# THE MULTIVARIATE GAUSSIAN DISTRIBUTION

Let $\underline{\mu} \in \mathbb{R}^N$ and $\Sigma \in \mathbb{R}^{N \times N}$ be symmetric and positive definite. A random variable $\underline{X}$ has a multivariate Gaussian (or normal) distribution iff its pdf is

$$\begin{aligned}\phi(\underline{x}) &= \phi(\underline{x}\,;\,\underline{\mu}, \Sigma) \\ &= (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})\right\}.\end{aligned}$$

How can we show that this is a valid density, i.e., that $\int_{\mathbb{R}^N} \phi(\underline{x})\, d\underline{x} = 1$ ?

(a)

Notation: $\underline{X} \sim N(\underline{\mu}, \Sigma)$

The MVG model is the most important and widely employed model in statistical signal processing. Reasons include:

- Tractability
- Estimators and detectors with intuitive forms and properties
- Often effective in practice
- Justification in terms of the central limit theorem.

## CLT

If $\underline{X} = \frac{1}{n} \sum_{i=1}^{n} \underline{Y}_i$, then for $n$ sufficiently large,

$$X \approx N(\underline{\mu}, \Sigma)$$

for some $\underline{\mu}, \Sigma$, regardless of the distribution of $\underline{Y}$.

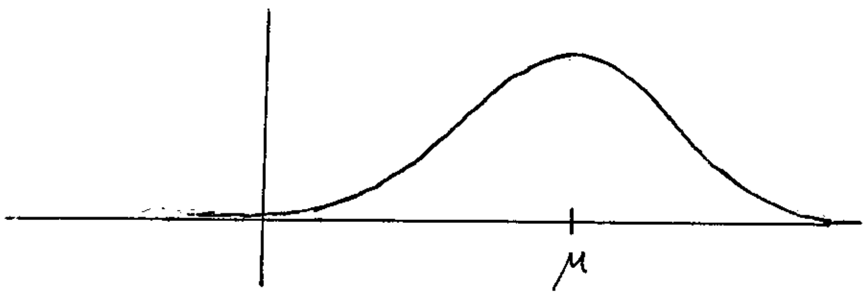
rough paraphrase

**Example** In communication systems, electronic noise is due to the aggregate effect of huge numbers of charge carriers undergoing random motion.

## Conceptualization

In 1-d, $\Sigma = [\sigma^2]$ $(1 \times 1)$ and

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



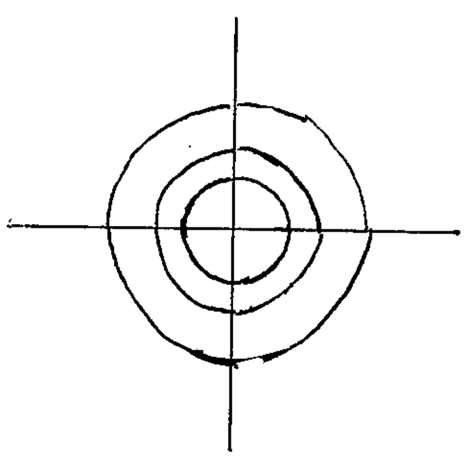

bell curve

In 2-d, let's consider 3 cases:

<u>Case 1</u>: $\Sigma = \sigma^2 I_{2\times 2} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$

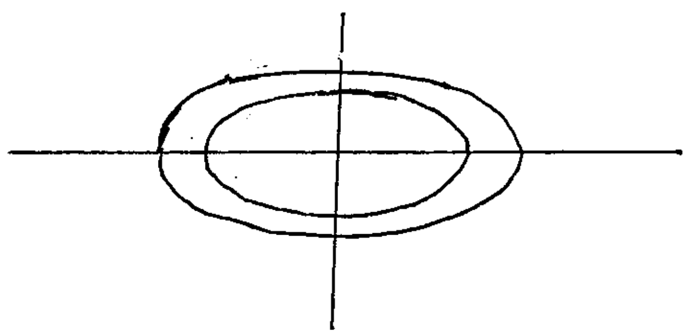
Then a contour of the density is a <u>circle</u>:

$$f(\underline{x}) = \gamma \iff \| \underline{x} - \underline{\mu} \|^2 = \gamma'$$



<u>Case 2</u> : $\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$ , say $\sigma_1 > \sigma_2$

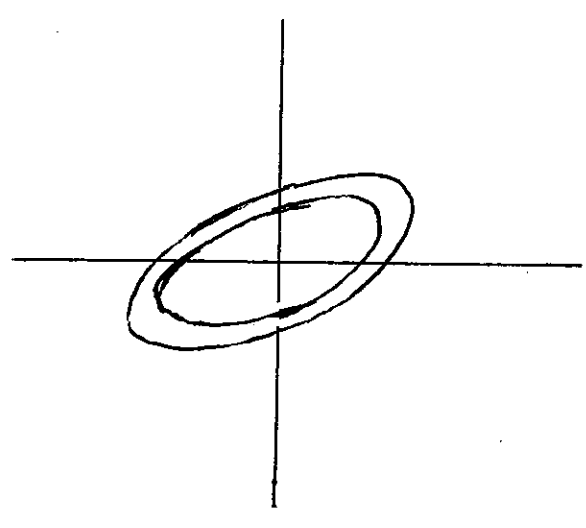
Then the density contours are <u>ellipses</u> whose axes align with the standard basis.



To see this, observe

$$f(\underline{x}) = \gamma \iff \frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} = \gamma'$$

<u>Case 3</u> : $\Sigma$ is arbitrary. Then density contours are ellipses with arbitrary orientation.



To see this, write

$$\Sigma = U \Lambda U^T$$

Then

$$(\underline{x}-\underline{\mu})^T \; {}^{-1} (\underline{x}-\underline{\mu})$$

$$= (\underline{x}-\underline{\mu})^T U \Lambda^{-1} U^T (\underline{x}-\underline{\mu})$$

$$= (\underline{x}'-\underline{\mu}')^T \Lambda^{-1} (\underline{x}'-\underline{\mu}')$$

$$\left[\text{where } \underline{x}' = U^T \underline{x}, \; \underline{\mu}' = U^T \underline{\mu}\right]$$

$$= \frac{(x_1'-\mu_1')^2}{\lambda_1} + \frac{(x_2'-\mu_2')^2}{\lambda_2}$$

which defines an ellipse in the rotated coordinate system.

Clearly this idea generalizes to multiple dimensions.

## Linear Transformations

One of the most important properties of the MVG distribution is that

"an linear transformation of a MVG is another MVG."

From this property we can deduce many other properties of MVGs.

Before we get there, we need the following technical device

### Characteristic Function

The characteristic function of an N dimensional random variable $\underline{X}$ with pdf $f(\underline{x})$ is

$$\Phi_{\underline{X}}(\underline{\omega}) = E\left[e^{-j\underline{\omega}^T\underline{X}}\right]$$

$$= \int e^{-j\underline{\omega}^T\underline{x}} f(\underline{x})\, d\underline{x}$$

The char. fun. is an N-dim Fourier transform of the density of $\underline{X}$. Thus it uniquely characterizes the random variable. The density may be recovered from $\Phi$ by taking the inverse Fourier transform.

For the MVG, $\underline{X} \sim \mathcal{N}(\underline{\mu}, \Sigma)$ we have

$$\Phi_{\underline{X}}(\underline{\omega}) = E\left[e^{-j\underline{\omega}^T \underline{X}}\right]$$

$$= \int e^{-j\underline{\omega}^T \underline{x}} \phi(\underline{x})\, d\underline{x}$$

$$= \int (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-j\underline{\omega}^T \underline{x} - \frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})\right\} d\underline{x}$$

$\boxed{\text{complete the square}} \rightarrow$

$$= e^{-j\underline{\omega}^T \underline{\mu} - \frac{1}{2}\underline{\omega}^T \Sigma \underline{\omega}} \times$$

$$\underbrace{\int (2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\underline{x}-\underline{\mu}+j\Sigma\underline{\omega})^T \Sigma^{-1} (\underline{x}-\underline{\mu}+j\Sigma\underline{\omega})\right\} d\underline{x}}_{\text{Gaussian density} \rightarrow 1}$$

$$= e^{-j\underline{\omega}^T \underline{\mu} - \frac{1}{2}\underline{\omega}^T \Sigma \underline{\omega}}$$

# Linear Transformations

**Proposition** If $\underline{X} \sim \mathcal{N}(\underline{\mu}, \Sigma)$ is $N$-dim, $A \in \mathbb{R}^{M \times N}$,

and $\underline{Y} = A\underline{X}$, then

$$\underline{Y} \sim \mathcal{N}(A\underline{\mu}, A\Sigma A^T)$$

**Proof**

$$\begin{aligned}
\Phi_{\underline{Y}}(\underline{\omega}) &= E\left[e^{-j\underline{\omega}^T \underline{Y}}\right] \\
&= E\left[e^{-j\underline{\omega}^T A \underline{X}}\right] \\
&= E\left[e^{-j(A^T\underline{\omega})^T \underline{X}}\right] \\
&= \Phi_{\underline{X}}(A^T\underline{\omega}) \\
&= e^{-j\underline{\omega}^T A\underline{\mu} - \frac{1}{2}\underline{\omega}^T A\Sigma A^T \underline{\omega}}
\end{aligned}$$

$$\Longrightarrow \quad \underline{Y} \sim \mathcal{N}(A\underline{\mu}, A\Sigma A^T)$$

since the characteristic function uniquely characterizes a distribution.

Additional Properties

## Mean

(b) If $\underline{X} \sim N(\underline{\mu}, \Sigma)$, then $E[\underline{X}] =$

Proof: $\phi$ is symmetric w.r.t $\mu$, i.e. $\phi(\mu+c) = \phi(\mu-c)$.

## Covariance

If $\underline{X} \sim N(\mu, \Sigma)$, then $\operatorname{cov}(\underline{X}) =$

(c) $E[(\underline{X}-\underline{\mu})(\underline{X}-\underline{\mu})^T] =$

Proof: First consider $\underline{Z} \sim N(0, I)$.

Easy to show $\operatorname{Cov}(\underline{Z}) = I$. By linear transform property,

$$\mathcal{U}\Lambda^{\frac{1}{2}}\underline{Z} + \underline{\mu}$$

has the same distribution as $\underline{X}$, where $\Sigma = \mathcal{U}\Lambda\mathcal{U}^T$. Now

$$\begin{aligned}
\operatorname{Cov}(\mathcal{U}\Lambda^{\frac{1}{2}}\underline{Z} + \mu) &= E[(\mathcal{U}\Lambda^{\frac{1}{2}}\underline{Z})(\mathcal{U}\Lambda^{\frac{1}{2}}\underline{Z})^T] \\
&= \mathcal{U}\Lambda^{\frac{1}{2}} \cdot E[\underline{Z}\underline{Z}^T]\Lambda^{\frac{1}{2}}\mathcal{U}^T \\
&= \mathcal{U}\Lambda\mathcal{U}^T \\
&= \Sigma
\end{aligned}$$

# Marginals

**Proposition** Let $\underline{X} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}$, $\underline{\mu} = \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix}$, $= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$.

If $\underline{X} \sim N(\underline{\mu}, \Sigma)$, then

$$\underline{X}_1 \sim N(\underline{\mu}_1, \Sigma_{11}).$$

**Exercise** Prove this.

Solution | Write $\underline{X}_1 = A\underline{X}$ where

$$A = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & & O \\ & & & 1 & \end{bmatrix} \qquad (p \times N)$$

assuming $\underline{X} \in \mathbb{R}^N$, $\underline{X}_1 \in \mathbb{R}^p$.

Then

$$A\underline{\mu} = \underline{\mu}_1$$

$$A \Sigma A^T = \Sigma_{11}$$

*Now apply the linear transformation property.*

## Conditioning

Proposition | Let $\underline{X} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}$, $\underline{\mu} = \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix}$, $= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$.

If $\underline{X} \sim N(\underline{\mu}, \Sigma)$, then

$$\underline{X}_2 \mid \underline{X}_1 = \underline{x}_1 \sim N(\tilde{\underline{\mu}}, \tilde{\Sigma})$$

where

$$\tilde{\underline{\mu}} = \underline{\mu}_2 + \Sigma_{21} \Sigma_{11}^{-1} (\underline{x}_1 - \underline{\mu}_1)$$

$$\tilde{\Sigma} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$$

Proof | Write out $f(\underline{x}_2 | \underline{x}_1) = \dfrac{f(\underline{x})}{f(\underline{x}_1)}$

and simplify. See Kay (Vol. I) or Moon and Stirling for details.

$\boxed{\text{Key}}$ (a) First, show ~~$\int \phi(\underline{y})\, d\underline{y} = 1$ for the case~~ $\Sigma = I$, which is easy because the integral into a product of integrals of 1-d Gaussian pdfs. Next apply the general transformation of variables result from multivariable calculus: If $g : \mathbb{R}^N \to \mathbb{R}$, $F : \mathbb{R}^M \to \mathbb{R}^N$ (invertible), then

$$\int_{\mathbb{R}^N} g(\underline{y})\, d\underline{y} = \int g(F(\underline{x})) \cdot |J_F(\underline{x})|\, dx$$

where $J_F$ is the Jacobian. If $F$ is linear, say $F(\underline{x}) = A\underline{x}$, then $J_F = A$. Now apply this with $A = \Sigma^{-\frac{1}{2}} = U\Lambda^{-\frac{1}{2}}$ and $g(\underline{y}) = \phi(\underline{y}; \mu, I)$.

(b) $\mu$

(c) $\Sigma$