Notes de cours Biostatistiques – MIV (L3)
Théorème central limite

M. Bailly-Bechet

Université Claude Bernard Lyon 1 – France

1 Développements limités, moments et théorème central limite

1.1 Développements limités

.

On aborde ici de manière très pratique la théorie des développements limités (notés DL par la suite). La formule à retenir est la suivante : si ϵ est petit, on a

$$f(x+\epsilon) = f(x) + \epsilon \left[\frac{df}{dx}\right]_x + \frac{\epsilon^2}{2!}\left[\frac{d^2f}{dx^2}\right]_0 + \frac{\epsilon^3}{3!}\left[\frac{d^3f}{dx^3}\right]_0 + \dots \tag{1}$$

Le symbole $[]_x$ signifie que les dérivées successives doivent être évaluées en x. Il faut bien sur que la fonction f soit dérivable n fois en x pour appliquer cette formule à l'ordre n.

Les idées clefs de cette formule sont :

- L'ordre de grandeur de chaque terme diminue, car les puissances de ϵ augmentent et ϵ est supposé petit.
- Si on ne garde que les deux premiers termes, on trouve une approximation courante, à savoir qu'on approxime localement une fonction par sa tangente
- On peut parfois trouver l'écriture $f(x+\epsilon) = f(x) + \epsilon\frac{df}{dx} + O(\epsilon^2)$, ce qui veut dire que les termes supplémentaires sont d'ordre de grandeur ϵ^2.

Cette formule, que l'on peut démontrer exactement (formule de Taylor, 1714, voir par exemple `http://fr.wikipedia.org/wiki/Théorème_de_Taylor`), est à l'origine de nombreuses simplifications connues en mathématiques. Par exemple, on dit souvent en terminale que $lim_{x->0}\frac{e^x-1}{x} = 1$. Ceci peut se comprendre à l'aide d'un développement à l'ordre 1 (en ne gardant que le terme dérivé) de la fonction exponentielle :

$$e^x \quad \simeq e^0 + x\left[\frac{de^x}{x}\right]_0 \quad \simeq 1 + x \tag{2}$$

Donc $\frac{e^x-1}{x} = \frac{1+x-1}{x} = \frac{x}{x} = 1$.

Finalement, une autre façon de noter cette formule est :

$$f(x+\epsilon) = \sum_{k=0}^{\infty} \frac{\epsilon^k}{k!}\left[\frac{d^{(k)}f}{dx^k}\right]_x \tag{3}$$

Exemple 1 : polynôme de degré 2 Prenons l'exemple d'un polynôme d'ordre 2 : $f(x) = ax^2 + bx + c$. Si on réalise son DL à l'ordre 1, on obtient :

$$\begin{align}
f(x+\epsilon) &= f(x) + \epsilon\left[\frac{df}{dx}\right]_x \tag{4}\\
&= ax^2 + bx + c + \epsilon(2ax+b) \tag{5}\\
&= ax^2 + (b+2a\epsilon)x + c + b\epsilon \tag{6}\\
& \tag{7}
\end{align}$$

À l'ordre 2, on obtient :

$$\begin{align}
f(x+\epsilon) &= f(x) + \epsilon\left[\frac{df}{dx}\right]_x + \frac{\epsilon^2}{2}\left[\frac{d^2f}{dx^2}\right]_x \tag{8}\\
&= ax^2 + bx + c + \epsilon(2ax+b) + a\epsilon^2 \tag{9}\\
&= ax^2 + (b+2a\epsilon)x + c + b\epsilon + a\epsilon^2 \tag{10}\\
&= ax^2 + 2a\epsilon + a\epsilon^2 + bx + \epsilon x + c \tag{11}\\
&= a(x+\epsilon)^2 + b(x+\epsilon) + c \tag{12}
\end{align}$$

On voit que le développement limité à l'ordre 2 d'un polynôme de degré 2 est le polynôme initial, et donc dans ce cas le DL est exact dès l'ordre 2. Ceci est normal dans le cas des polynômes, puisque toutes les dérivées d'ordre supérieur à n sont nulles pour un polynôme de degré n ; son DL à l'ordre n est donc exactement égal à lui même. Pour une fonction infiniment dérivable, son DL à l'ordre n est donc l'approximation polynomiale de degré n de la fonction en un point précis.

Exemple 2 : fonction exponentielle Le DL de la fonction exponentielle $f(x) = e^x$ est particulièrement intéréssant, car il nous donne une formule de développement de l'exponentielle en série entière particulièrement utile par la suite. On applique la formule générale du DL à la fonction exponentielle, en considérant un accroissement x et un point de départ 0 :

$$\begin{aligned} e^x = e^{0+x} = e^0 + x \left[\frac{d(e^x)}{dx}\right]_0 + \frac{x^2}{2}\left[\frac{d^2(e^x)}{dx^2}\right]_0 + \dots & \quad (13) \\ = \sum_{k=0}^{n} \frac{x^k}{k!}\left[\frac{d^k(e^x)}{dx^k}\right]_0 & \quad (14) \\ = \sum_{k=0}^{n} \frac{x^k}{k!}\left[e^x\right]_0 & \quad (15) \\ = \sum_{k=0}^{n} \frac{x^k}{k!} & \quad (16) \end{aligned}$$

Cette dernière formule est très employée dans les problèmes de combinatoire.

Exemple 3 : la fonction sinus Ici on représente simplement visuellement les DL à différents ordres de la fonction sinus, en $x = 0$, fig **??**. Seuls les ordres impairs sont représentés, car les DL de sinus à l'ordre pair n'apportent que des termes nuls (i.e en $\sin(0)$). On peut voir que quand l'ordre du DL augmente, la précision fait de même.

1.2 Moments d'une densité de probabilité

Les moments d'une v.a. X de loi de probabilité $p(x)$ sont définis comme :

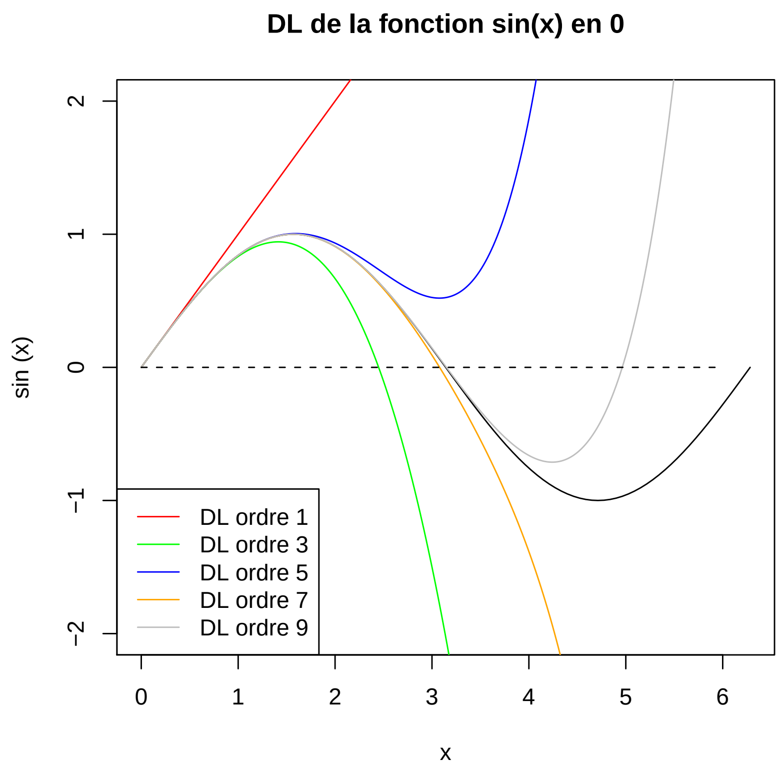

FIGURE 1 – Développements limités de $\sin(x)$ à différents ordres

$$\mu_X^n = \sum x^n p(x), \tag{17}$$

$$\mu_X^n = \int x^n p(x) dx. \tag{18}$$

On voit le lien entre le moment d'ordre 1 et l'espérance, le moment d'ordre 2 et la variance (le calcul montre que $\mu_X^1 = \mathbb{E}(X)$ et $\mu_X^2 = \mathbb{V}(X) + \mathbb{E}^2(X)$). Le moment d'ordre 0, dans le cas d'une distribution de probabilité, vaut 1 : c'est l'aire sous la courbe. Le moment d'ordre 3 est lié à l'asymétrie. Les interprétations géométriques des moments d'ordre supérieur sont possibles, mais moins intuitives. Un des usages pratiques des moments d'ordre 3 et 4 est de tester la normalité de données, à l'aide du test de Jarque-Bera.

Le théorème de Lévy (non démontré) nous dit que connaître tous les moments d'une fonction, c'est la connaître parfaitement, et donc on peut décrire une fonction à partir de ses moments. Intuitivement, cela implique que connaître parfaitement l'espérance, la variance, l'asymétrie ... d'une loid e probabilité est équivalent à en connaître la formule exacts. Ceci n'est pas très utile en pratique s'il faut connaître une infinité de valeurs pour décrire une fonction, mais on peut contourner cette difficulté à l'aide de la fonction génératrice des moments (FGM), notée $M_X(t)$. Cette fonction permet de retrouver tous les moments d'une distribution de probabilité, et permet donc de caractériser complètement une loi de probabilité. On présente ici les formules pour une variable continue (la généralisation au discret est laissée aux étudiants) :

$$M_X(t) = \mathbb{E}(e^{tx}) = \int e^{tx} p(x) dx. \tag{19}$$

On peut voir que :

$$\left[\frac{dM_X(t)}{dt}\right]_{t=0} = \int xp(x)dx = \mu_X^1, \tag{20}$$

$$\left[\frac{d^2M_X(t)}{dt^2}\right]_{t=0} = \int x^2p(x)dx = \mu_X^2, \tag{21}$$

$$\left[\frac{d^nM_X(t)}{dt^n}\right]_{t=0} = \int x^np(x)dx = \mu_X^n, \tag{22}$$

d'où le nom de génératrice des moments. D'après le théorème de Lévy, une fonction est définie par ses moments, et donc directement par sa fonction génératrice des moments. Notamment, deux fonctions ayant les mêmes FGM sont identiques, sous certaines conditions. Le calcul de la FGM est donc une transformation appliquée à une fonction qui maintient "l'identité" de la fonction. Cette propriété est très utile, car pour montrer que deux fonctions sont égales, on peut chercher soit à démontrer directement leur égalité, soit à démontrer celle de leurs FGM. Leurs FGM étant des espérances, leur calcul est assez aisé, et c'est à l'aide de ce type de calcul que l'on va démontrer le théorème central limite.

1.3 Moments de la loi normale

On rappelle l'écriture de la densité de probabilité de la loi normale dans le cas général, pour une moyenne μ et une variance σ^2 :

$$p(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}. \tag{23}$$

On calcule la FGM pour une loi normale centrée réduite (moyenne 0, écart-type 1) :

$$M_z(t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{zt} e^{-\frac{z^2}{2}} dz \tag{24}$$

On revient à la forme canonique en écrivant $(z - t)^2 = z^2 - 2tz + t^2$. On a alors :

$$\begin{aligned}
M_z(t) &= \textstyle\int_{\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2} - \frac{(z-t)^2}{2}} dz && (25)\\
&= e^{\frac{t^2}{2}} \textstyle\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} dz && (26)\\
&= e^{\frac{t^2}{2}} \text{ car on reconnaît une intégrale gaussienne.} && (27)
\end{aligned}$$

On calcule maintenant les moments d'ordre 1,2 et 3 pour une loi normale centrée réduite :

$$\begin{aligned}
\left[\frac{dM_z(t)}{dt}\right]_{t=0} &= \mu_z^1 = \left[te^{\frac{t^2}{2}}\right]_{t=0} &&= 0, && (28)\\
\left[\frac{d^2M_z(t)}{dt^2}\right]_{t=0} &= \mu_z^2 = \left[e^{\frac{t^2}{2}} + te^{\frac{t^2}{2}}\right]_{t=0} &&= 1, && (29)\\
\left[\frac{d^3M_z(t)}{dt^3}\right]_{t=0} &= \mu_z^3 = \left[e^{\frac{t^2}{2}} + 2te^{\frac{t^2}{2}} + t^3e^{\frac{t^2}{2}}\right]_{t=0} &&= 0. && (30)
\end{aligned}$$

On trouve bien les valeurs attendues pour une distribution de moyenne 0, d'écart-type 1 et symétrique. Attention, l'asymétrie n'est pas donnée directement par le moment d'ordre 3, mais est dépend de $\mu_z^3 - \mathbb{E}^3(X)$; ici l'espérance étant nulle, on trouve la synétrie pour $\mu_z^3 = 0$.

On peut retrouver la génératrice des moments d'une loi normale quelconque en utilisant la loi de transformation, en posant une nouvelle v.a. X de moyenne μ et d'écart-type σ^2 calculée à partir de la v.a. Z centrée réduite, par décentrage et dé-réduction $X = \sigma Z + \mu$. On peut alors vérifier aisément que :

$$M_X(t) = M_{\sigma Z+\mu}(t) = e^{\mu t} M_{\sigma Z}(t) = e^{\mu t + \sigma^2 \frac{t^2}{2}} \tag{31}$$

1.4 Énoncé et démonstration du théorème central limite

Un énoncé du théorème central limite (TCL) est : *Toute somme de n variables aléatoires indépendantes converge vers une loi normale avec n.*

On va ici démontrer un sous-cas, à savoir que la moyenne de n tirages d'une v.a. X suit une loi de probabilité normale. La démonstration du cas le plus général est plus complexe et ne sera pas abordée.

Démonstration numérique du TCL Ici on va montrer par un exemple numérique que la moyenne du tirage de n variables uniformes a une distribution normale. Pour bien comprendre comment cet effet se développe avec un n grandissant, on trace la distribution de la moyenne de n v.a. de distribution uniforme entre 0 et 1, pour $n = 1, 2, 10, 100$.

Pour démontrer le TCL numériquement on va générer sous R une moyenne de variables aléatoires uniformes. On va observer que, même si la loi de chaque variable est une loi uniforme sur $[0, 1]$, la loi de la moyenne est une loi normale.

On commence par générer une v.a. uniforme comprise entre 0 et 1 en utilisant la fonction `runif` de R :

```
> runif(1)

[1] 0.1509883

> x <- runif(1e+05)
> head(x)

[1] 0.03770831 0.42499234 0.70435300 0.79625655 0.80666984 0.44392315

> mean(x)

[1] 0.4995604
```

On regarde ensuite la distribution de la moyenne de n tirages, en réalisant un grand nombre de fois (50000) chacun de ces n tirages et en étudiant leur distribution (fig 2). Dans cette figure la densité de probabilité d'une loi normale est tracée en rouge par-dessus la densité de probabilité estimée pour une moyenne n v.a. uniformes, pour faciliter les comparaisons. On voit que

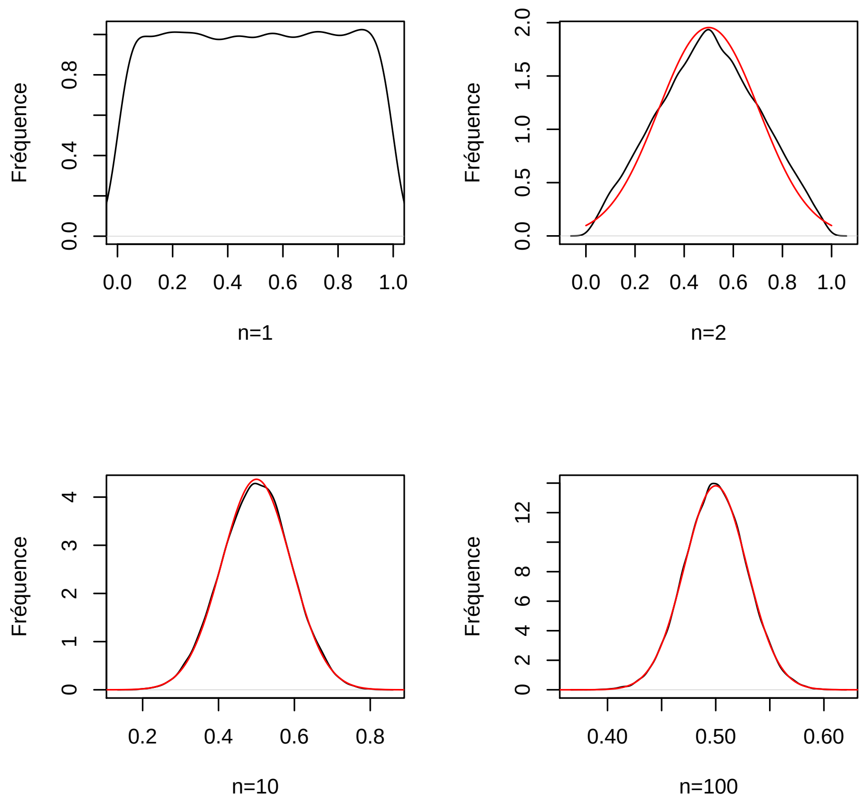

Figure 2 – Convergence de la distribution d'une moyenne de n v.a. uniformes vers la loi normale

dès que $n = 10$, l'approximation est très bonne. L'idée clef ici est que même si la distribution de X est complètement uniforme, la moyenne de nombreux tirages de X va forcément tendre vers la moyenne d'un tirage, et que la probabilité d'observer un grand nombre de tirages tous en-dessous ou au-dessus de la moyenne attendue est très faible et décroît expontentiellement avec n.

Par exemple, pour un dé à 6 faces, si on le jette une fois, on ne pourra rien dire sur le résultat. Mais on sait bien que si on additionne deux dés, on a plus de chances de faire 7, et très peu de chances de faire 2 ou 12. Quand n augmente, ceci devient de plus en plus vrai.

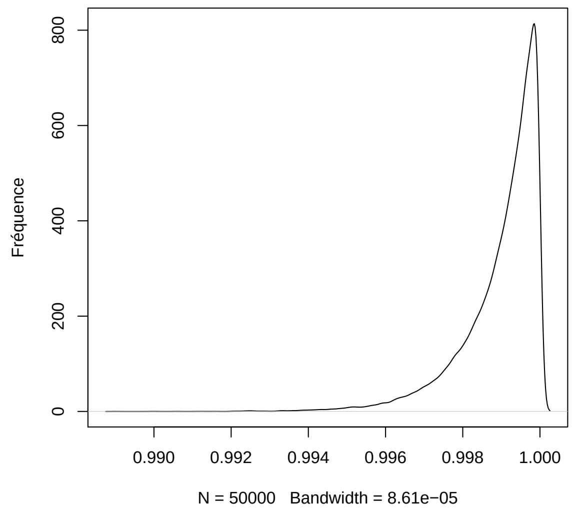

FIGURE 3 – Distribution du maximum du tirage de 1000 valeurs d'une v.a. uniformes

Il faut noter que cette convergence n'est valable que pour une *somme* de v.a. ;par exemple, si on regarde la distribution du maximum obtenu pour n tirages d'une v.a. uniforme, on obtient la figure 3, où l'on voit clairement que la distribution n'est pas une loi normale [1].

1.5 Démonstration du TCL par la méthode des moments

Soit X une v.a.de loi inconnue, de moyenne μ et de variance σ^2. On définit la v.a. Y comme la moyenne sur n tirages de X, à savoir $Y = \frac{1}{n}\sum_{i=1}^{n} x_i$. D'après les propriétés de l'espérance et de la variance, on a $\mathbb{E}(Y) = \mu$ et $\mathbb{V}(Y) = \frac{\sigma^2}{n}$, mais on ne peut a priori rien affirmer sur la distribution de Y, à part cela. On va donc calculer la fonction génératrice des moments de Y pour gagner plus d'informations.

1. La distribution des valeurs extrêmes est un champ de recherche très actif, notamment en assurances, en finances, et dans la prévention des catastrophes

$$
\begin{align}
M_Y(t) &= \mathbb{E}\left(\exp\left(\frac{t}{n}\sum_i x_i\right)\right) \tag{32}\\
&= \mathbb{E}\left(\prod_i \exp\left(\frac{tx_i}{n}\right)\right) \tag{33}\\
&= \prod_i \mathbb{E}\left(\exp\left(\frac{tx_i}{n}\right)\right) \tag{34}\\
&= \left[M_X\left(\frac{t}{n}\right)\right]^n \text{car les tirages sont indépendants.} \tag{35}
\end{align}
$$

On calcule $\left[M_X(\frac{t}{n})\right]$ par un développement limité d'ordre 3 (ce qui fonctionne car $\frac{t}{n} \to 0$ quand $n \to \infty$). On obtient :

$$
\begin{align}
M_X\left(\frac{t}{n}\right) = M_X(0) + \frac{t}{n}\left[\frac{dM_X}{dt}\right]_{t=0} + \frac{t^2}{2n^2}\left[\frac{d^2M_X}{dt^2}\right]_{t=0} + O\left(\left(\frac{t}{n}\right)^3\right) \tag{36}\\
= 1 + \frac{t}{n}\mu + \frac{t^2}{2n^2}\left(\sigma^2 + \mu^2\right) + O\left(\left(\frac{t}{n}\right)^3\right). \tag{37}
\end{align}
$$

On revient ensuite à la formule 35, et on prend le logarithme pour simplifier la puissance, et on utilise le DL à l'ordre 2 du logarithme $ln(1+x) = x - \frac{x^2}{2} + O(x^3)$:

$$
\begin{align}
\ln M_Y(t) = n \ln M_x(t) = n \ln\left(1 + \frac{t}{n}\mu + \frac{t^2}{2n^2}\left(\sigma^2 + \mu^2\right)\right) \tag{38}\\
= n\left(\frac{t}{n}\mu + \frac{t^2}{2n^2}\left(\sigma^2+\mu^2\right) - \frac{1}{2}\left(\frac{t}{n}\mu + \frac{t^2}{2n^2}\left(\sigma^2+\mu^2\right)\right)^2 + O\left(\left(\frac{t}{n}\right)^3\right)\right) \tag{39}
\end{align}
$$

Dans ce dernier terme, on voir que le développement de la parenthèse va donner des termes proportionnels à $\left(\frac{t}{n}\right)^2$, à $\left(\frac{t}{n}\right)^3$ et à $\left(\frac{t}{n}\right)^4$. Sachant qu'on a décidé de faire le DL en s'arrêtant à l'odre 3 non compris, on va négliger comme "petits" les termes d'ordre supérieur et obtenir la formule suivante :

$$\ln M_Y(t) = n\left(\frac{t}{n}\mu + \frac{t^2}{2n^2}\left(\sigma^2+\mu^2\right) - \frac{1}{2}\left(\frac{t}{n}\mu\right)^2 + O\left(\left(\frac{t}{n}\right)^3\right)\right) \tag{40}$$

$$= t\mu + \frac{t^2}{n}\sigma^2, \text{d'où} \tag{41}$$

$$M_Y(t) = e^{t\mu+\frac{\sigma^2}{n}}. \tag{42}$$

On retrouve bien la fonction génératrice des moments d'une loi normale de moyenne μ et de variance $\frac{\sigma^2}{n}$. D'après le théorème de Lévy, deux fonctions ont la même FGM si et seulement si elles sont égales. On a donc bien égalité en loi de la moyenne d'une v.a. quelconque X et d'une loi normale, si n est assez grand ; autrement dit "toute somme de v.a. indépendantes et identiquement distribuées converge vers une loi normale". Ceci explique l'importance de la loi normale dans toutes les applications statistiques, et pourquoi on peut souvent raisonnablement supposer que des données issues d'un processus complexe suivent une loi normale. La démonstration peut se généraliser à toute somme de v.a. indépendantes qui ont un moment d'ordre 2, même si elles sont différentes entre elles (c-à-d. pas identiquement distribuées).