Записки научных
семинаров ПОМИ
Том 298, 2003 г.

Н. Д. Кан, В. Б. Невзоров

О НЕКОТОРЫХ ПРЕДЕЛЬНЫХ ЗАКОНАХ В ЗАДАЧЕ КОЛЛЕКЦИОНЕРА

Введение

Классическая задача коллекционера использует следующую модель. В урне находятся n шаров, которые вытаскиваются по одному. Вытащенный шар окрашивается и возвращается в урну. Наиболее популярные задачи, связанные с этой моделью, сводятся к нахождению распределений и различных численных характеристик числа шаров, окрашенных после r извлечений, и числа извлечений, необходимых для того, чтобы получить k окрашенных шаров. Обзор обширной литературы, в которой изучается задача коллекционера и связанные с ней проблемы, можно найти в работе Штадье ([4]). В ряде работ ([3, 4]) рассматривалась схема, в которой шары извлекаются не по одному, а комплектами по L шаров. В данной заметке мы исследуем асимптотическое поведение случайной величины $S(n,k,L)$ – числа групп по L шаров, которые необходимо извлечь, чтобы окрашенными оказались не менее k (из n) шаров.

Постановка задачи и связь с классической моделью

В урне находятся n неокрашенных шаров. Случайно выбираем группу из L шаров. Если $L = 1$, то имеем классическую схему. Мы будем рассматривать случай, когда объем выборки L представляет некоторое фиксированное целое число. Будем также рассматривать схему серий, в которой $L = L(n)$ может расти с ростом n. Выбранные шары окрашиваются и возвращаются в урну. После этого повторяем процедуру до тех пор, пока не окрасим k шаров. Пусть $S(n,k,L)$ – число извлечений, необходимых для получения k окрашенных шаров. Через $S(n,k)$ будем обозначать величину $S(n,k,1)$. Асимптотическое поведение случай-

Работа поддержана грантами Президента Российской Федерации (НШ-2258.2003.1) и РФФИ (01-01-00031, 02-01-00779).

ной величины $S(n,k)$ детально изучено. При различных взаимоотношениях $k = k(n)$ и n надлежащим образом центрированная и нормированная величина $S(n,k)$ может иметь в пределе вырожденное, пуассоновское, нормальное распределение и распределение экстремальных значений. Соответствующие асимптотические распределения были получены Баумом и Биллингсли ([1]). Цель нашей заметки – показать, как результаты, полученные для классической модели коллекционера, могут быть перенесены на случай коллекционирования комплектами. Мы будем опираться на следующие утверждения Баума и Биллингсли, полученные для классической $(L = 1)$ схемы.

Теорема 1 (Baum, Billingsley ([1])). a) *Если при $n \to \infty$, $k = k(n)$ удовлетворяет соотношениям $\frac{k(n)}{n^{1/2}} \to \infty$ и $n - k(n) \to \infty$, то*

$$\sup_x \left| P\left\{ \frac{S(n,k) - a(n,k)}{b(n,k)} < x \right\} - \Phi(x) \right| \longrightarrow 0, \tag{1}$$

где

$$a(n,k) = ES(n,k) = n \sum_{i=n-k+1}^{n} \frac{1}{i}, \tag{2}$$

$$b^2(n,k) = DS(n,k) = n \sum_{i=n-k+1}^{n} \frac{n-i}{i^2}, \tag{3}$$

и $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ – функция распределения стандартного нормального закона;

b) *если $\frac{k(n)}{\sqrt{n}} \to \lambda$, где $0 < \lambda < \infty$, то для любого $m = 0, 1, 2, \ldots$ справедливо соотношение*

$$\left| P\left\{ S(n,k) - k = m \right\} - \Pi_{\frac{\lambda^2}{2}}(m) \right| \longrightarrow 0, \quad n \to \infty, tag4$$

где $\Pi_\mu(m) = \frac{e^{-\mu}\mu^m}{m!}$ обозначает пуассоновские вероятности;

c) *если $k(n) = n$, то*

$$\sup_x \left| P\left\{ \frac{S(n,n) - a(n,n)}{b(n,n)} < x \right\} - H(x) \right| \longrightarrow 0, \quad n \to \infty, \tag{5}$$

где $H(x) = \exp(e^{-x})$.

Задача об асимптотическом распределении величины $S(n,k,L)$ может быть сведена к уже решенной (для случая $L=1$) следующим образом. Пусть T совпадает с числом извлечений в классической схеме, необходимых для получения L различных шаров. Тогда независимые случайные величины $T_1, T_2, \ldots$ имеющие такое же распределение, что и величина T, соответствуют числу извлечений в классической схеме, обеспечивающих появление последовательных групп из L различных шаров. Число окрашенных шаров в классической схеме после извлечения $N(r)$ шаров, где $N(r) = T_1 + \ldots + T_r$, имеет такое же распределение, что и число окрашенных шаров в схеме коллекционирования комплектами после извлечения r групп по L шаров. Это построение, предложенное Селлке ([3]), приводит к соотношению

$$P\{S(n,k,L) \leq m\} = P\{S(n,k) \leq T_1 + T_2 + \ldots T_m\}, \tag{6}$$

которое позволяет исследовать асимптотическое распределение $S(n,k,L)$, зная соответствующие результаты для $S(n,k)$. При использовании соотношения (6) важен тот факт, что величины $T_1, T_2, \ldots$ независимы. В данной заметке мы ограничиваемся рассмотрением ситуации, когда величина L является постоянной для всех комплектов (а следовательно, величины $T_1, T_2, \ldots$ имеют одинаковое распределение), но доказательства для случая различных чисел шаров в группах могут быть проведены по той же схеме, что и для постоянных по величине комплектов.

Основные результаты

Ниже мы приведем доказательства следующих трех теорем, в формулировке которых

$$d = d(n,L) = ES(n,L) = n \sum_{n-L+1}^{n} \frac{1}{i} \tag{7}$$

и

$$\sigma_L^2(n) = DS(n,L) = n \sum_{i=n-L+1}^{n} \frac{n-i}{i^2}. \tag{8}$$

Теорема 2. *Если при $n \to \infty$, $k = k(n)$ удовлетворяет соотношениям $\frac{k(n)}{\sqrt{n}} \to \infty$ и $n - k(n) \to \infty$, то для любого фиксированного*

$L = 1, 2, \ldots$ *выполняется соотношение*

$$\sup_x \left| P\left\{ \frac{dS(n,k,L) - a(n,k)}{b(n,k)} < x \right\} - \Phi(x) \right| \longrightarrow 0, \tag{9}$$

где центрирующие и нормирующие константы $a(n,k)$ *и* $b(n,k)$ *определены в* (2) *и* (3).

Замечание 1. Утверждение теоремы 3 остается справедливым и в схеме серий, когда объем выборки $L = L(n) = \bar{O}(\sqrt{n})$ при $n \to \infty$. В более общем виде в схеме серий справедливость соотношения (9) гарантируется, если $L = L(n)$, $k = k(n)$ и n удовлетворяют соотношениям $\frac{k}{\sqrt{n}} \to \infty$, $n - k \to \infty$ и $\frac{L(n-k)}{kn} \to 0$ при $n \to \infty$.

Теорема 3. *Для любого фиксированного* $L = 1, 2, \ldots$ *выполняется соотношение*

$$\sup_x \left| P\left\{ \frac{dS(n,n,L) - a(n,n)}{b(n,n)} < x \right\} - \exp(-e^{-x}) \right\} \right| \longrightarrow 0 \tag{10}$$

при $n \to \infty$.

Замечание 2. Утверждение (10) остается справедливым и в схеме серий, когда $L = L(n) = \bar{o}\,(n)$, при $n \to \infty$.

Если формулировки теорем 2 и 3 близки к утверждениям a) и c) теоремы 1, то в случае b) мы получим несколько иную формулировку, чем в теореме 1.

Теорема 4. *Если* $k(n)/n^{1/2} \to \lambda$, *где* $0 < \lambda < \infty$, *то для любого фиксированного* $L = 2, 3, \ldots$ *и любого* $r = 0, 1, 2, \ldots$ *при* $n \to \infty$ *справедливо соотношение*

$$P\{S(n, Lk, L) - k = r\} \longrightarrow p_r, \tag{11}$$

где

$$p_0 = P\{\pi_\mu = 0\}/2$$

и

$$p_r = P\{(r-1)L < \pi_\mu < rL\} + P\{\pi_\mu = (r-1)L\} + P\{\pi_\mu = rL\}/2,$$
$$r = 1, 2, \ldots,$$

а π_μ *здесь обозначает пуассоновскую случайную величину с параметром* $\mu = (L\lambda)^2/2$.

Вспомогательные результаты

Исследуем поведение центрирующих и нормирующих констант $a(n,k)$, $b(n,k)$, $d(n,L)$ и $\sigma_L^2(n)$ при различных соотношениях между k и n, L и n. Справедливы следующие очевидные неравенства:
если $1 \le k \le 3n/4$, то

$$k \le a(k,n) = n \sum_{i=n-k+1}^{n} \frac{1}{i} \le 4k \tag{12}$$

и

$$(k-1)^2/2n \le b^2(n,k) = n \sum_{i=n-k+1}^{n} \frac{n-i}{i^2} \le 16k^2/n. \tag{13}$$

Если $k > 3n/4$, то

$$3n/4 \le a(n,k) \le n\Bigg(\int\limits_{n-k+1}^{n} dx/x + 1/(n-k+1)\Bigg)$$
$$= n\ln(n/(n-k+1)) + n/(n-k+1) \tag{14}$$

и

$$(n-2)^2/8n \le b^2(n,k) = b^2(n,n/2) + n \sum_{i=n-k+1}^{n/2} \frac{n-i}{i^2} \le 4n + n^2/(n-k+1). \tag{15}$$

Отметим также, что

$$a(n,n) \sim n\ln n \tag{16}$$

и

$$b^2(n,n) \sim n^2\pi^2/6, \tag{17}$$

при $n \to \infty$.

Далее,

$$d = d(n,L) = a(n,L)$$

и

$$L \le d \le nL/(n-L+1). \tag{18}$$

Поскольку

$$\sigma_L^2(n) = b^2(n,L),$$

то для $1 \le L \le n/2$ справедливо неравенство

$$(L-1)^2/2n \le \sigma_L^2(n) \le 16L^2/n. \tag{19}$$

Доказательство теоремы 2. Используя соотношение (6), получаем, что

$$P\left\{\frac{dS(n,k,L)-a(n,k)}{b(n,k)} \le x\right\} = P\{S(n,k) \le T_1 + \ldots + T_m\}, \tag{20}$$

где

$$m = [(xb(n,k) + a(n,k)/d].$$

Не умаляя общности, будем считать, что $(xb(n,k) + a(n,k))/d$ является целым числом. Обозначим

$$\eta_n = (T_1 + \ldots + T_m - md)/m^{1/2}\sigma_L(n). \tag{21}$$

Из соотношений (12)–(19) нетрудно убедиться, что для любого фиксированного x последовательность $m = m(n)$ стремится к бесконечности, если $k/L = k(n)/L(n)$ и $n - k(n)$ стремятся к бесконечности с ростом n. В частности, эти два условия выполняются, если $L(n) = \bar{O}(n^{1/2})$, $k(n)/n^{1/2} \to \infty$ и $n - k(n) \to \infty$ при $n \to \infty$. Получаем, что для каждого фиксированного значения x величины η_n имеют в пределе стандартное нормальное распределение. Перепишем правую часть (20) в виде

$$P\{S(n,k) \le T_1 + \ldots + T_m\} = P\{(S(n,k) - a(n,k)/b(n,k) \le x + r_n\eta_n\}, \tag{22}$$

где

$$r_n = m^{1/2}\sigma_L(n)/b(n,k). \tag{23}$$

Вновь используя соотношения (12)–(19), убеждаемся, что для любого фиксированного значения x величина r_n стремится к нулю, если $n \to \infty$. Следовательно, $r_n\eta_n$ сходится к нулю по вероятности и , учитывая утверждение а) теоремы 1, получаем, что предел правой части (22) при любом фиксированном x равен $\Phi(x)$, что и доказывает теорему 2. Более того, мы показали, что утверждение теоремы сохраняет силу, если позволить величине L (в схеме серий) расти с ростом n таким образом, чтобы $L = L(n) = O(n^{1/2})$, $n \to \infty$.

Доказательство теоремы 3. По аналогии с (20) можно написать, что

$$P\left\{\frac{dS(n,n,L)-a(n,n)}{b(n,n)} \le x\right\} = P\{S(n,n) \le T_1+\ldots+T_m\}, \quad (24)$$

где $m=(xb(n,n)+a(n,n))/d$. Для того, чтобы использовать утверждение c) теоремы 1, первую часть (24) перепишем в виде

$$P\{S(n,n) \le T_1+\ldots+T_m\} =$$

$$= P\left\{\frac{S(n,n)-a(n,n)}{b(n,n)} \le \frac{T_1+\ldots+T_m-md}{b(n,n)} + \frac{md-a(n,n)}{b(n,n)}\right\} =$$

$$= P\left\{\frac{S(n,n)-a(n,n)}{b(n,n)} \le \frac{T_1+\ldots+T_m-md}{b(n,n)} + x\right\}. \quad (25)$$

Остается убедиться, что для любого фиксированного x, случайная величина

$$\mu_0 = \frac{T_1+\ldots+T_m-md}{b(n,n)}$$

стремится к нулю по вероятности. Поскольку для любого $\varepsilon>0$

$$P\{|\mu_n|>\varepsilon\} \le D(\mu_n)/\varepsilon^2 = mDT/\varepsilon^2 b^2(n,n) = m\sigma_L^2(n)/\varepsilon^2 b^2(n,n), \quad (26)$$

нам достаточно проверить, что при любом фиксированном x правая часть в соотношении (26) стремится к нулю при $n\to\infty$. В нашем случае $m=(xb(n,n)+a(n,n))/d$ и из соотношений (16) и (17) видим, что

$$m \sim a(n,n)/d \sim n\ln n/d$$

и

$$m\sigma_L^2(n)/b^2(n,n) \sim 6\ln n\sigma_L^2/(n)/d\pi^2 n.$$

Отсюда, анализируя поведение $\sigma_L^2(n)$ и d при различных условиях на поведение $L=L(n)$, убеждаемся, что правая часть (23) стремится к нулю не только при фиксированных значениях L, но и для таких $L=L(n)$ в схеме серий, для которых выполняется соотношение $L(n)=o(n)$ при $n\to\infty$.

Доказательство теоремы 4. При соотношении $k(n)\sim\lambda n^{1/2}$ и фиксированных значениях L с большой вероятностью в каждой новой группе приходят L еще незакрашенных шаров. Поэтому имеет смысл в данной ситуации рассматривать лишь моменты $S(n, kL, L)$. Записав соотношение

$$P\{S(n, Lk, L) - k \le r\} = P\{S(n, k) \le T_1 + \ldots + T_m\} =$$
$$= P\{S(n, Lk) - Lk \le (T_1 + \ldots + T_m - mL) + Lr\}, \qquad (27)$$

где $m = k + r$, мы, используя уже приведенные методы, убеждаемся, что случайная величина $(T_1 + \ldots + T_m - mL)$ стремится по распределению к нулю. Отметим также, что вероятности $P\{T_1 + \ldots + T_m - dL < 0\}$ стремятся к 1/2. Здесь существенен тот факт, что $L > 1$ и величины T не являются вырожденными. Далее, распределение величины $S(n, Lk) - Lk$, как это следует из теоремы 1, имеет в пределе пуассоновское распределение с параметром $\mu = (L\lambda)^2/2$. Следовательно, вероятности $P\{S(n, Lk, L) - k \le r\}$ сходятся при $n \to \infty$ к вероятностям

$$(P\{\pi_\mu < rL\} + P\{\pi_\mu \le rL\})/2.$$

Отсюда получаем утверждение теоремы 4.

Литература

1. L. E. Baum and P. Billingsley, *Asymptotic distributions for the coupon collector's problem.* — Ann. Math. Statist. **36** (1965), 1835–1839.
2. N. D. Kan and V. B. Nevzorov, *On the convergence rates to asymptotic distributions for the coupon collector's problem.* — Proceedings of the 4th St. Petersburg Workshop on Simulation (2001), 285–289.
3. T. M. Sellke, *How many iid samples does it take to see all the balls in a box?* — Ann. Appl. Probab. **5**, (1) (1995), 294–309.
4. W. Stadje, *The collector's problem with group drawings.* — Adv. in Appl. Probab. **22**, (4) (1990), 866–882.

Kan N. D., Nevzorov V. B. On some limit laws in the coupon collector's problem.

Three limit distributions are obtained for waiting times in the collector's problem with group drawings.

С.-Петербургский государственный университет

Поступило 20 сентября 2003 г.