Aléatoire

MAP 311 - X2013

Leçon 8

Mots clés :

1. Statistique gaussienne : estimation de la moyenne
   - Variance inconnue
   - Estimateur du maximum de vraisemblance
2. Estimation non-paramétrique
3. Tests d'hypothèses dans le cas gaussien
4. Test du χ^2
5. Pour en savoir plus : estimation d'un jeu avec observation partielle

Statistique sur échantillon gaussien, estimation de moyenne

Modèle : $(Y_1, \dots, Y_n)$ v.a. i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$ (moyenne et variance inconnues)

Moyenne empirique : $\bar{Y}_n := \frac{1}{n}\sum_{i=1}^n Y_i \overset{\text{loi}}{=} \mathcal{N}(\mu, \frac{\sigma^2}{n})$

Variance empirique : $\sigma_n^2 := \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2$

Théorème. $\bar{Y}_n$ et σ_n^2 sont des estimateurs convergents et sans biais de (μ, σ^2).

Preuve. Pour $\bar{Y}_n$, voir leçon 7. Il reste à démontrer

- $\sigma_n^2 \overset{p.s.}{\to} \sigma^2$,
- $\mathbb{E}(\sigma_n^2) = \sigma^2$.

On a $\sigma_n^2 = \dfrac{n}{n-1}\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^n Y_i^2 - \bar{Y}_n^2\right)$.

- Par la LGN, $\sigma_n^2 \overset{p.s.}{\to} \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \sigma^2$.
- $\mathbb{E}(\sigma_n^2) = \dfrac{n}{n-1}\mathbb{E}(Y^2) - \dfrac{n}{n-1}\left[\dfrac{\mathbb{V}\mathrm{ar}(Y)}{n} + (\mathbb{E}(Y))^2\right]$
$= \dfrac{n}{n-1}\mathbb{V}\mathrm{ar}(Y) - \dfrac{n}{n-1}\dfrac{\mathbb{V}\mathrm{ar}(Y)}{n} = \mathbb{V}\mathrm{ar}(Y)$. □

Vitesse de convergence

Théorème. $\sqrt{n}\frac{(\bar{Y}_n-\mu)}{\sigma}$ et $(n-1)\frac{\sigma_n^2}{\sigma^2}$ sont indépendants, respectivement de loi $\mathcal{N}(0,1)$ et de loi $\chi^2(n-1)$.

Corollaire. $\sqrt{n}\frac{(\bar{Y}_n-\mu)}{\sigma_n} \overset{\text{loi}}{=} \frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}}$ avec numérateur et dénominateur indépendants, c.à-d. une v.a. de loi de Student avec paramètre $n-1$.

Intervalles de confiance. Avec probabilité 95%, on a (pour chaque n)

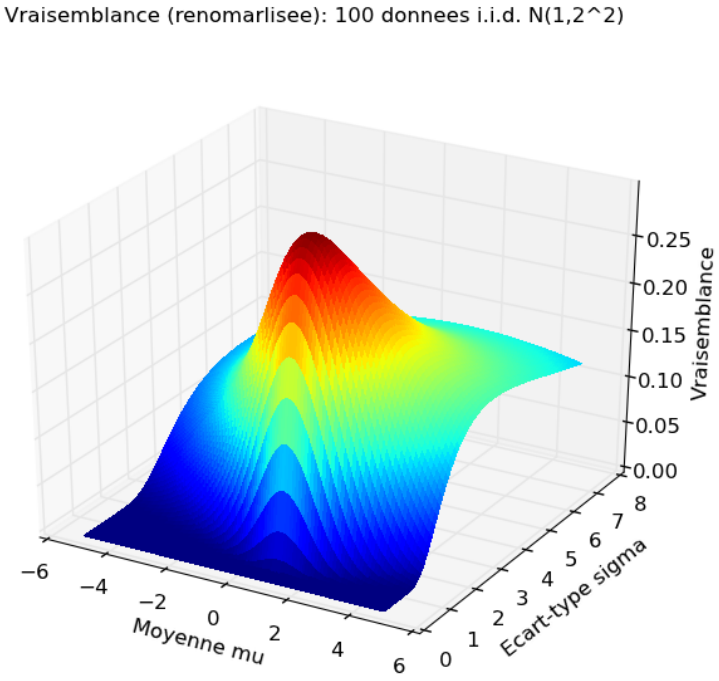
$$\bar{Y}_n - t_{n-1}(97.5\%)\frac{\sigma_n}{\sqrt{n}} \leq \mu \leq \bar{Y}_n + t_{n-1}(97.5\%)\frac{\sigma_n}{\sqrt{n}}.$$

n	10	20	100	$+\infty$
$t_{n-1}(97.5\%)$	2.262	2.093	1.984	1.96

Estimateur du maximum de vraisemblance (e.m.v.)

Etant donnés des paramètres (μ, σ^2), le vecteur aléatoire $(Y_1, \dots, Y_n)$ i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$ a une loi de densité

$$p_{\mu,\sigma^2}(y_1, \dots, y_n) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_j-\mu)^2}{2\sigma^2}}.$$



Définition. L'e.m.v. est défini par

$$(\hat{\mu}_\mathbf{n}, \hat{\sigma}^\mathbf{2}_\mathbf{n}) = \arg\max_{\mu,\sigma^\mathbf{2}} \mathbf{p}_{\mu,\sigma^\mathbf{2}}(\mathbf{Y_1}, \dots, \mathbf{Y_n})$$

(paramètres qui rendent l'observation la plus "probable").

Proposition. Dans ce cas gaussien, on trouve $\hat{\mu}_n = \bar{X}_n$ et $\hat{\sigma}^2_n = \frac{n-1}{n}\sigma^2_n$.

Remarque. En général, l'e.m.v. a de très bonnes propriétés de convergence et d'optimalité (voir cours de Statistique de 2A).

Estimation non paramétrique

$(X_n)_n$: variables aléatoires indépendantes, de même loi.

Proposition (LGN). Convergence (ponctuelle) des fonctions de répartition :

$$\mathbf{F_n(x) := \frac{1}{n}\sum_{i=1}^{n} 1_{X_i(\omega)\leq x} \to \mathbb{P}(X \leq x) := F(x), \quad \forall x \quad p.s.}$$

Proposition (TCL). $\sqrt{n}\Big(F_n(x) - F(x)\Big) \overset{loi}{\Rightarrow} \mathcal{N}\Big(0, F(x)(1-F(x))\Big)$.

Théorème (Glivenko-Cantelli, admis). Convergence (uniforme) :

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0 \quad p.s.$$

➠ Validation de la représentation en histogramme :

$$\mathbf{\frac{1}{n}\#\{i : 1 \leq i \leq n, a < X_i \leq b\} = F_n(b) - F_n(a) \overset{p.s.}{\to} F(b) - F(a).}$$

On suppose maintenant que la loi de X a une densité $f > 0$.

Théorème. $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)|$ a une loi qui ne dépend pas de la loi F :

$$\sup_{\mathbf{x}\in\mathbb{R}} |\mathbf{F_n(x) - F(x)}| \overset{\text{loi}}{=} \sup_{\mathbf{u}\in[0,1]} |\frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{1_{U_i \le u}} - \mathbf{u}|$$

où $(U_i)_i$ sont des v.a. i.i.d. de loi uniforme sur $[0, 1]$.

Preuve. $U_i = F(X_i)$ définit une suite de v.a. i.i.d. de loi uniforme sur $[0, 1]$, et F est bijective de $\mathbb{R}$ dans $[0, 1]$:

$$\sup_{x\in\mathbb{R}} |\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \le x} - F(x)| = \sup_{u\in[0,1]} |\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \le F^{-1}(u)} - u| = \sup_{u\in[0,1]} |\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{U_i \le u} - u|. \square$$

Théorème (Kolmogorov-Smirnov, admis).

La v.a. $\sqrt{n} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)|$ converge en loi vers une variable aléatoire B de fonction de répartition

$$\mathbb{P}(B \le x) = 1 + 2 \sum_{k\ge1} (-1)^k e^{-2k^2 x}, \quad x \ge 0.$$

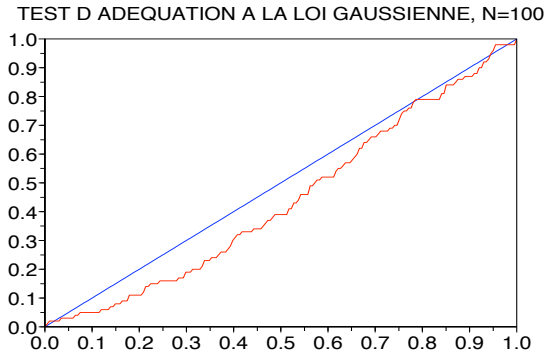
Comment tester si F (inconnue) est égale à F_0 donnée ?

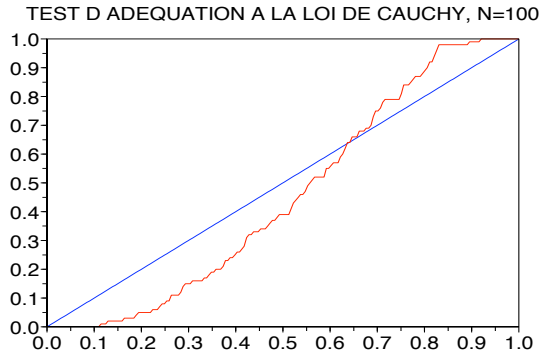
Posons $Q_0 = F_0^{-1}$. Si $F = F_0$, $u \mapsto F_n(Q_0(u)) \overset{p.s.}{\to} F(Q_0(u)) = u$: alignement asymptotique sur la diagonale.

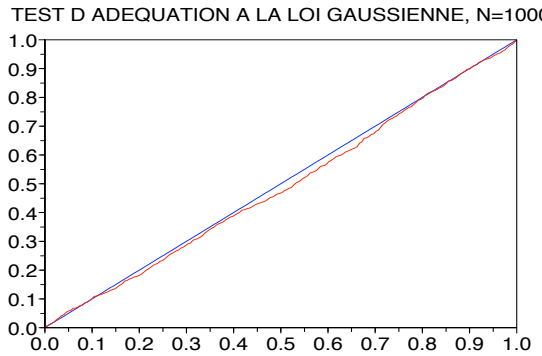
Exemple. Données de loi normale $\mathcal{N}(0,1)$.
Test d'adéquation à

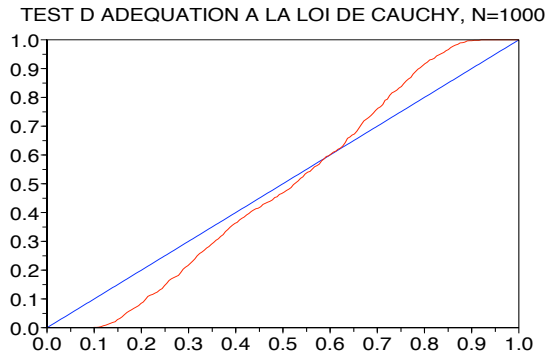
- la loi normale,
- la loi de Cauchy.

Représentation dite en Probabilité/Probabilité.









Utiliser le résultat de Kolmogorov-Smirnov permet d'obtenir des intervalles de confiance sur les fluctuations uniformes.

Pour en savoir plus : convergence des quantiles empiriques

A partir des observations $(X_i)_{1\leq i\leq n}$ i.i.d., on forme la statistique d'ordre :

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(i)} \leq \cdots \leq X_{(n)}.$$

Pour simplifier la discussion, on suppose que *la loi de X est à densité* : alors les $\leq$ sont des $<$ avec probabilité 1.

La fonction de répartition empirique F_n associée est alors telle que :

$$\mathbf{F_n(x) = \frac{i}{n}} \quad \text{si} \quad \mathbf{x \in [X_{(i)}, X_{(i+1)}[} \qquad (\text{avec } X_{(0)} := -\infty, X_{(n+1)} = +\infty).$$

Le quantile empirique est alors $\mathbf{Q_n(\alpha) = X_{(\lceil n\alpha \rceil)}}$.

Exemple. La 11ème donnée la plus grande sur 1000 données donne Q(99%).

Convergence - Loi des grands nombres et Théorème de la Limite Centrale

Théorème. $Q_n(\alpha) \to Q(\alpha)$.

Preuve. D'une part $F_n(Q_n(\alpha)) = \frac{\lceil n\alpha \rceil}{n} \to \alpha$.
D'autre part, par Glivenko-Cantelli, $F_n(Q_n(\alpha)) - F(Q_n(\alpha)) \to 0$. □

Théorème (fluctuations, admis). Supposons que X a une densité $f > 0$.

$$\sqrt{n}(Q_n(\alpha) - Q(\alpha)) \overset{loi}{\Longrightarrow} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f^2(Q(\alpha))}\right).$$

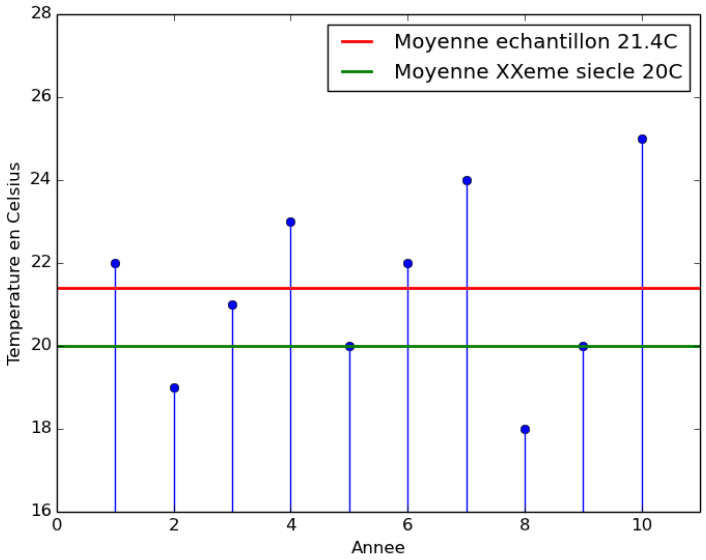
Tests d'hypothèses

On considère une modélisation paramétrique décrit par une famille de probabilités $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ pour $\Theta \subset \mathbb{R}^d$.

Exemple. Modèle gaussien $\mathcal{N}(\mu, \sigma^2)$ avec $\Theta = \{(\mu, \sigma) : \mu \in \{\mu_0, \mu_1\}, \sigma \in [\sigma_0, \sigma_1]\}$

Question : au vu de l'observation $(X_1, \ldots, X_n)$, le paramètre θ du modèle est-il ou non dans un sous-ensemble $\mathbf{H}_0$ (appelé *hypothèse nulle*) ?

Exemple. Température moyenne à Paris au XXème siècle : modélisé par $\mathcal{N}(20, 1)$.



Températures observées sur les 10 dernières années :

▷ peut-on décider que le modèle sur ces dernières années (supposé gaussien) a une moyenne inférieure ou supérieure à 20°C ?

▷ réchauffement ?

Définitions usuelles

▷ **Test d'hypothèse :** règle de décision qui, au vu de l'observation de X, permet de décider si $\theta \in \mathbf{H}_0$ (*hypothèse nulle*) ou si $\theta \in \mathbf{H}_1$ (*hypothèse alternative*) ?

Ici $\mathbf{H}_0$ et $\mathbf{H}_1$ partitionnent l'espace Θ des paramètres.

▷ **Région critique W :**

- si $\mathbf{X} := (X_1, \ldots, X_n) \in W$, on rejette $\mathbf{H}_0$ et on accepte $\mathbf{H}_1$
- si $\mathbf{X} \notin W$, on accepte $\mathbf{H}_0$ et on rejette $\mathbf{H}_1$


Notation : pour simplifier, on note de la même manière la région critique W et l'événement critique $\{\mathbf{X} \in W\}$ associé à la régle de décision.

▷ **Erreur/risque de 1ère espèce :** on rejette $\mathbf{H}_0$ à tort, avec probabilité

$$\mathbb{P}_\theta(W) \text{ pour } \theta \in \mathbf{H}_0 \ .$$

Niveau du test : $\sup_{\theta \in \mathbf{H}_0} \mathbb{P}_\theta(W)$ (typiquement 1%, 5%...)

▷ **Erreur/risque de 2nde espèce :** on rejette $\mathbf{H}_1$ à tort, avec probabilité :

$$\mathbb{P}_\theta(W^c) \text{ pour } \theta \in \mathbf{H}_1 \ .$$

On accepte à raison $\mathbf{H}_1$ avec $\mathbb{P}_\theta(W)$

Puissance du test : $\mathbb{P}_\theta(W)$ $(\theta \in \mathbf{H}_1\ )$

▷ **Objectif :** à un niveau de test fixé, trouver le test de puissance maximale

Dissymétrie entre $\mathbf{H}_0$ et $\mathbf{H}_1$

Exemple modèle gaussien $\mathcal{N}(\mu, \sigma^2)$: test de 2 moyennes avec σ^2 connu

$\mathbf{H}_0 = \{\mu = \mu_0\}$, $\mathbf{H}_1 = \{\mu = \mu_1\}$, $\mu_0 > \mu_1$, $(X_i)_i$ **de loi** $\mathcal{N}(\mu, \sigma^2)$

On propose un test qui accepte $\mathbf{H}_0$ si la moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est suffisamment grande.

➠ région critique $W = \{\bar{X}_n \leq a\}$

▷ **Niveau du test :** pour $\theta \in \mathbf{H}_0$

$$\mathbb{P}_\theta(W) = \mathbb{P}(\mathcal{N}(\mu_0, \frac{\sigma^2}{n}) \leq a) = \mathbb{P}(\mathcal{N}(0,1) \leq \frac{a - \mu_0}{\sigma/\sqrt{n}})$$

Pour un niveau de test α, prendre $a = \mu_0 + \mathcal{N}^{-1}(\alpha)\frac{\sigma}{\sqrt{n}} \overset{\alpha=5\%}{=} \mu_0 - 1.65\frac{\sigma}{\sqrt{n}}$

▷ **Puissance :** pour $\theta \in \mathbf{H}_1$

$$\mathbb{P}_\theta(W) = \mathbb{P}(\mathcal{N}(0,1) \leq \frac{a - \mu_1}{\sigma/\sqrt{n}}) = \mathbb{P}(\mathcal{N}(0,1) \leq \mathcal{N}^{-1}(\alpha) + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}) \underset{n\to+\infty}{\to} 1$$

La puissance du test dépend de la distance entre μ_1 et μ_0, et la taille de l'échantillon.

Encore un peu de terminologie

Définition. On se donne $(W_n)_n$ une suite de régions critiques (une suite de tests) où n est la taille de l'échantillon.

La suite de tests $(W_n)_n$ est dite

- convergente si $\mathbb{P}_\theta(W_n) \underset{n\to+\infty}{\to} 1$, pour $\theta \in \mathbf{H}_1$
- de niveau asympotique α si $\sup_{\theta\in\mathbf{H}_0} \mathbb{P}_\theta(W_n) \underset{n\to+\infty}{\to} \alpha$

Exemple (modèle gaussien). Test de 2 moyennes avec

$$W_n = \{\bar{X}_n \leq \mu_0 + \mathcal{N}^{-1}(\alpha)\frac{\sigma}{\sqrt{n}}\}.$$

Ce test est *exactement* de niveau α, il est convergent.

Modèle gaussien : test de moyenne

$\mathbf{H}_0 = \{\mu = \mu_0\}$ versus $\mathbf{H}_1 = \{\mu \neq \mu_0\}$

Rappels :

- sous $\mathbf{H}_0$, $\zeta_n = \sqrt{n}\frac{(\bar{X}_n - \mu_0)}{\sigma_n} \overset{\text{loi}}{=}$ Student $t(n-1)$
- sous $\mathbf{H}_0$, $\bar{X}_n \overset{p.s.}{\to} \mu_0$

➠ Région critique : $W_n = \{|\zeta_n| > a\}$

Théorème. En prenant $a_n = t_{n-1}^{-1}(1-\alpha/2)$, la suite de tests associés à $(W_n = \{|\zeta_n| > a_n\})_n$ est convergente et de niveau α.

Preuve.

- Niveau du test : $\mathbb{P}_{\mu_0}(|\zeta_n| > a_n) = 2t_{n-1}(-a_n) = 2(1 - t_{n-1}(a_n)) = \alpha$
- Convergence : pour tout $\mu \neq \mu_0$, $\mathbb{P}_\mu(|\zeta_n| > a_n) \to 1$ car
  - $|\zeta_n| \underset{n\to+\infty}{\sim} |\sqrt{n}\frac{(\mu-\mu_0)}{\sigma}| \to +\infty$ p.s. et $a_n \to \mathcal{N}^{-1}(1-\alpha/2)$
  - $\mathbf{1}_{|\zeta_n|>a_n} \overset{p.s.}{\to} 1$ et par convergence dominée $\mathbb{E}(\mathbf{1}_{|\zeta_n|>a_n}) \to 1$.

□

Réchauffement à Paris ?

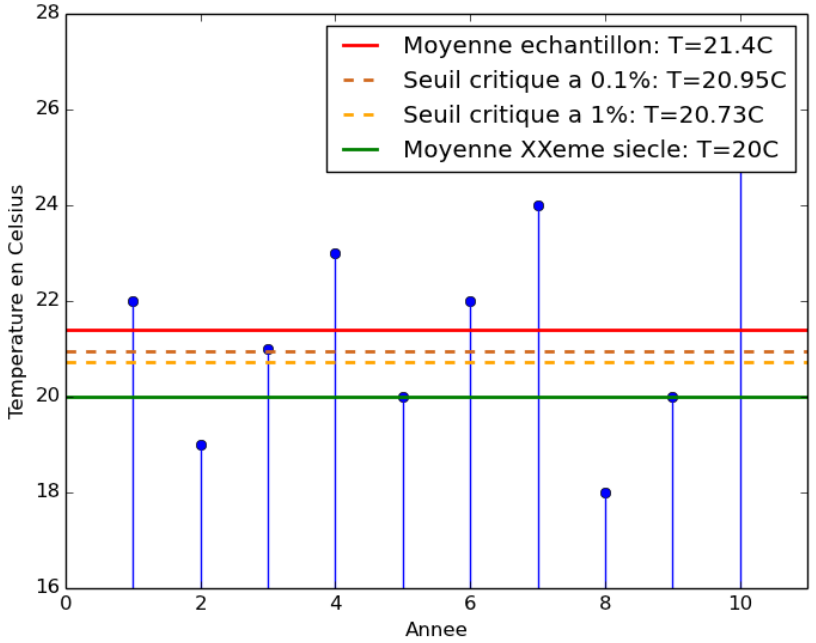
On teste $\mathbf{H}_0 = \{\mu \leq 20\}$ versus $\mathbf{H}_1 = \{\mu > 20\}$, $\sigma^2 = 1$ connue.

On pose le test $W_n = \{\sqrt{n}(\bar{X}_n - 20) \geq \mathcal{N}^{-1}(1-\alpha)\}$.

Proposition (à faire en exercice). C'est un test de niveau α et convergent.

Exemple. Sur 10 données et au niveau $\alpha = 1\%$, $W_n = \{\bar{X}_n \geq 20 + \frac{2.33}{\sqrt{10}} = 20.73\}$.

$x_n^{\text{obs}} = 21.4 > 20.73$ ➠ on rejette $\mathbf{H}_0$ et on accepte $\mathbf{H}_1$



Définition. La p-valeur du test est la probabilité que la statistique de test prenne des valeurs plus défavorables que la valeur observée, càd $\mathbb{P}_\mu(\bar{X}_n \geq x_n^{\text{obs}})$, $\mu \in \mathbf{H}_0$. Ici $p = 4.7 \times 10^{-6}$: on rejette $\mathbf{H}_0$ à tous les niveaux $\alpha \geq p$.

Test du χ^2

Contexte :

- on pose une question avec 3 réponses possibles : 1, 2, 3
- on interroge n personnes indépendamment : les réponses sont X_i.
$$\mathbb{P}(X_i = j) = p_j,\ j = 1, 2, 3.$$
- $\Theta = \{p = (p_1, p_2, p_3) \in]0, 1[^3 : p_1 + p_2 + p_3 = 1\}$
- On veut tester l'hypothèse nulle $\mathbf{H}_0 = \{p = p^0\}$ contre l'hypothèse alternative $\mathbf{H}_1 = \{p \neq p^0\}$

Plus généralement, k réponses possibles. Espace fini.

Fréquence empirique :

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i = j}, \quad 1 \leq j \leq k.$$

Définition. La distance du χ^2 est définie par $\boldsymbol{\zeta_n := n \sum_{j=1}^{k} \frac{(\hat{p}_j - p_j^0)^2}{p_j^0}}$.

Intuition :

- Quantifie la proximité entre $\hat{p}$ et p^0
- Normalisation par n pour tenir compte des fluctuations (TCL) en $1/\sqrt{n}$ pour les grands échantillons

Théorème. 1) La loi de ζ_n converge sous $\mathbf{H}_0$ vers la loi du $\chi^2(k-1)$.
2) Sous $\mathbf{H}_1$, $\zeta_n \overset{p.s.}{\to} +\infty$.

PREUVE. 1) (dans le cas $k = 2$) Admis dans le cas général (basé sur le TCL multidimensionnel).

- $\sqrt{n}(\hat{p}_1 - p_1^0) \overset{\mathcal{L}}{\to} \mathcal{N}(0, p_1^0(1-p_1^0))$
- $\zeta_n = n\left[\frac{(\hat{p}_1 - p_1^0)^2}{p_1^0} + \frac{(1-\hat{p}_1-(1-p_1^0))^2}{1-p_1^0}\right] = n\frac{(\hat{p}_1-p_1^0)^2}{p_1^0(1-p_1^0)} \overset{\mathcal{L}}{\to}$ carré d'une gaussienne standard $\overset{\text{loi}}{=} \chi^2(1)$

2) $\zeta_n \underset{n\to+\infty}{\sim} n\sum_{j=1}^{k} \frac{(p_j - p_j^0)^2}{p_j^0} \overset{p.s.}{\to} +\infty$. □

Corollaire. $W_n = \{\zeta_n > \chi^2_{k-1}(1-\alpha)\}$ définit une suite de tests convergents et de niveau asymptotique α.

Exemple. 100 lancers d'un dé à 6 faces.

j	1	2	3	4	5	6
$\hat{p}_j$	0.2	0.13	0.17	0.12	0.23	0.15

Le dé est-il pipé ?

- $p^0 = (\frac{1}{6}, \ldots, \frac{1}{6})$,
- $k = 6$, $n = 100$
- $\zeta_n^{\text{obs}} = 5.36$
- $\chi^2_{k-1}(1-5\%) = 11.07$

➠ au niveau 5%, on accepte l'hypothèse que le dé est équilibré.

Pour en savoir plus

Estimation d'un jeu avec observation partielle

Deux pièces A et B.

- La pièce A donne P avec probabilité α_A
- La pièce B donne P avec probabilité α_B

Deux individus : joueur, observateur

- n tours de jeu répétés :
  1. Le joueur choisit une pièce : $Z_i \in \{A, B\}$. Il choisit A avec proba β et B avec proba $1 - \beta$.
  2. Il la lance 3 fois.
  3. Il transmet le nombre de P obtenu à un observateur : $X_i \in \{0, 1, 2, 3\}$.

Objectif : l'observateur cherche les paramètres $\theta = (\alpha_A, \alpha_B, \beta)$ à partir des seules observations $X = (X_1, \cdots, X_n)$ (et pas $Z = (Z_1, \ldots, Z_n)$)

Maximum de vraisemblance : $\theta^* = \arg\max_\theta p_\theta(X)$.

Ici, optimisation directe délicate : $p_\theta(X) = \sum_{(z_1,\ldots,z_n)\in\{A,B\}^n} \prod_{i=1}^n p_\theta(X_i, z_i)$

Algorithme EM (*Expectation-Maximization*)

PRINCIPE DE L'ALGORITHME : remplacer $p_\theta(X)$ par une fonction plus facile à maximiser et qui a le même arg max.

Propriétés.

1. $\log p_\theta(X) - \log p_{\theta'}(X) \geq C(\theta, \theta') := \mathbb{E}_{\theta'}\left(\log\left(\frac{p_\theta(X,Z)}{p_{\theta'}(X,Z)}\right) \mid X\right)$.
2. $C(\theta, \theta) = 0$.

PREUVE. 2) est clair. Pour 1), par Bayes,

$$\begin{aligned}
\log p_\theta(X) - \log p_{\theta'}(X) &= \log\left(\sum_z p_\theta(X,z)\right) - \log p_{\theta'}(X) \\
&= \log\left(\sum_z p_{\theta'}(z \mid X)\frac{p_\theta(X \mid z)p_\theta(z)}{p_{\theta'}(z \mid X)}\right) - \sum_z p_{\theta'}(z \mid X)\log p_{\theta'}(X) \\
&\geq \sum_z p_{\theta'}(z \mid X)\log\left(\frac{p_\theta(X \mid z)p_\theta(z)}{p_{\theta'}(z \mid X)p_{\theta'}(X)}\right) \quad \text{(inégalité de Jensen)} \\
&= \sum_z p_{\theta'}(z \mid X)\log\left(\frac{p_\theta(X,z)}{p_{\theta'}(X,z)}\right) = \mathbb{E}_{\theta'}\left[\log\left(\frac{p_\theta(X,Z)}{p_{\theta'}(X,Z)}\right) \mid X\right].
\end{aligned}$$

□

Algorithme EM en deux étapes

- **Expectation (E) :** calculer

$$\theta \mapsto \hat{C}(\theta, \theta_k) = \mathbb{E}_{\theta_k}\left[\log p_\theta(X, Z) \mid X\right]$$

car $C(\theta, \theta_k) - \hat{C}(\theta, \theta_k) = \mathbb{E}_{\theta_k}[\log p_{\theta_k}(X, Z) \mid X]$ est indépendant de θ.

- **Maximisation (M) :**

$$\theta_{k+1} := \arg\max_\theta \hat{C}(\theta, \theta_k) = \arg\max_\theta C(\theta, \theta_k)$$

Théorème. Etant donné θ_0, le choix $\theta_{k+1} := \arg\max_\theta C(\theta, \theta_k)$ assure que $(p_{\theta_k}(X))_k$ est une suite croissante en k.

Preuve. $\log p_{\theta_{k+1}}(X) - \log p_{\theta_k}(X) \geq C(\theta_{k+1}, \theta_k) \geq C(\theta_k, \theta_k) \geq 0.$ □

➠ $(\theta_k)_k$ converge vers un maximum local ? global ?

Mise en œuvre sur l'exemple du jeu

Etape E :

$$\hat{C}(\theta,\theta_k) := \mathbb{E}_{\theta_k}\left[\log p_\theta\left((X_1,\cdots,X_n),(Z_1,\cdots,Z_n)\right) \mid X_1,\cdots X_n\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}_{\theta_k}\left[\log p_\theta(X_i,Z_i) \mid X_i\right] \quad \text{(indépendance des } (X_i,Z_i)_i\text{)}$$

$$= \sum_{i=1}^{n} p_{\theta_k}(A \mid X_i)\log p_\theta(X_i,A) + p_{\theta_k}(B \mid X_i)\log p_\theta(X_i,B),$$

$$= \sum_{i=1}^{n} p_{\theta_k}(A \mid X_i)\log\left(\beta\binom{3}{X_i}\alpha_A^{X_i}(1-\alpha_A)^{3-X_i}\right) + p_{\theta_k}(B \mid X_i)\log\left((1-\beta)\binom{3}{X_i}\alpha_B^{X_i}(1-\alpha_B)^{3-X_2}\right).$$

avec

$$p_{\theta_k}(A \mid X_i) = \frac{\beta_k p_{\theta_k}(X_i \mid A)}{\beta_k p_{\theta_k}(X_i \mid A) + (1-\beta_k)p_{\theta_k}(X_i \mid B)} \quad \text{(Bayes)},$$

$$p_{\theta_k}(B \mid X_i) = \frac{(1-\beta_k) p_{\theta_k}(X_i \mid B)}{\beta_k p_{\theta_k}(X_i \mid A) + (1-\beta_k)p_{\theta_k}(X_i \mid B)} \quad \text{(Bayes)}.$$

Etape M : le maximum de $\hat{C}(\theta, \theta_k)$ est $\theta_{k+1} = (\alpha_{A,k+1}, \alpha_{B,k+1}, \beta_{k+1})$ donné par

- $\alpha_{\mathbf{A,k+1}} = \dfrac{\sum_{\mathbf{i=1}}^{\mathbf{n}} \frac{\mathbf{X_i}}{\mathbf{3}} \mathbf{p}_{\theta_\mathbf{k}}(\mathbf{A} \mid \mathbf{X_i})}{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{p}_{\theta_\mathbf{k}}(\mathbf{A} \mid \mathbf{X_i})}$: moyenne des fréquences de P pondérée par la probabilité a posteriori d'avoir choisi la pièce A,
- $\alpha_{\mathbf{B,k+1}} = \dfrac{\sum_{\mathbf{i=1}}^{\mathbf{n}} \frac{\mathbf{X_i}}{\mathbf{3}} \mathbf{p}_{\theta_\mathbf{k}}(\mathbf{B} \mid \mathbf{X_i})}{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{p}_{\theta_\mathbf{k}}(\mathbf{B} \mid \mathbf{X_i})}$,
- $\beta_{\mathbf{k+1}} = \dfrac{\mathbf{1}}{\mathbf{n}} \displaystyle\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{p}_{\theta_\mathbf{k}}(\mathbf{A} \mid \mathbf{X_i})$.

A la limite $k \to +\infty$ (si convergence vers le max global)

▷ On retrouve $p_{\theta^*}(A \mid X_i)$ la probabilité a posteriori du choix de la pièce A au tirage i sachant l'observation X_i, pour le paramètre θ^* le plus vraisemblable.

▷ Classification des pièces choisies $Z_i = A$ ou B avec certaines probabilités.

Exemple numérique. Valeurs exactes : $\alpha_A = 0.3$, $\alpha_B = 0.5$, $\beta = 0.7$, $n = 503$

```
Initialisation :  alphaA=0.450, alphaB=0.550, beta=0.500
 Iteration   10:  alphaA=0.285, alphaB=0.404, beta=0.548
.....................................................
 Iteration  990:  alphaA=0.289, alphaB=0.467, beta=0.720
 Iteration 1000:  alphaA=0.289, alphaB=0.467, beta=0.722
============================================================
Probabilite a posteriori (exacte) de piece A sachant Xi:
 Xi=0 -> 86.5% | Xi=1 -> 73.3% | Xi=2 -> 54.0% | Xi=3 -> 33.5% |
Probabilite a posteriori (estimee) de piece A sachant Xi:
 Xi=0 -> 86.1% | Xi=1 -> 74.1% | Xi=2 -> 57.1% | Xi=3 -> 38.2% |
Probabilite a posteriori (aveugle) de piece A sachant Xi:
 Xi=0 -> 50.0% | Xi=1 -> 50.0% | Xi=2 -> 50.0% | Xi=3 -> 50.0% |
============================================================
```