

第三章

多維隨機變數

3.8 不等式

在數學裡, 不等式的角色很重要。機率及統計裡, 也有一些不等式常出現。本節我們介紹一些。

引理8.1 設 $u, v, p, q > 0$, 且

$$(8.1) \quad \frac{1}{p} + \frac{1}{q} = 1。$$

則

$$(8.2) \quad \frac{u^p}{p} + \frac{v^q}{q} \geq uv,$$

且等號只有在 $u = v^{1/(p-1)}$ 才成立。

證明. 首先由(8.1)式及 $p, q > 0$, 立即得 $p > 1, q > 1$ 。現固定 v , 並考慮函數

$$g(u) = \frac{u^p}{p} + \frac{v^q}{q} - uv。$$

欲求 $g(u)$ 之絕對極小值。令 $g'(u) = 0$, 得

$$u^{p-1} - v = 0 \Rightarrow u = v^{1/(p-1)}。$$

2 第三章 多維隨機變數

又 $g''(u) = (p-1)u^{p-2} > 0, \forall u > 0$, 故 $g(u)$ 在 $u = v^{1/(p-1)}$ 有唯一的相對極小值。且極小值為 (由 (8.1) 式得 $p/(p-1) = q$)

$$\begin{aligned} g(v^{1/(p-1)}) &= \frac{1}{p} v^{p/(p-1)} + \frac{1}{q} v^q - v^{1/(p-1)} v \\ &= \left(\frac{1}{p} + \frac{1}{q} - 1\right) v^q = 0。 \end{aligned}$$

由微積分中的結果知, 一函數的絕對極小值, 只能發生在該函數導數等於0的點、不可微的點, 或定義域的端點。而 $g(u)$ 定義在 $(0, \infty)$, 到處可微, 只有一導數為0的點, 且 $\lim_{u \rightarrow 0} g(u) = v^q/q > 0$, $\lim_{u \rightarrow \infty} g(u) = \infty$ 。故唯一的絕對極小值, 發生在 $u = v^{1/(p-1)}$, 且極小值為0。故 $g(u) \geq 0, \forall u \in R$, 且 $g(u) > 0, \forall u \neq v^{1/(p-1)}$ 。得證 (8.2) 式, 且等號只有在 $u = v^{1/(p-1)}$ 才成立。

利用上述引理即得下述不等式:

赫德爾不等式 (Hölder's inequality). 設 X, Y 為二隨機變數, p, q 滿足 (8.1) 式, 且設 $E(|X|^p)$ 及 $E(|X|^q)$ 皆存在。則

$$(8.3) \quad E(|XY|) \leq (E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}。$$

證明. 令

$$u = \frac{|X|}{(E(|X|^p))^{1/p}}, \quad v = \frac{|Y|}{(E(|Y|^q))^{1/q}},$$

則由引理 8.1, 得

$$\frac{1}{p} \frac{|X|^p}{E(|X|^p)} + \frac{1}{q} \frac{|Y|^q}{E(|Y|^q)} \geq \frac{|X||Y|}{(E|X|^p)^{1/p} (E|Y|^q)^{1/q}}。$$

將上式兩側取期望值, 則左側成為1, 再將右側的分母移至左側即得 (8.3) 式。證畢。

Hölder (1859-1937) 為德國數學家。赫德爾不等式之最著名的特例, 很可能是 $p = q = 2$ 的情況。此即在定理 5.4 中我們證過的**史瓦茲不等式**。

史瓦茲不等式. 對任二隨機變數 X, Y , 只要 $E(X^2), E(Y^2) < \infty$, 則

$$(8.4) \quad E(|XY|) \leq (E(X^2))^{1/2}(E(Y^2))^{1/2}。$$

4 第三章 多維隨機變數

赫德爾不等式尚有一些特例也常出現。例如，在(8.3)式中令 $Y \equiv 1$ ，得

$$(8.5) \quad E(|X|) \leq (E(|X|^p))^{1/p}, 1 < p < \infty。$$

在此對一隨機變數 Y ，及常數 c ， $Y \equiv c$ ，讀做“ Y 恆等於 c ”，其意義為 $P(Y = c) = 1$ ，即 Y 為一退化隨機變數，取值在 c 。

對 $1 < r < p$ ，若以 $|X|^r$ 取代(8.5)式中之 $|X|$ ，得

$$E(|X|^r) \leq (E(|X|^{pr}))^{1/p}。$$

次令 $s = rp$ (注意因 $p > 1$ ，故 $s > r$)，再將兩側開 r 次方，得

$$(8.6) \quad E(|X|^r)^{1/r} \leq (E(|X|^s))^{1/s}, 1 < r < s < \infty,$$

此即**黎波諾夫不等式**(Liapounov's inequality, Liapounov(1857-1918)，為俄國機率學家)。由黎波諾夫不等式知，若一隨機變數之高次動差存在，則較低次的動差亦存在。例如，若知 $E(|X|^4) = E(X^4)$ 存在，則知 $E(|X|)$ ， $E(|X|^2)$ 及 $E(|X|^3)$ 也皆存在，因此 $E(X)$ ， $E(X^2)$ 及 $E(X^3)$ 也皆存在。

底下的不等式精神上與赫德爾不等式類似，也可藉助赫德爾不等式得到。

明考夫斯基不等式(Minkowski's inequality)。設 X, Y 為二隨機變數，且 $1 \leq p < \infty$ 。則

$$(8.7) \quad (E(|X + Y|^p))^{1/p} \leq (E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p}。$$

證明。首先利用**三角形不等式**(Triangle inequality)，

$$|X + Y| \leq |X| + |Y|,$$

可得

$$\begin{aligned} E(|X + Y|^p) &= E(|X + Y||X + Y|^{p-1}) \\ &\leq E(|X||X + Y|^{p-1}) + E(|Y||X + Y|^{p-1})。 \end{aligned}$$

再利用赫德爾不等式於上述右側之每一期望值，得

$$E(|X + Y|^p) \leq E(|X|^p)^{1/p} (E(|X + Y|^{(p-1)q}))^{1/q} \\ + (E(|Y|^p))^{1/p} (E(|X + Y|^{(p-1)q}))^{1/q},$$

其中 q 滿足 $1/p + 1/q = 1$ 。上式兩側各除以 $(E(|X + Y|^{(p-1)q}))^{1/q}$ ，且因 $(p-1)q = p$ ， $1 - 1/q = 1/p$ ，即得(8.7)式。證畢。

Minkowski(1864-1909)亦為著名的俄國數學家。上述這些不等式，也有對應之數值的版本，即將期望值改為和。例如，對任意實數 $a_i, b_i, i = 1, \dots, n$,

$$(8.8) \quad \sum_{i=1}^n |a_i b_i| \leq \left(\sum_{i=1}^n |a_i|^p \right)^{1/p} \left(\sum_{i=1}^n |b_i|^q \right)^{1/q}, \quad \frac{1}{p} + \frac{1}{q} = 1, p, q > 0,$$

便對應赫德爾不等式。這只要在赫德爾不等式中，定義隨機變數 X, Y ，分別以 $1/n$ 的機率取值在 $(a_1, b_1), \dots, (a_n, b_n)$ ，便導致(8.8)式(習題第8題)。

明考夫斯基之數值的版本則為：對任意實數 $a_i, b_i, i = 1, \dots, n$ ，及 $p \geq 1$,

$$(8.9) \quad \left(\sum_{i=1}^n |a_i + b_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^n |a_i|^p \right)^{1/p} + \left(\sum_{i=1}^n |b_i|^p \right)^{1/p}。$$

(8.8)式一重要的特例是 $p = q = 2$ 的情況：

$$(8.10) \quad \left(\sum_{i=1}^n a_i b_i \right)^2 \leq \left(\sum_{i=1}^n |a_i b_i| \right)^2 \leq \left(\sum_{i=1}^n a_i^2 \right) \left(\sum_{i=1}^n b_i^2 \right)。$$

這是在數學中常見的關於數值之柯西-史瓦茲不等式。(8.10)式中若取 $b_i = 1, i = 1, \dots, n$ ，則得

$$(8.11) \quad \frac{1}{n} \left(\sum_{i=1}^n a_i \right)^2 \leq \frac{1}{n} \left(\sum_{i=1}^n |a_i| \right)^2 \leq \sum_{i=1}^n a_i^2。$$

有些不等式主要是利用實值函數的性質，而不是仰賴機率的性質。其中一個很有用的是間森不等式(Jensen's inequality)，適用於凸函數(convex function)。

定義8.1 函數 g 若滿足對 $\forall x, y \in R$ ，及 $0 < \lambda < 1$,

$$(8.12) \quad g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y),$$

便稱為**凸函數**。而若(8.12)式中之不等式反過來, g 便稱為**凹函數**(concave function)。可看出 g 為凹函數若且唯若 $-g$ 為凸函數。

(8.12)式成立, 表 $g(x)$ 之圖形由 u 至 v , 在 $(u, g(u)), (v, g(v))$ 連線的下方。至於凹函數則是函數圖形在前述連線段上方。例如, $g(x) = x^2$ 為凸函數, $g(x) = \log x$ 為凹函數。見圖8.1。

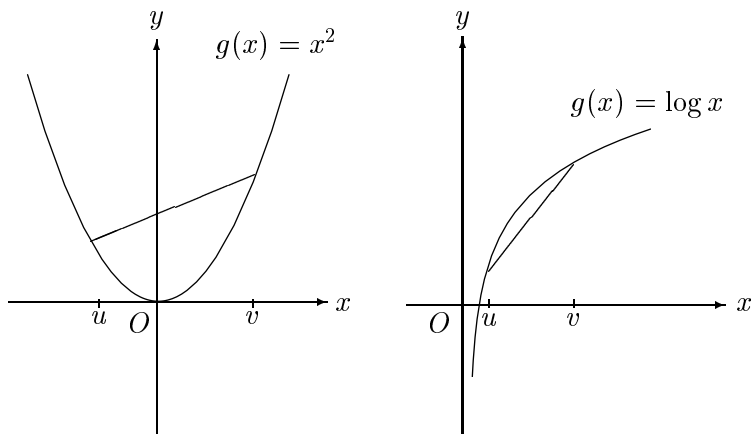


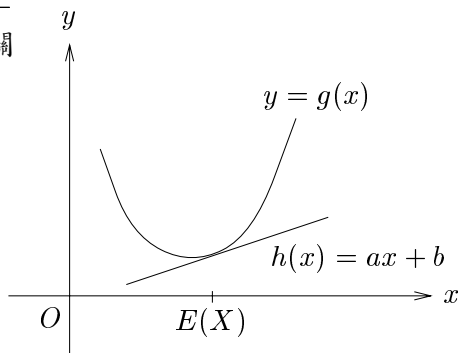
圖8.1 凸函數及凹函數之圖示

間森不等式. 設 $g(x)$ 為一凸函數, 且隨機變數 X 之 $E(X)$ 及 $E(g(X))$ 皆存在。則

$$(8.13) \quad E(g(X)) \geq g(E(X)).$$

證明. 考慮 $y = g(x)$ 之圖形, 令 $h(x) = ax + b$ 表圖形在 $(E(X), g(E(X)))$ 之切線。由於 $g(x)$ 為一凸函數, $g(x) \geq ax + b$ 。又因期望值會保持大小關係, 故

$$\begin{aligned} E(g(X)) &\geq E(ax + b) \\ &= aE(X) + b \\ &= h(E(X)) \\ &= g(E(X)), \end{aligned}$$



其中最後一等式成立乃是因直線 $y = h(x)$ 切圖形於 $(E(X), g(E(X)))$, 故 h 與 g 在 $E(X)$ 的值相等。

底下為間森不等式之一立即的應用(取 $g(x) = x^2$):

$$E(X^2) \geq (E(X))^2,$$

我們在1.6節便提過了。另外, 對 $x > 0$, 利用 $g(x) = 1/x$ 為凸函數, 則得

$$(8.14) \quad E(1/X) \geq 1/E(X)。$$

由微積分裡的知識得, 對一個二次可微的函數 g , 若 $g''(x) \geq 0, \forall x \in R$, 則 $g(x)$ 為凸函數; 而若 $g''(x) < 0, \forall x \in R$, $g(x)$ 便為凹函數。又對於凹函數 g , 因 $-g$ 為凸函數, 故

$$(8.15) \quad E(g(X)) \leq g(E(X))。$$

例8.1 設 $a_1, \dots, a_n$ 皆大於0, 令

$$\begin{aligned} M_A &= \frac{1}{n}(a_1 + \dots + a_n) = \frac{1}{n} \sum_{i=1}^n a_i, \\ M_G &= (a_1 \cdots a_n)^{1/n} = \left(\prod_{i=1}^n a_i \right)^{1/n}, \\ M_H &= \frac{1}{\frac{1}{n}(\frac{1}{a_1} + \dots + \frac{1}{a_n})} = \left(\frac{1}{n} \sum_{i=1}^n a_i^{-1} \right)^{-1}, \end{aligned}$$

分別表其**算術平均**(arithmetic mean)、**幾何平均**(geometric mean), 及**調和平均**(harmonic mean)。此三不等式之關係為

$$(8.16) \quad M_A \geq M_G \geq M_H。$$

底下我們看如何以間森不等式證明上述不等式。

設隨機變數 X 滿足 $P(X = a_i) = 1/n, i = 1, \dots, n$ 。因 $y = \log x$, $x > 0$, 為一凹函數, 間森不等式導致

$$E(\log X) \leq \log(E(X))。$$

故

$$\begin{aligned}\log M_G &= \frac{1}{n} \sum_{i=1}^n \log a_i = E(\log X) \leq \log(E(X)) \\ &= \log\left(\frac{1}{n} \sum_{i=1}^n a_i\right) = \log M_A.\end{aligned}$$

因此 $M_A \geq M_G$ 。其次因 $y = 1/x, x > 0$, 為凸函數, 故

$$\frac{1}{M_H} = \frac{1}{n} \sum_{i=1}^n \frac{1}{a_i} = E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)} = \frac{1}{\frac{1}{n} \sum_{i=1}^n a_i} = \frac{1}{M_A}.$$

因此 $M_A \geq M_H$ 。

至於 $M_G \geq M_H$, 可利用上述推導而得:

$$\frac{1}{M_H} = \frac{1}{n} \sum_{i=1}^n \frac{1}{a_i} \geq \left(\prod_{i=1}^n \frac{1}{a_i}\right)^{1/n} = \frac{1}{\left(\prod_{i=1}^n a_i\right)^{1/n}} = \frac{1}{M_G},$$

其中不等式, 我們是利用 $M_A \geq M_G$ 於正數 $\frac{1}{a_1}, \dots, \frac{1}{a_n}$ 。

柴比雪夫不等式(Chebyshev's inequality)可以說是機率裡一極重要且有用的不等式。此不等式的重要, 是因其適用性廣泛, 而非其精確性。Chebyshev(1821-1894, 有時寫成Chebychev, Shebychev, Tschebyshev等), 為十九世紀俄國極重要的一位機率學家。

柴比雪夫不等式. 設 X 為一隨機變數, $g(x)$ 為一非負函數, 且滿足 $E(g(X))$ 存在。則對任意 $c > 0$,

$$(8.17) \quad P(g(X) \geq c) \leq \frac{E(g(X))}{c}.$$

證明. 我們只證連續型的情況。設 X 為一連續型的隨機變數, 以 $f(x)$ 為其 p.d.f., 則

$$\begin{aligned}E(g(X)) &= \int_{-\infty}^{\infty} g(x)f(x)dx \\ &\geq \int_{\{x|g(x) \geq c\}} g(x)f(x)dx \\ &\geq c \int_{\{x|g(x) \geq c\}} f(x)dx \\ &= cP(g(X) \geq c).\end{aligned}$$

得證(8.17)式。

在柴比雪夫不等式中, 取 $g(x) = |x - \mu|^r$, 且以 c^r 取代 c , 其中 $\mu = E(X)$, $r > 0$, 則只要 $E(|X - \mu|^r)$ 存在, 便有

$$P(|X - \mu|^r \geq c^r) \leq \frac{E(|X - \mu|^r)}{c^r},$$

但 $P(|X - \mu|^r \geq c^r) = P(|X - \mu| \geq c)$ 。故得

$$(8.18) \quad P(|X - \mu| \geq c) \leq \frac{E|x - \mu|^r}{c^r}, r > 0。$$

由(8.18)式, 又得(取 $c = k\sigma$, $r = 2$, 其中 $\sigma^2 = \text{Var}(X)$)

$$(8.19) \quad P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, k > 0。$$

這是柴比雪夫不等式一極常出現的型式。可驗證(8.19)式與下式等價

$$(8.20) \quad P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \varepsilon > 0。$$

當然(8.19)及(8.20)式成立的先決條件是 X 的變異數要存在。(8.17)式是最一般的柴比雪夫不等式, 不過(8.18)、(8.19)及(8.20)三式也常被稱做柴比雪夫不等式。

例8.2 設 X 有 $\mathcal{P}(1)$ 分佈, 則 $E(X) = \text{Var}(X) = 1$ 。柴比雪夫不等式給出(利用(8.20)式)

$$P(|X - 1| \geq 2) \leq \frac{1}{4}。$$

另外, 經由實際計算, 可得

$$\begin{aligned} P(|X - 1| \geq 2) &= P(X \geq 3) = 1 - \sum_{i=0}^2 P(X = i) \\ &= 1 - \sum_{i=0}^2 \frac{e^{-1}}{i!} = 1 - e^{-1} \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{2} \right) \\ &= 1 - 2.5e^{-1} \doteq 0.080, \end{aligned}$$

的確是小於 $1/4 = 0.25$ ，但顯然誤差不小。換句話說，雖然 σ^2/ε^2 為 $P(|X - \mu| \geq \varepsilon)$ 之一上界，卻不一定很精確。但這還不是最差的情況。事實上(8.19)及(8.20)式之左側皆為機率值，不能超過1，但右側卻有可能大於1。此時不等式便未提供任何資訊。不過如前所述，柴比雪夫不等式並非因精確性而重要，而是因其普遍性。對於(8.19)及(8.20)二式，對於只要是變異數存在的隨機變數便皆成立。對一隨機變數 X ，(8.19)及(8.20)式，分別提供 X 會偏離期望值 k 個標準差及一固定距離 μ 之機率的一個上界。在1.6節我們說過，期望值像是隨機變數分佈的一個核心，隨機變數的可能值則散佈在期望值的左右。對不同的隨機變數，其可能值有時偏離期望值較遠，有時較近，但都會滿足(8.19)及(8.20)式。

(8.19)與(8.20)式分別等價於

$$(8.21) \quad P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}, k > 0,$$

$$(8.22) \quad P(|X - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}, \varepsilon > 0。$$

我們再以常態分佈來檢驗柴比雪夫不等式之精確性。

例8.3 設 $X \sim \mathcal{N}(\mu, \sigma^2)$ ， $Z \sim \mathcal{N}(0, 1)$ 。我們比較實際的機率及(8.21)式所給之機率下界。

$$P(|X - \mu| < \sigma) = P(|Z| < 1) \doteq 0.682, \quad 1 - \frac{1}{1^2} = 0,$$

$$P(|X - \mu| < 2\sigma) = P(|Z| < 2) \doteq 0.954, \quad 1 - \frac{1}{2^2} = 0.75,$$

$$P(|X - \mu| < 3\sigma) = P(|Z| < 3) \doteq 0.997, \quad 1 - \frac{1}{3^2} \doteq 0.889。$$