

Moments, fonctions génératrices, transformées de Laplace

1 Moments et variance

Théorème 6.1 Soit (Ω, \mathcal{A}, P) un espace de probabilité, et soit n un entier > 0 . Soit \mathcal{L}_n l'ensemble des v.a. X sur cet espace telles que l'espérance $m_n = \mathbf{E}(X^n)$, appelée **moment** d'ordre n , existe. Alors \mathcal{L}_n est un espace vectoriel, et on a

$$\mathcal{L}_1 \supset \mathcal{L}_2 \supset \cdots \supset \mathcal{L}_n.$$

Démonstration Puisque $f(x) = x^n$ définit une fonction convexe sur la demi-droite positive, on peut écrire pour x et y positif que

$$\left(\frac{x+y}{2}\right)^n \leq \frac{1}{2}(x^n + y^n),$$

et donc $|X+Y|^n \leq (|X|+|Y|)^n \leq 2^{n-1}(|X|^n + |Y|^n)$. Une autre méthode pour obtenir cette inégalité est de montrer que $g(t) = 2^{n-1}(t^n + 1) - (t+1)^n$ atteint son minimum sur $[0, +\infty[$ en $t = 1$ et de considérer $g(x/y)$.

Si maintenant les espérances de $|X|^n$ et de $|Y|^n$ sont finies, on en déduit d'après la fin du théorème 5.2 que l'espérance de $|X+Y|^n$ est finie et que $X+Y$ est dans \mathcal{L}_n quand X et Y y sont. Enfin, pour voir que si l'espérance de $|X|^n$ est finie il en est de même pour $|X|^{n-1}$, on utilise l'inégalité

$$|X|^{n-1} \leq 1 + |X|^n,$$

qu'on vérifie immédiatement en étudiant les cas $|X| \leq 1$ et $|X| \geq 1$. Le fait que $\mathcal{L}_{n-1} \supset \mathcal{L}_n$ s'en déduit.

Définition Le **moment centré** d'ordre n de la variable aléatoire X est défini par $\mathbf{E}[(X - m_1)^n]$ où $m_1 = \mathbf{E}(X)$.

Remarquons au passage que si le moment non centré m_n existe, alors le moment centré existe, puisque c'est l'espérance d'un polynôme en X de degré n et qu'on vient de voir que les moments de degré inférieur à n existaient.

Le cas particulier réellement important est le cas où $n = 2$.

Définition Soit X une variable aléatoire réelle. On appelle le moment centré d'ordre 2 de X la **variance** de X , et sa racine carrée positive l'**écart type** de X , encore appelé déviation standard. On note l'écart type $\sigma(X)$ et la variance $(\sigma(X))^2$, ou plus rarement $V(X)$.

Insistons sur le fait que l'écart type a la dimension de la variable aléatoire: si celle ci s'exprime en centimètres, l'écart type s'exprime en centimètres et la variance en centimètres carrés. Il faut connaître les deux formules suivantes:

Proposition 6.2 Si X a un moment d'ordre 2, alors pour λ réel

$$\sigma^2(\lambda X) = \lambda^2 \sigma^2(X),$$

et **Formule de Huyghens:**

$$\sigma^2(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2.$$

En particulier, $(\mathbf{E}(X))^2 \leq \mathbf{E}(X^2)$, avec égalité si et seulement si la loi de X est une probabilité de Dirac.

Démonstration La première formule est immédiate. Pour Huyghens:

$$\sigma^2(X) = \mathbf{E}(X^2 - 2m_1X + m_1^2) = \mathbf{E}(X^2) - 2m_1\mathbf{E}(X) + m_1^2 = \mathbf{E}(X^2) - (\mathbf{E}(X))^2.$$

Ici on a utilisé le fait que l'espérance d'une constante est la constante elle même et que $m_1 = \mathbf{E}(X)$. Quant à la dernière inégalité elle vient du fait qu'une variance est toujours positive ou nulle. Si la variance est nulle, alors appliquant le 5) du théorème 5.2 à la v.a. positive $Y = (X - m_1)^2$, alors la loi de Y est δ_0 et celle de X est donc δ_{m_1} .

Il y a également à connaître deux inégalités célèbres:

Proposition 6.3 Inégalité de Markov Si Y est une variable aléatoire positive ou nulle dont l'espérance existe, alors pour tout $y > 0$ on a

$$P(Y \geq y) \leq \frac{1}{y} \mathbf{E}(Y).$$

Inégalité de Tchebychev Si X est une variable aléatoire ayant un second moment, alors pour tout $t > 0$ on a

$$P(|X - \mathbf{E}(X)| \geq t) \leq \frac{1}{t^2} \sigma^2(X).$$

Démonstration

$$\mathbf{E}(Y) = \mathbf{E}(Y\mathbf{1}_{Y \geq y} + Y\mathbf{1}_{Y < y}) \geq \mathbf{E}(Y\mathbf{1}_{Y \geq y}) \geq$$

$$\mathbf{E}(y\mathbf{1}_{Y \geq y}) \geq y\mathbf{E}(\mathbf{1}_{Y \geq y}) = yP(Y \geq y),$$

ce qui est équivalent à l'inégalité de Markov en divisant les extrémités par y .

On applique ensuite Markov à $Y = (X - m_1)^2$ et à $y = t^2$. Comme

$$P(|X - m_1| \geq t) = P((X - m_1)^2 \geq t^2) \leq \frac{1}{t^2} \mathbf{E}((X - m_1)^2) = \frac{1}{t^2} \sigma^2(X),$$

l'inégalité de Tchebychev est aussi démontrée.

Finalement, la variance d'une somme de variables aléatoires indépendantes est la somme des variances. Plus précisément:

Proposition 6.4 Si X_1, X_2, \dots, X_N sont des variables aléatoires indépendantes ayant un second moment, alors

$$\sigma^2(X_1 + \dots + X_N) = \sigma^2(X_1) + \dots + \sigma^2(X_N).$$

Démonstration Procédons par récurrence sur N . C'est trivial pour $N = 1$. Montrons le pour $N = 2$. Notons pour simplifier $X = X_1 - \mathbf{E}(X_1)$ et $Y = X_2 - \mathbf{E}(X_2)$. Tous deux sont d'espérance nulle. Alors

$$\sigma^2(X_1 + X_2) = \mathbf{E}((X + Y)^2) = \mathbf{E}(X^2) + 2\mathbf{E}(XY) + \mathbf{E}(Y^2) = \sigma^2(X_1) + \sigma^2(X_2),$$

car $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y) = 0$ en utilisant l'indépendance de X et de Y . Ensuite, supposons le résultat vrai à l'ordre $N - 1$. Alors appliquant le résultat pour $N = 2$ au couple $X = X_1 + \dots + X_{N-1}$ et $Y = X_N$, puis l'hypothèse de récurrence, on arrive au résultat.

En corollaire, on a donc la *loi faible des grands nombres* qui dit que en un certain sens, si des variables aléatoires sont indépendantes et de même loi, alors leur moyenne arithmétique tend vers leur espérance commune. Plus précisément:

Théorème 6.5 Loi faible des grands nombres Soit X_1, X_2, \dots une suite infinie de v.a. indépendantes et de même loi, et possédant un second moment. Alors, pour tout nombre $\epsilon > 0$ fixé on a

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{X_1 + \dots + X_n}{n} - \mathbf{E}(X_1) \right| \geq \epsilon \right) = 0.$$

Démonstration Notons $S_n = X_1 + \dots + X_n$. Alors $\mathbf{E}(S_n/n) = \mathbf{E}(X_1)$ et

$$\sigma^2(S_n/n) = \sigma^2(S_n)/n^2 = (\sigma^2(X_1) + \dots + \sigma^2(X_n))/n^2 = \sigma^2(X_1)/n.$$

Ici on a utilisé successivement les propositions 6.2 puis 6.4, puis le fait que les X_j sont de même loi et ont donc même variance. Appliquons alors l'inégalité de Tchebychev à $X = S_n/n$ et à $t = \epsilon$; on obtient

$$P \left(\left| \frac{X_1 + \dots + X_n}{n} - \mathbf{E}(X_1) \right| \geq \epsilon \right) \leq \frac{1}{n\epsilon^2} \sigma^2(X_1),$$

qui tend bien vers 0 pour ϵ fixé.

Commentaires: l'importance philosophique de la loi des grands nombres est non négligeable: elle justifie la démarche que nous avons adoptée pour modéliser le calcul des probabilités. L'idée d'expérience décrite au début de ce cours est la sélection d'un point ω dans un espace d'observables Ω , mais par un procédé susceptible d'être répété ad libitum et dans les mêmes conditions. Soit S une partie de Ω , comptons le nombre de fois où S est réalisé en n essais, divisons ce nombre par n et notons par f_n la fraction, ou la fréquence, ainsi obtenue. L'idée de probabilité est basée sur la constatation physique que la suite des f_n converge vers un nombre $P(S)$ qu'on appellera probabilité de S . Si la théorie est bien faite, c'est à dire si les axiomes sont bien choisis, on doit retrouver cette constatation physique quelque part à l'état de théorème dans la théorie développée à partir de ces axiomes. C'est le cas. En effet, le Ω initial décrivant une expérience est remplacé par un produit infini $\prod_{j=1}^{\infty} \Omega_j$ où les Ω_j sont identiques à l' Ω initial, et sont les résultats possibles de l'expérience répétée à l'instant j . Les points de ce produit sont donc des suites infinies $\omega = (\omega_j)_{j=1}^{\infty}$. Quant à la probabilité sur le produit, elle est telle que toutes les fonctions $f_j(\omega) = \omega_j$ soient indépendantes. Ceci fait, notons $X_j(\omega) = 1$ si $\omega_j \in S$ et $X_j(\omega) = 0$ sinon. On a une suite de v.a. de Bernoulli indépendantes et de même loi d'espérance $p = P(S)$. La loi faible des grands nombres dit que $f_n = \frac{1}{n}(X_1 + \dots + X_n)$ converge vers $P(S)$, dans le sens décrit au théorème 6.5. Il existe un théorème avec une conclusion plus précise, appelé *loi forte des grands nombres*, que nous exposons maintenant.

Théorème 6.6 loi forte des grands nombres Soit X_1, \dots, X_n, \dots des variables aléatoires de Bernoulli indépendantes et de même loi $q\delta_0 + p\delta_1$, avec $0 < p = 1 - q < 1$. Alors

$$\Pr\left(\lim_{n \rightarrow \infty} \frac{1}{n}(X_1 + \dots + X_n) = p\right) = 1.$$

Démonstration Elle s'appuie sur le lemme de Borel:

Lemme de Lebesgue Si $(A_n)_{n \geq 1}$ est une suite d'évènements telle que $\sum_{n \geq 1} \Pr(A_n)$ converge, alors $\Pr(\bigcap_{k \geq 1} \bigcup_{n \geq k} A_n) = 0$.

La démonstration de ce lemme est à peu près triviale: Puisque la suite $(r_k)_{k \geq 1}$ des restes de la série convergente tend vers 0 et que pour tout entier k on peut écrire

$$\Pr(\bigcap_{k \geq 1} \bigcup_{n \geq k} A_n) \leq \Pr(\bigcup_{n \geq k} A_n) \leq \sum_{n \geq k} \Pr(A_n) = r_k,$$

le résultat s'ensuit en faisant tendre k vers l'infini.

On se fixe ensuite un nombre $\epsilon > 0$ et on note pour simplifier

$$U_n(\epsilon) = U_n = \frac{1}{n}(X_1 + \dots + X_n) - p - \epsilon,$$

$$A_n(\epsilon) = A_n = \{U_n > 0\},$$

$$B(\epsilon) = \{\overline{\lim}_{n \rightarrow \infty} U_n > 0\}.$$

Le point délicat de la démonstration est de montrer que pour tout $\epsilon > 0$ il existe un nombre $r_\epsilon = r \in]0,1[$ tel que $P(A_n) \leq r^n$. Admettons ce point quelques instants et achevons la démonstration. On remarque d'abord que

$$\bigcap_{k \geq 1} \bigcup_{n \geq k} A_n = \{\forall k, \exists n \geq k; U_n > 0\}.$$

Un point subtil est ensuite l'inclusion d'évènements:

$$\begin{aligned} \{\overline{\lim}_{n \rightarrow \infty} U_n > 0\} &\subset \{\forall k, \exists n \geq k; U_n > 0\} \\ &\subset \{\forall k, \exists n \geq k; U_n \geq 0\} \subset \{\overline{\lim}_{n \rightarrow \infty} U_n \geq 0\}. \end{aligned}$$

Il n'y a jamais égalité dans ces inclusions: il suffit de penser aux cas $U_n = 1/n$ et $U_n = -1/n$ pour s'en convaincre. Nous n'allons utiliser que la première inclusion. Ayant admis que $\Pr(A_n) < r^n$ avec $r \in]0,1[$, comme la série géométrique de raison r converge, le lemme de Borel est applicable et on en déduit que $\Pr(B(\epsilon)) = 0$.

Ensuite on observe que si $0 < \epsilon < \epsilon'$ on a $B(\epsilon) \supset B(\epsilon')$. Changeons un peu de notation en écrivant pour N entier $B_N = B(1/N)$. La suite d'évènements $(B_N)_{N \geq 1}$ est donc croissante. Mais comme tous les B_N sont de probabilité nulle, on a encore $\Pr(\bigcup_{N \geq 1} B_N) = 0$. Analysons alors l'évènement $\bigcup_{N \geq 1} B_N$. On a

$$\begin{aligned} \bigcup_{N \geq 1} B_N &= \{\exists N; \overline{\lim}_{n \rightarrow \infty} \frac{1}{n}(X_1 + \dots + X_n) > p + \frac{1}{N}\} = \\ &= \{\overline{\lim}_{n \rightarrow \infty} \frac{1}{n}(X_1 + \dots + X_n) > p\}. \end{aligned}$$

Nous avons donc montré que

$$\Pr(\overline{\lim}_{n \rightarrow \infty} \frac{1}{n}(X_1 + \dots + X_n) > p) = 0.$$

Appliquons ce résultat aux variables de Bernoulli $X'_n = 1 - X_n$. Elles sont de loi $p\delta_0 + q\delta_1$ et donc $\Pr(\overline{\lim}_{n \rightarrow \infty} \frac{1}{n}(X'_1 + \dots + X'_n) > q) = 0$. Cependant $\frac{1}{n}(X'_1 + \dots + X'_n) = 1 - \frac{1}{n}(X_1 + \dots + X_n)$ et donc

$$\Pr(\underline{\lim}_{n \rightarrow \infty} \frac{1}{n}(X_1 + \dots + X_n) < p) = 0.$$

L'union de deux évènements de probabilité nulle est nulle, le complémentaire de cette union est de probabilité 1. Cela entraîne:

$$\Pr\left(\overline{\lim}_{n \rightarrow \infty} \frac{1}{n}(X_1 + \dots + X_n) \leq p \leq \underline{\lim}_{n \rightarrow \infty} \frac{1}{n}(X_1 + \dots + X_n)\right) = 1.$$

Donc avec probabilité 1, les limites supérieure et inférieure sont égales à p . C'est le résultat annoncé.

Reste à montrer qu'il existe $r_\epsilon = r \in]0,1[$ tel que

$$\Pr(A_n) = \Pr\left(\frac{1}{n}(X_1 + \dots + X_n) > p + \epsilon\right) \leq r^n.$$

A l'aide d'un nombre $s > 0$ arbitraire, nous donnons d'abord une autre présentation de cet évènement:

$$A_n = \left\{ \left(\frac{1}{n} (X_1 + \dots + X_n) > p + \epsilon \right) \right\} = \left\{ e^{s(X_1 + \dots + X_n)} > e^{sn(p+\epsilon)} \right\}.$$

On applique alors l'inégalité de Markov (proposition 6.3) à $Y = e^{s(X_1 + \dots + X_n)}$ et $y = e^{sn(p+\epsilon)}$. On en tire

$$\begin{aligned} \Pr(A_n) &\leq \frac{1}{y} \mathbf{E}(Y) \\ &= e^{-sn(p+\epsilon)} \mathbf{E}(e^{s(X_1 + \dots + X_n)}) \\ &= (e^{-s(p+\epsilon)} \mathbf{E}(e^{sX_1}))^n \\ &= (e^{-s(p+\epsilon)} (q + pe^s))^n \\ &= (qe^{-sp-s\epsilon} + pe^{sq-s\epsilon})^n. \end{aligned}$$

Insistons sur le fait que cette inégalité est valable pour tout $s > 0$. Observons alors qu'il existe des valeurs de s telles que $s \mapsto \varphi(s) = qe^{-sp-s\epsilon} + pe^{sq-s\epsilon}$ soit < 1 . Une manière de le voir est de calculer $\varphi(0) = 1$ et $\varphi'(0) = -\epsilon$. Cela entraîne évidemment, puisque $-\epsilon = \varphi'(0) = \lim_{s \rightarrow 0} (1 - \varphi(s))/s$, qu'il existe $s_0 > 0$ proche de 0 tel que $r = \varphi(s_0) < 1$. Comme $\varphi > 0$ cela termine la démonstration.

Exercices sur 6.1

1. Soit X une variable aléatoire telles que $0 \leq X \leq 1$. Montrer que $\sigma^2(X) \leq \frac{1}{4}$.
Méthode: si $m = \mathbf{E}(X)$, écrire

$$\frac{1}{4} - (X - m)^2 = \left(\frac{1}{2} - m\right)^2 + X(1 - X)$$

et prendre l'espérance de chaque membre.

2 Les variables aléatoires à valeurs entières.

Nous allons nous concentrer pour un moment sur les variables à valeurs dans l'ensemble \mathbf{N} des entiers ≥ 0 . Dans ce cas les moments seront plus faciles à calculer grâce à l'introduction de la notion de *fonction génératrice* de X :

Théorème 6.6 Soit X une v.a. à valeurs dans \mathbf{N} de loi $P_X = \sum_{n=0}^{+\infty} p_n \delta_n$. On désigne par $f_X(z)$ la somme de la série entière

$$\sum_{n=0}^{+\infty} p_n z^n$$

de rayon de convergence R . Alors

1. $R \geq 1$ et, pour $|z| \leq 1$ on a $f_X(z) = \mathbf{E}(z^X)$.

2. Pour tout n on a $p_n = \frac{1}{n!} f_X^{(n)}(0)$. En particulier, la connaissance de f_X donne la connaissance de la loi de X .
3. Pour tout n le moment d'ordre n $\mathbf{E}(X^n)$ existe si et seulement si la dérivée à gauche d'ordre n au point 1 de la fonction $z \mapsto f_X(z)$ définie sur $[-1,1]$ existe et est finie. Dans ce cas,

$$\mathbf{E}(X(X-1)\cdots(X-n+1)) = f_X^{(n)}(1) = \sum_{k=n}^{\infty} k(k-1)\cdots(k-n+1)p_n;$$

en particulier $\mathbf{E}(X) = f_X'(1)$, $\mathbf{E}(X^2) = f_X''(1) + f_X'(1)$.

4. Si X_1, X_2, \dots, X_N sont des variables aléatoires indépendantes à valeurs dans \mathbf{N} et si $S = X_1 + X_2 + \cdots + X_N$ alors pour $|z| \leq 1$:

$$f_S(z) = f_{X_1}(z) \cdots f_{X_N}(z),$$

c'est-à-dire que la fonction génératrice d'une somme est le produit des fonctions génératrices.

Démonstration Il est clair que la série entière converge pour $z = 1$ puisque $\sum_{n=0}^{+\infty} p_n = 1$ et donc que $f_X(1) = 1$. Donc $R \geq 1$. Ensuite, si $|z| = 1$ la série est absolument convergente. Pour le 2), cela découle du lien entre la formule de Taylor et la somme d'une série entière.

Le 3) est plus délicat. Nous le montrons pour $n = 1$. Le principe pour n quelconque est le même. Supposons d'abord que $\mathbf{E}(X)$ existe, c'est-à-dire, d'après la proposition 5.4, que $\sum_{n=0}^{+\infty} np_n$ converge. Montrons qu'alors la dérivée à gauche en 1 de f_X existe et est finie. Celle-ci est définie comme la limite quand z croît vers 1 de la fonction

$$\frac{f_X(z) - f_X(1)}{z - 1} = \frac{1 - f_X(z)}{1 - z} = \sum_{n=0}^{+\infty} p_n \frac{1 - z^n}{1 - z} = \sum_{n=0}^{+\infty} p_n (1 + z + \cdots + z^{n-1}).$$

Or si $0 \leq z \leq 1$ on a $1 + z + \cdots + z^{n-1} \leq n$. Comme $\sum_{n=0}^{+\infty} np_n$ converge la série précédente converge normalement et sa limite est pour z tendant vers 1 est $\mathbf{E}(X)$.

Inversement, supposons que la dérivée à gauche en 1, notée $f_X'(1)$ existe. Appliquons le théorème des accroissements finis à l'intervalle $[z,1]$ et à la fonction f_X . Il existe donc $c \in]z,1[$ tel que

$$\frac{1 - f_X(z)}{1 - z} = f_X'(c) = \sum_{n=1}^{+\infty} np_n c^{n-1}.$$

Ceci tend vers une limite finie si z croît vers 1 par hypothèse. Il est clair puisque c tend vers 1 avec z , que cette limite est supérieure ou égale à toutes les sommes partielles de la série $\sum_{n=0}^{+\infty} np_n$, ce qui prouve que cette série converge. Enfin, trivialement,

$$\sum_{n=1}^{+\infty} p_n c^{n-1} \leq \sum_{n=1}^{+\infty} np_n,$$

ce qui montre finalement que $f'_X(1) = \mathbf{E}(X)$.

Le 4) est une conséquence immédiate du fait que si les X_j sont indépendants, alors les z^{X_j} sont indépendants, et que l'espérance du produit de variables indépendantes est le produit des espérances:

$$\begin{aligned} f_S(z) &= \mathbf{E}(z^{X_1+\dots+X_N}) = \mathbf{E}(z^{X_1} \dots z^{X_N}) = \\ &= \mathbf{E}(z^{X_1}) \dots \mathbf{E}(z^{X_N}) = f_{X_1}(z) \dots f_{X_N}(z). \end{aligned}$$

Commentaires: la démonstration du 3) n'est pas facile si $R = 1$, comme on l'a vu. Si $R > 1$, c'est simple et immédiat par le théorème de dérivation d'une série entière à l'intérieur de l'intervalle de convergence.

Nous étudions maintenant 4 exemples fondamentaux de lois sur \mathbf{N} .

Définition - Proposition La loi de Bernoulli $B_{1,p}$. Pour $0 < p < 1$ c'est la loi

$$B_{1,p} = (1-p)\delta_0 + p\delta_1.$$

Sa fonction génératrice est $f(z) = (1-p) + pz$, son espérance est p et sa variance est $(1-p)p$.

Définition - Proposition La loi binomiale $B_{N,p}$. C'est la loi du nombre de succès dans le schéma Succès Echec fini à N essais:

$$B_{N,p} = \sum_{k=0}^N C_N^k (1-p)^{N-k} p^k \delta_k.$$

Sa fonction génératrice est d'après la formule du binôme, $f(z) = ((1-p) + pz)^N$. Donc en prenant sa dérivée à l'ordre 1, son espérance est donc Np . Quant à sa variance, c'est $N(1-p)p$.

On remarque que si X et Y sont indépendantes et de lois respectives $B_{N,p}$ et $B_{M,p}$, alors la loi de $X + Y$ est $B_{N+M,p}$, comme on le voit par la fonction génératrice.

Un bon moyen de retenir ces résultats sur la loi binomiale est d'observer que si X_1, \dots, X_N sont des variables aléatoires indépendantes de même loi de Bernoulli $B_{1,p}$, alors $S = X_1 + \dots + X_N$ est de loi binomiale $B_{N,p}$ comme on le voit par la fonction génératrice f_S .

Définition - Proposition La loi de Poisson \mathcal{P}_λ . Pour $\lambda > 0$, c'est la loi définie par

$$\mathcal{P}_\lambda = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \delta_n.$$

Sa fonction génératrice est $f(z) = \exp(\lambda(z - 1))$, son espérance et sa variance sont toutes deux égales à λ .

On remarque que si X et Y sont indépendantes et de lois respectives \mathcal{P}_λ et \mathcal{P}_μ , alors la loi de $X + Y$ est $\mathcal{P}_{\lambda+\mu}$, comme on le voit par la fonction génératrice.

La manière la plus courante de rencontrer cette loi de Poisson dans la nature est en tant qu'approximation de la loi binomiale. En effet, la suite de lois $B_{N,\lambda/N}$ tend vers \mathcal{P}_λ dans le sens suivant: pour tout entier k on a

$$\lim_{N \rightarrow \infty} B_{N,\lambda/N}(\{k\}) = \mathcal{P}_\lambda(\{k\}).$$

Pour le voir, on observe que la suite du premier membre est

$$C_N^k \left(1 - \frac{\lambda}{N}\right)^{N-k} \left(\frac{\lambda}{N}\right)^k = \frac{N(N-1) \cdots (N-k+1)}{N^k} \left(1 - \frac{\lambda}{N}\right)^{-k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{N}\right)^N.$$

Le premier produit tend vers 1, comme quotient de deux polynômes de N de degré k ayant même terme de plus haut degré. Il est clair que toute l'expression tend vers $\frac{\lambda^k}{k!} e^{-\lambda}$ si N tend vers l'infini, par la formule connue $\lim_{N \rightarrow \infty} \left(1 + \frac{x}{N}\right)^N = \exp x$.

Définition - Proposition **La loi de Pascal et la loi négative binomiale.** Dans le schéma Succès Echec infini, intéressons nous à la loi du temps d'attente T_1 du premier succès, soit $T_1(\omega) = \inf \{n ; \omega_j = S\}$. La loi de T_1 se calcule facilement en remarquant que dire que $T_1 > n$ est dire que les n premiers essais ont été des échecs, un évènement de probabilité $(1-p)^n$. Donc, puisque

$$P(T_1 = n) = P(T_1 > n-1) - P(T_1 > n) = (1-p)^{n-1} - (1-p)^n = (1-p)^{n-1} p,$$

la loi de T_1 , dite **loi de Pascal**, ou **loi géométrique**, est

$$P_{T_1} = p\delta_1 + (1-p)p\delta_2 + \cdots + (1-p)^{n-1}p\delta_n + \cdots$$

Sa fonction génératrice est la fonction homographique $f_{T_1}(z) = \frac{pz}{1-(1-p)z}$, sa moyenne est $1/p$, un résultat qu'il est bon de retenir. Quant à sa variance, c'est $\sigma^2(T_1) = (1-p)/p^2$.

Si ensuite on s'intéresse au temps d'attente T_k du k ième succès, il est intuitivement clair, bien que pas si facile à montrer rigoureusement, que c'est la somme de k variables aléatoires indépendantes I_1, \dots, I_k , de même loi que T_1 : la v.a. I_k représente l'intervalle de temps entre les $k-1$ ième et k ième succès. La fonction génératrice est donc $f_{T_k}(z) = \left(\frac{pz}{1-(1-p)z}\right)^k$, la moyenne k/p et la variance $k(1-p)/p^2$. Toutefois, la loi de T_k est concentrée sur les entiers supérieurs ou égaux à k , et il y a avantage en vue d'une généralisation à considérer plutôt la loi de $T_k - k$, concentrée sur \mathbf{N} , de fonction génératrice

$$f_{T_k-k}(z) = \left(\frac{p}{1-(1-p)z}\right)^k = \sum_{n=0}^{\infty} \frac{1}{n!} k(k+1) \cdots (k+n-1) p^k (1-p)^n z^n,$$

en développant selon la formule du binôme de Newton. Cela entraîne donc que si $n \geq k$:

$$P(T_k = n) = P(T_k - k = n - k) = \frac{1}{(n - k)!} k(k + 1) \cdots (n - 1) p^k (1 - p)^{n - k} = C_{n-1}^{k-1} p^k (1 - p)^{n - k},$$

une formule difficile à retenir.

Maintenant, on peut généraliser la loi de $T_k - k$ en remplaçant le paramètre entier k par le paramètre continu positif λ . L'interprétation probabiliste disparaît, mais les formules demeurent. On introduit donc la loi dite négative-binomiale définie par :

Définition - Proposition **La loi négative binomiale** est la loi $NB_{\lambda,p}$ définie pour $\lambda > 0$ et $0 < p < 1$ par

$$NB_{\lambda,p} = \sum_{n=0}^{\infty} \frac{1}{n!} \lambda(\lambda + 1) \cdots (\lambda + n - 1) p^\lambda (1 - p)^n \delta_n.$$

Une variable aléatoire X qui suit une telle loi est donc telle que si $n \in \mathbb{N}$:

$$P(X = n) = \frac{1}{n!} \lambda(\lambda + 1) \cdots (\lambda + n - 1) p^\lambda (1 - p)^n,$$

sa fonction génératrice est $f_X(z) = \left(\frac{p}{1 - (1-p)z}\right)^\lambda$, sa moyenne est $\lambda(1 - p)/p$ et sa variance est $\lambda(1 - p)/p^2$.

Exercices sur 6.2

1. Montrer que si deux dés sont marqués sur leurs faces 1,2,3,4,5,6 il est impossible de les piper de sorte que la somme $X + Y$ de leur points soit telle que $P(X + Y = n) = \frac{1}{11}$ pour $n = 2, 3, \dots, 12$. Méthode: montrer que les fonctions génératrices $f_X(z)$ et $f_Y(z)$ sont telles que $f_X(z)/z$ et $f_X(z)/z$ sont des polynômes ayant au moins un zéro réel, et que $f_{X+Y}(z)/z^2$ n'a que des zéros imaginaires.
2. Une fonction génératrice f_X est telle que $f_X(z) = (1 - \sqrt{1 - z})/z$. Quelle est la probabilité pour que $X = n$? Est ce que $\mathbb{E}(X)$ existe?
3. Soit X et Y deux variables aléatoires indépendantes qui suivent des lois de Pascal pas nécessairement identiques. Soit $Z = \min(X, Y)$. Calculer pour n fixé $P(X > n)$, $P(Y > n)$, $P(Z > n)$, $P(Z = n)$. Montrer que Z suit une loi de Pascal. Exprimer sa moyenne en fonction des moyennes de X et Y .

3 Transformée de Laplace d'une variable aléatoire.

Théorème 6.7 Soit X une variable aléatoire. Soit I_X l'ensemble des z réels tels que $L_X(z) = \mathbb{E}(e^{zX})$ existe. La fonction $z \mapsto L_X(z)$ définie sur I_X est appelée la *transformée de Laplace* de X . Alors

1. L'ensemble I_X est un intervalle contenant 0.

2. Si 0 est dans l'intérieur de I_X , la transformée de Laplace est développable en série entière et les coefficients de cette série sont les $L_X^{(n)}(0)/n! = \mathbf{E}(X^n)/n!$:

$$L_X(z) = \sum_{n=0}^{\infty} \frac{\mathbf{E}(X^n)}{n!} z^n.$$

3. Si I_X est de longueur positive, la loi de X est caractérisée par sa transformée de Laplace. Plus précisément, si $I_X \cap I_Y$ est de longueur positive et si $L_X = L_Y$ sur cet intervalle, alors X et Y sont de même loi.
4. Si X et Y sont indépendantes, alors $I_{X+Y} = I_X \cap I_Y$ et, pour z dans cet intervalle: $L_{X+Y}(z) = L_X(z)L_Y(z)$.
5. Si a et b sont réels avec $a \neq 0$ alors $I_{aX+b} = \frac{1}{a}I_X$ et $L_{aX+b}(z) = \exp(bz)L_X(az)$.

Démonstration 1) Il est clair que $0 \in I_X$. Si $0 < s < z$ ou si $z < s < 0$ et si $z \in I_X$, montrons que $s \in I_X$. Cela vient du fait que $\exp(sX) \leq 1 + \exp(zX)$, comme on le voit en examinant les 4 cas $X \geq 0$ et $X < 0$, $z > 0$ et $z < 0$.

2) Si $[-a, a] \subset I_X$ avec $a > 0$, alors comme $\exp(a|X|) < \exp(aX) + \exp(-aX)$ on en déduit que $\mathbf{E}(\exp(a|X|))$ existe, et donc $\mathbf{E}(\exp|zX|)$ existe pour tout $|z| \leq a$. D'où pour un tel z

$$\left| L_X(z) - \sum_{n=0}^N \frac{\mathbf{E}(X^n)}{n!} z^n \right| = \left| \mathbf{E}(\exp(zX)) - \sum_{n=0}^N \frac{(Xz)^n}{n!} \right| = \left| \mathbf{E} \left(\sum_{n=N+1}^{\infty} \frac{(Xz)^n}{n!} \right) \right| \leq \mathbf{E} \left(\sum_{n=N+1}^{\infty} \frac{|Xz|^n}{n!} \right) = \mathbf{E} \left(\exp|zX| - \sum_{n=0}^N \frac{|Xz|^n}{n!} \right) = \mathbf{E}(Y_N).$$

La variable aléatoire Y_N décroît vers 0: un théorème de 3ème année dit que cela suffit pour entraîner que $\lim_{N \rightarrow \infty} \mathbf{E}(Y_N) = 0$; ce qui achève la démonstration du 2).

La partie 3) est beaucoup plus difficile et nous admettons ce résultat.

La partie 4) est une conséquence du théorème 5.8 appliqué à $N = 2$ et à $(X_1, X_2) = (\exp(zX), \exp(zY))$. La partie 5) est immédiate.

A cause du 2) on appelle parfois la transformée de Laplace la *fonction génératrice des moments*. C'est à éviter, pour ne pas confondre avec la fonction génératrice d'une variable aléatoire X à valeurs dans \mathbf{N} . D'ailleurs, pour un tel X , les deux notions sont reliées par $f_X(\exp z) = L_X(z)$ et l'intérieur de I_X est alors $] -\infty, \log R[$ où R est le rayon de convergence de la série entière de somme f_X . Les transformées de Laplace sont surtout utilisées pour caractériser des v.a. à densité. Nous en donnons 3 exemples importants.

Définition - Proposition **La loi normale N_{m, σ^2} .** C'est la loi la plus importante du calcul des probabilités. On l'appelle aussi une loi gaussienne, une loi de Laplace-Gauss, ou encore une seconde loi de Laplace. Si $m \in \mathbf{R}$ et si $\sigma > 0$, elle est définie par sa densité:

$$\frac{1}{\sigma\sqrt{2\pi}} \exp - \frac{(x - m)^2}{2\sigma^2}.$$

Le fait que ce soit une densité de probabilité n'est pas évident, car il faut vérifier que l'intégrale de cette fonction > 0 est 1. Si on l'admet pour le cas $m = 0$ et $\sigma = 1$, on se ramène facilement à ce cas particulier en posant $x = \sigma y + m$. Cette remarque permet alors de montrer que la transformée de Laplace d'une variable aléatoire Y de loi $N_{0,1}$ est

$$L_Y(z) = \mathbf{E}(e^{zY}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2} + zy} dy = e^{\frac{z^2}{2}}.$$

Pour voir cette dernière égalité il suffit d'écrire que la densité de $N_{z,1}$ est d'intégrale 1. Remarquons que l'intervalle d'existence est $I_Y = \mathbf{R}$

Ensuite, on remarque que si Y est de loi $N_{0,1}$, alors $X = \sigma Y + m$ est de loi N_{m,σ^2} . Pour le voir, il suffit d'écrire la fonction de répartition de X de la manière suivante:

$$F_X(x) = P(\sigma Y + m \leq x) = P(Y \leq \frac{x - m}{\sigma}) = F_Y(\frac{x - m}{\sigma});$$

on dérive alors les deux membres extrêmes de la ligne ci dessus: à gauche on obtient la densité cherchée de X , à droite en utilisant le théorème de dérivation des fonctions composées et le fait que la densité de Y est par hypothèse $\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$: ceci fournit pour X la densité de la loi N_{m,σ^2} comme annoncé.

Enfin, pour avoir la transformée de Laplace de X à partir de Y on utilise le 5) du théorème 6.7 pour obtenir que si X est de loi N_{m,σ^2} , alors

$$L_X(z) = \exp(\frac{\sigma^2 z^2}{2} + mz).$$

On déduit du 2) du théorème 6.7 qu'alors $\mathbf{E}(X) = m$ et que $\sigma^2(X) = \sigma^2$. On déduit aussi des 3) et 4) du théorème 6.7 que si X_1 et X_2 sont des variables aléatoires indépendantes et de lois respectives N_{m_1,σ_1^2} et N_{m_2,σ_2^2} , alors $X_1 + X_2$ est de loi $N_{m_1+m_2,\sigma_1^2+\sigma_2^2}$.

A propos de fonction de répartition, il faut noter que la fonction de répartition Φ de la loi $N_{0,1}$, soit

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy,$$

n'est pas élémentaire. Elle est tabulée dans tous les ouvrages.

On rencontre la loi $N_{0,1}$ dans la nature comme approximation de bien des lois. La plus ancienne est l'*approximation de Moivre Laplace de la loi binomiale*:

Théorème Approximation de Moivre Laplace de la loi binomiale Si X est de loi $B_{N,p}$, alors la loi de $\frac{X - Np}{\sqrt{Np(1-p)}}$ tend vers la loi $N_{0,1}$ dans le sens suivant: pour tout intervalle $[a,b]$ on a

$$\lim_{N \rightarrow \infty} P \left(a \leq \frac{X - Np}{\sqrt{Np(1-p)}} \leq b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{y^2}{2}} dy.$$

Une autre présentation de ce théorème de Moivre Laplace est donc

$$\lim_{N \rightarrow \infty} P \left(a\sqrt{Np(1-p)} + Np \leq X \leq b\sqrt{Np(1-p)} + Np \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{y^2}{2}} dy.$$

C'est dire que $P \left(a\sqrt{Np(1-p)} + Np \leq X \leq b\sqrt{Np(1-p)} + Np \right)$ est approchée par $\Phi(b) - \Phi(a)$. Cette approximation est à la base de la statistique.

La démonstration de ce résultat n'est pas élémentaire. Toutefois, l'usage des transformées de Laplace le rend plausible; avec le théorème 6.7, partie 5):

$$L_{\frac{X-Np}{\sqrt{Np(1-p)}}}(z) = \left(1-p + p \frac{z}{\sqrt{Np(1-p)}} \right)^N \exp \frac{-Npz}{\sqrt{Np(1-p)}} \rightarrow_{N \rightarrow \infty} \exp \frac{z^2}{2},$$

par un calcul de développement limité.

Définition - Proposition Les lois gamma $\gamma_{p,q}$. La loi exponentielle $\gamma_{1,q}$ de moyenne q est la plus importante des lois à densité après la loi normale. Elle est concentrée sur la demi droite positive, sa fonction de répartition est pour $x > 0$ $F(x) = 1 - \exp(-x/q)$ et en dérivant F , sa densité est

$$\frac{1}{q} \exp(-x/q) \mathbf{1}_{]0, +\infty[}(x).$$

On la rencontre dans la nature car c'est une loi sans mémoire: si X suit une loi exponentielle de moyenne q et si x et y sont > 0 , alors

$$P(X > x+y | X > y) = \frac{P(X > x+y)}{P(X > y)} = \frac{1 - F(x+y)}{1 - F(y)} = \exp(-x/q) = P(X > x).$$

Par exemple une ampoule électrique ne s'use pas, et le fait que nous sachions qu'elle a déjà duré un temps y ne nous donne aucune information pour savoir si elle va durer au moins un temps x à partir de maintenant.

La transformée de Laplace d'une variable aléatoire X de loi exponentielle existe sur $I_X =]-\infty, 1/q[$ et est égale à $L_X(z) = \frac{1}{1-qz}$. Ceci montre avec le théorème 6.7, 2), que $\mathbf{E}(X) = q$, $\mathbf{E}(X^2) = 2q^2$ et, par la formule de Huyghens, que $\sigma^2(X) = q^2$.

Si p est un nombre entier positif et si X_1, \dots, X_p sont des v.a. indépendantes et de même loi $\gamma_{1,q}$, la transformée de Laplace de $X_1 + \dots + X_p$ est donc $(\frac{1}{1-qz})^p$ sur $] -\infty, 1/q[$. Comme la transformée de Laplace détermine la loi, il suffit de montrer (par une intégration par parties qui permet de faire une récurrence sur p) que

$$\frac{1}{(p-1)!} \int_0^{+\infty} \exp(zx - x/q) q^{-p} x^{p-1} dx = \left(\frac{1}{1-qz} \right)^p$$

pour en déduire que la densité de $X_1 + \dots + X_p$ est

$$\frac{1}{(p-1)!} \int_0^{+\infty} \exp(-x/q) q^{-p} x^{p-1} \mathbf{1}_{]0,+\infty[}(x) :$$

c'est la densité de la loi $\gamma_{p,q}$.

En fait, comme pour la loi négative binomiale qui a été obtenue par une interpolation des entiers, il est possible dans la loi $\gamma_{p,q}$ de remplacer le paramètre entier par le paramètre $p > 0$. Pour cela on introduit une importante fonction de p appelée fonction Gamma d'Euler et définie pour $p > 0$ par

$$\Gamma(p) = \int_0^{+\infty} \exp(-x) x^{p-1} dx.$$

Une intégration par parties montre que $\Gamma(p+1) = p\Gamma(p)$. Comme $\Gamma(1) = 1$ on en tire que si p est entier $\Gamma(p) = (p-1)!$: cette fonction Gamma interpole les factorielles.

On définit alors la loi $\gamma_{p,q}$ pour $p > 0$ non nécessairement entier par sa densité :

$$\frac{1}{\Gamma} \int_0^{+\infty} \exp(-x/q) q^{-p} x^{p-1} \mathbf{1}_{]0,+\infty[}(x)$$

qui a pour transformée de Laplace $(\frac{1}{1-qz})^p$. On déduit de cette transformée de Laplace que la moyenne est pq et que la variance est pq^2 . On appelle p le paramètre de forme et q le paramètre d'échelon. En effet, on voit facilement, soit avec les fonctions de répartition, soit avec les transformées de Laplace, que si X est de loi $\gamma_{p,1}$ alors qX est de loi $\gamma_{p,q}$. Changer q est un simple changement d'unités de mesure, changer p change de façon importante l'allure de la densité.

Définition - Proposition **La loi uniforme sur $[a,b]$.** C'est la loi $U_{[a,b]}$, de densité $\frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$. Sa fonction de répartition $F(x)$ est nulle si $x < a$, égale à $\frac{x-a}{b-a}$ si $x \in [a,b]$ et égale 1 si $x > b$.

Il est facile de voir que si X est de loi $U_{[0,1]}$ alors $Y = a + (b-a)X$ est de loi $U_{[a,b]}$ (on dit aussi que Y est *uniformément répartie* sur $[a,b]$). La transformée de Laplace n'est pas spécialement remarquable. Pour $U_{[0,1]}$, c'est $L(z) = \frac{1}{z}(e^z - 1)$ si $z \neq 0$ et $L(0) = 1$. Le moment d'ordre n pour $U_{[0,1]}$ s'obtient directement à partir de la définition : c'est $1/(n+1)$. Les variables uniformes sont intensément utilisées en simulation.