基礎機率和統計

黃志勝

義隆電子 人工智慧研發部

國立陽明交通大學 AI學院 合聘助理教授

基礎機率和統計

- 基礎數學
- 基礎統計學

Deep Learning

An MIT Press book

Ian Goodfellow and Yoshua Bengio and Aaron Courville

Exercises Lectures External Links

The Deep Learning textbook is a resource intended to help students and practitioners enter the field of mac in general and deep learning in particular. The online version of the book is now complete and will remain online for free.

The deep learning textbook can now be ordered on Amazon.

For up to date announcements, join our mailing list.

Citing the book

To cite this book, please use this bibtex entry:

```
@book{Goodfellow-et-al-2016,
    title={Deep Learning},
    author={Ian Goodfellow and Yoshua Bengio and Aaron Courville},
    publisher={MIT Press},
    note={\url{http://www.deeplearningbook.org}},
    year={2016}
}
```

To write your own document using our LaTeX style, math notation, or to copy our notation page, downloac files.

Errata in published editions

Deep Learning

Reference book: Deep Learning by Goodfellow (2016)

基礎機率和統計

Deep Learning

Part I: Applied Math and Machine Learning Basics

2 Linear Algebra

3 Probability and Information Theory

4 Numerical Computation

5 Machine Learning Basics

基礎機率和統計

Deep Learning

Part I: Applied Math and Machine Learning Basics

Part II: Modern Practical Deep Networks

Part III: Deep Learning Research

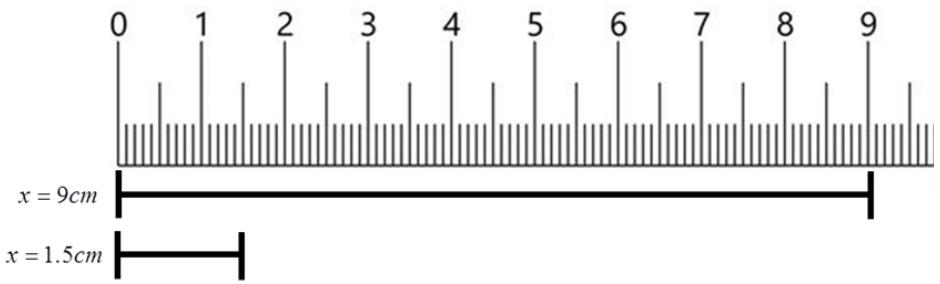
Basic linear algebra for Learning algorithms

- Scalar
- Vector
- Matrix
- Tensor
- Matrix Computation
- Matrix Transpose
- Inverse matrix

Scalar and vector

- **Scalar**

A quantity representable by a mathematical scalar, usually the field of real numbers, such as speed, length, mass



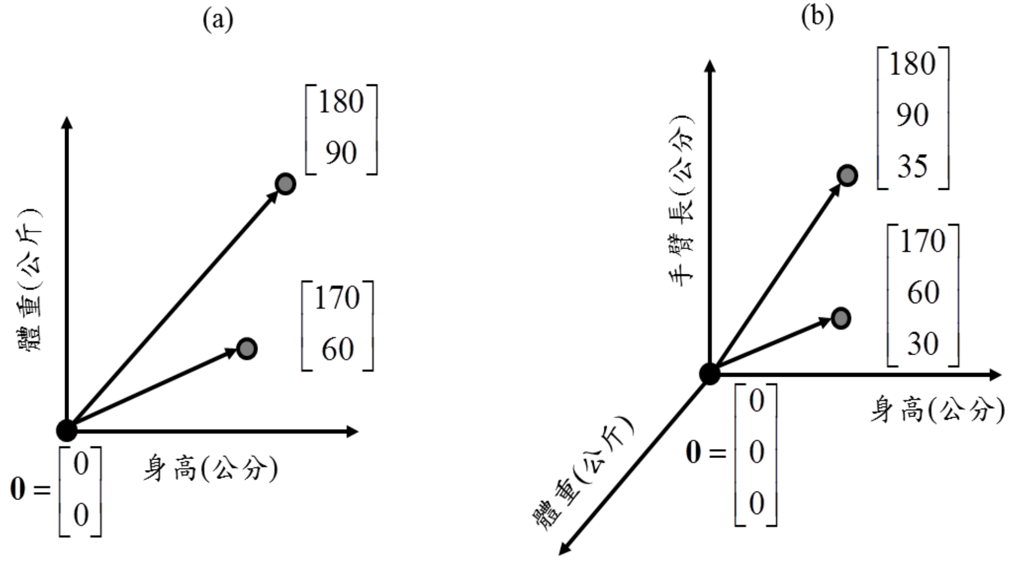
Vector

Array: Magnitude and direction

Magnitude $\|\vec{x}\| > \|\vec{y}\|$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \in \mathbb{R}^d$$

$\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^{d} x_i^2}$: Euclidean Distance



Matrix

$$\boldsymbol{X} = [\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_m] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{d1} & x_{d2} & \cdots & x_{dm} \end{bmatrix} \in \mathbb{R}^{\boldsymbol{d \times m}}$$

$$\boldsymbol{x_1} = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{d1} \end{bmatrix}, \boldsymbol{x_2} = \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{d2} \end{bmatrix}, \ldots, \boldsymbol{x_m} = \begin{bmatrix} x_{1m} \\ x_{2m} \\ \vdots \\ x_{dm} \end{bmatrix}$$

Tensor

- **Multidimensional array**

0-dimensional tensor: scalar

0.5

2-dimensional tensor: matrix

0.1	2	1
-3	1.2	5.1
-0.2	1	0

3×3

1-dimensional tensor: vector

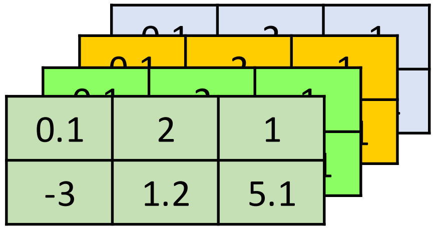
0.1
1
-3
-0.5

4×1

0.1	2	1

1×3

3-dimensional tensor



2×3×4

Tensor

$$\mathbf{X}_{3\times2\times2} = \left[R^{2\times2}, G^{2\times2}, B^{2\times2}\right] = \left[\begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}, \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}, \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\right]_{3\times2\times2}$$

$$\mathbf{X}_{4\times3\times2\times2}$$

$$= \left[\left[R_1^{2\times2}, G_1^{2\times2}, B_1^{2\times2}\right], \left[R_2^{2\times2}, G_2^{2\times2}, B_2^{2\times2}\right], \left[R_3^{2\times2}, G_3^{2\times2}, B_3^{2\times2}\right], \left[R_4^{2\times2}, G_4^{2\times2}, B_4^{2\times2}\right]\right]$$

$$\mathbf{X} \in \mathbb{R}^{n \times order \times ch \times w \times h}$$

Matrix Computation

- Scalar multiplication: $a\boldsymbol{x} = a\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} = \begin{bmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_d \end{bmatrix}$
- Vector multiplication: $\boldsymbol{xy}$, but $\left(\boldsymbol{x} \in \mathbb{R}^{d\times 1}, \boldsymbol{y} \in \mathbb{R}^{1\times d}\right)$ or $\left(\boldsymbol{x} \in \mathbb{R}^{1\times d}, \boldsymbol{y} \in \mathbb{R}^{d\times 1}\right)$
- $\left(\boldsymbol{x} \in \mathbb{R}^{d\times 1}, \boldsymbol{y} \in \mathbb{R}^{1\times d}\right)$: $\boldsymbol{x}\mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_d \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \dots & x_1 y_d \\ x_2 y_1 & x_2 y_2 & \dots & x_2 y_d \\ \vdots & \vdots & \ddots & \vdots \\ x_d y_1 & x_d y_1 & \dots & x_d y_d \end{bmatrix}$
- $\left(\boldsymbol{x} \in \mathbb{R}^{1\times d}, \boldsymbol{y} \in \mathbb{R}^{d\times 1}\right)$: $\boldsymbol{xy} = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix} = \left[x_1 y_1 + x_2 y_2 + \cdots + x_d y_d\right] = \sum_{i=1}^{d} x_i y_i$

Matrix Computation

$$\mathbf{Z}_{d\times n} = \mathbf{X}_{d\times m}\mathbf{Y}_{m\times n} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & \boxed{z_{22}} & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{d1} & z_{d2} & \cdots & z_{dn} \end{bmatrix}_{d\times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{d1} & x_{d2} & \cdots & x_{dm} \end{bmatrix}_{d\times m} \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}_{m\times n}$$

$$z_{22} = \begin{bmatrix} x_{21} & x_{22} & \cdots & x_{2m} \end{bmatrix} \begin{bmatrix} y_{12} \\ y_{22} \\ \vdots \\ y_{m2} \end{bmatrix} = \sum_{k=1}^{m} x_{2k} y_{k2}$$

Hadamard乘積(Hadamard product)

- 矩陣必須一樣大

$$\mathbf{Z}_{m\times n}=\mathbf{X}_{m\times n}\odot\mathbf{Y}_{m\times n}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}\odot\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix}=\begin{bmatrix} x_{11}y_{11} & x_{12}y_{12} & \cdots & x_{1n}y_{1n} \\ x_{21}y_{21} & x_{22}y_{22} & \cdots & x_{2n}y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1}y_{m1} & x_{m2}y_{m2} & \cdots & x_{mn}y_{mn} \end{bmatrix}$$

$$\mathbf{X}_{2\times1}=\begin{bmatrix} 1 \\ 2 \end{bmatrix},\ \mathbf{Y}_{2\times1}=\begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$$\mathbf{Z}_{2\times1}=\mathbf{X}_{2\times1}\odot\mathbf{Y}_{2\times1}=\begin{bmatrix} 1 \\ 2 \end{bmatrix}\odot\begin{bmatrix} 3 \\ 4 \end{bmatrix}=\begin{bmatrix} 1\times3 \\ 2\times4 \end{bmatrix}=\begin{bmatrix} 3 \\ 8 \end{bmatrix}$$

Inverse matrix

$$1/2 = 0.5$$

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} / 2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \times 2^{-1} = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix}$$

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} / \begin{bmatrix} 2 \\ 3 \end{bmatrix} = ?$$

- Inverse Matrix is used for this situation.

Inverse matrix

- Matrix must be square. $X \in \mathbb{R}^{d} = \mathbb{R}^{d\times d}$
- X is a square matrix, and there is a Y let

$$XY = YX = I$$

X is an invertible matrix, full rank, and nonsingular matrix.

$$Y = X^{-1}$$

I denote identity matrix.

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

- If X is singular (Not-Invertible), $\det(X) = 0$.

Inverse matrix

- If X is singular (Not-Invertible), $\det(X) = 0$.

假設$X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$且可逆，則$X^{-1} = \frac{1}{\det(X)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

所以

$$\det(X) = 0 \Rightarrow X^{-1} = \frac{1}{0}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} ?$$

二階矩陣的反矩陣推導

$$\mathbf{A}^{-1} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{cases} ax_{11} + bx_{21} = 1 \\ cx_{11} + dx_{21} = 0 \end{cases} \Rightarrow \begin{cases} ax_{11} + bx_{21} = 1 \Rightarrow \dfrac{-ad+bc}{c}x_{21} = 1 \\ x_{11} = -\dfrac{d}{c}x_{21} \end{cases} \Rightarrow \begin{cases} x_{11} = \dfrac{-d}{bc-ad} \\ x_{21} = \dfrac{c}{bc-ad} \end{cases}$$

$$\begin{cases} ax_{12} + bx_{22} = 0 \\ cx_{12} + dx_{22} = 1 \end{cases} \Rightarrow \begin{cases} x_{22} = -\dfrac{a}{b}x_{12} \\ bx_{12} + dx_{22} = 1 \Rightarrow \dfrac{bc-ad}{b}x_{12} = 1 \end{cases} \Rightarrow \begin{cases} x_{12} = \dfrac{b}{bc-ad} \\ x_{22} = \dfrac{-a}{bc-ad} \end{cases}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \begin{bmatrix} \dfrac{-d}{bc-ad} & \dfrac{b}{bc-ad} \\ \dfrac{c}{bc-ad} & \dfrac{-a}{bc-ad} \end{bmatrix} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{\det(\mathbf{A})}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

基礎統計學

基礎統計學- Distribution

- **離散數據:**

資料→數量

1, 1→ 2

2, 2 → 2

3 → 1

4, 4, 4 → 4

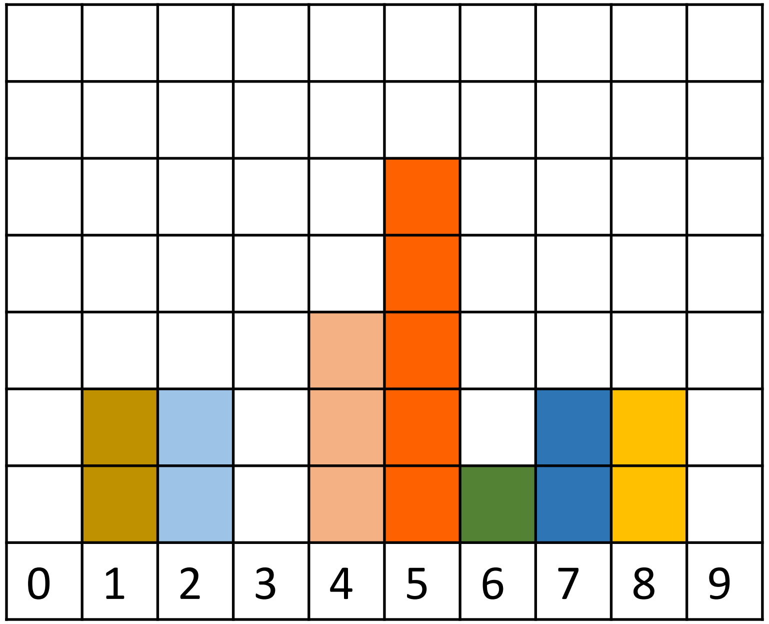
5, 5, 5, 5, 5 → 5

6 → 1

7, 7 → 2

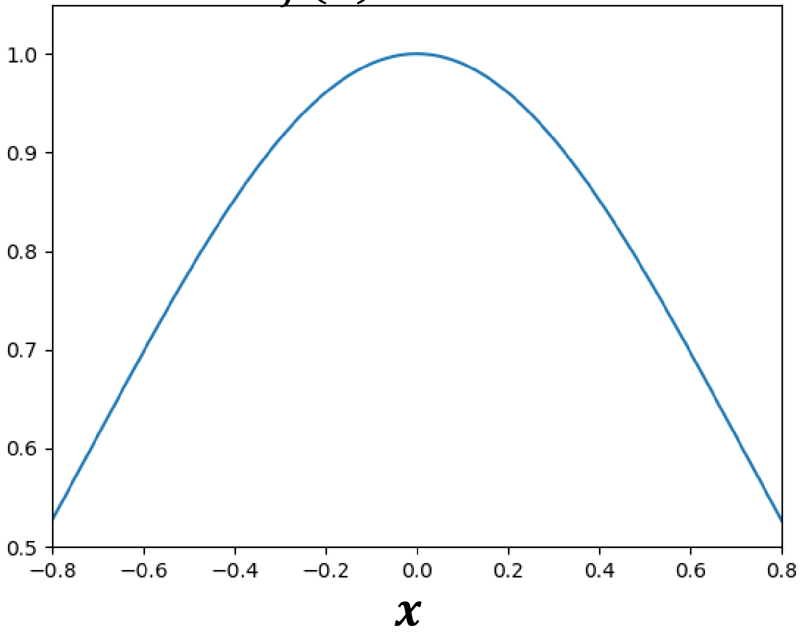
8, 8 → 2

9, 9 → 2



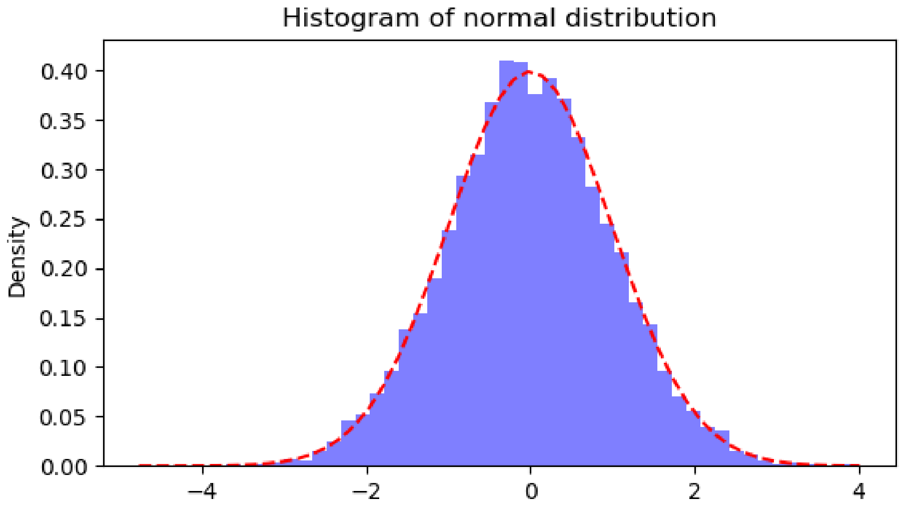
連續函數

$f(x) = e^{-x^2}$



基礎統計學- Distribution

Normal (Gaussian) Distribution



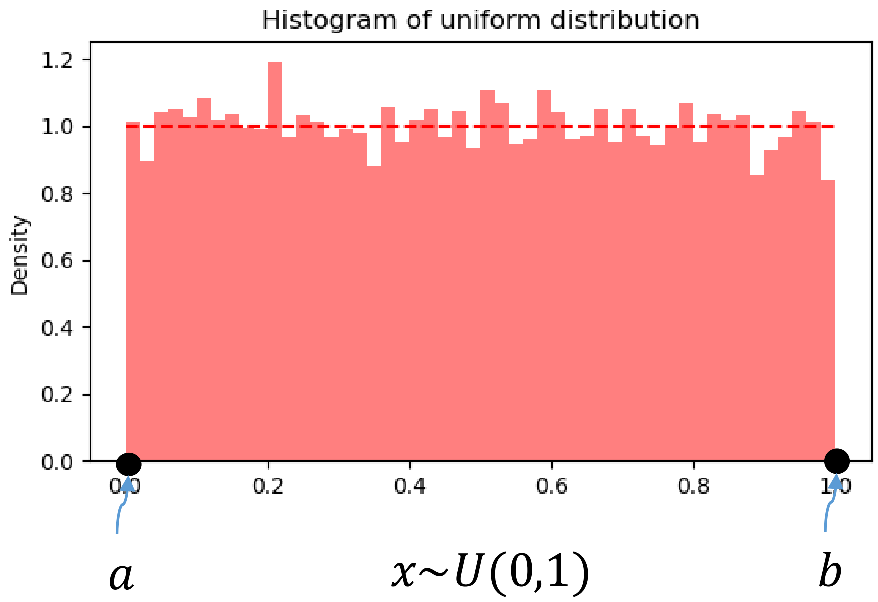
μ: 平均數
σ: 標準差

$$x \sim N(\mu, \sigma)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mu = 0, \sigma = 1$$
$$f(x = 0) = 0.3989$$

Uniform Distribution



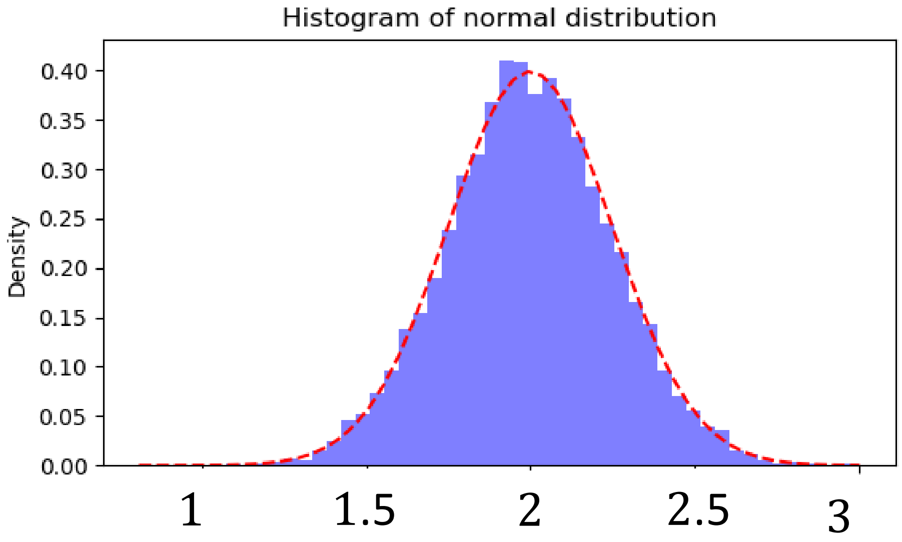
$$x \sim U(0,1)$$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & o.w. \end{cases}$$

$$a = 0, b = 1$$
$$f(x = 0.5) = 1$$

基礎統計學- Distribution

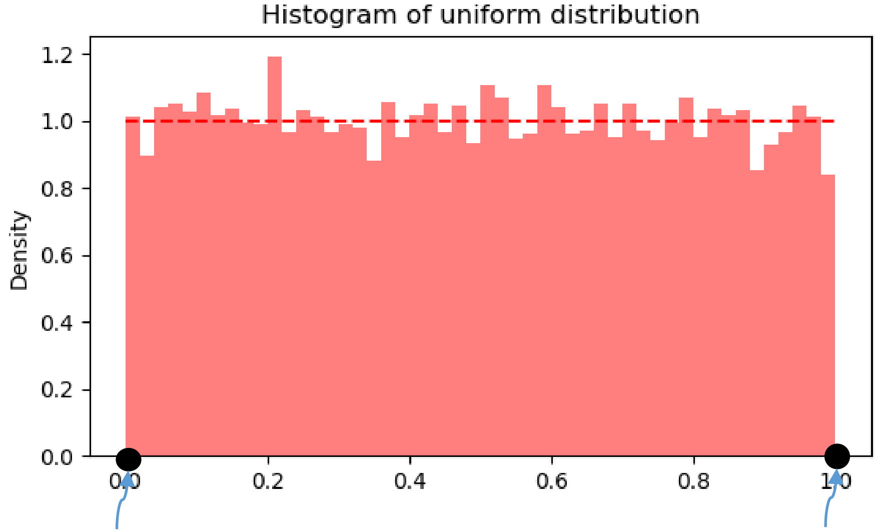
Normal (Gaussian) Distribution



$x \sim N(\mu = 2, \sigma = 0.1)$

手機生命週期
平均壽命: 兩年
標準差: 0.1年

Uniform Distribution



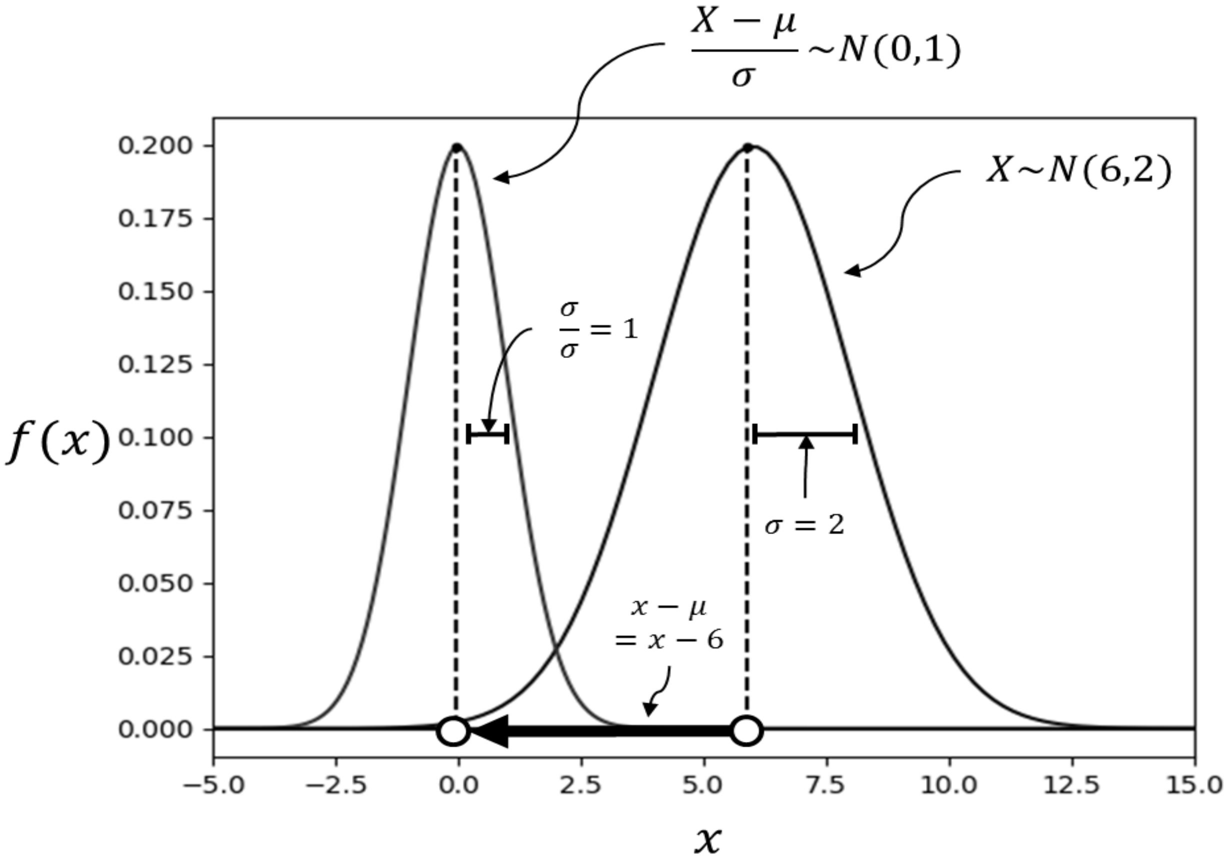
a $x \sim U(0,1)$ b

投骰子

點數1~6:
每個點數出現的機率都是一樣，
1/6

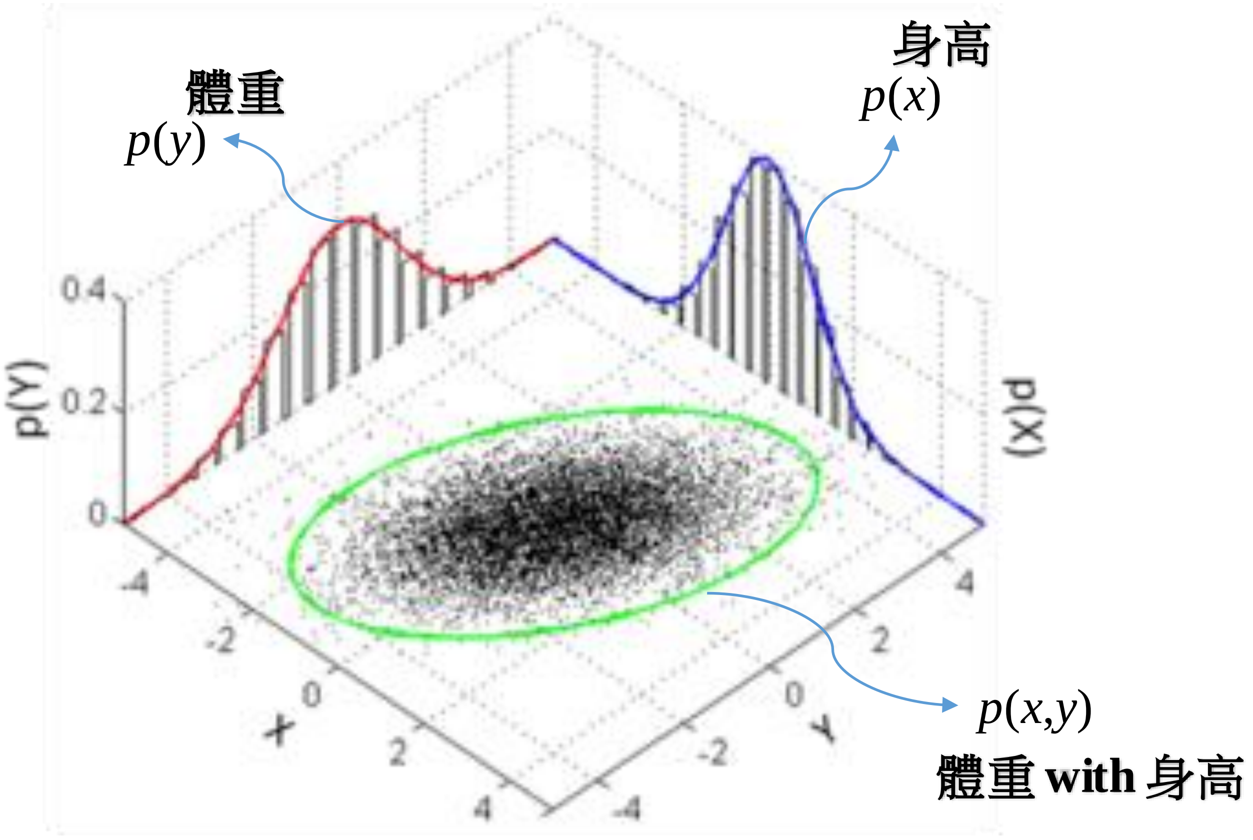
基礎統計學-標準常態分布

藉由z-score的操作可以將常態分布$N(\mu,\sigma)$轉換成標準常態分布$N(0,1)$



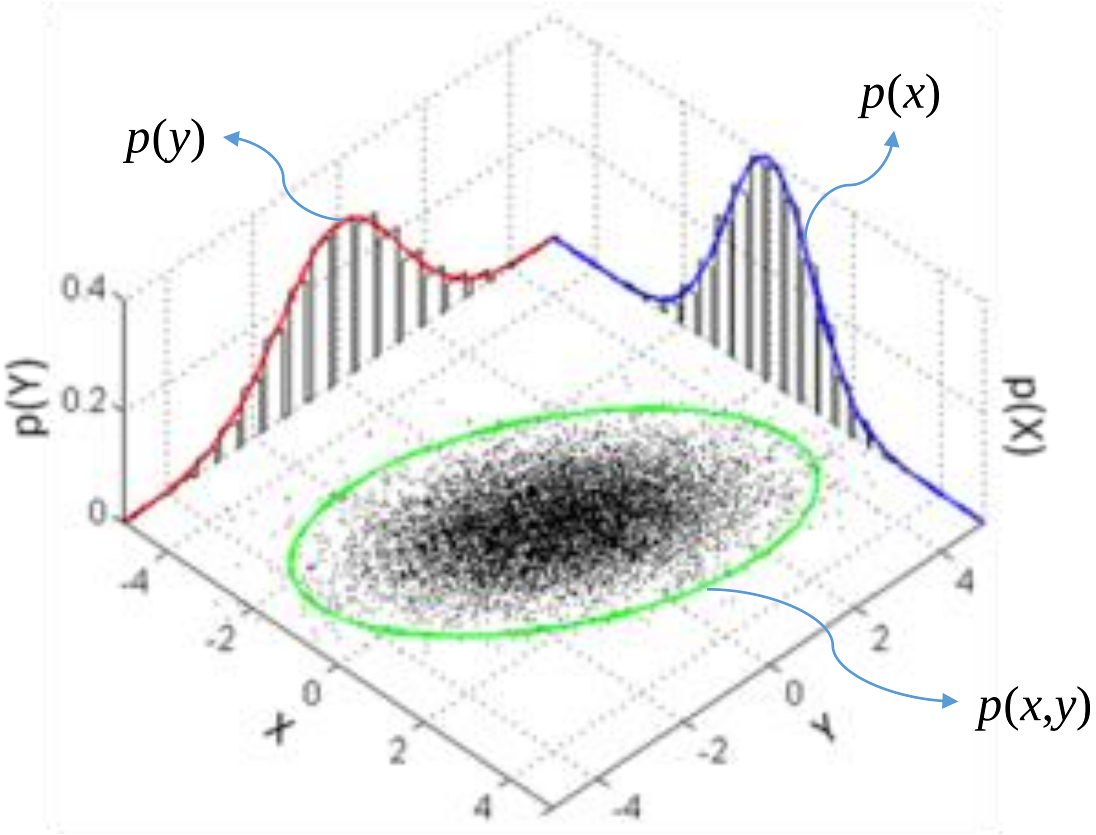
基礎統計學- Distribution
Joint probability

X: 身高
Y: 體重
(X,Y)=(180, 100)
(200, 120)
(160, 50)
(170, 60)
…



https://upload.wikimedia.org/wikipedia/commons/thumb/9/95/Multivariate_normal_sample.svg/300px-Multivariate_normal_sample.svg.png

基礎統計學- Distribution
Joint probability



https://upload.wikimedia.org/wikipedia/commons/thumb/9/95/Multivariate_normal_sample.svg/300px-Multivariate_normal_sample.svg.png

Discrete: probability mass function.

Probability:

$$p(x) = \sum_i p(x_i) = 1$$

Joint probability:

$$p(x, y) = \sum_i \sum_j p(x_i, y_i) = 1$$

Continuous: probability density function.

Probability:

$$p(x) = \int p(x_i) dx = 1$$

Joint probability:

$$p(x, y) = \iint p(x_i, y_i) = 1$$

基礎統計學- Distribution (multivariate normal distribution)

Single-variate normal distribution:

$$x \sim N(\mu, \sigma)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

multivariate normal distribution:

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \Sigma)$$

$$f(\boldsymbol{x}) = (2\pi)^{\frac{-p}{2}} |\Sigma|^{\frac{-1}{2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}$$

基礎統計學-期望值 (Expectation)

- 其目的是希望由過去蒐集的資料中，在事件未發生前，進行統計推論得到一個期望(預期)出現的值。

EX: 歷史數據顯示小明數學考試為100, 94, 90, 95, 90

那如果沒有任何影響下，我們期望小明下次數學考試的成績是多少?

$$\frac{1}{5}(100+94+90+95+90)=93.8$$

所以沒有特殊情況一般說的期望值=平均數

基礎統計學-期望值 (Expectation)

- 其目的是希望由過去蒐集的資料中，在事件未發生前，進行統計推論得到一個期望(預期)出現的值。

- 期望值的公式:

$$E(X) = \sum_{x} x f(x)$$

剛剛範例: $x = 100, 94, 90, 95, 90;\ f(x) = \frac{1}{5}$

$f(x)$: Uniform Distribution

常態分布的期望值

$$x \sim N(\mu, \sigma), f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$z = \frac{x-\mu}{\sigma} \sim N(0,1), f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$E(Z) = \int_{-\infty}^{\infty} z f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \Big|_{-\infty}^{\infty} = 0$$

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma} E(X) - \frac{\mu}{\sigma} = 0 \Rightarrow E(X) = \mu$$

基礎統計學-變異數(Variance)

- 變異數就是計算隨機變數和其中心點(平均數)取完差值後算平方的期望值。
- 變異數都是用在看資料的分散量大小使用，變異數越大代表資料分散程度越大。

$$var(x) = \frac{1}{n}\sum_{i=1}^{n}(x-\mu)^2$$

$$var\ (x) = E[(x-\mu)^2]$$

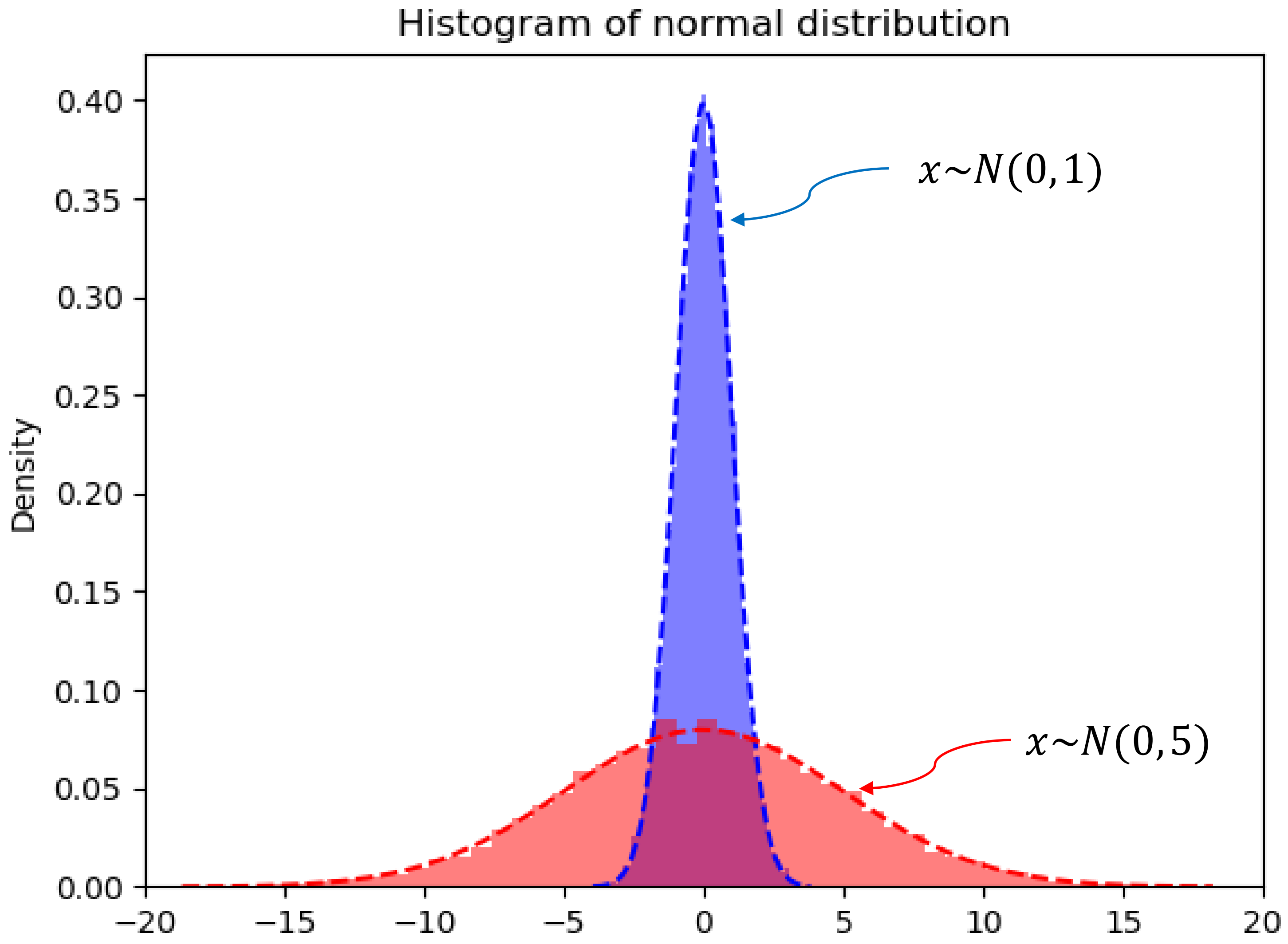
標準差: $std(x) = \sqrt{var(x)}$

基礎統計學-變異數(Variance)

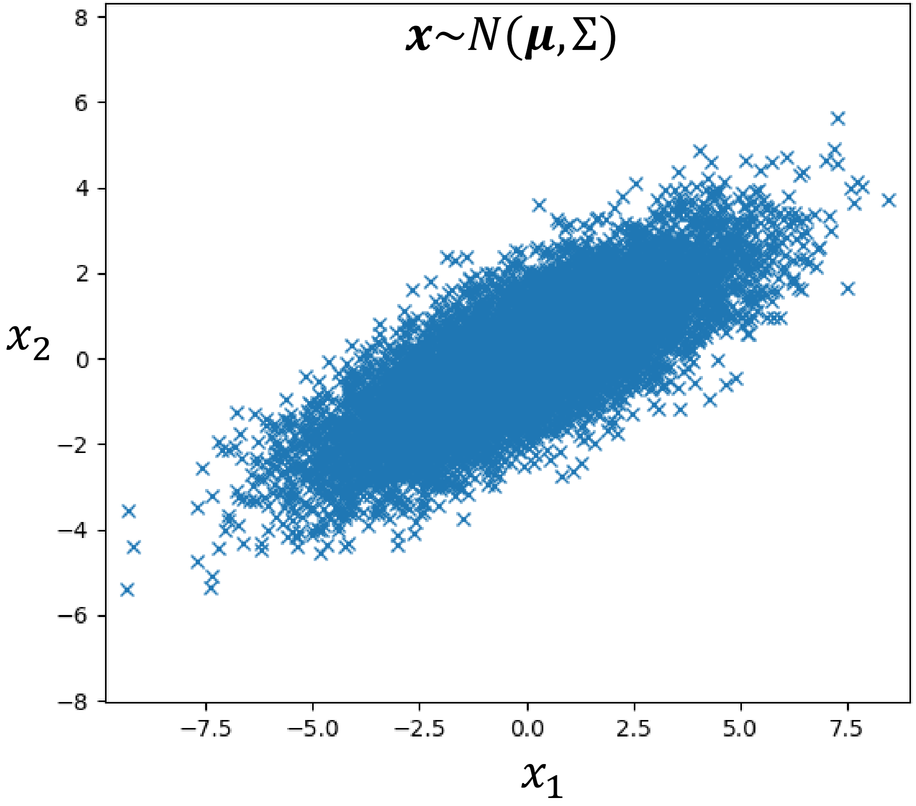
變異數(Variance)用來表示資料分布的瘦和胖，變異量越大資料越胖。

$$x \sim N(0,1)$$
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$x \sim N(0,5)$$
$$f(x) = \frac{1}{5\sqrt{2\pi}} e^{-\frac{x^2}{50}}$$



基礎統計學-共變異數



- 當資料維度>2，這時候就不能只用變異量來描述整筆資料，需要在加上

<u>共變異數(斜方差，Covariance)</u>

$$cov(x,y) = E[(x - E(x))(y - E(y))]$$
$$var(x) = cov(x,x)$$

共變異數則是用來表示兩個變數之間的共線性強度。

基礎統計學-共變異數

範例 1	
身高(*X*)，單位公分	體重(*Y*)，單位公斤
180	80
175	70
170	70
165	70
160	50

範例2	
價錢(*X*)，單位百萬	坪數(*Y*)
20	60
15	55
13	50
10	45
7	40

身高和體重的共變異數 $\frac{1}{5}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y) = 300/5 = 60$。

房子價錢和坪數的共變異數 $\frac{1}{5}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y) = 155/5 = 31$。

基礎統計學-共變異數

範例 1		範例2	
身高(*X*)，單位公分	體重(*Y*)，單位公斤	價錢(*X*)，單位百萬	坪數(*Y*)
180	80	20	60
175	70	15	55
170	70	13	50
165	70	10	45
160	50	7	40

男生的身高和體重的共變異數 60 。

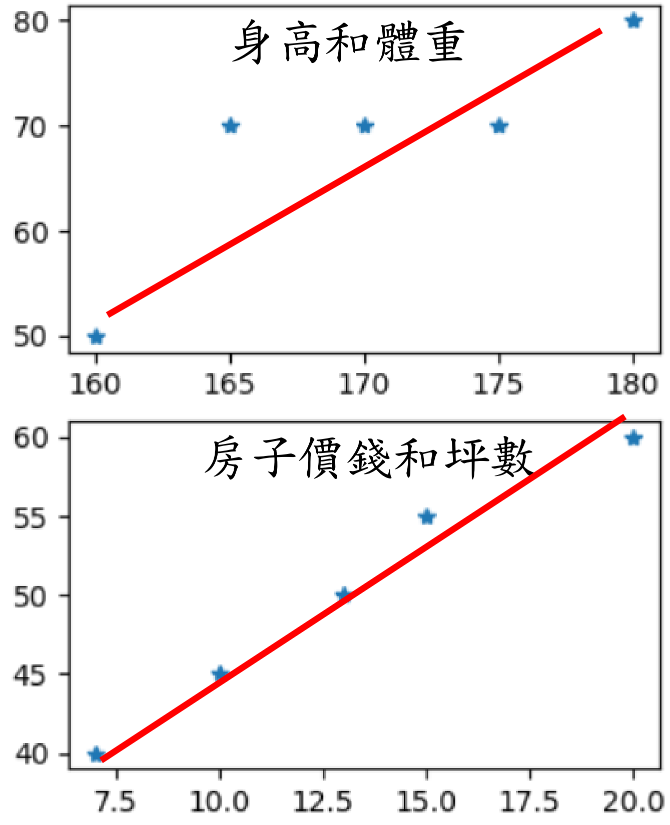
房子價錢和坪數的共變異數 31。

問題在?

ANS: 單位

範例 1: $cov(x, y)$(單位: 公分 * 公斤)

範例 2: $cov(x, y)$(單位: 百萬 * 坪數)



基礎統計學-相關係數

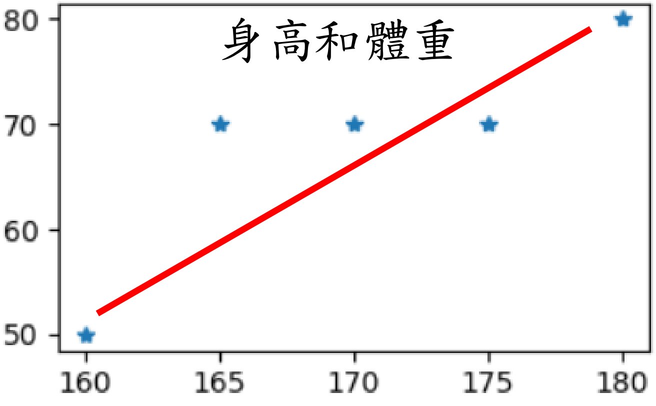
$$\rho = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sqrt{E\left((X-\mu_X)^2\right)}\sqrt{E\left((Y-\mu_Y)^2\right)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

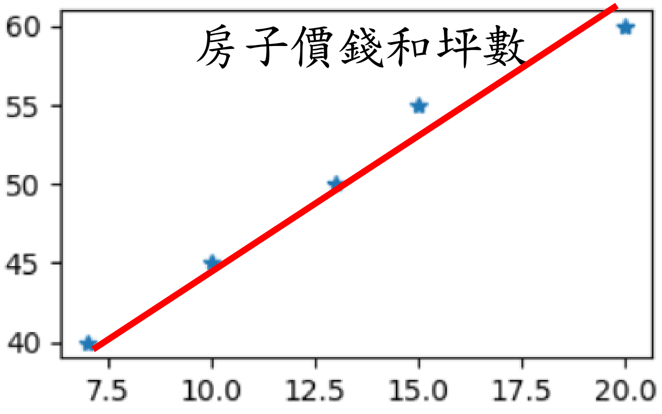
$$-1 \le \rho \le 1$$

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{Cov(X,Y)(\text{單位:}\cancel{\text{公分}} \times \cancel{\text{公斤}})}{\sigma_X(\text{單位:}\cancel{\text{公分}}) \times \sigma_Y(\text{單位:}\cancel{\text{公斤}})}$$

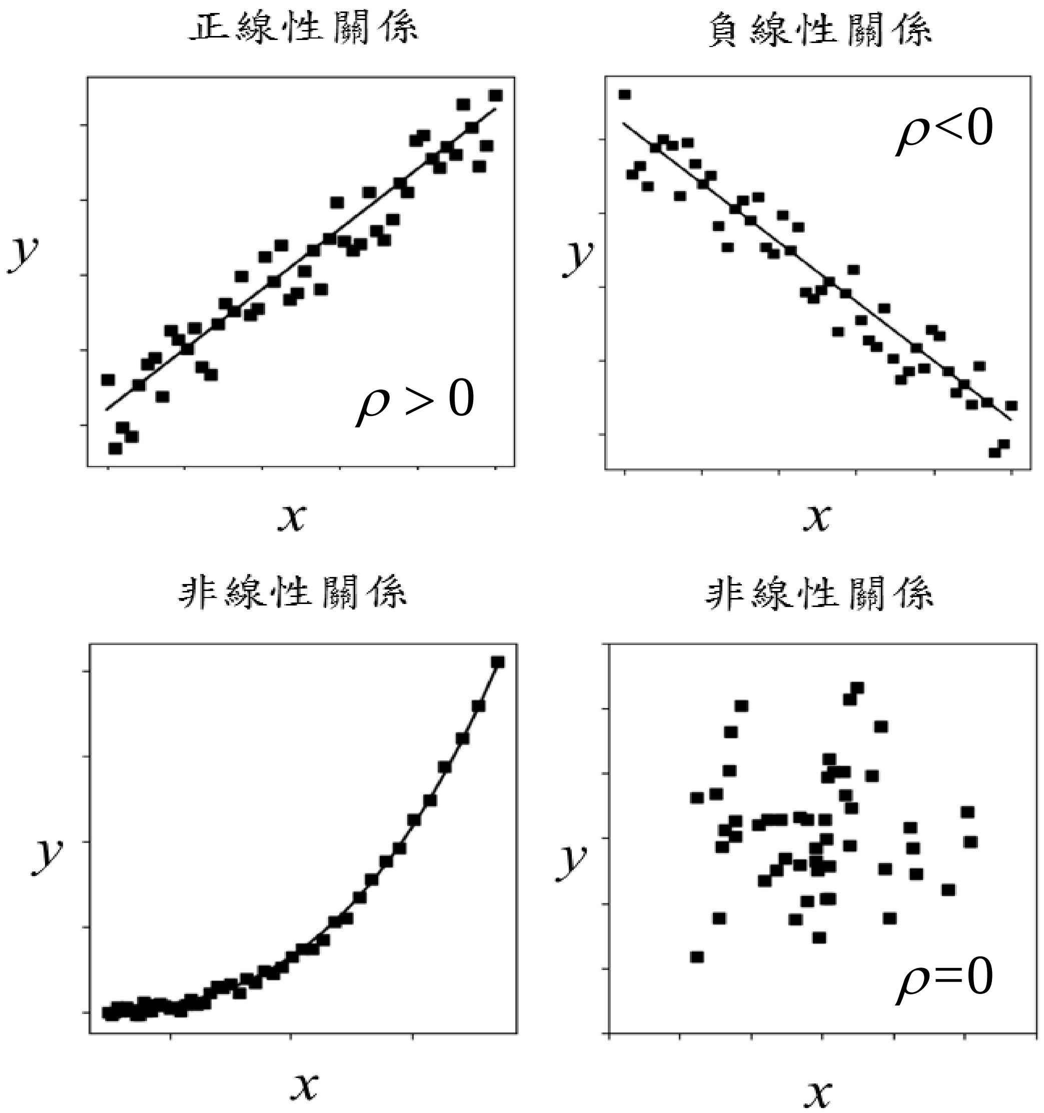
男生的身高和體重的相關係數 $\frac{60}{7.07*9.80} = 0.866$。

房子價錢和坪數的共變異 $\frac{31}{4.43*7.07} = 0.990$。





基礎統計學-相關係數



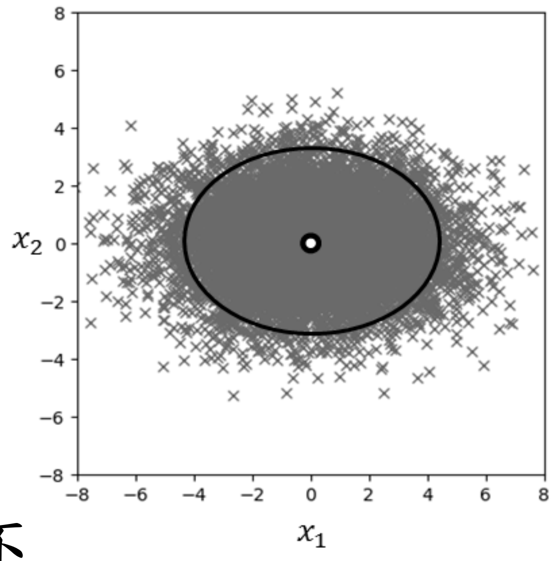
基礎統計學-共變異數矩陣

$$\Sigma = \begin{bmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Cov(X_2, X_2) \end{bmatrix} = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix}$$

單看每個特徵的變異數
這三個範例都一樣
X 軸變異量都是5
Y 軸變異量都是2

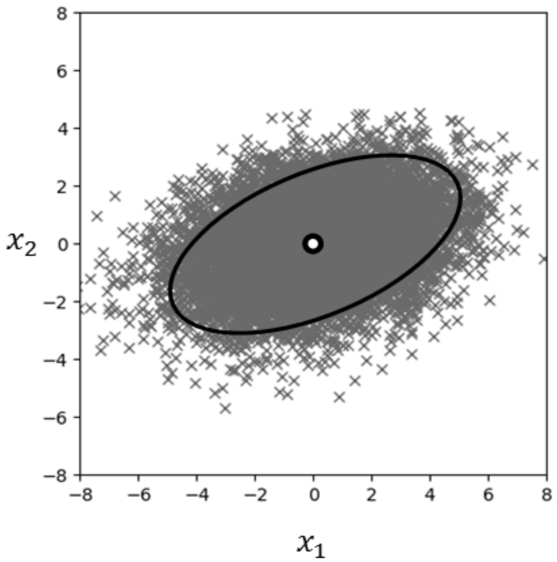
所以差異就在共變異量。
因此在看資料特性的過程不要忽略共變異量。

範例一
X_1, X_2統計獨立



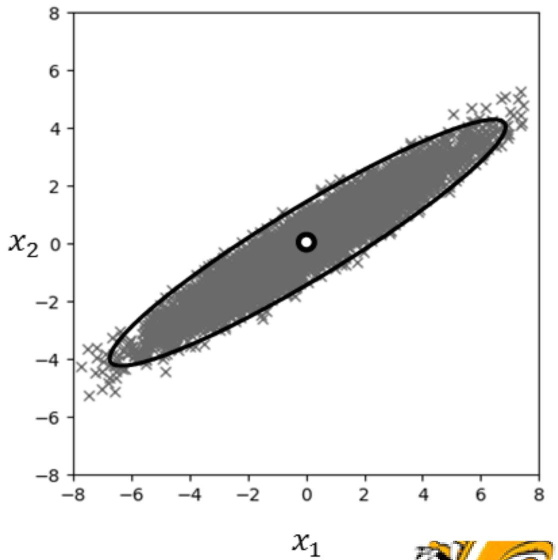
$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}$

範例二
X_1, X_2統計相依



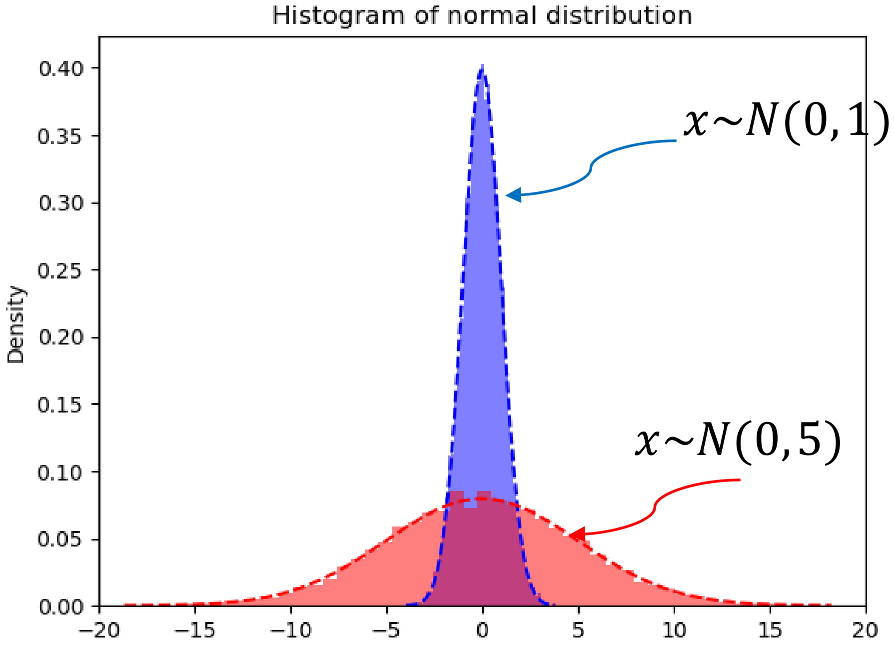
$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}$

範例三
X_1, X_2統計相依

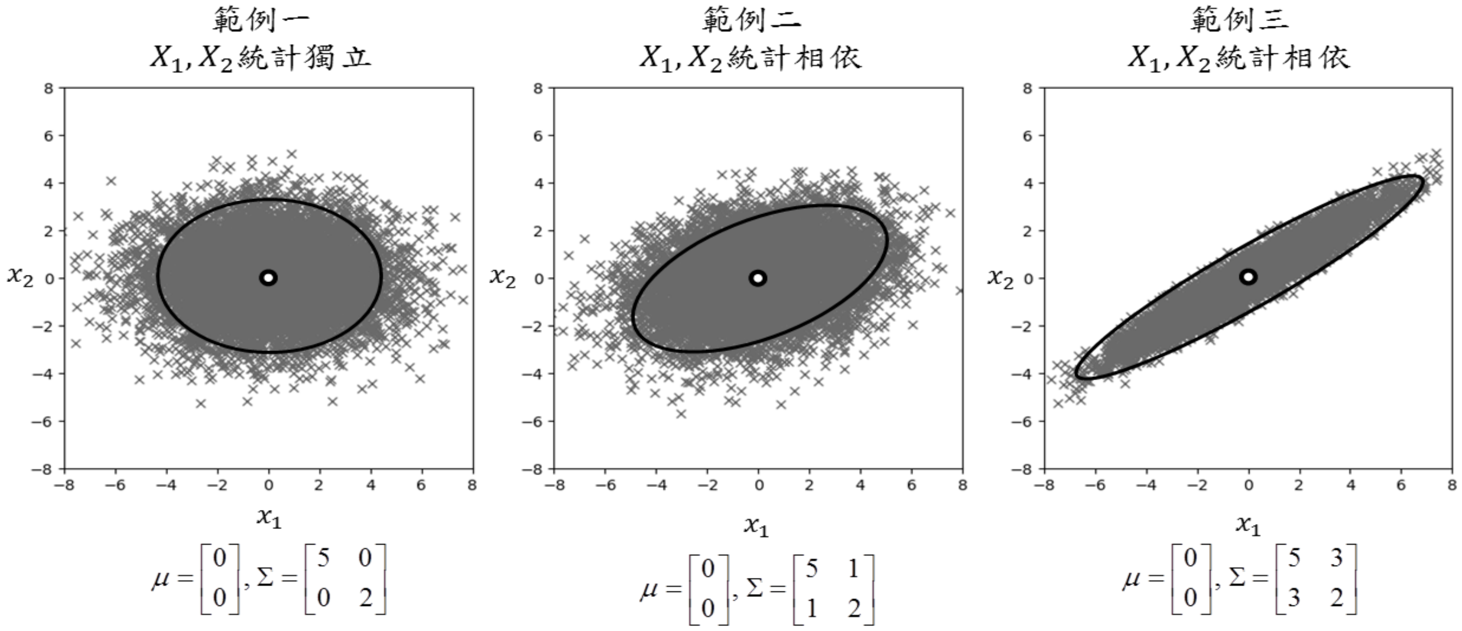


$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$

基礎統計學- Distribution



在一個維度的資料上
標準差就影響整個資料的分布型態
標準差越大，分布越寬



在多個維度的資料上
標準差就影響整個資料的分布型態，標準差越大，分布越寬
但因為有多個維度，所以要考慮到共變異數才能完整描述資料的分布

基礎統計學 - 統計量

- First raw moment (Expectation): mean

$$\mathrm{E}[x] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu$$

- n-order central moment

$$\mathrm{E}[(x-\mu)^n] = \int_{-\infty}^{\infty} (x-\mu)^n f(x)\, dx$$

- 1-order central moment:

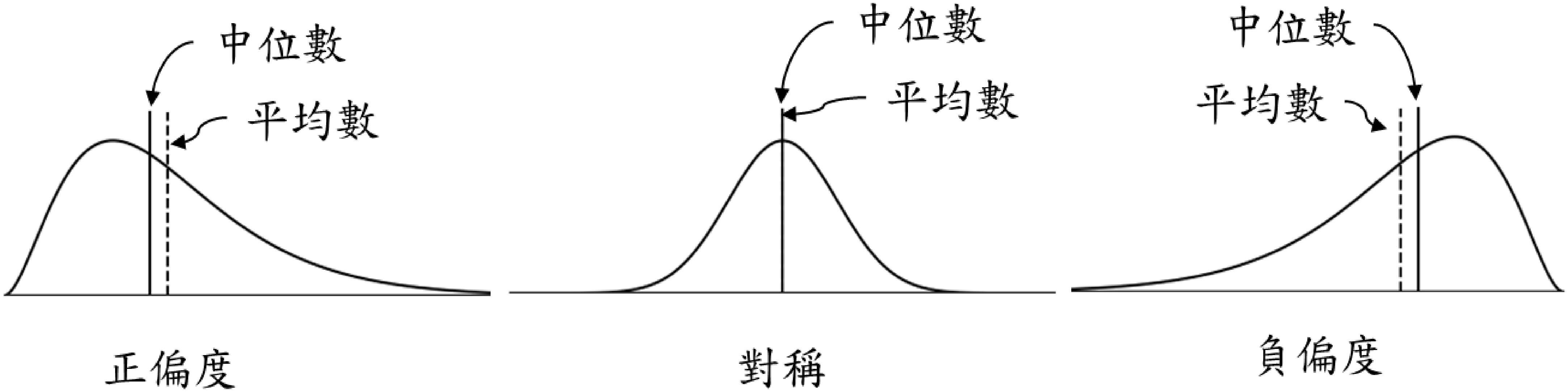
$$\mathrm{E}[(x-\mu)^1] = \mathrm{E}[x-\mu] = \mathrm{E}[x] - \mu = \mu - \mu = 0$$

- 2-order central moment (Variance):

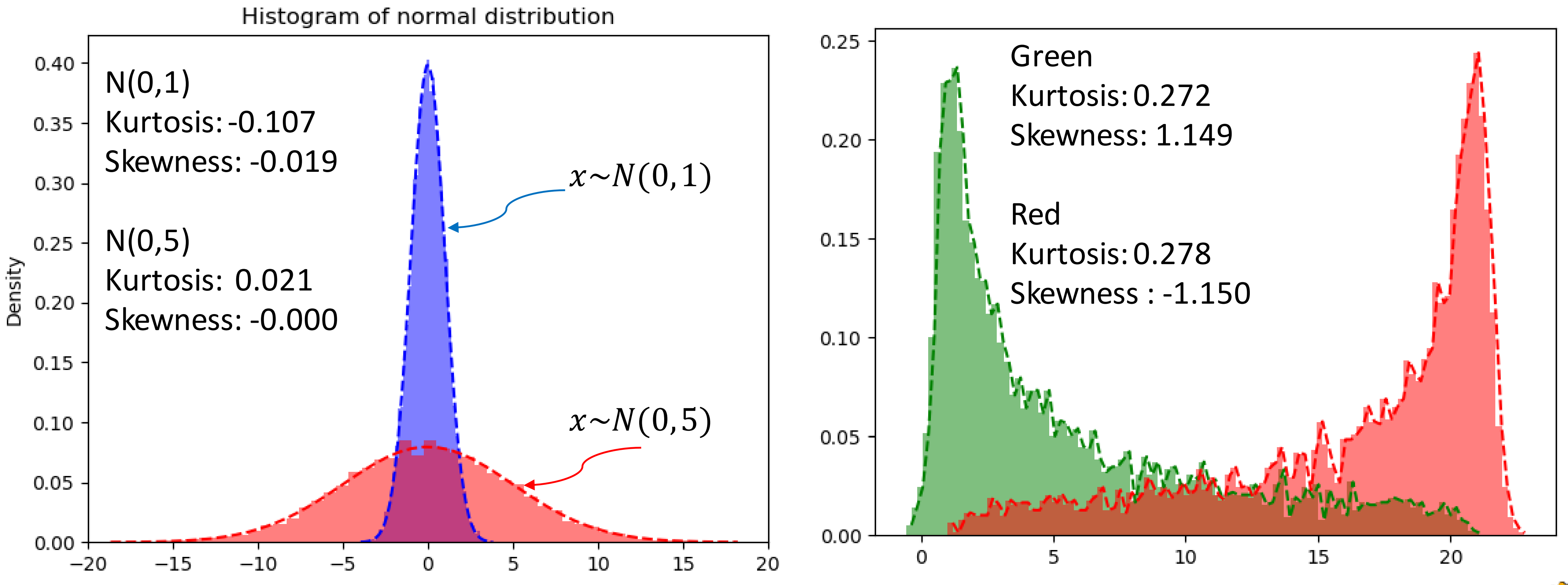
$$Var(x) = \mathrm{E}[(x-\mu)^2] = \mathrm{E}[x^2 + \mu^2 - 2x\mu] = \mathrm{E}[x^2] + \mu^2 - 2\mathrm{E}[x]\mu = \mathrm{E}[x^2] - \mathrm{E}[x]^2$$

- 3-order central moment : Skewness
- 4-order central moment : Kurtosis

基礎統計學 - 統計量



基礎統計學 - 統計量



Histogram of normal distribution
N(0,1)
Kurtosis: -0.107
Skewness: -0.019
N(0,5)
Kurtosis: 0.021
Skewness: -0.000
x~N(0,1)
x~N(0,5)
Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−20
−15
−10
−5
0
5
10
15
20
Green
Kurtosis: 0.272
Skewness: 1.149
Red
Kurtosis: 0.278
Skewness : -1.150
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20

Probability axioms

1. $p(x) \geq 0, \forall x \in A,$ A is a event space.
2. $p(\Omega) = 1$
3. Any countable sequence of disjoint sets $(E_1, E_2, \ldots)$

$$p(\bigcup_i^{\infty} E_i) = \sum_i^{\infty} p(E_i)$$

$P($  $)=0.51$

$P($  $)=0.49$

$P($  $)$ + $P($  $)$ =1

Bayes probability

Bayes probability:

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

$$P(A,B) = P(A \cap B)$$

If A and B are independent,

$$P(A,B) = P(A)P(B)$$

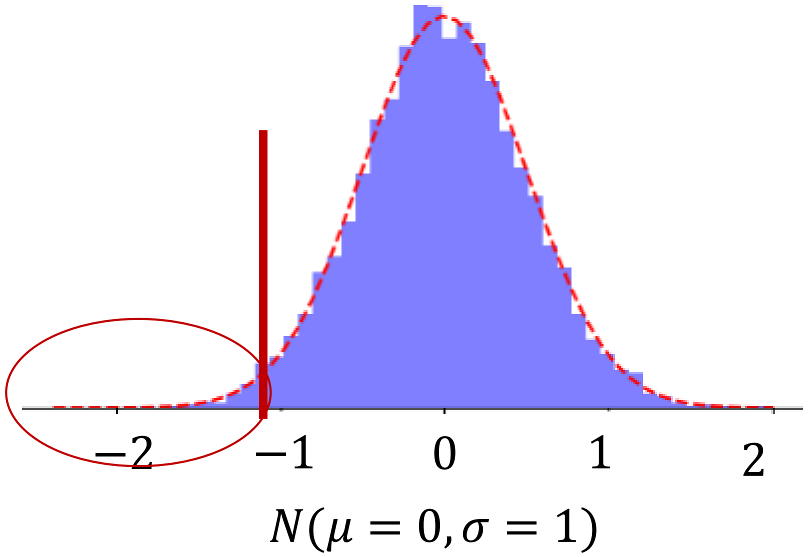
Distribution and probability

假設有A牌手機保固是1年 ，我們知道這台A牌手機的壽命服從常態分佈，平均壽命2年，標準差1年。

Q: 百分之多少的手機會在保固前死機?

假設X為手機壽命，$X \sim N(\mu = 2, \sigma = 1)$

$$p(x < 1) = p\left(z < \frac{1-2}{1}\right)$$
$$= p(z < -1) = 15.87\%$$



所以15.87%手機在過保前就壞了。

Q & A