

Notes de cours
Statistique inférentielle : tests
Master Statistique et Économétrie, Univ. Rennes 1

V. Monbet

Master 1 - 2013

Table des matières

| | | |
|----------|---|-----------|
| 1 | Tests d'hypothèses | 3 |
| 1.1 | Principe et définitions | 3 |
| 1.1.1 | Hypothèses de test | 3 |
| 1.1.2 | Principe général | 4 |
| 1.1.3 | Erreurs et puissance | 4 |
| 1.1.4 | Principe de Neyman | 5 |
| 1.2 | Exemples | 6 |
| 1.2.1 | Test du signe | 6 |
| 1.2.2 | Test pour la moyenne d'une loi de Gauss | 7 |
| 1.3 | Principe de Neyman et optimalité | 9 |
| 1.3.1 | Test randomisé | 9 |
| 1.3.2 | Tests uniformément plus puissants | 11 |
| 1.4 | Tests UPP pour les hypothèses composites | 15 |
| 1.4.1 | Test unilatéral $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_1$ | 15 |
| 1.4.2 | Test bilatéral $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ | 17 |
| 1.5 | Généralisation | 20 |
| 1.6 | Autres tests basés sur le maximum de vraisemblance | 24 |
| 1.6.1 | Test de Wald | 24 |
| 1.6.2 | Score test (ou test des multiplicateurs de Lagrange) | 24 |
| 1.7 | Tests classiques | 24 |
| 1.7.1 | Tests paramétriques pour des moyennes, des variances ou des corrélations | 24 |
| 1.7.2 | Tests non paramétriques pour des moyennes, des variances ou des corrélations | 25 |
| 1.7.3 | Test d'adéquation ou de comparaison de distribution | 26 |
| 1.7.4 | Estimer la fonction de répartition | 27 |
| 2 | Estimation par intervalles | 29 |
| 2.1 | Exemple | 29 |
| 2.2 | Méthode générale pour construire des intervalles de confiance | 30 |
| 2.3 | Lien avec les tests | 30 |

Chapitre 1

Tests d'hypothèses

1.1 Principe et définitions

On se place dans un modèle paramétrique $\mathbf{X} = (X_1, \dots, X_n)$ i.i.d suivant $(P_\theta, \theta \in \Theta)$. Supposons que $\Theta = \Theta_0 \cup \Theta_1$ où Θ_0 et Θ_1 sont deux ensembles disjoints non vides. Connaissant une réalisation (x_1, \dots, x_n) de \mathbf{X} , on voudrait décider si θ est dans Θ_0 ou Θ_1 . En pratique, on choisira toujours pour Θ_0 le plus petit des deux sous-espaces Θ_0, Θ_1 . Ainsi, $\theta \in \Theta_0$ correspond à la version la plus simple du modèle.

1.1.1 Hypothèses de test

Une hypothèse est une assertion sur la distribution d'une (ou de plusieurs) variable(s) aléatoire(s). On pose une *hypothèse nulle* notée

$$H_0 : \theta \in \Theta_0$$

contre une *hypothèse alternative* notée

$$H_1 : \theta \in \Theta_1$$

Exemple - Supposons que X_1, \dots, X_m et Y_1, \dots, Y_n sont des variables indépendantes telles que $X_i \sim \mathcal{N}(\mu_1, \sigma_2)$ et $Y_i \sim \mathcal{N}(\mu_2, \sigma_2)$. L'espace des paramètres est

$$\Theta = \{(\mu_1, \mu_2, \sigma) : -\infty < \mu_1, \mu_2 < \infty, \sigma > 0\}$$

Dans les applications, on cherche à déterminer si les X_i et les Y_i ont la même distribution (c'est à dire si $\mu_1 = \mu_2$). Par exemple, on administre 2 somminifères différents à 2 groupes de patients et on cherche à savoir si la durée moyenne du sommeil est la même dans les deux groupes. On définit alors Θ_0 par

$$\Theta_0 = \{(\mu_1, \mu_2, \sigma) : -\infty < \mu_1 = \mu_2 < \infty, \sigma > 0\}$$

et Θ_1 est le complémentaire de Θ_0 . On remarque ici de Θ et de dimension 3 alors que Θ_0 n'est plus que de dimension 2.

1.1.2 Principe général

On décide que θ est dans Θ_0 ou Θ_1 à l'aide des observations (x_1, \dots, x_n) . Pour cela on cherche une règle de décision qui prend la forme suivante.

- Si $S(x_1, \dots, x_n) \in \text{RC}$ alors on rejette H_0
- Si $S(x_1, \dots, x_n) \notin \text{RC}$ alors on ne rejette pas H_0

où S est une *statistique* (ou fonction) de *test* et RC est une *région critique*. Le plus souvent, S est une fonction d'un estimateur de θ .

Exemple - Supposons que X_1, \dots, X_n sont des variables i.i.d telles que $E(X_i) = \theta$. On souhaite tester

$$H_0 : \theta = 0 \text{ contre } H_1 : \theta \neq 0$$

On peut alors choisir $S(X_1, \dots, X_n) = \bar{X}$. Et on rejette H_0 si \bar{x} est éloigné de 0 c'est à dire si $\text{RC} =]\infty, -c] \cup [c, \infty[$ avec c une constante positive. On verra plus loin comment on détermine c .

1.1.3 Erreurs et puissance

Il est peu probable qu'une règle de décision soit parfaite. Ainsi, quand on définit une règle de décision, on doit regarder qu'elle est la probabilité de faire des erreurs quand on prend l'une ou l'autre décision en fonction de la valeur de $\theta \in \Theta$. On peut faire deux types d'erreur

- On fait une *erreur de première espèce*¹ quand on rejette H_0 à tort c'est à dire alors que $\theta \in \Theta_0$. On peut associer une probabilité à cette erreur :

$$P(S(X_1, \dots, X_n) \in \text{RC} | \theta \in \Theta_0)$$

On parle aussi parfois de *risque de première espèce*.

- On fait une *erreur de seconde espèce*² quand on accepte H_0 à tort c'est à dire alors que $\theta \in \Theta_1$. La probabilité associée est alors

$$P(S(X_1, \dots, X_n) \notin \text{RC} | \theta \in \Theta_1)$$

On pourrait être tenté de chercher à définir des statistiques de tests telles que ces deux erreurs soient uniformément petites pour tout $\theta \in \Theta$. On verra que c'est généralement impossible.

Définition 1 - On appelle *fonction de risque de première espèce* la fonction

$$\begin{aligned} \alpha & : \Theta_0 \rightarrow [0, 1] \\ & \theta \mapsto P_\theta(S(X_1, \dots, X_n) \in \text{RC}) \end{aligned}$$

On appelle *fonction de risque de seconde espèce* la fonction

$$\begin{aligned} \beta & : \Theta_1 \rightarrow [0, 1] \\ & \theta \mapsto P_\theta(S(X_1, \dots, X_n) \notin \text{RC}) \end{aligned}$$

1. en anglais : type I error
2. en anglais : type II error

Ainsi, α représente la probabilité de se tromper quand on est sous H_0 . On note parfois abusivement $\alpha(\theta) = P_\theta(H_1|H_0)$ et $\beta(\theta) = P_\theta(H_0|H_1)$.

Comme H_0 est conservative, une erreur de type I est généralement considérée plus sérieuse qu'une erreur de type II. par exemple, le jury d'une meurtre devrait prendre comme hypothèse nulle "l'accusé est innocent". En effet l'erreur de type I correspond au cas où un innocent est condamné et le vrai meurtrier n'est jamais pris. Alors que l'erreur de type II correspond au cas où le meurtrier est acquitté.

Définition 2 - On appelle fonction puissance la fonction

$$\begin{aligned} \Pi & : \Theta_1 \rightarrow [0, 1] \\ \theta & \mapsto P_\theta(S(X_1, \dots, X_n) \in RC) \end{aligned}$$

Propriété - On a pour tout $\theta \in \Theta_1$, $\Pi(\theta) = 1 - \beta(\theta)$

Définition 3 - $\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta)$ est appelé niveau du test³. C'est le risque de première espèce maximal.

1.1.4 Principe de Neyman

On voudrait trouver des procédures de test qui minimisent les 2 erreurs. Or il est facile de voir, que le plus souvent, si α diminue alors β augmente (voir par exemple le graphique de la puissance du test du signe ci-dessous).

Le principe de Neyman consiste à fixer le niveau α à une valeur petite (typiquement 5% ou 1 %) et à chercher une région critique qui minimise $\beta(\theta)$ à α fixé.

En pratique :

1. On fixe le niveau α .
2. On en déduit une région critique : $RC(\alpha)$. Si plusieurs régions sont possibles, on choisit celle qui minimise $\beta(\theta)$.
3. On conclut : si $S(x_1, \dots, x_n) \in RC(\alpha)$, on rejette H_0 .

On utilise parfois une alternative pour conclure. Au lieu de fixer α et de comparer la valeur de la statistique de test observée à la région critique $RC(\alpha)$, on estime un *degré de significativité* ou (p-value) :

$$\hat{\alpha}(X_1, \dots, X_n) = \inf\{\alpha \text{ tel que } S(X_1, \dots, X_n) \in RC(\alpha)\}$$

Ainsi le degré de significativité est le niveau le plus faible qu'on peut choisir pour conclure au rejet de H_0 . On dit parfois que c'est l'erreur que l'on fait quand on rejette H_0 . Concrètement, on compare $\hat{\alpha}$ au niveau α fixé.

3. signifiant level of the test

1.2 Exemples

1.2.1 Test du signe

Un aquaculteur a 114 poissons d'une certaine espèce dans un de ses bassins. On extrait un échantillon de 12 poissons afin de vérifier l'hypothèse selon laquelle la médiane de la longueur des poissons est de 220 mm. On observe les longueurs suivantes :

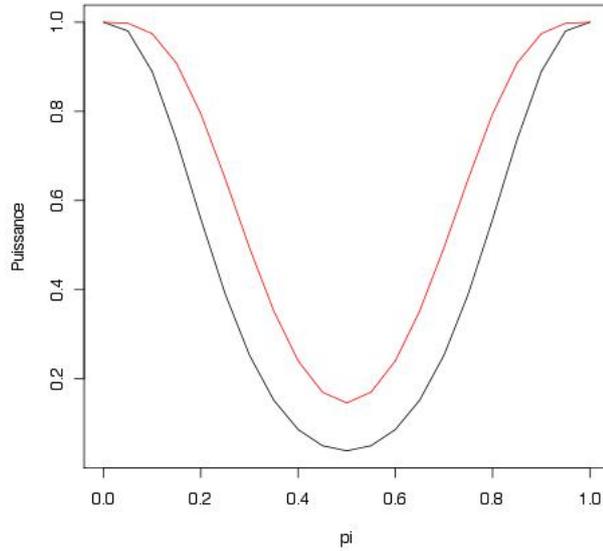
126 142 156 228 245 246 370 419 433 454 478 503

1. Hypothèses de test : en notant μ la médiane, $H_0 : \mu = 220$ contre $H_1 : \mu \neq 220$.
2. Risque de première espèce : on fixe, arbitrairement, $\alpha = 0.05$.
3. Statistique de test. On choisit ici un test naïf appelé *test du signe*. La statistique de test est construite de la façon suivante. Si la médiane est 220, il est également probable pour chaque poisson sélectionné d'être plus ou moins long que 220 mm. Puis on calcule S égale au nombre d'individus plus long que 220. (Implicitement, on associe à chaque individu un signe - si sa longueur est inférieure à 220 et un signe + sinon ; S est alors la somme des signes +).
4. Loi de la statistique de test. Il est facile de voir que la loi de la statistique de test est une loi binomiale de paramètres $n = 12$ et $\pi = 1/2$.
5. Région critique. On remarque aisément que l'on va rejeter H_0 si S est trop grande ou trop petite (dominance de signes +, ou de signes -) ; la région critique est donc de la forme $\{s \text{ tels que } s \notin [s_{\text{inf}}, s_{\text{sup}}]\}$. Les bornes s_{inf} et s_{sup} sont déterminées à l'aide de la loi binomiale de telle sorte que

$$P_{H_0}(S < s_{\text{inf}}) = \frac{\alpha}{2} \text{ et } P_{H_0}(S > s_{\text{sup}}) = \frac{\alpha}{2}$$

En s'aidant de la table 1.1, on trouve que $s_{\text{inf}} = 2$ et $s_{\text{sup}} = 10$ (car $P(x = 0) + P(x = 1) + P(x = 2) < .025$).

6. Puissance du test. On ne peut calculer $P_{H_1}(S \notin [s_{\text{inf}}, s_{\text{sup}}])$ que si on choisit une alternative ponctuelle pour H_1 c'est à dire si on fixe la valeur de la médiane sous H_1 . En pratique, on dispose généralement pas d'une telle information. On peut alors regarder comment varie la puissance pour différentes valeurs de la médiane. Par exemple, le graphique ci-dessous montre la puissance du test du signe quand π varie. La courbe en noir correspond à un risque de première espèce de 5% et la courbe en rouge à un risque de première espèce de 20%.



| | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|
| r | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| P | 0.000 | 0.003 | 0.016 | 0.054 | 0.121 | 0.193 | 0.226 |
| | | 7 | 8 | 9 | 10 | 11 | 12 |
| | | 0.193 | 0.121 | 0.054 | 0.016 | 0.003 | 0.000 |

TABLE 1.1 – Probabilités binomiales $P(S = k)$, $n = 12$, $\pi = \frac{1}{2}$

1.2.2 Test pour la moyenne d'une loi de Gauss

Un contrôle anti-dopage a été effectué sur 16 sportifs. On a mesuré la variable X de moyenne m , qui est le taux (dans le sang) d'une certaine substance interdite. Voici les données obtenues :

0.35 0.4 0.65 0.27 0.14 0.59 0.73 0.13
 0.24 0.48 0.12 0.70 0.21 0.13 0.74 0.18

La variable X est supposée gaussienne et de variance $\sigma^2 = 0.04$. On veut tester, au niveau 5% l'hypothèse selon laquelle le taux moyen dans le sang de la population des sportifs est égal à 0.4.

On pose des hypothèses de test unilatérales :

$$H_0 : m = m_0 = 0.4 \text{ contre } H_1 : m > 0.4$$

La statistique de test est la moyenne empirique. Si on note X_1, \dots, X_n l'échantillon de variables aléatoires de même loi que X , la moyenne empirique est donnée par

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Intuitivement, on comprend bien qu'on va rejeter H_0 si $\bar{X}_n - m_0$ est trop grande en valeur absolue c'est à dire si la moyenne empirique est trop éloignée de la moyenne sous H_0 .

Sous H_0 , $Z = \frac{\bar{X}_n - m_0}{\sigma/\sqrt{n}}$ suit une loi de Gauss de moyenne 0 et de variance 1. D'autre part, d'après la remarque faite plus haut on comprend qu'on rejette H_0 si $|Z| > z_0$. Pour construire la région critique, on cherche donc z_0 tel que

$$P(|Z| > z_0) = \alpha$$

soit encore

$$P(Z > z_0 \text{ ou } Z < -z_0) = P(Z > z_0) + P(Z < -z_0) = \alpha$$

or on a par symétrie de la loi de Gauss de moyenne 0 et de variance 1

$$P(Z > z_0) = P(Z < -z_0) = \Phi(-z_0) = 1 - \Phi(z_0)$$

où on note Φ la fonction de répartition de la loi Gauss de moyenne 0 et de variance 1. Ainsi z_0 est tel que

$$1 - \Phi(z_0) = \alpha/2$$

ce qui s'écrit encore

$$z_0 = \Phi^{-1}(1 - \alpha/2)$$

D'après la table de la fonction de répartition inverse de la loi normale, on en déduit que $z_0 = 1.96$ car $\alpha = 0.05$.

Finalement, on rejette donc H_0 si

$$|\bar{X}_n - m_0| > 1.96 \frac{\sigma}{\sqrt{n}}$$

.

Remarques

- Lorsque le nombre d'observations n est grand (supérieur à 30), d'après le théorème de limite centrale on a que la statistique de test

$$Z = \frac{\bar{X} - m_0}{\sigma/\sqrt{n}}$$

suit approximativement une loi de Gauss quelque soit la loi de la variable X considérée.

Si la variance est inconnue

Dans le cas où la variance n'est pas connue, on doit l'estimer en utilisant les observations. Et la statistique de test du test de la moyenne donnée par

$$Z = \frac{\bar{X} - m_0}{s/\sqrt{n}}$$

Elle ne suit plus une loi de Gauss car le dénominateur n'est plus une constante mais une réalisation de l'estimateur de la variance de la variable X . L'écart-type s est obtenu par

$$s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$

Par construction, S^2 suit une loi du χ^2 . Y est donc une v.a. suivant une de Student à $(n-1)$ degrés de libertés. Et on utilise alors une table de la loi de Student pour conclure le test.

Remarque : Lorsque le nombre d'observations n est grand (supérieur à 30), on peut utiliser le théorème de limite centrale pour approcher la loi de la statistique Z .

Calcul de la puissance du test

Dans le cas d'un test de Student, on peut calculer la puissance du test si on on peut donner une valeur de la moyenne sous l'hypothèse alternative.

$$H_0 : m = m_0 \text{ contre } H_1 : m = m_1$$

La puissance est définie par

$$\mathcal{P} = P(\text{rejeter } H_0 | H_0 \text{ est fausse})$$

Ainsi la puissance est la probabilité de la la région de rejet de H_0 sous la loi de H_1 .

$$\begin{aligned} \mathcal{P} &= P\left(Z > z_0 \mid \frac{Z - m_1}{\sigma/\sqrt{n}} \text{ suit une loi } \mathcal{N}(0, 1)\right) \\ &= P\left(\tilde{Z} > \frac{z_0 - m_1}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(\frac{z_0 - m_1}{\sigma/\sqrt{n}}\right) \end{aligned}$$

1.3 Principe de Neyman et optimalité

1.3.1 Test randomisé

Affinons le test du signe réalisé plus haut. En effet, nous avons remarqué qu'on ne peut pas toujours atteindre exactement le niveau α fixé.

Dans l'exemple des poissons, si on pose des hypothèses de test unilatérales, correspondants à la question : "les poissons pêchés sont-ils trop petits (de longueur inférieure à 22 mm) ?",

$$H_0 : \mu = 22 \text{ contre } H_1 : \mu < 22$$

alors la région critique est de la forme $RC = \{S \leq c\}$ où c doit vérifier $P_{H_0}(S \leq c) = \alpha$. Or sous H_0 , la statistique de test suit une loi binomiale (définie sur un ensemble discret) et il n'existe pas de c qui permette d'obtenir l'égalité : $P(S \leq 2) = 0.019$ et $P(S \leq 3) = 0.073$.

Une solution consiste à randomiser le test, c'est à dire qu'on tire au hasard la solution du test avec une certaine probabilité. Dans l'exemple des poissons,

- si $s_{obs} \geq 4$ on ne refuse pas H_0 ;
- si $s_{obs} \leq 2$ on refuse H_0 ;
- si $s_{obs} = 3$, on tire au sort H_1 avec une probabilité $\gamma \in [0, 1]$. On choisit γ telle que

$$\begin{aligned}\alpha = P_{H_0}(\text{RC}) &= 0 \times P_{H_0}(S \geq 4) + \gamma \times P_{H_0}(S = 3) + 1 \times P_{H_0}(S \leq 2) \\ &= \gamma P_{H_0}(S = 3) + P_{H_0}(S \leq 2)\end{aligned}$$

d'où $\gamma = (0.05 - 0.019)/0.054 = 0.57$.

Donc, si $s_{obs} = 3$ on décide H_1 avec une probabilité de 57%.

Définition 4 Soit (X_1, \dots, X_n) à valeur dans E^n . Un test est une fonction aléatoire Ψ de $E^n \rightarrow [0, 1]$.

Interprétation - La fonction Ψ représente la probabilité de décider H_1 .

Si $\Psi(X_1, \dots, X_n) = 0$ on conclut à H_0 .

Si $\Psi(X_1, \dots, X_n) = 1$ on conclut à H_1 .

Si $\Psi(X_1, \dots, X_n) \in]0, 1[$ on tire au hasard la décision H_1 avec la probabilité $\Psi(X_1, \dots, X_n)$.

Lorsque Ψ est à valeurs dans $\{0, 1\}$, on parle de test pur, c'est à dire non randomisé. C'est le cas de la plupart des tests classiques. Par exemple quand on teste, pour une variable aléatoire de loi de Gauss de moyenne μ ,

$$H_0 : \mu = \mu_0 \text{ contre } H_1 : \mu \neq \mu_0$$

La région critique est

$$\text{RC} = \left\{ \sqrt{n} \left| \frac{\bar{x}_n - \mu_0}{s_n} \right| > F_{n-1}^{-1}(1 - \alpha/2) \right\}$$

au niveau α avec F_{n-1} la fonction de répartition de la loi de Student à $n - 1$ degrés de liberté. C'est un test pur pour lequel $\Psi(x_1, \dots, x_n) = \mathbb{I}_{\{\text{RC}\}}$.

Définition 5 Pour le test $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$,

- le risque de 1ère espèce est la fonction $\alpha(\theta) = E_\theta(\Psi(X_1, \dots, X_n))$, $\forall \theta \in \Theta_0$;
- le risque de 2nde espèce est la fonction $\beta(\theta) = E_\theta(1 - \Psi(X_1, \dots, X_n))$, $\forall \theta \in \Theta_1$;
- le niveau est $\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta)$;
- la puissance du test est la fonction $\Pi(\theta) = 1 - \beta(\theta)$.

L'utilisation de tests randomisés permet de considérer des tests de niveau α pour tout $\alpha \in [0, 1]$. Ils existent d'après le lemme de Neyman-Pearson donné ci-dessous. On peut donc définir la notion de test le plus puissant parmi les tests de niveau α .

Définition 6 Un test associé à la fonction Ψ est un test uniformément plus puissant (UPP) au niveau α , si son niveau est inférieur ou égal à α et si pour tout test Ψ^* de niveau inférieur ou égal à α ,

$$\Pi_\theta(\Psi(X_1, \dots, X_n)) \geq \Pi_\theta(\Psi^*(X_1, \dots, X_n))$$

pour tout $\theta \in \Theta_1$.

1.3.2 Tests uniformément plus puissants

Le lemme de Neyman-Pearson est important car il suggère un principe pour trouver de "bons" tests au sens du compromis entre une puissance forte et une erreur de première espèce faible.

Théorème 1 - Lemme de Neyman-Pearson. Soit (X_1, \dots, X_n) un échantillon de vraisemblance $\mathcal{L}(\theta; X_1, \dots, X_n)$. Pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$, $\theta_0 \neq \theta_1$, pour tout $\alpha \in]0, 1[$ fixé, il existe $c > 0$ et $\gamma \in [0, 1[$ tels que le test

$$\Psi(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \\ \gamma & \text{si } \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} = c \\ 0 & \text{si } \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} < c \end{cases}$$

De plus ce test est uniformément plus puissant parmi les tests de niveau au plus α et c'est le seul. Ce test est appelé test de Neyman-Pearson associé à c et γ est déterminé par l'équation de test $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$. (γ n'est pas forcément unique.)

Preuve - 1. Nous montrons tout d'abord que le niveau est bien α .

$$\begin{aligned} E_{\theta_0}(\Psi(X_1, \dots, X_n)) &= 1 \times P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \right) \\ &+ \gamma \times P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} = c \right) \\ &+ 0 \times P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} < c \right) \end{aligned}$$

On veut trouver c et γ tels que $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$ pour tout $\alpha \in]0, 1[$. Soit F la fonction de répartition de $\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)}$

$$F(t) = P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} \leq t \right).$$

- Si F est continue, alors il existe c tel que $F(c) = 1 - \alpha$. En prenant ce c et $\gamma = 0$, on a

$$\begin{aligned} P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} > c \right) + \gamma P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = c \right) \\ = 1 - F(c) = 1 - (1 - \alpha) = \alpha \end{aligned}$$

- Si F n'est pas continue (cas discret), on note $c^+ = \min_t (F(t) \geq 1 - \alpha)$ et on a

$$F(c^+) = \lim_{x \rightarrow c^+, x > c^+} F(x), F(c^-) = \lim_{x \rightarrow c^+, x < c^+} F(x)$$

ainsi

$$F(c^+) - F(c^-) = \lim_{\substack{x \rightarrow c^+, x > c^+ \\ y \rightarrow c^+, y < c^+}} P \left(y < \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} \leq x \right) = P \left(\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = c^+ \right)$$

Or on cherche c et γ tels que $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$. En choisissant $c = c^+$, on obtient

$$1 - F(c^+) + \gamma(F(c^+) - F(c^-)) = \alpha$$

ce qui est équivalent à

$$\gamma = \frac{\alpha + F(c^+) - 1}{F(c^+) - F(c^-)}$$

Il reste à vérifier que γ appartient à $[0, 1[$.

$$\gamma \geq 0 \Leftrightarrow \alpha - 1 + F(c^+) \geq 0 \Leftrightarrow F(c^+) \geq 1 - \alpha$$

ce qui est vrai par définition de c^+ . Par ailleurs,

$$\gamma < 1 \Leftrightarrow \alpha - 1 + F(c^+) < F(c^+) - F(c^-) \Leftrightarrow F(c^-) < 1 - \alpha$$

ce qui est vrai par définition.

2. Montrons maintenant que le test est UPP.

Soit Ψ^* un test de niveau au plus α . On considère l'intégrale

$$\int_{E^n} (\Psi(x_1, \dots, x_n) - \Psi^*(x_1, \dots, x_n)(\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0))) d\mu(x_1, \dots, x_n) \quad (1.1)$$

où μ est la mesure de référence par rapport à laquelle $\mathcal{L}(\theta_0)$ et $\mathcal{L}(\theta_1)$ sont définies. Cette intégrale est toujours positive car

- si $\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0) > 0$ alors $\Psi(X_1, \dots, X_n) = 1$ par définition de Ψ et, $\Psi(X_1, \dots, X_n) \geq \Psi^*(X_1, \dots, X_n)$;
- si $\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0) < 0$ alors $\Psi(X_1, \dots, X_n) = 0$ par définition de Ψ et, $\Psi(X_1, \dots, X_n) \leq \Psi^*(X_1, \dots, X_n)$;
- si $\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0) = 0$ alors l'intégrale est nulle.

On a donc

$$\begin{aligned} & \int_{E^n} \Psi(x_1, \dots, x_n) \mathcal{L}(\theta_1) d\mu(x_1, \dots, x_n) - c \int_{E^n} \Psi(x_1, \dots, x_n) \mathcal{L}(\theta_0) d\mu(x_1, \dots, x_n) \\ & - \int_{E^n} \Psi^*(x_1, \dots, x_n) \mathcal{L}(\theta_1) d\mu(x_1, \dots, x_n) + c \int_{E^n} \Psi^*(x_1, \dots, x_n) \mathcal{L}(\theta_0) d\mu(x_1, \dots, x_n) \geq 0 \end{aligned}$$

Ce qui s'écrit aussi

$$E_{\theta_1}(\Psi) - cE_{\theta_0}(\Psi) - E_{\theta_1}(\Psi^*) + cE_{\theta_0}(\Psi^*) \geq 0$$

ou encore

$$E_{\theta_1}(\Psi) - E_{\theta_1}(\Psi^*) \geq c(E_{\theta_0}(\Psi) - E_{\theta_0}(\Psi^*))$$

et on reconnaît que

- $E_{\theta_1}(\Psi)$: puissance de Ψ
- $E_{\theta_1}(\Psi^*)$: puissance de Ψ^*
- $E_{\theta_0}(\Psi)$: niveau de Ψ qui est égal à α
- $E_{\theta_0}(\Psi^*)$: niveau de Ψ^* qui est inférieur ou égal à α

On en déduit donc que $E_{\theta_1}(\Psi) - E_{\theta_1}(\Psi^*) \geq 0$ et que la puissance Ψ est supérieure à celle de tout test Ψ^* de niveau au plus α .

3. On montre enfin que c'est le seul test UPP, à γ près.

Soit Ψ^* un test UPP au niveau au plus α . On a donc

$E_{\theta_1}(\Psi^*(X_1, \dots, X_n)) \geq E_{\theta_1}(\Psi(X_1, \dots, X_n))$ car Ψ^* est UPP,
et $E_{\theta_1}(\Psi(X_1, \dots, X_n)) \geq E_{\theta_1}(\Psi^*(X_1, \dots, X_n))$ car Ψ est UPP.

Donc, ces deux tests sont de même puissance

$$E_{\theta_1}(\Psi(X_1, \dots, X_n)) = E_{\theta_1}(\Psi^*(X_1, \dots, X_n))$$

Reprenons l'intégrale (1.1) :

$$\begin{aligned} \int_{E^n} (\Psi(x_1, \dots, x_n) - \Psi^*(x_1, \dots, x_n)(\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0))) d\mu(x_1, \dots, x_n) \\ = E_{\theta_1}(\Psi) - cE_{\theta_0}(\Psi) - E_{\theta_1}(\Psi^*) + cE_{\theta_0}(\Psi^*) \end{aligned}$$

On a noté que cette intégrale est positive donc $E_{\theta_0}(\Psi) - E_{\theta_0}(\Psi^*) \leq 0$. Or on a vu que $E_{\theta_0}(\Psi) = \alpha$ et $E_{\theta_0}(\Psi^*) \leq \alpha$. Donc $E_{\theta_0}(\Psi) = E_{\theta_0}(\Psi^*)$ et l'intégrale est nulle. Comme $\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0)$ est différent de 0, cela implique $Psi = Psi^*$, μ presque sûrement. Donc les tests coïncident (à γ près qui reste à déterminer). \diamond

Remarque - Dans le cas continu (μ est la mesure de Lebesgue), on retrouve un test pur de région critique

$$\text{RC} = \left\{ \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \right\}$$

En effet, dans ce cas,

$$E_{\theta_0}(\Psi(X)) = 1 \times P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X)}{\mathcal{L}(\theta_0; X)} > c \right) + \gamma \times P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X)}{\mathcal{L}(\theta_0; X)} = c \right)$$

or

$$P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X)}{\mathcal{L}(\theta_0; X)} = c \right) = 0$$

car la loi est continue.

Pour résoudre $E_{\theta_0}(\Psi(X)) = \alpha$, on peut choisir γ quelconque : on prend $\gamma = 0$ et le test devient

$$\Psi(X_1, \dots, X_n) = \mathbb{I}_{\left\{ \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \right\}}$$

autrement dit, un test pur de région critique

$$\text{RC} = \left\{ \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \right\}$$

Pour déterminer c , on résoud $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$ ce qui est équivalent à résoudre

$$P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \right) = \alpha$$

Exemple 1 - Soient X_1, \dots, X_n des v.a.i.i.d de loi de Gauss de moyenne θ et de variance un. On teste

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

à partir de (x_1, \dots, x_n) une réalisation de (X_1, \dots, X_n) . La vraisemblance s'écrit

$$\mathcal{L}(\theta; x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2\right)$$

Ainsi, le rapport des vraisemblances est

$$\frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} = \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \theta_1)^2 + \frac{1}{2} \sum_{i=1}^n (x_i - \theta_0)^2\right)$$

en passant au log, on obtient

$$\begin{aligned} \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c &\Leftrightarrow -\frac{1}{2} \sum_{i=1}^n (x_i - \theta_1)^2 + \frac{1}{2} \sum_{i=1}^n (x_i - \theta_0)^2 > \log(c) \\ &\Leftrightarrow (\theta_1 - \theta_0) \sum_{i=1}^n x_i - \frac{n\theta_1^2}{2} + \frac{n\theta_0^2}{2} > \log(c) \\ &\Leftrightarrow (\theta_1 - \theta_0) \sum_{i=1}^n x_i > \log(c) + \frac{n\theta_1^2}{2} - \frac{n\theta_0^2}{2} \end{aligned}$$

Si $\theta_1 > \theta_0$, l'inégalité devient

$$\sum_{i=1}^n x_i > \frac{1}{\theta_1 - \theta_0} \left(\log(c) + \frac{n\theta_1^2}{2} - \frac{n\theta_0^2}{2} \right)$$

et si au contraire, $\theta_1 < \theta_0$, elle s'écrit

$$\sum_{i=1}^n x_i < \frac{1}{\theta_1 - \theta_0} \left(\log(c) + \frac{n\theta_1^2}{2} - \frac{n\theta_0^2}{2} \right)$$

Notons

$$c' = \frac{1}{\theta_1 - \theta_0} \left(\log(c) + \frac{n\theta_1^2}{2} - \frac{n\theta_0^2}{2} \right)$$

Le test de Neyman-Pearson se ramène, dans le cas $\theta_1 > \theta_0$ à

$$\Psi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } \sum_{i=1}^n x_i > c' \\ \gamma & \text{si } \sum_{i=1}^n x_i = c' \\ 0 & \text{si } \sum_{i=1}^n x_i < c' \end{cases}$$

où c' et γ sont déterminés par $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$ ce qui est équivalent à

$$1 \times P_{\theta_0}\left(\sum_{i=1}^n X_i > c'\right) + \gamma \times P_{\theta_0}\left(\sum_{i=1}^n X_i = c'\right) = \alpha$$

Ici $P_{\theta_0}(\sum_{i=1}^n X_i = c') = 0$ car la loi est continue. On choisit $\gamma = 0$ et c' est obtenu en résolvant l'équation

$$P_{\theta_0}\left(\sum_{i=1}^n X_i > c'\right) = \alpha$$

Or, sous H_0 , $\sum_{i=1}^n X_i$ suit une loi de Gauss de moyenne $n\theta_0$ et de variance n . On peut donc écrire,

$$P_{\theta_0}\left(\frac{\sum_{i=1}^n X_i - n\theta_0}{\sqrt{(n)}} > \frac{c' - n\theta_0}{\sqrt{(n)}}\right) = \alpha$$

et on en déduit que

$$\frac{c' - n\theta_0}{\sqrt{(n)}} = \Phi^{-1}(1 - \alpha)$$

avec Φ la fonction de répartition de la loi de Gauss centrée et réduite. Autrement dit $\frac{c' - n\theta_0}{\sqrt{(n)}} = \Phi^{-1}(1 - \alpha)$ est le quantile d'ordre $1 - \alpha$ de la loi de Gauss de moyenne 0 et de variance 1 ; on note parfois $q(1 - \alpha)$. Finalement, on conclut que pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ le test de Neyman-Pearson est un test pur ($\gamma = 0$) de région critique :

$$\text{RC} = \left\{ \sum_{i=1}^n x_i > \sqrt{n}\Phi^{-1}(1 - \alpha) + n\theta_0 \right\}$$

Faut-il ajouter l'exemple sur le paramètre d'une loi binomiale ?

1.4 Tests UPP pour les hypothèses composites

Dans la partie précédente, nous avons montré des résultats (constructifs) d'existence et d'unicité pour des tests dont les hypothèses sont des singletons.

1.4.1 Test unilatéral $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_1$

Pour le test unilatéral $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$, on peut construire un test UPP mais en se restreignant à certaines familles de lois.

Définition 7 La famille $(P_\theta)_{\theta \in \Theta}$ est à rapport de vraisemblance monotone s'il existe une statistique $U(X_1, \dots, X_n)$ telle que pour tout $\theta < \theta_1$, le rapport

$$\frac{\mathcal{L}(\theta_1; x_1, \dots, x_n)}{\mathcal{L}(\theta_0; x_1, \dots, x_n)} = h(U(x_1, \dots, x_n))$$

avec une fonction h une fonction strictement monotone.

Remarque - On peut toujours supposer que h est strictement croissante quitte à considérer $-U(x_1, \dots, x_n)$ à la place de $U(x_1, \dots, x_n)$.

Exemple - la famille des lois exponentielles.
 Considérons le cas particulier

$$\mathcal{L}(x_1, \dots, x_n) = K(\theta) \exp(c(\theta)T(x_1, \dots, x_n))$$

ainsi pour tout $\theta_1 > \theta_0$,

$$\frac{\mathcal{L}(\theta_1; x_1, \dots, x_n)}{\mathcal{L}(\theta_0; x_1, \dots, x_n)} = \frac{K(\theta_1)}{K(\theta_0)} \exp((\theta_1 - \theta_0)T(x_1, \dots, x_n))$$

est une fonction croissante de $T(x_1, \dots, x_n)$. C'est donc une famille à rapport de vraisemblance monotone avec $U(x_1, \dots, x_n) = T(x_1, \dots, x_n)$.

Théorème 2 Soient X_1, \dots, X_n v.a.i.i.d. suivant $(P_\theta)_{\theta \in \Theta}$ où $(P_\theta)_{\theta \in \Theta}$ est une famille à rapport de vraisemblance monotone. Pour tester

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_1$$

il existe un test UPP de niveau α de la forme

$$\Psi(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } U(X_1, \dots, X_n) > c \\ \gamma & \text{si } U(X_1, \dots, X_n) = c \\ 0 & \text{si } U(X_1, \dots, X_n) < c \end{cases}$$

où U est la statistique de la définition 7 et γ et c sont définies par $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$.

Preuve - a. On vérifie d'abord que le test est de niveau α .

Soient les hypothèses $H_0 : \theta = \theta'$ contre $H_1 : \theta = \theta''$ avec $\theta' < \theta''$. On considère le test du théorème. Ce test est exactement le test de Neyman-Pearson pour tester $H_0 : \theta = \theta'$ contre $H_1 : \theta = \theta''$. Son niveau vaut $E_{\theta_0}(\Psi(X_1, \dots, X_n))$ que l'on note α' . On sait d'après le lemme de Neyman-Pearson que Ψ est UPP parmi tous les tests de niveau α' .

Soit maintenant le test $\psi(X_1, \dots, X_n) = \alpha'$ pour tout (X_1, \dots, X_n) . Pour tester $H_0 : \theta = \theta'$ contre $H_1 : \theta = \theta''$, $E_{\theta}(\psi(X_1, \dots, X_n)) = \alpha'$ donc ψ est de niveau α' et Ψ est plus puissant que ψ . Donc

$$E_{\theta''}(\Psi(X_1, \dots, X_n)) \geq E_{\theta''}(\psi(X_1, \dots, X_n)) = \alpha'$$

Ainsi pour tout $\theta' < \theta''$, $\alpha' = E_{\theta'}(\Psi(X_1, \dots, X_n)) \leq E_{\theta''}(\psi(X_1, \dots, X_n))$.

En particulier pour $\theta'' = \theta_0$, pour tout $\theta' < \theta_0$, on a

$$E_{\theta'}(\Psi(X_1, \dots, X_n)) \leq E_{\theta_0}(\Psi(X_1, \dots, X_n))$$

donc $E_{\theta'}(\Psi(X_1, \dots, X_n)) = \alpha$ et le niveau de Ψ est bien α .

b. On montre que Ψ est UPP.

Le test de Neyman-Pearson pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ où $\theta_1 > \theta_0$ est exactement *Psi*, sa forme ne dépend pas de θ_1 . Ce test est le plus puissant pour tester $\theta = \theta_0$ contre $\theta = \theta_1$ d'après le lemme de Neyman-Pearson. Si Ψ n'était pas UPP pour tester $\theta \leq \theta_0$ contre $\theta > \theta_0$ alors il existerait un test $\Psi^{(\alpha)}$ plus puissant que Ψ au moins pour un $\theta_1 > \theta_0$ c'est à dire $E_{\theta_1}(\Psi^{(\alpha)}(X_1, \dots, X_n)) \geq E_{\theta_1}(\Psi(X_1, \dots, X_n))$. Ce qui est impossible puisque Ψ est UPP pour tester $\theta = \theta_0$ contre $\theta = \theta_1$. \diamond

Exemple - Loi de Gauss de moyenne inconnue.

Soient X_1, \dots, X_n v.a.i.i.d. suivant une loi de Gauss de moyenne θ et de variance 1. On veut tester

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_0$$

La vraisemblance est

$$\mathcal{L}(\theta; X_1, \dots, X_n) = \frac{1}{(\sqrt{1\pi})^n} \exp\left(-\frac{1}{2} \sum_{i=1}^n (X_i - \theta)^2\right)$$

Il faut montrer que le modèle est à rapport de vraisemblance monotone.

On peut soit montrer que la loi appartient à la famille exponentielle, soit le montrer directement. Considérons $\theta' < \theta''$,

$$\begin{aligned} \frac{\mathcal{L}(\theta''; x_1, \dots, x_n)}{\mathcal{L}(\theta'; x_1, \dots, x_n)} &= \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \theta'')^2 + \frac{1}{2} \sum_{i=1}^n (x_i - \theta')^2\right) \\ &= \exp\left((\theta'' - \theta') \sum_{i=1}^n x_i + \frac{n}{2} ((\theta')^2 - (\theta'')^2)\right) \end{aligned}$$

C'est une fonction monotone de $U(x_1, \dots, x_n) = \sum_{i=1}^n x_i$ car $\theta'' > \theta'$. On peut appliquer le théorème précédent pour conclure que le test UPP au niveau α pour tester $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$ est

$$\Psi(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } \sum_{i=1}^n X_i > c \\ \gamma & \text{si } \sum_{i=1}^n X_i = c \\ 0 & \text{si } \sum_{i=1}^n X_i < c \end{cases}$$

où γ et c sont définies par $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$. Comme la loi est absolument continue par rapport à la mesure de Lebesgue, on peut choisir $\gamma = 0$ et le test se ramène à un test pour la région critique

$$\text{RC} = \left\{ \sum_{i=1}^n x_i > c \right\}$$

où c est déterminée par

$$E_{\theta_0}(\Psi(X_1, \dots, X_n)) = P_{\theta_0} \left(\sum_{i=1}^n X_i > c \right) = \alpha$$

Pour $\theta = \theta_0$, $\sum_{i=1}^n X_i \sim \mathcal{N}(n\theta_0, n)$ donc $c = n\theta_0 + \sqrt{n}\Phi^{-1}(1 - \alpha)$.

1.4.2 Test bilatéral $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$

Il n'existe pas en général de test UPP pour le test bilatéral $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$. En effet, pour être UPP, un test doit être le plus puissant pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ pour tout $\theta_1 \neq \theta_0$. Cependant, selon le lemme de Neyman-Pearson, la forme des tests les plus puissants diffère selon que $\theta_1 > \theta_0$ ou $\theta_1 < \theta_0$.

Soient X_1, \dots, X_n des v.a.i.i.d. de loi binomiale de paramètres n et θ et supposons que l'on veuille tester

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta \neq \theta_0$$

à un certain niveau α . Considérons tout d'abord, les hypothèses

$$H'_0 : \theta = \theta_0 \text{ contre } H'_1 : \theta = \theta_1$$

avec $\theta_1 \neq \theta_0$. Le lemme de Neyman-Pearson indique que le test UPP de H'_0 contre H'_1 est basé sur la statistique de test :

$$T = \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} = \left(\frac{1 - \theta_0}{1 - \theta_1} \right)^n \left(\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)} \right)^{\sum_{i=1}^n X_i}$$

Si $\theta_1 > \theta_0$, il est facile de vérifier que T est une fonction croissante de $\sum_{i=1}^n X_i$: donc un test plus puissant de H'_0 contre H'_1 rejettera H'_0 pour de grandes valeurs de $\sum_{i=1}^n X_i$. Mais, si $\theta_1 < \theta_0$, T est une fonction décroissante de $\sum_{i=1}^n X_i$ et un test plus puissant va rejeter H'_0 pour de petites valeurs de $\sum_{i=1}^n X_i$. On comprend donc qu'on ne pourra pas trouver de test UPP.

Une solution consiste à se restreindre à la classe des tests sans biais.

Définition 8 - *Un test est dit sans biais si sa puissance est toujours supérieure à son niveau, autrement dit si pour tester $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$, pour tout $\theta_1 \in \Theta_1$,*

$$E_{\theta_1}(\Psi(X_1, \dots, X_n)) \geq \sup_{\theta \in \Theta_0} E_{\theta}(\Psi(X_1, \dots, X_n))$$

c'est à dire qu'on a souvent raison quand on conclut H_1 .

Proposition 1 *Un test UPP est forcément sans biais.*

Preuve - Soit Ψ un test UPP de niveau α pour tester $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$. Soit $\psi(x_1, \dots, x_n) = \alpha$ pour tout (x_1, \dots, x_n) . ψ est de niveau α car

$$\sup_{\theta \in \Theta_0} E_{\theta}(\psi(X_1, \dots, X_n)) = \sup_{\theta \in \Theta_0} \alpha = \alpha$$

Comme Ψ est UPP parmi les tests de niveau α , pour tout $\theta_1 \in \Theta_1$,

$$E_{\theta_1}(\Psi(X_1, \dots, X_n)) = E_{\theta_1}(\psi(X_1, \dots, X_n)) = \alpha$$

Donc la puissance de Ψ est toujours supérieure à son niveau et donc Ψ est sans biais. \diamond

Il est parfois possible de construire des tests uniformément plus puissants parmi les tests sans biais. Supposons qu'on veuille tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ au niveau α , on peut construire un test UPP parmi les tests sans biais⁴ en combinant des tests UPP pour

4. UMPU : unbiased most powerfull test

tester $H'_0 : \theta \leq \theta_0$ contre $H''_0 : \theta \geq \theta_0$. Plus précisément, supposons que $\Phi_1(X_1, \dots, X_n)$ est un test UPP de niveau α_1 et $\Phi_2(X_1, \dots, X_n)$ est un test UPP de niveau α_2 tels que $\alpha_1 + \alpha_2 = \alpha$. Alors $\Phi = \Phi_1 + \Phi_2$ sera un test de niveau α si

$$\Phi_1(X_1, \dots, X_n) + \Phi_2(X_1, \dots, X_n) \leq 1$$

Ainsi, en choisissant bien α_1 et α_2 il est possible d'avoir Φ un test UMPU. Le choix naturel est $\alpha_1 = \alpha_2 = \alpha/2$ mais en général, ça ne conduit pas à un test sans biais.

Soit X une variable aléatoire continue de densité

$$f(x; \theta) = \theta x^{\theta-1} \text{ pour } 0 \leq x \leq 1$$

et supposons qu'on veuille tester

$$H_0 : \theta = 1 \text{ contre } H_1 : \theta_1 \neq 1$$

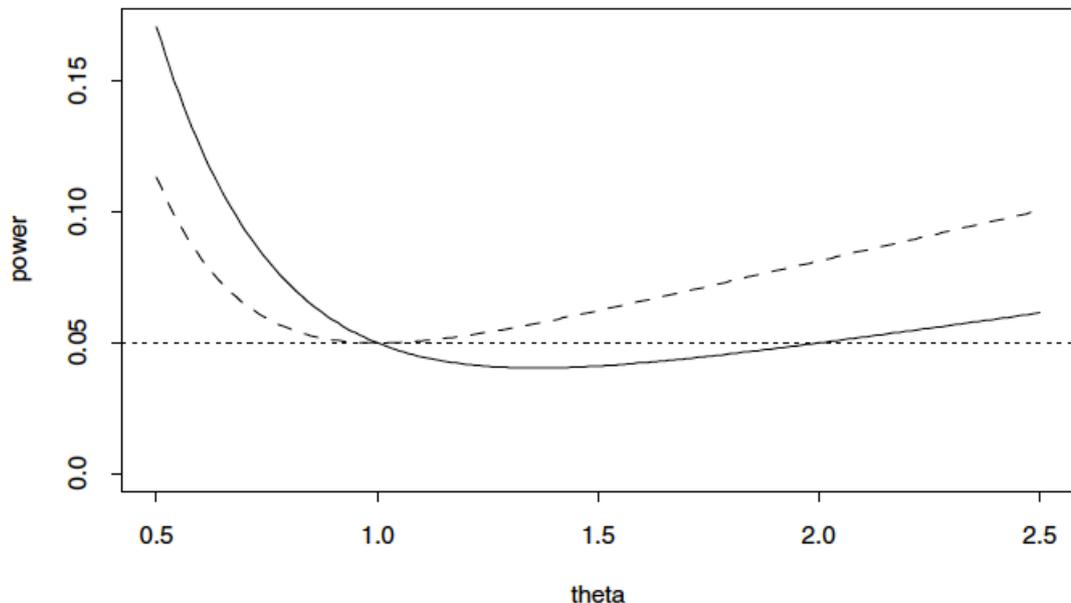
au niveau 5%. On va rejeter H_0 if $x \leq 0.025$ ou si $x \geq 0.975$; ce test est clairement de niveau $\alpha = 5\%$ puisque $P_{\theta=1}(X \leq 0.025) = P_{\theta=1}(X \geq 0.975) = 0.025$.

La fonction puissance est alors

$$\Pi(\theta) = \int_0^{0.025} \theta x^{\theta-1} dx + \int_{0.975}^1 \theta x^{\theta-1} dx = 1 + 0.025^\theta + 0.975^\theta$$

Si on évalue $\Pi(\theta)$ pour θ proche de 1 il est facile de voir que le test n'est pas sans biais. En effet $\Pi(\theta) < 0.05$ pour $1 < \theta < 2$.

Cependant, il est possible de trouver un test sans biais pour tester H_0 contre H_1 . Ce test rejette H_0 si $x \leq 0.0085$ ou si $x \geq 0.9585$. Les deux fonctions puissance sont représentées ci-dessous avec la fonction puissance du test sans biais en tirets :



Dans cet exemple, le test sans biais a une puissance plus grande pour $\theta > 1$ mais plus petite pour $\theta < 1$. Ceci illustre le fait qu'en choisissant un test sans biais, on sacrifie de la puissance dans certaines régions.

1.5 Généralisation

Jusqu'à présent, nous avons proposé des méthodes constructives permettant d'obtenir des tests UPP (ou localement UPP) dans le cas d'un paramètre unique. Ce type de test optimal n'existe pas en général pour les modèles à plus d'un paramètre. Il existe une méthode qui permet de développer des tests dans des cas plus généraux.

Considérons le test

$$H_0 : \theta \in \Theta_0 \text{ contre } \theta \in \Theta_1$$

avec $\Theta_0 \cap \Theta_1 = \emptyset$. On va utiliser le même type d'idée que dans le lemme de Neyman-Pearson.

Définition 9 La statistique Λ du rapport de vraisemblance est

$$\Lambda = \frac{\sup_{\theta \in \Theta} \mathcal{L}(X_1, \dots, X_n; \theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}(X_1, \dots, X_n; \theta)}.$$

Un test du rapport de vraisemblance pour $H_0 : \theta \in \Theta_0$ contre $\theta \in \Theta_1$ va rejeter H_0 pour de grandes valeurs de Λ .

Pour utiliser ces tests du rapport de vraisemblance, on a besoin de connaître (exactement ou approximativement) la distribution de la statistique Λ sous H_0 . Dans certains cas, Λ est fonction d'une autre statistique T dont on connaît la loi et on peut alors utiliser cette statistique. Sinon on utilise un résultat limite.

Considérons le test bilatéral

$$H_0 : \theta = \theta_0 \text{ contre } \theta \neq \theta_1$$

et X_1, \dots, X_n v.a.i.i.d. de densité (ou fonction de fréquence) f dans (P_θ) . La statistique Λ s'écrit

$$\Lambda_n = \prod_{i=1}^n \frac{f(X_i; \hat{\theta}_n)}{f(X_i; \theta_0)}$$

avec $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance.

Théorème 3 Soient X_1, \dots, X_n v.a.i.i.d. qui admettent une densité (ou une fonction de fréquence) vérifiant les hypothèses (A1) à (A5) (du chapitre précédent) avec $I(\theta) = J(\theta)$. Si l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ satisfait un théorème de limite centrale

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow Z \sim \mathcal{N}(0, 1/I(\theta))$$

alors la statistique Λ satisfait

$$2 \ln(\Lambda) \rightarrow V \sim \chi^2(1)$$

quand $H_0 : \theta = \theta_0$ est vraie.

Preuve - Notons $\ell(x; \theta) = \ln(f(x; \theta))$ et $\ell'(x; \theta)$ et $\ell''(x; \theta)$ ses dérivées par rapport à θ . Sous les hypothèses du théorème, sous H_0 ,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow Z \sim \mathcal{N}(0, 1/I(\theta_0))$$

En prenant le $\log(\Lambda_n)$ et en faisant un développement de Taylor, on a

$$\begin{aligned} \ln(\Lambda_n) &= \sum_{i=1}^n \left(\ell(X_i; \hat{\theta}_n) - \ell(X_i; \theta_0) \right) \\ &= (\theta_0 - \hat{\theta}_n) \sum_{i=1}^n \ell'(X_i; \hat{\theta}_n) - \frac{1}{2}(\hat{\theta}_n - \theta_0)^2 \ell''(X_i; \theta_n^*) \\ &= -\frac{1}{2}n(\hat{\theta}_n - \theta_0)^2 \frac{1}{n} \ell''(X_i; \theta_n^*) \end{aligned}$$

où θ_n^* est entre θ_0 et $\hat{\theta}_n$. Sous les hypothèses (A4) et (A5), du chapitre précédent, on a donc sous H_0

$$\frac{1}{n} \ell''(X_i; \theta_n^*) \rightarrow_P -E_{\theta_0}[\ell''(X_i; \theta_0)] = I(\theta_0)$$

On a aussi

$$n(\hat{\theta}_n - \theta_0)^2 \rightarrow_d \frac{V}{I(\theta_0)}$$

et on conclut avec le lemme de Slutsky. \diamond

Exemple - Soient X_1, \dots, X_n v.a.i.i.d. de loi de Gauss de moyenne μ et de variance σ^2 (toutes deux inconnues) et supposons qu'on veut tester

$$H_0 : \mu = \mu_0 \text{ contre } H_1 : \mu \neq \mu_0$$

Les estimateurs du maximum de vraisemblance de μ et σ^2 sont

$$\hat{\mu} = \bar{X} \text{ et } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Sous H_0 , l'estimateur du maximum de vraisemblance de σ s'écrit

$$\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2$$

Et on a donc

$$\begin{aligned} \Lambda_n &= \frac{(2\pi\sigma_0^2)^{n/2}}{(2\pi\hat{\sigma}^2)^{n/2}} \exp \left(-\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (X_i - \bar{X})^2 + \frac{1}{\hat{\sigma}_0^2} \sum_{i=1}^n (X_i - \mu_0)^2 \right) \\ &= \left(\frac{\sigma_0^2}{(\hat{\sigma}^2)} \right)^{n/2} \exp \left(\frac{1}{2\hat{\sigma}^2} n\hat{\sigma}^2 + \frac{1}{2\hat{\sigma}_0^2} n\hat{\sigma}_0^2 \right) \\ &= \left(\frac{\sigma_0^2}{(\hat{\sigma}^2)} \right)^{n/2} \end{aligned}$$

et la région critique est donc de la forme $\left\{ \left(\frac{\sigma_0^2}{\hat{\sigma}^2} \right)^{n/2} \geq c \right\}$ avec c déterminé par

$$P_{\theta_0} \left(\left(\frac{\sigma_0^2}{\hat{\sigma}^2} \right)^{n/2} \geq c \right) = \alpha$$

car le test est pur (loi continue). La distribution de Λ n'est pas triviale; cependant, on remarque que Λ est une fonction monotone de $\hat{\sigma}_0^2/\hat{\sigma}^2$ et

$$\begin{aligned} \frac{\sigma_0^2}{\hat{\sigma}^2} &= \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= 1 + \frac{n(\bar{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= 1 + \frac{1}{n-1} \left(\frac{n(\bar{X} - \mu_0)^2}{S^2} \right) \\ &= 1 + \frac{T^2}{n-1} \end{aligned}$$

où

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

et

$$T = \sqrt{n}(\bar{X} - \mu_0)/S.$$

Or on sait que sous H_0 , T suit une distribution de Student à $(n-1)$ degrés de liberté et donc que T^2 suit une distribution de Fisher à 1 et $(n-1)$ degrés de liberté.

On peut généraliser le théorème précédent aux de paramètres dans \mathbb{R}^p pour $p > 1$.

Théorème 4 Soient X_1, \dots, X_n v.a.i.i.d. qui admettent une densité (ou une fonction de fréquence) vérifiant les hypothèses A1 à A5 dans le cas d'un paramètre de dimension p et avec $I(\theta) = J(\theta)$ où $\theta = (\theta_1, \dots, \theta_p)$. Si l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ satisfait un théorème de limite centrale

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow Z \sim \mathcal{N}(0, 1/I(\theta))$$

alors la statistique Λ pour tester $H_0 : \theta_1 = \theta_{10}, \dots, \theta_r = \theta_{r0}$ satisfait

$$2 \ln(\Lambda) \rightarrow V \sim \chi^2(r)$$

quand H_0 est vraie.

La preuve du théorème repose sur le fait que la log-vraisemblance peut être approchée par une fonction quadratique près de la vraie valeur du paramètre.

Exemple - Soient X_1, \dots, X_m des variables aléatoires i.i.d. de loi exponentielle de paramètre λ et Y_1, \dots, Y_n des variables aléatoires i.i.d. de loi exponentielle de paramètre θ . On suppose que les X_i sont indépendants des Y_i . On veut tester

$$H_0 : \lambda = \theta \text{ contre } H_1 : \lambda \neq \theta$$

Les estimateurs du maximum de vraisemblance de λ et θ sont, dans le cas général,

$$\hat{\lambda} = 1/\bar{X} \text{ et } \hat{\theta} = 1/\bar{Y}$$

et sous H_0 ,

$$\hat{\lambda}_0 = \hat{\theta}_0 = \left(\frac{m\bar{X} + n\bar{Y}}{m + n} \right)^{-1}$$

On a donc

$$\Lambda = \left(\frac{m}{n+m} + \frac{n}{n+m} \frac{\bar{Y}}{\bar{X}} \right)^m \left(\frac{n}{n+m} + \frac{m}{n+m} \frac{\bar{X}}{\bar{Y}} \right)^n$$

On remarque que Λ ne dépend que de $T = \bar{X}/\bar{Y}$. On peut déduire un test de T ou construire un test asymptotique avec Λ .

Exemple - Soient $(X_1, Y_1), \dots, (X_n, Y_n)$ couples i.i.d. de variables aléatoires continues de densité jointe

$$f(x, y; \theta, \lambda, \alpha) = \frac{2\theta\lambda\alpha}{(\theta x + \lambda y + \alpha)^3} \text{ pour } x, y > 0$$

avec $\theta, \lambda, \alpha > 0$. Les densités marginales de X_i et Y_i sont

$$f_X(x; \theta, \alpha) = \frac{\theta\alpha}{(\theta x + \alpha)^2} \text{ pour } x > 0$$

$$f_Y(y; \lambda, \alpha) = \frac{\lambda\alpha}{(\lambda y + \alpha)^2} \text{ pour } y > 0$$

On veut tester $H_0 : \theta = \lambda$.

On peut reparamétriser ce problème de différentes façons. Par exemple, on peut définir $\eta_1 = \theta - \lambda$, $\eta_2 = \theta$ et $\eta_3 = \alpha$ ou $\eta_1 = \theta/\lambda$. On exprime alors H_0 en fonction de η_1 et on s'attend à ce que la statistique de test du rapport de vraisemblance suive approximativement une loi du χ^2 à 1 degré de liberté pour n grand.

1.6 Autres tests basés sur le maximum de vraisemblance

1.6.1 Test de Wald

$$H_0 : \theta = \theta_0 \text{ contre } \theta \neq \theta_0$$

Dans le test de Wald, l'estimateur du maximum de vraisemblance $\hat{\theta}$ du paramètre θ est comparé à la valeur θ_0 , sous l'hypothèse que la différence est distribuée approximativement selon une loi de Gauss. En pratique le carré de la différence est comparé à un seuil de la loi du chi 2. Dans le cas univarié, la statistique de Wald est

$$\frac{(\hat{\theta} - \theta_0)^2}{\text{var}(\hat{\theta})}$$

Si on compare la différence à un quantile de la loi de Gauss, la statistique de test est

$$\frac{\hat{\theta} - \theta_0}{\text{se}(\hat{\theta})}$$

où $\text{se}(\hat{\theta})$ est l'écart-type de l'estimateur du maximum de vraisemblance. Un estimateur raisonnable de cet écart-type est donné par $\frac{1}{\sqrt{I_n(MLE)}}$, où I_n est l'information de Fisher du paramètre.

Dans le cas univarié, un test sur plusieurs paramètres simultanément est réalisé en utilisant une matrice de variance. Par exemple, on utilise ce test pour une variable catégorielle recodée en plusieurs variables dichotomiques.

1.6.2 Score test (ou test des multiplicateurs de Lagrange)

Le test des multiplicateurs de Lagrange utilise le fait que si l'hypothèse nulle est fautive, alors, le gradient de la log vraisemblance ne doit pas être proche de 0. Plus précisément, en notant $S_i(\theta)$ le gradient de $\log(f(X_i; \theta))$ en fonction de θ , alors sous H_0 , on a

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n S_i(\hat{\theta})$$

tend en loi vers une variable de loi de Gauss centrée où $\hat{\theta}$ est l'estimateur du maximum de vraisemblance de θ sous H_0 .

Comme pour le test de Wald, on rejette H_0 pour les grandes valeurs de $S_n = \frac{1}{n} \left(\sum_{i=1}^n S_i(\hat{\theta}) \right)^2 / I_n(\theta)$ et, sous H_0 on a S_n tend en loi vers une variable qui suit un chi 2 à r degrés de liberté ($r = \dim(\theta)$).

1.7 Tests classiques

1.7.1 Tests paramétriques pour des moyennes, des variances ou des corrélations

- Test de Student : comparaison de moyennes dans le cadre de la loi de Gauss ou pour des grands échantillons.
- Test de Fisher : comparaison de moyennes dans le cadre de la loi de Gauss. Ce test permet notamment de tester l'égalité des variances de 2 échantillons.
- Analyse de la variance (variance inter-classe/ variance intra-classe)
- Test de Pearson : test de corrélation. $H_0 : \rho = 0$ contre $H_1 : \rho \neq 0$.

1.7.2 Tests non paramétriques pour des moyennes, des variances ou des corrélations

On les utilise quand on a de petits échantillons dont on ne connaît pas la distribution.

- Test du signe, test des signes et rangs de Wilcoxon ou mann-Whitney Wilcoxon
- Test de Kruskal-Wallis
- test de Spearman, Test du τ de Kendall

Test des signes et rangs de Wilcoxon

Hypothèses : Nous supposons que la distribution de la variable dans la population est *symétrique* et *continue*.

Etant donné un échantillon de n mesures indépendantes, nous pouvons au lieu de noter seulement les signes des écarts à la médiane spécifiée dans H_0 , relever aussi la grandeur de chaque écart. Si H_0 est vraie, les écarts d'une grandeur donnée ont autant de chance, pour une distribution symétrique, d'être positifs que négatifs ; et une valeur dépassant θ de 4 ou 5 unités a la même probabilité d'être observée qu'une valeur inférieure à θ de 4 à 5 unités. C'est sur cette idée que se base le test des **signes et rangs** de Wilcoxon⁵

Reprenons l'exemple de la taille des poissons. En notant θ la longueur médiane, les hypothèses de test sont

$$H_0 : \theta = 220 \text{ contre } H_1 : \theta \neq 220$$

Nous rappelons que nous avons observé l'échantillon suivant :

126 142 156 228 245 246 370 419 433 454 478 503

TABLE 1.2 – Longueur de 12 poissons

Formulation et postulat - Nous rangeons par ordre croissant les écarts à 220 (écarts en valeurs absolue), puis nous associons à chaque écart son signe (c'est à dire un signe + si l'observation correspondante est supérieure à la médiane spécifiée sous H_0 et un signe - sinon). On calculons la somme S_p des rangs des écarts positifs et la somme S_n des rangs des écarts négatifs. Si H_0 est vraie, on s'attend à ce que ces deux sommes soit presque égales. La statistique de test est la plus petite des deux sommes. Pour construire la région de rejet, nous utilisons la table des signes et rangs de Wilcoxon.

Le problème des ex aequo - Nous avons supposé que la distribution de la variable d'intérêt est continue dans la population. Or pour une distribution continue, la probabilité d'obtenir des observations égales est nulle de même que celle d'obtenir des observations égales à la médiane de la population. Cependant, en pratique, les observations ne sont pas strictement continues (arrondis ou précision limitée des appareils de mesure). Si une ou plusieurs valeurs coïncident avec la médiane spécifiée sous H_0 , nous leur attribuons le rang 0.

Si un grand échantillon comporte des valeurs égales à la médiane sous H_0 ou des ex aequo, on modifie Z de la façon suivante

$$Z = \frac{S - \frac{n(n+1)}{4} - d_0(d_0 + 1)}{\sqrt{n(n+1)(2n+1)/24 - d_0(d_0+1)(2d_0+1)/24 - \sum_{i=1}^{n_{ge}} (d_i^3 - d_i)/48}}$$

où d_0 est le nombre de valeurs égales à la médiane spécifiée sous H_0 , n_{ge} est le nombre de groupes d'ex aequo et d_i le nombre d'ex aequo dans le i ème groupe.

1.7.3 Test d'adéquation ou de comparaison de distribution

Test d'adéquation⁶.

5. En anglais, on dit *signed ranks test* ce qui est aussi traduit **test des rangs signés**.

6. En anglais : *Goodness-of-fit tests*

- Tests du χ^2 : loi discrète
- Test de Kolmogorov, Cramer-von Mises, : loi continue quelconque
- Test de Shapiro-Wilk : loi normale

Test du chi2

Pour une distribution discrète on utilise le test d'adéquation du chi 2.

Exemple : On suppose que le nombre de pièces défectueuses produites en un jour par une machine suit une loi de Poisson, de paramètre inconnu. Rappelons les caractéristiques de cette loi : si une variable aléatoire X suit une loi de Poisson de paramètre λ , alors $E(X) = \lambda$, $Var(X) = \lambda$, et pour tout $k \in \mathbb{N}$,

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

On pose les hypothèses de test

H_0 : X suit une loi de Poisson de paramètre 1.2 contre H_1 : X suit une autre loi

On observe 100 jours de production de cette machine, voici les résultats, regroupés en 5 classes.

| | | | | | |
|-------------------------------|------|------|------|------|-----------|
| Nombre de pièces défectueuses | 0 | 1 | 2 | 3 | 4 et plus |
| Nombre d'observations | 27 | 41 | 21 | 7 | 4 |
| Fréquence empirique | 0,27 | 0,41 | 0,21 | 0,07 | 0,04 |

On utilise une statistique de test du chi2 donnée par

$$T = \sum_{k=1}^K \frac{(\hat{n}f_k - np_k)^2}{np_k}$$

avec f_k et p_k les fréquences empirique et théorique de la classe k et K le nombre de classes. T suit une loi du chi 2 à $K - 1$ degrés de liberté.

Pour l'exemple considéré, les fréquences théoriques sont données ci-dessous.

| | | | | | |
|-------------------------------|------|------|------|------|-----------|
| Nombre de pièces défectueuses | 0 | 1 | 2 | 3 | 4 et plus |
| Fréquence théorique | 0,30 | 0,36 | 0,22 | 0,09 | 0,03 |

La statistique de test $T = 0,0112$; or d'après la table du chi 2 on rejette H_0 au risque 5% si $T > 5,99$.

Test de Kolmogorov-Smirnov

Soient x_1, \dots, x_n , n réalisations d'une variable aléatoire X . On se demande s'il est raisonnable de supposer que X suit la loi caractérisée par la fonction de répartition F et on pose les hypothèses de tests :

H_0 : X suit la loi F contre H_1 : X suit une autre loi

On propose de construire une statistique de test basée sur la distance entre les fonction F et une estimation de la fonction de répartition de X obtenue à partir des observations.

1.7.4 Estimer la fonction de répartition

On construit naturellement un estimateur de la fonction de répartition de X d'après l'équation (??).

$$F_n(x) = \frac{\text{Card}(\{i|x_i \leq x\})}{n}$$

Exercice - On considère les données d'un essai visant à déterminer la solidité d'une corde d'escalade. Un morceau de 1 m corde est mis sous tension jusqu'à cassure. On se demande si la corde pour casser à n'importe endroit. On obtient les résultats suivants :

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0.1 | 0.4 | 0.4 | 0.6 | 0.7 | 0.7 | 0.8 | 0.9 | 0.9 | 0.9 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

1. Tracer l'estimation de la fonction de répartition.
2. Ajouter sur le graphique la fonction de répartition théorique pour ce problème.

Kolmogorov propose d'utiliser la statistique de test suivante :

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$$

La loi de cette statistique de test est donnée dans la table de Kolmogorov.

Si on considère de nouveau les données de l'exercice, on obtient

| | | | | | | | | | | |
|--------------|------|------|------|------|------|------|------|-----|-----|-----|
| observations | 0.1 | 0.4 | 0.4 | 0.6 | 0.7 | 0.7 | 0.8 | 0.9 | 0.9 | 0.9 |
| F_n | 1/10 | 3/10 | 3/10 | 4/10 | 6/10 | 6/10 | 7/10 | 1 | 1 | 1 |
| F | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1 |
| $F - F_n$ | 0 | 0.1 | 0 | 0 | 0.1 | 0 | 0.2 | 0.1 | 0 | |

d'où $D_n = 0.2$ or on rejette H_0 au risque $\alpha = 5\%$ si $D_n > 0.369$. Ainsi ici on peut supposer que les observations sont issues d'une loi uniforme sur $[0, 1]$.

Cas de la loi normale

La version du test de Kolmogorov adaptée pour la loi de Gauss s'appelle le test de Lilliefors. Ce test est peu puissant et on lui préfère le test de Shapiro-Wilk. Ce dernier est basé sur une comparaison de deux estimateurs de la variance qui ne peuvent conduire à la même estimation que si les observations sont issues d'une loi de Gauss.

Pour vérifier qu'une série d'observation suit une loi normale, on peut en première approche utiliser une méthode graphique : la droite de Henry (*quantile-quantile plot* ou *qqplot*).

Soit $\{x_1, \dots, x_n\}$ une suite d'observations. Si cette suite constitue une suite de réalisation d'une variable gaussienne, alors les points de coordonnées $(x_i, \Phi^{-1}((i-1/2)/n))$ sont alignés sur la droite d'équation

$$y = \frac{x - \bar{x}}{\hat{\sigma}}$$

Cette droite est appelée *droite de Henry*.

Comparaison de deux distributions

On peut généraliser le test de Kolmogorov au cas de deux échantillons afin de comparer leurs distributions. Le test s'appelle alors test de Kolmogorov-Smirnov. L'hypothèse nulle est que les deux échantillons proviennent de la même distribution ; l'alternative est qu'ils

proviennent de distributions ayant des répartitions différentes. On ne spécifie aucune forme particulière pour leur différence. Et la statistique de test est basée sur un écart en valeur absolue entre la fonctions de répartition empiriques des deux suites d'observations.

Il existe d'autres tests permettant de comparer des distributions. Par exemple, le test de Cramér-von Mises repose sur la somme des carrés des écarts en valeurs absolue entre les deux fonctions de répartition. En notant, S_d^2 cette somme, la statistique de test est

$$T = \frac{nmS_d^2}{n+m}$$

avec m et n les nombres d'observation des deux groupes.

Pour un test bilatéral, on rejette H_0 au niveau de signification 5% (resp. 1%) si T est supérieur à 0.461 (resp. 0.743).

Le test de Cramér-von Mises est souvent plus puissant que le test de Kolmogorov-Smirnov et il est plus facile à utiliser grâce à la bonne approximation qui évite le recours à des tables.

Chapitre 2

Estimation par intervalles

On a vu dans le chapitre sur l'estimation qu'une statistique n'est en général pas exactement égale à la valeur du paramètre qu'elle est censée estimer (si la loi de la statistique a une densité, cet événement est même de probabilité nulle). Donc, il est important que toute procédure d'estimation soit accompagnée d'une indication sur la précision de l'estimation.

2.1 Exemple

On considère un n -échantillon X_1, \dots, X_n de la loi gaussienne de paramètres (μ, σ^2) . On suppose que σ^2 est connu. On estime μ . On a déjà vu toutes les qualités de l'estimateur $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. On sait que la variable $\sqrt{n}\sigma^{-1}(\bar{X}_n - \mu)$ est gaussienne standard. Sa loi ne dépend pas du paramètre μ , ce qu'on a déjà utilisé pour construire un test de niveau α pour tester $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$ ou contre $H_1 : \mu > \mu_0$. On sait par exemple que pour tout μ

$$P_\mu \left(\mu - \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n \leq \mu + \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha \quad (2.1)$$

Or ceci peut aussi se lire

$$P_\mu \left(\bar{X}_n - \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha \quad (2.2)$$

ce qui change tout car dans (2.2), l'intervalle est aléatoire (la probabilité qu'il contienne le paramètre est $1 - \alpha$), tandis que dans (2.1), l'intervalle est fixé et la variable aléatoire \bar{X}_n a une probabilité $1 - \alpha$ de se trouver dedans.

On dit que

$$I(X_1, \dots, X_n) = \left[\bar{X}_n - \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}}, \bar{X}_n + \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}} \right]$$

est un intervalle de confiance de niveau $1 - \alpha$ pour le paramètre μ .

Remarque - On a choisi un intervalle symétrique : est ce nécessaire ?

Dans une certaine mesure la réponse est oui. En effet, soit U une variable gaussienne

standard. Il est facile de vérifier que, $P(U \in [x, x + 2\Phi^{-1}(1 - \alpha/2)])$ est maximale (et vaut $1 - \alpha$) quand $x = -\Phi^{-1}(1 - \alpha/2)$. On en déduit que parmi les intervalles tels que $P(U \in [x, y]) = 1 - \alpha$, le plus court est $[-\Phi^{-1}(1 - \alpha/2), \Phi^{-1}(1 - \alpha/2)]$. Autrement dit l'intervalle proposé en (2.2) est le plus précis des intervalles de confiance de niveau $1 - \alpha$ pour le paramètre μ . Ceci dit, on peut avoir d'autres critères que la précision pour choisir l'intervalle de confiance. On peut vouloir par exemple une demi-droite de confiance si le seul souci est de garantir que le paramètre est suffisamment grand (ou suffisamment petit).

2.2 Méthode générale pour construire des intervalles de confiance

On cherche à estimer $\theta \in \Theta$ à partir de (X_1, \dots, X_n) , variables indépendantes et de même loi P_θ . Ici, $\Theta \subset \mathbb{R}^d$.

Définition 10 On appelle fonction pivotale, une fonction $h(\theta; X_1, \dots, X_n)$ à valeurs dans \mathbb{R}^k possédant les propriétés suivantes :

- Propriété 1 : quelque soit $\theta \in \Theta$, la fonction $h(\theta; X_1, \dots, X_n)$ est mesurable.
- Propriété 2 : la loi de $h(\theta; X_1, \dots, X_n)$ ne dépend pas de θ .

Il est facile de s'appuyer sur une fonction pivotale pour construire des intervalles de confiance de la façon suivante. Supposons qu'il existe B , un borélien de \mathbb{R}^k , tel que

$$P_\theta(h(\theta; X_1, \dots, X_n) \in B) = 1 - \alpha, \forall \theta \in \Theta \quad (2.3)$$

on définit alors la région de confiance

$$I(X_1, \dots, X_n) = \{\theta \in \Theta | h(\theta; X_1, \dots, X_n) \in B\}$$

Comme dans l'exemple introductif, l'ensemble $I(X_1, \dots, X_n)$ est aléatoire et a, sous P_θ , une probabilité $1 - \alpha$ de contenir θ . Dans l'exemple on avait

$$h(\mu; X_1, \dots, X_n) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

et le borélien B était l'intervalle $[-\Phi^{-1}(1 - \alpha/2), \Phi^{-1}(1 - \alpha/2)]$

Remarques -

- L'égalité (2.3) s'écrit aussi $\pi(B) = 1 - \alpha$ où π désigne la loi (ne dépendant pas de θ , d'après la Propriété 2) transportée de P_θ^n par $h(\theta, x)$. Bien sûr, un tel B peut ne pas exister. On recherchera alors un borélien tel que $\pi(B) > 1 - \alpha$.
- On peut aussi bien se servir de la fonction pivotale pour construire un test de l'hypothèse $\theta = \theta_0$. Pour cela, il suffit de prendre comme région de rejet l'ensemble des (X_1, \dots, X_n) tels que $h(\theta_0; X_1, \dots, X_n) \notin B$.

2.3 Lien avec les tests

On peut construire des intervalles de confiance à l'aide des tests. La région d'acceptation du test définit un intervalle de confiance.

Théorème 5 Soit $\mathbf{X} = X_1, \dots, X_n$ un n -échantillon de loi $(P_\theta)_{\theta \in \Theta}$. Pour chaque $\theta_0 \in \Theta$ on considère le problème de tester $H_0 : \theta = \theta_0$ au niveau α et soit $A(\theta_0) = RC(\theta_0)^c$ sa région d'acceptation.

Pour chaque $(x_1, \dots, x_n) \in \mathbb{R}^n$ on définit $C(x)$ par

$$C(x_1, \dots, x_n) = \{\theta_0 \in \Theta : \mathbf{x} \in A(\theta_0)\}.$$

Alors $C(X_1, \dots, X_n)$ est une région de confiance pour θ de niveau $1 - \alpha$.

Réciproquement, soit $C(X)$ une région de confiance pour θ de niveau $1 - \alpha$. Pour tout $\theta_0 \in \Theta$ on définit

$$A(\theta_0) = \{(x_1, \dots, x_n) : \theta_0 \in C(\mathbf{x})\}$$

Alors $A(\theta_0)$ est la région d'acceptation d'un tests de niveau α pour $H_0 : \theta = \theta_0$.

Preuve - Comme $A(\theta_0)$ est la région d'acceptation d'une test de niveau α , on a

$$P_{\theta_0}(\mathbf{X} \notin A(\theta_0)) \leq \alpha$$

ou encore

$$P_{\theta_0}(\mathbf{X} \in A(\theta_0)) \geq 1 - \alpha$$

Comme θ_0 est quelconque, on peut écrire θ à la place de θ_0 . Alors, d'après l'inégalité précédente,

$$P_\theta(\theta \in C(\mathbf{X})) = P_{\theta_0}(\mathbf{X} \in A(\theta_0)) \geq 1 - \alpha$$

donc C est une région de confiance pour θ de niveau $1 - \alpha$.

Par ailleurs, on voit que l'erreur de 1ère espèce pour H_0 avec pour région d'acceptation $A(\theta_0)$ est

$$P_{\theta_0}(\mathbf{X} \notin A(\theta_0)) = P_{\theta_0}(\theta_0 \notin C(\mathbf{X})) \leq \alpha$$

donc le test est de niveau α sous H_0 . \diamond

Références

Knight, K., *Mathematical Statistics*, Taylor Francis, 1999 - 504 pages.

Balabdaoui-Mohr F., Wintenberger O., *Statistique Mathématique*. Notes de cours.