# Cours Doctoral EDITE

# Champs de Markov - Convergence des échantillonneurs

Marc Sigelle

Année 2008-2009

Dernière mise a jour : 11 mars 2009

## Mesures de probabilité sur un ensemble $\Omega$ fini

- **espace $\mathcal{E}$ des fonctions réelles bornées sur $\Omega$:**

$$x \in \Omega \mapsto f(x) \in \mathbb{R}$$

- **les fonctions indicatrices:**

$$\forall y \in \Omega \ , \quad e_y(x) = 1\!\!1_{x=y}$$

$$\rightarrow \quad \{e_y\}_{y \in \Omega} \quad \text{est la base canonique de } \mathcal{E}$$

- **la fonction constante 1:**

$$x \mapsto 1 \qquad \forall x \in \Omega$$

## Mesures de probabilité sur $\Omega$ fini (suite)

- **forme linéaire sur $\mathcal{E}$ :**

$$f \in \mathcal{E} \mapsto P(f) \in \mathbb{R}$$

  * **positive** : $\forall x \in \Omega, \quad f(x) \geq 0 \quad \rightarrow \quad \mu(f) \geq 0$
  * **de masse totale** $1$ : $\mu(\mathbf{1}) = 1$

- **on écrira dans la suite (abus)**

$$\mu(x) = \mu(e_x)$$

- **masses de Dirac**

$$\delta_x(f) = f(x) \qquad \forall x \in \Omega, \forall f \in \mathcal{E}$$

## Distance entre mesures

- **(une) norme sur $\mathcal{E}$**

$$|| \, f \, ||_\infty = \sup_{x \in \Omega} | \, f(x) \, |$$

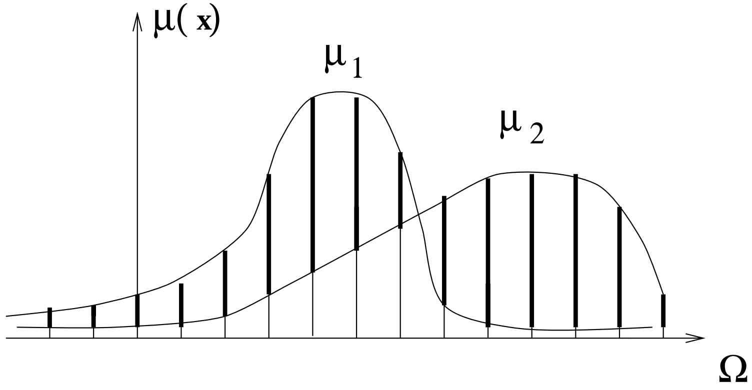
- **distance en variation entre deux mesures**

$$\begin{aligned} || \, \mu_1 - \mu_2 \, ||_1 &= \sum_{x \in \Omega} |\mu_1(x) - \mu_2(x)| \\ &= \sup_{|| \, f \, ||_\infty = 1} | \, \mu_1(f) - \mu_2(f) \, | \end{aligned}$$

- **on en tire**

$$| \, \mu_1(f) - \mu_2(f) \, | \leq || \, \mu_1 - \mu_2 \, ||_1 \; \times \; || \, f \, ||_\infty$$

# Distance entre mesures (suite)




○ **propriétés**

$$||\,\mu_1 - \mu_2\,||_1 \in [0.,\ 2.]$$

$$||\,\mu_1 - \mu_2\,||_1 = 0 \Leftrightarrow \mu_1 = \mu_2$$

$$||\,\mu_1 - \mu_2\,||_1 = 2 \Leftrightarrow \mu_1 \text{ et } \mu_2 \text{ sont à supports disjoints}$$

$$||\,\mu_1 - \mu_2\,||_1 = 2 - 2\sum_{x\in\Omega} \min(\mu_1(x), \mu_2(x))$$

$$\forall\ a, b \geq 0 \Rightarrow |a - b| = a + b - 2\ \min(a,\ b)$$

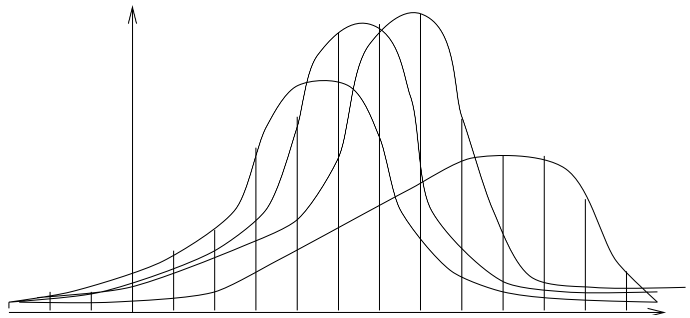
## Noyau $Q$ sur $\Omega$ (matrice stochastique)

- **linear application on $\mathcal{E}$:**

$$f \mapsto g = Qf \quad \text{with} \quad Q\,\mathbf{1} = \mathbf{1}$$

*i.e.* $Q(x,.)$ est une mesure $\forall x \in \Omega$.

$=$ famille de mesures sur $\Omega$ indicée par un élément de $\Omega$ lui-même !



- **interprétation en terme de probabilité de transition**

$$Q(x,\ y) = \pi(Y = y \ / \ X = x) = \pi(x \to y)$$

où $\pi(.)$ est une mesure sur $\Omega$

# Noyau $Q$ sur $\Omega$ (suite)

- **produit d'une mesure $\mu$ par un noyau $\rightarrow$ mesure définie par**

$$(\mu\ Q)(f) = \mu(Qf) \qquad \forall f \in \mathcal{E}$$

- **en particulier**

$$(\delta_x\ Q)(f) = \delta_x(Q\ f) = (Q\ f)(x) \Rightarrow (\delta_x\ Q) = Q(x, .\ )$$

- **on en tire**

$$(\mu\ Q)(x) = \sum_{y \in \Omega} \mu(y)\ Q(y,\ x) \qquad \forall x \in \Omega$$

- **produit de deux noyaux $\rightarrow$ noyau défini par**

$$(Q\ R)(x, y) = \sum_{z \in \Omega} Q(x,\ z)\ R(z,\ y) \qquad \forall x, y \in \Omega$$

- **conséquences**

$$(Q\ R)(x, .) = \delta_x(Q\ R) = (\delta_x\ Q)\ R = Q(x, .)\ R \qquad \forall x \in \Omega$$

# Coefficient de contraction de Dobrushin d'un noyau

◦ **définition**

$$c(Q) = \frac{1}{2} \max_{x,y \in \Omega} ||Q(x,.) - Q(y,.)||$$

◦ **comportement "physique"**

mesures "voisines" : $c(Q) \approx 0$ — deux mesures ou plus à supports disjoints : $c(Q) = 1$ — cas "normal" : $0 < c(Q) < 1$

◦ **→ "dispersion" des lois de probabilités d'un noyau**

échantillonnage des distributions de Gibbs

→ lié à la dispersion des différentes lois de transitions associées

## Coefficient de Dobrushin d'un noyau (suite)

◦ **propriétés fondamentales**

**a**) $c(Q) \in [0.,\ 1.]$

**b**) $c(Q) = 1 - \min_{x,y \in \Omega} \left( \sum_{z \in \Omega} \min(Q(x,z),\ Q(y,z)) \right) \leq 1 - |\Omega| \min_{a,b \in \Omega} Q(a,b)$

**c**) $||\, \mu_1\ Q - \mu_2\ Q \,||_1 \leq c(Q)\ ||\, \mu_1 - \mu_2 \,||_1$

**d**) $c(Q\ R) \leq c(Q)\ c(R)$

◦ **tout résulte de cela pour les MRFs, en particulier c) et d)**

## Coefficient de Dobrushin d'un noyau (suite)

◦ **exemple pour c) : deux masses de Dirac**

$$\mu_1 = \delta_a \quad \text{et} \quad \mu_2 = \delta_b$$

$$\begin{aligned}
\mu_1\, Q(x) &= Q(a,\ x) \\
\|\, \mu_1\, Q - \mu_2\, Q\,\|_1 &= \sum_{x\in\Omega} |Q(a,x) - Q(b,x)| = \|\, Q(a,.) - Q(b,.)\,\|_1 \\
&\leq \max_{x,y\in\Omega} \|\, Q(x,.) - Q(y,.)\,\|_1 = 2\ c(Q) \\
&= \|\, \mu_1 - \mu_2\,\|_1\ c(Q)
\end{aligned}$$

$\rightarrow$ égalité lorsque $(a,b) = \arg\min_{(x,y)} \|\, Q(x,.) - Q(y,.)\,\|_1 \quad (\Omega \text{ fini} \rightarrow \exists (a,b))$

## Coefficient de Dobrushin d'un noyau (suite)

◦ **propriété c) ⇒ d)**

$$\forall x,\ y \in \Omega \quad (QR)(x,.) - (QR)(y,.) = Q(x,.)\ R - Q(y,.)\ R$$

$$\rightarrow \forall x,\ y \in \Omega \quad ||\,(QR)(x,.) - (QR)(y,.)\,||_1 \leq c(R)\ ||\,Q(x,.) - Q(y,.)\,||_1$$

$$\Rightarrow \frac{1}{2} \max_{x,y\in\Omega} ||\,(QR)(x,.) - (QR)(y,.)\,||_1 \leq \frac{1}{2} \max_{x,y\in\Omega} ||\,Q(x,.) - Q(y,.)\,||_1\ c(R) \qquad \square$$

◦ **demonstration de c)**

$$(\mu_1\ Q)(f) - (\mu_2\ Q)(f) = \mu_1\ (Qf) - \mu_2\ (Qf)$$

$$Qf = \frac{\max Qf - \min Qf}{2}\ \phi + \frac{\max Qf + \min Qf}{2} \qquad (-1 \leq \phi \leq +1)$$

$$\Rightarrow |\ \mu_1\ (Qf) - \mu_2\ (Qf)\ | = \frac{Qf(M) - Qf(m)}{2}\ |\ \mu_1(\phi) - \mu_2(\phi)\ |$$

$$= \quad [\ \frac{Q_M(f) - Q_m(f)}{2}\ ]\ |\ \mu_1(\phi) - \mu_2(\phi)\ | \qquad (||\,\phi\,||_\infty = 1)$$

$$\leq \quad \frac{||\,Q(M,.) - Q(m,.)\,||_1}{2}\ ||\,f\,||_\infty\ ||\,\mu_1 - \mu_2\,||_1$$

$$\leq \quad c(Q)\ ||\,f\,||_\infty ||\,\mu_1 - \mu_2\,||_1 \qquad \forall f$$

## Mesure(s) invariante(s) associée(s) un noyau

◦ **définition**

$$(\mu Q)(x) = \mu(x) \ \ \forall x \in \Omega$$

$\longrightarrow$ "vecteurs propres" de la matrice stochastique $Q$ pour la valeur propre 1.

◦ **réversibilité**

$$\mu(x)\ Q(x,\ y) = \mu(y)\ Q(y,\ x) \ \ \forall x, y \in \Omega$$

$\longrightarrow$ $\mu$ mesure invariante pour $Q$ (sommation sur $x$ par exemple)

◦ **interprétation en terme de probabilités de transition** (Bayes)

$$\forall \pi \ \ , \pi(X = x)\ \pi(Y = y\ /\ X = x) = \pi(Y = y)\ \pi(X = x\ /\ Y = y) \quad \forall x, y \in \Omega$$

$i.e.\ \ \pi(X = x)\ \pi(x \to y) = \pi(Y = y)\ \pi(y \to x).$

$\Rightarrow$ $\pi(.)$ est réversible, donc invariante, *p.r.* à sa propre probabilité de transition $\pi(. \to .)$.

## Mesure(s) invariante(s) associée(s) un noyau (suite)

◦ **stricte positivité**

$$Q(x,y) > 0 \quad \forall x, y \in \Omega, \Leftrightarrow c(Q) < 1$$

◦ **irréductibilité**

$$\exists r \in \mathbb{N}^{+}\ t.q.\ Q^{r} > 0$$

◦ **théorème de Perron-Frobenius:** si $Q$ est stochastique $> 0$, $\exists$ une **unique measure invariante** $\mu$ de $Q$, avec une **convergence géométrique** de la chaîne de Markov associée $\forall$ la mesure initiale $P_0$

$$\forall M \geq 0, \quad ||\,(P_0\ Q^{M}) - \mu\,||_1 = ||\,(P_0\ Q^{M}) - (\mu\ Q^{M})\,||_1 \leq c(Q)^{M}\ ||\,P_0 - \mu\,||_1 \ ,$$

$$c(Q) < 1$$

## Mesure(s) invariante(s) associée(s) un noyau (suite)

- **la convergence tient aussi $Q$ si irréductible: $Q^r > 0$**

$$\mu\ Q = \mu \Rightarrow \mu\ Q^r = \mu$$

$$||\, P_0\ Q^{nr+p} - \mu\, ||_1 = ||\, P_0\ (Q^r)^n\ Q^p - \mu\ (Q^r)^n\, ||_1 \leq c(Q^r)^n\ ||\, P_0\ Q^p - \mu\, ||_1\ ,$$

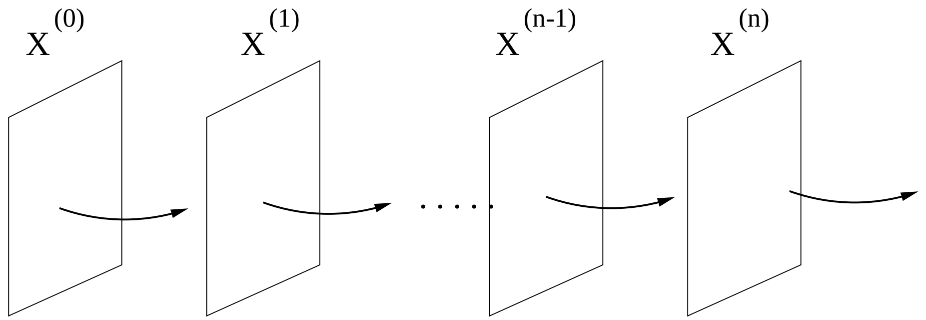
$$c(Q^r) < 1$$

- **loi des grands nombres (échantillons $X^{(n)}$ non indépendants!)**

$$\frac{\sum_{n=M}^{M+N} P_n(f)}{N} \underset{N\to+\infty}{\longrightarrow} \mu(f) \quad \Rightarrow \quad \frac{\sum_{n=M}^{M+N} f(X^{(n)})}{N} \underset{N\to+\infty}{\longrightarrow} \mathbf{E}_\mu[f(X)]$$

# Echantillonnage d'une distribution de Gibbs

- **chaîne de Markov de variables d'images !**




- **échantillonnage homogène : trouver $Q_T(x,\ y)$ tel que**

$$\lim_{n \to +\infty} \mathrm{P}(X^{(n)} = x) = \mathrm{P}_T(X = x) \quad \text{distribution de Gibbs à température } T$$

- **échantillonnage inhomogène : trouver $Q_n(x,\ y)$ tel que**

$$\lim_{n \to +\infty} \mathrm{P}(X^{(n)} = x) = \frac{1}{|\Omega^*|}\ \mathbb{1}_{x \in \Omega^*} \quad \text{distribution de Gibbs à température ``0''}$$

$\longrightarrow$ minimum globaux de $U(x)$

- **reformulation : trouver $Q_T(x, y)$ irréductible, réversible $p.r.$ $\mathbf{P}_T(.)$**

$\longrightarrow$ admettant donc $\mathrm{P}_T(.)$ comme mesure invariante $\forall T$

# Échantillonneur de Gibbs

◦ **descripteur au site $s$** $\xi \to \eta : \Delta U_{\text{ global}} = \Delta U(. \,/\, V_s)_{\text{ local}}$

$$\to U(x_s = \xi,\ x^s) + U(x_s = \eta \,/\, V_s) = U(x_s = \eta,\ x^s) + U(x_s = \xi \,/\, V_s)$$

$$\frac{\exp\ - U(x_s = \xi,\ x^s) \,/\, T}{Z_T} \cdot \frac{\exp\ - U(x_s = \eta \,/\, V_s) \,/\, T}{Z_T^s}$$

$$= \quad \frac{\exp\ - U(x_s = \eta,\ x^s) \,/\, T}{Z_T} \cdot \frac{\exp\ - U(x_s = \xi \,/\, V_s) \,/\, T}{Z_T^s} \qquad \forall T > 0$$

◦ **noyau de transition local**

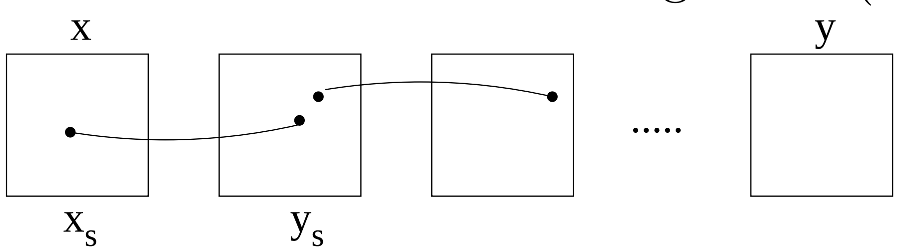
$$Q_T^s(x, y) = \mathbb{1}_{x_r = y_r, r \neq s} \,.\, \mathrm{P}_T(y_s \,/\, V_s) = \mathbb{1}_{x_r = y_r, r \neq s} \,.\, \frac{\exp\ - U(y_s \,/\, V_s) \,/\, T}{Z_T^s}$$

**réversible** (mais non strictement positif) *p.r.* la distribution de Gibbs

$$\mathrm{P}_T(x) = \frac{\exp\ - U(x) \,/\, T}{Z_T} \quad ,$$

# Échantillonneur de Gibbs (suite)

∘ **visite de l'ensemble des sites** : $\mathcal{T}$ = arrangement (= permutation) de $S$




∘ **noyau de transition global**

$$Q_T(x,\ y) = P(Y = y/X = x) = \prod_{s\in\mathcal{T}} \mathrm{P}_T(y_s\ /\ V_s) = \prod_{s\in\mathcal{T}} Q^s_T(x,\ y)$$

**strictement** $> 0$ et **réversible** (produit de noyaux réversibles).

∘ **convergence et loi des grands nombres :**

$$P(X^{(n)} = x) = P_0\ (Q_T)^n(x) \rightarrow \mathrm{P}_T(X = x) \quad \forall P_0$$

$$\frac{\sum_{n=M}^{M+N} f(X^{(n)})}{N} \underset{N\rightarrow+\infty}{\longrightarrow} \mathbf{E}_T[f(X)] \quad \text{ex: } \frac{\sum_{n=M}^{M+N} \mathbb{1}_{X_s^{(n)}=i}}{N} \underset{N\rightarrow+\infty}{\longrightarrow} \mathbf{E}_T[\mathbb{1}_{X_s=i}] = P_T(X_s = i)\ !$$

# Échantillonneur de Gibbs : vitesse de convergence

- **majoration du coefficient de Dobrushin**

$$\mathrm{P}_T(y_s \;/\; V_s) = \frac{1}{\displaystyle\sum_{\xi \in E} \exp\,[\; U(y_s \;/\; V_s) - U(\xi \;/\; V_s)\;] \;/\; T} \qquad \forall s \in S$$

$$\geq \frac{1}{|E|} \exp\, - \delta_s \;/\; T$$

$$\delta_s = \max_{V_s}\, (\max_{\eta \in E} U(\eta \;/\; V_s) - \min_{\xi \in E} U(\xi \;/\; V_s)) = \textbf{oscillation} \text{ de } U(. \;/\; .)$$

$$\Rightarrow Q_T(x,y) \geq \left( \frac{1}{|E|} \exp\, - \delta_s \;/\; T \right) \qquad \forall x, y \in \Omega$$

- **d'où il résulte :**

$$c(Q_\theta) \quad \leq \quad 1 - |\Omega| \, \frac{1}{|E|^{|S|}} \; \exp\, (- \sum_{s \in S} \delta_s \;/\; T \;)$$

$$= 1 - \exp\, (- \sum_{s \in S} \delta_s \;/\; T \;)$$

## Échantillonneur de Metropolis

◦ **niveau global**

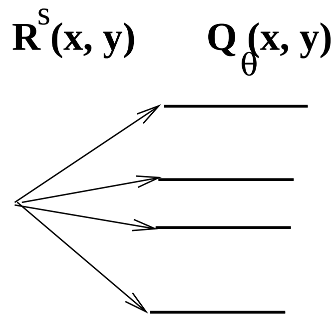
$$
\begin{array}{rrcll}
(\alpha) & Q_T(x,y) = & 1 & \text{si } U(y) < U(x) & (P_T(y) > P_T(x)) \\
 & Q_T(y,x) = & 1 & \text{si } U(y) > U(x) & + \\
\multicolumn{2}{r}{\mathrm{P}_T(x)\ Q_T(x,y) =} & \mathrm{P}_T(y)\ Q_T(y,x) & (\text{ réversibilité }) & \Rightarrow \\
(\beta) & Q_T(x,y) = & \dfrac{\mathrm{P}_T(y)}{\mathrm{P}_T(x)} & & \\
 & = & \exp -(U(y) - U(x)\ /\ T & \text{si } U(y) > U(x) &
\end{array}
$$

# Échantillonneur de Metropolis

◦ **niveau local : loi d'acceptance**

$$R^s(x,y) = R(x_s, y_s) \ \text{ symétrique } \ \text{e.g. } R(x_s, y_s) = \frac{1}{|E|} \ (\text{ loi uniforme sur } E \ )$$

$$S^s_T(x,y) = \mathbb{1}_{x_r = y_r, r \neq s} \ . \ R^s(x,y) \ Q_T(x,y)$$




réversible car $R$ symétrique

◦ **tour $\mathcal{T}$ comme précédemment**

$$S_T(x,y) = \prod_{s \in \mathcal{T}} S^s_T(x,y) = \prod_{s \in \mathcal{T}} [\ R^s(x,y) \ Q_T(x,y) \ ]$$

**réversible** *p.r.* $\mathrm{P}_\theta(x)$ et **strictement** $> 0 \Rightarrow$ **convergence :**

$$P(X^{(n)} = x) = P_0 \ (Q_T)^n(x) \to \mathrm{P}_T(X = x) \qquad \forall P_0$$

◦ **+ loi des grands nombres**

# Échantillonneurs : vitesse de convergence

◦ **Metropolis : majoration du coefficient de Dobrushin**

loi d'acceptation équidistibuée sur $E$ ,

$$S_T(x,y) \;\geq\; \frac{1}{|\Omega|} \exp - (\,\Delta \,/\, T\,) \qquad \text{avec } \Delta = \max_{y\in\Omega} U(y) - \min_{x\in\Omega} U(x)$$

$$\Rightarrow c(S_T) \;\leq\; 1 - \exp(-\,\Delta \,/\, T\,)$$

◦ **pour les deux échantillonneurs, on obtient une majoration du type**

$$c(Q_T) \leq 1 - \exp(-\,\Delta \,/\, T\,)$$

où $\Delta$ dépend de l'échantillonneur étudié.

## Echantillonnage inhomogène (recuit) : convergence

◦ **hypothèses et notations**

- une suite $T_n \rightarrow T = 0$
- une loi de probabilité initiale sur $\Omega$, notée $\mu_0$
- une suite de mesures invariantes notées $P_n = \mathrm{P}_{T_n}(.)$
- la génération d'échantillons avec des noyaux de transition notés $Q_n = Q_{T_n}(. \, , .)$

A chaque étape $n$

$$P(X^{(n)} = x) = \pi_n(x) = \mu_0 \; Q_1 \; Q_2 \ldots Q_n(x)$$

# Recuit simulé : convergence

◦ **utilisation du lemme d'Abel [Winkler(1995)]**

$P_n = P_n \; Q_n$ pour $\geq 1 \Rightarrow$ :

$$
\begin{aligned}
P_n - \pi_n = P_n - \mu_0 \; Q_1 \; Q_2 \ldots Q_n &= P_n \; Q_n - \mu_0 \; Q_1 \; Q_2 \ldots \; Q_{n-1} \; Q_n \\
&= (P_n - P_{n-1}) \; Q_n + (P_{n-1} - \pi_{n-1}) \; Q_n \\
& \quad \text{et par récurrence} \\
&= (P_n - P_{n-1}) \; Q_{n-1} + (P_{n-1} - P_{n-2}) \; Q_{n-1} \; Q_n \\
&+ (P_{n-2} - P_{n-3}) \; Q_{n-2} \; Q_{n-1} \ldots Q_n \quad \ldots \\
&+ (P_{n-p+1} - P_{n-p}) \; Q_{n-p+1} \ldots Q_n + (P_{n-p} - \pi_{n-p}) \; Q_{n-p+1} \ldots Q_n
\end{aligned}
$$

◦ **condition suffisante de convergence : on distingue**

1. les $p$ premiers termes : $\displaystyle\sum_k ||P_k - P_{k-1}|| < +\infty$

   (Cauchy pour la suite $\{P_k\}$ : $||P_m - P_q|| \leq \sum_{k=q}^{m} ||P_k - P_{k-1}||$)

2. le dernier terme : $\forall p \geq 1, \;\; c(Q_{n-p} Q_{n-p+1} \ldots Q_n) \rightarrow 0$

# Recuit simulé (suite)

- **1ère condition :** $\sum_k ||P_k - P_{k-1}|| < +\infty$ d'où $P_k \to P_0 = \frac{1}{|\Omega^*|} \mathbb{1}_{x \in \Omega*}$

$$\forall x \in \Omega \ , P_k(x) \nearrow \text{ ou } \searrow \text{ quand } T \to 0^+$$

- **2ème condition : suite de conditions suffisantes !**

$$\forall p \geq 1 \ , \ c(Q_{n-p} Q_{n-p+1} \dots Q_n) \to 0 \qquad (c(Q_n) \to 1!)$$

$$c(Q_{n-p} Q_{n-p+1} \dots Q_n) \leq \prod_{k=n-p}^{n} (1 - \exp(-\frac{\Delta}{T_k}))$$

$$\lim_{n \to +\infty} \prod_{k=1}^{n} (1 - \exp(-\frac{\Delta}{T_k})) = 0 \text{ produit infini ``diverge'' vers } 0 \text{ !!}$$

$$\sum_{n=1}^{+\infty} \exp -\frac{\Delta}{T_n} = +\infty \Leftarrow \boxed{T_n \geq \frac{\Delta}{\log n + 1}}$$

[Geman and Geman(1984)]

# References

[Geman and Geman(1984)] S. Geman and D. Geman. Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[Winkler(1995)] Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods. A Mathematical Introduction.* Applications of mathematics. Springer-Verlag.