Range-Renewal Speed and Entropy for I.I.D Models

Xin-Xing Chen[1], Jian-Sheng Xie[2,3] and Jiangang Ying[2]

1. Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200240, China

2. School of Mathematical Sciences, Fudan University, Shanghai 200433, China

3. Corresponding author. E-mail: jiansheng.xie@gmail.com

Abstract

In this note the relation between the range-renewal speed and entropy for i.i.d. models is discussed.

In [2] the authors build an SLLN

$$\lim_{n\to\infty} \frac{R_n}{\mathbb{E}R_n} = 1 \text{ almost surely} \tag{1}$$

with

$$\mathbb{E}R_n = \sum_x [1 - (1 - \pi_x)^n] \tag{2}$$

for n samples of a discrete distribution π, where R_n denotes the number of distinct values of the n samples. In this note we would like to study further the relation between entropy of the distribution and the range-renewal speed $\mathbb{E}R_n$, where the entropy of a (discrete) distribution π is defined as

$$S(\pi) := \sum_x -\pi_x \cdot \log \pi_x. \tag{3}$$

As is already well known, for our i.i.d. model, we always have

$$\lim_{n\to\infty} \frac{R_n}{n} = 0 \text{ almost surely.}$$

But an information of the entropy $S(\pi)$ being finite or infinite would pose a constriction on the range-renewal speed as the following:

Theorem 1 *For our i.i.d. range-renewal model, in general we have*

$$\lim_{n\to\infty}\frac{R_n}{n}=0 \tag{4}$$

almost surely. If the entropy $S(\pi)<\infty$, then

$$\lim_{n\to\infty}\frac{\log n}{n}\cdot R_n=0 \tag{5}$$

almost surely; Conversely, if the entropy $S(\pi)=\infty$, then almost surely

$$\overline{\lim_{n\to\infty}}\frac{(\log n)^{1+\varepsilon}}{n}\cdot R_n=\infty,\quad \forall\varepsilon>0. \tag{6}$$

Proof. We will always assume, for simplicity, that π is supported on $\mathbb{N}$ with

$$\pi_1\geq\pi_2\geq\cdots$$

and we would denote

$$\varphi^{-1}(n):=\#\{x:\pi_x>\frac{1}{n}\}$$

for each $n\geq 1$.

Eq. (4) can be proved easily via eq. (1) and (2) (to prove $\mathbb{E}R_n/n\to 0$); it can also be regarded as a consequence of the main result of [3] [4] (see also [1]). Hence the proof is omitted here.

For (5), first notice that

$$\begin{aligned}\frac{\log n}{n}\cdot\mathbb{E}R_n &= \sum_x[1-(1-\pi_x)^n]\cdot\frac{\log n}{n}\\ &= \Big\{\sum_{1/\pi_x\leq n}+\sum_{1/\pi_x>n}\Big\}[1-(1-\pi_x)^n]\cdot\frac{\log n}{n}=:I_1+I_2.\end{aligned}$$

For the second part of the above equation, we have

$$\begin{aligned}I_2 &= \sum_{1/\pi_x>n}[1-(1-\pi_x)^n]\cdot\frac{\log n}{n}\\ &\leq \sum_{1/\pi_x>n}\pi_x\cdot\log n\leq\sum_{1/\pi_x>n}-\pi_x\cdot\log\pi_x\to 0.\end{aligned}$$

For the first part, choose a large number $N \in \mathbb{N}$ (but still $N < n$). Noting $\phi(t) := -t \cdot \log t$ is increasing on $(0, e^{-1})$ (especially on $[1/n, 1/N]$ for all $n > N$), we have

$$\begin{aligned} I_1 &= \sum_{1/\pi_x \le n} [1-(1-\pi_x)^n] \cdot \frac{\log n}{n} \le \varphi^{-1}(N) \cdot \frac{\log n}{n} + \sum_{N<1/\pi_x \le n} \phi(\frac{1}{n}) \\ &\le \varphi^{-1}(N) \cdot \frac{\log n}{n} + \sum_{N<1/\pi_x \le n} \phi(\pi_x) \\ &\le \varphi^{-1}(N) \cdot \frac{\log n}{n} + \sum_{x>\varphi^{-1}(N)} -\pi_x \cdot \log \pi_x. \end{aligned}$$

First letting $n \to \infty$ then $N \to \infty$, we get the desired result.

For (6), it's equivalent to $\overline{\lim\limits_{n\to\infty}} \frac{(\log n)^{1+\varepsilon}}{n} \cdot \mathbb{E}R_n = \infty$. Suppose on the contrary that there exists some $\varepsilon > 0$ such that $\overline{\lim\limits_{n\to\infty}} \frac{(\log n)^{1+\varepsilon}}{n} \cdot \mathbb{E}R_n < \infty$. This clearly implies $\overline{\lim\limits_{n\to\infty}} (\log n)^{1+\varepsilon} \cdot \sum\limits_{\pi_x<1/n} \pi_x < \infty$ since $\mathbb{E}R_n = \sum_x [1-(1-\pi_x)^n]$. We write

$$a_k := \#\{x : \frac{1}{k+1} < \pi_x \le \frac{1}{k}\}, \quad k \ge 1.$$

Then the above implies $B_n := (\log n)^{1+\varepsilon} \cdot \sum\limits_{k=n}^{\infty} \frac{a_k}{k} \le C$ for some $C > 0$ and all $n \ge 1$. Hence

$$\frac{a_n}{n} = \frac{B_n}{(\log n)^{1+\varepsilon}} - \frac{B_{n+1}}{(\log(n+1))^{1+\varepsilon}}.$$

From this we shall derive the following result

$$\sum_k \frac{a_k}{k} \cdot \log k < \infty, \tag{7}$$

which implies $S(\pi) < \infty$, a contradiction. In fact,

$$\begin{aligned} \frac{a_n}{n} \cdot \log n &= \frac{B_n}{(\log n)^{\varepsilon}} - \frac{B_{n+1} \cdot \log n}{(\log(n+1))^{1+\varepsilon}} \\ &= [\frac{B_n}{(\log n)^{\varepsilon}} - \frac{B_{n+1}}{(\log(n+1))^{\varepsilon}}] + \frac{B_{n+1} \cdot \log(1+1/n)}{(\log(n+1))^{1+\varepsilon}} \\ &= [\frac{B_n}{(\log n)^{\varepsilon}} - \frac{B_{n+1}}{(\log(n+1))^{\varepsilon}}] + O(\frac{1}{n \cdot (\log n)^{1+\varepsilon}}), \end{aligned}$$

which surely implies (7). □

Remark 1 *(1) Let*

$$\pi_x := \frac{C}{x[\log(x+1)]^{\beta+1}}, \quad x = 1, 2, \cdots$$

with $\beta > 0$ and C being a normalizing constant. By the results in [2] we know

$$\mathbb{E}R_n = O(1) \cdot \frac{n}{(\log n)^\beta}$$

as $n \to +\infty$. When $0 < \beta \le 1$, we always have $S(\pi) = +\infty$, but

$$\lim_{n\to+\infty} \frac{\log n}{n} \cdot \mathbb{E}R_n = \begin{cases} c \in (0, +\infty), & \text{if } \beta = 1 \\ +\infty, & \text{if } 0 < \beta < 1 \end{cases}$$

with c being some positive constant. Therefore the result in (6) cannot be strengthened into the one with $\varepsilon = 0$;

(2) The $\overline{\lim}$ in (6) cannot be replaced by $\underline{\lim}$. There exists distributions π such that

$$S(\pi) = +\infty \text{ with } \underline{\lim}_{n\to\infty} \frac{(\log n)^{1+\varepsilon}}{n} \cdot R_n < \infty, \quad \forall 0 < \varepsilon < 1. \tag{8}$$

For the part (2) of the above remark, for example, let for any $k \ge 1$,

$$b_k := 2^{2^k},\ S_0 := 0,\ S_k := \sum_{\ell=1}^{k} \frac{2^{b_\ell}}{b_\ell}.$$

And for any $S_{k-1} < x \le S_k$, we set $\pi_x := A \cdot 2^{-b_k}$, where A is the normalizing constant. Obviously $2 < A < 4$. It is easily to see that $S(\pi) = \infty$ since

$$\begin{aligned} S(\pi) &= \sum_{k=1}^{\infty} \sum_{x=S_{k-1}+1}^{S_k} \pi_x \log(\pi_x^{-1}) = \sum_{k=1}^{\infty} \frac{2^{b_k}}{b_k} \cdot (A \cdot 2^{-b_k}) \log(\frac{2^{b_k}}{A}) \\ &= \sum_{k=1}^{\infty} \left[A \cdot \log 2 - \frac{A}{b_k} \log A \right] = \infty. \end{aligned}$$

The proof of (8) is as the following. For each $k \ge 1$, let $n_k = 2^{2b_k}$. Then $A \cdot 2^{-b_{k+1}} < \frac{1}{n_k} \le A \cdot 2^{-b_k}$ for sufficiently large k and $\#\{x : \pi_x \ge \frac{1}{n_k}\} = S_k$. And

$$\begin{aligned} \mathbb{E}R_{n_k} &= \sum_{\pi_x \ge n_k{}^{-1}} [1 - (1-\pi_x)^{n_k}] + \sum_{\pi_x < n_k{}^{-1}} [1 - (1-\pi_x)^{n_k}] \\ &\le \sum_{\pi_x \ge n_k{}^{-1}} 1 + n_k \cdot \sum_{\pi_x < n_k{}^{-1}} \pi_x = S_k + n_k \cdot \sum_{\pi_x \le A \cdot 2^{-b_{k+1}}} \pi_x \\ &= S_k + n_k \cdot \sum_{\ell=k+1}^{\infty} \frac{2^{b_\ell}}{b_\ell} \cdot (A \cdot 2^{-b_\ell}) = S_k + n_k \cdot \sum_{\ell=k+1}^{\infty} \frac{1}{b_\ell} \\ &\le \frac{2^{b_k+3}}{b_k} + \frac{2^{2b_k+1}}{b_{k+1}} = \frac{2^{b_k+3}}{b_k} + \frac{2^{2b_k+1}}{b_k^2}. \end{aligned}$$

Fix $0 < \varepsilon < 1$. Furthermore,

$$\begin{aligned} \frac{(\log_2 n_k)^{1+\varepsilon}}{n_k} \cdot \mathbb{E}R_{n_k} \leq & \quad \frac{(2b_k)^{1+\varepsilon}}{2^{2b_k}} \cdot \left(\frac{2^{b_k+3}}{b_k} + \frac{2^{2b_k+1}}{b_k^2} \right) \\ = & \quad \frac{2^{4+\varepsilon} b_k^{\varepsilon}}{2^{b_k}} + 2^{2+\varepsilon} b_k^{-1+\varepsilon} \to 0 \end{aligned}$$

as k tends to infinity. As a result, (8) holds.

Acknowledgements The second author would like to thank Prof. De-Jun Feng for helpful discussions. This work is in part supported by NSFC (No. 11001173, No. 11271255 and No. 11271077) and the Laboratory of Mathematics for Nonlinear Science, Fudan University.

References

[1] Athreya, K. B.; *On the Range of Recurrent Markov Chains*, Statist. Probab. Lett. **3** (1985), no. 3, pp. 143–145. MR0801860

[2] Chen, X.-X.; Xie, J.-S.; Ying, J.-G.: *Range-Renewal Processes: SLLN, Power Law and Beyonds*, arXiv:1305.1829.

[3] Chosid, Leo; Isaac, Richard; *On the Range of Recurrent Markov Chains*, Ann. Probab. **6** (1978), no. 4, pp. 680–687. MR0474507

[4] Chosid, Leo; Isaac, Richard; *Correction to: "On the range of recurrent Markov chains" [Ann. Probab.* ***6*** *(1978), no. 4, 680–687; MR 57 #14146].* Ann. Probab. **8** (1980), no. 5, pp. 1000. MR0600347