Grand classique

Somme de variables aléatoires indépendantes à densité

Définition 1 — Produit de convolution

Soient f et g deux fonctions définies sur $\mathbb{R}$ et continues sur $\mathbb{R}$ sauf éventuellement en un nombre fini de points.

Le **produit de convolution** de f et g est la fonction $x \longmapsto \int_{-\infty}^{+\infty} f(x-t)g(t)dt$.

On notera $f * g$ cette fonction lorsqu'elle est existe.

Petite remarque

Par changement de variable affine, on remarque que $\int_{-\infty}^{+\infty} f(x-t)g(t)dt = \int_{-\infty}^{+\infty} f(t)g(x-t)dt$.

D'où : $f * g = g * f$ (cohérent avec l'appellation).

Lemme 1 — Existence du produit de convolution

Soient f et g deux fonctions définies sur $\mathbb{R}$.
Si : f et g sont des densités de probabilités et que f (ou g) est bornée sur $\mathbb{R}$, alors la fonction $f * g$ est définie et continue sur $\mathbb{R}$.

⋆ Démonstration : Supposons que f et g sont deux densités de probabilités. Par symétrie de f et g, supposons que f est continue et bornée sur $\mathbb{R}$.

- **Existence de $f * g$.**

  Soit $x \in \mathbb{R}$. Montrons que l'intégrale $\int_{-\infty}^{+\infty} f(x-t)g(t)dt = \int_{-\infty}^{+\infty} f(t)g(x-t)dt$ est convergente.

  ⋆ Puisque f est bornée sur $\mathbb{R}$, il existe un réel $M \in \mathbb{R}$, que nous considérons ensuite, tel que : $\forall y \in \mathbb{R},\ f(y) \leqslant M$.
  Ainsi, f étant positive (c'est une densité), on a :

  $$\forall t \in \mathbb{R},\ 0 \leqslant f(x-t) \leqslant M$$

  D'où, g étant positive sur $\mathbb{R}$ (c'est une densité) :

  $$\forall t \in \mathbb{R},\ 0 \leqslant f(x-t)g(t) \leqslant Mg(t)$$

  ⋆ Mais g est une densité, donc $\int_{-\infty}^{+\infty} g(t)dt$ est convergente (et vaut même 1). D'où la convergence de $\int_{-\infty}^{+\infty} Mg(t)dt$.

  Par critère de comparaison sur les intégrales à intégrande positive, on en déduit que l'intégrale $\int_{-\infty}^{+\infty} f(x-t)g(t)dt$ est convergente.
  On a ainsi établi : pour tout $x \in \mathbb{R}$, $f * g(x)$ existe. Autrement dit, la fonction $f * g$ est définie sur $\mathbb{R}$.

- **Continuité de $f * g$.**
  On admet la continuité de $f * g$ sous les hypothèse faites (nécessite des connaissances qui sortent du programme d'ECG). ⋆

☞ Pour info...

En intégrant cet encadrement, on constate que $f * g$ est positive et majorée... par M !

Théorème 1 — Loi de la somme de deux VA indépendantes

Soient X et Y deux variables aléatoires de densités respectives f_X et f_Y.
Si X et Y sont indépendantes et que la fonction $f_X * f_Y$ existe sur $\mathbb{R}$, sauf éventuellement un nombre fini de points, alors la variable aléatoire $X+Y$ est à densité et la fonction $f_X * f_Y$ en est une densité.

Important !

D'après le lemme précédent, si f_X ou f_Y bornée sur $\mathbb{R}$, alors $X+Y$ est à densité et $f_X * f_Y$ en est une densité sur $\mathbb{R}$.

⋆ Démonstration : Largement hors programme... ⋆

Exemples 1

E1 Soient X et Y deux variables aléatoires indépendantes suivant toutes deux la loi exponentielle de paramètre 1. Montrons que $X+Y$ est à densité et donnons-en une densité.
Les variables aléatoires X et Y suivent toutes deux la loi exponentielle de paramètre 1, dont une densité est la fonction $f : x \longmapsto \begin{cases} e^{-x} & \text{si } x \geqslant 0 \\ 0 & \text{si } x < 0 \end{cases}$.
Notons $Z = X+Y$.

- Puisque X et Y suivent une loi exponentielle, on considère $X(\Omega) = \mathbb{R}^+$ et $Y(\Omega) = \mathbb{R}^+$. D'où : $Z(\Omega) \subset \mathbb{R}^+$.
- Puisque X et Y sont indépendantes et que f est bornée, on a :

  $$f * f \text{ est définie et continue sur } \mathbb{R}\text{; et } X+Y \text{ est à densité, de densité } f * f.$$

  Notons $f_Z = f * f$. Soit $x \in \mathbb{R}$. Distinguons deux cas.

  ⋆ Si $x \in]-\infty; 0[$:
  Puisque $Z(\Omega) \subset \mathbb{R}^+$, on a $f_Z(x) = 0$.

Pourquoi ?

f est constante sur $]-\infty; 0[$, positive et majorée par 1 (car décroissante...) sur $\mathbb{R}^+$.

Par le calcul :

Supposons $x < 0$.
$f * f(x) = \int_{-\infty}^{+\infty} f(x-t)f(t)dt = \int_{0}^{+\infty} f(x-t)f(t)dt$ (car f est nulle sur $\mathbb{R}_*^-$). Mais, comme $x < 0$, on a : $\forall t \geqslant 0,\ x-t < 0$. D'où : $\forall t \geqslant 0,\ f(x-t) = 0$. Et ainsi, $f * f(x) = 0$.

⋆ Si $x \in [0;+\infty[$:

$$\begin{aligned}
f_Z(x) &= f*f(x)\\
&= \int_{-\infty}^{+\infty} f(x-t)f(t)dt && f \text{ est nulle sur }]-\infty;0[\\
&= \int_{0}^{+\infty} f(x-t)f(t)dt && \forall t>x,\ f(x-t)=0\\
&= \int_{0}^{x} f(x-t)f(t)dt && x \geqslant 0, \text{ donc pour tout } t\in[0;x],\ t\geqslant 0 \text{ et } x-t\geqslant 0\\
&= \int_{0}^{x} \mathrm{e}^{-(x-t)}\mathrm{e}^{-t}dt\\
&= \int_{0}^{x} \mathrm{e}^{-x}dt\\
&= x\mathrm{e}^{-x}
\end{aligned}$$

Conclusion : la variable aléatoire $X+Y$ est à densité et a pour densité la fonction $f_Z : x \longmapsto \begin{cases} x\mathrm{e}^{-x} & \text{si } x\geqslant 0\\ 0 & \text{si } x<0\end{cases}$.

Vérification
On vérifie la continuité de f_Z sur $\mathbb{R}$...

E2 Soient X et Y deux variables aléatoires indépendantes suivant toutes deux la loi uniforme sur $[0;1]$. Montrons que $X+Y$ est à densité et donnons-en une densité.

Les variables aléatoires X et Y suivent toutes deux la loi uniforme sur $[0;1]$, dont une densité est $f : x \longmapsto \begin{cases} 1 & \text{si } x\in[0;1]\\ 0 & \text{sinon}\end{cases}$.

Notons $Z = X+Y$.

- Puisque X et Y suivent la loi uniforme sur $[0;1]$, on considère $X(\Omega)=[0;1]$ et $Y(\Omega)=[0;1]$. D'où : $Z(\Omega)\subset[0;2]$.
- Puisque X et Y sont indépendantes et que f est bornée, on a :

$$f*f \text{ est définie et continue sur } \mathbb{R}\text{; et } X+Y \text{ est à densité, de densité } f*f.$$

Notons $f_Z = f*f$. Soit $x\in\mathbb{R}$. Distinguons deux cas.

⋆ Si $x\in]-\infty;0[\cup]2;+\infty[$:
Puisque $Z(\Omega)\subset[0;2]$, on a $f_Z(x)=0$.

⋆ Si $x\in[0;2]$:

$$\begin{aligned}
f_Z(x) &= f*f(x)\\
&= \int_{-\infty}^{+\infty} f(x-t)f(t)dt && f \text{ est nulle en dehors de } [0;1]\\
&= \int_{0}^{1} f(x-t)f(t)dt\\
&= \int_{0}^{1} f(x-t)dt
\end{aligned}$$

✘ Attention !
On ne peut pas remplacer $f(x-t)$ par 1 pour tous les $t\in[0;1]$! En effet, si $x=2$ et $t=0,5$, alors $x-t=1,5\notin[0;1]$... Vigilance constante !

Or :

$$\begin{cases} 0\leqslant t\leqslant 1\\ 0\leqslant x-t\leqslant 1\end{cases} \iff \begin{cases} 0\leqslant t\leqslant 1\\ x-1\leqslant t\leqslant x\end{cases} \iff \max\big(0;x-1\big)\leqslant t\leqslant \min\big(1;x\big)$$

♣ Méthode !
• Pour calculer $\int_0^1 f(x-t)dt$, on a besoin de remplacer $f(x-t)$. On sait que $t\in[0;1]$. La question est donc de savoir quand $x-t\in[0;1]$ pour savoir quand $f(x-t)=1$...
• L'écriture avec min et max n'est pas nécessaire mais peut aider à la compréhension.

Distinguons des cas.

◇ Si $x\in[0;1]$:
Dans ce cas $x-1\leqslant 0$, et ainsi :

$$\max\big(0;x-1\big)=0 \quad ; \quad \min\big(1;x\big)=x$$

D'après ce qui précède, on a ainsi :

$$\begin{aligned}
\int_0^1 f(x-t)dt &= \int_0^x f(x-t)dt && \forall t\in[0;x],\ x-t\in[0;1], \text{ car } x\leqslant 1\\
&= \int_0^x 1dt\\
&= x
\end{aligned}$$

◇ Si $x\in]1;2]$:
Dans ce cas $x-1>0$, et ainsi :

$$\max\big(0;x-1\big)=x-1 \quad ; \quad \min\big(1;x\big)=1$$

D'après ce qui précède, on a ainsi :

$$\begin{aligned}
\int_0^1 f(x-t)dt &= \int_{x-1}^1 f(x-t)dt && \forall t\in[x-1;1],\ x-t\in[0;1], \text{ car } x>1\\
&= \int_{x-1}^1 1dt\\
&= 1-(x-1)\\
&= 2-x
\end{aligned}$$

Conclusion : la variable aléatoire $X+Y$ est à densité et a pour densité la fonction $f_Z : x \longmapsto \begin{cases} x & \text{si } x\in[0;1]\\ 2-x & \text{si } x\in]1;2]\\ 0 & \text{sinon}\end{cases}$.

Vérification
On vérifie la continuité de f_Z sur $\mathbb{R}$...

Propriétés 1 — Stabilité des lois normales

P1# Soient $\mu_1, \mu_2 \in \mathbb{R}$ et $\sigma_1, \sigma_2 \in \mathbb{R}_*^+$ ainsi que X_1 et X_2 deux variables aléatoires définies sur le même espace probabilisé.
On a :

$$\left.\begin{array}{l} X_1 \hookrightarrow \mathcal{N}(\mu_1; \sigma_1^2) \\ X_2 \hookrightarrow \mathcal{N}(\mu_2; \sigma_2^2) \\ X_1 \text{ et } X_2 \text{ sont indépendantes} \end{array}\right\} \Longrightarrow X_1 + X_2 \hookrightarrow \mathcal{N}(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2)$$

P2# Soient $(\mu_k)_{k\in\mathbb{N}^*}$ une suite de réels, $(\sigma_k)_{k\in\mathbb{N}^*}$ une suite de réels strictement positifs ainsi que $(X_k)_{k\in\mathbb{N}^*}$ une suite de variables aléatoires définies sur le même espace probabilisé.
On a :

$$\left.\begin{array}{l} \forall k \in \mathbb{N}^*,\ X_k \hookrightarrow \mathcal{N}(\mu_k; \sigma_k^2) \\ X_1, X_2, \ldots \text{ sont indépendantes} \end{array}\right\} \Longrightarrow \forall n \in [\![2; +\infty[\![,\ X_1+X_2+\ldots+X_n \hookrightarrow \mathcal{N}(\mu_1+\mu_2+\ldots+\mu_n; \sigma_1^2+\sigma_2^2+\ldots+\sigma_n^2)$$

Petite remarque
On retient que la somme de deux VA indépendantes suivant une loi normale suit encore une loi normale. Les propriétés sur l'espérance et la variance permettent de retrouver les paramètres...

★ **Démonstration** :

P1# Supposons : $\left\{\begin{array}{l} X_1 \hookrightarrow \mathcal{N}(\mu_1; \sigma_1^2) \\ X_2 \hookrightarrow \mathcal{N}(\mu_2; \sigma_2^2) \\ X_1 \text{ et } X_2 \text{ sont indépendantes} \end{array}\right.$. Notons f_1 et f_2 des densités respectives de X_1 et X_2. On a ainsi :

$$\forall x \in \mathbb{R},\ f_1(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) \quad ; \quad f_2(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right)$$

Notons $Z = X_1 + X_2$.

- Puisque X_1 et X_2 suivent des lois normales, on considère $X_1(\Omega) = \mathbb{R}$ et $X_2(\Omega) = \mathbb{R}$. D'où : $Z(\Omega) \subset \mathbb{R}$.
- Puisque X_1 et X_2 sont indépendantes et que f_1 (f_2 également) est bornée, on a :

$f_1 * f_2$ est définie et continue sur $\mathbb{R}$; et $X_1 + X_2$ est à densité, de densité $f_1 * f_2$.

Pourquoi ?
f_1 est positive, croissante sur $]-\infty; \mu_1]$ et décroissante sur $[\mu_1; +\infty[$, donc elle admet un maximum en μ_1.
Elle est donc bornée sur $\mathbb{R}$...

Notons $f_Z = f_1 * f_2$.

- ⋆ Commençons par traiter un cas particulier : si $\mu_1 = \mu_2 = 0$.
  Soit $x \in \mathbb{R}$. On a :

$$\begin{aligned} f_Z(x) &= f_1 * f_2(x) \\ &= \int_{-\infty}^{+\infty} f_1(x-t) f_2(t)\,dt \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\left(\frac{x-t}{\sigma_1}\right)^2\right) \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\frac{t}{\sigma_2}\right)^2\right) dt \qquad \sigma_1, \sigma_2 > 0 \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(\frac{-x^2}{2\sigma_1^2} + \frac{2xt}{2\sigma_1^2} - \frac{t^2}{2\sigma_1^2} - \frac{t^2}{2\sigma_2^2}\right) dt \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(\frac{-x^2}{2\sigma_1^2}\right) \int_{-\infty}^{+\infty} \exp\left(\frac{2xt}{2\sigma_1^2} - \frac{t^2}{2\sigma_1^2} - \frac{t^2}{2\sigma_2^2}\right) dt \end{aligned}$$

Petite remarque
L'argument $\sigma_1, \sigma_2 > 0$ n'est pas nécessaire, car on donne souvent une expression de la densité dans laquelle apparait $\dfrac{1}{\sqrt{2\pi}\sigma}$ au lieu de $\dfrac{1}{\sqrt{2\pi\sigma^2}}$...

Or, pour tout $t \in \mathbb{R}$:

$$\begin{aligned} \frac{2xt}{2\sigma_1^2} - \frac{t^2}{2\sigma_1^2} - \frac{t^2}{2\sigma_2^2} &= -\left(\frac{(\sigma_1^2+\sigma_2^2)t^2}{2\sigma_1^2\sigma_2^2} - \frac{2xt}{2\sigma_1^2}\right) \qquad \text{en posant } \sigma = \sqrt{\sigma_1^2+\sigma_2^2} \\ &= \frac{-1}{2}\left(\frac{\sigma^2}{\sigma_1^2\sigma_2^2}t^2 - 2\frac{x}{\sigma_1^2}t\right) \\ &= \frac{-1}{2}\left(\left(\frac{\sigma}{\sigma_1\sigma_2}t - \frac{x\sigma_2}{\sigma\sigma_1}\right)^2 - \frac{x^2\sigma_2^2}{\sigma^2\sigma_1^2}\right) \\ &= \frac{-1}{2}\left(\frac{\sigma}{\sigma_1\sigma_2}t - \frac{x\sigma_2}{\sigma\sigma_1}\right)^2 + \frac{x^2\sigma_2^2}{2\sigma^2\sigma_1^2} \end{aligned}$$

☞ Rappel...
Si $a \neq 0$:
$a^2t^2 - 2bt = \left(at - \dfrac{b}{a}\right)^2 - \left(\dfrac{b}{a}\right)^2$
C'est une sorte de mise sous forme canonique...

D'où :

$$\begin{aligned} f_Z(x) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(\frac{-x^2}{2\sigma_1^2} + \frac{x^2\sigma_2^2}{2\sigma^2\sigma_1^2}\right) \int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2}\left(\frac{\sigma}{\sigma_1\sigma_2}t - \frac{x\sigma_2}{\sigma\sigma_1}\right)^2\right) dt \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(\frac{x^2(-\sigma^2+\sigma_2^2)}{2\sigma^2\sigma_1^2}\right) \int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2}u^2\right) \frac{\sigma_1\sigma_2}{\sigma} du \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(\frac{-x^2}{2\sigma^2}\right) \frac{\sigma_1\sigma_2}{\sigma} \int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2}u^2\right) du \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-1}{2}\left(\frac{x}{\sigma}\right)^2\right) \end{aligned}$$

changement de variable $u = \dfrac{\sigma}{\sigma_1\sigma_2}t - \dfrac{x\sigma_2}{\sigma\sigma_1}$, licite car affine

$\sigma^2 = \sigma_1^2 + \sigma_2^2$

$\displaystyle\int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = \sqrt{2\pi}$

On reconnaît la densité de la loi normale $\mathcal{N}(0; \sigma^2)$.
Par conséquent : $Z \hookrightarrow \mathcal{N}(0; \sigma_1^2 + \sigma_2^2)$.

- ⋆ Cas général.
  Puisque $X_1 \hookrightarrow \mathcal{N}(\mu_1; \sigma_1^2)$ et $X_2 \hookrightarrow \mathcal{N}(\mu_2; \sigma_2^2)$, on a :

$$X_1 - \mu_1 \hookrightarrow \mathcal{N}(0; \sigma_1^2) \quad ; \quad X_2 - \mu_2 \hookrightarrow \mathcal{N}(0; \sigma_2^2)$$

D'après le cas précédent, on en déduit :

$$X_1 + X_2 - (\mu_1 + \mu_2) \hookrightarrow \mathcal{N}(0; \sigma_1^2 + \sigma_2^2)$$

Et ainsi :

$$X_1 + X_2 \hookrightarrow \mathcal{N}(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2)$$

Conclusion : $X_1 + X_2 \hookrightarrow \mathcal{N}(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2)$.

P2# Supposons que $\left\{\begin{array}{l} \forall k \in \mathbb{N}^*,\ X_k \hookrightarrow \mathcal{N}(\mu_k; \sigma_k^2) \\ X_1, X_2, \ldots \text{ sont indépendantes} \end{array}\right.$. Procédons ensuite par récurrence...

- **Initialisation.** Pour $n = 2$:
  C'est P1.
- **Hérédité.** Soit $n \in [\![2; +\infty[\![$. Supposons $X_1 + X_2 + \ldots + X_n \hookrightarrow \mathcal{N}(\mu_1 + \mu_2 + \ldots + \mu_n; \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2)$ et montrons $X_1 + X_2 + \ldots + X_{n+1} \hookrightarrow \mathcal{N}(\mu_1 + \mu_2 + \ldots + \mu_{n+1}; \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_{n+1}^2)$.
  Notons $Z_n = X_1 + X_2 + \ldots + X_n$. On a :
  $$X_1 + X_2 + \ldots + X_{n+1} = Z_n + X_{n+1}$$
  Mais :
  - ⋆ par hypothèse de récurrence, $Z_n \hookrightarrow \mathcal{N}(\mu_1 + \mu_2 + \ldots + \mu_n; \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2)$;
  - ⋆ les variables aléatoires $X_1, X_2, \ldots, X_n, X_{n+1}$ sont indépendantes, donc, par lemme des coalitions : Z_n et X_{n+1} sont indépendantes ;
  - ⋆ $X_{n+1} \hookrightarrow \mathcal{N}(\mu_{n+1}; \sigma_{n+1}^2)$

  D'où, d'après P1 :
  $$Z_n + X_{n+1} \hookrightarrow \mathcal{N}(\mu_1 + \mu_2 + \ldots + \mu_{n+1}; \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_{n+1}^2)$$
  L'hérédité est ainsi établie.

Conclusion : $\forall n \in [\![2; +\infty[\![,\ X_1 + X_2 + \ldots + X_n \hookrightarrow \mathcal{N}(\mu_1 + \mu_2 + \ldots + \mu_n; \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2)$.

⋆

Petite remarque

On est dans un cas classique de "si c'est vrai pour 2, alors c'est vrai pour $n \in [\![2; +\infty[\![$"... Que l'on démontre par récurrence, en utilisant le cas de 2 (et le lemme des coalitions ici).