

УДК 519.987

Асимптотика дисперсии оценки энтропии для симметричных мер Бернулли

Тимофеев Е.А.¹

Ярославский государственный университет им. П.Г.Демидова

e-mail: timofeeva@gmail.com

получена 20 июня 2009

Ключевые слова: энтропия, непараметрическая оценка, моменты, мера Бернулли

Найдена точная асимптотика для дисперсии оценки энтропии П. Грассбергера [1] в случае симметричной бернуллиевской меры (с равными вероятностями символов).

В работе П. Грассбергера [1] предложена статистическая оценка энтропии h меры, которая строится по n независимым заданным точкам, распределенным по этой мере. В [4] предложено обобщение этой оценки и доказано степенное (по n) убывание дисперсии для достаточно широкого класса мер. В [5] для симметричной бернуллиевской меры найдено смещение как оценки Грассбергера, так и ее обобщения.

В настоящей работе для этой же меры найдена асимптотическая формула для дисперсии оценки Грассбергера.

1. Постановка задачи и обозначения

Через $\Omega = \mathcal{A}^{\mathbb{N}}$ обозначим пространство правосторонних последовательностей символов из конечного алфавита $\mathcal{A}$, где $\mathbb{N} = \{1, 2, \dots\}$. Введем на Ω метрику

$$\rho(x, y) = \frac{1}{\max \{k : x_k \neq y_k\}}. \quad (1)$$

Пусть $\xi_0, \xi_1, \dots, \xi_n$ – независимые одинаково распределенные случайные величины, принимающие значения в Ω и имеющие общее распределение – меру μ .

Статистическая оценка П. Грассбергера [1] η_n строится следующими простыми вычислениями:

¹Работа поддержана РФФИ, грант №07-01-00702а, и проводилась при финансовой поддержке Федерального агентства по науке и инновациям, ФЦП "Научные и научно-педагогические кадры инновационной России на 2009-2013 годы".

- находим вспомогательные случайные величины

$$R_{n,i} = \left(\min_{j:j \neq i} \rho(\xi_i, \xi_j) \right)^{-1}; \quad (2)$$

- полагаем оценкой обратной энтропии случайную величину

$$\eta_n = \frac{1}{(n+1) \ln n} \sum_{i=0}^n R_{n,i}. \quad (3)$$

В статье будет рассматриваться симметричная бернуллиевская мера (с равными вероятностями символов). Для этой меры распределение расстояния от случайной точки ξ_0 до любой точки $x \in \Omega$ не зависит от x и находится как

$$\mathbf{P}\{\rho(x, \xi_0)^{-1} \leq k\} = 1 - e^{-hk}, \quad (4)$$

где через

$$h = \ln |\mathcal{A}| \quad (5)$$

обозначается энтропия симметричной меры Бернулли (натуральное основание логарифма выбрано для удобства вычислений).

Основной результат работы — доказательство следующей асимптотической формулы

$$\mathbf{D}\eta_n = \mathcal{O}((n \ln^2 n)^{-1}). \quad (6)$$

2. Совместное распределение величин $R_{n,i}$

Найдем $\mathbf{P}\{R_{n,0} \leq a, R_{n,1} \leq b\}$ для $a \leq b$.

По формуле полной вероятности и (4) имеем

$$\mathbf{P}\{R_{n,0} \leq a, R_{n,1} \leq b\} = \sum_{m=1}^a \mathbf{P}\{R_{n,0} \leq a, R_{n,1} \leq b \mid \rho(\xi_0, \xi_1) = 1/m\} e^{-hm} (e^h - 1).$$

В силу независимости и одинаковости распределенности величин $\xi_0, \xi_1, \dots, \xi_n$

$$\begin{aligned} \mathbf{P}\{R_{n,0} \leq a, R_{n,1} \leq b\} &= \\ &= \sum_{m=1}^a \mathbf{P}\{\rho(\xi_0, \xi_2) \geq 1/a, \rho(\xi_1, \xi_2) \geq 1/b \mid \rho(\xi_0, \xi_1) = 1/m\}^{n-1} e^{-hm} (e^h - 1). \end{aligned}$$

Нетрудно видеть, что для рассматриваемой метрики при $m \leq a \leq b$

$$\mathbf{P}\{\rho(\xi_0, \xi_2) < 1/a, \rho(\xi_1, \xi_2) < 1/b \mid \rho(\xi_0, \xi_1) = 1/m\} = 0.$$

Поэтому

$$\begin{aligned} \mathbf{P}\{\rho(\xi_0, \xi_2) \geq 1/a, \rho(\xi_1, \xi_2) \geq 1/b \mid \rho(\xi_0, \xi_1) = 1/m\} &= \\ = 1 - \mathbf{P}\{\rho(\xi_0, \xi_2) < 1/a \mid \rho(\xi_0, \xi_1) = 1/m\} - \mathbf{P}\{\rho(\xi_1, \xi_2) < 1/b \mid \rho(\xi_0, \xi_1) = 1/m\} &= \\ = 1 - e^{-ha} - e^{-hb}. \end{aligned}$$

Следовательно,

$$P\{R_{n,0} \leq a, R_{n,1} \leq b\} = (1 - e^{-ha})(1 - e^{-ha} - e^{-hb})^{n-1}, \quad a \leq b. \quad (7)$$

Поскольку задача нахождения дисперсии величины η_n , определенной в (3), сводится к нахождению $ER_{n,0}R_{n,1}$, $ER_{n,0}^2$ и $ER_{n,0}$, то для их нахождения нужно найти следующие суммы:

$$\begin{aligned} ER_{n,0}R_{n,1} &= \sum_{a=1}^{\infty} \sum_{b=1}^{\infty} P\{R_{n,0} \geq a, R_{n,1} \geq b\} = \\ &= \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} [1 - P\{R_{n,0} \leq a\} - P\{R_{n,1} \leq b\} + P\{R_{n,0} \leq a, R_{n,1} \leq b\}] = \\ &= \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} [1 - (1 - e^{-ha})^n - (1 - e^{-hb})^n + (1 - e^{-h \min\{a,b\}})(1 - e^{-ha} - e^{-hb})^{n-1}]. \quad (8) \end{aligned}$$

$$\begin{aligned} ER_{n,0}^2 &= \sum_{a=1}^{\infty} a^2 P\{R_{n,0} = a\} = \sum_{a=0}^{\infty} (2a+1) P\{R_{n,0} \geq a\} = \\ &= \sum_{a=0}^{\infty} (2a+1) [1 - (1 - e^{-ha})^n]. \quad (9) \end{aligned}$$

Отметим, что асимптотика суммы для $ER_{n,0}$ найдена в [5], но она здесь не приводится потому, что в следующем разделе будет найдена более общая сумма.

3. Асимптотика вспомогательных функций

В этом разделе будут найдены асимптотики четырех вспомогательных сумм с применением формулы суммирования Пуассона [2, (13.8.4)]

$$\sum_{k=0}^{\infty} f(k) = \int_0^{\infty} f(t) dt + \frac{1}{2}f(0) - \sum_{k \neq 0} \frac{1}{2\pi i k} \int_0^{\infty} f'(t) e^{2\pi i k t} dt. \quad (10)$$

Это суммы

$$\Phi_n(t) = \sum_{k=0}^{\infty} [1 - (1 - te^{-hk})^n], \quad (11)$$

$$\Phi_n^{(2)} = \sum_{k=0}^{\infty} k [1 - (1 - te^{-hk})^n], \quad (12)$$

$$F(x) = \sum_{k=0}^{\infty} [1 - e^{-xe^{-hk}}], \quad (13)$$

$$F(x) = \sum_{k=0}^{\infty} x \left[1 - e^{-xe^{-hk}} \right]. \quad (14)$$

Отметим, что $\Phi_n(1) = ER_{n,0}$, а $ER_{n,0}^2 = 2\Phi_n^{(2)} + \Phi_n(1)$.

Отметим также, что усреднение функции $\Phi_n(t)$ по распределению Пуассона с параметром x дает функцию $F(xt)$, поэтому при $n \rightarrow \infty$ функции $\Phi_n(t)$ и $F(nt)$ совпадают [6, гл.VII,(1.5)].

Утверждение 1. При $0 < t < 2$ и $n \rightarrow \infty$

$$\Phi_n(t) = \frac{1}{h}H_n + \frac{1}{h}\ln t + \frac{1}{2} + \sum_{k \neq 0} \frac{t^{2\pi ik/h} \Gamma(1 - 2\pi ik/h) \Gamma(n+1)}{2\pi ik \Gamma(n+1 - 2\pi ik/h)} + \mathcal{O}(n(1-t)^n), \quad (15)$$

где через H_s обозначаются гармонические числа

$$H_n = \sum_{k=1}^n \frac{1}{k}. \quad (16)$$

Доказательство. Применяя формулу (10) для функции $f(x) = 1 - (1 - te^{-xh})^n$, получим

$$\begin{aligned} \Phi_n(t) = \int_0^{\infty} [1 - (1 - te^{-xh})^n] dx + \frac{1}{2} - \frac{1}{2}(1-t)^n + \\ + \sum_{k \neq 0} \frac{nht}{2\pi ik} \int_0^{\infty} e^{2\pi ikx - hx} (1 - te^{-xh})^{n-1} dx. \end{aligned}$$

После замены $y = te^{-xh}$ первый интеграл вычисляется, а второй сводится к неполной бета-функции. В результате получим

$$\begin{aligned} \Phi_n(t) = \frac{1}{h}H_n - \frac{1}{h} \sum_{k=1}^n \frac{1}{k} (1-t)^k + \frac{1}{2} - \frac{1}{2}(1-t)^n + \\ + \sum_{k \neq 0} \frac{n}{2\pi ik} t^{2\pi ik/h} B(1 - 2\pi ik/h, n; t), \quad (17) \end{aligned}$$

где $B(a, b; t)$ – неполная бета-функция

$$B(a, b; t) = \int_0^t x^{a-1} (1-x)^{b-1} dx. \quad (18)$$

Легко проверяемое тождество

$$B(a, b; t) = \frac{t^a}{a} (1-t)^{b-1} + \frac{b-1}{a} B(a+1, b-1; t)$$

и оценка

$$|B(a, b; t) - B(a, b)| \leq \frac{1}{b} t^{Re a - 1} |t - 1|^b, \quad b > 1$$

дают

$$\left| B(1 - 2\pi ik/h, n; t) - \frac{t^{1-2\pi ik/h}(1-t)^{n-1}}{1-2\pi ik/h} - \frac{n-1}{1-2\pi ik/h} B(2-2\pi ik/h, n-1) \right| \leq \frac{t|1-t|^{n-1}}{|1-2\pi ik/h|}.$$

Подставляя эту оценку в (17), получим

$$\begin{aligned} \Phi_n(t) &= \frac{1}{h} H_n - \frac{1}{h} \sum_{k=1}^n \frac{1}{k} (1-t)^k + \frac{1}{2} + \\ &+ \sum_{k \neq 0} \frac{n(n-1)t^{2\pi ik/h}}{2\pi ik(1-2\pi ik/h)} B(2-2\pi ik/h, n-1) + \mathcal{O}(n(1-t)^n). \end{aligned}$$

Выражая бета-функцию через гамма-функции и подставляя разложение $\ln t$ в ряд по $1-t$, получим (15). $\square$

Утверждение 2. При $x \rightarrow \infty$

$$F(x) = \frac{\ln x}{h} + \frac{\mathcal{C}}{h} + \frac{1}{2} + \sum_{k \neq 0} \frac{x^{2\pi ik/h}}{2\pi ik} \Gamma(1-2\pi ik/h) + \mathcal{O}(e^{-x}), \quad (19)$$

где через $\mathcal{C}$ обозначается константа Эйлера.

Доказательство. Применяя формулу (10) для функции $f(t) = 1 - e^{-xe^{-th}}$, получим

$$F(x) = \int_0^\infty [1 - e^{-xe^{-ht}}] dt + \frac{1}{2} - \frac{1}{2}e^{-x} + \sum_{k \neq 0} \frac{hx}{2\pi ik} \int_0^\infty e^{2\pi ikt - ht} e^{-xe^{-ht}} dt.$$

После замены $y = xe^{-ht}$ второй интеграл сводится к неполной (нижней) гамма-функции $\gamma(1-2\pi ik/h, x)$, а первый приводится к следующему виду:

$$I(x) = \int_0^\infty [1 - e^{-xe^{-ht}}] dt = \frac{1}{h} \int_0^x \frac{1 - e^{-y}}{y} dy.$$

Применяя формулу [3, 8.367.12], получим

$$hI(x) = \ln x + \mathcal{C} + \int_x^\infty \frac{e^{-y}}{y} dy. \quad (20)$$

Итак,

$$F(x) = \frac{\ln x + \mathcal{C}}{h} + \frac{1}{h} \int_x^\infty \frac{e^{-y}}{y} dy + \frac{1}{2} - \frac{1}{2}e^{-x} + \sum_{k \neq 0} \frac{x^{2\pi ik/h}}{2\pi ik} \gamma(1-2\pi ik/h, x). \quad (21)$$

Поскольку

$$|\Gamma(1 - 2\pi ik/h) - \gamma(1 - 2\pi ik/h, x)| \leq \int_x^\infty e^{-t} dt = e^{-x},$$

то, подставляя эту оценку в (21), получим (19). $\square$

Следствие 1. При $n \rightarrow \infty$

$$\begin{aligned} ER_{n,0} = \Phi_n(1) = F(n) + \mathcal{O}(n^{-1}) = \\ = \frac{\ln n}{h} + \frac{C}{h} + \frac{1}{2} + \sum_{k \neq 0} \frac{n^{2\pi ik/h}}{2\pi ik} \Gamma(1 - 2\pi ik/h) + \mathcal{O}(n^{-1}). \end{aligned} \quad (22)$$

Утверждение 3. При $n \rightarrow \infty$

$$\begin{aligned} \Phi_n^{(2)} = \frac{1}{2h^2} H_n^2 + \frac{\pi^2}{12h^2} - \frac{1}{12} + \sum_{k \neq 0} \frac{n^{2\pi ik/h}}{4\pi^2 k^2} \Gamma(1 - 2\pi ik/h) + \\ + \ln n \sum_{k \neq 0} \frac{n^{2\pi ik/h}}{2\pi ik} \Gamma(1 - 2\pi ik/h) - \sum_{k \neq 0} \frac{n^{2\pi ik/h}}{2\pi ik} \Gamma'(1 - 2\pi ik/h) + \mathcal{O}(n^{-1}). \end{aligned} \quad (23)$$

Доказательство. Применяя формулу (10) для функции $f(x) = x[1 - (1 - e^{-xh})^n]$, получим

$$\begin{aligned} \Phi_n^{(2)} = \int_0^\infty x [1 - (1 - e^{-xh})^n] dx - \sum_{k \neq 0} \frac{1}{2\pi ik} \int_0^\infty e^{2\pi ikx} [1 - (1 - e^{-xh})^n] dx + \\ + \sum_{k \neq 0} \frac{nh}{2\pi ik} \int_0^\infty x e^{2\pi ikx - hx} (1 - e^{-xh})^{n-1} dx. \end{aligned}$$

После замены $y = e^{-xh}$ первый интеграл приводится к сумме известных интегралов [3, 4.2533.1]

$$-\frac{1}{h^2} \sum_{m=1}^n \int_0^1 \ln y \frac{1 - (1 - y)^n}{y} dy = \frac{1}{h^2} \sum_{m=1}^n \frac{H_m}{m} = \frac{1}{2h^2} \left(H_n^2 + \frac{\pi^2}{6} \right) + \mathcal{O}(n^{-1});$$

второй сводится к бета-функции

$$\begin{aligned} \frac{1}{h} \int_0^1 y^{-2\pi ik/h-1} [1 - (1 - y)^n] dy = -\frac{1}{2\pi ik} \int_0^1 [1 - (1 - y)^n] dy^{-2\pi ik/h} = \\ = -\frac{1}{2\pi ik} + \frac{n}{2\pi ik} B(1 - 2\pi ik/h, n); \end{aligned}$$

а третий по формуле [3, 4.2533.1] сводится к бета и пси-функциям

$$\begin{aligned} -\frac{1}{h^2} \int_0^1 \ln y \, y^{-2\pi ik/h} (1-y)^{n-1} dy &= \\ &= \frac{1}{h^2} B(1-2\pi ik/h, n) [\psi(n+1-2\pi ik/h) - \psi(1-2\pi ik/h)]. \end{aligned}$$

Подставив эти интегралы, получим

$$\begin{aligned} \Phi_n^{(2)} &= \frac{H_n^2}{2h^2} + \frac{\pi^2}{12h^2} - \sum_{k \neq 0} \frac{1}{4\pi^2 k^2} + \sum_{k \neq 0} \frac{n}{4\pi^2 k^2} B(1-2\pi ik/h, n) + \\ &+ \sum_{k \neq 0} \frac{n}{2\pi i k h} B(1-2\pi ik/h, n) [\psi(n+1-2\pi ik/h) - \psi(1-2\pi ik/h)] + \mathcal{O}(n^{-1}). \end{aligned}$$

Выражая бета-функцию через гамма-функции и подставляя разложение $\psi(n)$ в ряд по n , получим (23). $\square$

Утверждение 4. При $x \rightarrow \infty$

$$\begin{aligned} F^{(2)}(x) &= \frac{1}{2h^2} \ln^2 x + \frac{\mathcal{C} \ln x}{h^2} + \frac{\mathcal{C}^2}{2h^2} + \frac{\pi^2}{12h^2} - \frac{1}{12} + \sum_{k \neq 0} \frac{x^{2\pi ik/h}}{4\pi^2 k^2} \Gamma(1-2\pi ik/h) + \\ &+ \ln x \sum_{k \neq 0} \frac{x^{2\pi ik/h}}{2\pi i k} \Gamma(1-2\pi ik/h) - \sum_{k \neq 0} \frac{x^{2\pi ik/h}}{2\pi i k} \Gamma'(1-2\pi ik/h) + \mathcal{O}(x^{-1}). \end{aligned} \quad (24)$$

Доказательство. Применяя формулу (10) для функции $f(t) = t(1 - e^{-xe^{-ht}})$, получим

$$F^{(2)}(x) = \int_0^\infty t \left[1 - e^{-xe^{-ht}} \right] dt - \sum_{k \neq 0} \frac{1}{2\pi i k} \int_0^\infty e^{2\pi i k t} \left[1 - e^{-xe^{-ht}} - t x h e^{-ht} e^{-xe^{-ht}} \right] dt.$$

Перепишем эту формулу, как

$$F^{(2)}(x) = I_0 - \sum_{k \neq 0} \frac{1}{2\pi i k} I_1 + \sum_{k \neq 0} \frac{xh}{2\pi i k} I_2, \quad (25)$$

и найдем каждый из интегралов I_0, I_1, I_2 .

После замены $y = xe^{-ht}$ имеем

$$I_0 = -\frac{1}{h^2} \int_0^x \ln y \frac{1-e^{-y}}{y} dy + \frac{\ln x}{h^2} \int_0^x \frac{1-e^{-y}}{y} dy.$$

Проинтегрировав первый интеграл по частям (по $d \ln y$), получим

$$2 \int_0^x \ln y (1 - e^{-y}) d \ln y = \ln^2 x (1 - e^{-x}) - \int_0^x \ln^2 y e^{-y} dy.$$

Применяя формулу [3, 4.358.1], получим

$$\int_0^x \ln y (1 - e^{-y}) d \ln y = \frac{1}{2} \ln^2 x - \frac{\mathcal{C}^2}{2} - \frac{\pi^2}{12} + \mathcal{O}(x^{-1}).$$

Подставив $I(x)$ из (20), находим

$$I_0 = \frac{\ln^2 x}{2h^2} + \frac{\mathcal{C} \ln x}{h^2} + \frac{\mathcal{C}^2}{2h^2} + \frac{\pi^2}{12h^2} + \mathcal{O}(x^{-1}). \quad (26)$$

Так же после замены $y = xe^{-ht}$ имеем

$$I_1 = \frac{x^{2\pi ik/h}}{h} \int_0^x y^{-1-2\pi ik/h} (1 - e^{-y}) dy.$$

Проинтегрировав по частям и подставив гамма-функцию, получим

$$I_1 = -\frac{1}{2\pi ik} + \frac{x^{2\pi ik/h}}{2\pi ik} \Gamma(1 - 2\pi ik/h) + \mathcal{O}(x^{-1}). \quad (27)$$

Эта же замена $y = xe^{-ht}$ для третьего интеграла дает

$$I_2 = -\frac{x^{2\pi ik/h-1}}{h^2} \int_0^x \ln y y^{-2\pi ik/h} e^{-y} dy + \ln x \frac{x^{2\pi ik/h-1}}{h^2} \int_0^x y^{-2\pi ik/h} e^{-y} dy.$$

Подставив гамма-функцию и ее производную, получим

$$I_2 = -\frac{x^{2\pi ik/h-1}}{h^2} \Gamma'(1 - 2\pi ik/h) + \ln x \frac{x^{2\pi ik/h-1}}{h^2} \Gamma(1 - 2\pi ik/h) + \mathcal{O}(x^{-1}). \quad (28)$$

Подставляя в (25) найденные интегралы (26) – (28), получим (24). $\square$

Следствие 2. При $n \rightarrow \infty$

$$\mathbb{E}R_{n,0}^2 = 2\Phi_n^{(2)} + \Phi_n(1) = 2F^{(2)}(n) + F(n) + \mathcal{O}(n^{-1}). \quad (29)$$

4. Нахождение $\mathbb{E}R_{n,0}R_{n,1}$

Для нахождения $\mathbb{E}R_{n,0}R_{n,1}$ по формуле (8) предварительно найдем суммы

$$S_{n,m} = \sum_{a=0}^{\infty} [1 - (1 - e^{-ha})^n - (1 - e^{-ha-hm})^n + (1 - e^{-ha})(1 - e^{-ha} - e^{-ha-hm})^{n-1}]. \quad (30)$$

Заменив множитель $1 - e^{-ha}$ на

$$1 - e^{-ha} = \frac{1}{1 + e^{-hm}} (1 - e^{-ha} (1 + e^{-hm})) + \frac{e^{-hm}}{1 + e^{-hm}}$$

и используя обозначение (11), получим

$$S_{n,m} = \Phi_n(1) + \Phi_n(e^{-hm}) - \frac{1}{1+e^{-hm}}\Phi_n(1+e^{-hm}) - \frac{e^{-hm}}{1+e^{-hm}}\Phi_{n-1}(1+e^{-hm}).$$

Перепишем это тождество в более удобном для суммирования виде

$$S_{n,m} = \Phi_n(1) - \Phi_n(1+e^{-hm}) + \Phi_n(e^{-hm}) + \frac{e^{-hm}}{1+e^{-hm}} [\Phi_n(1+e^{-hm}) - \Phi_{n-1}(1+e^{-hm})]. \quad (31)$$

Основной трудностью в нахождении суммы по m является нахождение суммы первой разности, для определения асимптотики которой поступим следующим образом: заменим функцию $\Phi_n(t)$ на $F(nt)$, найдем соответствующую сумму для F и покажем, что разность этих сумм мала.

Итак, найдем

$$T_n = \sum_{m=1}^{\infty} [F(n(1+e^{-hm})) - F(n)]. \quad (32)$$

Подставляя (13), получим

$$T_n = \sum_{m=1}^{\infty} \sum_{k=0}^{\infty} [e^{-ne^{-hk}} - e^{-ne^{-hk} - ne^{-h(k+m)}}].$$

Заменяя m на $l = k + m$, получим

$$\begin{aligned} T_n &= \sum_{k=0}^{\infty} \sum_{l>k} [1 - (1 - e^{-ne^{-hk}})] [1 - e^{-ne^{-hl}}] = \\ &= \sum_{k=0}^{\infty} \sum_{l>k} [1 - e^{-ne^{-hl}}] - \frac{1}{2} \sum_{k=0}^{\infty} \sum_{l=0, l \neq k}^{\infty} [1 - e^{-ne^{-hk}}] [1 - e^{-ne^{-hl}}] = \\ &= F^{(2)}(n) - \frac{1}{2} (F^2(x) - 2F(x) + F(2x)). \quad (33) \end{aligned}$$

Лемма 1. При $n \rightarrow \infty$

$$T_n + \sum_{m=1}^{\infty} [\Phi_n(1) - \Phi_n(1+e^{-hm})] = \mathcal{O}(n^{-1}).$$

Доказательство. Обозначив через

$$Z_{n,m} = F(n(1+e^{-hm})) - F(n) + \Phi_n(1) - \Phi_n(1+e^{-hm})$$

и подставив (19), (15), получим

$$\begin{aligned} Z_{n,m} &= \sum_{k \neq 0} \frac{\Gamma(1 - 2\pi i k/h)}{2\pi i k} \left[n^{2\pi i k/h} - \frac{\Gamma(n+1)}{\Gamma(n+1 - 2\pi i k/h)} \right] [(1+e^{-hm})^{2\pi i k/h} - 1] + \\ &\quad + \mathcal{O}(ne^{-nmh}). \end{aligned}$$

По формуле Стирлинга

$$n^{2\pi ik/h} - \frac{\Gamma(n+1)}{\Gamma(n+1-2\pi ik/h)} = \mathcal{O}(n^{-1}). \quad (34)$$

Для оценивания суммы

$$\sum_{m=1}^{\infty} [(1 + e^{-hm})^{2\pi ik/h} - 1]$$

воспользуемся интегральным признаком. Соответствующий интеграл после замены переменных и интегрирования по частям принимает следующий вид:

$$\begin{aligned} \int_0^{\infty} [(1 + e^{-hx})^{ia} - 1] dx &= \\ &= \frac{1}{h} \int_0^1 \frac{(1+t)^{ia} - 1}{t} dt = \\ &= -\frac{ia}{h} \int_0^1 (1+t)^{ia-1} \ln t dt. \end{aligned}$$

Последний интеграл оценивается как $\frac{|a|}{h}$.

Поскольку гамма-функция убывает по мнимой оси экспоненциально, то ряды по k сходятся. $\square$

Итак,

$$\begin{aligned} \sum_{m=1}^{\infty} S_{n,m} &= -T_n + \sum_{m=1}^{\infty} \Phi_n(e^{-hm}) + \\ &+ \sum_{m=1}^{\infty} \frac{e^{-hm}}{1 + e^{-hm}} [\Phi_n(1 + e^{-hm}) - \Phi_{n-1}(1 + e^{-hm})] + \mathcal{O}(n^{-1}). \end{aligned}$$

Поскольку

$$\begin{aligned} \sum_{m=1}^{\infty} \Phi_n(e^{-hm}) &= \sum_{m=1}^{\infty} \sum_{k=m}^{\infty} [1 - (1 - e^{-hk})^n] = \Phi_n^{(2)}, \\ \Phi_n(1 + e^{-hm}) - \Phi_{n-1}(1 + e^{-hm}) &= \mathcal{O}(n^{-1}) \end{aligned}$$

то, подставляя (33), получим

$$\sum_{m=1}^{\infty} S_{n,m} = \frac{1}{2} (F^2(x) - 2F(x) + F(2x)) + \mathcal{O}(n^{-1}). \quad (35)$$

Формула (8) в обозначениях (30) выглядит следующим образом:

$$ER_{n,0}R_{n,1} = S_{n,0} + 2 \sum_{m=1}^{\infty} S_{n,m}.$$

Подставляя (35) и

$$S_{n,0} = 2\Phi_n(1) - \Phi_n(2) + \frac{1}{2} [\Phi_n(2) - \Phi_{n-1}(2)] = 2F(n) - F(2n) + \mathcal{O}(n^{-1}),$$

получим

$$\mathbb{E}R_{n,0}R_{n,1} = F^2(n) + \mathcal{O}(n^{-1}).$$

Следовательно,

$$\text{Cov}(R_{n,0}, R_{n,1}) = \mathcal{O}(n^{-1}). \quad (36)$$

Из (29) следует, что $\mathbb{D}R_{n,0} = \mathcal{O}(n^{-1})$. Таким образом,

$$\mathbb{D}\eta_n = \frac{1}{n^2 \ln^2 n} (n\mathbb{D}R_{n,0} + n(n-1)\text{Cov}(R_{n,0}, R_{n,1})) = \mathcal{O}((n \ln^2 n)^{-1}).$$

Основной результат — формула (6).

Список литературы

1. Grassberger P. Estimating the information content of symbol sequences and efficient codes // IEEE Trans. Inform. Theory. 1989. **IT-35**. P. 669-675.
2. Харди Г. Расходящиеся ряды. М.: ИЛ, 1951.
3. Градштейн И.С., Рыжик И.М. Таблицы интегралов, сумм, рядов и произведений. М.: Наука, 1971.
4. Тимофеев Е.А. Состоятельная оценка энтропии мер и динамических систем // Мат. заметки. 2005. Т. 77. №6. С. 903 – 916.
5. Тимофеев Е.А. Смещение статистической оценки энтропии для простейшей меры Бернулли // Модел. и анализ информ. систем. 2006. Т.13, №1. С. 41 – 45.
6. Феллер В. Введение в теорию вероятностей и ее приложения. Том 2. М. : Мир, 1984.

On asymptotics of the entropy estimator variance for symmetric Bernoulli measures

Timofeev E.A.

Keywords: entropy, nonparametric statistic, moments, Bernoulli measure

We consider Bernoulli measures with equally probable symbols and obtain a closed-form expression of the Grassberger entropy estimator variance.

Сведения об авторе: Тимофеев Евгений Александрович,
Ярославский государственный университет им. П.Г.Демидова,
доктор физ.-мат. наук, профессор