Boltzmann en dimension *un et quelques*

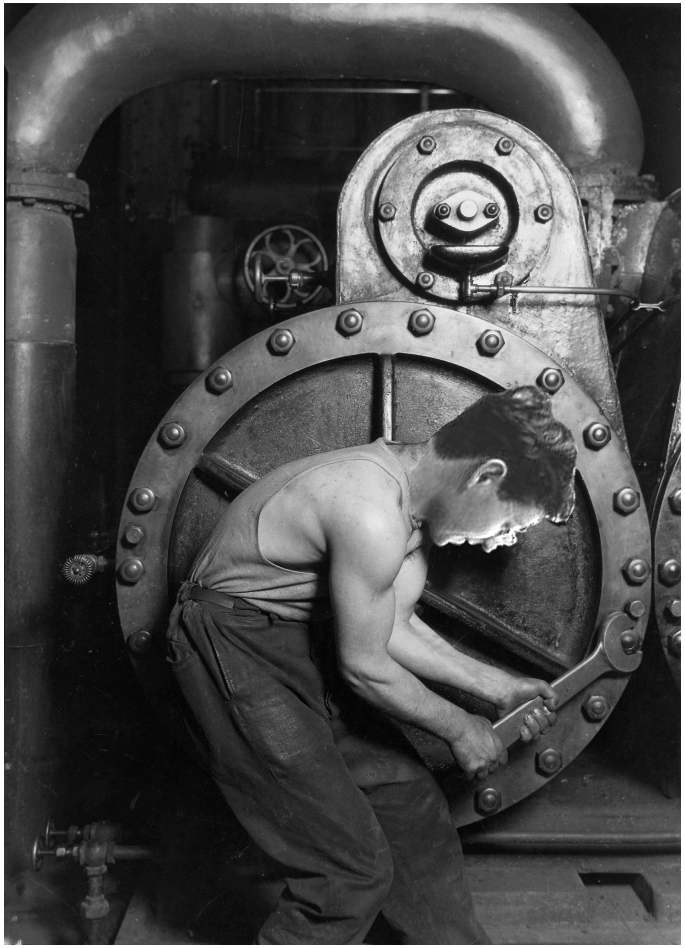
Yann Ponty Olivier Bodini

ANR GAMMA - LIAFA/LIP6/IGM/LRI/LIPN - Paris VII

UPMC - Paris universitas

March 20, 2009

Travail en cours



Lewis Hine, 1920 + O. Bodini

Encore pas mal de boulons à serrer ...

Contexte

Objectif

Génération aléatoire d'objets combinatoires ayant une composition contrainte $(n_1, \ldots, n_k)$ en plusieurs types d'atomes $(Z_1, \ldots, Z_k)$.

Motivation : Étude de classes difficilement spécifiables

- Langages contextuels

  Ex. : $\mathcal{L}' = \{\omega \in \mathcal{L} \mid |\omega|_a = |\omega|_b = |\omega|_c\}$, $\mathcal{L} \subseteq \{a, b, c\}^* \in \mathrm{Rat}$
- Marches (excursions) d'ordonnée finale déterminée
- Objets biologiques

Contexte

Objectif

Génération aléatoire d'objets combinatoires ayant une composition contrainte $(n_1, \ldots, n_k)$ en plusieurs types d'atomes $(Z_1, \ldots, Z_k)$.

Motivation : Étude de classes difficilement spécifiables

- Langages contextuels
- Marches (excursions) d'ordonnée finale déterminée

  Ex. : Excursions à pas $\nearrow$ $(+1, +a)$ et $\searrow$ $(+1, -b)$, ordonnée finale h
  $\Rightarrow$ Langages algébriques $+ \left\{ n_{\nearrow} = \frac{bn+h}{a+b}, n_{\searrow} = \frac{an-h}{a+b} \right\}$
- Objets biologiques

Contexte

Objectif

Génération aléatoire d'objets combinatoires ayant une composition contrainte $(n_1, \ldots, n_k)$ en plusieurs types d'atomes $(Z_1, \ldots, Z_k)$.

Motivation : Étude de classes difficilement spécifiables

- Langages contextuels
- Marches (excursions) d'ordonnée finale déterminée
- Objets biologiques
  ADN : Séquences de `A`, `T`, `G` et `C` respectant des fréquences observées.
  Structures secondaires d'ARN : Arbres unaires/binaires
  + Contraintes (Nombre,longueur) sur séquences de noeuds unaires

Contexte

Objectif

Génération aléatoire d'objets combinatoires ayant une composition contrainte $(n_1, \ldots, n_k)$ en plusieurs types d'atomes $(Z_1, \ldots, Z_k)$.

Motivation : Étude de classes difficilement spécifiables

- Langages contextuels
- Marches (excursions) d'ordonnée finale déterminée
- Objets biologiques

$\Rightarrow$ Boltzmann multivarié

Exemple : ARN

ARN $\approx$ Mots de Motzkin [1]

$$\mathcal{T} \quad = \quad Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + Z_c \times \mathcal{T} + \varepsilon$$

Ex : $Z_a Z_a Z_a Z_c Z_c Z_a Z_a Z_a Z_c Z_c Z_c Z_c Z_{a'} Z_{a'} Z_{a'} Z_{a'} Z_{a'} Z_{a'} Z_c Z_c$

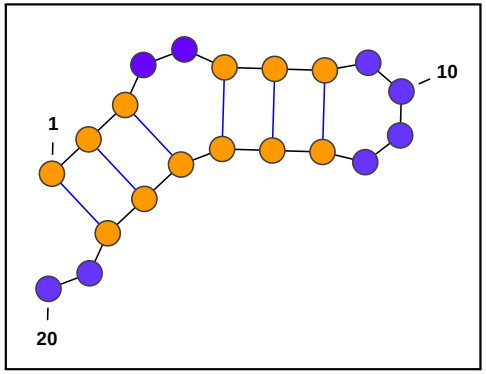
[1] **Remarque :** Normalement, absence de pics $\nearrow\searrow$, mais séries génératrices multivariées sur plusieurs pages.

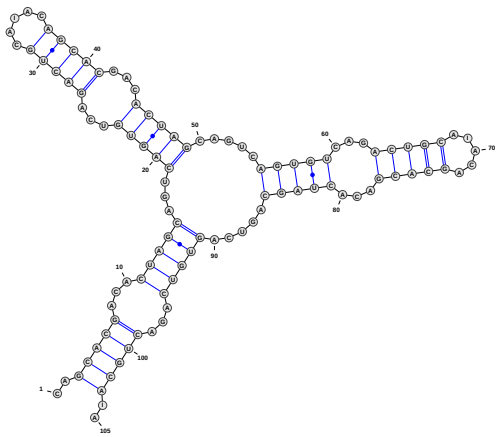
Exemple : ARN

ARN $\approx$ Mots de Motzkin [1]

$$\mathcal{T} \quad = \quad Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + Z_c \times \mathcal{T} + \varepsilon$$

Ex : $Z_a Z_a Z_a Z_c Z_c Z_a Z_a Z_a Z_c Z_c Z_c Z_c Z_{a'} Z_{a'} Z_{a'} Z_{a'} Z_{a'} Z_{a'} Z_c Z_c$





[1] **Remarque :** Normalement, absence de pics $\nearrow\searrow$, mais séries génératrices multivariées sur plusieurs pages.

Génération de Boltzmann

Principe général [DucLouFlaSch02]

- Tirage d'objets selon une probabilité $\frac{x^n}{C(x)}$, où :

  n est la taille de l'objet
  x est un paramètre *magique*
  $C(x)$ est la valeur de la série génératrice

- Rejet des objets de mauvaise taille

Choix judicieux du paramètre x :

$\Rightarrow$ Génération en $\mathcal{O}(n^2)$ en taille exacte n

$\Rightarrow$ Génération en $\mathcal{O}(n)$ en taille approchée $[n(1-\varepsilon), n(1+\varepsilon)]$

Difficultés :

- Calcul de l'oracle
- Détermination de x

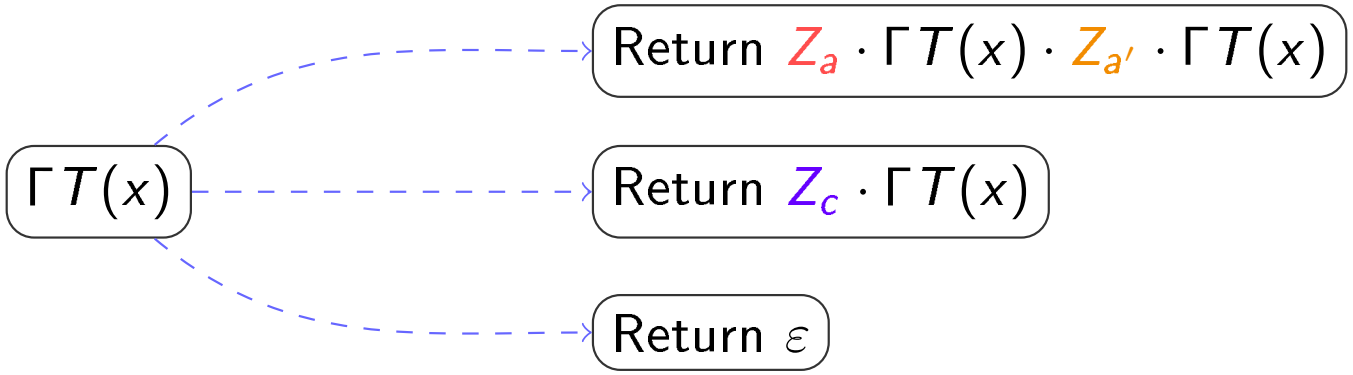
$\Rightarrow$ Itération numérique (Newton)
[Pivoteau, Salvy, Soria 08]

Génération de Boltzmann pour les ARN

Spécification/Série génératrice

$$\mathcal{T} = Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + Z_c \times \mathcal{T} + \varepsilon \quad \Rightarrow \quad T(z) = \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z^2}$$

Générateur de Boltzmann correspondant



Détermination du paramètre x

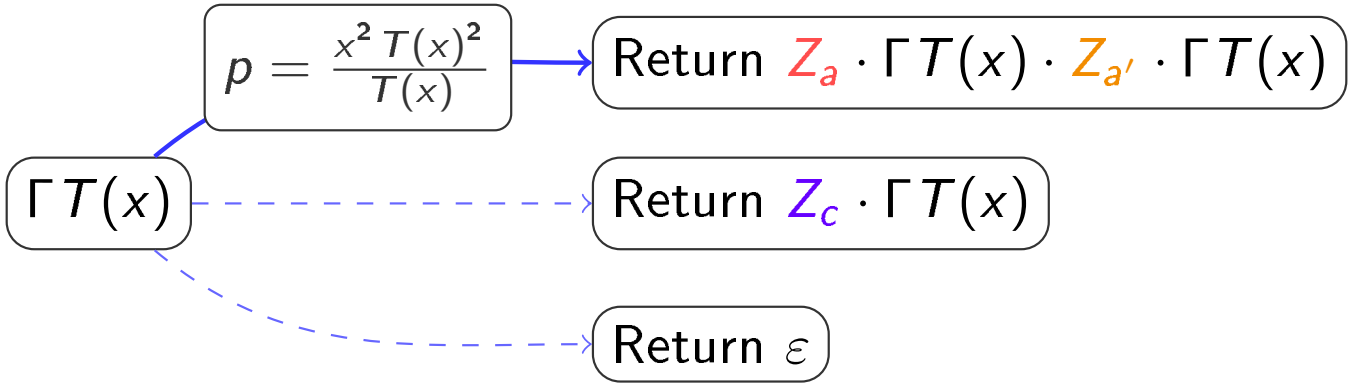
1. Choix de x tel que $n = x \frac{T'(x)}{T(x)}$ $\Rightarrow$ Pointage nécessaire
2. Générateur singulier : $x = \rho$ singularité dominante + Rejet anticipé

Génération de Boltzmann pour les ARN

Spécification/Série génératrice

$$\mathcal{T} = Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + Z_c \times \mathcal{T} + \varepsilon \quad \Rightarrow \quad T(z) = \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z^2}$$

Générateur de Boltzmann correspondant



Détermination du paramètre x

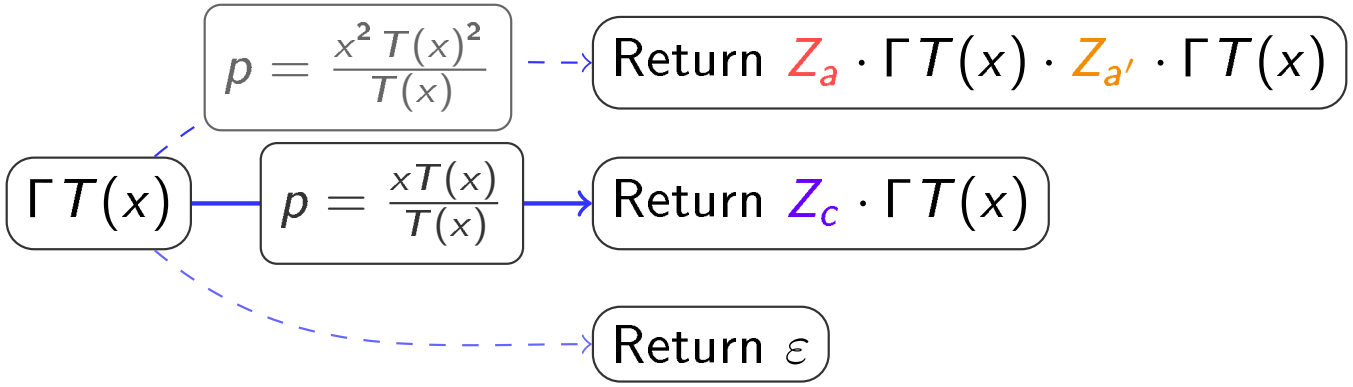
1. Choix de x tel que $n = x \frac{T'(x)}{T(x)}$ $\Rightarrow$ Pointage nécessaire
2. Générateur singulier : $x = \rho$ singularité dominante + Rejet anticipé

Génération de Boltzmann pour les ARN

Spécification/Série génératrice

$$\mathcal{T} = Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + Z_c \times \mathcal{T} + \varepsilon \quad \Rightarrow \quad T(z) = \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z^2}$$

Générateur de Boltzmann correspondant



Détermination du paramètre x

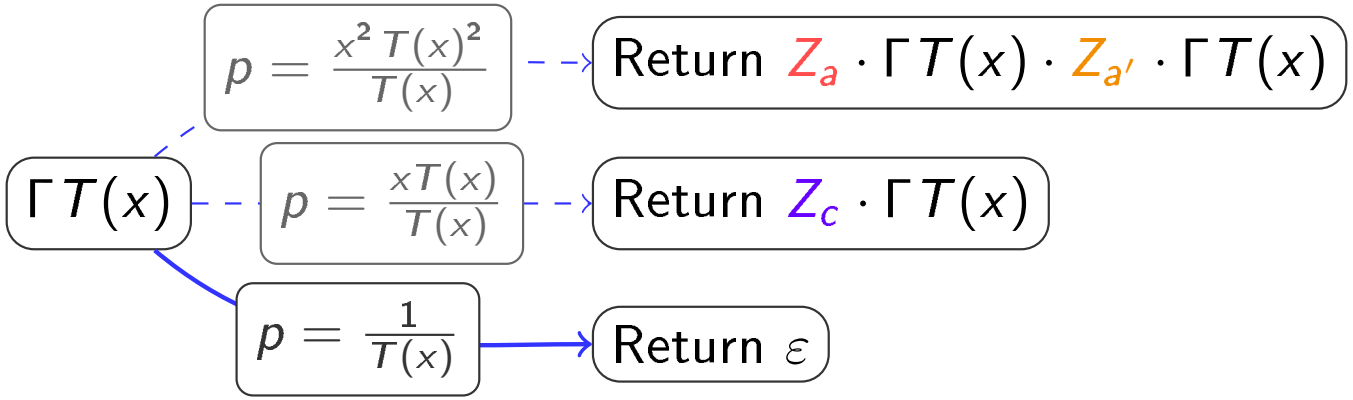
1. Choix de x tel que $n = x \frac{T'(x)}{T(x)}$ $\Rightarrow$ Pointage nécessaire
2. Générateur singulier : $x = \rho$ singularité dominante + Rejet anticipé

Génération de Boltzmann pour les ARN

Spécification/Série génératrice

$$\mathcal{T} = Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + Z_c \times \mathcal{T} + \varepsilon \quad \Rightarrow \quad T(z) = \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z^2}$$

Générateur de Boltzmann correspondant



Détermination du paramètre x

1. Choix de x tel que $n = x \frac{T'(x)}{T(x)}$ $\Rightarrow$ Pointage nécessaire
2. Générateur singulier : $x = \rho$ singularité dominante + Rejet anticipé

Boltzmann : Version multivariée

Atomes $\mathbf{Z} = (Z_1, \ldots, Z_K) \Rightarrow$ Paramètres $\mathbf{y} = (y_1, \ldots, y_k)$

Principe général

- Tirage d'objets selon une probabilité $\frac{x^n \prod_{i=1}^{k} y_i^{n_i}}{C(x,\mathbf{y})}$ où

n	Taille de l'objet
n_i	Nombre d'occurrences de Z_i dans l'objet
$x,\mathbf{y}$	Paramètres de la génération
$C(x,\mathbf{y})$	Valeur de la série génératrice multivariée

- Rejet des objets de taille ou composition inadéquate

Remarque 1 : Générateurs similaires :

$$C(x) \to C(x,\mathbf{y}) \qquad Z_i \to xy_i$$

Remarque 2 : À composition $(n; n_1, \ldots, n_k)$ fixée, génération **uniforme**

Exemple : ARN, ajout de paramètres

ARN $\approx$ Mots de Motzkin [2]

$$\mathcal{T} \quad = \quad Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + Z_c \times \mathcal{T} + \varepsilon$$

[2] **Remarque :** Normalement, absence de pics $\nearrow\searrow$, mais séries génératrices multivariées sur plusieurs pages.

Exemple : ARN, ajout de paramètres

ARN $\approx$ Mots de Motzkin [2]

$$\mathcal{T} \quad = \quad Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \boxed{Z_c \times \mathcal{T}} + \varepsilon$$

[2] **Remarque :** Normalement, absence de pics $\nearrow\searrow$, mais séries génératrices multivariées sur plusieurs pages.

Exemple : ARN, ajout de paramètres

ARN $\approx$ Mots de Motzkin [2]

$$\begin{aligned}\mathcal{T} &= Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + Z_c \times \mathcal{T} + \varepsilon \\ &= \mathrm{Seq}(Z_c) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon)\end{aligned}$$

[2] **Remarque :** Normalement, absence de pics $\nearrow\searrow$, mais séries génératrices multivariées sur plusieurs pages.

Exemple : ARN, ajout de paramètres

ARN $\approx$ Mots de Motzkin [2]

$$\begin{aligned}
\mathcal{T} &= Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + Z_c \times \mathcal{T} + \varepsilon \\
&= \mathsf{Seq}(Z_c) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon) \\
&= (\varepsilon + Z_b \times \mathsf{Seq}(Z_c)) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon)
\end{aligned}$$

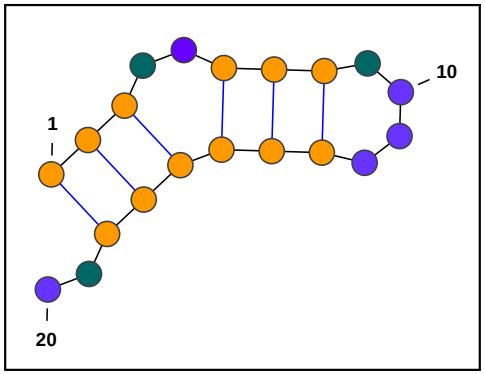
[2] **Remarque :** Normalement, absence de pics $\nearrow\searrow$, mais séries génératrices multivariées sur plusieurs pages.

Exemple : ARN, ajout de paramètres

ARN $\approx$ Mots de Motzkin [2]

$$\begin{aligned}
\mathcal{T} &= Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + Z_c \times \mathcal{T} + \varepsilon \\
&= \mathrm{Seq}(Z_c) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon) \\
&= (\varepsilon + Z_b \times \mathrm{Seq}(Z_c)) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon)
\end{aligned}$$

Ex : $Z_a Z_a Z_a Z_b Z_c Z_a Z_a Z_a Z_b Z_c Z_c Z_c Z_{a'} Z_{a'} Z_{a'} Z_{a'} Z_{a'} Z_{a'} Z_b Z_c$



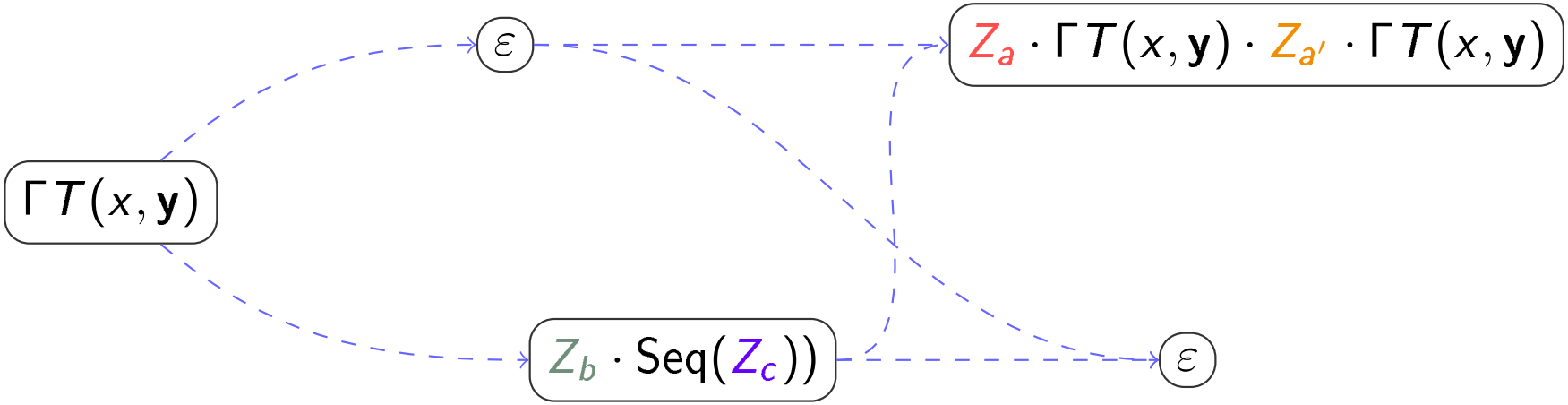
[2] **Remarque :** Normalement, absence de pics $\nearrow\searrow$, mais séries génératrices multivariées sur plusieurs pages.

Exemple : ARN, Boltzmann multivarié

Spécification/Série génératrice

$$\begin{aligned}
\mathcal{T} &= (\varepsilon + Z_b \times \mathrm{Seq}(Z_c)) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon) \\
T(z, a, b, c) &= \frac{1 - cz - \sqrt{\Delta^2 - \Delta'^2}}{2\,(1 + bz - cz)\,a^2 z^2} \\
\Delta := 1 - cz \qquad & \qquad \Delta' := 2az\,(1 + bz - cz)
\end{aligned}$$

Générateur de Boltzmann correspondant



Exemple : ARN, Boltzmann multivarié

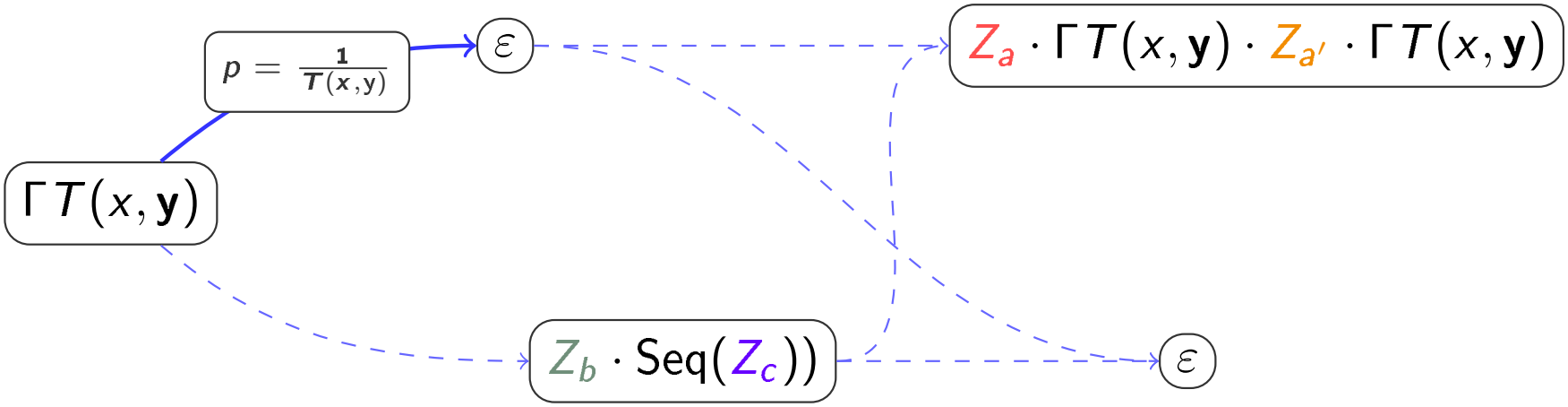
Spécification/Série génératrice

$$\mathcal{T} = (\varepsilon + Z_b \times \mathrm{Seq}(Z_c)) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon)$$

$$T(z, a, b, c) = \frac{1 - cz - \sqrt{\Delta^2 - \Delta'^2}}{2\,(1 + bz - cz)\,a^2 z^2}$$

$$\Delta := 1 - cz \qquad \Delta' := 2az\,(1 + bz - cz)$$

Générateur de Boltzmann correspondant



Exemple : ARN, Boltzmann multivarié

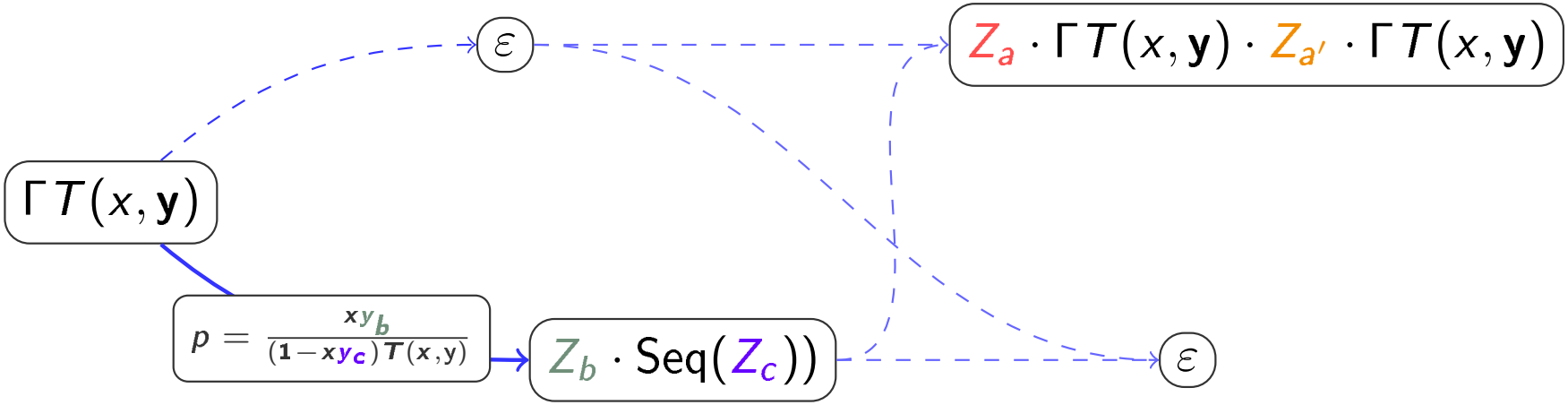
Spécification/Série génératrice

$$\mathcal{T} = (\varepsilon + Z_b \times \mathrm{Seq}(Z_c)) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon)$$

$$T(z, a, b, c) = \frac{1 - cz - \sqrt{\Delta^2 - \Delta'^2}}{2\,(1 + bz - cz)\,a^2 z^2}$$

$$\Delta := 1 - cz \qquad \Delta' := 2az\,(1 + bz - cz)$$

Générateur de Boltzmann correspondant



Exemple : ARN, Boltzmann multivarié

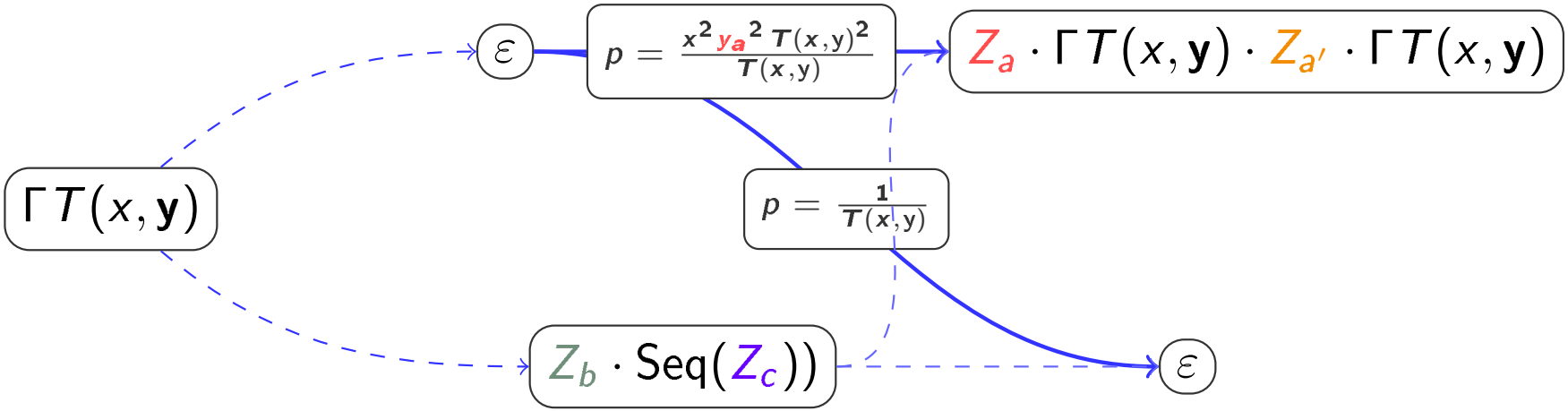
Spécification/Série génératrice

$$\mathcal{T} = (\varepsilon + Z_b \times \mathrm{Seq}(Z_c)) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon)$$

$$T(z, a, b, c) = \frac{1 - cz - \sqrt{\Delta^2 - \Delta'^2}}{2\,(1 + bz - cz)\,a^2 z^2}$$

$$\Delta := 1 - cz \qquad \Delta' := 2az\,(1 + bz - cz)$$

Générateur de Boltzmann correspondant



Exemple : ARN, Boltzmann multivarié

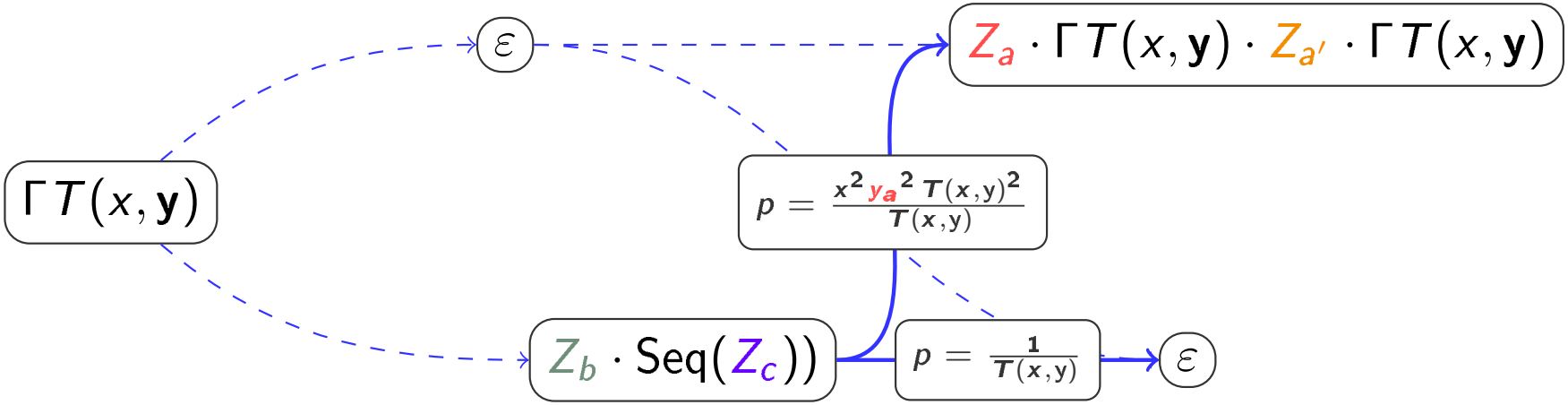
Spécification/Série génératrice

$$\mathcal{T} = (\varepsilon + Z_b \times \mathrm{Seq}(Z_c)) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon)$$

$$T(z, a, b, c) = \frac{1 - cz - \sqrt{\Delta^2 - \Delta'^2}}{2\,(1 + bz - cz)\,a^2 z^2}$$

$$\Delta := 1 - cz \qquad \Delta' := 2az\,(1 + bz - cz)$$

Générateur de Boltzmann correspondant



Exemple : ARN, Boltzmann multivarié

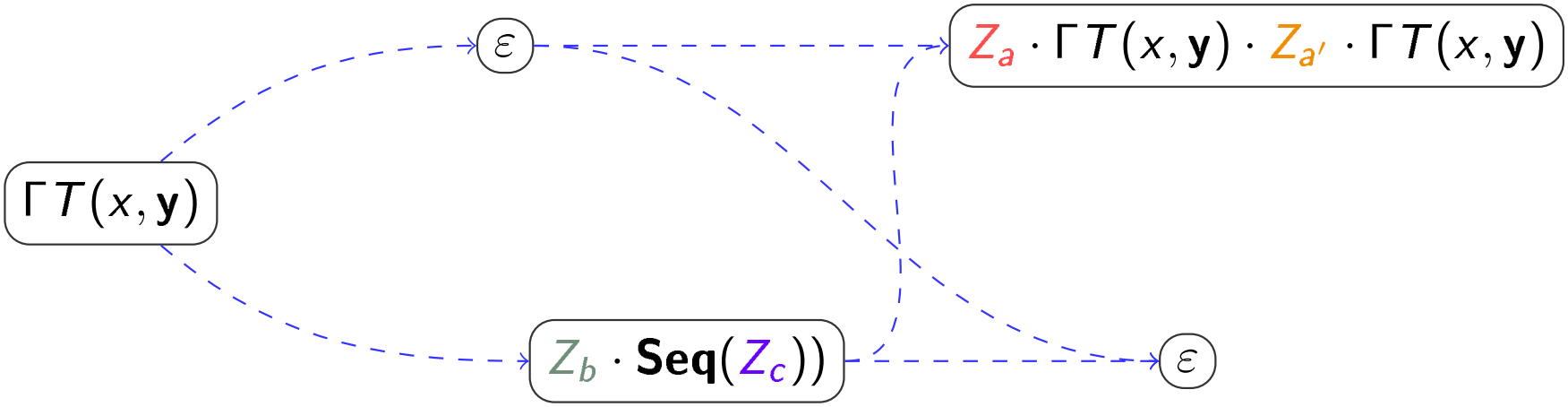
Spécification/Série génératrice

$$\mathcal{T} = (\varepsilon + Z_b \times \mathrm{Seq}(Z_c)) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon)$$

$$T(z, a, b, c) = \frac{1 - cz - \sqrt{\Delta^2 - \Delta'^2}}{2\,(1 + bz - cz)\,a^2 z^2}$$

$$\Delta := 1 - cz \qquad \Delta' := 2az\,(1 + bz - cz)$$

Générateur de Boltzmann correspondant



Boltzmann multivarié : Analyse

Validation d'un générateur

Correction : Respect de la distribution !
Efficacité : Analyse de la complexité $\Rightarrow$ Analyse de singularité

Supposons connues des valeurs pour $\mathbf{y}$ telles que

$$\mathbb{E}(N_i) = n_i, \ \forall i \in [1, k]$$

Analyse multivariée complexe $\Rightarrow$ Modification *technique* de l'algorithme :

Principe général

- Tirage d'objets selon une probabilité $\frac{x^n \prod_{i=1}^{k} y_i^{n_i}}{C(x,\mathbf{y})}$
- Rejet des objets de taille ou composition inadéquate

Boltzmann multivarié : Analyse

Validation d'un générateur

Correction : Respect de la distribution !
Efficacité : Analyse de la complexité $\Rightarrow$ Analyse de singularité

Supposons connues des valeurs pour $\mathbf{y}$ telles que

$$\mathbb{E}(N_i) = n_i, \ \forall i \in [1, k]$$

Analyse multivariée complexe $\Rightarrow$ Modification *technique* de l'algorithme :

Principe général

- Tirage d'objets selon une probabilité $\frac{x^n \prod_{i=1}^{k} y_i^{n_i}}{C(x,\mathbf{y})}$

Phase I Rejet des objets de taille inadéquate
Phase II Rejet des objets de composition inadéquate

$\Rightarrow$ Dimension un (taille) et *des poussières* (composition) ...

Phase I : Complexité de Boltzmann

Hypothèse : Valeurs de $\mathbf{y}$ indépendantes de n

Simplification

Classe combinatoire + Paramètres $\mathbf{y} \approx$ Classe combinatoire

Pour des valeurs rationnelles :

- Distribution stable par multiplication des valeurs par un réel
- Multiplication par PPCM des dénominateurs $\Rightarrow$ Valeurs entières
- Duplication des sous-expressions $\Rightarrow$ Spécification standard

Ex : $Z_a = 2/3$, $Z_{a'} = 1$ $Z_c = 2$ et $Z_b = 1/3$

$$\mathcal{T} = (\varepsilon + Z_b \times \mathrm{Seq}(Z_c)) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon)$$

Phase I : Complexité de Boltzmann

Hypothèse : Valeurs de $\mathbf{y}$ indépendantes de n

Simplification

Classe combinatoire + Paramètres $\mathbf{y} \approx$ Classe combinatoire

Pour des valeurs rationnelles :

- Distribution stable par multiplication des valeurs par un réel
- Multiplication par PPCM des dénominateurs $\Rightarrow$ Valeurs entières
- Duplication des sous-expressions $\Rightarrow$ Spécification standard

Ex : $Z_a = 2$, $Z_{a'} = 3$ $Z_c = 6$ et $Z_b = 1$

$$\mathcal{T} = (\varepsilon + Z_b \times \mathrm{Seq}(Z_c)) \times (Z_a \times \mathcal{T} \times Z_{a'} \times \mathcal{T} + \varepsilon)$$

Phase I : Complexité de Boltzmann

Hypothèse : Valeurs de $\mathbf{y}$ indépendantes de n

Simplification

Classe combinatoire + Paramètres $\mathbf{y} \approx$ Classe combinatoire

Pour des valeurs rationnelles :

- Distribution stable par multiplication des valeurs par un réel
- Multiplication par PPCM des dénominateurs $\Rightarrow$ Valeurs entières
- Duplication des sous-expressions $\Rightarrow$ Spécification standard

Ex : $Z_a = 2$, $Z_{a'} = 3$ $Z_c = 6$ et $Z_b = 1$

$$\mathcal{T} = (\varepsilon + Z_b \times \mathrm{Seq}(Z_c + Z_c + Z_c + Z_c + Z_c + Z_c)) \times ((Z_a + Z_a) \times \mathcal{T} \times (Z_{a'} + Z_{a'} + Z_{a'}) \times \mathcal{T} + \varepsilon)$$

Phase I : Complexité de Boltzmann

Hypothèse : Valeurs de $\mathbf{y}$ indépendantes de n

Simplification

Classe combinatoire + Paramètres $\mathbf{y} \approx$ Classe combinatoire

$\Rightarrow$ Génération en $\mathcal{O}(n^2)$ en taille exacte
$\mathcal{O}(n)$ avec tolérance ε.

Remarques :

- Influence des valeurs des paramètres $\mathbf{y}$
- Quid des valeurs des paramètres dépendantes de n ?
- Tirage + Phase I aussi obtenus par la méthode récursive [DeRoTe00]

Phase I : Complexité de Boltzmann

Hypothèse : Valeurs de $\mathbf{y}$ indépendantes de n

Simplification

Classe combinatoire + Paramètres $\mathbf{y} \approx$ Classe combinatoire

$\Rightarrow$ Génération en $\mathcal{O}(n^2)$ en taille exacte
$\mathcal{O}(n)$ avec tolérance ε.

Remarques :

- Influence des valeurs des paramètres $\mathbf{y}$
- Quid des valeurs des paramètres dépendantes de n ?
- Tirage + Phase I aussi obtenus par la méthode récursive [DeRoTe00]

Phase I : Complexité de Boltzmann

Hypothèse : Valeurs de $\mathbf{y}$ indépendantes de n

Simplification

Classe combinatoire + Paramètres $\mathbf{y} \approx$ Classe combinatoire

$\Rightarrow$ Génération en $\mathcal{O}(n^2)$ en taille exacte
$\mathcal{O}(n)$ avec tolérance ε.

Remarques :

- Influence des valeurs des paramètres $\mathbf{y}$
- Quid des valeurs des paramètres dépendantes de n ?
- Tirage + Phase I aussi obtenus par la méthode récursive [DeRoTe00]

Calcul des pondérations

But : Viser des fréquences $\mathbf{F} = F_1, \ldots, F_k$ pour les atomes.

Inversion de la relation paramêtres/espérances

- Plus proche des valeurs maximisant proba d'acceptation
- Adaptation *naturelle* à des cas pathologiques

Mais symbolique coûte cher $\Rightarrow$ Numérique

Asymptotiquement

En sing. dom. $\rho(y_1, \ldots, y_n)$,analyse de singularité

$\Rightarrow$ Système *simple* $\left\{ F_i = \frac{-\frac{\partial \rho(\boldsymbol{y_1}, \ldots, \boldsymbol{u} \cdot \boldsymbol{y_i}, \ldots, \boldsymbol{y_k})}{\partial \boldsymbol{u}}}{\boldsymbol{n} \rho(\boldsymbol{y_1}, \ldots, \boldsymbol{y_k})} \right\}_{i=1}^{k}$

Résoudre en $\mathbf{y}$ garantit des fréquences $\mathbf{F}$ à l'asympotique.

Mais problème quand plusieurs singularités en compétition.

ARN : Phase II – Pondérations asymptotiques

- Déterminer la singularité dominante

$$\rho(a, b, c) = \frac{c + 2\,a - \sqrt{c^2 - 4\,ca + 4\,a^2 + 8\,ab}}{4a\,(b - c)}$$

- Construire et résoudre la relation paramètres/fréquences ($F_a = n_a/n$, $F_b = n_b/n$, $F_c = 1 - (F_a + F_b)$)

$$y_c = \frac{y_b\,(F_a - F_b)\,(1 - (F_a + F_b))}{{F_b}^2}$$

$$y_a = \frac{y_b\,(1 - F_a)\,(F_a - F_b)^2}{2F_a\,{F_b}^2}$$

- Utiliser les valeurs dans le générateur et laisser tourner ...

ARN : Phase II – Pondérations asymptotiques

- Déterminer la singularité dominante

$$\rho(a, b, c) = \frac{c + 2\,a - \sqrt{c^2 - 4\,ca + 4\,a^2 + 8\,ab}}{4a\,(b - c)}$$

- Construire et résoudre la relation paramètres/fréquences ($F_a = n_a/n$, $F_b = n_b/n$, $F_c = 1 - (F_a + F_b)$)

$$\begin{aligned} y_c &= \frac{y_b\,(F_a - F_b)\,(1 - (F_a + F_b))}{{F_b}^2} \\ y_a &= \frac{y_b\,(1 - F_a)\,(F_a - F_b)^2}{2F_a\,{F_b}^2} \end{aligned}$$

- Utiliser les valeurs dans le générateur et laisser tourner ...

ARN : Phase II – Pondérations asymptotiques

- Déterminer la singularité dominante

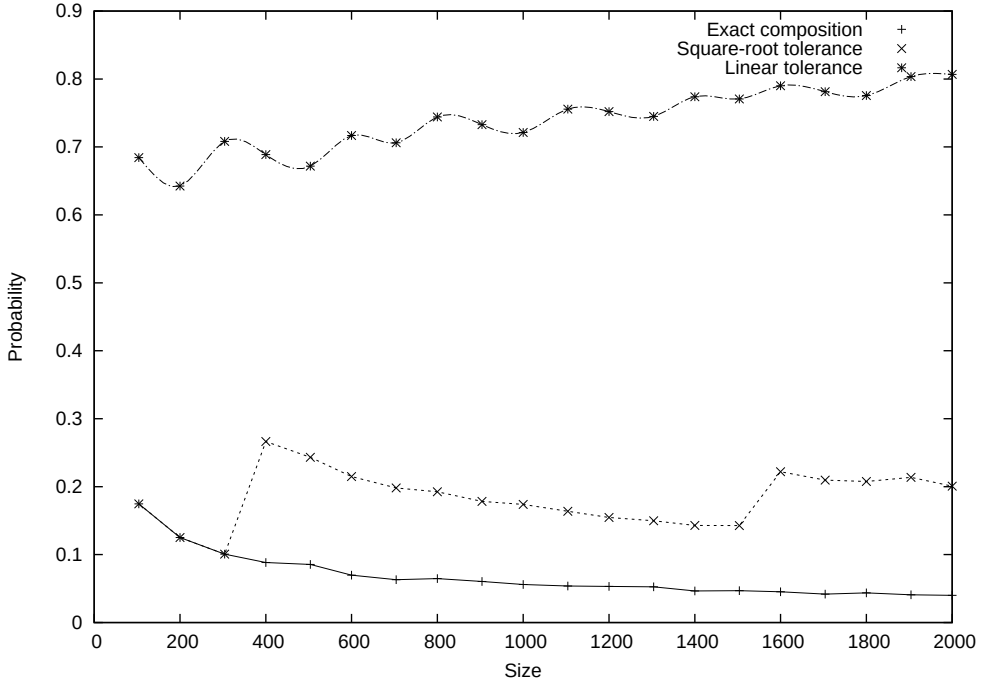
$$\rho(a, b, c) = \frac{c + 2\,a - \sqrt{c^2 - 4\,ca + 4\,a^2 + 8\,ab}}{4a\,(b - c)}$$

- Construire et résoudre la relation paramètres/fréquences ($F_a = n_a/n$, $F_b = n_b/n$, $F_c = 1 - (F_a + F_b)$)

$$\begin{aligned} y_c &= \frac{y_b\,(F_a - F_b)\,(1 - (F_a + F_b))}{{F_b}^2} \\ y_a &= \frac{y_b\,(1 - F_a)\,(F_a - F_b)^2}{2F_a\,{F_b}^2} \end{aligned}$$

- Utiliser les valeurs dans le générateur et laisser tourner ...

ARN : Phase II – Rejet avec tolérance



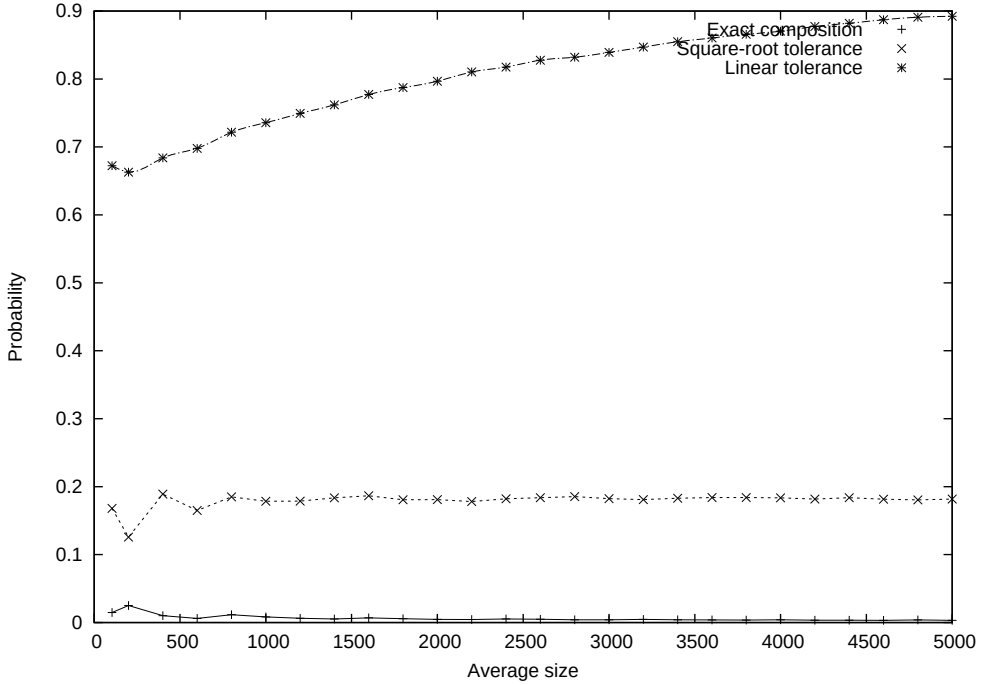
Probabilités d'acceptation selon la tolérance :

Haut Linéaire : $n_i^* \in [n_i - n_i\varepsilon, n_i + n_i\varepsilon]$

Milieu Racine : $n_i^* \in [n_i - \sqrt{n_i}\varepsilon, n_i + \sqrt{n_i}\varepsilon]$

Bas Zéro : $n_i^* = n_i$

ARN : Phase II – Rejet avec tolérance



Probabilités d'acceptation selon la tolérance :

Haut Linéaire : $n_i^* \in [n_i - n_i\varepsilon, n_i + n_i\varepsilon]$

Milieu Racine : $n_i^* \in [n_i - \sqrt{n_i}\varepsilon, n_i + \sqrt{n_i}\varepsilon]$

Bas Zéro : $n_i^* = n_i$

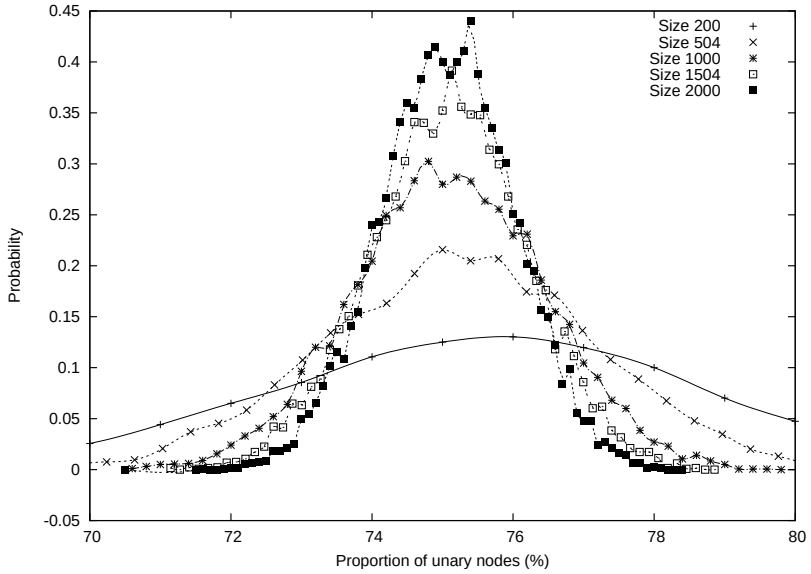
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ *Moyennes/variances linéaires [Drmota97] [FlaSed09] et rejet :*

- $\Theta(n^{k/2})$ *en composition exacte*
- $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



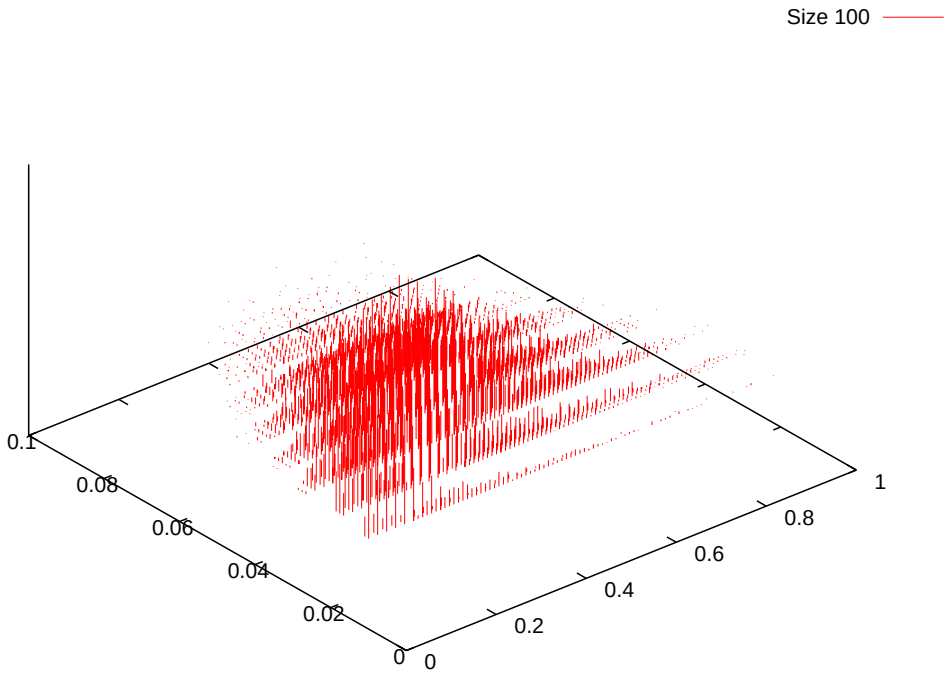
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



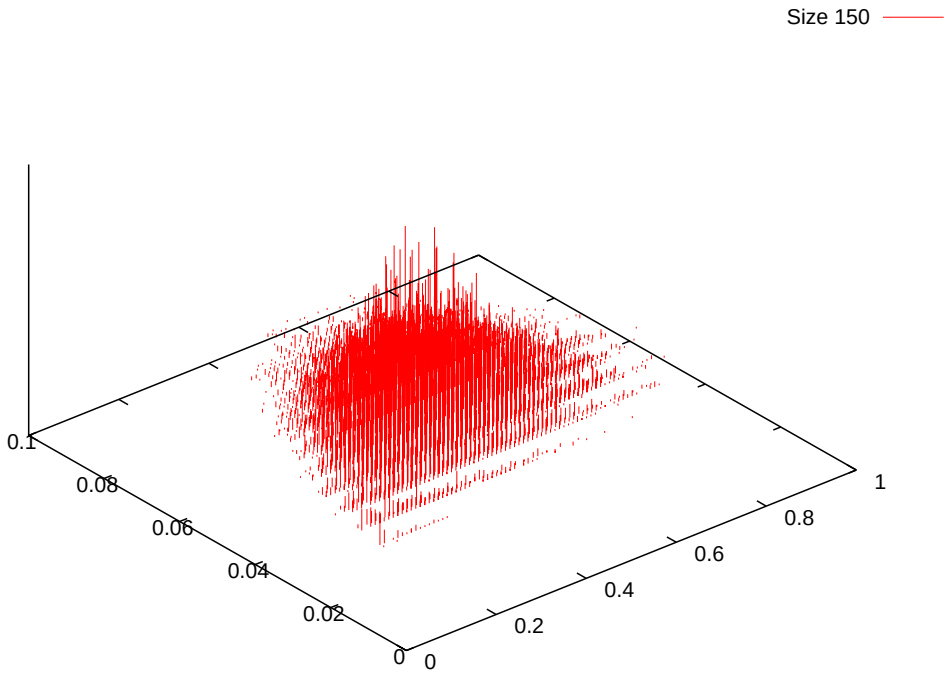
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



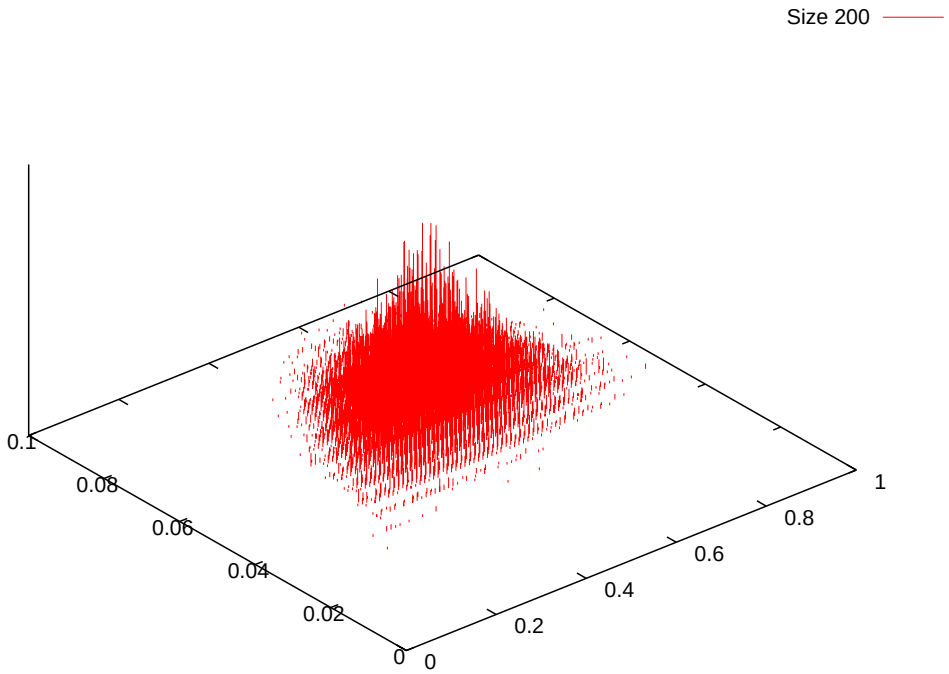
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



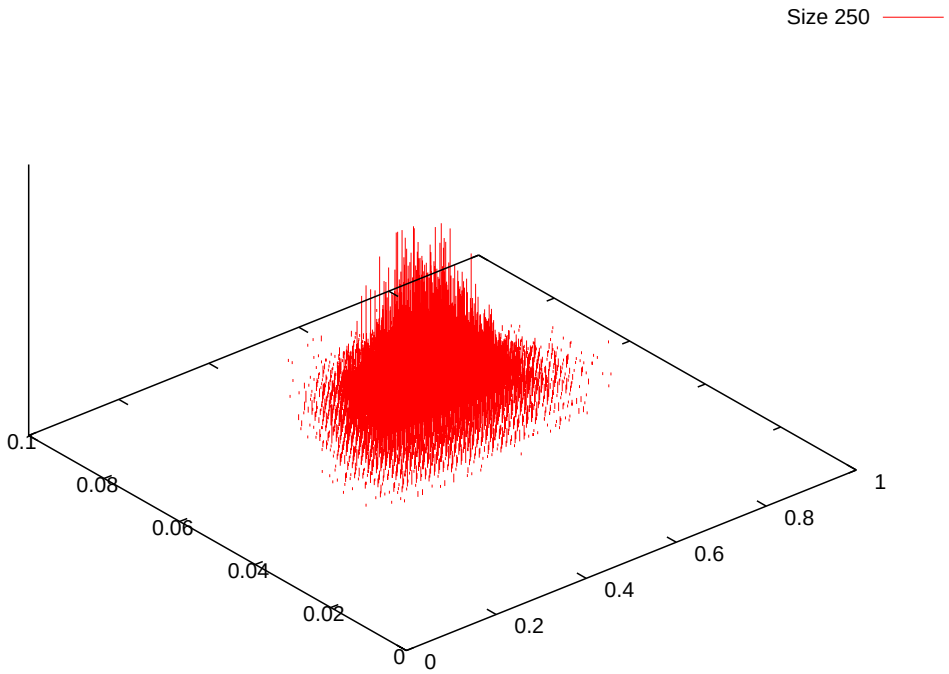
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



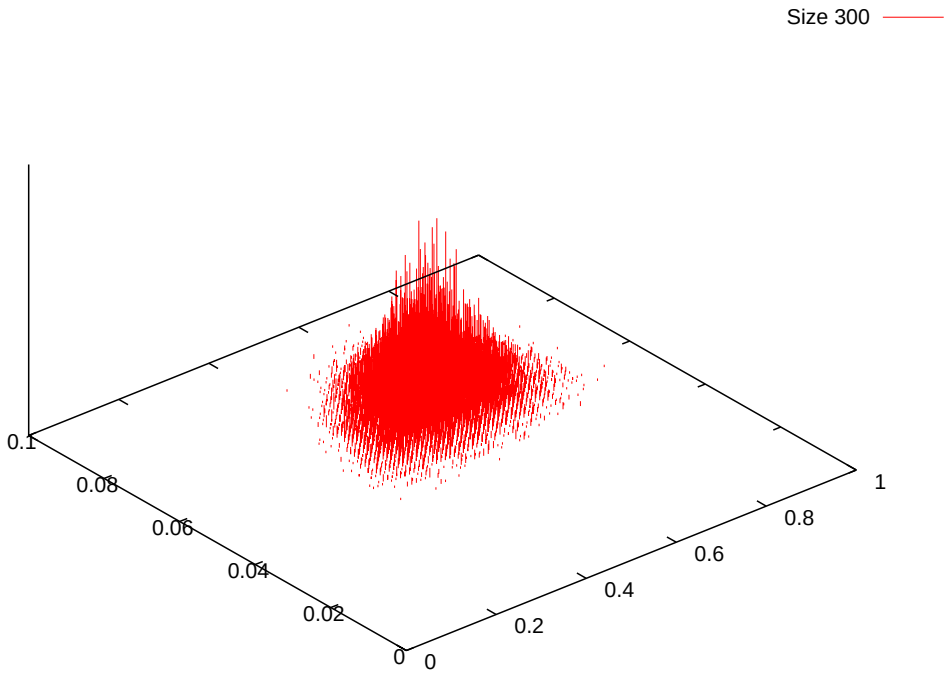
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



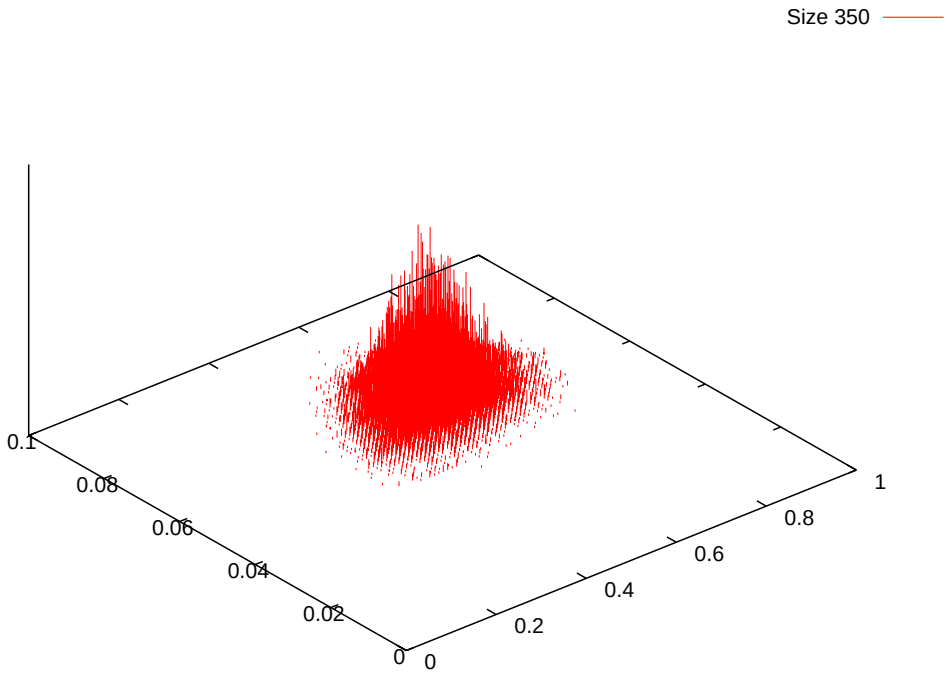
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



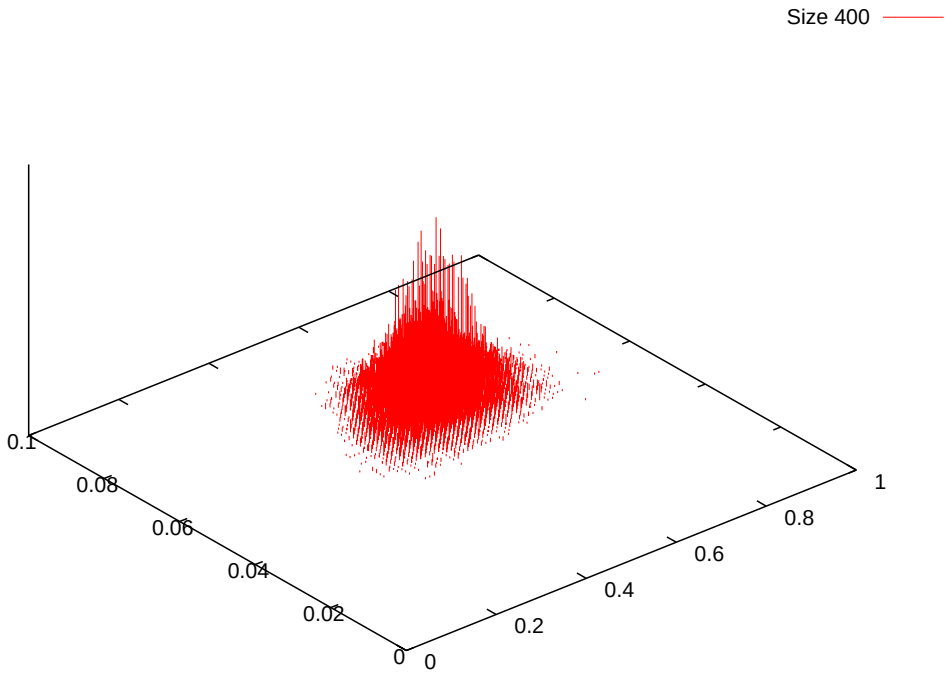
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ *+ Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



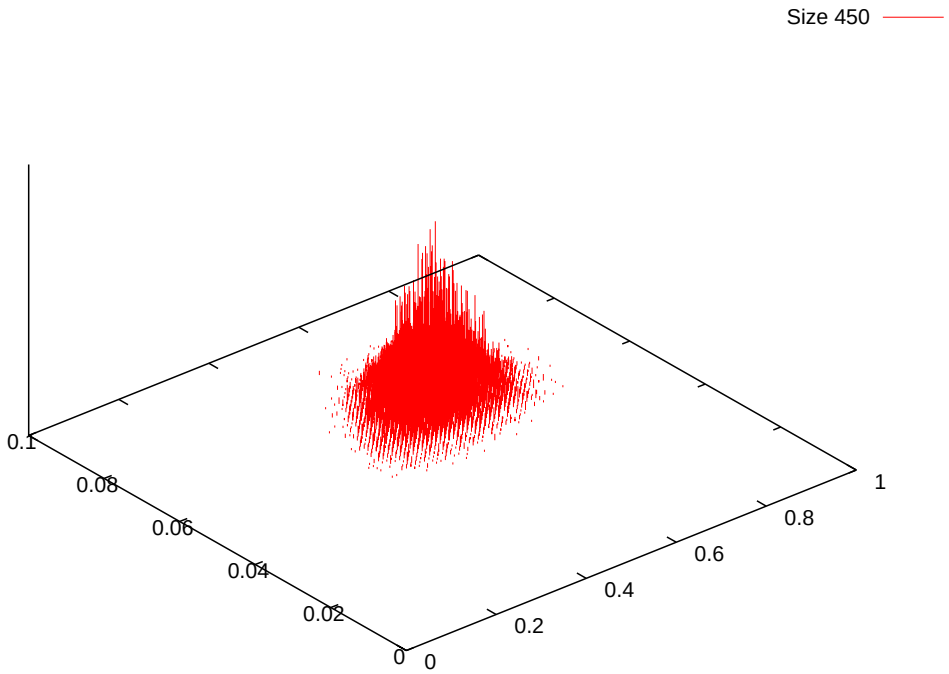
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



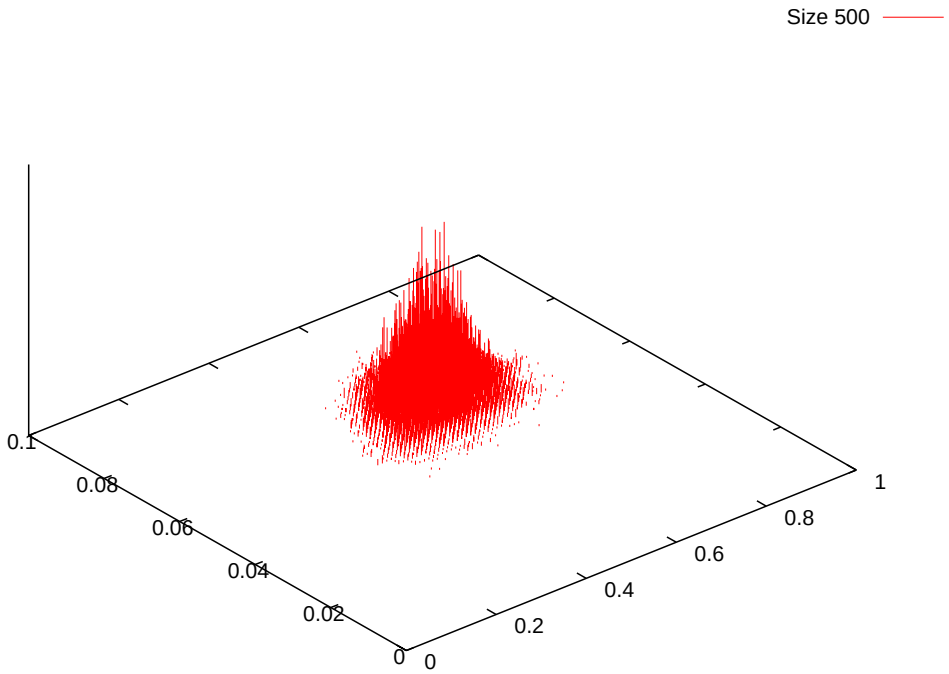
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



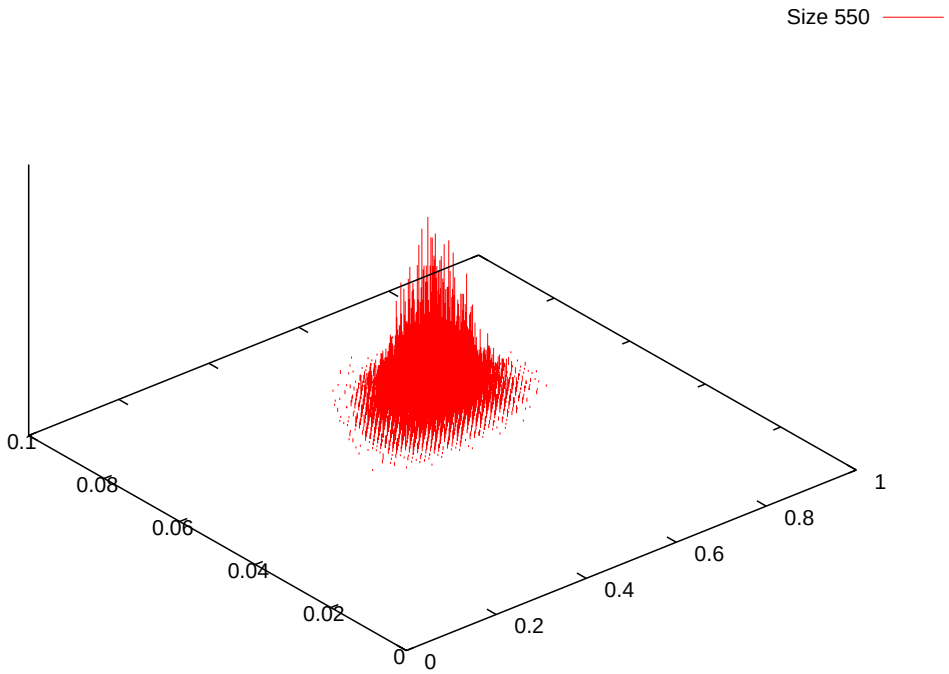
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



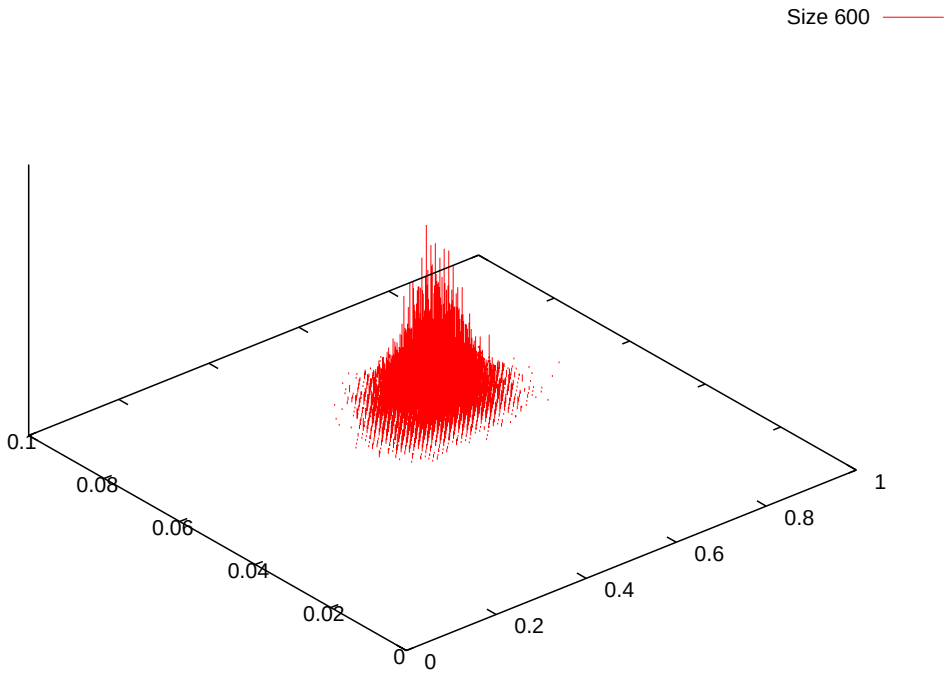
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



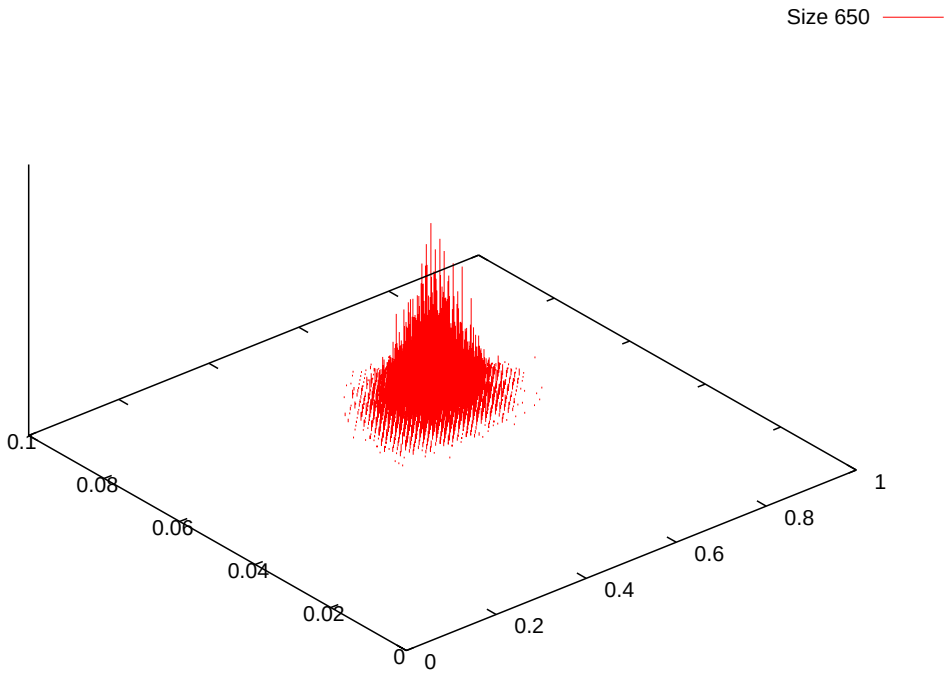
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



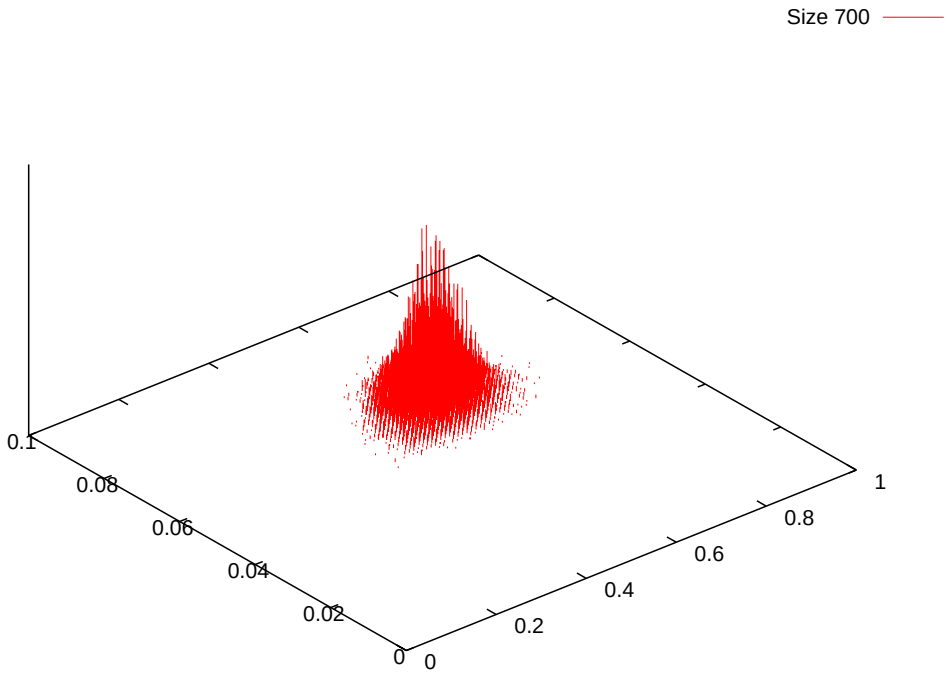
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



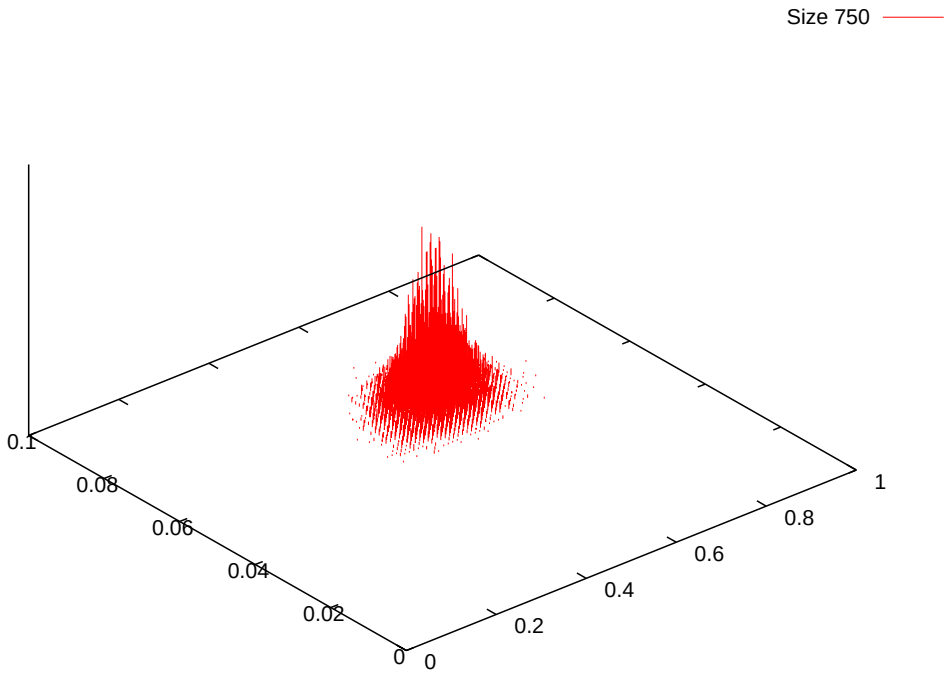
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



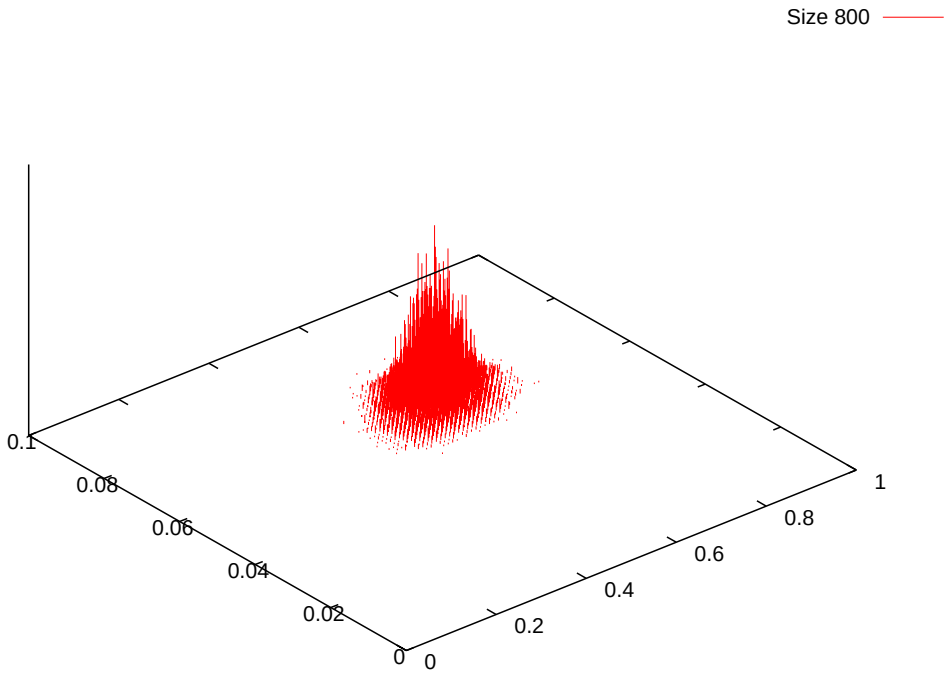
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



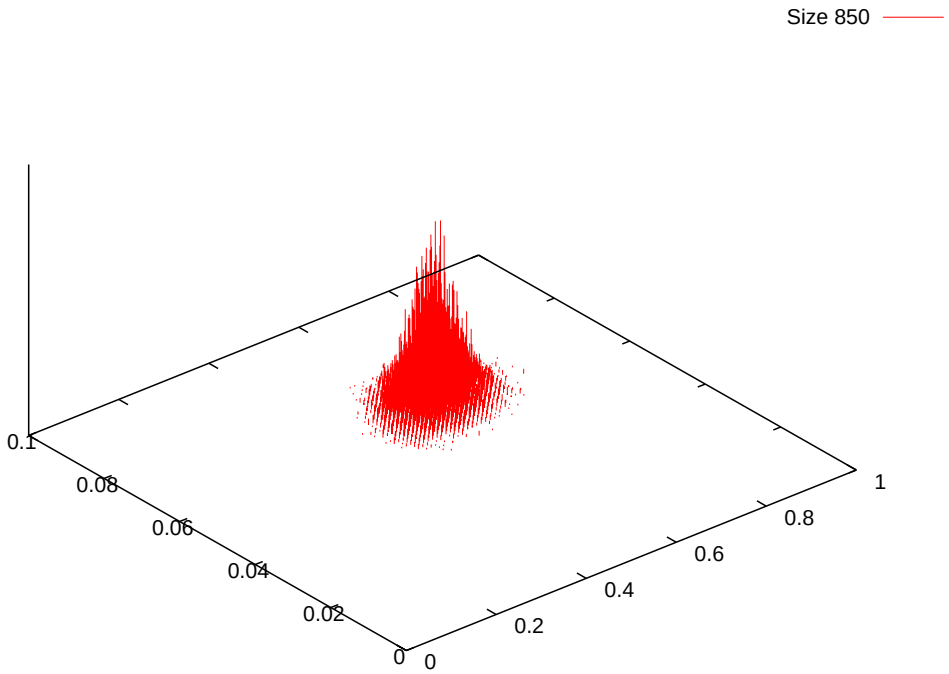
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



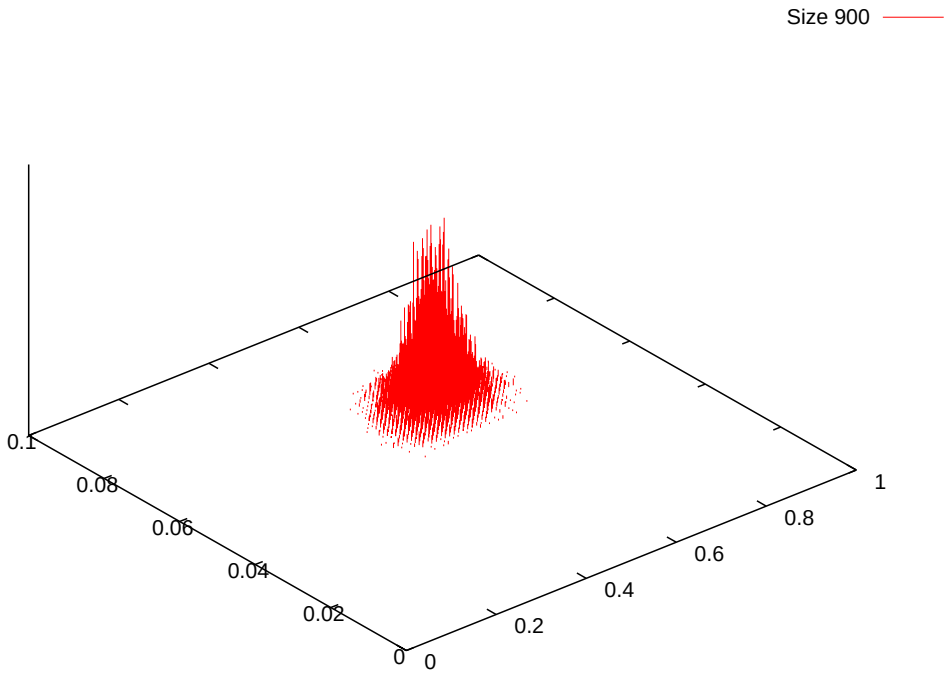
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



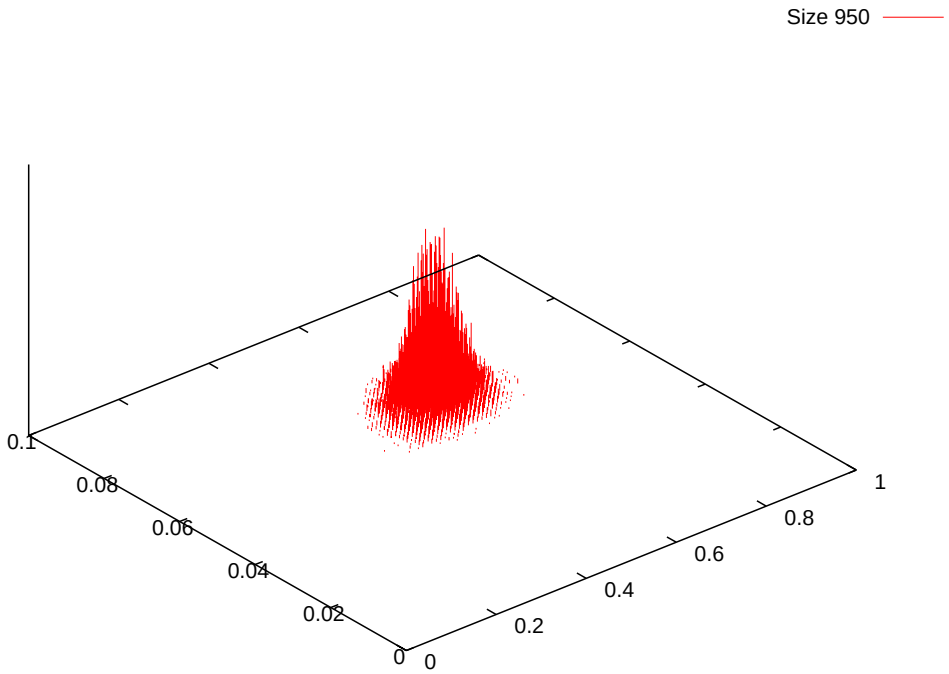
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



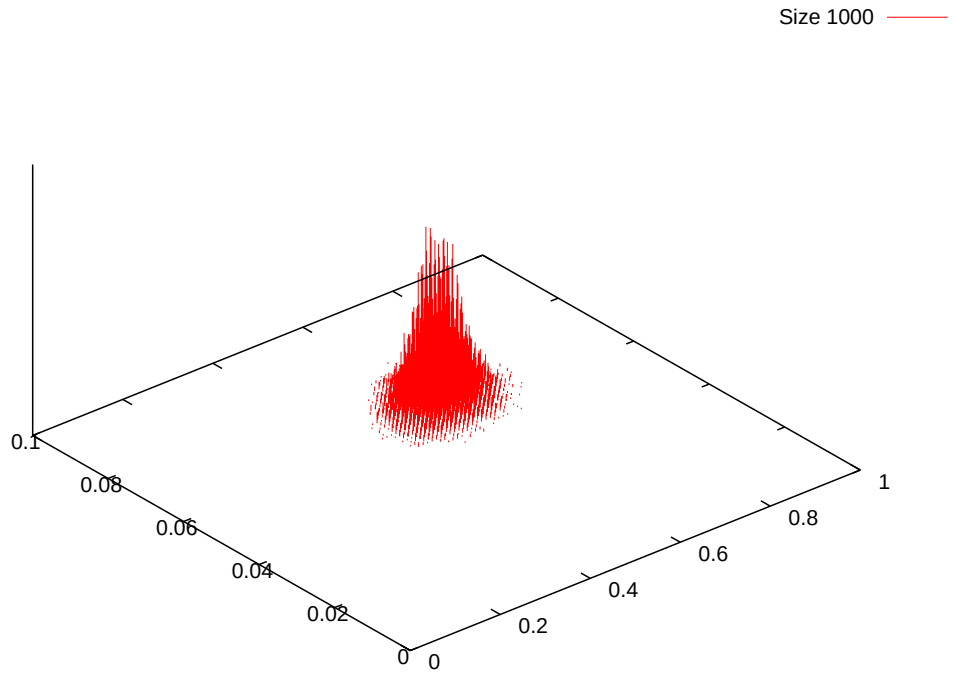
Phase II : Langages

Theorem (Complexité du rejet dans les langages)

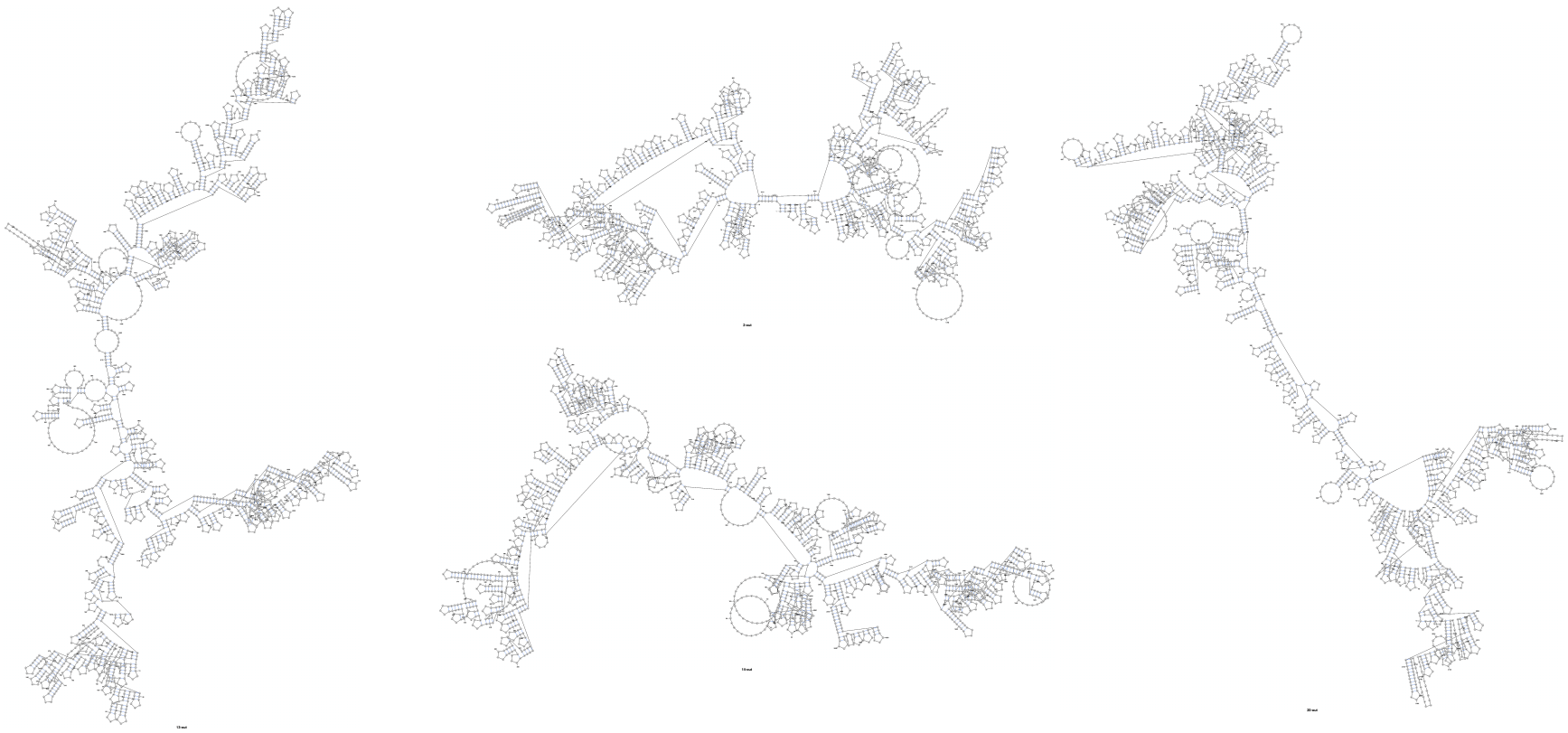
- *Opérateurs :* $\{\times, +, Sequence\}$ + *Récurrence (Langages)*
- *Grammaire fortement connexe/apériodique (Type simple)*
- *Fréquences visées non pathologiques :* $n_i = nF_i$

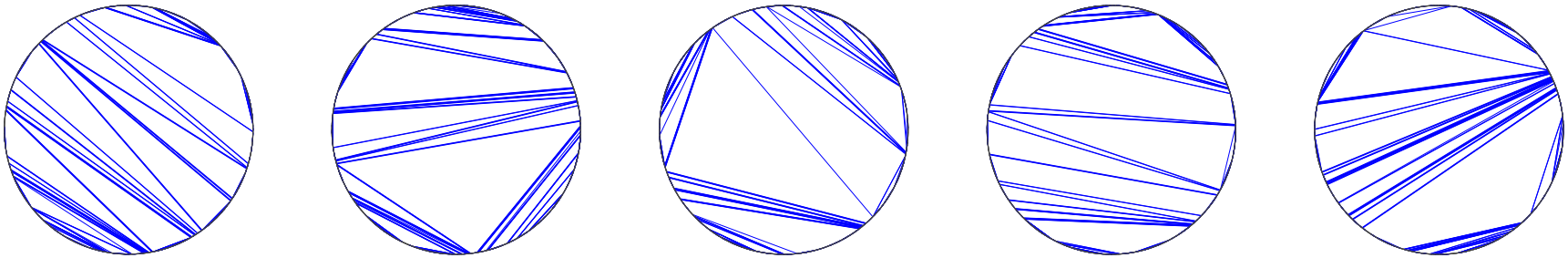
$\Rightarrow$ $\Theta(n^{k/2})$ *en composition exacte*

$\Rightarrow$ $\Theta(1)$ *avec une tolérance en* $\sqrt{n}$ *sur les différents paramètres*



Exemple : ARN





Approches existantes

n : Taille

m : Nombre d'échantillons

Récursive naive [DenRoqTer00]

Adaptation de la méthode récursive pour langages non-contextuels.
Complexités en $\mathcal{O}(mn^k)$ ou $\mathcal{O}(n^{2k} + mnlog(n))$

Manipulation de séries holonomes [RadMas04]

Cas rationnel bivarié, génération en $\mathcal{O}(mn)$.
Cas rationnel multivarié (k types d'atomes) en $\mathcal{O}(mn^{k-1})$.

Boltzmann multivarié

Non-contextuel : $\mathcal{O}(mn^{2+k/2})$ en taille et composition exacte,
$\mathcal{O}(mn)$ pour tolérances εn sur la taille et $\varepsilon' \sqrt{n}$ sur les différents atomes.

Approches existantes

n : Taille

m : Nombre d'échantillons

Récursive naive [DenRoqTer00]

Adaptation de la méthode récursive pour langages non-contextuels.
Complexités en $\mathcal{O}(mn^k)$ ou $\mathcal{O}(n^{2k} + mnlog(n))$

Manipulation de séries holonomes [RadMas04]

Cas rationnel bivarié, génération en $\mathcal{O}(mn)$.
Cas rationnel multivarié (k types d'atomes) en $\mathcal{O}(mn^{k-1})$.

Boltzmann multivarié

Non-contextuel : $\mathcal{O}(mn^{2+k/2})$ en taille et composition exacte,
$\mathcal{O}(mn)$ pour tolérances εn sur la taille et $\varepsilon' \sqrt{n}$ sur les différents atomes.

Approches existantes

n : Taille

m : Nombre d'échantillons

Récursive naive [DenRoqTer00]

Adaptation de la méthode récursive pour langages non-contextuels.
Complexités en $\mathcal{O}(mn^k)$ ou $\mathcal{O}(n^{2k} + mnlog(n))$

Manipulation de séries holonomes [RadMas04]

Cas rationnel bivarié, génération en $\mathcal{O}(mn)$.
Cas rationnel multivarié (k types d'atomes) en $\mathcal{O}(mn^{k-1})$.

Boltzmann multivarié

Non-contextuel : $\mathcal{O}(mn^{2+k/2})$ en taille et composition exacte,
$\mathcal{O}(mn)$ pour tolérances εn sur la taille et $\varepsilon' \sqrt{n}$ sur les différents atomes.

Conclusion

Boulons à serrer :

- Petites valeurs des n_i
  $\Rightarrow$ Modifier spécification ? Récursive ?
- Calcul des paramètres :
  Maximiser probabilité $\Leftrightarrow$ Fixer l'espérance ?
  Calcul numérique $\Rightarrow$ Hypothèse de monotonie
- Distribution des atomes dans des spécifications non-connexes :
  Concurrence de singularité ?
- Spécifications fortement connexes :
  Stabilité de la singularité dominante ?
- Extension à autres opérateurs ?

Conclusion

Boulons à serrer :

- Petites valeurs des n_i
  $\Rightarrow$ Modifier spécification ? Récursive ?
- Calcul des paramètres :
  Maximiser probabilité $\Leftrightarrow$ Fixer l'espérance ?
  Calcul numérique $\Rightarrow$ Hypothèse de monotonie
- Distribution des atomes dans des spécifications non-connexes :
  Concurrence de singularité ?
- Spécifications fortement connexes :
  Stabilité de la singularité dominante ?
- Extension à autres opérateurs ?

Conclusion

Boulons à serrer :

- Petites valeurs des n_i
  $\Rightarrow$ Modifier spécification ? Récursive ?
- Calcul des paramètres :
  Maximiser probabilité $\Leftrightarrow$ Fixer l'espérance ?
  Calcul numérique $\Rightarrow$ Hypothèse de monotonie
- Distribution des atomes dans des spécifications non-connexes :
  Concurrence de singularité ?
- Spécifications fortement connexes :
  Stabilité de la singularité dominante ?
- Extension à autres opérateurs ?

Conclusion

Boulons à serrer :

- Petites valeurs des n_i
  $\Rightarrow$ Modifier spécification ? Récursive ?
- Calcul des paramètres :
  Maximiser probabilité $\Leftrightarrow$ Fixer l'espérance ?
  Calcul numérique $\Rightarrow$ Hypothèse de monotonie
- Distribution des atomes dans des spécifications non-connexes :
  Concurrence de singularité ?
- Spécifications fortement connexes :
  Stabilité de la singularité dominante ?
- Extension à autres opérateurs ?

Conclusion

Boulons à serrer :

- Petites valeurs des n_i
  $\Rightarrow$ Modifier spécification ? Récursive ?
- Calcul des paramètres :
  Maximiser probabilité $\Leftrightarrow$ Fixer l'espérance ?
  Calcul numérique $\Rightarrow$ Hypothèse de monotonie
- Distribution des atomes dans des spécifications non-connexes :
  Concurrence de singularité ?
- Spécifications fortement connexes :
  Stabilité de la singularité dominante ?
- Extension à autres opérateurs ?