Discrete Probability Distributions

離散型機率分配

離散型機率分配

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　二項分配 多項分配　負二項分配 超幾何分配　幾何分配 伯努利分配　波松分配
	連續型 機率分配	連續均勻分配　常態分配 標準常態分配　指數分配 Gamma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

常見離散型機率分配

分配類型	適用情境說明	分配函數
均勻分配 discrete uniform distribution	無論隨機變數如何變化，其機率值為固定常數，例如投擲一次骰子，機率都是1/6	$f(x) = \begin{cases} \frac{1}{n}, x = 1,2,..,n \\ 0, \text{other} \end{cases}$
二項分配 binomial distribution	在成功與失敗的機率不變情況下，不斷重複試驗所得的機率分配函數	$f(x) = C_x^n\, p^x\, q^{n-x}$, x=0,1,..,n 其中p為成功機率；q為失敗機率
伯努利分配 bernoulli distribution	在二項分配下，不論成功或失敗，只做一次（n=1），算是二項分配的特例	$f(x) = p^x\, q^{1-x}$, x=0,1 其中p為成功機率；q為失敗機率
超幾何分配 hyper geometric distribution	從N物中採抽出不放回抽樣，抽出n個樣本，其中N物中，有S個相同，N-S個不同。所以每次抽樣都會受前次抽樣影響	$f(x) = \frac{C_x^S\, C_{n-x}^{N-S}}{C_n^N}$, $0 \le x \le \min(n,S)$
波松分配 Poisson distribution	在一個單位時段或區段內，某事件發生次數（X）的問題	$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$, x = 0,1,2 ... λ為平均次數

其他離散型機率分配

分配類型	適用情境說明	分配函數
多項分配 multinomial distribution	二項分配的延伸，二項分配是把母體分成二種類別，而多項分配則將母體分類超過兩個以上	$f(x_1, x_2, \ldots, x_k)$ $= \dfrac{n!}{x_1!\ x_2!\ \cdots\ x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$ $= \mathrm{C}^{\mathrm{n}}_{x_1, x_2, \ldots, x_k}\, p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$
負二項分配 negative binomial distribution	第 r 次成功所需試驗的次數 （每次都是伯努利試驗） （X表示第 r 次事件發生時所需試驗次數）	$f(x) = \ C^{\mathrm{x-1}}_{\mathrm{r-1}}\ \mathrm{p^r}\ \mathrm{q^{x-r}}$, $x = r,\ r+1,\ r+2,\ \ldots..$ 其中p為成功機率；q為失敗機率
幾何分配 geometric distribution	不斷進行伯努利分配，直到第1次成功為止所需試驗次數之機率函數	$f(x) = \ \mathrm{q^{x-1}p}$, x=0,1,2,…. 其中p為成功機率；q為失敗機率

discrete uniform distribution

均匀分配

函數定義（分配形式）

無論隨機變數如何變化，其機率值為固定常數，例如投擲一次骰子，機率都是1/6

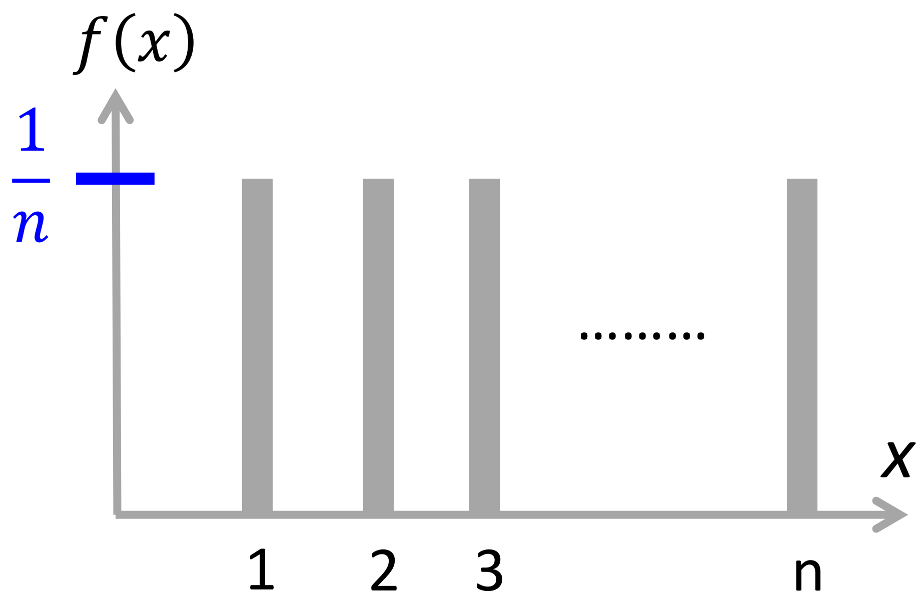
$$f(x)=\begin{cases}\dfrac{1}{\mathrm{n}} & ,\mathrm{x}=1,2,..,\mathrm{n}\\[2ex] 0 & ,\mathrm{other}\end{cases}$$

若我們對母體的情況未知，通常我們採用隨機方式抽樣時，最常就是預設事件發生的機率服從均勻分配。例如

抽籤時，每支籤的機率相同；

隨機調查大樓住戶的滿意度，每戶被抽到的機率相同；

擲骰子的機率、抽撲克牌的機率等



均勻分配的重要參數

期望值 $$E(x)=\frac{n+1}{2}$$

變異數 $$V(x)=\frac{n^2-1}{12}$$

證明：

$$E(x)=\sum_{x=1}^{n} xf(x)=$$

$$E\left(x^2\right)=\sum_{x=1}^{n} x^2 f(x)=$$

$$V(x)=\sigma^2=E\left(x^2\right)-〔E(x)〕^2=$$

binomial distribution

二項分配

函數定義（分配形式）

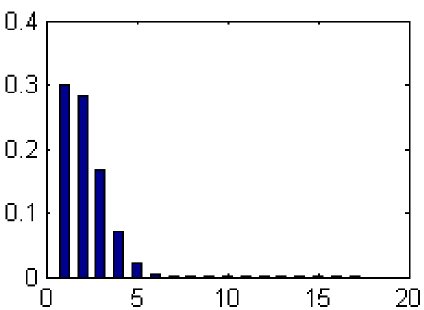
在成功與失敗的機率不變情況下，不斷重複試驗所得的機率分配函數

$$f(x) = C_{\mathrm{x}}^{\mathrm{n}}\, \mathrm{p}^{\mathrm{x}}\, \mathrm{q}^{\mathrm{n-x}}, \mathrm{x}=0,1,..,\mathrm{n}$$

其中 p 為成功機率；

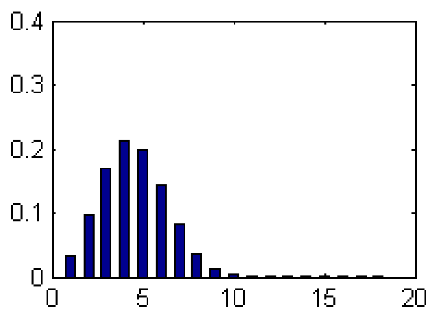
q=1-p 為失敗機率

P = 0.1



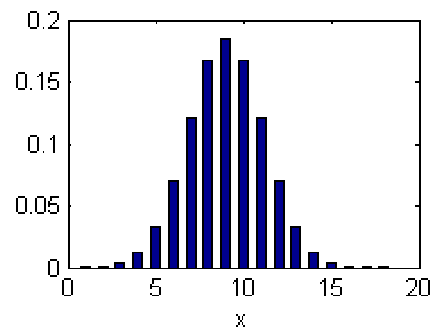
右偏

P = 0.25



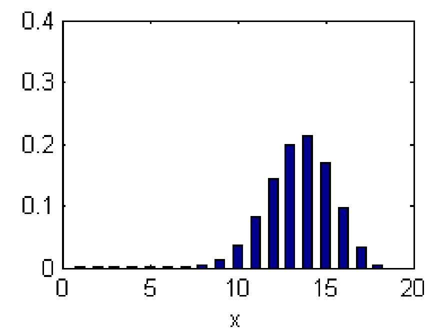
右偏

P = 0.5



對稱

P = 0.75



左偏

分配的重要參數

期望值 $E(x) = np$

變異數 $V(x) = npq = np(1-p)$

證明：

$$E(x) = \sum_{x=1}^{n} xf(x) =$$

$$E(x^2) = \sum_{x=1}^{n} x^2 f(x) =$$

$$V(x) = \sigma^2 = E(x^2) - 〔E(x)〕^2 =$$

案例

假設台北市約有40%的人們喜歡棒球運動，現於市區中隨機訪問1人，試問(a)此人喜歡棒球運動的期望值為何？變異數與標準差為何？(b)若隨機訪問5人，試問我們可以期望5人中喜歡棒球的人數與變異數、標準差為何？有2人喜歡棒球運動的機率為何？至少有3個人喜歡棒球運動的機率為何？

「現代統計學」p88，吳柏林 著，五南圖書

案例解說

(a)（隨機訪問1人）此人喜歡棒球運動的期望值為何？變異數與標準差為何？

令隨機變數 X 代表喜歡棒球與否，則 執行次數 n=1 ；p=0.4

期望值 E(X) =np=(1)(0.4)=0.4

變異數 V(X) =σ^2=pq=np(1-p)=(1)(0.4)(1-0.4)=0.24　　標準差 σ =$\sqrt{0.24}$ ≒ 0.49

(b)若隨機訪問5人，試問我們可以期望5人中喜歡棒球的人數與變異數、標準差為何？
有2人喜歡棒球運動的機率為何？至少有3個人喜歡棒球運動的機率為何？

令隨機變數 X 代表喜歡棒球的人數，則 執行次數 n=5 ；p=0.4

期望值 E(X) =np=(5)(0.4)=2

變異數 V(X) =σ^2=pq=np(1-p)=(5)(0.4)(1-0.4)=1.2　　標準差 σ =$\sqrt{1.2}$ ≒ 1.1

P(X=2)= $C_x^n\, p^x\, q^{n-x} = C_2^5\,(0.4)^2(1-0.4)^{5-2} = \frac{5\times 4}{1\times 2}\times 0.0346 = 0.346$

P(X≧3)=1-P(X≦2)=1-(P(X=0)+P(X=1)+P(X=2))=1- $\sum_{x=0}^{2} C_x^5 (0.4)^x (0.6)^{5-x}$ =0.317

bernoulli distribution

伯努利分配

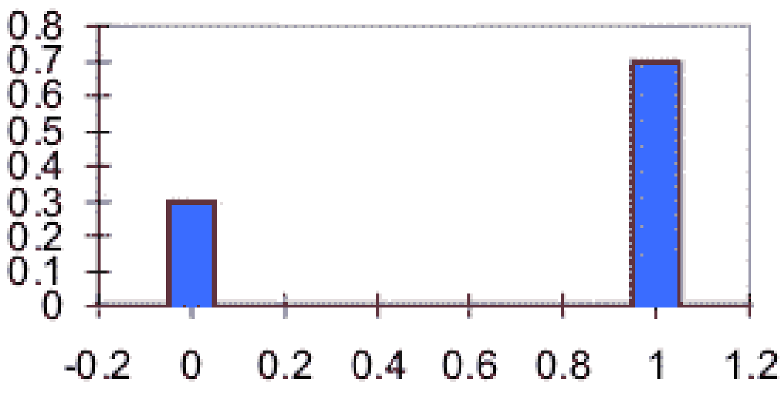
函數定義（分配形式）

在二項分配下，不論成功或失敗，只做一次（n=1），算是二項分配的特例

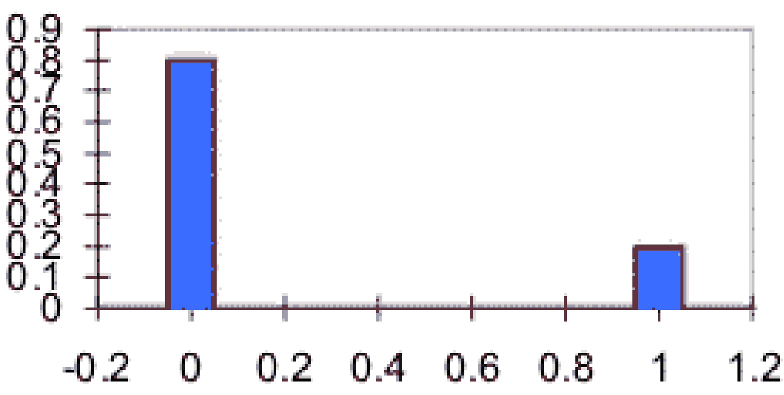
$$f(x) = \mathrm{p}^{\mathrm{x}}\,\mathrm{q}^{1-\mathrm{x}}, \mathrm{x}=0,1$$

其中 p 為成功機率；
q=1-p 為失敗機率

P = 0.7



P = 0.2



分配的重要參數

期望值　　$E(x) = p$

變異數　　$V(x) = pq = p(1-p)$

證明：

$$E(x) = \sum_{x=1}^{n} xf(x) =$$

$$E(x^2) = \sum_{x=1}^{n} x^2 f(x) =$$

$$V(x) = \sigma^2 = E(x^2) - 〔E(x)〕^2 =$$

hyper geometric distribution

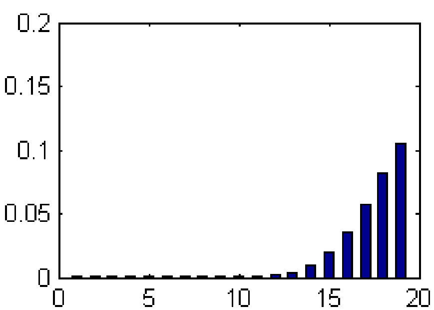
超幾何分配

函數定義（分配形式）

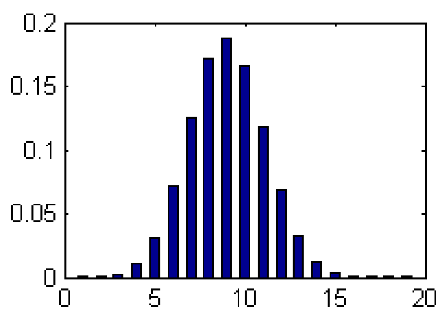
從 N 物中採抽出不放回抽樣，抽出 n 個樣本，其中 N 物中，有 S 個相同，N-S 個不同。所以每次抽樣都會受前次抽樣影響

$$f(x) = \frac{C_x^S \, C_{n-x}^{N-S}}{C_n^N} \qquad 0 \le x \le \min(n,S)$$

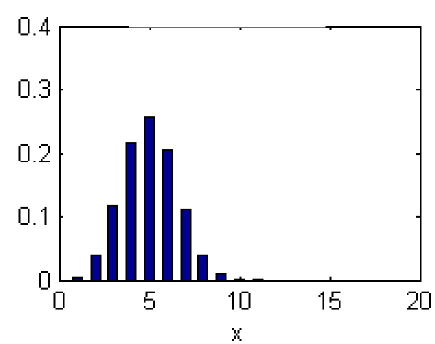
n = 50



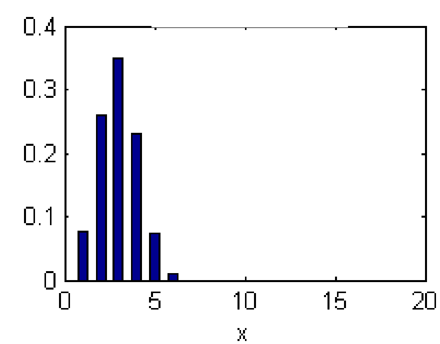
n = 20



n = 10



n = 5



分配的重要參數

期望值 $E(x) = np$ 其中 $p = \frac{n}{N}$；

$q=1-p$

變異數 $V(x) = \frac{N-n}{N-1} npq$

有限母體修正因子
(finite population correction)

證明：

$$E(x) = \sum_{x=1}^{n} xf(x) =$$

$$E(x^2) = \sum_{x=1}^{n} x^2 f(x) =$$

$$V(x) = \sigma^2 = E(x^2) - 〔E(x)〕^2 =$$

當母體總數很大時（N→∞），每次抽樣成功的機率都可以視為成功：

$$\lim_{N\to\infty} \frac{N-n}{N-1} = 1，V(x) = npq$$

此時超幾何分配會趨近二項分配，一般以 n/N ≤ 0.05 時作為判別

案例

1.從6男3女中隨機抽出4人組成委員會，求此委員會中男性人數與女性人數的機率分配？

2.一批產品共有10件，其中含有2件不良品？今隨機抽取3件，求均為良品的機率？

「應用統計學 二版」p147，李德治、童惠玲 著，博碩文化

案例解說

1）令 X 表委員會中男性人數、Y表委員會中女性人數，符合超幾何分配：
所以 N=6+3=9；S=6 ；n=4

$$f(x) = \frac{C_x^6\, C_{4-x}^{9-6}}{C_4^9}\ , x=1,2,3,4 \qquad f(y) = \frac{C_x^3\, C_{4-y}^{9-3}}{C_4^9}\ , y=0,1,2,3$$

2）令 X 表不良品件數，符合超幾何分配，所以 N=10；S=2 ；n=3 ，所以

$$f(x) = \frac{C_x^2\, C_{3-x}^{10-2}}{C_3^{10}}\ , x=0,1,2$$

均為良品的機率：

$$f(0) = \frac{C_0^2\, C_{3-0}^{10-2}}{C_3^{10}} = \frac{7}{15}$$

超幾何分配與二項分配

	超幾何分配	二項分配
選取方式	取後不放回	取後放回
是否獨立	否	是
期望值	np	np
變異數	$npq\frac{N-n}{N-1}$	npq

超幾何分配的近似分配

當 $n/N \leqq 0.05$ ，$(N-n)/(N-1)$ 趨近於1

超幾何分配 $\xrightarrow{n/N \leqq 0.05}$ 二項分配

Poisson distribution

波松分配

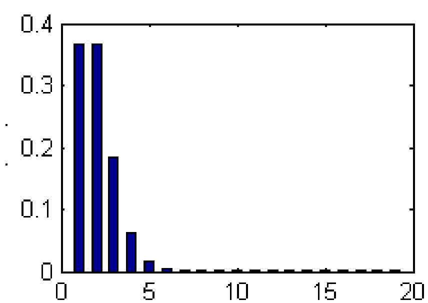
函數定義（分配形式）

在一個單位時段或區段內，某事件發生次數（X）的問題（常用於關於時間、長度、面積、...等問題）

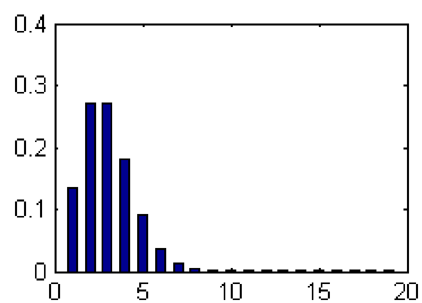
$$f(x) = \frac{e^{-\lambda}\,\lambda^{\mathrm{x}}}{\mathrm{x}\,!}\ ,\mathrm{x} = 0,1,2\ \ldots$$

λ 為平均次數

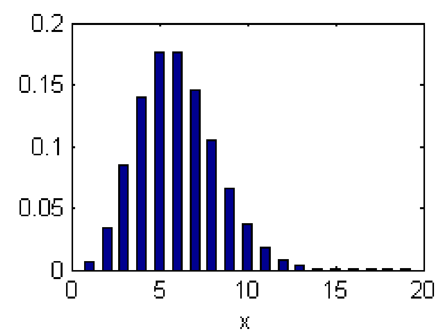
λ = 1



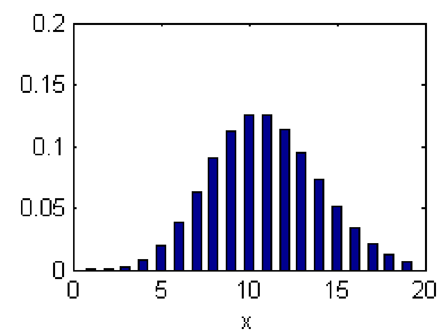
λ = 3



λ = 5



λ = 10



分配的重要參數

期望值 $E(x) = \lambda$

變異數 $V(x) = \lambda$

證明：

$$E(x) = \sum_{x=1}^{n} xf(x) =$$

$$E(x^2) = \sum_{x=1}^{n} x^2 f(x) =$$

$$V(x) = \sigma^2 = E(x^2) - 〔E(x)〕^2 =$$

案例

台灣地處亞熱帶，常受颱風過境的自然災害困擾。每年夏季6月至9月為颱風侵襲台灣的主要季節。中央氣象局統計資料指出，平均每年約有5個颱風過境台灣。試問明年沒有颱風過境台灣的機率為多少？明年將有5個颱風過境台灣的機率為多少？明年超過7個以上颱風過境台灣的機率為多少？

「現代統計學」p92，吳柏林 著，五南圖書

案例解說

(a)沒有颱風過境台灣的機率為多少？
令隨機變數 X 代表每年颱風過境台灣次數，本題符合 λ=5 的Poisson分配

$$P(X=0 \mid \lambda=5)=\frac{e^{-\lambda}\,\lambda^{x}}{x\,!}=\frac{e^{-5}\,5^{0}}{0\,!}=0.0067$$

(b)將有5個颱風過境台灣的機率為多少？

$$P(X=5 \mid \lambda=5)=\frac{e^{-\lambda}\,\lambda^{x}}{x\,!}=\frac{e^{-5}\,5^{5}}{5\,!}=0.1755$$

(c)超過7個以上颱風過境台灣的機率為多少？

$$P(X \geqq 7 \mid \lambda=5) = 1-P(X \leqq 6) = 1-〔P(X=0)+P(X=1)+P(X=2)+\ldots+P(X=6)〕$$

$$= 1-〔\textstyle\sum_{k=0}^{6}\frac{e^{-\lambda}\,\lambda^{k}}{k\,!}〕 = 1-〔\frac{e^{-5}\,5^{0}}{0\,!}+\frac{e^{-5}\,5^{1}}{1\,!}+\ldots+\frac{e^{-5}\,5^{6}}{6\,!}〕 = 1-0.7622$$

$$= 0.2378$$

Poisson的近似分配

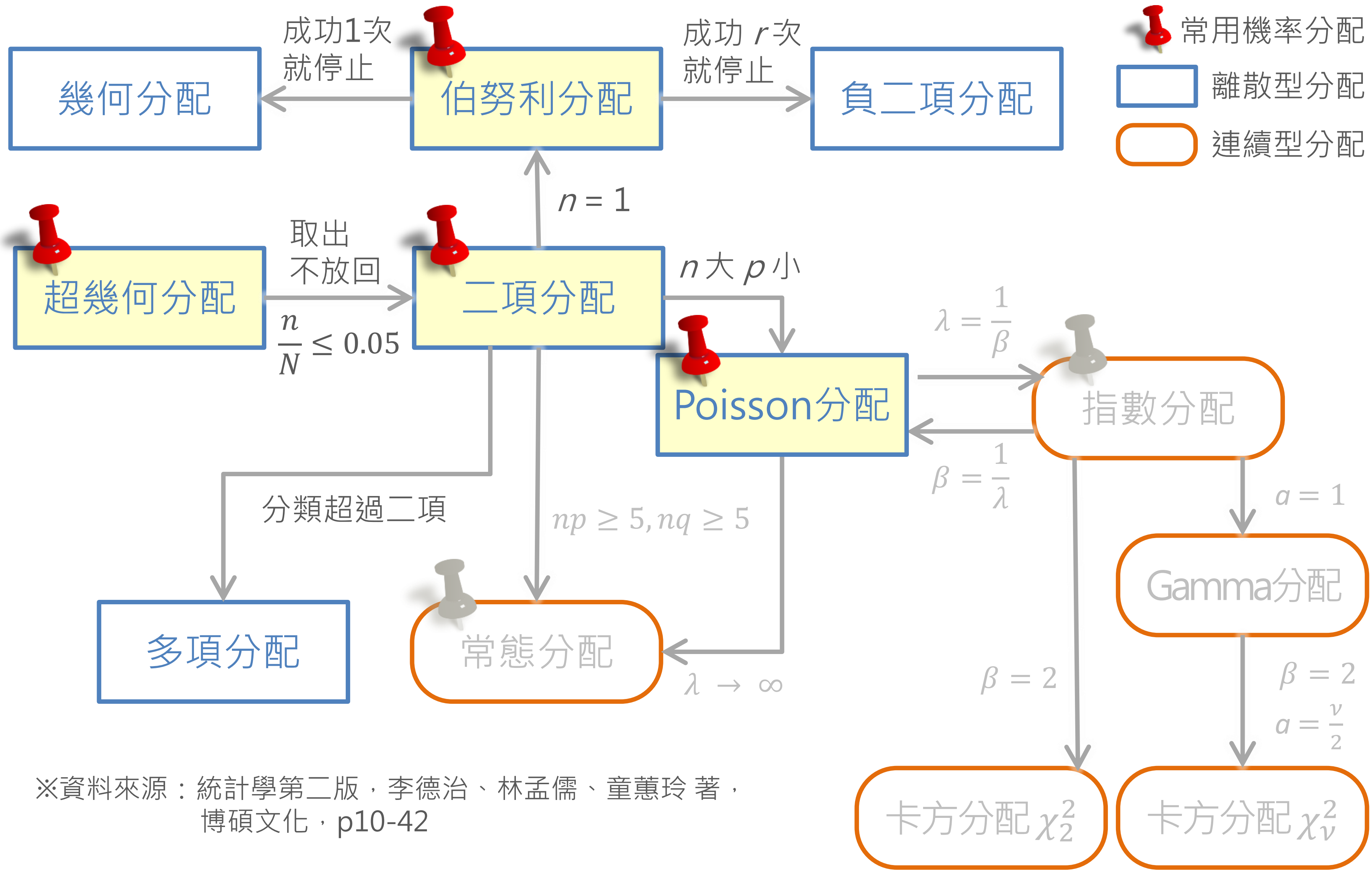
當二項分配成功機率 $p \to 0$ 且 $n \to \infty$ 時，二項分配會趨近Poisson分配。因為Poisson分配的機率計算較容易，所以當二項分配的成功機率很小時，就可以用Poisson分配來計算

表示成功機率很小時

$p \to 0$ 且 $n \to \infty$

$$\text{二項分配} \xrightarrow[\lambda = np]{} \text{Poisson分配}$$

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