Chap 3 : Théorie des tests

Organisation du chapitre 3

1. Tests statistiques
   - Un exemple d'introduction
   - Tests paramétriques
   - Test d'adéquation
   - Test d'indépendance
   - Test d'ajustement de Kolmogorov

1. Tests statistiques

1.1. Un exemple d'introduction

Exemple d'introduction : les faiseurs de pluie

- Des relevés (sur longue période) : le niveau naturel des pluies dans la Beauce par an (en mm) suit une loi $\mathcal{N}(\mu = 600, \sigma = 100)$.

- Des entrepreneurs prétendent pouvoir augmenter le niveau moyen de pluie (insémination des nuages-iodure d'argent). Essai entre 1951 et 1959 :

Année	1951	1952	1953	1954	1955	1956	1957	1958	1959
mm	510	614	780	512	501	534	603	788	650

- Deux hypothèses s'affrontent :

  - L'insémination était sans effet

  - Elle augmentait le niveau moyen de pluie de 50 mm.

othèses Si m désigne l'espérance mathématique de X v.a. égale au niveau annuel :

$$\begin{cases} H_0 & : \quad m = 600 \ mm \\ H_1 & : \quad m = 650 \ mm \end{cases}$$

1. Pour les agriculteurs : le coût ! Donc les faits observés doivent contredire nettement la validité de H_0 (l'hypothèse nulle).

Exemple d'introduction

veau α Niveau de probabilité ou ils sont prets à accepter H_1, i.e. le résultat obtenu faisait partie d'une éventualité de 5 sur 100 de se produire.

Ils prennent donc un risque de 5% de se tromper (évènements rares arrivent quand mÍme).

Test Tester la valeur m, on considère $\bar{X}$ (moyenne des observations) qui est donc la variable de décision

Si H_0 est vraie, $\bar{X}$ suit $\mathscr{N}(600, \frac{100}{\sqrt{9}})$

Si $\bar{x}$ est trop grand, i.e. $\bar{x} \geq k$, avec k tel que $P(Y \geq k) = \alpha$ oô $Y \sim \mathscr{N}(600, \frac{100}{\sqrt{9}})$. donc

$$k = 600 + \frac{100}{3} 1.64 = 655$$

écision Règle de décision :

- Si $\bar{X} > 655$, on rejette H_0 et accepter H_1
- Si $\bar{X} < 655$, on conserve H_0

Exemple d'introduction : les types d'erreur

Supposons que les faiseurs de pluie ont raison, alors $\bar{X} \sim \mathcal{N}(650, \frac{100}{3})$.

On commet une erreur à chaque fois que $\bar{x} < 655$, donc avec probabilité :

$$\beta := P\left(U < \frac{655 - 650}{100/3}\right) = P(U < 0.15) = 0.56$$

$\alpha = P[\text{accepter } H_1 \,; \text{alors que } H_0 \text{ vraie}]$: erreur de première espèce !

$\beta = P[\text{accepter } H_0 \,; \text{alors que } H_1 \text{ vraie}]$: erreur de deuxième espèce !

Décision\Vérité	H_0	H_1
H_0	$1-\alpha$	β
H_1	α	$1-\beta$

Remarque. Les deux hypothèses ne jouent pas de rôles symétriques : - k déterminé par H_0 et α - β est déterminé par la considération supplémentaire de H_1.

1.2. Tests paramétriques

Les Différents Tests paramétriques

Définition 4.1

Soit $X_1, ..., X_n$ un échantillon dont la loi est dans le modèle $(\Omega, \mathscr{A}, \{P_\theta\}_{\theta\in\Theta})$. On suppose Θ partitionné en Θ_0 et Θ_1 et on associe les hypothèses suivantes :

$$\left\{ \begin{array}{l} H_0 \,:\, \theta \in \Theta_0 \\ H_1 \,:\, \theta \in \Theta_1 \end{array} \right.$$

H_0 s'appelle l'hypothèse nulle et H_1 l'hypothèse alternative.

Test Simple $\left\{ \begin{array}{l} H_0 \,:\, \theta = \theta_0 \\ H_1 \,:\, \theta = \theta_1 \end{array} \right.$; Test unilateral $\left\{ \begin{array}{l} H_0 \,:\, \theta = \theta_0 \\ H_1 \,:\, \theta > \theta_0 \end{array} \right.$ ou

$\left\{ \begin{array}{l} H_0 \,:\, \theta = \theta_0 \\ H_1 \,:\, \theta < \theta_0 \end{array} \right.$; Test composite $\left\{ \begin{array}{l} H_0 \,:\, \theta \in [\theta_0, \theta_1] \\ H_1 \,:\, \theta \notin [\theta_0, \theta_1] \end{array} \right.$ ou

$\left\{ \begin{array}{l} H_0 \,:\, \theta \notin [\theta_0, \theta_1] \\ H_1 \,:\, \theta \in [\theta_0, \theta_1] \end{array} \right.$; Test bilateral $\left\{ \begin{array}{l} H_0 \,:\, \theta = \theta_0 \\ H_1 \,:\, \theta \neq \theta_0 \end{array} \right.$

Tests paramétriques

Définition 4.2

- Erreur de première espèce, notée α, est la probabilité de rejeter à tort l'hypothèse nulle, i.e.

$$\alpha = P(\text{choisir } H_1 | H_0 \text{ est vraie})$$

- Erreur de deuxième espèce, notée β, est la probabilité de conserver à tort l'hypothèse nulle, i.e.

$$\beta = P(\text{choisir } H_0 | H_1 \text{ est vraie})$$

- La puissance d'un test, notée $\mathscr{P}$, est la probabilité de rejeter H_0 lorsque H_1 est vraie.

Définition 4.3

La région de Rejet d'un test, $\mathscr{W}$, est l'ensemble des valeurs de la statistique de test qui conduisent à rejeter H_0 au profit de H_1.

Construction d'un test paramétrique

On a donc :

$$\alpha = P_{H_0}(\mathscr{W}), 1-\alpha = P_{H_0}(\mathscr{W}^c), \mathscr{P} = 1-\beta = P_{H_1}(\mathscr{W}).$$

Résumé :

Décision\Vérité	H_0	H_1
H_0	$1-\alpha$	β
H_1	α	$1-\beta$

Les démarches de la construction d'un test :

1- choix de H_0 et H_1

2- détermination de la statistique de test

3- forme de la région de rejet

4- détermination de la loi de la statistique de test sous H_0 et le calcul de la région de rejet en fonction de α

5- calcul de la puissance

6- calcul de la valeur expérimental de test

Le test simple : Méthode de Neyman-Pearson

Il s'agit de maximiser la puissance $(1-\beta)$ du test pour une valeur donnée de α risque de première espèce.

Cela revient à choisir la région critique optimale, i.e. un domaine de $\mathbb{R}^n$ parmi l'ensemble de toutes les réalisations possibles de l'échantillon $(X_1, \ldots, x_n)$ dont la forme définira ensuite une variable statistique.

Théorème 4.1 (Neyman Pearson)

La région critique optimale W est définie par l'ensemble des points de $\mathbb{R}^n$ tels que :

$$\frac{L(x;\theta_1)}{L(x;\theta_0)} > k_\alpha$$

ou k_α est une constante telle que $P_{H_0}(W) = \alpha$. Alors cette région W réalise le maximum de $1-\beta$.

Test Simple sur un paramètre : moyenne d'une loi normale

connu Echantillonnage gaussien $\mathcal{N}(\mu; \sigma_0^2); \mu \in \mathbb{R})$, variance connue.

Step 1 On choisit les hypothèses

$$\text{Test simple : } \begin{cases} H_0 \; : \; \mu = \mu_0 \\ H_1 \; : \; \mu = \mu_1 \end{cases}$$

avec $\mu_1 > \mu_0$.

tep 2-3 Le rapport de vraisemblance est :

$$\begin{aligned} Z_n = \frac{L(X;\mu_1)}{L(X;\mu_0)} &= \exp\left(-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^n (X_i-\mu_1)^2 - \sum_{i=1}^n (X_i-\mu_0)^2\right]\right) \\ &= \exp\left(\frac{1}{\sigma_0^2}(\mu_1-\mu_0)\sum_{i=1}^n X_i\right)\exp\left(-\frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2)\right). \end{aligned}$$

Z_n est une variable aléatoire continue sous $\mathbb{P}_{\mu_0}$. La région critique optimale au seuil α est

$$W = \left\{(x_1,\dots,x_n);\ \exp\left(-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^n (x_i-\mu_1)^2 - \sum_{i=1}^n (x_i-\mu_0)^2\right]\right) > k\right\}$$

Tests portant sur un paramètre : moyenne d'une loi normale

$$
\begin{aligned}
W &= \left\{(x_1,\ldots,x_n);\ \exp\left(\frac{1}{\sigma_0^2}(\mu_1-\mu_0)\sum_{i=1}^n x_i\right)\exp\left(-\frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2)\right) > k\right\} \\
&= \left\{(x_1,\ldots,x_n);\ \frac{1}{n}\sum_{i=1}^n x_i > c\right\}.
\end{aligned}
$$

On a donc : - statistique de test : $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ - région critique (de rejet) : $W = \{\bar{X}_n > K_\alpha\}$.

Step 4 Détermination de K_α : sous l'hypothèse H_0, $\bar{X}_n$ suit une loi $\mathcal{N}(\mu_0, \frac{\sigma_0^2}{n})$. Donc :

$$
\begin{aligned}
\mathbb{P}_{H_0}(W) &= \mathbb{P}_{H_0}(\frac{1}{n}\sum_{i=1}^n X_i > K_\alpha) = \mathbb{P}_{H_0}(\frac{\sqrt{n}(\bar{X}_n-\mu_0)}{\sigma_0} > \frac{\sqrt{n}(K_\alpha-\mu_0)}{\sigma_0}) \\
&= 1-\phi(\frac{\sqrt{n}(K_\alpha-\mu_0)}{\sigma_0}) = \alpha.
\end{aligned}
$$

oô ϕ est la fonction de répartition de la gaussienne centrée et réduite.

d'oô $K_\alpha = \mu_0 + \frac{\sigma_0}{\sqrt{n}}\phi^{-1}(1-\alpha)$.

Tests portant sur un paramètre : moyenne d'une loi normale

On rejette donc l'hypothèse H_0 si :

$$\bar{x}_n > \mu_0 + \frac{\sigma_0}{\sqrt{n}} \phi^{-1}(1-\alpha)$$

Step 5 **Puissance :** sous l'hypothèse H_1, $\bar{X}_n$ suit une loi $\mathcal{N}(\mu_1, \frac{\sigma_0^2}{n})$, la puissance du test est définie par :

$$\mathscr{P} = \mathbb{P}_{H_1}(W) \quad = \quad \mathbb{P}_{H_1}(\bar{X}_n > K_\alpha) = 1 - \phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma_0} + \phi^{-1}(1-\alpha)\right).$$

Tests portant sur un paramètre : moyenne d'une loi normale

Test Unilatéral

Step 1 On choisit les hypothèses

$$\text{Test unilatéral : } \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

tep 2-3 On a toujours - statistique de test : $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ - région critique (de rejet) : $W = \{\bar{X}_n > K_\alpha\}$.

Step 4 On rejette donc l'hypothèse H_0 si :

$$\bar{x}_n > \mu_0 + \frac{\sigma_0}{\sqrt{n}}\phi^{-1}(1-\alpha)$$

Step 5 Pour tout $\mu > \mu_0$, la puissance de test est définie donc par :

$$\mathscr{P}(\mu) = \mathbb{P}_{H_1}(W) \quad = \quad \mathbb{P}_{H_1}(\bar{X}_n > K_\alpha) = 1 - \phi\left(\frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} + \phi^{-1}(1-\alpha)\right).$$

Test Unilatéral

Step 1 On choisit les hypothèses

$$\text{Test unilatéral : } \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

Step 2-3 On a - statistique de test : $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ - région critique (de rejet) :

$W = \{\bar{X}_n < K_\alpha\}$.

Step 4 Détermination de K_α : sous l'hypothèse H_0, $\bar{X}_n$ suit une loi $\mathcal{N}(\mu_0, \frac{\sigma_0^2}{n})$. Il vient donc :

$$\begin{aligned} \mathbb{P}_{H_0}(W) &= \mathbb{P}_{H_0}(\bar{X}_n > K_\alpha) = \mathbb{P}_{H_0}\left(\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma_0} < \frac{\sqrt{n}(K_\alpha - \mu_0)}{\sigma_0}\right) \\ &= \phi\left(\frac{\sqrt{n}(K_\alpha - \mu_0)}{\sigma_0}\right) = \alpha. \end{aligned}$$

d'oô $K_\alpha = \mu_0 + \frac{\sigma_0}{\sqrt{n}}\phi^{-1}(\alpha)$.

Tests portant sur un paramètre : moyenne d'une loi normale

Step 5 **Puissance :** sous l'hypothèse H_1, Pour tout $\mu < \mu_0$, $\bar{X}_n$ suit une loi $\mathcal{N}(\mu, \frac{\sigma_0^2}{n})$, la puissance du test est définie par :

$$\mathscr{P}(\mu) = \mathbb{P}_{H_1}(W) = \mathbb{P}_{H_1}(\bar{X}_n < K_\alpha) = 1 - \phi\left(\frac{\sqrt{n}(\mu_0 - \mu)}{\sigma_0} + \phi^{-1}(\alpha)\right).$$

Test Bilatéral

Step 1 On choisit les hypothèses

$$\text{Test bilatéral : } \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

ep 2-3 On a - statistique de test : $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ - région critique (de rejet) : $W = \{|\bar{X}_n - \mu_0| > K_\alpha\}$.

Step 4 Détermination de K_α : sous l'hypothèse H_0, $\bar{X}_n$ suit une loi $\mathcal{N}(\mu_0, \frac{\sigma_0^2}{n})$.

$$\begin{aligned} \mathbb{P}_{H_0}(W) &= \mathbb{P}_{H_0}(|\bar{X}_n - \mu_0| > K_\alpha) = 1 - \mathbb{P}_{H_0}(\frac{\sqrt{n}K_\alpha}{\sigma_0} \leq \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma_0} \leq \frac{\sqrt{n}K_\alpha}{\sigma_0} \\ &= 2(1 - \phi(\frac{\sqrt{n}K_\alpha}{\sigma_0}) = \alpha. \end{aligned}$$

Tests portant sur un paramètre : moyenne d'une loi normale

d'ou $K_\alpha = \frac{\sigma_0}{\sqrt{n}}\phi^{-1}(1-\frac{\alpha}{2})$.

On rejette donc l'hypothèse H_0 si :

$$|\bar{x}_n - \mu_0| > \frac{\sigma_0}{\sqrt{n}}\phi^{-1}(1-\frac{\alpha}{2})$$

Step 5 **Puissance :** sous l'hypothèse H_1, Pour tout $\mu \neq \mu_0$, $\bar{X}_n$ suit une loi $\mathcal{N}(\mu, \frac{\sigma_0^2}{n})$, la puissance du test est définie par :

$$\mathscr{P}(\mu) = 1 - \phi\left(\frac{\sqrt{n}(\mu_0-\mu)}{\sigma_0} + \phi^{-1}(1-\frac{\alpha}{2})\right) + \phi\left(\frac{\sqrt{n}(\mu_0-\mu)}{\sigma_0} - \phi^{-1}(1-\frac{\alpha}{2})\right)$$

Tests simple sur un paramètre : moyenne d'une loi normale

nconnu On considère un modèle d'échantillonnage gaussien $\mathscr{P}\mathscr{N}(\mu;\sigma^2);^{\mu}\in\mathbb{R})$ à variance inconnue. On rappelle que S_n^2 est un estimateur sans biais de σ^2.

Step 1 On choisit les hypothèses

$$\text{Test simple : } \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases}$$

avec $\mu_1 > \mu_0$.

tep 2-3 On a - statistique de test : $\Lambda_n = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{S_n}$ - région critique (de rejet) : $W = \{\Lambda_n > K_\alpha\}$.

Step 4 Détermination de K_α : sous l'hypothèse H_0, Λ_n suit une loi T_{n-1} (loi de student à $n-1$ degrés de liberté). Il vient donc :

$$\mathbb{P}_{H_0}(W) = \mathbb{P}_{H_0}(\Lambda_n > K_\alpha) = 1 - F_{T_{n-1}}(K_\alpha) = \alpha.$$

oô $F_{T_{n-1}}$ désigne la fonction de répartition de T_{n-1}.

on a donc $K_\alpha = F^{-1}_{T_{n-1}}(1-\alpha)$. On rejette donc l'hypothèse H_0 si :

$$\bar{x}_n > \mu_0 + \frac{s_n}{\sqrt{n}} F^{-1}_{T_{n-1}}(1-\alpha)$$

Tests portant sur un paramètre : moyenne d'une loi normale

Step 5 **Puissance :** sous l'hypothèse H_1, Λ_n suit une loi T_{n-1}, la puissance du test est définie par :

$$\mathscr{P} = \mathbb{P}_{H_1}(W) \quad = \quad \mathbb{P}_{H_1}(\Lambda_n > K_\alpha) = 1 - F_{T_{n-1}}\left(\frac{\sqrt{n}(\mu_0 - \mu)}{s_n} + F^{-1}_{T_{n-1}}(1 -$$

Pour les autres tests : test unilatéral et test bilatéral, la méthode reste la mÍme.

Tests simple sur un paramètre : variance d'une loi normale

connu Soit $X_1, \dots, X_n$ un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$ avec μ connu.

$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ est un estimateur sans biais de σ^2.

Step 1 On choisit les hypothèses

$$\text{Test simple : } \begin{cases} H_0 : \sigma = \sigma_0 \\ H_1 : \sigma = \sigma_1 \end{cases}$$

avec $\sigma_1 > \sigma_0$.

ep 2-3 On a - statistique de test : $\Lambda_n = \sum_{i=1}^{n} (X_i - \mu)^2$ - région critique (de rejet) : $W = \{\Lambda_n > K_\alpha\}$.

Step 4 Détermination de K_α : sous l'hypothèse H_0, $\frac{\Lambda_n}{\sigma_0^2}$ suit une loi χ_n^2 :

$$\mathbb{P}_{H_0}(W) \;=\; \mathbb{P}_{H_0}(\Lambda_n > K_\alpha) = \mathbb{P}_{H_0}\left(\frac{\Lambda_n}{\sigma_0^2} > \frac{K_\alpha}{\sigma_0^2}\right) = 1 - F_{\chi_n^2}\left(\frac{K_\alpha}{\sigma_0^2}\right) = \alpha.$$

oô $F_{\chi^2_n}$ désigne la fonction de répartition de la loi χ^2_n, on a donc $K_\alpha = \sigma_0^2 F^{-1}_{\chi^2_n}(1-\alpha)$. On rejette donc l'hypothèse H_0 si :

$$\sum_{i=1}^{n}(x_i-\mu)^2 > \sigma_0^2 F^{-1}_{\chi^2_n}(1-\alpha)$$

Step 5 **Puissance :** sous l'hypothèse H_1, $\frac{\Lambda_n}{\sigma_1^2}$ suit une loi χ^2_n, la puissance du test est définie par :

$$\mathscr{P} = \mathbb{P}_{H_1}(W) \ = \ \mathbb{P}_{H_1}(\Lambda_n > K_\alpha) = \mathbb{P}_{H_1}(\frac{\Lambda_n}{\sigma_1^2} > \frac{K_\alpha}{\sigma_1^2}) = 1 - F_{\chi^2_n}\left(\frac{\sigma_0^2}{\sigma_1^2}F^{-1}_{\chi^2_n}(\right.$$

MÍme méthode pour les autres tests.

connu On utilise la statistique de test $\Lambda_n = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

1.3. Test d'adéquation

Test d'adéquation : test de χ^2

- On considère un échantillon $X_1, ..., X_n$ et P_θ une loi donnée oô le paramètre $\theta \in \Theta \subset \mathbb{R}^d$.

- L'objectif du test d'adéquation (ou test du χ^2) est répondre à la question : les observations suivent-elles bien la loi P_θ ?

- Le problème de test à étudier est le suivant :

$$\begin{cases} H_0 : \ X_1, ..., X_n \text{ suivent la loi } \ P_\theta \\ H_1 : \ X_1, ..., X_n \text{ ne suivent pas la loi } \ P_\theta \end{cases}$$

Test d'adéquation : test de χ^2 - construction du test

On répartit les n observations de la v.a. X en k classes : $[e_{i-1}, e_i[_{1\leq i\leq k}$ d'effectifs aléatoires N_i. On calcule :

$$p_i = P_{H_0}(X \in [e_{i-1}, e_i[)$$

$$D_n = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} = n\sum_{i=1}^{k} \frac{(\hat{p}_i - p_i)^2}{p_i}$$

arque - np_i représente l'effectif théorique - $\hat{p}_i$ représente la probabilité estimée : $\hat{p}_i = \frac{N_i}{n}$. - D représente une distance entre la loi théorique P_θ et la loi observée. - On aurait également pu définir D_n par

$$D_n = n\sum_{i=1}^{k} \frac{(\hat{p}_i - p_i)^2}{\hat{p}_i}$$

Dans le cas oô le paramètre θ est connu, on a le résultat suivant

orème Si $n \to \infty$, alors D_n est asymptotiquement distribué comme une variable de χ^2_{k-1} et ceci quelle que soit la loi de X.

Test d'adéquation : test de χ^2 - construction du test

narque On assimile $D_n \sim \chi^2_{k-1}$ si $np_i > 5$ pour toute classe. Dans le cas contraire, on procède à des regroupements. D'oô le test χ^2.

n rejet On rejettera H_0 si la valeur d_n constaté $\sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = n \sum_{i=1}^{k} \frac{(\hat{p}_i - p_i)^2}{p_i}$ est trop grand, c-à-d si

$$d_n > F^{-1}_{\chi^2_{k-1}}(1-\alpha)$$

Dans le cas ou le paramètre θ n'est pas connu, on a le résultat suivant :

orème Si $n \to \infty$, alors D_n est asymptotiquement distribué comme une variable de χ^2_{k-s-1} et ceci quelle que soit la loi de X, oô s désigne le nombre de paramètres estimés.

narque Dans le calcul des p_i, on utilise les valeurs estimés des paramètres $\hat{\theta}$.

Test d'adéquation : test de χ^2 - Exemple

xemple Le tableau suivant présente les résultats de 500 mesures de l'erreur de pointage en dérive lors du tir à partir d'un avion sur une cible terrestre.

mesures (en radian)	effectifs
$[-4,-3[$	6
$[-3,-2[$	25
$[-2,-1[$	72
$[-1,0[$	133
$[0,1[$	120
$[1,2[$	88
$[2,3[$	46
$[3,4[$	10

Testons l'adéquation des observations à une loi normale $\mathcal{N}(\mu,\sigma^2)$ avec un risque $\alpha = 0.05$.

Test d'adéquation : test de χ^2 - Exemple

- Pour $i = 1,...,8$ notons c_i le centre de la i-ème classe. Les estimateurs de μ et de σ^2 sont donnés par :

$$\hat{\mu} = \frac{1}{500}\sum_{i=1}^{8} n_i c_i = 0.168;\ \hat{\sigma}^2 = \frac{1}{500}\sum_{i=1}^{8} n_i (c_i - \hat{\mu})^2 = 2.098$$

- Sous H_0, on a

$$p_i = \phi\left(\frac{e_i - \hat{\mu}}{\hat{\sigma}}\right) - \phi\left(\frac{e_{i-1} - \hat{\mu}}{\hat{\sigma}}\right)$$

  oô ϕ désigne la fonction de répartition de $\mathcal{N}(0,1)$.

- On rejette donc l'hypothèse H_0 si $D > F^{-1}_{\chi^2_{8-2-1}}(1-0.05) = F^{-1}_{\chi^2_5}(0.95) = 11.07$.

- D'après calcul, on trouve $D = 3.524$. Donc, on accepte l'hypothèse H_0.

1.5. Test d'indépendance

Test d'indépendance

- Soit (X, Y) un couple de v.a à valeurs dans $\{1, .., r\} \times \{l, ..., s\}$. Soit $((X_i; Y_i); 1 \leq i \leq n)$ un n-échantillon de (X, Y).

- On note $p_{ij} = P_{H_0}(X_1 = i, Y_1 = j)$, et les marginales

$$p_{i.} = \sum_{j=1}^{s} p_{ij}, \quad p_{.j} = \sum_{i=1}^{r} p_{ij}$$

- On souhaite vérifier si les variables X et Y sont indépendantes.

- On étudie le test suivant

$$\begin{cases} H_0 : \; p_{ij} = p_{i.}p_{.j} \\ H_1 : \; p_{ij} \neq p_{i.}p_{.j} \end{cases}$$

- On note les occurrences

$$N_{ij} = \sum_{k=1}^{n} \mathbf{1}_{X_k=i, Y_k=j}; \quad N_{i.} = \sum_{k=1}^{n} \mathbf{1}_{X_k=i}, \quad N_{.j} = \sum_{k=1}^{n} \mathbf{1}_{Y_k=j}$$

Test d'indépendance : construction du test

- La statistique du test est :

$$D_n = \sum_{i=1}^{r}\sum_{j=1}^{s} D_{ij} = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(N_{ij} - n\hat{p}_{i.}\hat{p}_{.j})^2}{n\hat{p}_{i.}\hat{p}_{.j}} = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(N_{ij} - \frac{N_{i.}N_{.j}}{n})^2}{\frac{N_{i.}N_{.j}}{n}}$$

- D représente la distance entre le tableau observé et le tableau théorique. Sous H_0, D_n suit une loi de $\chi^2_{(s-1)(r-1)}$

- Pour $\alpha \in [01]$ donné, on rejette donc l'hypothèse H_0 si

$$D > F^{-1}_{X^2_{(r-1)(s-1)}}(1-\alpha)$$

1.6. d'ajustement de Kolmogorov

Test de Kolmogorov

- Test non-paramétrique d'adjustement à une distribution entièrement spécifiiée de fonction de répartition $F(x)$.

- Soit $X_1, ..., X_n$ un échantillon de loi inconnue P de fonction de répartition F supposée continue. L'objectif du test de Kolmogorov est l'ajustement de la loi inconnue P à une loi connue P_0 de fonction de répartition continue F_0.

- Le problème de test à étudier est le suivant :

$$\begin{cases} H_0 : & F = F_0 \ \ (F_0 \text{ connue}) \\ H_1 : & F \neq F_0 \end{cases}$$

- On introduit d'abord F_n la fonction de répartition empirique associée à l'échantillon $X_1, ..., X_n$.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{]-\infty, x[}(X_i)$$

Test de Kolmogorov : construction du test

- F_n est un estimateur sans biais de F, en effet :

$$E[F_n(x)] = \frac{1}{n}\sum_{i=1}^{n} E[\mathbf{1}_{]-\infty,x[}(X_i)] = P(X_1 < x) = F(x)$$

- D'après le Théorème Glivenko-Cantelli (admis)

$$\sup_{y\in\mathbb{R}} |F_n(y) - F(y)| \to 0, \text{ p.s. quand } \ n \to \infty.$$

- En particulier, si l'on suppose de plus que F est continue, alors

$$\sqrt{n} \sup_{y\in\mathbb{R}} |F_n(y) - F(y)| \to W \text{ en loi, quand } \ n \to \infty.$$

- W est indépendante de F et admet comme fonction de répartition :

$$K(y) := \sum_{i=-\infty}^{+\infty} (-1)^k \exp(-2k^2y^2)$$

Test de Kolmogorov : construction du test

- La statistique du test est :

$$D_n = \sqrt{n} \sup |F_n(x) - F_0(x)|$$

- Sous H_0, (d'après les résultats de Glivenko - Kolmogorov en théorie de l'échantillonnage) D_n est asymptotiquement distribué comme suit :

$$P(D_n < y) \to K(y) = \sum_{i=-\infty}^{+\infty} (-1)^k \exp(-2k^2y^2)$$

- La fonction K a été tabulée et fournit donc la région de rejet :

$$D_n > K^{-1}(1-\alpha)$$