УДК 519.224.24

В. А. Савушкин

Асимптотическое поведение обратных моментов некоторых дискретных распределений и их применение для вычисления дефектов

Рассматривается асимптотическое поведение обратных моментов для основных дискретных распределений. Доказана общая лемма, позволяющая вычислять отрицательные моменты для любой целочисленной случайной величины. Рассчитаны отрицательные моменты для разных распределений, которые дают возможность для вычисления дефектов одной статистики относительно другой статистики.

Ключевые слова: дефект, отрицательные моменты, асимптотические разложения, выборки случайного объема.

Об авторах

Савушкин Владислав Андреевич — ассистент кафедры прикладной математики и информатики Государственного университета «Дубна».

В современной математической статистике не всегда удается получить точные результаты. В связи с этим все более распространенным становится асимптотический подход. В частности, для статистик, у которых число наблюдений также случайная величина, как правило, не удается получить точных результатов. Таким образом, для статистик, основанных на выборках случайного объема, и большинства других нетривиальных статистик, как правило, применим только асимптотический подход. Несмотря на разработанность этого метода [5], остаются многие нерешенные проблемы.

Одной из важных задач является сравнение статистик. Для этого вводят понятие асимптотической относительной эффективности (АОЭ).

$$e \equiv \lim_{n\to\infty} \frac{n}{m(n)},$$

где n — число наблюдений; $m(n)$ — число наблюдений, необходимое для достижения другой статистикой такого же качества, что и первая. Как известно, во многих важных случаях значение АОЭ равно единице, т. е. такой подход оказывается слишком грубым для сравнения многих статистик [2; 4].

Поэтому используют более точный метод. А именно, для оценки «качества» статистик R_n вводят понятие дефекта этой статистики по отношению к статистике R_n^*:

$$d_n = m(n) - n.$$

То есть дефект показывает, сколько дополнительных наблюдений необходимо произвести для достижения такого же качества, что и первая статистика.

Во многих приложениях объем выборки сам является случайной величиной [3]. В таком случае для расчета дефекта необходимо знать асимптотическое поведение обратных моментов случайных величин случайного объема выборки [6]. Некоторое представление о порядке моментов случайных величин дает неравенство Иенсена вида

$$\mathrm{E}N_n^{-k} \ge \frac{1}{(\mathrm{E}N_n)^k} = \frac{1}{n^k}, \qquad k \in \mathrm{N}.$$

Кроме того, первый отрицательный момент уже нередко рассчитывался для некоторых распределений [1]. Однако для нашей задачи необходимо знать асимптотическое поведение по крайней мере первых двух отрицательных моментов. Специально для расчета любого отрицательного момента целочисленной случайной величины была получена лемма.

Лемма

Пусть неотрицательная целочисленная случайная величина N имеет производящую функцию $\psi(s) = \mathrm{E}s^N$, $|s| \le 1$, тогда k-й отрицательный момент рассчитывается по формуле

$$\mathrm{E}N^{-k} = \int_0^1 \frac{ds_1}{s_1} \int_0^{s_1} \frac{ds_2}{s_2} \cdots \int_0^{s_{k-1}} \frac{ds_k}{s_k} \psi(s_k), \ k \in \mathrm{N}.$$

Лемма легко доказывается с помощью определения момента k-го порядка и применением теоремы Фубини к полученному интегралу (k-1) раз. Применяя эту лемму к трем основным дискретным распределениям (смещенным на единицу), получаем следующие три теоремы, описывающие асимптотическое поведение первых двух моментов этих распределений.

Теорема 1

Пусть случайная величина $N = M + 1$, где M распределена биномиально, причем выполнена нормировка $\mathrm{E}N_n = n$, тогда справедливы формулы:

$$\mathrm{E}N^{-1} = \frac{1}{n} + \frac{m-1}{mn^2} + \mathrm{O}(n^{-3});$$

$$\mathrm{E}N^{-2} = \frac{1}{n^2} + \mathrm{O}(n^{-3}).$$

Доказательство

Как известно, производящая функция случайной величины N получается умножением производящей функции случайной величины M на s, поэтому имеем

$$\Psi_N(s) = \mathrm{E}s^N = s(ps+q)^n.$$

Заметим, что для выполнения нормировки, параметры биномиального распределения можно взять равными $n = m(n-1), n \ge 2$ и $p = \dfrac{1}{m}$, где $m \ge 2$ — натуральное фиксированное число. Применяя лемму и условия нормировки, получаем:

$$\mathrm{E}N^{-1} = \int_0^1 (ps+q)^n \, ds = \frac{1-q^{n+1}}{p(n+1)} = \frac{m(1-(1-m^{-1})^{m(n-1)+1}}{m(n-1)+1} = \frac{1}{n} + \frac{m-1}{mn^2} + \mathrm{O}(n^{-3}).$$

Аналогично поступаем для второго отрицательного момента:

$$\mathrm{E}N^{-2} = \int_0^1 \frac{ds_1}{s_1} \int_0^{s_1} (ps+q)^n \, ds_2 = \int_0^1 \frac{ds_1}{s_1} \frac{(ps_1+q)^{n+1} - q^{n+1}}{p(n+1)} = \frac{q^{n+1}}{p(n+1)} \int_0^{p/q} \sum_{i=0}^{n} (x+1)^i dx =$$

$$= \frac{q^{n+1}}{p(n+1)} \sum_{i=0}^{n} \frac{q^{-i-1}-1}{i+1} = \frac{1}{p(n+1)} \sum_{i=0}^{n} \frac{q^{n-i} - q^{n+1}}{i+1} = \frac{1}{p(n+1)} \sum_{i=0}^{n} \frac{q^{n-i}}{i+1} + \mathrm{O}(n^{-3}) =$$

$$= \frac{1}{p(n+1)} \sum_{i=0}^{n} \frac{q^k}{1+n-k} + \mathrm{O}(n^{-3}) = \frac{1}{p(n+1)^2} \sum_{i=0}^{n} \frac{q^k}{1 - \dfrac{k}{n+1}} + \mathrm{O}(n^{-3}) =$$

$$= \frac{1}{p(n+1)^2} \sum_{i=0}^{n} q^k \sum_{l=0}^{\infty} \left(\frac{k}{n+1} \right)^l + \mathrm{O}(n^{-3}) = \frac{1}{p(n+1)^2} \sum_{l=0}^{\infty} \frac{1}{(n+1)^l} \sum_{i=0}^{n} q^k k^l + \mathrm{O}(n^{-3}) =$$

$$= \frac{1}{pn^2} \sum_{k=0}^{n} q^k + \mathrm{O}(n^{-3}) = \frac{1}{p^2 n^2} + \mathrm{O}(n^{-3}) = \frac{1}{n^2} + \mathrm{O}(n^{-3}), \qquad n \to \infty.$$

Теорема доказана.

Теорема 2

Пусть случайная величина N=M+1, где M распределена геометрически, причем выполнена нормировка $\mathrm{E}N_n = n$, тогда справедливы формулы:

$$\mathrm{E}N^{-1} = \frac{\log n}{n} + \frac{\log n}{n^2} + \mathrm{O}\left(\frac{\log n}{n^3} \right);$$

$$\mathrm{E}N^{-2} = \frac{\pi^2}{6n} - \frac{\log n}{n^2} + \frac{1}{n^2}\left(\frac{\pi^2}{6} - 1\right) + \mathrm{O}\left(\frac{\log n}{n^3}\right).$$

Доказательство

Как известно, производящая функция случайной величины N получается умножением производящей функции случайной величины M на s, поэтому имеем:

$$\Psi_N(s) = \mathrm{E}s^N = \frac{ps}{1 - sq}.$$

Заметим, что для выполнения нормировки, параметр геометрического распределения можно взять равным $p = \frac{1}{n}$, где $n \geq 2, n \in \mathrm{N}$. Применяя лемму и условия нормировки, и раскладывая логарифм в ряд получаем:

$$\mathrm{E}N^{-1} = \int_0^1 \frac{p}{1 - sq} ds = -\frac{p}{q} \log p =$$

$$= \frac{\log n}{n - 1} = \frac{\log n}{n} + \frac{\log n}{n^2} + \mathrm{O}\left(\frac{\log n}{n^3}\right).$$

Аналогично поступаем для второго отрицательного момента:

$$\mathrm{E}N^{-2} = \int_0^1 \frac{ds_1}{s_1} \int_0^{s_1} \frac{p}{1 - s_2 q} ds_2 = -\frac{p}{q} \int_0^1 \frac{\log(1 - q s_1)}{s_1} ds_1 = -\frac{p}{q} \int_p^1 \frac{\log t}{1 - t} dt = -\frac{p}{q} \sum_{k=0}^{\infty} \int_p^1 \log t \cdot t^k dt =$$

$$= -\frac{p}{q} \sum_{k=0}^{\infty} \frac{1}{k+1} \int_p^1 \log t \cdot dt^{k+1} = \frac{p}{q} \sum_{k=0}^{\infty} \frac{1}{k+1} \left(p^{k+1} \log p + \frac{1}{k+1} - \frac{p^{k+1}}{k+1} \right) =$$

$$= \frac{p}{q} \left(\sum_{k=1}^{\infty} \frac{1}{k^2} + p \log p - p \right) + \mathrm{O}\left(p^3 \log p\right) = \frac{p}{q} \left(\frac{\pi^2}{6} + p \log p - p \right) + \mathrm{O}\left(p^3 \log p\right) =$$

$$= \frac{\pi^2 p}{6} + p^2 \left(\frac{\pi^2}{6} - 1 \right) + p^2 \log p + \mathrm{O}\left(p^3 \log p\right) =$$

$$= \frac{\pi^2}{6n} - \frac{\log n}{n^2} + \frac{1}{n^2}\left(\frac{\pi^2}{6} - 1\right) + \mathrm{O}\left(\frac{\log n}{n^3}\right), \qquad n \to \infty.$$

Теорема доказана.

Теорема 3

Пусть случайная величина $N = M + 1$, где M распределена пуассоновски, причем выполнена нормировка $\mathrm{E}N_n = n$, тогда справедливы формулы:

$$\mathrm{E}N^{-1} = \frac{1}{n} + \frac{1}{n^2} + \mathrm{O}\left(\frac{1}{n^3}\right);$$

$$\mathrm{E}N^{-2} = \frac{1}{n^2} + \mathrm{O}\left(\frac{1}{n^3}\right).$$

Доказательство

Как известно, производящая функция случайной величины N получается умножением производящей функции случайной величины M на s, поэтому имеем:

$$\Psi_N(s) = \mathrm{E}s^N = \frac{e^s - 1}{s}.$$

Заметим, что для выполнения нормировки параметр пуассоновского распределения можно взять равным $\lambda = n - 1$, где $n \geq 2, n \in N$. Применяя лемму и условия нормировки и раскладывая экспоненту в ряд, получаем:

$$\mathrm{E}N^{-1} = \int_0^1 e^{\lambda(s-1)} ds =$$

$$= \frac{1 - e^{-\lambda}}{\lambda} = \frac{1}{n} + \frac{1}{n^2} + \mathrm{O}\left(\frac{1}{n^3}\right).$$

Аналогично поступаем для второго отрицательного момента, применяя интегрирование по частям:

$$\mathrm{E}N^{-2} = \int_0^1 \frac{ds_1}{s_1}\int_0^{s_1} e^{\lambda(s_2-1)}ds_2 = \frac{e^{-\lambda}}{\lambda}\int_0^{\lambda}\frac{e^s-1}{s}ds = \frac{e^{-\lambda}}{\lambda}\int_0^1\frac{e^{\lambda t}-1}{t}dt = \frac{e^{-\lambda}}{\lambda}\sum_{k=0}^{\infty}\int_0^1\left(e^{\lambda t}-1\right)(1-t)^k dt =$$

$$= e^{-\lambda}\sum_{k=0}^{\infty}\frac{1}{k+1}\int_0^1 e^{\lambda t}(1-t)^{k+1}dt = \frac{1}{\lambda}\sum_{k=1}^{\infty}\frac{1}{k\lambda^k}\int_0^{\lambda}e^{-x}x^k dx = \frac{1}{n^2}+\mathrm{O}\left(\frac{1}{n^3}\right), \qquad n\to\infty.$$

Теорема доказана.

Полученные результаты позволяют легко рассчитывать асимптотические дефекты [2]. Например, при нормировке из теоремы 1 асимптотический дефект статистики, основанной на выборке биномиального объема относительно статистики неслучайного объема, имеет вид:

$$d(\theta) = \frac{m-1}{m}.$$

Аналогично, учитывая нормировку теоремы 3 для статистики, основанной на пуассоновской выборке, дефект имеет вид:

$$d(\theta) = 1.$$

Аналогично можно рассчитать асимптотический дефект критерия. Например, если рассмотреть аналог критерия Стьюдента для проверки статистических гипотез для случая выборок случайного объема, то получим, что асимптотический дефект статистики пуассоновского объема равен:

$$d_n = \frac{u_{1-\alpha}^4}{16n^2} + o\left(n^{-2}\right),$$

статистики биномиального объема равен:

$$d_n = \frac{u_{1-\alpha}^4(m-1)}{16mn^2} + o\left(n^{-2}\right),$$

и статистики геометрического объема равен:

$$d_n = \frac{u_{1-\alpha}^4}{16}\left(\frac{\log n}{n-1}-\frac{1}{n}\right) + o\left(\frac{\log n}{n}\right).$$

Библиографический список

1. Бенинг, В. Е. Асимптотические разложения для функций распределения статистик, построенных по выборкам случайного объёма / В.Е. Бенинг, Н.К. Галиева, В.Ю. Королев // Информатика и её применения. — 2013. — Т. 7, № 2. — С. 75—83.
2. Крамер, Г. Математические методы статистики / Г. Крамер. — Москва : Мир, 1976. — 648 с.
3. Королев, В. Ю. Математические основы теории риска / В.Ю. Королев, В.Е. Бенинг, С.Я. Шоргин. — Москва : ФизМатЛит, 2011. — 591 с.
4. Леман, Э. Теория точечного оценивания / Э. Леман. — Москва : Наука : ФизМатЛит, 1991. — 444 с.
5. Bening, V. E. Asymptotic Theory of Testing Statistical Hypotheses: Efficient Statistics Optimality, Power Loss, and Deficiency // VSP, Utrecht. — 2000. — 277 p.
6. Hodges, J. L. Deficiency / J.L. Hodges, E.L. Lehmann // Ann. Math. Statist. — 1970. — V. 41, N 5. — P. 783—801.

Поступила в редакцию 22.12.15