Chapitre 9. Tests d'hypothèses. Ajustement

Contenu

1 Introduction

Il arrive que l'expérimentateur dispose d'une hypothèse ("idée dominante du moment") et qu'il réalise une expérience destinée à remettre en question cette hypothèse. Pour être concret, supposons que l'on vous propose le jeu suivant. Vous gagnez 1$ si une pièce de monnaie attérrit sur Pile, mais vous perdez 1$ si elle attérrit sur Face. Si la pièce est équilibrée (non biaisée), vous pourriez vouloir jouer, sinon peut-être pas. Il faudra donc décider si la pièce est équilibrée (hypothèse initiale) ou non. En d'autres termes, il faudra décider si $\theta = p \geq 1/2$ ($p =$ probabilité d'obtenir Pile dans un lancer) ou non ($\theta < 1/2$). Cette décision sera basée sur l'observation d'une séquence de jets (échantillon). Vous allez donc conclure que la pièce n'est pas équilibrée si vous voyez trop de Faces appaître dans votre échantillon. Mais à partir de combien de Faces allez-vous allez-vous situer ce "trop" ? Il s'agit donc de trouver une valeur critique.

En général, l'hypothèse porte sur un modèle paramétrique :

$$X \backsim P_\theta, \quad \theta \in \Theta \quad \left(\Theta \subset \mathbb{R}^k\right)$$

($k = 1$ la plupart du temps). L'espace des paramètres des lois de probabilité est partitionné en deux ensembles disjoints :

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset$$

Hypothèse nulle :

$$H_0 : \theta \in \Theta_0$$

Hypothèse alternative (ou contre-hypothèse) :

$$H_1 : \theta \in \Theta_1$$

Souvent, l'hypothèse nulle ne représente aucun changement ou aucune différence (statu quo, absence d'effet d'un traitement), tandis que l'hypothèse alternative représente un changement ou une différence. L'hypothèse alternative est parfois appelée l'hypothèse du chercheur. On se donne alors un échantillon aléatoire $X_1, \cdots, X_n$ issu de la même loi que celle de X qui nous permettra de décider d'accepter ou de rejeter H_0.

Dans un certain sens, le problème des tests d'hypothèses est plus simple que celui de l'estimation (à laquelle nous avons consacré beaucoup de temps) puisqu'on ne cherche pas à estimer le paramètre θ, on veut juste savoir à quel ensemble il appartient. Alors, pourquoi présenter cette matière (tests d'hypothèses) après celle de l'estimation (plus difficile) ? La raison principale est que les concepts de vraisemblance, exhaustivité, propriétés asymptotiques, etc. sont plus naturellement compris dans le contexte de l'estimation.

Une hypothèse est dite **simple** si elle correspond à une seule loi de probabilité (une seule distribution). Elle est dite **composée** (composite) si elle contient plus d'une loi.

Exemple 1 $X \backsim Bin(n,p)$

$$\begin{aligned} H_0 &: p = 0.5 \qquad \textit{(simple)} \\ H_1 &: p = 0.7 \qquad \textit{(simple)} \end{aligned}$$

Exemple 2 $X \backsim Bin(n,p)$

$$\begin{aligned} H_0 &: p = 0.5 \qquad \textit{(simple)} \\ H_1 &: p > 0.5 \qquad \textit{(composée)} \end{aligned}$$

Exemple 3 $X \backsim N(\mu, \sigma_0^2)$, *$\sigma_0$ connue*

$$\begin{aligned} H_0 &: \mu = 0 \qquad \textit{(simple)} \\ H_1 &: \mu = 15 \qquad \textit{(simple)} \end{aligned}$$

$$\begin{aligned} H_0 &: \mu = 15 && \textit{(simple)} \\ H_1 &: \mu > 15 && \textit{(composée : unilatérale à droite)} \\ H_1 &: \mu < 15 && \textit{(composée : unilatérale à gauche)} \\ H_1 &: \mu \neq 15 && \textit{(composée : bilatérale)} \end{aligned}$$

Remarque 1 *Une définition formelle d'un test peut être présentée comme suit. C'est une fonction (de décision)*

$$\begin{aligned} d &: \Omega \to \{0,1\} \\ X &= (X_1, \ldots, X_n) \to d(X) \end{aligned}$$

où $d(X) = 1$ est associée au rejet de H_0 et

$$R = \{x \in \mathbb{R}^n : d(x) = u1\}$$

est la région du rejet (critique) du test (de H_0) et

$$R^c = \{x \in \mathbb{R}^n : d(x) = 0\}$$

(non rejet de H_0) ne signifie pas nécessairement acceptation de H_0, mais plutôt acceptation de H_0, faute de mieux : on n'a pas pu montrer que H_0 est fausse. Noter que la statistique (fonction de décision) $d(X)$ est une variable de Bernoulli. Dans la pratique, on dispose d'un seul vecteur d'observations $x = (x_1, \ldots, x_n)$.

Remarque 2 *Dans le cas de l'approche bayésienne : rapport de vraisemblance et distribution a priori (traités dans le prochain chapitre).*

2 Approche de Neyman-Pearson

2.1 Région critique. Puissance d'un test. Valeur-p

Dans l'approche de Neyman-Pearson (N-P), la décision de rejeter H_0 en faveur de H_1 est prise à l'aide d'une statistique $T(X_1, \ldots, X_n)$, appelée **statistique du test**. Supposons que la décision de rejeter H_0 est prise si

$$T(x_1, \ldots, x_n) \geq c$$

ou $x = (x_1, \ldots, x_n)$ est un vecteur d'observations. L'ensemble

$$R = \{x \in \mathbb{R}^n : T(x) \geq c\}$$

est la **région critique du test**. On rencontre alors deux types d'erreurs (de décision) :

		Θ : état de la nature	
		Θ_0 (H_0 vraie)	Θ_1 (H_0 fausse)
observé : $x \in$	R	Erreur de 1ère espèce (α)	Bonne décision ($1-\beta$)
	R^c	Bonne décision ($1-\alpha$)	Erreur de 2eme espèce (β)

Les probabilités associées sont :

$$\alpha(\theta) = P_\theta(x \in R), \quad \theta \in \Theta_0$$
$$\beta(\theta) = P_\theta(x \in R^c), \quad \theta \in \Theta_1$$

Ainsi,

$$\alpha = P(\text{rejet de } H_0 \,|H_0 \text{ vraie})$$

est le niveau de signification du test (probabilité de rejeter H_0 alors que H_0 est vraie),

$$\beta = P(\text{acceptation de } H_0 \,|H_0 \text{ fausse})$$

(probabilité de ne pas rejeter H_0 alors que H_0 est fausse) et

$$(1-\beta)(\theta) = P_\theta(x \in R), \quad \theta \in \Theta_1$$

est **la puissance du test**. Généralement, α est fixé par le chercheur (souvent $\alpha = 5\%$), mais β est compliqué car ce n'est pas un nombre fixe (H_1 est composite).

Exemple 4 *Un test médical est dit vrai négatif (VN) s'il indique qu'une personne n'est pas malade quand elle ne l'est pas et faux négatif (FN) lorsqu'il indique qu'une personne n'est pas malade aolors qu'elle l'est (cela correspond à l'erreur de première espèce). De même, le test est dit vrai positif (VP) lorsqu'il indique correctement qu'une personne est malade et faux positif (FP) lorsqu'il fait l'indication inverse (erreur de seconde espèce).*

Exemple 5 *Une autre situation similaire est un système de justice où l'on peut condamner à tort ou à raison une personne coupable ou innocente (quelles sont les quatres décisions possibles ?).*

L'approche de Neyman-Pearson rompt la symétrie entre H_0 et H_1 (on ne traite pas les deux risques de la même manière : compromis).

Remarque 3 *Il y a d'autres approches possibles : trouver $T(X_1, \dots, X_n)$ qui minimise $\alpha + \beta$, par exemple.*

Exemple 6 $X \backsim Bin(n = 10, p)$.

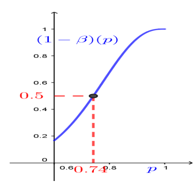
$$H_0 : p = 0.5$$
$$H_1 : p > 0.5$$

Supposons que $R = \{x : x \geq 8\}$. *Alors*

$$\alpha = P(X \geq 8 \,|p = 0.5) = 0.0547$$

On a

$$\begin{aligned}(1-\beta)(p) &= P(X \geq 8 \,|p) \\ &= \binom{10}{8} p^8 (1-p)^2 + \binom{10}{9} p^9 (1-p) + \binom{10}{10} p^{10} (1-p)^0 \\ &= 36p^{10} - 80p^9 + 45p^8\end{aligned}$$



p	0.6	0.7	0.8	0.9	0.99	0.999
$1-\beta$	0.167 29	0.382 78	0.677 80	0.929 81	0.999 89	1. 000 00

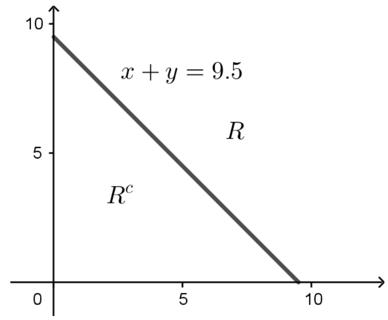
Pour obtenir une plus grande puissance (pour p proche de $0.5 : 0.6$*), il faut augmenter* n.

Exemple 7 $X \backsim Exp(\theta) : f(x\,|\theta) = \frac{1}{\theta} e^{-x/\theta}$, $x, \theta > 0$

$$H_0 : \theta = 2 \quad contre \quad H_1 : \theta = 4$$

$n = 2$ *observations. Supposons que* $R = \left\{x = (x_1, x_2) \in \mathbb{R}^2 : 9.5 < x_1 + x_2 < \infty\right\}$. *On a*

$$\begin{aligned}\alpha &= P(X_1 + X_2 > 9.5 \,|\theta = 2) \\ &= 1 - \int_0^{9.5} \int_0^{9.5 - x_2} \frac{1}{2} e^{-x_1/2} \frac{1}{2} e^{-x_2/2} dx_1 dx_2 \\ &= 1 - \frac{1}{4} \int_0^{9.5} e^{-x_2/2} \left(\int_0^{9.5 - x_2} e^{-x_1/2} dx_1 \right) dx_2 \\ &= 1 - \frac{1}{4} \int_0^{9.5} e^{-x_2/2} \left(2 - 2e^{\frac{1}{2}x_2} e^{-\frac{19}{4}} \right) dx_2 \\ &\overset{Calculs}{=} \frac{23}{4} e^{-\frac{19}{4}} \approxeq 0.05 \\ &\overset{R}{=} 1 - pgamma(9.5, rate = 0.5, shape = 2) = 0.04974725\end{aligned}$$



Maintenant,

$$\begin{aligned}
(1-\beta)_{|\theta=4} &= P\left(X_1+X_2>9.5\,|\theta=4\right)\\
&= 1-\int_0^{9.5}\int_0^{9.5-x_2}\frac{1}{4}e^{-x_1/4}\frac{1}{4}e^{-x_2/4}dx_1dx_2\\
&\overset{Calculs}{=} \frac{27}{8}e^{-\frac{19}{8}} \approxeq 0.31
\end{aligned}$$

<u>Note sur le calcul</u>. Sous H_0 ($\theta=2$),

$$\begin{aligned}
X &\backsim \chi^2_2 \Rightarrow X_1+X_2 \backsim \chi^2_4 \Rightarrow P\left(X_1+X_2>9.5\right)\\
&= 1-P\left(X_1+X_2\leq 9.5\right) \overset{Table}{=} 0.05\\
&\overset{R}{=} 1-pchisq(9.5,4)=0.04974725
\end{aligned}$$

Sous H_1 ($\theta=4$),

$$\begin{aligned}
\frac{X}{2} &\backsim \chi^2_2 \Rightarrow \frac{X_1+X_2}{2} \backsim \chi^2_4\\
&\Rightarrow P\left(X_1+X_2>9.5\right) = P\left(\frac{X_1+X_2}{2}>4.75\right)\\
&= \int_{4.75}^{\infty}\frac{1}{4}ue^{-u/2}du \approxeq 0.31 \text{ (ou table)}\\
&\overset{R}{=} 1-pchisq(9.5/2,4)=0.3139239
\end{aligned}$$

Exemple 8 *Test pour une moyenne (grand échantillon). Soit X une variable aléatoire de moyenne μ et de variance σ^2. On veut tester*

$$H_0:\mu=\mu_0 \quad \text{contre} \quad H_1:\mu>\mu_0$$

(μ_0 constante donnée). On prend un échantillon iid $X_1\ldots,X_n$ issu de la loi de X. Puisque $\overline{X}$ est un estimateur sans biais de μ, une règle de décision naturelle est de rejeter H_0 si $\overline{x}$ est beaucoup plus grand que μ_0. Comme il arrive souvent, on ne connaît pas la loi de $\overline{X}$ (ou elle est trop compliquée). Si l'échantillon est de grande taille, on peut invoquer TCL et utiliser le fait que la statistique

$$T=\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim N\left(0,1\right) \text{ approx.}$$

Ceci nous mène à la région critique (de rejet de H_0)

$$R = \left\{ x \in \mathbb{R}^n : \frac{\overline{X} - \mu_0}{s/\sqrt{n}} \geq z_\alpha \right\}$$

Ainsi, le test rejette H_0 si $\overline{x}$ est plus grand que μ_0 d'au moins $z_\alpha \dfrac{s}{\sqrt{n}}$ unités. Pour approximer la puissance du test, utilisons encore une fois TCL (avec σ à la place de S)

$$\begin{aligned} 1 - \beta(\mu) &= P\left(\overline{X} \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\right) = P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_\alpha\right) \\ &\approxeq 1 - \Phi\left(z_\alpha + \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right) = \Phi\left(-z_\alpha - \sqrt{n}\frac{\mu_0 - \mu}{\sigma}\right) \end{aligned}$$

Si nous avons une idée de la valeur approximative de σ, nous pourrons alors calculer approximativement $1 - \beta(\mu)$. C'est bien sûr une fonction strictement croissante de μ.

Noter que $X \sim N(\mu, \sigma^2)$, nous savons alors que $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$ et le test rejette H_0 si $t > t_{\alpha,n-1}$. Code R : t.test(x,mu=mu0,alt="greater").

2.1.1 Valeur-p (p-value)

Soit $T = T(X_1, \cdots, X_n)$ la statistique utilisée pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$ (ou $\theta < \theta_0$). Le test rejette H_0 pour $T \geq c$ (ou $T \leq c$) pour une certaine valeur c (voir région critique plus haut). Soit $T = t = T((x_1, \cdots, x_n))$ pour un échantillon particulier. Alors, on définit la valeur-p par

$$p - value(t) = P_{\theta_0}(T \leq t) \ \text{(resp. } p - value(t) = P_{\theta_0}(T \geq t)\text{)}$$

Que signifie ceci ? Ayant observé cette valeur particulière de la statistique du test, on se pose la question de savoir quelle est les probabilité (sous H_0) de faire pire (obtenir une observation au moins aussi extrême que le t observé). Si on obtient une petite probabilité (typiquement inférieure à α), ceci nous conforte dans notre idée de rejeter l'hypothèse nulle. C'est donc une sorte de mesure de la plausibilité de l'hypothèse nulle. Attention, ceci n'est pas la probabilité que H_0 est vraie ou fausse (car il n'y a pas de probabilité pour H_0 : elle est vraie ou fausse avec certitude), ni la probabilité qu'on commet (ou non) une erreur de type 1 (α, calculé en amont, avant l'observation). Elle ne fait que mesurer combien les données effectivement observées sont en faveur (ou non) de l'hypothèse nulle.

Exemple 9 *Soit $X_1, \cdots, X_n$ iid $\sim N(\mu, \sigma^2)$ (σ^2 inconnu) et on veut tester*

$$H_0 : \mu = 30 \text{ contre } H_1 : \mu < 30$$

Soit

$$T = \sqrt{n}\frac{\overline{X} - 30}{S}$$

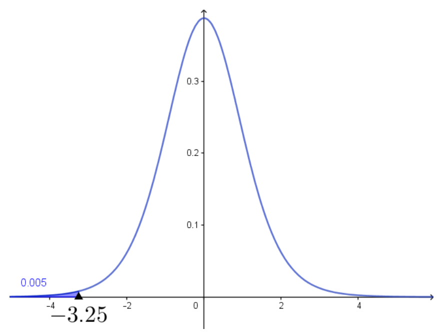
la statistique du test (on verra plus loin comment l'obtenir). Si on a obtenu $\overline{x} = 26.4$, $s = 3.5$ dans un échantillon de taille $n = 10$ donné, alors

$$t = \sqrt{10}\frac{26.4 - 30}{3.5} = -3.25$$

Étant donné le "<" dans H_1, la valeur-p de ce test est alors

$$p - value\,(-3.25) = P\left(T \leq -3.25\,|H_0\right) = P\left(t_{10-1} \leq -3.25\right) \overset{table}{\approx} 0.005$$
$$\overset{R}{=} pt(-3.25, 9) = 0.004998685$$

Ici t_{10-1} est une variable de Student à 9 degrés de liberté.



2.2 Vraisemblance relative

Exemple 10 *Revenons à $X \backsim Bin\,(n = 10, p)$ et posons*

$$p_0\,(x) = P\,(X = x\,|p = 0.5\,) = \binom{10}{x}\frac{1}{2^{10}}$$
$$p_1\,(x) = P\,(X = x\,|p = 0.6\,) = \binom{10}{x}0.6^x \times 0.4^{10-x}$$

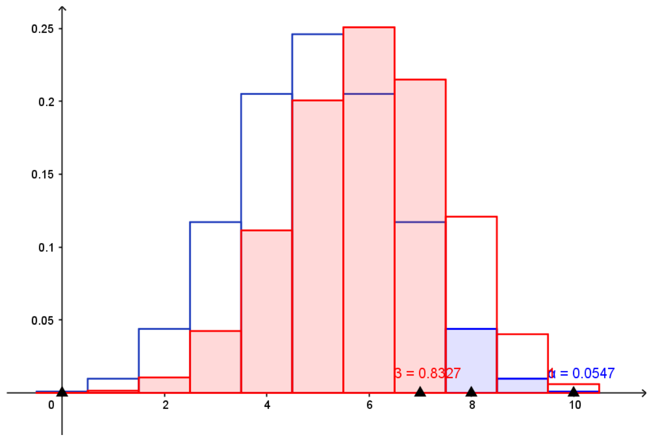
Examinons les différents possibilités

X	0	1	2	3	4	**5**	**6**	7	8	9	10
$10^3 p_0\,(x)$	0.9 8	9.78	43.9	117.2	205.1	246.1	205.1	117.2	43.9	9.78	0.98
$10^3 p_1\,(x)$	0.11	1.5 7	10.6	42.5	111.9	200.7	250.9	215.0	120.9	40.3	6.1
$p_0\,(x)\,/p_1\,(x)$	9. 3	6. 2	4.1	2.6	1. 8	**1. 2**	**0.8**	0.5	0.4	0.2	0.16

*Pour une valeur x observée, la **vraisemblance relative** de H_0 par rapport à H_1 est mesurée par $p_0\,(x)\,/p_1\,(x)$. On voit que plus x augmente, moins H_0 est vraisemblable. Pour $R = \{x : x \geq 8\}$, on a vu que*

$$\alpha = P\,(R\,|H_0\,) = P\,(X \geq 8\,|p = 0.5\,) = 0.0547$$
$$\beta = P\,(R^c\,|H_1\,) = P\,(X \leq 7\,|p = 0.6\,) = 0.8327$$

On contrôle l'erreur de type I : on choisit α le plus petit possible (autour de 1%, 5%). On peut faire baisser β en augmentant n.



Exemple 11 *Soit $X \backsim Bin\,(n=3,\theta)$*

$$H_0 : \theta = 0.5$$
$$H_1 : \theta = 0.75$$

X	0	1	2	3
$H_0 : P\left(X\,|\theta=0.5\right)$	$1/8$	$3/8$	$3/8$	$1/8$
$H_1 : P\left(X\,|\theta=0.75\right)$	$1/64$	$9/64$	$27/64$	$27/64$
p_0/p_1	8	$8/3$	$8/9$	$8/27$

Test de Neyman-Pearson (N-P) : rejet de H_0 si $p\left(x\,|\theta_0\right)/p\left(x\,|\theta_1\right) < c$ ($c =$ constante positive). Ceci revient à rejeter si

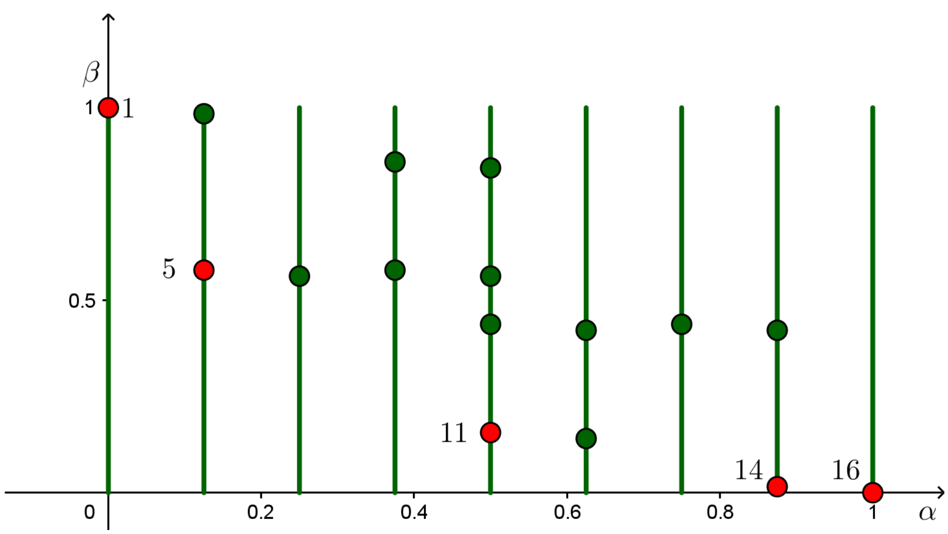
$$\binom{3}{x}\frac{1}{2^3} < c\binom{3}{x}0.75^x \times 0.25^{3-x}$$
$$\Rightarrow \frac{1}{8} < c\frac{3^x}{64} \Rightarrow 3^x > 8/c$$
$$\Rightarrow x > \frac{\ln\left(8/c\right)}{\ln 3} = c_1$$

Ceci nous mène aux 5 régions suivantes (selon la valeur de c_1) :

$$R : \{0,1,2,3\}\,,\{1,2,3\}\,,\{2,3\}\,,\{3\}\,,\emptyset$$

Examinons tous les tests possibles ($2^4 = 16$ tests : nombre de sous ensembles de $X\left(\Omega\right) = \{0,1,2,3\}$).

Test	R	$\alpha = P(R\,\vert\theta = 0.5)$	$\beta = P(R^c\,\vert\theta = 0.75)$
$\boxed{1}$	$\emptyset$	0	1
2	$\{0\}$	1/8	63/64
3	$\{1\}$	3/8	55/64
4	$\{2\}$	3/8	37/64
$\boxed{5}$	$\{3\}$	1/8	37/64
6	$\{0,1\}$	4/8	54/64
7	$\{0,2\}$	4/8	36/64
8	$\{0,3\}$	2/8	36/64
9	$\{1,2\}$	6/8	28/64
10	$\{1,3\}$	4/8	28/64
$\boxed{11}$	$\{2,3\}$	4/8	10/64
12	$\{0,1,2\}$	7/8	27/64
13	$\{0,1,3\}$	5/8	27/64
$\boxed{14}$	$\{1,2,3\}$	7/8	1/64
15	$\{0,2,3\}$	5/8	9/64
$\boxed{16}$	$\{0,1,2,3\}$	1	0



Les tests $1, 5, 11, 14$ *et* 16 *sont les tests de N-P. Ils sont les meilleurs parmi tous les tests de leur niveau* α*. On note qu'il y a d'autres tests qui ne sont pas N-P.*

2.3 Tests les plus puissants

Approche de N-P.
Supposons que nous soyons en présence de deux hypothèses **simples**.

1. On fixe $\alpha = P(R|\theta_0)$, α petit.
2. On cherche R tel que $\beta = P(R^c|\theta_1)$ soit la plus petite parmi tous les tests vérifiant $P(R|\theta_0) = \alpha$.

On a le

Théorème 1 *(Lemme : N-P). Suffisant. Le **test le plus puissant de niveau** α. Soit $X_1, \ldots, X_n \backsim f(x|\theta)$ $(x = (x_1, \ldots, x_n))$ avec les hypothèses simples*

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta = \theta_1$$

Soit $\alpha \in]0;1[$ fixé tel qu'il existe $c > 0$ avec

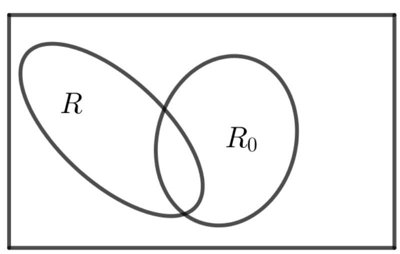
$$P_{\theta_0}(f(x|\theta_0) < cf(x|\theta_1)) = \alpha$$

*Alors, parmi tous les tests de niveau $\leq \alpha$, celui qui **maximise la puissance** $1-\beta$ est de la forme*

$$\begin{aligned} R_0 &= \{x = (x_1, \ldots, x_n) : f(x|\theta_0) < cf(x|\theta_1)\} \\ &= \left\{x = (x_1, \ldots, x_n) : \frac{L(\theta_0|x)}{L(\theta_1|x)} < c\right\} \end{aligned}$$

Preuve. *On donnera une preuve dans le cas continu. Soit R un test avec $P_{\theta_0}(R) \leq \alpha$. Si $R = R_0$, il n'y a rien à prouver. Sinon (si $R \neq R_0$), on a les unions disjointes*

$$\begin{aligned} &R_0 = (R_0 \cap R) \cup (R_0 \cap R^c) \\ &\text{et } R = (R \cap R_0) \cup (R \cap R_0^c) \end{aligned}$$



et

$$
\begin{aligned}
c\left(P_{\theta_1}\left(R_0^c\right)-P_{\theta_1}\left(R^c\right)\right) &= c\left(P_{\theta_1}\left(R\right)-P_{\theta_1}\left(R_0\right)\right) \qquad \textit{(complémentatition)}\\
&= c\int_R f\left(x\left|\theta_1\right.\right)dx - c\int_{R_0} f\left(x\left|\theta_1\right.\right)dx\\
&= \int_{R\cap R_0^c} \underbrace{cf\left(x\left|\theta_1\right.\right)}_{\leq f\left(x\left|\theta_0\right.\right)}dx - \int_{R_0\cap R^c} \underbrace{cf\left(x\left|\theta_1\right.\right)}_{>f\left(x\left|\theta_0\right.\right)}dx\\
&\leq \int_{R\cap R_0^c} f\left(x\left|\theta_0\right.\right)dx - \int_{R_0\cap R^c} f\left(x\left|\theta_0\right.\right)dx\\
&+ \int_{R\cap R_0} f\left(x\left|\theta_0\right.\right)dx - \int_{R_0\cap R} f\left(x\left|\theta_0\right.\right)dx\\
&= \int_R f\left(x\left|\theta_0\right.\right)dx - \int_{R_0} f\left(x\left|\theta_0\right.\right)dx \leq \alpha-\alpha = 0
\end{aligned}
$$

La dernière égalité vient des relations de Chasles

$$\int_{R\cap R_0^c} + \int_{R\cap R_0} = \int_R$$

et

$$\int_{R_0\cap R^c} + \int_{R_0\cap R} = \int_{R_0}$$

■

2.3.1 Quelques exemples

Exemple 12 *Soit* $X_1,\ldots,X_n iid \backsim N\left(\mu,\sigma_0^2\right)$, σ_0^2 *connue.*

$$H_0 : \mu=\mu_0 \text{ contre } H_1 : \mu=\mu_1 \ \left(\text{avec } \mu_1>\mu_0\right)$$

On a

$$\begin{aligned}
\frac{L(\mu_0 \,|x)}{L(\mu_1 \,|x)} = \frac{f(x\,|\mu_0)}{f(x\,|\mu_1)} &= \frac{(2\pi)^{-n/2}\,\sigma_0^{-n}\exp\left[-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\mu_0)^2\right]}{(2\pi)^{-n/2}\,\sigma_0^{-n}\exp\left[-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\mu_1)^2\right]} \\
&= \exp\left[-\frac{1}{2\sigma_0^2}\left(\sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\mu_1)^2\right)\right] < c \\
&\Leftrightarrow \sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\mu_1)^2 > c' \\
&\Leftrightarrow \sum_{i=1}^n \left[\left(\mu_0^2 - 2\mu_0 x_i + x_i^2\right) - \left(\mu_1^2 - 2\mu_1 x_i + x_i^2\right)\right] > c' \\
&\Leftrightarrow \sum_{i=1}^n \left[\mu_0^2 - \mu_1^2 - 2x_i(\mu_0-\mu_1)\right] > c' \\
&\Leftrightarrow 2n(\mu_1-\mu_0)\,\overline{x} + \underbrace{n\left(\mu_0^2-\mu_1^2\right)}_{Cte} > c' \\
&\Leftrightarrow \boxed{\overline{x} > c''}
\end{aligned}$$

On veut obtenir le test de N-P de niveau α (petit). On choisit c'' de sorte que $P_{\mu_0}\left(\overline{X} > c''\right) = \alpha$:

$$\frac{\sqrt{n}\left(\overline{X}-\mu_0\right)}{\sigma_0} \backsim N(0,1) \qquad \text{(sous } H_0 : \text{distribution nulle)}$$

Ainsi,

$$P\left(\frac{\sqrt{n}\left(\overline{X}-\mu_0\right)}{\sigma_0} > z_\alpha\right) = \alpha \Leftrightarrow P\left(\overline{X} > \underbrace{\mu_0 + \frac{\sigma_0}{\sqrt{n}} z_\alpha}_{c''}\right) = \alpha$$

Donc

$$\boxed{R = \left\{x \in \mathbb{R}^n : \overline{x} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_\alpha\right\}}$$

C'est le test de N-P de niveau α.

Exemple 13 *Soit $X_1, \ldots, X_n$ iid $\backsim Exp(\theta)$, $f(x\,|\theta) = \frac{1}{\theta}e^{-x/\theta}$ ($x > 0$)*

$$\begin{aligned}
&H_0 : \theta = \theta_0 \\
&H_1 : \theta = \theta_1 \qquad \text{(avec } \theta_1 > \theta_0)
\end{aligned}$$

On a

$$\frac{f(x\,|\theta_0)}{f(x\,|\theta_1)} = \frac{\theta_0^{-n}\exp\left(-\sum_{i=1}^n x_i/\theta_0\right)}{\theta_1^{-n}\exp\left(-\sum_{i=1}^n x_i/\theta_1\right)} = \left(\frac{\theta_0}{\theta_1}\right)^{-n}\exp\left[-n\overline{x}\left(\frac{1}{\theta_0}-\frac{1}{\theta_1}\right)\right] < c$$

$$\Leftrightarrow \underbrace{-n\ln\frac{\theta_0}{\theta_1}}_{cte} - n\overline{x}\left(\frac{1}{\theta_0}-\frac{1}{\theta_1}\right) < c'$$

$$\Leftrightarrow \overline{x}\underbrace{\left(\frac{1}{\theta_0}-\frac{1}{\theta_1}\right)}_{>0} > c'' \Leftrightarrow \overline{x} > c''' \Leftrightarrow \boxed{\sum_{i=1}^n x_i > cte = K}$$

Maintenant, sous H_0, soit α petit donné. On a

$$2\sum_{i=1}^n \frac{X_i}{\theta_0} \backsim \chi^2_{2n} \qquad \textit{(utiliser MGF par ex.)}$$

$$\Rightarrow \alpha = P\left(\sum_{i=1}^n X_i > K\,|\theta_0\right) = P\left(2\sum_{i=1}^n \frac{X_i}{\theta_0} > \frac{2K}{\theta_0}\,|\theta_0\right)$$

On prend donc

$$\boxed{K = \frac{\theta_0}{2}\chi^2_{2n,\alpha}}$$

Exemple 14 *Soit $X_1,\ldots,X_n$ iid $\backsim N\left(0,\sigma^2\right)$*

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 = \sigma_1^2 \qquad \left(\textit{avec } \sigma_1^2 > \sigma_0^2\right)$$

$$\frac{f(x\,|H_0)}{f(x\,|H_1)} = \frac{(2\pi)^{-n/2}\sigma_0^{-n}\exp\left[-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2\right]}{(2\pi)^{-n/2}\sigma_1^{-n}\exp\left[-\frac{1}{2\sigma_1^2}\sum_{i=1}^n x_i^2\right]}$$

$$= \exp\left[-\left(\frac{1}{2\sigma_0^2}-\frac{1}{2\sigma_1^2}\right)\sum_{i=1}^n x_i^2\right] < c$$

$$\Leftrightarrow \underbrace{-\left(\frac{1}{2\sigma_0^2}-\frac{1}{2\sigma_1^2}\right)}_{>0}\sum_{i=1}^n x_i^2 < c_1$$

$$\Leftrightarrow \sum_{i=1}^n x_i^2 > c_2$$

On sait que

$$\sum_{i=1}^n \frac{X_i^2}{\sigma_0^2} \backsim \chi^2_n \qquad \textit{(sous } H_0\textit{)}$$

Donc on rejette H_0 si

$$\boxed{\sum_{i=1}^{n} \frac{x_i^2}{\sigma_0^2} > \chi^2_{n,1-\alpha}}$$

Exemple 15 *Soit $X_1, \ldots, X_n iid \backsim Bernoulli\,(p)$. Posons $S = \sum_{i=1}^{n} X_i \backsim Bin\,(n, p)$ et*

$$H_0 : p = p_0 \text{ contre } H_1 : p = p_1 \ \ (\text{avec } p_1 > p_0)$$

On a

$$\begin{aligned}
\frac{p\,(x\,|p_0\,)}{p\,(x\,|p_1\,)} &= \frac{\binom{n}{x} p_0^x\,(1-p_0)^{n-x}}{\binom{n}{x} p_1^x\,(1-p_1)^{n-x}} < c \\
&\Leftrightarrow \left[\frac{p_0\,(1-p_1)}{p_1\,(1-p_0)}\right]^x < c_1 \\
&\Leftrightarrow x \underbrace{\ln \frac{p_0\,(1-p_1)}{p_1\,(1-p_0)}}_{<0} < c_2 \\
&\Leftrightarrow \boxed{x > c_3}
\end{aligned}$$

Ceci confirme ce que nous avons observé dans le cas $n = 3$.

NB. On a utilisé $\frac{p_0}{p_1} < 1$ et $\frac{1-p_1}{1-p_0} < 1 \Rightarrow \ln \frac{p_0\,(1-p_1)}{p_1\,(1-p_0)} < 0$.

Exemple 16 *Soit $X_1, \ldots, X_n iid \backsim U\,[0; \theta]$. On veut tester*

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1 \ \ (\text{avec } \theta_1 > \theta_0)$$

La fonction vraisemblance est donnée par

$$L\,(\theta\,|x\,) = \frac{1}{\theta^n} \text{ avec } 0 \leq x_{(1)} < x_{(n)} \leq \theta,\ \theta > 0.$$

Le test rejette H_0 si

$$\frac{L\,(\theta_0\,|x\,)}{L\,(\theta_1\,|x\,)} < c$$

c'est-à-dire si $x_{(n)}$ est grand :

$$x_{(n)} > k$$

($X_{(n)}$ est la statistique du test). Trouvons la région critique, une fois α petit choisi. Nous savons que la statistique $T\,(X_1, \ldots, X_n) = X_{(n)}$ a pour fonction de densité

$$f\,(t) = \frac{nt^{n-1}}{\theta^n}, t \in [0; \theta]$$

Alors (sous H_0),

$$\begin{aligned}
\alpha &= P\left(X_{(n)} > k\,|\theta_0\right) = \int_k^{\theta_0} \frac{nt^{n-1}}{\theta_0^n} dt = \frac{\theta_0^n - k^n}{\theta_0^n} \\
&\Rightarrow k = \theta_0\,(1-\alpha)^{1/n}
\end{aligned}$$

On en conclut que la région critique est donnée par

$$\boxed{R = \left\{ x \in \mathbb{R}^n : x_{(n)} > \theta_0 \left(1 - \alpha\right)^{1/n} \right\}}$$

Exemple 17 *Autre test. Soit $X_1, \ldots, X_n$ iid*

$$\begin{aligned} &H_0 : X \backsim U\,]0;1[\\ &H_1 : X \backsim Exp\,(\lambda = 1) \end{aligned}$$

On a

$$\frac{f\left(x\,|H_0\right)}{f\left(x\,|H_1\right)} = \frac{1}{\exp\left(-\sum_{i=1}^{n} x_i\right)} < c$$

$$\Leftrightarrow \sum_{i=1}^{n} x_i < c_1$$

La loi de $T\left(X_1, \ldots, X_n\right) = \sum_{i=1}^{n} X_i$ est très compliquée (sous H_0). On passera alors par le TCL. Rappelons que si $X \backsim U\,]0;1[$, $E\left(X\right) = 1/2$ et $VAR\left(X\right) = 1/12$ et

$$\sqrt{n}\frac{\overline{X} - 1/2}{\sqrt{1/12}} \overset{Loi}{\to} N\left(0, 1\right)$$

Pour α petit donné, on rejette H_0 si

$$\boxed{\sqrt{n}\frac{\overline{x} - 1/2}{\sqrt{1/12}} < -z_\alpha}$$

Exemple 18 *Reprenons l'exemple 7 ci-dessus : $X \backsim Exp\left(\theta\right) : f\left(x\,|\theta\right) = \frac{1}{\theta}e^{-x/\theta}$, $x, \theta > 0$ avec, cette fois-ci, l'hypothèse alternative suivante :*

$$H_0 : \theta = 2 \quad \textit{contre} \quad H_1 : \theta > 2$$

On dispose toujours de $n = 2$ observations et on supposons toujours que

$$R = \left\{ x = \left(x_1, x_2\right) \in \mathbb{R}^2 : 9.5 < x_1 + x_2 < \infty \right\}$$

On a vu que

$$\alpha = \frac{23}{4}e^{-\frac{19}{4}} \approxeq 0.05$$

Maintenant, on va voir que cette région R est la région optimale <u>pour ce choix de α</u>, et ce <u>quelle que soit la valeur de $\theta > 2$</u>. Soit donc $\theta = \theta_1 > 2$. Alors,

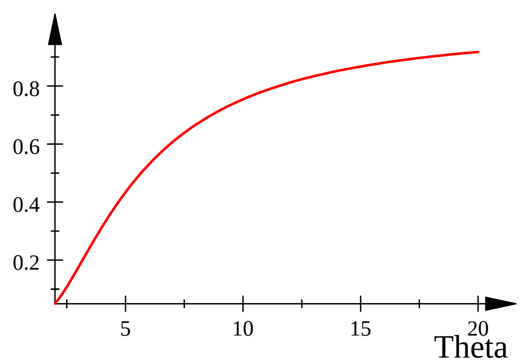
$$\frac{L\left(2\,|x_1, x_2\right)}{L\left(\theta\,|x_1, x_2\right)} = \frac{\left(1/2\right)^2 \exp\left(-\left(x_1 + x_2\right)/2\right)}{\left(1/\theta_1\right)^2 \exp\left(-\left(x_1 + x_2\right)/\theta_1\right)} < c$$

Prenons le log :

$$
\begin{aligned}
&2\ln\frac{1}{2} - 2\ln\theta_1 + (x_1+x_2)\left(-1/2+1/\theta_1\right) < \ln c \\
&\Leftrightarrow \\
&x_1 + x_2 > \frac{\ln c - 2\ln\frac{1}{2} + 2\ln\theta_1}{-1/2+1/\theta_1} = c'
\end{aligned}
$$

Donc on rejette H_0 quand $x_1 + x_2 > c'$ (c' est calculée à partir du niveau de signification α). Pour $\alpha = 0.05$, on a donc $c' = 9.5$ (comme calculé) et donc la région proposée est la meilleure, quel que soit la valeur de $\theta_1 > 2$. La puissance, pour cette région, est alors :

$$
\begin{aligned}
1-\beta(\theta) &= 1 - \int_0^{9.5}\int_0^{9.5-x_2} \frac{1}{\theta^2}\exp\left(-\frac{x_1+x_2}{\theta}\right) dx_1 dx_2 \\
&= \frac{\theta+9.5}{\theta} e^{-9.5/\theta}
\end{aligned}
$$



Puissance du test

Remarque 4 *Vous avez sans doute remarqué que dans chacun des exemples ci-dessus, le test de niveau α le plus puissant dépend d'une statistique exhaustive. Ceci n'est pas une coincidence. En effet, soit $T = T(X_1, \ldots, X_n)$ est une statistique exhaustive pour θ. D'après le théorème de factorisation de Neyman-Pearson, on peut écrire la fonction vraisemblance comme suit :*

$$L(\theta\,|x = (x_1, \ldots, x_n)) = g(T(x)\,|\theta)\,h(x)$$

où $h(x)$ ne dépend pas de θ. Maintenant, le test le plus puissant rejette $H_0 : \theta = \theta_0$ en faveur de $H_1 : \theta = \theta_1$ pour les petites valeurs de

$$\frac{L(\theta_0\,|x)}{L(\theta_1\,|x)} = \frac{g(T(x)\,|\theta_0)}{g(T(x)\,|\theta_1)}\,(< c)$$

ce qui est exactement le sens de cette remarque.

Remarque 5 *Le lemme de NP n'exige pas que les observables soient indépendantes ou identiquement distibuées. Dans la pratique, tout ce dont on a besoin, c'est une forme explicite de la fonction vraisemblance aussi bien sous H_0 que sous H_1. On en donne des exemples.*

Exemple 19 *Soit* $X_1 \sim N\left(\mu, \sigma^2\right)$ *et* $X_2 \sim N\left(\mu, 4\sigma^2\right)$ *indépendantes. On supposera* σ^2 *connu. Nous avons donc une situation de deux variables aléatoires indépendantes mais de lois différentes. On veut trouver le test le plus puissant de niveau* α *pour*

$$H_0 : \mu = \mu_0 \quad \textit{contre} \quad H_1 : \mu = \mu_1 \ (\mu_1 > \mu_0)$$

La fonction vraisemblance (une paire d'observations) est

$$\begin{aligned} L\left(\mu \,| x = (x_1, x_2)\right) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x_1 - \mu}{\sigma}\right)^2\right] \times \frac{1}{\sqrt{2\pi}\,(2\sigma)} \exp\left[-\frac{1}{2}\left(\frac{x_2 - \mu}{2\sigma}\right)^2\right] \\ &= \frac{1}{4\pi\sigma^2} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1 - \mu}{\sigma}\right)^2\right] + \left(\frac{x_2 - \mu}{2\sigma}\right)^2\right\} \end{aligned}$$

On a alors

$$\begin{aligned} \frac{L\left(\mu_0 \,| x\right)}{L\left(\mu_1 \,| x\right)} &= \frac{\exp\left\{-\frac{1}{2}\left[\left(\frac{x_1 - \mu_0}{\sigma}\right)^2 + \left(\frac{x_2 - \mu_0}{2\sigma}\right)^2\right]\right\}}{\exp\left\{-\frac{1}{2}\left[\left(\frac{x_1 - \mu_1}{\sigma}\right)^2 + \left(\frac{x_2 - \mu_1}{2\sigma}\right)^2\right]\right\}} \\ &= \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1 - \mu_0}{\sigma}\right)^2 + \left(\frac{x_2 - \mu_0}{2\sigma}\right)^2 - \left(\frac{x_1 - \mu_1}{\sigma}\right)^2 - \left(\frac{x_2 - \mu_1}{2\sigma}\right)^2\right]\right\} \\ &\overset{calculs}{=} \exp\left[-\frac{1}{4\sigma^2}\left(4x_1 + x_2\right)\left(\mu_1 - \mu_0\right)\right] \exp\left[-\frac{5}{8\sigma^2}\left(\mu_1^2 - \mu_0^2\right)\right] < c \\ &\Leftrightarrow 4x_1 + x_2 > k \end{aligned}$$

Donc la région critique est donnée par

$$R = \left\{(x_1, x_2) \in \mathbb{R}^2 : 4x_1 + x_2 > k\right\}$$

où la constante k *est déterminée par le choix du niveau* α*. Maintenant, la statistique*

$$T\left(X_1, X_2\right) = 4X_1 + X_2 \sim N\left(5\mu_0, 20\sigma^2\right) \ \textit{(sous } H_0\textit{)}$$

Par conséquent, la région critique est

$$R = \left\{(x_1, x_2) \in \mathbb{R}^2 : 4x_1 + x_2 > 5\mu_0 + z_\alpha \sqrt{20}\sigma\right\}$$

Exemple 20 *Soit* $X = (X_1, X_2)$ *suivant une loi binormale* $N\left(\mu_1 = \mu, \mu_2 = \mu, \sigma_1^2 = 1, \sigma_2^2 = 1, \rho = 1/\sqrt{2}\right)$*. On veut trouver le test le plus puissant de niveau* α *pour*

$$H_0 : \mu = \mu_0 \quad \textit{contre} \quad H_1 : \mu = \mu_1 \ (\mu_1 > \mu_0)$$

La fonction vraisemblance est

$$\begin{aligned} &L\left(\mu \,| x = (x_1, x_2)\right) \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2\left(1 - \rho^2\right)}\left[\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 - \frac{2\rho}{\sigma_1\sigma_2}\left(x_1 - \mu_1\right)\left(x_2 - \mu_2\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2\right]\right\} \\ &\overset{calculs}{=} \frac{1}{\pi\sqrt{2}} \exp\left\{-\frac{1}{2}\left[\left(x_1 - \mu_1\right)^2\right] - 2\left(x_1 - \mu_1\right)\left(x_2 - \mu_2\right) + \left(x_2 - \mu_2\right)^2\right\} \end{aligned}$$

On a alors

$$\frac{L(\mu_0 | x)}{L(\mu_1 | x)} \overset{calculs}{=} \exp\left[-(x_1+x_2)(\mu_1-\mu_0)\right] \exp\left[-\left(\mu_1^2-\mu_0^2\right)\right] < c$$
$$\Leftrightarrow x_1 + x_2 > k$$

Maintenant, la statistique

$$T(X_1, X_2) = X_1 + X_2 \sim N\left(2\mu_0, 2+\sqrt{2}\right) \ \textit{(sous } H_0\textit{)}$$

Par conséquent, la région critique est

$$R = \left\{(x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 > 2\mu_0 + z_\alpha \sqrt{2+\sqrt{2}}\right\}$$

2.3.2 Exercices

Exercice 1 *Soit $X_1, \ldots, X_n iid \sim Poi(\lambda)$. Trouver le test le plus puissant de niveau $\alpha \in]0;1[$ pour confronter*

$$H_0 : \lambda = \lambda_0 \ \textit{contre } H_1 : \lambda = \lambda_1 \ \textit{(avec } \lambda_1 > \lambda_0\textit{)}$$

Exercice 2 *Soit X une variable aléatoire observable de densité $f(x)$. On hésite entre deux modèles :*

$$f_0(x) = \begin{cases} \frac{3}{64}x^2 & si \quad x \in [0;4] \\ 0 & ailleurs \end{cases} \quad et \quad f_1(x) = \begin{cases} \frac{3}{16}\sqrt{x} & si \quad x \in [0;4] \\ 0 & ailleurs \end{cases}$$

Trouver le test le plus puissant de niveau $\alpha \in]0;1[$ pour confronter

$$H_0 : f(x) = f_0(x) \ \textit{ et } H_1 : f(x) = f_1(x)$$

Calculer la puissance du test pour les valeurs suivantes de α : $0.05; 0.01; 0.001$.

Exercice 3 *Quelle est la puissance de chacun des tests des exemples 19 et 20 ?*

2.4 Tests uniformément les plus puissants (UPP)

Nous allons voir que le théorème de Neyman-Pearson s'applique aux **tests composés** unilatéraux de la forme

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

ou $H_1 : \theta < \theta_0$.

Soit $X_1, \ldots, X_n iid \backsim N\left(\mu, \sigma_0^2\right)$, σ_0^2 connue.

$$H_0 : \mu = \mu_0 \text{ contre } H_1 : \mu > \mu_0$$

Ce test équivaut à

$$H_0 : \mu = \mu_0 \text{ contre } H_1 : \mu = \mu_1 \ \text{ (avec } \mu_1 > \mu_0\text{)}$$

Le calcul fait dans un exemple précédent est donc valide et on a obtenu le test uniformément le plus puissant (i.e. indépendant de μ, tant que $\mu > \mu_0$) :

$$R = \left\{x \in \mathbb{R}^n : \overline{x} > \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_\alpha\right\}$$

On a la défintion suivante.

Définition 1 *Dans le cas ou H_0 est simple et H_1 est composée, si, pour chacun des choix d'une hypothèse simple sous H_1, le test de N-P confrontant H_0 à cette hypothèse simple reste le même, alors ce test N-P est dit* ***uniformément le plus puissant*** *(UPP).*

Remarque 6 *Il n'existe pas de test UPP pour confronter*

$$H_0 : \mu = \mu_0 \text{ et } H_1 : \mu \neq \mu_0$$

2.4.1 Technique pour décider si un test est UPP

1. Choisir le niveau α du test.
2. Trouver le test N-P pour une valeur particulière de l'alternative composée (unilatérale).
3. Montrer que le test ne dépend pas du choix de cette valeur particulière.

Exemple 21 *Soit $X_1, \ldots, X_n iid \backsim Exp(\theta)$. Nous avons vu que le test N-P (pour α donné) pour*

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1 \ \ (\text{avec } \theta_1 > \theta_0)$$

rejette H_0 si

$$\frac{2n}{\theta_0}\overline{x} = 2\sum_{i=1}^{n}\frac{x_i}{\theta_0} > \chi^2_{2n,\alpha}$$

Ce résultat ne dépend pas de θ_1 (mais du fait que $\theta_1 > \theta_0$). On en conclut que ce test est UPP.

Exemple 22 *Soit $X_1, \ldots, X_n iid \backsim N\left(0, \theta = \sigma^2\right)$. Cherchons le test UPP pour confronter*

$$H_0 : \sigma = \sigma_0 \text{ et } H_1 : \sigma < \sigma_0$$

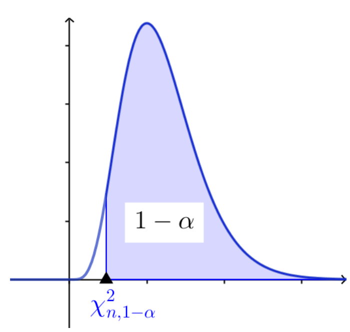
Soit $\sigma_1 < \sigma_0$ fixé et vérifions que le test

$$H_0 : \sigma = \sigma_0 \text{ contre } H_1 : \sigma = \sigma_1$$

est le plus puissant. Un petit calcul (exercice) montre que le test rejette H_0 si

$$\boxed{\sum_{i=1}^{n} X_i^2 < \sigma_0^2\chi^2_{n,1-\alpha}}$$

Le test ne dépend pas du choix particulier de $\sigma_1 (< \sigma_0)$. Il est donc UPP.



2.4.2 Propriété du Rapport de Vraisemblance Monotone (RVM)

Dans les exemples précédents, on a vu des tests UPP. On peut trouver des résultats généraux si les fonctions de densité possèdent la propriété du **R**apport de **V**raisemblance **M**onotone. Cette propriété nous servira pour construire dans certaines situations le test unilatéral UPP de niveau α.

Définition 2 *On dit que la famille $\{f(x_1,\ldots,x_n\,|\theta\,;\theta\in\Theta)\}$ a la propriété du RVM dans la statistique $T=T(X_1,\ldots,X_n)$ si $\forall\theta_0<\theta_1$ (deux valeurs fixes de θ), alors*

$$1.\quad \frac{L(\theta_0\,|x_1,\ldots,x_n)}{L(\theta_1\,|x_1,\ldots,x_n)}=\frac{f(x_1,\ldots,x_n\,|\theta_0)}{f(x_1,\ldots,x_n\,|\theta_1)}=g(T(x_1,\ldots,x_n))$$

(fonction de $T=t$) et

$$2.\quad g(t)\ \text{décroissante au sens large}$$

Noter que dans ce cas, la statistique T est exhaustive pour θ.

Exemple 23 *Soit $X_1,\ldots,X_n iid \backsim Exp(\theta)$. On a vu plus haut que, pour $T(X_1,\ldots,X_n)=\sum_{i=1}^n X_i$, on a*

$$\begin{aligned}\frac{f(x\,|\theta_0)}{f(x\,|\theta_1)}&=\left(\frac{\theta_1}{\theta_0}\right)^n\exp\left[-\left(\frac{1}{\theta_0}-\frac{1}{\theta_1}\right)t\right]\\&=g(t)\searrow\qquad\text{(décroissante)}\end{aligned}$$

Exemple 24 *Soit $X_1,\ldots,X_n iid \backsim N(\theta=\mu,\sigma^2)$, σ^2 connu. Soit $\mu_1>\mu$ donné (et arbitraire) et $T(X_1,\ldots,X_n)=\sum_{i=1}^n X_i$. On a*

$$\frac{L(\mu_0\,|x_1,\ldots,x_n)}{L(\mu_1\,|x_1,\ldots,x_n)}=\exp\left\{\left[-\frac{(\mu_1-\mu)}{\sigma^2}t\right]+\left[n\frac{(\mu_1^2-\mu^2)}{2\sigma^2}\right]\right\}$$

qui est une fonction décroissante de t.

Remarque 7 *Test bilatéral. Il n'existe en général pas de test UPP pour $H_0:\theta=\theta_0$ contre $H_1:\theta\neq\theta_1$. Pour fixer les idées mettons nous dans le cas continu (par exemple $N(\theta,1)$). Alors, en effet, nous avons un test UPP pour $H_0:\theta=\theta_0$ contre $H_1:\theta<\theta_1$ et pour $H_0:\theta=\theta_0$ contre $H_1:\theta>\theta_1$. Ces deux tests sont évidemment différents. S'il y avait un test UPP pour $H_0:\theta=\theta_0$ contre $H_1:\theta\neq\theta_1$, cela contredirait l'existence de ces deux derniers (pourquoi ?).*

2.4.3 Exercice

Exercice 4 *Considérons un test pour une loi exponentielle à deux paramètres :*

$$f(x\,|\lambda,\theta)=\begin{cases}\lambda\exp(-\lambda(x-\theta)) & si\quad x\geq\theta\\ 0 & sinon\end{cases}$$

et considérons un échantillon $X_1, \ldots, X_n iid \backsim f(x|\lambda = 1, \theta)$. La densité conjointe est donnée par

$$f(x_1, \cdots, x_n |\lambda = 1, \theta) = \begin{cases} \exp\left(-\sum_{i=1}^n (x_i - \theta)\right) & pour \quad x_{(1)} \geq \theta \\ \\ 0 & sinon \end{cases}$$

Soit

$$\theta_0 < \theta_1$$

Vérifier que $f(x|\lambda = 1, \theta)$ est RVM.

3 Test du rapport de vraisemblance (généralisé)

On a vu qu'obtenir des tests optimums n'était possible que dans certaines situations simples. Nous allons utiliser une approche (due à Neyman et Pearson) plus flexible, basée sur le maximum de vraisemblance. Pourquoi ? On sait (chapitre 8) que les EMV ont de bonnes propriétés asymptotiques et s'appliquent à un grand nombre de situations.

Contexte général : Soit donné $X_1, \cdots, X_n \sim f(x|\theta)$ où $\theta \in \Theta$ avec

$$\Theta = \Theta_0 \cup \Theta_1, \ \Theta_0 \cap \Theta_1 = \emptyset$$

On veut tester au niveau $\alpha \in]0; 1[$

$$H_0 : \theta \in \Theta_0 \text{ contre } H_1 : \theta \in \Theta_1$$

La fonction vraisemblance est

$$L(\theta) = L(\theta |x_1, \cdots, x_n)$$

On compare ensuite $\sup_{\theta\in\Theta_0} L(\theta)$, qui signifie qu'on cherche la valeur de θ la plus favorable à l'hypothèse nulle, à $\sup_{\theta\in\Theta} L(\theta)$ qui signifie qu'on cherche la valeur la plus favorable de θ indépendamment de toute restriction. Pour ce faire, la **statistique du test** est définie par

$$\Lambda = \frac{\sup_{\theta\in\Theta_0} L(\theta)}{\sup_{\theta\in\Theta} L(\theta)}$$

Noter qu'on a les inégalités évidentes $0 < \Lambda < 1$. Le **test rejettera H_0 si Λ est petit**. En d'autres termes, a région critique d'un test de rapport de vraisemblance est de la forme

$$R = \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : \Lambda \leq c\}$$

avec $c > 0$, où c (petit) est calculé selon le niveau α choisi. On a

$$\Lambda = \frac{f\left(x \left|\widehat{\theta}_0\right.\right)}{f\left(x \left|\widehat{\theta}\right.\right)}$$

où $\widehat{\theta}_0 =$ estimateur EMV sous H_0 (i.e. dans Θ_0) et $\widehat{\theta} =$ estimateur EMV usuel (i.e. dans Θ).

Remarque 8 *La condition que les X_i soient indépendantes n'est pas nécessaire.*

3.1 Quelques exemples

3.1.1 Test sur une moyenne (variance connue, population normale)

Soit $X \backsim N\left(\mu, \sigma^2\right)$, $\underline{\sigma^2 \text{ connu}}$. Test :

$$H_0 : \mu = \mu_0 \text{ contre } H_0 : \mu \neq \mu_0$$

On a

$$L\left(\mu \,|x\right) = (2\pi)^{-n/2}\left(\sigma^2\right)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

Puisque $\widehat{\mu}_0 = \mu_0$ et $\widehat{\mu} = \overline{X}$, alors

$$\begin{aligned}\Lambda = \frac{L\left(\mu_0\right)}{L\left(\overline{x}\right)} &= \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\overline{x})^2\right)\right] \\ &\overset{calculs}{=} \exp\left[-\frac{1}{2}\left(n\left(\frac{\overline{x}-\mu_0}{\sigma}\right)^2\right)\right] \leq c\end{aligned}$$

Prenons le log :

$$\begin{aligned}\ln\Lambda = -\frac{1}{2}\left(n\left(\frac{\overline{x}-\mu_0}{\sigma}\right)^2\right) \leq c_1 &\Leftrightarrow n\left(\frac{\overline{x}-\mu_0}{\sigma}\right)^2 \geq c_2 \\ &\Leftrightarrow \sqrt{n}\frac{\left|\overline{x}-\mu_0\right|}{\sigma} \geq c_3\end{aligned}$$

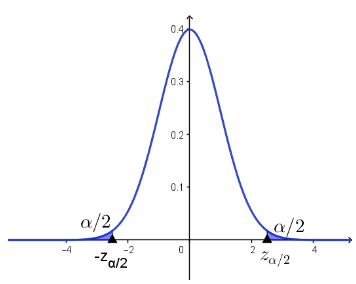
Finalement,

$$R = \left\{x \in \mathbb{R}^n : \sqrt{n}\frac{\left|\overline{x}-\mu_0\right|}{\sigma} \geq c\right\}$$

Distribution nulle (sous H_0) :

$$\sqrt{n}\frac{\overline{X}-\mu_0}{\sigma} \backsim N\left(0,1\right)$$

Donc, pour un niveau α donné,



$$\boxed{R = \left\{x \in \mathbb{R}^n : \sqrt{n}\frac{\left|\overline{x}-\mu_0\right|}{\sigma} > z_{\alpha/2}\right\}}$$

$$R^c = \left\{x \in \mathbb{R}^n : \sqrt{n}\frac{\left|\overline{x}-\mu_0\right|}{\sigma} \leq z_{\alpha/2}\right\} = \boxed{\left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} ; \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = R^c} = IC_{1-\alpha}\left(\mu\right)$$

Lorsque $x \in R$, on rejette H_0 (les données observées ne favorisent pas H_0) et lorsque $x \notin R$, il n'y a pas assez d'éléments pour rejeter H_0.

Remarque 9 *Constatons la dualité entre test et IC.*

Remarque 10 *Nous avons vu qu'il n'y a pas de test UPP dans ce cas.*

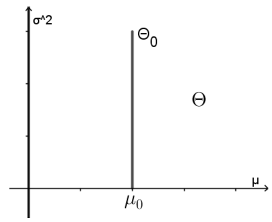
Remarque 11 *On aurrait pu utiliser* $Z^2 = \left(\frac{\overline{X}-\mu_0}{\sigma_0}\right)^2 \backsim \chi_1^2$.

Remarque 12 *L'approximation asymptotique* $-2\ln\Lambda \backsim \chi_1^2$ *est exacte ici. Nous y reviendrons plus loin dans ce chapitre.*

3.1.2 Test sur une moyenne (variance inconnue, population normale)

Soit $X_1, \cdots, X_n$ iid $\backsim N\left(\mu, \sigma^2\right)$, $\underline{\sigma^2 \text{ inconnu}}$. Test :

$$H_0 : \mu = \mu_0 \text{ contre } H_1 : \mu \neq \mu_0$$



Noter que H_0 n'est pas une hypothèse simple puisqu'elle dépend de σ^2 inconnu, i.e. $\theta_0 = \left(\mu_0, \sigma^2\right)$. On a

$$L\left(\mu, \sigma^2\right) = (2\pi)^{-n/2}\left(\sigma^2\right)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

Dénominateur : $L\left(\widehat{\mu}, \widehat{\sigma}^2\right)$ usuel :

$$\widehat{\mu} = \overline{X}$$
$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Numérateur :

$$\sup_{\sigma^2}\left((2\pi)^{-n/2}\left(\sigma^2\right)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\right]\right)$$

Log :

$$c - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2$$

$$\frac{\partial}{\partial\sigma^2} : -\frac{n}{2\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu_0)^2 = 0$$
$$\Rightarrow \begin{cases} \widehat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i-\mu_0\right)^2 \\ \widehat{\mu}_0 = \mu_0 \end{cases}$$

et alors

$$\Lambda = \frac{L\left(\widehat{\mu}_0, \widehat{\sigma}_0^2\right)}{L\left(\widehat{\mu}, \widehat{\sigma}^2\right)} \overset{calculs}{=} \left(1 + \frac{1}{n-1} T^2\right)^{-n/2}$$

où

$$T = \sqrt{n} \frac{\overline{X} - \mu_0}{S} \backsim t_{n-1}$$

Ainsi,

$$\Lambda = \left(1 + \frac{1}{n-1} T^2\right)^{-n/2} < c \Leftrightarrow |T| > c'$$

et

$$\boxed{R = \left\{ x \in \mathbb{R}^n : |T| = \sqrt{n} \frac{\left|\overline{X} - \mu_0\right|}{S} > t_{n-1, \alpha/2} \right\}}$$

C'est le test de Student.

3.1.3 Autre exemple

Soit $X_1, \cdots, X_n$ iid $\sim Exp\left(\theta = \lambda\right)$ et considérons

$$H_0 : \lambda \leq \lambda_0 \text{ contre } H_1 : \lambda > \lambda_0$$

La fonction de vraisemblance est

$$L\left(\lambda\right) = \lambda^n e^{-n\lambda\overline{x}} = \left(\lambda e^{-\lambda\overline{x}}\right)^n$$

La fonction

$$g\left(\lambda\right) = \lambda e^{-\lambda\overline{x}}, \quad \lambda > 0$$

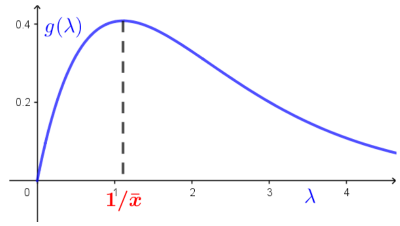
atteint son maximum en $\lambda = 1/\overline{x}$ car

$$\frac{d}{d\lambda}\left(\lambda e^{-\lambda\overline{x}}\right) = \left(1 - \lambda\overline{x}\right) e^{-\lambda\overline{x}} = 0 \Rightarrow \lambda = \frac{1}{\overline{x}}$$

$$\frac{d^2}{d\lambda^2}\left(\lambda e^{-\lambda\overline{x}}\right) = \overline{x}\left(-2 + \lambda\overline{x}\right) e^{-\lambda\overline{x}} \Rightarrow \frac{d^2}{d\lambda^2}\left(\lambda e^{-\lambda\overline{x}}\right)_{|\lambda = \frac{1}{\overline{x}}} = -e^{-1} < 0$$

On a le tableau de variation suivant

λ	0		$1/\overline{x}$		$2/\overline{x}$		$+\infty$
$g'\left(\lambda\right)$		$+$	0	$-$	$-$	$-$	
$g''\left(\lambda\right)$		$-$	$-$	$-$	0	$+$	
			$\left(1/\overline{x}\right) e^{-1}$				
$g\left(\lambda\right)$		$\nearrow$		$\searrow$	PI		
					$\left(2/\overline{x}\right) e^{-2}$		
						$\searrow$	
							0



$$g(\lambda) = \lambda e^{-\lambda \overline{x}}$$

Ainsi,

$$\widehat{\lambda}_{EVM} = \frac{1}{\overline{x}}$$

et

$$\sup_{\lambda \in \Theta} L(\lambda) = L\left(\widehat{\lambda}_{EVM}\right) = L\left(\frac{1}{\overline{x}}\right) = g\left(\frac{1}{\overline{x}}\right)^n = \left(\frac{1}{\overline{x}} e^{-1}\right)^n, \quad \Theta =]0; +\infty[$$

Puisque

$$\Theta_0 =]0; \lambda_0]$$

on a deux cas à considérer pour trouver $\sup_{\lambda \in \Theta_0} L(\lambda)$ (et par suite Λ).

Cas 1. $\lambda_0 \leq 1/\overline{x}$ (λ_0 à gauche du maximum de g, voir graphe). Alors

$$\sup_{\lambda \in \Theta_0} L(\lambda) = L(\lambda_0) = g(\lambda_0)^n = \left(\lambda_0 e^{-\lambda_0 \overline{x}}\right)^n$$

et on a

$$\Lambda = \frac{\sup_{\lambda \in \Theta_0} L(\lambda)}{\sup_{\lambda \in \Theta} L(\lambda)} = \frac{\left(\lambda_0 e^{-\lambda_0 \overline{x}}\right)^n}{\left(\frac{1}{\overline{x}} e^{-1}\right)^n} = \left(\lambda_0 \overline{x} e^{1-\lambda_0 \overline{x}}\right)^n$$

Cas 2. $\lambda_0 > 1/\overline{x}$ (λ_0 à droite du maximum de g, voir graphe). Alors

$$\sup_{\lambda \in \Theta_0} L(\lambda) = L\left(\frac{1}{\overline{x}}\right) = \sup_{\lambda \in \Theta} L(\lambda)$$

et

$$\Lambda = \frac{\sup_{\lambda \in \Theta_0} L(\lambda)}{\sup_{\lambda \in \Theta} L(\lambda)} = \frac{L(1/\overline{x})}{L(1/\overline{x})} = 1$$

Finalement,

$$\Lambda = \begin{cases} h(\overline{x})^n & si \quad \overline{x} \leq 1/\lambda_0 \\ 1 & si \quad \overline{x} > 1/\lambda_0 \end{cases}$$

où

$$h(t) = e\lambda_0 t e^{-\lambda_0 t}, \quad t > 0$$

Une petite comparaison avec la fonction $g(t) = te^{-t\overline{x}}$ montre que $h(t)$ est croissante sur $]0; \lambda_0]$ puis décroissante.

Maintenant, pour $c < 1$,

$$\Lambda \leq c \Leftrightarrow h(\overline{x}) \leq \sqrt[n]{c} \Leftrightarrow \overline{x} \leq c'$$

puisque la fonction h est croissante sur $]0; \lambda_0]$.

3.1.4 Exercices

Exercice 5 *Soit* $X_1, \cdots, X_n$ *iid* $\backsim N(\mu, \sigma^2)$, μ *et* σ^2 *inconnus. Tester au niveau* $\alpha \in]0;1[$

$$H_0 : \sigma = \sigma_0 \text{ contre } H_1 : \sigma \neq \sigma_0$$

Réponse. La région critique du test est donnée par

$$\left\{x \in \mathbb{R}^n : (n-1) S^2/\sigma_0^2 > \chi^2_{n-1,\alpha/2} \text{ ou } (n-1) S^2/\sigma_0^2 < \chi^2_{n-1,1-\alpha/2}\right\}$$

Noter qu'il n'existe pas de test UPP dans ce cas, même si μ *est connu.*

Exercice 6 *Le test de l'exemple 3.1.3 aurait-il été différent si on avait*

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

Ce test est-il UPP ?

Refaire le test du RVG, mais cette fois-ci pour une alternative bilatérale

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

Que constatez-vous ?

3.2 Théorème de Wilks

Dans les exemples précédents, on a pu obtenir la distribution nulle (sous H_0) exacte de la statistique du test. Quand ceci n'est pas possible (très souvent), on utilise le résultat suivant.

Théorème 2 *Sous certaines conditions de régularité, quand* $n \to \infty$, *la distribution nulle*

$$-2 \ln \Lambda \overset{Loi}{\to} \chi^2_r$$

où $r = \dim \Theta - \dim \Theta_0$.

Preuve. *Omise. Basée sur un développement limité d'ordre* 2 *de* $l(\theta)$ *autour de* θ_0. ■

Exemple 25 *Soit* $X \backsim N(\mu, \sigma_0^2)$, σ_0^2 *connu. On a vu que*

$$\Lambda = \exp\left[-\frac{1}{2}\left(n\left(\frac{\overline{x}-\mu_0}{\sigma_0}\right)^2\right)\right]$$

Alors,

$$\begin{aligned} \ln \Lambda &= -\frac{1}{2}\left(n\left(\frac{\overline{x}-\mu_0}{\sigma_0}\right)^2\right) \\ \Rightarrow -2\ln\Lambda &= n\left(\frac{\overline{x}-\mu_0}{\sigma_0}\right)^2 \chi^2_1 \qquad \text{(sous } H_0) \end{aligned}$$

Maintenant, $\dim \Theta = 1$ *(droite* μ*) et* $\dim \Theta_0 = 0$ *(point* μ_0*) et donc* $r = 1 - 0 = 1$.

4 Test d'ajustement du χ^2 (K. Pearson, 1900)

Motivation. Soit $X_1, \ldots, X_n$ indépendantes avec $X_i \backsim N\left(\mu_i, \sigma_i^2\right)$. La densité conjointe est ($x = (x_1, \ldots, x_n)$)

$$f(x) = \frac{1}{\sigma_1 \ldots \sigma_n (2\pi)^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right]$$

La variable aléatoire dans l'exposant (forme quadratique)

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \backsim \chi_n^2$$

Il peut être montré que (c'est essentiellement un problème d'algèbre linéaire) si les X_i sont <u>dépendantes</u>, alors

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \backsim \chi_k^2$$

où $k =$ rang de la matrice de variance-covariance. Cette observation est à la base du test du khi-deux.

Passons maintenant à une variable qui est asymptotiquement du khi-deux. Soit donc $X \backsim Bin(n,p)$. On a (TCL)

$$Y = \frac{X - np}{\sqrt{np(1-p)}} \overset{Loi}{\to} N(0,1)$$

Alors on va montrer que (ce qui semble intuitif)

$$Y^2 \overset{Loi}{\to} \chi_1^2$$

Soit $F_n(y)$ la FR de Y. On a donc

$$\lim_{n\to\infty} F_n(y) = \Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-u^2/2}du$$

Posons

$$Z = Y^2$$

et

$$G_n(z) = P(Z \leq z)$$

sa FR. On a (pour $z \geq 0$)

$$\begin{aligned}
G_n(z) &= P(Z \leq z) \\
&= P\left(-\sqrt{z} \leq Y \leq \sqrt{z}\right) \\
&= F_n\left(\sqrt{z}\right) - F_n\left(-\sqrt{z}^{-}\right)
\end{aligned}$$

où

$$F_n\left(-\sqrt{z}^{-}\right) = \lim_{t\to z^-} F_n\left(-\sqrt{t}\right)$$

Puisque $\Phi(y)$ est continue sur $\mathbb{R}$, on a alors

$$\begin{aligned}\lim_{n\to\infty} G_n(z) = \Phi\left(\sqrt{z}\right) - \Phi\left(-\sqrt{z}\right) &= 2\int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \\ &= \int_0^z \frac{1}{\Gamma(1/2)\, 2^{1/2}} v^{1/2-1} e^{-v/2} \qquad (u^2 = v)\end{aligned}$$

Ainsi,

$$Z = Y^2 \overset{Loi}{\longrightarrow} \backsim \chi_1^2$$

Maintenant,

$$\begin{aligned}Y^2 &= \frac{(X-np)^2}{np(1-p)} \\ &= \frac{(X-np)^2}{np} + \frac{(X-np)^2}{n(1-p)} \qquad \text{car } \frac{1}{p(1-p)} = \frac{1}{p} + \frac{1}{1-p} \\ &= \frac{(X-np)^2}{np} + \frac{((n-X)-n(1-p))^2}{n(1-p)} \\ &= \frac{(X_1-np_1)^2}{np_1} + \frac{(X_2-np_2)^2}{np_2} \quad \text{(forme quadratique)}\end{aligned}$$

où $X_1 = X$, $X_2 = n - X$, $p_1 = p$, $p_2 = 1-p$.

Généralisons à une multinomiale et faisons le lien avec les tests.

Soit $(X_1, \ldots, X_m) \backsim mult\,(n; p_1, \ldots, p_m)$. L'EVM de $p = (p_1, \ldots, p_m)$ est

$$\widehat{p} = (\widehat{p}_1, \ldots, \widehat{p}_m) = (X_1/n, \ldots, X_m/n)$$

$$H_0 : p = p(\theta) \text{ contre } H_1 : p \neq p(\theta)$$

$$\Theta = \left\{ p = (p_1, \ldots, p_m) : 0 < p_i < 1, \sum_{i=1}^m p_i = 1 \right\}$$

On a

$$P(X_1 = x_1, \ldots, X_m = x_m) = \frac{n!}{\prod_{i=1}^m x_i!} p_1^{x_1} \times \ldots \times p_m^{x_m} = L$$

L'EVM de θ est $\widehat{\theta}$ dans le modèle $p = p(\theta)$ (sous H_0) :

$$\Lambda = \frac{\prod_{i=1}^m p_i\left(\widehat{\theta}\right)^{x_i}}{\prod_{i=1}^m \widehat{p}_i^{x_i}} = \prod_{i=1}^m \left(\frac{p_i\left(\widehat{\theta}\right)}{\widehat{p}_i} \right)^{x_i}$$

$$\boxed{-2\ln\Lambda =} - 2\sum_{i=1}^{m} x_i \ln \frac{p_i\left(\widehat{\theta}\right)}{\widehat{p}_i} = -2\sum_{i=1}^{m} n\widehat{p}_i \ln \frac{np_i\left(\widehat{\theta}\right)}{n\widehat{p}_i}$$
$$= 2\sum_{i=1}^{m} n\widehat{p}_i \ln \frac{n\widehat{p}_i}{np_i\left(\widehat{\theta}\right)}$$
$$= \boxed{2\sum_{i=1}^{m} O_i \ln \frac{O_i}{E_i} \overset{Loi}{\to} \chi_r^2} \qquad \text{sous } H_0,\ r = m-1-k$$

Ici, $O_i = n\widehat{p}_i$ = fréquence observée et $E_i = np_i\left(\widehat{\theta}\right)$ = fréquence espérée sous H_0 et $k = \dim\theta$. Finalement,

$$\boxed{R = \left\{x \in \mathbb{R}^n : -2\ln\Lambda > \chi^2_{m-1-k,\alpha}\right\}}$$

Simplification heuristique.
Soit $g(x) = x\ln(x/x_0)$. On a

$$g'(x) = x\frac{x_0}{x}\frac{1}{x_0} + \ln\frac{x}{x_0} = 1 + \ln\frac{x}{x_0}$$
$$g'(x_0) = 1$$
$$g''(x) = \frac{x_0}{x}\frac{1}{x_0} = \frac{1}{x}$$
$$g''(x_0) = \frac{1}{x_0}$$

Ainsi,

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + g''(x_0)\frac{(x-x_0)^2}{2!} + \dots$$
$$= (x - x_0) + \frac{1}{2x_0}(x-x_0)^2 + \dots$$

Si H_0 est vraie et n est grand, $O_i \approxeq E_i$ et

$$-2\ln\Lambda = 2\sum_{i=1}^{m} O_i \ln\frac{O_i}{E_i} \approxeq 2\sum_{i=1}^{m}\left[(O_i - E_i) + \frac{1}{2}\frac{(O_i-E_i)^2}{E_i}\right]$$
$$\approxeq \sum_{i=1}^{m}\frac{(O_i-E_i)^2}{E_i} = \chi^2_{calculé} \overset{L}{\to} \chi^2_{m-1-k} \text{ (Pearson)}$$

Exemple 26 *Génotypes AA, Aa et aa avec probabilités associées* $(1-\theta)^2$, $2\theta(1-\theta)$, θ^2 *(*$0 < \theta < 1$*),* $n = 1029$

Type	AA	Aa	aa	*Total*
O_i	342	500	187	1029
E_i	340.6	502.8	185.6	1029.0

$$\widehat{\theta} = \frac{x_2 + 2x_3}{2n} \approxeq 0.4247$$

et $E_i = np_i\left(\widehat{\theta}\right)$:

$$
\begin{aligned}
E_1 &= 1029\left(1-0.4247\right)^2 = 340.6\\
E_2 &= 1029 \times 2 \times 0.4247 \times (1-0.4247) = 502.8\\
E_3 &= 1029 \times 0.4247^2 = 185.6
\end{aligned}
$$

Donc

$$\chi^2_{calculé} = \sum_{i=1}^{3} \frac{\left(O_i - E_i\right)^2}{E_i} = 0.0319$$

Vérifions (valeur exacte) : $-2\ln\Lambda = 2\sum_{i=1}^{3} O_i \ln\frac{O_i}{E_i} = 0.0319$. *On a*

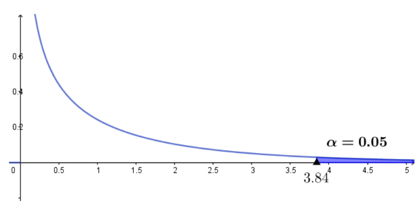
$$
\begin{aligned}
&H_0 : p_1 = (1-\theta)^2, p_2 = 2\theta(1-\theta), p_3 = \theta^2\\
&H_1 : \textit{Toutes les alternatives}
\end{aligned}
$$

Prenons $\alpha = 0.05$. *Alors*

$$R = \left\{x : \chi^2_{calculé} > \chi^2_{1,0.05} = 3.84\right\}$$

Conclusion : il n'y a aucune raison de rejeter le modèle.

Cherchons la valeur-p $= P\left(\chi^2_1 > 0.0319\right) = 0.86 >> 0.05 = \alpha$. *Même conclusion.*



Remarque sur le test d'ajustement du χ^2. Le cas précédent concernait l'ajustement à une multinomiale. On peut l'adapter pour tester l'ajustement d'un échantillon provenant d'une distribution quelconque :

$$H_0 : X \backsim F(x)$$

On partionne l'ensemble fondamental Ω en m cellules $C_1, \ldots, C_m$ et on pose

$$p_{j0} = P\left(X \in C_j\right) \qquad \text{(sous } H_0\text{)}$$

Si $X_1, \ldots, X_n iid \backsim F(x)$, on a $E_j = np_{j0}$, ce qui nous ramène au cas multinomial.

Deux questions se posent.

1. Comment choisir les C_j ? Souvent, il y a un ordre naturel : le plus petit possible pour augmenter les dl.
2. Principe : chaque cellule doit contenir $E_j \geq 5$ pour que l'approximation du χ^2 soit correcte.