Partie I

DU THÉORÈME DE MOIVRE-LAPLACE (TML)

AU THÉORÈME-LIMITE CENTRAL (TLC)

Journée académique "Terminale"
Besançon, octobre 2012

Yves DUCEL & Bruno SAUSSEREAU
Institut de recherche sur l'enseignement des mathématiques
Université de Franche-Comté

Travail IREM - Université de Franche-Comté

- **Yves DUCEL** et **Bruno SAUSSEREAU**, enseignants-chercheurs à l'Université de Franche-Comté.
- Cet exposé est le fruit d'une réflexion menée dans le cadre du groupe de travail *Probabilités et statistique* de l'IREM de Franche-Comté.
- Le diaporama sera disponible.

1 Objectifs et notations

2 Le théorème de Moivre-Laplace (TML)
- Le TML et son contexte
- Une approche graphique du TML

3 Le théorème-limite central (TLC)
- Un autre énoncé du TML
- L'énoncé du TLC

4 Deux exemples d'utilisation du TLC
- La montre
- Un calcul d'erreur

5 Conclusion sur l'intérêt du TML

6 Bibliographie

Objectifs et notations

Objectifs de la partie I

- Situer le théorème de De Moivre-Laplace dans son contexte mathématique.
- Pointer les difficultés du passage du "discret" au "continu" qu'il implique.
- Illustrer le TML par une approche graphique.
- Présenter le TLC, généralisation du TML.
- Montrer sur deux exemples l'intérêt du TLC.
- Commenter l'intérêt actuel du TML au regard du TLC.

Notations

Dans toute la suite de l'exposé :

- les échantillons sont obtenus par répétition n fois d'une même expérience de Bernoulli (Succès/Échec) ;
- n (n entier naturel non nul) est la taille de l'échantillon et p ($0 < p < 1$) le paramètre de l'expérience de Bernoulli ;
- la variable aléatoire F_n désigne **la fréquence empirique** ;
- la variable aléatoire $X_n = nF_n$, de loi binomiale $\mathcal{B}(n, p)$, désigne **l'effectif empirique** ;
- l'abréviation "v.a." est mise pour "variable aléatoire".

Le théorème de Moivre-Laplace (TML)

Le TML et son contexte

Théorème de De Moivre-Laplace (TML)

- Posons, pour tout entier $n \geq 1$, $Z_n = \frac{X_n - np}{\sqrt{np(1-p)}} = \frac{F_n - p}{\sqrt{\frac{p(1-p)}{n}}}$.
- **Théorème de De Moivre-Laplace (TML)** : *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires de lois binomiales respectives* $\mathcal{B}(n,p)$ *avec* $p \in]0,1[$ *fixé. Alors, pour tous réels a et b avec* $a < b$,

$$\lim_{n\to+\infty} \mathbb{P}\left(a \leq Z_n \leq b\right) = \mathbb{P}(a \leq Z \leq b)$$

où Z *est une variable aléatoire normale centrée-réduite.*

♣ Pour une démonstration du cas $p = 1/2$ avec les outils du lycée, voir [MEN/DESCO-1].

Le contexte mathématique

- **Jacques Bernoulli** (1654-1705) : *Ars Conjectandi* (1713, édition posthume)
  - *Pour tout réel* $a > 0$, $\mathbb{P}\left(|F_n - p| \geq a\right) \leq \frac{p(1-p)}{na^2}$.
- **Abraham de Moivre** (1667-1754) : *The Doctrine of Chances* (2ième édition, 1738)
  - Démonstration en 1733 du TML pour le cas $p = 1/2$.
- **Pierre-Simon de Laplace** (1749-1827) : *Théorie analytique des probabilités* (1812)
  - Généralisation du TML vers 1789 au cas p quelconque.
  - Premières formulation et démonstration en 1809 de la forme générale du théorème-limite central généralisant le TML.

Interprétation du TML

- Le résultat de Bernoulli entraîne une certaine convergence (en probabilité) des v.a. F_n vers p.
- Les travaux de De Moivre sont motivés par l'étude plus poussée de la convergence.
- Autre écriture de la conclusion du TML :

$$\lim_{n\to+\infty} \mathbb{P}\left(a\sqrt{\frac{p(1-p)}{n}} \leq F_n - p \leq b\sqrt{\frac{p(1-p)}{n}}\right) = \mathbb{P}(a \leq Z \leq b)$$

- Le TML donne une information sur le comportement asymptotique de la loi de probabilité de l'erreur $|F_n - p|$.

♣ Pour une étude d'histoire des maths sur le TML, voir [LANIER] (PDF librement téléchargeable en ligne).

Majoration uniforme de l'erreur

- **Théorème de Uspensky** (1937) : Si $np(1-p) \geq 25$, Alors, pour tous réels a et b avec $a < b$,

$$|\mathbb{P}\,(a \leq Z_n \leq b) - \mathbb{P}(a \leq Z \leq b)| \leq \frac{0,588}{\sqrt{np(1-p)}} \,.$$

♣ Pour plus de détails, voir [SUQUET] (PDF librement téléchargeable en ligne).

Une approche graphique du TML

Rappels : Lois discrète ou continue

- La loi de probabilité d'une v.a. X **discrète à valeurs dans** $\mathbb{N}$ se représente graphiquement par un diagramme en bâtons :
  La probabilité $\mathbb{P}(u \leq X \leq v)$ s'interprète géométriquement comme une somme de longueurs : Voir graphique

- La loi de probabilité d'une v.a. X **continue à valeurs dans** $\mathbb{R}$ se représente graphiquement par la courbe de la fonction-densité de probabilité.
  La probabilité $\mathbb{P}(u \leq X \leq v)$ s'interprète géométriquement comme l'aire d'une certaine surface : Voir graphique

V.a. impliquées dans le TML

Le TML implique deux familles de v.a. :

- des v.a. discrètes, X_n et Z_n. Pour n fixé :
  - La v.a. X_n (binomiale) peut prendre les seules valeurs $0,\ 1,\ 2,\ \ldots,\ (n-1),\ n$.
    L'écart entre deux bâtons consécutifs est 1.
  - La v.a. Z_n (non binomiale) peut prendre les seules valeurs
    $$\frac{0-np}{\sqrt{np(1-p)}},\ \frac{1-np}{\sqrt{np(1-p)}},\ \ldots,\ \frac{(n-1)-np}{\sqrt{np(1-p)}},\ \frac{n-np}{\sqrt{np(1-p)}}.$$
    L'écart entre deux bâtons consécutifs est $1/\sigma$ avec $\sigma = \sqrt{np(1-p)}$.
- et une v.a. continue Z (gaussienne) qui peut prendre toutes les valeurs de $\mathbb{R}$.

Signification géométrique du TML

- Rappel TML : $\lim_{n \to +\infty} \mathbb{P}(a \leq Z_n \leq b) = \mathbb{P}(a \leq Z \leq b)$
- Le nombre $\mathbb{P}(a \leq Z \leq b)$ se représente géométriquement par l'aire d'une surface sous la courbe de Gauss : Voir graphique
- Le nombre $\mathbb{P}(a \leq Z_n \leq b)$ se représente géométriquement par une somme de longueurs de bâtons du diagramme de Z_n : Voir graphique
- Le TML pose la question du passage du discret au continu, via l'interprétation d'une longueur en termes d'aire.

Illustration graphique du TML

- Interprétation en tant qu'aire du nombre $\mathbb{P}(u \leq X \leq v)$ où X v.a. binomiale quelconque (espacement entre bâtons $= 1$) :
  Voir graphique
- Interprétation en tant qu'aire du nombre $\mathbb{P}(Y = \frac{x-m}{\sigma})$ où $Y = \frac{X-m}{\sigma}$ (espacement entre bâtons $= 1/\sigma$, détail d'un bâton) : Voir graphique
- Interprétation en tant qu'aire du nombre $\mathbb{P}(Z_n = \frac{k-np}{\sqrt{npq}})$ où $Z_n = \frac{X_n-np}{\sqrt{npq}}$ (espacement entre bâtons $= 1/\sqrt{npq}$, détail d'un bâton) : Voir graphique
- Interprétation en tant qu'aire du nombre $\mathbb{P}(a \leq Z_n \leq b)$ et approximation par $\mathbb{P}(a \leq Z \leq b)$: Voir graphique

Le théorème-limite central (TLC)

Un autre énoncé du TML

Modèle binomial

- Une urne $\mathcal{U}$ contient des boules rouges et vertes, avec une proportion p de boules rouges.
- L'expérience de Bernoulli considérée consiste à prélever une boule de l'urne $\mathcal{U}$, à noter sa couleur et à la remettre dans l'urne.
- On note Ω l'ensemble des suites de n boules (rouges ou vertes) obtenues par répétition n fois de cette expérience de Bernoulli.
- Une issue $\omega \in \Omega$ est une suite de n boules (rouges ou vertes).
- La probabilité $\mathbb{P}$ considérée ici est l'équiprobabilité sur Ω.
- La v.a. X_n, qui à toute suite ω de n boules associe le nombre $X_n(\omega)$ de boules rouges dans la suite ω, est de loi $\mathcal{B}(n, p)$.

Somme de v.a. de Bernoulli

- Pour tout $i = 1, 2, ..., n$, on définit la v.a. U_i qui, à toute suite ω de n boules, associe le nombre $U_i(\omega)$ défini par
$$U_i(\omega) = \begin{cases} 1 & \text{si la i-ième boule de } \omega \text{ est rouge,} \\ 0 & \text{si la i-ième boule de } \omega \text{ est verte.} \end{cases}$$
- Les v.a. U_i sont des v.a. de Bernoulli de paramètre $p = \mathbb{P}(U_i = 1)$, d'espérance $\mathbb{E}(U_i) = p$ et de variance $\mathbb{V}(U_i) = p(1 - p)$ (ces valeurs ne dépendent pas de i).
- Les v.a. $U_1, U_2, ..., U_n$ sont indépendantes du fait de la répétition, n fois à l'identique, de la <u>même</u> expérience aléatoire.
- On vérifie que $X_n = U_1 + U_2 + ... + U_n$.

Autre énoncé du TML

- Pour tout entier $n \geq 1$, on peut écrire :

$$Z_n = \frac{X_n - np}{\sqrt{np(1-p)}} = \frac{(U_1 + U_2 + \ldots + U_n) - n\mathbb{E}(U_1)}{\sqrt{n\mathbb{V}(U_1)}}.$$

- **Autre énoncé du TML** : *Soit $(U_n)_{n\geq 1}$ une suite de v.a. indépendantes et de Bernoulli de même paramètre p avec $p \in]0,1[$ fixé. On pose $m = \mathbb{E}(U_1)$ et $\sigma^2 = \mathbb{V}(U_1)$. Alors, pour tous réels a et b avec $a < b$, on a*

$$\lim_{n\to+\infty} \mathbb{P}\left(a \leq \frac{(U_1 + U_2 + \ldots + U_n) - nm}{\sqrt{n\sigma^2}} \leq b\right) = \mathbb{P}(a \leq Z \leq b)$$

où Z est une variable aléatoire normale centrée-réduite.

L'énoncé du TLC

Du TML au TLC

- La conclusion du TML, énoncé sous cette forme, reste vraie même si on affaiblit les hypothèses qui portent sur la loi commune des v.a. $U_1, U_2, \ldots, U_n, \ldots$, en remplaçant la contrainte "*de Bernoulli de même paramètre p avec $p \in]0, 1[$ fixé*" par "*de même loi de probabilité possédant un moment d'ordre* 2".
- On obtient alors le **théorème-limite central (du calcul des probabilités)** - en abrégé TLC - appelé ainsi en référence à l'article de George Pólya (1920) : *Über den zentralen Grenzwertsatz des Wahrscheinlichkeitsrechnung und das Momentenproblem.*

♣ Pour une démonstration relativement accessible du TLC, voir [DALANG].

TLC

Théorème-limite central (TLC) :

Soit $(U_n)_{n\geq 1}$ une suite de v.a. indépendantes et de même loi de probabilité possédant un moment d'ordre 2.
On pose $m = \mathbb{E}(U_1)$ et $\sigma^2 = \mathbb{V}(U_1)$.

Alors, pour tous réels a et b avec $a < b$, on a

$$\lim_{n\to+\infty} \mathbb{P}\left(a \leq \frac{(U_1 + U_2 + \ldots + U_n) - nm}{\sqrt{n\sigma^2}} \leq b\right) = \mathbb{P}(a \leq Z \leq b)$$

où Z est une variable aléatoire normale centrée-réduite.

Signification du TLC

- La connaissance de la loi de probabilité commune aux v.a. $(U_n)_{n\geq 1}$ n'est pas exigée. On exige seulement que cette loi soit la même pour chacun des U_n.
- Le TLC signifie pratiquement que, pour tout entier n assez grand :
  * on peut faire l'approximation suivante :

    $$\mathbb{P}\left(a \leq \frac{(U_1 + U_2 + \ldots + U_n) - nm}{\sqrt{n\sigma^2}} \leq b\right) \simeq \mathbb{P}(a \leq Z \leq b)$$

    où Z est une variable aléatoire normale centrée-réduite ;
  * la loi de la v.a. $\frac{(U_1+U_2+\ldots+U_n)-nm}{\sqrt{n\sigma^2}}$ est pratiquement la loi normale centrée-réduite.

Intérêt du TLC

- Le TLC présente un intérêt à la fois pratique et théorique : Bien qu'on ne connaisse pas la loi commune des U_i, on connaît cependant de façon approchée la loi de probabilité de $U_1 + U_2 + ... + U_n$.
- Les approximations du TLC sont communément acceptées pour $n \geq 30$ (certains préfèrent $n \geq 60$).

Deux exemples d'utilisation du TLC

La montre

La montre : le problème

Situation : On considère une montre qui présente chaque jour un décalage (avance ou retard) avec l'heure exacte d'au plus une demi-minute.
On considère que les décalages journaliers suivent une loi uniforme et qu'ils sont indépendants les uns des autres.

Question : Quelle est la probabilité que le décalage total au bout d'une année de 365 jours soit d'au plus 15 minutes ?

Modélisation de la situation

- Ω désigne l'ensemble des années considérées et $\mathbb{P}$ désigne l'équiprobabilité sur Ω.
- Pour tout $i = 1, 2, ..., 365$, on note U_i la v.a. qui, à toute année ω, associe le décalage $U_i(\omega)$ que présente la montre le jour i de l'année ω.
- Les applications U_i sont des v.a. indépendantes avec pour loi de probabilité commune la loi uniforme sur $[-1/2\ ;\ 1/2]$.
- Pour tout $i = 1, 2, ..., 365$, $m = \mathbb{E}(U_i) = 0$ et $\sigma^2 = \mathbb{V}(U_i) = 1/12$.
- La v.a. S, qui à toute année ω associe le décalage total que présente la montre au bout de l'année ω, vérifie $S = U_1 + U_2 + ... + U_{365}$.

Formalisation du problème

- L'événement "*Le décalage total au bout d'une année de* 365 *jours est d'au plus* 15 *minutes*" se formalise par $\{-15 \leq S \leq 15\}$.
- La probabilité cherchée est $\mathbb{P}(-15 \leq S \leq 15)$.
- Il faut déterminer la loi de la v.a. $S = U_1 + U_2 + ... + U_{365}$.
- Problème : bien que les lois des Y_i soient connues, celle de la v.a. $S = U_1 + U_2 + ... + U_{365}$ est compliquée à déterminer.

Utilisation du TLC

- Comme $n = 365 \geq 30$, on peut utiliser l'approximation du TLC.
- Par le TLC, la loi de $Z = \frac{S-0}{\sqrt{365/12}}$ est approximativement $\mathcal{N}(0,1)$.
- On a l'égalité entre événements :

$$\{-15 \leq S \leq 15\} = \left\{ \frac{-15-0}{\sqrt{365/12}} \leq Z \leq \frac{15-0}{\sqrt{365/12}} \right\} .$$

- La probabilité cherchée est pratiquement donnée par $\mathbb{P}(-15 \leq S \leq 15) \simeq \mathbb{P}(-2,72 \leq Z \leq 2,72) \simeq 0,9934$.

Un calcul d'erreur

Calcul d'erreur : le problème

Situation (d'après [MEN/DESCO-2]) : On souhaite mesurer de façon très précise la résistance d'un conducteur ohmique avec un ohmmètre relié à un ordinateur muni d'un programme d'acquisition de données. On effectue 2000 mesurages de cette résistance dans les mêmes conditions. On sait que l'écart type de la dispersion due aux modalités du mesurage est $\sigma_0 = 200$ microohms.

Question : On choisit de prendre comme valeur de cette résistance, la moyenne observée sur les 2000 mesurages. Peut-on quantifier l'erreur ainsi commise par rapport à la vraie valeur (inconnue) y_0 de cette résistance ?

Un peu de vocabulaire

- Un **mesurande** est une grandeur particulière (longueur, masse, intensité, résistance, ...) qu'on souhaite mesurer.
- Un **mesurage** est un ensemble d'opérations permettant de déterminer une valeur du mesurande.
- La **vraie valeur**, ou valeur **théorique**, d'un mesurande est la valeur (inconnue) qu'on obtiendrait dans un mesurage parfait.

Modélisation de la situation

- On considère ici qu'il n'y a pas d'erreur systématique ou de biais.
- On considère ici que l'erreur est uniquement aléatoire.
- On note Ω l'ensemble d'un grand nombre de mesurages de la résistance considérée, effectués dans les mêmes conditions, et $\mathbb{P}$ l'équiprobabilité sur Ω.
- On note Y la v.a. qui à un mesurage ω associe la valeur $Y(\omega)$ du mesurande obtenue par ce mesurage.

Hypothèses sur la loi de Y

- La loi de probabilité de la v.a. Y est inconnue, mais on fait les hypothèses que :
  - la v.a. Y fluctue aléatoirement autour de y_0 : $\mathbb{E}(Y) = y_0$;
  - la dispersion de la v.a. Y autour de y_0 est une caractéristique du type de mesurage utilisé : $\mathbb{V}(Y) = \sigma_0^2$.
- Pour tout $i = 1, 2, ..., 2000$, on note Y_i la v.a. qui, à une série de 2000 mesurages de Ω, associe la valeur obtenue par le i-ième mesurage.
- Comme les 2000 mesurages se font dans les mêmes conditions, on peut considérer que :
  - les v.a. Y_i sont indépendantes ;
  - les v.a. Y_i ont la même loi de probabilité (non connue) avec $\mathbb{E}(Y_i) = y_0$ et $\mathbb{V}(Y_i) = \sigma_0^2$.

Utilisation du TLC

- On pose $S = Y_1 + Y_2 + \dots + Y_{2000}$.
- Comme $n = 2000 \geq 30$, par le TLC la loi de $Z = \dfrac{S - ny_0}{\sigma_0\sqrt{n}}$ est approximativement la loi $\mathcal{N}(0,1)$.
- On sait que $\mathbb{P}(-1,96 \leq Z \leq 1,96) \simeq 0,95$.
- On a l'égalité entre événements :

$$\{-1,96 \leq Z \leq 1,96\} = \left\{ \frac{S}{n} - 1,96\frac{\sigma_0}{\sqrt{n}} \leq y_0 \leq \frac{S}{n} + 1,96\frac{\sigma_0}{\sqrt{n}} \right\}.$$

- On a donc :

$$\mathbb{P}\left(\frac{S}{n} - 1,96\frac{\sigma_0}{\sqrt{n}} \leq y_0 \leq \frac{S}{n} + 1,96\frac{\sigma_0}{\sqrt{n}} \right) \simeq 0,95$$

.

Interprétation et application numérique

- A priori, quand on effectue 2000 mesurages, on a 95% de chances que la vraie valeur y_0 de la résistance soit dans l'intervalle aléatoire :

$$\left[\frac{S}{n} - 1,96\frac{\sigma_0}{\sqrt{n}} \ , \ \frac{S}{n} + 1,96\frac{\sigma_0}{\sqrt{n}}\right] .$$

- La moyenne des valeurs observées est la réalisation de la v.a. $\frac{S}{n}$ pour les 2000 mesurages.
- On considérera que l'erreur commise, quand on prend la moyenne des valeurs observées sur 2000 mesurages pour estimer la vraie valeur y_0, est inférieure ou égale à : $1,96\ \sigma_0/\sqrt{n} = 9$ microohms.
- En prenant cette valeur de l'erreur, on a environ 5% de chances de se tromper.

Conclusion sur l'intérêt du TML

Intérêt du TML

- Le TML n'est pas intéressant du point de vue théorique car :
  * la loi commune des Y_i est connue : loi de Bernoulli de paramètre p ;
  * la loi de la v.a. $Y_1 + Y_2 + ... + Y_n$ est également connue de façon exacte et simple : loi binomiale $\mathcal{B}(n, p)$.
- Le TML a été intéressant du point de vue pratique mais ne l'est plus actuellement, car aujourd'hui
  * les outils de calcul scientifique permettent d'effectuer des calculs pour les lois binomiales avec n très grand ou avec p petit ;
  * l'approximation de loi binomiale par une loi normale n'est plus utile.
- Cependant le TML présente un intérêt historique et, du fait de la simplicité de son énoncé, un intérêt pédagogique, qui préparent à l'introduction ultérieure du TLC.

Bibliographie

Bibliographie pour le lycée

MEN/DESCO-1,
Probabilités et statistique : « Le théorème de Moivre-Laplace : une démonstration complète dans le cas $p = 1/2$ »,
Ressources pour la classe terminale générale et technologique : dossier annexe, téléchargeable sur le site Web Eduscol :
http ://eduscol.education.fr/prog, février 2012.

MEN/DESCO-2,
Mesure et incertitudes,
Ressources pour le cycle terminal général et technologique, téléchargeable sur le site Web Eduscol :
http ://eduscol.education.fr/prog, juin 2012.

Bibliographie sur le TML

Lanier D., Trotoux D.,
La loi des grands nombres, le théorème de De Moivre-Laplace, *Contribution à une approche historique de l'enseignement des mathématiques*, Actes de la 6ième université d'été interdisciplinaire sur l'histoire des mathématiques, Besançon, 8-13 juillet 1995, pages 259-294, Les Publications de l'IREM de Besançon, (Fichier pdf disponible sur http ://www.math.ens.fr/culturemath/), 1995.

Bibliographie sur le TLC

Suquet C.,
Théorème-limite central,
Cours agrégation externe de maths, 2005-2006, 63 pages, Université des sciences et technologies de Lille, (Fichier pdf disponible sur internet), 2005.

Dalang R.C.,
Une démonstration élémentaire du théorème central limite,
Elemente der Mathematik, 60, pages 1-9, Swiss Mathematical Society, 2005.

Merci de votre attention

Pour nous contacter ...

- **Yves Ducel, Bruno Saussereau**
  Institut de recherche sur l'enseignement des mathématiques (IREM)
  YD : 03 81 66 62 32 / BS : 03 81 66 63 00
  yves.ducel@univ-fcomte.fr ; bruno.saussereau@univ-fcomte.fr

- **Site Web de l'IREM** : http ://www-irem.univ-fcomte.fr/

- **Site Web du groupe IREM "Proba-Stat"** :
  http ://bsauss.perso.math.cnrs.fr/IREM_FC_GrouProbaStat/grouprobastat.html