Theorem 3.1. *Пусть дан p-мерный случайный вектор ξ, такой, что $E\xi = 0, E(\xi\xi^T) = \Sigma$. Тогда существует линейное ортоганальное преобразование $\eta = B^T\xi$, при котором $E\eta\eta^T = \Lambda$,*

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix},$$

где $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ - корни уравнения $|\Sigma - \lambda \cdot E| = 0$ и вектора $\beta^{(k)}$, образующие матрицу B удовлетворяют уравнению: $(\Sigma - \lambda \cdot E)\beta = 0$.

Доказательство. Пусть $b = (\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(p)})$ - матрица коэффициентов преобразования. Тогда систему уравнений $\Sigma\beta^{(k)} = \lambda_k\beta^{(k)},\ k = 1, ..., p$ можно записать в матричной форме: $\Sigma \cdot B = B \cdot \Lambda$, а условия нормировки ($\beta^{(k)T}\,\beta^{(k)} = 1$) и некоррелированности $\beta^{(k)T}\,\beta^{(i)} = 0,\ k \neq i$ можно переписать в виде $B^TB = E$. Тогда из этих двух соотношений имеем:

$$B^T\Sigma B = B^TB \cdot \Lambda = \Lambda.$$

Используя свойства определителя:

$$|\Sigma - \lambda \cdot E| = |B^T||\Sigma - \lambda \cdot E||B| = |B^T\Sigma B - \lambda B^TB| = |\Lambda - \lambda \cdot E| = \prod_{i=1}^{p}(\lambda_i - \lambda).$$

Находя из этого уравнения характеристические корни, получаем, что они являются диагональными элементами матрицы Λ. □

Theorem 3.2. *(Теорема о сохранении дисперсии) ортоганальное преобразование $\eta = B\xi$ случайного вектора ξ не меняет обобщенную дисперсию и сумму дисперсий компонент.*

Доказательство. Так как вектор математических ожиданий $E\xi = 0$, то и вектор математических ожиданий главных компонент $E\eta = 0$. Пусть Σ ковариационная матрица вектора исходных признаков, то есть $E\xi\xi^T = \Sigma$. Тогда ковариационная матрица вектора главных компонент:

$$E\eta\eta^T = E(B\xi(B\xi)^T) = E(B\xi\xi^TB^T) = BE(\xi\xi^T)B^T = B\Sigma B^T.$$

Тогда обобщенная дисперсия (определитель ковариационно матрицы вектора главных компонент) равна:

$$|B\Sigma B^T| = |B||\Sigma||B^T| = |\Sigma||BB^T| = |\Sigma|$$

Последнее равенство получается в силу условия некоррелированности и нормировки главных компонент.

Вычислим теперь суммарную дисперсию главных компонент:

$$\sum_{i=1}^{p} E\eta_i^2 = tr(B\Sigma B^T) = tr(\Sigma B \cdot B^T) = tr(\Sigma E) = tr(\Sigma) = \sum_{i=1}^{p} E\xi_i^2.$$

□

Выводы: Задача отыскания главных компонент сводится к задаче на собственные значения и собственные векторы матрицы ковариаций Σ. Коэффициенты преобразования $\beta^{(i)}$, $i = \overline{1,p}$ удовлетворяющие условиям задачи поиска главных компонент являются собственными векторами матрицы ковариаций Σ, соответствующими собственным значениям $\lambda_1, \lambda_2, ..., \lambda_p$, упорядоченным по убыванию. Причем

$$D\left(\eta_i\right) = \lambda_i$$

,

$$\sum_{i=1}^{p} D(\xi_i) = \sum_{i=1}^{p} D(\eta_i) = \sum_{i=1}^{p} \lambda_i.$$

Теперь необходимо разобрать, как находить главные компоненты на практике, если нам даны лишь наблюдения за вектором исходных признаков ξ. Ясно, что ковариационная матрица не задана, поэтому мы можем ее только оценить по данным наблюдениям за вектором исходных признаков ξ. Следующая теорема дает способ получения оценок для главных компонент и их дисперсий.

Theorem 3.3. *Пусть для вектора ξ получена выборка $X = \{(X_1^{(1)}, X_2^{(1)}, \ldots, X_p^{(1)}), (X_1^{(2)}, X_2^{(2)}, \ldots, X_p^{(2)}),$ объема n из p-мерной нормальной совокупности $N_p(\vec{\mu}, \Sigma)$. Тогда оценки ММП для дисперсий главных компонент являются корнями алгебраического уравнения: $|\widehat{\Sigma} - kE| = 0$, упорядоченными по убыванию, а соответствующие им собственные векторы $b^{(1)}, b^{(2)}, \ldots, b^{(p)}$ являются оценками векторов коэффициентов $\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(p)}$ и находятся из уравнения:*

$$\left(\widehat{\Sigma} - k_i E\right) \cdot b^{(i)} = 0,$$

где $b^{(i)T} b^{(i)} = 1$, $\widehat{\Sigma}$ – оценка ковариационной матрицы вектора ξ, полученная по ММП, $k_i, i = 1, ..., p$ – собственные значения матрицы $\widehat{\Sigma}$.

Примем данную теорему без доказательства (приведено в Андерсон "Введение в многомерный статистический анализ").

Таким образом, оценки главных компонент на основе выборочных данных строится с помощью выборочной матрицы ковариаций. Оценки собственных значений, являющиеся собственными числами выборочной матрицы ковариаций, в случае нормального распределения генеральной совокупности являются оценками максимального правдоподобия.

Замечание 3.4. *Если единицы измерения исходных признаков различаются или их значения сильно различаются, то лучше использовать при нахождении оценок главных компонент вместо выборочной матрицы ковариаций выборочную корреляционную матрицу.*

Какие из главных компонент теперь стоит оставить? Строгих математических критериев отбора главных компонент нет, существуют лишь эвристические методы. Рассмотрим их.

1) Зная $tr\,\bar{\Sigma} = \sum_{i=1}^{p} \lambda_i$, можно выбрать те компоненты η_i, $i = \overline{1, m}$ из общего набора, которые бы объясняли не менее некоторой заданной доли q **суммарной доли дисперсии признаков**. Обычно q полагают не менее 0,7.

2) **Критерий Кайзера**, который предполагает использование для нахождения оценок собственных значений выборочной матрицы корреляций (можно применять и в случае использования матрицы ковариаций, следует лишь в этом случае нормировать (умножить) все собственные значения на величину $p/tr(\widehat{\Sigma})$. Согласно данному критерию оставляют только те главные компоненты, дисперсия которых больше 1. По существу, это означает, что если фактор не выделяет дисперсию, эквивалентную, по крайней мере, дисперсии одной переменной, то он опускается.

3) Графический **критерий каменистой осыпи Кэттелла**. Критерий каменистой осыпи состоит в поиске точки, где убывание собственных значений замедляется наиболее сильно. Справа от этой точки должна находится, по-видимому, только "факторная осыпь"("осыпь-это геологический термин для обломков, которые скапливаются в нижней части каменистого склона). Таким образом, число выделенных факторов не должно превышать количество факторов слева от этой точки.

4) **Правило сломанной трости.** Набор нормированных собственных чисел ($k_i/tr\widehat{\Sigma}$, $i = \overline{1, p}$) сравнивается с распределением длин обломков трости единичной длины, сломанной в $p-1$ случайно выбранной точке (точки разлома выбираются независимо и равномерно распределенными по длине трости). Пусть l_i ($i = \overline{1, p}$) - длины полученных кусков трости, занумерованные в порядке убывания длины: $l_1 \geq l_2 \geq \ldots \geq l_p$. Тогда математическое ожидание l_i: $L_i = M(l_i) = \frac{1}{p}\sum_{j=i}^{p} \frac{1}{j}$. Правило сломанной трости: i-й собственный вектор (в порядке убывания собственных чисел λ_i) сохраняется в списке главных компонент, если $\frac{k_1}{tr\widehat{\Sigma}} > L_1,\ \frac{k_2}{tr\widehat{\Sigma}} > L_2,\ \ldots,\ \frac{k_i}{tr\widehat{\Sigma}} > L_i$.

Алгоритм применения метода главных компонент.