cours 23, le lundi 3 mai 2010

Rappel. On a démontré à l'aide de la formule d'inversion de Fourier (et de Fubini) la formule suivante $(*)$ qui nous sera utile dans ce cours : si X est une variable aléatoire, si φ_{X} est sa fonction caractéristique et si f est une fonction de classe C^2 sur $\mathbb{R}$, à support compact, on a

$$(*) \qquad \mathrm{E}\, f(\mathrm{X}) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(t) \varphi_{\mathrm{X}}(t)\, \mathrm{d}t.$$

Il en résulte que la fonction caractéristique détermine la loi : pour tous $a < b$, on peut approcher (au sens de la limite simple) l'indicatrice de l'intervalle semi-ouvert $]a, b]$ par une suite (f_n) de fonctions bosse de classe C^2,

$$\forall x \in \mathbb{R}, \quad f_n(x) \underset{n}{\longrightarrow} \mathbf{1}_{]a,b]}(x), \quad |f_n(x)| \le 1\,;$$

par convergence dominée, on voit que

$$\mathrm{F}_{\mathrm{X}}(b) - \mathrm{F}_{\mathrm{X}}(a) = \mathrm{P}\bigl(\{a < \mathrm{X} \le b\}\bigr) = \int_{\Omega} \mathbf{1}_{]a,b]}\bigl(\mathrm{X}(\omega)\bigr)\, \mathrm{dP}(\omega)$$

$$= \lim_n \int_{\Omega} f_n\bigl(\mathrm{X}(\omega)\bigr)\, \mathrm{dP}(\omega) = \lim_n \mathrm{E}\, f_n(\mathrm{X}) = \lim_n \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f_n}(t) \varphi_{\mathrm{X}}(t)\, \mathrm{d}t$$

s'exprime en termes de φ_{X}.

Exemple. On a vu que si G est une variable aléatoire gaussienne, suivant la loi gaussienne centrée réduite

$$\mathrm{d}\gamma(x) = \mathrm{e}^{-x^2/2}\, \frac{\mathrm{d}x}{\sqrt{2\pi}},$$

la fonction caractéristique de G est

$$\varphi_{\mathrm{G}} : t \in \mathbb{R} \longrightarrow \mathrm{e}^{-t^2/2}\,.$$

La détermination de la loi au moyen de la fonction caractéristique permet de trouver facilement la loi de combinaisons linéaires de gaussiennes *indépendantes* : si les $(\mathrm{G}_j)_{j \geqslant 1}$ sont des variables aléatoires indépendantes de loi γ et si les coefficients réels (a_j) vérifient l'égalité $\sum_{j=1}^n a_j^2 = 1$, la variable aléatoire $\mathrm{X} = \sum_{j=1}^n a_j \mathrm{G}_j$ suit aussi la loi γ. En particulier, la loi de

$$\frac{\mathrm{G}_1 + \mathrm{G}_2 + \cdots + \mathrm{G}_n}{\sqrt{n}}$$

reste égale à la loi gaussienne γ, pour tout $n \ge 1$: c'est un cas très particulier du *théorème de la limite centrale* qu'on verra un peu plus loin.

Pour montrer que la loi de cette v.a. X est égale à γ, il suffit de montrer que la fonction caractéristique de X est égale à celle des gaussiennes centrées réduites. On a pour tout t réel, en utilisant l'indépendance des (G_j),

$$\mathrm{E}\, \mathrm{e}^{\mathrm{i}t\mathrm{X}} = \mathrm{E}\, \mathrm{e}^{\mathrm{i} \sum_{j=1}^n t a_j \mathrm{G}_j} = \mathrm{E}\Bigl(\prod_{j=1}^n \mathrm{e}^{\mathrm{i}t a_j \mathrm{G}_j}\Bigr) = \prod_{j=1}^n \Bigl(\mathrm{E}\, \mathrm{e}^{\mathrm{i}t a_j \mathrm{G}_j}\Bigr) =$$

$$= \prod_{j=1}^n \varphi_{\mathrm{G}_j}(t a_j) = \prod_{j=1}^n \mathrm{e}^{-(t a_j)^2/2} = \mathrm{e}^{-t^2 \left(\sum_{j=1}^n a_j^2\right)/2} = \mathrm{e}^{-t^2/2},$$

ce qu'il fallait démontrer.

Lemme 1. *Soit* (Z_n) *une suite de variables aléatoires réelles sur un espace de probabilité* $(\Omega, \mathcal{F}, \mathrm{P})$ *; si la fonction caractéristique de* Z_n *converge simplement sur* $\mathbb{R}$ *vers celle d'une v.a.* Z*, alors pour toute fonction bosse* f *de classe* C^2 *sur* $\mathbb{R}$ *on a*

$$\mathrm{E}\, f(\mathrm{Z}_n) \xrightarrow[n]{} \mathrm{E}\, f(\mathrm{Z}).$$

Preuve. — On utilise le théorème de convergence dominée et la formule $(*)$: pour tout entier n,

$$\mathrm{E}\, f(\mathrm{Z}_n) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{f}(t) \varphi_{\mathrm{Z}_n}(t)\, \mathrm{d}t\,;$$

on a vu que les fonctions caractéristiques sont bornées par 1 en module sur $\mathbb{R}$; on voit donc d'après les hypothèses du lemme que

$$\forall t \in \mathbb{R}, \quad \widehat{f}(t)\varphi_{\mathrm{Z}_n}(t) \xrightarrow[n]{} \widehat{f}(t)\varphi_{\mathrm{Z}}(t),$$

avec la domination

$$|\widehat{f}(t)\varphi_{\mathrm{Z}_n}(t)| \le |\widehat{f}(t)|$$

par la fonction $|\widehat{f}|$ qui est intégrable sur $\mathbb{R}$ (transformée de Fourier d'une fonction de classe C^2 à support compact). Le résultat découle du théorème de convergence dominée.

Théorème de la limite centrale

Avant d'attaquer le théorème, rappelons que la fonction caractéristique φ_{X} d'une v.a. X est de classe C^2 sur $\mathbb{R}$ quand $\mathrm{E}\,\mathrm{X}^2 < +\infty$. On a vu que dans ce cas, on a

$$\varphi'_{\mathrm{X}}(0) = \mathrm{i}\, \mathrm{E}\, \mathrm{X} \quad \text{et} \quad \varphi''_{\mathrm{X}}(0) = -\, \mathrm{E}\, \mathrm{X}^2.$$

Théorème de la limite centrale, version 0. *Soit* (Y_n) *une suite de variables aléatoires indépendantes, centrées, de même loi, et avec un moment d'ordre deux* $\mathrm{E}\,\mathrm{Y}_n^2$ *qui soit tel que* $\mathrm{E}\,\mathrm{Y}_n^2 = 1$ *; la suite des variables aléatoires*

$$s_n = \frac{\mathrm{Y}_1 + \cdots + \mathrm{Y}_n}{\sqrt{n}}$$

converge en loi vers la loi gaussienne centrée réduite, c'est-à-dire que pour tous $a < b$ *on a*

$$\mathrm{P}\big(\{a \le s_n \le b\}\big) \xrightarrow[n]{} \int_a^b \mathrm{e}^{-x^2/2}\, \frac{\mathrm{d}x}{\sqrt{2\pi}}.$$

Preuve. — Les variables aléatoires (Y_n) ayant toutes la même loi, elles ont la même fonction caractéristique qu'on appellera φ_{Y} (on a $\varphi_{\mathrm{Y}} = \varphi_{\mathrm{Y}_1} = \varphi_{\mathrm{Y}_j}$, pour tout $j > 1$). Comme Y_1 est de carré intégrable, la fonction caractéristique φ_{Y} est de classe C^2, donc au voisinage de 0 on peut écrire par Taylor-Young

$$\varphi_{\mathrm{Y}}(t) = \varphi_{\mathrm{Y}}(0) + t\varphi'_{\mathrm{Y}}(0) + \frac{t^2}{2}\, \varphi''_{\mathrm{Y}}(0) + t^2 \varepsilon_1(t),$$

où $\varepsilon_1(t)$ tend vers 0 quand $t \to 0$. On a ici

$$\varphi'_{\mathrm{Y}}(0) = \mathrm{i}\, \mathrm{E}\, \mathrm{Y} = 0, \quad \varphi''_{\mathrm{Y}}(0) = -\, \mathrm{E}\, \mathrm{Y}^2 = -1,$$

donc, en remplaçant, on obtient le développement limité

$$\varphi_{\mathrm{Y}}(t) = 1 - \frac{t^2}{2} + t^2 \varepsilon_1(t).$$

Par l'indépendance des v.a. (Y_j), pour tout $t \in \mathbb{R}$ fixé, on obtient

$$\varphi_{s_n}(t) = \mathrm{E}\, \mathrm{e}^{\mathrm{i} t s_n} = \mathrm{E}\Bigl(\prod_{j=1}^n \mathrm{e}^{\mathrm{i} t \mathrm{Y}_j/\sqrt{n}}\Bigr) = \prod_{j=1}^n \Bigl(\mathrm{E}\, \mathrm{e}^{\mathrm{i} t \mathrm{Y}_j/\sqrt{n}}\Bigr) =$$

$$= \Bigl(\mathrm{E}\, \mathrm{e}^{\mathrm{i}(t/\sqrt{n})\mathrm{Y}_1}\Bigr)^n = \varphi_{\mathrm{Y}}(t/\sqrt{n})^n.$$

On a par ailleurs

$$\ln(1+u) = u + u\varepsilon_2(u)$$

au voisinage de 0, où $\varepsilon_2(u)$ tend vers 0 quand $u \to 0$. On a donc

$$\ln \varphi_{s_n}(t) = n \ln \varphi_{\mathrm{Y}}(t/\sqrt{n}) = n \ln\Bigl(1 - \frac{t^2}{2n} + \frac{t^2}{n} \varepsilon_1(t/\sqrt{n})\Bigr).$$

Posons

$$u_n = -\frac{t^2}{2n} + \frac{t^2}{n} \varepsilon_1(t/\sqrt{n}),$$

qui tend vers 0 avec n, ce qui entraîne que $\varepsilon_2(u_n)$ tend vers 0 ; de plus

$$n u_n = -t^2/2 + t^2 \varepsilon_1(t/\sqrt{n})$$

tend vers $-t^2/2$ quand $n \to +\infty$; on en déduit que

$$\ln \varphi_{s_n}(t) = n \ln(1+u_n) = n u_n + n u_n \varepsilon_2(u_n)$$

tend vers $-t^2/2$. Donc pour tout t réel, $\varphi_{s_n}(t)$ converge vers $\mathrm{e}^{-t^2/2} = \varphi_{\mathrm{G}}(t)$, en désignant par G une v.a. gaussienne centrée réduite.

On va conclure grâce au lemme 1. Donnons nous $a < b$, et $\varepsilon > 0$ tel que $2\varepsilon < b - a$, de sorte que $a + \varepsilon < b - \varepsilon$. Considérons deux bosses f_1, f_2 de classe C^2 telles que

$$\mathbf{1}_{(a+\varepsilon, b-\varepsilon)} \le f_1 \le \mathbf{1}_{(a,b)} \le f_2 \le \mathbf{1}_{(a-\varepsilon, b+\varepsilon)}.$$

D'après le lemme 1, $\mathrm{E}\, f_i(s_n)$ tend vers $\mathrm{E}\, f_i(\mathrm{G})$, $i = 1, 2$. Pour $n \ge n_0$, on aura

$$\Bigl|\mathrm{E}\, f_i(s_n) - \mathrm{E}\, f_i(\mathrm{G})\Bigr| < \varepsilon, \quad i = 1, 2\,;$$

en utilisant le fait que la densité $(2\pi)^{-1/2}\, \mathrm{e}^{-x^2/2}$ de la loi gaussienne γ est < 1, on obtient l'inégalité $\mathrm{P}(x \le \mathrm{G} \le x + \varepsilon) \le \varepsilon$ pour tout réel x, donc pour $n \ge n_0$

$$\mathrm{P}(a \le \mathrm{G} \le b) \le \mathrm{P}(a + \varepsilon \le \mathrm{G} \le b - \varepsilon) + 2\varepsilon = \mathrm{E}\, \mathbf{1}_{(a+\varepsilon, b-\varepsilon)}(\mathrm{G})$$

$$\le \mathrm{E}\, f_1(\mathrm{G}) + 2\varepsilon < \mathrm{E}\, f_1(s_n) + 3\varepsilon \le \mathrm{P}(a \le s_n \le b) + 3\varepsilon$$

et de même pour l'autre côté,

$$\mathrm{P}(a \le \mathrm{G} \le b) \ge \mathrm{P}(a - \varepsilon \le \mathrm{G} \le b + \varepsilon) - 2\varepsilon$$

$$\ge \mathrm{E}\, f_2(\mathrm{G}) - 2\varepsilon > \mathrm{E}\, f_2(s_n) - 3\varepsilon \ge \mathrm{P}(a \le s_n \le b) - 3\varepsilon.$$

Pour $n \ge n_0$, on a donc

$$\bigl|\mathrm{P}(a \le s_n \le b) - \mathrm{P}(a \le \mathrm{G} \le b)\bigr| < 3\varepsilon,$$

ce qui termine la preuve.

Théorème de la limite centrale, version 1. *Soit* $(\mathrm{X}_n)_{n\geqslant 1}$ *une suite de variables aléatoires indépendantes, centrées, de même loi, et avec un moment d'ordre deux* $\mathrm{E}\,\mathrm{X}_n^2$ *fini ; posons* $\sigma^2 = \mathrm{E}(\mathrm{X}_1 - \mathrm{E}\,\mathrm{X}_1)^2 = \mathrm{Var}(\mathrm{X}_1)$ *et supposons* $\mathrm{Var}(\mathrm{X}_1) > 0$. *La suite*

$$s_n = \frac{\mathrm{X}_1 + \cdots + \mathrm{X}_n - n\,\mathrm{E}\,\mathrm{X}_1}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\mathrm{X}_j - \mathrm{E}\,\mathrm{X}_j}{\sigma}$$

converge en loi vers la loi gaussienne centrée réduite.

Preuve. — Comme les X_i sont de carré intégrable par rapport à la mesure finie P sur Ω, elles sont aussi P-intégrables ; comme elles ont la même loi, elles ont la même intégrale $m = \mathrm{E}\,\mathrm{X}_i = \mathrm{E}\,\mathrm{X}_1$. Posons

$$\mathrm{Y}_i = \frac{\mathrm{X}_i - m}{\sigma} \ ;$$

ces variables sont indépendantes comme fonctions $\mathrm{Y}_i = g(\mathrm{X}_i)$ de variables indépendantes, où

$$g(x) = \frac{x - m}{\sigma},$$

et les (Y_i) ont la même loi (elles sont obtenues par la *même* fonction g à partir de v.a. de même loi). On a $\mathrm{E}\,\mathrm{Y}_i = \big(\mathrm{E}\,\mathrm{X}_i - m\big)/\sigma = 0$ et

$$\mathrm{E}\,\mathrm{Y}_i^2 = \frac{1}{\sigma^2}\,\mathrm{E}(\mathrm{X}_i - m)^2 = \frac{1}{\mathrm{E}(\mathrm{X}_1 - \mathrm{E}\,\mathrm{X}_1)^2}\,\mathrm{E}(\mathrm{X}_i - \mathrm{E}\,\mathrm{X}_i)^2 = 1.$$

Le résultat annoncé découle de la version 0, appliquée aux (Y_i).

Convergence des lois, topologie étroite

Pour une probabilité μ sur $(\mathbb{R}, \mathcal{B})$, on définit la fonction de répartition F_μ par

$$\forall x \in \mathbb{R}, \quad \mathrm{F}_\mu(x) = \mu\big(]-\infty, x]\big).$$

On a envie de se servir de la fonction de répartition pour introduire une notion de convergence des lois de probabilité sur $\mathbb{R}$. Par exemple, si (μ_n) est une suite de probabilités sur $\mathbb{R}$ et si les fonctions de répartition F_{μ_n} convergent uniformément sur $\mathbb{R}$ vers la fonction de répartition F_ν d'une probabilité ν, alors certainement on a envie de dire que (μ_n) tend vers ν. Mais se limiter à ce cas serait trop restrictif. En effet, si la suite (x_n) de réels converge vers x, on veut dire que les mesures de Dirac (δ_{x_n}) convergent vers δ_x, par exemple parce qu'on veut sûrement dire que la loi de la v.a. constante $\mathrm{X}_n = x_n$ converge vers la loi de la constante $\mathrm{X} = x$, dans la mesure où les v.a. X_n convergent (uniformément) vers la v.a. $\mathrm{X} = x$. Or, quand $x_n \neq x$, on a

$$\|\mathrm{F}_{\delta_{x_n}} - \mathrm{F}_{\delta_x}\|_\infty = 1.$$

Pour s'en sortir, on va accepter de comparer les fonctions de répartition en des points légèrement différents.

Définition. On dit qu'une suite (μ_n) de probabilités sur $(\mathbb{R}, \mathcal{B})$ *converge étroitement* vers une probabilité ν si pour tout $\varepsilon > 0$, il existe n_0 tel que pour tout $n \geq n_0$ on ait

$$\forall x \in \mathbb{R}, \quad \mathrm{F}_\nu(x - \varepsilon) - \varepsilon \leq \mathrm{F}_{\mu_n}(x) \leq \mathrm{F}_\nu(x + \varepsilon) + \varepsilon.$$

On peut définir une distance sur l'ensemble des probabilités sur $(\mathbb{R}, \mathcal{B})$ qui corresponde à cette notion de convergence ; l'ensemble des $\varepsilon \geq 0$ tels que

$$\forall x \in \mathbb{R}, \quad \mathrm{F}_\mu(x) \leq \mathrm{F}_\nu(x+\varepsilon)+\varepsilon$$

est fermé parce que F_ν est continue à droite, on peut donc définir le plus petit nombre $\varepsilon(\mu, \nu) \geq 0$ tel que

$$\forall x \in \mathbb{R}, \quad \mathrm{F}_\mu(x) \leq \mathrm{F}_\nu\big(x+\varepsilon(\mu, \nu)\big)+\varepsilon(\mu, \nu)$$

et poser

$$d(\mu, \nu)=\min \big(\varepsilon(\mu, \nu), \varepsilon(\nu, \mu)\big).$$

C'est une distance, et la convergence des suites de probabilités sur $(\mathbb{R}, \mathcal{B})$ pour cette distance est la même convergence que celle qui est donnée dans la définition précédente.

Exemple. On peut voir que

$$d(\delta_x, \delta_y)=\min (1,|x-y|).$$

Attention ! Si on considère la suite des mesures de Dirac (δ_n) aux points $n \in \mathbb{N}$, la suite des fonctions de répartition F_{δ_n} converge simplement sur $\mathbb{R}$ vers la fonction nulle, mais la fonction nulle n'est pas une fonction de répartition de probabilité. Il n'y a pas convergence étroite dans ce cas. De même on n'a pas que

$$\mu_n=\frac{1}{2}\big(\delta_0+\delta_n\big)$$

converge étroitement vers $\frac{1}{2} \delta_0$, qui n'est pas une probabilité, bien que la suite des fonctions de répartition F_{μ_n} converge simplement vers $\frac{1}{2} \mathrm{F}_{\delta_0}$.

En fait, si les (μ_n) sont des probabilités et si on a, sans dire d'avance que ν est une probabilité, que pour tout $\varepsilon>0$, on a pour n assez grand

$$\forall x \in \mathbb{R}, \quad \mathrm{F}_\nu(x-\varepsilon)-\varepsilon \leq \mathrm{F}_{\mu_n}(x) \leq \mathrm{F}_\nu(x+\varepsilon)+\varepsilon,$$

il en résulte que

$$\limsup_{x \rightarrow+\infty}|1-\mathrm{F}_\nu(x)| \leq \varepsilon,$$

pour tout $\varepsilon>0$, donc F_ν tend vers 1 en $+\infty$ et ν est une probabilité sur $(\mathbb{R}, \mathcal{B})$.

Remarque. *Si (μ_n) tend étroitement vers ν et si x_0 est un réel tel que $\nu(\{x_0\})=0$, on a*

$$\mathrm{F}_{\mu_n}(x_0) \longrightarrow \mathrm{F}_\nu(x_0).$$

Preuve. — Sous cette hypothèse sur x_0, la fonction F_ν est continue au point x_0 ; étant donné $\varepsilon>0$, il existe α tel que $0<\alpha<\varepsilon$ et tel que

$$|\mathrm{F}_\nu(y)-\mathrm{F}_\nu(x_0)|<\varepsilon$$

dès que $|y-x_0| \leq \alpha$. Pour $n \geq n_0$ convenable, n_0 dépendant de $\alpha>0$, on aura

$$\mathrm{F}_{\mu_n}(x_0) \leq \mathrm{F}_\nu(x_0+\alpha)+\alpha \leq \mathrm{F}_\nu(x_0)+\varepsilon+\alpha \leq \mathrm{F}_\nu(x_0)+2 \varepsilon$$

et de même

$$\mathrm{F}_{\mu_n}(x_0) \geq \mathrm{F}_\nu(x_0-\alpha)-\alpha \geq \mathrm{F}_\nu(x_0)-\varepsilon-\alpha \geq \mathrm{F}_\nu(x_0)-2\varepsilon,$$

ce qui prouve l'affirmation.

Théorème. *Si ν et les (μ_n) sont des probabilités sur $(\mathbb{R},\mathcal{B})$, les conditions suivantes sont équivalentes :*

1. — *la suite (μ_n) converge étroitement vers ν ;*

2. — *pour toute fonction continue* bornée *f sur $\mathbb{R}$, réelle ou complexe,*

$$\int_{\mathbb{R}} f\,\mathrm{d}\mu_n \xrightarrow[n]{} \int_{\mathbb{R}} f\,\mathrm{d}\nu\,;$$

3. — *il y a convergence simple sur $\mathbb{R}$ des transformées de Fourier :*

$$\forall t\in\mathbb{R},\quad \widehat{\mu}_n(t)\xrightarrow[n]{}\widehat{\nu}(t).$$

Esquisse de preuve. — Traitons d'abord $\mathbf{1} \Rightarrow \mathbf{2}$; supposons que μ_n tende vers ν étroitement ; il existe un ensemble D au plus dénombrable de points y tels que $\nu(\{y\})>0$. Alors $\mathbb{R}\setminus\mathrm{D}$ est dense dans $\mathbb{R}$ (Cantor). Soit f continue et bornée, disons par 1 pour fixer les idées ; on peut trouver $a<b$ tels que $a,b\notin\mathrm{D}$ et que l'intervalle $]a,b]$ soit suffisamment grand pour que

$$\mathrm{F}_\nu(b)-\mathrm{F}_\nu(a)=\nu(]a,b])>1-\varepsilon.$$

D'après la remarque, pour n assez grand on a $\mathrm{F}_{\mu_n}(b)-\mathrm{F}_{\mu_n}(a)=\mu_n(]a,b])>1-\varepsilon$. On a donc pour ces valeurs de n

$$\mu_n\bigl(\mathbb{R}\backslash]a,b]\bigr)<\varepsilon,\quad \nu\bigl(\mathbb{R}\backslash]a,b]\bigr)<\varepsilon.$$

On a donc aussi pour ces valeurs de n

$$\Bigl|\int_{\mathbb{R}} \mathbf{1}_{\mathbb{R}\backslash]a,b]} f\,\mathrm{d}\mu_n\Bigr| \leq \int_{\mathbb{R}} \mathbf{1}_{\mathbb{R}\backslash]a,b]}|f|\,\mathrm{d}\mu_n \leq \mu_n\bigl(\mathbb{R}\backslash]a,b]\bigr)<\varepsilon,$$

et de même

$$\Bigl|\int_{\mathbb{R}} \mathbf{1}_{\mathbb{R}\backslash]a,b]} f\,\mathrm{d}\nu\Bigr| \leq \nu\bigl(\mathbb{R}\backslash]a,b]\bigr)<\varepsilon.$$

La fonction f est uniformément continue sur le compact $[a,b]$, il existe donc $\delta>0$ tel que $|f(x)-f(y)|<\varepsilon$ dès que $a\leq x,y\leq b$ et $|x-y|\leq\delta$; on peut découper $[a,b]$ en $a=a_0<a_1<\ldots<a_{\mathrm{N}}=b$ au moyen de points (a_j) qui ne sont pas dans D, et avec un pas de subdivision vérifiant $\max_{0<j\leqslant\mathrm{N}}(a_j-a_{j-1})<\delta$; alors la fonction étagée

$$\varphi=\sum_{j=1}^{\mathrm{N}} f(a_j)\mathbf{1}_{]a_{j-1},a_j]}$$

est à distance uniforme $\leq\varepsilon$ de $\mathbf{1}_{]a,b]}f$, donc

$$\Bigl|\int_{\mathbb{R}} \mathbf{1}_{]a,b]} f\,\mathrm{d}\mu_n-\int_{\mathbb{R}}\varphi\,\mathrm{d}\mu_n\Bigr|\leq\varepsilon \quad\text{et}\quad \Bigl|\int_{\mathbb{R}} \mathbf{1}_{]a,b]} f\,\mathrm{d}\nu-\int_{\mathbb{R}}\varphi\,\mathrm{d}\nu\Bigr|\leq\varepsilon,$$

alors que

$$\Big|\int_{\mathbb{R}} \mathbf{1}_{\mathbb{R}\setminus]a,b]} f \,\mathrm{d}\mu_n\Big| < \varepsilon, \quad \Big|\int_{\mathbb{R}} \mathbf{1}_{\mathbb{R}\setminus]a,b]} f \,\mathrm{d}\nu\Big| < \varepsilon.$$

on a donc

$$\Big|\int_{\mathbb{R}} f \,\mathrm{d}\mu_n - \int_{\mathbb{R}} f \,\mathrm{d}\nu\Big| \leq \Big|\int_{\mathbb{R}} \varphi \,\mathrm{d}\mu_n - \int_{\mathbb{R}} \varphi \,\mathrm{d}\nu\Big| + 4\varepsilon.$$

Par ailleurs

$$\int_{\mathbb{R}} \varphi \,\mathrm{d}\mu_n = \sum_{j=1}^{\mathrm{N}} f(a_j)\mu_n(]a_{j-1}, a_j]) = \sum_{j=1}^{\mathrm{N}} f(a_j)\big(\mathrm{F}_{\mu_n}(a_j) - \mathrm{F}_{\mu_n}(a_{j-1})\big)$$

converge d'après la remarque vers

$$\sum_{j=1}^{\mathrm{N}} f(a_j)\big(\mathrm{F}_{\nu}(a_j) - \mathrm{F}_{\nu}(a_{j-1})\big) = \int_{\mathbb{R}} \varphi \,\mathrm{d}\nu.$$

On en déduit que **1** $\Rightarrow$ **2**.

Comme $x \to \mathrm{e}^{-\mathrm{i}tx}$ est continue bornée sur $\mathbb{R}$, on déduit de **2** la convergence simple des transformées de Fourier. Pour terminer avec **3** $\Rightarrow$ **1**, on utilise des bosses de classe C^2, comme on a fait pour la preuve du TLC.