Adaptive compression over countable alphabets.

D. Bontemps, S. Boucheron, A. Garivier, E. Gassiat & M. Ohanessian

Paul Sabatier, Paris-Diderot, Paris-Sud

May 1, 2013

Outline

Universal compression

Lossless compression

Mapping sequences of symbols on sequences of $\{0, 1\}$ so as to minimize the expected length of codewords.

Example

If the source statistics is known, Huffmann coding provides an optimal solution, Arithmetic coding provides an almost optimal encoding.

Universality

The ability to encode efficiently sources with different statistics

Example

i) Lempel-Ziv, 1977-78 (zip, gzip, ...)

ii) Burrows-Wheeler transform, move to front (bzip)

iii) Context-Tree Weighting. (see O. Catoni, Saint-Flour Lecture Notes)

Non-ambiguous codes

Prefix codes

No codeword shall be a strict prefix of another codeword :

$$f(\omega) \cdot y = f(\omega') \Rightarrow y = \epsilon \wedge \omega = \omega'$$

Prefix codes are a special case of more general non-ambiguous codes (self-delimited codes).
Sequences of codewords from non-ambiguous codes can be parsed into codewords in a unique way.

Binary expansions of integers do not form a non-ambiguous binary code for integers.

Coding probabilities

Kraft-McMillan inequality

$\lambda : A \longrightarrow \mathbb{N}_+$ satisfies

$$\sum_{\omega \in A} |\mathcal{X}|^{-\lambda(\omega)} \leq 1,$$

iff
there is a prefix code $f : A \longrightarrow \mathcal{X}^*$
with

$$\ell[f(\omega)] = \lambda(\omega)$$

$\hookrightarrow$ No prefix code with codelength $\lfloor \log_2 x \rfloor + 1$ for integers.

Coding probabilities

Kraft-McMillan inequality

$\lambda : A \longrightarrow \mathbb{N}_+$ satisfies

$$\sum_{\omega \in A} |\mathcal{X}|^{-\lambda(\omega)} \leq 1,$$

iff
there is a prefix code $f : A \longrightarrow \mathcal{X}^*$
with

$$\ell[f(\omega)] = \lambda(\omega)$$

$\hookrightarrow$ No prefix code with codelength $\lfloor \log_2 x \rfloor + 1$ for integers.

Arithmetic coding w.r.t. Q^n

encodes $x_{1:n} = (x_1, \ldots, x_n) \in \mathcal{X}^n$
with codewordlength at most
$-\log Q^n(x_{1:n}) + 1$,

Coding probabilities

Kraft-McMillan inequality

$\lambda : A \longrightarrow \mathbb{N}_+$ satisfies

$$\sum_{\omega \in A} |\mathcal{X}|^{-\lambda(\omega)} \leq 1,$$

iff
there is a prefix code $f : A \longrightarrow \mathcal{X}^*$
with

$$\ell[f(\omega)] = \lambda(\omega)$$

$\hookrightarrow$ No prefix code with codelength $\lfloor \log_2 x \rfloor + 1$ for integers.

Arithmetic coding w.r.t. Q^n

encodes $x_{1:n} = (x_1, \ldots, x_n) \in \mathcal{X}^n$
with codewordlength at most
$-\log Q^n(x_{1:n}) + 1$,

Elias penultimate code

$l_1(0) = 1$,
$l_1(x) = \lfloor \log_2(x) \rfloor + 1, \quad \text{for} x > 0,$
Codeword length:

$$2l_1(l_1(x)) + l_1(x)$$

Redundancies

Definition (Redundancy of coding probability Q^n with respect to source P^n)

Expected difference between codelengths obtained by feeding an arithmetic coder with $Q^n(\mathbf{x})$ rather than with $P^n(\mathbf{x})$

$$D(P^n, Q^n) = \mathbb{E}_{P^n} \log \frac{P^n(X_{1:n})}{Q^n(X_{1:n})}$$

Redundancies

Definition (Redundancy of coding probability Q^n with respect to source P^n)

Expected difference between codelengths obtained by feeding an arithmetic coder with $Q^n(\mathbf{x})$ rather than with $P^n(\mathbf{x})$

$$D(P^n, Q^n) = \mathbb{E}_{P^n} \log \frac{P^n(X_{1:n})}{Q^n(X_{1:n})}$$

Definition (Minimax redundancy)

$$\begin{aligned} R^+(Q^n, \Lambda^n) &= \sup_{P \in \Lambda} D(P^n, Q^n) \\ R^+(\Lambda^n) &= \inf_{Q} \sup_{P \in \in \Lambda} D(P^n, Q^n) \end{aligned}$$

Redundancies

Definition (Redundancy of coding probability Q^n with respect to source P^n)

Expected difference between codelengths obtained by feeding an arithmetic coder with $Q^n(\mathbf{x})$ rather than with $P^n(\mathbf{x})$

$$D(P^n, Q^n) = \mathbb{E}_{P^n} \log \frac{P^n(X_{1:n})}{Q^n(X_{1:n})}$$

Definition (Minimax redundancy)

$$R^+(Q^n, \Lambda^n) = \sup_{P \in \Lambda} D(P^n, Q^n)$$

$$R^+(\Lambda^n) = \inf_Q \sup_{P \in \in \Lambda} D(P^n, Q^n)$$

Definition (Maximin redundancy)

π : prior distribution on sources

$$R_+(\Lambda^n) = \sup_\pi \inf_Q \mathbb{E}_\pi D(P^n, Q^n)$$

Redundancy corresponds to cumulative logarithmic/entropy/self-information loss in statistics and individual sequences analysis.

Cesa-Bianchi & Lugosi, Chap. 9 of Prediction, Learning and Games, 2006.

Finite alphabet with cardinality k

Λ : memoryless sources over finite alphabet with cardinality k

Minimax redundancy

$$R^+(\Lambda^n) = \frac{k-1}{2} \log \frac{n}{2\pi e} + O(1)$$

Rissanen, Ryabko, Shtarkov, Krichevsky, Trofimov, Barron, Clarke, Xie et al..

Krichevsky-Trofimov coding is asymptotically maximin and approximately minimax

$$\mathbb{KT}\left(X_{n+1} = a | X_{1:n} = x_{1:n}\right) = \frac{n_a(x_{1:n}) + \frac{1}{2}}{n + \frac{k}{2}},$$

$n_a =$ number of a's in $x_{1:n}$.

KT-coding

KT-coding mixes coding probabilities using Jeffrey's (least favorable) prior.

$\hookrightarrow$ no need to explicitly estimate the source

In statistical parlance

Data compression

- lossless coding
- redundancy
- universality
- envelope conditions
- computational efficiency
- online processing

Statistics

- estimating a density over $\mathbb{N}$
- cumulative relative entropy risk
- adaptivity
- thresholding
- metric entropy conditions

Envelop classes

Definition (Envelop function : $f : \mathbb{N} \to \mathbb{R}_+$ with $1 < \sum_{j>0} f(j) < \infty$.)

$\Lambda_f = \left\{ \mathbb{P} \ : \ \forall x \in \mathbb{N},\ \mathbb{P}^1\{x\} \leq f(x) \text{ and } \mathbb{P} \text{ is stationary and memoryless.} \right\}$.

$F(k) = 1 - \sum_{j>k} f(j)$ for $k \geq l_f = \max\{k : \ \sum_{j \geq k} f(j) \geq 1\}$.

$\overline{F} = 1 - F$. $U(t) = \inf\{x : F(x) \geq 1 - 1/t\}$

Envelop classes

Definition (Envelop function : $f : \mathbb{N} \to \mathbb{R}_+$ with $1 < \sum_{j>0} f(j) < \infty$.)

$\Lambda_f = \left\{ \mathbb{P} \ : \ \forall x \in \mathbb{N},\ \mathbb{P}^1\{x\} \leq f(x) \text{ and } \mathbb{P} \text{ is stationary and memoryless.} \right\}$.
$F(k) = 1 - \sum_{j>k} f(j)$ for $k \geq l_f = \max\{k : \ \sum_{j \geq k} f(j) \geq 1\}$.
$\overline{F} = 1 - F$. $U(t) = \inf\{x : F(x) \geq 1 - 1/t\}$

F_c has piecewise constant hazard rate, and satisfies $\overline{F}_c(n) = \overline{F}(n)$

$U_c(t) = \inf\{x : 1/\overline{F}_c(x) \geq t\}$.
If $X \sim F_c$ then $\lfloor X \rfloor + 1 \sim F$ and $U(t) = \lfloor U_c(t) \rfloor + 1$ for $t > 1$.

Envelop classes

Definition (Envelop function : $f : \mathbb{N} \to \mathbb{R}_+$ with $1 < \sum_{j>0} f(j) < \infty$.)

$\Lambda_f = \left\{ \mathbb{P} \ : \ \forall x \in \mathbb{N},\ \mathbb{P}^1\{x\} \leq f(x) \text{ and } \mathbb{P} \text{ is stationary and memoryless.} \right\}$.
$F(k) = 1 - \sum_{j>k} f(j)$ for $k \geq l_f = \max\{k : \ \sum_{j \geq k} f(j) \geq 1\}$.
$\overline{F} = 1 - F$. $U(t) = \inf\{x : F(x) \geq 1 - 1/t\}$

F_c has piecewise constant hazard rate, and satisfies $\overline{F}_c(n) = \overline{F}(n)$

$U_c(t) = \inf\{x : 1/\overline{F}_c(x) \geq t\}$.
If $X \sim F_c$ then $\lfloor X \rfloor + 1 \sim F$ and $U(t) = \lfloor U_c(t) \rfloor + 1$ for $t > 1$.

Lemma [Stochastic comparison by quantile coupling]

There exists a probability space where $X \sim G \in \Lambda_f$, $Y \sim F_c$ such that

$$\mathbb{P}\{X \leq Y\} = 1$$

An upper-bound on minimax redundancy

Theorem [BGG, 2009]

If Λ is a class of memoryless sources, with the tail envelope distribution function $\bar{F}_{\Lambda^1}(u) = \sum_{k>u} \hat{p}(k)$, then:

$$R^*(\Lambda^n) \leq \inf_{u:u\leq n} \left[n\bar{F}_{\Lambda^1}(u) \log_2 e + \frac{u-1}{2} \log_2 n \right] + 2\,.$$

Suggestion

If the envelop is known, choose threshold τ as the solution of $\bar{F}_{\Lambda^1}(u) = \frac{u}{n}$.

i) Encode symbols over threshold using Elias penultimate code

ii) Encode other symbols using Krichevsky-Trofimov mixture over alphabet $\{1, \ldots, \tau\}$.

A lower bound

For any prior μ on $\Lambda^1(f)$

$$\begin{aligned} R^*(\Lambda^n) &\geq \inf_{Q^n} \mathbb{E}_\mu D\left(P^n \mid Q^n\right) \\ &= \mathbb{E}_\mu D\left(P^n \mid \mathbb{E}_\mu P^n\right) \\ &= I\left(\theta; X_{1:n}\right) \end{aligned}$$

A lower bound

The Bayes risk coincides with the mutual information between the parameter and the observation.

For any prior μ on $\Lambda^1(f)$

$$\begin{aligned} R^*(\Lambda^n) &\geq \inf_{Q^n} \mathbb{E}_\mu D\left(P^n \mid Q^n\right) \\ &= \mathbb{E}_\mu D\left(P^n \mid \mathbb{E}_\mu P^n\right) \\ &= I\left(\theta; X_{1:n}\right) \end{aligned}$$

A lower bound

For any prior μ on $\Lambda^1(f)$

$$\begin{aligned} R^*(\Lambda^n) &\geq \inf_{Q^n} \mathbb{E}_\mu D\left(P^n \mid Q^n\right) \\ &= \mathbb{E}_\mu D\left(P^n \mid \mathbb{E}_\mu P^n\right) \\ &= I\left(\theta; X_{1:n}\right) \end{aligned}$$

The Bayes risk coincides with the mutual information between the parameter and the observation.

Two options

- Design of priors and ad hoc lower bounds on mutual information
- Design of analogs of Fano's lemma

$$H(\theta \mid \widehat{\theta}(X_{1:n})) \leq p_e \log |\Theta| + h(p_e)$$

with $p_e = \mathbb{P}(\theta \neq \widehat{\theta})$ and $h(p_e)$ entropy of Bernoulli distribution with parameter p_e

Cumulative entropy risk and metric entropy

Haussler & Opper, AoS, 1997

Minimax redundancy can be lower bounded using metric entropy of Λ_f^1 under Hellinger metric.

Hellinger distance

$H^2(P_1, P_2) = \sum_{k\in\mathbb{N}} \left(\sqrt{p_1(k)} - \sqrt{p_2(k)}\right)^2.$

ϵ-entropy

$$\mathcal{H}_\epsilon(\Lambda) = \ln \mathcal{D}_\epsilon(\Lambda)$$

where $\mathcal{D}_\epsilon(\Lambda)$: cardinality of the smallest finite partition of Λ^1 into sets of diameter at most ϵ.

Cumulative entropy risk and metric entropy

Haussler & Opper, AoS, 1997

Minimax redundancy can be lower bounded using metric entropy of Λ_f^1 under Hellinger metric.

Hellinger distance

$H^2(P_1, P_2) = \sum_{k\in\mathbb{N}} \left(\sqrt{p_1(k)} - \sqrt{p_2(k)}\right)^2$.

ϵ-entropy

$$\mathcal{H}_\epsilon(\Lambda) = \ln \mathcal{D}_\epsilon(\Lambda)$$

where $\mathcal{D}_\epsilon(\Lambda)$: cardinality of the smallest finite partition of Λ^1 into sets of diameter at most ϵ.

Haussler, Opper, AoS 1997

For any prior π on Λ_1

$$R^+(\Lambda^n) \quad \geq \quad \mathbb{E}_\pi \left[-\log \mathbb{E}_\pi e^{-n\frac{H^2(P_1,P_2)}{2}} \right]$$

Cumulative entropy risk and metric entropy

Haussler & Opper, AoS, 1997

Minimax redundancy can be lower bounded using metric entropy of Λ_f^1 under Hellinger metric.

Hellinger distance

$H^2(P_1, P_2) = \sum_{k\in\mathbb{N}} \left(\sqrt{p_1(k)} - \sqrt{p_2(k)}\right)^2$.

ϵ-entropy

$$\mathcal{H}_\epsilon(\Lambda) = \ln \mathcal{D}_\epsilon(\Lambda)$$

where $\mathcal{D}_\epsilon(\Lambda)$: cardinality of the smallest finite partition of Λ^1 into sets of diameter at most ϵ.

Haussler, Opper, AoS 1997

For any prior π on Λ_1

$$R^+(\Lambda^n) \geq \mathbb{E}_\pi \left[-\log \mathbb{E}_\pi e^{-n\frac{H^2(P_1,P_2)}{2}} \right]$$

Consequence

$$\begin{aligned} R^+(\Lambda^n) &\geq -\log\left(\frac{1}{\mathcal{D}_\epsilon(\Lambda_1)} + \exp\left(-n\frac{\epsilon^2}{2}\right)\right) \\ &\geq \log e \sup_\epsilon \min\left(H_\epsilon(\Lambda_1), \frac{n\epsilon^2}{2}\right) - 1 \end{aligned}$$

Haussler-Opper lower bound (proof)

Repeatedly using Fubini's theorem and Jensen's inequality.
Fubini

$$\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P_{\theta^*}^n(x_{1:n}) \frac{\int_\Theta \mathrm{d}\pi(\tilde{\theta}) \frac{\mathrm{d}P_{\tilde{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}}{\int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}}} = 1$$

Haussler-Opper lower bound (proof)

Fubini

$$\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P^n_{\theta^*}(x_{1:n}) \frac{\int_\Theta \mathrm{d}\pi(\tilde{\theta}) \frac{\mathrm{d}P_{\tilde{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}}{\int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}}} = 1$$

$$-\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P^n_{\theta^*}(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\tilde{\theta}) \frac{\mathrm{d}P_{\tilde{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}$$

Haussler-Opper lower bound (proof)

Fubini

$$\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P^n_{\theta^*}(x_{1:n}) \frac{\int_\Theta \mathrm{d}\pi(\tilde{\theta}) \frac{\mathrm{d}P_{\tilde{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}}{\int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}}} = 1$$

$$-\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P^n_{\theta^*}(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\tilde{\theta}) \frac{\mathrm{d}P_{\tilde{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}$$

$$\geq \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P^n_{\theta^*}(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}}$$

Haussler-Opper lower bound (proof)

$$
\begin{aligned}
& -\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P^n_{\theta^*}(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\tilde{\theta}) \frac{\mathrm{d}P_{\tilde{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})} \\
\geq \quad & -\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P^n_{\theta^*}(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}} \\
\geq \quad & -\int_\Theta \mathrm{d}\pi(\theta^*) \log \int_{\mathcal{X}^n} \mathrm{d}P^n_{\theta^*}(x_{1:n}) \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}}
\end{aligned}
$$

Haussler-Opper lower bound (proof)

$$
\begin{aligned}
& -\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P_{\theta^*}^n(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\tilde{\theta}) \frac{\mathrm{d}P_{\tilde{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})} \\
\geq & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P_{\theta^*}^n(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}} \\
\geq & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \log \int_{\mathcal{X}^n} \mathrm{d}P_{\theta^*}^n(x_{1:n}) \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}} \\
\geq & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \int_{\mathcal{X}^n} \sqrt{\mathrm{d}P_{\widehat{\theta}}(X_{1:n}) \mathrm{d}P_{\theta^*}(X_{1:n})}
\end{aligned}
$$

Haussler-Opper lower bound (proof)

$$
\begin{aligned}
& -\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P_{\theta^*}^n(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\tilde{\theta}) \frac{\mathrm{d}P_{\tilde{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})} \\
\geq & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P_{\theta^*}^n(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}} \\
\geq & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \log \int_{\mathcal{X}^n} \mathrm{d}P_{\theta^*}^n(x_{1:n}) \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}} \\
\geq & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \int_{\mathcal{X}^n} \sqrt{\mathrm{d}P_{\widehat{\theta}}(X_{1:n}) \mathrm{d}P_{\theta^*}(X_{1:n})} \\
= & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \alpha_H(P_{\widehat{\theta}}, P_{\theta^*})^n
\end{aligned}
$$

Haussler-Opper lower bound (proof)

$$
\begin{aligned}
& -\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P_{\theta^*}^n(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\tilde{\theta}) \frac{\mathrm{d}P_{\tilde{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})} \\
\geq & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \int_{\mathcal{X}^n} \mathrm{d}P_{\theta^*}^n(x_{1:n}) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}} \\
\geq & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \log \int_{\mathcal{X}^n} \mathrm{d}P_{\theta^*}^n(x_{1:n}) \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \sqrt{\frac{\mathrm{d}P_{\widehat{\theta}}(X_{1:n})}{\mathrm{d}P_{\theta^*}(X_{1:n})}} \\
\geq & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \int_{\mathcal{X}^n} \sqrt{\mathrm{d}P_{\widehat{\theta}}(X_{1:n}) \mathrm{d}P_{\theta^*}(X_{1:n})} \\
= & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \alpha_H(P_{\widehat{\theta}}, P_{\theta^*})^n \\
\geq & \quad -\int_\Theta \mathrm{d}\pi(\theta^*) \log \int_\Theta \mathrm{d}\pi(\widehat{\theta}) \exp\left(-n \frac{H^2(P_{\widehat{\theta}}, P_{\theta^*})}{2}\right)
\end{aligned}
$$

Metric entropy of envelope classes

Let $h(t) = H_{1/t}(\Lambda_1)$

Generic lower bound

Define r_n as the solution of $r_n = h\left(\sqrt{n/r_n}\right)$, then $R^+(\Lambda^n) \geq \log(e) r_n - 1$.

Questions

- The Haussler-Opper bound may or may not be tight.
- Connection between ϵ-entropy Λ_f and the tail behavior of the envelope distribution function $\overline{F}$

Flavors of adaptivity

For collections of small classes

Definition [Asymptotic adaptivity]

$(Q^n)_n$ is **asymptotically adaptive** with respect to $(\Lambda_m)_{m\in\mathcal{M}}$ if

$$\forall m \in \mathcal{M}, \quad R^+(Q^n, \Lambda_m^n) \quad = \quad \sup_{\mathbb{P}\in\Lambda_m} D(\mathbb{P}^n, Q^n) \leq (1+o(1))R^+(\Lambda_m^n)$$

For collections of massive envelop classes

Definition [Weak asymptotic adaptivity]

$(Q^n)_n$ is **asymptotically weakly adaptive** with respect to $(\Lambda_m)_{m\in\mathcal{M}}$

$$\forall m \in \mathcal{M}, \quad R^+(Q^n, \Lambda_m^n) \leq o(\log n)R^+(\Lambda_m^n)\,.$$

Light-tailed envelopes

The AC-code is adaptive with respect to source classes defined by envelopes with finite and non-decreasing hazard rate.

Theorem [BBG, 2012]

Q^n : the coding probability associated with the AC-code,
If f is an envelope with non-decreasing hazard rate,

$$R^+(Q^n; \Lambda_f^n) \leq (1 + o(1)) R^+(\Lambda_f^n)$$

while

$$R^+(\Lambda_f^n) = (1 + o(1))(\log e) \int_1^n \frac{U_c(x)}{2x} \mathrm{d}x$$

AC-code and sub-exponential envelope classes

Sub-exponential envelope class $\Lambda(\alpha, \beta, \gamma)$

with $\alpha \geq 1$, $\beta > 0$ and $\gamma > 1$, defined by envelope

$$f(k) = \gamma e^{-\left(\frac{k}{\beta}\right)^{\alpha}}.$$

Corollary

The AC-code is adaptive with respect to sub-exponential envelope classes.
For all $\alpha \geq 1$, $\beta > 0$ and $\gamma > 1$

$$R^+(Q^n; \Lambda^n(\alpha, \beta, \gamma)) \quad \leq \quad (1 + o(1)) R^+(\Lambda^n(\alpha, \beta, \gamma))$$

$$R^+(\Lambda^n(\alpha, \beta, \gamma)) = \frac{\alpha}{2(\alpha+1)} \beta \left(\ln(2)\right)^{1/\alpha} (\log n)^{1+1/\alpha} (1 + o(1)).$$

AC-code: sketch

Thresholding above last record

$m_i = \max_{1 \le j \le i} x_j$.
The j^{th} record is denoted by $\widetilde{m}_j$ ($\widetilde{m}_0 = 0$)
Let $\widetilde{\mathbf{m}} = (\widetilde{m}_i - \widetilde{m}_{i-1} + 1)1$.
Symbols from $\widetilde{\mathbf{m}}$ encoded using Elias penultimate code.

Progressive KT coding below the last record

$\widetilde{x}_i = x_i \mathbb{I}_{x_i \le m_{i-1}}$.
C_M : progressive $\mathbb{KT}$- encoding of $\widetilde{x}_{1:n}0$

$$Q_{i+1}(\widetilde{X}_{i+1} = j | X_{1:i} = x_{1:i}) = \frac{n_i^j + \frac{1}{2}}{i + \frac{m_i+1}{2}} \quad \text{if} \quad 1 \le j \le m_i,$$

$$Q_{i+1}(\widetilde{X}_{i+1} = 0 | X_{1:i} = x_{1:i}) = \frac{1/2}{i + \frac{m_i+1}{2}},$$

where n_i^j is the number of occurrences of symbol j in $x_{1:i}$, $n_i^0 = 0$.

Example

$x_{1:n}$

5 15 8 1 30 7 1 2 1 8 4 7 15 1 5 17 13 4 12 12

$m_{1:n}$

5 15 15 15 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30 30

$\tilde{x}_{1:n} \curvearrowright$ progressive KT encoding

0 0 8 1 0 7 1 2 1 8 4 7 15 1 5 17 13 4 12 12

$\tilde{m} \curvearrowright$ Elias encoding

6 11 16

Decomposing redundancy of AC-code

Decomposing pointwise redundancy

$$-\log Q^n(X_{1:n}) + \log \mathbb{P}^n(X_{1:n}) \quad = \quad \underbrace{\ell(C_E)}_{\text{i}} + \underbrace{\ell(C_M) + \log \mathbb{P}^n(X_{1:n})}_{\text{ii}} \, .$$

Establishing main theorem in [BBG, 2012]

$\hookrightarrow$

- (i) (Elias encoding of increments between records) is negligible with respect to $R^+(\Lambda_f^n)$, uniformly for $\mathbb{P} \in \Lambda_f$,
- The expected value of (ii) is upper bounded, uniformly for $\mathbb{P} \in \Lambda_f$, by a term which is equivalent to $R^+(\Lambda_f^n)$.

Decomposing redundancy of AC-code

Decomposing pointwise redundancy

$$-\log Q^n(X_{1:n}) + \log \mathbb{P}^n(X_{1:n}) \quad = \quad \underbrace{\ell(C_E)}_{\text{i}} + \underbrace{\ell(C_M) + \log \mathbb{P}^n(X_{1:n})}_{\text{ii}} .$$

Establishing main theorem in [BBG, 2012]

$\hookrightarrow$

- (i) (Elias encoding of increments between records) is negligible with respect to $R^+(\Lambda_f^n)$, uniformly for $\mathbb{P} \in \Lambda_f$,
- The expected value of (ii) is upper bounded, uniformly for $\mathbb{P} \in \Lambda_f$, by a term which is equivalent to $R^+(\Lambda_f^n)$.

Decomposing redundancy of AC-code

Decomposing pointwise redundancy

$$-\log Q^n(X_{1:n}) + \log \mathbb{P}^n(X_{1:n}) \quad = \quad \underbrace{\ell(C_E)}_{\text{i}} + \underbrace{\ell(C_M) + \log \mathbb{P}^n(X_{1:n})}_{\text{ii}} \, .$$

Establishing main theorem in [BBG, 2012]

$\hookrightarrow$

- (i) (Elias encoding of increments between records) is negligible with respect to $R^+(\Lambda_f^n)$, uniformly for $\mathbb{P} \in \Lambda_f$,
- The expected value of (ii) is upper bounded, uniformly for $\mathbb{P} \in \Lambda_f$, by a term which is equivalent to $R^+(\Lambda_f^n)$.

Bounding the length of Elias encoding

Proposition

Envelop f with non-decreasing hazard rate. Then, for all $\mathbb{P} \in \Lambda_f$,

$$\mathbb{E}\left[\ell(C_E)\right] \quad \leq \quad (2\log(e) + \rho)(U_c(\exp(H_n)) + 1)$$

where $\rho \leq 2$.

Bounding the length of Elias encoding

Proposition

Envelop f with non-decreasing hazard rate. Then, for all $\mathbb{P} \in \Lambda_f$,

$$\mathbb{E}\left[\ell(C_E)\right] \quad \leq \quad (2\log(e) + \rho)(U_c(\exp(H_n)) + 1)$$

where $\rho \leq 2$.

Argument:

$$\begin{aligned}
\ell(C_E) &\leq \sum_{i=1}^{n_n^0} (2\log\left(1 + \widetilde{m}_i - \widetilde{m}_{i-1}\right) + \rho) \\
&\leq \sum_{i=1}^{n_n^0} 2\log(e)\left(\widetilde{m}_i - \widetilde{m}_{i-1}\right) + \rho n_n^0 \\
&\leq (2\log(e) + \rho) M_n
\end{aligned}$$

Controlling (II)

Proposition (pointwise bound)

Let $i_0 = 1 \vee \lfloor M_n/4 \rfloor$, then

$$-\ln Q^n(\widetilde{X}_{1:n}) + \ln \mathbb{P}^n(X_{1:n}) \leq \underbrace{\frac{M_n(\ln(M_n)+10)}{2} + \frac{\ln n}{2}}_{\text{(a.i)}} + \underbrace{\sum_{i=i_0}^{n-1}\left(\frac{M_i}{2i+1}\right)}_{\text{(a.ii)}}$$

Controlling (II)

Proposition (pointwise bound)

Let $i_0 = 1 \vee \lfloor M_n/4 \rfloor$, then

$$-\ln Q^n(\widetilde{X}_{1:n}) + \ln \mathbb{P}^n(X_{1:n}) \leq \underbrace{\frac{M_n(\ln(M_n)+10)}{2} + \frac{\ln n}{2}}_{\text{(a.i)}} + \underbrace{\sum_{i=i_0}^{n-1}\left(\frac{M_i}{2i+1}\right)}_{\text{(a.ii)}}$$

$$\begin{aligned} &-\ln Q^n(\widetilde{X}_{1:n}) + \ln \mathbb{P}^n(X_{1:n}) \\ &= \underbrace{-\ln \mathbb{KT}_{M_n+1}(\widetilde{X}_{1:n}) + \ln \mathbb{P}^n(X_{1:n})}_{\text{(A)}} \underbrace{-\ln Q^n(\widetilde{X}_{1:n}) + \ln \mathbb{KT}_{M_n+1}(\widetilde{X}_{1:n})}_{\text{(B)}} \end{aligned}$$

Controlling (II)

Proposition (pointwise bound)

Let $i_0 = 1 \vee \lfloor M_n/4 \rfloor$, then

$$-\ln Q^n(\widetilde{X}_{1:n}) + \ln \mathbb{P}^n(X_{1:n}) \leq \underbrace{\frac{M_n(\ln(M_n)+10)}{2} + \frac{\ln n}{2}}_{\text{(a.i)}} + \underbrace{\sum_{i=i_0}^{n-1}\left(\frac{M_i}{2i+1}\right)}_{\text{(a.ii)}}$$

$$-\ln Q^n(\widetilde{X}_{1:n}) + \ln \mathbb{P}^n(X_{1:n})$$
$$= \underbrace{-\ln \mathbb{KT}_{M_n+1}(\widetilde{X}_{1:n}) + \ln \mathbb{P}^n(X_{1:n})}_{\text{(A)} \leq \frac{M_n+1}{2}\ln(n)+2\ln(2)} \underbrace{-\ln Q^n(\widetilde{X}_{1:n}) + \ln \mathbb{KT}_{M_n+1}(\widetilde{X}_{1:n})}_{\text{(B)}}$$

Controlling (II)

Proposition (pointwise bound)

Let $i_0 = 1 \vee \lfloor M_n/4 \rfloor$, then

$$-\ln Q^n(\widetilde{X}_{1:n}) + \ln \mathbb{P}^n(X_{1:n}) \leq \underbrace{\frac{M_n(\ln(M_n)+10)}{2} + \frac{\ln n}{2}}_{\text{(a.i)}} + \underbrace{\sum_{i=i_0}^{n-1}\left(\frac{M_i}{2i+1}\right)}_{\text{(a.ii)}}$$

$$\begin{aligned} &-\ln Q^n(\widetilde{X}_{1:n}) + \ln \mathbb{P}^n(X_{1:n}) \\ &= \underbrace{-\ln \mathbb{KT}_{M_n+1}(\widetilde{X}_{1:n}) + \ln \mathbb{P}^n(X_{1:n})}_{(\mathrm{A}) \leq \frac{M_n+1}{2}\ln(n) + 2\ln(2)} \underbrace{-\ln Q^n(\widetilde{X}_{1:n}) + \ln \mathbb{KT}_{M_n+1}(\widetilde{X}_{1:n})}_{(\mathrm{B}) = -\sum_{i=1}^{n-1}\ln\left(\frac{2i+1+M_n}{2i+1+M_i}\right) \leq 0} \end{aligned}$$

Controlling (II), continued

Proposition

$f : \mathbb{N}_+ \to [0,1]$: envelope with finite and non-decreasing hazard rate.

$$\mathbb{E}\left[\ell(C_M) + \log \mathbb{P}(X_{1:n})\right] \quad \leq \quad \log(e) \int_1^n \frac{U_c(x)}{2x} \mathrm{d}x \, (1 + o(1))$$

as $n \nearrow \infty$.

Nuts and bolts

Maximal inequalities

Let $Y_1, \ldots, Y_n \sim_{i.i.d.} F$ with density $f = F'$ on $[1, \infty)$ and $f/\overline{F} \nearrow$.
Let b be the infimum of the hazard rate.
Let $U(t) = \inf\{x : F(x) \geq 1 - 1/t\} = F^{\leftarrow}(1 - 1/t)$.
Let $Y_{(1)} \geq \ldots \geq Y_{(n)}$ be the order statistics.

$$\begin{aligned} \mathbb{E}[Y_{(1)}] &\leq U(\exp(H_n)) \\ \mathbb{E}[Y_{(1)} \ln(Y_{(1)})] &\leq (\mathbb{E}Y_{(1)}) \ln(\mathbb{E}Y_{(1)}) + 2/b^2 \,. \end{aligned}$$

where $H_n = \sum_{i=1}^{n} 1/i$.

Nuts and bolts

Maximal inequalities

Let $Y_1, \ldots, Y_n \sim_{i.i.d.} F$ with density $f = F'$ on $[1, \infty)$ and $f/\overline{F} \nearrow$.
Let b be the infimum of the hazard rate.
Let $U(t) = \inf\{x : F(x) \geq 1 - 1/t\} = F^{\leftarrow}(1 - 1/t)$.
Let $Y_{(1)} \geq \ldots \geq Y_{(n)}$ be the order statistics.

$$\begin{aligned} \mathbb{E}[Y_{(1)}] &\leq U(\exp(H_n)) \\ \mathbb{E}[Y_{(1)} \ln(Y_{(1)})] &\leq (\mathbb{E}Y_{(1)}) \ln(\mathbb{E}Y_{(1)}) + 2/b^2 . \end{aligned}$$

where $H_n = \sum_{i=1}^{n} 1/i$.

Corollary

Let $X_1, \ldots, X_n \sim_{i.i.d.} P \in \Lambda_f^1$, let $M_n = \max(X_1, \ldots, X_n)$, then,

$$\begin{aligned} \mathbb{E}M_n &\leq U_c(en) + 1 \\ \mathbb{E}[M_n \log M_n] &\leq [U_c(en) + 1] \log[U_c(en) + 1] + 2/b^2. \end{aligned}$$

Nuts and bolts

Maximal inequalities

Let $Y_1, \ldots, Y_n \sim_{i.i.d.} F$ with density $f = F'$ on $[1, \infty)$ and $f/\overline{F} \nearrow$.
Let b be the infimum of the hazard rate.
Let $U(t) = \inf\{x : F(x) \geq 1 - 1/t\} = F^{\leftarrow}(1 - 1/t)$.
Let $Y_{(1)} \geq \ldots \geq Y_{(n)}$ be the order statistics.

$$\begin{aligned} \mathbb{E}[Y_{(1)}] &\leq U(\exp(H_n)) \\ \mathbb{E}[Y_{(1)} \ln(Y_{(1)})] &\leq (\mathbb{E}Y_{(1)}) \ln(\mathbb{E}Y_{(1)}) + 2/b^2 \,. \end{aligned}$$

where $H_n = \sum_{i=1}^{n} 1/i$.

Ingredients of proof

- Rényi's representation of order statistics & concavity of $U \circ \exp$
- Sub-additivity of relative entropy (see Ledoux, 2001, Massart, 2006)
- The entropy method provides us with sharp tail and moment bounds for order statistics (B. & Thomas, 2012)

Lower bounds : back to metric entropy

Corollary of Haussler & Opper, AoS, 1997

Assume there exists a (very) slowly varying function h such that:

$$\mathcal{H}_\epsilon(\Lambda) = h\left(\frac{1}{\epsilon}\right)(1+o(1)) \quad \text{as } \epsilon \searrow 0.$$

Then

$$R^+(\Lambda^n) = (\log e)h(\sqrt{n})(1+o(1)) \quad \text{as } n \nearrow +\infty.$$

Entropy of envelope classes with finite and non-decreasing hazard rate.

$$\mathcal{H}_\epsilon(\Lambda_f) = (1+o(1))\int_0^{1/\epsilon^2} \frac{U_c(x)}{2x}\mathrm{d}x \quad \text{as } \epsilon \searrow 0 \ .$$

Envelops with heavier tails

If the tail envelope distribution is heavier than exponential, thresholding at maximum does not lead to (weakly) adaptive coding.

Ideal threshold

$\lfloor m_c(n) \rfloor$ where $m_c(t)$ solution of

$$t\overline{F}_c(u) = \frac{u}{2}\log t$$

or $u = U_c\left(\frac{2t}{u \log t}\right)$

Proxy threshold

$m_c(t)$ solution of $t\overline{F}_c(u) = u$ or $u = U_c\left(\frac{t}{u}\right)$

Properties

- Ideal and proxy thresholds are implicitly defined functions.
- m_c is non-decreasing.
- $m_c(t) \nearrow \infty$
- $m_c(t)/t \searrow 0$

Envelopes belonging to max-domains of attraction

Von Mises assumption

$U = (1/\overline{F})^{\leftarrow}$ is twice differentiable and

$$\frac{tU''(t)}{U'(t)} \to_{t\nearrow\infty} \gamma - 1\,.$$

for $\gamma \in \mathbb{R}$.
For $\gamma > 0$ entails $\lim_{t\nearrow\infty} \frac{tF'(t)}{\overline{F}(t)} = \frac{1}{\gamma}$

If U_c satisfies Von Mises condition, F_c belongs to a max-domain of attraction (Gumbel if $\gamma = 0$, Frechet if $\gamma > 0$).

If F_c has non-decreasing hazard rate, U_c satisfies Von-Mises condition with $\gamma = 0$.

Envelopes belonging to max-domains of attraction

Von Mises assumption

$U = (1/\overline{F})^{\leftarrow}$ is twice differentiable and

$$\frac{tU''(t)}{U'(t)} \to_{t\nearrow\infty} \gamma - 1\,.$$

for $\gamma \in \mathbb{R}$.
For $\gamma > 0$ entails $\lim_{t\nearrow\infty} \frac{tF'(t)}{\overline{F}(t)} = \frac{1}{\gamma}$

If U_c satisfies Von Mises condition, F_c belongs to a max-domain of attraction (Gumbel if $\gamma = 0$, Frechet if $\gamma > 0$).

If F_c has non-decreasing hazard rate, U_c satisfies Von-Mises condition with $\gamma = 0$.

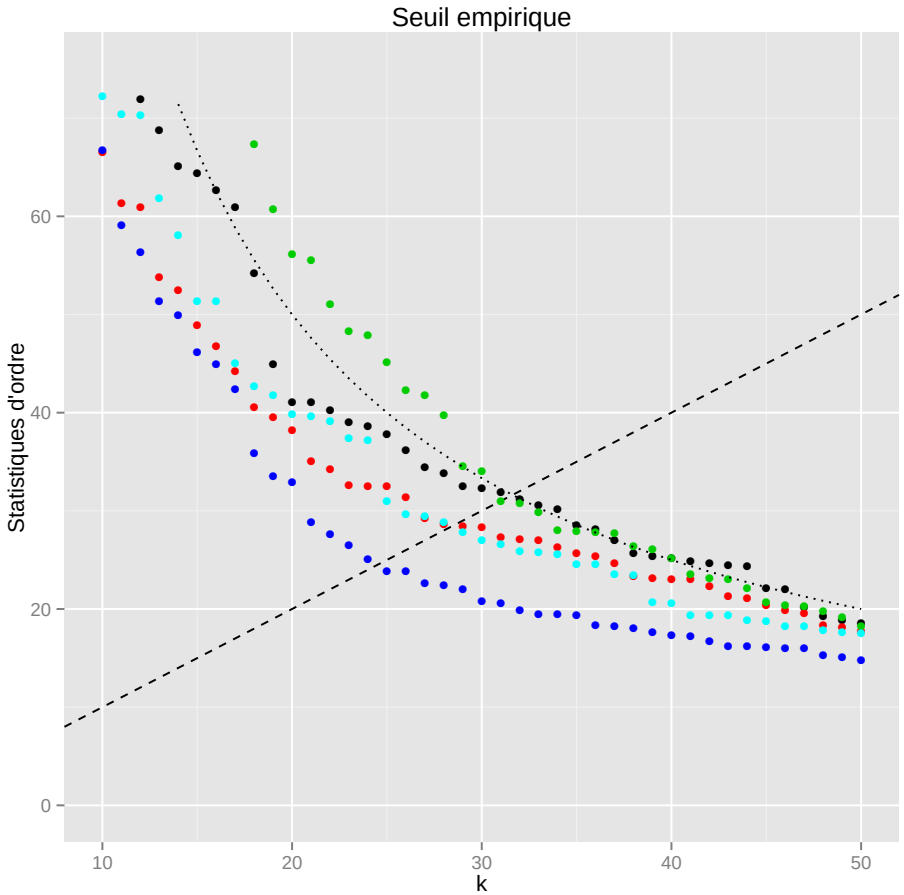
Proposition : m_c is regularly varying

$$\lim_{t\nearrow\infty} \frac{m_c(tx)}{m_c(t)} = x^{\frac{\gamma}{\gamma+1}}$$

Proposition

Conditional excess distributions converge towards generalized Pareto distribution.

Adaptive thresholding



$$M_n = \min\left(n, \{k \ : \ X_{k,n} \leq k\}\right)$$

$F_c \in \mathrm{MDA}(\gamma), \gamma > 0$

- $\frac{M_n}{m_n} \xrightarrow{P} 1\,.$
- $\frac{X_{M_n,n}}{m_c(n)} \xrightarrow{P} 1\,.$

M_n is self-bounded

$$\mathbb{P}\{|M_n - \mathbb{E}M_n| \geq t\} \leq 2e^{\left(-\frac{t^2}{2(\mathbb{E}M_n + t)}\right)}\,.$$

WAC-code

$$\widetilde{x}_i = \begin{cases} x_i & \text{if } i \geq 2 \text{ and } x_i \leq x_{M_{i-1},i-1} \\ 0 & \text{if } i \geq 2 \text{ and } x_i > x_{M_{i-1},i-1} \\ x_1 & \text{if } i = 1 \end{cases}$$

C_M : arithmetic encoding of $\widetilde{x}_{1:n}0$ using

$Q^{n+1}(\widetilde{x}_{1:n}0) = Q_{n+1}(0 \mid x_{1:n}) \prod_{i=0}^{n-1} Q_{i+1}(\widetilde{x}_{i+1} \mid x_{1:i})$.

with $Q_{i+1}\left(\widetilde{X}_{i+1} = j \mid X_{1:i} = x_{1:i}\right) = \frac{n_i^j + \frac{1}{2}}{i + \frac{x_{M_i,i}+1}{2}}$

The sequence of Elias codewords forms C_E

$\widetilde{\mathbf{m}}$: the sequence $x_{i+1} - x_{M_i,i} + 1$, for i s.t. $x_{i+1} > x_{M_i,i}$

Interleaving...

KT encoding of subsequences of non-censored symbols terminated by 0, with Elias encoding of the censored symbols.

Upper bounds on redundancy of WAC-code

$$
\begin{aligned}
&\ell(C_M) + \log_2 \mathbb{P}^n(X_{1:n}) \\
&\quad = \underbrace{-\log_2 \mathbb{KT}_{X_{M_n,n+1}}(\widetilde{X}_{1:n}) + \log_2 \mathbb{P}^n(X_{1:n})}_{\leq \frac{X_{M_n,n+1}}{2}\log_2(n)+2} \\
&\qquad \underbrace{-\log_2 Q^n(\widetilde{X}_{1:n}) + \log_2 \mathbb{KT}_{X_{M_n,n+1}}(\widetilde{X}_{1:n})}_{\leq 0}
\end{aligned}
$$

$$
\ell(C_E) \leq 2 \sum_{i=1}^{n-1} \mathbb{I}_{X_{i+1} > X_{M_i,i}} \left\{ \log_2 \left(1 + X_{i+1} - X_{M_i,i}\right) + \rho \right\}.
$$

Bounding redundancy of mixture encoding

If $\overline{F}_c \in MDA(-1/\gamma)$ with $\gamma > 0$,

$\forall \epsilon > 0$, for sufficiently large n,

$$\mathbb{E} X_{M_n,n} \leq m_n(1+\epsilon)\,.$$

$$\mathbb{E}\left[\ell(C_M) + \log_2 \mathbb{P}^n(X_{1:n})\right] \leq (1+\epsilon)\frac{m_n}{2}\log n + 2$$

Bounding length of Elias encoding

If $\overline{F}_c \in MDA(-1/\gamma)$ with $\gamma > 0$,

$\forall \epsilon > 0$, for sufficiently large n,

$$\mathbb{E}\ell(C_E) \leq 2(1+\epsilon)\sum_{i=1}^{n} \frac{m_i \log m_i}{i} \leq 2(1+\epsilon)\int_1^n \frac{m_t \log m_t}{t}\mathrm{d}t \leq \frac{2\gamma}{\gamma+1} m_n \log m_n$$

Cooking lower bounds

For an ad hoc prior

$$I(\theta; X_{1:n}) \geq \mathbb{E} Z_n$$

where Z_n is the number of distinct symbols in $X_{1:n}$

$$\mathbb{E} Z_n \geq m_n$$

References

- S. Boucheron and E. Gassiat : A Bernstein-von Mises theorem for discrete probability distributions Electronic Journal of Statistics. **3** (2009) 114-148.
- S. Boucheron and A. Garivier and E. Gassiat : Coding over Infinite Alphabets IEEE Trans. on Information Theory **55** (2009) 358 - 373.
- D. Bontemps : Universal coding on infinite alphabets: exponentially decreasing envelopes. IEEE Trans. Inform. Theory 57 (2011), no. 3, 1466--1478.
- D. Bontemps, S. Boucheron and E. Gassiat : Adaptive compression against a countable alphabet. 23rd Intern. Meeting on Probabilistic, Combinatorial, and Asymptotic Methods for the Analysis of Algorithms (AofA'12), 201--218, Discrete Math. Theor. Comput. Sci. Proc.
- S. Boucheron, E. Gassiat & M. Ohanessian : Weakly adaptive compression against a countable alphabet. 2013
- S. Boucheron, M. Thomas : Concentration inequalities for order statistics. Electronic Communications in Probability. 17 (2012). 1-12