The Multivariate Normal Distribution

Defn: $Z \in \mathbb{R}^1 \sim N(0,1)$ iff

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} .$$

Defn: $\mathbf{Z} \in \mathbb{R}^p \sim MVN_p(0,I)$ if and only if $\mathbf{Z} = (Z_1,\ldots,Z_p)^T$ with the Z_i independent and each $Z_i \sim N(0,1)$.

In this case according to our theorem

$$\begin{aligned} f_{\mathbf{Z}}(z_1,\ldots,z_p) &= \prod \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} \\ &= (2\pi)^{-p/2} \exp\{-z^T z/2\} ; \end{aligned}$$

superscript t denotes matrix transpose.

Defn: $\mathbf{X} \in \mathbb{R}^p$ has a multivariate normal distribution if it has the same distribution as $\mathbf{AZ}+\boldsymbol{\mu}$ for some $\boldsymbol{\mu} \in \mathbb{R}^p$, some $p \times q$ matrix of constants $\mathbf{A}$ and $Z \sim MVN_q(0,I)$.

$p = q$, $\mathbf{A}$ singular: $\mathbf{X}$ does not have a density.

$\mathbf{A}$ invertible: derive multivariate normal density by change of variables:

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu} \Leftrightarrow \mathbf{Z} = \mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\mu})$$

$$\frac{\partial \mathbf{X}}{\partial \mathbf{Z}} = \mathbf{A} \qquad \frac{\partial \mathbf{Z}}{\partial \mathbf{X}} = \mathbf{A}^{-1}\,.$$

So

$$\begin{aligned} f_{\mathbf{X}}(x) &= f_{\mathbf{Z}}(\mathbf{A}^{-1}(x - \boldsymbol{\mu}))|\det(\mathbf{A}^{-1})| \\ &= \frac{\exp\{-(x-\boldsymbol{\mu})^T(\mathbf{A}^{-1})^T\mathbf{A}^{-1}(x-\boldsymbol{\mu})/2\}}{(2\pi)^{p/2}|\det \mathbf{A}|}\,. \end{aligned}$$

Now define $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$ and notice that

$$\boldsymbol{\Sigma}^{-1} = (\mathbf{A}^T)^{-1}\mathbf{A}^{-1} = (\mathbf{A}^{-1})^T\mathbf{A}^{-1}$$

and

$$\det \boldsymbol{\Sigma} = \det \mathbf{A} \det \mathbf{A}^T = (\det \mathbf{A})^2\,.$$

Thus $f_{\mathbf{X}}$ is

$$\frac{\exp\{-(x-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(x-\boldsymbol{\mu})/2\}}{(2\pi)^{p/2}(\det \boldsymbol{\Sigma})^{1/2}}\,;$$

the $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ density. Note density is the same for all $\mathbf{A}$ such that $\mathbf{A}\mathbf{A}^T = \boldsymbol{\Sigma}$. This justifies the notation $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

For which $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ is this a density?

Any $\boldsymbol{\mu}$ but if $x \in \mathbb{R}^p$ then

$$\begin{aligned} x^T \boldsymbol{\Sigma} x &= x^T \mathbf{A}\mathbf{A}^T x \\ &= (\mathbf{A}^T x)^T (\mathbf{A}^T x) \\ &= \sum_1^p y_i^2 \geq 0 \end{aligned}$$

where $y = \mathbf{A}^T x$. Inequality strict except for $y = 0$ which is equivalent to $x = 0$. Thus $\boldsymbol{\Sigma}$ is a positive definite symmetric matrix.

Conversely, if $\boldsymbol{\Sigma}$ is a positive definite symmetric matrix then there is a square invertible matrix $\mathbf{A}$ such that $\mathbf{A}\mathbf{A}^T = \boldsymbol{\Sigma}$ so that there is a $MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. ($\mathbf{A}$ can be found via the Cholesky decomposition, e.g.)

When $\mathbf{A}$ is singular $\mathbf{X}$ will not have a density: $\exists a$ such that $P(a^T \mathbf{X} = a^T \boldsymbol{\mu}) = 1$; $\mathbf{X}$ is confined to a hyperplane.

Still true: distribution of $\mathbf{X}$ depends only on $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^T$: if $\mathbf{A}\mathbf{A}^T = BB^T$ then $\mathbf{A}\mathbf{Z} + \boldsymbol{\mu}$ and $B\mathbf{Z} + \boldsymbol{\mu}$ have the same distribution.

Expectation, moments

Defn: If $\mathbf{X} \in \mathbb{R}^p$ has density f then

$$\mathsf{E}(g(\mathbf{X})) = \int g(x) f(x)\, dx\,.$$

any g from $\mathbb{R}^p$ to $\mathbb{R}$.

FACT: if $Y = g(X)$ for a smooth g (mapping $\mathbb{R} \to \mathbb{R}$)

$$\begin{aligned}
\mathsf{E}(Y) &= \int y f_Y(y)\, dy \\
&= \int g(x) f_Y(g(x)) g'(x)\, dx \\
&= \mathsf{E}(g(X))
\end{aligned}$$

by change of variables formula for integration. This is good because otherwise we might have two different values for $\mathsf{E}(e^X)$.

Linearity: $\mathsf{E}(aX + bY) = a\mathsf{E}(X) + b\mathsf{E}(Y)$ for real X and Y.

Defn: The r^{th} moment (about the origin) of a real rv X is $\mu_r' = \mathsf{E}(X^r)$ (provided it exists). We generally use μ for $\mathsf{E}(X)$.

Defn: The r^{th} central moment is

$$\mu_r = \mathsf{E}[(X-\mu)^r]$$

We call $\sigma^2 = \mu_2$ the variance.

Defn: For an $\mathbb{R}^p$ valued random vector $\mathbf{X}$

$$\boldsymbol{\mu}_{\mathbf{X}} = \mathsf{E}(\mathbf{X})$$

is the vector whose i^{th} entry is $\mathsf{E}(X_i)$ (provided all entries exist).

Fact: same idea used for random matrices.

Defn: The $(p \times p)$ variance covariance matrix of $\mathbf{X}$ is

$$\mathsf{Var}(\mathbf{X}) = \mathsf{E}\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\right]$$

which exists provided each component X_i has a finite second moment.

Example moments: If $Z \sim N(0,1)$ then

$$\begin{aligned} \mathsf{E}(Z) &= \int_{-\infty}^{\infty} z e^{-z^2/2} dz/\sqrt{2\pi} \\ &= \left. \frac{-e^{-z^2/2}}{\sqrt{2\pi}} \right|_{-\infty}^{\infty} \\ &= 0 \end{aligned}$$

and (integrating by parts)

$$\begin{aligned} \mathsf{E}(Z^r) &= \int_{-\infty}^{\infty} z^r e^{-z^2/2} dz/\sqrt{2\pi} \\ &= \left. \frac{-z^{r-1} e^{-z^2/2}}{\sqrt{2\pi}} \right|_{-\infty}^{\infty} \\ &\quad + (r-1) \int_{-\infty}^{\infty} z^{r-2} e^{-z^2/2} dz/\sqrt{2\pi} \end{aligned}$$

so that

$$\mu_r = (r-1)\mu_{r-2}$$

for $r \geq 2$. Remembering that $\mu_1 = 0$ and

$$\mu_0 = \int_{-\infty}^{\infty} z^0 e^{-z^2/2} dz/\sqrt{2\pi} = 1$$

we find that

$$\mu_r = \begin{cases} 0 & r \text{ odd} \\ (r-1)(r-3)\cdots 1 & r \text{ even}. \end{cases}$$

If now $X \sim N(\mu, \sigma^2)$, that is, $X \sim \sigma Z + \mu$, then $\mathsf{E}(X) = \sigma \mathsf{E}(Z) + \mu = \mu$ and

$$\mu_r(X) = \mathsf{E}[(X-\mu)^r] = \sigma^r \mathsf{E}(Z^r)$$

In particular, we see that our choice of notation $N(\mu, \sigma^2)$ for the distribution of $\sigma Z + \mu$ is justified; σ is indeed the variance.

Similarly for $\mathbf{X} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ we have $\mathbf{X} = \mathbf{AZ} + \mu$ with $\mathbf{Z} \sim MVN(0, I)$ and

$$\mathsf{E}(\mathbf{X}) = \boldsymbol{\mu}$$

and

$$\begin{aligned} \mathsf{Var}(\mathbf{X}) &= \mathsf{E}\left\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\right\} \\ &= \mathsf{E}\left\{\mathbf{AZ}(\mathbf{AZ})^T\right\} \\ &= \mathbf{A}\mathsf{E}(\mathbf{ZZ}^T)\mathbf{A}^T \\ &= \mathbf{A}I\mathbf{A}^T = \boldsymbol{\Sigma}\,. \end{aligned}$$

Note use of easy calculation: $\mathsf{E}(\mathbf{Z}) = 0$ and

$$\mathsf{Var}(\mathbf{Z}) = \mathsf{E}(\mathbf{ZZ}^T) = I\,.$$

Moments and independence

Theorem: If $X_1, \ldots, X_p$ are independent and each X_i is integrable then $X = X_1 \cdots X_p$ is integrable and

$$\mathsf{E}(X_1 \cdots X_p) = \mathsf{E}(X_1) \cdots \mathsf{E}(X_p)\,.$$

Moment Generating Functions

Defn: The moment generating function of a real valued X is

$$M_X(t) = \mathsf{E}(e^{tX})$$

defined for those real t for which the expected value is finite.

Defn: The moment generating function of $\mathbf{X} \in \mathbb{R}^p$ is

$$M_{\mathbf{X}}(u) = \mathsf{E}[e^{u^T \mathbf{X}}]$$

defined for those vectors u for which the expected value is finite.

Example: If $Z \sim N(0,1)$ then

$$\begin{aligned} M_Z(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(z-t)^2/2+t^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2+t^2/2} du \\ &= e^{t^2/2} \end{aligned}$$

Theorem: ($p = 1$) If M is finite for all t in a neighbourhood of $\mathbf{0}$ then

1. Every moment of X is finite.
2. M is C^∞ (in fact M is analytic).
3. $\mu'_k = \frac{d^k}{dt^k} M_X(0)$.

Note: C^∞ means has continuous derivatives of all orders. Analytic means has convergent power series expansion in neighbourhood of each $t \in (-\epsilon, \epsilon)$.

The proof, and many other facts about mgfs, rely on techniques of complex variables.

Characterization & MGFs

Theorem: Suppose $\mathbf{X}$ and $\mathbf{Y}$ are $\mathbb{R}^p$ valued random vectors such that

$$M_{\mathbf{X}}(\mathbf{u}) = M_{\mathbf{Y}}(\mathbf{u})$$

for $\mathbf{u}$ in some open neighbourhood of $\mathbf{0}$ in $\mathbb{R}^p$. Then $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution.

The proof relies on techniques of complex variables.

MGFs and Sums

If $X_1, \ldots, X_p$ are independent and $Y = \sum X_i$ then mgf of Y is product mgfs of individual X_i:

$$\mathsf{E}(e^{tY}) = \prod_i \mathsf{E}(e^{tX_i})$$

or $M_Y = \prod M_{X_i}$. (Also for multivariate X_i.)

Example: If $Z_1, \ldots, Z_p$ are independent $N(0,1)$ then

$$\begin{aligned}\mathsf{E}(e^{\sum a_i Z_i}) &= \prod_i \mathsf{E}(e^{a_i Z_i}) \\ &= \prod_i e^{a_i^2/2} \\ &= \exp(\sum a_i^2/2)\end{aligned}$$

Conclusion: If $\mathbf{Z} \sim MNV_p(0, I)$ then

$$M_Z(u) = \exp(\sum u_i^2/2) = \exp(\mathbf{u}^T\mathbf{u}/2).$$

Example: If $X \sim N(\mu, \sigma^2)$ then $X = \sigma Z + \mu$ and

$$M_X(t) = \mathsf{E}(e^{t(\sigma Z + \mu)}) = e^{t\mu} e^{\sigma^2 t^2/2}.$$

Theorem: Suppose $\mathbf{X} = \mathbf{A}\mathbf{Z} + \boldsymbol{\mu}$ and $\mathbf{Y} = \mathbf{A}^*\mathbf{Z}^* + \boldsymbol{\mu}^*$ where $\mathbf{Z} \sim MVN_p(0, I)$ and $\mathbf{Z}^* \sim MVN_q(0, I)$. Then $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution if and only iff the following two conditions hold:

1. $\boldsymbol{\mu} = \boldsymbol{\mu}^*$.
2. $\mathbf{A}\mathbf{A}^T = \mathbf{A}^*(\mathbf{A}^*)^T$.

Alternatively: if $\mathbf{X}$, $\mathbf{Y}$ each MVN then $\text{E}(\mathbf{X}) = \text{E}(\mathbf{Y})$ and $\text{Var}(\mathbf{X}) = \text{Var}(\mathbf{Y})$ imply that $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution.

Proof: If 1 and 2 hold the mgf of $\mathbf{X}$ is

$$\begin{aligned}
\text{E}\left(e^{t^T\mathbf{X}}\right) &= \text{E}\left(e^{t^T(\mathbf{A}\mathbf{Z}+\boldsymbol{\mu}}\right) \\
&= e^{t^T\boldsymbol{\mu}}\text{E}\left(e^{(\mathbf{A}^Tt)^T\mathbf{Z}}\right) \\
&= e^{t^T\boldsymbol{\mu}+(\mathbf{A}^Tt)^T(\mathbf{A}^Tt)} \\
&= e^{t^T\boldsymbol{\mu}+t^T\boldsymbol{\Sigma}t}
\end{aligned}$$

Thus $M_{\mathbf{X}} = M_{\mathbf{Y}}$. Conversely if $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution then they have the same mean and variance.

Thus mgf is determined by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Theorem: If $\mathbf{X} \sim MVN_p(\mu, \boldsymbol{\Sigma})$ then there is $\mathbf{A}$ a $p \times p$ matrix such that $\mathbf{X}$ has same distribution as $\mathbf{AZ} + \boldsymbol{\mu}$ for $\mathbf{Z} \sim MVN_p(0, I)$.

We may assume that $\mathbf{A}$ is symmetric and non-negative definite, or that $\mathbf{A}$ is upper triangular, or that Ba is lower triangular.

Proof: Pick any $\mathbf{A}$ such that $\mathbf{AA}^T = \boldsymbol{\Sigma}$ such as $\mathbf{PD}^{1/2}\mathbf{P}^T$ from the spectral decomposition. Then $\mathbf{AZ} + \boldsymbol{\mu} \sim MVN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

From the symmetric square root can produce an upper triangular square root by the Gram Schmidt process: if $\mathbf{A}$ has rows $a_1^T, \ldots, a_p^T$ then let v_p be $a_p/\sqrt{a_p^T a_p}$. Choose v_{p-1} proportional to $a_{p-1} - bv_p$ where $b = a_{p-1}^T v_p$ so that v_{p-1} has unit length. Continue in this way; you automatically get $a_j^T v_k = 0$ if $j < k$. If $\mathbf{P}$ has columns $v_1, \ldots, v_p$ then $\mathbf{P}$ is orthogonal and $\mathbf{AP}$ is an upper triangular square root of $\boldsymbol{\Sigma}$.

Variances, Covariances, Correlations

Defn: The covariance between $\mathbf{X}$ and $\mathbf{Y}$ is

$$\text{Cov}(\mathbf{X},\mathbf{Y}) = \text{E}\left\{(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{Y}-\boldsymbol{\mu}_{\mathbf{Y}})^T\right\}$$

This is a matrix.

Properties:

- $\text{Cov}(\mathbf{X},\mathbf{X}) = \text{Var}(\mathbf{X})$.

- Cov is bilinear:

$$\begin{aligned}\text{Cov}(\mathbf{AX}+\mathbf{BW},\mathbf{Y}) =& \mathbf{A}\text{Cov}(\mathbf{X},\mathbf{Y}) \\ &+ \mathbf{B}\text{Cov}(\mathbf{W},\mathbf{Y})\end{aligned}$$

  and

$$\begin{aligned}\text{Cov}(\mathbf{X},\mathbf{CY}+\mathbf{DZ}) =& \text{Cov}(\mathbf{X},\mathbf{Y})\mathbf{C}^T \\ &+ \text{Cov}(\mathbf{X},\mathbf{Z})\mathbf{D}^T\end{aligned}$$

Properties of the MVN distribution

1: All margins are multivariate normal: if

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

and

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}$$

then $\mathbf{X} \sim MVN(\boldsymbol{\mu}, \mathbf{\Sigma}) \Rightarrow \mathbf{X}_1 \sim MVN(\boldsymbol{\mu}_1, \mathbf{\Sigma}_{11})$.

2: $\mathbf{MX} + \boldsymbol{\nu} \sim MVN(\mathrm{M}\boldsymbol{\mu} + \boldsymbol{\nu}, \mathrm{M}\Sigma\mathrm{M}^T)$: affine transformation of MVN is normal.

3: If

$$\mathbf{\Sigma}_{12} = \mathrm{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \mathbf{0}$$

then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent.

4: All conditionals are normal: the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = x_2$ is $MVN(\boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(x_2 - \boldsymbol{\mu}_2), \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})$

Proof of ($\mathbf{1}$): If $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$ then

$$\mathbf{X}_1 = [I|\mathbf{0}]\,\mathbf{X}$$

for I the identity matrix of correct dimension.

So

$$\mathbf{X}_1 = ([I|\mathbf{0}]\,\mathbf{A})\,\mathbf{Z} + [I|\mathbf{0}]\,\boldsymbol{\mu}$$

Compute mean and variance to check rest.

Proof of ($\mathbf{2}$): If $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$ then

$$\mathbf{MX} + \boldsymbol{\nu} = \mathbf{MAZ} + \boldsymbol{\nu} + \mathbf{M}\boldsymbol{\mu}$$

Proof of ($\mathbf{3}$): If

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$$

then

$$M_{\mathbf{X}}(u) = M_{\mathbf{X}_1}(\mathbf{u}_1) M_{\mathbf{X}_2}(\mathbf{u}_2)$$

Proof of (**4**): first case: assume $\mathbf{\Sigma}_{22}$ has an inverse.

Define

$$\mathbf{W} = \mathbf{X}_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{X}_2$$

Then

$$\begin{bmatrix} \mathbf{W} \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$$

Thus $(\mathbf{W}, \mathbf{X}_2)^T$ is $MVN(\boldsymbol{\mu}_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2, \mathbf{\Sigma}^*)$ where

$$\mathbf{\Sigma}^* = \begin{bmatrix} \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{22} \end{bmatrix}$$

Now joint density of $\mathbf{W}$ and $\mathbf{X}$ factors

$$f_{\mathbf{W},\mathbf{X}_2}(w,x_2) = f_{\mathbf{W}}(w) f_{\mathbf{X}_2}(x_2)$$

By change of variables joint density of $\mathbf{X}$ is

$$f_{\mathbf{X}_1,\mathbf{X}_2}(x_1,x_2) = c f_{\mathbf{W}}(x_1 - \mathbf{M}x_2) f_{\mathbf{X}_2}(x_2)$$

where $c = 1$ is the constant Jacobian of the linear transformation from $(\mathbf{W},\mathbf{X}_2)$ to $(\mathbf{X}_1,\mathbf{X}_2)$ and

$$\mathbf{M} = \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}$$

Thus conditional density of $\mathbf{X}_1$ given $\mathbf{X}_2 = x_2$ is

$$\frac{f_{\mathbf{W}}(x_1 - \mathbf{M}x_2) f_{\mathbf{X}_2}(x_2)}{f_{\mathbf{X}_2}(x_2)} = f_{\mathbf{W}}(x_1 - \mathbf{M}x_2)$$

As a function of x_1 this density has the form of the advertised multivariate normal density.

Specialization to bivariate case:

Write

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

where we define

$$\rho = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}}$$

Note that

$$\sigma_i^2 = \text{Var}(X_i)$$

Then

$$W = X_1 - \rho\frac{\sigma_1}{\sigma_2}X_2$$

is independent of X_2. The marginal distribution of W is $N(\mu_1 - \rho\sigma_1\mu_2/\sigma_2, \tau^2)$ where

$$\begin{aligned}\tau^2 =& \text{Var}(X_1) - 2\rho\frac{\sigma_1}{\sigma_2}\text{Cov}(X_1, X_2) \\ &+ \left(\rho\frac{\sigma_1}{\sigma_2}\right)^2 \text{Var}(X_2)\end{aligned}$$

This simplifies to

$$\sigma_1^2(1-\rho^2)$$

Notice that it follows that

$$-1 \le \rho \le 1$$

More generally: any X and Y:

$$\begin{aligned} 0 &\le \text{Var}(X-\lambda Y) \\ &= \text{Var}(X) - 2\lambda\text{Cov}(X,Y) + \lambda^2\text{Var}(Y) \end{aligned}$$

RHS is minimized at

$$\lambda = \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}$$

Minimum value is

$$0 \le \text{Var}(X)(1-\rho_{XY}^2)$$

where

$$\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

defines the correlation between X and Y.

Multiple Correlation

Now suppose $\mathbf{X}_2$ is scalar but $\mathbf{X}_1$ is vector.

Defn: Multiple correlation between $\mathbf{X}_1$ and X_2

$$R^2(\mathbf{X}_1, X_2) = \max |\rho_{\mathbf{a}^T\mathbf{X}_1, X_2}|^2$$

over all $\mathbf{a} \neq 0$.

Thus: maximize

$$\frac{\text{Cov}^2(\mathbf{a}^T\mathbf{X}_1, X_2)}{\text{Var}(\mathbf{a}^T\mathbf{X}_1)\text{Var}(X_2)} = \frac{\mathbf{a}^T\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{21}\mathbf{a}}{\left(\mathbf{a}^T\boldsymbol{\Sigma}_{11}\mathbf{a}\right)\boldsymbol{\Sigma}_{22}}$$

Put $b = \boldsymbol{\Sigma}_{11}^{1/2}\mathbf{a}$. For $\boldsymbol{\Sigma}_{11}$ invertible problem is equivalent to maximizing

$$\frac{\mathbf{b}^T\mathbf{Q}\mathbf{b}}{\mathbf{b}^T\mathbf{b}}$$

where

$$\mathbf{Q} = \boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}$$

Solution: find largest eigenvalue of $\mathbf{Q}$.

Note

$$\mathbf{Q} = \mathbf{v}\mathbf{v}^T$$

where

$$\mathbf{v} = \mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}$$

is a vector. Set

$$\mathbf{v}\mathbf{v}^T\mathbf{x} = \lambda\mathbf{x}$$

and multiply by $\mathbf{v}^T$ to get

$$\mathbf{v}^T\mathbf{x} = 0 \text{ or } \lambda = \mathbf{v}^T\mathbf{v}$$

If $\mathbf{v}^T\mathbf{x} = 0$ then we see $\lambda = 0$ so largest eigenvalue is $\mathbf{v}^T\mathbf{v}$.

Summary: maximum squared correlation is

$$R^2(\mathbf{X}_1, X_2) = \frac{\mathbf{v}^T\mathbf{v}}{\mathbf{\Sigma}_{22}} = \frac{\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}}{\mathbf{\Sigma}_{22}}$$

Achieved when eigenvector is $\mathbf{x} = \mathbf{v} = \mathbf{b}$ so

$$\mathbf{a} = \mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$$

Notice: since R^2 is squared correlation between two scalars ($\mathbf{a}^t\mathbf{X}_1$ and X_2) we have

$$0 \le R^2 \le 1$$

Equals 1 iff X_2 is linear combination of $\mathbf{X}_1$.

Correlation matrices, partial correlations:

Correlation between two scalars X and Y is

$$\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

If $\mathbf{X}$ has variance $\mathbf{\Sigma}$ then the correlation matrix of $\mathbf{X}$ is $\mathbf{R_X}$ with entries

$$R_{ij} = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)\text{Var}(X_j)}} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

If $\mathbf{X}_1, \mathbf{X}_2$ are MVN with the usual partitioned variance covariance matrix then the conditional variance of $\mathbf{X}_1$ given $\mathbf{X}_2$ is

$$\mathbf{\Sigma}_{11\cdot 2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$$

From this define **partial correlation matrix**

$$\mathbf{R}_{11\cdot 2} = \frac{(\mathbf{\Sigma}_{11\cdot 2})_{ij}}{\sqrt{\mathbf{\Sigma}_{11\cdot 2})_{ii}\mathbf{\Sigma}_{11\cdot 2})_{jj}}}$$

Note: these are used even when $\mathbf{X}_1, \mathbf{X}_2$ are NOT MVN