III - Max & Min de variables aléatoires

Soient n un entier naturel non nul et $X_1, \ldots, X_n$ des variables aléatoires à valeurs réelles, indépendantes et de mêmes lois. On note F la fonction de répartition commune à ces variables aléatoires.
Les calculs suivants sont classiques et il est important de savoir les reproduire.

I - Maximum de variables aléatoires

Théorème 1 - Maximum

Soit $Z_n = \max\{X_1, \ldots, X_n\}$. En notant F_n la fonction de répartition de Z_n, alors

$$\forall\, x \in \mathbb{R},\ F_n(x) = F(x)^n.$$

Remarque 1

Soit $x \in \mathbb{R}$. L'événement $Z_n \leqslant x$ est réalisé si et seulement si $\max\{X_1, \ldots, X_n\} \leqslant x$. Alors, toutes les variables aléatoires sont inférieures à x et

$$\begin{aligned}
[Z_n \leqslant x] &= [X_1 \leqslant x] \cap \cdots \cap [X_n \leqslant x] \\
\mathbf{P}(X_n \leqslant x) &= \mathbf{P}([X_1 \leqslant x] \cap \cdots \cap [X_n \leqslant x]) \\
&= \mathbf{P}(X_1 \leqslant x) \cdots \mathbf{P}(X_n \leqslant x)\,, \text{ par indépendance} \\
&= \mathbf{P}(X_1 \leqslant x)^n\,, \text{ car mêmes lois} \\
F_n(x) &= F(x)^n.
\end{aligned}$$

Proposition 1

* Si $X_1, \ldots, X_n$ sont des variables aléatoires à densité de densité f, alors Z_n est une variable aléatoire de densité f_n définie par

$$f_n(x) = F_n'(x) = nf(x)F(x)^{n-1}.$$

* Si $X_1, \ldots, X_n$ sont des variables aléatoires discrètes et $X_1(\Omega) = \{x_1, \ldots, x_p\}$, où $x_1 \leqslant \cdots \leqslant x_p$, alors pour tout $i \in [\![1, p]\!]$, en notant $F(x_0) = F_n(x_0) = 0$,

$$\mathbf{P}(Z_n = x_i) = F_n(x_i) - F_n(x_{i-1}) = F(x_i)^n - F(x_{i-1})^n.$$

Exemple 1 - Maximum de deux dés

Soient X_1 et X_2 les résultats indépendants obtenus par le lancer d'un dé équilibré à 6 faces. On pose $Z = \max\{X_1, X_2\}$. Alors, $Z(\Omega) = [\![1, 6]\!]$.
Pour tout $k \leqslant 6$,

$$\begin{aligned}
\mathbf{P}(X_1 \leqslant k) &= \sum_{j=1}^{k} \mathbf{P}(X_1 = j) = \sum_{j=1}^{k} \frac{1}{6} \\
&= \frac{k}{6}.
\end{aligned}$$

De manière analogue, $\mathbf{P}(X_2 = k) = \frac{k}{6}$.

Ainsi,

$$\begin{aligned}
\mathbf{P}(Z \leqslant k) &= \mathbf{P}(\max\{X_1, X_2\} \leqslant k) \\
&= \mathbf{P}([X_1 \leqslant k] \cap [X_2 \leqslant k]) \\
&= \mathbf{P}(X_1 \leqslant k)\,\mathbf{P}(X_2 \leqslant k)\,, \quad \text{par indépendance} \\
&= \left(\frac{k}{6}\right)^2.
\end{aligned}$$

Alors,

$$\begin{aligned}
\mathbf{P}(Z=k) &= \mathbf{P}(Z \leqslant k) - \mathbf{P}(Z \leqslant k-1) \\
&= \left(\frac{k}{6}\right)^2 - \left(\frac{k-1}{6}\right)^2 \\
&= \frac{k^2-(k-1)^2}{36} \\
&= \frac{2k-1}{36}.
\end{aligned}$$

II - Minimum de variables aléatoires

Théorème 2 - Minimum

Soit $Y_n = \min\{X_1, \dots, X_n\}$. En notant G_n la fonction de répartition de Y_n, alors

$$\forall\, x \in \mathbb{R},\ G_n(x) = 1 - (1 - F(x))^n.$$

Remarque 2

Soit $x \in \mathbb{R}$. L'événement $Y_n \leqslant x$ est réalisé si et seulement si $\min\{X_1, \dots, X_n\} \leqslant x$. Ainsi, l'une des variables aléatoires doit être inférieure ou égale à x. On pourrait ainsi écrire :

$$\mathbf{P}(Y_n \leqslant x) = \mathbf{P}([X_1 \leqslant x] \cup \dots \cup [X_n \leqslant x]).$$

Cependant, cette réunion n'est pas disjointe ! On va donc adopter une autre stratégie en utilisant l'événement complémentaire. En effet, $\min\{X_1, \dots, X_n\} > x$ si et seulement si toutes les variables aléatoires sont strictement supérieures à x :

$$\begin{aligned}
\mathbf{P}(Y_n \leqslant x) &= 1 - \mathbf{P}(Y_n > x) \\
&= 1 - \mathbf{P}(\min\{X_1, \dots, X_n\} > x) \\
&= 1 - \mathbf{P}([X_1 > x] \cap \dots \cap [X_n > x]) \\
&= 1 - \mathbf{P}(X_1 > x) \cdots \mathbf{P}(X_n > x), \quad \text{par indépendance} \\
&= 1 - \mathbf{P}(X_1 > x)^n, \quad \text{car mêmes lois} \\
&= 1 - (1 - \mathbf{P}(X_1 \leqslant x))^n, \quad \text{par complémentaire} \\
&= 1 - (1 - F(x))^n.
\end{aligned}$$

Proposition 2

* Si $X_1, \dots, X_n$ sont des variables aléatoires à densité de densité f, alors Y_n est une variable aléatoire de densité g_n définie par

$$g_n(x) = G_n'(x) = nf(x)(1 - F(x))^{n-1}.$$

* Si $X_1, \dots, X_n$ sont des variables aléatoires discrètes et $X_1(\Omega) = \{x_1, \dots, x_p\}$, où $x_1 \leqslant \dots \leqslant x_p$, alors pour tout $i \in [\![1, p]\!]$, en notant $F(x_0) = G_n(x_0) = 0$,

$$\begin{aligned}
\mathbf{P}(Y_n = x_i) &= G_n(x_i) - G_n(x_{i-1}) \\
&= (1 - F(x_{i-1})^n) - (1 - F(x_{i-1})^n).
\end{aligned}$$

Exemple 2 - Minimum de deux lois géométriques

Soient X_1 une variable aléatoire de loi géométrique de paramètre p_1 et X_2 une variable aléatoire de loi géométrique de paramètre p_2. On suppose X_1 et X_2 indépendantes. On pose $Y = \min\{X_1, X_2\}$.

D'après la définition des lois géométriques, pour tout $k \in \mathbb{N}^*$,

$$\begin{aligned}
\mathbf{P}(X_1 \leqslant k) &= \sum_{j=1}^{k} \mathbf{P}(X_1 = j) = \sum_{j=1}^{k} p_1(1-p_1)^{j-1} \\
&= p_1 \sum_{j=0}^{k-1} (1-p_1)^j = p_1 \times \frac{1-(1-p_1)^k}{1-(1-p_1)} \\
&= 1 - (1-p_1)^k.
\end{aligned}$$

De manière analogue, $\mathbf{P}(X_2 \leqslant k) = 1 - (1-p_2)^k$.

Soit $k \in \mathbb{N}^*$. En reprenant l'idée du théorème précédent,

$$\begin{aligned}\mathbf{P}(Y \leqslant k) &= 1 - \mathbf{P}(Y > k) = 1 - \mathbf{P}(\min\{X_1, X_2\} > k) \\ &= 1 - \mathbf{P}([X_1 > k] \cap [X_2 > k]) \\ &= 1 - \mathbf{P}(X_1 > k)\,\mathbf{P}(X_2 > k), \text{ par indépendance} \\ &= 1 - (1-p_1)^k(1-p_2)^k.\end{aligned}$$

Ainsi,

$$\begin{aligned}\mathbf{P}(Y = k) &= \mathbf{P}(Y \leqslant k) - \mathbf{P}(Y \leqslant k-1) \\ &= (1-p_1)^{k-1}(1-p_2)^{k-1} - (1-p_1)^k(1-p_1)^k \\ &= [(1-p_1)(1-p_2)]^{k-1}(1-(1-p_1)(1-p_2)).\end{aligned}$$

Ainsi, Y suit une loi géométrique de paramètre $1-(1-p_1)(1-p_2)$.

III - Généralisons !

Théorème 3 - Statistique d'ordre k

Soit $k \in [\![1, n]\!]$. On ordonne $X_1, \ldots, X_n$ par ordre croissant et on note U_k la valeur du k^{e} réel ainsi obtenu. On pose H_k la fonction de répartition de U_k. Alors,

$$\forall\, x \in \mathbb{R},\ H_k(x) = \sum_{i=k}^{n} \binom{n}{i} F(x)^i (1-F(x))^{n-i}.$$

Remarque 3

* Lorsque $k = n$, on obtient $U_n = \max\{X_1, \ldots, X_n\}$.
* Lorsque $k = 1$, on obtient $U_1 = \min\{X_1, \ldots, X_n\}$.
* Soit $x \in \mathbb{R}$. Dans le cas général, on note $N(x)$ le nombre de variables aléatoires $X_1, \ldots, X_n$ inférieures ou égales à x. $N(x)$ compte le nombre de succès dans un schéma de Bernoulli dont la probabilité de succès vaut $\mathbf{P}(X_1 \leqslant x) = F(x)$. Ainsi, $N(x) \hookrightarrow \mathscr{B}(n, F(x))$.
  $[U_k \leqslant x]$ signifie que la k^{e} plus petite variable aléatoire est inférieure à x, donc que le nombre de variables aléatoires inférieures ou égales à x est supérieur à k :

$$\begin{aligned}H_k(x) &= \mathbf{P}(U_k \leqslant x) \\ &= \mathbf{P}(N(x) \geqslant k) \\ &= \sum_{i=k}^{n} \mathbf{P}(N(x) = i) \\ &= \sum_{i=k}^{n} \binom{n}{i} F(x)^i (1-F(x))^{n-i}.\end{aligned}$$