История опытов с бросанием монеты
Часть 3. Опыт Джевонса[1]

Г.И.Фалин, д.ф.м.н., проф.
кафедра теории вероятностей
механико-математический факультет
МГУ им М.В.Ломоносова (Москва)
http://mech.math.msu.su/~falin

Во третьей части статьи мы кратко рассказываем о жизни и научной работе английского математика и логика Джевонса. Используя оригинальную публикацию этого выдающегося учёного, мы описываем его опыт с бросанием монеты. Строгий статистический анализ результатов этого опыта (который заключался в бросании комплекта из 10 монет) показывает, что его нельзя интерпретировать как эксперимент с бросанием одной монеты.

Ключевые слова: бросание комплекта монет, опыты, устойчивость частот

[1] Г.И.Фалин. История опытов с бросанием монеты. Часть 3. Опыт Джевонса. *Математика в школе*, 2015, №1, стр.52-58.

Уильям Стенли Джевонс (William Stanley Jevons, 1.09.1835 – 13.08.1882) – английский экономист и логик. Интересные детали жизни Джевонса можно найти в его «Письмах и журнале» [1], изданных в 1886 году его женой.

Уильям Стенли Джевонс родился в семье ливерпульского торговца железом Томаса Джевонса, который увлекался техническими новшествами и привил сыну любовь к науке. В 1850 году его отправили учиться в Университетский колледж Лондона, где его больше всего интересовали химия и ботаника. В конце 1853 года он получил предложение о работе пробирщиком на монетном дворе в Австралии и на следующий год уехал туда на 5 лет. Решение это было естественным, т.к. в январе 1848 года фирма, которой владели его дед и отец, разорилась и материальное положение семьи ухудшилось. В 1859 году Джевонс вернулся в Лондон для окончания университетского образования, но теперь решил специализироваться по экономике, в развитие которой он внёс выдающийся вклад. Джевонс является одним из основоположников применения математических методов для изучения экономических явлений. В самой знаменитой своей книге по экономике, «Теория политэкономии», он пишет: «…экономика, если она вообще хочет быть наукой, должна быть математической наукой». В память о заслугах Джевонса в развии экономической науки Институт изучения закона конкуренции и экономики, основанный в 2006 году в Университетском колледже Лондона, получил его имя. Джевонс сохранил интерес и к естественным наукам и, в частности, в 1869 году сконструировал механический компьютер для вывода логических следствий – «логическое пианино» Джевонса (см. рис. 2). Это «пианино» сейчас хранится в Музее истории науки в Оксфорде.

В 1874 году Джевонс опубликовал одну из самых знаменитых своих книг – «Основы науки: трактат о логике и научном методе» [2] (её титульная страница приведена на рис.3). В 1881 году эту книгу перевели на русский язык, а недавно русский перевод переиздали [3].

Именно в этой книге, на стр. 238 (стр. 200 русского перевода [3]), описан опыт с бросанием монеты, упоминаемый в школьных учебниках (см.рис.4). Что же мы там видим? В отличие от того, что пишут в школьных учебниках, Джевонс вовсе не бросал монету 20480 раз и не подсчитывал число выпадений орла. На самом деле Джевонс провёл две серии экспериментов. Каждая серия состояла из n=1024 экспериментов, которые заключались в бросании комплекта из 10 монет. После каждого бросания Джевонс регистрировал общее число выпавших орлов. Таким образом, в качестве элементарных событий рассматривались события A_k=«выпало k орлов и $10-k$ решек», $k=0,1,\ldots 10$. Их теоретические вероятности образуют так называемое биномиальное распределение: $P\left(A_k\right)=\frac{1}{2^{10}}C_{10}^k\equiv\frac{1}{2^{10}}\cdot\frac{10\cdot 9\cdot\ldots\cdot(10-k+1)}{1\cdot 2\cdot\ldots\cdot k}$ (в числителе и знаменателе последней дроби стоит по k сомножителей), так что, например, $P\left(A_0\right)=\frac{1}{2^{10}}$, $P\left(A_1\right)=\frac{10}{2^{10}}$, $P\left(A_2\right)=\frac{45}{2^{10}}$, $P\left(A_3\right)=\frac{120}{2^{10}}$ и т.д.. Ожидаемое число N_k^* наступлений события A_k можно вычислить как произведение числа проведённых экспериментов $n=1024=2^{10}$ на вероятность $P\left(A_k\right)$: $N_k^*=\frac{10\cdot 9\cdot\ldots\cdot(10-k+1)}{1\cdot 2\cdot\ldots\cdot k}$ (число n было взято равным 1024 именно потому, что в знаменателях дробей, которые задают вероятности $P\left(A_k\right)$, стоит число $2^{10}=1024$) . Величины N_k^* приведены во втором столбце Таблицы 1. В третьем и четвёртом столбцах указано, сколько раз Джевонс наблюдал появление событий A_k в первой и второй сериях экспериментов соответственно.

Конечно, бросание n=1024 раза комплекта из десяти монет можно *интерпретировать* как бросание $10n = 10240$ раз одной монеты, а две таких серии экспериментов – как бросание одной монеты 20480 раз (позже мы выясним, что на самом деле это не всегда законно). Поскольку вероятность появления орла в результате одного бросания равна 0,5, ожидаемое из теории число орлов при 10240 бросаниях монеты равно 5120, а при 20480 бросаниях – 10240. Наблюдаемое по результатам серии экспериментов общее число появлений орла будет равно $0 \cdot N_0 + 1 \cdot N_1 + \ldots + 10 \cdot N_{10}$. В опыте Джевонса общее число выпавших орлов равно 5130 для первой серии бросаний и 5222 для второй, так что общее число орлов в двух сериях равно 10352 (отметим, что в книге Джевонса перепутаны первая и вторая серия и, кроме того, ошибочно указано, что в первой серии общее число орлов равно 5131, а общее число орлов в двух сериях равно 10353; эти опечатки остались и в русском переводе [3]). Частота выпадения орла в первой серии экспериментов равна равна $\frac{5130}{10240} \approx 50.1\%$, во второй – $\frac{5222}{10240} \approx 51\%$, а по результатам двух серий – $\frac{10352}{20480} \approx 50.55\%$. Кстати, в школьных учебниках сказано, что в опыте Джевонса по результатам двух серий частота выпадения орла равна 50,68% – я думаю, авторы переписали это число (как и всю таблицу, приведённую в первой части нашей статьи) из книги [4], стр.28, не потрудившись посмотреть книгу самого Джевонса и самим во всём разобраться.

Результаты предыдущего параграфа показывают, что, как пишет Джевонс, «согласие с теорией довольно хорошее».

Но не всё так просто. Согласие с теорией хорошее, *если* эксперимент Джевонса *действительно* заключался в бросании одной монеты 20480 раз. Но, во-первых, он проводил *две* серии экспериментов, а не одну. Во-вторых, и это самое главное, бросался комплект из 10 монет. Казалось бы, какая разница? «Очевидно», что бросание n=1024 раза комплекта из десяти монет *можно интерпретировать* как бросание $10n = 10240$ раз одной монеты, а две таких серии экспериментов – как бросание одной монеты 20480 раз. Но не всем это было очевидно. На стр. 244-245 своей книги [2] (глава X «Теория вероятностей», раздел «Трудности теории») Джевонс даже пишет: «Очень любопытно, как часто самые проницательные и мощные умы сбивались при расчёте вероятностей… Д'Аламбер… постоянно впадал в ошибки, которые должны умалять вес его мнений. Он не мог понять, например, что когда монеты подбрасывают последовательно вероятности такие же, как когда монеты бросают одновременно.» Но из курса теории вероятностей хорошо известно, что точка зрения Д'Аламбера ближе к истине. Как пишет В.Феллер в своём классическом учебнике [5], стр. 85, «Мы столкнёмся с теоретическими заключениями, которые окажутся не просто неожиданными, но будут прямо противоречить интуиции и здравому смыслу. Они покажут, что широко распространённые представления о случайных флуктуациях лишены основания и что смысл закона больших чисел часто неправильно истолковывается. Например, в различных приложениях предполагается, что наблюдения за результатами бросаний одной и той же монеты в течение длительного промежутка времени будут давать те же статистические характеристики, что и наблюдение результатов очень большого числа независимых бросаний в данный момент. Это не так.»

Имея в виду ограниченную задачу о частоте появления орла, в этом заочном споре с Д'Аламбером Джевонс прав, *но только при условии, что каждая из десяти монет выпадает «орлом» вверх независимо от других* (иначе говоря, монеты хорошо перемешиваются), *причём для каждой из десяти монет вероятность появления «орла» равна* 0.5. Если эта гипотеза верна, то вероятности событий A_k=«выпало k орлов и $10-k$ решек», $k = 0,1,\ldots 10$, должны образовывать биномиальное распределение:

$P\left(A_k\right)=\frac{1}{2^{10}}C_{10}^k$. Проверим, подтверждают ли результаты двух серий экспериментов Джевонса эту гипотезу. При этом мы будем работать с числами, полученными Джевонсом, никак их не интерпретируя.

Как мы рассказывали во второй части нашей статьи, посвящённой опыту де Моргана, степень близости наблюдаемых частот и соответствующих вероятностей характеризуется суммарной относительной погрешностью $X^2=\sum_i \frac{1}{N_i^*}\left(N_i-N_i^*\right)^2$. Если величина X^2 слишком велика, то гипотезу о том, что распределение числа орлов при бросании комплекта из 10 монет имеет биномиальное распределение $P\left(A_k\right)=\frac{1}{2^{10}}C_{10}^k$, следует отвергнуть. Иначе говоря, если величина X^2 слишком велика, то есть веские основания сомневаться в том, что при в эксперименте независимо друг от друга бросались 10 правильных монет. Точная формулировка этого критерия связана с тем, что при большом числе экспериментов распределение случайной величины X^2 зависит только от числа r анализируемых событий и не зависит от их вероятностей; оно называется распределением χ^2 («хи-квадрат») с $r-1$ степенями свободы. Для опыта Джевонса число степеней свободы равно 10, а значение X^2 равно 27.62857 для первой серии экспериментов и 41.75784 для второй (эти числа стоят в правой нижней части Таблицы 1).

Таблица 1. Результаты опыта Джевонса

число орлов, i	ожидаемое число экспериментов с данным числом орлов, N_i^*	число экспериментов с данным числом орлов, первая серия, N_i	число экспериментов с данным числом орлов, вторая серия, N_i	«относительная погрешность», первая серия, $\frac{\left(N_i-N_i^*\right)^2}{N_i^*}$	«относительная погрешность», вторая серия, $\frac{\left(N_i-N_i^*\right)^2}{N_i^*}$
0	1	0	1	1	0
1	10	21	15	12.1	2.5
2	45	52	50	1.088889	0.555556
3	120	111	119	0.675	0.008333
4	210	201	197	0.385714	0.804762
5	252	257	232	0.099206	1.587302
6	210	181	190	4.004762	1.904762
7	120	129	123	0.675	0.075
8	45	57	73	3.2	17.42222
9	10	12	23	0.4	16.9
10	1	3	1	4	0
Итого	**1024**	**1024**	**1024**	$X^2=\mathbf{27.62857}$	$X^2=\mathbf{41.75794}$

Чтобы понять, большие ли это значения, нужно сделать небольшое отступление.

В теории вероятностей, рассматриваемой как наука о специфических закономерностях реального мира, фраза «событие A имеет вероятность $P(A)$» означает не только

1. возможность неограниченного повторения в неизменных условиях комплекса условий (эксперимента), относительно которого рассматривается событие A,

2. близость частоты наступления события A к числу $P(A)$ при большом числе n повторений эксперимента,

но и ещё одно утверждение *фундаментальной* важности:

3. если вероятность события близка к 1, то при проведении единичного эксперимента можно быть практически уверенным, что это событие наступит. Если же вероятность события близка к 0, то при проведении единичного эксперимента можно быть практически уверенным, что это событие не наступит.

Вот как эту фундаментальную мысль подчёркивают классики теории вероятностей и математической статистики:

А.Н.Колмогоров [6], §2: «Применение теории вероятностей к действительному миру опыта происходит по следующей схеме… Если P(A) очень мало, то можно практически быть уверенным, что при однократной реализации условий σ событие A не будет иметь места…»

Б.Л.ван дер Варден [7], стр. 37-38: «События, имеющие очень маленькую вероятность, следует рассматривать как почти невозможные; при однократной реализации условий, для которых эти события теоретически возможны, нельзя рассчитывать на их осуществление. На этом принципе основано вообще каждое практическое применение теории вероятностей.»

Вероятность ε, которой мы пренебрегаем на практике, называется *уровнем значимости*, а вероятность $\alpha = 1-\varepsilon$ – *коэффициентом доверия* (или *доверительной вероятностью*). На практике мы считаем, что событие, вероятность которого больше доверительной, обязательно происходит в единичном испытании. Точное значение уровня значимости зависит от характера решаемой задачи. Во многих естественно-научных, технических и экономических приложениях пренебрегают возможностью наступления событий, вероятность которых не превосходит 1%. В ряде случаев уровень значимости должен быть много меньше (например, если речь идёт о надёжности авиационной техники). Важно понимать, что какой бы маленькой ни была вероятность наступления события, оно всё равно может произойти, но происходит это настолько редко, что мы пренебрегаем этой возможностью.

Предположим, что для появившегося в нашем исследовании распределения «хи-квадрат» с 10 степенями свободы мы выбрали уровень значимости 1%. С помощью стандартной функции CHIINV пакета Micsoft Excel легко найти, что с вероятностью 99% наступает событие X^2<23.20925 (для вычисления числа 23.20925 нужно в ячейку таблицы Excel внести формулу =CHIINV(0.01,10) и нажать клавишу Enter). И для первой, и для второй серии экспериментов значения величины X^2 выходят за границу 23.20925. Поэтому с уровнем доверия 99% мы должны отклонить гипотезу о том, что в опыте Джевонса распределение числа орлов при бросании комплекта из 10 монет имеет биномиальное распределение $P\left(A_k\right)=\frac{1}{2^{10}}C_{10}^k$. Иначе говоря, можно быть практически уверенным в том, что в эксперименте Джевонса нельзя говорить о независимом друг от друга бросании 10 правильных монет. Его подтверждает и беглый взгляд на Таблицу 1 – слишком часто получалось большое число орлов или решек. Видимо, это связано с тем, что монеты плохо перемешивались.

Этот вывод имеет даже большую достоверность, чем 99%. С помощью стандартной функции CHIDIST пакета Micsoft Excel легко найти, что вероятность события $X^2 \geq 27.62857$ (справа стоит значение X^2 для первой серии экспериментов) примерно равна 0.2%, а вероятность события $X^2 \geq 41.75794$ (справа стоит значение X^2 для второй серии экспериментов) даже меньше 0.001%. Если представить, что кто-то каждый день n=1024 раз бросает комплект из десяти монет, то за тысячу лет будет лишь три дня когда статистика Пирсона $X^2 = \sum_i \frac{1}{N_i^*}\left(N_i - N_i^*\right)^2$ превысит уровень 41.75794. Вряд ли Джевонсу так не повезло – гораздо разумнее признать, что проверяемая гипотеза о характере проведённого эксперимента ошибочна.

Итак, имеются более чем веские основания отвергнуть гипотезу о том, что в опыте Джевонса каждая из десяти монет выпадает «орлом» вверх независимо от других. Следовательно, *интерпретировать* две серии опытов Джевонса как два эксперимента с бросанием 10240 раз одной монеты неправомерно (так что прав Д'Аламбер, а не Джевонс). Основанный на результатах этой *интерпретации* вывод о том, что «согласие с теорией довольно хорошее», в сущности, является результатом произвольного манипулирования опытными данными.

Литература

1. William Stanley Jevons. Letters and Journal of W. Stanley Jevons, edited by his Wife Harriet A. Jevons. London: Macmillan and Co., 1886.

2. W. Stanley Jevons. The Principles of Science: A Treatise on Logic and Scientific Method. Macmillan and Co, Special American Edition, New York, 1874.

3. Джевонс У.С. Основы науки: Трактат о логике и научном методе. Пер. с англ. Изд-во Либроком, 2011. ISBN 978-5-397-01852-4

4. Колмогоров А.Н., Журбенко И.Г., Прохоров А.В. Введение в теорию вероятностей. М.: Наука, 1982.

5. Феллер В. Введение в теорию вероятностей и её приложения. Том 1. М.:Мир, 1984.

6. А.Н.Колмогоров. Основные понятия теории вероятностей. 1-е изд. 1933 (нем), 1936 (рус.), 2-е изд., 1974г., 3-е изд. М.:Фазис, 1998.

7. Б.Л.ван дер Варден. Математическая статистика. М.: Изд-во иностранной литературы, 1960.

Рис. 1. Уильям Стенли Джевонс

Рис. 2 «Логическое пианино» Джевонса

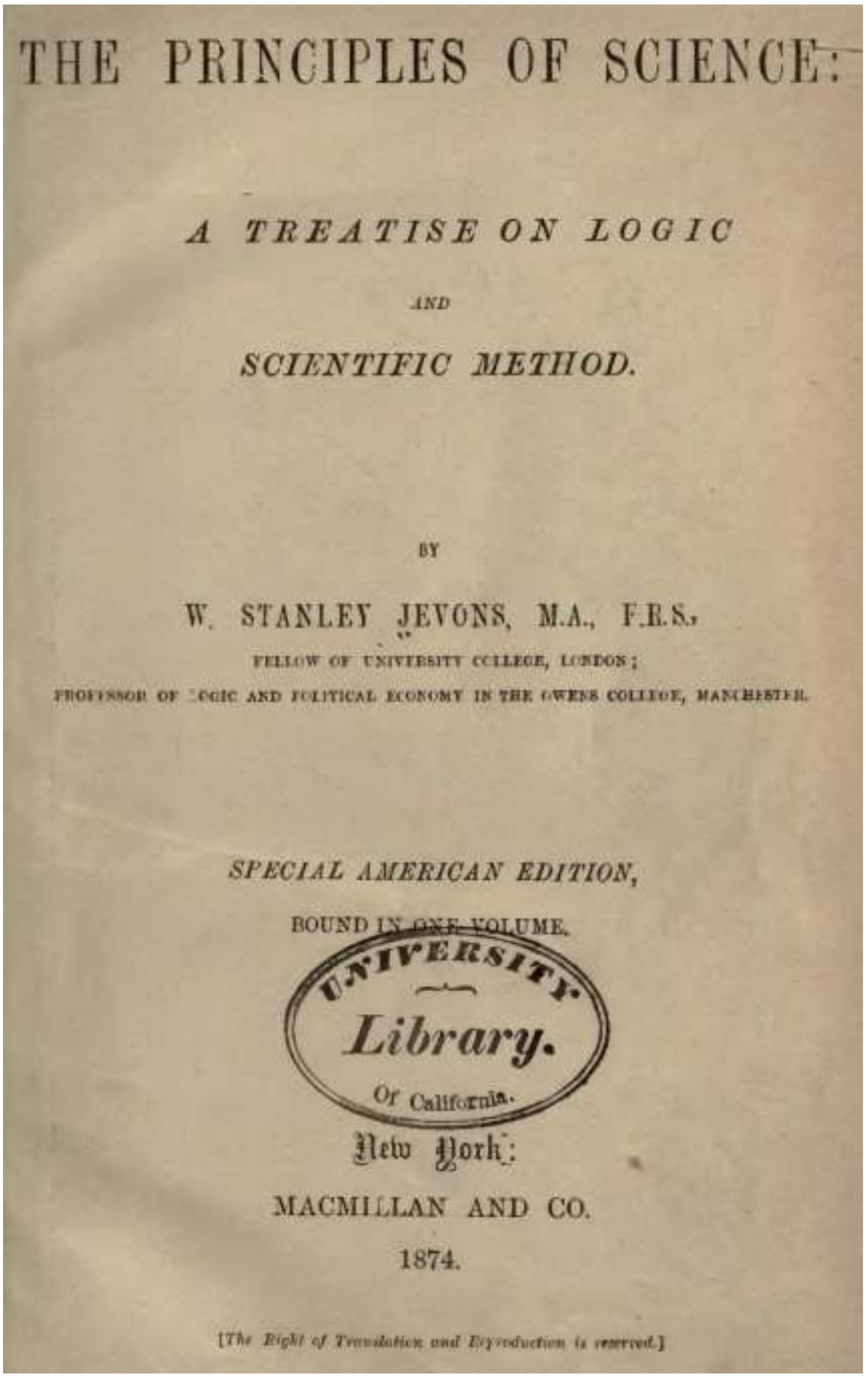

Рис. 3 Титульный лист книги Джевонса «Основы науки: трактат о логике и научном методе»

I have made a series of experiments in a third manner, which seemed to me even more interesting, and capable of more extensive trial. Taking a handful of ten coins, usually shillings, I threw them up time after time, and registered the numbers of heads which appeared each time. Now the probability of obtaining 10, 9, 8, 7, &c., heads is proportional to the number of combinations of 10, 9, 8, 7, &c., things out of 10 things. Consequently the results ought to approximate to the numbers in the eleventh line of the Arithmetical Triangle. I made altogether 2048 throws, in two sets of 1024 throws each, and the numbers obtained are given in the following table :—

Character of Throw.	Theoretical Numbers.	First Series.	Second Series.	Average.	Divergence.
10 Heads 0 Tail	1	3	1	2	+ 1
9 " 1 "	10	12	23	$17\frac{1}{2}$	+ $7\frac{1}{2}$
8 " 2 "	45	57	73	65	+20
7 " 3 "	120	129	123	126	+ 6
6 " 4 "	210	181	190	$185\frac{1}{2}$	−$25\frac{1}{2}$
5 " 5 "	252	257	232	$244\frac{1}{2}$	− $7\frac{1}{2}$
4 " 6 "	210	201	197	199	−11
3 " 7 "	120	111	119	115	− 5
2 " 8 "	45	52	50	51	+ 6
1 " 9 "	10	21	15	18	+ 8
0 " 10 "	1	0	1	$\frac{1}{2}$	− $\frac{1}{2}$
Totals.	1024	1024	1024	1024	− 1

The whole number of single throws of coins amounted to 10×2048 or 20,480 in all, one half of which or 10,240 should theoretically give head. The total number of heads obtained was actually 10,353, or 5222 in the first series, and 5131 in the second. The coincidence with theory is pretty close, but considering the large number of throws there is some reason to suspect a tendency in favour of heads.

Рис. 4 Часть страницы из книги Джевонса «Основы науки: трактат о логике и научном методе», где описан его опыт