Detecting alterations in genomic profiles : the infinite population case.

MP. Etienne[1],
joint work with L. Decreusefond[2], G. Lang[3], S. Robin[4] and P. Vallois[5]

Rochebrune 2024

([1]) Institut Agro - IRMAR
Telecom ParisTech
([2]) Telecom ParisTech
([3]) MIA Paris Saclay
([4]) LPSM
([5]) IECL

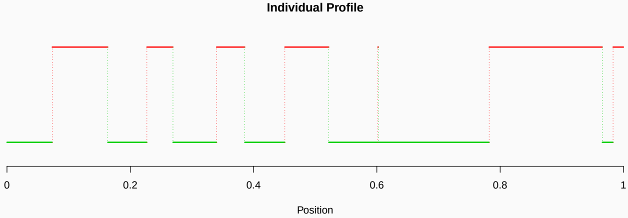
Modelling individual genomic profile, large L

Pure Jump Continuous-time Markov model:

Each profile X_i is a continuous time Markov process $X_i = \{X_i(t)\}_{0 \leq t \leq L}$:

States:	$0 = \text{normal}$	$1 = \text{alteration}$
Transition rates:	$\lambda : 0 \to 1$	$\mu : 1 \to 0.$

Consequence.
Mean length of normal regions = $1/\lambda$ and altered regions = $1/\mu$.



Cumulative Cohort profiles

We are interested in the alterations shared by a large proportion of patients.

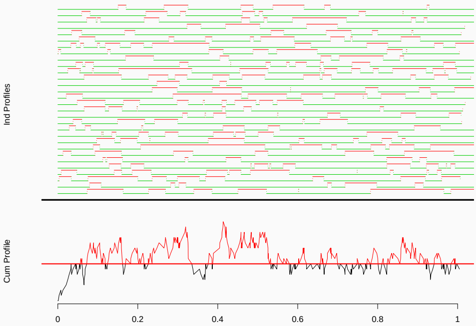
Define the cumulative cohort profile as $S_n = \sum_{i=1}^{n} X_i$.

Focus on the length $E_1, E_2, \ldots.$ of the excursions above a of S_n on $[0,1]$.

Goal : characterizing the distribution of

$$E_*^{a,S_n}(1) = \max_i E_i(1)$$

$$\mathbb{P}\left(E_*^{a,S_n} > l\right) = ??$$



A True Sea Snake

Let's Dare to say it in French *Un véritable serpent de mer*

Rats do not leave the ship

Theoretical behaviour with large number of patients

Cumulative process S_n

The process $\{S_n(t), t \in [0,1]\}$ counts the number of altered profiles at time t. Assuming $\lambda = \mu$, the jumps $\mathbf{U} = (U_1^{(n)}, U_2^{(n)}, \ldots)$ occur at exponential times λn.

Cumulative process S_n

The process $\{S_n(t), t \in [0,1]\}$ counts the number of altered profiles at time t. Assuming $\lambda = \mu$, the jumps $\mathbf{U} = (U_1^{(n)}, U_2^{(n)}, \ldots)$ occur at exponential times λn.
It is a Markovian birth and death process with states $\{0, 1 \ldots, n\}$ and transition rates

$$i \to i+1 \quad \text{('birth' of an alteration) :} \qquad \lambda_i = (n-i) \times \lambda$$

$$i \to i-1 \quad \text{(death = back to normal) :} \qquad \mu_i = i \times \lambda$$

Cumulative process S_n

The process $\{S_n(t), t \in [0,1]\}$ counts the number of altered profiles at time t. Assuming $\lambda = \mu$, the jumps $\mathbf{U} = (U_1^{(n)}, U_2^{(n)}, \ldots)$ occur at exponential times λn.
In the stationary case,

$$X_i(0) \sim \mathcal{B}\left(\frac{1}{2}\right), \quad i = 1, \ldots n,$$

Define $\tau = 2\lambda$, $p = 0.5$, we have:

$$\mathbb{E}\left[S_n(t)\right] = \frac{n}{2}, \quad \mathbb{V}\left[S_n(t)\right] = \frac{n}{4}, \quad \mathbb{C}\mathrm{ov}\left[S_n(s), S_n(s+t)\right] = \frac{n}{4} e^{-\tau t}$$

Consider

$$Z^{(n)}(t) = \frac{S_n(t) - n/2}{\sqrt{n/4}},$$

The perfect suspect $\mathbf{Z} = (Z(t), 0 \le t \le 1)$ solution to

$$dZ(t) = -\tau Z(t)dt + \sqrt{2\tau}dW(t), \quad Z(0) \sim \mathcal{N}(0,1).$$

Cumulative process $S^{(n)}$

Sketch of proof :

- $(Z^{(n)}(t_1), \dots, Z^{(n)}(t_k))$ converges to a Gaussian vector,

$$\sum_{j=1}^{k} \alpha_j Z(t_j) = \sum_{i=1}^{n} \sum_{j=1}^{k} \alpha_j \frac{X_i(t_j) - 0.5}{\sqrt{n/4}} \quad \forall (\alpha_1, \dots, \alpha_C) \in \mathbb{R}^k$$

- The covariance function is

$$\rho(s, s+t) = \mathbb{C}\mathrm{ov}(Z^{(n)}(s), Z^{(n)}(s+t)) = \rho(t) = e^{-\tau t}$$

- The unique stationnary Gaussian process with covariance function $\rho(t) = e^{-\tau t}$ is the Ornstein Uhlenbeck process.
- Tightness

Cumulative process $S^{(n)}$

Sketch of proof :

- $(Z^{(n)}(t_1), \ldots, Z^{(n)}(t_k))$ converges to a Gaussian vector,

$$\sum_{j=1}^{k} \alpha_j Z(t_j) = \sum_{i=1}^{n} \sum_{j=1}^{k} \alpha_j \frac{X_i(t_j) - 0.5}{\sqrt{n/4}} \quad \forall (\alpha_1, \ldots, \alpha_C) \in \mathbb{R}^k$$

- The covariance function is

$$\rho(s, s+t) = \mathbb{C}\mathrm{ov}(Z^{(n)}(s), Z^{(n)}(s+t)) = \rho(t) = e^{-\tau t}$$

- The unique stationnary Gaussian process with covariance function $\rho(t) = e^{-\tau t}$ is the Ornstein Uhlenbeck process.
- Tightness

From a practical point of view, **the rate would be useful** !

Step 0 : Understanding Kubilius

5 years of scientific wandering and convergence rate as rewarding as $N^{1/23}$, which might require large cancer cohort.

Finally, getting inspired by Kęstutis Kubilius[a] (6 citations in Google Scholar).

Theorem ([Kub94])

Let $(\xi_k^{(n)}, \mathcal{F}_k^{(n)})$ be a sequence of square-integrable martingale differences a.s bounded by $n^{-1/2}M$. Denote

$$W^n(t) = \sum_i^n \xi_i^{(n)} + \frac{tV_n^{(n)^2} - V_k^{(n)^2}}{V_{k+1}^{(n)^2} - V_k^{(n)^2}} \xi_i^{(n)}, \quad \text{if } \frac{V_k^{(n)^2}}{V_n^{(n)^2}} < t \leq \frac{V_{k+1}^{(n)^2}}{V_n^{(n)^2}}$$

where $V_k^{(n)^2} = \sum_{i=1}^k \mathbb{E}\left\{(\xi_i^{(n)})^2 | \mathcal{F}_{i-1}^{(n)}\right\}$. There exists a constant C such that

$$\pi(W^{(n)}, W) < C \ln n \left\{ n^{-1/4} + inf_{0<\varepsilon<1} \left(\varepsilon + \mathbb{P}(|V_n^{(n)^2} - 1| > \varepsilon^2) \right) \right\}.$$

π stands for the Prokhorov distance.

[a] K Kubilius. "Rate of convergence in the invariance principle for martingale difference arrays". In: Lithuanian Mathematical Journal 34.4 (1994), pp. 383–392.

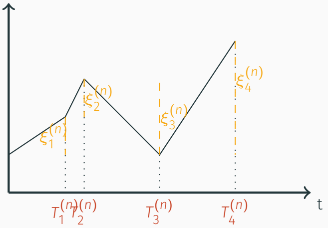
- $\xi_k^{(n)}$ square-integrable martingale differences:

$$\mathbb{E}\left[\sum_{k=1}^{j} \xi_k^{(n)} | \mathcal{F}_{j-1}^{(n)}\right] = \mathbb{E}\left[\xi_j^{(n)} | \mathcal{F}_{j-1}^{(n)}\right] + \sum_{k=1}^{j-1} \xi_k^{(n)}$$

$$\mathbb{E}\left[\xi_j^{(n)} | \mathcal{F}_{j-1}^{(n)}\right] = 0$$

- Handling time

$$T_k^{(n)} = \frac{V_k^{(n)2}}{V_n^{(n)2}}, \quad V_k^{(n)2} = \sum_{i=1}^{k} \mathbb{E}\left[(\xi_i^{(n)})^2 | \mathcal{F}_{i-1}^{(n)}\right]$$



Step 0 : Understanding Kubilius

- $\xi_k^{(n)}$ square-integrable martingale differences:

$$\mathbb{E}\left[\sum_{k=1}^{j} \xi_k^{(n)} | \mathcal{F}_{j-1}^{(n)}\right] = \mathbb{E}\left[\xi_j^{(n)} | \mathcal{F}_{j-1}^{(n)}\right] + \sum_{k=1}^{j-1} \xi_k^{(n)}$$

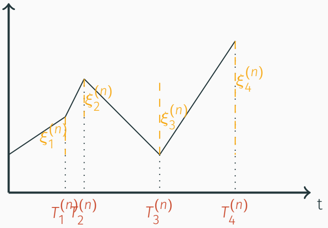
$$\mathbb{E}\left[\xi_j^{(n)} | \mathcal{F}_{j-1}^{(n)}\right] = 0$$

- The Prokhorov distance π, U et V deux variables à valeurs dans le même espace,

$$\pi(U,V) \leq \varepsilon \text{ iff } \mathbb{P}\{U \in A\} \leq \mathbb{P}\{V \in A^{\varepsilon}\} + \varepsilon$$

- Handling time

$$T_k^{(n)} = \frac{{V_k^{(n)}}^2}{{V_n^{(n)}}^2}, \quad {V_k^{(n)}}^2 = \sum_{i=1}^{k} \mathbb{E}\left[(\xi_i^{(n)})^2 | \mathcal{F}_{i-1}^{(n)}\right]$$



Action Plan

- $Z_k^{(n)} = S^{(n)}(U_k^{(n)})$ for $k = 1, \dots m$ (m the number of jumps)

Action Plan

- $Z_k^{(n)} = S^{(n)}(U_k^{(n)})$ for $k = 1, \dots m$ (m the number of jumps)
- $Z_{k+1}^{(n)} = Z_k^{(n)} + \frac{2}{\sqrt{n}} v_{k+1}, \quad \mathbb{P}(v_{k+1} = -1) = 1 - \mathbb{P}(v_{k+1} = 1) = \frac{Z_k^{(n)}}{2\sqrt{n}} + \frac{1}{2},$

Action Plan

- $Z_k^{(n)} = S^{(n)}(U_k^{(n)})$ for $k = 1, \ldots m$ (m the number of jumps)
- $Z_{k+1}^{(n)} = Z_k^{(n)} + \frac{2}{\sqrt{n}} \upsilon_{k+1}, \quad \mathbb{P}(\upsilon_{k+1} = -1) = 1 - \mathbb{P}(\upsilon_{k+1} = 1) = \frac{Z_k^{(n)}}{2\sqrt{n}} + \frac{1}{2},$
- $\mathbb{E}\left[Z_{k+1}^{(n)} | \mathcal{F}_k^{(n)}\right] = \left(1 - \frac{2}{n}\right) Z_k^{(n)}$

Action Plan

- $Z_k^{(n)} = S^{(n)}(U_k^{(n)})$ for $k = 1, \dots m$ (m the number of jumps)
- $Z_{k+1}^{(n)} = Z_k^{(n)} + \frac{2}{\sqrt{n}} v_{k+1}, \quad \mathbb{P}(v_{k+1} = -1) = 1 - \mathbb{P}(v_{k+1} = 1) = \frac{Z_k^{(n)}}{2\sqrt{n}} + \frac{1}{2},$
- $\mathbb{E}\left[Z_{k+1}^{(n)} | \mathcal{F}_k^{(n)}\right] = \left(1 - \dfrac{2}{n}\right) Z_k^{(n)}$
- $\xi_{k+1}^{(n,m)} = \frac{1}{2}\sqrt{\frac{n}{m}}\left(Z_{k+1}^{(n)} - \left(1 - \frac{2}{n}\right) Z_k^{(n)}\right), \quad 0 \leq k \leq m-1$
- Prokhorov:

$$\begin{aligned}\mathbb{P}(\omega_n \in A) =& \mathbb{P}(\omega_n \in A, \|\omega_n - \omega\|_\infty \leq \varepsilon) + \mathbb{P}(\omega_n \in A, \|\omega_n - \omega\|_\infty > \varepsilon) \\ \leq& \mathbb{P}(\omega \in A^\varepsilon) + \mathbb{P}(\|\omega_n - \omega\|_\infty > \varepsilon)\end{aligned}$$

Step 1: Some random process converges to a Brownian motion

Theorem
Let n and m be two integers satisfying

$$\lambda_1 m \leq n \leq \lambda_2 m, \quad \text{with } 0 < \lambda_1 \leq \lambda_2 \leq n^{1/4},$$

$$W_k^{(n,m)} = \sum_{i=1}^{k} \xi_i^{(n,m)}, \quad 1 \leq k \leq m, \quad W_0^{(n,m)} = 0.$$

Then, there exists a constant $C_1(\lambda_2)$ such that

$$\pi\left(\boldsymbol{W}^{(n,m)}, \mathrm{W}\right) \leq C_1(\lambda_2) n^{-1/4} \ln(n).$$

Step 2: Getting closer from the original process

Let's consider $\boldsymbol{Y}^{(\boldsymbol{n},\boldsymbol{m})} = (Y^{(n,m)}(t), 0 \leq t < 1)$:

$$Y^{(n,m)}(t) = Z_k^{(n)} + \frac{t - t_k^{(n,m)}}{t_{k+1}^{(n,m)} - t_k^{(n,m)}} \left(Z_{k+1}^{(n)} - Z_k^{(n)} \right), \quad t_k^{(n,m)} \leq t < t_{k+1}^{(n,m)}.$$

Step 2: Getting closer from the original process

Let's consider $\boldsymbol{Y^{(n,m)}} = (Y^{(n,m)}(t), 0 \leq t < 1)$:

$$\boldsymbol{Y^{(n,m)}}(t) = Z_k^{(n)} + \frac{t - t_k^{(n,m)}}{t_{k+1}^{(n,m)} - t_k^{(n,m)}} \left(Z_{k+1}^{(n)} - Z_k^{(n)} \right), \quad t_k^{(n,m)} \leq t < t_{k+1}^{(n,m)}.$$

The integral representation of a Ornstein uhlenbeck process

$$Z(t) = Z_0 e^{-\tau t} + \sqrt{2\tau} W(t) - \sqrt{2}\tau^{3/2} \int_0^t e^{-\tau(t-s)} W(s) ds, \quad Z_0 \sim \mathcal{N}(0,1).$$

Step 2: Getting closer from the original process

Let's consider $\boldsymbol{Y}^{(\boldsymbol{n},\boldsymbol{m})} = (Y^{(n,m)}(t), 0 \le t < 1)$:

$$\boldsymbol{Y}^{(\boldsymbol{n},\boldsymbol{m})}(t) = Z_k^{(n)} + \frac{t - t_k^{(n,m)}}{t_{k+1}^{(n,m)} - t_k^{(n,m)}} \left(Z_{k+1}^{(n)} - Z_k^{(n)} \right), \quad t_k^{(n,m)} \le t < t_{k+1}^{(n,m)}.$$

The integral representation of a Ornstein uhlenbeck process

$$Z(t) = Z_0 e^{-\tau t} + \sqrt{2\tau} W(t) - \sqrt{2}\tau^{3/2} \int_0^t e^{-\tau(t-s)} W(s) ds, \quad Z_0 \sim \mathcal{N}(0,1).$$

Let's consider

$$\begin{aligned} F_{n,m}(\omega, x)(t_k^{(n,m)}) &= x\left(1 - \frac{2}{n}\right)^k + 2\sqrt{\frac{m}{n}}\left(1 - \frac{2}{n}\right)^{-1} \omega(t_k^{(n,m)}) \\ &\quad - 4\left(1 - \frac{2}{n}\right)^{-1} \sqrt{\frac{m}{n}} \left(\frac{1}{n} \sum_{i=1}^{k} \left(1 - \frac{2}{n}\right)^{k-i} \omega(t_i^{(n,m)}) \right) \end{aligned} \tag{1}$$

Step 2: Getting closer from the original process

Theorem
Given the previous notations

$$Y^{(n,m)} = F_{n,m}(W^{(n,m)}, Z_0),$$

and if $\left|\frac{M}{n} - 1\right| \leq n^{-1/3}$*, then there exists a constant C such that*

$$\pi(Y^{(n,m)}, \mathrm{Y}) \leq Cn^{-1/4} \ln(n).$$

Step 3: Finally ....

By controling the distance between $t_k^{(n,m)}$ and $U_k^{(n)}$, we finally have

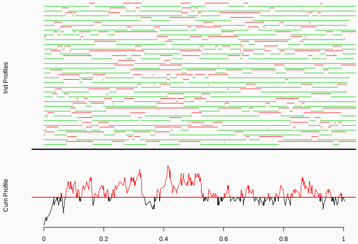
Theorem

For any κ in $]0, 1/4[$, there exists a constant C which only depends on κ such that

$$\pi\left(\mathsf{Z}^{(\mathsf{n})}, \mathsf{Z}\right) \leq \frac{C}{n^{\kappa}}.$$

Did we make any progress ?

Original problem



Characterizing the distribution of

$$E_*^{a,S_n}(1) = \max_i E_i(1)$$

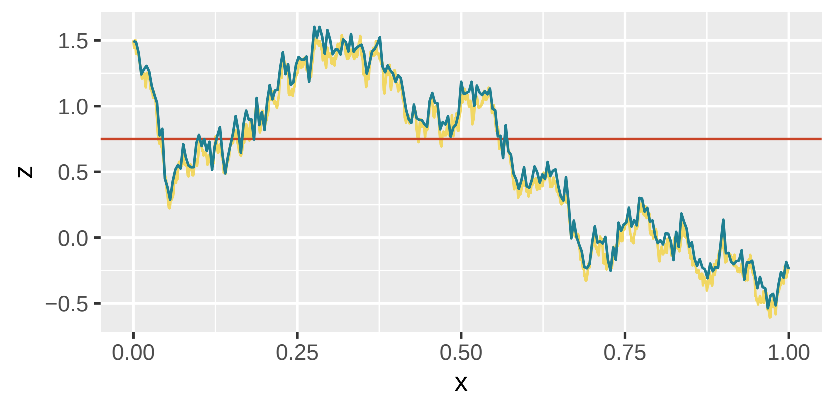
$$\mathbb{P}\left(E_*^{a,S_n} > l\right)$$

Convergence of the longest excursion

Some care is required as the excursion is not obviously a continuous functionnal of the process

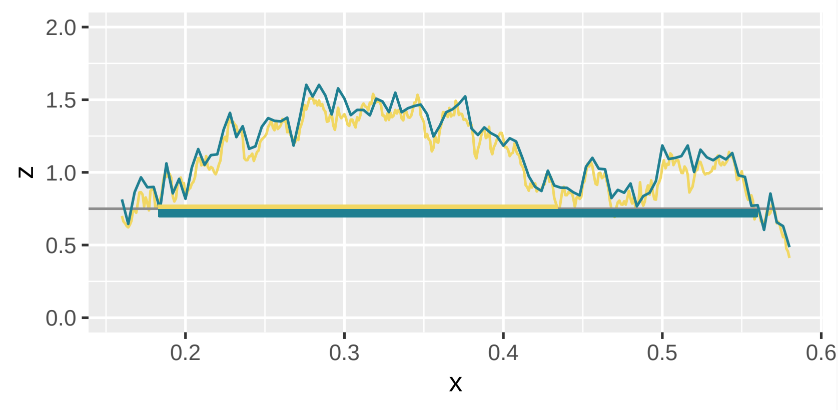
Convergence of the longest excursion

Some care is required as the excursion is not obviously a continuous functionnal of the process



Convergence of the longest excursion

Some care is required as the excursion is not obviously a continuous functionnal of the process



Convergence of the longest excursion

Some care is required as the excursion is not obviously a continuous functionnal of the process

$$\mathbb{P}\left(E_*^{a,S_n} > \ell\right) = \mathbb{P}\left(\sup_{0\leq s\leq t-l}\left\{\inf_{s\leq u<s+l} S_n(u)\right\} > a\right)$$

This proves the convergence as the event of interest is expressed through a continuous functionnal, but does nt control the rate

Convergence of the longest excursion

Let's denote $H(\omega) = \sup_{0\leq s\leq t-l} \left\{ \inf_{s\leq u<s+l} \omega(u) \right\}$,

We hope $\pi\left(H(\mathsf{Z}^{(\mathsf{n})}), H(\mathsf{Z})\right) \leq \varepsilon_n$, i.e. $\forall$ closed set A, $\mathbb{P}(H(\mathsf{Z}^{(\mathsf{n})}) \in A) \leq \mathbb{P}(H(\mathsf{Z}) \in A^{\varepsilon})$

$$\begin{aligned}
\mathbb{P}(H(\mathsf{Z}^{(\mathsf{n})}) > a) &= \mathbb{P}(\mathsf{Z}^{(\mathsf{n})} \in A), \quad \text{with } A = \{\omega, H(\omega) > a\} \\
&\leq \mathbb{P}\left(\mathsf{Z} \in A^{\varepsilon_n}\right) + \varepsilon_n, \\
&\leq \mathbb{P}\left(\mathsf{Z} \in A\right) + \mathbb{P}\left(\mathsf{Z} \in A^{\varepsilon_n} \cap \bar{A}\right) + \varepsilon_n, \\
&\leq \mathbb{P}\left(H(\mathsf{Z}) > a\right) + \mathbb{P}\left(\mathsf{Z} \in A^{\varepsilon_n} \cap \bar{A}\right) + \varepsilon_n.
\end{aligned}$$

Convergence of the longest excursion

Let's denote $H(\omega) = \sup_{0 \le s \le t-l} \left\{ \inf_{s \le u < s+l} \omega(u) \right\}$,

We hope $\pi \left(H(\mathsf{Z}^{(\mathsf{n})}), H(\mathsf{Z}) \right) \le \varepsilon_n$, i.e. $\forall$ closed set A, $\mathbb{P}(H(\mathsf{Z}^{(\mathsf{n})}) \in A) \le \mathbb{P}(H(\mathsf{Z}) \in A^{\varepsilon})$

$$\begin{aligned}
\mathbb{P}(H(\mathsf{Z}^{(\mathsf{n})}) > a) &= \mathbb{P}(\mathsf{Z}^{(\mathsf{n})} \in A), \quad \text{with } A = \{\omega, H(\omega) > a\} \\
&\le \mathbb{P}\left(\mathsf{Z} \in A^{\varepsilon_n}\right) + \varepsilon_n, \\
&\le \mathbb{P}\left(\mathsf{Z} \in A\right) + \mathbb{P}\left(\mathsf{Z} \in A^{\varepsilon_n} \cap \bar{A}\right) + \varepsilon_n, \\
&\le \mathbb{P}\left(H(\mathsf{Z}) > a\right) + \mathbb{P}\left(\mathsf{Z} \in A^{\varepsilon_n} \cap \bar{A}\right) + \varepsilon_n.
\end{aligned}$$

$\left\{\mathsf{Z} \in A^{\varepsilon_n} \cap \bar{A}\right\} = \{\exists \omega, H(\omega) > a, \|\omega - \mathsf{Z}\| \le \varepsilon_n, H(\mathsf{Z}) \le a\} \subset \{a - \varepsilon_n < H(\mathsf{Z}) \le a\}$

Convergence of the longest excursion

Let's denote $H(\omega) = \sup_{0\leq s\leq t-l} \left\{ \inf_{s\leq u<s+l} \omega(u) \right\}$,

We hope $\pi\left(H(\mathsf{Z}^{(\mathsf{n})}), H(\mathsf{Z})\right) \leq \varepsilon_n$, i.e. $\forall$ closed set A, $\mathbb{P}(H(\mathsf{Z}^{(\mathsf{n})}) \in A) \leq \mathbb{P}(H(\mathsf{Z}) \in A^{\varepsilon})$

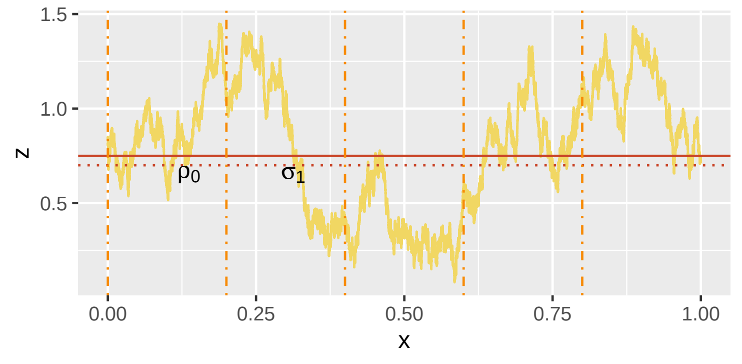
$$\begin{aligned}
\mathbb{P}(H(\mathsf{Z}^{(\mathsf{n})}) > a) &= \mathbb{P}(\mathsf{Z}^{(\mathsf{n})} \in A), \quad \text{with } A = \{\omega, H(\omega) > a\} \\
&\leq \mathbb{P}\left(\mathsf{Z} \in A^{\varepsilon_n}\right) + \varepsilon_n, \\
&\leq \mathbb{P}\left(\mathsf{Z} \in A\right) + \mathbb{P}\left(\mathsf{Z} \in A^{\varepsilon_n} \cap \bar{A}\right) + \varepsilon_n, \\
&\leq \mathbb{P}\left(H(\mathsf{Z}) > a\right) + \mathbb{P}\left(\mathsf{Z} \in A^{\varepsilon_n} \cap \bar{A}\right) + \varepsilon_n.
\end{aligned}$$

$$\left\{\mathsf{Z} \in A^{\varepsilon_n} \cap \bar{A}\right\} = \{\exists \omega, H(\omega) > a, \|\omega - \mathsf{Z}\| \leq \varepsilon_n, H(\mathsf{Z}) \leq a\} \subset \{a - \varepsilon_n < H(\mathsf{Z}) \leq a\}$$

Proposition
There exists $c > 0$ and a sequence ε_n decreasing to zero such that

$$\mathbb{P}^{\lambda}(a - \varepsilon_n < H(\mathsf{Z}) \leq a) \leq c\varepsilon_n$$

Let's denote $F_y^s(a) = \mathbb{P}_{Z(0)=y}\left(\forall t \leq s - l, \inf_{t \leq u \leq t+l} Z(u) \leq a\right)$,
We control $F_y^s(a) - F_y^s(a - \varepsilon)$, reasonning by induction and conditioning on $Z(kl)$



Finally

Theorem
There exists C and $\kappa \in [0, 1/4[$, *such that*

$$\pi(H(\mathsf{Z}^{(\mathsf{n})}), H(\mathsf{Z})) \leq \frac{C}{n^{\kappa}}$$

Conclusion

The bad news

- We prove the convergence of the pure jumps Markov Process to an Ornstein Uhlenbeck Process,
- as well as the convergence of the longest excursion, with the same rate of convergence
- with a descent rate
- in only 10 years,

Conclusion

The bad news

- We prove the convergence of the pure jumps Markov Process to an Ornstein Uhlenbeck Process,
- as well as the convergence of the longest excursion, with the same rate of convergence
- with a descent rate
- in only 10 years,

The good news

- We aim at approximating $\mathbb{P}\left(E_*^{a,S_n} > l\right)$ by $\mathbb{P}\left(E_*^{a,Z} > l\right)$,
- but $\mathbb{P}\left(E_*^{a,Z} > l\right)$ is only known for $a = 0$,

Conclusion

The bad news

- We prove the convergence of the pure jumps Markov Process to an Ornstein Uhlenbeck Process,
- as well as the convergence of the longest excursion, with the same rate of convergence
- with a descent rate
- in only 10 years,

The good news

- We aim at approximating $\mathbb{P}\left(E_*^{a,S_n} > l\right)$ by $\mathbb{P}\left(E_*^{a,Z} > l\right)$,
- but $\mathbb{P}\left(E_*^{a,Z} > l\right)$ is only known for $a = 0$,
- We have good plan for the 10 years to come.

THANK YOU !

See You in 10 years !

Références

Références

[Kub94] K Kubilius. "Rate of convergence in the invariance principle for martingale difference arrays". In: Lithuanian Mathematical Journal 34.4 (1994), pp. 383–392.

Appendix

Practical approach

Practical evaluation of $\mathbb{P}(E_*^{a,Z} > s)$ via MC approach

Several alternatives:

1. Sample a discretized version, named Z^δ of Z,
2. Sample the first hitting time σ_a according to [Alili+2005] and then discretized Z,

Practical evaluation of $\mathbb{P}(E_*^{a,Z} > s)$ via MC approach

Several alternatives:

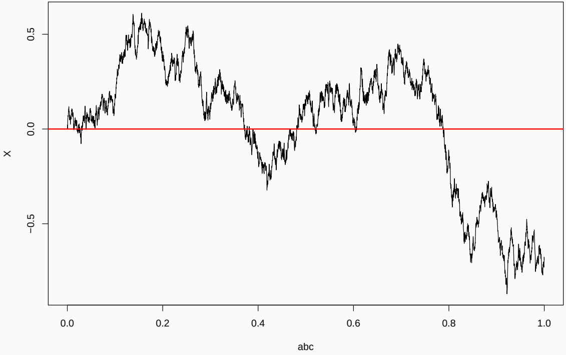
1. Sample a discretized version, named Z^δ of Z,
2. Sample the first hitting time σ_a according to *[Alili+2005]* and then discretized Z,
3. Use splitting rare events technics to sample σ_a and then use previous discretized Z,

Practical evaluation of $\mathbb{P}(E_*^{a,Z} > s)$ via MC approach

Several alternatives:

1. Sample a discretized version, named Z^δ of Z,
2. Sample the first hitting time σ_a according to [Alili+2005] and then discretized Z,
3. Use splitting rare events technics to sample σ_a and then use previous discretized Z,
4. Use splitting rare events technics to sample σ_a and use theoretical work for Importance sampling approach.

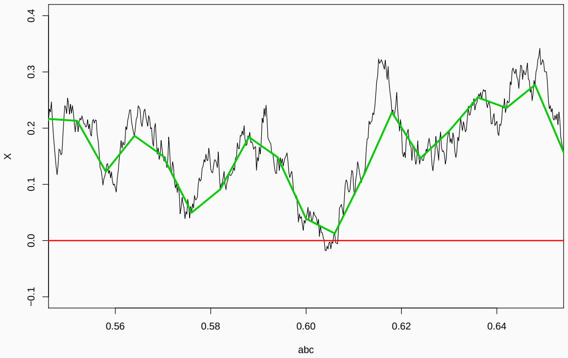
Validity of the approximation via E_*^{a,Z^δ}



Validity of the approximation via E_*^{a,Z^δ}

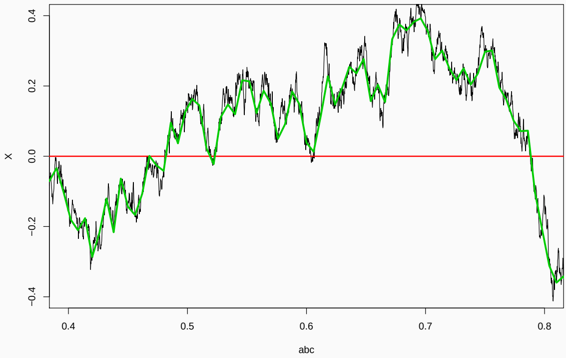
Validity of the approximation via E_*^{a,Z^δ}

Validity of the approximation via E_*^{a,Z^δ}



Validity of the approximation via E_*^{a,Z^δ}

Validity of the approximation via E_*^{a,Z^δ}



Under regularity conditions, [Azais1990] proves the convergence of the excursion of Z^δ to the excursions of Z, when Z is a Gaussian processes.

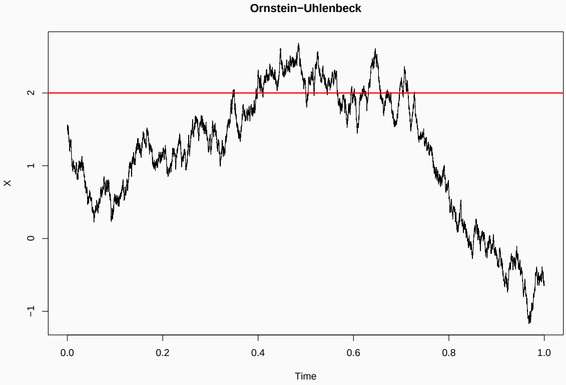
Validity of the approximation via E_*^{a,Z^δ}

But poor efficiency

Sample the first hitting time using the density

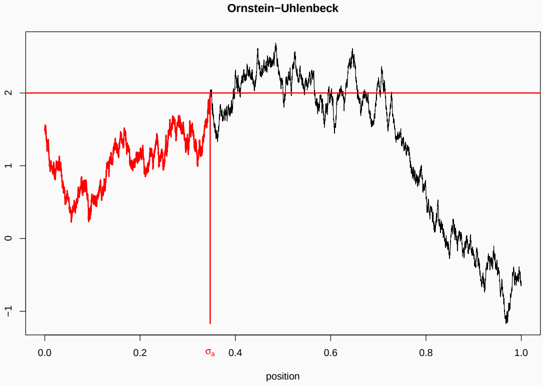
Sample the first hitting time using the density

Key idea



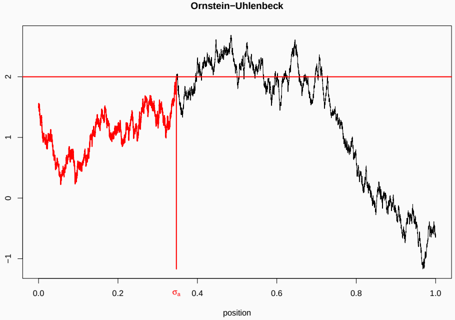
Sample the first hitting time using the density

Key idea



Sample the first hitting time using the density

Key idea



Density of σ_a starting from x.

$$p_{x,\,a}(r) = \frac{|a-x|}{\sqrt{2\pi r^3}} \,\mathbf{exp}\left(-\frac{\tau}{2}(a^2 - x^2 - r) - \frac{(a-x)^2}{2r}\right)$$

$$\times\, \mathbb{E}\mathbf{exp}\left(-\frac{\tau^2}{2}\int_0^r (\mathrm{Bes}_{0,a-x,r}(u) - a)^2 \,\mathrm{d}u\right),$$

where $Bes_{0,a-x,r}$ is a three dimensional Bessel bridge over [0, r] between 0 and $a - x$.

Sample the first hitting time using multilevel splitting

Sample the first hitting time using multilevel splitting

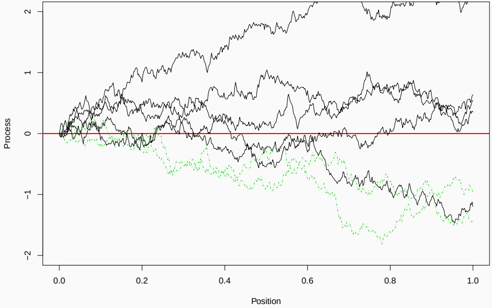
[Cerou+2005] presents a particle algorithm for rare events to sample σ_a.

Let $a_1 < a_2 < \ldots < a_K = a$, $\mathbb{P}\left(\sigma_a < t\right) = \mathbb{P}(\sigma_{a_1} < t)\mathbb{P}(\sigma_{a_2} < t|\sigma_{a_1} < t)\ldots\mathbb{P}(\sigma_a < t|\sigma_{a_{K-1}}|t)$

Sample the first hitting time using multilevel splitting

[Cerou+2005] presents a particle algorithm for rare events to sample σ_a.

Let $a_1 < a_2 < \ldots < a_K = a$, $\mathbb{P}\left(\sigma_a < t\right) = \mathbb{P}(\sigma_{a_1} < t)\mathbb{P}(\sigma_{a_2} < t|\sigma_{a_1} < t)\ldots\mathbb{P}(\sigma_a < t|\sigma_{a_{K-1}}|t)$

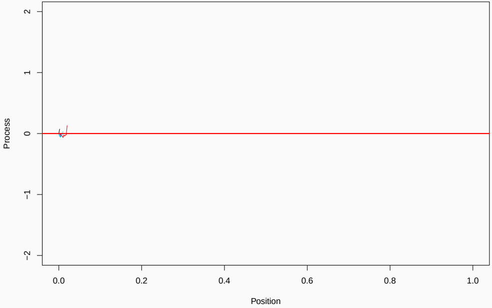


sampling with fixed success: G trajectories have to reach the next level.

Sample the first hitting time using multilevel splitting

[Cerou+2005] presents a particle algorithm for rare events to sample σ_a.

Let $a_1 < a_2 < \ldots < a_K = a$, $\mathbb{P}\left(\sigma_a < t\right) = \mathbb{P}(\sigma_{a_1} < t)\mathbb{P}(\sigma_{a_2} < t|\sigma_{a_1} < t)\ldots\mathbb{P}(\sigma_a < t|\sigma_{a_{K-1}}|t)$

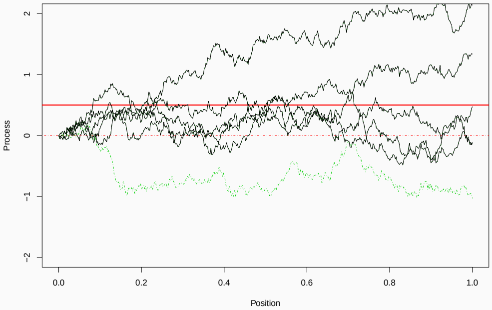


sampling with fixed success: G trajectories have to reach the next level.

Sample the first hitting time using multilevel splitting

[Cerou+2005] presents a particle algorithm for rare events to sample σ_a.

Let $a_1 < a_2 < \ldots < a_K = a$, $\mathbb{P}(\sigma_a < t) = \mathbb{P}(\sigma_{a_1} < t)\mathbb{P}(\sigma_{a_2} < t | \sigma_{a_1} < t) \ldots \mathbb{P}(\sigma_a < t | \sigma_{a_{K-1}} | t)$

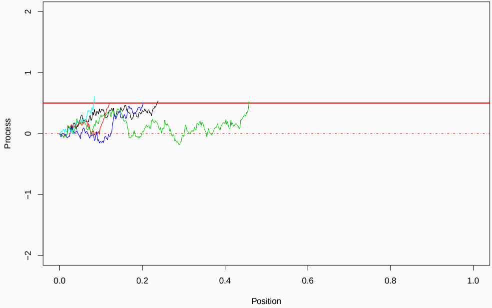


sampling with fixed success: G trajectories have to reach the next level.

Sample the first hitting time using multilevel splitting

[Cerou+2005] presents a particle algorithm for rare events to sample σ_a.

Let $a_1 < a_2 < \ldots < a_K = a$, $\mathbb{P}(\sigma_a < t) = \mathbb{P}(\sigma_{a_1} < t)\mathbb{P}(\sigma_{a_2} < t|\sigma_{a_1} < t)\ldots\mathbb{P}(\sigma_a < t|\sigma_{a_{K-1}}|t)$

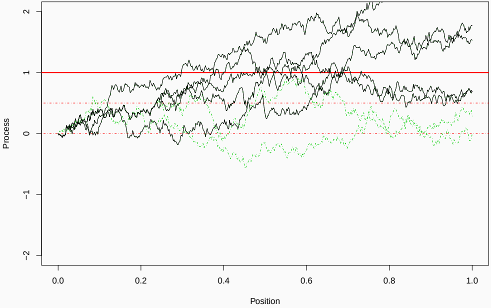


sampling with fixed success: G trajectories have to reach the next level.

Sample the first hitting time using multilevel splitting

[Cerou+2005] presents a particle algorithm for rare events to sample σ_a.

Let $a_1 < a_2 < \ldots < a_K = a$, $\mathbb{P}(\sigma_a < t) = \mathbb{P}(\sigma_{a_1} < t)\mathbb{P}(\sigma_{a_2} < t|\sigma_{a_1} < t)\ldots\mathbb{P}(\sigma_a < t|\sigma_{a_{K-1}}|t)$



sampling with fixed success: G trajectories have to reach the next level.