Matriz de Variâncias e Covariâncias Amostral

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & & s_{2p} \\ \vdots & & & \\ s_{p1} & s_{p2} & & s_{pp} \end{bmatrix} \qquad s_{jk} = \frac{1}{n}\sum_{i=1}^{n}(X_{ij}-\bar{X}_j)(X_{ik}-\bar{X}_k)$$

$j, k = 1, 2 \dots p$

s_{jk} - Covariância amostral entre as variáveis X_j e X_k.

Ex: $n=5 \quad p=2$

$$\underset{\sim}{X}_1 = \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix} \quad \underset{\sim}{X}_2 = \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix} \quad \underset{\sim}{X}_3 = \begin{bmatrix} x_{31} \\ x_{32} \end{bmatrix} \quad \underset{\sim}{X}_4 = \begin{bmatrix} x_{41} \\ x_{42} \end{bmatrix} \quad \underset{\sim}{X}_5 = \begin{bmatrix} x_{51} \\ x_{52} \end{bmatrix}$$

$$\bar{X} = \begin{bmatrix} \frac{1}{5}\sum_{i=1}^{5} X_{i1} = \bar{X}_1 \\ \frac{1}{5}\sum_{i=1}^{5} X_{i2} = \bar{X}_2 \end{bmatrix}$$

$$s_{11} = \frac{1}{5}\sum_{i=1}^{5}(X_{i1}-\bar{X}_1)(X_{i1}-\bar{X}_1) = \frac{1}{5}\sum_{i=1}^{5}(X_{i1}-\bar{X}_1)^2 \rightarrow$$ Variância amostral de X_1

$$s_{12} = \frac{1}{5}\sum_{i=1}^{5}(X_{i1}-\bar{X}_1)(X_{i2}-\bar{X}_2) \rightarrow$$ Covariância amostral entre X_1 e X_2.

S pode ser escrita como

$$\frac{1}{n}\sum_{i=1}^{n}(\underset{\sim}{X}_i - \bar{X})(\underset{\sim}{X}_i - \bar{X})' \quad \text{pois}$$

$$\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} X_{i1}-\bar{X}_1 \\ X_{i2}-\bar{X}_2 \\ \vdots \\ X_{ip}-\bar{X}_p \end{bmatrix}\begin{bmatrix} X_{i1}-\bar{X}_1 & X_{i2}-\bar{X}_2 & \cdots & X_{ip}-\bar{X}_p \end{bmatrix} =$$

$$\frac{1}{n}\begin{bmatrix} \sum(X_{i1}-\bar{X}_1)(X_{i1}-\bar{X}_1) & \sum(X_{i1}-\bar{X}_1)(X_{i2}-\bar{X}_2) \cdots & \sum(X_{i1}-\bar{X}_1)(X_{ip}-\bar{X}_p) \\ & & \\ & & \end{bmatrix}_{p\times p}$$

$$= \begin{bmatrix} s_{11} & s_{12} \cdots & s_{1p} \\ \vdots & & \\ s_{p1} & s_{p2} \cdots & s_{pp} \end{bmatrix} = S.$$

Outra forma: $S = \frac{1}{n}\left[\sum_{i=1}^{n}\underset{\sim}{X}_i\underset{\sim}{X}_i' - n\bar{X}\bar{X}'\right]$

demonstrado na aula passada.

Resultado 1

Y vetor aleatório com $E(Y) = \mu_Y$ e $Var(Y) = \Sigma_Y$.

Nestas condições, $E(YY') = \Sigma_Y + \mu_Y \mu_Y'$.

Resultado 2

Seja $\underset{\sim}{X}_1, \underset{\sim}{X}_2 \ldots \underset{\sim}{X}_n$ uma amostra aleatória da densidade $f(\underset{\sim}{x}) = f(x_1, x_2 \ldots x_p)$ com vetor de médias μ e matriz de covariância Σ. Nestas condições,

a) $\bar{X} \doteq \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}$ é um estimador não viciado de μ.

b) $Var(\bar{X}) = \frac{1}{n} \Sigma$.

c) $E(S) = \frac{n-1}{n} \Sigma$.

Correspondentes unidimensionais

$E(\bar{X}) = \mu$.

$Var(\bar{X}) = \frac{\sigma^2}{n}$

$E\left(\frac{\Sigma(X_i - \bar{X})^2}{n-1}\right) = \sigma^2$

↳ não viciado p/ σ^2

$E\left(\frac{\Sigma(X_i - \bar{X})^2}{n}\right) = \frac{n-1}{n} \sigma^2$.

Resultado 3.1

Johnson e Wichern

pag 121

Estudar a demonstração lá.

S, matriz de variância e covariância amostral, é um estimador viciado de Σ.

Um estimador não viciado de $Var(X) = \Sigma$ é

$$S^* = \frac{1}{n-1} \sum_{i=1}^{n} (\underset{\sim}{X}_i - \bar{X})(\underset{\sim}{X}_i - \bar{X})' = \frac{nS}{n-1}.$$

Variância Generalizada

X - matriz de dados

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & & x_{np} \end{bmatrix}$$

n pontos no espaço p-dimensional

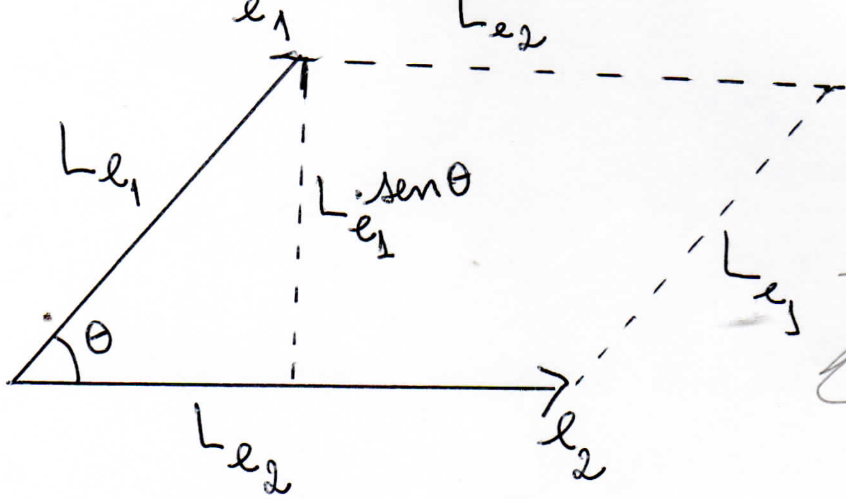
$$S^* = (s^*_{jk}) \qquad s^*_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$$

$$j, k = 1, 2, \dots p$$

Para resumir a variabilidade expressa em S^* através de um único número utiliza-se

Variância Amostral Generalizada $= |S^*|$

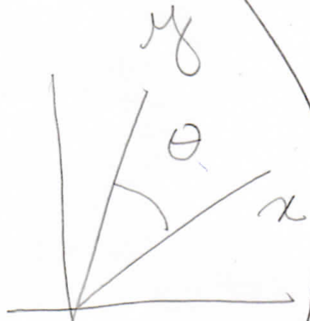
Para $p=2$ $\quad e_1 = \begin{bmatrix} x_{11} - \bar{x}_1 \\ x_{21} - \bar{x}_1 \\ \vdots \\ x_{n1} - \bar{x}_1 \end{bmatrix} \qquad e_2 = \begin{bmatrix} x_{12} - \bar{x}_2 \\ x_{22} - \bar{x}_2 \\ \vdots \\ x_{n2} - \bar{x}_2 \end{bmatrix}$



Para e_1 e_2 fixo

$$\frac{\sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sqrt{\sum (x_{i1} - \bar{x}_1)^2} \sqrt{\sum (x_{i2} - \bar{x}_2)^2}} = r(X_1, X_2)$$

$$\text{área} = L_{e_2} L_{e_1} \mathrm{sen}\,\theta = L_{e_1} L_{e_2} \sqrt{1 - \cos^2\theta}$$



$$L_{e_1} = \sqrt{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2} = \sqrt{(n-1)\, s_{11}}$$

$$L_{e_2} = \sqrt{\sum_{i=1}^{n} (x_{i2} - \bar{x}_2)^2} = \sqrt{(n-1)\, s_{22}}$$

$$\cos\theta = \frac{x'y}{\sqrt{y'y}\sqrt{x'x}}$$

$$\cos\theta = r_{12}$$

$$\text{área} = (n-1)\sqrt{s_{11}}\ \sqrt{s_{22}}\ \sqrt{1 - r_{12}^2}$$

$$|S^*| = \begin{vmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{vmatrix} = \begin{vmatrix} s_{11} & \sqrt{s_{11}}\sqrt{s_{22}}\, r_{12} \\ \sqrt{s_{11}}\sqrt{s_{22}}\, r_{12} & s_{22} \end{vmatrix}$$

$$= s_{11} s_{22} - s_{11} s_{22} r_{12}^2 = s_{11} s_{22} (1 - r_{12}^2)$$

$$\therefore \; |S^*| = \frac{\text{Área}^2}{(n-1)^2}$$

De modo geral $|S^*| = \dfrac{(\text{volume})^2}{(n-1)^p}$

Volume do sólido gerado pelos p vetores de desvios

Variância generalizada é zero $\Longleftrightarrow$ pelo menos um dos vetores de desvios é combinação linear dos demais

Ex: $X = \begin{bmatrix} 1 & 2 & 5 \\ 4 & 1 & 6 \\ 4 & 0 & 4 \end{bmatrix}$ $\quad \bar{X} = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$

$\quad\quad\quad X_1 \quad X_2 \quad X_3$ $\quad\quad n = 3 \quad p = 3$

$$e_1 = \begin{bmatrix} 1-3 \\ 4-3 \\ 4-3 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \quad e_2 = \begin{bmatrix} 2-1 \\ 1-1 \\ 0-1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad e_3 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

Observar que $e_3 = e_1 + 2e_2 \Rightarrow \text{Volume} = 0$

$$X_3 = X_1 + 2X_2$$

Cálculo de S^*

$$s^*_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$$

ou $\quad e'_j e_k = (n-1)\, s^*_{jk}$

$$e_1' e_1 = (-2)^2 + 1^2 + 1^2 = 6 = 2 s_{11}^* \Rightarrow s_{11}^* = 3$$

$$e_1' e_2 = -2 . 1 + 1 . 0 + 1(-1) = -3 = 2 s_{12}^*$$

$$\Rightarrow s_{12}^* = -1,5$$

$$S^* = \begin{bmatrix} 3 & -1,5 & 0 \\ -1,5 & 1 & 0,5 \\ 0 & 0,5 & 1 \end{bmatrix}$$

$$|S^*| = 3 \begin{vmatrix} 1 & 0,5 \\ 0,5 & 1 \end{vmatrix} + 1,5 \begin{vmatrix} -1,5 & 0,5 \\ 0 & 1 \end{vmatrix}$$

$$= 3\left(1 - \frac{1}{4}\right) + \frac{3}{2}\left(-\frac{3}{2} - 0\right) = 0$$

Obs:

1 - O volume aumenta com o comprimento de $\forall e_j$, $j = 1, 2, .., p$ e com a "ortogonalidade" entre eles (ver figura).

2 - Se for utilizada a matriz S, a variância Generalizada é $|S| = \text{volume}^2 / n^p$.

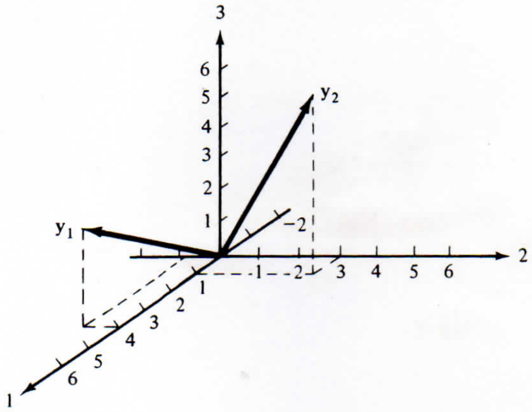


Figure 3.2 A plot of the data matrix $\mathbf{X}$ as $p = 2$ vectors in $n = 3$ space.

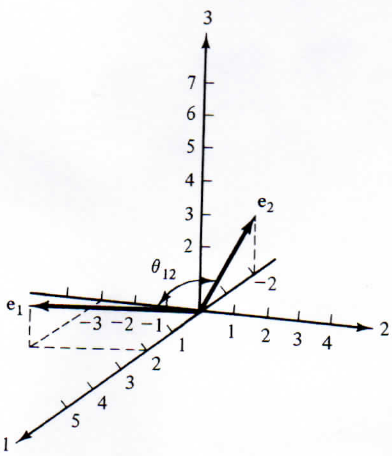


Figure 3.5 The vectors $\mathbf{e}_1$ and $\mathbf{e}_2$.

5- Distribuição Normal Multivariada

A v.a. unidimensional X tem distribuição normal se sua função densidade de probabilidades é

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

$$\sigma > 0 \qquad E(X) = \mu \qquad Var(X) = \sigma^2 \qquad X \sim N(\mu, \sigma^2)$$

Se X_j, $j = 1, 2 \dots p$ são v.a. independentes, $X_j \sim N(\mu_j, \sigma_j^2)$ então a função densidade conjunta de $X_1, X_2 \dots X_p$ é

$$f(x_1, x_2 \dots x_p) = f(x_1) f(x_2) \dots f(x_p) =$$

$$= \frac{1}{(2\pi)^{p/2} \sigma_1 \dots \sigma_p} \exp\left[-\frac{1}{2} \sum_{j=1}^{p} \left(\frac{x_j - \mu_j}{\sigma_j} \right)^2 \right]$$

Escrevendo

$$x' = [x_1, x_2 \dots x_p]$$

$$\mu' = [\mu_1, \mu_2 \dots \mu_p]$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & & 0 \\ 0 & \sigma_2^2 & & 0 \\ & & \ddots & \\ 0 & 0 & & \sigma_p^2 \end{bmatrix}$$

, $f(x_1, x_2 \dots x_p)$ pode ser escrita como

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu) \Sigma (x-\mu)\right\}$$

$$|\Sigma| = \sigma_1^2 \sigma_2^2 \ldots \sigma_p^2$$

$$(x-\mu)' \Sigma^{-1} (x-\mu) = [x_1-\mu_1, x_2-\mu_2 \ldots x_p-\mu_p] \begin{bmatrix} 1/\sigma_1^2 & & 0 \\ & 1/\sigma_2^2 & \\ 0 & & \ddots 1/\sigma_p^2 \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ \vdots \\ x_p-\mu_p \end{bmatrix}$$

$$= \left[\frac{x_1-\mu_1}{\sigma_1^2} \quad \frac{x_2-\mu_2}{\sigma_2^2} \quad \ldots \quad \frac{x_p-\mu_p}{\sigma_p^2}\right] \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ \vdots \\ x_p-\mu_p \end{bmatrix} = \frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} +$$

$$\ldots + \frac{(x_p-\mu_p)^2}{\sigma_p^2} = \sum_{j=1}^{p} \left(\frac{x_j-\mu_j}{\sigma_j}\right)^2.$$

A generalização dessa densidade é o caso em que Σ é qualquer matriz simétrica, positiva definida $p \times p$. A densidade de $X' = [X_1 \; X_2 \ldots X_p]$ fica

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)\right\}.$$

Verifica-se que

-) $f(x)$ é positiva para todo $x' = (x_1, x_2 \ldots x_p)$.

e $\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x)\, dx_1 \ldots dx_p = 1.$

ii) $\mu' = [\mu_1 \ \mu_2 \ldots \mu_p]$ e $E(X_j) = \mu_j$, $j = 1, 2, \ldots p$

iii) $\Sigma = Var(X)$ é a matriz de variância e covariânc[ia]
de X

iv) Se $\sigma_{ij} = 0$ para todo $i \neq j$, $X_1, X_2 \ldots X_p$ são variáve[is]
aleatórias independentes.

v) Notação: $X \sim N_p(\mu, \Sigma)$

Caso Particular $p = 2$

Distribuição Normal Bivariada

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

$$\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix},$$

Seja $\rho_{12} = Corr(X_1, X_2) \Rightarrow \sigma_{12} = \rho_{12} \sqrt{\sigma_{11}} \sqrt{\sigma_{22}}$

$$(x-\mu)' \Sigma^{-1} (x-\mu) = [x_1 - \mu_1 \quad x_2 - \mu_2] \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}$$

$$\times \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} =$$

Obs: $\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x & y \\ y & w \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} =$

$$= \begin{bmatrix} ax + by & ay + bw \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 x + aby + aby + b \ldots$$
$$= a^2 x + 2aby + b^2 \ldots$$

$\therefore$

$$(x-\mu)' \Sigma^{-1} (x-\mu) =$$

$$\frac{\sigma_{22}(x_1-\mu_1)^2 + \sigma_{11}(x_2-\mu_2)^2 - 2\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1-\mu_1)(x_2-\mu \ldots}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}$$

$$= \frac{1}{1-\rho_{12}^2}\Bigg[\underbrace{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2}_{X_1 \text{ e } X_2 \text{ padronizadas}} - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right) \ldots$$

Substituindo em $f(x)$

$$f(x) = \frac{1}{2\pi\sqrt{\underbrace{\sigma_{11}\sigma_{22} - \rho_{12}^2\sigma_{11}\sigma_{22}}_{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}}} \exp\left\{ -\frac{1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \right.\right.$$

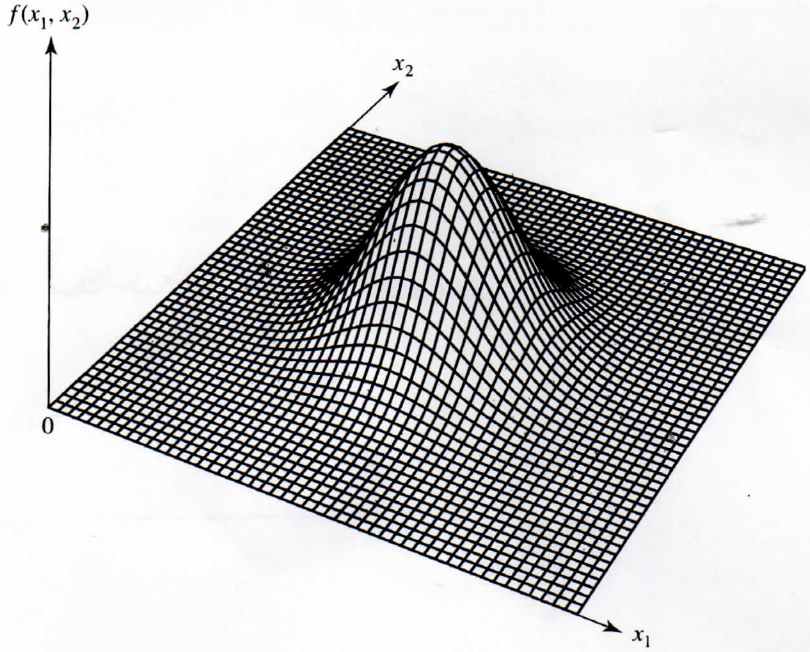
$$\left.\left.\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}$$

Se $\rho_{12} = 0$ $(\Leftrightarrow \sigma_{12} = 0)$

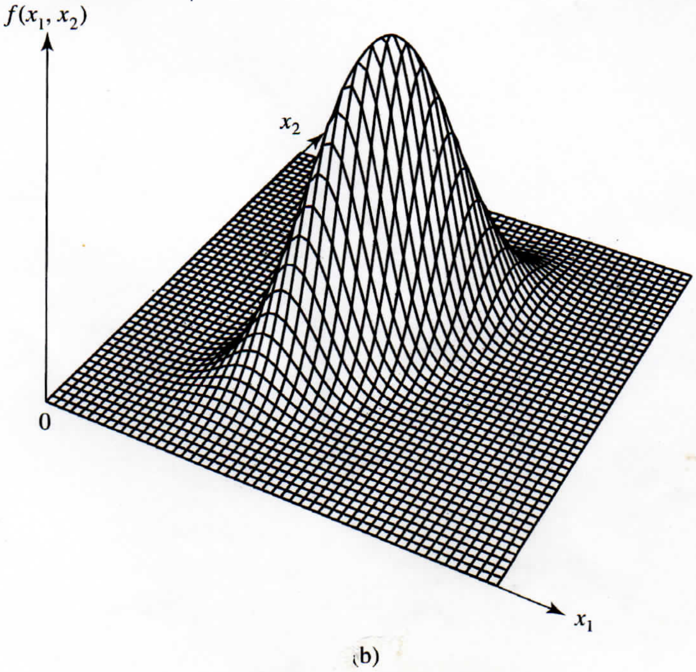
$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}}} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]\right\}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}}} e^{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2} \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{22}}} e^{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2}$$

$$= f(x_1)\, f(x_2) \qquad X_1 \sim N(\mu_1, \sigma_{11}) \quad X_2 \sim N(\mu_2, \sigma_{22})$$



(a)



(b)

Figure 4.2 Two bivariate normal distributions. (a) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = 0$. (b) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = .75$.