Глава 6. Многомерные гауссовские статистические линейные модели.

План этой главы таков.

В $1 напоминается о классических результатах, касающихся оценивания параметров распределения и проверки линейных гипотез, для обычной гауссовской модели из n одномерных наблюдений; также делается небольшое введение в методы построения аналогичной, но многомерной модели. Этой последней задаче, собственно, и посвящена вся настоящая глава.

В $2 на пространстве $\mathbb{R}_n^p$ матриц размера $p \times n$ вводятся некоторые простые, но важные линейные структуры; фактически там строится своя как бы линейная алгебра, аналогичная всем известной линейной алгебре на n-мерном действительном линейном пространстве. Интересно, что $\mathbb{R}_n^p$ оказывается не $np-$мерным пространством, как следовало бы ожидать, а, в некотором смысле, n-мерным!

В $3 вводится дополнительная конструкция - так называемая скобка Тюрина на $\mathbb{R}_n^p$, аналог скалярного произведения в $\mathbb{R}^n$, и определяются понятия ортогональности и ортогональной проекции.

В $4 уже на базе построенного аналога линейной алгебры в $\mathbb{R}_n^p$ рассматриваются важные статистические модели, очень похожие на одномерные. Строго формулируются две задачи, аналогичные двум одномерным задачам - несмещенное оценивание математического ожидания и матрицы ковариаций, а также проверка линейных гипотез. Напоминается понятие статистики Уишарта из главы 2 - многомерного аналога распределения χ^2.

В $5 мы уже непосредственно приступаем к решению этих задач. Этот параграф посвящен первой из них - оцениванию. Основой для разрешения этой задачи станет лемма об ортогональном разложении, почти такая же, как и в классическом одномерном случае.

В $6, заключительном, мы прикасаемся к, пожалуй, самому сложному и мало изученному разделу этой главы - проверке гипотез. Вводятся некоторые статистики, отдаленно напоминающие об известной статистике - дроби с распределением Фишера из одномерного случая, и формулируются искомые правило проверки гипотез. Но ситуация здесь далеко не так прозрачна, как в одномерном случае.

$1. Предварительные замечания.

Напомним, в чем состоит классическая одномерная гауссовская линейная статистическая модель. Пусть $X = (X_1, \ldots, X_n)^T \backsim \mathcal{N}_n(l, \sigma^2 I_n)$ - n-мерный случайный вектор наблюдений, где $\sigma > 0$ - неизвестное среднеквадратическое отклонение, $l = \mathbf{E}X$ - неизвестный n-мерный вектор - столбец; единственное, что нам известно про него - он принимает значения в некотором фиксированном линейном подпространстве $L \subset \mathbb{R}^n$, не совпадающем со всем пространством $\mathbb{R}^n$. (А это подпространство L нам известно.)

Эти наблюдения $X_k,\ k = \overline{1,n}$ некоррелированы, т.к. $\operatorname{Var} X = \sigma^2 I_n$ - диагональная матрица. Но они имеют совместное гауссовское распределение - ведь вектор X распределен по нормальному закону - а, значит, по утверждению 1.13 главы 1 они независимы. Наблюдение $X_k,\ k = \overline{1,n}$ выбирается из нормального закона распределения $N(l_k, \sigma^2)$, где l_k есть k-я компонента вектора l. (Это последнее утверждение следует из упражнения 1.6 в конце главы 1.) Таким образом, случайные величины, входящие в вектор X, имеют, вообще говоря, разные математические ожидания $l_k,\ k = \overline{1,n}$, но одну и ту же дисперсию σ^2.

Или, более формально: пусть $\mathfrak{X} := \mathbb{R}^n,\ \mathfrak{F} := \mathfrak{B}(\mathbb{R}^n)$ - борелевская σ-алгебра на $\mathbb{R}^n$, $\Theta := \{(l, \sigma) | l \in L, \sigma > 0\}$, и для $\theta = (l, \sigma) \in \Theta\ \ \mathbf{P}_\theta = \mathcal{N}_n(l, \sigma^2 I_n)$ - вероятностная мера на $\mathfrak{F}$. Тогда имеем статистическое пространство, базовое для этой модели:

$$(\mathfrak{X}, \mathfrak{F}, \{\mathbf{P}_\theta | \theta \in \Theta\}) = (\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), \{\mathcal{N}_n(l, \sigma^2 I_n) | \sigma > 0, l \in L\}).$$

О том, что такое статистическое пространство, см. \$1 главы 2, определение 1.

Мы можем задать случайные величины - наблюдения X_k на данном статистическом пространстве следующим образом: $X_k(x) := x_k$ при $x = (x_1, \ldots, x_n)^T \in \mathfrak{X} = \mathbb{R}^n$. Тогда случайный вектор наблюдений будет задан так: $X(x) := x$. В теории вероятностей, когда случайная функция есть тождественное отображение, ее называют **непосредственно заданной**. Например, если вам известно доказательство теоремы Колмогорова, то вы знаете, что там случайный процесс строится также как непосредственно заданный. Таким образом, наш случайный вектор непосредственно задан.

Напомним еще известную задачу **одномерной гауссовской линейной регрессии**.

Пусть мы проводим n независимых экспериментов, причем в $k-$м эксперименте вводим исходные данные $X_{kj},\ j = \overline{1, m}$ (будем называть их **факторами**), где m есть число факторов, оно одно и то же для каждого эксперимента. Пусть в k-м эксперименте мы фиксируем результат - действительное число Y_k, называемое **откликом**. При этом нам заведомо известно, что отклик, с точностью до случайной **несистематической** (т.е. с нулевым математическим ожиданием) ошибки, зависит от факторов линейно, с некоторыми неизвестными нам коэффициентами. Т.е., записывая это в виде формул, имеем:

$$Y_k = \sum_{j=1}^{m} a_j X_{kj} + \varepsilon_k$$

при всех $k = \overline{1, n}$, где $a_j,\ j = \overline{1, m}$ - вышеупомянутые неизвестные коэффициенты, $\varepsilon_k \backsim \mathcal{N}(0, \sigma^2),\ k = \overline{1, n}$ - независимые одинаково распределенные случайные величины, причем $\sigma > 0$ нам не известно. Такая регрессия называется одномерной в силу того, что отклик одномерен; линейной, т.к. отклик зависит от факторов линейно с точностью до случайной ошибки ε_k; гауссовской, т.к. единственные случайные слагаемые в этих формулах - это ε_k, т.е. гауссовсие случайные величины.

Эта терминология хорошо согласована с определением регрессии и линейной регрессии, данных в \$6 главы 1. Напомним эти определения.

Регрессия случайной величины (или случайного вектора) X_1 относительно случайной величины (или случайного вектора) X_2 - это $\mathbf{E}(X_1|X_2) = f(X_2)$, где f - некоторая (борелевская) функция. Тогда $X_1 = f(X_2) + \varepsilon$, где ε - случайная величина (или случайный вектор) с $\mathbf{E}(\varepsilon|X_2) = 0$, называемая **случайной ошибкой**. Если f линейна, то регрессия называется **линейной**.

А в данном случае, если x - случайный m-мерный вектор - столбец факторов (в k-м эксперименте он принимал значение $(X_{k1}, \ldots, X_{km})$, $k = \overline{1, n}$), а Y - случайная величина - отклик (в k-м эксперименте он принимал значение Y_k, $k = \overline{1, n}$), то $Y = a^T X + \varepsilon, \varepsilon \backsim \mathcal{N}(0, \sigma^2)$. (Здесь $a := (a_1, \ldots, a_n)^T$.) Предполагая, что ошибка ε не зависит от X - от данных, подаваемых на вход, получаем: $\mathbf{E}(\varepsilon|X) = \mathbf{E}\varepsilon = 0$. Значит, $\mathbf{E}(Y|X) = \mathbf{E}(a^T X + \varepsilon|X) = a^T X$ - линейная функция от X, т.е. имеем как раз линейную регрессию в смысле определения из главы 1.

Наша задача - оценить эти коэффициенты. На самом деле эта задача сводится к сформулированной выше. Действительно, при всех $k = \overline{1, n}$ случайная величина Y_k распределена по закону $\mathcal{N}(\sum\limits_{j=1}^{m} a_j X_{kj}, \sigma^2)$, и $Y_k, k = \overline{1, n}$ независимы. Отсюда получаем: случайный вектор $Y = (Y_1, \ldots, Y_n)^T$ имеет распределение $\mathcal{N}_n(Xa, \sigma^2 I_n)$, где через X мы обозначили матрицу размера $m \times n$, $kj-$й элемент которой равен X_{kj}. Т.е. $Y \backsim \mathcal{N}_n(l, \sigma^2 I_n)$, и нам известно про $n-$мерный вектор $l = Xa$ только то, что он лежит в линейном подпространстве $L \subset \mathbb{R}^n$, порожденном векторами $(X_{1j}, \ldots, X_{nj})^T, j = \overline{1, m}$.

Оценивание коэффициентов $a_j, j = \overline{1, m}$, равносильно оцениванию l, т.к. если ранг матрицы X равен m, мы можем однозначно разрешить систему уравнений $Xa =$

$\hat{l}$ (предварительно получив $\hat{l} \in L$ - оценку l) и получить решение этой системы - оценку вектора a. (Вопрос: почему при указанных условиях решение этой системы существует и единственно?) Также мы можем оценить и σ^2; как - изложено ниже.

Можно поставить следующие задачи:

1. Построить несмещенные оценки для l, σ^2.

2. Проверить при заданном уровне значимости $\alpha \in (0,1)$ линейную гипотезу $H : l \in L_0$, где $L_0 \subset L$ - заданное линейное подпространство, не совпадающее с L. Можно записать эту гипотезу по-другому: $\theta \in \Theta_0$, где $\Theta_0 := \{(l,\sigma) | l \in L_0, \sigma > 0\} = \{(l,\sigma) \in \Theta | l \in L_0\}$.

Эти задачи уже решены в математической статистике. Пусть далее $m := \dim L, m_0 := \dim L_0$. Напомним результаты:

1. $\hat{l} := \operatorname{proj}_L X$ - несмещенная оценка для l, и притом наилучшая и даже эффективная. Кроме того,

$$\hat{\sigma}^2 := \frac{1}{n-m} |\operatorname{proj}_{L^\perp} X|^2 \quad - \tag{6.1}$$

несмещенная оценка для дисперсии σ^2.

Замечание 6.1. Напомним, что если $x \in \mathbb{R}^n$, $M \subset \mathbb{R}^n$ - линейное подпространство, то $\operatorname{proj}_M x$ - это ортогональная проекция вектора x на подпространство M, а $M^\perp$ - это ортогональное дополнение к подпространству M.

2. Пусть задан уровень значимости $\alpha \in (0,1)$. Мы отвергаем гипотезу $H : l \in L_0$ на этом уровне, если и только если

$$T(X) = \frac{\frac{1}{m-m_0} |\operatorname{proj}_{L_1} X|^2}{\frac{1}{n-m} |\operatorname{proj}_{L^\perp} X|^2} > F_{1-\alpha}(m-m_0, n-m). \tag{6.2}$$

Здесь L_1 - ортогональное дополнение подпространства L_0 в пространстве L, т.е $L_1 \perp L_0$ и $L_1 \bigoplus L_0 = L$. А $F_\varepsilon(k,l)$ - это обозначение для $\varepsilon-$квантиля распределения Фишера (или, как еще говорят, распределения Снедекора) с k и l степенями свободы, где $\varepsilon \in (0;1), k, l \in \mathbb{N}$.

Замечание 6.2. В формулах (6.1), (6.2) знаменатели $n-m$, $m-m_0$ не обращаются в ноль, т.к. $L \neq \mathbb{R}^n$ и $m = \dim L < n$, $L_0 \neq L$ и $m_0 = \dim L_0 < \dim L = m$.

Это правило оправдано тем, что статистика T имеет распределение Фишера $F(m-m_0, n-m, \triangle)$ с $m-m_0$ и $n-m$ степенями свободы и параметром нецентральности $\triangle := \frac{1}{\sigma^2} |\operatorname{proj}_{L_1} l|^2$. А если гипотеза H верна, то $l \in L_0$ и в силу $L_1 \perp L_0$ имеем: $\operatorname{proj}_{L_1} l = 0$, т.е. $\triangle = 0$ и статистика T имеет центральное распределение Фишера $F(m-m_0, n-m)$. Т.е. для $\theta \in \Theta_0$ вероятность $\mathbf{P}_\theta$ отвергнуть гипотезу H (а при этих θ она как раз верна), т.е. совершить фатальную ошибку, называемую ошибкой 1 рода, равна α.

Можно ли обобщить эти важные и интересные результаты на случай, когда имеем не n одномерных наблюдений $X_1, \ldots, X_n$, а n многомерных, скажем, p-мерных наблюдений? Оказывается, можно. Этому вопросу и посвящена вся данная глава.

Но ключевую роль в рассмотренной выше одномерной модели играет ортогональная проекция - ведь именно с ее использованием построены все вышеприведенные статистики. Как обобщить понятие ортогональной проекции для случая уже не n-мерных векторов, а совокупности (или, иначе говоря, таблицы или матрицы) из n p-мерных наблюдений?

Поставим более общий вопрос. Как определить понятие ортогональности для этих матриц? Ведь чтобы знать, что есть ортогональная проекция, надо прежде понять, что же такое, собственно, ортогональность. В линейной алгебре ортогональность определяется через скалярное произведение, т.е. функцию, сопоставляющую двум векторам число из $\mathbb{R}$ или $\mathbb{C}$.

В той теории, которую мы построим, также будет присутствовать скалярное произведение на пространстве матриц, и две матрицы будут ортогональны, если их скалярное произведение равно нулю. Однако оно будет сопоставлять двум матрицам, как ни странно, не число, а другую матрицу. Такое скалярное произведение носит название **скобка Тюрина** и определяется в $3.

$2. Базовое пространство матриц. Подмодули. Каноническая биекция.

Пусть имеем n p-мерных наблюдений $X_1, \ldots, X_n \in \mathbb{R}^p$, $n, p \in \mathbb{N}$. Это - векторы - столбцы; запишем их в виде матрицы $p \times n$:

$$T = (X_1|X_2|\ldots|X_n),$$

т.е. пусть при $k = \overline{1,n}$ на месте $k-$го столбца матрицы T размера $p \times n$ будет стоять вектор X_k. Иногда пишут: $T = \{X_k,\ k = \overline{1,n}\}$. Так в многомерной статистике естественным образом возникает пространство матриц $p \times n$; будем обозначать его так: $\mathbb{R}^p_n$. В частности, будем обозначать пространство k-мерных строк через $\mathbb{R}^1_k$ - ведь такие строки есть на самом деле матрицы размера $1 \times k$. Пространство k-мерных столбцов будем обозначать, как и раньше, $\mathbb{R}^k$ (а не $\mathbb{R}^k_1$).

Какие же операции можно производить над матрицами из $\mathbb{R}^p_n$?

1. Сложение, умножение на число.

Это - обычные операции над матрицами фиксированного размера $p \times n$; относительно них, как известно, $\mathbb{R}^p_n$ есть линейное $np-$мерное пространство над полем $\mathbb{R}$.

2. Умножение на матрицу $k \in \mathbb{R}^p_p$

Произведение квадратной матрицы k порядка p и матрицы $T \in \mathbb{R}^p_n$ размера $p \times n$ дает матрицу kT, которая также имеет размер $p \times n$, т.е. $kT \in \mathbb{R}^p_n$. Если снабдить $\mathbb{R}^p_n$ таким произведением, получим модуль над кольцом $\mathbb{R}^p_p$. (Доказать самостоятельно.)

Напомним используемые нами понятия из высшей алгебры.

Определение 6.3. Если K - ассоциативное кольцо с единицей 1_K, а $(V, +)$ - абелева группа, и задано отображение $(x, v) \mapsto xv$ $(x \in K, v \in V)$ из $K \times V$ в V, удовлетворяющее условиям

$$x(u + v) = xu + xv,\ (x + y)u = xu + yu,\ (xy)u = x(yu),\ 1_K \cdot v = v$$

для всех $x, y \in K,\ u, v \in V$, то V называется **левым K-модулем или левым модулем над кольцом** K. А операция $(x, v) \mapsto xv$ называется **умножением.** Понятие модуля над кольцом аналогично понятию линейного пространства над полем и является его обобщением. См., например, книгу [5], глава 4, $ 3. Всюду ниже, говоря: кольцо K, мы будем иметь в виду ассоциативное кольцо с единицей, обозначаемой 1_K.

Определение 6.4. Подмодуль модуля V - это такое подмножество $U \subset V$, являющееся подгруппой $(V, +)$ (т.е. при $u, v \in U$ $u + v,\ -u \in U$), что при $x \in K,\ u \in U$ $xu \in U$. Говоря по-другому - это подмножество V, замкнутое относительно операций сложения и умножения слева на любой элемент кольца.

Утверждение 6.5. Любой подмодуль $\mathcal{L}$ модуля $\mathbb{R}^p_n$ имеет следующую структуру: найдутся такое $m \in \mathbb{N}$ и такие вектора $e_k \in \mathbb{R}^1_n, k = \overline{1,m}$, что

$$\mathcal{L} = \left\{ \sum_{k=1}^{m} \alpha_k e_k \,\middle|\, \alpha_k \in \mathbb{R}^p, k = \overline{1,m} \right\}. \tag{6.3}$$

И обратно, множество матриц из $\mathbb{R}^p_n$, заданное формулой (6.1), есть подмодуль модуля $\mathbb{R}^p_n$.

□ Второе утверждение проверьте самостоятельно - оно очень простое. Докажем первое, более содержательное утверждение. Пусть $\mathcal{L} \subset \mathbb{R}^p_n$ - подмодуль $\mathbb{R}^p_n$. Докажем следующее: найдутся матрицы $T_l \in \mathcal{L}, l = \overline{1,m}$, для которых

$$\mathcal{L} = \left\{ \sum_{l=1}^{m} k_l T_l \middle| k_l \in \mathbb{R}^p_p, l = \overline{1,m} \right\}. \tag{6.4}$$

Действительно, если $\mathcal{L}$ - нулевой подмодуль, т.е. $\mathcal{L} = \{0\}$, то достаточно положить $m = 1, T_1 = 0$. В противном случае найдется матрица $T_1 \in \mathcal{L}, T_1 \neq 0$. Если теперь $\mathcal{L} = \{k_1 T_1 | k_1 \in \mathbb{R}^p_p\}$, то все доказано; иначе существует матрица $T_2 \in \mathcal{L}, T_2 \notin \{k_1 T_1 | k_1 \in \mathbb{R}^p_p\}$. Далее, если $\mathcal{L} = \{k_1 T_1 + k_2 T_2 | k_1, k_2 \in \mathbb{R}^p_p\}$, то утверждение доказано. В противном случае найдется $T_3 \in \mathcal{L}, T_3 \notin \{k_1 T_1 + k_2 T_2 | k_1, k_2 \in \mathbb{R}^p_p\}$ и т.д.

На $m+1$ шаге ($m = 0, 1, 2, \ldots$) происходит следующее. Мы построили матрицы $T_1, \ldots, T_m$. При этом, заметим, они линейно независимы как элементы линейного пространства $\mathbb{R}^p_n$ над полем $\mathbb{R}$ - как это вытекает из условий $T_1 \neq 0$ и $T_l \notin \{k_1 T_1 + \ldots + k_{l-1} T_{l-1} | k_1, \ldots, k_{l-1} \in \mathbb{R}^p_p\}$ при $l = \overline{2,m}$? Если

$$\mathcal{L} = \left\{ \sum_{l=1}^{m} k_l T_l \middle| k_l \in \mathbb{R}^p_p, l = \overline{1,m} \right\},$$

то все доказано; если нет, то найдется матрица $T_{m+1} \in \mathcal{L}$, для которой

$$T_{m+1} \notin \left\{ \sum_{l=1}^{m} k_l T_l | k_l \in \mathbb{R}^p_p, l = \overline{1,m} \right\}.$$

Но бесконечно этот процесс продолжаться не может; он оборвется не позже, чем через np шагов. Докажем это; предположим противное - процесс совершил $np + 1$ шагов. Тогда мы построили матрицы $T_1, \ldots, T_{np+1}$, причем они линейно независимы - это отмечено выше. Но линейное пространство матриц $\mathbb{R}^p_n$ над полем $\mathbb{R}$ имеет размерность np, т.е. в нем могут существовать не более, чем np независимых матриц. Получили противоречие. Итак, на некотором шаге процесс построения матриц $T_1, T_2, \ldots$ оборвется, и мы получим представление (6.4).

На языке высшей алгебры мы только что доказали, что модуль $\mathcal{L}$ конечнопорожден, т.е. представляется в виде конечной суммы циклических подмодулей. Нам не понадобятся эти алгебраические понятия, но заинтересованного читателя отсылае к все той же книге [5], главе 4, \$3 или к любому другому достаточно полному учебнику по высшей алгебре.

Осталось теперь получить (6.3) из (6.4). Делаем это так: если при $s = \overline{1,p}, l = \overline{1,m}$ матрица $k_l \in \mathbb{R}^p_p$ имеет $\alpha_l^{(s)}$ в качестве $s-$го столбца, а T_l имеет $e_l^{(s)}$ в качестве $s-$й строки, то

$$\sum_{l=1}^{m} k_l T_l = \sum_{l=1}^{m} \sum_{s=1}^{p} \alpha_l^{(s)} e_l^{(s)}.$$

Но если матрицы $k_l \in \mathbb{R}^p_p$ пробегают все значения из $\mathbb{R}^p_p$, то их столбцы $\alpha_l^{(s)}$ пробегают все значения из $\mathbb{R}^p$. Это и дает нам искомое представление (6.3). ■

Определение 6.6. Подмодуль в формуле (6.3) называется **порожденным векторами** $e_k, k = \overline{1,m}$. Обозначение: $\mathcal{L} = \mathcal{L}(e_1, \ldots, e_m)$. Если вектора $e_k, k = \overline{1,m}$ линейно независимы, то их совокупность $\{e_k | k = \overline{1,m}\}$ называется **базисом** $\mathcal{L}$.

Лемма 6.7. Если $\mathcal{L}$ порожден базисом $\{e_k | k = \overline{1,m}\}$, то представление любого элемента $T \in \mathcal{L}$ в виде $T = \sum\limits_{k=1}^{m} \alpha_k e_k, \alpha_k \in \mathbb{R}^p, k = \overline{1,m}$ единственно. (Доказать самостоятельно.)

Замечание 6.8. Для подмодулей $U_k, k = \overline{1,m}$ модуля V над кольцом K, так же как и для линейных подпространств над полем, определены операции суммы и пересечения. А именно: $\bigcap\limits_{k=1}^{m} U_k$, т.е. пересечение всех подмодулей U_k, есть также подмодуль. (Это верно, кстати говоря, не только для конечного семейства подмодулей, но и для их семейства произвольной мощности - обоснуйте самостоятельно.) Далее, можно ввести понятие суммы подмодулей точно так же, как и суммы линейных подпространств:

Определение 6.9. Пусть $U_k, k = \overline{1,m}$ - подмодули модуля V над кольцом K. Тогда их **суммой** (обозначаемой $U_1+U_2+\ldots+U_m$ или $\sum\limits_{k=1}^{m} U_k$) называется множество $\{u_1 + \ldots + u_m | u_k \in U_k, k = \overline{1,m}\}$. И это - также подмодуль V. (Доказать самостоятельно.)

Операция взятия суммы подмодулей в $\mathbb{R}_n^p$ естественным образом согласована с представлением подмодулей в $\mathbb{R}_n^p$ в виде (6.1) следующим образом:

Лемма 6.10. Если $\mathcal{L}_k, k = \overline{1,m}$ - подмодули модуля $\mathbb{R}_n^p$, причем модуль $\mathcal{L}_k$ порожден системой $E_k := \{e_{kj} | j = \overline{1,m_k}\} \subset \mathbb{R}_n^1$, то сумма $\sum\limits_{k=1}^{m} U_k$ порождается системой $\bigcup\limits_{k=1}^{m} E_k = \{e_{kj} | k = \overline{1,m}, j = \overline{1,m_k}\}$. (Предоставляется для самостоятельного обоснования.)

Сейчас мы введем очень важное понятие, которое сразу во многом прояснит структуру модуля $\mathbb{R}_n^p$. Оно будет лежать в основе всех дальнейших построений.

Определение 6.11. Каноническая биекция - это соответствие между подмодулями в $\mathbb{R}_n^p$ и линейными подпространствами в $\mathbb{R}_n^1$, определяемое следующим образом: любой подмодуль в $\mathbb{R}_n^p$ имеет вид $\mathcal{L}(e_1, \ldots, e_m), e_k \in \mathbb{R}_n^1, k = \overline{1,m}$, и ему сопоставляется линейное пространство $< e_1, \ldots, e_m >$, порожденное векторами - строками $e_k, k = \overline{1,m}$.

Теорема 6.12 (основная). 1. Такое соответствие определено корректно, т.е. если $\mathcal{L}(e_1, \ldots, e_m) = \mathcal{L}(e'_1, \ldots, e'_l), e_k, e'_q \in \mathbb{R}_n^1, k = \overline{1,m}, q = \overline{1,l}$, то $< e_1, \ldots, e_m >=< e'_1, \ldots, e'_l >$. Т.е., проще говоря, если две различных конечных системы векторов - строк из $\mathbb{R}_n^1$ порождают один и тот же подмодуль из $\mathcal{L}$, то они порождают одно и то же линейное подпространство в $\mathbb{R}_n^1$.

(Обозначение $< e_1, \ldots, e_m >$ употребляется для линейной оболочки векторов $e_k, k = \overline{1,m}$.)

2. Это соответствие - действительно биекция. (Иначе бы название **каноническая биекция** не было оправдано.)

3. Это соответствие сохраняет отношение вложения: если L_1, L_2 - линейные подпространства $\mathbb{R}_n^1$, а $\mathcal{L}_1, \mathcal{L}_2$ - соответствующие им подмодули в $\mathbb{R}_n^p$, то $L_1 \subset L_2$ тогда и только тогда, когда $\mathcal{L}_1 \subset \mathcal{L}_2$. (Иначе говоря, совокупность линейных подпространств в $\mathbb{R}_n^1$ и совокупность подмодулей в $\mathbb{R}_n^p$ образуют множества, частично упорядоченные по включению, и построенная биекция осуществляет изоморфизм этих частично упорядоченных множеств.)

4. Модулю $\mathbb{R}_n^p$, рассматриваемому как подмодуль самого себя, соответствует пространство $\mathbb{R}_n^1$, рассматриваемое как подпространство самого себя. А нулевому подмодулю соответствует нулевое подпространство.

5. Каноническая биекция сохраняет сумму и пересечение подмодулей и линейных подпространств. Более точно: если $L_k, k = \overline{1,m}$ - линейные подпространства в $\mathbb{R}_n^1$, а $\mathcal{L}_k, k = \overline{1,m}$ - соответствующие им подмодули в $\mathbb{R}_n^p$, то пересечению $\bigcap\limits_{k=1}^{m} L_k$ соответствует пересечение $\bigcap\limits_{k=1}^{m} \mathcal{L}_k$, а сумме $\sum\limits_{k=1}^{m} L_k$ соответствует сумма $\sum\limits_{k=1}^{m} \mathcal{L}_k$.

6. У любого ненулевого подмодуля найдется (непустой) базис. (Доказательство всех шести пунктов предоставляется читателю.)

Замечание 6.13. Что следует из этой теоремы? Мы научились отождествлять подмодули в $\mathbb{R}_n^p$ и линейные подпространства в $\mathbb{R}_n^1$. Но структура $\mathbb{R}_n^1$ - самого обычного арифметического пространства n-мерных строк - великолепно изучена и разработана в линейной алгебре. А, значит, мы можем переносить идеи, методы и результаты уже построенной теории подпространств в $\mathbb{R}_n^1$ на наш основной объект рассмотрения - модуль $\mathbb{R}_n^p$.

И такой подход окажется в дальнейшем очень плодотворным - он позволит нам в \$3 развить в духе классической линейной алгебры структуры в $\mathbb{R}_n^p$ настолько, что потом мы уже сможем в \$4 проводить теоретико - вероятностные и статистические построения в $\mathbb{R}_n^p$.

Кстати, каноническая биекция приведет нас к удивительному результату: окажется, что линейных моделей в $\mathbb{R}_n^p$ в некотором смысле столько же, сколько и в $\mathbb{R}_n^1$. Изначально это совершенно не очевидно - ведь пространство $\mathbb{R}_n^p$ гораздо шире пространства $\mathbb{R}_n^1$, так что, казалось бы, гауссовский статистический анализ в первом пространстве должен быть богаче, чем уже известный нам гауссовский статистический анализ во втором!

Замечание 6.14. Эту каноническую биекцию можно построить и по-другому. Зафиксируем вектор - строку $z \in \mathbb{R}_p^1, z \neq 0$. Сопоставим каждой матрице $T \in \mathbb{R}_n^p$ $n-$мерную вектор - строку zT. Получили отображение $\mathbb{R}_n^p \to \mathbb{R}_n^1$. Это не биекция (почему?), но оно индуцирует естественным образом соответствие, при котором подмодулю $\mathcal{L}$ из $\mathbb{R}_n^p$ сопоставляется множество $z\mathcal{L} := \{zT | T \in \mathcal{L}\}$. Это последнее множество есть линейное подпространство в $\mathbb{R}_n^1$, и построенное соответствие на самом деле не зависит от z и совпадает с канонической биекцией, построенной в определении 6.10. (Доказать самостоятельно.)

Но нам требуется ввести еще важное понятие размерности подмодуля.

Определение 6.15. Размерность подмодуля $\mathcal{L} = \mathcal{L}(e_1, \ldots, e_m)$, порожденного своим базисом $\{e_k, k = \overline{1,m}\}$, по определению, равна m. Это определение годится только для ненулевых подмодулей; размерность нулевого, по определению, полагаем равной 0. Обозначение: $\dim \mathcal{L}$.

Теорема 6.16. 1) Это определение корректно, т.е. любые два базиса любого подмодуля состоят из одного и того же числа элементов.

2) Размерность подмодуля совпадает с размерностью линейного подпространства в $\mathbb{R}_n^1$, соответствующего этому подмодулю при канонической биекции.

3) Размерность любого подмодуля лежит в пределах от 0 до n, причем для подмодуля $\mathbb{R}_n^p \subset \mathbb{R}_n^p$ она равна n, для нулевого - нулю, а для остальных - одному из чисел $\overline{1, n-1}$. (Доказать самостоятельно.)

В качестве завершения данного параграфа введем еще простое понятие векторизации.

Определение 6.17. Если $T \in \mathbb{R}_n^p$, то **векторизацией матрицы** T называется $np-$мерный вектор - столбец (обозначаемый $\operatorname{vec} T$), построенный так: первые p его элементов образуют первый столбец матрицы T, следующие p его элементов - второй столбец T и т.д., последние (самые нижние) p элементов образуют n–й, последний столбец T. Более формально, при $i = \overline{1,p}, j = \overline{1,n}$ ij-й элемент матрицы T равен $p(j-1)+i$-му элементу столбца $\mathrm{vec}\,T$.

Лемма 6.18. Отображение $\mathrm{vec} : \mathbb{R}_n^p \to \mathbb{R}^{np}$ осуществляет изоморфизм линейных пространств $\mathbb{R}_n^p$ и $\mathbb{R}^{np}$. (Доказать самостоятельно.)

\$3. Скобка Тюрина. Ортогональное проецирование на подмодули.

Определение 6.19. Сопоставим двум матрицам $T, S \in \mathbb{R}_n^p$ матрицу $TS^T \in \mathbb{R}_p^p$ размера $p \times p$. Обозначим ее так: $< T, S >$ и назовем **скобкой Тюрина**. Матрицу $< T, T >$, где $T \in \mathbb{R}_n^p$, обозначим символом $|T|^2$.

Теорема 6.20. Эта операция обладает следующими свойствами:
1. $< T, S >=< S, T >^T$;
2. $< \alpha_1 T_1 + \alpha_2 T_2, S >= \alpha_1 < T_1, S > + \alpha_2 < T_2, S >$;
3. $< T, \alpha_1 S_1 + \alpha_2 S_2 >= \alpha_1 < T, S_1 > + \alpha_2 < T, S_2 >$;
4. $< T, 0 >=< 0, T >= 0$;
5. $< kT, S >= k < T, S >$;
6. $< T, kS >=< T, S > k^T$;
7. $< T, T >= |T|^2 \geq 0$;
8. $< T, T >= 0$ тогда и только тогда, когда $T = 0$;
9. $< TC, SC >=< T, S >$,

где $C \in \mathbb{R}_n^n$ - ортогональная матрица, $T, T_1, T_2, S, S_1, S_2 \in \mathbb{R}_n^p, \alpha_1, \alpha_2 \in \mathbb{R}, k \in \mathbb{R}_p^p, 0 \in \mathbb{R}_n^p$ - нулевая матрица, а $\geq$ в п. 7 означает, что матрица $< T, T >$ неотрицательно определена. (Доказать самостоятельно.)

Замечание 6.21. Как видим, скобка Тюрина есть аналог обычного скалярного произведения $(x, y) = xy^T = \sum_{k=1}^n x_k y_k$ векторов - строк $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}_n^1$. Ведь модуль над кольцом - это аналог линейного пространства над полем, и скалярное произведение на линейном пространстве - это билинейная функция, оба аргумента которой лежат в этом пространстве, а значение - в поле, в то время, как скобка Тюрина - это билинейная функция, аргументы которой лежат в модуле $\mathbb{R}_n^p$, а значение - в кольце $\mathbb{R}_p^p$. И свойства скобки Тюрина, перечисленные в теореме 6.20, очень похожи на свойства обычного скалярного произведения.

Для чего нам нужна скобка Тюрина? Оказывается, с ее помощью мы сможем ввести понятия ортогональности и ортогональной проекции.

Определение 6.22. Матрицы $T, S \in \mathbb{R}_n^p$ называются **ортогональными**, если $< T, S >= 0$. Пишут: $T \perp S$. Матрица $T \in \mathbb{R}_n^p$ и подмножество $\mathcal{L} \subset \mathbb{R}_n^p$ называются **ортогональными**, если $T \perp S$ при всех $S \in \mathcal{L}$. Пишут: $T \perp \mathcal{L}$. Подмножества $\mathcal{L}_1, \mathcal{L}_2 \subset \mathbb{R}_n^p$ называются **ортогональными**, если $T \perp S$ при всех $T \in \mathcal{L}_1, S \in \mathcal{L}_2$. Пишут: $\mathcal{L}_1 \perp \mathcal{L}_2$.

Лемма 6.23. (Теорема Пифагора.) Если $S, T \in \mathbb{R}_n^p$, $S \perp T$, то $|S + T|^2 = |S|^2 + |T|^2$. Более общо: если $\{T_k | k = \overline{1,m}\} \subset \mathbb{R}_n^p$ - конечная система попарно ортогональных матриц, то $\left|\sum\limits_{k=1}^m T_k\right|^2 = \sum\limits_{k=1}^m |T_k|^2$. (Доказать самостоятельно.)

Оказывается, только что введенное понятие ортогональности в модуле $\mathbb{R}_n^p$ и давно известное нам понятие ортогональности в обычном линейном пространстве $\mathbb{R}_n^1$ (относительно стандартного скалярного произведения $(x, y) := xy^T, x, y \in \mathbb{R}_n^1$) прекрасным образом согласованы, и притом через каноническую биекцию, введенную в \$2.

Утверждение 6.24. Подмодули $\mathcal{L}_1, \mathcal{L}_2$ из $\mathbb{R}_n^p$ ортогональны тогда и только тогда, когда ортогональны линейные подпространства L_1, L_2 из $\mathbb{R}_n^1$, соответствующие этим подмодулям при канонической биекции. (Обоснование этого предоставляется читателю.)

Определение 6.25. Базис $\{e_k | k = \overline{1,m}\}$ подмодуля $\mathcal{L}$ называется **ортогональным (ортонормированным)**, если эта система векторов - строк ортогональна (ортонормирована) в обычном смысле этого слова, т.е. $e_k e_l = \delta_{kl}$ при $k, l = \overline{1,m}$.

Лемма 6.26. В любом ненулевом подмодуле найдется ортонормированный базис. (Существование какого-то базиса, не обязательно ортонормированного, указано в п.6 основной теоремы 6.12.)

Определение 6.27. Сумма подмодулей $\mathcal{L}_k, k = \overline{1,m}$ из $\mathbb{R}_n^p$ называется **прямой** и обозначается так: $\mathcal{L}_1 \bigoplus \mathcal{L}_2 \bigoplus \ldots \bigoplus \mathcal{L}_m$, или так: $\bigoplus\limits_{k=1}^{m} \mathcal{L}_k$, если $\mathcal{L}_i \perp \mathcal{L}_j$ при $i, j = \overline{1,m}, i \neq j$.

Лемма 6.28. Если $\mathcal{L} = \bigoplus\limits_{k=1}^{m} \mathcal{L}_k$ - разложение подмодуля $\mathcal{L}$ в прямую сумму подмодулей $\mathcal{L}_k, k = \overline{1,m}$, а $E_k \subset \mathbb{R}_n^1$ - ортонормированный базис в $\mathcal{L}_k$, $k = \overline{1,m}$, то $\bigcup\limits_{k=1}^{m} E_k$ - ортонормированный базис $\mathcal{L}$. (Доказать самостоятельно.)

Утверждение 6.29. Если $\mathcal{L}_1, \mathcal{L}_2$ - подмодули в $\mathbb{R}_n^p$, причем $\mathcal{L}_2 \subset \mathcal{L}_1$, то существует и единственен такой подмодуль $\mathcal{L}_0$ в $\mathbb{R}_n^p$, что $\mathcal{L}_0 \subset \mathcal{L}_1$ и $\mathcal{L}_0 \bigoplus \mathcal{L}_2 = \mathcal{L}_1$. Этот подмодуль называется **ортогональным дополнением** подмодуля $\mathcal{L}_2$ в подмодуле $\mathcal{L}_1$. В частности, если положить $\mathcal{L}_2 = \mathcal{L}, \mathcal{L}_1 = \mathbb{R}_n^p$, получаем: существует и единственен такой подмодуль (будем обозначать его $\mathcal{L}^{\perp}$ и называть **ортогональным дополнением** подмодуля $\mathcal{L}$), что $\mathcal{L}^{\perp} \bigoplus \mathcal{L} = \mathbb{R}_n^p$. (Доказать самостоятельно.)

Лемма 6.30. Прямая сумма подмодулей переходит при канонической биекции в прямую сумму ортогональных линейных подпространств. (Доказать самостоятельно - впрочем, это очевидным образом следует из утверждений 6.24 и п.5 6.12.)

В этом параграфе нам осталось ввести последнее важное понятие, без которого немыслима многомерная гауссовская статистическая теория.

Определение 6.31. Проекцией матрицы $T \in \mathbb{R}_n^p$ **на подмодуль** $\mathcal{L} \subset \mathbb{R}_n^p$ называется такая матрица $S \in \mathcal{L}$, что $T - S \perp \mathcal{L}$.

Теорема 6.32. Пусть $T \in \mathbb{R}_n^p$, $\mathcal{L}$ - подмодуль $\mathbb{R}_n^p$. Тогда:

1. Проекция T на $\mathcal{L}$ существует и единственна. Будем в дальнейшем обозначать ее так: $\operatorname{proj}_{\mathcal{L}} T$. Если $\mathcal{L}$ - нулевой подмодуль, то эта проекция равна нулевой матрице. В противном случае у $\mathcal{L}$, согласно п.6 основной теоремы 6.12 найдется базис $e_k, k = \overline{1,m}$, и тогда проекцию можно в явном виде выразить следующей формулой:

$$\operatorname{proj}_{\mathcal{L}} T = TF^T(FF^T)^{-1}F,$$

где F - матрица размера $m \times n$, k-я строка которой равна e_k, $k = \overline{1,m}$. В частности, если $\mathcal{L} = \mathbb{R}_n^p$, то $\operatorname{proj}_{\mathcal{L}} T = T$.

2. Проекция - линейная операция, т.е.

$$\operatorname{proj}_{\mathcal{L}}(\alpha T + \beta S) = \alpha \operatorname{proj}_{\mathcal{L}} T + \beta \operatorname{proj}_{\mathcal{L}} S,$$

если $T, S \in \mathbb{R}_n^p, \alpha, \beta \in \mathbb{R}$, и, кроме того,

$$\operatorname{proj}_{\mathcal{L}}(kT) = k \operatorname{proj}_{\mathcal{L}} T,$$

если $T \in \mathbb{R}^p_n, k \in \mathbb{R}^p_p$. Т.е. proj есть эндоморфизм модуля $\mathbb{R}^p_n$. (**Гомоморфизм** левых модулей U, V над кольцом K есть такое отображение $f : U \to V$, что $f(u_1 + u_2) = f(u_1) + f(u_2), f(ku) = kf(u)$ при любых $u_1, u_2, u \in U, k \in K$. **Эндоморфизм** левого модуля V над кольцом K есть гомоморфизм V в V.)

3. Если $S \in \mathcal{L}$, то $|T - S|^2 \geq |T - \operatorname{proj}_{\mathcal{L}}|^2$, причем равенство достигается только при $S = \operatorname{proj}_{\mathcal{L}} T$. Иными словами, оператор из подмодуля $\mathcal{L}$ в пространство симметричных квадратных матриц размера $p \times p$, сопоставляющий матрице S матрицу $|T - S|^2$, достигает строгого глобального минимума при $S = \operatorname{proj}_{\mathcal{L}} T$. (Напомним, что мы умеем сравнивать симметричные квадратные матрицы из $\mathbb{R}^p_p$ - см. \$6 гл. 1.)

4. Если $T \in \mathbb{R}^p_n$, $\mathcal{L}$ - подмодуль $\mathbb{R}^p_n$, причем $T \perp \mathcal{L}$, то $\operatorname{proj}_{\mathcal{L}} T = 0$. В частности, если $\mathcal{L}_1, \mathcal{L}_2$ - подмодули $\mathbb{R}^p_n$, $\mathcal{L}_1 \perp \mathcal{L}_2$, $T \in \mathcal{L}_1$, то $\operatorname{proj}_{\mathcal{L}_2} T = 0$.

□ Докажем п. 1. Пп. 2, 3, 4, а также то утверждение, что если $\mathcal{L} = \mathbb{R}^p_n$, то $\operatorname{proj}_{\mathcal{L}} T = T$, предоставляются читателю.

Пусть пока $\mathcal{L}$ - ненулевой модуль. Он имеет базис $e_k, k = \overline{1, m}$ - см. п.6 основной теоремы 6.12. Значит, его структура может быть задана следующим образом: $S \in \mathcal{L}$ тогда и только тогда, когда найдутся $\alpha_k \in \mathbb{R}^p, k = \overline{1, m}$, для которых $S = \sum\limits_{k=1}^{m} \alpha_k e_k$, т.е. тогда и только тогда, когда существует матрица $A \in \mathbb{R}^p_m$ (составленная из столбцов α_k так: $k-$й столбец матрицы A равен α_k при $k = \overline{1, m}$), для которой $S = AF$. Итак, $\mathcal{L} = \{AF | A \in \mathbb{R}^p_m\}$.

Т.е. S - проекция T на $\mathcal{L}$, если и только если $S \in \mathcal{L}$ и $T - S \perp \mathcal{L}$, т.е. если найдется $A \in \mathbb{R}^p_m$, для которого $S = AF$, и притом

$$(T - AF)(BF)^T = (T - S)(BF)^T =< T - S, BF >= 0$$

при любой матрице $B \in \mathbb{R}^p_m$. Перепишем полученное равенство так: $(T - AF)F^T B^T = 0$. Это выполнено при всех $B \in \mathbb{R}^p_m$ тогда и только тогда, когда $(T - AF)F^T = 0$ (почему?), т.е. если и только если $TF^T = AFF^T$. Последнее равенство запишем в виде $TF^T(FF^T)^{-1} = A$. (Мы используем то, что FF^T - матрица Грама векторов - строк $e_k, k = \overline{1, m}$, и она невырождена в силу их линейной независимости, которая, свою очередь, следует из того, что они образуют базис $\mathcal{L}$.)

Значит, S - проекция T на $\mathcal{L}$ тогда и только тогда, когда $S = AF$, где $A = TF^T(FF^T)^{-1}$, т.е. тогда и только тогда, когда $S = TF^T(FF^T)^{-1}F$. Этим заодно доказано существование и единственность проекции.

Случай нулевого подмодуля $\mathcal{L}$ разобрать самостоятельно. ■

Замечание 6.33. Напомним, что если $x \in \mathbb{R}^l, L$ - ненулевое линейное подпространство $\mathbb{R}^l$ с базисом $e_k, k = \overline{1, s}$, а матрица $F \in \mathbb{R}^l_s$ имеет e_k в качестве k-го столбца, $k = \overline{1, s}$, то проекция x на L выражается в явном виде:

$$\operatorname{proj}_L x = F(F^T F)^{-1} F^T x.$$

Доказанная в теореме 6.32 формула есть многомерный аналог только что выписанной формулы. Кроме того, в одномерном случае проекция также существует и единственна, является линейным оператором $\mathbb{R}^l \to \mathbb{R}^l$ и минимизирует функционал $L \to \mathbb{R}$, $y \mapsto |y - x|^2$.

Утверждение 6.34. Пусть $\mathcal{L}$ - ненулевой подмодуль $\mathbb{R}^p_n$, не совпадающий со всем модулем $\mathbb{R}^p_n$. Тогда по лемме 6.27 ненулевые подмодули $\mathcal{L}, \mathcal{L}^\perp$ имеют ортонормированные базисы $\{e_k | k = \overline{1, m}\}, \{e_k | k = \overline{m+1, n}\}$. Их объединение - это ортонормированный базис $\{e_k | k = \overline{1, n}\}$ всего модуля $\mathbb{R}^p_n$. (Почему?) Если дана матрица $T \in \mathbb{R}^p_n$, то ее можно единственным образом (см. лемму 6.7) представить в виде

$T = \sum\limits_{k=1}^{n} \alpha_k e_k, \alpha_k \in \mathbb{R}^p$, и тогда ортогональная проекция T на $\mathcal{L}$ выражается так:

$$\operatorname{proj}_{\mathcal{L}} T = \sum_{k=1}^{m} \alpha_k e_k.$$

(Доказать самостоятельно.)

\$4. Статистические гауссовские модели на $\mathbb{R}^p_n$ - общие положения. Статистика Уишарта.

Теперь мы развили структуры на модуле $\mathbb{R}^p_n$, аналогичные структурам из классической линейной алгебры,в достаточной степени, чтобы перейти уже к статистическим моделям на этом модуле.

Формально построим базовое статистическое пространство. (Определение этого понятия, напомним, см. в начале \$1 главы 2.) Пусть $\mathfrak{X} := \mathbb{R}^p_n, \mathfrak{F} := \mathfrak{B}(\mathbb{R}^p_n)$ (борелевская $\sigma-$ алгебра вводится на $\mathbb{R}^p_n$ так же, как и на пространстве $\mathbb{R}^{np}$, отождествляемом с $\mathbb{R}^p_n$ путем естественного изоморфизма - векторизации, построенной в \$2). Пусть далее $\Theta := \{(T,\Sigma)|T \in \mathcal{L}, \Sigma \in \mathbb{R}^p_p, \Sigma > 0\}$, где $\mathcal{L}$ - фиксированный подмодуль $\mathbb{R}^p_n$, не совпадающий со всем модулем $\mathbb{R}^p_n$. А для $\theta = (T,\Sigma) \in \Theta$ вероятностная мера $\mathbf{P}_\theta$ определяется так: $\mathbf{P}_\theta := N_p(l_1,\Sigma) \times N_p(l_2,\Sigma) \times \ldots \times N_p(l_n,\Sigma) = N_{pn}(\operatorname{vec} T, \Sigma \bigotimes I_n)$, где $l_k - k-$й столбец матрицы T. (Доказать последнее равенство!)

Тогда $(\mathfrak{X}, \mathfrak{F}, \{\mathbf{P}_{theta} | \theta \in \Theta\})$ - статистическое пространство, являющееся основой для дальнейших рассмотрений.

Замечание 6.35. $A \bigotimes B$ для матрицы A размера $k_1 \times l_1$ и для матрицы B размера $k_2 \times l_2$ есть, по определению, матрица размера $k_1k_2 \times l_1l_2$, задаваемая как блочная матрица, состоящая из $k_1 \times l_1$ блоков размера $k_2 \times l_2$. При всех $i = \overline{1,k_1}, j = \overline{1,l_1}$ $ij-$й блок равен $a_{ij}B$, где a_{ij} есть ij-й элемент матрицы A. Такая матрица называется **кронекеровским произведением** матриц A и B.

Неформально говоря, мы делаем n p-мерных независимых наблюдений, причем $k-$е наблюдение ($k = \overline{1,n}$) взято из распределения $N_p(l_k,\Sigma)$. Независимость $X_1, \ldots, X_n$ следует из того, что $\operatorname{vec} X$ имеет распределение, являющееся прямым произведением вероятностных мер $\mathcal{N}_p(l_k,\Sigma)$ (где X - случайная матрица размера $p \times n$, k-й столбец которой равен X_k, $k = \overline{1,n}$.) Т.е. эти случайные векторы имеют, вообще говоря, разные математические ожидания, но одну и ту же матрицу ковариаций (не известную нам) Σ. Кроме этого, нам известно лишь, что матрица $(l_1|l_2|\ldots|l_n)$ лежит в фиксированном подмодуле $\mathcal{L}$. Ситуация, как видим, полностью аналогична одномерному случаю.

Мы можем задать эти наблюдения X_k как случайные вектора на построенном выше статистическом пространстве. А именно: при $k = \overline{1,n}$ $X_k(T) := T_k$, где T_k есть k-й столбец матрицы $T \in \mathbb{R}^p_n$. В таком случае имеем: случайная матрица $X = (X_1|\ldots|X_n)$ задается на нашем статистическом пространстве так: $X(T) := T$ при всех $T \in \mathbb{R}^p_n$. В этом случае - напомним \$1 данной главы - говорят, что случайная матрица X, в которую мы собрали все наши наблюдения, непосредственно задана.

Можно ввести в рассмотрение также **многомерную линейную гауссовскую регрессию**, полностью аналогичную той, которая обсуждалась в \$1 данной главы. Пусть мы, как и в \$1, проводим серию из n независимых экспериментов, в k-м эксперименте мы вводим как исходные данные m факторов $X_{kj}, j = \overline{1,m}$, а на выходе получаем уже не один, а p откликов (оттого и регрессия называется многомерной) $Y_{jk}, j = \overline{1,p}$. При этом нам известно, что зависимость вектора откликов от вектора факторов линейна, с точностью до случайной несистематической ошибки (т.е., напомним, ошибки, математическое ожидание которой равно 0). Или, на языке формул:

$$Y_{jk} = \sum_{q=1}^{m} a_{jq} X_{qk} + \varepsilon_{kj}$$

для всех $k = \overline{1,n}, j = \overline{1,p}$, где $a_{jq}, j = \overline{1,p}, q = \overline{1,m}$ - искомые коэффициенты линейной зависимости, ε_{kj} - случайные ошибки, причем вектора $\varepsilon_k = (\varepsilon_{k1}, \ldots, \varepsilon_{kp})^T \backsim \mathcal{N}_p(0, \Sigma)$ независимы и одинаково распределены. В матричном виде:

$$Y = AX + \varepsilon,$$

где X - матрица размера $m \times n$, qk-й элемент которой равен X_{qk}, Y - матрица размера $p \times n$, $jk-$й элемент которой равен Y_{jk}, ε - случайная матрица размера $p \times n$, k-й столбец которой равен ε_k. Наша задача - оценить матрицу $p \times m$ коэффициентов A, jq-й элемент которой равен $a_{jq}, j = \overline{1,p}, q = \overline{1,m}$. Совершенно аналогично одномерной регрессии из \$1 показывается, что мы получили частный случай многомерной модели, описанной в начале этого параграфа. А именно: случайная матрица Y такова, что ее столбцы независимы и l-й столбец имеет распределение $N_p(T_l, \Sigma)$, где T_l есть l-й столбец матрицы средних $T = AX$. Т.е., эквивалентно, $\mathrm{vec}\, Y \backsim \mathcal{N}_{np}(\mathrm{vec}\, T, \Sigma \bigoplus I_p)$, и нам известно про матрицу средних T только то, что она принадлежит подмодулю $\{BX | B \in \mathbb{R}^p_m\}$ модуля $\mathbb{R}^p_n$. Оценив матрицу T (как - см. ниже) и решив соответствующие системы линейных уравнений - их решение существует и единственно, если ранг матрицы X равен m - получим оценки параметров $a_{jq}, j = \overline{1,p}, q = \overline{1,m}$. (Проведите это последнее рассуждение подробнее.)

Поставим следующие задачи:

1. Построить несмещенные оценки для $T = \mathbf{E}X$ и Σ.

2. Проверить при данном уровне значимости линейную гипотезу $H : T \in \mathcal{L}_0$, где $\mathcal{L}_0 \subset \mathcal{L}$ - подмодуль $\mathbb{R}^p_n$. Формально запишем эту гипотезу в следующем виде: $\theta \in \Theta_0$, где $\Theta_0 \subset \Theta$, $Theta_0 := \{(T, \Sigma) | T \in \mathcal{L}_0, \Sigma \in \mathbb{R}^p_p, \Sigma > 0\} = \{(T, \Sigma) \in \Theta | T \in \mathcal{L}_0\}$.

Теперь нам полностью понятно замечание 6.13 об одинаковом числе линейных моделей в одномерном случае, т.е. в $\mathbb{R}^n$, и в многомерном случае, т.е. в $\mathbb{R}^p_n$. Ведь каждая линейная модель в $\mathbb{R}^p_n$ задается подмодулем $\mathcal{L}$ (или двумя подмодулями $\mathcal{L}, \mathcal{L}_0, \mathcal{L}_0 \subset \mathcal{L}$, если речь идет о проверке линейных гипотез). А линейная модель в $\mathbb{R}^n$ задается линейным подпространством L (или двумя подпространствами $L, L_0, L_0 \subset L$, если речь идет о проверке линейных гипотез). Отождествляя линейные подпространства и подмодули (и вспоминая то, что согласно п.3 основной теоремы 6.12 это отождествление сохраняет включение), получаем отождествление лиейных моделей в $\mathbb{R}^n$ и $\mathbb{R}^p_n$.

Отметим еще важный частный случай только что сформулированной линейной гипотезы. В модели многомерной гауссовской линейной регрессии можно поставить задачу о проверке гипотезы $H : a_{jq} = 0, j = \overline{1,p}, q = \overline{r+1,m}$. Формально говоря, эта гипотеза соответствует частному случаю линейной гипотезы $H : T \in \mathcal{L}_0$, сформулированной выше, где $\mathcal{L}_0 := \{AX | A = (a_{jq})_{j=\overline{1,p}}^{q=\overline{1,m}} \in \mathbb{R}^p_m, a_{jq} = 0, j = \overline{1,p}, q = \overline{r+1,m}\}$ - подмодуль в $\mathbb{R}^p_n$, лежащий в подмодуле $\mathcal{L} = \{AX | A \in \mathbb{R}^p_m\}$. (Проверьте то, что $\mathcal{L}_0$ - действительно подмодуль!) Неформально это означает, что гипотеза выполнена тогда и только тогда, когда все коэффициенты при всех факторах, кроме r первых, равны нулю, т.е., попросту говоря, когда остальные факторы не влияют на результаты эксперимента.

Итак, сформулированные две задачи полностью аналогичны задачам из классической одномерной модели - см. \$1. Приступим к их решению. Напомним сначала важное понятие.

Определение 6.36. Центральная p-мерная статистика Уишарта (Wishart) с m степенями свободы и матрицей ковариаций Q (где $m, p \in \mathbb{N}$, Q - неотрицательная симметричная квадратная матрица порядка p - заданные параметры) по определению, задается так: если $\xi_k, k = \overline{1,m}$ - независимые одинаково распределенные случайные вектора с распределением $\mathcal{N}_p(0,Q)$, то

$$\mathcal{W}_p(m,Q) := \sum_{k=1}^{m} \xi_k \xi_k^T \ -$$

данная статистика. Иногда ее называют просто статистикой Уишарта с m степенями свободы и матрицей ковариаций Q. **Нецентральная p-мерная статистика Уишарта с m степенями свободы, матрицей ковариаций Q и параметром нецентральности $\triangle$** - это, по определению,

$$\mathcal{W}_p(m,Q,\triangle) := \sum_{k=1}^{m} (\xi_k + a_k)(\xi_k + a_k)^T,$$

где $a_k \in \mathbb{R}^p, k = \overline{1,m}$ - такие постоянные (неслучайные) вектора, что $\triangle = \sum\limits_{k=1}^{m} a_k a_k^T = AA^T =< A, A >, A = ||a_1|a_2|\ldots|a_m||$. Можно определить нецентральную статистику Уишарта и так:

$$\mathcal{W}_p(m,Q,\triangle) := \sum_{k=1}^{m} \zeta_k \zeta_k^T,$$

где $\zeta_k \backsim \mathcal{N}_p(a_k, Q)$ при $k = \overline{1,m}$ - независимые случайные векторы. (Почему эти два определения эквивалентны?)

Это понятие есть просто многомерный аналог распределения χ^2 с m степенями свободы, которое, кстати, также бывает центральным и нецентральным. Имеет место важное, но простое утверждение.

Лемма 6.37. Распределение случайной матрицы $\sum\limits_{k=1}^{m} (\xi_k + a_k)(\xi_k + a_k)^T$, где $\xi_k \backsim \mathcal{N}_p(0,Q), k = \overline{1,m}$, независимы и одинаково распределены, зависит только от величины $< A, A >= \sum\limits_{k=1}^{m} a_k a_k^T$.

(Докажите это самостоятельно. Это - опять-таки многомерный аналог того факта, что распределение $\sum\limits_{k=1}^{m} (\eta_k + b_k)^2$ при независимых одинаково распределенных $\eta_k \backsim \mathcal{N}(0,1)$ не зависит от самих $b_k \in \mathbb{R}$, а только от параметра нецентральности - суммы их квадратов $\sum\limits_{k=1}^{m} b_k^2$. Эти-то утверждения и делают осмысленным понятие параметра нецентральности и в одномерном, и в многомерном случаях.)

Лемма 6.38. 1. Имеет место равенство распределений: $\mathcal{W}_p(m,\Sigma) = \Sigma^{1/2}\mathcal{W}_p(m,I_p)\Sigma^{1/2}$.

2. Если $p > m$ или $\det\Sigma = 0$, то случайная матрица $W_p(m,\Sigma)$ вырождена с вероятностью 1 (но неотрицательно определена), а если $p \le n, \det\Sigma > 0$, то она невырождена и положительно определена почти наверное. (Доказать самостоятельно.)

Лемма 6.39. Математическое ожидание матрицы - статистики Уишарта $\mathcal{W}_p(m,Q)$ - есть mQ. (Доказать самостоятельно.)

5. Лемма об ортогональной проекции. Несмещенные оценки.

1. Напомним еще полезную лемму.

Лемма 2.13. Пусть $X_i, i = \overline{1,n}$ независимые гауссовские вектора, $X_i \backsim \mathcal{N}_p(\mu_i, Q), i = \overline{1,n}$. Пусть, далее, C - ортогональная матрица порядка n, $c_{\alpha\beta} - \alpha\beta$-й элемент матрицы C,

$$Y_\alpha = \sum_{\beta=1}^{n} c_{\alpha\beta} X_\beta$$

при $\alpha = \overline{1,n}$. Тогда:

(a) $(Y_1, \ldots, Y_n)^T$ - гауссовский вектор;

(b) $\mathbf{E} Y_\alpha = \sum\limits_{\beta=1}^{n} c_{\alpha\beta} \mathbf{E} X_\beta$;

(c) $\operatorname{cov}(Y_\alpha, Y_\beta) = \delta_{\alpha\beta} Q$ (следствие: при $\alpha \neq \beta$ Y_α, Y_β некоррелированы, а, значит, в силу утверждения 1.10 и п.(a) данного утверждения, независимы, а при $\alpha = \beta$ имеем: $\operatorname{Var} Y_\alpha = \operatorname{cov}(Y_\alpha, Y_\beta) = Q$);

(d) $\sum\limits_{i=1}^{n} X_i X_i^T = \sum\limits_{\alpha=1}^{n} Y_\alpha Y_\alpha^T$.

Напомним, она была сформулирована в главе 2, \$3. Переформулируем ее на языке матриц из $\mathbb{R}_n^p$. Соотношение (6.2) можно переписать в виде $Y = XC$, где X, Y - случайные матрицы размера $p \times n$, $k-$е столбцы которых равны X_k, Y_k соответственно, $k = \overline{1,n}$. Утверждения (a), (b), (d) леммы можно переписать так:

(a) $\operatorname{vec} Y$ - гауссовский ($np-$мерный) вектор;

(b) $\mathbf{E} Y = (\mathbf{E} X) C$;

(d) $|X|^2 =< X, X >= |Y|^2 =< Y, Y >$.

Теперь сформулируем и докажем одну общую лемму, из которой, как частный случай, будут следовать выражения для несмещенных оценок параметров T и Σ.

Лемма 6.40. (об ортогональном разложении) Пусть $m \geq 2$, $\mathbb{R}_n^p = \bigoplus\limits_{k=1}^{m} \mathcal{L}_k$ - прямая сумма ненулевых подмодулей $\mathcal{L}_k, k = \overline{1,m}$. Пусть X - случайная матрица размера $p \times n$, причем $\operatorname{vec} X \backsim N_{pn}(\operatorname{vec} T, \Sigma \bigotimes I_n)$. (Т.е., эквивалентно, столбцы матрицы X - независимые случайные вектора и $k-$й столбец распределен как $N_p(l_k, \Sigma)$, где l_k есть $k-$й столбец матрицы T. Докажите эту эквивалентность.)

Тогда $\operatorname{proj}_{\mathcal{L}_k} X, k = \overline{1,m}$ - независимые случайные матрицы, и их векторизации распределены по нормальному закону, причем $\mathbf{E} \operatorname{proj}_{\mathcal{L}_k} X = \operatorname{proj}_{\mathcal{L}_k} T$ для всех $k = \overline{1,m}$.

Кроме того,

$$|\operatorname{proj}_{\mathcal{L}_k} X|^2 \backsim \mathcal{W}_p(n_k, \Sigma, |\operatorname{proj}_{\mathcal{L}_k} T|^2),$$

где $n_k := \dim \mathcal{L}_k$, $k = \overline{1,m}$.

$\square$ Пусть $\{e_s^{(k)} | s = \overline{1,n_k}\}$ - ортонормированный базис в $\mathcal{L}_k$, $k = \overline{1,m}$. (Его можно выбрать по лемме 6.26.) Тогда объединение этих базисов, т.е. $\{e_s^{(k)} | s = \overline{1,n_k}, k = \overline{1,m}\}$ - ортонормированный базис в $\mathbb{R}_n^p$ по лемме 6.28. Т.е. для случайной матрицы X найдутся такие вектора - столбцы (функции на исходном измеримом пространстве, на котором задана сама X, например, на $\mathbb{R}_n^p$, если X непосредственно задана; позднее мы получим, что это - случайные вектора) $Y_s^{(k)}, k = \overline{1,m}, s = \overline{1,n_k}$, для которых

$$X = \sum_{k=1}^{m} \sum_{s=1}^{n_k} Y_s^{(k)} e_s^{(k)}.$$

Составим матрицу Y размера $p \times n$ из столбцов $Y_s^{(k)}$ - сначала запишем столбец $Y_1^{(1)}$, затем столбец $Y_2^{(1)}, \ldots, Y_{n_1}^{(1)}, Y_1^{(2)}, Y_2^{(2)}, \ldots, Y_{n_2}^{(2)}, \ldots, Y_1^{(m)}, Y_2^{(m)}, \ldots, Y_{n_m}^{(m)}$. Говоря более точно, пусть $Y_s^{(k)}$ составляет $\sum\limits_{j=1}^{k-1} m_j + s$-й столбец матрицы Y для $k = \overline{1,m}, s = \overline{1,n_k}$. Тогда легко проверяется, что $Y = XC$, где C - матрица (не случайная, а постоянная) размера $n \times n$, у которой $\sum\limits_{j=1}^{k-1} m_j + s$-я строка равна $e_s^{(k)}$. Т.к. X - случайная матрица, то и $Y = XC$ - случайная матрица, т.е. ее столбцы $Y_s^{(k)}$ - случайные векторы.

Пусть F - матрица $n_k \times n$, l-я строка которой равна $e_l^{(k)}$, $l = \overline{1, n_k}$. Тогда $\operatorname{proj}_{\mathcal{L}_k} X = XF^T(FF^T)^{-1}F$, $\operatorname{proj}_{\mathcal{L}_k} T = TF^T(FF^T)^{-1}F$ по теореме 6.32, а $\mathbf{E}X = T$. Отсюда сразу видим, что $\mathbf{E}\operatorname{proj}_{\mathcal{L}_k} X = \operatorname{proj}_{\mathcal{L}_k} T$. Далее, $\operatorname{proj}_{\mathcal{L}_k} X, k = \overline{1, m}$ независимы, т.к. $\operatorname{vec}(\operatorname{proj}_{\mathcal{L}_k} X) = \operatorname{vec}(\sum\limits_{s=1}^{n_k} Y_s^{(k)} e_s^{(k)})$ - используем утверждение 6.34, а $Y_s^{(k)}$ при $s = \overline{1, n_k}, k = \overline{1, m}$ независимы по лемме 2.13.

Нормальность распределения $\operatorname{proj}_{\mathcal{L}_k} X$ докажите, пожалуйста, самостоятельно.

Осталось обосновать, что $|\operatorname{proj}_{\mathcal{L}_k} X|^2 \backsim \mathcal{W}_p(n_k, \Sigma, |\operatorname{proj}_{\mathcal{L}_k} T|^2)$.

По лемме 6.34, $\operatorname{proj}_{\mathcal{L}_k} X = \sum\limits_{s=1}^{n_k} Y_s^{(k)} e_s^{(k)}$. Отсюда

$$|\operatorname{proj}_{\mathcal{L}_k} X|^2 = < \sum_{s=1}^{n_k} Y_s^{(k)} e_s^{(k)}, \sum_{s=1}^{n_k} Y_s^{(k)} e_s^{(k)} > = \sum_{s,q=1}^{n_k} < Y_s^{(k)} e_s^{(k)}, Y_q^{(k)} e_q^{(k)} > =$$

$$= \sum_{s,q=1}^{n_k} (Y_s^{(k)} e_s^{(k)})(Y_q^{(k)} e_q^{(k)})^T = \sum_{s,q=1}^{n_k} Y_s^{(k)} e_s^{(k)} (e_q^{(k)})^T Y_q^{(k)} = \sum_{s,q=1}^{n_k} Y_s^{(k)} (e_s^{(k)}, e_q^{(k)}) Y_q^{(k)} =$$

$$= \sum_{s,q=1}^{n_k} Y_s^{(k)} \delta_{sq} Y_q^{(k)},$$

где скобки $(\cdot, \cdot)$ обозначают стандартное скалярное произведение n-мерных векторов - строк. В последнем равенстве мы использовали ортонормированность базиса $\{e_s^{(k)} | s = \overline{1, n_k}\}$. Продолжая выкладки, имеем:

$$\sum_{s,q=1}^{n_k} Y_s^{(k)} \delta_{sq} Y_q^{(k)} = \sum_{s=1}^{n_k} Y_s^{(k)} Y_s^{(k)} \backsim \mathcal{W}_p(n_k, \Sigma, |\operatorname{proj}_{\mathcal{L}_k} T|^2),$$

т.к. по лемме 2.13 $\operatorname{cov} Y_s^{(k)} = \Sigma$, а параметр нецентральности данного распределения Уишарта равен (мы в этих выкладках опять будем использовать то, что $e_s^{(k)} (e_q^{(k)})^T = (e_s^{(k)}, e_q^{(k)}) = \delta_{sq}$)

$$\sum_{s=1}^{n_k} \mathbf{E}Y_s^{(k)} (\mathbf{E}Y_s^{(k)})^T = \sum_{s,q=1}^{n_k} \mathbf{E}Y_s^{(k)} e_s^{(k)} (e_q^{(k)})^T (\mathbf{E}Y_s^{(k)})^T = \sum_{s,q=1}^{n_k} (\mathbf{E}Y_s^{(k)} e_s^{(k)})(\mathbf{E}Y_q^{(k)} e_q^{(k)})^T =$$

$$= \sum_{s,q=1}^{n_k} < \mathbf{E}Y_s^{(k)} e_s^{(k)}, \mathbf{E}Y_q^{(k)} e_q^{(k)} > = < \sum_{s=1}^{n_k} \mathbf{E}Y_s^{(k)} e_s^{(k)}, \sum_{s=1}^{n_k} \mathbf{E}Y_s^{(k)} e_s^{(k)} > = |\sum_{s=1}^{n_k} \mathbf{E}Y_s^{(k)} e_s^{(k)}|^2.$$

А полученное выражение как раз равно $\mathbf{E}|\operatorname{proj}_{\mathcal{L}_k} X|^2$. Итак, это и означает, что $|\operatorname{proj}_{\mathcal{L}_k} X|^2 \backsim \mathcal{W}_p(n_k, \Sigma, |\operatorname{proj}_{\mathcal{L}_k} T|^2)$. ■

Следствие 6.41. В построенной выше линейной многомерной гауссовской модели $\operatorname{proj}_{\mathcal{L}} X$ есть несмещенная оценка для T, а $\frac{1}{n-m} |\operatorname{proj}_{\mathcal{L}^\perp} X|^2$ - несмещенная оценка для Σ.

(Это совсем несложно вывести из только что доказанной леммы - сделайте это самостоятельно. Подсказка: используйте утверждение п.4 теоремы 6.32.)

$6. Проверка линейных гипотез.

Будем решать задачу, сформулированную в начале $4 текущей главы в п.2.

Пусть $S_1 := |\operatorname{proj}_{\mathcal{L}_1} X|^2$, $S_2 := |\operatorname{proj}_{\mathcal{L}^\perp} X|^2$. Здесь $\mathcal{L}_1$ - ортогональное дополнение подмодуля $\mathcal{L}_0$ в подмодуле $\mathcal{L}$ - по лемме 6.29 такое существует. Матрицы S_1, S_2, как легко видеть - аналоги числителя и знаменателя (с точностью до умножения на константы) в статистике (6.2) для одномерной модели. Можно было бы рассмотреть статистику $S_1 S_2^{-1}$, но ее распределение при разных $\Sigma > 0$ (если $T \in \mathcal{L}_0$) разное. Как говорят, эта статистика не имеет **свободное распределение**; по-английски - **free distribution**.

Почему? Потому, что при $T \in \mathcal{L}_0$ $S_1 \backsim \mathcal{W}_p(m - m_0, \Sigma) = \Sigma^{1/2}\mathcal{W}_p(m - m_0, I_p)\Sigma^{1/2}$, а $S_2 \backsim \mathcal{W}_p(n - m, \Sigma) = \Sigma^{1/2}\mathcal{W}_p(n - m, I_p)\Sigma^{1/2}$ по п.1 леммы 6.38. Мы использовали то, что $\mathbf{E}X = T \in \mathcal{L}_0$, но $\mathcal{L}_0 \perp \mathcal{L}_1, \mathcal{L}_0 \perp \mathcal{L}^\perp$, т.е. верны равенства $\operatorname{proj}_{\mathcal{L}_1} T = \operatorname{proj}_{\mathcal{L}^\perp} T = 0$ по теореме 6.32. - а, значит, параметры нецентральности $|\operatorname{proj}_{\mathcal{L}_1} T|^2, |\operatorname{proj}_{\mathcal{L}^\perp} T|^2$ распределений Уишарта матриц S_1, S_2 равны нулю.

При этом S_1, S_2 независимы.

(Чтобы получить эти факты, достаточно рассмотреть разложение модуля $\mathbb{R}^p_n$ в прямую сумму $\mathcal{L}_0 \bigoplus \mathcal{L}_1 \bigoplus \mathcal{L}^\perp$ и применить лемму об ортогональном разложении.)

Отсюда получаем: $S_1 S_2{-1}$ имеет распределение

$$(\Sigma^{1/2}\mathcal{W}_p(m-m_0, I_p)\Sigma^{1/2})(\Sigma^{1/2}\mathcal{W}_p(m-m_0, I_p)\Sigma^{1/2})^{-1} = \Sigma^{1/2}\mathcal{W}_p(m-m_0, I_p)\mathcal{W}_p^{-1}(n-m, I_p)\Sigma^{-1/2}.$$

Итак, это распределение, к сожалению, зависит от Σ.

Замечание 6.42. Что такое $\Sigma^{-1/2}$? По определению, это $(\Sigma^{1/2})^{-1}$. Согласно п.3 упражнения 1.5 в конце главы 1, из невырожденности Σ следует невырожденность $\Sigma^{1/2}$, а, значит, существование обратной матрицы - $(\Sigma^{1/2})^{-1}$.

Замечание 6.43. Если $p > n - m$, то матрица S_2 вырождена почти наверное - см. лемму 6.38, п. 2, т.е. вышеприведенные выкладки не имеют смысла. Но если $p \le n - m$, то S_2 невырождена с вероятностью 1 (по той же лемме), и все эти выкладки верны.

Т.е. статистика $S_1 S_2^{-1}$ не годится для проверки гипотезы. Но попытаемся сконструировать из матриц S_1 и S_2 какие-нибудь статистики, которые уже имеют свободное распределение.

1. Статистика Хотеллинга. Проще всего заметить следующее: при разных $\Sigma > 0$ матрицы $S_1 S_2^{-1}$ имеют вид $\Sigma^{1/2}\mathcal{W}_p(m - m_0, I_p)\mathcal{W}_p^{-1}(n - m, I_p)\Sigma^{-1/2}$, т.е. они подобны той матрице $S_1 S_2^{-1}$, которая получается при $\Sigma = I_p$, а, значит, все они подобны друг другу. Отсюда сразу же получаем, что их след - один и тот же при всех Σ. Получаем статистику $T^2(X) = \operatorname{tr}(S_1 S_2^{-1})$.

Она обозначается T^2 - так же, как и статистика Хотеллинга (см. главу 2, конец \$2), потому что на самом деле можно доказать, что она совпадает со статистикой Хотеллинга. Распределение ее в случае, когда выполнена гипотеза, известно (оно, напомним, найдено самим Хотеллингом), и для проверки гипотезы мы можем поступить стандартным образом: рассмотреть квантиль этого распределения, соответствующий уровню $1 - \alpha$, $\alpha \in (0; 1)$, и отвергать гипотезу на уровне α тогда и только тогда, когда значение статистики T^2 больше этого квантиля.

2. Статистика Уилкса. Можно составить и другую статистику: при $n - m \ge p$ отношение

$$\Lambda = \frac{\det S_1}{\det(S_1 + S_2)}$$

не зависит от Σ (почему?) и называется **статистикой Уилкса (Wilks)**. При этом такая запись корректна, т.к. $\det(S_1 + S_2) > 0$ почти наверное - ведь $S_2 \ge 0$ и S_2 невырождена, т.е. $S_2 > 0$; далее, $S_1 \ge 0$, а, значит, $S_1 + S_2 > 0$ (почему?), $\det(S_1 + S_2) > 0$. С помощью нее опять-таки можно проверить данную гипотезу, поступая совершенно так же, как и со статистикой Хотеллинга.

3. Статистика Роя. Это статистика

$$V = \max_{u \in \mathbb{R}^p, u \neq 0} \frac{u^T S_2 u}{u^T S_1 u}.$$

Можно записать ее и в таком виде:

$$V = \max_{u \in \mathbb{R}^p, u^T S_1 u = 1} u^T S_2 u.$$

Мы предполагаем, что $m - m_0 \geq p$, т.к. тогда и только тогда матрица S_1 невырождена и положительно определена почти наверное, т.е. с вероятностью 1 $u^T S_1 u > 0$ при всех $u \in \mathbb{R}^p, u \neq 0$. Этот максимум конечен и достигается (почему? мы пользуемся тем, что функция $u \mapsto u^T S_2 u$ непрерывна на компакте $\{u \in \mathbb{R}^p | u^T S_1 u = 1\}$ ограничена и достигает максимума.) Кроме того, эта статистика, как и две предыдущих, не зависит от Σ - имеет свободное распределение. (Обосновать это.)

Т.е. и она годится в качестве инструмента для проверки нашей гипотезы.

4. Собственные значения. Так как матрицы $S_1 S_2^{-1}$ подобны при всех Σ, то их собственные значения $\lambda_k, k = \overline{1,p}$, остаются постоянными при изменении $\Sigma > 0$. Вектор, составленный из них, также является статистикой со свободным распределением. Эти собственные значения положительны с вероятностью 1, т.к. матрицы S_1, S_2, а, значит, и $S_1 S_2^{-1}$, положительно определены почти наверное - если только $n - m, m - m_0 \geq p$, см. лемму 6.38, п. 2.

Утверждение 6.44. (без доказательства). Пусть $\lambda_k, k = \overline{1,p}$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ - собственные значения для случайной матрицы $\mathcal{W}_p(\nu_1, I_p)\mathcal{W}_p^{-1}(\nu_2, I_p)$ при $\nu_1, \nu_2 \geq p$ - эти условия необходимы, чтобы эта запись имела смысл почти наверное. Тогда плотность совместного распределения $(\lambda_1, \ldots, \lambda_p)$ равна

$$f(\lambda_1, \ldots, \lambda_p) = \mathrm{const} \prod_{j=1}^{p} \lambda_j^{\frac{1}{2}(\nu_1 - p - 1)} \prod_{j=1}^{p} (1 + \lambda_j)^{-\frac{1}{2}(\nu_1 + \nu_2)} \prod_{1 \leq i < j \leq p} (\lambda_i - \lambda_j).$$

К сожалению, маргинальные распределения, т.е. распределения отдельных собственных значений λ_j, неизвестны.