

第三章

多維隨機變數

3.6 二變數常態分佈

在機率及統計裡, 兩個變數以上之常態分佈很重要, 本節我們介紹二變數常態分佈。此分佈具有大部分多維常態分佈(multivariate normal distribution)的特性, 不但是一很重要的多維常態分佈, 且二變數之相關係數, 亦為其聯合p.d.f.之一參數。先給一定義。

定義6.1 設隨機向量 (X, Y) 為有參數 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 之二變數常態分佈, 以 $\mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ (或 $\mathcal{BVN}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$)表之, 其中 $\mu_1, \mu_2 \in R$, $\sigma_1, \sigma_2 > 0, -1 < \rho < 1$, 若其聯合p.d.f.為

$$(6.1) \quad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-q/2}, x, y \in R,$$

其中

$$(6.2) \quad q = \frac{1}{1-\rho^2} \left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{x-\mu_1}{\sigma_1} \frac{y-\mu_2}{\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right)。$$

2 第三章 多維隨機變數

雖然上述p.d.f.看起來是如此不可親, 不過此為一極常出現之二變數的分佈。此分佈並有不少好的性質, 我們列出一些如下。

(i) X, Y 之邊際分佈, 分別為 $\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)$ 。

(ii) $\rho(X, Y) = \rho$ 。

(iii) 對任二常數 $a, b \in R$, 且 $a^2 + b^2 \neq 0$, $aX + bY$ 有 $\mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)$ 分佈。

(iv) X 與 Y 獨立, 若且唯若 $\rho = 0$ 。

性質(iii)為二變數常態分佈之一特性。即只有二變數常態分佈才有性質(iii)。因此性質(iii)可刻劃(characterize)二變數常態分佈。

性質(iv)的證明可藉助定理2.1。由於 (X, Y) 之聯合p.d.f.如(6.1)式, 可看出要能寫成 $g(x)h(y)$ 的型式, 若且唯若 $\rho = 0$ 。性質(iv)是比較特別的, 因 ρ 為 X, Y 之相關係數, 而我們已數次強調, 二隨機變數之相關係數為0, 不表此二變數獨立。但對二變數之常態分佈, 其獨立與否, 就依 ρ 是否為0而定。可見 ρ 在此分佈中之重要角色。

底下我們來證明(6.1)式所定義之 $f(x, y)$ 確為一p.d.f.。首先 $f(x, y)$ 可改寫為

$$\begin{aligned} (6.3) \quad f(x, y) &= \frac{1}{\sigma_2 \sqrt{1 - \rho^2} \sqrt{2\pi}} e^{-\frac{1}{2\sigma_2^2(1-\rho^2)}(y - (\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1)))^2} \\ &\quad \cdot \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2}(x - \mu_1)^2} \\ &= A(x, y)B(x), \end{aligned}$$

這只要經過一些代數運算即可, 其證明留在習題中(第2題)。可看出在給定 x 之下, $A(x, y)$ 為 $\mathcal{N}(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1), \sigma_2^2(1 - \rho^2))$ 分佈之p.d.f., 而 $B(x)$ 為 $\mathcal{N}(\mu_1, \sigma_1^2)$ 分佈之p.d.f.。即

$$\int_{-\infty}^{\infty} A(x, y)dy = \int_{-\infty}^{\infty} B(x)dx = 1。$$

故

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dydx$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x, y) B(x) dy dx \\
&= \int_{-\infty}^{\infty} B(x) \left(\int_{-\infty}^{\infty} A(x, y) dy \right) dx \\
&= \int_{-\infty}^{\infty} B(x) dx = 1.
\end{aligned}$$

得證 $f(x, y)$ 爲一 p.d.f.。

在上述推導中, 因 X 之邊際 p.d.f. 爲

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

故 X 之邊際分佈爲 $\mathcal{N}(\mu_1, \sigma_1^2)$ 。同理可得 Y 之邊際分佈爲 $\mathcal{N}(\mu_2, \sigma_2^2)$ 。得證性質(i)。

由於 $Y|X = x$ 之 p.d.f. 爲

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{A(x, y)B(x)}{B(x)} = A(x, y).$$

即得二變數常態分佈之另一性質:

(v) $Y|X = x$ 及 $X|Y = y$ 等二條件分佈, 皆仍爲常態分佈, 且 $Y|X = x \sim \mathcal{N}(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1), \sigma_2^2(1 - \rho^2))$, $X|Y = y \sim \mathcal{N}(\mu_1 + \rho(\sigma_1/\sigma_2)(y - \mu_2), \sigma_1^2(1 - \rho^2))$ 。

由性質(v)再度看出, 若 X 與 Y 獨立, 則 ρ 須爲 0。又

$$(6.4) \quad E(Y|X = x) = \mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1),$$

$$(6.5) \quad \text{Var}(Y|X = x) = \sigma_2^2(1 - \rho^2).$$

此二式可與定理 5.5 相對應。當然亦可寫出 $E(X|Y = y)$ 及 $\text{Var}(X|Y = y)$ 。注意此處條件變異數 $\text{Var}(Y|X = x)$ 與 x 無關。由 (6.5) 式即得

$$E(\text{Var}(Y|X)) = \text{Var}(Y|X = x) = \sigma_2^2(1 - \rho^2).$$

故 (6.5) 式較 (5.23) 式的結果還強。又在統計學裡, $E(Y|X = x)$ 有時稱爲**迴歸函數**(regression function)。而 $E(Y|X = x)$ 之圖形則稱爲給定 $X = x$ 下

之迴歸曲線(regression curve)。(6.4)式指出, 對二變數常態分佈, 有線性迴歸函數。

我們常以(6.4)式右側, 做為給定 $X = x$, Y 之預測值。當 $\sigma_1 = \sigma_2$, 且 $0 < \rho < 1$, 則 Y 之預測值 $y = \mu_2 + \rho(x - \mu_1)$ 滿足

$$0 < y - \mu_2 < x - \mu_1, \quad x > \mu_1,$$

$$0 < \mu_2 - y < \mu_1 - x, \quad x < \mu_1。$$

即 Y 之預測值 y 有向 Y 之期望值 μ_2 靠近之傾向。此現象稱為迴歸效應(regression effect)。第一次考試考較好的學生, 第二次考試成績較可能下降; 第一次考試考較差的學生, 第二次考試成績則較可能上升。不過原先成績較高的, 其預測值仍然較高。又只有當 $\rho = 1$ 時, $y - \mu_2 = x - \mu_1$, 即 x, y 與期望值之差距不會改變。

這並不是太奇怪的事。考試考得好, 除了實力外, 總是運氣也不錯, 下一次考試不見得又有這麼好的運氣, 因此成績下降之可能性較大。考試考差的情況則反過來。

上述討論是在正相關之下。至於若負相關($\rho < 0$), 則第一次考高分, 第二次考低分就反而是較可能發生了。又上述討論雖是在 $\sigma_1 = \sigma_2$ 之下, 當 $0 < \rho < 1$, 即使 $\sigma_1 \neq \sigma_2$, 仍有

$$0 < \frac{y - \mu_2}{\sigma_2} < \frac{x - \mu_1}{\sigma_1}, \quad x > \mu_1,$$

$$0 < \frac{\mu_2 - y}{\sigma_2} < \frac{\mu_1 - x}{\sigma_1}, \quad x < \mu_1。$$

到底 x 或 y , 何者較接近期望值, 考慮相對差距(即各除以其標準差)是較合理的。

由於 $|\rho| < 1$, 除非 $\rho = 0$ (此時 X 與 Y 獨立), 否則

$$(6.6) \quad \text{Var}(Y|X = x) \leq \text{Var}(Y),$$

顯示知道 X 之值後, 對 Y 提供了一些資訊。由於對 Y 了解較多, 因此使變異數 $\text{Var}(Y|X = x)$ 減小。又當 $|\rho|$ 接近1時, 條件變異數 $\sigma_2^2(1 - \rho^2)$ 接近0。此

時條件分佈 $Y|X = x$ 集中在直線 $y = \mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1)$ 附近。此印證之前我們所說的, 當相關係數較接近1或-1時, X 與 Y 有較強的線性關係。

利用性質(v)及總機率法則, 可證明前述性質(ii):

$$\begin{aligned}
 E(XY) &= E(E(XY|X)) \\
 &= E(XE(Y|X)) \\
 &= E(X(\mu_2 + \rho(\sigma_2/\sigma_1)(X - \mu_1))) \\
 &= \mu_2 E(X) + \rho(\sigma_2/\sigma_1)E(X^2) - \rho\mu_1(\sigma_2/\sigma_1)E(X) \\
 &= \mu_1\mu_2 + \rho(\sigma_2/\sigma_1)(\sigma_1^2 + \mu_1^2) - \rho\mu_1^2(\sigma_2/\sigma_1) \\
 &= \mu_1\mu_2 + \rho\sigma_1\sigma_2,
 \end{aligned}$$

故

$$\begin{aligned}
 \text{Cov}(X, Y) &= E(XY) - \mu_1\mu_2 \\
 &= (\mu_1\mu_2 + \rho\sigma_1\sigma_2) - \mu_1\mu_2 \\
 &= \rho\sigma_1\sigma_2.
 \end{aligned}$$

由此即得 $\rho(X, Y) = \rho$ 。

(vi) (X, Y) 之聯合特徵函數為

$$\begin{aligned}
 (6.7) \quad \phi(s, t) &= E(e^{isX+itY}) \\
 &= \exp\{i\mu_1s + i\mu_2t - (\sigma_1^2s^2 + \sigma_2^2t^2 + 2\sigma_1\sigma_2\rho st)/2\}, s, t \in R,
 \end{aligned}$$

動差母函數則為

$$\begin{aligned}
 (6.8) \quad M(s, t) &= E(e^{sX+tY}) \\
 &= \exp\{\mu_1s + \mu_2t + (\sigma_1^2s^2 + \sigma_2^2t^2 + 2\sigma_1\sigma_2\rho st)/2\}, s, t \in R.
 \end{aligned}$$

(6.7)式與(6.8)式等價。底下我們利用性質(v)來證明(6.8)式:

首先由總機率法則,

$$\begin{aligned}
 M(s, t) &= E(e^{sX+tY}) \\
 &= E(E(e^{sX+tY}|X)) \\
 &= E(e^{sX}E(e^{tY}|X)).
 \end{aligned}$$

6 第三章 多維隨機變數

而因 $Y|X = x \sim \mathcal{N}(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1), \sigma_2^2(1 - \rho^2))$, 利用第二章(3.25)式, 得

$$E(e^{tY}|X = x) = \exp\{(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1))t + \sigma_2^2(1 - \rho^2)t^2/2\}。$$

故

$$\begin{aligned} M(s, t) &= E(\exp\{sX + (\mu_2 + \rho(\sigma_2/\sigma_1)(X - \mu_1))t + \sigma_2^2(1 - \rho^2)t^2/2\}) \\ &= \exp\{\mu_2 t - \rho(\sigma_2/\sigma_1)\mu_1 t + \sigma_2^2(1 - \rho^2)t^2/2\} E(\exp\{(s + \rho(\sigma_2/\sigma_1)t)X\}) \\ &= \exp\{\mu_2 t - \rho(\sigma_2/\sigma_1)\mu_1 t + \sigma_2^2(1 - \rho^2)t^2/2\} \\ &\quad \cdot \exp\{\mu_1(s + \rho(\sigma_2/\sigma_1)t) + \sigma_1^2(s + \rho(\sigma_2/\sigma_1)t)^2/2\} \\ &= \exp\{\mu_1 s + \mu_2 t + (\sigma_1^2 s^2 + \sigma_2^2 t^2 + 2\sigma_1\sigma_2\rho st)/2\}。 \end{aligned}$$

證畢。

由性質(vi)可得性質(ii): $aX + bY$ 之動差母函數為

$$\begin{aligned} &E(e^{t(aX+bY)}) \\ &= E(e^{atX+btY}) \\ &= \exp\{\mu_1 at + \mu_2 bt + (\sigma_1^2 a^2 t^2 + \sigma_2^2 b^2 t^2 + 2\sigma_1\sigma_2\rho abt^2)/2\} \\ &= \exp\{(a\mu_1 + b\mu_2)t + (a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)t^2/2\}。 \end{aligned}$$

與第二章(3.25)式比較, 即得 $aX + bY \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)$ 。

利用(6.8)式, 及二變數之聯合動差母函數可唯一決定其聯合分佈, 即得下述性質, 其證明留在習題中(第8題)。此亦為性質(iii)之推廣。

(vii) 令 $Z_1 = a_1X + b_1Y$, $Z_2 = a_2X + b_2Y$, 且 $a_1b_2 - a_2b_1 \neq 0$, 則 (Z_1, Z_2) 有 $\mathcal{N}(\eta_1, \eta_2, \tau_1^2, \tau_2^2, \gamma)$ 分佈, 其中

$$\begin{aligned} (6.9) \quad \eta_1 &= a_1\mu_1 + b_1\mu_2, \quad \eta_2 = a_2\mu_1 + b_2\mu_2, \\ \tau_1^2 &= a_1^2\sigma_1^2 + b_1^2\sigma_2^2 + 2a_1b_1\rho\sigma_1\sigma_2, \quad \tau_2^2 = a_2^2\sigma_1^2 + b_2^2\sigma_2^2 + 2a_2b_2\rho\sigma_1\sigma_2, \\ \gamma &= (a_1a_2\sigma_1^2 + b_1b_2\sigma_2^2 + (a_1b_2 + a_2b_1)\rho\sigma_1\sigma_2)/(\tau_1\tau_2)。 \end{aligned}$$

由此也可看出 Z_1 與 Z_2 獨立, 若且唯若

$$(6.10) \quad a_1 a_2 \sigma_1^2 + b_1 b_2 \sigma_2^2 + (a_1 b_2 + a_2 b_1) \rho \sigma_1 \sigma_2 = 0。$$

又若 X 與 Y 獨立, 即 $\rho = 0$, 則上述 τ_1, τ_2, γ 可簡化不少。

利用定理5.5, 此處檢驗 γ 確如(6.9)式中所給。首先

$$\begin{aligned} \text{Cov}(Z_1, Z_2) &= \text{Cov}(a_1 X + b_1 Y, a_2 X + b_2 Y) \\ &= a_1 a_2 \text{Var}(X) + b_1 b_2 \text{Var}(Y) + (a_1 b_2 + a_2 b_1) \text{Cov}(X, Y) \\ &= a_1 a_2 \sigma_1^2 + b_1 b_2 \sigma_2^2 + (a_1 b_2 + a_2 b_1) \rho \sigma_1 \sigma_2。 \end{aligned}$$

而

$$\rho(Z_1, Z_2) = \frac{\text{Cov}(Z_1, Z_2)}{\sqrt{\text{Var}(Z_1)}\sqrt{\text{Var}(Z_2)}} = \frac{\text{Cov}(Z_1, Z_2)}{\tau_1 \tau_2}。$$

故 $\gamma = \rho(Z_1, Z_2)$ 確如(6.9)式中所給。

雖然二變數常態分佈, 其邊際分佈必皆亦為常態分佈, 但其逆不真。即若 X, Y 之邊際分佈皆有常態分佈, 並不一定導致 (X, Y) 有二變數常態分佈。當然此時這兩個變數必不獨立。原因是兩獨立的常態分佈, 其聯合 p.d.f. 為邊際 p.d.f. 之乘積, 故有(6.1)式的型式(且 $\rho = 0$), 見習題第6題。

例如, 設隨機變數 X, Y 之聯合 p.d.f. 為

$$(6.11) \quad f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \left(1 - \frac{xy}{(1+x^2)(1+y^2)}\right), x, y \in R。$$

則 (X, Y) 並無二變數常態分佈, 但可證明 X, Y 之邊際分佈皆為常態(習題第14題)。由(6.11)式, 讀者不難寫出許多邊際分佈為常態, 但並非二變數常態分佈之聯合 p.d.f.。

Anderson(1984)曾以幾何的方法給下述著名的例子。設 (X, Y) 有二變數常態分佈, 且期望值皆為0。在圖6.1中, 有四個正方形, A與D, C與B分別對稱於 y 軸; A與C, D與B分別對稱於 x 軸。將A中之機率移至C中, B中之機率移至D中, 則得到的兩個變數 U, V , 其邊際分佈仍分別與 X, Y 相同, 也就是皆有常態分佈。但 (U, V) 之聯合分佈不為二變數常態分佈。

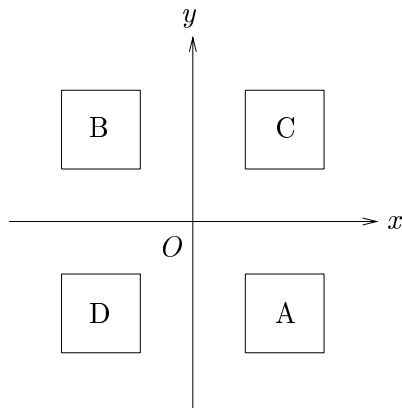


圖6.1 非二變數之常態分佈但邊際分佈為常態

另外，一般而言，若 X, Y 為二無相關之隨機變數，且皆有常態分佈，則 X 與 Y 並不一定獨立。由性質(iv)知，此時 (X, Y) 必非二變數之常態分佈。下例為Broffitt(1986)所給。

例6.1 設隨機變數 X 有 $\mathcal{N}(0, 1)$ 分佈， Z 與 X 獨立，且 $P(Z = 1) = P(Z = -1) = 1/2$ ，又令 $Y = ZX$ 。由

$$\begin{aligned} P(Y \leq y | Z = 1) &= P(ZX \leq y | Z = 1) = P(X \leq y | Z = 1) \\ &= P(X \leq y), \\ P(Y \leq y | Z = -1) &= P(ZX \leq y | Z = -1) = P(-X \leq y | Z = -1) \\ &= P(-X \leq y) = P(X \leq y), \end{aligned}$$

注意 $X \stackrel{d}{=} (-X)$ ，得

$$P(Y \leq y | Z = 1) = P(Y \leq y | Z = -1), y \in R。$$

故 Y 與 Z 獨立(此亦為一兩個變數隨機的獨立，但不是函數的獨立(functionally independent)之例)。因此

$$P(Y \leq y) = P(Y \leq y | Z = 1) = P(X \leq y),$$

即 Y 亦有 $\mathcal{N}(0, 1)$ 分佈。現因 $E(X) = E(Y) = E(Z) = 0$ ，且

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(ZX^2) = E(Z)E(X^2) = 0,$$

故 X 與 Y 無相關，但明顯地 X 與 Y 並不獨立。因此 (X, Y) 無二變數常態分佈，雖然其邊際分佈皆為常態。

在一維的情況，若 $X \sim \mathcal{N}(\mu, \sigma^2)$ ，令 $Z = (X - \mu)/\sigma$ ，則 $Z \sim \mathcal{N}(0, 1)$ 。因此只需一個標準常態分佈之機率值表，便能求出任一常態分佈之機率值。在二維的情況，設 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 。令 $Z_1 = (X - \mu_1)/\sigma_1$ ， $Z_2 = (Y - \mu_2)/\sigma_2$ ，則 $(Z_1, Z_2) \sim \mathcal{N}(0, 0, 1, 1, \rho)$ (習題第12題)。 Z_1, Z_2 皆有標準常態分佈，而 Z_1, Z_2 之相關係數不變仍為 ρ 。因此

$$\begin{aligned} F_{X,Y}(a, b) &= P(X \leq a, Y \leq b) \\ &= P\left(\frac{X - \mu_1}{\sigma_1} \leq \frac{a - \mu_1}{\sigma_1}, \frac{Y - \mu_2}{\sigma_2} \leq \frac{b - \mu_2}{\sigma_2}\right) \\ &= F_{Z_1, Z_2}\left(\frac{a - \mu_1}{\sigma_1}, \frac{b - \mu_2}{\sigma_2}\right). \end{aligned}$$

任一二變數常態之隨機變數 X, Y ，經由

$$Z_1 = \frac{X - \mu_1}{\sigma_1}, \quad Z_2 = \frac{Y - \mu_2}{\sigma_2}$$

之變數代換，可轉為標準二變數常態隨機變數 Z_1, Z_2 。雖然只有一個一維的標準常態分佈，卻有無限多個標準二變數常態分佈，每一相關係數 ρ 便有一個。所以二變數標準常態分佈之機率值表是較麻煩的，因 $F_{Z_1, Z_2}(h, k)$ 有三個變數： h, k, ρ 。本書並不給出這種表。但對有些特殊情況，可以求出有關二變數之機率。見下二例。

例6.2 某地區男性的身高可以 $\mathcal{N}(170, 7.2^2)$ 當做其分佈，女性身高可以 $\mathcal{N}(165, 5.4^2)$ 當做其分佈。男性與女性的身高假設為相互獨立。任選一男性及一女性。試求女性身高高於男性之機率。

解. 設 $X \sim \mathcal{N}(170, 7.2^2)$ ， $Y \sim \mathcal{N}(165, 5.4^2)$ ， X 與 Y 獨立。本例即要求 $P(Y > X)$ 。由定理2.4-(v)，得 $Y - X \sim \mathcal{N}(-5, 9^2)$ 。故

$$P(Y > X) = P(Y - X > 0)$$

$$\begin{aligned}
&= P\left(\frac{Y - X - (-5)}{9} > \frac{0 - (-5)}{9}\right) \\
&= P\left(Z > \frac{5}{9}\right) \doteq 1 - P(Z \leq 0.556) \\
&\doteq 1 - 0.7109 = 0.2891,
\end{aligned}$$

其中 $Z \sim \mathcal{N}(0, 1)$ 。

例6.3 承上例。設 $\rho(X, Y) = 0.6$ 。例如 X 可表一家庭中男孩之身高, Y 表女孩之身高, 通常 X 與 Y 不獨立。試求 $P(Y > X)$ 。

解. 由性質 (iii), $Y - X \sim \mathcal{N}(-5, 7.2^2 + 5.4^2 - 2 \cdot 0.6 \cdot 7.2 \cdot 5.4) = \mathcal{N}(-5, 34.344)$ 。因此

$$\begin{aligned}
P(Y > X) &= P(Y - X > 0) \\
&= P\left(\frac{Y - X - (-5)}{\sqrt{34.344}} > \frac{0 - (-5)}{\sqrt{34.344}}\right) \\
&\doteq P(Z > 0.853) \doteq 1 - P(Z \leq 0.853) \\
&\doteq 1 - 0.8031 = 0.1969。
\end{aligned}$$

一般對二變數常態分佈, 若涉及的本質上只是一維的常態分佈, 如 $aX + bY$, 便可求出其機率值。例6.3中所求即為關於一維變數 $-X + Y$ 之機率。