Electrical Engineering 229A Lecture 12 Notes

Daniel Raban

October 5, 2021

1 Lempel-Ziv Coding for Ergodic Processes

1.1 Intuition behind Lempel-Ziv coding

Last time, we discussed a comma-free binary representation of natural numbers using $\log n + 2 \log \log n + k$ bits ($k = 5$). To send n, send $\lceil \log n \rceil$ bits (tells us t$n \in 1, \dots, 2^{\lceil \log n \rceil}$). To send $\lceil \log n \rceil$, send $\lceil \log \lceil \log n \rceil \rceil$ bits (same idea). Send $\lceil \log \lceil \log n \rceil \rceil$ as $\lceil \log \lceil \log n \rceil \rceil$ 0s, followed by a 1.

Example 1.1. To send $n = 17$, we have $\lceil \log n \rceil = 5$ and $\lceil \log 5 \rceil = 3$. Then transmit

$$0001 \quad \underbrace{101}_{\text{represents } 5} \quad 10001.$$

Example 1.2. To send $n = 14$, we have $\lceil \log n \rceil = 4$ and $\lceil \log 5 \rceil = 2$. Then transmit

$$00011001110,$$

which can be parsed as

$$0001 \quad 100 \quad 1110.$$

To motivate the LZ'77 scheme (which compresses to the entropy rate for any stationary ergodic process), let's consider i.i.d.

$$\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots$$

at the level of blocks of size L. The situation is that $\dots, X_{-3}, X_{-2}, X_{-1}$ is common knowledge to the compressor and decompressor (or the transmitter and receiver). We need to send $(X_0, X_1, \dots, X_{L-1})$. We do this by finding

$$\inf\{m \geq 1 : (X_0, X_1, \dots, X_{L-1}) = (X_{-mL} X_{-mL+1}, \dots, X_{-mL+L-1}\}$$

and sending m using the comma-free encoding of $\mathbb{N}$. Since the blocks of length L of the type $(X_{-jL}, X_{-jL+1}, \dots, X_{-jL+L-1})$ are independent, m will be geometrically distributed, conditioned on $(X_0, X_1, \dots, X_{L-1})$. Then

$$\mathbb{P}(m = j \mid (X_0, \dots, X_{L-1}) = x_0^{L-1}) = p(x_0^{L-1})(1 - p(x_0)^{L-1})^{j-1}, \qquad j = 1, 2, \dots.$$

So the conditional expectation on this event is

$$\mathbb{E}[m \mid (X_0, \dots, X_{L-1}) = x_0^{L-1}] = \frac{1}{p(x_0^{L-1})}$$

Also, for all $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}(m > \widetilde{K}\frac{1}{p(x_0^{L-1})} \mid X_0^{L-1} = x_0^{L-1}) &= \sum_{j=\lceil \widetilde{K}\frac{1}{p(x_0^{L-1})}\rceil} p(x_0^{L-1})(1 - p(x_0^{L-1}))^{j-1} \\ &\leq (1 - p(x_0^{L-1}))^{\lceil \widetilde{K}(1/p(x_0^{L-1}))\rceil - 1} \\ &\lesssim e^{-\widetilde{K}}\end{aligned}$$

as $L \to \infty$.

The upshot is that we can, with probability close to 1, convey m with $\log \frac{\widetilde{K}}{p(x_0^{L-1})} + \log\log \frac{\widetilde{K}}{p(x_0^{L-1})} + k$ bits (conditioned on $X_0^{L-1} = x_0^{L-1}$) for any $\widetilde{K}$, as $L \to \infty$. Note that

$$\sum_{x_0^{L-1}} p(x_0^{k-1}) \left(\log \frac{\widetilde{K}}{p(x_0^{L-1})} + \log\log \frac{\widetilde{K}}{p(x_0^{k-1})} + k \right) \asymp H(X_0, \dots, X_{L-1})$$

as $L \to \infty$.

1.2 Ergodicity and Kac's theorem

Definition 1.1. A two-sided process $(X_n, n \in \mathbb{Z})$ with $X_n \in \mathscr{X}$ for finite $\mathscr{X}$ is **ergodic** if

1. The process is stationary.
2. Every shift-invariant event should have probability 0 or probability 1.

By **shift-invariant**, we mean

$$\{(\dots, X_{-1}, X_0, X_1, \dots) \in A\} = \{(\dots, X_{-2}, X_{-1}, X_0, \dots) \in A\}.$$

Shift-invariant events can be very interesting.

Example 1.3. The event {there are infinitely many 1s in the sequence} is shift-invariant.

Example 1.4. The event $\{$the $\limsup$ of the sequence is 1$\}$ is shift-invariant.

Theorem 1.1 (Pointwise ergodic theorem, Birkhoff)**.** *If $(X_n, n \in \mathbb{Z})$ is ergodic and $\phi : \mathscr{X}^k \to \mathbb{R}$, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \phi(X_t, X_{t+1}, \dots, X_{t+k-1}) = \mathbb{E}[\phi(X_0, \dots, X_{k-1})]$$

almost surely.

To look back in the past in the general ergodic case, we use the following theorem:

Theorem 1.2 (Kac)**.** *Let $(X_n, n \in \mathbb{Z})$ be an ergodic process with $X_n \in \mathscr{X}$ for all n, where $\mathscr{X}$ is finite. Let*

$$Q_b(i) = \mathbb{P}(X_{-i} = b, X_j \neq b \text{ for } -i+1 \leq j \leq -1 \mid X_0 = b).$$

Then

$$\sum_{i=1}^{\infty} iQ_b(i) = \frac{1}{\mathbb{P}(X_0 = b)}.$$

Proof. Fix $b \in \mathscr{X}$. Define the events

$$A_{j,k} := \{X_{-j} = b, X_{-j+1} \neq b, \dots, X_{k-1} \neq b, X_k = b\}, \qquad k \geq 0, j \geq 1.$$

These events are disjoint. We claim that

$$\mathbb{P}\left(\bigcup_{j,k} A_{j,k}\right) = 1$$

if $\mathbb{P}(X_0 = b) > 0$. This is because b occurs some finite time in the future and some time in the past; we can see this from, for example, looking at the sample averages of the ergodic theorem with ϕ as the indicator of $\{b\}$.

Hence,

$$\sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \mathbb{P}(A_{j,k}) = 1.$$

But this equals

$$\sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \mathbb{P}(X_k = b) Q_b(j+k) = \mathbb{P}(X_0 = b) \sum_{i=0}^{\infty} iQ_b(i)$$

because $\mathbb{P}(X_k = b) = \mathbb{P}(X_0 = b)$ by stationarity and because the number of ways to get $j + k = i$ is i. $\square$

Now for LZ'77, assume that $(X_n, n \in \mathbb{Z})$ is an ergodic process. For any fixed $L \geq 1$, define

$$R_L(X_0, X_1, \ldots, X_{L-1}) := \min\{j \geq 1 : (X_{-j}, X_{-j+1}, \ldots, X_{-j+L-1}) = (X_0, \ldots, X_{L-1})\}.$$

By Kac's theorem,

$$\mathbb{E}[R_L(X_0, X_1, \ldots, X_{L-1}) \mid X_0^{L-1} = x_0^{L-1}] = \frac{1}{p(x_0^{L-1})}.$$

The transmitter will send $R_L(X_0, X_1, \ldots, X_{L-1})$ by comma-free encoding (in order to convey X_0). Let

$$\lambda_L(x_0^{L-1}) = \log R_L(X_0^{L-1}) + \log\log R_L(X_0^{L-1}) + 5.$$

Next time, we will show that

$$\frac{1}{L}\,\mathbb{E}[\lambda_L(X_0^{L-1})] \xrightarrow{L\to\infty} H,$$

the entropy rate of the process.