Содержание

1 Описательная статистика

Здравствуйте, уважаемые слушатели. В предыдущих двух лекциях мы познакомились с моделями реальных случайных явлений и процессов, которые могут быть изучены с помощью теории вероятностей, а также способами их построения. Резюмируя, наверное можно сказать, что основная задача теории вероятностей – это прогноз, а именно: по построенной модели определить вероятность интересующего события; распределение той или иной случайной величины, системы случайных величин, их характеристики; распределение процесса случайных величин, изменяющегося во времени и проч. Вроде бы в теории все хорошо, не так ли? А на практике?

На практике нас часто ждет неизвестность, ведь обычно в результате эксперимента мы наблюдаем только проявления некоторой закономерности, а сама закономерность (или вероятностная модель этой закономерности) остается неизвестной. Так, мы можем измерить рост каждого ребенка в конкретной группе детского сада или в конкретно взятой волейбольной команде; мы можем получить данные об изменении курса доллара за последний год, посмотрев банковские сводки; можем узнать статистику забитых голов некоторой команды на нескольких последних чемпионатах мира. Зачем? Какую цель мы при этом преследуем?

Часто оказывается, что наша цель – получение каких-то сведений о (на данный момент) неизвестном будущем. Так, рост детей в группе детского сада (то есть детей конкретного возраста) может помочь нам в выборе мебели: ведь если взять чересчур высокие столы, то дети просто не смогут до них дотянуться и за ними обедать, а если чересчур высокие стулья – не смогут на них забраться. Анализируя изменение курса доллара, мы, конечно, хотим знать: когда выгодно купить валюту, а когда продать, и цель этого ясна – мы хотим получить максимальную прибыль. Статистика голов любимой команды может помочь нам выиграть дружеский спор об исходе матча на следующем чемпионате мира, и так далее.

На чем основаны наши надежды и ожидания при изучении прошлого? Конечно, на каком-то повторении, на наличии закономерности, на том, что в рассматриваемых ситуациях что-то да предопределено, на принципе: «если проводить эксперимент в одинаковых условиях, то и результат должен быть одинаков», так хорошо известном из физики. Иными словами, мы надеемся, что рассматриваемые случайные события подчиняются какой-то вероятностной закономерности, которую мы, правда, не знаем. Но вот если узнаем, то сразу сможем очень многое, ведь аппарат теории вероятностей изучен: мы сможем вычислять вероятности тех или иных событий, сравнивать их, вычислять характеристики вроде среднего, разброса и многое-многое другое, короче говоря, мы сможем прогнозировать и анализировать.

Выявление этих закономерностей (или отсутствия оных, что тоже очень ценно), – и есть одна из основных задач математической статистики. Изучением различных методов математической статистики мы и займемся в следующих лекциях.

1.1 Выборка и генеральная совокупность

Давайте начнем, наверное, с самой непростой ситуации, которую мы уже описали ранее, с ситуации, когда практически ничего неизвестно. Итак, у нас есть некоторый набор собранных данных и больше ничего. Что мы можем сделать, как это интерпретировать и толковать? Давайте начнем с такого вот примера.

Пример 1.1.1 *Молодой человек Петя собрал данные о времени (в минутах), на которое его девушка Даша опаздывала на свиданиях. Данные представлены набором чисел через запятую (в виде числового вектора) и для краткости обозначены X:*

$$X = (0, 11, 2, 3, 9, 2, 8, 6, 3.4, 8, 7.5, 9, 4, 8, 6).$$

Петя предполагает, что опоздания Даши хоть и носят случайных характер, но описываются какой-то случайной величиной ξ, имеющей какое-то неизвестное нам распределение.

Что же у нас есть? Согласно тому, что мы уже обсуждали в предыдущей лекции, можно сказать, что перед нами не что иное, как выборка из некоторой генеральной совокупности ξ. Основная же цель Пети – это, по имеющейся выборке X понять, как ведет себя генеральная совокупность ξ, а затем научиться прогнозировать и делать какие-то выводы об опозданиях Даши.

Напомним, что в прошлой лекции мы под выборкой понимали набор из n чисел $X = (x_1, x_2, ..., x_n)$. В нашем примере выборка состоит из 15 элементов, а Петя собрал данные, представленные в выборке, побывав на 15 свиданиях с Дашей. Но если бы он «обнулил счетчик» и начал бы собирать данные заново, то получил бы, скорее всего, другую выборку, хотя Даша-то не изменилась. И как тут строить какую-то статистику, данные же все время меняются.

Дело в том, что ранее введенное нами определение выборки – это так называемая выборка «после эксперимента». На самом же деле, чтобы строить какую-то статистику, разумно принять и вот какое определение выборки.

Определение 1.1.1 *Пусть ξ – рассматриваемая нами случайная величина. Выборкой $X = (X_1, X_2, ..., X_n)$ называется n независимых случайных величин, имеющих распределение такое же, как и ξ.*

Итак, говоря про выборку до конкретного эксперимента, мы будем понимать под ней набор независимых случайных величин и обозначать, как

$X = (X_1, X_2, ..., X_n)$. Если же эксперимент произошел и перед нами набор конкретных значений генеральной совокупности, то под выборкой мы будем понимать числовой вектор $X = (x_1, x_2, ..., x_n)$. В дальнейшем мы не будем конкретизировать то, что мы понимаем под выборкой (числовой вектор или вектор из случайных величин), это должно будет быть понятно из контекста.

1.2 Эмпирическое распределение

Итак, наша цель – узнать что-то о генеральной совокупности ξ. Раз ξ – случайная величина, то, наверное, логично построить по выборке тоже какую-то случайную величину, распределение которой и будет «приближать» истинное распределение ξ (а значит и будет приближать ее различные характеристики, как математическое ожидание, дисперсию и прочее). Мы уже подробно обсуждали в прошлой лекции, что хороший кандидат на роль «приближения» – так называемая эмпирическая (то есть «опытная») случайная величина ξ^*. Для удобства сейчас, и для приложений чуть позже, введем следующее важное определение.

Определение 1.2.1 *Пусть имеется выборка $X = (X_1, X_2, ..., X_n)$. Если элементы выборки упорядочить по возрастанию, то новый набор случайных величин, удовлетворяющий неравенствам*

$$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)},$$

называется вариационным рядом.

Понятно, что, например, $X_{(1)} = \min\{X_1, ..., X_n\}$, $X_{(n)} = \max\{X_1, ..., X_n\}$. Кроме того отметим, что k-ый член вариационного ряда часто называют k-ой порядковой статистикой.

Сразу вернемся к нашему примеру. По выборке опозданий Даши легко строится следующий вариационный ряд:

$$(0, 2, 2, 3, 3.4, 4, 6, 6, 7.5, 8, 8, 8, 9, 9, 11).$$

Вариационный ряд хорошо визуализирует небольшое количество данных и уже дает нам какую-то «статистику», ведь можно сразу сделать вывод, что не опоздала Даша всего раз, дважды опоздала на две минуты, один раз на три ну и так далее.

Не нужно думать, что вариационный ряд по конкретной выборке строится руками. Для этого в каждом инструменте, используемом при работе с данными, есть функция упорядочивания. Например, в Excel для этого во вкладке Данные реализована функция сортировки.

Имея вариационный ряд, удобно строить эмпирическое распределение случайной величины ξ^*. Группируя одинаковые значения и приписывая им вероятности, пропорциональные частоте встречи значения в выборке, получаем

ξ^*	0	2	3	3.4	4	6	7.5	8	9	11
P	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$

Эмпирическое распределение нам дает очень многое. Например, зная эмпирическое распределение случайной величины ξ^*, можно построить эмпирическую функцию распределения – функцию распределения ξ^*. В нашем случае она задается соотноешнием

$$F_n^*(t) = \begin{cases} 0, & t \leq 0 \\ 1/15, & 0 < t \leq 2, \\ 3/15, & 2 < t \leq 3, \\ 4/15, & 3 < t \leq 3.4, \\ \dots & \dots, \\ 14/15, & 9 < t \leq 11, \\ 1, & t > 11, \end{cases},$$

а ее график представлен на рисунке 1.

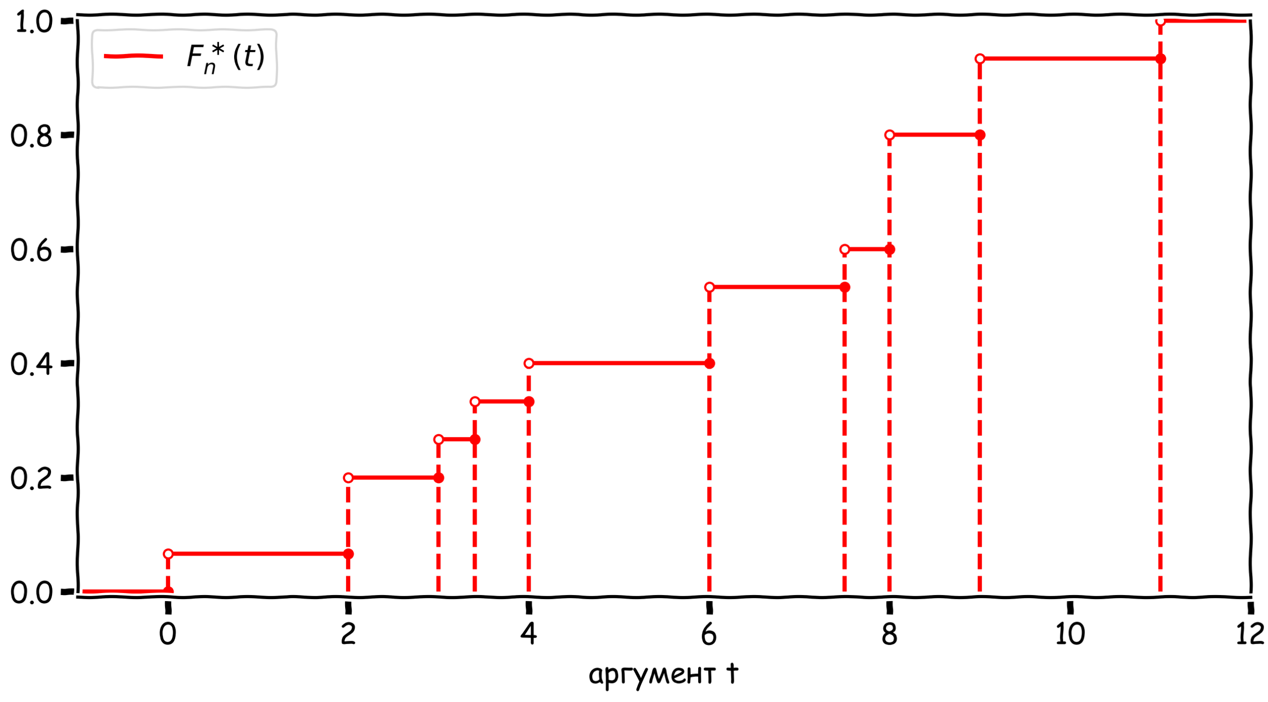

Рис. 1: Эмпирическая функция распределения, построенная по выборке

Теперь можно проводить какие-то элементарные оценки и предсказания. Скажем, как оценить вероятность события, что купленные для Даши круассан и стаканчик кофе не остынут? Иными словами, на языке вероятностей, какова вероятность события, что девушка опоздает не более, чем на 5 минут? Это мы умеем вычислять:

$$\mathsf{P}(\xi^* \le 5) = F_n^*(5+0) = \frac{6}{15} = 0.4,$$

что, кстати, не так уж и много: надежнее, быть может, купить букет цветов, чтобы уж точно порадовать подругу, ну или прихватить с собой термосумку.

Часто, однако, бывает достаточно оценить какие-то глобальные характеристики генеральной совокупности ξ (еще называемые мерами центральной тенденции) такие, как математическое ожидание $\mathsf{E}\xi$, дисперсию $\mathsf{D}\xi$, среднеквадратическое отклонение σ_ξ и так далее. Давайте научимся это делать.

1.3 Выборочные моменты

Итак, логично считать, что если распределение эмпирической случайной величины ξ^* приближает истинное распределение генеральной совокупности ξ, то и математическое ожидание, дисперсия и прочие характеристики эмпирической случайной величины являются хорошими аналогами тех же характеристик, но уже генеральной совокупности.

Так как в общем случае распределение ξ^* задается таблицей

$$\begin{array}{c|c|c|c|c|c} \xi^* & X_1 & X_2 & X_3 & ... & X_n \\ \hline \mathsf{P} & \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & ... & \frac{1}{n} \end{array},$$

где значение каждого элемента выборки равновероятно (снова, так как мы не отдаем ни одному из них какого-либо предпочтения), то

$$\mathsf{E}\xi^* = X_1 \cdot \frac{1}{n} + X_2 \cdot \frac{1}{n} + X_3 \cdot \frac{1}{n} + ... + X_n \cdot \frac{1}{n} = \frac{X_1 + X_2 + X_3 + ... + X_n}{n}.$$

Что мы получили? А то, что математическое ожидание построенной случайной величины есть не что иное, как среднее арифметическое элементов выборки. Именно поэтому эту характеристику даже назвали по-особому.

Определение 1.3.1 *Пусть дана выборка $X = (X_1, X_2, ..., X_n)$. Величину, равную среднему арифметическому элементов выборки, называют выборочным средним и обозначают $\overline{X}$. Иными словами,*

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Итак, еще раз, какой смысл несет в себе выборочное среднее $\overline{X}$? Так как распределение ξ^* приближает истинное распределение ξ, то и математическое ожидание $\mathsf{E}\xi^*$ (которое и есть выборочное среднее $\overline{X}$) приближает истинное математическое ожидание $\mathsf{E}\xi$. Последнее же несет в себе простой смысл: это среднее вероятностное значение случайной величины ξ.

Вернемся к примеру с опозданиями Даши. Выборочное среднее вычисляется, как

$$\overline{X} = \frac{0+11+2+3+9+2+8+6+3.4+8+7.5+9+4+8+6}{15} \approx 5.793.$$

Что это значит? Это значит, что в среднем Пете приходится ждать Дашу почти что 6 минут. На рисунке 2 синие точки – это элементы выборки. Красная точка отвечает выборочному среднему $\overline{X}$. Видно, что красная точка не совпадает ни с одним элементом выборки, но и правда располагается где-то посередине.

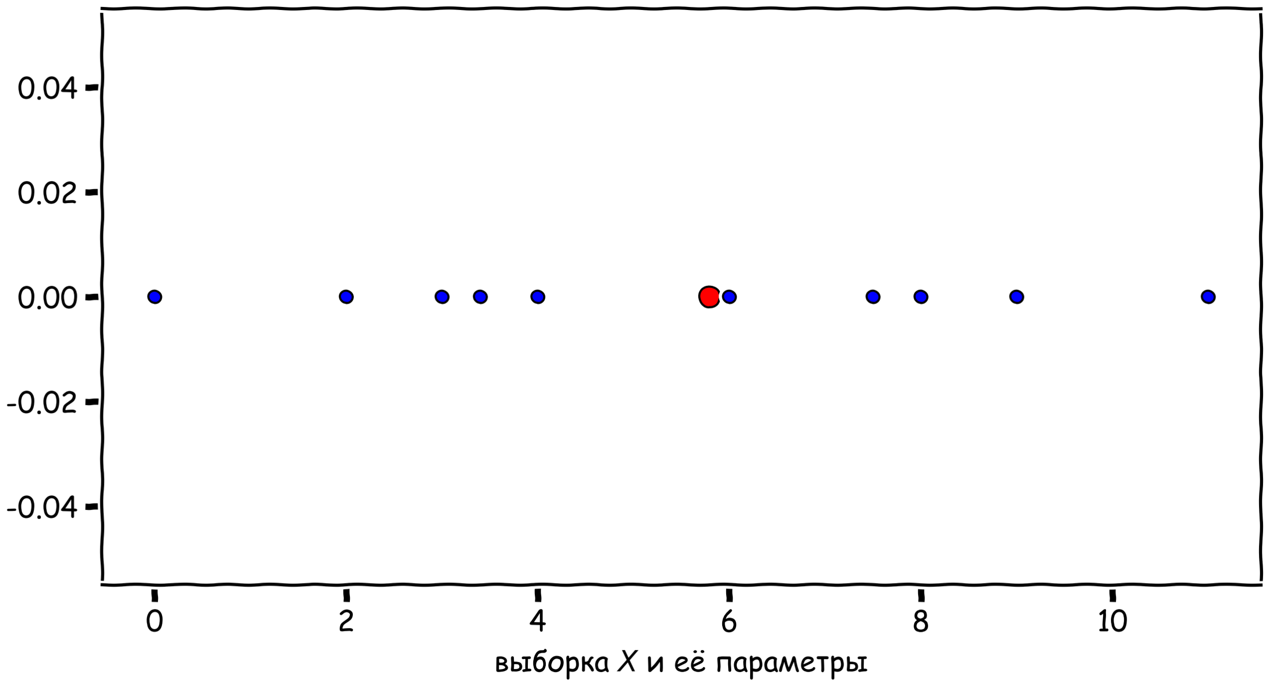

Рис. 2: Выборка X и ее среднее $\overline{X}$

Не нужно думать, что выборочное среднее нужно вычислять вручную. В том же Excel есть функция СРЗНАЧ, аргументом которой являются те числовые данные, среднее которых нужно вычислить.

По аналогичным причинам, дисперсия эмпирической случайной величины ξ^* должна неплохо приближать истинную дисперсию $\mathsf{D}\xi$. Согласно определению дисперсии и используя то, что $\mathsf{E}\xi^* = \overline{X}$, получаем

$$\mathsf{D}\xi^* = \mathsf{E}\left(\xi^* - \mathsf{E}\xi^*\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2.$$

Дисперсию эмпирической случайной величины тоже выделяют особо.

Определение 1.3.2 *Пусть дана выборка* $X = (X_1, X_2, ..., X_n)$*. Величину, равную дисперсии* $\mathsf{D}\xi^*$ *эмпирической случайной величны* ξ^**, построенной по выборке* X*, называют выборочной дисперсией и обозначают* S^2*. Иными словами,*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2.$$

Обратим внимание на смысл выборочной дисперсии S^2? Так как распределение ξ^* приближает истинное распределение ξ, то и дисперсия $\mathsf{D}\xi^*$ (которая и есть выборочная дисперсия S^2) приближает истинную дисперсию $\mathsf{D}\xi$. Последняя же несет в себе простой смысл: это средний квадрат разброса значений случайной величины ξ от ее математического ожидания $\mathsf{E}\xi$.

Рассмотрим все ту же выборку

$$X = (0, 11, 2, 3, 9, 2, 8, 6, 3.4, 8, 7.5, 9, 4, 8, 6).$$

Давайте посмотрим, насколько сильно отличаются в среднем опоздания Даши. Легко понять, что $S^2 \approx 9.625$. На рисунке 3 элементы выборки изображены синими точками, красная точка – выборочное среднее, а красная линия между зелеными точками показывает интервал

$$\left(\overline{X} - S, \overline{X} + S\right).$$

Что такое S? Это величина среднего разброса значений эмпирической случайной величины от среднего. Ее разумно рассматривать, как оценку среднеквадратического отклонения σ_ξ случайной величины ξ.

В нашем случае, как легко понять, $S \approx 3.1$, так что разброс от среднего, равного, примерно, 5.8, достаточно велик. Какой отсюда можно сделать вывод? А такой, что Пете самому не стоит опаздывать больше, чем на где-то две минуты, так как опоздания Даши носят весьма «разный» характер, и он может опоздать больше, чем она. Ну а если Даша опаздывает больше, чем на 9 минут, то можно начинать волноваться.

Замечание 1.3.1 *Отметим, что в статистике часто используется не только введенная выше выборочная дисперсия* S^2*, но и так называемая несмещенная выборочная дисперсия* S_0^2*, которая по выборке* $X = (X_1, X_2, ..., X_n)$ *определяется, как*

$$S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2.$$

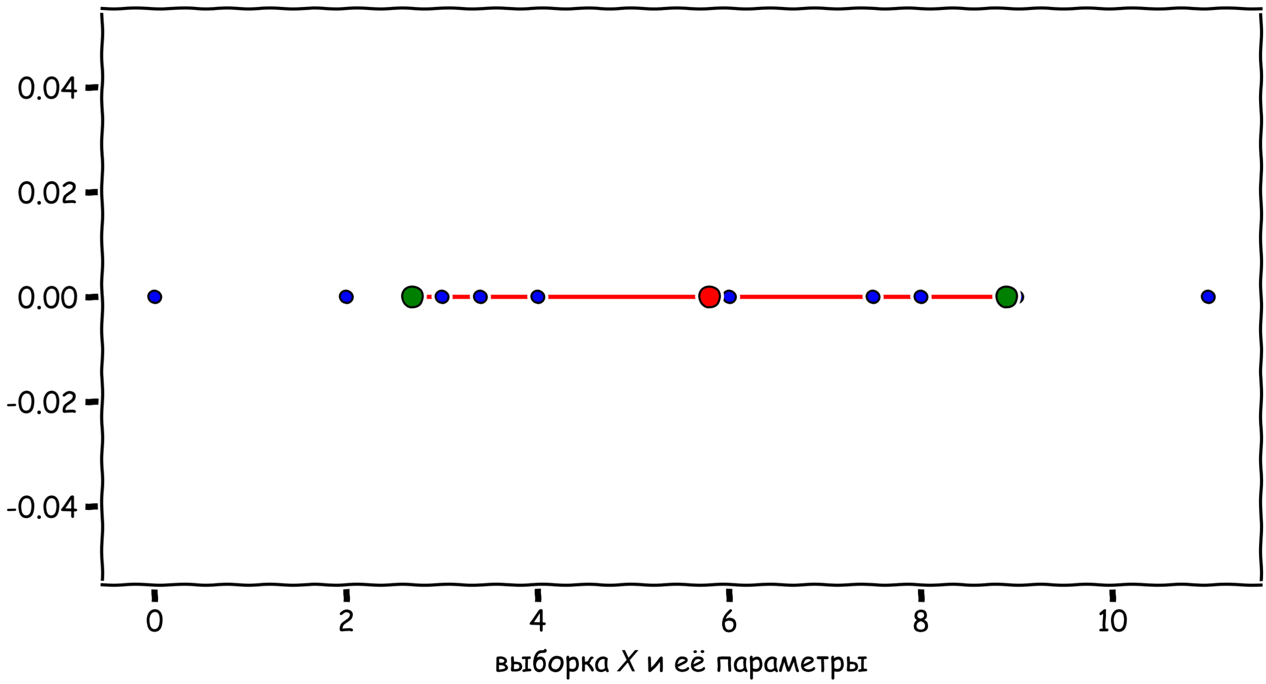

Рис. 3: Выборка X, ее выборочное среднее $\overline{X}$ и S

Несмещенная выборочная дисперсия S_0^2 алгебраически отличается от выборочной дисперсии S^2 сомножителем перед суммой: $\frac{1}{n}$ меняется на $\frac{1}{n-1}$. Благодаря этому наблюдению легко видеть, что

$$S_0^2 = \frac{n}{n-1} S^2$$

и, так как при $n \to +\infty$

$$\lim_{n \to +\infty} \frac{n}{n-1} = 1,$$

то при больших объемах выборки S_0^2 и S^2 практически не отличаются и могут использоваться равноправно. На малых же объемах выборки предпочтительнее использовать S_0^2, так как она точнее. Почему это так, а также что означает несмещенность, мы узнаем несколько позже.

В нашем примере $S_0^2 \approx 10.312$ и видно, что это значение достаточно сильно (почти на 0.8) отличается от S^2.

Ровно по тем же соображениям, по которым при больших n оценку S_0^2 можно считать разумной оценкой дисперсии генеральной совокупности ξ, величину S_0 можно считать разумной оценкой среднеквадратического отклонения σ_ξ. В нашем случае $S_0 \approx 3.21$.

Пример 1.3.1 *Рассмотрим еще один пример, иллюстрирующий отличие S^2 и S_0^2 на синтетических выборках из генеральной совокупности, имеющей известную дисперсию, равную 9 (рисунок 4).*

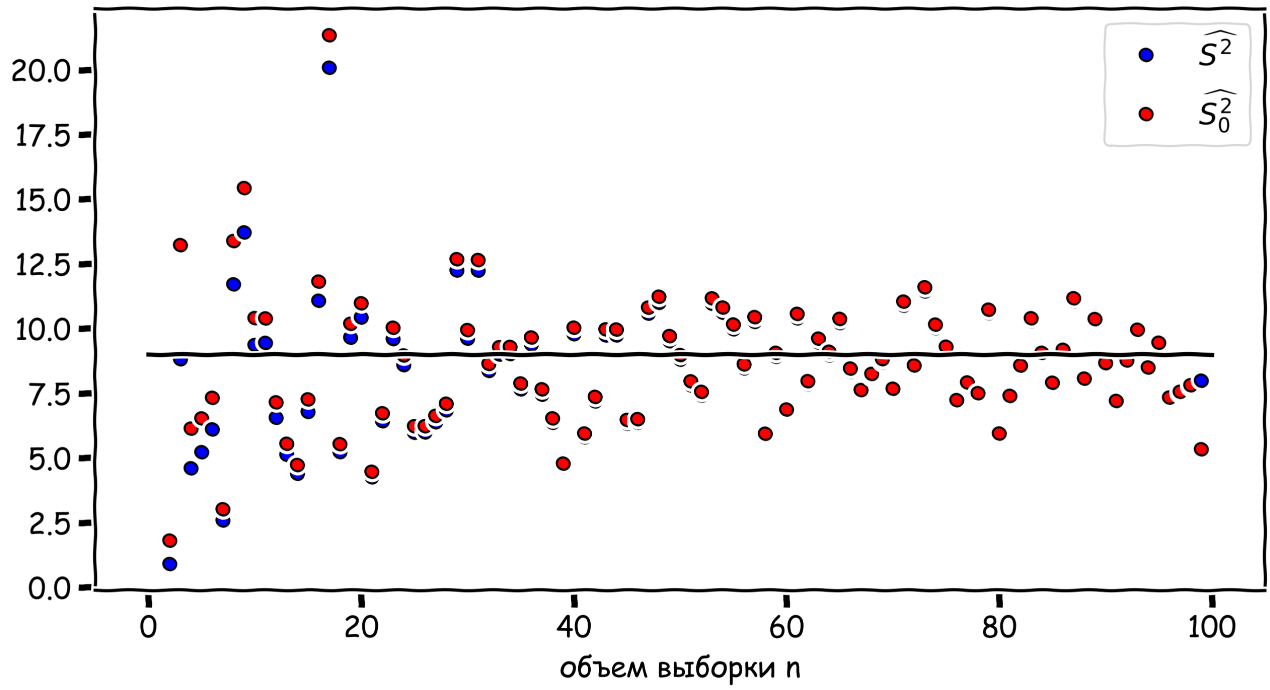

Рис. 4: Сравнение S^2 и S_0^2

Как можно видеть, на малых объемах выборки несмещенная дисперсия S_0^2 оказывается в среднем ближе к истинному значению дисперсии. Однако, с ростом количества элементов выборки, различия между S^2 и S_0^2 становятся все меньше и точки практически сливаются.

Выборочная дисперсия, как и многие другие характеристики, «вшиты» в большинство пакетов для анализа данных. В Excel есть функция ДИСП.Г или ДИСПРА, аргументом которой являются те числовые данные, выборочную дисперсию S^2 которых нужно вычислить. Обратите внимание, что похожие функции ДИСП.В и ДИСПА выдают несмещенную выборочную дисперсию S_0^2.

1.4 Медиана. Выборочная медиана.

Сейчас нам придется немного отключиться от статистики, и совсем ненадолго окунуться в вероятностный аппарат, а именно – изучить еще одну вероятностную характеристику случайной величины, называемую медианой. Давайте сначала дадим строгое определение новому для нас понятию, а потом поясним ее смысл, а также сравним с математическим ожиданием.

Определение 1.4.1 *Число* $\mathsf{med}\ \xi$ *называется медианой случайной величины* ξ*, если*

$$\mathsf{P}(\xi \leq \mathsf{med}\ \xi) \geq \frac{1}{2} \text{ и } \mathsf{P}(\xi \geq \mathsf{med}\ \xi) \geq \frac{1}{2}.$$

По сути своей медиана – это такое число, что случайная величина как минимум с вероятностью $\frac{1}{2}$ не больше и не меньше нее. Похоже на математическое ожидание, не так ли? На самом деле не все так просто. Например оказывается, что медиана всегда существует, но не всегда является единственной. Почему? Давайте подумаем и проиллюстрируем это таким простым примером.

Пример 1.4.1 *Пусть эксперимент заключается в подбрасывании правильной монеты. Тогда распределение случайной величины ξ, дающей единицу, если выпал орел, и ноль, если решка, задается следующей таблицей*

$$\begin{array}{c|c|c} \xi & 0 & 1 \\ \hline \mathsf{P} & \frac{1}{2} & \frac{1}{2} \end{array}.$$

Ясно, что любое число $\mathsf{med}\ \xi$ *из диапазона* $[0,1]$ *служит медианой случайной величины* ξ*, ведь*

$$\mathsf{P}(\xi \leq \mathsf{med}\ \xi) \geq \frac{1}{2} \text{ и } \mathsf{P}(\xi \geq \mathsf{med}\ \xi) \geq \frac{1}{2}.$$

Почему так случилось? Потому, что в данном случае исходы равновозможны и медиана, грубо говоря, «не понимает», что выбрать.

А чему в данном случае равно математическое ожидание? Легко проверить, что оно равно

$$\mathsf{E}\xi = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

Чаще всего, если медиана не единственна (а значит этих медиан сразу целый отрезок), то в качестве медианы берут середину получившегося отрезка. В нашем случае $\mathsf{med}\ \xi$ совпадает с математическим ожиданием и равна $\frac{1}{2}$. Так бывает не всегда.

Пример 1.4.2 *Рассмотрим такой простой пример. Предположим, что имеется некоторая фирма, в которой работает* 100 *человек, один из которых начальник. Заработная плата начальника равна* 101000 *долларов в месяц, а заработная плата каждого работника равна* 1000 *долларов в месяц. Пусть случайная величина* ξ *(или генеральная совокупность) – зарплата работника, тогда эмпирическое распределение, построенное по данной нам выборке, может быть задано, как*

$$\begin{array}{c|c|c} \xi^* & 1000 & 101000 \\ \hline \mathsf{P} & \frac{99}{100} & \frac{1}{100} \end{array}$$

Легко понять, что математическое ожидание случайной величины ξ^* *равно*

$$\mathsf{E}\xi^* = 1000 \cdot \frac{99}{100} + 101000 \cdot \frac{1}{100} = 2000,$$

то есть средняя зарплата равна 2000 *долларов в месяц. В то же время, медиана* $\mathsf{med}\ \xi^*$ *равна* 1000 *и медианная зарплата равна* 1000.

Ясно, что в этом примере гораздо более «честной» оценкой средней зарплаты является медиана, нежели математическое ожидание, из-за такого сильного выброса в заработной плате начальника.

Итак, снова мысленно повторяя аргументы, приводимые нами ранее, скажем, что медиана $\mathsf{med}\ \xi^*$ должна хорошо приближать истинную медиану $\mathsf{med}\ \xi$, а потому примем следующее определение.

Определение 1.4.2 *Пусть дана выборка* $X = (X_1, X_2, ..., X_n)$. *Выборочной медианой* $\widehat{\mathsf{med}\ \xi}$ *генеральной совокупности* ξ *называется медиана выборочной случайной величины, то есть*

$$\widehat{\mathsf{med}\ \xi} = \mathsf{med}\ \xi^*.$$

Имея выборку $X = (X_1, X_2, ..., X_n)$, и построенный по ней вариационный ряд

$$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)},$$

выборочная медиана легко может быть найдена из соотношений

$$\widehat{\mathsf{med}\ \xi} = \begin{cases} X_{([n/2]+1)}, & n \text{ нечетно}, \\ \frac{X_{([n/2])}+X_{([n/2]+1)}}{2}, & n \text{ четно}, \end{cases}$$

где $[x]$ обозначают целую часть от x. Например, $[2.7] = 2$, $[4.2] = 4$.

А как найти медиану по выборке (или по вариационному пряду)? По вариационному ряду выборочная медиана ищется следующим образом. Она равна серединному элементу вариационного ряда, если он определен однозначно (то есть если n нечетно), и полусумме серединных элементов, если n четно.

Возвращаясь к примеру с опозданиями Даши получим, что, согласно вариационному ряду

$$(0, 2, 2, 3, 3.4, 4, 6, 6, 7.5, 8, 8, 8, 9, 9, 11),$$

выборочная медиана равна $X_{([15/2]+1)} = X_{(8)} = 6$. Это значит, что Даша с «одинаковыми» (в нашем приближении) вероятностями опаздывает как менее, чем на 6 минут, так и более. Напомним, что выборочное среднее $\overline{X}$ было равно примерно 5.8, так что в нашем примере полученные параметры близки, а значит предполагаемое распределение генеральной совокупности ξ достаточно симметрично относительно как математического ожидания, так и близкой к нему медианы.

Еще раз особо отметим, что медиана, по сравнению с выборочным средним, устойчива к выбросам. Предположим, что Петя случайно отвлекся и приписал лишний ноль к очередному измерению, тем самым вместо 5 минутного опоздания записал значение в 50 минут. Тогда новый вариационный ряд будет содержать уже 16 членов:

$$(0, 2, 2, 3, 3.4, 4, 6, 6, 7.5, 8, 8, 8, 9, 9, 11, 50),$$

а выборочное среднее $\overline{X}$ окажется примерно равным $\overline{X} = 8.566$ и будет отличаться от предыдущего значения примерно на 2.8. А вот новая выборочная медиана $\widehat{\mathsf{med}\ \xi}$ будет равна $\widehat{\mathsf{med}\ \xi} = 6.75$ и отличается от предыдущего значения всего на 0.75.

Замечание 1.4.1 *Выбросы в данных, как показывает практика, – это достаточно частое явление. Иногда может барахлить прибор, записывающий измерения, могут возникать редкие аномалии (как невнимательность Пети), влияющие на эксперимент. Все это портит прогнозы, если появляется с заядлой периодичностью, потому устойчивые (или робастные) оценки характеристик оказываются очень полезными.*

Теперь что касается конкретного счета. Большинство пакетов анализа данных умеют вычислять медиану. В Excel это может быть сделано посредством функции МЕДИАНА, аргументом которой выступает вся выборка.

1.5 Гистограмма, как оценка плотности

Напомним, что наша цель – узнать что-то о генеральной совокупности ξ. При этом случайная величина ξ может быть и непрерывной. А что очень хотелось бы знать о непрерывной случайной величине? Ну, наверное, ее плотность. Эмпирическим аналогом плотности является так называемая гистограмма.

Гистограмма строится достаточно просто. На первом этапе определяется множество значений выборки. Используя вариационный ряд, это множество значений будет отрезком

$$A = [X_{(1)}, X_{(n)}].$$

Далее этот отрезок делят на некоторое количество непересекающихся отрезков, интервалов, полуинтервалов. Затем подсчитывают количество значений, попавших в каждый такой промежуток.

Пусть $A_1, ..., A_k$ – эти отрезки, одинаковой длины l. Пусть ν_j – количество элементов выборки, попавших в отрезок A_j, $n = \sum\limits_{j=1}^{k} \nu_j$. Заменим истинную плотность на промежутке A_j длины l прямоугольником высоты $h_j = \frac{\nu_j}{nl}$. Заметим, что сумма площадей всех прямоугольников равна 1, а это значит, что полученная неотрицательная функция может трактоваться, как плотность распределения некоторой случайной величины. Действительно, площадь S_j j-ого прямоугольника равна

$$S_j = h_j \cdot l = \frac{\nu_j}{nl} \cdot l = \frac{\nu_j}{n},$$

а тогда суммарная площадь построенных прямоугольников равна

$$\sum_{j=1}^{k} S_j = \sum_{j=1}^{k} \frac{\nu_j}{n} = \frac{1}{n} \sum_{j=1}^{k} \nu_j = 1.$$

Таким образом построенная ступенчатая фигура, являющаяся объединением прямоугольников, называется гистограммой.

Пример 1.5.1 *Снова рассмотрим выборку, собранную Петей, из нашего примера. Напомним, что вариационный ряд имеет вид*

$$(0, 2, 2, 3, 3.4, 4, 6, 6, 7.5, 8, 8, 8, 9, 9, 11).$$

Разобьем для начала отрезок $[0, 11]$ *на* $n = 4$ *части одинаковой длины (*$l = 2.75$*), тем самым получим множества*

$$A_1 = [0, 2.75),\ A_2 = [2.75, 5.5),\ A_3 = [5.5, 8.25),\ A_4 = [8.25, 11].$$

В множество A_1 *попадает 3 элемента выборки, значит* $\nu_1 = 3$. *Аналогичными рассуждениями получаем, что* $\nu_2 = 3$, $\nu_3 = 6$, $\nu_4 = 3$. *Гистограмма, отвечающая такому разбиению, представлена на рисунке 5. Если взять* $n = 5$, *то картина меняется, см. рисунок 6.*

Замечание 1.5.1 *Большинство пакетов по обработке данных поддерживают построение гистограмм. В частности в* Excel *это можно сделать при помощи пакета анализа данных или посредством встроенной функции ЧАСТОТА().*

1.6 Многомерная выборка. Выборочная корреляция

Предположим теперь, что мы наблюдаем не одну, а несколько случайных величин. Эта ситуация не менее жизненна, чем те, что мы рассматривали ранее. Один из самых интересных вопросов – есть ли какая-то зависимость

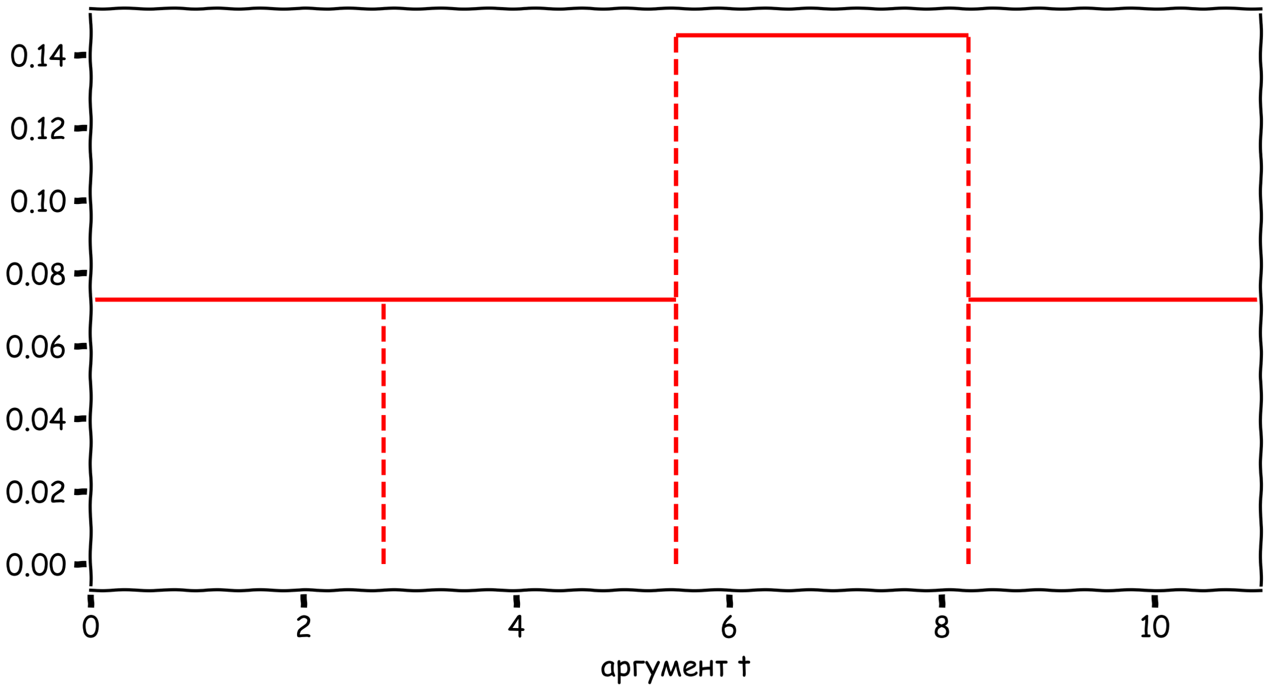

Рис. 5: Гистограмма при $n = 4$

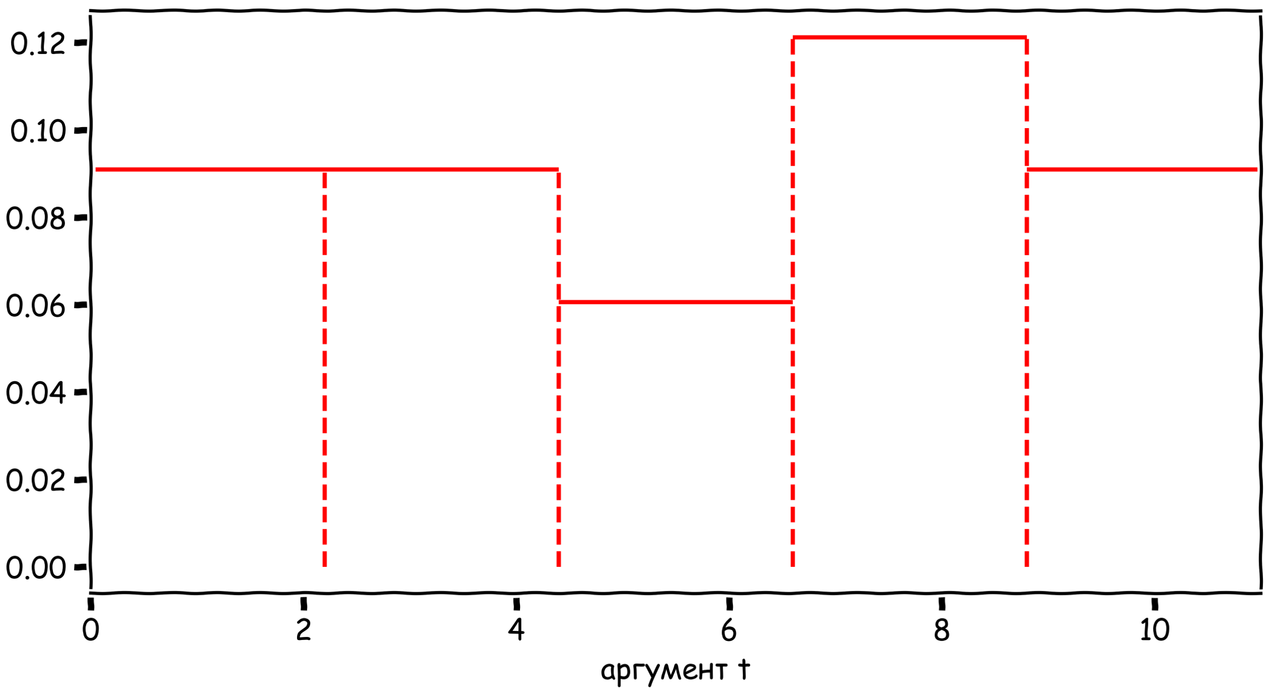

Рис. 6: Гистограмма при $n = 5$

между наблюдаемыми значениями выборки? Как мы знаем, индикатором зависимости в теории вероятностей выступают ковариация и коэффициент корреляции. Давайте определим их выборочные аналоги, но для начала введем понятие многомерной выборки.

Итак, пусть генеральная совокупность – это случайный вектор $\vec{\xi} = (\xi_1, \xi_2)$, состоящий из двух компонент (для простоты). Тогда, аналогично тому, как было сделано в одномерном случае, резонно ввести следующее определение.

Определение 1.6.1 *Двумерной выборкой* $(X, Y) = ((X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n))$ *объема* n *называется набор из* n *независимых одинаково распределенных пар случайных величин* (X_i, Y_i), *каждая из которых имеет такое же совместное распределение, как и пара* $\vec{\xi} = (\xi_1, \xi_2)$.

Напомним, что так как ковариация двух случайных величин ξ_1 и ξ_2 определяется, как

$$\mathsf{cov}(\xi_1, \xi_2) = \mathsf{E}(\xi_1 - \mathsf{E}\xi_1)(\xi_2 - \mathsf{E}\xi_2) = \mathsf{E}(\xi_1\xi_2) - \mathsf{E}\xi_1\mathsf{E}\xi_2,$$

то логично принять следующее определение.

Определение 1.6.2 *Выборочной ковариацией, построенной по выборке* $(X, Y) = ((X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n))$, *называется*

$$k(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

Часто вводят в рассмотрение и несмещенную выборочную ковариацию, которая определяется из соотношения

$$k_0(X, Y) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

Мотивировка для рассмотрения несмещенной выборочной ковариации ничем не отличается от мотивировки для рассмотрения несмещенной выборочной дисперсии: при больших n они ведут себя одинаково, а при маленьких ошибка у несмещенной оценки меньше. Об этом подробнее мы еще будем говорить в дальнейшем.

Для вычисления ковариации в Excel предусмотрены две функции: КОВАРИАЦИЯ.В для несмещенной ковариации и КОВАРИАЦИЯ.Г – для смещенной.

Для рассмотрения зависимости интереснее получить так называемую выборочную корреляцию. Так как корреляция ξ_1 и ξ_2 определяется, как

$$\rho(\xi_1, \xi_2) = \frac{\mathsf{cov}(\xi_1, \xi_2)}{\sigma_{\xi_1}\sigma_{\xi_2}},$$

то выборочная корреляция определяется следующим образом.

Определение 1.6.3 *Выборочной корреляцией, построенной по выборке* $(X, Y) = ((X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n))$, *называют величину*

$$r(X, Y) = \frac{k(X, Y)}{S_X S_Y} = \frac{k_0(X, Y)}{S_{0X} S_{0Y}},$$

где S_X^2, S_Y^2 *– смещенные выборочные дисперсии, а* S_{0X}^2, S_{0Y}^2 *– несмещенные выборочные дисперсии, построенные по выборкам* $X = (X_1, X_2, ..., X_n)$ *и* $Y = (Y_1, Y_2, ..., Y_n)$, *соответственно.*

Пример 1.6.1 *Среднемесячная заработная плата (тыс. руб.) в Ленинградской области в 2010-2011 годах составила по отраслям.*

Отрасль	*ЖКХ*	*Здравоохранение*	*Образование*	*ИТ*
2010 *год*	13.1	9.2	10.2	23.6
2011 *год*	16.2	11.6	13	28.8

Найдем несмещенную ковариацию при помощи Excel. *Для этого используем соответствующую функцию КОВАРИАЦИЯ.В.*

$$k_0(X, Y) \approx 51.6.$$

Для нахождения коэффициента корреляции используем соответствующую функцию КОРРЕЛ.

$$r(X, Y) \approx 0.99$$

2 Оценивание параметров вероятностной модели

2.1 Наводящий пример

Ситуация, которую мы только что рассматривали, то есть ситуация, когда ничего (кроме выборки) неизвестно, наверное, самая плохая, но не единственная. Давайте рассмотрим такой модельный пример.

Предположим, что заядлый посетитель тира производит 10 выстрелов по мишени из одного и того же пистолета. Выборка результатов стрельбы, где 1 отвечает попаданию, а 0 – промаху, такова:

$$(1, 0, 0, 1, 0, 1, 0, 0, 1, 1).$$

Конечно, можно считать, что кроме выборки никакой другой информации нет, но можно рассуждать и иначе. Наверное, нетрудно понять, что каждый эксперимент (которых проводилось всего 10) имеет всего два исхода: попал и не попал, и вероятность благоприятного исхода в каждом испытании одинакова и равна p (ведь раз посетитель постоянный, то глаз у него должен быть «наметан», пристрелян, и так далее). Значит, можно допустить, что случайная величина, показывающая результат эксперимента, имеет распределение Бернулли B_p с неизвестным параметром p.

А что это нам дает? А дает вот что. Если мы сможем оценить p, используя данную нам выборку, то мы сразу получим вероятностную модель эксперимента, а это, как мы не раз говорили, очень много. Но как же оценить p? Оказывается, мы для этого уже все подготовили, смотрите.

Пусть $\xi \sim \mathsf{B}_p$, тогда $\mathsf{E}\xi = p$. Но хорошей оценкой математического ожидания $\mathsf{E}\xi$ является выборочное среднее $\overline{X}$, значит, вестимо, его можно считать и оценкой $\widehat{p}$ параметра p. Итак,

$$\widehat{p} = \overline{X} = 0.5$$

Теперь можно и прогнозировать.

2.2 Оценки параметров некоторых стандартных распределений

Итак, давайте конспективно приведем оценки параметров некоторых стандартных распределений. Чтобы подчеркнуть, что вычисленная нами по выборке величина является оценкой, мы над ней будем ставить крышку.

1. Рассмотрим распределение Бернулли B_p. Так как математическое ожидание $\mathsf{E}\xi$ случайной величины ξ, имеющей распределение Бернулли, равно p, то

$$\widehat{p} = \overline{X}.$$

Проведем численный эксперимент на синтетических выборках разного объема из распределения Бернулли с параметром $p = 0.6$. Как видно из рисунка 7, с ростом числа n выборочное среднее все лучше приближает истинное значение параметра $p = 0.6$

2. Рассмотрим биномиальное распределение $\mathsf{Bin}(m, p)$ с неизвестными параметрами m, p. Мы знаем, что математическое ожидание $\mathsf{E}\xi$ случайной величины, имеющей биномиальное распределение, равно mp, а дисперсия $\mathsf{D}\xi = mp(1-p)$. Тогда (учитывая, что оценка математического ожидания – это $\overline{X}$, а оценка дисперсии – S^2), решая систему,

$$\begin{cases} mp = \overline{X} \\ mp(1-p) = S^2 \end{cases}$$

приходим к оценкам

$$\widehat{p} = 1 - \frac{S^2}{\overline{X}}, \quad \widehat{m} = \frac{\overline{X}^2}{\overline{X} - S^2}.$$

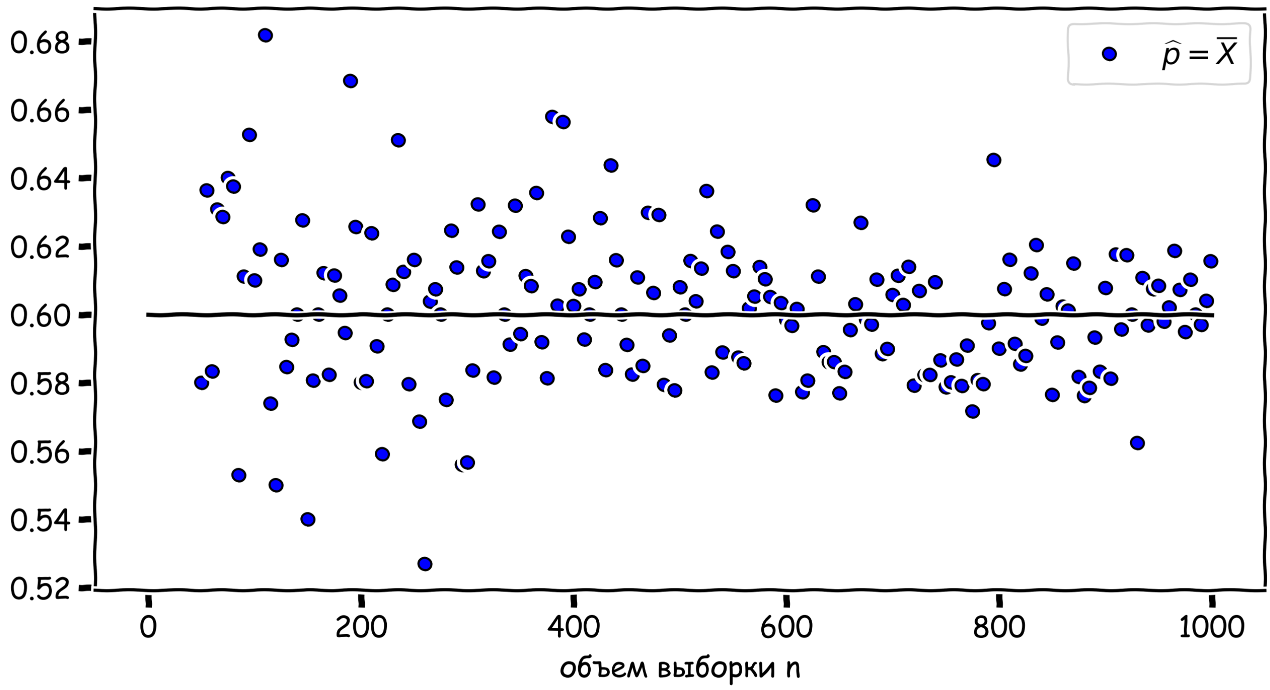

Рис. 7: Зависимость $\overline{X}$ от объема выборки

Если использовать для оценки дисперсии несмещенную выборочную дисперсию S_0^2, то

$$\widehat{p} = 1 - \frac{S_0^2}{\overline{X}}, \quad \widehat{m} = \frac{\overline{X}^2}{\overline{X} - S_0^2}.$$

На самом деле, так как m – натуральное число, то в качестве оценки стоит брать ближайшее целое число к тому, что получается по формулам, приведенным выше.

В случае, если m известно, формулы упрощаются и

$$\widehat{p} = \frac{\overline{X}}{m}.$$

Если же известно p, то

$$\widehat{m} = \frac{\overline{X}}{p}$$

и в качестве оценки стоит снова брать ближайшее целое число.

Пример оценок значения p, построенных по выборкам из биномиального распределения с параметром $p = 0.6$, представлен на рисунке 8. Можно убедиться, что если известен параметр m, то оценка параметра p получается более точной, чем в случае, когда неизвестно ничего.

3. Рассмотрим распределение Пуассона Π_λ с неизвестным параметром $\lambda > 0$. Так как математическое ожидание $\mathsf{E}\xi$ случайной величины, имеющей

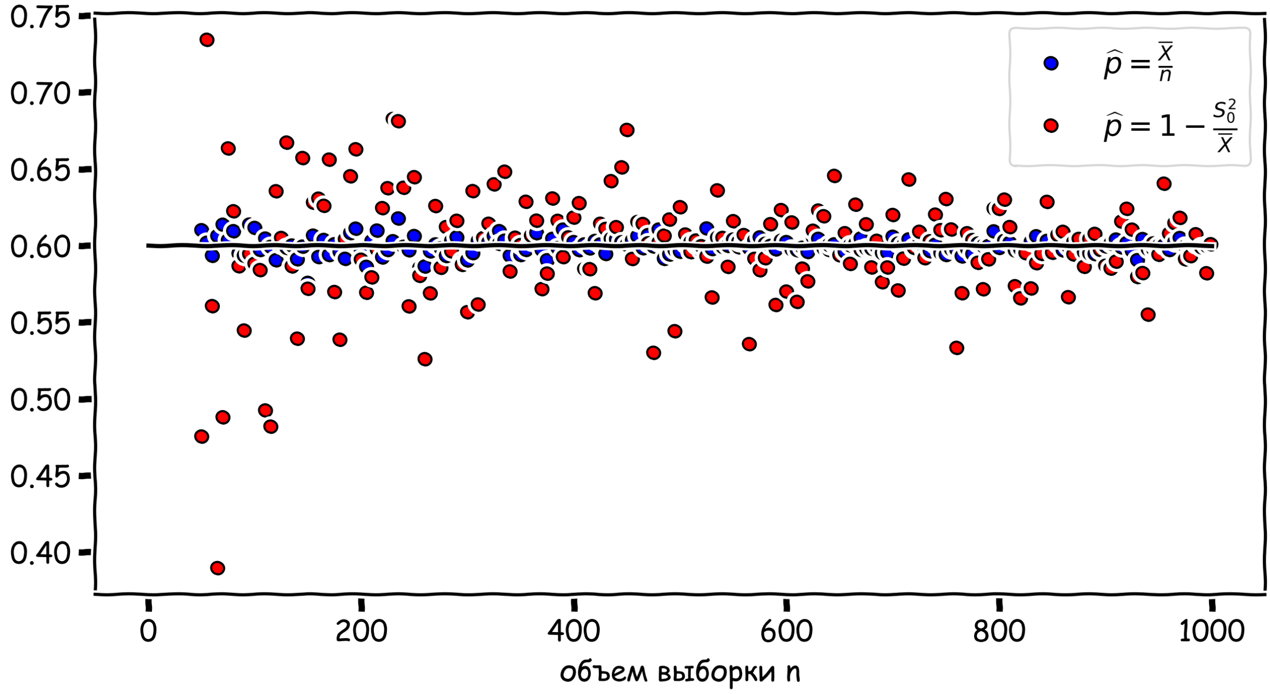

Рис. 8: Зависимость $\widehat{p}$ от объема выборки

распределение Пуассона, равно λ, то

$$\widehat{\lambda} = \overline{X}.$$

Пример оценок значения λ, построенных по выборкам из распределения Пуассона с параметром $\lambda = 1$, представлен на рисунке 9.

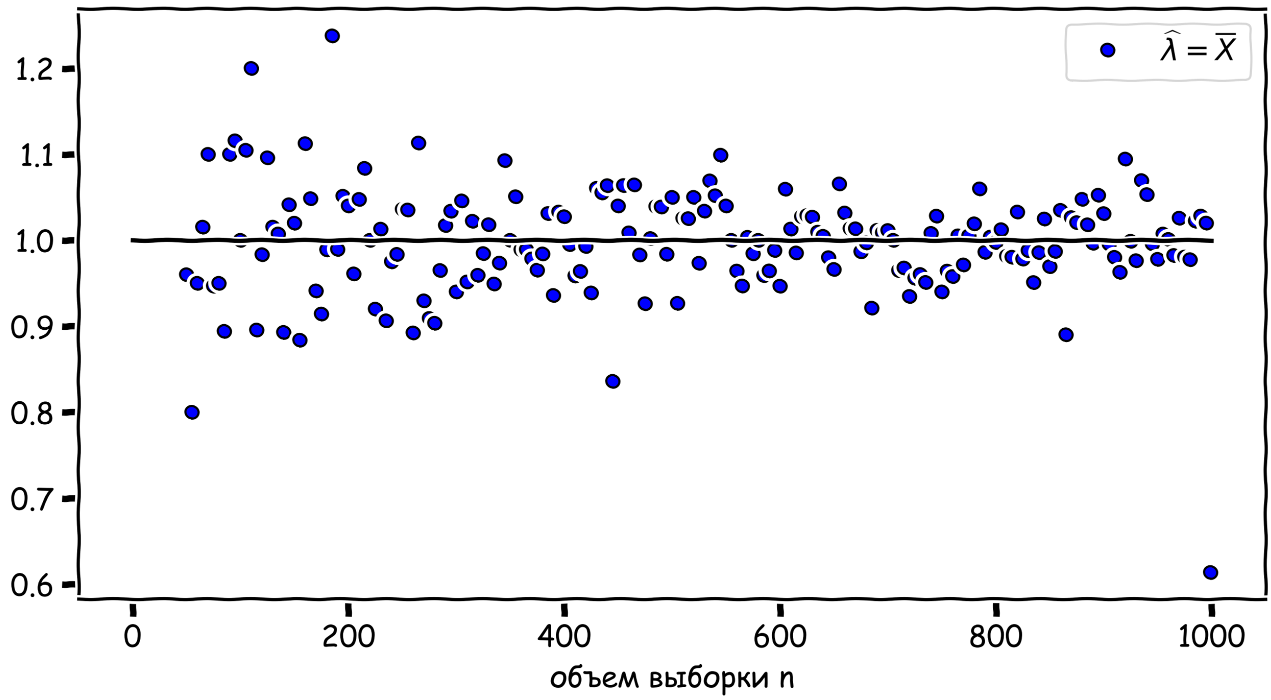

Рис. 9: Зависимость $\widehat{\lambda}$ от объема выборки

4. Рассмотрим равномерное распределение $\mathsf{U}_{a,b}$ с неизвестными параметрами $a < b$. Оказывается, что оценку этих параметров выгоднее всего производить, используя 1-ую и n-ую порядковые статистики, а именно:

$$\widehat{a} = X_{(1)}, \quad \widehat{b} = X_{(n)}.$$

Пример оценок значений a, b, построенных по выборкам из биномиального распределения с параметрами $a = 0, b = 10$, представлен на рисунке 10. Можно заметить, что оценки достаточно хорошо справляются со своей задачей.

5. Рассмотрим показательное распределение Exp_λ с неизвестным параметром $\lambda > 0$. Так как математическое ожидание $\mathsf{E}\xi$ случайной величины, имеющей показательное распределение, равно $\frac{1}{\lambda}$, то

$$\frac{1}{\widehat{\lambda}} = \overline{X} \Rightarrow \widehat{\lambda} = \frac{1}{\overline{X}}.$$

Пример оценок значения λ, построенных по выборкам из экспоненциального распределения с параметром $\lambda = \frac{1}{3}$, представлен на рисунке 11.

6. Рассмотрим нормальное распределение N_{a,σ^2} с неизвестными параметрами a, σ^2. Так как математическое ожидание $\mathsf{E}\xi$ случайной величины, имеющей нормальное распределение, равно a, а дисперсия $\mathsf{D}\xi$ равна σ^2, то

$$\widehat{a} = \overline{X}, \quad \widehat{\sigma^2} = S^2 \text{ или } \widehat{\sigma^2} = S_0^2.$$

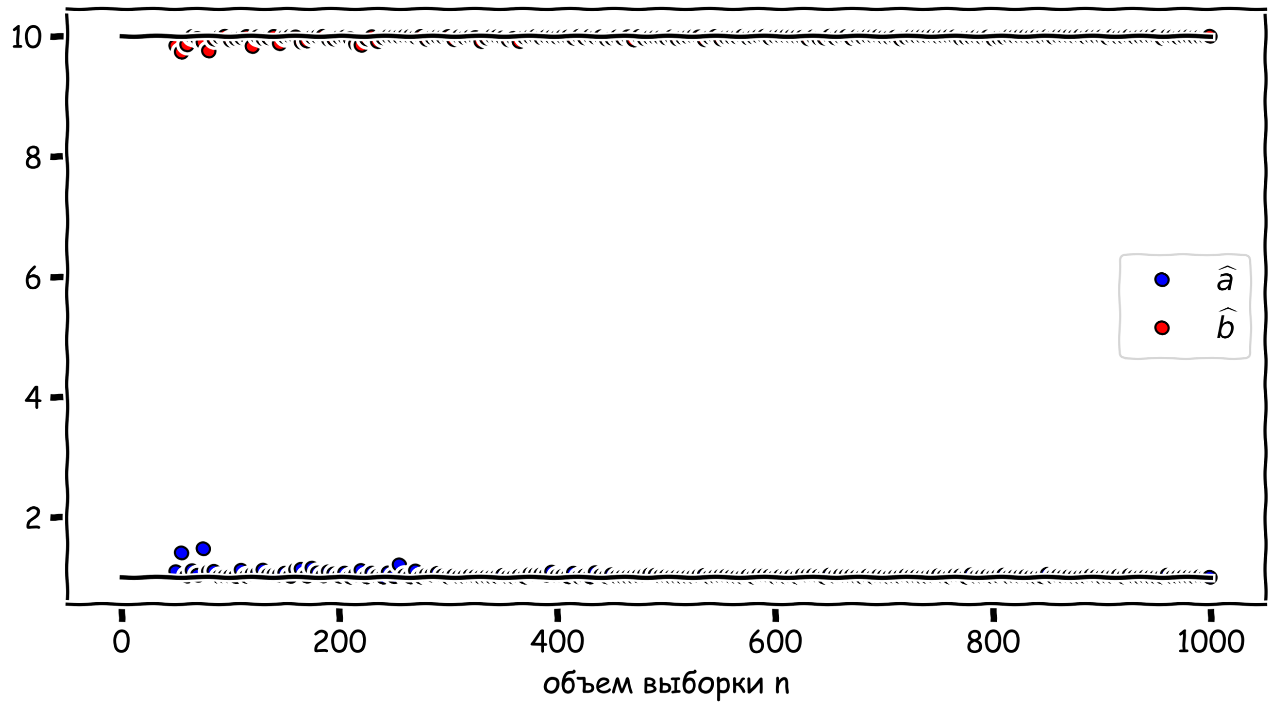

Рис. 10: Зависимость $\widehat{a}, \widehat{b}$ от объема выборки

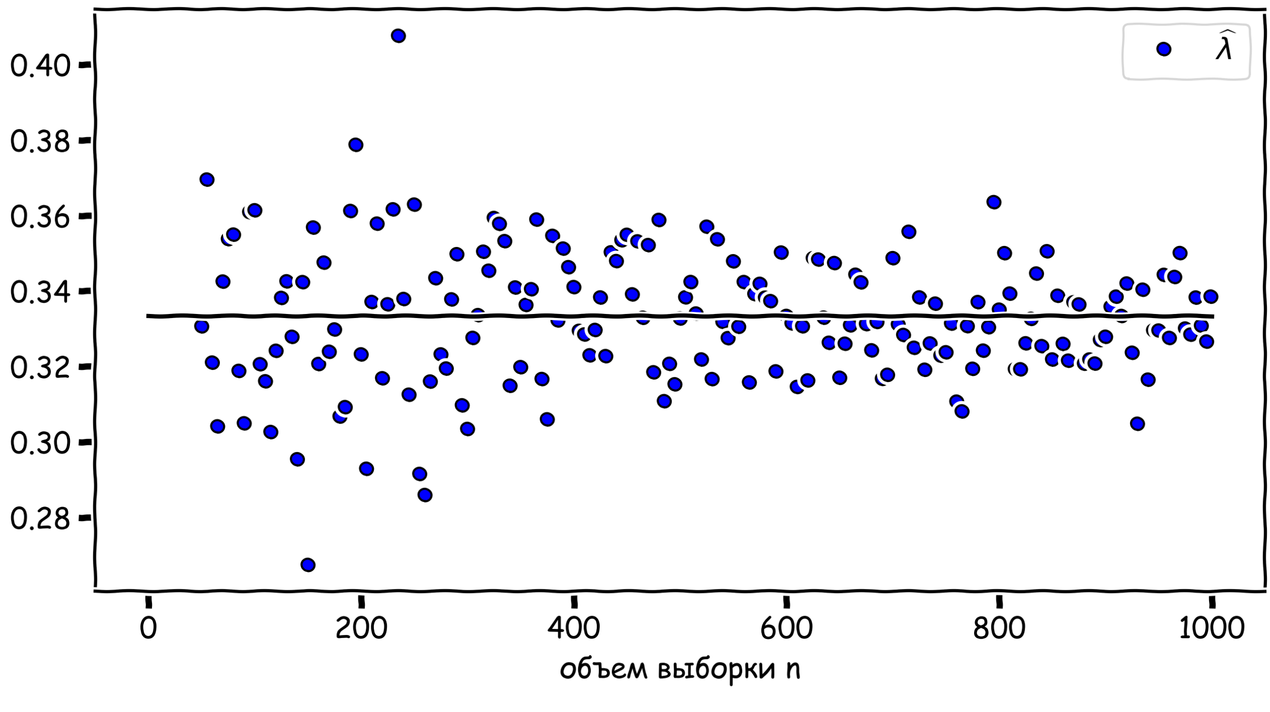

Рис. 11: Зависимость $\widehat{\lambda}$ от объема выборки

Пример оценок значений a, σ^2, построенных по выборкам из нормального распределения с параметрами $a = 2, \sigma^2 = 9$, представлен на рисунке 11.

Пример представлен на рисунке 12.

Популярные пакеты обработки и анализа данных позволяют генериро-

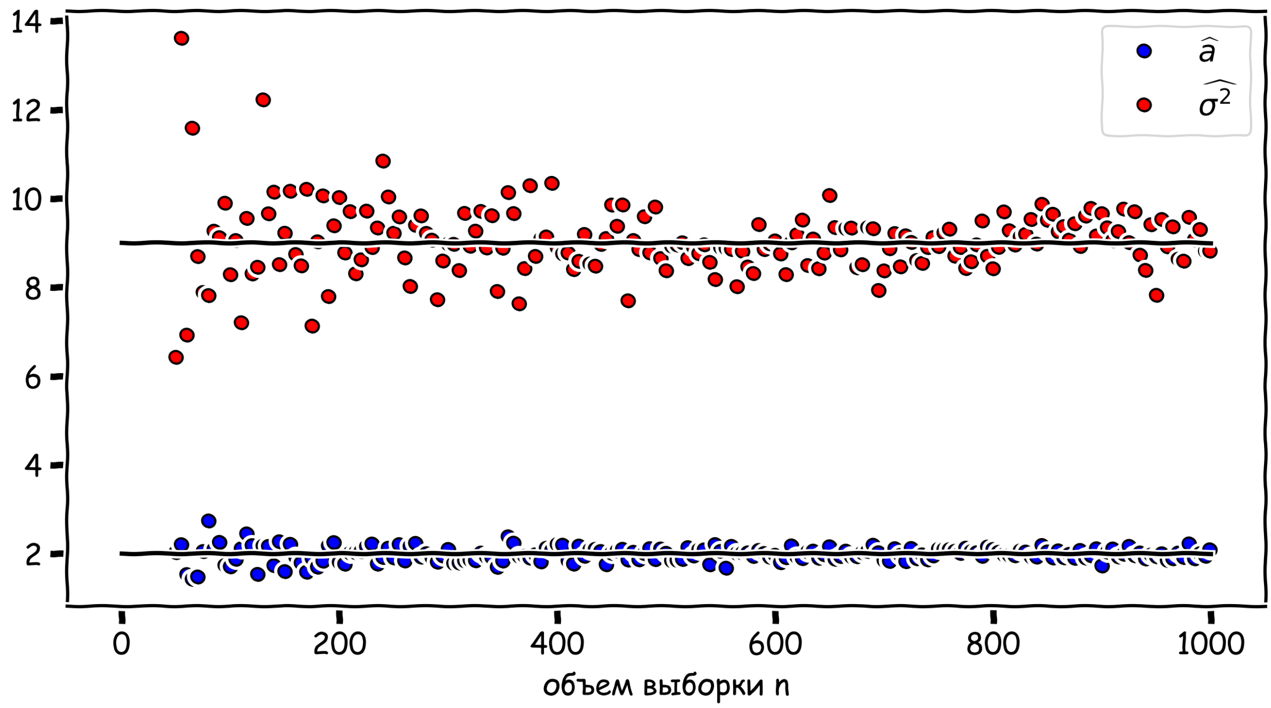

Рис. 12: Зависимость $\widehat{a}, \widehat{\sigma^2}$ от объема выборки

вать выборки из известных распределений с заданными параметрами. Например, в Excel это можно реализовать, используя надстройку «Анализ данных».

3 Закон больших чисел и свойства оценок

Итак, мы изучили практическое применение введенного аппарата, но до сих пор у нас нет никаких формальных доказательств того, что все рассказанное нами выше и правда работает. Смешно сказать, но мы, то и дело употребляя термин «оценка», ни разу даже не ввели соответствующего определения: а что такое оценка? Более того, говоря, что какая-то выборочная характеристика приближает истинную, мы тоже ни разу не конкретизировали: а в каком смысле приближает? Давайте попробуем разобраться.

3.1 А что вообще такое оценка?

Итак, начнем с того, а что такое оценка. Пусть имеется выборка $X = (X_1, X_2, ..., X_n)$ из генеральной совокупности ξ, а θ – некоторый параметр, который характеризует (или присущ) распределению случайной величины ξ. Так, в качестве θ может выступать математическое ожидание $\mathsf{E}\xi$, дисперсия $\mathsf{D}\xi$, медиана $\mathsf{med}\ \xi$ и так далее. Параметр θ может быть и каким-то более общим параметром распределения. Скажем, можно рассматривать семейство распределений $\mathsf{U}_{0,\theta}$, и тогда θ – это не самая явная характеристика генеральной совокупности (на самом деле $\theta = 2\mathsf{E}\xi$, подумайте почему).

Определение 3.1.1 *Оценкой параметра θ называется произвольная функция $\widehat{\theta}$, зависящая от выборки X, то есть*

$$\widehat{\theta} = \widehat{\theta}(X_1, X_2, ..., X_n).$$

Отметим также, что $\widehat{\theta}$ – это случайная величина, так как она является функцией от случайных величин $(X_1, X_2, ..., X_n)$.

Все оценки, которые мы вводили ранее, ровно-таки и были функциями от выборки, смотрите:

1. Выборочное среднее $\overline{X}$ – это средне арифметическое элементов выборки:

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}.$$

2. Выборочная дисперсия S^2 и несмещенная выборочная дисперсия S_0^2 – более хитрые функции, но тоже зависящие только лишь от выборки:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2, \quad S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

и так далее.

Однако достаточно очевидно, что не всякая функция от выборки является «хорошей» оценкой рассматриваемого параметра. Например, в качестве оценки $\widehat{\theta}$ можно взять тождественный ноль: $\widehat{\theta} \equiv 0$. Такая функция вполне подходит под определение, но является ли она разумной оценкой?

Конечно, разумная оценка – это та, которая, как мы уже не раз повторяли, приближает истинное значение параметра. А что значит приближает? Чтобы дать четкое определение, нам потребуется ввести понятие сходимости по вероятности.

3.2 Сходимость по вероятности и состоятельность

Итак, рассмотрим последовательность случайных величин ξ_n.

Определение 3.2.1 *Говорят, что последовательность случайных величин ξ_n сходится по вероятности к случайной величине ξ, если*

$$\forall\, \varepsilon > 0 \;\Rightarrow\; \mathsf{P}\left(|\xi_n - \xi| \geq \varepsilon\right) \xrightarrow[n\to+\infty]{} 0.$$

Обозначают сходимость по вероятности следующим образом:

$$\xi_n \xrightarrow[n\to+\infty]{\mathsf{P}} \xi.$$

Как можно пояснить это определение? На самом деле оно не зря выглядит пугающим, ведь для понимания оно совсем не тривиально. Утверждение, данное в определении, можно формулировать для себя так: с ростом n вероятность события, что ξ_n хоть как-то отличается от ξ, стремится к нулю.

Резонно задать и еще один вопрос: а при чем тут последовательность? Посмотрите, любая оценка $\widehat{\theta}$ – это тоже последовательность, ведь она зависит от объема выборки n:

$$\widehat{\theta} = \widehat{\theta}(X_1, X_2, ..., X_n).$$

Скажем, посмотрите на выборочное среднее $\overline{X}$:

$$\overline{X} = \frac{X_1 + X_2 + ... + X_n}{n}.$$

Сразу видно, что в выражении для выборочного среднего от n зависит как знаменатель, так и число слагаемых в числителе! Наверное, теперь можно догадаться, какая оценка будет считаться хорошей? Конечно та, к которая по вероятности будет сходиться к истинному значению параметра. Такие оценки называются состоятельными.

Определение 3.2.2 *Оценка $\widehat{\theta} = \widehat{\theta}(X_1, X_2, ..., X_n)$ называется состоятельной оценкой параметра θ, если*

$$\widehat{\theta} \xrightarrow[n \to +\infty]{\mathsf{P}} \theta.$$

Итак, состоятельность – понятное слово даже с точки зрения русского языка. Ведь если с ростом n, то есть с ростом объема выборки, значения нашей оценки не приближают истинного значения параметра, то оценка оказывается крайне несостоятельной: ведь она просто-напросто не состоялась, как оценка. За этой, казалось бы, игрой слов, на самом деле кроется достаточно большой смысл, постарайтесь его уловить.

Хорошо, а как доказать состоятельность хотя бы одной из наших оценок? В этом нам поможет так называемый закон больших чисел.

3.3 Закон больших чисел

Рассмотрим случайный эксперимент и некоторую случайную величину ξ, которая наблюдается в ходе этого эксперимента. Предположим, что математическое ожидание этой случайной величины равно a.

Если повторять этот эксперимент большое количество раз, наблюдая за ξ, то мы получим последовательность независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \ldots$ с математическим ожиданием a. Заранее предсказать, какие значения будут принимать в ходе экспериментов случайные величины ξ_i невозможно, так как эти значения зависят от случая. Однако, как показывает практика, среднее арифметическое значений случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, полученных в результате n экспериментов, приближается с ростом n к числу a. Этот эмпирический факт говорит о том, что поведение суммы большого числа случайных величин становится в определённом смысле закономерным.

Математическая формулировка явления носит название «Закон больших чисел» (ЗБЧ).

Теорема 3.3.1 (Закон больших чисел в форме Хинчина) *Пусть имеется последовательность $\xi_1, \xi_2, ..., \xi_n, ...$ независимых и одинаково распределенных случайных величин, имеющих математическое ожидание a. Тогда*

$$\forall \varepsilon > 0 \;\; \mathsf{P}\left(\left|\frac{\xi_1 + \xi_2 + ... + \xi_n}{n} - a\right| \geq \varepsilon\right) \xrightarrow[n \to +\infty]{} 0.$$

Итак, закон больших чисел утверждает, что в случае независимых и одинаково распределенных случайных величин, вероятность того, что их среднее арифметическое отклоняется от математического ожидания, стремится к нулю (по вероятности) с ростом n.

В введенных нами обозначениях это может быть записано, как

$$\frac{\xi_1 + \xi_2 + ... + \xi_n}{n} \xrightarrow[n \to +\infty]{\mathsf{P}} a = \mathsf{E}\xi_i = \mathsf{E}\xi_1,$$

так как раз все случайные величины одинаково распределены, то они имеют одно и то же математическое ожидание.

Именно на законе больших чисел основываются статистические методы оценивания неизвестных параметров распределения.

Теорема 3.3.2 (Состоятельность выборочного среднего) *Пусть существует математическое ожидание генеральной совокупности $\mathsf{E}\xi$. Тогда $\overline{X}$ – состоятельная оценка $\mathsf{E}\xi$.*

Доказательство. Пусть $X = (X_1, X_2, ..., X_n)$ – выборка из генеральной совокупности ξ. Тогда $X_1, X_2, ...$ – последовательность независимых случайных величин, имеющих такое же распределение, как и ξ. По условию существует математическое ожидание $\mathsf{E}\xi$, тогда, согласно закону больших чисел в форме Хинчина,

$$\frac{X_1 + X_2 + ... + X_n}{n} \xrightarrow[n \to +\infty]{\mathsf{P}} \mathsf{E}\xi.$$

□

Оказывается, что и все остальные введенные нами оценки: выборочная дисперсия, несмещенная выборочная дисперсия, выборочная медиана, выборочная корреляция – состоятельные оценки соответствующих характеристик генеральной совокупности ξ, а значит они и правда пригодны к использованию.

3.4 Несмещенность оценок

Последний долг, который у нас остался – это определить понятие несмещенности оценки.

Определение 3.4.1 *Оценка $\widehat{\theta} = \widehat{\theta}(X_1, X_2, ..., X_n)$ называется несмещенной оценкой параметра θ, если*

$$\mathsf{E}\widehat{\theta} = \theta.$$

Иными словами, свойство несмещенности означает, что в среднем наша оценка совпадает с оцениваемым параметром, или, точнее, что «в среднем» ее значения совпадают со значением оцениваемого параметра.

Теорема 3.4.1 *Пусть существует математическое ожидание генеральной совокупности $\mathsf{E}\xi$. Тогда $\overline{X}$ – несмещенная оценка $\mathsf{E}\xi$.*

Доказательство. Пусть $X = (X_1, X_2, ..., X_n)$ – выборка из генеральной совокупности ξ. Тогда $X_1, X_2, ...$ – последовательность независимых случайных величин, имеющих такое же распределение, как и ξ. По свойству математического ожидания,

$$\mathsf{E}\overline{X} = \mathsf{E}\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}X_i = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}\xi = \frac{n\mathsf{E}\xi}{n} = \mathsf{E}\xi.$$

□

Оказывается, что несмещенная выборочная дисперсия, несмещенная выборочная ковариация и корреляция являются несмещенными оценками.

Выборочная дисперсия же – смещенная оценка. Можно показать, что в предположении, что существует дисперсия генеральной совокупности $\mathsf{D}\xi$,

$$\mathsf{E}S^2 = \frac{n-1}{n}\mathsf{D}\xi.$$

На практике несмещенные оценки встречаются достаточно редко. Куда чаще возникают так называемые асимптотически несмещенные оценки.

Определение 3.4.2 *Оценка $\widehat{\theta} = \widehat{\theta}(X_1, X_2, ..., X_n)$ называется асимптотически несмещенной оценкой параметра θ, если*

$$\lim_{n\to+\infty}\mathsf{E}\widehat{\theta} = \theta.$$

Выборочная дисперсия, очевидно, является асимптотически несмещенной оценкой дисперсии генеральной совокупности $\mathsf{D}\xi$, так как

$$\lim_{n\to+\infty}\frac{n-1}{n} = 1.$$

Отсюда и следуют те умозаключения, что мы приводили выше. Смещенные, но асимптотически несмещенные оценки, хорошо работают на больших объемах выборки. На маленьких же объемах они могут давать существенные ошибки. На маленьких объемах предпочтительнее использовать несмещенные оценки.

4 Резюме

Итак, в этой лекции мы узнали, как по выборке можно оценить математическое ожидание, дисперсию и плотность генеральной совокупности, научились вычислять выборочную ковариацию и применять полученные оценки для оценок параметров распределений. Более того, мы попытались подкрепить все наши эмпирические соображения некоторой теорией: разобрали важность состоятельных оценок, поняли разницу между смещенными и несмещенными оценками, а также важный принцип: где и какие оценки применять. Однако осталась еще одна большая брешь, которую мы будем закрывать в следующей лекции: а как понять, что полученная нами оценка близка к истинному значению параметра? Скажем, получив в качестве оценки математического ожидания $\overline{X} = 5$, можем ли мы быть удовлетворены? Ведь, поменяв выборку, мы получим какое-то другое значение $\overline{X}$. И какое из них лучше? И далеко ли от них истинное?

С этими вопросами нам поможет разобраться так называемое интервальное оценивание, которое мы и будем рассматривать в следующей лекции.