INÉGALITÉ DE CRAMER-DARMOIS-FRÉCHET-RAO (G04, G06, H01, H05)
(30 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

L'**inégalité de CRAMER - DARMOIS - FRÉCHET - RAO** est une inégalité classique exprimant, sous certaines **conditions de régularité**, l'impossibilité de diminuer la **dispersion** (**variance**) d'un estimateur en deçà d'une certaine « borne ». Dans le **contexte** d'un **modèle dominé** quelconque, cette inégalité :

(a) relie la variance d'un estimateur et sa quantité d'**information** ;

(b) permet aussi de définir une notion importante : l'**efficacité** d'un estimateur.

Elle est aussi appelée **inégalité de CRAMER - RAO** dans les pays de langue anglaise, et **inégalité de DARMOIS - FRÉCHET** en France.

(i) Soit $(\mathcal{X}, \mathcal{B}, P_\theta^X)_{\theta \in \Theta}$ un **modèle image** dans lequel $\Theta \subset \mathbf{R}$ (**paramètre** scalaire). On définit la **vraisemblance** (ou **dérivée de NIKODYM-RADON**) :

(1) $\quad f(x, \theta) = dP_\theta^X / d\mu\,(x), \qquad \forall\,(x, \theta) \in \mathcal{X} \times \Theta,$

où μ est une **mesure positive** sur $\mathcal{B}$, qui ne dépend pas de θ. Etant donné une fonction **mesurable** $g : \Theta \mapsto \mathbf{R}$, on considère un **estimateur** $T : \Omega \mapsto \mathbf{R}$ de $\tau = g(\theta)$ basé sur $t : \mathcal{X} \mapsto \mathbf{R}$, ie défini par $T = t(X)$.

Par suite, si les trois conditions suivantes :

(a) $f(x, \theta) > 0,\ \forall\,(x, \theta) \in \mathcal{X} \times \Theta$;

(b) $D^2 f$ est de classe C^1 et l'on peut dériver une fois sous le signe somme $\int$;

(c) $T \in \mathcal{L}_{\mathbf{R}}^2(\Omega, \mathcal{T}, P_\theta),\ \forall\,\theta$, et T est **sans biais** (ie $E_\theta\, t(X) = g(\theta),\ \forall\,\theta \in \Theta$) ;

sont vérifiées, alors T vérifie l'**inégalité de C.H. CRAMER - G.E. DARMOIS - R.M. FRÉCHET - C.R. RAO** suivante :

(2) $\quad V_\theta T = V_\theta\, t(X) \geq (g'(\theta))^2 / I(\theta), \qquad \forall\,\theta \in \Theta,$

dans laquelle $V_\theta T = \int \{t(x) - g(\theta)\}^2\, dP_\theta^X(x)$ est la variance de T calculée avec la loi P_θ^X, $g'(\theta) = D\,g(\theta)$ est la **dérivée** de g et $I(\theta)$ est la **quantité d'information de R.A. FISHER**, définie selon (cf **information de FISHER**) :

(3) $\quad I(\theta) = V_\theta \{D_2 \operatorname{Log} f(X, \theta)\} = E_\theta \{D_2 f(X, \theta) / f(X, \theta)\}^2,$

où D_2 désigne la **dérivation** pr au second argument θ.

Si l'on peut dériver deux fois sous le signe $\int$, l'expression (3) s'écrit alors :

(4) $I(\theta) = - E_\theta D_2^2 \operatorname{Log} f(X, \theta)$.

Lorsque T est un estimateur de $\tau = g(\theta)$ tq :

(5) $V_\theta T = (g'(\theta))^2 / I(\theta), \quad \forall \theta \in \Theta$,

on dit que T est un **estimateur (sans biais) de variance minimum**.

Si S = s (X) est un autre estimateur (sans biais) de τ, on appelle **efficacité (absolue)** de S le rapport e_S entre sa variance et celle de T (ie la borne inférieure définie dans l'inégalité (2)) :

(6) e_S ou $e(S) = (g'(\theta))^2 / (I(\theta) \cdot V_\theta S) \in [0, 1]$.

(ii) L'inégalité précédente se généralise à un paramètre vectoriel. On suppose alors que $(\mathcal{X}, \mathcal{B}, P_\theta^X)_{\theta \in \Theta}$ est un **modèle statistique paramétrique régulier**, avec $\Theta \in \mathbf{R}^Q$, que $g : \Theta \mapsto \mathbf{R}^K$ est une fonction mesurable donnée et que $T : \Omega \mapsto g(\Theta)$ est une **statistique régulière**, fondée sur une **application mesurable** $t : \mathcal{X} \mapsto g(\Theta)$, et supposée être un **estimateur sans biais** de $\tau = g(\theta)$.

L'**inégalité de CRAMER - DARMOIS - FRÉCHET - RAO** prend alors la forme d'une **inégalité matricielle** (ie d'une inégalité définie par les **formes quadratiques** associées : $A \geq B \Leftrightarrow h' A h \geq h' B h, \forall h \in \mathbf{R}^K$) :

(7) $V_\theta T = V_\theta t(X) \geq \Delta_\theta I(\theta)^{-1} \Delta_\theta'$,

dans laquelle :

(a) $I(\theta)$ est la (Q,Q)-matrice d'**information de FISHER** ;

(b) Δ_θ est la (K,Q)-matrice **jacobienne** associée à la dérivée $D g(\theta)$, et de terme général :

(8) $\delta_{kq}(\theta) = (\partial g_k / \partial \theta_q)(\theta), \quad \forall (k, q) \in N_K^* \times N_Q^*$.

Comme précédemment, lorsque l'inégalité (7) est saturée à l'égalité, on dit que T est un **estimateur efficace** de $\tau = g(\theta)$.

Une expression (scalaire) analogue à (7) s'exprime en termes de **variance généralisée** selon :

(9) $\text{Dét}(V_\theta T) \geq \text{Dét}(\Delta_\theta I(\theta)^{-1} \Delta_\theta')$.

De plus, si Δ_θ est carrée (ie si K = Q), on obtient :

(10) $\text{Dét}(V_\theta T) \geq \text{Dét}(\Delta_\theta^2) / \text{Dét}\, I(\theta), \quad \forall \theta \in \Theta$.

(iii) On peut étendre les formules (2), (7), (9) et (10) précédentes au cas où T possède un **biais**.

(iv) Lorsque T est un **estimateur convergent** (en probabilité) de $\tau = g(\theta)$ et que sa **loi asymptotique** est **gaussienne**, avec pour **matrice de covariances** la borne $\Delta_\theta \, I(\theta) \, \Delta_\theta'$ de (7), on dit que T est le **meilleur estimateur asymptotiquement normal** (en anglais *« best asymptotically normal estimator »*, ou ban) de $\tau = g(\theta)$.

A titre d'exemple, on montre, sous certaines hypothèses, que l'**estimateur du maximum de vraisemblance** est de ce type.

Néanmoins, un tel estimateur n'est pas en général unique : pour les comparer, on se réfère souvent à leur **efficacité (relative)**, dans le cadre d'un **modèle d'échantillonnage** à distance finie (eg efficacité au second ordre de RAO), ou à leur **efficacité (relative) asymptotique**, dans le cadre d'un modèle d'échantillonnage asymptotique (cf **modèle asymptotique**).