Ve 22 mars, 15h30 - 17h30 : DS d'informatique

Pour me : 302.9, 303.8, 303.27

2 Fonctions génératrices

$X(\Omega) \subset \mathbb{N}$

On s'intéresse dans ce paragraphe aux variables aléatoires qui sont à valeurs dans $\mathbb{N}$. Typiquement, celles qui apparaissent dans des situations de comptage.

2.1 Définition

Lemme. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. La série entière :

$$\sum_{n \geqslant 0} P(X = n) t^n$$

converge normalement sur $[-1, 1]$, et son rayon de convergence satisfait : $R_X \geqslant 1$.

Preuve : On note $f_n(t) = P(X=n) t^n$

$\forall t \in [-1,1] \quad |f_n(t)| \leq P(X=n)$ indep de t

tg série cv de somme 1.

donc $\sum f_n$ cv normalement sur $[-1,1]$

donc $R \geqslant 1$.

série génératrice

Définition. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On définit la **fonction génératrice** de X par :

$$G_X : t \mapsto \sum_{n \geqslant 0} P(X = n) t^n$$

(au moins pour $t \in [-1,1]$)

Remarque. *$G_X(1) = 1$ et $G_X(t) = \mathrm{E}(t^X)$ par la formule de transfert.*

On note $f : x \longmapsto t^x$ $\quad f(X)$ est une va (fct de va)

$$E(f(X)) = \sum_{\substack{x \in X(\Omega) \\ \cap \\ \mathbb{N}}} \underbrace{f(x)}_{= t^x} P(X=x) \quad \text{par transfert}$$

Proposition. La loi d'une variable aléatoire à valeurs dans $\mathbb{N}$ est caractérisée par sa fonction génératrice.

<u>Explicat°</u> (va tq $X(\Omega) \subset \mathbb{N}$)

donner la loi d'une va X,

c'est donner : $\cdot$ $X(\Omega) \subset \mathbb{N}$

$\cdot$ $\forall n \in X(\Omega)$, $P(X=n) = \dots$

En donnant $G_X \; t \longmapsto \sum_{n=0}^{+\infty} a_n t^n$, on donne

les coeff du DSE a_n qui sont les $P(X=n)$

<u>Preuve:</u> $\boxed{\Rightarrow}$ Si X est donnée par sa loi, G_X est définie

de façon unique

$\boxed{\Leftarrow}$ Si X et Y deux va de même fct génératrice

$G : t \longmapsto \sum_{n=0}^{+\infty} a_n t^n$

par unicité des coeff d'une SE

$a_n = P(X=n) = P(Y=n)$

donc $X \sim Y$

Fonctions génératrices des lois usuelles.

- Si $X \sim \mathscr{U}([\![1, n]\!])$, alors $G_X(t) = \dfrac{1}{n}(t + t^2 + \cdots + t^n)$.

$$G_X(t) = E(t^X)$$
$$= \sum_{k=1}^{n} P(X=k)\, t^k \quad \text{par transfert}$$
$$= \frac{1}{n} \sum_{k=1}^{n} t^k \quad \text{(polynôme)}$$
$$= \frac{1}{n} \cdot t \cdot \frac{1-t^n}{1-t} \quad \text{pour } t \neq 1$$

- Si $X \sim \mathscr{B}(p)$, alors $G_X(t) = pt + (1-p)$.

$$X(\Omega) = \{0, 1\} \subset \mathbb{N}$$
$$G_X(t) = E(t^X)$$
$$= t^0 P(X=0) + t^1 P(X=1)$$
$$= (1-p) + pt$$

- Si $X \sim \mathscr{B}(n, p)$, alors $G_X(t) = (pt + (1-p))^n$.

$$X(\Omega) = [\![0, n]\!]$$
$$G_X(t) = E(t^X)$$
$$= \sum_{k=0}^{n} t^k P(X=k)$$
$$= \sum_{k=0}^{n} t^k \binom{n}{k} p^k (1-p)^{n-k}$$
$$= (pt + (1-p))^n \quad \text{par le binôme.}$$

- Si $X \sim \mathscr{G}(p)$, alors $G_X(t) = \dfrac{pt}{1-(1-p)t}$.

$X(\Omega) = \mathbb{N}^*$ $\qquad P(X=n) = q^{n-1} p \quad \forall n \in \mathbb{N}^*$

$$G_X(t) = E(t^X)$$

$$= \sum_{n=1}^{+\infty} t^n q^{n-1} p \qquad \text{par transfert}$$

$$= pt \cdot \frac{1}{1-qt} \qquad \text{somme géom de raison } qt$$

$\forall t \in [-1,1] \subset]-\frac{1}{q}, \frac{1}{q}[$

- Si $X \sim \mathscr{P}(\lambda)$, alors $G_X(t) = \mathrm{e}^{\lambda(t-1)}$.

$X(\Omega) = \mathbb{N}$ $\qquad P(X=n) = e^{-\lambda} \dfrac{\lambda^n}{n!} \quad \forall n \in \mathbb{N}$

$$G_X(t) = E(t^X)$$

$$= \sum_{n=0}^{+\infty} t^n e^{-\lambda} \frac{\lambda^n}{n!} \qquad \forall t \in [-1,1] \subset]-\infty, +\infty[$$

$$= e^{-\lambda} \cdot e^{t\lambda}$$

$$= e^{\lambda(t-1)}$$

2.2 Propriétés

Proposition. On conserve les notations précédentes. G_X est continue sur $[-1,1]$.

Proposition. On conserve les notations précédentes.
X admet une espérance finie si et seulement si G_X est dérivable en 1 (à gauche).
Dans ce cas :

$$\mathrm{E}(X) = G'_X(1)$$

Proposition. On conserve les notations précédentes.
X admet une variance si et seulement si G_X est deux fois dérivable en 1 (à gauche).
Dans ce cas :

$$G''_X(1) = \mathrm{E}(X(X-1))$$

Remarque. *De cette égalité, il faut savoir retrouver rapidement l'expression de la variance à l'aide de G :*

$$\mathrm{V}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$$

Exemple. Retrouver par les fonctions génératrices espérance et variance des lois usuelles.

Preuve.

On y croit ?

$$G_X(t) = \sum_{n=0}^{+\infty} P(X=n)\,t^n$$

Formellement

$$G'_X(t) \overset{?}{=} \sum_{n=1}^{+\infty} P(X=n)\,n\,t^{n-1} \qquad \forall t \in]-R,R[$$

Formellement

$$G'_X(1) = \sum_{n=\cancel{1}\,0}^{+\infty} n P(X=n)$$

$$= E(X)$$

Si $R>1$ $\longrightarrow$ facile

Si $R=1$ on ne peut pas faire le calcul précédent.

$\boxed{\Rightarrow}$ On suppose X d'espérance finie ie $\sum_{n\geq 0} n P(X=n)$ CV

Mq G_X est dérivable en 1 (à gauche)

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

$$= \sum_{n=0}^{+\infty} P(X=n) \underbrace{\frac{t^n-1}{t-1}}_{u_n(t)}$$

$$\xrightarrow[t \to 1^-]{} \ ?$$

Appliquons le th de double limite :

* $u_n(t) = P(X=n)\left(1 + t + \dots + t^{n-1}\right)$

$$\xrightarrow[t \to 1^-]{} P(X=n)\,(1+1+\dots+1) = n\,P(X=n)$$

* $\forall t \in [0,1[$

$|u_n(t)| = P(X=n)\,(1+t+\dots+t^{n-1})$

$\leq P(X=n)\,n$ indep de t

tg d'une série cv par hyp.

donc $\sum u_n$ cv normalement, donc uniformément

sur $[0,1[$.

Par double limite

$$\frac{G_X(t) - G_X(1)}{t-1} \xrightarrow[t \to 1^-]{} \sum_{n=0}^{+\infty} n\,P(X=n) = E(X)$$

donc G_X est dérivable en 1 (à gauche)

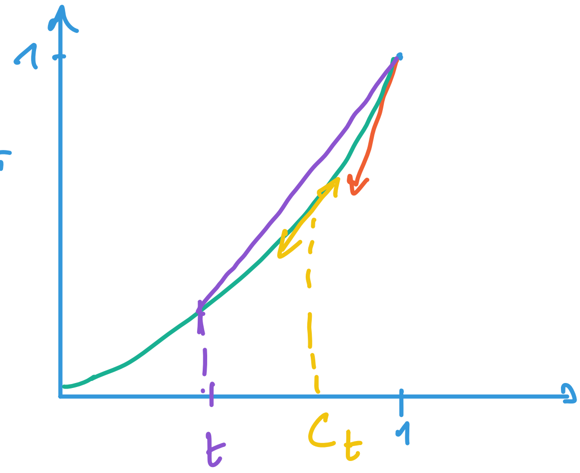
et $G_X'(1) = E(X)$.

$\boxed{\Leftarrow}$ On suppose G_X dérivable en 1 (à gauche)

Soit $t \in [0,1[$

Appliquons le th des accroissements finis à G_X entre t et 1.

D'où l'existence de



$c_t \in]t,1[$ tq

$$G_X'(c_t) = \frac{G_X(t) - G_X(1)}{t-1}$$

$$\| \quad \sum_{n=1}^{+\infty} P(X=n)\, n\, c_t^{n-1}$$

somme de termes ≥ 0

Prenons $N \in \mathbb{N}$, passons par les sommes partielles.

$$\sum_{n=1}^{N} P(X=n)\, n\, c_t^{n-1} \leq \frac{G_X(t) - G_X(1)}{t-1}$$

à N fixé, on passe à la limite pour $t \xrightarrow{<} 1$ dans l'inégalité large (somme d'un nb fini de termes)

$$\sum_{n=1}^{N} P(X=n)\, n \leq G_X'(1)$$

Donc la somme partielle de la série à termes positifs $\sum n P(X=n)$ est majorée, donc $\sum n P(X=n)$ converge si X est d'espérance finie.

On peut reprendre $\boxed{\Rightarrow}$ et obtenir $E(X) = G_X'(1)$

Proposition. On conserve les notations précédentes.
X admet une variance si et seulement si G_X est deux fois dérivable en 1 (à gauche).
Dans ce cas :

$$G_X''(1) = \mathrm{E}(X(X-1))$$

Remarque. *De cette égalité, il faut savoir retrouver rapidement l'expression de la variance à l'aide de G :*

$$\mathrm{V}(X) = G_X''(1) + G_X'(1) - (G_X'(1))^2$$

Exemple. Retrouver par les fonctions génératrices espérance et variance des lois usuelles.

Pour $R > 1$, $\forall t \in]-R, R[$

$$G_X'(t) = \sum_{n=1}^{+\infty} P(X=n)\, n\, t^{n-1}$$

$$G_X''(t) = \sum_{n=2}^{+\infty} P(X=n)\, n(n-1) t^{n-2}$$

donc $G_X''(1) = \displaystyle\sum_{n=\cancel{2}\ 0}^{+\infty} n(n-1) P(X=n) = E(X(X-1))$

par transfert

Exemple. Retrouver par les fonctions génératrices espérance et variance des lois usuelles.

- Soit $X \sim \mathcal{G}(p)$

$$G_X(t) = \frac{pt}{1-qt}$$

~~définie~~ $\mathcal{C}^\infty$ sur $]-\frac{1}{q}, \frac{1}{q}[$

en particulier G_X est 2 fois dérivable en 1

donc X est d'espérance finie et X^2 est d'esp. finie

$$G_X'(t) = \frac{p(1-qt) - pt(-q)}{(1-qt)^2}$$

$$= \frac{p}{(1-qt)^2} \qquad \text{donc } G'_X(1) = \frac{p}{(1-q)^2} = \frac{1}{p}$$

$$\text{donc } E(X) = \frac{1}{p}$$

$$G''_X(t) = p \cdot (-2) \cdot (-q) \cdot \frac{1}{(1-qt)^3}$$

$$\text{donc } G''_X(1) = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2}$$

$$\text{Donc } V(X) = E(X^2) - E(X)^2$$

$$= E(X(X-1)) + E(X) - E(X)^2$$

$$= G''_X(1) + \frac{1}{p} - \left(\frac{1}{p}\right)^2$$

$$= \frac{2q}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2$$

- Soit $X \sim \mathcal{P}(\lambda)$

$$\forall t \in \mathbb{R}, \quad G_X(t) = e^{\lambda(t-1)} \qquad \mathcal{C}^\infty$$

$$G'_X(t) = \lambda e^{\lambda(t-1)}$$

$$G''_X(t) = \lambda^2 e^{\lambda(t-1)}$$

$$\text{donc } E(X) = G'_X(1) = \lambda$$

$$E(X(X-1)) = G''_X(1) = \lambda^2 \qquad V(X) = \lambda$$

2.3 Fonction génératrice et somme

Proposition. Soit X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$. Si X et Y sont indépendantes, alors pour tout $t \in]-1,1[$:

$$G_{X+Y}(t) = G_X(t)\,G_Y(t)$$

où $G_X(t)\,G_Y(t)$ est le produit de Cauchy des deux séries entières.

$$G_{X+Y}(t) = E\left(t^{X+Y}\right) \qquad \forall t \in [-1,1]$$

$$= E\left(t^X . t^Y\right)$$

$$= E\left(t^X\right) E\left(t^Y\right) \quad \text{par indép de } t^X \text{ et } t^Y$$

$$= G_X(t)\; G_Y(t)$$

Proposition. Soit X et Y deux variables aléatoires indépendantes suivant des lois de Poisson de paramètres respectifs λ et μ. Alors :

$$X + Y \sim \mathscr{P}(\lambda + \mu)$$

$X \sim \mathcal{P}(\lambda)$ $\qquad X(\Omega) = \mathbb{N}$ $\qquad P(X=n) = e^{-\lambda}\dfrac{\lambda^n}{n!}$

$Y \sim \mathcal{P}(\mu)$ $\qquad Y(\Omega) = \mathbb{N}$ $\qquad P(Y=n) = e^{-\mu}\dfrac{\mu^n}{n!}$

$Z = X+Y$ $\qquad$ loi de Z ?

$\boxed{M1}$ <u>Loi d'une somme de 2 va ?</u>

- $Z(\Omega) = \mathbb{N}$
- Pour $n \in \mathbb{N}$

$$P(X+Y=n) = \sum_{k=0}^{+\infty} P(X+Y=n, X=k)$$

$$= \sum_{k=0}^{+\infty} P(Y=n-k, X=k)$$

$$= \sum_{k=0}^{+\infty} P(X=k)\; P(Y=n-k)$$

par indépendance

$$= \sum_{k=0}^{n} e^{-\lambda}\frac{\lambda^{k}}{k!} \; e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}$$

$$= \frac{e^{-\lambda}e^{-\mu}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda^{k} \mu^{n-k}$$

$$= e^{-(\lambda+\mu)} \cdot \frac{1}{n!} (\lambda+\mu)^{n}$$

par le binôme.

donc $X+Y \sim \mathcal{P}(\lambda+\mu)$

$\boxed{M2}$ <u>Par les fonctions génératrices.</u>

$\forall t \in [-1,1]$

$$G_{X+Y}(t) = G_X(t)\; G_Y(t)$$

par indépendance

$$= e^{\lambda(t-1)}\; e^{\mu(t-1)}$$

$$= e^{(\lambda+\mu)(t-1)}$$

on reconnait la fct génératrice de $\mathcal{P}(\lambda+\mu)$

donc $\underline{X+Y \sim \mathcal{P}(\lambda+\mu)}$

Proposition. Soit $X_1, X_2, \dots, X_n$ des variables aléatoires à valeurs dans $\mathbb{N}$. Si elles sont indépendantes, alors pour tout $t \in]-1,1[$:

$$G_{X_1+X_2+\cdots+X_n}(t) = G_{X_1}(t)\, G_{X_2}(t)\, \dots\, G_{X_n}(t)$$

Preuve: par récurrence sur $n \in \mathbb{N}^*$.

- clair pour $n=1$.
- on suppose le résultat vrai pour $n \in \mathbb{N}^*$.

ie: $\forall X_1 \dots X_n$ va indép, $G_{X_1+\dots+X_n}(t) = G_{X_1}(t) \dots G_{X_n}(t)$

Soit $X_1 \dots X_n, X_{n+1}$ va indépendantes

Pour $t \in [-1,1]$:

$$G_{\underbrace{(X_1 + \dots + X_n)}_{Y_1} + \underbrace{(X_{n+1})}_{Y_2}}(t)$$

Y_1, Y_2 indép par coalition

donc par la prop précédente

$$G_{(X_1+\dots+X_n)+(X_{n+1})}(t) = G_{X_1+\dots+X_n}(t) \times G_{X_{n+1}}(t)$$

$$= G_{X_1}(t) \dots G_{X_n}(t) \times G_{X_{n+1}}(t)$$

par HR

3 Loi faible des grands nombres

Loi faible des grands nombres.

Soit $(X_n)_{n\in\mathbb{N}}$ une suite i.i.d. de variables aléatoires de variance finie. On note $m = \mathrm{E}(X_1)$ (les espérances sont toutes égales), $\sigma = \sigma(X_1)$ (les écarts-types sont tous égaux) et :

$$S_n = X_1 + X_2 + \cdots + X_n$$

Alors, pour tout $\varepsilon > 0$:

$$P\left(\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{n\varepsilon^2}$$

et donc :

$$P\left(\left|\frac{S_n}{n} - m\right| \geqslant \varepsilon\right) \xrightarrow[n\to+\infty]{} 0$$

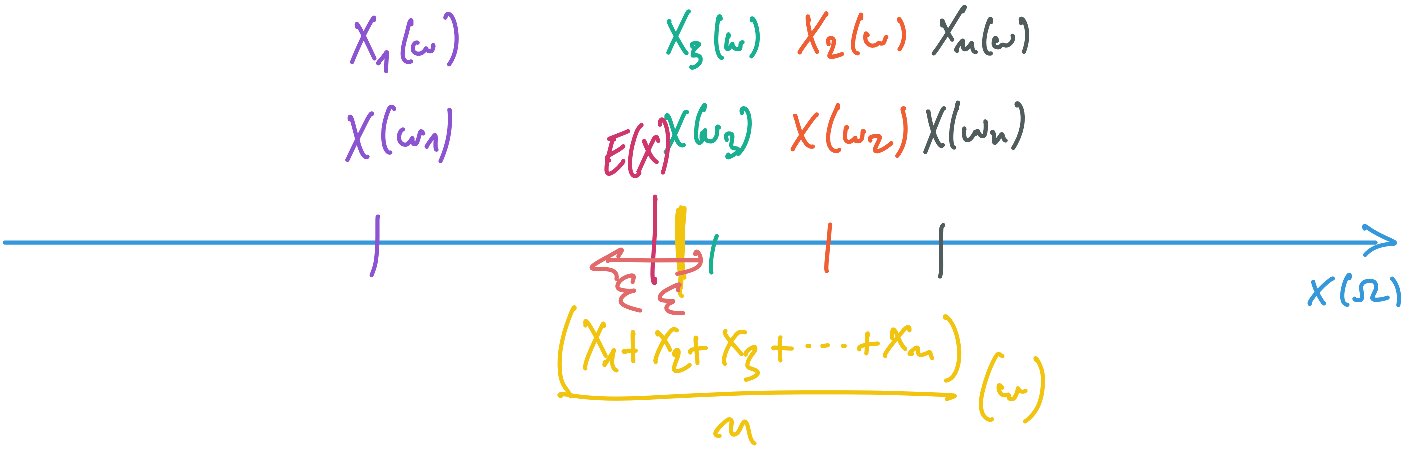
Remarque. *Pour déterminer la probabilité d'un événement A, l'idée courante consiste à répéter l'expérience aléatoire un grand nombre de fois, et observer le nombre d'apparition de l'événement A. Lorsque le nombre d'expérience augmente, la fréquence d'apparition de A devrait se rapprocher de la probabilité de A.*
La loi faible des grands nombres formalise et quantifie cette intuition : En notant $p = P(A)$ et X_k l'indicatrice de A, on a $X_k \sim \mathscr{B}(p)$ et X_k représente le nombre de fois que A a été observé à l'expérience k. Ainsi, la fréquence d'apparition de A au cours des n répétitions est la variable aléatoire :

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Par la loi faible des grands nombres,

$$\forall\varepsilon > 0,\ P(|M_n - p| < \varepsilon) \xrightarrow[n\to+\infty]{} 1$$

On peut dire que l'événement « M_n tend vers $p = P(A)$ » est quasi-certain.



Preuve: On applique l'inég de Bienaymé-Tchebycheff à $\frac{S_n}{n}$

$$E\left(\frac{S_n}{n}\right) = E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)$$

$$= \frac{1}{n}\sum_{k=1}^{n} E(X_k) \quad \text{par linéarité}$$

$$= m$$

$$V\left(\frac{S_n}{n}\right) = \frac{1}{n^2} V(X_1 + \dots + X_n)$$

$$= \frac{1}{n^2}\left(V(X_1) + \dots + V(X_n)\right)$$

par indépendance

$$= \frac{V(X)}{n}$$

donc $$P\left(\left|\frac{S_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{V(X)}{n\varepsilon^2} \xrightarrow[n \to +\infty]{} 0$$

Remarque: on peut approcher numériquement $E(X)$ en faisant la moyenne des valeurs de X sur un grand nb de réalisations de l'exp. aléatoire