Stochastic chains with memory of variable length

Antonio Galves

Universidade de São Paulo

AofA 2008

Chains with memory of variable length

- Introduced by Rissanen (1983) as a universal system for data compression.
- He called this model a *finitely generated source* or a *tree machine*.
- Statisticians call it *variable length Markov chain* (Bühlman and Wyner 1999).
- Also called *prediction suffix tree* in bio-informatics (Bejerano and Yona 2001).

Chains with memory of variable length

- Introduced by Rissanen (1983) as a universal system for data compression.
- He called this model a *finitely generated source* or a *tree machine*.
- Statisticians call it *variable length Markov chain* (Bühlman and Wyner 1999).
- Also called *prediction suffix tree* in bio-informatics (Bejerano and Yona 2001).

Chains with memory of variable length

- Introduced by Rissanen (1983) as a universal system for data compression.
- He called this model a *finitely generated source* or a *tree machine*.
- Statisticians call it *variable length Markov chain* (Bühlman and Wyner 1999).
- Also called *prediction suffix tree* in bio-informatics (Bejerano and Yona 2001).

Chains with memory of variable length

- Introduced by Rissanen (1983) as a universal system for data compression.
- He called this model a *finitely generated source* or a *tree machine*.
- Statisticians call it *variable length Markov chain* (Bühlman and Wyner 1999).
- Also called *prediction suffix tree* in bio-informatics (Bejerano and Yona 2001).

Heuristics

When we have a symbolic chain describing

Heuristics

When we have a symbolic chain describing

a syntatic structure,

Heuristics

When we have a symbolic chain describing

a syntatic structure,

a prosodic contour,

Heuristics

When we have a symbolic chain describing

a syntatic structure,

a prosodic contour,

a protein,....

Heuristics

When we have a symbolic chain describing

a syntatic structure,

a prosodic contour,

a protein,....

it is natural to assume that each symbol depends only on a **finite suffix** of the past

Heuristics

When we have a symbolic chain describing

a syntatic structure,

a prosodic contour,

a protein,....

it is natural to assume that each symbol depends only on a **finite suffix** of the past

whose **length depends on the past**.

Warning!

We are not making the usual **markovian assumption**:

Warning!

We are not making the usual **markovian assumption**:

at each step we are under the influence of a suffix of the past whose **length depends on the past itsel.**

Warning!

We are not making the usual **markovian assumption**:

at each step we are under the influence of a suffix of the past whose **length depends on the past itsel.**

Even if it is finite, in general the length of the relevant part of the past is not bounded above!

Warning!

We are not making the usual **markovian assumption**:

at each step we are under the influence of a suffix of the past whose **length depends on the past itsel**.

Even if it is finite, in general the length of the relevant part of the past is not bounded above!

This means that in general these are chains of infinite order, not Markov chains.

Contexts

- Call the relevant suffix of the past a **context.**
- The set of all contexts should have the **suffix property**:
- **Suffix property:** no context is a proper suffix of another context.
- This means that we can identify the end of each context without knowing what happened sooner.
- The suffix property implies that the set of all contexts can be represented as a **rooted tree with finite branches.**

Contexts

- Call the relevant suffix of the past a **context.**
- The set of all contexts should have the **suffix property**:
- **Suffix property:** no context is a proper suffix of another context.
- This means that we can identify the end of each context without knowing what happened sooner.
- The suffix property implies that the set of all contexts can be represented as a **rooted tree with finite branches**.

Contexts

- Call the relevant suffix of the past a **context.**
- The set of all contexts should have the **suffix property**:
- **Suffix property:** no context is a proper suffix of another context.
- This means that we can identify the end of each context without knowing what happened sooner.
- The suffix property implies that the set of all contexts can be represented as a **rooted tree with finite branches**.

Contexts

- Call the relevant suffix of the past a **context.**
- The set of all contexts should have the **suffix property**:
- **Suffix property:** no context is a proper suffix of another context.
- This means that we can identify the end of each context without knowing what happened sooner.
- The suffix property implies that the set of all contexts can be represented as a **rooted tree with finite branches**.

Contexts

- Call the relevant suffix of the past a **context.**
- The set of all contexts should have the **suffix property**:
- **Suffix property:** no context is a proper suffix of another context.
- This means that we can identify the end of each context without knowing what happened sooner.
- The suffix property implies that the set of all contexts can be represented as a **rooted tree with finite branches.**

Chains with variable length memory

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

- The tree of all contexts.
- A family of transition probabilities associated to each context.

Chains with variable length memory

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

- The tree of all contexts.
- A family of transition probabilities associated to each context.

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

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

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

Example: the renewal process on $\mathbb{Z}$

$$\mathcal{A} = \{0, 1\}$$

$$\tau = \{1, 10, 100, 1000, \ldots\}$$

$$p(1 \mid 0^k 1) = q_k$$

where $0 < q_k < 1$, for any $k \geq 0$, and

$$\sum_{k \geq 0} q_k = +\infty \, .$$

Contexts, partitions and stoping times

The set of all contexts should define a partition of the set of all possible infinite pasts

Contexts, partitions and stoping times

The set of all contexts should define a partition of the set of all possible infinite pasts

Given an infinite past $x_{-\infty}^{-1}$ its context $x_{-\ell}^{-1}$ is the only element of τ which is a suffix of the sequence $x_{-\infty}^{-1}$.

Contexts, partitions and stoping times

The set of all contexts should define a partition of the set of all possible infinite pasts

Given an infinite past $x_{-\infty}^{-1}$ its context $x_{-\ell}^{-1}$ is the only element of τ which is a suffix of the sequence $x_{-\infty}^{-1}$.

The length of the context $\ell = \ell(x_{-\infty}^{-1})$ is a function of the sequence.

Contexts, partitions and stoping times

The set of all contexts should define a partition of the set of all possible infinite pasts

Given an infinite past $x_{-\infty}^{-1}$ its context $x_{-\ell}^{-1}$ is the only element of τ which is a suffix of the sequence $x_{-\infty}^{-1}$.

The length of the context $\ell = \ell(x_{-\infty}^{-1})$ is a function of the sequence.

More precisely, the event

$$\{\ell(X_{-\infty}^{-1}) = k\}$$

is measurable with respect to the σ-algebra generated by X_{-k}^{-1}.

Probabilistic context trees

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

- τ 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 (τ, p) with

- τ 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 and chains

A stationary stochastic chain (X_n) is *compatible* with a probabilistic context tree (τ, p) if

for any infinite past $x_{-\infty}^{-1}$ and any symbol $a \in \mathcal{A}$ we have

Probabilistic context trees and chains

A stationary stochastic chain (X_n) is *compatible* with a probabilistic context tree (τ, 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 τ which is a suffix of the sequence $x_{-\infty}^{-1}$.

A first mathematical question

Given a probabilistic context tree (τ, p) does it exist at least (at most) one stationary chain (X_n) compatible with it?

A first mathematical question

Given a probabilistic context tree (τ, p) does it exist at least (at most) one stationary chain (X_n) compatible with it?

First answer: verify if the infinite order transition probabilities defined by (τ, p) satisfy the sufficient conditions which assure the existence and uniqueness of a chain of infinite order.

Type A probabilistic context trees

A **type A** probabilistic context tree (τ, 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** $\beta(k) \to 0$ as, $k \to \infty$, where

$$\beta(k) := \sup |p(a \mid w) - p(a \mid v)|,$$

  and the sup is taken wih respect to all $a \in \mathcal{A}, v \in \tau, w \in \tau$ with $w_{-k}^{-1} = v_{-k}^{-1}$.

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

Type A probabilistic context trees

A **type A** probabilistic context tree (τ, 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** $\beta(k) \to 0$ as, $k \to \infty$, where

$$\beta(k) := \sup |p(a \mid w) - p(a \mid v)|,$$

  and the sup is taken wih respect to all $a \in \mathcal{A}, v \in \tau, w \in \tau$ with $w_{-k}^{-1} = v_{-k}^{-1}$.

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

Type A probabilistic context trees

A **type A** probabilistic context tree (τ, 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** $\beta(k) \to 0$ as, $k \to \infty$, where

$$\beta(k) := \sup |p(a \mid w) - p(a \mid v)|,$$

  and the sup is taken wih respect to all $a \in \mathcal{A}, v \in \tau, w \in \tau$ with $w_{-k}^{-1} = v_{-k}^{-1}$.

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

A uniqueness result

For a probabilistic suffix tree of type A

.

A uniqueness result

For a probabilistic suffix tree of type A

with summable continuity rate,

.

A uniqueness result

For a probabilistic suffix tree of type A

with summable continuity rate,

the maximal coupling argument used in Fernández and Galves (2002)

.

A uniqueness result

For a probabilistic suffix tree of type A

with summable continuity rate,

the maximal coupling argument used in Fernández and Galves (2002)

implies the uniqueness of the law of the chain compatible with it.

Given a sample is it possible to estimate the smallest probabilistic context tree generating it ?

A basic statistical question

Given a sample is it possible to estimate the smallest probabilistic context tree generating it ?

In the case of finite context trees, Rissanen (1983) introduced the *algorithm Context* to estimate in a consistent way the probabilistic context tree out from a sample.

The algorithm Context

Starting with a finite sample $(X_0, \ldots, X_{n-1})$ the goal is to estimate the context at step n.

- Start with a candidate context $(X_{n-k(n)}, \ldots, X_{n-1})$, where $k(n) = C_1 \log n$.
- 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 $(X_0, \ldots, X_{n-1})$ the goal is to estimate the context at step n.

- Start with a candidate context $(X_{n-k(n)}, \ldots, X_{n-1})$, where $k(n) = C_1 \log n$.
- 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 $(X_0, \ldots, X_{n-1})$ the goal is to estimate the context at step n.

- Start with a candidate context $(X_{n-k(n)}, \ldots, X_{n-1})$, where $k(n) = C_1 \log n$.
- 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 $w_{-j}^{-1} = (w_{-j}, \ldots, w_{-1})$, 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 $w_{-j}^{-1} = (w_{-j}, \ldots, w_{-1})$, 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 $w_{-j}^{-1} = (w_{-j}, \ldots, w_{-1})$, 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 (τ, p) against the alternative that it is consistent with (τ', p') where τ and τ' 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 (τ, p) against the alternative that it is consistent with (τ', p') where τ and τ' 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. (Rissanen 1983) Given a realization $X_0, \ldots, X_{n-1}$ of a probabilistic suffix tree (τ, p) with **finite height**, then

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

as $n \to \infty$.

Extending the algorithm Context

Is it possible to extend the algorithm Context to the case of unbounded probabilistic context trees?

Is it possible to extend the algorithm Context to the case of unbounded probabilistic context trees?

How fast does the algorithm Context converge?

A theorem for unbounded trees.

Theorem. (Duarte, Galves and Garcia 2006)
Let $(X_0, X_2, \ldots, X_{n-1})$ be a sample from a type A unbounded probabilistic suffix tree (τ, p)

A theorem for unbounded trees.

Theorem. (Duarte, Galves and Garcia 2006)
Let $(X_0, X_2, \ldots, X_{n-1})$ be a sample from a type A unbounded probabilistic suffix tree (τ, p)

with continuity rate

$$\beta(j) \leq f(j) \exp\{-j\},$$

with $f(j) \to 0$ as $j \to \infty$.

A theorem for unbounded trees.

Theorem. (Duarte, Galves and Garcia 2006)
Let $(X_0, X_2, \ldots, X_{n-1})$ be a sample from a type A unbounded probabilistic suffix tree (τ, 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 the constants C_1 and C_2 defining the algorithm we have

A theorem for unbounded trees.

Theorem. (Duarte, Galves and Garcia 2006)
Let $(X_0, X_2, \ldots, X_{n-1})$ be a sample from a type A unbounded probabilistic suffix tree (τ, 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 the constants C_1 and C_2 defining the algorithm we have

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

where D is a positive constant.

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 compatible with a type A probabilistic suffix tree (τ, 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 canonical Markov approximation

Theorem.(Fernández and Galves 2002)

- Let $(X_t)_{t\in\mathbb{Z}}$ be a chain compatible with a type A probabilistic suffix tree (τ, 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 canonical Markov approximation

Theorem.(Fernández and Galves 2002)

- Let $(X_t)_{t\in\mathbb{Z}}$ be a chain compatible with a type A probabilistic suffix tree (τ, 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 canonical Markov approximation

Theorem.(Fernández and Galves 2002)

- Let $(X_t)_{t\in\mathbb{Z}}$ be a chain compatible with a type A probabilistic suffix tree (τ, 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 algorithm 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 χ^2 is random variable with distribution chi-square with $|\mathcal{A}| - 1$ degrees of freedom.

The chi-square approximation

- At each step of the algorithm 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 χ^2 is random variable with distribution chi-square with $|\mathcal{A}| - 1$ degrees of freedom.

The chi-square approximation

- At each step of the algorithm 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 χ^2 is random variable with distribution chi-square with $|\mathcal{A}| - 1$ degrees of freedom.

The chi-square approximation

- At each step of the algorithm 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 χ^2 is random variable with distribution chi-square with $|\mathcal{A}| - 1$ degrees of freedom.

- The paper with Duarte and Garcia can be downloaded from `www.ime.usp.br/~galves/artigos/uvlmc.pdf`
- My review paper with Eva Löcherbach can be downloaded from `www.ime.usp.br/~galves/artigos/rissanen.pdf`

- The paper with Duarte and Garcia can be downloaded from
  `www.ime.usp.br/~galves/artigos/uvlmc.pdf`
- My review paper with Eva Löcherbach can be downloaded from
  `www.ime.usp.br/~galves/artigos/rissanen.pdf`