3. Théorie des tests

3.1. **Un exemple : les faiseurs de pluie (cf. poly).**

3.2. **Notions générales.** On dispose de la réalisation $(x_1, ..., x_n)$ d'un échantillon $(X_1, ..., X_n)$ d'une variable aléatoire X (réelle ou vectorielle). Un test statistique définit une règle de décision pour choisir entre deux hypothèses H_0 et H_1 faites sur la loi de X au vu de données recueillies. Les hypothèses H_0 et H_1 ne jouent pas le même rôle, l'hypothèse H_0 est celle à laquelle on tient le plus, qu'on ne veut rejeter qu'avec une faible probabilité de le faire à tort. De plus, pour pouvoir procéder à un test il faut impérativement être capable de faire des calculs sous l'hypothèse H_0, elle doit donc être suffisamment précise alors que l'hypothèse H_1 peut être relativement vague (la négation de H_0 par exemple). Bien sûr les hypothèses H_0 et H_1 doivent s'exclure mutuellement.

Construction et utilisation du test :

(1) On fixe $\alpha > 0$ petit (*risque de première espèce*), la probabilité de rejeter H_0 à tort.

(2) On détermine une *région de rejet de* H_0, $W \in \mathcal{B}(\mathbb{R}^n)$, telle que

$$\mathbf{P}\left[(X_1, ..., X_n) \in W | H_0\right] = \alpha.$$

Cette région dépend fortement des hypothèses que l'on considère. En particulier, elle dépend de H_1 en ce sens que l'on souhaite que la probabilité

$$1 - \beta = \mathbf{P}\left[(X_1, ..., X_n) \in W | H_1\right]$$

soit la plus grande possible. Le paramètre $1 - \beta$ (*puissance du test*) mesure la probabilité que les données soient dans la région de rejet de H_0 lorsque H_1 est vraie.

(3) Règle de décision : si la réalisation $(x_1, ..., x_n)$ de notre échantillon est dans W, on rejette H_0; sinon, on conserve H_0.

Finalement, construire un test, c'est se donner les hypothèses H_0 et H_1, le seuil de risque α petit, la région de rejet W de H_0 et, si on peut la calculer, la puissance du test $1 - \beta$.

Remarque. Le paramètre $\beta = \mathbf{P}\left[(X_1, ..., X_n) \in W^c | H_1\right]$ (*risque de seconde espèce*) est la probabilité de conserver H_0 alors que H_1 est vraie. Ce risque doit être aussi petit que possible à α fixé.

Remarque. Heuristiquement, il est assez facile de se convaincre que, lorsqu'on diminue α, on diminue la taille de la région de rejet W et donc on diminue également la puissance du test (ou on augment le risque de seconde espèce). Par conséquent, on ne peut choisir α trop petit. Les valeurs usuelles de α sont 0.1, 0.05, voire 0.01.

Qualité d'un test : Si $1 - \beta > \alpha$, on dit que le test est sans biais. Si $1 - \beta \to 1$ lorsque la taille de l'échantillon n tend vers l'infini, on dit que le test est convergent.

Classification des tests. On distingue les *tests paramétriques* (qui portent sur la valeur d'un ou plusieurs paramètres de la loi de X) des *tests non paramétriques*. Si un même test convient pour différentes lois, on dit que le test est *robuste* (comme les tests de moyenne, par exemple). Parmi les tests *non paramétriques* (qui sont robustes), on trouve les *tests d'ajustement* à une loi donnée. Enfin, il existe des tests de comparaison entre plusieurs échantillons qui permettent de déterminer si des échantillons sont issus d'une même population.

3.3. **Tests paramétriques.** On cherche à faire des tests sur certaines valeurs d'un paramètre θ de la loi d'une v.a. X. Pour cela, on dispose de la réalisation d'un échantillon $(X_1, ..., X_n)$ de la v.a. X. On note

$$((x_1, ..., x_n) \longmapsto L(x_1, ..., x_n; \theta))_{\theta \in \Theta}$$

la vraisemblance de l'échantillon.

Les hypothèses que l'on peut formuler sont de deux types :

- *hypothèse simple* : $[\theta = \theta_0]$ où θ_0 est une valeur fixée du paramètre;
- *hypothèse composite* : $[\theta \in A]$ où A est une partie de $\mathbb{R}$ non réduite à un point.

Ces notions peuvent être généralisées à un paramètre vectoriel. Noter que, lorsque le paramètre est réel, une hypothèse composite a souvent la forme $[\theta < \theta_0]$, $[\theta > \theta_0]$ ou $[\theta \neq \theta_0]$ pour une valeur fixée θ_0 du paramètre.

Remarque. Nous supposerons toujours que l'hypothèse H_0 est une hypothèse simple, pour pouvoir faire tous les calculs.

Test entre deux hypothèses simples : la méthode de Neyman et Pearson. On suppose

$$H_0 \ : \ [\theta = \theta_0], \quad H_1 \ : \ [\theta = \theta_1]$$

où θ_0 et θ_1 sont deux valeurs fixées du paramètre.

Théorème 3.1. *(Lemme de Neyman et Pearson)*. Soit $\alpha \in]0,1[$. La région de rejet optimale (celle qui maximise la puissance du test) est de la forme

$$W = \left\{ (x_1, ..., x_n) \in \mathbb{R}^n / \ L(x_1, ..., x_n; \theta_0) > 0 \text{ et } \frac{L(x_1, ..., x_n; \theta_1)}{L(x_1, ..., x_n; \theta_0)} \geq k \right\}$$

pour une constante k à déterminer en fonction de α.

Remarque. On ne peut pas toujours trouver k pour que l'égalité $\mathbf{P}\left[(X_1, ..., X_n) \in W | H_0\right] = \alpha$ soit satisfaite (en particulier pour une variable parente X discrète). Dans ce cas, on cherche le W telle que l'égalité ci-dessus soit approchée au mieux.

La *méthode de Neyman et Pearson* consiste à construire le test avec la région de rejet suggérée par le lemme du même nom. Les tests construits par cette méthode sont sans biais et convergents.

Test d'une hypothèse simple contre une hypothèse composite : la fonction puissance. On suppose

$$H_0 \ : \ [\theta = \theta_0], \quad H_1 \ : \ [\theta \in A]$$

où θ_0 est une valeur fixée du paramètre et A une partie de $\mathbb{R}$ ne contenant pas θ_0.

Même si l'on connaît la loi de la variable parente X, on ne peut calculer la puissance d'un test car H_1 n'est pas assez précise. Par contre, pour tout $\theta_1 \in A$, on peut calculer la puissance d'un test pour les hypothèses

$$H_0 \ : \ [\theta = \theta_0], \quad H_1 \ : \ [\theta = \theta_1].$$

On appelle alors *fonction puissance du test* la fonction, définie sur A, $\theta_1 \in A \longmapsto 1 - \beta(\theta_1)$. On recherche alors le test *uniformément le plus puissant* (UPP en abrégé), c'est-à-dire, s'il existe, celui tel que, pour tout $\theta_1 \in A$, sa puissance en θ_1 est supérieure à celle de tout autre test.

Lorsqu'un test construit par la méthode de Neyman et Pearson produit une région de rejet qui ne dépend pas explicitement de θ_1, on peut utiliser celle-ci pour le test entre une hypothèse simple et une hypothèse composite avec comme hypothèse $H_1 \ : \ [\theta > \theta_0]$ ou $[\theta < \theta_0]$. Dans ce cas, ce test est UPP.

Lorsque $H_1 \ : \ [\theta \neq \theta_0]$, on peut encore utiliser la méthode de Neyman et Pearson (lorsque le test correspondant entre deux hypothèses simples ne dépend pas explicitement de θ_1) de la façon suivante : on construit la région de rejet W_1 pour $H_1' \ : \ [\theta > \theta_0]$ et un risque de première espèce $\alpha/2$ et la région de rejet W_2 pour $H_1'' \ : \ [\theta < \theta_0]$ pour le même risque. Après avoir vérifié que $W_1 \cap W_2 = \emptyset$, on choisit finalement $W = W_1 \cup W_2$.