Chapitre 5

Approximation des lois de probabilités

5.1 Approximation de la loi binomiale vers la loi de Poisson

La loi de Poisson est souvent utilisée comme approximation de certaines lois binomiales pour de grands échantillons, i.e. des lois binomiales correspondant ‘a des grands nombres n d'épreuves de Bernoulli. Il y a bien sûr quelques restrictions dont nous tairons ici les justifications théoriques, et le paramètre de la loi approximante doit être choisi de sorte que l'espérance soit celle de la loi binomiale approximée.

Soient $n \geq 1,\ 0 < p < 1$ et X suit la loi binomial $B(n,p)$ et supposons que le produit np tends vers $\lambda > 0$ lorsque n tends vers $+\infty$.[11]

Dans ce cas, pour n assez grand on peut remplacer p par $\dfrac{\lambda}{n}$. D'autre part si $k \in \{0, 1, \cdots, n\}$, alors :

$$P_X(k) = P(X = k) = C_n^k p^k q^{n-k}.$$

Ainsi pour tout $k \in \{0, 1, \cdots, n\}$, on a :

$$\begin{aligned}
P_X(k) &= \frac{n(n-1)\cdots(n-k+1)}{k!} p^k q^{n-k} \\
&= \frac{n^k\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{k!}\left(\frac{p}{q}\right)^k q^n \\
&= \frac{(np)^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^n
\end{aligned}$$

Mais

$$\lim_{n\to+\infty}\frac{(np)^k}{k!}=\frac{\lambda^k}{k!}\quad et\quad \lim_{n\to+\infty}\left(1-\frac{\lambda}{n}\right)^n=e^{-\lambda}.$$

Par conséquent

$$\lim_{n\to+\infty}P_X(k)=e^{-\lambda}\frac{\lambda^k}{k!},$$

qui est la fonction de densité de la loi de Poisson du paramètre λ.

Théorème 5.1.1.

Soit $n\geq 1,\ 0<p<1$ et X une variable aléatoire de la loi $B\left(n,\frac{\lambda}{n}\right)$, supposons que $np\to\lambda$ lorsque $n\to+\infty$ et $p\to 0$, alors :

$$B\left(n,\frac{\lambda}{n}\right)=P(\lambda).$$

Preuve 31.

On a :

$$\begin{aligned}P(X_n=k)&=C_n^k\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\&=\frac{\lambda^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k}e^{(n-k)\ln\left(1-\frac{\lambda}{n}\right)}\end{aligned}$$

Or,

$$n(n-1)\ldots(n-k+1)\underset{n\to+\infty}{\approx}n^k.$$

De plus, comme

$$\frac{\lambda}{n}\to 0\ et\ \left(1-\frac{\lambda}{n}\right)\underset{n\to+\infty}{\approx}-\frac{\lambda}{n}.$$

Donc

$$\lim_{n\to+\infty}P(X_{nb}=k)=\frac{\lambda^n}{n!}e^{-\lambda}$$

Ainsi, (X_n) converge bien en loi vers une variable qui suit la loi de Poisson de paramètre λ.

Remarque 5.1.2.

En pratique, on considère que si $n\geq 30$ et p est petit ($p\leq 0.1$), on peut approcher la loi $B(n,p)$ par la loi $P(np)$.

Exemple 5.1.1.

Soit X une variable aléatoire suivant la loi $B(100, 0.05)$. On souhaite calculer $P(X = 2)$.

Cas n=01 :valeur exacte par la loi binomiale.

$$P(X = 2) = C_{100}^{2}(0,05)^2(0,95)^2 \approx 0,0812.$$

Cas n=02 :valeur approchée par la loi $P(5)$.

$$P(X = 2)\frac{5^2}{2!}e^{-5} \approx 0,0843.$$

Exemple 5.1.2.

Dans un pays le pourcentage des personne atteints d'une certaine maladie génétique est de 0,4%.

Quelle est la probabilité de trouver au plus 2 personnes atteintes de cette maladie dans un village de 250 personnes ?.

Solution 19

Soit X la variable aléatoire qui donne le nombre de personnes atteinte de cette maladie dans ce village de 250 personnes. Il est clair qu'on cherche $P(X \leq 2)$ et que $X \hookrightarrow B(250, 0.004)$. Ainsi n et assez grand et p est petit, de plus le produit

$$np = 250 \times 0.004 = 1, \text{ donc si } \tilde{X} \hookrightarrow P(1), \text{ on a :}$$

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &\simeq P(\tilde{X} = 0) + P(\tilde{X} = 1) + P(\tilde{X} = 2) \\ &= e^{-1} + e^{-1} + \frac{e^1}{2} \\ &= \frac{5}{2}e^{-1}. \end{aligned}$$

Remarque 5.1.3.

Dans la pratique, ce résultat nous permet de remplacer la loi $B(n, p)$ par la loi $P(np)$ lorsque le produit np est très petit. L'avantage de la loi de poisson est qu'elle ne dépend que d'un seul paramètre.

5.2 Approximation d'une loi binomiale par une loi normale

Soit une loi binomiale $B(n,p)$ de moyenne $m = np$, d'écart-type $\sigma = \sqrt{npq}$.
On montre que l'on peut approximer la loi binomiale $B(n,p)$ par :
La loi normale $N(m = np, \sigma = \sqrt{npq})$, si $n \geq 15$, p et q étant non voisins de 0.
Dans la pratique, l'approximation est admise si $n \geq 20$, $np \geq 10, nq \geq 10$.

Théorème 5.2.1.
Soit $p \in]0,1[$ et $q = 1-p$. Soit (S_n) une suite de variables aléatoires telle que S_n suit la loi binomiale $B(n,p)$. Alors, la suite de variables aléatoires $S_n^* = \dfrac{S_n - np}{\sqrt{npq}}$ converge en loi vers la loi normale centrée $N(0,1)$.

Preuve 32.

Toute variable binomiale de paramètre (n,p) peut être considérée comme la somme de n variables aléatoires de Bernoulli de paramètre p mutuellement indépendantes. On a donc :

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

où X_k est une variable de Bernoulli de paramètre p. Comme les variables de Bernoulli admettent une espérance p et une variance pq, le Théorème de la limite centrée s'applique bien et donne le résultat voulu.

Remarque 5.2.2.

En pratique, on considère que si $n \geq 30$, $np \geq 5$ et $nq \geq 5$, on peut approcher la loi $B(n, p)$ par la loi $N(np, npq)$.

Si S_n suit une loi $B(n, p)$ avec n et p tels qu'on puisse approcher S_n par N_n qui suit la loi $N(np, npq)$, on devrait écrire :

$$\forall k \in [0, n], \; P(S_n = k) \approx P(S_n = k).$$

Mais comme N_n est une variable à densité, on a $P(N_n = k) = 0$, donc notre approximation ci-dessus n'est pas vraiment bonne...On écrira plutôt

$$P(S_n = k) \approx P(k - 0,5 \leq N_n \leq k + 0,5).$$

Et on appelle cette méthode utilisé la correction de continuité.

Exemple 5.2.1.

Soit X une variable qui suit la loi $B(900, 0.5)$. On cherche à calculer $P(405 \leq X \leq 495)$ Pour le calcul exact, il nous faudrait calculer des combinaisons avec de très grands nombres, ce qui nécessite un ordinateur et ne donne parfois qu'une valeur approchée. On remarque ici que l'on est dans les conditions où l'on peut approchée notre loi binomiale par une loi normale $N(450, 225)$ car $900 > 30$ et $900 \times 0.5 = 450 > 5$. Cependant pour la loi normale $N(450, 225)$, il n'est pas facile de calculer $P(405 < X < 495)$. On se ramène donc à une loi normale centrée réduite. On pose

$$X^* = \frac{X - 450}{\sqrt{225}},$$

et on a que X^* suit approximativement la loi normale centrée réduite. De plus,

$$P(405 < X_n < 495) = P\Big(-3 \leq \frac{X - 450}{15} \leq 3\Big) = P(-3 \leq X^* \leq +3) = \phi(3) - \phi(-3).$$

5.3 Approximation d'une loi du poisson par une loi normale

La loi de Poisson de paramètre λ est approximée par la loi normale de moyenne λ et variance λ, pour $\lambda \to \infty$.

En pratique, cette approximation est faite si $\lambda \geq 20$.

Si X est une variable aléatoire de loi Poisson de paramètre λ, avec $\lambda \geq 20$. Alors, on a l'approximation suivante :

$$P(a \leq X \leq b) \approx \Phi\Big(\frac{b-\lambda}{\sqrt{\lambda}}\Big) - \Phi\Big(\frac{a-\lambda}{\sqrt{\lambda}}\Big).$$

Théorème 5.3.1.

Soit α un réel strictement positif et soit (S_n) une suite de variables aléatoires telle que Sn suit la loi de Poisson $P(n\alpha)$. Alors, la suite de variables aléatoires $S_n^* = \dfrac{S_n - n\alpha}{\sqrt{n\alpha}}$ converge en loi vers la loi normale centrée réduite $N(0,1)$.

Preuve 33.

Toute variable de Poisson de paramètre $n\alpha$ peut être considérée comme la somme de n variables aléatoires de Poisson de paramètre α mutuellement indépendantes. On a donc :

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

où X_k est une variable de Poisson de paramètre α. Comme les variables de Bernoulli admettent une espérance α et une variance α , le Théorème de la limite centrée s'applique bien et donne le résultat voulu.

Remarque 5.3.2.

En pratique, on considère que si $\lambda \geq 18$, on peut approcher la loi $P(\lambda)$ par la loi $N(\lambda, \lambda)$.

Exemple 5.3.1.

Soit X une variable aléatoire suivant une loi $P(64)$ et on cherche $P(X \leq 74)$ Pour faire le calcul exact, il faut obligatoirement un ordinateur. On a donc intérêt à approcher la loi de X par la loi $N(64, 64)$ et donc $\frac{X-64}{8}$ suit la loi normale centrée réduite. On a donc :

$$\begin{aligned} P(X \leq 74) &= P\Big(\frac{X-64}{8} \leq \frac{74-64}{8}\Big) \\ &= P\Big(\frac{X-64}{8} \leq 1,25\Big) \\ &= \phi(1,25) \\ &= 0,8944. \end{aligned}$$

5.4 Approximation d'une loi du Khi-deux par une loi normale

La loi de Khi-deux à m degrés de liberté est approximée par la loi normale de moyenne m et variance $2m$, pour $m > 30$.

5.4.1 Résumé sur les approximations de lois

- $H(N, p, n) \approx B(n, p)$ pour $N > 10n$.
- $B(n, p) \approx P(\lambda = np)$ pour $n \geq 30,\ p \leq 0.1$ et $np \leq 10$.
- $B(n, p) \approx N(m = np,\ \sqrt{npq}$ avec $\begin{cases} n \geq 30 \\ 0,1 < p < 0,9 \end{cases}$ $ou \begin{cases} np \geq 10 \\ npq > 3. \end{cases}$
- $P(\lambda = np \approx) N(m = np,\ \sqrt{npq}$ pour $np \geq 10$.

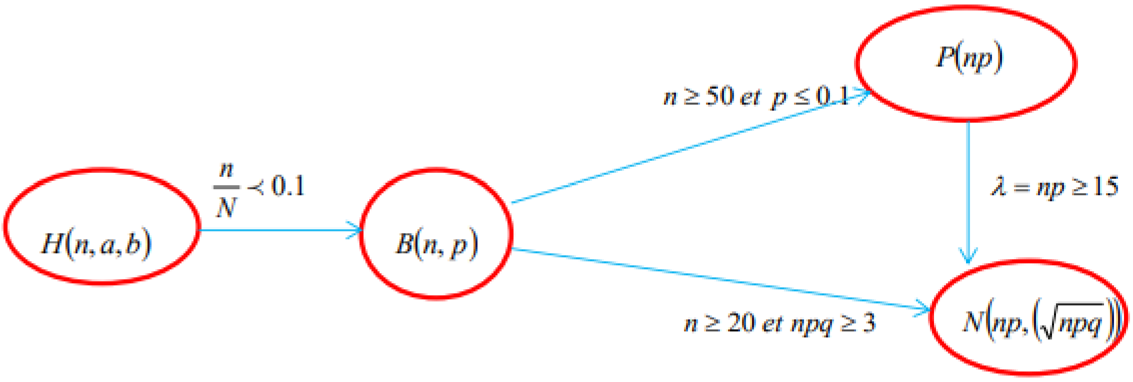

Tableau comparatif :

Dans ce tableaux on trouve une comparaison entre les déférents caractéristique des de types de variables aléatoire (discrète et continues).

X	Variable discrète	Variable à densité f
$X(\Omega)$	$\{x_1, x_2, \cdots, x_k, \cdots\}$	$\mathbb{R}$ ou un intervalle
$P(a \leq X \leq b)$	$\displaystyle\sum_{a \leq x_k \leq b} P(X = x_k)$	$\displaystyle\int_a^b f(t)dt$
$F(x) = P(X \leq x)$	$\displaystyle\sum_{x_k \leq x} P(X = x_k)$	$\displaystyle\int_{-\infty}^x f(t)dt$
$E[X]$	$\displaystyle\sum_{k=1}^{+\infty} x_k P(X = x_k)$	$\displaystyle\int_{-\infty}^{+\infty} tf(t)dt$
$E[g(X)]$	$\displaystyle\sum_{k=1}^{+\infty} g(x_k) P(X = x_k)$	$\displaystyle\int_{-\infty}^{+\infty} g(t)f(t)dt$
$E[X^2]$	$\displaystyle\sum_{k=1}^{+\infty} x_k^2 P(X = x_k)$	$\displaystyle\int_{-\infty}^{+\infty} t^2 f(t)dt$
$Var(X)$	$\displaystyle\sum_{k=1}^{+\infty} (x_k - E[X])^2 P(X = x_k)$	$\displaystyle\int_{-\infty}^{+\infty} (t - E[X])^2 f(t)dt$

5.5 Série 5 (Approximations des Lois de Probabilités)

- **Exercice 1**(Corrigé) :

Une usine fabrique des billes de diamètre 8mm. Les erreurs d'usinage provoquent des variations de diamètre. On estime, sur les données antérieures, que l'erreur est une variable aléatoire qui obeit à une loi normale les paramètres étant : moyenne : 0mm, écart-type : 0.02mm. On rejette les pièces dont le diamètre n'est pas compris entre 7.97mm et 8.03mm.

Quelle est la proportion de billes rejetées ?.

- **Exercice 2**(Corrigé) :

Des machines fabriquent des crêpes destinées à être empilées dans des paquets de 10. Chaque crêpe a une épaisseur qui suit une loi normale de paramètres $m = 0.6$mm et $\sigma = 0.1$. Soit X la variable aléatoire ≪épaisseur du paquet en mm≫.

Calculez la probabilité pour que X soit compris entre 6.3mm et 6.6mm.

- **Exercice 3**(Corrigé) :

Sur un grand nombre de personnes on a constaté que la répartition du taux de cholestérol suit une loi normale avec les résultats suivants : 56% ont un taux inférieur à 165 cg ; 34% taux compris entre 165 cg et 180 cg ; 10% ont un taux supérieur à 180 cg. Quelle est le nombre de personnes qu'il faut prévoir de soigner dans une population de 10 000 personnes, si le taux maximum toléré sans traitement est de 182 cg ?.

- **Exercice 4**(Corrigé) :

Pour chacune des variables aléatoires qui sont décrites ci-dessous, indiquez quelle est la loi exacte avec les paramètres éventuels (espérance, variance) et indiquez éventuellement une loi approchée.

1. Nombre annuel d'accidents à un carrefour donné où la probabilité d'accident par jour est estimée à 4 /365 .
2. Nombre de garçons dans une famille de 6 enfants ; nombre de filles par jour dans une maternité où naissent en moyenne 30 enfants par jour.
3. Dans un groupe de 21 personnes dont 7 femmes, le nombre de femmes dans une délégation de 6 personnes tirées au hasard.

5.5.1 Solution de série 5

• **Exercice 1 :**Énoncé.

La probabilité qu'une bille soit rejetée est, en notant D la variable aléatoire "diamètre",

$$p = 1 - P(7,97 \leq D \leq 8,03).$$

Or

$$\begin{aligned} P(7,97 \leq D \leq 8,03) &= P\Big(-\frac{0,03}{0,02} \leq \frac{D-8}{0,02} \leq \frac{0,03}{0,02}\Big) \\ &= F(1,5) - F(-1,5) \\ &= 0,8664. \end{aligned}$$

La proportion de billes rejetée est donc $p = 13,4\%$.

• **Exercice 2 :**Énoncé.

Par des méthodes analogues on trouve que la probabilité pour X soit compris entre 6,3mm et 6,6mm est 14,3.

• **Exercice 3 :**Énoncé.

Si X est de moyenne m et d'écrit-type σ, alors $Y = \dfrac{X-m}{\sigma}$ suit une loi centrée réduite. Donc,

$$\text{si } P(X \leq 165) \text{ alors } P\Big(\frac{X-m}{\sigma} \leq \frac{165-m}{\sigma}\Big) = 0,56.$$

Or on peut lire dans la table de Gauss $F(0,15) = 0,5596$.

De meme, si $P(X \geq 180)$ alors :

$$P\Big(\frac{X-m}{\sigma} \geq \frac{180-m}{\sigma}\Big) = 0,1.$$

Donc,

$$P\Big(\frac{X-m}{\sigma} \leq \frac{180-m}{\sigma}\Big) = 0,9.$$

et l'on peut lire de méme $F(1,28) = 0,8997$.

Pour trouver m et σ il suffit de résoudre le système d'équation :

$$\frac{165-m}{\sigma} = 0,15 \text{ et } \frac{180-m}{\sigma} = 1,28$$

d'où $\sigma \simeq 13,27$. $m \simeq 163cg$.Alors :

$$P(X \geq 182) = P\Big(\frac{X-m}{\sigma} \geq \frac{182-m}{\sigma}\Big) = 1 - F(1,43) = 0,0764.$$

Sur 10000 personnes on estime le nombre de personnes à soigner de l'ordre de 764 personnes, en fait la théorie de l'estimation donnera une fourchette.

• **Exercice 4 :** Énoncé.

1. Loi binomiale $B(365; \frac{4}{365})$, approchée par la loi de poisson de paramètre 4, d'espérance et variance 4.
2. Loi binomiale $B(6; \frac{1}{2})$ d'espérance 3 et variance $\frac{3}{2}$.
3. Loi hypergéométrique.

Bibliographie

[1] L-P. Arguin (2022). A first Course in Stochastic Calculs. AMS. — AMERICAN MATHEMATICAL SOCIETY, USA, 2022.

[2] E.Cantoni, P. Huber, E. Ronchetti (2006). Maîtriser l'aléatoire. Exercices résolus de probabilités et statistique. Springer-Verlag, France, Paris, 2006.

[3] H. Carrieu (2008). Probabilité. Exercices Corrigés. EDP Sciences, 2008.

[4] M. Cottrell, Ch. Duhamel et V. Genon-Catalot (1980). Exercices de Probabilités avec rappels de cours. Librairie classique Eugène Belin, 1980.

[5] Marie Cottrell, Valentine Genon-Catalot, Christian Duhamel et Thierry Meyre (1999). Exercices de probabilités Licence- maitrise- école d'ingénieurs. CASSINI, PARIS, 1999.

[6] D. Foata, A. Fuchs(1998). Calcul des probabilités. Dunod, 1998.

[7] M. Lejeune (2010). Statistique. La Théorie et ses applications. Deuxième édition. — Springer, 2010.

[8] M. Loève (1978). Probability Theory II. 4th Edition. — Springer-Verlag, New York, 1978.

[9] J. Neveu (1994). Introduction aux probabilités. École Polytechnique, Paris, 1994.

[10] C. Reidcher, R. Leblanc, B. Rémillard, D. Larocque (2002). Théorie des probabilities Problèmes et Solutions. Presses de l'Université du Québec, 2002.

[11] Tortrat (1971). Calcul des probabilités et introduction aux processus aléatoires. MASSON, 1971.

[12] Saporta, Gilbert (2006). probabilités, analyse des données et statistique, 2006. Editions technip.