25 – UN CALCUL ÉLÉMENTAIRE DE P(X<Y) QUAND X ET Y SONT DIFFUSES

www.daniel-saada.eu

Dans toute cette note, X et Y sont des variables aléatoires réelles et indépendantes sur $(\Omega, \mathcal{T}, p)$.
Le calcul de $p(X<Y)$ ne pose pas de problémes théoriques si l'une des deux variables est discrète : si $Y(\Omega)$ est une suite (y_n), $p(X<Y)=\sum_n p(X<Y \text{ et } Y=y_n)=\sum_n p(X<y_n).p(Y=y_n)$. Par exemple si X et Y suivent des lois de Poisson de paramètres respectifs a et b, $p(X<Y)=e^{-(a+b)}\sum_{n=1}^{\infty}\left(\sum_{i=0}^{n-1}\frac{a^i}{i!}\right)\frac{b^n}{n!}$; pour $a=1$ et $b=2$, le calcul numérique donne $p(X<Y)=0{,}606$, $p(X>Y)=0{,}259$ et $p(X=Y)=0{,}135$.
Il en va tout autrement quand X et Y sont diffuses car $p(X=x)=p(Y=y)=0$ pour tous x et y. Le but de cette note est de proposer une méthode de calcul de $p(X<Y)$ qui contourne cette difficulté conceptuelle.

Notations

On appelle F et G les fonctions de répartition respectives de X et Y : comme X et Y sont diffuse, les fonctions positives et croissantes F et G sont continues sur $\mathbb{R}$ et même *uniformément continues* car elles ont des limites finies aux deux infinis. On désignera par f une densité (si elle existe) de X, par g une densité éventuelle de Y.

1) $p(X<Y)=\bigcup_{r\in\mathbb{Q}}(X<r \text{ et } Y>r)$

En effet, deux réels a et b vérifient $a<b$ si et seulement si il existe un rationnel r tel que $a<r<b$.

2) Comme $\mathbb{Q}$ est dénombrable et p $\sigma-$additive, $p(X<Y)$ est donc la borne supérieure des réels

$$\left\{p\left(\bigcup_{r\in\mathcal{F}}(X<r \text{ et } Y>r)\right) : \mathcal{F} \textit{ finie et incluse dans } \mathbb{Q}\right\}.$$

3) Soit $\mathcal{F}$ composée de n rationnels rangés par ordre croissant : $r_1<r_2<...<r_n$.

Nous allons calculer $p\left(\bigcup_{i=1}^n(X<r_i \text{ et } Y>r_i\right)$ en décomposant l'événement $\bigcup_{i=1}^n(X<r_i \text{ et } Y>r_i$ en réunion disjointe. Posons $A_i=(X<r_i)$ et $B_i=(Y>r_i)$: $p(A_i)=F(r_i)$ car X est diffuse, $p(B_i)=1-G(r_i)$ et grâce à l'indépendance

$$p(A_i\cap B_i)=F(r_i)(1-G(r_i)).$$

De plus, $A_1\subset A_2\subset...\subset A_n$ et $B_n\subset B_{n-1}\subset...\subset B_1$, de sorte que

$$A_i=A_1+(A_2-A_1)+...+(A_i-A_{i-1}) \text{ (somme disjointe) et } B_i=\bigcup_{j=i}^n B_j\,.$$

Alors $\bigcup_{i=1}^n A_i\cap B_i$ s'écrit $A_1\cap B_1+(A_1+A_2-A_1)\cap B_2+...+(A_1+A_2-A_1+...A_n-A_{n-1})\cap B_n$,
puis $A_1\cap(B_1\cup...\cup B_n)+(A_2-A_1)\cap(B_2\cup...\cup B_n)+...+(A_n-A_{n-1})\cap B_n$,
soit $A_1\cap B_1+(A_2-A_1)\cap B_2+...+(A_n-A_{n-1})\cap B_n$,
ou encore $A_1\cap(B_1-B_2)+A_2\cap(B_2-B_3)+...+A_{n-1}\cap(B_{n-1}-B_n)+A_n\cap B_n$.
Comme ces deux dernières réunions sont disjointes :

$$p\left(\bigcup_{i=1}^n(X<r_i \text{ et } Y>r_i\right)=F(r_1)(1-G(r_1))+G(r_2)[F(r_2)-F(r_1)]+...+G(r_n)[F(r_n)-F(r_{n-1})]$$
$$=F(r_1)[G(r_2)-G(r_1)]+...+F(r_{n-1})[G(r_n)-G(r_{n-1})]+F(r_n)(1-G(r_n)).$$

Pour obtenir une expression simple de la borne supérieure des $p\left(\bigcup_{i=1}^{n}(X<r_i \text{ et } Y>r_i\right)$ nous supposons désormais que X a une densité f ou que Y a une densité g. Si c'est Y, nous allons minorer assez finement

$$F(r_1)[G(r_2)-G(r_1)]+...+F(r_{n-1})[G(r_n)-G(r_{n-1})]+F(r_n)(1-G(r_n))$$

de façon à ce que le majorant obtenu soit la borne supérieure cherchée.

4) D'abord, $F(r_i)[G(r_{i+1})-G(r_i)]=F(r_i)\int_{r_i}^{r_{i+1}}g$, et comme F est croissante, il vient

$$F(r_i)[G(r_{i+1})-G(r_i)]\le\int_{r_i}^{r_{i+1}}Fg\le F(r_{i+1})[G(r_{i+1})-G(r_i)]$$

d'où

$$0\le\int_{r_i}^{r_{i+1}}F(x)g(x)dx-F(r_i)[G(r_{i+1})-G(r_i)]\le[F(r_{i+1})-F(r_i)][G(r_{i+1})-G(r_i)]$$

et

$$\sum_{i=1}^{n-1}F(r_i)[G(r_{i+1})-G(r_i)]\le\int_{r_1}^{r_n}Fx)g(x)dx\le\int_{r_1}^{r_n}F(x)g(x)dx+\sum_1^{n-1}[F(r_{i+1})-F(r_i)][G(r_{i+1})-G(r_i)].$$

Ensuite, parce que $1-G(r_n)=\int_{r_n}^{+\infty}g$:

$$F(r_1)[G(r_2)-G(r_1)]+...+F(r_{n-1})[G(r_n)-G(r_{n-1})]+F(r_n)(1-G(r_n))\le\int_{r_i}^{r_n}F(x)g(x)dx+\int_{r_n}^{+\infty}Fg=\int_{r_i}^{+\infty}Fg$$

et donc enfin $\boxed{p\left(\bigcup_{r\in\mathcal{F}}(X<r \text{ et } Y>r)\right)\le\int_{-\infty}^{+\infty}F(x)g(x)dx \text{ pour toute } \mathcal{F} \text{ finie dans } \mathbb{Q}.}$

5) $p(X<Y)=\int_{\mathbb{R}}Fg$

On se donne $\varepsilon>0$ et on veut trouver une partie finie $\mathcal{F}$ composée de rationnels telle que

$$p\left(\bigcup_{r\in\mathcal{F}}(X<r \text{ et } Y>r)\right)\ge\int_{-\infty}^{+\infty}F(x)g(x)dx-\varepsilon.$$

a) Il existe un *rationnel* $r<0$ tel que $\int_r^{+\infty}F(x)g(x)dx\ge\int_{\mathbb{R}}Fg-\varepsilon$: il suffit que $\int_{-\infty}^{r}Fg\le\varepsilon$, ce qui est possible.

b) Il existe un *rationnel* $s>0$ tel que $F(s)(1-G(s))\ge\int_s^{+\infty}Fg-\varepsilon$: équivalent à $\int_s^{+\infty}(F(x)-F(s))g(x)dx\le\varepsilon$,

or $\int_s^{+\infty}(F(x)-F(s))g(x)dx\le(1-F(s))$ qui tend vers 0 quand s tend vers +infini.

c) Il existe un *entier* n tel que :

$F(r_1)[G(r_2)-G(r_1)]+...+F(r_{n-1})[G(r_n)-G(r_{n-1})]+F(r_n)(1-G(r_n))\ge\int_{r_i}^{r_n}F(x)g(x)dx-\varepsilon$: on peut trouver

n tel si $|x-y|\le\dfrac{s-r}{n}$ alors $|G(x)-G(y)|\le\varepsilon$ car g est uniformément sur $\mathbb{R}$, d'où

$$\sum_1^{n-1}[F(r_{i+1})-F(r_i)][G(r_{i+1})-G(r_i)]\le\varepsilon\sum_1^{n-1}[F(r_{i+1})-F(r_i)]=\varepsilon[F(r_n)-F(r_1)]\le\varepsilon.$$

Au total, $p\left(\bigcup_{r\in\mathcal{F}}(X<r \text{ et } Y>r)\right)\ge\int_{-\infty}^{+\infty}F(x)g(x)dx-3\varepsilon$, avec $\mathcal{F}$ finie de premier terme r, de dernier terme s et $r_i=r+\dfrac{i(s-r)}{n}$ comme points intermédiaires (i allant de 1 à $n-1$) : $\mathcal{F}$ est bien incluse dans $\mathbb{Q}$.

Conclusion. Soit X et Y des variables aléatoires réelles diffuses et indépendantes :

- si X a une densité f , $p(X < Y) = \int_{\mathbb{R}} p(Y > x) f(x) dx$,

- si Y a une densité g, $p(X < Y) = \int_{\mathbb{R}} p(X < y) g(y) dy$.

Exemple. Supposons que X et Y suivent des lois exponentielles de paramètres respectifs a et b. Alors,

$p(X < y) = \int_0^y a e^{-at} dt = 1 - e^{-ay}$, $g(y) = b e^{-by}$, $p(X < Y) = \dfrac{a}{a+b}$.