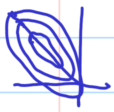
- contour for the case $\sigma_{11} = \sigma_{22},\ \sigma_{12} \neq 0$
  - $0 = |\boldsymbol{\Sigma} - \lambda\mathbf{I}| = \begin{vmatrix} \sigma_{11}-\lambda & \sigma_{12} \\ \sigma_{12} & \sigma_{11}-\lambda \end{vmatrix} = (\sigma_{11}-\lambda)^2 - \sigma_{12}^2$
    $= (\lambda - \sigma_{11} - \sigma_{12})(\lambda - \sigma_{11} + \sigma_{12})$
  - the eigenvalues are $\lambda_1 = \sigma_{11} + \sigma_{12}$ and $\lambda_2 = \sigma_{11} - \sigma_{12}$
  - $\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{11} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = (\sigma_{11}+\sigma_{12}) \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \Rightarrow \begin{matrix} \cancel{\sigma_{11}e_1} + \sigma_{12}e_2 = (\cancel{\sigma_{11}} + \sigma_{12})e_1 \\ \sigma_{12}e_1 + \cancel{\sigma_{11}e_2} = (\cancel{\sigma_{11}} + \sigma_{12})e_2 \end{matrix}$ $e_1 = e_2$

Σ^{-1}:
eigenvectors: same as the eigenvector of Σ
eigenvalue: $\frac{1}{\lambda_1}, \frac{1}{\lambda_2}$

$\lambda_1 = \sigma_{11} + \sigma_{12}, \quad \mathbf{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

  - $\lambda_2 = \sigma_{11} - \sigma_{12}$ yields the eigenvector $\mathbf{e}_2' = [1/\sqrt{2}, -1/\sqrt{2}]$. $e_1 = -e_2$
  - When the covariance σ_{12} (or correlation ρ_{12}) is positive,
    $\underset{\lambda_1}{\sigma_{11}+\sigma_{12}} > \underset{\lambda_2}{\sigma_{11}-\sigma_{12}} \Rightarrow$ major axis: e_1, minor axis: e_2 } length ratio: major : minor $= \sqrt{\lambda_1} : \sqrt{\lambda_2}$ $= \left(\frac{1}{1/\lambda_1}\right)^{1/2} : \left(\frac{1}{1/\lambda_2}\right)^{1/2}$

    When the covariance (correlation) is negative,
    $\underset{\lambda_2}{\sigma_{11}-\sigma_{12}} > \underset{\lambda_1}{\sigma_{11}+\sigma_{12}} \Rightarrow$ major axis: e_2, minor axis: e_1 } length ratio major : minor $= \sqrt{\lambda_2} : \sqrt{\lambda_1}$



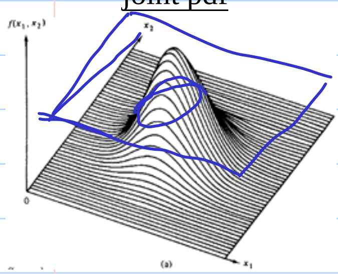
What if $|\rho_{12}|$ increase?

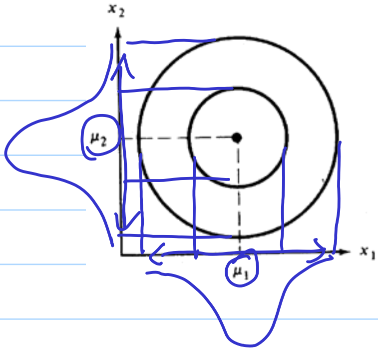
small ρ_{12} large ρ_{12}

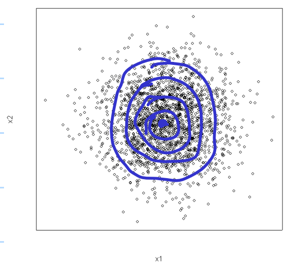
  - when $\rho_{12} = 0$
    $0 = |\Sigma - \lambda I| = (\lambda - \sigma_{11})^2$, eigenvalue: σ_{11}, σ_{11}
    any vector is eigenvector. ⇒ ellipse becomes circle

(a) $\sigma_1 = \sigma_2,\ \rho = 0 \Rightarrow$ independent and equal variance

joint pdf | contour lines of the pdf | data generated from the pdf

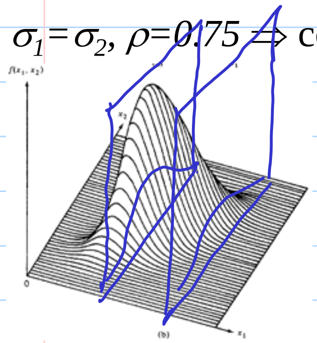


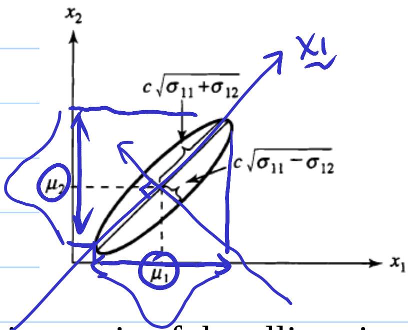


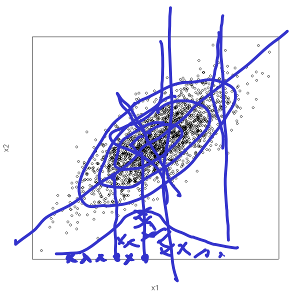


Q: what should the contour lines look like if $\sigma_1 \neq \sigma_2$?

(b) $\sigma_1 = \sigma_2,\ \rho = 0.75 \Rightarrow$ correlated and equal variance







when $\sigma_1 = \sigma_2,\ \rho \neq 0$, the major/minor axis of the ellipse is parallel to $x_1 = x_2$ or $x_1 = -x_2$

$\boxed{\text{contour of Normal pdf is an ellipse because it can be expressed as } (\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu}) = c}$

- random sample from a multivariate normal distribution

Let us assume that the $p \times 1$ vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ represent a random sample from a multivariate normal population with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Since $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are mutually independent and each has distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ parameter, the joint density function of all the observations is the product of the marginal normal densities:

$$f(x_1,\ldots,x_n|\mu,\Sigma) = \left\{ \begin{array}{c} \text{Joint density} \\ \text{of } \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n \end{array} \right\} = \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu})/2} \right\}$$

$$= \frac{1}{(2\pi)^{np/2}} \frac{1}{|\boldsymbol{\Sigma}|^{n/2}} e^{-\sum_{j=1}^{n} (\mathbf{x}_j - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu})/2}$$

$$= \mathcal{L}(\mu, \Sigma | x_1, \ldots, x_n) \Leftarrow \text{likelihood.}$$

joint pdf.

- **Q**: why normal?
  - While real data are never exactly multivariate normal, the normal density is often a useful approximation to the "true" population distribution
  - The multivariate normal density is mathematically tractable and "nice" results can be obtained
  - The distribution of many multivariate statistics are approximately normal, regardless of the form of the parent population because of a *central limit* effect

    although the dist. of $X_1, \ldots, X_n$ might not be multivariate normal, the stat. based on $X_1, \ldots, X_n$ could approximate the dist. based multivariate normal.

• Some properties of multivariate normal distribution

- **Result 4.3.** If $\mathbf{X}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the q linear combinations

$$\underset{(q\times p)(p\times 1)}{\mathbf{A}\ \mathbf{X}} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{bmatrix}$$

are distributed as $N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$. Also, $\underset{(p\times 1)}{\mathbf{X}} + \underset{(p\times 1)}{\mathbf{d}}$, where $\mathbf{d}$ is a vector of constants, is distributed as $N_p(\boldsymbol{\mu} + \mathbf{d}, \boldsymbol{\Sigma})$.

irrelevant to d.

Alternative definition of multivariate normal distribution: Let $z_1, \ldots, z_p$ be i.i.d from $N(0,1)$ and $Z = [z_1, \ldots, z_p]^T$. For a p-dim vector μ and a $p \times p$ symmetric, positive definite matrix Σ, X is said to have a multivariate normal distribution $N_p(\mu, \Sigma)$ if it has the same distribution as

$$\Sigma^{1/2} Z + \mu$$

any multivariate normal distribution can be generated from i.i.d. N(0,1).

*sketch of proof: of $X = \Sigma^{1/2} Z + \mu$,

$$AX + d = A(\Sigma^{1/2} Z + \mu) + d = \underline{(A\Sigma^{1/2})} Z + \underline{A\mu + d} \leftarrow \text{mean of } AX+d$$

$$\Sigma_{AX+d} = (A\Sigma^{1/2})(A\Sigma^{1/2})' = A\Sigma A'$$

- $\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu})$ has a $N_p(\mathbf{0}, \mathbf{I}_p)$ distribution

normalization $X \sim N(\mu, \sigma^2)$, $\frac{X-\mu}{\sigma} \sim N(0,1)$

$$\underbrace{(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})}_{\text{quadratic form}} = \left[(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-\frac{1}{2}}\right]\left[\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu})\right]$$

$$= \mathbf{Z}'\mathbf{Z} = Z_1^2 + Z_2^2 + \cdots + Z_p^2 \leftarrow z_1, \ldots, z_p \text{ iid } N(0,1)$$

➢ **Result 4.4.** All subsets of $\mathbf{X}$ are normally distributed. If we respectively partition $\mathbf{X}$, its mean vector $\boldsymbol{\mu}$, and its covariance matrix $\boldsymbol{\Sigma}$ as

$$\underset{(p\times 1)}{\mathbf{X}} = \begin{bmatrix} \underset{(q\times 1)}{\mathbf{X}_1} \\ \hdashline \underset{((p-q)\times 1)}{\mathbf{X}_2} \end{bmatrix} \qquad \underset{(p\times 1)}{\boldsymbol{\mu}} = \begin{bmatrix} \underset{(q\times 1)}{\boldsymbol{\mu}_1} \\ \hdashline \underset{((p-q)\times 1)}{\boldsymbol{\mu}_2} \end{bmatrix}$$

and

symmetric ... not symmetric in general

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \left[\begin{array}{c:c} \underset{(q\times q)}{\boldsymbol{\Sigma}_{11}} & \underset{(q\times(p-q))}{\boldsymbol{\Sigma}_{12}} \\ \hdashline \underset{((p-q)\times q)}{\boldsymbol{\Sigma}_{21}} & \underset{((p-q)\times(p-q))}{\boldsymbol{\Sigma}_{22}} \end{array}\right]$$

Symmetric ... symmetric

then $\mathbf{X}_1$ is distributed as $N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

* sketch of proof: $X = \Sigma^{1/2}Z + \mu$

$X_1 = [I|0]X = [I|0]\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = A(\Sigma^{1/2}Z+\mu) = (A\Sigma^{1/2})Z + A\mu$

$A \equiv \left[\begin{array}{ccc|ccc} 1 & & 0 & 0 & \cdots & 0 \\ 0 & & 1 & 0 & \cdots & 0 \end{array}\right]$

$A\mu = [I|0]\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mu_1$

$A\Sigma A' = [I|0]\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\begin{bmatrix} I \\ 0 \end{bmatrix} = \Sigma_{11}$.

➢ **Result 4.5.**

(a) If $\underset{(q_1\times 1)}{\mathbf{X}_1}$ and $\underset{(q_2\times 1)}{\mathbf{X}_2}$ are independent, then $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{0}$, a $q_1 \times q_2$ matrix of zeros.

(b) If $\begin{bmatrix} \mathbf{X}_1 \\ \hdashline \mathbf{X}_2 \end{bmatrix}$ is $N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{bmatrix}, \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$, then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

uncorrelated is equivalent to indep. for multivariate normal.

(c) If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent and are distributed as $N_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $N_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$, respectively, then $\begin{bmatrix} \mathbf{X}_1 \\ \hdashline \mathbf{X}_2 \end{bmatrix}$ has the multivariate normal distribution

$$N_{q_1+q_2}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{bmatrix}, \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \hdashline \mathbf{0}' & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$$

Textbook, p.202, problem 4.8

Note. Suppose that X_1 and X_2 are multivariate normal. $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ may not be a multivariate normal.

* sketch of proof for (b).

$M_X(\underset{\sim}{t}) = E e^{t_1 X_1 + \cdots + t_{q_1+q_2} X_{q_1+q_2}} \overset{\Sigma_{12}=0}{=\!=} M_{X_1}(\underset{\sim}{t_1}) \cdot M_{X_2}(\underset{\sim}{t_2})$

$(\underset{\sim}{t_1}, \underset{\sim}{t_2})$

➢ **Result 4.6.** Let $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \hdashline \mathbf{X}_2 \end{bmatrix}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{bmatrix}$, $\boldsymbol{\Sigma} = \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]$, and $|\boldsymbol{\Sigma}_{22}| > 0$. Then the conditional distribution of $\mathbf{X}_1$, given that $\mathbf{X}_2 = \mathbf{x}_2$, is normal and has

coefficient of X_2 in LM.

$[\beta_{ij}]$

$$\text{Mean} = \boldsymbol{\mu}_1 + \underset{\beta}{\underbrace{\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}}}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

and

$$\text{Covariance} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

Note that the covariance does not depend on the value $\mathbf{x}_2$ of the conditioning variable.

* sketch of proof:

Let $W = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$.

$$\begin{bmatrix} W \\ X_2 \end{bmatrix} = \underbrace{\begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix}}_{\overset{\shortparallel}{A}} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{bmatrix}, \Sigma^* \right)$$

$$\Sigma^* = \left[\begin{array}{c|c} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ \hline 0 & \Sigma_{22} \end{array} \right]$$

$\therefore$ W & X_2 are independent

$\Rightarrow f_{W,X_2}(w, x_2) = f_W(w) \cdot f_{X_2}(x_2)$

$$A^{-1} = \begin{bmatrix} I & \Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix} \Rightarrow \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = A^{-1} \begin{bmatrix} W \\ X_2 \end{bmatrix}$$

$$f_{X_1X_2}(x_1, x_2) = f_W(x_1 - \Sigma_{12}\Sigma_{22}^{-1}x_2) \cdot f_{X_2}(x_2) \cdot \underbrace{|J|}_{} = 1$$

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)} = \frac{f_W(x_1 - \Sigma_{12}\Sigma_{22}^{-1}x_2)\ \cancel{f_{X_2}(x_2)}}{\cancel{f_{X_2}(x_2)}}$$

in pdf of f_W: $(w - [\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2])$

$= x_1 - \Sigma_{12}\Sigma_{22}^{-1}x_2 - \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\mu_2$

mean of $x_1|x_2$ $= x_1 - \underline{[\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)]}$

cov of $x_1|x_2$

- Example: conditional density of a bivariate normal distribution
  - The conditional density of X_1, given that $X_2 = x_2$, is

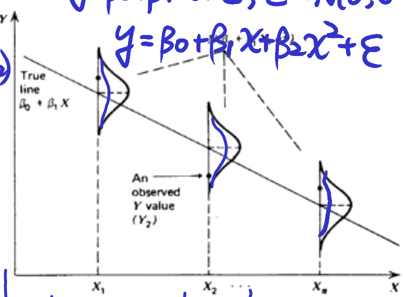
$$\frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)} = N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right) \text{ (exercise)}$$

$y \sim N(\beta_0 + \beta_1 x, \sigma^2)$

$y = \beta_0 + \beta_1 x + \varepsilon,\ \varepsilon \sim N(0, \sigma^2)$

$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$

$f(x_1 x_2)$



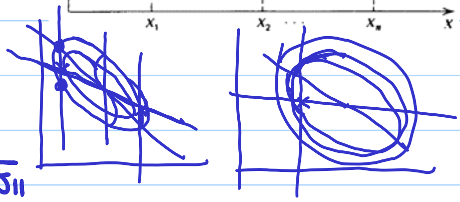
Here $\sigma_{11} - \sigma_{12}^2/\sigma_{22} = \sigma_{11}(1 - \rho_{12}^2)$.

mean of $x_1|x_2$

$\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) = \mu_1 + (\sigma_{12}/\sigma_{22})(x_2 - \mu)$

variance of $x_1|x_2$

$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} = \sigma_{11} - \frac{(\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}})^2}{\sigma_{22}}$



  - conditional density and regression model

$\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2)$

β_2 $= \left(\mu_1 - \frac{\sigma_{12}}{\sigma_{22}}\mu_2\right) + \underbrace{\frac{\sigma_{12}}{\sigma_{22}}}_{\beta_1} x_2$

$\mu_1 + \frac{\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}{\sigma_{22}}(x_2 - \mu_2)$

$= \mu_1 + \left[\rho_{12}\left(\frac{x_2 - \mu_2}{\sqrt{\sigma_{22}}}\right)\right] \cdot \sqrt{\sigma_{11}}$

➢ **Result 4.7.** Let $\mathbf{X}$ be distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $|\boldsymbol{\Sigma}| > 0$. Then

(a) $(\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ is distributed as χ_p^2, where χ_p^2 denotes the chi-square distribution with p degrees of freedom.

(b) The $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution assigns probability $1 - \alpha$ to the solid ellipsoid $\{\mathbf{x}: (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \le \chi_p^2(\alpha)\}$, where $\chi_p^2(\alpha)$ denotes the upper (100α)th percentile of the χ_p^2 distribution.

* sketch of proof: $(x-\mu)'\Sigma^{-1}(x-\mu) = [(x-\mu)'\Sigma^{-1/2}][\underbrace{\Sigma^{-1/2}(x-\mu)}_{\sim N(0, I)}]$