The Multivariate Normal Distribution[1]

STA 302 Fall 2020

[1] See last slide for copyright information.

Overview

1. Joint Moment-generating Functions
2. Definition of the Multivariate Normal Distribution
3. Properties of the Multivariate Normal
4. χ^2 and t Distributions

Joint moment-generating function

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

- $M_{\mathbf{x}}(\mathbf{t}) \stackrel{def}{=} E\left(e^{\mathbf{t}'\mathbf{x}}\right)$ Compare $M_x(t) = E(e^{xt})$.
- For example,

$$\begin{aligned} M_{(x_1,x_2,x_3)}(t_1,t_2,t_3) &= E\left(e^{x_1t_1+x_2t_2+x_3t_3}\right) \\ &= \iiint e^{x_1t_1+x_2t_2+x_3t_3} f(x_1,x_2,x_3)\, dx_1 dx_2 dx_3 \end{aligned}$$

- Just write $M(\mathbf{t})$ if there is no ambiguity.

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

Uniqueness
Proof omitted

Joint moment-generating functions correspond uniquely to joint probability distributions.

- $M(\mathbf{t})$ is a function of $F(\mathbf{x})$.
  - Step One: $f(\mathbf{x}) = \frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_p} F(\mathbf{x})$.
  - For example, $\frac{\partial}{\partial x_1} \frac{\partial}{\partial x_2} \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(y_1, y_2)\, dy_1 dy_2$
  - Step Two: $M(\mathbf{t}) = \int \cdots \int e^{\mathbf{t}'\mathbf{x}} f(\mathbf{x})\, d\mathbf{x}$
  - Could write $M(\mathbf{t}) = g\left(F(\mathbf{x})\right)$.
- Uniqueness says the function g is one-to-one, so that $F(\mathbf{x}) = g^{-1}\left(M(\mathbf{t})\right)$.

$g^{-1}(M(\mathbf{t})) = F(\mathbf{x})$

A two-variable example

$$g^{-1}(M(\mathbf{t})) = F(\mathbf{x})$$

$$g^{-1}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{x_1 t_1 + x_2 t_2} f(x_1, x_2)\, dx_1 dx_2\right) = \int_{-\infty}^{x_2}\int_{-\infty}^{x_1} f(y_1, y_2)\, dy_1 dy_2$$

Theorem

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.

Proof

Two random vectors are independent if and only if the moment-generating function of their joint distribution is the product of their moment-generating functions.

Independence therefore the MGFs factor is an exercise.

$$\begin{aligned}
M_{x_1,x_2}(t_1,t_2) &= M_{x_1}(t_1)M_{x_2}(t_2) \\
&= \left(\int_{-\infty}^{\infty} e^{x_1t_1} f_{x_1}(x_1)\,dx_1\right)\left(\int_{-\infty}^{\infty} e^{x_2t_2} f_{x_2}(x_2)\,dx_2\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{x_1t_1} e^{x_2t_2} f_{x_1}(x_1) f_{x_2}(x_2)\,dx_1dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{x_1t_1+x_2t_2} f_{x_1}(x_1) f_{x_2}(x_2)\,dx_1dx_2
\end{aligned}$$

Proof continued

Have $M_{x_1,x_2}(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{x_1 t_1 + x_2 t_2} f_{x_1}(x_1) f_{x_2}(x_2)\, dx_1 dx_2$.

Using $F(\mathbf{x}) = g^{-1}(M(\mathbf{t}))$,

$$\begin{aligned} F(x_1,x_2) &= g^{-1}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{x_1 t_1 + x_2 t_2} f_{x_1}(x_1) f_{x_2}(x_2)\, dx_1 dx_2\right) \\ &= \int_{-\infty}^{x_2}\int_{-\infty}^{x_1} f_{x_1}(y_1) f_{x_2}(y_2)\, dy_1 dy_2 \\ &= \int_{-\infty}^{x_2} f_{x_2}(y_2)\left(\int_{-\infty}^{x_1} f_{x_1}(y_1)\, dy_1\right) dy_2 \\ &= \int_{-\infty}^{x_2} f_{x_2}(y_2) F_{x_1}(x_1)\, dy_2 \\ &= F_{x_1}(x_1)\int_{-\infty}^{x_2} f_{x_2}(y_2)\, dy_2 \\ &= F_{x_1}(x_1)\, F_{x_2}(x_2) \end{aligned}$$

So that x_1 and x_2 are independent. ■

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 $\mathbf{y} = \left(\frac{\mathbf{y}_1}{\mathbf{y}_2}\right) = \left(\frac{g_1(\mathbf{x}_1)}{g_2(\mathbf{x}_2)}\right)$ and $\mathbf{t} = \left(\frac{\mathbf{t}_1}{\mathbf{t}_2}\right)$. Then

$$
\begin{aligned}
M_{\mathbf{y}}(\mathbf{t}) &= E\left(e^{\mathbf{t}'\mathbf{y}}\right) = E\left(e^{(\mathbf{t}_1'|\mathbf{t}_2')\left(\frac{\mathbf{y}_1}{\mathbf{y}_2}\right)}\right) \\
&= E\left(e^{\mathbf{t}_1'\mathbf{y}_1+\mathbf{t}_2'\mathbf{y}_2}\right) = E\left(e^{\mathbf{t}_1'\mathbf{y}_1}e^{\mathbf{t}_2'\mathbf{y}_2}\right) = E\left(e^{\mathbf{t}_1'g_1(\mathbf{x}_1)}e^{\mathbf{t}_2'g_2(\mathbf{x}_2)}\right) \\
&= \int\int e^{\mathbf{t}_1'g_1(\mathbf{x}_1)}e^{\mathbf{t}_2'g_2(\mathbf{x}_2)} f_{\mathbf{x}_1}(\mathbf{x}_1) f_{\mathbf{x}_2}(\mathbf{x}_2)\, d\mathbf{x}_1 d\mathbf{x}_2 \\
&= \int e^{\mathbf{t}_2'g_2(\mathbf{x}_2)} f_{\mathbf{x}_2}(\mathbf{x}_2)\left(\int e^{\mathbf{t}_1'g_1(\mathbf{x}_1)} f_{\mathbf{x}_1}(\mathbf{x}_1)\, d\mathbf{x}_1\right) d\mathbf{x}_2 \\
&= \int e^{\mathbf{t}_2'g_2(\mathbf{x}_2)} f_{\mathbf{x}_2}(\mathbf{x}_2) M_{g_1(\mathbf{x}_1)}(\mathbf{t}_1) d\mathbf{x}_2 \\
&= M_{g_1(\mathbf{x}_1)}(\mathbf{t}_1) M_{g_2(\mathbf{x}_2)}(\mathbf{t}_2) = M_{\mathbf{y}_1}(\mathbf{t}_1) M_{\mathbf{y}_2}(\mathbf{t}_2)
\end{aligned}
$$

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

$M_{\mathbf{Ax}}(\mathbf{t}) = M_{\mathbf{x}}(\mathbf{A}'\mathbf{t})$

Analogue of $M_{ax}(t) = M_x(at)$

Recalling $M_{\mathbf{x}}(\mathbf{t}) \stackrel{def}{=} E\left(e^{\mathbf{t}'\mathbf{x}}\right)$,

$$\begin{aligned} M_{\mathbf{Ax}}(\mathbf{t}) &= E\left(e^{\mathbf{t}'\mathbf{Ax}}\right) \\ &= E\left(e^{\left(\mathbf{A}'\mathbf{t}\right)'\mathbf{x}}\right) \\ &= M_{\mathbf{x}}(\mathbf{A}'\mathbf{t}) \end{aligned}$$

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

$M_{\mathbf{x}+\mathbf{c}}(\mathbf{t}) = e^{\mathbf{t}'\mathbf{c}} M_{\mathbf{x}}(\mathbf{t})$

Analogue of $M_{x+c}(t) = e^{ct} M_x(t)$

$$
\begin{aligned}
M_{\mathbf{x}+\mathbf{c}}(\mathbf{t}) &= E\left(e^{\mathbf{t}'(\mathbf{x}+\mathbf{c})}\right) \\
&= E\left(e^{\mathbf{t}'\mathbf{x}+\mathbf{t}'\mathbf{c}}\right) \\
&= e^{\mathbf{t}'\mathbf{c}} \, E\left(e^{\mathbf{t}'\mathbf{x}}\right) \\
&= e^{\mathbf{t}'\mathbf{c}} M_{\mathbf{x}}(\mathbf{t})
\end{aligned}
$$

Definition of the Multivariate Normal Distribution

Not in the text.

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

$$
\begin{aligned}
M_y(t) &= E(e^{yt}) \\
&= \sum_{\{y:\ p(y)>0\}} e^{yt} p(y) \\
&= e^{y_0 t} \cdot p(y_0) \\
&= e^{y_0 t} \cdot 1 \\
&= e^{y_0 t}
\end{aligned}
$$

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 $z_1, \ldots, z_p \overset{i.i.d.}{\sim} N(0, 1)$.

$$\mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_p \end{pmatrix}$$

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

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

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

$$
\begin{aligned}
M_{\mathbf{z}}(\mathbf{t}) &= \prod_{j=1}^{p} M_{z_j}(t_j) \\
&= \prod_{j=1}^{p} e^{\frac{1}{2}t_j^2} \\
&= e^{\frac{1}{2}\sum_{j=1}^{p} t_j^2} \\
&= e^{\frac{1}{2}\mathbf{t}'\mathbf{t}}
\end{aligned}
$$

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

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

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 independent normals are 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: $M_{\mathbf{Ax}}(\mathbf{t}) = M_{\mathbf{x}}(\mathbf{A}'\mathbf{t})$ and $M_{\mathbf{x}+\mathbf{c}}(\mathbf{t}) = e^{\mathbf{t}'\mathbf{c}} M_{\mathbf{x}}(\mathbf{t})$ and $M_{\mathbf{z}}(\mathbf{t}) = e^{\frac{1}{2}\mathbf{t}'\mathbf{t}}$

$$
\begin{aligned}
M_{\mathbf{y}}(\mathbf{t}) &= M_{\boldsymbol{\Sigma}^{1/2}\mathbf{z}+\boldsymbol{\mu}}(\mathbf{t}) \\
&= e^{\mathbf{t}'\boldsymbol{\mu}} \, M_{\boldsymbol{\Sigma}^{1/2}\mathbf{z}}(\mathbf{t}) \\
&= e^{\mathbf{t}'\boldsymbol{\mu}} \, M_{\mathbf{z}}(\boldsymbol{\Sigma}^{1/2\prime}\mathbf{t}) \\
&= e^{\mathbf{t}'\boldsymbol{\mu}} \, M_{\mathbf{z}}(\boldsymbol{\Sigma}^{1/2}\mathbf{t}) \\
&= e^{\mathbf{t}'\boldsymbol{\mu}} \, e^{\frac{1}{2}(\boldsymbol{\Sigma}^{1/2}\mathbf{t})'(\boldsymbol{\Sigma}^{1/2}\mathbf{t})} \\
&= e^{\mathbf{t}'\boldsymbol{\mu}} \, e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{t}} \\
&= e^{\mathbf{t}'\boldsymbol{\mu}} \, e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} \\
&= e^{\mathbf{t}'\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}
\end{aligned}
$$

So *define* a multivariate normal random variable $\mathbf{y}$ as one with moment-generating function $M_{\mathbf{y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$.

Compare univariate and multivariate normal moment-generating functions

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

Multivariate $M_{\mathbf{y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$

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 $\mathbf{y} = \mathbf{\Sigma}^{1/2}\mathbf{z} + \boldsymbol{\mu}$.

$$
\begin{aligned}
E(\mathbf{y}) &= \boldsymbol{\mu} \\
cov(\mathbf{y}) &= \mathbf{\Sigma}^{1/2} cov(\mathbf{z}) \mathbf{\Sigma}^{1/2\prime} \\
&= \mathbf{\Sigma}^{1/2}\,\mathbf{I}\,\mathbf{\Sigma}^{1/2} \\
&= \mathbf{\Sigma}
\end{aligned}
$$

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

Note that because $M_{\mathbf{y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}$, $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ completely determine the distribution.

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

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

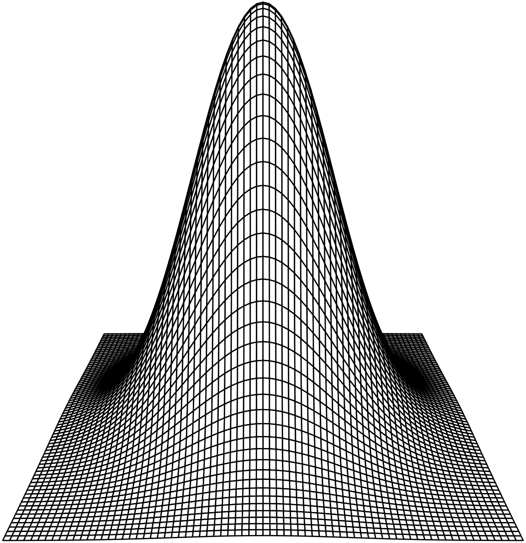
It is easy to write down the density of $\mathbf{z} \sim N_p(\mathbf{0}, I)$ as a product of standard normals.

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

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

This is usually how the multivariate normal is defined.

Bivariate Normal Density ($p = 2$)



$\boldsymbol{\Sigma}$ positive definite?

- Positive definite means that for any non-zero $p \times 1$ vector $\mathbf{a}$, we have $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} > 0$.
- Since the one-dimensional random variable $w = \sum_{i=1}^{p} a_i y_i$ may be written as $w = \mathbf{a}'\mathbf{y}$ and $Var(w) = cov(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$, 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: $\boldsymbol{\Sigma}$ is positive *semi*-definite.

Or "non-negative definite"

Suppose there is $\mathbf{a} \neq \mathbf{0}$ with $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = 0$. Let $w = \mathbf{a}'\mathbf{y}$.

- Then $Var(w) = cov(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = 0$. 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.
- Including the singular case saves 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 $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means $M_{\mathbf{y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$

Let $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\mathbf{w} = \mathbf{Ay}$, where $\mathbf{A}$ is an $r \times p$ matrix.

$$\begin{aligned} M_{\mathbf{w}}(\mathbf{t}) &= M_{\mathbf{Ay}}(\mathbf{t}) \\ &= M_{\mathbf{y}}(\mathbf{A}'\mathbf{t}) \\ &= e^{(\mathbf{A}'\mathbf{t})'\boldsymbol{\mu}} \, e^{\frac{1}{2}(\mathbf{A}'\mathbf{t})'\boldsymbol{\Sigma}(\mathbf{A}'\mathbf{t})} \\ &= e^{\mathbf{t}'(A\boldsymbol{\mu})} \, e^{\frac{1}{2}\mathbf{t}'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')\mathbf{t}} \\ &= e^{\mathbf{t}'(\mathbf{A}\boldsymbol{\mu}) + \frac{1}{2}\mathbf{t}'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')\mathbf{t}} \end{aligned}$$

Recognize moment-generating function and conclude

$$\mathbf{w} \sim N_r(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$$

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

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

$$\begin{aligned} M_{\mathbf{y}}(\mathbf{t}) &= E\left(e^{\mathbf{t}'\mathbf{y}}\right) \\ &= E\left(e^{\left(\begin{array}{c} \mathbf{t}_1 \\ \hline \mathbf{t}_2 \end{array} \right)' \mathbf{y}} \right) \\ &= \ldots \end{aligned}$$

Continuing the calculation: $M_{\mathbf{y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu}+\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$

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

$$
\begin{aligned}
M_{\mathbf{y}}(\mathbf{t}) &= E\left(e^{\left(\frac{\mathbf{t}_1}{\mathbf{t}_2}\right)' \mathbf{y}} \right) \\
&= \exp\left\{ (\mathbf{t}_1' | \mathbf{t}_2') \left(\frac{\boldsymbol{\mu}_1}{\boldsymbol{\mu}_2}\right) \right\} \exp\left\{ \frac{1}{2} (\mathbf{t}_1' | \mathbf{t}_2') \left(\begin{array}{c|c} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Sigma}_2 \end{array}\right) \left(\frac{\mathbf{t}_1}{\mathbf{t}_2}\right) \right\} \\
&= e^{\mathbf{t}_1'\boldsymbol{\mu}_1 + \mathbf{t}_2'\boldsymbol{\mu}_2} \exp\left\{ \frac{1}{2} \left(\mathbf{t}_1'\boldsymbol{\Sigma}_1 | \mathbf{t}_2'\boldsymbol{\Sigma}_2 \right) \left(\frac{\mathbf{t}_1}{\mathbf{t}_2}\right) \right\} \\
&= e^{\mathbf{t}_1'\boldsymbol{\mu}_1 + \mathbf{t}_2'\boldsymbol{\mu}_2} \exp\left\{ \frac{1}{2} \left(\mathbf{t}_1'\boldsymbol{\Sigma}_1\mathbf{t}_1 + \mathbf{t}_2'\boldsymbol{\Sigma}_2\mathbf{t}_2 \right) \right\} \\
&= e^{\mathbf{t}_1'\boldsymbol{\mu}_1} \; e^{\mathbf{t}_2'\boldsymbol{\mu}_2} \; e^{\frac{1}{2}(\mathbf{t}_1'\boldsymbol{\Sigma}_1\mathbf{t}_1)} \; e^{\frac{1}{2}(\mathbf{t}_2'\boldsymbol{\Sigma}_2\mathbf{t}_2)} \\
&= e^{\mathbf{t}_1'\boldsymbol{\mu}_1 + \frac{1}{2}(\mathbf{t}_1'\boldsymbol{\Sigma}_1\mathbf{t}_1)} \; e^{\mathbf{t}_2'\boldsymbol{\mu}_2 + \frac{1}{2}(\mathbf{t}_2'\boldsymbol{\Sigma}_2\mathbf{t}_2)} \\
&= M_{\mathbf{y}_1}(\mathbf{t}_1) M_{\mathbf{y}_2}(\mathbf{t}_2)
\end{aligned}
$$

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.

$$\left(\begin{array}{c} w_1 \\ w_2 \end{array} \right) = \left(\begin{array}{ccc} 1 & 1 & 0 \\ 0 & 1 & 1 \end{array} \right) \left(\begin{array}{c} y_1 \\ y_2 \\ y_3 \end{array} \right)$$

$\mathbf{w} = \mathbf{Ay} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

$\mathbf{w} = \mathbf{Ay} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A}')$

$y_1 \sim N(1, 2)$, $y_2 \sim N(2, 4)$ and $y_3 \sim N(6, 3)$ are independent

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

$$\begin{aligned} \mathbf{A\Sigma A}' &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 6 & 4 \\ 4 & 7 \end{pmatrix} \end{aligned}$$

Marginal distributions are multivariate normal

$\mathbf{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, so $\mathbf{w} = \mathbf{A}\mathbf{y} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$

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

Of $\mathbf{Ay}$

$$
\begin{aligned}
cov(\mathbf{Ay}) &= \mathbf{A\Sigma A}' \\
&= \left(\begin{array}{cccc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \left(\begin{array}{cccc} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \sigma_{3,4} \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 \end{array}\right) \left(\begin{array}{cc} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{array}\right) \\
&= \left(\begin{array}{cccc} \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} \\ \sigma_{1,4} & \sigma_{2,4} & \sigma_{3,4} & \sigma_4^2 \end{array}\right) \left(\begin{array}{cc} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{array}\right) \\
&= \left(\begin{array}{cc} \sigma_2^2 & \sigma_{2,4} \\ \sigma_{2,4} & \sigma_4^2 \end{array}\right)
\end{aligned}
$$

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, $\mathbf{Ax} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.
- 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.

χ^2 and t Distributions

Need the multivariate normal for this.

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

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

$$\begin{aligned}
\mathbf{x} &\sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\mathbf{y} = \mathbf{x}-\boldsymbol{\mu} &\sim N(\mathbf{0},\ \boldsymbol{\Sigma}) \\
\mathbf{z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{y} &\sim N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right) \\
&= N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}\ \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}}\right) \\
&= N(\mathbf{0}, \mathbf{I}_p)
\end{aligned}$$

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

$$\begin{aligned}
w &= (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \\
&= \mathbf{y}'\boldsymbol{\Sigma}^{-1}\mathbf{y} = \mathbf{y}'\boldsymbol{\Sigma}^{-1/2\prime}\boldsymbol{\Sigma}^{-1/2}\mathbf{y} \\
&= (\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{y})'(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{y}) \\
&= \mathbf{z}'\mathbf{z} \\
&= \sum_{j=1}^{p} z_i^2 \sim \chi^2(p) \qquad \blacksquare
\end{aligned}$$

Show $\overline{x}$ and s^2 independent

$x_1, \ldots, x_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \sim N\left(\mu\mathbf{j}, \sigma^2\mathbf{I}\right) \qquad \mathbf{y} = \begin{pmatrix} x_1 - \overline{x} \\ \vdots \\ x_n - \overline{x} \\ \\ \overline{x} \end{pmatrix} = \mathbf{Ax}$$

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

The argument

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \left(\begin{array}{c} x_1 - \overline{x} \\ \vdots \\ x_n - \overline{x} \\ \\ \overline{x} \end{array} \right) = \left(\begin{array}{c} \mathbf{y}_2 \\ \hline \\ \overline{x} \end{array} \right)$$

- $\mathbf{y}$ is multivariate normal because $\mathbf{x}$ 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. $\blacksquare$

Leads to the t distribution

If

- $z \sim N(0,1)$ and
- $y \sim \chi^2(\nu)$ and
- z and y are independent, then we say

$$t = \frac{z}{\sqrt{y/\nu}} \sim t(\nu)$$

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