

СОДЕРЖАНИЕ И ТЕХНОЛОГИИ
МАТЕМАТИЧЕСКОГО ОБРАЗОВАНИЯ В ВУЗЕ

УДК 519.21

КАК ОЦЕНИТЬ ТОЧНОСТЬ ОЦЕНКИ СРЕДНЕГО?

А. В. Зорин

*Нижегородский государственный университет им. Н. И. Лобачевского
Нижегород, Россия*

`andrei.zorine@itmm.unn.ru`

Предлагаются простые вычислительные эксперименты, демонстрирующие, что формулу выборочной дисперсии следует применять с крайней осторожностью при построении доверительного интервала для среднего, если нет уверенности, что выборка получена в результате наблюдений над независимыми случайными величинами. Приводятся простые эксперименты в среде (на языке) программирования R, демонстрирующие этот эффект.

Ключевые слова: выборочная дисперсия, независимая повторная выборка, анализ временных рядов, преподавание теории вероятностей и математической статистики.

Введение

Математическая статистика входит в единый курс «Теория вероятностей и математическая статистика» во многих направлениях вузовской подготовки по математике и информационным технологиям, физике, экономике, социальным наукам. Зачастую это единственный курс, в рамках которого студенты имеют возможность ознакомиться со статистическими методами. Поэтому основные формулы из курса математической статистики часто становятся универсальным инструментом, который пытаются применить во всех ситуациях, порой даже забывая *условия* применимости этих формул.

Часто возникает задача об оценке среднего и об оценке погрешности выборочного среднего. Как решается эта задача в классической математической статистике? Рассмотрим две связанные друг с другом задачи.

Пусть дана повторная выборка $x_1, x_2, \dots, x_n$ из некоторого распределения вероятностей $F(x)$. Допустим, что у данного распределения существуют конечное математическое ожидание m и конечная дисперсия σ^2 . Требуется:

1) оценить по выборке числовые характеристики m и σ^2 ; 2) построить доверительный интервал для среднего m с заданной доверительной вероятностью γ , $0 < \gamma < 1$.

Ответ на первый вопрос содержится в формулах для выборочного среднего, выборочной дисперсии и несмещённой оценки для дисперсии:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad s_0^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Для ответа на второй вопрос предположим, что объём выборки n велик. Тогда, по центральной предельной теореме для независимых, одинаково распределённых случайных величин $X_1, X_2, \dots, X_n$ и среднего арифметического $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$ величина

$$\sqrt{n}(\bar{X} - m) \quad (1)$$

имеет приближённо нормальное распределение с математическим ожиданием ноль и дисперсией σ^2/n . И хотя для малых n замена неизвестной величины σ^2 её оценками s^2 или s_0^2 недопустима, при больших n практически можно использовать доверительный интервал вида

$$\left(\bar{x} - z \frac{\sigma}{\sqrt{n}}, \bar{x} + z \frac{\sigma}{\sqrt{n}} \right) \approx \left(\bar{x} - z \frac{s}{\sqrt{n}}, \bar{x} + z \frac{s}{\sqrt{n}} \right), \quad (2)$$

где $z = z_{(1+\gamma)/2}$ — квантиль стандартного гауссовского распределения уровня $(1 + \gamma)/2$.

Мы приведём примеры, демонстрирующие, что если выборка $x_1, x_2, \dots, x_n$ порождена зависимыми, а следовательно и коррелированными случайными величинами, то величина s^2 уже может быть плохой оценкой для дисперсии σ^2 , а асимптотическая дисперсия для нормированного отклонения (1) уже не имеет такого простого выражения, и может не быть связана с дисперсией одного слагаемого. Приводимые ниже примеры относятся к разделу случайных процессов и анализа временных рядов и не являются тайной для специалистов. Однако эти разделы, как правило, не читаются в базовом курсе теории вероятностей и математической статистики. Поэтому многие студенты остаются при ложном убеждении об универсальности формулы (2).

Тем не менее, все приводимые ниже примеры легко могут быть реализованы в виде небольших компьютерных программ и интересующий нас эффект станет наблюдаемым непосредственно в эксперименте. Иногда для практиков лучше сто раз увидеть, чем один раз доказать математически. Мы приведём тексты программ для языка программирования R [1], который является свободно распространяемым языком для прикладных статистических вычислений и набирает популярность при преподавании дисциплин по статистике и анализу данных. Таким образом, предлагаемые примеры могут быть использованы при чтении курса теории вероятностей и математической статистики

как части лабораторных работ или как самостоятельные исследовательские домашние задания.

Пример 1: Скользящее среднее

Пусть q — целое положительное число, $\{W_n; n = -q, 1 - q, 2 - q, \dots\}$ — последовательность независимых гауссовских случайных величин с математическим ожиданием ноль и дисперсией σ_w^2 (гауссовский «белый шум»). Пусть, далее,

$$X_n = W_n + \psi_1 W_{n-1} + \psi_2 W_{n-2} + \dots + \psi_q W_{n-q}, \quad n = 0, 1, \dots \quad (3)$$

есть так называемый процесс *скользящего среднего порядка q* .

Попробуем выполнить программу на Рис. 1 на языке R. Функция `ma.run` (строки 3–6) генерирует реализацию (временной ряд) длины `n`, используя при этом коэффициенты, переданные в виде вектора через переменную `ma`. При этом дисперсия шума $\sigma_w^2 = 1$. Основная работа по генерации происходит в функции `arima.sim`, взятой из библиотеки `astsa`. Затем вычисляются выборочное среднее $\bar{x}$ и выборочная дисперсия s^2 , которые возвращаются в точку вызова. В строке 8 устанавливаются коэффициенты $\psi_1 = -0.9$, $\psi_2 = 0.1$ как элементы вектора `ma1`, в строке 9 выбирается длина реализации. В строке 10 производится многократный запуск функции `ma.run` при одних и тех же значениях параметров, но при разных реализациях датчика псевдослучайных чисел.

```

1 library(astsa)
2
3 ma.run <- function(n=100, ma=c(-0.9)){
4   x<-arima.sim(list(ma=ma), n)
5   c(mean(x), mean( (x-mean(x))^2 ))
6 }
7
8 ma1 = c(-0.9, 0.1)
9 n = 10000
10 s1 <- replicate(10000, ma.run(n, ma1))
11 cat('Дисперсия_выборочного_среднего_', var(s1[1,]),
12     'наивная_оценка_', mean(s1[2,])/n, '\n')
```

Рис. 1. Программа для сравнения дисперсии выборочного среднего и дисперсии стационарного распределения в процессе скользящего среднего. Номера строк слева не являются частью программы и добавлены для удобства ссылок

В результате работы этой программы мы получили выборочную дисперсию для выборочного среднего $3.949622 \cdot 10^{-6}$, а «наивную» оценку по формуле σ^2/n получили 0.0001819824, что весьма близко к $(1 + \psi_1^2 + \psi_2^2)/n = 0.000182$ (пояснение этой формулы будет дано ниже). Как видно, разница

между реальной дисперсией выборочного среднего и величин σ^2/n составила два порядка, причём «наивная» оценка слишком пессимистична. В алгоритме ошибки нет. В чём же дело?

Проведём теоретический анализ. Последовательность $\{X_n; n = 0, 1, \dots\}$ состоит из коррелированных случайных величин, так как её автоковариационная функция имеет вид (положим $\psi_0 = 1$):

$$\begin{aligned}\Gamma(k) &= \mathbf{cov}(X_n, X_{n+k}) = \\ &= \begin{cases} (\psi_k + \psi_1\psi_{k+1} + \psi_2\psi_{k+2} + \dots + \psi_{q-k}\psi_q)\sigma_w^2, & k = 0, 1, \dots, q; \\ 0, & k = q + 1, q + 2, \dots \end{cases}\end{aligned}$$

Очевидно, все величины $\{X_n; n = 0, 1, \dots\}$ имеют одно и то же гауссовское распределение, задаваемое параметрами

$$m = \mathbf{M}(X_n) = 0, \quad \sigma^2 = \mathbf{D}(X_n) = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_q^2)\sigma_w^2.$$

Поскольку $\{X_n; n = 0, 1, \dots\}$ — стационарная случайная последовательность, для оценки математического ожидания m можно использовать выборочное среднее $\bar{x}$. Однако дисперсия выборочного среднего равна

$$\frac{1}{n} \left(\sigma^2 + 2 \sum_{k=1}^{n-1} \frac{n-k}{n} \Gamma(k) \right),$$

значит, при $n \rightarrow \infty$ распределение величины $\sqrt{n}(\bar{X} - m)$ стремится к нормальному с дисперсией

$$\sigma^2 + 2 \sum_{k=1}^{\infty} \Gamma(k).$$

Оценки дисперсии величинами $s^2(X_1, X_2, \dots, X_n)$ или $s_0^2(X_1, X_2, \dots, X_n)$ в этом случае также являются смещёнными, например, математическое ожидание выборочной дисперсии $s^2(X_1, X_2, \dots, X_n)$ равно (см. [2])

$$\sigma^2 - \frac{1}{n} \left(\sigma^2 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n} \right) \Gamma(k) \right).$$

Таким образом ошибка в среднем имеет порядок $1/n$.

Данный пример может выглядеть достаточно искусственным для начинающего изучать теорию вероятностей и математическую статистику. Оставляя для узкоспециальной литературы вопрос о реальных временных рядах, где адекватной моделью является именно процесс скользящего среднего, отметим, что этот пример позволяет аналитически рассчитать многие характеристики и сравнить их с экспериментальными наблюдениями.

Пример 2: Муха на подоконнике

Рассмотрим одну специальную цепь Маркова, которой можно дать следующую занимательную формулировку. Пусть на подоконнике находится муха¹. Длина подоконника равна единице. Если положение мухи (отсчитываемое, например, от левого края) задано координатой x , то она перелетает через некоторое время в новую точку y , которая имеет, по предположению, равномерное распределение на большем из отрезков $[0, x]$ и $[x, 1]$.

Моделирование скачков мухи, оценка среднего значения, дисперсия выборочного среднего и наивная оценка дисперсии среднего через дисперсию членов ряда приведены на Рис. 2.

```

1 fly.run <- function(n){
2   x <- rep(0, n)
3   u <- runif(n)
4   x[1] <- u[1]
5   for (i in 2:n)
6     if(x[[i-1]]<0.5)
7       x[[i]] <- u[[i]]+(1-u[[i]])*x[[i-1]]
8     else
9       x[[i]] <- u[[i]]*x[[i-1]]
10    c(mean(x), mean( (x-mean(x))^2 ), x[[n]])
11 }
12 n <- 1000
13 s2 <- replicate( 1000, fly.run(n) )
14 cat('Выборочное_среднее:', mean(s2[1,]),
15     ',_выборочная_дисперсия:', var(s2[3,]), '\n')
16 cat('Дисперсия_выборочного_среднего_', var(s2[1,]),
17     '_наивная_оценка_', mean(s2[2,])/n, '\n')
```

Рис. 2. Программа для сравнения дисперсии выборочного среднего и дисперсии стационарного распределения в процессе Марковской цепи скачков по единичному отрезку. Номера строк слева не являются частью программы и добавлены для удобства ссылок

Моделирование одной траектории скачков выполняется функцией `fly.run` (строки 1–11), которая возвращает затем выборочное среднее, выборочную дисперсию по этой последовательности скачков, а также последнее положение мухи. Запоминание последнего положения мухи позволит контролировать правильность моделирования, так как ниже мы получим точные значения для стационарного среднего и стационарной дисперсии. Их выборочные оценки находятся в строках 14–15. У нас получилось выборочное

¹По-видимому, когда студенту скучно, он начинает отвлекаться на всякую всячину. Допустим, в этот раз, во время обсуждения процессов скользящего среднего, ему на глаза попала муха!

среднее 0.5003033, выборочная дисперсия 0.06971212. При этом, выборочная дисперсия выборочного среднего оказалась равна $3.54815 \cdot 10^{-5}$, а «наивная» оценка оказалась равна $6.976128 \cdot 10^{-5}$. Снова несовпадение.

Если обозначить через X_n координату мухи после $n = 0, 1, \dots$ скачков, то случайная последовательность $\{X_n; n = 0, 1, \dots\}$ будет цепью Маркова с несчётным числом состояний. Поскольку такие цепи Маркова в учебной литературе практически не представлены, приведём здесь все необходимые вычисления.

Переходная плотность вероятности $p(x, y)$ равна

$$p(x, y) = \begin{cases} (1-x)^{-1} I_{(x,1)}(y), & x \leq 0.5, \\ x^{-1} I_{(0,x)}(y), & x > 0.5, \end{cases}$$

где индикаторная функция на множестве A определяется равенством

$$I_A(x) = \begin{cases} 0, & x \notin A, \\ 1, & x \in A. \end{cases}$$

Стационарная плотность $\pi(x)$ удовлетворяет интегральному уравнению ([3])

$$\pi(y) = \int_0^1 \pi(x) p(x, y) dx$$

или

$$\pi(y) = \int_0^{0.5} \pi(x) \frac{I_{(x,1)}(y)}{1-x} dx + \int_{0.5}^1 \pi(x) \frac{I_{(0,x)}(y)}{x} dx.$$

Пусть, для определённости, $y < 0.5$. Тогда правая часть равна

$$\int_0^y \frac{\pi(x)}{1-x} dx + \int_{0.5}^1 \frac{\pi(x)}{x} dx.$$

При $y > 0.5$ правая часть равна

$$\int_0^{0.5} \frac{\pi(x)}{1-x} dx + \int_y^1 \frac{\pi(x)}{x} dx.$$

Следовательно, $\pi(x) = \pi(1-x)$ и математическое ожидание по стационарному распределению равно 0.5.

Рассмотрим интервал $y < 0.5$. Второй интеграл в сумме не зависит от y ,

$$\int_{0.5}^1 \frac{\pi(x)}{x} dx = C > 0.$$

Значит, в этом диапазоне плотность удовлетворяет интегральному уравнению

$$\pi(y) = C + \int_0^y \frac{\pi(x)}{1-x} dx.$$

Дифференцируем по y :

$$\pi'(y) = \frac{\pi(y)}{1-y},$$

откуда (уравнение с разделяющимися переменными)

$$\ln \pi(y) = -\ln(1-y) + C_1, \quad \pi(y) = \frac{C}{1-y}.$$

Условие нормировки примет вид:

$$\frac{1}{2} = \int_0^{1/2} \frac{C}{1-x} dx = C \ln 2.$$

Итак, стационарная плотность распределения есть

$$\pi(x) = \begin{cases} (\ln 4 \cdot (1-x))^{-1}, & 0 < x < \frac{1}{2}, \\ (\ln 4 \cdot x)^{-1}, & \frac{1}{2} < x < 1, \end{cases}$$

а дисперсия распределения вероятностей с этой плотностью равна $0.069663\dots$. Это значение согласуется с результатом моделирования 0.07283708 .

Чтобы найти дисперсию выборочного среднего, надо применить подходящий вариант центральной предельной теоремы для статистически зависимых величин. Общая формулировка имеет вид: для функции $f(x): \mathbb{R} \rightarrow \mathbb{R}$ от состояния цепи Маркова $\{X_n; n = 0, 1, \dots\}$, $X_n \in \mathbb{R}$ (с произвольным начальным распределением, т.е. не обязательно со стационарным), со стационарной плотностью распределения $\pi(x)$, имеет место сходимость по распределению

$$\sqrt{n}(\overline{f(X)} - \pi(f)) \rightarrow \mathcal{N}(0, \sigma_f^2), \quad \frac{1}{n}(f(X_0) + \dots + f(X_{n-1})) \sim \mathcal{N}(\pi(f), \sigma_f^2/n),$$

где

$$\pi(f) = \int f(x)\pi(x) dx, \quad \overline{f(X)} = \frac{1}{n} \sum_{i=1}^n f(X_i),$$

и весь вопрос сводится к вычислению предельной дисперсии σ_f^2 . Наиболее цитируемый вариант ЦПТ для цепи Маркова восходит к работе [4] и выражает дисперсию для предельного распределения величины $\sqrt{n}(\overline{f(X)} - M)$ через стационарную автоковариационную функцию цепи Маркова $\{X_n; n = 0, 1, \dots\}$. Отметим, что для цепей Маркова вопрос об автоковариационных функциях ставится редко, видимо, из-за его аналитической трудности. Та-

ким образом, путь первого примера нам не подходит. К счастью, в последнее время появился вариант центральной предельной теоремы на основе теории мартингалов. Приведём его здесь, следуя монографиям [5, 6]. Для вычисления дисперсии σ_f^2 необходимо решить уравнение Пуассона:

$$g(x) = \left(f(x) - \int f(y)\pi(y) dy \right) + \int p(x, y)g(y) dy.$$

Применительно к нашему случаю, оно принимает вид:

$$\begin{aligned} g(y) &= y - \frac{1}{2} + \int_y^1 \frac{g(s)}{1-y} ds, & y < 1/2, \\ g(y) &= y - \frac{1}{2} + \int_0^y \frac{g(s)}{y} ds, & y > 1/2. \end{aligned}$$

Умножаем первое уравнение на $(1-y)$, затем дифференцируем по y :

$$-g(y) + (1-y)g'(y) = \left((1-y)\left(y - \frac{1}{2}\right) \right)' - g(y),$$

откуда

$$g'(y) = \frac{(3-4y)}{2(1-y)}, \quad g(y) = 2y + \frac{1}{2} \ln(1-y), \quad g\left(\frac{1}{2}\right) = 1 - \frac{1}{2} \ln 2.$$

Поскольку решение уравнения Пуассона определяется с точностью до аддитивной константы, мы считаем что $g(0) = 0$.

При $y > 0.5$ делаем аналогичные вещи со вторым уравнением:

$$(yg(y))' = \left(y\left(y - \frac{1}{2}\right) \right)' + g(y), \quad g(y) + yg'(y) = 2y - \frac{1}{2} + g(y),$$

откуда

$$g'(y) = \frac{4y-1}{2y}, \quad g(y) = C_1 + \int_{1/2}^y \frac{4x-1}{2x} dx = 2y - \frac{1}{2} \ln y + C_2.$$

Константа C_2 должна быть такая, чтобы функция удовлетворяла уравнению. Подстановка в уравнение даёт $C_2 = -1$.

Дисперсия σ_f^2 окончательно находится по формуле

$$\sigma_f^2 = \int_0^1 g^2(x) d\pi(x) - \int_0^1 \left(\int_0^1 p(x, y)g(y) dy \right)^2 d\pi(x).$$

Она оказывается равна 0.03570256.

Необходимо подчеркнуть одно важное отличие этого примера от первого примера: начальное распределение может отличаться от стационарного. Так что марковская цепь $\{X_n; n = 0, 1, \dots\}$ не обязана образовывать стационарную последовательность, и поэтому формулы на основе стационарных автоковариаций напрямую неприменимы. Также отметим, что стационарная автоковариационная функция для этой цепи Маркова, по-видимому, также довольно сложна. Автору ни разу не попадалась задача на вычисление автоковариационных функций для цепей Маркова (за исключением тривиальных авторегрессий первого порядка).

Пример 3: Времена ожидания

Рассмотрим систему массового обслуживания (далее — СМО), в которую поступает один входной поток клиентов с независимыми, одинаково распределёнными интервалами между поступлениями с функцией распределения $A(x)$, и независимыми, одинаково распределёнными длительностями обслуживания каждого клиента, заданными функцией распределения $B(x)$. Клиенты ждут в очереди с неограниченным числом мест, направляются на обслуживание в порядке поступления. Обозначим через X_n время ожидания начала обслуживания n -м клиентом, τ_n^e момент времени прихода n -го клиента, τ_n^s длительность обслуживания n -го клиента. Известно, что длительности ожидания определяются рекуррентным соотношением Линдли:

$$X_{n+1} = \max\{0, X_n + \tau_n^s - (\tau_{n+1}^e - \tau_n^e)\}, \quad n = 1, 2, \dots, \quad (4)$$

которое удобно использовать для имитационного моделирования этого случайного процесса.

Пример программы приведен на Рис. 3. Здесь принято, что интервалы между клиентами имеют экспоненциальное распределение с параметром λ (параметр `me` у функции `wait.run`), то есть

$$A(x) = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-\lambda x}, & x > 0, \end{cases}$$

а длительности обслуживания распределены по закону Эрланга порядка r (параметр `eo`) и средним значением μ (параметр `ms`), то есть

$$B(x) = \begin{cases} 0, & x \leq 0, \\ \int_0^x \frac{u^{r-1} r^r}{\mu^r \Gamma(r)} e^{-ru/\mu} du, & x > 0. \end{cases}$$

Поскольку распределение Эрланга есть частный случай гамма-распределения, здесь вызывается функция `rgamma()`. Мы получили, что для входного потока

с математическим ожиданием интервала 0.5 и средней длительностью обслуживания $\mu = 1$ выборочные среднее равно 0.6154766, выборочная дисперсия выборочного среднего равна 0.008542407, а «наивная» оценка дисперсии равна приближённо 0.001000222, то есть в несколько раз меньше. Следовательно, ширина доверительного интервала на основе выборочной дисперсии также в разы меньше, то есть точность оценки будет безосновательно завышена!

```

1 wait.run <- function(n=100, me=0.5, ms=1.0, eo=4){
2   x <- numeric(n)
3   x[0]<-0
4   for(i in 2:n){
5     # shape=eo, scale=eo/ms
6     x[[i]]<-max(c(0, x[[i-1]]+
7                 rgamma(1,eo,eo/ms)-rexp(1,me)))
8   }
9   c(mean(x), mean( (x-mean(x))^2 ))
10 }
11
12 n=1000
13 s3 <- replicate(5000, wait.run(n))
14 cat('Выборочное_среднее:', mean(s3[1,]), '\n')
15 cat('Дисперсия_выборочного_среднего_', var(s3[1,]),
16     'наивная_оценка_', mean(s3[2,])/n, '\n')
```

Рис. 3. Программа для сравнения дисперсии выборочного среднего и дисперсии стационарного распределения в процессе времён ожидания клиентами начала обслуживания. Номера строк слева не являются частью программы и добавлены для удобства ссылок

Случайная последовательность $\{X_n; n = 0, 1, \dots\}$, при сделанных предположениях, также будет цепью Маркова с несчётным множеством значений. Для существования стационарного распределения этой цепи необходимо и достаточно, чтобы выполнялось неравенство $\mathbf{M}(\tau_{n+1}^e - \tau_n^e) > \mathbf{M}(\tau_n^s)$ [7].

Чтобы проверить точность вычислений, приведём здесь вид характеристической функции времени ожидания клиентом начала обслуживания, следуя цитированной выше монографии. Пусть $\mathbf{i} = \sqrt{-1}$ — мнимая единица,

$$w_1(t) = \frac{(\lambda + \mathbf{i}t)^r}{\prod_{i=1}^r (t - t_i)}, \quad -\infty < t < \infty,$$

где $t_1, t_2, \dots, t_r$ есть корни уравнения

$$1 - \frac{\lambda r^r}{\mu^r (\lambda + \mathbf{i}t)(r - \mathbf{i}t\mu)^r} = 0$$

с отрицательной мнимой частью.

Тогда характеристическая функция времени ожидания обслуживания клиентом равна

$$\varphi(t) = \frac{w_1(t)}{w_1(0)},$$

а математическое ожидание и дисперсия находятся по формуле

$$m = \varphi'(0)/i, \quad \sigma^2 = (\varphi'(0))^2 - \varphi''(0).$$

По этой формуле, наивная оценка должна быть равна $\sigma^2/n = 0.001015625$. Кстати, теоретическое математическое ожидание равно 0.625. Таким образом, убеждаемся, что результаты имитационного моделирования согласуются с аналитически предсказанными результатами.

Заключение

Современные студенты, кажется, более заинтересованы в компьютеризации преподавания всех наук. На этом пути представляется рациональным ставить перед ними любопытную проблему сначала в виде вычислительного эксперимента, и лишь затем переходить к математическому анализу обнаруженного феномена.

Мы предложили три разные содержательные задачи, в которых проявляется один эффект — дисперсия выборочного среднего не определяется просто по дисперсии слагаемых. Хочется отметить, что пространство экспериментов в каждой из этих задач неограниченно. Например, можно провести исследование того, как быстро распределение выборочного среднего приближается к нормальному (скорость сходимости в центральной предельной теореме), как быстро распределение членов последовательности приближается к стационарному распределению (скорость сходимости к стационарному распределению для марковских цепей). Читатель наверняка сможет придумать ещё какие-то опыты.

И ещё одна тема может быть поднята в аудитории после обсуждения этих примеров: о соотношении между теорией вероятностей и математической статистикой. Предположим, что статистику принесли временной ряд, порождённый процессом из второго или третьего примера. Какие шансы, что статистик, не зная физики задачи *и не построив модель средствами теории вероятностей и теории случайных процессов*, сможет правильно угадать аналитическое выражение для законов распределения, породивших эти временные ряды?

Литература

- [1] R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [2] Андерсон Т. Статистический анализ временных рядов. — М.: Мир, 1976. — 755 с.

- [3] В.Феллер. Введение в теорию вероятностей и ее приложения. Т.2. —М.:Мир, 1984. — 738 с.
- [4] Ибрагимов И.А., Линник Ю.В. Независимые и стационарно связанные величины. — М.: Наука, 1965. — 524 с.
- [5] Meyn S., Tweedie R.L. Markov chains and Stochastic Stability. 2nd ed. — Cambridge: Cambridge University Press, 2009. — 624 p.
- [6] Asmussen S. Applied Probabilities and Queues. 2nd ed. — New York: Springer-Verlag, 2003. — 451 p.
- [7] Боровков А.А. Стохастические процессы в теории массового обслуживания. — М.: Наука, 1971. — 368 с.

Поступила 30.11.2022

HOW TO ESTIMATE THE ACCURACY OF AN ESTIMATOR FOR A MEAN VALUE?

A. V. Zorine

Several simple computational experiments are suggested which demonstrate that classical confidence interval formula should be used with precautions when it's not sure wheather the observations come from statistically independent random variables. Also, simple experiments by means of a programming language R (also a computating environment) are given making this effect visible.

Keywords: sample variance, independent sampling with replacement, time-series analysis, teaching probability theory and mathematical statistics.