Объем куба радиуса r в R^k.

В случае четного k=2n $V_{2n}=\frac{\pi^n}{n!}r^{2n}$.

В случае нечетного k=2n+1 $V_{2n+1}=\frac{2(2\pi)^n}{(2n+1)!!}r^{2n+1}$

Анализ главных компонент

Рассмотрим случайный вектор $X_{1,}X_{2,}\ldots,X_k$

Задача 1. Найти $Y_1=a_{11}X_1+a_{12}X_2+\ldots+a_{1k}X_k$ такую, что $D(Y_1)$ максимальна.
Дополнительное условие 1: $\vec{a}_1\vec{a}_1^T=1$, где $\vec{a}_1=(a_{11,}a_{12,}\ldots,a_{1k})$

Задача 2. Найти $Y_2=a_{21}X_1+a_{22}X_2+\ldots+a_{2k}X_k$ такую, что $D(Y_2)$ максимальна.
Дополнительное условие 2.1: $\vec{a}_2\vec{a}_2^T=1$, где $\vec{a}_2=(a_{21,}a_{22,}\ldots,a_{2k})$
Дополнительное условие 2.2: $corr(Y_{2,}Y_1)=0$

Задача 3. Найти $Y_3=a_{31}X_1+a_{32}X_2+\ldots+a_{3k}X_k$ такую, что $D(Y_3)$ максимальна.
Дополнительное условие 3.1: $\vec{a}_3\vec{a}_3^T=1$, где $\vec{a}_3=(a_{31,}a_{32,}\ldots,a_{3k})$
Дополнительное условие 3.2: $corr(Y_{3,}Y_1)=0$
Дополнительное условие 3.3: $corr(Y_{3,}Y_2)=0$

Задача k. Найти $Y_k=a_{k1}X_1+a_{k2}X_2+\ldots+a_{kk}X_k$ такую, что $D(Y_k)$ максимальна.
Дополнительное условие k.1: $\vec{a}_k\vec{a}_k^T=1$, где $\vec{a}_k=(a_{k1,}a_{k2,}\ldots,a_{kk})$
Дополнительное условие k.2: $corr(Y_k,Y_1)=0$
Дополнительное условие k.3: $corr(Y_k,Y_2)=0$
…
Дополнительное условие k.k: $corr(Y_k,Y_{k-1})=0$

Обозначение
R – матрица ковариаций (корреляций) случайного вектора ***X***.

$$R\vec{a}=\lambda\vec{a}$$
$$D(Y_i)=\lambda$$

$$\sum_{i=1}^{k}\lambda_i=trace(\boldsymbol{R})$$

Факторный анализ

$\boldsymbol{X}^T = (X_{1,} X_{2,} \ldots , X_p)$

$\boldsymbol{F}^T = (F_{1,} F_{2,} \ldots F_k)$ - факторы (общие факторы, латентные факторы, factors, common factors)

$$k < p$$

$X_i = a_{i1} F_1 + a_{i2} F_2 + a_{ik} F_k + U_i \qquad i = 1, \ldots , p$

$\boldsymbol{X} = \boldsymbol{A} \boldsymbol{F} + \boldsymbol{U}$

где $\boldsymbol{A} = (a_{ij}) \qquad i = 1 \ldots p , j = 1 \ldots k$
$\boldsymbol{U}^T = (U_{1,} U_{2,} \ldots U_p)$

Дополнительные предположения

1 $\quad E\boldsymbol{X} = \boldsymbol{0}$
Не теряем общности, так как будем рассматривать корреляции.

2 $\quad E\boldsymbol{F} = \boldsymbol{0}$, $\quad D\boldsymbol{F} = \boldsymbol{I}$
Желаем, чтобы факторы объясняли разное, то есть были независимы, на самом деле достаточно некоррелированности…

3 Остатки, т.е элементы вектора ***U*,** не коррелируют друг с другом и с факторами.

<u>Определение.</u>
Элементы матрица А называются факторными нагрузками (factor loadings).
Элементы вектора ***U*** называются уникальными факторами (specific variates).

<u>Утверждение</u>

$D(X_i) = \sum_{j=1}^{k} a_{ij}{}^2 + D(u_i)$

$cov(X_i , X_j) = \sum_{m=1}^{k} a_{im} a_{jm}$

<u>Определения</u>

Общность (communality) $\quad h_i^2 = \sum_{j=1}^{k} a_{ij}^2$ - часть дисперсии переменной, объясненной факторами.

Уникальность (specific or unique variance) - часть дисперсии переменной, объясненной уникальными факторами.

Обозначение
R – матрица ковариаций (корреляций) случайного вектора ***X***.

Утверждение
$$\boldsymbol{R}=\boldsymbol{A}\,\boldsymbol{A}^T+diag\left(D\left(\boldsymbol{U}\right)\right)$$

Вопрос: где тут данные, наблюдения?
Ответ: пока их нет.

На практике вместо матрицы ***R*** используем выборочную матрицу ковариаций (корреляций)

Утверждение.
Решение (матрица А) определено с точностью до ортогонального преобразования.

Следствие. Среди всех решений (отличающихся поворотом) можно выбирать самое "симпатичное".

Плохая статистика, fishing

Вращение факторов (factor rotation)

Если не хочется вращать, хочется, чтобы было как в методе главных компонент, вводим условие:
матрица $\boldsymbol{A}^T diag\left(D\left(U_i\right)\right)\boldsymbol{A}$ да будет диагональной, элементы на главной диагонали расположены в убывающем порядке.

Если хочется вращать, идея метода varimax

$a^2+b^2\rightarrow max$ при условии, что $a+b=1$

• Each row or the factor loading matrix should contain at least one zero.
• Each column of the loading matrix should contain at least k zeros.
• Every pair of columns of the loading matrix should contain several variables whose loadings vanish in one column but not in the other.
• If the number of factors is four or more, every pair of columns should contain a large number of variables with zero loadings in both columns.
• Conversely, for every pair of columns of the loading matrix only a small number of variables should have non-zero loadings in both columns.