Pairwise graphical models and Markov random fields

Definition

For undirected graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, finite alphabet $\mathcal{X}$, functions $\psi_i : \mathcal{X} \to \mathbb{R}_+$, $i \in \mathcal{V}$, $\psi_e : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$, $e \in \mathcal{E}$, the probability distribution μ on $\mathcal{X}^{\mathcal{V}}$ defined by $\mu(x) := \frac{1}{Z} \prod_{i \in \mathcal{V}} \psi_i(x_i) \prod_{e=(i,j) \in \mathcal{E}} \psi_{i,j}(x_i, x_j)$

is a **pairwise graphical model** with underlying graph $\mathcal{G}$

The normalization constant $Z = \sum_{x \in \mathcal{X}^{\mathcal{V}}} \prod_{i \in \mathcal{V}} \psi_i(x_i) \prod_{e=(i,j) \in \mathcal{E}} \psi_{i,j}(x_i, x_j)$ is known as its partition function.

Example

Ising model: $\mathcal{X} = \{-1, 1\}$, $\psi_i(x_i) = e^{h_i x_i}$, $\psi_{ij}(x_i, x_j) = e^{J_{ij} x_i x_j}$.
X_i: spin at site i; h_i: external field at i; J_{ij}: coupling coefficient between sites i and j

Definition

For undirected graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, finite alphabet $\mathcal{X}$, probability measure μ on $\mathcal{X}^{\mathcal{V}}$, $(\mathcal{G}, \mu)$: Markov field if there exist functions $\psi_K : \mathcal{X}^K \to \mathbb{R}_+$ indexed by **cliques** K of $\mathcal{G}$ and Z such that
$\mu(x) = \frac{1}{Z} \prod_K \psi_K(x_K), x \in \mathcal{X}^{\mathcal{V}}$,
where $x_K := (x_i)_{i \in K}$.

Easy property: For any pairwise graphical model μ with underlying graph $\mathcal{G}$, $(\mu, \mathcal{G})$ is a Markov random field.

Definition

For subsets $A, B, C \subset \mathcal{V}$, C separates A and B in $\mathcal{G}$ if any path in $\mathcal{G}$ from A to B traverses C. We denote this $A \underset{\mathcal{G}}{\overset{C}{--}} B$.

Hammersley-Clifford Theorem

Theorem

For Markov field $(\mu, \mathcal{G})$*, and* A, B, C *such that* $A \overset{C}{\underset{\mathcal{G}}{--}} B$*, then under* μ*,* X_A *and* X_B *are independent conditionally on* X_C*.*

Conversely, for any probability measure μ *on* $\mathcal{X}^{\mathcal{V}}$ *such that* $\forall x \in \mathcal{X}^{\mathcal{V}},\ \mu(x) > 0$ *and for all* A, B, C *such that* $A \overset{C}{\underset{\mathcal{G}}{--}} B$*, under* μ*,* X_A *and* X_B *are independent conditionally on* X_C*, then* $(\mu, \mathcal{G})$ *is a Markov field.*

Markov random field distributions: $\mu(x) = \frac{1}{Z}\psi(x), \quad x \in \mathcal{X}^{\mathcal{V}}$.
For $\psi > 0$, $\mu(x) = \frac{1}{Z}e^{-E(x)/T}$, Boltzmann distribution with energy $E(x)$ and temperature T

Definition

Gibbs **free energy** functional
$\mathbb{G} : M(\mathcal{X}^{\mathcal{V}}) =$ probability measures on $\mathcal{X}^{\mathcal{V}} \to \mathbb{R}$
$\mathbb{G}(\nu) := -H(\nu) - \mathbb{E}_\nu \ln \psi(x)$, where
$H(\nu) := \sum_x \nu(x) \ln(1/\nu(x))$: Shannon entropy of ν

Proposition

$\mathbb{G}$ strictly convex on $M(\mathcal{X}^{\mathcal{V}})$, minimal at $\nu = \mu$, the Boltzmann-Gibbs distribution, and $\mathbb{G}(\mu) = -\ln(Z)$.

Remark

Recall Kullback-Leibler divergence between distributions on discrete $\mathcal{X}$:
$D(p\|q) := \sum_{x\in\mathcal{X}} p(x) \ln\left(\frac{p(x)}{q(x)}\right)$. *Then* $\mathbb{G}(\nu) = -\ln(Z) + D(\nu\|\mu)$.

Tree Markov fields and belief propagation

$(\mathcal{G}, \mu)$: $\mathcal{G}$ tree graph, $\mu(x) = \frac{1}{Z} \prod_{i \in \mathcal{V}} \psi_i(x_i) \prod_{(ij) \in \mathcal{E}} \psi_{ij}(x_i, x_j), \quad x \in \mathcal{X}^{\mathcal{V}}$

For edge (ij) of $\mathcal{G}$ with orientation $i \to j$, $\mathcal{T}_{i\to j} = (\mathcal{V}_{i\to j}, \mathcal{E}_{i\to j})$: subtree of $\mathcal{G}$ = connected component of i after removal of edge (ij)

Let $\mu_{i\to j}(x_{\mathcal{V}_{i\to j}}) = \frac{1}{Z_{i\to j}} \prod_{k \in \mathcal{V}_{i\to j}} \psi_k(x_k) \prod_{(k\ell) \in \mathcal{E}_{i\to j}} \psi_{k\ell}(x_k, x_\ell)$:
Pairwise graphical model induced on $\mathcal{T}_{i\to j}$

Let:
$b_{i\to j}(x_i)$= marginal of X_i under $\mu_{i\to j}$,
$\mu(x_i)$= marginal of X_i under μ,
$\mu_{ij}(x_i, x_j)$= marginal of (X_i, X_j) under μ

Recursive computations on trees

For leaf node i with incident edge (ij) in $\mathcal{G}$,
$b_{i\to j}(x_i) = \frac{1}{Z_{i\to j}}\psi_i(x_i),\ Z_{i\to j} = \sum_{x_i\in\mathcal{X}}\psi_i(x_i)$.

Induction for non-leaf node i:

$$\begin{aligned}
b_{i\to j}(x_i) &= \frac{1}{Z_{i\to j}} \sum_{x_k, k\in\mathcal{V}_{i\to j}\setminus\{i\}} \prod_{k\in\mathcal{V}_{i\to j}} \psi_k(x_k) \prod_{(k\ell)\in\mathcal{E}_{i\to j}} \psi_{k\ell}(x_k, x_\ell) \\
&= \frac{1}{Z_{i\to j}}\psi_i(x_i) \prod_{k\sim i, k\neq j} \sum_{x_k} \psi_{ki}(x_k, x_i) Z_{k\to i} b_{k\to i}(x_k).
\end{aligned}$$

$$\begin{aligned}
Z_{i\to j} &= \sum_{x_k, k\in\mathcal{V}_{i\to j}} \prod_{k\in\mathcal{V}_{i\to j}} \psi_k(x_k) \prod_{(k\ell)\in\mathcal{E}_{i\to j}} \psi_{k\ell}(x_k, x_\ell) \\
&= \sum_{x_i\in\mathcal{X}} \psi_i(x_i) \prod_{k\sim i, k\neq j} \sum_{x_k} \psi_{ki}(x_k, x_i) Z_{k\to i} b_{k\to i}(x_k).
\end{aligned}$$

Formulas for marginals

$$\mu_i(x_i) = \frac{1}{Z}\psi_i(x_i)\prod_{k\sim i}\sum_{x_k}\psi_{ki}(x_k, x_i)Z_{k\to i}b_{k\to i}(x_k),$$

$$\mu_{ij}(x_i, x_j) = \frac{1}{Z}\psi_{ij}(x_i, x_j)Z_{i\to j}b_{i\to j}(x_i)Z_{j\to i}b_{j\to i}(x_j),$$

$$Z = \sum_{x_i\in\mathcal{X}}\psi_i(x_i)\prod_{k\sim i}\sum_{x_k}\psi_{ki}(x_k, x_i)Z_{k\to i}b_{k\to i}(x_k).$$

Computational cost for obtaining Z and all marginals, with recursive approach: $O(|\mathcal{X}|^2)$ cost per oriented edge, hence $O(|\mathcal{V}|\cdot|\mathcal{X}|^2)$ operations in total.
Compare to brute force computation of Z: sum of $|\mathcal{X}|^{|\mathcal{V}|}$ terms...

Belief Propagation (a.k.a. Product-Sum) Algorithm

Iteratively update *beliefs*, or *messages* $b_{i\to j}$ by letting

$$b_{i\to j}(x_i) \propto \psi_i(x_i) \prod_{k\sim i, k\neq j} \sum_{x_k} \psi_{ki}(x_k, x_i) b_{k\to i}(x_k)$$

- Can be run synchronously or asynchronously
- Converges after each vector $b_{i\to j}$ updated finite number of times
- Limit values allow to determine marginals and normalization constants

Tree Markov fields, further properties

$$\begin{aligned}\mu(x) &= \prod_{i\in\mathcal{V}} \mu_i(x_i) \prod_{(ij)\in\mathcal{E}} \frac{\mu_{ij}(x_i,x_j)}{\mu_i(x_i)\mu_j(x_j)} \\ &= \prod_{i\in\mathcal{V}} \mu_i(x_i)^{1-d_i} \prod_{(ij)\in\mathcal{E}} \mu_{ij}(x_i,x_j)\end{aligned}$$

Proof.

Induction on $|\mathcal{V}|$: for leaf node i, and incident edge (ij), write
$\mu(x) = \frac{\mu_{ij}(x_i,x_j)}{\mu_j(x_j)}\mathbb{P}_\mu(X_{\mathcal{V}-i} = x_{\mathcal{V}-i})$
Note that $\mathbb{P}_\mu(X_{\mathcal{V}-i} = \cdot)$: tree Markov field on $\mathcal{T}_{j\to i}$... $\square$

Belief propagation in general (non-tree) Markov fields

Recall Gibbs free energy $\mathbb{G}(\nu) = -H(\nu) - \mathbb{E}_\nu \ln \psi(x)$ minimized by Boltzmann-Gibbs distribution $\mu(x) \propto \psi(x)$.

For $\psi(x) = \prod_{i \in \mathcal{V}} \psi_i(x_i) \prod_{(ij) \in \mathcal{E}} \psi_{ij}(x_i, x_j)$, "energy" term:

$$\mathbb{E}_\nu \ln(\psi(x)) = \sum_{i \in \mathcal{V}} \sum_{x_i \in \mathcal{X}} \nu_i(x_i) \ln(\psi_i(x_i)) + \sum_{(ij) \in \mathcal{E}} \sum_{x_i, x_j \in \mathcal{X}} \nu_{ij}(x_i, x_j) \ln(\psi_{ij}(x_i, x_j)),$$

where ν_i, ν_{ij}: marginals of ν.

If $\mathcal{G}$ = tree, and $(\mathcal{G}, \nu)$ = tree Markov field, entropy term:

$$\begin{aligned} H(\nu) = & \sum_{(ij) \in \mathcal{E}} \sum_{x_i, x_j \in \mathcal{X}} \nu_{ij}(x_i, x_j) \ln\left(\frac{1}{\nu_{ij}(x_i, x_j)}\right) \\ & + \sum_{i \in \mathcal{V}} \sum_{x_i \in \mathcal{X}} (1 - d_i) \nu_i(x_i) \ln\left(\frac{1}{\nu_i(x_i)}\right) \end{aligned}$$

Bethe free energy

Using expression for tree-field entropy, free energy becomes
$\mathbb{G}_{\text{Bethe}}(\{\nu_i\}_{i\in\mathcal{V}}, \{\nu_{ij}\}_{(ij)\in\mathcal{E}}) := \cdots$

$$\cdots \sum_{i\in\mathcal{V}} \sum_{x_i\in\mathcal{X}} \nu_i(x_i) \left[-\ln(\psi_i(x_i)) + (1-d_i)\ln(\nu_i(x_i))\right]$$
$$+ \sum_{(ij)\in\mathcal{E}} \sum_{x_i,x_j\in\mathcal{X}} \nu_{ij}(x_i,x_j) \left[-\ln(\psi_{ij}(x_i,x_j)) + \ln(\nu_{ij}(x_i,x_j))\right]$$

Natural constraints:
$$\begin{array}{l} \sum_{x_i\in\mathcal{X}} \nu_i(x_i) = 1,\ i\in\mathcal{V}, \\ \sum_{x_i,x_j\in\mathcal{X}} \nu_{ij}(x_i,x_j) = 1,\ (ij)\in\mathcal{E}, \\ \nu_i(x_i) = \sum_{x_j\in\mathcal{X}} \nu_{ij}(x_i,x_j),\ i\in\mathcal{V},\ x_i\in\mathcal{X},\ (ij)\in\mathcal{E}. \end{array}$$

Remark

1) Number of variables $|\mathcal{X}|^{|\mathcal{V}|}$ *in Gibbs free energy* $\to$ $|\mathcal{V}|\cdot|\mathcal{X}| + |\mathcal{E}|\cdot|\mathcal{X}|^2$ *in Bethe free energy*

2) For marginal distributions ν_i, ν_{ij} *satisfying constraints, existence of* ν *on* $\mathcal{X}^{\mathcal{V}}$ *with these marginals is not guaranteed*

Bethe Free Energy minimization and belief propagation

Lagrangian for BFE minimization:

$$\begin{aligned}
&\mathcal{L}((\nu_i,\nu_{ij}),(\alpha_i,\beta_{ij},\lambda_{i\to j})) = \mathbb{G}_{\text{Bethe}}(\{\nu_i\},\{\nu_{ij}\}) \\
&+\sum_{i\in\mathcal{V}}\alpha_i\big(\sum_{x_i\in\mathcal{X}}\nu_i(x_i)-1\big) \\
&+\sum_{(ij)\in\mathcal{E}}\beta_{ij}\big(\sum_{x_i,x_j\in\mathcal{X}}\nu_{ij}(x_i,x_j)-1\big) \\
&+\sum_{(i\to j)}\sum_{x_i\in\mathcal{X}}\lambda_{i\to j}(x_i)\big[\nu_i(x_i)-\sum_{x_j\in\mathcal{X}}\nu_{ij}(x_i,x_j)\big].
\end{aligned} \tag{1}$$

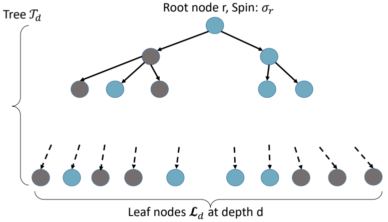
Stationarity conditions: $\partial\mathcal{L}/\partial\alpha_i = \partial\mathcal{L}/\partial\beta_{ij} = \partial\mathcal{L}/\partial\lambda_{i\to j}(x_i) = 0$, and

$$\begin{aligned}
&-\ln\psi_{ij}(x_i,x_j)+1+\ln\nu_{ij}(x_i,x_j)-\lambda_{i\to j}(x_i)-\lambda_{j\to i}(x_j)+\beta_{ij}=0, \\
&-\ln\psi_i(x_i)+(1-d_i)(1+\ln\nu_i(x_i))+\alpha_i+\sum_{j\sim i}\lambda_{i\to j}(x_i)=0.
\end{aligned}$$

Theorem

Assume that $\psi_i(x_i)$ and $\psi_{ij}(x_i,x_j)>0$ for all $(i,j)\in\mathcal{E}$ and $x_i,x_j\in\mathcal{X}$. There is then a one-to-one correspondence between stationary points of Lagrangian (1) associated with BFE minimization and fixed points $b_{i\to j}(x_i)$ of BP, obtained via relationships $b_{i\to j}(x_i)=\exp(\lambda_{i\to j}(x_i))$.

The tree reconstruction problem



Tree $\mathcal{T}$, root r. $\mathcal{L}_d$: nodes in generation d (at distance d from r).
Tree of nodes of generations $0, \ldots, d$: $\mathcal{T}_d = (V_d, E_d)$.

$\sigma_i \in [q]$: "trait" of individual i. $p(i)$: parent of i.

Probabilistic transmission: $\mathbb{P}(\sigma_{\mathcal{L}_d} = s_{\mathcal{L}_d} | \mathcal{T}, \sigma_{V_{d-1}}) = \prod_{i \in \mathcal{L}_d} P_{\sigma_{p(i)} s_i}$ where P: stochastic matrix, assumed irreducible, with invariant distribution ν on $[q]$

The tree reconstruction problem

Assume root spin $\sigma_r \sim \nu$. Then $\mathbb{P}(\sigma_{V_d} = s_{V_d} | \mathcal{T}) = \nu_{s_r} \prod_{(i,j) \in E_d, i=p(j)} P_{s_i s_j}$

$\to$ A tree Markov field.

Special case: $P_{\tau\tau} = p$, $P_{\tau s} = \frac{1-p}{q-1}$, $s \neq \tau$: symmetric Potts model ($q = 2$: Ising model).

Question: can one estimate σ_r from $\mathcal{T}_d, \sigma_{\mathcal{L}_d}$ non-trivially as $d \to \infty$?

Assumptions on $\mathcal{T}$: either deterministic, or Galton-Watson branching tree.

Information theory background

Shannon entropy

$H(\nu) = \sum_s \nu_s \ln(1/\nu_s)$ $\rightarrow$ also denote $H(X)$ for r.v. X

Conditional entropy

$$\begin{aligned} H(X|Y) &:= \sum_{x,y} p_{X,Y}(x,y) \ln\left(\frac{1}{p_{X|Y}(x|y)}\right) \\ &= H(\{X,Y\}) - H(Y) \\ &= \sum_y p_Y(y) H(\mathcal{L}(X|Y=y)). \end{aligned}$$

Mutual information

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(\{X,Y\}) \\ &= \sum_{x,y} p_{X,Y}(x,y) \ln\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right) \\ &= D(p_{X,Y} \| p_X \otimes p_Y). \end{aligned}$$

Non-negativity of KL divergence $\Rightarrow I(X;Y) \geq 0$.

Information theory background

Conditional mutual information

$$\begin{aligned} I(X;Y|Z) &:= H(X|Z) + H(Y|Z) - H(\{X,Y\}|Z) \\ &= \sum_z p_Z(z) D(p_{\{X,Y\}|Z=z} \| p_{X|Z=z} \otimes p_{Y|Z=z}) \end{aligned}$$

$\Rightarrow I(X;Y|Z) \geq 0.$

Chain rule: $I(X;\{Y,Z\}) = I(X;Z) + I(X;Y|Z)$.

Lemma

Data processing inequality: Let X, Y, Z such that X, Y independent conditional on Z. Then $I(X;Y) \leq I(X;Z)$.

Proof.

$I(X;(Y,Z)) = I(X;Z) + I(X;Y|Z) = I(X;Z)$ (conditional independence)
$I(X;(Y,Z)) = I(X;Y) + I(X;Z|Y) \geq I(X;Y)$ (non-negativity of mutual information) □

Non-trivial tree reconstruction

Let $\mathcal{F}_d = \sigma(\mathcal{T}_d, \sigma_{V_d})$, $\mathcal{G}_d = \sigma(\mathcal{T}_d, \sigma_{\mathcal{L}_d})$, $\hat{\nu}_{s,d} = \mathbb{P}(\sigma_r = s | \mathcal{G}_d)$, $s \in [q]$.
Then

$$\begin{aligned} I(\sigma_r; \mathcal{G}_d) &= H(\sigma_r) - H(\sigma_r | \mathcal{G}_d) \\ &= \mathbb{E} \sum_{s\in[q]} \hat{\nu}_{s,d} \ln(\hat{\nu}_{s,d}/\nu_s) \\ &= \mathbb{E} D(\{\hat{\nu}_{s,d}\}_{s\in[q]} \| \nu). \end{aligned}$$

For $\mathcal{T}$: deterministic or Galton-Watson, conditionally on $\mathcal{G}_d$, σ_r and $\mathcal{G}_{d+1}$ are independent.
$\Rightarrow$ (data processing inequality) $I(\sigma_r; \mathcal{G}_d) \geq I(\sigma_r; \mathcal{G}_{d+1})$.
$\Rightarrow \exists \lim_{d\to\infty} I(\sigma_r; \mathcal{G}_d)$.

Definition

tree reconstruction is feasible if and only if $\lim_{d\to\infty} I(\sigma_r; \mathcal{G}_d) > 0$.

Proposition

Non-reconstructibility, i.e. $\lim_{d\to\infty} I(\sigma_r; \mathcal{G}_d) = 0$ *is equivalent to* $\forall s \in [q]$, $\hat{\nu}_{s,d} \overset{d\to\infty}{\to} \nu_s$ *in probability.*

Proof: By irreducibility, $\inf_{s\in[q]} \nu_s > 0$.
Map $f : M([q]) \to \mathbb{R}_+$, $f(\mu) := D(\mu\|\nu)$ is thus continuous, null only for $\mu = \nu$.
Recall that $I(\sigma_r; \mathcal{G}_d) = \mathbb{E}f(\{\hat{\nu}_{\cdot,d}\})$. Thus if $I(\sigma_r; \mathcal{G}_d) \overset{d\to\infty}{\to} 0$, then for all $\epsilon > 0$, $\mathbb{P}(\{\hat{\nu}_{\cdot,d}\} \in \{\mu \in M([q]) : f(\mu) \geq \epsilon\}) \overset{d\to\infty}{\to} 0$. Hence $\{\hat{\nu}_{\cdot,d}\} \to \nu$ in probability.
Conversely if $\{\hat{\nu}_{\cdot,d}\} \to \nu$ in probability, by continuity of f and compactness of $M([q])$, $I(\sigma_r; \mathcal{G}_d) \to f(\nu) = 0$.

Illustration of non-reconstructibility

Estimate $\hat{\sigma}_r = \phi(\mathcal{G}_d)$ maximizing $\mathbb{P}(\hat{\sigma}_r = \sigma_r)$, i.e. maximizing
$\mathbb{E}(\mathbb{I}_{\hat{\sigma}_r=\sigma_r}) = \mathbb{E}\sum_{s\in[q]} \hat{\nu}_{s,d}\mathbb{I}_{\hat{\sigma}_r=s}$,
achieved by $\hat{\sigma}_r \in \arg\max\{\hat{\nu}_{s,d}\}_{s\in[q]}$, yielding
$\mathbb{P}(\hat{\sigma}_r = \sigma_r) = \mathbb{E}\max_{s\in[q]}\hat{\nu}_{s,d}$.

For ν: uniform on $[q]$,

non-reconstructibility $\Leftrightarrow \max_{s\in[q]}\hat{\nu}_{s,d} \underset{d\to\infty}{\overset{\text{prob.}}{\to}} \frac{1}{q} \Leftrightarrow \mathbb{E}\max_{s\in[q]}\hat{\nu}_{s,d} \overset{d\to\infty}{\to} \frac{1}{q}$

Hence non-reconstructibility equivalent to $\mathbb{P}(\hat{\sigma}_r = \sigma_r) \overset{d\to\infty}{\to} \frac{1}{q}$,
performance that can be achieved trivially
(take fixed or random σ_r, independent of $\mathcal{G}_d$)

Census reconstructibility

Define generation d's **census**: $X_d = \{X_{s,d}\}_{s\in[q]}$ where $X_{s,d} := \sum_{i\in\mathcal{L}_d} \mathbb{I}_{\sigma_i = s}$.

Definition

Census reconstructibility holds if $\lim_{d\to\infty} I(\sigma_r; X_d) > 0$.

Remark

Data processing inequality ensures $\lim_{d\to\infty} I(\sigma_r; X_d)$ *exists for Galton-Watson tree* $\mathcal{T}$*.*
It also ensures $I(\sigma_r; \mathcal{G}_d) \geq I(\sigma_r; X_d)$*. Hence census reconstructibility implies (tree) reconstructibility.*

Assume $\mathcal{T}$: Galton-Watson, with r.v. Z: number of children verifying $\mathbb{E}Z = \alpha > 1$ and $\mathbb{E}Z^2 < \infty$.
For transition matrix $P_{s\tau} := \mathbb{P}(\sigma_i = \tau | \sigma_{p(i)} = s)$, let $\lambda_2(P)$: eigenvalue of P with second largest modulus ($\lambda_1(P) = 1$).

Census reconstructibility and Kesten-Stigum threshold

Theorem

If $\alpha|\lambda_2(P)|^2 > 1$, census reconstructibility holds.

Proof: Let $\{x_s\}_{s\in[q]}$: eigenvector of P for $\lambda_2(P)$. Necessarily, x non-constant, since constant vector: eigenvector for $\lambda_1(P) = 1$.

Let $Z_d = \phi(X_d) = \sum_{s\in[q]} x_s X_{s,d}(\alpha\lambda_2)^{-d}$.

We shall show that $\liminf I(\sigma_r; Z_d) > 0$, using

Lemma

$\{Z_d\}$: uniformly integrable $\mathcal{F}_d$-martingale, where $\mathcal{F}_d = \sigma(\mathcal{T}_d, \sigma_{V_d})$.

Background on martingales

Definition

Family of random variables $\{M_d\}_{d\geq 0}$ is a martingale for filtration $\{\mathcal{G}_d\}_{d\geq 0}$ iff $\forall d \geq 0,\ \mathbb{E}(M_{d+1}|\mathcal{G}_d) = M_d$.
It is uniformly integrable iff $\lim_{A\to+\infty} \sup_d \mathbb{E}(|M_d|\mathbb{I}_{|M_d|>A}) = 0$.

Theorem

For uniformly integrable $\{\mathcal{G}_d\}$-martingale $\{M_d\}$, there exists $\mathcal{G}_\infty$-measurable random variable M_∞ such that:
$M_d = \mathbb{E}(M_\infty|\mathcal{G}_d),\ d \geq 0;$
$M_d \stackrel{d\to\infty}{\to} M_\infty$ *almost surely and in* $\mathbb{L}^1$.

Theorem

Let $\{\mathcal{H}_d\}$: decreasing sequence of σ-fields ($\mathcal{H}_{d+1} \subseteq \mathcal{H}_d$). For integrable r.v. X, then $X_d := \mathbb{E}(X|\mathcal{H}_d)$ verifies
$X_d \stackrel{d\to\infty}{\to} \mathbb{E}(X|\mathcal{H}_\infty)$ *almost surely and in* $\mathbb{L}^1$.

Lemma's proof

$$
\begin{aligned}
\mathbb{E}(Z_d|\mathcal{F}_{d-1}) &= (\alpha\lambda_2)^{-d}\mathbb{E}\left[\sum_{i\in\mathcal{L}_{d-1}}\sum_{j:p(j)=i} x_{\sigma_j}|\mathcal{F}_{d-1}\right] \\
&= (\alpha\lambda_2)^{-d}\sum_{i\in\mathcal{L}_{d-1}}\alpha\sum_{s\in[q]} P_{\sigma_i s}x_s \\
&= Z_{d-1}.
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var} Z_d &= \mathrm{Var}(\mathbb{E}(Z_d|\mathcal{F}_{d-1})) + \mathbb{E}(\mathrm{Var}(Z_d|\mathcal{F}_{d-1})) \\
&= \mathrm{Var}(Z_{d-1}) + |\alpha\lambda_2|^{-2d}\mathbb{E}[\mathrm{Var}(\sum_{i\in\mathcal{L}_d} x_{\sigma_i}|\mathcal{F}_{d-1})] \\
&= \mathrm{Var}(Z_{d-1}) + |\alpha\lambda_2|^{-2d}\mathbb{E}[\sum_{i\in\mathcal{L}_{d-1}}\mathrm{Var}(\sum_{j:p(j)=i} x_{\sigma_i}|\mathcal{F}_{d-1})] \\
&\le \mathrm{Var}(Z_{d-1}) + |\alpha\lambda_2|^{-2d}\mathbb{E}(|\mathcal{L}_{d-1}|)\sup_s |x_s|^2\mathbb{E}(Z^2) \\
&\le \mathrm{Var}(Z_{d-1}) + C(\alpha|\lambda_2|^2)^{-d}.
\end{aligned}
$$

Hence $\sup_{d>0}\mathrm{Var}(Z_d) \le \frac{C'}{\alpha|\lambda_2|^2-1} < +\infty$.

Uniform integrability follows: $\sup_d \mathbb{E}(|Z_d|\mathbb{I}_{|Z_d|>A}) \le \sup_d \frac{1}{A}\mathbb{E}|Z_d|^2 \le \frac{C''}{A}$.

By martingale convergence theorem, $Z_d \xrightarrow[d\to\infty]{\text{a.s.},\ \mathbb{L}^1} Z_\infty$.

Necessarily $\mathbb{E}[Z_\infty|\mathcal{F}_0] = Z_0 = x_{\sigma_r}$.
Assume that for Lebesgue almost all $t \in \mathbb{R}$, $\mathbb{I}_{Z_\infty \le t}$ independent of σ_r, and thus $\forall s \in [q]$, $\mathbb{P}(Z_\infty \le t|\sigma_r = s) = \mathbb{P}(Z_\infty \le t)$.
Then $\mathbb{E}(Z_\infty|\sigma_r = s)$ independent of $s \in [q]$, a contradiction.

Thus there exists a set of positive Lebesgue measure of $t \in \mathbb{R}$ such that Z_∞, σ_r non-independent.

Choose such t that is also a continuity point of $\mathbb{P}(Z_\infty \le t|\sigma_r = s)$ for all $s \in [q]$
For such t, a.s. $\lim_{d\to\infty} \mathbb{P}(\sigma_r = s|Z_d \le t) = \mathbb{P}(\sigma_r = s|Z_\infty \le t)$
$\Rightarrow \lim_{d\to\infty} I(\sigma_r; \mathbb{I}_{Z_d \le t}) = I(\sigma_r; \mathbb{I}_{Z_\infty \le t}) > 0$
$\Rightarrow \lim_{d\to\infty} I(\sigma_r; Z_d) > 0$, i.e. census reconstructibility