Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Information Processing Letters 89 (2004) 209–213
Information Processing Letters
www.elsevier.com/locate/ipl

On some properties of the NML estimator for Bernoulli strings ☆

Ciprian Doru Giurcăneanu

Institute of Signal Processing, Tampere University of Technology P.O. Box 553, FIN-33101 Tampere, Finland

Received 15 October 2001; received in revised form 15 October 2003

Communicated by P.M.B. Vitányi

Abstract

This paper proves the monotonicity of the sequence $C_n/\sqrt{n}$, where C_n denotes the normalization coefficient in the universal Normalized Maximum Likelihood (NML) model for the Bernoulli class. The main result is used to find a non-asymptotic estimation of $\log C_n$.

Keywords: Algorithms; Combinatorial problems; Monotonicity; Bounds; Non-asymptotic estimation

1. Introduction

The universal NML model [1–3] for the Bernoulli class $\{P(x^n;\theta) = \theta^{n_0}(1-\theta)^{n-n_0}\colon \theta \in (0,1)\}$ is

$$\widehat{P}(x^n) \triangleq \frac{\left(\frac{n_0}{n}\right)^{n_0}\left(\frac{n-n_0}{n}\right)^{n-n_0}}{2+\sum_{m=1}^{n-1}\binom{n}{m}\left(\frac{m}{n}\right)^m\left(\frac{n-m}{n}\right)^{n-m}}, \tag{1}$$

where the entries of the observed sequence $x^n = (x_1, x_2, \ldots, x_n)$ are independently generated as outcomes of a Bernoulli process with parameter $\theta = \Pr(0)$, and n_0 denotes the number of 0's in the string x^n. It solves Shtarkov's minimax problem [3]

$$\min_q \max_{x^n} \log \frac{P(x^n; (\hat{\theta}(x^n))}{q(x^n)},$$

where q ranges over the set of all nonsingular distributions and $\hat{\theta}(x^n)$ is the maximum likelihood estimate of parameter θ. It is proven in [4] that the NML distribution solves also the problem

$$\inf_q \sup_g E_g \log \frac{P(x^n; \hat{\theta}(x^n))}{q(x^n)},$$

☆ This work has been supported by Academy of Finland, project No. 44876 (Finnish Centre of Excellence Program (2000–2005)).
E-mail address: cipriand@cs.tut.fi (C.D. Giurcăneanu).

0020-0190/$ – see front matter
doi:10.1016/j.ipl.2003.10.011

where q and g range over all nonsingular distributions. The importance of the normalization coefficient, defined by

$$C_n \triangleq 2 + \sum_{m=1}^{n-1} \binom{n}{m} \left(\frac{m}{n}\right)^m \left(\frac{n-m}{n}\right)^{n-m}, \tag{2}$$

was emphasized in [4] by showing that $\log C_n$ is the amount of information in the data which can be learned with the agreed model class. The problem studied in this paper is the monotonicity of the sequence given by $\frac{C_n}{\sqrt{n}}$, and the main result is used to find a non-asymptotic estimation of the normalization coefficient C_n. An asymptotic approximation is deduced in [5] $\log C_n = \frac{1}{2} \log \frac{n^3}{n_0(n-n_0)} - \log \sqrt{2\pi} + \mathrm{O}(1/n)$, and another one $\log C_n = \log \sqrt{\frac{\pi n}{2}} + \mathrm{o}(1)$ in [2].

2. Main results

The main result of this note is to show that $\frac{C_n}{\sqrt{n}}$, $n \geqslant 1$, is a monotone decreasing sequence.

Lemma 1. *The sequence* $a_n \triangleq \frac{n! \mathrm{e}^n}{n^n \sqrt{n}}$, $n \geqslant 1$, *is monotone decreasing.*

This property of the sequence $(a_n)_{n \geqslant 1}$ is sometimes included in the demonstration for Stirling's approximation. The complete proof can be found in [6].

Lemma 2. *The sequence* $b_n \triangleq \frac{1}{\mathrm{e}^n} \sum_{k=0}^{n} \frac{n^k}{k!}$, $n \geqslant 1$, *is monotone decreasing.*

The proof is deferred to Appendix A.

Theorem 3. *The sequence* $\frac{C_n}{\sqrt{n}}$, $n \geqslant 1$, *is monotone decreasing.*

Proof. We use the identity $C_n = \frac{n!}{n^n} \sum_{k=0}^{n} \frac{n^k}{k!}$, which was proven in [7]. It follows that

$$\frac{C_n}{\sqrt{n}} = \frac{n!}{n^n \sqrt{n}} \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{n! \mathrm{e}^n}{n^n \sqrt{n}} \times \frac{1}{\mathrm{e}^n} \sum_{k=0}^{n} \frac{n^k}{k!} = a_n b_n$$

and the monotonicity property results from Lemmas 1 and 2. □

2.1. Relation to prior work

The expression for C_n plays an important role in coding theory, average case analysis of algorithms, combinatorics, which is the reason for the interest in its properties. We briefly review in this section some results on asymptotic expansion of C_n. In general, these results have been obtained by applying non-elementary techniques, while the proof of Theorem 3 is elementary. The monotonicity property of the sequence with the general term given by $\frac{C_n}{\sqrt{n}}$ is seen to not be a direct consequence of the previously known bounds and asymptotic expansions.

We note the relationship of the sum of C_n to a famous conjecture of Ramanujan, namely, that if n is a positive integer then $\frac{1}{2}\mathrm{e}^n = \sum_{i=0}^{n-1} \frac{n^i}{i!} + \frac{n^n}{n!}\tau(n)$, where $\tau(n)$ lies between $1/2$ and $1/3$. There are different proofs of this conjecture; Karamata shows that

$$f_n \triangleq 1 + \frac{1}{2}\frac{n!\mathrm{e}^n}{n^n} - \frac{n!}{n^n}\sum_{i=0}^{n}\frac{n^i}{i!} \tag{3}$$

is a monotone decreasing function from $1/2$ to $1/3$, when n increases from 1 to $+\infty$ [8]. The result of Ramanujan's conjecture is used in [9] to prove that

$$\sqrt{\frac{\pi n}{2}} + \frac{1}{2} < C_n < \sqrt{\frac{\pi n}{2}} + \frac{2}{3} + \frac{\alpha}{\sqrt{n}},$$

where $\alpha = (\mathrm{e}^{1/12} - 1)\sqrt{\frac{\pi}{2}}$. Moreover, in [9] an asymptotic expansion for C_n is conjectured and the following enhanced version of this conjecture is proved in [10]:

$$C_n = \sqrt{\frac{\pi n}{2}} + \frac{2}{3} + \frac{\sqrt{2\pi}}{24}\frac{1}{\sqrt{n}} - \frac{4}{135}\frac{1}{n} + \mathrm{O}\big(1/n^{3/2}\big).$$

By applying the definition of f_n (3) we can write

$$\frac{C_n}{\sqrt{n}} = \left[\frac{1}{\sqrt{n}} + \frac{1}{2}\frac{n!\mathrm{e}^n}{n^n\sqrt{n}}\right] - \frac{f_n}{\sqrt{n}}, \tag{4}$$

which is the difference of two monotone decreasing sequences, but this does not prove the monotonicity of the sequence $\frac{C_n}{\sqrt{n}}$. After some elementary calculations, we can rewrite (3) as $b_n = \frac{1}{2} + \frac{n^n}{n!\mathrm{e}^n}[1 - f_n]$. Since the sequence with the general term given by $\frac{n^n}{n!\mathrm{e}^n}$ is monotone decreasing, while the sequence $1 - f_n$ is positive and monotone increasing, the equation does not prove the monotonicity of the sequence b_n.

We show how the result of Theorem 3 can be used to find bounds for $\log C_n$.

Corollary 4. *For any* $n \geqslant 1$, $\log C_n$ *satisfies*

$$\log\sqrt{\frac{\pi n}{2}} < \log C_n \leqslant \log 2\sqrt{n}.$$

Proof. Theorem 3 implies that $\frac{C_n}{\sqrt{n}} \leqslant \frac{C_1}{\sqrt{1}} = 2$ for any $n \geqslant 1$. Now the upper bound for $\log C_n$ results from the observation $\log C_n \leqslant \log 2\sqrt{n}$. Since the monotone decreasing sequence $\frac{C_n}{\sqrt{n}}$, $n \geqslant 1$, is positive $\lim_{n\to\infty} \frac{C_n}{\sqrt{n}}$ is finite and nonnegative. The identity (4) implies $\lim_{n\to\infty} \frac{C_n}{\sqrt{n}} = \frac{1}{2}\lim_{n\to\infty} a_n$. The lower bound results from the limit of the sequence in Lemma 1 $\lim_{n\to\infty} a_n = \sqrt{2\pi}$, [6]. □

For n greater than any fixed positive integer N, the upper bound can be sharpened as $\log \frac{C_N}{\sqrt{N}}\sqrt{n}$. Also, since $\log \frac{C_n}{\sqrt{n}} = \log C_n - \frac{1}{2}\log n$, we see that the distance from $\log C_n$ to the usual approximation $\frac{1}{2}\log n$ decreases monotonely in n.

Appendix A. Proof of Lemma 2

Consider the difference

$$
\begin{aligned}
b_{n+1}-b_n &= \frac{1}{\mathrm{e}^{n+1}}\sum_{k=0}^{n+1}\frac{(n+1)^k}{k!}-\frac{1}{\mathrm{e}^n}\sum_{k=0}^{n}\frac{n^k}{k!}\\
&= \frac{1}{\mathrm{e}^{n+1}}\frac{(n+1)^{n+1}}{(n+1)!}+\frac{1}{\mathrm{e}^n}\sum_{k=0}^{n}\frac{1}{k!}\left(\frac{(n+1)^k}{\mathrm{e}}-n^k\right)\\
&= \frac{1}{\mathrm{e}^{n+1}}\left[\frac{(n+1)^n}{n!}+\sum_{k=0}^{n}\frac{(n+1)^k}{k!}-\mathrm{e}\sum_{k=0}^{n}\frac{n^k}{k!}\right]\\
&= \frac{1}{\mathrm{e}^{n+1}}\left[\sum_{p=0}^{n}\frac{n^p}{p!(n-p)!}+\sum_{k=0}^{n}\sum_{q=0}^{k}\frac{n^q}{q!(k-q)!}-\mathrm{e}\sum_{k=0}^{n}\frac{n^k}{k!}\right]\\
&= \frac{1}{\mathrm{e}^{n+1}}\left[\sum_{r=0}^{n}\frac{n^r}{r!}\left(\left(\sum_{s=0}^{n-r}\frac{1}{s!}\right)+\frac{1}{(n-r)!}-\mathrm{e}\right)\right].
\end{aligned}
$$

With the notation

$$
T(r)=\left(\sum_{s=0}^{n-r}\frac{1}{s!}\right)+\frac{1}{(n-r)!}-\mathrm{e},\quad r\in\{0,1,\ldots,n\},
$$

we see that for $r=n$, $T(n)=2-\mathrm{e}<0$. For e we have from [6], for any $r\in\{0,1,\ldots,n-1\}$, the identity

$$
\mathrm{e}=1+\frac{1}{1!}+\frac{1}{2!}+\cdots+\frac{1}{(n-r)!}+\frac{\tau(n-r)}{(n-r)!(n-r)},
$$

where $\tau(n-r)$ is a constant such that $0<\tau(n-r)<1$. For $0\leqslant r\leqslant n-1$, $T(r)$ becomes

$$
\begin{aligned}
T(r)&=\left(\sum_{s=0}^{n-r}\frac{1}{s!}\right)+\frac{1}{(n-r)!}-\left(\sum_{s=0}^{n-r}\frac{1}{s!}\right)-\frac{\tau(n-r)}{(n-r)!(n-r)}\\
&=\frac{1}{(n-r)!}\left(1-\frac{\tau(n-r)}{n-r}\right).
\end{aligned}
$$

Since the constant $\tau(n-r)$ is upper bounded by 1 we get the inequality $T(r)>0$ for $r\in\{0,1,\ldots,n-1\}$. Observe that the sequence $\frac{n^r}{r!}$, indexed by r, is increasing for $r\in\{0,1,\ldots,n-1\}$. These two observations together with $\frac{n^n}{n!}=\frac{n^{n-1}}{(n-1)!}$ imply:

$$
\begin{aligned}
b_{n+1}-b_n &< \frac{1}{\mathrm{e}^{n+1}}\frac{n^{n-1}}{(n-1)!}\left[\sum_{r=0}^{n}\left(\left(\sum_{s=0}^{n-r}\frac{1}{s!}\right)+\frac{1}{(n-r)!}-\mathrm{e}\right)\right]\\
&= \frac{1}{\mathrm{e}^{n+1}}\frac{n^{n-1}}{(n-1)!}\left[\sum_{r=0}^{n}\left(\left(\sum_{s=0}^{n-r}\frac{1}{s!}\right)+\frac{1}{(n-r)!}-\sum_{t=0}^{\infty}\frac{1}{t!}\right)\right]\\
&= \frac{1}{\mathrm{e}^{n+1}}\frac{n^{n-1}}{(n-1)!}\left[\sum_{r=0}^{n}\left(\frac{1}{(n-r)!}-\sum_{t=n-r+1}^{\infty}\frac{1}{t!}\right)\right]\\
&= \frac{1}{\mathrm{e}^{n+1}}\frac{n^{n-1}}{(n-1)!}\left[1-\sum_{s=2}^{n}\left(\frac{s-1}{s!}\right)-(n+1)\sum_{t=n+1}^{\infty}\frac{1}{t!}\right]\\
&= \frac{1}{\mathrm{e}^{n+1}}\frac{n^{n-1}}{(n-1)!}\left[1-\sum_{s=2}^{n}\left(\frac{s-1}{s!}\right)-\frac{1}{n!}-(n+1)\sum_{t=n+2}^{\infty}\frac{1}{t!}\right]
\end{aligned}
$$

$$= \frac{1}{e^{n+1}} \frac{n^{n-1}}{(n-1)!} \left[1 - 1 - (n+1) \sum_{t=n+2}^{\infty} \frac{1}{t!} \right]$$

$$= -\frac{n+1}{e^{n+1}} \frac{n^{n-1}}{(n-1)!} \sum_{t=n+2}^{\infty} \frac{1}{t!} < 0. \tag{A.1}$$

This inequality concludes the proof of the monotone decreasing property of the sequence $(b_n)_{n \geqslant 1}$. Note that in (A.1) the sum $\sum_{s=2}^{n} (\frac{s-1}{s!}) + \frac{1}{n!}$ was computed recursively by grouping the terms:

$$\sum_{s=2}^{n} \left(\frac{s-1}{s!} \right) + \frac{1}{n!} = \sum_{s=2}^{n-1} \left(\frac{s-1}{s!} \right) + \underbrace{\frac{n-1}{n!} + \frac{1}{n!}}_{\frac{1}{(n-1)!}}$$

$$= \sum_{s=2}^{n-2} \left(\frac{s-1}{s!} \right) + \underbrace{\frac{n-2}{(n-1)!} + \frac{1}{(n-1)!}}_{\frac{1}{(n-2)!}}$$

$$= \cdots = \frac{1}{2!} + \frac{1}{2!} = 1. \qquad \square$$

References

[1] A. Barron, J. Rissanen, B. Yu, The minimum description length principle in coding and modeling, IEEE Trans. Inform. Theory 44 (1998) 2743–2760.

[2] J. Rissanen, Fisher information and stochastic complexity, IEEE Trans. Inform. Theory 42 (1996) 40–47.

[3] Y. Shtarkov, Universal coding of single messages, Problems Inform. Transmission 23 (3) (1987) 3–17.

[4] J. Rissanen, Strong optimality of the normalized ML models as universal codes and information in data, IEEE Trans. Inform. Theory 47 (5) (2001) 1712–1717.

[5] J. Rissanen, Stochastic Complexity in Statistical Inquiry, World Scientific, Singapore, 1989.

[6] G. Fikhtengolts, The Fundamentals of Mathematical Analysis, Vols. 1–2, Pergamon Press, Oxford, 1965.

[7] J. Riordan, N. Sloane, The enumeration of rooted trees by total height, J. Australian Math. Soc. 10 (1969) 278–282.

[8] J. Karamata, Sur quelques problemes poses par Ramanujan, J. Indian Math. Soc. (N.S.) 24 (1960) 343–365.

[9] T. Klove, Bounds on the worst case probability of undetected error, IEEE Trans. Inform. Theory 41 (1) (1995) 298–300.

[10] W. Szpankowski, On asymptotics of certain sums arising in coding theory, IEEE Trans. Inform. Theory 41 (6) (1995) 2087–2090.