Grammaires stochastiques ou quantiques en génomique

Dimitri PETRITIS

Institut de Recherche Mathématique

Université de Rennes 1 et CNRS (UMR 6625)

France

dimitri.petritis@univ-rennes1.fr

http://name.math.univ-rennes1.fr/dimitri.petritis

“Vous visitez un musée de peinture, et vous vous promenez parmi les toiles françaises du début de XXe siècle. Vous voyez ici un Renoir somptueux, là sans erreur possible c’est un Modigliani, là encore ce sont de fleurs peintes par van Gogh, ou des fruits de Cézanne. Plus loin vous appercevez un Picasso, à moins que ce ne soit un Braque. Sans doute est-ce la première fois que vous voyez ces peintures, mais vous n’avez le plus souvent aucun doute quant à l’artiste à qui on les doit.[...] Comment les distinguez vous? Eh bien la peinture n’est pas appliquée de la même manière, et les sujets sont différents. Mais il y a quelque chose d’autre, qui est plus difficile à exprimer et que cependant on saisit immédiatement, un quelque chose qui dépend du choix des formes et de l’équilibre de couleurs.”

–David RUELLE, Hasard et chaos

Quelques observations

- **Première observation:** Une séquence de symboles codant message avec contenu relatif à un individu contient la signature statistique de l'individu.

Quelques observations

- **Première observation:** Une séquence de symboles codant message avec contenu relatif à un individu contient la signature statistique de l'individu.
- "Zij $\omega(\cdot)$ een lineaire afbeelding van een reële Hilbertruimte $\mathcal{K}$ naar de ruimte van stochasten op een kansruimte met eindige momenten."

Quelques observations

- **Première observation:** Une séquence de symboles codant message avec contenu relatif à un individu contient la signature statistique de l'individu.
- "Zij $\omega(\cdot)$ een lineaire afbeelding van een reële Hilbertruimte $\mathcal{K}$ naar de ruimte van stochasten op een kansruimte met eindige momenten."
- **Seconde observation:** Confiance en la signature statistique renforcée lorsque message concerne large classe d'individus.

Queulques observations

- **Troisème observation:** Au delà de variété combinatoire des phrases individuelles d'une classe, il existe une structure commune sous-jacente, une *grammaire*, permettant la distinction entre phrases grammaticallement correctes et fausses et nous aidant à donner un sens:

Qeulques observations

- **Troisème observation:** Au delà de variété combinatoire des phrases individuelles d'une classe, il existe une structure commune sous-jacente, une *grammaire*, permettant la distinction entre phrases grammaticallement correctes et fausses et nous aidant à donner un sens:
  - Un crocodile mord un homme (grammaticalement correcte)

Qeulques observations

- **Troisème observation:** Au delà de variété combinatoire des phrases individuelles d'une classe, il existe une structure commune sous-jacente, une *grammaire*, permettant la distinction entre phrases grammaticallement correctes et fausses et nous aidant à donner un sens:
  - Un crocodile mord un homme (grammaticalement correcte)
  - Un homme mord un crocodile (grammaticalement correcte)

Qeulques observations

- **Troisème observation:** Au delà de variété combinatoire des phrases individuelles d'une classe, il existe une structure commune sous-jacente, une *grammaire*, permettant la distinction entre phrases grammaticallement correctes et fausses et nous aidant à donner un sens:
  - Un crocodile mord un homme (grammaticalement correcte)
  - Un homme mord un crocodile (grammaticalement correcte)
  - Un crocodile un homme mord (grammaticalement fausse)

Qeulques observations

- **Troisème observation:** Au delà de variété combinatoire des phrases individuelles d'une classe, il existe une structure commune sous-jacente, une *grammaire*, permettant la distinction entre phrases grammaticallement correctes et fausses et nous aidant à donner un sens:
  - Un crocodile mord un homme (grammaticalement correcte)
  - Un homme mord un crocodile (grammaticalement correcte)
  - Un crocodile un homme mord (grammaticalement fausse)
- **Quatrième observation:** Une *grammaire* est un *ensemble fini* de règles permettant formation et compréhenseion d'une infinité de phrases.

Un résumé de biologie moléculaire

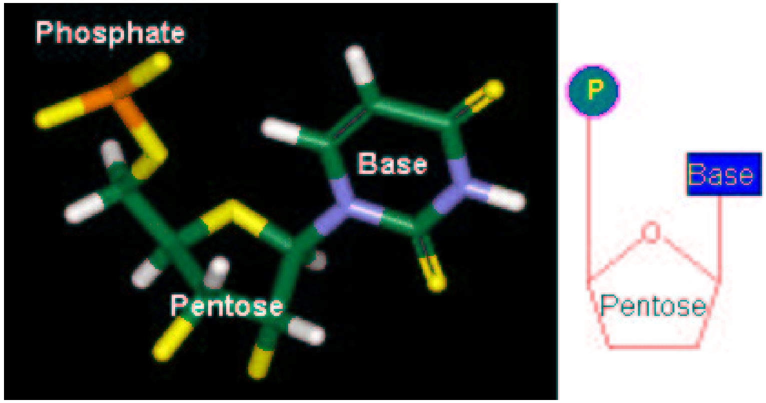
- Dans la cellule il y a 2 types de molécules:
  - les courtes: sucres, acides aminés, nucléotides, etc.
  - les longues: ADN (acide desoxyribonucleïque), ARN (acide ribonucleïque) et les protéines
- ADN code la représentation des protéines
- Dogme central de la biologie: l'information se transmet suivant les flèches:

$$\circlearrowleft \text{ADN} \longrightarrow \text{ARN} \longrightarrow \text{PROTÉINE}$$

appeées *replication, transcription et traduction.*

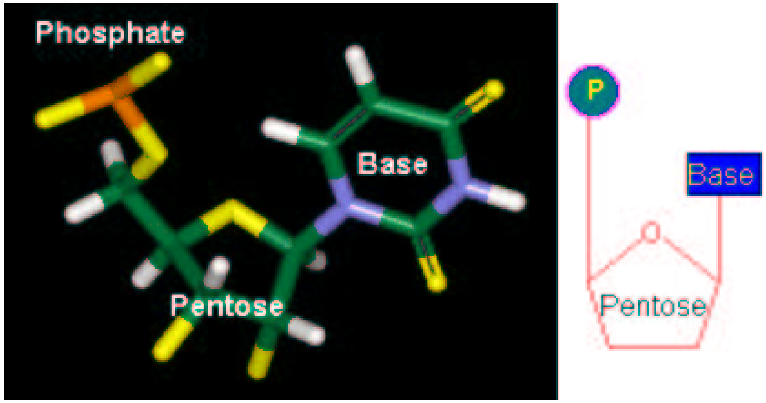
La structure de la molécule d'ADN

- Nucléotides: *E. coli* 5×10^6; *H. sapiens* 3×10^9

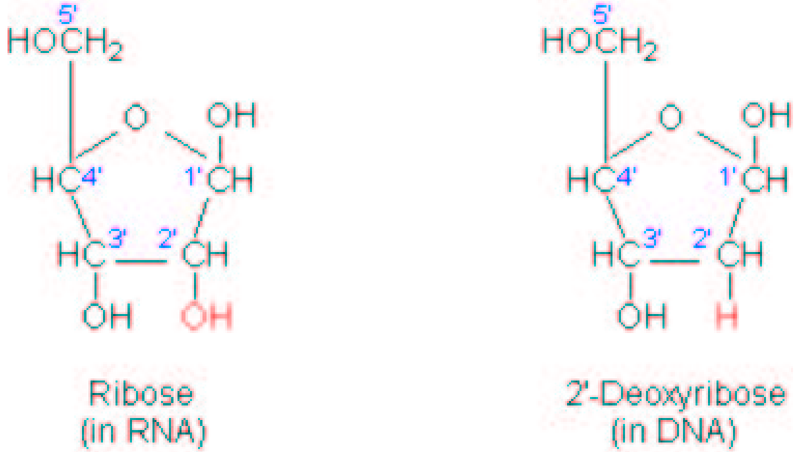


La structure de la molécule d'ADN

- Nucléotides: *E. coli* 5×10^6; *H. sapiens* 3×10^9

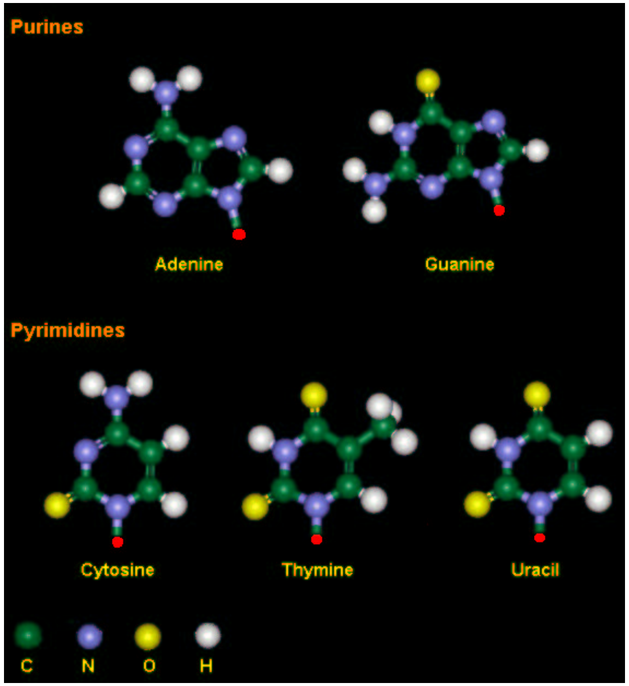


- Squelette de pentoses



La structure de la molécule d'ADN

- Bases

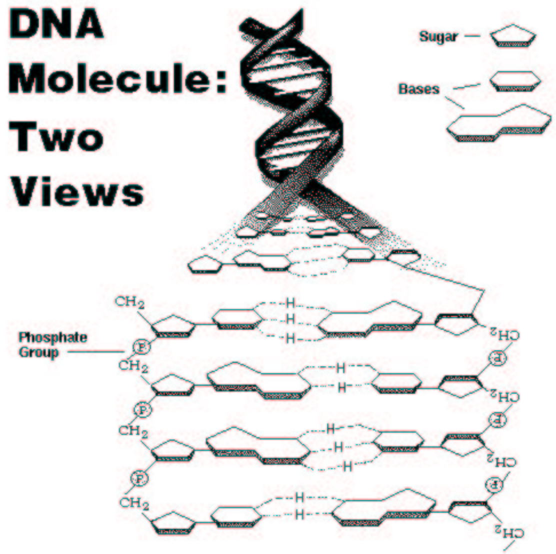


$B = \{A, C, G, T(U)\}$; G-C et A-T(U) complémentaires

- L'ensemble $B = \{A, C, G, T(U)\}$ indexe les nucléeotides

La structure de la molécule d'ADN

- ADN long polymère branché. Stéreo-chimiquement: double hélice. Distance entre monomères 0.34 nm



Cinquième observation: L'ADN est une molécule chimique standard. Covalence des liens et distance interbase typique indiquent qu'interactions quantiques ne sont pas négligeables.

Le code génétique

- A ensemble de 20 acides aminés

Acide	nnn	n	Acide	nnn	n
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Acide aspartique	Asp	D	Méthionine	Met	M
Asparginine	Asn	N	Pheńylanine	Phe	F
Cystéine	Cys	C	Proline	Pro	P
Acide glutique acid	Glu	E	Sérine	Ser	S
Glutamine	Gln	Q	Thréonine	Thr	T
Glycine	Gly	G	Tryptophane	Trp	W
Histine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

- 3 nucléotides, le *codon*, pour coder un acide aminé
- Direction de lecture $5' \to 3'$

Le code génétique

Code génétique: $g : B^3 \to A \cup \{\textsc{stop}\}$. Codon début: AUG

1st	2nd				3rd	1st	2nd				3rd
	U	C	A	G			U	C	A	G	
U	Phe	Ser	Tyr	Cys	U	A	Ile	Thr	Asn	Ser	U
U	Phe	Ser	Tyr	Cys	C	A	Ile	Thr	Asn	Ser	C
U	Leu	Ser	STOP	STOP	A	A	Ile	Thr	Lys	Arg	A
U	Leu	Ser	STOP	Trp	G	A	Met	Thr	Lys	Arg	G
C	Leu	Pro	His	Arg	U	G	Val	Ala	Asp	Glu	U
C	Leu	Pro	His	Arg	C	G	Val	Ala	Asp	Glu	C
C	Leu	Pro	Gln	Arg	A	G	Val	Ala	Glu	Glu	A
C	Leu	Pro	Gln	Arg	G	G	Val	Ala	Glu	Glu	G

Gènes et génome

- Gène: région d'ADN impliquée dans synthèse d'une protéine
- Gène séparé en régions codantes (*exons*) et non codantes (*introns*)
  - 5% ADN "utile" (contient information pour synthèse des protéines)
  - 95% ADN "poubelle" (signature de l'individu)
- Génome: ADN contenu dans la cellule. Longueur de la molécule d'ADN $= 10^5$ diamètres cellulaires!

Dernière observation et une thèse

Sixième observation: L'ADN de chaque individu est une collection de phrases ayant un sens précis.

Une thèse: Il existe un grammaire sous-jacente (probablement quantique) pour le langage d'ADN.

- Formalisme mathématiquement cohérent
- Inventer d'algorithmes et expériences *in vivo* pour vérifier pertinence biologique de la thèse

Grammaires et langages

- $\mathbb{A}_t$ alphabet fini de *symboles terminaux*
  $\mathbb{A}_{sc}$ alphabet fini de *catégories syntaxiques*
  $\mathbb{A}_t \cap \mathbb{A}_{sc} = \emptyset$, $\mathbb{A} = \mathbb{A}_t \cup \mathbb{A}_{sc}$, $\mathbb{A}^* = \cup_{n \geq 0} \mathbb{A}^n$

Grammaires et langages

- $\mathbb{A}_t$ alphabet fini de *symboles terminaux*
  $\mathbb{A}_{sc}$ alphabet fini de *catégories syntaxiques*
  $\mathbb{A}_t \cap \mathbb{A}_{sc} = \emptyset$, $\mathbb{A} = \mathbb{A}_t \cup \mathbb{A}_{sc}$, $\mathbb{A}^* = \cup_{n \geq 0} \mathbb{A}^n$
- $\Pi \subset (\mathbb{A}^+ \setminus \mathbb{A}_t^*) \times \mathbb{A}^*$ *productions*

Grammaires et langages

- $\mathbb{A}_t$ alphabet fini de *symboles terminaux*
  $\mathbb{A}_{sc}$ alphabet fini de *catégories syntaxiques*
  $\mathbb{A}_t \cap \mathbb{A}_{sc} = \emptyset$, $\mathbb{A} = \mathbb{A}_t \cup \mathbb{A}_{sc}$, $\mathbb{A}^* = \cup_{n \geq 0} \mathbb{A}^n$
- $\Pi \subset (\mathbb{A}^+ \setminus \mathbb{A}_t^*) \times \mathbb{A}^*$ *productions*
- $s_0 \in \mathbb{A}_{sc}$ symbole initial

Grammaires et langages

- $\mathbb{A}_t$ alphabet fini de *symboles terminaux*
  $\mathbb{A}_{sc}$ alphabet fini de *catégories syntaxiques*
  $\mathbb{A}_t \cap \mathbb{A}_{sc} = \emptyset$, $\mathbb{A} = \mathbb{A}_t \cup \mathbb{A}_{sc}$, $\mathbb{A}^* = \cup_{n \geq 0} \mathbb{A}^n$
- $\Pi \subset (\mathbb{A}^+ \setminus \mathbb{A}_t^*) \times \mathbb{A}^*$ *productions*
- $s_0 \in \mathbb{A}_{sc}$ symbole initial
- *Grammaire:* $\Gamma = (\mathbb{A}_t, \mathbb{A}_{sc}, \Pi, s_0)$

Grammaires et langages

- $\mathbb{A}_t$ alphabet fini de *symboles terminaux*
  $\mathbb{A}_{sc}$ alphabet fini de *catégories syntaxiques*
  $\mathbb{A}_t \cap \mathbb{A}_{sc} = \emptyset$, $\mathbb{A} = \mathbb{A}_t \cup \mathbb{A}_{sc}$, $\mathbb{A}^* = \cup_{n \geq 0} \mathbb{A}^n$
- $\Pi \subset (\mathbb{A}^+ \setminus \mathbb{A}_t^*) \times \mathbb{A}^*$ *productions*
- $s_0 \in \mathbb{A}_{sc}$ symbole initial
- *Grammaire:* $\Gamma = (\mathbb{A}_t, \mathbb{A}_{sc}, \Pi, s_0)$
- *Forme sententielle:* tout mot de $\mathbb{A}^*$

Grammaires et langages

- $\mathbb{A}_t$ alphabet fini de *symboles terminaux*
  $\mathbb{A}_{sc}$ alphabet fini de *catégories syntaxiques*
  $\mathbb{A}_t \cap \mathbb{A}_{sc} = \emptyset$, $\mathbb{A} = \mathbb{A}_t \cup \mathbb{A}_{sc}$, $\mathbb{A}^* = \cup_{n \geq 0} \mathbb{A}^n$
- $\Pi \subset (\mathbb{A}^+ \setminus \mathbb{A}_t^*) \times \mathbb{A}^*$ *productions*
- $s_0 \in \mathbb{A}_{sc}$ symbole initial
- *Grammaire:* $\Gamma = (\mathbb{A}_t, \mathbb{A}_{sc}, \Pi, s_0)$
- *Forme sententielle:* tout mot de $\mathbb{A}^*$
- Si $\alpha\beta\gamma \in \mathbb{A}^+$ et $\beta \to \delta \in \Pi$, alors $\alpha\delta\gamma \in \mathbb{A}^*$

Grammaires et langages

- $\mathbb{A}_t$ alphabet fini de *symboles terminaux*
  $\mathbb{A}_{sc}$ alphabet fini de *catégories syntaxiques*
  $\mathbb{A}_t \cap \mathbb{A}_{sc} = \emptyset$, $\mathbb{A} = \mathbb{A}_t \cup \mathbb{A}_{sc}$, $\mathbb{A}^* = \cup_{n \geq 0} \mathbb{A}^n$
- $\Pi \subset (\mathbb{A}^+ \setminus \mathbb{A}_t^*) \times \mathbb{A}^*$ *productions*
- $s_0 \in \mathbb{A}_{sc}$ symbole initial
- *Grammaire:* $\Gamma = (\mathbb{A}_t, \mathbb{A}_{sc}, \Pi, s_0)$
- *Forme sententielle:* tout mot de $\mathbb{A}^*$
- Si $\alpha\beta\gamma \in \mathbb{A}^+$ et $\beta \to \delta \in \Pi$, alors $\alpha\delta\gamma \in \mathbb{A}^*$
- *Phrase* de Γ: tout mot sans catégories syntaxiques

Grammaires et langages

- $\mathbb{A}_t$ alphabet fini de *symboles terminaux*
  $\mathbb{A}_{sc}$ alphabet fini de *catégories syntaxiques*
  $\mathbb{A}_t \cap \mathbb{A}_{sc} = \emptyset$, $\mathbb{A} = \mathbb{A}_t \cup \mathbb{A}_{sc}$, $\mathbb{A}^* = \cup_{n \geq 0} \mathbb{A}^n$
- $\Pi \subset (\mathbb{A}^+ \setminus \mathbb{A}_t^*) \times \mathbb{A}^*$ *productions*
- $s_0 \in \mathbb{A}_{sc}$ symbole initial
- *Grammaire:* $\Gamma = (\mathbb{A}_t, \mathbb{A}_{sc}, \Pi, s_0)$
- *Forme sententielle:* tout mot de $\mathbb{A}^*$
- Si $\alpha\beta\gamma \in \mathbb{A}^+$ et $\beta \to \delta \in \Pi$, alors $\alpha\delta\gamma \in \mathbb{A}^*$
- *Phrase* de Γ: tout mot sans catégories syntaxiques
- *Langage* $L(\Gamma)$: ensemble de phrases de Γ.

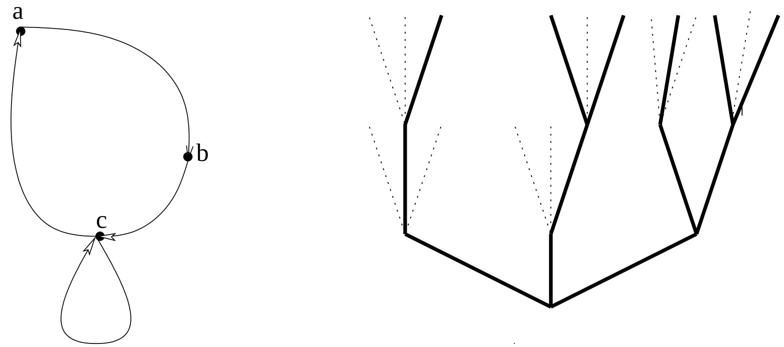
Ex. de grammaire réguilère

- $\mathbb{A}_t = \{0, 1, \ldots, 9, .\}$
- $\mathbb{A}_{sc} = \{n, f, d, i\}$
- $s_0 = n$
- Productions sous forme de Backus-Naur

$$\begin{array}{rcl} n & ::= & i \mid f \mid i\ f \\ f & ::= & .\ i \\ i & ::= & d \mid d\ i \\ d & ::= & 0 \mid 1 \mid \cdots \mid 9 \end{array}$$

- Immediat que 137.28 est **reconnu** comme n par la grammaire mais pas 137.28.57.204.

Graphe orienté ⇔ grammaire



$$\begin{array}{rcl} \mathbb{A}_{sc} & = & \{a, b, c\} \\ a & ::= & b \\ b & ::= & c \\ c & ::= & c \mid a \end{array}$$

Ensemble de mots de longueur 6 commençant par a: abcccc, abccca, abccab, abcabc

Matrice d'adjacence des mots

En général graphe orienté ayant

- $\mathbb{A}^*$ (dénombrable) comme ensemble de vertex
- ensemble d'arêtes défini par *matrice d'adjacence de mots* $W : \mathbb{A}^* \times \mathbb{A}^* \to \{0, 1\}$ définie par

$$W(\alpha, \beta) = \begin{cases} 1 & \text{si } (\alpha, \beta) \in \Pi \\ 0 & \text{sinon} \end{cases}$$

- Mots accessibles de α

$$D_\alpha = \{\beta \in \mathbb{A}^* : W(\alpha, \beta) = 1\}, \alpha \in \mathbb{A}^*$$

Types d'évolution en informatique théorique

- *Déterministe:* $\forall \alpha \in \mathbb{A}^* : \sharp D_\alpha = 1$. Alors $\mathbb{A}^*$ se sépare en plusieurs parties disjointes de trajectoires calculatoires.

Types d'évolution en informatique théorique

- *Déterministe:* $\forall \alpha \in \mathbb{A}^* : \sharp D_\alpha = 1$. Alors $\mathbb{A}^*$ se sépare en plusieurs parties disjointes de trajectoires calculatoires.
- *Non-déterministe:* $\forall \alpha \in \mathbb{A}^* : 1 < \sharp D_\alpha < \infty$. Plusieurs trajectoires atteignent le même résultat.

Types d'évolution en informatique théorique

- *Déterministe:* $\forall \alpha \in \mathbb{A}^* : \sharp D_\alpha = 1$. Alors $\mathbb{A}^*$ se sépare en plusieurs parties disjointes de trajectoires calculatoires.
- *Non-déterministe:* $\forall \alpha \in \mathbb{A}^* : 1 < \sharp D_\alpha < \infty$. Plusieurs trajectoires atteignent le même résultat.
- *Stochastique:* non-déterministe et pour tout α vecteur probabilité $\mathbf{p}_\alpha = (p_{\alpha,\beta}, \beta \in D_\alpha)$ avec $p_{\alpha,\beta} \geq 0$ et $\sum_{\beta \in D_\alpha} p_{\alpha,\beta} = 1$. Chaque trajectoire acquiert poids de probabilité.

Types d'évolution en informatique théorique

- *Déterministe:* $\forall \alpha \in \mathbb{A}^* : \sharp D_\alpha = 1$. Alors $\mathbb{A}^*$ se sépare en plusieurs parties disjointes de trajectoires calculatoires.
- *Non-déterministe:* $\forall \alpha \in \mathbb{A}^* : 1 < \sharp D_\alpha < \infty$. Plusieurs trajectoires atteignent le même résultat.
- *Stochastique:* non-déterministe et pour tout α vecteur probabilité $\mathbf{p}_\alpha = (p_{\alpha,\beta}, \beta \in D_\alpha)$ avec $p_{\alpha,\beta} \geq 0$ et $\sum_{\beta \in D_\alpha} p_{\alpha,\beta} = 1$. Chaque trajectoire acquiert poids de probabilité.
- *Quantique:* non-déterministe et pour α vecteur unitaire $\mathbf{U}_\alpha = (U_{\alpha,\beta}, \beta \in D_\alpha)$ with $U_{\alpha,\beta} \in \mathbb{C}$ et $\sum_{\beta \in D_\alpha} |U_{\alpha,\beta}|^2 = 1$. Chaque trajectoire acquiert amplitude complexe de probabilité.

Hiérarchie des grammaires de Chomsky

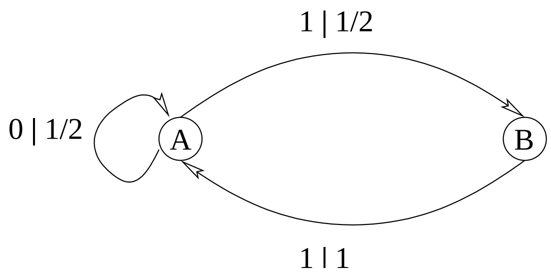
Type	Grammaire	Productions de la forme	Réconnaissance
0	recursivement énumerable	$\beta \to \beta'$	eTM
1	contextuelle	$\alpha b \gamma \to \alpha\beta\gamma$, $b \in \mathbb{A}_{sc}, \beta \neq \kappa$	eLBA
2	non-contextuelle	$b \to \beta$, $b \in \mathbb{A}_{sc}$, $\beta \in \mathbb{A}^*$	ePDA
3	régulière	$b ::= c\beta \vert \beta$, $b, c \in \mathbb{A}_{sc}$, $\beta \in \mathbb{A}_t^*$	eDFA

Préfixe e pour type d'évolution:

$$e \in \{D, N, S, Q\}.$$

Machine (automate) $=$ système dynamique

L'exemple de chaînes de Markov cachées



$$
\begin{array}{ccccc}
 & 0A & 1A & 0B & 1B \\
0A & 1/2 & 0 & 0 & 1/2 \\
1A & 1/2 & 0 & 0 & 1/2 \\
0B & 0 & 0 & 1 & 0 \\
1B & 0 & 1 & 0 & 0
\end{array}
$$

Phrases du type: $001101100011110110111 0\ldots$

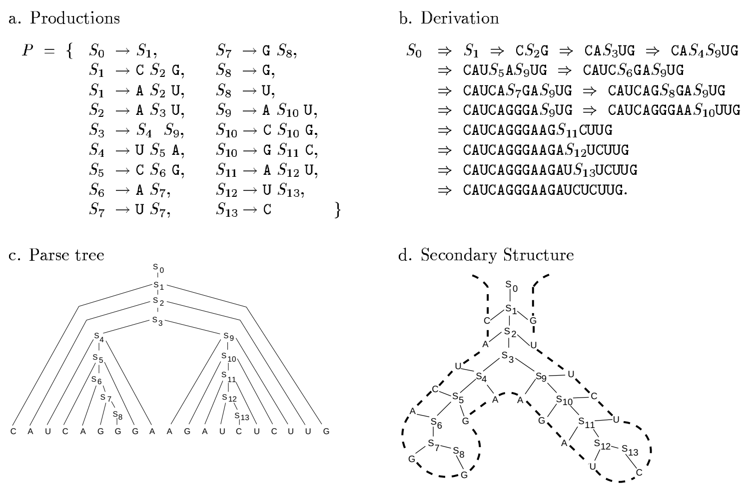
Théorème: CMC et grammaires régulières stochastiques sont équivalentes.

Ex. grammaire régulière stochastique

```
                1         11        21        31        41        51                  71        81        91        101       111       121       131
   Subdomain    ..........<-I-----*-*-----*------------>..........<-II----------*-----------------------.<-III-------------*--------------->....<-IV----
   PROSITE      ................AAAAAAAA.A..........................................................................................................
   X-ray        ..........BBBB.BB.........BBBBB..BB..................B......BBBBBBB.BAAA..AA.A..........A..A.AAA.........A.A.AA......AAA..AA.A............
   X-ray        ..........1111.11.........22222..22..................3......3333333.3BBB..BB.B..........B..C.CCC.........C.C.CC......CCC..CC.C............
 1 CAPK-ALPHA   ..........FERI.KTLGTGSFGR.VMLVK..HK.....E..........TG.N......HYAMKIL.DKqk..VV.K..........LK.Q.IEH.........T.L.NE......KRI..LQ.AV....N...F...
 2 WEE1+        ..........FRNV.TLLGSGEFSE.VFQVE..DP.....V..........EK.Tl.....KYAVKKL.KVk...FS.G..........PK.E.RNR.........L.L.QE......VSI..QR.AL....Kg..H...
 3 TIK          ..........FEDI.EEIGLGGFGQ.VFKAK..HR.....I..........DG.K......RYAIKRV.KY....--.N..........TE.K.AE-.........-.-.HE......VQA..LA.EL....N...H...
 4 SPK1         df........SIID.EVVGQGAFAT.VKKAI..ER.....T..........TG.K......TFAVKII.SKrk..VI.G..........NM.DgVT-.........-.-.RE......LEV..LQ.KL....N...H...
 5 RSK1-N       ..........FELL.KVLGQGSFGK.VFLVR..KVtrp..D..........SG.H......LYAMKVL.KKa...TL.K..........VR.D.RVR.........T.K.ME......RDI..LA.DV....N...H...
 6 PYT          ..........YSIL.KQIGSGGSSK.VFQVL..NE.....K..........K-.Q......IYAIKYV.NLe...EA.D..........NQ.T.LDS.........Y.R.NE......IAY..LN.KL....Qqh.S...
 7 PKC-ALPHA    ..........FNFL.MVLGKGSFGK.VMLAD..RK.....G..........TE.E......LYAIKIL.KKdv..VI.Q..........DD.D.VEC.........T.M.VE......KRV..LA.LL....Dk..P...
 8 PDGFR-B      ..........LVLG.RTLGSGAFGQ.VVEAT..AH.....G..........LS.Hsqatm.KVAVKML.KS....TA.R..........SS.E.KQA.........L.M.SE......LKI..MS.HL....Gp..H...
 9 PBS2         ..........LEFL.DELGHGNYGN.VSKVL..HK.....P..........TN.V......IMATKEV.RL....EL.D..........EA.K.FRQ.........I.L.ME......LEV..LH.KC....N...S...
10 MIK1         ..........FQQV.KPIHESDFSF.VYHVS..SInp...P..........TE.T......VYVVKML.KKn...AA.K..........FT.G.KER.........H.L.QE......VSI..LQ.RL....Qa..C...
11 MCK1         ..........VKEY.RKIGRGAFGT.VVQAY..LT.....Q..........DK.Knwlg..PFAIKKV.PA....--.H..........TE.Y.KS-.........-.-.RE......LQI..LR.IA....D...H...
12 INS.R        ..........ITLL.RELGQGSFGM.VYEGN..AR.....D..........II.Kgeaet.RVAVKTV.NE....SA.S..........LR.E.RIE.........F.L.NE......ASV..MK.GF....T...C...
13 HSVK         ..........FTIH.GALTPGSEGC.VFDSS..HP.....D..........YP.Q......RVIVKAG.WY....-T.S..........TS.-.---.........-.-.HE......ARL..LR.RL....D...H...
14 ERK1         ..........YTQL.QYIGEGAYGM.VSSAY..DH.....V..........RK.T......RVAIKKI.SP....FE.H..........QT.Y.CQR.........T.L.RE......IQI..LL.GF....R...H...
15 EGFR         ..........FKKI.KVLGSGAFGT.VYKGL..WI.....P..........EG.Ekvki..PVAIKEL.RE....AT.S..........PK.A.NKE.........I.L.DE......AYV..MA.SV....D...N...
16 ECK          ..........VTRQ.KVIGAGEFGE.VYKGM..LKt....S..........SG.Kkev...PVAIKTL.KA....GY.T..........EK.Q.RVD.........F.L.GE......AGI..MG.QF....S...H...
17 DPYK1        ne........LEFG.QTIGKGFFGE.VKRGY..WR.....E..........T-.-......DVAIKII.YRdq..FK.T..........KS.S.LVM.........F.Q.NE......VGI..LS.KL....R...H...
18 CLK          ar........YEIV.DTLGEGAFGK.VVECI..DH.....K..........VGgR......RVAVKIV.KN....--.V..........DR.Y.CEA.........A.Q.SE......IQV..LE.HL....Ntt.D...
19 CDC2HS       ..........YTKI.EKIGEGTYGV.VYKGR..HK.....T..........TG.Q......VVAMKKI.RLe...SE.E..........EG.V.PST.........A.I.RE......ISL..LK.EL....R...H...
20 CAMII-ALPHA  ..........YQLF.EELGKGAFSV.VRRCV..KV.....L..........AG.Q......EYAAKII.NTk...KL.S..........AR.D.HQK.........L.E.RE......ARI..CR.LL....K...H...
21 C-SRC        ..........LRLE.VKLGQGCFGE.VWMGT..WN.....G..........T-.T......RVAIKTL.KP....--.G..........TM.S.PEA.........F.L.QE......AQV..MK.KL....R...H...
22 C-RAF        ..........VMLS.TRIGSGSFGT.VYKGK..WH.....G..........D-.-......-VAVKIL.KVv...DP.T..........PE.Q.FQA.........F.R.NE......VAV..LR.KT....R...H...
23 KLSK_HUMAN   mgcgcss244LKLV.ERLGAGQFGE.VWMGY..YN.....G..........H-.T......KVAVKSL.KQ....--.G..........SM.S.PDA.........F.L.AE......ANL..MK.QL....Q...H...
24 KLSK_MOUSE   mgcvcss244LKLV.ERLGAGQFGE.VWMGY..YN.....G..........H-.T......KVAVKSL.KQ....--.G..........SM.S.PVP.........F.L.AE......ANL..MK.QL....Q...H...
25 ARKB_HUMAN   madleav190FSVH.RIIGRGGFGE.VYGCR..KR.....D..........TG.K......MYAMKCL.DKk...RI.K..........MK.Q.GET.........LaL.NE......RIM..LS.LVstg.D...C...
26 ARKB_BOVIN   madleav190FSVH.RIIGRGGFGE.VYGCR..KA.....D..........TG.K......MYAMKCL.DKk...RI.K..........MK.Q.GET.........LaL.NE......RIM..LS.LVstg.D...C...
27 BYR1_SCHPO   mfkrrrnp65LEVV.RHLGEGNGGA.VSLVK..HR.....-..........-N.I......FMARKTV.YV....GS.D..........SK.L.QKQ.........I.L.RE......LGV..LH.HC....R...S...
28 CYGR_ARBPU   mattrl1579QQIF.ATIG------.----T..YR.....G..........T-.-......ICAIHAVhKN....--.H..........ID.L.TRA.........V.R.TE......LKL..MR.DM....R...H...
29 ANPA_RAT     mpgsrrv536GSLL.TT--EGQF-Q.VFAKT..AY.....Y..........KG.N......LVAVKRVnRK....--.R..........IE.L.TRK.........V.L.FE......LKH..MR.DV....Q...N...
30 ANPA_HUMAN   mpgprrp471LEVL.ALVGSLS---.-LLGI..LI.....Vsffiyrkm70KG.N......LVAVKRVnRK....--.R..........IE.L.TRK.........V.L.FE......LKH..MR.DV....Q...N...
31 ANPB_HUMAN   malpsll512SRLT.LSLRGSSYGS.LMTAH..GKyqifaN..........TG.Hfkgn..VVAIKHV.NK....-K.R..........IE.L.TRQ.........V.L.FE......LKH..MR.DV....Q...F...
32 ANPA_MOUSE   mprsrrv536GSLL.TT--EGQF-Q.VFAKT..AY.....Y..........KG.N......LVAVKRVnRK....--.R..........IE.L.TRK.........V.L.FE......LKH..MR.DV....Q...N...
33 ANPB_RAT     malpsll512SRLT.LSLRGSSYGS.LMTAH..GKyqifaN..........TG.Hfkgn..VVAIKHV.NK....-K.R..........IE.L.TRQ.........V.L.FE......LKH..MR.DV....Q...
```

Ex.grammire non-contextuelles stochastique

Y. Sakakibara. M. Brown, R. Hughey, I. S. Mian, K. Sjölander, R. C. Underwood, and D. Haussler, Stochastic Context-Free Grammars for tRNA modeling", Nucleic Acids Research, 22:5112-5120 (1994)



Grammaires contextuelles stochastiques

Compilation de plusiers articles sur des sujete différents:
Gravitation quantique: Malyshev (98)
Chaos multiplicatif: Kahane-Peyrière (78), Collet-Koukiou (88), Liu (98), Menshikov-P (01)

Rwre sur des arbres: Comets-Menshikov-Popov (98), Menshikov-P (01),

Menshikov-P-Popov (02)

- Productions $\beta a \gamma := \delta \to \delta' := \beta \alpha \gamma$
- Chaîne de Markov à temps continu

$$\mathbb{P}(X_t = \delta' | X_0 = \delta) = (\exp(-tH))_{\delta\delta'}, \delta, \delta' \in \mathbb{A}^*$$

- Processus induit sur $\mathbb{A}^{\mathbb{Z}}$ par limite thermodynamique
- **Théorème:** Sous conditions sur productions, la limite thermodynamique pour le processus sur $\mathbb{A}^{\mathbb{Z}}$ existe.

Grammaires stochastiques (précisions)

$$\Omega = \{\omega : [0,\infty[\to \mathbb{A}^{\mathbb{Z}} \text{ t.q. } \omega \text{ admissible } \}$$

Admissible signifie:

- continu à droite
- compatible avec productions, *i.e.* si $\omega(s^-) = \alpha\beta\gamma$ and $\omega(s) = \alpha\beta'\gamma$ alors $(\beta,\beta') \in \Pi$

Pour $\alpha \in \mathbb{A}^{\mathbb{Z}}$ noter $\alpha_{[-N,N]}$ sa restriction dans $\{-N,-N+1,\ldots,N\}$. Soit $-N \leq i < 0 \leq j \leq N$ et

$$\begin{aligned}
\Omega^{(N)}(\alpha) &= \{\omega \in \Omega : \omega(0)_{[-N,N]} = \alpha_{[-N,N]}\} \\
\Omega^{(N)}_{ij;t}(\alpha) &= \{\omega \in \Omega^{(N)}(\alpha) : \ \alpha_i \text{ et } \alpha_j \text{ ne changent pas pour } s \in [0,t]\}
\end{aligned}$$

Grammaires contextuelles stochastiques

$$\lim_{N\to\infty} \mathbb{P}(\Omega_{ij;t}^{(N)}(\alpha)) = P(t;i,j), \qquad \sum_{(i,j):i<0\leq j} P(t;i,j) = 1$$

s_4	...	A	G	G	C	·	T	T	C	A	T	A	C	G	...
s_3	...	A	G	G	C	·	T	T	C	T	T	A	C	G	...
s_2	...	A	G	G	C	·	T	T	C	A	T	A	C	G	...
s_1	...	A	G	G	C	T	T	T	C	T	T	A	C	G	...
0	...	A	G	G	C	T	A		C	T	T	A	C	G	...
		$-N$			i				j					N	

Grammaire contextuelles quantiques

- Espace de Hilbert $\mathbb{H} = \ell^2(\mathbb{A}^*)$, avec base vérifiant $\langle e_\alpha \,|\, e_\beta \rangle = \delta_{\alpha,\beta}$
- Productions

$$\pi := (\beta a \gamma := \zeta \to \zeta' := \beta \alpha \gamma), |\beta| = j - 1$$

engendrées par des opérateurs

$$A_\pi(j) e_\zeta = e_{\zeta'}; \quad A_\pi^*(j) e_{\zeta'} = e_\zeta$$

- Hamiltonien formel

$$H = \sum_{\pi \in \Pi} \sum_{j \in \mathbb{N}} (\lambda_\pi A_\pi(j) + \overline{\lambda}_\pi A_\pi^*(j)).$$

Grammaires contectuelles quantiques

- Volume fini: $\mathbb{H}_N = \mathsf{span}\{e_\alpha : |\alpha| \leq N\}$.
- Operateurs en volume fini: $A_{\pi,N}(j) = P_N A_\pi(j) P_N$
- $\mathfrak{A}_N$ est la C^*-algèbre agissant sur $\mathbb{H}$ engendrée par $(P_N A_\pi(j) P_N, \pi \in \Pi; j \in \mathbb{N})$
- $\mathfrak{A}_N \simeq \mathcal{M}_{\frac{r^{N+1}-1}{r-1}}$, où $r = |\mathbb{A}|$
- $\phi_N : \mathfrak{A}_N \to \mathfrak{A}_{N+1}$ défini par $\phi_N : B \mapsto \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}$
- *Algèbre locale:* $\mathfrak{A}_\infty = \varinjlim(\mathfrak{A}_N, \phi_N)$.
- *Algèbre quasi-locale:* $\mathfrak{A} = \overline{\mathfrak{A}_\infty}$

Grammaires contextuelles quantiques

- Groupe d'automorphismes sur $\mathfrak{A}$:

$$\tau_t^{(N)}(B) = \exp(itH_N)B\exp(-itH_N).$$

- **Théorème:** $\exists t_0 > 0$: $B \in \mathfrak{A}$ et $\forall t$ avec $|t| < t_0$ la norme $\lim_{N\to\infty} \tau_t^{(N)}(B)$ existe et définit unique groupe $(\tau_t)_t$ d'automorphismes sur $\mathfrak{A}$

États KMS

- **Définition:** Soit $(\mathfrak{A}, \tau)$ un C^*-système dynamique. L'état ϕ sur $\mathfrak{A}$ est un **état τ-KMS à température inverse** β si

$$\phi(A\tau_{i\beta}(B)) = \phi(BA).$$

- Naïvement états KMS:

$$\langle\, B \,\rangle_\beta = \lim_N \langle\, B \,\rangle_{\beta,N} = \lim_N \frac{\mathsf{tr}_N(B \exp(-\beta H_N))}{Z_N}.$$

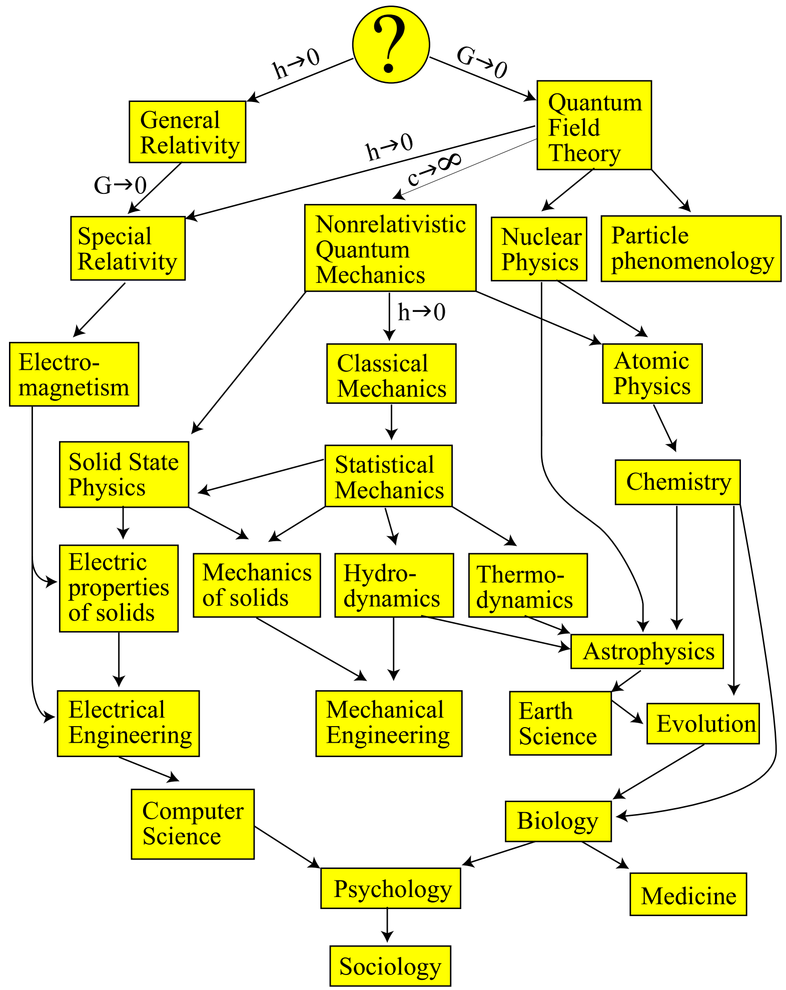
Mais

- **Lemme:** Pour β petit
  1. $\langle\, B \,\rangle_\beta = 0$
  2. La pression existe

États KMS

- Introduire subsititutions triviales $a \to a, a \in \mathbb{A}$ et noter $A_a(j)$ les opérateurs de production correspondant
- Soit $P_{=N}$ la projection orthogonale sur $\mathbb{H}_N \ominus \mathbb{H}_{N-1}$
- Pour $\mu > 0$, $\mu H^0 = \mu \sum_{N=0}^{\infty} N P_{=N} = \mu \sum_{a \in \mathbb{A}} \sum_j A_a(j)$
- Utiliser l'Hamiltonien formel $H' = H + \mu H^0$ pour définir états KMS
- Il est naturel de s'attendre à une transition de phase à μ_{cr}

Interdépendence des disciplines scientifiques



Max Tegmark, John Archibald Wheeler: 100 years of the Quantum, Scientific American

"I think of my lifetime in physics as divided into three periods. In the first period [. . .] I was in the grip of the idea that Everything is Particles. [. . .] I call my second period Everything is Fields. [. . .] Now I am in the grip of a new vision, that Everything is Information."

– John Archibald Wheeler

Geons, Black Holes, and Quantum Foam