О.М. Дюсуше

К ВОПРОСУ О ПРОБЛЕМЕ БЕРЕНСА-ФИШЕРА: ПРИМЕНЕНИЕ ПОДХОДА НЕЙМАНА-ПИРСОНА

В отличие от тестирования на значимость гипотезы равенства двух средних приближенными методами и построения доверительных границ (подход Беренса-Фишера-Уэлча) применяется метод тестирования простых альтернативных гипотез (подход Неймана-Пирсона) на основе построения и предварительного анализа точных распределений статистик. Примеры рассчитаны с использованием данных Леманна.

O. Dusouchet

Applying of Neyman-Pearson approach to the problem of Behrence-Fisher

In contrast to test of significance of hypothesis of equality two means for constructing confidence interval by approximate methods (Behrens-Fisher-Welch approach) here is used method of testing simple alternative hypotheses (Neyman-Pearson approach), based on selecting the appropriate statistics and constructing exact distributions. Lehmann's data are used to calculate some examples.

Введение

Беренс (1929) рассматривал задачу о равенстве генеральных средних по малым выборкам из двух независимых нормальных совокупностей $N\{\mu_1, \sigma_1\}$, $N\{\mu_2, \sigma_2\}$. Вопрос построения доверительных границ оказался более сложным, чем в случае Госсета (Стьюдента), что привлекло внимание Фишера и многих других ведущих статистиков к решению этой достаточно актуальной задачи. Введение в проблематику можно найти в Википедии - общедоступных мультиязычных универсальных интернет-энциклопедиях, например, на русском, немецком, английском языках. Обзор и обсуждение основных исследований задачи «о двух средних» периода до середины 60-х есть втором томе монографии Кендалла и Стьюарта (1973, гл. 21). Критерий сравнения средних с соответствующими таблицами можно найти в таблицах Большева и Смирнова (1983). Ким и Кохен (1998) в обзоре более 80-ти публикаций (включая анализ 10-таблиц) по проблеме Беренса-Фишера отмечали, что она стала одной из фокальных точек противоречия трех теоретических подходов: частотного (Нейман-Пирсон (1928, 1933) против фидуциального (Фишер (1935, 1939) и байесовского (Джеффри (1940), заключив, что не существует никаких конечных решений в рамках рассмотренных работ. Основанием подобного заключения стала серия работ Линника (1963, 1964 и др., англ. 1968/1966) и других авторов, в которых было показано, в том числе, на примере теста Вальда (1955), что однородный наиболее мощный критерий не существует в контексте задачи Беренса-Фишера.

Фишер (1935), будучи сторонником (и создателем) *теста на значимость*, отмечал, что в статистических задачах нулевая гипотеза никогда не может быть доказана или установлена, но может быть опровергнута[1]. Также Фишер (1934) утверждал, что, если существует такая достаточная статистика, что функция максимального правдоподобия выборки включает только эту статистику, и ее распределение выражается в терминах функции правдоподобия, то только эти случаи позволяют получать тесты на значимость типа равномерно наиболее мощных тестов Неймана-Пирсона относительно класса альтернатив. Если такой статистики не существует, Фишер рекомендовал искать подходящую статистику, обеспечивающую использование всей необходимой информации, содержащейся в выборке.

Настоящий анализ здесь мотивируется целью построения и сравнения *точных распределений* известных статистик Беренса-Фишера-Уэлча и Вальда в конечном виде, и применения подхода Неймана-Пирсона (1933) для *тестирования простых альтернативных гипотез.* Рей и Питмен [9] получили точное (центральное) распределение статистики Беренса-Фишера для нечетных объемов двумерной выборки в виде взвешенной суммы

[1] Цитируется по “Earliest Uses of Symbols in Probability and Statistics” со ссылкой на работу Фишера “The Design of Experiments” (1935, с. 19). Используемые термины и символы - там же, http://jeff560.tripod.com/h.html см. раздел H – “Hypothesis testing”, “Test of significance” и др.

распределений Стьюдента. Здесь распределения в конечном виде были построены методами Вентцель и Овчарова (1969, гл. 8) путем эквивалентных преобразований монотонных функций от одной или двух независимых случайных величин – к функциям от случайных величин с известными распределениями. Это позволило определить зависимость функции мощности от параметров распределений, в том числе от, так называемого, «мешающего» параметра λ и показать неоднородность ошибки второго рода по паре параметров U и λ. Три рассматриваемые статистики преобразованы одним способом, поэтому вывод функции плотности дан только для первой статистики.

1. Распределение статистики *u* Уэлча

Распределения статистики *u,* рассмотренной в статье Уэлча (1938, с.350), и составляющих компонент V и S после ее преобразования проводятся ниже:

$$u = (\overline{x_1} - \overline{x_2}) / \sqrt{s_1^2/n_1 + s_2^2/n_2}\ , \tag{1}$$

где $\overline{x_1} = \dfrac{\sum_{i=1}^{n_1} x_1^i}{n_1}$; $\overline{x_2} = \dfrac{\sum_{i=1}^{n_2} x_2^i}{n_2}$; $s_1^2 = \dfrac{\sum_{i=1}^{n_1}(x_1^i - \overline{x_1})^2}{n_1 - 1}$; $s_2^2 = \dfrac{\sum_{i=1}^{n_2}(x_2^i - \overline{x_2})^2}{n_2 - 1}$ - оценки, состоятельные и несмещенные, неизвестных параметров μ_1, μ_2, D_1, D_2 двух нормальных распределений.

Распределение числителя $\boldsymbol{V = (z+U)\sqrt{D_V}}$:

$$u = \left(\frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{D_1/n_1 + D_2/n_2}} + \frac{(\mu_1 - \mu_2)}{\sqrt{D_1/n_1 + D_2/n_2}} \right) \frac{\sqrt{D_V}}{\sqrt{\lambda c w_1 + w_2}} = \frac{(z+U)\sqrt{D_V}}{\sqrt{\lambda c w_1 + w_2}} = \frac{V}{Z},$$

$$\lambda = \frac{D_1}{D_2},\ c = \frac{n_2(n_2-1)}{n_1(n_1-1)},\ D_V(\lambda) = \frac{D_1/n_1 + D_2/n_2}{D_2/((n_2-1)n_2)} = \lambda c(n_1 - 1) + (n_2 - 1),$$

$$z \in N(0,1),\quad U = const,\quad V \in N(U\sqrt{D_V}, D_V),\ f_V(V) = \frac{\exp\left(-\frac{1}{2}(V - U\sqrt{D_V})^2\right)}{\sqrt{2pD_V}}. \tag{2}$$

Плотности распределений выражения $\boldsymbol{Z=S^2}$ ***и знаменателя*** $\boldsymbol{Z=S}$ ***для квантили*** *u*

$$w_i = \frac{s_i^2(n_i - 1)}{D_i} \in \chi^2_{vi},\ f_{wi}(w_i) = \frac{w_i^{\frac{vi-2}{2}}}{2^{\frac{vi}{2}}\Gamma(v_i/2)},\ v_i = n_i - 1,\ i = 1,\ 2,\quad f_{\lambda cw1}(y) = \frac{\chi^2_{v1}(y/\lambda c)}{\lambda c},\ \text{где}$$

$w > 0$, Γ(.)– гамма-функция; Плотность распределения функции $y = \varphi(x)$ с распределением случайной величины *x*: *f(x),* равна: $f_y(y)=f(\psi(y))\left|\psi'(y)\right|$, где $\psi(y)$ – функция обратная к функции $y(x)$; $\left|\psi'(y)\right|$ – модуль производной от функции $\psi(y)$. Плотности распределения случайных, независимых величин под корнем S^2 в (1) и величины $Z = \sqrt{S^2}$ равны:

$$f_{S^2}(S^2) = \int_0^{S^2} f_{\lambda cw1}(y) f_{w2}(S^2 - y)dy = \int_0^{S^2} \frac{(y/\lambda c)^{\frac{(n_1-3)}{2}} \exp\left(-\frac{y}{2\lambda c}\right)}{\lambda c \cdot 2^{\frac{(n_1-1)}{2}} \Gamma\left(\frac{n_1-1}{2}\right)} \frac{(S^2-y)^{\frac{(n_2-3)}{2}} \exp\left(-\frac{S^2-y}{2}\right)}{2^{\frac{(n_2-3)}{2}} \Gamma\left(\frac{n_2-1}{2}\right)} dy$$

$$f_Z(Z) = Z' f_Z(Z) = 2Z f_{S^2}(Z^2) = 2Z \frac{\int_0^{Z^2} y^{\frac{n_1-3}{2}} (Z^2 - y)^{\frac{n_2-3}{2}} \exp\left(-\frac{y}{2\lambda c} - \frac{Z^2 - y}{2}\right) dy}{(\lambda c)^{\frac{n_1-1}{2}} 2^{\frac{n_1+n_2-2}{2}} \Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)} \tag{3}$$

Плотность распределения статистики u **(1):** $f_u(u) = \int_0^{+\infty} Z \cdot f_V(uZ) f_Z(Z) dZ =$

$$
= \frac{\int_0^{\infty}\int_0^{Z^2} Z^2 y^{\frac{n_1-3}{2}} (Z^2-y)^{\frac{n_2-3}{2}} \exp\left(-\frac{1}{2}\left(\frac{u\cdot Z}{\sqrt{\lambda c(n_1-1)+n_2-1}}-U\right)^2-\frac{y}{2\lambda c}-\frac{Z^2-y}{2}\right)dydZ}{\sqrt{\pi(\lambda c(n_1-1)+n_2-1)}\,(\lambda c)^{\frac{n_1-1}{2}}\,2^{\frac{n_1+n_2-3}{2}}\,\Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)} \qquad (4)
$$

УТВЕРЖДЕНИЕ: Плотность распределения статистики (1) имеет особенность в окрестности $\delta=\mu_1-\mu_2=0$, или $U=0$, где плотность распределения определена как функция от одного параметра – отношения дисперсий λ : $f_u(u, \lambda \backslash n_1, n_2)$; а при $\delta \neq 0$ плотность распределения статистики (1) может быть определена как параметрическая функция от генеральной квантили для выборочных средних U и отношения дисперсий λ : $f_u(u, U, \lambda \backslash n_1, n_2)$, и в общем случае имеет вид:

$$
f_u(u) = \frac{\int_0^{\infty}\int_0^{S^2} K(\lambda) S x^{\frac{n_1-3}{2}} (S^2-x)^{\frac{n_2-3}{2}} \exp\left(-\frac{1}{2}\left(u\frac{S}{\sqrt{D_V(\lambda)}}-U(\delta,\lambda,\sigma_2)\right)^2-\frac{x}{2\lambda c}-\frac{(S^2-x)}{2}\right)dx\,dS}{\sqrt{\pi(\lambda c(n_1-1)+n_2-1)}\,(\lambda c)^{\frac{n_1-1}{2}}\,2^{\frac{n_1+n_2-3}{2}}\,\Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)} \qquad (5)
$$

Параметрическое представление функции плотности позволяет анализировать альтернативы $U=0$ и $U \neq 0$, которые определяют $\delta=0$ и $\delta \neq 0$.

$$
\int_0^{100}\int_0^{S^2} S^2 \cdot \frac{1}{y^{\frac{(n1-1)}{2}}\sqrt{y\cdot cc2+n2-1}\cdot AK}\cdot\exp\left[\frac{-ss}{2\cdot cc1\cdot y}-\frac{S^2-ss}{2}-\left[\left(U1-\frac{U1\cdot S}{\sqrt{y\cdot cc2+n2-1}}\right)^2\right]\cdot 0.5\right]\cdot ss^{\frac{n1-3}{2}}\cdot\left(S^2-ss\right)^{\frac{n2-3}{2}}\,dss\,dS
$$

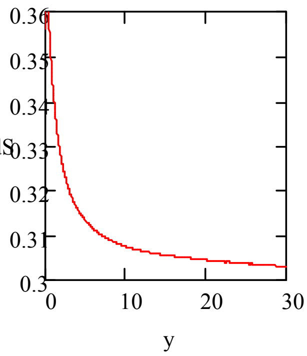


Рис. 1. Зависимость плотности u (λ) в точке $u=U1$ (μ_1- μ_2, D_1, D_2). Расчеты Mathcad-14 ($\lambda=y$)

2. Распределение статистики *v* Беренса- Фишера

Статистика v Беренса-Фишера, рассмотренная в таблицах [4, с. 204]имеет вид:

$$
v=((\overline{x_1}-\overline{x_2})-(\mu_1-\mu_2))/\sqrt{s_1^2/n_1+s_2^2/n_2}\,. \qquad (6)
$$

Эквивалентными преобразованиями статистика v (6) приводится к виду, зависящему только от параметра отношения дисперсий λ:

$$
v=\frac{z\sqrt{D_V}}{\sqrt{\lambda c w_1+w_2}}, \quad z\in N(0,1), \quad w_i\in\chi^2_{n_i-1}, i=1,2\,,
$$

Аналогично порядку, показанному в п.1 получена плотность распределения $f(v)$, которая имеет вид:

$$
f_v(v,\lambda\backslash n_1,n_2)=\int_0^{\infty}\int_0^{S^2} K(\lambda) S x^{\frac{n_1-3}{2}}(S^2-x)^{\frac{n_2-3}{2}}\exp\left(-\frac{1}{2}\left(u\frac{S}{\sqrt{D_V(\lambda)}}\right)^2-\frac{x}{2\lambda c}-\frac{(S^2-x)}{2}\right)dxdS \qquad (7)
$$

Сравнение распределения статистик u и v: (5) и (7) показывает, что в выражении (7) параметр U=0, δ=0, и плотность распределения зависит только от отношения дисперсий, но не их размера.

3. Тест Вальда

Вальд [11] исследовал проблему равенства средних и критический регион для статистики, удовлетворяющей требованиям достаточности и инвариантности, в виде неравенства:

$$(\overline{x_1}-\overline{x_2})/\sqrt{s_1^2+s_2^2} \ge \varphi\left(s_2^2/s_1^2\right) \quad . \tag{8}$$

Эквивалентные преобразования статистики W (8) и распределения ее компонент показывают, что $D_v > D_w$, $c = c_w n_2/n_1$. Соотношение рассеивания статистик (1) и (8) зависит от соотношения объемов выборки. В примере Леманна рассеивание статистики (8) меньше чем у (1), т.е. статистика (8) более эффективна.

$$u_w = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{s_1^2+s_2^2}} = \left(\frac{\overline{x_1}-\overline{x_2}-(\mu_1-\mu_2)}{\sqrt{D_1/n_1+D_2/n_2}} + \frac{(\mu_1-\mu_2)}{\sqrt{D_1/n_1+D_2/n_2}}\right)\cdot\frac{\sqrt{D_1/n_1+D_2/n_2}}{\sqrt{D_2/(n_2-1)}\sqrt{C_w w_{v1}+w_{v2}}} = \frac{(z+U)\sqrt{D_w}}{\sqrt{c_w w_1+w_2}}$$

$$\lambda=\frac{D_1}{D_2},\ u=\frac{(z+U)\sqrt{D_w}}{\sqrt{c_w w_1+w_2}}=\frac{W}{S},\ c_w(\lambda)=\lambda\frac{(n_2-1)}{(n_1-1)}=\lambda c_w,\ D_w(\lambda)=\lambda c_w\frac{(n_1-1)}{n_1}+\frac{(n_2-1)}{n_2}\ ,$$

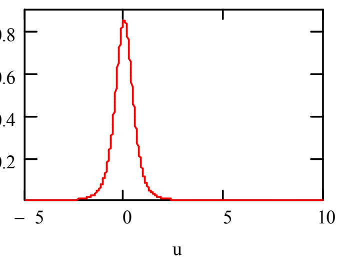
Аналогичные выкладки позволяют получить плотность распределения статистики Вальда (9)

$$f_W(W,U,\lambda)=\frac{\int\limits_0^\infty\int\limits_0^{S^2} S^2 y^{\frac{n_1-3}{2}}(S^2-y)^{\frac{n_2-3}{2}}\exp\left(-\frac{1}{2}\left(\frac{uS}{\sqrt{\left(\lambda c_w\frac{n_1-1}{n_1}+\frac{n_2-1}{n_2}\right)}}-U\right)^2-\frac{y}{2\lambda c_w}-\frac{S^2-y}{2}\right)dydS}{\sqrt{\pi\left(\lambda c_w\frac{n_1-1}{n_1}+\frac{n_2-1}{n_2}\right)}(\lambda c_w)^{\frac{n_1-1}{2}}2^{\frac{n_1+n_2-3}{2}}\Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)}.$$

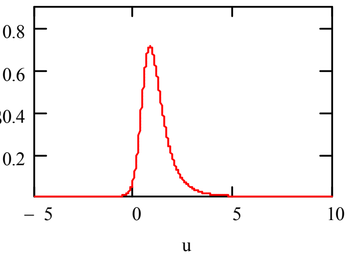
Распределение статистики φ в правой части неравенства (8) зависит только от величины, имеющей распределение Фишера и от параметра отношения дисперсий λ, то есть, принадлежит к однопараметрическому семейству. Статистика Вальда в левой части критерия (8) аналогично статистике Уэлча может рассматриваться в двухмерном $\{U, \lambda\}$ параметрическом пространстве. Фактически, в неопубликованной задаче Вальд должен был прийти к ограничению функции мощности, определенной в двухмерном пространстве, посредством критерия, определенного в одномерном пространстве. В исследованиях Линника и Шалаевских было показано, что в задаче Беренса-Фишера критериев подобных по размеру критерия и параметрам (σ_1, σ_2) не существует, что соответствует заключению Линника о невозможности существования теста Вальда подобного относительно счетного ограниченного множества значений $\lambda=\sigma_1^2/\sigma_2^2$.

Функция плотности статистики Вальда позволяет определить математическое ожидание и дисперсию. Графически можно показать, что при прочих равных распределение f_W более компактно, и поэтому статистика Вальда предпочтительнее статистики Уэлча. На рис. 2. приведены распределения плотности Вальда и Уэлча, соответствующие гипотезам: $H_0:\ U_0=0,\ \ \lambda=30.644$, и $H_1:\ U_1=2.143,\ \ \lambda=30.644$.

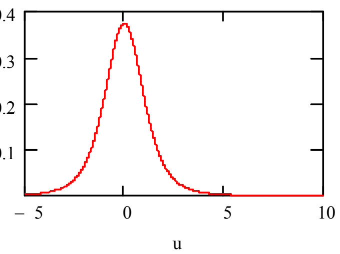
$$\int_0^{200}\int_0^{S^2} S^2\cdot KKw\cdot\exp\left[\frac{-x}{2\cdot cc}-\frac{S^2-x}{2}-\left[\left(-U+u\cdot\frac{S}{SVw}\right)^2\right]\cdot 0.5\right]\cdot x^{\frac{n1-3}{2}}\cdot\left(S^2-x\right)^{\frac{n2-3}{2}}\,dx\,dS$$



$$\int_0^{200}\int_0^{S^2} S^2\cdot KKw\cdot\exp\left[\frac{-x}{2\cdot cc}-\frac{S^2-x}{2}-\left[\left(-U1+u\cdot\frac{S}{SVw}\right)^2\right]\cdot 0.5\right]\cdot x^{\frac{n1-3}{2}}\cdot\left(S^2-x\right)^{\frac{n2-3}{2}}\,dx\,dS$$



$$\int_0^{200}\int_0^{S^2} S^2\cdot KK\cdot\exp\left[\frac{-x}{2\cdot cc}-\frac{S^2-x}{2}-\left[\left(-U+u\cdot\frac{S}{SV}\right)^2\right]\cdot 0.5\right]\cdot x^{\frac{n1-3}{2}}\cdot\left(S^2-x\right)^{\frac{n2-3}{2}}\,dx\,dS$$



$$\int_0^{200}\int_0^{S^2} S^2\cdot KK\cdot\exp\left[\frac{-x}{2\cdot cc}-\frac{S^2-x}{2}-\left[\left(-U1+u\cdot\frac{S}{SV}\right)^2\right]\cdot 0.5\right]\cdot x^{\frac{n1-3}{2}}\cdot\left(S^2-x\right)^{\frac{n2-3}{2}}\,dx\,dS$$

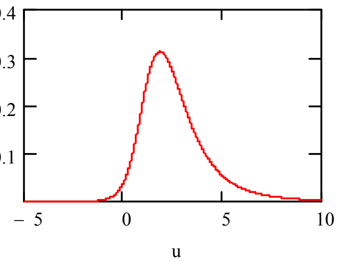


Рис. 2 Распределения плотности статистик (1) и (8) при U=0, и $U1=(d/S)=2.143$ в примере Леманна

4. Проверка простых альтернативных гипотез в задаче Беренса-Фишера на основе подхода Неймана-Пирсона

Простой гипотезой относительно распределения случайных величин согласно Крамеру [6, с.574] называется *точка* в пространстве параметров распределения этих величин. Простые альтернативные гипотезы как точки в 2-х мерном пространстве параметров, позволяют определить гипотезы и *критерий тестирования* двух распределений с параметрами, соответствующими гипотезам H_0 и H_1. Константа k ограничивает область интегрирования по u, и определяет уровень ошибок первого и второго рода. Если выборочная квантиль u не попадает в область вокруг нуля, это означает, что альтернативная гипотеза более вероятна, чем нулевая. Достаточно малые ошибки должны обеспечиваться соответствием объемов выборки и параметров адекватной и приемлемой альтернативы(*):

$$\begin{cases} H_0: \quad \delta = 0, \quad \lambda = \lambda_0 \Rightarrow \quad \text{если} \quad |u| < k, \quad \text{принимают} \quad H_0; \\ H_1: \quad \delta = \delta_1, \quad \lambda = \lambda_1, \sigma_2 = \sigma_{21} \Rightarrow \quad \text{если} \quad |u| \geq k, \quad \text{принимают} \quad H_1. \end{cases}$$

Система уравнений рисков при *размере* (α) и *мощности* (1- β) *критерия* имеет вид:

$$\begin{cases} \int_{-k}^{k} f_W(u, n_1, n_2, \lambda \setminus H_0) du = 1 - \alpha \\ \int_{-k}^{k} f_W(u, n_1, n_2, \lambda, \sigma_2 \setminus H_1) du = \beta \end{cases}$$

Решение уравнения при выполнении условия (*) в общем случае целесообразно объединить с целочисленным решением задачи планирования оптимальных объемов испытаний, т.к. задание уровня ошибок первого и второго рода, обоснование альтернативной гипотезы и выбор объемов испытаний взаимосвязаны.

Интервал [-k, k] определяет дополнение к критической области. Нейман и Пирсон [11] при заданном объеме испытаний советуют задать ошибку первого рода, и проверять приемлемость ошибки второго рода. В этой задаче мы вынуждены поступить наоборот – определить k=2.87 из условия β=0.05 (при H_1 : $U1$= 2,5):

$$\int_{-k}^{k}\int_{0}^{200}\int_{0}^{S^2} S^2 \cdot KKw \cdot \exp\left[\frac{-x}{2 \cdot cc} - \frac{S^2 - x}{2} - \left[\left(-U1 + u \cdot \frac{S}{SVw}\right)^2\right] \cdot 0.5\right] \cdot x^{\frac{n1-3}{2}} \cdot \left(S^2 - x\right)^{\frac{n2-3}{2}} dx\, dS\, du = 0.9505976$$

$$\int_{-k}^{k}\int_{0}^{200}\int_{0}^{S^2} S^2 \cdot KKw \cdot \exp\left[\frac{-x}{2 \cdot cc} - \frac{S^2 - x}{2} - \left[\left(u \cdot \frac{S}{SVw}\right)^2\right] \cdot 0.5\right] \cdot x^{\frac{n1-3}{2}} \cdot \left(S^2 - x\right)^{\frac{n2-3}{2}} dx\, dS\, du = 0.9976267$$

В этом случае k=2.87 и Вер (u>k/ H_1)=1- β=0.95 и Вер (u>k/ H_0)=α=0.002. Принимается нулевая гипотеза если выборочная квантиль ограничена значением k.

Почему не существует однородного наиболее мощного критерия для проверки абсолютной разности средних? Во-первых, из функции правдоподобия двумерной выборки при четырех неизвестных параметрах не удается выделить зависимость от разности средних, $\delta = \mu_1 - \mu_2$. Согласно рекомендациям Фишера рассматриваются критерии, основанные не на функции правдоподобия, а на функции от достаточных статистик выборки. В качестве «хорошей» статистики выбрана статистика Вальда (1955)[2].

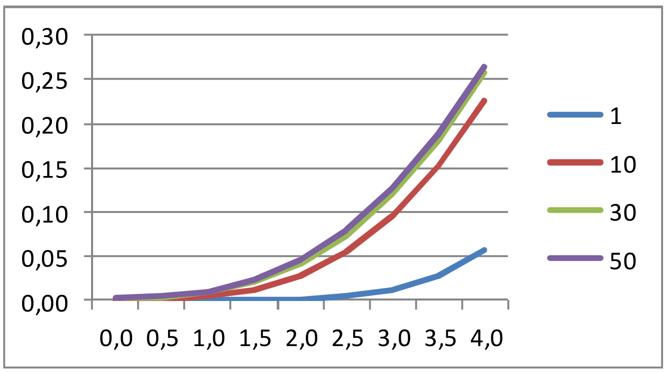


Рис.3. Функции мощности критерия в зависимости от квантили U по для 4-х значений λ.

Во вторых, если из некоторых прагматических соображений исследователь выбирает альтернативную гипотезу H_1, то не для всех параметров λ равномерно обеспечивается

[2] Статистика Уэлча описывает вероятность непревышения средних значений Вер($\overline{x_2} < \overline{x_1}$), а статистика Вальда вероятность непревышения единичных реализаций Вер ($x_2 < x_1$). Оба распределения зависят от параметра разности генеральных средних $\mu_1 - \mu_2$.

размер критерия α. Невозможно построить равномерно мощный критерий для какой-либо области λ (например, по горизонтали уровня 0,05).

Библиографический список

1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983.
2. Вентцель Е.С., Овчаров Л.А. Теория вероятностей. Гл.8. – М.: Наука, 1969.
3. Кендалл М. Дж., Стьюарт А. Статистические выводы и связи. – М.: Наука, 1973.
4. Крамер Г. Математические методы статистики. – М.: Мир, 1975.
5. Линник Ю.В., Шалаевский О.В. К аналитической теории тестов для проблемы Беренса-Фишера Доклады Академии наук СССР, 1963, Том 150, №.1.
6. Линник Ю.В. О тесте А. Вальда для сравнения двух нормальных выборок Теория вероятностей и ее применение. – 1964. – № 9. – С. 16-30.
7. Behrens, W.V. Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen// *Landwirtsch. Jahrbücher* 68, 1929. – pp. 807–837.
8. Fisher R. Two new properties of mathematical likelihood// Proceedings of the Royal Society, 1934. – A, 144: 285-307.
9.. Fisher R. Fiducial argument in statictical inference// Annals of Eugenics, 1935, 6: 391-398. http://digital.library.adelaide.edu.au/dspace/handle/2440/15222
10. Kim S-H, Cohen A.S. On the Behrens-Fisher Problem: A Review // Journal of Educational and Behavioral Statistics, Winter. – 1998. – Vol. 23. – No. 4. – pp. 356-377.
11. Neyman J., E.S. Pearson. On the Problem of the Most Efficient Test of Statistical Hypotheses// Philosophical Transactions of the Royal Society, Series A. (1934) 1933. – Vol. 231, pp. 289-337.
12. Ray, W. D. and Pitman, E. N. T. An Exact Distribution of the Fisher-Behrens-Welch Statistic// Journal of the Royal Statistical Society, 23 (1961).
13. Wald A. Testing the difference between the means of two normal populations with unknown standard deviations// Selected papers in statistics and probability by Abraham Wald. – pp. 669–695. McGraw-Hill Book, Co., Inc.New York-Toronto-London, 1955. Цит. по Mathematical Reviews on the Web (MathSciNet, MR0070918).
14. Welch, B.L. The Significance of the Difference Between Two Means when the Population Variances are Unequal// Biometrika, 1938. – Vol. 29. – No. – Feb. – pp. 350-362.