भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Probability and Stochastic Processes

Lecture 28: Moment Generating Functions, Chernoff Bound, Characteristic Functions, Joint MGF and Characteristic Functions, Multivariate Gaussians

Karthik P. N.
Assistant Professor, Department of AI
Email: pnkarthik@ai.iith.ac.in

18 November 2025

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Moment Generating Function (MGF)

Definition (Moment Generating Function)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let X be a random variable.
The **moment generating function (MGF)** of X is a function $M_X : \mathbb{R} \to [0, +\infty]$ defined as

$$M_X(t) := \mathbb{E}[e^{tX}], \qquad t \in \mathbb{R}.$$

The **region of convergence (ROC)** of MGF is defined as the set

$$\mathrm{ROC}(M_X) = \Big\{ t \in \mathbb{R} : \; M_X(t) < +\infty \Big\}.$$

Examples

- If $X \sim \text{Exponential}(\mu)$, then

$$M_X(t) = \begin{cases} \frac{\mu}{\mu - t}, & t < \mu, \\ +\infty, & t \geq \mu. \end{cases}$$

- If $X \sim \mathcal{N}(0, 1)$, then

$$M_X(t) = e^{\frac{t^2}{2}}, \qquad t \in \mathbb{R}.$$

- If $f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}$, $x \in \mathbb{R}$, then

$$M_X(t) = \begin{cases} 1, & t = 0, \\ +\infty, & t \neq 0. \end{cases}$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Non-Uniqueness of MGF

- Consider the PDFs f and g given by

$$f(x) = \frac{2}{\pi} \cdot \frac{1}{1+x^2}, \quad x > 0, \qquad\qquad g(x) = \frac{c}{|x|^3}, \quad |x| > 1.$$

- If $X \sim f$ and $Y \sim g$, then

$$M_X(t) = M_Y(t) = \begin{cases} 1, & t = 0, \\ +\infty, & t \neq 0. \end{cases}$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

MGF and Uniqueness of the Underlying Distribution

Theorem (MGF and Underlying Distribution)

1. Suppose there exists $\varepsilon > 0$ such that

$$M_X(t) < +\infty \qquad \forall\, t \in (-\varepsilon, \varepsilon).$$

   Then, $M_X(t)$ determines the CDF of X **uniquely**.

2. If X and Y are random variables such that $M_X(t) = M_Y(t) < +\infty$ for all $t \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$, then X and Y have the same CDF.

Properties of MGF

- $M_X(0) = 1$

- **(Moment generating property)**
  Suppose $M_X(t) < +\infty$ for all $t \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$. Then,

$$\left.\frac{\mathrm{d}^k}{\mathrm{d}t^k} M_X(t)\right|_{t=0} = \mathbb{E}[X^k] \qquad \forall k \in \mathbb{N}.$$

  In particular, for $k = 1$, we have

$$\left.\frac{\mathrm{d}}{\mathrm{d}t} M_X(t)\right|_{t=0} = \mathbb{E}[X].$$

Properties of MGF

- If $Y = aX + b$, then
$$M_Y(t) = e^{bt} M_X(at).$$
As a corollary, it follows that if $Y = \sigma X + \mu$, where $X \sim \mathcal{N}(0, 1)$, then
$$M_Y(t) = e^{\mu t} M_X(\sigma t) = e^{\mu t} e^{t^2\sigma^2/2}.$$

- If $X \perp\!\!\!\perp Y$, then
$$M_{X+Y}(t) = M_X(t) M_Y(t).$$
For example, if $X_1 \sim \text{Exponential}(\mu_1)$, $\quad X_2 \sim \text{Exponential}(\mu_2)$, $\quad$ and $\quad X_1 \perp\!\!\!\perp X_2$, $\quad$ then
$$M_{X_1+X_2}(t) = \begin{cases} \frac{\mu_1}{\mu_1 - t} \cdot \frac{\mu_2}{\mu_2 - t}, & t < \min\{\mu_1, \mu_2\}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Properties of MGF

- Let $Y = \sum_{i=1}^{N} X_i$, where $X_1, X_2, \ldots$ are IID and N is positive integer-valued and independent of $\{X_1, X_2, \ldots\}$. Then,

  $$M_Y(t) = G_N(M_X(t)) = M_N(\log M_X(t)),$$

  where G_N is the PGF of N.

  As a corollary, suppose $X_1, X_2, \ldots \overset{\text{IID}}{\sim} \text{Exponential}(\mu)$ and $N \sim \text{Geometric}(p)$, then

  $$M_Y(t) = \begin{cases} \frac{\mu p}{\mu p - t}, & t < \mu p, \\ +\infty, & t \geq \mu p. \end{cases}$$

Chernoff Bound

Theorem (Chernoff Bound)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let X be a random variable with $M_X(t) < +\infty$ for all $t \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$. Then,

$$\mathbb{P}(\{X > \alpha\}) \leq \frac{M_X(t)}{e^{t\alpha}}, \qquad \forall \alpha \in \mathbb{R}, \ \ t > 0.$$

Optimising over t, we get

$$\mathbb{P}(\{X > \alpha\}) \ \leq \ \inf_{t>0} \frac{M_X(t)}{e^{t\alpha}} \qquad \forall \alpha \in \mathbb{R}.$$

Proof:

- For any $t > 0$, we have

$$\mathbb{P}(\{X > \alpha\}) = \mathbb{P}(\{t\,X > t\,\alpha\}) = \mathbb{P}(\{e^{t\,X} > e^{t\,\alpha}\}).$$

- Apply Markov's inequality to the non-negative RV $e^{t\,X}$, and optimise over $t > 0$ to arrive at the answer

Example: Bernoulli Distribution

- Suppose $X_1, X_2, \ldots \overset{\text{IID}}{\sim} \text{Bernoulli}(p)$
- Then, we have

$$M_{X_1}(t) = \mathbb{E}[e^{t\,X}] = (1-p)\,e^0 + p\,e^t = 1 - p + p\,e^t, \qquad t \in \mathbb{R}.$$

- If $S_n = \sum_{\ell=1}^{n} X_\ell$ for $n \in \mathbb{N}$, then

$$M_{S_n}(t) = \mathbb{E}[e^{t\,S_n}] = \left(1 - p + p\,e^t\right)^n, \qquad t \in \mathbb{R}.$$

Comparison of Markov's, Chebyshev's, and Chernoff's Bound

Upper Bound

Suppose $X_1, X_2, \cdots \overset{\text{IID}}{\sim} \text{Ber}(0.5)$. Let $S_n = \sum_{\ell=1}^{n} X_\ell$ for $n \in \mathbb{N}$. Upper bound the probability

$$\mathbb{P}\left(\left\{S_n \geq \frac{3n}{4}\right\}\right).$$

- **Markov's inequality:**

$$\mathbb{P}\left(\left\{S_n \geq \frac{3n}{4}\right\}\right) \leq \frac{\mathbb{E}[S_n]}{3n/4} = \frac{n/2}{3n/4} = \frac{2}{3}.$$

- **Chebyshev's Inequality:**

$$\mathbb{P}\left(\left\{S_n \geq \frac{3n}{4}\right\}\right) = \mathbb{P}\left(\left\{S_n - \mathbb{E}[S_n] \geq \frac{3n}{4} - \mathbb{E}[S_n]\right\}\right) = \mathbb{P}\left(\left\{S_n - \frac{n}{2} \geq \frac{n}{4}\right\}\right) \leq \mathbb{P}\left(\left\{\left|S_n - \frac{n}{2}\right| \geq \frac{n}{4}\right\}\right)$$
$$\leq \frac{\text{Var}(S_n)}{(n/4)^2} = \frac{n/4}{n^2/16} = \frac{4}{n}.$$

Comparison of Markov's, Chebyshev's, and Chernoff's Bound

Upper Bound

Suppose $X_1, X_2, \cdots \overset{\text{IID}}{\sim} \text{Ber}(0.5)$. Let $S_n = \sum_{\ell=1}^{n} X_\ell$ for $n \in \mathbb{N}$. Upper bound the probability

$$\mathbb{P}\left(\left\{S_n \geq \frac{3n}{4}\right\}\right).$$

- **Chernoff's Bound:**

$$\mathbb{P}\left(\left\{S_n \geq \frac{3n}{4}\right\}\right) \leq \inf_{t>0} \frac{M_{S_n}(t)}{e^{\frac{3nt}{4}}} = \inf_{t>0} \frac{\left(0.5 + 0.5e^t\right)^n}{e^{\frac{3nt}{4}}} = \inf_{t>0} \exp\left(n \log(0.5 + 0.5\, e^t) - \frac{3nt}{4}\right)$$

$$= \exp\left(\inf_{t>0} \left\{n \log(0.5 + 0.5\, e^t) - \frac{3nt}{4}\right\}\right) = \exp\left(-\frac{nC}{4}\right), \qquad C = \log\frac{27}{16} > 0.$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Characteristic Functions

Characteristic Function

Definition (Characteristic Function

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let X be a random variable.
The **characteristic function** of X is a function $C_X : \mathbb{R} \to \mathbb{C}$ defined as

$$C_X(s) := \mathbb{E}[e^{jsX}] = \mathbb{E}[\cos sX] + j\,\mathbb{E}[\sin sX], \qquad s \in \mathbb{R}.$$

Remark:

$$\left|C_X(s)\right| \leq 1 \qquad \forall s \in \mathbb{R}.$$

Examples

- If $X \sim \text{Exponential}(\mu)$, then

$$C_X(s) = \frac{\mu}{\mu - js}, \qquad s \in \mathbb{R}.$$

- **(Fourier Duality):**

$$\begin{aligned} f(x) &\quad\longleftrightarrow\quad C(s) \\ C(x) &\quad\longleftrightarrow\quad 2\,\pi\, f(-s) \end{aligned}$$

- **(Characteristic Function of Cauchy distribution via Fourier duality):**

$$\begin{aligned} e^{-\gamma\,|x|} &\quad\longleftrightarrow\quad \frac{2\,\gamma}{\gamma^2 + s^2} \\ \frac{2\,\gamma}{\gamma^2 + x^2} &\quad\longleftrightarrow\quad 2\,\pi\, e^{-\gamma\,|s|} \end{aligned}$$

If $f_X(x) = \frac{1}{\pi}\,\frac{1}{1+x^2}, \quad x \in \mathbb{R},$ then using the above property, we have

$$C_X(s) = e^{-|s|}, \qquad s \in \mathbb{R}.$$

Properties of Characteristic Functions

- If $Y = aX + b$, then
$$C_Y(s) = e^{jbs}\, C_X(as), \qquad s \in \mathbb{R}.$$
- If $X \perp\!\!\!\perp Y$, then
$$C_{X+Y}(s) = C_X(s)\, C_Y(s) \qquad \forall s \in \mathbb{R}.$$
  As a corollary, it follows that if X, Y are IID Cauchy, then $X + Y$ is also Cauchy (albeit with a different parameter).
- If $M_X(t) < +\infty$ for all $t \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$, then
$$C_X(s) = M_X(js) \qquad \forall s \in \mathbb{R}.$$

Properties of Characteristic Functions

- If $C_X(s) = C_Y(s)$ for all $s \in \mathbb{R}$, then X and Y have the same CDF, i.e.,

$$F_X(x) = F_Y(x) \qquad \forall x \in \mathbb{R}.$$

- **(Recovering moments from characteristic function)**

  For $k \in \mathbb{N}$, if $\left| \frac{\mathrm{d}^k}{\mathrm{d}s^k} C_X(s) \Big|_{s=0} \right| < +\infty$, then

$$\mathbb{E}[X^k] = (-j)^k \frac{\mathrm{d}^k}{\mathrm{d}s^k} C_X(s) \Big|_{s=0}.$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Joint MGF and Joint Characteristic Functions

Joint MGF and Joint Characteristic Function

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $X_1, \ldots, X_n$ be random variables.

1. The **joint MGF** of $X_1, \ldots, X_n$ is a function $M_{X_1,\ldots,X_n} : \mathbb{R}^n \to [0, +\infty]$, defined as

$$M_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = \mathbb{E}[e^{t_1 X_1 + \cdots + t_n X_n}] = \mathbb{E}[e^{\mathbf{t}^\top \mathbf{X}}],$$

where $\mathbf{t} = [t_1 \ \cdots \ t_n]^\top$ and $\mathbf{X} = [X_1 \ \cdots \ X_n]^\top$.

2. The **joint characteristic function** of $X_1, \ldots, X_n$ is a function $C_{X_1,\ldots X_n} : \mathbb{R}^n \to \mathbb{C}$, defined as

$$C_{X_1,\ldots,X_n}(s_1, \ldots, s_n) = \mathbb{E}[j(s_1 X_1 + \cdots + j_n X_n)] = \mathbb{E}[e^{j\mathbf{s}^\top \mathbf{X}}],$$

where $\mathbf{s} = [s_1 \ \cdots \ s_n]^\top$.

Independence and Joint MGF/CF

Theorem (Independence and Joint MGF/CF)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $X_1, \ldots, X_n$ be random variables. Suppose that $M_{X_1,\ldots,X_n}(t_1, \ldots, t_n) < +\infty$ for all $(t_1, \ldots, t_n) \in B(\mathbf{0}, \varepsilon)$ for some $\varepsilon > 0$, where $B(\mathbf{0}, \varepsilon)$ denotes a ball in $\mathbb{R}^n$ centered at the origin $\mathbf{0}$ and having radius ε. Then,

$$X_1, \ldots, X_n \text{ mutually independent} \iff M_{X_1,\ldots,X_n}(t_1, \ldots, t_n) = \prod_{i=1}^{n} M_{X_i}(t_i) \quad \forall\, (t_1, \ldots, t_n) \in \mathbb{R}^n,$$

$$X_1, \ldots, X_n \text{ mutually independent} \iff C_{X_1,\ldots,X_n}(s_1, \ldots, s_n) = \prod_{i=1}^{n} C_{X_i}(s_i) \quad \forall (s_1, \ldots, s_n) \in \mathbb{R}^n.$$

Caution

Caution

To check that two random variables X and Y are independent, it **DOES NOT suffice** to check that

$$C_{X,Y}(s,s) = C_X(s)\, C_Y(s) \qquad \forall s \in \mathbb{R}.$$

Example:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{4}(1 + xy(x^2 - y^2)), & |x| < 1,\ |y| < 1, \\ 0, & \text{otherwise.} \end{cases}$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Multivariate Gaussians

Standard Bivariate Gaussian Random Variables

Definition (Standard Bivariate Gaussian Random Variables)

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let X and Y be random variables.
X and Y are said to be **standard bivariate Gaussian** if:

1. X and Y are jointly continuous, and
2. The joint PDF of X and Y is given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right), \qquad x, y \in \mathbb{R},$$

for some $\rho \in (-1, 1)$.

Properties of Standard Bivariate RVs

Proposition

Let X and Y be standard bivariate Gaussian RVs with parameter $\rho \in (-1, 1)$. Then, the following hold.

- $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$.
- Conditioned on $\{Y = y\}$, X is distributed according to $\mathcal{N}(\rho y, 1 - \rho^2)$. Consequently, $\mathbb{E}[X|Y] = \rho Y$.
- $\rho_{X,Y} = \rho$.
- If $\rho = 0$, then $X \perp\!\!\!\perp Y$.
  That is, **uncorrelatedness implies independence**.

Some notations going forward:

$$\mathbf{X} = \begin{bmatrix} X \\ Y \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

General Bivariate Gaussian RVs

Definition (Bivariate Gaussian RVs)

We say X and Y are **bivariate Gaussian** RVs or **jointly Gaussian** if

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi\sqrt{\det(K)}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \qquad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2,$$

for some $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \in \mathbb{R}^2$ and a **positive definite** matrix K.

Notation: $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$

Exercises:

- If $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$, then

$$\mathbb{E}[X] = \mu_1, \qquad \mathbb{E}[X] = \mu_2, \qquad K = \mathbb{E}\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\top\right] = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(X,Y) & \mathrm{Var}(Y) \end{bmatrix}.$$

Caution!

Caution!

If X and Y are individually Gaussian, then they **need not be** jointly Gaussian.

- Let Y_1, Y_2 be IID with PDF
$$f(y) = \sqrt{\frac{2}{\pi}} e^{-y^2/2}, \qquad y \geq 0.$$
- Let $W \perp\!\!\!\perp Y_1, Y_2$, with $\mathbb{P}(\{W = 1\}) = \mathbb{P}(\{W = -1\}) = \frac{1}{2}$. Let
$$X = W Y_1, \qquad Y = W Y_2.$$
- **Exercise:** $\quad X \sim \mathcal{N}(0, 1), \quad Y \sim \mathcal{N}(0, 1)$
- However, X and Y are **NOT** jointly Gaussian:
$$X \geq 0 \iff Y \geq 0 \qquad\qquad X \leq 0 \iff Y \leq 0.$$

The joint PDF of X and Y has probability concentrated only in first and third quadrants!

Multivariate Gaussian RVs: Definition 1

Definition 1 (Multivariate Gaussian RVs)

Fix $n \in \mathbb{N}$. Random variables $X_1, \ldots, X_n$ are said to be **multivariate Gaussian** if:

- $X_1, \ldots, X_n$ are jointly continuous, and
- The joint PDF of $\mathbf{X} = (X_1, \ldots, X_n)$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(K)}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1} (\mathbf{x}-\boldsymbol{\mu})}{2}\right), \qquad \mathbf{x} \in \mathbb{R}^n,$$

for some $\boldsymbol{\mu} \in \mathbb{R}^n$ and a **positive definite** matrix K of size $n \times n$.

Notation: $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu},\ K)$

Multivariate Gaussian RVs: Definition 2

Definition 2 (Multivariate Gaussian RVs)

Fix $n \in \mathbb{N}$. Random variables $X_1, \ldots, X_n$ are said to be **multivariate Gaussian** if

$$\mathbf{X} = D\mathbf{W} + \boldsymbol{\mu}$$

for some matrix $D \in \mathbb{R}^{n \times m}$ and some real vector $\boldsymbol{\mu} \in \mathbb{R}^n$, where $\mathbf{W} = (W_1, \ldots, W_m)$ with $W_1, \ldots, W_m \overset{\text{IID}}{\sim} \mathcal{N}(0, 1)$.

Multivariate Gaussian RVs: Definition 3

Definition 3 (Multivariate Gaussian RVs)

Fix $n \in \mathbb{N}$. Random variables $X_1, \ldots, X_n$ are said to be **multivariate Gaussian** if:
for every non-zero $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, the random variable

$$\mathbf{a}^\top \mathbf{X} = a_1 X_1 + \cdots + a_n X_n$$

is (one-dimensional) Gaussian distributed.

Equivalence of Definitions $1, 2, 3$

Definition 1 $\implies$ **Definition 2:**

- Suppose $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$ for some $\boldsymbol{\mu} \in \mathbb{R}^n$ and positive definite matrix K (i.e., $\det(K) > 0$)
- **Spectral decomposition** of K: Because K is symmetric ($K^\top = K$), we can write it as

$$K = \sum_{i=1}^{n} \lambda_i \, \mathbf{z}_i \, \mathbf{z}_i^\top = U \Lambda U^\top,$$

  where $\lambda_1, \ldots, \lambda_n > 0$ are eigenvalues, and $\mathbf{z}_1, \ldots, \mathbf{z}_n$ are orthonormal eigenvectors, U is a matrix with columns as eigenvectors, Λ is a diagonal matrix with eigenvalues on the diagonal

- Let $D = U\Lambda^{1/2}U^\top$. Then, we have:
  - $D^\top = D$
  - $DD^\top = D^2 = D^\top D = K$
  - $\det(D) = \prod_{i=1}^n \sqrt{\lambda_i} > 0$
  - D^{-1} exists

Equivalence of Definitions $1, 2, 3$

Definition 1 $\implies$ Definition 2:

- Let $\mathbf{W} = D^{-1}(\mathbf{X} - \boldsymbol{\mu})$
- Clearly, $\mathbb{E}[\mathbf{W}] = \mathbf{0}$, and

$$\mathrm{Cov}(\mathbf{W}, \mathbf{W}) = \mathbb{E}[\mathbf{W}\mathbf{W}^\top] = \mathbb{E}[D^{-1}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\top D^{-1}] = D^{-1} K D^{-1} = I.$$

- **Exercise:** Using the Jacobian transformations formula,

$$f_{\mathbf{W}}(\mathbf{w}) = \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{\mathbf{w}^\top \mathbf{w}}{2}\right), \qquad \mathbf{w} \in \mathbb{R}^n,$$

  thus proving that $W_1, \ldots, W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$
- Thus, we have

$$\mathbf{X} = D\mathbf{W} + \boldsymbol{\mu}, \qquad D = \sqrt{K}.$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Equivalence of Definitions $1, 2, 3$

Definition 2 $\implies$ **Definition 3:**

- Suppose there exists $D \in \mathbb{R}^{n \times m}$ and $\boldsymbol{\mu} \in \mathbb{R}^n$ such that
$$\mathbf{X} = D\mathbf{W} + \boldsymbol{\mu}$$
where $\mathbf{W} = (W_1, \ldots, W_m)$, with $W_1, \ldots, W_m \overset{\text{IID}}{\sim} \mathcal{N}(0, 1)$
- Given a non-zero $\mathbf{a} \in \mathbb{R}^n$, we have
$$a^\top \mathbf{X} = a^\top D\mathbf{W} + \mathbf{a}^\top \boldsymbol{\mu}$$
- **Notations:**
$$\mathbf{b} = \mathbf{a}^\top D, \qquad \mathbf{a}^\top D\,\mathbf{W} = \mathbf{b}^\top\,\mathbf{W} = \sum_{\ell=1}^{m} b_\ell\, W_\ell.$$
- The MGF of $Y_\mathbf{a} = \mathbf{a}^\top \mathbf{X}$ is given by
$$M_{Y_\mathbf{a}}(t) = \mathbb{E}[e^{t\,Y_\mathbf{a}}] = e^{t\mathbf{a}^\top \boldsymbol{\mu}} \cdot \mathbb{E}[e^{t\mathbf{a}^\top D\mathbf{W}}] = e^{t\mathbf{a}^\top \boldsymbol{\mu}} \cdot \prod_{i=1}^{m} \mathbb{E}[e^{t\,b_i\,W_i}] = e^{t\mathbf{a}^\top \boldsymbol{\mu}} \cdot \prod_{i=1}^{m} e^{t^2\,b_i^2/2}.$$
From the above MGF expression, we conclude that $Y_\mathbf{a} \sim \mathcal{N}(\alpha_\mathbf{a}, \sigma_\mathbf{a}^2)$, with $\alpha_\mathbf{a} = \mathbf{a}^\top \boldsymbol{\mu}$ and $\sigma_\mathbf{a}^2 = \mathbf{a}^\top DD^\top \mathbf{a}$

Joint MGF

- So far, we have seen
$$\text{Definition 1} \implies \text{Definition 2} \implies \text{Definition 3}$$
- Therefore, we have
$$\text{Definition 1} \implies \text{Definition 3}$$
- We can use the above implication to derive the joint MGF of $(X_1, \ldots, X_n) \sim \mathcal{N}(\boldsymbol{\mu}, K)$

Joint MGF

- Suppose $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu},\ K)$
- For any non-zero $\mathbf{s} \in \mathbb{R}^n$,

$$M_{\mathbf{X}}(\mathbf{s}) = \mathbb{E}[e^{\mathbf{s}^\top \mathbf{X}}] = M_{\mathbf{s}^\top \mathbf{X}}(1)$$

- From Definition 3, we know that $Y_{\mathbf{s}} = \mathbf{s}^\top \mathbf{X}$ is Gaussian with mean and variance

$$\mathbb{E}[Y_{\mathbf{s}}] = \mathbb{E}[\mathbf{s}^\top \mathbf{X}] = \mathbf{s}^\top \boldsymbol{\mu}, \qquad \mathrm{Var}(Y_{\mathbf{s}}) = \mathbb{E}[(\mathbf{s}^\top (\mathbf{X} - \boldsymbol{\mu}))^2] = \mathbf{s}^\top K \mathbf{s}.$$

- Therefore, we have

$$M_{\mathbf{X}}(\mathbf{s}) = M_{Y_{\mathbf{s}}}(1) = e^{\mathbf{s}^\top \boldsymbol{\mu}} \cdot e^{\mathbf{s}^\top K \mathbf{s}/2}$$

Equivalence of Definitions $1, 2, 3$

Definition 3 $\implies$ **Definition 1**

- Suppose that $Y_{\mathbf{a}} = \mathbf{a}^\top \mathbf{X}$ is Gaussian for every non-zero $\mathbf{a} \in \mathbb{R}^n$
- Assume $\mathbb{E}[\mathbf{X}] = \mu = \mathbf{0}$ (w.l.o.g.)
- Let $K = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$
- Two cases: K invertible or not invertible

Equivalence of Definitions $1, 2, 3$

Definition 3 $\implies$ **Definition 1:** **(assuming $K = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ is invertible)**

- Let $D = \sqrt{K}$
- K invertible $\implies$ D invertible
- Define $\mathbf{W} = D^{-1}\mathbf{X}$
- $\mathbb{E}[\mathbf{W}] = \mathbf{0}, \qquad \mathbb{E}[\mathbf{W}\mathbf{W}^\top] = D^{-1}KD^{-1} = I$
- For any non-zero $\mathbf{s} \in \mathbb{R}^n$,
$$M_{\mathbf{W}}(\mathbf{s}) = \mathbb{E}[e^{\mathbf{s}^\top \mathbf{W}}] = M_{\mathbf{s}^\top \mathbf{W}}(1).$$
- From Definition 3, we know that $Y = \mathbf{s}^\top \mathbf{W}$ is Gaussian with mean and variance
$$\mathbb{E}[Y] = \mathbb{E}[\mathbf{s}^\top \mathbf{W}] = 0, \qquad \mathrm{Var}(Y) = \mathbb{E}[(\mathbf{s}^\top \mathbf{W})^2] = \mathbf{s}^\top \mathbf{s}.$$
- Therefore, $M_{\mathbf{W}}(\mathbf{s}) = M_Y(1) = e^{\mathbf{s}^\top \mathbf{s}/2} \implies W_1, \ldots, W_n \overset{\text{IID}}{\sim} \mathcal{N}(0, 1)$

Equivalence of Definitions $1, 2, 3$

Definition 3 $\implies$ **Definition 1:** **(assuming $K = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ is invertible)**

- Thus, $\mathbf{X} = D\mathbf{W}$, $\quad D = \sqrt{K}$
- **Exercise:** Using Jacobian transformations formula with $\mathbf{X} = g(\mathbf{W})$, $\quad g(\mathbf{w}) = D\mathbf{w}$,

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{f_{\mathbf{W}}(g^{-1}(\mathbf{x}))}{\left|\det(J_g(g^{-1}(\mathbf{x})))\right|} = \frac{f_{\mathbf{W}}(D^{-1}\mathbf{x})}{\det(D)}$$

$$= \frac{1}{\sqrt{(2\pi)^n \det(K)}} \exp\left(-\frac{\mathbf{x}^\top K^{-1}\mathbf{x}}{2}\right)$$

Equivalence of Definitions $1, 2, 3$

Definition 3 $\implies$ Definition 1: **(assuming $K = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ is NOT invertible)**

- Suppose $\det(K) = 0$
- There exists $\mathbf{a} \neq \mathbf{0}$ such that $K\mathbf{a} = \mathbf{0}$, and therefore

$$\mathbf{a}^\top K\mathbf{a} = \mathbf{0}.$$

- But $\mathbf{a}^\top K\mathbf{a} = \mathbb{E}[(\mathbf{a}^\top \mathbf{X})^2]$, so we have $\mathbb{E}[(\mathbf{a}^\top \mathbf{X})^2] = 0$, which implies

$$\mathbb{P}(\{a^\top \mathbf{X} = 0\}) = 1.$$

- With probability 1, one of the components of $\mathbf{X}$ is linearly dependent on the others

Equivalence of Definitions $1, 2, 3$

Definition 3 $\implies$ Definition 1: **(assuming $K = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ is NOT invertible)**

- W.l.o.g., let X_n be a linear combination of $(X_1, \ldots, X_{n-1})$
- Let K_1 be the covariance matrix of $(X_1, \ldots, X_{n-1})$
- If $\det(K_1) = 0$, repeat the process till we arrive at a non-singular covariance matrix
- After suitable reordering of coordinates, $\mathbf{X}$ may be expressed as

$$\mathbf{X} = (\mathbf{Y}, \mathbf{Z}),$$

  in which $\mathbf{Y}$ has non-singular covariance matrix $K_{\mathbf{Y}}$, and $\mathbf{Z} = A\mathbf{Y}$ for some matrix A
- Let $K_{\mathbf{Y}}$ be of size $k \times k$
- Let $D = \sqrt{K_{\mathbf{Y}}}$; $\qquad D$ is also of size $k \times k$
- Because $K_{\mathbf{Y}}$ is invertible, we have

$$\mathbf{Y} = D\mathbf{W}, \qquad \mathbf{W} \sim \mathcal{N}(\mathbf{0}, I_{k \times k})$$

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Equivalence of Definitions $1, 2, 3$

Definition 3 $\Longrightarrow$ Definition 1: **(assuming $K = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$ is NOT invertible)**

- **Exercise:** Using Jacobian transformations formula, we have

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, D^2) = \mathcal{N}(0, K_{\mathbf{Y}}).$$

- Noting $\mathbf{Y} = D\mathbf{W}$, $\mathbf{Z} = AY = AD\mathbf{W}$, we can write $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{Z} \end{bmatrix} = \begin{bmatrix} D & \mathbf{O}_{k\times k} \\ AD & \mathbf{O}_{n-k\times n-k} \end{bmatrix} \begin{bmatrix} \mathbf{W} \\ \overline{\mathbf{W}} \end{bmatrix},$$

where $\overline{\mathbf{W}}$ consists of $(n-k)$ i.i.d. $\mathcal{N}(0, 1)$ RVs

Final Remarks

If the components of $\mathbf{X} = (X_1, \ldots, X_n)$ are linearly dependent, then $\mathbf{X}$ does not have a joint PDF and therefore $X_1, \ldots, X_n$ are **NOT** jointly Gaussian.
We can find a subset of components, say $\mathbf{Y}$, which admits a joint PDF and is jointly Gaussian.