香港中文大學
The Chinese University of Hong Kong

Chapter 10
Differential Entropy

The Chinese University of Hong Kong

In this chapter

In this chapter

- Real-valued random vectors

In this chapter

- Real-valued random vectors
- Symmetric, positive definite, and covariance matrices

In this chapter

- Real-valued random vectors
- Symmetric, positive definite, and covariance matrices
- Differential entropy and mutual information

In this chapter

- Real-valued random vectors
- Symmetric, positive definite, and covariance matrices
- Differential entropy and mutual information
- AEP and informational divergence

In this chapter

- Real-valued random vectors
- Symmetric, positive definite, and covariance matrices
- Differential entropy and mutual information
- AEP and informational divergence
- Gaussian distribution

Real Random Variables

Real Random Variables

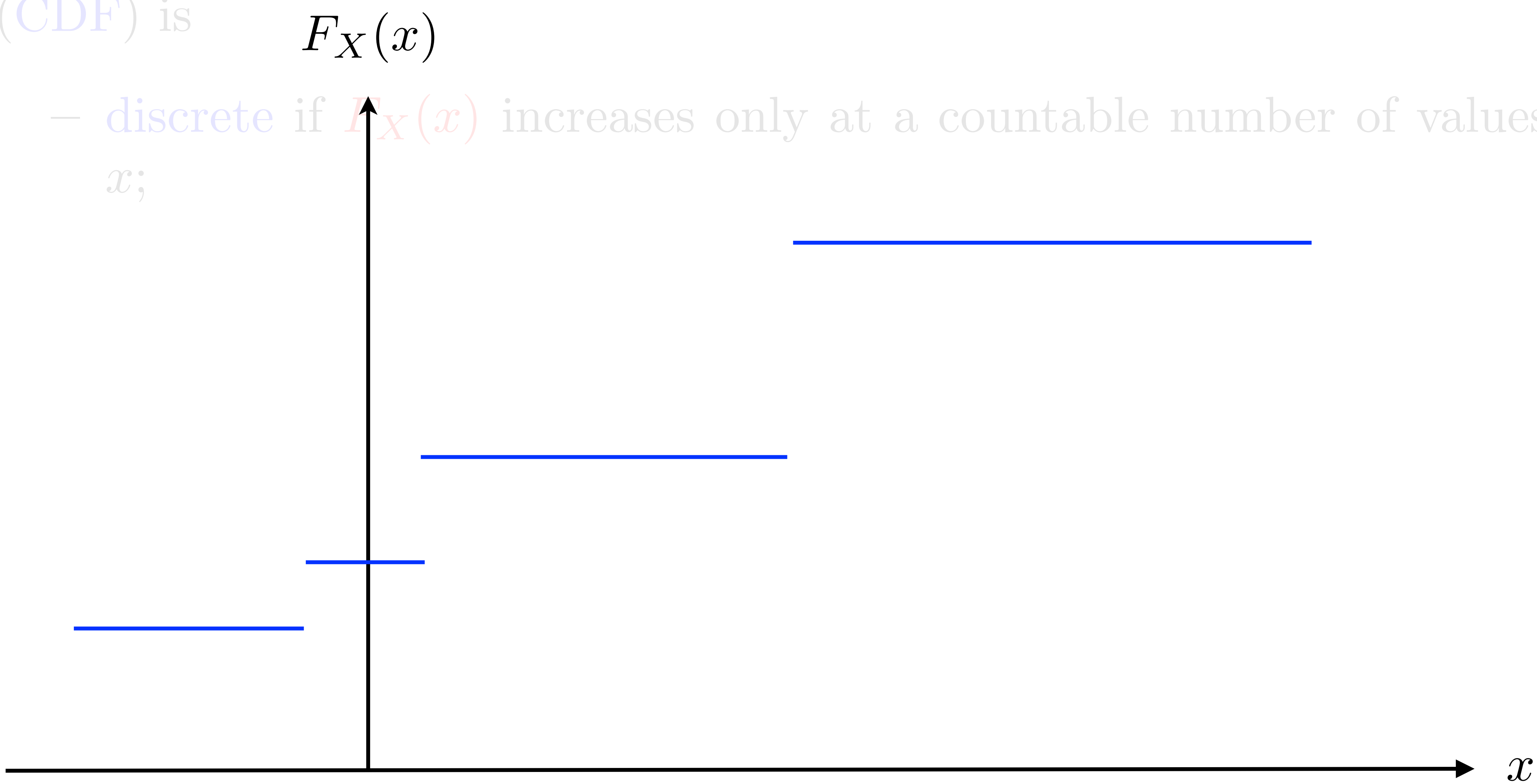
- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is

Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;

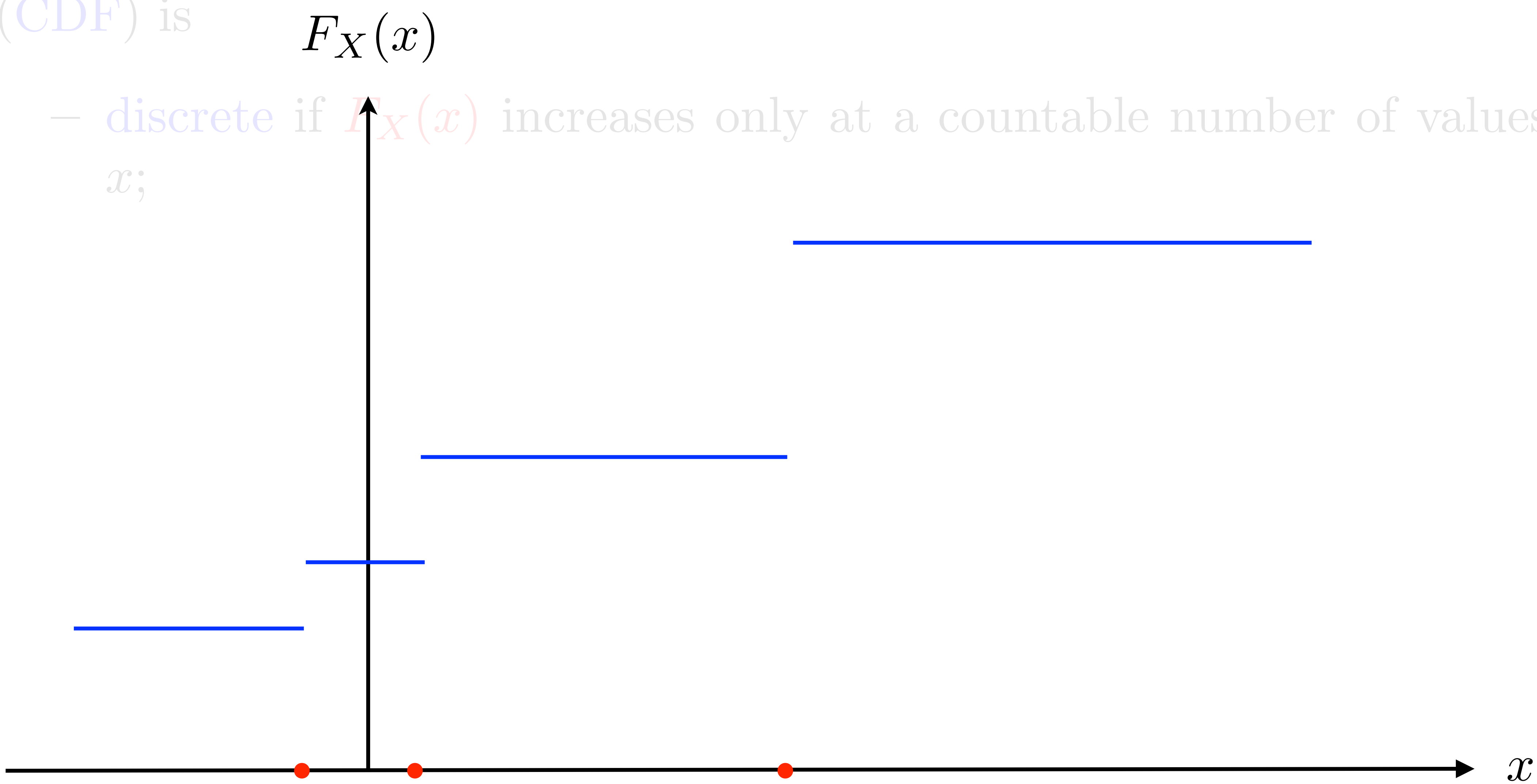
Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;



Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;



Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;

Real Random Variables

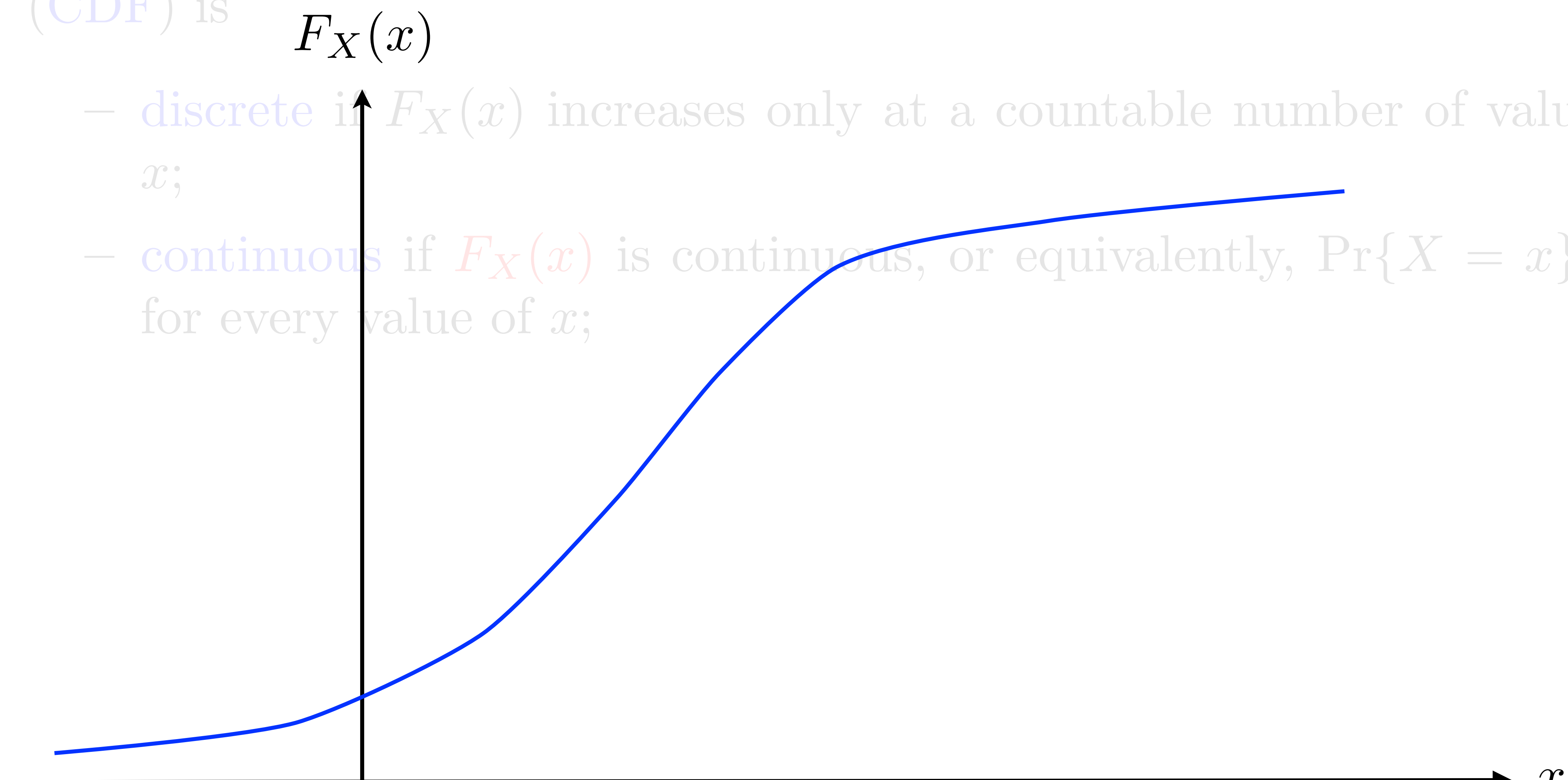
- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;

Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;

Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;



Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;

Real Random Variables

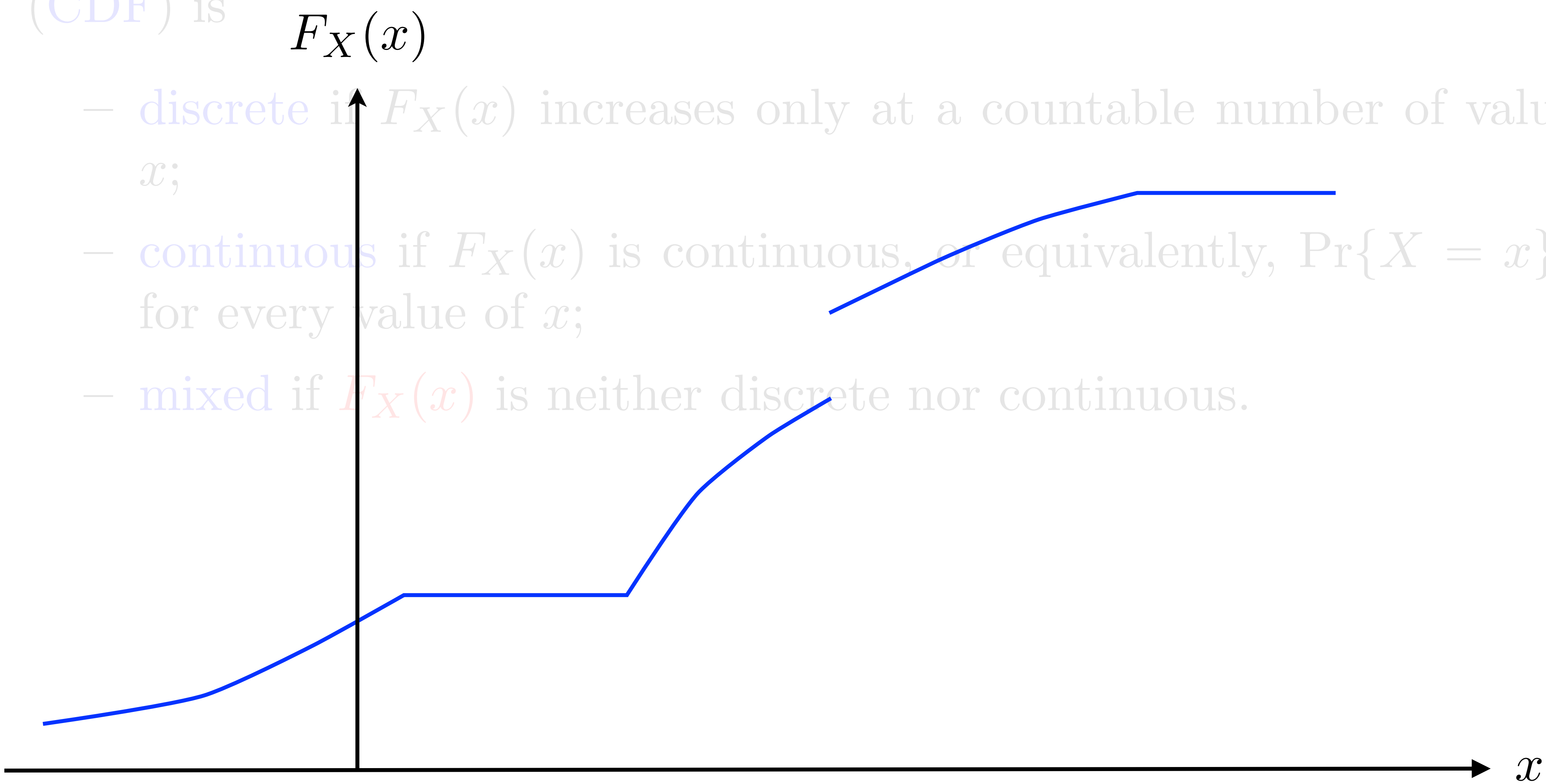
- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;

Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.

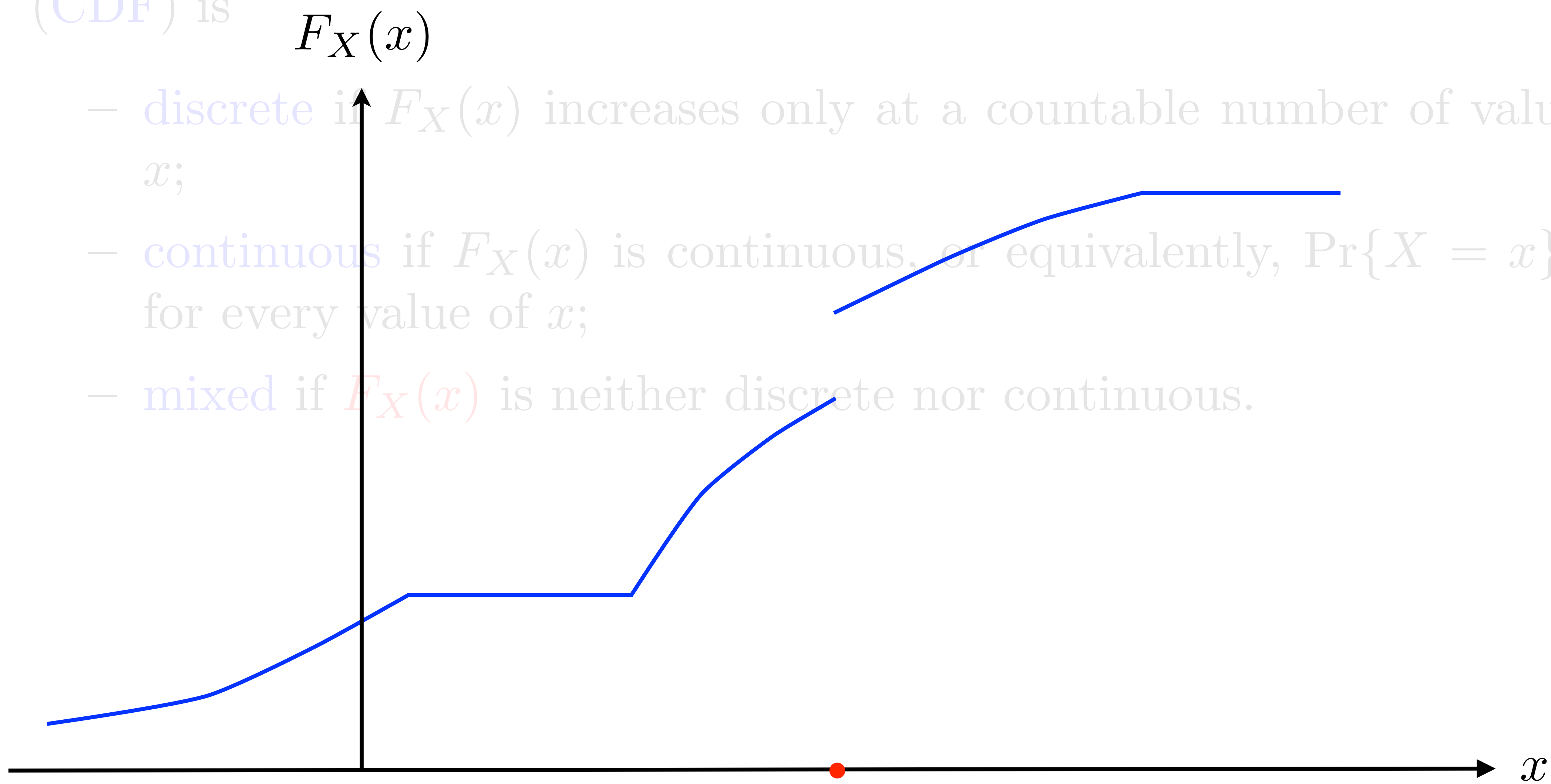
Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.



Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.



Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.

Real Random Variables

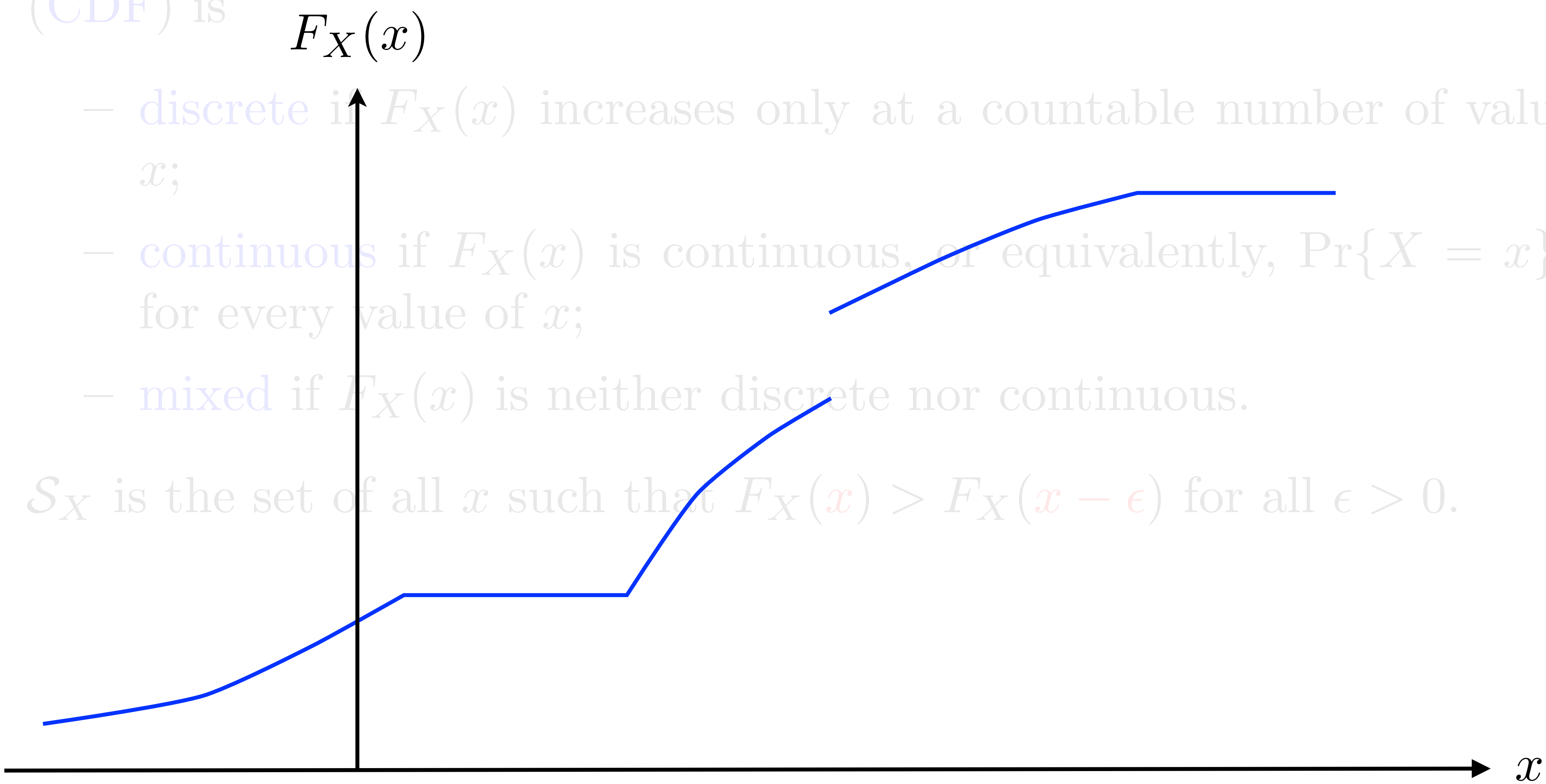
- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.

Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.
- $\mathcal{S}_X$ is the set of all x such that $F_X(x) > F_X(x - \epsilon)$ for all $\epsilon > 0$.

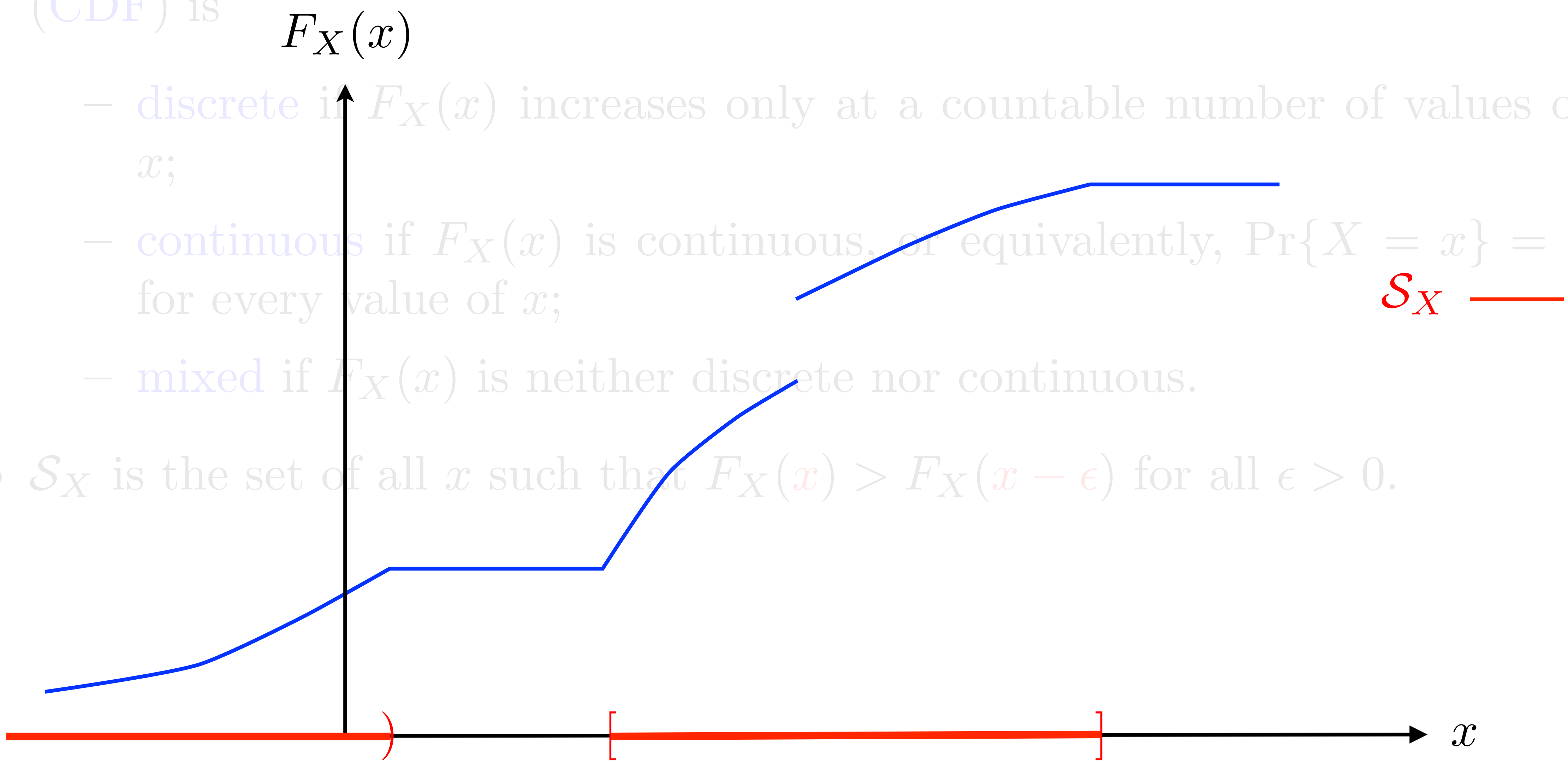
Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.
- $\mathcal{S}_X$ is the set of all x such that $F_X(x) > F_X(x - \epsilon)$ for all $\epsilon > 0$.



Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \le x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.
- $\mathcal{S}_X$ is the set of all x such that $F_X(x) > F_X(x - \epsilon)$ for all $\epsilon > 0$.



Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.
- $\mathcal{S}_X$ is the set of all x such that $F_X(x) > F_X(x - \epsilon)$ for all $\epsilon > 0$.

Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.
- $\mathcal{S}_X$ is the set of all x such that $F_X(x) > F_X(x - \epsilon)$ for all $\epsilon > 0$.

Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.
- $\mathcal{S}_X$ is the set of all x such that $F_X(x) > F_X(x - \epsilon)$ for all $\epsilon > 0$.
- $$Eg(X) = \int_{\mathcal{S}_X} g(x)\, dF_X(x),$$

  where the right hand side is a Lebesgue-Stieltjes integration which covers all cases (i.e., discrete, continuous, and mixed) for the CDF $F_X(x)$.

Real Random Variables

Real Random Variables

- A nonnegative function $f_X(x)$ is called a probability density function (pdf) of X if

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$$

for all x.

Real Random Variables

- A nonnegative function $f_X(x)$ is called a probability density function (pdf) of X if
$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$$
for all x.

- By the fundamental theorem of calculus,
$$\frac{d}{dx}F_X(x) = \frac{d}{dx}\int_{-\infty}^{x} f_X(u)\,du = f_X(x).$$

Real Random Variables

- A nonnegative function $f_X(x)$ is called a probability density function (pdf) of X if
$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$$
for all x.

- By the fundamental theorem of calculus,
$$\frac{d}{dx} F_X(x) = \frac{d}{dx} \int_{-\infty}^{x} f_X(u)\,du = f_X(x).$$

- If X has a pdf, then X is continuous, but not vice versa.

Real Random Variables

Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.
- $\mathcal{S}_X$ is the set of all x such that $F_X(x) > F_X(x - \epsilon)$ for all $\epsilon > 0$.
- $$Eg(X) = \int_{\mathcal{S}_X} g(x)\, dF_X(x),$$

  where the right hand side is a Lebesgue-Stieltjes integration which covers all cases (i.e., discrete, continuous, and mixed) for the CDF $F_X(x)$.

Real Random Variables

- A real r.v. X with cumulative distribution function $F_X(x) = \Pr\{X \leq x\}$ (CDF) is
  - discrete if $F_X(x)$ increases only at a countable number of values of x;
  - continuous if $F_X(x)$ is continuous, or equivalently, $\Pr\{X = x\} = 0$ for every value of x;
  - mixed if $F_X(x)$ is neither discrete nor continuous.
- $\mathcal{S}_X$ is the set of all x such that $F_X(x) > F_X(x - \epsilon)$ for all $\epsilon > 0$.
-
$$Eg(X) = \int_{\mathcal{S}_X} g(x)\, dF_X(x),$$

  $f_X(x)\,dx$

  where the right hand side is a Lebesgue-Stieltjes integration which covers all cases (i.e., discrete, continuous, and mixed) for the CDF $F_X(x)$.

Jointly Distributed Random Variables

Jointly Distributed Random Variables

- Let X and Y be two real random variables with joint CDF $F_{XY}(x, y) = \Pr\{X \leq x, Y \leq y\}$.

Jointly Distributed Random Variables

- Let X and Y be two real random variables with joint CDF $F_{XY}(x, y) = \Pr\{X \leq x, Y \leq y\}$.
- Marginal CDF of X: $F_X(x) = F_{XY}(x, \infty)$

Jointly Distributed Random Variables

- Let X and Y be two real random variables with joint CDF $F_{XY}(x, y) = \Pr\{X \leq x, Y \leq y\}$.
- Marginal CDF of X: $F_X(x) = F_{XY}(x, \infty)$
- A nonnegative function $f_{XY}(x, y)$ is called a joint pdf of X and Y if

Jointly Distributed Random Variables

- Let X and Y be two real random variables with joint CDF $F_{XY}(x, y) = \Pr\{X \leq x, Y \leq y\}$.
- Marginal CDF of X: $F_X(x) = F_{XY}(x, \infty)$
- A nonnegative function $f_{XY}(x, y)$ is called a joint pdf of X and Y if

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v) \, dv \, du$$

Jointly Distributed Random Variables

- Let X and Y be two real random variables with joint CDF $F_{XY}(x,y) = \Pr\{X \leq x, Y \leq y\}$.

- Marginal CDF of X: $F_X(x) = F_{XY}(x,\infty)$

- A nonnegative function $f_{XY}(x,y)$ is called a joint pdf of X and Y if

$$F_{XY}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u,v)\,dv\,du$$

- Conditional pdf of Y given $\{X = x\}$:

Jointly Distributed Random Variables

- Let X and Y be two real random variables with joint CDF $F_{XY}(x, y) = \Pr\{X \leq x, Y \leq y\}$.
- Marginal CDF of X: $F_X(x) = F_{XY}(x, \infty)$
- A nonnegative function $f_{XY}(x, y)$ is called a joint pdf of X and Y if

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v)\, dv\, du$$

- Conditional pdf of Y given $\{X = x\}$:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Jointly Distributed Random Variables

- Let X and Y be two real random variables with joint CDF $F_{XY}(x, y) = \Pr\{X \leq x, Y \leq y\}$.
- Marginal CDF of X: $F_X(x) = F_{XY}(x, \infty)$
- A nonnegative function $f_{XY}(x, y)$ is called a joint pdf of X and Y if

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v)\, dv\, du$$

- Conditional pdf of Y given $\{X = x\}$:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

- Conditional CDF of Y given $\{X = x\}$:

Jointly Distributed Random Variables

- Let X and Y be two real random variables with joint CDF $F_{XY}(x,y) = \Pr\{X \leq x, Y \leq y\}$.
- Marginal CDF of X: $F_X(x) = F_{XY}(x, \infty)$
- A nonnegative function $f_{XY}(x,y)$ is called a joint pdf of X and Y if

$$F_{XY}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u,v)\, dv\, du$$

- Conditional pdf of Y given $\{X = x\}$:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

- Conditional CDF of Y given $\{X = x\}$:

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} f_{Y|X}(v|x)\, dv$$

Variance and Covariance

- Variance of X:

$$\mathrm{var} X = E(X - EX)^2 = EX^2 - (EX)^2$$

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

$$\mathrm{var}X \quad = \quad E(X - EX)^2$$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

$$\begin{aligned}\text{var}X &= E(X - EX)^2 \\ &= E[X^2 - 2(X)(EX) + (EX)^2]\end{aligned}$$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

$$\begin{aligned} \text{var}X &= E(X - EX)^2 \\ &= E[X^2 - 2(X)(EX) + (EX)^2] \\ &= EX^2 - 2(EX)(EX) + (EX)^2 \end{aligned}$$

Variance and Covariance

- Variance of X:

$$\mathrm{var} X = E(X - EX)^2 = EX^2 - (EX)^2$$

$$\begin{aligned}
\mathrm{var} X &= E(X - EX)^2 \\
&= E[X^2 - 2(X)(EX) + (EX)^2] \\
&= EX^2 - 2(EX)(EX) + (EX)^2 \\
&= EX^2 - (EX)^2
\end{aligned}$$

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Variance and Covariance

- Variance of X:

$$\mathrm{var} X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\boxed{\mathrm{cov}(X, X) = \mathrm{var}X}$$

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\text{var}(X + Y) = \text{var}X + \text{var}Y + 2\text{cov}(X, Y)$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\text{var}(X + Y) = \text{var}X + \text{var}Y + 2\text{cov}(X, Y)$

$\text{var}(X + Y)$

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}Y + 2\mathrm{cov}(X, Y)$

$$\begin{aligned} &\mathrm{var}(X + Y) \\ &\quad = \quad E(X + Y)^2 - [E(X + Y)]^2 \end{aligned}$$

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}Y + 2\mathrm{cov}(X, Y)$

$$\begin{aligned}
&\mathrm{var}(X + Y) \\
&= E(X + Y)^2 - [E(X + Y)]^2 \\
&= [EX^2 + 2E(XY) + EY^2] - [(EX)^2 + 2(EX)(EY) + (EY)^2]
\end{aligned}$$

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}Y + 2\mathrm{cov}(X, Y)$

$$\begin{aligned} &\mathrm{var}(X + Y) \\ &= E(X + Y)^2 - [E(X + Y)]^2 \\ &= [\underline{EX^2} + 2E(XY) + EY^2] - [(EX)^2 + 2(EX)(EY) + (EY)^2] \end{aligned}$$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\text{cov}(X,Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\text{var}(X+Y) = \text{var}X + \text{var}Y + 2\text{cov}(X,Y)$

$$\begin{aligned} &\text{var}(X+Y) \\ &\quad = \quad E(X+Y)^2 - [E(X+Y)]^2 \\ &\quad = \quad [\underline{EX^2} + 2E(XY) + EY^2] - [\underline{(EX)^2} + 2(EX)(EY) + (EY)^2] \end{aligned}$$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\text{var}(X + Y) = \text{var}X + \text{var}Y + 2\text{cov}(X, Y)$

$$\begin{aligned}
&\text{var}(X + Y) \\
&\quad = \quad E(X + Y)^2 - [E(X + Y)]^2 \\
&\quad = \quad [\underline{EX^2} + 2E(XY) + EY^2] - [\underline{(EX)^2} + 2(EX)(EY) + (EY)^2] \\
&\quad = \quad \underline{\text{var}X} + \text{var}Y + 2\text{cov}(X, Y)
\end{aligned}$$

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}Y + 2\mathrm{cov}(X, Y)$

$$\begin{aligned} &\mathrm{var}(X + Y) \\ &= E(X+Y)^2 - [E(X+Y)]^2 \\ &= [EX^2 + 2E(XY) + \underline{EY^2}] - [(EX)^2 + 2(EX)(EY) + (EY)^2] \\ &= \mathrm{var}X + \mathrm{var}Y + 2\mathrm{cov}(X, Y) \end{aligned}$$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\text{cov}(X,Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\text{var}(X+Y) = \text{var}X + \text{var}Y + 2\text{cov}(X,Y)$

$$\begin{aligned}
&\text{var}(X+Y) \\
&\quad = \quad E(X+Y)^2 - [E(X+Y)]^2 \\
&\quad = \quad [EX^2 + 2E(XY) + \underline{EY^2}] - [(EX)^2 + 2(EX)(EY) + \underline{(EY)^2}] \\
&\quad = \quad \text{var}X + \text{var}Y + 2\text{cov}(X,Y)
\end{aligned}$$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\text{var}(X + Y) = \text{var}X + \text{var}Y + 2\text{cov}(X, Y)$

$$\begin{aligned} &\text{var}(X+Y) \\ &= E(X+Y)^2 - [E(X+Y)]^2 \\ &= [EX^2 + 2E(XY) + \underline{EY^2}] - [(EX)^2 + 2(EX)(EY) + \underline{(EY)^2}] \\ &= \text{var}X + \underline{\text{var}Y} + 2\text{cov}(X, Y) \end{aligned}$$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\text{var}(X + Y) = \text{var}X + \text{var}Y + 2\text{cov}(X, Y)$

$$\begin{aligned} &\text{var}(X+Y) \\ &= E(X+Y)^2 - [E(X+Y)]^2 \\ &= [EX^2 + 2\underline{E(XY)} + EY^2] - [(EX)^2 + 2(EX)(EY) + (EY)^2] \\ &= \text{var}X + \text{var}Y + 2\text{cov}(X, Y) \end{aligned}$$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\text{var}(X + Y) = \text{var}X + \text{var}Y + 2\text{cov}(X, Y)$

$$\begin{aligned}
&\text{var}(X + Y) \\
&\quad = \quad E(X + Y)^2 - [E(X + Y)]^2 \\
&\quad = \quad [EX^2 + 2\underline{E(XY)} + EY^2] - [(EX)^2 + 2\underline{(EX)(EY)} + (EY)^2] \\
&\quad = \quad \text{var}X + \text{var}Y + 2\text{cov}(X, Y)
\end{aligned}$$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\text{cov}(X,Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\text{var}(X+Y) = \text{var}X + \text{var}Y + 2\text{cov}(X,Y)$

$$\begin{aligned} &\text{var}(X+Y) \\ &= E(X+Y)^2 - [E(X+Y)]^2 \\ &= [EX^2 + 2\underline{E(XY)} + EY^2] - [(EX)^2 + 2\underline{(EX)(EY)} + (EY)^2] \\ &= \text{var}X + \text{var}Y + 2\underline{\text{cov}(X,Y)} \end{aligned}$$

Variance and Covariance

- Variance of X:

$$\text{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\text{var}(X + Y) = \text{var}X + \text{var}Y + 2\text{cov}(X, Y)$

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}Y + 2\mathrm{cov}(X, Y)$
2. If $X \perp Y$, then $\mathrm{cov}(X, Y) = 0$, or X and Y are uncorrelated. However, the converse is not true.

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$(EX)(EY)$

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}Y + 2\mathrm{cov}(X, Y)$
2. If $X \perp Y$, then $\mathrm{cov}(X, Y) = 0$, or X and Y are uncorrelated. However, the converse is not true.

Variance and Covariance

- Variance of X:

$$\mathrm{var}X = E(X - EX)^2 = EX^2 - (EX)^2$$

- Covariance between X and Y:

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

Remarks

1. $\mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}Y + 2\mathrm{cov}(X, Y)$
2. If $X \perp Y$, then $\mathrm{cov}(X, Y) = 0$, or X and Y are uncorrelated. However, the converse is not true.
3. If $X_1, X_2, \cdots, X_n$ are mutually independent, then

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{var}X_i$$

Random Vectors

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^n$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^{n}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \operatorname{cov}(X_i, X_j) \,]_{i,j=1}^{n}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \underline{\mathrm{cov}(X_i, X_j)} \,]_{i,j=1}^{n}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$\mathrm{cov}(X_i, X_i) = \mathrm{var} X_i$

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \underline{\mathrm{cov}(X_i, X_j)} \,]_{i,j=1}^{n}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^{n}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^{n}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j)\,]_{i,j=1}^{n}$$

$$E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top \quad = \quad E\, [\, (X_i - EX_i)(X_j - EX_j)\,]_{i,j=1}^{n}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j)\,]_{i,j=1}^{n}$$

$$E\underline{(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top} \quad = \quad E\,[\,(X_i - EX_i)(X_j - EX_j)\,]_{i,j=1}^{n}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^n$$

$$E\underline{(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top} \quad = \quad E\,[\, \underline{(X_i - EX_i)(X_j - EX_j)} \,]_{i,j=1}^n$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^n$$

$$\begin{aligned} E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top &= E\,[\,(X_i - EX_i)(X_j - EX_j)\,]_{i,j=1}^n \\ &= [\,E(X_i - EX_i)(X_j - EX_j)\,] \end{aligned}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^{n}$$

$$\begin{aligned} E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top &= E\,[\,(X_i - EX_i)(X_j - EX_j)\,]_{i,j=1}^{n} \\ &= [\,E(X_i - EX_i)(X_j - EX_j)\,] \\ &= [\,\mathrm{cov}(X_i, X_j)\,] \end{aligned}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \, X_2 \, \cdots \, X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^n$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \, X_2 \, \cdots \, X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^n$$

- Correlation matrix: $\tilde{K}_{\mathbf{X}} = E\mathbf{X}\mathbf{X}^\top = [\, EX_iX_j \,]$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \, X_2 \, \cdots \, X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^n$$

- Correlation matrix: $\tilde{K}_{\mathbf{X}} = E\mathbf{X}\mathbf{X}^\top = [\, EX_iX_j \,]$
- Relations between $K_{\mathbf{X}}$ and $\tilde{K}_{\mathbf{X}}$:

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^n$$

- Correlation matrix: $\tilde{K}_{\mathbf{X}} = E\mathbf{X}\mathbf{X}^\top = [\, E X_i X_j \,]$
- Relations between $K_{\mathbf{X}}$ and $\tilde{K}_{\mathbf{X}}$:

$$K_{\mathbf{X}} \;\; = \;\; \tilde{K}_{\mathbf{X}} - (E\mathbf{X})(E\mathbf{X})^\top$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \, X_2 \, \cdots \, X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^n$$

- Correlation matrix: $\tilde{K}_{\mathbf{X}} = E\mathbf{X}\mathbf{X}^\top = [\, EX_iX_j \,]$
- Relations between $K_{\mathbf{X}}$ and $\tilde{K}_{\mathbf{X}}$:

$$\begin{aligned} K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}} - (E\mathbf{X})(E\mathbf{X})^\top \\ K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}-E\mathbf{X}} \end{aligned}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^{n}$$

- Correlation matrix: $\tilde{K}_{\mathbf{X}} = E\mathbf{X}\mathbf{X}^\top = [\, EX_iX_j \,]$
- Relations between $K_{\mathbf{X}}$ and $\tilde{K}_{\mathbf{X}}$:

$$\begin{aligned} K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}} - (E\mathbf{X})(E\mathbf{X})^\top \\ K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}-E\mathbf{X}} \end{aligned}$$

- These are vector generalizations of

$$\begin{aligned} \mathrm{var}X &= EX^2 - (EX)^2 \\ \mathrm{var}X &= E(X - EX)^2 \end{aligned}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^{n}$$

- Correlation matrix: $\tilde{K}_{\mathbf{X}} = E\mathbf{X}\mathbf{X}^\top = [\, EX_iX_j \,]$
- Relations between $K_{\mathbf{X}}$ and $\tilde{K}_{\mathbf{X}}$:

$$\begin{aligned} K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}} - (E\mathbf{X})(E\mathbf{X})^\top \\ K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}-E\mathbf{X}} \end{aligned}$$

- These are vector generalizations of

$$\begin{aligned} \mathrm{var}X &= EX^2 - (EX)^2 \\ \mathrm{var}X &= E(X - EX)^2 \end{aligned}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^{n}$$

- Correlation matrix: $\tilde{K}_{\mathbf{X}} = E\mathbf{X}\mathbf{X}^\top = [\, EX_iX_j \,]$
- Relations between $K_{\mathbf{X}}$ and $\tilde{K}_{\mathbf{X}}$:

$$\begin{aligned} K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}} - (E\mathbf{X})(E\mathbf{X})^\top \\ K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}-E\mathbf{X}} \end{aligned}$$

- These are vector generalizations of

$$\begin{aligned} \mathrm{var} X &= EX^2 - (EX)^2 \\ \mathrm{var} X &= E(X - EX)^2 \end{aligned}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \; X_2 \; \cdots \; X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^n$$

- Correlation matrix: $\tilde{K}_{\mathbf{X}} = E\mathbf{X}\mathbf{X}^\top = [\, EX_iX_j \,]$
- Relations between $K_{\mathbf{X}}$ and $\tilde{K}_{\mathbf{X}}$:

$$\begin{aligned} K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}} - (E\mathbf{X})(E\mathbf{X})^\top \\ K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}-E\mathbf{X}} \end{aligned}$$

- These are vector generalizations of

$$\begin{aligned} \mathrm{var}X &= EX^2 - (EX)^2 \\ \mathrm{var}X &= E(X - EX)^2 \end{aligned}$$

Random Vectors

- Let $\mathbf{X} = [\, X_1 \, X_2 \, \cdots \, X_n \,]^\top$.
- Covariance matrix:

$$K_{\mathbf{X}} = E(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})^\top = [\, \mathrm{cov}(X_i, X_j) \,]_{i,j=1}^{n}$$

- Correlation matrix: $\tilde{K}_{\mathbf{X}} = E\mathbf{X}\mathbf{X}^\top = [\, EX_iX_j \,]$
- Relations between $K_{\mathbf{X}}$ and $\tilde{K}_{\mathbf{X}}$:

$$\begin{aligned} K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}} - (E\mathbf{X})(E\mathbf{X})^\top \\ K_{\mathbf{X}} &= \tilde{K}_{\mathbf{X}-E\mathbf{X}} \end{aligned}$$

- These are vector generalizations of

$$\begin{aligned} \mathrm{var}X &= EX^2 - (EX)^2 \\ \mathrm{var}X &= E(X - EX)^2 \end{aligned}$$

Gaussian Distribution

Gaussian Distribution

- $\mathcal{N}(\mu, \sigma^2)$ – Gaussian distribution with mean μ and variance σ^2:

Gaussian Distribution

- $\mathcal{N}(\mu, \sigma^2)$ – Gaussian distribution with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

Gaussian Distribution

- $\mathcal{N}(\mu, \sigma^2)$ – Gaussian distribution with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

Gaussian Distribution

- $\mathcal{N}(\mu, \sigma^2)$ – Gaussian distribution with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- $\mathcal{N}(\boldsymbol{\mu}, K)$ – multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix K, i.e., the joint pdf of the distribution is given by

Gaussian Distribution

- $\mathcal{N}(\mu, \sigma^2)$ – Gaussian distribution with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- $\mathcal{N}(\boldsymbol{\mu}, K)$ – multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix K, i.e., the joint pdf of the distribution is given by

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \Re^n$$

Gaussian Distribution

- $\mathcal{N}(\mu, \sigma^2)$ – Gaussian distribution with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- $\mathcal{N}(\boldsymbol{\mu}, K)$ – multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix K, i.e., the joint pdf of the distribution is given by

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \Re^n$$

where K is a symmetric positive definite matrix.

Gaussian Distribution

- $\mathcal{N}(\mu, \sigma^2)$ – Gaussian distribution with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- $\mathcal{N}(\boldsymbol{\mu}, K)$ – multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix K, i.e., the joint pdf of the distribution is given by

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \Re^n$$

where K is a symmetric positive definite matrix.

Gaussian Distribution

- $\mathcal{N}(\mu, \sigma^2)$ – Gaussian distribution with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- $\mathcal{N}(\boldsymbol{\mu}, K)$ – multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix K, i.e., the joint pdf of the distribution is given by

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \underline{K^{-1}}(\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \Re^n$$

where K is a symmetric positive definite matrix.

Gaussian Distribution

- $\mathcal{N}(\mu, \sigma^2)$ – Gaussian distribution with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- $\mathcal{N}(\boldsymbol{\mu}, K)$ – multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix K, i.e., the joint pdf of the distribution is given by

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \Re^n$$

where K is a symmetric positive definite matrix.

Gaussian Distribution

- $\mathcal{N}(\mu, \sigma^2)$ – Gaussian distribution with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- $\mathcal{N}(\boldsymbol{\mu}, K)$ – multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix K, i.e., the joint pdf of the distribution is given by

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n |K|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top K^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad \mathbf{x} \in \Re^n$$

where K is a symmetric positive definite matrix.

10.1 Preliminaries

Definition 10.1 A square matrix K is symmetric if $K^\top = K$.

Definition 10.1 A square matrix K is symmetric if $K^\top = K$.

Definition 10.2 An $n \times n$ matrix K is positive definite if

Definition 10.1 A square matrix K is symmetric if $K^\top = K$.

Definition 10.2 An $n \times n$ matrix K is positive definite if

$$\mathbf{x}^\top K \mathbf{x} > 0$$

for all nonzero column n-vector $\mathbf{x}$,

Definition 10.1 A square matrix K is symmetric if $K^\top = K$.

Definition 10.2 An $n \times n$ matrix K is positive definite if

$$\mathbf{x}^\top K \mathbf{x} > 0$$

for all nonzero column n-vector $\mathbf{x}$, and is positive semidefinite if

$$\mathbf{x}^\top K \mathbf{x} \geq 0$$

for all column n-vector $\mathbf{x}$.

Definition 10.1 A square matrix K is symmetric if $K^\top = K$.

Definition 10.2 An $n \times n$ matrix K is positive definite if

$$\mathbf{x}^\top K \mathbf{x} > 0$$

for all nonzero column n-vector $\mathbf{x}$, and is positive semidefinite if

$$\mathbf{x}^\top K \mathbf{x} \geq 0$$

for all column n-vector $\mathbf{x}$.

Proposition 10.3 A covariance matrix is both symmetric and positive semidefinite.

Diagonalization

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q \Lambda Q^\top$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q \Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q \Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

Let $\mathbf{q}_i$ be the ith column of Q. Then

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

Let $\mathbf{q}_i$ be the ith column of Q. Then

$$\mathbf{q}_i \mathbf{q}_j^\top = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

Let $\mathbf{q}_i$ be the ith column of Q. Then

$$\mathbf{q}_i \mathbf{q}_j^\top = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Therefore,

$$QQ^\top = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} — & \mathbf{q}_1^\top & — \\ & \vdots & \\ — & \mathbf{q}_n^\top & — \end{bmatrix} = [\mathbf{q}_i \mathbf{q}_j^\top] = I.$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

Let $\mathbf{q}_i$ be the ith column of Q. Then

$$\mathbf{q}_i \mathbf{q}_j^\top = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Therefore,

$$QQ^\top = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} — & \mathbf{q}_1^\top & — \\ & \vdots & \\ — & \mathbf{q}_n^\top & — \end{bmatrix} = \underline{[\mathbf{q}_i \mathbf{q}_j^\top]} = I.$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

Let $\mathbf{q}_i$ be the ith column of Q. Then

$$\mathbf{q}_i \mathbf{q}_j^\top = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Therefore,

$$QQ^\top = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} - & \mathbf{q}_1^\top & - \\ & \vdots & \\ - & \mathbf{q}_n^\top & - \end{bmatrix} = \left[\mathbf{q}_i \mathbf{q}_j^\top\right] = I.$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

Let $\mathbf{q}_i$ be the ith column of Q. Then

$$\mathbf{q}_i \mathbf{q}_j^\top = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Therefore,

$$QQ^\top = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} — & \mathbf{q}_1^\top & — \\ & \vdots & \\ — & \mathbf{q}_n^\top & — \end{bmatrix} = [\mathbf{q}_i \mathbf{q}_j^\top] = I.$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

Let $\mathbf{q}_i$ be the ith column of Q. Then

$$\mathbf{q}_i \mathbf{q}_j^\top = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Therefore,

$$QQ^\top = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} — & \mathbf{q}_1^\top & — \\ & \vdots & \\ — & \mathbf{q}_n^\top & — \end{bmatrix} = \left[\mathbf{q}_i \mathbf{q}_j^\top\right] = I.$$

Hence, $Q^{-1} = Q^\top$.

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q \Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1.$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

  where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1$.

Consider

$$|Q|^2 = |Q||Q^\top| = |QQ^\top| = |I| = 1.$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1$.

Consider

$$|Q|^2 = |Q||Q^\top| = |QQ^\top| = |I| = 1.$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

  where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1.$

Consider

$$|Q|^2 = |Q||Q^\top| = |QQ^\top| = |I| = 1.$$

Then

$$|Q| = |Q^\top| = \pm 1.$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

  where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1.$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

  where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1$.

- Let $\lambda_i = i$th diagonal element of Λ and $\mathbf{q}_i = i$th column of Q.

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

  where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1$.
- Let $\lambda_i = i$th diagonal element of Λ and $\mathbf{q}_i = i$th column of Q.
- Then $KQ = (Q\Lambda Q^\top)Q = Q\Lambda(Q^\top Q) = Q\Lambda$, or

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1$.
- Let $\lambda_i = i$th diagonal element of Λ and $\mathbf{q}_i = i$th column of Q.
- Then $KQ = (Q\Lambda Q^\top)Q = Q\Lambda(Q^\top Q) = Q\Lambda$, or

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q \Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1$.
- Let $\lambda_i = i$th diagonal element of Λ and $\mathbf{q}_i = i$th column of Q.
- Then $KQ = (Q\Lambda Q^\top)Q = Q\Lambda(Q^\top Q) = Q\Lambda$, or

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1$.
- Let $\lambda_i = i$th diagonal element of Λ and $\mathbf{q}_i = i$th column of Q.
- Then $KQ = (Q\Lambda Q^\top)Q = Q\Lambda(Q^\top Q) = Q\Lambda$, or

$$K\mathbf{q}_i = \lambda_i \mathbf{q}_i.$$

Diagonalization

Diagonalization

Consider

$$KQ = K \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} K\mathbf{q}_1 & \cdots & K\mathbf{q}_n \end{bmatrix}$$

Diagonalization

Consider

$$KQ = K \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} K\mathbf{q}_1 & \cdots & K\mathbf{q}_n \end{bmatrix}$$

and

$$Q\Lambda = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{q}_1 & \cdots & \lambda_n\mathbf{q}_n \end{bmatrix}.$$

Diagonalization

Consider

$$KQ = K \left[\begin{array}{ccc} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{array} \right] = \left[\begin{array}{ccc} K\mathbf{q}_1 & \cdots & K\mathbf{q}_n \end{array} \right]$$

and

$$Q\Lambda = \left[\begin{array}{ccc} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{array} \right] \left[\begin{array}{ccc} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{array} \right] = \left[\begin{array}{ccc} \lambda_1\mathbf{q}_1 & \cdots & \lambda_n\mathbf{q}_n \end{array} \right].$$

Therefore $KQ = Q\Lambda$ is equivalent to

$$K\mathbf{q}_i = \lambda_i\mathbf{q}_i$$

for all i.

Diagonalization

Consider

$$KQ = K \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} \underline{K\mathbf{q}_1} & \cdots & K\mathbf{q}_n \end{bmatrix}$$

and

$$Q\Lambda = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{q}_1 & \cdots & \lambda_n\mathbf{q}_n \end{bmatrix}.$$

Therefore $KQ = Q\Lambda$ is equivalent to

$$K\mathbf{q}_i = \lambda_i\mathbf{q}_i$$

for all i.

Diagonalization

Consider

$$KQ = K \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} = \begin{bmatrix} \underline{K\mathbf{q}_1} & \cdots & K\mathbf{q}_n \end{bmatrix}$$

and

$$Q\Lambda = \begin{bmatrix} | & & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \begin{bmatrix} \underline{\lambda_1\mathbf{q}_1} & \cdots & \lambda_n\mathbf{q}_n \end{bmatrix}.$$

Therefore $KQ = Q\Lambda$ is equivalent to

$$K\mathbf{q}_i = \lambda_i\mathbf{q}_i$$

for all i.

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

  where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1$.
- Let $\lambda_i = i$th diagonal element of Λ and $\mathbf{q}_i = i$th column of Q.
- Then $KQ = (Q\Lambda Q^\top)Q = Q\Lambda(Q^\top Q) = Q\Lambda$, or

$$K\mathbf{q}_i = \lambda_i \mathbf{q}_i.$$

Diagonalization

- A symmetric matrix K can be diagonalized as

$$K = Q\Lambda Q^\top$$

where Λ is a diagonal matrix and Q (also $Q^\top$) is an orthogonal matrix, i.e.,

$$Q^{-1} = Q^\top.$$

- $|Q| = |Q^\top| = \pm 1$.
- Let $\lambda_i = i$th diagonal element of Λ and $\mathbf{q}_i = i$th column of Q.
- Then $KQ = (Q\Lambda Q^\top)Q = Q\Lambda(Q^\top Q) = Q\Lambda$, or

$$K\mathbf{q}_i = \lambda_i \mathbf{q}_i.$$

- That is, $\mathbf{q}_i$ is an eigenvector of K with eigenvalue λ_i.

Proposition 10.4 The eigenvalues of a positive semidefinite matrix are non-negative.

Proof

Proposition 10.4 The eigenvalues of a positive semidefinite matrix are non-negative.

Proof

1. Consider eigenvector $\mathbf{q} \neq 0$ and corresponding eigenvalue λ of K, i.e.,

$$K\mathbf{q} = \lambda\mathbf{q}.$$

Proposition 10.4 The eigenvalues of a positive semidefinite matrix are non-negative.

Proof

1. Consider eigenvector $\mathbf{q} \neq 0$ and corresponding eigenvalue λ of K, i.e.,

$$K\mathbf{q} = \lambda\mathbf{q}.$$

2. Since K is positive semidefinite,

$$0 \leq \mathbf{q}^\top K\mathbf{q} = \mathbf{q}^\top(\lambda\mathbf{q}) = \lambda(\mathbf{q}^\top\mathbf{q}).$$

Proposition 10.4 The eigenvalues of a positive semidefinite matrix are non-negative.

Proof

1. Consider eigenvector $\mathbf{q} \neq 0$ and corresponding eigenvalue λ of K, i.e.,

$$K\mathbf{q} = \lambda\mathbf{q}.$$

2. Since K is positive semidefinite,

$$\underline{0 \leq \mathbf{q}^\top K\mathbf{q}} = \mathbf{q}^\top(\lambda\mathbf{q}) = \lambda(\mathbf{q}^\top\mathbf{q}).$$

Proposition 10.4 The eigenvalues of a positive semidefinite matrix are non-negative.

Proof

1. Consider eigenvector $\mathbf{q} \neq 0$ and corresponding eigenvalue λ of K, i.e.,

$$K\mathbf{q} = \lambda\mathbf{q}.$$

2. Since K is positive semidefinite,

$$0 \leq \mathbf{q}^\top \underline{K\mathbf{q}} = \mathbf{q}^\top(\lambda\mathbf{q}) = \lambda(\mathbf{q}^\top\mathbf{q}).$$

Proposition 10.4 The eigenvalues of a positive semidefinite matrix are non-negative.

Proof

1. Consider eigenvector $\mathbf{q} \neq 0$ and corresponding eigenvalue λ of K, i.e.,

$$K\mathbf{q} = \lambda\mathbf{q}.$$

2. Since K is positive semidefinite,

$$0 \leq \mathbf{q}^\top \underline{K\mathbf{q}} = \mathbf{q}^\top (\underline{\lambda\mathbf{q}}) = \lambda(\mathbf{q}^\top \mathbf{q}).$$

Proposition 10.4 The eigenvalues of a positive semidefinite matrix are non-negative.

Proof

1. Consider eigenvector $\mathbf{q} \neq 0$ and corresponding eigenvalue λ of K, i.e.,

$$K\mathbf{q} = \lambda\mathbf{q}.$$

2. Since K is positive semidefinite,

$$0 \leq \mathbf{q}^\top K\mathbf{q} = \mathbf{q}^\top(\lambda\mathbf{q}) = \underline{\lambda(\mathbf{q}^\top\mathbf{q})}.$$

Proposition 10.4 The eigenvalues of a positive semidefinite matrix are non-negative.

Proof

1. Consider eigenvector $\mathbf{q} \neq 0$ and corresponding eigenvalue λ of K, i.e.,

$$K\mathbf{q} = \lambda\mathbf{q}.$$

2. Since K is positive semidefinite,

$$0 \leq \mathbf{q}^\top K\mathbf{q} = \mathbf{q}^\top(\lambda\mathbf{q}) = \lambda(\mathbf{q}^\top\mathbf{q}).$$

3. $\lambda \geq 0$ because $\mathbf{q}^\top\mathbf{q} = \|\mathbf{q}\|^2 > 0$.

Proposition 10.4 The eigenvalues of a positive semidefinite matrix are non-negative.

Proof

1. Consider eigenvector $\mathbf{q} \neq 0$ and corresponding eigenvalue λ of K, i.e.,

$$K\mathbf{q} = \lambda\mathbf{q}.$$

2. Since K is positive semidefinite,

$$0 \leq \mathbf{q}^\top K\mathbf{q} = \mathbf{q}^\top(\lambda\mathbf{q}) = \lambda(\underline{\mathbf{q}^\top\mathbf{q}}).$$

3. $\lambda \geq 0$ because $\mathbf{q}^\top\mathbf{q} = \|\mathbf{q}\|^2 > 0$.

Proposition 10.4 The eigenvalues of a positive semidefinite matrix are nonnegative.

Proof

1. Consider eigenvector $\mathbf{q} \neq 0$ and corresponding eigenvalue λ of K, i.e.,

$$K\mathbf{q} = \lambda\mathbf{q}.$$

2. Since K is positive semidefinite,

$$0 \leq \mathbf{q}^\top K\mathbf{q} = \mathbf{q}^\top(\lambda\mathbf{q}) = \lambda(\mathbf{q}^\top\mathbf{q}).$$

3. $\lambda \geq 0$ because $\mathbf{q}^\top\mathbf{q} = \|\mathbf{q}\|^2 > 0$.

Remark Since a covariance matrix is both symmetric and positive semidefinite, it is diagonalizable and its eigenvalues are nonnegative.

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = A K_{\mathbf{X}} A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A \tilde{K}_{\mathbf{X}} A^{\top}.$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$K_{\mathbf{Y}} = E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$K_{\mathbf{Y}} = E\mathbf{Y}\underline{\mathbf{Y}^{\top}} - (E\mathbf{Y})(E\mathbf{Y})^{\top}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$\begin{aligned} K_{\mathbf{Y}} &= E\mathbf{Y}\underline{\mathbf{Y}^{\top}} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\ &= E(A\mathbf{X})(\underline{\mathbf{X}^{\top}A}) - (EA\mathbf{X})(E\mathbf{X}^{\top}A) \end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$\begin{aligned} K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\ &= E\underline{(A\mathbf{X})(\mathbf{X}^{\top}A)} - (EA\mathbf{X})(E\mathbf{X}^{\top}A) \end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$\begin{aligned} K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\ &= E(A\mathbf{X})(\mathbf{X}^{\top}A) - (EA\mathbf{X})(E\mathbf{X}^{\top}A) \\ &= E(A\mathbf{X}\mathbf{X}^{\top}A^{\top}) - A(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top} \end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$\begin{aligned}
K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\
&= E(A\mathbf{X})(\mathbf{X}^{\top}A) - (EA\mathbf{X})(E\mathbf{X}^{\top}A) \\
&= E(A\mathbf{X}\mathbf{X}^{\top}A^{\top}) - A(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top}
\end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^\top$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^\top.$$

$$\begin{aligned} K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^\top - (E\mathbf{Y})(E\mathbf{Y})^\top \\ &= E(A\mathbf{X})(\mathbf{X}^\top A) - (\underline{E}A\mathbf{X})(E\mathbf{X}^\top A) \\ &= E(A\mathbf{X}\mathbf{X}^\top A^\top) - A(\underline{E}\mathbf{X})(E\mathbf{X}^\top)A^\top \end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = A K_{\mathbf{X}} A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A \tilde{K}_{\mathbf{X}} A^{\top}.$$

$$\begin{aligned} K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\ &= E(A\mathbf{X})(\mathbf{X}^{\top}A) - (EA\mathbf{X})(E\mathbf{X}^{\top}A) \\ &= E(A\mathbf{X}\mathbf{X}^{\top}A^{\top}) - A(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top} \end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$\begin{aligned} K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\ &= E(A\mathbf{X})(\mathbf{X}^{\top}A) - (EA\mathbf{X})(\underline{E}\mathbf{X}^{\top}A) \\ &= E(A\mathbf{X}\mathbf{X}^{\top}A^{\top}) - A(E\mathbf{X})(\underline{E}\mathbf{X}^{\top})A^{\top} \end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$\begin{aligned} K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\ &= E(A\mathbf{X})(\mathbf{X}^{\top}A) - (EA\mathbf{X})(E\mathbf{X}^{\top}A) \\ &= E(A\mathbf{X}\mathbf{X}^{\top}A^{\top}) - A(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top} \end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^\top$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^\top.$$

$$\begin{aligned}
K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^\top - (E\mathbf{Y})(E\mathbf{Y})^\top \\
&= E(A\mathbf{X})(\mathbf{X}^\top A) - (EA\mathbf{X})(E\mathbf{X}^\top A) \\
&= \underline{E}(A\mathbf{X}\mathbf{X}^\top A^\top) - A(E\mathbf{X})(E\mathbf{X}^\top)A^\top \\
&= A(\underline{E}\mathbf{X}\mathbf{X}^\top)A^\top - A(E\mathbf{X})(E\mathbf{X}^\top)A^\top
\end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$\begin{aligned}
K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\
&= E(A\mathbf{X})(\mathbf{X}^{\top}A) - (EA\mathbf{X})(E\mathbf{X}^{\top}A) \\
&= E(A\mathbf{X}\mathbf{X}^{\top}A^{\top}) - A(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top} \\
&= \underline{A}(E\mathbf{X}\mathbf{X}^{\top})A^{\top} - \underline{A}(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top}
\end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = A K_{\mathbf{X}} A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A \tilde{K}_{\mathbf{X}} A^{\top}.$$

$$\begin{aligned}
K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\
&= E(A\mathbf{X})(\mathbf{X}^{\top} A) - (EA\mathbf{X})(E\mathbf{X}^{\top} A) \\
&= E(A\mathbf{X}\mathbf{X}^{\top} A^{\top}) - A(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top} \\
&= \underline{A}(E\mathbf{X}\mathbf{X}^{\top})A^{\top} - \underline{A}(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top} \\
&= \underline{A}\left[E\mathbf{X}\mathbf{X}^{\top} - (E\mathbf{X})(E\mathbf{X})^{\top}\right]A^{\top}
\end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$\begin{aligned} K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\ &= E(A\mathbf{X})(\mathbf{X}^{\top}A) - (EA\mathbf{X})(E\mathbf{X}^{\top}A) \\ &= E(A\mathbf{X}\mathbf{X}^{\top}A^{\top}) - A(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top} \\ &= A(E\mathbf{X}\mathbf{X}^{\top})\underline{A^{\top}} - A(E\mathbf{X})(E\mathbf{X}^{\top})\underline{A^{\top}} \\ &= A\left[E\mathbf{X}\mathbf{X}^{\top} - (E\mathbf{X})(E\mathbf{X})^{\top}\right]A^{\top} \end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$\begin{aligned}
K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\
&= E(A\mathbf{X})(\mathbf{X}^{\top}A) - (EA\mathbf{X})(E\mathbf{X}^{\top}A) \\
&= E(A\mathbf{X}\mathbf{X}^{\top}A^{\top}) - A(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top} \\
&= A(E\mathbf{X}\mathbf{X}^{\top})\underline{A^{\top}} - A(E\mathbf{X})(E\mathbf{X}^{\top})\underline{A^{\top}} \\
&= A\left[E\mathbf{X}\mathbf{X}^{\top} - (E\mathbf{X})(E\mathbf{X})^{\top}\right]\underline{A^{\top}}
\end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^\top$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^\top.$$

$$\begin{aligned}
K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^\top - (E\mathbf{Y})(E\mathbf{Y})^\top \\
&= E(A\mathbf{X})(\mathbf{X}^\top A) - (EA\mathbf{X})(E\mathbf{X}^\top A) \\
&= E(A\mathbf{X}\mathbf{X}^\top A^\top) - A(E\mathbf{X})(E\mathbf{X}^\top)A^\top \\
&= A(E\mathbf{X}\mathbf{X}^\top)A^\top - A(E\mathbf{X})(E\mathbf{X}^\top)A^\top \\
&= A\left[\underline{E\mathbf{X}\mathbf{X}^\top - (E\mathbf{X})(E\mathbf{X})^\top}\right]A^\top
\end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^\top$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^\top.$$

$$\begin{aligned}
K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^\top - (E\mathbf{Y})(E\mathbf{Y})^\top \\
&= E(A\mathbf{X})(\mathbf{X}^\top A) - (EA\mathbf{X})(E\mathbf{X}^\top A) \\
&= E(A\mathbf{X}\mathbf{X}^\top A^\top) - A(E\mathbf{X})(E\mathbf{X}^\top)A^\top \\
&= A(E\mathbf{X}\mathbf{X}^\top)A^\top - A(E\mathbf{X})(E\mathbf{X}^\top)A^\top \\
&= A\left[\underline{E\mathbf{X}\mathbf{X}^\top - (E\mathbf{X})(E\mathbf{X})^\top}\right]A^\top \\
&= A\underline{K_{\mathbf{X}}}A^\top
\end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

$$\begin{aligned}
K_{\mathbf{Y}} &= E\mathbf{Y}\mathbf{Y}^{\top} - (E\mathbf{Y})(E\mathbf{Y})^{\top} \\
&= E(A\mathbf{X})(\mathbf{X}^{\top}A) - (EA\mathbf{X})(E\mathbf{X}^{\top}A) \\
&= E(A\mathbf{X}\mathbf{X}^{\top}A^{\top}) - A(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top} \\
&= A(E\mathbf{X}\mathbf{X}^{\top})A^{\top} - A(E\mathbf{X})(E\mathbf{X}^{\top})A^{\top} \\
&= A\left[E\mathbf{X}\mathbf{X}^{\top} - (E\mathbf{X})(E\mathbf{X})^{\top}\right]A^{\top} \\
&= AK_{\mathbf{X}}A^{\top}
\end{aligned}$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^\top$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^\top.$$

Proposition 10.6 (Decorrelation) Let $\mathbf{Y} = Q^\top\mathbf{X}$, where $K_{\mathbf{X}} = Q\Lambda Q^\top$.

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

Proposition 10.6 (Decorrelation) Let $\mathbf{Y} = Q^{\top}\mathbf{X}$, where $K_{\mathbf{X}} = Q\Lambda Q^{\top}$. Then $K_{\mathbf{Y}} = \Lambda$, i.e.,

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^\top$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^\top.$$

Proposition 10.6 (Decorrelation) Let $\mathbf{Y} = Q^\top \mathbf{X}$, where $K_{\mathbf{X}} = Q\Lambda Q^\top$. Then $K_{\mathbf{Y}} = \Lambda$, i.e.,

1. the random variables in $\mathbf{Y}$ are uncorrelated

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

Proposition 10.6 (Decorrelation) Let $\mathbf{Y} = Q^{\top}\mathbf{X}$, where $K_{\mathbf{X}} = Q\Lambda Q^{\top}$. Then $K_{\mathbf{Y}} = \Lambda$, i.e.,

1. the random variables in $\mathbf{Y}$ are uncorrelated
2. $\operatorname{var} Y_i = \lambda_i$ for all i

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q \Lambda Q^\top$ is a diagonalization of $K_{\mathbf{X}}$. Then $K_{\mathbf{Y}} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q\Lambda Q^\top$ is a diagonalization of $K_\mathbf{X}$. Then $K_\mathbf{Y} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

1. By Proposition 10.5,

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q\Lambda Q^\top$ is a diagonalization of $K_\mathbf{X}$. Then $K_\mathbf{Y} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\text{var}\, Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

1. By Proposition 10.5,

$$K_\mathbf{Y} = Q^\top K_\mathbf{X} Q$$

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q\Lambda Q^\top$ is a diagonalization of $K_\mathbf{X}$. Then $K_\mathbf{Y} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\mathrm{var}\, Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

1. By Proposition 10.5,

$$K_\mathbf{Y} = Q^\top \underline{K_\mathbf{X}} Q$$

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q\Lambda Q^\top$ is a diagonalization of $K_\mathbf{X}$. Then $K_\mathbf{Y} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\text{var}\, Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

1. By Proposition 10.5,

$$\begin{aligned} K_\mathbf{Y} &= Q^\top \underline{K_\mathbf{X}} Q \\ &= Q^\top (\underline{Q\Lambda Q^\top}) Q \end{aligned}$$

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q\Lambda Q^\top$ is a diagonalization of $K_\mathbf{X}$. Then $K_\mathbf{Y} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\mathrm{var}\, Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

1. By Proposition 10.5,

$$\begin{aligned} K_\mathbf{Y} &= Q^\top K_\mathbf{X} Q \\ &= Q^\top (Q\Lambda Q^\top) Q \\ &= (Q^\top Q)\Lambda(Q^\top Q) \end{aligned}$$

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q\Lambda Q^\top$ is a diagonalization of $K_\mathbf{X}$. Then $K_\mathbf{Y} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\text{var}\, Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

1. By Proposition 10.5,

$$\begin{aligned} K_\mathbf{Y} &= Q^\top K_\mathbf{X} Q \\ &= Q^\top (Q\Lambda Q^\top) Q \\ &= (Q^\top Q)\Lambda(Q^\top Q) \end{aligned}$$

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q\Lambda Q^\top$ is a diagonalization of $K_\mathbf{X}$. Then $K_\mathbf{Y} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\text{var}\, Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

1. By Proposition 10.5,

$$\begin{aligned} K_\mathbf{Y} &= Q^\top K_\mathbf{X} Q \\ &= Q^\top (Q\Lambda Q^\top) Q \\ &= (Q^\top Q)\Lambda(Q^\top Q) \end{aligned}$$

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q\Lambda Q^\top$ is a diagonalization of $K_{\mathbf{X}}$. Then $K_{\mathbf{Y}} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\text{var}\, Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

1. By Proposition 10.5,

$$\begin{aligned} K_{\mathbf{Y}} &= Q^\top K_{\mathbf{X}} Q \\ &= Q^\top (Q\Lambda Q^\top) Q \\ &= (Q^\top Q)\Lambda(Q^\top Q) \\ &= \Lambda. \end{aligned}$$

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q\Lambda Q^\top$ is a diagonalization of $K_{\mathbf{X}}$. Then $K_{\mathbf{Y}} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\operatorname{var} Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

1. By Proposition 10.5,

$$\begin{aligned} K_{\mathbf{Y}} &= Q^\top K_{\mathbf{X}} Q \\ &= Q^\top (Q\Lambda Q^\top) Q \\ &= (Q^\top Q)\Lambda(Q^\top Q) \\ &= \Lambda. \end{aligned}$$

2. Since $K_{\mathbf{Y}} = \Lambda$ is a diagonal matrix, the random variables in $\mathbf{Y}$ are uncorrelated because

$$\operatorname{cov}(Y_i, Y_j) = 0$$

for $i \neq j$.

Lemma 10.6 Let $\mathbf{X}$ and $\mathbf{Y}$ be column vectors of n random variables such that

$$\mathbf{Y} = Q^\top \mathbf{X},$$

where $Q\Lambda Q^\top$ is a diagonalization of $K_\mathbf{X}$. Then $K_\mathbf{Y} = \Lambda$, i.e., the random variables in $\mathbf{Y}$ are uncorrelated and $\text{var}\, Y_i = \lambda_i$, the ith diagonal element of Λ.

Proof

1. By Proposition 10.5,

$$\begin{aligned} K_\mathbf{Y} &= Q^\top K_\mathbf{X} Q \\ &= Q^\top (Q\Lambda Q^\top) Q \\ &= (Q^\top Q)\Lambda(Q^\top Q) \\ &= \Lambda. \end{aligned}$$

2. Since $K_\mathbf{Y} = \Lambda$ is a diagonal matrix, the random variables in $\mathbf{Y}$ are uncorrelated because

$$\text{cov}(Y_i, Y_j) = 0$$

for $i \neq j$.

3. Furthermore, the variance of Y_i is given by the ith diagonal element of $K_\mathbf{Y} = \Lambda$, i.e., λ_i. The proposition is proved.

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

Proposition 10.6 (Decorrelation) Let $\mathbf{Y} = Q^{\top}\mathbf{X}$, where $K_{\mathbf{X}} = Q\Lambda Q^{\top}$. Then $K_{\mathbf{Y}} = \Lambda$, i.e.,

1. the random variables in $\mathbf{Y}$ are uncorrelated
2. $\operatorname{var} Y_i = \lambda_i$ for all i

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^\top$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^\top.$$

Proposition 10.6 (Decorrelation) Let $\mathbf{Y} = Q^\top\mathbf{X}$, where $K_{\mathbf{X}} = Q\Lambda Q^\top$. Then $K_{\mathbf{Y}} = \Lambda$, i.e.,

1. the random variables in $\mathbf{Y}$ are uncorrelated

2. $\operatorname{var} Y_i = \lambda_i$ for all i

Corollary 10.7 Any random vector $\mathbf{X}$ can be written as a linear transformation of an uncorrelated vector. Specifically, $\mathbf{X} = Q\mathbf{Y}$, where $K_{\mathbf{X}} = Q\Lambda Q^\top$.

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

Proposition 10.6 (Decorrelation) Let $\mathbf{Y} = Q^{\top}\mathbf{X}$, where $K_{\mathbf{X}} = Q\Lambda Q^{\top}$. Then $K_{\mathbf{Y}} = \Lambda$, i.e.,

1. the random variables in $\mathbf{Y}$ are uncorrelated
2. $\operatorname{var} Y_i = \lambda_i$ for all i

Corollary 10.7 Any random vector $\mathbf{X}$ can be written as a linear transformation of an uncorrelated vector. Specifically, $\mathbf{X} = Q\mathbf{Y}$, where $K_{\mathbf{X}} = Q\Lambda Q^{\top}$.

Proof In Proposition 10.6, $\mathbf{Y} = Q^{\top}\mathbf{X}$ implies $Q\mathbf{Y} = QQ^{\top}\mathbf{X}$, or $\mathbf{X} = Q\mathbf{Y}$.

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^{\top}$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^{\top}.$$

Proposition 10.6 (Decorrelation) Let $\mathbf{Y} = Q^{\top}\mathbf{X}$, where $K_{\mathbf{X}} = Q\Lambda Q^{\top}$. Then $K_{\mathbf{Y}} = \Lambda$, i.e.,

1. the random variables in $\mathbf{Y}$ are uncorrelated
2. $\operatorname{var} Y_i = \lambda_i$ for all i

Corollary 10.7 Any random vector $\mathbf{X}$ can be written as a linear transformation of an uncorrelated vector. Specifically, $\mathbf{X} = Q\mathbf{Y}$, where $K_{\mathbf{X}} = Q\Lambda Q^{\top}$.

Proof In Proposition 10.6, $\mathbf{Y} = Q^{\top}\mathbf{X}$ implies $Q\mathbf{Y} = QQ^{\top}\mathbf{X}$, or $\mathbf{X} = Q\mathbf{Y}$.

Proposition 10.5 Let $\mathbf{Y} = A\mathbf{X}$. Then

$$K_{\mathbf{Y}} = AK_{\mathbf{X}}A^\top$$

and

$$\tilde{K}_{\mathbf{Y}} = A\tilde{K}_{\mathbf{X}}A^\top.$$

Proposition 10.6 (Decorrelation) Let $\mathbf{Y} = Q^\top\mathbf{X}$, where $K_{\mathbf{X}} = Q\Lambda Q^\top$. Then $K_{\mathbf{Y}} = \Lambda$, i.e.,

1. the random variables in $\mathbf{Y}$ are uncorrelated
2. $\operatorname{var} Y_i = \lambda_i$ for all i

Corollary 10.7 Any random vector $\mathbf{X}$ can be written as a linear transformation of an uncorrelated vector. Specifically, $\mathbf{X} = Q\mathbf{Y}$, where $K_{\mathbf{X}} = Q\Lambda Q^\top$.

Proof In Proposition 10.6, $\mathbf{Y} = Q^\top\mathbf{X}$ implies $Q\mathbf{Y} = QQ^\top\mathbf{X}$, or $\mathbf{X} = Q\mathbf{Y}$.

Proposition 10.8 Let $\mathbf{X}$ and $\mathbf{Z}$ be independent and $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. Then

$$K_{\mathbf{Y}} = K_{\mathbf{X}} + K_{\mathbf{Z}}.$$

Proposition 10.8 Let $\mathbf{X}$ and $\mathbf{Z}$ be independent and $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. Then

$$K_{\mathbf{Y}} = K_{\mathbf{X}} + K_{\mathbf{Z}}.$$

Remarks

Proposition 10.8 Let $\mathbf{X}$ and $\mathbf{Z}$ be independent and $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. Then

$$K_{\mathbf{Y}} = K_{\mathbf{X}} + K_{\mathbf{Z}}.$$

Remarks

1. If $\mathbf{Y} = \sum_{i=1}^{n} \mathbf{X}_i$ where $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ are mutually independent, then

$$K_{\mathbf{Y}} = \sum_{i=1}^{n} K_{\mathbf{X}_i}.$$

Proposition 10.8 Let $\mathbf{X}$ and $\mathbf{Z}$ be independent and $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. Then

$$K_{\mathbf{Y}} = K_{\mathbf{X}} + K_{\mathbf{Z}}.$$

Remarks

1. If $\mathbf{Y} = \sum_{i=1}^{n} \mathbf{X}_i$ where $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ are mutually independent, then

$$K_{\mathbf{Y}} = \sum_{i=1}^{n} K_{\mathbf{X}_i}.$$

2. When $\mathbf{X}_i$ are scalars, this reduces to

$$\operatorname{var} Y = \sum_{i=1}^{n} \operatorname{var} X_i.$$

Proposition 10.9 (Preservation of Energy) Let $\mathbf{Y} = Q\mathbf{X}$, where Q is an orthogonal matrix. Then

$$E \sum_{i=1}^{n} Y_i^2 = E \sum_{i=1}^{n} X_i^2.$$

Proposition 10.9 Let $\mathbf{Y} = Q\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors of n random variables and Q is an orthogonal matrix. Then

$$E\sum_{i=1}^{n} Y_i^2 = E\sum_{i=1}^{n} X_i^2.$$

Proof

Proposition 10.9 Let $\mathbf{Y} = Q\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors of n random variables and Q is an orthogonal matrix. Then

$$E \sum_{i=1}^{n} Y_i^2 = E \sum_{i=1}^{n} X_i^2 .$$

Proof

1. Consider

$$\sum_{i=1}^{n} Y_i^2 \quad = \quad \mathbf{Y}^\top \mathbf{Y}$$

Proposition 10.9 Let $\mathbf{Y} = Q\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors of n random variables and Q is an orthogonal matrix. Then

$$E\sum_{i=1}^{n} Y_i^2 = E\sum_{i=1}^{n} X_i^2.$$

Proof

1. Consider

$$\sum_{i=1}^{n} Y_i^2 = \mathbf{Y}^\top \mathbf{Y}$$

Proposition 10.9 Let $\mathbf{Y} = Q\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors of n random variables and Q is an orthogonal matrix. Then

$$E\sum_{i=1}^{n} Y_i^2 = E\sum_{i=1}^{n} X_i^2.$$

Proof

1. Consider

$$\begin{aligned}\sum_{i=1}^{n} Y_i^2 &= \mathbf{Y}^\top \mathbf{Y} \\ &= (Q\mathbf{X})^\top (Q\mathbf{X})\end{aligned}$$

Proposition 10.9 Let $\mathbf{Y} = Q\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors of n random variables and Q is an orthogonal matrix. Then

$$E \sum_{i=1}^{n} Y_i^2 = E \sum_{i=1}^{n} X_i^2.$$

Proof

1. Consider

$$\begin{aligned}
\sum_{i=1}^{n} Y_i^2 &= \mathbf{Y}^\top \mathbf{Y} \\
&= (Q\mathbf{X})^\top (Q\mathbf{X}) \\
&= (\mathbf{X}^\top Q^\top)(Q\mathbf{X})
\end{aligned}$$

Proposition 10.9 Let $\mathbf{Y} = Q\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors of n random variables and Q is an orthogonal matrix. Then

$$E\sum_{i=1}^{n} Y_i^2 = E\sum_{i=1}^{n} X_i^2.$$

Proof

1. Consider

$$\begin{aligned}
\sum_{i=1}^{n} Y_i^2 &= \mathbf{Y}^\top \mathbf{Y} \\
&= (Q\mathbf{X})^\top (Q\mathbf{X}) \\
&= (\mathbf{X}^\top Q^\top)(Q\mathbf{X}) \\
&= \mathbf{X}^\top (Q^\top Q)\mathbf{X}
\end{aligned}$$

Proposition 10.9 Let $\mathbf{Y} = Q\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors of n random variables and Q is an orthogonal matrix. Then

$$E\sum_{i=1}^{n} Y_i^2 = E\sum_{i=1}^{n} X_i^2.$$

Proof

1. Consider

$$\begin{aligned}
\sum_{i=1}^{n} Y_i^2 &= \mathbf{Y}^\top \mathbf{Y} \\
&= (Q\mathbf{X})^\top (Q\mathbf{X}) \\
&= (\mathbf{X}^\top Q^\top)(Q\mathbf{X}) \\
&= \mathbf{X}^\top (Q^\top Q)\mathbf{X}
\end{aligned}$$

Proposition 10.9 Let $\mathbf{Y} = Q\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors of n random variables and Q is an orthogonal matrix. Then

$$E\sum_{i=1}^{n} Y_i^2 = E\sum_{i=1}^{n} X_i^2.$$

Proof

1. Consider

$$\begin{aligned}
\sum_{i=1}^{n} Y_i^2 &= \mathbf{Y}^\top \mathbf{Y} \\
&= (Q\mathbf{X})^\top (Q\mathbf{X}) \\
&= (\mathbf{X}^\top Q^\top)(Q\mathbf{X}) \\
&= \mathbf{X}^\top (Q^\top Q)\mathbf{X} \\
&= \mathbf{X}^\top \mathbf{X}
\end{aligned}$$

Proposition 10.9 Let $\mathbf{Y} = Q\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors of n random variables and Q is an orthogonal matrix. Then

$$E \sum_{i=1}^{n} Y_i^2 = E \sum_{i=1}^{n} X_i^2.$$

Proof

1. Consider

$$\begin{aligned}
\sum_{i=1}^{n} Y_i^2 &= \mathbf{Y}^\top \mathbf{Y} \\
&= (Q\mathbf{X})^\top (Q\mathbf{X}) \\
&= (\mathbf{X}^\top Q^\top)(Q\mathbf{X}) \\
&= \mathbf{X}^\top (Q^\top Q)\mathbf{X} \\
&= \mathbf{X}^\top \mathbf{X} \\
&= \sum_{i=1}^{n} X_i^2.
\end{aligned}$$

Proposition 10.9 Let $\mathbf{Y} = Q\mathbf{X}$, where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors of n random variables and Q is an orthogonal matrix. Then

$$E\sum_{i=1}^{n} Y_i^2 = E\sum_{i=1}^{n} X_i^2.$$

Proof

1. Consider

$$\begin{aligned}
\sum_{i=1}^{n} Y_i^2 &= \mathbf{Y}^\top \mathbf{Y} \\
&= (Q\mathbf{X})^\top (Q\mathbf{X}) \\
&= (\mathbf{X}^\top Q^\top)(Q\mathbf{X}) \\
&= \mathbf{X}^\top (Q^\top Q)\mathbf{X} \\
&= \mathbf{X}^\top \mathbf{X} \\
&= \sum_{i=1}^{n} X_i^2.
\end{aligned}$$

2. The proposition is proved upon taking expectation on both sides.