

(Public-A02)

Résumé : Nous nous intéressons au comportement de la vitesse de propagation d'un polluant dans un écoulement d'eau en sous-sol rocheux stratifié. On considère un modèle plan qui est constitué de strates horizontales homogènes. Un modèle stochastique est proposé pour deux mécanismes qui définissent l'évolution du polluant : les collisions avec les particules du fluide d'une part, le transport passif dû à l'écoulement d'autre part.

Mots clefs : Marche aléatoire, théorèmes limites, variables non indépendantes.

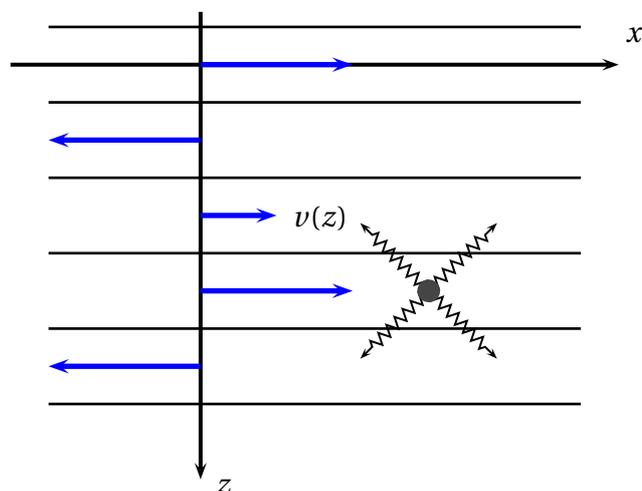
- *Il est rappelé que le jury n'exige pas une compréhension exhaustive du texte. La présentation, bien que totalement libre, doit être organisée et le jury apprécie qu'un plan soit annoncé en préliminaire. L'exposé doit être construit en évitant la paraphrase et mettant en lumière les connaissances, à partir des éléments du texte. Il doit contenir des illustrations informatiques réalisées sur ordinateur, ou, à défaut, des propositions de telles illustrations. Des pistes de réflexion, indicatives et largement indépendantes les unes des autres, vous sont proposées en fin de texte.*

1. Un polluant dans un écoulement hétérogène

Nous considérons un écoulement d'eau dans un sous-sol stratifié. Dans cet écoulement, nous nous intéressons à l'évolution d'une particule de polluant. Nous discrétisons d'abord l'espace : notre écoulement est caractérisé par un "champ de vitesses" $V(x, z) \in \mathbb{R}^2$ en chaque point $(x, z) \in \mathbb{R} \times \mathbb{Z}$. Cela veut dire qu'une particule de polluant transportée *passivement* dans cet écoulement a une position $r(t) = (x(t), z(t))$ qui satisfait l'équation différentielle

$$(1) \quad \frac{dr(t)}{dt} = V(r(t)).$$

L'axe horizontal est celui des x et l'axe vertical, orienté vers le bas, est celui des z , représentant la profondeur. On suppose que le champ de vitesses est parallèle à l'axe horizontal et de la forme $V(x, z) = (v(z), 0)$ (voir figure). Comme les détails du sous-sol rocheux nous sont inconnus, nous modélisons les valeurs de la fonction $z \mapsto v(z)$ par une suite de variables aléatoires $\{\eta(z), z \in \mathbb{Z}\}$, que nous supposons indépendantes et centrées.



Une simplification supplémentaire résulte de la discrétisation du temps. Ainsi, la position du polluant au temps $n > 0$, est représentée par un vecteur $r_n = (x_n, z_n)$. Le polluant est transporté passivement si

$$(2) \quad r_n = r_{n-1} + V(r_{n-1}) = (x_{n-1} + \eta(z_{n-1}), z_{n-1}) \quad (\text{et } r_0 = (0, 0)).$$

On voudrait aussi tenir compte des interactions entre le polluant et les particules du fluide : nous modélisons les collisions comme pouvant déplacer le polluant uniformément au hasard dans une des quatre directions possibles. Soient $\{\alpha_k, k \in \mathbb{N}\}$ et $\{\beta_k, k \in \mathbb{N}\}$ deux suites de variables aléatoires indépendantes identiquement distribuées (i.i.d.). On suppose que la suite $\{\alpha_k, k \in \mathbb{N}\}$ est indépendante de la suite $\{\beta_k, k \in \mathbb{N}\}$, que ces deux suites sont indépendantes de la suite $\{\eta(z), z \in \mathbb{Z}\}$, et que pour tout $k \in \mathbb{N}$,

$$(3) \quad P(\alpha_k = 1) = P(\alpha_k = -1) = P(\beta_k = 1) = P(\beta_k = -1) = \frac{1}{2},$$

alors la position du polluant à l'instant n , disons R_n , s'obtient à partir de R_{n-1} par les relations de récurrence suivantes, enchaînant collisions puis transport passif :

$$(4) \quad \begin{aligned} \tilde{R}_{n-1} &= R_{n-1} + (\alpha_n, \beta_n), \\ R_n &= \tilde{R}_{n-1} + V(\tilde{R}_{n-1}), \quad \forall n > 0, \quad \text{avec } R_0 = (0, 0). \end{aligned}$$

En écrivant ce qu'il advient pour chaque coordonnée de $R_n := (X_n, Z_n)$, $n \geq 1$, on obtient

$$(5) \quad X_n = \sum_{k=1}^n \alpha_k + \sum_{k=1}^n \eta(Z_k), \quad \text{et} \quad Z_n = \sum_{k=1}^n \beta_k.$$

Pour alléger les notations, on posera

$$(6) \quad Y_n := \eta(Z_1) + \dots + \eta(Z_n).$$

Ainsi, la position horizontale comporte deux contributions :

- $\alpha_1 + \dots + \alpha_n$, qui est une somme de variables aléatoires i.i.d. centrées.
- Y_n , qui est une somme de variables centrées mais pas indépendantes.

La position verticale Z_n est une marche aléatoire simple. Dans la suite, les symboles E et P désigneront toujours des espérances et probabilités par rapport aux deux aléas que sont les vitesses η et les collisions (α, β) .

2. Sur-diffusion

Nous allons supposer dans cette section que $\{\eta(z), z \in \mathbb{Z}\}$ est une suite de variables i.i.d. indépendantes de $\{Z_n, n \in \mathbb{N}\}$, telle que

$$(7) \quad \forall z \in \mathbb{Z}, \quad P(\eta(z) = 1) = P(\eta(z) = -1) = \frac{1}{2}.$$

La suite $\{Z_n, n \in \mathbb{N}\}$ vérifie le théorème limite central, ce qui implique qu'en un temps de l'ordre de n , Z_n est de l'ordre de \sqrt{n} . Par contre Y_n a un autre comportement.

Proposition 1. *Il existe une constante $\kappa > 0$ telle que*

$$(8) \quad E[Y_n^2] \geq \kappa n^{3/2}.$$

Ébauche de preuve. On note que

$$(9) \quad \begin{aligned} E[Y_n^2] &= \sum_{k=1}^n E[\eta^2(Z_k)] + 2 \sum_{1 \leq k < k' \leq n} E[\eta(Z_k)\eta(Z_{k'})] \\ &= n + 2 \sum_{1 \leq k < k' \leq n} P(Z_k = Z_{k'}), \end{aligned}$$

et, pour $k' > k$, on obtient par un raisonnement combinatoire

$$(10) \quad P(Z_k = Z_{k'}) = P(\beta_1 + \dots + \beta_{k'-k} = 0) = \begin{cases} 0 & \text{si } k' - k \text{ est impair,} \\ 2^{-2m} \binom{2m}{m} \sim \frac{1}{\sqrt{\pi m}} & \text{si } k' - k = 2m. \end{cases}$$

La notation $u_n \sim v_n$ signifie que $u_n/v_n \rightarrow 1$ lorsque $n \rightarrow \infty$. Le résultat (8) découle de (9) et (10).

3. Grandes déviations

Supposons maintenant les variables $\{\eta(z), z \in \mathbb{Z}\}$ de loi gaussienne centrée réduite $\mathcal{N}(0, 1)$ et indépendantes de $\{Z_n, n \in \mathbb{N}\}$.

On voudrait montrer que la probabilité que Y_n soit *très grand* est bien plus élevée que pour une somme de variables indépendantes. Par exemple, on peut voir qu'il existe une constante strictement positive I telle que

$$(11) \quad P\left(\sum_{i=1}^n \beta_i \geq n\right) \leq \exp(-nI).$$

Une telle majoration n'est pas vérifiée pour la suite $\{Y_n, n \in \mathbb{N}\}$, en vertu de la proposition suivante.

Proposition 2. *Il existe des constantes $C, \gamma > 0$ telles que*

$$(12) \quad P(Y_n > n) \geq C \exp(-\gamma n^{4/7}).$$

Ébauche de preuve. Récrivons Y_n , de manière à *découpler* les deux aléas :

$$(13) \quad Y_n = \eta(Z_1) + \dots + \eta(Z_n) = \sum_{z \in \mathbb{Z}} \eta(z) l_n(z), \quad \text{où } l_n(z) = \mathbb{1}_{\{Z_1=z\}} + \dots + \mathbb{1}_{\{Z_n=z\}}.$$

Ainsi $l_n(z)$ représente le nombre de visites du point z jusqu'au temps n . Donc, conditionnellement à $\sigma(Z_1, \dots, Z_n)$,

$$(14) \quad \sum_{z \in \mathbb{Z}} \eta(z) l_n(z) \quad \text{a la même loi que} \quad \sqrt{\sum_{z \in \mathbb{Z}} l_n^2(z)} Z,$$

où $Z \sim \mathcal{N}(0, 1)$ est indépendant de $\{Z_n, n \in \mathbb{N}\}$. Alors, en notant que

$$(15) \quad \sum_{z \in \mathbb{Z}} l_n^2(z) \geq l_n^2(0),$$

on peut montrer, pour tout entier $k \in \{1, \dots, n\}$, les inégalités

$$(16) \quad P\left(\sqrt{\sum_{z \in \mathbb{Z}} l_n^2(z)} Z > n\right) \geq P(l_n(0) Z > n) \geq P\left(Z > \frac{n}{k}\right) P(l_n(0) \geq k).$$

On peut montrer par une intégration par parties que $\int_y^\infty e^{-z^2/2} dz \sim \frac{1}{y} e^{-y^2/2}$ pour $y \rightarrow +\infty$. On déduit alors de (16) l'existence d'une constante $0 < C_0 < 1$ telle que

$$(17) \quad P(Y_n > n) \geq C_0 \frac{\exp(-\frac{n^2}{2k^2})}{\sqrt{2\pi}(n/k)} P(l_n(0) \geq k).$$

Il nous reste donc à évaluer $P(l_n(0) \geq k)$. Pour visiter au moins k fois l'origine avant l'instant n , il suffit que les retours en 0 soient espacés de moins de n/k unités. Ainsi, en introduisant le temps d'arrêt $T = \inf\{n \geq 1 : Z_n = 0\}$, on peut écrire

$$(18) \quad P(l_n(0) \geq k) \geq \left[P\left(T \leq \frac{n}{k}\right)\right]^k.$$

On admettra qu'il existe une constante c positive telle que, pour $t \geq 1$,

$$(19) \quad P(T > t) \leq c/\sqrt{t}.$$

L'étude de la fonction $u \mapsto \frac{\ln(1-u)}{-u}$ sur $]0, 1[$ permet de montrer qu'il existe $D > 0$ telle que

$$(20) \quad P(l_n(0) \geq k) \geq \exp\left(-\frac{Dk\sqrt{k}}{\sqrt{n}}\right).$$

Finalement, on obtient

$$(21) \quad P(Y_n > n) \geq C_0 \frac{\exp\left(-\frac{n^2}{2k^2} - \frac{Dk\sqrt{k}}{\sqrt{n}}\right)}{\sqrt{2\pi}(n/k)}.$$

En fixant k à la partie entière de $n^{5/7}$, on en déduit qu'il existe une constante $E > 0$ telle que

$$(22) \quad \frac{k\sqrt{k}}{\sqrt{n}} \leq n^{4/7} \quad \text{et} \quad \frac{n}{k} \leq E n^{2/7},$$

d'où se déduit le résultat annoncé.

Suggestions et pistes de réflexion

- ▶ *Les pistes de réflexion suivantes ne sont qu'indicatives et il n'est pas obligatoire de les suivre. Vous pouvez choisir d'étudier, ou non, certains des points proposés, de façon plus ou moins approfondie, mais aussi toute autre question à votre initiative. Vos investigations comporteront une partie traitée sur ordinateur et, si possible, des représentations graphiques de vos résultats. À défaut, si vos illustrations informatiques n'ont pas abouti, il est conseillé d'expliquer ce que vous auriez souhaité mettre en œuvre.*
- *Développements mathématiques.*
 - Les variables $\{\eta(Z_k), k \geq 1\}$ ne sont pas toutes indépendantes. On peut essayer de préciser ce point et voir aussi si elles forment une chaîne de Markov.
 - En utilisant l'approximation de Stirling, on pourra démontrer l'équation (10) et établir la proposition 1.
 - Vous pouvez démontrer les inégalités (11) et (16).
 - On pourra rendre rigoureuse la démonstration de l'inégalité (18).
- *Modélisation.*
 - L'hypothèse de centrage du champ de vitesses vous paraît-elle réaliste?
 - Que se passerait-il si les $\{\eta(z), z \in \mathbb{Z}\}$ n'étaient pas centrées?
 - On pourra formuler un modèle tri-dimensionnel raisonnable d'écoulement dans un sous-sol stratifié.
- *Étude numérique.*
 - On pourra simuler une réalisation de Y_n .
 - Il est également possible de vérifier par simulation que $E[Y_n^2]$ a bien un comportement en $n^{3/2}$.
 - En simulant diverses trajectoires de Z_n , on pourra obtenir (19).