test : un exemple simple

Dans un sac j'ai mis deux types de pièces. Je tire une pièce dans le sac, je la lance 10 fois et j'obtiens k piles. De quelle pièce s'agit-il ?

1. j'ai mis deux pièces dans un sac
   - une équilibrée ($p = 0,5$) et l'autre pas ($p = 0,7$)
   - hypothèses/décision, vraisemblance, règle de décision
2. j'ai mis mille et une pièces dans un sac
   - mille équilibrées ($p = 0,5$) et l'une seule qui ne l'est pas ($p = 0,7$)
   - hypothèses déséquilibrées, rapport de vraisemblance + seuil
3. j'ai mis deux pièces dans un sac, mais une que je ne connais pas.
   - une équilibrée ($p = 0,5$) et l'autre non, mais p est inconnu.
   - hypothèses déséquilibrées, rapport de vraisemblance + 2 seuils

Règle de décision du maximum a posteriori

Règle de décision du maximum a posteriori (ce que l'on veut)

On décide H_0 quand :

$$\mathbb{P}(H_0, p = 0,5 \,|\, X_1, \ldots, X_n) \geq \mathbb{P}(H_1, p = 0,7 \,|\, X_1, \ldots, X_n)$$

soit quand

$$\frac{\mathbb{P}(p = 0,5 \,|\, X_1, \ldots, X_n)}{\mathbb{P}(p = 0,7 \,|\, X_1, \ldots, X_n)} \geq 1$$

or

$$\frac{\mathbb{P}(p = 0,5 \,|\, X_1, \ldots, X_n)}{\mathbb{P}(p = 0,7 \,|\, X_1, \ldots, X_n)} = \frac{\mathbb{P}(X_1, \ldots, X_n \,|\, p = 0,5)\,\mathbb{P}(p = 0,7)}{\mathbb{P}(X_1, \ldots, X_n \,|\, p = 0,7)\,\mathbb{P}(p = 0,5)}$$

ce qui est donc équivalent à

Règle du rapport de vraisemblance (ce que l'on a)

$$\frac{\mathbb{P}(X_1, \ldots, X_n \,|\, p = 0,5)}{\mathbb{P}(X_1, \ldots, X_n \,|\, p = 0,7)} \geq \frac{\mathbb{P}(p = 0,5)}{\mathbb{P}(p = 0,7)} = \text{seuil}$$

Vraisemblances et rapport de Vraisemblances

$$\mathbb{P}(X_1, \ldots, X_n \mid p = 0,5)$$

$$\mathbb{P}(X_1, \ldots, X_n \mid p = 0,7),$$

$$\frac{\mathbb{P}(X_1, \ldots, X_n \mid p = 0,5)}{\mathbb{P}(X_1, \ldots, X_n \mid p = 0,7)}$$

k	0	1	2	3	4	5	6	7	8	9	10
p 0.5	.001	.010	.044	.117	.205	.246	.205	.117	.044	.010	.001
p 0.7	.000	.000	.001	.009	.037	.103	.200	.267	.233	.121	.028
$L0/L1$	165	70.8	30.4	13.0	5.58	2.39	1.02	.439	.188	0.08	0.03

Règle de décision : On décide H_0 quand :

$$\frac{\mathbb{P}(X_1, \ldots, X_n \mid p = 0,5)}{\mathbb{P}(X_1, \ldots, X_n \mid p = 0,7)} \geq \text{seuil}$$

Exemple : le cas gaussien

La question du test

- Laquelle de ces deux variétés s'agit-il Varonne ou Touselles de Mayan ?
  -> seule la variété Varonne permet de faire du bon whisky

- Le taux de naissance des deux genres est égal (Laplace 1778)
  réponses : oui ou non

- La durée de vie de ce téléphone portable est inférieure à 1000 jours
  réponses : oui ou non

- Ce nouveau vaccin est efficace (plus efficace que l'ancien) ?
  actions : si oui on va le distribuer

- Ce candidat va gagner les élections
  réponses : oui, non, on ne peu pas dire

Attention aux raisonnements faux

- Laquelle de ces deux variétés s'agit-il Varonne ou Touselles de Mayan ?
  -> seule la variété Varonne permet de faire du bon whisky

- Le taux de naissance des deux genres est égal (Laplace 1778)
  non parce qu'on à observé 502 naissance de garçon et 498 de filles (et $502 \neq 498$)

- La durée de vie de ce téléphone portable est inférieure à 1000 jours
  non parce que le mien est tombé en pane au bout de 978 jours

- Ce nouveau vaccin est efficace (plus efficace que l'ancien) ?
  non parce qu'on a observé des personnes malades qui ont été vaccinées

- Ce candidat va gagner les élections
  oui parce que les sondage lui prédisent un score de 51 %

La question du test : deux hypothèses s'affrontent

La durée de vie de ce téléphone portable est inférieure à 1000 jours -> réponses : oui ou non
du point de vue du client : H0 : non H1 : oui On n'y croit pas
Ce nouveau vaccin est efficace (plus efficace que l'ancien) ? -> actions : si oui on va le distribuer H0 : non H1 : oui On ne va pas se lancer dans le fabrication d'une nouveau vaccin pour rien.
Ce candidat va gagner les élections -> réponses : oui, non, on ne peu pas dire
il faut faire deux tests
L'hypothèse nulle notée H0
L'hypothèse alternative notée H1 est l'hypothèse complémentaire à l'hypothèse nulle

hypothèses, modèle statistique et p-valeur

Definition (p-valeur)

C' est la probabilité pour un modèle statistique donné sous l'hypothèse nulle d'obtenir la même valeur ou une valeur encore plus extrême que celle observée.

La valeur p n'est pas la probabilité que l'hypothèse H_0 soit vraie.
La valeur p indique dans quelle mesure les données sont conformes à l'hypothèse H_0

erreur à priori : α
erreur a posteriori : p-valeur

Test = hypothèses + décision

Soient

- X une variable aléatoire.
- On considère des hypothèses qui portent sur la nature d'une variable aléatoire
- Deux hypothèses non symétriques : H_0 et H_1
- On cherche à tester H_0 à partir d'un Échantillon $(X_1, \ldots, X_n)$

Definition (Test statistique)

Tester deux hyothèses H_0 et H_1 c'est définir une fonction de décision D

$$D : \begin{array}{ccc} \mathbb{R}^n & \longmapsto & \{0,1\} \\ (x_1, \ldots, x_n) & \longrightarrow & D(x_1, \ldots, x_n) \end{array}$$

si $D(x_1, \ldots, x_n) = 1$ on rejette H_0 (on décide H_1)

Cette fonction D réalise une partition de l'espace des échantillons

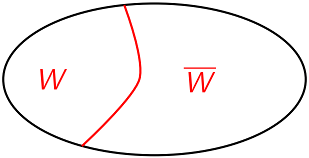
Hypothèses et décisions

La fonction de décision réalise une partition de l'ensemble des échantillons

Definition (Régions de décision)

Région d'acceptation : l'ensemble $\overline{W}$ des réalisations $(x_1, \ldots, x_n)$ pour lesquelles on garde H_0 $(D(x_1, \ldots, x_n) = 0)$

Région critique : l'ensemble W des réalisations $(x_1, \ldots, x_n)$ pour lesquelles on rejette H_0 $(D(x_1, \ldots, x_n) = 1)$



On a deux hypothèses et deux décisions possibles

Mesures d'erreur et d'efficacité d'un test

4 situations :

	H_0	H_1
$D(x_1, \ldots, x_n) = 0$	$1 - \alpha$	β
$D(x_1, \ldots, x_n) = 1$	α	$1 - \beta$

2 erreurs :

α probablité de rejeter H_0 ($D(x_1, \ldots, x_n) = 0$) alors que H_0 est vraie

$$\alpha = \mathbb{P}(D = 1 | H_0) = \int_{(x_1, \ldots, x_n) \in W} L(x_1, \ldots, x_n, H_0) \, dx_1, \ldots, dx_n$$

β probabilité de garder H_0 ($D(x_1, \ldots, x_n) = 0$) alors que H_1 est vraie

$$\beta = \mathbb{P}(D = 0 | H_1) = \int_{(x_1, \ldots, x_n) \in \overline{W}} L(x_1, \ldots, x_n, H_1) \, dx_1, \ldots, dx_n$$

Definition (puissance d'un test (d'une règle de décision))

$1 - \beta$ la probabilité de rejeter hypothèses nulle avec raison

Méthodes de construction des tests

Comment construire D ?

Cas des tests à hypothèses simples (HS = une hypothèse paramétrique qui caractérise entièrement la loi de probabilité)

Definition (Tests du rapport de vraisemblance)

$$W = \left\{ x_1, \ldots, x_n \mid \frac{L(x_1, \ldots, x_n, H_1)}{L(x_1, \ldots, x_n, H_0)} > k \right\}$$

$$D(x_1, \ldots, x_n) = \begin{cases} 0 & \text{si } \frac{L(x_1, \ldots, x_n, H_1)}{L(x_1, \ldots, x_n, H_0)} \leq k \\ 1 & \text{sinon} \end{cases}$$

Exemple : tester deux valeurs pour l'espérance d'une loi normale

Comment trouver k ?

Pour un test de risque α on a

$$\alpha = \mathbb{P}(D = 1 | H_0) = \mathbb{P}\left(\frac{L(x_1, \ldots, x_n, H_1)}{L(x_1, \ldots, x_n, H_0)} > k \;\middle|\; H_0 \right)$$

La méthode de Neyman–Pearson

Definition (Le principe de Neyman–Pearson)

pour un α fixé, trouver la fonction de décision qui minimise β
(ou qui maximise $1 - \beta$ la puissance du test)

$$\min_D \quad \mathbb{P}(D = 0|H_1) = \int_{(x_1,\ldots,x_n)\in\overline{W}} L(x_1, \ldots, x_n, H_1)\, dx_1, \ldots, dx_n$$

$$\text{avec} \quad \alpha = \mathbb{P}(D = 1|H_0) = \int_{(x_1,\ldots,x_n)\in W} L(x_1, \ldots, x_n, H_0)\, dx_1, \ldots, dx_n$$

La méthode de Neyman–Pearson

Definition (L'idée de Neyman–Pearson)

Le test du rapport de vraisemblance

$$W = \left\{ x_1, \ldots, x_n \mid \frac{L(x_1, \ldots, x_n, H_1)}{L(x_1, \ldots, x_n, H_0)} > k \right\}$$

tel que k soit fixé de sorte que

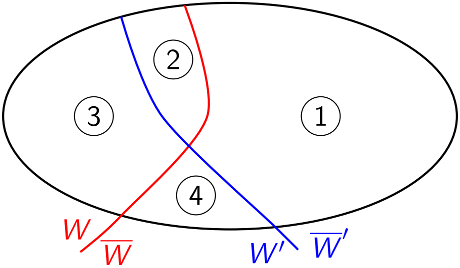
$$\alpha = \int_{(x_1, \ldots, x_n) \in W} L(x_1, \ldots, x_n, H_0)\, dx_1, \ldots, dx_n$$

est optimal au sens du principe de Neyman–Pearson (parmi tous les tests de risque α, c'est celui de puissance maximale).

Preuve : soit W' un autre test de risque α on a donc

$$\alpha = \int_W L(x_1, \ldots, x_n, H_0)\, dx_1, \ldots, dx_n = \int_{W'} L(x_1, \ldots, x_n, H_0)\, dx_1, \ldots, dx_n$$

Comparaison de deux fonctions de décision



Zone 1 $(\overline{W}$ et $\overline{W}') = \dfrac{L(x_1,\ldots,x_n,H_1)}{L(x_1,\ldots,x_n,H_0)} \leq k$ et $\overline{W}'$

Zone 2 $(W$ et $\overline{W}') = \dfrac{L(x_1,\ldots,x_n,H_1)}{L(x_1,\ldots,x_n,H_0)} > k$ et $\overline{W}'$

Zone 3 $(W$ et $W') = \dfrac{L(x_1,\ldots,x_n,H_1)}{L(x_1,\ldots,x_n,H_0)} > k$ et W'

Zone 4 $(\overline{W}$ et $W') = \dfrac{L(x_1,\ldots,x_n,H_1)}{L(x_1,\ldots,x_n,H_0)} \leq k$ et W'

Comparaison de deux fonctions de décision

W et W' sont tout les deux de même risque α :

$$\begin{aligned}
\alpha &= \int_W L(x_1,\ldots,x_n,H_0)\,dx_1,\ldots,dx_n &&= \int_{W'} L(x_1,\ldots,x_n,H_0)\,dx_1,\ldots,dx_n\\
&= \int_{2} L(0)\,dx + \int_{3} L(0)\,dx &&= \int_{3} L(0)\,dx + \int_{4} L(0)\,dx\\
&= \int_{2=\frac{L(\mathbf{1})}{L(\mathbf{0})}>k\cap\overline{W}'} L(0)\,dx &&= \int_{4=\frac{L(\mathbf{1})}{L(\mathbf{0})}\leq k\cap W'} L(0)\,dx
\end{aligned}$$

et

$$\begin{aligned}
\beta_W - \beta_{W'} &= \int_{\overline{W}} L(x_1,\ldots,x_n,H_1)\,dx_1,\ldots,dx_n - \int_{\overline{W}'} L(x_1,\ldots,x_n,H_1)\,dx_1,\ldots,dx_n\\
&= \int_{1} L(1)\,dx + \int_{4} L(1)\,dx - \int_{1} L(1)\,dx - \int_{2} L(1)\,dx\\
&= \int_{4=\frac{L(\mathbf{1})}{L(\mathbf{0})}\leq k\cap W'} \frac{L(1)}{L(0)}L(0)\,dx - \int_{2=\frac{L(\mathbf{1})}{L(\mathbf{0})}>k\cap\overline{W}'} \frac{L(1)}{L(0)}L(0)\,dx\\
&\leq k\int_{4=\frac{L(\mathbf{1})}{L(\mathbf{0})}\leq k\cap W'} L(0)\,dx - k\int_{2=\frac{L(\mathbf{1})}{L(\mathbf{0})}>k\cap\overline{W}'} L(0)\,dx = 0
\end{aligned}$$

Le théorème de Neyman–Pearson

Theorem (Neyman–Pearson)

Soient H0 et H1 deux hypothèses simples. Soit UPP le test qui rejette H0 chaque fois que le rapport de vraisemblance est inférieur à c et au niveau de signification α. Alors tout autre test dont le niveau de signification est inférieur ou égal à α a une puissance inférieure ou égale à celle du test UPP (celui du rapport de vraisemblance).

Definition (hypothèses simples)

Une hypothèse simple est une hypothèse qui spécifie complètement la distribution de probabilité sous-jacente. Cela signifie que l'hypothèse ne contient aucun paramètre libre ou inconnu qui doit être estimé.

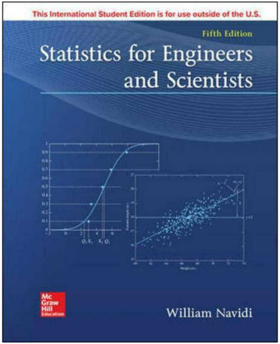
Une hypothèse qui n'est pas simple est dite **composite**

Exemples : $X \sim \mathcal{N}(\mu; \sigma^2 = 1)$ (σ^2 connu)

$$H_a : \mu = \mu_0 \qquad H_b : \mu \leq \mu_0 \qquad H_c : \mu \neq \mu_0$$

Lecture road map

1. La question du test
2. Formalisation des test d'hypothèse
3. Le cas de l'espérance d'une loi normale



https://moodle.insa-rouen.fr/course/view.php?id=93

Loi normale : tests de la moyenne

Selon les hypothèses

$$\left\{ \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \ (> \mu_0) \end{array} \right. \quad \left\{ \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \ (> \mu_0) \end{array} \right. \quad \left\{ \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{array} \right.$$

Forme de la région critique W

$$\overline{X} > k \qquad\qquad \overline{X} < k \qquad\qquad \overline{X} < k_1 \ \cup \ k_2 < \overline{X}$$

La détermination de k dépend du niveau de signification α et de la loi de $\overline{X}$

Tests de la moyenne : Z-test et T-test

La détermination de k dépend du niveau de signification α et de la loi de $\overline{X}$
Par hypothèse on a $X \sim \mathcal{N}(\mu, \sigma^2)$,
et donc si σ^2 est connu :

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

et sinon

$$\frac{\overline{X} - \mu}{\sqrt{\frac{\widehat{\sigma}^2}{n}}} \sim \mathcal{T}_{n-1}$$

avec $\widehat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$
En pratique, σ^2 est presque toujours inconnu.

t-test

On t
ttest_

scipy.stats.ttest_1samp

`scipy.stats.ttest_1samp(a, popmean, axis=0, nan_policy='propagate', alternative='two-sided', *, keepdims=False)` [source]

Calculate the T-test for the mean of ONE group of scores.

This is a test for the null hypothesis that the expected value (mean) of a sample of independent observations *a* is equal to the given population mean, *popmean*.

Parameters:

a : *array_like*

Sample observations.

popmean : *float or array_like*

Expected value in null hypothesis. If array_like, then its length along *axis* must equal 1, and it must otherwise be broadcastable with *a*.

axis : *int or None, default: 0*

If an int, the axis of the input along which to compute the statistic. The statistic of each axis-slice (e.g. row) of the input will appear in a corresponding element of the output. If `None`, the input will be raveled before computing the statistic.

nan_policy : *{'propagate', 'omit', 'raise'}*

Defines how to handle input NaNs.

- `propagate`: if a NaN is present in the axis slice (e.g. row) along which the statistic is computed, the corresponding entry of the output will be NaN.
- `omit`: NaNs will be omitted when performing the calculation. If insufficient data remains in the axis slice along which the statistic is computed, the corresponding entry of the output will be NaN.
- `raise`: if a NaN is present, a `ValueError` will be raised.

alternative : *{'two-sided', 'less', 'greater'}, optional*

Defines the alternative hypothesis. The following options are available (default is 'two-sided'):

- 'two-sided': the mean of the underlying distribution of the sample is different than the given population mean (*popmean*)
- 'less': the mean of the underlying distribution of the sample is less than the given population mean (*popmean*)
- 'greater': the mean of the underlying distribution of the sample is greater than the given population mean (*popmean*)

keepdims : *bool, default: False*

If this is set to True, the axes which are reduced are left in the result as dimensions with size one. With this option, the result will broadcast correctly against the input array.

Returns:

result : `TtestResult`

An object with the following attributes:

statistic : *float or array*

The t-statistic.

pvalue : *float or array*

The p-value associated with the given alternative.

df : *float or array*

The number of degrees of freedom used in calculation of the t-statistic; this is one less than the size of the sample (`a.shape[axis]`).

Dualité tests d'hypothèses et intervalles de confiance : l'exemple de la loi normale

Intervalle de confiance :

$$\mathbb{P}\left(\overline{X} - 1,96\,\frac{\sigma}{\sqrt{n}} \;\leqslant\; \mu \;\leqslant\; \overline{X} + 1,96\,\frac{\sigma}{\sqrt{n}}\right) = 0,95$$

Pour le test

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

On a la région d'acceptation suivante :

$$\overline{W} = \left\{(X_1, \ldots, X_n) \mid \mu_0 - 1,96\,\frac{\sigma}{\sqrt{n}} \;\leqslant\; \overline{X} \;\leqslant\; \mu + 1,96\,\frac{\sigma}{\sqrt{n}}\right\}$$

L'intervalle de confiance autour de $\overline{X}$ est constitué de toutes les valeurs de μ_0 pour lesquelles l'hypothèse nulle $H_0 : \mu = \mu_0$ est conservée.