The Multivariate Normal Distribution[1]

STA 442/2101 Fall 2014

[1] See last slide for copyright information.

Overview

Joint moment-generating function

Of a p-dimensional random vector $\mathbf{X}$

- $M_{\mathbf{X}}(\mathbf{t}) = E\left(e^{\mathbf{t}^\top \mathbf{X}}\right)$
- For example, $M_{(X_1,X_2,X_3)}(t_1,t_2,t_3) = E\left(e^{X_1t_1+X_2t_2+X_3t_3}\right)$

Section 4.3 of *Linear models in statistics* has some material on moment-generating functions (optional).

Two big theorems

Proof omitted

1. Joint moment-generating functions correspond uniquely to joint probability distributions.
2. Two random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if the moment-generating function of their joint distribution is the product of their moment-generating functions.

These results assume only that the moment-generating functions exist in a neighborhood of $\mathbf{t} = \mathbf{0}$. Nothing else is required.

A helpful distinction

- If X_1 and X_2 are independent,

$$M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t)$$

- X_1 and X_2 are independent if and only if

$$M_{X_1,X_2}(t_1,t_2) = M_{X_1}(t_1)M_{X_2}(t_2)$$

Theorem: Functions of independent random vectors are independent

Show $\mathbf{X}_1$ and $\mathbf{X}_2$ independent implies that $\mathbf{Y}_1 = g_1(\mathbf{X}_1)$ and $\mathbf{Y}_2 = g_2(\mathbf{X}_2)$ are independent.

Let

$$\mathbf{Y} = \left(\frac{\mathbf{Y}_1}{\mathbf{Y}_2} \right) = \left(\frac{g_1(\mathbf{X}_1)}{g_2(\mathbf{X}_2)} \right) \text{ and } \mathbf{t} = \left(\frac{\mathbf{t}_1}{\mathbf{t}_2} \right). \text{ Then}$$

$$\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) &= E\left(e^{\mathbf{t}^\top \mathbf{Y}}\right) \\
&= E\left(e^{\mathbf{t}_1^\top \mathbf{Y}_1 + \mathbf{t}_2^\top \mathbf{Y}_2}\right) = E\left(e^{\mathbf{t}_1^\top \mathbf{Y}_1} e^{\mathbf{t}_2^\top \mathbf{Y}_2}\right) \\
&= E\left(e^{\mathbf{t}_1^\top g_1(\mathbf{X}_1)} e^{\mathbf{t}_2^\top g_2(\mathbf{X}_2)}\right) \\
&= \int \int e^{\mathbf{t}_1^\top g_1(\mathbf{x}_1)} e^{\mathbf{t}_2^\top g_2(\mathbf{x}_2)} f_{\mathbf{X}_1}(\mathbf{x}_1) f_{\mathbf{X}_2}(\mathbf{x}_2)\, d(\mathbf{x}_1) d(\mathbf{x}_2) \\
&= M_{g_1(\mathbf{X}_1)}(\mathbf{t}_1) M_{g_2(\mathbf{X}_2)}(\mathbf{t}_2)
\end{aligned}$$

$M_{\mathbf{AX}}(\mathbf{t}) = M_{\mathbf{X}}(\mathbf{A}^\top \mathbf{t})$

Analogue of $M_{aX}(t) = M_X(at)$

$$\begin{aligned} M_{\mathbf{AX}}(\mathbf{t}) &= E\left(e^{\mathbf{t}^\top \mathbf{AX}}\right) \\ &= E\left(e^{\left(\mathbf{A}^\top \mathbf{t}\right)^\top \mathbf{X}}\right) \\ &= M_{\mathbf{X}}(\mathbf{A}^\top \mathbf{t}) \end{aligned}$$

Note that $\mathbf{t}$ is the same length as $\mathbf{Y} = \mathbf{AX}$: The number of rows in $\mathbf{A}$.

$M_{\mathbf{X}+\mathbf{c}}(\mathbf{t}) = e^{\mathbf{t}^\top \mathbf{c}} M_{\mathbf{X}}(\mathbf{t})$

Analogue of $M_{X+c}(t) = e^{ct} M_X(t)$

$$
\begin{aligned}
M_{\mathbf{X}+\mathbf{c}}(\mathbf{t}) &= E\left(e^{\mathbf{t}^\top(\mathbf{X}+\mathbf{c})}\right) \\
&= E\left(e^{\mathbf{t}^\top\mathbf{X}+\mathbf{t}^\top\mathbf{c}}\right) \\
&= e^{\mathbf{t}^\top\mathbf{c}}\, E\left(e^{\mathbf{t}^\top\mathbf{X}}\right) \\
&= e^{\mathbf{t}^\top\mathbf{c}} M_{\mathbf{X}}(\mathbf{t})
\end{aligned}
$$

Distributions may be defined in terms of moment-generating functions

Build up the multivariate normal from univariate normals.

- If $Y \sim N(\mu, \sigma^2)$, then $M_Y(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
- Moment-generating functions correspond uniquely to probability distributions.
- So *define* a normal random variable with expected value μ and variance σ^2 as a random variable with moment-generating function $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.
- This has one surprising consequence ...

Degenerate random variables

A *degenerate* random variable has all the probability concentrated at a single value, say $Pr\{Y = y_0\} = 1$. Then

$$\begin{aligned} M_Y(t) &= E(e^{Yt}) \\ &= \sum_y e^{yt} p(y) \\ &= e^{y_0 t} \cdot p(y_0) \\ &= e^{y_0 t} \cdot 1 \\ &= e^{y_0 t} \end{aligned}$$

If $Pr\{Y = y_0\} = 1$, then $M_Y(t) = e^{y_0 t}$

- This is of the form $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ with $\mu = y_0$ and $\sigma^2 = 0$.
- So $Y \sim N(y_0, 0)$.
- That is, degenerate random variables are "normal" with variance zero.
- Call them *singular* normals.
- This will be surprisingly handy later.

Independent standard normals

Let $Z_1, \ldots, Z_p \overset{i.i.d.}{\sim} N(0,1)$.

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix}$$

$$E(\mathbf{Z}) = \mathbf{0} \qquad cov(\mathbf{Z}) = \mathbf{I}_p$$

Moment-generating function of $\mathbf{Z}$

Using $e^{\mu t+\frac{1}{2}\sigma^2 t^2}$

$$
\begin{aligned}
M_{\mathbf{z}}(\mathbf{t}) &= \prod_{j=1}^{p} M_{Z_j}(t_j) \\
&= \prod_{j=1}^{p} e^{\frac{1}{2}t_j^2} \\
&= e^{\frac{1}{2}\sum_{j=1}^{p} t_j^2} \\
&= e^{\frac{1}{2}\mathbf{t}^\top \mathbf{t}}
\end{aligned}
$$

Transform $\mathbf{Z}$ to get a general multivariate normal

Remember: $\mathbf{A}$ non-negative definite means $\mathbf{v}^\top \mathbf{A}\mathbf{v} \geq 0$

Let $\mathbf{\Sigma}$ be a $p \times p$ symmetric *non-negative definite* matrix and $\boldsymbol{\mu} \in \mathbb{R}^p$. Let $\mathbf{Y} = \mathbf{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$.

- The elements of $\mathbf{Y}$ are linear combinations of independent standard normals.
- Linear combinations of normals should be normal.
- $\mathbf{Y}$ has a multivariate distribution.
- We'd like to call $\mathbf{Y}$ a *multivariate normal.*

Moment-generating function of $\mathbf{Y} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$

Remember: $M_{\mathbf{AX}}(\mathbf{t}) = M_{\mathbf{X}}(\mathbf{A}^\top\mathbf{t})$ and $M_{\mathbf{X}+\mathbf{c}}(\mathbf{t}) = e^{\mathbf{t}^\top\mathbf{c}} M_{\mathbf{X}}(\mathbf{t})$ and $M_{\mathbf{Z}}(\mathbf{t}) = e^{\frac{1}{2}\mathbf{t}^\top\mathbf{t}}$

$$
\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) &= M_{\mathbf{Y}=\boldsymbol{\Sigma}^{1/2}\mathbf{Z}+\boldsymbol{\mu}}(\mathbf{t}) \\
&= e^{\mathbf{t}^\top\boldsymbol{\mu}} M_{\boldsymbol{\Sigma}^{1/2}\mathbf{Z}}(\mathbf{t}) \\
&= e^{\mathbf{t}^\top\boldsymbol{\mu}} M_{\mathbf{Z}}(\boldsymbol{\Sigma}^{1/2\,\top}\mathbf{t}) \\
&= e^{\mathbf{t}^\top\boldsymbol{\mu}} M_{\mathbf{Z}}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}) \\
&= e^{\mathbf{t}^\top\boldsymbol{\mu}} e^{\frac{1}{2}(\boldsymbol{\Sigma}^{1/2}\mathbf{t})^\top(\boldsymbol{\Sigma}^{1/2}\mathbf{t})} \\
&= e^{\mathbf{t}^\top\boldsymbol{\mu}} e^{\frac{1}{2}\mathbf{t}^\top\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{t}} \\
&= e^{\mathbf{t}^\top\boldsymbol{\mu}} e^{\frac{1}{2}\mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t}}
\end{aligned}
$$

So *define* a multivariate normal random variable $\mathbf{Y}$ as one with moment-generating function $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^\top\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t}}$.

Compare univariate and multivariate normal moment-generating functions

Univariate $M_Y(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

Multivariate $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^\top \boldsymbol{\mu}} \, e^{\frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}}$

So the univariate normal is a special case of the multivariate normal with $p = 1$.

Mean and covariance matrix

For a univariate normal, $E(Y) = \mu$ and $Var(Y) = \sigma^2$

Recall $\mathbf{Y} = \mathbf{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$.

$$
\begin{aligned}
E(\mathbf{Y}) &= \boldsymbol{\mu} \\
cov(\mathbf{Y}) &= \mathbf{\Sigma}^{1/2} cov(\mathbf{Z}) \mathbf{\Sigma}^{1/2\top} \\
&= \mathbf{\Sigma}^{1/2}\,\mathbf{I}\,\mathbf{\Sigma}^{1/2} \\
&= \mathbf{\Sigma}
\end{aligned}
$$

We will say $\mathbf{Y}$ is multivariate normal with expected value $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{\Sigma}$, and write $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$.

Probability density function of $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Remember, $\boldsymbol{\Sigma}$ is only positive *semi*-definite.

It is easy to write down the density of $\mathbf{Z} \sim N_p(\mathbf{0}, \mathbf{I})$ as a product of standard normals.

If $\boldsymbol{\Sigma}$ is strictly positive definite (and not otherwise), the density of $\mathbf{Y} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$ can be obtained using the Jacobian Theorem as

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

This is usually how the multivariate normal is defined.

$\boldsymbol{\Sigma}$ positive definite?

- Positive definite means that for any non-zero $p \times 1$ vector $\mathbf{a}$, we have $\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} > 0$.
- Since the one-dimensional random variable $W = \sum_{i=1}^{p} a_i Y_i$ may be written as $W = \mathbf{a}^\top \mathbf{Y}$ and $Var(W) = cov(\mathbf{a}^\top \mathbf{Y}) = \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}$, it is natural to require that $\boldsymbol{\Sigma}$ be positive definite.
- All it means is that every non-zero linear combination of $\mathbf{Y}$ values has a positive variance. Often, this is what you want.

Singular normal: Σ is positive *semi*-definite.

Suppose there is $\mathbf{a} \neq \mathbf{0}$ with $\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} = 0$. Let $W = \mathbf{a}^\top \mathbf{Y}$.

- Then $Var(W) = Var(\mathbf{a}^\top \mathbf{Y}) = \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} = 0$. That is W has a degenerate distribution (but it's still still normal).
- In this case we describe the distribution of $\mathbf{Y}$ as a *singular* multivariate normal.
- Excluding the singular case creates a lot of extra work in later proofs.
- We will insist that a singular multivariate normal is still multivariate normal, even though it has no density.

Distribution of $\mathbf{AY}$

Recall $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}}$

Let $\mathbf{A}$ be an $r \times p$ matrix, and $\mathbf{W} = \mathbf{AY}$.

$$
\begin{aligned}
M_{\mathbf{W}}(\mathbf{t}) &= M_{\mathbf{AY}}(\mathbf{t}) \\
&= M_{\mathbf{Y}}(\mathbf{A}^\top \mathbf{t}) \\
&= e^{(\mathbf{A}^\top \mathbf{t})^\top \boldsymbol{\mu}} \, e^{\frac{1}{2}(\mathbf{A}^\top \mathbf{t})^\top \boldsymbol{\Sigma} (\mathbf{A}^\top \mathbf{t})} \\
&= e^{\mathbf{t}^\top (\mathbf{A}\boldsymbol{\mu})} \, e^{\frac{1}{2}\mathbf{t}^\top (\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)\mathbf{t}} \\
&= e^{\mathbf{t}^\top (\mathbf{A}\boldsymbol{\mu}) + \frac{1}{2}\mathbf{t}^\top (\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)\mathbf{t}}
\end{aligned}
$$

Recognize moment-generating function and conclude

$$
\mathbf{W} \sim N_r(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)
$$

Exercise

Use moment-generating functions, of course.

Let $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Show $\mathbf{Y} + \mathbf{c} \sim N_p(\boldsymbol{\mu} + \mathbf{c}, \boldsymbol{\Sigma})$.

Zero covariance implies independence for the multivariate normal.

- Independence always implies zero covariance.
- For the multivariate normal, zero covariance also implies independence.
- The multivariate normal is the only continuous distribution with this property.

Show zero covariance implies independence

By showing $M_{\mathbf{Y}}(\mathbf{t}) = M_{\mathbf{Y}_1}(\mathbf{t}_1)M_{\mathbf{Y}_2}(\mathbf{t}_2)$

Let $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\mathbf{Y} = \left(\frac{\mathbf{Y}_1}{\mathbf{Y}_2}\right) \quad \boldsymbol{\mu} = \left(\frac{\boldsymbol{\mu}_1}{\boldsymbol{\mu}_2}\right) \quad \boldsymbol{\Sigma} = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Sigma}_2 \end{array}\right) \quad \mathbf{t} = \left(\frac{\mathbf{t}_1}{\mathbf{t}_2}\right)$$

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= E\left(e^{\mathbf{t}^\top \mathbf{Y}}\right) \\ &= E\left(e^{(\mathbf{t}_1^\top|\mathbf{t}_2^\top)\mathbf{Y}}\right) \\ &= M_{\mathbf{Y}}\left((\mathbf{t}_1^\top|\mathbf{t}_2^\top)^\top\right) \\ &= \ldots \end{aligned}$$

Continuing the calculation: $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma}\mathbf{t}}$

$\mathbf{Y} = \left(\frac{\mathbf{Y}_1}{\mathbf{Y}_2}\right) \quad \boldsymbol{\mu} = \left(\frac{\boldsymbol{\mu}_1}{\boldsymbol{\mu}_2}\right) \quad \boldsymbol{\Sigma} = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Sigma}_2 \end{array}\right) \quad \mathbf{t} = \left(\frac{\mathbf{t}_1}{\mathbf{t}_2}\right)$

$$
\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) &= M_{\mathbf{Y}}\left((\mathbf{t}_1^\top | \mathbf{t}_2^\top)^\top\right) \\
&= \exp\left\{(\mathbf{t}_1^\top | \mathbf{t}_2^\top)\left(\frac{\boldsymbol{\mu}_1}{\boldsymbol{\mu}_2}\right)\right\} \exp\left\{\frac{1}{2}(\mathbf{t}_1^\top | \mathbf{t}_2^\top)\left(\begin{array}{c|c} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Sigma}_2 \end{array}\right)\left(\frac{\mathbf{t}_1}{\mathbf{t}_2}\right)\right\} \\
&= e^{\mathbf{t}_1^\top \boldsymbol{\mu}_1 + \mathbf{t}_2^\top \boldsymbol{\mu}_2} \exp\left\{\frac{1}{2}\left(\mathbf{t}_1^\top \boldsymbol{\Sigma}_1 | \mathbf{t}_2^\top \boldsymbol{\Sigma}_2\right)\left(\frac{\mathbf{t}_1}{\mathbf{t}_2}\right)\right\} \\
&= e^{\mathbf{t}_1^\top \boldsymbol{\mu}_1 + \mathbf{t}_2^\top \boldsymbol{\mu}_2} \exp\left\{\frac{1}{2}\left(\mathbf{t}_1^\top \boldsymbol{\Sigma}_1 \mathbf{t}_1 + \mathbf{t}_2^\top \boldsymbol{\Sigma}_2 \mathbf{t}_2\right)\right\} \\
&= e^{\mathbf{t}_1^\top \boldsymbol{\mu}_1} \; e^{\mathbf{t}_2^\top \boldsymbol{\mu}_2} \; e^{\frac{1}{2}(\mathbf{t}_1^\top \boldsymbol{\Sigma}_1 \mathbf{t}_1)} \; e^{\frac{1}{2}(\mathbf{t}_2^\top \boldsymbol{\Sigma}_2 \mathbf{t}_2)} \\
&= M_{\mathbf{Y}_1}(\mathbf{t}_1) M_{\mathbf{Y}_2}(\mathbf{t}_2)
\end{aligned}
$$

So $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent.

An easy example

If you do it the easy way

Let $Y_1 \sim N(1,2)$, $Y_2 \sim N(2,4)$ and $Y_3 \sim N(6,3)$ be independent, with $W_1 = Y_1 + Y_2$ and $W_2 = Y_2 + Y_3$. Find the joint distribution of W_1 and W_2.

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

$$\mathbf{W} = \mathbf{AY} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$$

$\mathbf{W} = \mathbf{AY} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$

$Y_1 \sim N(1,2)$, $Y_2 \sim N(2,4)$ and $Y_3 \sim N(6,3)$ are independent

$$\mathbf{A}\boldsymbol{\mu} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 3 \\ 8 \end{pmatrix}$$

$$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 4 \\ 4 & 7 \end{pmatrix}$$

Marginal distributions are multivariate normal

$\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, so $\mathbf{W} = \mathbf{AY} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$

Find the distribution of

$$\left(\begin{array}{cccc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right) \left(\begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{array} \right) = \left(\begin{array}{c} Y_2 \\ Y_4 \end{array} \right)$$

Bivariate normal. The expected value is easy.

Covariance matrix

$$
\begin{aligned}
cov(\mathbf{AY}) &= \mathbf{A\Sigma A}^\top \\
&= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \sigma_{3,4} \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \sigma_2^2 & \sigma_{2,4} \\ \sigma_{2,4} & \sigma_4^2 \end{pmatrix}
\end{aligned}
$$

Marginal distributions of a multivariate normal are multivariate normal, with the original means, variances and covariances.

Summary

- If $\mathbf{c}$ is a vector of constants, $\mathbf{X} + \mathbf{c} \sim N(\mathbf{c} + \boldsymbol{\mu}, \boldsymbol{\Sigma})$
- If $\mathbf{A}$ is a matrix of constants, $\mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$
- Linear combinations of multivariate normals are multivariate normal.
- All the marginals (dimension less than p) of $\mathbf{X}$ are (multivariate) normal, but it is possible in theory to have a collection of univariate normals whose joint distribution is not multivariate normal.
- For the multivariate normal, zero covariance implies independence. The multivariate normal is the only continuous distribution with this property.

Multivariate normal likelihood

For reference

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}} (2\pi)^{\frac{p}{2}}} \exp\left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Sigma}|^{-n/2} (2\pi)^{-np/2} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Sigma}|^{-n/2} (2\pi)^{-np/2} \exp -\frac{n}{2} \left\{ tr(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu}) \right\},
\end{aligned}
$$

where $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^\top$ is the sample variance-covariance matrix.

Showing $(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \sim \chi^2(p)$

$\boldsymbol{\Sigma}$ has to be positive definite this time

$$
\begin{aligned}
\mathbf{X} &\sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu} &\sim N(\mathbf{0}, \ \boldsymbol{\Sigma}) \\
\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y} &\sim N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right) \\
&= N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}} \ \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right) \\
&= N(\mathbf{0}, \mathbf{I})
\end{aligned}
$$

So $\mathbf{Z}$ is a vector of p independent standard normals, and

$$
\mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} = (\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y})^\top(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{Y}) = \mathbf{Z}'\mathbf{Z} = \sum_{j=1}^{p} Z_i^2 \sim \chi^2(p) \qquad \blacksquare
$$

$\overline{X}$ and S^2 independent

$X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N\left(\mu \mathbf{1}, \sigma^2 \mathbf{I}\right) \qquad \mathbf{Y} = \begin{pmatrix} X_1 - \overline{X} \\ \vdots \\ X_n - \overline{X} \\ \\ \overline{X} \end{pmatrix} = \mathbf{AX}$$

Note $\mathbf{A}$ is $(n+1) \times n$, so $cov(\mathbf{AY}) = \sigma^2 \mathbf{A}\mathbf{A}^\top$ is $(n+1) \times (n+1)$, singular.

The argument

$$\mathbf{Y} = \mathbf{AX} = \begin{pmatrix} X_1 - \overline{X} \\ \vdots \\ X_n - \overline{X} \\ \\ \overline{X} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_2 \\ \hline \\ \overline{X} \end{pmatrix}$$

- $\mathbf{Y}$ is multivariate normal.
- $Cov\left(\overline{X}, (X_j - \overline{X})\right) = 0$ (Exercise)
- So $\overline{X}$ and $\mathbf{Y}_2$ are independent.
- So $\overline{X}$ and $S^2 = g(\mathbf{Y}_2)$ are independent. ■

Leads to the t distribution

If

- $Z \sim N(0,1)$ and
- $Y \sim \chi^2(\nu)$ and
- Z and Y are independent, then

$$T = \frac{Z}{\sqrt{Y/\nu}} \sim t(\nu)$$

Random sample from a normal distribution

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Then

- $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim N(0,1)$ and
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ and
- These quantities are independent, so

$$\begin{aligned} T &= \frac{\sqrt{n}(\overline{X}-\mu)/\sigma}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} \\ &= \frac{\sqrt{n}(\overline{X}-\mu)}{S} \sim t(n-1) \end{aligned}$$

Copyright Information