

Autour de la loi de Poisson

Notations - Introduction

Dans tout ce problème, $(\Omega, \mathcal{T}, \mathbb{P})$ est un espace probabilisé. Toutes les variables aléatoires considérées ici sont des variables aléatoires définies sur Ω . Si X est une telle variable aléatoire, on note $\mathbb{E}(X)$ et $\mathbb{V}(X)$ l'espérance et la variance de X lorsque celles-ci existent.

Pour $n \in \mathbb{N}^*$ et $p \in]0, 1[$ on notera $\mathcal{B}(p)$ la loi de Bernoulli de paramètre p et $\mathcal{B}(n, p)$ la loi binomiale de paramètres n et p . On rappelle que la variable aléatoire X suit $\mathcal{B}(p)$ si elle prend les valeurs 0 et 1 avec

$$\mathbb{P}(X = 0) = 1 - p \quad \text{et} \quad \mathbb{P}(X = 1) = p$$

et que Y suit $\mathcal{B}(n, p)$ si elle prend les valeurs entières de 0 à n avec

$$\forall k \in \{0, \dots, n\}, \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Pour $\lambda > 0$ on note $\mathcal{P}(\lambda)$ la loi de Poisson de paramètre λ . On rappelle que X suit $\mathcal{P}(\lambda)$ si elle prend toutes les valeurs de \mathbb{N} avec

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Si X est une variable aléatoire prenant ses valeurs dans \mathbb{N} , alors sa fonction génératrice est la fonction $g_X : t \mapsto \mathbb{E}(t^X) = \sum_{k \geq 0} \mathbb{P}(X = k) t^k$.

Dans la première partie on révisé le programme concernant la loi de Poisson et son utilisation comme approximation de la loi binomiale. Dans la deuxième partie on obtient une estimation plus fine de cette approximation. Dans la troisième partie on étudie quelques propriétés en lien avec la fonction génératrice.

1 Loi de Poisson - révisions

1. Soit $\lambda > 0$ et Z une variable aléatoire suivant $\mathcal{P}(\lambda)$, calculer $\mathbb{E}(Z)$ et $\mathbb{V}(Z)$.

2. Soit X et Y deux variables aléatoires indépendantes suivant respectivement les lois $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$ pour λ et μ strictement positifs. Démontrer que $X + Y$ suit la loi $\mathcal{P}(\lambda + \mu)$.
3. **Approximation de la loi binomiale par la loi de Poisson.** Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires telle que X_n suit la loi $\mathcal{B}(n, p_n)$ pour tout $n \geq 1$, où $(p_n)_{n \geq 1}$ est une suite dans $]0, 1[$. On suppose que la suite $(np_n)_n$ converge vers $\lambda > 0$.
 - a) Montrer que la suite $(X_n)_n$ converge en loi vers une variable aléatoire Z suivant $\mathcal{P}(\lambda)$, c'est-à-dire

$$\forall k \in \mathbb{N}, \quad \lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

- b) Démontrer que $\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}(Z)$ et $\lim_{n \rightarrow +\infty} \mathbb{V}(X_n) = \mathbb{V}(Z)$.
- c) *Application.* On suppose qu'il apparaît en moyenne deux étoiles filantes toutes les 5 minutes dans le ciel d'une nuit de la première semaine d'août. On choisit au hasard un intervalle de 5 minutes. Soit Y la variable aléatoire associant à l'intervalle de 5 minutes choisi le nombre d'étoiles filantes observées. Pour déterminer la loi de probabilité de Y , on "discrétise" le problème : on partage les cinq minutes en n intervalles de temps suffisamment petits pour contenir au plus une apparition d'étoile filante, et on suppose que les apparitions des étoiles filantes au cours du temps sont des événements indépendants. Ainsi, n est "grand" et la probabilité d'apparition d'une étoile filante dans l'un des n intervalles de temps est de l'ordre de $\frac{2}{n}$. Donner une approximation des probabilités suivantes :

$$\mathbb{P}(Y = 0) \quad ; \quad \mathbb{P}(Y = 1) \quad \text{et} \quad \mathbb{P}(Y \geq 2)$$

Quelle serait la probabilité de voir au moins 2 étoiles filantes en 10 minutes ?

2 Convergence forte de la loi binomiale vers la loi de Poisson

1. Soit $p \in [0, \frac{4}{5}]$, déterminer la loi du vecteur aléatoire (X, Y) tel que
 - (X, Y) prend les valeurs $(0, 0)$, $(0, 1)$, $(1, 1)$ et $(k, 0)$ pour tout $k \geq 2$ (remarque : (X, Y) ne prend pas la valeur $(1, 0)$),
 - la loi marginale X est $\mathcal{P}(p)$ et la loi marginale Y est $\mathcal{B}(p)$.
 On notera $\mathcal{M}(p)$ la loi de ce vecteur aléatoire. Dédurre du calcul précédent que $\mathbb{P}(X \neq Y) \leq 2p^2$ en utilisant $e^{-p} \geq 1 - p$.
2. Soit (X, Y) un vecteur aléatoire prenant ses valeurs dans \mathbb{N}^2 et A une partie de \mathbb{N} , montrer que

$$|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \leq \mathbb{P}(X \neq Y)$$

3. Soit $(X_1, Y_1), \dots, (X_n, Y_n)$ des vecteurs aléatoires suivant respectivement les lois $\mathcal{M}(p_1), \dots, \mathcal{M}(p_n)$ avec $p_i \in [0, \frac{4}{5}]$ pour tout i . Soit A une partie de \mathbb{N} , montrer que

$$|\mathbb{P}(X_1 + \dots + X_n \in A) - \mathbb{P}(Y_1 + \dots + Y_n \in A)| \leq 2 \sum_{i=1}^n p_i^2$$

4. Soit $\lambda > 0$ et n un entier tel que $4n \geq 5\lambda$. Soit X et Z deux variables aléatoires suivant respectivement la loi $\mathcal{B}(n, \frac{\lambda}{n})$ et la loi $\mathcal{P}(\lambda)$.
- a) Soit A une partie de \mathbb{N} , montrer que

$$|\mathbb{P}(X \in A) - \mathbb{P}(Z \in A)| \leq 2 \frac{\lambda^2}{n}$$

- b) Donner une condition sur n et p pour que l'approximation de la loi $\mathcal{B}(n, p)$ par la loi $\mathcal{P}(np)$ soit correcte à 10^{-2} près pour toute valeur $k \in \{0, \dots, n\}$.
- c) En considérant les ensembles $E = \{k : \mathbb{P}(X = k) \geq \mathbb{P}(Z = k)\}$ et $\mathbb{N} \setminus E$, montrer que

$$\sum_{k=0}^n \left| \binom{n}{k} p^k (1-p)^{1-k} - \frac{\lambda^k}{k!} e^{-\lambda} \right| + \sum_{k \geq n+1} \frac{\lambda^k}{k!} e^{-\lambda} \leq 4 \frac{\lambda^2}{n}$$

3 Loi de Poisson et fonction génératrice

Dans cette partie, X est une variable aléatoire à valeurs entières.

- Démontrer que la fonction génératrice de X a un rayon de convergence supérieur ou égal à 1.
- Dans cette question, on suppose que X suit la loi $\mathcal{P}(\lambda)$ pour un certain $\lambda > 0$. Calculer la fonction génératrice g_X de X .
- Démontrer que

$$\forall n \geq 0, \quad \sum_{k=0}^n k \mathbb{P}(X = k) \leq \liminf_{t \rightarrow 1^-} \frac{g_X(t) - g_X(1)}{t - 1}$$

En déduire que X admet une espérance si et seulement si g_X admet une dérivée à gauche $g'_X(1^-)$ en 1, et que dans ce cas $\mathbb{E}(X) = g'_X(1^-)$.

- Retrouver à l'aide des questions précédentes l'espérance de la loi $\mathcal{P}(\lambda)$.
- Démontrer que X admet une variance si et seulement si g_X admet une dérivée seconde à gauche $g''_X(1^-)$ en 1, et que dans ce cas $\mathbb{V}(X) = g''_X(1^-) + g'_X(1^-) - (g'_X(1^-))^2$. Retrouver la variance de la loi $\mathcal{P}(\lambda)$.
- On suppose que les variables aléatoires X et Y sont indépendantes, démontrer que $g_{X+Y} = g_X g_Y$. Retrouver le résultat de la question 1.2.

4 Correction succincte.

Les parties 1 et 3 sont classiques et sont basées essentiellement sur du cours. On peut donc en trouver l'essentiel dans les ouvrages usuels (Cottrell et al., Dantzer, Ouvrard, etc), à l'exception de l'application 3c de la partie 1. La partie 2 est plus originale et précise la convergence de la loi binomiale vers la loi de Poisson.

4.1 Correction de la partie 1 : Loi de Poisson - révisions

1. Les séries à termes positifs $\sum k\mathbb{P}(Z = k)$ et $\sum k^2\mathbb{P}(Z = k)$ sont convergentes, donc Z admet une espérance et une variance, et on trouve : $\mathbb{E}(Z) = \mathbb{V}(Z) = \lambda$.
2. Puisque X et Y suivent des lois de Poisson, elles prennent leurs valeurs dans \mathbb{N} , et comme elles sont indépendantes on obtient pour tout entier n :

$$\begin{aligned}\mathbb{P}(X + Y = n) &= \sum_{k=0}^n \mathbb{P}(X = k \text{ et } Y = n - k) = \sum_{k=0}^n \frac{\lambda^k}{k!} e^{-\lambda} \frac{\mu^{n-k}}{(n-k)!} e^{-\mu} \\ &= \frac{1}{n!} e^{-(\lambda+\mu)} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} = \frac{(\lambda + \mu)^n}{n!} e^{-(\lambda+\mu)}.\end{aligned}$$

Ceci étant valable pour tout $n \in \mathbb{N}$, on en déduit que $X + Y$ suit la loi $\mathcal{P}(\lambda + \mu)$.

3. **Approximation de la loi binomiale par la loi de Poisson.** Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires telle que X_n suit la loi $\mathcal{B}(n, p_n)$ pour tout $n \geq 1$, où $(p_n)_{n \geq 1}$ est une suite dans $]0, 1[$. On suppose que la suite $(np_n)_n$ converge vers $\lambda > 0$.

a) Soit $k \in \mathbb{N}$, pour tout $n \geq k$ on a

$$\begin{aligned}\mathbb{P}(X_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \frac{(np_n)^k}{k!} \left(1 - \frac{np_n}{n}\right)^n \times \frac{n(n-1)\dots(n-k+1)}{n^k} (1 - p_n)^{-k}\end{aligned}$$

or on a

$$\left(1 - \frac{np_n}{n}\right)^n = \exp(-np_n(1 + o(1))) \rightarrow e^{-\lambda}$$

et comme k est fixé et $p_n \rightarrow 0$ on obtient bien $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}$.

- b) Pour tout n on a $\mathbb{E}(X_n) = np_n$ et $\mathbb{V}(X_n) = np_n(1-p_n)$, donc on a $\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \lambda = \mathbb{E}(Z)$ et $\lim_{n \rightarrow +\infty} \mathbb{V}(X_n) = \lambda = \mathbb{V}(Z)$.

- c) D'après l'énoncé, Y suit la loi $\mathcal{B}(n, \frac{2}{n})$, et comme n est supposé grand on peut approcher cette loi par la loi de Poisson $\mathcal{P}(2)$, on trouve donc

$$\mathbb{P}(Y = 0) \simeq e^{-2} = 13,5\%$$

$$\mathbb{P}(Y = 1) \simeq 2e^{-2} = 27\%$$

$$\mathbb{P}(Y \geq 2) = 1 - (\mathbb{P}(Y = 0) + \mathbb{P}(Y = 1)) \simeq 1 - (e^{-2} + 2e^{-2}) = 59\%$$

Dans un intervalle de 10 minutes il y a $2n$ petits intervalles, dans ce cas le nombre d'étoiles filantes observées suit $\mathcal{B}(2n, \frac{2}{n})$ ce qu'on peut approcher par la loi $\mathcal{P}(4)$ et donc la probabilité demandée est environ $1 - (e^{-4} + 4e^{-4}) = 91\%$.

4.2 Correction de la partie 2 : Convergence forte de la loi binomiale vers la loi de Poisson

Cette partie est tirée de l'exercice 406 du livre **Intégration et probabilités - Analyse de Fourier** de Gérard Letac (Masson, 1997).

1. Comme la marginale X suit $\mathcal{P}(p)$ on a :

$$\mathbb{P}[(X, Y) = (0, 0) \text{ ou } (X, Y) = (0, 1)] = e^{-p}$$

$$\mathbb{P}[(X, Y) = (1, 1)] = pe^{-p}$$

$$\forall k \geq 2, \quad \mathbb{P}[(X, Y) = (k, 0)] = \frac{\lambda^k}{k!} e^{-p}$$

et comme Y suit $\mathcal{B}(p)$ on a

$$\mathbb{P}[(X, Y) = (0, 1) \text{ ou } (X, Y) = (1, 1)] = p$$

et donc on en conclut que

$$\mathbb{P}[(X, Y) = (0, 1)] = p - pe^{-p} \quad \text{et} \quad \mathbb{P}[(X, Y) = (0, 0)] = (1 + p)e^{-p} - p$$

en utilisant le fait que les événements $((X, Y) = (0, 0))$, $((X, Y) = (0, 1))$ et $((X, Y) = (1, 1))$ sont incompatibles. Puisque $p \in [0, \frac{4}{5}]$ on a bien que ces deux probabilités sont positives.

Enfin on remarque pour ce vecteur aléatoire on a

$$\begin{aligned} \mathbb{P}(X \neq Y) &= 1 - \mathbb{P}(X = Y) = 1 - \mathbb{P}[(X, Y) = (0, 0) \text{ ou } (X, Y) = (1, 1)] \\ &= 1 - ((1 + p)e^{-p} - p + pe^{-p}) \leq 1 - (1 + 2p)(1 - p) + p = 2p^2. \end{aligned}$$

2. On remarque que $(X \in A)$ est inclus dans l'évènement $(Y \in A \text{ ou } X \neq Y)$: en effet si $X(\omega) \in A$ alors soit $X(\omega) = Y(\omega)$ et donc $Y(\omega) \in A$, soit $X(\omega) \neq Y(\omega)$. Par conséquent

$$\mathbb{P}(X \in A) \leq \mathbb{P}(Y \in A \text{ ou } X \neq Y) \leq \mathbb{P}(Y \in A) + \mathbb{P}(X \neq Y)$$

donc

$$\mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \leq \mathbb{P}(X \neq Y)$$

En échangeant les rôles de X et Y on obtient l'inégalité souhaitée.

Autre manière d'écrire la même preuve : On calcule

$$\begin{aligned} \mathbb{P}(X \in A) &= \mathbb{P}[(X \in A \text{ et } X = Y) \text{ ou } (X \in A \text{ et } X \neq Y)] \\ &= \mathbb{P}[(Y \in A \text{ et } X = Y) \text{ ou } (X \in A \text{ et } X \neq Y)] \leq \mathbb{P}(Y \in A) + \mathbb{P}(X \neq Y) \end{aligned}$$

et on conclut de même.

3. On applique les deux questions précédentes pour calculer

$$\begin{aligned} |\mathbb{P}(X_1 + \dots + X_n \in A) - \mathbb{P}(Y_1 + \dots + Y_n \in A)| &\leq \mathbb{P}((X_1 + \dots + X_n) \neq (Y_1 + \dots + Y_n)) \\ &\leq \mathbb{P}((X_1 \neq Y_1) \text{ ou } \dots \text{ ou } (X_n \neq Y_n)) \\ &\leq \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i) \leq 2 \sum_{i=1}^n p_i^2 \end{aligned}$$

où la deuxième inégalité découle du fait que $(X_1 + \dots + X_n)(\omega) \neq (Y_1 + \dots + Y_n)(\omega)$ implique que $X_i(\omega) \neq Y_i(\omega)$ pour au moins un indice i .

4. a) On applique le résultat de la question précédente avec des variables aléatoires X_1, \dots, X_n indépendantes suivant $\mathcal{B}(\frac{\lambda}{n})$ et Y_1, \dots, Y_n indépendantes suivant $\mathcal{P}(\frac{\lambda}{n})$, et telles que chaque couple (X_i, Y_i) suit $\mathcal{M}(\frac{\lambda}{n})$. Dans ce cas $X = X_1 + \dots + X_n$ suit $\mathcal{B}(n, \frac{\lambda}{n})$ et $Z = Y_1 + \dots + Y_n$ suit $\mathcal{P}(\lambda)$ et on a

$$|\mathbb{P}(X \in A) - \mathbb{P}(Z \in A)| \leq 2 \sum_{i=1}^n \left(\frac{\lambda}{n}\right)^2 = 2 \frac{\lambda^2}{n}$$

b) En prenant $\lambda = np$ dans ce qui précède, il suffit d'avoir $2np^2 \leq 10^{-2}$.

c) On applique la question 4a à l'ensemble $E = \{k : \mathbb{P}(X = k) \geq \mathbb{P}(Z = k)\}$ et on obtient

$$\begin{aligned} 2 \frac{\lambda^2}{n} &\geq |\mathbb{P}(X \in E) - \mathbb{P}(Z \in E)| = \sum_{k \in E} (\mathbb{P}(X = k) - \mathbb{P}(Z = k)) \\ &= \sum_{k \in E} \left(\binom{n}{k} p^k (1-p)^{1-k} - \frac{\lambda^k}{k!} e^{-\lambda} \right) = \sum_{k \in E} \left| \binom{n}{k} p^k (1-p)^{1-k} - \frac{\lambda^k}{k!} e^{-\lambda} \right| \end{aligned}$$

où on a utilisé la définition de E et le fait que $E \subset \{0, \dots, n\}$. En appliquant la question 4a avec $\mathbb{N} \setminus E$ il vient

$$\begin{aligned} 2 \frac{\lambda^2}{n} &\geq |\mathbb{P}(X \in E) - \mathbb{P}(Z \in E)| = \sum_{k \in \mathbb{N} \setminus E} (\mathbb{P}(Z = k) - \mathbb{P}(X = k)) \\ &= \sum_{\substack{0 \leq k \leq n \\ k \in \mathbb{N} \setminus E}} \left(\frac{\lambda^k}{k!} e^{-\lambda} - \binom{n}{k} p^k (1-p)^{1-k} \right) + \sum_{k \geq n+1} \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{\substack{0 \leq k \leq n \\ k \in \mathbb{N} \setminus E}} \left| \frac{\lambda^k}{k!} e^{-\lambda} - \binom{n}{k} p^k (1-p)^{1-k} \right| + \sum_{k \geq n+1} \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

On conclut en additionnant ces deux inégalités.

4.3 Correction de la partie 3 : Loi de Poisson et fonction génératrice

1. Pour $t = 1$, la somme de la série entière $\sum P(X = k)t^k$ est convergente et de valeur 1 puisque X prend ses valeurs dans \mathbb{N} : la fonction génératrice de X a donc un rayon de convergence supérieur ou égal à 1.

2. Lorsque X suit $\mathcal{P}(\lambda)$ on trouve $g_X(t) = \mathbb{E}(t^X) = e^{\lambda(t-1)}$, et la fonction génératrice a un rayon de convergence infini.
3. Soit $n \geq 0$ fixé, alors pour tout $t \in [0, 1[$ on a

$$\frac{g_X(t) - g_X(1)}{t - 1} = \sum_{k=0}^{+\infty} \mathbb{P}(X = k) \frac{t^k - 1}{t - 1} \geq \sum_{k=0}^n \mathbb{P}(X = k) \frac{t^k - 1}{t - 1}$$

d'où on déduit

$$\sum_{k=0}^n k \mathbb{P}(X = k) \leq \liminf_{t \rightarrow 1^-} \frac{g_X(t) - g_X(1)}{t - 1} \quad (1)$$

Ainsi, si X n'admet pas d'espérance, alors la série à termes positifs $\sum k \mathbb{P}(X = k)$ est divergente et g_X n'est pas dérivable à gauche en 1. Par contre, si X admet une espérance alors cette série est convergente et pour tout $t \in [0, 1[$ on a

$$\begin{aligned} \frac{g_X(t) - g_X(1)}{t - 1} &= \sum_{k=0}^{+\infty} \mathbb{P}(X = k) \frac{t^k - 1}{t - 1} \\ &= \sum_{k=1}^{+\infty} \mathbb{P}(X = k) (1 + t + \dots + t^{k-1}) \leq \sum_{k=1}^{+\infty} \mathbb{P}(X = k) k \end{aligned}$$

donc $\limsup_{t \rightarrow 1^-} \frac{g_X(t) - g_X(1)}{t - 1} \leq \mathbb{E}(X)$. En passant à la limite quand $n \rightarrow +\infty$ dans

(1) on obtient $\mathbb{E}(X) \leq \liminf_{t \rightarrow 1^-} \frac{g_X(t) - g_X(1)}{t - 1}$, et on conclut donc que dans ce cas g_X est dérivable à gauche en 1 et $\mathbb{E}(X) = g'_X(1^-)$.

4. Si X suit la loi $\mathcal{P}(\lambda)$ alors sa fonction génératrice $g_X : t \mapsto e^{\lambda(t-1)}$ est dérivable en 1 de dérivée λ , qui est bien l'espérance attendue.
5. En raisonnant de la même manière, on trouve les inégalités (pour $t \in [0, 1[$)

$$\sum_{k=1}^n k \mathbb{P}(X = k) \frac{t^{k-1} - 1}{t - 1} \leq \frac{g'_X(t) - g'_X(1)}{t - 1} \leq \sum_{k=2}^{+\infty} \mathbb{P}(X = k) k(k-1)$$

d'où on déduit que X admet un moment d'ordre 2 si et seulement si g_X admet une dérivée seconde à gauche $g''_X(1^-)$ en 1, et que dans ce cas

$$g''_X(1^-) = \sum_{k=0}^{+\infty} \mathbb{P}(X = k) k(k-1) = \mathbb{V}(X) + \mathbb{E}(X)^2 - \mathbb{E}(X) = \mathbb{V}(X) - (g'_X(1^-))^2 + g'_X(1^-)$$

6. Soit $t \in]-1, 1[$, alors on peut écrire

$$\begin{aligned} g_{X+Y}(t) &= \sum_{n=0}^{+\infty} \mathbb{P}(X + Y = n) t^n = \sum_{n=0}^{+\infty} \sum_{k=0}^n \mathbb{P}(X = k \text{ et } Y = n - k) t^n \\ &= \sum_{n=0}^{+\infty} \sum_{k=0}^n \mathbb{P}(X = k) \mathbb{P}(Y = n - k) t^n \\ &= \sum_{n=0}^{+\infty} \sum_{k=0}^n \mathbb{P}(X = k) t^k \mathbb{P}(Y = n - k) t^{n-k} = g_X(t) g_Y(t) \end{aligned}$$

où on a utilisé l'indépendance de X et Y et la formule du produit de Cauchy de deux séries convergentes. Cette égalité étant vraie sur un voisinage de 0, on a bien $g_{X+Y} = g_X g_Y$ sur le disque de convergence commun.

Dans le cas où X suit $\mathcal{P}(\lambda)$ et Y suit $\mathcal{P}(\mu)$ on obtien alors pour tout réel t

$$g_{X+Y}(t)g_X(t)g_Y(t) = e^{\lambda(t-1)}e^{\mu(t-1)} = e^{(\lambda+\mu)(t-1)}$$

et donc $X + Y$ suit la loi $\mathcal{P}(\lambda + \mu)$.