

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



N° Ref :.....

Centre Universitaire de Mila

Institut des Sciences et de la Technologie

Département de math et informatique

Mémoire préparé En vue de l'obtention du diplôme de licence

En : - Filière: Mathématiques fondamentale

Thème

Théorème central limite

Préparé par : Mokrani Yousra
Khenouche Zahra
Menaâ Amel

Encadré par : Mehazzem Allal

Année universitaire : 2013/2014

REMERCIEMENTS

Louange à Dieu tout puissant de nous avoir aidé, éclairer le chemin pour achever notre travail et nos études.

Nous voudrions exprimer notre remerciement à Centre universitaire de Mila et Mr Mehazzem Allal, notre encadreur, pour l'attention et l'écoute dont il a su faire preuve. Nous le remercions très sincèrement pour ses conseils et ses commentaires, pertinents et constructifs, qui nous permis de progresser dans notre travail et pour les documentations, relectures et corrections de ce mémoire.

Nous exprimons aussi nos remerciements à tous les enseignants de l'institut des Sciences et de la Technologie.

Enfin, nous remercions tous les amis pour leur encouragement durant la réalisation de notre projet.

Table des matières

Introduction Générale	2
1 Généralités	3
1.1 Quelques lois des probabilités	3
1.1.1 Loi continue	3
1.1.2 Lois discrètes	4
1.2 Généralité sur les échantillons	6
1.2.1 Population et échantillon	6
1.2.2 Les différents types d'échantillons	6
1.2.3 Les grandes échantillons	7
1.3 Estimations et estimateurs	7
1.3.1 Définitions	7
1.3.2 Les deux propriétés d'un échantillon	7
1.4 Fonction caractéristique	8
1.4.1 Définitions	8
1.4.2 Propriétés :	9
1.5 Le théorème de de Moivre-Laplace	10
2 Théorèmes Central Limites (TCL)	12
2.1 Illustration	12
2.2 Théorème central limite	14
2.2.1 Théorème du TCL :	14
2.2.2 Démonstration du TCL	15
2.2.3 Utilisation du TCL :	16
2.3 Réciproque du TLC i.i.d	17
2.3.1 Théorème	17

2.4	Vitesse de convergence	18
2.4.1	Théorème (Berry-Esséen, 1941–42)	18
2.5	Le TCL dans \mathbb{R}^d	19
2.5.1	Vecteurs aléatoires	19
2.5.2	Le TCL iid dans \mathbb{R}^d	20
2.6	TCL pour les sommes d’un nombre aléatoire de termes	21
2.6.1	Théorèmes	21
	Bibliographie	22

Introduction Générale

La théorie des probabilités est l'étude mathématique des phénomènes caractérisés par le hasard et l'incertitude. Elle forme avec la statistique les deux sciences du hasard qui sont partie intégrante des mathématiques.

Dans notre travail, On a appliqué un théorème de probabilité s'appelle le **Théorème Central Limite** (aussi appelé **Théorème de la Limite Centrale** ou **centrée**) établit la convergence en loi de la somme d'une suite de variables aléatoires vers la loi normale. Intuitivement, ce résultat affirme que toute somme de variables aléatoires indépendantes et identiquement distribuées tend vers une variable aléatoire gaussienne.

La première démonstration de ce théorème, publiée en 1809, est due à *Pierre-Simon de Laplace*, mais le cas particulier où les variables suivent la loi de Bernoulli de paramètre $p = 0,5$ était connu depuis les travaux de *De Moivre*, en 1733.

Historiquement, et conformément à la traduction du titre, c'est donc bien le théorème qui est central, ce qui rend l'appellation « théorème de la limite centrale » impropre en toute rigueur. Elle est cependant souvent employée, les mathématiciens français considérant que l'adjectif « central » s'applique au centre de la distribution, par opposition à sa queue.

Le Théorème Central Limite admet plusieurs généralisations qui donnent la convergence des sommes des variables aléatoires sous des hypothèses beaucoup plus faibles. Ces généralisations ne nécessitent pas des lois identiques mais font appel à des conditions qui assurent qu'aucune des variables n'exerce une influence significativement plus importante que les autres.

Pour cela notre travail a été réparti sur deux chapitres comme suit :

Un premier Chapitre dans lequel nous allons expliquer quelques généralités de probabilité qui servir la théorie.

Fondamentalement, au deuxième chapitre et nous avons expliqué la théorie par définition, démonstration et l'utilisation...ect.

Ainsi, ce théorème (TCL) et ses généralisations offrent une explication de l'omniprésence de la loi normale dans la nature : de nombreux phénomènes sont dus à l'addition d'un grand nombre de petites perturbations aléatoires.

Chapitre 1

Généralités

1.1 Quelques lois des probabilités

1.1.1 Loi continue

Loi normale ou loi de la Place-Gauss

Si l'on considère des variables X_i , $i = 1, \dots, n$, indépendantes, obéissant à des lois de probabilités quelconques, on a démontré que pourvu que toutes les variables X_i soient du même ordre de grandeur, la somme des variables tend à obéir à une loi de probabilité unique s'appelle **loi normale**.

Définition 1.1.1 : une variable aléatoire absolument continue X suit une loi normale de paramètre (μ, δ) si sa densité de probabilité est donnée par : $\forall x \in \mathbb{R}$, $f(x) = \frac{1}{\delta\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\delta}\right)^2\right]$ avec $\mu \in \mathbb{R}$ et $\delta \in \mathbb{R}^+$

Notation : $X \sim N(\mu, \delta)$.

Propriétés :

propriétés de $f(x)$:

1. $\lim_{x \rightarrow \pm\infty} f_X(x) = 0$.
2. $f_X(\mu + x) = f_X(\mu - x)$ (symétrie par rapport à l'axe $x = \mu$).
3. f_X atteint son maximum en $x = \mu$ (μ est le mode de X).
4. Les points d'inflexion du graphe de f_X sont $x = \mu \pm \delta$.

Si $X \sim N(\mu, \delta^2)$ alors :

1. $p(X < \mu - x) = p(X < \mu + x)$.
2. $F_X(\mu - x) = 1 - F_X(\mu + x)$

Moyenne et variance de la loi normale :

Si $X \sim N(\mu, \delta^2)$ alors :

1. $E(X) = \mu$.
2. $V(X) = \delta^2$.

Lorsque $\mu = 0$ et $\delta^2 = 1$. La loi normale $N(0, 1)$ est appelée **centrée réduite** et on la densité par z .

Sa fonction de densité est :

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

Sa fonction de répartition est :

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{t^2}{2}\right) dt.$$

Loi normale centrée réduite :

On sait que l'intégrale définie par $\int_a^b f(x) dx$ est représenté par la surface de l'aire limitée par la courbe $y = f(x)$. L'axe O_X et deux droite $X = a$, $X = b$. Ici cet intégrale est la probabilité $p(a \leq X \leq b)$ c'est pourquoi on va étudier la fonction sous intégrale $Y = f(x)$.

On introduit la nouvelle variable $T = \frac{X-m}{\delta}$ qui s'appelle **variable aléatoire réduite**. D'ou on a $T = \frac{X-m}{\delta}$ on en déduit $dx = \delta dt$.

$$\begin{aligned} p(a \leq x \leq b) &= \int_a^b \frac{1}{\delta\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{(x-m)^2}{\delta^2}\right)\right] dx = \int_{t_1}^{t_2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \\ &= \int_{t_1}^{t_2} p(t) dt = p(t_1 \leq T \leq t_2) \end{aligned}$$

où $t_1 = \frac{a-m}{\delta}$, $t_2 = \frac{b-m}{\delta}$, $p(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$.

En introdui sont la variable réduite $T = \frac{x-m}{\delta}$, on passe à calculer la probabilité $p(t_1 \leq T \leq t_2)$ de cette nouvelle T avec la nouvelle densité de probabilité $p(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$.

On dit que la variable aléatoire réduite suit la loi normale centrée réduite (puisque sa moyenne est égale à 0, son écart-type est égale à 1).

1.1.2 Lois discrètes

Loi de Bernoulli :

Définition 1.1.2 : soit un univer Ω constitué de deux éventualités, S pour succès et E pour echec $\Omega = \{E, S\}$

Sur le quel on construit une variable aléatoire discrète telle que où cours d'une épreuve si S est réalisée, $X = 1$.

si E est réalisée, $X = 0$.

On appelle **variable de Bernoulli** une variable indicatrice la variable aléatoire X telle que :

$$X : \Omega \rightarrow \mathbb{R}$$

$$X(\Omega) = \{0, 1\}$$

La loi de probabilité associée à la variable de Bernoulli X telle que : $p(X = 1) = p$, $p(X = 0) = q$, avec $p + q = 1$ est appelée **loi de Bernoulli** notée $B(p)$ on a :

$$E(X) = \sum_{i=1} x_i p_i(X = x_i) = p$$

$$Var(X) = E(X^2) - (E(X))^2$$

$$E(X^2) = 1 \times p^2 = p$$

$$Var(X) = p - p^2 = p(1 - p) = p \times q$$

Loi Binomiale :

Définition 1.1.3 : soit l'application $S = \sum_{i=1}^n X_i$ où $(X_i)_i$ est une suite de variable aléatoire indépendants de Bernoulli de paramètre p .

On dit que S suite une **loi Binomiale** de paramètre n et p la probabilité que $S = k$ c'est-à-dire l'obtention de k succès au cours de n épreuves indépendantes est :

$$p(S = k) = C_n^k \times p^k \times q^{n-k}.$$

$$E(S) = n \times p.$$

$$V(S) = n \times p \times q.$$

Loi de Poisson :

Définition 1.1.4 : une variable aléatoire X à valeurs dans \mathbb{N} suite une loi Poisson de paramètre $\lambda (\lambda > 0)$ si :

$$p(X = k) = \exp(-\lambda) \frac{\lambda^k}{k!}$$

On noté : $X \sim p(\lambda)$.

$$E(X) = \lambda$$

$$\text{donc } V(X) = E(X^2) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

1.2 Généralité sur les échantillons

1.2.1 Population et échantillon

Définition 1.2.1 : lorsqu'on veut étudier les données relatives aux caractéristiques d'un ensemble d'individus ou d'objets (ex : nombre d'enfants de sexe féminin dans une famille de n enfants, taille des étudiants de l'université, poids de nouveau nés, etc...) il est difficile d'observer toute les données lorsque leur nombre est élevé. Au lieu d'examiner l'ensemble qu'on appelle **population**, on examine un nombre restreint qu'on appelle **échantillon**.

Pour être représentatif, l'échantillon doit-être pris au hasard. Une population peut être finie ou infinie.

1.2.2 Les différents types d'échantillons

Définition 1.2.2 : un **échantillon probabiliste** est un échantillon qui est construit à partir d'une méthode d'échantillonnage où chaque unité statistique de la population a une chance d'être sélectionnée. Il est parfois nommé échantillon aléatoire en raison de la manière aléatoire dont les unités statistiques sont choisies afin de garantir une représentation sans biais de l'ensemble de la population.

Définition 1.2.3 : une **échantillon non aléatoire** est un échantillon sélectionné par une méthode non probabiliste.

Définition 1.2.4 : un **échantillon exhaustif** est un échantillon pour lequel le tirage a été réalisé sans remise.

Définition 1.2.5 : un **échantillon non-exhaustif** est un échantillon pour lequel le tirage a été réalisé avec remise.

Définition 1.2.6 : un **échantillon représentatif** est un échantillon qui reproduit les caractéristiques d'une population de manière à ce que les conclusions obtenues avec cet échantillon se généralisent à la population.

Pour finir, les notions d'échantillon indépendants et appariés vont être présentés. Par exemple, vous aurez besoin de comparer deux population à l'aide de deux échantillons.

Définition 1.2.7 : si les $k \geq 2$ échantillons que vous étudiez sont composés de séries d'unités statistiques différentes, il s'agit alors d'**échantillons indépendants**.

Définition 1.2.8 : si les $k \geq 2$ échantillons que vous étudiez sont composés de la même séries d'unités statistiques, il s'agit alors d'**échantillons appariés**.

1.2.3 Les grandes échantillons

Distribution de la moyenne d'un échantillon

On considère une population de moyenne m et d'écart-type δ .

A partir de cette population on tire des échantillons de taille n .

1er échantillon X_1^1, \dots, X_n^1 de moyenne $m_1 = \sum X_i^1/n$

2ème échantillon X_1^2, \dots, X_n^2 de moyenne m_2

.....
 k ième échantillon X_1^k, \dots, X_n^k de moyenne m_k .

La moyenne de ces différents échantillons est une variable aléatoire qui varie d'un échantillon à un autre. On la notera \bar{m} et s'appelle le **moyenne d'échantillon**. \bar{m} suit une loi normale $N(\mu, \frac{\delta}{\sqrt{n}})$. $\bar{m} \sim N(\mu, \frac{\delta}{\sqrt{n}})$ lorsque $n \geq 30$

1.3 Estimations et estimateurs

1.3.1 Définitions

Définition 1.3.1 : un estimateur est un paramètre (le plus souvent la moyenne ou l'écart-type) d'une loi de probabilité théorique (loi de Poisson, loi Normale...etc) que l'on suppose suivie par une population.

Les vrais paramètres de cette loi étant inconnus (et pour cause....) , ils sont estimés à partir d'un échantillon aléatoire.

Exemple 1.3.2 : vous analysez un échantillon tiré aléatoirement et vous observez une moyenne de 100. Vous savez que la distribution de la population s'écarte peu de la normalité. Personne ne vous interdit de penser que la moyenne sur la population totale s'établit à 101. Qui est peut-être la médiane de l'échantillon ; sauf qu'il y a plus de chances qu'elle soit aussi égale à 100. La première étape de calcul des statistiques inférentielles consiste à utiliser l'estimateur ponctuel adéquat afin d'obtenir la meilleure estimation.

Définition 1.3.3 : l'estimation par intervalle donne un ensemble de valeurs susceptible d'être prises par le paramètre inconnu, avec une borne inférieure et une borne supérieure qui sont les limites de l'intervalle. Cet intervalle prend le nom d'intervalle de confiance et on lui affecte un coefficient de crédibilité appelée **niveau de confiance**.

1.3.2 Les deux propriétés d'un échantillon

On se doute que l'estimateur de moyenne d'une population est probablement meilleur lorsqu'il s'agit de la moyenne de l'échantillon plutôt que si on l'estime par : $(\min + \max)/2$.

On demande à un estimateur digne de ce nom de posséder les deux propriétés suivantes : être **convergent** et être **sans biais**.

Sans biais (unbiased estimator) :

Le biais est une erreur systématique. Un estimateur est dit "sans biais" si son espérance est égale au vrai paramètre de la population.

Evidemment, ce dernier est inconnu et c'est par une démonstration mathématique que l'on sait si tel estimateur est par construction, biaisé ou non.

En pratique, cette erreur est parfois négligeable et il n'est alors pas fondamental de chercher à l'éliminer. Plus l'échantillon est grand, plus le biais est petit (attention, je ne parle pas d'un éventuel biais sur les données).

L'estimateur est alors **asymptotiquement sans biais**.

Exemple 1.3.4 : sur un échantillon de mille individus, les estimateurs "variance" et "variance sans biais" ne montreront pas une différence sensible !.

La convergence (consistent estimator) :

L'autre qualité souhaitée est la convergence : si l'échantillon augmente indéfiniment, où du moins jusqu'à la taille de la population, l'estimation finit par être égale à la vraie valeur du paramètre, avec un intervalle de confiance qui diminue au fur et à mesure que la taille de l'échantillon augmente.

On démontre à l'aide de l'inégalité de *Bienayme-Tchebychert* qu'un estimateur dont la taille de l'échantillon tend vers l'infini.

Tout ceci est assez intuitif. Ce qui l'est moins, c'est qu'il peut exister plusieurs estimateurs pour un même paramètre et qu'ils ne convergent pas forcément de la même façon selon la taille de l'échantillon. L'estimateur le plus précis sur un échantillon de dix individus n'est pas toujours le plus précis sur un échantillon de taille 1000.

1.4 Fonction caractéristique

1.4.1 Définitions

Définition 1.4.1 : la fonction caractéristique d'une variable aléatoire réelle X est la fonction à valeurs complexes définie sur \mathbb{R} par :

$$\begin{aligned}\Phi_X(t) &= E[\exp(itX)] \\ &= E[\cos(tX)] + iE[\sin(tX)].\end{aligned}$$

Si cette variable aléatoire possède une densité, disons f_X alors :

$$\Phi_X(t) = \int_{\mathbb{R}} f_X(x) \exp(itx) dx$$

Ainsi, dans le cas d'une variable aléatoire à densité, la fonction caractéristique est la transformée de Fourier (à un facteur 2π près suivant la convention) de la densité. Probablement pour cette raison, il arrive que l'on choisisse une convention différente, à savoir : $\Phi_X(t) = E[\exp(2i\pi tX)]$.

Plus généralement, la fonction caractéristique d'une variable aléatoire X à valeurs dans \mathbb{R}^d est la fonction à valeurs complexes définie sur \mathbb{R}^d par : $\Phi_X(u) = E[\exp i \langle u, X \rangle]$ où $\langle u, X \rangle$ est le produit scalaire de u avec X .

Lorsque la variable aléatoire X est discrète, on définit sa fonction génératrice par $G(Z) = E[Z^X]$ avec Z complexe (quand cela a un sens). Avec les notations précédentes on a donc $\Phi_X(t) = G(\exp(it))$ cette fonction G est donc en fait un prolongement de φ_X .

Définition 1.4.2 : on peut écrire $N = Y_1 + \dots + Y_n$, où $(Y_i)_{i \leq n}$ sont n vecteurs aléatoires indépendamment distribués (iid) définis par :

$$\forall 1 \leq i \leq n, \forall 1 \leq j \leq k, p \left[Y_i = \begin{bmatrix} 0 \\ \vdots \\ 1_{i=j} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right] = p_j$$

donc pour tout $u \in \mathbb{R}^k$ on a :

$$\begin{aligned} \varphi_N(u) &= [\varphi_{Y_1}(u)]^n \\ &= \left(\sum_{j=1}^k p_j \exp(i \langle u, e_j \rangle) \right)^n \\ &= \left(\sum_{j=1}^k p_j \exp(iu_j) \right)^n \end{aligned}$$

1.4.2 Propriétés :

*Elle détermine de façon unique la loi d'une variable aléatoire au sens où $\varphi_X = \varphi_Y$ (égalité de fonction) équivaut à « X et Y ont la même loi».

*Si X et Y sont deux variables aléatoires indépendantes, $\varphi_{X+Y} = \varphi_X \varphi_Y$ plus généralement, si X_1, \dots, X_n sont des variables aléatoires indépendantes dans leur ensemble, alors : $\varphi_{X_1 + \dots + X_n} = \varphi_{X_1} \dots \varphi_{X_n}$.

En appliquant alors la transformée de Fourier à φ_{X+Y} permet de retrouver la loi de $X + Y$.

*Il y a aussi une relation entre les moments et la fonction caractéristique d'une variable aléatoire. Lorsque les moments existent et que la série converge :

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{i^k m_k(X)}{k!} t^k$$

où $m_k(X)$ est le moment d'ordre k de X .

*Cette relation sert parfois pour calculer la moyenne (premier moment) et la variance d'une variable aléatoire plus explicitement :

$$\Phi_X^{(k)}(0) = i^k m_k(X).$$

donc :

$$E[X] = -i\Phi_X'(0).$$

$$E[X^2] = -\Phi_X''(0).$$

$$Var(X) = -\Phi_X''(0) + \Phi_X'^2(0).$$

*La relation suivante sert, par exemple, à calculer la fonction caractéristique d'une variable centrée réduite, à partir de la fonction caractéristique de la variable de départ :

$$\Phi_{aX+b}(t) = \Phi_X(at) \exp(itb)$$

1.5 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 :

$$S_n^* = \frac{S_n - n\mu}{\delta} = \frac{S_n - np}{\sqrt{npq}} \quad (*)$$

Théorème 1.5.1 (de Moivre-Laplace). Soit S_n une variable aléatoire de loi binomiale de paramètres n et p et S_n^* définie par (*). Pour tous réels $c < d$ fixés :

$$\lim p(c < S_n^* \leq d) = \Phi(d) - \Phi(c) = \int_c^d \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

On a la même limite pour $P(c \leq S_n^* \leq d)$, $P(c \leq S_n^* < d)$ et $P(c < S_n^* < d)$. 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 $x \rightarrow (x - np) / \sqrt{npq}$ étant croissante, on a l'équivalence :

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

On en déduit que pour tout $n \geq 1$:

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

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 (*). L'idée générale est que la vitesse de convergence dans le théorème de de 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).

Exemple 1.5.2 : on lance 3600 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 $B(3600, 1/6)$. On a $E(S) = 600$ et $Var(S) = 500$. En notant $S^* = (S - E(S))/\delta(S)$ 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 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 :

$$p(540 < S < 660) \simeq 2\Phi(2.68) - 1 \simeq 0.9926 \simeq 0.99$$

Chapitre 2

Théorèmes Central Limites (TCL)

2.1 Illustration

Ce théorème est évident si les variables aléatoires suivent une loi normale d'espérance (ou moyenne) μ : on imagine bien que la somme de n variables puisse suivre une loi normale de paramètre $n\mu$.

Dans le cas de variables ne suivant pas une loi normale, le théorème peut sembler étonnant au premier abord. Nous allons donc en faire une illustration ne nécessitant pas de connaissance particulière en statistiques, mais uniquement du dénombrement.

Considérons le jeu de pile ou face et mettons des valeurs sur les faces de la pièce, par exemple 0 pour pile et 1 pour face, on s'intéresse à la somme de n tirages. La pièce est équilibrée, chaque face a une chance sur deux d'être tirée. Si l'on fait un seul tirage, nous avons donc le tirage n°1 (et aucun autre), et son résultat peut être 0 ou 1, nous faisons la somme d'une seule valeur.

Résultats d'un tirage

Résultat tirage n°1	Somme
0	0
1	1

Nous avons donc $n = 2$ possibilités pour la valeur de la somme, apparaissant avec les fréquences suivantes :

Fréquences pour tirage

valeurs de la somme	Nombre d'apparitions	Fréquence
0	1	$\frac{1}{2} = 0.5$ (50 %)
1	1	$\frac{1}{2} = 0.5$ (50 %)

Avec deux tirages, chaque tirage peut donner 0 ou 1, ce qui donne le tableau suivant :

Résultats de deux tirages

Résultat tirage n°1	Résultat tirage n°2	Somme
0	0	0
0	1	1
1	0	1
1	1	2

nous avons $n = 4$ possibilités, soit le tableau des fréquences.

Fréquences pour deux tirages

valeurs de la somme	Nombre d'apparitions	Fréquence
0	1	$\frac{1}{4} = 0.25$ (25 %)
1	2	$\frac{2}{4} = 0.5$ (50 %)
2	1	$\frac{1}{4} = 0.25$ (25 %)

Et ainsi de suite :

Résultats et fréquences de trois tirages

Résultat tiragen°1	Résultat tiragen°2	Résultat tiragen°3	Somme	Valeurs de la somme	Nombre d'apparitions	Fréquence
0	0	0	0	0	1	0.125 (12.5 %)
0	0	1	1	1	3	0.375 (37.5 %)
0	1	0	1	2	3	0.375 (37.5 %)
0	1	1	2	3	1	0.125 (12.5 %)
1	0	0	1			
1	0	1	2			
1	1	0	2			
1	1	1	3			

On ne s'intéresse pas au tirage en lui-même, mais à la somme du tirage. De ce point de vue, plusieurs tirages sont équivalents, donc une valeur de somme peut être obtenue par plusieurs tirages, par exemple : pour deux dés à six faces, on peut obtenir 7 par $1 + 6$, $2 + 5$, $3 + 4$, $4 + 3$, $5 + 2$ et $6 + 1$, il y a six tirages équivalents. Or, il y a toujours plus de combinaisons permettant d'obtenir une valeur moyenne qu'une valeur extrême, ce qui donne la courbe en cloche.

2.2 Théorème central limite

2.2.1 Théorème du TCL :

Théorème 2.2.1 : soit X_1, X_2, \dots une suite des variables aléatoires réelles définies sur le même espace de probabilité, indépendantes et identiquement distribuées suivant la même loi D . Supposons que l'espérance μ et l'écart-type δ de D existent et soient finis avec $\delta \neq 0$.

Considérons la somme :

$$S_n = X_1 + X_2 + \dots + X_n.$$

Alors :

- l'espérance de S_n est $n\mu$.
- son écart-type vaut $\delta\sqrt{n}$.

De plus, quand n est assez grand, la loi normale $N(n\mu, n\delta^2)$ est une bonne approximation de la loi de S_n .

Afin de formuler mathématiquement cette approximation, nous allons poser :

$$X_n = S_n/n = (X_1 + X_2 + \dots + X_n)/n$$

et

$$Z_n = \frac{S_n - n\mu}{\delta\sqrt{n}} = \frac{\overline{X_n} - \mu}{\delta/\sqrt{n}}$$

De sorte que l'espérance et l'écart-type de Z_n valent respectivement 0 et 1 : la variable est ainsi dite **centrée et réduite**.

Le théorème central limite (TCL) stipule alors que la suite des variables aléatoires $Z_1, Z_2, \dots, Z_n, \dots$ converge en loi vers une variable aléatoire Z , définie sur le même espace probabilisé, et de loi normale centrée réduite lorsque n tend vers l'infini.

Cela signifie que si Φ est la fonction de répartition de $N(0, 1)$, alors pour tout réel z :

$$\lim_{n \rightarrow \infty} p(Z_n \leq z) = \Phi(z)$$

ou, de façon équivalente :

$$\lim_{n \rightarrow \infty} \mathbf{P}\left(\frac{\overline{X_n} - \mu}{\delta/\sqrt{n}} \leq \mathbf{z}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp(-u^2/2) \mathbf{d}u = \Phi(\mathbf{z})$$

2.2.2 Démonstration du TCL

Pour un théorème d'une telle importance en statistiques et en probabilité appliquée, il existe une démonstration particulièrement simple utilisant les fonctions caractéristiques. Cette démonstration ressemble à celle d'une des lois des grands nombres. Pour une variable aléatoire Y d'espérance 0 et de variance 1, la fonction caractéristique de Y admet le développement limité :

$$\varphi_Y(t) = 1 - \frac{t^2}{2} + \theta(t^2) \quad , \quad t \rightarrow 0$$

Si Y_i vaut $\frac{X_i - \mu}{\delta}$, il est facile de voir que la moyenne centrée réduite des observations : X_1, X_2, \dots, X_n est simplement :

$$Z_n = \frac{\overline{X_n} - \mu}{\delta/\sqrt{n}} = \sum_{i=1}^n \frac{Y_i}{\sqrt{n}}$$

D'après les propriétés élémentaires des fonctions caractéristiques, la fonction caractéristique de Z_n est :

$$\left[\varphi_Y \left(\frac{t}{\sqrt{n}} \right) \right]^n = \left[1 - \frac{t^2}{2n} + \theta(t^2/n) \right]^n \rightarrow \exp \left(\frac{-t^2}{2} \right) \text{ lorsque } n \rightarrow \infty.$$

Mais cette limite est la fonction caractéristique de la loi normale centrée réduite $N(0, 1)$, d'où l'on déduit le théorème central limite grâce au théorème de continuité de Lévy, qui affirme que la convergence simple des fonctions caractéristiques implique la convergence en loi.

2.2.3 Utilisation du TCL :

Nous allons utiliser ce théorème sous une hypothèse A un peu forte que celle indiquée dans l'énoncé.

Remarque 2.2.2 : hypothèse A : $E[|X_1|^3] < +\infty$.

On cherche souvent à déterminer une valeur numérique approximative pour des expressions du type $p\{S_n \in A\}$

Pour ce faire, on exprime d'abord cette probabilité comme somme où différence de probabilités de la forme $p\{S_n \in I\}$, où I est un intervalle. Dans le cas où, par exemple, $I =]a, b]$ est un intervalle ouvert à gauche et fermé à droite, on a :

$$p\{a < S_n \leq b\} = p\left\{ \frac{a - n\mu}{\delta\sqrt{n}} < \frac{S_n - n\mu}{\delta\sqrt{n}} \leq \frac{b - n\mu}{\delta\sqrt{n}} \right\}.$$

D'après le **théorème limite central**, le membre de droite est approximativement égal à :

$$P\left\{ \frac{a - n\mu}{\delta\sqrt{n}} < z \leq \frac{b - n\mu}{\delta\sqrt{n}} \right\}$$

où z est une v.a. $N(0, 1)$. Cette probabilité s'écrit donc :

$$\Phi\left(\frac{b - n\mu}{\delta\sqrt{n}}\right) - \Phi\left(\frac{a - n\mu}{\delta\sqrt{n}}\right)$$

L'opération qui consiste à soustraire $n\mu$ et à diviser par $\delta\sqrt{n}$ est appelée **normalisation** et on parle de somme normalisée pour se référer à l'expression $\frac{S_n - n\mu}{\delta\sqrt{n}}$.

Exemple 2.2.3 : a la roulette, vous jouez 81 fois, pariant chaque fois 1 dollar sur noir. Quelle est la probabilité qu'à la fin, votre bénéfice soit positif ?

Pour répondre à cette question, soit X_i le bénéfice à la mise i , de sorte que :

$$E(X_i) = -\frac{1}{19} = \mu, p\{X_i = +1\} = \frac{18}{38}, p\{X_i = -1\} = \frac{20}{38}$$

et

$$\text{var}(X_i) = E(X_i^2) - E(X_i)^2 = 1 - \left(-\frac{1}{19}\right)^2 = 0.9972 = \delta^2.$$

Le bénéfice après 81 jeux est $S_{81} = X_1 + \dots + X_{81}$. On cherche $p\{S_{81} \geq 0\}$.

Cette probabilité est égale à :

$$p\left\{S_{81} - 81 \times \left(-\frac{1}{19}\right) \geq 0 - 81 \times \left(-\frac{1}{19}\right)\right\} = p\left\{\frac{S_{81} - 81 \times \left(-\frac{1}{19}\right)}{\sqrt{0.9972} \sqrt{81}} \geq \frac{\frac{81}{19}}{\sqrt{0.9972} \sqrt{81}}\right\}.$$

D'après le **TLC**, cette probabilité est approximativement égale à :

$$\begin{aligned} p\left\{Z \geq \frac{9}{19\sqrt{0.9972}}\right\} &= p\{Z \geq 0.47\} = 1 - p\{Z < 0.47\} \\ &= 1 - 0.6808 = 0.3192. \end{aligned}$$

Exemple 2.2.4 : on lance une pièce de monnaie 400 fois. Quelle est la probabilité d'avoir au moins 220 piles ? (On s'attend à 200 piles).

Pour $i = 1, \dots, 400$, soit $X_i = 1$ si le $i^{\text{ème}}$ jet tombe sur pile, $X_i = 0$ sinon.

Alors X_i suit une loi de Bernoulli de paramètre $\frac{1}{2}$ et nous avons :

$$E(X_i) = \frac{1}{2} \text{ et } \text{var}(X_i) = \frac{1}{2} \left(1 - \frac{1}{2}\right) = \frac{1}{4}.$$

Posons $S_n = X_1 + \dots + X_n$, de sorte que S_{400} représente le nombre de piles après 400 jets. Nous cherchons

$$p\{S_{400} \geq 220\} = p\left\{\frac{S_{400} - 400 \times \frac{1}{2}}{\sqrt{\frac{1}{4} \sqrt{400}}} \geq \frac{220 - 400 \times \frac{1}{2}}{\sqrt{\frac{1}{4} \sqrt{400}}}\right\} = p\left\{\frac{S_{400} - 400 \times \frac{1}{2}}{\sqrt{\frac{1}{4} \sqrt{400}}} \geq 2\right\}.$$

D'après le **TCL**, cette probabilité est approximativement égale à :

$$p\{Z \geq 2\} = 1 - p\{Z \leq 2\} = 1 - 0.9772 = 0.0228.$$

Remarque 2.2.5 : si la question était «quelle est la probabilité d'avoir au plus 219 piles ?», on aurait calculé :

$$p\{S_{400} \leq 219\} = p\left\{\frac{S_{400} - 200}{\sqrt{\frac{1}{4} \sqrt{400}}} \leq \frac{219 - 200}{\sqrt{\frac{1}{4} \sqrt{400}}}\right\} = p\left\{Z \leq \frac{219 - 200}{10}\right\} \simeq p\{Z \leq 1.9\} = 0.9713.$$

La somme de ces deux probabilités devrait être égale à 1. Cependant, d'après nos approximations :

$$p\{S_{400} \leq 219\} + p\{S_{400} \geq 220\} = 0.0228 + 0.9713 = 0.9941 < 1.$$

Pour pallier ce problème, lié au fait qu'on approxime une loi discrète par une loi continue, on utilise la correction de continuité, que nous introduisons ci-dessous.

2.3 Réciproque du TLC i.i.d

2.3.1 Théorème

Théorème 2.3.1 : 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 $N(0, 1)$. Alors X_1 est de carré intégrable, $E(X_1^2) = 1$ et $E(X_1) = 0$.

2.4 Vitesse de convergence

La vitesse de convergence dans le théorème du **TCL** 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.

2.4.1 Théorème (Berry-Esséen, 1941–42)

Théorème 2.4.1 : soit $(X_i)_{i \geq 1}$ une suite des variables aléatoires i.i.d. telle que $E(X_1) = 0$, $E|X_1|^3 < +\infty$. On note $\delta^2 = E(X_1^2)$ ($\delta > 0$). Il existe alors une constante universelle $C > 0$ telle que pour tout $n \geq 1$:

Définition 2.4.2

$$\Delta_n = \sup_{x \in \mathbb{R}} \left| p \left(\frac{S_n}{\delta\sqrt{n}} \leq x \right) - \Phi(x) \right| \leq c \frac{E|X_1|^3}{\delta^3} \frac{1}{\sqrt{n}}$$

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, c'est ce qui explique la présence de $E|X_1|^3$ 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*, donc avec $X_1 = Y_1 - p$, où Y_1 suit une loi de *Bernoulli* de paramètre p . On trouve alors :

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

Exemple 2.4.3 : 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 n . Prenons X_1 de loi de *Bernoulli* de paramètre $1/2$. On a alors :

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

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

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

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

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

2.5 Le TCL 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.

2.5.1 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 (Ω, F, P) dans E .

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

$$\begin{aligned}
 \varphi_X & : E' \rightarrow \mathbb{C} \\
 u & \rightarrow \varphi_X(u) = E \exp(iu(X))
 \end{aligned}$$

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

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

Comme en dimension 1, la fonctionnelle caractéristique caractérise la loi et la convergence ponctuelle des fonctionnelles caractéristiques équivaut à la convergence en loi. Le

lemme de *Cramér-Wold* est une conséquence immédiate de ces remarques.

Lemme 2.5.1 :(*Cramér-Wold*). La suite $(X_n)_{n \geq 1}$ des 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 des variables aléatoires réelles $u(X_n)_{n \geq 1}$ converge en loi dans \mathbb{R} vers $u(X)$.

2.5.2 Le TCL iid dans \mathbb{R}^d

Théorème 2.5.2 : soit (X_k) une suite des vecteurs aléatoires de \mathbb{R}^d , indépendants, de même loi et de carré intégrable (i.e. $E \|X_1\|^2 < +\infty$) et $S_n = \sum_{k=1}^n X_k$. Alors :

$$\frac{S_n - \mu}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} N(\cdot, k).$$

où k est l'opérateur de covariance de X_1 .

Commentaire :

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

Une application importante du théorème précédent 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.

Exemple 2.5.3 :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 $j_1 + j_2 + \dots + j_d = n$,

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

2.6 TCL pour les sommes d'un nombre aléatoire de termes

2.6.1 Théorèmes

Théorème 2.6.1 : toutes les variables aléatoires considérées étant définies sur le même espace probabilisé, on suppose que :

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

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

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

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

3. Pour tout $j \geq 1$, N_j est indépendante de la suite $(X_k)_{k \geq 1}$. Dans ces conditions :

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

Théorème 2.6.2 : toutes les variables aléatoires considérées étant définies sur le même espace probabilisé, on suppose que :

1. $Y_n \xrightarrow[n \rightarrow +\infty]{loi} Y$.
2. $N_n \xrightarrow[n \rightarrow +\infty]{pr} +\infty$.
3. Pour tous $j, n \in \mathbb{N}^*$, Y_j et N_n sont indépendantes. Alors :

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

Bibliographie

- [1] Statistique et Probabilités : cour, exercices et problème corrigé. Par Dominique Foata et Aimé Fuchs. 2^{ème} édition Pour la nouvelle présentation. Dunod, Paris, 2003.
- [2] Initiation à la statistique. Par Frédéric Bertrand et Myriam Maumy. Dunod, Paris, 2010.
- [3] Statistique : cour et exercices. Par Dr.Prof.Ali-Rachedi. 9^{ème} édition office des publication universitaires : 02-2009.
- [4] Statistique et Probabilités : exercices et problèmes. Par Thérèse Phan et Pierre Roweczyk. Dunod, Paris, 2007.
- [5] Introduction à la théorie des probabilités. Par Robert c.Dalang et Daniel Conus. presses polytechniques et universitaires Romandes, 2008.
- [6] Statistique et Probabilités : manuel et exercices corrigée. Par Jean-Pierre locoutre. 4^{ème} édition maître de conférences à l'université Panthéon-Assas-Dunod, Paris 2009.
- [7] [Fr.wikipedia.org/wiki/Théorème_central_limite](http://fr.wikipedia.org/wiki/Théorème_central_limite).
- [8] Math.univ-lille.fr/suquet/Polys/TLC.pdf.