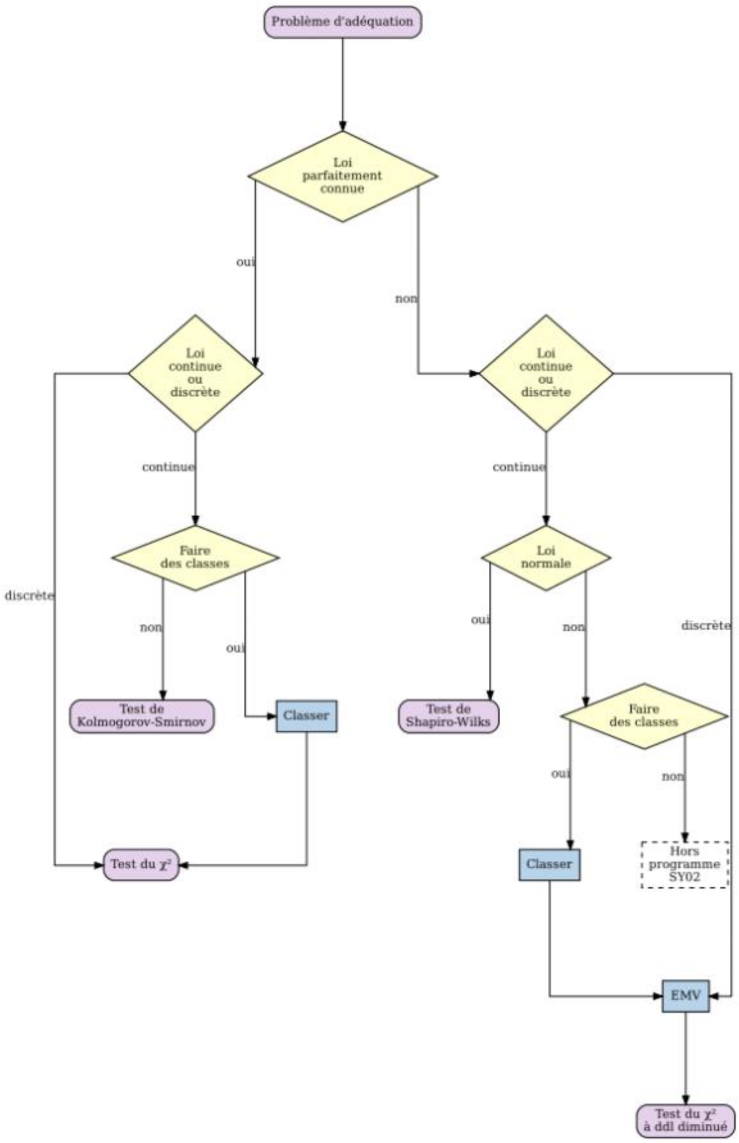
12&13 Tests d'adéquation et Tests d'Indépendance

mercredi 28 décembre 2022

1. Test d'adéquation 一致性 (**entre 1 modèle**)

-1- Introduction

1-



-2- Test du χ^2 d'adéquation

1- vector aléatoire de numbre : $N = (N_1, \dots, N_K)$

2- $H_0 : p_k = p_{k_0} \forall k$

3- Écart entre l'échantillon et l'hypothèse H_0

1> $D^2 = (\chi^2) = \sum_{i=1}^{K} \frac{\left(N_i - np_{i_0}\right)^2}{np_{i_0}} \sim \chi^2_{K-1}, - -n \; est \; numbre \; total$

a) $\Rightarrow D^2 = \sum^{K} \frac{N_i^2}{np_{i_0}} - n$

2> Remplacer θ par son EMV $\hat{\theta}$, si $H_0 = (F_\theta)_{\theta\in\Theta\subset\mathbb{R}^p}$

a) $D^2 = \sum^{K} \frac{\left(N_i - n\hat{p}_{i_0}\right)^2}{n\hat{p}_{i_0}} \sim \chi^2_{K-p-1}$

4- Région critique

1> $W = \left\{D^2 > \chi^2_{K-1;1-\alpha^*}\right\}$

-3- Test de Kolmogorov-Smirnov (K-S) (loi continue)

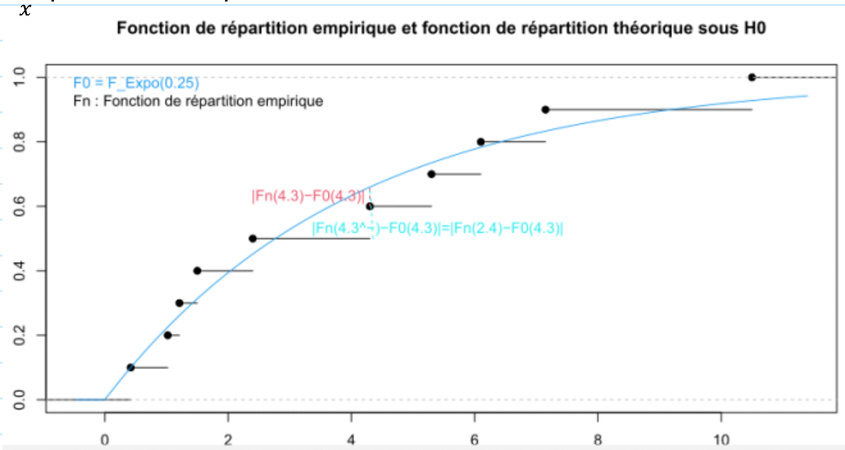
1- $H_0 : F_X = F_0$

2- Fonction de répartition empirique correspondante

1> $F_X(x) \approx^{n\to+\infty} \hat{F}(x) = \frac{1}{n} card\{i \in (1, \dots, n) | x_i \le x\}$

3- Écart entre l'échantillon et l'hypothèse H_0

1> $D_n = \sup_x \left|\hat{F}(x) - F_0(x)\right|$

a) 

2> Car il est discontinue

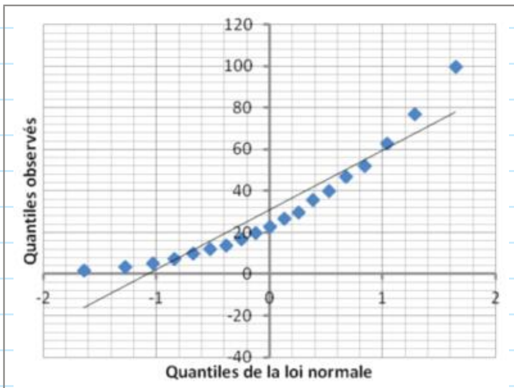
a) $\Rightarrow D_n = \max_{1\le i\le n} max\left(\left|\hat{F}(x_i) - F_0(x_i)\right|, \left|\hat{F}(x_i^-) - F_0(x_i)\right|\right), --\hat{F}(x_i^-) = \lim_{x\to x_i} \hat{F}(x)$

4- Région critique

1> $W = \left\{D_n > d_{n;1-\alpha^*}\right\} --distribution\ de\ K-S$

-4- Test de Shapiro-Wilk/normalité (**loi inconnue**)

1- Diagramme quantile–quantile

1> 

1. X : Quantile de la loi normale
2. Y : Quantile observée

2- Statistique de test de Shapiro-Wilk

1> $w(X_1, \dots, X_n) = \dfrac{\left(\sum a_i X_{(i)}\right)^2}{\sum (X_i - \bar{X})^2}$

a) a_i sont des valeurs obtenues à partir des moments théorique des statistiques d'ordre d'une loi $\mathcal{N}(0,1)$, ils sont antisymétrique反对称关系 $(\forall k, a_{1+k} = -a_{n-k})$

b) $W(X_1, \dots, X_n) = W(aX_1 + b, \dots, aX_n + b)$

a) Ce test étant invariante aux changements de paramètres μ et σ^2.

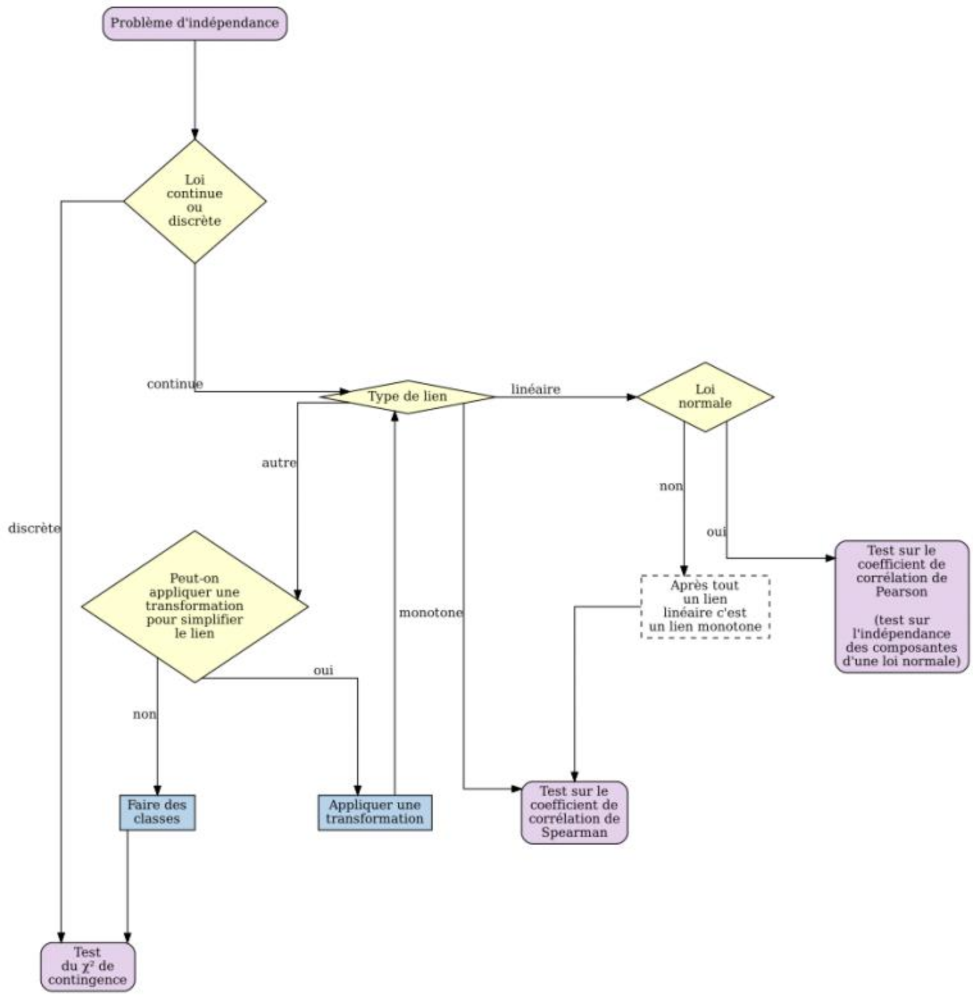
c) Région critique

a) $\{W(X_1, \dots, X_n) < w_{n,\alpha^*}\}$

2. Test d'indépendance (**entre 2 échantillon**)

-1- Introduction

1-



-2- Test du χ^2 d'indépendance

1- χ^2 de contingences

1>

X Y	1 ... j ... J	
1	⋮	
⋮	⋮	
i	... ... N_{ij}	$N_{i.}$
⋮		
I		
	$N_{.j}$	

2- $D^2 = \sum\sum \frac{\left(N_{ij} - np_i p_j\right)^2}{np_i p_j}$

3- $H_0: D^2 \sum\sum \frac{\left(N_{ij} - \frac{N_i N_j}{n}\right)^2}{\frac{N_i N_j}{n}} = \sum\sum \frac{N_{ij}^2}{\left(\frac{N_i N_j}{n}\right)} - n \sim \chi^2_{(I-1)(J-1)}$

4- RC

1> $W = \left\{D^2 > \chi^2_{(I-1)(J-1);1-\alpha^*}\right\}$

-3- Coefficient de correlation de Pearson (**Gaussien**)

1- $R = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}}$

2- $H_0: R\sqrt{\frac{n-2}{1-R^2}} \sim T_{n-2}$

3- RC

1> $W = \left\{\left|R\sqrt{\frac{n-2}{1-R^2}}\right| > t_{n-2,1-\frac{\alpha^*}{2}}\right\}$

4- Remarque

1> Le coefficient de corrélation de Pearson ne doit être utilisé que quand le vecteur aléatoire (X, Y) suit au moins approximativement **une loi normale bidimensionnelle**.

2> Le coefficient de corrélation de Pearson n'indique qu'une dépendance linéaire entre deux variables.

-4- Coefficient de corrélation de Spearman : X_i *et* Y_j suivent des lois continues quelconques (**non Gaussien et continu**)

1- Notion

1> Noter $(R_1, \dots, R_n)$ les rangs de $(X_1, \dots, X_n)$ et $(S_1, \dots, S_n)$ les rangs de $(Y_1, \dots, Y_n)$

a) $Ex: si\ (X_i) = (11.1, 13.2, 7.8)\ alors\ (R_i) = (2,3,1)$

2- Coefficient de corrélation de Spearman

1> $R_s = 1 - \frac{6\sum D_i^2}{n(n^2-1)} - -D_i = R_i - S_i$

2> Coefficient de Spearman mesure le **degré de dépendance monotne** entre X et Y.

3> Fonction pivotale (c'est mieux)

a） H_0: $\left(\sqrt{\frac{n-3}{1.06}}\right)\frac{1}{2}ln\left(\frac{(1+R_s)}{1-R_s}\right)\sim^{app}N(0,1)$

b） RC

a） $W=\left\{\left|\left(\sqrt{\frac{n-3}{1.06}}\right)\frac{1}{2}ln\left(\frac{(1+R_s)}{1-R_s}\right)\right|>u_{1-\frac{\alpha^*}{2}}\right\}$