Leçon 249 : Loi binomiale, loi de Poisson. Applications.

Cadre : $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

I. Loi binomiale

1) Définition et premières propriétés.

Déf : Une v.a.r X suit une loi de Bernoulli de paramètre $p \in [0,1]$ notée $\mathcal{B}(p)$ si $\mathbb{P}(X=1) = p$ et $\mathbb{P}(X=0) = 1-p$.

Soient $X_1, \dots, X_n$ des v.a.r i.i.d. de loi $\mathcal{B}(p)$, $S = X_1 + \dots + X_n$ est une v.a.r à valeurs dans $\{0, \dots, n\}$ vérifiant pour $k \in \{0, \dots, n\}$:

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

On appelle sa loi, la loi binomiale et on la note $\mathcal{B}(n,p)$.

Rem : Cette formule justifie le terme " binomiale "

Si $X \sim \mathcal{B}(n,p)$ alors $n - X \sim \mathcal{B}(n, 1-p)$.

Prop : On a $\mathbb{E}(S) = np$; $\mathbb{V}(S) = np(1-p)$ et pour tout $t \in \mathbb{R}$.

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = (pe^{it} + (1-p))^n.$$

Corollaire : Soient X_1, X_2 deux v.a.r indépendantes telles que $X_1 \sim \mathcal{B}(n_1, p)$ et $X_2 \sim \mathcal{B}(n_2, p)$ Alors $X_1 + X_2 \sim \mathcal{B}(n_1 + n_2, p)$.

Exemples : - On lance n fois une pièce de manière indépendante. On note $X_1, \dots, X_n$ les v.a valant 1 si le résultat du k-ième lancé est pile, 0 sinon. Ces v.a.r sont indépendantes et suivent une loi $\mathcal{B}(p)$ avec $p \in [0,1]$. Le nombre de pile $S = X_1 + \dots + X_n$ suit une loi binomiale $\mathcal{B}(n,p)$.

- Soient $X_1, \dots, X_n$ des v.a.r indépendantes et identiquement distribuées. Soit $A \in \mathcal{B}(\mathbb{R})$ un borélien, alors $\mathbb{1}_{\{X_1 \in A\}}, \dots, \mathbb{1}_{\{X_n \in A\}}$ sont des v.a.r indépendantes et de même loi $\mathcal{B}(\mathbb{P}(X_1 \in A))$. On obtient ainsi que : la v.a.r $S = \mathbb{1}_{\{X_1 \in A\}} + \dots + \mathbb{1}_{\{X_n \in A\}}$ suit une loi $\mathcal{B}(n, \mathbb{P}(X_1 \in A))$.

Application : Soient $X_1, \dots, X_n$ des v.a.r i.i.d. de loi sans atome. On note F leur fonction de répartition. (elle est continue), $(X_{(1)}, \dots, X_{(n)})$ la statistique d'ordre associée : on a alors pour tout $y \in \mathbb{R}$, pour tout $k \in \{1, \dots, n\}$

$$\mathbb{P}(X_{(k)} \leq y) = \sum_{i=k}^{n} \binom{n}{i} [F(y)]^i [1-F(y)]^{n-i}.$$

2) Loi multinomiale

Déf : Soient $p_1, \dots, p_k$ des réels tels que $0 < p_i < 1$ et $\sum_{i=1}^{k} p_i = 1$
$Z = (Z_1, \dots, Z_k)$ une v.a à valeur dans $\mathbb{N}^k$ telle que

$$\mathbb{P}(Z = (n_1, \dots, n_k)) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k} \quad \text{où } 0 \leq n_i \leq n \text{ et } \sum_{i=1}^{n} n_i = n.$$

On dit que Z suit une loi multinomiale de paramètre $p = (p_1, \dots, p_k)$ et n.

Prop : Soient $X_1, \dots, X_n$ des v.a.r i.i.d et $\mathbb{R} = \bigsqcup_{i=1}^{k} A_i$ une partition de $\mathbb{R}$ par des boréliens. On note $S_1 = \sum_{l=1}^{n} \mathbb{1}_{\{X_l \in A_1\}}, \dots, S_k = \sum_{l=1}^{n} \mathbb{1}_{\{X_l \in A_k\}}$
Alors $S = (S_1, \dots, S_k)$ suit une loi multinomiale de paramètre $p = (\mathbb{P}(X_1 \in A_1), \dots, \mathbb{P}(X_1 \in A_k))$ et n.

Exemple : Une population de taille N est divisée en classes $E_1, \dots, E_k$ (socioprofessionnelle) de taille $Np_1, \dots, Np_k$. On considère n individus pris au hasard avec remise et on compte leur nombre par classes $Z = (Z_1, \dots, Z_k)$ alors Z suit une loi multinomiale de paramètres $p = (p_1, \dots, p_k)$ et n.

Application : (Test du χ^2). Soit $(Z^n)=(Z_1^n, \dots, Z_k^n)$ une suite de v.a de loi multinomiale de paramètres $p = (p_1, \dots, p_k)$ et n Alors

$$\sum_{i=1}^{k} \frac{(Z_i^n - np_i)^2}{np_i} \xrightarrow[n \to \infty]{\mathcal{L}} \chi^2(k-1)$$

rem : Chaque marginale suit une loi $\mathcal{B}(n, p_i)$.

3) Théorèmes limites.

Déf : Soient $r \in \mathbb{N}^*$, $r_1 \in \{0, \dots, r\}$ et $n \in \{1, \dots, r\}$. Une v.a.r X à valeurs dans $\{0, \dots, n\}$ suit une loi hypergéométrique de paramètres n, r et r_1 si

$$\mathbb{P}(X = k) = \frac{\binom{r_1}{k}\binom{r-r_1}{n-k}}{\binom{r}{n}} \quad \text{pour } k \in \{0, \dots, n\} \text{ notée } \mathcal{H}(n, r, r_1).$$

Application : Une population est formée de r personnes dont r_1 vont voter A pour un référendum. On effectue un sondage sur n personnes choisies au hasard. On note X le nombre de personnes votant A. X suit alors une loi hypergéométrique de paramètres r, r_1 et n.

Prop : Soit $(X_j)_{j \geq 1}$ une suite de v.a.r de loi $\mathcal{H}(n, r_j, r_j^1)$ telle que $\lim_{j \to +\infty} \frac{r_j^1}{r_j} = p \in]0,1[$

Alors $X_j \xrightarrow{\mathcal{L}} X \sim \mathcal{B}(n, p)$.

rem : Si la taille de la population est suffisamment grande, on assimile la loi de X dans le sondage à une loi binomiale de paramètres $n, \frac{r_1}{r}$.

On cherche à estimer le paramètre $p \in]0,1[$ grâce à $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ la moyenne empirique lorsque $X_1, \dots X_n$ sont iid de loi $\mathcal{B}(p)$.

Thm : Soit $(X_n)_{n \geq 1}$ une suite de v.a iid de loi $\mathcal{B}(p)$. On note $S_n = X_1 + \dots + X_n$

Alors $\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > t \right) \leq \frac{pq}{nt^2} \quad \forall t > 0$ (loi faible des grands nombres)

Application - Thm de Weierstrass : Soit $f : [0,1] \to \mathbb{C}$ une fonction continue, ω son module de continuité. On pose pour $n \geq 1$ $x \in [0,1]$

$B_n(f, x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right)$ le n-ième polynôme de Bernstein Alors

1) B_n converge vers f uniformément sur $[0,1]$

2) $\|B_n - f\| \leq \frac{3}{2}\, \omega\left(\frac{1}{\sqrt{n}}\right)$ pour $n \geq 1$

- $\overline{X}$ est un estimateur consistant.

Thm : (grande déviations) Soit $(S_n)_{n \geq 1}$ des v.a de loi $\mathcal{B}(n, p)$ $p \in]0,1[$

Pour $\varepsilon > 0$, on a $\mathbb{P}\left(\frac{S_n}{n} \geq p + \varepsilon\right) \leq \exp(-n h(p, \varepsilon))$

$\mathbb{P}\left(\frac{S_n}{n} \leq p - \varepsilon\right) \leq \exp(-n h(p, \varepsilon))$ où $h(p, \varepsilon), h(p, \varepsilon) > 0$.

Thm : Sous les mêmes hypothèses, on a $\frac{S_n}{n} \xrightarrow{p.s.} p$ (loi forte des grands nombres)

rem : $\overline{X}$ est fortement consistant.

Thm : Sous les mêmes hypothèses, on a $\frac{S_n - np}{\sqrt{npq}} \xrightarrow{\mathcal{L}} X \sim \mathcal{N}(0,1)$ (thm central limite).

rem : $\overline{X}$ est asymptotiquement normal. On peut approximer une loi binomiale par une loi normale pour n grand et p voisin de $\frac{1}{2}$, en pratique pour $n\,p(1-p) > 18$.

Application : Intervalle de confiance : $\mathbb{P}\left(|\overline{X} - p| \geq \frac{1{,}96\sqrt{pq}}{\sqrt{n}} \right) \xrightarrow[n\infty]{} \beta \leq 0{,}05$.

II. Loi de Poisson

Déf : Soit X une v.a.r à valeur entière telle que $\mathbb{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$ $\forall x \in \mathbb{N}$

On dit que X suit une loi de Poisson de paramètre λ : $\mathcal{P}(\lambda)$

Prop : On a $\mathbb{E}(X) = \mathbb{V}(X) = \lambda$, pour $t \in \mathbb{R}$ $\varphi_X(t) = \exp(\lambda(e^{it} - 1))$

Thm : Soit (X_n) une suite de v.a de loi $\mathcal{B}(n, p_n)$ telle que $n p_n \to \lambda$

Alors $X_n \xrightarrow{\mathcal{L}} X \sim \mathcal{P}(\lambda)$; $\mathbb{E}(X_n) \to \mathbb{E}(X)$; $\mathbb{V}(X_n) \to \mathbb{V}(X)$.

rem : Lorsque n est assez grand et p petit, on peut approximer la loi $\mathcal{B}(n, p)$ à $\mathcal{P}(np)$ en pratique, pour $p \leq 0{,}10$ et $n > 50$.

Applications : étude de la répartition spaciale des impacts de bombes dans le sud de Londres pendant la seconde guerre mondiale.

- voir les applications des processus de Poisson.

Thm : $X_1, \dots X_n, \dots$ une suite de v.a iid de loi $\mathcal{P}(\lambda)$ Alors.

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s} \lambda$ et $\sqrt{n}\left(\frac{\overline{X}_n - \lambda}{\sqrt{\lambda}}\right) \xrightarrow{\mathcal{L}} X \sim \mathcal{N}(0,1)$.

Application : $e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} \xrightarrow[n\infty]{} \frac{1}{2}$.

III - Processus de Poisson

Déf : Un processus aléatoire $(N(t))_{t\geq 0}$ est dit ponctuel si :
- pour tout $t \geq 0$, la v.a. $N(t)$ est à valeurs dans $\mathbb{N}$
- presque sûrement, la trajectoire $t \longrightarrow N(t)$ est croissante par sauts d'amplitude 1, continue à droite et telle que $N(0) = 0$.

Déf : Un processus ponctuel $(N(t))_{t\geq 0}$ est appelé processus de Poisson d'intensité $\lambda > 0$ si :
- il est à accroissements indépendants :
$$\forall s,t \geq 0,\ N(t+s) - N(s) \perp\!\!\!\perp \sigma(N(u) ; u \leq s)$$
- $\forall s,t \geq 0$, $N(t+s) - N(s) \sim \mathcal{P}(\lambda t)$

Th : Soit $(N(t))_{t\geq 0}$ un processus ponctuel, à accroissements indépendants, tel que :
$$\forall s,t \geq 0,\ N(t+s) - N(s) \sim N(t) \quad \text{(stationnarité)}$$
Alors c'est un processus de Poisson. (définition qualitative)

Soit $S_n = \inf\{t \geq 0 ; N(t) = n\}$ pour $n \in \mathbb{N}^*$, et $S_0 = 0$

Posons, pour $n \in \mathbb{N}^*$, $T_n = S_n - S_{n-1}$

Prop : la suite $(T_n)_{n\geq 1}$ est formée de v.a iid de loi $\mathcal{E}(\lambda)$

Rq : $S_n = T_1 + \dots + T_n$, donc $S_n \sim \gamma(n,\lambda)$

Existence du processus de Poisson :

Soit $(T_n)_{n\geq 1}$ suite de v.a. iid de loi $\mathcal{E}(\lambda)$, $\lambda > 0$

Posons $S_0 = 0$ et $S_n = T_1 + \dots + T_n$ pour $n \geq 1$

$N(t) = \sum_{n\geq 1} \mathbb{1}_{\{S_n \leq t\}}$ définit un processus de Poisson d'intensité λ.

Estimation du paramètre λ :

Prop : soit $(N(t))_{t\geq 0}$ un processus de Poisson d'intensité λ

Alors : $\dfrac{N(t)}{t} \xrightarrow{p.s.} \lambda$ et $\dfrac{N(t) - \lambda t}{\sqrt{\lambda t}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$

Ex : Temps d'arrivée à un guichet
- $N(t)$ désigne le nombre de clients arrivés au guichet à l'instant t.
  T_n désigne le temps écoulé entre l'arrivée du $(n-1)^{è}$ client et celle du $n^{è}$.
  S_n représente la date d'arrivée du $n^{è}$ client.
- Supposons que l'on a observé N arrivées de clients pendant une durée T.
  Alors $\hat{\lambda} = \dfrac{N}{T}$ est un estimateur de λ, sans biais et fortement consistant.

Références :
- Dacunha-Castelle, Duflo, problèmes à temps fixe
- Foata, Fuchs, processus stochastiques
- Ouvrard, probabilités (1 & 2)

Développements :
- loi faible des grands nombres et grandes déviations.
- définition qualitative d'un processus de Poisson.