# Tests d'hypothèses

Les exemples sont développés en cours...

## 1 Cadre

Comme dans le chapitre précédent, on s'intéresse à des données $x_{\text{obs}} \in \mathscr{X}$. On suppose que ces données ont été engrendré par une loi $P_\theta$ qui dépend d'un paramètre $\theta \in \mathscr{T}$ dont on ne connait pas la valeur. On note $X$ l'objet aléatoire qui modélise les données : $X \sim P_\theta$.

On partitionne l'espace des valeurs possibles du paramètre en deux :

$$\mathscr{T} = \mathscr{T}_0 \cup \mathscr{T}_1, \quad \text{avec } \mathscr{T}_0 \cap \mathscr{T}_1 = \varnothing.$$

Pour des questions techniques, il est courant que $\mathscr{T}_0$ soit une partie fermée de $\mathscr{T}$. Ainsi, la frontière $\partial\mathscr{T}_0$ de $\mathscr{T}_0$ est incluse dans $\mathscr{T}_0$.

L'objectif est de savoir si $\theta$ est dans $\mathscr{T}_0$ ou $\mathscr{T}_1$. Il s'agit donc de déterminer, parmi ces deux hypothèses, laquelle est juste. On s'intéresse donc au paramètre

$$\psi = \Psi(\theta) = \mathbf{1}\{\theta \in \mathscr{T}_1\} = \begin{cases} 1 & \text{si } \theta \in \mathscr{T}_1, \\ 0 & \text{si } \theta \in \mathscr{T}_0. \end{cases}$$

On veut donc construire une statistique $\hat{\psi} = s(X)$ qui estime $\psi$. On dit parfois que $\hat{\psi}$ est une **règle de décision**. Pour évaluer cet estimateur, on introduit la fonction de coût

$$d(\hat{\psi}, \psi) = \mathbf{1}\{\hat{\psi} = \psi\}.$$

Dans ce cas, le risque est donné par :

$$R_\theta(\hat{\psi}) = \mathbb{P}_\theta(\hat{\psi} = \psi) = \begin{cases} \mathbb{P}_\theta(\hat{\psi} = 1) & \text{si } \theta \in \mathscr{T}_1, \\ \mathbb{P}_\theta(\hat{\psi} = 0) & \text{si } \theta \in \mathscr{T}_0. \end{cases}$$

## 2 Dissymétrie des hypothèses

Pour construire une statistique dont on contrôle partiellement le risque, le principe de Neyman-Pearson dissymétrise les deux hypothèses. On appelle **hypothèse nulle** le fait $\theta \in \mathscr{T}_0$ et **hypothèse alternative** $\theta \in \mathscr{T}_1$. Dans le jargon des tests, on dit que l'on teste

$$(H_0) \ : \ \theta \in \mathscr{T}_0 \quad \text{contre} \quad (H_1) \ : \ \theta \in \mathscr{T}_1.$$

On peut commettre deux types d'erreur :
- une erreur de type I (ou de première espèce) : décider en faveur de $(H_1)$ alors que $(H_0)$ est vraie,
- une erreur de type II (ou de seconde espèce) : décider en faveur de $(H_0)$ alors que $(H_1)$ est vraie.

On dissymétrise les deux hypothèses en décidant de contrôler le risque d'une erreur de type I. Cette erreur est contrôlée au travers du pire cas. Ainsi, on veut construire une règle de décision $\hat{\psi}$ dont on contrôle le risque de première espèce maximal

$$\sup_{\theta \in \mathscr{T}_0} R_\theta(\hat{\psi}).$$

**Définition 1.** Le risque de première espèce du test $\hat{\psi}$ est la fonction

$$\theta \in \mathscr{T}_0 \mapsto \alpha(\theta) = R_\theta(\hat{\psi}) = \mathbb{P}_\theta(\hat{\psi} = 1).$$

Le test $\hat{\psi}$ est de niveau $\alpha$ si

$$\sup_{\theta \in \mathcal{T}_0} R_\theta(\hat{\psi}) \le \alpha.$$

Et, il est dit de taille $\alpha$ si

$$\sup_{\theta \in \mathcal{T}_0} R_\theta(\hat{\psi}) = \alpha.$$

La ou les valeurs de $\theta$ pour lesquelles ce maximum est atteint s'appellent les pires cas du test. Dans toutes les situations intéressantes, ces pires cas sont à la frontière de l'hypothèse nulle, notée $\partial T_0$. Enfin, la pratique courante par défaut est de choisir $\alpha = 0.05$.

**Définition 2.** Le risque de de deuxième espèce du test $\hat{\psi}$ est la fonction

$$\theta \in \mathcal{T}_1 \mapsto \beta(\theta) = R_\theta(\hat{\psi}) = \mathbb{P}_\theta(\hat{\psi} = 0).$$

La fonction puissance de ce test est la fonction

$$\theta \in \mathcal{T}_1 \mapsto 1 - \beta(\theta) = \mathbb{P}_\theta(\hat{\psi} = 1).$$

On peut comparer deux tests de même taille $\alpha$ en comparant leur fonction puissance. Si $\hat{\psi}$ et $\tilde{\psi}$ sont deux tests de niveau $\alpha$, on dit que $\hat{\psi}$ est plus puissant que $\tilde{\psi}$ si, pour tout $\theta \in \mathcal{T}_1$,

$$\mathbb{P}_\theta(\hat{\psi} = 1) \ge \mathbb{P}_\theta(\tilde{\psi} = 1).$$

À taille de test fixée à $\alpha$, on essaie toujours de trouver le test $\hat{\psi}$ le plus puissant.

Une fois que l'on a contrôlé le risque de première espèce maximal par $\alpha$, on ne peut en règle générale pas contrôler le risque de seconde espèce uniformément. En effet, supposons que $\mathbb{P}_\theta(\psi = 1)$ soit une fonction continue de $\theta$. Et regardons la limite du risque de seconde espèce, lorsque $\theta \in \mathcal{T}_1$ tend à se rapprocher de $\mathcal{T}_0$. Dans ce cas, $\lim \theta \in \partial \mathcal{T}_0 \subset \mathcal{T}_0$. Et le risque de second espèce vérifie

$$\lim \beta(\theta) = \lim \mathbb{P}_\theta(\hat{\psi} = 0) = 1 - \lim \mathbb{P}_\theta(\hat{\psi} = 1) \ge 1 - \alpha.$$

On sait donc que, lorsque $\theta$ est dans $\mathcal{T}_1$, mais proche de $\mathcal{T}_0$, on aura tendance à décider en faveur de $(H_0)$ avec grande probabilité. La dissymétrie des deux hypothèses implique donc que **l'hypothèse nulle est celle à laquelle il faut croire, sauf si les données nous prouvent le contraire**.

# 3 Construction d'un test en pratique

La règle de décision $\hat{\psi}$ est construit à partir d'une première statistique $T = t(X)$, qui est constuite à partir d'un estimateur de la ou des coordonnées de $\theta$ qui nous intéresse, ou bien qui évalue la distance entre $\theta$ et $\mathcal{T}_0$.

**Définition 3.** La zone de rejet $\mathcal{R}$ est l'ensemble des valeurs de $T$ pour lesquelles on va rejetter $(H_0)$. Autrement dit, la règle de décision $\hat{\psi}$ est

$$\hat{\psi} = \mathbf{1}\{t(X) \in \mathcal{R}\}.$$

Pour constuire $\hat{\psi}$, on suit les étapes ci-dessous.

1. Modélisation du phénomène ayant engendré les données à l'aide d'un modèle paramétrique $(P_\theta, \theta \in \mathcal{T})$.
2. Détermination des hypothèses nulle $(H_0)$ et alternative $(H_1)$, en tenant compte de la dissymétrie de Neyman-Pearson.
3. Choix d'une statistique de test : $T = t(X)$ avec $t : \mathcal{X} \to \mathbb{R}$,
   (a) que le peut calculer sans connaître la vraie valeur de $\theta$,
   (b) dont le comportement sous $(H_0)$ et sous $(H_1)$ est différent,

   (c) dont la loi sous les pires cas est entièrement connue,
   (d) dont le comportement est connu lorsque l'on s'éloigne de $(H_0)$.
4. Détermination de la forme de la zone de rejet à l'aide du comportement de $T$ lorsque l'on s'éloigne de $(H_0)$.
5. Détermination des valeurs critiques (des seuils) de la zone de rejet à l'aide du comportement de $T$ sous le(s) pire(s) cas pour que le test soit de taille $\alpha$.

Une fois le test construit, et la conclusion tirée sur les données $x_{\text{obs}}$, il est important d'étudier la puissance du test, par exemple à l'aide d'une méthode de Monte-Carlo.

# 4 $p$-value

On suppose ici que l'on s'appuit sur une statistique de test $T = t(X)$ comme proposer ci-dessus, pour construire le test. Dans ce cas, on peut construire toute une famille de règles de décision : $(\hat{\psi}_\alpha, 0 < \alpha < 1)$, la règle de décision $\hat{\psi}_\alpha$ ayant la taille $\alpha$. Dans ce cas, pour tout $\alpha$,

$$\sup_{\theta \in \mathcal{T}_0} \mathbb{P}_\theta(\hat{\psi}_\alpha = 1) = \alpha.$$

On note $\hat{\psi}_\alpha = s_\alpha(X)$.

**Définition 4.** La $p$-value est la statistique définie comme la plus petite valeur de $\alpha$ pour laquelle on rejette $(H_0)$ à la vue des données $X$ :

$$p(X) = \inf\Big\{\alpha \ : \ s_\alpha(X) = 1\Big\}.$$

Par nature, la $p$-value est un nombre entre $0$ et $1$. Mais **ce N'est PAS la probabilité de** $(H_0)$ ! Une fois que l'on connait la $p$-value, il est facile de retrouver la décision pour la taille $\alpha$ :

$$\hat{\psi}_\alpha = \mathbf{1}\{p(X) < \alpha\}.$$

Autrement dit, on décide en faveur de $(H_1)$ si la $p$-value est inférieure à $\alpha$. La $p$-value est donc un outil auquel on recourt lorsque l'on veut reporter la décision (et donc le choix de la taille du test) le plus tardivement possible dans la procédure..

On peut montrer les résultats ci-dessous.

**Proposition 5.** — *Lorsque l'on s'éloigne de $(H_0)$, la $p$-value tend vers $0$.*

— *Lorsque $\theta \in \mathcal{T}_0$, c'est-à-dire sous $(H_0)$, on a, pour tout $u \in ]0;1[$,*

$$\mathbb{P}_\theta\Big(p(X) \le u\Big) \le u.$$

— *Si $\theta_0$ est un pire cas commun à toutes les règles de décision,*

$$\mathbb{P}_{\theta_0}\Big(p(X) \le u\Big) = u.$$

*Autrement dit, sous ce pire cas, la loi de $p(X)$ est la uniforme sur l'intervalle $[0;1]$.*

# 5 Approximation asymptotique

Dans de nombreux cas, on ne pas trouver de statistique de test ayant une loi connue exactement. Il est courant alors de s'appuyer sur une approximation asymptotique de la loi d'une statistique, lorsque le nombre d'observations $n$ du jeu de données est grand...