

Dépendance faible et études de quelques paramètres sur les arbres aléatoires et de quelques exemples de processus de fragmentation

RAFIK AGUECH
Faculté des Sciences de Monastir

- 1 Poursuite adaptative.
- 2 Version du Théorème limite centrale.
- 3 Distance entre noeuds sur deux types d'arbres binaires.
- 4 Poids d'un parcours dans ABR.
- 5 Urnes et Processus de Polya.
- 6 Processus de Fragmentation.
- 7 Modèle de Moran.
- 8 Perspectives.

Traîtement de signal

Présentation du problème :

$$\theta_{n+1} = \theta_n + w_n$$

w_n est un bruit.

$$Y_{n+1} = \varphi_n^T \theta_n + v_{n+1}, \quad n \in \mathbb{N}.$$

Y_{n+1} : mesure faite à l'instant $n + 1$ connue

φ_n : suite de vecteur aléatoire connue

v_n : le bruit d'observation supposé indépendant de φ_n

Présentation du problème :

$$\theta_{n+1} = \theta_n + w_n$$

w_n est un bruit.

$$Y_{n+1} = \varphi_n^T \theta_n + v_{n+1}, \quad n \in \mathbb{N}.$$

Y_{n+1} : mesure faite à l'instant $n + 1$ connue

φ_n : suite de vecteur aléatoire connue

v_n : le bruit d'observation supposé indépendant de φ_n

Présentation du problème :

$$\theta_{n+1} = \theta_n + w_n$$

w_n est un bruit.

$$Y_{n+1} = \varphi_n^T \theta_n + v_{n+1}, \quad n \in \mathbb{N}.$$

Y_{n+1} : mesure faite à l'instant $n + 1$ connue

φ_n : suite de vecteur aléatoire connue

v_n : le bruit d'observation supposé indépendant de φ_n

Présentation du problème :

$$\theta_{n+1} = \theta_n + w_n$$

w_n est un bruit.

$$Y_{n+1} = \varphi_n^T \theta_n + v_{n+1}, \quad n \in \mathbb{N}.$$

Y_{n+1} : mesure faite à l'instant $n + 1$ connue

φ_n : suite de vecteur aléatoire connue

v_n : le bruit d'observation supposé indépendant de φ_n

Algorithme de poursuite :

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \mu L_n \varphi_n (Y_{n+1} - \varphi_n^T \hat{\theta}_n)$$

L_n suite de matrices aléatoires $\sigma(Y_i, \varphi_i, i \leq n)$ -mesurable.

μ est le pas d'adaptation connu déterministe.

Algorithme de poursuite :

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \mu L_n \varphi_n (Y_{n+1} - \varphi_n^T \hat{\theta}_n)$$

L_n suite de matrices aléatoires $\sigma(Y_i, \varphi_i, i \leq n)$ -mesurable.

μ est le pas d'adaptation connu déterministe.

Erreur de poursuite : Erreur à l'instant n : $\tilde{\theta}_n = \hat{\theta}_n - \theta_n$

Cette erreur vérifie l'équation générique suivante :

$$\eta_{n+1} = (I - \mu F_n) \eta_n + \xi_{n+1}$$

F_n une matrice carrée qui est aléatoire.

ξ_{n+1} est un bruit.

Problème 1 : Traitement de signal

Erreur de poursuite : Erreur à l'instant n : $\tilde{\theta}_n = \hat{\theta}_n - \theta_n$

Cette erreur vérifie l'équation générique suivante :

$$\eta_{n+1} = \left(I - \mu F_n \right) \eta_n + \xi_{n+1}$$

F_n une matrice carrée qui est aléatoire.

ξ_{n+1} est un bruit.

Problème 1 : Traitement de signal

Erreur de poursuite : Erreur à l'instant n : $\tilde{\theta}_n = \hat{\theta}_n - \theta_n$

Cette erreur vérifie l'équation générique suivante :

$$\eta_{n+1} = (I - \mu F_n) \eta_n + \xi_{n+1}$$

F_n une matrice carrée qui est aléatoire.

ξ_{n+1} est un bruit.

Problème 1 : Traitement de signal

Développement en perturbation :

$(I - \mu F_n) \longrightarrow (I - \mu \bar{F}_n)$, \bar{F}_n est une matrice carrée déterministe.

$$\begin{aligned}\eta_n &= J_n^0 + H_n^0, \\ J_{n+1}^0 &= (I - \mu \bar{F}_n) J_n^0 + \xi_n, \\ H_{n+1}^0 &= (I - \mu F_n) H_n^0 + \mu(F_n - \bar{F}_n) J_n^0\end{aligned}$$

m itérations, on obtient :

$$\eta_n = J_n^0 + J_n^1 + \dots + J_n^{(m)} + H_n^{(m)},$$

$$\begin{aligned}J_{n+1}^{(r)} &= (I - \mu \bar{F}_n) J_n^{(r)} + \mu(F_n - \bar{F}_n) J_n^{(r-1)}, \quad 0 \leq r \leq m \\ H_{n+1}^{(m)} &= (I - \mu F_n) H_n^{(m)} + \mu Z_n J_n^{(m)}\end{aligned}$$

Problème 1 : Traitement de signal

Développement en perturbation :

$(I - \mu F_n) \longrightarrow (I - \mu \bar{F}_n)$, \bar{F}_n est une matrice carrée déterministe.

$$\begin{aligned}\eta_n &= J_n^0 + H_n^0, \\ J_{n+1}^0 &= (I - \mu \bar{F}_n) J_n^0 + \xi_n, \\ H_{n+1}^0 &= (I - \mu F_n) H_n^0 + \mu(F_n - \bar{F}_n) J_n^0\end{aligned}$$

m itérations, on obtient :

$$\eta_n = J_n^0 + J_n^1 + \dots + J_n^{(m)} + H_n^{(m)},$$

$$\begin{aligned}J_{n+1}^{(r)} &= (I - \mu \bar{F}_n) J_n^{(r)} + \mu(F_n - \bar{F}_n) J_n^{(r-1)}, \quad 0 \leq r \leq m \\ H_{n+1}^{(m)} &= (I - \mu F_n) H_n^{(m)} + \mu Z_n J_n^{(m)}\end{aligned}$$

Développement en perturbation :

$(I - \mu F_n) \longrightarrow (I - \mu \bar{F}_n)$, \bar{F}_n est une matrice carrée déterministe.

$$\begin{aligned}\eta_n &= J_n^0 + H_n^0, \\ J_{n+1}^0 &= (I - \mu \bar{F}_n) J_n^0 + \xi_n, \\ H_{n+1}^0 &= (I - \mu F_n) H_n^0 + \mu(F_n - \bar{F}_n) J_n^0\end{aligned}$$

m itérations, on obtient :

$$\eta_n = J_n^0 + J_n^1 + \dots + J_n^{(m)} + H_n^{(m)},$$

$$\begin{aligned}J_{n+1}^{(r)} &= (I - \mu \bar{F}_n) J_n^{(r)} + \mu(F_n - \bar{F}_n) J_n^{(r-1)}, \quad 0 \leq r \leq m \\ H_{n+1}^{(m)} &= (I - \mu F_n) H_n^{(m)} + \mu Z_n J_n^{(m)}\end{aligned}$$

Développement en perturbation :

$$(I - \mu F_n) \longrightarrow (I - \mu \bar{F}_n), \bar{F}_n \text{ est une matrice carrée déterministe.}$$

$$\begin{aligned}\eta_n &= J_n^0 + H_n^0, \\ J_{n+1}^0 &= (I - \mu \bar{F}_n) J_n^0 + \xi_n, \\ H_{n+1}^0 &= (I - \mu F_n) H_n^0 + \mu(F_n - \bar{F}_n) J_n^0\end{aligned}$$

m itérations, on obtient :

$$\eta_n = J_n^0 + J_n^1 + \dots + J_n^{(m)} + H_n^{(m)},$$

$$\begin{aligned}J_{n+1}^{(r)} &= (I - \mu \bar{F}_n) J_n^{(r)} + \mu(F_n - \bar{F}_n) J_n^{(r-1)}, 0 \leq r \leq m \\ H_{n+1}^{(m)} &= (I - \mu F_n) H_n^{(m)} + \mu Z_n J_n^{(m)}\end{aligned}$$

Développement en perturbation :

$(I - \mu F_n) \longrightarrow (I - \mu \bar{F}_n)$, \bar{F}_n est une matrice carrée déterministe.

$$\begin{aligned}\eta_n &= J_n^0 + H_n^0, \\ J_{n+1}^0 &= (I - \mu \bar{F}_n) J_n^0 + \xi_n, \\ H_{n+1}^0 &= (I - \mu F_n) H_n^0 + \mu(F_n - \bar{F}_n) J_n^0\end{aligned}$$

m itérations, on obtient :

$$\eta_n = J_n^0 + J_n^1 + \dots + J_n^{(m)} + H_n^{(m)},$$

$$\begin{aligned}J_{n+1}^{(r)} &= (I - \mu \bar{F}_n) J_n^{(r)} + \mu(F_n - \bar{F}_n) J_n^{(r-1)}, \quad 0 \leq r \leq m \\ H_{n+1}^{(m)} &= (I - \mu F_n) H_n^{(m)} + \mu Z_n J_n^{(m)}\end{aligned}$$

Développement en perturbation :

$(I - \mu F_n) \longrightarrow (I - \mu \bar{F}_n)$, \bar{F}_n est une matrice carrée déterministe.

$$\begin{aligned}\eta_n &= J_n^0 + H_n^0, \\ J_{n+1}^0 &= (I - \mu \bar{F}_n) J_n^0 + \xi_n, \\ H_{n+1}^0 &= (I - \mu F_n) H_n^0 + \mu(F_n - \bar{F}_n) J_n^0\end{aligned}$$

m itérations, on obtient :

$$\eta_n = J_n^0 + J_n^1 + \dots + J_n^{(m)} + H_n^{(m)},$$

$$\begin{aligned}J_{n+1}^{(r)} &= (I - \mu \bar{F}_n) J_n^{(r)} + \mu(F_n - \bar{F}_n) J_n^{(r-1)}, \quad 0 \leq r \leq m \\ H_{n+1}^{(m)} &= (I - \mu F_n) H_n^{(m)} + \mu Z_n J_n^{(m)}\end{aligned}$$

Problème 1 : Traitement de signal

Résultats : Sous des conditions sur les processus $(F_n)_n$ et $(\xi_n)_n$ (des conditions du type mélange fort et stabilité exponentielle)

Théorème (Aguéch, Moulines, Priouret 2001)

Il existe $p \geq 2$ et une constante $C > 0$ tels que

$$\sup_{s \geq 1} \|J_s^{(r)}\|_p \leq C\mu^{(r-1)/2}$$

$$\sup_{s \geq 0} \|H_s^{(m)}\|_p \leq C\mu^{m/2}$$

Résultats : Sous des conditions sur les processus $(F_n)_n$ et $(\xi_n)_n$ (des conditions du type mélange fort et stabilité exponentielle)

Théorème (Aguech, Moulines, Priouret 2001)

Il existe $p \geq 2$ et une constante $C > 0$ tels que

$$\sup_{s \geq 1} \|J_s^{(r)}\|_p \leq C\mu^{(r-1)/2}$$

$$\sup_{s \geq 0} \|H_s^{(m)}\|_p \leq C\mu^{m/2}$$

Résultats : Sous des conditions sur les processus $(F_n)_n$ et $(\xi_n)_n$ (des conditions du type mélange fort et stabilité exponentielle)

Théorème (Aguech, Moulines, Priouret 2001)

Il existe $p \geq 2$ et une constante $C > 0$ tels que

$$\sup_{s \geq 1} \|J_s^{(r)}\|_p \leq C\mu^{(r-1)/2}$$

$$\sup_{s \geq 0} \|H_s^{(m)}\|_p \leq C\mu^{m/2}$$

Une version du Théorème limite centrale

Problème 2 : Une version du Théorème limite centrale

Problème : Etant donné une suite de variables aléatoires stationnaires triangulaires $(X_{i,n})_{0 \leq i \leq n}$ centrées et dans L^2 . Construire une suite de v.a. $(\tilde{X}_{i,n})_{0 \leq i \leq n}$ telle que

$S_n = \sum_{i=1}^n X_{i,n}$ et $\tilde{S}_n = \sum_{i=1}^n \tilde{X}_{i,n}$ aient asymptotiquement la même loi.

La suite $(\tilde{X}_{i,n})_{0 \leq i \leq n}$ sera appelée "asymptotic independant representation" notée : a.i.r.

Définition

Un vecteur aléatoire $X = (X_1, \dots, X_n)$ est dit associé si, pour toutes fonctions f et g définies sur \mathbb{R}^n et croissantes par composantes, on a

$$\text{Cov}(f(X), g(X)) \geq 0.$$

Problème 2 : Une version du Théorème limite centrale

Problème : Etant donné une suite de variables aléatoires stationnaires triangulaires $(X_{i,n})_{0 \leq i \leq n}$ centrées et dans L^2 . Construire une suite de v.a. $(\tilde{X}_{i,n})_{0 \leq i \leq n}$ telle que

$S_n = \sum_{i=1}^n X_{i,n}$ et $\tilde{S}_n = \sum_{i=1}^n \tilde{X}_{i,n}$ aient asymptotiquement la même loi.

La suite $(\tilde{X}_{i,n})_{0 \leq i \leq n}$ sera appelée "asymptotic independant representation" notée : a.i.r.

Définition

Un vecteur aléatoire $X = (X_1, \dots, X_n)$ est dit associé si, pour toutes fonctions f et g définies sur \mathbb{R}^n et croissantes par composantes, on a

$$\text{Cov}(f(X), g(X)) \geq 0.$$

Problème 2 : Une version du Théorème limite centrale

Problème : Etant donné une suite de variables aléatoires stationnaires triangulaires $(X_{i,n})_{0 \leq i \leq n}$ centrées et dans L^2 . Construire une suite de v.a. $(\tilde{X}_{i,n})_{0 \leq i \leq n}$ telle que

$S_n = \sum_{i=1}^n X_{i,n}$ et $\tilde{S}_n = \sum_{i=1}^n \tilde{X}_{i,n}$ aient asymptotiquement la même loi.

La suite $(\tilde{X}_{i,n})_{0 \leq i \leq n}$ sera appelée "asymptotic independant representation" notée : a.i.r.

Définition

Un vecteur aléatoire $X = (X_1, \dots, X_n)$ est dit associé si, pour toutes fonctions f et g définies sur \mathbb{R}^n et croissantes par composantes, on a

$$\text{Cov}(f(X), g(X)) \geq 0.$$

Problème 2 : Une version du Théorème limite centrale

Problème : Etant donné une suite de variables aléatoires stationnaires triangulaires $(X_{i,n})_{0 \leq i \leq n}$ centrées et dans L^2 . Construire une suite de v.a. $(\tilde{X}_{i,n})_{0 \leq i \leq n}$ telle que

$S_n = \sum_{i=1}^n X_{i,n}$ et $\tilde{S}_n = \sum_{i=1}^n \tilde{X}_{i,n}$ aient asymptotiquement la même loi.

La suite $(\tilde{X}_{i,n})_{0 \leq i \leq n}$ sera appelée "asymptotic independant representation" notée : a.i.r.

Définition

Un vecteur aléatoire $X = (X_1, \dots, X_n)$ est dit associé si, pour toutes fonctions f et g définies sur \mathbb{R}^n et croissantes par composantes, on a

$$\text{Cov}(f(X), g(X)) \geq 0.$$

Problème 2 : Une version du Théorème limite centrale

Inégalité de Newman : $(X_{i,n})_{0 \leq i \leq n}$ est associé, on a

$$\left| \mathbf{E} \left(\exp(itS_n) \right) - \prod_{j=1}^n \left(\mathbf{E}(itX_{j,n}) \right) \right| \leq \frac{t^2}{2} \left(\mathbf{V}S_n - n\mathbf{V}X_{0,n} \right)$$

Lemme

$(\tilde{X}_{i,n})_{0 \leq i \leq n}$ i.i.d de même loi que $(X_{0,n})$.

sous la condition

$$\lim_n n \sum_{r=1}^n \mathbf{Cov}(X_{0,n}, X_{r,n}) = 0 \quad (1)$$

la suite $(\tilde{X}_{i,n})_{0 \leq i \leq n}$ est une a.i.r de $(X_{i,n})_{0 \leq i \leq n}$

Problème 2 : Une version du Théorème limite centrale

Inégalité de Newman : $(X_{i,n})_{0 \leq i \leq n}$ est associé, on a

$$\left| \mathbf{E} \left(\exp(itS_n) \right) - \prod_{j=1}^n \left(\mathbf{E}(\exp(itX_{j,n})) \right) \right| \leq \frac{t^2}{2} \left(\mathbf{V}S_n - n\mathbf{V}X_{0,n} \right)$$

Lemme

$(\tilde{X}_{i,n})_{0 \leq i \leq n}$ i.i.d de même loi que $(X_{0,n})$.

sous la condition

$$\lim_n n \sum_{r=1}^n \mathbf{Cov}(X_{0,n}, X_{r,n}) = 0 \quad (1)$$

la suite $(\tilde{X}_{i,n})_{0 \leq i \leq n}$ est une a.i.r de $(X_{i,n})_{0 \leq i \leq n}$

Problème 2 : Une version du Théorème limite centrale

Coefficient de Cox-Grimmett :

$$R(N) = \limsup_n \frac{1}{n} \sum_{r=N}^n \mathbf{Cov}(X_{0,n}, X_{r,n}),$$

Soit $N_0 := \inf\{N, R(N) = 0\} \in [0, +\infty]$.

On considère la famille suivante :

$$\mathcal{F}_{0,C} = \left\{ x \mapsto \cos(zx) - 1 + \frac{z^2 x^2}{2}, x \mapsto \sin(zx) - zx, z \in \mathbb{R} \right\}.$$

Problème 2 : Une version du Théorème limite centrale

Coefficient de Cox-Grimmett :

$$R(N) = \limsup_n \frac{1}{n} \sum_{r=N}^n \mathbf{Cov}(X_{0,n}, X_{r,n}),$$

Soit $N_0 := \inf\{N, R(N) = 0\} \in [0, +\infty]$.

On considère la famille suivante :

$$\mathcal{F}_{0,C} = \left\{ x \mapsto \cos(zx) - 1 + \frac{z^2 x^2}{2}, x \mapsto \sin(zx) - zx, z \in \mathbb{R} \right\}.$$

Coefficient de Cox-Grimmett :

$$R(N) = \limsup_n \frac{1}{n} \sum_{r=N}^n \mathbf{Cov}(X_{0,n}, X_{r,n}),$$

Soit $N_0 := \inf\{N, R(N) = 0\} \in [0, +\infty]$.

On considère la famille suivante :

$$\mathcal{F}_{0,C} = \left\{ x \mapsto \cos(zx) - 1 + \frac{z^2 x^2}{2}, x \mapsto \sin(zx) - zx, z \in \mathbb{R} \right\}.$$

Problème 2 : Une version du Théorème limite centrale

Décomposition de S_n : Pour $\eta > 0$, on définit les deux processus suivants :

$$\begin{aligned}X_{i,n}(\eta) &= f_\eta(X_{i,n}) - \mathbf{E}\left(f_\eta(X_{i,n})\right), \\X_{i,n}^{(\eta)} &= X_{i,n} - X_{i,n}(\eta),\end{aligned}$$

avec $f_\eta(x) = (x \wedge \eta) \vee (-\eta)$.

On définit aussi

$$S_n^{(\eta)} = \sum_{i=1}^n X_{i,n}^{(\eta)}, \quad S_n(\eta) = \sum_{i=1}^n X_{i,n}(\eta).$$

On a bien $S_n = S_n^{(\eta)} + S_n(\eta)$.

Décomposition de S_n : Pour $\eta > 0$, on définit les deux processus suivants :

$$\begin{aligned}X_{i,n}(\eta) &= f_\eta(X_{i,n}) - \mathbf{E}\left(f_\eta(X_{i,n})\right), \\X_{i,n}^{(\eta)} &= X_{i,n} - X_{i,n}(\eta),\end{aligned}$$

avec $f_\eta(x) = (x \wedge \eta) \vee (-\eta)$.

On définit aussi

$$S_n^{(\eta)} = \sum_{i=1}^n X_{i,n}^{(\eta)}, \quad S_n(\eta) = \sum_{i=1}^n X_{i,n}(\eta).$$

On a bien $S_n = S_n^{(\eta)} + S_n(\eta)$.

Problème 2 : Une version du Théorème limite centrale

Résultats : $S_{N,n}^{(\eta)} := \sum_{i=1}^N X_{i,n}^{(\eta)}$

Théorème (Aguech, Louhichi, Louhichi (2005))

Soient $(X_{i,n})_{0 \leq i \leq n}$ une suite triangulaire de variables aléatoires stationnaires, centrées, dans L^2 et associée et satisfait $\sup_n \mathbf{E}(X_{0,n}^2) < \infty$ et $(\tilde{X}_{i,n})$ une suite i.i.d telles que

$$\lim_{N \rightarrow N_0} \limsup_{\eta \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{h \in \mathcal{F}_{0,C}} n \left| \mathbf{E}h(\tilde{X}_{0,n}^{(\eta)}) - \mathbf{E} \left[h(S_{N,n}^{(\eta)}) - h(S_{N-1,n}^{(\eta)}) \right] \right| = 0. \quad (2)$$

Sous les hypothèses

$$\lim_{n \rightarrow \infty} n \mathbf{V} \tilde{X}_{0,n} = \sigma^2$$

et

$$\lim_{N \rightarrow N_0} R(N) = 0, \quad \lim_{n \rightarrow \infty} \mathbf{V} S_n = \sigma^2,$$

$(\tilde{X}_{i,n})$ est une a.i.r de $(X_{i,n})$.

TLC :

Corollaire

Soit $(X_{i,n})_{i,n \in \mathbb{N}}$ une suite triangulaire de variables aléatoires stationnaires, associée de L^2 et centrées. On suppose que

$$\lim_{n \rightarrow \infty} \mathbf{E}[X_{0,n}^2] = 0,$$

que

$$\lim_{N \rightarrow N_0} R(N) = 0, \quad \lim_{n \rightarrow \infty} \mathbf{V}[S_n] = 1,$$

et que

$$\lim_{n \rightarrow \infty} n \mathbf{E} \left[X_{0,n}^2 \mathbb{I}_{\{|X_{0,n}| \geq \varepsilon\}} \right] = 0, \quad \forall \varepsilon > 0,$$

alors

$$S_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Distance sur les arbres digitaux

Présentation :

Définition

- *Les arbres digitaux de recherche [DST]*
- *Les tries*

Distance entre deux noeuds : Soit Δ_n la variable aléatoire représentant la distance entre deux noeuds choisis au hasard dans un arbre digital de taille n : c'est le nombre d'arêtes entre les deux noeuds.

Distance sur DST :

Présentation :

Définition

- *Les arbres digitaux de recherche [DST]*
- *Les tries*

Distance entre deux noeuds : Soit Δ_n la variable aléatoire représentant la distance entre deux noeuds choisis au hasard dans un arbre digital de taille n : c'est le nombre d'arêtes entre les deux noeuds.

Distance sur DST :

Présentation :

Définition

- *Les arbres digitaux de recherche [DST]*
- *Les tries*

Distance entre deux noeuds : Soit Δ_n la variable aléatoire représentant la distance entre deux noeuds choisis au hasard dans un arbre digital de taille n : c'est le nombre d'arêtes entre les deux noeuds.

Distance sur DST :

Présentation :

Définition

- *Les arbres digitaux de recherche [DST]*
- *Les tries*

Distance entre deux noeuds : Soit Δ_n la variable aléatoire représentant la distance entre deux noeuds choisis au hasard dans un arbre digital de taille n : c'est le nombre d'arêtes entre les deux noeuds.

Distance sur DST :

Présentation :

Définition

- *Les arbres digitaux de recherche [DST]*
- *Les tries*

Distance entre deux noeuds : Soit Δ_n la variable aléatoire représentant la distance entre deux noeuds choisis au hasard dans un arbre digital de taille n : c'est le nombre d'arêtes entre les deux noeuds.

Distance sur DST :

Problème 3 : Distance sur les arbres digitaux

Expression conditionnelle de $\Delta_n(\text{DST})$: R_n (resp. L_n) le nombre de noeuds dans le sous arbre droit (resp. gauche). On a bien $L_n + R_n + 1 = n$. Etant donné R_n on a

$$\Delta_n/R_n = \begin{cases} \Delta_{L_n}, & \text{avec proba } \binom{L_n}{2} / \binom{n}{2}, \\ \tilde{\Delta}_{R_n}, & \text{avec proba } \binom{R_n}{2} / \binom{n}{2}, \\ (\delta_{L_n} + 1) + (\tilde{\delta}_{R_n} + 1), & \text{avec proba } L_n R_n / \binom{n}{2}, \\ \delta_{L_n} + 1, & \text{avec proba } L_n / \binom{n}{2}, \\ \tilde{\delta}_{R_n} + 1, & \text{avec proba } R_n / \binom{n}{2}, \end{cases}$$

Problème 3 : Distance sur les arbres digitaux

Expression conditionnelle de $\Delta_n(\text{DST})$: R_n (resp. L_n) le nombre de noeuds dans le sous arbre droit (resp. gauche). On a bien $L_n + R_n + 1 = n$. Etant donné R_n on a

$$\Delta_n/R_n = \begin{cases} \Delta_{L_n}, & \text{avec proba } \binom{L_n}{2} / \binom{n}{2}, \\ \tilde{\Delta}_{R_n}, & \text{avec proba } \binom{R_n}{2} / \binom{n}{2}, \\ (\delta_{L_n} + 1) + (\tilde{\delta}_{R_n} + 1), & \text{avec proba } L_n R_n / \binom{n}{2}, \\ \delta_{L_n} + 1, & \text{avec proba } L_n / \binom{n}{2}, \\ \tilde{\delta}_{R_n} + 1, & \text{avec proba } R_n / \binom{n}{2}, \end{cases}$$

Problème 3 : Distance sur les arbres digitaux

Transformée de Laplace de $\Delta_n(\text{DST})$: $\phi_X(t) := \mathbf{E}[\exp(tX)]$,

$$\begin{aligned} \binom{n}{2} \phi_{\Delta_n}(t) &= \sum_{l=0}^{n-1} \binom{l}{2} \phi_{\Delta_l}(t) \binom{n-1}{l} p^{n-1-l} q^l \\ &+ \sum_{r=0}^{n-1} \binom{r}{2} \phi_{\Delta_r}(t) \binom{n-1}{r} p^r q^{n-1-r} \\ &+ e^{2t} \sum_{l=0}^{n-1} \binom{n-1}{l} l(n-1-l) \phi_{\delta_l}(t) \phi_{\delta_{n-1-l}}(t) p^{n-1-l} q^l \\ &+ e^t \sum_{l=0}^{n-1} l \binom{n-1}{l} \phi_{\delta_l}(t) p^{n-1-l} q^l \\ &+ e^t \sum_{r=0}^{n-1} r \binom{n-1}{r} \phi_{\delta_r}(t) q^{n-1-r} p^r. \end{aligned} \quad (3)$$

Problème 3 : Distance sur les arbres digitaux

Transformée de Laplace de $\Delta_n(\text{DST})$: $\phi_X(t) := \mathbf{E}[\exp(tX)]$,

$$\begin{aligned} \binom{n}{2} \phi_{\Delta_n}(t) &= \sum_{l=0}^{n-1} \binom{l}{2} \phi_{\Delta_l}(t) \binom{n-1}{l} p^{n-1-l} q^l \\ &+ \sum_{r=0}^{n-1} \binom{r}{2} \phi_{\Delta_r}(t) \binom{n-1}{r} p^r q^{n-1-r} \\ &+ e^{2t} \sum_{l=0}^{n-1} \binom{n-1}{l} l(n-1-l) \phi_{\delta_l}(t) \phi_{\delta_{n-1-l}}(t) p^{n-1-l} q^l \\ &+ e^t \sum_{l=0}^{n-1} l \binom{n-1}{l} \phi_{\delta_l}(t) p^{n-1-l} q^l \\ &+ e^t \sum_{l=0}^{n-1} r \binom{n-1}{r} \phi_{\delta_r}(t) q^{n-1-r} p^r. \end{aligned} \quad (3)$$

Problème 3 : Distance sur les arbres digitaux

Poissonisation :

Définition

Soient $(g_n)_n$ une suite réelle (ou complexe) et N une variable aléatoire de loi de Poisson de paramètre $z \geq 0$. La transformée de poisson de $(g_n)_n$ est la fonction $G(z)$ définie par

$$G(z) = \mathbf{E}[g_N] = \sum_{n \geq 0} g_n \frac{z^n}{n!} e^{-z}.$$

On introduit $\Psi(t, z)$, la transformée de poisson de la suite $\left(\binom{n}{2} \phi_{\Delta_n}(t)\right)_n$

$$\Psi(t, z) = e^{-z} \Phi(t, z) = \mathbf{E} \left[\binom{N(z)}{2} \exp(\Delta_{N(z)} t) \right].$$

Equation (3) \implies

$$\frac{d}{dz} \Psi(t, z) + \Psi(t, z) = \Psi(t, pz) + \Psi(t, qz) + e^{2t} \psi(t, pz) \psi(t, qz) \quad (4)$$

Problème 3 : Distance sur les arbres digitaux

Poissonisation :

Définition

Soient $(g_n)_n$ une suite réelle (ou complexe) et N une variable aléatoire de loi de Poisson de paramètre $z \geq 0$. La transformée de poisson de $(g_n)_n$ est la fonction $G(z)$ définie par

$$G(z) = \mathbf{E}[g_N] = \sum_{n \geq 0} g_n \frac{z^n}{n!} e^{-z}.$$

On introduit $\Psi(t, z)$, la transformée de poisson de la suite $(\binom{n}{2} \phi_{\Delta_n}(t))_n$

$$\Psi(t, z) = e^{-z} \Phi(t, z) = \mathbf{E} \left[\binom{N(z)}{2} \exp(\Delta_{N(z)} t) \right].$$

Equation (3) \implies

$$\frac{d}{dz} \Psi(t, z) + \Psi(t, z) = \Psi(t, pz) + \Psi(t, qz) + e^{2t} \psi(t, pz) \psi(t, qz) \quad (4)$$

Problème 3 : Distance sur les arbres digitaux

Poissonisation :

Définition

Soient $(g_n)_n$ une suite réelle (ou complexe) et N une variable aléatoire de loi de Poisson de paramètre $z \geq 0$. La transformée de poisson de $(g_n)_n$ est la fonction $G(z)$ définie par

$$G(z) = \mathbf{E}[g_N] = \sum_{n \geq 0} g_n \frac{z^n}{n!} e^{-z}.$$

On introduit $\Psi(t, z)$, la transformée de poisson de la suite $(\binom{n}{2} \phi_{\Delta_n}(t))_n$

$$\Psi(t, z) = e^{-z} \Phi(t, z) = \mathbf{E} \left[\binom{N(z)}{2} \exp(\Delta_{N(z)} t) \right].$$

Equation (3) \implies

$$\frac{d}{dz} \Psi(t, z) + \Psi(t, z) = \Psi(t, pz) + \Psi(t, qz) + e^{2t} \psi(t, pz) \psi(t, qz) \quad (4)$$

Problème 3 : Distance sur les arbres digitaux

La moyenne de $\Delta_n(\text{DST})$: Pour obtenir la moyenne on dérive par rapport à t la fonction $\Psi(t, z)$. Soit

$$X(z) = \frac{d}{dt} \Psi(t, z)|_{t=0} - \frac{z^2}{2} = \mathbf{E} \left[\binom{N(z)}{2} \Delta_{N(z)} \right] - \frac{z^2}{2}.$$

Equation (4) \implies

$$X'(z) + X(z) = X(pz) + X(qz) + pzX(qz) + qzX(pz) = x(pz) + x(qz) + pqz^2.$$

$$x(z) = \sum_{n=1}^{\infty} n \mathbf{E}[\delta_n] z^n e^{-z} / n!$$

Problème 3 : Distance sur les arbres digitaux

La moyenne de Δ_n (DST) : Pour obtenir la moyenne on dérive par rapport à t la fonction $\Psi(t, z)$. Soit

$$X(z) = \frac{d}{dt} \Psi(t, z)|_{t=0} - \frac{z^2}{2} = \mathbf{E} \left[\binom{N(z)}{2} \Delta_{N(z)} \right] - \frac{z^2}{2}.$$

Equation (4) \implies

$$X'(z) + X(z) = X(pz) + X(qz) + pzx(qz) + qzx(pz) = x(pz) + x(qz) + pqz^2.$$

$$x(z) = \sum_{n=1}^{\infty} n \mathbf{E}[\delta_n] z^n e^{-z} / n!$$

Problème 3 : Distance sur les arbres digitaux

La variance de $\Delta_n(\text{DST})$: On pose

$$W(z) = \mathbf{E} \left[\binom{N(z)}{2} \Delta_{N(z)}^2 \right] - \frac{z^2}{2}.$$

On a

$$\begin{aligned} W'(z) + W(z) &= W(pz) + W(qz) \\ &+ 4pzx(qz) + 4qzx(pz) + pzw(qz) + qzw(pz) \\ &+ 2 \left(x(pz) - \frac{p^2 z^2}{2} \right) + 2 \left(x(qz) - \frac{q^2 z^2}{2} \right) \\ &+ \left(w(pz) - \frac{p^2 z^2}{2} \right) + \left(w(qz) - \frac{q^2 z^2}{2} \right) \\ &+ 2x(pz)x(qz) + \frac{3}{2} z^2. \end{aligned}$$

Problème 3 : Distance sur les arbres digitaux

Transformation de Mellin :

Définition

f une fonction réelle, sa transformée de Mellin f^* , lorsque elle existe, est donnée par

$$f^*(s) = \int_0^{+\infty} f(x)x^{s-1}dx.$$

Mellin $\implies X^*(s), W^*(s)$

Théorème (Mellin inverse)

Pour tout c tel que $f^*(c)$ existe, on a

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(s)x^{-s}ds.$$

Mellin inverse $\implies X(z), W(z)$

Problème 3 : Distance sur les arbres digitaux

Transformation de Mellin :

Définition

f une fonction réelle, sa transformée de Mellin f^* , lorsque elle existe, est donnée par

$$f^*(s) = \int_0^{+\infty} f(x)x^{s-1}dx.$$

Mellin $\implies X^*(s), W^*(s)$

Théorème (Mellin inverse)

Pour tout c tel que $f^(c)$ existe, on a*

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(s)x^{-s}ds.$$

Mellin inverse $\implies X(z), W(z)$

Problème 3 : Distance sur les arbres digitaux

Transformation de Mellin :

Définition

f une fonction réelle, sa transformée de Mellin f^* , lorsque elle existe, est donnée par

$$f^*(s) = \int_0^{+\infty} f(x)x^{s-1}dx.$$

Mellin $\implies X^*(s), W^*(s)$

Théorème (Mellin inverse)

Pour tout c tel que $f^*(c)$ existe, on a

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(s)x^{-s}ds.$$

Mellin inverse $\implies X(z), W(z)$

Problème 3 : Distance sur les arbres digitaux

Transformation de Mellin :

Définition

f une fonction réelle, sa transformée de Mellin f^* , lorsque elle existe, est donnée par

$$f^*(s) = \int_0^{+\infty} f(x)x^{s-1}dx.$$

Mellin $\implies X^*(s), W^*(s)$

Théorème (Mellin inverse)

Pour tout c tel que $f^*(c)$ existe, on a

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(s)x^{-s}ds.$$

Mellin inverse $\implies X(z), W(z)$

Dépoissonisation :

Théorème (Jaquet, Szpankowski (1997))

S'il existe des constantes $A, B, R > 0$, $\alpha, \beta < 1$ et $\theta < \pi/2$, telles que sur le cône S_θ , on a les conditions suivantes :

- *Pour tout $z \in S_\theta$ tel que $|z| > R$ on a $|G(z)| \leq B|z|^\beta$,*
- *pour $z \in S_\theta^c$ et $|z| > R$ on a $|G(z)e^z| \leq A \exp(\alpha|z|)$.*

alors,

$$g_n = G(n) + O(n^{\beta-1}).$$

$$\text{Dépoissonisation} \implies E[\Delta_n], E[\Delta_n^2]$$

Dépoissonisation :

Théorème (Jaquet, Szpankowski (1997))

S'il existe des constantes $A, B, R > 0$, $\alpha, \beta < 1$ et $\theta < \pi/2$, telles que sur le cône S_θ , on a les conditions suivantes :

- *Pour tout $z \in S_\theta$ tel que $|z| > R$ on a $|G(z)| \leq B|z|^\beta$,*
- *pour $z \in S_\theta^c$ et $|z| > R$ on $|G(z)e^z| \leq A \exp(\alpha|z|)$.*

alors,

$$g_n = G(n) + O(n^{\beta-1}).$$

$$\text{Dépoissonisation} \implies \mathbf{E}[\Delta_n], \mathbf{E}[\Delta_n^2]$$

Problème 3 : Distance sur les arbres digitaux

Convergence en loi : Méthode de contraction : Pour étudier les lois limites, on utilise une technique développée par Neininger et Rüschendorf (2004), qui s'appelle "méthode de contraction".

Soit $(Y_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires vérifiant

$$Y_n \stackrel{\mathcal{L}}{=} \sum_{r=1}^K A_r(n) Y_{I_r^{(n)}}^{(r)} + b_n, \quad n \geq 0 \quad (5)$$

$(A_1(n), \dots, A_K(n), b_n, I^{(n)}, (Y_n^{(1)}), \dots, (Y_n^{(K)}))$ sont indépendantes.
 $I_r^{(n)} \in \{0, \dots, n\}$ et $Y_n^{(1)}, \dots, Y_n^{(K)}$ sont indépendantes de même loi que (Y_n) .

Problème 3 : Distance sur les arbres digitaux

Convergence en loi : Méthode de contraction : Pour étudier les lois limites, on utilise une technique développée par Neininger et Rüschendorf (2004), qui s'appelle "méthode de contraction".

Soit $(Y_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires vérifiant

$$Y_n \stackrel{\mathcal{L}}{=} \sum_{r=1}^K A_r(n) Y_{I_r^{(n)}}^{(r)} + b_n, \quad n \geq 0 \quad (5)$$

$(A_1(n), \dots, A_K(n), b_n, I^{(n)}, (Y_n^{(1)}), \dots, (Y_n^{(K)}))$ sont indépendantes.
 $I_r^{(n)} \in \{0, \dots, n\}$ et $Y_n^{(1)}, \dots, Y_n^{(K)}$ sont indépendantes de même loi que (Y_n) .

Problème 3 : Distance sur les arbres digitaux

Convergence en loi : Méthode de contraction : Pour étudier les lois limites, on utilise une technique développée par Neininger et Rüschendorf (2004), qui s'appelle "méthode de contraction".

Soit $(Y_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires vérifiant

$$Y_n \stackrel{\mathcal{L}}{=} \sum_{r=1}^K A_r(n) Y_{I_r^{(n)}}^{(r)} + b_n, \quad n \geq 0 \quad (5)$$

$(A_1(n), \dots, A_K(n), b_n, I^{(n)}), (Y_n^{(1)}), \dots, (Y_n^{(K)})$ sont indépendantes.

$I_r^{(n)} \in \{0, \dots, n\}$ et $Y_n^{(1)}, \dots, Y_n^{(K)}$ sont indépendantes de même loi que (Y_n) .

Problème 3 : Distance sur les arbres digitaux

Convergence en loi : Méthode de contraction : Pour étudier les lois limites, on utilise une technique développée par Neininger et Rüschendorf (2004), qui s'appelle "méthode de contraction".

Soit $(Y_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires vérifiant

$$Y_n \stackrel{\mathcal{L}}{=} \sum_{r=1}^K A_r(n) Y_{I_r^{(n)}}^{(r)} + b_n, \quad n \geq 0 \quad (5)$$

$(A_1(n), \dots, A_K(n), b_n, I^{(n)}), (Y_n^{(1)}), \dots, (Y_n^{(K)})$ sont indépendantes.
 $I_r^{(n)} \in \{0, \dots, n\}$ et $Y_n^{(1)}, \dots, Y_n^{(K)}$ sont indépendantes de même loi que (Y_n) .

Problème 3 : Distance sur les arbres digitaux

Méthode de contraction : Normalisation On considère deux suites déterministes convenablement choisies (m_n) et (c_n) (si $Y_n \in L^2$ alors, généralement, on choisit $m_n = \mathbf{E}[Y_n]$, $c_n = \mathbf{V}[Y_n]$)

On définit le nouveau processus (X_n) par

$$X_n = \frac{Y_n - m_n}{\sqrt{c_n}}.$$

en utilisant (5), (X_n) satisfait

$$X_n \stackrel{\mathcal{L}}{=} \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq 0, \quad (6)$$

avec

$$A_r^{(n)} = c_n^{1/2} A_r(n) c_{I_r^{(n)}}^{1/2}, \quad b^{(n)} = c_n^{-1/2} \left(b_n - m_n + \sum_{r=1}^K A_r(n) m_{I_r^{(n)}} \right).$$

Problème 3 : Distance sur les arbres digitaux

Méthode de contraction : Normalisation On considère deux suites déterministes convenablement choisies (m_n) et (c_n) (si $Y_n \in L^2$ alors, généralement, on choisit $m_n = \mathbf{E}[Y_n]$, $c_n = \mathbf{V}[Y_n]$)

On définit le nouveau processus (X_n) par

$$X_n = \frac{Y_n - m_n}{\sqrt{c_n}}.$$

en utilisant (5), (X_n) satisfait

$$X_n \stackrel{\mathcal{L}}{=} \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq 0, \quad (6)$$

avec

$$A_r^{(n)} = c_n^{1/2} A_r(n) c_{I_r^{(n)}}^{1/2}, \quad b^{(n)} = c_n^{-1/2} \left(b_n - m_n + \sum_{r=1}^K A_r(n) m_{I_r^{(n)}} \right).$$

Problème 3 : Distance sur les arbres digitaux

Méthode de contraction : Normalisation On considère deux suites déterministes convenablement choisies (m_n) et (c_n) (si $Y_n \in L^2$ alors, généralement, on choisit $m_n = \mathbf{E}[Y_n]$, $c_n = \mathbf{V}[Y_n]$)

On définit le nouveau processus (X_n) par

$$X_n = \frac{Y_n - m_n}{\sqrt{c_n}}.$$

en utilisant (5), (X_n) satisfait

$$X_n \stackrel{\mathcal{L}}{=} \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq 0, \quad (6)$$

avec

$$A_r^{(n)} = c_n^{1/2} A_r(n) c_{I_r^{(n)}}^{1/2}, \quad b^{(n)} = c_n^{-1/2} \left(b_n - m_n + \sum_{r=1}^K A_r(n) m_{I_r^{(n)}} \right).$$

Problème 3 : Distance sur les arbres digitaux

Méthode de contraction : Normalisation On considère deux suites déterministes convenablement choisies (m_n) et (c_n) (si $Y_n \in L^2$ alors, généralement, on choisit $m_n = \mathbf{E}[Y_n]$, $c_n = \mathbf{V}[Y_n]$)

On définit le nouveau processus (X_n) par

$$X_n = \frac{Y_n - m_n}{\sqrt{c_n}}.$$

en utilisant (5), (X_n) satisfait

$$X_n \stackrel{\mathcal{L}}{=} \sum_{r=1}^K A_r^{(n)} X_{I_r^{(n)}}^{(r)} + b^{(n)}, \quad n \geq 0, \quad (6)$$

avec

$$A_r^{(n)} = c_n^{1/2} A_r(n) c_{I_r^{(n)}}^{1/2}, \quad b^{(n)} = c_n^{-1/2} \left(b_n - m_n + \sum_{r=1}^K A_r(n) m_{I_r^{(n)}} \right).$$

Problème 3 : Distance sur les arbres digitaux

Méthode de contraction : Théorème

Théorème (Théorème de contraction : N., R. (2004))

On suppose que $X_n \in L^3$ et que

$$(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}) \xrightarrow{L^3} (A_1^*, \dots, A_K^*, b^*),$$

$$\sum_{r=1}^K \mathbf{E}[|A_r^*|^3] < 1,$$

$$\mathbf{E} \left[\mathbb{1}_{\{I_r^{(n)} \leq l\} \cup \{I_r^{(n)} = n\}} |A_r^{(n)}|^3 \right] \longrightarrow 0$$

pour tout $l \in \mathbb{N}$ et $r = 1, \dots, K$.

X_n converge en loi vers une variable aléatoire X qui est l'unique loi vérifiant

$$X \stackrel{L}{=} \sum_{r=1}^K A_r^* X^{(r)} + b^*.$$

Problème 3 : Distance sur les arbres digitaux

Méthode de contraction : Théorème

Théorème (Théorème de contraction : N., R. (2004))

On suppose que $X_n \in L^3$ et que

$$(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}) \xrightarrow{L^3} (A_1^*, \dots, A_K^*, b^*),$$

$$\sum_{r=1}^K \mathbf{E}[|A_r^*|^3] < 1,$$

$$\mathbf{E} \left[\mathbb{1}_{\{I_r^{(n)} \leq l\} \cup \{I_r^{(n)} = n\}} |A_r^{(n)}|^3 \right] \longrightarrow 0$$

pour tout $l \in \mathbb{N}$ et $r = 1, \dots, K$.

X_n converge en loi vers une variable aléatoire X qui est l'unique loi vérifiant

$$X \stackrel{L}{=} \sum_{r=1}^K A_r^* X^{(r)} + b^*.$$

Problème 3 : Distance sur les arbres digitaux

Méthode de contraction : Théorème

Théorème (Théorème de contraction : N., R. (2004))

On suppose que $X_n \in L^3$ et que

$$(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}) \xrightarrow{L^3} (A_1^*, \dots, A_K^*, b^*),$$

$$\sum_{r=1}^K \mathbf{E}[|A_r^*|^3] < 1,$$

$$\mathbf{E} \left[\mathbb{1}_{\{I_r^{(n)} \leq l\} \cup \{I_r^{(n)} = n\}} |A_r^{(n)}|^3 \right] \longrightarrow 0$$

pour tout $l \in \mathbb{N}$ et $r = 1, \dots, K$.

X_n converge en loi vers une variable aléatoire X qui est l'unique loi vérifiant

$$X \stackrel{\mathcal{L}}{=} \sum_{r=1}^K A_r^* X^{(r)} + b^*.$$

Problème 3 : Distance sur les arbres digitaux

Résultats : Soit Δ_n la distance entre deux neuds choisis au hasard dans un arbre digital \mathcal{T}_n à n noeuds.

Théorème (Aguech, Lasmar, Mahmoud (2006))

Si \mathcal{T}_n est un arbre digital de recherche on a

$$\frac{\Delta_n - \frac{2}{h_p} \ln n}{\sqrt{\ln n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2\sigma_p^2)$$

$$h_p = -p \ln p - q \ln q, \sigma_p > 0.$$

Théorème (Aguech, Lasmar, Mahmoud (2006))

Si \mathcal{T}_n est un tries

$$\frac{\Delta_n - \frac{2}{h_p} \ln n}{\sqrt{\ln n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2\alpha_p^2)$$

$$h_p = -p \ln p - q \ln q, \alpha_p > 0.$$

Poids sur les arbres binaires de recherches

Présentation du problème : On considère un ABR construit à partir d'une permutation de $\{1, 2, \dots, n\}$.

$W_1(n)$ représente le poids de l'arête tout à gauche de l'arbre : c'est la somme de toutes les clés joignant la racine est le noeud de clef 1.

$W_n(n)$ le poids du parcours joignant la racine et le noeud d'étiquette n .

Présentation du problème : On considère un ABR construit à partir d'une permutation de $\{1, 2, \dots, n\}$.

$W_1(n)$ représente le poids de l'arête tout à gauche de l'arbre : c'est la somme de toutes les clés joignant la racine est le noeud de clef 1.

$W_n(n)$ le poids du parcours joignant la racine et le noeud d'étiquette n .

Présentation du problème : On considère un ABR construit à partir d'une permutation de $\{1, 2, \dots, n\}$.

$W_1(n)$ représente le poids de l'arête tout à gauche de l'arbre : c'est la somme de toutes les clés joignant la racine est le noeud de clef 1.

$W_n(n)$ le poids du parcours joignant la racine et le noeud d'étiquette n .

Résultats :

Théorème (Aguech, Lasmar, Mahmoud (2007))

On a

$$\mathbf{E}[W_1(n)] = n, \quad \mathbf{V}[W_1(n)] = \frac{n(n+1)}{2},$$

et

$$\frac{W_1(n)}{n} \xrightarrow{\mathcal{L}} 1 + X,$$

où X une v.a. de *Dichman*.

Par une méthode de réflexion, et en utilisant un résultat sur la longueur du cheminement interne on montre

Théorème (Aguech, Lasmar, Mahmoud (2007))

$$\frac{W_n(n) - n \ln n}{n\sqrt{\ln n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Résultats :

Théorème (Aguech, Lasmar, Mahmoud (2007))

On a

$$\mathbf{E}[W_1(n)] = n, \quad \mathbf{V}[W_1(n)] = \frac{n(n+1)}{2},$$

et

$$\frac{W_1(n)}{n} \xrightarrow{\mathcal{L}} 1 + X,$$

où X une v.a. de *Dichman*.

Par une méthode de réflexion, et en utilisant un résultat sur la longueur du cheminement interne on montre

Théorème (Aguech, Lasmar, Mahmoud (2007))

$$\frac{W_n(n) - n \ln n}{n \sqrt{\ln n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Les urnes de Polya

Problème 5 : Les urnes de Polya

Présentation du problème : Une urne contenant des boules blanches W et des boules noires B .

Initialement, au temps 0 on a W_0 boules blanches et B_0 boules noires.

On effectue des tirages de boules à chaque instant.

Si à un instant k on a tiré une boule du type W on la remet dans l'urne et on ajoute X_k boules du type W .

Si on a tiré une du type B on la remet ds l'urne et on ajoute Y_k boules W et Z_k boules B

X_k, Y_k, Z_k sont independantes entre elles et sont des copies des v.a.r X, Y, Z .

Problème 5 : Les urnes de Polya

Présentation du problème : Une urne contenant des boules blanches W et des boules noires B .

Initialement, au temps 0 on a W_0 boules blanches et B_0 boules noires.

On effectue des tirages de boules à chaque instant.

Si à un instant k on a tiré une boule du type W on la remet dans l'urne et on ajoute X_k boules du type W .

Si on a tiré une du type B on la remet ds l'urne et on ajoute Y_k boules W et Z_k boules B

X_k, Y_k, Z_k sont independantes entre elles et sont des copies des v.a.r X, Y, Z .

Problème 5 : Les urnes de Polya

Présentation du problème : Une urne contenant des boules blanches W et des boules noires B .

Initialement, au temps 0 on a W_0 boules blanches et B_0 boules noires.

On effectue des tirages de boules à chaque instant.

Si à un instant k on a tiré une boule du type W on la remet dans l'urne et on ajoute X_k boules du type W .

Si on a tiré une du type B on la remet ds l'urne et on ajoute Y_k boules W et Z_k boules B

X_k, Y_k, Z_k sont independantes entre elles et sont des copies des v.a.r X, Y, Z .

Problème 5 : Les urnes de Polya

Présentation du problème : Une urne contenant des boules blanches W et des boules noires B .

Initialement, au temps 0 on a W_0 boules blanches et B_0 boules noires.

On effectue des tirages de boules à chaque instant.

Si à un instant k on a tiré une boule du type W on la remet dans l'urne et on ajoute X_k boules du type W .

Si on a tiré une du type B on la remet ds l'urne et on ajoute Y_k boules W et Z_k boules B

X_k, Y_k, Z_k sont independantes entre elles et sont des copies des v.a.r X, Y, Z .

Problème 5 : Les urnes de Polya

Présentation du problème : Une urne contenant des boules blanches W et des boules noires B .

Initialement, au temps 0 on a W_0 boules blanches et B_0 boules noires.

On effectue des tirages de boules à chaque instant.

Si à un instant k on a tiré une boule du type W on la remet dans l'urne et on ajoute X_k boules du type W .

Si on a tiré une du type B on la remet ds l'urne et on ajoute Y_k boules W et Z_k boules B

X_k, Y_k, Z_k sont independantes entre elles et sont des copies des v.a.r X, Y, Z .

Problème 5 : Les urnes de Polya

Objectif : Le but est d'étudier la composition asymptotique de l'urne.

Historique : Dans le cas où X, Y, Z sont des déterministes le problème a été étudié par Svante Janson (2005).

Problème 5 : Les urnes de Polya

Objectif : Le but est d'étudier la composition asymptotique de l'urne.

Historique : Dans le cas où X, Y, Z sont des déterministes le problème a été étudié par Svante Janson (2005).

Problème 5 : Les urnes de Polya

Techniques utilisées (Plongement de l'urne en temps continu) :

Chaque Boule, indépendamment de toutes les autres, est munie d'une horloge exponentielle de paramètre 1

A la sonnerie d'une horloge d'une boule du type W on ajoute à l'urne un nombre aléatoire de même loi que $X + 1$.

Si une boule du type B sonne on ajoute à l'urne un nombre aléatoire de même loi que Y de boules W et un nombre de même loi que $Z + 1$ de boules B .

Techniques utilisées (Plongement de l'urne en temps continu) :

Chaque Boule, indépendamment de toutes les autres, est munie d'une horloge exponentielle de paramètre 1

A la sonnerie d'une horloge d'une boule du type W on ajoute à l'urne un nombre aléatoire de même loi que $X + 1$.

Si une boule du type B sonne on ajoute à l'urne un nombre aléatoire de même loi que Y de boules W et un nombre de même loi que $Z + 1$ de boules B .

Techniques utilisées (Plongement de l'urne en temps continu) :

Chaque Boule, indépendamment de toutes les autres, est munie d'une horloge exponentielle de paramètre 1

A la sonnerie d'une horloge d'une boule du type W on ajoute à l'urne un nombre aléatoire de même loi que $X + 1$.

Si une boule du type B sonne on ajoute à l'urne un nombre aléatoire de même loi que Y de boules W et un nombre de même loi que $Z + 1$ de boules B .

Techniques utilisées (Plongement de l'urne en temps continu) :

Chaque Boule, indépendamment de toutes les autres, est munie d'une horloge exponentielle de paramètre 1

A la sonnerie d'une horloge d'une boule du type W on ajoute à l'urne un nombre aléatoire de même loi que $X + 1$.

Si une boule du type B sonne on ajoute à l'urne un nombre aléatoire de même loi que Y de boules W et un nombre de même loi que $Z + 1$ de boules B .

Problème 5 : Les urnes de Polya

Processus en temps continu :

On obtient un processus en temps continu $\mathcal{X}(t) = (W(t), B(t), t \geq 0)$.
 $\tau_1 < \tau_2 < \dots < \tau_n < \dots$ instants de sonnerie

Lemme (Athreya et Ney (1972))

*Les deux processus $(W_n, B_n)_{n \geq 0}$ et $(\mathcal{X}(\tau_n))_{n \geq 0} = (W(\tau_n), B(\tau_n))_{n \geq 0}$ ont même loi
et $\tau_n \rightarrow \infty$ presque sûrement.*

Processus en temps continu :

On obtient un processus en temps continu $\mathcal{X}(t) = (W(t), B(t), t \geq 0)$.
 $\tau_1 < \tau_2 < \dots < \tau_n < \dots$ instants de sonnerie

Lemme (Athreya et Ney (1972))

*Les deux processus $(W_n, B_n)_{n \geq 0}$ et $(\mathcal{X}(\tau_n))_{n \geq 0} = (W(\tau_n), B(\tau_n))_{n \geq 0}$ ont même loi
et $\tau_n \rightarrow \infty$ presque sûrement.*

Problème 5 : Les urnes de Polya

Hypothèses (H) :

(H1) X, Y, Z des v.a. entières indépendantes, telles

$$(*) \sigma_X + \sigma_Y + \sigma_Z < \infty,$$

$$(**) \mu_X \geq \mu_Z,$$

$$(***) \mu_X(\mu_Y + \mu_Z) > 0.$$

(H2) la composition initiale de l'urne est (W_0, B_0) , avec $B_0 > 0$.

Une martingale utile :

$$A = \begin{pmatrix} \mu_X & \mu_Y \\ 0 & \mu_Z \end{pmatrix}, \mathcal{Y}(t) := e^{-tA} \mathcal{X}^T(t), \mathcal{F}_t := \sigma(W(s), B(s), s \leq t).$$

Théorème (Aguech (2009))

Sous (H), le processus $\{\mathcal{Y}(t), \mathcal{F}_t, t \geq 0\}$ est une martingale bornée dans L^2 , donc converge presque sûrement et dans L^2 vers une limite

$$\tilde{\mathcal{Y}} := \begin{pmatrix} W \\ B \end{pmatrix}, W \text{ et } B \text{ sont deux variables aléatoires.}$$

Problème 5 : Les urnes de Polya

Hypothèses (H) :

(H1) X, Y, Z des v.a. entières indépendantes, telles

$$(*) \quad \sigma_X + \sigma_Y + \sigma_Z < \infty,$$

$$(**) \quad \mu_X \geq \mu_Z,$$

$$(***) \quad \mu_X(\mu_Y + \mu_Z) > 0.$$

(H2) la composition initiale de l'urne est (W_0, B_0) , avec $B_0 > 0$.

Une martingale utile :

$$A = \begin{pmatrix} \mu_X & \mu_Y \\ 0 & \mu_Z \end{pmatrix}, \quad \mathcal{Y}(t) := e^{-tA} \mathcal{X}^T(t), \quad \mathcal{F}_t := \sigma(W(s), B(s), s \leq t).$$

Théorème (Aguech (2009))

Sous (H), le processus $\{\mathcal{Y}(t), \mathcal{F}_t, t \geq 0\}$ est une martingale bornée dans L^2 , donc converge presque sûrement et dans L^2 vers une limite

$$\tilde{\mathcal{Y}} := \begin{pmatrix} W \\ B \end{pmatrix}, \quad W \text{ et } B \text{ sont deux variables aléatoires.}$$

Résultat 1 (Urne Triangulaire) :

Théorème (Aguech (2009))

$\rho = \frac{\mu_Z}{\mu_X}$ et $K = (\mu_X)^\rho B \left(W + \frac{\mu_Y}{\mu_X - \mu_Z} B \right)^{-\rho}$, sous **H** et $\mu_Y > 0$, on a

① Si $\mu_X > \mu_Z > 0$,

① presque sûrement

$$W_n = \mu_X n + o(n), \quad B_n = Kn^\rho + o(n^\rho)$$

② si $\rho \leq \frac{1}{2}$, alors, $\frac{W_n - n\mu_X}{\sqrt{n}} - \frac{K}{\mu_Z} (\mu_Y - \mu_X) \mathbb{1}_{\{\rho = \frac{1}{2}\}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_X^2)$,

③ si $\frac{1}{2} < \rho < 1$,

$$\frac{W_n - n\mu_X}{n^\rho} \xrightarrow{\mathbb{P}} \frac{K}{\mu_Z} (\mu_Y - \mu_X).$$

② si $\mu_X = \mu_Z > 0$, presque sûrement, on a

$$W_n = \mu_X n + o(n), \quad B_n = \frac{\mu_X^2}{\mu_Y} \frac{n}{\ln n} + o\left(\frac{n}{\ln n}\right).$$

Résultat 2 ($Z \stackrel{p.s.}{=} 0$) :

Théorème (Aguech (2009))

Sous (H) on a,

- 1 presque sûrement : $W_n = \mu_X n + o(n)$,
- 2 si $\sigma_X \neq 0$, on a un TLC,

$$\frac{W_n - \mu_X n}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_X^2),$$

- 3 si $X = \alpha \neq 0$ déterministe, alors

$$\frac{1}{\sqrt{\ln n}} \left(W_n - n\alpha - \frac{\mu_Y - \alpha}{\alpha} B_0 \ln n \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \left(\sigma_Y^2 + (\mu_Y - \alpha)^2 \right) \frac{B_0}{\alpha} \right).$$

Problème 5 : Les urnes de Polya

Résultat 3 (Urne diagonale) :

Théorème (Aguech (2009))

$\rho = \frac{\mu_Z}{\mu_X}$ and $D = \mu_X^\rho B W^{-\rho}$, alors

① si $\mu_X > \mu_Z > 0$,

① $W_n \stackrel{p.s.}{=} \mu_X n + o(n)$, $B_n \stackrel{p.s.}{=} D n^\rho + o(n^\rho)$,

② si $0 < \rho \leq \frac{1}{2}$, alors lorsque $n \rightarrow +\infty$

$$\frac{W_n - n\mu_X}{\sqrt{n}} + \sqrt{\mu_X} \frac{\mu_X}{\mu_Z} B W^{-1/2} \mathbb{I}_{\{\rho=\frac{1}{2}\}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_X^2),$$

③ si $\frac{1}{2} < \rho < 1$, $\frac{W_n - n\mu_X}{n^\rho} \xrightarrow{\mathbb{P}} -\mu_X^\rho \frac{\mu_X}{\mu_Z} B W^{-\rho}$.

② $\mu_X = \mu_Z$

$$W_n \stackrel{p.s.}{=} \mu_X \frac{W}{W+B} n + o(n), \quad B_n \stackrel{p.s.}{=} \mu_X \frac{B}{W+B} n + o(n).$$

Résultat 3 (Urne diagonale) :

Théorème (Aguech (2009))

$\rho = \frac{\mu_Z}{\mu_X}$ and $D = \mu_X^\rho B W^{-\rho}$, alors

① si $\mu_X > \mu_Z > 0$,

① $W_n \stackrel{p.s.}{=} \mu_X n + o(n)$, $B_n \stackrel{p.s.}{=} D n^\rho + o(n^\rho)$,

② si $0 < \rho \leq \frac{1}{2}$, alors lorsque $n \rightarrow +\infty$

$$\frac{W_n - n\mu_X}{\sqrt{n}} + \sqrt{\mu_X} \frac{\mu_X}{\mu_Z} B W^{-1/2} \mathbb{I}_{\{\rho = \frac{1}{2}\}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_X^2),$$

③ si $\frac{1}{2} < \rho < 1$, $\frac{W_n - n\mu_X}{n^\rho} \xrightarrow{\mathbb{P}} -\mu_X^\rho \frac{\mu_X}{\mu_Z} B W^{-\rho}$.

② $\mu_X = \mu_Z$

$$W_n \stackrel{p.s.}{=} \mu_X \frac{W}{W+B} n + o(n), \quad B_n \stackrel{p.s.}{=} \mu_X \frac{B}{W+B} n + o(n).$$

Fragmentation aléatoire

Problème 6 : Fragmentation aléatoire

Description du problème : Au départ on a un segment de longueur x , avec une probabilité $p(x)$ (qui peut ne pas dépendre de x), ce segment se fragmente en b morceaux de longueurs $(x\xi_j)_{1 \leq j \leq b}$ (Les $(\xi_j)_j$ sont appelés loi de reproduction et ne dépendent pas de x).

Ce segment reste indéfiniment stable (ne se fragmente jamais) avec la probabilité $q(x) = 1 - p(x)$.

Ainsi de suite chaque sous-fragment subit, d'une façon indépendante et indépendamment de tout l'historique, le même phénomène.

Problème 6 : Fragmentation aléatoire

Description du problème : Au départ on a un segment de longueur x , avec une probabilité $p(x)$ (qui peut ne pas dépendre de x), ce segment se fragmente en b morceaux de longueurs $(x\xi_j)_{1 \leq j \leq b}$ (Les $(\xi_j)_j$ sont appelés loi de reproduction et ne dépendent pas de x).

Ce segment reste indéfiniment stable (ne se fragmente jamais) avec la probabilité $q(x) = 1 - p(x)$.

Ainsi de suite chaque sous-fragment subit, d'une façon indépendante et indépendamment de tout l'historique, le même phénomène.

Problème 6 : Fragmentation aléatoire

Description du problème : Au départ on a un segment de longueur x , avec une probabilité $p(x)$ (qui peut ne pas dépendre de x), ce segment se fragmente en b morceaux de longueurs $(x\xi_j)_{1 \leq j \leq b}$ (Les $(\xi_j)_j$ sont appelés loi de reproduction et ne dépendent pas de x).

Ce segment reste indéfiniment stable (ne se fragmente jamais) avec la probabilité $q(x) = 1 - p(x)$.

Ainsi de suite chaque sous-fragment subit, d'une façon indépendante et indépendamment de tout l'historique, le même phénomène.

Problème 6 : Fragmentation aléatoire

Objectif : Etude de $N(x)$ représentant le nombre de segments stables à la fin du processus de fragmentation.

Historique :

- Sibuya, Itoh (1987) \implies

$$b = 2, \quad p(x) = \mathbb{I}_{\{x \geq 1\}}, \quad \xi_1 \text{ de loi uniforme sur } [0, 1],$$

- Janson, Neininger (2007) \implies

$$p(x) = \mathbb{I}_{\{x \geq 1\}}.$$

Problème 6 : Fragmentation aléatoire

Objectif : Etude de $N(x)$ représentant le nombre de segments stables à la fin du processus de fragmentation.

Historique :

- Sibuya, Itoh (1987) \implies

$$b = 2, \quad p(x) = \mathbb{I}_{\{x \geq 1\}}, \quad \xi_1 \text{ de loi uniforme sur } [0, 1],$$

- Janson, Neininger (2007) \implies

$$p(x) = \mathbb{I}_{\{x \geq 1\}}.$$

Problème 6 : Fragmentation aléatoire

Objectif : Etude de $N(x)$ représentant le nombre de segments stables à la fin du processus de fragmentation.

Historique :

- Sibuya, Itoh (1987) \implies

$$b = 2, \quad p(x) = \mathbb{1}_{\{x \geq 1\}}, \quad \xi_1 \text{ de loi uniforme sur } [0, 1],$$

- Janson, Neininger (2007) \implies

$$p(x) = \mathbb{1}_{\{x \geq 1\}}.$$

Problème 6 : Fragmentation aléatoire

Hypothèses : On suppose que le segment initial est de longueur x (assez grand), $p(x) = 1 - e^{-x}$, $b = 2$, ξ_1 suit la loi uniforme sur $[0, 1]$.

Motivation :

Problème 6 : Fragmentation aléatoire

Hypothèses : On suppose que le segment initial est de longueur x (assez grand), $p(x) = 1 - e^{-x}$, $b = 2$, ξ_1 suit la loi uniforme sur $[0, 1]$.

Motivation :

Problème 6 : Fragmentation aléatoire

Méthode de contraction :

$$N(x) \stackrel{p.s.}{=} \mathbb{I}_{\{x \text{ stable}\}} + \mathbb{I}_{\{x \text{ instable}\}} \sum_{j=1}^2 N^{(j)}(\xi_j x)$$



La méthode de contraction adaptée en temps continu
(Janson Neininger(2007))



$$N_*(x) := \frac{N(x) - m(x)}{\sqrt{v(x)}}.$$



$$N_*(x) = \sum_{j=1}^2 N_*^{(j)}(\xi_j x) \sqrt{\frac{v(\xi_j x)}{v(x)}} + D(x), \quad (7)$$

Problème 6 : Fragmentation aléatoire

Méthode de contraction :

$$N(x) \stackrel{p.s.}{=} \mathbb{I}_{\{x \text{ stable}\}} + \mathbb{I}_{\{x \text{ instable}\}} \sum_{j=1}^2 N^{(j)}(\xi_j x)$$



La méthode de contraction adaptée en temps continu
(Janson Neininger(2007))



$$N_*(x) := \frac{N(x) - m(x)}{\sqrt{v(x)}}.$$



$$N_*(x) = \sum_{j=1}^2 N_*^{(j)}(\xi_j x) \sqrt{\frac{v(\xi_j x)}{v(x)}} + D(x), \quad (7)$$

Problème 6 : Fragmentation aléatoire

Méthode de contraction :

$$N(x) \stackrel{p.s.}{=} \mathbb{I}_{\{x \text{ stable}\}} + \mathbb{I}_{\{x \text{ instable}\}} \sum_{j=1}^2 N^{(j)}(\xi_j x)$$



La méthode de contraction adaptée en temps continu
(Janson Neininger(2007))



$$N_*(x) := \frac{N(x) - m(x)}{\sqrt{v(x)}}.$$



$$N_*(x) = \sum_{j=1}^2 N_*^{(j)}(\xi_j x) \sqrt{\frac{v(\xi_j x)}{v(x)}} + D(x), \quad (7)$$

Problème 6 : Fragmentation aléatoire

Résultat1 (Les moments de $N(x)$) :

Théorème (Aguech (2008))

Asymptotiquement, pour x assez grand

$$m(x) := \mathbf{E}[N(x)] = 2\gamma x + O(1),$$

$\gamma > 0$ *constante explicite.*

Théorème (Aguech (2008))

$$v(x) := \text{Var}[N(x)] = \lambda x + O(1). \quad (8)$$

$\lambda > 0$ *constante connue.*

Résultat2 (Loi limite de $N_*(x)$) :

Théorème

$N_*(x)$ converge en loi vers une variable aléatoire N_* , qui vérifie l'égalité en loi suivante

$$N_* \stackrel{\mathcal{L}}{=} N_*^{(1)} \sqrt{\xi_1} + N_*^{(2)} \sqrt{1 - \xi_1}, \quad (9)$$

avec $N_*^{(1)}$ et $N_*^{(2)}$ sont des copies indépendantes de N_* .



Théorème (Aguech (2008))

$N_*(x)$ converge en loi vers $\mathcal{N}(0, 1)$.

Résultat2 (Loi limite de $N_*(x)$) :

Théorème

$N_*(x)$ converge en loi vers une variable aléatoire N_* , qui vérifie l'égalité en loi suivante

$$N_* \stackrel{\mathcal{L}}{=} N_*^{(1)} \sqrt{\xi_1} + N_*^{(2)} \sqrt{1 - \xi_1}, \quad (9)$$

avec $N_*^{(1)}$ et $N_*^{(2)}$ sont des copies indépendantes de N_* .



Théorème (Aguech (2008))

$N_*(x)$ converge en loi vers $\mathcal{N}(0, 1)$.

Un model de Moran, pattern matching

Problème 7 : Un model de Moran, pattern matching

Présentation du problème : *Une population de m gamètes qui évoluent en fonction du temps d'âges initiales : a_1, a_2, \dots, a_m .*

À chaque instant, une de ces gamètes (choisie au hasard) meurt et est remplacée par un nouveau gamète, les autres restent en vie.

Application à la théorie de panne : Ce modèle est analogue à supposer qu'on a une machine composée de m composantes toutes neuves et que à chaque unité de temps on choisit, au hasard, une composante et on la remplace par une neuve.

Autres applications : propagation de vague [Mahmoud & Itoh, 2004],
algorithme de mise en cache

Problème 7 : Un model de Moran, pattern matching

Présentation du problème : *Une population de m gamètes qui évoluent en fonction du temps d'âges initiales : a_1, a_2, \dots, a_m .*

À chaque instant, une de ces gamètes (choisie au hasard) meurt et est remplacée par un nouveau gamète, les autres restent en vie.

Application à la théorie de panne : Ce modèle est analogue à supposer qu'on a une machine composée de m composantes toutes neuves et que à chaque unité de temps on choisit, au hasard, une composante et on la remplace par une neuve.

Autres applications : propagation de vague [Mahmoud & Itoh, 2004],
algorithme de mise en cache

Problème 7 : Un model de Moran, pattern matching

Présentation du problème : *Une population de m gamètes qui évoluent en fonction du temps d'âges initiales : a_1, a_2, \dots, a_m .*

À chaque instant, une de ces gamètes (choisie au hasard) meurt et est remplacée par un nouveau gamète, les autres restent en vie.

Application à la théorie de panne : Ce modèle est analogue à supposer qu'on a une machine composée de m composantes toutes neuves et que à chaque unité de temps on choisit, au hasard, une composante et on la remplace par une neuve.

Autres applications : propagation de vague [Mahmoud & Itoh, 2004],
algorithme de mise en cache

Problème 7 : Un model de Moran, pattern matching

Présentation du problème : *Une population de m gamètes qui évoluent en fonction du temps d'âges initiales : a_1, a_2, \dots, a_m .*

À chaque instant, une de ces gamètes (choisie au hasard) meurt et est remplacée par un nouveau gamète, les autres restent en vie.

Application à la théorie de panne : Ce modèle est analogue à supposer qu'on a une machine composée de m composantes toutes neuves et que à chaque unité de temps on choisit, au hasard, une composante et on la remplace par une neuve.

Autres applications : propagation de vague [Mahmoud & Itoh, 2004],
algorithme de mise en cache

Problème 7 : Un model de Moran, pattern matching

Présentation du problème : *Une population de m gamètes qui évoluent en fonction du temps d'âges initiales : a_1, a_2, \dots, a_m .*

À chaque instant, une de ces gamètes (choisie au hasard) meurt et est remplacée par un nouveau gamète, les autres restent en vie.

Application à la théorie de panne : Ce modèle est analogue à supposer qu'on a une machine composée de m composantes toutes neuves et que à chaque unité de temps on choisit, au hasard, une composante et on la remplace par une neuve.

Autres applications : propagation de vague [Mahmoud & Itoh, 2004],
algorithme de mise en cache

modèle (Modèle de Moran généralisé) : Dans la machine à m composantes, initialement d'âges respectives a_1, \dots, a_m , à chaque unité de temps on choisit au hasard une composante, ensuite on décide avec probabilité p de la remplacer par une neuve et on ne touche pas aux autres.

modèle (Modèle de Moran généralisé) : Dans la machine à m composantes, initialement d'ages respectives a_1, \dots, a_m , à chaque unité de temps on choisit au hasard une composante, ensuite on décide avec probabilité p de la remplacer par une neuve et on ne touche pas aux autres.

Problème 7 : Un modèle de Moran, pattern matching

Exemple (modèle voisin en Pattern Matching) : Etant donné un mot ω et m mots de longueur n (leurs lettres sont i.i.d de loi connue),

- Quelle est la longueur de la plus longue séquence commune avec ω ?
- Quelle est la longueur maximale de la dernière séquence commune ω ?

Exemple avec $m = 2$ et $n = 24$:

$\omega = abcea**cbddacbcc**acabb**aaaa**,$

$abceb**cbddacbcc**bcabb**acaa**$

$abce**ccbcddac**acccccac**aaaa**.$

La longueur de la plus longue séquence $H_n = 10$

La longueur maximale de la dernière séquence commune est $L_n = 5$

Problème 7 : Un modèle de Moran, pattern matching

Exemple (modèle voisin en Pattern Matching) : Etant donné un mot ω et m mots de longueur n (leurs lettres sont i.i.d de loi connue),

- Quelle est la longueur de la plus longue séquence commune avec ω ?
- Quelle est la longueur maximale de la dernière séquence commune ω ?

Exemple avec $m = 2$ et $n = 24$:

$\omega = abcea**cbcd**dacbccacabb**baaaa**,$

$abceb**cbcd**dacbccbcabbacaa$

$abceccbcddacacccac**baaaa**.$

La longueur de la plus longue séquence $H_n = 10$

La longueur maximale de la dernière séquence commune est $L_n = 5$

Problème 7 : Un modèle de Moran, pattern matching

Exemple (modèle voisin en Pattern Matching) : Etant donné un mot ω et m mots de longueur n (leurs lettres sont i.i.d de loi connue),

- Quelle est la longueur de la plus longue séquence commune avec ω ?
- Quelle est la longueur maximale de la dernière séquence commune ω ?

Exemple avec $m = 2$ et $n = 24$:

$\omega = abcea**cbcd**dacbccacab**baaaa**,$

$abceb**cbcd**dacbccbcabbacaa$

$abceccbcddacacccac**baaaa**.$

La longueur de la plus longue séquence $H_n = 10$

La longueur maximale de la dernière séquence commune est $L_n = 5$

Problème 7 : Un modèle de Moran, pattern matching

Exemple (modèle voisin en Pattern Matching) : Etant donné un mot ω et m mots de longueur n (leurs lettres sont i.i.d de loi connue),

- Quelle est la longueur de la plus longue séquence commune avec ω ?
- Quelle est la longueur maximale de la dernière séquence commune ω ?

Exemple avec $m = 2$ et $n = 24$:

$\omega = abcea**bcddacbcc**acabb**baaaa**,$

$abceb**bcddacbcc**bcabbacaa$

$abceccbcddacacccac**baaaa**.$

La longueur de la plus longue séquence $H_n = 10$

La longueur maximale de la dernière séquence commune est $L_n = 5$

Résultat (cas $m = 1$) :

Theorem (Knuth 1998)

La hauteur H_n associée au modèle de Moran avec $m = 1$, vérifie, pour n assez grand

$$\mathbb{P}\left(H_n \leq \lfloor \ln_2 n \rfloor + h\right) = \exp(-a(n)2^{-h-1}) + O\left(\frac{\ln n}{\sqrt{n}}\right), \quad a(n) = 2^{\{\ln_2 n\}}.$$

En particulier, l'espérance de H_n est donnée, asymptotiquement, par

$$\mathbb{E}[H_n] = \frac{\ln n}{\ln 2} + \frac{\gamma}{\ln 2} - \frac{3}{2} + P(\ln_2(n)) + O\left(\frac{\ln^2 n}{\sqrt{n}}\right)$$

P est une fonction continue périodique et

$$\mathbf{V}(H_n) = O(1).$$

Résultat1 (Altitude finale ($m = 2$)) : Soit (X_n, Y_n) les âges à la n^{me} génération.

X_n et Y_n ont la même loi.

Théorème

Asymptotiquement, on a la convergence en loi

$$X_n, Y_n \longrightarrow \text{Geo}(p/2) + 1$$

Résultat1 (Altitude finale ($m = 2$)) : Soit (X_n, Y_n) les âges à la n^{me} génération.

X_n et Y_n ont la même loi.

Théorème

Asymptotiquement, on la convergence en loi

$$X_n, Y_n \longrightarrow \text{Geo}(p/2) + 1$$

Problème 7 : Un modèle de Moran, pattern matching

Age maximal atteint ($m = 2$) : Pour $h \in \mathbb{N}^*$, soit $F^{\leq h}(t, x, y)$ la série génératrice des marches qui ne dépassent pas h :

$$F^{\leq h}(t, x, y) = \sum_{n \geq 0} f_n^{\leq h}(x, y) t^n, \text{ où}$$

$$f_n^{\leq h}(x, y) = \mathbf{E} \left[x^{X_n} y^{Y_n} \mathbb{I}_{\{\max(X_1, Y_1) \leq h, \dots, \max(X_n, Y_n) \leq h\}} \right].$$

Proposition

$$\begin{aligned} F^{\leq h}(t, x, y) &= 1 + tqxy F^{\leq h}(t, x, y) \\ &+ tpxy/2 \left[F^{\leq h}(t, x, 1) + F^{\leq h}(t, 1, y) \right] \\ &- qxy \left(x^h y^h [x^h y^h] F^{\leq h}(t, x, y) \right) \\ &- tpxy/2 \left[x^h [x^h] F^{\leq h}(t, x, 1) - y^h [y^h] F^{\leq h}(t, 1, y) \right], \end{aligned}$$

Problème 7 : Un modèle de Moran, pattern matching

Age maximal atteint ($m = 2$) : Pour $h \in \mathbb{N}^*$, soit $F^{\leq h}(t, x, y)$ la série génératrice des marches qui ne dépassent pas h :

$$F^{\leq h}(t, x, y) = \sum_{n \geq 0} f_n^{\leq h}(x, y) t^n, \text{ où}$$

$$f_n^{\leq h}(x, y) = \mathbf{E} \left[x^{X_n} y^{Y_n} \mathbb{I}_{\{\max(X_1, Y_1) \leq h, \dots, \max(X_n, Y_n) \leq h\}} \right].$$

Proposition

$$\begin{aligned} F^{\leq h}(t, x, y) &= 1 + tqxy F^{\leq h}(t, x, y) \\ &+ tpxy/2 \left[F^{\leq h}(t, x, 1) + F^{\leq h}(t, 1, y) \right] \\ &- qxy \left(x^h y^h [x^h y^h] F^{\leq h}(t, x, y) \right) \\ &- tpxy/2 \left[x^h [x^h] F^{\leq h}(t, x, 1) - y^h [y^h] F^{\leq h}(t, 1, y) \right], \end{aligned}$$

Problème 7 : Un modèle de Moran, pattern matching

Age maximal atteint ($m = 2$) : Pour $h \in \mathbb{N}^*$, soit $F^{\leq h}(t, x, y)$ la série génératrice des marches qui ne dépassent pas h :

$$F^{\leq h}(t, x, y) = \sum_{n \geq 0} f_n^{\leq h}(x, y) t^n, \text{ où}$$

$$f_n^{\leq h}(x, y) = \mathbf{E} \left[x^{X_n} y^{Y_n} \mathbb{I}_{\{\max(X_1, Y_1) \leq h, \dots, \max(X_n, Y_n) \leq h\}} \right].$$

Proposition

$$\begin{aligned} F^{\leq h}(t, x, y) &= 1 + tqxy F^{\leq h}(t, x, y) \\ &+ tpxy/2 \left[F^{\leq h}(t, x, 1) + F^{\leq h}(t, 1, y) \right] \\ &- qxy \left(x^h y^h [x^h y^h] F^{\leq h}(t, x, y) \right) \\ &- tpxy/2 \left[x^h [x^h] F^{\leq h}(t, x, 1) - y^h [y^h] F^{\leq h}(t, 1, y) \right], \end{aligned}$$

Problème 7 : Un modèle de Moran, pattern matching

Age maximal atteint ($m = 2$) : Pour $h \in \mathbb{N}^*$, soit $F^{\leq h}(t, x, y)$ la série génératrice des marches qui ne dépassent pas h :

$$F^{\leq h}(t, x, y) = \sum_{n \geq 0} f_n^{\leq h}(x, y) t^n, \text{ où}$$

$$f_n^{\leq h}(x, y) = \mathbf{E} \left[x^{X_n} y^{Y_n} \mathbb{I}_{\{\max(X_1, Y_1) \leq h, \dots, \max(X_n, Y_n) \leq h\}} \right].$$

Proposition

$$\begin{aligned} F^{\leq h}(t, x, y) &= 1 + tqxy F^{\leq h}(t, x, y) \\ &+ tpxy/2 \left[F^{\leq h}(t, x, 1) + F^{\leq h}(t, 1, y) \right] \\ &- qxy \left(x^h y^h [x^h y^h] F^{\leq h}(t, x, y) \right) \\ &- tpxy/2 \left[x^h [x^h] F^{\leq h}(t, x, 1) - y^h [y^h] F^{\leq h}(t, 1, y) \right], \end{aligned}$$

Hauteur H_n ($m = 2$) :

Pour étudier H_n il suffit d'avoir

$$F^{\leq h}(t, 1, 1) = \sum_{n \geq 0} \mathbb{P}(H_n \leq h) t^n.$$

On a

Proposition

$$F^{\leq h}(t, 1, 1) = \frac{1 - t + \left(\frac{t}{2}\right)^h - 2\left(\frac{t}{2}\right)^{h+1}}{1 - t + \left(\frac{t}{2}\right)^h}.$$

$$\mathbb{P}(H_n \leq h) = [t^n] F^{\leq h}(t, 1, 1).$$

Mais

$$\mathbb{E}[H_n] = \sum_{h \geq 0} \mathbb{P}(H_n > h)$$

Hauteur H_n ($m = 2$) :

Pour étudier H_n il suffit d'avoir

$$F^{\leq h}(t, 1, 1) = \sum_{n \geq 0} \mathbb{P}(H_n \leq h) t^n.$$

On a

Proposition

$$F^{\leq h}(t, 1, 1) = \frac{1 - t + \left(\frac{t}{2}\right)^h - 2\left(\frac{t}{2}\right)^{h+1}}{1 - t + \left(\frac{t}{2}\right)^h}.$$

$$\mathbb{P}(H_n \leq h) = [t^n] F^{\leq h}(t, 1, 1).$$

Mais

$$\mathbf{E}[H_n] = \sum_{h \geq 0} \mathbb{P}(H_n > h)$$

Résultat 2 (H_n) ($m = 2$) :

Théorème (Aguech, Banderier (2009))

$$\mathbf{E}[H_n] = \frac{\ln(n)}{\ln(2)} + O(1).$$

Remarque : Résultat asymptotique complet peut être obtenu via Mellin.

Perspectives

Problème de Parking de Rényi (1958) : Parking de longueur x .

Chaque voiture, supposée de même largeur 1, se gare d'une façon aléatoire uniforme, donc coupe le parking en deux parties de longueurs $U(x - 1)$ et $(1 - U)(x - 1)$, U suit la loi uniforme sur $[0, 1]$.

Ainsi de suite chaque voiture choisit aléatoirement une partie de longueur grand 1 et se gare d'une façon uniforme et indépendamment de tout l'historique.

Question : Combien de voitures peuvent trouver une place au parking??

Généralisation : Si chaque segment de longueur $y > 1$ est coupé selon une loi d'un vecteur (ξ_1, \dots, ξ_b) en b morceaux de longueurs $((x - 1)\xi_1, \dots, (x - 1)\xi_b)$,

Que sera le comportement asymptotique du nombre de segments obtenus à la fin du processus de fragmentation??

Problème de Parking de Rényi (1958) : Parking de longueur x .

Chaque voiture, supposée de même largeur 1, se gare d'une façon aléatoire uniforme, donc coupe le parking en deux parties de longueurs $U(x - 1)$ et $(1 - U)(x - 1)$, U suit la loi uniforme sur $[0, 1]$.

Ainsi de suite chaque voiture choisit aléatoirement une partie de longueur grand 1 et se gare d'une façon uniforme et indépendamment de tout l'historique.

Question : Combien de voitures peuvent trouver une place au parking??

Généralisation : Si chaque segment de longueur $y > 1$ est coupé selon une loi d'un vecteur (ξ_1, \dots, ξ_b) en b morceaux de longueurs $((x - 1)\xi_1, \dots, (x - 1)\xi_b)$,

Que sera le comportement asymptotique du nombre de segments obtenus à la fin du processus de fragmentation??

Poids sur les ABR :

- Poids d'un parcours de l'arbre choisi au hasard.
- Poids total sur l'arbre = la somme des poids de tous les parcours.

Poids sur les ABR :

- Poids d'un parcours de l'arbre choisi au hasard.
- Poids total sur l'arbre = la somme des poids de tous les parcours.

Fragmentation non uniforme : Généraliser les résultats obtenus concernant la fragmentation non uniforme avec condition d'arrêt aléatoire : un segment de longueur y reste indéfiniment stable avec la probabilité $q(y) = e^{-y}$.

Chaque fragment non stable de longueur y se fragmente selon une loi donnée d'un vecteur aléatoire (ξ_1, \dots, ξ_b) en b morceaux de longueurs $(y\xi_1, \dots, y\xi_b)$

Question : Comportement asymptotique en loi du nombre de segments stables lorsque la taille du premier segment x tend vers l'infini.

Fragmentation non uniforme : Généraliser les résultats obtenus concernant la fragmentation non uniforme avec condition d'arrêt aléatoire : un segment de longueur y reste indéfiniment stable avec la probabilité $q(y) = e^{-y}$.

Chaque fragment non stable de longueur y se fragmente selon une loi donnée d'un vecteur aléatoire (ξ_1, \dots, ξ_b) en b morceaux de longueurs $(y\xi_1, \dots, y\xi_b)$

Question : Comportement asymptotique en loi du nombre de segments stables lorsque la taille du premier segment x tend vers l'infini.

Modèle de Moran généralisé :

- Etude de la variance de H_n .
- Etude de la loi limite de H_n .
- Généraliser le problème pour $m \geq 2$.
- Existe-t-il un problème de transition de phase en m ?

Modèle de Moran généralisé :

- Etude de la variance de H_n .
- Etude de la loi limite de H_n .
- Généraliser le problème pour $m \geq 2$.
- Existe-t-il un problème de transition de phase en m ?

Modèle de Moran généralisé :

- Etude de la variance de H_n .
- Etude de la loi limite de H_n .
- Généraliser le problème pour $m \geq 2$.
- Existe-t-il un problème de transition de phase en m ?

Modèle de Moran généralisé :

- Etude de la variance de H_n .
- Etude de la loi limite de H_n .
- Généraliser le problème pour $m \geq 2$.
- Existe-t-il un problème de transition de phase en m ?

Urnes de Polya :

- Avoir la loi jointe du couple (W_n, B_n) .
- Caractérisation de la loi limite de la Martingale $\mathcal{Y}(t)$.
- Généraliser le travail à plus que deux couleurs.

Urnes de Polya :

- Avoir la loi jointe du couple (W_n, B_n) .
- Caractérisation de la loi limite de la Martingale $\mathcal{Y}(t)$.
- Généraliser le travail à plus que deux couleurs.

Urnes de Polya :

- Avoir la loi jointe du couple (W_n, B_n) .
- Caractérisation de la loi limite de la Martingale $\mathcal{Y}(t)$.
- Généraliser le travail à plus que deux couleurs.

MERCI