

MÉTHODE DU CHI-DEUX MINIMUM (H3)

(10 / 05 / 2020, © Monfort, Dicostat2005, 2005-2020)

La **méthode du chi-deux minimum** est une méthode d'**estimation** d'un **paramètre** associée à la notion de **loi du chi-deux** : la **distance entre lois** considérée est définie par une **variable du chi-deux** (cf **estimateur à distance minimale**).

(i) Dans le cadre du **test du chi-deux**, on désigne par $X = (X_1, \dots, X_K)$ un **vecteur aléatoire** tq $e_K' X = \sum_{k=1}^K X_k = N \in \mathbf{N}^*$.

On pose $p = (p_1, \dots, p_K)$, avec $p_k \geq 0$ ($\forall k$) et $\sum_{k=1}^K p_k = 1$, ie $p \in S_K$ (**simplexe** de $\mathbf{R}^K$), et l'on suppose que p dépend d'un paramètre $\theta \in \Theta$, avec $\Theta \subset \mathbf{R}$, ie $p(\theta) = (p_1(\theta), \dots, p_K(\theta))$. La **statistique du chi-deux** :

$$(1) \quad K^2 = \sum_{k=1}^K (N p_k)^{-1} (X_k - N p_k)^2,$$

dépend donc aussi de θ par l'intermédiaire des p_k .

La **méthode du chi-deux minimum** consiste à choisir pour estimateur $\theta_N^\#$ de θ une valeur qui rend minimum la fonction quadratique :

$$(2) \quad \theta \in \Theta \mapsto K^2(\theta) = \sum_{k=1}^K (N p_k(\theta))^{-1} (X_k - N p_k(\theta))^2.$$

Cette valeur est donc solution de l'équation (condition du premier ordre) :

$$(3) \quad dK^2(\theta) / d\theta = 0.$$

(ii) On établit la **convergence en loi** suivante :

$$(4) \quad \mathcal{L}(\theta_N^\#) \rightarrow_{N \rightarrow \infty} \mathcal{N}_1(\theta^*, (I(\theta^*))^{1/2}) \quad (\text{loi normale}),$$

où $I(\theta)$ est la quantité d'**information de FISHER** au point θ et θ^* la « **vraie valeur** » (inconnue) de θ .

(ii) La méthode s'étend directement au cas où $\Theta \subset \mathbf{R}^Q$ (paramètre vectoriel réel).

Les calculs sont généralement moins simples que ceux impliqués par la **méthode du maximum de vraisemblance** : une simplification réside dans la **méthode du chi-deux minimum modifiée**, indiquée infra.

(iii) Plus généralement, soit $(\Omega, \mathcal{F}, P_\theta)_{\theta \in \Theta}$ un **modèle statistique**, I un ensemble d'**indices** fini et, $\forall i \in I$, soit $\xi_i : \Omega \mapsto \mathbf{R}$ une **vars** dont on observe un **échantillon** $(X_{i,1}, \dots, X_{i,N(i)})$ de taille N_i (où les $n(i)$ désignent les indices n_i).

Soit $\Pi_i = \{R_{i1}, \dots, R_{i,K(i)}\}$ ($\forall i \in I$) une **suite** (finie) de **partitions** de la droite réelle $\mathbf{R}$. On pose :

$$(5) \quad p_{ik}(\theta) = P_{\theta}([\xi_i \in R_{ik}]), \quad \forall (i, k) \in I \times \{1, \dots, K_i\}.$$

Par suite, $\sum_{k=1}^{K(i)} p_{ik}(\theta) = 1, \forall i \in I$.

Soit enfin N_{ik} le nombre d'**observations** X_{ik} appartenant à R_{ik} et $f_{ik} = N_{ik} / N_{i\cdot}$, avec $N_{i\cdot} = \sum_{k=1}^{K(i)} N_{ik}$.

La **méthode du chi-deux minimum** consiste alors à minimiser la **statistique du chi-deux** :

$$(6) \quad K^2(\theta) = \sum_{i \in I} \left\{ N_{i\cdot} \sum_{k=1}^{K(i)} \{p_{ik}(\theta)\}^{-1} (f_{ik} - p_{ik}(\theta))^2 \right\}$$

pr à $\theta \in \Theta$.

(iv) Pour simplifier les calculs, la **méthode du chi-deux minimum modifiée** (H. JEFFREYS) consiste à minimiser pr à θ l'expression (cf **statistique du chi-deux modifiée**) :

$$(7) \quad K^2(\theta)^* = \sum_{i \in I} \left\{ N_{i\cdot} \sum_{k=1}^{K(i)} (f_{ik})^{-1} (f_{ik} - p_{ik}(\theta))^2 \right\}.$$

dans laquelle les f_{ik} remplacent les $p_{ik}(\theta)$ des dénominateurs de la statistique (6).

(v) Les deux estimateurs précédents (solutions resp de (6) et (7)) sont, sous des hypothèses assez larges, convergents en probabilité, asymptotiquement normaux et efficaces (cf **estimateur convergent**, **normalité asymptotique**, **efficacité**, **efficacité asymptotique**).

En pratique, on leur préfère généralement l'**estimateur du maximum de vraisemblance**.