

Национальный исследовательский университет

Высшая школа экономики

Кафедра математической экономики
и эконометрики

КУРСОВАЯ РАБОТА
НА ТЕМУ: «АНАЛИЗ ИНТЕРНЕТ ДАННЫХ»

Студент группы И-212
Петров Никита Александрович

Научный руководитель
Старший преподаватель
кафедры математической экономики
и эконометрики
Демешев Борис Борисович

Москва
2012

1 Введение

Сегодня мы можем часто встретить задачи, для выполнения которых необходимо обработать информацию из сети Интернет. Не смотря на то, что она находится в свободном доступе, работать с ней не всегда удобно.

Зачастую всё дело в том, что данные в сети представлены не в виде таблиц или списков. Более того, элементы одной выборки могут располагаться на разных страницах или даже на разных сайтах. Можно собирать эти данные вручную, но гораздо проще автоматизировать процесс. Поэтому существуют специальные методы, автоматизирующие процесс сбора и обработки информации из сети Интернет.

2 Задачи

Один из таких способов — использовать язык программирования Python.

Python - язык программирования, созданный в 1990 году. Особенностью является акцент на производительность разработчика и читаемость кода. Распространяется под свободной лицензией «Python Software Foundation License», официальный сайт — <http://www.python.org/>. Версия Python, которую мы будем использовать — **2.7.3**.

С помощью Python мы будем заниматься парсингом сайтов, то есть программа и специальные библиотеки к ней будут структурировать html-код страницы чтобы собирать информацию автоматически.

Библиотека, которую мы используем — BeautifulSoup. Эта библиотека распространяется под свободной MIT лицензией и легко подключается к Python. С ее помощью можно заниматься парсингом сайтов, официальный сайт — <http://www.crummy.com/software/BeautifulSoup/>. Версия BeautifulSoup, которую мы будем использовать — **3.2.1**

Вторая часть работы состоит в некотором анализе собранных данных. С помощью библиотек к Python мы визуально представим собранную информацию. Библиотека, которую мы будем использовать — Matplotlib, версия — **1.1.0** - тоже распространяется бесплатно, официальный сайт — <http://matplotlib.sourceforge.net/>

В ходе работы мы обработаем два источника информации. Первый — данные о ставках на аукционе нестандартного типа, второй — доходы депутатов государственной думы и их супруг/супругов.

3 Аукцион

Рассмотрим аукцион типа «минимальная уникальная ставка» или «lowest unique bid». Пример, с которым мы будем работать — <http://www.uniquebidhomes.com/lub/index.php>. Суть аукциона такова.

1. Игроки поочередно или одновременно делают ставки. Обычно это суммы от 0.01\$ до 5.00\$ на этом аукционе.
2. Чтобы сделать каждую ставку необходимо заплатить так называемый взнос «fee».
3. Просматриваются ставки, сделанные за отведенное на аукцион время.
4. Победившим признается игрок, сделавший уникальную, но при этом самую маленькую ставку.

Цель аукциона состоит сразу из двух подцелей — «сделать уникальную ставку» и «сделать минимальную ставку». Должно быть выполнено оба условия. Мы рассмотрим поведение игроков,

а именно плотность ставок, то есть как часто были сделаны ставки каждого номинала от 0.01\$ до 5.00\$.

В каждом завершенном аукционе предоставляется статистика: время — игрок — размер ставки. Победившая ставка обозначается звездочкой.

Например:

04/10/11 11:34	7700	\$0.76
04/10/11 21:04	7700	\$0.27
04/10/11 21:05	7700	★ \$0.34
04/10/11 21:32	7700	\$0.29
04/10/11 21:33	7700	\$0.28
04/10/11 21:37	7700	\$0.33
04/10/11 21:46	7700	\$0.31
04/10/11 21:49	7700	\$0.30
04/10/11 21:49	7700	\$0.32
04/10/11 21:50	7700	\$0.38

Рис. 1: Статистика по ставкам на сайте

Часть html-кода (предпоследняя строка) выглядит так:¹

```
<div id="mem_table_rows" style="margin-bottom:10px;">
<table width="695" border="0" cellpadding="0"
  cellspacing="1" bgcolor="#E0DFB4" align="center">
[...]
```

04/10/11 21:49	7700	\$0.32
----------------	------	--------

```
</table>
</div>
```

Теперь обратимся к библиотеке Beautiful Soup. Она позволяет из структурированного командой

«BeautifulSoup(urllib2.urlopen("http://www.uniquebidhomes.com/lub/results.php"))» содержимого страницы (в данном случае со страницы со списком завершенных аукционов) найти любую информацию. Для поиска информации используем команды «find» и «findAll» (для списков). Допустим, нам надо собрать с этой страницы все (или несколько) адресов аукционов, чтобы затем перейти по ним и работать непосредственно на страницах аукционов.

```
item_url = "http://www.uniquebidhomes.com/lub/" \
+ items.find("table", {"class" : "content_table"}).\
findAll("a")[i]["href"]
```

Где

1. «item_url» — название переменной, которая является адресом страницы с аукционом.
2. «items» — переменная, которая хранит структурированную под BeautifulSoup страницу.
3. «find("table", {"class" : "content_table"})» — команда которая ищет таблицу («table») с параметром «class» равным «content_table».

¹код представлен в удобном для просмотра и печати виде

4. точка показывает, что дальше есть команда для работы над найденной таблицей
5. «`findAll("a")[i]["href"]`» — команда, похожая на «`find`» за исключением того, что она ищет **все** разделы «a» (обычно это разделы с ссылками), но берёт конкретную по очереди (а именно `i`-ую), а затем берёт в ней показатель «`href`» — то есть адрес.

Пройдя циклом и собрав все ссылки, циклом же пройдем теперь уже по всем страницам с аукционами. С каждой страницы с помощью такого кода список ставок:

```
page = BeautifulSoup(urllib2.urlopen(item_url))
q = len(page.find("div", {"id" : "mem_table_rows"}).\
    findAll("td", {"align" : "right"}))
for x in range(0, q - 1):
    bid = page.find("div", {"id" : "mem_table_rows"}).\
        findAll("td", {"align" : "right"})[x].text
    if bid.count("&nbsp;") == 2:
        bid = bid.replace("&nbsp;", "")
        bid = bid.replace(" ", "")
        bid = bid.replace("$", "")
        bid = float(bid)
        winner = bid
    else:
        bid = bid.replace("&nbsp;", "")
        bid = bid.replace(" ", "")
        bid = bid.replace("$", "")
        bid = float(bid)
    list.append(bid)
```

В этой части кода:

1. `q` - количество строк, которые мы считаем командой `len(***)`
2. `bid` - собственно значение ставки, строка.
3. Ищем в таблице с параметрами «`id`» = «`mem_table_rows`» элементы «`td`» с параметрами «`align`» = «`right`» и берём из этого значение «`text`»
4. а затем «обрезаем» лишние символы и переводим в тип «`float`».
5. И добавляем в список со ставками «`list`», с которым в последствии будем работать

Все параметры и элементы, необходимые для кода мы узнали из html-кода страницы ставок. И так для необходимого количества аукционов.

Соберем информацию с первой страницы аукционов, а именно первые 29 аукционов. Получили следующие данные:

Все ставки на аукционах:

http://cdn.bitbucket.org/nikitapetrov/lub_auctions/downloads/Data29.txt

Ставки, победившие на аукционах:

http://cdn.bitbucket.org/nikitapetrov/lub_auctions/downloads/DataW.txt

Затем строим плотность ставок на каждом аукционе. Для каждого значения на оси $0x$ мы ставим значение на оси $0y$, высчитываемое как сумма всех плотностей нормальных распределений. Где количество этих распределений равно числу ставок, $\sigma = 0.03\$$ для каждого, а μ является собственно величиной ставки для каждого аукциона. Используем такой код:

```

maxt = max(general_list[t])
p = numpy.arange(0, maxt, 0.01)
ptnorm = 0
for h in range(len(general_list[t])):
    ptnorm += mlab.normpdf(p, general_list[t][h], sigma)
ptnorm = ptnorm / len(general_list[t])

```

Здесь:

1. `t` — номер аукциона, с которым мы работаем, список его ставок
2. «`maxt`» — максимальная ставка на аукционе (обычно в пределах 5.00\$).
3. «`p = numpy.arange(0, maxt, 0.01)`» строит список для оси $0x$
4. «`ptnorm += mlab.normpdf(p, general_list[t][h], sigma)`» высчитывает значения y

Полный код, с некоторыми техническими дополнениями, представлен в Приложении. Построим графики для первых 6 аукционов.

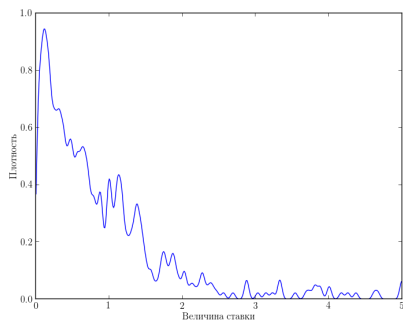


Рис. 2: Плотность

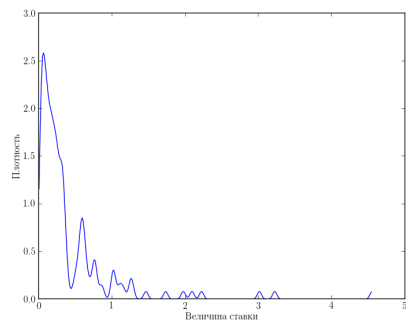


Рис. 3: Плотность

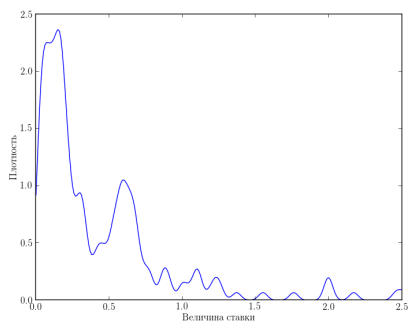


Рис. 4: Плотность

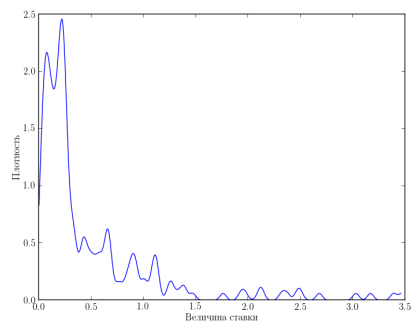


Рис. 5: Плотность

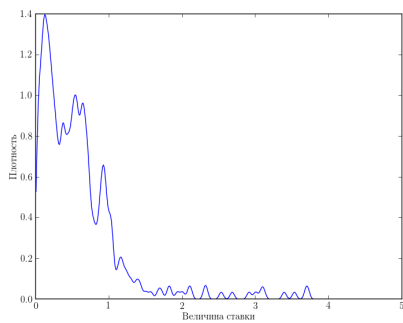


Рис. 6: Плотность

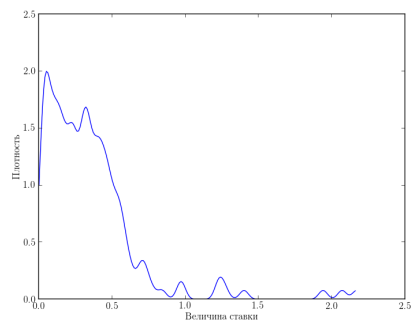


Рис. 7: Плотность

Визуально графики в общих чертах напоминают экспоненциальное распределение. Распределение должно быть асимметричным и длиннохвостым. Формула плотности экспоненциального распределения:

$$p(x) = \begin{cases} 0 & , \text{ при } x < 0 \\ \lambda e^{-\lambda x} & , \text{ при } x \geq 0 \end{cases} \quad (1)$$

Методом максимального правдоподобия оценим параметр λ . Находим функцию правдоподобия:

$$F(x_1, x_2, \dots, x_n, \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} \quad (2)$$

Находим логарифмическую функцию правдоподобия:

$$L(x_1, x_2, \dots, x_n, \lambda) = n \ln \lambda - \lambda \sum_{i=1}^n x_i \quad (3)$$

Дифференцируем по λ и приравниваем к 0:

$$\frac{dL}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0 \quad (4)$$

Получаем λ^* :

$$\lambda^* = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}} \quad (5)$$

Поэтому код, с помощью которого мы вычисляем параметр λ экспоненциального распределения `lam[t]`, а затем строим графики плотности ставок и экспоненциального распределения будет выглядеть так:

```
for r in range(len(p)):
    expon[t].append(0)
lam[t] = float(sum(general_list[t])) / len(general_list[t])
print lam[t]
lam[t] = 1 / lam[t]
for r in range(len(p)):
    expon[t][r] = lam[t] * math.exp(-lam[t]*p[r])
pylab.plot(p, ptnorm)
pylab.plot(p, expon[t])
pylab.show()
```

В итоге, для тех же 6 графиков получим следующие экспоненциальные распределения:

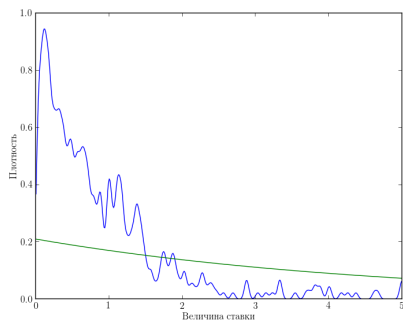


Рис. 8: Плотность распределения ставок на аукционе

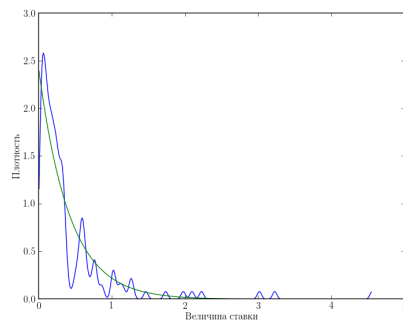


Рис. 9: Плотность распределения ставок на аукционе

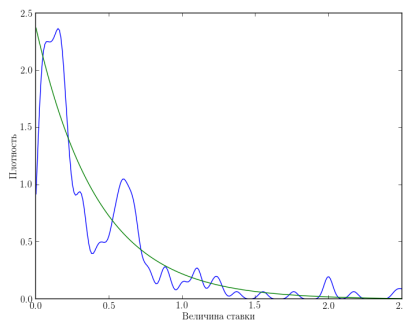


Рис. 10: Плотность распределения ставок на аукционе

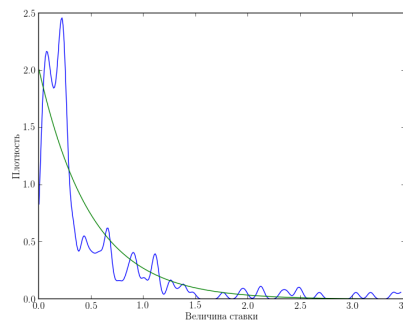


Рис. 11: Плотность распределения ставок на аукционе

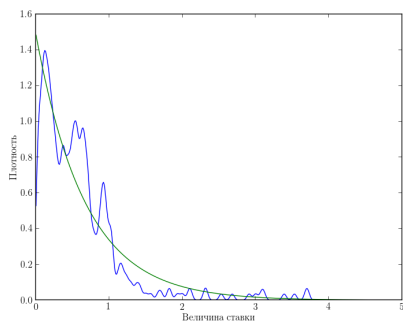


Рис. 12: Плотность распределения ставок на аукционе

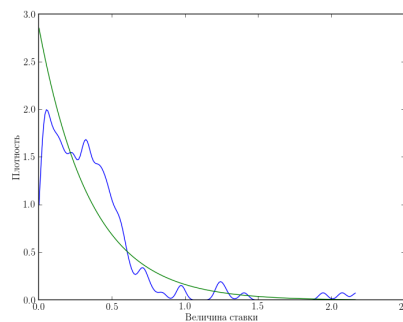


Рис. 13: Плотность распределения ставок на аукционе

По графикам видно, что экспоненциальное распределение в общем отражает поведение игроков на аукционе. Исключения: 1 аукцион, где очень много было сделано ставок выше 5.00\$, и аукцион 6, распределение которого оказалось не таким остроконечным, как предполагало экспоненциальное распределение. Теперь посмотрим на статистику ставок, с которыми игроки победили на аукционах. Код сохраняет информацию, о том, являлась ли ставка победной (имела ли «звездочку в таблице») и сохраняла в отдельный список.

Строим плотность распределения победивших ставок (аналогично плотности всех ставок на каждом аукционе).

Посмотрим экспоненциальное распределение (методом максимального правдоподобия, аналогично для распределения ставок на каждом аукционе).

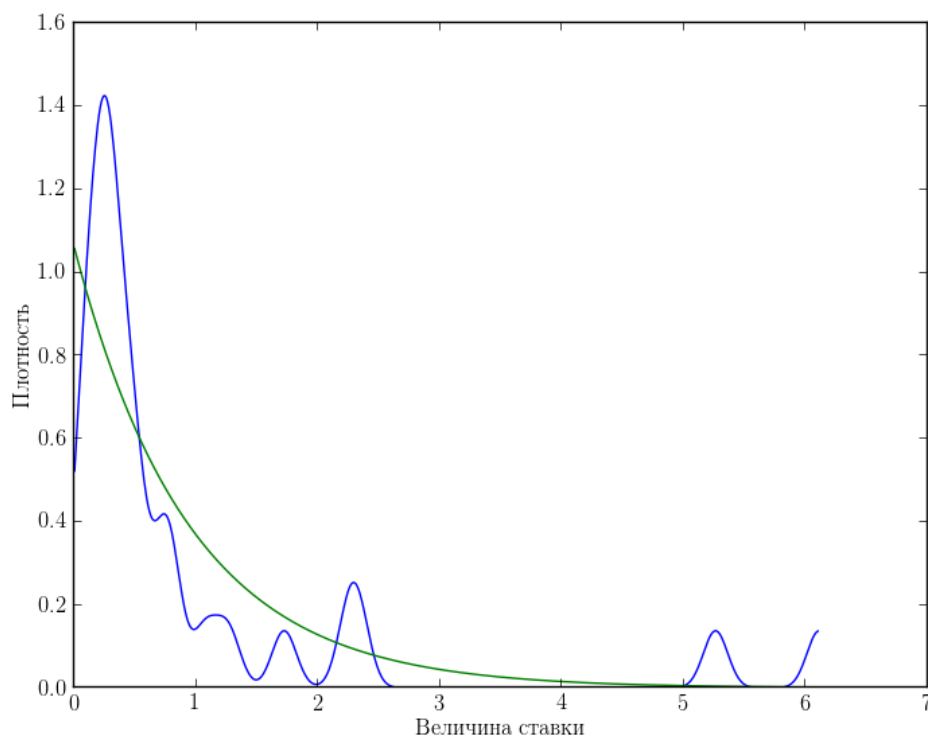


Рис. 14: Плотность распределения победивших ставок

Видно, что распределение не совсем близко к данному экспоненциальному. Причиной может быть некоторые ставки больше 5, исключив которые, построим новое распределение. Распределения все еще не достаточно близки, но уже более похожи. Причиной такого трудного подбора функции распределения для распределения победивших ставок можно считать наличие всего 29 наблюдений.

Графики, для большего числа аукционов, а также полный код представлен в Приложении.

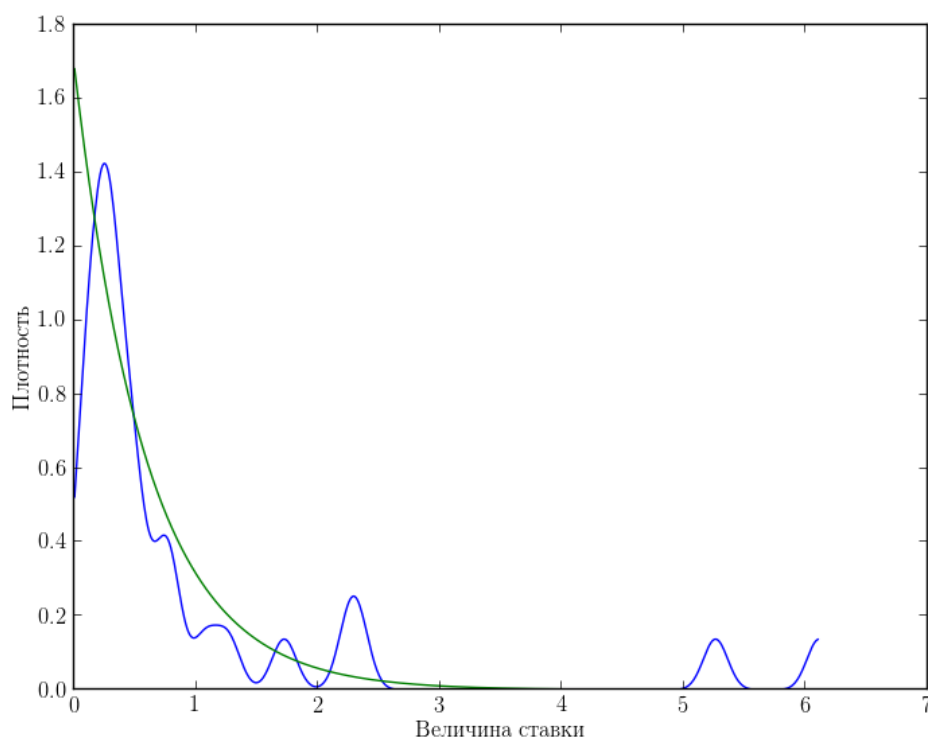


Рис. 15: Плотность распределения победивших ставок

4 Доходы членов Государственной Думы

Следующий тип сбора и представления информации из сети Интернет — построение двумерной плотности распределения заявленных доходов депутатов и их супруг/супругов. Информация находится в свободном доступе на сайте Государственной Думы. Сайт — <http://www.duma.gov.ru/structure/deputies/>. Алгоритм сбора информации похож на сбор данных с аукционов. Код просматривает список всех депутатов, затем заходит на страницу каждого, читает таблицу доходов. Доход записывает в двумерный вектор дохода в первую переменную. Если есть супруг или супруга, записывает в вектор дохода во вторую переменную это значение, иначе ставит 0.

В случаях ошибки поиска по таблице, когда у депутата просто не заявлен доход, пропускает. В результате собраны следующие данные: <http://cdn.bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/Data.txt>.

Данные характеризуют следующие параметры:

Всего депутатов — 447

Всего наблюдений — 442

Количество депутатов, у которых не заявлен доход — 5

	Доходы всех, руб.	Доходы депутатов, руб.	Доходы супруг/супругов депутатов, руб.
Минимальное значение	0.00	99 185.85	0.00
Максимальное значение	2 712 131 087.00	2 712 131 087.00	1 305 119 348.00
Среднее значение	16 654 365.34	25 715 467.92	7 593 262.77

Чтобы примерно представить себе, как распределены доходы депутатов, построим график, где на оси Oy отложен доход.

На оси Ox номер депутата/супруги в ранжированном по доходу списке. Видно, что только с 800-го начался серьезный рост доходов.

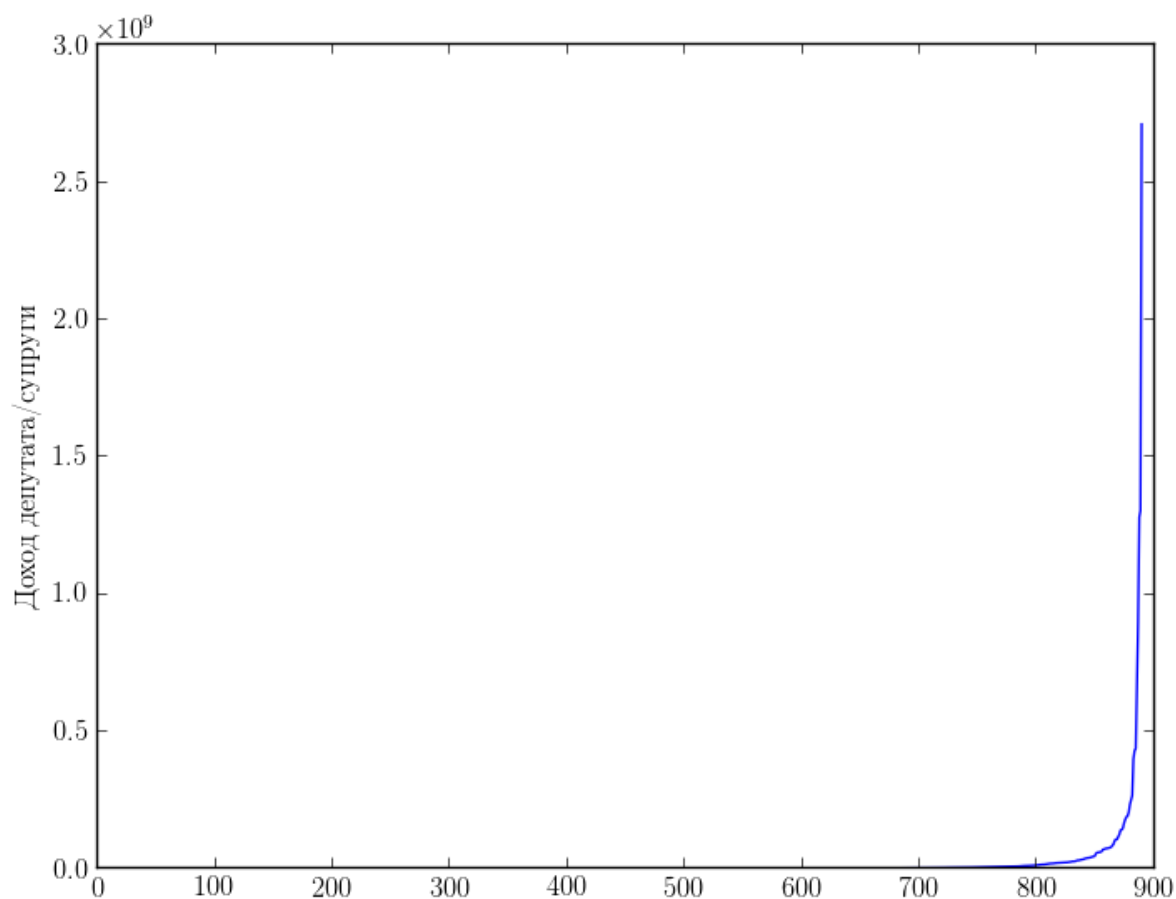


Рис. 16: Доход депутата/супруги

Поэтому, при построении двумерной плотности (аналогично аукционам) были заданы следующие параметры:

$$\sigma_x, \sigma_y = 450\,000$$

$$\sigma_{xy} = 0$$

Максимальное значение на шкале дохода — 10 000 000, потому что встречались значения выше, из-за которых двумерная плотность становилась менее наглядной. Хотя среднее больше 16 000 000 рублей, оно возникло из-за больших значений максимума.

Получили следующую поверхность:

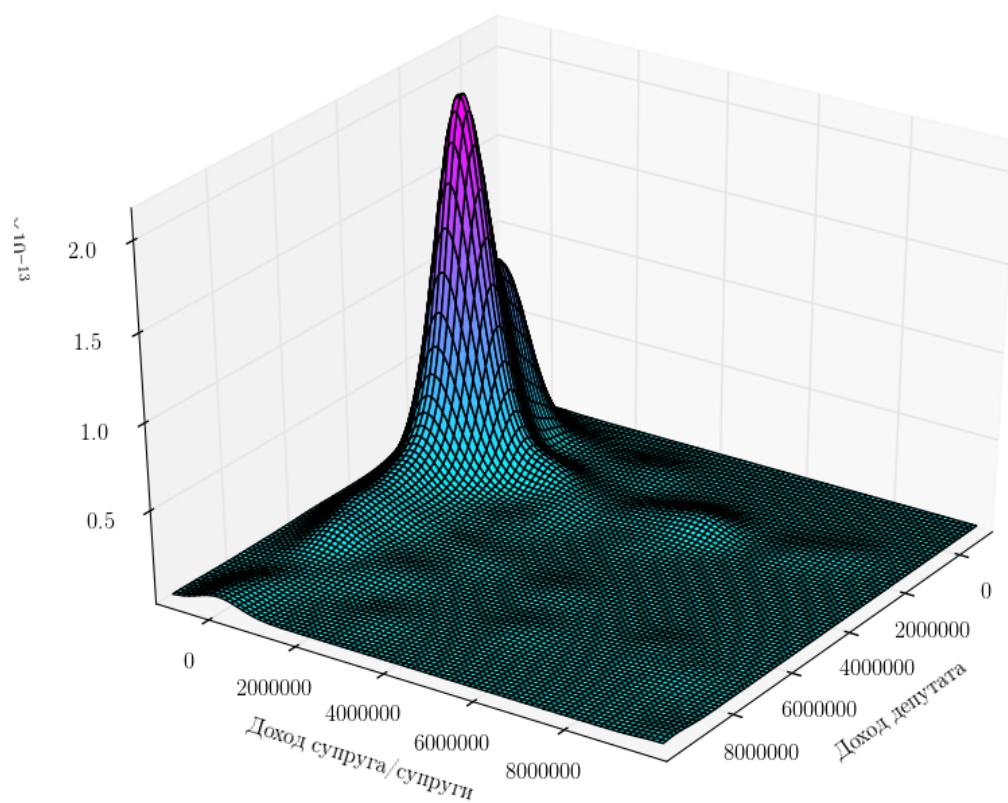


Рис. 17: Двумерная плотность

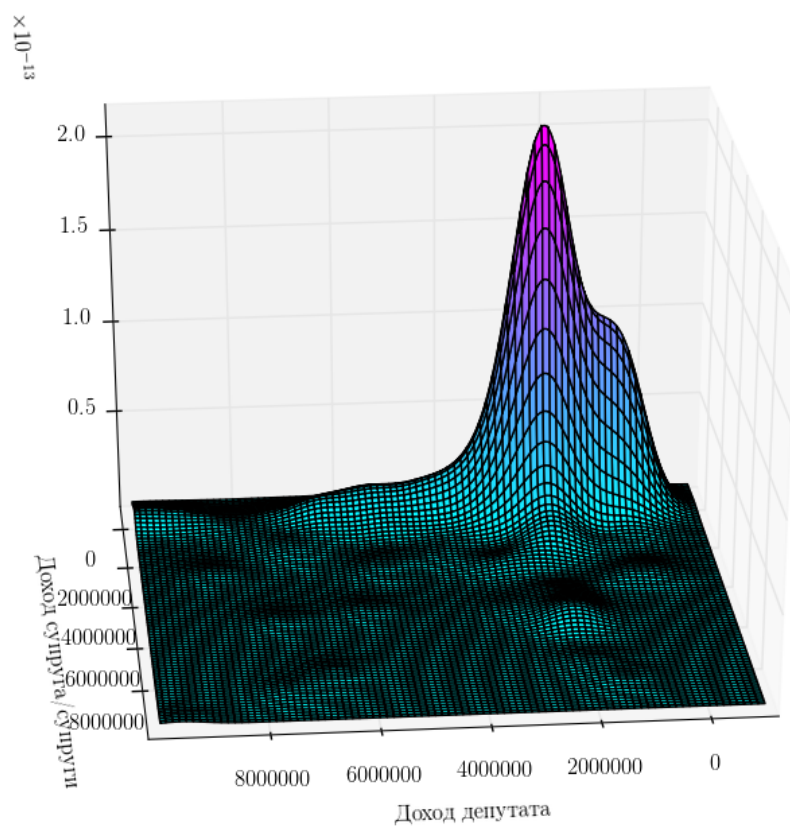


Рис. 18: Двумерная плотность

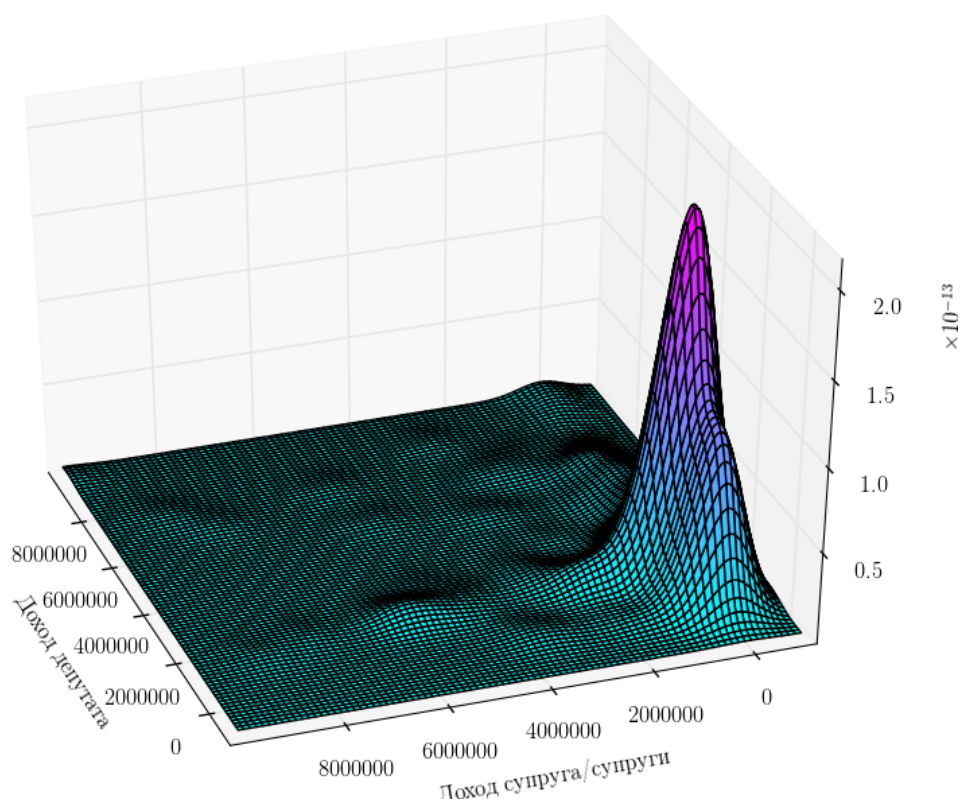


Рис. 19: Двумерная плотность

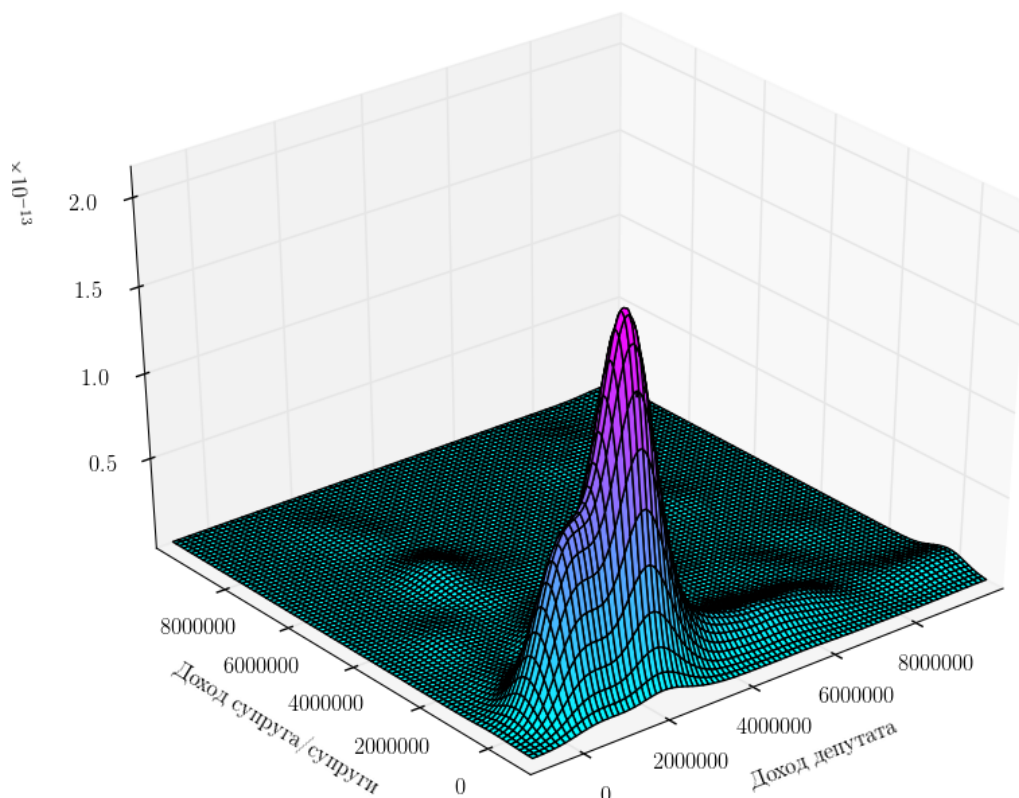


Рис. 20: Двумерная плотность

Таким образом, мы получили распределение, наибольшая плотность у которого достигается в окрестности точки $(2\,000\,000, 0)$, что показывает, что большинство депутатов имеют высокие доходы, но либо у них нет супруг/супругов, либо у них не заявлено доходов.

Также видно, что некоторое повышение плотности идет вдоль прямых $x = 2\,000\,000$ и $y = 0$, и намного меньше по прямой $y = x$ и вообще в области между прямыми $x = 2\,000\,000$ и $y = 0$.

В свою очередь, есть депутаты, у которых либо нет супруг/супругов, либо доход последних равен 0. Проанализируем эту ситуацию. Для этого обратимся к одномерной плотности. Для семей депутатов, доход в которых формируется из доходов обоих супругов, получим доход как сумма доходов. Для депутатов, которые являются единственным источником в семье, доход семьи будет тождественно равным доходу депутата. Таким образом, получим данные:

1. Данные по **всем** семьям —
<http://cdn.bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/DataBoth.txt>
2. Данные по доходам семьи, депутаты в которых являются **единственным** источником дохода в семье —
<http://cdn.bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/DataSolely.txt>
таких семей — 116
3. Данные по доходам семьи, депутаты в которых **не** являются **единственным** источником дохода в семье —
<http://cdn.bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/DataNotSolely.txt>
таких семей — 326

В результате получим плотности распределения доходов депутатов в зависимости от «типа семьи». На следующем графике взяты плотности по всей шкале дохода, даже с экстремально большим доходом.

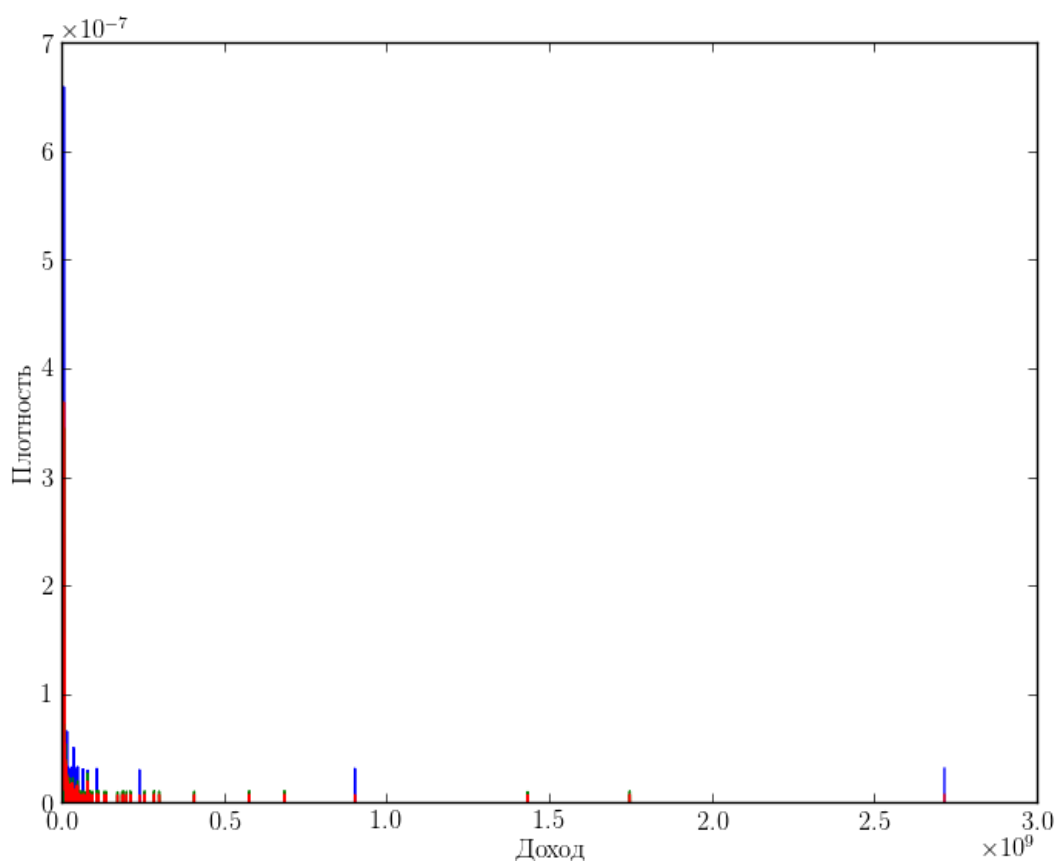


Рис. 21: Плотность распределения дохода семей депутатов, полная шкала доходов

Как видно из графика, учёт экстремально больших доходов снижает наглядность графика. Поэтому на следующем графике рассмотрим шкалу до 10 000 000 руб. дохода. Как видно из графиков, доход в размере около 2 000 000 руб. характерен для всех типов семей, что также подтверждается и двумерной плотностью.

Плотность в доходах до 2 000 000 руб. намного выше, чем выше 2 000 000. Особенно это касается депутатов-«одиночек». Более того, семьи, где зарабатывают оба супруга показывают более постоянную плотность. Не смотря на это, поведение распределения плотности доходов двух типов семей (где депутат зарабатывает один и где зарабатывают оба супруга) очень похожи.

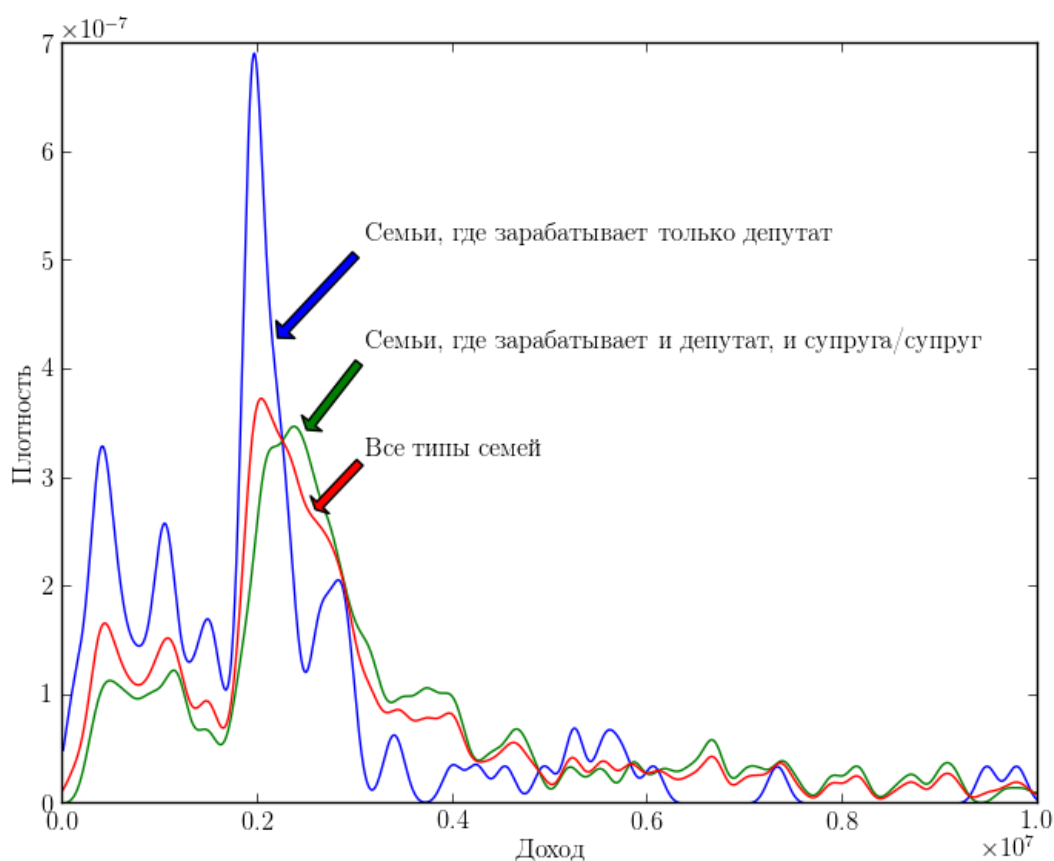


Рис. 22: Плотность распределения дохода семей депутатов, ограниченная шкала доходов

Теперь обратимся к партийной принадлежности депутатов. Добавим в код поиск партии на странице. Найдя в html-коде фразу типа «Депутат Государственной Думы избран в составе федерального списка кандидатов, выдвинутого Политической партией «Коммунистическая партия Российской Федерации», добавим по ней поиск. Закодируем так, что если депутат — член партии «Единая Россия», то у него будет (теперь уже в трёхмерном векторе дохода) третья переменная — «1», если «КПРФ» — «2», «ЛДПР» — «3», «Справедливая Россия» — «4».

Получим новые данные —

<http://cdn.bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/DataParties.txt>

Проанализируем по партиям.

	Единая Россия	КПРФ	ЛДПР	Справедливая Россия
Количество членов	23	92	55	64
Максимальное значение	2 712 131 087.00	403 269 294.29	167 143 780.00	1 743 518 894.0
Среднее значение	45 379 369.64	12 184 069.99	11 618 146.45	38 748 313.75

Из таблицы видно, что наибольший максимальный и средний доходы у членов партии «Единая Россия».

Члены партии ЛДПР, наоборот, отличаются минимальными значениями этих показателей.

Теперь строим плотности доходов семей депутатов в зависимости от партийной принадлежности. Первый график показывает плотность при всех уровнях доходов (шкала от 0 до 2 800 000 000 руб.). $\sigma = 100\,000$ руб.

Из него видно, что, как и раньше, вся плотность скапливается на уровне дохода в 2 000 000 руб. На участках с доходом выше преобладают доходы членов партии «Единая Россия» относительно равномерно на всем участке, «Справедливая Россия» — «всплеск» около 1 700 000

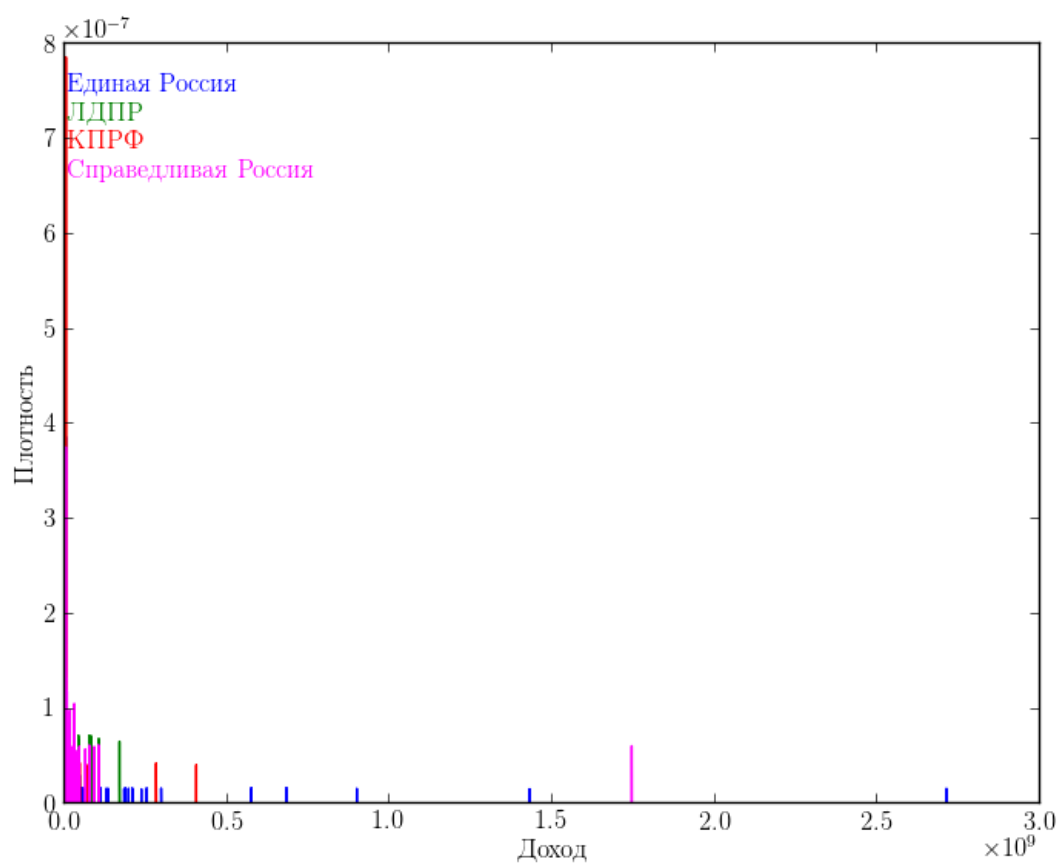


Рис. 23: Плотность распределения дохода семей депутатов, в зависимости от партийной принадлежности, шкала не ограничена

000 руб. Относительно меньше доходы у членов КПРФ — меньше 500 000 000 руб. и ЛДПР — меньше 200 000 000 руб.

Теперь посмотрим плотность доходов на отрезке от 0 до 10 000 000 руб. $\sigma = 100\,000$ руб.

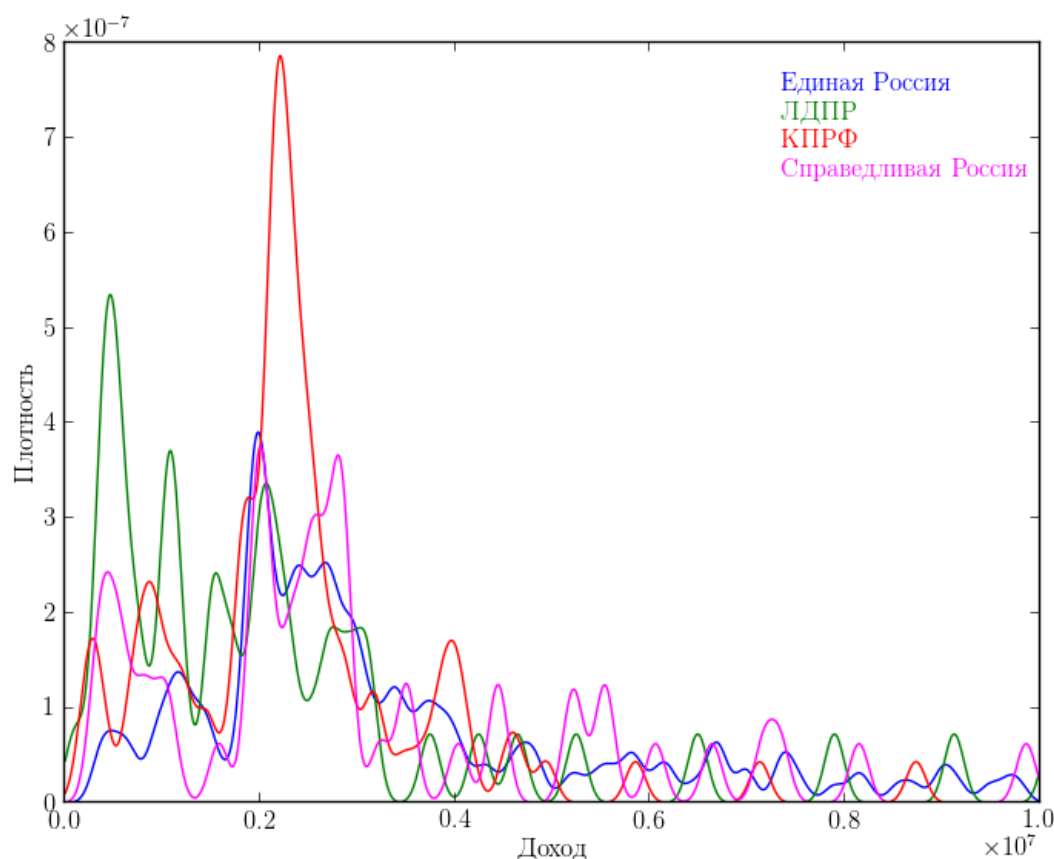


Рис. 24: Плотность распределения дохода семей депутатов, в зависимости от партийной принадлежности, доходы до 10 000 000 руб.

Наблюдаем, что тот типичный уровень в 2 000 000 руб. из предыдущего анализа показывают доходы членов партии КПРФ, в свою очередь, те «заниженные» показатели (высокая плотность ниже 2 000 000 руб.) — результат влияния доходов членов партии ЛДПР. А доходы членов партий «Единая Россия» и «Справедливая Россия» относительно равномерно распределены на этом отрезке с типичным всплеском в доходе — 2 000 000 руб.

Вкупе с предыдущим рисунком можно попытаться сделать выводы, что доходы членов партии «Единая Россия» распространены относительно равномерно, в то же время они всех выше. Менее равномерно распределены и меньше по величине доходы членов партии «Справедливая Россия». Доходы членов партии КПРФ распределены наименее равномерно, с резкими всплесками в районе 2 000 000 руб., а также около 300 000 000 руб. и 400 000 000 руб. Менее всего — доходы членов партии ЛДПР, их распределение имеет свой максимум в районе 500 000 руб.

5 Выводы

Итак, в ходе этой работы мы рассмотрели один из методов сбора информации из сети Интернет. У него есть ряд преимуществ, в том числе: использование языка программирования Python, который не сложен для изучения и подходит для начальных этапов работы с информацией, также к нему можно загружать библиотеки для обработки данных и представления их в наглядном виде. Мы рассмотрели поведение игроков на аукционе типа «минимальная уникальная ставка». Распределение ставок близко к экспоненциальному распределению - оно асимметрично и имеет длинный хвост - действительно, пик ставок приходится на минимальные ставки, так как обязательно условие победы - чтобы ставка была наименьшей, длинный хвост показывает, что люди ставят также ставки выше, чтобы удовлетворить следующее условие - уникальность ставки. Также стремление к уникальности показывают «выбросы» на ставках больше минимальных, выше экспоненциального распределения.

Далеко не все аукционы показывают близкое к экспоненциальному распределение. Ряд аукционов имеет более толстый хвост, или не остроконечный пик.

Рассмотрение доходов семей членов Государственной Думы показывает нам, что преобладает доход непосредственно депутата Государственной Думы. Хотя есть случаи, когда доход супруга/-супруги выше. Случаи же когда доход в семье распределяется между супругами относительно редки.

Проанализировав поведение распределения дохода в зависимости от типа семьи (где один источник дохода — депутат, и где два источника дохода), мы увидели, что с учетом некоторых изменений (а именно более сглаженных графиков у семей второго типа) распределение плотности доходов схожи.

При анализе с учетом партийной принадлежности, мы выделили характерные особенности распределения дохода членов каждой партии. Можно выделить относительно «более богатые» и «менее богатые» партии. Хотя выделить какие-то конкретные тенденции нельзя.

С другой стороны, мы не можем делать окончательные выводы, потому что распределение доходов очень широкое и трудно поддающееся анализу.

6 Список литературы

1. Python. Самое необходимое. - СПб.: БХВ-Петербург, 2011. - 416 с.
2. <http://docs.python.org/tutorial/>
3. <http://www.crummy.com/software/BeautifulSoup/bs4/doc/>
4. <http://matplotlib.sourceforge.net/contents.html>

7 Приложение

Ссылки на части кода:

1. Код, который парсит сайт с аукционами, собирает информацию, которую впоследствии представляет графически в виде плотности. Кроме того, выводит собранные данные по аукционам и выигрышным ставкам.
https://bitbucket.org/nikitapetrov/lub_auctions/downloads/auction.py
2. Код, который обрабатывает информацию по выигрышным ставкам.
https://bitbucket.org/nikitapetrov/lub_auctions/downloads/win_analyses.py

3. Код, который парсит сайт Государственный Думы, собирает информацию и выводит данные.
<https://bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/duma.py>
4. Код, который также парсит сайт Государственной Думы, кроме дохода сохраняет еще и партийную принадлежность
https://bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/duma_parties.py
5. Код, который выводит статистику по данным, полученным за счет парсинга сайта Государственной Думы. Также выводит отдельно данные по разным типам семей.
https://bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/duma_data_stat.py
6. Код, который обрабатывает и выводит графически информацию по доходам в зависимости от типа семей.
https://bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/duma_distr_fam.py
7. Код, который обрабатывает информацию и выводит графически в зависимости от партийной принадлежности.
https://bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/duma_parties_analyses.py

Данные, собранные и используемые в процессе работы. Актуальные на 26 июня 2012 г..

1. https://bitbucket.org/nikitapetrov/lub_auctions/downloads/Data29.txt — данные по 29 аукционам.
2. https://bitbucket.org/nikitapetrov/lub_auctions/downloads/DataW.txt — данные по выигрышным ставкам.
3. <https://bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/Data.txt> — двумерные данные по доходам членов Государственной Думы и их супруг/супругов.
4. <https://bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/DataParties.txt> — двумерные данные по доходам членов Государственной Думы и их супруг/супругов с учетом партийной принадлежности.
5. <https://bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/DataNotSolely.txt> — одномерные данные по семьям, где депутат — не единственный источник дохода.
6. <https://bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/DataSolely.txt> — одномерные данные по семьям, где депутат — единственный источник дохода.
7. <https://bitbucket.org/nikitapetrov/dumadeputiesincomes/downloads/DataBoth.txt> — одномерные данные по обоим типам семей.

Графики распределения ставок на аукционах:

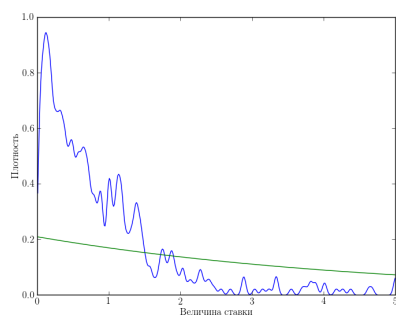


Рис. 25: Плотность распределения ставок на аукционе

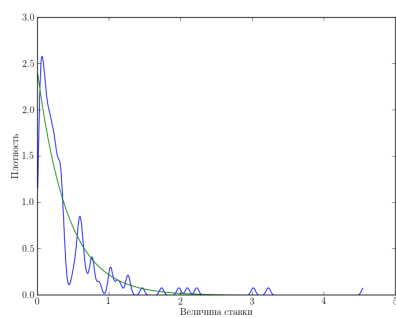


Рис. 26: Плотность распределения ставок на аукционе

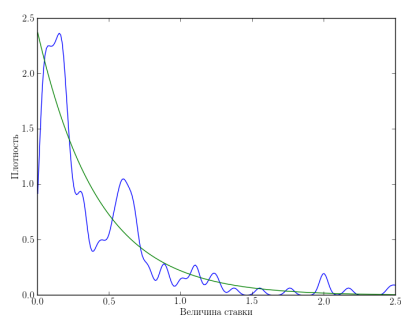


Рис. 27: Плотность распределения ставок на аукционе

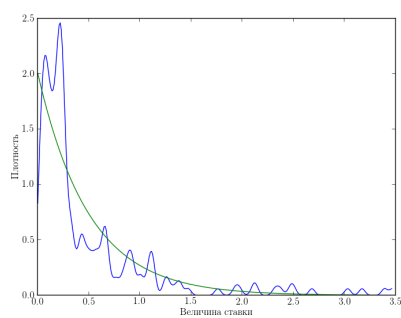


Рис. 28: Плотность распределения ставок на аукционе

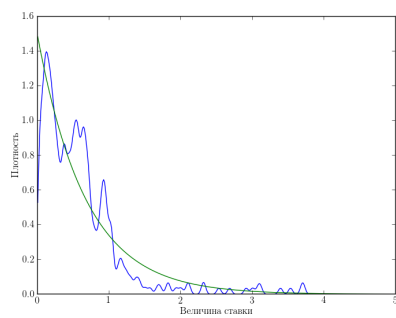


Рис. 29: Плотность распределения ставок на аукционе

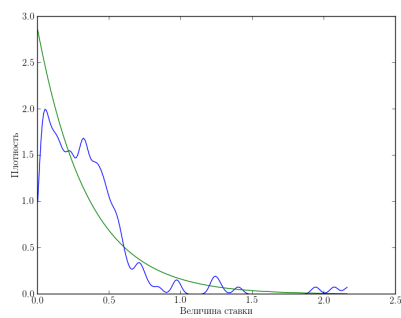


Рис. 30: Плотность распределения ставок на аукционе

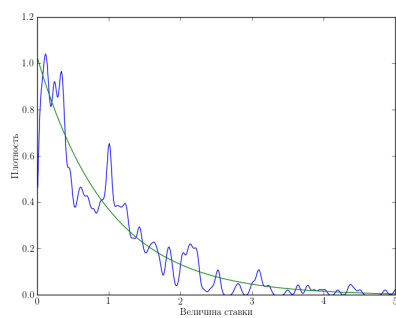


Рис. 31: Плотность распределения ставок на аукционе

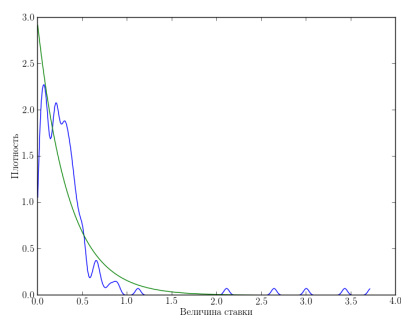


Рис. 32: Плотность распределения ставок на аукционе

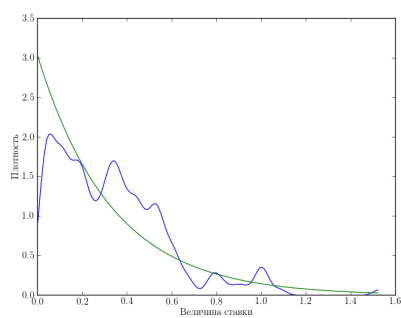


Рис. 33: Плотность распределения ставок на аукционе

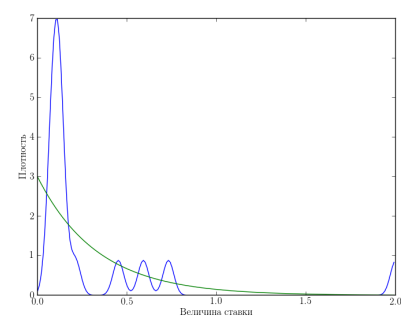


Рис. 34: Плотность распределения ставок на аукционе

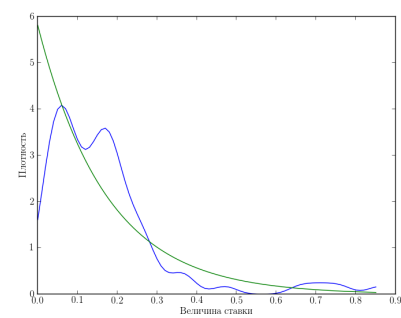


Рис. 35: Плотность распределения ставок на аукционе

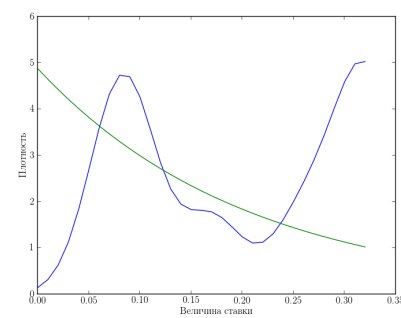


Рис. 36: Плотность распределения ставок на аукционе

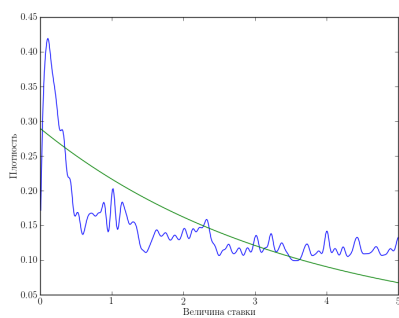


Рис. 37: Плотность распределения ставок на аукционе

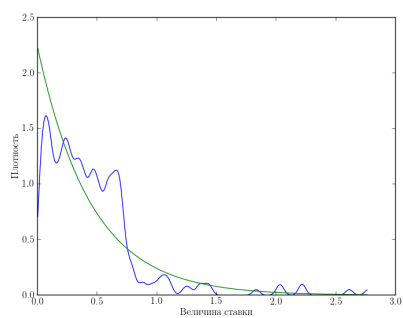


Рис. 38: Плотность распределения ставок на аукционе

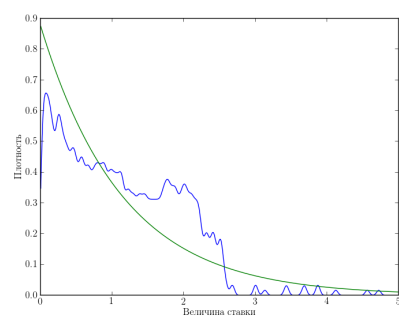


Рис. 39: Плотность распределения ставок на аукционе

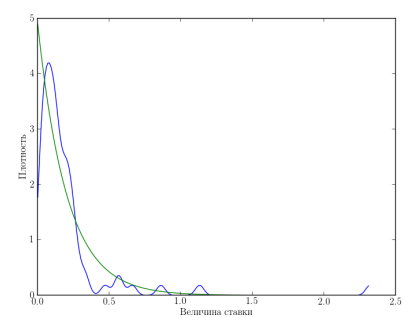


Рис. 40: Плотность распределения ставок на аукционе

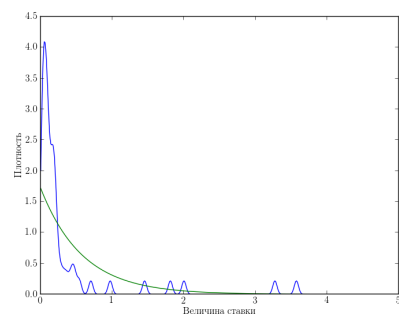


Рис. 41: Плотность распределения ставок на аукционе

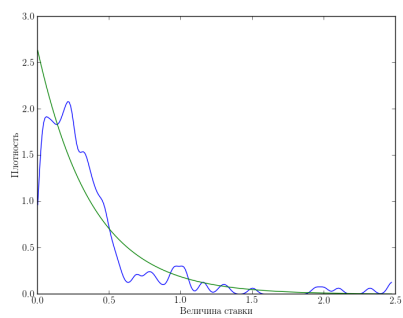


Рис. 42: Плотность распределения ставок на аукционе

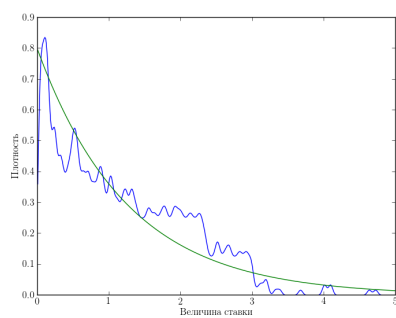


Рис. 43: Плотность распределения ставок на аукционе

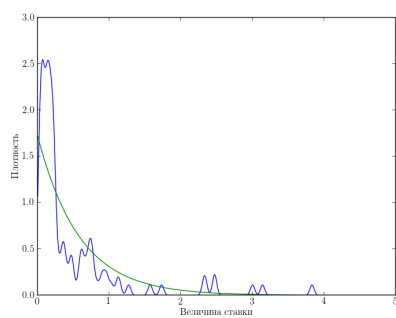


Рис. 44: Плотность распределения ставок на аукционе

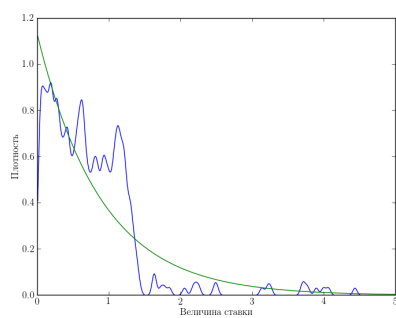


Рис. 45: Плотность распределения ставок на аукционе

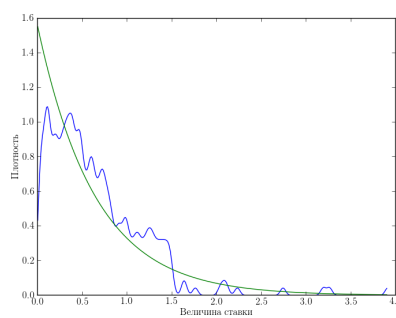


Рис. 46: Плотность распределения ставок на аукционе

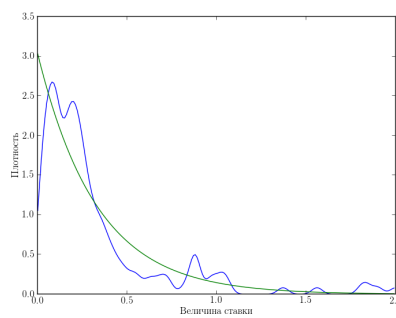


Рис. 47: Плотность распределения ставок на аукционе

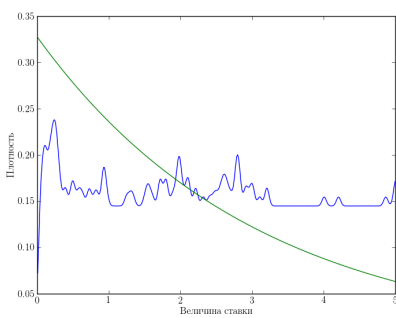


Рис. 48: Плотность распределения ставок на аукционе

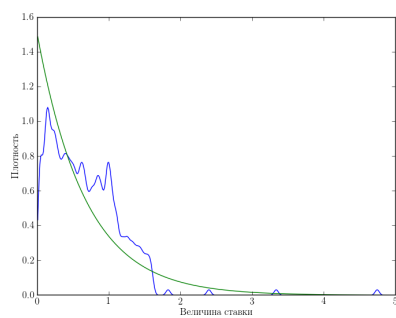


Рис. 49: Плотность распределения ставок на аукционе

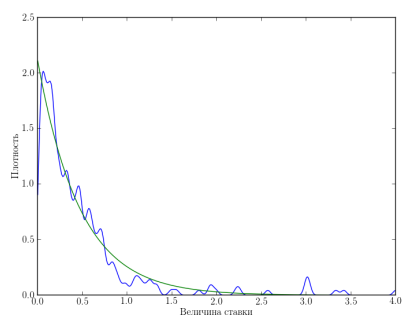


Рис. 50: Плотность распределения ставок на аукционе

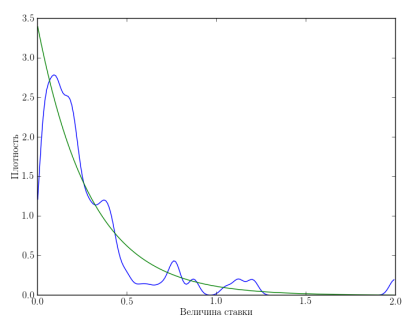


Рис. 51: Плотность распределения ставок на аукционе

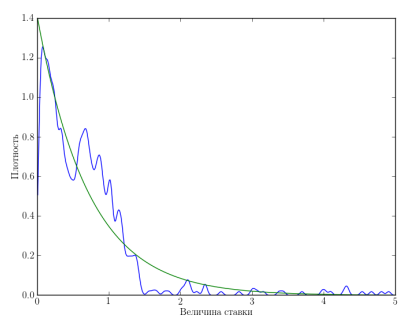


Рис. 52: Плотность распределения ставок на аукционе

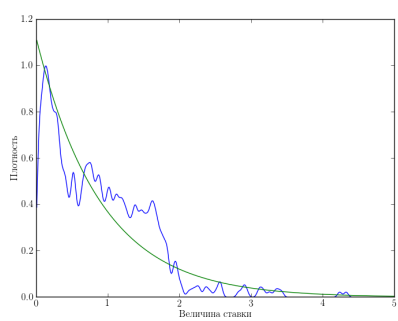


Рис. 53: Плотность распределения ставок на аукционе