Математическая статистика

Деменков П.С.,
Иванисенко В.А.

Области применения мат. статистики

- имеется случайный эксперимент, свойства которого частично или полностью неизвестны
- мы умеем воспроизводить этот эксперимент в одних и тех же условиях некоторое (а лучше — какое угодно) число раз.

Основные понятия выборочного метода

Пусть $\xi : \Omega \to R$ — случайная величина, наблюдаемая в случайном эксперименте. Предполагается, что вероятностное пространство задано (и не будет нас интересовать).

Будем считать, что, проведя n раз этот эксперимент в одинаковых условиях, мы получили числа $X_1, X_2, \ldots, X_n$ — значения этой случайной величины в первом, втором, и т.д. экспериментах. Случайная величина ξ имеет некоторое распределение F, которое нам частично или полностью неизвестно.

Рассмотрим набор $X = (X_1, X_2, \ldots, X_n)$.

В серии уже произведенных экспериментов выборка — это набор чисел. Но если эту серию экспериментов повторить еще раз, то вместо этого набора мы получим новый набор чисел. Вместо числа X_1 появится другое число — одно из значений случайной величины ξ.

Определение: **Выборка объема n** **— это набор $X = (X_1, X_2, \ldots, X_n)$ из независимых и одинаково распределенных случайных величин («копий ξ»), имеющих, как и ξ, распределение F.**

Эмпирическая функция распределения

Поскольку неизвестное распределение *F* можно описать, например, его функцией распределения $F(y) = P(X_1 < y)$, построим по выборке «оценку» для этой функции.

Определение: **Эмпирической функцией распределения, построенной по выборке $X = (X_1, X_2, \ldots, X_n)$ объема n, называется случайная функция**

$$F_n^* : R \times \Omega \to [0,1]$$

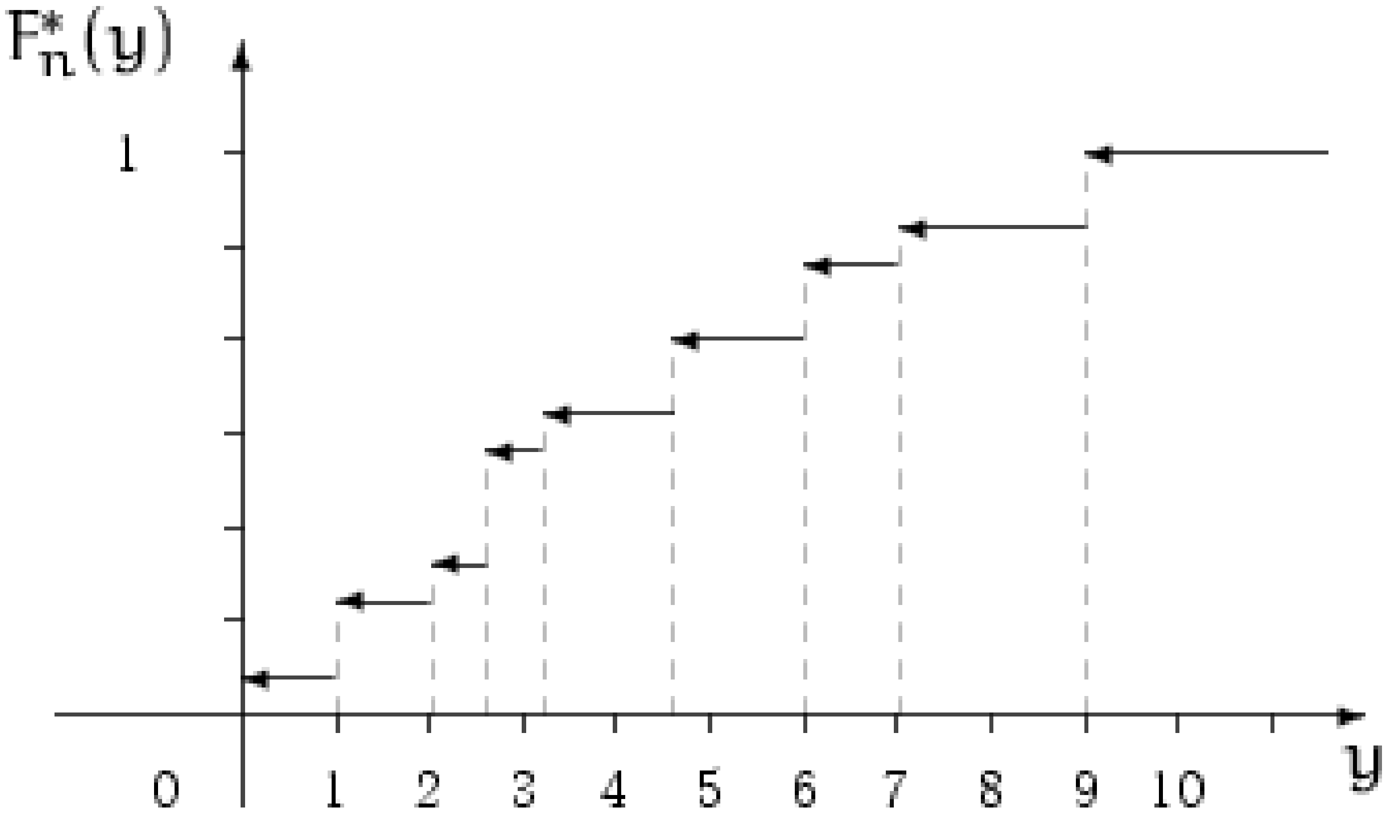
, при каждом $y \in R$ равная

$$F_n^*(y) = \frac{количество\_X_i \in (-\infty, y)}{n}$$

Пример эмпирической функции распределения

Выборка:

X = (0; 2; 1; 2,6; 3,1; 4,6; 1; 4,6; 6; 2,6; 6; 7; 9; 2,6)



Гистограмма

Другой характеристикой распределения является плотность

Определение: Под <u>плотностью распределения</u> будем понимать вероятность попасть в интервал *dx*

Для дискретных распределений плотность понимается как вероятность попасть в точку, а записывать ее удобно в таблицу.

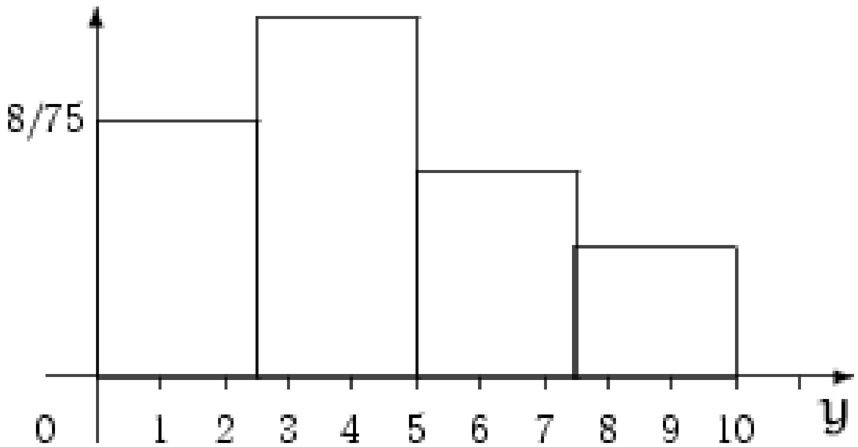
Эмпирическим, или выборочным аналогом плотности является так называемая гистограмма.

Гистограмма строится по группированным данным. Предполагаемую область значений случайной величины (или область выборочных данных) делят независимо от выборки на некоторое количество интервалов (не обязательно одинаковых), а затем вычисляют количество значений попавших в построенные интервалы.

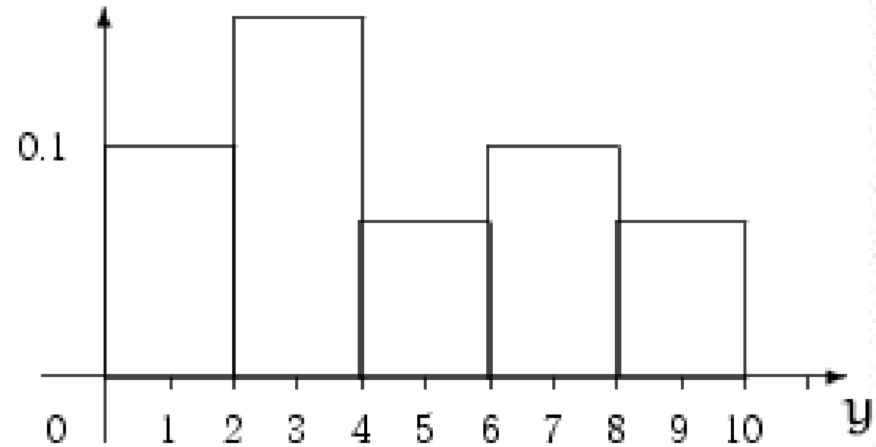
Пример построения гистограммы

Выборка:

X = (0; 2; 1; 2,6; 3,1; 4,6; 1; 4,6; 6; 2,6; 6; 7; 9; 2,6)



Интервал [0, 10] разбит на 4 равные части



Интервал [0, 10] разбит на 5 равных частей

Выборочные моменты

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Выборочное среднее

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

Выборочная дисперсия

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

Несмещенная выборочная дисперсия

$$\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Выборочный k-ый момент

Определение: **Статистикой** называется произвольная функция $\theta^* = \theta^*(X_1, \ldots, X_n)$ от элементов выборки.

Статистика есть функция от эмпирических данных, но никак не от параметра θ. Статистика, как правило, предназначена именно для оценивания неизвестного параметра θ (поэтому ее иначе называют «оценкой»), и уже поэтому от него зависеть не может.

Несмещенность статистик

Определение: Статистика $\theta^* = \theta^*(X_1, \ldots, X_n)$ называется <u>несмещенной оценкой</u> параметра θ, если для любого θ выполнено равенство

$$E_\theta \theta^* = \theta$$

Несмещенность — свойство оценок при фиксированном n. Означает это свойство отсутствие ошибки «в среднем», т.е. при систематическом использовании данной оценки.

Состоятельность статистик

Определение: Статистика $\theta^* = \theta^*(X_1, \ldots, X_n)$ называется <u>состоятельной оценкой</u> параметра θ, если для любого θ из области определения имеет место сходимость $\theta^* \xrightarrow{P} \theta$ при $n \to \infty$

Свойство состоятельности означает, что последовательность оценок приближается к неизвестному параметру при увеличении количества данных. Понятно, что при отсутствии этого свойства оценка совершенно «несостоятельна» как оценка.

Методы нахождения оценок: метод моментов

Метод моментов заключается в следующем:

любой момент случайной величины X_1 (например, k-й) зависит, часто функционально, от параметра θ. Но тогда и параметр θ может оказаться функцией от теоретического k-го момента. Подставив в эту функцию вместо неизвестного теоретического k -го момента его выборочный аналог, получим вместо параметра θ оценку θ^*.

Метод моментов

Пусть $X_1, X_2, \ldots, X_n$ — выборка объема n из параметрического семейства распределений F_θ. Выберем некоторую функцию $g(y)$ так, чтобы существовал момент $E_\theta g(X_1) = h(\theta)$ и функция h была обратима в области определения θ. Тогда в качестве оценки θ^* для θ возьмем решение уравнения $\overline{g(X)} = h(\theta^*)$

$$\theta = h^{-1}\left(E_\theta g(X_1)\right)$$

$$\theta^* = h^{-1}\left(\overline{g(X_1)}\right) = h^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right)$$

Чаще всего в качестве функции $g(y)$ берут $g(y)=y^k$

Состоятельность оценок метода моментов

Теорема: **Пусть** $\theta^* = h^{-1}\left(\overline{g(X_1)}\right)$ **— оценка параметра** θ**, полученная по методу моментов, причем функция** h^{-1} **непрерывна. Тогда** θ^* **<u>состоятельна</u>.**

Пример

Пусть $X_1, X_2, \ldots, X_n$ — выборка объема n из нормального распределения $N_{a,1}$ с неотрицательным средним $a \geq 0$.

Ищем оценку для по первому моменту:

$$E_a X_1 = a, \text{ поэтому } a^* = \overline{X}$$

Однако по условию $a \geq 0$, тогда как $\overline{X}$ может быть и отрицательно. В этом случае оценку корректируют. Например, в качестве ОММ берут ближайшую к построенной из области определения.

Итого: $a^* = \max\left\{0, \overline{X}\right\}$ — «исправленная» оценка метода моментов.

Методы нахождения оценок: метод максимального правдоподобия

Метод максимального правдоподобия — еще один разумный способ построения оценки неизвестного параметра. Состоит он в том, что в качестве «наиболее правдоподобного» значения параметра берут значение θ, максимизирующее вероятность получить при n опытах данную выборку $X = (X_1, X_2, \ldots, X_n)$. Это значение параметра θ зависит от выборки и является искомой оценкой.

Функция максимального правдоподобия

Определение: Функция (случайная величина при фиксированном θ)

$$f(X,\theta)=f_\theta(X_1)\cdot f_\theta(X_2)\cdot\ldots\cdot f_\theta(X_n)=\prod_{i=1}^{n} f_\theta(X_i)$$

называется функцией правдоподобия.

Функция

$$L(X,\theta)=\ln f(X,\theta)=\sum_{i=1}^{n}\ln(f_\theta(X_i))$$

называется логарифмической функцией правдоподобия.

$f_\theta(X_i)$ **- плотность распределения F_θ**

Оценка максимального правдоподобия

Определение: **Оценкой максимального правдоподобия** $\hat{\theta}$ **неизвестного параметра** θ **называют значение** θ**, при котором функция** $f(X, \theta)$ **достигает максимума (как функция от** θ **при фиксированных** $X_1, X_2, \ldots, X_n$**):**

$$\hat{\theta} = \arg\max_{\theta}\left(f\left(X, \theta\right)\right)$$

Пример

Пусть $X_1, X_2, \ldots, X_n$ — выборка объема n из нормального распределения N_{a,σ^2}, где $a \in R,\ \sigma > 0$; и оба параметра a, σ^2 неизвестны.

Плотность распределения $f_{(a,\sigma^2)}(y) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\dfrac{-(y-a)^2}{2\sigma^2} \right)$

Функция правдоподобия

$$f(X, a, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - a)^2}{2\sigma^2} \right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\sum_{i=1}^{n} (X_i - a)^2}{2\sigma^2} \right)$$

Логарифмическая функция правдоподобия

$$L(X, a, \sigma^2) = \ln f(X, a, \sigma^2) = -\ln(2\pi)^{n/2} - \frac{n}{2} \ln\sigma^2 - \frac{\sum_{i=1}^{n} (X_i - a)^2}{2\sigma^2}$$

Пример

В точке экстремума (по a,σ^2) гладкой функции L обращаются в нуль обе частные производные:

$$\frac{\partial}{\partial a}L(X,a,\sigma^2)=\frac{2\sum_{i=1}^{n}(X_i-a)}{2\sigma^2}=\frac{n\bar{X}-na}{\sigma^2};\qquad \frac{\partial}{\partial\sigma^2}L(X,a,\sigma^2)=-\frac{n}{2\sigma^2}+\frac{\sum_{i=1}^{n}(X_i-a)^2}{2\sigma^4}$$

Оценка максимального правдоподобия $(\hat{a},\hat{\sigma}^2)$ **для** (a,σ^2) **— решение системы уравнений**

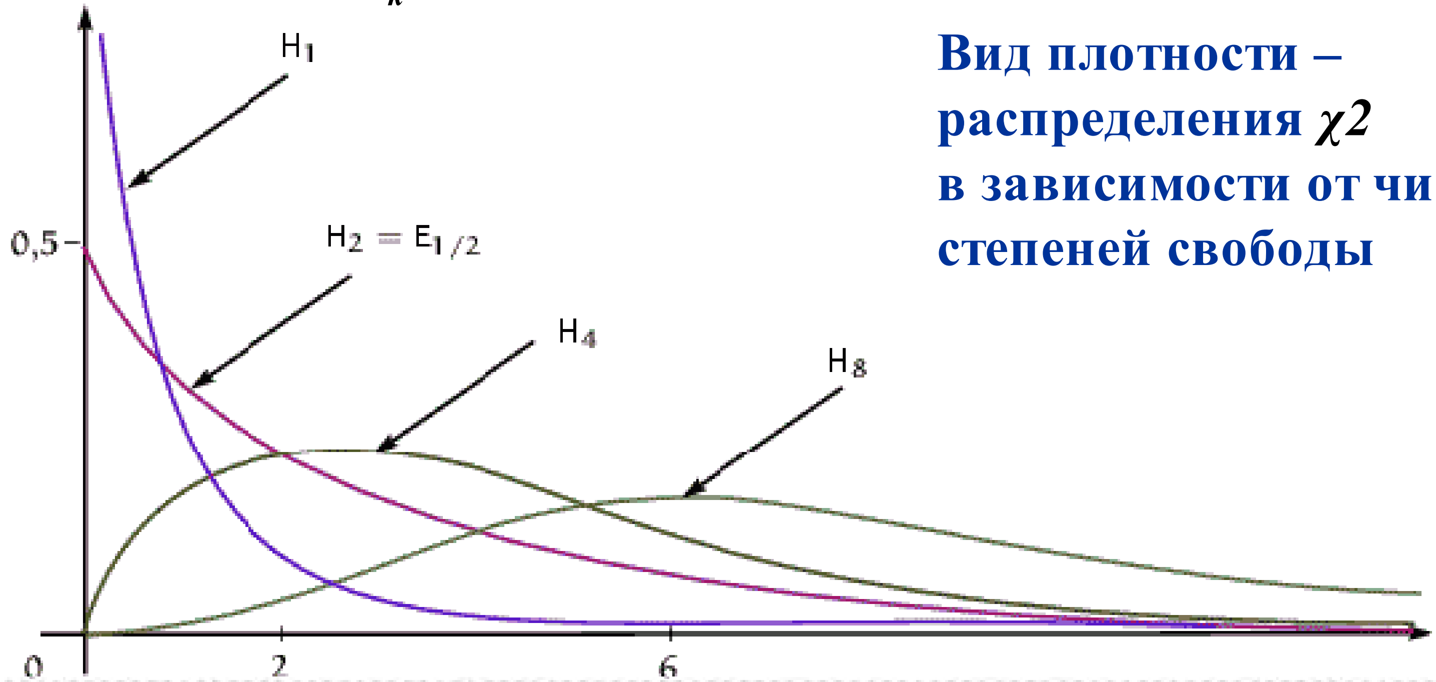
$$\frac{n\bar{X}-na}{\sigma^2}=0;\qquad -\frac{n}{2\sigma^2}+\frac{\sum_{i=1}^{n}(X_i-a)^2}{2\sigma^4}=0$$

Решением системы являются

$$\hat{a}=\bar{X}\qquad \sigma^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

Распределение «хи-квадрат»

Определение: **Распределение суммы *k* квадратов независимых стандартных нормальных случайных величин называют распределением «хи-квадрат» с *k* степенями свободы и обозначают H_k.**



Вид плотности – распределения $\chi 2$ в зависимости от числа степеней свободы

Свойства χ^2 распределения

1. **Устойчивость по суммированию**

   Пусть случайная величина ξ имеет распределение H_k, случайная величина η имеет распределение H_m, причем эти случайные величины независимы. Тогда их сумма имеет распределение H_{k+m}.

3. **Моменты распределения $\chi 2$**

   Если ξ имеет распределение H_k, то $E\xi = k$, $D\xi = 2k$

5. **Если ξ_1, ξ_2,…, ξ_k независимые случайные величины с нормальным распределением N_{a,σ^2}, то**

$$\xi = \sum_{i=1}^{k} \left(\frac{\xi_i - a}{\sigma} \right)^2$$

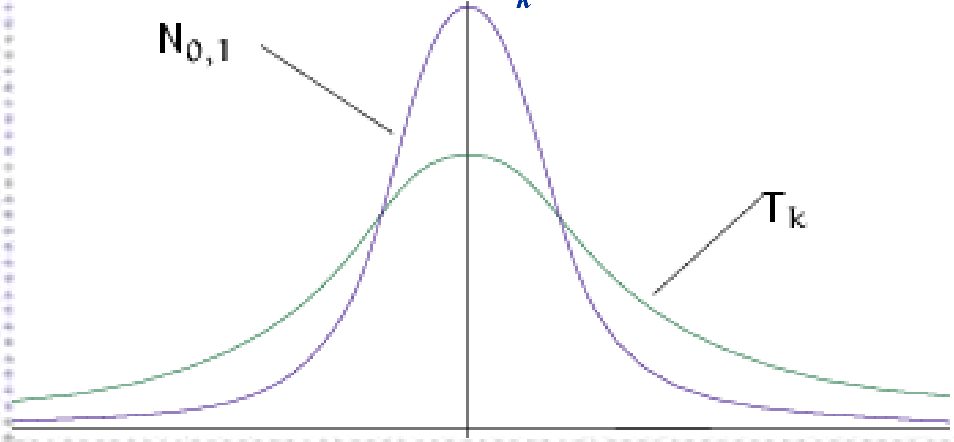
имеет $\chi 2$ распределение с k степенями свободы

Распределение Стьюдента

Определение: **Пусть ξ_0, $\xi_1, \ldots, \xi_k$ независимы и имеют стандартное нормальное распределение. Тогда распределение случайной величины**

$$t_k = \frac{\xi_0}{\sqrt{\frac{1}{k}\left(\xi_1^2 + \ldots + \xi_k^2\right)}} = \frac{\xi_0}{\sqrt{\frac{\chi_k^2}{k}}}$$

называют <u>распределением Стьюдента с k степенями свободы</u> и обозначают T_k.



Плотность распределения Стьюдента по сравнению с плотностью стандартного нормального распределения

Свойства распределения Стьюдента

1. Симметричность.

   Если случайная величина ξ имеет распределение Стьюдента с k степенями свободы, то и $-\xi$ имеет такое же распределение.

3. Асимптотическая нормальность.

   Распределение Стьюдента T_k слабо сходится к стандартному нормальному распределению при $k \to \infty$.

Проверка гипотез

Пусть дана выборка $X_1, X_2, \ldots, X_n$ из распределения F. Если не оговорено противное, считается, что все наблюдения имеют одно и то же распределение.

Определение: <u>Гипотезой</u> (H) называется любое предположение о распределении наблюдений:

$$H = \{F = F_1\} \text{ или } H = \left\{F \in \left\{ \hat{F} \right\}\right\}$$

Гипотеза H называется простой, если она однозначно определяет распределение, иначе H называется сложной. Сложная гипотеза предполагает, что распределение F — одно из некоторого множества распределений $\{\hat{F}\}$.

Если гипотез всего две, то одну из них принято называть основной, а другую — альтернативой или отклонением от основной гипотезы.

Типичные постановки задач

- **Выбор из нескольких простых гипотез:**
  $H_1 = \{F = F_1\},\ldots,H_k = \{ F = F_k\}$ **(и другие предположения невозможны).**
- **Простая основная гипотеза и сложная альтернатива:**
  $H_1 = \{F = F_1\}$ **и** $H_2 = \{F \neq F_1\}$
- **Сложная основная гипотеза и сложная альтернатива, например гипотеза нормальности** $H_1 = \{F \text{ нормальное с параметрами } a, \sigma^2\}$ **и** $H_2 = \{H_1 \text{ неверна}\}$
- **Гипотеза независимости.**

Виды ошибок

Определение: **Для заданного критерия** $\delta : R^n \to \{H_1, \dots, H_k\}$ **будем говорить, что произошла** <u>**ошибка** i**-го рода**</u>**, если гипотеза** H_i **отвергнута критерием, в то время как она верна. Вероятностью ошибки** i**-го рода критерия** δ **называется**

$$\alpha_i(\delta) = P_{H_i}(\delta(X) \neq H_i)$$

Если гипотеза H_i **простая, т.е. указывает ровно на одно возможное распределение выборки, то** $\alpha_i(\delta)$**— число. Если же** H_i**— сложная гипотеза, то** $\alpha_i(\delta)$ **будет зависеть от того, при каком именно из распределений** F_i**, отвечающих** H_i**, вычисляется вероятность:**

$$\alpha_i(\delta) = \alpha_i(\delta, F_i) = P_{F_i}(\delta(X) \neq H_i)$$

Замечания

1. Статистический критерий не отвечает на вопрос, верна или нет проверяемая гипотеза. Он лишь решает, противоречат или не противоречат выдвинутой гипотезе выборочные данные, можно ли принять или следует отвергнуть данную гипотезу.
2. Если есть одна основная гипотеза, а все остальное — нежелательные отклонения от нее, то вывод «данные противоречат гипотезе» всегда весомее, нежели вывод «данные не противоречат гипотезе».
3. Нам неизвестно, какая из гипотез верна в действительности, поэтому следует считаться с гипотетическими вероятностями ошибок критерия. Если много раз применять критерий к выборкам из распределения, для которого гипотеза H_i верна, то примерно доля α_i таких выборок будет признана противоречащей гипотезе H_i.

Критерии согласия

Критериями согласия называют критерии, предназначенные для проверки простой гипотезы $H_1=\{F=F_1\}$ при сложной альтернативе $H_1=\{H_1 \text{ неверна}\}$

Мы рассмотрим более широкий класс основных гипотез, включающий и сложные гипотезы, а критериями согласия будем называть любые критерии, устроенные по одному и тому же принципу.

Пусть задана некоторая функция отклонения эмпирического распределения от теоретического, распределение которой существенно разнится в зависимости от того, верна или нет основная гипотеза. Критерии согласия принимают или отвергают основную гипотезу исходя из величины этой функции отклонения.

Пусть дана выборка $X = (X_1, X_2, \ldots, X_n)$ из распределения F. Будем проверять простую гипотезу $H1=\{F=F1\}$ при сложной альтернативе $H_1=\{F\neq F_1\}$.

К1: Пусть возможно задать функцию отклонения $\rho(X)$, обладающую свойствами:

- **a) если гипотеза H_1 верна, $\rho(X)\Rightarrow G$, где G – непрерывное распределение**
- **b) Если гипотеза H_1 неверна, то** $\left|\rho(X)\right| \xrightarrow{P} \infty$ *при* $n \to \infty.$

К2: Пусть такая функция $\rho(X)$ задана. Для случайной величины η из распределения G определим постоянную C из равенства $P(|\eta| \geq C)=\varepsilon$

$$\delta(X)=\begin{cases} H_1, & \text{если } \left|\rho(X)\right| < C, \\ H_2, & \text{если } \left|\rho(X)\right| \geq C. \end{cases}$$

Определение: **Говорят, что критерий δ для проверки простой гипотезы H_1 является критерием асимптотического размера ε, если его размер приближается к ε с ростом n:**

$$\alpha_1(\delta) = P_{H_1}(\delta(X) \neq H_1) \to \varepsilon \; при \; n \to \infty.$$

Определение: **Критерий δ для проверки гипотезы H_1 против сложной альтернативы H_2 называется состоятельным, если для любого распределения F_2, отвечающего альтернативе H_2, вероятность ошибки второго рода стремится к нулю с ростом объема выборки:**

$$\alpha_2(\delta, F_2) = P_{F_2}(\delta(X) = H_1) \to 0 \; при \; n \to \infty.$$

Свойства критериев согласия

1. $\alpha_1(\delta) = P_{H_1}(|\rho(X)| \geq C) \to P(|\eta| \geq C) = \varepsilon$ *при* $n \to \infty$.

2. $\alpha_2(\delta, F_2) = P_{F_2}(|\rho(X)| < C) \to 0$ *при* $n \to \infty$ **для любого распределения $F_2 \neq F_1$**

Таким образом, построенный критерий имеет асимптотический размер ε и состоятелен.

Пусть дана выборка $X = (X_1, X_2, \dots, X_n)$ из распределения F. Проверяется простая гипотеза $H_1 = \{ F = F_1 \}$ против сложной альтернативы $H_2 = \{ F \neq F_1 \}$.

Построим разбиение области значений случайной величины с распределением F_1 на интервалы $A_1, \dots, A_k$.

Обозначения:

$\nu_j = \{$ *число* X_i *попавших в* $A_j \}$

$$p_j = P_{H_1}(X_1 \in A_j) > 0$$

$p_1 + \dots + p_k = 1$ **- необходимое условие**

Как правило, длины интервалов выбирают так, чтобы

$p_1 = \dots = p_k = 1/k$

критерий χ^2 Пирсона

$$\rho(X)=\sum_{j=1}^{k}\frac{(\nu_j-np_j)^2}{np_j}$$

Свойство К1(b) выполнено далеко не для всех альтернатив. Если распределение выборки $F_2 \neq F_1$ имеет такие же, как у F_1, вероятности попадания в каждый из интервалов A_j, то по данной функции ρ эти распределения различить невозможно.

Теорема(Пирсона): Если верна гипотеза H_1, то при фиксированном k и при $n\to\infty$

$$\rho(X)=\sum_{j=1}^{k}\frac{(\nu_j-np_j)^2}{np_j}\Rightarrow H_{k-1}$$

Проверка гипотезы независимости: критерий хи-квадрат Пирсона

Дана выборка $(X,Y)=((X_1,Y_1),\dots,(X_n,Y_n))$ значений двух наблюдаемых совместно случайных величин X и Y в n независимых экспериментах. Проверяется гипотеза $H_1=\{X \text{ и } Y \text{ независимы}\}$

Построим интервалы $A_1,\dots,A_k$ для значений X и $B_1,\dots, B_m$ для значений Y

Обозначим:

$\nu_{i,j}=\{$*число пар* (X,Y) *попавших в* $(A_i \times B_j)\}$

$\nu_{\cdot,j}=\{$*число* Y_l *попавших в* $B_j\}$

$\nu_{i,\cdot}=\{$*число* X_l *попавших в* $A_i\}$

Проверка гипотезы независимости

Если гипотеза H_1 верна, то теоретические вероятности попадания пары *(X,Y)* в любую из областей *(A_i x B_j)* равны произведению вероятностей: для всех *i* и *j*.

$$p_{i,j} = P\left(\left(X,Y\right) \in A_i \times B_j\right) = P\left(X \in A_i\right) \cdot P\left(Y \in B_j\right) = p_i^x \cdot p_i^y$$

Теорема: **Если гипотеза H_1 верна, то**

$$\rho\left(X,Y\right) = n \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\left(\nu_{i,j} - \left(\nu_{i,\bullet} \nu_{\bullet,j}\right) / n\right)^2}{\nu_{i,\bullet} \nu_{\bullet,j}} \Rightarrow H_{(k-1)(m-1)} \text{ при } n \to \infty$$

Критерий согласия строится обычным образом

Статистическая зависимость

Часто требуется определить, как зависит наблюдаемая случайная величина от одной или нескольких других величин. Самый общий случай такой зависимости — зависимость статистическая: например, $X=\xi+\eta$ и $Z=\xi+\chi$

Для зависимых случайных величин имеет смысл рассмотреть математическое ожидание одной из них при фиксированном значении другой (других). Такое условное математическое ожидание показывает, как влияет на среднее значение первой величины изменение значений второй.

Математическая модель регрессии

Определение: Функция $f(t)=E(X \mid Z=t)$ называется линией регрессии X на Z , а уравнение $x=f(t)$ - регрессионным уравнением.

После n экспериментов, в которых Z последовательно принимает значения $t_1,\ldots,t_n$, получим значения наблюдаемой величины X, равные $X_1,\ldots,X_n$.

Обозначим $\varepsilon_i = X_i - f(t) = X_i - E(X \mid Z = t_i)$ – ошибку наблюдения

Про распределение $\varepsilon_1,\ldots, \varepsilon_n$ обычно что-либо известно или предполагается: например, что вектор ошибок ε состоит из независимых и одинаково нормально распределенных случайных величин с нулевым средним.

Требуется по значениям $t_1,\ldots,t_n$ и $X_1,\ldots,X_n$ оценить как можно точнее функцию $f(t)$.

Будем предполагать, что функция $f(t)$ полностью определяется неизвестными параметрами $\theta_1,\ldots,\theta_k$

Метод максимального правдоподобия

Будем выбирать неизвестные параметры так, чтобы максимизировать функцию правдоподобия случайного вектора $X_1, X_2, \ldots, X_n$.

Пусть все ε_i имеют плотность распределения $h(x)$. Так как X_i зависит от ε_i линейно, то распределение у X_i будет таким же как у ε_i, но с центром не в нуле, а в точке $f(t_i)$. Поэтому $h(x-f(t_i))$ – плотность распределения X_i.

Так как $X_1, X_2, \ldots, X_n$ независимы, то функция правдоподобия

$$f(X_1, X_2, \ldots, X_n; \theta_1, \theta_2, \ldots, \theta_k) = h(x - f(t_1)) \cdot \ldots \cdot h(x - f(t_n)) = h(\varepsilon_1) \cdot \ldots \cdot h(\varepsilon_n)$$

Метод максимального правдоподобия предписывает находить оценки неизвестных параметров θ_i функции $f(t)$ и оценки неизвестной дисперсии $D\varepsilon_i$, максимизируя по этим параметрам функцию правдоподобия

Метод наименьших квадратов

Предположим, что вектор ошибок *ε* состоит из независимых случайных величин с нормальным распределением N_{0,σ^2} **.**

Функция правдоподобия имеет вид

$$f(X;\theta)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(X_1-f(t_1))^2}{2\sigma^2}\right\}\bullet\ldots\bullet\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(X_1-f(t_1))^2}{2\sigma^2}\right\}=$$

$$=\frac{1}{\sigma^n(2\pi)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-f(t_i))^2\right\}$$

Очевидно, что при любом фиксированном σ^2 максимум функции правдоподобия достигается при наименьшем значении суммы квадратов ошибок.

$$\sum_{i=1}^{n}(X_i-f(t_i))^2=\sum_{i=1}^{n}\varepsilon^2$$

Оценка метода наименьших квадратов

Определение: Оценкой метода наименьших квадратов для неизвестных параметров $\theta_1,\ldots,\theta_k$ уравнения регрессии называется набор значений параметров, доставляющий минимум сумме квадратов отклонений

$$\sum_{i=1}^{n}\left(X_i - f(t_i)\right)^2 = \sum_{i=1}^{n}\varepsilon^2$$

Найдя оценки для θ_i, найдем тем самым оценку $\hat{f}(t)$ для $f(t)$. Обозначим через $\hat{\varepsilon}_i = X_i - f(t_i)$ соответствующие оценки ошибок. Оценка максимального правдоподобия для σ^2, она же точка максимума по функции правдоподобия, равна

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \hat{f}(t_i)\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$