M1 MEEF Mathématiques

TD6 - Loi des grands nombres - Théorème de De Moivre Laplace - Théorème de convergence de la loi binomiale vers la loi de Poisson - Intervalles de fluctuation

Le cours

- **Loi faible des grands nombres.** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles définies sur un même espace probabilisé $(\Omega,\, \mathfrak{F},\, \mathbb{P})$, indépendantes, de même loi et intégrables. Alors $(X_1+\cdots+X_n)/n$ converge en probabilité vers $\mathbb{E}(X_1)$, i.e., pour tout $\varepsilon>0$,
$$\lim_{n\to\infty}\mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n}-\mathbb{E}(X_1)\right|\geq\varepsilon\right)=0$$
Remarque : Dans le cas de variables aléatoires de carré intégrable ($\mathbb{E}({X_1}^2)<+\infty$*), une preuve basée sur l'inégalité de Bienaymé-Tchebychev donne une estimation de la vitesse de convergence :* $\mathbb{P}(|(X_1+\cdots+X_n)/n-\mathbb{E}(X_1)|\geq\varepsilon)\leq\sigma^2/(n\varepsilon^2)$.
- **Loi forte des grands nombres.** On fait les mêmes hypothèses que ci-dessus. Alors, pour presque tout $\omega\in\Omega$, $X_1(\omega)+\cdots+X_n(\omega)/n$ tend vers $\mathbb{E}(X_1)$ quand $n\to\infty$.
Remarque : Ce théorème est plus fort que le précédent. Cependant, la loi faible des grands nombres est intéressante pour sa preuve très simple dans le cas de variables aléatoires de carré intégrable, ainsi que pour l'estimation de la vitesse de convergence dans ce cas.
- **Convergence de la loi binomiale vers la loi de Poisson.** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles telles que $X_n\sim\mathcal{B}(n,\, p_n)$ et telles que $\lim_{n\to\infty}np_n=\lambda$. Alors, pour tout $k\in\mathbb{N}$,
$$\lim_{n\to\infty}\mathbb{P}(X_n=k)=\frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\,.$$
Aplications numériques : Dans la pratique, on admet que l'approximation est satisfaisante si $n\geq 30$ *et si* $\inf\{p,\, 1-p\}\leq 0,1$ *et* $\inf\{np,\, n(1-p)\}\leq 10$.
- **Quelques propriétés de la loi normale**
  – Si $X\sim\mathcal{N}(m,\,\sigma)$, alors $\frac{X-m}{\sigma}\sim\mathcal{N}(0,\,1)$.
  – Si $X\sim\mathcal{N}(m,\,\sigma)$, alors
$$\mathbb{P}(|X-m|\geq\sigma)\simeq 0,683,\quad \mathbb{P}(|X-m|\geq 2\sigma)\simeq 0,954,\quad \mathbb{P}(|X-m|\geq 3\sigma)\simeq 0,997\,.$$
  – Si $X_1\sim\mathcal{N}(m_1,\,\sigma_1)$ et $X_2\sim\mathcal{N}(m_2,\,\sigma_2)$ sont deux variables aléatoires normales indépendantes, alors
$$X_1+X_2\sim\mathcal{N}(m_1+m_2,\,\sqrt{\sigma_1^2+\sigma_2^2})\quad\text{et}\quad X_1-X_2\sim\mathcal{N}(m_1-m_2,\,\sqrt{\sigma_1^2+\sigma_2^2})\,.$$
  Ces propriétés se généralisent au cas de n variables aléatoires normales indépendantes.
- **Théorème de De Moivre-Laplace.** Soient $p\in]0,\,1[$ et $X\sim\mathcal{B}(n\,p)$, alors
$$\mathbb{P}(X=k)=\frac{1}{\sqrt{2\pi p(1-p)}}\,\mathrm{e}^{-\frac{(k-np)^2}{2np(1-p)}}\,(1+\mathcal{O}(1/\sqrt{n}))\,,$$
si $(k-np)/\sqrt{np(1-p)}$ reste borné.
On a aussi
$$\mathbb{P}\left(a<\frac{X-np}{\sqrt{np(1-p)}}\leq b\right)=\frac{1}{\sqrt{2\pi}}\int_a^b\mathrm{e}^{-z^2/2}\mathrm{d}z\,(1+\mathcal{O}(1/\sqrt{n}))\,.$$
Aplications numériques : Dans la pratique, on admet que l'approximation est satisfaisante si $n\geq 30$ *et* $\inf\{np,\, n(1-p)\}\geq 10$.
Remarque : Ce résultat admet une généralisation au cas de variables aléatoires quelconques. C'est le résultat connu sous le nom de théorème de la limite centrale

- **Théorème de la limite centrale.** Soient $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires réelles sur un espace probabilisé $(\Omega,\, \mathfrak{F},\, \mathbb{P})$, indépendantes, et de même loi. Alors, pour tout $-\infty \leq a < b \leq +\infty$, on a

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_1+X_2+\cdots+X_n-n\mathbb{E}(X_1)}{\sqrt{n\sigma^2}} \in [a,\, b]\right) = \frac{1}{\sqrt{2\pi}}\int_a^b \mathrm{e}^{-z^2/2}\mathrm{d}z\,. \tag{1}$$

Remarques : i) Ce théorème est fondamental en théorie des probabilités. Dans certaines applications numériques de ce théorème, on considère que la différence entre la limite quand n tend vers l'infini dans le membre de droite de (1) *et la valeur pour un n donné "suffisamment grand", est négligeable. Il reste cependant clair que le théorème énoncé ici ne permet nullement d'estimer cette différence. Dans la pratique, on remplace la loi de $\frac{X_1+X_2+\cdots+X_n-n\mathbb{E}(X_1)}{\sqrt{n\sigma^2}}$ par la loi normale centrée réduite lorsque $n \geq 30$.*

*ii) Lorsqu'on approxime la loi de $\frac{X_1+X_2+\cdots+X_n-n\mathbb{E}(X_1)}{\sqrt{n\sigma^2}}$ par la loi normale centrée réduite, dans le cas où les variables aléatoires X_i suivent une loi discrète, on approxime une v.a. discrète par une v.a. à densité, ce qui implique une erreur d'approximation de plus selon la probabilité à estimer. On peut alors appliquer une **correction de continuité** pour atténuer cette erreur. Un cas typique : Si $X_i \sim \mathcal{B}(1, 1/3)$ et $U \sim \mathcal{N}(0,\,1)$, on a $\mathbb{P}\left(\frac{X_1+X_2+\cdots+X_{300}-300\times 1/3}{\sqrt{300\times 1/3\times 2/3}} = 8\sqrt{\frac{3}{2}}\right) \neq 0$ alors que $\mathbb{P}(U = 8\sqrt{\frac{3}{2}}) = 0$. On utilise alors une correction de continuité en tenant compte du "pas" entre deux valeurs possibles. Ici, le pas entre deux valeurs possibles de $\frac{X_1+X_2+\cdots+X_n-n\mathbb{E}(X_1)}{\sqrt{n\sigma^2}}$ est de $\frac{1}{10}\sqrt{\frac{3}{2}}$. Donc on approxime $\mathbb{P}\left(\frac{X_1+X_2+\cdots+X_{300}-300\times 1/3}{\sqrt{300\times 1/3\times 2/3}} = 8\sqrt{\frac{3}{2}}\right)$ par $\mathbb{P}(U \in [8\sqrt{\frac{3}{2}} - \frac{1}{20}\sqrt{\frac{3}{2}},\, 8\sqrt{\frac{3}{2}} + \frac{1}{20}\sqrt{\frac{3}{2}}])$.*

*iii) Une des applications directes du théorème de la limite centrale est la dérivation de **la loi d'échantillonage de la moyenne ou de la fréquence pour des échantillons non exhaustifs** (c.f. exercice 8).*

iv) Une autre application du théorème de la limite centrale est l'estimation par intervalle de confiance. Ceci sera abordé ultérieurement.

Les exercices

Autour de la loi des grands nombres

Exercice 1 (Méthode de Monte-Carlo)**.** Soit g une fonction réelle continue sur $[0,1]$, et soit $\{X_n\}_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées, suivant la loi uniforme sur $[0,1]$. Expliciter $\mathbb{E}(g(X_n))$. Montrer :

$$\mathbb{P}\left[\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n g(X_k(\omega)) = \int_0^1 g(x)dx\right] = 1\ .$$

En quoi cette formule peut-elle être utile pour le calcul de $\int_0^1 g(x)dx$?

Convergence de la loi binomiale vers la loi de Poisson

Exercice 2. Un hotelier loue ses 56 chambres à l'avance pour le mois d'août. Ayant remarqué que chaque année 6% de ses clients se décommandent au dernier moment, il décide d'accepter 60 réservations.

i) Donner une valeur aprochée de la probabilité que tous les clients qui se présentent soient logés.

ii) De combien de chambres devrait-il disposer pour être sûr à 95% de faire face à ses engagements (toujours dans le cas de 60 réservations).

Exercice 3 (Extrait de V. Girardin, N. Limnios, Probabilité, "Variables, Vecteurs et Suites aléatoires", Vuibert)**.** Un indiviu vacciné sur mille fait une réaction allergique. On vaccine 2000 personnes. Quelle est la probabilité d'observer k réactions. Donner la valeur numérique exacte pour $k = 12$. Donner une approximation numérique de cette probabilité pour $k = 12$.

Théorème de la limite centrale

Exercice 4 (Savoir lire les tables de la loi normale).
i) Soit U une variable aléatoire normale centrée et réduite.

Calculer $\mathbb{P}[U < 2]$, $\mathbb{P}[U > 2]$, $\mathbb{P}[U < -2]$, $\mathbb{P}[-1 < U < 0.5]$ et $\mathbb{P}[4U \geq 3]$. Déterminer a et b tels que $\mathbb{P}[|U| < a] = 0.82$ et $\mathbb{P}[U < -b] = 0.6$

ii) Une variable aléatoire X suit une loi de Laplace-Gauss de moyenne 5 et de variance $\sigma^2 = 9$.

Si $U = \mathcal{N}(0,1)$, pour a donné, exprimer $\mathbb{P}(X \leq a)$ en fonction de $\mathbb{P}(U \leq b)$, où b est un réel qui dépend de a et que l'on déterminera.

Calculer les probabilités des événements suivants.

X inférieur à 8.
X supérieur à 2.
X compris entre -1 et 11.
X extérieur à l'intervalle $[-4, 14]$.

Exercice 5. (Appliquer le théorème de De Moivre - Laplace)

Parmi les pièces produites par une chaîne de montage, 10% sont défectueuses. Estimer la probabilité que parmi 400 pièces, plus de 50 soient défectueuses.

Exercice 6. (Appliquer le théorème de De Moivre - Laplace)

Un astronome souhaite mesurer la distance d, en années lumière, entre son observatoire et une étoile lointaine. Bien qu'il connaisse une technique de mesure, il sait aussi que chaque résultat ne constitue qu'une valeur approchée de la distance réelle d, en raison des influences atmosphériques et des erreurs des appareils de mesure. Par conséquent, notre astronome prévoit d'effectuer un nombre N de mesures et d'accepter leur moyenne comme estimation de la distance réelle. Il a des raisons de penser que les différentes valeurs mesurées sont des variables aléatoires indépendantes et identiquement distribuées, d'espérance commune d (d est la distance réelle), et de variance commune estimée à 4 (l'unité étant l'année lumière).

i) Soit X_i la variable aléatoire qui décrit la $i$ème mesure. A l'aide des X_i et de d, écrire l'événement : “La valeur moyenne de N observations est distante de d de 0,5 au plus”.

ii) A l'aide du théorème de la limite centrale, estimer la valeur N minimale pour que la moyenne des N mesures effectuées s'écarte de la valeur réelle d au plus de 0,5 années lumière, avec une probabilité de 95%.

Exercice 7. (Appliquer le théorème de De Moivre - Laplace)

Une compagnie aérienne donne des réservations sur le vol d'un appareil de 400 places. La probabilité qu'un passager ayant réservé pour ce vol ne se présente pas à l'embarquement est de $0,08 = 8\%$.

i) Si la compagnie accorde 420 réservations sur ce vol, quel est le risque de “surbooking”, i.e., quelle est la probabilité que se présentent plus de passagers que les 400 qui pourront embarquer ? (On commencera par modéliser ce problème à l'aide d'une famille de variables aléatoires X_i, $i = 1, ..., 420$, indépendantes et de même loi, et dont on choisira la loi en s'aidant de la question a)).

ii) Quel est le nombre maximum de billet que doit vendre la compagnie pour être sûre à 99% que tous les passagers qui se présenteront à l'embarquement avec un billet aient une place dans l'avion ?

Echantillonage - Intervalles de fluctuations (Programmes Lycée-BTS)

Exercice 8 (Loi d'échantillonage de la fréquence pour des échantillons non exhaustifs - (F. Roche, F. Barny, “Mathématiques” BTS, Hachette)).

Une entreprise fabrique en grand nombre des ampoules électriques. Des études statistiques ont montré que 5% des ampoules fabriquées sont impropres à la vente. L'entreprise effectue des tests de contrôle : pour cela, elle prélève dans la production des échantillons non exhaustifs de n ampoules. On désigne par X la variable aléatoire qui à tout échantillon aléatoire associe le pourcentage d'ampoules impropres à la vente.

a) Montrer que la loi suivie par X peut être approximée par $\mathcal{N}(p, \sqrt{p(1-p)/n})$ où $p = 0,05$.

b) Dans cette question, on prend $n = 500$; on considère un nombre d'échantillons de 500 ampoules très grand. Quel est le nombre moyen d'ampoules impropres à la vente dans un échantillon ? Déterminer la probabilit{e $\mathbb{P}(X \leq 0,055)$.

c) Déterminer la valeur minimale n_0 de la taille d'un échantillon pour que la probabilité d'avoir une fréquence d'ampoules impropres à la vente appartenant à l'intervalle $[0,049; 0,051]$ dépasse 95%.

Exercice 9 (Échantillonage et intervalle de fluctuation)**.**

En classe de seconde, la propriété suivante est admise :

" Pour un échantillon de grande taille ($n \geq 30$) *d'une population ayant une proportion* p *d'un caractère comprise entre* 0,2 *et* 0,8*, l'intervalle*

$$I_{2^{nde}} = [p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}] \tag{2}$$

est une bonne approximation de l'intervalle de fluctuation au seuil de 95% *de la fréquence* f_n^* *du caractère observée sur l'échantillon".*

(1) Qu'est ce qu'un intervalle de fluctuation ?

(2) Rappeler l'énoncé du théorème de *De Moivre-Laplace.*

(3) Soit $p \in (0, 1)$. Montrer que

$$\frac{1}{\sqrt{p(1-p)}} \geq 2\,.$$

(4) On considère la famille de variables aléatoires (X_i) où X_i vaut 1 si le caractère est observé sur le i-ème individu et vaut 0 sinon.

  (a) Pour $p = 0,3$ et $n = 100$, calculer explicitement (i.e. sans utiliser l'intervalle $I_{2^{nde}}$ ci-dessus) un intervalle I_0 centré en p tel que la fréquence f_n^* sur un échantillon de taille n appartienne à cet intervalle I_0 avec probabilité supérieure ou égale à 0,95 (Considérer la loi binomiale et utiliser un tableur).

  (b) En utilisant (1) et (2), démontrer que $I_{2^{nde}}$ est bien un intervalle de fluctuation "approché" au seuil de $0,95$.

  Pour $p = 0,3$ et $n = 100$, calculer l'intervalle $I_{2^{nde}}$ donné par l'égalité (2) et comparez-le avec I_0.

(5) En utilisant la définition (2) de $I_{2^{nde}}$, déterminer la taille minimale d'un échantillon pour être sûr à 95% que la fréquence mesurée sur l'échantillon soit distante de la fréquence du caractère dans la population totale de moins de 1%. Même question avec $0,1\%$.

(6) Pour $p = 0,3$, à l'aide d'un tableur, effectuer un tirage d'échantillon de taille 100 et donner la fréquence f_n^* correspondante. Recommencez l'opération 100 fois et comptez le nombre de fois où vous avez obtenu une fréquence dans l'intervalle de confiance.

Exercice 10. Voici un extrait d'un document d'accompagnement des programmes de premières (cf. eduscol.education.fr/prog) :

L'intervalle de fluctuation au seuil de 95% d'une fréquence F, correspondant à la réalisation sur un échantillon aléatoire de taille n, de la variable aléatoire X égale à nF et de loi binomiale de paramètres n et p, est l'intervalle $\left[\frac{a}{n}, \frac{b}{n}\right]$ défini par le système de conditions suivant :

- a est le plus grand entier tel que $\mathbb{P}(X < a) \leq 0,025$,

- b est le plus petit entier tel que $\mathbb{P}(X > b) \leq 0,025$.

ou encore par le système de conditions équivalents :

- a est le plus petit entier tel que $\mathbb{P}(X \leq a) > 0,025$,
- b est le plus petit entier tel que $\mathbb{P}(X \leq b) \geq 0,975$.

Nous reprenons les mêmes notations. On a donc X qui suit la loi binomiale de paramètre n et p. On suppose que $0 < p < 1$.

(1) Montrer que les deux systèmes de conditions sur a et b ci-dessus sont équivalents.

(2) Avec a et b vérifiant les systèmes ci-dessus, montrer soigneusement que, pour tout $n \in \mathbb{N}^*$, on peut en déduire une inégalité entre $P(\frac{X}{n} \in [\frac{a}{n}, \frac{b}{n}])$ et 0,95 ?

(3) Toujours avec les mêmes nombres a et b, montrer soigneusement que pour tout $n \in \mathbb{N}^*$, on obtient une autre inégalité entre $P(\frac{X}{n} \in [\frac{a+1}{n}, \frac{b-1}{n}])$ et 0,95.

(4) Donner une expression de $P(X \in [a, b]) - P(X \in [a+1, b-1])$ en fonction de a, b, n et p.

Exercice 11 (Intervalle de fluctuation asymptotique au seuil $1 - \alpha$; prise de décision)**.** Dans le document “éduscol : Ressources pour la classe terminale générale et technologique, Probabilités et statistique”, on trouve l'énoncé suivant :

Théorème : Si la variable aléatoire X_n suit la loi $\mathcal{B}(n; p)$, avec $p \in]0, 1[$, alors pour tout réel α dans l'intervalle $]0, 1[$ on a :

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{X_n}{n} \in I_n\right) = 1 - \alpha\,,$$

où I_n désigne l'intervalle $\left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ et u_α désigne l'unique réel tel que $\mathbb{P}(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$ où Z suit la loi normale $\mathcal{N}(0, 1)$.

(1) A l'aide du théorème de De Moivre - Laplace, démontrer ce résultat.

(2) Prise de décision. Résoudre l'exercice joint en document annexe extrait de l'ouvrage “Odyssée Mathématique, 1ère S, Collection Hatier”.