भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Probability and Stochastic Processes

Gaussian Random Vectors (or Multivariate Gaussian RVs), Equivalent Definitions for Multivariate Gaussian RVs, Convergence of Sequences of Random Variables (Intro)

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

21/23 November 2024

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Gaussian Random Vectors
OR
Jointly Gaussian Random Variables
OR
Multivariate Gaussian Random Variables

Standard Bivariate Gaussian Random Variables

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let X and Y be random variables w.r.t. $\mathscr{F}$.

Definition (Standard Bivariate Gaussian Random Variables)

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 random variables with parameter $\rho \in (-1, 1)$. Then, the following hold.

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

General Bivariate Gaussian RVs

Definition (Bivariate Gaussian RVs)

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

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

for some $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and a positive definite matrix K.

Remark:

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

Caution

Caution!

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

Example: Let Y_1, Y_2 be i.i.d. 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

$$X_1 = WY_1, \qquad X_2 = WY_2.$$

Clearly, $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(0, 1)$. Furthermore,

$$X_1 \geq 0 \iff X_2 \geq 0, \qquad X_1 \leq 0 \iff X_2 \leq 0.$$

So, the joint density of X_1 and X_2 has mass only in first and third quadrants

Multivariate Gaussian RVs

Let $X_1, \ldots, X_n$ be random variables. Let $\mathbf{X} = [X_1 \;\; X_2 \;\; \cdots \;\; X_n]^\top$.

Definition 1 (Multivariate Gaussian RVs)

The random variables $X_1, \ldots, X_n$ are said to be multivariate Gaussian if

- 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.

Notation:

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

Multivariate Gaussian RVs

Let $X_1, \ldots, X_n$ be random variables. Let $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}^\top$.

Definition 2 (Multivariate Gaussian RVs)

The random variables $X_1, \ldots, X_n$ are said to be multivariate Gaussian if

- $\mathbf{X} = (X_1, \ldots, X_n)$ can be expressed as

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

  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 $W_1, \ldots, W_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Multivariate Gaussian RVs

Let $X_1, \ldots, X_n$ be random variables. Let $\mathbf{X} = [X_1 \;\; X_2 \;\; \cdots \;\; X_n]^\top$.

Definition 3 (Multivariate Gaussian RVs)

The 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 Gaussian distributed.

Equivalence of Definitions $1, 2, 3$

Definition 1 $\implies$ Definition 2

- Suppose $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, K)$, where $\det(K) > 0$
- Spectral decomposition of K:

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

  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:
  - $D^\top = D$
  - $DD^\top = D^2 = D^\top D = K$
  - $\det(D) = \prod_{i=1}^{n} \sqrt{\lambda_i} > 0$

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

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

- 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

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

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 $W_1, \ldots, W_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$
- Given a non-zero $\mathbf{a} \in \mathbb{R}^n$, we have

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

- The MGF of $Y = \mathbf{a}^\top \mathbf{X}$ is given by

$$M_Y(t) = \mathbb{E}[e^{tY}] = e^{t\mathbf{a}^\top \boldsymbol{\mu}} \cdot \mathbb{E}[e^{t\mathbf{a}^\top D\mathbf{W}}] = e^{t\mathbf{a}^\top \boldsymbol{\mu}} \cdot \prod_{i=1}^{m} \mathbb{E}[e^{t\, b_i\, W_i}] = e^{t\mathbf{a}^\top \boldsymbol{\mu}} \cdot \prod_{i=1}^{m} e^{t^2\, b_i^2/2},$$

  where $b_i = (\mathbf{a}^\top D)_i$. From the above MGF expression, we conclude that $Y \sim \mathcal{N}(\alpha, \sigma^2)$, with $\alpha = \mathbf{a}^\top \boldsymbol{\mu}$ and $\sigma^2 = \mathbf{a}^\top D D^\top \mathbf{a}$

Joint MGF

- 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 this 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 $Y = \mathbf{s}^\top \mathbf{X}$ is Gaussian with mean and variance

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

- Therefore, we have

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

Equivalence of Definitions $1, 2, 3$

Definition 3 $\implies$ Definition 1

- Suppose that $Y = \mathbf{a}^\top \mathbf{X}$ is Gaussian for every non-zero $\mathbf{a} \in \mathbb{R}^n$
- Assume $\mathbb{E}[\mathbf{X}] = \mathbf{0}$ (w.l.o.g.)
- Let $K = \mathbb{E}[\mathbf{X}\mathbf{X}^\top]$

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, $M_{\mathbf{W}}(\mathbf{s}) = M_Y(1) = e^{\mathbf{s}^\top \mathbf{s}/2} \implies W_1, \ldots, W_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$

Equivalence of Definitions $1, 2, 3$

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

- Thus, $\mathbf{X} = D\mathbf{W}, \qquad D = \sqrt{K}$
- 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 not invertible)

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

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

- 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

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

- One of the components of $\mathbf{X}$ is linearly dependent on the others

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Equivalence of Definitions $1, 2, 3$

Definition 3 $\quad\Longrightarrow\quad$ Definition 1
(Assuming K 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 $D = \sqrt{K_{\mathbf{Y}}}$; D is also of size $k \times k$
- Because $K_{\mathbf{Y}}$ is invertible, we have

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

Equivalence of Definitions $1, 2, 3$

Definition 3 $\implies$ Definition 1
(Assuming K not invertible)

- Using Jacobian transformations formula, we can show that

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

- 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

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Convergence of Sequences of Random Variables

Objective

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables defined w.r.t. $\mathscr{F}$.
Let X be another random variable defined w.r.t. $\mathscr{F}$. We allow X to take $\pm\infty$

Objective

To define the following forms of convergence.

1. Pointwise convergence; notation: $X_n \xrightarrow{\text{pointwise}} X$.
2. Almost-sure convergence; notation: $X_n \xrightarrow{\text{a.s.}} X$.
3. Mean-squared convergence; notation: $X_n \xrightarrow{\text{m.s.}} X$.
4. Convergence in probability; notation: $X_n \xrightarrow{\text{p}} X$.
5. Convergence in distribution; notation: $X_n \xrightarrow{\text{d}} X$.

Pointwise Convergence

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.
Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables defined w.r.t. $\mathscr{F}$.
Let X be another random variable defined w.r.t. $\mathscr{F}$. We allow X to take $\pm\infty$

Definition (Pointwise Convergence)

We say that the sequence $\{X_n\}_{n=1}^{\infty}$ converges to X pointwise if

$$\forall \omega \in \Omega, \qquad \lim_{n\to\infty} X_n(\omega) = X(\omega).$$

Notation:

$$X_n \xrightarrow{\text{pointwise}} X.$$

Example

$\Omega = [0,1], \mathscr{F} = \mathscr{B}([0,1]), \mathbb{P} = \lambda.$
For each $n \in \mathbb{N}$, let

$$X_n(\omega) = \begin{cases} 1, & \omega \in \left[0, \frac{1}{n}\right), \\ 0, & \text{otherwise.} \end{cases}$$

Identify the limit random variable X to which the above sequence converges pointwise.

Convergence in Distribution

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables defined w.r.t. $\mathscr{F}$.

Let X be another random variable defined w.r.t. $\mathscr{F}$. We allow X to take $\pm\infty$

Definition (Convergence in Distribution)

We say that the sequence $\{X_n\}_{n=1}^{\infty}$ converges to X in distribution if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \qquad \forall x \in C_{F_X},$$

where C_{F_X} denotes the points of continuity of F_X.

Notation:

$$X_n \xrightarrow{\mathrm{d}} X.$$

Example

Let $X_n = U$ for all $n \in \mathbb{N}$, with $U \sim \text{Unif}([0, 1])$.
Let $X = 1 - U$.
Show that does not converge to X pointwise, but $X_n \xrightarrow{\text{d}} X$.