Multivariate Gaussians

COMS 4721 Spring 2022
Daniel Hsu

Motivation

Multivariate Gaussian (normal) distributions:

- One of the most basic statistical models for multivariate data
  - Gaussian distributions are plausible for variables that arise from an averaging process
  - With several such variables, it's natural to ask how they are statistically related
- Connections to linear regression
- Entry-point to broader world of *graphical models*

Basic multivariate statistics

Review: Variance and covariance

Let X and Y be random variables, and let $a, b \in \mathbb{R}$ be fixed scalars

- $\mathrm{var}(aX + b) =$
- $\mathrm{cov}(X + a, Y + b) =$
- $\mathrm{cov}(aX, bY) =$

Covariance matrix

For a random vector $\vec{X} = (X_1, \ldots, X_d)$, its **covariance matrix** is

$$\operatorname{cov}(\vec{X}) := \begin{bmatrix} \operatorname{var}(X_1) & \operatorname{cov}(X_1, X_2) & \cdots & \operatorname{cov}(X_1, X_d) \\ \operatorname{cov}(X_2, X_1) & \operatorname{var}(X_2) & \cdots & \operatorname{cov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_d, X_1) & \operatorname{cov}(X_d, X_2) & \cdots & \operatorname{var}(X_d) \end{bmatrix}$$

Covariance matrix

For a random vector $\vec{X} = (X_1, \ldots, X_d)$, its **covariance matrix** is

$$\operatorname{cov}(\vec{X}) := \begin{bmatrix} \operatorname{var}(X_1) & \operatorname{cov}(X_1, X_2) & \cdots & \operatorname{cov}(X_1, X_d) \\ \operatorname{cov}(X_2, X_1) & \operatorname{var}(X_2) & \cdots & \operatorname{cov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_d, X_1) & \operatorname{cov}(X_d, X_2) & \cdots & \operatorname{var}(X_d) \end{bmatrix}$$

$$= \begin{bmatrix} \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))^2\right] & \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))\right] & \cdots & \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))(X_d - \mathbb{E}(X_d))\right] \\ \mathbb{E}\left[(X_2 - \mathbb{E}(X_2))(X_1 - \mathbb{E}(X_1))\right] & \mathbb{E}\left[(X_2 - \mathbb{E}(X_2))^2\right] & \cdots & \mathbb{E}\left[(X_2 - \mathbb{E}(X_2))(X_d - \mathbb{E}(X_d))\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\left[(X_d - \mathbb{E}(X_d))(X_1 - \mathbb{E}(X_1))\right] & \mathbb{E}\left[(X_d - \mathbb{E}(X_d))(X_2 - \mathbb{E}(X_2))\right] & \cdots & \mathbb{E}\left[(X_d - \mathbb{E}(X_d))^2\right] \end{bmatrix}$$

Covariance matrix

For a random vector $\vec{X} = (X_1, \ldots, X_d)$, its **covariance matrix** is

$$\operatorname{cov}(\vec{X})$$

$$:= \begin{bmatrix} \operatorname{var}(X_1) & \operatorname{cov}(X_1, X_2) & \cdots & \operatorname{cov}(X_1, X_d) \\ \operatorname{cov}(X_2, X_1) & \operatorname{var}(X_2) & \cdots & \operatorname{cov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_d, X_1) & \operatorname{cov}(X_d, X_2) & \cdots & \operatorname{var}(X_d) \end{bmatrix}$$

$$= \begin{bmatrix} \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))^2\right] & \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))\right] & \cdots & \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))(X_d - \mathbb{E}(X_d))\right] \\ \mathbb{E}\left[(X_2 - \mathbb{E}(X_2))(X_1 - \mathbb{E}(X_1))\right] & \mathbb{E}\left[(X_2 - \mathbb{E}(X_2))^2\right] & \cdots & \mathbb{E}\left[(X_2 - \mathbb{E}(X_2))(X_d - \mathbb{E}(X_d))\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\left[(X_d - \mathbb{E}(X_d))(X_1 - \mathbb{E}(X_1))\right] & \mathbb{E}\left[(X_d - \mathbb{E}(X_d))(X_2 - \mathbb{E}(X_2))\right] & \cdots & \mathbb{E}\left[(X_d - \mathbb{E}(X_d))^2\right] \end{bmatrix}$$

$$= \mathbb{E} \begin{bmatrix} (X_1 - \mathbb{E}(X_1))^2 & (X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2)) & \cdots & (X_1 - \mathbb{E}(X_1))(X_d - \mathbb{E}(X_d)) \\ (X_2 - \mathbb{E}(X_2))(X_1 - \mathbb{E}(X_1)) & (X_2 - \mathbb{E}(X_2))^2 & \cdots & (X_2 - \mathbb{E}(X_2))(X_d - \mathbb{E}(X_d)) \\ \vdots & \vdots & \ddots & \vdots \\ (X_d - \mathbb{E}(X_d))(X_1 - \mathbb{E}(X_1)) & (X_d - \mathbb{E}(X_d))(X_2 - \mathbb{E}(X_2)) & \cdots & (X_d - \mathbb{E}(X_d))^2 \end{bmatrix}$$

Covariance matrix

For a random vector $\vec{X} = (X_1, \dots, X_d)$, its **covariance matrix** is

$$
\begin{aligned}
\operatorname{cov}(\vec{X}) \\
&:= \begin{bmatrix} \operatorname{var}(X_1) & \operatorname{cov}(X_1, X_2) & \cdots & \operatorname{cov}(X_1, X_d) \\ \operatorname{cov}(X_2, X_1) & \operatorname{var}(X_2) & \cdots & \operatorname{cov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(X_d, X_1) & \operatorname{cov}(X_d, X_2) & \cdots & \operatorname{var}(X_d) \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))^2\right] & \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2))\right] & \cdots & \mathbb{E}\left[(X_1 - \mathbb{E}(X_1))(X_d - \mathbb{E}(X_d))\right] \\ \mathbb{E}\left[(X_2 - \mathbb{E}(X_2))(X_1 - \mathbb{E}(X_1))\right] & \mathbb{E}\left[(X_2 - \mathbb{E}(X_2))^2\right] & \cdots & \mathbb{E}\left[(X_2 - \mathbb{E}(X_2))(X_d - \mathbb{E}(X_d))\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\left[(X_d - \mathbb{E}(X_d))(X_1 - \mathbb{E}(X_1))\right] & \mathbb{E}\left[(X_d - \mathbb{E}(X_d))(X_2 - \mathbb{E}(X_2))\right] & \cdots & \mathbb{E}\left[(X_d - \mathbb{E}(X_d))^2\right] \end{bmatrix} \\
&= \mathbb{E} \begin{bmatrix} (X_1 - \mathbb{E}(X_1))^2 & (X_1 - \mathbb{E}(X_1))(X_2 - \mathbb{E}(X_2)) & \cdots & (X_1 - \mathbb{E}(X_1))(X_d - \mathbb{E}(X_d)) \\ (X_2 - \mathbb{E}(X_2))(X_1 - \mathbb{E}(X_1)) & (X_2 - \mathbb{E}(X_2))^2 & \cdots & (X_2 - \mathbb{E}(X_2))(X_d - \mathbb{E}(X_d)) \\ \vdots & \vdots & \ddots & \vdots \\ (X_d - \mathbb{E}(X_d))(X_1 - \mathbb{E}(X_1)) & (X_d - \mathbb{E}(X_d))(X_2 - \mathbb{E}(X_2)) & \cdots & (X_d - \mathbb{E}(X_d))^2 \end{bmatrix} \\
&= \mathbb{E}\left[(\vec{X} - \mathbb{E}(\vec{X}))\,(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}}\right]
\end{aligned}
$$

Linear functions

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For a fixed vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, what are the mean and variance of $\vec{\alpha} \cdot \vec{X}$?

Linear functions

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For a fixed vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, what are the mean and variance of $\vec{\alpha} \cdot \vec{X}$?

$$\mathbb{E}(\vec{\alpha} \cdot \vec{X}) \quad = \quad \mathbb{E}\left(\alpha_1 X_1 + \cdots + \alpha_d X_d\right)$$

Linear functions

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For a fixed vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, what are the mean and variance of $\vec{\alpha} \cdot \vec{X}$?

$$\begin{aligned} \mathbb{E}(\vec{\alpha} \cdot \vec{X}) &= \mathbb{E}\left(\alpha_1 X_1 + \cdots + \alpha_d X_d\right) \\ &= \alpha_1 \, \mathbb{E}(X_1) + \cdots + \alpha_d \, \mathbb{E}(X_d) \end{aligned}$$

Linear functions

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For a fixed vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, what are the mean and variance of $\vec{\alpha} \cdot \vec{X}$?

$$\begin{aligned} \mathbb{E}(\vec{\alpha} \cdot \vec{X}) &= \mathbb{E}\left(\alpha_1 X_1 + \cdots + \alpha_d X_d\right) \\ &= \alpha_1 \, \mathbb{E}(X_1) + \cdots + \alpha_d \, \mathbb{E}(X_d) \\ &= \vec{\alpha} \cdot \mathbb{E}(\vec{X}) \end{aligned}$$

Linear functions

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For a fixed vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, what are the mean and variance of $\vec{\alpha} \cdot \vec{X}$?

$$\begin{aligned} \mathbb{E}(\vec{\alpha} \cdot \vec{X}) &= \mathbb{E}\left(\alpha_1 X_1 + \cdots + \alpha_d X_d\right) \\ &= \alpha_1 \, \mathbb{E}(X_1) + \cdots + \alpha_d \, \mathbb{E}(X_d) \\ &= \vec{\alpha} \cdot \mathbb{E}(\vec{X}) \end{aligned}$$

$$\mathrm{var}(\vec{\alpha} \cdot \vec{X}) = \mathbb{E}\left[(\vec{\alpha} \cdot \vec{X} - \vec{\alpha} \cdot \mathbb{E}(\vec{X}))^2\right]$$

Linear functions

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For a fixed vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, what are the mean and variance of $\vec{\alpha} \cdot \vec{X}$?

$$\begin{aligned} \mathbb{E}(\vec{\alpha} \cdot \vec{X}) &= \mathbb{E}\left(\alpha_1 X_1 + \cdots + \alpha_d X_d\right) \\ &= \alpha_1 \, \mathbb{E}(X_1) + \cdots + \alpha_d \, \mathbb{E}(X_d) \\ &= \vec{\alpha} \cdot \mathbb{E}(\vec{X}) \end{aligned}$$

$$\begin{aligned} \mathrm{var}(\vec{\alpha} \cdot \vec{X}) &= \mathbb{E}\left[(\vec{\alpha} \cdot \vec{X} - \vec{\alpha} \cdot \mathbb{E}(\vec{X}))^2\right] \\ &= \mathbb{E}\left[\vec{\alpha}^{\mathsf{T}} (\vec{X} - \mathbb{E}(\vec{X}))(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}} \vec{\alpha}\right] \end{aligned}$$

Linear functions

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For a fixed vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, what are the mean and variance of $\vec{\alpha} \cdot \vec{X}$?

$$\begin{aligned}
\mathbb{E}(\vec{\alpha} \cdot \vec{X}) &= \mathbb{E}\left(\alpha_1 X_1 + \cdots + \alpha_d X_d\right) \\
&= \alpha_1 \, \mathbb{E}(X_1) + \cdots + \alpha_d \, \mathbb{E}(X_d) \\
&= \vec{\alpha} \cdot \mathbb{E}(\vec{X})
\end{aligned}$$

$$\begin{aligned}
\mathrm{var}(\vec{\alpha} \cdot \vec{X}) &= \mathbb{E}\left[(\vec{\alpha} \cdot \vec{X} - \vec{\alpha} \cdot \mathbb{E}(\vec{X}))^2\right] \\
&= \mathbb{E}\left[\vec{\alpha}^{\mathsf{T}} (\vec{X} - \mathbb{E}(\vec{X}))(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}} \vec{\alpha}\right] \\
&= \vec{\alpha}^{\mathsf{T}} \, \mathbb{E}\left[(\vec{X} - \mathbb{E}(\vec{X}))(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}}\right] \vec{\alpha}
\end{aligned}$$

Linear functions

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For a fixed vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, what are the mean and variance of $\vec{\alpha} \cdot \vec{X}$?

$$\begin{aligned}
\mathbb{E}(\vec{\alpha} \cdot \vec{X}) &= \mathbb{E}\left(\alpha_1 X_1 + \cdots + \alpha_d X_d\right) \\
&= \alpha_1 \, \mathbb{E}(X_1) + \cdots + \alpha_d \, \mathbb{E}(X_d) \\
&= \vec{\alpha} \cdot \mathbb{E}(\vec{X})
\end{aligned}$$

$$\begin{aligned}
\mathrm{var}(\vec{\alpha} \cdot \vec{X}) &= \mathbb{E}\left[(\vec{\alpha} \cdot \vec{X} - \vec{\alpha} \cdot \mathbb{E}(\vec{X}))^2\right] \\
&= \mathbb{E}\left[\vec{\alpha}^{\mathsf{T}}(\vec{X} - \mathbb{E}(\vec{X}))(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}}\vec{\alpha}\right] \\
&= \vec{\alpha}^{\mathsf{T}} \, \mathbb{E}\left[(\vec{X} - \mathbb{E}(\vec{X}))(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}}\right] \vec{\alpha} \\
&= \vec{\alpha}^{\mathsf{T}} \, \mathrm{cov}(\vec{X}) \vec{\alpha}
\end{aligned}$$

Linear functions (again)

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For fixed vectors $\vec{\alpha} \in \mathbb{R}^d$ and $\vec{\beta} \in \mathbb{R}^d$, what is the covariance between $\vec{\alpha} \cdot \vec{X}$ and $\vec{\beta} \cdot \vec{X}$?

Linear functions (again)

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For fixed vectors $\vec{\alpha} \in \mathbb{R}^d$ and $\vec{\beta} \in \mathbb{R}^d$, what is the covariance between $\vec{\alpha} \cdot \vec{X}$ and $\vec{\beta} \cdot \vec{X}$?

$$\text{cov}(\vec{\alpha} \cdot \vec{X}, \vec{\beta} \cdot \vec{X}) \quad = \quad \mathbb{E}\left[(\vec{\alpha} \cdot \vec{X} - \vec{\alpha} \cdot \mathbb{E}(\vec{X}))(\vec{\beta} \cdot \vec{X} - \vec{\beta} \cdot \mathbb{E}(\vec{X}))\right]$$

Linear functions (again)

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For fixed vectors $\vec{\alpha} \in \mathbb{R}^d$ and $\vec{\beta} \in \mathbb{R}^d$, what is the covariance between $\vec{\alpha} \cdot \vec{X}$ and $\vec{\beta} \cdot \vec{X}$?

$$\begin{aligned} \operatorname{cov}(\vec{\alpha} \cdot \vec{X}, \vec{\beta} \cdot \vec{X}) &= \mathbb{E}\left[(\vec{\alpha} \cdot \vec{X} - \vec{\alpha} \cdot \mathbb{E}(\vec{X}))(\vec{\beta} \cdot \vec{X} - \vec{\beta} \cdot \mathbb{E}(\vec{X}))\right] \\ &= \mathbb{E}\left[\vec{\alpha}^{\mathsf{T}}(\vec{X} - \mathbb{E}(\vec{X}))(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}}\vec{\beta}\right] \end{aligned}$$

Linear functions (again)

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For fixed vectors $\vec{\alpha} \in \mathbb{R}^d$ and $\vec{\beta} \in \mathbb{R}^d$, what is the covariance between $\vec{\alpha} \cdot \vec{X}$ and $\vec{\beta} \cdot \vec{X}$?

$$
\begin{aligned}
\operatorname{cov}(\vec{\alpha} \cdot \vec{X}, \vec{\beta} \cdot \vec{X}) &= \mathbb{E}\left[(\vec{\alpha} \cdot \vec{X} - \vec{\alpha} \cdot \mathbb{E}(\vec{X}))(\vec{\beta} \cdot \vec{X} - \vec{\beta} \cdot \mathbb{E}(\vec{X}))\right] \\
&= \mathbb{E}\left[\vec{\alpha}^{\mathsf{T}}(\vec{X} - \mathbb{E}(\vec{X}))(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}}\vec{\beta}\right] \\
&= \vec{\alpha}^{\mathsf{T}}\, \mathbb{E}\left[(\vec{X} - \mathbb{E}(\vec{X}))(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}}\right] \vec{\beta}
\end{aligned}
$$

Linear functions (again)

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For fixed vectors $\vec{\alpha} \in \mathbb{R}^d$ and $\vec{\beta} \in \mathbb{R}^d$, what is the covariance between $\vec{\alpha} \cdot \vec{X}$ and $\vec{\beta} \cdot \vec{X}$?

$$
\begin{aligned}
\operatorname{cov}(\vec{\alpha} \cdot \vec{X}, \vec{\beta} \cdot \vec{X}) &= \mathbb{E}\left[(\vec{\alpha} \cdot \vec{X} - \vec{\alpha} \cdot \mathbb{E}(\vec{X}))(\vec{\beta} \cdot \vec{X} - \vec{\beta} \cdot \mathbb{E}(\vec{X}))\right] \\
&= \mathbb{E}\left[\vec{\alpha}^{\mathsf{T}}(\vec{X} - \mathbb{E}(\vec{X}))(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}}\vec{\beta}\right] \\
&= \vec{\alpha}^{\mathsf{T}}\, \mathbb{E}\left[(\vec{X} - \mathbb{E}(\vec{X}))(\vec{X} - \mathbb{E}(\vec{X}))^{\mathsf{T}}\right] \vec{\beta} \\
&= \vec{\alpha}^{\mathsf{T}} \operatorname{cov}(\vec{X})\vec{\beta}
\end{aligned}
$$

Linear transformations

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For a fixed matrix $A \in \mathbb{R}^{d \times k}$, what are $\mathbb{E}(A^\top \vec{X})$ and $\operatorname{cov}(A^\top \vec{X})$?

Linear transformations

Let $\vec{X} = (X_1, \ldots, X_d)$ be a random vector

Question: For a fixed matrix $A \in \mathbb{R}^{d \times k}$, what are $\mathbb{E}(A^\intercal \vec{X})$ and $\mathrm{cov}(A^\intercal \vec{X})$?

$$\begin{aligned} \mathbb{E}(A^\intercal \vec{X}) &= A^\intercal \, \mathbb{E}(\vec{X}) \\ \mathrm{cov}(A^\intercal \vec{X}) &= A^\intercal \, \mathrm{cov}(\vec{X}) \, A \end{aligned}$$

Just think of A as specifying k linear functions, one per column . . .

$$A = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{\alpha}_1 & \vec{\alpha}_2 & \cdots & \vec{\alpha}_k \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

Example: Dartmouth data

$\vec{X} = (X_1, X_2, X_3) = (\texttt{SAT Score}, \texttt{HS GPA}, \texttt{College GPA})$

Example: Dartmouth data

$\vec{X} = (X_1, X_2, X_3) = (\texttt{SAT Score}, \texttt{HS GPA}, \texttt{College GPA})$

(Sample) mean and covariance:

$$\widehat{\mathbb{E}}[\vec{X}] = \begin{bmatrix} 1028.59 \\ 3.18324 \\ 2.46031 \end{bmatrix}, \qquad \widehat{\text{cov}}(\vec{X}) = \begin{bmatrix} 20602.8 & 34.2707 & 51.6484 \\ 34.2707 & 0.30857 & 0.226406 \\ 51.6484 & 0.226406 & 0.556484 \end{bmatrix}$$

Example: Dartmouth data

$\vec{X} = (X_1, X_2, X_3) = (\texttt{SAT Score}, \texttt{HS GPA}, \texttt{College GPA})$

(Sample) mean and covariance:

$$\widehat{\mathbb{E}}[\vec{X}] = \begin{bmatrix} 1028.59 \\ 3.18324 \\ 2.46031 \end{bmatrix}, \qquad \widehat{\text{cov}}(\vec{X}) = \begin{bmatrix} 20602.8 & 34.2707 & 51.6484 \\ 34.2707 & 0.30857 & 0.226406 \\ 51.6484 & 0.226406 & 0.556484 \end{bmatrix}$$

(Sample) correlation matrix

$$\widehat{\text{cor}}(\vec{X}) = \begin{bmatrix} 1.0000 & 0.4298 & 0.4824 \\ 0.4298 & 1.0000 & 0.5464 \\ 0.4824 & 0.5464 & 1.0000 \end{bmatrix} \quad \text{where} \quad \widehat{\text{cor}}(\vec{X})_{i,j} = \frac{\widehat{\text{cov}}(\vec{X})_{i,j}}{\sqrt{\widehat{\text{cov}}(\vec{X})_{i,i}\,\widehat{\text{cov}}(\vec{X})_{j,j}}}$$

Example: Dartmouth data

$\vec{X} = (X_1, X_2, X_3) = (\texttt{SAT Score}, \texttt{HS GPA}, \texttt{College GPA})$

(Sample) mean and covariance:

$$\widehat{\mathbb{E}}[\vec{X}] = \begin{bmatrix} 1028.59 \\ 3.18324 \\ 2.46031 \end{bmatrix}, \qquad \widehat{\mathrm{cov}}(\vec{X}) = \begin{bmatrix} 20602.8 & 34.2707 & 51.6484 \\ 34.2707 & 0.30857 & 0.226406 \\ 51.6484 & 0.226406 & 0.556484 \end{bmatrix}$$

(Sample) correlation matrix

$$\widehat{\mathrm{cor}}(\vec{X}) = \begin{bmatrix} 1.0000 & 0.4298 & 0.4824 \\ 0.4298 & 1.0000 & 0.5464 \\ 0.4824 & 0.5464 & 1.0000 \end{bmatrix} \quad \text{where} \quad \widehat{\mathrm{cor}}(\vec{X})_{i,j} = \frac{\widehat{\mathrm{cov}}(\vec{X})_{i,j}}{\sqrt{\widehat{\mathrm{cov}}(\vec{X})_{i,i}\,\widehat{\mathrm{cov}}(\vec{X})_{j,j}}}$$

Question: Let Y_1 be average of `SAT Score` and `HS GPA`, and let Y_2 be average of `HS GPA` and `College GPA`. What is $\mathrm{cov}(Y_1, Y_2)$?

Example: Dartmouth data

$\vec{X} = (X_1, X_2, X_3) = (\texttt{SAT Score}, \texttt{HS GPA}, \texttt{College GPA})$

(Sample) mean and covariance:

$$\widehat{\mathbb{E}}[\vec{X}] = \begin{bmatrix} 1028.59 \\ 3.18324 \\ 2.46031 \end{bmatrix}, \qquad \widehat{\mathrm{cov}}(\vec{X}) = \begin{bmatrix} 20602.8 & 34.2707 & 51.6484 \\ 34.2707 & 0.30857 & 0.226406 \\ 51.6484 & 0.226406 & 0.556484 \end{bmatrix}$$

(Sample) correlation matrix

$$\widehat{\mathrm{cor}}(\vec{X}) = \begin{bmatrix} 1.0000 & 0.4298 & 0.4824 \\ 0.4298 & 1.0000 & 0.5464 \\ 0.4824 & 0.5464 & 1.0000 \end{bmatrix} \quad \text{where} \quad \widehat{\mathrm{cor}}(\vec{X})_{i,j} = \frac{\widehat{\mathrm{cov}}(\vec{X})_{i,j}}{\sqrt{\widehat{\mathrm{cov}}(\vec{X})_{i,i}\,\widehat{\mathrm{cov}}(\vec{X})_{j,j}}}$$

Question: Let Y_1 be average of `SAT Score` and `HS GPA`, and let Y_2 be average of `HS GPA` and `College GPA`. What is $\mathrm{cov}(Y_1, Y_2)$?

$$\begin{bmatrix} 1/2 & 1/2 & 0 \end{bmatrix} \widehat{\mathrm{cov}}(\vec{X}) \begin{bmatrix} 0 \\ 1/2 \\ 1/2 \end{bmatrix} = 21.6135$$

Isotropic and anisotropic random vectors

A random vector $\vec{X} = (X_1, \ldots, X_d)$ is **isotropic** if

- $\mathrm{var}(X_i) = 1$ for all i, and
- $\mathrm{cov}(X_i, X_j) = 0$ for all $i \neq j$

Isotropic and anisotropic random vectors

A random vector $\vec{X} = (X_1, \ldots, X_d)$ is **isotropic** if

- $\mathrm{var}(X_i) = 1$ for all i, and
- $\mathrm{cov}(X_i, X_j) = 0$ for all $i \neq j$

Shorthand: $\mathrm{cov}(\vec{X}) = I_d$ (we'll sometimes include subscript d on I to emphasize $I_d \in \mathbb{R}^{d \times d}$)

Isotropic and anisotropic random vectors

A random vector $\vec{X} = (X_1, \ldots, X_d)$ is **isotropic** if

- $\mathrm{var}(X_i) = 1$ for all i, and
- $\mathrm{cov}(X_i, X_j) = 0$ for all $i \neq j$

Shorthand: $\mathrm{cov}(\vec{X}) = I_d$ (we'll sometimes include subscript d on I to emphasize $I_d \in \mathbb{R}^{d \times d}$)

Note: Isotropy of $\vec{X} \not\Rightarrow X_1, \ldots, X_d$ are independent

Isotropic and anisotropic random vectors

A random vector $\vec{X} = (X_1, \ldots, X_d)$ is **isotropic** if

- $\mathrm{var}(X_i) = 1$ for all i, and
- $\mathrm{cov}(X_i, X_j) = 0$ for all $i \neq j$

Shorthand: $\mathrm{cov}(\vec{X}) = I_d$ (we'll sometimes include subscript d on I to emphasize $I_d \in \mathbb{R}^{d \times d}$)

Note: Isotropy of $\vec{X} \not\Rightarrow X_1, \ldots, X_d$ are independent

(**anisotropic** = "not isotropic")

Summary

- Covariance between X and Y: How do X and Y “vary” together, on average?
- Covariance matrix of $\vec{X}$:
  - Lets you determine variance of *any affine function* of $\vec{X}$
  - Lets you determine covariance between *any pair of affine functions* of $\vec{X}$

Multivariate Gaussians

Standard multivariate Gaussian

Random vector $\vec{Z} = (Z_1, \ldots, Z_d)$ follows the **standard (isotropic) multivariate Gaussian distribution** in $\mathbb{R}^d$ if

$$Z_1, \ldots, Z_d \ \sim_{\text{iid}} \ \mathrm{N}(0, 1)$$

i.e., Z_i's are IID, and marginal distribution of each Z_i is standard (univariate) Gaussian distribution

Standard multivariate Gaussian

Random vector $\vec{Z} = (Z_1, \ldots, Z_d)$ follows the **standard (isotropic) multivariate Gaussian distribution** in $\mathbb{R}^d$ if

$$Z_1, \ldots, Z_d \ \sim_{\text{iid}} \ \text{N}(0, 1)$$

i.e., Z_i's are IID, and marginal distribution of each Z_i is standard (univariate) Gaussian distribution

- Shorthand: $\vec{Z} \sim \text{N}(\vec{0}, I_d)$

Standard multivariate Gaussian

Random vector $\vec{Z} = (Z_1, \dots, Z_d)$ follows the **standard (isotropic) multivariate Gaussian distribution** in $\mathbb{R}^d$ if

$$Z_1, \dots, Z_d \ \sim_{\mathrm{iid}} \ \mathrm{N}(0, 1)$$

i.e., Z_i's are IID, and marginal distribution of each Z_i is standard (univariate) Gaussian distribution

- Shorthand: $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$
- $\mathbb{E}(\vec{Z}) = \vec{0}$

Standard multivariate Gaussian

Random vector $\vec{Z} = (Z_1, \ldots, Z_d)$ follows the **standard (isotropic) multivariate Gaussian distribution** in $\mathbb{R}^d$ if

$$Z_1, \ldots, Z_d \ \sim_{\text{iid}} \ \mathrm{N}(0, 1)$$

i.e., Z_i's are IID, and marginal distribution of each Z_i is standard (univariate) Gaussian distribution

- Shorthand: $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$
- $\mathbb{E}(\vec{Z}) = \vec{0}$
- $\mathrm{cov}(\vec{Z}) = I_d$

Standard multivariate Gaussian

Random vector $\vec{Z} = (Z_1, \ldots, Z_d)$ follows the **standard (isotropic) multivariate Gaussian distribution** in $\mathbb{R}^d$ if

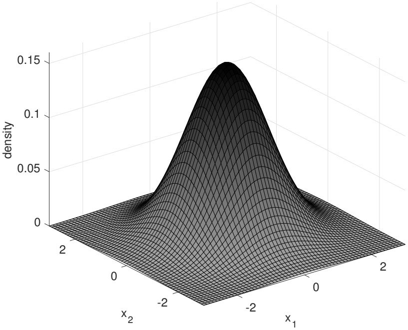
$$Z_1, \ldots, Z_d \ \sim_{\text{iid}} \ \mathrm{N}(0, 1)$$

i.e., Z_i's are IID, and marginal distribution of each Z_i is standard (univariate) Gaussian distribution

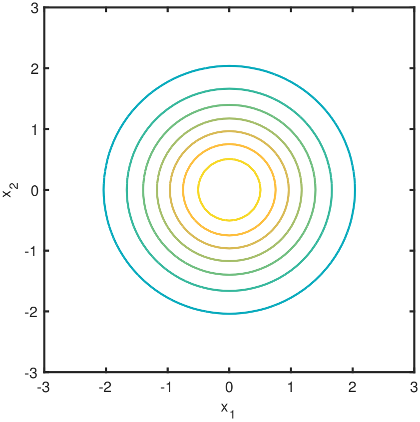
- Shorthand: $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$
- $\mathbb{E}(\vec{Z}) = \vec{0}$
- $\mathrm{cov}(\vec{Z}) = I_d$
- Probability density function (PDF) $p_{\vec{Z}} \colon \mathbb{R}^d \to \mathbb{R}$:

$$p_{\vec{Z}}(\vec{z}) \ = \ \prod_{i=1}^{d} p_{Z_i}(z_i) \ = \ \prod_{i=1}^{d} \left\{ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_i^2}{2} \right) \right\} \ = \ \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|\vec{z}\|_2^2}{2} \right)$$

Density function for standard bivariate Gaussian distribution



PDF



Level sets of PDF

Affine functions

If $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then for any $\mu \in \mathbb{R}$ and $\vec{\alpha} \in \mathbb{R}^d$, the random variable

$$X \;:=\; \mu + \vec{\alpha} \cdot \vec{Z} \;=\; \mu + \sum_{i=1}^{d} \alpha_i Z_i$$

follows a (univariate) Gaussian distribution!

Affine functions

If $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then for any $\mu \in \mathbb{R}$ and $\vec{\alpha} \in \mathbb{R}^d$, the random variable

$$X := \mu + \vec{\alpha} \cdot \vec{Z} = \mu + \sum_{i=1}^{d} \alpha_i Z_i$$

follows a (univariate) Gaussian distribution!

- Mean: μ

Affine functions

If $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then for any $\mu \in \mathbb{R}$ and $\vec{\alpha} \in \mathbb{R}^d$, the random variable

$$X \;:=\; \mu + \vec{\alpha} \cdot \vec{Z} \;=\; \mu + \sum_{i=1}^{d} \alpha_i Z_i$$

follows a (univariate) Gaussian distribution!

- Mean: μ
- Variance: $\sigma^2 := \|\vec{\alpha}\|_2^2$

Affine functions

If $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then for any $\mu \in \mathbb{R}$ and $\vec{\alpha} \in \mathbb{R}^d$, the random variable

$$X \;:=\; \mu + \vec{\alpha} \cdot \vec{Z} \;=\; \mu + \sum_{i=1}^{d} \alpha_i Z_i$$

follows a (univariate) Gaussian distribution!

- Mean: μ
- Variance: $\sigma^2 := \|\vec{\alpha}\|_2^2$
- Shorthand: $X \sim \mathrm{N}(\mu, \sigma^2)$

Affine functions

If $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then for any $\mu \in \mathbb{R}$ and $\vec{\alpha} \in \mathbb{R}^d$, the random variable

$$X \;:=\; \mu + \vec{\alpha} \cdot \vec{Z} \;=\; \mu + \sum_{i=1}^{d} \alpha_i Z_i$$

follows a (univariate) Gaussian distribution!

- Mean: μ
- Variance: $\sigma^2 := \|\vec{\alpha}\|_2^2$
- Shorthand: $X \sim \mathrm{N}(\mu, \sigma^2)$
- PDF:

$$p_X(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Affine functions (again)

If $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then for any $\mu \in \mathbb{R}$, $\nu \in \mathbb{R}$, $\vec{\alpha} \in \mathbb{R}^d$, $\vec{\beta} \in \mathbb{R}^d$, the random variables

$$X_1 \;:=\; \mu + \vec{\alpha} \cdot \vec{Z} \quad \text{and} \quad X_2 \;:=\; \nu + \vec{\beta} \cdot \vec{Z}$$

jointly follow a (bivariate) Gaussian distribution, and

$$\mathrm{cov}(X_1, X_2) \;=\; \vec{\alpha} \cdot \vec{\beta}$$

Affine functions (again)

If $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then for any $\mu \in \mathbb{R}$, $\nu \in \mathbb{R}$, $\vec{\alpha} \in \mathbb{R}^d$, $\vec{\beta} \in \mathbb{R}^d$, the random variables

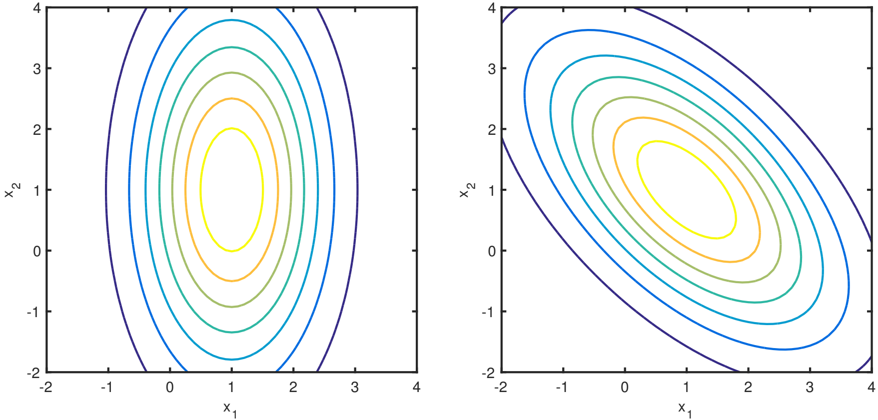
$$X_1 \;:=\; \mu + \vec{\alpha} \cdot \vec{Z} \quad \text{and} \quad X_2 \;:=\; \nu + \vec{\beta} \cdot \vec{Z}$$

jointly follow a (bivariate) Gaussian distribution, and

$$\operatorname{cov}(X_1, X_2) \;=\; \vec{\alpha} \cdot \vec{\beta}$$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mu \\ \nu \end{bmatrix}, \begin{bmatrix} \|\vec{\alpha}\|_2^2 & \vec{\alpha} \cdot \vec{\beta} \\ \vec{\beta} \cdot \vec{\alpha} & \|\vec{\beta}\|_2^2 \end{bmatrix} \right)$$

Examples of bivariate Gaussian distributions



Level sets of PDFs for two bivariate Gaussian distributions

General multivariate Gaussian distributions

If $\vec{X} := \vec{\mu} + A^{\mathsf{T}}\vec{Z}$ for some $\vec{\mu} \in \mathbb{R}^k$ and $A \in \mathbb{R}^{d\times k}$ with $\operatorname{rank}(A) = k$, and $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then

General multivariate Gaussian distributions

If $\vec{X} := \vec{\mu} + A^{\mathsf{T}}\vec{Z}$ for some $\vec{\mu} \in \mathbb{R}^k$ and $A \in \mathbb{R}^{d \times k}$ with $\operatorname{rank}(A) = k$, and $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then

- $\vec{X}$ follows the **multivariate Gaussian distribution** in $\mathbb{R}^k$ with mean $\vec{\mu}$ and covariance $\Sigma := A^{\mathsf{T}}A$

General multivariate Gaussian distributions

If $\vec{X} := \vec{\mu} + A^\intercal \vec{Z}$ for some $\vec{\mu} \in \mathbb{R}^k$ and $A \in \mathbb{R}^{d \times k}$ with $\operatorname{rank}(A) = k$, and $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then

- $\vec{X}$ follows the **multivariate Gaussian distribution** in $\mathbb{R}^k$ with mean $\vec{\mu}$ and covariance $\Sigma := A^\intercal A$
- Shorthand: $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$

General multivariate Gaussian distributions

If $\vec{X} := \vec{\mu} + A^{\mathsf{T}}\vec{Z}$ for some $\vec{\mu} \in \mathbb{R}^k$ and $A \in \mathbb{R}^{d \times k}$ with $\operatorname{rank}(A) = k$, and $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then

- $\vec{X}$ follows the **multivariate Gaussian distribution** in $\mathbb{R}^k$ with mean $\vec{\mu}$ and covariance $\Sigma := A^{\mathsf{T}}A$
- Shorthand: $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$
- (If $\operatorname{rank}(A) < k$, then distribution of $\vec{X}$ is degenerate)

General multivariate Gaussian distributions

If $\vec{X} := \vec{\mu} + A^{\mathsf{T}}\vec{Z}$ for some $\vec{\mu} \in \mathbb{R}^k$ and $A \in \mathbb{R}^{d \times k}$ with $\operatorname{rank}(A) = k$, and $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then

- $\vec{X}$ follows the **multivariate Gaussian distribution** in $\mathbb{R}^k$ with mean $\vec{\mu}$ and covariance $\Sigma := A^{\mathsf{T}}A$
- Shorthand: $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$
- (If $\operatorname{rank}(A) < k$, then distribution of $\vec{X}$ is degenerate)

If $k = d$ and A is invertible, PDF of $\vec{X}$ can be obtained using "change-of-variables" formula:

$$p_{\vec{X}}(\vec{x}) = \frac{1}{|\det(A)|} \cdot p_{\vec{Z}}\left(A^{-\mathsf{T}}(\vec{x} - \vec{\mu})\right)$$

General multivariate Gaussian distributions

If $\vec{X} := \vec{\mu} + A^{\mathsf{T}}\vec{Z}$ for some $\vec{\mu} \in \mathbb{R}^k$ and $A \in \mathbb{R}^{d \times k}$ with $\operatorname{rank}(A) = k$, and $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then

- $\vec{X}$ follows the **multivariate Gaussian distribution** in $\mathbb{R}^k$ with mean $\vec{\mu}$ and covariance $\Sigma := A^{\mathsf{T}}A$
- Shorthand: $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$
- (If $\operatorname{rank}(A) < k$, then distribution of $\vec{X}$ is degenerate)

If $k = d$ and A is invertible, PDF of $\vec{X}$ can be obtained using "change-of-variables" formula:

$$
\begin{aligned}
p_{\vec{X}}(\vec{x}) &= \frac{1}{|\det(A)|} \cdot p_{\vec{Z}}\left(A^{-\mathsf{T}}(\vec{x} - \vec{\mu})\right) \\
&= \frac{1}{\det(\Sigma)^{1/2}} \cdot \frac{1}{(2\pi)^{d/2}} \cdot \exp\left(-\frac{1}{2}\left\|A^{-1}(\vec{x} - \vec{\mu})\right\|_2^2\right)
\end{aligned}
$$

General multivariate Gaussian distributions

If $\vec{X} := \vec{\mu} + A^{\mathsf{T}}\vec{Z}$ for some $\vec{\mu} \in \mathbb{R}^k$ and $A \in \mathbb{R}^{d\times k}$ with $\operatorname{rank}(A) = k$, and $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then

- $\vec{X}$ follows the **multivariate Gaussian distribution** in $\mathbb{R}^k$ with mean $\vec{\mu}$ and covariance $\Sigma := A^{\mathsf{T}}A$
- Shorthand: $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$
- (If $\operatorname{rank}(A) < k$, then distribution of $\vec{X}$ is degenerate)

If $k = d$ and A is invertible, PDF of $\vec{X}$ can be obtained using "change-of-variables" formula:

$$\begin{aligned} p_{\vec{X}}(\vec{x}) &= \frac{1}{|\det(A)|} \cdot p_{\vec{Z}}\left(A^{-\mathsf{T}}(\vec{x} - \vec{\mu})\right) \\ &= \frac{1}{\det(\Sigma)^{1/2}} \cdot \frac{1}{(2\pi)^{d/2}} \cdot \exp\left(-\frac{1}{2}\left\|A^{-1}(\vec{x} - \vec{\mu})\right\|_2^2\right) \\ &= \frac{1}{\det(\Sigma)^{1/2}} \cdot \frac{1}{(2\pi)^{d/2}} \cdot \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^{\mathsf{T}}\Sigma^{-1}(\vec{x} - \vec{\mu})\right) \end{aligned}$$

General multivariate Gaussian distributions

If $\vec{X} := \vec{\mu} + A^{\mathsf{T}}\vec{Z}$ for some $\vec{\mu} \in \mathbb{R}^k$ and $A \in \mathbb{R}^{d \times k}$ with $\operatorname{rank}(A) = k$, and $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$, then

- $\vec{X}$ follows the **multivariate Gaussian distribution** in $\mathbb{R}^k$ with mean $\vec{\mu}$ and covariance $\Sigma := A^{\mathsf{T}}A$
- Shorthand: $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$
- (If $\operatorname{rank}(A) < k$, then distribution of $\vec{X}$ is degenerate)

If $k = d$ and A is invertible, PDF of $\vec{X}$ can be obtained using "change-of-variables" formula:

$$\begin{aligned} p_{\vec{X}}(\vec{x}) &= \frac{1}{|\det(A)|} \cdot p_{\vec{Z}}\left(A^{-\mathsf{T}}(\vec{x} - \vec{\mu})\right) \\ &= \frac{1}{\det(\Sigma)^{1/2}} \cdot \frac{1}{(2\pi)^{d/2}} \cdot \exp\left(-\frac{1}{2}\left\|A^{-1}(\vec{x} - \vec{\mu})\right\|_2^2\right) \\ &= \frac{1}{\det(\Sigma)^{1/2}} \cdot \frac{1}{(2\pi)^{d/2}} \cdot \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^{\mathsf{T}}\Sigma^{-1}(\vec{x} - \vec{\mu})\right) \end{aligned}$$

If $k < d$ and $\operatorname{rank}(A) = k$, formula for $p_{\vec{X}}(\vec{x})$ is final line shown above

Marginalization

Let $\vec{X} = (X_1, \ldots, X_d) \sim \mathrm{N}(\vec{\mu}, \Sigma)$

Question: What is the marginal distribution of $(X_1, \ldots, X_k)$ for some $k < d$?

Marginalization

Let $\vec{X} = (X_1, \ldots, X_d) \sim \mathrm{N}(\vec{\mu}, \Sigma)$

Question: What is the marginal distribution of $(X_1, \ldots, X_k)$ for some $k < d$?

- A Gaussian distribution in $\mathbb{R}^k$

Marginalization

Let $\vec{X} = (X_1, \ldots, X_d) \sim \mathrm{N}(\vec{\mu}, \Sigma)$

Question: What is the marginal distribution of $(X_1, \ldots, X_k)$ for some $k < d$?

- A Gaussian distribution in $\mathbb{R}^k$
- Mean and covariance are obtained by taking appropriate sub-vector of $\vec{\mu}$ and sub-matrix of Σ

Marginalization

Let $\vec{X} = (X_1, \ldots, X_d) \sim \mathrm{N}(\vec{\mu}, \Sigma)$

Question: What is the marginal distribution of $(X_1, \ldots, X_k)$ for some $k < d$?

- A Gaussian distribution in $\mathbb{R}^k$
- Mean and covariance are obtained by taking appropriate sub-vector of $\vec{\mu}$ and sub-matrix of Σ

Why?

Marginalization

Let $\vec{X} = (X_1, \ldots, X_d) \sim \mathrm{N}(\vec{\mu}, \Sigma)$

Question: What is the marginal distribution of $(X_1, \ldots, X_k)$ for some $k < d$?

- A Gaussian distribution in $\mathbb{R}^k$
- Mean and covariance are obtained by taking appropriate sub-vector of $\vec{\mu}$ and sub-matrix of Σ

Why?

- $\vec{X}$ is obtained by taking d affine functions of $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$

Marginalization

Let $\vec{X} = (X_1, \ldots, X_d) \sim \mathrm{N}(\vec{\mu}, \mathit{\Sigma})$

Question: What is the marginal distribution of $(X_1, \ldots, X_k)$ for some $k < d$?

- A Gaussian distribution in $\mathbb{R}^k$
- Mean and covariance are obtained by taking appropriate sub-vector of $\vec{\mu}$ and sub-matrix of $\mathit{\Sigma}$

Why?

- $\vec{X}$ is obtained by taking d affine functions of $\vec{Z} \sim \mathrm{N}(\vec{0}, I_d)$
- $(X_1, \ldots, X_k)$ just takes the first k of them!

Conditioning

Partition $\{1, \dots, d\} = A \cup B$, with $A = \{1, \dots, p\}$ and $B = \{p+1, \dots, p+q\}$ where $p + q = d$

Conditioning

Partition $\{1, \ldots, d\} = A \cup B$, with $A = \{1, \ldots, p\}$ and $B = \{p+1, \ldots, p+q\}$ where $p + q = d$

Let $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$ be multivariate Gaussian random vector in $\mathbb{R}^d$, and let

$$\vec{X} = \begin{bmatrix} \vec{X}_A \\ \vec{X}_B \end{bmatrix}, \qquad \vec{\mu} = \begin{bmatrix} \vec{\mu}_A \\ \vec{\mu}_B \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{A,A} & \Sigma_{A,B} \\ \Sigma_{B,A} & \Sigma_{B,B} \end{bmatrix}$$

Conditioning

Partition $\{1, \ldots, d\} = A \cup B$, with $A = \{1, \ldots, p\}$ and $B = \{p+1, \ldots, p+q\}$ where $p + q = d$

Let $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$ be multivariate Gaussian random vector in $\mathbb{R}^d$, and let

$$\vec{X} = \begin{bmatrix} \vec{X}_A \\ \vec{X}_B \end{bmatrix}, \qquad \vec{\mu} = \begin{bmatrix} \vec{\mu}_A \\ \vec{\mu}_B \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{A,A} & \Sigma_{A,B} \\ \Sigma_{B,A} & \Sigma_{B,B} \end{bmatrix}$$

Question: For any $\vec{x}_B \in \mathbb{R}^q$, what is the conditional distribution of $\vec{X}_A$ given $\vec{X}_B = \vec{x}_B$?

Conditioning

Partition $\{1, \ldots, d\} = A \cup B$, with $A = \{1, \ldots, p\}$ and $B = \{p+1, \ldots, p+q\}$ where $p + q = d$

Let $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$ be multivariate Gaussian random vector in $\mathbb{R}^d$, and let

$$\vec{X} = \begin{bmatrix} \vec{X}_A \\ \vec{X}_B \end{bmatrix}, \qquad \vec{\mu} = \begin{bmatrix} \vec{\mu}_A \\ \vec{\mu}_B \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{A,A} & \Sigma_{A,B} \\ \Sigma_{B,A} & \Sigma_{B,B} \end{bmatrix}$$

Question: For any $\vec{x}_B \in \mathbb{R}^q$, what is the conditional distribution of $\vec{X}_A$ given $\vec{X}_B = \vec{x}_B$?

- Miracle 1: It is a Gaussian distribution in $\mathbb{R}^p$

Conditioning

Partition $\{1, \ldots, d\} = A \cup B$, with $A = \{1, \ldots, p\}$ and $B = \{p+1, \ldots, p+q\}$ where $p + q = d$

Let $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$ be multivariate Gaussian random vector in $\mathbb{R}^d$, and let

$$\vec{X} = \begin{bmatrix} \vec{X}_A \\ \vec{X}_B \end{bmatrix}, \qquad \vec{\mu} = \begin{bmatrix} \vec{\mu}_A \\ \vec{\mu}_B \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{A,A} & \Sigma_{A,B} \\ \Sigma_{B,A} & \Sigma_{B,B} \end{bmatrix}$$

Question: For any $\vec{x}_B \in \mathbb{R}^q$, what is the conditional distribution of $\vec{X}_A$ given $\vec{X}_B = \vec{x}_B$?

- Miracle 1: It is a Gaussian distribution in $\mathbb{R}^p$
- Miracle 2: Conditional mean is affine function of $\vec{x}_B$

Conditioning

Partition $\{1, \ldots, d\} = A \cup B$, with $A = \{1, \ldots, p\}$ and $B = \{p+1, \ldots, p+q\}$ where $p + q = d$

Let $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$ be multivariate Gaussian random vector in $\mathbb{R}^d$, and let

$$\vec{X} = \begin{bmatrix} \vec{X}_A \\ \vec{X}_B \end{bmatrix}, \qquad \vec{\mu} = \begin{bmatrix} \vec{\mu}_A \\ \vec{\mu}_B \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \Sigma_{A,A} & \Sigma_{A,B} \\ \Sigma_{B,A} & \Sigma_{B,B} \end{bmatrix}$$

Question: For any $\vec{x}_B \in \mathbb{R}^q$, what is the conditional distribution of $\vec{X}_A$ given $\vec{X}_B = \vec{x}_B$?

- Miracle 1: It is a Gaussian distribution in $\mathbb{R}^p$
- Miracle 2: Conditional mean is affine function of $\vec{x}_B$
- Miracle 3: Conditional covariance does not depend on $\vec{x}_B$

Special case: Inference about a single variable

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

Special case: Inference about a single variable

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\mathbb{E}[X_A \mid \vec{X}_B = \vec{x}_B]$?

Special case: Inference about a single variable

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\mathbb{E}[X_A \mid \vec{X}_B = \vec{x}_B]$?

  Miracle 2 says: There's some $\vec{w} \in \mathbb{R}^{d-1}$ s.t. $\mathbb{E}[X_A \mid \vec{X}_B = \vec{x}_B] = \vec{x}_B \cdot \vec{w}$ for all $\vec{x}_B \in \mathbb{R}^{d-1}$

Special case: Inference about a single variable

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\mathbb{E}[X_A \mid \vec{X}_B = \vec{x}_B]$?

  Miracle 2 says: There's some $\vec{w} \in \mathbb{R}^{d-1}$ s.t. $\mathbb{E}[X_A \mid \vec{X}_B = \vec{x}_B] = \vec{x}_B \cdot \vec{w}$ for all $\vec{x}_B \in \mathbb{R}^{d-1}$

  Hence, this $\vec{w}$ is the minimizer of

$$J(\vec{w}) = \mathbb{E}[(\vec{X}_B \cdot \vec{w} - X_A)^2]$$

Special case: Inference about a single variable

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\mathbb{E}[X_A \mid \vec{X}_B = \vec{x}_B]$?

  Miracle 2 says: There's some $\vec{w} \in \mathbb{R}^{d-1}$ s.t. $\mathbb{E}[X_A \mid \vec{X}_B = \vec{x}_B] = \vec{x}_B \cdot \vec{w}$ for all $\vec{x}_B \in \mathbb{R}^{d-1}$

  Hence, this $\vec{w}$ is the minimizer of

  $$J(\vec{w}) = \mathbb{E}[(\vec{X}_B \cdot \vec{w} - X_A)^2]$$

  The minimizer of J via solving population normal equations:

  $$\begin{aligned} \vec{w} &= \mathbb{E}[\vec{X}_B \vec{X}_B^\intercal]^{-1} \mathbb{E}[\vec{X}_B X_A] \\ &= \Sigma_{B,B}^{-1} \Sigma_{B,A} \end{aligned}$$

Special case: Inference about a single variable

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\mathbb{E}[X_A \mid \vec{X}_B = \vec{x}_B]$?

  Miracle 2 says: There's some $\vec{w} \in \mathbb{R}^{d-1}$ s.t. $\mathbb{E}[X_A \mid \vec{X}_B = \vec{x}_B] = \vec{x}_B \cdot \vec{w}$ for all $\vec{x}_B \in \mathbb{R}^{d-1}$

  Hence, this $\vec{w}$ is the minimizer of

  $$J(\vec{w}) = \mathbb{E}[(\vec{X}_B \cdot \vec{w} - X_A)^2]$$

  The minimizer of J via solving population normal equations:

  $$\begin{aligned} \vec{w} &= \mathbb{E}[\vec{X}_B \vec{X}_B^{\intercal}]^{-1} \mathbb{E}[\vec{X}_B X_A] \\ &= \Sigma_{B,B}^{-1} \Sigma_{B,A} \end{aligned}$$

  So the conditional mean is:

  $$\mathbb{E}[X_A \mid \vec{X}_B = \vec{x}_B] = \Sigma_{A,B} \Sigma_{B,B}^{-1} \vec{x}_B$$

Special case: Inference about a single variable (continued)

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\mathrm{var}(X_A \mid \vec{X}_B = \vec{x}_B)$?

Special case: Inference about a single variable (continued)

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\operatorname{var}(X_A \mid \vec{X}_B = \vec{x}_B)$?

  Miracle 3 says: It doesn't depend on $\vec{x}_B$, so

$$\operatorname{var}(X_A \mid \vec{X}_B = \vec{x}_B) \;=\; \mathbb{E}[\operatorname{var}(X_A \mid \vec{X}_B)]$$

Special case: Inference about a single variable (continued)

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\mathrm{var}(X_A \mid \vec{X}_B = \vec{x}_B)$?

  Miracle 3 says: It doesn't depend on $\vec{x}_B$, so

$$\mathrm{var}(X_A \mid \vec{X}_B = \vec{x}_B) \;=\; \mathbb{E}[\mathrm{var}(X_A \mid \vec{X}_B)]$$

  Recall: "Law of total variance"

$$\mathrm{var}(X_A) \;=\; \mathbb{E}[\mathrm{var}(X_A \mid \vec{X}_B)] + \mathrm{var}(\mathbb{E}[X_A \mid \vec{X}_B])$$

Special case: Inference about a single variable (continued)

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\operatorname{var}(X_A \mid \vec{X}_B = \vec{x}_B)$?

  Miracle 3 says: It doesn't depend on $\vec{x}_B$, so

  $$\operatorname{var}(X_A \mid \vec{X}_B = \vec{x}_B) \;=\; \mathbb{E}[\operatorname{var}(X_A \mid \vec{X}_B)]$$

  Recall: "Law of total variance"

  $$\operatorname{var}(X_A) \;=\; \mathbb{E}[\operatorname{var}(X_A \mid \vec{X}_B)] + \operatorname{var}(\mathbb{E}[X_A \mid \vec{X}_B])$$

  Hence (with $\vec{w}$ from previous slide),

  $$\operatorname{var}(X_A \mid \vec{X}_B = \vec{x}_B) \quad = \quad \operatorname{var}(X_A) - \operatorname{var}(\mathbb{E}[X_A \mid \vec{X}_B])$$

Special case: Inference about a single variable (continued)

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\mathrm{var}(X_A \mid \vec{X}_B = \vec{x}_B)$?

  Miracle 3 says: It doesn't depend on $\vec{x}_B$, so

  $$\mathrm{var}(X_A \mid \vec{X}_B = \vec{x}_B) \;=\; \mathbb{E}[\mathrm{var}(X_A \mid \vec{X}_B)]$$

  Recall: "Law of total variance"

  $$\mathrm{var}(X_A) \;=\; \mathbb{E}[\mathrm{var}(X_A \mid \vec{X}_B)] + \mathrm{var}(\mathbb{E}[X_A \mid \vec{X}_B])$$

  Hence (with $\vec{w}$ from previous slide),

  $$\begin{aligned} \mathrm{var}(X_A \mid \vec{X}_B = \vec{x}_B) &= \mathrm{var}(X_A) - \mathrm{var}(\mathbb{E}[X_A \mid \vec{X}_B]) \\ &= \mathrm{var}(X_A) - \mathrm{var}(\vec{X}_B \cdot \vec{w}) \end{aligned}$$

Special case: Inference about a single variable (continued)

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\mathrm{var}(X_A \mid \vec{X}_B = \vec{x}_B)$?

  Miracle 3 says: It doesn't depend on $\vec{x}_B$, so

  $$\mathrm{var}(X_A \mid \vec{X}_B = \vec{x}_B) \;=\; \mathbb{E}[\mathrm{var}(X_A \mid \vec{X}_B)]$$

  Recall: "Law of total variance"

  $$\mathrm{var}(X_A) \;=\; \mathbb{E}[\mathrm{var}(X_A \mid \vec{X}_B)] + \mathrm{var}(\mathbb{E}[X_A \mid \vec{X}_B])$$

  Hence (with $\vec{w}$ from previous slide),

  $$\begin{aligned} \mathrm{var}(X_A \mid \vec{X}_B = \vec{x}_B) &= \mathrm{var}(X_A) - \mathrm{var}(\mathbb{E}[X_A \mid \vec{X}_B]) \\ &= \mathrm{var}(X_A) - \mathrm{var}(\vec{X}_B \cdot \vec{w}) \\ &= \mathrm{var}(X_A) - \vec{w}^{\mathsf{T}} \mathrm{cov}(\vec{X}_B)\vec{w} \end{aligned}$$

Special case: Inference about a single variable (continued)

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

- What is $\text{var}(X_A \mid \vec{X}_B = \vec{x}_B)$?

  Miracle 3 says: It doesn't depend on $\vec{x}_B$, so

$$\text{var}(X_A \mid \vec{X}_B = \vec{x}_B) \ = \ \mathbb{E}[\text{var}(X_A \mid \vec{X}_B)]$$

  Recall: "Law of total variance"

$$\text{var}(X_A) \ = \ \mathbb{E}[\text{var}(X_A \mid \vec{X}_B)] + \text{var}(\mathbb{E}[X_A \mid \vec{X}_B])$$

  Hence (with $\vec{w}$ from previous slide),

$$\begin{aligned} \text{var}(X_A \mid \vec{X}_B = \vec{x}_B) &= \text{var}(X_A) - \text{var}(\mathbb{E}[X_A \mid \vec{X}_B]) \\ &= \text{var}(X_A) - \text{var}(\vec{X}_B \cdot \vec{w}) \\ &= \text{var}(X_A) - \vec{w}^{\mathsf{T}} \operatorname{cov}(\vec{X}_B)\vec{w} \\ &= \Sigma_{A,A} - \Sigma_{A,B}\Sigma_{B,B}^{-1}\Sigma_{B,A} \end{aligned}$$

Special case: Inference about a single variable (conclusion)

Special case: $A = \{1\}$, $B = \{2, \dots, d\}$, $\vec{\mu} = \vec{0}$.

Conclusion:

$$(X_A \mid \vec{X}_B = \vec{x}_B) \;\sim\; \mathrm{N}\left(\Sigma_{A,B}\Sigma_{B,B}^{-1}\vec{x}_B,\; \Sigma_{A,A} - \Sigma_{A,B}\Sigma_{B,B}^{-1}\Sigma_{B,A}\right)$$

Special case: Inference about a single variable (conclusion)

Special case: $A = \{1\}$, $B = \{2, \ldots, d\}$, $\vec{\mu} = \vec{0}$.

Conclusion:

$$(X_A \mid \vec{X}_B = \vec{x}_B) \;\sim\; \mathrm{N}\left(\Sigma_{A,B}\Sigma_{B,B}^{-1}\vec{x}_B,\; \Sigma_{A,A} - \Sigma_{A,B}\Sigma_{B,B}^{-1}\Sigma_{B,A}\right)$$

Generalization to arbitrary A and B is similar (and allowing $\vec{\mu} \neq \vec{0}$):

$$(\vec{X}_A \mid \vec{X}_B = \vec{x}_B) \;\sim\; \mathrm{N}\left(\vec{\mu}_A + \Sigma_{A,B}\Sigma_{B,B}^{-1}(\vec{x}_B - \vec{\mu}_B),\; \Sigma_{A,A} - \Sigma_{A,B}\Sigma_{B,B}^{-1}\Sigma_{B,A}\right)$$

Estimation

IID sample from a multivariate Gaussian distribution in $\mathbb{R}^d$
$\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n \sim_{\text{iid}} \mathrm{N}(\vec{\mu}, \Sigma)$

- Parameters: $\vec{\mu} \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$
- Maximum likelihood estimation:
  Given data $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n \in \mathbb{R}^d$, MLE for $\vec{\mu}$ and Σ are

$$\vec{\mu}_{\text{mle}} := \frac{1}{n} \sum_{i=1}^{n} \vec{x}_i, \qquad \Sigma_{\text{mle}} := \frac{1}{n} \sum_{i=1}^{n} (\vec{x}_i - \vec{\mu}_{\text{mle}})(\vec{x}_i - \vec{\mu}_{\text{mle}})^{\mathsf{T}}$$

Example: Dartmouth data

$\vec{X} = (X_1, X_2, X_3) = (\texttt{SAT Score}, \texttt{HS GPA}, \texttt{College GPA})$

Example: Dartmouth data

$\vec{X} = (X_1, X_2, X_3) = (\texttt{SAT Score}, \texttt{HS GPA}, \texttt{College GPA})$

(Sample) mean and covariance:

$$\vec{\mu} = \begin{bmatrix} 1028.59 \\ 3.18324 \\ 2.46031 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 20602.8 & 34.2707 & 51.6484 \\ 34.2707 & 0.30857 & 0.226406 \\ 51.6484 & 0.226406 & 0.556484 \end{bmatrix}$$

Example: Dartmouth data

$$\vec{X} = (X_1, X_2, X_3) = (\texttt{SAT Score}, \texttt{HS GPA}, \texttt{College GPA})$$

(Sample) mean and covariance:

$$\vec{\mu} = \begin{bmatrix} 1028.59 \\ 3.18324 \\ 2.46031 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 20602.8 & 34.2707 & 51.6484 \\ 34.2707 & 0.30857 & 0.226406 \\ 51.6484 & 0.226406 & 0.556484 \end{bmatrix}$$

If we pretend $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$,

$$\left(\begin{bmatrix} \texttt{HS GPA} \\ \texttt{College GPA} \end{bmatrix} \mid \texttt{SAT Score} = 1600 \right) \sim \mathrm{N}\left(\begin{bmatrix} 4.1337 \\ 3.8928 \end{bmatrix}, \begin{bmatrix} 0.2516 & 0.1405 \\ 0.1405 & 0.4270 \end{bmatrix} \right)$$

Example: Dartmouth data

$\vec{X} = (X_1, X_2, X_3) = (\texttt{SAT Score}, \texttt{HS GPA}, \texttt{College GPA})$

(Sample) mean and covariance:

$$\vec{\mu} = \begin{bmatrix} 1028.59 \\ 3.18324 \\ 2.46031 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 20602.8 & 34.2707 & 51.6484 \\ 34.2707 & 0.30857 & 0.226406 \\ 51.6484 & 0.226406 & 0.556484 \end{bmatrix}$$

If we pretend $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$,

$$\left(\begin{bmatrix} \texttt{HS GPA} \\ \texttt{College GPA} \end{bmatrix} \mid \texttt{SAT Score} = 1600 \right) \sim \mathrm{N}\left(\begin{bmatrix} 4.1337 \\ 3.8928 \end{bmatrix}, \begin{bmatrix} 0.2516 & 0.1405 \\ 0.1405 & 0.4270 \end{bmatrix} \right)$$

Conditional correlation matrix

$$\begin{bmatrix} 1.0000 & 0.4287 \\ 0.4287 & 1.0000 \end{bmatrix}$$

Example: Dartmouth data

$\vec{X} = (X_1, X_2, X_3) = (\texttt{SAT Score}, \texttt{HS GPA}, \texttt{College GPA})$

(Sample) mean and covariance:

$$\vec{\mu} = \begin{bmatrix} 1028.59 \\ 3.18324 \\ 2.46031 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 20602.8 & 34.2707 & 51.6484 \\ 34.2707 & 0.30857 & 0.226406 \\ 51.6484 & 0.226406 & 0.556484 \end{bmatrix}$$

If we pretend $\vec{X} \sim \mathrm{N}(\vec{\mu}, \Sigma)$,

$$\left(\begin{bmatrix} \texttt{HS GPA} \\ \texttt{College GPA} \end{bmatrix} \mid \texttt{SAT Score} = 1600 \right) \sim \mathrm{N}\left(\begin{bmatrix} 4.1337 \\ 3.8928 \end{bmatrix}, \begin{bmatrix} 0.2516 & 0.1405 \\ 0.1405 & 0.4270 \end{bmatrix} \right)$$

Conditional correlation matrix

$$\begin{bmatrix} 1.0000 & 0.4287 \\ 0.4287 & 1.0000 \end{bmatrix}$$

(But, of course, the Gaussian distribution is not reasonable here)

Summary

- Multivariate Gaussians: Affine transformations of standard (isotropic) multivariate Gaussian
- Marginalization and conditioning $\longrightarrow$ resulting distribution is Gaussian!