

# **Inférence statistique (UGA M1 S1) TESTS STATISTIQUES(2) <sup>1</sup>**

Michal W. Urdanivia\*

\*Université de Grenoble Alpes, Faculté d'Économie, GAEL,  
e-mail: [michal.wong-urdanivia@univ-grenoble-alpes.fr](mailto:michal.wong-urdanivia@univ-grenoble-alpes.fr)

8 novembre 2022

---

1. Cette parite du cours utilise une partie du cours de Guillaume Lecué de Statistique à l'ENSAE. Il est vivement remercié.

# Cours précédent (rappels)

1. vocabulaire : accepter, rejeter, zone de rejet, erreur de 1ère et 2-ième espèces, niveau d'un test, fonction puissance, p-value, test UPP
2. résultat : construction du test de Neyman-Pearson grâce au rapport de vraisemblance pour les problèmes de tests à hypothèses simples – propriété UPP des tests de Neyman-Pearson

# Test UPP dans une famille à rapport de vraisemblance monotone

# Rappels sur Neyman-Pearson

On a montré que le test de Neyman-Pearson est UPP pour un test de la forme :

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

ce test est de la forme

$$\varphi_\alpha(Z) = \begin{cases} H_0 & \text{si } f(\theta_1, Z) \leq c_\alpha f(\theta_0, Z) \\ H_1 & \text{sinon} \end{cases}$$

où  $c_\alpha$  est tel que (quand il y a une solution)

$$\mathbb{P}_{\theta_0} [f(\theta_1, Z) > c_\alpha f(\theta_0, Z)] = \alpha$$

Alors :

1.  $\varphi_\alpha$  est de niveau  $\alpha$
2.  $\varphi_\alpha$  est le test le plus puissant parmi tous les tests de niveau  $\alpha$

# Famille à rapport de vraisemblance monotone (1/2)

On va construire un test UPP pour les problèmes de test de la forme :

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_0$$

quand le rapport de vraisemblance a une structure particulière.

On considère le cadre suivant :

- ▶  $\mathcal{E} = (\mathfrak{Z}, \mathcal{Z}, \{\mathbb{P}_\theta : \theta \in \Theta\})$  : une expérience statistique
- ▶  $Z$  : une observation dans cette expérience
- ▶ on suppose que  $\Theta \subset \mathbb{R}$  est un intervalle ouvert
- ▶ le modèle est dominé par  $\mu$  tel que  $f(\theta, z) = \frac{d\mathbb{P}_\theta}{d\mu}(z) > 0$ , pour  $\mu$ -presque tout  $z \in \mathfrak{Z}$ .

# Famille à rapport de vraisemblance monotone (2/2)

## Définition

On dit que la famille de densités  $\{f(\theta, \cdot) : \theta \in \Theta\}$  est à **rapport de vraisemblance monotone** s'il existe une statistique  $T(Z)$  telle que

$$\forall \theta_1 < \theta_2, \quad \frac{f(\theta_2, Z)}{f(\theta_1, Z)} \text{ est une fonction monotone de } T(Z)$$

ex. : quand  $Z = (X_1, \dots, X_n)$  où  $X_i \stackrel{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$ , le rapport de vraisemblance pour  $\theta_1 < \theta_2$

$$\frac{f(\theta_2, Z)}{f(\theta_1, Z)} = \exp\left(\frac{n}{\sigma^2}(\theta_2 - \theta_1)\overline{X}_n\right) \exp\left(\frac{-n}{2\sigma^2}(\theta_2^2 - \theta_1^2)\right)$$

est une fonction croissante de  $T(Z) = \overline{X}_n$ .

# Test UPP – théorème de Lehmann

Soit  $\{f(\theta, \cdot) : \theta \in \mathbb{R}\}$  une famille vérifiant les hypothèses précédentes et à **rapport de vraisemblance monotone**, par exemple croissant en  $T(Z)$ . Soit  $\alpha \in (0, 1)$ . On suppose que pour  $\theta_0 \in \Theta$ , il existe  $\rho(\theta_0, \alpha)$  tel que

$$\mathbb{P}_{\theta_0} [T(Z) > \rho(\theta_0, \alpha)] = \alpha$$

alors le test de région de rejet

$$\mathcal{R}^* = \{z \in \mathfrak{Z} : T(z) > \rho(\theta_0, \alpha)\}$$

est :

1. de niveau  $\alpha$
2. UPP (le plus puissant parmi tous les tests de niveau  $\alpha$ )

pour le problème de test

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_0$$

## Exemple de test UPP

Pour  $Z = (X_1, \dots, X_n)$  où  $X_i \stackrel{i.i.d.}{\sim} \mathcal{N}(\theta, 1)$ , le rapport de vraisemblance

$$\frac{f(\theta_2, Z)}{f(\theta_1, Z)} = \exp\left(\frac{n}{\sigma^2}(\theta_2 - \theta_1)\bar{X}_n\right) \exp\left(\frac{-n}{2\sigma^2}(\theta_2^2 - \theta_1^2)\right)$$

est une fonction croissante de  $T(Z) = \bar{X}_n$ .

Alors pour le problème de test

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_0$$

le test de région de rejet

$$\mathcal{R}^* = \{(x_1, \dots, x_n)^\top \in \mathbb{R}^n : \bar{x}_n > \theta_0 + t_{n,\alpha}\}$$

où  $t_{n,\alpha}$  est tel que  $\mathbb{P}_{\theta_0}[\bar{X}_n > \theta_0 + t_{n,\alpha}] = \alpha$  (càd  $t_{n,\alpha} = q_{1-\alpha}/\sqrt{n}$ ), est un test de niveau  $\alpha$ , UPP.



# objectifs : niveau asymptotique et consistance

idée : Les tests UPP sont très rares. En général, on se contentera de construire des tests  $\varphi$  satisfaisant deux propriétés quand le nombre  $n$  d'observations tend vers  $+\infty$  :

1. on dit qu'un test est de **niveau asymptotique  $\alpha$**  quand

$$\forall \theta \in \Theta_0, \quad \limsup_{n \rightarrow \infty} \mathbb{P}_\theta[\varphi = H_1] \leq \alpha$$

2. on dit qu'un test est **consistant** ou **convergeant** quand

$$\forall \theta \in \Theta_1, \quad \lim_{n \rightarrow \infty} \mathbb{P}_\theta[\varphi = H_1] = 1$$

# Test de Wald

# Le test de Wald : hypothèse nulle simple

On considère le problème de test

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

Hypothèse : on dispose d'un estimateur  $\hat{\theta}_n$  **asymptotiquement normal**

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, v(\theta))$$

et  $\theta \mapsto v(\theta) > 0$  est **continue**.

Sous  $H_0$  (càd quand  $Z \sim \mathbb{P}_{\theta_0}$ ) : on a **la convergence**

$$\sqrt{\frac{n}{v(\hat{\theta}_n)}}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, 1)$$

Rem. : on peut prendre  $v(\theta_0)$  à la place de  $v(\hat{\theta}_n)$  (car  $H_0$  est une hypothèse simple).

# Test de Wald : comportement sous $H_0$ et $H_1$

sous  $H_0$  :

$$T_n = n \frac{(\hat{\theta}_n - \theta_0)^2}{v(\hat{\theta}_n)} \xrightarrow{d} \chi^2(1)$$

sous  $H_1$  :  $\forall \theta \in \Theta_1$  (càd  $\theta \neq \theta_0$ ),

$$T_n \xrightarrow{\mathbb{P}_\theta} +\infty$$

On va utiliser  $T_n$  comme statistique de test et, au vue du comportement de  $T_n$  sous  $H_1$ , on considère des tests de la forme

$$\varphi_\alpha = \begin{cases} H_0 & \text{quand } T_n \leq t_{n,\alpha} \\ H_1 & \text{sinon} \end{cases}$$

Pour déterminer le seuil  $t_{n,\alpha}$ , on regarde le comportement sous  $H_0$  : soit  $q_{1-\alpha}^{\chi^2(1)}$ , le quantile d'ordre  $1 - \alpha$  d'une  $\chi^2(1)$ , on choisit

$$t_{n,\alpha} = q_{1-\alpha}^{\chi^2(1)}$$

# Propriétés du test de Wald

## Proposition

*Le test de Wald de l'hypothèse simple  $\theta = \theta_0$  contre l'alternative  $\theta \neq \theta_0$  basé sur l'estimateur a.n.  $\hat{\theta}_n$  est*

- ▶ *de niveau asymptotique  $\alpha$  :*

$$\mathbb{P}_{\theta_0} [\varphi_\alpha = H_1] \longrightarrow \alpha$$

- ▶ *convergent (ou consistant) : pour tout  $\theta \neq \theta_0$*

$$\mathbb{P}_\theta [\varphi_\alpha = H_1] \longrightarrow 1$$

# Preuve

- ▶ niveau asymptotique  $\alpha$  : **par construction**
- ▶ Contrôle de la puissance : Soit  $\theta \neq \theta_0$ . On a, sous  $\mathbb{P}_\theta$ ,

$$\frac{(\hat{\theta}_n - \theta_0)}{v(\hat{\theta}_n)} \xrightarrow{\mathbb{P}_\theta} \frac{(\theta - \theta_0)^2}{v(\theta_0)} > 0$$

et comme

$$T_n = n \frac{(\hat{\theta}_n - \theta_0)}{v(\hat{\theta}_n)}$$

alors  $T_n \xrightarrow{\mathbb{P}_\theta} +\infty$  et donc, quand  $n$  tends vers  $+\infty$

$$\mathbb{P}_\theta [\varphi_\alpha = H_1] = \mathbb{P}_\theta [T_n > q_{1-\alpha}^{\chi^2(1)}] \longrightarrow 1$$

# Test de Wald : hypothèse nulle composite

- ▶ Même contexte :  $\Theta \subset \mathbb{R}^d$  et on dispose d'un estimateur  $\hat{\theta}_n$  asymptotiquement normal :

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V(\theta))$$

où  $V(\theta)$  est définie positive et continue en  $\theta$ .

- ▶ But : Tester  $H_0 : \theta \in \Theta_0$  contre  $H_1 : \theta \notin \Theta_0$ , où

$$\Theta_0 = \{\theta \in \Theta, g(\theta) = 0\}$$

et

$$g : \mathbb{R}^d \rightarrow \mathbb{R}^m$$

est régulière.

# Test de Wald : comportement sous $H_0$

- **Hypothèse** : le gradient  $\nabla g(\theta)$  de  $g$  est tel que

$$\theta \in \Theta \mapsto \Sigma_g(\theta) = \nabla g(\theta)^\top V(\theta) \nabla g(\theta)$$

est continue et inversible en tout point  $\theta$ .

## Proposition

Pour tout  $\theta \in \Theta_0$  (i.e. **sous  $H_0$** ), on a :



$$\sqrt{n}g(\hat{\theta}_n) \xrightarrow{d} \mathcal{N}(0, \Sigma_g(\theta))$$



$$T_n = ng(\hat{\theta}_n)^\top \Sigma_g(\hat{\theta}_n)^{-1} g(\hat{\theta}_n) \xrightarrow{d} \chi^2(m)$$

- Preuve : méthode « delta » multidimensionnelle.



# Test de Wald : propriétés

## Proposition

*Sous les hypothèses précédentes, le test*

$$\varphi_\alpha = \begin{cases} H_0 & \text{si } T_n \leq q_{1-\alpha}^{\chi^2(m)} \\ H_1 & \text{sinon} \end{cases}$$

*pour  $q_{1-\alpha}^{\chi^2(m)}$ , quantile d'ordre  $1 - \alpha$  d'une  $\chi^2(m)$ , est*

- ▶ de **niveau asymptotique  $\alpha$**  :  $\forall \theta \in \Theta_0$ ,

$$\mathbb{P}_\theta [\varphi_\alpha = H_1] \rightarrow \alpha$$

- ▶ **Convergeant** :  $\forall \theta \notin \Theta_0$ ,

$$\mathbb{P}_\theta [\varphi_\alpha = H_1] \rightarrow 1$$

# Tests d'adéquation

# Problème de test d'adéquation

Situation : On observe un  $n$ -échantillon de loi (cdf)  $F$  inconnue

$$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} F$$

et on se donne une loi (cdf)  $F_0$  (ex. :  $F_0$  cdf d'une  $\mathcal{N}(0, 1)$ )

But : Tester

$$H_0 : F = F_0 \text{ contre } H_1 : F \neq F_0$$

Ce type de problème de test est appelé : **problème de test d'adéquation**

# Test (d'adéquation) de Kolmogorov-Smirnov

# Test de Kolmogorov-Smirnov

- Rappel : Si la fonction de répartition  $F$  est continue alors

$$\sqrt{n} \left\| \widehat{F}_n - F \right\|_{\infty} \xrightarrow{d} K$$

où la loi de  $K$  ne dépend pas de  $F$  (cf. cours1)

## Proposition (Test de Kolmogorov-Smirnov)

Soit  $q_{1-\alpha}^K$  tel que  $\mathbb{P}[K > q_{1-\alpha}^K] = \alpha$ . Le test défini par

$$\varphi_{\alpha} = \begin{cases} H_0 & \text{si } T_n \leq q_{1-\alpha}^K \\ H_1 & \text{sinon} \end{cases} \quad \text{pour } T_n = \sqrt{n} \|\widehat{F}_n - F_0\|_{\infty}$$

est *de niveau asymptotique  $\alpha$*  :  $\mathbb{P}_{F_0}[\varphi_{\alpha} = H_1] \rightarrow \alpha$  et *consistant* :

$$\forall F \neq F_0 : \mathbb{P}_F[\varphi_{\alpha} = H_1] \rightarrow 1$$

# Tests d'adéquation du $\chi^2$

# Test du Chi-deux

- ▶  $X$  variable qualitative :  $X \in \{1, \dots, d\}$

$$\mathbb{P}[X = \ell] = p_\ell, \ell = 1, \dots, d.$$

- ▶ La loi de  $X$  est caractérisée par  $\mathbf{p} = (p_1, \dots, p_d)^T$
- ▶ Notation

$$\mathcal{M}_d = \left\{ \mathbf{p} = (p_1, \dots, p_d)^T : p_\ell \geq 0, \sum_{\ell=1}^d p_\ell = 1 \right\}$$

- ▶ **Objectif** : A partir d'un  $n$ -échantillon  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbf{p}$ , et d'une loi  $\mathbf{q} \in \mathcal{M}_d$  donnée on veut tester

$$H_0 : \mathbf{p} = \mathbf{q} \text{ contre } H_1 : \mathbf{p} \neq \mathbf{q}$$

# Construction « naturelle » d'un test

- Comparaison des fréquences empiriques

$$\hat{p}_{n,\ell} = \frac{1}{n} \sum_{i=1}^n I(X_i = \ell) \quad \text{proche de } q_\ell, \quad \ell = 1, \dots, d ?$$

- Loi forte des grands nombres : sous  $H_0$ ,

$$(\hat{p}_{n,1}, \dots, \hat{p}_{n,d}) \xrightarrow{p.s.} (p_1, \dots, p_d) = \mathbf{p}.$$

- Théorème central-limite

$$\mathbf{U}_n(\mathbf{p}) = \sqrt{n} \left( \frac{\hat{p}_{n,1} - p_1}{\sqrt{p_1}}, \dots, \frac{\hat{p}_{n,d} - p_d}{\sqrt{p_d}} \right) \xrightarrow{d} ?$$

- Composante par composante : oui. Convergence globale plus délicate (coordonnées dépendantes) : **TCL multidimensionnel**



# Statistique du Chi-deux

## Proposition

Si les composantes de  $\mathbf{p}$  sont toutes non-nulles

- ▶ On a la *convergence en loi* sous  $\mathbb{P}_{\mathbf{p}}$

$$\mathbf{U}_n(\mathbf{p}) \xrightarrow{d} \mathcal{N}(0, V(\mathbf{p}))$$

avec  $V(\mathbf{p}) = \text{Id}_d - \sqrt{\mathbf{p}}(\sqrt{\mathbf{p}})^T$  et  $\sqrt{\mathbf{p}} = (\sqrt{p_1}, \dots, \sqrt{p_d})^T$

- ▶ *De plus*

$$\|\mathbf{U}_n(\mathbf{p})\|_2^2 = n \sum_{\ell=1}^d \frac{(\hat{p}_{n,\ell} - p_\ell)^2}{p_\ell} \xrightarrow{d} \chi^2(d-1)$$

# Preuve de la normalité asymptotique

- Pour  $i = 1, \dots, n$  et  $1 \leq \ell \leq d$ , on pose

$$Y_\ell^i = \frac{1}{\sqrt{p_\ell}} (I(X_i = \ell) - p_\ell)$$

- Les vecteurs  $\mathbf{Y}_i = (Y_1^i, \dots, Y_d^i)$  sont **i.i.d.** et

$$\mathbf{U}_n(\mathbf{p}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{Y}_i,$$

$$\mathbb{E}[Y_\ell^i] = 0, \mathbb{E}[(Y_\ell^i)^2] = 1 - p_\ell, \mathbb{E}[Y_\ell^i Y_{\ell'}^i] = -(p_\ell p_{\ell'})^{1/2}$$

- On applique le **TCL vectoriel**

# Convergence de la norme au carré

- ▶ On a donc, sous  $H_0$ ,  $\mathbf{U}_n(\mathbf{p}) \xrightarrow{d} \mathcal{N}(0, V(\mathbf{p}))$
- ▶ et aussi, par [Cochran](#),

$$\|\mathbf{U}_n(\mathbf{p})\|_2^2 \xrightarrow{d} \|\mathcal{N}(0, V(\mathbf{p}))\|_2^2 \sim \chi^2(d-1)$$

$V(\mathbf{p}) = \text{Id}_d - \sqrt{\mathbf{p}}(\sqrt{\mathbf{p}})^T$  est la projection orthogonale sur  $\text{vect}\{\sqrt{\mathbf{p}}\}^\perp$  qui est de dimension  $d-1$

# Test d'adéquation du $\chi^2$

- ▶ « distance » du  $\chi^2$  :

$$\chi^2(\mathbf{p}, \mathbf{q}) = \sum_{\ell=1}^d \frac{(p_{\ell} - q_{\ell})^2}{q_{\ell}}.$$

- ▶ Avec ces notations  $\|\mathbf{U}_n(\mathbf{q})\|_2^2 = n\chi^2(\hat{\mathbf{p}}_n, \mathbf{q})$ .

## Proposition

Pour  $\mathbf{q} \in \mathcal{M}_d$  le test défini par

$$\varphi_{\alpha} = \begin{cases} H_0 & \text{si } n\chi^2(\hat{\mathbf{p}}_n, \mathbf{q}) \leq q_{1-\alpha}^{\chi^2(d-1)} \\ H_1 & \text{sinon} \end{cases}$$

est de *niveau asymptotique  $\alpha$*  et *consistant* pour tester

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

# Exemple de mise en oeuvre : expérience de Mendel

- ▶ Soit  $d = 4$  et

$$\mathbf{q} = \left( \frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16} \right).$$

- ▶ fréquence empirique :  $n = 556$

$$\hat{\mathbf{p}}_{556} = \frac{1}{556} (315, 101, 108, 32).$$

- ▶ Calcul de la statistique du  $\chi^2$

$$556 \times \chi^2(\hat{\mathbf{p}}_{556}, \mathbf{q}) = 0,47.$$

- ▶ On a  $q_{95\%}^{\chi^2(3)} = 7.815$ .
- ▶ **Conclusion** : Puisque  $0.47 < 7.815$ , on accepte l'hypothèse  $\mathbf{p} = \mathbf{q}$  au niveau  $\alpha = 5\%$ . De plus, la p-value vaut 0.93 : très grand niveau de confiance en l'acceptation.

# Test d'indépendance du $\chi^2$

# Problématique

données/modèle : on observe un  $n$ -échantillon de couples de variables aléatoires **qualitatives**

$$(X_1, Y_1), \dots, (X_n, Y_n) \stackrel{i.i.d.}{\sim} (X, Y) \in \{1, \dots, d_1\} \times \{1, \dots, d_2\}$$

La loi  $\mathbf{p}$  de  $(X, Y)$  appartient à

$$\mathcal{M}_{d_1, d_2} = \left\{ \mathbf{p} = (p_{\ell, \ell'})_{\substack{1 \leq \ell \leq d_1 \\ 1 \leq \ell' \leq d_2}} : 0 \leq p_{\ell, \ell'}, \sum_{\ell, \ell'} p_{\ell, \ell'} = 1 \right\}$$

On considère le problème de test suivant :

$$H_0 : \mathbb{P}_{(X, Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y \text{ contre } H_1 : \mathbb{P}_{(X, Y)} \neq \mathbb{P}_X \otimes \mathbb{P}_Y$$

# principe du test d'indépendance du $\chi^2$ (1/2)

On a l'équivalence :

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y \Leftrightarrow \forall \ell, \ell', p_{\ell, \ell'} = p_{\ell, \bullet} \times p_{\bullet, \ell'}$$

où

$$p_{\ell, \ell'} = \mathbb{P}[(X, Y) = (\ell, \ell')], \quad p_{\ell, \bullet} = \mathbb{P}[X = \ell], \quad p_{\bullet, \ell'} = \mathbb{P}[Y = \ell']$$

On estime ces quantités par leurs fréquences empiriques :

$$\hat{p}_{\ell, \ell'}^{(n)} = \frac{1}{n} \sum_{i=1}^n I[(X_i, Y_i) = (\ell, \ell')]$$
$$\hat{p}_{\ell, \bullet}^{(n)} = \frac{1}{n} \sum_{i=1}^n I[X_i = \ell] \quad \hat{p}_{\bullet, \ell'}^{(n)} = \frac{1}{n} \sum_{i=1}^n I[Y_i = \ell']$$



## principe du test d'indépendance du $\chi^2$ (2/2)

Idée :

1. On aimerait pouvoir tester si la loi du couple  $(X, Y)$  est égale à  $\mathbb{P}_X \otimes \mathbb{P}_Y$ , on ferait alors un test d'adéquation.
2. Mais, on ne connaît pas la loi  $\mathbb{P}_X \otimes \mathbb{P}_Y$ , on va donc l'estimer par

$$(\hat{p}_{\ell, \bullet}^{(n)} \times \hat{p}_{\bullet, \ell'}^{(n)})_{\substack{1 \leq \ell \leq d_1 \\ 1 \leq \ell' \leq d_2}}$$

3. on peut alors faire un test d'adéquation du  $\chi^2$  par rapport à cette loi estimée. La statistique de test est

$$n\chi^2\left((\hat{p}_{\ell, \ell'}^{(n)})_{\ell, \ell'}, (\hat{p}_{\ell, \bullet}^{(n)} \times \hat{p}_{\bullet, \ell'}^{(n)})_{\ell, \ell'}\right)$$

## Comportement de la statistique de test sous $H_0$ et sous $H_1$

On utilise la statistique de test

$$T_n = n\chi^2\left(\left(\hat{p}_{\ell,\ell'}^{(n)}\right)_{\ell,\ell'}, \left(\hat{p}_{\ell,\bullet}^{(n)} \times \hat{p}_{\bullet,\ell'}^{(n)}\right)_{\ell,\ell'}\right)$$

pour construire le test d'indépendance du  $\chi^2$ . Son comportement sous les deux hypothèses est donné par :

Sous  $H_0$  :

$$T_n \xrightarrow{d} \chi^2((d_1 - 1)(d_2 - 1))$$

Sous  $H_1$  :

$$T_n \xrightarrow{p.s.} +\infty$$

On regarde donc des tests de la forme

$$\varphi_\alpha = \begin{cases} H_0 & \text{si } T_n \leq t_{n,\alpha} \\ H_1 & \text{sinon} \end{cases}$$

et le seuil  $t_{n,\alpha}$  est donné par la loi limite de  $T_n$  sous  $H_0$

# test d'indépendance du $\chi^2$

## Theorem

Pour tout  $\alpha \in (0, 1)$ , le test

$$\varphi_\alpha = \begin{cases} H_0 & \text{si } T_n \leq q_{1-\alpha}^{\chi^2((d_1-1)(d_2-1))} \\ H_1 & \text{sinon} \end{cases}$$

où  $q_{1-\alpha}^{\chi^2((d_1-1)(d_2-1))}$  est le quantile d'ordre  $1 - \alpha$  d'une  $\chi^2((d_1 - 1)(d_2 - 1))$ , est de **niveau asymptotique  $\alpha$**  et **consistant**

## Exemple : test d'indépendance entre revenus et nombre d'enfants

Catégories de revenus : I (faible) à IV (élevé)

nb. enfants	I	II	III	IV	total ligne
0	2161	3577	2184	1636	9558
1	2755	5081	2222	1052	11110
2	936	1753	640	306	3635
3	225	419	96	38	778
$\geq 4$	39	98	31	14	182
tot. col.	6116	10928	5173	3016	25263

On obtient :

$$\begin{aligned} T_n &= n\chi^2\left((\hat{p}_{\ell,\ell'}^{(n)})_{\ell,\ell'}, (\hat{p}_{\ell,\bullet}^{(n)} \times \hat{p}_{\bullet,\ell'}^{(n)})_{\ell,\ell'}\right) \\ &= n \sum_{\substack{\ell \in \{0,1,2,3, \geq 4\} \\ \ell' \in \{I,II,III,IV\}}} \frac{(\hat{p}_{\ell,\ell'}^{(n)} - \hat{p}_{\ell,\bullet}^{(n)} \hat{p}_{\bullet,\ell'}^{(n)})^2}{\hat{p}_{\ell,\bullet}^{(n)} \hat{p}_{\bullet,\ell'}^{(n)}} = 568.5 \end{aligned}$$

## Table des quantiles d'une $\chi^2(12)$

On a ici  $(d_1 - 1)(d_2 - 1) = (5 - 1)(4 - 1) = 12$ . On regarde alors les quantiles

$$q_{1-\alpha}^{\chi^2(12)}$$

$\alpha$	0,2	0,1	0,05	0,02	0,01	0,005	0,001
$q_{1-\alpha}^{\chi^2(12)}$	15,8	18,5	21,0	24,0	26,2	28,2	32,9

Python :

```
> from scipy.stats import chi2  
> 1-chi2.cdf(100, 12)  
5.5e-16
```

Ici la statistique du test vaut  $T_n = 568.5$  donc la p-value est plus petite que  $5.5 \cdot 10^{-16}$ . Alors on rejete avec confiance. Il y a donc une très forte dépendance entre le nombre d'enfants et les revenus.