III. Approximation des lois de probabilités

III.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 à 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 $n \geq 1, 0 \prec p \prec 1$ et X suit la loi binomial B(n,p) et supposons que le produit np tends vers λ>0 lorsque n tends vers $+\infty$. Dans ce cas, pour n assez grand on peut remplacer p par $\frac{\lambda}{n}$. D'autre part si $k \in \{0,1,\dots,n\}$, alors $P_X(k) = P(X = k) = C_n^k p^k q^{n-k}$

$$P_X(k) = \frac{n(n-1)\dots(n-k+1)}{k!} p^k q^{n-k}$$

Ainsi pour tout $k \in \{0,1,\dots,n\} = \dfrac{n^k\left(1-\frac{1}{n}\right)\dots\left(1-\frac{k-1}{n}\right)}{k!}\left(\dfrac{p}{q}\right)^k q^n$

$$= \frac{(np)^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^n$$

Mais $\lim\limits_{n\to+\infty} \dfrac{(np)^k}{k!} = \dfrac{\lambda^k}{k!}$ *et* $\lim\limits_{n\to+\infty}\left(1-\dfrac{\lambda}{n}\right)^n = e^{-\lambda}$

Par conséquent $\lim\limits_{n\to+\infty} P_X(k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$, qui est la fonction de densité de la loi de Poisson du paramètre λ.

Théorème :

Soit $n \geq 1, 0 \prec p \prec 1$ et X une variable aléatoire de la loi B (n,p).Supposons que $np \to \lambda$ lorsque $n \to +\infty$ et $p \to 0$,alors $B(n,p) \cong P(\lambda)$

Exemple :

Comparaison des fonctions de répartitions d'une loi B (100,0.1) et celle d'une loi de Poisson P(10)

Exemple 4.3. 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:

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 $I\!P(X \leq 2)$ et que $X \hookrightarrow \mathcal{B}(250, 0.004)$. Ainsi n et assez grand et p est petit, de plus le produit $np = 250 \times 0.004 = 1$, donc si $\widetilde{X} \hookrightarrow \mathcal{P}(1)$, on a

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

Remarque 4.5. Dans la pratique, ce résultat nous permet de remplacer la loi $\mathcal{B}(n,p)$ par la loi $\mathcal{P}(np)$ lorsque le produit np est très petit. L'avantage de la loi de Poisson est qu'elle ne dépend que d'un seul paramètre. Bien évidement ceci

III.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 ≥ 15, p et q étant non voisins de 0.

Dans la pratique, l'approximation est admise si n ≥ 20, np ≥ 10, nq ≥ 10.

III.3 Approximation d'une loi 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 λ ≥ 20.

Si X est une variable aléatoire de loi Poisson de paramètre λ, avec λ ≥ 20 ; alors, on a l'approximation suivante :

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

III.4 Approximation d'une loi 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.

Résumé sur les approximations de lois

- $H(N, p, n) \approx B(n, p) \quad pour\ N \succ 10n$
- $B(n, p) \approx P(\lambda = np)\ pour\ n \geq 30, p \leq 0.1\ et\ np \leq 10$
- $B(n, p) \approx N(m = np, \sigma = \sqrt{npq})\ avec \begin{cases} n \geq 30 \\ 0.1 \prec p \prec 0.9 \end{cases} ou \begin{cases} np \geq 10 \\ nq \geq 10 \end{cases} ou npq \succ 3$
- $P(\lambda = np) \approx N(m = np, \sigma = \sqrt{npq})\ pour\ np \geq 10$

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