Définition 11 La suite (X_n) converge presque sûrement vers a si $\forall\varepsilon$ et η, $\exists n_0$ tel que $n > n_0$ entraine :

$$P(|X_n - a| > \varepsilon) \cup \ldots \cup (|X_{n+i} - a| > \varepsilon) \cup \ldots) < \eta$$

On note
- $X_n \xrightarrow{p.s.} a$;
- $X_n \xrightarrow{p.s.} X \Rightarrow (X_n - X) \xrightarrow{p.s.}$.

Condition : $X_n \xrightarrow{p.s.} a \Rightarrow \max_i |X_{n+i} - a| \xrightarrow{P} 0$

Définition 12 La suite (X_n) converge dans L^p vers X si $\lim_{n\to+\infty} E(|X_n - X|^p) = 0$

Définition 13 La suite (X_n) converge en loi vers X de fonction de répartition F si en tout point de continuité de F la suite (F_n) des fonctions de répartition de (X_n) converge vers F. On note $(X_n) \xrightarrow{L} X$

Proposition 8 On a $(X_n) \xrightarrow{L} X$ et $(Y_n) \xrightarrow{L} C$.
- Cvg p.s. ⇒ Cvg en P ⇒ Cvg en Loi
- $X_n + Y_n \xrightarrow{L} X + C$
- $X_n Y_n \xrightarrow{L} CX$
- $\frac{X_n}{Y_n} \xrightarrow{L} \frac{X}{C}$ Si $C \neq 0$

Théorème 1 (Loi des grands nombres) Si (X_n) est une suite de v.a. indépendantes de même loi et d'espérance μ, alors $\frac{X_1+\cdots+X_n}{n}$ converge p.s. vers μ.

Théorème 2 Théorême central limite : (X_n) suite de v.a. indépendantes de même loi, d'espérance μ et de variance σ^2 finie, alors

$$\frac{\frac{1}{n}\sum_{i=1}^n X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{L} \mathcal{N}(0,1)$$

quand $n \to \infty$

Pour la loi Binomiale, on a : $P(X = x) \simeq \Phi\left(\frac{x - np + 0.5}{\sqrt{np(1-p)}}\right)$

Proposition 9 Autres approximations :
- Si $X_\lambda \sim P(\lambda)$ alors $\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow{L} \mathcal{N}(0,1)$ quand $\lambda \to \infty$
- Si $X_n \sim \chi^2_n$ alors $\frac{X_n - n}{\sqrt{2n}} \xrightarrow{L} \mathcal{N}(0,1)$ quand $n \to \infty$

3 Echantillonnage

3.1 Notation

Définition 14 On appellera échantillon le n-uplet de variables aléatoires $(X_1, \ldots, X_n)$. Une réalisation de cet échantillon est alors le n-uplet $(x_1, \ldots, x_n)$.

3.2 Statistique $\overline{X}$

Définition 15 La statistique $\overline{X}$ ou moyenne empirique est $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Si X a une espérance μ, alors $E(\overline{X}) = \mu$. Si X a une variance σ^2, alors $Var(\overline{X}) = \frac{\sigma^2}{n}$. Si X a une espérance μ, alors $\overline{X} \xrightarrow{p.s.} \mu$.

3.3 Statistiques S^2 et S^{*2}

Définition 16 La statistique S^2 ou variance empirique est
- $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$. $E(S^2) = \frac{n-1}{n}\sigma^2 = E[X_i^2] - E[\overline{X}^2]$;
- $S^{*2} = \frac{n}{n-1} S^2$;
- $E[S^{*2}] = \sigma^2$.

S^{*2} est un estimateur sans biais de σ^2.

Proposition 10 Si X a une variance σ^2, alors $S^2 \xrightarrow{p.s.} \sigma^2$.

3.4 Généralisation : moments empiriques

Définition 17
- Moment empirique centré d'ordre k :

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^k$$

- Moment empirique non centré d'ordre k :

$$\hat{m}_k = \frac{1}{n}\sum_{i=1}^n (X_i)^k$$

Les moments empiriques d'un échantillon iid $(X_1, \ldots, X_n)$ convergent p.s. vers les moments théoriques correspondants de X.

3.5 Fonction de répartition empirique

Définition 18 Fonction de répartition empirique : $\hat{F}(x) = \frac{1}{n}\sum_{i=1}^n 1_{(X_i \le x)}$. On a $\forall x \in \mathbb{R}$ $\hat{F}(x) \xrightarrow{ps} F(x)$. De plus $D_n = \sup_x |\hat{F}(x) - F(x)|$ est une v.a. qui ne dépend plus de X.

3.6 Echantillon gaussiens

On suppose dans ce paragraphe que $X \sim \mathcal{N}(\mu, \sigma^2)$

Définition 19
- $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
- $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$. $\overline{X}$ et S^2 sont indépendantes.

3.7 loi jointe et vraisemblance

Définition 20 Soit $x_1, \ldots, x_n$ une réalisation de $X_1, \ldots, X_n$ d'une v.a. X. X suit une probabilité dépendant d'un paramètre réel θ.
- X est continue : $L(\theta; x_1, \ldots, x_n) = f(x_1, \ldots, x_n; \theta) = \Pi_{i=1}^n f(x_i; \theta)$
- X est discrète : $L(\theta; x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n; \theta) = \Pi_{i=1}^n P(X_i = x_i; \theta)$
- L'application $\theta \to L(\theta; x_1, \ldots, x_n)$ est appelée vraisemblance de θ. C'est une fonction de vraisemblance.
- $x_1, \ldots, x_n \to L(\theta; x_1, \ldots, x_n)$ est une probabilité ou une densité.

4 Estimation

4.1 formalisation et notation

Définition 21 On appelle estimateur du paramètre de θ toute fonction de l'échantillon $X_1, \ldots X_n$.

Définition 22 Si $T = \varphi(X_1, \ldots X_n)$ est un estimateur, $t = \varphi(x_1, \ldots, x_n)$, réalisation de cet estimateur, est appelée estimation.

4.2 Estimateur convergent

Définition 23 Un estimateur T du paramètre θ est convergent si T converge en probabilité vers θ quand $n \to \infty$, c-à-d si $\forall\epsilon > 0$, $\lim_{n\to\infty} P(|T - \theta| > \epsilon) = 0$.

4.3 Estimateur sans biais

Définition 24
- Un estimateur T est dit sans biais si quelle que soit la taille n de l'échantillon, et θ la valeur du paramètre, $E(T) = \theta$.
- La quantité $E(T) - \theta$ est appelée biais de l'estimateur T.
- Un estimateur est dit asymptotiquement sans biais si quel que soit θ, $\lim_{n\to\infty} E(T) = \theta$.
- Si T est un estimateur sans biais ou asymptotiquement sans biais et si $\lim_{n\to\infty} Var(T) = 0$, alors T est convergent.

4.4 Précision

Définition 25 La précision d'un estimateur T est mesurée par $R(T, \theta) = E[(T - \theta)^2] = Var(T) + (E(T) - \theta)^2$. T_1 est plus précis que T_2 si $R(T_1, \theta) < R(T_2, \theta)$.

4.5 Exhaustivité

Définition 26 Une statistique T est exhaustive si et seulement si il existe 2 fonctions g et h telles que :
- $L(\theta; x_1, \ldots, x_n) = g(t, \theta) h(x_1, \ldots, x_n)$;
- ou encore $L(\theta; X_1, \ldots, X_n) = g(T, \theta) h(X_1, \ldots, X_n)$.

4.6 Estimateur sans biais et de variance minimale

Théorème 3
- S'il existe un estimateur T de θ sans biais de variance minimale, alors il est unique presque sûrement.
- S'il existe une statistique exhaustive U pour le paramètre de θ, alors l'estimateur T de θ sans biais de variance minimale ne dépend que de U.

4.7 Inégalité de Fréchet–Darmois–Cramer–Rao et estimateur efficace

Théorème 4 Si les conditions suivantes de Cramer-Rao :
- Le support de X (ensemble pour lequel $f(x, \theta) > 0$) est indépendant de θ
- $\frac{\partial L}{\partial \theta}(\theta; x_1, \ldots, x_n)$ existe et est continue par rapport à θ
- $I_n(\theta) = E[(\frac{\partial \ln L}{\partial \theta}(\theta; X_1, \ldots, X_n))^2]$ finie
- $\frac{\partial L}{\partial \theta}$ et $\hat{u}\frac{\partial L}{\partial \theta}$ intégrables par rapport à l'échantillon

sont vérifiées, alors $Var(\hat{u}) \geq \frac{(u'(\theta))^2}{I_n(\theta)}$ où u est une fonction quelconque. et l'on a $I_n(\theta) = E[(\frac{\partial \ln L}{\partial \theta})^2] = -E(\frac{\partial^2 \ln L}{\partial \theta^2})$.

Proposition 11 Si les conditions de Cramer-Rao sont vérifiées, on a :
- $E[\frac{\partial \ln L}{\partial \theta}(\theta; X_1, \ldots, X_n)] = 0$;
- $I_n(\theta) = E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]$.

Définition 27 Les conditions de Cramer-Rao étant vérifiées, un estimateur sans biais dont la variance atteint la borne de Cramer-Rao est appelé estimateur efficace.

Théorème 5 Les conditions de Cramer-Rao étant vérifiées, alors $\hat{u}$ est un estimateur efficace de $u(\theta) \Longleftrightarrow \exists$ une fonction A indépendante de $X_1, \ldots, X_n$ telle que $\frac{\partial \ln L}{\partial \theta}(\theta; X_1, \ldots, X_n) = A(n, \theta)(\hat{u} - u(\theta))$.

Définition 28 Un estimateur sans biais est dit asymptotiquement efficace si les conditions de Cramer-Rao sont vérifiées et si la variance tend vers la borne de Cramer-Rao lorsque la taille de l'échantillon tends vers l'infini.

Proposition 12 Si $\hat{u}$ est un estimateur efficace de $u(\theta)$, alors $Var(\hat{u}) = \frac{u'(\theta)}{A(n,\theta)}$. De plus un estimateur efficace est exhaustif.

5 Méthodes d'estimation

5.1 Estimation de l'espérance et de la variance

Estimateur sans biais, convergents, a distribution asymptotique normale :
- de l'espérance : $\hat{\theta} = \overline{X}$
- de la variance : $\hat{\theta} = S^{*2}$

5.2 Maximum de vraisemblance

$$\begin{cases} \frac{\partial \ln L}{\partial \theta}(\theta; x_1, x_2, \ldots, x_n) = 0 \\ \frac{\partial^2 \ln L}{\partial \theta^2}(\theta; x_1, x_2, \ldots, x_n) < 0 \end{cases} \Longleftrightarrow \theta = \hat{\theta}_{MV}$$

5.3 Propriétés

Proposition 13 Soit $\hat{\theta}$ un estimateur du maximum de vraisemblance de θ et u une application de $\mathbb{R} \mapsto \mathbb{R}$. Alors, sous certaines conditions de régularité, $u(\hat{\theta})$ est un estimateur du maximum de vraisemblance.

Proposition 14 S'il existe un estimateur efficace de θ, il est identique (p.s.) à l'unique solution de l'équation de vraisemblance.

Proposition 15 S'il existe une statistique exhaustive T, alors l'estimateur du maximum de vraisemblance est fonction de T.

Proposition 16 Soit θ_0 la vraie valeur du paramètre. Sous certaines conditions, il existe une suite $(\hat{\theta})_{n\geq 1}$ de solutions de l'équation de vraisemblance qui converge en probabilité vers θ_0 : $\hat{\theta}_n \xrightarrow{P} \theta_0$

Proposition 17 Sous certaines conditions de régularité, pour toute suite $(\hat{\theta}_n)_{n\geq 1}$ de solutions de l'équation de vraisemblance t.q. $\hat{\theta} \xrightarrow{P} \theta_0$, on a :

$$\frac{\hat{\theta}_n - \theta_0}{\sqrt{1/I_n(\theta_0)}} \xrightarrow{L} \mathcal{N}(0,1)$$

5.4 Méthode des moments

5.4.1 Principe

On écrit : $(m_1, \ldots, m_k) = g(\theta)$
Si g est inversible on en déduit : $\theta = g^{-1}(m_1, \ldots, m_k)$
La méthode des moments consiste à prendre :

$$\hat{\theta}_m = g^{-1}(m_1, \ldots, m_k)$$

8.3.1 Test du signe

On compte le nombre de $D_i > 0$ (soit $Y_i > X_i$). Nous avons donc une $B(n,p)$. Le test d'égalité est donc $p = 1/2$.

$$\begin{cases} H_0 : p = 1/2 (\Longleftrightarrow \text{médiane } M = 0) \\ H_1 : p > 1/2 (\Longleftrightarrow M > 0) \end{cases}$$

En cas de valeurs égales, on supprime simplement les couples correspondant à ces égalité.

8.3.2 Test de Wilcoxon signé

On ordonne les valeur D_i par ordre de valeur absolues croissantes (en supprimant les ex-aequo), puis à calculer la somme W_+ des rangs associés aux valeurs positives.
Sous H_0 : $E(W_+) = \frac{n(n+1)}{4}$ et $Var(W_+) = \frac{n(n+1)(2n+1)}{24}$
Pour la région critique, on utilise les tables statistiques, avec $W : W^+ > A$. Pour $n > 19$, on utilise le fait que W_+ est asymptotiquement gaussien.

9 Tests d'adéquation

$$\begin{cases} H_0 : X \sim L \\ H_1 : X \text{ ne suit pas } L \end{cases}$$

9.1 Test du χ^2

9.1.1 Loi multinomiale

Définition 31 On considère une expérience aléatoire pouvant prendre K valeurs avec les probabilité $p_1, \ldots, p_K$. Ayant réalisé n fois cette expérience, il est possible de définir le vecteur aléatoire $(N_1, \ldots, N_K)$ où N_K est le nombre de fois où la valeur k a été obtenue. Ce vecteur suit alors par définition une loi multinomiale de paramètre $p_1, \ldots, p_K, n$. Sa loi de probabilité est définie par $P(N_1 = n_1, \ldots, N_K = n_K) = \frac{n!}{n_1! \ldots n_K!} p_1^{n_1} \ldots p_K^{n_K}$.

Théorème 8 Si $(N_1, \ldots, N_k)$ est un vecteur aléatoire suivant une loi multinomiale de paramètres $p_1, \ldots, p_K, n$, alors $\sum_{k=1}^K \frac{(N_k - np_k)^2}{np_k}$ suit asymptotiquement une loi du χ^2 à $K - 1$ degrés de liberté.

9.1.2 Principe du test

La première étape de ce test consiste à définir K classes en répartissant les valeurs possibles de la v.a. X en K sous-ensemble. On peut alors associer à un échantillons de taille n le vecteur aléatoire $(N_1, \ldots, N_k)$. Il est alors possible de calculer les probabilités p_k de chacune des classes à partir de la définition des classes et de la loi de probabilité L.

Ce qui donne : $\begin{cases} H_0 : \text{les paramètres sont les } p_k \\ H_1 : \text{les paramètres ne sont pas les } p_k \end{cases}$

Fonction pivotale : $D^2 = \sum_{k=1}^K \frac{(N_k - np_k)^2}{np_k} \sim \chi^2_{K-1}$

Région critique : $D^2 = \sum_{k=1}^K \frac{(N_k - np_k)^2}{np_k} > \chi^2_{K-1,1-\alpha}$

9.1.3 Remarques

- Ceci est applicable uniquement si $np_k > 5$. Sinon, il faut faire des regroupement par classe de façon à ce qu'on aie $np_k > 5$.
- Si certains paramètres de la loi de l'hypothèse H_0 ne sont pas connus, il est possible de les remplacer par leur estimateur du maximum de vraissemblance, mais alors $D^2 = \sum_{k=1}^K \frac{(N_k - np_k)^2}{np_k} \sim \chi^2_{K-1-r}$ où r est le nombre de paramètres réel estimés.
- En pratique, on utilise souvent l'expression suivante : $D^2 = \sum_{k=1}^K \frac{N_k^2}{np_k} - n$

9.1.4 Application aux tables de contingence

On considère 2 variables aléatoires X et Y ne pouvant prendre respectivement que r et s valeurs (variables quantitatives). La donnée d'un échantillon $(X_1, Y_1), \ldots, (X_n, Y_n)$ permet alors de définir une table de contingence $(N_{ij}, i \in [1, r], j \in [i, s])$ où la modalité i à été prise simultanément avec la modalité j. On définit aussi les marges $N_{i.} = \sum_{j\in[1,s]} N_{ij}$ et $N_{.j} = \sum_{i\in[1,s]} N_{ij}$:

X^Y	1	$\cdots$	j	$\cdots$	s	
1	N_{11}	$\cdots$	N_{1j}	$\cdots$	N_{1s}	$N_{1.}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
i	N_{i1}	$\cdots$	N_{ij}	$\cdots$	N_{is}	$N_{i.}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
r	N_{r1}	$\cdots$	N_{rj}	$\cdots$	N_{rs}	$N_{r.}$
	$N_{.1}$	$\cdots$	$N_{.j}$	$\cdots$	$N_{.s}$	

On cherche à verifier si deux variables X et Y sont indépendantes. Pour ceci, on effectue le test :
$\begin{cases} H_0 : X \text{ et } Y \text{ sont indépendantes} \\ H_1 : X \text{ et } Y \text{ ne sont pas indépendantes} \end{cases}$
On a donc

$$D^2 = \sum_i \sum_j \frac{(N_{ij} - \frac{N_{i.} N_{.j}}{n})^2}{\frac{N_{i.} N_{.j}}{n}} \sim \chi^2_{(r-1)(s-1)}$$

9.2 Test de Kolmogorov

Soit F la fonction de répartition empirique dont est issu l'échantillon et F_0 la fonction de répartition de X sous H_0, alors le test est : $\begin{cases} H_0 : F = F_0 \\ H_0 : F \neq F_0 \end{cases}$

$D_n = \sup_x |\hat{F}(x) - F_0(x)|$ a une distribution qui, sous l'hypothèse H_0, ne dépend plus de la loi de X.
Région critique : $D_n > d_{n,1-\alpha^*}$ ($d_{n,1-\alpha^*}$ sur les tables statistiques)

$$D_n = \max_i \max\left(|\hat{F}(x_i) - F_0(x_i)|, |\hat{F}(x_i^-) - F_0(x_i)|\right)$$

Remarques :
- Le test n'est applicable que pour les v.a. continues.
- Si le test de K. est applicable, il est plus rigoureux que le test du χ^2 puisqu'il ne necessite aucune approximation.
- La loi de référence doit être parfaitement connue. Lorsqu'on veut vérifier si une distribution est gaussienne sans préciser plus les paramètres de la loi, le test de K. n'est donc pas applicable ; toutefois, un test de normalité s'appuyant sur la statistique D_n a pu être défini de la façon suivante : on estime les paramètres de la distribution par $\overline{X}$ et S^{*2} et on prend comme région critique :
  - si $\alpha = 0.05$ $(\sqrt{n} + \frac{0.85}{\sqrt{n}} - 0.01)D_n > 0.895$;
  - si $\alpha = 0.01$ $(\sqrt{n} + \frac{0.85}{\sqrt{n}} - 0.01)D_n > 1.035$.

On a alors $D_n^* = \sup_x |\hat{F}(x) - \Phi\frac{x - \overline{x}}{S^*}|$
Ce test de normalité est appelé test de Stephens (ou de Lilliefors). De la même façon, on peut faire un test d'exponentialité : on détermine D_n en prenant $\overline{X}$ comme estimateur du paramètre de la loi exponentielle et on prend comme région critique :

- si $\alpha = 0.05$ $(\sqrt{n} + \frac{0.5}{\sqrt{n}} - 0.26)D_n > 1.094$;
- si $\alpha = 0.01$ $(\sqrt{n} + \frac{0.5}{\sqrt{n}} - 0.26)D_n > 1.308$.

10 Analyse de la variance

10.1 Estimation des paramètres du modèle

Soient k échantillons de taille n_k et de loi :

$$X_k^i \sim \mathcal{N}(\mu_k^i, \sigma^2), \forall i \in [0, \ldots, n_k]$$

La méthode du maximum de vraisemblance donne :
- $\mu_{k_{MV}} = \overline{X_k}$
- $\hat{\sigma}^2_{MV} = \frac{1}{N}\sum_{k=1}^K \sum_{i=1}^{n_k} (X_k^i - \overline{X_k})^2 = \frac{1}{N}\sum_{k=1}^K (n_k - 1) S_k^{*2}$

Mais il est préférable de retenir comme estimateur sans biais de l'espérance : $\frac{1}{N-K}\sum_k^K (n_k - 1) S_k^{*2}$, appelé MSW
On on déduit également que : $\overline{X} = \frac{1}{N}\sum_{k=1}^K n_k \overline{X_k}$

10.2 Formule de l'analyse de la variance

$$\underbrace{\sum_k \sum_i (X_k^i - \overline{X})^2}_{SST} = \underbrace{\sum_k \sum_i (X_k^i - \overline{X_k})^2}_{SSW} + \underbrace{\sum_k (\overline{X} - \overline{X_k})^2}_{SSB}$$

Dispersion totale = Dispersion intra-classe + Dispersion inter-classe

10.3 Moyenne des dispersions

- variance totale : $MST = \frac{SST}{N-1}$;
- variance intra-modalités : $MSW = \frac{SSW}{N-K}$;
- variance inter-modalités : $MSB = \frac{SSB}{K-1}$.

10.4 vérification des hypothèses du modèle

- test de normalité : test de Kolmogorov ou du χ^2
- test d'égalité des variances (Bartlett) :

$$\begin{cases} H_0 : \sigma_1 = \sigma_2 = \ldots = \sigma_K \\ H_1 : \exists k, l \quad \sigma_k \neq \sigma_l, \end{cases}$$

On rejette H_0 si $B \geq \chi^2_{K-1,1-\alpha^*}$, avec :

$$B = (N - K)\ln(MSW) - \sum_{k=1}^K (n_k - 1)\ln S_k^{*2}$$

10.5 Test de l'analyse de la variance

Pour tester si les K population ont la même distribution, il suffit de faire le test : $\begin{cases} H_0 : \sigma_1 = \sigma_2 = \ldots = \sigma_K \\ H_1 : \exists k, l \quad \sigma_k \neq \sigma_l, \end{cases}$
et la région critique est de la forme : $\frac{MSB}{MSW} > F_{K-1,N-K,1-\alpha}$
On présente souvent les calculs intermédiaires dans un «tableau d'analyse de la variance» :

Variation	Inter-modalité	Intra-modalité	total
° de liberté	$K - 1$	$N - K$	$N - 1$
$\sum$ des ²	SSB	SSW	SST
Moy. des ²	$MSB = \frac{SSB}{K-1}$	$MSW = \frac{SSW}{N-K}$	
Rapport	$F = \frac{MSB}{MSW}$		

10.6 Effet du facteur : procédure LSD

- si on accepte H_0, le facteur n'a pas d'effet, et la procédure s'arrête.
- sinon, il est interessant de tester l'égalité des moyennes 2 a 2. La région critique pour la comparaison des niveaux k et l est :

$$W_{kl} : \quad \frac{|\overline{X_k} - \overline{X_l}|}{\sqrt{MSW\left(\frac{1}{n_k} + \frac{1}{n_l}\right)}} > t_{N-K;1-\frac{\alpha}{2}}$$

10.7 Le calcul du début à la fin

- Calculer les $\overline{X_k}, S_k^{*2}$, $k = 1, \ldots, K$. Rappel : $\overline{X} = \frac{1}{N}\sum_k n_k \overline{X_k}$.
- $SSW = \sum_k (n_k - 1) S_k^{*2}$, $MSW = \frac{SSW}{N-K}$.
- $SSB = \sum_k n_k(\overline{X_k} - \overline{X}) = \sum_k n_k \overline{X_k^2} - N\overline{X^2}$, $MSB = \frac{SSB}{K-1}$.
- La région critique est W : $F = \frac{MSB}{MSW} >$ (ou $<$ selon le test) $F_{K-1,N-K,1-\alpha^*}$.
- Éventuellement : tester la validité du modèle (test de normalité de chaque échantillon avec *Lillefors-Stephen*, test d'égalité des variances avec *Fisher* pour deux échantillons, *Bartlett* pour plus : voir **10.4**).

10.8 Le test de Kruskal-Wallis

C'est un équivalent non paramétrique de l'analyse de la variance, qui ne suppose pas la normalité des échantillons. Le test repose sur la statistique $H = \frac{SSB}{MST}$ calculée sur les rangs : on mélange toutes les valeurs, on les range par ordre croissant et on leur attribue un rang.
Les valeurs de SSB et SST deviennent :

$$\begin{cases} SSB = \sum_{k-1}^K \frac{1}{n_k}(\sum_{i=1}^{n_k} R_i^k)^2 - \frac{N(N+1)^2}{4} \\ SST = \frac{N(N+1)(N-1)}{12} \end{cases}$$

et donc $MST = \frac{SST}{N-1} = \frac{N(N+1)}{12}$ d'où :

$$H = \frac{SSB}{MST} = \frac{12}{N(N+1)}\sum_{k-1}^K \frac{1}{n_k}(\sum_{i=1}^{n_k} R_i^k)^2 - 3(N+1)$$

Si N n'est pas trop petit, H suit approximativement une loi χ^2_{K-1}. La région critique du test est donc :

$$W : \quad H > \chi^2_{K-1;1-\alpha}$$