

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ – ПРОЦЕССОВ УПРАВЛЕНИЯ
КАФЕДРА УПРАВЛЕНИЯ МЕДИКО-БИОЛОГИЧЕСКИХ СИСТЕМ

Ордин Леонид Иванович

Выпускная квалификационная работа

Анализ вероятностных моделей размещения

Направление 010400

Прикладная математика и информатика

Заведующий кафедрой,
доктор физ.-мат. наук,
профессор

Александров А.Ю.

Научный руководитель,
кандидат физ.-мат. наук,
доцент

Корников В.В.

Санкт-Петербург

2018

Оглавление

Введение	3
Глава 1. Общий дискретный случай.....	5
Глава 2. Общий непрерывный случай	10
Глава 3. Анализ распределения семян по Нейману.....	13
Глава 4. Анализ распределений Неймана на многовершинность.....	16
Заключение	21
Список литературы	22
Приложение 1.....	23
Приложение 2.....	24
Приложение 3.....	25
Приложение 4.....	26
Приложение 5.....	27
Приложение 6.....	28
Приложение 7.....	29
Приложение 8.....	30

Введение

Целью данной работы является анализ вероятностных моделей размещения семян растений с точки зрения модели управления капиталом семян.

Будут рассматриваться такие случаи, в которых капитал семян восполняется исключительно естественным образом, в силу таких факторов, как удалённость фермы от инфраструктуры, экзотичность выращиваемой культуры или высокая цена новых семян на рынке. Также в исследовании не рассматриваются многолетние культуры, продолжающие приносить семена каждый год после посадки.

Примерами описываемых ситуаций являются выращивание:

- картошки на Марсе (удалённость)
- новых генно-модифицированных культур (экзотичность)
- дорогих импортных цветов (высокая цена новых семян)
- любых культур в самом бизнесе производства семян

Вначале рассмотрим задачу управления капиталом в общем дискретном и непрерывном случаях. Затем перейдём к изучению частного случая размещения семян, основанного на обобщении контагиозных распределений Неймана [2], которые изначально были выведены Нейманом [3] для описания популяций вирусов, бактерий, личинок насекомых, а впоследствии используемые и в удалённых от биологии сферах, таких как описание страховых случаев [6].

Завершим работу анализом вышеназванного класса распределений на многовершинность.

Вычисления и анализ проводились в среде MATLAB, широко используемой научным сообществом. Плюсом данной платформы является большое количество расширений для решения задач оптимизации и

статистических расчётов. Также, язык MATLAB оптимизирован для сложных подсчётов высокой точности.

Несмотря на то, что работа проводилась на примере агрокультуры, практическая значимость исследования не ограничивается данным направлением, так как все приведённые рассуждения можно перенести на ряд аналогичных задач из других сфер науки.

Глава 1. Общий дискретный случай

Имеется некоторый капитал семян S .

Можно посадить любую долю из него, $a \in [0,1]$, и в конце года собрать семена урожая. Изменение капитала, x , определяется вероятностью $p(x,a)$.

Какая доля a обеспечивает наискорейший рост капитала семян в год?

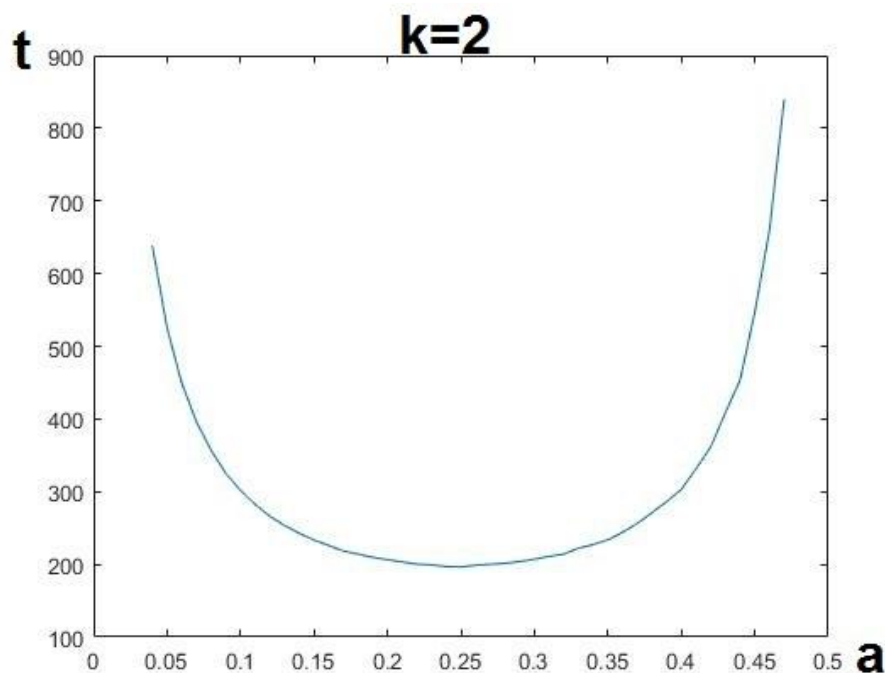
Для наглядности, вначале разберём частный случай этой задачи: допустим, есть всего две равновероятные возможности: потерять все посаженные семена, либо собрать в $k+1$ раз больше семян, чем было посажено:

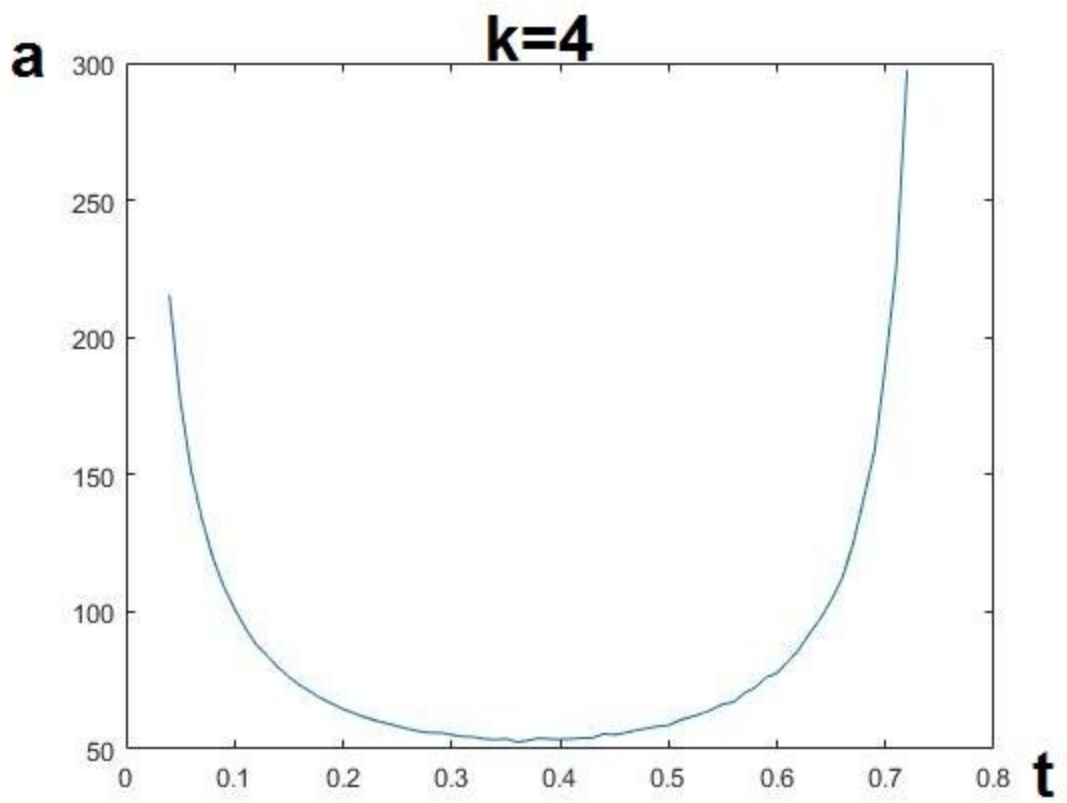
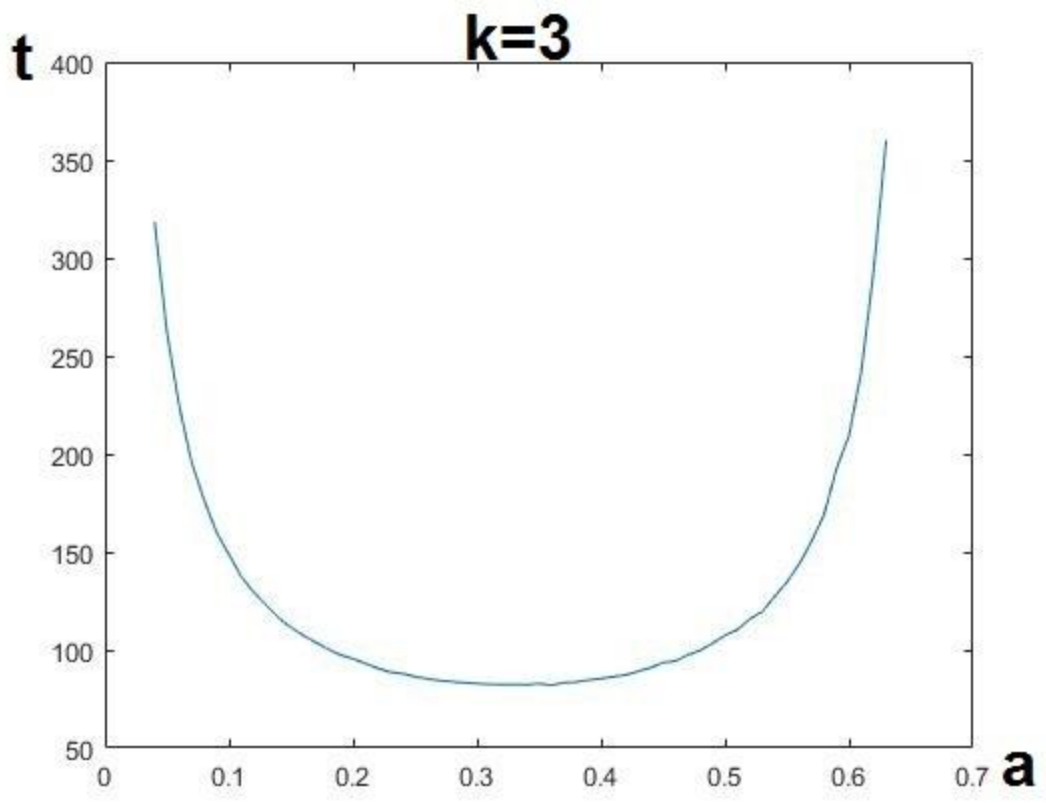
$$x_1 = -aS, \quad p_1 = 0.5$$

$$x_2 = kaS, \quad p_2 = 0.5$$

Сколько в среднем лет, t , занимает преумножение начального капитала в 100,000 раз, в зависимости от a ?

Экспериментально полученные графики $t(a)$ [приложение 1: `disc_k(10000, 0.01, 2 / 3 / 4, 100000)`] показывают, что при $k=2, 3, 4$, минимумы $t \approx 195.5, 80, 51.5$ достигаются при $a \approx 0.25, 0.333, 0.375$ соответственно:





Для вычисления ответа аналитически, перепишем аддитивные изменения капитала, x , мультипликативными, z :

$$S + x \rightarrow Sz = Se^{\ln z}; \quad z = \frac{S + x}{S}$$

Таким образом, изменения капитала добавляются соответствующими логарифмами в сумму над экспонентой, вне зависимости от их порядка.

$$z_1 = \frac{S + x_1}{S} = \frac{S - aS}{S} = 1 - a;$$

$$z_2 = \frac{S + x_2}{S} = \frac{S + kaS}{S} = 1 + ka$$

Как только сумма логарифмов исходов w_i i -тых урожаев от 1 до t удовлетворяет:

$$Se^{\sum_1^t \ln z_{w_i}} > 100,000S$$

$$e^{\sum_1^t \ln z_{w_i}} > 100,000$$

— цель преумножения капитала в 100,000 раз достигается.

Избавившись от влияния порядка исходов, посчитаем, при каком a среднее преумножение капитала в год, E , максимально:

$$E = e^{\frac{1}{2}\ln(z_1) + \frac{1}{2}\ln(z_2)} = e^{\ln\sqrt{z_1 z_2}} = e^{\ln\sqrt{(1-a)(1+ka)}} = \sqrt{(1-a)(1+ka)} \rightarrow \max$$

$$\max(E) = \sqrt{\frac{(k+1)^2}{4k}}, \quad \text{при} \quad a = \frac{1}{2}\left(1 - \frac{1}{k}\right); \quad \bar{t} \approx \frac{\ln 100,000}{\ln \max(E)}$$

Графики выше поддерживают проведённые вычисления.

Стоит отметить, что максимальное E не гарантирует минимальное \bar{t} , в силу того, что в последний год вся прибыль выше цели не учитывается.

Разница особо выделяется при низких целях. Тем не менее, оно гарантирует максимальную скорость роста капитала в год в принципе.

Интересное свойство, что каким бы высоким не было k , оптимальное a всегда меньше 0.5

Аналогично, среднее преумножение капитала в год, $E(z)$, можно определить на любом произвольном дискретном распределении преумножений капитала z_i , с соответствующими вероятностями p_i :

$$E = \exp\left(\sum_i \frac{1}{p_i} \ln z_i\right)$$

$$E = \exp\left(\ln \prod_i z_i^{p_i}\right) = \prod_i z_i^{p_i}$$

Отметим свойства:

1. E является ничем иным, как средним геометрическим случайной величины z .
2. $E(yz) = E(y) * E(z)$;
3. $E(cz) = c * E(z)$
4. Если $M(x) < 0$, то $E(z) < 1$, где M — математическое ожидание:

$$\prod_i z_i^{p_i} \leq \sum_i \frac{1}{p_i} z_i = \sum_i \frac{1}{p_i} \frac{S + x_i}{S} = 1 + \sum_i \frac{1}{p_i} \frac{x_i}{S} = 1 + \frac{M}{S} < 1$$

5. $E(z) \approx \sqrt{M(z)^2 - D(z)}$, где $D(z)$ — дисперсия [8].

Ради интереса, разрешим Санкт-Петербургский парадокс [1] с помощью E .

Рассматривается следующая ситуация: для вступления в игру, игрок платит b рублей, и затем подбрасывает честную монетку, пока не выпадет орёл. При выпадении орла на i -том броске игра заканчивается, а игрок получает выигрыш, равный 2^i рублей.

Вопрос: при каком вступительном взносе b игра становится справедливой?

Нетрудно убедиться, что математическое ожидание исхода этой игры бесконечно:

$$M = \sum_{i=1}^{\infty} \frac{1}{2^i} 2^i = \infty, \text{ иными словами, игра выгодная, вне зависимости от}$$

конечного вступительного взноса. Но в реальном мире никто не согласится платить миллионы ради игры, в которой так легко можно уйти всего лишь с двумя рублями выигрыша.

Разрешим парадокс, учитывая начальный капитал игрока S :

$$z_i = \frac{S + x_i}{S} = \frac{S - b + 2^i}{S} = 1 - \frac{b}{S} + \frac{2^i}{S}$$

$$E = \prod_{i=1}^{\infty} z_i^{p_i}$$

Посмотрим на границу $E(S,b)=1$:

S	4	8	16	32	64	128	256	512	1024
b	4	4.73	5.50	6.33	7.21	8.12	9.05	10.01	10.98

Она задаёт справедливую ставку b от текущего капитала S , относительно среднего преумножения капитала игрока в ход. Для максимизации прибыли, если игроку предложена вступительная ставка меньше табличной $b(S)$, ему стоит вступить в игру, иначе нет.

Глава 2. Общий непрерывный случай

Для полноты расширим E на непрерывный случай.

$f(z)$ — плотность распределения вероятности преумножения капитала z .

Тогда среднее геометрическое в пределе принимает вид [5]:

$$E = \exp\left(\int_{-\infty}^{+\infty} \ln(z) f(z) dz\right)$$

В качестве примера, посчитаем E , когда z распределено равномерно в промежутке $[a, b]$:

$$\ln(E) = \int_a^b \ln(z) \frac{1}{b-a} dz = \frac{1}{b-a} (z \ln z - z) \Big|_a^b = \frac{b \ln b - a \ln a}{b-a} - 1$$

$$E = \exp\left(\frac{b \ln b - a \ln a}{b-a} - 1\right)$$

Найдём справедливую b (удовлетворяющую $E=1$) при $a=0$:

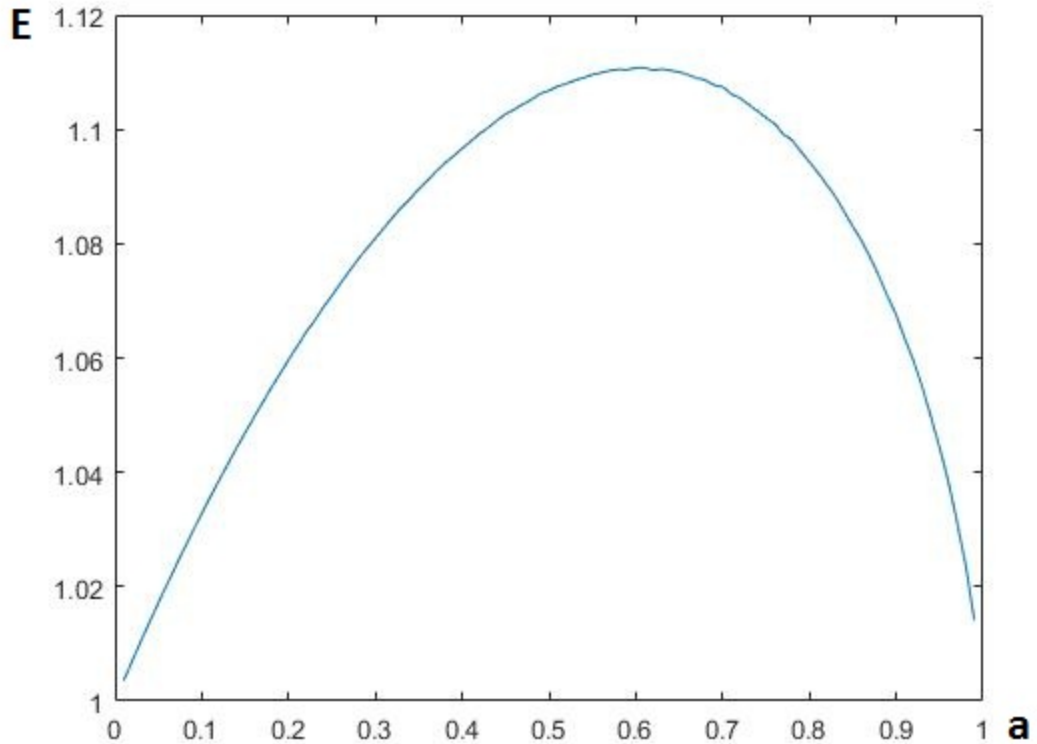
$$\ln(E) = 0$$

$$\frac{b \ln b - 0}{b - 0} - 1 = 0$$

$$b = e$$

Рассмотрим частный случай общей задачи, когда x равномерно распределён между $[-aS, (e - 1)aS]$.

Посмотрим на график $E(a)$, $0 < a < 1$, полученный экспериментально [приложение 2: `perg_ravn(100000, 0.01, 100000)`], как ранее считая количество лет до преумножения капитала в 100,000 раз:



Как и было посчитано в примере выше, $E(a) = 1$ при $a = 1$ ($z \in [0, e]$).

Найдём вершину аналитически:

$x \in [-aS, (e - 1)aS]$ *равномерно*

$$z = \frac{S + x}{S}$$

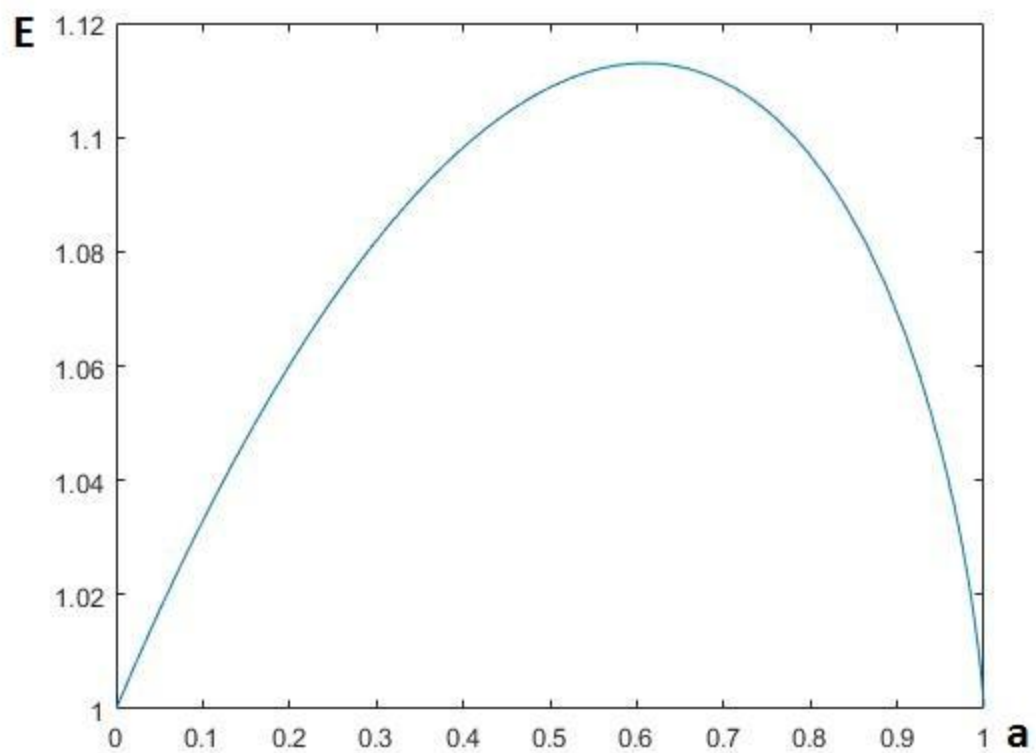
$z \in [1 - a, 1 + (e - 1)a]$ *равномерно*

Используя формулу для равномерного распределения, полученную выше, получаем:

$$E = \exp\left(\frac{(1 + (e - 1)a) \ln(1 + (e - 1)a) - (1 - a) \ln(1 - a)}{ae} - 1\right)$$

— $\max(E) \approx 1.1129$ при $a \approx 0.6094$.

Аналитически полученный график $E(a)$ [приложение 3] —



— сходится с графиком, полученным экспериментально.

Глава 3. Анализ распределения семян по Нейману

Рассмотрим более реалистичную модель разведения растений.

Имеется некоторый капитал семян, S , любую долю $a \in [0,1]$ которого можно посадить, а в конце года собрать урожай новых семян. Изменение капитала $x = -aS + k$, где k определяется вероятностью $p(k)$ согласно обобщению контагиозных распределений Неймана [2] с параметрами m_1 , m_2 , α и β , где:

- m_1 — среднее количество выросших растений, распределённых согласно распределению Пуассона, зависит от количества посаженных семян aS ;
- m_2 — среднее количество семян каждого растения, распределённых одинаково для каждого растения по Пуассону;
- α и β — параметры бета-распределения доли успешно созревших и собранных семян каждого растения.

Какая доля a обеспечивает наискорейший рост капитала семян в год, в зависимости от функциональной зависимости $m_1(aS)$ и постоянных m_2 , α и β ?

Математическое ожидание обобщения контагиозных распределений Неймана:

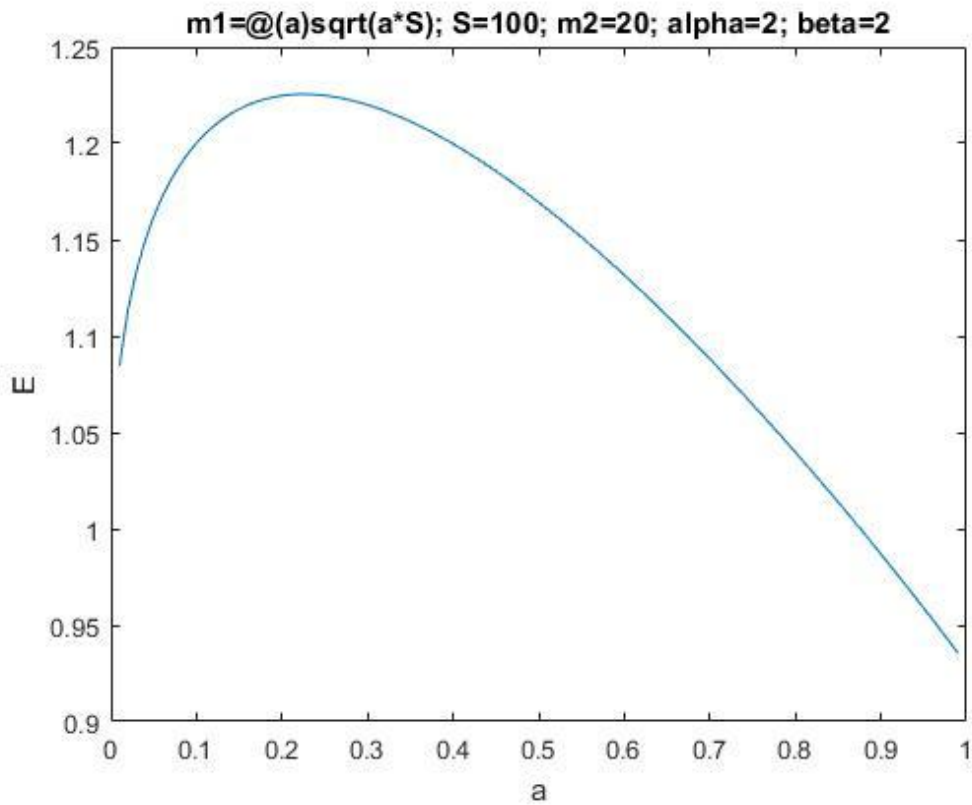
$$M = m_1 m_2 \frac{\alpha}{\alpha + \beta}$$

Так что если $m_1(aS)$ не удовлетворяет $M > aS$, чтобы математическое ожидание x было положительным, оптимальная a равна нулю, согласно четвёртому свойству E , выделенному выше.

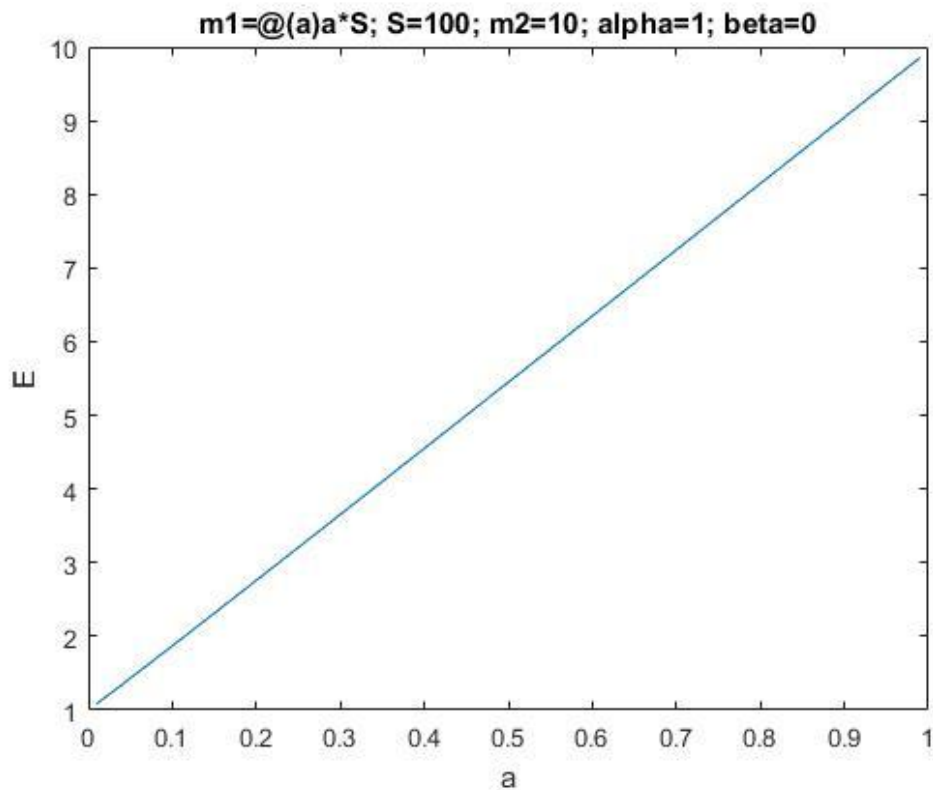
В силу сложности разрешения аналитически итогового бесконечного произведения z_i -ых из явного вида распределения, в среде MATLAB была разработана программа [приложения 4-7], которая по введённым функции $m_1(aS)$ и постоянным S , m_2 , α и β с заданной точностью выводит график $E(a)$.

Для каждого a от 0 до 1 с заданным шагом программа итеративно считает распределение $P(k)$ от введённых (α , β , $m_1(aS)$, m_2) по формулам указанным в [1] начиная с $k=0$ до тех пор, пока сумма всех посчитанных $P(k)$ не превысит 1-eps, и затем добавляет небольшой хвостик, чтобы сумма всех $P(k)$ равнялась единице. В конце каждого цикла считается E полученного распределения.

Пример вывода $E(a)$ при параметрах:



Интересно, что линейная зависимость $m1(aS)$ возвращает практически линейную зависимость $E(a)$, но в итоге $E(1)$ всегда резко загибается к нулю.

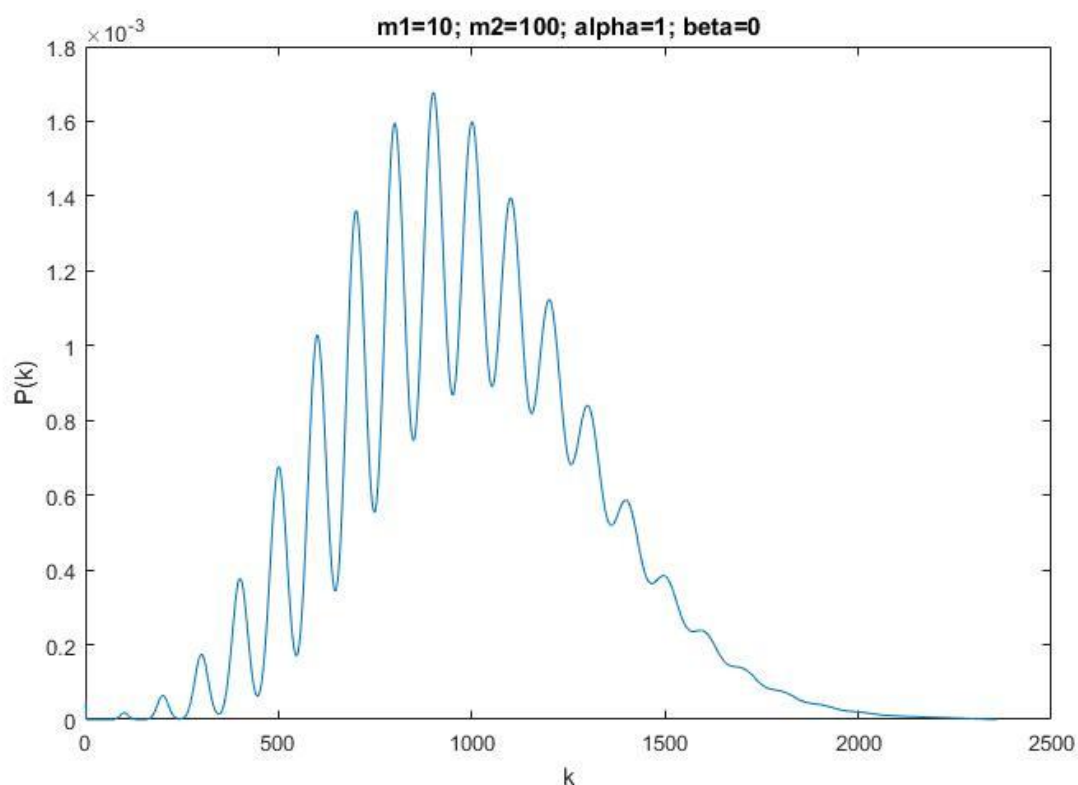


Глава 4. Анализ распределений Неймана на многовершинность

Особый интерес представляет исследовать распределения Неймана на многовершинность. В частности, задачу третьей главы можно было бы существо упростить, заменив рассматриваемое распределение приближённым отрицательным биномиальным [9]. Также ярко выраженная многовершинность существенно увеличивает разброс результатов, усложняя прогнозы. Например, когда в распределении два пика: в нуле и десяти, математическое ожидание может быть равно 5, но на практике чаще всего результат будет либо 0, либо около 10, и очень редко около 5.

Откуда возникает многовершинность в этих распределениях?

Рассмотрим пример распределения Неймана типа А при параметрах $m_1=10$, $m_2=100$ [приложение 5: $P(1,0,10,100)$]:



На графике видно, что вершины формируются в точках кратным m_2 , и это не случайно.

Когда среднее количество семян с одного растения m_2 становится достаточно большим, по сравнению со средним количеством самих растений m_1 , итоговое количество растений начинает играть большую роль на сформировавшееся количество семян.

Действительно, получить ноль семян в итоге просто отсутствием растений становится гораздо вероятнее, чем одним растением, с которого не было собрано никакого урожая семян.

В то время как одно семя можно получить только когда хотя бы одно растение всё же возшло, но оно дало очень маловероятный бедный урожай в одну семечку, вместо средних ста. Как следствие, образуется пик в точке $k=0$.

Схожим образом образуются остальные пики: например, вероятнее получить 99 или 101 семя от одного растения, чем от двух неудачных, давших всего половину ожидаемого урожая с каждого. Однако с одного растения вероятнее собрать 100 семян, чем 99 или 101, поэтому вокруг сотни образуется второй локальный пик.

Следующий пик в $k=200$ образован суммой двух пуассоновских распределений по 100, но уже немного больше размыт, потому что на него начинают сильнее влиять возможные скудные урожаи с трёх растений, давших примерно по две трети среднего каждый, а также очень успешный урожай с одного.

Таким образом, следующие вершины постепенно сглаживаются соседними.

Понимая природу многовершинности этого распределения, мы можем заключить, что если распределение вероятностей многовершинно, увеличение m_2 не изменит многовершинности, и если распределение одновершинно, уменьшение m_2 не создаст многовершинность. Значит, можно выделить границу $m_2(m_1)$, с одной стороны от которой распределение одновершинно, а с другой — многовершинно.

На основании этого факта, напишем программу [приложение 8], которая

будет искать эту границу по заданным параметрам α и β .

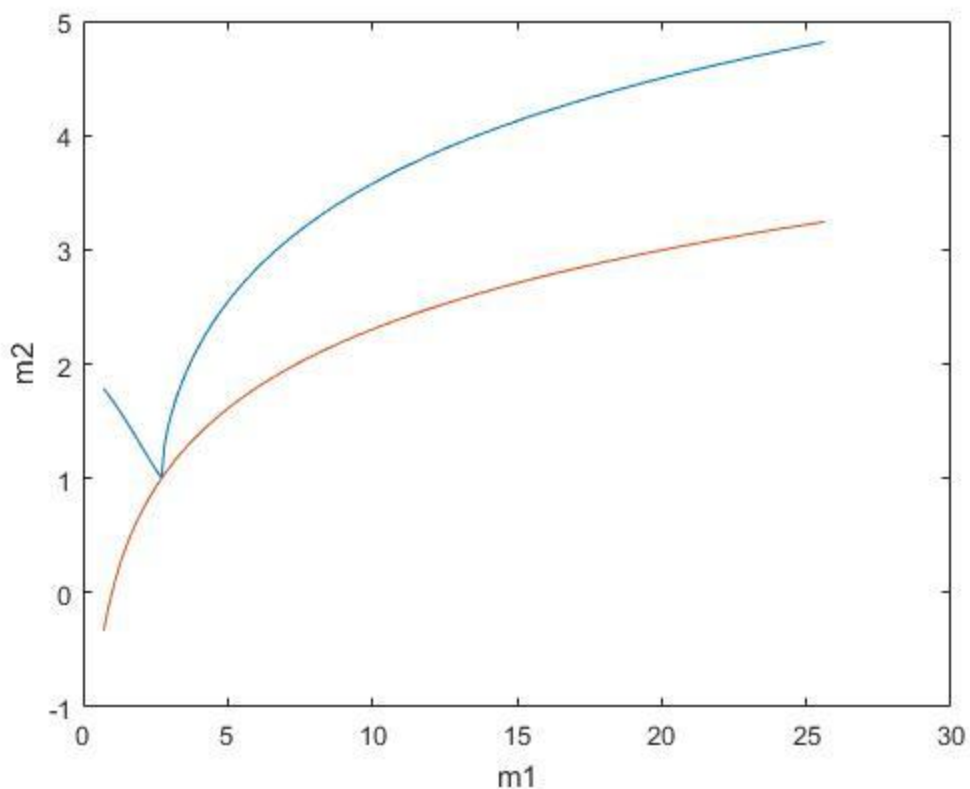
Из приложения 5 мы уже имеем программу, считающую $P(\alpha, \beta, m_1, m_2)$.

Проходя циклом по m_1 от начальной до конечной точки с заданным шагом, бинарным поиском с заданной точностью находим граничную m_2 , с одной стороны которой распределение $P(\alpha, \beta, m_1, m_2)$ одновершинно, а с другой многовершинно.

Количество вершин находилось количеством перемен знаков в последовательности $\{P(k+1)-P(k)\}$, $k=1, \dots, n-1$.

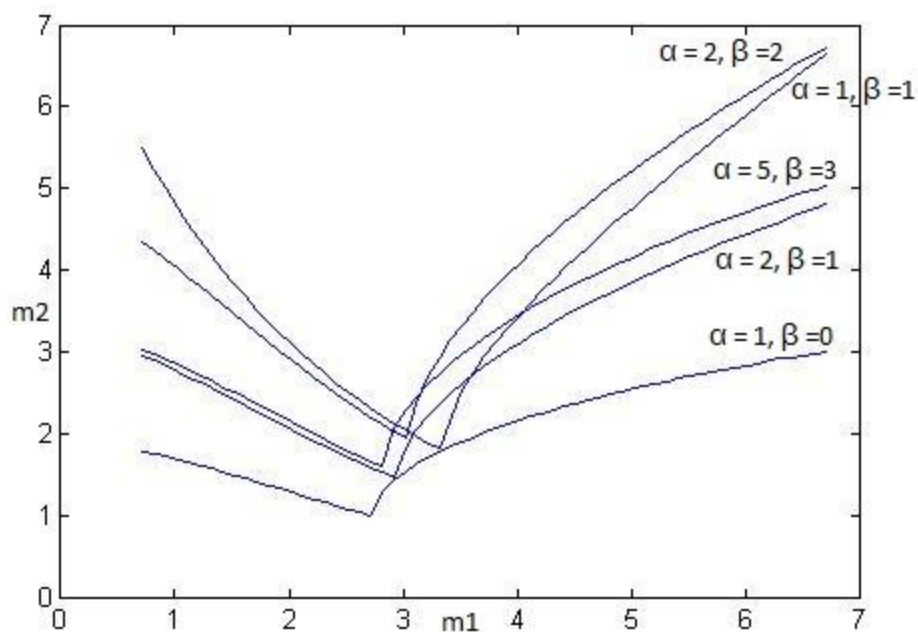
В статье [4] аналитически показывается, что распределение Неймана типа А ($\alpha = 1, \beta = 0$) одновершинно, когда $m_2 \leq \ln(m_1)$;

Давайте посмотрим на график граничной $m_2(m_1)$, полученный программой (синим), против найденной в статье $m_1 = \ln(m_2)$ (красным):



Графики касаются в точке $m_1 = e, m_2 = 1$, однако настоящая область унимодальности лежит существенно выше.

Закономерно, граница унимодальности распределения Неймана типа А лежит ниже всех, потому что в ней успешно созревают и собираются все семена с каждого растения.



Интересно также отметить, что во всех найденных многовершинных распределениях, первая вершинка всегда была в точке $k=0$, что может быть очень полезным для последующего изучения унимодальности распределений Неймана.

Заключение

В нашем исследовании было получено аналитическое решение общего случая задачи управления капиталом, впоследствии применённое на ряде частных случаев моделей размещения семян растений.

В силу сложности в ходе работы не удалось аналитически найти решение модели управления капиталом семян при размещении семян согласно обобщению контагиозных распределений Неймана, а также аналитически найти границы унимодальности вышеназванного класса распределений. Для чего были разработаны программы, решающие эти задачи численно для любых конкретных параметров.

Во втором случае, программа позволила пролить свет на возможное расширение границы унимодальности распределения Неймана типа А, полученной аналитически в статье [4], а также на возможный новый способ анализа многовершинности распределений Неймана по вершине в точке $k = 0$, что является возможным направлением для развития исследования в будущем.

Также в работе было представлено интересное разрешение Санкт-Петербургского парадокса с учётом доступного игроку начального капитала.

Список литературы

1. Martin, Robert. "The St. Petersburg Paradox". The Stanford Encyclopedia of Philosophy, 2001.
2. Gurland, John. A Generalized Class of Contagious Distributions / Biometrics, Vol. 14, No. 2, 1958. — pp. 229-249.
3. Neyman, Jerzy. On a new class of "contagious" distributions applicable in entomology and bacteriology / Annals of Mathematical Statistics 10, 1939. — pp. 35-57.
4. Massé Jean–Claude, Theodorescu Radu. Neyman Type A Distribution Revisited, December 10, 2004.
5. Айвазян С. А. Прикладная статистика: Основы моделирования и первичная обработка данных, М.: Финансы и статистика, 1983. — с. 139.
6. Panjer Harry, Willmot Gordon. Insurance risk models / Society of Actuaries, Schaumburg, 1992. — pp. 97, 258-259.
7. Barton, D. E., The modality of Neyman's contagious distribution of type A, Trabajos de Estadística 8, 1957. — pp. 13–22.
8. Винс Ральф. Математика управления капиталом: методы анализа риска для трейдеров и портфельных менеджеров. М.: 2007. — с. 67.
9. Rahman M.S., Gupta R.P. Identification and approximation of distribution by moment ratios, 2007, pp.59-65.

Приложение 1

```
function [n, step, k, S_target] = disc_k(n, step, k, S_target)
    Plot = [];
    for a = 0.04:step:1
        t = 0;
        for i = 1:n
            S = 1;
            while (S < S_target)
                S = S * (1 - a + (randi(2) - 1) * (k + 1) * a);
                t = t + 1;
            end
        end
        disp([a t / n])
        Plot = [Plot; a t / n];
        if (t / n > 1.05 * Plot(1, 2))
            break
        end
    end
    plot(Plot(:, 1), Plot(:, 2))
end
```

Приложение 2

```
function [n, step, S_target] = nepr_ravn(n, step, S_target)
    Plot = [];
    for a = step:step:(1 - step)
        j = 0;
        for i = 1:n
            S = 1;
            while (S < S_target)
                S = S * (1 - a + exp(1) * rand * a);
                j = j + 1;
            end
        end
        disp([a exp(log(S_target) * n / j)])
        Plot = [Plot; a exp(log(S_target) * n / j)];
    end
    plot(Plot(:, 1), Plot(:, 2))
end
```


Приложение 3

```
E=@(a)exp(((1 + (exp(1) - 1) * a) * log(1 + (exp(1) - 1) * a) -  
(1 - a) * log(1 - a)) / (a * exp(1)) - 1)  
fplot(E, [0 1])  
ylim([1 1.12])
```

Приложение 4

```
clear;
clc;

alpha = 2;
beta = 2;
S = 100;
m1 = @(a) sqrt(a * S);
m2 = 20;
step = 0.01;

Plot = [];
for a = 0:step:1 - step
    p = P(alpha, beta, m1(a), m2);
    X = - a * S + [1:size(p)].' - 1;
    Z = (S + X) / S;
    disp([a E(Z, p)])
    Plot = [Plot; a E(Z, p)];
end
plot(Plot(:, 1), Plot(:, 2))
title(strcat('m1=', func2str(m1), sprintf('; S=%d; m2=%d;
alpha=%d; beta=%d', S, m2, A, B)))
xlabel('a')
ylabel('E')
```

Приложение 5

```
function P = P(a, b, m1, m2)
e = 0.001;

r = 0;
F = [];
p = exp(-m1) * exp(m1 * hypergeom(a, a + b, -m2));
big = false;
while (sum(p) < 1 - e)
    if ~big
        fk = Fk(r, a, b, m2);
        if isnan(fk) || fk == Inf
            big = true;
            fk = 0;
        end
    end
    F = [F; fk];
    r = r + 1;
    s = 0;
    for k = 0:r - 1
        s = s + F(k + 1) * p(r - k);
    end
    p=[p; m1 / r * s];
end

x = ceil(2 * (1 - sum(p)) / p(end));
h = 2 * (1 - sum(p)) / x / (x + 1);

for i = 1:1:x
    p = [p; h * (x - i + 1)];
end

P = p;
```

Приложение 6

```
function Fk = Fk(k, a, b, m2)
%Fk = m2 ^ (k + 1) / factorial(k) * factorial(a + k) /
factorial(a - 1) * factorial(a + b - 1) / factorial(a + b + k) *
hypergeom(a + k + 1, a + b + k + 1, -m2);
Fk = m2 ^ (k + 1) * prod(k + 1:1:a + k) / factorial(a - 1) /
prod(a + b:1:a + b + k) * hypergeom(a + k + 1, a + b + k + 1, -
m2);
```

Приложение 7

```
function E = E(Z, P)
prod = 1;
for i = 1:size(Z)
    prod = prod * Z(i) ^ P(i);
end
E = prod;
```

Приложение 8

```
clear;
clc;

alpha = 1;
beta = 0;
e = 10^-4;
m1_start = 0.7182;
m1_end = 10.7182;
step = 0.1;

m1 = m1_start;
direction = - 1;
Plot = [];
exception = 0;
h = 4;
m2_left = h;
m2_right = 30;

while (m1 <= m1_end)
    [n,m] = size(Plot);
    if (n >= 2)
        direction = sign(Plot(n, 2) - Plot(n - 1, 2));
        h = abs(Plot(n, 2) - Plot(n - 1, 2)) + 3 * e;
    end
    if (direction == - 1)
        m2_left = m2_left - h * 2;
    end
    if (direction == 1)
        m2_right = m2_right + h * 2;
    end
    end
    if strcmp(exception, 'reached the boarder')
        m2_left = 0;
        m2_right = 30;
    end
    m2_l = m2_left;
    m2_r = m2_right;

    while (m2_right - m2_left > e)
        m2 = (m2_left + m2_right) / 2;
        p = P(alpha, beta, m1, m2);
        sig = sign(p(2:end) - p(1:end - 1));
        sig_changes = sum(abs(sign(sig(2:end) - sig(1:end -
1)))));
        if (sig_changes >= 2)
            m2_right = m2;
        elseif (sig_changes < 2)
            m2_left = m2;
        else

```

```

        exception = 'P out of reach'
        break
    end
end

exception = 0;
if (abs(m2_left - m2_l) < e) || (abs(m2_right - m2_r) < e)
    exception = 'reached the boarder'
else
    Plot = [Plot; m1 m2];
    disp([m1 m2])
    m1 = m1 + step;
end
end

f=@(x)log(x);
plot(Plot(:, 1), Plot(:, 2), Plot(:, 1), f(Plot(:, 1)))
%plot(Plot(:, 1), Plot(:, 2))
xlabel('m1')
ylabel('m2')
```