Statistique & Apprentissage

Paul-Henry Cournède

Amphi 5

III - Test d'hypothèses statistiques

III.1 - Un exemple introductif

Problème : Une entreprise pharmaceutique veut tester l'effet d'un nouveau médicament pour une maladie. Or pour cette maladie : effet placébo connu donne un taux de guérison de $\theta_0 = 0.2$.

Modèle : Soit X v.a. définie par :

$$X = \begin{cases} 1 \text{ si un patient traité par le nouveau médicament guérit} \\ 0 \text{ sinon.} \end{cases}$$

On suppose donc que $X \sim$ Bernoulli(θ), θ est donc le taux de guérison pour les patients sous traitement, paramètre inconnu.

Tests d'hypothèses statistiques : On formule un test d'hypothèses :

$$\begin{aligned} H_0 : \theta = \theta_0, &\text{ hypothèse nulle ou conservative} \\ H_1 : \theta > \theta_0, &\text{ hypothèse alternative} \end{aligned}$$

Expérimentation : On met en œuvre une expérimentation clinique pour sélectionner l'une ou l'autre des hypothèses.

- On traite N patients malades avec le médicament : correspond à un N-échantillon $(X_1, \ldots, X_N)$ de variables aléatoires i.i.d. selon Bernoulli(θ)
- On obtient N_g guérisons, $N_g = \sum_{i=1}^N X_i$.

Critère de Décision : Intuitivement, on va rejeter H_0 si N_g est assez grand, $N_g > N_0$. On parle de région de rejet. Comment définir ce seuil N_0 ? Deux erreurs possibles :

(i) erreur de type I : on rejette H_0 alors qu'elle est vraie

(ii) erreur de type II : on accepte H_0 alors que H_1 est vraie

Critère de Décision : Intuitivement, on va rejeter H_0 si N_g est assez grand, $N_g > N_0$. On parle de région de rejet. Comment définir ce seuil N_0 ? Deux erreurs possibles :
(i) erreur de type I : on rejette H_0 alors qu'elle est vraie
(ii) erreur de type II : on accepte H_0 alors que H_1 est vraie
$\Longrightarrow$ On fixe en général un risque de première espèce α, qui est le risque limite qu'on est prêt à accepter pour l'erreur de type I.
$\Longrightarrow$ On cherche donc N_0 tel que $\mathbb{P}(N_g > N_0) \leq \alpha$ quand H_0 est vraie.

Mise en œuvre du test : $\Longrightarrow$ N_g est une variable aléatoire binomiale, $N_g \sim B(N, \theta)$.
Sous H_0, $N_g \sim B(N, \theta_0)$.
On rappelle la définition du quantile d'ordre r : $q_r = \inf\{x \in \mathbb{R}, F(x) \geq r\}$. Si on prend $N_0 = q_{1-\alpha}$, quantile d'ordre $1-\alpha$ pour la loi Binomiale $B(N, \theta_0)$, on a bien :

$$\mathbb{P}(N_g > q_{1-\alpha}) \leq \alpha$$

Application Numérique : $N = 1000$, $\alpha = 0.05$ $\Longrightarrow$ $q_{0.95}^{B(1000,0.2)} = N_0 = 221$.

- Exemple 1 : essai clinique donne $N_g = 218$ $\Longrightarrow$ on ne peut pas rejeter H_0 : l'effet du nouveau médicament n'est pas concluant.

- Exemple 2 : essai clinique donne $N_g = 232$ $\Longrightarrow$ on peut rejeter H_0 et dire que l'effet du médicament est signficativement meilleur que l'effet placébo.

Remarque : Le niveau de risque 5% fixé de manière arbitraire. Jusqu'à où pouvait-on réduire le risque, et encore rejetter l'hypothèse conservative dans le cas 2 ?

$\Longrightarrow$ Le N_0 maximal pour lequel on rejette H_0 est donc $N_0 = 231$, soit un risque minimal associé (p-valeur) de $\mathbb{P}(N_g > 231) \approx 0.0071$.

$\Longrightarrow$ Niveau faible : on est alors très confiant quant à notre décision de rejeter H_0.

III.2 - Cadre général des tests statistiques

Soit un modèle statistique $\mathcal{M}$ sur $(\mathcal{X}, \mathcal{A})$, et soit $P^* \in \mathcal{M}$ loi de probabilité inconnue.

Définition : Une hypothèse statistique sur une loi $P^* \in \mathcal{M}$ est une affirmation du type $P^* \in \mathcal{M}_0$, où $\mathcal{M}_0 \subset \mathcal{M}$ est une sous-famille de lois de $\mathcal{M}$.

Exemples variés :

- $\mathcal{M}_0$ est un modèle paramétrique, $\mathcal{M}_0 = \left\{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\right\}$
- $\mathcal{M}_0$ est un singleton : $\mathcal{M}_0 = \{\mathcal{N}(0, 10)\}$

Objectif d'un test : Décider si une hypothèse statistique peut être considérée comme vraie à partir d'une expérience aléatoire associée à P^*.

$\Longrightarrow$ Nous considérons le cadre où on confronte deux hypothèses statistiques disjointes

Définition : Soient $\mathcal{M}_0 \subset \mathcal{M}$, $\mathcal{M}_1 \subset \mathcal{M}$, $\mathcal{M}_0 \cap \mathcal{M}_1 = \emptyset$.

Un test d'hypothèses statistiques est une procédure qui confronte deux hypothèses statistiques :

$$\begin{array}{lll} H_0 : P^* \in \mathcal{M}_0 & \Longrightarrow & \text{hypothèse nulle ou conservative} \\ H_1 : P^* \in \mathcal{M}_1 & \Longrightarrow & \text{hypothèse alternative} \end{array}$$

et définit une règle de décision permettant à partir de l'observation d'un N-échantillon i.i.d. pour P^* d'indiquer si H_0 peut être acceptée (supposée vraie) ou rejetée (supposée fausse) au profit de H_1.

La règle de décision est définie par une région de rejet ou région critique $\mathcal{R} \subset \mathcal{X}^N$: soit l'échantillon d'observations $(x_1, \dots, x_N) \in \mathcal{X}^N$, si $(x_1, \dots, x_N) \notin \mathcal{R}$ alors on accepte H_0, si $(x_1, \dots, x_N) \in \mathcal{R}$ alors on rejette H_0

Formellement, un test sur $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, est ainsi défini par le triplet $(\mathcal{M}_0, \mathcal{M}_1, \mathcal{R})$ où $\mathcal{M}_0, \mathcal{M}_1$ sont deux familles de lois de probabilité sur $(\mathcal{X}, \mathcal{A})$ disjointes et $\mathcal{R} \subset \mathcal{X}^N$ la région de rejet.

Remarques :

• La région de rejet $\mathcal{R}$ conduisant au rejet de l'hypothèse H_0 est construite à partir d'une statistique de test $T : \mathcal{X}^N \to \mathbb{R}$, dont on connait la loi, et la région de rejet sera de la forme :

$$\mathcal{R} = \left\{ (x_1, \dots, x_N) \in \mathcal{X}^N : T(x_1, \dots, x_N) \in \mathcal{W} \right\}, \text{ avec } \mathcal{W} \subset \mathbb{R} .$$

• Dissymétrie des hypothèses H_0 et H_1 : analogie avec la présomption d'innocence.

• Deux types d'erreur : l'erreur de première espèce, qui consiste à rejeter H_0 alors qu'elle est vraie (condamner un innocent), et l'erreur de deuxième espèce qui consiste à accepter H_0 alors que H_1 est vraie (relâcher un coupable).
$\Longrightarrow$ La région de rejet sera construite de façon à contrôler le risque lié à ces erreurs.

Définition : Soit $(\mathcal{M}_0, \mathcal{M}_1, \mathcal{R})$, un test d'hypothèse pour P^* sur $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$.

On appelle risque de première espèce du test la probabilité de rejeter H_0 alors qu'elle est vraie, c'est à dire $\mathbb{P}((X_1, \dots, X_N) \in \mathcal{R})$ alors que $P^* \in \mathcal{M}_0$.

On appelle risque de deuxième espèce du test la probabilité d'accepter H_0 alors que H_1 est vraie, c'est à dire $\mathbb{P}((X_1, \dots, X_N) \notin \mathcal{R})$ alors que $P^* \in \mathcal{M}_1$.

La puissance du test est alors (1 - le risque de deuxième espèce), c'est à dire la probabilité de rejeter H_0 alors que H_1 est vraie

Remarques :

• Plus la puissance du test est grande, plus celui-ci est capable de remettre en cause l'hypothèse conservative quand elle est fausse.

• En général P^* inconnu, donc on ne pourra pas calculer des risques, mais des majorants.

Définition : Pour un test $(\mathcal{M}_0, \mathcal{M}_1, \mathcal{R})$ sur $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, on appelle taille du test

$$\alpha = \sup_{P \in \mathcal{M}_0} P(\mathcal{R}) \ .$$

Tout majorant de la taille du test sera dit un niveau du test.

$\Longrightarrow$ La taille du test est donc un majorant du risque de première espèce.

2 stratégies :

• On fixe un risque de première espèce acceptable α (typiquement $\alpha = 0.01$, 0.05, 0.1) et on construit un test de niveau α $\Longrightarrow$ on construit $\mathcal{R}_\alpha$ telle que $\sup_{P \in \mathcal{M}_0} P(\mathcal{R}_\alpha) \leq \alpha$

• Pour un échantillon, on regarde quel était le risque d'obtenir une valeur aussi extrême que celle obtenue, c'est à dire la plus petite taille d'un test qui conduirait à rejeter H_0 $\Longrightarrow$ p-valeur.

Définition :Soit $\{(\mathcal{M}_0, \mathcal{M}_1, \mathcal{R}_t), t \in \mathcal{T}\}$ une collection de tests sur $\left(\mathcal{X}^N, \mathcal{A}^{\otimes N}\right)$ indicés par $t \in \mathcal{T} \subset \mathbb{R}$, et α_t la taille du test de région de rejet $\mathcal{R}_t$.
La p-valeur α^* est la statistique définie sur $\mathcal{X}^N$ par :

$$\alpha^*(X_1, \ldots, X_N) = \inf \{\alpha_t, t \in \mathcal{T} : (X_1, \ldots, X_N) \in \mathcal{R}_t\}$$

En pratique, on tire les conclusions décrites ci-dessous en fonction de la p-valeur :

p-valeur	évidence
$\alpha^*(x) < 0.01$	très forte évidence contre H_0
$0.01 \leq \alpha^*(x) < 0.05$	forte évidence contre H_0
$0.05 \leq \alpha^*(x) < 0.1$	faible évidence contre H_0
$0.1 \leq \alpha^*(x)$	aucune évidence contre H_0

Attention ! La p-valeur n'est pas la probabilité que H_0 soit vraie !
$\Longrightarrow$ H_0 est vraie avec une probabilité 1 ou 0

Différents types de test :

• $\mathcal{M}$ est un modèlé statistique paramétrique et test sur les paramètres de P^* $\Longrightarrow$ test paramétrique

• $\mathcal{M}_0$ est une famille de loi de probabilités $\Longrightarrow$ test d'adéquation ou d'ajustement : par ex., est-ce-que la loi de l'échantillon est Gaussienne ?

• 2 sous-échantillons $(x_1, \dots, x_N)$, $(y_1, \dots, y_M)$ sont-ils issus de la même population ? $\Longrightarrow$ test de comparaisons d'échantillons

III.3 - Tests paramétriques

Soit un modèle statistique paramétrique $\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$.

Définition : Soit $P_\theta \in \mathcal{M}_\Theta$ la loi inconnue. Un test paramétrique est un test de la forme :

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$$

où Θ_0 et Θ_1 sont deux sous-ensembles disjoints de Θ.

Terminologie :

Le test paramétrique d'hypothèses simples prend Θ_0 et Θ_1 comme singletons :

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

Tout autre test paramétrique est dit composite (ou d'hypothèses composites).

$$\begin{array}{lcll} H_0 : \theta \in \Theta_0 & \text{vs} & H_1 : \theta \in \Theta_1 = \Theta_0^c & \\ H_0 : \theta = \theta_0 & \text{vs} & H_1 : \theta \neq \theta_0 & \Longrightarrow \text{test bilatère} \\ H_0 : \theta = \theta_0 & \text{vs} & H_1 : \theta > \theta_0 & \Longrightarrow \text{test unilatère} \end{array}$$

III.3.a - Construction de la zone de rejet

Méthode directe :

Soit le test d'hypothèses : $H_0 : \theta \in \Theta_0 \quad$ vs $\quad H_1 : \theta \in \Theta_1$.
Soit α, un niveau de risque.

S'il existe une statistique T, telle que la loi de $T(X_1, \ldots, X_N)$ soit connue quand $(X_1, \ldots, X_N)$ échantillon i.i.d. pour P_θ.
$\Longrightarrow$ Alors on peut déterminer $\mathcal{W}_\alpha$ tel que :

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left(T(X_1, \ldots, T_N) \in \mathcal{W}_\alpha \right) \leq \alpha$$

et la région de rejet est donnée par :

$$\mathcal{R}_\alpha = \left\{ (x_1, \ldots, x_N) \in \mathcal{X}^N : T(x_1, \ldots, x_N) \in \mathcal{W}_\alpha \right\}$$

Exemple :

Pour la population française, le taux de glycémie à jeun noté suit une loi Normale de moyenne $\mu_0 = 4,8$ mmol/l et d'écart type $\sigma_0 = 0,4$ mmol/l.
$\Longrightarrow$ On veut tester si ce taux est différent dans une sous-population identifée : par exemple les adolescents de 14 à 16 ans, les habitants d'une région particulière, les sportifs, les fumeurs...

Modélisation : Soit X correspondant à la mesure du taux de glycémie à jeun pour un individu dans la sous-population d'intérêt. On suppose donc $X \sim \mathcal{N}(\mu, \sigma^2)$. Comme la population est plus homogène, on ne peut pas supposer que $\sigma = \sigma_0 \Longrightarrow \sigma$ inconnu.

Hypothèses : $H_0 : \mu = \mu_0 \quad$ vs $\quad H_1 : \mu \neq \mu_0$

Expérimentation : On réalise $N = 50$ mesures sur des individus.

Statistique de test pour construire la région de rejet ?

Exemple : On veut tester si le taux de glycémie est différent dans une sous-population identifée, et dans cette sous-population $X \sim \mathcal{N}(\mu, \sigma^2)$, σ inconnu.

Hypothèses : $H_0 : \mu = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$

Expérimentation : On réalise $N = 50$ mesures sur des individus « indépendants ».

Statistique de test :

Si $X \sim \mathcal{N}(\mu, \sigma^2) \Longrightarrow \dfrac{\sqrt{N}(\overline{X} - \mu)}{\sigma} \sim \mathcal{N}(0, 1)$ et $\dfrac{NS^2}{\sigma^2} \sim \chi^2(N-1)$, $\overline{X}$ et S^2 indépendantes.

$\Longrightarrow \dfrac{\sqrt{N-1}\,(\overline{X} - \mu)}{S} \sim \text{Student}(N-1).$

$\Longrightarrow$ sous H_0, $T(X_1, \ldots, X_N) = \dfrac{\sqrt{N-1}\,(\overline{X} - \mu_0)}{S} \sim \text{Student}(N-1).$

Région de rejet pour un niveau de risque fixé, $\alpha = 0.05$:

$$\mathbb{P}_{\mu_0}\left(-q_{1-\alpha/2}^{Student(N-1)} \leq T(X_1, \ldots, X_N) \leq q_{1-\alpha/2}^{Student(N-1)}\right) = 1 - \alpha, \text{ et donc :}$$

$$\mathcal{R}_\alpha = \left\{(x_1, \ldots, x_N) : \frac{\sqrt{N-1}\,(\overline{x} - \mu_0)}{s} < -q_{1-\alpha/2}^{Student(N-1)} \text{ ou } \frac{\sqrt{N-1}\,(\overline{x} - \mu_0)}{s} > q_{1-\alpha/2}^{Student(N-1)}\right\}$$

Région symétrique car test bilatère !

Application numérique : $q_{0.975}^{Student(49)} = 2.01$; sur notre échantillon : $\overline{X} = 4.6$ mmol/l, $s = 0.3$ mmol/l, d'où $T(x_1, \ldots, x_N) = -4.67$.

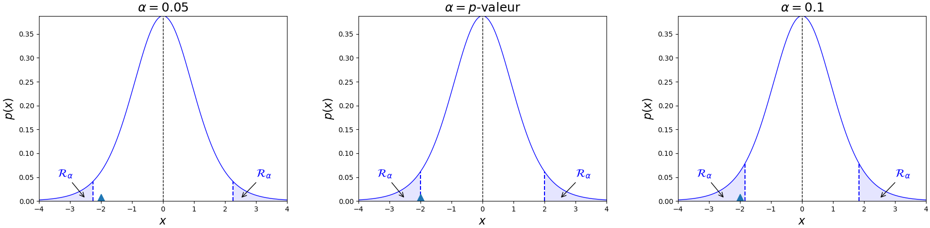
$\Longrightarrow$ On rejette H_0, la sous-population est significativement différente au niveau de risque 0.05.

$\Longrightarrow$ Calcul de la p-valeur : la borne inférieure du risque tel qu'on rejette H_0 est donc obtenu pour : $-q_{1-\alpha/2}^{Student(N-1)} = -4.67$, soit $1 - \alpha/2 = F(4.67)$, $\alpha = 2(1 - F(4.67)) = 0.000024$

$\Longrightarrow$ Très forte évidence contre H_0

Remarque : si seulement 10 mesures, $T(X_1, \ldots, X_10) = -2.0$ et $-q_{1-\alpha/2}^{Student(9)} \approx -2.26$

$\Longrightarrow$ ne permet pas de rejeter H_0 !



Méthode asymptotique : Test de Wald

Soit $X \sim P_\theta$, avec $\theta \in \mathbb{R}$, et le test sur un échantillon i.i.d. $(X_1, \ldots, X_N)$ des hypothèses :

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

On suppose que $\hat{\theta}_N = \hat{\theta}(X_1, \ldots, X_N)$ est un estimateur asymptotiquement normal de θ et $\tau^2(\theta)$ la variance asymptotique associée. Alors

$$\frac{\sqrt{N}(\hat{\theta}_N - \theta_0)}{\tau(\theta_0)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

$\Longrightarrow$ $\dfrac{\sqrt{N}(\hat{\theta}_N - \theta_0)}{\tau(\theta_0)}$ est dite statistique de Wald.

Soit α un niveau de risque donné : en notant q_r le quantile d'ordre r de la Gaussienne normalisée, $\forall \gamma \in [0;\alpha]$:

$$\mathcal{R}_{\alpha,\gamma} = \left\{ (x_1, \ldots, x_N) : \hat{\theta}(x_1, \ldots, x_N) < \theta_0 + \frac{q_\gamma \tau(\theta_0)}{\sqrt{N}} \text{ ou } \hat{\theta}(x_1, \ldots, x_N) > \theta_0 + \frac{q_{1-\alpha+\gamma}\tau(\theta_0)}{\sqrt{N}} \right\}$$

définit une région de rejet de taille asymptotique α.

En effet :

$$\begin{aligned}\mathbb{P}_{\theta_0}\Big((X_1, \ldots, X_N) \in \mathcal{R}_{\alpha,\gamma}\Big) &= \mathbb{P}_{\theta_0}\left(\tfrac{\sqrt{N}(\hat{\theta}_N - \theta_0)}{\tau(\theta_0)} < q_\gamma\right) + \mathbb{P}_{\theta_0}\left(\tfrac{\sqrt{N}(\hat{\theta}_N - \theta_0)}{\tau(\theta_0)} > q_{1-\alpha+\gamma}\right)\\ &\underset{N\to+\infty}{\longrightarrow} F^{\mathcal{N}(0,1)}(q_\gamma) + 1 - F^{\mathcal{N}(0,1)}(q_{\gamma+1-\alpha})\\ &= \gamma + 1 - \gamma - 1 + \alpha = \alpha\end{aligned}$$

$\Longrightarrow$ En pratique $\gamma = 0$ ou $\gamma = \alpha$ pour tests unilatères, $\gamma = \alpha/2$ pour tests unilatères.

Remarque : De façon générale, on peut construire des tests asymptotiques, c'est à dire des tests dont les régions de rejet vérifient les niveaux de risque asymptotiquement.

III.3.b - Puissance d'un test et Comparaison de tests

Définition : Soit un test paramétrique $(\Theta_0, \Theta_1, \mathcal{R})$ sur $\left(\mathcal{X}^N, \mathcal{A}^{\otimes N}\right)$. On définit π la fonction puissance du test, $\forall \theta \in \Theta_0 \cup \Theta_1$:

$$\pi(\theta) = \mathbb{P}_\theta\Big((X_1, \ldots, X_N) \in \mathcal{R}\Big),$$

Remarque : La taille du test α est alors donnée par : $\alpha = \sup_{\theta \in \Theta_0} \pi(\theta)$

Pour un test d'hypothèses simples, $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$, on obtient :

(i) $\alpha = \pi(\theta_0)$ est la taille du test et le risque de première espèce,

(ii) $\beta = 1 - \pi(\theta_1)$ est le risque de deuxième espèce.

(iii) $\pi(\theta_1)$ est la puissance du test.

Définition : Soient $\mathcal{T} = (\Theta_0, \Theta_1, \mathcal{R})$ et $\mathcal{T}' = (\Theta_0, \Theta_1, \mathcal{R}')$ deux tests paramétriques et de fonctions puissances respectives π et π'.

Si $\Theta_1 = \{\theta_1\}$: $\mathcal{T}$ est plus puissant que $\mathcal{T}'$ si $\pi(\theta_1) > \pi'(\theta_1)$

Si Θ_1 est quelconque : $\mathcal{T}$ est uniformément plus puissant que $\mathcal{T}'$ si $\pi(\theta) > \pi'(\theta)$, $\forall \theta \in \Theta_1$.

Exemple : Même exemple : taux de glycémie dans une sous-population, sauf qu'on suppose cette fois $\sigma = \sigma_0$ connu : $X \sim \mathcal{N}(\mu, \sigma_0^2)$.

Hypothèses : $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$

Expérimentation : On réalise $N = 50$ mesures sur des individus « indépendants ».

Statistique de test : Sous H_0 $\dfrac{\sqrt{N}(\overline{X} - \mu_0)}{\sigma_0} \sim \mathcal{N}(0,1)$ et

$$\mathcal{R}_\alpha = \left\{ (x_1, \ldots, x_N) : \frac{\sqrt{N}\,(\overline{x} - \mu_0)}{\sigma_0} < -q_{1-\alpha/2}^{\mathcal{N}(0,1)} \text{ ou } \frac{\sqrt{N}\,(\overline{x} - \mu_0)}{\sigma_0} > q_{1-\alpha/2}^{\mathcal{N}(0,1)} \right\}$$

Exemple : Même exemple : taux de glycémie dans une sous-population, sauf qu'on suppose cette fois $\sigma = \sigma_0$ connu : $X \sim \mathcal{N}(\mu, \sigma_0^2)$.

Hypothèses : $H_0 : \mu = \mu_0 \quad$ vs $\quad H_1 : \mu \neq \mu_0$

Expérimentation : On réalise $N = 50$ mesures sur des individus « indépendants ».

Statistique de test : Sous H_0 $\dfrac{\sqrt{N}(\overline{X} - \mu_0)}{\sigma_0} \sim \mathcal{N}(0,1)$ et

$$\mathcal{R}_\alpha = \left\{ (x_1, \ldots, x_N) : \frac{\sqrt{N}\,(\overline{x} - \mu_0)}{\sigma_0} < -q_{1-\alpha/2}^{\mathcal{N}(0,1)} \text{ ou } \frac{\sqrt{N}\,(\overline{x} - \mu_0)}{\sigma_0} > q_{1-\alpha/2}^{\mathcal{N}(0,1)} \right\}$$

Calcul de la Puissance : On considère désormais que $X \sim \mathcal{N}(\mu, \sigma_0^2)$, $\forall \mu$, et on calcule : $\mathbb{P}_\mu\Big((x_1, \ldots, x_N) \in \mathcal{R}_\alpha\Big)$.

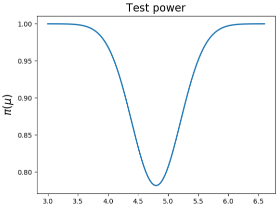
$$\text{On a : } \mathcal{R}_\alpha = \left\{ (x_1, \ldots, x_N) : \overline{x} < \mu_0 - \frac{\sigma_0\, q_{1-\alpha/2}^{\mathcal{N}(0,1)}}{\sqrt{N}} \text{ ou } \overline{x} > \mu_0 - \frac{\sigma_0\, q_{1-\alpha/2}^{\mathcal{N}(0,1)}}{\sqrt{N}} \right\}$$

Soit : pour $\mu_0 = 4.8$, $\sigma_0 = 0.4$, $N = 50$, $q_{0.975}^{\mathcal{N}(0,1)} = 1.96$

$$\mathcal{R}_\alpha = \Big\{ (x_1, \ldots, x_N) : \overline{x} < a_{min} \approx 4.69 \text{ ou } \overline{x} > a_{max} \approx 4.91 \Big\}$$

Si $X \sim \mathcal{N}(\mu, \sigma_0^2)$, $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma_0^2}{N})$, en notant $\Psi_\mu = F^{\mathcal{N}(\mu, \frac{\sigma_0^2}{N})}$:

Puissance du test $\pi(\mu) = \Psi_\mu(a_{min}) + 1 - \Psi_\mu(a_{max})$



III.3.c - Test du rapport de vraisemblance et méthode de Neyman-Pearson

On considère un test d'hypothèses simples sur le N-échantillon $(X_1, \ldots, X_N)$ i.i.d. pour la loi inconnue de densité p_θ :

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

Définition : On appelle test du rapport de vraisemblances de niveau α un test qui utilise la statistique de test :

$$\lambda(X_1, \ldots, X_N) = \frac{\mathcal{L}(\theta_1; X_1, \ldots, X_N)}{\mathcal{L}(\theta_0; X_1, \ldots, X_N)}$$

et définit la région critique par

$$\mathcal{R}^*_\alpha = \left\{(x_1, \ldots, x_N) \in \mathcal{X}^N \ , \ \lambda(x_1, \ldots, x_N) > c_\alpha\right\} \ ,$$

où c_α est choisi tel que $\mathbb{P}_{\theta_0}\Big((X_1, \ldots, X_N) \in \mathcal{R}^*_\alpha\Big) \leq \alpha$.

Lemme de Neyman-Pearson Le test du rapport de vraisemblance de taille α est le test le plus puissant parmi les tests de niveau α.

Exemple : On reprend l'exemple de la mise sur marché d'un médicament. Le laboratoire annonce un taux de guérison de $\theta = \theta_1 = 0.25$. L'Agence Nationale de Sécurité du Médicament doit donner son autorisation de mise sur le marché. Elle conduit ses propres essais cliniques indépendants sur N individus, et met en place le test :

Hypothèses : $H_0 : \theta = \theta_0 = 0.2 \quad \text{vs} \quad H_1 : \theta = \theta_1 = 0.25$

Recherche de la statistique de test par Neyman-Pearson : On note $N_g = \sum_{i=1}^N X_i$.

On sait que $\mathcal{L}(\theta; (X_1, \ldots, X_N)) = \prod_{i=1}^N \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{N_g}(1-\theta)^{N-N_g}$

$$\Longrightarrow \lambda(X_1, \ldots, X_N) = \left(\frac{\theta_1}{\theta_0}\right)^{N_g} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{N-N_g}$$

Exemple : On reprend l'exemple de la mise sur marché d'un médicament. Le laboratoire annonce un taux de guérison de $\theta = \theta_1 = 0.25$. L'Agence Nationale de Sécurité du Médicament doit donner son autorisation de mise sur le marché. Elle conduit ses propres essais cliniques indépendants sur N individus, et met en place le test :

Hypothèses : $H_0 : \theta = \theta_0 = 0.2 \quad \text{vs} \quad H_1 : \theta = \theta_1 = 0.25$

Recherche de la statistique de test par Neyman-Pearson : On note $N_g = \sum_{i=1}^N X_i$.

On sait que $\mathcal{L}(\theta; (X_1, \ldots, X_N)) = \prod_{i=1}^N \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{N_g}(1-\theta)^{N-N_g}$

$$\Longrightarrow \lambda(X_1, \ldots, X_N) = \left(\frac{\theta_1}{\theta_0}\right)^{N_g} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{N-N_g}$$

On sait que la zone de rejet est de la forme :

$$\mathcal{R}_\alpha^* = \left\{(x_1, \ldots, x_N) \in \mathcal{X}^N \ : \ \lambda(x_1, \ldots, x_N) > c_\alpha\right\}$$

$$\begin{aligned}
\lambda(x_1, \ldots, x_N) > c_\alpha &\Leftrightarrow \left(\frac{\theta_1}{\theta_0}\right)^{N_g} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{N-N_g} > c_\alpha \\
&\Leftrightarrow N_g \ln\left(\frac{\theta_1}{\theta_0}\right) + (N - N_g)\ln\left(\frac{1-\theta_1}{1-\theta_0}\right) > k_1 \\
&\Leftrightarrow N_g \ln\left(\frac{\theta_1}{\theta_0}\right) - N_g \ln\left(\frac{1-\theta_1}{1-\theta_0}\right) > k_2 \\
&\Leftrightarrow N_g \left[\ln\left(\frac{\theta_1}{\theta_0}\right) + \ln\left(\frac{1-\theta_0}{1-\theta_1}\right)\right] > k_3 \\
&\Leftrightarrow N_g > k_4, \ \text{comme } \theta_1 > \theta_0 \text{ et } 1-\theta_0 > 1-\theta_1
\end{aligned}$$

Et donc on prend la région de rejet de la forme :

$$\mathcal{R}_\alpha^* = \left\{(x_1, \ldots, x_N) \in \mathcal{X}^N \ : \ N_g > k_\alpha\right\}$$

III.3.d - Test du χ^2 de Pearson pour le modèle Multinomial

Définition : Soit $N \in \mathbb{N}^*$, $p \in]0;1[^k$ tel que $\sum_{i=1}^k p_i = 1$. On appelle loi multinomiale de paramètres (N, p), la loi de probabilité sur $\{0;1;\ldots;N\}^k$ définie par la fonction de masse :

$$P(x_1, \ldots, x_k) = \frac{N!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k p_i^{x_i},$$

pour tout $(x_1, \ldots, x_k) \in \{0;1;\ldots;N\}^k$ tel que $\Sigma_{i=1}^k x_j = N$. On note $X \sim M(N, p)$.

$\Longrightarrow$ On considère une urne avec des boules de k couleurs différentes en proportions $p = (p_1, \ldots, p_k)$. On effectue N tirages avec remise et on relève le nombre X_{i} de tirages de boules de chaque couleur dans le vecteur $X = (X_1, \ldots, X_k)$, on a $\Sigma_{i=1}^k X_{\mathrm{i}} = N$.

Définition Soit $X \sim M(N, p)$, $p \in]0;1[^k$, $\sum_{i=1}^k p_i = 1$.
On appelle statistique de Pearson la statistique T définie par

$$T(X) = \sum_{i=1}^k \frac{(X_i - Np_i)^2}{Np_i} .$$

$\Longrightarrow$ T est une mesure relative de l'écart entre entre les effectifs réalisés et théoriques.

Proposition : Soit $X \sim M(N, p)$, $p \in]0;1[^k$, $\sum_{i=1}^k p_i = 1$ et $T(X)$ la statistique de Pearson.
Alors, la loi limite de $T(X)$ est une loi du chi-deux à $k-1$ degrés de liberté :

$$T(X) = \sum_{i=1}^k \frac{(X_i - Np_i)^2}{Np_i} \xrightarrow{\mathcal{L}} \chi^2(k-1)$$

Corollaire : Test de Pearson pour le modèle multinomial Soit $X \sim M(N,p)$, N connu, p inconnu. Soit p_0 donné, $p_0 \in]0;1[^k$, $\sum_{i=1}^k p_{0i} = 1$. On considère le test :

$$H_0 : p = p_0 \quad \text{contre} \quad H_1 : p \neq p_0$$

Alors, sous H_0 : $T(X) = \sum_{i=1}^{k} \frac{(X_i - Np_{0i})^2}{Np_{0i}} \xrightarrow{\mathcal{L}} \chi^2(k-1)$.

Pour un niveau α donnée, on en déduit la région asymptotique de rejet :

$$\mathcal{R}_\alpha = \left\{ x : T(x) > q_{1-\alpha}^{\chi^2(k-1)} \right\}$$

avec $q_{1-\alpha}^{\chi^2(k-1)}$ le quantile d'ordre $(1-\alpha)$ pour $\chi^2(k-1)$.

III.4 - Tests d'ajustement ou Tests d'adéquation

Soit $(X_1, \ldots, X_N)$, un échantillon i.i.d de loi P^* inconnue : nous souhaitons vérifier si P^* appartient à une famille paramétrique particulière $\mathcal{M}_\Theta = \{P_\theta, \theta \in \Theta\}$:

$$H_0 : \; P^* \in \mathcal{M}_\Theta \quad \text{contre} \quad H_1 : \; P^* \notin \mathcal{M}_\Theta$$

$\Longrightarrow$ phase préliminaire d'estimation : pour un échantillon donné $(x_1, \ldots, x_N)$, parmi tous les $P_\theta \in \mathcal{P}_\Theta$, on choisit celui qui maximise $P_\theta(x_1, \ldots, x_N)$, donc en prenant $P_{\hat{\theta}(x_1,\ldots,x_N)}$, avec $\hat{\theta}(x_1, \ldots, x_N)$ l'estimation du maximum de vraisemblance pour θ, et on se ramène au test :

$$H_0 : \; P^* = P_{\theta_0} \quad \text{contre} \quad H_1 : \; P^* \neq P_{\theta_0}$$

III.4.a -Test du χ^2

$\Longrightarrow$ Le test d'ajustement du χ^2 adapte le test défini pour la loi multinomiale.

III.4.a -Test du χ^2

$\Longrightarrow$ Le test d'ajustement du χ^2 adapte le test défini pour la loi multinomiale.

Soit $\mathcal{M}_\Theta$ un modèle paramétrique, et $\theta_0 \in \Theta$, le test du χ^2 permet de tester :

$$H_0 : \ P^* = P_{\theta_0} \ \text{ contre } \ H_1 : \ P^* \neq P_{\theta_0}$$

Définition : Soit $(X_1, \ldots, X_N)$ un échantillon i.i.d. de loi P^*.
Soit $(I_1, \ldots, I_k)$, une partition du support de P^* telle que $\forall j, 1 \leq j \leq k$, $P^*(I_j) > 0$. Soit $p_j = P^*(I_j)$ et $p = (p_1, \ldots, p_k)$.
Soit $Y = (Y_1, \ldots, Y_k)$, la statistique à valeurs dans $\{0; 1; \ldots; N\}^k$ qui compte le nombre de X_i dans tous les I_j, $1 \leq j \leq k$: $Y_j(X_1, \ldots, X_N) = \sum_{i=1}^{N} \mathbb{I}_{I_j}(X_i)$.

Alors $Y \sim M(N, p)$ et la statistique de Pearson généralisée est donnée par :

$$\tilde{T}(X_1, \ldots, X_N) := T(Y(X_1, \ldots, X_N)) = \sum_{i=1}^{k} \frac{(Y_i - Np_i)^2}{Np_i}$$

Test du χ^2 : Sous l'hypothèse H_0, $Y \sim M(N, p_0)$ avec $p_{0j} = P_{\theta_0}(X_1 \in I_j)$, $\forall 1 \leq j \leq k$. Et donc,

$$T(Y(X_1, \ldots, X_N)) = \sum_{i=1}^{k} \frac{(Y_i - Np_{0i})^2}{Np_{0i}} \xrightarrow{\mathcal{L}} \chi^2(k-1) \ .$$

Pour un niveau de test donné α, on forme la région de rejet asymptotique :

$$\mathcal{R}_\alpha = \left\{ (x_1, \ldots, x_N) : T(Y(x_1, \ldots, x_N)) > q_{1-\alpha}^{\chi^2(k-1)} \right\} ,$$

avec $q_{1-\alpha}^{\chi^2(k-1)}$ le quantile d'ordre $(1-\alpha)$ du $\chi^2(k-1)$.

Test du χ^2 : Sous l'hypothèse H_0, $Y \sim M(N, p_0)$ avec $p_{0j} = P_{\theta_0}(X_1 \in I_j)$, $\forall 1 \leq j \leq k$. Et donc,

$$T(Y(X_1, \ldots, X_N)) = \sum_{i=1}^{k} \frac{(Y_i - Np_{0i})^2}{Np_{0i}} \xrightarrow{\mathcal{L}} \chi^2(k-1) \ .$$

Pour un niveau de test donné α, on forme la région de rejet asymptotique :

$$\mathcal{R}_\alpha = \left\{ (x_1, \ldots, x_N) : T(Y(x_1, \ldots, x_N)) > q_{1-\alpha}^{\chi^2(k-1)} \right\} \ ,$$

avec $q_{1-\alpha}^{\chi^2(k-1)}$ le quantile d'ordre $(1-\alpha)$ du $\chi^2(k-1)$.

Remarques :

• Le test du χ^2 est souvent précédé d'une phase d'estimation de θ_0. Si Θ est de dimension q, q degrés de liberté sont ainsi perdus, et nous avons :

$$T(Y(X_1, \ldots, X_N)) = \sum_{i=1}^{k} \frac{(Y_i - Np_{0i})^2}{Np_{0i}} \xrightarrow{\mathcal{L}} \chi^2(k-1-q) \ .$$

• On considère que l'approximation asymptotique est valide dés que $Np_{0i} \geq 5$ pour toutes les classes. Sinon on regroupe certaines classes.

• Autres tests d'ajustement : Kolmogorov-Smirnov et Cramer Von-Mises (sur les écarts entre fonction de répartition théorique et fonction de répartition empirique).