Предмет и задачи математической статистики. Понятие выборки. Вариационный ряд.

В теории вероятностей рассматриваются случайные величины с заданным распределением или случайные эксперименты, свойства известны. Предмет теории вероятностей – свойства и взаимосвязи этих распределений.

Однако, часто эксперимент представляет собой черный ящик, он выдает только некие результаты, по которым нужно понять свойства эксперимента.

Чтобы пользоваться мат. статистикой, нужно:

- иметь случайный эксперимент, свойства которого частично или полностью неизвестны.
- эксперимент воспроизводим в одних и тех же условиях некоторое (а лучше - какое угодно) число раз.

Def. Выборка $\vec{X} = (X_1, \dots, X_n)$ объема n из распределения $\mathcal{F}$ — это набор из n независимых и одинаково распределенных величин, имеющих распределение $\mathcal{F}$

Def. Если упорядочить элементы выборки по возрастанию, то полученный набор случайных величин наз. вариационным рядом:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

Случайная величина $X_{(k)}$ наз. k-ой порядковой статистикой

Эмпирическая функция распределения. Гистограмма и полигон частот.

Def. Эмпирической функцией распределения называется

$$F_n^*(y) = \frac{\nu(y)}{n}$$

где $\nu(y)$ – число наблюдений X_i таких, что $X_i < y$

Гистограмма

$\forall i \in [1, n] \quad X_i \in [a, b]$

Разбиваем на k интервалов: $a = t_0 < \ldots < t_k = b$

На равные интервалы, шириной h $(t_j - t_{j-1} = h)$

$h = \frac{b-a}{k}$, ν_j – число наблюдений попавших в j-ый интервал

$$\nu_1 + \ldots + \nu_n = n$$

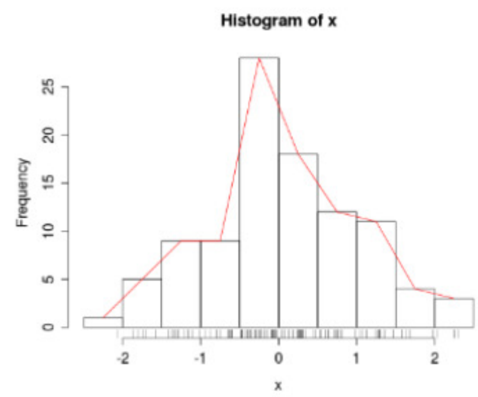
Высота ступеньки $L_j = \frac{\nu_j}{nh}$

Тогда сумма площадей всех ступенек: $S_1 + \ldots + S_k = \sum_j L_j h = \frac{\nu_1 + \ldots + \nu_k}{n} = 1$

Ф-ла Стерджесса для кол-ва интервалов $k = [\log_2 n] + 1$

Полигон частот

Соединим отрезками середины верха ступенек. Данная непрерывная кривая наз. полигоном частот.



Гистограмма и полигон частот

Пример:

П р и м е р 3. Имеется вариационный ряд из примера 2:

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

Разобьём отрезок [0, 10] на четыре равных отрезка. Отрезку [0, 2,5) принадлежат четыре элемента выборки, отрезку [2,5, 5) — шесть, отрезку [5, 7,5) — три, и отрезку [7,5, 10] — два элемента выборки. Строим гистограмму (рис. 2). На рис. 3 — гистограмма для той же выборки, но при разбиении области на пять равных отрезков.

$n = 15$, $h = \frac{10}{4} = 2{,}5$

$L_1 = \frac{4}{15 \cdot 2{,}5} = \frac{8}{75}$

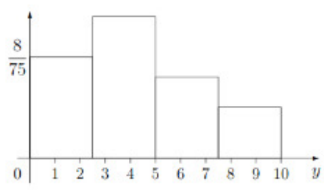


Рис. 2. Гистограмма при $k = 4$

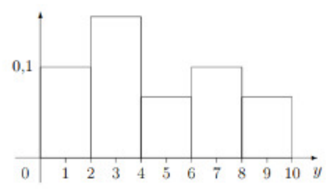


Рис. 3. Гистограмма при $k = 5$

Теорема Гливенко-Кантелли.

Прежде поговорим о свойствах Эмпирической Функции Распределения

Пусть $X_1, \dots, X_n$ - выборка из распределения $\mathcal{F}$ с функцией распределения F и пусть F_n^* - эмпирическая функция, построенная по этой выборке. Тогда

$$\forall y \in \mathbb{R} \quad F_n^*(y) \xrightarrow{p} F(y)$$

Доказательство: По определению, $F_n^*(y) = \frac{1}{n}\sum_{i=1}^{n} I(X_i < y)$

Случайные величины $I(X_i < y)$ независимы и одинаково распределены.
Чтобы применить ЗБЧ, осталось проверить что их мат. ожидание конечно.

$$\mathbb{E}\, I(X_1 < y) = 1 \cdot P(X_1 < y) + 0 \cdot P(X_1 \geq y) = P(X_1 < y) = F(y) < \infty$$

Тогда, по ЗБЧ: $F_n^*(y) = \dfrac{I(X_1<y) + \dots + I(X_n<y)}{n} \xrightarrow{p} \mathbb{E}\, I(X_1<y) = F(y)$

Следующая теорема говорит о том, что с ростом объема выборки наибольшее из расхождений стремится к нулю.

Теорема Гливенко - Кантелли

Пусть $X_1, \dots, X_n$ - выборка из распределения $\mathcal{F}$ с функцией распределения F и пусть F_n^* - эмпирическая функция, построенная по этой выборке. Тогда

$$\sup_{y \in \mathbb{R}} |F_n^*(y) - F(y)| \xrightarrow[n\to\infty]{p} 0$$

Без доказательства.

- $F_n^*(y)$ - несмещенная оценка для $F(y)$

$$\Delta \quad \mathbb{E} F_n^*(y) = \mathbb{E}\,\frac{\sum_{i=1}^{n} I(X_i<y)}{n} = \frac{n\,\mathbb{E}(I(X_1<y))}{n} = P(X_1<y) = F(y)$$

Свойства выборочных моментов.

Def. Выборочный момент порядка k — $\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

Свойства: 1) $\mathbb{E}\overline{X^k} = \mathbb{E}X_1^k$ (если момент сущ.)

Δ $\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right) = \frac{1}{n}\left(\mathbb{E}X_1^k + \dots + \mathbb{E}X_n^k\right) = \frac{n\,\mathbb{E}X_1^k}{n} = \mathbb{E}X_1^k$ $\blacktriangle$

2) $\overline{X^k} \xrightarrow{P} \mathbb{E}X_1^k$ (если момент сущ.)

Δ по ЗБЧ $\frac{X_1^k + \dots + X_n^k}{n} \xrightarrow{P} \mathbb{E}X_1^k$ $\blacktriangle$

3) $\mathbb{D}\overline{X} = \frac{\mathbb{D}X_1}{n}$

Δ $\mathbb{D}\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\sum \mathbb{D}X_i = \frac{1}{n^2}\, n\,\mathbb{D}X_1 = \frac{\mathbb{D}X_1}{n}$ $\blacktriangle$

Def. Смещенной выборочной дисперсией наз.

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

Def. Несмещенной выборочной дисперсией наз.

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

Свойства: 1) $S^2 = \overline{X^2} - (\overline{X})^2$

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

$= \overline{X^2} - 2(\overline{X})^2 + (\overline{X})^2 = \overline{X^2} - (\overline{X})^2$ $\blacktriangle$

2) $\mathbb{E} S^2 = \frac{n-1}{n} \mathbb{D} X_1$

Δ $\mathbb{E} S^2 = \mathbb{E}(\overline{X^2} - (\overline{X})^2) = \mathbb{E}\overline{X^2} - \mathbb{E}(\overline{X})^2 = \mathbb{E}\overline{X^2} - (\mathbb{D}\overline{X} + (\mathbb{E}\overline{X})^2) =$

$\mathbb{D}\overline{X} = \mathbb{E}(\overline{X})^2 - (\mathbb{E}\overline{X})^2$

$= \mathbb{E}X_1^2 - \frac{\mathbb{D}X_1}{n} - (\mathbb{E}\overline{X}_1)^2 = \mathbb{D}X_1 - \frac{\mathbb{D}X_1}{n} = \mathbb{D}X_1 \cdot \frac{n-1}{n}$ $\blacktriangle$

3) $\mathbb{E} S_0^2 = \mathbb{D} X_1$

4) $S^2 \xrightarrow{P} \mathbb{D} X_1$

Δ $S^2 = \overline{X^2} - (\overline{X})^2 \xrightarrow{P} \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \mathbb{D}X_1$ $\blacktriangle$

$g = t - s^2$ – непр. ф-я

5) $S_0^2 \xrightarrow{P} \mathbb{D} X_1$

Δ т.к. $\frac{n}{n-1} \xrightarrow[n\to\infty]{} 1$, то $S_0^2 = \frac{n}{n-1} S^2 \xrightarrow{P} \mathbb{D}X_1$

Задача оценивания неизвестных параметров. Несмещенность, состоятельность оценок.

Def Статистика — любая функция от выборки.

Пусть дана выборка из распределения с неизвестным параметром. Точно узнать его невозможно. Однако можно приблизить.

Def Оценкой неизвестного параметра θ наз. любая функция от выборки $\theta^* = g(X_1, \dots, X_n)$ в том или ином смысле приближающая θ.

Def Оценка θ^* наз. несмещенной, если $\mathbb{E}\theta^* = \theta$

Def Оценка θ^* наз. состоятельной, если $\theta^* \xrightarrow[n\to\infty]{P} \theta$

Метод моментов, примеры. Состоятельность оценок, полученных методом моментов.

Метод моментов

Он заключается в следующем:

- Любой момент случайной величины X_1 (пусть напр. k-ый) является функцией от θ (неизвестного параметра распределения)
- Тогда и θ может оказаться функцией от теоретического k-ого момента.
- Выразим θ и теоретический момент заменим выборочным.

Алгоритм: $\vec{X} \in F_\theta$, $\theta \in \mathbb{R}$

чаще всего $g(y) = y^k$

1. Выбираем пробную функцию $g(y)$ такую, чтобы момент $\mathbb{E}g(X_1)$ существовал и $m(\theta) = \mathbb{E}g(X_1)$ была обратима
2. Выражаем θ: $\theta = m^{-1}\left(\mathbb{E}g(X_1)\right)$
3. Заменяем истинный момент выборочным: $\theta^* = m^{-1}\left(\overline{g(X)}\right)$

θ^* – О.М.М. – оценка методом моментов.

Примеры:

1) $\vec{X} \in U_{[0,\theta]}$, $\theta > 0$

- $g(y) = y$, $\mathbb{E}g(X_1) = \mathbb{E}X_1 = \frac{\theta}{2}$
- $\theta = 2\mathbb{E}X_1$
- $\theta^* = 2\overline{X}$

Попробуем $g(y) = y^k$

- $\mathbb{E}g(X_1) = \mathbb{E}X_1^k = \int_{-\infty}^{+\infty} t^k \frac{1}{\theta} dt = \frac{1}{\theta}\int_0^{\theta} t^k dt = \frac{t^{k+1}}{\theta(k+1)}\Big|_0^{\theta} = \frac{\theta^k}{k+1}$
- $\theta = \sqrt[k]{(k+1)\cdot \mathbb{E}X_1^k}$
- $\theta^* = \sqrt[k]{(k+1)\cdot \overline{X^k}}$

2) $\vec{X} \in \Pi_\lambda$

- $g(y) = y$ $\quad \mathbb{E}g(X_1) = \mathbb{E}X_1 = \lambda$
- $\lambda^* = \overline{X}$

3) $\vec{X} \in E_\alpha$

- $g(y) = y$ $\quad \mathbb{E}X_1 = \frac{1}{\alpha}$
- $\alpha = \frac{1}{\mathbb{E}X_1}$ $\quad$ а) Сост: $\frac{1}{\overline{X}} = \frac{n}{X_1 + \ldots + X_n} \xrightarrow{P} \frac{1}{\mathbb{E}X_1} = \frac{1}{\alpha}$
- $\alpha^* = \frac{1}{\overline{X}}$ $\quad$ б) Несмещ: $\mathbb{E}\left(\frac{1}{\overline{X}}\right) = n\,\mathbb{E}\left(\frac{1}{X_1 + \ldots + X_n}\right) = n \cdot \frac{\alpha}{n-1}$

$X_i \in \Gamma_{\alpha,1} \Rightarrow \sum X_n \in \Gamma_{\alpha,n}$ $\quad \gamma_{\alpha,\lambda} = \frac{\alpha^\lambda t^{\lambda-1} \cdot e^{-\alpha t}}{\Gamma(\lambda)}$ $\quad t = \frac{z}{\alpha}$, $z = \alpha t$, $dz = \alpha dt$

$$\mathbb{E}\left(\frac{1}{x}\right) = \int_0^{+\infty} t^{-1} \frac{t^{n-1} \cdot \alpha^n e^{-\alpha t}}{\Gamma(n)} dt = \frac{\alpha^n}{\Gamma(n)} \int_0^{+\infty} t^{n-2} e^{-\alpha t} dt = \frac{\alpha^n}{\Gamma(n)} \cdot \frac{1}{\alpha^{n-1}} \int_0^{\infty} z^{n-2} e^{-z} dz$$

$\Gamma(n) = (n-1)\Gamma(n-1)$

$$= \frac{\alpha}{\Gamma(n)} \cdot \Gamma(n-1) = \frac{\alpha}{n-1}$$

Теорема о состоятельности ОММ.

Пусть $\mathbb{D} g(X_1) < \infty$, $m(t)$ – обратима и непрерывна. Тогда $\theta^* = m^{-1}(\overline{g(X)})$ состоятельна

Доказательство: $\overline{g(X)} \xrightarrow{P} \mathbb{E} g(X_1)$

$$\theta^* = m^{-1}(\overline{g(X)}) \xrightarrow{P} m^{-1}(\mathbb{E} g(X_1)) = \theta$$

Неравенство Йенсена

Если $g(u)$ - выпуклая вниз $(\smile)$, то $\mathbb{E} g(x) \geq g(\mathbb{E}X)$ если $\mathbb{E}X < \infty$

Причем равенство достигается $\iff \left[\begin{array}{l} g \text{ – линейная функция} \\ X \in I_c \end{array}\right.$

Пример использования:

10.16. Используя метод моментов, оценить параметр θ равномерного распределения на отрезке:
а) $[-\theta;\ \theta]$, $\theta > 0$; б) $[\theta;\ \theta+1]$.
Исследовать полученные оценки на несмещенность и состоятельность.

б) I $\mathbb{E}X_1 = \dfrac{(\theta+1)^2 - \theta^2}{2} = \dfrac{2\theta+1}{2}$

II. $\theta + \frac{1}{2} = \overline{X}$

III. $\theta^* = \overline{X} - \frac{1}{2}$

Сост: $\theta^* \xrightarrow{P} \mathbb{E}X_1 - \frac{1}{2} = \theta + \frac{1}{2} - \frac{1}{2} = \theta$ ✓

Несмещ: $\mathbb{E}\theta^* = \mathbb{E}\left(\overline{X} - \frac{1}{2}\right) \underset{\text{лин.}}{=} \mathbb{E}X_1 - \frac{1}{2} = \theta$ ✓

Метод максимального правдоподобия, примеры.

Def. Функцией правдоподобия для выборки $\vec{X}$ наз. функция

$$\Psi(\theta) = \Pi(\vec{X}, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

где $f(x_i, \theta) = \begin{cases} f_{X_i}(t) & \text{, абс. непр.} \\ P(X_i = t) & \text{, дискр.} \end{cases}$

Def. Оценкой максимального правдоподобия наз. значение $\theta(\vec{X})$ при котором ф-я правдоподобия принимает наибольшее значение

$$\hat{\theta} = \max_{\theta} \Pi(\theta, \vec{X}) = \max_{\theta} \Psi(\theta)$$

Def. Для упрощения поиска максимума используется логарифмическая ОМП

$$L(\vec{X}, \theta) = \ln \Pi(\vec{X}, \theta)$$

Пример: $\vec{X} \in B_p$, $p \in (0,1)$

$$f(t) = \begin{cases} p & , t = 1 \\ 1-p & , t = 0 \end{cases} = p^t (1-p)^{1-t}$$

$$\Psi(p) = \prod_{i=1}^{n} p^{x_i} \cdot (1-p)^{1-x_i} = p^{\sum x_i} \cdot (1-p)^{n - \sum x_i}$$

$$L(p) = \sum x_i \cdot \ln(p) + (n - \sum x_i)(1-p)$$

$$L'(p) = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = \sum x_i - p \cdot \sum x_i - np + p \sum x_i = 0$$

$$\sum x_i = np$$

$$p = \frac{\sum x_i}{n} = \overline{X}$$

Сравнение оценок. Понятие эффективной оценки.

Def: Функцией среднеквадратического отклонения оценки θ^* наз.

$$\delta_{\theta^*}(\theta) = E(\theta^* - \theta)^2$$

если $\mathbb{D}\theta^* < \infty$

Def: Говорят, что θ^* не хуже, чем θ^{**} (в среднеквадратическом смысле)

если $\forall \theta \quad \delta_{\theta^*}(\theta) \leq \delta_{\theta^{**}}(\theta)$

- Для несмещенных оценок $\delta_{\theta^*}(\theta) = \mathbb{D}\theta^*$

Доказательство: $\mathbb{E}(\theta^* - \theta)^2 = \mathbb{E}(\theta^* - \theta + \mathbb{E}\theta^* - \mathbb{E}\theta^*)^2 = \mathbb{E}((\theta^* - \mathbb{E}\theta^*) + (\mathbb{E}\theta^* - \theta))^2 =$

$$= \underbrace{\mathbb{E}(\theta^* - \mathbb{E}\theta)^2}_{\mathbb{D}\theta^*} + 2\underbrace{\mathbb{E}[(\theta^* - \mathbb{E}\theta^*)(\mathbb{E}\theta^* - \theta)]}_{=0} + \underbrace{\mathbb{E}(\mathbb{E}\theta^* - \theta)^2}_{b(\theta)\text{ - смещение}} = \mathbb{D}\theta^* + \underset{0}{\underset{\|}{b_{\theta^*}(\theta)}} = \mathbb{D}\theta^* \ \blacktriangle$$

т.к. $\mathbb{E}(\theta^* - \mathbb{E}\theta^*) = \mathbb{E}\theta^* - \mathbb{E}\theta^* = 0$

т.к. оценка несмещ. и $\mathbb{E}\theta^* = \theta$

Пример: $\vec{X} \in U_{[0,\theta]}$ $\quad \theta^* = 2\overline{X}$, $\quad \hat{\theta} = X_{(n)}$ $\qquad \mathbb{D}X_1 = \frac{b^2 - a^2}{12} = \frac{\theta^2}{12}$

- $\delta_{2\overline{X}}(\theta) = \mathbb{D}(2\overline{X}) \overset{\text{несмещ.}}{=} 4\mathbb{D}\overline{X} = 4\frac{\mathbb{D}X_1}{n} = \frac{4\theta^2}{12n} = \frac{4\theta^2}{3n}$

Надо найти $\mathbb{E}$ и $\mathbb{E}X_{(n)}^2$ для $X_{(n)}$:

$$F_{X_{(n)}}(y) = P(X_{(n)} < y) = P(X_1 < y, X_2 < y, \dots, X_n < y) = P^n(X < y) = \begin{cases} 0, & x < 0 \\ \frac{y^n}{\theta^n}, & x \in [0, \theta] \\ 1, & x > \theta \end{cases}$$

$$f_{X_{(n)}}(y) = \begin{cases} 0, & x \notin [0, \theta] \\ \frac{n y^{n-1}}{\theta^n}, & x \in [0, \theta] \end{cases}$$

$$\mathbb{E}X_{(n)} = \int_0^\theta y \cdot \frac{n \cdot y^{n-1}}{\theta^n} dy = \frac{n}{\theta^n} \frac{y^{n+1}}{n+1}\Big|_0^\theta = \frac{n\theta}{n+1}$$

$$\mathbb{E}X_{(n)}^2 = \int_0^\theta y^2 \frac{n \cdot y^{n-1}}{\theta^n}\,dy = \frac{n}{\theta^n}\frac{y^{n+2}}{n+2}\Big|_0^\theta = \frac{n\theta^2}{n+2}$$

$-n^2\theta^2 \qquad -n\theta^2$

$$n^2\theta + n\theta - 2n^2\theta^2 - 4n\theta^2 + \theta^2 n^2 + 3n\theta^2 + 2\theta^2 = 2\theta^2$$

$$\mathbb{E}(X_{(n)}-\theta)^2 = \mathbb{E}X_{(n)}^2 - 2\theta\,\mathbb{E}X_{(n)} + \theta^2 = \frac{n\theta}{n+2} - \frac{2n\theta^2}{n+1} + \theta^2 = \frac{n\theta(n+1) - 2n\theta^2(n+2) + \theta^2\overbrace{(n+2)(n+1)}^{n^2+3n+2}}{(n+1)(n+2)} = \frac{2\theta^2}{(n+1)(n+2)}$$

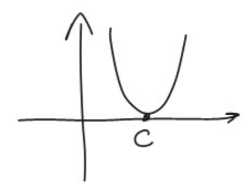
$$\delta_{2\bar{X}}(\theta) = \frac{4\theta^2}{3n} \qquad \delta_{X_{(n)}}(\theta) = \frac{2\theta^2}{(n+1)(n+2)}$$

При $n=1,2$ они равны, при $n \geq 3$ $X_{(n)}$ лучше, т.к. $\frac{4\theta^2}{3n} > \frac{2\theta^2}{(n+1)(n+2)}$

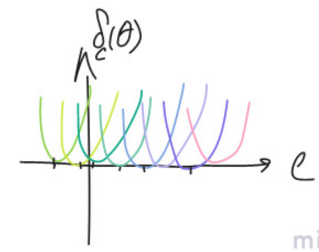
Теорема Пусть $\vec{X} \in F_\theta$. Среди всех оценок с конечной дисперсией наилучшей в среднеквадратическом смысле не существует.

Доказательство: Пусть $\theta^* \equiv C$ $(C \in \mathbb{R})$

$\delta_{\theta^*}(\theta) = \mathbb{E}(C-\theta)^2 = (C-\theta)^2$ – парабола



Всевозможные вырожденные оценки будут пересекаться (т.к. C пробегает всё $\mathbb{R}$)



Чтобы найти наилучшую исходя из геом. смысла, нужно найти огибающую всех этих парабол. Но они покрывают все $\mathbb{R}$, поэтому огибающая одна – $\delta_{\theta^*_{opt}} \equiv 0$

$\delta_{\theta^*_{opt}}(\theta) \equiv 0 \Rightarrow \mathbb{E}(\theta^*_{opt} - \theta) \equiv 0$

$\theta^*_{opt} \equiv \theta$ – вырожденная оценка, в точности угадывает неизвестный параметр.

Def Оценка наз. эффективной, если она наилучшая среди всех несмещенных оценок в среднеквадратическом смысле.

Распределения, связанные с нормальным (хи-квадрат, Стьюдента, Фишера).

Гамма-распределение

Вспомним свойство устойчивости по суммированию:

- Пусть $\xi_1, \dots, \xi_n$ – независ. и $\xi_i \in \Gamma_{\alpha, \lambda_i}$ $\forall i \in [1, n]$ тогда $\xi_1 + \dots + \xi_n \in \Gamma_{\alpha, \lambda_1 + \dots + \lambda_n}$

И докажем еще одно свойство $\Gamma_{\alpha, \lambda}$:

- Если $\xi \in N_{0,1}$ то $\xi^2 \in \Gamma_{\frac{1}{2}, \frac{1}{2}}$

Доказательство:

При $y < 0$: $F_{\xi^2}(y) = P(\xi^2 < y) = 0 \Rightarrow f_{\xi^2}(y) = 0$

При $y > 0$: $F_{\xi^2}(y) = P(\xi^2 < y) = P(-\sqrt{y} < \xi < \sqrt{y}) = N_{0,1}(\sqrt{y}) - N_{0,1}(-\sqrt{y})$

$$f_{\xi^2}(y) = \frac{dF_{\xi^2}(y)}{dy} = (N_{0,1}(\sqrt{y}))' \cdot (\sqrt{y})' - (N_{0,1}(-\sqrt{y}))' \cdot (-\sqrt{y})' =$$

из симметр плотности $N_{0,1}$

$$\gamma_{0,1}(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + \gamma_{0,1}(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{2\sqrt{y}} \left(\gamma_{0,1}(\sqrt{y}) + \gamma_{0,1}(\sqrt{y}) \right)$$

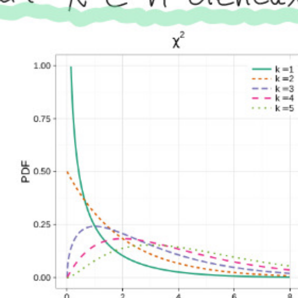
$$= \frac{\gamma_{0,1}(\sqrt{y})}{\sqrt{y}} = \frac{1}{\sqrt{2\pi} \cdot \sqrt{y}} \cdot e^{-\frac{y}{2}} = \frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}}{\underset{\sqrt{\pi}}{\underset{\shortparallel}{\Gamma\left(\frac{1}{2}\right)}}} y^{\frac{1}{2} - 1} \cdot e^{-\frac{1}{2} \cdot y} \in \Gamma_{\frac{1}{2}, \frac{1}{2}}$$

Распределение χ^2 Пирсона

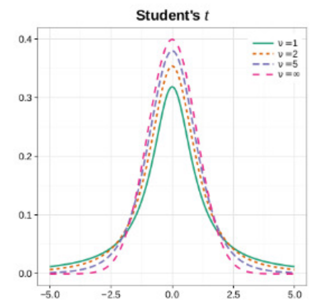
Из предыдущих двух свойств следует, что если $\xi_1, \dots, \xi_n$ независимые и $\xi_i \in N_{0,1}$ $\forall i \in [1, n]$ то $\chi^2 = \xi_1^2 + \dots + \xi_n^2 \in \Gamma_{\frac{1}{2}, \frac{n}{2}}$. Это называется распределением χ^2 с n степенями свободы

Обозначение: $\chi^2 \in \chi^2_n$

- Свойство: Если $Z_1 \in \chi^2_n$, $Z_2 \in \chi^2_m$ $\Rightarrow Z_1 + Z_2 \in \chi^2_{n+m}$



Распределение Стьюдента



Def Пусть $\xi_0, \xi_1, \dots, \xi_k$ – независимы и $\xi_i \in N_{0,1}$ $\forall i \in [0,n]$

Распределение случайной величины $t_k = \dfrac{\xi_0}{\sqrt{\dfrac{\xi_1^2 + \dots + \xi_k^2}{k}}}$ наз распределением Стьюдента с k степенями свободы и обозначается T_k

Свойство: $T_n \underset{n\to\infty}{\Rightarrow} N_{0,1}$. Доказательство: • $\dfrac{\chi_n^2}{n} \xrightarrow{P} 1$ т.к по ЗБЧ $\dfrac{\xi_1^2 + \dots + \xi_n^2}{n} \to \mathbb{E}\xi_1^2 = 1$ $\left(\xi_1 \in \Gamma_{\frac{1}{2},\frac{1}{2}},\ \mathbb{E}\xi_1 = \frac{1}{2}\cdot\frac{2}{1} = 1\right)$

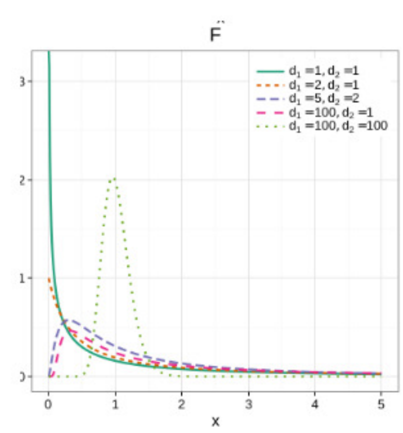
• $t_n = \dfrac{\xi}{\sqrt{\dfrac{\chi_n^2}{n}}} \Rightarrow \xi \in N_{0,1}$

Распределение Фишера

Def Пусть $\left.\begin{matrix} Z_1 \in \chi_n^2 \\ Z_2 \in \chi_m \end{matrix}\right\}$ независимы.

Распределение случайной величины $f_{n,m} = \dfrac{\frac{Z_1}{n}}{\frac{Z_2}{m}} = \dfrac{m}{n}\dfrac{Z_1}{Z_2}$ имеет распределение Фишера с n и m степенями свободы. ($F_{n,m}$)

Свойство: $F_{n,m} \underset{n,m\to\infty}{\Rightarrow} I_1$



Очевидно, т.к. $\dfrac{Z_1}{n} = \dfrac{\xi_1^2 + \dots + \xi_n^2}{n} \xrightarrow{P} \mathbb{E}\xi_1^2 = 1$ и $\dfrac{Z_2}{m} = \dfrac{\eta_1^2 + \dots + \eta_m^2}{m} \to \mathbb{E}\eta_1^2 = 1$

Лемма Фишера.

Лемма Фишера

Пусть случайные величины $\vec{X}=(X_1,..,X_n)$, $\forall i\in[1,n]$ $X_i\in N_{0,1}$, $\vec{Y}=A\vec{X}$

A – ортогональная матрица. Тогда $\forall r\in[1,n-1]$

$$Q = X_1^2+X_2^2+\ldots+X_n^2 - Y_1^2-\ldots-Y_r^2 \in \chi^2_{n-r}$$

И Q не зависит от $Y_1,..,Y_r$.

Доказательство:

- Покажем, что $\forall i\in[1,n]$ $Y_i\in N_{0,1}$.

$\vec{Y}$ – нормальный вектор по опр. (т.к. преобразование линейное) у $N_{0,1}$ она единичная

Рассмотрим ковариационную матрицу: $C(\vec{Y}) = C(A\vec{X}) = A\,C(\vec{X})A^T = A\cdot A^T = E$

по свойству $C(\vec{X})$ т.к A - ортогональная

т.е $\vec{Y}\in N(0,\vec{E})$

- Так как A - ортогональная, а умножение на орт. матрицу не меняет норму вектора, то

$$\|\vec{X}\| = \|\vec{Y}\| \Rightarrow \|\vec{X}\|^2 = \|\vec{Y}\|^2 \Rightarrow \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i^2$$

т.к все Y_i независимы, то эта сумма не зависит от $Y_1,...,Y_r$

Тогда верно $X_1^2+\ldots+X_n^2 - Y_1^2-\ldots-Y_r^2 = \underbrace{Y_{r+1}^2+\ldots+Y_n^2}_{\text{по опр } \in \chi^2_{n-r}}$

Теорема о свойствах выборочного среднего и выборочной дисперсии для выборок из нормальной совокупности

Теорема о свойствах нормальных выборок

Пусть $\vec{X} \in N_{a,\sigma^2}$, тогда

1) $\frac{\bar{X} - a}{\sigma}\sqrt{n} \in N_{0,1}$ 2) $\frac{nS^2}{\sigma^2} \in \chi^2_{n-1}$ 3) $\sum_{i=1}^{n}\left(\frac{X_i - a}{\sigma}\right)^2 = \frac{nS_1^2}{\sigma^2} \in \chi^2_n$

4) $\bar{X}$ и S - независимые 5) $\sqrt{n}\left(\frac{\bar{X} - a}{S_0}\right) \in T_{n-1}$

Доказательство:

1) $\vec{X} \in N_{a,\sigma^2}$ $\quad \bar{X} = \frac{\overbrace{X_1 + \dots + X_n}^{\in N_{n\cdot a,\, n\sigma^2}}}{n} = \underbrace{(\sqrt{n}\sigma)\cdot\frac{1}{n}}_{A}\underbrace{\left(\frac{\sum X_i - na}{\sqrt{n}\sigma}\right)}_{\in N_{0,1}} + \underbrace{\frac{1}{n}\cdot na}_{B} \in N_{a,\frac{\sigma^2}{n}}$

Следствие 1. *Если* $X \in \Phi_{\alpha,\sigma^2}$, *то* $Y = (X - \alpha)/\sigma \in \Phi_{0,1}$.
Следствие 2. *Если* $Y \in \Phi_{0,1}$, *то* $X = \sigma Y + \alpha \in \Phi_{\alpha,\sigma^2}$.
Следствие 3. *Если* $X \in \Phi_{\alpha,\sigma^2}$, *то* $Y = AX + B \in \Phi_{A\alpha+B,\sigma^2A^2}$.
Доказательство. Утверждения первых двух следствий прямо вытекают из формулы (2). Для доказательства третьего утверждения удобно сначала представить

$$AX + B = \sigma A\left(\frac{X-\alpha}{\sigma}\right) + A\alpha + B,$$

$\bar{X} \in N_{a,\frac{\sigma^2}{n}}$

$\left(\frac{\bar{X} - a}{\sigma}\right)\sqrt{n} \in N_{0,1}$

2) $nS_1^2 = \sum_{i=1}^{n}(X_i - a)^2$

$\frac{nS_1^2}{\sigma^2} = \sum_{i=1}^{n}\Bigg(\underbrace{\frac{X_i - a}{\sigma}}_{\in N_{0,1} \text{ т.к } X_i \in N_{a,\sigma^2}}\Bigg)^2 \in \chi^2_n$

3) $\frac{nS^2}{\sigma^2} = \frac{n}{n}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X} + a - a}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i - a}{\sigma} - \frac{\bar{X} - a}{\sigma}\right)^2 =$

Обозначим $z_i = \frac{X_i - a}{\sigma}$, тогда $\bar{z} = \frac{X_1 - a + X_2 - a + \dots + X_n - a}{n\sigma} = \frac{X_1 + \dots + X_n - na}{n\sigma} = \frac{\overbrace{\frac{X_1 + \dots + X_n}{n}}^{\bar{X}} - a}{\sigma}$

Тогда $= \sum_{i=1}^{n}(z_i - \bar{z})^2 = \sum_{i=1}^{n}(z_i^2 - 2z_i\bar{z} + \bar{z}^2) = \sum_{i=1}^{n} z_i^2 - 2n\bar{z}^2 + n\bar{z}^2 = \sum_{i=1}^{n} z_i^2 - n\bar{z}^2 =$

$$= \sum_{i=1}^{n} z_i - (\sqrt{n}\,\overline{z})^2$$

$$\sqrt{n}\,z = \begin{pmatrix} \frac{1}{\sqrt{n}} \dots & \frac{1}{\sqrt{n}} \end{pmatrix} \begin{pmatrix} z_1 \\ z_1 \\ \vdots \\ z_n \end{pmatrix}$$

$\left\| \frac{1}{\sqrt{n}} \dots \frac{1}{\sqrt{n}} \right\| = 1$, этот вектор можно достроить до ортогональной матрицы, в которой он будет первой строкой

$$\begin{pmatrix} \frac{1}{\sqrt{n}} & \dots & \frac{1}{\sqrt{n}} \\ \vdots & \ddots & \vdots \\ \vdots & \dots & \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} \sqrt{n}\,\overline{z} \\ \vdots \\ \vdots \end{pmatrix}$$

Применим Лемму Фишера:

$$\frac{nS^2}{\sigma^2} = z_1^2 + \dots + z_n^2 - (n\sqrt{z})^2 \in \chi^2_{n-1} \quad (*)$$

miro

4) По Л. Фишера, в $(*)$ $\frac{nS^2}{\sigma^2}$ не зависит от $(\sqrt{n}\,\overline{z})$ но тогда S и $\overline{z} = \frac{\overline{X} - a}{\sigma}$ не зав.

5) $T_{n-1} = \dfrac{\xi_0}{\sqrt{\frac{z_{n-1}}{n-1}}}$ $\quad \left.\begin{matrix} \xi_0 \in N_{0,1} \\ z_{n-1} \in \chi^2_{n-1} \end{matrix}\right\}$ незав. $\quad \sqrt{n}\left(\frac{\overline{X} - a}{S_0}\right) \in T_{n-1}$

$$\left.\begin{matrix} \xi_0 = \left(\frac{\overline{X} - a}{\sigma}\right)\sqrt{n} \in N_{0,1} \\ z_{n-1} = \frac{nS^2}{\sigma^2} \in \chi^2_{n-1} \end{matrix}\right\} \quad \eta = \frac{\left(\frac{\overline{X} - a}{\sigma}\right)\sqrt{n}}{\sqrt{\frac{nS^2}{\sigma^2 (n-1)}}} = \frac{\overline{X} - a}{S_0}\sqrt{n} \in T_{n-1}$$

S_0^2

miro

Построение доверительных интервалов для среднего нормальной совокупности.

Интервальные оценки

Идея: построить 2 ф-ции от неизвестного параметра чтобы они оценивали его сверху и снизу. (по возможности задавали маленький интервал). Будем называть это доверительным интервалом.

Def Интервал $(\theta^-(\vec{X}), \theta^+(\vec{X}))$ такой, что

$$P(\theta^-(\vec{X}) < \theta < \theta^+(\vec{X})) \geqslant 1-\varepsilon$$

наз. доверительным интервалом уровня $1-\varepsilon$

Def Доверительный интервал наз. точным доверительным интервалом уровня $1-\varepsilon$

$$P(\theta^-(\vec{X}) < \theta < \theta^+(\vec{X})) = 1-\varepsilon$$

Def Асимптотический доверительный интервал уровня $1-\varepsilon$:

$$\lim_{n\to\infty} P(\theta^-(\vec{X}) < \theta < \theta^+(\vec{X})) \geqslant 1-\varepsilon \quad (n - \text{объем выборки})$$

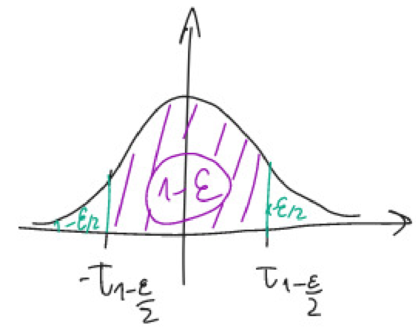
Def Асимптотически точный доверительный интервал уровня $1-\varepsilon$:

$$\lim_{n\to\infty} P(\theta^-(\vec{X}) < \theta < \theta^+(\vec{X})) = 1-\varepsilon$$

Доверительные интервалы для среднего нормальной совокупности

$\vec{X} \in N_{a,\sigma^2}$

1) σ известно. (I) $G = \frac{\bar{X} - a}{\sigma}\sqrt{n} \in N_{0,1}$



(II) $P\left(-t < \frac{\bar{X}-a}{\sigma}\sqrt{n} < t\right) = 1-\varepsilon$

$P\left(-\frac{t\sigma}{\sqrt{n}} < \bar{X} - a < \frac{t\sigma}{\sqrt{n}}\right) = 1-\varepsilon$

$P\left(-\frac{t\sigma}{\sqrt{n}} - \bar{X} < -a < \frac{t\sigma}{\sqrt{n}} - \bar{X}\right) = 1-\varepsilon$

$P\left(\bar{X} - \frac{t\sigma}{\sqrt{n}} < a < \bar{X} + \frac{t\sigma}{\sqrt{n}}\right) = 1-\varepsilon$

$P\left(\bar{X} - \tau_{1-\frac{\varepsilon}{2}}\frac{\sigma}{n} < a < \bar{X} + \tau_{1-\frac{\varepsilon}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\varepsilon$

2) σ неизвестно. (I) $G(\vec{X}, a) = \frac{\bar{X}-a}{S_0}\sqrt{n} \in T_{n-1}$

(II) $P\left(-t < \frac{\bar{X}-a}{S_0}\sqrt{n} < t\right) = 1-\varepsilon$

$P\left(\bar{X} - t_{1-\frac{\varepsilon}{2}}\frac{S_0}{\sqrt{n}} < a < \bar{X} + t_{1-\frac{\varepsilon}{2}}\frac{S_0}{\sqrt{n}}\right) = 1-\varepsilon$

Построение доверительных интервалов дисперсии нормальной совокупности.

Доверительные интервалы для дисперсии нормальной совокупности

$\vec{X} \in N_{a,\sigma^2}$

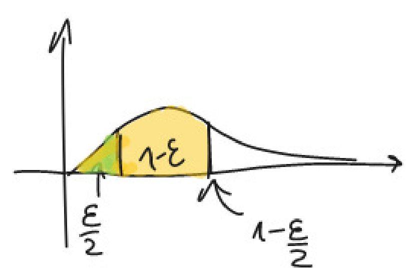
1) a известно. $G(\sigma, \vec{X}) = \frac{nS_1^2}{\sigma^2} \in \chi^2_n$ $\qquad nS_1^2 = \sum_{i=1}^{n} (X_i - a)^2$

$$P\left(q_1 < \frac{\sum_i (X_i - a)^2}{\sigma^2} < q_2 \right) = 1 - \varepsilon$$

$$P\left(\frac{q_1}{\sum (X_i - a)^2} < \frac{1}{\sigma^2} < \frac{q_2}{\sum (X_i - a)^2} \right) = 1 - \varepsilon$$

$$P\left(\frac{\sum (X_i - a)^2}{q_2} < \sigma^2 < \frac{\sum (X_i - a)^2}{q_1} \right) = 1 - \varepsilon$$

$$\begin{cases} q_1 = \chi_n^{-1}\left(\frac{\varepsilon}{2}\right) \\ q_2 = \chi_n^{-1}\left(1 - \frac{\varepsilon}{2}\right) \end{cases}$$



2) σ неизвестно. $G(\sigma, \vec{X}) = \frac{nS^2}{\sigma^2} \in \chi^2_{n-1}$ $\qquad nS^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2$

$$P\left(q_1 < \frac{\sum (X_i - \overline{X})^2}{\sigma^2} < q_2 \right) = 1 - \varepsilon$$

$$P\left(\frac{\sum (X_i - \overline{X})^2}{q_2} < \sigma^2 < \frac{\sum (X_i - \overline{X})^2}{q_1} \right) = 1 - \varepsilon$$

$$\begin{cases} q_1 = \chi_{n-1}^{-1}\left(\frac{\varepsilon}{2}\right) \\ q_2 = \chi_{n-1}^{-1}\left(1 - \frac{\varepsilon}{2}\right) \end{cases}$$

Построение доверительных интервалов с помощью нормального приближения

Построение Д.И. с помощью нормального приближения

$\vec{X} \in F_\theta$, θ - неизв

1) Строим $G(\theta, \vec{x}) \Rightarrow \eta \in N_{0,1}$

2) $\lim\limits_{n\to\infty} P(t_1 < G(\theta, \vec{x}) < t_2) = 1-\varepsilon$

3) Решить неравенство относительно θ

Пример $\vec{X} \in B_p$, p - неизв.

Знаем, что $$\frac{\sqrt{n}\,(\bar{X}-p)}{\sqrt{p(1-p)}} \underset{\substack{\uparrow \\ \text{Ц.П.Т}}}{\Rightarrow} \eta \in N_{0,1}$$

$$P\left(-\tau_{1-\frac{\varepsilon}{2}} \le \frac{\sqrt{n}\,(\bar{X}-p)}{\sqrt{p(1-p)}} \le \tau_{1-\frac{\varepsilon}{2}}\right) \to 1-\varepsilon$$

$$\parallel$$

$$\frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{\bar{X}(1-\bar{X})}} \cdot \frac{\sqrt{n}\,(\bar{X}-p)}{\sqrt{p(1-p)}} \qquad \xrightarrow{p} 1 \text{ (по ЗБЧ)}$$

$$P\left(-\tau_{1-\frac{\varepsilon}{2}} \le \frac{\sqrt{n}\,(\bar{X}-p)}{\sqrt{\bar{X}(1-\bar{X})}} \le \tau_{1-\frac{\varepsilon}{2}}\right) \to 1-\varepsilon$$

$$P\left(\bar{X} - \tau_{1-\frac{\varepsilon}{2}} \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}} \le p \le \bar{X} + \tau_{1-\frac{\varepsilon}{2}} \frac{\sqrt{\bar{X}(1-\bar{X})}}{\sqrt{n}}\right) \to 1-\varepsilon$$

Критерий Колмогорова

Def Статистическим критерием наз $\delta(\vec{X}) = \begin{cases} 0, & \vec{X} \notin \mathbb{R}^n \setminus K \\ 1, & \vec{X} \in K \end{cases}$, K - критическая область критерия

- $\delta(\vec{X}) = 0$ – принимаем H_0, отвергаем H_a
- $\delta(\vec{X}) = 1$ – принимаем H_a, отвергаем H_0

	Верна H_0	Верна H_a
$\delta = 0$	хорошо	ошибка 2го рода
$\delta = 1$	ошибка 1го рода	хорошо

Def $\alpha_1(\delta) = P_{H_0}(\delta = 1)$ – вероятность ошибки 1го рода
размер критерия

$\alpha_2(\delta) = P_{H_a}(\delta = 0)$ – вер-ть ошибки 2го рода

$\beta(\delta) = 1 - \alpha_2(\delta)$ – мощность критерия

Def Критерий δ состоятелен если $\beta(\delta) \xrightarrow[n\to\infty]{} 1$

Есть так называемые критерии согласия. В общем виде они выглядят так:

$\vec{X} \in F \qquad H_0 = \{F = F_0\}$
$\qquad\qquad H_a = \{F \neq F_0\}$

Нужно придумать функцию, которая бы представляла собой меру близости эмпирической и предполагаемой функций распределения.

Назовем её $d(F_0, F_n^*)$. Она должна удовлетворять условиям:

к1) при верной H_0 $\quad d(F_0, F_n^*) \underset{n\to\infty}{\Rightarrow} \eta \in G$ (полностью известное абсолютно непр. распределение)

к2) при верной H_a $\quad |d(F_0, F_n^*)| \xrightarrow[n\to\infty]{P} \infty$

Тогда, критерием согласия называют $\delta = \begin{cases} 0, & d(F_0, F_n^*) < C \\ 1, & d(F_0, F_n^*) \geq C \end{cases}$, где $G(C) = 1 - \varepsilon$

т.е. C – квантиль уровня $1-\varepsilon$ распр. G

Теорема Колмогорова

Пусть $X \in F$, F - непрер. $\forall t>0$ $\quad d_K = \sqrt{n}\, \sup_t |F_n^*(t) - F_0(t)| \Rightarrow \eta \in K$ (при верной H_0)

$$\delta = \begin{cases} 0, & \sqrt{n}\, \sup |F_n^*(t) - F_0(t)| < C \\ 1, & \text{иначе} \end{cases} \qquad \begin{matrix} H_0 = \{F = F_0\} \\ H_a = \{F \neq F_0\} \end{matrix}$$

Этот критерий состоятелен:

По Т. Гливенко-Кантелли

$$\alpha_2(\delta) = P_{H_a}(d_K < C) = P_{H_a}\Big(\sqrt{n}\, \sup_t |\overset{\nearrow F(t)}{F_n^*(t)} - F_0(t)| < C\Big) = [\text{верна } H_a \Rightarrow F_0(t) \neq F(t)] \Rightarrow$$

$$\Rightarrow \exists \tilde{t} \quad |F(\tilde{t}) - F_0(\tilde{t})| > 0 \Rightarrow \sup_t |F(t) - F_0(t)| \geq |F(\tilde{t}) - F_0(\tilde{t})| = \sup > 0 \text{, тогда}$$

$$P_{H_a}\Big(\underbrace{\sqrt{n}\, \underbrace{\sup|-\text{"}-|}_{>0}}_{\xrightarrow[n\to\infty]{} \infty} < C\Big) \to 0$$

Def Реально достигнутый уровень значимости - $\varepsilon^* = P_{H_0}\big(d(F_n^*, F_0) \geq \overset{\eta \in G}{\tilde{d}}\big)$

числовое значение которое мы сравнивали с C.

Эта вероятность имеет следующий смысл: Это вероятность, взяв выборку из распределения F_0, получить по ней большее отклонение, чем по проверяемой выборке.
Большие значения ε^* свидетельствуют в пользу H_0,
Малые, напротив, в пользу H_a. Почему?

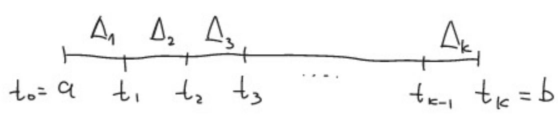
Если, например $\varepsilon^* = 0{,}2$, то в среднем 20 % контрольных выборок (у которых $F = F_0$) будут иметь большее отклонение чем проверяемая. Отсюда можно сделать вывод что тестируемая выборка ведет себя не хуже чем 20% "правильных" выборок

- $\varepsilon^* \leq 0{,}05 \Rightarrow H_a$
- $\varepsilon^* \geq 0{,}1 \Rightarrow H_0$

Критерий хи-квадрат.

Критерий χ^2 Пирсона

Основывается на группированных данных. Область значений F_0 делят на некоторое число интервалов, затем строят функцию отклонения Ψ по разностям теоретических вероятностей попадания в интервалы и эмпирических частот.



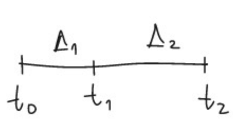
$$\nu_i = \sum_{i=1}^{n} I(x \in \Delta_i) \qquad p_i = P(x \in \Delta_i) = F_0(t_i) - F_0(t_{i-1})$$

$$\nu_1 + \ldots + \nu_n = n \qquad p_1 + \ldots + p_n = 1$$

$$d(F_n^*, F_0) = (\chi^2)^* = \sum_{i=1}^{n} \frac{(\nu_i - np_i)^2}{np_i} \Rightarrow \eta \in \chi^2_{k-1} \text{ (при верной } H_0)$$

$$\delta = \begin{cases} 0, & (\chi^2)^* < c \\ 1, & (\chi^2)^* \geq c \end{cases}, \text{ где } \chi^2_{k-1}(c) = 1 - \varepsilon$$

Докажем теорему Пирсона для $k = 2$



$$\Psi = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(\nu_2 - np_2)^2}{np_2} = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(\overbrace{(n-\nu_1) - n(1-p_1)}^{n-\nu_1-n+np_1})^2}{n(1-p_1)} =$$

$\nu_1 + \nu_2 = n$

$p_1 + p_2 = 1$

$$= \frac{(\nu_1 - np_1)^2(1 - \cancel{p_1}) + \cancel{(np_1 - \nu_1)^2} \cdot p_1}{np_1(1-p_1)} = \left(\frac{\nu_1 - \overset{\text{EX Бернулли}}{np_1}}{\sqrt{\underbrace{np(1-p)}_{\text{DX Бернулли}}}}\right)^2 \Rightarrow \xi^2, \quad \xi \in N_{0,1}$$

Сх. Бернулли: попадаем в Δ_1 – успех (p_1)

в Δ_2 – неудача ($1-p_1$) $\qquad \nu_1$ – число успехов

Состоятельность χ^2

$H_0 = \{\vec{X} \in F_0\}$

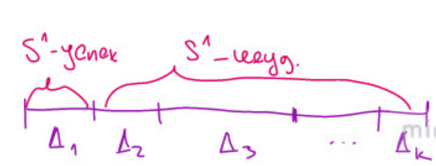
$H_a = \{\vec{X} \notin F_0\}$

– Для этих гипотез не всегда будет выполнятся K_2.
Если распределение $\vec{X}$ $\mathcal{F}_1 \neq \mathcal{F}_0$ имеет такие же как у $\mathcal{F}_0$ вероятности p_i попадания в каждый из Δ_i, то по данной ф-ции $(\chi^2)^*$ эти распределения различить невозможно.

Поэтому введем более корректные гипотезы:

$$H_0' = \{\mathcal{F} : \forall i \quad P(X_1 \in \Delta_i) = p_i\}$$

$$H_a' = \{\mathcal{F} : \exists i : \quad P(X_1 \in \Delta_i) \neq p_i\}$$



Введём случайную величину $S^i(x) = \begin{cases} 1, & x \in \Delta_i \\ 0, & x \notin \Delta_i \end{cases}$ – Сх. Бернулли $\begin{pmatrix} \text{успех} - p_i \\ \text{неуд.} - (1-p_i) \end{pmatrix}$

Тогда по ЗБЧ, $\dfrac{S_1^i + \ldots + S_n^i}{n} \xrightarrow{P} \mathbb{E} S_1^i = p_i$

$$\alpha_2 = P_{H_a}\left(\sum_{j=1}^{k} \frac{(\nu_j - np_j)^2}{np_j} < C \right)$$

Если верна $H_a \Rightarrow \exists i: \; P(x \in \Delta_i) \neq p_i$

по ЗБЧ

$\dfrac{S_1^i + \ldots + S_n^i}{n} \xrightarrow{P} \mathbb{E} S_n^i = P(x \in \Delta_i) = p_k \neq p_i$

$\parallel$

Тогда, рассмотрим i-ое слагаемое: $\dfrac{(\nu_i - np_i)^2}{np_i} = \dfrac{n}{p_i}\left(\dfrac{\nu_i}{n} - p_i\right)^2 \xrightarrow{P} \dfrac{n}{p_i}\underbrace{\left(p_k - p_i\right)^2}_{>0} \xrightarrow[n\to\infty]{} \infty$

Поэтому

$$\alpha_2(\delta) = P_{H_a}\left(\sum_{j=1}^{n} \underbrace{\frac{(\nu_j - np_j)^2}{np_j}}_{\searrow \infty} < C \right) \to 0$$

Построение критерия с помощью доверительного интервала

$\vec{X} \in F_\theta$

$H_0 = \{\theta = \theta_0\}$

$H_a = \{\theta \neq \theta_0\}$

Пусть дан точный Д.И. уровня $1-\varepsilon$: $P(\theta^- < \theta_0 < \theta^+) = 1-\varepsilon$

$$\delta = \begin{cases} 0, & \theta_0 \in [\theta^-, \theta^+] \\ 1, & \theta_0 \notin [\theta^-, \theta^+] \end{cases}$$

$$\alpha_1(\delta) = P_{\theta=\theta_0}(\theta_0 \notin [\theta^-, \theta^+]) = 1 - P_{\theta=\theta_0}(\theta^- \leq \theta_0 \leq \theta^+) = 1-(1-\varepsilon) = \varepsilon$$

Проверка гипотез о совпадении средних двух нормальных совокупностей.

Критерий Стьюдента

$\vec{X} \in N_{a_1, \sigma^2} \quad \vec{Y} \in N_{a_2, \sigma^2}$

Если у выборок дисперсии совпали, то можно проверить совпадение мат. ожиданий

$H_0 = \{a_1 = a_2\}$

$H_a = \{a_1 \neq a_2\}$

1) $\dfrac{nS^2(\vec{X})}{\sigma^2} \in \chi^2_{n-1} \qquad \dfrac{mS^2(\vec{Y})}{\sigma^2} \in \chi^2_{m-1}$

$$\frac{nS^2(\vec{X}) + mS^2(\vec{Y})}{\sigma^2} \in \chi^2_{n+m-2}$$

2) $\overline{X} \in N_{a_1, \frac{\sigma^2}{n}} \qquad \overline{Y} \in N_{a_2, \frac{\sigma^2}{m}}$

$\overline{X} - \overline{Y} \in N_{a_1 - a_2,\ \frac{\sigma^2}{n} + \frac{\sigma^2}{m}}$. Стандартизуем: $\dfrac{(\overline{X} - \overline{Y}) - (a_1 - a_2)}{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \in N_{0,1}$

3) $$\frac{(\overline{X} - \overline{Y}) - (a_1 - a_2)}{\sigma \sqrt{\frac{n+m}{n \cdot m}}} \cdot \sqrt{\frac{\sigma^2(n+m-2)}{nS^2(\vec{X}) + mS^2(\vec{Y})}} \in T_{n+m-2}$$

$$d_T = \frac{(\overline{X} - \overline{Y})\sqrt{nm}\sqrt{n+m+2}}{\sqrt{n+m}\sqrt{nS^2(\vec{X}) + mS^2(\vec{Y})}} \in T_{n+m-2} \text{ при верной } H_0$$

$$\delta = \begin{cases} 0, & |d_T| < t_{1-\frac{\varepsilon}{2}} \\ 1, & |d_T| \geq t_{1-\frac{\varepsilon}{2}} \end{cases}$$ — Критерий Стьюдента

Состоятельность:

$$d = \frac{\sqrt{nm}}{\sqrt{n+m}} \cdot \frac{(\bar{X} - \bar{Y})\sqrt{n+m-2}}{\sqrt{nS^2(X) + mS^2(Y)}} \xrightarrow[\substack{n\to\infty \\ m\to\infty}]{} \infty$$

$\sigma_1 = \sigma_2$ по условию

$$(\bar{X} - \bar{Y}) \xrightarrow[\substack{n\to\infty \\ m\to\infty}]{P} const$$

$\bar{X} \to EX = a_1$, $\bar{Y} \to EY = a_2$

$$\frac{n\overset{\xrightarrow{P}\sigma_1^2}{S^2(X)} + m\overset{\xrightarrow{P}\sigma_2^2}{S^2(Y)}}{n+m-2} \xrightarrow[\substack{n\to\infty \\ m\to\infty}]{P} \frac{\sigma_1^2 (n+m)}{n+m-2} \longrightarrow \underset{const}{\sigma_1^2}$$

$$P_{H_a}(|\overset{\to\infty}{d}| < C) \to 0$$

Проверка гипотез о совпадении дисперсий двух нормальных совокупностей.

Критерий Фишера

$\vec{X} \in N_{a_1, \sigma_1^2} \quad \vec{Y} \in N_{a_2, \sigma_2^2}$

$H_0 = \{\sigma_1 = \sigma_2\}$

$H_a = \{\sigma_1 \neq \sigma_2\}$

$$\left.\begin{aligned} \frac{nS^2(\vec{X})}{\sigma_1} &\in \chi^2_{n-1} \\ \frac{mS^2(\vec{Y})}{\sigma_2} &\in \chi^2_{m-1} \end{aligned}\right\} \quad \overbrace{\frac{nS^2(X)}{\sigma_1 \cdot (n-1)}}^{S_0^2(\vec{X})} \cdot \overbrace{\frac{\sigma_2 (m-1)}{mS^2(\vec{Y})}}^{\frac{1}{S_0^2(\vec{Y})}} \in F_{n-1, m-1}$$

Теорема Фишера: $d_F = \dfrac{S_0(\vec{X})}{S_0(\vec{Y})} \in F_{n-1, m-1}$ при верной H_0.

$$\delta = \begin{cases} 0, & f_{\frac{\varepsilon}{2}} \leq d_F \leq f_{1-\frac{\varepsilon}{2}} \\ 1, & \text{иначе} \end{cases}$$

miro

Сост. кр. Фишера

$H_0 = \{\sigma_1^2 = \sigma_2^2\}$

$H_a = \{\sigma_1^2 \neq \sigma_2^2\}$

$\vec{X} = (X_1, \dots, X_n) \quad \vec{Y} = (Y_1, \dots, Y_n)$

$$d_F = \frac{S_0^2(\vec{X})}{S_0^2(\vec{Y})} \in F_{n-1, m-1}$$

$$\delta = \begin{cases} 0, & f_{\frac{\varepsilon}{2}} \leq \dfrac{S_0(\vec{X})}{S_0(\vec{Y})} \leq f_{1-\frac{\varepsilon}{2}} \\ 1, & \text{иначе} \end{cases}$$

Убедимся, что посл-ть квантилей $f_\delta = f_\delta(n,m)$ распр $F_{n,m}$ $\forall$ уровня $\delta \in (0,1)$ сходится к 1 при $n, m \to \infty$:

miro

Пусть $\xi_{n,m} \in F_{n,m}$ по опр квантиля $\underline{P}(\xi_{n,m} < f_\delta) = \delta$

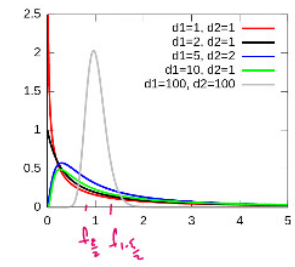
$\underline{P}(\xi_{n,m} > f_\delta) = 1-\delta$

Опр. $Y \in F_{n,m}$ если $Y = \dfrac{\frac{Z_1}{m}}{\frac{Z_2}{k}} = \dfrac{k}{m} \cdot \dfrac{Z_1}{Z_2}$

$Z_1 \in \chi^2_m$

$Z_2 \in \chi^2_k$

$$\boxed{\frac{Z_1}{m} = \frac{\xi_1^2 + \ldots + \xi_m^2}{m} \xrightarrow{P} \underline{E\xi_1^2 = 1}}$$



т.е. $\xi_{n,m} \xrightarrow[n,m\to\infty]{P} 1$

Поэтому, $\underline{P}(\xi < 1-\varepsilon) \xrightarrow[n,m\to\infty]{} 0$

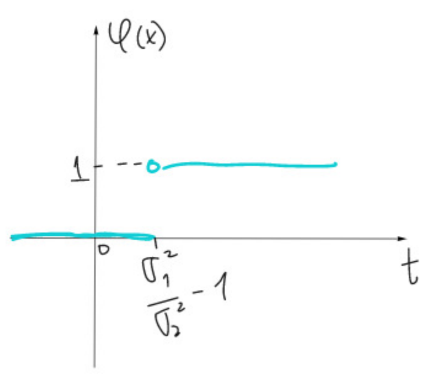
$\underline{P}(\xi > 1-\varepsilon) \xrightarrow[n,m\to\infty]{} 0$

$$\left.\begin{matrix} \underline{P}(\xi < 1-\varepsilon) \xrightarrow[n,m\to\infty]{} 0 \\ \underline{P}(\xi > 1-\varepsilon) \xrightarrow[n,m\to\infty]{} 0 \end{matrix}\right\} \Rightarrow \underbrace{1-\varepsilon < f_\delta < 1+\varepsilon}_{\text{при больших } n,m} \Rightarrow f_\delta \xrightarrow[n,m]{} 1 \quad (*)$$

Далее, т.к. верна H_a, то $\sigma_1 \neq \sigma_2 \Rightarrow \dfrac{S_0^2(\vec{X})}{S_0^2(\vec{Y})} \xrightarrow[n,m\to\infty]{P} \dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$

И по $(*)$ можем сказать, что $\dfrac{S_0^2(\vec{X})}{S_0^2(\vec{Y})} - f_{1-\varepsilon} \to \underbrace{\dfrac{\sigma_1^2}{\sigma_2^2} - 1}_{>0}$

$\varphi(x) = \underline{P}\left(\dfrac{\sigma_1^2}{\sigma_2^2} - 1 < x\right)$ — непр в $x=0$ и очевидно, $\underline{P}\left(\dfrac{\sigma_1^2}{\sigma_2^2} - 1 < 0\right) = 0$



$$\alpha_2 = \underline{P}_{H_a}\left(\frac{S_0^2(\vec{X})}{S_0^2(\vec{Y})} < f_{1-\varepsilon}\right) =$$

$$= \underline{P}\left(\frac{S_0^2(\vec{X})}{S_0^2(\vec{Y})} - f_{1-\varepsilon} < 0\right) \xrightarrow{P} \underline{P}\left(\frac{\sigma_1^2}{\sigma_2^2} - 1 < 0\right) = 0$$

Критерий Колмогорова-Смирнова однородности двух выборок.