# Stochastic chains with variable length memory

Antonio Galves

Instituto de Matemática e Estatística
Universidade de São Paulo

Séminaire Parisien de Statistiques, IHP, le 12 février 2007

# Chains with variable length memory

- Rissanen (1983) introduced the notion of *stochastic chains with variable length memory* as a universal system for data compression
- He called this model a *finitely generated source* or a *tree machine*.
- Recently this model has become popular in the Statistics literature under the name of *variable length Markov chain* coined by Bühlman and Wyner (1999).
- in Bio-informatics it was also called a *probabilistic suffix tree* (Bejerano and Yona 2001).

# Chains with variable length memory

- Rissanen (1983) introduced the notion of *stochastic chains with variable length memory* as a universal system for data compression
- He called this model a *finitely generated source* or a *tree machine*.
- Recently this model has become popular in the Statistics literature under the name of *variable length Markov chain* coined by Bühlman and Wyner (1999).
- in Bio-informatics it was also called a *probabilistic suffix tree* (Bejerano and Yona 2001).

# Chains with variable length memory

- Rissanen (1983) introduced the notion of *stochastic chains with variable length memory* as a universal system for data compression
- He called this model a *finitely generated source* or a *tree machine*.
- Recently this model has become popular in the Statistics literature under the name of *variable length Markov chain* coined by Bühlman and Wyner (1999).
- in Bio-informatics it was also called a *probabilistic suffix tree* (Bejerano and Yona 2001).

# Chains with variable length memory

- Rissanen (1983) introduced the notion of *stochastic chains with variable length memory* as a universal system for data compression
- He called this model a *finitely generated source* or a *tree machine*.
- Recently this model has become popular in the Statistics literature under the name of *variable length Markov chain* coined by Bühlman and Wyner (1999).
- in Bio-informatics it was also called a *probabilistic suffix tree* (Bejerano and Yona 2001).

# Variable Length Markov Chains

A VLMC is a stationary stochastic chain $(X_n)$ taking values on a finite alphabet $\mathcal{A}$ and characterized by two elements:

- The set of all contexts.

  A context $X_{n-\ell}, \ldots, X_{n-1}$ is the finite portion of the past $X_{-\infty}, \ldots, X_{n-1}$ which is relevant to predict the next symbol $X_n$.

- A family of transition probabilities associated to each context.

  Given a context, its associated transition probability gives the distribution of occurrence of the next symbol immediately after the context.

# Variable Length Markov Chains

A VLMC is a stationary stochastic chain $(X_n)$ taking values on a finite alphabet $\mathcal{A}$ and characterized by two elements:

- The set of all contexts.

  A context $X_{n-\ell}, \ldots, X_{n-1}$ is the finite portion of the past $X_{-\infty}, \ldots, X_{n-1}$ which is relevant to predict the next symbol $X_n$.

- A family of transition probabilities associated to each context.

  Given a context, its associated transition probability gives the distribution of occurrence of the next symbol immediately after the context.

# Context trees

Notation: $w_m^n = (w_m, \ldots, w_n)$.
A subset $\tau$ of $\cup_{k=1}^{\infty} \mathcal{A}^{\{-k,\ldots,-1\}}$ is a **complete tree with finite branches** if

- **Suffix property.** For no $w_{-k}^{-1} \in \tau$, there exists $u_{-j}^{-1} \in \tau$ with $j < k$ such that $w_{-i} = u_{-i}$ for $i = 1, \ldots, j$.
- **Completeness.** $\tau$ defines a **partition** of $A^{\{\ldots,-2,-1\}}$. Each element of the partition coincides with the set of the sequences in $A^{\{\ldots,-2,-1\}}$ having $w_{-k}^{-1}$ as suffix, for some $w_{-k}^{-1} \in \tau$.

# Context trees

Notation: $w_m^n = (w_m, \ldots, w_n)$.
A subset $\tau$ of $\cup_{k=1}^{\infty} \mathcal{A}^{\{-k,\ldots,-1\}}$ is a **complete tree with finite branches** if

- **Suffix property.** For no $w_{-k}^{-1} \in \tau$, there exists $u_{-j}^{-1} \in \tau$ with $j < k$ such that $w_{-i} = u_{-i}$ for $i = 1, \ldots, j$.
- **Completeness.** $\tau$ defines a **partition** of $A^{\{\ldots,-2,-1\}}$. Each element of the partition coincides with the set of the sequences in $A^{\{\ldots,-2,-1\}}$ having $w_{-k}^{-1}$ as suffix, for some $w_{-k}^{-1} \in \tau$.

A **probabilistic context tree** on $\mathcal{A}$ is an ordered pair $(\tau, p)$ with

- $\tau$ is a complete tree with finite branches; and
- $p = \{p(\cdot|w); w \in \tau\}$ is a family of probability measures on $\mathcal{A}$.

# Probabilistic context trees

A **probabilistic context tree** on $\mathcal{A}$ is an ordered pair $(\tau, p)$ with

- $\tau$ is a complete tree with finite branches; and
- $p = \{p(\cdot|w); w \in \tau\}$ is a family of probability measures on $\mathcal{A}$.

A stationary stochastic chain $(X_n)$ is *consistent* with a probabilistic context tree $(\tau, p)$ if for any infinite past $x_{-\infty}^{-1}$ and any symbol $a \in \mathcal{A}$ we have

$$\mathbb{P}\left\{X_0 = a \mid X_{-\infty}^{-1} = x_{-\infty}^{-1}\right\} = p(a \mid x_{-\ell}^{-1}),$$

where $x_{-\ell}^{-1}$ is the only element of $\tau$ which is a suffix of the sequence $x_{-\infty}^{-1}$.

- The suffix $x_{-\ell}^{-1}$ is called the *context* of the sequence $x_{-\infty}^{-1}$
- The length of the context $\ell = \ell(x_{-\infty}^{-1})$ is a function of the sequence.
- The suffix property implies that the set $\{\ell(X_{-\infty}^{-1}) = k\}$ is measurable with respect to the $\sigma$-algebra generated by $X_{-k}^{-1}$.

# The suffix property

- The suffix $x_{-\ell}^{-1}$ is called the *context* of the sequence $x_{-\infty}^{-1}$
- The length of the context $\ell = \ell(x_{-\infty}^{-1})$ is a function of the sequence.
- The suffix property implies that the set $\{\ell(X_{-\infty}^{-1}) = k\}$ is measurable with respect to the $\sigma$-algebra generated by $X_{-k}^{-1}$.

# The suffix property

- The suffix $x_{-\ell}^{-1}$ is called the *context* of the sequence $x_{-\infty}^{-1}$
- The length of the context $\ell = \ell(x_{-\infty}^{-1})$ is a function of the sequence.
- The suffix property implies that the set $\{\ell(X_{-\infty}^{-1}) = k\}$ is measurable with respect to the $\sigma$-algebra generated by $X_{-k}^{-1}$.

# Bounded and unbounded trees

- The length of a context $w = w_k^{-1}$ denoted by $|w| = k$.
- The *height* of the tree $\tau$ is defined as $|\tau| = \max\{|w|; w \in \tau\}$.
- If $\ell$ is bounded, $\tau$ is finite and the process $(X_n)$ is a Markov chain;
- If $\ell$ is unbounded,the process $(X_n)$ is a chain of *infinite order*.
- Rissanen, Bühlman and Wyner only considered the bounded case.

# Bounded and unbounded trees

- The length of a context $w = w_k^{-1}$ denoted by $|w| = k$.
- The *height* of the tree $\tau$ is defined as $|\tau| = \max\{|w|; w \in \tau\}$.
- If $\ell$ is bounded, $\tau$ is finite and the process $(X_n)$ is a Markov chain;
- If $\ell$ is unbounded,the process $(X_n)$ is a chain of *infinite order*.
- Rissanen, Bühlman and Wyner only considered the bounded case.

# Bounded and unbounded trees

- The length of a context $w = w_k^{-1}$ denoted by $|w| = k$.
- The *height* of the tree $\tau$ is defined as $|\tau| = \max\{|w|; w \in \tau\}$.
- If $\ell$ is bounded, $\tau$ is finite and the process $(X_n)$ is a Markov chain;
- If $\ell$ is unbounded,the process $(X_n)$ is a chain of *infinite order*.
- Rissanen, Bühlman and Wyner only considered the bounded case.

# Bounded and unbounded trees

- The length of a context $w = w_k^{-1}$ denoted by $|w| = k$.
- The *height* of the tree $\tau$ is defined as $|\tau| = \max\{|w|; w \in \tau\}$.
- If $\ell$ is bounded, $\tau$ is finite and the process $(X_n)$ is a Markov chain;
- If $\ell$ is unbounded,the process $(X_n)$ is a chain of *infinite order*.
- Rissanen, Bühlman and Wyner only considered the bounded case.

# Bounded and unbounded trees

- The length of a context $w = w_k^{-1}$ denoted by $|w| = k$.
- The *height* of the tree $\tau$ is defined as $|\tau| = \max\{|w|; w \in \tau\}$.
- If $\ell$ is bounded, $\tau$ is finite and the process $(X_n)$ is a Markov chain;
- If $\ell$ is unbounded,the process $(X_n)$ is a chain of *infinite order*.
- Rissanen, Bühlman and Wyner only considered the bounded case.

# Why variable length memory chains are interesting?

- They constitute an interesting class of chains of infinite order;
- They are able to model candidates to model rhythmic contours in natural languages, or families of proteins!
- This is due to the fact that the tree of contexts $\tau$ describes structural dependencies present in the data.
- More precisely, Rissanen's algorithm *Context* captures structural dependencies of the data.

# Why variable length memory chains are interesting?

- They constitute an interesting class of chains of infinite order;
- They are able to model candidates to model rhythmic contours in natural languages, or families of proteins!
- This is due to the fact that the tree of contexts $\tau$ describes structural dependencies present in the data.
- More precisely, Rissanen's algorithm *Context* captures structural dependencies of the data.

# Why variable length memory chains are interesting?

- They constitute an interesting class of chains of infinite order;
- They are able to model candidates to model rhythmic contours in natural languages, or families of proteins!
- This is due to the fact that the tree of contexts $\tau$ describes structural dependencies present in the data.
- More precisely, Rissanen's algorithm *Context* captures structural dependencies of the data.

# Why variable length memory chains are interesting?

- They constitute an interesting class of chains of infinite order;
- They are able to model candidates to model rhythmic contours in natural languages, or families of proteins!
- This is due to the fact that the tree of contexts $\tau$ describes structural dependencies present in the data.
- More precisely, Rissanen's algorithm *Context* captures structural dependencies of the data.

# The algorithm Context

- Rissanen (1983) not only introduced the class of variable length memory chains;
- he also introduced a consistent algorithm to estimate the probabilistic tree which produced the data.

# The algorithm Context

- Rissanen (1983) not only introduced the class of variable length memory chains;
- he also introduced a consistent algorithm to estimate the probabilistic tree which produced the data.

# The algorithm Context

- Starting with a finite sample $X_1, \ldots, X_n$
- First construct a candidate context $X_{n-k(n)}^{n-1}$ where $k(n) = C \log n$
- Then decide to shorten or not this candidate context using some *gain function*. For instance the log-likelihood ratio statistics.
- The intuitive reason behind the choice of the upper bound length $C \log n$ is the impossibility of estimating the probability of sequences of length longer than $\log n$ based on a sample of length $n$.

# The algorithm Context

- Starting with a finite sample $X_1, \ldots, X_n$
- First construct a candidate context $X_{n-k(n)}^{n-1}$ where $k(n) = C \log n$
- Then decide to shorten or not this candidate context using some *gain function*. For instance the log-likelihood ratio statistics.
- The intuitive reason behind the choice of the upper bound length $C \log n$ is the impossibility of estimating the probability of sequences of length longer than $\log n$ based on a sample of length $n$.

# The algorithm Context

- Starting with a finite sample $X_1, \ldots, X_n$
- First construct a candidate context $X_{n-k(n)}^{n-1}$ where $k(n) = C \log n$
- Then decide to shorten or not this candidate context using some *gain function*. For instance the log-likelihood ratio statistics.
- The intuitive reason behind the choice of the upper bound length $C \log n$ is the impossibility of estimating the probability of sequences of length longer than $\log n$ based on a sample of length $n$.

# Estimation of the probability transitions

- For any finite string $w_{-j}^{-1}$ with $j \leq n$, denote $N_n(w_{-j}^{-1})$ the number of occurrences of the string in the sample
- $$N_n(w_{-j}^{-1}) = \sum_{t=0}^{n-j} \mathbf{1}\left\{X_t^{t+j-1} = w_{-j}^{-1}\right\}.$$
- If $\sum_{b \in \mathcal{A}} N_n(w_{-k}^{-1}b) > 0$, we define the estimator of the transition probability $p$ by

$$\hat{p}_n(a|w_{-k}^{-1}) = \frac{N_n(w_{-k}^{-1}a)}{\sum_{b \in \mathcal{A}} N_n(w_{-k}^{-1}b)}.$$

# Estimation of the probability transitions

- For any finite string $w_{-j}^{-1}$ with $j \leq n$, denote $N_n(w_{-j}^{-1})$ the number of occurrences of the string in the sample
- $$N_n(w_{-j}^{-1}) = \sum_{t=0}^{n-j} \mathbf{1}\left\{X_t^{t+j-1} = w_{-j}^{-1}\right\} .$$
- If $\sum_{b \in \mathcal{A}} N_n(w_{-k}^{-1}b) > 0$, we define the estimator of the transition probability $p$ by

$$\hat{p}_n(a|w_{-k}^{-1}) = \frac{N_n(w_{-k}^{-1}a)}{\sum_{b \in \mathcal{A}} N_n(w_{-k}^{-1}b)} .$$

# Estimation of the probability transitions

- For any finite string $w_{-j}^{-1}$ with $j \leq n$, denote $N_n(w_{-j}^{-1})$ the number of occurrences of the string in the sample
- $$N_n(w_{-j}^{-1}) = \sum_{t=0}^{n-j} \mathbf{1}\left\{X_t^{t+j-1} = w_{-j}^{-1}\right\} .$$
- If $\sum_{b \in \mathcal{A}} N_n(w_{-k}^{-1}b) > 0$, we define the estimator of the transition probability $p$ by

$$\hat{p}_n(a|w_{-k}^{-1}) = \frac{N_n(w_{-k}^{-1}a)}{\sum_{b \in \mathcal{A}} N_n(w_{-k}^{-1}b)} .$$

# Log-likelihood ratio statistic

- We also define

$$\Lambda_n(i,w) = -2 \sum_{w_{-i}\in\mathcal{A}} \sum_{a\in\mathcal{A}} N_n(w_{-i}^{-1}a) \log\left[\frac{\hat{p}_n(a|w_{-i}^{-1})}{\hat{p}_n(a|w_{-i+1}^{-1})}\right].$$

- $\Lambda_n(i,w)$ is the log-likelihood ratio statistic for testing the consistency of the sample with a probabilistic suffix tree $(\tau, p)$ against the alternative that it is consistent with $(\tau', p')$ where $\tau$ and $\tau'$ differ only by one set of sibling nodes branching from $w_{-i+1}^{-1}$.

# Log-likelihood ratio statistic

- We also define

$$\Lambda_n(i,w) \;=\; -2 \sum_{w_{-i}\in\mathcal{A}} \sum_{a\in\mathcal{A}} N_n(w_{-i}^{-1}a)\log\left[\frac{\hat{p}_n(a|w_{-i}^{-1})}{\hat{p}_n(a|w_{-i+1}^{-1})}\right] .$$

- $\Lambda_n(i,w)$ is the log-likelihood ratio statistic for testing the consistency of the sample with a probabilistic suffix tree $(\tau, p)$ against the alternative that it is consistent with $(\tau', p')$ where $\tau$ and $\tau'$ differ only by one set of sibling nodes branching from $w_{-i+1}^{-1}$.

# Length of the estimated current context

$$\hat{\ell}(X_0^{n-1}) = \max\left\{ i = 2, \ldots, k(n) : \Lambda_n(i, X_{n-k(n)}^{n-1}) \; > \; C_2 \log n \right\} \; ,$$

where $C_2$ is any positive constant.

# Rissanen's theorem

**Theorem.** Given a realization $X_0, \ldots, X_{n-1}$ of a probabilistic suffix tree $(\tau, p)$ with **finite height**, then

$$\mathbb{P}\left\{\hat{\ell}(X_0^{n-1}) \neq \ell(X_0^{n-1})\right\} \longrightarrow 0$$

as $n \to \infty$.

# Unbounded context trees

In the unbounded case, the compactness of $\mathcal{A}^{\mathbb{Z}}$ assures that there is at least one stationary stochastic chain consistent with a continuous probabilistic suffix tree. Uniqueness requires further conditions.

# Type A probabilistic context trees

A **type A** probabilistic context tree $(\tau, p)$ on $\mathcal{A}$ satisfies the conditions:

- **Weakly non-nullness**, that is

$$\sum_{a \in \mathcal{A}} \inf_{w \in \tau} p(a \mid w) > 0\,;$$

- **Continuity**

$$\beta(k) := \max_{a \in \mathcal{A}} \sup\{|p(a \mid w) - p(a \mid v)|, v \in \tau, w \in \tau \text{ with } w_{-k}^{-1} = v$$

as $k \to \infty$.

- $\{\beta(k)\}_k \in \mathbb{N}$ is called the **continuity rate**.

# Type A probabilistic context trees

A **type A** probabilistic context tree $(\tau, p)$ on $\mathcal{A}$ satisfies the conditions:

- **Weakly non-nullness**, that is

$$\sum_{a\in\mathcal{A}} \inf_{w\in\tau} p(a \mid w) > 0\,;$$

- **Continuity**

$$\beta(k) \;:=\; \max_{a\in\mathcal{A}} \sup\{|p(a \mid w) - p(a \mid v)|, v \in \tau, w \in \tau \text{ with } w_{-k}^{-1} = v_{-k}^{-1}\}$$

as $k \to \infty$.

- $\{\beta(k)\}_k \in \mathbb{N}$ is called the **continuity rate.**

# Type A probabilistic context trees

A **type A** probabilistic context tree $(\tau, p)$ on $\mathcal{A}$ satisfies the conditions:

- **Weakly non-nullness**, that is

$$\sum_{a \in \mathcal{A}} \inf_{w \in \tau} p(a \mid w) > 0\,;$$

- **Continuity**

$$\beta(k) \ := \ \max_{a \in \mathcal{A}} \sup\{|p(a \mid w) - p(a \mid v)|, v \in \tau, w \in \tau \ \text{with} \ w_{-k}^{-1} = v_{-}$$

as $k \to \infty$.

- $\{\beta(k)\}_k \in \mathbb{N}$ is called the **continuity rate.**

# A theorem for unbounded trees.

**Theorem.** (Duarte, Galves and Garcia) Let $X_0, X_2, \ldots, X_{n-1}$ be a sample from a type A unbounded probabilistic suffix tree $(\tau, p)$ with continuity rate $\beta(j) \leq f(j) \exp\{-j\}$, with $f(j) \to 0$ as $j \to \infty$. Then, for any choice of positive constants $C_1$ and $C_2$ there exist positive constants $C$ and $D$ such that

$$\mathbb{P}\left\{\hat{\ell}(X_0^{n-1}) \neq \ell(X_0^{n-1})\right\} \leq C_1 \log n(n^{-C_2} + D/n) + Cf(C_1 \log n) \, .$$

# Ingredients of the proof

- The proof has two ingredients:
- the first ingredient is the convergence of the log-likelihood ratio statistics of a *finite order* Markov chain.
- The problem is that an unbounded probabilistic context tree defines a chain of infinite order, not a Markov chain!
- That's why we need a second ingredient which is the canonical Markov approximation to chains of infinite order.

# Ingredients of the proof

- The proof has two ingredients:
- the first ingredient is the convergence of the log-likelihood ratio statistics of a *finite order* Markov chain.
- The problem is that an unbounded probabilistic context tree defines a chain of infinite order, not a Markov chain!
- That's why we need a second ingredient which is the canonical Markov approximation to chains of infinite order.

# Ingredients of the proof

- The proof has two ingredients:
- the first ingredient is the convergence of the log-likelihood ratio statistics of a *finite order* Markov chain.
- The problem is that an unbounded probabilistic context tree defines a chain of infinite order, not a Markov chain!
- That's why we need a second ingredient which is the canonical Markov approximation to chains of infinite order.

# Ingredients of the proof

- The proof has two ingredients:
- the first ingredient is the convergence of the log-likelihood ratio statistics of a *finite order* Markov chain.
- The problem is that an unbounded probabilistic context tree defines a chain of infinite order, not a Markov chain!
- That's why we need a second ingredient which is the canonical Markov approximation to chains of infinite order.

# The canonical Markov approximation

**Theorem.**(fernández and Galves 2002) Let $(X_t)_{t\in\mathbb{Z}}$ be a chain consistent with a type A probabilistic suffix tree $(\tau, p)$ with summable continuity rate, and let $(X_t^{[k]})$ be its canonical Markov approximation of order $k$. Then there exists a coupling between $(X_t)$ and $(X_t^{[k]})$ and a constant $C > 0$ such that

$$\mathbb{P}\left\{X_0 \neq X_0^{[k]}\right\} \leq C\beta(k)\,.$$

# The chi-square approximation

- At each step of the algoithm Context we perform at most $k(n)$ sequential tests, where $k(n) \to \infty$ as $n$ diverges.
- To control the error in the chi-square approximation we use a well-known asymptotic expansion for the distribution of $\Lambda_n(i, w)$ due to Hayakawa (1970) which implies that

$$\mathbb{P}\left\{\Lambda_n(i, w) \leq x \mid H_0^i\right\} = \mathbb{P}\left\{\chi^2 \leq x\right\} + D/n \,,$$

where $D$ is a positive constant and $\chi^2$ is random variable with distribution chi-square with $|\mathcal{A}| - 1$ degrees of freedom.

# The chi-square approximation

- At each step of the algoithm Context we perform at most $k(n)$ sequential tests, where $k(n) \to \infty$ as $n$ diverges.
- To control the error in the chi-square approximation we use a well-known asymptotic expansion for the distribution of $\Lambda_n(i, w)$ due to Hayakawa (1970) which implies that

$$\mathbb{P}\left\{\Lambda_n(i, w) \leq x \mid H_0^i\right\} = \mathbb{P}\left\{\chi^2 \leq x\right\} + D/n\,,$$

where $D$ is a positive constant and $\chi^2$ is random variable with distribution chi-square with $|\mathcal{A}| - 1$ degrees of freedom.

# Another version of the algorithm Context

- In a recent paper with Véronique Maume and Bernard Schmitt we propose to use as *gain* function
- $$\Delta_n(j) = \max_{a \in A} |\hat{p}_n(a|X_{n-j}^{n-1}) - \hat{p}_n(a|X_{n-(j-1)}^{n-1})|,$$

  where $1 \leq j \leq k(n)$;
  and define $\hat{\ell}(X_0^{n-1})$ as

  $$\max\{j = 1, \ldots, k(n) : \Delta_n(j) < \delta\},$$

  where $\delta > 0$ is any fixed threshold.
- If the contexts are bounded, then any $\delta > 0$ would do the job.

# Another version of the algorithm Context

- In a recent paper with Véronique Maume and Bernard Schmitt we propose to use as *gain* function
- $$\Delta_n(j) = \max_{a \in A} |\hat{p}_n(a|X_{n-j}^{n-1}) - \hat{p}_n(a|X_{n-(j-1)}^{n-1})|,$$

  where $1 \leq j \leq k(n)$;
  and define $\hat{\ell}(X_0^{n-1})$ as

  $$\max\{j = 1, \ldots, k(n) : \Delta_n(j) < \delta\},$$

  where $\delta > 0$ is any fixed threshold.
- If the contexts are bounded, then any $\delta > 0$ would do the job.

# Another version of the algorithm Context

- In a recent paper with Véronique Maume and Bernard Schmitt we propose to use as *gain* function
-
$$\Delta_n(j) = \max_{a \in A} |\hat{p}_n(a|X_{n-j}^{n-1}) - \hat{p}_n(a|X_{n-(j-1)}^{n-1})|,$$

  where $1 \leq j \leq k(n)$;
  and define $\hat{\ell}(X_0^{n-1})$ as

$$\max\{j = 1, \ldots, k(n) : \Delta_n(j) < \delta\},$$

  where $\delta > 0$ is any fixed threshold.
- If the contexts are bounded, then any $\delta > 0$ would do the job.

# Following Dedecker and Prieur (2005)

- For all $t > 0$,

$$\mathbb{P}(|N_n(a_0^j) - np(a_0^j)| > t) \leq e^{\frac{1}{e}} \exp\left(\frac{-t^2 \beta p_{\min}}{2enp(a_0^j)}\right),$$

- where

$$p_{\min} = \min_{w \in \tau} p(w),$$

- and $\beta$ is defined using Dobrushin's coefficient...

# Following Dedecker and Prieur (2005)

- For all $t > 0$,

$$\mathbb{P}(|N_n(a_0^j) - np(a_0^j)| > t) \leq e^{\frac{1}{e}} \exp\left(\frac{-t^2 \beta p_{\min}}{2enp(a_0^j)}\right),$$

- where

$$p_{\min} = \min_{w \in \tau} p(w),$$

- and $\beta$ is defined using Dobrushin's coefficient...

# Following Dedecker and Prieur (2005)

- For all $t > 0$,

$$\mathbb{P}(|N_n(a_0^j) - np(a_0^j)| > t) \leq e^{\frac{1}{e}} \exp\left(\frac{-t^2 \beta p_{\min}}{2enp(a_0^j)}\right),$$

- where

$$p_{\min} = \min_{w \in \tau} p(w)\,,$$

- and $\beta$ is defined using Dobrushin's coefficient...

- The paper with Bernard and Véronique can be downloaded from
  `www.ime.usp.br/~galves/artigos/arbres.pdf`
- The paper with Denise and Nancy can be downloaded from
  `www.ime.usp.br/~galves/artigos/uvlmc.pdf`

- The paper with Bernard and Véronique can be downloaded from `www.ime.usp.br/~galves/artigos/arbres.pdf`
- The paper with Denise and Nancy can be downloaded from `www.ime.usp.br/~galves/artigos/uvlmc.pdf`