



Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées
Bât. M2, F-59655 Villeneuve d'Ascq Cedex

Théorème limite central

Charles SUQUET

Théorème limite central

Les X_k étant des variables aléatoires réelles définies sur le même espace probabilisé et indépendantes, on note

$$S_n := \sum_{k=1}^n X_k,$$

et on étudie la convergence *en loi* de

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\text{Var } S_n}},$$

lorsque cette quantité est définie. On envisagera ensuite le cas où les X_k sont des vecteurs aléatoires de \mathbb{R}^d .

1 Cas d'une suite i.i.d.

1.1 De Moivre-Laplace

Théorème 1. *Si les X_k sont indépendantes et de même loi de Bernoulli de paramètre $p \in]0, 1[$, avec $q := 1 - p$, on a*

$$S_n^* := \frac{S_n - np}{\sqrt{npq}} = \sqrt{\frac{n}{pq}} \left(\frac{S_n}{n} - p \right) \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, 1).$$

Comme la fonction de répartition Φ de $\mathfrak{N}(0, 1)$ est continue sur \mathbb{R} , ceci équivaut à

$$\forall x \in \mathbb{R}, \quad \mathbf{P}(S_n^* \leq x) \xrightarrow[n \rightarrow +\infty]{} \Phi(x) = \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) \frac{dt}{\sqrt{2\pi}}.$$

Preuve. La démonstration historique de ce théorème repose sur un bon contrôle des coefficients binomiaux via la formule de Stirling. On pourra la consulter en annexe (cf. B.4). L'intérêt de cette approche « élémentaire » est de donner une idée de la vitesse de convergence. La preuve moderne par les fonctions caractéristiques est quasi immédiate, car tout le travail a été fait dans la preuve du théorème de Lévy sur l'équivalence entre la convergence en loi et la convergence des fonctions caractéristiques. \square

Une application directe de ce théorème est la construction d'*intervalles de confiance* pour l'estimation d'une probabilité inconnue p à partir de l'observation d'un *échantillon* de n variables de Bernoulli indépendantes de paramètre p . Considérons pour $t > 0$ l'événement

$$A_{n,t} := \left\{ \omega \in \Omega; -t \leq \sqrt{\frac{n}{pq}} \left(\frac{S_n(\omega)}{n} - p \right) \leq t \right\}.$$

Le théorème de de Moivre-Laplace nous dit que pour n assez grand, on peut utiliser l'approximation :

$$\mathbf{P}(A_{n,t}) \simeq \Phi(t) - \Phi(-t) = 2\Phi(t) - 1.$$

Ceci peut se réécrire

$$\mathbf{P}\left(\frac{S_n}{n} - t\sqrt{\frac{pq}{n}} \leq p \leq \frac{S_n}{n} + t\sqrt{\frac{pq}{n}}\right) = 2\Phi(t) - 1 + \varepsilon_n.$$

On ignore la valeur de p , donc a fortiori celle de \sqrt{pq} . Heureusement, il est possible de la majorer car $p(1-p)$ est maximal pour $p = 1/2$. D'où

$$\sqrt{pq} \leq \sqrt{\frac{1}{4}} = \frac{1}{2}, \quad (1)$$

de sorte qu'en notant

$$B_{n,t} := \left\{ \omega \in \Omega; \frac{S_n(\omega)}{n} - \frac{t}{2\sqrt{n}} \leq p \leq \frac{S_n(\omega)}{n} + \frac{t}{2\sqrt{n}} \right\},$$

l'inclusion $A_{n,t} \subset B_{n,t}$ nous donne :

$$\mathbf{P}(B_{n,t}) \geq 2\Phi(t) - 1 + \varepsilon_n. \quad (2)$$

En pratique, n est fixé et on a observé des valeurs numériques explicites x_1, \dots, x_n que l'on interprète comme les valeurs de $X_1(\omega), \dots, X_n(\omega)$ pour *un même* ω tiré au sort (suivant \mathbf{P}). On est donc en présence d'une valeur numérique explicite, $S_n(\omega)/n = (x_1 + \dots + x_n)/n$, disons pour fixer les idées $S_n(\omega)/n = 0,53$. Proposer pour le paramètre inconnu p l'intervalle de confiance

$$I_{n,t} = \left[0,53 - \frac{t}{2\sqrt{n}}; 0,53 + \frac{t}{2\sqrt{n}} \right],$$

c'est faire le *pari* que le ω observé est bien dans $B_{n,t}$. La probabilité de gagner ce pari est minorée par $2\Phi(t) - 1 + \varepsilon_n$. On dit que $I_{n,t}$ est un intervalle de confiance pour p avec un *niveau*¹ d'au moins $2\Phi(t) - 1 + \varepsilon_n$. En pratique, on laisse tomber le ε_n et on

1. Il y a ici un piège sémantique : supposons qu'on ait trouvé $I_{n,t} = [0,51; 0,55]$ avec un niveau de confiance de 95%. Il est tout à fait incorrect d'écrire « $\mathbf{P}(p \in [0,51; 0,55]) \geq 0,95$ ». En effet, p n'a rien d'aléatoire, c'est une constante. L'aléatoire concerne notre ignorance sur sa valeur. Même si on considère p comme une variable aléatoire constante, la probabilité de son appartenance à $[0,51; 0,55]$ vaut 0 ou 1, et comme on ne peut pas exclure le premier cas, on ne peut pas minorer cette probabilité par 0,95.

détermine t de façon approchée grâce à la tabulation de Φ . Par exemple pour un niveau de confiance de 95%, on est ramené à la résolution de l'équation $\Phi(t) = 1,95/2 = 0,975$ d'où $t \simeq 1,96$, ce qui nous donne l'intervalle

$$I_n = \left[\frac{S_n(\omega)}{n} - \frac{1,96}{2\sqrt{n}}; \frac{S_n(\omega)}{n} + \frac{1,96}{2\sqrt{n}} \right], \quad \text{au niveau de confiance 95\%}.$$

En fait les statisticiens préfèrent une variante de cette méthode pour obtenir des intervalles de confiance plus étroits, notamment quand p n'est pas trop proche de $1/2$. L'idée est de remplacer la variance inconnue pq de X_1 par un *estimateur* au lieu de la majorer de façon certaine par (1). Ainsi en estimant pq par $M_n(1 - M_n)$ où $M_n := S_n/n$, on obtient au niveau de confiance 95% l'intervalle

$$J_n = \left[M_n(\omega) - 1,96\sqrt{\frac{M_n(\omega)(1 - M_n(\omega))}{n}}; M_n(\omega) + 1,96\sqrt{\frac{M_n(\omega)(1 - M_n(\omega))}{n}} \right].$$

La construction des intervalles de confiance pose naturellement la question de la *vitesse de convergence* dans le théorème de de Moivre-Laplace. En effet on aimerait pouvoir contrôler l'erreur commise ci-dessus en laissant tomber le ε_n . Pour mesurer cette vitesse de convergence, introduisons

$$\Delta_n := \sup_{x \in \mathbb{R}} |\mathbf{P}(S_n^* \leq x) - \Phi(x)|.$$

Il est de bon ton de savoir que la vitesse de convergence vers 0 de Δ_n est en $O(n^{-1/2})$ et que la constante du O explose quand p tend vers 0 ou vers 1. Des résultats plus précis sont présentés dans l'annexe B, section B.5 (voir aussi la preuve du théorème de de Moivre-Laplace pour justifier le $O(n^{-1/2})$).

Remarque 2. Lorsque np est « petit », l'approximation gaussienne de la binomiale contenue dans le théorème de de Moivre-Laplace est avantageusement remplacée par l'approximation poissonienne. Plus précisément, si $p = p(n)$ et si pour une certaine constante $\lambda > 0$, $np(n) \rightarrow \lambda$, S_n converge en loi vers $\text{Pois}(\lambda)$. Cette convergence équivaut² à :

$$\forall k \in \mathbb{N}, \quad \mathbf{P}(S_n = k) \xrightarrow{n \rightarrow +\infty} \frac{e^{-\lambda} \lambda^k}{k!}.$$

1.2 Le théorème limite central dans le cas i.i.d.

Théorème 3. Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes, de même loi et de carré intégrable (et non constantes). Notons $\mu = \mathbf{E}X_1$, $\sigma^2 := \text{Var} X_1$ avec $\sigma > 0$. Alors

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\text{Var} S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, 1).$$

2. Pour justifier cette équivalence, il suffit d'utiliser la caractérisation de la convergence en loi par la convergence des f.d.r. en tout point de continuité de la loi limite.

Corollaire 4. Si Y_n suit une loi de Poisson de paramètre $\lambda = n\alpha$ (α constante),

$$Y_n^* := \frac{Y_n - \lambda}{\sqrt{\lambda}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, 1).$$

Commentaire : Ceci explique que l'approximation poissonnienne d'une binomiale finisse par devenir gaussienne quand λ augmente. En pratique, on préfère l'approximation gaussienne dès que $\lambda \geq 20$. On peut en proposer une illustration graphique (cf. T.P.).

Preuve. Il suffit de remarquer que Y_n a même loi que $X_1 + \dots + X_n$ où les X_i sont indépendantes et de même loi $\text{Pois}(\alpha)$. Par le théorème 3, le S_n^* bâti sur la suite des X_i converge en loi vers $\mathfrak{N}(0, 1)$. Comme $\mathbf{E}Y_n = \lambda$ et $\text{Var} Y_n = \lambda$, on voit que Y_n^* a même loi que S_n^* . Il en résulte que Y_n^* converge en loi vers la même limite que S_n^* . \square

Avant d'examiner d'autres applications du théorème limite central, il est opportun de rappeler le résultat suivant.

Lemme 5 (Slutsky). Soient (X_n) et (Y_n) deux suites de variables aléatoires réelles définies sur le même espace probabilisé et convergentes en loi respectivement vers la variable aléatoire X et la constante C . Alors la suite de vecteurs aléatoires (X_n, Y_n) converge en loi vers (X, C) . Voici trois utiles conséquences :

- a) Pour toute fonction continue $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, $g(X_n, Y_n)$ converge en loi vers $g(X, C)$.
- b) Si X_n converge en loi vers X et Y_n converge en probabilité vers 0, alors $X_n + Y_n$ converge en loi vers X .
- c) Si X_n converge en loi vers X et Y_n converge en probabilité vers C , alors $X_n Y_n$ converge en loi vers CX .

Commentaire : On pourra trouver une démonstration sous forme d'exercice corrigé dans [6, pp. 347–350]. Comme la convergence en loi vers une constante C équivaut à la convergence en probabilité vers C , les propriétés b) et c) ne sont que des cas particuliers du a).

Corollaire 6 (Convergence de la statistique t_n de Student). Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. définies sur le même espace et ayant un moment d'ordre 2. On note $\mathbf{E}X_1 = \mu$ et

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i.$$

Alors

$$T_n := \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, 1). \quad (3)$$

Commentaire : Lorsqu'on veut construire des intervalles de confiance pour l'estimation d'une espérance inconnue μ par la moyenne arithmétique \bar{X}_n de l'échantillon, la première idée est d'utiliser le théorème 3 en remarquant que

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\text{Var} S_n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\text{Var} X_1}}.$$

L'ennui c'est qu'en général $\text{Var } X_1$ est inconnue. On la remplace alors par un estimateur sans biais

$$W_n := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

qui n'est autre que le « σ_{n-1}^2 » des calculatrices. Le corollaire 6 fournit la légitimation théorique de cette recette.

Preuve. Avant de démontrer (3), vérifions que W_n est un estimateur fortement consistant et sans biais.

En interprétant

$$V_n(\omega) := \frac{1}{n} \sum_{i=1}^n (X_i(\omega) - \bar{X}_n(\omega))^2$$

comme la variance de la mesure $\nu_n(\omega) := n^{-1} \sum_{i=1}^n \delta_{X_i(\omega)}$ (mesure empirique, dont l'espérance est $\int_{\mathbb{R}} x d\nu_n(\omega)(x) = \bar{X}_n(\omega)$), la formule de Koenig nous donne

$$V_n(\omega) = \frac{1}{n} \sum_{i=1}^n X_i^2(\omega) - (\bar{X}_n(\omega))^2. \quad (4)$$

Par une double application de la loi forte des grands nombres, on voit que

$$V_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{E}X_1^2 - (\mathbf{E}X_1)^2 = \text{Var } X_1,$$

en utilisant à nouveau la formule de Koenig, mais dans l'autre sens et pour la loi de X_1 au lieu de $\nu_n(\omega)$. Comme $W_n = (n/(n-1))V_n$, on en déduit

$$W_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \text{Var } X_1. \quad (5)$$

Vérifions maintenant que W_n est sans biais.

$$\begin{aligned} \mathbf{E} \left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) &= n \mathbf{E} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2 \right) \quad (\text{grâce à (4)}) \\ &= \mathbf{E} \left(\sum_{i=1}^n X_i^2 \right) - n \mathbf{E} \left(\frac{1}{n^2} \left(\sum_{i=1}^n X_i \right)^2 \right) \\ &= n \mathbf{E}X_1^2 - \frac{1}{n} \mathbf{E} \left(\sum_{i=1}^n X_i^2 + \sum_{1 \leq i \neq j \leq n} X_i X_j \right) \\ &= (n-1) \mathbf{E}X_1^2 - \frac{1}{n} \sum_{1 \leq i \neq j \leq n} \mathbf{E}(X_i X_j) \\ &= (n-1) \mathbf{E}X_1^2 - \frac{n^2 - n}{n} (\mathbf{E}X_1)^2 \quad (\text{car } \mathbf{E}(X_1 X_2) = (\mathbf{E}X_1)^2) \\ &= (n-1) (\mathbf{E}X_1^2 - (\mathbf{E}X_1)^2) \\ &= (n-1) \text{Var } X_1. \end{aligned}$$

On a donc bien $\mathbf{E}W_n = \text{Var } X_1$.

Pour établir la convergence en loi de T_n , on écrit

$$T_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sigma W_n^{-1/2},$$

où $\sigma = \sqrt{\text{Var } X_1}$. Par (5), $\sigma W_n^{-1/2}$ converge p.s. et donc *a fortiori* en loi vers 1. Le théorème 3 et le lemme de Slutsky c) nous donnent la convergence de T_n vers $\mathfrak{N}(0, 1)$. \square

Corollaire 7 (La δ -méthode). *Supposons que*

a) $Z_n := \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z$, où Z suit la loi $\mathfrak{N}(0, 1)$;

b) f est une fonction définie sur un voisinage de μ , dérivable au point μ et $f'(\mu) \neq 0$.

Alors

$$\frac{\sqrt{n}}{\sigma|f'(\mu)|}(f(\bar{X}_n) - f(\mu)) \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, 1). \quad (6)$$

Ce résultat est présenté sous forme d'exercice dans Billingsley [2, Ex. 27.10]. Il est utile pour construire des intervalles de confiance pour des fonctions non linéaires de l'espérance inconnue μ .

Preuve. Les variables X_i , \bar{X}_n et Z sont définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$. Par le théorème de Skorokhod, il existe un espace probabilisé $(\Omega', \mathcal{A}', \mathbf{P}')$ et des variables aléatoires Z'_n, Z' définies sur cet espace, telles que pour chaque $n \geq 1$, Z_n et Z'_n ont même loi, Z et Z' ont même loi et

$$Z'_n \xrightarrow[n \rightarrow +\infty]{\mathbf{P}'\text{-p.s.}} Z'. \quad (7)$$

Pour bien comprendre la preuve en cours, une petite digression sur le théorème de Skorokhod n'est pas superflue. Barbe Ledoux [1, p. 129, (iv)] ou Foata Fuchs [5, p. 217] sont des références accessibles pour sa démonstration. En voici une esquisse. On prend $\Omega' = [0, 1]$, \mathcal{A}' sa tribu borélienne et \mathbf{P}' la mesure de Lebesgue. On choisit *une seule* variable aléatoire U définie sur cet espace et de loi uniforme sur $[0, 1]$, par exemple la fonction identité. On pose ensuite $Z'_n = F_n^{-1}(U)$ et $Z' = F^{-1}(U)$ où F_n^{-1} et F^{-1} sont les inverses généralisés respectifs des fonctions de répartition de Z_n et Z . Par hypothèse Z_n converge en loi vers Z et donc $F_n(x)$ converge vers $F(x)$ en tout point x de continuité de F . Les discontinuités de F forment un ensemble au plus dénombrable. La partie technique (et purement analytique) est d'en déduire que $F_n^{-1}(u)$ converge vers $F^{-1}(u)$ pour \mathbf{P}' -presque tout réel $u \in]0, 1[$. L'esquisse ci-dessus permet de répondre à la question que le jury ne manquera pas de poser à tout candidat invoquant le théorème de Skorokhod : est-ce que (Z'_n, Z'_{n+1}) a même loi que (Z_n, Z_{n+1}) ? Il est clair que non puisque Z'_{n+1} est une fonction déterministe de Z'_n , ce qui n'est pas le cas pour Z_n et Z_{n+1} .

Revenons à la preuve du corollaire et posons

$$Y'_n := \sigma n^{-1/2} Z'_n + \mu.$$

Comme $\bar{X}_n = \sigma n^{-1/2} Z_n + \mu$, il est clair que Y'_n et \bar{X}_n ont même loi. En réécrivant (7) sous la forme $\sqrt{n}(Y'_n(\omega) - \mu) \rightarrow \sigma Z'(\omega)$, on voit que $Y'_n(\omega)$ converge³ vers μ pour \mathbf{P}' -presque tout $\omega \in \Omega'$.

L'hypothèse b) sur f nous donne

$$\exists \alpha > 0, \quad \forall x \in]\mu - \alpha, \mu + \alpha[, \quad f(x) - f(\mu) = f'(\mu)(x - \mu)(1 + \varepsilon(x)), \quad \lim_{x \rightarrow \mu} \varepsilon(x) = 0.$$

Comme Y'_n converge \mathbf{P}' -p.s. vers μ , on a pour \mathbf{P}' -presque tout $\omega \in \Omega'$, un entier $n_0(\omega, \alpha)$ tel que

$$\forall n \geq n_0(\omega, \alpha), \quad \frac{f(Y'_n(\omega)) - f(\mu)}{f'(\mu)} = (Y'_n(\omega) - \mu)(1 + \varepsilon(Y'_n(\omega))).$$

On en déduit que pour \mathbf{P}' -presque tout $\omega \in \Omega'$,

$$\forall n \geq n_0(\omega, \alpha), \quad \frac{\sqrt{n}(f(Y'_n(\omega)) - f(\mu))}{\sigma f'(\mu)} = Z'_n(\omega)(1 + \varepsilon(Y'_n(\omega))).$$

Par conséquent,

$$\frac{\sqrt{n}(f(Y'_n) - f(\mu))}{\sigma f'(\mu)} \xrightarrow[n \rightarrow +\infty]{\mathbf{P}'\text{-p.s.}} Z'.$$

Cette convergence presque sûre entraîne la convergence en loi. Pour conclure, on remarque que $\sqrt{n}(f(Y'_n) - f(\mu))/(\sigma f'(\mu))$ et $\sqrt{n}(f(\bar{X}_n) - f(\mu))/(\sigma f'(\mu))$ ont même loi. Comme la convergence en loi ne dépend que des lois et pas des variables, on aboutit à

$$\frac{\sqrt{n}(f(\bar{X}_n) - f(\mu))}{\sigma f'(\mu)} \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, 1),$$

ce qui est la conclusion recherchée, en notant que l'absence de valeur absolue pour la constante $f'(\mu)$ au dénominateur est sans importance puisque la loi limite $\mathfrak{N}(0, 1)$ est symétrique. \square

Application à l'estimation du paramètre d'une loi exponentielle.

Soit X_1 de loi exponentielle de paramètre $a > 0$ inconnu. La densité de X_1 est $g(t) = ae^{-at}\mathbf{1}_{\mathbb{R}^+}(t)$. On a immédiatement $\mathbf{E}X_1 = 1/a$, de sorte que par la loi forte des grands nombres \bar{X}_n converge p.s. vers $1/a$ et que cette convergence implique la convergence p.s. de $1/\bar{X}_n$ vers a (noter que $\mathbf{P}(\bar{X}_n = 0) = 0$). L'estimateur $1/\bar{X}_n$ de a est donc fortement consistant⁴. Pour construire un intervalle de confiance correspondant, on applique la δ -méthode avec $f(x) = 1/x$ d'où $|f'(1/a)| = a^2$. Comme

$$\mathbf{E}X_1^2 = \frac{2}{a^2}, \quad \text{Var } X_1 = \frac{1}{a^2}, \quad \sigma = \frac{1}{a},$$

3. Attention au piège : cette convergence p.s. ne peut s'obtenir par la loi forte des grands nombres comme pour \bar{X}_n . En effet bien que Y'_n et \bar{X}_n aient même loi, le théorème de Skorokhod ne nous permet pas d'écrire Y'_n comme la moyenne arithmétique de variables aléatoires i.i.d. définies sur Ω' . Il est même clair que pour la construction de Z'_n présentée ci-dessus, cette représentation est impossible puisque Y'_{n+1} est une fonction déterministe de Y'_n !

4. Il est clair qu'il est biaisé puisque $\mathbf{E}(1/\bar{X}_n) \neq 1/\mathbf{E}\bar{X}_n = a$.

le corollaire 7 nous donne

$$\frac{\sqrt{n}}{a} \left(\frac{1}{\overline{X}_n} - a \right) \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, 1).$$

Si l'on connaît un majorant c de a , on peut alors en déduire des intervalles de confiance au niveau α pour a de la forme

$$I = \left[\frac{1}{\overline{X}_n(\omega)} - \frac{ct_\alpha}{\sqrt{n}}; \frac{1}{\overline{X}_n(\omega)} + \frac{ct_\alpha}{\sqrt{n}} \right],$$

où t_α est la solution de l'équation $2\Phi(t) - 1 = \alpha$ (exemple : $t_{0,95} = 1,96$). Si l'on ne souhaite pas imposer *a priori* un majorant de a , il suffit de combiner le corollaire 7 avec le lemme de Slutsky c) pour obtenir

$$\overline{X}_n \sqrt{n} \left(\frac{1}{\overline{X}_n} - a \right) \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, 1).$$

Les intervalles de confiance correspondants seront de la forme

$$J = \left[\frac{1}{\overline{X}_n(\omega)} - \frac{t_\alpha}{\overline{X}_n(\omega)\sqrt{n}}; \frac{1}{\overline{X}_n(\omega)} + \frac{t_\alpha}{\overline{X}_n(\omega)\sqrt{n}} \right].$$

1.3 Réciproque du TLC i.i.d.

Théorème 8. *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. définies sur le même espace probabilisé et telle que S_n/\sqrt{n} converge en loi vers $\mathfrak{N}(0, 1)$. Alors X_1 est de carré intégrable, $\mathbf{E}X_1^2 = 1$ et $\mathbf{E}X_1 = 0$.*

Commentaires : Ce théorème contient essentiellement la réciproque du théorème 3. Plus précisément on en déduit que s'il existe $\mu \in \mathbb{R}$ et $\sigma > 0$ constantes⁵ telles que $(\sigma n)^{-1/2}(S_n - n\mu)$ converge en loi vers $\mathfrak{N}(0, 1)$, alors $X_1 \in L^2(\Omega)$, $\mathbf{E}X_1 = \mu$ et $\text{Var } X_1 = \sigma^2$. En effet, il suffit d'appliquer le théorème 8 aux variables aléatoires $X'_i = (X_i - \mu)/\sigma$. Ainsi les théorèmes 3 et 8 nous donnent l'équivalence entre la convergence en loi vers une gaussienne de $n^{-1/2}(S_n - n\mu)$ (pour une certaine constante μ) et l'existence d'un moment d'ordre 2 pour X_1 . La démonstration du théorème 8 se trouve dans Foata Fuchs [5, p. 249] avec l'hypothèse supplémentaire $X_1 \in L^1(\Omega)$ et $\mathbf{E}X_1 = 0$. En réalité ces hypothèses sont superflues, comme nous le montre la preuve ci-dessous qui à part cela, suit [5].

Preuve. La démonstration comporte 5 étapes dont deux lemmes. On commence par réduire le problème à la preuve de l'appartenance de X_1 à $L^2(\Omega)$. Le lemme 9 établit ensuite un lien entre le comportement de la fonction caractéristique de X_1 au voisinage de 0 et $\mathbf{E}X_1^2$. On peut alors achever la preuve dans le cas particulier où X_1 est de loi symétrique. Le lemme 10 est utile pour le passage du cas symétrique au cas général.

5. Bien sûr sans supposer *a priori* que ces constantes ont un lien avec d'éventuels moments de X_1 !

Réduction à la preuve de $\mathbf{E}X_1^2 < +\infty$.

Supposons établie l'appartenance de X_1 à $L^2(\Omega)$. On a alors $\mathbf{E}|X_1| < +\infty$ donc par la loi forte des grands nombres, $n^{-1}S_n$ converge p.s. vers $\mathbf{E}X_1$, donc aussi en loi vers la même limite. D'autre part en écrivant $n^{-1}S_n = n^{-1/2}(n^{-1/2}S_n)$, l'hypothèse de convergence de $n^{-1/2}S_n$ vers Z de loi $\mathfrak{N}(0, 1)$ et le lemme de Slutsky (sous une forme dégénérée!) nous donnent la convergence en loi de $n^{-1}S_n$ vers $0 \times Z = 0$. Par unicité de la loi limite, les deux variables aléatoires constantes $\mathbf{E}X_1$ et 0 doivent avoir même loi. Ceci implique $\mathbf{E}X_1 = 0$ (pour des masses de Dirac, $\delta_a = \delta_b$ équivaut à $a = b$).

Ensuite par le T.L.C. (Th. 3), $n^{-1/2}S_n$ converge en loi vers $\mathfrak{N}(0, \mathbf{E}X_1^2)$. Par unicité de la loi limite, on doit donc avoir $\mathfrak{N}(0, 1) = \mathfrak{N}(0, \mathbf{E}X_1^2)$, d'où $\mathbf{E}X_1^2 = 1$.

Lemme 9. Soit X une variable aléatoire réelle de fonction caractéristique φ . Alors

$$\lim_{u \rightarrow 0} \frac{2(1 - \operatorname{Re} \varphi(u))}{u^2} = \int_{\mathbb{R}} x^2 dP_X(x) = \mathbf{E}X^2 \in \overline{\mathbb{R}}_+.$$

Preuve. Pour $u \neq 0$, on a

$$\begin{aligned} I(u) &:= \frac{2(1 - \operatorname{Re} \varphi(u))}{u^2} = \frac{2}{u^2} \left(1 - \int_{\mathbb{R}} \cos(ux) dP_X(x) \right) \\ &= 2 \int_{\mathbb{R}} \frac{1 - \cos(ux)}{u^2} dP_X(x). \end{aligned}$$

On remarque alors que :

$$\frac{1 - \cos(ux)}{u^2} = \frac{2 \sin^2(ux/2)}{u^2} \leq \frac{2}{u^2} \left(\frac{ux}{2} \right)^2 = \frac{x^2}{2}. \quad (8)$$

De plus on a $\lim_{u \rightarrow 0} (1 - \cos(ux))/u^2 = x^2/2$. On résiste ici à la tentation d'appliquer le théorème de convergence dominée pour conclure car on ignore si $\int_{\mathbb{R}} x^2 dP_X(x) = \mathbf{E}X^2$ est fini. Par contre on a toujours

$$\int_{\mathbb{R}} x^2 dP_X(x) \leq \liminf_{u \rightarrow 0} I(u) \leq \limsup_{u \rightarrow 0} I(u) \leq \int_{\mathbb{R}} x^2 dP_X(x),$$

où la première inégalité résulte du lemme de Fatou et la dernière de (8). Le lemme est prouvé. \square

Cas où X_1 a une loi symétrique.

Notons $\varphi(u) := \mathbf{E} \exp(iuX_1)$ la fonction caractéristique de X_1 et φ_n celle de S_n/\sqrt{n} . Comme les X_k sont i.i.d., $\varphi_n(u) = \varphi(u/\sqrt{n})^n$. De plus, X_1 étant symétrique, φ et donc φ_n sont réelles. L'hypothèse de convergence en loi de S_n/\sqrt{n} vers $\mathfrak{N}(0, 1)$ équivaut à la convergence ponctuelle de $\varphi_n(u)$ vers $\exp(-u^2/2)$. En particulier pour $u = 1$ on a $\varphi(1/\sqrt{n})^n \rightarrow \exp(-1/2)$ et $\varphi(1/\sqrt{n})^n > 0$ pour $n \geq n_0$. On peut passer au logarithme et en déduire $\ln \varphi(1/\sqrt{n}) \sim -1/(2n)$, puis $\varphi(1/\sqrt{n}) - 1 \sim -1/(2n)$. En réécrivant ceci sous la forme

$$\lim_{n \rightarrow +\infty} \frac{2 \left(1 - \varphi\left(\frac{1}{\sqrt{n}}\right) \right)}{\frac{1}{n}} = 1,$$

le lemme 9 nous donne $\mathbf{E}X_1^2 = 1$.

Lemme 10. Si X et Y sont deux variables aléatoires indépendantes telles que $X + Y \in L^2(\Omega)$, alors $X \in L^2(\Omega)$ et $Y \in L^2(\Omega)$.

Preuve. Compte tenu de l'indépendance de X et Y , une simple application du théorème de Fubini nous donne

$$\int_{\mathbb{R}^2} (x + y)^2 P_{X,Y}(dx, dy) = \int_{\mathbb{R}^2} (x + y)^2 P_X(dx) P_Y(dy) = \int_{\mathbb{R}} P_X(dx) \int_{\mathbb{R}} (x + y)^2 P_Y(dy).$$

Par conséquent,

$$\mathbf{E}(X + Y)^2 = \int_{\mathbb{R}} P_X(dx) \mathbf{E}(x + Y)^2 < +\infty,$$

d'où $\mathbf{E}(x + Y)^2 < +\infty$ pour P_X -presque tout $x \in \mathbb{R}$, donc au moins pour un x_0 . Ainsi $x_0 + Y$ est dans $L^2(\Omega)$ et comme les constantes sont dans l'espace vectoriel $L^2(\Omega)$, il en va de même pour $Y = (x_0 + Y) - x_0$. Enfin l'égalité $X = (X + Y) - Y$ donne l'appartenance de X à $L^2(\Omega)$. \square

Cas général.

Considérons les variables aléatoires

$$Z_n := \frac{(X_1 - X_2) + (X_3 - X_4) + \cdots + (X_{2n-1} - X_{2n})}{\sqrt{2n}} = \frac{1}{\sqrt{2}} \left(\frac{S'_n}{\sqrt{n}} - \frac{S''_n}{\sqrt{n}} \right),$$

où

$$S'_n := \sum_{k=1}^n X_{2k-1}, \quad S''_n := \sum_{k=1}^n X_{2k}.$$

Par hypothèse, $n^{-1/2}S'_n$ et $n^{-1/2}S''_n$ convergent en loi vers $\mathfrak{N}(0, 1)$. Comme elles sont indépendantes on en déduit la convergence en loi du vecteur aléatoire (S'_n, S''_n) vers (Z', Z'') où Z' et Z'' sont deux variables aléatoires indépendantes de même loi $\mathfrak{N}(0, 1)$. On en déduit par image continue que

$$Z_n = \frac{1}{\sqrt{2}} \left(\frac{S'_n}{\sqrt{n}} - \frac{S''_n}{\sqrt{n}} \right) \xrightarrow[n \rightarrow +\infty]{\text{loi}} \frac{1}{\sqrt{2}} (Z' - Z'').$$

Il est immédiat de vérifier (regarder la fonction caractéristique) que $(Z' - Z'')/\sqrt{2}$ a pour loi $\mathfrak{N}(0, 1)$. Notons maintenant

$$Y_k := \frac{X_{2k-1} - X_{2k}}{\sqrt{2}}, \quad k \geq 1$$

et remarquons que puisque X_{2k-1} et X_{2k} sont indépendantes et de même loi, la loi de Y_k est symétrique. On vient donc de montrer que

$$Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, 1).$$

Par le cas symétrique ci-dessus on en déduit que $\mathbf{E}Y_1^2 < +\infty$. Le lemme 10 appliqué à $X = X_1$ et $Y = -X_2$ (on n'a pas supposé dans ce lemme que X et Y ont même loi) nous donne $\mathbf{E}X_1^2 < +\infty$, ce qui achève la preuve. \square

1.4 Vitesse de convergence

La vitesse de convergence dans le théorème 3 est dans les bons cas en $O(n^{-1/2})$, comme pour le théorème de de Moivre-Laplace. Plus précisément on a le résultat suivant.

Théorème 11 (Berry-Esséen, 1941–42). *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. telle que $\mathbf{E}X_1 = 0$, $\mathbf{E}|X_1|^3 < +\infty$. On note $\sigma^2 := \mathbf{E}X_1^2$ ($\sigma > 0$). Il existe alors une constante universelle $C > 0$ telle que pour tout $n \geq 1$,*

$$\Delta_n := \sup_{x \in \mathbb{R}} \left| \mathbf{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) - \Phi(x) \right| \leq C \frac{\mathbf{E}|X_1|^3}{\sigma^3} \frac{1}{\sqrt{n}}.$$

On pourra trouver une démonstration de ce théorème dans Feller [4, T.2, XVI.5]. La clé du problème est l'obtention d'une formule d'inversion permettant de contrôler la différence des fonctions de répartition à l'aide de la différence des fonctions caractéristiques. Une fois celle-ci acquise, on effectue un développement à l'ordre 3 comme dans la preuve du T.L.C., c'est ce qui explique la présence de $\mathbf{E}|X_1|^3$ dans le majorant. L'obtention de la meilleure constante C a été l'objet d'une longue quête. La valeur initiale de Esséen était $C = 7,59$. Feller propose $C = 3$. Une valeur plus moderne et proche de l'optimale est $C = 0,7975$ (Van Beek (1972)).

Il est intéressant de regarder ce que donne le théorème de Berry-Esséen pour le cas de de Moivre-Laplace, donc avec $X_1 = Y_1 - p$, où Y_1 suit une loi de Bernoulli de paramètre p . On trouve alors

$$\Delta_n \leq C \frac{p^2 + q^2}{\sqrt{pq}} \frac{1}{\sqrt{n}}.$$

Des résultats plus précis concernant ce cas particulier se trouvent dans Uspensky [10] (voir l'annexe B.5 ci-dessous).

Voici un exemple tout à fait élémentaire permettant de comprendre qu'il n'y a pas lieu d'espérer une vitesse de convergence meilleure que $O(n^{-1/2})$ pour Δ_n . Prenons X_1 de loi de Bernoulli de paramètre $1/2$. On a alors

$$S_{2n} \sim \text{Bin}(2n, \frac{1}{2}), \quad \mathbf{E}S_{2n} = 2n \frac{1}{2} = n.$$

On cherche un équivalent de $\mathbf{P}(S_{2n}^* \leq 0) - \Phi(0)$. Remarquons d'abord que

$$\{S_{2n}^* < 0\} = \{0 \leq S_{2n} < n\} \quad \text{et} \quad \{S_{2n}^* > 0\} = \{n < S_{2n} \leq 2n\}.$$

En raison de la symétrie des coefficients binomiaux ($C_{2n}^k = C_{2n}^{2n-k}$),

$$\mathbf{P}(S_{2n}^* < 0) = \sum_{k=0}^{n-1} C_{2n}^k 2^{-2n} = \sum_{j=n+1}^{2n} C_{2n}^j 2^{-2n} = \mathbf{P}(S_{2n}^* > 0).$$

On a ainsi $2\mathbf{P}(S_{2n}^* < 0) + \mathbf{P}(S_{2n}^* = 0) = 1$ d'où l'on tire $\mathbf{P}(S_{2n}^* < 0) = \frac{1}{2} - \frac{1}{2}\mathbf{P}(S_{2n}^* = 0)$ et $\mathbf{P}(S_{2n}^* \leq 0) = \frac{1}{2} + \frac{1}{2}\mathbf{P}(S_{2n}^* = 0)$. En rappelant que $\Phi(0) = \frac{1}{2}$, on aboutit à

$$\mathbf{P}(S_{2n}^* \leq 0) - \Phi(0) = \frac{1}{2}\mathbf{P}(S_{2n}^* = 0) = \frac{1}{2}\mathbf{P}(S_{2n} = n) = C_{2n}^n 2^{-2n-1}.$$

Par la formule de Stirling ($n! \sim \sqrt{2\pi n} n^{n+1/2} e^{-n}$), on obtient l'équivalent

$$\mathbf{P}(S_{2n}^* \leq 0) - \Phi(0) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2n}}.$$

Comme $(2\pi)^{-1/2} > 0,398\ 9$, on a pour $n \geq n_0$, $|\mathbf{P}(S_{2n}^* \leq 0) - \Phi(0)| \geq 0,398(2n)^{-1/2}$, minorant à comparer avec le majorant uniforme $\Delta_{2n} \leq 0,798(2n)^{-1/2}$ fourni dans ce cas par le théorème de Berry-Esséen⁶.

Revenons à la situation générale du théorème 11. Que se passe-t-il dans la zone intermédiaire où X_1 vérifie le TLC (i.e. $\mathbf{E}X_1^2 < +\infty$, cf. th. 8) mais n'a pas de moment d'ordre 3? On a toujours une vitesse de convergence, pourvu que l'intégrabilité de X_1 soit un peu plus forte que la seule existence d'un moment d'ordre 2. Le résultat précis est le suivant (voir Petrov [7, Ch. V, th. 5, 6]).

Théorème 12 (Katz 1963, Petrov 1965). *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. telle que $\mathbf{E}X_1 = 0$, $\mathbf{E}X_1^2 g(X_1) < +\infty$, où g est une fonction positive, paire, croissante sur $]0, +\infty[$ et telle que $x/g(x)$ soit croissante sur $]0, +\infty[$. Il existe alors une constante universelle $A > 0$ telle que pour tout $n \geq 1$,*

$$\Delta_n := \sup_{x \in \mathbb{R}} \left| \mathbf{P}\left(\frac{S_n}{\sqrt{n}} \leq x\right) - \Phi(x) \right| \leq A \frac{\mathbf{E}(X_1^2 g(X_1))}{\sigma^2 g(\sigma\sqrt{n})}.$$

En particulier si $\mathbf{E}|X_1|^{2+\delta} < +\infty$ pour un $\delta \in]0, 1]$,

$$\Delta_n \leq A \frac{\mathbf{E}|X_1|^{2+\delta}}{\sigma^{2+\delta}} \frac{1}{n^{\delta/2}}.$$

2 Le TLC sans équidistribution

On envisage maintenant la question du TLC pour des variables aléatoires indépendantes n'ayant pas forcément même loi. Le résultat essentiel est le théorème de Lindeberg dont on discute la signification avant d'en examiner plusieurs applications.

Théorème 13 (Lindeberg). *Considérons le « tableau triangulaire » de variables aléatoires de ligne numéro n ($n \in \mathbb{N}^*$) :*

$$X_{n,1}, \dots, X_{n,i}, \dots, X_{n,k_n},$$

où ces k_n variables sont définies sur le même $(\Omega_n, \mathcal{F}_n, \mathbf{P}_n)$, indépendantes, centrées, de carrés intégrables. On note

$$\sigma_{n,i}^2 := \text{Var } X_{n,i}, \quad s_n^2 := \sum_{i=1}^{k_n} \sigma_{n,i}^2, \quad S_n := \sum_{i=1}^{k_n} X_{n,i}.$$

6. Signalons cependant que dans le cas de la loi de Bernoulli de paramètre $p = 1/2$, il est possible d'obtenir une vitesse de convergence en $O(n^{-1})$ pour une variante de la distance Δ_n obtenue en appliquant la *correction de continuité*, voir le théorème 32 dans l'annexe B.5.

On suppose $s_n > 0$ pour tout n . Si de plus le tableau vérifie la condition de Lindeberg :

$$\forall \varepsilon > 0, \quad \frac{1}{s_n^2} \sum_{i=1}^{k_n} \int_{\{|X_{n,i}| > \varepsilon s_n\}} X_{n,i}^2 d\mathbf{P}_n \xrightarrow{n \rightarrow +\infty} 0, \quad (9)$$

alors S_n/s_n converge en loi vers $\mathfrak{N}(0, 1)$.

Commentaires : Ce théorème dont on trouvera la preuve dans Billingsley [2, Th. 27.2] est la version générale du TLC lorsque l'on ne suppose plus l'équadistribution. L'appellation « tableau triangulaire » est un abus de langage commode, la forme triangulaire n'apparaissant réellement que dans le cas particulier où $k_n = cn$. Remarquons qu'avec un tableau triangulaire, toutes les variables peuvent changer quand on passe de S_n à S_{n+1} , c'est pour cela qu'il n'est pas nécessaire de supposer que des variables situées sur des lignes différentes du tableau sont définies sur le même espace probabilisé. Bien sûr le théorème de Lindeberg contient le théorème 3 (cas d'une suite i.i.d. $(X_i)_{i \geq 1}$). Pour le vérifier, considérons le tableau triangulaire où $k_n = n$ et où la ligne n est : X_1, X_2, \dots, X_n . Ces variables étant indépendantes et de même loi, $s_n^2 = n\sigma^2$ et la condition de Lindeberg se réduit à :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \int_{\{|X_1| > \varepsilon \sigma \sqrt{n}\}} X_1^2 d\mathbf{P} = 0,$$

ce qui découle immédiatement de l'existence d'un moment d'ordre 2 par convergence dominée.

En revenant au cas général, il convient de remarquer que la condition (9) implique

$$\frac{1}{s_n^2} \max_{1 \leq k \leq k_n} \sigma_{n,k}^2 \xrightarrow{n \rightarrow +\infty} 0. \quad (10)$$

En effet on peut écrire pour tout $\varepsilon > 0$ et $k \leq k_n$,

$$\begin{aligned} \frac{1}{s_n^2} \sigma_{n,k}^2 &= \frac{1}{s_n^2} \int_{\Omega_n} X_{n,k}^2 d\mathbf{P}_n = \frac{1}{s_n^2} \int_{\{|X_{n,k}| \leq \varepsilon s_n\}} X_{n,k}^2 d\mathbf{P}_n + \frac{1}{s_n^2} \int_{\{|X_{n,k}| > \varepsilon s_n\}} X_{n,k}^2 d\mathbf{P}_n \\ &\leq \frac{\varepsilon^2 s_n^2}{s_n^2} + \frac{1}{s_n^2} \sum_{i=1}^{k_n} \int_{\{|X_{n,i}| > \varepsilon s_n\}} X_{n,i}^2 d\mathbf{P}_n. \end{aligned}$$

On en déduit

$$\frac{1}{s_n^2} \max_{1 \leq k \leq k_n} \sigma_{n,k}^2 \leq \varepsilon^2 + \frac{1}{s_n^2} \sum_{i=1}^{k_n} \int_{\{|X_{n,i}| > \varepsilon s_n\}} X_{n,i}^2 d\mathbf{P}_n,$$

puis en faisant tendre n vers $+\infty$, on aboutit à (10) via :

$$\forall \varepsilon > 0, \quad \limsup_{n \rightarrow +\infty} \frac{1}{s_n^2} \max_{1 \leq k \leq k_n} \sigma_{n,k}^2 \leq \varepsilon^2.$$

Une première conséquence de (10) est que dans le théorème de Lindeberg, k_n tend vers l'infini avec n , même si cela n'est pas mentionné explicitement dans les hypothèses. En

effet sinon (k_n) aurait une sous-suite bornée $(k_{n'})$ indexée par une partie infinie \mathbb{N}' de \mathbb{N} ($\forall n' \in \mathbb{N}', k_{n'} \leq M$). On aurait alors $s_{n'}^2 \leq M \max_{1 \leq k \leq k_{n'}} \sigma_{n',k}^2$, d'où

$$\forall n' \in \mathbb{N}', \quad \frac{1}{s_{n'}^2} \max_{1 \leq k \leq k_{n'}} \sigma_{n',k}^2 \geq \frac{1}{M} > 0,$$

en contradiction avec (10).

Voici une deuxième conséquence importante : (10) combinée avec l'inégalité de Tchebycheff nous donne (rappelons que les $X_{n,k}$ sont centrées) :

$$\forall \varepsilon > 0, \quad \max_{1 \leq k \leq k_n} \mathbf{P} \left(\left| \frac{X_{n,k}}{s_n} \right| \geq \varepsilon \right) \xrightarrow{n \rightarrow +\infty} 0. \quad (11)$$

Cette condition signifie intuitivement qu'aucune des variables $X_{n,k}$ n'apporte individuellement une contribution substantielle à S_n . Lorsque (11) est satisfaite, on dit que les variables $X_{n,k}/s_n$ forment un *tableau infinitésimal* ou qu'elles sont *uniformément asymptotiquement négligeables*. Selon un théorème de Feller [2, Th. 27.4], si le tableau triangulaire de variables aléatoires indépendantes $X_{n,k}$ est infinitésimal et si S_n/s_n converge en loi vers $\mathfrak{N}(0, 1)$, alors la condition de Lindeberg est vérifiée. Ainsi (9) est une condition optimale pour qu'un tableau infinitésimal vérifie le TLC.

Pour compléter cette discussion du théorème de Lindeberg, voici un exemple simple de suite de variables aléatoires indépendantes vérifiant le TLC mais pas la condition de Lindeberg. Prenons les X_k indépendantes de loi $\mathfrak{N}(0, \sigma_k^2)$, avec $\sigma_k^2 = 2^{k-1}$. Ainsi $s_n^2 = \sum_{k=1}^n 2^{k-1} = 2^n - 1$ et

$$\max_{1 \leq k \leq n} \frac{\sigma_k^2}{s_n^2} = \frac{2^{n-1}}{2^n - 1} \xrightarrow{n \rightarrow +\infty} \frac{1}{2} \neq 0,$$

de sorte que le tableau triangulaire $(X_k, 1 \leq k \leq n)_{n \geq 1}$ ne vérifie pas (10) ni par conséquent la condition de Lindeberg. Par ailleurs il est évident directement que S_n est gaussienne de loi $\mathfrak{N}(0, s_n^2)$ donc S_n/s_n est pour tout n de loi $\mathfrak{N}(0, 1)$, ce qui rend triviale la convergence en loi de S_n/s_n vers $\mathfrak{N}(0, 1)$.

Application du théorème de Lindeberg à un test d'égalité de deux espérances.

On dispose d'un échantillon X_1, \dots, X_l de la loi inconnue P_X (les X_i sont i.i.d. de même loi qu'une v.a. aléatoire générique X) et d'un échantillon Y_1, \dots, Y_m de la loi P_Y . On voudrait tester l'hypothèse

$$(\mathcal{H}_0) : \quad \mathbf{E}X = \mathbf{E}Y \quad \text{contre} \quad (\mathcal{H}_1) : \quad \mathbf{E}X \neq \mathbf{E}Y.$$

Cette situation se présente notamment en médecine quand on veut tester l'efficacité d'un nouveau médicament et où l'on observe les durées de guérisons X_i d'un premier échantillon de patients auxquels on a administré le nouveau médicament, tandis que les Y_i sont les durées de guérison des patients d'un autre échantillon ayant reçu un médicament ancien (ou un placebo!). A priori on ne connaît pas les variances des lois P_X et P_Y . On construit une statistique de test

$$T := \frac{\bar{Y} - \bar{X}}{\sqrt{\frac{1}{l}S_X^2 + \frac{1}{m}S_Y^2}}, \quad (12)$$

où \bar{X} , \bar{Y} sont les moyennes et S_X^2 , S_Y^2 les variances empiriques :

$$\bar{X} := \frac{1}{l} \sum_{i=1}^l X_i, \quad S_X^2 := \frac{1}{l} \sum_{i=1}^l (X_i - \bar{X})^2, \quad \text{etc.}$$

Pour l et m « grands », on considère que sous (\mathcal{H}_0) , T doit être approximativement gaussienne $\mathfrak{N}(0, 1)$. Ceci amène à définir un test de seuil⁷ ε en décidant de rejeter (\mathcal{H}_0) si l'on observe $|T| > t_\varepsilon$, où t_ε est défini par $\mathbf{P}(|Z| > t_\varepsilon) = \varepsilon$, avec $Z \sim \mathfrak{N}(0, 1)$.

La légitimation théorique de ce test est contenue dans le théorème suivant qui est une simple application du théorème de Lindeberg.

Théorème 14. *On suppose que X et Y sont de carré intégrable (avec $\sigma_X, \sigma_Y > 0$) et que $l = l(n)$ et $m = m(n)$ tendent vers l'infini avec n . Alors sous (\mathcal{H}_0) la statistique de test $T = T_n$ définie par (12) converge en loi vers $\mathfrak{N}(0, 1)$ lorsque n tend vers l'infini. Par contre, sous (\mathcal{H}_1) , $|T_n|$ tend p.s. vers l'infini.*

Preuve. Que l'on soit sous (\mathcal{H}_0) ou (\mathcal{H}_1) , l'indépendance et l'équadistribution des X_i d'une part et des Y_j d'autre part nous donnent par la loi forte des grands nombres les convergences presque sûres de \bar{X} vers $\mathbf{E}X$, de S_X^2 vers σ_X^2 , de \bar{Y} vers $\mathbf{E}Y$ et de S_Y^2 vers σ_Y^2 . Ainsi le numérateur de T_n converge p.s. vers $\mathbf{E}X - \mathbf{E}Y$ et le dénominateur p.s. vers 0. Il est alors clair que sous (\mathcal{H}_1) , $|T_n|$ tend p.s. vers l'infini.

Sous (\mathcal{H}_0) on a une forme indéterminée du type « 0/0 » pour la convergence p.s. et on étudie la convergence en loi. Considérons pour cela le tableau triangulaire de n -ième ligne

$$\frac{-X_1}{l}, \dots, \frac{-X_l}{l}, \frac{Y_1}{m}, \dots, \frac{Y_m}{m}. \quad (13)$$

Ici $k_n = l + m$ et la somme de la ligne est $S_n = \bar{Y} - \bar{X}$. Grâce à l'indépendance, sa variance est

$$s_n^2 = \text{Var } S_n = \frac{1}{l^2} \sum_{i=1}^l \text{Var } X_i + \frac{1}{m^2} \sum_{j=1}^m \text{Var } Y_j = \frac{1}{l} \sigma_X^2 + \frac{1}{m} \sigma_Y^2.$$

Par la loi forte des grands nombres, on vérifie facilement que

$$\frac{\frac{1}{l(n)} \sigma_X^2 + \frac{1}{m(n)} \sigma_Y^2}{\frac{1}{l(n)} S_X^2 + \frac{1}{m(n)} S_Y^2} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 1.$$

En écrivant T_n sous la forme

$$T_n = \left(\frac{\frac{1}{l} \sigma_X^2 + \frac{1}{m} \sigma_Y^2}{\frac{1}{l} S_X^2 + \frac{1}{m} S_Y^2} \right)^{1/2} \times \frac{S_n}{s_n}$$

⁷ Le seuil est un majorant de la probabilité (inconnue!) de rejeter à tort (\mathcal{H}_0) . On néglige dans cette affirmation l'erreur due à l'approximation gaussienne.

et en invoquant le lemme de Slutsky, on réduit la preuve du théorème à la vérification de la condition de Lindeberg pour le tableau triangulaire (13). Il s'agit donc de montrer pour tout $\varepsilon > 0$ la convergence vers 0 de

$$\begin{aligned} A_n(\varepsilon) &:= \frac{1}{s_n^2} \left[\sum_{i=1}^l \int_{\{|u| \geq \varepsilon s_n\}} u^2 P_{(-X_i/l)}(du) + \sum_{j=1}^m \int_{\{|v| \geq \varepsilon s_n\}} v^2 P_{(Y_j/m)}(dv) \right] \\ &= \frac{l}{s_n^2} \int_{\{|u| \geq \varepsilon s_n\}} u^2 P_{(-X_1/l)}(du) + \frac{m}{s_n^2} \int_{\{|v| \geq \varepsilon s_n\}} v^2 P_{(Y_1/m)}(dv). \end{aligned}$$

Les changements de variable $u = -x/l$ et $v = y/m$ nous donnent comme mesures images respectives P_{X_1} et P_{Y_1} et comme nouveaux ensembles d'intégration $\{|x| \geq \varepsilon l s_n\}$ et $\{|y| \geq \varepsilon m s_n\}$. On voit ainsi que

$$A_n(\varepsilon) = \frac{1}{l s_n^2} \int_{\{|x| \geq \varepsilon l s_n\}} x^2 P_{X_1}(dx) + \frac{1}{m s_n^2} \int_{\{|y| \geq \varepsilon m s_n\}} y^2 P_{Y_1}(dy).$$

Pour majorer $A_n(\varepsilon)$, on remarque que $l s_n^2 > \sigma_X^2$ d'où l'inclusion

$$\{|x| \geq \varepsilon l s_n\} = \{x^2 \geq \varepsilon^2 l^2 s_n^2\} \subset \{x^2 \geq l \varepsilon^2 \sigma_X^2\}.$$

En faisant de même avec m et σ_Y^2 , on obtient

$$A_n(\varepsilon) \leq \frac{1}{\sigma_X^2} \int_{\{x^2 \geq l \varepsilon^2 \sigma_X^2\}} x^2 P_{X_1}(dx) + \frac{1}{\sigma_Y^2} \int_{\{y^2 \geq m \varepsilon^2 \sigma_Y^2\}} y^2 P_{Y_1}(dy).$$

Comme $l = l(n)$ et $m = m(n)$ tendent vers $+\infty$, ce majorant tend vers 0 par convergence dominée puisque X_1 et Y_1 sont de carré intégrable. \square

Théorème 15 (Lyapounov). *Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires définies sur le même espace probabilisé, indépendantes (mais pas forcément de même loi), centrées, ayant toutes un moment d'ordre $2 + \delta$ ($\delta > 0$). On note $s_n^2 := \text{Var}(S_n)$ et on suppose $s_n > 0$. On suppose de plus vérifiée la condition de Lyapounov :*

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^n \mathbf{E}|X_k|^{2+\delta} \xrightarrow{n \rightarrow +\infty} 0. \quad (14)$$

Alors S_n/s_n converge en loi vers $\mathfrak{N}(0, 1)$.

Preuve. On applique le théorème de Lindeberg au tableau triangulaire $X_{n,k} := X_k$, $1 \leq k \leq n$. Il suffit alors de vérifier que la condition de Lyapounov (14) entraîne celle de Lindeberg (9). En notant que sur l'évènement $\{|X_k| \geq \varepsilon s_n\}$ on a $|X_k|^\delta / (\varepsilon s_n)^\delta \geq 1$, la majoration

$$\int_{\{|X_k| \geq \varepsilon s_n\}} X_k^2 d\mathbf{P} \leq \int_{\{|X_k| \geq \varepsilon s_n\}} X_k^2 \left(\frac{|X_k|}{\varepsilon s_n} \right)^\delta d\mathbf{P} \leq \frac{1}{\varepsilon^\delta s_n^\delta} \int_{\Omega} |X_k|^{2+\delta} d\mathbf{P},$$

donne la conclusion escomptée via l'inégalité

$$\frac{1}{s_n^2} \sum_{k=1}^n \int_{\{|X_k| \geq \varepsilon s_n\}} X_k^2 d\mathbf{P} \leq \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{k=1}^n \mathbf{E}|X_k|^{2+\delta}.$$

□

On trouvera des applications du théorème de Lyapounov dans Foata Fuchs [5, p. 248–249]

3 Le TLC dans \mathbb{R}^d

Nous étudions dans cette section le TLC en dimension finie. Par souci de simplicité, on se limitera essentiellement au TLC iid dans \mathbb{R}^d . On verra qu'en fait il y a toujours moyen de se ramener à la dimension 1 grâce au lemme de Cramér-Wold (« Cramér-Wold device » dans la littérature anglo-saxonne). Néanmoins, cela ne dispense pas de la possession d'un minimum de connaissances sur les vecteurs aléatoires (fonctionnelles caractéristiques, structure de covariance, lois gaussiennes en dimension d). Il est commode de présenter ces « rappels » dans le cadre plus abstrait d'un espace vectoriel E , d'une part pour éviter de faire jouer prématurément un rôle à la structure euclidienne de \mathbb{R}^d , d'autre part pour ouvrir la voie en direction des lois gaussiennes et du TLC en dimension infinie.

3.1 Vecteurs aléatoires

Soit donc E un \mathbb{R} -espace vectoriel de dimension finie d et E' son dual. Rappelons qu'il n'y a qu'une seule topologie d'espace vectoriel sur E (i.e. rendant continues l'addition des vecteurs et la multiplication d'un vecteur par un scalaire) et que toutes les normes sur E sont équivalentes et métrisent chacune cette topologie. Dans la suite, E sera muni de sa tribu borélienne, c'est-à-dire la tribu engendrée par les ouverts de E . Un vecteur aléatoire X dans E est une application mesurable d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbf{P})$ dans E .

La fonctionnelle caractéristique de X (ou de sa loi) est l'application

$$\varphi_X : E' \rightarrow \mathbb{C}, \quad u \mapsto \varphi_X(u) := \mathbf{E} \exp(iu(X)).$$

Le lien avec les fonctions caractéristiques en dimension 1 est donné par

$$\forall u \in E', \forall t \in \mathbb{R}, \quad \varphi_X(tu) = \mathbf{E} \exp(itu(X)) = \varphi_{u(X)}(t).$$

Comme en dimension 1, la fonctionnelle caractéristique caractérise la loi et la convergence ponctuelle des fonctionnelles caractéristiques équivaut à la convergence en loi⁸. Le lemme de Cramér-Wold est une conséquence immédiate de ces remarques.

8. Cette dernière propriété ne subsiste pas en dimension infinie, disons dans le cas où E est un espace de Banach séparable et où φ_X obéit à la même définition formelle que ci-dessus, en prenant pour E' le dual *topologique* de E .

Lemme 16 (Cramér-Wold). *La suite $(X_n)_{n \geq 1}$ de vecteurs aléatoires dans l'espace vectoriel de dimension finie E converge en loi dans E vers le vecteur aléatoire X si et seulement si pour toute forme linéaire $u \in E'$, la suite de variables aléatoires réelles $(u(X_n))_{n \geq 1}$ converge en loi dans \mathbb{R} vers $u(X)$.*

Passons à la définition de l'espérance d'un vecteur aléatoire X . On suppose pour cela que pour tout $u \in E'$, $\mathbf{E}|u(X)|$ est finie⁹ et on considère l'application

$$\theta : E' \rightarrow \mathbb{R}, \quad u \mapsto \mathbf{E}u(X).$$

En raison de la linéarité de l'espérance, θ est une forme linéaire sur E' donc un élément du bidual E'' . Il existe alors un unique élément $x_0 \in E$ tel que $\theta(u) = u(x_0)$. C'est cet élément que l'on définit¹⁰ comme le *vecteur déterministe* $\mathbf{E}X$. Ainsi $\mathbf{E}X$ est l'unique élément de E vérifiant

$$\forall u \in E', \quad u(\mathbf{E}X) = \mathbf{E}u(X). \quad (15)$$

Pour donner une expression plus familière de $\mathbf{E}X$, choisissons une base (e_1, \dots, e_d) de E et notons e'_k les formes coordonnées associées ($e'_k(e_j) = \delta_{j,k}$, symbole de Kronecker) de sorte que (e'_1, \dots, e'_d) est une base de E' et tout $x \in E$ s'écrit $x = \sum_{k=1}^d e'_k(x)e_k$. En appliquant (15) avec $u = e'_k$ ($k = 1, \dots, d$), on obtient en notant que $X = \sum_{k=1}^d e'_k(X)e_k$,

$$\mathbf{E}X = \sum_{k=1}^d (\mathbf{E}e'_k(X))e_k = (\mathbf{E}e'_1(X), \dots, \mathbf{E}e'_d(X)).$$

Supposons maintenant que pour tout $u \in E'$, $u(X)$ soit de carré intégrable¹¹ et définissons

$$B : E' \times E' \rightarrow \mathbb{R}, \quad (u, v) \mapsto B(u, v) := \text{Cov}(u(X), v(X)).$$

B est clairement une forme bilinéaire sur $E' \times E'$ et en utilisant l'identification de E'' avec E , on a pour chaque $v \in E'$ un unique élément $y = Kv \in E$ tel que $u(y) = B(u, v)$. L'opérateur $K : E' \rightarrow E$ est linéaire, autoadjoint ($u(Kv) = v(Ku)$) et positif ($u(Ku) \geq 0$). On l'appelle *l'opérateur de covariance de X* . Il est défini de manière intrinsèque par

$$\forall u, v \in E', \quad u(Kv) = \text{Cov}(u(X), v(X)). \quad (16)$$

On peut donner une représentation matricielle de K en munissant l'espace d'arrivée E d'une base (e_1, \dots, e_d) et l'espace de départ E' de la base duale (e'_1, \dots, e'_d) définie

9. En dimension finie, cette condition équivaut à $\mathbf{E}\|X\| < +\infty$ (peu importe pour quelle norme sur E puisqu'elles sont toutes équivalentes). Vérification laissée en exercice.

10. Cette définition indirecte de l'espérance d'un vecteur aléatoire correspond à l'intégrale vectorielle faible ou intégrale de Pettis. On pourrait aussi construire une intégrale vectorielle dite forte (ou de Bochner) par approximation à partir de fonctions vectorielles ne prenant qu'un nombre fini de valeurs en imitant la construction de l'intégrale de Lebesgue. En dimension finie, les deux types d'intégrale coïncident et leur existence équivaut à $\mathbf{E}\|X\| < +\infty$. Ce n'est plus vrai en dimension infinie : l'existence de l'intégrale de Bochner équivaut à $\mathbf{E}\|X\|^2 < +\infty$ et implique celle de l'intégrale de Pettis (les deux ayant alors même valeur). La réciproque est fautive.

11. En dimension finie cette condition équivaut à $\mathbf{E}\|X\|^2 < +\infty$.

comme ci-dessus. En choisissant $u = e'_i$ et $v = e'_j$ dans (16), on voit que la matrice de K a pour terme général

$$K_{i,j} = \text{Cov}(e'_i(X), e'_j(X)), \quad 1 \leq i, j \leq d. \quad (17)$$

L'opérateur de covariance donne une information sur le support de la loi de X . Définissons en effet $H := \{u \in E'; u(Ku) = 0\}$. Comme l'appartenance de u à H équivaut à $\text{Var } u(X) = 0$, il est facile de vérifier que H est un sous-espace vectoriel de E' . Considérons alors $H^0 := \{x \in E; \forall u \in H, u(x) = 0\}$. Il est clair que H^0 est un s.e.v. de E . La loi de X est supportée par le sous-espace affine $\mathbf{E}X + H^0$, ce qui signifie que $\mathbf{P}(X - \mathbf{E}X \in H^0) = 1$. Pour le vérifier, notons d_0 la dimension de H et choisissons une base quelconque $(f_i)_{1 \leq i \leq d_0}$ de H . On a alors

$$x \notin H^0 \Leftrightarrow \exists i \in \{1, \dots, d_0\}, f_i(x) \neq 0,$$

d'où

$$\mathbf{P}(X - \mathbf{E}X \notin H^0) = \mathbf{P}\left(\bigcup_{i=1}^{d_0} \{f_i(X - \mathbf{E}X) \neq 0\}\right) \leq \sum_{i=1}^{d_0} \mathbf{P}(f_i(X - \mathbf{E}X) \neq 0). \quad (18)$$

Or par définition de H , pour tout $i = 1, \dots, d_0$, $\text{Var } f_i(X - \mathbf{E}X) = 0$, ce qui implique que la variable aléatoire réelle $f_i(X - \mathbf{E}X)$ qui est d'espérance nulle, vaut 0 avec probabilité 1. Le membre de droite de (18) est donc nul et on a bien $X \in \mathbf{E}X + H^0$ presque sûrement¹².

On dit que le vecteur aléatoire X est *gaussien* si pour tout $u \in E'$, la variable aléatoire réelle $u(X)$ est gaussienne¹³. Ceci implique que $u(X)$ est de carré intégrable pour tout u et d'après ce qui précède, X a donc une espérance $m \in E$ et un opérateur de covariance $K : E' \rightarrow E$. La loi de X est alors notée $\mathfrak{N}(m, K)$. Comme $\mathbf{E}u(X) = u(m)$ et $\text{Var } u(X) = u(Ku)$, la loi gaussienne de $u(X)$ est $\mathfrak{N}(u(m), u(Ku))$. On en déduit que la fonctionnelle caractéristique de $\mathfrak{N}(m, K)$ s'écrit :

$$\varphi_X(u) = \exp(iu(m) - u(Ku)/2), \quad u \in E'.$$

Une conséquence immédiate de cette définition est que l'image $T(X)$ d'un vecteur gaussien de E par une application linéaire T de E dans F est un vecteur gaussien de F ayant pour opérateur de covariance TKT' où $T' : F' \rightarrow E'$ est l'adjoint de T , défini par

$$\forall x \in E, \forall v \in F', \quad v(T(x)) = (T'(v))(x).$$

Dans le cas usuel $E = \mathbb{R}^d$, on identifie E et E' et on note Σ la matrice de K pour la base canonique de \mathbb{R}^d . Lorsque K est défini positif (le sous-espace H de E' défini ci-dessus

12. On peut d'ailleurs montrer que $\mathbf{E}X + H^0$ est le plus petit sous-espace affine de E supportant la loi de X , cf. Toulouse [9, p. 37].

13. Rappelons que la famille des lois gaussiennes en dimension 1 contient les masses de Dirac ou lois des variables aléatoires constantes qui correspondent au cas dégénéré d'une variance $\sigma^2 = 0$.

est donc réduit au vecteur nul et K est inversible) la loi gaussienne $\mathfrak{N}(m, \Sigma)$ a alors pour densité par rapport à la mesure de Lebesgue canonique¹⁴ de \mathbb{R}^d

$$x \mapsto \frac{1}{(2\pi)^{d/2}(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2} {}^t(x - m)\Sigma^{-1}(x - m)\right).$$

Dans ce cas, le support de la loi gaussienne $\mathfrak{N}(m, K)$ est tout l'espace \mathbb{R}^d . Si K est seulement semi-défini positif, la loi $\mathfrak{N}(m, K)$ existe encore, mais est supportée par le sous espace affine $m + H^0$, où H^0 est maintenant l'orthogonal de H (après identification de $E = \mathbb{R}^d$ et E') et a une densité par rapport à la mesure de Lebesgue canonique de ce sous-espace, sauf dans le cas complètement dégénéré où $H = \mathbb{R}^d$ et où $\mathfrak{N}(m, K)$ est la masse de Dirac au point m .

3.2 TLC iid !

Théorème 17. *Soit (X_k) une suite de vecteurs aléatoires de \mathbb{R}^d , indépendants, de même loi et de carré intégrable (i.e. $\mathbf{E}\|X_1\|^2 < +\infty$) et $S_n := \sum_{k=1}^n X_k$. Alors*

$$\frac{S_n - \mathbf{E}S_n}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, K), \quad (19)$$

où K est l'opérateur de covariance de X_1 .

Commentaire : Malgré la ressemblance formelle de cet énoncé avec celui du théorème 3, il n'est pas possible ici de modifier la normalisation \sqrt{n} par $\sigma\sqrt{n}$ pour avoir toujours la même loi limite $\mathfrak{N}(0, I)$, où I est l'opérateur de matrice identité par rapport à la base canonique de \mathbb{R}^d et à sa duale.

Preuve. Par le lemme de Cramér-Wold, il suffit de vérifier que pour toute forme linéaire u sur \mathbb{R}^d , $u(n^{-1/2}(S_n - \mathbf{E}S_n))$ converge en loi vers $u(Z)$ où Z désigne un vecteur aléatoire gaussien de loi $\mathfrak{N}(0, K)$. Notons d'emblée que $\mathbf{E}u(Z) = u(\mathbf{E}Z) = u(0) = 0$ et $\text{Var } u(Z) = u(Ku) = \text{Var } u(X_1)$.

Par linéarité de u ,

$$u\left(\frac{S_n - \mathbf{E}S_n}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}} \sum_{k=1}^n (u(X_k) - \mathbf{E}u(X_k)).$$

Les $u(X_k)$ sont des variables aléatoires réelles indépendantes, de même loi, de carré intégrable ($\mathbf{E}|u(X_1)|^2 \leq \|u\|_{E'}^2 \mathbf{E}\|X_1\|_E^2$), on peut donc appliquer le théorème 3 qui donne ici

$$u\left(\frac{S_n - \mathbf{E}S_n}{\sqrt{n}}\right) \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, \text{Var } u(X_1)).$$

La loi limite est bien celle de $u(Z)$. □

14. Celle qui donne la masse 1 au cube unité construit sur la base canonique de \mathbb{R}^d ou à tout cube unité construit sur une base orthonormale de \mathbb{R}^d pour le produit scalaire usuel.

Grâce au théorème 8, il est facile de vérifier (exercice laissé au lecteur) que la condition $\mathbf{E}\|X_1\|^2 < +\infty$ est une C.N.S. pour la convergence (19).

Pour illustrer graphiquement le théorème dans \mathbb{R}^2 , on a choisi de générer 600 échantillons indépendants et de taille 500 d'une loi uniforme sur un triangle. Pour chacun de ces échantillons on a calculé $n^{-1/2}(S_n - \mathbf{E}S_n)$ et tracé le point correspondant. On obtient ainsi un nuage de 600 points qui se comporte approximativement comme un échantillon de taille 600 de la loi gaussienne limite. Les lignes de niveau tracées sur ce nuage de points sont celles de la densité de cette loi gaussienne (figures 1.1 et 1.2).

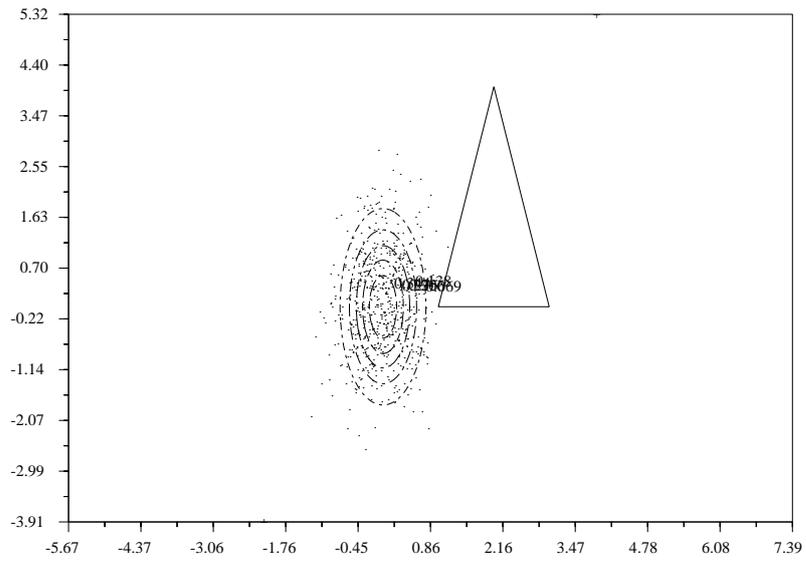


FIGURE 1.1 – $n^{-1/2}(S_n - \mathbf{E}S_n)$ pour 600 échantillons de la loi uniforme sur le triangle

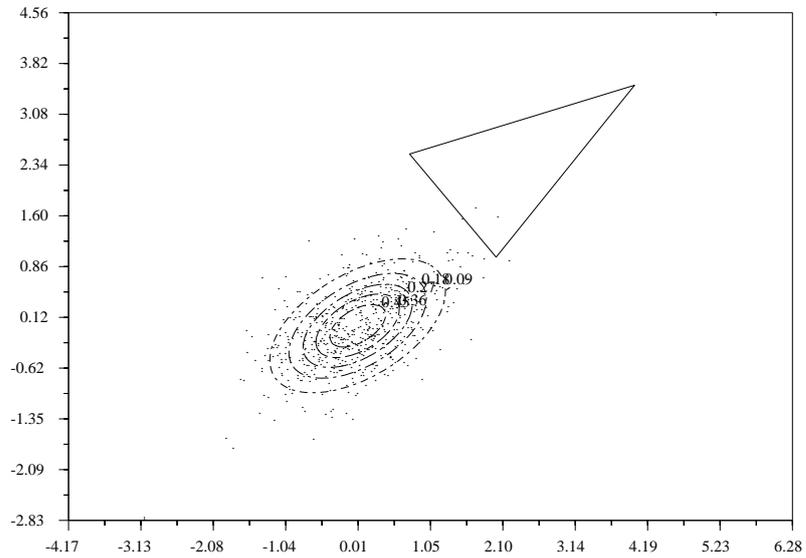


FIGURE 1.2 – $n^{-1/2}(S_n - \mathbf{E}S_n)$ pour 600 échantillons de la loi uniforme sur le triangle

Une application importante du théorème 17 est la convergence d'une loi multinomiale vers une loi gaussienne, ce qui en un certain sens, généralise le théorème de de Moivre Laplace.

Rappelons que la loi multinomiale sert à modéliser le total des résultats observés pour chaque type dans une suite d'épreuves répétées indépendantes ayant chacune d types de résultats possibles. Par exemple si on lance 200 fois un dé, on obtient un vecteur de dimension 6 dont la i -ème composante est le nombre total d'apparitions de la face numéro i au cours des 200 lancers. Ce vecteur suit la loi multinomiale de paramètres 200 et $(p_1, p_2, p_3, p_4, p_5, p_6)$, où les p_i valent tous $1/6$ si le dé est équilibré. Plus formellement, le vecteur aléatoire N suit la loi multinomiale de paramètres n et (p_1, \dots, p_d) où $n \in \mathbb{N}^*$ et les p_i sont strictement positifs et de somme 1, si pour tout d -uple (j_1, j_2, \dots, j_d) d'entiers tels que $j_1 + j_2 + \dots + j_d = n$,

$$\mathbf{P}\{N = (j_1, j_2, \dots, j_d)\} = \frac{n!}{j_1! j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d}.$$

Théorème 18. Si $(N_n)_{n \geq 1}$ est une suite de vecteurs aléatoires de \mathbb{R}^d de lois multinomiales de paramètres n et $p = (p_1, \dots, p_d)$,

$$\frac{1}{\sqrt{n}}(N_n - np) \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, K), \quad (20)$$

où la matrice de l'opérateur de covariance K dans la base canonique de \mathbb{R}^d a pour terme général

$$K_{i,j} = p_i \delta_{i,j} - p_i p_j, \quad 1 \leq i, j \leq d.$$

Preuve. Le vecteur aléatoire N_n a même loi que $S_n := \sum_{k=1}^n X_k$, où les X_k sont des vecteurs aléatoires de \mathbb{R}^d , indépendants et de même loi donnée par

$$\mathbf{P}\{X_k = (0, \dots, 0, \underset{\substack{\uparrow \\ i}}{1}, 0, \dots, 0)\} = p_i, \quad 1 \leq i \leq d.$$

Autrement dit, on se ramène à un modèle d'épreuves répétées indépendantes et le vecteur aléatoire X_k est un codage binaire du résultat de la k -ième épreuve. En notant $X_{k,i} = e'_i(X_k)$ la i -ème composante de X_k , on voit que cette variable aléatoire suit la loi de Bernoulli de paramètre p_i et que $\sum_{i=1}^d X_{k,i} = 1$. L'espérance de X_k est clairement le vecteur p . Le vecteur X_k étant borné a bien un moment d'ordre 2. On est ainsi dans les conditions de validité du théorème 17 qui nous donne

$$T_n := \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - p) \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, K), \quad (21)$$

où le terme général de la matrice de K par rapport à la base canonique de \mathbb{R}^d est

$$K_{i,j} = \text{Cov}(X_{1,i}, X_{1,j}) = \mathbf{E}X_{1,i}X_{1,j} - p_i p_j.$$

Si $i \neq j$, $X_{1,i}X_{1,j} = 0$ comme produit d'indicatrices d'évènements disjoints. Pour $i = j$, $X_{1,i}^2 = X_{1,i}$ car $X_{1,i}$ ne peut prendre que la valeur 0 ou 1. On a donc bien dans les deux cas $K_{i,j} = p_i\delta_{i,j} - p_ip_j$ et on conclut en remarquant que T_n et $n^{-1/2}(N_n - np)$ ont même loi. \square

Remarque 19. La loi limite dans ce théorème est un exemple naturel de loi gaussienne sur \mathbb{R}^d sans densité par rapport à λ_d , la mesure de Lebesgue sur \mathbb{R}^d . On montre en effet que

$$\mathbf{P}(Z_1 + Z_2 + \dots + Z_d = 0) = 1 \quad \text{où} \quad Z = (Z_1, \dots, Z_d) \sim \mathfrak{N}(0, K). \quad (22)$$

Pour cela, considérons la forme linéaire

$$u(x) := x_1 + x_2 + \dots + x_d, \quad x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d.$$

La continuité de u et (21) entraînent la convergence en loi de $u(T_n)$ vers $u(Z)$. D'autre part, $u(T_n) = n^{-1/2} \sum_{k=1}^n (u(X_k) - u(p))$ par linéarité de u . Comme $u(X_k) = 1 = u(p)$, $u(T_n)$ est la variable aléatoire nulle, sa loi est la masse de Dirac en 0 et il en est de même pour la loi limite quand n tend vers $+\infty$. Mais dire que $u(Z)$ a pour loi δ_0 revient exactement à (22). Ainsi toute la masse de la loi de Z est portée par l'hyperplan $x_1 + x_2 + \dots + x_d = 0$ qui est de λ_d -mesure nulle et la loi de Z ne peut avoir de densité par rapport à λ_d .

Le théorème 18 a une application incontournable au célèbre test du khi-deux.

Corollaire 20 (Th. de Pearson). *Si le vecteur aléatoire $N_n = (N_{n,1}, \dots, N_{n,d})$ suit la loi multinomiale de paramètres n et $p = (p_1, p_2, \dots, p_d)$,*

$$Q_n := \sum_{i=1}^d \frac{(N_{n,i} - np_i)^2}{np_i} \xrightarrow[n \rightarrow +\infty]{\text{loi}} \chi^2(d-1). \quad (23)$$

Commentaires : Rappelons que $\chi^2(d-1)$ est la loi d'une somme $\sum_{k=1}^{d-1} Y_k^2$, où les Y_k sont des variables aléatoires indépendantes de même loi $\mathfrak{N}(0, 1)$. En observant que Q_n est l'image de $n^{-1/2}(N_n - np)$ par l'application continue $g : \mathbb{R}^d \rightarrow \mathbb{R}^+$, $x \mapsto \sum_{i=1}^d x_i^2/p_i$, (20) donne immédiatement la convergence en loi de Q_n vers $g(Z) = Z_1^2/p_1 + \dots + Z_d^2/p_d$. La partie technique de la preuve est de montrer que la loi de $g(Z)$ est précisément $\chi^2(d-1)$. On pourra trouver une preuve détaillée dans Ouvrard [6, pp. 326–330] ou Revuz [8, p. 166].

Une autre application du théorème 18 est la convergence des lois de dimension finie du processus empirique. Ce processus s'obtient en centrant et en normalisant par \sqrt{n} la fonction de répartition empirique. Il joue un grand rôle en statistique, on le retrouve notamment derrière les tests de Kolomogorov-Smirnov et de Cramér-von Mises. Sa loi limite est celle d'un processus gaussien à temps continu appelé « pont brownien ». La convergence en loi correspondante est une convergence en dimension infinie, dans un espace fonctionnel adéquat. Le premier pas pour prouver une telle convergence est d'établir la convergence des lois de dimension finie.

Corollaire 21 (lois fini-dimensionnelles du processus empirique). Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires indépendantes et de même loi de fonction de répartition F . On note F_n la fonction de répartition empirique

$$F_n(t) := \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k \leq t\}}, \quad t \in \mathbb{R}$$

et ξ_n le processus empirique

$$\xi_n(t) := \sqrt{n}(F_n(t) - F(t)), \quad t \in \mathbb{R}.$$

Pour tout $d \in \mathbb{N}^*$ et tous réels $-\infty < t_1 < t_2 < \dots < t_d < +\infty$,

$$(\xi_n(t_1), \dots, \xi_n(t_d)) \xrightarrow[n \rightarrow +\infty]{\text{loi}} (B(t_1), \dots, B(t_d)), \quad (24)$$

où B est un pont brownien associé à F , c'est-à-dire une processus gaussien centré de covariance

$$\Gamma(s, t) = \text{Cov}(B(s), B(t)) = \min(F(s), F(t)) - F(s)F(t), \quad s, t \in \mathbb{R}. \quad (25)$$

Preuve. Fixons $t_1 < t_2 < \dots < t_d$ et posons $t_0 = -\infty$, $t_{d+1} = +\infty$ avec la convention $F(-\infty) := 0$ et $F(+\infty) := 1$. Il s'agit de montrer la convergence en loi de $(\xi_n(t_1), \dots, \xi_n(t_d))$ vers $\mathfrak{N}(0, K)$ où la matrice de K a pour terme général $K_{i,j} = \Gamma(t_i, t_j)$ ($1 \leq i, j \leq d$). Notons $U_n(t) := nF_n(t)$ le nombre d'observations inférieures ou égales à t dans le n -échantillon X_1, \dots, X_n . Le vecteur

$$V_n := (U_n(t_1), U_n(t_2) - U_n(t_1), \dots, U_n(t_d) - U_n(t_{d-1}), n - U_n(t_d))$$

représente les effectifs de l'échantillon dans les intervalles *disjoints* $]t_{i-1}, t_i]$ ($1 \leq i \leq d+1$). Il suit donc clairement la loi multinomiale de paramètres n et

$$p = (p_1, p_2, \dots, p_d, p_{d+1}) := (F(t_1), F(t_2) - F(t_1), \dots, F(t_d) - F(t_{d-1}), 1 - F(t_d)).$$

Remarquons au passage que

$$\sum_{i=1}^k p_i = F(t_k), \quad 1 \leq k \leq d+1. \quad (26)$$

Par le théorème 18,

$$n^{-1/2}(V_n - np) \xrightarrow[n \rightarrow +\infty]{\text{loi}} Y, \quad Y \sim \mathfrak{N}(0, R), \quad R_{i,j} = p_i \delta_{i,j} - p_i p_j \quad (1 \leq i, j \leq d+1).$$

Considérons alors l'application

$$g : \mathbb{R}^{d+1} \rightarrow \mathbb{R}^d, \quad (x_1, \dots, x_d, x_{d+1}) \mapsto (x_1, x_1 + x_2, \dots, x_1 + \dots + x_d).$$

Par linéarité et continuité de g on voit que

$$g\left(\frac{1}{\sqrt{n}}(V_n - np)\right) = (\xi_n(t_1), \dots, \xi_n(t_d)) \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z := g(Y),$$

où Z est encore un vecteur gaussien centré comme image linéaire d'un vecteur gaussien centré. Pour achever la preuve, il ne reste plus alors qu'à identifier la structure de covariance de Z . Pour $1 \leq k \leq l \leq d$, on obtient en utilisant (26) :

$$\begin{aligned} \text{Cov}(Z_k, Z_l) &= \text{Cov}\left(\sum_{i=1}^k Y_i, \sum_{j=1}^l Y_j\right) = \sum_{i=1}^k \sum_{j=1}^l \text{Cov}(Y_i, Y_j) \\ &= \sum_{i=1}^k \sum_{j=1}^l (p_i \delta_{i,j} - p_i p_j) \\ &= \sum_{i=1}^k p_i - \sum_{i=1}^k \sum_{j=1}^l p_i p_j \\ &= F(t_k) - (p_1 + \dots + p_k)(p_1 + \dots + p_l) \\ &= F(t_k) - F(t_k)F(t_l). \end{aligned}$$

Ainsi $\text{Cov}(Z_k, Z_l) = \Gamma(t_k, t_l)$ et la covariance de Z est bien celle escomptée. \square

4 Compléments

4.1 Théorèmes de limite locale

Le TLC sur \mathbb{R} fait converger la masse d'un intervalle pour la loi de S_n^* vers la masse du même intervalle pour la loi gaussienne standard $\mathfrak{N}(0, 1)$. Les théorèmes de limite locale concernent une convergence uniforme vers la *densité* de $\mathfrak{N}(0, 1)$. La quantité censée converger vers cette densité sera soit la densité de S_n^* lorsqu'elle existe, soit la masse convenablement normalisée des *atomes* de la loi de S_n^* lorsque la loi de X_1 est supportée par un réseau discret de \mathbb{R} . De ce point de vue, la démonstration historique du théorème de Moivre Laplace est le premier exemple de théorème de limite locale. Deux références accessibles pour ces théorèmes sont Feller [4, XV.5, pp. 515–521] et Revuz [8, IV 5,5].

Théorème 22. *Soit $(X_k)_{k \geq 1}$ une suite i.i.d. de variables aléatoires réelles centrées et de variance 1. On suppose que la fonction caractéristique $\varphi(t) = \mathbf{E} \exp(itX_1)$ vérifie*

$$\int_{\mathbb{R}} |\varphi(t)| dt < +\infty.$$

Alors S_n/\sqrt{n} a une densité f_n qui converge uniformément sur \mathbb{R} vers celle de $\mathfrak{N}(0, 1)$.

Théorème 23. *Soit $(X_k)_{k \geq 1}$ une suite i.i.d. de variables aléatoires réelles centrées et de variance 1. On suppose de plus que la loi de X_1 est supportée par un réseau $b+h\mathbb{Z}$ ($b \in \mathbb{R}$*

et $h \in \mathbb{R}^{+*}$ fixés). Dans la suite h désigne le plus grand réel tel que $\mathbf{P}(X_1 \in b + h\mathbb{Z}) = 1$. Les atomes de la loi de S_n/\sqrt{n} sont alors parmi les réels $(nb + kh)/\sqrt{n}$, $k \in \mathbb{Z}$. On note A_n l'ensemble de ces atomes et on définit

$$p_n(x) := \mathbf{P}\left(\frac{S_n}{\sqrt{n}} = x\right), \quad x \in A_n.$$

Alors en notant $f(x) := (2\pi)^{-1/2} \exp(-x^2/2)$,

$$\sup_{x \in A_n} \left| \frac{\sqrt{n}}{h} p_n(x) - f(x) \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Commentaires : La conclusion s'écrit aussi

$$p_n(x) = \frac{hf(x)}{\sqrt{n}} + \frac{\varepsilon_n(x)}{\sqrt{n}}, \quad \sup_{x \in A_n} |\varepsilon_n(x)| \xrightarrow{n \rightarrow +\infty} 0. \quad (27)$$

Examinons l'application de ce résultat à l'approximation gaussienne d'une loi binomiale. Soit donc $(X_k)_{k \geq 1}$ une suite i.i.d. de variables aléatoires de Bernoulli de paramètre p . $S_n := X_1 + \dots + X_n$ suit alors la loi $\text{Bin}(n, p)$. En notant $q = 1 - p$, on a

$$S_n^* = \frac{S_n - \mathbf{E}S_n}{\sqrt{\text{Var } S_n}} = \frac{S_n - np}{\sqrt{npq}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n X'_k = \frac{S'_n}{\sqrt{n}},$$

où $X'_k = (X_k - p)/\sqrt{pq}$ vérifie les hypothèses du théorème 23, que nous appliquons donc à S'_n . Les atomes de la loi de S'_n/\sqrt{n} sont les x tels que $\mathbf{P}(S'_n = x\sqrt{n}) > 0$. En posant $x = y/\sqrt{npq}$, on a

$$\mathbf{P}(S'_n = x\sqrt{n}) = \mathbf{P}(S_n - np = y)$$

et comme les masses ponctuelles de la loi binomiale de S_n sont les entiers $k = 0, 1, 2, \dots, n$, on voit que

$$A_n = \left\{ \frac{k - np}{\sqrt{npq}}; k = 0, 1, 2, \dots, n \right\}.$$

Ainsi $b = -\sqrt{p/q}$, $h = 1/\sqrt{pq}$ et (27) s'écrit

$$\mathbf{P}(S_n = k) = \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k - np)^2}{2npq}\right) + \frac{\varepsilon_{n,k}}{\sqrt{n}}, \quad \max_{0 \leq k \leq n} |\varepsilon_{n,k}| \xrightarrow{n \rightarrow +\infty} 0.$$

Ce résultat est à comparer avec le théorème 31 de l'annexe B qui donne la convergence uniforme de $\varepsilon_{n,k}$ sur le sous ensemble des k tels que $|k - np| \leq c_0\sqrt{npq}$.

4.2 TLC pour les sommes d'un nombre aléatoire de termes

Théorème 24. *Toutes les variables aléatoires considérées étant définies sur le même espace probabilisé, on suppose que*

a) $(X_k)_{k \geq 1}$ est une suite de variables aléatoires telle que

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, \sigma^2), \quad (\sigma > 0).$$

b) $(N_n)_{n \geq 1}$ est une suite de variables aléatoires à valeurs dans \mathbb{N}^* tendant vers $+\infty$ en probabilité, ce qui signifie :

$$\forall A > 0, \quad \lim_{n \rightarrow +\infty} \mathbf{P}(N_n > A) = 1.$$

c) Pour tout $j \geq 1$, N_j est indépendante de la suite $(X_k)_{k \geq 1}$.

Dans ces conditions,

$$\frac{1}{\sqrt{N_n}} \sum_{k=1}^{N_n} X_k \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathfrak{N}(0, \sigma^2).$$

Commentaire : Voici une illustration de ce théorème dans le domaine de l'assurance ¹⁵. N_n représente le nombre de déclarations de sinistres reçues par une compagnie au cours de n unités de temps (disons lors des n jours ouvrables depuis une date origine). X_k est le montant des dédommagements payés par la compagnie pour le k -ième sinistre ainsi déclaré. Le montant total des remboursements pour la période considérée est donc $\sum_{k=1}^{N_n} X_k$. Le théorème 24 peut servir à prévoir une fourchette de remboursements sur une longue période, pour peu que l'on sache estimer σ^2 et la vitesse de convergence vers l'infini de N_n . Pour prouver le théorème 24, il est commode de le voir comme une conséquence immédiate du théorème plus général suivant (cf. Foata-Fuchs [5, p. 250]).

Théorème 25. *Toutes les variables aléatoires considérées étant définies sur le même espace probabilisé, on suppose que*

a) $Y_n \xrightarrow[n \rightarrow +\infty]{\text{loi}} Y.$

b) $N_n \xrightarrow[n \rightarrow +\infty]{Pr} +\infty.$

c) Pour tous $j, n \in \mathbb{N}^*$, Y_j et N_n sont indépendantes.

Alors

$$Y_{N_n} \xrightarrow[n \rightarrow +\infty]{\text{loi}} Y.$$

Preuve. Il suffit de montrer qu'en tout point de continuité x de la fonction de répartition F de Y ,

$$\mathbf{P}(Y_{N_n} \leq x) \xrightarrow[n \rightarrow +\infty]{} F(x). \quad (28)$$

Pour cela, on commence par partitionner suivant les valeurs possibles de N_n pour obtenir

15. On peut appliquer le même modèle pour les remboursements de soins par une caisse de Sécurité Sociale.

grâce à c)

$$\begin{aligned}\mathbf{P}(Y_{N_n} \leq x) &= \sum_{j=1}^{+\infty} \mathbf{P}(Y_{N_n} \leq x, N_n = j) = \sum_{j=1}^{+\infty} \mathbf{P}(Y_j \leq x, N_n = j) \\ &= \sum_{j=1}^{+\infty} \mathbf{P}(Y_j \leq x) \mathbf{P}(N_n = j).\end{aligned}$$

Par a), on a pour tout $\varepsilon \in]0, 1[$ un $j_0 = j_0(\varepsilon, x)$ tel que

$$\forall j > j_0, \quad |\mathbf{P}(Y_j \leq x) - F(x)| \leq \varepsilon.$$

En découpant alors

$$\mathbf{P}(Y_{N_n} \leq x) - F(x) = \left\{ \sum_{j=1}^{j_0} + \sum_{j=j_0+1}^{+\infty} \right\} (\mathbf{P}(Y_j \leq x) - F(x)) \mathbf{P}(N_n = j)$$

et en majorant $|\mathbf{P}(Y_j \leq x) - F(x)|$ par 1 pour $j \leq j_0$ et par ε pour $j > j_0$, on aboutit à

$$|\mathbf{P}(Y_{N_n} \leq x) - F(x)| \leq \mathbf{P}(N_n \leq j_0) + \varepsilon \mathbf{P}(N_n > j_0).$$

En raison de b), les deux termes de ce majorant tendent respectivement vers 0 et vers ε quand n tend vers $+\infty$. On en déduit

$$\forall \varepsilon \in]0, 1[, \quad \limsup_{n \rightarrow +\infty} |\mathbf{P}(Y_{N_n} \leq x) - F(x)| \leq \varepsilon,$$

ce qui donne (28) et termine la preuve. □

Annexe A

Quelques inégalités sur les lois binomiales et de Poisson

Proposition 26. Soient S_n une variable binomiale de paramètres n et p et Z une variable de Poisson de paramètre $\lambda = np$.

(i) Pour tout entier $k \geq 2\lambda - 1$,

$$P(Z > k) < P(Z = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

(ii) Pour tout entier $k \geq 2\lambda + 1$, on a $P(S_n = k) \leq P(Z = k)$ et

$$P(S_n > k) \leq P(Z > k) < P(Z = k).$$

Ces deux résultats sont démontrés dans les exercices A.3 et A.4 ci-dessous. L'exercice A.1 est utile pour la preuve élémentaire (sans fonctions caractéristiques!) du théorème de de Moivre-Laplace. L'exercice A.2 pointe une erreur classique.

Ex A.1 Le terme central de la loi binomiale

Soit $p \in]0, 1[$ et $q = 1 - p$. Dans tout ce qui suit on note :

$$b(k, n, p) = C_n^k p^k q^{n-k} \quad \text{pour } 0 \leq k \leq n,$$

k et n étant entiers. Ainsi $b(k, n, p)$ est la probabilité qu'une variable binomiale de paramètres n et p prenne la valeur k .

1) Vérifier que pour $k \geq 1$, on a :

$$\frac{b(k, n, p)}{b(k-1, n, p)} = 1 + \frac{(n+1)p - k}{kq}.$$

En déduire que :

$$\begin{aligned} \text{si } k \leq (n+1)p, & \quad b(k, n, p) \geq b(k-1, n, p), \\ \text{si } k > (n+1)p, & \quad b(k, n, p) < b(k-1, n, p). \end{aligned}$$

Pour n et p fixés, le maximum de $b(k, n, p)$ est donc atteint en $k = m$, m étant défini comme l'unique entier tel que :

$$(n + 1)p - 1 < m \leq (n + 1)p.$$

Si $(n + 1)p$ n'est pas entier, l'inégalité de droite est stricte et il y a un seul maximum. Si $(n + 1)p$ est un entier on a $m = (n + 1)p$ et $b(m - 1, n, p) = b(m, n, p)$ réalisent tous deux le maximum. Dans tous les cas, le nombre $b(m, n, p)$ est appelé *terme central de la loi binomiale*. On se propose d'en donner un équivalent simple lorsque n tend vers $+\infty$.

2) Etudier les variations de la fonction :

$$f : [0, 1] \rightarrow [0, 1], \quad x \mapsto x^m(1 - x)^{n-m}.$$

3) Vérifier que $np \in](n + 1)p - 1, (n + 1)p[$. Montrer que :

$$\text{si } np \leq m \leq (n + 1)p, \quad b\left(m, n, \frac{m}{n+1}\right) \leq b(m, n, p) \leq b\left(m, n, \frac{m}{n}\right). \quad (\text{A.1})$$

4) Montrer de même que :

$$\text{si } (n + 1)p - 1 < m < np, \quad b\left(m, n, \frac{m+1}{n+1}\right) \leq b(m, n, p) \leq b\left(m, n, \frac{m}{n}\right). \quad (\text{A.2})$$

5) On rappelle la formule de Stirling : pour tout $j \in \mathbb{N}^*$,

$$j! = \sqrt{2\pi}j^{j+1/2}e^{-j}e^{\theta_j} \quad \text{où} \quad \frac{1}{12j+1} < \theta_j < \frac{1}{12j}.$$

En utilisant cette formule, donner un équivalent des bornes des encadrements (A.1) et (A.2). En déduire que :

$$b(m, n, p) \sim \frac{1}{\sqrt{2\pi npq}}, \quad (n \rightarrow +\infty). \quad (\text{A.3})$$

En remplaçant les équivalents par des encadrements, et en se fatiguant un peu plus, on pourrait montrer qu'il existe une constante C indépendante de p et un entier n_0 dépendant de p tels que :

$$\forall n \geq n_0, \quad b(m, n, p) = \frac{1}{\sqrt{2\pi npq}}(1 + \varepsilon_n) \quad \text{avec} \quad |\varepsilon_n| \leq \frac{C}{npq}. \quad (\text{A.4})$$

Ex A.2 Que pensez vous de l'affirmation suivante : « Si on lance un grand nombre (pair) de fois une pièce équilibrée, il y a une forte probabilité d'obtenir exactement autant de piles que de faces » ?

Ex A.3 La queue de la loi de Poisson¹

1) Si X suit la loi de Poisson de paramètre $\lambda > 0$, montrer que :

$$\forall k > \lambda - 1, \quad P(X \geq k) < P(X = k) \frac{k+1}{k+1-\lambda}.$$

Indication : On part de :

$$\sum_{j=k}^{+\infty} \frac{\lambda^j}{j!} = \frac{\lambda^k}{k!} \left(1 + \frac{\lambda}{k+1} + \frac{\lambda^2}{(k+1)(k+2)} + \frac{\lambda^3}{(k+1)(k+2)(k+3)} + \dots \right)$$

et on majore la parenthèse par la somme d'une série géométrique...

2) En déduire que :

$$\forall k \geq 2\lambda - 1, \quad P(X > k) < P(X = k).$$

Ex A.4 Contrôle de l'erreur dans l'approximation poissonnienne

On se propose de donner des résultats quantitatifs sur l'approximation de la probabilité binomiale $b(k, n, p)$ par $e^{-\lambda} \lambda^k / k!$ où $\lambda = np$.

1) Justifier l'encadrement suivant :

$$\forall u \in [0, 1[, \quad \exp\left(-u - \frac{u^2}{2(1-u)}\right) \leq 1 - u \leq \exp(-u). \quad (\text{A.5})$$

Indication : Dans le développement en série entière de $\ln(1-u)$, contrôler le reste de rang 2 par une série géométrique.

2) En déduire que si $0 \leq k \leq n$,

$$C_n^k = \frac{n^k}{k!} \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) \leq \frac{n^k}{k!} \exp\left(-\frac{(k-1)k}{2n}\right).$$

3) En déduire que si $n \geq 2$ et $0 \leq k \leq n$:

$$b(k, n, p) \leq \frac{e^{-\lambda} \lambda^k}{k!} \exp\left(\frac{k}{2n}(2\lambda + 1 - k)\right).$$

En particulier :

$$\forall k \geq 2\lambda + 1, \quad b(k, n, p) \leq \frac{e^{-\lambda} \lambda^k}{k!}.$$

En combinant cette inégalité avec le résultat de l'exercice A.3, on en déduit la majoration suivante de la queue de la loi binomiale :

$$\forall k \geq 2\lambda + 1, \quad \sum_{j=k+1}^n b(j, n, p) \leq \frac{e^{-\lambda} \lambda^k}{k!}. \quad (\text{A.6})$$

4) Montrer en utilisant l'encadrement de la question 1 et la formule de Stirling (cf. exercice A.1) que l'on a pour $0 \leq k < n$ la minoration :

$$b(k, n, p) \geq \frac{e^{-\lambda} \lambda^k}{k!} \exp\left\{ \frac{1}{2n} \left[(2\lambda + 1)k - 2k^2 - \lambda^2 \frac{n-k}{n-\lambda} - \frac{12k+1}{72(n-k)} \right] \right\}.$$

1. La « queue » de la loi d'une v.a. X est la fonction $t \mapsto P(X > t)$.

Annexe B

Approximation gaussienne de la loi binomiale*

Nous connaissons déjà l'approximation d'une loi binomiale $\mathcal{B}(n, p)$ par une loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = np$ lorsque n est « grand » et np « petit ». Nous étudions dans ce chapitre une approximation utilisable lorsque np ne peut être considéré comme « petit ». Le résultat théorique qui justifie cette approximation est le théorème de de Moivre-Laplace qui est lui même un cas particulier du théorème limite central. Ce dernier est, avec la loi des grands nombres, certainement le plus important théorème du calcul des probabilités. L'approximation qui nous intéresse fait intervenir une famille de fonctions appelées densités gaussiennes (ou normales) liées à la célèbre *courbe en cloche* de Gauss.

B.1 La courbe en cloche

Définition 27. Soit $m \in \mathbb{R}$ et $\sigma \in]0, +\infty[$. On appelle densité gaussienne ou normale $f_{m,\sigma}$ sur \mathbb{R} la fonction :

$$f_{m,\sigma} : \mathbb{R} \longrightarrow \mathbb{R}^+ \quad t \longmapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right)$$

La fonction $f_{0,1}$ est appelée densité normale ou gaussienne standard.

Dans cette famille de fonctions, m est un paramètre de localisation ou de position (c'est la valeur où $f_{m,\sigma}$ atteint son maximum). Le paramètre σ est un paramètre d'échelle, il caractérise l'aplatissement de la courbe. Les courbes représentatives $C_{m,\sigma}$ de ces fonctions se déduisent toutes de la courbe $C_{0,1}$ par translations et changements d'échelle. On passe de $C_{0,1}$ à $C_{m,\sigma}$ par la transformation : $(x, y) \mapsto (\sigma x + m, \frac{y}{\sigma})$ et de $C_{m,\sigma}$ à $C_{0,1}$ par $(x, y) \mapsto (\frac{x-m}{\sigma}, \sigma y)$. La courbe $C_{0,1}$ (figure B.1) est très populaire sous le nom de courbe en cloche de Gauss.

Bien que $f_{0,1}(t)$ soit strictement positif pour tout réel t , la densité $f_{0,1}$ semble être à support dans $[-4, +4]$. Cela provient de la décroissance rapide de $\exp(-\frac{t^2}{2})$ quand t tend vers $+\infty$ et des limites de résolution du dessin.

*. Reproduction du chapitre 7 de *Introduction au Calcul des Probabilités*, cours de Deug, Ch. Suquet.

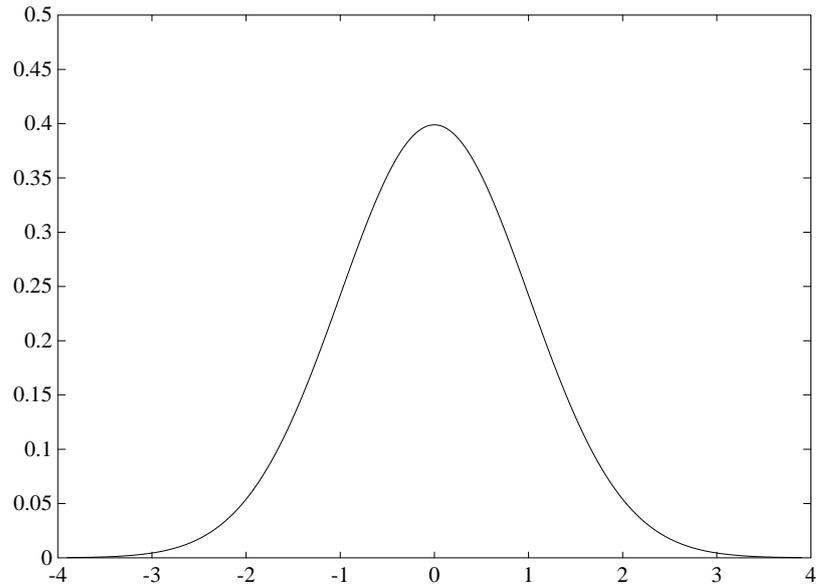


FIGURE B.1 – Densité $f_{0,1}$

L'influence des paramètres m et σ est illustrée par les figures B.2 et B.3.

Une propriété importante de $C_{0,1}$ est que l'aire qu'elle délimite avec l'axe des abscisses vaut 1 :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt = 1. \quad (\text{B.1})$$

L'exercice B.1 présente une démonstration de cette relation. Les autres fonctions $f_{m,\sigma}$ ont aussi une intégrale généralisée sur \mathbb{R} qui vaut 1. Pour le voir, soient $a < b$ deux réels quelconques. Par le changement de variable $u = (t - m)/\sigma$:

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right) dt = \int_{a^*}^{b^*} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du, \quad (\text{B.2})$$

où les nouvelles bornes sont :

$$a^* = \frac{a-m}{\sigma} \quad \text{et} \quad b^* = \frac{b-m}{\sigma}.$$

Lorsque a et b tendent respectivement vers $-\infty$ et $+\infty$, il en est de même pour a^* et b^* . On en déduit :

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right) dt = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = 1.$$

L'aire délimitée par $C_{m,\sigma}$, l'axe des abscisses et les droites d'équation $t = a$, $t = b$ joue un rôle important dans l'approximation des probabilités binomiales $P(a < S_n \leq b)$. Par

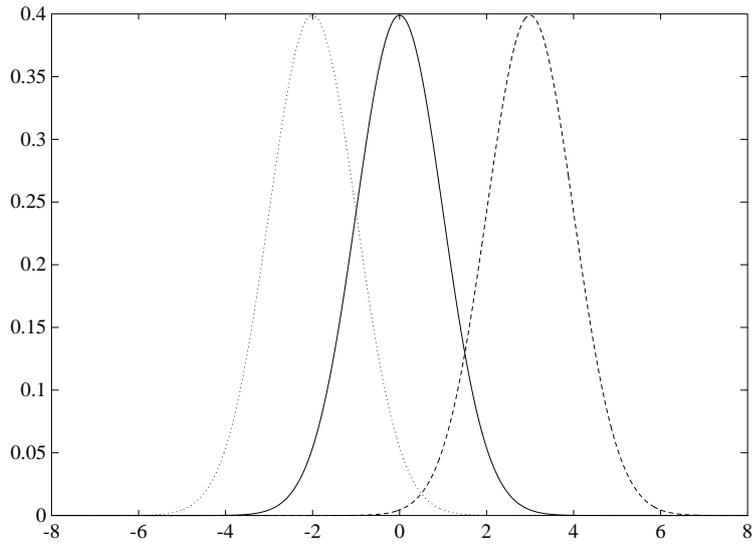


FIGURE B.2 – Densités $f_{-2,1}$, $f_{0,1}$, $f_{3,1}$

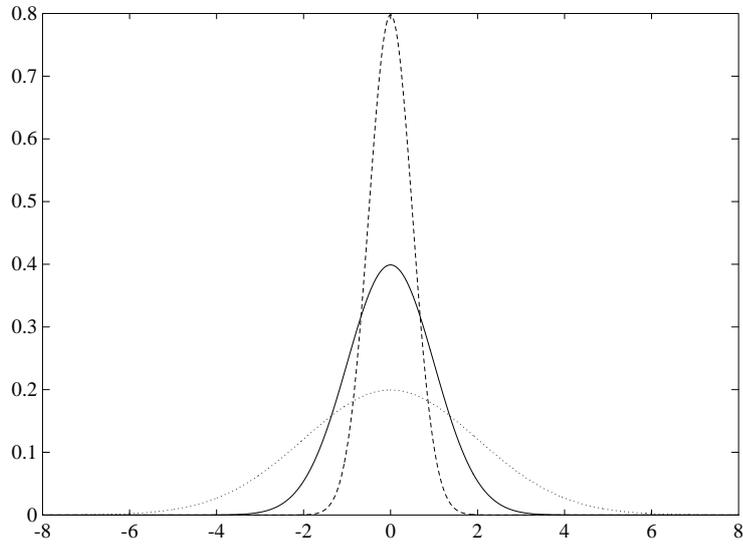


FIGURE B.3 – Densités $f_{0,1}$, $f_{0,0.5}$, $f_{0,2}$

changement de variable, le calcul de cette aire se ramène à celui de l'aire correspondante pour $C_{0,1}$ avec a^* et b^* à la place de a et b . Elle peut donc s'écrire sous la forme $\Phi(b^*) - \Phi(a^*)$ où la fonction Φ est définie par :

$$\forall x \in \mathbb{R}, \quad \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt. \quad (\text{B.3})$$

Cette intégrale ne peut s'exprimer à l'aide des fonctions usuelles. On peut en calculer une valeur approchée avec toute précision souhaitée (voir exercice B.2). La figure B.4 représente le graphe de Φ .

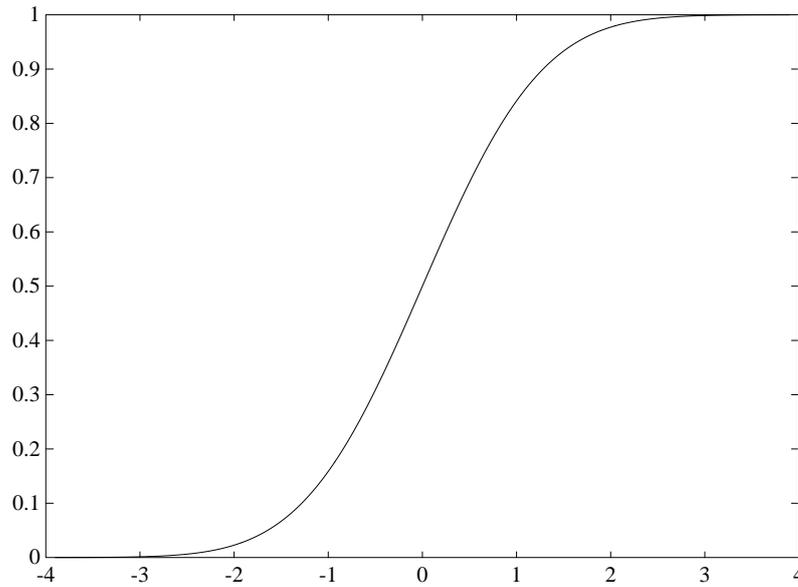


FIGURE B.4 – Fonction $\Phi : x \mapsto \int_{-\infty}^x f_{0,1}(t) dt$

La table en annexe donne les valeurs de $\Phi(x)$ par *pas* de 0.01 pour x compris entre 0 et 3 et quelques valeurs pour x compris entre 3 et 4.5. Pour x négatif, la parité de la densité entraîne la relation $\Phi(x) = 1 - \Phi(-x)$. Pour les « très grandes valeurs de x », (i.e. $|x| \geq 4$), on dispose du résultat suivant qui donne une évaluation de la « queue » de la loi normale :

Lemme 28. *Pour tout $x > 0$, on a l'encadrement :*

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) < 1 - \Phi(x) < \frac{1}{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \quad (\text{B.4})$$

Preuve. La dérivation de chacun des 3 membres A , B et C de (B.4) donne :

$$A'(x) = \left(-1 + \frac{3}{x^4}\right) \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad B'(x) = -\frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad C'(x) = \left(-1 - \frac{1}{x^2}\right) \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

Il suffit alors d'intégrer sur $[x, +\infty[$ l'encadrement suivant vrai pour tout y :

$$\left(-1 - \frac{1}{y^2}\right) \frac{e^{-y^2/2}}{\sqrt{2\pi}} < -\frac{e^{-y^2/2}}{\sqrt{2\pi}} < \left(-1 + \frac{3}{y^4}\right) \frac{e^{-y^2/2}}{\sqrt{2\pi}},$$

pour obtenir (B.4). □

B.2 Étude graphique

Soient X_1, \dots, X_n n v.a. de Bernoulli indépendantes de même paramètre p . On pose $S_n = X_1 + \dots + X_n$. La loi de S_n est alors une binomiale $B(n, p)$:

$$P(S_n = k) = C_n^k p^k (1-p)^{n-k} \quad (0 \leq k \leq n)$$

Les histogrammes¹ ci-dessous représentent cette loi pour différentes valeurs des paramètres n et p . Bien que l'histogramme de la loi $B(n, p)$ soit constitué théoriquement de $n + 1$ rectangles, seule une partie d'entre eux est visible sur le dessin, les autres correspondant à des probabilités trop petites. Le nombre ε représente pour chaque figure un majorant de la probabilité correspondant à la réunion de ces rectangles « invisibles ». Sur chaque figure, la courbe en pointillés représente la densité $f_{m,\sigma}$ dont les paramètres sont donnés par : $m = \mathbf{E}S_n = np$ et $\sigma^2 = \text{Var } S_n = npq$.

1. Les rectangles de l'histogramme ont une aire proportionnelle aux nombres $P(S_n = k)$ et ont pour axes de symétrie les droites d'équation $x = k$ correspondantes.

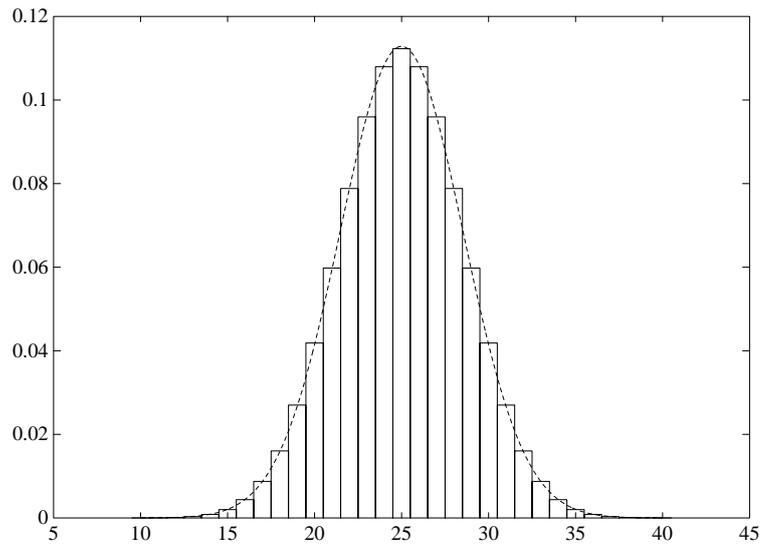


FIGURE B.5 – Binomiale $n = 50$, $p = 0,5$ ($\varepsilon = 3,1 \times 10^{-4}$)

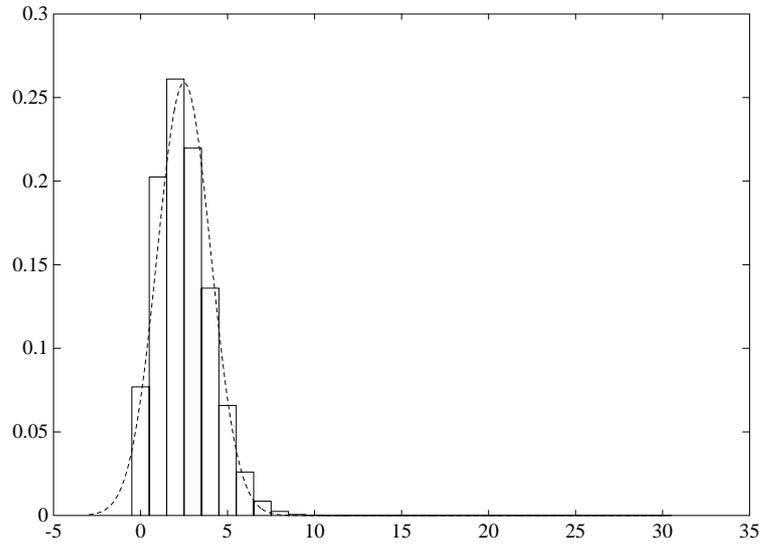


FIGURE B.6 – Binomiale $n = 50$, $p = 0,05$ ($\varepsilon = 1,6 \times 10^{-4}$)

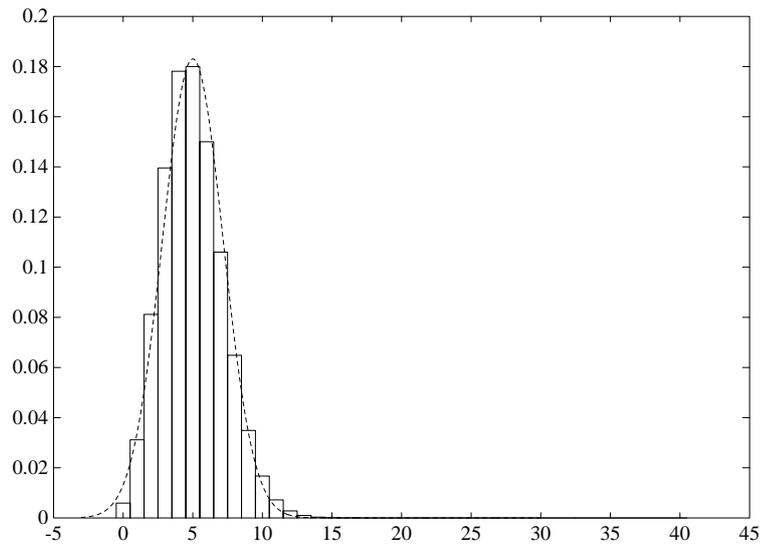


FIGURE B.7 – Binomiale $n = 100$, $p = 0,05$ ($\varepsilon = 1,4 \times 10^{-4}$)

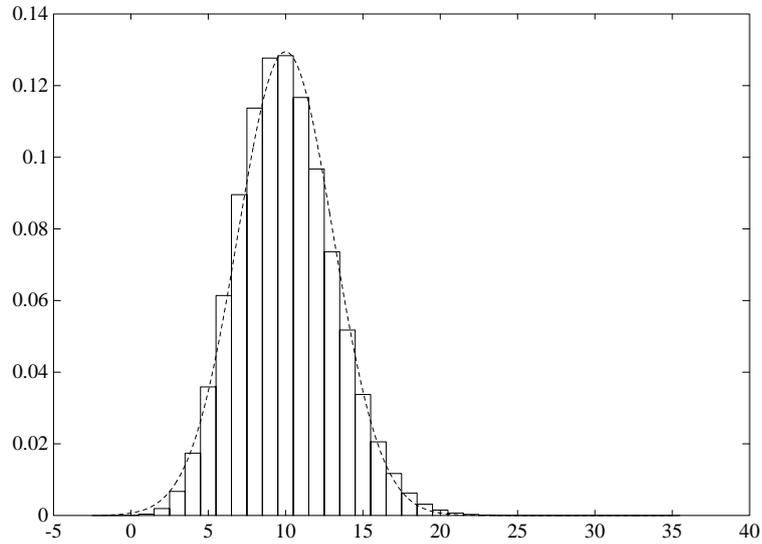


FIGURE B.8 – Binomiale $n = 200$, $p = 0,05$ ($\varepsilon = 2,3 \times 10^{-4}$)

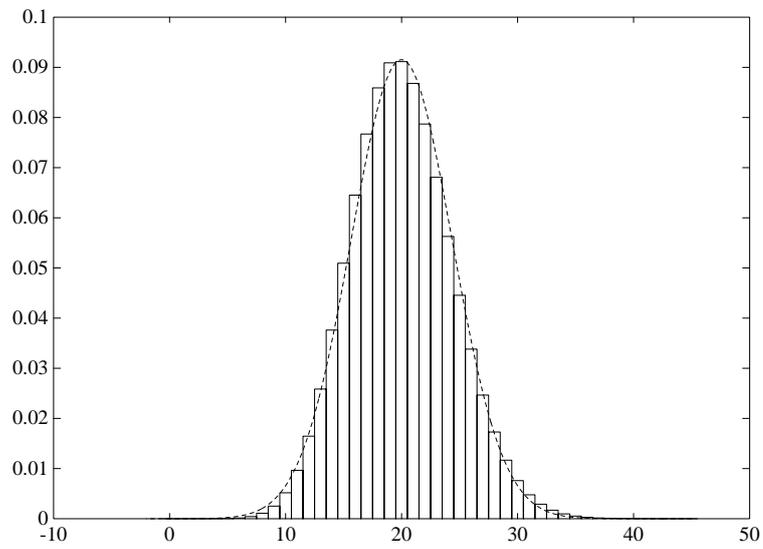


FIGURE B.9 – Binomiale $n = 400$, $p = 0,05$ ($\varepsilon = 4,9 \times 10^{-4}$)

B.3 Le théorème de de Moivre-Laplace

Ces figures laissent entrevoir la possibilité d'approximer une binomiale en un sens que nous précisons maintenant. Il est commode pour cela d'effectuer un changement d'origine de façon à centrer les variables aléatoires et un changement d'échelle de façon à ce qu'elles aient toutes pour variance 1. La nouvelle v.a. obtenue par ce procédé est :

$$S_n^* = \frac{S_n - \mathbf{E}(S_n)}{\sigma(S_n)} = \frac{S_n - np}{\sqrt{npq}} \quad (\text{B.5})$$

Théorème 29 (de Moivre-Laplace). *Soit S_n une variable aléatoire de loi binomiale de paramètres n et p et S_n^* définie par (B.5). Pour tous réels $c < d$ fixés :*

$$\lim_{n \rightarrow +\infty} P(c < S_n^* \leq d) = \Phi(d) - \Phi(c) = \int_c^d \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

On a la même limite pour $P(c \leq S_n^* \leq d)$, $P(c \leq S_n^* < d)$ et $P(c < S_n^* < d)$. De Moivre a donné la première version de ce théorème² en 1733. Laplace (1812) prouva plus tard le même résultat par une méthode différente en obtenant une évaluation de la vitesse de convergence. La démonstration du théorème (cf. section B.4) repose sur un encadrement adéquat des coefficients binomiaux et la formule de Stirling. Elle pourra sans inconvénient être passée en première lecture, l'important étant plutôt de comprendre la signification et l'utilisation pratique du théorème de de Moivre-Laplace. Comme dans le cas de l'approximation par une loi de Poisson, ce théorème permet l'approximation de $P(a < S_n \leq b)$ pour les grandes valeurs de n . Pour cela il suffit de remarquer que la fonction $x \mapsto (x - np)/\sqrt{npq}$ étant croissante, on a l'équivalence :

$$a < S_n(\omega) \leq b \Leftrightarrow \frac{a - np}{\sqrt{npq}} < \frac{S_n(\omega) - np}{\sqrt{npq}} = S_n^*(\omega) \leq \frac{b - np}{\sqrt{npq}}.$$

On en déduit que pour tout $n \geq 1$

$$P(a < S_n \leq b) = P\left(\frac{a - np}{\sqrt{npq}} < S_n^* \leq \frac{b - np}{\sqrt{npq}}\right)$$

Lorsque n est « grand » le théorème de de Moivre-Laplace nous dit que le second membre peut être approximé par :

$$P\left(\frac{a - np}{\sqrt{npq}} < S_n^* \leq \frac{b - np}{\sqrt{npq}}\right) \simeq \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right).$$

Il suffit alors d'utiliser la table des valeurs de Φ pour calculer cette probabilité.

Pour que l'approximation soit légitime, il convient de savoir majorer l'erreur commise en fonction de n , p (et aussi de a et b). Une telle majoration sera donnée sans démonstration section B.5. L'idée générale est que la vitesse de convergence dans le théorème de de

2. Publiée en 1738 dans la seconde édition de son livre *The Doctrine of Chances*.

Moivre-Laplace est, comme pour l'inégalité de Tchebycheff, en $O(n^{-1/2})$. La constante sous-entendue dans le O est d'autant meilleure que p est proche de $1/2$ (et se dégrade fortement quand p est proche de 0 ou 1). Cette dépendance par rapport à p est illustrée par les figures B.5 à B.9.

Exemple 1. On lance 3 600 fois un dé. Evaluer la probabilité que le nombre d'apparitions du 1 soit compris strictement entre 540 et 660.

Soit S le nombre d'apparitions du 1. Cette variable aléatoire suit la loi binomiale $\mathcal{B}(3\,600, 1/6)$. On a $\mathbf{E}S = 600$ et $\text{Var } S = 500$. En notant $S^* = (S - \mathbf{E}S)/\sigma(S)$ la variable centrée réduite associée à S :

$$P(540 < S < 660) = P\left(\frac{540 - 600}{\sqrt{500}} < S^* < \frac{660 - 600}{\sqrt{500}}\right).$$

Comme n est grand on peut utiliser l'approximation liée au théorème de Moivre-Laplace :

$$P\left(\frac{-60}{10\sqrt{5}} < S^* < \frac{60}{10\sqrt{5}}\right) \simeq \Phi\left(\frac{6}{\sqrt{5}}\right) - \Phi\left(\frac{-6}{\sqrt{5}}\right).$$

En utilisant la parité de la densité $f_{0,1}$, on a pour tout $a > 0$: $\Phi(a) - \Phi(-a) = 2\Phi(a) - 1$. En utilisant la table des valeurs de Φ on obtient donc :

$$P(540 < S < 660) \simeq 2\Phi(2,68) - 1 \simeq 0,992\,6 \simeq 0,99.$$

(en raison de l'approximation de $6/\sqrt{5}$ par 2,68, il n'y a pas lieu de conserver les deux derniers chiffres).

Il est intéressant de comparer ce résultat avec ce qu'aurait donné l'inégalité de Tchebycheff :

$$P(540 < S < 660) \geq 1 - \frac{\text{Var } S}{60^2} = 1 - \frac{500}{60^2} = \frac{31}{36} > 0,86.$$

Pour être honnête, il convient néanmoins de noter que les deux résultats sont de nature différente. Le deuxième est une minoration dont on est certain, pas une approximation. Pour pouvoir affirmer que le théorème de Moivre-Laplace nous donne ici un meilleur résultat que l'inégalité de Tchebycheff, il faut donc vérifier que l'erreur d'approximation est inférieure à $0,99 - 0,86 = 0,13$. (À suivre...)

Exemple 2. Une entreprise emploie 500 personnes qui déjeunent à la cantine à l'un ou l'autre des deux services avec une probabilité égale de manger au premier ou au second. Si le gérant veut avoir une probabilité supérieure à 95% de disposer d'assez de couverts, combien devra-t-il en prévoir pour chacun des deux services ?

Voici une modélisation du problème. On numérote les 500 personnes par ordre alphabétique³. On note X_i la variable aléatoire valant 1 si la i -ème personne de la liste mange au premier service, 0 si elle mange au deuxième. Les X_i sont des variables de Bernoulli

3. Ou tout autre ordre *fixé à l'avance*.

de même loi : $P(X_i = 0) = P(X_i = 1) = 1/2$. On suppose les X_i indépendantes, hypothèse raisonnable dans le cas d'une numérotation par ordre alphabétique⁴. Bien que l'énoncé ne le précise pas il s'agit évidemment de trouver le nombre *minimum* de couverts à disposer à chaque service, sinon avec 500 couverts à chaque service, le gérant ne prendrait aucun risque! Notons k ce nombre minimum. Le nombre (aléatoire) de personnes mangeant au premier service est : $S_n = \sum_{i=1}^{500} X_i$ (avec $n = 500$). Le nombre de personnes mangeant au deuxième service est par conséquent $500 - S_n$ (on suppose que tout le monde mange). Le problème revient donc à trouver k minimal tel que :

$$P(S_n \leq k \text{ et } 500 - S_n \leq k) \geq 0,95.$$

Ce qui s'écrit encore :

$$P(500 - k \leq S_n \leq k) \geq 0,95. \quad (\text{B.6})$$

La loi de S_n est la binomiale $\mathcal{B}(500, 1/2)$, mais comme n est grand, il est légitime d'utiliser l'approximation de cette loi par le théorème de de Moivre-Laplace. Pour ce faire, on normalise S_n en la centrant puis en divisant par l'écart type. L'espérance et la variance de S_n sont $\mathbf{E}S_n = n\frac{1}{2} = 250$ et $\text{Var } S_n = n(\frac{1}{2})(1 - \frac{1}{2}) = 125$. On peut écrire (B.6) sous la forme équivalente :

$$P\left(\frac{500 - k - 250}{\sqrt{125}} \leq S_n^* \leq \frac{k - 250}{\sqrt{125}}\right) \geq 0,95.$$

où $S_n^* = (S_n - 250)/\sqrt{125}$. En négligeant l'erreur d'approximation, le problème revient à trouver k minimal tel que :

$$\Phi\left(\frac{k - 250}{\sqrt{125}}\right) - \Phi\left(\frac{250 - k}{\sqrt{125}}\right) \geq 0,95.$$

En utilisant comme dans l'exemple précédent la relation $\Phi(a) - \Phi(-a) = 2\Phi(a) - 1$ avec ici $a = (k - 250)/\sqrt{125}$, on est ramené à la résolution de l'inéquation : $2\Phi(a) - 1 \geq 0,95$ équivalente à $\Phi(a) \geq 0,975$. La table nous donne $\Phi(1,96) = 0,9750$. On prendra donc pour k le plus petit entier tel que $(k - 250)/\sqrt{125} \geq 1,96$ d'où : $k \geq 250 + 1,96\sqrt{125} \approx 271,91$. Ainsi il suffit de prévoir 272 couverts par service pour avoir une probabilité supérieure à 0,95 que chacun puisse manger au service de son choix. On constate qu'en acceptant un risque faible de surcharge de la cantine, il y a moyen de réaliser une économie considérable en place et en mobilier.

B.4 Preuve du théorème de de Moivre-Laplace

Nous donnons dans cette section une démonstration du théorème de de Moivre-Laplace (cas général : p quelconque dans $]0, 1[$). Les techniques utilisées relèvent uniquement du programme d'analyse du DEUG. Signalons qu'il existe une démonstration beaucoup plus rapide⁵ basée sur la transformation de Fourier (cf. cours de Licence).

4. Ce ne serait bien sûr pas le cas si on les avait numérotées d'après leur ordre d'arrivée à la cantine, mais cet ordre est lui même aléatoire et nous l'avons exclu.

5. Mais moins instructive.

La preuve que nous proposons n'est pas la démonstration classique. Elle s'inspire de Feller⁶. En fait nous allons faire un peu plus que prouver le théorème de de Moivre-Laplace. Nous allons montrer l'existence d'un entier $n_0 = n_0(p)$ et d'une constante C tels que :

$$\forall n \geq n_0, \quad |P(c \leq S_n^* \leq d) - (\Phi(d) - \Phi(c))| \leq \frac{C}{\sqrt{npq}}, \quad (\text{B.7})$$

ce qui prouvera le théorème tout en nous donnant une idée de la vitesse de convergence⁷. Nous notons dans la suite :

$$b(k, n, p) = C_n^k p^k q^{n-k}.$$

Il s'agit d'étudier le comportement asymptotique de :

$$P(c \leq S_n^* \leq d) = \sum_{c \leq x_k \leq d} b(k, n, p), \quad \text{où } x_k = \frac{k - np}{\sqrt{npq}}. \quad (\text{B.8})$$

Remarquons que dans cette somme, le nombre de termes (en gros $(d - c)\sqrt{npq}$) tend vers l'infini avec n , donc k peut très bien tendre vers l'infini avec n , mais pas n'importe comment. Il doit vérifier une condition du type :

$$\left| \frac{k - np}{\sqrt{npq}} \right| \leq c_0 = \max(|c|, |d|). \quad (\text{B.9})$$

La démonstration se fait en deux temps :

- On cherche un encadrement de $b(k, n, p)$ permettant d'obtenir un équivalent de la forme $(2\pi npq)^{-1/2} \exp(-x_k^2/2)$ lorsque n tend vers $+\infty$ et k reste astreint à la condition (B.9).
- On traite alors le second membre de (B.8) comme une somme de Riemann que l'on approxime par l'intégrale $\Phi(d) - \Phi(c)$ en contrôlant l'erreur commise.

B.4.1 Évaluation asymptotique de $b(k, n, p)$

Notre point de départ est l'évaluation asymptotique du terme central de la loi binomiale pour laquelle nous renvoyons à l'exercice A.1 p. 31 dont nous rappelons les résultats :

Le maximum de $b(k, n, p)$ à n et p fixés est atteint pour $k = m$ défini comme l'unique entier tel que :

$$(n + 1)p - 1 < m \leq (n + 1)p.$$

Il existe un entier $n_1 = n_1(p)$ et une constante c_1 tels que

$$\forall n \geq n_1, \quad b(m, n, p) = \frac{1}{\sqrt{2\pi npq}}(1 + \varepsilon_n) \quad \text{avec } |\varepsilon_n| \leq \frac{c_1}{npq}. \quad (\text{B.10})$$

6. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I, ex. 19–21 p. 182 (2^e édition).

7. On ne cherchera pas à optimiser la valeur de C , ni même à l'explicitier.

On raccroche l'évaluation asymptotique de $b(k, n, p)$ à celle de $b(m, n, p)$ en écrivant :

$$\text{si } k > m, \quad b(k, n, p) = b(m, n, p) \prod_{j=m+1}^k \frac{b(j, n, p)}{b(j-1, n, p)}, \quad (\text{B.11})$$

$$\text{si } k < m, \quad b(k, n, p) = b(m, n, p) \prod_{j=k+1}^m \frac{b(j-1, n, p)}{b(j, n, p)}. \quad (\text{B.12})$$

Nous sommes ainsi conduits à la recherche d'un encadrement du quotient :

$$\frac{b(j, n, p)}{b(j-1, n, p)} = \frac{(n+1-j)p}{jq}. \quad (\text{B.13})$$

Pour des raisons de simplification calculatoire, on « centre » j par $(n+1)p$ et on « normalise » en divisant par $\sqrt{(n+1)pq}$. Pour cela, posons :

$$j = (n+1)p + \delta_j \quad \text{et} \quad (n+1)pq = \tau_n^2 \quad (\tau > 0).$$

Avec ces notations, on a :

$$(n+1-j)p = (n+1)p - (n+1)p^2 - p\delta_j = (n+1)p(1-p) - p\delta_j = \tau_n^2 - p\delta_j.$$

En reportant dans (B.13), il vient :

$$\frac{b(j, n, p)}{b(j-1, n, p)} = \frac{\tau_n^2 - p\delta_j}{\tau_n^2 + q\delta_j} = \frac{1 - pu_j}{1 + qu_j} \quad \text{avec} \quad u_j = \frac{\delta_j}{\tau_n^2}. \quad (\text{B.14})$$

Examinons d'abord le cas où $k > m$: on a alors $j > m$ et $\delta_j > 0$ donc $u_j > 0$. Nous pouvons alors utiliser avec $x = qu_j$ l'encadrement suivant :

$$\forall x > 0, \quad 1 - x < \frac{1}{1+x} < 1 - x + x^2.$$

En effet, pour $x \geq 1$, cet encadrement est trivial, pour $0 < x < 1$, on utilise le développement $(1+x)^{-1} = 1 - x + x^2 - x^3 + \dots$ en série alternée dont le terme général décroît en valeur absolue. La somme de la série est donc encadrée par deux sommes partielles consécutives.

Minoration du quotient :

$$\frac{1 - pu_j}{1 + qu_j} > (1 - pu_j)(1 - qu_j) = 1 - (p+q)u_j + pq u_j^2 > 1 - u_j.$$

Majoration du quotient :

$$\begin{aligned} \frac{1 - pu_j}{1 + qu_j} &< (1 - pu_j)(1 - qu_j + q^2 u_j^2) \\ &= 1 - (p+q)u_j + (pq + q^2)u_j^2 - pq^2 u_j^3 \\ &< 1 - u_j + qu_j^2 \\ &< 1 - u_j + u_j^2. \end{aligned}$$

Nous obtenons ainsi l'encadrement :

$$\text{Si } j > m, \quad 1 - u_j < \frac{b(j, n, p)}{b(j-1, n, p)} < 1 - u_j + u_j^2. \quad (\text{B.15})$$

Dans le cas où $k < m$, on s'intéresse au quotient inverse :

$$\frac{b(j-1, n, p)}{b(j, n, p)} = \frac{1 + qu_j}{1 - pu_j}.$$

Cette fois u_j est négatif. En échangeant les rôles de p et q et en remplaçant u_j par $-u_j$, on est donc ramené au cas précédent, d'où :

$$\text{Si } j < m, \quad 1 + u_j < \frac{b(j-1, n, p)}{b(j, n, p)} < 1 + u_j + u_j^2. \quad (\text{B.16})$$

L'étape suivante est d'encadrer les produits figurant dans (B.11) et (B.12). Pour cela, il est commode d'utiliser le lemme suivant qui procure un encadrement de chaque facteur par des exponentielles.

Lemme 30.

$$\forall t \in]0, 1[, \quad \exp\left(-t - \frac{t^2}{2(1-t)}\right) < 1 - t < \exp(-t).$$

Preuve. L'inégalité de droite provient de la convexité de la fonction $t \mapsto e^{-t}$: la courbe représentative est toujours au dessus de sa tangente au point d'abscisse 0 (on peut aussi considérer le développement en série entière de e^{-t} et remarquer que pour $0 < t < 1$, il s'agit d'une série alternée dont le terme général décroît en valeur absolue...).

Pour l'inégalité de gauche, on utilise le développement en série entière :

$$\forall t \in]0, 1[, \quad \ln(1-t) = -t - \frac{t^2}{2} - \frac{t^3}{3} - \dots - \frac{t^k}{k} - \dots = -t - \sum_{k=2}^{+\infty} \frac{t^k}{k}.$$

En minorant chaque dénominateur par 2 dans la série, on exploite la formule explicite pour la somme d'une série géométrique pour obtenir :

$$\sum_{k=2}^{+\infty} \frac{t^k}{k} < \frac{t^2}{2}(1 + t + t^2 + \dots) = \frac{t^2}{2(1-t)}.$$

On en déduit :

$$\forall t \in]0, 1[, \quad \ln(1-t) > -t - \frac{t^2}{2(1-t)},$$

et il ne reste plus qu'à prendre l'exponentielle des deux membres. □

Notons maintenant que d'après (B.9), u_j tend vers 0 comme $n^{-1/2}$ lorsque n tend vers $+\infty$ (et même uniformément par rapport à j) : on peut trouver un entier $n_2 = n_2(p, c, d)$ tel que :

$$\forall n \geq n_2, \quad 0 \leq |u_j| \leq \frac{1}{2},$$

pour tous les j concernés par notre démonstration. Si $t \in [0, 1/2]$, $1/(1-t) \leq 2$. Combinée à cette remarque, l'application du lemme 30 avec $t = u_j$ nous fournit :

$$\forall n \geq n_2, \quad \forall j > m, \quad \exp(-u_j - u_j^2) < 1 - u_j.$$

Le même lemme avec $t = u_j - u_j^2$ nous donne :

$$\forall n \geq n_2, \quad \forall j > m, \quad 1 - u_j + u_j^2 < \exp(-u_j + u_j^2).$$

Ces deux inégalités nous permettent de remplacer les bornes de l'encadrement (B.15) par des exponentielles. On en déduit pour tout $n \geq n_2$ et tout $k > m$:

$$\exp \left[- \sum_{j=m+1}^k u_j - \sum_{j=m+1}^k u_j^2 \right] < \prod_{j=m+1}^k \frac{b(j, n, p)}{b(j-1, n, p)} < \exp \left[- \sum_{j=m+1}^k u_j + \sum_{j=m+1}^k u_j^2 \right].$$

Evaluation de la somme des u_j

La somme des termes d'une progression arithmétique est égale au nombre de termes multiplié par la moyenne du premier et du dernier. En appliquant cette règle, puis en centrant par rapport à $(n+1)p$, nous obtenons :

$$\begin{aligned} \sum_{j=m+1}^k u_j &= \frac{1}{\tau_n^2} \sum_{j=m+1}^k (j - (n+1)p) \\ &= \frac{1}{2\tau_n^2} (k - m)(k + m + 1 - 2(n+1)p) \\ &= \frac{1}{2\tau_n^2} \{ (k - (n+1)p) - (m - (n+1)p) \} \\ &\quad \times \{ (k - (n+1)p) + (m - (n+1)p) + 1 \} \\ &= \frac{(k - (n+1)p)^2 - (m - (n+1)p)^2}{2\tau_n^2} + \frac{k - m}{2\tau_n^2} \\ &= \frac{(k - (n+1)p)^2}{2\tau_n^2} - \frac{(m - (n+1)p)^2}{2\tau_n^2} + \frac{k - m}{2\tau_n^2}. \end{aligned} \tag{B.17}$$

Comme $(m - (n+1)p)^2$ est majoré par 1, la somme des deux derniers termes de (B.17) est comprise entre $(k-1-m)/(2\tau_n^2)$ et $(k-m)/(2\tau_n^2)$. Il existe donc une constante c_2 (par exemple $c_2 = (1 + c_0)/2$) telle que :

$$\frac{(k - (n+1)p)^2}{2\tau_n^2} - \frac{c_2}{\tau_n} \leq \sum_{j=m+1}^k u_j \leq \frac{(k - (n+1)p)^2}{2\tau_n^2} + \frac{c_2}{\tau_n}, \tag{B.18}$$

pour tout $n \geq n_2$ et tous les $k > m$ vérifiant (B.9).

Majoration de la somme des u_j^2

A la différence de la somme des u_j qui est le terme principal, celle des u_j^2 est un terme d'erreur. Comme nous voulons seulement avoir une idée de la vitesse de convergence dans le théorème de de Moivre-Laplace, une majoration grossière nous suffit. On l'obtient en majorant chaque terme de la somme par le plus grand d'entre eux :

$$\sum_{j=m+1}^k u_j^2 = \frac{1}{\tau_n^4} \sum_{j=m+1}^k (j - (n+1)p)^2 \leq \frac{(k-m)^3}{\tau_n^4}.$$

Il existe une constante c_3 (par exemple $c_3 = (1+c_0)^3$) telle que pour tout $n \geq n_2$ et tous les $k > m$ vérifiant (B.9) :

$$0 \leq \sum_{j=m+1}^k u_j^2 \leq \frac{c_3}{\tau_n}. \quad (\text{B.19})$$

Posons pour alléger les écritures :

$$t_k = \frac{(k - (n+1)p)^2}{\tau_n}.$$

De (B.18) et (B.19), on tire :

$$\exp\left(-\frac{t_k^2}{2} - \frac{c_2 + c_3}{\tau_n}\right) \leq \prod_{j=m+1}^k \frac{b(j, n, p)}{b(j-1, n, p)} \leq \exp\left(-\frac{t_k^2}{2} + \frac{c_2 + c_3}{\tau_n}\right). \quad (\text{B.20})$$

En combinant (B.10), (B.11) et (B.20), on vérifie facilement qu'il existe une constante c_4 et un entier $n_3 = n_3(p, c, d)$ tels que pour tout $n \geq n_3$ et tous les $k > m$ vérifiant (B.9) :

$$b(k, n, p) = \frac{e^{-t_k^2/2}}{\sqrt{2\pi npq}} (1 + \varepsilon'_{n,k}), \quad \text{avec } |\varepsilon'_{n,k}| \leq \frac{c_4}{\tau_n}. \quad (\text{B.21})$$

Pour le cas $j < m$, u_j est négatif et on a les mêmes estimations que pour $j > m$ en remplaçant u_j par $-u_j$ (en prenant garde que dans l'évaluation de la somme des u_j , le nombre de termes est $(m-k)$ au lieu de $(k-m)$). On vérifie ainsi facilement⁸ que (B.21) reste valable pour $k < m$.

Avant de passer à la suite de notre programme, il convient de mettre le résultat obtenu sous une forme plus commode en revenant à la normalisation et au centrage traditionnels. Posons :

$$\sigma_n = \sqrt{npq} \quad \text{et} \quad x_k = \frac{k - np}{\sigma_n}. \quad (\text{B.22})$$

Notons que $1/\tau_n \leq 1/\sigma_n$. On vérifiera à titre d'exercice que :

$$\exp(-t_k^2/2) = \exp(-x_k^2/2)(1 + \varepsilon''_{n,k}) \quad \text{avec } |\varepsilon''_{n,k}| \leq \frac{c_5}{\sigma_n},$$

pour $n \geq n_4$ et tous les k vérifiant (B.9). Le résultat de cette partie peut donc s'énoncer :

8. Pour les courageux et les sceptiques pas encore épuisés...s'il en reste.

Théorème 31.

Avec les notations (B.22) ci-dessus, il existe une constante A et un entier $n_0 = n_0(p, c_0)$ tels que pour tout $n \geq n_0$ et tout k tel que $|k - np| \leq c_0 \sqrt{npq}$:

$$b(k, n, p) = \frac{\exp(-x_k^2/2)}{\sigma_n \sqrt{2\pi}} (1 + \varepsilon_{n,k}) \quad \text{avec} \quad |\varepsilon_{n,k}| \leq \frac{A}{\sigma_n}. \quad (\text{B.23})$$

B.4.2 Sommes de Riemann

Disposant d'une évaluation asymptotique de chaque terme $b(k, n, p)$, nous pouvons passer à celle de leur somme indexée par la condition :

$$c \leq \frac{k - np}{\sqrt{npq}} \leq d.$$

Nous notons respectivement k_1 et k_2 le plus petit et le plus grand entier vérifiant cette condition. Nous nous proposons dans cette partie d'évaluer à l'aide d'une intégrale la somme :

$$P(c \leq S_n^* \leq d) = \sum_{k=k_1}^{k_2} b(k, n, p) = \frac{1}{\sigma_n} \sum_{c \leq x_k \leq d} f(x_k) (1 + \varepsilon_{n,k}),$$

où l'on pose :

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Le nombre de termes dans cette somme est $k_2 - k_1 + 1$ qui est majoré par $(d - c)\sigma_n + 1$. Compte tenu de (B.23), on en déduit :

$$\frac{1}{\sigma_n} \sum_{k=k_1}^{k_2} |\varepsilon_{n,k}| \leq \frac{A((d - c)\sigma_n + 1)}{\sigma_n^2} \leq \frac{c_6}{\sigma_n}.$$

Par conséquent :

$$\left| P(c \leq S_n^* \leq d) - \frac{1}{\sigma_n} \sum_{c \leq x_k \leq d} f(x_k) \right| \leq \frac{c_6}{\sigma_n}. \quad (\text{B.24})$$

On est ainsi ramené à l'évaluation de la somme de Riemann de terme général $\sigma_n^{-1} f(x_k)$. On considère pour cela $\sigma_n^{-1} f(x_k)$ comme l'aire d'un rectangle de base $[s_k, s_{k+1}]$ et de hauteur $f(x_k)$, où les s_k constituent un partage de l'intervalle $[c, d]$ tel que $s_{k+1} - s_k = \sigma_n^{-1}$ et $s_k \leq x_k \leq s_{k+1}$ pour tout k . A priori, x_k peut être placé n'importe où⁹ sur $[s_k, s_{k+1}]$. Nous admettrons (cf. exercice B.10) que le meilleur choix pour une fonction f de classe C^2 , est de prendre x_k au milieu de $[s_k, s_{k+1}]$, ce qui revient à poser :

$$s_k = x_k - \frac{1}{2\sigma_n} \quad \text{et} \quad s_{k+1} = x_k + \frac{1}{2\sigma_n} \quad (k_1 \leq k \leq k_2).$$

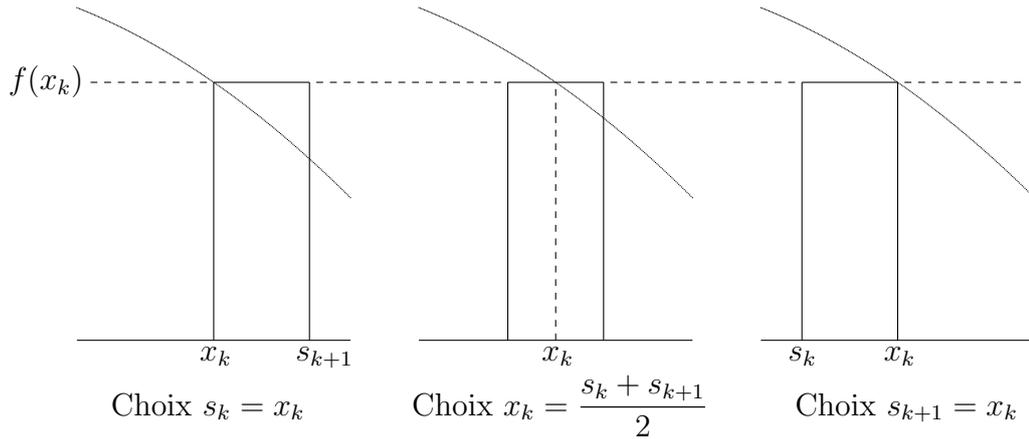


FIGURE B.10 – Divers choix de s_k

Lorsque f est monotone sur $[x_{k-1}, x_{k+1}]$ (ce qui est bien le cas ici sauf peut-être pour l'un de ces intervalles), il est facile de se convaincre que ce choix est meilleur que de prendre comme partage la suite des x_k elle-même (figure B.10).

Avec notre choix des s_k , on a (cf. exercice B.10) :

$$\left| \frac{1}{\sigma_n} f(x_k) - \int_{s_k}^{s_{k+1}} f(x) dx \right| \leq \frac{1}{24\sigma_n^3} \sup_{[s_k, s_{k+1}]} |f''|.$$

La fonction f'' est donnée ici par $f''(x) = (2\pi)^{-1/2}(x^2 - 1) \exp(-x^2/2)$. Elle est bornée sur \mathbb{R} , le maximum de $|f''|$ est atteint en 0 et vaut $(2\pi)^{-1/2}$. Nous pouvons donc écrire :

$$\frac{1}{\sigma_n} \sum_{k=k_1}^{k_2} f(x_k) = \int_{x_{k_1}-1/(2\sigma_n)}^{x_{k_2}+1/(2\sigma_n)} f(x) dx + \Delta_n, \quad (\text{B.25})$$

où

$$|\Delta_n| \leq \frac{1}{\sqrt{2\pi}} \times \frac{k_2 - k_1 + 1}{24\sigma_n^3} \leq \frac{(d - c)\sigma_n + 1}{24\sqrt{2\pi} \sigma_n^3} \leq \frac{c_7}{\sigma_n^2}. \quad (\text{B.26})$$

Enfin dans l'évaluation de notre somme de Riemann par $\Phi(d) - \Phi(c)$, nous devons tenir compte d'une dernière source d'erreur : les deux termes de bord générés par la non coïncidence (en général) des bornes c et $x_{k_1} - 1/(2\sigma_n)$ à gauche et d et $x_{k_2} + 1/(2\sigma_n)$ à droite (cf. figure B.11). Comme à chaque fois la distance entre les deux bornes concernées n'excède pas $1/(2\sigma_n)$ et la fonction f est majorée sur \mathbb{R} par $1/\sqrt{2\pi}$, on contrôle ces deux termes d'erreur par l'inégalité :

$$\left| \int_{x_{k_1}-1/(2\sigma_n)}^c f(x) dx \right| + \left| \int_{x_{k_2}+1/(2\sigma_n)}^d f(x) dx \right| \leq \frac{1}{\sqrt{2\pi} \sigma_n}. \quad (\text{B.27})$$

9. Le choix de la position de l'un des x_k détermine celle de tous les autres par translation.

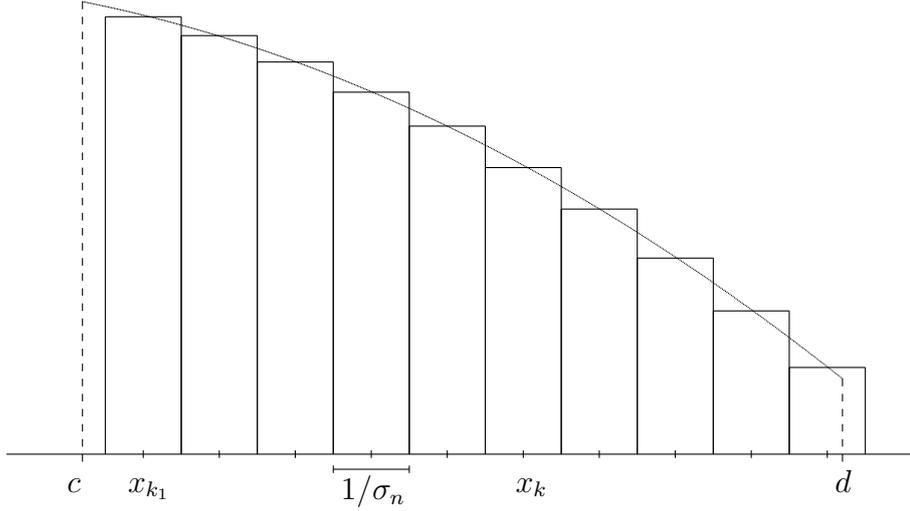


FIGURE B.11 – Approximation par $\Phi(d) - \Phi(c)$

Conclusion

L'inégalité triangulaire nous permet de déduire de (B.24), (B.25), (B.26) et (B.27) l'existence d'une constante C et d'un entier $n_0 = n_0(p, c, d)$ tels que :

$$\forall n \geq n_0, \quad \left| P(c \leq S_n^* \leq d) - \int_c^d \frac{\exp(-x^2/2)}{\sqrt{2\pi}} dx \right| \leq \frac{C}{\sqrt{npq}}. \quad (\text{B.28})$$

Ceci achève la preuve du théorème de de Moivre-Laplace.

Remarque :

Lorsqu'on applique ce théorème d'un point de vue pratique (i.e. avec n fixé), on peut éliminer l'erreur due aux termes de bord (B.27) en remplaçant c par $x_{k_1} - 1/(2\sigma_n)$ et d par $x_{k_2} + 1/(2\sigma_n)$. On peut voir que cela ne change rien pour S_n^* , en effet :

$$\begin{aligned} x_{k_1} - \frac{1}{2\sigma_n} \leq S_n^* \leq x_{k_2} + \frac{1}{2\sigma_n} &\Leftrightarrow \frac{k_1 - 1/2 - np}{\sqrt{npq}} \leq \frac{S_n - np}{\sqrt{npq}} \leq \frac{k_2 + 1/2 - np}{\sqrt{npq}} \\ &\Leftrightarrow k_1 - \frac{1}{2} \leq S_n \leq k_2 + \frac{1}{2} \\ &\Leftrightarrow k_1 \leq S_n \leq k_2, \end{aligned}$$

puisque S_n ne prend que des valeurs entières. Ce procédé d'approximation s'appelle *correction de continuité*.

B.5 Vitesse de convergence

La démonstration du théorème de de Moivre-Laplace nous a permis de voir¹⁰ que la vitesse de convergence était en $O(n^{-1/2})$, plus précisément nous avons montré que l'erreur commise en utilisant l'approximation :

$$P(a \leq S_n \leq b) \simeq \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right)$$

est majorée par C/\sqrt{npq} . Par ailleurs l'évaluation du terme dominant de la loi binomiale par la formule de Stirling nous montre qu'on ne peut pas espérer en général avoir une vitesse de convergence meilleure que $O(n^{-1/2})$. De nombreux travaux de l'école russe du début du siècle ont été consacrés à la détermination d'une valeur explicite pour la constante C et plus généralement à la recherche d'un contrôle fin de l'erreur tenant compte des valeurs de p , a et b . Les calculs sont trop compliqués pour trouver leur place dans ce polycopié. Nous nous contenterons de présenter sans démonstration des résultats dûs à Uspensky¹¹.

Le résultat le plus fin concerne l'approximation obtenue en effectuant la « correction de continuité ». Dans cette approximation, chaque probabilité binomiale $P(S_n = k)$ est représentée graphiquement par le rectangle élémentaire de base $[k - 1/2, k + 1/2]$ et de hauteur $P(S_n = k)$. C'est exactement la convention adoptée sur les figures B.5 à B.9 pour représenter la loi de S_n . De plus les bornes de variation de Φ sont ajustées de façon à prendre en compte seulement un nombre entier de ces rectangles (plus de « chutes » aux bords). La réduction des bornes (en abscisse) du rectangle élémentaire par la transformation $t \mapsto (t - np)/\sqrt{npq}$ nous donne comme nouvelles bornes : $(k - 1/2 - np)/\sqrt{npq}$ et $(k + 1/2 - np)/\sqrt{npq}$. Si k_1 et k_2 sont des entiers, l'approximation avec correction de continuité s'écrit ainsi :

$$\begin{aligned} P(k_1 \leq S_n \leq k_2) &= P\left(k_1 - \frac{1}{2} \leq S_n \leq k_2 + \frac{1}{2}\right) \\ &= P\left(\frac{k_1 - 0,5 - np}{\sqrt{npq}} \leq S_n^* \leq \frac{k_2 + 0,5 - np}{\sqrt{npq}}\right) \\ &\simeq \Phi\left(\frac{k_2 + 0,5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - 0,5 - np}{\sqrt{npq}}\right). \end{aligned}$$

Notons pour alléger :

$$z_1 = \frac{k_1 - 0,5 - np}{\sqrt{npq}}, \quad z_2 = \frac{k_2 + 0,5 - np}{\sqrt{npq}}. \quad (\text{B.29})$$

Voici maintenant une évaluation de l'erreur d'approximation pour la correction de continuité :

10. La lecture de cette section ne présuppose pas celle de la précédente. Il y a donc forcément quelques redites...

11. J. V. Uspensky, *Introduction to mathematical probability*, McGraw-Hill 1937.

Théorème 32 (Uspensky).

Si S_n suit la loi binomiale $\mathcal{B}(n, p)$ et si $npq \geq 25$, alors pour tous entiers k_1 et k_2 tels que $0 \leq k_1 \leq k_2 \leq n$, on a :

$$P(k_1 \leq S_n \leq k_2) = \Phi(z_2) - \Phi(z_1) + \frac{q-p}{6\sqrt{2\pi npq}} \left[(1-t^2) \exp\left(-\frac{t^2}{2}\right) \right]_{z_1}^{z_2} + \varepsilon,$$

où z_1 et z_2 sont définis par (B.29) et où le terme complémentaire ε vérifie :

$$|\varepsilon| < \frac{0,13 + 0,18|p-q|}{npq} + \exp\left(-\frac{3}{2}\sqrt{npq}\right). \quad (\text{B.30})$$

Le terme d'erreur dominant est en $1/\sqrt{npq}$, il se dégrade quand p est proche de 0 ou de 1 ($q = 1-p$ est alors proche de 0).

Dans le cas particulier où $p = 1/2$, le terme en $1/\sqrt{npq}$ s'annule et l'erreur d'approximation est en $O(1/(npq))$. De même si p est quelconque mais $z_1 = -z_2$ (intervalle $[k_1, k_2]$ symétrique autour de np).

Lorsque l'on n'effectue pas la correction de continuité, les résultats sont moins bons, mais on a toujours en général un terme d'erreur dominant en $1/\sqrt{npq}$. Quelques notations supplémentaires sont nécessaires à leur exposition. Nous souhaitons évaluer $P(x_1 \leq S_n^* \leq x_2)$ pour x_1 et x_2 réels fixés. Pour $x \in \mathbb{R}$, on appelle partie fractionnaire de x le réel $\theta = x - [x]$ (où $[x]$ désigne la partie entière de x , plus grand entier inférieur ou égal à x). On note alors :

$$\theta_1 \text{ la partie fractionnaire de } nq - x_1\sqrt{npq}, \quad (\text{B.31})$$

$$\theta_2 \text{ la partie fractionnaire de } np - x_2\sqrt{npq}. \quad (\text{B.32})$$

Théorème 33 (Uspensky).

Si S_n suit la loi binomiale $\mathcal{B}(n, p)$ et si $npq \geq 25$, alors pour tous réels x_1 et x_2 avec $x_1 < x_2$, on a :

$$P(x_1 \leq S_n^* \leq x_2) = \Phi(x_2) - \Phi(x_1) + \frac{A}{\sqrt{2\pi npq}} + \frac{q-p}{6\sqrt{2\pi npq}} \left[(1-t^2)e^{-t^2/2} \right]_{x_1}^{x_2} + \varepsilon',$$

où θ_1 et θ_2 sont définis par (B.31) et (B.32), A par :

$$A = \left(\frac{1}{2} - \theta_1\right) \exp\left(-\frac{x_1^2}{2}\right) + \left(\frac{1}{2} - \theta_2\right) \exp\left(-\frac{x_2^2}{2}\right)$$

et le terme ε' est majoré par :

$$|\varepsilon'| < \frac{0,20 + 0,25|p-q|}{npq} + \exp\left(-\frac{3}{2}\sqrt{npq}\right). \quad (\text{B.33})$$

On peut faire des commentaires analogues à ceux qui suivent le théorème 32. Il est intéressant de déduire du théorème 33 une majoration uniforme (par rapport à x_1 et x_2) de l'erreur d'approximation du type C/\sqrt{npq} . Ceci permet de se faire rapidement une première idée et de voir s'il y a lieu d'affiner l'approximation à l'aide de l'un des résultats ci-dessus.

Corollaire 34 (Majoration uniforme de l'erreur).

Si $npq \geq 25$, pour tous réels $x_1 < x_2$,

$$|P(x_1 \leq S_n^* \leq x_2) - (\Phi(x_2) - \Phi(x_1))| \leq \frac{0,588}{\sqrt{npq}}.$$

Preuve. Comme θ_1 et θ_2 sont dans $[0, 1[$, il est clair que $|A| \leq 1$. D'autre part l'étude des variations de la fonction $t \mapsto (1 - t^2)e^{-t^2/2}$ montre que sur \mathbb{R} son maximum est 1 (atteint en $t = 0$) et son minimum $-2e^{-1,5} > -0,446\ 3$ (atteint en $t = -\sqrt{3}$ et $t = \sqrt{3}$). On a donc :

$$\left| \frac{q-p}{6\sqrt{2\pi npq}} \left[(1-t^2)e^{-t^2/2} \right]_{x_1}^{x_2} \right| \leq \frac{1}{\sqrt{npq}} \frac{1+0,446\ 3}{6\sqrt{2\pi}}.$$

Par hypothèse, $\sqrt{npq} \geq 5$, d'où :

$$\frac{0,20 + 0,25|p-q|}{npq} \leq \frac{1}{\sqrt{npq}} \frac{0,45}{5} = \frac{0,09}{\sqrt{npq}}.$$

Enfin, pour le dernier terme, on écrit :

$$\exp\left(-\frac{3}{2}\sqrt{npq}\right) = \frac{1}{\sqrt{npq}} \left(\sqrt{npq} \exp\left(-\frac{3}{2}\sqrt{npq}\right) \right) \leq \frac{1}{\sqrt{npq}} \sup_{x \geq 5} (xe^{-3x/2}).$$

L'étude des variations de la fonction $x \mapsto xe^{-3x/2}$ montre qu'elle est décroissante et positive sur $[2/3, +\infty[$, le supremum ci-dessus est donc atteint en $x = 5$ et vaut $5e^{-7,5} < 0,002\ 8$. Finalement en rassemblant toutes ces majorations partielles, on voit que l'on a bien un majorant du type C/\sqrt{npq} et que l'on peut prendre :

$$C = \frac{1}{\sqrt{2\pi}} \left(1 + \frac{1,446\ 3}{6} \right) + 0,09 + 0,002\ 8 < 0,588\ 0.$$

□

Exemple 3 (suite de l'exemple 1).

On lance 3 600 fois un dé et on s'intéresse à la probabilité que le nombre d'apparitions du 1 soit compris strictement entre 540 et 660. D'après l'approximation gaussienne utilisée à l'exemple 1, cette probabilité vaut *environ* 0,992 6. D'autre part, l'inégalité de Tchebycheff nous dit qu'elle est certainement supérieure à 0,86. On voudrait savoir si l'approximation gaussienne donne réellement un meilleur résultat.

D'après le corollaire 34, l'erreur Δ commise en faisant l'approximation est majorée par :

$$\Delta \leq \frac{0,588}{\sqrt{npq}} = \frac{0,588}{\sqrt{500}} < 0,026\ 3.$$

On peut donc affirmer que

$$P(540 < S_n < 660) \geq 0,992\ 6 - 0,026\ 3 > 0,966\ 2$$

(en diminuant d'une unité le quatrième chiffre après la virgule pour prendre en compte la précision de 10^{-4} dans la table des valeurs de Φ). L'approximation gaussienne donne donc bien un meilleur résultat que l'inégalité de Tchebycheff.

B.6 Exercices

Ex B.1 *Un calcul d'intégrale*

- 1) Montrer la convergence de l'intégrale généralisée :

$$I = \int_0^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx.$$

- 2) Vérifier que :

$$I^2 = \int_0^{+\infty} \int_0^{+\infty} \exp\left(-\frac{(x^2 + y^2)}{2}\right) dx dy.$$

- 3) Calculer cette intégrale double en passant en coordonnées polaires.
4) En déduire que :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 1.$$

Ex B.2 *Comment construire la table des valeurs de Φ*

Utiliser le lemme sur l'évaluation de la queue de la loi normale pour donner une méthode pratique de construction d'une table des valeurs de Φ avec une précision au moins égale à 10^{-d} . *Indication* : On considérera le développement en série de Taylor de la densité gaussienne standard sur un intervalle compact $[-a, +a]$ convenablement choisi et on remarquera qu'il s'agit d'une série alternée. Il est donc facile d'encadrer les sommes partielles et de majorer l'erreur commise en remplaçant la série par l'une de ses sommes partielles. On est ainsi ramené à un calcul de polynôme.

Ex B.3 Trouver un entier k tel qu'avec une probabilité d'environ 0,5 le nombre de faces obtenues en 1 000 lancers d'une pièce soit compris au sens large entre 490 et k .

Ex B.4 Un central téléphonique dessert 5 000 abonnés. A un instant donné, chaque abonné a une probabilité égale à 2 % d'utiliser son téléphone et les appels des abonnés sont supposés indépendants. Quel nombre minimal d'appels doit pouvoir traiter simultanément le central pour que sa probabilité d'être saturé à un instant donné soit inférieure à 2.5 % ?

Ex B.5 *Surréservation aérienne*

Il arrive assez souvent que le nombre de réservations pour une liaison aérienne soit supérieur au nombre de passagers se présentant effectivement le jour du vol. Cela est dû à des empêchements imprévisibles de certains passagers et à une politique systématique de certains d'entre eux qui réservent des places sur plusieurs vols de façon à choisir au dernier moment celui qui leur convient le mieux (en raison de la concurrence, les compagnies ne pénalisent pas les clients qui se désistent et ne font payer effectivement que ceux qui embarquent). Pour compenser ce phénomène, une compagnie aérienne exploitant un avion de 300 places décide de faire de la surréservation (*surbooking*) en prenant pour chaque vol un nombre $n > 300$ de réservations. S'il se présente plus de 300 passagers à

l'embarquement, les 300 premiers arrivés prennent leur vol et les autres sont dédommagés financièrement.

1) On considère que les passagers sont mutuellement indépendants et que la probabilité de désistement de chacun d'eux est 10%. On note n le nombre de réservations prises par la compagnie pour un vol donné et S_n le nombre (aléatoire) de passagers se présentant à l'embarquement pour ce vol. Donner la loi de S_n , $\mathbf{E}S_n$ et $\text{Var } S_n$.

2) Le directeur commercial de la compagnie aimerait connaître la valeur maximale de n telle que : $P(S_n \leq 300) \geq 0,99$. En utilisant le théorème de De Moivre-Laplace, proposez une solution approchée de ce problème.

Ex B.6 Une entreprise de vente par correspondance envoie un lundi 2 100 colis. La probabilité qu'un colis soit distribué le lendemain est 0,7, celle qu'il soit retourné pour erreur d'adresse est de 0,001. On note X le nombre de colis distribués le lendemain et Y le nombre de colis retournés pour erreur d'adresse.

1) Donner la loi exacte, l'espérance et la variance de X . Même question pour Y .

2) Proposer une valeur approchée de la probabilité $P(1\,358 \leq X \leq 1\,442)$ en expliquant soigneusement la méthode utilisée.

3) Par quelle loi discrète classique peut-on approcher la loi de Y ? Utiliser cette approximation pour évaluer $P(Y \geq 4)$.

Ex B.7 *Un modèle simple pour le mélange de deux gaz*

Un récipient hermétique de 2 litres est séparé en deux parties symétriques « gauche » et « droite » par une cloison hermétique munie d'une vanne à large ouverture. La vanne étant fermée, la partie gauche du récipient contient au départ 1 litre d'oxygène et sa partie droite 1 litre d'azote le tout à la pression atmosphérique. On ouvre la vanne de la cloison intermédiaire et on laisse s'effectuer le mélange, puis au bout d'un temps suffisant on ferme la vanne. On mesure alors la proportion d'azote et d'oxygène dans la partie gauche et on constate *expérimentalement* qu'elles sont égales. Le but de cet exercice est l'étude d'un modèle simple permettant de prévoir ce phénomène.

Les molécules étant agitées d'un mouvement incessant, on suppose qu'après fermeture de la vanne, chaque molécule a une probabilité 1/2 de se trouver dans la partie gauche du récipient. On dispose de n molécules d'oxygène et n d'azote. On indexe les molécules d'oxygène de 1 à n et celles d'azote de $n + 1$ à $2n$. On note X_i la variable de Bernoulli indicatrice de l'événement *la i -ème molécule se trouve dans la partie gauche après fermeture de la vanne*. On suppose les X_i indépendantes. On pose :

$$S_n = \sum_{i=1}^n X_i, \quad T_n = \sum_{i=n+1}^{2n} X_i.$$

La variable S_n est donc le nombre aléatoire de molécules d'oxygène et T_n celui des molécules d'azotes dans la partie gauche après fermeture.

1) Quelle est la loi exacte de S_n , son espérance et sa variance (en fonction de n)? On a clairement les mêmes résultats pour T_n .

2) Soit $x > 0$ un nombre fixé. On considère l'événement

$$A = \left\{ \frac{n}{2} - \frac{x}{2}\sqrt{n} \leq S_n \leq \frac{n}{2} + \frac{x}{2}\sqrt{n} \right\}.$$

En utilisant le théorème de de Moivre-Laplace, montrer que l'on peut approximer $P(A)$ par $2\Phi(x) - 1$ où Φ est la fonction de répartition de la loi normale $\mathcal{N}(0, 1)$. On a bien sûr le même résultat pour $P(B)$, avec

$$B = \left\{ \frac{n}{2} - \frac{x}{2}\sqrt{n} \leq T_n \leq \frac{n}{2} + \frac{x}{2}\sqrt{n} \right\}.$$

3) On suppose désormais que $n - x\sqrt{n} > 0$. On s'intéresse à l'événement

$$C = \left\{ \frac{n - x\sqrt{n}}{n + x\sqrt{n}} \leq \frac{T_n}{S_n} \leq \frac{n + x\sqrt{n}}{n - x\sqrt{n}} \right\}.$$

Montrer que $A \cap B \subset C$.

4) En négligeant l'erreur due à l'approximation gaussienne, proposer à l'aide de $\Phi(x)$ une majoration de $P(A^c \cup B^c)$. En déduire une *minoration* de $P(C)$. On exprimera simplement le résultat final en fonction de la quantité $R(x) = (1 - \Phi(x))$.

5) Application numérique : $n = 10^{22}$, $x = 10$. On utilisera la formule d'encadrement de $1 - \Phi(x)$ pour les « très grandes valeurs de x » fournie à la fin des tables. Commentez le résultat obtenu.

Ex B.8 Une compagnie d'assurances assure sur une année n personnes contre un certain risque. On note X_i la somme qu'aura à verser cette année la compagnie au i -ème client. C'est une variable aléatoire qui prend la valeur 0 lorsque le client n'est pas sinistré. On peut en général considérer que les variables aléatoires X_1, \dots, X_n sont indépendantes. Supposons qu'elles obéissent toutes à une même loi d'espérance mathématique μ et de variance σ^2 . Soit x la prime demandée à chacun des n clients. Comment la compagnie doit-elle fixer x pour qu'avec une probabilité supérieure à $1 - \varepsilon$ la différence entre l'encaissement des primes et les remboursements sur l'année reste supérieure à un montant b déterminé par ses frais de gestion et le bénéfice minimum qu'elle désire faire ?

Ex B.9 Montrer que le théorème de de Moivre-Laplace implique la loi faible des grands nombres pour les fréquences. *Indication* : pour $\varepsilon > 0$ fixé, contrôler l'erreur commise écrivant l'approximation :

$$P\left(\left|\frac{S_n}{n} - p\right| \leq \varepsilon\right) \simeq \Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) - \Phi\left(-\varepsilon\sqrt{\frac{n}{pq}}\right).$$

Ex B.10 *Sur l'approximation d'une somme de Riemann*

Soit f une fonction deux fois continûment dérivable sur un intervalle $[a, b]$ et (s_k) un partage quelconque de $[a, b]$. Soit $x_k \in [s_k, s_{k+1}]$.

1) Montrer qu'en général on a :

$$\left| \frac{1}{s_{k+1} - s_k} f(x_k) - \int_{s_k}^{s_{k+1}} f(x) dx \right| \leq \frac{c}{(s_{k+1} - s_k)^2}.$$

2) Lorsque x_k est le milieu du segment $[s_k, s_{k+1}]$:

$$\left| \frac{1}{s_{k+1} - s_k} f(x_k) - \int_{s_k}^{s_{k+1}} f(x) dx \right| \leq \frac{m_k}{24(s_{k+1} - s_k)^3},$$

où m_k est le maximum de $|f''|$ sur $[s_k, s_{k+1}]$.

Indication : utiliser la formule de Taylor-Lagrange.

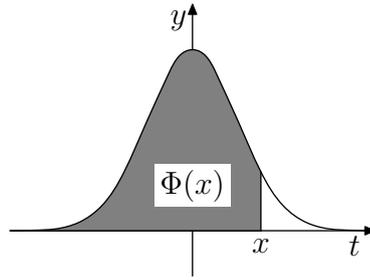
Ex B.11 En reprenant l'exemple 3, donner une majoration plus fine de l'erreur d'approximation.

Ex B.12 *Retour sur le problème des deux dés*

1) Combien de fois faut-il lancer une paire de dés pour qu'avec une probabilité supérieure à 0,999, la fréquence d'apparition du double six soit plus proche de $1/36$ que de $1/21$? (On supposera que la *probabilité* d'obtention d'un double six est effectivement $1/36$...).

2) Reprendre la même question en supposant que la probabilité d'apparition du double six est $1/21$.

3) Définir et discuter une procédure permettant de répondre expérimentalement à la question posée à la fin du chapitre 1.

Table des valeurs de Φ (loi normale standard)

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6627	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7122	0.7156	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7356	0.7389	0.7421	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8414	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8622
1.1	0.8643	0.8665	0.8687	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9083	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9193	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9895	0.9898	0.9901	0.9903	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9944	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

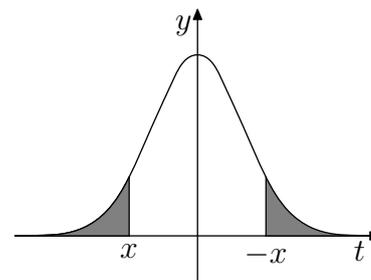
Table pour les grandes valeurs de x

x	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.8	4.0	4.5
$\Phi(x)$	0.99865	0.99904	0.99931	0.99952	0.99966	0.99976	0.999841	0.999928	0.999968	0.999997

La table donne les valeurs de $\Phi(x)$ pour x positif. Lorsque x est négatif, on utilise la relation

$$\Phi(x) = 1 - \Phi(-x)$$

qui résulte de la parité de la densité gaussienne $N(0, 1)$. Exemple : pour $x = -1, 8$, on trouve : $\Phi(x) = 1 - 0, 9641 = 0, 0359$.



Bibliographie

- [1] P. BARBE et M. LEDOUX. *Probabilité*. Espaces 34, Belin, 1998.
- [2] P. BILLINGSLEY. *Probability and measure*. Wiley, third edition 1995.
- [3] W. FELLER. *An Introduction to Probability Theory and its Applications*, Vol. I. Wiley.
- [4] W. FELLER. *An Introduction to Probability Theory and its Applications*, Vol. II. Wiley.
- [5] D. FOATA et A. FUCHS. *Calcul des Probabilités*. Dunod, 1998.
- [6] J.-Y. OUVRARD. *Probabilités tome 2, Maîtrise–Agrégation*. Cassini, 2000.
- [7] V. V. PETROV. *Sums of Independent Random Variables*. Springer, 1975.
- [8] D. REVUZ. *Probabilités*. Hermann, 1997.
- [9] P. S. TOULOUSE. *Thèmes de Probabilités et Statistique*. Dunod, 1999.
- [10] J. V. USPENSKY, *Introduction to mathematical probability*. McGraw-Hill, 1937.