Fadoua Balabdaoui-Mohr
Université Paris Dauphine
U.F.R. Mathématiques de la Décision
L3 - Statistiques Mathématiques

Année 2009

Solution de l'exercice 10 du TD1

On sait que $X_1, ..., X_{n_1}$ et $Y_1, ..., Y_{n_2}$ sont des v.a. i.i.d. telles que $Var(X_1) < \infty$, $Var(Y_1) < \infty$ et X_i et Y_j sont indépendantes pour tout $i = 1, ..., n_1$ et $j = 1, ..., n_2$. Posons $\mu = E[X_1] = E[Y_1]$ (les moyennes sont égales par hypothèse). On aimerait bien dire que

$$\begin{pmatrix} \frac{\sqrt{n_1}(\bar{X}_{n_1}-\mu)}{\sigma_X} \\ \frac{\sqrt{n_2}(\bar{Y}_{n_2}-\mu)}{\sigma_Y} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right) \stackrel{\mathcal{L}}{=} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

quand $n_1 \to \infty$ et $n_2 \to \infty$, et où Z_1 et Z_2 sont des v.a. i.i.d $\sim \mathcal{N}(0, 1)$
Problème : Le TCL multi-dimensionnel ne marche pas avec des tailles d'echantillon différentes comme c'est le cas ici ($n_1 \neq n_2$). Le résultat est quand-même **vrai**. En effet, notons

$$Z_{n_1} = \frac{\sqrt{n_1}(\bar{X}_{n_1} - \mu)}{\sigma_X} \text{ et } Z_{n_2} = \frac{\sqrt{n_2}(\bar{Y}_{n_2} - \mu)}{\sigma_Y}.$$

et soit $(t_1, t_2) \in \mathbb{R}^2$. La fonction caractéristique du couple (Z_{n_1}, Z_{n_2}) au point (t_1, t_2) est donnée par

$$\begin{aligned} \Phi_{(Z_{n_1}, Z_{n_2})}(t_1, t_2) &= E\left[e^{i(Z_{n_1}t_1 + Z_{n_2}t_2)}\right] \\ &= E\left[e^{iZ_{n_1}t_1} \times e^{iZ_{n_2}t_2}\right] \\ &= E\left[e^{iZ_{n_1}t_1}\right] E\left[e^{iZ_{n_2}t_2}\right] \text{ par indépendance} \\ &= \Phi_{Z_{n_1}}(t_1) \times \Phi_{Z_{n_2}}(t_2). \end{aligned}$$

Par le TCL ($Var(X_1) < \infty$ et $Var(Y_1) < \infty$), on a que

$$Z_{n_1} = \frac{\sqrt{n_1}(\bar{X}_{n_1} - \mu)}{\sigma_X} \xrightarrow{\mathcal{L}} Z_1 \sim \mathcal{N}(0, 1)$$

et

$$Z_{n_2} = \frac{\sqrt{n_2}(\bar{Y}_{n_2} - \mu)}{\sigma_Y} \xrightarrow{\mathcal{L}} Z_2 \sim \mathcal{N}(0, 1)$$

avec Z_1 et Z_2 sont définies sur le même espace de probabilité que les $X_i, i = 1, ..., n_1$ et les $Y_j, j = 1, ..., n_2$. Par le Théorème de Paul-Lévy, il s'en suit que

$$\lim_{n_1\to\infty} \Phi_{Z_{n_1}}(t_1) = \Phi_{Z_1}(t_1) \text{ et } \lim_{n_2\to\infty} \Phi_{Z_{n_2}}(t_2) = \Phi_{Z_2}(t_2).$$

On conclut que

$$\lim_{n_1,n_2\to\infty} \Phi_{(Z_{n_1},Z_{n_2})}(t_1,t_2) = \Phi_{Z_1}(t_1) \times \Phi_{Z_2}(t_2) = e^{-t_1^2/2} \times e^{-t_2^2/2} = e^{-(t_1^2+t_2^2)/2}$$

qui est aussi la fonction caractéristique d'un vecteur Gaussien d'espérance $(0,0)$ et de matrice variance-covariance égale à la matrice identité de dimension 2×2. Il vient que

$$\begin{pmatrix} \frac{\sqrt{n_1}(\bar{X}_{n_1}-\mu)}{\sigma_X} \\ \frac{\sqrt{n_2}(\bar{Y}_{n_2}-\mu)}{\sigma_Y} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right).$$

Voir aussi une autre démonstration de la convegence en loi de la suite $(Z_{n_1}, Z_{n_2})_{n_1,n_2}$ à la page 4.

D'autre part, on a que

$$\begin{aligned} S^2_{n_1,n_2} &= \frac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1}(X_i-\bar{X}_{n_1})^2 + \sum_{i=1}^{n_2}(Y_i-\bar{Y}_{n_2})^2\right) \\ &= \frac{n_1}{n_1+n_2}\frac{\sum_{i=1}^{n_1}(X_i-\bar{X}_{n_1})^2}{n_1} + \frac{n_2}{n_1+n_2}\frac{\sum_{i=1}^{n_2}(Y_i-\bar{Y}_{n_2})^2}{n_2} \end{aligned}$$

Or, on sait que (cf. Question 1 - Exercice 6)

$$\begin{aligned} \frac{\sum_{i=1}^{n_1}(X_i-\bar{X}_{n_1})^2}{n_1} &\xrightarrow{p.s.} \sigma_X^2 \text{ quand } n_1 \to \infty \\ \frac{\sum_{i=1}^{n_2}(Y_i-\bar{Y}_{n_2})^2}{n_2} &\xrightarrow{p.s.} \sigma_Y^2 \text{ quand } n_2 \to \infty. \end{aligned}$$

Maintenant du fait que $\lim_{n_1\to\infty,n_2\to\infty} n_1/(n_1+n_2) = \lambda$, il s'ent suit que

$$S^2_{n_1,n_2} \xrightarrow{p.s.} \lambda\sigma_X^2 + (1-\lambda)\sigma_Y^2 \tag{1}$$

Pour le voir, considérons l'ensemble

$$E = \left\{\lim_{n_1,n_2\to\infty} S^2_{n_1,n_2} = \lambda\sigma_X^2 + (1-\lambda)\sigma_Y^2\right\}$$

Il est clair que

$$\left\{\lim_{n_1\to\infty}\frac{\sum_{i=1}^{n_1}(X_i-\bar{X}_{n_1})^2}{n_1} = \sigma_X^2 \text{ et } \lim_{n_2\to\infty}\frac{\sum_{i=1}^{n_2}(Y_i-\bar{Y}_{n_2})^2}{n_2} = \sigma_Y^2\right\} \subset E.$$

Or,

$$
\begin{aligned}
&P\left(\left\{\lim_{n_1\to\infty}\frac{\sum_{i=1}^{n_1}(X_i-\bar{X}_{n_1})^2}{n_1}=\sigma_X^2 \text{ et } \lim_{n_2\to\infty}\frac{\sum_{i=1}^{n_2}(Y_i-\bar{Y}_{n_2})^2}{n_2}=\sigma_Y^2\right\}\right)\\
&=1-P\left(\left\{\lim_{n_1\to\infty}\frac{\sum_{i=1}^{n_1}(X_i-\bar{X}_{n_1})^2}{n_1}\neq\sigma_X^2\right\} \text{ ou } \left\{\lim_{n_2\to\infty}\frac{\sum_{i=1}^{n_2}(Y_i-\bar{Y}_{n_2})^2}{n_2}\neq\sigma_Y^2\right\}\right)\\
&\geq 1-\left[P\left(\left\{\lim_{n_1\to\infty}\frac{\sum_{i=1}^{n_1}(X_i-\bar{X}_{n_1})^2}{n_1}\neq\sigma_X^2\right\}\right)+P\left(\left\{\lim_{n_2\to\infty}\frac{\sum_{i=1}^{n_2}(Y_i-\bar{Y}_{n_2})^2}{n_2}\neq\sigma_Y^2\right\}\right)\right]\\
&=1-(0+0)=1
\end{aligned}
$$

d'où $P(E)=1$ et la convergence presque sûre en (1) s'en suit. Par continuité de la fonction $g(x)=1/\sqrt{x}, x>0$, on déduit que

$$\frac{1}{S_{n_1,n_2}}\xrightarrow{p.s.}\frac{1}{\sqrt{\lambda\sigma_X^2+(1-\lambda)\sigma_Y^2}}.$$

Maintenant, soit

$$A_{n_1,n_2}=\left(\frac{\sqrt{n_2}}{\sqrt{n_1+n_2}}\frac{\sigma_X}{S_{n_1,n_2}},-\frac{\sqrt{n_1}}{\sqrt{n_1+n_2}}\frac{\sigma_Y}{S_{n_1,n_2}}\right)$$

On a que

$$A_{n_1,n_2}\xrightarrow{p.s.}\left(\frac{\sqrt{1-\lambda}\;\sigma_X}{\sqrt{\lambda\sigma_X^2+(1-\lambda)\sigma_Y^2}},-\frac{\sqrt{\lambda}\;\sigma_Y}{\sqrt{\lambda\sigma_X^2+(1-\lambda)\sigma_Y^2}}\right)$$

(la démonstration se base sur des argments similaires à ceux utilisés pour montrer la convergence presque sûre en (1), et est laissée donc en exercice). On en déduit que

$$A_{n_1,n_2}\xrightarrow{P}\left(\frac{\sqrt{1-\lambda}\sigma_X}{\sqrt{\lambda\sigma_X^2+(1-\lambda)\sigma_Y^2}},-\frac{\sqrt{\lambda}\sigma_Y}{\sqrt{\lambda\sigma_X^2+(1-\lambda)\sigma_Y^2}}\right)=(A_1,A_2)$$

et donc par le Théorème de Slutsky il vient que

$$A_{n_1,n_2}\begin{pmatrix}\frac{\sqrt{n_1}(\bar{X}_{n_1}-\mu)}{\sigma_X}\\ \frac{\sqrt{n_2}(\bar{Y}_{n_2}-\mu)}{\sigma_Y}\end{pmatrix}=\sqrt{\frac{n_1n_2}{n_1+n_2}}\frac{\bar{X}_{n_1}-\bar{Y}_{n_2}}{S_{n_1,n_2}}\xrightarrow{\mathcal{L}}A\begin{pmatrix}Z_1\\ Z_2\end{pmatrix}=A_1Z_1+A_2Z_2$$

avec

$$A_1Z_1+A_2Z_2\sim\mathcal{N}(0,\sigma^2)$$

et

$$\sigma^2=A_1^2+A_2^2=\frac{(1-\lambda)\sigma_X^2+\lambda\sigma_Y^2}{\lambda\sigma_X^2+(1-\lambda)\sigma_Y^2}. \qquad \square$$

Une autre démonstration pour la convergence en loi de $(Z_{n_1}, Z_{n_2})_{n_1,n_2}$**.**
on utilisera la définition suivante de la convergence en loi : La suite de v.a. $(X_n)_n$ converge en loi vers X si pour toute fonction g continue et à support compact on a que

$$\lim_{n\to\infty} E[g(X_n)] = E[g(X)].$$

(cette définition peut bien évidemment être généralisée à une suite $(X_{n_1,n_2})_{n_1,n_2}$ convergeant en loi vers X quand $n_1, n_2 \to \infty$).
On aura ensuite besoin du résultat suivant :
Soit g une fonction continue à support compact définie sur $\mathbb{R}^2$. Pour tout $\epsilon > 0$, il existe $N_\epsilon \in \mathbb{N}^*$ tel que

$$\sup_{(x,y)\in\mathbb{R}^2} \left| g(x,y) - \sum_{j=1}^{N_\epsilon} f_j(x)h_j(y) \right| \leq \epsilon.$$

où f_i et g_j sont continues à support compact. On posera $g_\epsilon(x,y) = \sum_{j=1}^{N_\epsilon} f_j(x)h_j(y)$.
Soit g une fonction continue à support compact, et soient $\tilde{Z}_1$ et $\tilde{Z}_2$ des v.a. $\sim \mathcal{N}(0,1)$ telles que $\tilde{Z}_1$ et $\tilde{Z}_2$ sont indépendantes. On a que

$$\begin{aligned}
\left| E[g(Z_{n_1}, Z_{n_2})] - E[g(\tilde{Z}_1, \tilde{Z}_2)] \right| &\leq \left| E[g(Z_{n_1}, Z_{n_2})] - E[g_\epsilon(Z_{n_1}, Z_{n_2})] \right| \\
&\quad + \left| E[g_\epsilon(Z_{n_1}, Z_{n_2})] - E[g_\epsilon(\tilde{Z}_1, \tilde{Z}_2)] \right| \\
&\quad + \left| E[g_\epsilon(\tilde{Z}_1, \tilde{Z}_2)] - E[g(\tilde{Z}_1, \tilde{Z}_2)] \right| \\
&\leq E\left[\left| g(Z_{n_1}, Z_{n_2})] - g_\epsilon(Z_{n_1}, Z_{n_2}) \right| \right] \\
&\quad + E\left[\left| g(\tilde{Z}_1, \tilde{Z}_2] - g_\epsilon(\tilde{Z}_1, \tilde{Z}_2) \right| \right] \\
&\quad + \left| E[g_\epsilon(Z_{n_1}, Z_{n_2})] - E[g_\epsilon(\tilde{Z}_1, \tilde{Z}_2)] \right| \\
&\leq 2\epsilon + \left| E[g_\epsilon(Z_{n_1}, Z_{n_2})] - E[g_\epsilon(Z_1, Z_2)] \right| \\
&= 2\epsilon + \sum_{j=1}^{N_\epsilon} \left| E[f_j(Z_{n_1})h_j(Z_{n_2})] - E[f_j(\tilde{Z}_1)h_j(\tilde{Z}_2)] \right| \\
&= 2\epsilon + \sum_{j=1}^{N_\epsilon} \left| E[f_j(Z_{n_1})]E[h_j(Z_{n_2})] - E[f_j(\tilde{Z}_1)]E[h_j(\tilde{Z}_2)] \right| \\
&\qquad \text{par indépendance de } Z_{n_1} \text{ et } Z_{n_2} \text{ et de } \tilde{Z}_1 \text{ et } \tilde{Z}_2.
\end{aligned}$$

Or, en appliquant la définition de la convergence en loi de Z_{n_1} vers $\tilde{Z}_1$ et Z_{n_2} vers $\tilde{Z}_2$ (notons que $\tilde{Z}_1 \stackrel{\mathcal{L}}{=} Z_1$ et $\tilde{Z}_2 \stackrel{\mathcal{L}}{=} Z_2$) , on a que

$$\begin{aligned} \lim_{n_1 \to \infty} E[f_j(Z_{n_1})] &= E[f_j(\tilde{Z}_1)] \text{ et} \\ \lim_{n_2 \to \infty} E[f_j(Z_{n_2})] &= E[f_j(\tilde{Z}_2)] \end{aligned}$$

et ce pour $j = 1, ..., N_\epsilon$. On en déduit que

$$\left| E[g(Z_{n_1}, Z_{n_2})] - E[g(\tilde{Z}_1, \tilde{Z}_2)] \right| \leq 3\epsilon$$

pour n_1 et n_2 assez grands, c-à-d

$$\lim_{n_1, n_2 \to \infty} E[g(Z_{n_1}, Z_{n_2})] = E[g(\tilde{Z}_1, \tilde{Z}_2)].$$

Comme g a été arbitrairement choisie dans la classe des fonctions continues à support compact, on en déduit que (Z_{n_1}, Z_{n_2}) converge en loi vers un vecteur Gaussien bidimensionnel dont les composantes sont centrées réduites et indépendantes entre elles. $\square$