

УДК 517

С. П. Яковлев

О структуре гипергеометрических распределений

(Представлено академиком В.С. Захаряном 1/X 2008)

Ключевые слова: гипергеометрическое распределение, регулярное распределение, выпуклость, вогнутость

1. Введение. Семейство регулярных гипергеометрических распределений $\{p_n\}$

$$p_0 = \frac{\Gamma(q - \tilde{p}_1)\Gamma(q - \tilde{p}_2)}{\Gamma(q - \tilde{p}_1 - \tilde{p}_2)\Gamma(q)}, \quad p_n = p_0 \prod_{k=0}^{n-1} \frac{(\tilde{p}_1 + k)(\tilde{p}_2 + k)}{(1 + k)(q + k)}, \quad n = 1, 2, \dots, \quad (1)$$

где $\Gamma(\cdot)$ — гамма-функция Эйлера, введено в [1], с. 161-173 (см. также [2]).
 Найдены: производящая функция, факториальные моменты, доказано правильное изменение с показателем $(-\rho)$, где

$$\rho = q + 1 - \tilde{p}_1 - \tilde{p}_2 \in (1, +\infty); \quad (2)$$

установлено, что $\{p_n\}$ убывает, лог-вогнуто и унимодально с модой p_0 при

$$\min(\tilde{p}_1, \tilde{p}_2) < 1. \quad (3)$$

Свойства монотонности и лог-выпуклости/лог-вогнутости $\{p_n\}$ при $\tilde{p}_1 > 1$, $\tilde{p}_2 > 1$ обсуждены в [3].

Определение 1. Пусть $\{\epsilon_n\}$ — последовательность положительных чисел.

(1) $\{\epsilon_n\}$ выпукла (вогнута) для индекса n , если $\epsilon_{n+2} - \epsilon_{n+1} < (>) \epsilon_{n+1} - \epsilon_n$.

(2) $\{\epsilon_n\}$ лог-выпукла (лог-вогнута) для индекса n , если $(\epsilon_n/\epsilon_{n+1}) < (>) (\epsilon_{n+1}/\epsilon_{n+2})$.

Лемма 1. На массиве индексов $n = \underline{n}, \underline{n} + 1, \dots, \bar{n}$ выпуклость $\{\epsilon_n\}$ влечет лог-выпуклость $\{\epsilon_n\}$, лог-вогнутость $\{\epsilon_n\}$ влечет вогнутость $\{\epsilon_n\}$.

Доказательство. $\{\epsilon_n\}$ состоит из не более чем счетного числа монотонных кусков. Без ограничения общности считаем, что на каждом куске (массиве индексов) $\{\epsilon_n\}$ строго монотонно. Докажем лемму 1 на массиве с возрастающей $\{\epsilon_n\}$. Случай убывания $\{\epsilon_n\}$ доказывается аналогично.

Если на заданном массиве $\{\epsilon_n\}$ выпукла, то

$$(\epsilon_{n+1}/\epsilon_n) > 1, \quad \frac{\epsilon_{n+1}}{\epsilon_n} \left(\frac{\epsilon_{n+2}}{\epsilon_{n+1}} - 1 \right) < \frac{\epsilon_{n+1}}{\epsilon_n} - 1. \quad (4)$$

Допустим, что $\{\epsilon_n\}$ для n_0 из массива не является лог-выпуклой, т.е.

$$(\epsilon_{n_0+1}/\epsilon_{n_0}) - 1 \leq (\epsilon_{n_0+2}/\epsilon_{n_0+1}) - 1. \quad (5)$$

Из первого неравенства в (4) для n_0 и из (5) получаем

$$\frac{\epsilon_{n_0+1}}{\epsilon_{n_0}} \left(\frac{\epsilon_{n_0+2}}{\epsilon_{n_0+1}} - 1 \right) > \frac{\epsilon_{n_0+1}}{\epsilon_{n_0}} - 1,$$

что противоречит второму неравенству в (4).

Если на заданном массиве $\{\epsilon_n\}$ лог-вогнута, то $(\epsilon_{n+1}/\epsilon_n) > 1$ и $(\epsilon_{n+2}/\epsilon_{n+1}) - 1 > (\epsilon_{n+1}/\epsilon_n) - 1$, что после перемножения неравенств дает $\epsilon_{n+2} - \epsilon_{n+1} > \epsilon_{n+1} - \epsilon_n$.

По лемме 1, существование пары «выпуклость и лог-вогнутость» для индекса n невозможно. Возможность сосуществования пары «вогнутость и лог-выпуклость» доказана в [4], с. 61-62.

По лемме 1, $\{p_n\}$ при условии (3) будучи лог-вогнутым вогнутс.

Цель настоящей статьи заключается в изучении свойств монотонности и двух типов выпуклости/вогнутости распределений (1). В силу вышесказанного изучению подлежит случай $\hat{p}_1 > 1, \hat{p}_2 > 1$. При этом найдется натуральное $n_0 = 0, 1, 2, \dots$ такое, что

$$n_0\rho < (\hat{p}_1 - 1)(\hat{p}_2 - 1) \leq (n_0 + 1)\rho. \quad (6)$$

2. Монотонность. Число n_0 в (6) имеет определенный смысл.

Теорема 1. 1. Если $(\hat{p}_1 - 1)(\hat{p}_2 - 1) < \rho$, то $\{p_n\}$ убывает.

2. Если выполнено (6) со строгим неравенством, с $n_0 > 0$, то $\{p_n\}$ убывает при $n = n_0, n_0 + 1, \dots$, возрастает при $n = 0, 1, \dots, n_0$.

Доказательство основано на формуле (см. (1))

$$p_{n+1} = p_n \tau_n, \quad \text{где} \quad \tau_n = \frac{(\hat{p}_1 + n)(\hat{p}_2 + n)}{(1 + n)(q + n)}, \quad n = 0, 1, 2, \dots \quad (7)$$

В силу (2) эквивалентны неравенства

$$\hat{p}_1 \cdot \hat{p}_2 > (<)(=)q \quad \text{и} \quad (\hat{p}_1 - 1)(\hat{p}_2 - 1) > (<)(=)\rho. \quad (8)$$

С помощью (7) легко проверяется неравенство

$$\frac{p_{n+1}}{p_n} = 1 - \frac{\rho}{q+n} + \tau'_n, \quad n = 0, 1, 2, \dots, \quad (9)$$

где $\tau'_n = \frac{(\bar{p}_1-1)(\bar{p}_2-1)}{(1+n)(q+n)}$.

Случай $\bar{p}_1 = 1$ (или $\bar{p}_2 = 1$) с заменой q на $q+1$ приводит к изученному распределению Уорринга (см. [1]) и исключается из рассмотрения.

1. В силу (8) $0 > \bar{p}_1\bar{p}_2 - q - (q+1 - \bar{p}_1 - \bar{p}_2) = (\bar{p}_1+n)(\bar{p}_2+n) - (1+n)(q+n)$ при $n = 0, 1, 2, \dots$, или $\tau_n < 1$. Это из-за (7) означает $p_0 > p_1 > p_2 > \dots$.

2. Запишем (9) в виде $\frac{p_{n+1}}{p_n} = 1 - \frac{1}{1+n} \frac{1}{q+n} \{ (1+n)\rho - (\bar{p}_1-1)(\bar{p}_2-1) \}$, что, по (6), влечет $\frac{p_{n+1}}{p_n} \begin{cases} > 1 & \text{при } n = 0, 1, \dots, n_0-1, \\ < 1 & \text{при } n = n_0, n_0+1, \dots \end{cases}$.

Замечание 1. В случае равенства справа в (6) с $n_0 \geq 0$ утверждения теоремы 1 сохраняются, разве лишь $p_{n_0} = p_{n_0+1}$.

3. **Лог-выпуклость/лог-вогнутость.** Вначале рассмотрению подлежит случай

$$e \stackrel{\text{def}}{=} \frac{1}{\rho}(\bar{p}_1-1)(\bar{p}_2-1) - 1 \leq 0, \quad \bar{p}_1 > 1, \quad \bar{p}_2 > 1. \quad (10)$$

Покажем, что при условии (10): $\{p_n\}$ либо лог-вогнуто; либо найдется индекс $n_1 > 0$ такой, что $\{p_n\}$ лог-выпукло при $n = 0, 1, \dots, n_1-1$, лог-вогнуто при $n = n_1, n_1+1, \dots$.

В силу (9) последовательность $\{\tau'_n\}$ состоит из положительных членов. Основным член формулы (9) — последовательность $\{1 - (\rho/(q+n))\}$ возрастает и вогнута, а $\{\tau'_n\}$ убывает и $\lim_{n \rightarrow +\infty} \tau'_n = 0$. Очевидно «непродолжительное» воздействие $\{\tau'_n\}$ на $\{p_{n+1}/p_n\}$. Этим и обосновано наличие не более чем двух кусков в графика последовательности $\{p_{n+1}/p_n\}$, последний из которых является выпуклым.

Условие (10) эквивалентно условию $\bar{p}_1, \bar{p}_2 < q, \bar{p}_1 > 1, \bar{p}_2 > 1$, что следует из (8). Согласно (7) $\tau_0 = 1 - 2\frac{\bar{p}_1\bar{p}_2}{q} + \frac{\bar{p}_1\bar{p}_2(\bar{p}_1+1)(\bar{p}_2+1)}{2(q+1)}$. По этой формуле возможно показать, что все ситуации, связанные с графиком $\{p_{n+1}/p_n\}$, могут осуществиться, т.е. $\tau_0 > 0, \tau_0 < 0, \tau_0 = 0$.

Примеры. Пусть $\bar{p}_1 = 1, \bar{p}_2 = 2, q = 4, \rho = 1.5$. Тогда $\tau_0 = \frac{2.5}{40} > 0$.

Пусть $\bar{p}_1 = 1.5, \bar{p}_2 = 4, q = 7, \rho = 2.5$. Тогда $\tau_0 = -\frac{2.5}{56} < 0$.

Пусть $\bar{p}_1 = 2, \bar{p}_2 = 2, q = 5, \rho = 2$. Тогда $\tau_0 = 0$.

Так как τ_0 — непрерывная относительно $\bar{p}_1$ функция, то в первом примере можно взять $\bar{p}_1$ «незначительно» больше единицы и получить $\tau_0 > 0$.

Ниже мы имеем дело с условием

$$\bar{p}_1 > 1, \quad \bar{p}_2 > 1, \quad (11)$$

выполненным (6) с $n_0 \geq 0$.

Пусть $n_0 > 0$. Согласно (6) и (11), $n_0 - 1 < e < n_0$. Обозначим

$$f = (e + 1)(\rho + \bar{p}_1 + \bar{p}_2), \quad t_1 = e + \sqrt{e^2 + f - 1}, \quad \bar{n}_1 = [t_1], \quad (12)$$

где $[x]$ есть целая часть числа $x > 0$. Из (10) и (11) легко выводится, что $\bar{n}_1 \geq n_0$. Равенство может иметь место лишь при $n_0 = 1$.

Теорема 2. При условии (11) с $n_0 > 0$ $\{p_n\}$ лог-выпукло при $n = 0, 1, \dots, n_1$, лог-вогнуто при $n = n_1 + 1, n_1 + 2, \dots$, где n_1 равно $\bar{n}_1 - 1$ или $\bar{n}_1$, а $\bar{n}_1$ дано в (12).

Доказательство. Мы покажем, что $\{p_n\}$ лог-выпукло при $n = 0, 1, \dots, \bar{n}_1 - 1$, лог-вогнуто при $n = \bar{n}_1 + 1, \bar{n}_1 + 2, \dots$. Линейный случай интерпретируем либо как лог-выпуклый, либо как лог-вогнутый при $p_{\bar{n}_1} = p_{\bar{n}_1+1}$ для индекса $\bar{n}_1$.

Введем следующий непрерывный аналог последовательности:

$$g(t) = 1 - \frac{\rho}{q+1} + \frac{(\bar{p}_1 - 1)(\bar{p}_2 - 1)}{(1+t)(q+t)}, \quad t \in [0, +\infty). \quad (13)$$

Дифференцируя $g(t)$ и приравнявая к нулю, получаем квадратное уравнение $t^2 - 2et - (f - 1) = 0$, из корней которого лишь t_1 (см. 12) положителен.

Проверяется, что

$$g'(0) = \frac{1}{q^2} \{ \rho - (q+1)(\bar{p}_1 - 1)(\bar{p}_2 - 1) \} \leq \frac{1}{q^2} \{ \rho - (q+1)\rho \} = -\frac{\rho}{q} < 0,$$

где использовано первое неравенство в (10). Следовательно, $g(t)$ убывает в $[0, t_1)$, возрастает в $(t_1, +\infty)$. Тогда $\{p_{n+1}/p_n\}$ убывает при $n = 0, 1, \dots, n_1 - 1$; лог-вогнуто при $n = \bar{n}_1 + 1, \bar{n}_1 + 2, \dots$. Значит, $\{p_n\}$ лог-выпукло при $n = 0, 1, \dots, n_1 - 1$, лог-вогнуто при $n = \bar{n}_1 + 1, \bar{n}_1 + 2, \dots$.

Осталось рассмотреть случай индекса $\bar{n}_1$. Проверяем значения $p_{\bar{n}_1+1}/p_{\bar{n}_1}$ и $p_{\bar{n}_1}/p_{\bar{n}_1+1}$. Если первое больше второго, то $\{p_n\}$ лог-выпукло для индекса $\bar{n}_1$. В противном случае $\{p_n\}$ лог-вогнуто для индекса $\bar{n}_1$. Оба случая не меняют структуру графика $\{\log p_n\}$. Он образован из двух кусков.

4. Вспомогательные леммы. В силу теоремы 2 и леммы 1 $\{p_n\}$ вогнуто для индексов $n = n_1 + 1, n_1 + 2, \dots$. Так что проблему выпуклости/вогнутости $\{p_n\}$ осталось решить для индексов $n = 0, 1, \dots, n_1$ при условии (11).

Обозначим

$$h_n = \frac{1}{q+n} \left\{ \frac{(\bar{p}_1 - 1)(\bar{p}_2 - 1)}{1+n} - \rho \right\}. \quad (14)$$

При $n_0 \geq 2$ из (14) имеем $h_n > \frac{(n_0 - n - 1)\rho}{(1+n)(q+n)} > 0$ для $n = 0, 1, \dots, n_0 - 2$.

Далее, пусть

$$\bar{h}_n = -h_n \quad \text{при} \quad n = n_0 + 1, \dots \quad (15)$$

Тогда $h_n > \frac{(n - n_0)\rho}{(1 + n)(q + n)} \geq 0$ при $n = n_0, n_0 + 1, \dots$.

Лемма 2. Пусть выполнено условие (11). Тогда

$$\begin{cases} \frac{p_{n+2} - p_{n+1}}{p_{n+1} - p_n} = h_{n+1} + \frac{h_{n+1}}{h_n}, & n = 0, 1, \dots, n_0 - 2, n_0 \geq 2, \\ \frac{p_{n+1} - p_{n+2}}{p_n - p_{n+1}} = -\hat{h}_{n+1} + \frac{\hat{h}_{n+1}}{\hat{h}_n}, & n = n_0, n_0 + 1, \dots \end{cases} \quad (16)$$

Доказательство. Произведем выкладки с учетом (7)

$$p_{n+1} - p_n = p_n \left\{ \frac{(\hat{p}_1 + n)(\hat{p}_2 + n)}{(1 + n)(q + n)} - 1 \right\} = p_n \frac{1}{q + n} \left\{ \frac{(\hat{p}_1 - 1)(\hat{p}_2 - 1)}{1 + n} - \rho \right\}, \quad n = 0, 1, 2, \dots$$

Используя (14) и (15), получаем

$$\begin{cases} \frac{p_{n+2} - p_{n+1}}{p_{n+1} - p_n} = \frac{p_{n+1}}{p_n} \frac{h_{n+1}}{h_n}, & n = 0, 1, \dots, n_0 - 2, n_0 \geq 2, \\ \frac{p_{n+1} - p_{n+2}}{p_n - p_{n+1}} = \frac{p_{n+1}}{p_n} \frac{\hat{h}_{n+1}}{\hat{h}_n}, & n = n_0, n_0 + 1, \dots \end{cases} \quad (17)$$

Сравнивая $\{p_{n+1}/p_n\}$ из (9) и $\{h_n\}$, $\{\hat{h}_n\}$ из (14) и (15), заключаем, что $(p_{n+1}/p_n) = 1 + h_n$ при $n = 0, 1, \dots, n_0 - 2, n_0 \geq 2$ и $(p_{n+1}/p_n) = 1 - \hat{h}_n$ при $n = n_0, n_0 + 1, \dots, n_1$.

(Отсюда и из (17) следует (16).)

Лемма 3. Пусть выполнено условие (11). Тогда $\{h_n\}$ убывает и вогнуто при $n = 0, 1, \dots, n_0 - 2, n_0 \geq 2$, $\{\hat{h}_n\}$ возрастает и выпукло при $n = n_0, n_0 + 1, \dots, n_1 - 1$.

Доказательство. Рассмотрим непрерывные аналоги $h(t) = g(t) - 1$ и $\hat{h}(t) = 1 - g(\cdot)$, $t \in [0, +\infty)$ последовательностей $\{h_n\}$ и $\{\hat{h}_n\}$, где $g(t)$ задано формулой (13).

Из свойств $g(t)$ следует, что: $h(t)$ убывает на $[0, n_0 - 1]$, $n_0 > 0$, $\hat{h}(t)$ возрастает на $[n_0, n_1]$. Далее, легко видеть, что

$$h'(t) = -\frac{1}{q + t} \left\{ h(t) + \frac{(\bar{p}_1 - 1)(\bar{p}_2 - 1)}{(1 + t)^2} \right\}, \quad \hat{h}'(t) = \frac{1}{q + t} \left\{ -\hat{h}(t) + \frac{(\hat{p}_1 - 1)(\hat{p}_2 - 1)}{(1 - t)^2} \right\}.$$

Функции $(1/(q + t))$ и $(\bar{p}_1 - 1)(\bar{p}_2 - 1)/(1 + t)^2$ убывают на $[0, +\infty)$. Поэтому с учетом монотонности $h(t)$ и $\hat{h}(t)$ заключаем, что: $h'(t)$ возрастает на $[0, n_0 - 1]$, $n_0 > 0$, $\hat{h}'(t)$ убывает на $[n_0, n_1]$. Из вышесказанного вытекает лемма 3.

Из лемм 2 и 3 вытекает

Следствие 1. Пусть выполнено условие (11). Тогда: для выпуклости (вогнутости) $\{p_n\}$ при $n = 0, 1, \dots, n_0 - 2, n_0 \geq 2$ необходимо и достаточно

$$\frac{1}{h_{n+1}} - \frac{1}{h_n} > (<) 1; \quad (18)$$

для выпуклости (вогнутости) $\{p_n\}$ при $n = n_0, n_0 + 1, \dots, n_1 - 1$ необходимо и достаточно

$$\frac{1}{h_n} - \frac{1}{h_{n+1}} < (>) 1. \quad (19)$$

5. **Выпуклость/вогнутость.** Сформулируем основной результат работы.

Теорема 3. Пусть выполнено условие (11).

(а) Если $n_0 = 0$, то либо $\{p_n\}$ вогнуто, либо найдется n_2 , $n_0 < n_2 \leq n_1$ такое, что $\{p_n\}$ выпукло при $n = 0, 1, \dots, n_2 - 1$, вогнуто при $n = n_2, n_2 + 1, \dots$.

(б) Если $n_0 = 1$, то имеет место вторая ситуация в (а).

(в) Если $n_0 \geq 2$, то либо осуществляется вторая ситуация в (а), либо найдутся индексы n_2 , $0 \leq n_2 \leq n_0 - 2$ и n_3 , $n_3 \geq n_2 + 1$ такие, что $\{p_n\}$ вогнуто при $n = 0, 1, \dots, n_2$, выпукло при $n = n_2 + 1, n_2 + 2, \dots, n_3 - 1$, вогнуто при $n = n_3, n_3 + 1, \dots$.

Таким образом, график $\{p_n\}$ состоит из не более чем трех выпуклых/вогнутых кусков.

Доказательство. Из условия (6), обозначения (10) имеем

$$\begin{cases} e - n > 0 & \text{при } n = 0, 1, \dots, n_0 - 1, n_0 > 0, \\ e - n < 0 & \text{при } n = n_0, n_0 + 1, \dots \end{cases} \quad (20)$$

Нетрудно вычислениями проверить, что неравенство (18) эквивалентно неравенству

$$\rho < (>) I_n \stackrel{\text{def}}{=} \frac{(1+n)(q+n)}{(e-(1+n))(e-n)} + \frac{q+2n+1}{e-(1+n)}, \quad (21)$$

а неравенство (19) — неравенству

$$\rho > (<) I_n = \frac{1}{(1+n)-e} - 1 + \frac{(e+1)(e+q)}{((1+n)-e)(n-e)}, \quad (22)$$

где I_n определено в (21). По (20), оба слагаемых в правой части (21) при $n = 0, 1, \dots, n_0 - 2, n_0 \geq 2$ положительны. Они возрастают по n . Поэтому

$$I_0 - \rho < I_1 - \rho < \dots < I_{n_0-2} - \rho. \quad (23)$$

Аналогично из (20) и (22) следует, что

$$I_{n_0} - \rho > I_{n_0+1} - \rho > I_{n_0+2} - \rho > \dots \quad (24)$$

В силу (23) и (24) эти последовательности могут менять знак не более чем один раз. Рассмотрение в отдельности возможных ситуаций с учетом следствия 1 приводит к теореме 3. Разве лишь проверим, что при $n_0 \geq$

2 упомянутые в теореме 3 все ситуации для $n = 0, 1, \dots, n_2 - 1$ могут осуществиться.

Действительно, в силу (1) справедливо равенство

$$\frac{p_2 - p_1}{p_1 - p_0} = \{(\hat{p}_1 - 1)(\hat{p}_2 - 1) - 2\rho\} \frac{3q + 1}{2q(q + 1)} + \frac{\rho}{q}. \quad (25)$$

Примеры. Пусть $\hat{p}_1 = 4$, $\hat{p}_2 = 4$, $\rho = 2$, $q = 9$. Условие (6) выполнено с $n_0 = 4$:

$$\frac{p_2 - p_1}{p_1 - p_0} = 1 + \frac{35}{36} > 1.$$

Пусть $\hat{p}_1 = 7$, $\hat{p}_2 = 2$, $\rho = 2$, $q = 10$. Условие (6) выполнено с $n_0 = 3$:

$$\frac{p_2 - p_1}{p_1 - p_0} = \frac{27}{55} < 1.$$

Пусть $\hat{p}_1 = 4 + \sqrt{61/31}$, $\hat{p}_2 = 4 - \sqrt{61/31}$, $\rho = 3$, $q = 10$. Условие (6) выполнено с $n_0 = 2$:

$$\frac{p_2 - p_1}{p_1 - p_0} = 1.$$

Ереванский государственный университет

С. П. Яковлев

О структуре гипергеометрических распределений

Рассмотрено семейство гипергеометрических распределений. Изучены монотонность, лог-выпуклость/вогнутость и выпуклость/вогнутость в обычном смысле.

Ս. Պ. Յակովլև

Տիպերերկրաչափական բաշխումների կառուցվածքի մասին

Աշխատանքում դիտարկված է ռեգուլյար հիպերերկրաչափական բաշխումների ընտանիքը: Ուսումնասիրված են ընտանիքի բաշխումների մոնոտոնությունը, լոգ-ուռուցիկությունը դեպի վերև/ներքև, ուռուցիկությունը դեպի վերև/ներքև:

On the Structure of Hypergeometric Distributions

The family of Regular Hypergeometric Distributions is considered. The properties of monotonicity, log-downward/log-upward convexity, downward/upward convexity of such distributions are deeply investigated.

Литература

1. *Astola J., Danielian E.* - Tampere: TICSP Series, 2006. #31. 251p.
2. *Danielian E., Astola J.* - Tampere: TICSP Series. 2006. #34. P. 127-132.
3. Арскелян А.Г. Устойчивость частотных распределений в биомолекулярных моделях. Автореф. канд. дис. Ереван. ЕГУ. 2007.
4. *Astola J., Danielian E.* - Tampere: TICSP Series. 2004. #27. 94p.