
–**Texte**–

Records sportifs

1 Questions et modélisation

On se demande si les records sportifs (comme le temps du 100 mètres ou la hauteur atteinte par les sauteurs à la perche) seront toujours battus (quitte à attendre assez longtemps) et si oui, à quel rythme. On souhaite aussi se faire une idée de la valeur de ces records.

On propose, pour étudier cette question, le modèle suivant. On se donne une suite de v.a. $(X_n)_{n \in \mathbb{N}}$ indépendantes et identiquement distribuées de loi μ sur \mathbb{R} , de densité f et de fonction de répartition F .

Pour tout $n \geq 1$, on note R_n le rang relatif de X_n dans l'échantillon X_1, \dots, X_n . Il est donné par $R_n = 1 + \sum_{k=1}^n \mathbf{I}_{\{X_k > X_n\}}$. Le nombre de records jusqu'au temps n est alors noté $Z_n = \sum_{i=1}^n \mathbf{I}_{\{R_i=1\}}$. On définit également les temps de records par récurrence :

$$T_1 = 1 \quad \text{et, pour } j \geq 1 \quad T_{j+1} = \inf \{n > T_j, X_n > X_i, \text{ pour } i < n\}.$$

Notons tout de suite des résultats immédiats.

Proposition 1.1. *Les variables aléatoires $(R_n)_{n \geq 1}$ sont indépendantes et, pour tout $n \in \mathbb{N}^*$, R_n suit la loi uniforme sur $\{1, \dots, n\}$. De plus,*

$$\mathbb{E}(Z_n) = \sum_{k=1}^n \frac{1}{k} = \log n + \gamma + o(1),$$

où γ désigne la constante d'Euler.

2 Le nombre de records est logarithmique

On souhaite à présent établir que le nombre de records jusqu'à une date n est asymptotiquement logarithmique. Plus précisément, la proposition suivante donne le comportement asymptotique de Z_n .

Proposition 2.1. *La suite $(Z_n)_n$ converge presque sûrement vers ∞ à vitesse logarithmique : plus précisément,*

$$\frac{Z_n}{\log n} \xrightarrow[n \rightarrow \infty]{} 1 \quad p.s.$$

De plus, on peut donner une vitesse pour cette convergence :

$$\frac{Z_n - \log n}{\sqrt{\log n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Nous suggérons ici deux pistes différentes pour démontrer ce résultat, l'une directe, l'autre basée sur la théorie des martingales, la première consistant essentiellement à redémontrer les résultats de convergence des martingales dans le cas particulier présent.

Remarque 2.2. Ce théorème peut permettre d'invalider le modèle.

Proposition 2.3. Soit $(Y_i)_{i \geq 1}$ une suite de v.a. indépendantes possédant un moment d'ordre 2. Soit $(b_n)_n$ une suite croissante de réels tendant vers ∞ tels que

$$\sum_{j=1}^n \frac{\mathbb{V}(R_j)}{b_n^2} < \infty.$$

Alors, si $S_n = Y_1 + \dots + Y_n$,

$$\frac{S_n - \mathbb{E}(S_n)}{b_n} \xrightarrow[n \rightarrow \infty]{} 0 \quad p.s.$$

La preuve de ce résultat s'appuie sur les deux lemmes suivants.

Lemme 2.4 (Toeplitz). Soit (a_n) une suite de réels positifs et $b_n = a_1 + \dots + a_n$ tels que $b_n > 0$ pour $n \geq 1$ et b_n tend vers ∞ . Soit $(x_n)_n$ une suite réelle qui converge vers x . Alors

$$\frac{1}{b_n} \sum_{j=1}^n a_j x_j \xrightarrow[n \rightarrow \infty]{} x.$$

Ce lemme est une généralisation du lemme de Cesaro (prendre $a_n = 1$ pour tout n).

Preuve du lemme 2.4. Soit $\varepsilon > 0$. Il existe n_0 tel que pour tout $n \geq n_0$, $|x_n - x| \leq \varepsilon$. Soit $n_1 > n_0$ tel que

$$\frac{1}{b_{n_1}} \sum_{j=1}^{n_0} a_j |x_j - x| < \varepsilon.$$

Alors, pour tout $n > n_1$,

$$\begin{aligned} \left| \frac{1}{b_n} \sum_{j=1}^n a_j x_j - x \right| &\leq \frac{1}{b_n} \sum_{j=1}^n a_j |x_j - x| = \frac{1}{b_n} \sum_{j=1}^{n_0} a_j |x_j - x| + \frac{1}{b_n} \sum_{j=n_0+1}^n a_j |x_j - x| \\ &\leq \frac{1}{b_{n_1}} \sum_{j=1}^{n_0} a_j |x_j - x| + \frac{1}{b_n} \sum_{j=n_0+1}^n a_j |x_j - x| \\ &\leq \varepsilon + \frac{b_n - b_{n_0}}{b_n} \varepsilon \leq 2\varepsilon. \end{aligned}$$

□

Lemme 2.5 (Kronecker). *Soit (b_n) une suite croissante de réels strictement positifs qui tend vers ∞ et (x_n) une suite de réels tels que $\sum x_n$ converge. Alors*

$$\frac{1}{b_n} \sum_{j=1}^n b_j x_j \xrightarrow{n \rightarrow \infty} 0.$$

Démonstration. Soit $b_0 = 0$, $S_0 = 0$ et $S_n = x_1 + \dots + x_n$ pour $n \geq 1$. Alors,

$$\sum_{j=1}^n b_j x_j = \sum_{j=1}^n b_j (S_j - S_{j-1}) = b_n S_n - b_0 S_0 - \sum_{j=1}^n S_{j-1} (b_j - b_{j-1}).$$

Ainsi,

$$\frac{1}{b_n} \sum_{j=1}^n b_j x_j = S_n - \frac{1}{b_n} \sum_{j=1}^n S_{j-1} (b_j - b_{j-1}) \xrightarrow{n \rightarrow \infty} 0,$$

puisque, si $S_n \rightarrow s$, alors d'après le lemme de Toeplitz,

$$\frac{1}{b_n} \sum_{j=1}^n S_{j-1} (b_j - b_{j-1}) \xrightarrow{n \rightarrow \infty} s.$$

□

Preuve de la proposition 2.3. On réécrit la quantité dont on veut calculer la limite de façon à utiliser le lemme de Kronecker :

$$\frac{S_n - \mathbb{E}(S_n)}{b_n} = \frac{1}{b_n} \sum_{j=1}^n b_j \frac{Y_j - \mathbb{E}(Y_j)}{b_j}.$$

Il suffit de montrer que la série de terme général

$$Z_k = (Y_k - \mathbb{E}(Y_k))/b_k$$

converge presque sûrement. Par hypothèse,

$$\mathbb{E}(Z_k^2) = \mathbb{V}(Z_k) = \frac{\mathbb{V}(Y_k)}{b_k^2} < \infty$$

est le terme général d'une série convergente. La preuve est achevée en utilisant le résultat suivant.

Théorème 2.6. *Soit (Z_n) une suite de variables aléatoires indépendantes centrées telle que*

$$\sum_{n=1}^{\infty} \mathbb{E}(Z_k^2) < \infty,$$

alors la série $\sum Z_n$ converge avec probabilité 1.

Ceci achève la preuve de la proposition 2.3. □

Pour en déduire la première partie de la proposition 2.1, il suffit de remarquer que $\mathbb{E}(Z_n)$ est équivalent à $\log n$ et que le choix $b_n = \log n$ est possible dans la proposition 2.3.

La deuxième partie du théorème découle du théorème limite central de Lindeberg-Lévy appliqué aux variables $\mathbf{I}_{\{R_n=1\}} - 1/n$.

3 Quels sommets atteindront les records ?

On se pose à présent la question de la valeur des records à une date n . Ceci dépend bien sûr fortement de la loi des variables $(X_i)_{i \in \mathbb{N}}$. Après n tentatives, le record courant est égal à $X_{(n)} = \max(X_i, 1 \leq i \leq n)$.

3.1 Loi uniforme

Proposition 3.1. *Si la loi μ des variables aléatoires $(X_i)_i$ est la loi uniforme sur $[0, \theta]$ alors, le comportement de la suite $(X_{(n)})_n$ est le suivant :*

$$X_{(n)} \xrightarrow[n \rightarrow \infty]{p.s.} \theta \quad \text{et} \quad \frac{n}{\theta}(X_{(n)} - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Y,$$

où $-Y$ suit une loi exponentielle de paramètre 1.

Remarque 3.2. Si θ est inconnu, $X_{(n)}$ est son estimateur du maximum de vraisemblance construit à partir de l'échantillon X_1, \dots, X_n , c'est-à-dire que la densité de cet échantillon prise comme fonction de θ est maximale au point $X_{(n)}$. Notons que cet estimateur est biaisé.

3.2 Loi exponentielle

Proposition 3.3. *Si la loi μ des variables aléatoires $(X_i)_i$ est la loi exponentielle de paramètre $\lambda > 0$ alors, le comportement de la suite $(X_{(n)})_n$ est le suivant :*

$$\frac{X_{(n)}}{\log n} \xrightarrow[n \rightarrow \infty]{p.s.} \frac{1}{\lambda} \quad \text{et} \quad X_{(n)} - \frac{\log n}{\lambda} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Y,$$

où Y désigne une variable aléatoire de fonction de répartition

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

4 Records plus fréquents

Dans la pratique, on constate dans certains sports ou pour des enregistrements de grandeurs climatiques un nombre de records bien plus importants que celui prédit par le modèle précédent (Z_n de l'ordre de $\log n$). On souhaite donc proposer un modèle qui présente un nombre de records de l'ordre de n .

Pour cela, on suppose que les variables aléatoires $(X_i)_{i \in \mathbb{N}}$ sont indépendantes de fonctions de répartition respectives $(F_i)_{i \in \mathbb{N}}$. Pour simplifier le paramétrage du modèle, on choisit F_i de la forme suivante

$$F_i(x) = F(x)^{\gamma^i}$$

où F est une fonction de répartition continue et γ est un réel strictement supérieur à 1.

Proposition 4.1. *Pour tout $n \in \mathbb{N}^*$,*

$$\mathbb{P}(R_n = 1) = \frac{\gamma^n}{\gamma + \dots + \gamma^n} = \frac{1 - 1/\gamma}{1 - 1/\gamma^n} \xrightarrow[n \rightarrow \infty]{} 1 - 1/\gamma.$$

5 Suggestions

1. On pourra commenter le modèle et discuter d'amménagements éventuels.
2. On pourra illustrer par la simulation la proposition 2.1.
3. On pourra commenter la remarque 2.2.
4. On pourra démontrer la proposition 2.3.
5. On pourra démontrer les résultats de la section 3.1.
6. On pourra démontrer les résultats de la section 3.2 (un peu plus délicat que la suggestion précédente).