Agrégation externe de mathématiques
Tests d'hypothèses

Bibliographie :
- Statistique en Action, G. Stoltz et V. Rivoirard
- Statistique et Analyse de Données, G. Saporta
- P. S. Toulouse,
- Mathematical statistics. Basic ideas and Selected topics, P.J. Bickel et K.A. Docksum

1 Notions générales sur les tests statistiques

On manipulera au niveau de l'agrégation des tests paramétriques le plus souvent, même s'il existe également des tests non paramétrique (exemple des tests d'ajustement du χ^2 par exemple) valable pour des variables aléatoires dont on ne connait pas la loi, ce qui est en pratique très fréquent.

1.1 Définitions

On donne donc un modèle $(\Omega, \mathcal{A}, \mathbb{P}_\theta, \theta \in \Theta)$. L'objectif est de donner à la vue des observations des décisions sur la valeur de θ. Si $\Theta = \Theta_0 \cup \Theta_1$, l'hypothèse $\mathcal{H}_0$ est définie par $\theta \in \Theta_0$ et est à tester contre l'hypothèse alternative $\mathcal{H}_1$ définie par $\theta \in \Theta_1$.

Definition 1.1 (Test d'hypothèse) *On appelle test de l'hypothèse $\mathcal{H}_0$ contre $\mathcal{H}_1$ toute fonction $\phi(X_1, \dots X_n)$ à valeurs dans $\{0; 1\}$ où ϕ est mesurable et est calculée d'un n-échantillon i.i.d. de la variable aléatoire X observée de loi $\mathbb{P}_\theta$.*

Si $\phi(X) = 0$, on conserve $\mathcal{H}_0$ et sinon on accepte $\mathcal{H}_1$.

Definition 1.2 (Hypothèse composite) *L'hypothèse $\mathcal{H}_j$ est dite simple si Θ_j est réduit à un singleton, sinon $\mathcal{H}_j$ est dite composite.*

Definition 1.3 (Zone de Rejet) *On définit la zone de Rejet du test basé sur ϕ par $\phi^{-1}(1)$.*

Attention, nous verrons par la suite que $\mathcal{H}_0$ et $\mathcal{H}_1$ ne jouent pas un rôle symétrique.

Definition 1.4 (Niveau du test, puissance du test) *Le niveau du test (ou risque de première espèce) est défini par :*

$$\alpha = P_{\mathcal{H}_0}(\phi(X_1, \dots X_n)) = 1.$$

Le risque de seconde espèce est défini par :

$$\beta = P_{\mathcal{H}_1}(\phi(X_1, \dots X_n)) = 0.$$

La puissance du test est alors la quantité $1 - \beta$. Un test sera dit de niveau α si son risque de première espèce est inférieur ou égal à α.

Le risque de première espèce permet de mesurer la probabilité de rejetter à tort $\mathcal{H}_0$.

Exercice 1.5 *On suppose observé un n-échantillon i.i.d. de loi $\mathcal{N}(\theta, 1)$ pour θ inconnu réel. On veut tester $\mathcal{H}_0$: "$\theta \geq 1$" contre $\mathcal{H}_1$: "$\theta < 1$". Que penser du test basé sur la statistique de test :*

$$\phi(X_1, \dots X_n) = 1_{\{\bar{X}_n < 1\}}$$

L'exercice précédent n'est pas réellement représentatif puisque en général, il est essentiel de ne pas rejeter à tort $\mathcal{H}_0$. Pour l'exemple précédent, construire un test statistique basé sur $\bar{X}_n$ qui rejette à tort $\mathcal{H}_0$ au niveau α.

Definition 1.6 (Propriétés des tests) *Comme pour les estimateurs, on peut définir des propriétés asymptotiques sur les tests (le nombre d'observations tendant virtuellement vers $+\infty$).*

Un test est <u>sans biais</u> lorsque la puissance du test $1-\beta$ est supérieure au niveau α sur Θ_1.

Un test est consistant si la suite des risques de seconde espèce tend vers 0 tandis que les tests se maintiennent au niveau α.

Definition 1.7 (Test Uniformément plus puissant (U.P.P.)) *Pour deux statistiques de test ϕ et $\tilde{\phi}$ de niveau α, on dit que ϕ est plus puissant que $\tilde{\phi}$ si la puissance de ϕ est supérieure à la puissance de $\tilde{\phi}$.*

Un test est dit uniformément plus puissant de niveau α s'il est uniformément plus puissant que tout autre test de niveau α.

Il n'existe pas toujours de test U.P.P.(α) mais nous verrons plus loin pour quels types d'hypothèses on peut en construire.

1.2 Protocole Statistique

Le protocole de test suit généralement le schéma suivant :

1. Choix des hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$.
2. Détermination de la statistique de test.
3. Détermination de la région d'acceptation de $\mathcal{H}_0$ en fonction de α.
4. Calcul de la puissance du test.
5. Calcul de la statistique de test.
6. Acceptation ou rejet de $\mathcal{H}_0$.

2 Méthode de Neyman et Pearson

C'est pas vraiment très récent (1935-1938) !

2.1 Test du rapport de vraisemblance

Pour construire des tests U.P.P.P(α), on utilise souvent le test du rapport de vraisemblance. La statistique du test est alors définie en toute généralité par

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

Il faut comprendre intuitivement que ϕ a tendance à être grand lorsque X suit une loi pour $\theta \in \Theta_1$: le numérateur est alors grand tandis que le dénominateur est faible. Ainsi, la région naturelle de rejet de l'hypothèse $\mathcal{H}_0$ est de la forme $\phi(X) > k_\alpha$ où k_α est à calibrer en fonction des valeurs de θ_0, θ_1 ainsi que des lois paramétriques sur X et bien sûr de α.

Exercice 2.1 *Définir le test du rapport de vraisemblance pour l'exercice précédent à partir de la statistique h du rapport de vraisemblance.*

2.2 Test entre deux hypothèses simples

Il s'agit ici de tester $\mathcal{H}_0$ contre $\mathcal{H}_1$ avec $\mathcal{H}_0$: "$\theta = \theta_0$" et $\mathcal{H}_1$: "$\theta = \theta_1$". Le résultat suivant assure l'optimalité des tests de rapport de vraisemblance dans le cadre du test de deux hypothèses simples.

Théorème 2.2 (Lemme de Neyman-Pearson) *Pour $\alpha \in]0;1[$ donné, lorsqu'il existe k_α tel que le test de rapport de vraisemblance $\phi(X) = 1_{h(X)>k_\alpha}$ vérifie*

$$\mathbb{P}_{\theta_0}[\phi(X) = 1] = \alpha,$$

alors ce test est U.P.P.(α).

On démontre également que le test de Neyman-Pearson est sans biais, et consistant. Si α diminue, le seuil k_α augmente donc diminuer le risque de première espèce fait automatiquement augmenter le risque de seconde espèce β car $1 - \beta = \mathbb{P}_{\theta_1}(L_X(\theta_1)/L_X(\theta_0) > k_\alpha)$.

Remarque 2.3 (Lien avec les statistiques exhaustives) *La considération d'une statistique exhaustive simplifie considérablement la pratique du test car alors la région de rejet ou d'acceptation en dépend exclusivement. En effet, si T est exhaustive pour θ, on a*

$$L_X(\theta) = g(T, \theta)h(X)$$

et le test de Neyman-Pearson se réduit alors à

$$\frac{g(T, \theta_1)}{g(T, \theta_0)} > k_\alpha.$$

Exercice 2.4 *Tester la moyenne d'une loi normale de variance connue σ pour deux valeurs possibles des moyennes m_1 et m_2.*

2.3 Test entre hypothèses composites

2.3.1 Simple Vs Composite

Il s'agit ici de tester $\mathcal{H}_0$ contre $\mathcal{H}_1$ avec $\mathcal{H}_0$: "$\theta = \theta_0$" et $\mathcal{H}_1$: "$\theta > \theta_0$" ou bien encore $\mathcal{H}_1$: "$\theta \neq \theta_0$".
Bien entendu, le risque de première espèce ne change pas, par contre la puissance du test dépend de la valeur de θ sous $\mathcal{H}_1$. On dit qu'un test est U.P.P.(α) si quelle que soit la valeur de θ appartenant à l'hypothèse alternative $\mathcal{H}_1$, sa puissance $1 - \beta(\theta)$ est supérieure à celle de tout autre test de niveau α.

2.3.2 Test entre deux hypothèses composites

Si $\mathcal{H}_0$ est elle-même composite, α dépend de θ et on exige un $\alpha(\theta) \leq \alpha$, pour tout θ. Le théorème de Lehmann démontre qu'il existe toujours un test U.P.P.(α) pour les hypothèses de la forme :

$$\mathcal{H}_0 : "\theta < \theta_0" \qquad \mathcal{H}_1 : "\theta \geq \theta_0"$$

et

$$\mathcal{H}_0 : "\theta \leq \theta_1 \quad \text{ou} \quad \theta \geq \theta_2" \qquad \mathcal{H}_1 : "\theta_1 < \theta < \theta_2"$$

Ici, il s'agit donc de faire des tests soit unilatères, soit bilatères. C'est assez facile à voir dès lors que le modèle est exponentiel canonique :

$$f_\theta(x) = h(x)C(\theta)e^{\theta t(x)}.$$

En effet, on peut alors définir le test de vraisemblance monotone basé sur l'hypothèse suivante (vraie dans le cas d'un modèle exponentiel) :

Definition 2.5 (Modèle à rapport de vraisemblance monotone) *Ces modèles vérifient l'existence d'une statistique T telle que $d\mathbb{P}_\theta(X) = f_\theta(T(X))d\mu(X)$ avec en plus :*

$$\theta < \theta' \qquad \Rightarrow \qquad \frac{f_{\theta'}(T)}{f_\theta(T)} \qquad \textit{croissant en } T.$$

Exercice 2.6 *Traiter le cas des tests sur les moyennes pour le cas Gaussien.*

Théorème 2.7 (Tests unilatères dans les modèles à vraisemblance monotone) *Soit un modèle statistique $(\Omega, \mathcal{A}, \mathbb{P}_\theta)$ à rapport de vraisemblance monotone pour la statistique réelle T. Alors, $\forall \alpha \in]0,1[, \exists C \in \mathbb{R}, \exists \gamma \in [0,1[$, tels que le test*

$$\phi(X) = 1_{T(X)>C} + \gamma 1_{T(X)=C}$$

est U.P.P.(α). De plus, la fonction $\theta \mapsto \mathbb{E}_\theta[\phi]$ est croissante et $\alpha = \mathbb{E}_\theta[\phi]$. Ainsi ce test est sans biais.

Il existe aussi une version de ce théorème dans le cas des modèles exponentiels avec les tests bilatères.

3 Exercices

Exercice 1

On considère un échantillon $(X_1, \ldots, X_n)$ de loi $\mathcal{N}(\theta, \sigma^2)$. On souhaite tester l'hypothèse H_0 : « $\theta = \theta_0$ »contre l'alternative H_1 : « $\theta = \theta_1$ avec $\theta_0 < \theta_1$

1. Montrer que le test de rapport de vraisemblance est fondé sur la statistique $S = X_1 + \cdots + X_n$.
2. Sachant que si Y suit une loi $\mathcal{N}(0,1)$, alors $\mathbb{P}(Y > 1.64) = 0.05$, expliciter la région de rejet du test de niveau 5% dans notre cas.
3. Proposer une estimation par minoration de la puissance de ce test.
4. Que suggérez-vous pour augmenter la puissance du test ?
5. On désire cette fois tester l'hypothèse $\mathcal{H}_0$ « $\theta \leq 0$ »contre l'hypothèse $\mathcal{H}_1$ « $\theta \geq 0$ ». Montrer que le modèle est une famille à rapport de vraisemblances monotones.
6. Donner le test de puissance maximale de niveau α donné.
7. Un radar actif de surveillance aérienne a des caractéristiques telles qu'une éventuelle cible réfléchit 20 impulsions lors d'un balayage. À l'aide d'un traitement adapté, ces N impulsions réfléchies en cas de présence de la cible fournissent un vecteur d'observations (z_i) avec

$$\mathcal{H}_1 : z_i = A + b_i \qquad \text{en présence de cible}$$

$$\mathcal{H}_0 : z_i = b_i \qquad \text{en l'absence de cible}$$

   où les b_i sont des variables aléatoires gaussiennes $\mathcal{N}(0, \sigma^2)$ indépendantes modélisant les divers bruit.

   (a) Donner le test de Neyman Pearson de niveau α de $\mathcal{H}_0$ contre $\mathcal{H}_1$. Application numérique $A = 1, \sigma = 0.6$ et $\alpha = 10^{-6}$.

   (b) Écrire un code matlab simulant n échantillons, chacun de taille N sous l'hypothèse nulle. Appliquer pour chacun des échantillons le test de N-P et calculer la fréquence empirique des fausses détections.

   (c) Faire la même chose sous l'hypothèse $\mathcal{H}_1$ et calculer la fréquence empirique des cibles non détectées.

8. Une antenne de communication cherche à détecter l'émission d'un signal. À cause des perturbations atmosphériques, même en l'absence de signal, l'antenne perçoit du « bruit ». On procède à n mesures et on modélise le bruit par des v.a.i.i.d. de loi gaussienne $\mathcal{N}(0, \sigma^2)$ et le signal par des v.a.i.i.d. $\mathcal{N}(0, \xi^2)$, signal et bruit étant considérés indépendants. On veut tester l'hypothèse $\mathcal{H}_0$ « pas de signal »contre l'alternative « signal ».
   (a) Formaliser le problème de test dans notre situation.
   (b) Déterminer la statistique de test du rapport de vraisemblance. En déduire que le test est fondé sur une statistique de loi du χ^2.
   (c) Comment déterminer la région de rejet pour obtenir un test de niveau α donné et calculer sa puissance.

Exercice 2

Soient X et Y deux variables aléatoires. On veut tester l'hypothèse $\mathcal{H}_0$: « X et Y ont même loi »contre $\mathcal{H}_1$: « X et Y n'ont pas même loi ». On dispose pour cela d'un échantillon $(X_1, \dots X_m)$ de loi celle de X et d'un échantillon $(Y_1, \dots Y_n)$ de loi celle de Y.

Toutes ces $m+n$ variables aléatoires étant indépendantes. On range par ordre croissant les $m+n$ valeurs effectivement observées et après effacement des indices on obtient un « mot »de la forme $XXYXYYYXYXX$. On compte alors le nombre R de blocs de lettres identiques (les longueurs ou « runs »ou composantes connexes) dans ce mot, (sur cet exemple, $R = 7$ soit 4 longueurs de X et 3 de Y).

L'idée intuitive de ce test est que c'est sous $\mathcal{H}_0$ que le mélange entre les valeurs de X et celles de Y sera le plus intense et donc le nombre de longueur le plus élevé.

1. Sous $\mathcal{H}_0$, calculer la loi de R par les formules suivantes où l'on suppose $m \leq n$:

$$P(R = 2k) \qquad 1 \leq k \leq m$$

$$P(R = 2k+1) \qquad 1 \leq k < m$$

$$P(R = 2m+1)$$

2. Lorsque m et n sont grands, quelle est la distribution (sous $\mathcal{H}_0$) asymptotiquement obtenue pour R (répondre grâce à Matlab) ? On admettra que

$$\mathbb{E}[R] = 1 + \frac{2mn}{m+n} \qquad Var(R) = \frac{2mn(2mn - m - n)}{(m+n)^2(m+n-1)}$$

3. Définir un test de niveau ϵ de rejet de $\mathcal{H}_0$ à partir de R.
4. Écrire une fonction Matlab qui calcule R pour les échantillons X et Y.
5. Écrire une fonction Matlab qui réalise le test des longueurs au niveau asymptotique α pour m et n grands.
6. Écrire une fonction Matlab qui calcule la fonction de répartition de R. La comparer avec la fonction de répartition de la loi gaussienne correspondant à l'approximation utilisée ci-dessus.

Exercice 3

Θ ouvert de $\mathbb{R}$, on considère $H_0 = \{\theta \in \Theta \quad | \quad \theta \leq \theta_0\}$ et $H_1 = \{\theta \in \Theta \quad | \quad \theta > \theta_0\}$. On suppose que la famille des densités est de la forme

$$p(x, \theta) = C(\theta)h(x)e^{Q(\theta)S(x)}$$

où Q est strictement monotone. On pose alors $S_n(x) = \sum_{i=1}^{n} S(x_i)$. On a alors la proposition :

Proposition 3.1 *Soit $\alpha \in]0;1[$, il existe un test UPP de niveau α pour tester H_0 contre H_1 défini par*

$$\phi(x) = 1 \quad si \quad S_n(x) > c \qquad et \qquad \phi(x) = 0 \quad si \quad S_n(x) < c \qquad et \qquad \phi(x) = \gamma \quad si \quad S_n(x) = c$$

Les constantes γ et c sont déterminées par la condition $\mathbb{E}_{\theta_0}[\phi] = \alpha$.

Remarque 3.2 *Il n'existe pas en général de test UPP pour tester H_0 contre une hypothèse alternative bilatère du type $H_1 = \{\theta \neq \theta_0\}$.*

1. Démontrer ce résultat dans le cas particulier du modèle gaussien de moyenne inconnue et de variance fixée.
2. On admettra le théorème :

   Théorème 3.3 (Tests UPPS) *Il existe un test UPPS au seuil α pour tester $H_0 =]\theta_1;\theta_2[$ contre $H_1 = \Theta \backslash]\theta_1;\theta_2[$. Ce test est défini par*

   $$\phi(x) = 1 \quad si \quad S_n(x) \notin [c_1;c_2] \qquad et \qquad \phi(x) = \gamma_i \quad si \quad S_n(x) = c_i \qquad et$$

   Les constantes sont alors déterminées par les conditions $\mathbb{E}_{\theta_1}[\phi] = \alpha$ et $\mathbb{E}_{\theta_1}[\phi] = \alpha$.

   Déterminer un test pour le modèle gaussien d'hypothèse nulle $H_0 = \{\mu = \mu_0\}$ et d'hypothèse alternative $H_1 = \{\mu \neq \mu_0\}$ de niveau $\alpha = 5\%$ et de puissance maximale parmi les tests sans biais.

Application numérique

Une machine fabrique des ailettes de turbo avec pour caractéristique la longueur d'une ailette $\sim \mathcal{N}(L_0, \sigma_0)$. On souhaite avoir $L_0 = 785mm$ et $\sigma_0 = 2mm$. Une trop grande dispersion des caractéristiques peut avoir deux conséquences :
– Les ailettes trop longues ne sont pas montables.
– Les ailettes trop courtes nuisent aux performances du turbo.
On a donc étudié la variance et σ_0 peut être considéré comme fixé et invariable. Par contre, la longueur moyenne a tendance à varier. On désire vérifier que les caractéristiques de la machine n'ont pas trop dérivé, *i.e.* $L \in]L_0 - \delta L; L_0 + \delta L[$ avec $\delta L = 1.5mm$. On procède à l'analyse de $n = 200$ ailettes, ce qui conduit à $\hat{L_n} = 788.3mm$.

1. Vérifier que le modèle est exponentiel, construire un estimateur de L.
2. On souhaite tester $H_0 = L \in]L_0 - \delta L; L_0 + \delta L[$ contre $H_1 = L \notin]L_0 - \delta L; L_0 + \delta L[$. Construire un test bilatéral UPPS au seuil α pour tester H_0 contre H_1.
3. Faire l'application numérique pour $\alpha = 5\%$.
4. Calculer un intervalle de confiance centré à 95% sur L et le comparer à $L_0 + \delta L$.

Exercice 4

On considère un échantillon $(X_1, Y_1), \ldots, (X_n, Y_n)$ de v.a. de lois $\mathcal{N}((\mu_1, \mu_2); \Sigma_2)$ avec

$$\Sigma_2 = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

où σ_1 et σ_2 sont inconnus. Le paramètre θ est donc $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$.

1. Donner le test de Wald pour tester si $\mu_1 = \mu_2$ et donner sa région critique de niveau $\alpha = 5\%$.
2. On suppose $\sigma_1 = \sigma_2$, donner une région critique de niveau exact 5%, construite à l'aide de la même statistique de test que celle utilisée précédemment. Application numérique pour $n = 15$.
3. On dispose de 2 échantillons de tailles différentes $(X_1; \ldots; X_{n_1})$ et $(Y_1; \ldots; Y_{n_2})$ de lois respectives $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$ où μ_i et σ_i^2 sont inconnus. Déduire des questions précédentes un test asymptotique en $\min(n_1, n_2)$ pour l'hypothèse nulle $\mu_1 = \mu_2$ contre son alternative $\mu_1 \neq \mu_2$.