5. Random Vectors

- probabilities
- characteristic function
- cross correlation, cross covariance
- Gaussian random vectors
- functions of random vectors

Random vectors

we denote $\mathbf{X}$ a random vector

$\mathbf{X}$ is a function that maps each outcome ζ to a vector of real numbers

an n-dimensional random variable has n components:

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

also called a *multivariate* or *multiple* random variable

Probabilities

Joint CDF

$$F(\mathbf{x}) \triangleq F_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

Joint PMF

$$p(\mathbf{x}) \triangleq p_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

Joint PDF

$$f(\mathbf{x}) \triangleq f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F(\mathbf{x})$$

Marginal PMF

$$p_{X_j}(x_j) = P(X_j = x_j) = \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$$

Marginal PDF

$$f_{X_j}(x_j) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)\, dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_n$$

Conditional PDF: the PDF of X_n given $X_1, \ldots, X_{n-1}$ is

$$f(x_n | x_1, \ldots, x_{n-1}) = \frac{f_{\mathbf{X}}(x_1, \ldots, x_n)}{f_{X_1, \ldots, X_{n-1}}(x_1, \ldots, x_{n-1})}$$

Characteristic Function

the characteristic function of an n-dimensional RV is defined by

$$\begin{aligned}\Phi(\omega) = \Phi(\omega_1, \dots, \omega_n) &= \mathbf{E}[e^{\mathrm{i}(\omega_1 X_1 + \cdots + \omega_n X_n)}] \\ &= \int_{\mathbf{x}} e^{\mathrm{i}\omega^T \mathbf{x}} f(\mathbf{x}) \mathbf{dx}\end{aligned}$$

where

$$\omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$\Phi(\omega)$ is the n-dimensional Fourier transform of $f(\mathbf{x})$

Independence

the random variables $X_1, \ldots, X_n$ are **independent** if

the joint pdf (or pmf) is equal to the product of their marginal's

Discrete

$$p_{\mathbf{X}}(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$$

Continuous

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

we can specify an RV by the characteristic function in place of the pdf,

$X_1, \ldots, X_n$ are *independent* if

$$\Phi(\omega) = \Phi_1(\omega_1) \cdots \Phi_n(\omega_n)$$

Example: White noise signal in communication

the n samples $X_1, \ldots X_n$ of a noise signal have the joint pdf:

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \frac{e^{-(x_1^2 + \cdots + x_n^2)/2}}{(2\pi)^{n/2}} \quad \text{for all } x_1, \ldots, x_n$$

the joint pdf is the n-product of one-dimensional Gaussian pdf's

thus, $X_1, \ldots, X_n$ are independent Gaussian random variables

Expected Values

the expected value of a function

$$g(\mathbf{X}) = g(X_1, \ldots, X_n)$$

of a vector random variable $\mathbf{X}$ is defined by

$$\mathbf{E}[g(\mathbf{X})] = \int_{\mathbf{x}} g(\mathbf{x}) f(\mathbf{x}) \mathbf{dx} \qquad \text{Continuous}$$

$$\mathbf{E}[g(\mathbf{X})] = \sum_{\mathbf{x}} g(\mathbf{x}) p(\mathbf{x}) \qquad \text{Discrete}$$

Mean vector

$$\mu = \mathbf{E}[\mathbf{X}] = \mathbf{E}\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{E}[X_1] \\ \mathbf{E}[X_2] \\ \vdots \\ \mathbf{E}[X_n] \end{bmatrix}$$

Correlation and Covariance matrices

Correlation matrix has the second moments of $\mathbf{X}$ as its entries:

$$\mathbf{R} \triangleq \mathbf{E}[\mathbf{X}\mathbf{X}^T] = \begin{bmatrix} \mathbf{E}[X_1X_1] & \mathbf{E}[X_1X_2] & \cdots & \mathbf{E}[X_1X_n] \\ \mathbf{E}[X_2X_1] & \mathbf{E}[X_2X_2] & \cdots & \mathbf{E}[X_2X_n] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}[X_nX_1] & \mathbf{E}[X_nX_2] & \cdots & \mathbf{E}[X_nX_n] \end{bmatrix}$$

with

$$\mathbf{R}_{ij} = \mathbf{E}[X_iX_j]$$

Covariance matrix has the second-order central moments as its entries:

$$\mathbf{C} \triangleq \mathbf{E}[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^T]$$

with

$$\mathbf{C}_{ij} = \mathbf{cov}(X_i, X_j) = \mathbf{E}[(X_i-\mu_i)(X_j-\mu_j)]$$

Symmetric matrix

$A \in \mathbf{R}^{n \times n}$ is called *symmetric* if $A = A^T$

Facts: if A is symmetric

- all eigenvalues of A are real
- all eigenvectors of A are orthogonal
- A admits a decomposition

$$A = UDU^T$$

  where $U^TU = UU^T = I$ (U is unitary) and D is diagonal

(of course, the diagonals of D are eigenvalues of A)

Unitary matrix

a matrix $U \in \mathbf{R}^{n\times n}$ is called **unitary** if

$$U^TU = UU^T = I$$

example: $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Facts:

- a real unitary matrix is also called **orthogonal**
- a unitary matrix is always invertible and $U^{-1} = U^T$
- columns vectors of U are mutually orthogonal
- norm is preserved under a unitary transformation:

$$y = Ux \quad \Longrightarrow \quad \|y\| = \|x\|$$

Positive definite matrix

a symmetric matrix A is **positive semidefinite**, written as $A \succeq 0$ if

$$x^T A x \geq 0, \quad \forall x \in \mathbf{R}^n$$

and **positive definite**, written as $A \succ 0$ if

$$x^T A x > 0, \quad \text{for all } \textit{nonzero } x \in \mathbf{R}^n$$

Facts: $A \succeq 0$ if and only if

- all eigenvalues of A are non-negative
- all principle minors of A are non-negative

example: $A = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \succeq 0$ because

$$\begin{aligned} x^T A x &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= x_1^2 + 2x_2^2 - 2x_1 x_2 \\ &= (x_1 - x_2)^2 + x_2^2 \geq 0 \end{aligned}$$

or we can check from

- eigenvalues of A are 0.38 and 2.61 (real and positive)
- the principle minors are 1 and $\begin{vmatrix} 1 & -1 \\ -1 & 2 \end{vmatrix} = 1$ (all positive)

note: $A \succeq 0$ does not mean all entries of A are positive!

Properties of correlation and covariance matrices

let $\mathbf{X}$ be a (real) n-dimensional random vector with mean μ

Facts:

- $\mathbf{R}$ and $\mathbf{C}$ are $n \times n$ symmetric matrices
- $\mathbf{R}$ and $\mathbf{C}$ are positive semidefinite
- If $X_1, \ldots, X_n$ are independent, then $\mathbf{C}$ is diagonal
- the diagonals of $\mathbf{C}$ are given by the variances of X_k
- if $\mathbf{X}$ has zero mean, then $\mathbf{R} = \mathbf{C}$
- $\mathbf{C} = \mathbf{R} - \mu\mu^T$

Cross Correlation and Cross Covariance

let $\mathbf{X}, \mathbf{Y}$ be vector random variables with means μ_X, μ_Y respectively

Cross Correlation

$$\mathbf{cor}(\mathbf{X}, \mathbf{Y}) = \mathbf{E}[\mathbf{X}\mathbf{Y}^T]$$

if $\mathbf{cor}(\mathbf{X}, \mathbf{Y}) = 0$ then $\mathbf{X}$ and $\mathbf{Y}$ are said to be **orthogonal**

Cross Covariance

$$\begin{aligned} \mathbf{cov}(\mathbf{X}, \mathbf{Y}) &= \mathbf{E}[(\mathbf{X} - \mu_X)(\mathbf{Y} - \mu_Y)^T] \\ &= \mathbf{cor}(\mathbf{X}, \mathbf{Y}) - \mu_X \mu_Y^T \end{aligned}$$

if $\mathbf{cov}(\mathbf{X}, \mathbf{Y}) = 0$ then $\mathbf{X}$ and $\mathbf{Y}$ are said to be **uncorrelated**

Affine transformation

let $\mathbf{Y}$ be an affine transformation of $\mathbf{X}$:

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

where $\mathbf{A}$ and $\mathbf{b}$ are deterministic matrices

- $\mu_Y = \mathbf{A}\mu_X + \mathbf{b}$

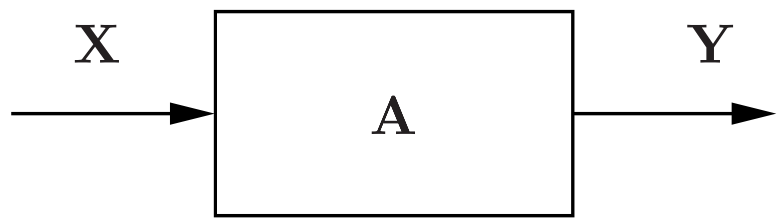
$$\mu_Y = \mathbf{E}[\mathbf{A}\mathbf{X} + \mathbf{b}] = \mathbf{A}\mathbf{E}[\mathbf{X}] + \mathbf{E}[\mathbf{b}] = \mathbf{A}\mu_X + \mathbf{b}$$

- $\mathbf{C}_Y = \mathbf{A}\mathbf{C}_X\mathbf{A}^T$

$$\begin{aligned}\mathbf{C}_Y &= \mathbf{E}[(\mathbf{Y} - \mu_Y)(\mathbf{Y} - \mu_Y)^T] = \mathbf{E}[(\mathbf{A}(\mathbf{X} - \mu_X))(\mathbf{A}(\mathbf{X} - \mu_X))^T] \\ &= \mathbf{A}\mathbf{E}[(\mathbf{X} - \mu_X)(\mathbf{X} - \mu_X)^T]\mathbf{A}^T = \mathbf{A}\mathbf{C}_X\mathbf{A}^T\end{aligned}$$

Diagonalization of covariance matrix

suppose a random vector $\mathbf{Y}$ is obtained via a linear transformation of $\mathbf{X}$



- the covariance matrices of $\mathbf{X}, \mathbf{Y}$ are $\mathbf{C}_X, \mathbf{C}_Y$ respectively
- $\mathbf{A}$ may represent linear filter, system gain, etc.
- the covariance of $\mathbf{Y}$ is $\mathbf{C}_Y = \mathbf{A}\mathbf{C}_X\mathbf{A}^T$

Problem: choose $\mathbf{A}$ such that $\mathbf{C}_Y$ becomes 'diagonal'

in other words, the variables $Y_1, \ldots, Y_n$ are required to be **uncorrelated** since $\mathbf{C}_X$ is symmetric, it has the decomposition:

$$\mathbf{C}_X = UDU^T$$

where

- D is diagonal and its entries are eigenvalues of $\mathbf{C}_X$
- U is unitary and the columns of U are eigenvectors of $\mathbf{C}_X$

diagonalization: pick $\mathbf{A} = U^T$ to obtain

$$\mathbf{C}_Y = \mathbf{A}\mathbf{C}_X\mathbf{A}^T = \mathbf{A}UDU^T\mathbf{A}^T = U^TUDU^TU = D$$

as desired

one can write $\mathbf{X}$ in terms of $\mathbf{Y}$ as

$$\mathbf{X} = UU^T\mathbf{X} = U\mathbf{Y} = \begin{bmatrix} U_1 & U_2 & \cdots & U_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{k=1}^{n} Y_k U_k$$

this equation is called **Karhunen-Loéve expansion**

- $\mathbf{X}$ can be expressed as a weighted sum of the eigenvectors U_k
- the weighting coefficients are *uncorrelated* random variables Y_k

example: $\mathbf{X}$ has the covariance matrix $\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$

design a transformation $\mathbf{Y} = \mathbf{A}\mathbf{X}$ s.t. the covariance of $\mathbf{Y}$ is diagonal

the eigenvalues of $\mathbf{C}_\mathbf{X}$ and the corresponding eigenvectors are

$$\lambda_1 = 6, \;\; u_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \lambda_2 = 2, \;\; u_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

u_1 and u_2 are orthogonal, so if we normalize u_k so that $\|u_k\| = 1$ then

$$U = \begin{bmatrix} \frac{u_1}{\sqrt{2}} & \frac{u_2}{\sqrt{2}} \end{bmatrix} \text{ is unitary}$$

therefore, $\mathbf{C}_\mathbf{X} = UDU^T$ where

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}, \quad D = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$$

thus, if we choose $\mathbf{A} = U^T$ then $\mathbf{C}_\mathbf{Y} = D$ which is diagonal as desired

Whitening transformation

we wish to find a transformation $\mathbf{Y} = \mathbf{A}\mathbf{X}$ such that

$$\mathbf{C}_{\mathbf{Y}} = I$$

- a white noise property: the covariance is the identity matrix
- all components in $\mathbf{Y}$ are all **uncorrelated**
- the variances of Y_k are **normalized** to 1

from $\mathbf{C}_{\mathbf{Y}} = \mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}^T$ and use the eigenvalue decomposition in $\mathbf{C}_{\mathbf{X}}$

$$\mathbf{C}_{\mathbf{Y}} = \mathbf{A}UDU^T\mathbf{A}^T = \mathbf{A}UD^{1/2}D^{1/2}U^T\mathbf{A}^T$$

denote $D^{1/2}$ the square root of D with $D \succeq 0$, *i.e.*, $D^{1/2}D^{1/2} = D$

$$D = \mathbf{diag}(d_1, \ldots, d_n) \quad \Longrightarrow \quad D^{1/2} = \mathbf{diag}(\sqrt{d_1}, \ldots, \sqrt{d_n})$$

can you find $\mathbf{A}$ that makes $\mathbf{C}_{\mathbf{Y}}$ the identity matrix ?

Gaussian random vector

$X_1, \ldots, X_n$ are said to be **jointly Gaussian** if their joint pdf is given by

$$f(\mathbf{x}) \triangleq f_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp -\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)$$

μ is the mean ($n \times 1$) and $\Sigma \succ 0$ is the covariance matrix ($n \times n$):

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1n} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{n1} & \Sigma_{n2} & \cdots & \Sigma_{nn} \end{bmatrix}$$

and

$$\mu_k = \mathbf{E}[X_k], \quad \Sigma_{ij} = \mathbf{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

example: the joint density function of $\mathbf{x}$ (not normalized) is given by

$$f(x_1, x_2, x_3) = \exp -\frac{x_1^2 + 3x_2^2 + 2(x_3 - 1)^2 + 2x_1(x_3 - 1)}{2}$$

- f is an exponential of *negative quadratic* in $\mathbf{x}$ so $\mathbf{x}$ must be a Gaussian

$$f(x_1, x_2, x_3) = \exp -\frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \\ x_3 - 1 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 - 1 \end{bmatrix}$$

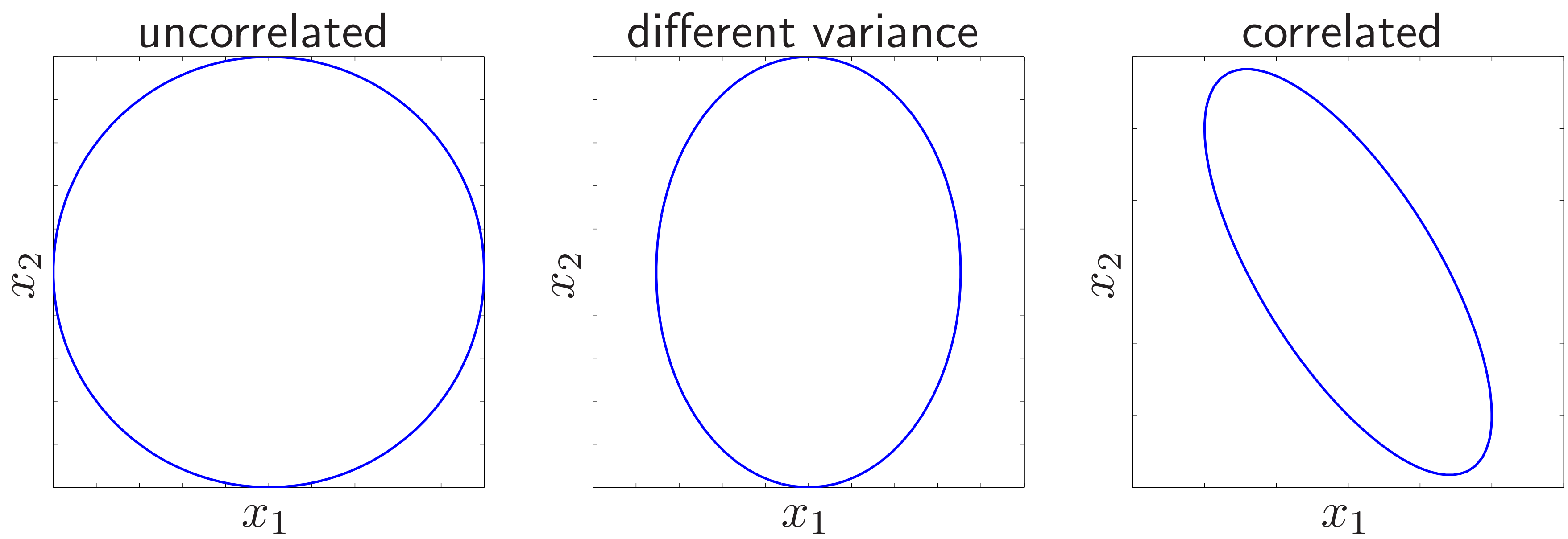
- the mean vector is $(0, 0, 1)$ and the covariance matrix is

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 1/3 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

- the variance of x_1 is highest while x_2 is smallest
- x_1 and x_2 are uncorrelated, so are x_2 and x_3

examples of Gaussian density contour (the exponent of exponential)

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1$$



$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

Properties of Gaussian variables

many results on Gaussian RVs can be obtained analytically:

- marginal's of $\mathbf{X}$ is also Gaussian
- conditional pdf of X_k given the other variables is a Gaussian distribution
- uncorrelated Gaussian random variables are *independent*
- any affine transformation of a Gaussian is also a Gaussian

these are well-known facts

and more can be found in the areas of estimation, statistical learning, etc.

Characteristic function of Gaussian

$$\Phi(\omega) = \Phi(\omega_1, \omega_2, \ldots, \omega_n) = e^{\mathrm{i}\mu^T \omega} \; e^{-\frac{\omega^T \Sigma \omega}{2}}$$

Proof. By definition and arranging the quadratic term in the power of exp

$$\begin{aligned}
\Phi(\omega) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int_{\mathbf{x}} e^{\mathrm{i}\mathbf{x}^T \omega} \; e^{-\frac{(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)}{2}} \mathbf{dx} \\
&= \frac{e^{\mathrm{i}\mu^T \omega} \; e^{-\frac{\omega^T \Sigma \omega}{2}}}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int_{\mathbf{x}} e^{-\frac{(\mathbf{x}-\mu-\mathrm{i}\Sigma\omega)^T \Sigma^{-1}(\mathbf{x}-\mu-\mathrm{i}\Sigma\omega)}{2}} \mathbf{dx} \\
&= \text{exp} \; (\mathrm{i}\mu^T \omega) \; \text{exp} \; (-\frac{1}{2}\omega^T \Sigma \omega)
\end{aligned}$$

(the integral equals 1 since it is a form of Gaussian distribution)

for one-dimensional Gaussian with zero mean and variance $\Sigma = \sigma^2$,

$$\Phi(\omega) = e^{-\frac{\sigma^2 \omega^2}{2}}$$

Affine Transformation of a Gaussian is Gaussian

let $\mathbf{X}$ be an n-dimensional Gaussian, $X \sim \mathcal{N}(\mu, \Sigma)$ and define

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

where $\mathbf{A}$ is $m \times n$ and $\mathbf{b}$ is $m \times 1$ (so $\mathbf{Y}$ is $m \times 1$)

$$\begin{aligned}
\Phi_{\mathbf{Y}}(\omega) &= \mathbf{E}[e^{\mathrm{i}\omega^T \mathbf{Y}}] = \mathbf{E}[e^{\mathrm{i}\omega^T (\mathbf{A}\mathbf{X}+\mathbf{b})}] \\
&= \mathbf{E}[e^{\mathrm{i}\omega^T \mathbf{A}\mathbf{X}} \cdot e^{\mathrm{i}\omega^T \mathbf{b}}] = e^{\mathrm{i}\omega^T \mathbf{b}} \Phi_{\mathbf{X}}(\mathbf{A}^T \omega) \\
&= e^{\mathrm{i}\omega^T \mathbf{b}} \cdot e^{\mathrm{i}\mu^T A^T \omega} \cdot e^{-\omega^T \mathbf{A}\Sigma\mathbf{A}^T \omega/2} \\
&= e^{\mathrm{i}\omega^T (\mathbf{A}\mu+\mathbf{b})} \cdot e^{-\omega^T A\Sigma\mathbf{A}^T \omega/2}
\end{aligned}$$

we read off that $\mathbf{Y}$ is Gaussian with mean $\mathbf{A}\mu + \mathbf{b}$ and covariance $\mathbf{A}\Sigma\mathbf{A}^T$

Marginal of Gaussian is Gaussian

the k^{th} component of $\mathbf{X}$ is obtained by

$$X_k = \begin{bmatrix} 0 & \cdots & 1 & 0 \end{bmatrix} \mathbf{X} \quad \triangleq \quad \mathbf{e}_k^T \mathbf{X}$$

($\mathbf{e}_k$ is a standard unit column vector; all entries are zero except the k^{th} position)

hence, X_k is simply a linear transformation (in fact, a projection) of $\mathbf{X}$

X_k is then a Gaussian with mean

$$\mathbf{e}_k^T \mu = \mu_k$$

and covariance

$$\mathbf{e}_k^T \, \Sigma \, \mathbf{e}_k = \Sigma_{kk}$$

Uncorrelated Gaussians are independent

suppose $(\mathbf{X}, \mathbf{Y})$ is a jointly Gaussian vector with

$$\text{mean } \mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \quad \text{and covariance} \quad \begin{bmatrix} \mathbf{C}_X & 0 \\ 0 & \mathbf{C}_Y \end{bmatrix}$$

in otherwords, X and Y are *uncorrelated* Gaussians:

$$\mathbf{cov}(X, Y) = \mathbf{E}[XY^T] - \mathbf{E}[X]\mathbf{E}[Y]^T = 0$$

the joint density can be written as

$$f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^n |\mathbf{C}_X|^{1/2} |\mathbf{C}_Y|^{1/2}} \text{ exp } -\frac{1}{2} \begin{bmatrix} \mathbf{x} - \mu_x \\ \mathbf{y} - \mu_y \end{bmatrix}^T \begin{bmatrix} \mathbf{C}_X^{-1} & 0 \\ 0 & \mathbf{C}_Y^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x} - \mu_x \\ \mathbf{y} - \mu_y \end{bmatrix}$$

$$= \frac{1}{(2\pi)^{n/2} |\mathbf{C}_X|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_x)^T \mathbf{C}_X^{-1} (\mathbf{x}-\mu_x)} \cdot \frac{1}{(2\pi)^{n/2} |\mathbf{C}_Y|^{1/2}} e^{-\frac{1}{2}(\mathbf{y}-\mu_y)^T \mathbf{C}_Y^{-1} (\mathbf{y}-\mu_y)}$$

proving the independence

we can also see from the characteristic function

$$
\begin{aligned}
\Phi(\omega_1, \omega_2) &= \mathbf{E}\left[\text{exp}\ \left(\mathrm{i}\begin{bmatrix}\omega_1\\ \omega_2\end{bmatrix}^T \begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix}\right)\right] \\
&= \text{exp}\ \left(\mathrm{i}\begin{bmatrix}\omega_1\\ \omega_2\end{bmatrix}^T \begin{bmatrix}\mu_y\\ \mu_y\end{bmatrix}\right) \cdot \text{exp}\ -\frac{1}{2}\begin{bmatrix}\omega_1\\ \omega_2\end{bmatrix}^T \begin{bmatrix}\mathbf{C}_X & 0\\ 0 & \mathbf{C}_Y\end{bmatrix}\begin{bmatrix}\omega_1\\ \omega_2\end{bmatrix} \\
&= \text{exp}\ (\mathrm{i}\omega_1^T\mu_x)\ \text{exp}\ -\frac{\omega_1^T\mathbf{C}_{\mathbf{X}}\omega_1}{2} \cdot \text{exp}\ (\mathrm{i}\omega_2^T\mu_y) \cdot \text{exp}\ -\frac{\omega_2^T\mathbf{C}_{\mathbf{Y}}\omega_2}{2} \\
&\triangleq \Phi_1(\omega_1)\cdot\Phi_2(\omega_2)
\end{aligned}
$$

proving the independence

Conditional of Gaussian is Gaussian

let $\mathbf{Z}$ be an n-dimensional Gaussian which can be decomposed as

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{bmatrix} \right)$$

the conditional pdf of $\mathbf{X}$ given $\mathbf{Y}$ is also Gaussian with conditional mean

$$\mu_{\mathbf{X}|\mathbf{Y}} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{Y} - \mu_y)$$

and conditional covariance

$$\Sigma_{\mathbf{X}|\mathbf{Y}} = \Sigma_x - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T$$

Proof:

from the **matrix inversion lemma**, Σ^{-1} can be written as

$$\Sigma^{-1} = \begin{bmatrix} S^{-1} & -S^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} \\ -\Sigma_{yy}^{-1}\Sigma_{xy}^T S^{-1} & \Sigma_{yy}^{-1} + \Sigma_{yy}^{-1}\Sigma_{xy}^T S^{-1}\Sigma_{xy}\Sigma_{yy}^{-1} \end{bmatrix}$$

where S is called the **Schur complement of** Σ_{xx} **in** Σ and

$$\begin{aligned} S &= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T \\ \det\Sigma &= \det S \cdot \det\Sigma_{yy} \end{aligned}$$

we can show that $\Sigma \succ 0$ if any only if $S \succ 0$ and $\Sigma_{yy} \succ 0$

from $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}, \mathbf{y})/f_{\mathbf{Y}}(\mathbf{y})$, we calculate the exponent terms

$$\begin{bmatrix} \mathbf{x} - \mu_x \\ \mathbf{y} - \mu_y \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} \mathbf{x} - \mu_x \\ \mathbf{y} - \mu_y \end{bmatrix} - (\mathbf{y} - \mu_y)^T \Sigma_{yy}^{-1} (\mathbf{y} - \mu_y)$$

$$\begin{aligned}
&= (\mathbf{x} - \mu_x)^T S^{-1} (\mathbf{x} - \mu_x) - (\mathbf{x} - \mu_x)^T S^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} (\mathbf{y} - \mu_y) \\
&\quad - (\mathbf{y} - \mu_y)^T \Sigma_{yy}^{-1} \Sigma_{xy}^T S^{-1} (\mathbf{x} - \mu_x) \\
&\quad + (\mathbf{y} - \mu_y)^T (\Sigma_{yy}^{-1} \Sigma_{xy}^T S^{-1} \Sigma_{xy} \Sigma_{yy}^{-1}) (\mathbf{y} - \mu_y) \\
&= [\mathbf{x} - \mu_x - \Sigma_{xy} \Sigma_{yy}^{-1} (\mathbf{y} - \mu_y)]^T S^{-1} [\mathbf{x} - \mu_x - \Sigma_{xy} \Sigma_{yy}^{-1} (\mathbf{y} - \mu_y)] \\
&\triangleq (\mathbf{x} - \mu_{\mathbf{X}|\mathbf{Y}})^T \Sigma_{\mathbf{X}|\mathbf{Y}}^{-1} (\mathbf{x} - \mu_{\mathbf{X}|\mathbf{Y}})
\end{aligned}$$

$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$ is an exponential of quadratic function in $\mathbf{x}$

so it has a form of Gaussian

Standard Gaussian vectors

for an n-dimensional Gaussian vector $X \sim \mathcal{N}(\mu, \mathbf{C})$ with $\mathbf{C} \succ 0$

let $\mathbf{A}$ be an $n \times n$ invertible matrix such that

$$\mathbf{A}\mathbf{A}^T = \mathbf{C}$$

($\mathbf{A}$ is called a **factor** of $\mathbf{C}$)

then the random vector

$$\mathbf{Z} = \mathbf{A}^{-1}(\mathbf{X} - \mu)$$

is a standard Gaussian vector, $i.e.$,

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

(obtain $\mathbf{A}$ via eigenvalue decomposition or Cholesky factorization)

Functions of random vectors

- minimum and maximum of random variables
- general transformation
- affine transformation

Minimum and Maximum of RVs

let $X_1, X_2, \ldots, X_n$ be independent RVs

define the minimum and maximum of RVs by

$$Y = \min(X_1, X_2, \ldots, X_n), \quad Z = \max(X_1, X_2, \ldots, X_n)$$

the maximum of $X_1, X_2, \ldots, X_n$ is less than z iff $X_i \leq z$ for all i, so

$$F_Z(z) = P(X_1 \leq z)P(X_2 \leq z)\cdots P(X_n \leq z) = (F_X(z))^n$$

the minimum of $X_1, X_2, \ldots, X_n$ is greater than y iff $X_i \geq y$ for all i, so

$$1 - F_Y(y) = P(X_1 > y)P(X_2 > y)\cdots P(X_n > y) = (1 - F_X(y))^n$$

and

$$F_Y(y) = 1 - (1 - F_X(y))^n$$

Maximum of uniform samples

let $X_1, X_2, \ldots, X_n$ are i.i.d. samples of $\mathcal{U}(-a, a)$

$$Z = \max(X_1, X_2, \ldots, X_n)$$

find the pdf of X

$$F_Z(z) = (F_X(z))^n = \left(\frac{z+a}{2a}\right)^n, \quad -a \leq z \leq a$$

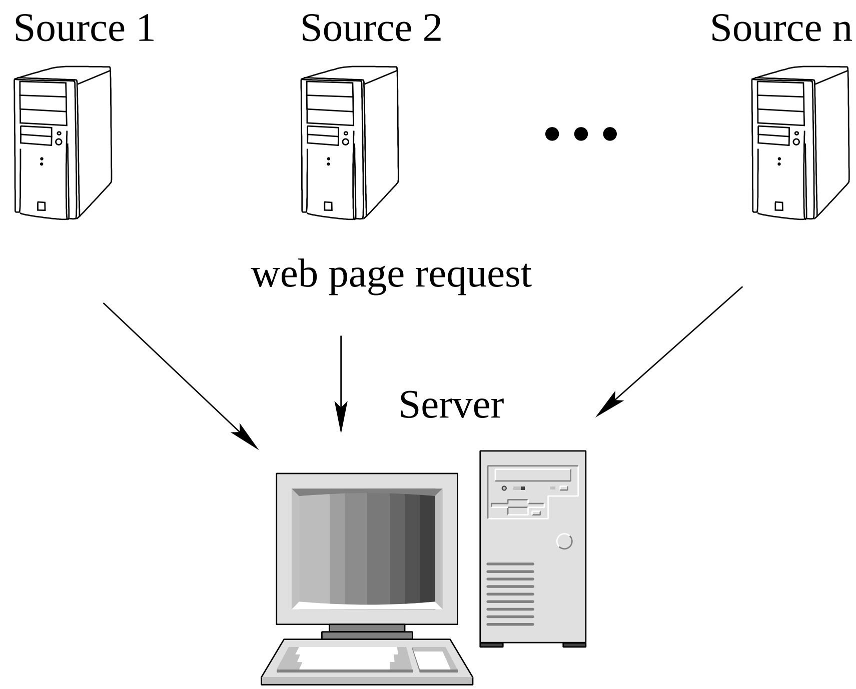
where $F_Z(z) = 0$ if $z \leq -a$ and $F_Z(z) = 1$ when $z \geq a$

the pdf can be obtained by

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{n(z+a)^{n-1}}{(2a)^n}, \quad -a \leq z \leq a$$

note: for $X_i \sim \mathcal{U}(a, b)$ with $a = 0, b = 1$, the pdf of Z is Beta$(n, 1)$

Example: Merging of independent Poisson arrivals



- T_i denotes the interarrival times for source i
- T_i has exponential distribution with rate λ_i
- find the distribution of the interarrival times between consecutive requests at server

each T_i satisfies the memoryless property, so the time that has elapsed since the last arrival is irrelevant

the time until the next arrival at the multiplexer is

$$Z = \min(T_1, T_2, \dots, T_n)$$

therefore, the cdf of Z can be computed by:

$$\begin{aligned}
1 - F_Z(z) &= P(\min(T_1, T_2, \dots, T_n) > z) \\
&= P(T_1 > z)P(T_2 > z)\cdots P(T_n > z) \\
&= (1 - F_{T_1}(z))(1 - F_{T_2}(z))\cdots(1 - F_{T_n}(z)) \\
&= e^{-\lambda_1 z}\cdots e^{-\lambda_n z} = e^{-(\lambda_1+\lambda_2+\cdots+\lambda_n)z}
\end{aligned}$$

the interarrival time is an exponential RV with rate $\lambda_1 + \lambda_2 + \cdots + \lambda_n$

General transformation

let $\mathbf{X}$ be a vector random variable

define $\mathbf{Z} = g(\mathbf{X}) : \mathbb{R}^n \to \mathbb{R}^n$ and assume that g is invertible

so that for $\mathbf{Z} = \mathbf{z}$ we can solve for $\mathbf{x}$ uniquely:

$$\mathbf{x} = g^{-1}(\mathbf{z})$$

then the joint pdf of $\mathbf{Z}$ is given by

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{f_{\mathbf{X}}(g^{-1}(\mathbf{z}))}{|\det J|}$$

where $\det J$ is the determinant of the Jacobian matrix:

$$J = \begin{bmatrix} \partial g_1/\partial x_1 & \cdots & \partial g_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial g_n/\partial x_1 & \cdots & \partial g_n/\partial x_n \end{bmatrix}$$

Affine transformation

if $\mathbf{X}$ is a continuous random vector and $\mathbf{A}$ is an invertible matrix

then $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ has pdf

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det(\mathbf{A})|} f_{\mathbf{X}}(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}))$$

Gaussian case: let $\mathbf{X} \sim \mathcal{N}(0, \Sigma)$

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{(2\pi)^{n/2}|\det \mathbf{A}||\Sigma|^{1/2}} \text{exp } -\frac{1}{2}(\mathbf{y} - \mathbf{b})^T \mathbf{A}^{-T} \Sigma^{-1} \mathbf{A}^{-1} (\mathbf{y} - \mathbf{b}) \\ &= \frac{1}{(2\pi)^{n/2}|\mathbf{A}\Sigma\mathbf{A}^T|^{1/2}} \text{exp } -\frac{1}{2}(\mathbf{y} - \mathbf{b})^T (\mathbf{A}\Sigma\mathbf{A}^T)^{-1} (\mathbf{y} - \mathbf{b}) \end{aligned}$$

we read off that $\mathbf{Y}$ is also Gaussian with mean $\mathbf{b}$ and covariance $\mathbf{A}\Sigma\mathbf{A}^T$

this agrees with the result in page 5-16 and 5-27

Example: Sum of jointly Gaussian

a special case of linear transformation is

$$Z = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

where $X_1, \ldots, X_n$ are jointly Gaussian

Z can be written as

$$Z = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \triangleq \mathbf{AX}$$

Z is simply a linear transformation of a Gaussian therefore, Z is Gaussian with mean

$$\mathbf{E}[Z] = \mathbf{A}\mu = \sum_{i=1} a_i \mathbf{E}[X_i]$$

and variance

$$\mathbf{var}(Z) = \mathbf{cov}(Z) = \mathbf{A}\Sigma\mathbf{A}^T = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \,\mathbf{cov}(X_i, X_j)$$

if $X_1, \ldots, X_n$ are *independent* Gaussian, i.e.,

$$\mathbf{cov}(X_i, X_j) = 0$$

then the variance of Z is reduced to

$$\mathbf{var}(Z) = \sum_{i=1}^{n} a_i^2 \,\mathbf{cov}(X_i, X_i) = \sum_{i=1}^{n} a_i^2 \,\mathbf{var}(X_i)$$

References

Chapter 6 in
A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd edition, Pearson Prentice Hall, 2009