LEMMES DE LE CAM (E, I02)
(24 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

Les **lemmes de LE CAM** sont des résultats de base en **théorie des tests** non paramétriques. Ils sont fondés sur la notion de **suite de lois contigües** (cf **contiguïté probabiliste**).

(i) Soit $(\Omega_n , \mathcal{T}_n , P_n)_{n \in \mathbf{N}}$ une **suite** d'**espaces probabilisés**, $(\mathcal{X}_n , \mathcal{B}_n)_{n \in \mathbf{N}}$ une suite d'**espaces d'observation** et $(X_n)_{n \in \mathbf{N}}$ une suite de **va** $X_n : \Omega_n \mapsto \mathcal{X}_n$. Soit $(P_n^{X(n)})_{n \in \mathbf{N}}$ la suite des **lp** $P_n^{X(n)} = X_n (P_n)$ correspondantes, $\forall$ n $\in$ **N**, et soit $(Q_n^{X(n)})_{n \in \mathbf{N}}$ une autre suite de lois (par commodité, X(n) désigne X_n , etc).

On définit la suite des **vraisemblances** (ie des **densités** pr à des **mesures σ-finies** μ_n définies resp sur $\mathcal{B}_n$) selon (cf **famille de lois dominée**) :

(1) $\quad f_n = dP_n^{X(n)} / d\mu_n , \qquad g_n = dQ_n^{X(n)} / d\mu_n , \qquad \forall n \in \mathbf{N}.$

La suite des **rapports de vraisemblance** est alors définie comme suite des **va** (ou **statistiques**) suivantes :

(2) $$L_n = \begin{cases} g_n / f_n \text{ si } f_n > 0, \\ 1 \text{ si } f_n = g_n = 0, \\ +\infty \text{ si } f_n = 0 \text{ et } g_n > 0, \end{cases}$$

et on lui associe la suite des **fr** :

(3) $\quad F_n^{L(n)} (\lambda) = P_n (L_n (\lambda)), \qquad \forall \lambda \in \mathbf{R}_+ , \qquad \forall n \in \mathbf{N}.$

(ii) **Premier lemme de L. LE CAM**. Ce lemme tient en deux propositions :

(a) s'il existe une fr F_∞ tq :

(4) $\quad F_n^{X(n)} \to_{n \to +\infty} F_\infty$ et $\int_{\mathbf{R}+} \lambda \, dF_\infty (\lambda) = 1,$

alors $(Q_n)_{n \in \mathbf{N}}$ est une suite de lois contigüe à la suite $(P_n)_{n \in \mathbf{N}}$;

(b) si la suite $(P_n)_{n \in \mathbf{N}}$ est tq :

(5) $\quad \mathcal{L} (\text{Log } L_n) \to^{(\text{sous } P(n))}_{n \to +\infty} \mathcal{N}_1 (- \sigma^2 / 2, \sigma^2)$ $\qquad$ (**loi normale**),

alors la suite $(Q_n)_{n \in \mathbf{N}}$ est contigüe à la suite $(P_n)_{n \in \mathbf{N}}$.

(iii) **Deuxième lemme de L. LE CAM**. On suppose que $\mathcal{X}_n = \mathbf{R}^{N(n)}$ (où N(n) désigne N_n), que $x = (x_1 ,..., x_{N(n)})$ et que f_n et g_n s'expriment sous la forme :

(6) $f_n(x) = \Pi_{\alpha=1}^{N(n)} p_n^{\alpha}(x_\alpha)$

et :

$g_n(x) = \Pi_{\alpha=1}^{N(n)} q_n^{\alpha}(x_\alpha)$,

ce qui entraîne :

(7) $\text{Log } L_n(x) = \Sigma_{\alpha=1}^{N(n)} \{\text{Log } q_n^{\alpha}(x_\alpha) - \text{Log } p_n^{\alpha}(x_\alpha)\}$.

On montre alors que les deux hypothèses suivantes :

(a) $\max_{\alpha=1}^{N(n)} P_n \{|(q_n^{\alpha}(X_\alpha) / p_n^{\alpha}(X_\alpha)) - 1| > \varepsilon\} \to_{n \to \infty} 0$;

(b) il existe une suite de statistiques $(S_n)_{n \in \mathbf{N}}$, fondée sur la suite de fonctions mesurables $(s_n)_{n \in \mathbf{N}}$ définies sur $\mathcal{X}_n$ et à valeurs dans **R** (ie tq $S_n = s_n(X_n)$), tq :

(8) $\mathcal{L}(S_n) \to^{(\text{sous } P(n))}{}_{n \to +\infty} \mathcal{N}_1(-\sigma^2 / 4, \sigma^2)$,

impliquent que la suite de va (ou statistiques) Log L_n vérifie les deux propriétés suivantes :

(9)
$$\lim_{n \to +\infty} P_n(|\text{Log } L_n - S_n + (\sigma^2 / 4)| > \varepsilon) = 0,$$
$$\mathcal{L}(\text{Log } L_n) \to^{(\text{sous } P(n))}{}_{n \to +\infty} \mathcal{N}_1(-\sigma^2 / 2, \sigma^2).$$

(iv) **Troisième lemme de L. LE CAM**. Ce lemme se déduit des précédents. Si $(T_n)_{n \in \mathbf{N}}$ est une suite de **statistiques** fondée sur la suite de fonctions mesurables $t_n : \mathcal{X}_n \mapsto \mathbf{R}$ (ie tq $T_n = t_n(X_n)$, $\forall n \in \mathbf{N}$), et si la suite $(T_n, L_n)_{n \in \mathbf{N}}$ vérifie la propriété suivante :

(10) $\mathcal{L}\{(T_n, \text{Log } L_n)'\} \to^{(\text{sous } P(n))}{}_{n \to +\infty} \mathcal{N}_2(\mu, \Sigma)$,

avec $\Sigma = (\sigma_{ij})_{(i,j)}$, $\sigma_{ii} = \sigma_i^2$, $\mu = (\mu_1, \mu_2)'$ et $\mu_2 = \sigma_2^2 / 2$, alors la **loi marginale** asymptotique de T_n (cf **loi asymptotique**) est gaussienne (sous la suite des lois Q_n) :

(11) $\mathcal{L}(T_n) \to^{(\text{sous } Q(n))}{}_{n \to +\infty} \mathcal{N}_1(\mu_1 + \sigma_1^2, \sigma_1^2)$.