Continuous Probability Distributions

連續型機率分配

連續型機率分配

統計對象	型態	分配類型	
母體分配 樣本分配	離散型 機率分配	均勻分配 多項分配 超幾何分配 伯努利分配	二項分配 負二項分配 幾何分配 波松分配
	連續型 機率分配	連續均勻分配 標準常態分配 Gamma分配	常態分配 指數分配 卡方分配
抽樣分配	機率分配	樣本比例差（ $\lvert\hat{p}-p\rvert$ ） 平均數～ Z分配 t分配 變異數～ 單母體變異數 卡方分配 兩母體變異數比 F分配	

常見連續型機率分配

分配類型	適用情境說明	分配函數
連續均勻分配 uniform distribution	隨機變數（X）在某連續區間內所發生的機率都相同	$f(x) = \frac{1}{b-a}$，$a \le x \le b$
常態分配 normal distribution	存在於大自然間的各種現象或狀態，都可以將母體視為常態分配	$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$ $-\infty < x < \infty$, μ為平均數, σ為標準差
標準常態分配 standard normal distribution	透過 $z = \frac{x-\mu}{\sigma}$ 的「標準化」，將原本要利用微積分計算求值，轉換成可以利用查表得到結果	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, $-\infty < z < \infty$
指數分配 exponential distribution	與Poisson隨機變數相反，指數分配的隨機變數（X）是描述連續兩事件發生的間隔時間	$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$, $x \ge 0$ x 為第一次發生事件所需時間； β 為事件發生的平均時間；

其他連續型機率分配

分配類型	適用情境說明	分配函數
Gamma 分配 Gamma distribution	隨機變數 (*X*) 表示事件第 *a* 次發生所需的時間。因此若「*X>t*」表示事件第 *a* 次發生至少需要 *t*個時間單位；另個說法是，在 *t* 時間內事件至少發生 *(a-1)* 次的機率	$f(x) = \frac{1}{\beta\Gamma(a)}(\frac{x}{\beta})^{a-1}\, e^{-\frac{x}{\beta}}$, $0 < x < \infty, a > 0, \beta > 0$ x 為第 *a* 次發生所需時間； β 表發生一次所需的時間；
卡方分配 chi-square distribution	卡方分配可以算是Gamma分配的特例（$a = \frac{\nu}{2}$, $\beta = 2$），在統計應用上，可進行單一母體變異數 σ^2 的統計推論；可用來做適合度檢定（goodness-of-fit test）、獨立性檢定（test of independence）與變異數齊一性檢定（test of homogeneity）	$f(x) = \frac{1}{2\Gamma\left(\frac{\nu}{2}\right)}(\frac{x}{2})^{\frac{\nu}{2}-1}\, e^{-\frac{x}{2}}$, $0 < x$, ν為卡方分配的自由度

uniform distribution

連續均勻分配

函數定義（分配形式）

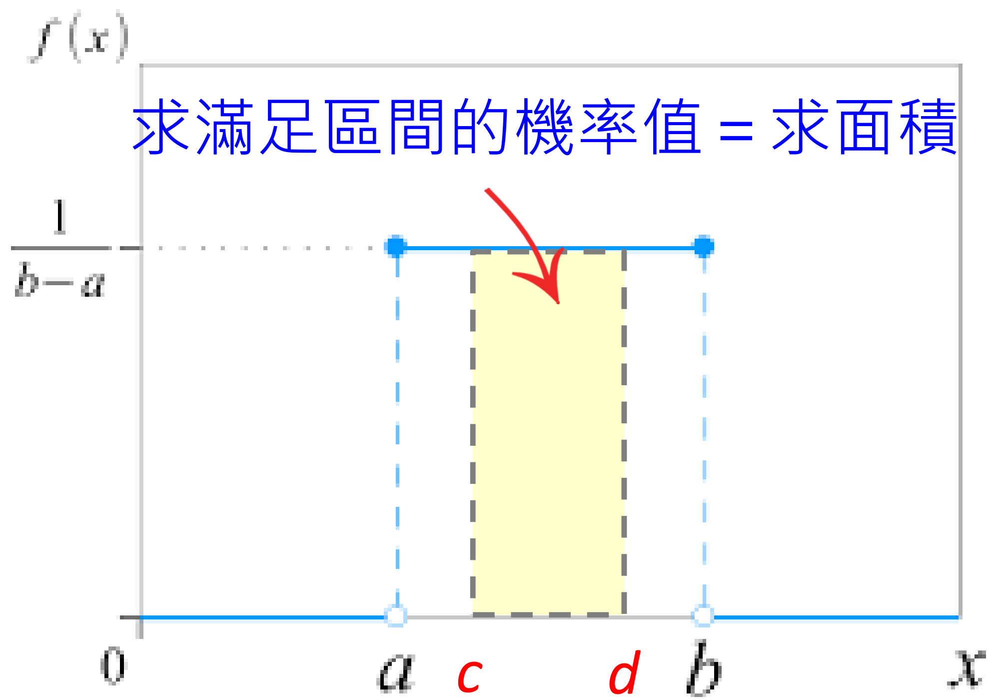
隨機變數（X）在某連續區間〔a,b〕內所發生的機率都相同，例如等待公車到達機率，或是約會碰面機率等。

$$f(x) = \frac{l}{b-a}\ ,\qquad a \le x \le b$$

$$P(c<\mathrm{x}<\mathrm{d}) = \int_c^d f(x)\,dx$$

$$= \int_c^d \frac{1}{b-a}\,dx$$

$$= \frac{1}{b-a}(d-\mathrm{c})$$



分配的重要參數

期望值 $$E(x)=\frac{a+b}{2}$$

變異數 $$V(x)=\frac{(b-a)^2}{12}$$

證明：

$$E(x)=\int_a^b xf(x)\,dx=$$

$$E(x^2)=\int_a^b x^2f(x)\,dx=$$

$$V(x)=\sigma^2=E(x^2)-〔E(x)〕^2=$$

案例

已知台北捷運系統大約5分鐘發一班車，且為均勻分配。假設某人欲到站搭乘捷運，若此人完全不知道任何列車到站時間的訊息，請問：

(1)此人等待時間小於2分鐘的機率？

(2)此人至少等待3分鐘以上的機率？

(3)此人等待時間介於1到4分鐘的機率？

(4)此人平均等待時間為何？

(5)此人等待時間的變異數為何？

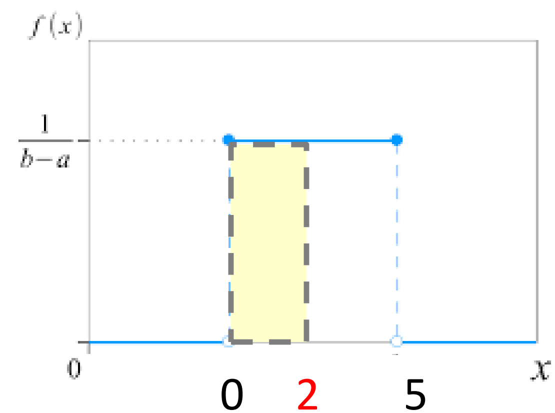
「統計學 二版」p10-4，李德治、林孟濡、童惠玲 著，博碩文化

案例解說 1

約5分鐘發一班車　（開始思考……）

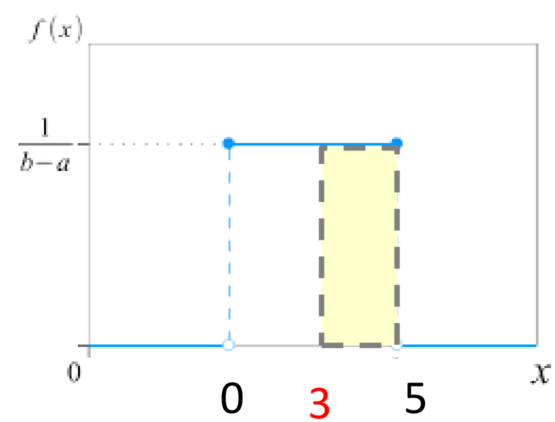
→（情境一）到達車站時，若是剛錯過一班，要再等5分鐘才能搭上車
→（情境二）到達車站時，剛好遇到車進站，所以不用等就能上車
→（情境三）到達車站時，看到顯示，還有幾分鐘到站
→→到達車站，還要再等待1分鐘、2分鐘、...或5分鐘的機率，每個等待發生的機率應該都是1/5的機率，所以

$$f(x) = \frac{l}{5}\ ,\qquad 0 \le x \le 5$$



(1)此人等待時間小於2分鐘的機率？

$$P(x < 2) = \int_0^2 \frac{1}{5}\,dx = \frac{1}{5}(x)|_0^2 = \frac{2}{5}$$



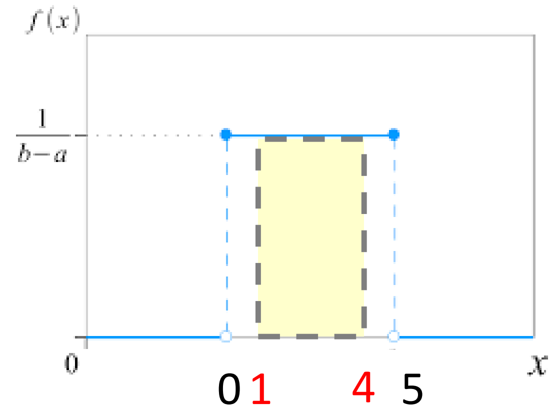
(2)此人至少等待3分鐘以上的機率？

$$p(x \ge 3) = \int_3^5 \frac{1}{5}\,dx = \frac{1}{5}(x)\Big|_3^5 = \frac{1}{5}(5-3) = \frac{2}{5}$$

案例解說 2

(3)此人等待時間介於1到4分鐘的機率？

$$P(1 < x < 4) = \int_{1}^{4} \frac{1}{5}\,dx = \frac{1}{5}(x)|_{1}^{4} = \frac{1}{5}(4-1) = \frac{3}{5}$$



(4)此人平均等待時間為何？

→平均等待時間就表示求期望值，所以

$$E(x) = \frac{a+b}{2} = \frac{0+5}{2} = \frac{5}{2} \quad (分鐘)$$

(5)此人等待時間的變異數為何？

$$V(x) = \frac{(b-a)^2}{12} = \frac{(5-0)^2}{12} = \frac{25}{12}$$

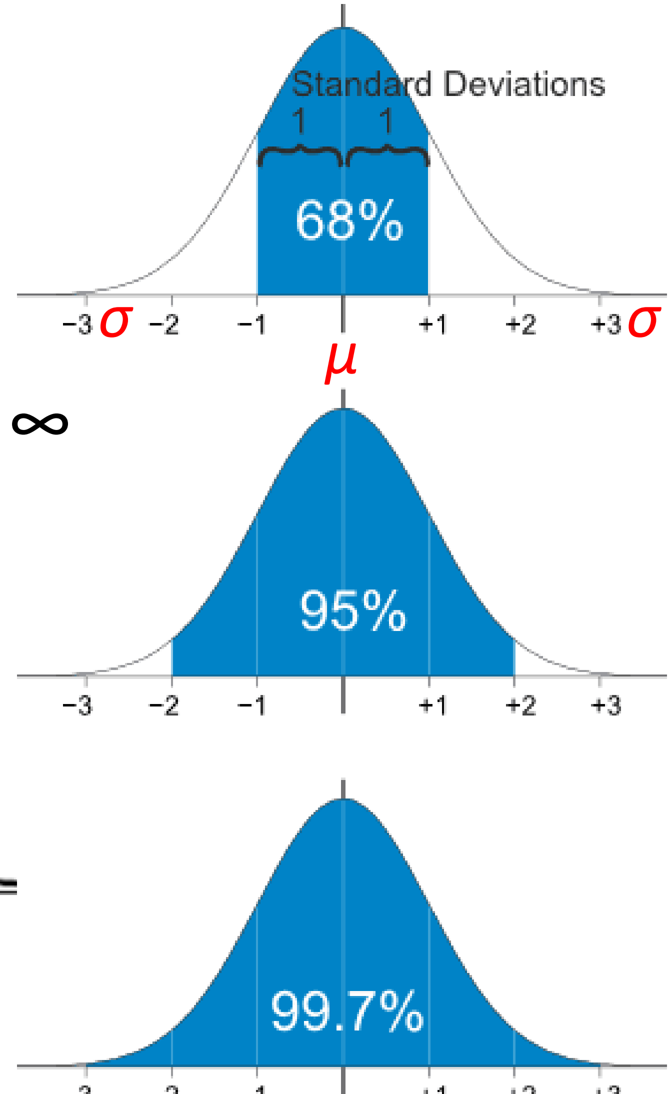
normal distribution

常態分配

函數定義（分配形式）

最早由法國數學家De Moiver於1773年提出，隨後高斯（Carl Gauss）在重複測量的誤差研究中，亦導出曲線方程式，所以常態分配又稱為高斯分配。

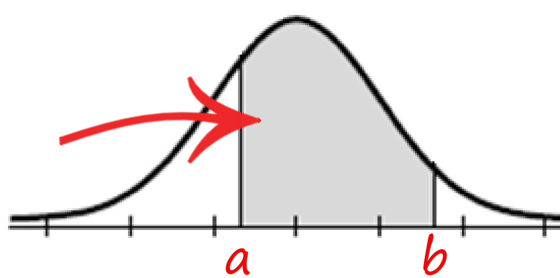
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}, \ -\infty < x < \infty$$



通常以符號 $X \sim N(\mu, \sigma^2)$ 表示具常態分配

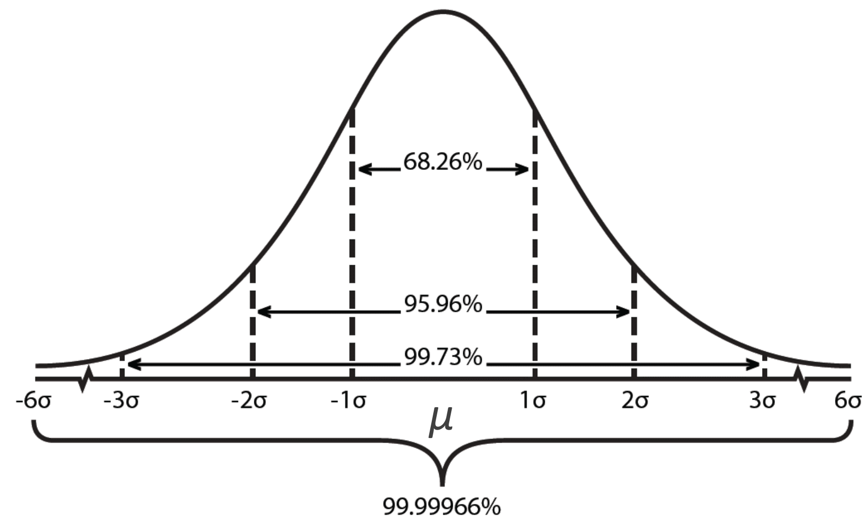
$P(a \leqq x \leqq b)$

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



函數定義（分配形式）2

存在於大自然間的各種現象或狀態，都可以將母體視為常態分配。例如身高、體重、智商等，常態分配機率函數曲現呈鐘形（bell shaped），此曲線稱為常態曲線（normal curve）；適合以平均數來代表母體的中央集中趨勢。

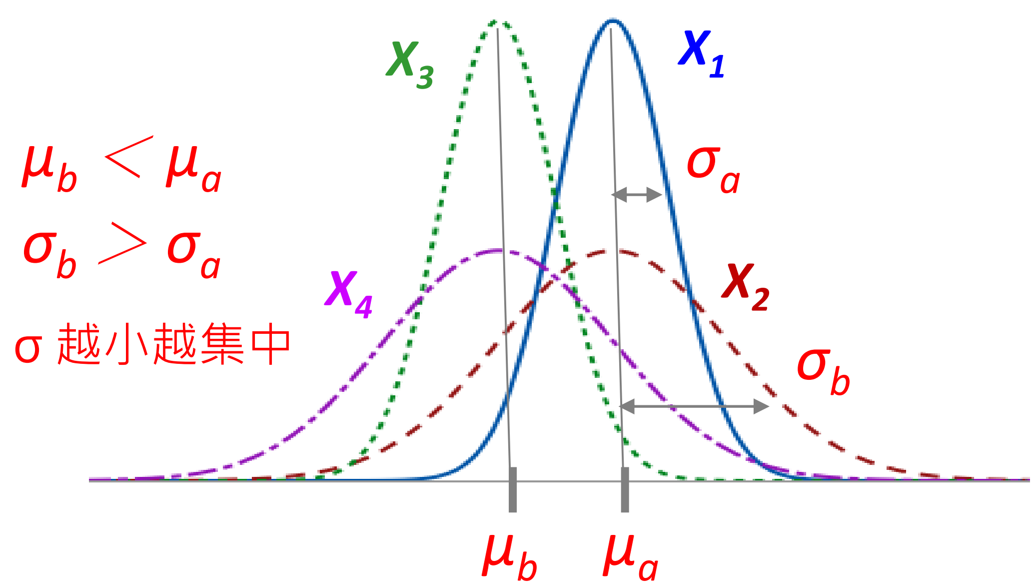


$X_1 \sim N(\mu_a , \sigma_a^2)$

$X_2 \sim N(\mu_a , \sigma_b^2)$

$X_3 \sim N(\mu_b , \sigma_a^2)$

$X_4 \sim N(\mu_b , \sigma_b^2)$



分配的重要參數與特性

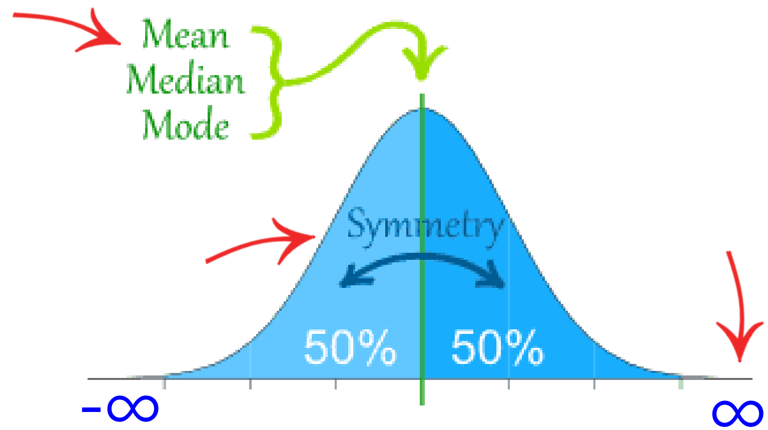
期望值　$E(x) = \mu$

變異數　$V(x) = \sigma^2$

常態分配特性：

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

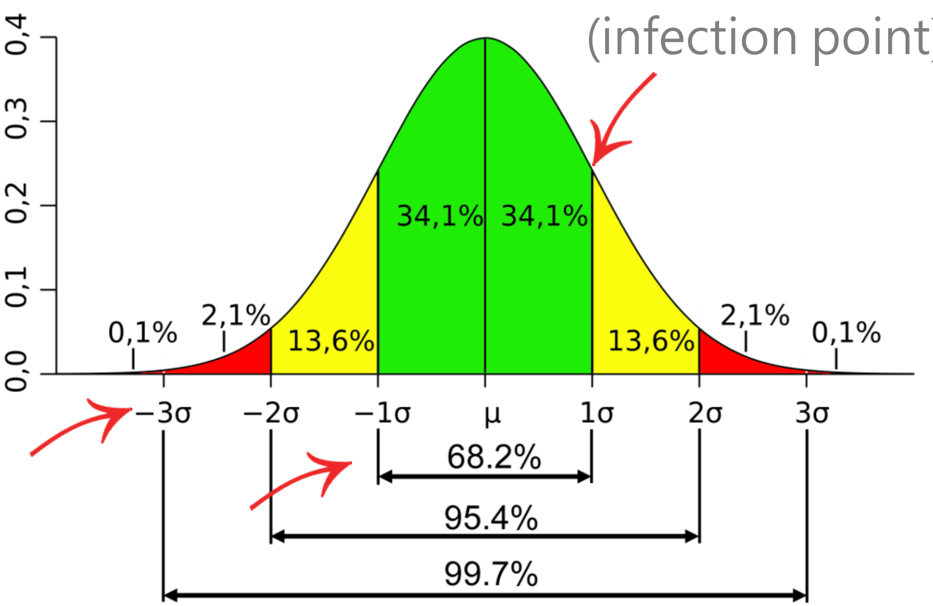
(2)



(3) 偏態係數=0；
峰度係數=3；

(4)

反曲點
(infection point)



案例

製藥廠製造一種抗胃癌新藥，假設每顆藥丸重量符合常態分配N（$0.3 , 0.01^2$），單位：g，因為此藥含有某成分比率的珍貴藥材與劇毒，製造產品需要較嚴格的品管。製藥廠認為此抗癌新藥的重量應為（0.3 ± 0.02）之間才安全。試問，此種抗癌新藥產品不被接受的比率有多少？

案例解說

〝重量符合常態分配N（0.3 , 0.01^2）〞 → μ = 0.3；σ = 0.01

〝重量應為（0.3 ± 0.02）之間〞

→ 要求P（(0.3-0.02) ≦ X ≦ (0.3+0.02)）的值

方法一（用定義計算）：

$$P(0.28 \leqq X \leqq 0.32) = \int_{0.28}^{0.32} \frac{1}{\sqrt{2\pi}(0.01)} e^{\frac{(x-0.3)^2}{2(0.01)^2}} = \ldots\ldots$$（現場任務）

方法二（利用「Z」查表）：

$$P(0.28 \leqq X \leqq 0.32) = P\left(\frac{0.28-0.3}{0.01} \leq \frac{x-0,3}{0.01} \leq \frac{0,32-0,3}{0,01}\right)$$

= P(-2≦ Z ≦2) = 2 P(Z ≦ 2) – 1

= 2 × 0.9772 – 1 → 查表可得「0.9772」的值

= <u>0.9544</u> → 〝安全〞，所以不被接受的比率有4.56%(1-0.9544)

standard normal distribution

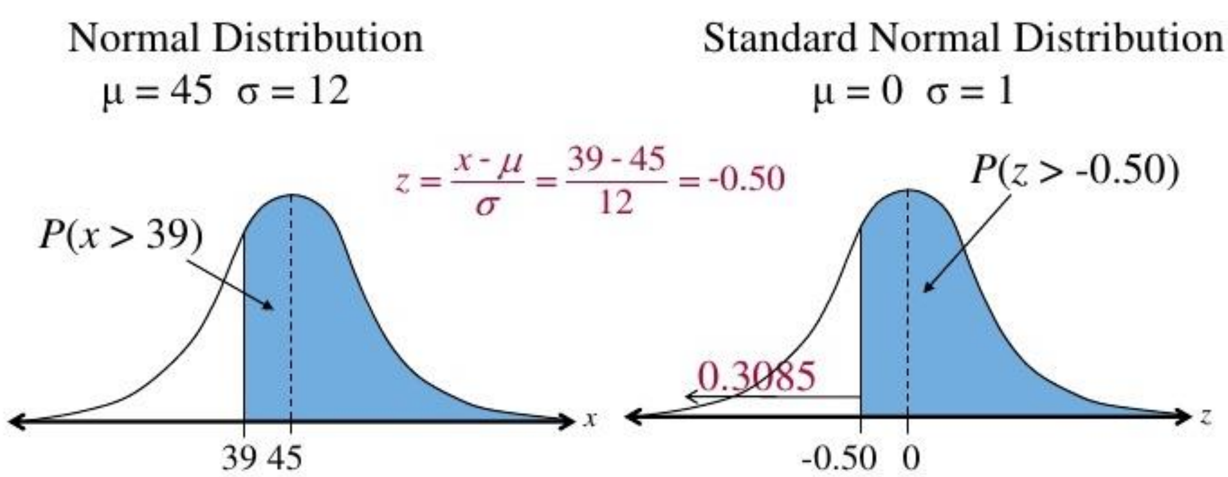
標準常態分配

函數定義（分配形式）

要計算常態分配就必須要進行積分運算，在多數實際的情況下，積分運算不容易求得答案結果，此時可以運用數值分析的方法來求得近似值。但是數學家更想到一種方式，利用「座標變換」（變數變換）的技巧，將常態分配轉換為「標準常態分配」（standard normal distribution），並且事先將各種可能的機率值製作成「查表」；利用這個「查表」方式，就可以不用積分計算，或是計算機操作，就可以求得對應的數值。

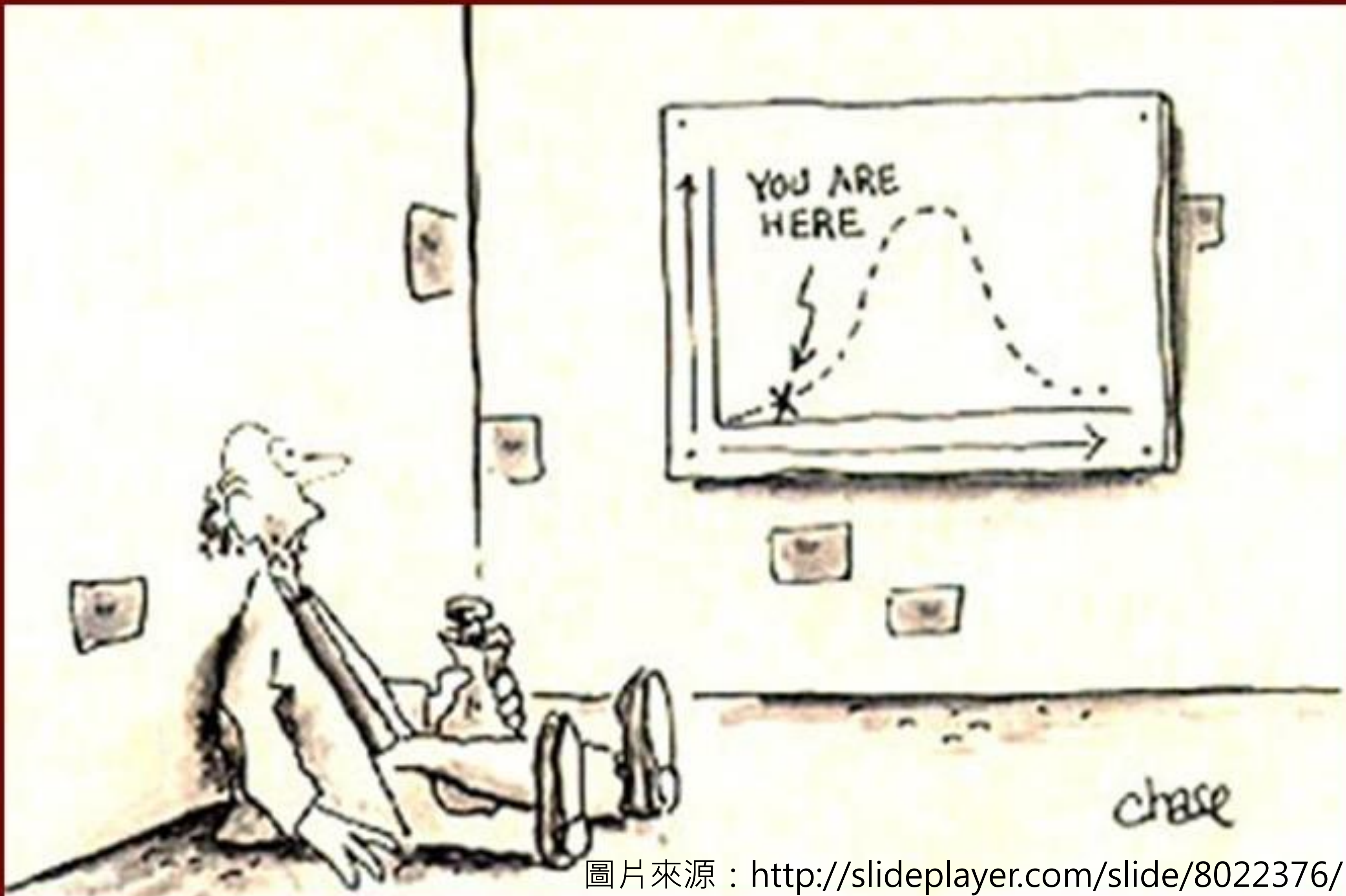
標準化： $z = \dfrac{x - \mu}{\sigma}$　　　$f(z) = \dfrac{1}{\sqrt{2\pi}} e^{\frac{z^2}{2}}$, $-\infty < z < \infty$

$\rightarrow \mu_z = 0$　；　$\sigma_z = 1$



圖片來源：http://www.slideshare.net/raj_2952/normal-distribution-and-sampling-distribution

請參閱專章進一步說明



圖片來源：http://slideplayer.com/slide/8022376/

exponential distribution

指數分配

函數定義（分配形式）1

與Poisson隨機變數相反，指數分配的隨機變數（X）是描述連續兩事件發生的間隔時間，常用在計算生命長度、顧客打電話時間、電子產品失效等問題。指數分配與Poisson分配可以彼此相互轉換，也就是說指數分配的問題也可用Poisson分配去計算機率，反之亦然：

Poisson分配	指數分配
1小時內，平均20部車子開進停車場（λ=20輛/小時）	平均每隔3分鐘有1部車子開進停車場（β=3分鐘/輛）
高速公路上每1公里平均種10棵樹（λ=10輛/1小時）	高速公路上平均每隔100公尺有1棵樹（β=100公/棵）
血液中每1cm^{3}中有1000個紅血球（λ=1000個/cm^3）	血液中每0.001cm^{3}中有1個紅血球（β=0.001cm^3/個）

「統計學 二版」10-25，李德治、林孟濡、童惠玲 著，博碩文化

函數定義（分配形式）2

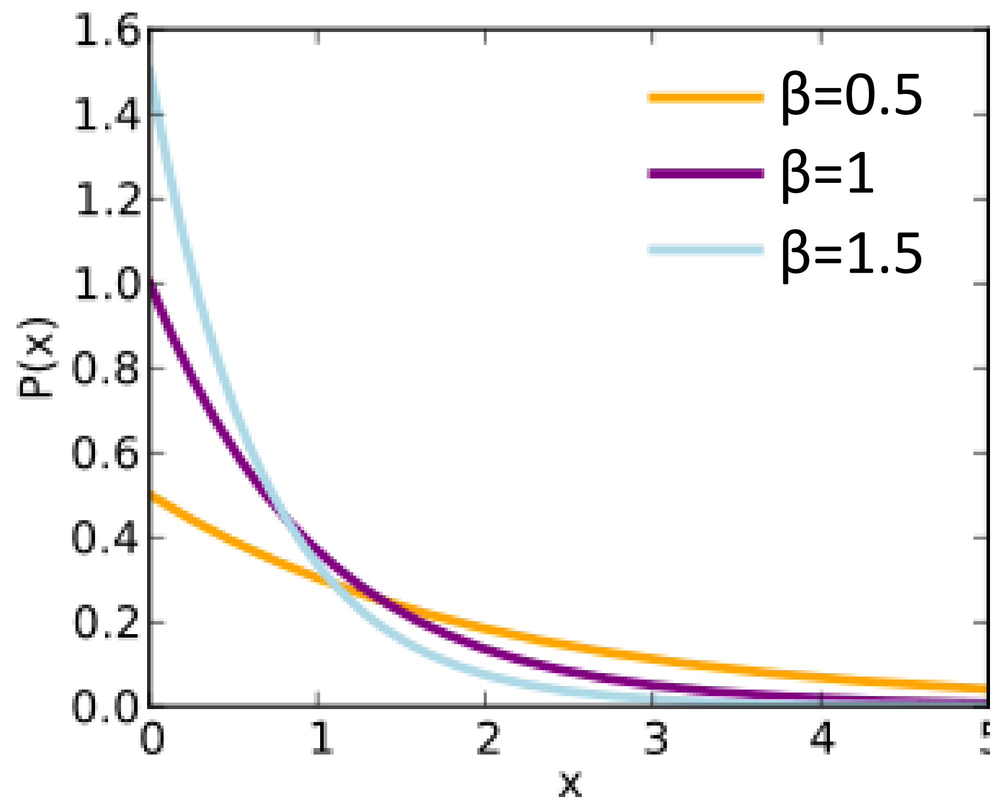
$$f(x) = \frac{1}{\beta}\, e^{-\frac{x}{\beta}}, x \geq 0$$

x 為第一次發生事件所需時間；

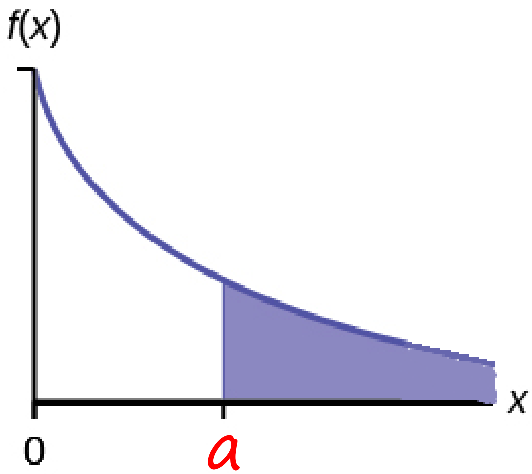
β 為事件發生的平均時間；

期望值 $E(x) = \beta$

變異數 $V(x) = \beta^2$



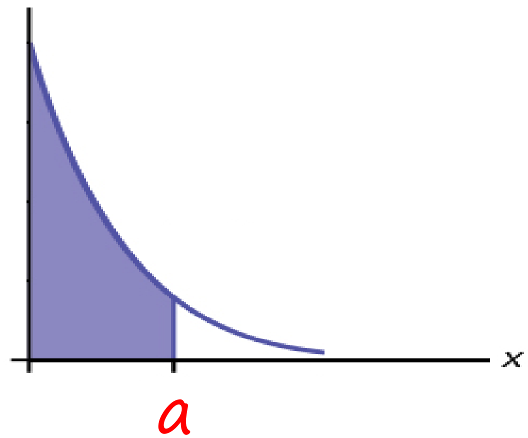
常用之指數分配計算公式



$$P(X > a) = \int_a^{\infty} f(x)\,dx$$

$$= \int_a^{\infty} \frac{1}{\beta} e^{-\frac{x}{\beta}}\,dx$$

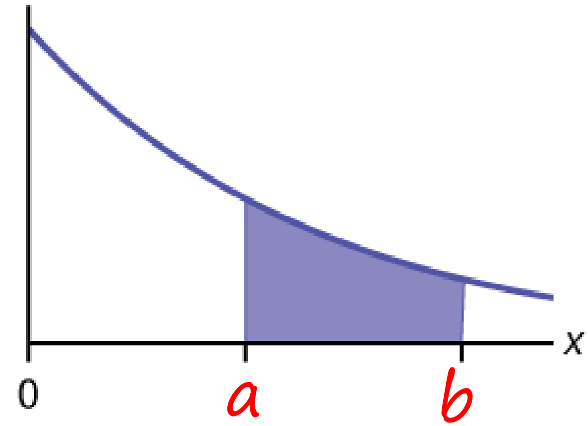
$$= e^{-\frac{a}{\beta}}$$



$$P(X < a)$$

$$= \int_0^{a} \frac{1}{\beta} e^{-\frac{x}{\beta}}\,dx$$

$$= 1 - e^{-\frac{a}{\beta}}$$



$$P(a < X < b)$$

$$= \int_a^{b} \frac{1}{\beta} e^{-\frac{x}{\beta}}\,dx$$

$$= e^{-\frac{a}{\beta}} - e^{-\frac{b}{\beta}}$$

案例 1

假設門診醫生看病時間平均每分鐘看0.2人，若看病時間為指數分配，試問看一病人超過5分鐘的機率有多少？最多10分鐘的機率有多少？

「現代統計學」p99，吳柏林 著，五南圖書

案例 1 解說

設隨機變數 X 表醫生對每個病人看病時間，平均每分鐘看 0.2病人 → 平均 5 分鐘看1個病人 → 服從指數分配，β=5

(1) $P(X > 5) = \int_{5}^{\infty} \frac{1}{5} e^{-\frac{x}{5}} dx = -e^{-\frac{x}{5}} \Big|_{5}^{\infty} = \mathrm{e}^{-1} = 0.37$

(2) $P(X < 10) = \int_{0}^{10} \frac{1}{5} e^{-\frac{x}{5}} d = -e^{-\frac{x}{5}} \Big|_{0}^{10} = 1 - \mathrm{e}^{-2} = 0.86$

案例 2

假設某廠牌彩色電視機其壽命時間呈指數分配，

且平均壽命10年，試求以下機率值：

(1)壽命長達12年以上？

(2)1年內即發生故障而報廢？

(3)壽命時間介於2至10年？

(4)求電視壽命的變異數？

「應用統計學 二版」p173，李德治、童惠玲 著，博碩文化

案例 2 解說

設隨機變數 X 表電視機壽命時間，平均壽命10年 $\rightarrow$ β=10

(1) $P(X > 12) = \int_{12}^{\infty} \frac{1}{10} e^{-\frac{x}{10}} dx = e^{-\frac{12}{10}} = e^{-\frac{6}{5}}$

(2) $P(X < 1) = \int_{0}^{1} \frac{1}{10} e^{-\frac{x}{10}} dx = 1 - e^{-\frac{1}{10}}$

(3) $P(2 < X < 10) = \int_{2}^{10} \frac{1}{10} e^{-\frac{x}{10}} dx = e^{-\frac{2}{10}} - e^{-\frac{10}{10}} = e^{-\frac{1}{5}} - e^{-1}$

(4) $V(X) = \beta^2 = 10^2 = 100$

Gamma distribution

Gamma分配

函數定義（分配形式）

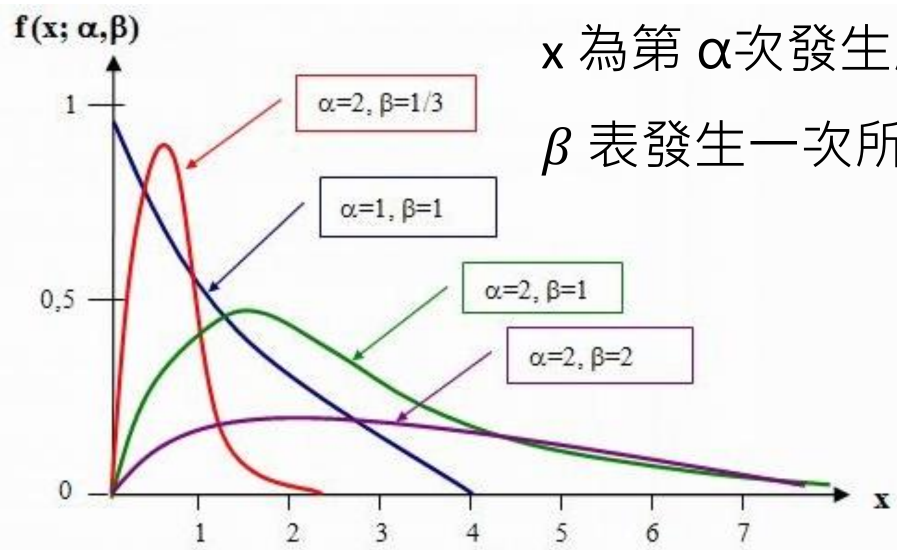
Gamma分配主要用在描述第 α 次事件發生所需的時間，而指數分配是在描述第一次發生所需的時間，所以可以說指數分配是Gamma分配的特例。Gamma分配應用在工業上的用途十分廣泛，例如在增加系統的可靠度上，會設計類似 α-1 個備份元件藉以延長壽命。這種直到 α 次事件發生損壞所需的時間，就服從Gamma分配。

$$f(x) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}},$$

$$0 < x < \infty, \alpha > 0, \beta > 0$$

x 為第 α次發生所需時間；

β 表發生一次所需的時間；



Gamma 函數

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$$

Gamma 函數性質：

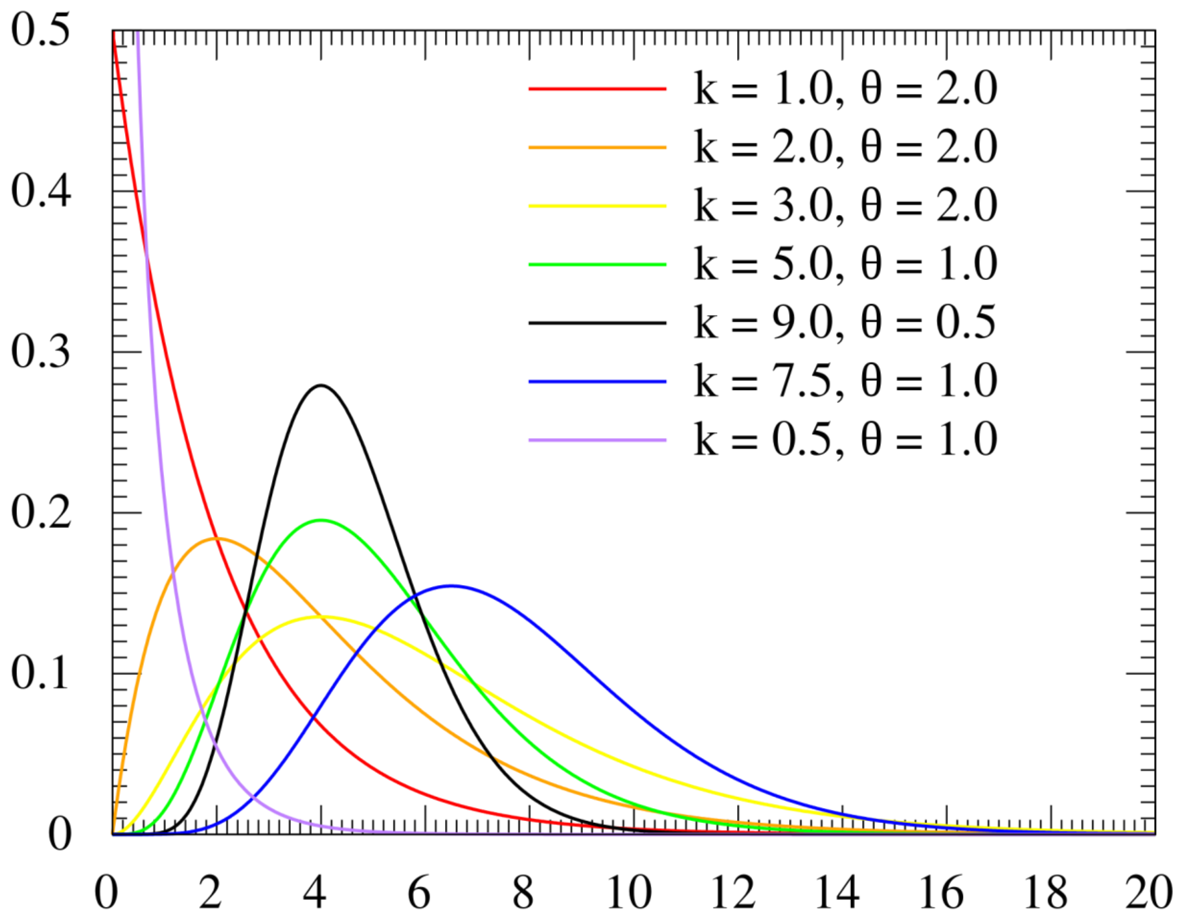
$$\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha)$$

$$\Gamma(\alpha+1) = \alpha\,!$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

分配的重要參數與特性

期望值 $E(x) = \alpha \cdot \beta$ 變異數 $V(x) = \alpha \cdot \beta^2$



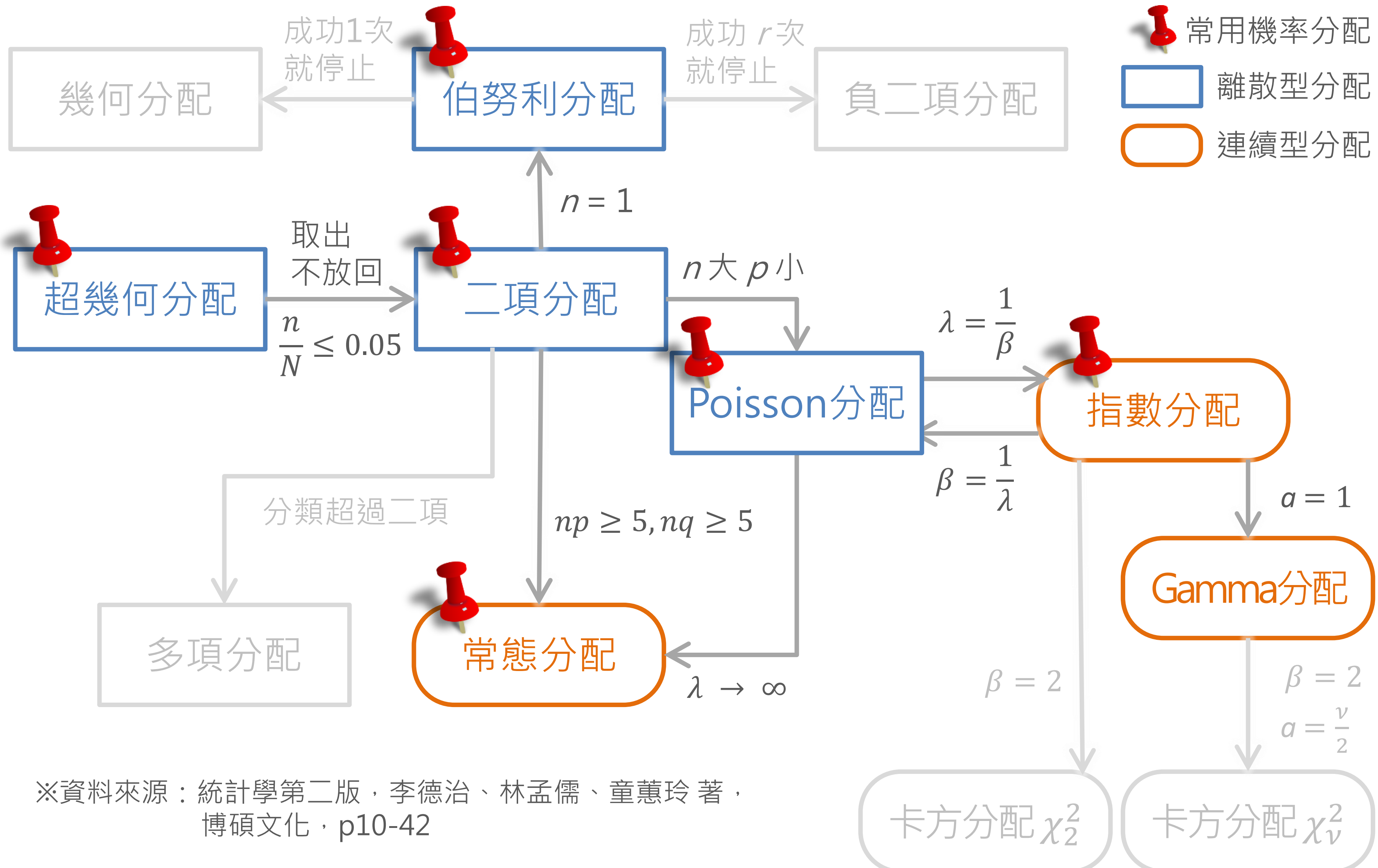
$k \rightarrow \alpha$

$\theta \rightarrow \beta$

about probability distribution

關於機率分配的一些事

各種機率分配的關係



※資料來源：統計學第二版，李德治、林孟儒、童蕙玲 著，
博碩文化，p10-42

The End