第六章　連續型隨機變數及其常用的機率分配

6.1 連續型隨機變數(Continuous type random variables)

隨機變數的可能值為不可數無限多時，其可能值的集合為一區間，稱為連續型隨機變數。

- 機率密度函數(probability density function , p.d.f.)

連續型隨機變數所產生的機率分配稱為機率分配函數，通常以 f(y)表示，具有如下性質：

(1)f(y)≧0　，y∈R

(2)令隨機變數的可能值集合為[a,b]，則

$\int_{-\infty}^{\infty} f(y)dy = \int_{a}^{b} f(y)dy = 1$(機率和為 1)

(3)如果[c,d]⊂[a,b]則

$$P(c \le Y \le d) = \int_{c}^{d} f(y)dy$$

因此可知

P(Y=e)=$\int_{e}^{e} f(y)dy = 0$

對連續型隨機變數而言，單點發生的機率為 0

6.2 期望值(expected value)
變異數(variance)

● 期望值

連續型隨機變數 Y 的期望值 E(Y)定義為

$$\mu = E(Y) = \int_{-\infty}^{\infty} yf(y)dy$$

若 k(Y)為連續型隨機變數 Y 之函數，則期望值為

$$E[k(Y)] = \int_{-\infty}^{\infty} k(y)f(y)dy$$

● 變異數

若連續型隨機變數 Y 之期望值 E(Y)= μ，則 Y 之變異數為

$$\sigma^2 = V(Y) = E[(Y-\mu)^2]$$
$$= \int_{-\infty}^{\infty}(y-\mu)^2 f(y)dy$$

且標準差為

$$\sigma = \sqrt{\sigma^2} = \sqrt{V(Y)}$$

● 期望值及變異數基本定理

若 Y 為一連續型隨機變數，期望值 E(Y)= μ ，a,b 為常數且 $k_1(Y), k_2(Y), \dots, k_m(Y)$ 為隨機變數 Y 之函數，則

(1)E(aY+b)=aE(Y)+b

(2)V(aY+b)=V(aY)= $a^2V(Y)$

(3) $\mathrm{E}[k_1(Y)+k_2(Y)+\cdots+k_m(Y)]$

$= \mathrm{E}[k_1(Y)]+\mathrm{E}[k_2(Y)]+\cdots+\mathrm{E}[k_m(Y)]$

(4) $\mathrm{V}(\mathrm{Y})=\mathrm{E}(Y^2)-[E(Y)]^2=E(Y^2)-\mu^2$

6.3 常態分配(normal distribution)

連續型隨機變數 Y，若其機率密度函數為

$$f(y)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(y-\mu)^2}{2\sigma^2}}\quad,\mu\in R,\sigma>0,y\in R$$

則 Y 的機率分配稱為常態分配，簡單表示為 $Y\sim N(\mu,\sigma^2)$

且 $E(Y)=\mu$

$V(Y)=\sigma^2$

● 常態分配性質

(1)常態分配曲線兩端尾巴與橫軸漸漸接近，但絕不與橫軸相交。

(2)常態分配是以 μ 為中心的左右對稱分配

(3) $P(\mu-\sigma<Y<\mu+\sigma)=0.683$

$P(\mu-2\sigma<Y<\mu+2\sigma)=0.954$

$P(\mu-3\sigma<Y<\mu+3\sigma)=0.997$

上述(3)之性質可用來簡略檢驗一組資料是否來自常態分配。

● 標準常態分配(standard normal

distribution)

當常態分配期望值為 0，變異數為 1，則稱為標準常態分配，其機率分配函數為

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}},\ z\in R$$

簡略表示為 Z~N(0,1)

(習慣用隨機變數 Z 來表示具有標準常態分配)

● 標準常態分配性質

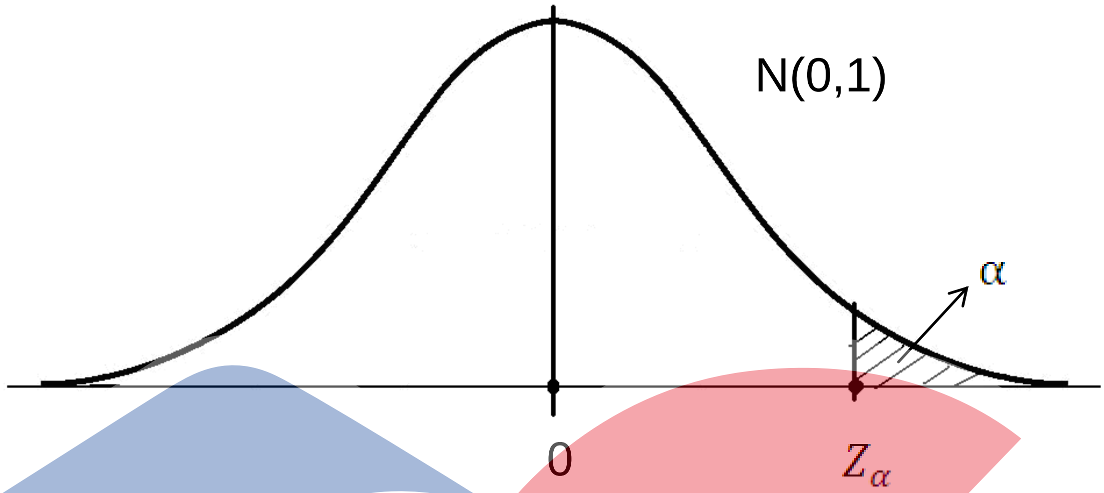
(1)P(Z>a)=P(Z<-a)

(2)P(Z>0)=P(Z<0)=$\frac{1}{2}$

(3)P(a<Z<b)=P(Z>a)-P(Z>b)

(4)P(Z>Z_{α})=α(查表重要表示法)

參考課本附錄 A-10



$P(Z>Z_\alpha)=\alpha$

其中　1. $Z \sim N(0,1)$

2. $>$ 必須大於

3. Z_α 必須正值

在上述三個條件下，方可查表

●標準化

隨機變數 Y 具有期望值 E(Y)及變異數 V(Y)，則$\frac{Y-E(Y)}{\sqrt{V(Y)}}$稱為隨機變數 Y 之標準化

若 Y~N(μ,σ^2)，則

$$Z=\frac{Y-E(Y)}{\sqrt{V(Y)}}=\frac{Y-\mu}{\sigma}\sim N(0,1)$$

即隨機變數 Y 具有常態分配，隨機變數 Y 標準化後具有標準常態分配。

Ex. Y~N(μ,σ^2)具有如下特性：

$P(\mu-\sigma<Y<\mu+\sigma)=0.683$

$P(\mu-2\sigma<Y<\mu+2\sigma)=0.954$

$P(\mu-3\sigma<Y<\mu+3\sigma)=0.997$

試證明之

Sol:

$P(\mu-\sigma<Y<\mu+\sigma)$

$=P(\frac{\mu-\sigma-\mu}{\sigma}<\frac{Y-\mu}{\sigma}<\frac{\mu+\sigma-\mu}{\sigma})$

$=P(-1<Z<1)$

$=0.6826\approx0.683$

$P(\mu-2\sigma<Y<\mu+2\sigma)$

$=P(\frac{\mu-2\sigma-\mu}{\sigma}<\frac{Y-\mu}{\sigma}<\frac{\mu+2\sigma-\mu}{\sigma})$

$=P(-2<Z<2)$

$=0.9544\approx0.954$

同理

$P(\mu-3\sigma<Y<\mu+3\sigma)\approx0.997$

給定一組樣本資料可計算樣本平均

數$\bar{X}$及樣本標準差 S，如果

資料落在$[\bar{X} - S, \bar{X} + S]$比例約 0.683

資料落在$[\bar{X} - 2S, \bar{X} + 2S]$比例約 0.954

資料落在$[\bar{X} - 3S, \bar{X} + 3S]$比例約 0.997

則可簡略檢驗該組資料來自常態分配。

●標準常態分配查表常用技巧

上回我們談及標準常態分配查表準則

$$P(Z > Z_{\alpha}) = \alpha$$

N(0,1)　　大於　　正值

給定Z_{α}值查表獲得機率α

或給定機率值α查表獲得Z_{α}

Ex： Z~N(0,1)

P(Z>1.96)=0.025

因此給定$Z_{\alpha} = 1.96$可查表獲得α=0.025

反之給定機率 0.025 可查表獲得Z_{α}=1.96

- 中央極限定理(Central Limit Theorem)

當$X_1, X_2, \ldots, X_n$是從具有母體平均數μ及母體變異數σ^2的任何分配取出的隨機樣本，則當 n 趨近於無限大時，$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$的累積分配函數會趨近於標準常態分配的累積分配函數，亦即當 n 趨近無限大時，$\bar{X}$的分配趨近於常態分配 $N(\mu, \frac{\sigma^2}{n})$。

一般而言，連續型機率分配樣本大小 n≥30 時，會有不錯近似效果(常態分配是連續型機率分配)，對於離散型機率分配可以使用連續型修正降低誤差。

● 二項分配近似於常態分配

若 Y~B(n,p)，且 np≥10,n(1-p)≥10，此時$\frac{Y-E(Y)}{\sqrt{V(Y)}}=\frac{Y-np}{np(1-p)}$之機率分配近似於標準常態分配，藉由連續性修正，a, b 為正整數，則

$$P(Y\le b)=P(Y\le b+\tfrac{1}{2})$$

$$=P\left(\frac{Y-np}{\sqrt{np(1-p)}}\le\frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

$$\approx P\left(Z\le\frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

$P(a \le Y \le b)$

$=P(a-\frac{1}{2} \le Y \le b+\frac{1}{2})$

$=P(\frac{\left(a-\frac{1}{2}\right)-np}{\sqrt{np(1-p)}} \le \frac{Y-np}{\sqrt{np(1-p)}} \le \frac{\left(b+\frac{1}{2}\right)-np}{\sqrt{np(1-p)}})$

$\approx P(\frac{\left(a-\frac{1}{2}\right)-np}{\sqrt{np(1-p)}} \le Z \le \frac{\left(b+\frac{1}{2}\right)-np}{\sqrt{np(1-p)}})$

$P(Y=a)$

$=P(a \le Y \le a)$

$=P(a-\frac{1}{2} \le Y \le a+\frac{1}{2})$

$=P(\frac{\left(a-\frac{1}{2}\right)-np}{\sqrt{np(1-p)}} \le \frac{Y-np}{\sqrt{np(1-p)}} \le \frac{\left(a+\frac{1}{2}\right)-np}{\sqrt{np(1-p)}})$

$\approx P(\frac{\left(a-\frac{1}{2}\right)-np}{\sqrt{np(1-p)}} \le Z \le \frac{\left(a+\frac{1}{2}\right)-np}{\sqrt{np(1-p)}})$