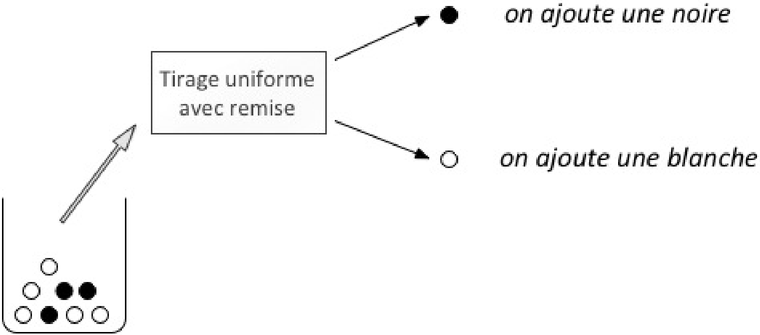
Urnes de Pólya
Analyse d'une classe algébrique

Basile Morcrette,
avec Philippe Flajolet

Projet Algorithms, INRIA Rocquencourt.

10 mars 2011

1. Modèles d'urnes



- Une urne contenant des boules de deux couleurs différentes
- Des règles fixées pour l'évolution de l'urne

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Urne de Pólya équilibrée

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \quad \alpha, \delta \in \mathbb{Z}, \qquad \beta, \gamma \in \mathbb{N}$$

Urne équilibrée : $\boxed{\alpha + \beta = \gamma + \delta}$ (nombre total de boules déterministe)

Configuration initiale fixée (a_0, b_0) : a_0 boules $\bullet$ (comptées par x)

b_0 boules $\circ$ (comptées par y)

Définition

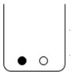
Histoire de longueur n : une suite de n évolutions (n tirages)

$$H(x, y, z) = \sum_{n,a,b} H_{n,a,b} \, x^a y^b \frac{z^n}{n!}$$

$H_{n,a,b}$: nombre d'histoires de longueur n, débutant en (a_0, b_0), et terminant en (a, b)

Compter les histoires - Exemple

Prenons l'urne $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $(a_1, b_1) = (1, 1)$.

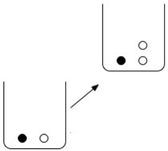


$$H(x, y, z) =$$

$$xyz$$

Compter les histoires - Exemple

Prenons l'urne $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $(a_1, b_1) = (1, 1)$.

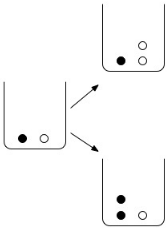


$$H(x, y, z) =$$

$$xyz$$

Compter les histoires - Exemple

Prenons l'urne $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $(a_1, b_1) = (1, 1)$.



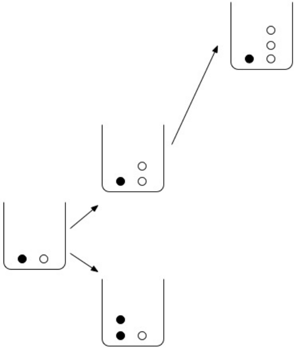
$$H(x,y,z) =$$

$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

Compter les histoires - Exemple

Prenons l'urne $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $(a_1, b_1) = (1, 1)$.



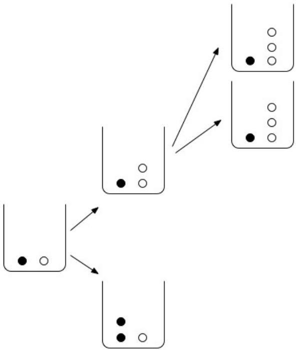
$$H(x, y, z) =$$

$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

Compter les histoires - Exemple

Prenons l'urne $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $(a_1, b_1) = (1, 1)$.



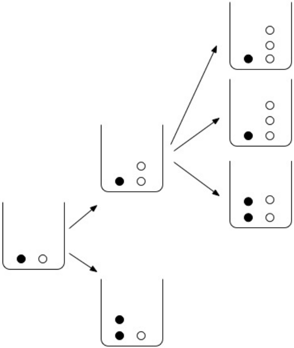
$$H(x, y, z) =$$

$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

Compter les histoires - Exemple

Prenons l'urne $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $(a_1, b_1) = (1, 1)$.



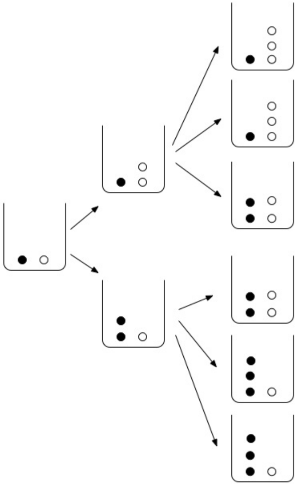
$$H(x, y, z) =$$

$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

Compter les histoires - Exemple

Prenons l'urne $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $(a_1, b_1) = (1, 1)$.



$$H(x, y, z) =$$

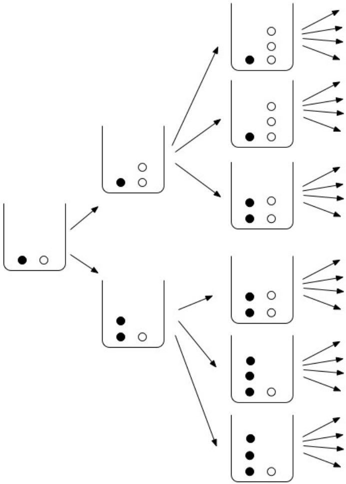
$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

$$+ \quad (2xy^3 + 2x^2y^2 + 2x^3y)\frac{z^3}{6}$$

Compter les histoires - Exemple

Prenons l'urne $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $(a_1, b_1) = (1, 1)$.



$$
\begin{aligned}
H(x,y,z) = & \\
& xyz \\
+ & (xy^2 + x^2y)\frac{z^2}{2} \\
+ & (2xy^3 + 2x^2y^2 + 2x^3y)\frac{z^3}{6} \\
+ & \dots
\end{aligned}
$$

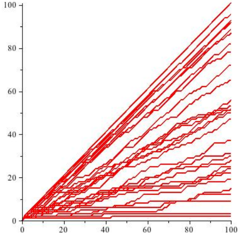
Des comportements très divers

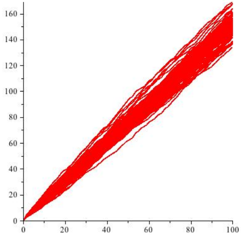
Problème Comprendre la composition de l'urne après n étapes, lorsque n tend vers ∞.

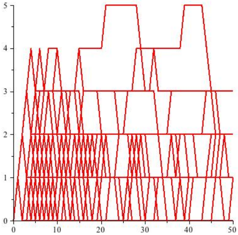
Des comportements très divers

Problème Comprendre la composition de l'urne après n étapes, lorsque n tend vers ∞.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$







Urne de Pólya — **Urne à croissance préférentielle** — Urne triangulaire 3×3

Résultats probabilistes

Urne $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ Rapport $\rho = \dfrac{\alpha - \gamma}{\alpha + \beta}$

- Petites urnes : $\rho \leqslant \frac{1}{2}$

  Loi limite **gaussienne** [Smythe96] [Janson04]

- Grandes urnes $\rho > \frac{1}{2}$

  Lois **non gaussiennes** [Mahmoud] [Janson04] [Chauvin–Pouyanne–Sahnoun11]

Outils :

- Plongement en temps continu [Jan04] [ChPoSa11]
- Martingales, théorème de la limite centrale

Urnes équilibrées et analyse

- Premiers pas : [Flajolet–Gabarro–Pekari05] : *Analytic urns*
- [Flajolet–Dumas–Puyhaubert06], sur les urnes avec coefficients négatifs, et cas triangulaire
- [Kuba–Panholzer–Hwang07], urnes non équilibrées

Approche analytique Théorème [FlDuPu06]

Urne $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ et $\left\{ \begin{array}{c} (a_0, b_0) \\ \alpha + \beta = \gamma + \delta \end{array} \right.$ $\Longrightarrow$ $H = X^{a_0} Y^{b_0}$ où $\left\{ \begin{array}{rcl} \dot{X} & = & X^{\alpha+1} Y^{\beta} \\ \dot{Y} & = & X^{\gamma} Y^{\delta+1} \end{array} \right.$

Travaux réalisés / Cadre

- Urnes équilibrées à coefficients positifs
- Étude d'une classe à deux paramètres

2. Urnes à croissance préférentielle

Motivation : Caractériser les urnes 2×2 additives (coefficients positifs).
Approche : Avoir une classe d'urnes avec des fonctions génératrices "exploitables".

Théorème [M.Fl11]

La classe d'urnes équilibrées $\begin{pmatrix} 2\alpha & \beta \\ \alpha & \alpha+\beta \end{pmatrix}$, avec $\alpha > 0$, $\beta \geqslant 0$, a une fonction génératrice bivariée algébrique.

La FG des histoires $H(x, 1, z)$ est racine du polynôme en Y

$$(z - a - b(x))\, Y^{2\alpha+\beta} \;+\; b(x)\, Y^{\alpha} \;+\; a$$

$$\text{avec } b(x) = \frac{x^{-\alpha} - 1}{\alpha + \beta} \text{ et } a = (2\alpha+\beta)^{-1}$$

Preuve

Système différentiel : $\left\{ \begin{array}{rcl} \dot{X} & = & X^{2\alpha+1}\, Y^{\beta} \\ \dot{Y} & = & X^{\alpha}\, Y^{\alpha+\beta+1} \end{array} \right.$ $\qquad \dot{X} = \frac{\partial}{\partial z} X$

$$\frac{\dot{X}}{X^{\alpha+1}} = \frac{\dot{Y}}{Y^{\alpha+1}} = X^{\alpha} Y^{\beta}$$

$$X^{-\alpha} - Y^{-\alpha} = x^{-\alpha} - y^{-\alpha}$$

$$\frac{\dot{Y}}{Y^{\alpha+\beta+1}} (Y^{-\alpha} + x^{-\alpha} - y^{-\alpha}) = 1$$

$$\frac{1}{2\alpha+\beta} Y^{-(2\alpha+\beta)} + \frac{x^{-\alpha} - y^{-\alpha}}{\alpha+\beta} Y^{-(\alpha+\beta)} = -\left(z - \frac{x^{-\alpha} - y^{-\alpha}}{\alpha+\beta} - \frac{1}{2\alpha+\beta} \right)$$

Urne équilibrée $a + b = a_0 + b_0 + n\sigma$. On pose $y = 1$.

$$\boxed{\left(z - \frac{x^{-\alpha} - 1}{\alpha+\beta} - \frac{1}{2\alpha+\beta} \right) Y^{2\alpha+\beta} + \frac{x^{-\alpha} - 1}{\alpha+\beta} Y^{\alpha} + \frac{1}{2\alpha+\beta} = 0}$$

Premières observations

Équilibre $\sigma = 2\alpha + \beta$ $\qquad \rho = \frac{\alpha}{2\alpha+\beta} \leqslant \frac{1}{2}$

Pour $x = 1$, l'équation devient : $(z - \sigma^{-1})Y^{\sigma} + \sigma^{-1} = 0$

Ainsi pour $(a_0, b_0) = (0, 1)$

$$H(1,1,z) = (1 - \sigma z)^{-1/\sigma} \qquad h_n \sim \frac{\sigma^n n^{1/\sigma - 1}}{\Gamma(1/\sigma)}$$

Proposition

Soit X_n la variable aléatoire comptant le nombre de boules de couleur $\mathtt{x}$ dans l'urne au temps n. Alors

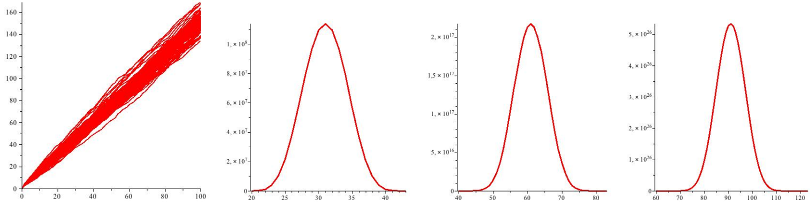
$$\mathbb{E}(X_n) = \frac{\alpha(2\alpha+\beta)}{\alpha+\beta} n + \frac{\alpha}{\alpha+\beta} \frac{\Gamma(\frac{1}{2\alpha+\beta})}{\Gamma(\frac{\alpha+1}{2\alpha+\beta})} n^{\frac{\alpha}{2\alpha+\beta}} + \frac{\alpha}{\alpha+\beta} + O\left(n^{\frac{\alpha}{2\alpha+\beta}-1}\right),$$

$$\mathbb{V}(X_n) = \frac{\alpha^3(2\alpha+\beta)}{(\alpha+\beta)^2} n + O\left(n^{\frac{\alpha+\beta}{2\alpha+\beta}}\right).$$

Exemple $\alpha = 1$, $\beta = 1$

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad \begin{array}{lcl} \mathtt{x} & \rightarrow & \mathtt{x\,x\,y} \\ \mathtt{y} & \rightarrow & \mathtt{x\,y\,y} \end{array}$$

croissance préférentielle



$$\left(z - \frac{x^{-1}-1}{2} - \frac{1}{3}\right) Y^3 + \frac{x^{-1}-1}{2} Y + \frac{1}{3} = 0$$

Méthode de col pour x=1

$$\left(z - \frac{1}{3}\right) Y^3 + \frac{1}{3} = 0$$

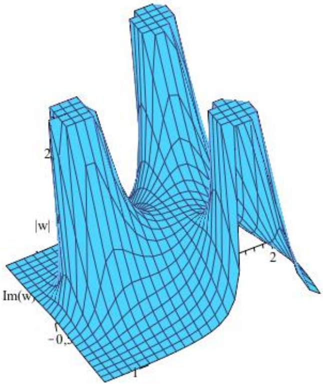
$$y_n = \frac{1}{2i\pi} \oint \frac{Y(z)}{z^{n+1}} \mathrm{d}z$$

$$\boxed{y_n = \frac{3^{n+1}}{2i\pi} \oint a(w) h(w)^{n+1} \mathrm{d}w}$$

$$\begin{cases} a(w) & = & 1 - w \\ h(w) & = & \dfrac{1}{w(w^2 - 3w + 3)} \end{cases}$$

$$h'(w) = \frac{-3(w-1)^2}{w^2(w^2 - 3w + 3)^2}$$

évaluer l'intégrale selon le bon contour...



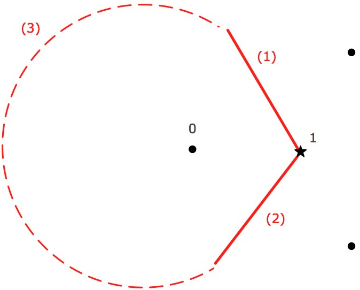
$w \mapsto |h(w)|$

3 pôles

1 col double en $w = 1$

Méthode de col x=1 (suite)

$t \in [0..L]$
(1) $w(t) = 1 + te^{i2\pi/3}$
(2) $w(t) = 1 + te^{-i2\pi/3}$



$$h(w(t))^n = \exp\left(-n(t^3 + O\left(t^6\right)\right)$$

Choisir L... $nL^3 \to \infty$ et $nL^6 \to 0$
Prenons $L \sim n^{-1/4}$

$$\int_{(1)} + \int_{(2)} : \int_0^\infty ue^{-u^3} du \ \text{ et } \ \int_{(3)} \text{exponentiellement négligeable}$$

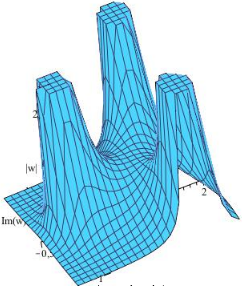
$$\boxed{y_n = \frac{3^n}{\Gamma(1/3)} \left(n^{-2/3} + O\left(n^{-11/12}\right)\right)}$$

Méthode de col $x \neq 1$

$$\left(z - \frac{x^{-1}-1}{2} - \frac{1}{3}\right) Y^3 + \frac{x^{-1}-1}{2} Y + \frac{1}{3} = 0$$

$$y_n = \frac{3^{n+1}}{2i\pi} \oint a_x(w) h_x(w)^{n+1} \mathrm{d}w$$

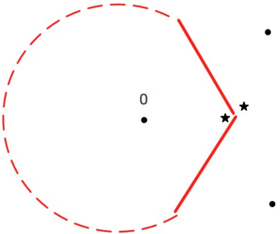
$$h'_x(1) = h'(x^{-1}) = 0$$



$w \mapsto |h_x(w)|$

3 pôles

2 cols en $w = 1$ et $w = x^{-1}$



$$x = 1 + \frac{\tilde{x}}{\sqrt{n}}, \quad |\tilde{x}| < 1$$

$$y_n(x) \sim \frac{3^n n^{-2/3}}{\Gamma(1/3)} \exp\left(\frac{3}{2}\sqrt{n}\tilde{x} - \frac{3}{8}\tilde{x}^2\right)$$

$$\boxed{p_n(x) = \frac{y_n(x)}{y_n(1)} \sim \exp\left(\tfrac{3}{2}\sqrt{n}\tilde{x} - \tfrac{3}{8}\tilde{x}^2\right)}$$

Loi normale limite

Soit X_n la variable aléatoire comptant le nombre de boules $\bullet$ dans l'urne après n étapes.

Théorème

$$\mathbb{P}\left\{\frac{X_n - \frac{3}{2}n}{\sqrt{\frac{3n}{4}}} \leqslant t\right\} = \Phi(t) + O\left(\frac{1}{\sqrt{n}}\right).$$

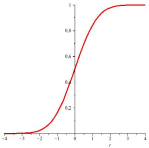
Loi Locale Limite

Théorème

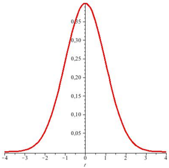
Notons $p_{n,k} = \mathbb{P}\{X_n = k\}$. La distribution des X_n satisfait une loi locale limite de type gaussienne avec vitesse de convergence $O\left(\frac{1}{\sqrt{n}}\right)$, i.e.

$$\sup_{t\in\mathbb{R}} \left| \frac{\sqrt{3n}}{2} p_{n,\lfloor 3n/2 + t\sqrt{3n}/2 \rfloor} - \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \right| \leqslant \frac{1}{\sqrt{n}} .$$

$$\mathbb{P}\left\{ \frac{X_n - \frac{3}{2}n}{\sqrt{\frac{3n}{4}}} \leqslant t \right\} \underset{n\to\infty}{\longrightarrow}$$



$$\frac{\sqrt{3n}}{2}\,\mathbb{P}\left\{ X_n = \left\lfloor \frac{3n}{2} + t\frac{\sqrt{3n}}{2} \right\rfloor \right\} \underset{n\to\infty}{\longrightarrow}$$

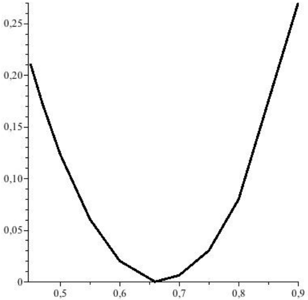


Grandes déviations

- Borne exponentiellement petite sur la grande déviation par rapport à la valeur moyenne : quantification sur les événements rares

Théorème

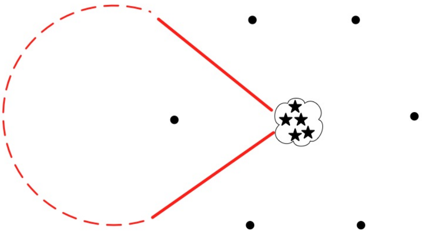
- si $0.42 < t < 2/3$, $\mathbb{P}(X_n \leqslant tn) \approx e^{-nW(t)}$ (queue gauche)
- si $2/3 < t < 0.73$, $\mathbb{P}(X_n \geqslant tn) \approx e^{-nW(t)}$ (queue droite)



Cas général

$$\left(z - \frac{x^{-\alpha}-1}{\alpha+\beta} - \frac{1}{2\alpha+\beta}\right) Y^{2\alpha+\beta} + \frac{x^{-\alpha}-1}{\alpha+\beta} Y^{\alpha} + \frac{1}{2\alpha+\beta} = 0$$

$$y_n(x) = \frac{\sigma^{n+1}}{2i\pi} \oint a_x(w) h_x(w)^{n+1} \mathrm{d}w$$



$h_x(w)$: $\sigma = 2\alpha+\beta$ pôles
Col en 1 avec multiplicité $\alpha+\beta-1$
On a α autres cols en $1-(1-x^{-\alpha})^{1/\alpha}$
$x \sim 1 + O(n^{-1/2})$
$L \sim n^{-\frac{1}{\sigma+1}}$

$$y_n(x) \sim \frac{\sigma^n n^{\frac{1-\sigma}{\sigma}}}{\Gamma(1/\sigma)} \exp\left(\frac{\alpha\sigma}{\alpha+\beta}\sqrt{n}\tilde{x} - \frac{\alpha^3\sigma}{2(\alpha+\beta)^2}\tilde{x}^2\right)$$

$$p_n(x) \sim \exp\left(\frac{\alpha\sigma}{\alpha+\beta}\sqrt{n}\tilde{x} - \frac{\alpha^3\sigma}{2(\alpha+\beta)^2}\tilde{x}^2\right)$$

Conclusion et perspectives

- Premiers résultats sur les petites urnes additives, avec des méthodes analytiques.
- Lien avec l'analyse de singularités ...
- Une autre piste : vers les urnes non équilibrées – EDP

Conclusion et perspectives

- Premiers résultats sur les petites urnes additives, avec des méthodes analytiques.
- Lien avec l'analyse de singularités ...
- Une autre piste : vers les urnes non équilibrées – EDP

Merci !