Analyse statistique d'une chaîne de fragmentation auto-similaire conservative.

Marc Hoffmann[1] et Nathalie Krell[2]

LAMA Université Paris-Est, LPMA Université Paris 6 et MAP5 Université Paris 5.

Journées de Probabilités à Lilles
4 septembre 2008

[1]marc.hoffmann@univ-mlv.fr
[2]nathalie.krell@upmc.fr

Plan

Plan

Définition heuristique d'une fragmentation.

On va considérer les fragmentations à valeurs dans l'espace

$$\mathcal{S}^{\downarrow} := \left\{ \mathbf{s} = (s_1, s_2, \ldots),\ s_1 \geq s_2 \geq \ldots \geq 0,\ \sum_{i=1}^{\infty} s_i \leq 1 \right\}.$$

Une fragmentation X est un processus de Markov à valeur dans $\mathcal{S}^{\downarrow}$ vérifiant les deux propriétés.

1. **La propriété de branchement**: différents fragments ont des évolutions indépendantes
2. **La propriété d'auto-similarité**: la dislocation va se produire de manière auto-similaire.

Définition d'une fragmentation.

La fragmentation va être totalement caractérisée par $\alpha \in \mathbb{R}$ et une mesure de dislocation ν positive sur $\mathcal{S}^{\downarrow}$.
La manière dont un fragment se disloque va être déterminée par la mesure de dislocation ν.
La constante α gouverne la vitesse à laquelle le fragment va se disloquer.

Cadre : ν est une probabilité et $\alpha \geq 0$.

Cadre : ν est une probabilité et $\alpha \geq 0$.

Definition

Soient deux familles indépendantes de variables i.i.d. indexées par $\mathbb{T} := \cup_{n\in\mathbb{N}}\mathbb{N}^n$*:* $(\overline{\xi}_u, u \in \mathcal{U})$ *et* $(\mathbf{e}_u, u \in \mathcal{U})$*, où pour* $u \in \mathcal{U}$ $\overline{\xi}_u = (\widetilde{\xi}_{ui})_{i\in\mathbb{N}}$ *a pour loi* ν*, et* $(\mathbf{e}_{ui})_{i\in\mathbb{N}}$ *est une suite de variables exponentielles i.i.d. de paramètre 1. On définit par récurrence*

$$\xi_\emptyset := 1, \qquad a_\emptyset := 0, \qquad \zeta_\emptyset := x^{-\alpha}\mathbf{e}_\emptyset,$$

et pour $u \in \mathcal{U}$ *et* $i \in \mathbb{N}$*:*

$$\xi_{ui} := \widetilde{\xi}_{ui}\xi_u, \qquad a_{ui} := a_u + \zeta_u, \qquad \zeta_{ui} := \xi_{ui}^{-\alpha}\mathbf{e}_{ui}.$$

Cadre : ν est une probabilité et $\alpha \geq 0$.

Definition

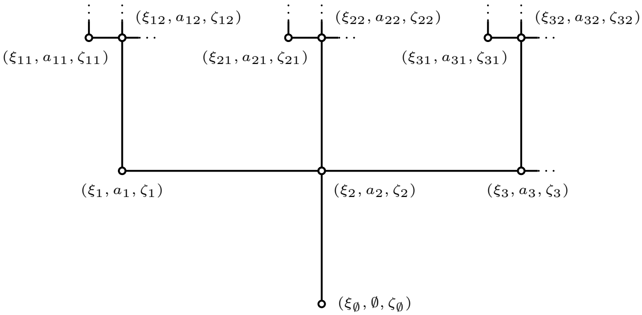
Soient deux familles indépendantes de variables i.i.d. indexées par $\mathbb{T} := \cup_{n\in\mathbb{N}}\mathbb{N}^n$*:* $(\overline{\xi}_u, u \in \mathcal{U})$ *et* $(\mathbf{e}_u, u \in \mathcal{U})$*, où pour* $u \in \mathcal{U}$ $\overline{\xi}_u = (\widetilde{\xi}_{ui})_{i\in\mathbb{N}}$ *a pour loi* ν*, et* $(\mathbf{e}_{ui})_{i\in\mathbb{N}}$ *est une suite de variables exponentielles i.i.d. de paramètre 1. On définit par récurrence*

$$\xi_\emptyset := 1, \qquad a_\emptyset := 0, \qquad \zeta_\emptyset := x^{-\alpha}\mathbf{e}_\emptyset,$$

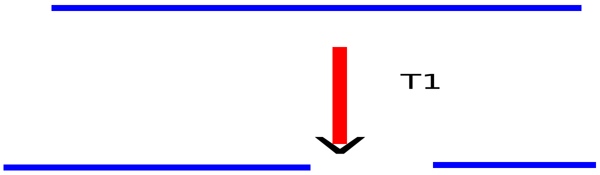
et pour $u \in \mathcal{U}$ *et* $i \in \mathbb{N}$*:*

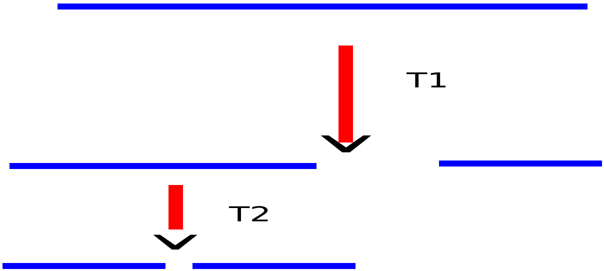
$$\xi_{ui} := \widetilde{\xi}_{ui}\xi_u, \qquad a_{ui} := a_u + \zeta_u, \qquad \zeta_{ui} := \xi_{ui}^{-\alpha}\mathbf{e}_{ui}.$$

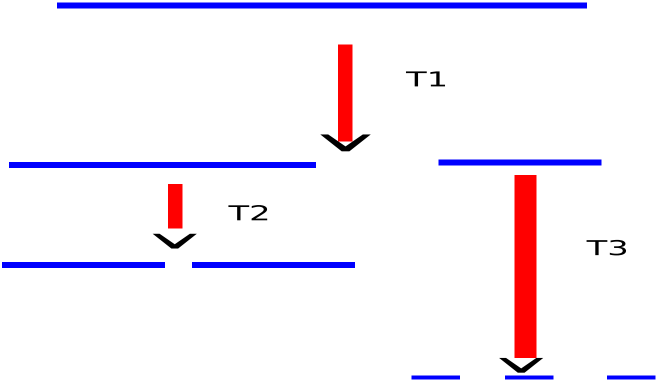
On va uniquement considérer les fragmentations "propres", c.à.d. $\sum_{i=1}^{\infty} s_i = 1 \quad \nu - \text{p.p.}$

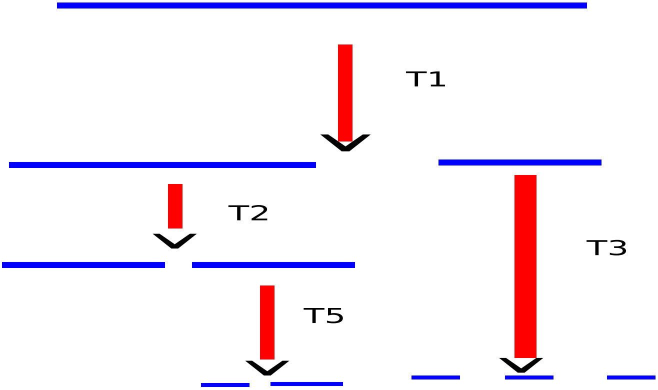

$(\xi_{12}, a_{12}, \zeta_{12})$
$(\xi_{22}, a_{22}, \zeta_{22})$
$(\xi_{32}, a_{32}, \zeta_{32})$
$(\xi_{11}, a_{11}, \zeta_{11})$
$(\xi_{21}, a_{21}, \zeta_{21})$
$(\xi_{31}, a_{31}, \zeta_{31})$
(ξ_1, a_1, ζ_1)
(ξ_2, a_2, ζ_2)
(ξ_3, a_3, ζ_3)
$(\xi_\emptyset, \emptyset, \zeta_\emptyset)$

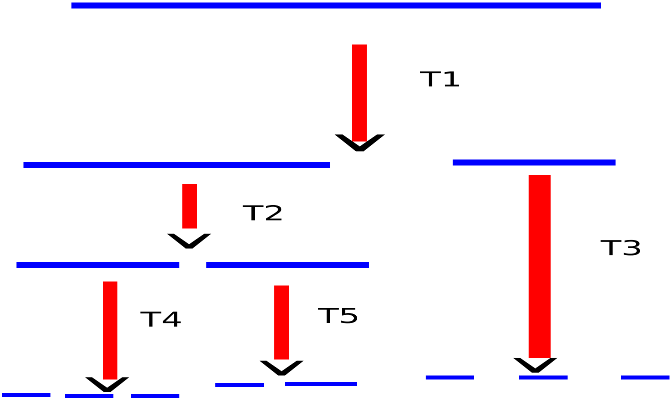
Un exemple.

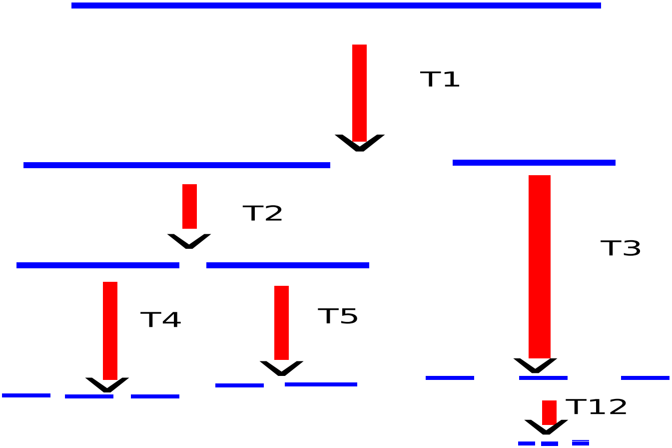

T1

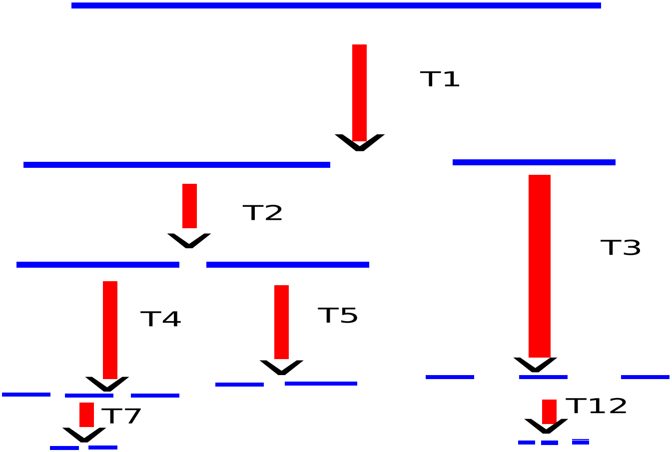

T1
T2

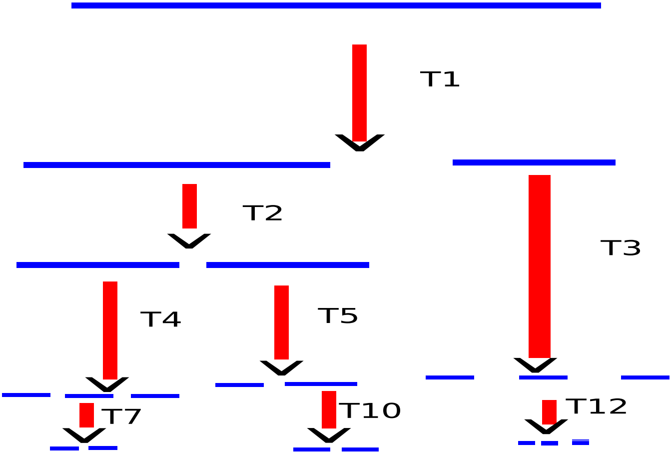

T1
T2
T3

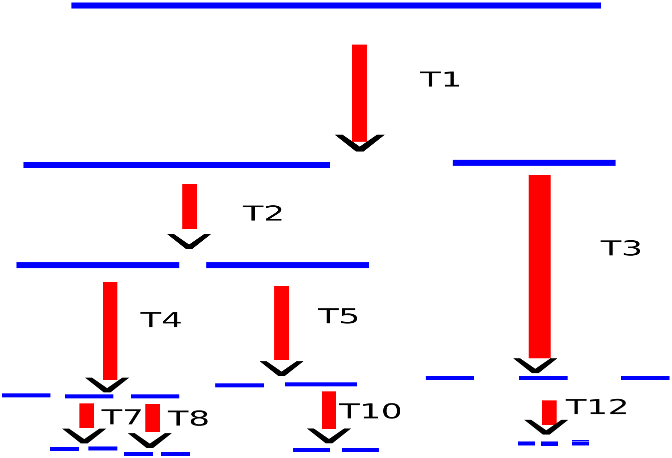

T1
T2
T3
T5

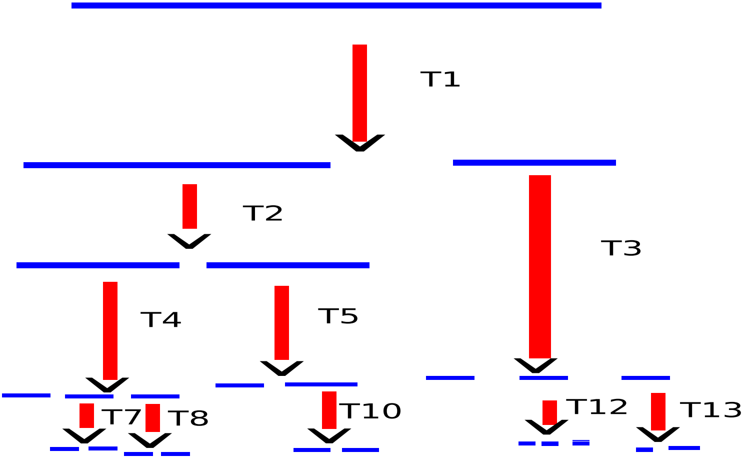

T1
T2
T3
T4
T5

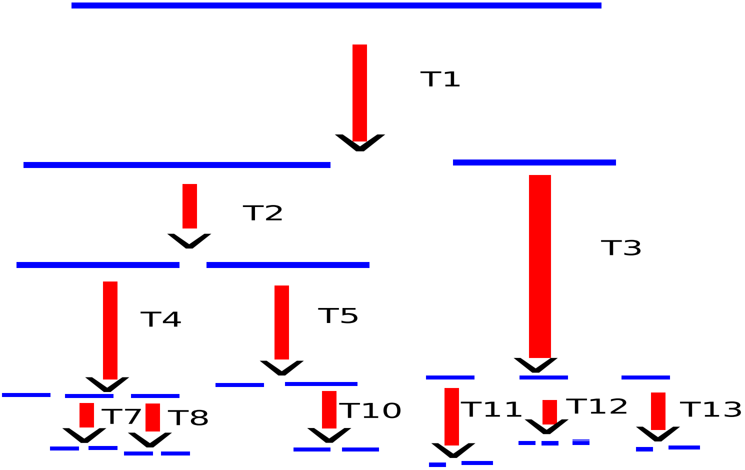

T1
T2
T3
T4
T5
T12

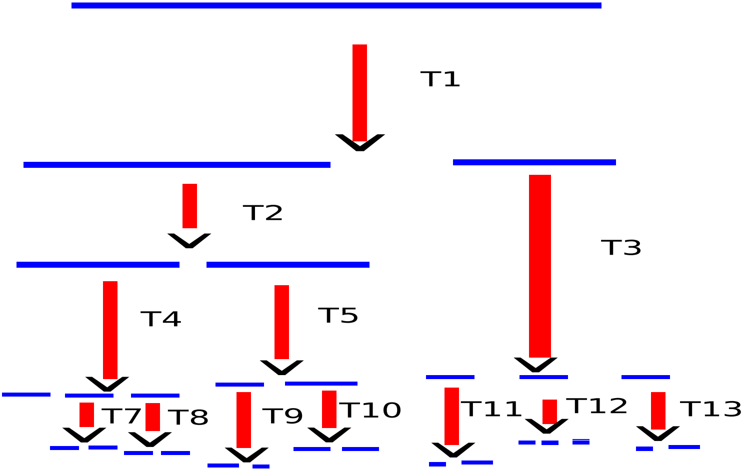

T1
T2
T3
T4
T5
T7
T12

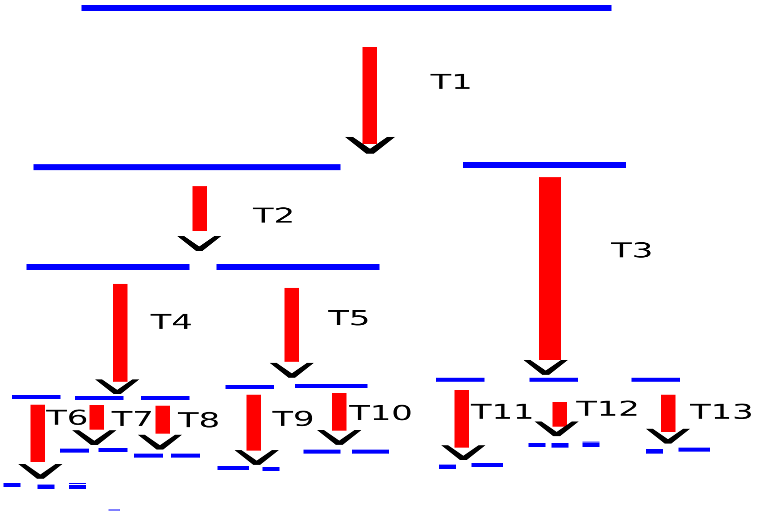

T1
T2
T3
T4
T5
T7
T10
T12


T1
T2
T3
T4
T5
T7
T8
T10
T12


T1
T2
T3
T4
T5
T7
T8
T10
T12
T13

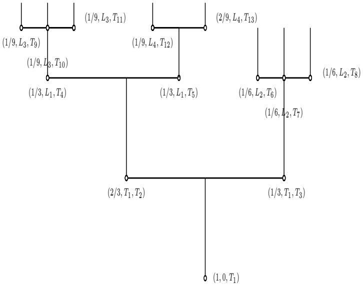

T1
T2
T3
T4
T5
T7
T8
T10
T11
T12
T13

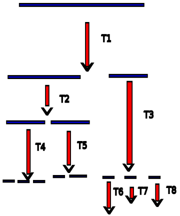

T1
T2
T3
T4
T5
T7
T8
T9
T10
T11
T12
T13


T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
T11
T12
T13

Lien fragmentation et arbre généalogique.

On utilisera donc la suite le fait que l'on peut réécrire les $\{X_i(t),\ i \in \mathbb{N},\ t \geq 0\}$ sous la forme $\{\xi_u, u \in \mathcal{U}\}$, avec $\mathcal{U} = \cup_{n=0}^{\infty} \mathbb{N}^n$.





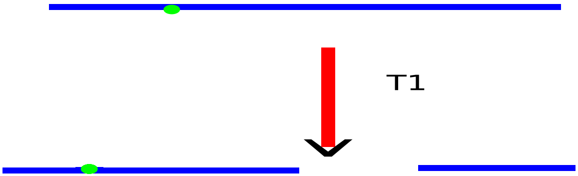
Un subordinateur important.

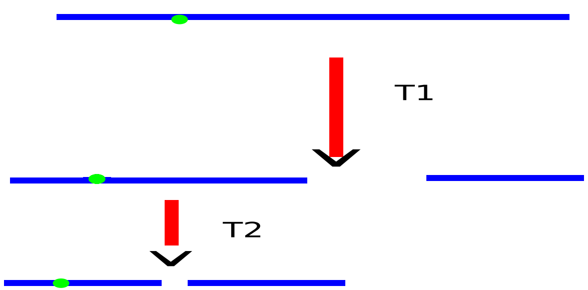
Soit $\chi(t) =$ la taille du fragment marqué. C'est à dire que $\chi(t) = X_{n(t)}(t)$ est un fragment choisi de manière biaisée par la taille: $\mathbb{P}(n(t) = l|\ X(t)) = X_l(t)$.
Bertoin a montré que

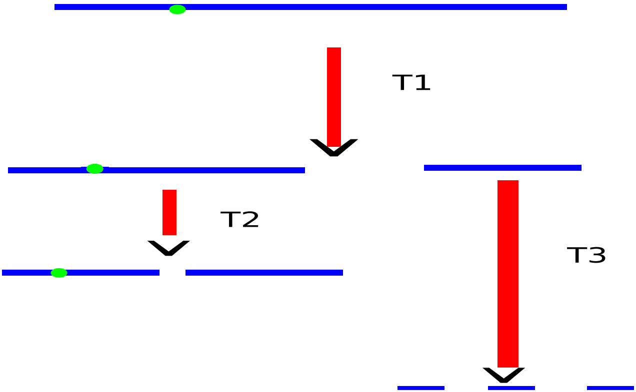
$$\xi_t := -\log \chi(t), \quad t \geq 0, \tag{1}$$

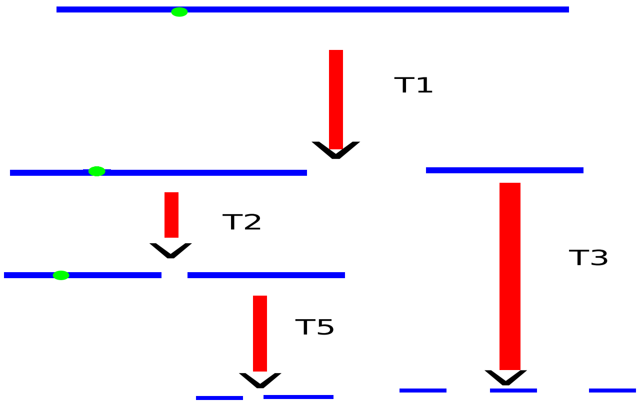
est un subordinateur.

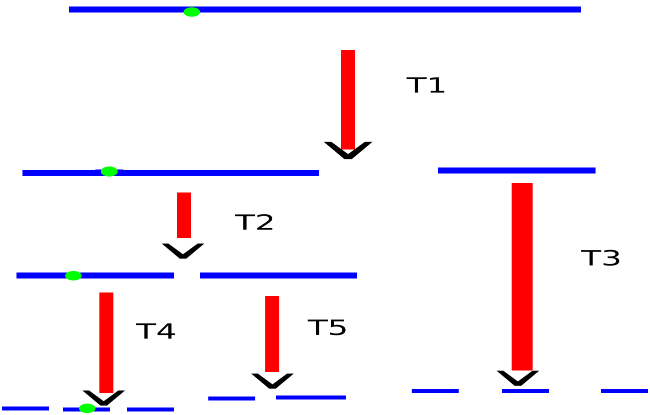
Evolution du fragment marqué.

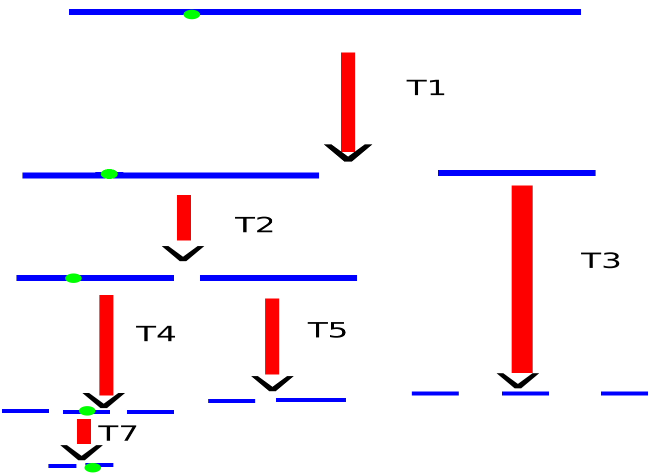

T1

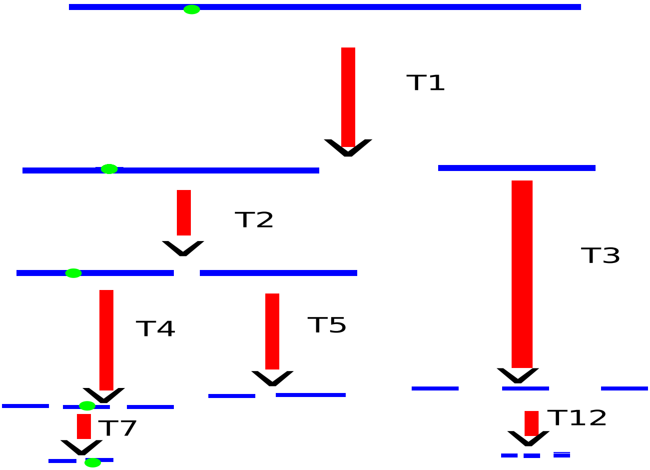

T1
T2

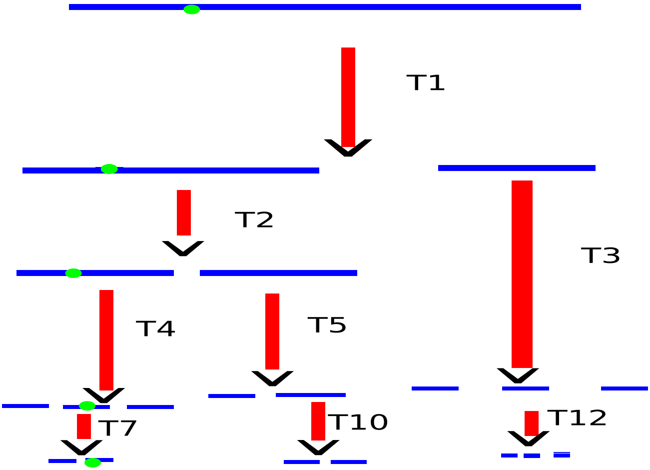

T1
T2
T3

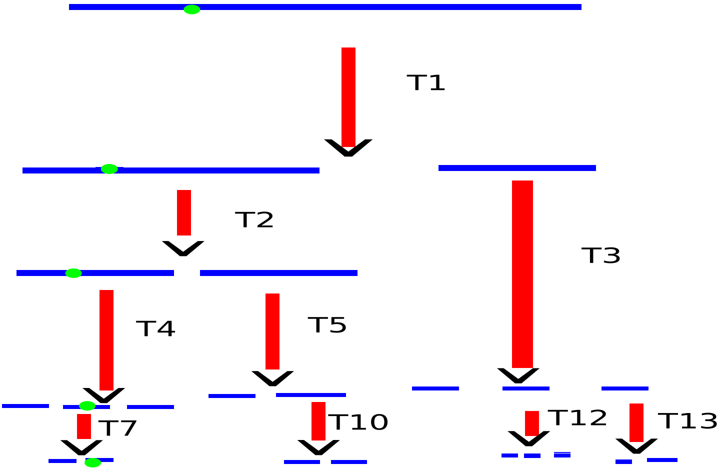

T1
T2
T3
T5

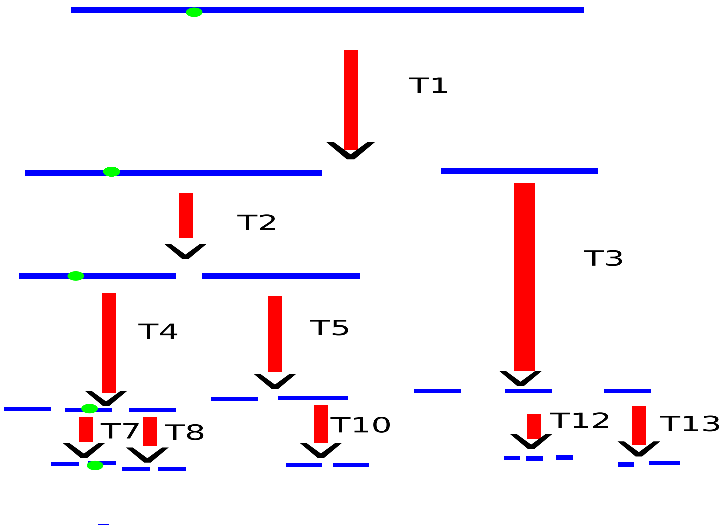

T1
T2
T3
T4
T5

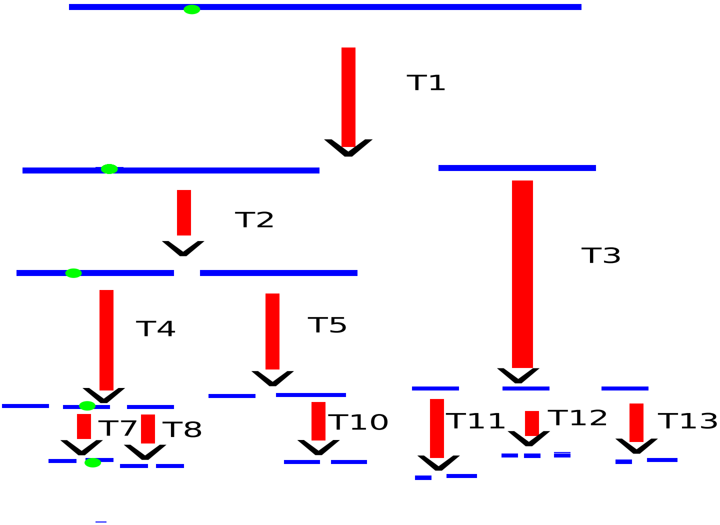

T1
T2
T3
T4
T5
T7

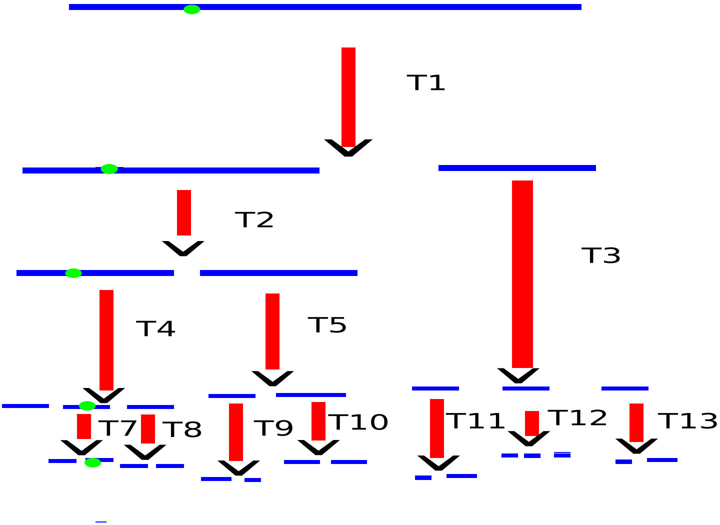

T1
T2
T3
T4
T5
T7
T12

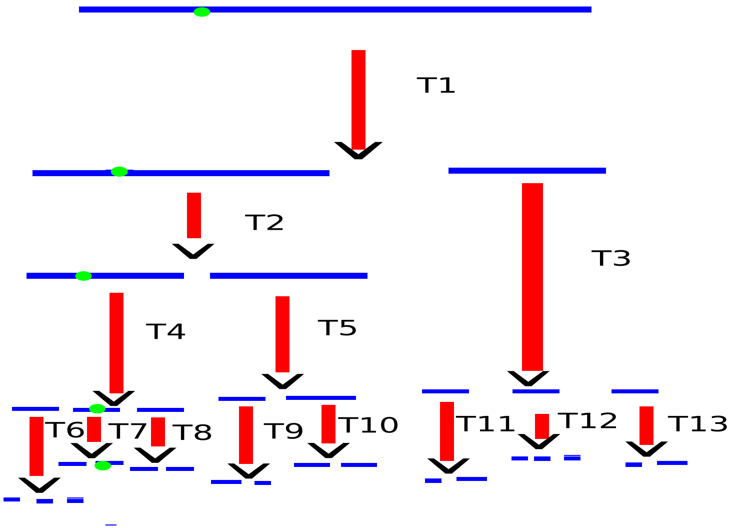

T1
T2
T3
T4
T5
T7
T10
T12


T1
T2
T3
T4
T5
T7
T10
T12
T13

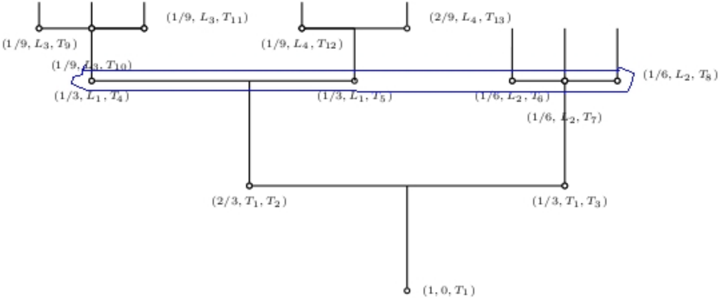

T1
T2
T3
T4
T5
T7
T8
T10
T12
T13


T1
T2
T3
T4
T5
T7
T8
T10
T11
T12
T13

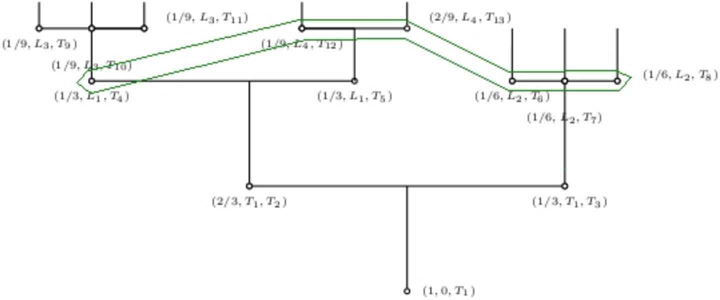

T1
T2
T3
T4
T5
T7
T8
T9
T10
T11
T12
T13


T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
T11
T12
T13

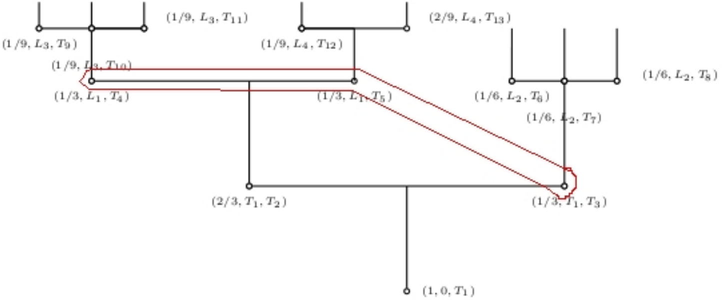
Ligne d'arrêt: la 2-ième génération



Ligne d'arrêt: le fragment en vie au temps $t \in [T_1 + T_2 + T_5, T_1 + T_2 + T_4)$.



Ligne d'arrêt: le fragment gelé dès qu'il a une taille plus petite que 0.34.



Ligne d'arrêt L et le subordinateur.

Pour L une ligne d'arrêt, $\chi(t)$ le fragment marqué et $\xi(t) = -\log(\chi(t))$, on obtient $\mathbb{E}\left[\sum_{v\in L} \xi_v \, g(\xi_v)\right] = \mathbb{E}^{\star}\left[g\left(\chi(T_L)\right)\right] = \mathbb{E}^{\star}\left[g\left(e^{-\xi(T_L)}\right)\right]$.

Ligne d'arrêt L et le subordinateur.

Pour L une ligne d'arrêt, $\chi(t)$ le fragment marqué et $\xi(t) = -\log(\chi(t))$, on obtient
$\mathbb{E}\left[\sum_{v\in L} \xi_v\, g(\xi_v)\right] = \mathbb{E}^\star\left[g\left(\chi(T_L)\right)\right] = \mathbb{E}^\star\left[g\left(e^{-\xi(T_L)}\right)\right]$.

Par exemple pour la ligne: la n-ième génération:
$\mathbb{E}\left[\sum_{|v|=n} \xi_v\, g(\xi_v)\right] = \mathbb{E}^\star\left[g\left(\chi_n\right)\right] = \mathbb{E}^\star\left[g\left(e^{-\xi_n}\right)\right]$.

Ligne d'arrêt L et le subordinateur.

Pour L une ligne d'arrêt, $\chi(t)$ le fragment marqué et $\xi(t) = -\log(\chi(t))$, on obtient
$\mathbb{E}\left[\sum_{v\in L} \xi_v \, g(\xi_v)\right] = \mathbb{E}^\star\left[g\left(\chi(T_L)\right)\right] = \mathbb{E}^\star\left[g\left(e^{-\xi(T_L)}\right)\right]$.

Par exemple pour la ligne: la n-ième génération:
$\mathbb{E}\left[\sum_{|v|=n} \xi_v \, g(\xi_v)\right] = \mathbb{E}^\star\left[g\left(\chi_n\right)\right] = \mathbb{E}^\star\left[g\left(e^{-\xi_n}\right)\right]$.

Et pour la ligne associé au fragment en vie au temps t:
$\mathbb{E}\left[\sum_{v:\ t\in[a_u, a_u+\zeta_u)} \xi_v \, g(\xi_v)\right] = \mathbb{E}^\star\left[g\left(\chi(t)\right)\right] = \mathbb{E}^\star\left[g\left(e^{-\xi(t)}\right)\right]$.

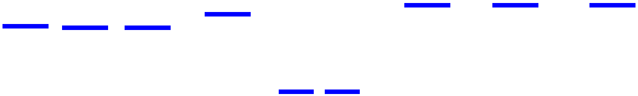
L'observable.

On observe la taille du fragment dès que qu'elle a atteint une taille inférieure à $\varepsilon > 0$: $X_\varepsilon := (\xi_u, u \in \mathcal{U}_\varepsilon)$ avec

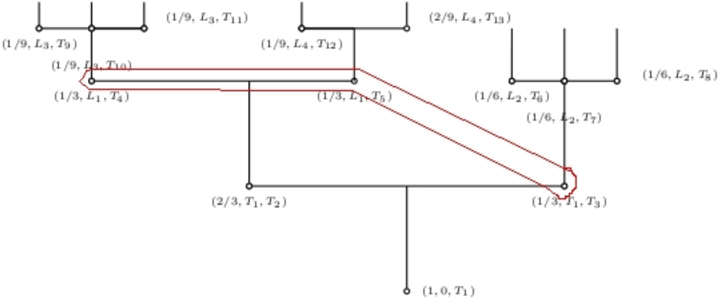
$$\mathcal{U}_\varepsilon := \{u \in \mathcal{U},\ \ \xi_{u-} \geq \varepsilon, \xi_u < \varepsilon\},$$

et $u-$ le parent du fragment d'étiquette u.
On va considérer l'asymptotique lorsque $\varepsilon \to 0$.

Un exemple.



Une ligne d'arrêt: le fragment est gelé dès qu'il a une taille plus petite que 0.34.



Une observable plus réaliste.

On s'est aussi intéressé à une observation un peu plus fine qui correspond au fait que l'on observe X_ε avec une perte d'information dûe à un bruit systématique d'intensité σ petit devant ε, c'est à dire que l'observable est

$$X_{\varepsilon,\sigma} := \left(\xi_u^\sigma, \quad \xi_{u-}^\sigma \geq \varepsilon,\ t_\sigma \leq \xi_u^\sigma < \varepsilon\right)$$

avec

$$\xi_u^\sigma := \xi_u + \sigma U_u.$$

avec $(U_u, u \in \mathcal{U})$ i.i.d. et indépendantes de X_ε et telles que pour tout $u \in \mathcal{U}$,

$$|U_u| \leq 1 \ \text{ et } \ \mathbb{E}[U_u] = 0$$

et t_σ tel que $0 < t_\sigma \leq \sigma \ll \varepsilon$.

Mesure empirique.

Ainsi on va considérer d'abord la mesure empirique

$$\mathcal{E}_\varepsilon(g) := \sum_{u \in \mathcal{U}_\varepsilon} \xi_u \, g(\xi_u/\varepsilon),$$

avec $g(\cdot)$ une fonction test.
Bertoin et Martinez ont montré dans [2] que, sous des hypothèses assez faibles sur $\nu(\cdot)$, la mesure $\mathcal{E}_\varepsilon(g)$ converge vers

$$\mathcal{E}(g) := \frac{1}{c(\nu)} \int_0^1 \frac{g(a)}{a} \int_{\mathcal{S}^\downarrow} \sum_{i=1}^{\infty} s_i \, 1_{\{s_i < a\}} \nu(d\mathbf{s}) da$$

en probabilité quand $\varepsilon \to 0$, avec $c(\nu) = -\int_{\mathcal{S}^\downarrow} \sum_{i=1}^{\infty} s_i \log s_i \, \nu(d\mathbf{s})$.

Un outil crucial.

On associe au subordinateur $\xi(t) = -\log \chi(t)$ le temps d'arrêt

$$T_\varepsilon := \{t \geq 0, \chi(t) < \varepsilon\},$$

Le valeur du subordinateur considéré au premier temps où il devient plus petit que $\log(1/\varepsilon)$ est importante comme le montre l'égalité suivante:

$$\mathbb{E}\left[\sum_{v:\, \xi_{u-} \geq \varepsilon, \xi_u < \varepsilon} \xi_v\, g(\xi_v)\right] = \mathbb{E}^\star\left[g\left(\chi(T_\varepsilon)\right)\right] = \mathbb{E}^\star\left[g\left(e^{-\xi(T_\varepsilon)}\right)\right].$$

D'où

$$\mathbb{E}\left[\mathcal{E}_\varepsilon(g)\right] = \mathbb{E}\left[\sum_{u \in \mathcal{U}} \mathbb{1}_{\{\xi_{u-} \geq \varepsilon, \xi_u < \varepsilon\}} \xi_u\, g(\xi_u/\varepsilon)\right] = \mathbb{E}^\star\left[g(\chi(T_\varepsilon)/\varepsilon)\right]$$

But.

Notre but va être d'estimer

$$\pi(dx) := e^{-x} \sum_{i=1}^{\infty} \nu(-\log s_i \in dx).$$

la mesure de Lévy du subordinateur $\xi(t)$.

But.

Notre but va être d'estimer

$$\pi(dx) := e^{-x} \sum_{i=1}^{\infty} \nu(-\log s_i \in dx).$$

la mesure de Lévy du subordinateur $\xi(t)$.
α est un paramètre nuisible.

But.

Notre but va être d'estimer

$$\pi(dx) := e^{-x} \sum_{i=1}^{\infty} \nu(-\log s_i \in dx).$$

la mesure de Lévy du subordinateur $\xi(t)$.
α est un paramètre nuisible.
La mesure π intervient explicitement dans l'énergie $\mathcal{E}(g)$ comme le montrent les égalités suivantes:

$$\begin{aligned}\mathcal{E}(g) &= \frac{1}{c(\pi)} \int_0^1 \frac{g(a)}{a} \pi(-\log a, +\infty)\, da \\ &= \frac{1}{c(\pi)} \int_0^{+\infty} g(e^{-x}) \pi(x, +\infty)\, dx.\end{aligned}$$

Stratégie.

- Etablir la vitesse de convergence de $\mathcal{E}_\varepsilon(g)$.

Stratégie.

• Etablir la vitesse de convergence de $\mathcal{E}_\varepsilon(g)$.

• Choisir les bonnes fonctions tests g afin de savoir ce qui est contenu dans la limite $\mathcal{E}(g)$.

Stratégie.

• Etablir la vitesse de convergence de $\mathcal{E}_\varepsilon(g)$.

• Choisir les bonnes fonctions tests g afin de savoir ce qui est contenu dans la limite $\mathcal{E}(g)$.

• En déduire des estimateurs paramétriques et non-paramétriques.

Stratégie.

- Etablir la vitesse de convergence de $\mathcal{E}_{\varepsilon}(g)$.

- Choisir les bonnes fonctions tests g afin de savoir ce qui est contenu dans la limite $\mathcal{E}(g)$.

- En déduire des estimateurs paramétriques et non-paramétriques.

- Généraliser à des observables plus réalistes.

Plan

Une vitesse.

On suppose que $\int_{[0,+\infty)} e^{\kappa x}\,\pi(dx) < +\infty$ pour un $\kappa > 0$.

Théorème

• *Pour tout* $m > 0$ *et* $0 < \mu < \kappa$

$$\sup_{g\in\mathcal{C}(m)} \mathbb{E}\Big[\big(\mathcal{E}_\varepsilon(g) - \mathcal{E}(g)\big)^2\Big] = o\big(\varepsilon^{\mu/(\mu+1)}\big).$$

avec $\mathcal{C}(m)$ *la classe des fonctions continues*
$\mathcal{C}(m) := \big\{g(\cdot) : [0,1] \to \mathbb{R},\ \ \|g\|_\infty := \sup_x |g(x)| \le m\big\}$.

Une vitesse.

On suppose que $\int_{[0,+\infty)} e^{\kappa x}\,\pi(dx) < +\infty$ pour un $\kappa > 0$.

Théorème

• *Pour tout* $m > 0$ *et* $0 < \mu < \kappa$

$$\sup_{g\in\mathcal{C}(m)} \mathbb{E}\Big[\big(\mathcal{E}_\varepsilon(g) - \mathcal{E}(g)\big)^2\Big] = o\big(\varepsilon^{\mu/(\mu+1)}\big).$$

avec $\mathcal{C}(m)$ *la classe des fonctions continues*
$\mathcal{C}(m) := \big\{g(\cdot) : [0,1] \to \mathbb{R},\ \ \|g\|_\infty := \sup_x |g(x)| \leq m\big\}$.

• *Le résultat reste valide si on suppose que* $|supp(g)| \leq b_\varepsilon$ *(et que* $\pi(dx) = \pi(x)dx$ *est absolument continue) avec un gain de* b_ε *dans la vitesse de convergence.*

Sous les mêmes hypothèses pour le cas de l'observation plus réaliste, on obtient un résultat analogue pour $\mathcal{E}_{\varepsilon,\sigma}(g) := \sum_{u\in\mathcal{U}_{\varepsilon,\sigma}} 1_{\{\xi_u^{(\sigma)} \geq t_\varepsilon\}} \xi_u^{(\sigma)} g\big(\xi_u^{(\sigma)}/\varepsilon\big)$:

Théorème

Pour tout $m > 0$ *et* $0 < \mu < \kappa$

$$\sup_{g\in\mathcal{C}'(m)} \mathbb{E}\Big[\big(\mathcal{E}_{\varepsilon,\sigma}(g) - \mathcal{E}_\varepsilon(g)\big)^2\Big] = \mathcal{O}(\sigma^2\varepsilon^{-2}).$$

avec $\mathcal{C}'(m)$ *la classe des fonctions continûment différentiables* $g(\cdot)$ *telles que* $g' \in \mathcal{C}(m)$.

Sous les mêmes hypothèses pour le cas de l'observation plus réaliste, on obtient un résultat analogue pour $\mathcal{E}_{\varepsilon,\sigma}(g) := \sum_{u\in\mathcal{U}_{\varepsilon,\sigma}} 1_{\{\xi_u^{(\sigma)} \geq t_\varepsilon\}} \xi_u^{(\sigma)} g\big(\xi_u^{(\sigma)}/\varepsilon\big)$:

Théorème

Pour tout $m > 0$ et $0 < \mu < \kappa$

$$\sup_{g\in\mathcal{C}'(m)} \mathbb{E}\Big[\big(\mathcal{E}_{\varepsilon,\sigma}(g) - \mathcal{E}_\varepsilon(g)\big)^2\Big] = \mathcal{O}(\sigma^2\varepsilon^{-2}).$$

avec $\mathcal{C}'(m)$ la classe des fonctions continûment différentiables $g(\cdot)$ telles que $g' \in \mathcal{C}(m)$.

Remarque

On obtient donc le même résultat que précédemment si $\sigma = \circ(\varepsilon^{3/2})$.

Une idée de la preuve.

$$\mathbb{E}\big[\mathcal{E}_\varepsilon(g)\big] = \mathbb{E}\left[\sum_{u\in\mathcal{U}} \mathbb{1}_{\{\xi_{u-}\geq\varepsilon,\xi_u<\varepsilon\}}\xi_u\, g(\xi_u/\varepsilon)\right]$$

En conditionnant par rapport aux parents de u: $u- = v$, et grâce à la propriété de branchement de la fragmentation, on obtient:

$$\begin{aligned}&\mathbb{E}\Big[\sum_{v\in\mathcal{U}} \xi_v \sum_{i=1}^{\infty} \mathbb{1}_{\left\{\widehat{\xi}_i\xi_v\geq\varepsilon\right\}}\widehat{\xi}_i g\big(\widehat{\xi}_i\xi_v/\varepsilon\big)\Big]\\ =&\mathbb{E}\Big[\sum_{v\in\mathcal{U}} \xi_v H_g\big(-\log\xi_v \;-\log(1/\varepsilon)\big)\Big]\end{aligned}$$

avec

$$H_g(a) := \mathbb{1}_{\{a\geq 0\}}\mathbb{E}\left[\sum_{i=1}^{\infty} \mathbb{1}_{\{\widehat{\xi}_i<e^{-a}\}}\widehat{\xi}_i g\big(\widehat{\xi}_i e^a\big)\right].$$

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{v\in\mathcal{U}}\xi_v H_g\big(-\log\xi_v \;-\log(1/\varepsilon)\big)\Big]\\
&=\sum_{n=0}^{\infty}\mathbb{E}\left[\sum_{|v|=n}\xi_v H_g\big(-\log\xi_v-\log(1/\varepsilon)\big)\right]\\
&=\sum_{n=0}^{\infty}\mathbb{E}^*\left[H_g\big(-\log\chi_n-\log(1/\varepsilon)\big)\right]
\end{aligned}$$

avec χ_n la taille du fragment marqué à la n-ième génération. On a donc $-\log\chi_n$ qui est une marche aléatoire de fonction de pas π.

Théorème de renouvellement.

Soit S_n une marche aléatoire (non discrète) de fonction de pas F et de moyenne positive m. $H = \sum_{n=0}^{\infty} F^{*n}$ la fonction de renouvellement.

Théorème (Théorème de renouvellement)

Le vitesse de convergence de

$$H * g(t) = \sum_{n=0}^{\infty} \mathbb{E}(g(S_n - t)) \quad \textit{vers} \quad \frac{1}{m} \int_0^{\infty} g(x)dx$$

est uniforme en g, avec $|g(x)| \leq r(x)$ avec r la fonction dominante (résultat de Sgibnev).

Théorème de renouvellement.

Soit S_n une marche aléatoire (non discrète) de fonction de pas F et de moyenne positive m. $H = \sum_{n=0}^{\infty} F^{*n}$ la fonction de renouvellement.

Théorème (Théorème de renouvellement)

Le vitesse de convergence de

$$H * g(t) = \sum_{n=0}^{\infty} \mathbb{E}(g(S_n - t)) \quad \textit{vers} \quad \frac{1}{m}\int_0^{\infty} g(x)dx$$

est uniforme en g, avec $|g(x)| \leq r(x)$ avec r la fonction dominante (résultat de Sgibnev).

D'où

$$\sum_{n=0}^{\infty} \mathbb{E}^* \left[H_g\big(-\log \chi_n - \log(1/\varepsilon)\big)\right] \to \frac{1}{\mathbb{E}^*(-\log \chi_1)} \int_0^{\infty} H_g(x)dx.$$

Plan

Estimation statistique.

•On reparamétrise $\pi(.)$ par la transformation logarithmique $[0,+\infty) \leftrightarrow (0,1]$:

$$x \rightsquigarrow \pi(x) \;\leftrightarrow\; a \rightsquigarrow \beta(a) := a^{-1}\pi(-\log a), \quad a \in (0,1).$$

• **Estimation semi-paramétrique**

$$m_k(\pi) := \int_0^{+\infty} x^k\,\pi(x)dx = \int_0^1 \log(1/a)^k \beta(a)da$$

• **Estimation non-paramétrique**

$$(\beta(a), a \in (0,1)).$$

Estimation paramétrique.

- $f(\cdot) : [0,1] \to \mathbb{R}$ une fonction régulière telle que $f(1) = 0$, soit $g(a) := -af'(a)$.

$$\begin{aligned}\mathcal{E}(g) &= \frac{1}{c(\pi)} \int_0^1 \frac{g(a)}{a} \pi(-\log a, +\infty)\, da \\ &= -\frac{1}{m_1(\pi)} \int_0^1 f'(a) \int_0^a \beta(u) du\, da \\ &= \frac{1}{m_1(\pi)} \int_0^1 f(a)\beta(a) da.\end{aligned}$$

- On va choisir f_ε qui va approximer 1 et soit $g_\varepsilon(a) := -af'_\varepsilon(a)$. On définit:

$$\widehat{m}_{1,\varepsilon} := \frac{1}{\mathcal{E}_\varepsilon(g_\varepsilon)}.$$

On suppose que $\pi(dx) = \pi(x)dx$ est absolument continue, que $x^{1-\lambda}\pi(x)$ reste borné en l'origine et $\int_{[0,+\infty)} e^{\kappa x}\,\pi(dx) < +\infty$.

Théorème

- *Pour* $1 \leq \mu < \kappa$, *la famille*

$$\left(\sqrt{\varepsilon}^{\frac{\mu}{(\mu+1)}\,\frac{2\lambda}{(2\lambda+1)}}\right)^{-1}\left(\widehat{m}_{1,\varepsilon} - m_1(\pi)\right)$$

est $\mathbb{P}_1$ *tendue.*

On suppose que $\pi(dx) = \pi(x)dx$ est absolument continue, que $x^{1-\lambda}\pi(x)$ reste borné en l'origine et $\int_{[0,+\infty)} e^{\kappa x}\,\pi(dx) < +\infty$.

Théorème

- *Pour* $1 \leq \mu < \kappa$, *la famille*

$$\Big(\sqrt{\varepsilon}^{\frac{\mu}{(\mu+1)}\frac{2\lambda}{(2\lambda+1)}}\Big)^{-1}\big(\widehat{m}_{1,\varepsilon} - m_1(\pi)\big)$$

est $\mathbb{P}_1$ *tendue.*

- *De même pour estimer* $m_{k,\varepsilon}$ *par* $\widehat{m}_{k,\varepsilon} := \mathcal{E}_\varepsilon(-\cdot[f_\varepsilon(1-\cdot)\log(1/\cdot)^k]')/\widehat{m}_{1,\varepsilon}$.

On suppose que $\pi(dx) = \pi(x)dx$ est absolument continue, que $x^{1-\lambda}\pi(x)$ reste borné en l'origine et $\int_{[0,+\infty)} e^{\kappa x}\,\pi(dx) < +\infty$.

Théorème

• *Pour* $1 \leq \mu < \kappa$, *la famille*

$$\Big(\sqrt{\varepsilon}^{\frac{\mu}{(\mu+1)}\frac{2\lambda}{(2\lambda+1)}}\Big)^{-1}\big(\widehat{m}_{1,\varepsilon} - m_1(\pi)\big)$$

est $\mathbb{P}_1$ *tendue.*
• *De même pour estimer* $m_{k,\varepsilon}$ *par*
$\widehat{m}_{k,\varepsilon} := \mathcal{E}_\varepsilon(-\cdot\,[f_\varepsilon(1-\cdot)\log(1/\cdot)^k]')/\widehat{m}_{1,\varepsilon}$.
• *Le même résultat reste vrai lorsque l'on remplace* $\mathcal{E}_\varepsilon$ *par* $\mathcal{E}_{\varepsilon,\sigma}$ *dès que* $\sigma\varepsilon^{-3}$ *reste borné.*

Estimation non paramétrique.

• Si $g(a) := -af'(a)$ avec $f(1) = 0$, alors

$$\mathcal{E}(g) = \frac{1}{m_1(\pi)} \int_0^1 f(a)\beta(a)da.$$

• On va choisir une fonction noyau $\varphi_{\gamma_\varepsilon,a}(\cdot) = \gamma_\varepsilon \varphi((\cdot - a)/\gamma_\varepsilon)$ ayant suffisament de régularité et s'annulant de manière propice.
• On définit

$$\widehat{\beta}_\varepsilon(a) := \widehat{m}_{1,\varepsilon}\, \mathcal{E}_{\varepsilon,\sigma}\big(- \cdot \varphi'_{\gamma_\varepsilon,a}(\cdot)\big) \;\; a \in (0,1).$$

On suppose que π vérifie les mêmes hypothèses que pour le théorème précédent, et que $\beta(\cdot)$ est höldérienne d'indice de régularité s.

Théorème

- *Pour* $1 \leq \mu < \kappa/2$, *la famille*

$$\left(\sqrt{\varepsilon}^{\frac{\mu}{(\mu+1)}\frac{2s}{(2s+3)}}\right)^{-1}\left(\widehat{\beta}_{\varepsilon}(a) - \beta(a)\right)$$

est tendue sous $\mathbb{P}_1$, *pour* $s < \kappa - 1$.

On suppose que π vérifie les mêmes hypothèses que pour le théorème précédent, et que $\beta(\cdot)$ est höldérienne d'indice de régularité s.

Théorème

• *Pour* $1 \leq \mu < \kappa/2$, *la famille*

$$\left(\sqrt{\varepsilon}^{\frac{\mu}{(\mu+1)}\frac{2s}{(2s+3)}}\right)^{-1}\left(\widehat{\beta}_\varepsilon(a) - \beta(a)\right)$$

est tendue sous $\mathbb{P}_1$, *pour* $s < \kappa - 1$.
• *Le même résultat reste vrai lorsque l'on remplace* $\mathcal{E}_\varepsilon$ *par* $\mathcal{E}_{\varepsilon,\sigma}$ *dès que* $\sigma\varepsilon^{-3}$ *reste borné.*

Plan

1. Préliminaires.
   - Définition d'une fragmentation.
   - Ligne d'arrêt.
   - L'observable.
   - Mesure empirique.
   - Stratégie.
2. Convergence de la mesure empirique.
   - Une vitesse.
   - Une idée de la preuve.
3. Estimation statistique.
   - Estimation paramétrique.
   - Estimation non paramétrique.
4. La bibliographie.

La bibliographie.

- J. Bertoin (2006). *Random fragmentation and coagulation processes*. Cambridge Univ. Pr.
- J. Bertoin and S. Martinez (2005). Fragmentation energy. *Adv. Appl. Probab.* **37** 553-570.
- M. Hoffmann and N. Krell (2008). Self-similar branching Markov chains. *Soumis*
- M. S. Sgibnev (2002). Stone's decomposition of the renewal measure via Banach-algebraic techniques. *Proc. of the American Math. Soc.*, **130**, 2425–2430.