第二章 統計概念回顧: 機率部份

September 29, 2010

There are three kinds of lies: lies, damned lies, and statistics — Benjamin Disraeli (1804-1881), 英國國會議員與作家.

觀念架構：機率與分配

1. 統計學是處理「不確定性 (uncertainty)」的一門科學。
   計學可略分「敘述統計」與「推論統計」二類，前者只是利用一些數字將資料組織與彙整，而後者則是發展各種方法，對資料的特性與內涵做深度的探討。
2. Question:「擲骰子」是否為一個隨機變數？

觀念架構: 機率與分配

1. 統計學是處理「不確定性 (uncertainty)」的一門科學。
   計學可略分「敘述統計」與「推論統計」二類, 前者只是利用一些數字將資料組織與彙整, 而後者則是發展各種方法, 對資料的特性與內涵做深度的探討。
2. Question:「擲骰子」是否為一個隨機變數?
   Answer: No! 「擲骰子」是一個隨機實驗 (random experiment)。

隨機變數

隨機實驗	擲一骰子
⇩	⇩
出象	一至六點的點數
⇩	⇩
出象空間	{一點、二點、⋯、六點}
⇩	⇩
事件	{一點、三點、五點}，{二點、四點、六點}
⇩	⇩
隨機變數	$\begin{cases} x(\text{奇數點}) = 1 \\ x(\text{偶數點}) = -1 \end{cases}$

Figure: 機率的架構

隨機變數

1. 隨機實驗 (random experiment): 例如, 擲一骰子。
2. 出象 (outcome): 一至六點。
3. 樣本空間 (sample space) 或出象空間:$\Omega = \{1, 2, 3, 4, 5, 6\}$。
4. 事件 (event): 即樣本空間的子集合 (subset)。例如, 奇數點:$A = \{1, 3, 5\}$, 或偶數點:$B = \{2, 4, 6\}$。
5. 隨機變數 (random variable): 隨機變數為由一個樣本空間映至 (map) 實數象限的函數。

事件的關係

- 二個常見的事件關係為互斥事件與補集事件
- 迪摩根定律 (De Morgan's law):
  令A與B為一出象空間中的二個事件，則下列為真:
  $(A \cup B)^c = A^c \cap B^c$
  $(A \cap B)^c = A^c \cup B^c$

機率測度

一隨機試驗的機率測度P, 須滿足下列的條件:

1. 對任一事件A, 其發生的機率至少為零, 至多為一, 亦即$0 \leq P(A) \leq 1$。
2. 總機率的和為一,$P(\Omega) = 1$。
3. 對互斥事件的事件:$A_1, A_2, A_3, \cdots, A_n$, 其發生的聯合機率等於其個別事件機率之加總。亦即:

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

條件機率

所謂的條件機率(conditional probability),$P(A|B)$, 即給定B事件發生, A事件發生的機率。定義如下:

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

非條件機率

我們稱$P(A)$為非條件機率 (unconditional probability)(有時也稱邊際機率), 這是因為

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

也就是$P(A)$是所有A可能出現的機率的加總, 包括與B同時發生, 及A發生而B不發生下的機率。

獨立事件(independent events)

令A與B為二事件, 若此二事件為獨立的事件, 則表示B事件的發生與否, 不會影響A事件的發生。

$$P(A|B) = P(A|B^c) = P(A)$$

聯合機率(joint probability)

若A與B非為獨立事件, 則其聯合機率為:

$$P(A \cap B) = P(A) + P(B) - P(A \cup B).$$

貝氏定理(Bayes Theorem)

貝氏定理主要被用來用來做機率的修正 (probability updating)。
若$P(A) \neq \phi$ 且 $P(B) \neq \phi$, 則

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

貝氏定理(Bayes Theorem)

若$A_1, A_2, \cdots A_n$為互斥事件, 且$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$, 則對任一事件$A_i, i = 1, 2, \cdots, n$, 下式成立:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_3)P(B|A_3)}$$

貝氏定理圖示

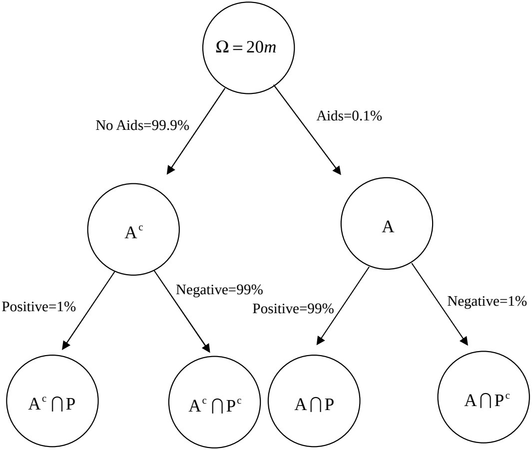

Figure: 貝氏定理圖示

機率分配的架構

從歷史發展的角度來說, 機率的決定主要有二個方式。一是視機率為客觀機率; 一出象的機率取決於其發生的相對頻率(relative frequency), 這就是古典的機率概念, 也是最主流的機率概念。第二個則是貝氏機率, 基本上是由研究者主觀賦予的主觀機率 (subjective probability)。古典的方法基本上只使用客觀的「資料」, 而貝氏機率的優點則是允許研究者將主觀的事前資訊 (prior information) 納入分析中, 並與資料合併求得所謂的事後機率 (posterior probability)。

機率分配的架構

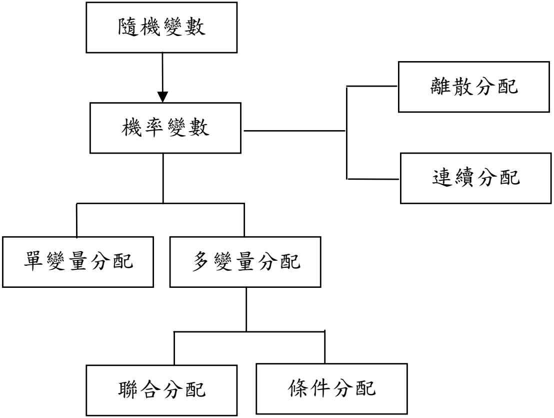

Figure: 機率分配的架構

離散隨機變數

離散隨機變數: 一個隨機變數可能發生的數值為有限多個(finite) 或可數的無窮多 (countably infinite), 稱之。
令x是一個離散的隨機變數, 其值為$x_1, x_2, x_3 \cdots$, 則我們可表達x的機率函數 (probability function)f_x為:

$$f(x = x_i) = Pr(x = x_i)$$

離散隨機變數

我們可定義一隨機變數的期望值$E(x)$為:

$$E(x) \equiv \sum_i x_i P(x_i).$$

我們也可計算x的函數的期望值。令$g(x)$為x的一個函數, 則其期望值為:

$$E(g(x)) \equiv \sum_i g(x_i) P(x_i).$$

離散隨機變數的動差

離散隨機變數的動差(moments):x的r階動差, m_r, $r = 1, 2, \ldots$:

$$m_r \equiv E(x^r) \equiv \sum_i x_i^r f(x_i)$$

中央動差就是將動差「平移」至平均值的動差, 亦即:

$$\mu_r \equiv E((x-\mu)^r) = \sum_i (x_i - \mu)^r f(x_i)$$

白努利分配

白努利分配(Bernoulli distribution): 出象非零即一的隨機變數的分配, 其機率分配函數如下:

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

白努利分配

白努利分配只有一個參數p, 其平均值與變異數分別如下:

$$
\begin{aligned}
E(x) &= 1 \cdot f_x(1,p) + 0 \cdot f_x(0,p) \\
&= 1 \times p + 0 \times (1-p) \\
&= p.
\end{aligned}
$$

$$
\begin{aligned}
var(x) &= (1-p)^2 f_x(1,p) + (0-p)^2 f_x(0,p) \\
&= (1-p)^2 p + p^2 (1-p) \\
&= p(1-p)(1-p+p) \\
&= p(1-p).
\end{aligned}
$$

二項分配

二項分配(binomial distribution): 若$x_1, \cdots, x_n$為n個獨立的白努利隨機變數, 則其加總服從二項分配, 表達為:

$$x = \sum_{i=1}^{n} x_i \sim B(n;p)$$

二項式分配的機率函數如下:

$$Pr(X = x) = C_x^n p^x (1-p)^{n-x}, \quad x = 1, ..., n.$$

其平均值與變異數分別為np與$np(1-p)$。

波氏分配

波氏分配(Possion distribution): 波氏分配捕捉一段固定長度的時間內, 某事件發生的次數的特性。波氏分配來自波松過程 (Poisson process), 其分配如下:

$$Pr(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!},\quad x=0,1,2\ldots$$

連續隨機變數

當隨機變數的值不是離散的, 而是連續的, 我們就稱之為連續隨機變數。我們可定義其累積機率函數(CDF) 如下:

$$F_x(a) = Pr(x \leq a)$$

而其機率密度, 如下:

$$f_x(a) = \lim_{s \to 0} \frac{F_x(a+s) - F_x(a)}{s}$$

連續隨機變數的動差

連續變數x的r階原始動差定義為:

$$m_r \equiv E(x^r) \equiv \int_{-\infty}^{\infty} x^r f_x(x)dx.$$

與離散變數的差別只是將加總符號$\sum$換成積分符號$\int$而已。

偏態係數與峰態係數

偏態係數(coefficient of skewness):

$$\text{skew} \equiv \frac{E(x-\mu)^3}{\sigma_x^3}$$

峰態係數 (coefficient of kurtosis):

$$\kappa \equiv \frac{E(x-\mu)^4}{\sigma_x^4}$$

常態峰、平闊峰與高狹峰

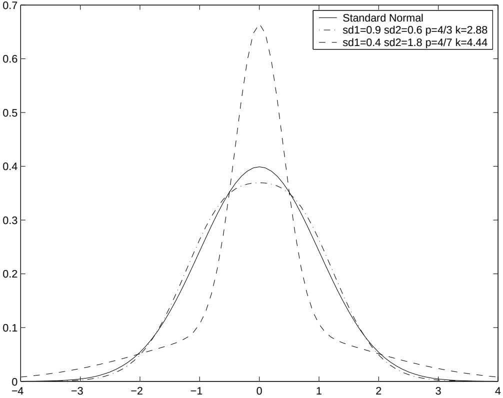

Figure: 常態峰、平闊峰與高狹峰: 以雙常態混合分配爲例

均勻分配

一均勻分佈於$[a, b]$區間的隨機變數之 pdf 如下:

$$f(x) = \frac{1}{b-a}, \quad \text{若 } a \leq x \leq b.$$

指數分配

Gamma分配的 pdf 如下:

$$f(x) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda x} x^{P-1}, \quad x \geq 0, \lambda > 0, P > 0$$

當$\lambda = 1/2$且$P = n/2$, 則 Gamma 分配就變成所謂的卡方分配, 而當$P = 1$時, 就成為指數分配,pdf 如下:

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

指數分配的平均值與變異數分別為λ^{-1}與λ^{-2}。

常態分配

常態分配(normal distribution) 又稱高斯分配 (Gaussian distribution), 其機率密度函數為:

$$f(x|\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2],\quad -\infty<x<\infty$$

其中μ與σ^2分別為其平均數與變異數。

卡方(χ^2) 分配

自由度為一的卡方分配如下：若 $z \sim \mathcal{N}(0,1)$，則

$$k \equiv z^2 \sim \chi^2(1)$$

亦即，自由度為1的卡方隨機變數為標準常態隨機變數的平方。若$k_1, k_2, \cdots, k_n \overset{iid}{\sim} \chi^2(1)$為$n$個互相獨立的自由度為一的卡方隨機變數，則其加總

$$k_1 + \cdots + k_n \sim \chi^2(n)$$

服從自由度為n的卡方分配 ($\chi^2(n)$)。

t分配

t分配又稱 Student's t distribution。若 $z \sim \mathcal{N}(0,1)$, $k \sim \chi^2(\nu)$, 且z與k為相互獨立的隨機變數, 則

$$t \equiv \frac{z}{\sqrt{k/\nu}} \sim t_\nu$$

t分配

t分配主要是拿來與「標準常態」比較，其主要特色如下:

1. 當$\nu > 1$時,t 分配的均數$E(t)$才存在，且$E(t) = 0$。
2. 當$\nu > 2$時,$var(t)$存在,$var(t) = \frac{\nu}{\nu-2} > 1$。
3. 當$\nu > 3$時，第三階動差才存在,$\frac{E[x-E(x)]^3}{\sigma_x^3} = 0$。
4. 當$\nu > 4$時，第四階動差才存在，

$$\begin{aligned}\frac{E[x-E(x)]^4}{\sigma_x^4} &= \frac{3\nu^2}{(\nu-2)(\nu-4)} \\ &> 3\end{aligned}$$

顯示尾端較標準常態肥厚 (fat tails)。

5. 當自由度ν趨近於無窮大時,t分配會趨近標準常態分配，也就是 $t \xrightarrow{p} \mathcal{N}(0,1)$ 當 $\nu \to \infty$。

F 分配

$x \sim \chi^2(m)$,$y \sim \chi^2(n)$, 且 x 與y 為互相獨立之隨機變數, 則

$$F \equiv \frac{x/m}{y/n} \sim F(m,n)$$

m與n為F分配的二個參數, 分別為分子與分母的自由度。

對數常態分配

若一變數x在取自然對數後, 服從常態分配, 則稱此變數服從對數常態分配 (lognormal distribution)。因此, 顯然x恆為正, 亦即 $x \in R^+$。對數常態分配的 pdf 如下:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp(-\frac{1}{2}[(\ln\ x - \mu)/\sigma]^2), \quad x > 0$$

聯合分配

以連續的雙變數為例,X與Y的聯合分配 (joint distribution),$F(x,y)$, 定義為

$$\begin{aligned} F(x,y) &= Pr(X \le x, Y \le y) \\ &= \int_{-\infty}^{y}\int_{-\infty}^{x} f(s,t)dsdt \end{aligned}$$

上面$f(x,y)$代表的是X,Y聯合 pdf (joint pdf)。
聯合分配同樣要服從我們在前面所討論過的機率所應有的條件, 也就是

$$\begin{aligned} f(x,y) &\ge 0 \qquad \forall x,y \in R, \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy &= 1. \end{aligned}$$

邊際分配

一般探討的單變量分配, 其實就是不考慮其他變數下的分配, 也就是將其他變數的各種可能性“積分掉”。

x 的分配可表示如下:

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

這樣的分配也稱為「邊際分配 (marginal distribution)」其實就是x的「非條件分配 (unconditional distribution)」, 亦即在無任何y的資訊下,x的分配。

聯合動差

定義x, y的第(p, q)階聯合中央動差$m_{xy}^{p,q}$為:

$$
\begin{aligned}
m_{xy}^{p,q} &\equiv E((x-\mu_x)^p(y-\mu_y)^q) \\
&= \int\int (x-\mu_x)^p(y-\mu_y)^q f(x,y)dxdy, \quad p, q = 1, 2, \cdots
\end{aligned}
$$

當$p = q = 1$時,

$$
\begin{aligned}
m_{xy} &= E[(x-\mu_x)(y-\mu_y)] \\
&\equiv cov(x, y)
\end{aligned}
$$

就是一般熟知的共變異數 (covariance), 也常以σ_{xy}表達。

相關係數

共變異數σ_{xy}捕捉x與y的一次式的關聯性 (也就是線性相關程度)。σ_{xy}的值會受到衡量單位的影響, 所以一般也會將其"標準化"為「相關係數 (correlation coefficient)」。

$$\rho_{xy} \equiv \frac{\sigma_{xy}}{\sigma_x \sigma_y},$$

相關係數的特色是其值恆介於-1與1之間, $\sigma_{xy} \in [-1, 1]$。當$\rho_{xy} < 0$時, 表示x, y間是負的線性相關, $\rho_{xy} > 0$則表示二者有正的線性相關, 而$\rho_{xy} = 0$則表示二者無線性相關。

相關與獨立的討論

變數間的相關係數為零是否就表示變數間是彼此是獨立的? 答案是否定的。要記得「獨立」的觀念與隨機變數的「值」大小無關。「獨立」指的是變數背後的"事件"的發生是"無關"的, 而「相關係數」只衡量線性相關程度。

假如x服從$(-1,1)$的均勻分配,$x \sim U(-1,1)$, 而$y = s\sqrt{1-x^2}$, 其中s為與x獨立的隨機變數, 且服從機率$p = \frac{1}{2}$的白努利分配, 但值為1與-1。所以y的分配如下:

$$y = \begin{cases} \sqrt{1-x^2}, & \text{機率} = \frac{1}{2}, \\ -\sqrt{1-x^2}, & \text{機率} = \frac{1}{2}. \end{cases}$$

相關與獨立的討論

因為$y^2 = 1 - x^2$, 所以$x^2 + y^2 = 1$。可知x, y的相關係數為0, 表示二者無線性相關, 但是因為$y^2 = 1 - x^2$, 則可知x^2與y^2的相關係數為-1, 因為

$$
\begin{aligned}
cov(x^2, y^2) &= cov(x^2, 1 - x^2) = -var(x^2), \\
var(y^2) &= var(1 - x^2) = var(x^2) \\
\rho_{x^2 y^2} &= \frac{cov(x^2, y^2)}{\sqrt{var(x^2) var(y^2)}} = -1.
\end{aligned}
$$

可見x與y並不獨立!

相關與獨立的討論

「獨立」指的是變數背後的“事件”的發生是“無關”的，而「相關係數」只衡量線性相關程度。若變數間是獨立的，則也必定是無相關的，但反之則未必成立。

例如在財務與經濟的實證上，最常見的現象就是不同時點的資產報酬率(例如$r_t = \ln p_t - \ln p_{t-1}$)，或是經濟變數的差分 (difference)，例如$\Delta y_t = y_t - y_{t-1}$，是無線性相關的，但是其平方卻是相關的，亦即

$$cov(r_{t-1}, r_t) \approx 0$$

$$cov(r_{t-1}^2, r_t^2) > 0$$

顯示變數間是不相關的，但卻不是獨立的。

條件分配與條件動差

以x,y二變數為例, 所謂條件分配 (conditional distribution) 就是給定另一個變數的資訊下, 一個變數的分配。例如, 給定$X = x$的資訊下,y的條件分配,$f(y \mid x)$, 為:

$$f(y \mid x) = \frac{f(x, y)}{f_x(x)}$$

如果x與y互相獨立的話, 則其條件分配與非條件分配相同:

$$f(y \mid x) = f_y(y)$$

給定$X = x$下,y的條件期望值為:

$$E(y \mid x) = \int y f(y \mid x) dy$$

條件分配與條件動差

幾個常用的定理如下:

1. 遞迴期望法則 (Law of iterated expectations)

$$E(y) = E_x(E(y \mid x))$$

2. 若$E(y \mid x) = \alpha + \beta x$, 則

$$\begin{aligned} \beta &= \frac{cov(x, y)}{var(x)} \\ \alpha &= E(y) - \beta E(x) \end{aligned}$$

3. 變異數分解 (decomposition of variance)

$$var(y) = var_x(E_y(y \mid x)) + E_x(var_y(y \mid x))$$

多變量常態分配

多變量分配, 指的只是超過一個以上的隨機變數的聯合分配(joint distribution)。若$x_1, x_2, \cdots, x_n$有聯合常態分配, 我們稱$x = (x_1, ..., x_n)'$服從多變量常態分配 (multivariate normal distribution), 表達為:

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

而其 pdf 可證明為:

$$f(x|\mu,\Sigma) = \left(\frac{1}{\sqrt{2\pi}}\right)^n |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu))$$

多變量常態分配

幾個重要的性質如下:

- 若$n = 1$, 上式就會縮減為一般的單變量常態 pdf:

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}}(\sigma^2)^{-\frac{1}{2}} \exp(-\frac{1}{2}(x-\mu)\sigma^{-2}(x-\mu)) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)
\end{aligned}
$$

- x_i的邊際分配 (marginal distribution) 亦為常態, 亦即$x_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$。
- $cx \sim \mathcal{N}(c\mu, c\Sigma c')$, 亦即多變量常態變數之線性組合仍為常態。
- 在常態分配下,「不相關」等於「獨立」。
- 若$x \sim \mathcal{N}(\mu, \Sigma)$, 則每一個元素也會服從常態分配, 亦即: $x_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$。但反之則未必成立, 也就是說, 即便個別的隨機變數都服從(邊際) 常態分配, 但其聯合分配未必是(多維) 常態。

條件常態分配

條件常態分配(conditional normal distribution):
若 $x = (x_1', x_2')'$, x為 $(n \times 1)$, x_1 為 $(n_1 \times 1)$, 且 x_2為 $(n_2 \times 1)$, 又$n = n_1 + n_2$, 亦即:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

平均值與共變異矩陣分別為:

$$\begin{aligned} \mu &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \\ \Sigma &= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix} \end{aligned}$$

則給定x_2資訊下,x_1的條件平均值與條件共變異矩陣分別為:

$$\begin{aligned} E(x_1|x_2) &= \mu_1 + \Sigma_{22}^{-1}\Sigma_{12}(x_2 - \mu_2) \\ cov(x_1|x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}' \end{aligned}$$

條件常態分配

若x與y有二元聯合常態分配, 則給定x的值下,y的條件期望值為:

$$
\begin{aligned}
E(y|x) &= \mu_y + \frac{cov(x,y)}{var(x)}(x-\mu_x) \\
&= \alpha + \beta x
\end{aligned}
$$

其中

$$
\begin{aligned}
\beta &= \frac{cov(x,y)}{var(x)} \\
\alpha &= \mu_y - \beta\mu_x
\end{aligned}
$$

條件變異為:

$$
\begin{aligned}
var(y|x) &= var(y) - \frac{cov(x,y)^2}{var(x)} \\
&= var(y) - \beta^2 var(x).
\end{aligned}
$$

常態分配二次式之分配

若x服從多維標準常態分配 (multivariate standard normal distribution), 表達為:

$$x \sim \mathcal{N}(0, I)$$

則其二次式$x'x$服從卡方分配, 自由度為n:

$$x'x = \sum x_i^2 \sim \chi^2(n)$$

即, 若$x \sim \mathcal{N}(\mu, \Sigma)$, 其中$\Sigma$為一正定矩陣則

$$(x-\mu)'\Sigma^{-1}(x-\mu) \sim \chi^2(n).$$

常態分配二次式之分配

因為Σ為對稱且正定, 因此存在一矩陣 P, $(n \times n)$, 使得$P\Sigma P' = I$, 且$P'P = \Sigma^{-1}$。所以,

$$x - \mu \sim \mathcal{N}(0, \Sigma)$$

令$Z = P(x - \mu)$, 則

$$Z \sim \mathcal{N}(0, P\Sigma P') \sim \mathcal{N}(0, I)$$

因此

$$Z'Z = (x - \mu)'P'P(x - \mu) = (x - \mu)'\Sigma^{-1}(x - \mu) = \sum_{i=1}^{n} z_i^2 \sim \chi_n^2.$$

常態分配二次式之分配

因為Σ可「對角化」為 $\Sigma = C\Lambda C'$，令$P = \Lambda^{-\frac{1}{2}}C'$，則

$$P' = (\Lambda^{-\frac{1}{2}}C')' = C(\Lambda^{-\frac{1}{2}})' = C\Lambda^{-\frac{1}{2}}$$

其中$\Lambda^{-\frac{1}{2}} = diag(1/\sqrt{\lambda_1}, \cdots, 1/\sqrt{\lambda_n})$。所以，

$$\begin{aligned} P\Sigma P' &= \Lambda^{-\frac{1}{2}}C'C\Lambda C'C\Lambda^{-\frac{1}{2}} \\ &= \Lambda^{-\frac{1}{2}} \cdot I \cdot \Lambda \cdot I \cdot \Lambda^{-\frac{1}{2}} \\ &= \Lambda^{-\frac{1}{2}} \cdot \Lambda \cdot \Lambda^{-\frac{1}{2}} \\ &= I_n \end{aligned}$$

而且

$$\begin{aligned} P'P &= C\Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}C' \\ &= C\Lambda^{-1}C' \\ &= \Sigma^{-1} \end{aligned}$$

常態分配二次式之分配

因為$\Sigma = C\Lambda C'$, 所以

$$\begin{aligned}
\Sigma^{-1} &= (C\Lambda C')^{-1} \\
&= (C')^{-1}\Lambda^{-1}C^{-1} \\
&= (C^{-1})^{-1}\Lambda^{-1}C^{-1} \\
&= C\Lambda^{-1}C'.
\end{aligned}$$

注意因為C為正交矩陣, 所以$C' = C^{-1}$。