8. Component Analysis

8.1 Singular Value Decomposition

If **A** is a square matrix, then we can perform eigenvector-eigenvalue decomposition for **A**:

$$\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^{-1}$$

$$\mathbf{A} = \lambda_1 \mathbf{e_1}\mathbf{f_1^H} + \lambda_2 \mathbf{e_2}\mathbf{f_2^H} + \cdots + \lambda_{N-1}\mathbf{e_{N-1}}\mathbf{f_{N-1}^H} + \lambda_N \mathbf{e_N}\mathbf{f_N^H}$$

where

$$\mathbf{E} = \begin{bmatrix} \mathbf{e_1} & \mathbf{e_2} & \cdots & \mathbf{e_N} \end{bmatrix}, \quad \mathbf{E}^{-1} = \begin{bmatrix} \mathbf{f_1^H} \\ \mathbf{f_2^H} \\ \vdots \\ \mathbf{f_N^H} \end{bmatrix} \qquad \mathbf{A}\mathbf{e_m} = \lambda_m \mathbf{e_m}$$

If $|\lambda_m|$ is the largest, then

$$\lambda_m \mathbf{e_m}\mathbf{f_m^H}$$

is the most important component of **A**.

8.1.1 Singular Value Decomposition Process

Q: How do we perform eigenvector-eigenvalue decomposition for **A** if **A** is not a square matrix?

$$size(\mathbf{A}) = M \times N, \quad M \neq N$$

We can apply the singular value decomposition (SVD) process as follows.

(Step 1) Generate **B** and **C**

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} \qquad \mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{H}}$$

(Note): Since **B** is an NxN square matrix,
C is an MxM square matrix,
therefore, it is possible to derive the eigenvector sets for **B** and **C**.

$$\mathbf{B} = \mathbf{A}^H\mathbf{A} \qquad \mathbf{C} = \mathbf{A}\mathbf{A}^H$$

(Step 2) Perform Eigenvector-Eigenvalue Decomposition for **B** and **C**

$$\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1} \qquad \mathbf{C} = \tilde{\mathbf{U}}\boldsymbol{\Omega}\tilde{\mathbf{U}}^{-1}$$

(Note): Since $\mathbf{B}^H = \mathbf{B}$, $\mathbf{C}^H = \mathbf{C}$, **B** and **C** have orthogonal eigenvector sets and $\tilde{\mathbf{U}}$ and **V** are orthogonal matrices.

(i) It is necessary to normalize $\tilde{\mathbf{U}}$ and **V** properly such that

$$\mathbf{V}^H\mathbf{V} = \mathbf{I} \qquad \tilde{\mathbf{U}}^H\tilde{\mathbf{U}} = \mathbf{I}$$

then

$$\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^H \qquad \mathbf{C} = \tilde{\mathbf{U}}\boldsymbol{\Omega}\tilde{\mathbf{U}}^H$$

(ii) It is proper to sort the eigenvalues of **B** and **C** from large to small.

The eigenvectors are also sorted according to eigenvalues.

(Step 3) Then, we calculate

$$\mathbf{S_1} = \mathbf{\tilde{U}^H A V}$$

$\mathbf{S_1}$ will be an MxN diagonal matrix

$$S_1[m,n] = 0 \quad if\ m \neq n$$

(Step 4) Varying the sign of $\mathbf{S_1}$ and $\mathbf{\tilde{U}}$

$$S[m,n] = |S_1[m,n]|$$

$$U[m,n] = \tilde{U}[m,n] \quad \text{if } S_1[n,n] \geq 0,$$

$$U[m,n] = -\tilde{U}[m,n] \quad \text{if } S_1[n,n] < 0,$$

(varying the sign of the n^{th} column)

(Note): With sign change,

$\mathbf{S} = \mathbf{U^H A V}$ and $\mathbf{C} = \mathbf{U \Omega U^{-1}}$

are still satisfied.

(Step 5) Then, the SVD of **A** is

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

eigenvector matrix
of $\mathbf{A}\mathbf{A}^{\mathbf{H}}$, size: MxM

diagonal matrix,
size: MxN

eigenvector matrix
of $\mathbf{A}^{\mathbf{H}}\mathbf{A}$, size: NxN

If

$$\mathbf{U} = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \cdots & \mathbf{u_M} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} & \cdots & \mathbf{v_N} \end{bmatrix}$$

then

$$\mathbf{A} = \underline{s_1 \mathbf{u_1} \mathbf{v_1^H}} + \underline{s_2 \mathbf{u_2} \mathbf{v_2^H}} + \cdots + s_{K-1} \mathbf{u_{K\text{-}1}} \mathbf{v_{K\text{-}1}^H} + s_K \mathbf{u_K} \mathbf{v_K^H}$$

$$s_1 \geq s_2 \geq \cdots \geq s_{K-1} \geq s_K$$

most
significant

second
significant

where $K = \min(M, N)$

$$s_n = S[n, n]$$

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_N \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

if $M > N$

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_M & 0 & \cdots & 0 \end{bmatrix}$$

if $M < N$

s_k is call the singular value

[Example 1] Perform the SVD for the following matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \\ -1 & 1 \end{bmatrix}$$

(Solution): First, we determine

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} = \begin{bmatrix} 6 & 2 \\ 2 & 6 \end{bmatrix} \qquad \mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{H}} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix}$$

Then, we perform eigenvector-eigenvalue decomposition for **B** and **C**:

$$\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathbf{H}} \quad \text{where} \quad \mathbf{V} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} 8 & 0 \\ 0 & 4 \end{bmatrix}$$

(with normalization)

$$\mathbf{C} = \tilde{\mathbf{U}}\boldsymbol{\Omega}\tilde{\mathbf{U}}^{\mathbf{H}}$$

where $\tilde{\mathbf{U}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$ $\quad \boldsymbol{\Omega} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

(with normalization)

Note: The eigenvectors should be (i) normalized and (ii) sorted according to the magnitudes of the eigenvalues.

Then,

$$\mathbf{S_1} = \tilde{\mathbf{U}}^{\mathbf{H}}\mathbf{AV} = \begin{bmatrix} \sqrt{8} & 0 \\ 0 & -2 \\ 0 & 0 \end{bmatrix}$$

Then,

$$\mathbf{S} = \begin{bmatrix} \left|\sqrt{8}\right| & 0 \\ 0 & \left|-2\right| \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{8} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}$$

Since $S_1[2, 2] < 0$, we change the sign of the 2nd column of $\tilde{\mathbf{U}}$ and obtain

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Therefore,

$$\mathbf{A} = \mathbf{USV}^{\mathbf{H}} \qquad \text{where}$$

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} \sqrt{8} & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \quad \mathbf{V} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Note that

$$\mathbf{A} = s_1\mathbf{u_1}\mathbf{v_1^H} + s_2\mathbf{u_2}\mathbf{v_2^H}$$

$$\mathbf{A} = \sqrt{8}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} + 2\begin{bmatrix} 0 \\ -1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$$

principal component

minor component

(Note):

(1) In fact, the eigenvalues of **B** and **C** has a close relation to the singular values of **A**.

$$\mathbf{S}^{\mathbf{H}}\mathbf{S} = \mathbf{D} \qquad\qquad \mathbf{S}\mathbf{S}^{\mathbf{H}} = \mathbf{\Omega}$$

$$S^2[n,n] = D[n,n] = \Omega[n,n]$$

Since

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} = \mathbf{V}\mathbf{S}^{\mathbf{H}}\mathbf{U}^{\mathbf{H}}\mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}} = \mathbf{V}\mathbf{S}^{\mathbf{H}}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

(Note):

(2) Even when $M = N$ (i.e., **A** is a square matrix), the SVD may not be the same as the eigenvector-eigenvalue decomposition.

For the SVD, **U** and **V** are both orthonormal matrices and the singular values are non-negative.

However, for a square matrix, the eigenvectors may not be orthogonal and the eigenvalues can be negative (even complex).

(3) Moreover, since **U** and **V** are usually different and $\mathbf{V}^{\mathbf{H}} \neq \mathbf{U}^{-1}$, one cannot use the SVD to compute the power of a matrix.

[Example 2] Determine the eigenvector-eigenvalue decomposition and the SVD of **A**.

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 0 & -1 \end{bmatrix}$$

(Solution): The eigenvalues of **A** are 2 and -1.

The eigenvectors corresponding to 2 is $[1\ 0]^T$
The eigenvectors corresponding to -1 is $[1, 3]^T$

Therefore,

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1/3 \\ 0 & 1/3 \end{bmatrix}$$

To perform SVD for **A**,

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} \qquad \mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{H}} = \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{B} = \overset{\mathbf{V}}{\begin{bmatrix} 0.8507 & 0.5257 \\ -0.5257 & 0.8507 \end{bmatrix}} \overset{\mathbf{D}}{\begin{bmatrix} 5.2361 & 0 \\ 0 & 0.7639 \end{bmatrix}} \overset{\mathbf{V^H}}{\begin{bmatrix} 0.8507 & -0.5257 \\ 0.5257 & 0.8507 \end{bmatrix}}$$

$$\mathbf{C} = \overset{\tilde{\mathbf{U}}}{\begin{bmatrix} 0.9732 & -0.2298 \\ 0.2298 & 0.9732 \end{bmatrix}} \overset{\boldsymbol{\Omega}}{\begin{bmatrix} 5.2361 & 0 \\ 0 & 0.7639 \end{bmatrix}} \overset{\tilde{\mathbf{U}}^{\mathbf{H}}}{\begin{bmatrix} 0.9732 & 0.2298 \\ -0.2298 & 0.9732 \end{bmatrix}}$$

$$\begin{bmatrix} 0.9732 & -0.2298 \\ 0.2298 & 0.9732 \end{bmatrix}^{\mathbf{H}} \mathbf{A} \begin{bmatrix} 0.8507 & 0.5257 \\ -0.5257 & 0.8507 \end{bmatrix} = \begin{bmatrix} 2.2882 & 0 \\ 0 & -0.8740 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 0.9732 & 0.2298 \\ 0.2298 & -0.9732 \end{bmatrix} \begin{bmatrix} 2.2882 & 0 \\ 0 & 0.8740 \end{bmatrix} \begin{bmatrix} 0.8507 & -0.5257 \\ 0.5257 & 0.8507 \end{bmatrix}$$

8.1.2 Generalized Inverse Using the SVD

Suppose that the SVD of **A** is

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

Then the generalized inverse of **A** is

$$\mathbf{A}^{+} = \mathbf{V}\mathbf{S}^{+}\mathbf{U}^{\mathbf{H}}$$

where

$$S^{+}[n,n] = 1/S[n,n] \quad \mathit{if}\ S[n,n] \neq 0$$

$$S^{+}[n,n] = 0 \quad \mathit{if}\ S[n,n] = 0$$

$$\mathit{size}\left(\mathbf{S}^{+}\right) = N \times M \quad \mathit{if}\ \mathit{size}\left(\mathbf{S}\right) = M \times N$$

(Proof):

$$\mathbf{AA^{+}A} = \mathbf{USV^{H}VS^{+}U^{H}USV^{H}} = \mathbf{USS^{+}SV^{H}}$$

If

$$\mathbf{S_2} = \mathbf{S^{+}S}$$

then

$$S_2[n,n] = 1 \quad if\ S[n,n] \neq 0 \qquad\qquad S_2[n,n] = 0 \quad if\ S[n,n] = 0$$

Therefore,

$$\mathbf{S} = \mathbf{SS^{+}S}$$

$$\mathbf{AA^{+}A} = \mathbf{USV^{H}} = \mathbf{A}$$

(1) $\mathbf{AA^{+}A} = \mathbf{A}$ is satisfied.

Note: The generalized inverse derived from the SVD is in fact the pseudo inverse since

(2) $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+} = \mathbf{A}^{+}$

(3) $\left(\mathbf{A}\mathbf{A}^{+}\right)^{H} = \mathbf{A}\mathbf{A}^{+}$

(4) $\left(\mathbf{A}^{+}\mathbf{A}\right)^{H} = \mathbf{A}^{+}\mathbf{A}$

are all satisfied.

(Try to prove them)

[Example 3] Determine the generalized inverse of the following matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 1 & -2 & 1 \\ -1 & 2 & -1 \end{bmatrix}$$

Note: Since the 1st and the 3rd columns are dependent, we cannot use the method of

$$(\mathbf{A}^{\mathbf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}$$

to determine the generalized inverse. Instead, we should apply the SVD method.

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}} \qquad \mathbf{A}^{+} = \mathbf{V}\mathbf{S}^{+}\mathbf{U}^{\mathbf{H}}$$

(Solution): Since

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 1 & -2 & 1 \\ -1 & 2 & -1 \end{bmatrix}$$

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} = \begin{bmatrix} 10 & 4 & 10 \\ 4 & 16 & 4 \\ 10 & 4 & 10 \end{bmatrix} \qquad \mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{H}} = \begin{bmatrix} 12 & 12 & 0 & 0 \\ 12 & 12 & 0 & 0 \\ 0 & 0 & 6 & -6 \\ 0 & 0 & -6 & 6 \end{bmatrix}$$

$$\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathbf{H}}$$

where $\mathbf{V} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \end{bmatrix}$ $\quad \mathbf{D} = \begin{bmatrix} 24 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$\mathbf{C} = \tilde{\mathbf{U}}\mathbf{\Lambda}\tilde{\mathbf{U}}^{\mathbf{H}}$ where

$$\tilde{\mathbf{U}} = \begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 0 & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix} \qquad \mathbf{\Lambda} = \begin{bmatrix} 24 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Then

$$\mathbf{S_1} = \tilde{\mathbf{U}}^{\mathbf{H}}\mathbf{AV} = \begin{bmatrix} \sqrt{24} & 0 & 0 \\ 0 & \sqrt{12} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Since all entries of $\mathbf{S_1}$ are non-negative,

$$\mathbf{S} = \mathbf{S_1} \qquad \mathbf{U} = \tilde{\mathbf{U}}$$

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

$$\mathbf{U} = \begin{bmatrix} 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 0 & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} \sqrt{24} & 0 & 0 \\ 0 & \sqrt{12} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \end{bmatrix}$$

$$\mathbf{A}^{+} = \mathbf{V}\mathbf{S}^{+}\mathbf{U}^{\mathbf{H}}$$

$$\mathbf{S}^{+} = \begin{bmatrix} 1/\sqrt{24} & 0 & 0 & 0 \\ 0 & 1/\sqrt{12} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{A}^{+} = \begin{bmatrix} 1/12 & 1/12 & 1/12 & -1/12 \\ 1/12 & 1/12 & -1/6 & 1/6 \\ 1/12 & 1/12 & 1/12 & -1/12 \end{bmatrix}$$

8.2 Principal Component Analysis

Principal component analysis (PCA) is to find the principal component of a set of data.

Principal components: Corresponding to larger singular values for SVD

[Process of PCA]

Suppose that there is a set of data. The number of data is M and each data has the length of N.

$$\mathbf{x_m} = \begin{bmatrix} x_{m,1} & x_{m,2} & x_{m,3} & \cdots & x_{m,N} \end{bmatrix}$$

$m = 1, 2, \ldots, M$

(In usual, $M >> N$)

(1) First, we subtract each entry by $\bar{x}_n = \frac{1}{M}\sum_{m=1}^{M} x_{m,n}$

$$\mathbf{a_m} = \begin{bmatrix} a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,N} \end{bmatrix}$$

where $a_{m,n} = x_{m,n} - \bar{x}_n$, $\quad \bar{x}_n = \frac{1}{M}\sum_{m=1}^{M} x_{m,n}$

(2) Then, construct an MxN matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \vdots \\ \mathbf{a_M} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

(3) Then, perform SVD for $\mathbf{A}$

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{H}}$$

second important

$(N\text{-}1)^{\text{th}}$ important

N^{th} important

(4) Then

$$\mathbf{A} = s_1 \mathbf{u_1} \mathbf{v_1^H} + s_2 \mathbf{u_2} \mathbf{v_2^H} + \cdots + s_{N-1} \mathbf{u_{N\text{-}1}} \mathbf{v_{N\text{-}1}^H} + s_N \mathbf{u_N} \mathbf{v_N^H}$$

most important

where $s_n = S[n,n]$

$$\mathbf{U} = [\mathbf{u_1} \quad \mathbf{u_2} \quad \cdots \quad \mathbf{u_M}], \quad \mathbf{V} = [\mathbf{v_1} \quad \mathbf{v_2} \quad \cdots \quad \mathbf{v_N}]$$

If we want to reduce the component from N to L due to the consideration of compression or feature selection, then

$$\mathbf{A} \cong \mathbf{A_1} = s_1\mathbf{u_1}\mathbf{v_1^H} + s_2\mathbf{u_2}\mathbf{v_2^H} + \cdots + \mathbf{s_L}\mathbf{u_L}\mathbf{v_L^H}$$

Note:

$$\mathbf{x_m} \cong c_{m,1}\mathbf{v_1^H} + c_{m,2}\mathbf{v_2^H} + \cdots + c_{m,L}\mathbf{v_L^H} + \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_L \end{bmatrix}$$

where $c_{m,n} = s_n u_n[m] \longleftarrow$ m^{th} entry of $\mathbf{u_n}$

$\mathbf{v_1^H}, \mathbf{v_2^H}, \cdots, \mathbf{v_L^H}$ can be viewed as the most important L axes

In general,

$$\mathbf{x} \cong c_1\mathbf{v_1^H} + c_2\mathbf{v_2^H} + \cdots + c_L\mathbf{v_L^H} + \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_L \end{bmatrix}$$

$$c_n \in (-\infty, \infty)$$

Main Applications of the PCA

(1) Dimensionality reduction (i.e., feature selection) for pattern recognition and machine learning

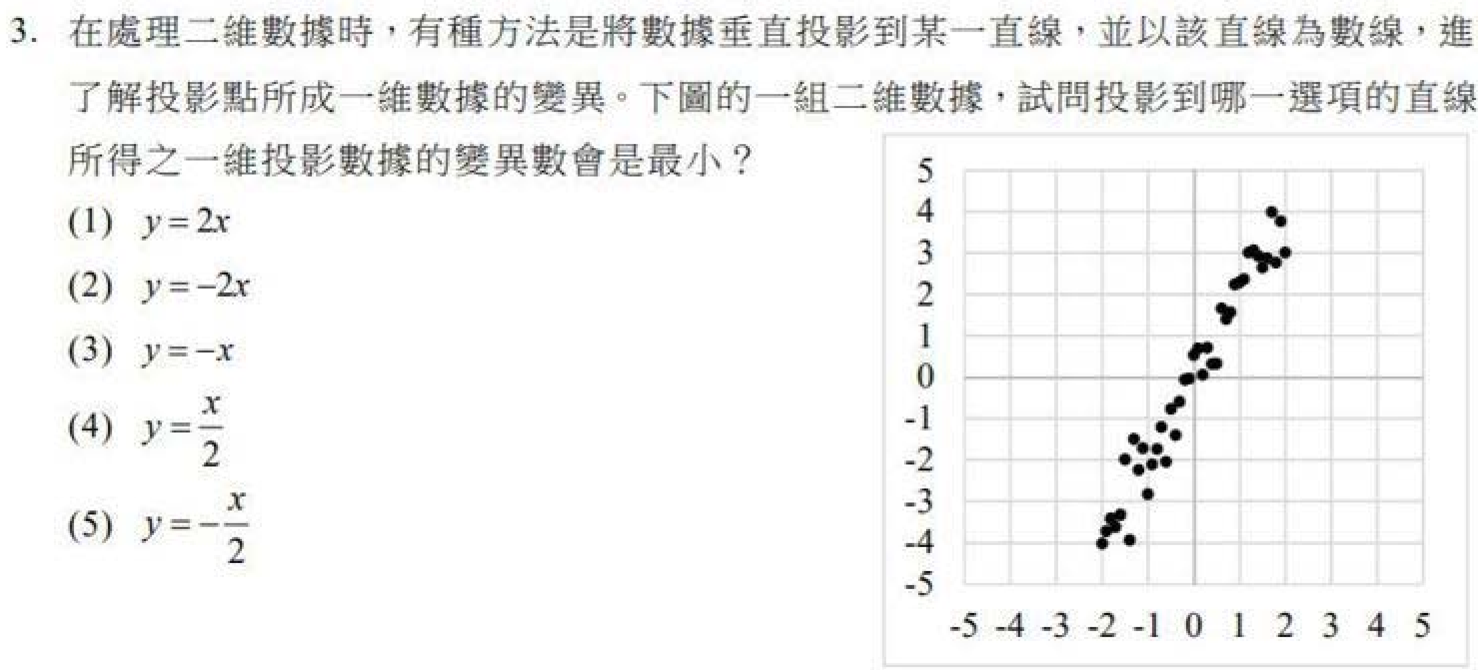
(2) Data compression

(3) Data mining

(4) Identifying the principal axis of an object in an image

(5) Line approximation

Example of PCA



3. 在處理二維數據時，有種方法是將數據垂直投影到某一直線，並以該直線為數線，進而了解投影點所成一維數據的變異。下圖的一組二維數據，試問投影到哪一選項的直線，所得之一維投影數據的變異數會是最小？

(1) $y=2x$

(2) $y=-2x$

(3) $y=-x$

(4) $y=\frac{x}{2}$

(5) $y=-\frac{x}{2}$

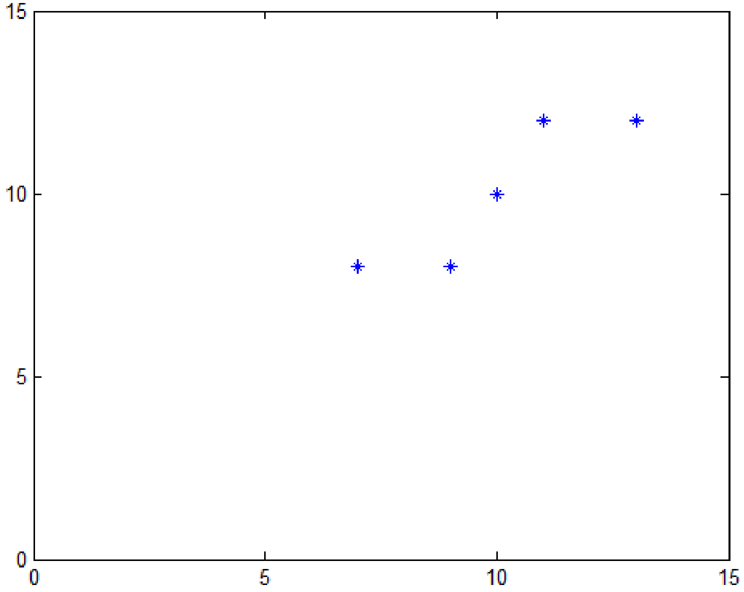
From 2022 大考中心官網

[Example 1] Suppose that there are 5 points in a 2-D space and their coordinates are

(7,8), (9,8), (10, 10), (11,12), (13,12)

Try to find a line that can approximate these points.

(Note): $M = 5$, $N = 2$



(Solution):

First, since the mean of these 5 points is

(10, 10),

we subtract these points by (10, 10) and obtain

(-3, -2), (-1 -2), (0, 0), (1, 2), (3, 2)

(-3, -2), (-1 -2), (0, 0), (1, 2), (3, 2)

Then, we construct a 5x2 matrix **A**:

$$\mathbf{A}=\begin{bmatrix}-3 & -2\\ -1 & -2\\ 0 & 0\\ 1 & 2\\ 3 & 2\end{bmatrix}$$

Then, we perform SVD for **A**:

$$\mathbf{B}=\mathbf{A}^{\mathbf{H}}\mathbf{A}=\begin{bmatrix}20 & 16\\ 16 & 16\end{bmatrix}$$

$$\mathbf{B}=\mathbf{V}\mathbf{D}\mathbf{V}^{-1}\qquad \mathbf{V}=\begin{bmatrix}0.7497 & -0.6618\\ 0.6618 & 0.7497\end{bmatrix}\qquad \mathbf{D}=\begin{bmatrix}34.1245 & 0\\ 0 & 1.8755\end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} -3 & -2 \\ -1 & -2 \\ 0 & 0 \\ 1 & 2 \\ 3 & 2 \end{bmatrix}$$

$$\mathbf{C} = \mathbf{A}\mathbf{A}^{\mathbf{H}}$$

$$\mathbf{C} = \tilde{\mathbf{U}}\boldsymbol{\Lambda}\tilde{\mathbf{U}}^{\mathbf{H}}$$

$$\tilde{\mathbf{U}} = \begin{bmatrix} -0.6116 & -0.3549 & 0 & 0.0393 & 0.7060 \\ -0.3549 & 0.6116 & 0 & 0.7060 & -0.0393 \\ 0 & 0 & 1 & 0 & 0 \\ 0.3549 & -0.6116 & 0 & 0.7060 & -0.0393 \\ 0.6116 & 0.3549 & 0 & 0.0393 & 0.7060 \end{bmatrix}$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} 34.1245 & 0 & 0 & 0 & 0 \\ 0 & 1.8755 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{S_1} = \mathbf{\tilde{U}^H A V} = \begin{bmatrix} 5.8416 & 0 \\ 0 & -1.3695 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{S} = \left|\mathbf{S_1}\right| = \begin{bmatrix} 5.8416 & 0 \\ 0 & 1.3695 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{U} = \begin{bmatrix} -0.6116 & 0.3549 & 0 & 0.0393 & 0.7060 \\ -0.3549 & -0.6116 & 0 & 0.7060 & -0.0393 \\ 0 & 0 & 1 & 0 & 0 \\ 0.3549 & 0.6116 & 0 & 0.7060 & -0.0393 \\ 0.6116 & -0.3549 & 0 & 0.0393 & 0.7060 \end{bmatrix}$$

$$\mathbf{A} = \mathbf{USV}^{\mathbf{H}}$$

$$\mathbf{U} = \begin{bmatrix} -0.6116 & 0.3549 & 0 & 0.0393 & 0.7060 \\ -0.3549 & -0.6116 & 0 & 0.7060 & -0.0393 \\ 0 & 0 & 1 & 0 & 0 \\ 0.3549 & 0.6116 & 0 & 0.7060 & -0.0393 \\ 0.6116 & -0.3549 & 0 & 0.0393 & 0.7060 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 5.8416 & 0 \\ 0 & 1.3695 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} 0.7497 & -0.6618 \\ 0.6618 & 0.7497 \end{bmatrix}$$

Then, **A** can be expanded by

$$\mathbf{A} = \underset{\underset{\text{principal component}}{\uparrow}}{5.8416} \begin{bmatrix} -0.6116 \\ -0.3549 \\ 0 \\ 0.3549 \\ 0.6116 \end{bmatrix} \begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix} + \underset{\underset{\text{secondary component}}{\uparrow}}{1.3695} \begin{bmatrix} 0.3549 \\ -0.6116 \\ 0 \\ 0.6116 \\ -0.3549 \end{bmatrix} \begin{bmatrix} -0.6618 & 0.7497 \end{bmatrix}$$

Therefore,

$$\mathbf{A} \cong 5.8416 \begin{bmatrix} -0.6116 \\ -0.3549 \\ 0 \\ 0.3549 \\ 0.6116 \end{bmatrix} \begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix} = \begin{bmatrix} -3.5726 \\ -2.0733 \\ 0 \\ 2.0733 \\ 3.5726 \end{bmatrix} \begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix}$$

$$\begin{bmatrix} 7 & 8 \\ 9 & 8 \\ 0 & 0 \\ 11 & 12 \\ 13 & 12 \end{bmatrix} \cong \begin{bmatrix} -3.5726 \\ -2.0733 \\ 0 \\ 2.0733 \\ 3.5726 \end{bmatrix} \begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix} + \begin{bmatrix} 10 & 10 \\ 10 & 10 \\ 10 & 10 \\ 10 & 10 \\ 10 & 10 \end{bmatrix}$$

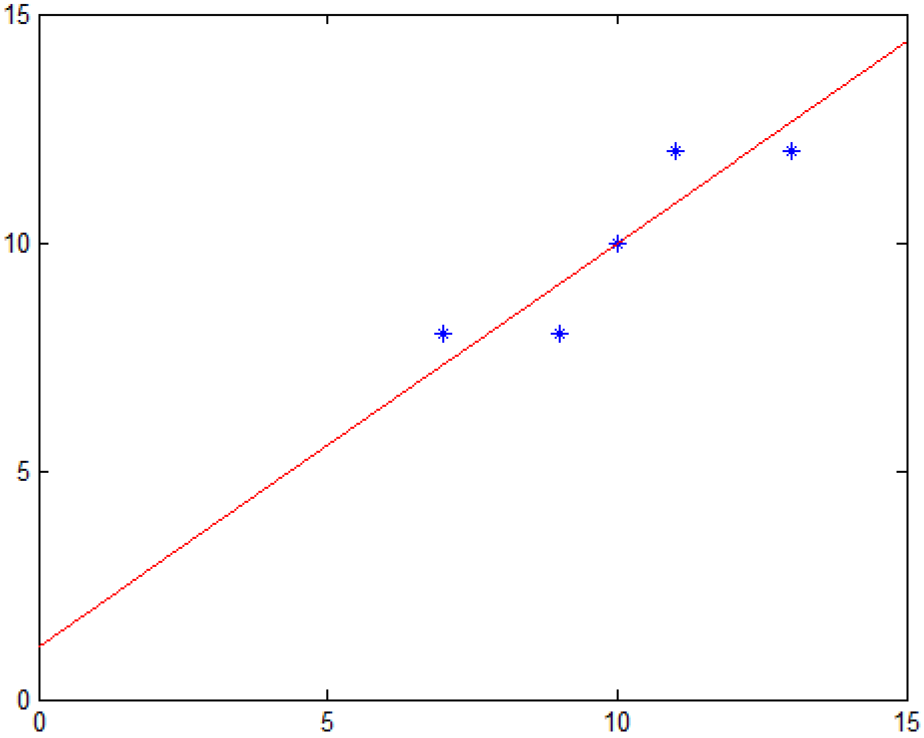
Approximation line:

$$\begin{bmatrix} 10 & 10 \end{bmatrix} + c \begin{bmatrix} 0.7497 & 0.6618 \end{bmatrix} \qquad c \in (-\infty, \infty)$$

Approximation line:

$$[10 \quad 10] + c[0.7497 \quad 0.6618]$$

$c \in (-\infty, \infty)$



[Simplification for Computation]

Suppose that we only want to find the most important *L* axes of the data. (It is usually the case for practical applications).

If *M* is very large, then the *M*x*M* matrix **U is unnecessary to be computed**. One only has to perform eigenvector-eigenvalue decomposition for **B** and obtain the *N*x*N* matrix **V**:

$$\mathbf{B} = \mathbf{A}^{\mathbf{H}}\mathbf{A} \qquad \mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$$

If *D*[*n*, *n*] is larger than other diagonal entries of **D**, then the nth column of **V** is the principal axis.

附錄十一 Some Common Mathematical Notations

(1) Commutator

$$[\mathbf{A},\mathbf{B}] = \mathbf{AB} - \mathbf{BA}$$

(2) Trace

$$tr(\mathbf{A}) = \sum_{n=1}^{N} A(n,n)$$

(3) Bras and Kets Notations

$$\langle \mathbf{A} \,|\, \mathbf{B} \rangle = \underbrace{\begin{bmatrix} \mathbf{a}_1^* & \mathbf{a}_2^* & \cdots & \mathbf{a}_\mathrm{N}^* \end{bmatrix}}_{\mathbf{A}^\mathrm{H}} \underbrace{\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_\mathrm{N} \end{bmatrix}}_{\mathbf{B}}$$

A and **B** are column vectors.

$$\langle \mathbf{A} | = \begin{bmatrix} \mathbf{a}_1^* & \mathbf{a}_2^* & \cdots & \mathbf{a}_\mathrm{N}^* \end{bmatrix} \qquad | \mathbf{B} \rangle = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_\mathrm{N} \end{bmatrix}$$

(4) sup: supremum (the least upper bound，上確界)

$$\sup\{x \mid 1 < x < 2\} = 2$$

$$\sup\left\{(-1)^n - 1/n \mid n \in N\right\} = 1$$

(5) inf: infimum (the greatest lower bound，下確界)

$$\inf\{x \mid 1 < x < 2\} = 1$$

$$\inf\left\{e^{-x} \mid x \in R\right\} = 0$$

(6) card: the number of elements in a set

$$card\left(\{x, y\}\right) = 2$$

$$card\left(\left\{x^2, y^2, xy, x, y, 1\right\}\right) = 6$$

9. Advanced Probability

Section 9.1 Moment and Correlation

Section 9.2 Probability Model (只教不考)

Section 9.3 Entropy

Section 9.4 Kullback-Leibler Divergence (KL Divergence)

Section 9.5 Basic Concepts of Random Process (只教不考)

Section 9.6 Independent Component Analysis (只教不考)

9.1 Moment and Correlation

9.1.1 Review for Probability

1. Discrete Case

(a) Probability (Probability Mass Function, PMF)

$$P_X(n) = P\{X = n\} = \frac{number\ of\ cases\ where\ X = n}{total\ number\ of\ cases}$$

Note: $\sum_n P_X(n) = 1$

(b) Joint Probability

$$P_{X,Y}(n,m) = P\{(X = n)\ and\ (Y = m)\}$$

$$= \frac{number\ of\ cases\ where\ X = n\ and\ Y = m}{total\ number\ of\ cases}$$

(c) Conditional Probability

$$P_{X|Y}(n\,|\,m) = P\{(X = n)\,|\,(Y = m)\}$$

$$= \frac{number\ of\ cases\ where\ X = n\ and\ Y = m}{number\ of\ cases\ where\ Y = m}$$

$$P_{X|Y}(n\,|\,m) = \frac{P_{X,Y}(n,m)}{P_Y(m)}$$

An Example of the Discrete Probability Distribution

Binomial Distribution

$$P_X(n) = \binom{N}{n} p^n (1-p)^{N-n} \quad \text{for } n = 0,\ 1,\ 2,\ \ldots,\ N$$

$$P_X(n) = 0$$

Other examples are shown in Section 9.2.1

2. Continuous Case

In the continuous case, it is hard to determine $\text{Prob}\{X = n\}$

(a) Cumulative Distribution Function (CDF)

$$F_X(x) = Prob(X \le x)$$

(b) Probability Density Function (PDF)

i,e., $f_X(x) = \frac{d}{dx} F_X(x)$

$$f_X(x) = \lim_{\Delta \to 0} \frac{F_X(x+\Delta) - F_X(x-\Delta)}{2\Delta}$$

$$f_X(x) = \lim_{\Delta \to 0} \frac{Prob(x-\Delta < X \le x+\Delta)}{2\Delta} \qquad f_X(x) \ne \text{Prob}\{X = x\}$$

Note: $F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx$

$$\lim_{x \to \infty} F_X(x) = \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

2. Continuous Case

(c) Joint Cumulative Distribution Function

$$F_{X,Y}(x,y) = Prob\big((X \le x)\ and\ (Y \le y)\big)$$

(d) Joint Probability Density Function

$$f_{X,Y}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{X,Y}(x,y)$$

(e) Conditional Probability

$$f_{X|Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{X,Y}(x,y)}{\frac{d}{dy}F_Y(y)}$$

(f) Integral Probability

$$f_X(x) = \int f_{X,Y}(x,y)\,dy$$

$$f_Y(y) = \int f_{X,Y}(x,y)\,dx$$

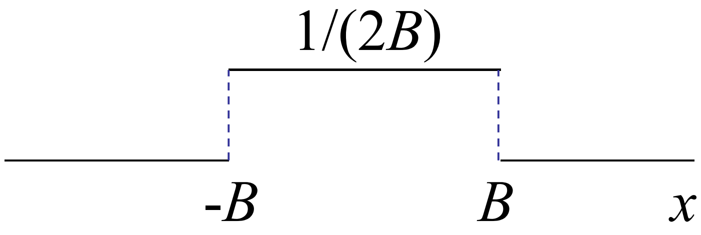
Examples of the Probability Density Function in the Continuous Case

Uniform Distribution

(Quantization error is a special case of $B = 0.5$)

$$f_X(x) = \frac{1}{2B} \quad \text{when } |x| < B$$

$$f_X(x) = 0 \quad \text{otherwise}$$



Normal Distribution

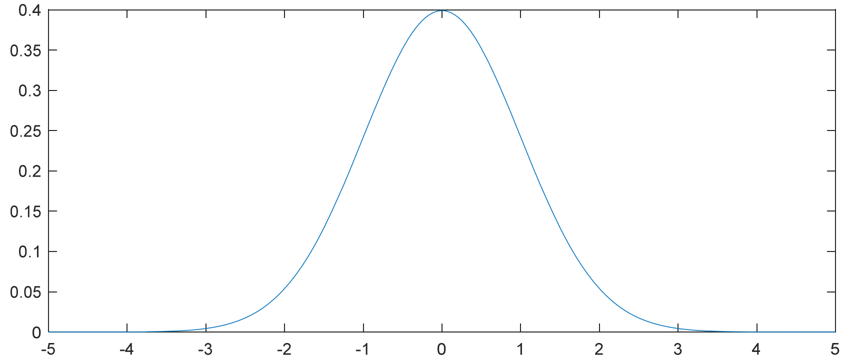
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$-\infty < x < \infty$

μ : mean

σ : standard deviation

when $\mu = 0$, $\sigma = 1$



Other examples are shown in Section 9.2.2

3. Expected Value

(discrete case)

$$E(g[X]) = \sum_n g[n]P(X = n) = \sum_n g[n]P_X(n)$$

(continuous case)

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

It is usually written as

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) dF_X(x)$$

9.1.2 Moments

Mean

$$\mu_X = E(X) = \sum_n n P_X(n)$$ (discrete case)

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$$ (continuous case)

Variance (var)

$$var_X = E\left((X-\mu_X)^2\right) = \sum_n (n-\mu_X)^2 P_X(n)$$ (discrete case)

$$var_X = E\left((X-\mu_X)^2\right) = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\, dx$$ (continuous case)

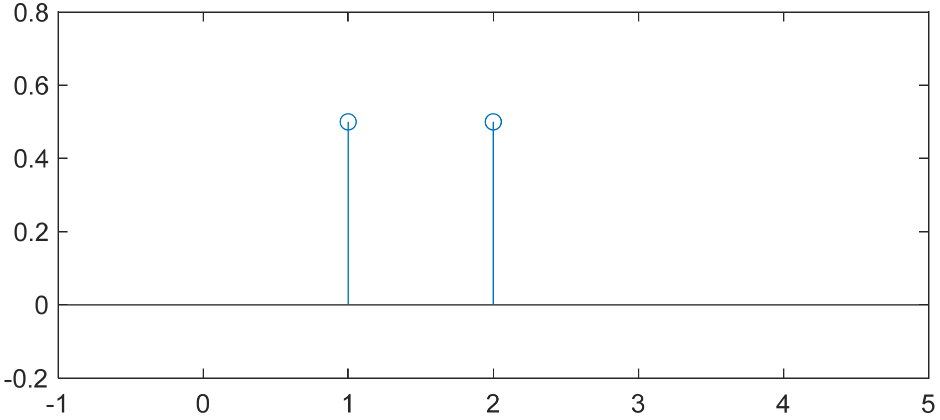
Standard Deviation (std)

$$\sigma_X = \sqrt{var_X} = \sqrt{E\left((X-\mu_X)^2\right)}$$ (discrete case)

$$\sigma_X = \sqrt{var_X} = \sqrt{E\left((X-\mu_X)^2\right)}$$ (continuous case)

[Example 1]

(1) $P_X(n)$

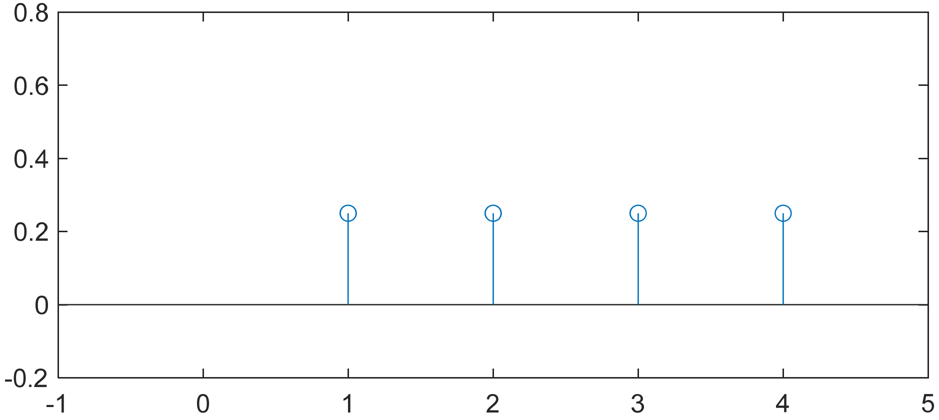


$$\mu_X = 1\cdot 0.5 + 2\cdot 0.5 = 1.5$$

$$var_X = 0.5(1-1.5)^2 + 0.5(2-1.5)^2 = 0.25$$

$$\sigma_X = \sqrt{var_X} = 0.5$$

(2) $P_X(n)$



$\mu_X = 2.5$ $var_X = 1.25$ $\sigma_X = \sqrt{5}/2$

(1) Moment (raw form)

$$\hat{m}_k = E\left(X^k\right) = \sum_n n^k P_X(n) \qquad \text{(discrete case)}$$

$$\hat{m}_k = E\left(X^k\right) = \int_{-\infty}^{\infty} x^k f_X(x)\,dx \qquad \text{(continuous case)}$$

(2) Moment (central form)

$$m_k = E\left((X-\mu_X)^k\right) = \sum_n (n-\mu_X)^k P_X(n) \qquad \text{(discrete case)}$$

$$m_k = E\left((X-\mu_X)^k\right) = \int_{-\infty}^{\infty} (x-\mu_X)^k f_X(x)\,dx \qquad \text{(continuous case)}$$

(3) Moment (standardized form)

$$v_k = \frac{m_k}{\sigma_X^k} = \frac{E\left((X-\mu_X)^k\right)}{\left[E\left((X-\mu_X)^2\right)\right]^{k/2}} = \frac{\sum_n (n-\mu_X)^k P_X(n)}{\left[\sum_n (n-\mu_X)^2 P_X(n)\right]^{k/2}}$$

(discrete case)

$$v_k = \frac{m_k}{\sigma_X^k} = \frac{E\left((X-\mu_X)^k\right)}{\left[E\left((X-\mu_X)^2\right)\right]^{k/2}} = \frac{\int_{-\infty}^{\infty} (x-\mu_X)^k f_X(x)\,dx}{\left[\int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\,dx\right]^{k/2}}$$

(continuous case)

Order	Moment (raw)	Moment (central)	Moment (standardized)
$k = 1$	Mean	0	0
$k = 2$		Variance	1
$k = 3$			Skewness
$k = 4$			Kurtosis
$k = 5$			Hyperskewness

Skewness (偏度)

It indicates the relative locations of the high-probability region and the long tail to the mean.

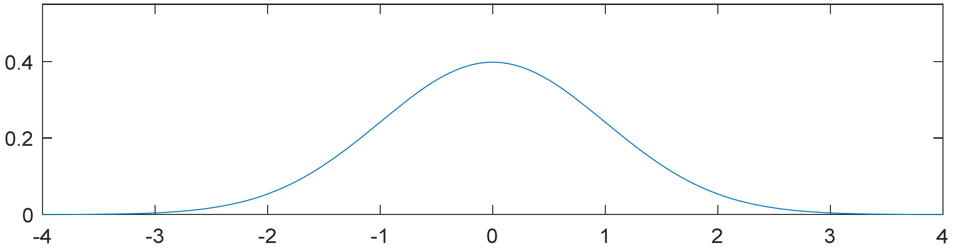
$$skewness = \frac{\sum_n (n-\mu_X)^3 P_X(n)}{\left[\sum_n (n-\mu_X)^2 P_X(n)\right]^{3/2}} \quad or \quad \frac{\int_{-\infty}^{\infty} (x-\mu_X)^3 f_X(x)\,dx}{\left[\int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\,dx\right]^{3/2}}$$

skewness > 0: The high-probability region is in the left of the mean.

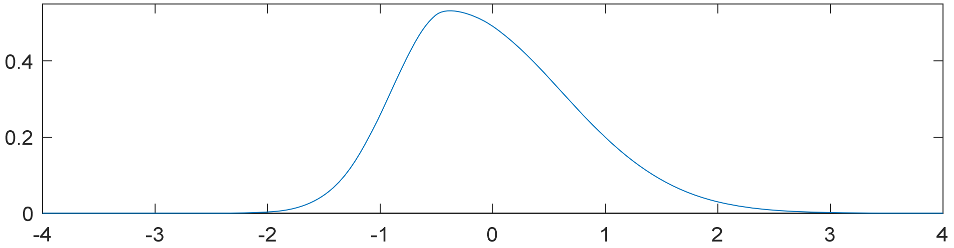
skewness < 0: The high-probability region is in the right of the mean.

skewness $= 0$: Symmetry

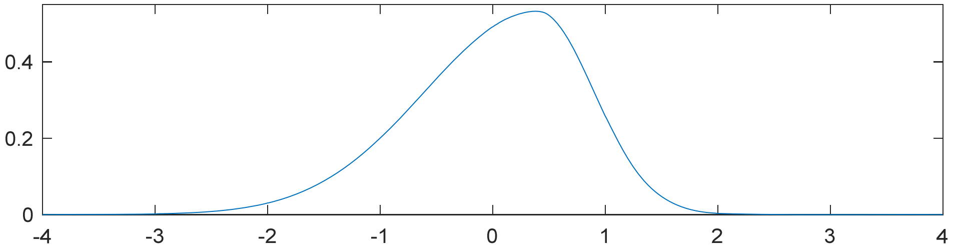
skewness = 0
mean = 0



skewness = 0.4992
mean = 0



skewness = -0.4992
mean = 0



Kurtosis (峰度)

It indicates how sharp the high-probability region is.

$$kurtosis = \frac{\sum_n (n-\mu_X)^4 P_X(n)}{\left[\sum_x (x-\mu_X)^2 P_X(x)\right]^2} \qquad or \ \frac{\int_{-\infty}^{\infty} (x-\mu_X)^4 f_X(x)dx}{\left[\int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx\right]^2}$$

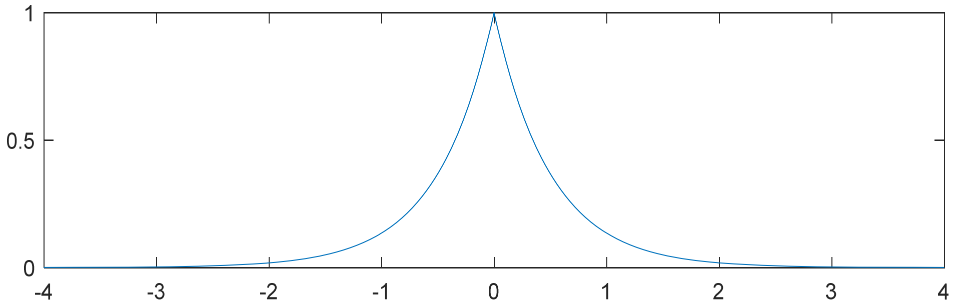
kurtosis ≥ 0

large kurtosis: The high probability region is sharp.

$$f_X(x) = e^{-2|x|}$$

kurtosis = 6

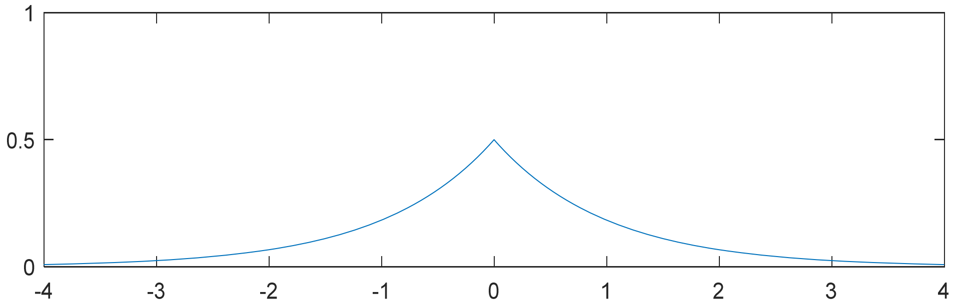
std = 0.707



$$f_X(x) = \frac{1}{2} e^{-|x|}$$

kurtosis = 5.86

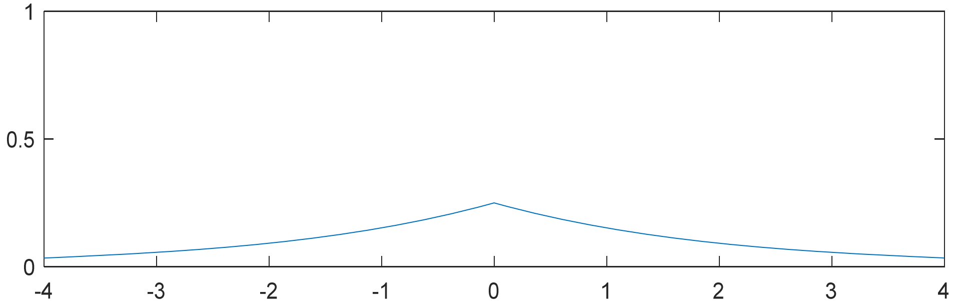
std = 1.414



$$f_X(x) = \frac{1}{4} e^{-|x|/2}$$

kurtosis = 4.38

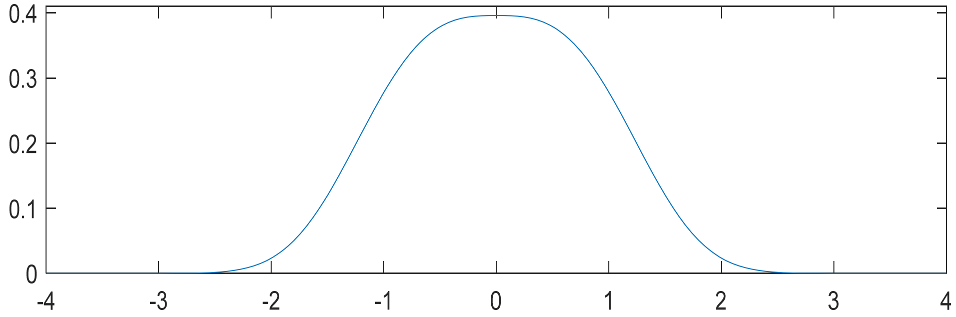
std = 2.647



$$f_X\left(x\right) = 0.396\, e^{-|x|^3/2^{3/2}}$$

kurtosis = 2.4184

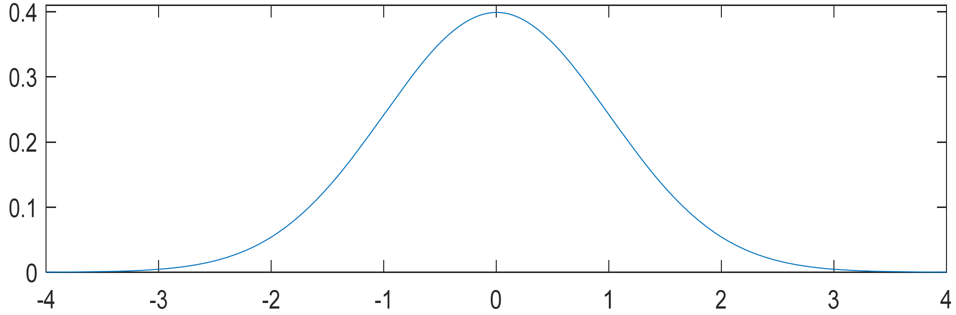
std = 0.864



$$f_X\left(x\right) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

kurtosis = 3

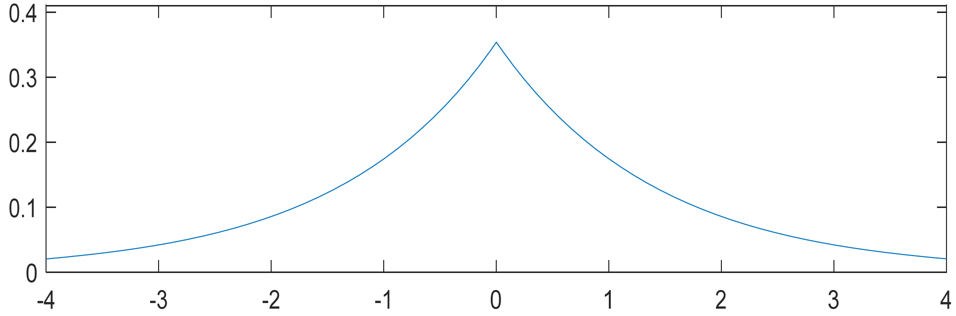
std = 1



$$f_X\left(x\right) = \frac{1}{2\sqrt{2}} e^{-|x|/\sqrt{2}}$$

kurtosis = 5.2903

std = 1.9726



[Example 2] Suppose that X is uniformly distributed in $x \in [0, 10]$. Determine the central moments, the standardized moments, the variance, the skewness, and the kurtosis of X.

(Solution): $f_X(x) = \frac{1}{10} \quad for\ 0 < x < 10$

$$f_X(x) = 0 \quad otherwise$$

$$\mu_X = \int_0^{10} x \frac{1}{10} dx = 5$$

$$\sigma_X = \sqrt{\int_0^{10} (x-5)^2 \frac{1}{10} dx} = \sqrt{\frac{1}{10}\int_{-5}^{5} x^2 \, dx} = \frac{5}{\sqrt{3}}$$

Central moment:

$$m_k = \int_0^{10} (x-5)^k \frac{1}{10} dx = \frac{1}{10}\int_{-5}^{5} x^k \, dx = \begin{cases} 0 & if\ \ k\ is\ odd \\ \dfrac{5^k}{k+1} & if\ \ k\ is\ even \end{cases}$$

Standardized moment:

$$v_k = \frac{m_k}{\sigma_X^k} = 3^{k/2}\frac{m_k}{5^k} = \begin{cases} 0 & if\ k\ is\ odd \\ \dfrac{3^{k/2}}{k+1} & if\ k\ is\ even \end{cases}$$

Variance: $var_X = m_2 = \frac{25}{3}$

Skewness = $v_3 = 0$

Kurtosis = $v_4 = \frac{9}{5}$

9.1.3 Correlation

Covariance

$$cov_{X,Y} = E\left((X-\mu_X)(Y-\mu_Y)\right)$$

$$cov_{X,Y} = \sum_x \sum_y (x-\mu_X)(y-\mu_Y)P_{X,Y}\left(x,y\right)$$ (discrete case)

$$cov_{X,Y} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{X,Y}\left(x,y\right)dxdy$$ (continuous case)

Note: (1) When $X = Y$,

$$cov_{X,X} = var_X$$

(2) $cov_{X,Y} = E\left(XY\right)-\mu_X\mu_Y$

Correlation

$$corr_{X,Y} = \frac{E\left((X-\mu_X)(Y-\mu_Y)\right)}{\sigma_X\sigma_Y} = \frac{cov_{X,Y}}{\sigma_X\sigma_Y}$$

$$corr_{X,Y} = \frac{\sum_x\sum_y (x-\mu_X)(y-\mu_Y)P_{X,Y}\left(x,y\right)}{\sqrt{\sum_x (x-\mu_X)^2 P_X\left(x\right)\sum_y (y-\mu_Y)^2 P_Y\left(y\right)}}$$

(discrete case)

$$corr_{X,Y} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f_{X,Y}\left(x,y\right) dxdy}{\sqrt{\int_{-\infty}^{\infty} (x-\mu_X)^2 f_X\left(x\right)dx \int_{-\infty}^{\infty} (y-\mu_Y)^2 f_Y\left(y\right)dy}}$$

(continuous case)

$$corr_{X,Y} = \frac{E\left((X-\mu_X)(Y-\mu_Y)\right)}{\sigma_X \sigma_Y}$$

Note: (1)

$$-1 \le corr_{X,Y} \le 1$$

(2) When $Y = cX + d$, c is a positive constant, d is a constant,

$$corr_{X,Y} = 1$$

(Proof): $Y = cX + d, \qquad \mu_Y = c\mu_X + d$

$$E\left((X-\mu_X)(Y-\mu_Y)\right) = E\left((X-\mu_X)(cX+d-c\mu_X-d)\right)$$

$$= E\left(c(X-\mu_X)^2\right) = c\,\mathrm{var}_X$$

$$\sigma_Y = \sqrt{E\left((cX+d-c\mu_X-d)^2\right)} = \sqrt{c^2 E\left((X-\mu_X)^2\right)} = |c|\sigma_X$$

$$corr_{X,Y} = \frac{c\,\mathrm{var}_X}{|c|\sigma_X\sigma_X} = \frac{c}{|c|}$$

(3) When $Y = cX + d$, c is a negative constant, d is a constant,

$$corr_{X,Y} = -1$$

(4) If Y is independent of X,

$$corr_{X,Y} = \frac{E\left((X-\mu_X)(Y-\mu_Y)\right)}{\sigma_X \sigma_Y} = \frac{E\left(X-\mu_X\right)E\left(Y-\mu_X\right)}{\sigma_X \sigma_Y} = 0$$

(i) Full Correlation:

$$|corr_{X,Y}| \geq 0.9$$

(ii) High Correlation:

$$0.6 \leq |corr_{X,Y}| < 0.9$$

(iii) Middle Correlation:

$$0.3 \leq |corr_{X,Y}| < 0.6$$

(iv) Low Correlation:

$$|corr_{X,Y}| < 0.3$$

(v) Positive Correlation:

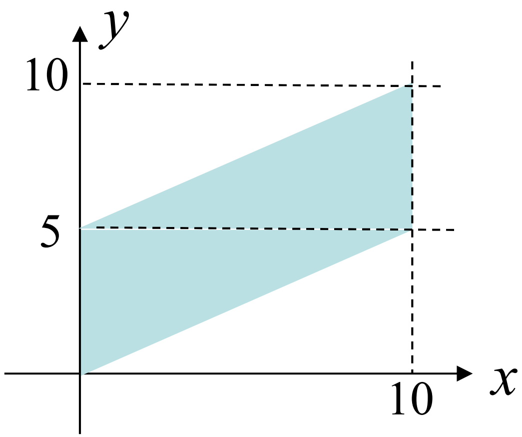
$$corr_{X,Y} > 0$$

(vi) Negative Correlation:

$$corr_{X,Y} < 0$$

[Example 3] Determine the Covariance and the Correlation of X and Y if

$$f_{X,Y}(x,y)=\begin{cases}\dfrac{1}{50} & 0<x<10,\quad \dfrac{x}{2}<y<\dfrac{x}{2}+5\\ 0 & otherwise\end{cases}$$



(Solution):

Note that

$$f_X(x)=\int f_{X,Y}(x,y)dy=\frac{1}{50}\int_{x/2}^{x/2+5}dy=\frac{1}{10}$$

$$f_Y(y)=\int f_{X,Y}(x,y)dx=\frac{1}{50}\int_{\max(0,2y-10)}^{\min(10,2y)}dx$$

$$f_Y(y)=\begin{cases}y/25 & for\ 0<y<5\\ (10-y)/25 & for\ 5<y<10\\ 0 & otherwise\end{cases}$$

$$\mu_X = \int x f_X(x)\,dx = \frac{1}{10}\int_0^{10} x\,dx = 5$$

$$\mu_Y = \int y f_Y(y)\,dy = \int_0^5 \frac{y^2}{25}dy + \int_5^{10} \frac{10y - y^2}{25}dy = 5$$

To determine the covariance,

$$\begin{aligned} cov_{X,Y} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f_{X,Y}(x,y)\,dxdy \\ &= \frac{1}{50}\int_0^{10}\int_{x/2}^{x/2+5} (x-5)(y-5)\,dydx \\ &= \frac{1}{50}\int_0^{10} (x-5)\left(\frac{5x-25}{2}\right)dx \\ &= \frac{25}{6} \end{aligned}$$

To determine the correlation, first, we determine the variance

$$\sigma_X^2 = \int (x-\mu_X)^2 f_X(x)\,dx = \frac{1}{10}\int_0^{10}(x-5)^2\,dx = \frac{25}{3}$$

$$\sigma_Y^2 = \int (y-\mu_Y)^2 f_Y(y)\,dy = \int_0^5 \frac{(y-5)^2 y}{25}dy + \int_5^{10}\frac{(y-5)^2(10-y)}{25}dy$$

$$= \int_0^5 \frac{(y-5)^2 y}{25}dy + \int_5^{0}\frac{(y_1-5)^2 y_1}{25}(-dy_1) = 2\int_0^5 \frac{(y-5)^2 y}{25}dy$$

(set y_1 = 10-y)

$$= \frac{25}{6}$$

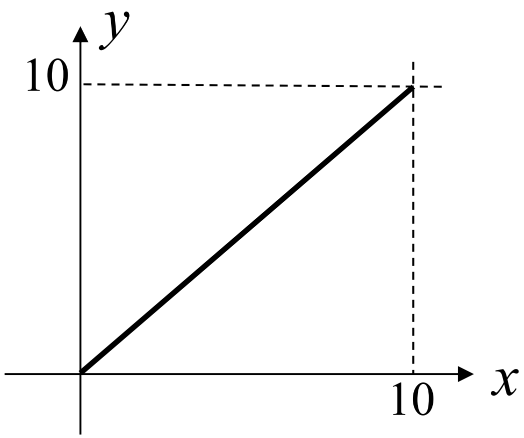
Therefore, $\sigma_X = \frac{5}{\sqrt{3}}$, $\sigma_Y = \frac{5}{\sqrt{6}}$

$$corr_{X,Y} = \frac{cov_{X,Y}}{\sigma_X \sigma_Y} = \frac{25}{6}\frac{\sqrt{18}}{25} = \frac{1}{\sqrt{2}} = 0.707$$

(X and Y are highly correlated)

[Example 4] Note that if

$$f_{X,Y}(x,y)=\begin{cases}\frac{1}{10}\delta(x-y) & 0<x<10\\ 0 & otherwise\end{cases}$$



then

$$f_X(x)=\frac{1}{10}\int_{-\infty}^{\infty}\delta(x-y)dy=\frac{1}{10}\int_{-\infty}^{\infty}\delta(y-x)dy$$

(page 343(1))

$$=\frac{1}{10}\int_{-\infty}^{\infty}\delta(y)dy=\frac{1}{10}$$

Similarly

$$f_Y(y)=\frac{1}{10}\int_{-\infty}^{\infty}\delta(x-y)dx=\frac{1}{10}$$

$$\mu_X=\frac{1}{10}\int_0^{10}xdx=5 \qquad \mu_Y=\frac{1}{10}\int_0^{10}ydy=5$$

$$cov_{X,Y} = \int_0^{10}\int_{-\infty}^{\infty}(x-5)(y-5)\frac{1}{10}\delta\left(x-y\right)dydx$$

$$= \int_0^{10}\int_{-\infty}^{\infty}(x-5)(y-5)\frac{1}{10}\delta\left(y-x\right)dydx$$

$$= \frac{1}{10}\int_0^{10}(x-5)^2\,dx = \frac{25}{3}$$

(page 344(2))

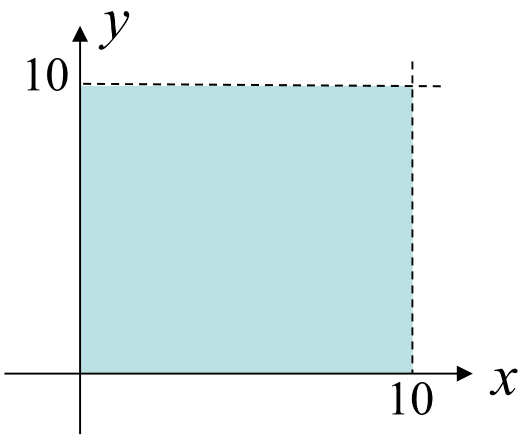
$$\sigma_X = \sqrt{\frac{1}{10}\int_0^{10}(x-5)^2\,dx} = \sqrt{\frac{25}{3}}$$

$$\sigma_Y = \sqrt{\frac{1}{10}\int_0^{10}(y-5)^2\,dy} = \sqrt{\frac{25}{3}}$$

$$corr_{X,Y} = \frac{cov_{X,Y}}{\sigma_X\sigma_Y} = 1$$

[Example 5] If

$$f_{X,Y}(x,y)=\begin{cases}\frac{1}{100} & 0<x<10 \; and \; 0<y<10 \\ 0 & otherwise\end{cases}$$



then

$$f_X(x)=\int_0^{10}\frac{1}{100}dy=\frac{1}{10} \qquad f_Y(y)=\int_0^{10}\frac{1}{100}dx=\frac{1}{10}$$

$$\mu_X=\frac{1}{10}\int_0^{10}xdx=5 \qquad \mu_Y=\frac{1}{10}\int_0^{10}ydy=5$$

$$cov_{X,Y}=\int_0^{10}\int_0^{10}(x-5)(y-5)\frac{1}{100}dydx=0$$

(odd symmetry with respect to (5, 5))

$$corr_{X,Y}=\frac{cov_{X,Y}}{\sigma_X\sigma_Y}=0$$