Выборочные оценки параметров распределения

Резюмируя, важно подчеркнуть, что, с точки зрения экспериментатора, функции распределения и статистические характеристики случайной величины (среднее, дисперсия и т.д.) делятся на два класса:

- *генеральные* – те, которые как бы «внутренне» присущи случайной величине (они закрыты для нашего восприятия и мы можем только предполагать, основываясь на каких-либо физических предпосылках, что, к примеру, закон распределения будет иметь определенный тип);
- *выборочные* – те, что вычисляются на основе данных, полученных в ходе эксперимента.

Вообще говоря, типовые задачи математической статистики связаны с получением тех или иных интервальных и точечных оценок различных параметров случайной выборки. В частности, формулы (1-3) дают алгоритмы вычисления точечных оценок среднего и дисперсии.

Приведем теперь в качестве примера задачу интервального оценивания дисперсии нормальной случайной величины, которая ставится следующим образом. Требуется определить числовой интервал (L,U), внутри которого будет лежать с вероятностью $1-\alpha$ дисперсия нормальной случайной величины, исходя из объема выборки в n чисел. Указанный интервал называется $(1-\alpha)$–*доверительным интервалом*. Мы будем использовать данные рис. 6 (курс доллара минус его линейный тренд) и выберем для определенности доверительный интервал $1-\alpha=75\%$.

Поставленная задача решается в статистике с помощью χ^2-распределения, плотность распределения $p(y)\equiv\chi^2(y)$ для которого приведена на рис. 22, а функция распределения F(y) – на рис. 23. Как видно из графиков, χ^2–распределение имеет единственный параметр n, называемый *степенью свободы*. Известно, что сумма квадратов выборки из n независимых случайных чисел с нормальным распределением имеет закон распределения χ^2 с n степенями свободы. Благодаря такой связи с гауссовым распределением, χ^2–распределение широко используется в статистике для построения доверительных интервалов.

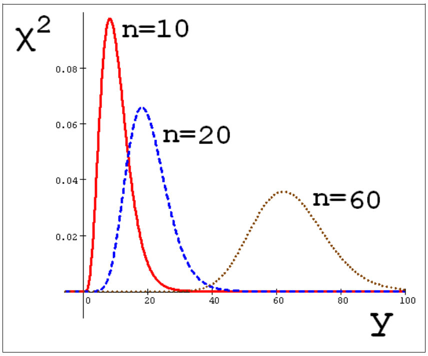

Рис. 22. *Плотность вероятности χ^2–распределения*

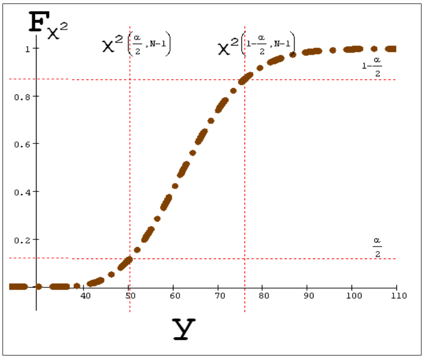

Рис. 23. *Функция распределения для χ^2–распределения*

Для решения задачи следует вычислить квантиль χ^2 – распределения с n-1 степенью свободы от аргументов $\alpha/2$ и $(1-\alpha/2)$. Соответствующие точки, которые мы обозначим символами q_1 и q_2, отложены на рис. 23 вертикальными пунктирными линиями. Затем надо расчитать по формуле (2) дисперсию выборки D, чтобы, наконец, получить оценку:

$$(L,U)=\left(\frac{D\cdot(n-1)^2}{q_1\cdot n},\frac{D\cdot(n-1)^2}{q_2\cdot n}\right) \quad . \qquad (5)$$

Результат оценивания показан на рис. 24, наряду с точками «курс доллара минус тренд». Заливкой мы выделили интервалы полученной 75%-й оценки стандартного отклонения. Подчеркнем, что решение мы получили, пользуясь предположением о нормальном распределении исходных данных (утверждение, вообще говоря, неправильное), которое дало нам право воспользоваться соответствующей формулой математической статистики (5).

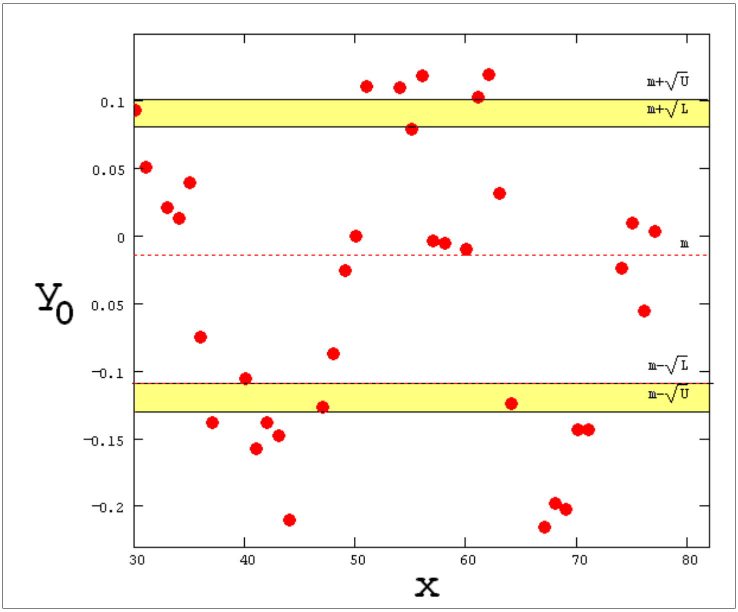

Рис. 24. *Интервальная оценка стандартного отклонения*

Проверка статистических гипотез

В статистике рассматривается огромное число задач, связанных с проверкой тех или иных гипотез H. Разберем пример простой гипотезы. Пусть имеется выборка N чисел с нормальным законом распределения и неизвестными дисперсией и математическим ожиданием. Требуется принять или отвергнуть гипотезу H о том, что математическое ожидание закона распределения равно некоторому числу m^0.

Задачи проверки гипотез требуют задания уровня критерия проверки α, который описывает вероятность ошибочного отклонения истинной гипотезы H. Если взять α очень малым, то гипотеза, даже если она ложная, будет почти всегда приниматься; если, напротив, взять a близким к 1, то критерий будет очень строгим, и гипотеза, даже верная, скорее всего, будет отклонена.

Рассмотрим, ради определенности, задачу исследования того же массива данных «курс доллара минус тренд», гипотеза для которой состоит в том, что $m^0=0.1$, а альтернатива – $m^0\neq 0.1$. Оценка математического ожидания, как следует из курса классической статистики, решается с помощью *распределения Стьюдента* (рис. 25) S(y) с параметром n-1 (этот параметр называется *степенью свободы* распределения). Для проверки гипотезы рассчитывается (1-α/2)-квантиль распределения Стьюдента T, который служит критическим значением для принятия или отклонения гипотезы. Если соответствующее выборочное значение

$$t \simeq \frac{m-m^0}{\sqrt{D/n}} \qquad (6)$$

по модулю окажется меньше T, то гипотеза принимается. В противном случае гипотезу следует отвергнуть. Символами m и D мы обозначили выборочное среднее и дисперсию, вычисляемые по данным эксперимента в соответствии с формулами (1-2).

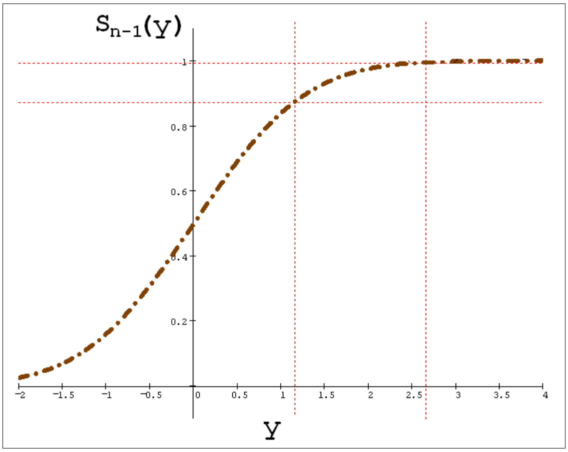

Рис. 25. *Функция распределения Стьюдента*

Если вычислить (6) для данных «курс доллара минус тренд», представленных на рис. 24, то мы увидим, что |t|=8.7, а квантиль T=1.16, т.е. гипотеза должна быть отвергнута. Если же взять пробную оценку $m^0=0$, то |t|=1.03, и гипотеза принимается.

Мы привели только два характерных примера, однако, пользуясь соответствующими формулами математической статистики и действуя примерно так же, можно без особых затруднений решать и другие задачи.