Интервальное оценивание параметров

Распределение χ^2 (Хи-квадрат),
t **- распределение (Стюдента),**
F – распределение (Фишера)

(Ахметов С.К.)

Три теоремы математической статистики

Сначала рассмотрим три теоремы математической статистики. Их суть состоит в определении закона распределения для СВ, которая является функцией других СВ

- **Распределение χ^2 (Хи-квадрат)**
- **t - распределение (Стьюдента)**
- **F – распределение (Фишера)**

Распределение χ^2 (Хи-квадрат)

Теорема 1. Если X_i - независимые СВ, подчиняющиеся нормальному закону распределения и у которых m_x равно нулю, а σ_x равно единице, то СВ

$$\chi^2 = \sum_{i=1}^{v} X_i^2$$

подчиняется распределению χ^2 (хи – квадрат) с v степенями свободы.

Распределение χ^2 определяется одним параметром v, который называется числом степеней свободы (значение v равно числу независимых СВ под знаком суммы)

Распределение χ^2 (Хи-квадрат)

Плотность вероятности распределения χ^2 равна

$$f(\chi^2) = \begin{cases} \left[2^{\frac{\nu}{2}} \Gamma(\nu/2)\right]^{-1} e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1} & \text{при } x > 0; \\ 0 & \text{при } x \le 0, \end{cases}$$

где ***Г(•)*** – гамма – функция; ***x*** – значение СВ χ^2.

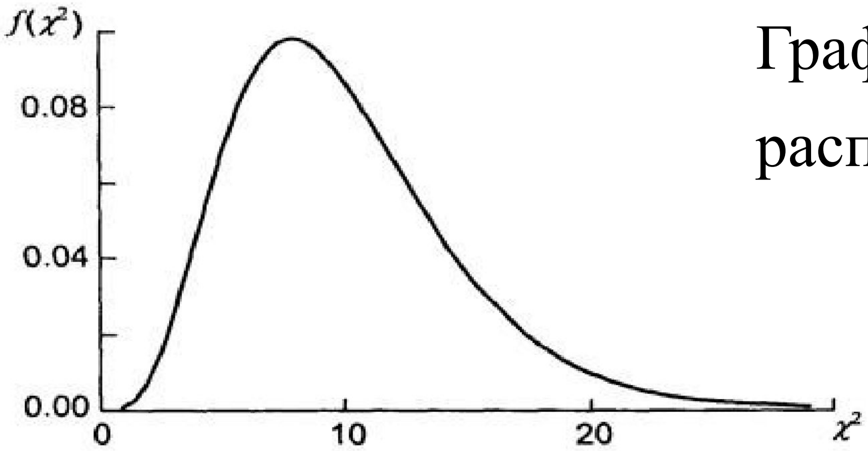


График плотности вероятности распределения *хи* - квадрат

Распределение χ^2 (Хи-квадрат)

- Математическое ожидание и дисперсия распределения χ^2 равны: $m_x = v$ и $D_x = 2v$
- Медиана может быть определена приближенным равенством: $Me = v - 0,66$
- Мода при $v \geq 2$ равна: $Mo = v - 2$

При $v = 1$ мода отсутствует, так как $f_v = \infty$ при $x = 0$

При увеличении значения v распределение χ^2 приближается к нормальному распределению

Распределение χ^2 (Хи-квадрат)

 В случае, если ***v > 30***, то можно использовать формулу

$$\chi_p^2(\nu) \approx \frac{1}{2}\left(\sqrt{2\nu - 1} + t'_p\right)^2$$

где ***t'ₚ*** – квантиль нормального распределения с ***m_t = 0*** и ***σ_t = 1***

p – вероятность не превышения

Это приближение не подходит при ***p,*** близких к ***0*** или ***100%.*** В этих случаях рекомендуется формула

$$\chi_p^2(\nu) \approx \nu\left[1 - 2/(9\nu) + t'_p\sqrt{2/(9\nu)}\right]^3$$

Распределение χ^2 (Хи-квадрат)

В конечном итоге из изложенной выше теоремы следует, что

$$(n-1)[S^2_x/\sigma^2_x]$$

имеет распределение χ^2 с $(n-1)$ степенями свободы,
где S^2_x и σ^2_x – соответственно выборочная и теоретическая дисперсии)

Значения квантилей χ^2 распределения даются в таблицах

t - распределение (Стьюдента)

Теорема 2. Если Z – нормированная нормально распределенная СВ, а U – независимая от Z СВ, подчиненная распределению χ^2 с v степенями свободы, тогда СВ $t = Z\sqrt{v/U}$ подчиняется распределению Стьюдента с v степенями свободы

Распределение Стьюдента называется также t – распределением.

t - распределение (Стьюдента)

Плотность вероятности этого распределения определяется равенством

$$f_{\nu}(t) = c(\nu)\left[1 + t^{2}/\nu\right]^{-(\nu+1)/2}$$

где ***c(ν)*** - параметр, зависящий от числа степеней свободы:

$$c(\nu) = \Gamma\left[(\nu+1)/2\right]/\left[\sqrt{\nu\pi}\,\Gamma(\nu/2)\right]$$

Г(•) – гамма – функция; ***π*** – число «пи».
Распределение Стьюдента симметрично.

t - распределение (Стьюдента)

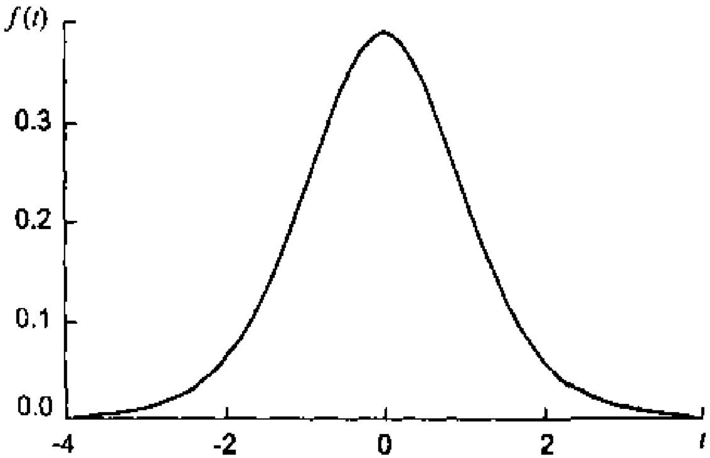


График функции плотности вероятности

Математическое ожидание m_t дисперсия D_t и среднее квадратичное отклонение σ_t равны: $m_t = 0$; $v = 1$; $D_t = \sigma_t^2 = v/(v - 2)$, $v > 2$.

С увеличением v распределение Стьюдента асимптотически приближается к нормальному распределению с параметрами $m_t = 0$ и $\sigma_t = 1$.

t - распределение (Стьюдента)

Из этой теоремы следует, что величина

$$(x_{ср.} - m_{x)}/(S/\sqrt{n})$$

имеет распределение Стьюдента,

где $x_{ср.}$ и S – выборочное среднее и СКО

n – длина выборки.

F – распределение (Фишера)

Теорема 3. Если Z и U независимые СВ, обладающие χ^2 распределением с v_1 и v_2 степенями свободы, то СВ $F = (Z/v_1)/(U/v_2)$ имеет распределение Фишера с v_1 и v_2 степенями свободы. Это распределение также называется F – распределением.

F – распределение (Фишера)

Плотность вероятности ***F*** – распределения имеет вид

$$f(F)=\begin{cases} c_1(\nu_1,\nu_2)F^{\frac{\nu_1-2}{2}}\Big/\left(1+\dfrac{\nu_1}{\nu_2}F\right)^{\frac{\nu_1+\nu_2}{2}} & \text{при } x>0; \\ 0 & \text{при } x\le 0, \end{cases}$$

где $c_1(\nu_1,\nu_2)$ – параметр, зависящий от ν_1 и ν_2.

$$c_1(\nu_1,\nu_2)=\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}\Gamma\left[\frac{\nu_1+\nu_2}{2}\right]\Big/\left[\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)\right]$$

F – распределение (Фишера)

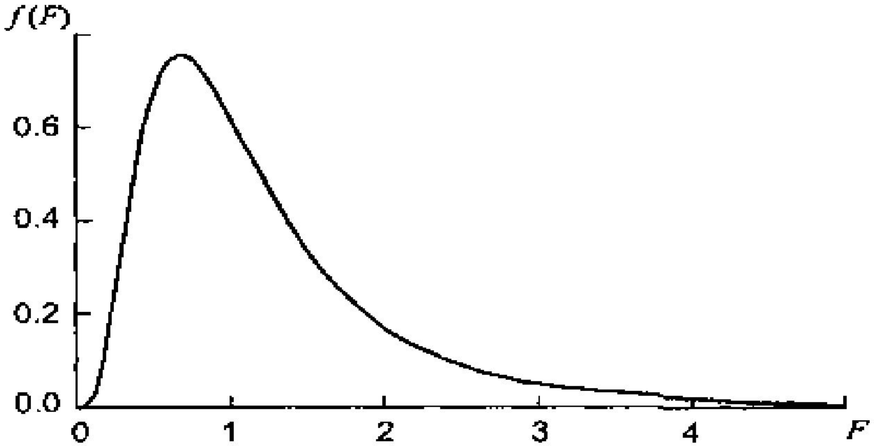


График плотности вероятности $f(F)$

Математическое ожидание, дисперсия и мода соответственно равны

$$m_F = v_2/(v_2 - 2), \quad v_2 > 2$$
$$D_F = 2\, v_2^2(v_1 + v_2 - 2)[v_1(v_2 - 2)^2(v_2 - 4)], \; v_2 > 4$$
$$M_0 = v_2(v_1 - 2)[v_1(v_2 + 2)], \; v > 1$$

Из этой теоремы следует, что отношение выборочных дисперсий S_1^2/S_2^2 двух выборок длиной m и n будет иметь F – распределение с числом степеней свободы соответственно $v_1 = (m-1)$ и $v_2 = (n-1)$

Интервальные оценки параметров распределения

Интервальной оценкой параметра G называется интервал, границы которого l_1^* и l_2^* являются функциями выборочных значений $x_1, x_2 \ldots . x_n$ и который с заданной вероятностью p накрывает оцениваемый параметр G.

$$P\{ l_1^* < G \le l_2^* \} = P$$

Интервал $(l_1^*, l_2^*]$ называется доверительным интервалом, а величина p – доверительной вероятностью. В качестве p наиболее часто используются значения: 0.9; 0.95 и 0.99.

Интервальные оценки параметров распределения

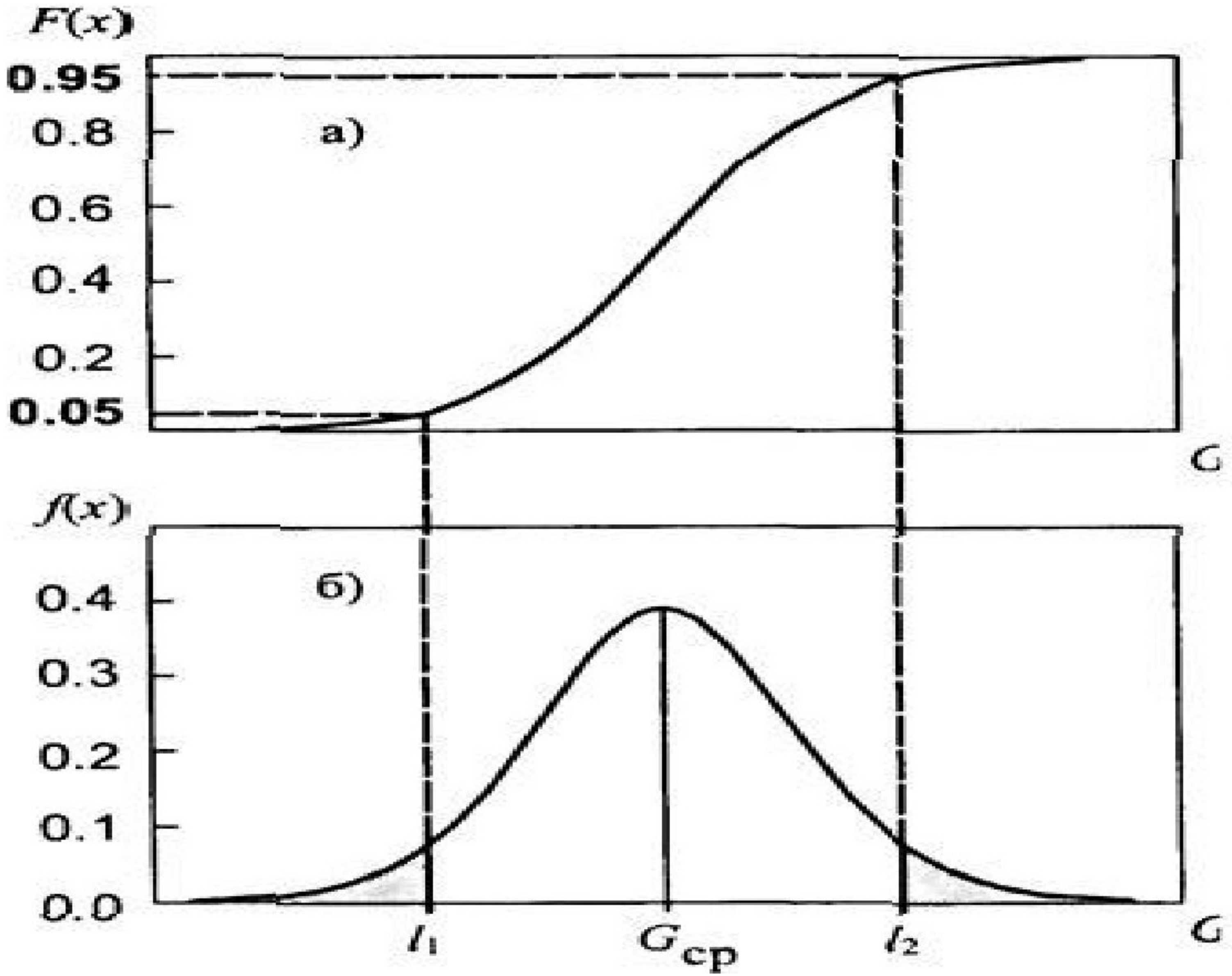
Используя функцию распределения выборочных значений параметра ***G***, можно записать вероятности не превышения для l_1^* и l_2^*

$$P\{ G^* \le l_1\} = F(l_1) = (1-p)/2,$$

$$P\{ G^* \le l_2\} = F(l_2) = p + (1-p)/2 = (1+p)/2$$

Например, если рассматривается 90%-ный доверительный интервал ($p = 0.9$), то $F(l_1) = 0.05$, $F(l_2) = 0.95$ или соответственно 5 и 95%. Как это показано на верхнем рисунке ниже. А на нижнем рисунке ниже показан тот же доверительный интервал на графике функции плотности вероятности. Не заштрихованная площадь на этом рисунке составляет 90% от общей площади графика.

Интервальные оценки параметров распределения



Интервальная оценка математического ожидания

На основании теоремы 2 выводится формула для интервальной оценки математического ожидания, а именно

$$t'_{(1-p)/2} \le [(x-m_x)/(S_x/\sqrt{n})] < t'_{(1+p)/2}$$

где $t'_{(1-p)/2}$ и $t'_{(1+p)/2}$ - квантили распределения Стьюдента, соответствующие вероятностям $(1-p)/2$ и $(1+p)/2$. Поскольку распределение Стьюдента симметрично относительно нуля, то $t_{(1-p)/2} = - t_{(1+p)/2}$.

Интервальная оценка математического ожидания

Следовательно

$$- t_{(1+p)/2} \le [(x-m_x)/(S_x/\sqrt{n})] < t_{(1+p)/2}$$

После преобразования получаем

$$x - \frac{t_{(1+p)/2} S_x}{\sqrt{n}} \le m_x < x + \frac{t_{(1+p)/2} S_x}{\sqrt{n}}$$

Интервальная оценка дисперсии

Исходя из теоремы 1 можно записать, что

$$\chi^2_{(1-p)/2} \leq \frac{(n-1)S_x^2}{\sigma_x^2} < \chi^2_{(1+p)/2}$$

где $\boldsymbol{S_x^{\,2} = D^*}$ - выборочная дисперсия; $\boldsymbol{\sigma_x^{\,2} = D}$ – фактическая дисперсия, $\boldsymbol{n}$ – длина ряда. После преобразований получим

$$\frac{(n-1)S_x^2}{\chi^2_{(1+p)/2}} < D \leq \frac{(n-1)S_x^2}{\chi^2_{(1-p)/2}}$$

Интервальная оценка дисперсии

Из этого выражения можно получить также интегральную оценку СКО.

$$S_x\sqrt{(n-1)/\chi^2_{(1+p)/2}} < \sigma \le S_x\sqrt{(n-1)/\chi^2_{(1-p)/2}}$$

СПАСИБО ЗА ВНИМАНИЕ!