1　多次元正規分布

多次元正規分布は、正規分布の多次元版であり、多変量正規分布ともいわれる。

1.1　多次元標準正規分布

最初に、平均 0 で分散 1 の標準正規分布に従う互いに独立な確率変数ベクトルについて考えてみる。ちなみに「互いに独立な確率変数」であれば、以下のように同時密度関数が周辺密度の積に分解できる。

定義 1.1.　【確率変数ベクトルの独立】
n 個の確率変数 $z_1, z_2, \cdots, z_n$ が独立であるとは、任意の実数 $a_1, a_2, \cdots, a_n$ に対して以下の式が成立する事である。

$$P(z_1 < a_1, z_2 < a_2, \cdots, z_n < a_n) = P(z_1 < a_1)P(z_2 < a_2)\cdots P(z_n < a_n)$$

確率密度関数でいえば、$\boldsymbol{z}$ を確率変数ベクトルとした時、同時密度関数 $f_{\boldsymbol{Z}}(\boldsymbol{z})$ が、周辺密度関数の積 $g(z_1)g(z_2)\cdots g(z_n)$ に分解されるとき独立であるという。

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = g(z_1)g(z_2)\cdots g(z_n) \tag{1}$$

この多次元標準正規分布の確率密度関数は以下のようになる。

定義 1.2.　【多次元標準正規分布の確率密度関数】
$\boldsymbol{Z}$ を n 次元標準正規分布 $N(\boldsymbol{o}, \boldsymbol{I})$ に従う確率変数とするとき、以下のように確率変数ベクトルを $\boldsymbol{z}$、ゼロベクトルを $\boldsymbol{o}$、そして単位行列を I としたとき、

$$\boldsymbol{z} = \begin{pmatrix} z_1 \\ z_1 \\ \vdots \\ z_n \end{pmatrix}, \qquad \boldsymbol{o} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad \boldsymbol{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

確率密度関数は以下で表す事ができる。

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{\sqrt{2\pi}^n}\ e^{-\frac{1}{2}||\boldsymbol{z}||^2} \tag{2}$$

この $||\boldsymbol{z}||^2$ をベクトルの内積で表示すると

$$||\boldsymbol{z}||^2 = \left(\sqrt{z_1^2 + z_2^2 + \cdots + z_n^2}\ \right)^2 = \boldsymbol{z}^T\boldsymbol{z}$$

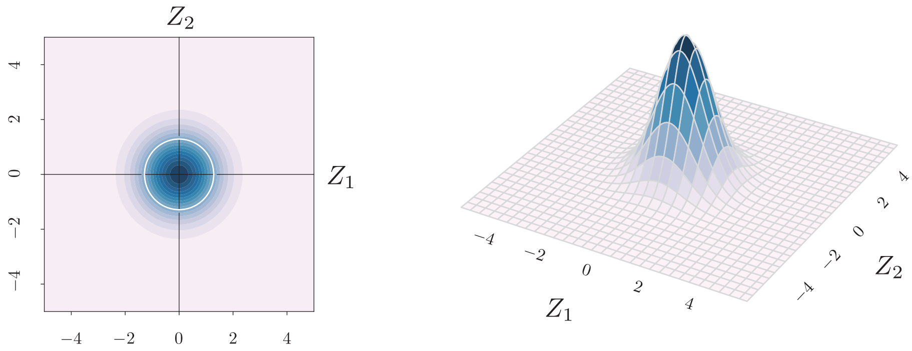

図 1　2 次元標準正規分布の確率密度関数

図 1 は二次元の標準正規分布の確率密度関数を表す。なお図 1 の左図の白丸は分散を表す。またこの図を描くプログラムをリスト 1 に示す。

この多次元標準正規分布 $N(\boldsymbol{o}, \boldsymbol{I})$ の期待値ベクトルと共分散行列は以下のようになる。これは各変数 $z_1, z_2, \cdots, z_n$ が平均 0 でばらつき 1 の標準正規分布に従う事から当然。また各変数は独立なので $Cov[Z_i, Z_j] = 0\ (i \neq j)$ となる。

$$E[\boldsymbol{Z}] = \begin{pmatrix} E[Z_1] \\ E[Z_2] \\ \vdots \\ E[Z_n] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \boldsymbol{o}$$

$$V[\boldsymbol{Z}] = \begin{pmatrix} V[Z_1] & Cov[Z_1, Z_2] & \cdots & Cov[Z_1, Z_n] \\ Cov[Z_2, Z_1] & V[Z_2] & \cdots & Cov[Z_2, Z_n] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[Z_n, Z_1] & Cov[Z_n, Z_2] & \cdots & V[Z_n] \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \boldsymbol{I}$$

ソースコード 1　2 次元標準正規分布の確率密度関数を描くプログラム

```
import numpy as np
import matplotlib.pyplot as plt

# データ列を作成する
a = 6;    sls = 0.1
x = np.arange(-a, a , sls)
y = np.arange(-a, a , sls)
X, Y = np.meshgrid(x, y)
Z = X.astype(np.float64)

# 一次変換した標準正規分布を計算する
mx = 2;    my = 1              # 平衡移動 x-mx y-my
Wt = np.deg2rad(0)             # 回転行列
Dr = np.array([[np.cos(Wt), -np.sin(Wt)], [np.sin(Wt), np.cos(Wt)]])
Dt = np.array([[1.2, 0], [0, 1.2]])           # 拡大縮小行列
D = np.linalg.inv(Dr @ Dt)
for i in range(Y[0].size):
    for j in range(X[0].size):
        Wk = D @ np.array([X[i, j] - mx, Y[i, j] - my])
        Z[i, j] = np.exp(-(Wk[0] ** 2 + Wk[1] ** 2) / 2) / (2 * np.pi)

# グラフ表示
fig = plt.figure()
ax = fig.add_subplot(122, projection='3d')
bx = fig.add_subplot(121)
x_min = -a ; x_max = a
y_min = -a ; y_max = a
ax.set_xlim(x_min, x_max) ; ax.set_ylim(y_min, y_max) ;    ax.set_zlim(0, 0.2)
bx.set_xlim(x_min, x_max) ; bx.set_ylim(y_min, y_max)
# 右の三次元サーフェス
ax.grid(False) ; ax.set_zticks([])
ax.plot_surface(X, Y, Z, cmap="PuBu" , rstride=4, cstride=4, edgecolor="lightgray" ,linewidth=0.5)
ax.set_xlabel("X", fontsize=9) ; ax.set_ylabel("Y", fontsize=9)
# 左の濃度マップ
bx.contourf(X, Y, Z,levels=15, cmap='PuBu' )
bx.plot([x_min,x_max],[0,0],color="black", linewidth=0.6)
bx.plot([0,0], [y_min,y_max],color="black", linewidth=0.6)
bx.set_xlabel("X", fontsize=20) ; bx.set_ylabel("Y", fontsize=20)
# 信頼楕円の描画
theta = np.linspace(0, 2*np.pi, 20)
EX = np.cos(theta) ;    EY = np.sin(theta)
EX, EY = Dr @ Dt @ np.array([EX, EY])
bx.plot(EX + mx, EY + my, color="w")

plt.show()
```

1.2　多次元の標準正規分布を一次変換して様々な正規分布をつくる

一般の多次元の正規分布は、標準正規分布を一次変換して、スケールを変えたり、原点を変えたり、回転させたりして得る事ができる。その様子を見ていこう。

■スケーリングとシフト

先の多次元標準正規分布に従う n 個の確率変数ベクトル $\boldsymbol{Z}$ に対して以下のような一次変換をする事を考えてみる。ここで、σ は正のスカラー定数、μ は n 次元の定数の縦ベクトルとする。

$$\boldsymbol{X} = \sigma\boldsymbol{Z} + \boldsymbol{\mu}$$

この期待値と分散行列は

$$E[\boldsymbol{X}] = E[\sigma\boldsymbol{Z}+\boldsymbol{\mu}] = \sigma E[\boldsymbol{Z}] + \boldsymbol{\mu} = \sigma\begin{pmatrix}0\\0\\\vdots\\0\end{pmatrix} + \begin{pmatrix}\mu_1\\\mu_2\\\vdots\\\mu_n\end{pmatrix} = \boldsymbol{\mu}$$

$$\begin{aligned}V[\boldsymbol{X}] &= E[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T] \\ &= E[(\sigma\boldsymbol{Z}+\boldsymbol{\mu}-\boldsymbol{\mu})(\sigma\boldsymbol{Z}+\boldsymbol{\mu}-\boldsymbol{\mu})^T] = \sigma^2E[\boldsymbol{Z}\boldsymbol{Z}^T] = \sigma^2V[\boldsymbol{X}] = \begin{pmatrix}\sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2\end{pmatrix} = \sigma^2\boldsymbol{I}\end{aligned}$$

この一次変換によって平均 $\boldsymbol{\mu}$、分散 $\sigma^2\boldsymbol{I}$ の正規分布になる。この多次元正規分布を $N(\boldsymbol{\mu},\sigma^2\boldsymbol{I})$ と表す。図 2 は具体的に標準正規分布 $\boldsymbol{Z}$ に対して以下の式で一次変換した $N(\boldsymbol{\mu},1.1^2\boldsymbol{I})$ の 2 次元正規分布のグラフである。

$$\begin{pmatrix}X_1\\X_2\end{pmatrix} = 1.1^2\begin{pmatrix}Z_1\\Z_2\end{pmatrix} + \begin{pmatrix}2\\1\end{pmatrix}$$

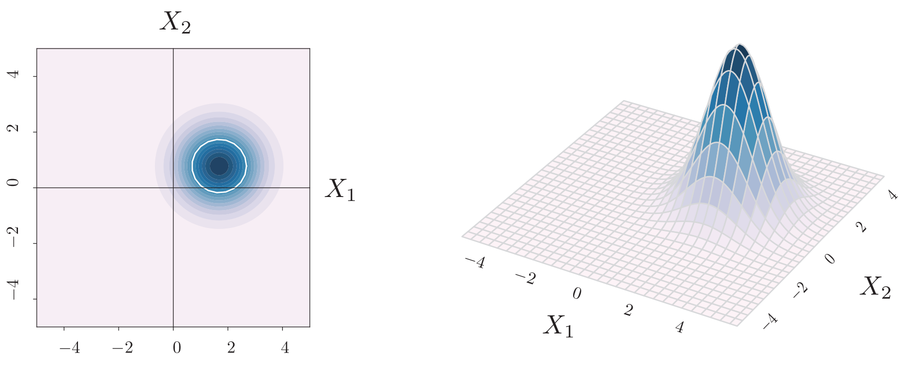

図 2　多次元正規分布 $N(\boldsymbol{\mu},\sigma^2\boldsymbol{I})$

■縦横の伸縮

スケーリングでは全方向に均等に σ 倍したが、軸によって伸縮の倍率を変えると図 3 のように楕円状の分布になる。これは n 個の多次元標準正規分布に従う確率変数ベクトル $\boldsymbol{Z}$ に対しての以下の式のように各成分を別々の定数で伸縮した事になる。

$$\boldsymbol{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}, \qquad X = \begin{pmatrix} \sigma_1 Z_1 \\ \sigma_2 Z_2 \\ \vdots \\ \sigma_n Z_n \end{pmatrix}$$

これを行列で表すと以下のようになる。

$$\boldsymbol{X} = D\boldsymbol{Z}, \qquad D = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix}$$

この期待値と分散行列は

$$E[\boldsymbol{X}] = E[D\boldsymbol{Z}] = DE[\boldsymbol{Z}] = D\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \boldsymbol{o}$$

$$V[\boldsymbol{X}] = E[(D\boldsymbol{Z})(D\boldsymbol{Z})^T] = DE[\boldsymbol{Z}\boldsymbol{Z}^T]D^T = DV[\boldsymbol{Z}]D^T = D\boldsymbol{I}D^T = D^2$$

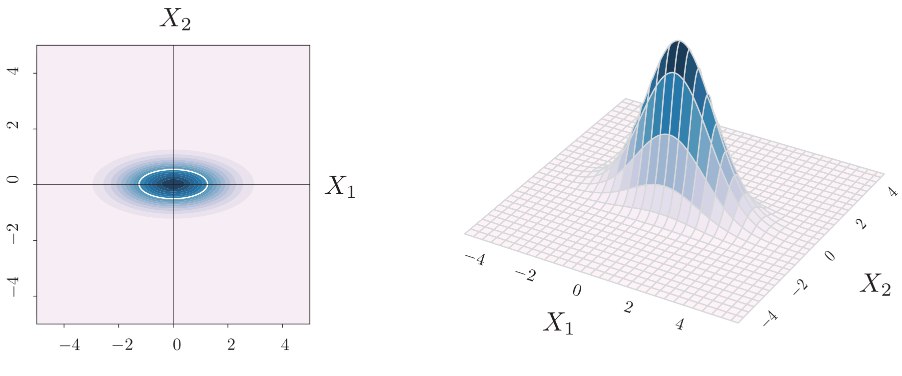

図 3　多次元正規分布 $N(\boldsymbol{o}, D^2)$

■回転

図 3 の多次元正規分布をさらに時計周りに 45° 回転させたものを考えると図 4 のようになる。一般的に、原点を中心とする期待値が $\boldsymbol{o}$ の多次元正規分布 $N(\boldsymbol{o}, \boldsymbol{V})$ はこのような形をしている。

回転という操作を行列で表すと、直交行列をかけるという操作になる。直交行列とは以下の式を満たす n 次正方行列 Q の事 (??ページの節??を参照) であり、その列ベクトルはお互いに正規直交基底で出来ている。。

$$Q^t Q = I$$

この直交行列による写像は、２つのベクトルのなす角度と長さの両方を変えない写像であるという特徴がある。つまり図形を合同な図形に移す写像であり、そういった写像は幾何学的にいうと回転または鏡映（裏表を変える）となる。

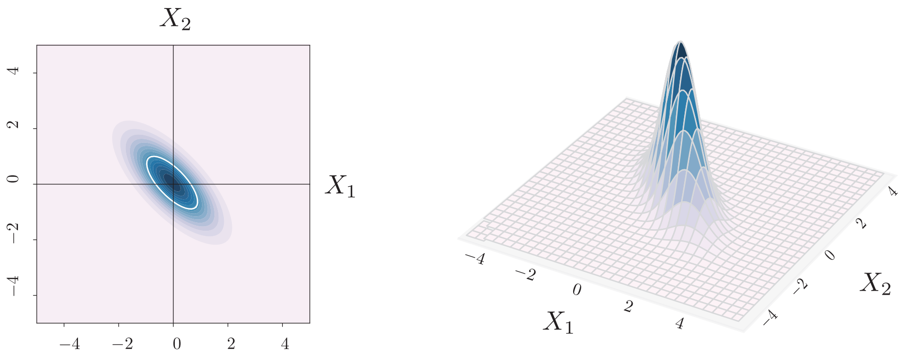

図 4　多次元正規分布 $N(\boldsymbol{o}, \boldsymbol{V})$

この一般的な原点を中心とする期待値が $\boldsymbol{o}$ の多次元正規分布 $N(\mathbf{0}, \boldsymbol{V})$ をつくり出す過程を振り返ってみる。

1. $N(\boldsymbol{o}, I)$ の確率変数ベクトルをつくる
   多次元標準正規分布 $N(\boldsymbol{o}, I)$ に従う確率変数ベクトル $\boldsymbol{Z}$ を持ってくる

$$\boldsymbol{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}$$

   この期待値と分散行列は

$$E[\boldsymbol{Z}] = \boldsymbol{o}\,, \quad V[\boldsymbol{Z}] = I$$

2. 各変数毎に拡大縮小して $N(\boldsymbol{o}, D^2)$ の確率変数ベクトルにする
   その $\boldsymbol{Z}$ に下式のように対角行列 D をかける事で、各成分毎に拡大縮小率を変えた新しい確率変数ベクトル $\boldsymbol{X}$ をつくる

$$\boldsymbol{X} = D\boldsymbol{Z} = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} = \begin{pmatrix} \sigma_1 Z_1 \\ \sigma_2 Z_2 \\ \vdots \\ \sigma_n Z_n \end{pmatrix}$$

   この期待値と分散行列は

$$E[\boldsymbol{X}] = \boldsymbol{o}\,, \quad V[\boldsymbol{X}] = D^2$$

3. 全体を回転して $N(\boldsymbol{o}, \boldsymbol{V})$ の確率変数ベクトルにする
その $\boldsymbol{X}$ に直交行列 Q をかけて $\boldsymbol{Y} = Q\boldsymbol{X}$ をつくる。この期待値と分散行列を求めよう。

期待値 $E[\boldsymbol{Y}]$ は、$E[\boldsymbol{X}] = \boldsymbol{o}$ より

$$E[\boldsymbol{Y}] = E[Q\boldsymbol{X}] = QE[\boldsymbol{X}] = \boldsymbol{o}$$

共分散行列 $V[\boldsymbol{Y}]$ は、$V[\boldsymbol{X}] = D^2$ より

$$V[\boldsymbol{Y}] = E[(Q\boldsymbol{X})(Q\boldsymbol{X})^T] = E[Q(XX^T)Q^T] = QV[\boldsymbol{X}]Q^T = QD^2Q^t$$

このように原点を中心とした多次元の標準正規分布に従う確率ベクトル $\boldsymbol{Z}$ を変数毎に伸縮し、さらに回転する事で、一般の多次元正規分布をつくりだす事ができ、その共分散行列は以下のように表す事ができる。

$$\boldsymbol{V} = Q\,D^2\,Q^T \tag{3}$$

逆に、今 $\boldsymbol{V}$ という共分散行列が与えられたとする。分散行列は $Cov[Z_1, Z_2] = Cov[Z_2, Z_1]$ なのでお互いの対角要素が同じ値の対称行列[*1]になる。この対称行列である $\boldsymbol{V}$ という共分散行列を対角行列 D と直交行列 Q によって上記のように分解する事ができれば、分布状況の見通しがわかりやすくなる。

いま上記の式 (3) を満たすような直交行列 Q が見つかったと過程すると $Q^TQ = QQ^T = I$ なので、式 (3) の両辺に左から Q^T をかけて右から Q をかけると以下のような対角行列が抽出できる。

$$Q^T\boldsymbol{V}Q = Q^T(Q\,D^2\,Q^T)Q = D^2$$

この D^2 の要素は各確率変数の分散になる。

$$D^2 = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{pmatrix} = \begin{pmatrix} V[Z_1] & & & \\ & V[Z_2] & & \\ & & \ddots & \\ & & & V[Z_n] \end{pmatrix}$$

このように対称行列 V を直交行列 Q を用いて対角行列 D に変換する事を対角化と呼ぶ。分散行列だけをみていると n 個の確率変数が複雑に絡み合っているように見える場合でも、対角化することによって複雑に見えた関係が実は独立な n 個の変数が合成された結果であると捉え直す事が可能になる。

[*1] 対称行列とは、任意の行列 A の行と列を入れ替えたとき、元の行列 A と等しくなるもの

1.3 分散行列の対角化

分散行列のように要素が実数の対称行列は必ず対角化できる。いったんこれを前提にする。いま分散行列 V に対して、ある適切な直交行列 Q を持ってきて、Q^TVQ が対角行列になるようにできたとし、できた対角行列を以下のように Λ と表すものとする。

$$Q^TVQ = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} = \Lambda$$

この両辺に Q をかける。Q は直交行列で $QQ^T = I$ なので、左辺は $QQ^TVQ = VQ$ となり左辺は $Q\Lambda$。つまり以下のように表す事ができる。

$$VQ = Q\Lambda$$

これを成分表示すると

$$\begin{pmatrix} - & \boldsymbol{v}_1 & - \\ & \vdots & \\ - & \boldsymbol{v}_n & - \end{pmatrix} \begin{pmatrix} | & & | \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

この両辺の 1 列目を取り出すと

$$\begin{pmatrix} - & \boldsymbol{v}_1 & - \\ & \vdots & \\ - & \boldsymbol{v}_n & - \end{pmatrix} \begin{pmatrix} | \\ \boldsymbol{q}_1 \\ | \end{pmatrix} = \lambda_1 \begin{pmatrix} | \\ \boldsymbol{q}_1 \\ | \end{pmatrix}$$

列ごとにみると以下のような構造。

$$\begin{pmatrix} - & \boldsymbol{v}_1 & - \\ & \vdots & \\ - & \boldsymbol{v}_n & - \end{pmatrix} \begin{pmatrix} | \\ \boldsymbol{q}_1 \\ | \end{pmatrix} = \lambda_1 \begin{pmatrix} | \\ \boldsymbol{q}_1 \\ | \end{pmatrix} \quad ,\cdots, \quad \begin{pmatrix} - & \boldsymbol{v}_1 & - \\ & \vdots & \\ - & \boldsymbol{v}_n & - \end{pmatrix} \begin{pmatrix} | \\ \boldsymbol{q}_n \\ | \end{pmatrix} = \lambda_n \begin{pmatrix} | \\ \boldsymbol{q}_n \\ | \end{pmatrix}$$

つまり、求めたい直交行列 Q の各列は以下のような行列 V の固有ベクトルを求める式で構成されている[*2]。

$$V\boldsymbol{q}_n = \lambda \boldsymbol{q}_n$$

以上の事から、直交行列 Q は分散行列 V の固有ベクトル $\boldsymbol{q}_n$ を求めて、その固有ベクトルを列ベクトルとして並べる事によってつくる事ができる事がわかる。具体的な手順は以下。

1. 与えられた分散行列 V の固有値 $\lambda_1, \cdots, \lambda_n$ を求める
2. 各固有値 λ_n に対応する固有ベクトル $\boldsymbol{p}_n$ を求める
3. 各固有ベクトルの長さを 1 にそろえる。つまり $\boldsymbol{q}_n = \boldsymbol{p}_n / ||\boldsymbol{p}_n||$

[*2] 固有値及び固有ベクトルについては、**??**ページの固有値と固有ベクトルの定義を参照

4. 固有ベクトルを以下のように列方向に並べて Q をつくる

$$Q = \begin{pmatrix} | & & | \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ | & & | \end{pmatrix}$$

5. 対応する固有値を対角成分に並べた行列 Λ をつくる

$$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

この直交行列 Q と対角行列 Λ を用いれば分散行列は $VQ = Q\Lambda$ と表せ、この両辺に左から Q^T をかける事で

$$Q^t VQ = \Lambda$$

というように対角化できる。