Закон Ципфа–Парето и процесс Юла

Плетнев Никита Вячеславович

Московский физико-технический институт
Факультет управления и прикладной математики

23 декабря 2020 г.

Постановка задачи

Исследуемая система

- Неограниченно много жителей, которые собирают монеты. Обозначение: $\overrightarrow{n} = (n_1, n_2, \ldots)^T$, где n_s — это количество жителей, имеющих по s монет.
- c_s — это доля жителей, имеющих ровно s монет (среди имеющих монеты, поскольку n_0 по условию не ограничено).
- На первом шаге $(t = 1)$ монету получает один житель: $n_1(1) = 1$.
- На каждом следующем шаге одну монетку получает новый житель, а другую — один из старых. С каждым элементом вектора $\overrightarrow{n}$ в момент времени t сопоставим случайную величину:

$$\xi_{s,t} = \begin{cases} 1, & \text{если монету получил тот, у кого было } s, \\ 0 & \text{иначе.} \end{cases}$$

Постановка задачи

Исследуемая система

- Тогда можно выписать формулы перехода:

$$\begin{cases} n_1(t+1) = n_1(t) + 1 - \xi_{1,t}, \\ n_s(t+1) = n_s(t) + \xi_{s-1,t} - \xi_{s,t}, \quad s \geq 2 \end{cases}$$

- Из условия следует, что $\sum\limits_{k \geq 1} n_k(t) = t$, $\sum\limits_{k \geq 1} k n_k(t) = 2t - 1$.
- Очевидно, в момент времени t происходит ровно одно из событий "$\xi_{s,t} = 1$".
- Вероятности событий:
  $P(uniform) = \alpha$, $P(preferential) = 1 - \alpha$;
  $P(\xi_{s,t} = 1 | uniform) = \frac{n_s(t)}{t}$; $P(\xi_{s,t} = 1 | preferential) = \frac{s n_s(t)}{2t-1}$;

$$P(\xi_{s,t} = 1) = \alpha \frac{n_s(t)}{t} + (1 - \alpha) \frac{s n_s(t)}{2t - 1}.$$

Постановка задачи

Цели исследования

- Найти предельную форму $\overrightarrow{c}$;
- Установить скорость сходимости распределения к пределу.

Используемые методы

- Прямое вычисление в предположении «среднего поля»: $\overrightarrow{n}(t) = \mathbb{E}\overrightarrow{n(t)}$
- Проведение серии экспериментов с усреднением.

Приближение «среднего поля»

Случайные величины заменяются на свои мат.ожидания.

$$\mathbb{E}\xi_{s,t} = P(\xi_{s,t} = 1) = \alpha\frac{n_s(t)}{t} + (1-\alpha)\frac{sn_s(t)}{2t-1}.$$

Теоретическое решение

Формулы перехода в приближении «среднего поля»

$$\begin{cases} n_1(t+1) = n_1(t) + 1 - \frac{\alpha n_1(t)}{t} - \frac{(1-\alpha) n_1(t)}{2t-1}; \\ \ldots; \\ n_s(t+1) = n_s(t) + \frac{\alpha n_{s-1}(t)}{t} + \frac{(1-\alpha)(s-1) n_{s-1}(t)}{2t-1} - \frac{\alpha n_s(t)}{t} - \frac{(1-\alpha) s n_s(t)}{2t-1}. \end{cases}$$

Обозначения для сокращения записи: $\beta_t := \frac{\alpha}{t}$, $\gamma_t := \frac{1-\alpha}{2t-1}$

Краткая запись формул перехода

$$\begin{cases} n_1(t+1) = 1 + (1 - \beta_t - \gamma_t) n_1(t); \\ \ldots; \\ n_s(t+1) = (\beta_t + (s-1)\gamma_t) n_{s-1}(t) + (1 - \beta_t - s\gamma_t) n_s(t). \end{cases}$$

Теоретическое решение

Формулы перехода в матричной форме

$$\overrightarrow{n}(t+1) = \overrightarrow{e_1} + A_t \overrightarrow{n}(t),$$

где $\overrightarrow{e_1} = (1, 0, \ldots, 0, \ldots)^T$,

$$A_t = \begin{pmatrix} 1-\beta_t-\gamma_t & 0 & 0 & \ldots \\ \beta_t+\gamma_t & 1-\beta_t-2\gamma_t & 0 & \ldots \\ 0 & \beta_t+2\gamma_t & 1-\beta_t-3\gamma_t & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Те же формулы для долей

$$\overrightarrow{c}(t+1) = \frac{1}{t+1}\overrightarrow{e_1} + \frac{t}{t+1} A_t \overrightarrow{c}(t);$$

$$\overrightarrow{c}(t+1) = \frac{1}{t+1}\left(E + A_t + A_t A_{t-1} + \ldots + A_t A_{t-1} \ldots A_1\right)\overrightarrow{e_1}.$$

Вопрос о существовании предела при $t \to \infty$.

Теоретическое решение

Существование предела

$$\overrightarrow{c}(t+1) = \frac{1}{t+1}\left(E + A_t + A_t A_{t-1} + \ldots + A_t A_{t-1} \ldots A_1\right)\overrightarrow{e_1}.$$

Наибольшее собственное значение матрицы A_t — это $1 - \beta_t - \gamma_t$, монотонно возрастает и стремится к 1. Поэтому операторные нормы всех слагаемых не превосходят 1, и это выражение — ограниченное.

Более того, на каждом шаге t оператор $\frac{t}{t+1}A_t$ имеет норму меньше 1.

Если есть стационарная точка, то она и является пределом данной последовательности.

Поэтому просто найдём такую точку.

Теоретическое решение

Поиск стационарной точки

Если доли людей с каждым числом монет не меняются, то $n_s(t) = c_s t$. Подставим в формулу перехода:
$c_s = \alpha(c_{s-1} - c_s) + (1-\alpha)\frac{t}{2t-1}((s-1)c_{s-1} - sc_s)$.
При устремлении t к бесконечности дробь превращается в $\frac{1}{2}$.
$c_s(1 + \alpha + \frac{1-\alpha}{2}s) = c_{s-1}(\alpha + \frac{1-\alpha}{2}(s-1))$;
$\frac{c_s}{c_{s-1}} = \frac{s+\frac{3\alpha-1}{1-\alpha}}{s+\frac{2\alpha+2}{1-\alpha}} = 1 - \frac{\frac{3-\alpha}{1-\alpha}}{s+\frac{2\alpha+2}{1-\alpha}}$.

Результат

$$c_s = c_1 \prod_{i=2}^{s} \left(1 - \frac{\frac{3-\alpha}{1-\alpha}}{s + \frac{2\alpha+2}{1-\alpha}}\right),$$

c_1 определяется из условия нормировки $\sum\limits_{s=1}^{\infty} = 1$.

Теоретическое решение

Приближённое решение

Точное вычисление данного произведения не представляется возможным, но для некоторого достаточно большого s_0 можно выписать:

$$\ln c_s = \ln C + \sum_{i=s_0}^{s} \ln\left(1 - \frac{\frac{3-\alpha}{1-\alpha}}{s+\frac{2\alpha+2}{1-\alpha}}\right) = \ln B + o(1) - k\sum_{i=s_0}^{s}\frac{1}{s+c},$$

где $c = \frac{2\alpha+2}{1-\alpha}$, $k = \frac{3-\alpha}{1-\alpha}$.

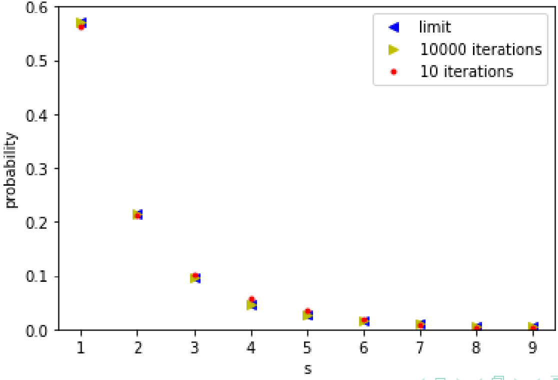
При больших s сумма приближается интегралом, поэтому $\ln c_s = \ln A - k\ln(s+c) + o(1)$, то есть $c_s \sim A(s+c)^{-k}$.

Результат

При достаточно больших s предельное распределение приблизительно равно $c_s^* \sim s^{-\frac{3-\alpha}{1-\alpha}}$.

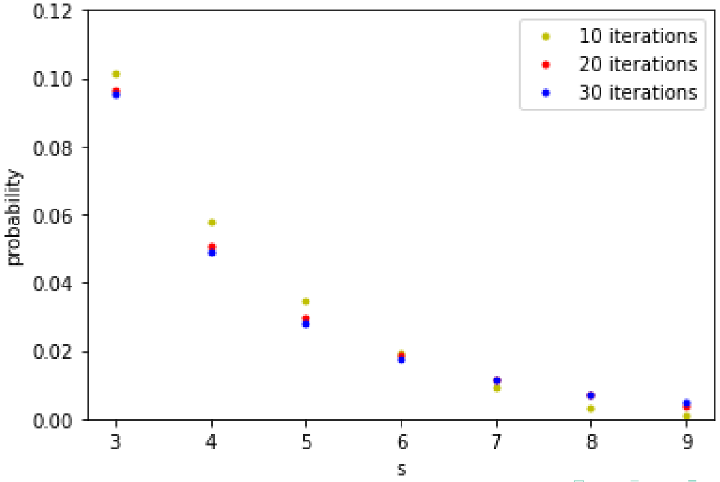
Вид предельного графика

Эксперименты показывают, что форма графика похожа на предельную уже после нескольких десятков шагов решения системы для «среднего поля». Близость графиков определяется c-нормой. Здесь — всего 0.01.



Несколько графиков

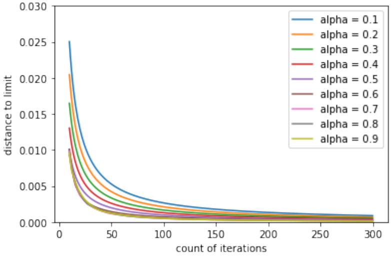
Видно, что происходит перераспределение между разными s.

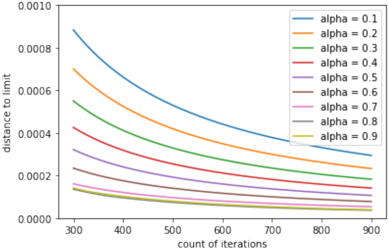


Эксперименты

Проверка точного теоретического решения

Степенное приближение при малых s некорректно, а при больших s возникают проблемы с точностью вычислений.





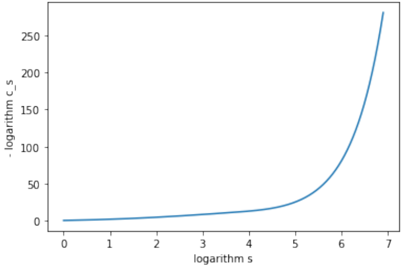
Вывод

Теоретическое решение в приближении «среднего поля» найдено верно: распределение сходится к нему достаточно быстро, причём чем больше α, тем быстрее сходимость.

Эксперименты

Проверка приближённого теоретического решения

Построим график зависимости доли людей от количества монет в логарифмическом масштабе.



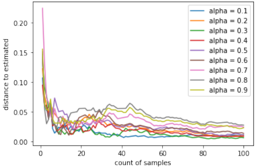
Вывод

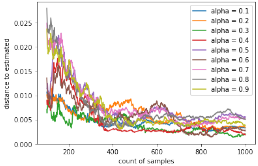
Данный график не имеет линейного вида ни на каком участке. Вероятно, это объясняется тем, что при тех s, для которых степенное приближение имеет смысл, c_s слишком малы, и их представление в языке Python не обеспечивает приемлемую точность вычислений.

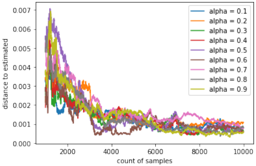
Эксперименты

Проверка приближения «среднего поля»

Сначала проверим, как среднее арифметическое реализаций распределения сходится к приближению «среднего поля» после 10 итераций.







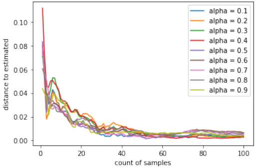
Результат

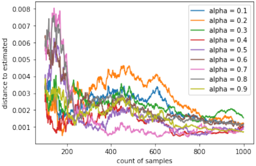
С ростом количества образцов среднее арифметическое распределений действительно сходится к распределению, полученному в приближении «среднего поля».

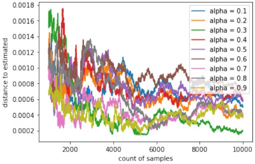
Эксперименты

Проверка приближения «среднего поля»

Та же самая проверка для 100 итераций.





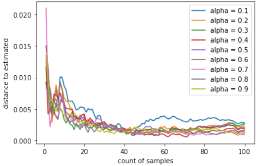


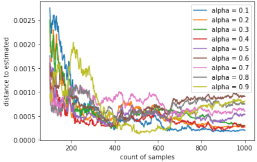
Результат

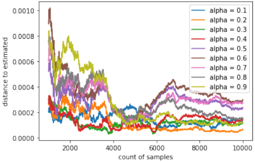
Очень похожая картина сходимости. Даже немного быстрее.

Эксперименты

И последняя проверка — для 1000 итераций.





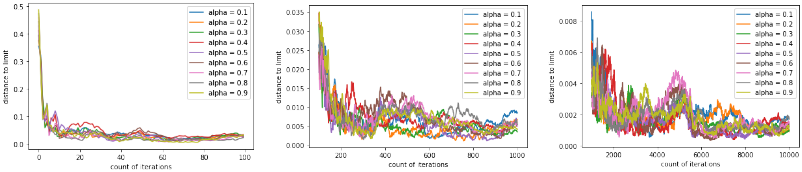


Выводы

- Среднее арифметическое реализаций стремится к результату «среднего поля» при любом числе итераций.
- Для «среднего поля» результат стремится к предельному с ростом числа итераций.
- Из неравенства треугольника можно заключить, что среднее распределение сходится к предельному с ростом числа итераций и усредняемых результатов.

Количественные характеристики сходимости

Исследование сходимости среднего арифметического реализаций к предельному распределению для 10 реализаций.



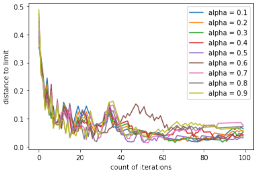
Результаты

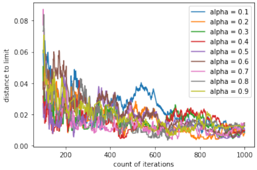
Даже для среднего из 10 реализаций наблюдается сходимость, причём отклонение в 0.01 достигается примерно за 600 итераций, а 0.002 — за 6000. Это значительно медленнее, чем для приближения «среднего поля».

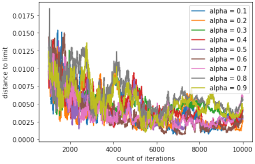
Эксперименты

Количественные характеристики сходимости

Эксперимент для единственной системы.







Результаты

Наблюдается сходимость, причём отклонение в 0.01 достигается примерно за 2000 итераций, а 0.005 — за 10000. Это ещё медленнее, чем для среднего из 10 итераций.

Выводы

- Найдено теоретически и проверено экспериментально предельное распределение. Выяснено, что степенное приближение на практике не является применимым.
- Изучен характер сходимости распределения к предельному. Для измерения погрешности использована c-норма, то есть максимальное отклонение.
- В приближении «среднего поля» сходимость быстрая: 0.005 за 50 итераций, 0.0005 за 600.
- Для единственной системы сходимость медленная: 0.01 за 2000 итераций, 0.005 за 10000.
- При усреднении по 10 независимым системам сходимость быстрее: 0.01 за 600 итераций, 0.002 за 6000.
- С ростом количества усредняемых распределений сходимость ускоряется.