Dyskretne procesy stacjonarne o nieskończonej entropii nadwyżkowej

Łukasz Dębowski
ldebowsk@ipipan.waw.pl

Instytut Podstaw Informatyki PAN

Co to jest entropia nadwyżkowa?

Niech $(X_i)_{i\in\mathbb{Z}}$ będzie procesem stacjonarnym, $X_{k:l} = (X_i)_{k\le i\le l}$.

Entropia bloku długości n:

$$H(n) = H(X_{1:n}) = -\mathbb{E} \log P(X_{1:n}).$$

Informacja wzajemna między przyległymi blokami długości n:

$$E(n) = I(X_{-n+1:0}; X_{1:n}) = 2H(n) - H(2n).$$

Entropia nadwyżkowa:

$$E = I(X_{-\infty:0}; X_{1:\infty}) = \lim_{n\to\infty} E(n).$$

Entropia nadwyżkowa jest miarą pamięci procesu.

Motywacja lingwistyczna

Hilberg (1990) przypuścił, że dla języka naturalnego zachodzi

$$\mathbf{E(n)} \propto \mathbf{n}^{\boldsymbol{\beta}}, \quad \boldsymbol{\beta} \approx \mathbf{0.5}.$$

Interesują mnie procesy, dla których **E(n)** rozbiega potęgowo (oraz ewentualne interpretacje lingwistyczne tych procesów).

Dwie intuicje

- Niech $(X_i)_{i\in\mathbb{Z}}$ — ukryty proces Markowa, tzn. $X_i = f(Y_i)$, gdzie $(Y_i)_{i\in\mathbb{Z}}$ — stacjonarny proces Markowa.

$$\begin{aligned} E(n) = I(X_{-n+1:0}; X_{1:n}) &\le I(Y_{-n+1:0}; Y_{1:n}) \\ &= I(Y_0; Y_1) \le H(Y_1) \end{aligned}$$

$H(Y_1) < \infty$, gdy Y_i przybierają skończoną liczbę wartości.

- Niech $\mathcal{F}$ — algebra niezmiennicza procesu $(X_i)_{i\in\mathbb{Z}}$.

$$E = H(\mathcal{F}) + I(X_{-\infty:0}; X_{1:\infty}|\mathcal{F})$$

$H(\mathcal{F}) = \infty$, gdy istnieje ciągła zmienna rzeczywista mierzalna względem algebry niezmienniczej procesu $(X_i)_{i\in\mathbb{Z}}$ (parametr procesu w sensie statystyki bayesowskiej). Procesy takie nazywam mocno nieergodycznymi.

Rozkład zmiennych ukrytych

Przypuśćmy, że $\mathbf{Y_i : \Omega \to \mathbb{Y}}$, gdzie $\mathbb{Y} = \{\sigma_{nk}\}_{1\le k\le r(n), n\ge 2}$. Oznaczmy „poziomy”

$$\mathbf{T_n := \{\sigma_{nk}\}_{1\le k\le r(n)}},$$

oraz przypuśćmy, że wskaźnik poziomu

$$\mathbf{N_i := n \iff Y_i \in T_n}$$

ma rozkład

$$\mathbf{P(N_i = n) = \frac{C}{n \log^{\alpha} n}}.$$

Dla $\alpha \in (1, 2]$ mamy $\mathbf{H(Y_i) \ge H(N_i) = \infty}$.

Ograniczenie informacji wzajemnej

Twierdzenie

Przypuśćmy, że $Y_i : \Omega \to \mathbb{Y}$, gdzie $\mathbb{Y} = \{\sigma_{nk}\}_{1 \le k \le r(n), n \ge 2}$, a funkcja $r(n)$ spełnia $r(n) = O(n^p)$. Ponadto załóżmy, że

$$P(Y_i = \sigma_{nk}) = \frac{1}{r(n)} \cdot \frac{C}{n \log^\alpha n},$$

gdzie $\alpha \in (1, 2]$ i $C^{-1} = \sum_{n=2}^{\infty} (n \log^\alpha n)^{-1}$.
Niech $f : \mathbb{Y} \to \mathbb{X}$, gdzie $\mathbb{X} = \{0, 1, ..., D-1\}$, oraz $X_i = f(Y_i)$.
Wówczas

$$E(n) = I(X_{-n+1:0}; X_{1:n}) = \begin{cases} O\left(n^{2-\alpha}\right), & \alpha \in (1, 2), \\ O\left(\log n\right), & \alpha = 2. \end{cases}$$

Szkic dowodu

Niech $\mathbf{B}$ będzie zdarzeniem. Mamy

$$\mathbf{E(n) \leq P(B)I(X_{-n+1}^{0}; X_1^n|B) + P(B^c)I(X_{-n+1}^{0}; X_1^n|B^c) + 1.}$$

Połóżmy $\mathbf{B = (N_o \leq 2^n)}$, gdzie $\mathbf{N_0}$ jest wskaźnikiem poziomu $\mathbf{Y_0}$. Ponieważ $\mathbf{(Y_i)_{i\in\mathbb{Z}}}$ jest procesem Markowa,

$$\mathbf{I(X_{-n+1}^{0}; X_1^n|B) \leq I(Y_{-n+1}^{0}; Y_1^n|B) \leq H(Y_0|B).}$$

Z drugiej strony

$$\mathbf{I(X_{-n+1}^{0}; X_1^n|B^c) \leq H(X_{-n+1}^{0}|B^c) \leq n \log |\mathbb{X}| .}$$

Można policzyć, że

$$\mathbf{P(B)H(Y_0|B) = \begin{cases} \Theta\left(n^{2-\alpha}\right), & \alpha \in (1,2), \\ \Theta\left(\log n\right), & \alpha = 2, \end{cases}}$$

$$\mathbf{nP(B^c) = \Theta\left(n^{2-\alpha}\right).}$$

Heavy Tailed Periodic Mixture I

Połóżmy $\mathbb{Y} = \{\sigma_{nk}\}_{1\le k\le n, n\ge 2}$,

$$P(Y_i = \sigma_{nk}) = \frac{1}{n} \cdot \frac{C}{n \log^\alpha n},$$

$$P(Y_{i+1} = \sigma_{nk} | Y_i = \sigma_{ml}) = \begin{cases} 1_{\{n=m, k=l+1\}}, & 1 \le l \le m-1, \\ 1_{\{n=m, k=1\}}, & l = m, \end{cases}$$

$$X_i = \begin{cases} 0, & Y_i = \sigma_{nk},\ 1 \le k \le n-1, \\ 1, & Y_i = \sigma_{nn}. \end{cases}$$

Wówczas

$$E(n) = \begin{cases} \Theta(\log^{2-\alpha} n), & \alpha \in (1, 2), \\ \Theta(\log \log n), & \alpha = 2. \end{cases}$$

Szkic dowodu

Dowód polega na skonstruowaniu zmiennych $\mathbf{D_n}$, które są funkcjami zarówno $\mathbf{X^0_{-n+1}}$ jak $\mathbf{X^n_1}$.
Korzystając z tej własności, otrzymujemy

$$\begin{aligned}\mathbf{E(n)} &= \mathbf{I(X^0_{-n+1}, D_n; X^n_1) = I(D_n; X^n_1) + I(X^0_{-n+1}; X^n_1|D_n)} \\ &= \mathbf{H(D_n) + I(X^0_{-n+1}; X^n_1|D_n).}\end{aligned}$$

W dalszej kolejności ograniczamy $\mathbf{H(D_n)}$ oraz $\mathbf{H(X^n_1|D_n)}$.
W przypadku procesu z poprzedniego slajdu kładziemy

$$\mathbf{D_n} = \begin{cases} \mathbf{N_0}, & \mathbf{2N_0 \le n} \\ \mathbf{0}, & \mathbf{2N_0 > n} \end{cases} = \begin{cases} \mathbf{N_1}, & \mathbf{2N_1 \le n,} \\ \mathbf{0}, & \mathbf{2N_1 > n.} \end{cases}$$

Heavy Tailed Periodic Mixture II

Niech $\mathbf{s(n)}$ — długość rozwinięcia binarnego liczby $\mathbf{n}$ oraz $\mathbf{b(n,k)}$ — $\mathbf{k}$-ta cyfra rozwinięcia binarnego liczby $\mathbf{n}$. Połóżmy $\mathbb{Y} = \{\sigma_{nk}\}_{1\le k\le s(n), n\ge 2}$,

$$P(Y_i = \sigma_{nk}) = \frac{1}{s(n)} \cdot \frac{C}{n \log^\alpha n},$$

$$P(Y_{i+1} = \sigma_{nk} | Y_i = \sigma_{ml}) = \begin{cases} 1_{\{n=m,k=l+1\}}, & 1 \le l \le s(m) - 1, \\ 1_{\{n=m,k=1\}}, & l = s(m), \end{cases}$$

$$X_i = \begin{cases} 2, & Y_i = \sigma_{n1}, \\ b(n,k), & Y_i = \sigma_{nk}, 2 \le k \le s(n). \end{cases}$$

Wówczas

$$E(n) = \begin{cases} \Theta(n^{2-\alpha}), & \alpha \in (1,2), \\ \Theta(\log n), & \alpha = 2. \end{cases}$$

Heavy Tailed Mixing Copy

Niech $\mathbf{s(n)}$ — długość rozwinięcia binarnego liczby $\mathbf{n}$
oraz $\mathbf{b(n,k)}$ — $\mathbf{k}$-ta cyfra rozwinięcia binarnego liczby $\mathbf{n}$.
Połóżmy $\mathbb{Y} = \{\sigma_{nk}\}_{1\le k\le 3s(n), n\ge 2}$,

$$P(Y_i = \sigma_{nk}) = \frac{1}{3s(n)} \cdot \frac{C}{n \log^\alpha n},$$

$$P(Y_{i+1} = \sigma_{nk} | Y_i = \sigma_{ml}) = \begin{cases} 1_{\{n=m, k=l+1\}}, & 1 \le l \le 3s(m) - 1, \\ p(n) 1_{\{k=1\}}, & l = 3s(m), \end{cases}$$

$$p(n) \propto \frac{1}{3s(n)} \cdot \frac{1}{n \log^\alpha n},$$

$$X_i = \begin{cases} 2, & Y_i = \sigma_{n1}, \\ b(n,k), & Y_i = \sigma_{nk},\ 2 \le k \le s(n), \\ 3, & Y_i = \sigma_{nk},\ s(n) + 1 \le k \le 2s(n) + 1, \\ b(n, k - 2s(n)), & Y_i = \sigma_{nk},\ 2s(n) + 2 \le k \le 3s(n). \end{cases}$$

Binarny proces wymienialny

Rozważmy rodzinę binarnych rozkładów IID

$$P(X_{1:n} = x_{1:n}|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}.$$

Skonstruujmy proces $(X_i)_{i\in\mathbb{Z}}$ taki, że

$$P(X_{1:n} = x_{1:n}) = \int_0^1 P(X_{1:n} = x_{1:n}|\theta)\pi(\theta)d\theta$$

dla rozkładu a priori $\pi(\theta) > 0$. Dla $Y = \lim_n n^{-1}\sum_{i=1}^{n} X_i$ mamy

$$P(Y \le y) = \int_0^y \pi(\theta)d\theta.$$

Proces $(X_i)_{i\in\mathbb{Z}}$ jest mocno nieergodyczny, ponieważ Y ma rozkład ciągły. Jednakże blok $X_{1:n}$ jest warunkowo niezależny od $X_{n+1:2n}$ względem sumy $S_n := \sum_{i=1}^{n} X_i$. Zatem

$$E(n) = I(X_{1:n}; X_{n+1:2n}) = I(S_n; X_{n+1:2n}) \le H(S_n) \le \log(n+1).$$

Procesy Santa Fe

Proces $(X_i)_{i\in\mathbb{Z}}$ postaci

$$X_i := (K_i, Z_{K_i}),$$

gdzie $(K_i)_{i\in\mathbb{Z}}$ i $(Z_k)_{k\in\mathbb{N}}$ są niezależnymi procesami IID,

$$P(K_i = k) = k^{-1/\beta}/\zeta(\beta^{-1}), \qquad \beta \in (0,1),$$
$$P(Z_k = z) = \frac{1}{2}, \qquad z \in \{0,1\}.$$

$Y = \sum_{k=1}^{\infty} 2^{-k} Z_k$ — mierzalna względem algebry niezmienniczej.

Interpretacja lingwistyczna

Proces $(X_i)_{i\in\mathbb{Z}}$ jest ciągiem losowych stwierdzeń niesprzecznie opisujących stan „wcześniej" wylosowanego obiektu $(Z_k)_{k\in\mathbb{N}}$.
$X_i = (k, z)$ stwierdza, że k-ty bit $(Z_k)_{k\in\mathbb{N}}$ ma wartość $Z_k = z$.

E(n) dla procesu Santa Fe

$$
\begin{aligned}
E(n) &= I\left(X_{1:n}; X_{n+1:2n}\right) \\
&= \sum_{k=1}^{\infty} I(X_{1:n}; X_{n+1:2n}; Z_k) \\
&= \sum_{k=1}^{\infty} (1 - [1 - P(K_i = k)]^n)^2 \\
&\approx \int_1^{\infty} \left(1 - \left(1 - \frac{k^{-1/\beta}}{\zeta(\beta^{-1})}\right)^n\right)^2 dk \\
&\to \frac{n^\beta}{[\zeta(\beta^{-1})]^\beta} \int_0^1 \frac{(1-u)^2}{u(-\ln u)^{\beta+1}} du \\
&= \frac{(2-2^\beta)\Gamma(1-\beta)}{[\zeta(\beta^{-1})]^\beta} \cdot n^\beta
\end{aligned}
$$

Kodowanie stacjonarne zmiennej długości

- Funkcję $f : \mathbb{X} \to \mathbb{Y}^+$ rozszerzamy do funkcji $f^{\mathbb{Z}} : \mathbb{X}^{\mathbb{Z}} \to \mathbb{Y}^{\mathbb{Z}}$,
$$f^{\mathbb{Z}}((x_i)_{i\in\mathbb{Z}}) := \dots f(x_{-1})f(x_0).f(x_1)f(x_2)\dots, \quad x_i \in \mathbb{X}.$$
- Dla miary AMS ν na $(\mathbb{Y}^{\mathbb{Z}}, \mathcal{Y}^{\mathbb{Z}})$ średnia stacjonarna to
$$\bar{\nu}(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \nu \circ T^{-i}(A), \quad T((y_i)_{i\in\mathbb{Z}}) := (y_{i+1})_{i\in\mathbb{Z}}.$$
- $(X_i)_{i\in\mathbb{Z}}$ — proces stacjonarny o rozkładzie
$$P((X_i)_{i\in\mathbb{Z}} \in \cdot) = \mu.$$
- $(Y_i)_{i\in\mathbb{Z}} = f^{\mathbb{Z}}((X_i)_{i\in\mathbb{Z}})$ — proces AMS rozkładzie
$$P((Y_i)_{i\in\mathbb{Z}} \in \cdot) = \nu = \mu \circ \left(f^{\mathbb{Z}}\right)^{-1}.$$
- $(\bar{Y}_i)_{i\in\mathbb{Z}}$ — proces stacjonarny o rozkładzie
$$P((\bar{Y}_i)_{i\in\mathbb{Z}} \in \cdot) = \bar{\nu} = \overline{\mu \circ (f^{\mathbb{Z}})^{-1}}.$$

E(m) dla kodowania stacjonarnego procesu Santa Fe

- Weźmy
$$\mathbf{f(k,z) := b(k)z2},$$
gdzie $\mathbf{1b(k)} \in \{0,1\}^+$ jest rozwinięciem binarnym liczby $\mathbf{k}$.
- $(\mathbf{X_i})_{\mathbf{i}\in\mathbb{Z}}$ — proces Santa Fe o rozkładzie μ.
- $(\mathbf{Y_i})_{\mathbf{i}\in\mathbb{Z}}$ — proces o rozkładzie $\nu = \mu \circ \left(\mathbf{f}^{\mathbb{Z}}\right)^{-1}$.
- $(\bar{\mathbf{Y}}_\mathbf{i})_{\mathbf{i}\in\mathbb{Z}}$ — proces o rozkładzie $\bar{\nu} = \overline{\mu \circ (\mathbf{f}^{\mathbb{Z}})^{-1}}$.
- Wszystkie trzy procesy są mocno nieergodyczne.
- Połóżmy $\mathbf{L} = \mathbb{E}\,|\mathbf{f(X_i)}|$ oraz $\mathbf{E(m)} = \mathbf{I}\left(\bar{\mathbf{Y}}_{\mathbf{1:n}};\bar{\mathbf{Y}}_{\mathbf{n+1:2n}}\right)$. Mamy

$$\lim_{m\to\infty}\frac{\mathbf{E(m)}}{\mathbf{m}^\beta} = \frac{1}{\mathbf{L}^\beta}\frac{(2-2^\beta)\Gamma(1-\beta)}{[\zeta(\beta^{-1})]^\beta}.$$

Uogólniony proces Santa Fe

Proces $(X_i)_{i\in\mathbb{Z}}$ postaci

$$X_i := (K_i, Z_{i,K_i}),$$

gdzie $(K_i)_{i\in\mathbb{Z}}$ i $(Z_{ik})_{i\in\mathbb{Z}}$, $k \in \mathbb{N}$, są procesami niezależnymi,

$$P(K_i = k) = k^{-1/\beta}/\zeta(\beta^{-1}), \qquad (K_i)_{i\in\mathbb{Z}} \sim \text{IID},$$

zaś $(Z_{ik})_{i\in\mathbb{Z}}$ są łańcuchami Markowa o rozkładzie

$$P(Z_{ik} = z) = \frac{1}{2},$$

$$P(Z_{ik} = z | Z_{i-1,k} = z) = 1 - p_k.$$

Proces $(X_i)_{i\in\mathbb{Z}}$ jest procesem mieszającym dla $p_k \in (0,1)$.

Interpretacja lingwistyczna

Obiekt $(Z_{ik})_{k\in\mathbb{N}}$ opisywany w tekście $(X_i)_{i\in\mathbb{Z}}$ jest funkcją czasu i.

$\mathbf{E(n)}$ dla uogólnionego procesu Santa Fe

Połóżmy $\mathbf{E(n) = I\left(X_{1:n}; X_{n+1:2n}\right)}$. Mamy

$$\limsup_{n\to\infty} \frac{E(n)}{n^\beta} \le \frac{(2-2^\beta)\Gamma(1-\beta)}{[\zeta(\beta^{-1})]^\beta}.$$

Dolne granice w szczególnych przypadkach są następujące:

1. Jeżeli $\mathbf{p_k \le P(K_i = k)}$, to

$$\liminf_{n\to\infty} \frac{E(n)}{n^\beta} \ge A(\beta).$$

2. Jeżeli $\mathbf{\lim_{k\to\infty} p_k / P(K_i = k) = 0}$, to

$$\lim_{n\to\infty} \frac{E(n)}{n^\beta} = \frac{(2-2^\beta)\Gamma(1-\beta)}{[\zeta(\beta^{-1})]^\beta}.$$

$\mathbf{E(m)}$ dla kodowania stacjonarnego

Rozpatrzmy kodowanie takie samo jak poprzednio.
$(\mathbf{Y_i})_{i\in\mathbb{Z}}$ oraz $(\bar{\mathbf{Y}}_i)_{i\in\mathbb{Z}}$ są procesami ergodycznymi.
Połóżmy $\mathbf{L} = \mathbb{E}\,|\mathbf{f}(\mathbf{X_i})|$ oraz $\mathbf{E(m)} = \mathbf{I}\left(\bar{\mathbf{Y}}_{1:n}; \bar{\mathbf{Y}}_{n+1:2n}\right)$. Mamy

$$\limsup_{n\to\infty} \frac{E(m)}{m^\beta} \le \frac{1}{L^\beta} \frac{(2-2^\beta)\Gamma(1-\beta)}{[\zeta(\beta^{-1})]^\beta}.$$

Dolne granice w szczególnych przypadkach są następujące:

1. Jeżeli $p_k \le P(K_i = k)$, to

$$\liminf_{n\to\infty} \frac{E(m)}{m^\beta} \ge \frac{A(\beta)}{L^\beta}.$$

2. Jeżeli $\lim_{k\to\infty} p_k/P(K_i = k) = 0$, to

$$\lim_{n\to\infty} \frac{E(m)}{m^\beta} = \frac{1}{L^\beta} \frac{(2-2^\beta)\Gamma(1-\beta)}{[\zeta(\beta^{-1})]^\beta}.$$

Podsumowanie

Podałem przykłady procesów o nieskończonej entropii nadwyżkowej:

1. $\mathbf{E(n)} \asymp \mathbf{\log n}$ dla nieergodycznego ukrytego procesu Markowa.
2. $\mathbf{E(n)} \asymp \mathbf{n}^{\beta}$ dla nieergodycznego ukrytego procesu Markowa.
3. $\mathbf{E(n)} \asymp \mathbf{n}^{\beta}$ dla ergodycznego ukrytego procesu Markowa.
4. $\mathbf{E(n)} \asymp \mathbf{\log n}$ dla procesu mocno nieergodycznego.
5. $\mathbf{E(n)} \asymp \mathbf{n}^{\beta}$ dla mocno nieergodycznego procesu Santa Fe nad nieskończonym alfabetem.
6. $\mathbf{E(n)} \asymp \mathbf{n}^{\beta}$ dla mocno nieergodycznego procesu Santa Fe nad skończonym alfabetem.
7. $\mathbf{E(n)} \asymp \mathbf{n}^{\beta}$ dla mieszającego procesu Santa Fe nad nieskończonym alfabetem.
8. $\mathbf{E(n)} \asymp \mathbf{n}^{\beta}$ dla ergodycznego procesu Santa Fe nad skończonym alfabetem.

Moje prace

- Ł. Dębowski, (2012). Mixing, Ergodic, and Nonergodic Processes with Rapidly Growing Information between Blocks. *IEEE Transactions on Information Theory*, **58**:3392-3401.
- Ł. Dębowski, (2013). On Hidden Markov Processes with Infinite Excess Entropy. *Journal of Theoretical Probability*, w druku. (`http://arxiv.org/abs/1211.0834`)

`www.ipipan.waw.pl/~ldebowsk`