IV. Tests d'une hypothèse scalaire sur un paramètre

- $(Y_1,\ldots,Y_n) \approx$ i.i.d. loi $\ell(y;\theta)$
- L'ensemble des valeurs possibles du paramètre θ est Θ.
- Soit Θ_o une partie de Θ. On veut décider si θ est dans Θ_o ou non. On dit que l'on teste
  $H_o = \{\theta \in \Theta_o\}$ contre $H_1=\{\theta \notin \Theta_o\}$

Le choix se fait entre deux décisions possibles, correspondant à une partition de l'ensemble des lois possibles des y, et donc à une partition de l'ensemble des valeurs possibles du paramètre. On désigne par H_o l'hypothèse selon laquelle la vraie loi des observations correspond à une valeur de θ dans Θ_o ..

Exemple : Le nombre annuel d'accidents à un carrefour est une v.a. de Poisson de paramètre λ. L'administration (la DDE, ou Direction Départementale de l'Equipement) se demande si $\lambda \geq 1$ ou non. Son problème de décision est : faut-il aménager le carrefour (indispensable si $\lambda \geq 1$), ou peut-on le laisser en l'état (tolérable si $\lambda < 1$) ?

Elle dispose du nombre d'accidents survenus chaque année pendant 5 ans.

Modèle statistique : $Y_1,\ldots, Y_5$ sont ≈i.i.d. Poisson(λ), Θ = {réels > 0} et Θ_o = {réels ≥ 1}.

Remarque : on suppose donc, en particulier, que l'état du carrefour ne se dégrade pas pendant les cinq années d'observation (λ reste constant) et qu'il n'y a pas de téléscopages en série (« événements rares et indépendants »). Ce modèle ne serait pas acceptable si le problème était lié à la dégradation du revêtement ou si on étudiait la sécurité sur une autoroute.

Hypothèse de base, région critique

- L'hypothèse de base $H_o = \{\theta \in \Theta_o\}$ est celle dont le rejet à tort a les conséquences les plus fâcheuses.
- Règle de décision (ou « test ») : règle qui fait correspondre à chaque résultat $(y_1, \ldots, y_n)$ une des deux décisions « acceptation de H_o » ou « refus de H_o »
- Région critique du test : ensemble des observations possibles qui conduisent à la décision « refus de H_o ».

Deux types d'erreurs sont possibles : aménager le carrefour alors que $\{0 < \lambda < 1\}$ et laisser le carrefour en l'état alors qu'il est dangereux. La DDE va choisir le terme de l'alternative dont le rejet à tort a les conséquences les plus fâcheuses : c'est $\{\lambda \geq 1\}$.

La DDE se demande si $\{\lambda \geq 1\}$ ou non. On dit qu'elle teste l'hypothèse $\{\lambda \geq 1\}$ contre l'hypothèse $\{0 < \lambda < 1\}$.

La règle de décision est mise au point avant de faire les observations : on construit à l'avance le mécanisme de choix entre les deux hypothèses.

La région critique d'un test (elle caractérise le test) est la région de refus de l'hypothèse de base.

Par exemple : λ inconnu étant estimé par λ^* = moyenne empirique de l'échantillon, une règle envisageable est de refuser $\{\lambda \geq 1\}$ si $\lambda^* < 1$ (sur 5 ans, au plus 4 accidents constatés).

On pourrait aussi ne refuser $\{\lambda \geq 1\}$ que si $\lambda^* < 0{,}4$ (sur 5 ans, au plus 1 accident constaté).

La seconde règle conduit à de plus faibles probabilités de refuser Ho alors qu'elle est vraie, mais également à de plus faibles probabilités de refuser Ho alors qu'elle fausse.

On pourrait fonder la règle sur une autre statistique que la moyenne empirique des observations..

Nous allons exposer les critères qui font préférer une règle de décision à une autre et expliquer dans un cas simple la méthode de calcul de la statistique optimale à utiliser.

IV.1. Définition et risques associés à un test

- région critique W = ensemble des $(y_1,\ldots,y_n)$ pour lesquels la règle conduit à rejeter H_o
- risques de première espèce pour θ dans H_o: $\alpha(\theta) = P_\theta\{W\}$ pour chaque θ dans H_o.
- Puissances, pour θ dans H_1:
  $\eta(\theta) = P_\theta\{W\}$ pour chaque θ dans H_1.
- Seuil du test W : le $\alpha(\theta)$ maximum
  seuil α = Max $\alpha(\theta)$ parmi les θ dans H_o.

32

W est la région de rejet de Ho.

La probabilité de W se calcule pour chaque loi du modèle :

si la loi *P* fait partie de Ho, *P*(W) est une probabilité de se tromper en refusant Ho : on l'appelle un risque de première espèce.

si la loi *P* fait partie de H1, *P*(W) est une probabilité de ne pas se tromper en refusant Ho : on l'appelle une puissance.

Le seuil du test est le plus grand risque de première espèce.

On appelle parfois $(1-\eta(\theta)) = \beta(\theta)$ un risque de deuxième espèce : c'est en effet un risque d'accepter Ho alors qu'elle est fausse.

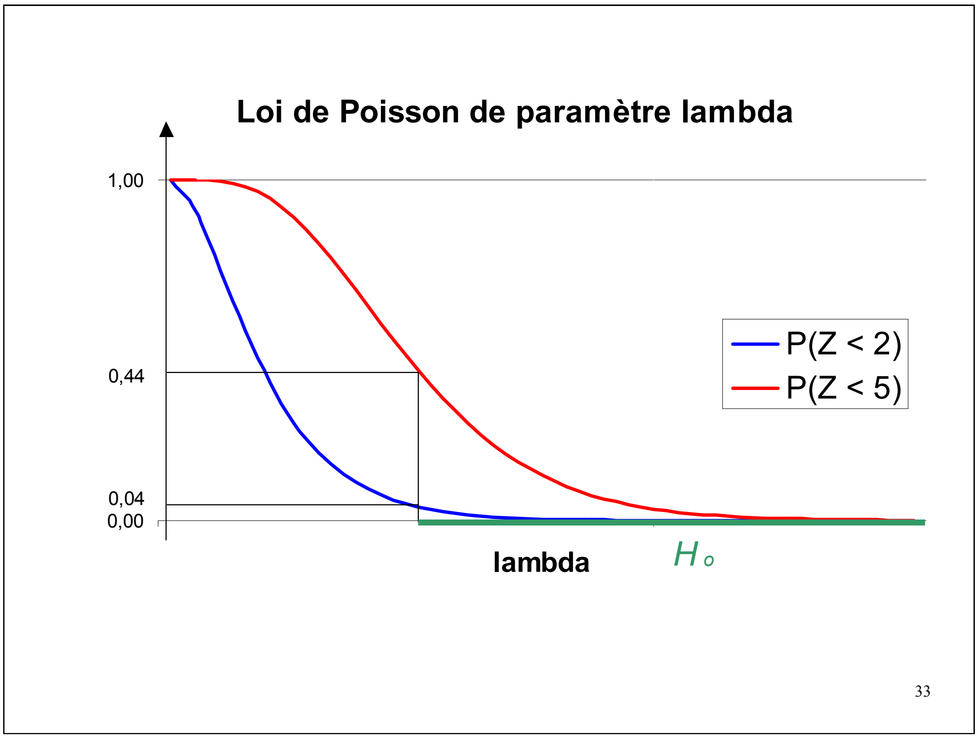
Exemple, suite

Considérons la règle suivante : on refuse Ho (i.e. on considère que $\lambda < 1$, i.e. on ne procède à aucun aménagement du carrefour) si le nombre total d'accidents survenus pendant 5 ans est inférieur ou égal à 1.

si $Y_1,\ldots,Y_5 \approx i.i.d.P(\lambda)$, alors $\sum_{i=1}^{5} Y_i \approx P(5\lambda)$

On en déduit la courbe des $P_\lambda\left\{\sum_i Y_i \leq 1\right\} = e^{-5\lambda}\dfrac{(5\lambda)^0}{1} + e^{-5\lambda}\dfrac{(5\lambda)^1}{1} = e^{-5\lambda}(1+5\lambda)$

D'où le seuil α et les puissances du test



Remarquer que :

• les deux courbes correspondent à deux tests différents :

> • au dessous, en bleu, la règle « je renonce à considérer le carrefour comme dangereux s'il n'y a pas eu plus d'un accident en 5 ans »
>
> • au dessus, en rouge : « je renonce à considérer le carrefour comme dangereux si le nombre d'accidents en 5 ans reste inférieur à 5 »

• pour $\{\lambda \geq 1\}$, les valeurs correspondent à des risques de première espèce. La valeur maximum atteinte sous Ho est celle prise en $\lambda = 1$, car ces risques sont décroissants en λ. Les seuils observés sont respectivement 0.040 et 0.440.

• pour $\{0 < \lambda < 1\}$, les valeurs correspondent à des puissances $\eta(\lambda)$.

• Aucun des deux tests n'est « meilleur » que l'autre : lorsque les puissances sont plus grandes, les risques de première espèce et le seuil le sont aussi. Le choix correspond au choix (arbitraire) du seuil.

• Enfin, le choix de $\{\lambda \geq 1\}$ pour hypothèse Ho est lui aussi crucial : dans l'optique inverse, ce sont les valeurs $1 - \eta(\lambda)$ qui auraient représenté les risques de première espèce, conduisant à des « seuils » respectifs de 0.960 et 0.560.

L 'hypothèse de base est celle dont le rejet à tort a les conséquences les plus fâcheuses pour le décideur

c'est elle dont on limite a priori les probabilités de rejet à tort.

IV.2. Propriétés d'un test

- Test sans biais = sa puissance ne peut être inférieure à son seuil :

$\forall\ \theta_o$ de Θ_o et θ_1 de Θ_1 : $\quad P_{\theta o}(W) \leq P_{\theta 1}(W)$

- Test W_1 meilleur que W_2 = W_1 et W_2 sont de même seuil, et la puissance de W_1 est toujours supérieure à celle de W_2 :

  $\forall\ \theta_1$ de Θ_1 : $\quad P_{\theta 1}(W_1) \geq P_{\theta 1}(W_2)$

- Un test W est admissible s'il n'existe pas de test meilleur que lui.

Exemple : test sur l'espérance d 'une variable Normale de variance 1 : on teste { m = 0} contre {m ≠0}.

Règles choisies :

W1 : on refuse {m = 0}si $\quad \overline{Y}_n > \dfrac{1.645}{\sqrt{n}}$

W2 : on refuse {m = 0}si $\quad \overline{Y}_n > \dfrac{1.96}{\sqrt{n}}$ ou $\overline{Y}_n < -\dfrac{1.96}{\sqrt{n}}$

$P_o(W1) = P_o(W2) = 5\%$

$P_m(W1) = P\left(U > 1.645 - m\sqrt{n}\right)$ tend vers 0 lorsque m tend vers -∞: W1 est un test biaisé.

alors que $P_m(W2) = P\left(U > 1.96 - m\sqrt{n}\right) + P\left(U < -1.96 - m\sqrt{n}\right)$ tend vers 1.

La puissance de W2 est plus faible que celle de W1 pour les m > 0, mais W2 est sans biais : voir la courbe dans l'exemple traité dans les TD d'entraînement.

IV.3 Test de $\{\theta = \theta_o\}$ contre $\{\theta = \theta_1\}$

- Il n 'y a ici le choix qu'entre deux lois, correspondant à deux vraisemblances de θ :

  $L(Y_1,\ldots,Y_n;\theta_o)$ et $L(Y_1,\ldots,Y_n;\theta_1)$.

- Théorème de Neyman : la région critique W du test le plus puissant parmi ceux de seuil α est définie par :

$$\begin{cases} W = \left\{ (Y_1,\ldots,Y_n) \middle| \ \dfrac{L(Y_1,\ldots Y_n;\theta_o)}{L(Y_1,\ldots,Y_n;\theta_1)} \leq k \right\} \\ P\left\{ (Y_1,\ldots,Y_n) \in W \middle| \theta_o \right\} = \alpha \end{cases}$$

Hypothèse simple contre hypothèse simple : chaque hypothèse ne contient qu'une seule loi (le modèle ne contient donc que 2 lois possibles).

(i) La première inégalité sert à déterminer la forme de la région critique : on résout l'inégalité de manière à faire apparaître une statistique simple (si possible)

(ii) calcul de la région critique qui correspond au seuil choisi.

Exemple : test sur le paramètre d'une loi de Poisson, on reprend l'exemple précédent. On suppose que λ ne peut prendre que deux valeurs : 0.5 ou 1

Test de λ = 1 contre λ = 0.5 : développement complet du calcul, pour trouver que la règle est fondée sur la somme des Yi, et détermination de A=2 pour avoir seuil donné 12.47%

Autres tests usuels : pour loi normale (test sur m quand σ est connu), loi exponentielle, loi de Bernoulli.

Tableau récapitulatif : théorème que l'on peut appliquer...

A NOTER :

Il ne sera pas demandé en partiel de savoir résoudre l'inégalité (étape (i)), mais il faudra connaître le résultat de ce calcul dans les cas usuels cités, savoir calculer la région critique correspondant à un seuil donné et savoir calculer la puissance du test ainsi défini.

Tests de Neyman usuels

- Bernoulli $B(1,p)$: $\{p = p_o\} \mid \{p = p_1\}$
  où $p_o < p_1$: $W = \{F_n > A\}$
- Poisson $P(\lambda)$: $\{\lambda = \lambda_o\} \mid \{\lambda = \lambda_1\}$
  où $\lambda_o < \lambda_1$: $W = \{\Sigma Y_i > A\}$
- exponentielle $\gamma(a)$: $\{a = a_o\} \mid \{a = a_1\}$
  où $a_o < a_1$: $W = \{\overline{Y}_n > A\}$
- Normale $N(m,\sigma^2)$, σ^2 connu
  $\{m = m_o\} \mid \{m = m_1\}$, où $m_o < m_1$:
  $W = \{ \overline{Y}_n > A\}$

36

Remarquer le sens des inégalités, qui sont cohérentes avec l'intuition la plus élémentaire : on refuse la plus petite valeur du paramètre si les observations conduisent à une valeur estimée trop grande.

De même, on aura des régions critiques dans l'autre sens si la valeur du paramètre dans l'hypothèse de base est la plus grande des deux :

• Bernoulli : lorsque $p_o > p_1$, on refuse ($p = p_o$) si $F_n < A$.

• Poisson : lorsque $\lambda_o > \lambda_1$, on refuse ($\lambda = \lambda_o$) si $\Sigma Y_i < A$.

• Exponentielle : lorsque $a_o > a_1$, on refuse ($a = a_o$) si $\overline{Y_n} < A$.

• Normale : lorsque $m_o > m_1$, on refuse ($m = m_o$) si $\overline{Y_n} < A$.

IV.4. Test d'une hypothèse simple contre une hypothèse multiple

Test de $\{\theta = \theta_o\}$ contre $\{\theta \in \Theta_1\}$

Un test W est u.p.p. (uniformément le plus puissant) parmi les tests de seuil α si :

- *W* est de seuil α
- pour tout $\theta \in \Theta_1$, et pour tout test *W** de seuil α, la puissance de *W* en θ est supérieure à celle de *W** en θ : $\eta(\theta) \quad \eta^*(\theta)$

On peut ainsi constater que les tests usuels décrits précédemment sont en fait u.p.p. pour tester $\{\theta=\theta_o\}$ contre $\{\theta > \theta_o\}$: la région trouvée est optimale pour toute valeur de θ_1, et elle est indépendante de θ_1 tant que θ_1 reste supérieur à θ_o.

Détermination d'un test u.p.p. s'il existe

Soit θ_1 une valeur choisie dans Θ_1 .

On construit le test le plus puissant de seuil α pour tester $\{\theta = \theta_o\}$ contre $\{\theta = \theta_1\}$.

si la région critique obtenue est $W(\theta_1) = W$ indépendante de θ_1, le test W est <u>u.p.p.</u> parmi les tests de seuil α.

Exemple : loi Normale (m;1), test de $\{m = 0\}$contre $\{m > 0\}$

soit m_1 une valeur > 0 de m. On construit le test de Neyman pour tester au seuil 5% $\{m = 0\}$contre $\{m = m_1\}$avec un échantillon de taille 16 :

$L_o/L_1 < k$ donne, pour tout $m_1 > 0$, une région critique de la forme $W_1 = \overline{Y}_{16} > A$

Pour calculer A, seule intervient la valeur de m dans l 'hypothèse de base. La condition d'un seuil 5% entraîne:

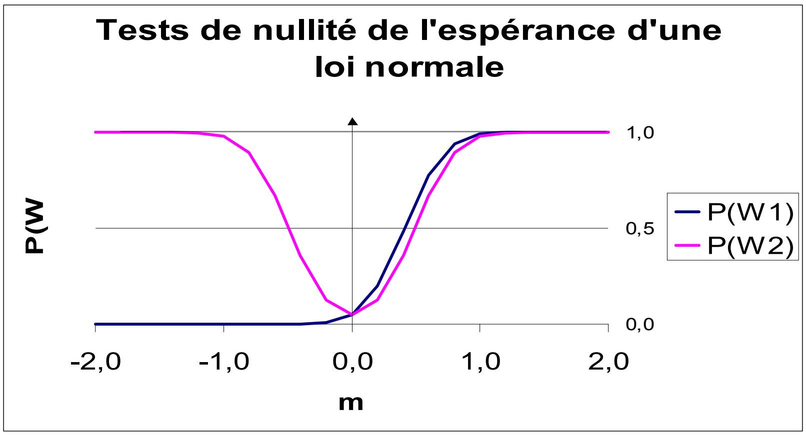
$$\frac{A - 0}{\sigma / \sqrt{16}} = 1.645 \Rightarrow W_1 = \overline{Y}_{16} > \frac{1.645}{\sqrt{16}} = 0.411$$

Cette région critique ne dépend pas de la valeur de m_1 tant m_1 reste supérieur à 0. Le test de région critique W1 est donc de seuil 5% et uniformément le plus puissant pour tester $\{m = 0\}$ contre $\{m > 0\}$.

Le graphique suivant trace la puissance de ce test, courbe bleue.

Echantillon de taille 16 de N(*m*;1)

$$W_1 = \overline{Y}_{16} > \frac{1.645}{\sqrt{n}} \qquad W_2 = \left\{ \overline{Y}_{16} > \frac{1.960}{\sqrt{n}} \right\} \cup \left\{ \overline{Y}_{16} < -\frac{1.960}{\sqrt{n}} \right\}$$



39

Par contre, si on teste {m=0}contre {m≠0}, la courbe d'efficacité de W1 se continue dans les valeurs négatives de m, où elle est inférieure au seuil 5%. Le test est biaisé.

On peut montrer que le test uniformément le plus puissant parmi les tests sans biais de seuil 5% est obtenu en prenant les valeurs de la moyenne correspondant aux plus petites valeurs de la densité.

La forme unimodale de la densité de la loi normale entraîne que W2 doit être formée des valeurs trop petites et des valeurs trop grandes de la moyenne. C'est une région critique bilatérale, correspondant à « une moyenne trop grande en valeur absolue » :

$$W_2 = \left\{ \overline{Y}_{16} > B_1 \right\} \cup \left\{ \overline{Y}_{16} < -B_2 \right\}$$

La symétrie de la loi normale (autour de 0 sous H0) entraîne alors B1 = B2 = B. Le seuil étant fixé à 5%, B = 1.960.

La courbe rouge montre la fonction d'efficacité de W2. On constate effectivement qu'elle est toujours au dessus de 5% (test sans biais) et qu'elle est au dessous de W1 pour les valeurs positives de m.

IV.5. Tests d'hypothèses multiples usuels

5.1. Test sur l'espérance d'une loi Normale de variance inconnue

$Y_1, \ldots, Y_n \approx$ i.i.d. $N(m ; \sigma^2)$, m réel, $\sigma^2 > 0$

$H_o = \{m = m_o\}$ contre $H_1 = \{m > m_o\}$

$$\left| \frac{\overline{Y}_n - m_o}{\sqrt{\sigma^2 / n}} \right| > u_\alpha \quad \text{devient:} \quad \left| \frac{\overline{Y}_n - m_o}{\sqrt{S^2 / n}} \right| > t_\alpha$$

où $P\{|U| > u_\alpha\} = \alpha$, $U \approx N(0;1)$

où $P\{|T| > t_\alpha\} = \alpha$, $T \approx T(n-1)$ loi de Student

La variance est remplacée par son estimation sans biais (exactement même problème que pour les intervalles de confiance)

Le degré de liberté est n-1 car un seul paramètre est estimé dans l'espérance de Y :

$$S_2 = \frac{\sum \left(Y_i - \overline{Y}_n\right)^2}{n-1} \quad \text{où} \quad \sum \left(Y_i - \overline{Y}_n\right) = 0$$

Dans le cas du modèle linéaire, plusieurs paramètres devront être estimés, et le degré de liberté sera d'autant plus petit.

Exemple : test de {m = 0} contre {m > 0} lorsque σ^2 n 'est pas connue, sur un échantillon de taille 16.

Table de STUDENT à (16-1) = 15 degrés de liberté, seuil 5%

$t_\alpha = 2.131$

$$\frac{\overline{Y}_n - 0}{S / \sqrt{16}} > 2{,}131 \Rightarrow W = \left\{ 4 \frac{\overline{Y}_n}{S} > 2{,}131 \right\}$$

Exemple

- $H_o = \{m = 0\}$ contre $H_1 = \{m \neq 0\}$, sur un échantillon de taille 16 d'une N(m, σ^2)

STUDENT à 15 degrés de liberté, seuil 5% :

$t_\alpha = 2.131$ $$W = \left\{ \left| \frac{\overline{Y}_n}{S} \right| > 0.533 \right\}$$

- Observations : $\overline{y}_n = 2.58$, $s^2 = 11.90$

$$\frac{\overline{y}_n}{s} = \frac{2.58}{\sqrt{11.90}} = 0.748$$

Conclusion : on décide $\{m \neq 0\}$

41

Remarquons ici une autre façon de définir la règle de décision de seuil α donné :

On calcule $\frac{\overline{y}_n}{s}$ la valeur observée de la statistique $\frac{\overline{Y}_n}{S}$

On calcule $PROB = P\left\{ \left| \frac{\overline{Y}_n}{S} \right| > \left| \frac{\overline{y}_n}{s} \right| \right\}$

On refuse H_o si $PROB < \alpha$

En effet, $P\left\{ \left| \frac{\overline{Y}_n}{S} \right| > C \right\}$ est une fonction décroissante de C.

Nous avons déterminé que sa valeur pour 0.533 est égal à α, si PROB est plus petite que α, c'est que la statistique observée est supérieur à 0.533 et se trouve donc dans la région critique.

Cette méthode de calcul est utilisée dans les logiciels, ce qui évite de consulter une table : il suffit de comparer PROB au seuil que l'on s'est fixé.

Cette probabilité est parfois désignée sous le nom de P-value.

Dans notre exemple,

P{| STUDENT(15) | > 0.748}

Test de nullité d'un coefficient dans un modèle linéaire

- $Y_1, \ldots, Y_{16}$ indépendantes≈ N($ax_i + b$; σ^2)
- Estimation MCO : $y_i = \underset{(0.0205)}{0.0512}\, x_i + 1.347$
- Test de $\{a = 0\}$contre $\{a \neq 0\}$:

  (16 - 2) = 14 degrés de liberté : $t_\alpha = 2.145$

  $t_{observé} = 0.0512/0.0205 = 2.498 > 2.145$

  Conclusion : on décide que $\{a \neq 0\}$

42

On a déjà rencontré la loi des estimateurs de a et b, pour calculer des intervalles de confiance pour les coefficients. On fait ici un test bilatéral. C'est le cas le plus fréquent, c'est pourquoi la table de Student usuelle donne directement la valeur t_α qui a la probabilité α d'être dépassée en valeur absolue.

La valeur t_α lue dans la table est la même pour tous les coefficients du modèle linéaire.

Remarque : vérifier que l'intervalle de confiance η sur a est l'ensemble des valeurs a_o qui ne pourraient être refusées si on testait $a = a_o$ contre $a \neq a_o$ au seuil $\alpha = 1 - \eta$.

Procédure de test

- modèle statistique
  (Observations, famille des lois, ens. des paramètres)
- hypothèses testées : H_o contre H_1
- statistique Z utilisée et forme de la région critique W
- loi de Z (au moins sous H_o)
- seuil α choisi, et région W_α correspondante
- valeur observée Z_{obs} de Z
- Conclusion

43

• modèle statistique : observations, famille des lois, ensemble des paramètres

On parle aussi de modèle structurel, ou de structure du test

Les hypothèses ne sont pas remises en cause par le test, mais doivent être justifiées ou testées si possible avant la mise en œuvre du test.

• hypothèses testées : H_o contre H_1

dans la formalisation du problème posé, H_1 est aussi importante que H_o

• statistique Z utilisée et forme de la région critique W

énoncé du type de procédure envisagée.

• loi de Z (au moins sous H_o)

pour pouvoir passer au point suivant

• seuil α choisi, et région W_α correspondante

Le choix de α est dicté par les coûts des différentes erreurs possibles.

• calcul de la valeur observée Z_{obs} de Z

en général, il s'agit d'extraire cette valeur d'un listing : les « sorties » d'un calcul commandé à un logiciel sont en général très prolixes, et fournissent systématiquement de nombreux chiffres. A vous de savoir lesquels regarder, et dans quel ordre.

• Conclusion

Si $Z_{obs} \in W_\alpha$: « au risque α, les observations conduisent à *rejeter Ho* » : on explicite complètement H1.

Si $Z_{obs} \notin W_\alpha$: «au risque α, les observations ne permettent pas de *rejeter Ho* »

Les mots « Ho » ou « H1 » ne doivent pas figurer dans la conclusion, mais être remplacés par la condition explicite qui leur correspond pour le paramètre.

5.2. Comparaison de fréquences

- Deux échantillons indépendants de Bernoulli

$X_1,\ldots, X_{n_1} \approx \mathrm{B}(1, p_1)$ $\qquad$ $Y_1,\ldots, Y_{n_2} \approx \mathrm{B}(1, p_2)$

- test de $\{p_2 = p_1\}$ contre $\{p_2 > p_1\}$

[ou $\{p_2 < p_1\}$, ou $\{p_2 \neq p_1\}$]

$n_1 p_1(1- p_1) > 15$

$n_2 p_2(1- p_2) > 15$ $\qquad$ $Z = \dfrac{F_2 - F_1}{\sqrt{F(1-F)\left[\dfrac{1}{n_1} + \dfrac{1}{n_2}\right]}} \underset{H_o}{\approx} N(0;1)$

- refus de H_o si $Z > u_\alpha$

[ou $Z < - u_\alpha$, ou $|Z| > u_\alpha$]

44

p1 est la probabilité théorique de l'événement A dans le premier tirage

F1 est la fréquence d 'occurrence de A dans le premier échantillon.

p1 est la probabilité théorique de l'événement A dans le deuxième tirage

F1 est la fréquence d'occurrence de A dans le deuxième échantillon.

F est la fréquence d'occurrence de A dans l'échantillon global (sous Ho, un seul échantillon d'une même loi de Bernoulli)

On fait ici un test « asymptotique », valable lorsque

$n_1p_1(1-p_1)>15$ et $n_2p_2(1-p_2)>15$

Il existe des tests exacts (la région critique est l'ensemble des observations possibles correspondant aux plus petites probabilités de Z sous Ho).

F_1 # $N(p_1 ; p_1(1-p_1)/ n_1)$

F_2 # $N(p_2 ; p_2(1-p_2)/ n_2)$

F_1-F_2 # $N(p_1 - p_2 ; p_1(1-p_1)/ n_1 + p_2(1-p_2)/ n_2)$

sous $H_o : p_1 = p_2 = p$:

F_1-F_2 # $N[0 ; p(1-p)(1/ n_1 + 1/ n_2)]$

p est estimé par $F = (n_1F_1 + n_2 F_2)/ (n_1+ n_2)$

d'où la statistique Z et sa loi sous Ho.

Exemple de comparaison de fréquences

Dans un premier sondage : 52% de satisfaits

Dans un deuxième sondage : 50% de satisfaits

Peut-on en conclure que le nombre de satisfaits a diminué ?
on ne peut rien conclure pour l'instant

- Sondages indépendants, respectivement sur 50 et 60 individus, risque accepté 5%
- test de $\{p_2 = p_1\}$ contre $\{p_2 < p_1\}$
- refus de H_o au risque 5% si $Z < -1.645$
- observation de $Z = -0.209$
- au risque de 5%, les observations ne mettent pas en évidence une diminution du nombre de satisfaits.

45

Le simple énoncé des fréquences observées ne permet pas de conclure quoi que ce soit.

Les deux échantillons sont indépendants (on aurait pu réinterroger les mêmes personnes systématiquement : cela relève d'un autre test, de comparaison d'échantillons « appariés »)

La taille des deux échantillons est précisée.

On peut alors faire le test de la diminution du nombre de satisfaits (on suppose alors aussi que la taille de la population totale n'a pas varié entre les deux sondages)

On constate que -0.209 n'est pas inférieure à -1.645 :

« la diminution observée n'est pas significative au seuil de 5% »

« au seuil de 5%, les observations n'ont pas mis en évidence une diminution de la proportion de satisfaits dans la population »

En fait, Prob(Z<-0.209) = 0.417 : même au seuil de 40%, on n'aurait pu refuser l 'hypothèse d'égalité des proportions de satisfaits entre les deux populations.

Pour que cette différence de 2% soit significative au seuil de 5%, toutes proportions gardées, il aurait fallu que l 'observation porte sur $(1.645/0.209)^2$ fois plus d 'observations, soit au moins 3100 et 3719 observations.

5.3. Comparaison d'échantillons Normaux

- Deux échantillons indépendants de variables Normales

$X_1,\ldots, X_{n_1} \approx N(m_1, \sigma_1^2)$

$Y_1,\ldots, Y_{n_2} \approx N(m_2, \sigma_2^2)$

- test de $\{m_2 = m_1\}$ contre $\{m_2 > m_1\}$

  [ou $\{m_2 < m_1\}$, ou $\{m_2 \neq m_1\}$]

- Si les variances sont inconnues, le test ne peut se faire que si $\sigma_1^2 = \sigma_2^2$

m_1 est estimé par : $\overline{X}_{n_1} \approx N\left(m_1; \frac{\sigma_1^2}{n_1}\right)$

et m_2 est estimé par : $\overline{Y}_{n_2} \approx N\left(m_2; \frac{\sigma_2^2}{n_2}\right)$

m_2 - m_1 est estimé par : $\overline{Y}_{n_2} - \overline{X}_{n_1} \approx N\left(m_2 - m_1; \frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}\right)$

Sous Ho : $$\frac{\overline{Y}_{n_2} - \overline{X}_{n_1}}{\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}} \approx N(0;1)$$

Si les variances ne sont pas connues, on ne peut éliminer les deux paramètres de variance. On pourra le faire lorsque les deux variances sont égales à σ^2 .

Cas où les variances sont connues

$$\frac{\overline{Y}_{n_2} - \overline{X}_{n_1}}{\sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}} \underset{\text{sous } H_o}{\approx} N(0;1)$$

Décider $\{m_2 > m_1\}$ si $\quad \overline{Y}_{n_2} - \overline{X}_{n_1} > u_\alpha \sqrt{\frac{\sigma_2^2}{n_2} + \frac{\sigma_1^2}{n_1}}$

où u_α est lu dans une table de loi N(0;1), tel que :

$P\{U > u_\alpha\} = \alpha$

Ce test n'est théoriquement possible que si les variances sont connues, mais il est applicable lorsque les effectifs sont grands (>30) en remplaçant les variances inconnues par leurs estimations.

$$\frac{\overline{Y}_{n_2} - \overline{X}_{n_1}}{\sqrt{\frac{S_2^2}{n_2} + \frac{S_1^2}{n_1}}} \xrightarrow[\text{sous } H_o]{\text{loi}} N(0;1)$$

Cas où les variances sont inconnues mais égales

$$U = \frac{\overline{Y}_{n_2} - \overline{X}_{n_1}}{\sqrt{\sigma^2\left(\frac{1}{n_2} + \frac{1}{n_1}\right)}} \underset{\text{sous } H_o}{\approx} N(0;1)$$

$$Z = \frac{\sum_{i=1}^{n_1}\left(X_i - \overline{X}_{n_1}\right)^2 + \sum_{j=1}^{n_2}\left(Y_j - \overline{Y}_{n_2}\right)^2}{\sigma^2} \approx \chi^2_{(n_1+n_2-2)}$$

U et Z sont indépendantes

$$S^2 = \frac{\sum_{i=1}^{n_1}\left(X_i - \overline{X}_{n_1}\right)^2 + \sum_{j=1}^{n_2}\left(Y_j - \overline{Y}_{n_2}\right)^2}{n_1 + n_2 - 2}$$

$$\boxed{\Delta = \frac{\overline{Y}_{n_2} - \overline{X}_{n_1}}{\sqrt{S^2\left(\frac{1}{n_2} + \frac{1}{n_1}\right)}} \underset{\text{sous } H_o}{\approx} T(n_1 + n_2 - 2)}$$

Décider $m_2 > m_1$ si $\Delta > t_\alpha$

où t_α est lu dans la table de STUDENT ($n_1 + n_2$ - 2)

Δ est appelé l'écart réduit.

Si les variances empiriques de chaque échantillon sont S_1^2 et S_2^2, on a :

$$S^2 = [(n_1-1)\, S_1^2 + (n_2-1)\, S_2^2]/(n_1 + n_2 - 2)$$

Les deux chi-deux sont indépendants, leur somme est un chi-deux dont le degré de liberté est la somme des degrés de liberté :

$$(n_1-1) + (n_2-1) = n_1 + n_2 - 2$$

Les hypothèses importantes sont la normalité des échantillons et l 'indépendance des variables X et Y.

Si on met en doute l'hypothèse de normalité des variables X et Y, le test reste asymptotiquement valable (on ne parle plus de loi de STUDENT, mais de loi Normale asymptotique).

Test d'égalité des variances

$\{\sigma_1 = \sigma_2\}$ contre $\{\sigma_1 \neq \sigma_2\}$

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} = \frac{\sum_{i=1}^{n_1}\left(X_i - \overline{X}_{n_1}\right)^2}{\sigma_1^2} \approx \chi^2(n_1-1)$$

$$\frac{(n_2-1)S_2^2}{\sigma_2^2} = \frac{\sum_{j=1}^{n_2}\left(Y_j - \overline{Y}_{n_2}\right)^2}{\sigma_2^2} \approx \chi^2(n_2-1)$$

indépendantes en probabilité

$$\frac{S_1^2}{S_2^2} \underset{\mathrm{si}\,\sigma_1=\sigma_2}{\approx} Fisher(n_1-1; n_2-1)$$

Décider $\{\sigma_1 \neq \sigma_2\}$ si

$$\left\{\frac{S_1^2}{S_2^2} > A\right\} \cup \left\{\frac{S_1^2}{S_2^2} < B\right\}$$

Le test est bilatéral: réunion de deux régions. On devrait calculer A et B tels que la région contienne les points de densité minimum, c'est-à-dire avec densité(A) = densité(B), et la somme des deux probabilités égale au seuil choisi. Il existe des tables donnant A et B pour α donné.

On se contentera d'attribuer $\alpha/2$ à chaque demi région.

Les tables de Fisher sont données en fait pour les valeurs A (>1). On utilise alors pour calculer B le fait que l'inverse d'un Fisher(n_1;n_2) est un Fisher(n_2,n_1)

La règle est alors simple : on calcule le rapport de la plus grande sur la plus petite variance empirique, et on compare à celui des deux chiffres A1 ou A2 qui correspond au rapport effectué.

Exemple

- Deux échantillons indépendants :

$$\bar{x}_{10} = 103.4 \qquad s_1^2 = 31.2$$

$$\bar{y}_{12} = 108.2 \qquad s_2^2 = 42.7$$

- Test d'égalité des variances au seuil de 10%

$F_{5\%}(9;11) = 2.90$ $\qquad$ $F_{5\%}(11;9) = 3.10$

$F_{observé} = 1.37 < 3.10 \Rightarrow \{\sigma_1 = \sigma_2\}$

- Test de $\{m_2 = m_1\}$ contre $\{m_2 \neq m_1\}$, seuil 5%

$T_{5\%}(20) = 2.086$

$s^2 = 37.525 \quad \Delta = 1.83 < 2.086 \Rightarrow \{m_1 = m_2\}$

• Test d'égalité des variances :

$s_2 > s_1$: le F observé est 42,7/31,2 (= 1,37), à comparer à 3,10.

Conclusion du test : au seuil de 10%, les variances empiriques ne sont pas significativement différentes.

• Test d'égalité des espérances mathématiques :

le test effectué ici est le test bilatéral : si nous avions testé $\{m_1 = m_2\}$ contre $\{m_1 < m_2\}$, la valeur limite de Δ aurait été lue dans la table de student à 20 degrés de liberté, mais dans la colonne 10% pour un test unilatéral de seuil 5% : nous décidons que $m_1 < m_2$ si $\Delta > 1{,}725$.

Au seuil de 5%, la différence observée de 4.8 n'est pas significativement différente de zéro.

Utilisation de PROB = $P\{|T(20)| > 1.83\} = 0.082$.

C'est une autre façon de définir la règle de décision. Elle est très utilisée lorsqu'on dispose d'une feuille de calcul ou dans les logiciels usuels. Elle est équivalente à celle exposée.

Règle : on refuse $\{m_1 = m_2\}$ au seuil α choisi si PROB < α .

Ici, 8.2% est supérieur à 5% : on ne peut refuser $\{m_1 = m_2\}$ au seuil de 5%