Random Vectors

Yuxin Chen

Princeton University, Fall 2018

Outline

- Random vectors
- Gaussian random vectors

Random vectors

- Let $X_1, X_2, \ldots, X_n$ be r.v.s defined on the same sample space. We define a random vector (RV) as

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

- $\boldsymbol{X}$ is completely specified by its joint CDF for $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$:

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}\{X_1 \leq x_1,\, X_2 \leq x_2,\, \ldots,\, X_n \leq x_n\}, \quad \boldsymbol{x} \in \mathbb{R}^n$$

Random vectors

- If $\boldsymbol{X}$ is continuous, i.e. $F_{\boldsymbol{X}}(\boldsymbol{x})$ is a continuous function of $\boldsymbol{x}$, then $\boldsymbol{X}$ can be specified by its joint PDF

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$$

- If $\boldsymbol{X}$ is discrete then it can be specified by its joint PMF

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = p_{X_1,\cdots,X_n}(x_1,\cdots,x_n)$$

Random vectors

- A marginal CDF (PDF, PMF) is the joint CDF (PDF, PMF) for a subset of $\{X_1, \ldots, X_n\}$; e.g. for

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

the marginals are

$$f_{X_1}(x_1), \quad f_{X_2}(x_2), \quad f_{X_3}(x_3)$$
$$f_{X_1,X_2}(x_1, x_2), \quad f_{X_1,X_3}(x_1, x_3), \quad f_{X_2,X_3}(x_2, x_3)$$

- The marginals can be obtained from the joint in the usual way. For the previous example,

$$F_{X_1}(x_1) = \lim_{x_2, x_3 \to \infty} F_{\boldsymbol{X}}(x_1, x_2, x_3)$$
$$f_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2,X_3}(x_1, x_2, x_3) \mathrm{d}x_3$$

Random vectors

- Conditional CDF (PDF, PMF) can also be defined in the usual way; e.g. the conditional PDF of $\boldsymbol{X}_{k+1}^n = (X_{k+1}, \dots, X_n)$ given $\boldsymbol{X}_1^k = (X_1, \dots, X_k)$ is

$$f_{\boldsymbol{X}_{k+1}^n | \boldsymbol{X}_1^k}(\boldsymbol{x}_{k+1}^n | \boldsymbol{x}_1^k) = \frac{f_{\boldsymbol{X}}(x_1, x_2, \dots, x_n)}{f_{\boldsymbol{X}_1^k}(x_1, x_2, \dots, x_k)} = \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{\boldsymbol{X}_1^k}(\boldsymbol{x}_1^k)}$$

- Chain rule:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) \cdots f_{X_n|\boldsymbol{X}_1^{n-1}}(x_n|\boldsymbol{x}_1^{n-1})$$

Independence and conditional independence

- Independence is defined in the usual way: $X_1, \cdots, X_n$ are independent if

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} f_{X_i}(x_i) \quad \text{for all } x_1, \cdots, x_n$$

- $X_1, \cdots, X_n$ are said to be independent, identically distributed (i.i.d.) if they are independent and have the same marginals
  - Example: if we flip a coin n times independently, we generate i.i.d. Bern(p) r.v.s. $X_1, \cdots, X_n$

Mean and covariance matrix

- The mean of the random vector $\boldsymbol{X}$ is defined as

$$\mathbb{E}[\boldsymbol{X}] = \Big[\mathbb{E}[X_1], \mathbb{E}[X_2], \cdots, \mathbb{E}[X_n]\Big]^\top$$

- Denote the covariance between X_i and X_j, $\mathsf{Cov}(X_i, X_j)$, by σ_{ij} (so the variance of X_i is denoted by σ_{ii}, $\mathsf{Var}(X_i)$, or $\sigma_{X_i}^2$)

- The covariance matrix of $\boldsymbol{X}$ is defined as

$$\mathsf{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}_{\boldsymbol{X}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

- For $n = 2$, $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ can be expressed via correlation coefficients:

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{X_1}^2 & \rho_{X_1,X_2}\sigma_{X_1}\sigma_{X_2} \\ \rho_{X_1,X_2}\sigma_{X_1}\sigma_{X_2} & \sigma_{X_2}^2 \end{bmatrix}$$

Properties of covariance matrix

- $\mathbf{\Sigma}_{\boldsymbol{X}}$ is real and symmetric (since $\sigma_{ij} = \sigma_{ji}$)
- $\mathbf{\Sigma}_{\boldsymbol{X}}$ is positive semidefinite ($\mathbf{\Sigma}_{\boldsymbol{X}} \succeq \mathbf{0}$), i.e. the quadratic form

$$\boldsymbol{a}^\top \mathbf{\Sigma}_{\boldsymbol{X}} \boldsymbol{a} \geq 0 \quad \text{for every real vector } \boldsymbol{a}$$

  This implies that all eigenvalues of $\mathbf{\Sigma}_{\boldsymbol{X}}$ are nonnegative, and also all leading principal minors are nonnegative

  - The kth leading principal minor of a matrix is the determinant of its upper-left k by k sub-matrix

Properties of covariance matrix

- To show that $\boldsymbol{\Sigma}_{\boldsymbol{X}} \succeq \mathbf{0}$, we write

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^\top\right]$$

  Thus

$$\begin{aligned}
\boldsymbol{a}^\top \boldsymbol{\Sigma}_{\boldsymbol{X}} \boldsymbol{a} &= \boldsymbol{a}^\top \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^\top\right] \boldsymbol{a} \\
&= \mathbb{E}\left[\boldsymbol{a}^\top (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^\top \boldsymbol{a}\right] \\
&= \mathbb{E}\left[(\boldsymbol{a}^\top (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}]))^2\right] \ \geq \ 0
\end{aligned}$$

Which of the following can be a covariance matrix?

1. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{bmatrix}$

4. $\begin{bmatrix} -1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

5. $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}$

6. $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$

Square root of covariance matrix

Let $\boldsymbol{\Sigma}$ be a covariance matrix. Then there exists an $n \times n$ positive semidefinite matrix $\boldsymbol{\Sigma}^{1/2}$ such that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\Sigma}^{1/2})$. The matrix $\boldsymbol{\Sigma}^{1/2}$ is called the square root of $\boldsymbol{\Sigma}$

- Since $\boldsymbol{\Sigma} \succeq \mathbf{0}$, its eigendecomposition can be written as $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$, where each column of the *orthonormal* matrix $\boldsymbol{U}$ is an eigenvector of $\boldsymbol{\Sigma}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues.
- Then $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^\top$ is the square root of $\boldsymbol{\Sigma}$

  Proof: Clearly, $\boldsymbol{M} \succeq \mathbf{0}$. Since $\boldsymbol{U}$ is orthonormal,

$$\begin{aligned}\boldsymbol{M}\boldsymbol{M} &= \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^\top = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^\top \\ &= \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top\end{aligned}$$

Coloring and whitening

- Let $\boldsymbol{X}$ be white RV, i.e. it has zero mean and $\boldsymbol{\Sigma}_{\boldsymbol{X}} = a\boldsymbol{I}$ for some $a > 0$.
- Let $\boldsymbol{\Sigma}$ be a covariance matrix, then the RV $\boldsymbol{Y} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{X}$ has covariance matrix $\boldsymbol{\Sigma}$

  Proof:

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{Y}} &= \mathbb{E}\left[(\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}])\,(\boldsymbol{Y} - \mathbb{E}[\boldsymbol{Y}])^\top\right] \\
&= \mathbb{E}\left[\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{X} - \mathbb{E}[\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{X}]\right)\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{X} - \mathbb{E}[\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{X}]\right)^\top\right] \\
&= \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])\,(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^\top\right]\boldsymbol{\Sigma}^{\frac{1}{2}} \\
&= \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{I}\boldsymbol{\Sigma}^{\frac{1}{2}} = \boldsymbol{\Sigma}
\end{aligned}
$$

  Hence we can generate a RV with any prescribed covariance from a white RV

Coloring and whitening

Whitening: Given a zero mean RV $\boldsymbol{Y}$ with nonsingular covariance matrix $\boldsymbol{\Sigma}$, then the RV $\boldsymbol{X} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}$ is white

- Hence, we can generate a white RV from any RV with nonsingular covariance matrix
- Coloring and whitening have applications in simulations, detection, and estimation

Gaussian random variables

A Gaussian r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$ has PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Gaussian has many nice properties
  - If $Y = aX + b$ for some constants a and b, then

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

  - If $X_1, \cdots, X_n$ are independent and $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, then

$$\sum_i a_i X_i \sim \mathcal{N}\left(\sum_i a_i \mu_i, \sum_i a_i^2 \sigma_i^2\right)$$

Central limit theorem (optional)

Let $X_1, \cdots, X_n$ be a sequence of i.i.d. r.v.s with mean μ_i and variance σ^2. Then the normalized sum

$$Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - \mathbb{E}[X_k]}{\sigma}$$

can be well-approximated by a standard Gaussian variable $\mathcal{N}(0,1)$ when $n \to \infty$

- This is an important reason for the popularity of Gaussians

Gaussian random vectors

- A random vector $\boldsymbol{X} = [X_1, \cdots, X_n]^\top$ is a Gaussian random vector (GRV) (or $X_1, \cdots, X_n$ are jointly Gaussian r.v.s) if the joint PDF is of the form

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\underbrace{(2\pi)^{\frac{n}{2}} \det(\boldsymbol{\Sigma})^{\frac{1}{2}}}_{\text{normalizing constant}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})},$$

  where $\boldsymbol{\mu}$ is the mean and $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{X}$, and $\det(\boldsymbol{\Sigma}) > 0$, i.e. $\boldsymbol{\Sigma}$ is positive definite

- Notation: $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a GRV with given mean and covariance matrix

Gaussian random vectors

- Since $\boldsymbol{\Sigma}$ is positive definite, $\boldsymbol{\Sigma}^{-1}$ is positive definite. Thus if $\boldsymbol{x} - \boldsymbol{\mu} \neq \mathbf{0}$,
$$(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) > 0\,,$$
which means that the contours of equal pdf are ellipsoids

- The GRV $\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, a\boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix and $a > 0$, is called white; its contours of equal joint PDF are spheres centered at the origin

Properties of Gaussian random vectors

Property 1: For a GRV, uncorrelation implies independence

- This can be verified by substituting $\sigma_{ij} = 0$ for all $i \neq j$ in the joint PDF.

  Then $\mathbf{\Sigma}$ becomes diagonal and so does $\mathbf{\Sigma}^{-1}$, and the joint PDF reduces to the product of the marginals $X_i \sim \mathcal{N}(\mu_i, \sigma_{ii})$

  For the white GRV $\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, a\boldsymbol{I})$, the r.v.s are i.i.d. $\mathcal{N}(0, a)$

Properties of Gaussian random vectors

Property 2: Linear transformation of a GRV yields a GRV, i.e. given any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ $(m \leq n)$ that has full rank m, then

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu},\, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top)$$

Properties of Gaussian random vectors

- **Example:** Let

$$\boldsymbol{X} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}\right)$$

Find the joint PDF of

$$\boldsymbol{Y} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{X}$$

- **Solution:** From Property 2, we conclude that

$$\boldsymbol{Y} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\right) = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 7 & 3 \\ 3 & 2 \end{bmatrix}\right)$$

Properties of Gaussian random vectors

Before we prove Property 2, let us show that

$$\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{A}\boldsymbol{\mu} \quad \text{and} \quad \boldsymbol{\Sigma}_{\boldsymbol{Y}} = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top$$

- **Proof:** These follow from linearity of expectation. First, expectation:

$$\mathbb{E}[\boldsymbol{Y}] = \mathbb{E}[\boldsymbol{A}\boldsymbol{X}] = \boldsymbol{A}\,\mathbb{E}[\boldsymbol{X}] = \boldsymbol{A}\boldsymbol{\mu}$$

Next consider the covariance matrix:

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{Y}} &= \mathbb{E}\left[(\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y}))(\boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y}))^\top\right] \\ &= \mathbb{E}\left[(\boldsymbol{A}\boldsymbol{X} - \boldsymbol{A}\boldsymbol{\mu})(\boldsymbol{A}\boldsymbol{X} - \boldsymbol{A}\boldsymbol{\mu})^\top\right] \\ &= \boldsymbol{A}\,\mathbb{E}\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top\right]\boldsymbol{A}^\top = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top\end{aligned}$$

Of course these are insufficient to show that $\boldsymbol{Y}$ is a GRV — we must also show that the joint PDF has the right form. We do so using the characteristic function for a random vector

Aside: characteristic function

Definition: If $\boldsymbol{X} \sim f_{\boldsymbol{X}}$, the characteristic function of $\boldsymbol{X}$ is

$$\Phi_{\boldsymbol{X}}(\boldsymbol{w}) \stackrel{\text{def}}{=} \mathbb{E}\left[e^{i\boldsymbol{w}^\top \boldsymbol{X}}\right],$$

where $\boldsymbol{w} \in \mathbb{R}^n$ and $i = \sqrt{-1}$

- It is seen that

$$\Phi_{\boldsymbol{X}}(\boldsymbol{w}) = \int f_{\boldsymbol{X}}(\boldsymbol{x}) e^{i\boldsymbol{w}^\top \boldsymbol{x}} \, \mathrm{d}\boldsymbol{x}$$

  This is the multi-dimensional inverse Fourier transform of $f_{\boldsymbol{X}}(\boldsymbol{x})$.

Aside: characteristic function

- **Uniqueness:** the characterstic function $\Phi_{\boldsymbol{X}}(\boldsymbol{w})$ uniquely specifies the joint distribution $f_{\boldsymbol{X}}(\boldsymbol{x})$ of $\boldsymbol{X}$.

- The joint PDF can be found by taking the Fourier transform of $\Phi_{\boldsymbol{X}}(\boldsymbol{w})$, i.e.

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \int \frac{1}{(2\pi)^n} \Phi_{\boldsymbol{X}}(\boldsymbol{w}) e^{-i\boldsymbol{w}^\top \boldsymbol{x}} \, \mathrm{d}\boldsymbol{w}$$

Aside: characteristic function

- **Example:** The characteristic function for $X \sim \mathcal{N}(\mu, \sigma^2)$ is

$$\Phi_X(\omega) = e^{-\frac{1}{2}\omega^2\sigma^2 + i\mu\omega},$$

  and for a GRV $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

$$\Phi_{\boldsymbol{X}}(\boldsymbol{w}) = e^{-\frac{1}{2}\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w} + i\boldsymbol{w}^\top \boldsymbol{\mu}}$$

- Joint CDF of a GRV is completely determined by the mean and the covariance matrix.

Characteristic function for Gaussians (optional)

Proof: Let $\boldsymbol{y} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{\mu})$, then $\boldsymbol{x} = \underbrace{\boldsymbol{\Sigma}^{\frac{1}{2}}}_{\text{Jacobian matrix}} \boldsymbol{y} + \boldsymbol{\mu}$.

Thus,

$$\mathrm{d}\boldsymbol{x} = \det(\boldsymbol{\Sigma}^{\frac{1}{2}})\mathrm{d}\boldsymbol{y} = \det\left(\boldsymbol{\Sigma}\right)^{\frac{1}{2}}\mathrm{d}\boldsymbol{y} \quad \text{(change of variables)}$$

and hence

$$\begin{aligned}\Phi_{\boldsymbol{X}}(\boldsymbol{w}) &= \frac{1}{(2\pi)^{\frac{n}{2}}\det(\boldsymbol{\Sigma})^{\frac{1}{2}}}\int \exp\left(i\boldsymbol{w}^\top \boldsymbol{x}\right)\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)\mathrm{d}\boldsymbol{x} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}}\int \exp\left(i\boldsymbol{w}^\top\left(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{y}+\boldsymbol{\mu}\right)\right)\exp\left(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{y}\right)\mathrm{d}\boldsymbol{y}\end{aligned}$$

Characteristic function for Gaussians (optional)

Further, setting $\boldsymbol{u} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{w}$ gives

$$
\begin{aligned}
\Phi_{\boldsymbol{X}}(\boldsymbol{w}) &= \frac{\exp\left(i\boldsymbol{w}^\top\boldsymbol{\mu}\right)}{(2\pi)^{\frac{n}{2}}}\int \exp\left(i\boldsymbol{u}^\top\boldsymbol{y}\right)\exp\left(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{y}\right)\mathrm{d}\boldsymbol{y} \\
&= \frac{\exp\left(i\boldsymbol{w}^\top\boldsymbol{\mu}\right)\exp\left(-\frac{1}{2}\boldsymbol{u}^\top\boldsymbol{u}\right)}{(2\pi)^{\frac{n}{2}}}\int \exp\left(-\frac{1}{2}(\boldsymbol{y}-i\boldsymbol{u})^\top(\boldsymbol{y}-i\boldsymbol{u})\right)\mathrm{d}\boldsymbol{y} \\
&= \exp\left(i\boldsymbol{w}^\top\boldsymbol{\mu}\right)\exp\left(-\frac{1}{2}\boldsymbol{u}^\top\boldsymbol{u}\right)\prod_{k=1}^{n}\int\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(y_k-iu_k)^2\right)\mathrm{d}y_k \\
&= \exp\left(i\boldsymbol{w}^\top\boldsymbol{\mu}\right)\exp\left(-\frac{1}{2}\boldsymbol{u}^\top\boldsymbol{u}\right) \\
&= \exp\left(i\boldsymbol{w}^\top\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}\right)
\end{aligned}
$$

Properties of Gaussian random vectors

Proof of Property 2: Compute the characteristic function of $\boldsymbol{Y}$ as follows

$$
\begin{aligned}
\Phi_{\boldsymbol{Y}}(\boldsymbol{w}) &= \mathbb{E}\left[e^{i\boldsymbol{w}^\top \boldsymbol{Y}}\right] \\
&= \mathbb{E}\left[e^{i\boldsymbol{w}^\top \boldsymbol{A}\boldsymbol{X}}\right] \\
&= \Phi_{\boldsymbol{X}}(\boldsymbol{A}^\top \boldsymbol{w}) \\
&= e^{-\frac{1}{2}(\boldsymbol{A}^\top \boldsymbol{w})^\top \boldsymbol{\Sigma}(\boldsymbol{A}^\top \boldsymbol{w}) + i\boldsymbol{w}^\top \boldsymbol{A}\boldsymbol{\mu}} \\
&= e^{-\frac{1}{2}\boldsymbol{w}^\top (\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top)\boldsymbol{w} + i\boldsymbol{w}^\top \boldsymbol{A}\boldsymbol{\mu}}
\end{aligned}
$$

The uniqueness property shows that $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^\top)$

Properties of Gaussian random vectors

- An equivalent definition of GRV: $\boldsymbol{X}$ is a GRV iff for any real vector $\boldsymbol{a} \neq 0$, the r.v. $Y = \boldsymbol{a}^\top \boldsymbol{X}$ is Gaussian

- Whitening transforms a GRV to a white GRV; conversely, coloring transforms a white GRV to a GRV with prescribed covariance matrix

Properties of Gaussian random vectors

Property 3: Marginals of a GRV are Gaussian, i.e. if $\boldsymbol{X}$ is GRV then for any subset $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$ of indexes, the RV

$$\boldsymbol{Y} = \begin{bmatrix} X_{i_1} \\ X_{i_2} \\ \vdots \\ X_{i_k} \end{bmatrix}$$

is a GRV

Properties of Gaussian random vectors

- To show Property 3, we use Property 2. For example, let $n = 3$ and $\boldsymbol{Y} = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$

  We can express $\boldsymbol{Y}$ as a linear transformation of $\boldsymbol{X}$:

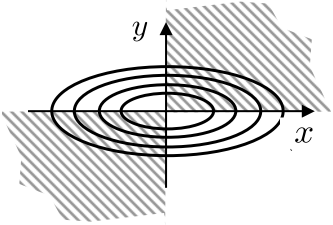
$$\boldsymbol{Y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$$

  Therefore

$$\boldsymbol{Y} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix} \right)$$

Properties of Gaussian random vectors

The converse of Property 3 does not hold in general, i.e. Gaussian marginals do NOT necessarily imply that the r.v.s are jointly Gaussian



- **Example:** consider the following joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} \exp\left\{-\frac{x^2+y^2}{2}\right\}, & \text{if } xy \geq 0 \\ 0, & \text{else} \end{cases}$$

Clearly, it is not 2-D Gaussian but still has Gaussian marginals.

Properties of Gaussian random vectors

Property 4: Conditionals of a GRV are Gaussian, more specifically, if

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ -- \\ \boldsymbol{X}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ -- \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & | & \boldsymbol{\Sigma}_{12} \\ -- & | & -- \\ \boldsymbol{\Sigma}_{21} & | & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) ,$$

where $\boldsymbol{X}_1$ is a k-dim RV and $\boldsymbol{X}_2$ is an $(n-k)$-dim RV, then

$$\boldsymbol{X}_2 \mid \{\boldsymbol{X}_1 = \boldsymbol{x}\} \sim \mathcal{N}\left(\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_2 \, , \, \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \right)$$

- Compare this to the case of $n = 2$ and $k = 1$:

$$X_2 \mid \{X_1 = x\} \sim \mathcal{N}\left(\frac{\sigma_{21}}{\sigma_{11}}(x - \mu_1) + \mu_2 \, , \, \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}} \right)$$

Properties of Gaussian random vectors

- **Example:**

$$\begin{bmatrix} X_1 \\ -- \\ X_2 \\ X_3 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ -- \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & | & 2 & 1 \\ -- & | & -- & -- \\ 2 & | & 5 & 2 \\ 1 & | & 2 & 9 \end{bmatrix} \right)$$

From Property 4, it follows that

$$\mathbb{E}[\boldsymbol{X}_2 \mid X_1 = x] = \begin{bmatrix} 2 \\ 1 \end{bmatrix} (x-1) + \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2x \\ x+1 \end{bmatrix}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{\{\boldsymbol{X}_2 \mid X_1 = x\}} &= \begin{bmatrix} 5 & 2 \\ 2 & 9 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix} \end{aligned}$$

Reference

[1] "*Lecture notes for Statistical Signal Processing*," A. El Gamal.

[2] "*Stochastic Processes: Theory for Applications*," R. Gallager, *Cambridge University Press*, 2013.

[3] "*Introduction to probability (2nd Edition)*," D. Bertsekas, J. Tsitsiklis, *Athena Scientific*, 2008.

[4] "*Probability and Computing (2nd Edition)*," M. Mitzenmacher, E. Upfal, *Cambridge University Press*, 2017.

[5] "*Statistical Inference (2nd Edition)*," G. Casella, R. Berger, *Cengage*, 2002.