Multivariate Normal Distribution

Edps/Soc 584, Psych 594

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Spring 2017

Outline

- Motivation
- The multivariate normal distribution
- The Bivariate Normal Distribution
- More properties of multivariate normal
- Estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$
- Central Limit Theorem

Reading: Johnson & Wichern pages 149–176

Motivation

- To be able to make inferences about populations, we need a model for the distribution of random variables $\longrightarrow$ We'll use the multivariate normal distribution, because. . .
- It's often a good population model. It's a reasonably good approximation of many phenomenon. A lot of variables are approximately normal (due to the central limit theorem for sums and averages).
- The sampling distribution of (test) statistics are often approximately multivariate or univariate normal due to the central limit theorem.
- Due to it's central importance, we need to thoroughly understand and know it's properties.

Introduction to the Multivariate Normal

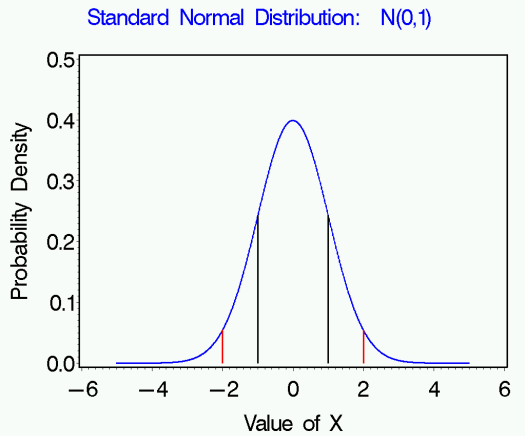
- The probability density function of the Univariate normal distribution ($p = 1$ variables):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \qquad \text{for } -\infty < x < \infty$$

- The parameters that completely characterize the distribution:
  - $\mu = E(X) =$ mean
  - $\sigma^2 = \text{var}(X) =$ variance

Introduction to the Multivariate Normal (continued)

Area corresponds to probability:
68% area between $\mu \pm \sigma$ and 95% between $\mu \pm 1.96\sigma$:



Generalization to Multivariate Normal

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$$

A squared statistical distance between x & μ in standard deviation units.

Generalization to $p > 1$ variables:

- We have $\mathbf{x}_{p\times 1}$ and parameters $\boldsymbol{\mu}_{p\times 1}$ and $\boldsymbol{\Sigma}_{p\times p}$.
- The exponent term for multivariate normal is

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

where $-\infty < x_i < \infty$ for $i = 1, \ldots, p$.

  - This is a scalar and reduces to what's at the top for $p = 1$.
  - It is a squared statistical distance of $\mathbf{x}$ to $\boldsymbol{\mu}$ (if $\boldsymbol{\Sigma}^{-1}$ exists). It takes into consideration both variability and covariability.
  - Integrating

$$\int_{x_1} \cdots \int_{x_p} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) = (2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}$$

Proper Distribution

Since the sum of probabilities over all possible values must add up to 1, we need to divide by $(2\pi)^{p/2}|\mathbf{\Sigma}|^{1/2}$ to get a "proper" density function.

Multivariate Normal density function:

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

where $-\infty < x_i < \infty$ for $i = 1, \ldots, p$.

To denote this, we use

$$\mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$$

For $p = 1$, this reduces to the univariate normal p.d.f.

Bivariate Normal: $p = 2$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad E(\mathbf{x}) = \begin{pmatrix} E(x_1) \\ E(x_2) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \boldsymbol{\mu}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

and

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

If we replace σ_{12} by $\rho_{12}\sqrt{\sigma_{11}\sigma_{22}}$, then we get

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22}(1 - \rho_{12}^2)} \begin{pmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}\sigma_{22}} & \sigma_{11} \end{pmatrix}$$

Using this, let's look at the statistical distance of $\mathbf{x}$ from $\boldsymbol{\mu}$. . .

Bivariate Normal & Statistical Distance

The quantity in the exponent of the bivariate normal is

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

$$\begin{aligned}
&= ((x_1-\mu_1),(x_2-\mu_2))\left(\frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}\right) \\
&\quad \times \begin{pmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}\sigma_{22}} - \rho_{12}\sqrt{\sigma_{11}\sigma_{22}} \\ \sigma_{11} & \end{pmatrix} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} \\
&= \frac{1}{1-\rho_{12}^2}\left\{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right\} \\
&= \frac{1}{1-\rho_{12}^2}\left\{z_1^2+z_2^2-2\rho_{12}z_1z_2\right\}
\end{aligned}$$

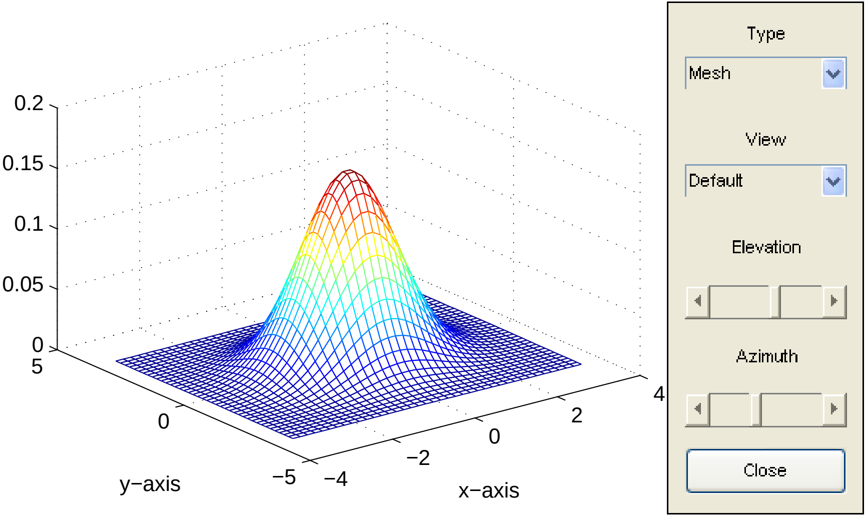
Bivariate Normal & Independence

$$f(\mathbf{x}) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}}} \exp\left[\frac{-1}{2(1-\rho_{12}^2)}\left\{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right\}\right]$$

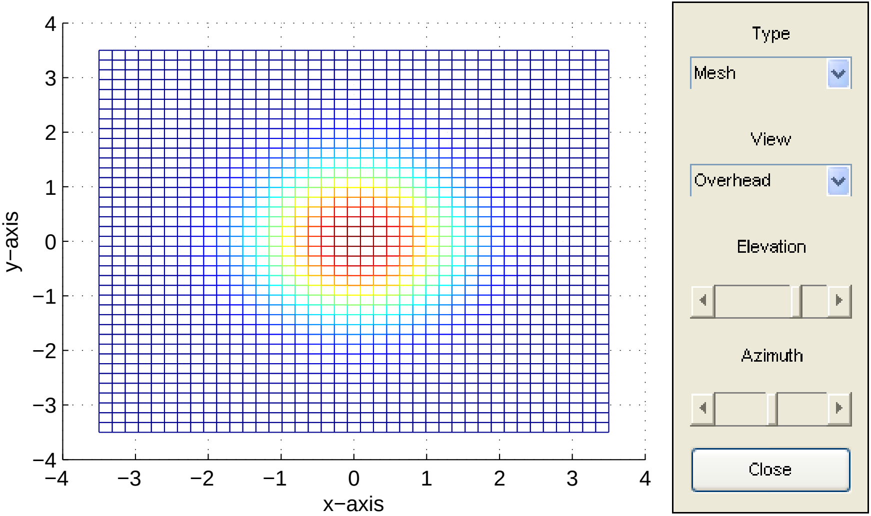
If $\sigma_{12}=0$ or equivalently $\rho_{12}=0$, then X_1 and X_2 are uncorrelated.
For bivariate normal, $\sigma_{12}=0$ implies that X_1 and X_2 are statistically independent, because the density factors

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}}} \exp\left[\frac{-1}{2}\left\{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right\}\right] \\ &= \frac{1}{\sqrt{2\pi\sigma_{11}}} \exp\left[\frac{-1}{2}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2\right] \frac{1}{\sqrt{2\pi\sigma_{22}}} \exp\left[\frac{-1}{2}\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right] \\ &= f_1(x_1)\times f_2(x_2) \end{aligned}$$

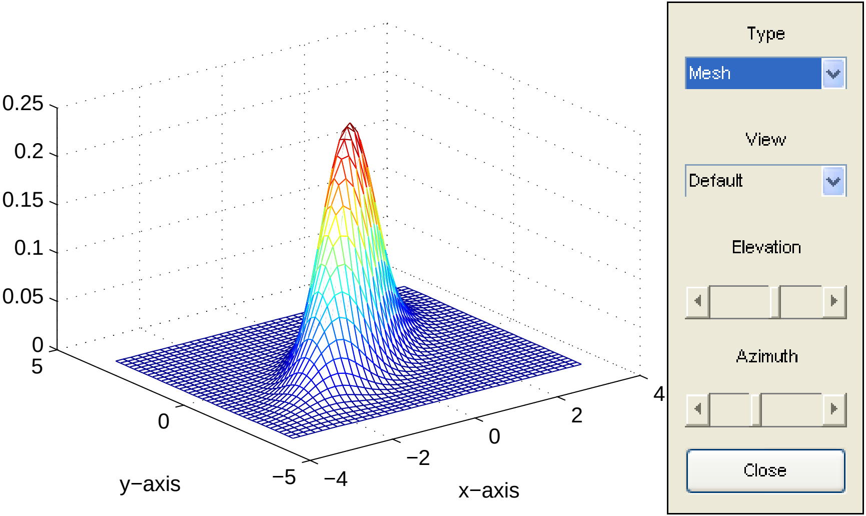
Picture: $\mu_k = 0$, $\sigma_{kk} = 1$, $r = 0.0$



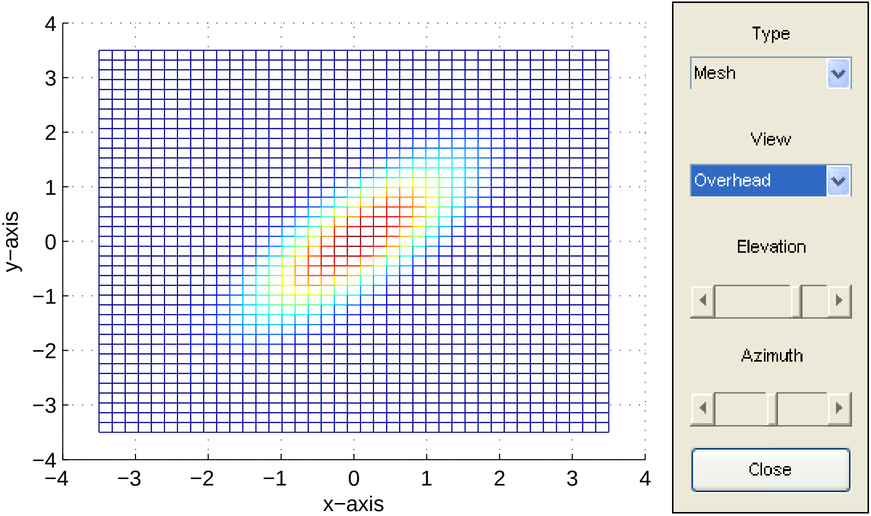
Overhead $\mu_k = 0$, $\sigma_{kk} = 1$, $r = 0.0$



Picture: $\mu_k = 0$, $\sigma_{kk} = 1$, $r = 0.75$



Overhead: $\mu_k = 0$, $\sigma_{kk} = 1$, $r = 0.75$



Summary: Comparing $r = 0.0$ vs $r = 0.75$

For the figures shown, $\mu_1 = \mu_2 = 0$ and $\sigma_{11} = \sigma_{22} = 1$:

- With $r = 0.0$,
  - $\mathbf{\Sigma} = \text{diag}(\sigma_{11}, \sigma_{22})$, a diagonal matrix.
  - Density is "random" in the x-y plane.
  - When take a slice parallel to x-y, you get a circle.
- When $r = .75$,
  - $\mathbf{\Sigma}$ is not a diagonal .
  - Density is not random in x-y plane.
  - There is a linear tilt (ie., density is concentrated on a line).
  - When you take a slice you get an ellipse that's tilted.
  - Tilt depends on relative values of σ_{11} and σ_{22} (and scale used in plotting).
- When $\mathbf{\Sigma} = \sigma^2\mathbf{I}$ (i.e., diagonal with equal variances), it's "spherical normal".

Real Time Software Demo

- binormal.m (Peter Dunn)
- Graph_Bivariate_.R (http://www.stat.ucl.ac.be/ISpersonnel/lecoutre/stats/fichiers/~gall

Slices of Multivariate Normal Density

- For bi-variate normal, you get an ellipse whose equation is

$$(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c^2$$

  which gives all (x_1, x_2) pairs with constant probability.
- The ellipses are call contours and all are centered around $\boldsymbol{\mu}$.
- Definition:

A constant probability contour equals

$$\begin{aligned} &= \{\text{all } \mathbf{x} \text{ such that } (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c^2\} \\ &= \{\text{surface of ellipsoid centered at } \boldsymbol{\mu}\} \end{aligned}$$

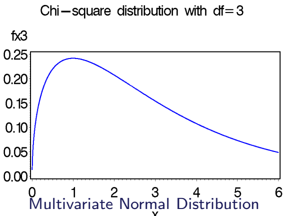
Probability Contours: Axes of ellipsoid

Important Points:

- $(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \sim \chi^2_p \qquad (\text{if } |\boldsymbol{\Sigma}| > 0)$
- The solid ellipsoid of values $\mathbf{x}$ that satisfy

$$(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \le c^2 = \chi^2_{p(\alpha)}$$

has probability $(1-\alpha)$ where $\chi^2_{p(\alpha)}$ is the $(1-\alpha)^{th}100\%$ point of the chi-square distribution with p degrees of freedom.



Example: Axses of Ellipses & Prob. Contours

Back to the example where $\mathbf{x} \sim N_2$ with

$$\boldsymbol{\mu} = \begin{pmatrix} 5 \\ 10 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 9 & 16 \\ 16 & 64 \end{pmatrix} \rightarrow \rho = .667$$

and we want the "95% probability contour".

The upper 5% point of the chi-square distribution with 2 degrees of freedom is $\chi^2_{2(.05)} = 5.9915$, so $c = \sqrt{5.9915} = 2.4478$

Axes: $\boldsymbol{\mu} \pm c\sqrt{\lambda_i}\mathbf{e}_i$ where $(\lambda_i, \mathbf{e}_i)$ is the i^{th} $(i = 1, 2)$ eigenvalue/eigenvector pair of $\boldsymbol{\Sigma}$.

$$\lambda_1 = 68.316 \qquad \mathbf{e}_1' = (.2604, .9655)$$
$$\lambda_2 = 4.684 \qquad \mathbf{e}_2' = (.9655, -.2604)$$

Major Axis

Using the largest eigenvalue and corresponding eigenvector:

$$\underbrace{\begin{pmatrix} 5 \\ 10 \end{pmatrix}}_{\boldsymbol{\mu}} \pm \underbrace{2.45}_{\sqrt{\chi^2_{2(.05)}}} \underbrace{\sqrt{68.316}}_{\lambda_1} \underbrace{\begin{pmatrix} .2604 \\ .9655 \end{pmatrix}}_{\mathbf{e}_1}$$

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} \pm 20.250 \begin{pmatrix} .2604 \\ .9655 \end{pmatrix}$$

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} \pm \begin{pmatrix} 5.273 \\ 19.551 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} -.273 \\ -9.551 \end{pmatrix}, \quad \begin{pmatrix} 10.273 \\ 29.551 \end{pmatrix}$$

Minor Axis

Same process but now use λ_2 and $\mathbf{e}_2$, the smallest eigenvalue and corresponding eigenvector:

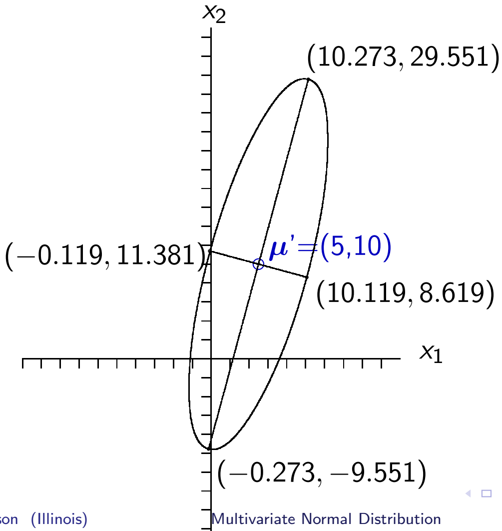
$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} \pm 2.45\sqrt{4.684}\begin{pmatrix} .9655 \\ -.2604 \end{pmatrix}$$

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} \pm 5.30\begin{pmatrix} .9655 \\ -.2604 \end{pmatrix}$$

$$\begin{pmatrix} 5 \\ 10 \end{pmatrix} \pm \begin{pmatrix} 5.119 \\ -1.381 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} -.119 \\ 11.381 \end{pmatrix}, \quad \begin{pmatrix} 10.119 \\ 8.619 \end{pmatrix}$$

Graph of 95% Probability Contour



Example: Equation for Contour

Equation for Contour:

$$(\mathbf{x}-\boldsymbol{\mu})' \quad \boldsymbol{\Sigma}^{-1} \quad (\mathbf{x}-\boldsymbol{\mu}) \quad \leq \quad 5.99$$

$$((x_1-5),(x_2-10))\begin{pmatrix} 9 & 16 \\ 16 & 64 \end{pmatrix}^{-1}\begin{pmatrix} (x_1-5) \\ (x_2-10) \end{pmatrix} \leq 5.99$$

$$((x_1-5),(x_2-10))\begin{pmatrix} .200 & -.050 \\ -.050 & .028 \end{pmatrix}\begin{pmatrix} (x_1-5) \\ (x_2-10) \end{pmatrix} \leq 5.99$$

$$.2(x_1-5)^2+.028(x_2-10)^2-.1(x_1-5)(x_2-10) \leq 5.99$$

$(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ is a quadratic form, which is equation for a polynomial

Points inside or outside?

Are the following points inside or outside the 95% probability contour?

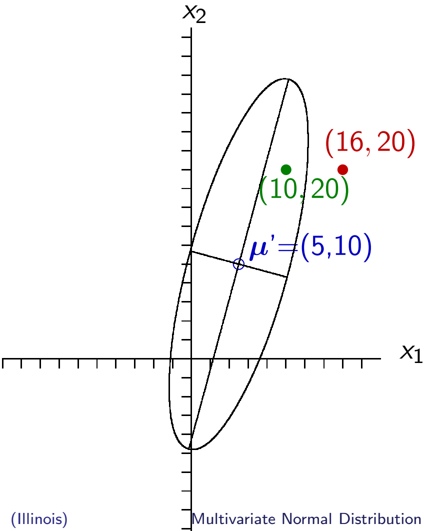
- Is the point (10,20) inside or outside the 95% probability contour?

$$\begin{aligned}(10,20) \quad \longrightarrow \quad & .2(10-5)^2 + .028(20-10)^2 - .1(10-5)(20-10) \\ & = .2(25) + .028(100) - .1(50) \\ & = 2.8\end{aligned}$$

- Is the point (16,20) inside or outside the 95% probability contour?

$$\begin{aligned}(16,20) \quad \longrightarrow \quad & .2(16-5)^2 + .028(20-10)^2 - .1(16-5)(20-10) \\ & .2(121) + .028(100) - .1(11)(10) \\ & = 16\end{aligned}$$

Points Inside and Outside



More Properties that we'll Expand on

If $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

- Linear combinations of components of $\mathbf{X}$ are (multivariate) normal.
- All sub-sets of the components of $\mathbf{X}$ are (multivariate) normal.
- Zero covariance implies that the corresponding components of $\mathbf{X}$ are statistical independent.
- The conditional distributions of the components of $\mathbf{X}$ are (multivariate) normal.

1: Linear Combinations

If $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any linear combination

$$\mathbf{a}'\mathbf{X} = a_1X_1 + a_2X_2 + \cdots + a_pX_p$$

is distributed as

$$\mathbf{a}'\mathbf{X} \sim \mathcal{N}_1(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$$

Also, If $\mathbf{a}'\mathbf{X}$ is normal $\mathcal{N}(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ for all possible $\mathbf{a}$, then $\mathbf{X}$ must be $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

$$\mathbf{X} \sim \mathcal{N}\left(\begin{pmatrix} 5 \\ 10 \end{pmatrix}, \begin{pmatrix} 16 & 12 \\ 12 & 36 \end{pmatrix}\right) \qquad \begin{array}{l} \mathbf{a}' = (3, 2) \\ Y = \mathbf{a}'\mathbf{X} = 3X_1 + 2X_2 \end{array}$$

$$\mu_Y = (3, 2)\begin{pmatrix} 5 \\ 10 \end{pmatrix} = 35 \quad \text{and } \sigma_Y^2 = (3, 2)\begin{pmatrix} 16 & 12 \\ 12 & 36 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \end{pmatrix} = 432$$

$$Y \sim \mathcal{N}(35, 432)$$

More Linear Combinations

If $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the q linear combinations

$$\mathbf{Y}_{q\times 1} = \mathbf{A}_{q\times p}\mathbf{X} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{q1} & a_{q2} & \cdots & a_{qp} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

is distributed as $\mathcal{N}_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

Also, if

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{d},$$

where $\mathbf{d}_{q\times 1}$ is a vector constants, then

$$\mathbf{Y} = \mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{d}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}').$$

Numerical Example with Multiple Combinations

$$\mathbf{X} \sim \mathcal{N}_2\left(\begin{pmatrix} 5 \\ 10 \end{pmatrix}\begin{pmatrix} 16 & 12 \\ 12 & 36 \end{pmatrix}\right)$$

$$\begin{array}{l} Y_1 = X_1 + X_2 \\ Y_2 = X_1 - X_2 \end{array} \quad \text{so} \quad \mathbf{A}_{2\times 2} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

$$\mu_Y = \mathbf{A}\boldsymbol{\mu} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 5 \\ 10 \end{pmatrix} = \begin{pmatrix} 15 \\ -5 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_Y = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 16 & 12 \\ 12 & 36 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 76 & -20 \\ -20 & 28 \end{pmatrix}$$

So

$$\mathbf{Y} \sim \mathcal{N}_2\left(\begin{pmatrix} 15 \\ -5 \end{pmatrix}, \begin{pmatrix} 76 & -20 \\ -20 & 28 \end{pmatrix}\right)$$

Multiple Regression as an Example

This example will use what we know about linear combinations and now what we know about the distribution of linear combinations.

Linear Regression Model

- Y = response variable.
- $Z_1, Z_2, \ldots, Z_r$ are predictor/explanatory variables, which are considered to be fixed.
- The model is

$$Y = \beta_o + \beta_1 Z_1 + \beta_2 Z_2 + \ldots + \beta_r Z_r + \epsilon$$

- The error of prediction ϵ is viewed as a random variable.

Multiple Regression as an Example

Suppose we have n observations on Y and have values of Z_i for all $i = 1, \ldots, n$; that is,

$$\begin{aligned} Y_1 &= \beta_o + \beta_1 Z_{11} + \beta_2 Z_{12} + \ldots + \beta_r Z_{1r} + \epsilon_1 \\ Y_2 &= \beta_o + \beta_1 Z_{21} + \beta_2 Z_{22} + \ldots + \beta_r Z_{2r} + \epsilon_2 \\ \vdots & \quad \vdots \\ Y_n &= \beta_o + \beta_1 Z_{n1} + \beta_2 Z_{n2} + \ldots + \beta_r Z_{nr} + \epsilon_n \end{aligned}$$

where $E(\epsilon_j) = 0$, $\text{var}(\epsilon_j) = \sigma^2$ (a constant), and $\text{cov}(\epsilon_j, \epsilon_k) = 0$ for $j \neq k$.

In terms of matrices,

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & Z_{11} & Z_{12} & \ldots & Z_{1r} \\ 1 & Z_{21} & Z_{22} & \ldots & Z_{2r} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & Z_{n1} & Z_{n2} & \ldots & Z_{nr} \end{pmatrix} \begin{pmatrix} \beta_o \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \epsilon \quad \text{where } E(\epsilon) = \mathbf{0} \text{ and } \text{cov}(\epsilon) = \sigma^2 \mathbf{I}.$$

Distribution of Y

$$\mathbf{Y} = \underbrace{\mathbf{Z}\boldsymbol{\beta}}_{\text{vector of constants}} + \underbrace{\epsilon}_{\text{random}} \quad \text{where } E(\epsilon) = \mathbf{0} \text{ and } \text{cov}(\epsilon) = \sigma^2\mathbf{I}.$$

So $\mathbf{Y}$ is a linear combination of a multivariate normally distributed variable, ϵ.

- Mean of $\mathbf{Y}$:

$$\boldsymbol{\mu}_{\mathbf{Y}} = E(\mathbf{Y}) = E(\mathbf{Z}\boldsymbol{\beta} + \epsilon) = \mathbf{Z}\boldsymbol{\beta} + E(\epsilon) = \mathbf{Z}\boldsymbol{\beta}$$

- Covariance of $\mathbf{Y}$:

$$\boldsymbol{\Sigma}_{\mathbf{Y}} = \sigma^2\mathbf{I}$$

(the same as ϵ).

- Distribution of $\mathbf{Y}$ is multivariate normal because ϵ is multivariate normal:

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{Z}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

Least Square Estimation

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \epsilon \quad \text{where } E(\epsilon) = \mathbf{0} \text{ and } \text{cov}(\epsilon) = \sigma^2 \mathbf{I}$$

$\boldsymbol{\beta}$ and σ^2 are unknown parameters that need to be estimated from data.

Let $y_1, y_2, \ldots, y_n$ be a random sample with values $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_r$ on the explanatory variables. The least squares estimate of $\boldsymbol{\beta}$ is the vector $\mathbf{b}$ that minimizes

$$\begin{aligned} \sum_{j=1}^{n} (y_j - \mathbf{z}_j'\mathbf{b})^2 &= \sum_{j=1}^{n} (y_j - b_o - b_1 z_{j1} - b_2 z_{j2} - \ldots - b_r z_{jr})^2 \\ &= (\mathbf{y} - \mathbf{Z}\mathbf{b})'(\mathbf{y} - \mathbf{Z}\mathbf{b}) \\ &= \epsilon'\epsilon \end{aligned}$$

where $\mathbf{z}_j'$ is the j^{th} row of $\mathbf{Z}$, and $\mathbf{b} = (b_o, b_1, b_2, \ldots, b_r)'$.
If $\mathbf{Z}$ has full rank (i.e., the rank of $\mathbf{Z}$ is $r + 1 \leq n$), then the least squares estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

What's the distribution of $\hat{\beta}$?

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \mathbf{A}\mathbf{y}$$

We showed that $\mathbf{Y} \sim \mathcal{N}_n(\mathbf{Z}\boldsymbol{\beta}, \sigma^2\mathbf{I})$.

- Mean of $\hat{\boldsymbol{\beta}}$:

$$\begin{aligned} \mu_{\hat{\boldsymbol{\beta}}} = E(\hat{\boldsymbol{\beta}}) &= E(\mathbf{A}\mathbf{Y}) \\ &= \mathbf{A}E(\mathbf{Y}) \\ &= \mathbf{A}\mathbf{Z}\boldsymbol{\beta} \\ &= \underbrace{(\mathbf{Z}'\mathbf{Z})^{-1}}\underbrace{\mathbf{Z}'\mathbf{Z}}\boldsymbol{\beta} = \boldsymbol{\beta} \end{aligned}$$

- Covariance matrix for $\hat{\boldsymbol{\beta}}$

$$\begin{aligned} \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} &= \mathbf{A}\boldsymbol{\Sigma}_Y\mathbf{A}' \\ &= ((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')(\sigma^2\mathbf{I})(\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}) \\ &= \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \\ &= \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1} \end{aligned}$$

- The distribution of $\hat{\boldsymbol{\beta}}$: $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1})$.

The distribution of $\hat{Y}$

The "fitted values" or predicted values are

$$\hat{\mathbf{y}} = \mathbf{Z}\hat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y}$$

where $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$. The matrix $\mathbf{H}$ is the "hat" matrix.

- We just showed that $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1})$, and so $\hat{\mathbf{y}}$ is a linear combination of a vector that's multivariate normal.
- Mean of $\hat{\mathbf{Y}}$:
$$\mu_{\hat{\mathbf{Y}}} = E(\mathbf{Z}\hat{\boldsymbol{\beta}}) = \mathbf{Z}E(\hat{\boldsymbol{\beta}}) = \mathbf{Z}\boldsymbol{\beta}$$
- Covariance matrix for $\hat{\mathbf{Y}}$
$$\mathbf{Z}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}\mathbf{Z}' = \mathbf{Z}(\sigma^2\underbrace{(\mathbf{Z}'\mathbf{Z})^{-1})}\underbrace{\mathbf{Z}'} = \sigma^2\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} = \sigma^2\mathbf{I}$$
- Distribution of $\hat{\mathbf{Y}}$:
$$\hat{\mathbf{Y}} \sim \mathcal{N}(\mathbf{Z}\boldsymbol{\beta}, \sigma^2\mathbf{I})$$

The distribution of $\hat{\epsilon}$

The estimated residuals are

$$\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

and they contain the information necessary to estimate σ^2.

The least squares estimate of σ^2 is

$$s^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n - (r+1)}$$

The estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\epsilon}$ are uncorrelated.

Multivariate Normality Assumption $\epsilon \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$ and what we know about linear combinations of random variables allowed us to derive the distribution of various random variables.

The distribution of $\hat{\epsilon}$

Last few comments on this example:

- The least squares estimates of $\boldsymbol{\beta}$ and ϵ are also the maximum likelihood estimates.
- The maximum likelihood estimate of σ^2 is $\hat{\sigma}^2 = \hat{\epsilon}'\hat{\epsilon}/n$
- $\hat{\boldsymbol{\beta}}$ and $\hat{\epsilon}$ are statistically independent.

2: Sub-sets of Variables

If $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then all sub-sets of $\mathbf{X}$ are (multivariate) normally distributed.

For example, let's partition $\mathbf{X}$ into two sub-sets

$$\mathbf{X}_{p\times 1} = \begin{pmatrix} X_1 \\ \vdots \\ X_q \\ \hline X_{q+1} \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{1(q\times 1)} \\ \hline \mathbf{X}_{2((p-q)\times 1)} \end{pmatrix} \text{ and } \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_q \\ \hline \mu_{q+1} \\ \vdots \\ \mu_p \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_{1(q\times 1)} \\ \hline \boldsymbol{\mu}_{2((p-q)\times 1)} \end{pmatrix}$$

$$\boldsymbol{\Sigma}_{p\times p} = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11(q\times q)} & \boldsymbol{\Sigma}_{12(q\times(p-q))} \\ \hline \boldsymbol{\Sigma}_{21((p-q)\times p)} & \boldsymbol{\Sigma}_{22((p-q)\times(p-q))} \end{array}\right) = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right)$$

Sub-sets of Variables continued

Then for

$$\mathbf{X} = \left(\frac{\mathbf{X}_{1(q\times 1)}}{\mathbf{X}_{2((p-q)\times 1)}} \right)$$

The distributions of the sub-sets are

$$\mathbf{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \qquad \text{and} \qquad \mathbf{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$$

The result means that

- Each of the $\mathbf{X}_i$'s are univariate normals (next page)
- All possible sub-sets are multivariate normal.
- All marginal distributions are (multivariate) normal.

Little Example on Sub-sets

Suppose that

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathcal{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Due to the result on sub-sets of multivariate normals,

$$\begin{aligned} X_1 &\sim \mathcal{N}(\mu_1, \sigma_{11}) \\ X_2 &\sim \mathcal{N}(\mu_2, \sigma_{22}) \\ X_3 &\sim \mathcal{N}(\mu_3, \sigma_{33}) \end{aligned}$$

Also

$$\begin{pmatrix} X_2 \\ X_3 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \begin{pmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} \end{pmatrix}\right)$$

3: Zero Covariance & Statistical Independence

There are three parts to this one:

- If $\mathbf{X}_1$ is $(q_1 \times 1)$ and $\mathbf{X}_2$ is $(q_2 \times 1)$ are statistically independent, then $\text{cov}(\mathbf{X}_1, \mathbf{X}_2) = \boldsymbol{\Sigma}_{12} = \mathbf{0}$.
- If
$$\begin{pmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{pmatrix} \sim \mathcal{N}_{q_1+q_2}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{pmatrix}, \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array} \right) \right),$$
Then $\mathbf{X}_1$ and $\mathbf{X}_2$ are statistically independent if and only if $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' = \mathbf{0}$.
- If $\mathbf{X}_1$ and $\mathbf{X}_2$ are statistically independent and distributed as $\mathcal{N}_{q_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\mathcal{N}_{q_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, respectively, then
$$\begin{pmatrix} \mathbf{X}_1 \\ \hline \mathbf{X}_2 \end{pmatrix} \sim \mathcal{N}_{q_1+q_2}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{pmatrix}, \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{array} \right) \right).$$

Example

$$\mathbf{Y}_{4\times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} \text{ and } \mathbf{\Sigma}_{\mathbf{Y}} = \begin{pmatrix} 2 & 1 & 0 & .5 \\ 1 & 3 & 0 & .5 \\ 0 & 0 & 4 & 0 \\ .5 & .5 & 0 & 1 \end{pmatrix}$$

and $\mathbf{Y} \sim \mathcal{N}_4(\boldsymbol{\mu}, \mathbf{\Sigma})$.
Let's take $\mathbf{X}_1' = (Y_1, Y_2, Y_4)$ and $\mathbf{X}_2' = (Y_3)$.

Then

$$\left(\frac{\mathbf{X}_1}{\mathbf{X}_2}\right) \sim \mathcal{N}_4\left(\left(\begin{array}{c} \mu_1 \\ \mu_2 \\ \mu_4 \\ \hline \mu_3 \end{array}\right), \left(\begin{array}{ccc|c} 2 & 1 & .5 & 0 \\ 1 & 3 & .5 & 0 \\ .5 & .5 & 1 & 0 \\ \hline 0 & 0 & 0 & 4 \end{array}\right)\right)$$

So set $\mathbf{X}_1$ is statistically independent of $\mathbf{X}_2$.

4: Conditional Distributions

Let $\mathbf{X}' = (\mathbf{X}'_{1(q1\times 1)}, \mathbf{X}'_{2(q_2\times 1)})$ be distributed at $\mathcal{N}_{q1+q2}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \left(\frac{\boldsymbol{\mu}_1}{\boldsymbol{\mu}_2}\right) \quad \text{and} \quad \boldsymbol{\Sigma} = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right)$$

and $|\boldsymbol{\Sigma}| > 0$ (i.e., positive definite). Then the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is (multivariate) normal with mean and covariance matrix

$$\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \quad \text{and} \quad \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

Let's look more closely at this for a simple case of $q_1 = q_2 = 1$.

Conditional Distribution for $q_1 = q_2 = 1$

Bivariate normal distribution

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right)$$

$$f(x_1|x_2) \quad \text{is} \quad \mathcal{N}_1\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \sigma_{12}\left(\frac{\sigma_{12}}{\sigma_{22}}\right)\right)$$

Notes:

- $\sigma_{12} = \rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$
- $\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1} = \sigma_{12}/\sigma_{22} = \rho_{12}(\sqrt{\sigma_{11}}/\sqrt{\sigma_{22}})$
- $\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} = \sigma_{11} - \sigma_{12}^2/\sigma_{22} = \sigma_{11}(1 - \rho_{12}^2)$

Alternative way to write $f(x_1|x_2)$:

$$f(x_1|x_2) \quad \text{is} \quad \mathcal{N}_1\left(\mu_1 + \rho_{12}\frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}(x_2 - \mu_2), \sigma_{11}(1 - \rho_{12}^2)\right)$$

Multiple Regression as a Conditional Dist.

Consider the case where $q_1 = 1$ and $q_2 > 1$.

- All conditional distributions are normal.
- The conditional covariance matrix $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ does not depend on the values of the conditioning variables.
- The conditional means have the following form:

$$\text{Let } \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} = \boldsymbol{\beta}_{q_1 \times q_2} = \begin{pmatrix} \beta_{1,q_1+1} & \beta_{1,q_1+2} & \cdots & \beta_{1,q_1+q_2} \\ \beta_{2,q_1+1} & \beta_{2,q_1+2} & \cdots & \beta_{2,q_1+q_2} \\ \cdots & \cdots & \ddots & \cdots \\ \beta_{q_1,q_1+1} & \beta_{q_1,q_1+2} & \cdots & \beta_{q_1,q_1+q_2} \end{pmatrix}$$

$$\text{Condtional means} \quad \begin{pmatrix} \mu_1 + \sum_{i=q_1+1}^{q_1+q_2} \beta_{1i}(x_i - \mu_i) \\ \mu_2 + \sum_{i=q_1+1}^{q_1+q_2} \beta_{2i}(x_i - \mu_i) \\ \vdots \\ \mu_{q_1} + \sum_{i=q_1+1}^{q_1+q_2} \beta_{q_1 i}(x_i - \mu_i) \end{pmatrix}$$

Estimation of μ and Σ

& sampling distribution of estimators.

Suppose we have a p dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Take n observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ (these are each $(p \times 1)$ vectors).

$$\mathbf{X}_j \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad j = 1, 2, \ldots, n \text{ and independent}$$

For $p = 1$, we know that the MLEs are

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \sim \mathcal{N}\left(\mu, \frac{1}{n}\sigma^2\right)$$

And $$n\hat{\sigma}^2 = \sum_{j=1}^{n}(x_j - \bar{x})^2 \text{ and } \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \bar{x})^2 \sim \chi^2_{(n-1)}$$

Or $$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 \sim \sigma^2\chi^2_{(n-1)}$$

Estimation of μ and Σ: Multivariate Case

The maximum likelihood estimator of $\boldsymbol{\mu}$ is

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

and the ML estimator of $\boldsymbol{\Sigma}$ is

$$\hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S}^2 = \mathbf{S}_n = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{X}_j - \hat{\mu})(\mathbf{X}_j - \hat{\mu})'$$

Sampling Distribution of $\hat{\boldsymbol{\mu}}$:

The estimator is a linear combination of normal random vectors each from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ $\quad i.i.d.$:

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = \frac{1}{n}\mathbf{X}_1 + \frac{1}{n}\mathbf{X}_2 + \cdots + \frac{1}{n}\mathbf{X}_n$$

So $\hat{\boldsymbol{\mu}}$ also has a normal distribution,

Sampling Distribution of $\hat{\Sigma}$

$$\hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S}$$

The matrix

$$(n-1)\mathbf{S} = \sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$$

is distributed as a Wishart random matrix with $(n-1)$ degrees of freedom.

Whishart distribution:

- A multivariate analogue to the chi-square distribution.
- It's defined as

$$\begin{aligned} W_m(\cdot|\boldsymbol{\Sigma}) &= \text{Wishart distribution with } m \text{ degrees of freedom} \\ &= \text{The distribution of } \sum_{j=1}^{m} \mathbf{Z}_j\mathbf{Z}_j' \end{aligned}$$

where $\mathbf{Z}_j \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ and independent.

Note : $\bar{\mathbf{X}}$ and $\mathbf{S}$ are independent.

Law of Large Numbers

Data are not always (multivariate) normal

The Law of Large Numbers (for multivariate data):

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from a population with mean $E(\mathbf{X}) = \boldsymbol{\mu}$.

Then $\bar{\mathbf{X}} = (1/n)\sum_{j=1}^{n} \mathbf{X}_j$ converges in probability to μ as n gets large; that is,

$$\bar{\mathbf{X}} \rightarrow \boldsymbol{\mu} \text{ for large samples}$$

And

$$\mathbf{S}(\text{or } \mathbf{S}_n) \text{ approach } \boldsymbol{\Sigma} \text{ for large samples}$$

These are true regardless of the true distribution of the $\mathbf{X}_j$'s.

Central Limit Theorem

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from a population with mean $E(\mathbf{X}) = \boldsymbol{\mu}$ and finite (non-singular, full rank), covariance matrix $\boldsymbol{\Sigma}$.

Then $\sqrt{n}(\mathbf{X} - \boldsymbol{\mu})$ has an approximate $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ distribution if $n >> p$ (i.e., "much larger than").

So, for "large" n

$$\bar{\mathbf{X}} = \text{ Sample mean vector } \approx \mathcal{N}(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}),$$

regardless of the underlying distribution of the $\mathbf{X}_j$'s.

What if $\boldsymbol{\Sigma}$ is unknown? If n is large "enough", $\mathbf{S}$ will be close to $\boldsymbol{\Sigma}$, so

$$\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \approx \mathcal{N}_p(\mathbf{0}, \mathbf{S}) \text{ or } \bar{\mathbf{X}} \approx \mathcal{N}_p(\boldsymbol{\mu}, \frac{1}{n}\mathbf{S}).$$

Since $n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim \chi^2_p$,

$$n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \approx \chi^2_p$$

Few more comments

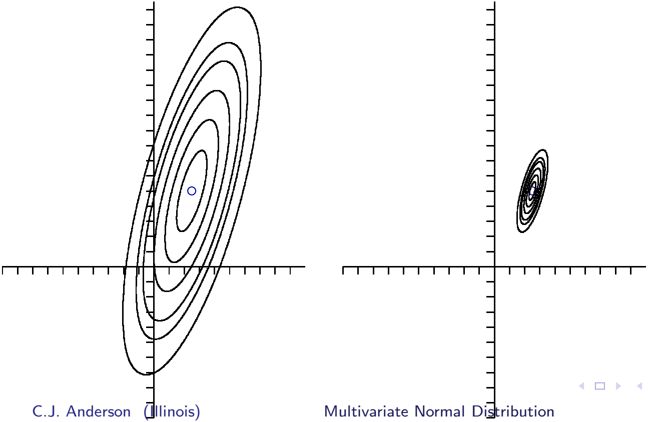
- Using $\mathbf{S}$ instead of $\mathbf{\Sigma}$ does not seriously effect approximation.
- n must be large relative to p; that is, $(n - p)$ is large.
- The probability contours for $\bar{\mathbf{X}}$ are tighter than those for $\mathbf{X}$ since we have $(1/n)\mathbf{\Sigma}$ for $\bar{\mathbf{X}}$ rather than $\mathbf{\Sigma}$ for $\mathbf{X}$.

  See next slide for an example of the latter.

Comparison of Probability Contours

Returning to our example and pretending we have $n = 20$. Below are contours for 99%, 95%, 90%, 75%, 50% and 20%:

Contours for $\mathbf{X}_j$ Contours for $\bar{\mathbf{X}}$



Why So Much a Difference with Only 20?

For $\mathbf{X}_j$

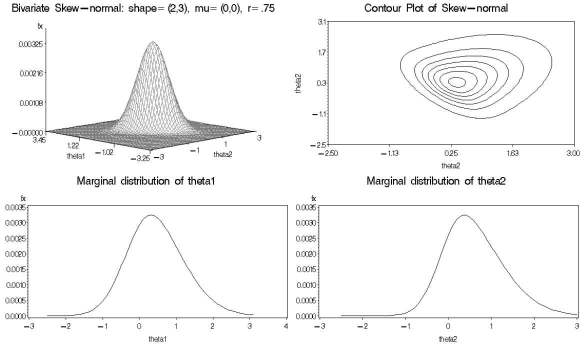
$$\boldsymbol{\Sigma} = \begin{pmatrix} 9 & 16 \\ 16 & 64 \end{pmatrix} \longrightarrow \lambda_1 = 68.316 \text{ and } \lambda_2 = 4.684$$

For $\bar{\mathbf{X}}$ with $n = 20$

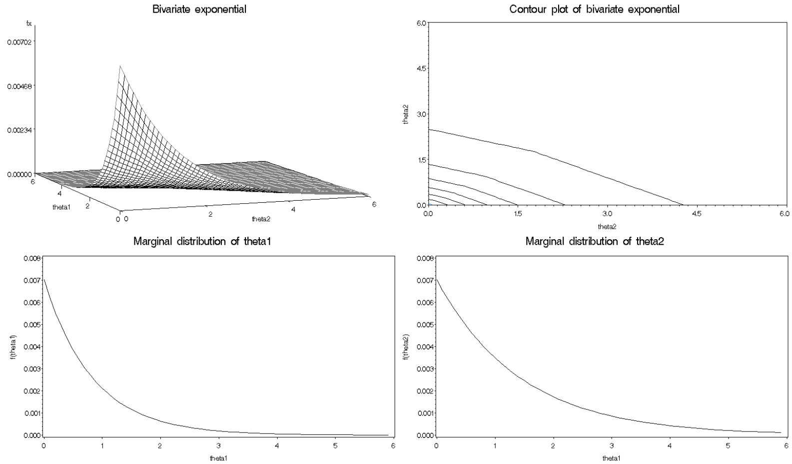
$$\boldsymbol{\Sigma} = \frac{1}{20}\begin{pmatrix} 9 & 16 \\ 16 & 64 \end{pmatrix} = \begin{pmatrix} 0.45 & 0.80 \\ 0.80 & 3.20 \end{pmatrix} \longrightarrow \lambda_1 = 3.42 \text{ and } \lambda_2 = 0.23$$

Note that $68.316/20 = 3.42$ and $4.684/20 = 0.23$.

Other Multivariate Distributions: Skew-Normal



Marshall-Olkin bivariate exponential



Contours for 4 different ones