МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования

«Брестский государственный университет имени А. С. Пушкина»

Физико-математический факультет

Кафедра алгебры, геометрии и математического моделирования

**Курсовая работа**

## **КОРРЕКТИРОВКА БУТСТРАПОВСКОЙ ИНТЕРВАЛЬНОЙ ОЦЕНКИ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ РАВНОМЕРНО РАСПРЕДЕЛЕННОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ**

Хайбулин Илья Марсельевич,

студент 3 курса

специальности «Экономическая кибернетика»

Брест 2014

ВВЕДЕНИЕ

Эконометрика и прикладная статистика бурно развиваются последние десятилетия. Серьезным (хотя, разумеется, не единственным и не главным) стимулом является стремительно растущая производительность вычислительных средств. Поэтому понятен острый интерес к статистическим методам, интенсивно использующим компьютеры. Одним из таких методов является так называемый "бутстрап", предложенный в 1977 г. Б. Эфроном из Станфордского университета (США).

Что же такое бутстрап?

Бутстрап - это практический компьютерный метод определения статистик вероятностных распределений, основанный на многократной генерации выборок методом Монте-Карло на базе имеющейся выборки. Позволяет просто и быстро оценивать самые разные статистики (доверительные интервалы, дисперсию, корреляцию и так далее) для сложных моделей.

В истории эконометрики было несколько более или менее успешно осуществленных рекламных кампаний. В каждой из них "раскручивался" тот или иной метод, который, как правило, отвечал нескольким условиям:

 по мнению его пропагандистов, полностью решал актуальную научную задачу;

 был понятен (при постановке задачи, при ее решении и при интерпретации результатов) широким массам потенциальных пользователей;

 использовал современные возможности вычислительной техники.

В стране с условиями отсутствия систематического эконометрического образования подобные рекламные кампании находили особо благоприятную почву, поскольку у большинства затронутых ими специалистов не было достаточных знаний в области методологии построения эконометрических моделей для того, чтобы составить самостоятельное квалифицированное мнение.

Речь идет о таких методах как бутстрап, нейронные сети, метод группового учета аргументов, робастные оценки по Тьюки-Хуберу, асимптотика пропорционального роста числа параметров и объема данных и др. Бывают локальные всплески энтузиазма, например, московские социологи в 1980-х годах пропагандировали так называемый "детерминационный анализ" - простой эвристический метод анализа таблиц сопряженности, хотя в Новосибирске в это время давно уже было разработано продвинутое программное обеспечение анализа векторов разнотипных.

Однако на фоне всех остальных рекламных кампаний судьба бутстрапа исключительна. Во-первых, признанный его автор Б. Эфрон с самого начала признавался, что он ничего принципиально нового не сделал. Его исходная статья называлась: "Бутстрап-методы: новый взгляд на методы складного ножа". Во вторых, сразу появились статьи и дискуссии в научных изданиях, публикации рекламного характера, и даже в научно-популярных журналах. Бурные обсуждения на конференциях, спешный выпуск книг. В 1980-е годы финансовая подоплека всей этой активности, связанная с выбиванием грантов на научную деятельность, содержание учебных заведений и т.п. была мало понятна отечественным специалистам.

1. ПОЛУЧЕНИЕ ИНТЕРВАЛЬНОЙ ОЦЕНКИ

1.1 Основные понятия и определения интервального оценивания

Задача интервального оценивания состоит в следующем: По данным выборки построить числовой интервал, относительно которого с заранее выбранной вероятностью можно сказать, что внутри этого интервала находится оцениваемый параметр.

Интервальная оценка - оценка, которая определяется двумя числами, а именно - концами интервала (), покрывающего оцениваемый параметр .

Требования, предъявляемые к статическим оценкам:

Для того чтобы статические оценки давали хорошее приближение оцениваемых параметров, они должны удовлетворять определенным требованиям:

. Несмещенность оценки (асимптотическая несмещенность);

Оценка  называется несмещенной оценкой параметра , если .

Оценка  называется асимптотически несмещенной оценкой , если

.

2. Состоятельность оценки;

Оценка  называется состоятельной оценкой параметра , если

, т.е. .

3. Эффективность оценки;

Оценка  называется эффективной оценкой , если она имеет наименьшую дисперсию среди всех несмещенных оценок данного параметра .

Пусть  - статистика, где  - точечная оценка неизвестного параметра . Чем меньше абсолютная величина разности, тем оценка точнее. Т.е. соотношение  определяет следующее: - называется точностью оценки. Чем меньше  - тем оценка точнее.

Доверительной вероятностью (надежностью оценки) называется вероятность , с которой выполняется соотношение

. [1, с. 22].

Статистические методы не позволяют категорически утверждать, что оценка  удовлетворяет неравенству . Можно лишь говорить о вероятности , с которой это соотношение  выполняется. Обычно  определяется в статистических таблицах() и задается в задачах заранее.

Доверительным интервалом называется интервал  который покрывает неизвестный параметр  с надежностью . Число  называют уровнем значимости.

.2 Построение доверительного интервала. Пусть  получены при n независимых наблюдений, проведенных при одинаковых условиях над генеральной совокупностью .

Математическое ожидание



Отсюда следует что доверительный интервал для неизвестного математического ожидания равен

(.

2. БУТСТРАП КАК МЕТОД ОПРЕДЕЛЕНИЯ СТАТИСТИК ВЕРОЯТНОСТНЫХ РАСПРЕДЕЛЕНИЙ

2.1 Возникновение бутстрапа

В институтах студентов учат интегрировать аналитически, а потом обнаруживается, что на практике интегралы почти все считают численными методами или проверяют, таким образом, аналитическое решение. В статистике тоже есть «нечестный» метод, который позволяет получить примерный ответ на многие практические вопросы без анализа, грубой компьютерной силой: бутстрап. Придумал и опубликовал его в 1979 году Брэдли Эфрон.

Суть метода: Допустим, есть у нас интернет-магазин, где мы торгуем разным товаром и привлекаем клиентов разными способами. Понятное дело, что мы постоянно что-то тестируем - расположение картинок и кнопок на странице, рекламный текст, баннеры на сайтах партнёров и так далее. В конечном счете, мы получаем свежие результаты - в тестовой группе из 893 пришедших у нас что-то купили 34, а в контрольной группе из 923 пришедших что-то купили 28.

Возникает вопрос - идти к начальству и говорить: «в тестовой группе соотношение числа купивших у нас что-либо к числу всех посетивших - 3.81%, в контрольной группе - 3.03%, налицо улучшение на 26%, где моя премия?» или продолжать сбор данных, потому что разница в 6 человек - ещё не статистика?

Эту задачу несложно решить аналитически. Видим две случайные величины (проценты в тестовой и контрольной группах). При большом количестве наблюдений биномиальное распределение похоже на нормальное. Нас интересует разность. Нормальное распределение бесконечно делимо, вычитаем математические ожидания и складываем дисперсии, получаем:

. математическое ожидание: 34/893-28/923 = 0.77%;

. дисперсию (34/893)\*(1-34/893)/893+(28/923)\*(1-28/923)/923.

Стандартное отклонение равно корню из дисперсии, в нашем случае 0.85%. Истинное значение с 95% вероятностью лежит в пределах плюс-минус двух стандартных отклонений от математического ожидания, то есть между -0.93% и 2.48%. Так что премия пока не будет, надо продолжать собирать данные.

Теперь решим эту же задачу методом бутстрапа. Основная идея такова: хорошо бы повторить наш эксперимент много раз и посмотреть на распределение результатов. Но мы это сделать не можем, поэтому будем действовать «нечестно» - «надёргаем» выборок из имеющихся данных и сделаем вид, что каждая из них - результат повторения нашего эксперимента.

2.2 Алгоритм бутстрапа

1. Выбираем наугад одно наблюдение из имеющихся.

. Повторяем пункт 1 столько раз, сколько у нас есть наблюдений. При этом некоторые из них мы выберем несколько раз, некоторые не выберем вообще - это нормально.

. Считаем интересующие нас метрики по этой новой выборке. Запоминаем результат.

Повторяем пункты 1-3 много раз. Например, 10 тысяч. Можно меньше, но точность будет хуже. Можно больше, но долго будет считать.

Теперь у нас есть распределение, на которое мы можем посмотреть или что-то по нему посчитать. Например, доверительный интервал, медиану или стандартное отклонение.

Следует обратить внимание на то, что мы не делаем никаких предположений о распределении чего-либо. Распределения могут быть разные. Алгоритм от этого не меняется. Однако если у распределения нет математического ожидания (такие встречаются) - бутстрап его не найдёт. То есть он найдёт математическое ожидание выборки, но не генеральной совокупности. То же касается ситуации, когда выборка маленькая.

Рассмотрим приведенный ниже пример написания бутстрапа на C++:

**#include** "stdafx.h"

**#include** <iostream>

**#include** <iomanip>

**#include** <math.h>

**#include** <stdlib.h>

**#include** <conio.h>

**typedef int** Data\_t;

**#define** ARRAY\_SIZE(x) **sizeof**(x)/**sizeof**(x[0])

**static** **double** bootstrap(**const** Data\_t\* data, **unsigned** n)

{

**unsigned** i;

**double** sum = 0;

**for** (i = 0; i < n; i++) {+= data[rand() % n];

}

**return** sum / n;

}

**static** **int** compare(**const** **void**\* a, **const** **void**\* b)

{

**if** (\*(**double**\*)a > \*(**double**\*)b) **return** 1;

**if** (\*(**double**\*)a < \*(**double**\*)b) **return** -1;

**return** 0;

}

**int** main(**int** argc, **char**\* argv[])

{\_t test[893] = { 0 };\_t control[923] = { 0 };

**unsigned** i;

**for** (i = 0; i < 34; i++) {[i] = 1;

}

**for** (i = 0; i < 28; i++) {[i] = 1;

}

**if** (argc == 2) {(atoi(argv[1]));

}

**double** t\_minus\_c[10000];

**for** (i = 0; i < ARRAY\_SIZE(t\_minus\_c); i++) {\_minus\_c[i] = bootstrap(test, ARRAY\_SIZE(test))

bootstrap(control, ARRAY\_SIZE(control));

}(t\_minus\_c, ARRAY\_SIZE(t\_minus\_c), **sizeof**(**double**), compare);("LCL=%g%%\n", 100. \* t\_minus\_c[250]);("UCL=%g%%\n", 100. \* t\_minus\_c[9750]);

\_getch();

**return** 0;

}

Опишем выше показанный код. Функция, которая делает выборку и считает по ней процент конверсии. Правильнее было бы использовать более точный алгоритм вычисления среднего значения, но для нашего примера это не важно.

**static double** bootstrap(**const** Data\_t\* data, **unsigned** n)

{

**unsigned** i;

**double** sum = 0;

**for** (i = 0; i < n; i++) {+= data[rand() % n];

}

**return** sum / n;

}

Функция сравнения для сортировки результатов

**static int** compare(**const void**\* a, **const void**\* b)

{

**if** (\*(**double**\*)a > \*(**double**\*)b) **return** 1;

**if** (\*(**double**\*)a < \*(**double**\*)b) **return** -1;

**return** 0;

}

Исходные данные

**int** main (**int** argc, **char**\* argv[])

{\_t test[893] = { 0 };\_t control[923] = { 0 };

**unsigned** i;

**for** (i = 0; i < 34; i++) {[i] = 1;

}

**for** (i = 0; i < 28; i++) {[i] = 1;

}

Инициализируем генератор псевдослучайных чисел параметром из командной строки. Если мы всё сделали правильно, то результаты не должны сильно плавать при изменении этого параметра.

**if** (argc == 2) {(atoi(argv[1]));

}

Сюда будем складывать результаты

**double** t\_minus\_c[10000];

Главный цикл

**for** (i = 0; i < ARRAY\_SIZE(t\_minus\_c); i++) {\_minus\_c[i] = bootstrap(test, ARRAY\_SIZE(test))

bootstrap(control, ARRAY\_SIZE(control));

}

Определяем 95% доверительный интервал: сортируем результаты, отбрасываем 2.5% снизу и столько же сверху, показываем результат.

qsort(t\_minus\_c, ARRAY\_SIZE(t\_minus\_c), sizeof(**double**), compare);("LCL=%g%%\n", 100. \* t\_minus\_c[250]);("UCL=%g%%\n", 100. \* t\_minus\_c[9750]);

\_getch();

**return** 0;

}

Проверяем найденное решение:



Похоже на теоритический результат (от -0,93% до 2,48%).

У этой задачи есть простое аналитическое решение, но у многих реальных задач его или нет вообще, или оно есть, но очень сложное. Представьте, что вместо процента конверсии нас интересует отношение прибыли от клиента к затратам на его привлечение. Распределение такой метрики вряд ли будет нормальным, и формулы перестанут укладываться в пару строчек. А бутстрап будет работать точно так же, достаточно поменять Data\_t на double и положить туда новые данные [4, с. 22].

3. ПОСТАНОВКА ЗАДАЧИ И РЕАЛИЗАЦИЯ АЛГОРИТМА

3.1 Постановка задачи

Интервальный метод оценивания параметров распределения случайных величин заключается в определении интервала (а не единичного значения), в котором с заданной степенью достоверности будет заключено значение оцениваемого параметра. Интервальная оценка характеризуется двумя числами - концами интервала, внутри которого предположительно находится истинное значение параметра. Иначе говоря, вместо отдельной точки для оцениваемого параметра можно установить интервал значений, одна из точек которого является своего рода "лучшей" оценкой. Интервальные оценки являются более полными и надежными по сравнению с точечными, они применяются как для больших, так и для малых выборок. Совокупность методов определения промежутка, в котором лежит значение параметра Т, получила название методов интервального оценивания.

Постановка задачи интервальной оценки параметров заключается в следующем:

Имеется выборка наблюдений () за нормально распределенной случайной величиной Х. Объем выборки n фиксирован. Объем бутстраповской выборки k=10 одинаков для любого объема выборки. Известно математическое ожидание m и стандартное отклонение  принимают значения 1 и 10.

Ограничения: выборка представительная, ее объем достаточен для оценки границ интервала. Представительная выборка - выборка, которая является (или считается) истинным отражением родительской популяции, то есть имеет тот же профиль признаков, например, возрастную структуру, классовую структуру, уровень образования.

3.2 Теоритическая реализация

Эта задача решается путем построения доверительного утверждения, которое состоит в том, что интервал  накрывает истинное значение параметра  с доверительной вероятностью не менее . Величины  и называются нижней и верхней доверительными границами (НДГ и ВДГ соответственно). Доверительные границы интервала выбирают так, чтобы выполнялось условие .

В доверительном утверждении считается, что статистики  и  являются случайными величинами и изменяются от выборки к выборке. Это означает, что доверительные границы определяются неоднозначно, существует бесконечное количество вариантов их установления.

На практике применяют два варианта задания доверительных границ:

. устанавливают симметрично относительно оценки параметра, т.е.

, ,

где  - стандартная ошибка среднего;  выбирают так, чтобы выполнялось доверительное утверждение. Следовательно, величина абсолютной погрешности оценивания  равна половине доверительного интервала;

. устанавливают из условия равенства вероятностей выхода за верхнюю и нижнюю границу: .

В общем случае величина  не равна . Для симметричных распределений случайного параметра  в целях минимизации величины интервала значения  и  равны, следовательно, в таких случаях оба варианта эквивалентны.

Нахождение доверительных интервалов требует знания вида и параметров закона распределения случайной величины . Для ряда практически важных случаев этот закон можно определить из теоретических соображений.

3.3 Реализация на языке программирования Pascal

бутстрап компьютерный интервал статистика

**Program** Ravn\_V\_K; **Uses** CRT;

{Исследование бутстраповских интервальных оценок для выборки с равномерным распределением}

**Const** T5 = 1.67655089261685; T2\_5 = 2.00957523712924;{t-статистики на 49 ст. св}

{Все константы расчитаны на выборку объёмом 50 элементов}

**Function** Ravn(A,B : real) : real;

**Begin**

Result:=A + (B-A)\*Random;

**end**;{Ravn}

**Var** X : **array** [1..1000] **of** real; {Выборка}, S,A\_,B\_ : real; {Параметры выборки}

N, M, G, B : integer; {Параметры эксперимента}

PA : **array** [1..3] **of** integer; {Оценки "Классические"}: **array** [1..3, 0..20] **of** integer; {Бутстраповские оценки}

Ab : **array** [1..10000] **of** real; {Бутстраповские статистики},J,K,L,Q,H : integer; {Вспомогательные переменные}\_, S\_, R : real; {Вспомогательные переменные}: Text; {Файл с входными, а затем и выходными данными}

St : String;

**BEGIN**

{Настройка модели}(Dat,'Param.txt'); ReSet(Dat);(Dat, N); ReadLn(Dat, A); ReadLn(Dat, S); ReadLn(Dat, M);(Dat, G); ReadLn(Dat, B); Close(Dat);('Бутстрап "равномерная" модель ('+intToStr(N)+' '+floatToStr(A)+' '+floatToStr(S)+')');; Randomize;

{Инициализация счетчиков}

**For** I:=1 **to** 3 **do begin**

PA[I]:=0;

**For** J:=0 **to** 20 **do** PbA[I,J]:=0;

**end**;\_ := A - S\*sqrt(3); B\_ := A + S\*sqrt(3);

{Прогоны модели}

**For** I := 1 **to** M **do Begin**(I **mod** 10)=0 **then Begin**

GoToXY(33,12); Write(I)

**end**;

{Генерируем выборку}

**For** J:=1 **to** N **do** X[J]:=Ravn(A\_,B\_);

{Реализуем классику}\_:=0;

**For** J:=1 **to** N **do Begin**

X\_+=X[J]; S\_+=Sqr(X[J])

**End**;\_/=N; S\_:=(S\_-Sqr(X\_)\*N)/(N-1);

**If** (A>X\_-Sqrt(S\_/N)\*T5) **then** Inc(PA[1]);

**If** (A<X\_+Sqrt(S\_/N)\*T5) **then** Inc(PA[2]);

**If** (A>X\_-Sqrt(S\_/N)\*T2\_5) **and** (A<X\_+Sqrt(S\_/N)\*T2\_5) **then** Inc(PA[3]);

{Реализуем бутстрап}

**For** K:=0 **to** 20 **do begin**

Q:=K+B;{Объём бутстраповской выборки}

**For** L := 1 **to** G **do begin**

Ab[L]:=0;

**for** J:=1 **to** Q **do Begin**

H:=Random(N)+1; Ab[L]+=X[H];

**end**;[L]/=Q;

**end**;

{Сортировка результатов}

**For** L:=1 **to** G-1 **do begin**

R:=Ab[L]; H:=L;

**for** J:=L+1 **to** G **do**R>Ab[J] **then begin**

R:=Ab[J]; H:=J

**end**;[H]:=Ab[L]; Ab[L]:=R

**end**;

**IF** (A>Ab[trunc(G\*0.05)]) **then** inc(PbA[1,K]);

**IF** (A<Ab[trunc(G\*0.95)]) **then** inc(PbA[2,K]);

**IF** (A>Ab[trunc(G\*0.025)]) **and** (A<Ab[trunc(G\*0.975)]) **then** inc(PbA[3,K]);

**end**;

**End**;{Конец прогонов}

{Вывод результатов}

St:='Левосторонняя';

**For** I:=1 **to** 3 **do begin**

Assign(Dat,'Result'+chr(48+I)+'.txt'); ReWrite(Dat);

WriteLn(Dat,St+' критическая область');(Dat,N,#9,'<- Объём выборки');

WriteLn(Dat,A,#9,'<- Матожидание');

WriteLn(Dat,S,#9,'<- Стандартное отклонение');(Dat,M,#9,'<- Количество прогонов');(Dat,G,#9,'<- Глубина бутстрапирования');

WriteLn(Dat);(Dat,#9,'A');

WriteLn(Dat,#9,PA[I]/M,#9,'<- Классическая оценка');

WriteLn(Dat,'Объём бутстрапа');

**For** J:=0 **to** 20 **do**

WriteLn(Dat,J+B,#9,PbA[I,J]/M, #9,'0,95');(Dat);

**If** i=1 **then** St:='Правосторонняя' **else** St:='Двухсторонняя'

**end**;

**END**.

.4 Результаты исследования

Рассмотрим результаты при заданных параметрах:

1. объем выборки: 10;

2. математическое ожидание: 0;

3. стандартное отклонение: 1;

4. количество прогонов: 100 000;

5. глубина бутстрапа: 1 000;

6. минимальный объем бутстрапа: 10.

После произведенных вычислений программным способом получим:

Таблица 1 - Оценка математического ожидания

|  |
| --- |
| Левосторонняя критическая область |
| 10 | <- Объём выборки |
| 0 | <- Матожидание |
| 1 | <- Стандартное отклонение |
| 100000 | <- Количество прогонов |
| 1000 | <- Глубина бутстрапирования |
|  |  |  |
|  | A |  |
|  | 0,94885 | <- Классическая оценка |
| Объём бутстрапа |  |  |
| 5 | 0,97452 | 0,95 |
| 6 | 0,9658 | 0,95 |
| 7 | 0,95614 | 0,95 |
| 8 | 0,94701 | 0,95 |
| 9 | 0,93801 | 0,95 |
| 10 | 0,92939 | 0,95 |
| 11 | 0,92124 | 0,95 |
| 12 | 0,91307 | 0,95 |
| 13 | 0,9047 | 0,95 |
| 14 | 0,89689 | 0,95 |
| 15 | 0,88915 | 0,95 |
| 16 | 0,88278 | 0,95 |
| 17 | 0,87627 | 0,95 |
| 18 | 0,87004 | 0,95 |
| 19 | 0,86427 | 0,95 |
| 20 | 0,85785 | 0,95 |
| 21 | 0,85284 | 0,95 |
| 22 | 0,84724 | 0,95 |
| 23 | 0,84215 | 0,95 |
| 24 | 0,83696 | 0,95 |
| 25 | 0,83269 | 0,95 |



Рисунок 1 - График значений математического ожидания

Где при объеме бутстрапа, который равен 7, мы получаем границу доверительного интервала, максимально приближенный к истинному значению. Проделаем данные наблюдения для разных объемов выборки, математических ожиданий и стандартных отклонений. Найдя все точки пересечения с надежностью, и построив по ним графики, мы можем увидеть зависимость оптимального объема бутстрапа от объема исследуемой выборки для разных параметров. К примеру, ниже приведен график зависимости для параметров 0 1 л:

Таблица 2 - Значения показателей объемов для параметров: 0 1 л

|  |  |  |  |
| --- | --- | --- | --- |
| 10 | 8 | 0,94701 | 0,95 |
| 20 | 18 | 0,94853 | 0,95 |
| 30 | 27 | 0,95038 | 0,95 |
| 40 | 38 | 0,95018 | 0,95 |
| 50 | 47 | 0,95084 | 0,95 |
| 60 | 58 | 0,9503 | 0,95 |
| 70 | 68 | 0,95029 | 0,95 |
| 80 | 79 | 0,95002 | 0,95 |
| 90 | 90 | 0,95053 | 0,95 |
| 100 | 98 | 0,95033 | 0,95 |
| 10 | 8 | 0,94701 | 0,95 |



Рисунок 2 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 1 1 л

Следовательно, для данных параметров объем бустрапа нужно брать приблизительно на 3 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение. В таком случае оценка получается несмещенной. Ниже будут приведены следующие зависимости объемов бутстрапа от объемов исследуемых выборок:

1. 0 1 л (Приложение A)

2. 0 1 п (Приложение Б)

3. 0 1 лп (Приложение В)

4. 0 10 л (Приложение Г)

5. 0 10 п (Приложение Д)

6. 0 10 лп (Приложение Е)

7. 1 1 л (Приложение Ж)

8. 1 1 п (Приложение И)

9. 1 1 лп (Приложение К)

10. 1 10 л (Приложение Л)

11. 1 10 п (Приложение М)

12. 1 10 лп (Приложение Н)

13. 10 1 л (Приложение П)

14. 10 1 п (Приложение Р)

15. 10 1 лп (Приложение С)

16. 10 10 л (Приложение Т)

17. 10 10 п (Приложение У)

18. 10 10 лп (Приложение Ф)

ЗАКЛЮЧЕНИЕ

В ходе выполненного исследования мы выяснили, что каждую нормально распределенную выборку можно бутстрапировать. Однако при этом объем бутстрапа нужно брать в среднем на 3 единицы меньше, чем объем исходной выборки. Только выполнив это условие, мы получим доверительный интервал, соответствующий заданной вероятности.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Баврин, И.И. Теория вероятностей и математическая статистика / И.И. Баврин. - М.: Высш. шк., 2005. - 160 с.

. Максимов, Ю.Д. Вероятностные разделы математики / Ю.Д. Максимов. - Изд.: Иван Федоров, 2001. - 592 с.

. Пугачев B.C. Теория вероятностей и математическая статистика: Учеб. Пособие / В.С. Пугачев. - 2-е изд., исправл. и дополн. - М.: Физматлит,2002. - 496 с.

. Электронная библиотека [Электронный ресурс] - Режим доступа: http://habrahabr.ru - Дата доступа: 25.01.2014.

ПРИЛОЖЕНИЯ

Приложение А

Таблица 3 - Значения показателей объемов для параметров: 0 1 л

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 8 | 0,94701 | 0,95 |
| 20 | 18 | 0,94853 | 0,95 |
| 30 | 27 | 0,95038 | 0,95 |
| 40 | 38 | 0,95018 | 0,95 |
| 50 | 47 | 0,95084 | 0,95 |
| 60 | 58 | 0,9503 | 0,95 |
| 70 | 68 | 0,95029 | 0,95 |
| 80 | 79 | 0,95002 | 0,95 |
| 90 | 90 | 0,95053 | 0,95 |
| 100 | 98 | 0,95033 | 0,95 |



Рисунок 3 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 1 1 л

Для данных параметров объем бутстрапа нужно брать приблизительно на 2 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.

Приложение Б

Таблица 4 - Значения показателей объемов для параметров: 0 1 п

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 8 | 0,9478 | 0,95 |
| 20 | 18 | 0,94803 | 0,95 |
| 30 | 27 | 0,95098 | 0,95 |
| 40 | 38 | 0,95018 | 0,95 |
| 50 | 47 | 0,95084 | 0,95 |
| 60 | 57 | 0,95062 | 0,95 |
| 70 | 66 | 0,95097 | 0,95 |
| 80 | 77 | 0,94962 | 0,95 |
| 90 | 86 | 0,95031 | 0,95 |
| 100 | 96 | 0,94995 | 0,95 |



Рисунок 4 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 0 1 п

Для данных параметров объем бутстрапа нужно брать приблизительно на 3-4 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.

Приложение В

Таблица 5 - Значения показателей объемов для параметров: 0 1 лп

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 7 | 0,94961 | 0,95 |
| 20 | 17 | 0,95131 | 0,95 |
| 30 | 27 | 0,95075 | 0,95 |
| 40 | 37 | 0,95023 | 0,95 |
| 50 | 47 | 0,95 | 0,95 |
| 60 | 57 | 0,95029 | 0,95 |
| 70 | 66 | 0,95075 | 0,95 |
| 80 | 76 | 0,95186 | 0,95 |
| 90 | 88 | 0,95032 | 0,95 |
| 100 | 96 | 0,94988 | 0,95 |



Рисунок 5 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 0 1 лп

Для данных параметров объем бутстрапа нужно брать на 3 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.

Приложение Г

Таблица 6 - Значения показателей объемов для параметров: 0 10 л

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 7 | 0,95189 | 0,95 |
| 20 | 17 | 0,95039 | 0,95 |
| 30 | 27 | 0,95073 | 0,95 |
| 40 | 37 | 0,94958 | 0,95 |
| 50 | 46 | 0,94981 | 0,95 |
| 60 | 57 | 0,9498 | 0,95 |
| 70 | 66 | 0,95082 | 0,95 |
| 80 | 77 | 0,94965 | 0,95 |
| 90 | 86 | 0,94998 | 0,95 |
| 100 | 96 | 0,95028 | 0,95 |



Рисунок 6 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 10 1 п

Для данных параметров объем бутстрапа нужно брать приблизительно на 3 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.

Приложение Д

Таблица 7 - Значения показателей объемов для параметров: 10 1 лп

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 7 | 0,94406 | 0,95 |
| 20 | 17 | 0,94675 | 0,95 |
| 30 | 27 | 0,94812 | 0,95 |
| 40 | 37 | 0,94814 | 0,95 |
| 50 | 46 | 0,94984 | 0,95 |
| 60 | 57 | 0,94936 | 0,95 |
| 70 | 66 | 0,94955 | 0,95 |
| 80 | 76 | 0,94979 | 0,95 |
| 90 | 86 | 0,94913 | 0,95 |
| 100 | 96 | 0,95028 | 0,95 |



Рисунок 7 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 10 1 лп

Для данных параметров объем бутстрапа нужно брать приблизительно на 3 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.

Приложение Е

Таблица 8 - Значения показателей объемов для параметров: 1 10 л

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 7 | 0,95176 | 0,95 |
| 20 | 18 | 0,94667 | 0,95 |
| 30 | 27 | 0,95034 | 0,95 |
| 40 | 38 | 0,94957 | 0,95 |
| 50 | 47 | 0,95046 | 0,95 |
| 60 | 58 | 0,94984 | 0,95 |
| 70 | 68 | 0,9505 | 0,95 |
| 80 | 78 | 0,95047 | 0,95 |
| 90 | 87 | 0,95007 | 0,95 |
| 100 | 96 | 0,94991 | 0,95 |



Рисунок 8 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 1 10 л

Для данных параметров объем бутстрапа нужно брать приблизительно на 2 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.

Приложение Ж

Таблица 9 - Значения показателей объемов для параметров: 1 10 п

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 7 | 0,95096 | 0,95 |
| 20 | 17 | 0,95013 | 0,95 |
| 30 | 27 | 0,95033 | 0,95 |
| 40 | 37 | 0,94946 | 0,95 |
| 50 | 47 | 0,94947 | 0,95 |
| 60 | 56 | 0,95068 | 0,95 |
| 70 | 66 | 0,95004 | 0,95 |
| 80 | 77 | 0,94988 | 0,95 |
| 90 | 86 | 0,9508 | 0,95 |
| 100 | 96 | 0,95071 | 0,95 |



Рисунок 9 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 1 10 п

Для данных параметров объем бутстрапа нужно брать приблизительно на 3 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.

Приложение И

Таблица 10 - Значения показателей объемов для параметров: 1 10 лп

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 7 | 0,94286 | 0,95 |
| 20 | 17 | 0,94693 | 0,95 |
| 30 | 27 | 0,94745 | 0,95 |
| 40 | 37 | 0,94852 | 0,95 |
| 50 | 46 | 0,94994 | 0,95 |
| 60 | 57 | 0,94942 | 0,95 |
| 70 | 66 | 0,95086 | 0,95 |
| 80 | 77 | 0,9496 | 0,95 |
| 90 | 86 | 0,95088 | 0,95 |
| 100 | 96 | 0,94988 | 0,95 |



Рисунок 10 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 1 10 лп

Для данных параметров объем бутстрапа нужно брать приблизительно на 3 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.

Приложение К

Таблица 11 - Значения показателей объемов для параметров: 10 10 л

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 7 | 0,95263 | 0,95 |
| 20 | 17 | 0,95026 | 0,95 |
| 30 | 28 | 0,94856 | 0,95 |
| 40 | 37 | 0,95063 | 0,95 |
| 50 | 47 | 0,95086 | 0,95 |
| 60 | 58 | 0,94924 | 0,95 |
| 70 | 68 | 0,9496 | 0,95 |
| 80 | 77 | 0,94997 | 0,95 |
| 90 | 88 | 0,94986 | 0,95 |
| 100 | 97 | 0,94955 | 0,95 |



Рисунок 11 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 10 10 л

Для данных параметров объем бутстрапа нужно брать приблизительно на 3 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.

Приложение Л

Таблица 12 - Значения показателей объемов для параметров: 10 10 п

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 7 | 0,9533 | 0,95 |
| 20 | 17 | 0,9497 | 0,95 |
| 30 | 27 | 0,95042 | 0,95 |
| 40 | 37 | 0,9499 | 0,95 |
| 50 | 46 | 0,95056 | 0,95 |
| 60 | 57 | 0,94914 | 0,95 |
| 70 | 66 | 0,94989 | 0,95 |
| 80 | 76 | 0,94989 | 0,95 |
| 90 | 86 | 0,95062 | 0,95 |
| 100 | 95 | 0,95053 | 0,95 |



Рисунок 12 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 10 10 п

Для данных параметров объем бутстрапа нужно брать приблизительно на 2-3 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.

Приложение М

Таблица 13 - Значения показателей объемов для параметров: 10 10 лп

|  |  |  |  |
| --- | --- | --- | --- |
| Объем выборки | Объем бутстрапа |  |  |
| 10 | 7 | 0,94438 | 0,95 |
| 20 | 17 | 0,94528 | 0,95 |
| 30 | 27 | 0,9484 | 0,95 |
| 40 | 37 | 0,94852 | 0,95 |
| 50 | 46 | 0,95064 | 0,95 |
| 60 | 56 | 0,95019 | 0,95 |
| 70 | 66 | 0,94986 | 0,95 |
| 80 | 75 | 0,95056 | 0,95 |
| 90 | 86 | 0,95007 | 0,95 |
| 100 | 96 | 0,94979 | 0,95 |



Рисунок 13 - Зависимость оптимального объема бутстрапа от объема исследуемой выборки для параметров: 10 10 лп

Для данных параметров объем бутстрапа нужно брать приблизительно на 3 единицы меньше объема основной выборки, чтобы получить доверительный интервал, в котором будет находиться истинное значение.