Méthodes – Probabilités

✓ Fonctions génératrices : cas où elles prévalent

Rappelons ce qu'est la fonction génératrice d'une variable aléatoire X à valeurs dans $\mathbb{N}$:

$$\forall t \in [-1{,}1], \quad G_X(t) = \sum_{n=0}^{+\infty} \mathrm{P}(X = n)t^n.$$

Si on la connaît explicitement, alors les coefficients de son développement en série entière donnent la loi de X, tandis que ses dérivées successives en 1 permettent de calculer l'espérance et la variance de X.

En soi, dès qu'on connaît la loi d'une variable aléatoires à valeurs dans $\mathbb{N}$, on pourrait en toutes circonstances étudier sa fonction génératrice pour en déduire son espérance et sa variance. Mais ce n'est pas forcément plus intéressant que par application directe des définitions.

Le cas où, par contre, la fonction génératrice apporte une réelle valeur ajoutée, **est quand nous étudions une somme de variables aléatoires INDÉPENDANTES** (attention, cette hypothèse est essentielle). Dans ce cas songez-y, même si l'énoncé ne vous y invite pas. L'intérêt est que dans ce cas, la fonction génératrice est facile à calculer : si X et Y sont INDÉPENDANTES, alors :

$$\forall t \in [-1{,}1], \quad G_{X+Y}(t) = G_X(t) \cdot G_Y(t).$$

Si l'on connaît les lois de X et Y, on en déduit leurs fonctions génératrices G_X et G_Y (c'est direct si X et Y suivent des lois usuelles), et donc on en déduit leur produit G_{X+Y} (ne pas passer par la définition du produit de Cauchy : ce n'est pas assez explicite pour ce qu'on veut en général).

La connaissance de G_{X+Y} permet d'en déduire ensuite **la loi de probabilité de** $Z = X + Y$, en développant en série entière en 0 l'application $t \mapsto G_Z(t)$ qu'on a explicitée grâce à la formule ci-dessus ; si on parvient à l'écrire sous la forme : $G_Z(t) = \sum\limits_{n=0}^{+\infty} a_n t^n$ (pour tout t dans un voisinage de 0), où les a_n sont CONNUS, alors du fait que par définition on ait aussi : $\forall t \in [-1{,}1]$, $G_Z(t) = \sum\limits_{n=0}^{+\infty} \mathrm{P}(Z = n)t^n$, on en déduit par unicité des coefficients : $\forall n \in \mathbb{N}$, $a_n = \mathrm{P}(Z = n)$.

> Si vous avez des facteurs devant la somme de série entière, rentrez-les dans la somme : il faut les compter dans l'expression du coefficient devant t^n ! Si $G_Z(t) = \lambda \sum\limits_{n=0}^{+\infty} b_n t^n$, alors il est faux d'en déduire qu'on a $b_n = \mathrm{P}(Z = n)$ pour tout $n \in \mathbb{N}$: on a $\lambda b_n = \mathrm{P}(Z = n)$ pour tout $n \in \mathbb{N}$, car $G_Z(t) = \sum\limits_{n=0}^{+\infty} \lambda b_n t^n = \sum\limits_{n=0}^{+\infty} \mathrm{P}(Z = n)t^n$.

Je ne parlerai pas tant que cela de l'espérance ni la variance, bien qu'on connaisse les relations :

$$\mathrm{E}(Z) = G'_Z(1), \text{ et : } \mathrm{V}(Z) = G''_Z(1) + G'_Z(1) - (G'_Z(1))^2,$$

puisque *dans le cas précis d'une somme de variables aléatoires*, on la trouve plus directement par linéarité pour l'espérance (nul besoin d'indépendance), et dans le cas de la variance on a $\mathrm{V}(X + Y) = \mathrm{V}(X) + \mathrm{V}(Y)$ si X et Y sont indépendantes : nul besoin de fonction génératrice.

Cette méthode d'utilisation de la fonction génératrice se généralise au cas d'une somme de n variables aléatoires indépendantes. Plus il y a de variables aléatoires, et plus elle devient rentable.

Attention, si vous ne mentionnez pas l'indépendance au moment d'utiliser cette formule, vous faites un oubli crucial et le raisonnement est faux : vous perdez des points. Ce n'est pas une bête hypothèse technique.

Pour ensuite développer en série entière en 0 la fonction génératrice G_Z, songez aux méthodes répertoriées dans le document *Méthodes* du chapitre sur les séries entières, section 3 (*Développer en série entière*).

Exemple 1. Soient X et Y deux variables aléatoires INDÉPENDANTES. On suppose : $X \sim \mathcal{G}(p)$, $Y \sim \mathcal{G}(q)$, avec p et q *différents*. On cherche la loi de probabilité de $Z = X + Y$.

Le fait d'avoir une somme, et l'hypothèse d'indépendance, incite immédiatement à utiliser la fonction génératrice, même si l'énoncé ne le suggère pas. On a, sous ces hypothèses :

$$\forall t \in [-1,1], \quad G_Z(t) = G_X(t)G_Y(t) = \frac{pt}{1-t(1-p)} \cdot \frac{qt}{1-t(1-q)}.$$

On développe cette fonction en série entière en 0. Pour cela, on décompose en éléments simples, et on trouve :

$$\begin{aligned}\forall t \in [-1,1], \quad G_Z(t) &= \frac{pqt^2}{q-p}\left(\frac{1-p}{1-t(1-p)} - \frac{1-q}{1-t(1-q)}\right) \\ &= \frac{pqt^2}{q-p}\left((1-p)\sum_{n=0}^{+\infty}(1-p)^n t^n - (1-q)\sum_{n=0}^{+\infty}(1-q)^n t^n\right) \\ &= \sum_{n=0}^{+\infty} \frac{pq}{q-p}\left((1-p)^{n+1} - (1-q)^{n+1}\right) t^{n+2} \\ &= \sum_{n=2}^{+\infty} \frac{pq}{q-p}\left((1-p)^{n-1} - (1-q)^{n-1}\right) t^n.\end{aligned}$$

Or : $\forall t \in [-1,1]$, $G_Z(t) = \sum_{n=0}^{+\infty} \mathrm{P}(Z = n)t^n$. L'unicité des coefficients d'une somme de série entière implique donc :

$$\forall n \in \mathbb{N} \setminus \{0,1\}, \quad \mathrm{P}(Z = n) = \frac{pq}{q-p}\left((1-p)^{n-1} - (1-q)^{n-1}\right),$$

tandis que $\mathrm{P}(Z = 0) = \mathrm{P}(Z = 1) = 0$ (car le coefficient de t^n, pour $n = 0$ et $n = 1$, est nul dans le développement en série entière explicite de G_Z). On a obtenu la loi de probabilité de Z.

Exercice 1. Trouver la variance de Z à l'aide de la fonction génératrice (nous vous conseillons de décomposer en éléments simples G_Z pour faciliter le calcul des dérivées, en tenant compte du facteur t^2 au numérateur).

Exercice 2. S'inspirer de ce qui précède pour donner la loi de probabilité de la somme de *trois* variables aléatoires indépendantes suivant des lois géométriques de paramètres tous différents.

Exercice 3.

1. Reprendre l'exemple ci-dessus, mais dans le cas où $p = q$.
2. Généraliser à une somme de n variables aléatoires indépendantes, suivant la même loi géométrique de paramètre p.