2.2 Построение доверительных областей для вектора параметров μ многомерного нормального распределения.

Перейдем теперь к построению доверительной области для компонент вектора математических ожиданий случайного вектора $\vec{X} = \{X_1, X_2, \ldots, X_p\}$ при заданной доверительной вероятности α.

1) Доверительная область для вектора $\mu = (m_1, m_2, \ldots, m_p)^T$, если ковариационная матрица Σ известна.

Для одномерной нормальной случайной величины при построении доверительного интервала для среднего μ при известном σ использовалась статистика:

$$\eta = \sqrt{n}\frac{\bar{X} - m}{\sigma}, \quad \eta \in N(0,1).$$

Если возвести η в квадрат, получим статистику, распределенную по закону хи-квадрат с одной степенью свободы:

$$\eta^2 = n\frac{(\bar{X} - m)^2}{\sigma^2} = n(\bar{X} - m)^T \left(\sigma^2\right)^{-1} (\bar{X} - m) \Rightarrow \; \eta^2 \in \chi^2_1.$$

В случае многомерного распределения генеральной совокупности можно построить аналогичную статистику:

$$\eta^2 = n(\bar{X} - \vec{m})^T \Sigma^{-1} (\bar{X} - \vec{m}), \tag{11}$$

здесь $\bar{X}$ - вектор выборочных средних многомерной случайной величины $\vec{X}$, Σ матрица ковариаций. Статистика η^2 в случае p-мерного нормального распределения генеральной совокупности будет иметь распределение хи-квадрат с p-степенями свободы.

Пусть χ^2_α - критическая точка распределения хи-квадрат с p-степенями свободы уровня α ($\alpha = 1 - \beta$). Так как матрица ковариаций положительно определенная квадратичная форма (при отсутствии линейной зависимости между компонентами), то уравнение

$$n(\bar{X} - \vec{\mu})^T \Sigma^{-1} (\bar{X} - \vec{\mu}) = \chi^2_\alpha$$

определяет p-мерный эллипсоид в системе координат $\{m_1, m_2, \ldots, m_k\}$ с центром $\bar{X}$. Соответственно, доверительной областью уровня β будет область, ограниченная этим эллипсоидом, т.е. область определяемая неравенством:

$$n(\bar{X} - \vec{\mu})^T \Sigma^{-1} (\bar{X} - \vec{\mu}) \leq \chi^2_\alpha. \tag{12}$$

2) Пусть теперь ковариационная матрица Σ случайной величины $\vec{X}$ неизвестна. Для одномерной случайной величины, для построения доверительного интервала для среднего μ в этом случае использовалась статистика

$$T = \sqrt{n}\,\frac{\bar{X} - \mu}{s}, \quad t \in t_{n-1}.$$

Если возвести T в квадрат получим статистику:

$$T^2 = n(\bar{X} - m)^T \left(s^2\right)^{-1} (\bar{X} - m) = n\frac{(\bar{X} - m)^2}{s^2},$$

которая будет иметь распределение Фишера с $\nu_1 = 1$ и $\nu_2 = n - 1$ степенями свободы ($T^2 \in F_{1,n-1}$).

В случае многомерного распределения генеральной совокупности можно построить аналогичную статистику:

$$T^2 = n(\bar{X} - \vec{m})^T S^{-1}(\bar{X} - \vec{m}), \tag{13}$$

которая в случае p-мерного нормального распределения генеральной совокупности будет иметь распределение Хотеллинга с $\nu_1 = p$, $\nu_2 = n - p$ степенями свободы

Лемма 2.2. ***(о связи распределения Хотеллинга и распределения Фишера)*** *Пусть $X_1, X_2, \ldots, X_n$ выборка из p- мерного нормального распределения с параметрами $\vec{\mu}$ и Σ и $T^2 = n(\bar{X} - \vec{\mu})^T S^{-1}(\bar{X} - \vec{\mu})$. Тогда величина $\dfrac{T^2}{n-1}\dfrac{n-p}{p}$ имеет F-распределение с p и $n - p$ степенями свободы.*

Доказательство можно посмотреть в книге Андерсона "Введение в многомерный статистический анализ."М.: Физматгиз, 1963

Из леммы следует, что случайную величину, имеющую распределение Хотеллинга можно получить из случайной величины, имеющей распределение Фишера очень просто:

$$T^2_{p,n-p} = \frac{p\,n}{n-p} F_{p,n-p}$$

Пусть T^2_α - критическая точка распределения Хотеллинга с $\nu_1 = p$, $\nu_2 = n - p$ степенями свободы. Так как

$$P\{n(\bar{X} - \mu)^T S^{-1}(\bar{X} - \mu) \le T^2_\alpha\} = \alpha, \tag{14}$$

доверительная область уровня α будет определяться неравенством:

$$n(\bar{X} - \mu)^T S^{-1}(\bar{X} - \mu) \le T^2_\alpha. \tag{15}$$

Пример 1. По выборке объема $n = 40$ из двумерной нормальной совокупности $\{X_1, X_2\}$ найден вектор выборочных средних $\bar{X} = \begin{pmatrix} -0,063 \\ -1,218 \end{pmatrix}$ и выборочная ковариационная матрица $S = \begin{pmatrix} 1,497 & 1,569 \\ 1,569 & 4,917 \end{pmatrix}$. Построить доверительную область уровня $\alpha = 0,9$ для вектора математических ожиданий $\vec{\mu} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$, если матрица ковариаций:

а) известна и равна:

$$\Sigma = \begin{pmatrix} 1 & 0,8 \\ 0,8 & 4 \end{pmatrix}$$

;

б) неизвестна.

Если ковариационная матрица известна, то доверительная область определяется неравенством:

$$n(\bar{X} - \vec{\mu})^T \Sigma^{-1} (\bar{X} - \vec{\mu}) \leq \chi^2_\alpha,$$

где χ^2_α - квантиль распределения хи-квадрат с 2 степенями свободы уровня $\alpha = 0,9$. Из таблиц (или используя статистические функции пакетов прикладных программ) находим $\chi^2_\alpha \approx 4,605$. Подставляя значения в неравенство, получим:

$$40 \cdot \begin{pmatrix} -0,063 - m_1 & -1,218 - m_2 \end{pmatrix} \begin{pmatrix} 1 & 0,8 \\ 0,8 & 4 \end{pmatrix}^{-1} \begin{pmatrix} -0,063 - m_1 \\ -1,218 - m_2 \end{pmatrix} \leq 4,605,$$

или:

$$47,619 m_1^2 - 19,048 m_1 m_2 + 11,905 m_2^2 - 17,2 m_1 + 27,8 m_2 \leq 11,783.$$

Полученная область изображена на рисунке 1 (область заключена внутри синего эллипса).

б) Если ковариационная матрица неизвестна, то доверительная область определяется неравенством:

$$n(\bar{X} - \vec{\mu})^T S^{-1} (\bar{X} - \vec{\mu}) \leq T^2_\alpha,$$

где T^2_α - квантиль распределения Хотеллинга с $\nu_1 = p = 2$, $\nu_2 = n - p = 38$ степенями свободы. Квантиль распределения Хотеллинга T^2_α находим через соответствующий квантиль распределения Фишера:

$$T^2_\alpha = \frac{p(n-1)}{n-p} F_\alpha$$

Для $\nu_1 = 2$, $\nu_2 = 38$ и $\alpha = 0,9$ квантиль распределения Фишера $F_\alpha \approx 2,448$, соответственно $T^2_\alpha \approx 5,025$.

Подставляя значения в неравенство, определяющее доверительную область, получим:

$$40 \cdot \begin{pmatrix} -0,063 - m_1 & -1,218 - m_2 \end{pmatrix} \begin{pmatrix} 1,497 & 1,569 \\ 1,569 & 4,917 \end{pmatrix}^{-1} \begin{pmatrix} -0,063 - m_1 \\ -1,218 - m_2 \end{pmatrix} \leq 5,025$$

или:

$$40,147 m_1^2 - 25,622 m_1 m_2 + 12,223 m_2^2 - 26,149 m_1 + 28,161 m_2 \leq 11,301.$$

Полученная область изображена на рисунке 1 (область заключена внутри красного эллипса).

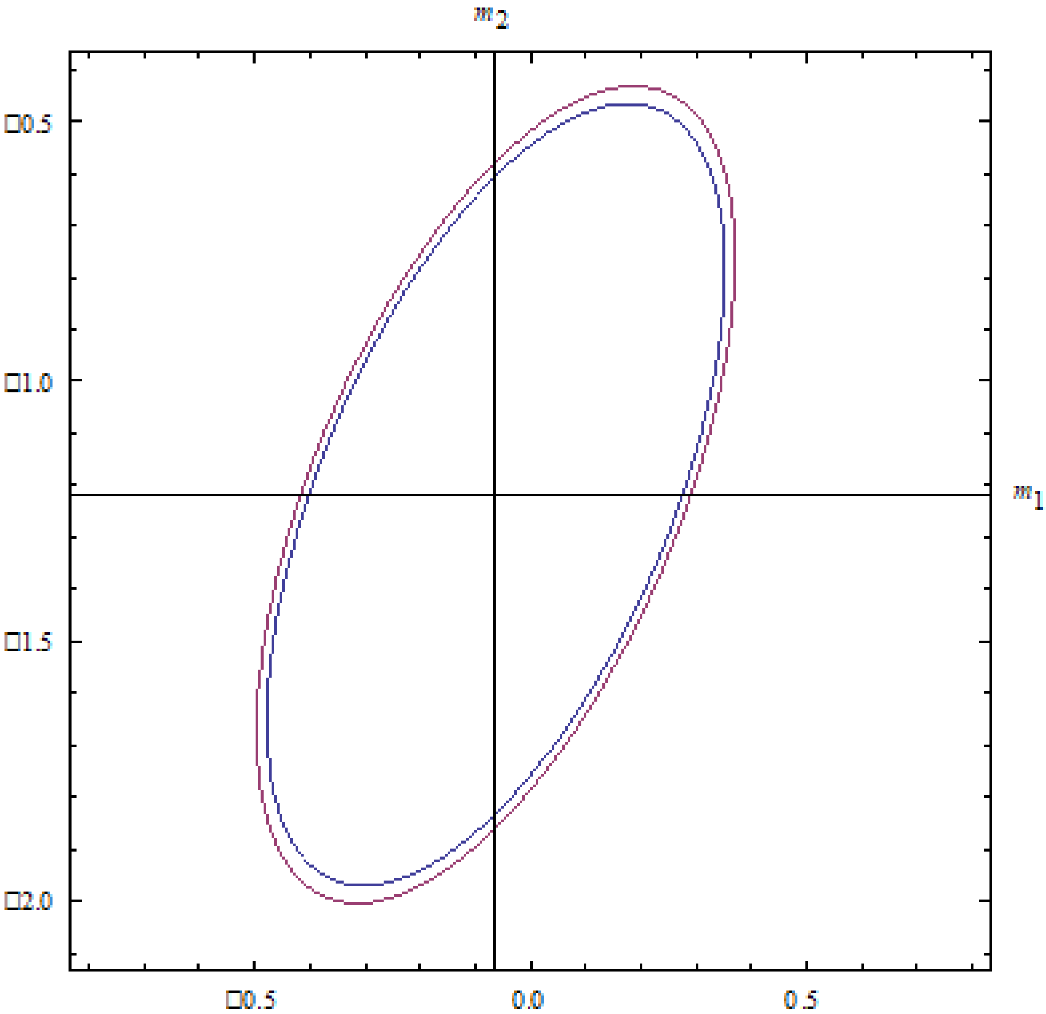

Рис. 1: Доверительная область для математических жиданий двумерного случайного вектора при известной матрице ковариаций и при неизвестной

2.3 Критерии для проверки гипотез о параметрах многомерной нормальной случайной величины на основе отношения правдоподобия

Гипотеза о равенстве матриц ковариаций

Пусть имеются выборки $X_1, X_2, \ldots, X_q$ объемов n_1, n_2, $\ldots$,n_q соответственно из p-мерных нормальных совокупностей $N(\vec{\mu}_1, \Sigma_1)$, $N(\vec{\mu}_2, \Sigma_2)$, $\ldots$, $N(\vec{\mu}_q, \Sigma_q)$. Проверяется гипотеза H_0 : $\Sigma_1 = \Sigma_2 = \ldots = \Sigma_q$.

Отношение правдоподобия в этом случае будет равно:

$$\lambda = \frac{\prod_{i=1}^{q} |S_i|^{\frac{n_i-1}{2}}}{|S|^{\frac{n-q}{2}}},$$

где $S = \frac{1}{n-q} \sum_{i=1}^{q} (n_i - 1) S_i$, а S_i ОМП ковариационной матрицы для i-ой выборки.