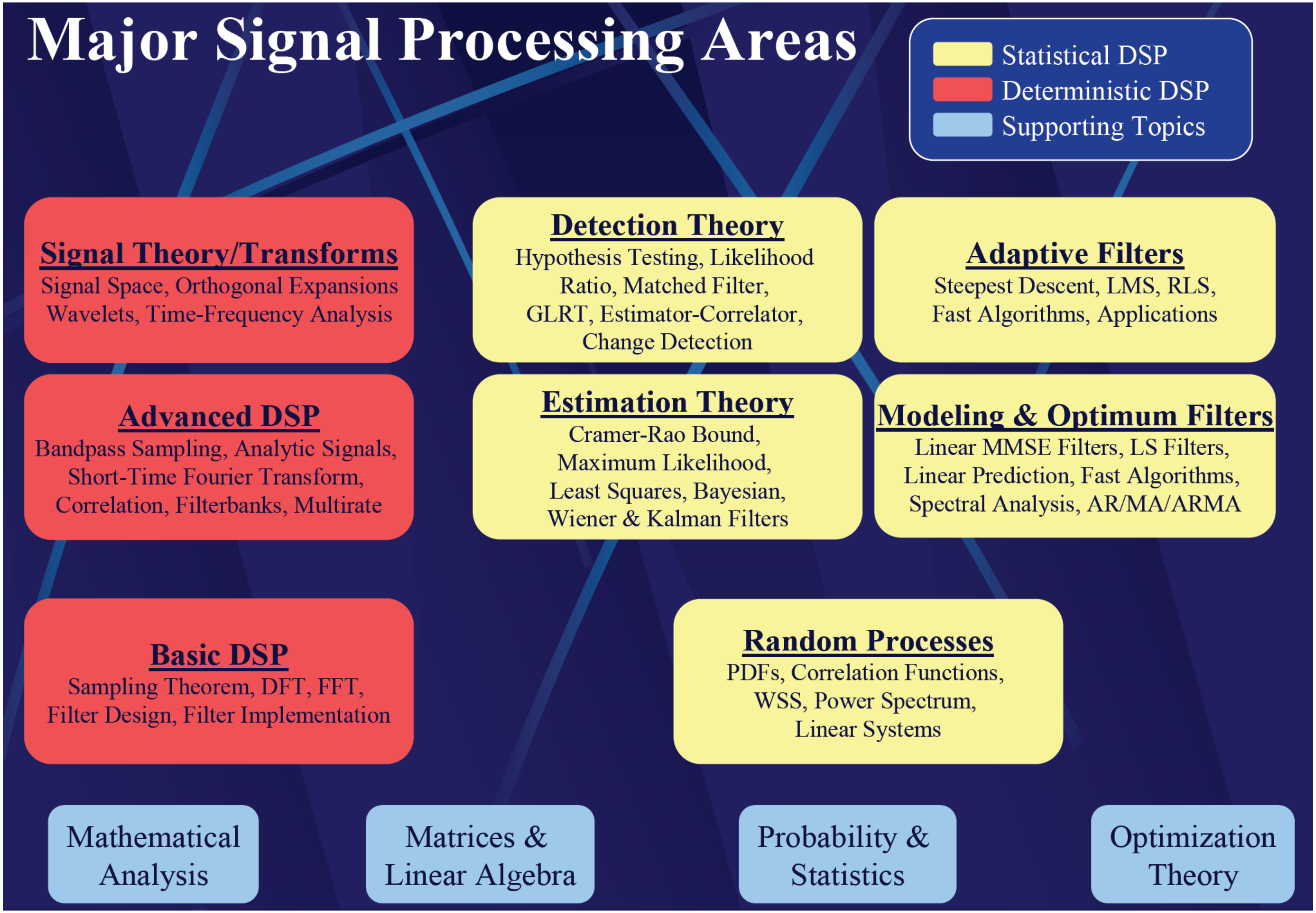
Stochastic Signal Processing

Dr.-Ing. Tuan Do-Hong

Department of Telecommunications Engineering
Faculty of Electrical and Electronics Engineering
HoChiMinh City University of Technology

E-mail: do-hong@hcmut.edu.vn


Major Signal Processing Areas
Statistical DSP
Deterministic DSP
Supporting Topics
Signal Theory/Transforms
Signal Space, Orthogonal Expansions
Wavelets, Time-Frequency Analysis
Detection Theory
Hypothesis Testing, Likelihood Ratio, Matched Filter, GLRT, Estimator-Correlator, Change Detection
Adaptive Filters
Steepest Descent, LMS, RLS, Fast Algorithms, Applications
Advanced DSP
Bandpass Sampling, Analytic Signals, Short-Time Fourier Transform, Correlation, Filterbanks, Multirate
Estimation Theory
Cramer-Rao Bound, Maximum Likelihood, Least Squares, Bayesian, Wiener & Kalman Filters
Modeling & Optimum Filters
Linear MMSE Filters, LS Filters, Linear Prediction, Fast Algorithms, Spectral Analysis, AR/MA/ARMA
Basic DSP
Sampling Theorem, DFT, FFT, Filter Design, Filter Implementation
Random Processes
PDFs, Correlation Functions, WSS, Power Spectrum, Linear Systems
Mathematical Analysis
Matrices & Linear Algebra
Probability & Statistics
Optimization Theory

Outline (1)

Chapter 1: Discrete-Time Signal Processing

z-Transform.

Linear Time-Invariant Filters.

Discrete Fourier Transform (DFT).

Chapter 2: Stochastic Processes and Models

Review of Probability and Random Variables.

Stochastic Models.

Stochastic Processes.

Chapter 3: Spectrum Analysis

Spectral Density.

Spectral Representation of Stochastic Process.

Outline (2)

Chapter 4: Eigenanalysis

Properties of Eigenvalues and Eigenvectors.

Chapter 5: Wiener Filters

Minimum Mean-Squared Error.

Wiener-Hopf Equations.

Linearly Constrained Minimum Variance (LCMV) Filter.

Examples: Fixed Weight Beamforming

Chapter 6: Linear Prediction

Forward Linear Prediction.

Backward Linear Prediction.

Levinson-Durbin Algorithm.

Chapter 7: Kalman Filters

Kalman Algorithm.

Applications of Kalman Filter: Tracking Trajectory of Object and System Identification.

Outline (3)

Chapter 8: Linear Adaptive Filtering

Adaptive Filters and Applications.
Method of Steepest Descent.
Least-Mean-Square Algorithm.
Examples of Adaptive Filtering: Adaptive Equalization and Adaptive Beamforming.

Chapter 9: Estimation Theory

Fundamentals.
Minimum Variance Unbiased Estimators.
Maximum Likelihood Estimation.
Spectral Estimation.

References

[1] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996 (3rd Ed.), 2002 (4th Ed.), 2014 (5th Ed.).

[2] Steven M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.

[3] Alan V. Oppenheim, Ronald W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, 1989.

[4] Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1991 (3rd Ed.), 2001 (4th Ed.).

[5] Dimitris G. Manolakis, Vinay K. Ingle, Stephen M. Kogon, *Statistical and Adaptive Signal Processing*, Artech House, 2005.

Goal of the Course

Introduction to the **theory and algorithms** used for the **analysis** and **processing** of **random signals** (**stochastic signals**) and their **applications to communications problems.**

- <u>Understanding the random signals</u>: via statistical description (theory of probability, random variables and stochastic processes), modeling and the dependence between the samples of one or more discrete-time random signals.

- <u>Developing the theoretically methods/practical techniques</u> for processing random signals to achieve a predefined application-dependent objective.
  - Major applications in communications: signal modeling, spectral estimation, frequency-selective filtering, adaptive filtering, array processing...

Optimum Receiver Design (1)

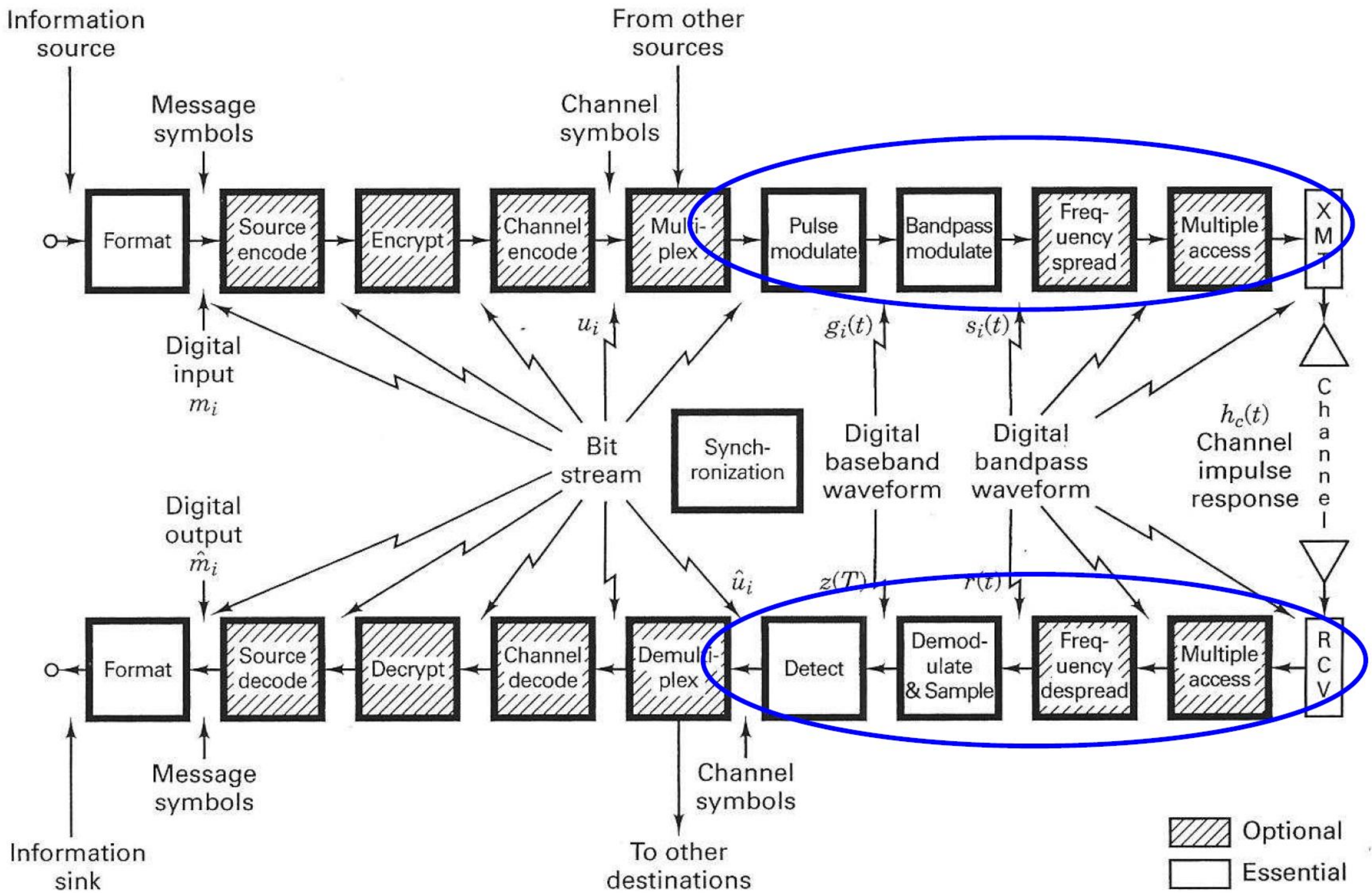
- Communication signal processing is applied in the design and optimization of
  - **transmitter**
  - **receiver**

  baseband parts with respect to user needs and **communication channel**, which includes
  - transmitter and receiver radio frequency (RF) parts
  - antennas
  - the radio channel itself time-varying.

  Therefore, transceivers must be **adaptive**.

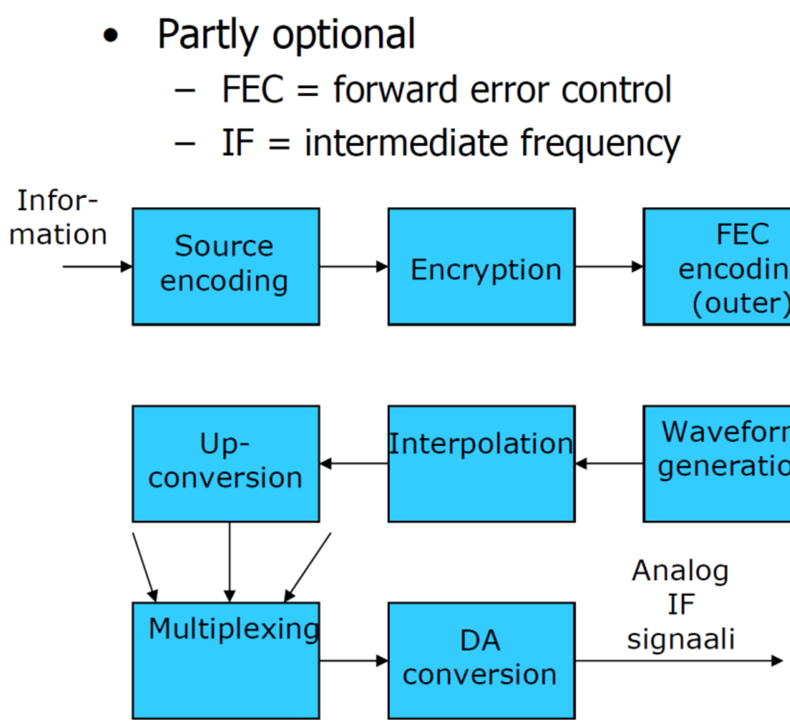
Optimum Receiver Design (2)

- Block diagram of a typical **digital communication system**:



Optimum Receiver Design (3)

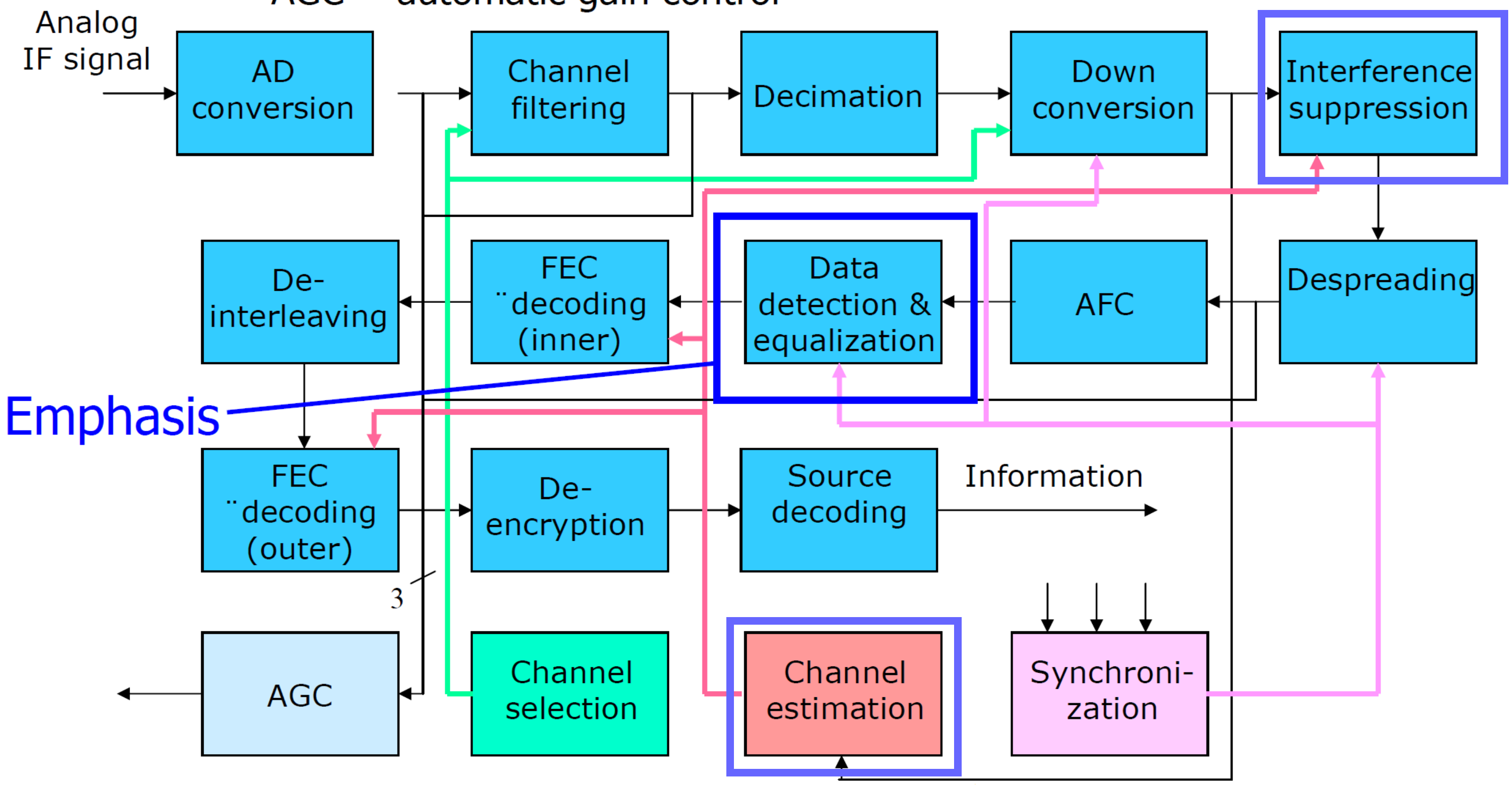
- A typical **transmitter baseband processing chain** example:



Optimum Receiver Design (4)

- A typical **receiver baseband processing chain** example:

  – AFC = automatic frequency correction
  – AGC = automatic gain control



Optimum Receiver Design (5)

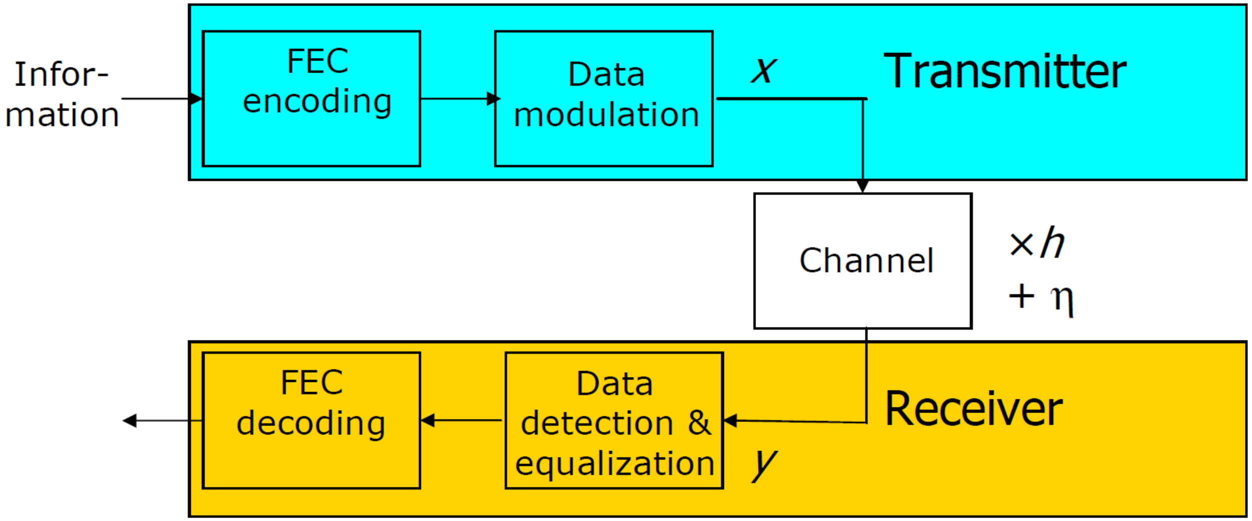
- The communication receiver design problem can be formulated as an **estimation** or **detection** (**classification**) **problem**.

  $\rightarrow$ **stochastic signal processing** used in **detector** and **equalizer optimization**.

- We formulate a typical optimization framework herein that is applicable for many scenarios.

Optimum Receiver Design (6)

- A simplified transmission system model:



- A fully discrete-time (all-digital) baseband model based on sampling and perfect synchronization.
- Two channel models used as an illustrative example:
  - frequency-selective multi-path channel.
  - multiple-input multiple-output (MIMO) multi-antenna channel.

Optimum Receiver Design (8)

- **Mathematical system model for modulator**:

  Assume a linear modulation method, e.g.,

  – phase shift keying (PSK)

  – quadrature amplitude modulation (QAM).

  Let the **transmitted data symbol** be $x \in \Omega$ where Ω is the modulations symbol alphabet.

  – Example, QPSK: $\Omega = \{-1,+1,-j,+j\}/\sqrt{2}$.

Optimum Receiver Design (9)

- **Mathematical system model for channel**:

  Assume that the channel:

  - causes a complex amplitude scaling hx, where $h = |h|e^{j\theta}$ is a complex gain factor with
    - – amplitude (absolute value) $|h|$
    - – phase shift θ
  - adds an additive noise term η.

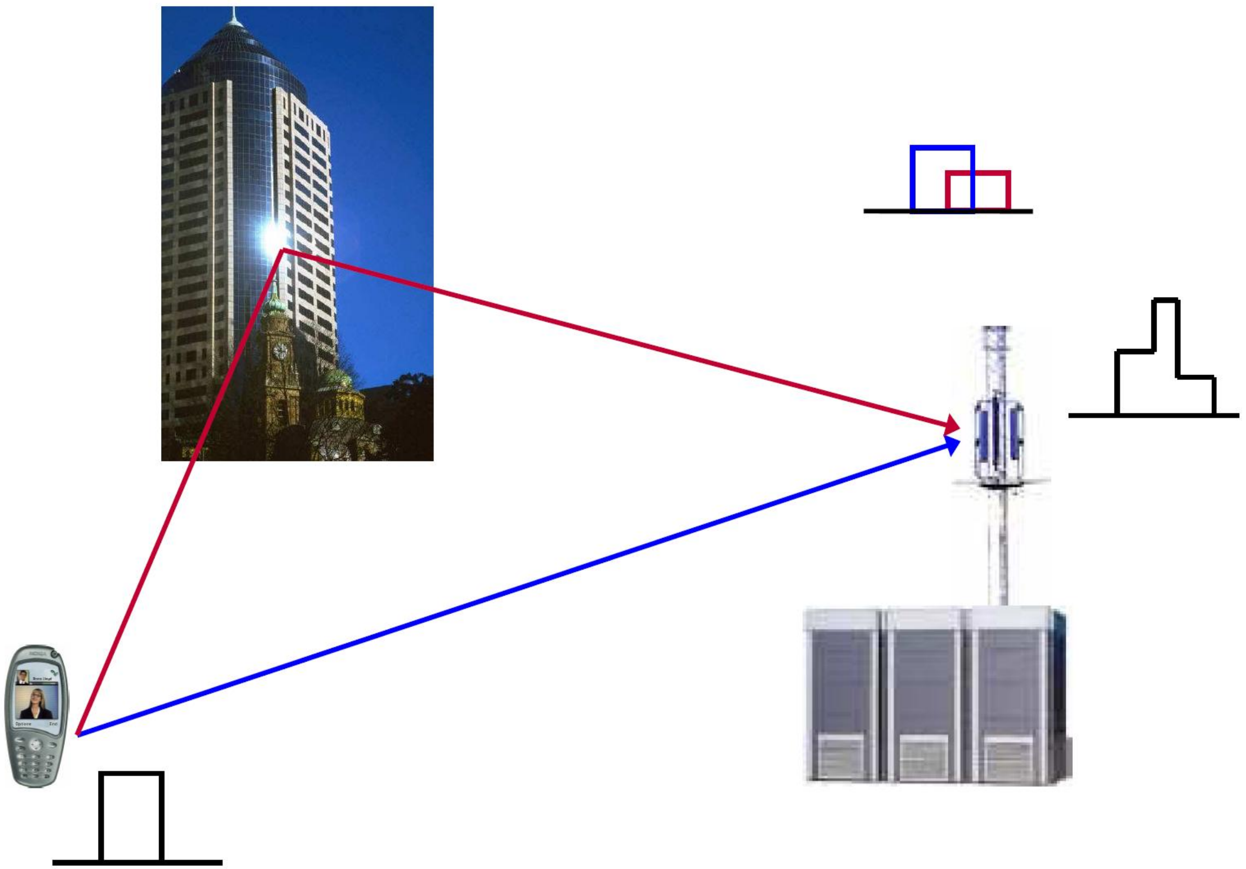
  Therefore, the **received signal sample** for discrete time interval n is:

$$y(n) = h(n)x(n) + \eta(n).$$

  It allows a time-varying channel gain $h(n)$.

Optimum Receiver Design (10)

- **Example: Frequency-selective channel**



Optimum Receiver Design (11)

Frequency-selective channel: **Mathematical model**

If the channel is frequency-selective $\rightarrow$ intersymbol interference (ISI), since multiple consequent transmitted symbols will overlap. Therefore, the received signal sample:

$$y(n)=\sum_{l=0}^{L-1} h(n;l)\,x(n-l)+\eta(n)=\mathbf{h}^{T}(n)\,\mathbf{x}(n)+\eta(n)$$

where L is the length of the channel impulse response and

$$\mathbf{x}(n)=\left[x(n-L+1)\quad x(n-L+2)\quad \cdots \quad x(n-1)\quad x(n)\right]^{T}\in C^{L}$$

$$\mathbf{h}(n)=\left[h(n;L-1)\quad h(n;L-2)\quad \cdots \quad h(n;1)\quad h(n;0)\right]^{T}\in C^{L}$$

Optimum Receiver Design (12)

Frequency-selective channel: Block model

Processing a block of N_B consequent symbols, then the received signal sample vector is

$$\mathbf{y} = \mathbf{H}^T\mathbf{x} + \boldsymbol{\eta}$$

where $\mathbf{x} = \begin{bmatrix} x(0) & x(1) & \cdots & x(N_B - 1) \end{bmatrix}^T \in C^{N_B}$ is the vector of transmitted symbols, and the vector of received symbol-spaced samples is

$$\mathbf{y} = \begin{bmatrix} y(0) & y(1) & \cdots & y(N_B - 1)\, y(N_B) & \cdots & y(N_B + L - 1) \end{bmatrix}^T \in C^{N_B + L - 1}$$

The channel matrix $\mathbf{H}$ can be decomposed as

$\mathbf{H} = [\mathbf{H}_{PRE}\ \mathbf{H}_{SS}\ \mathbf{H}_{POST}]$, where
$\mathbf{H}_{PRE}$, $\mathbf{H}_{POST}$ denote the transient parts and
$\mathbf{H}_{SS}$ denotes the steady-state part.

Optimum Receiver Design (13)

Frequency-selective channel: Pre-cursor transient channel matrix:

$$\mathbf{H}_{\text{PRE}} = \begin{bmatrix} h(0;0) & h(1;1) & \cdots & h(L-3;L-3) & h(L-2;L-2) \\ 0 & h(1;0) & \cdots & h(L-3;L-4) & h(L-2;L-3) \\ \vdots & 0 & \cdots & \vdots & \vdots \\ 0 & \vdots & & h(L-3;1) & h(L-2;2) \\ 0 & 0 & & h(L-3;0) & h(L-2;1) \\ 0 & 0 & \cdots & 0 & h(L-2;0) \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \in C^{N_{\text{B}} \times L-1}$$

Optimum Receiver Design (14)

Frequency-selective channel: Steady-state channel matrix:

$$\mathbf{H}_{\mathrm{SS}} = \begin{bmatrix} h(L-1;L-1) & 0 & \cdots & 0 & 0 \\ h(L-1;L-2) & h(L;L-1) & & \vdots & \vdots \\ \vdots & h(L;L-2) & \cdots & 0 & 0 \\ h(L-1;1) & \vdots & & h(N_{\mathrm{B}}-2;L-1) & 0 \\ h(L-1;0) & h(L;1) & \cdots & h(N_{\mathrm{B}}-2;L-2) & h(N_{\mathrm{B}}-1;L-1) \\ 0 & h(L;0) & & \vdots & h(N_{\mathrm{B}}-1;L-2) \\ 0 & 0 & \cdots & h(N_{\mathrm{B}}-2;1) & \vdots \\ \vdots & \vdots & & h(N_{\mathrm{B}}-2;0) & h(N_{\mathrm{B}}-1;1) \\ 0 & 0 & \cdots & 0 & h(N_{\mathrm{B}}-1;0) \end{bmatrix}$$

$$\in C^{N_{\mathrm{B}}\times(N_{\mathrm{B}}-L+1)}.$$

Optimum Receiver Design (15)

Frequency-selective channel: Post-cursor transient channel matrix:

$$\mathbf{H}_{\text{POST}} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & 0 & \cdots & \vdots & \vdots \\ 0 & \vdots & & 0 & 0 \\ h(N_{\text{B}};L-1) & 0 & & 0 & 0 \\ h(N_{\text{B}};L-2) & h(N_{\text{B}}+1;L-1) & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ h(N_{\text{B}};2) & h(N_{\text{B}}+1+2;3) & & h(N_{\text{B}}+L-3;L-1) & 0 \\ h(N_{\text{B}};1) & h(N_{\text{B}}+1+2;2) & \cdots & h(N_{\text{B}}+L-3;L-2) & h(N_{\text{B}}+L-2;L-1) \end{bmatrix}$$

$$\in C^{N_{\text{B}} \times L-1}.$$

Optimum Receiver Design (16)

Example: Time-invariant channel with $N_B = 8$ and $L = 3$:

$$\mathbf{H}_{SS} = \begin{bmatrix} h(2) & 0 & 0 & 0 & 0 & 0 \\ h(1) & h(2) & 0 & 0 & 0 & 0 \\ h(0) & h(1) & h(2) & 0 & 0 & 0 \\ 0 & h(0) & h(1) & h(2) & 0 & 0 \\ 0 & 0 & h(0) & h(1) & h(2) & 0 \\ 0 & 0 & 0 & h(0) & h(1) & h(2) \\ 0 & 0 & 0 & 0 & h(0) & h(1) \\ 0 & 0 & 0 & 0 & 0 & h(0) \end{bmatrix} \in C^{8\times 6}$$

Optimum Receiver Design (17)

$$\mathbf{H}_{\mathrm{PRE}} = \begin{bmatrix} h(0) & h(1) \\ 0 & h(0) \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \in C^{8\times 2}, \mathbf{H}_{\mathrm{POST}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ h(2) & 0 \\ h(1) & h(2) \end{bmatrix} \in C^{8\times 2}$$

Optimum Receiver Design (18)

$$\mathbf{H} = \left[\begin{array}{cc|cccccc|cc}
h(0) & h(1) & h(2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & h(0) & h(1) & h(2) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & h(0) & h(1) & h(2) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & h(0) & h(1) & h(2) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & h(0) & h(1) & h(2) & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & h(0) & h(1) & h(2) & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & h(0) & h(1) & h(2) & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & h(0) & h(1) & h(2)
\end{array}\right] \in C^{8\times 10}$$

Optimum Receiver Design (19)

- **Example: Multi-antenna channel**

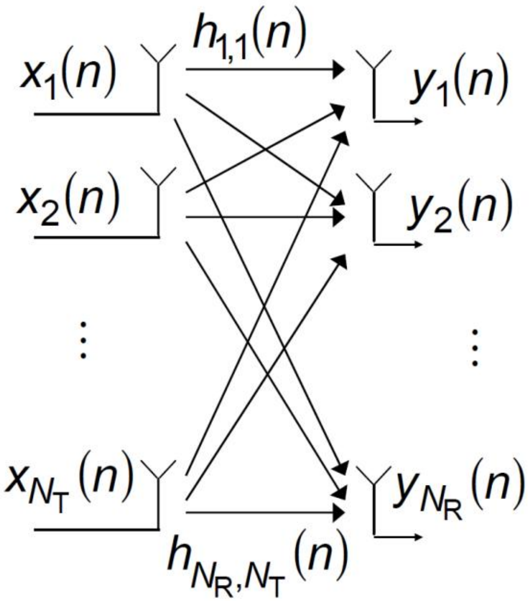
Signal $x_i(n)$ is transmitted at time interval n from antenna i ($i = 1, 2, \ldots, N_T$).

Signal $y_j(n)$ is received at time interval n at antenna j ($j = 1, 2, \ldots, N_R$):

$$y_j(n) = \sum_{i=1}^{N_T} h_{ij}(n) x_i(n) + \eta_j(n),$$

where $h_{ij}(n)$ is the complex channel gain with

$$\mathrm{E}\left(\left| h_{ij}(n) \right|^2 \right) = 1$$



MIMO channel model.

Optimum Receiver Design (20)

Matrix formulation of MIMO channel model:

The signal received at all antennas:

$$\mathbf{y}(n) = \mathbf{H}(n)\mathbf{x}(n) + \boldsymbol{\eta}(n),$$

where

$$\mathbf{x}(n) = \begin{bmatrix} x_1(n) & x_2(n) & \cdots & x_{N_\mathrm{T}}(n) \end{bmatrix}^\mathrm{T} \in C^{N_\mathrm{T}},$$

$$\mathbf{y}(n) = \begin{bmatrix} y_1(n) & y_2(n) & \cdots & y_{N_\mathrm{R}}(n) \end{bmatrix}^\mathrm{T} \in C^{N_\mathrm{R}},$$

$$\mathbf{H}(n) = \begin{bmatrix} h_{1,1}(n) & h_{1,2}(n) & \cdots & h_{1,N_\mathrm{T}}(n) \\ h_{2,1}(n) & h_{2,2}(n) & \cdots & h_{2,N_\mathrm{T}}(n) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_\mathrm{R},1}(n) & h_{N_\mathrm{R},2}(n) & \cdots & h_{N_\mathrm{R},N_\mathrm{T}}(n) \end{bmatrix} \in C^{N_\mathrm{R} \times N_\mathrm{T}}$$

Optimum Receiver Design (21)

- **Receiver design problem:**
  - The receiver must **detect/decode** the transmitted data.

    → The **most likely possible transmitted data symbol** should be found.
  - Due to outer FEC encoding the data bits and symbols are dependent.

    → Joint detection and decoding should is optimal.

    – Due to complexity reasons, detection and equalization is usually separated from decoding.

    – Soft output (log-likelihood ratio) information is provided to the decoder.

Optimum Receiver Design (22)

- **Received signal as random variable:**

The received signal vector $\mathbf{y} = \mathbf{Hx} + \boldsymbol{\eta}$ is a random.

– The noise term $\boldsymbol{\eta}$ is usually modeled as **zero mean complex Gaussian** with **covariance matrix** Σ_η often also **white**, i.e., $\Sigma_\eta = \sigma_n^2\mathbf{I}$

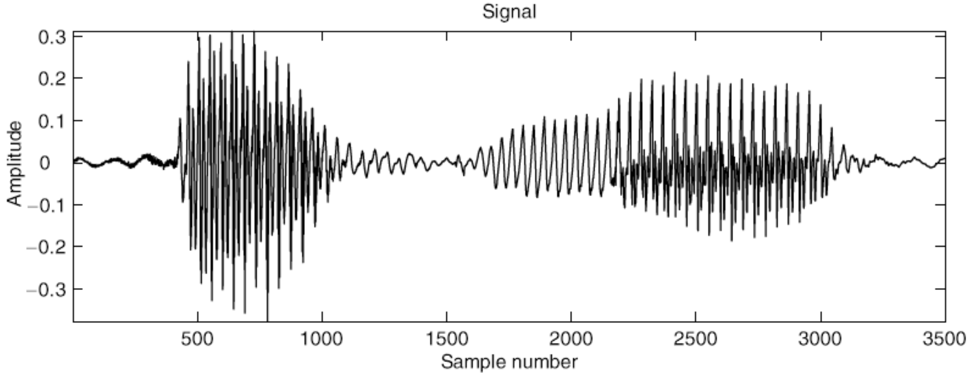
– From the receiver perspective, the modulated symbol vector $\mathbf{x}$ is a random. Detector sees it usually as uniformly distributed with independent elements ($\Sigma_x = E_x\mathbf{I}$) over the modulation alphabet Ω, where E_x is the symbol energy.

– The channel matrix $\mathbf{H}$ is often also a complex Gaussian with possibly some correlation between the elements. A slowly fading channel can also be realistically modeled as constant (known or unknown).

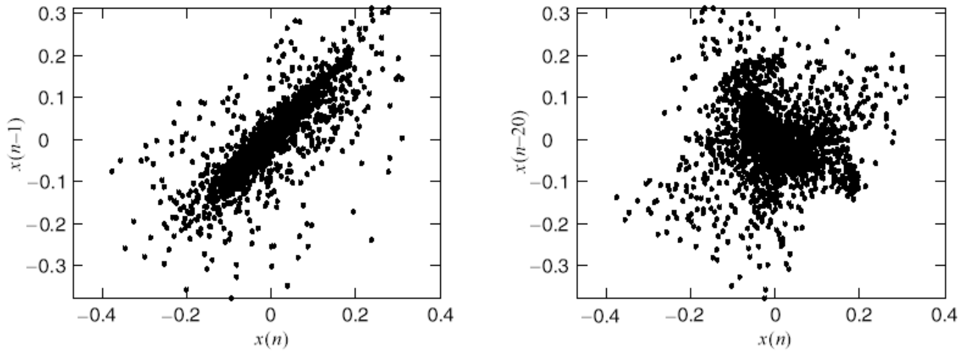
Based on the **probability density function** (**pdf**) of received signal vector, we are able to detect the transmitted signal, e.g. using optimum **maximum a posteriori (MAP) detector**, **maximum likelihood (ML) detector**.

Random Signals vs. Deterministic Signals (1)

Example: Discrete-time random signals (a) and the dependence between the samples (b), [5].



(a)



(b)

(a) The waveform for the speech signal "signal"; (b) two scatter plots for successive samples and samples separated by 20 sampling intervals.

Random Signals vs. Deterministic Signals (2)

- A **random signal** is not (precisely) predictable, it is not able to find a mathematical formula that provides its values as a function of time.

  A **white noise** is random signal when every sample is independent of all other samples. Such a signal is completely unpredictable.

- When signal samples are dependent and can be predicted precisely, it is **deterministic signal**.

Techniques for Processing Random Signals

- **Random signal analysis** (**signal modeling**, **spectral estimation**): Primary goal is to extract useful information for understanding and classifying the signals.

  Typical applications: Detection of useful information from receiving signals, system modeling/identification, detection and classification of radar and sonar targets, speech recognition, signal representation for data compression…

- **Random signal filtering** (**frequency-selective filtering**, **adaptive filtering**, **array processing**): Main objective is to improve the quality of a signal according to a criterion of performance.

  Typical applications: Noise and interference cancellation, echo cancellation, channel equalization, active noise control…

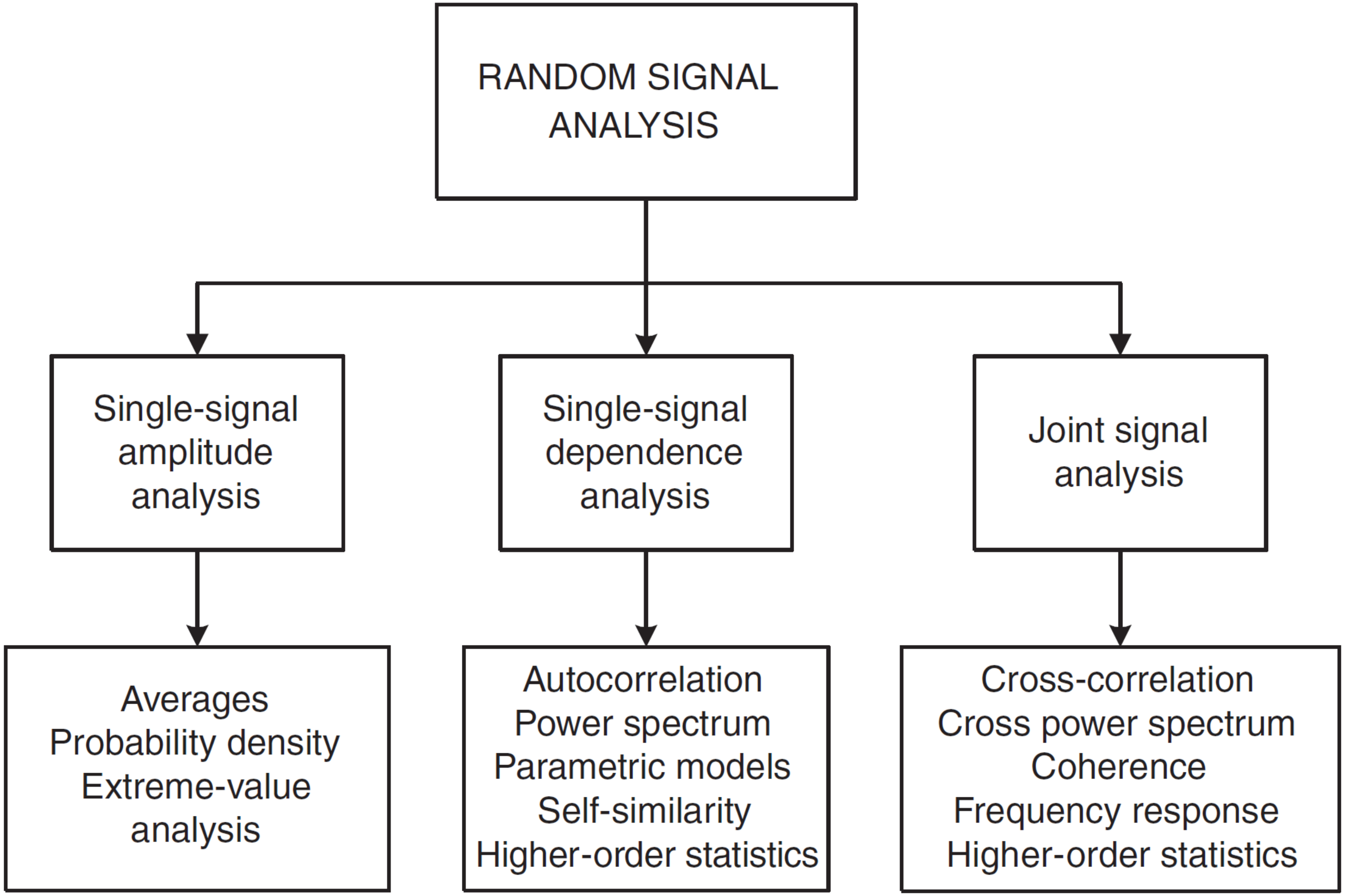
Random Signal Analysis (1)

- The **objective of signal analysis** is the development of quantitative techniques to study the properties of a signal and the differences and similarities between two or more signals from the same or different sources.

- The prominent tool in signal analysis is **spectral estimation**, which is a generic term for a multitude of techniques used to estimate the distribution of energy or power of a signal from a set of observations.

- The major areas of random signal analysis are:
  - Statistical analysis of signal amplitude (i.e., the sample values);
  - Analysis and modeling of the correlation among the samples of an individual signal;
  - Joint signal analysis (i.e., simultaneous analysis of two signals in order to investigate their interaction or interrelationships).

Random Signal Analysis (2)

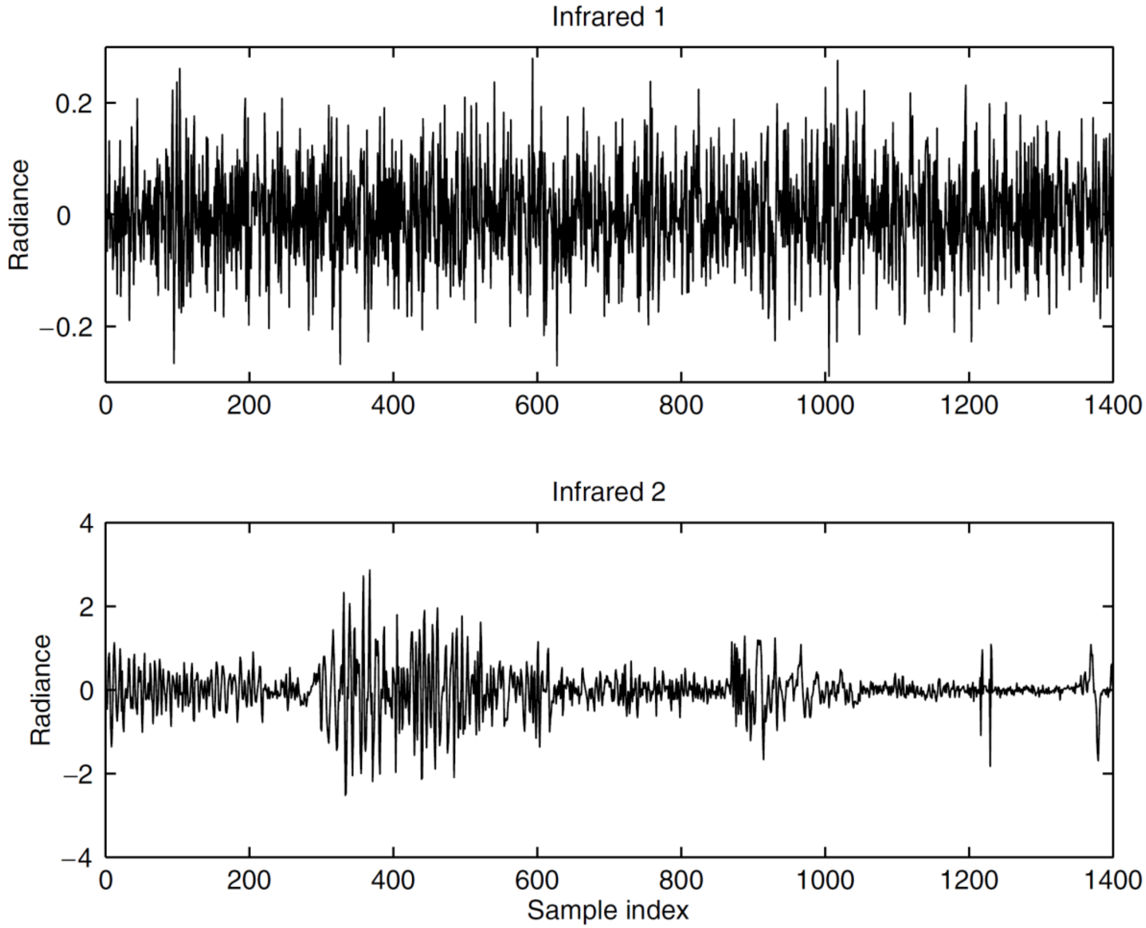


Random Signal Analysis (3)

- **Amplitude distribution:** The range of values taken by the samples of a signal and how often the signal assumes these values together determine the **signal variability**. The signal variability can be quantified by the **histogram** (**probability density**) of the signal samples, which shows the percentage of the signal amplitude values within a certain range. The numerical description of signal variability, which depends only on the value of the signal samples and not on their ordering, involves quantities such as **mean value**, **median**, **variance**, and **dynamic range**.

Random Signal Analysis (4)

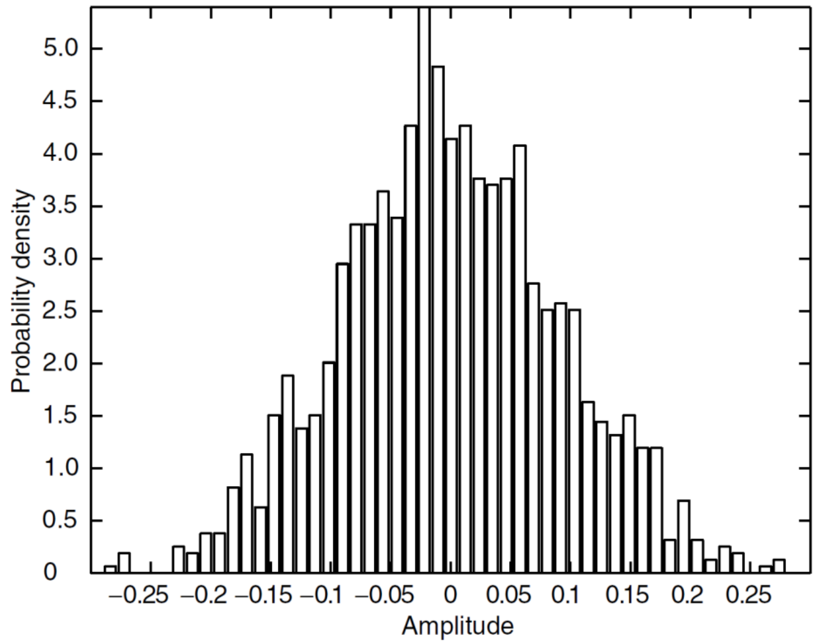
Example:



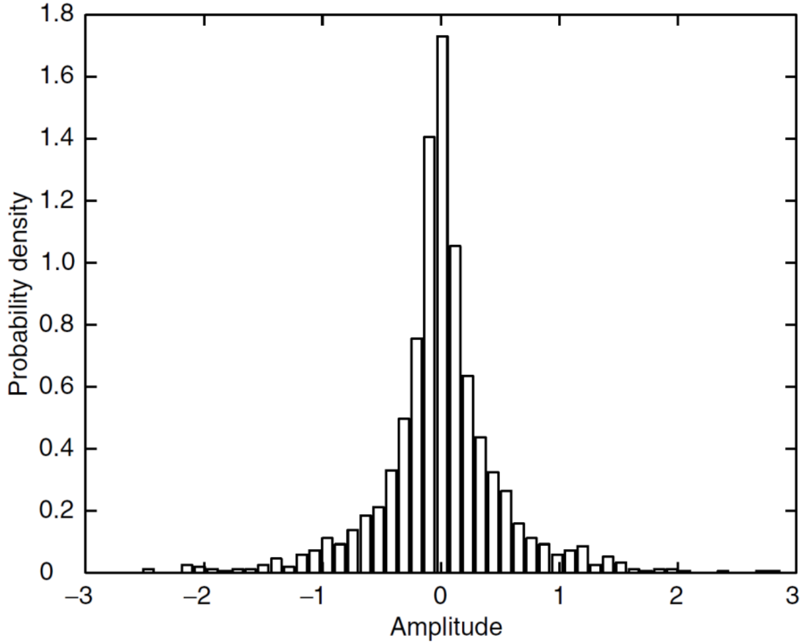
Two random signals (infrared 1 and infrared 2)

Random Signal Analysis (5)

Histograms for infrared 1 and infrared 2:



(a) Infrared 1



(b) Infrared 2

Random Signal Analysis (6)

- **Correlation and spectral analysis:** The scatter plot (see Slide 7) can be used to illustrate the existence of correlation, however, to obtain quantitative information about the correlation structure of a time series signal $x(n)$ with zero mean value, we use the **empirical normalized autocorrelation sequence:**

$$\hat{\rho}(l) = \frac{\sum_{n=l}^{N-1} x(n)x^*(n-l)}{\sum_{n=0}^{N-1} |x(n)|^2}$$

The **spectral density function** shows the distribution of signal power or energy as a function of frequency. <u>The autocorrelation and the spectral density of a signal form a Fourier transform pair</u> and hence contain the same information.

Random Signal Analysis (7)

- **Joint signal analysis.** In many applications, we are interested in the relationship between two different random signals. There are two cases of interest:
  - In the first case, the two signals are of the same or similar nature, and we want to ascertain and describe the similarity or interaction between them.
  - In the second case, we may have reason to believe that there is a causal relationship between the two signals. For example, one signal may be the input to a system and the other signal the output. The task in this case is to find an accurate description of the system, that is, a description that allows accurate estimation of future values of the output from the input. This process is known as **system modeling** or **system identification** and has many practical applications (including understanding the operation of a system in order to improve the design of new systems or to achieve better control of existing systems).

Random Signal Modeling (1)

- In many theoretical and practical applications, we are interested in generating random signals with certain properties or obtaining an efficient representation of real-world random signals that captures a desired set of their characteristics (e.g., correlation or spectral features) in the best possible way. We use the term **model** to refer to a mathematical description that provides an efficient representation of the "essential" properties of a signal.

  Example: A finite segment of any signal $\{x(n)\}_{n=0}^{N-1}$ can be approximated by a linear combination of constant ($\lambda_k = 1$) or exponentially fading ($0 < \lambda_k < 1$) sinusoids:

$$x(n) \simeq \sum_{k=1}^{M} a_k \lambda_k^n \cos\left(\omega_k n + \phi_k\right)$$

  where $\{a_k, \lambda_k, \omega_k, \phi_k\}_{k=1}^{M}$ are the model parameters.

Random Signal Modeling (2)

- From a practical viewpoint, we are most interested in **parametric models**, which assume a given functional form completely specified by a finite number of parameters. In contrast, **nonparametric models** do not put any restriction on the functional form or the number of model parameters.

- In practice, signal modeling involves the following steps:
  - Selection of an appropriate model.
  - Selection of the "right" number of parameters.
  - Fitting of the model to the actual data.
  - Model testing to see if the model satisfies the user requirements for the particular application.

Random Signal Modeling (3)

- If we can develop a successful parametric model for the behavior of a signal, then we can use the model for various applications:
  - To achieve a better understanding of the physical mechanism generating the signal.
  - To track changes in the source of the signal and help identify their cause.
  - To synthesize artificial signals similar to the natural ones (e.g., speech, infrared backgrounds, natural scenes, data network traffic).
  - To extract parameters for pattern recognition applications (e.g., speech and character recognition).
  - To get an efficient representation of signals for data compression (e.g., speech, audio, and video coding).
  - To forecast future signal behavior.

Adaptive Filtering (1)

- Conventional frequency-selective digital filters with **fixed coefficients** are designed to have a given frequency response chosen to alter the spectrum of the input signal in a desired manner. Their key features are as follows:
  - The filters are linear and time-invariant.
  - The design procedure uses the desired passband, transition bands, passband ripple, and stopband attenuation. We do *not* need to know the sample values of the signals to be processed.
  - Since the filters are frequency-selective, they work best when the various components of the input signal occupy non-overlapping frequency bands. For example, it is easy to separate a signal and additive noise when their spectra do not overlap.
  - The filter coefficients are chosen during the design phase and are held constant during the normal operation of the filter.

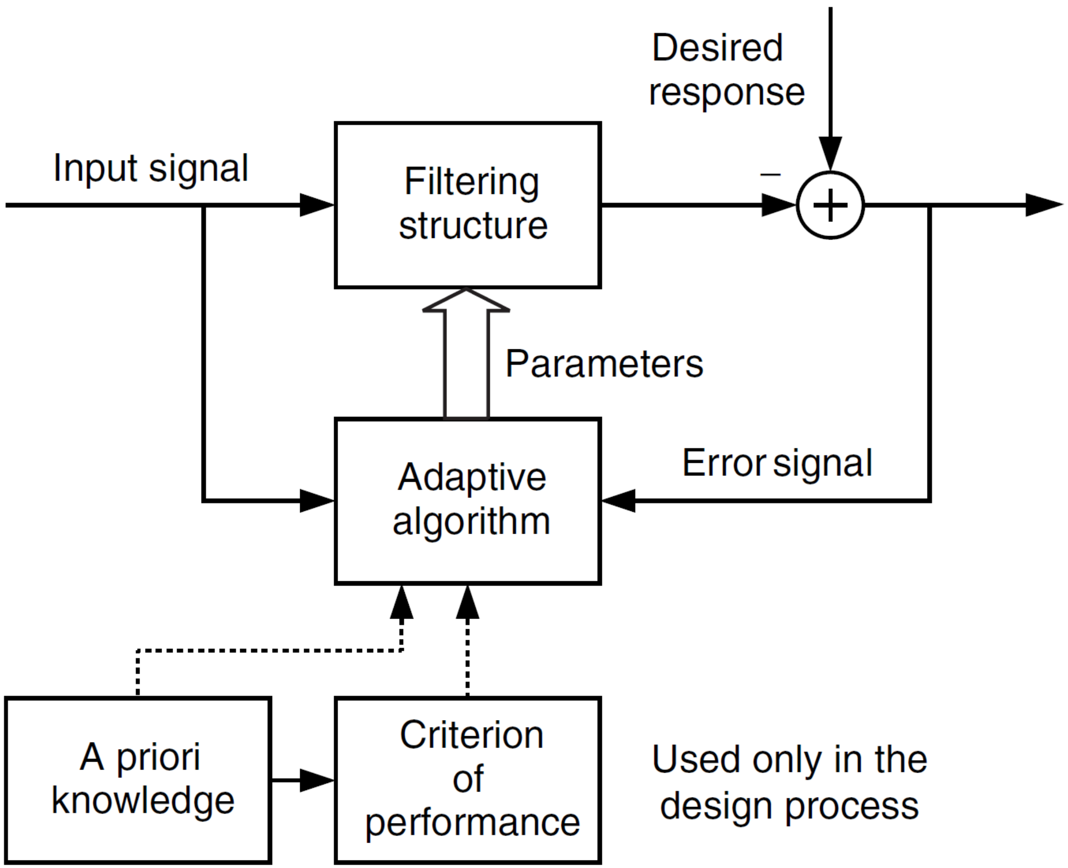
Adaptive Filtering (2)

- However, there are many practical application problems that cannot be successfully solved by using fixed digital filters because either we do not have sufficient information to design a digital filter with fixed coefficients or the design criteria change during the normal operation of the filter.

  Most of these applications can be successfully solved by using special "smart" filters known collectively as **adaptive filters**. The distinguishing feature of adaptive filters is that they <u>can modify their response to improve performance during operation without any intervention from the user</u>.

Adaptive Filtering (3)

- Basis elements of a general adaptive filters:



Adaptive Filtering (4)

- **Filtering structure.** This module forms the output of the filter using measurements of the input signal(s). The filtering structure is linear if the output is obtained as a linear combination of the input measurements; otherwise, it is said to be nonlinear. The structure is fixed by the designer, and its parameters are adjusted by the adaptive algorithm.
- **Criterion of performance (COP).** The output of the adaptive filter and the desired response (when available) are processed by the COP module to assess its quality with respect to the requirements of the particular application.
- **Adaptive algorithm.** The adaptive algorithm uses the value of the criterion of performance, or some function of it, and the measurements of the input and desired response (when available) to decide how to modify the parameters of the filter to improve its performance.

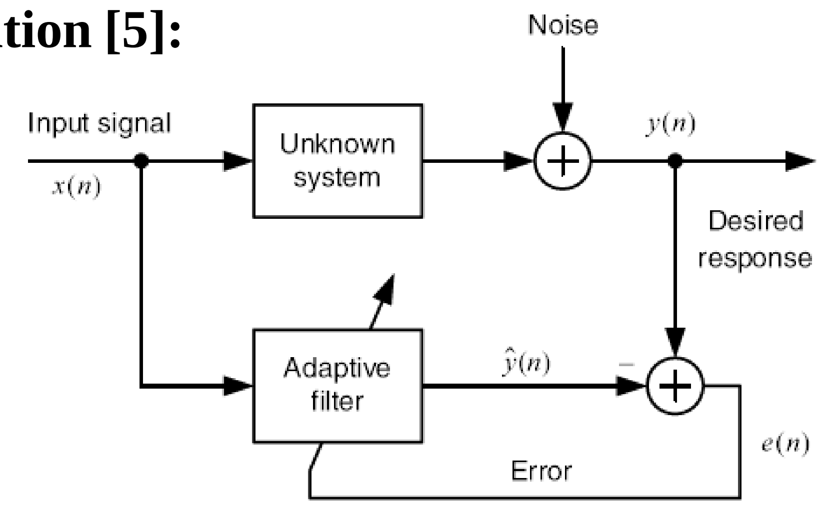
Adaptive Filtering (5)

- Typical application problems that can be effectively solved by using an adaptive filter:

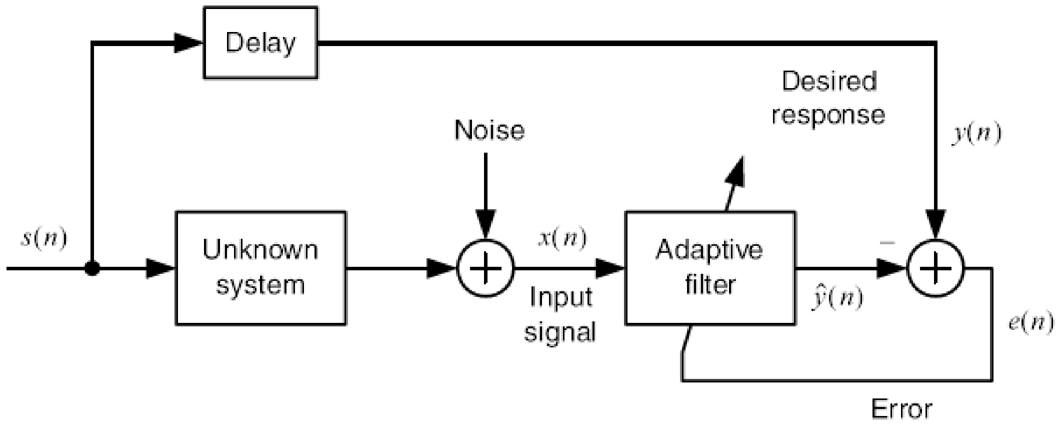
Application class	Examples
System identification	Echo cancelation Adaptive control Channel modeling
System inversion	Adaptive equalization Blind deconvolution
Signal prediction	Adaptive predictive coding Change detection Radio frequency interference cancelation
Multisensor interference cancelation	Acoustic noise control Adaptive beamforming

Adaptive Filtering (6)

System identification [5]:

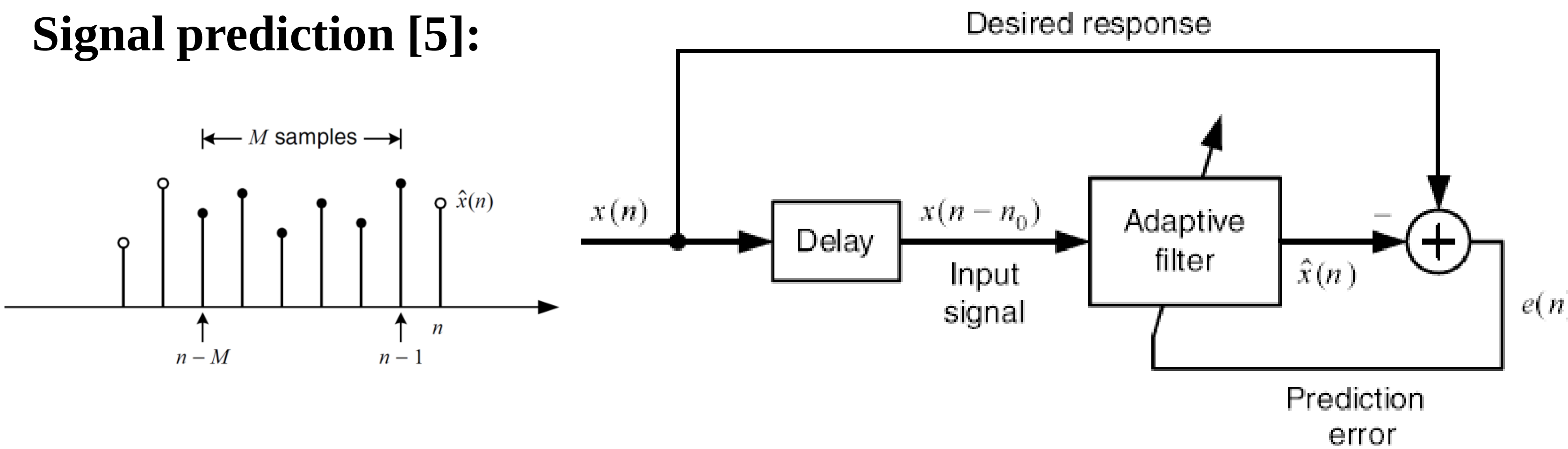


System inversion [5]:

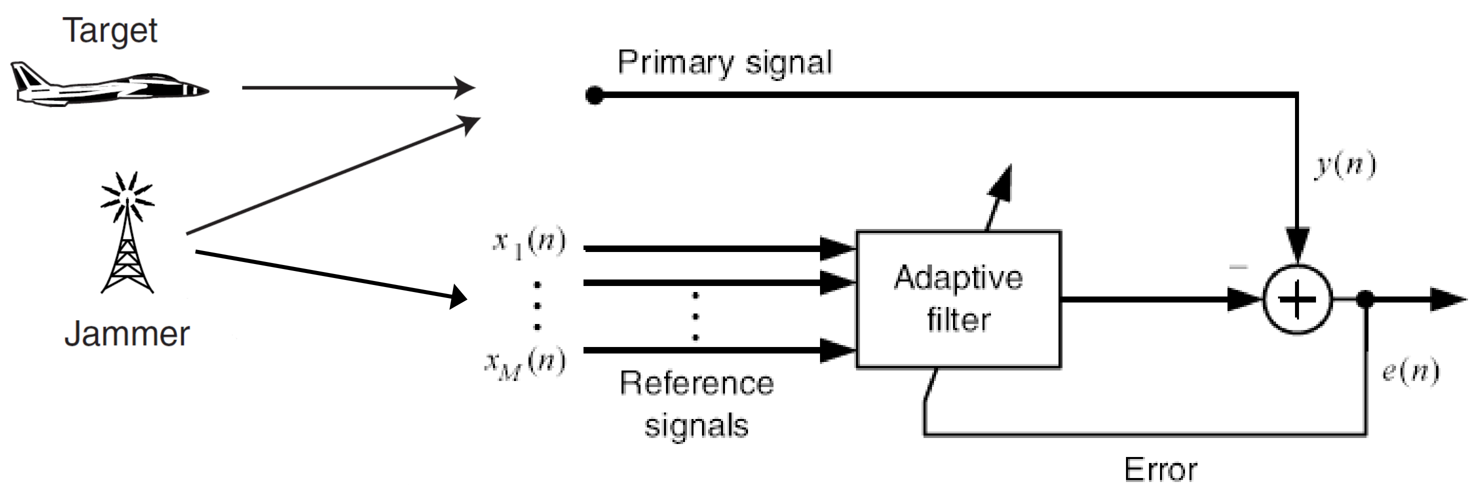


Adaptive Filtering (7)

Signal prediction [5]:

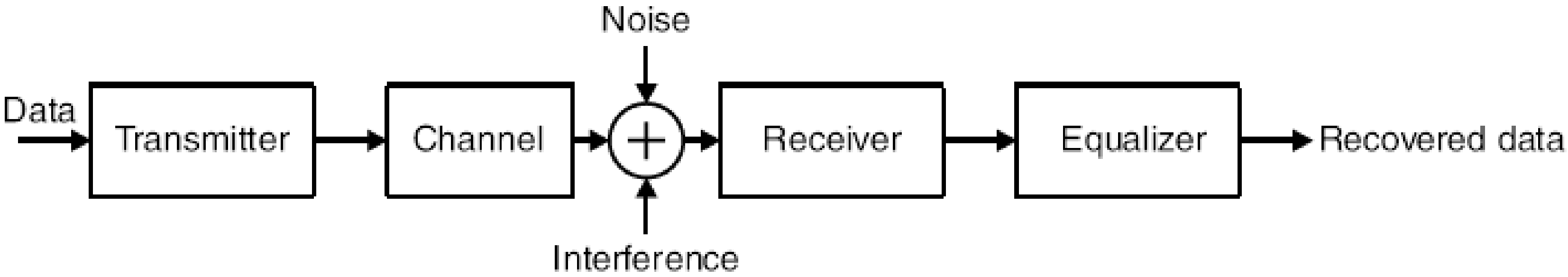


Interference cancellation [5]:



Adaptive Filtering (8)

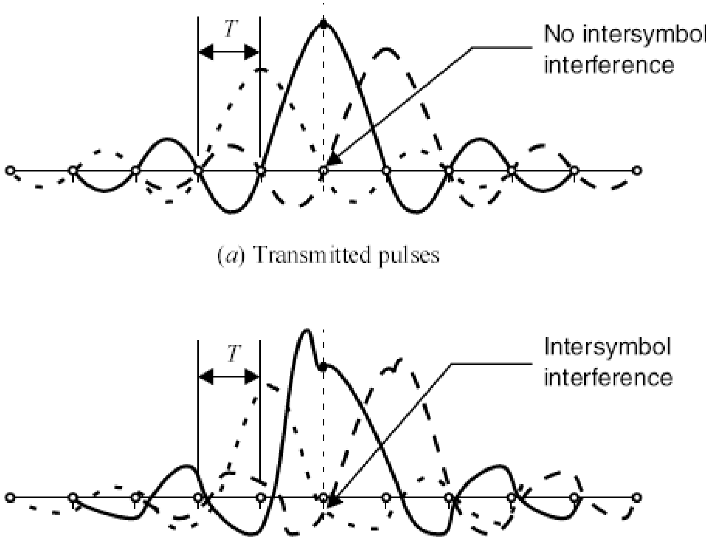
Simple model of a digital communications system with <u>channel equalization</u> [5]:



Pulse trains: (a) without intersymbol interference (ISI) and (b) with ISI.

ISI distortion: The tails of adjacent pulses interfere the current pulse and can lead to an incorrect decision.

The **equalizer** can compensate for the ISI distortion.

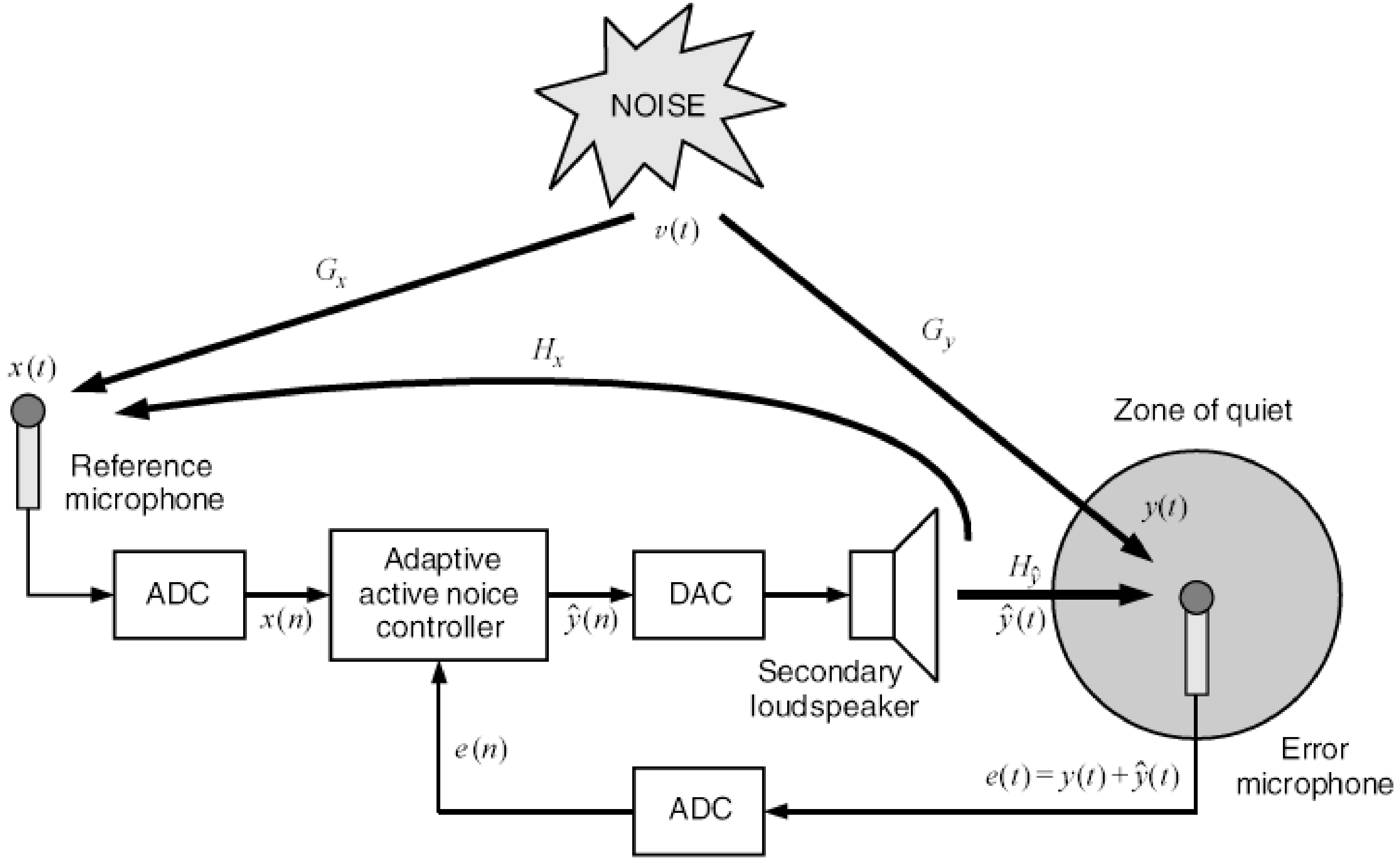


(a) Transmitted pulses

(b) Distorted pulses

Adaptive Filtering (9)

Block diagram of the basic components of an active noise control system [5]:



Array Processing (1)

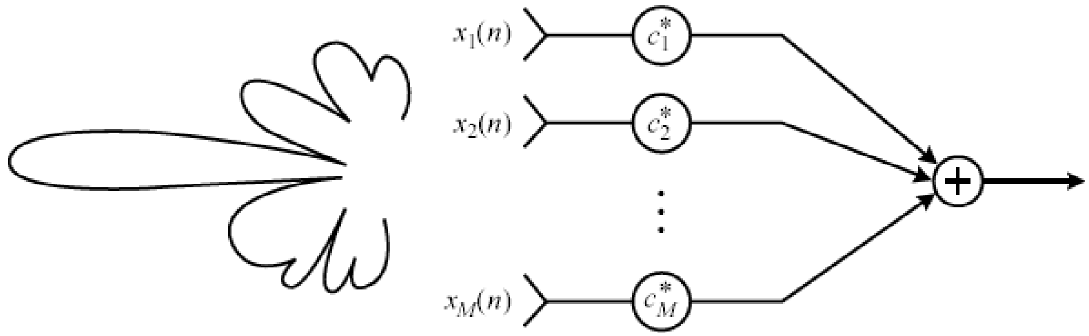
- **Array processing** deals with techniques for the analysis and processing of signals collected by a **group of sensors** (**sensor array**).

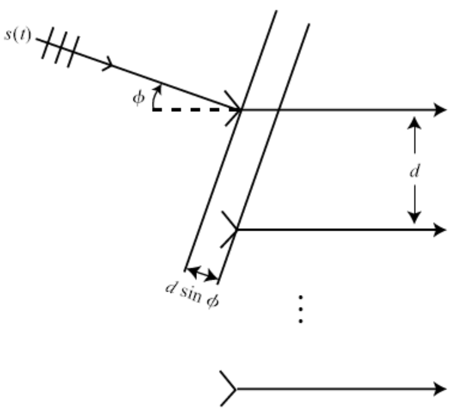
  The collection of sensors makes up the array, and the manner in which the signals from the sensors are combined and handled constitutes the processing.

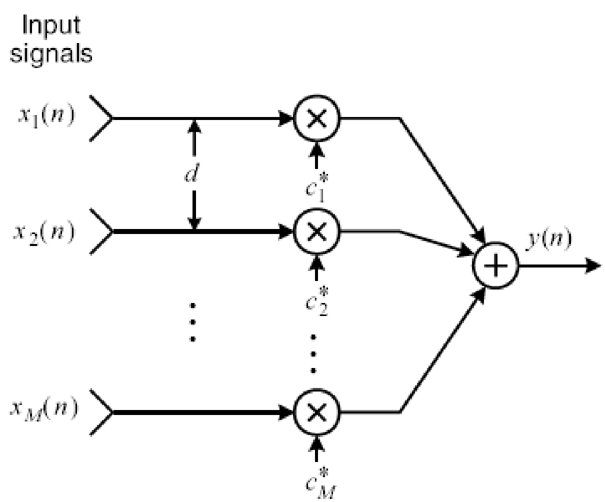
  The type of processing is dictated by the needs of the particular application. Array processing has found widespread application in a large number of areas, including radar, sonar, communications, seismology, geophysical prospecting for oil and natural gas, diagnostic ultrasound, and multi-channel audio systems.

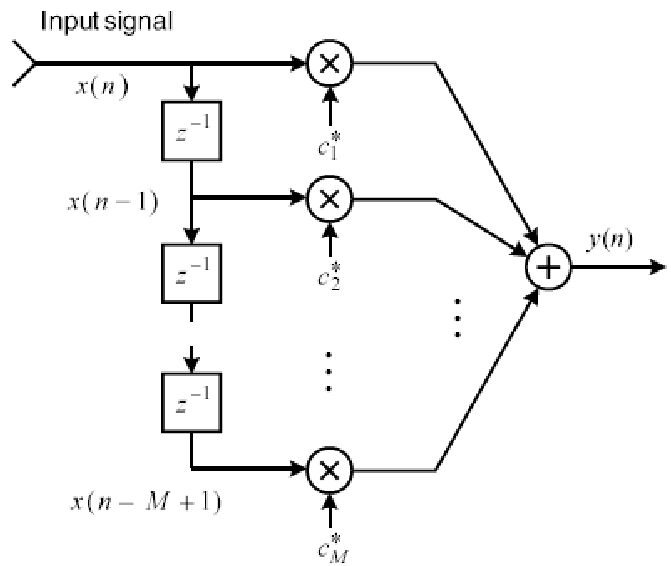
Array Processing (2)

Spatial filtering (Beamforming) [5]:

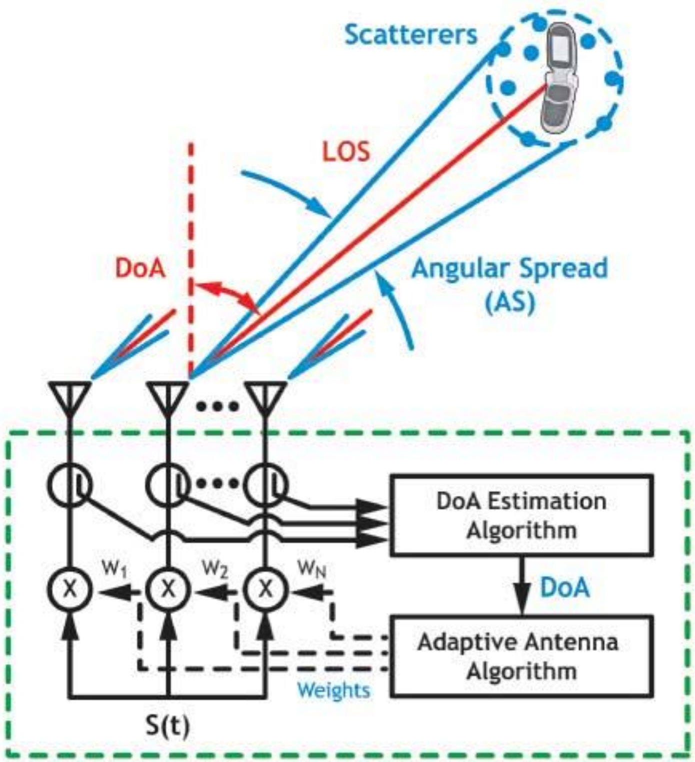




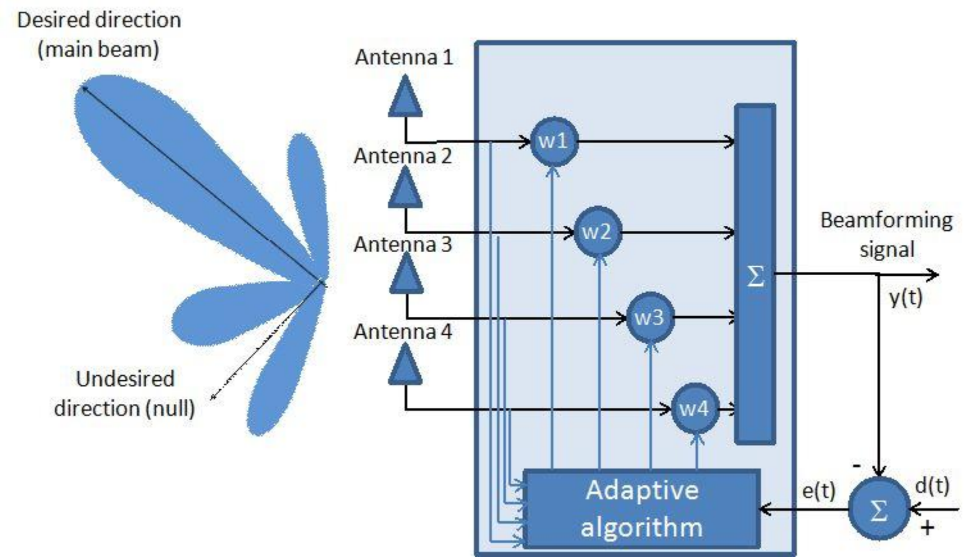




Array Processing (3)



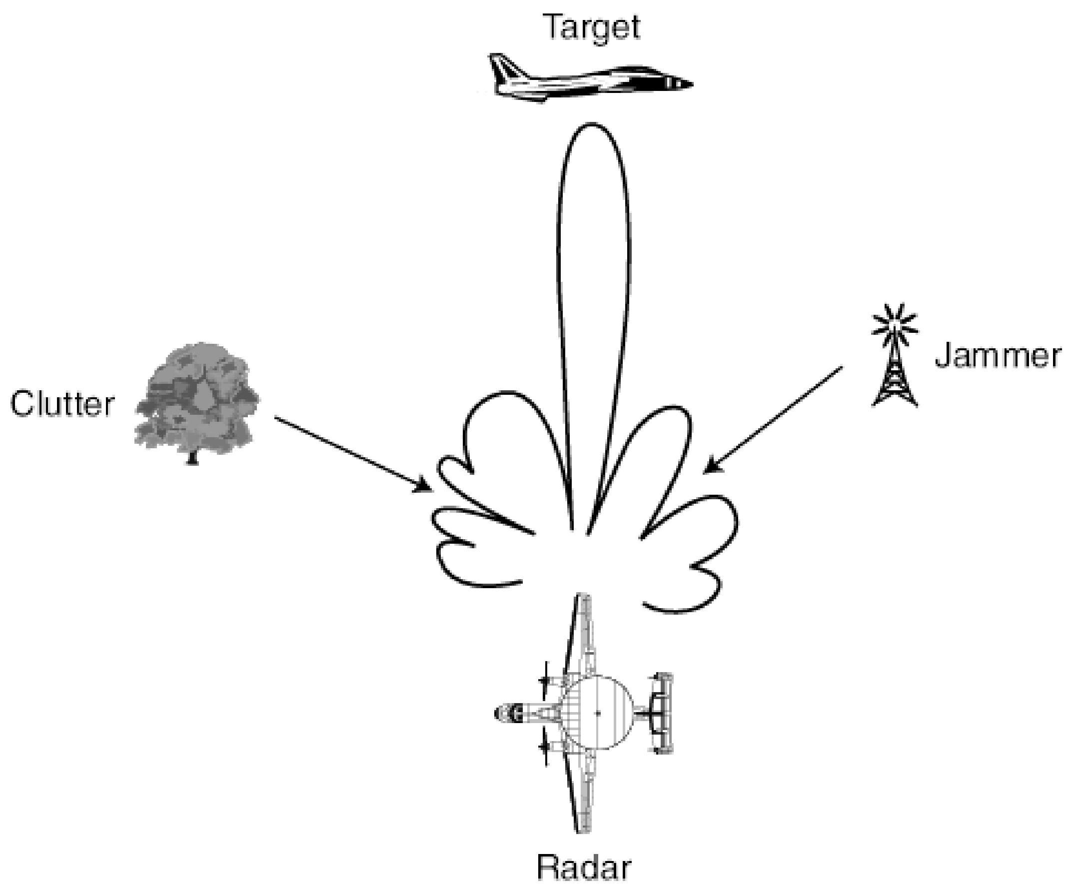
Beamforming based on
DoA estimation (e.g. at Tx)



Beamforming based on
Adaptive algorithms (e.g. at Rx)

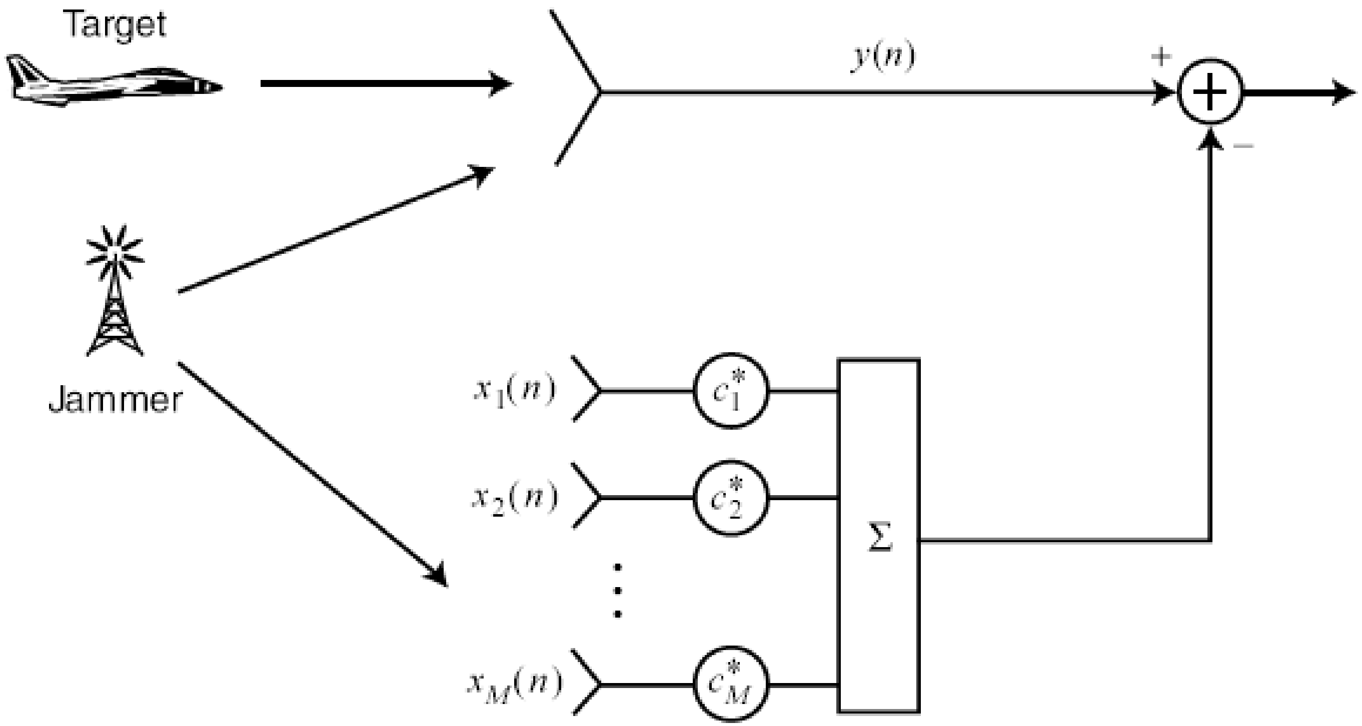
Array Processing (4)

Example of (adaptive) beamforming with an airborne for interference mitigation [5]:



Array Processing (5)

(Adaptive) Sidelobe canceler [5]:



Chapter 1:
Discrete-Time Signal Processing

- *z*-Transform.
- Linear Time-Invariant Filters.
- Discrete Fourier Transform (DFT).

1. *Z*-transform (1)

- **Discrete-time signals** $\rightarrow$ signals described as a *time series*, consisting of *sequence of uniformly spaced samples* whose *varying amplitudes carry the useful information content* of the signal.

- Consider time series (sequence) $\{u(n)\}$ or $u(n)$ denoted by samples:
  $u(n)$, $u(n\text{-}1)$, $u(n\text{-}2)$,…, n: discrete time.
  - **Z-transform** of $u(n)$:

$$U(z) = z[u(n)] = \sum_{n=-\infty}^{\infty} u(n) z^{-n} \qquad (1.1)$$

    z: complex variable.

    z-transform pair: $u(n) \leftrightarrow U(z)$.
  - Region of convergence (ROC): set of values of z for which $U(z)$ is uniformly convergent.

1. Z-transform (2)

- **Properties:**
  - Linear transform (superposition):

$$au_1(n) + bu_2(n) \leftrightarrow aU_1(z) + bU_2(z) \quad (1.2)$$

ROC of (1.2): intersection of ROC of $U_1(z)$ and ROC of $U_2(z)$.

  - Time-shifting:

$$u(n) \leftrightarrow U(z) \Rightarrow u(n - n_0) \leftrightarrow z^{n_0} U(z) \quad (1.3)$$

n_0: integer.

ROC of (1.3): same as ROC of $U(z)$.

Special case: $n_0 = 1 \Rightarrow u(n-1) \leftrightarrow z^{-1} U(z)$

z^{-1}: unit-delay element.

  - Convolution theorem:

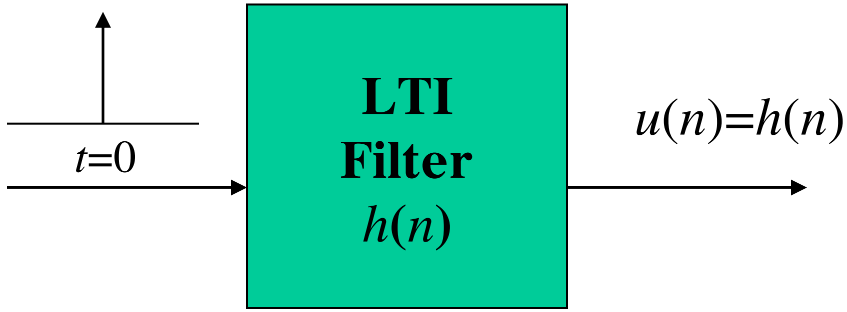
$$\sum_{i=-\infty}^{\infty} u_1(n) u_2(n-i) \leftrightarrow U_1(z) U_2(z) \quad (1.4)$$

ROC of (1.4): intersection of ROC of $U_1(z)$ and ROC of $U_2(z)$.

1. Linear Time-Invariant (LTI) Filters (1)

- **Definition:**

- **Impulse response $h(n)$:**



For arbitrary input $v(n)$: *convolution sum*

$$u(n)=\sum_{i=-\infty}^{\infty}h(i)v(n-i) \qquad (1.5)$$

1. LTI Filters (2)

- **Transfer function:**
  - Applying z-transform to both side of (1.5)

$$U(z) = H(z)V(z)$$
$$U(z) \leftrightarrow u(n),\ V(z) \leftrightarrow v(n),\ H(z) \leftrightarrow h(n) \qquad (1.6)$$

H(*z*): *transfer function* of the filter

$$H(z) = \frac{U(z)}{V(z)} \qquad (1.7)$$

  - When input sequence *v*(*n*) and output sequence *u*(*n*) are related by difference equation of order *N*:

$$\sum_{j=0}^{N} a_j u(n-j) = \sum_{j=0}^{N} b_j v(n-j) \qquad (1.8)$$

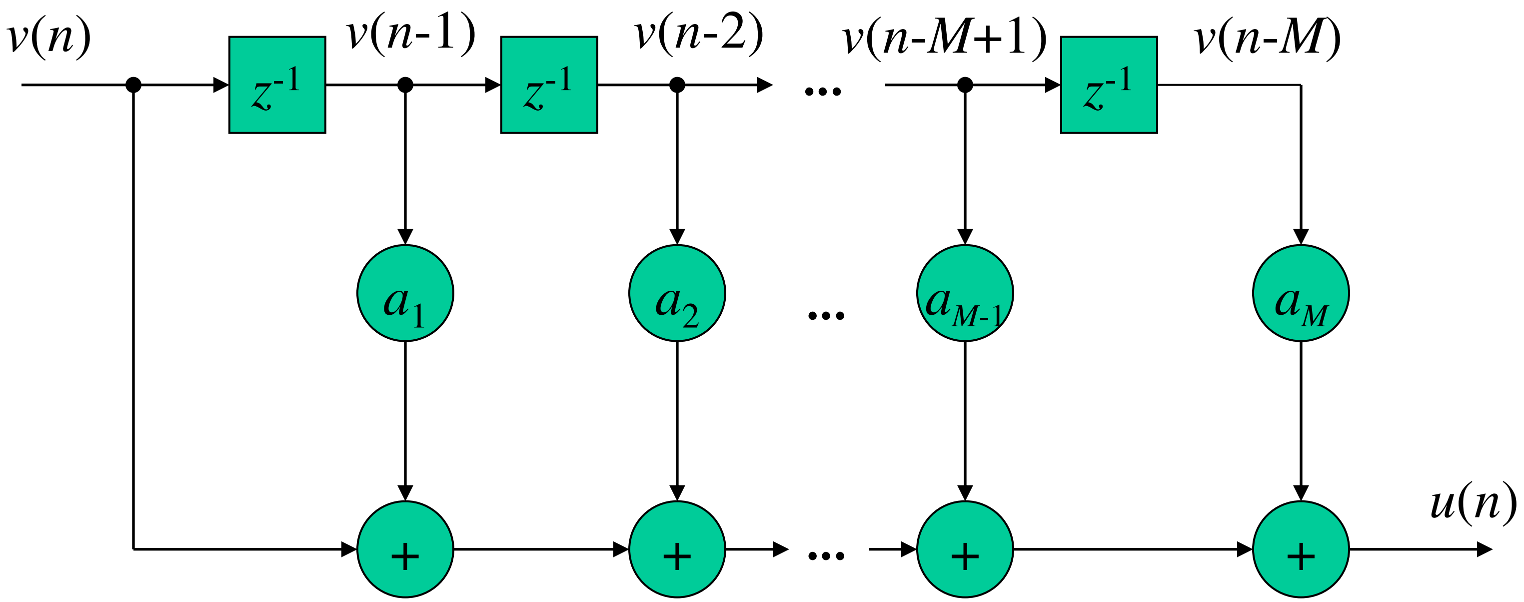
a_j, b_j: constant coefficients. Applying z-transform, obtaining:

$$H(z) = \frac{U(z)}{V(z)} = \frac{\sum_{j=0}^{N} a_j z^{-j}}{\sum_{j=0}^{N} b_j z^{-j}} = \frac{a_0}{b_0} \frac{\prod_{k=1}^{N}(1 - c_k z^{-1})}{\prod_{k=1}^{N}(1 - d_k z^{-1})} \qquad (1.9)$$

1. LTI Filters (3)

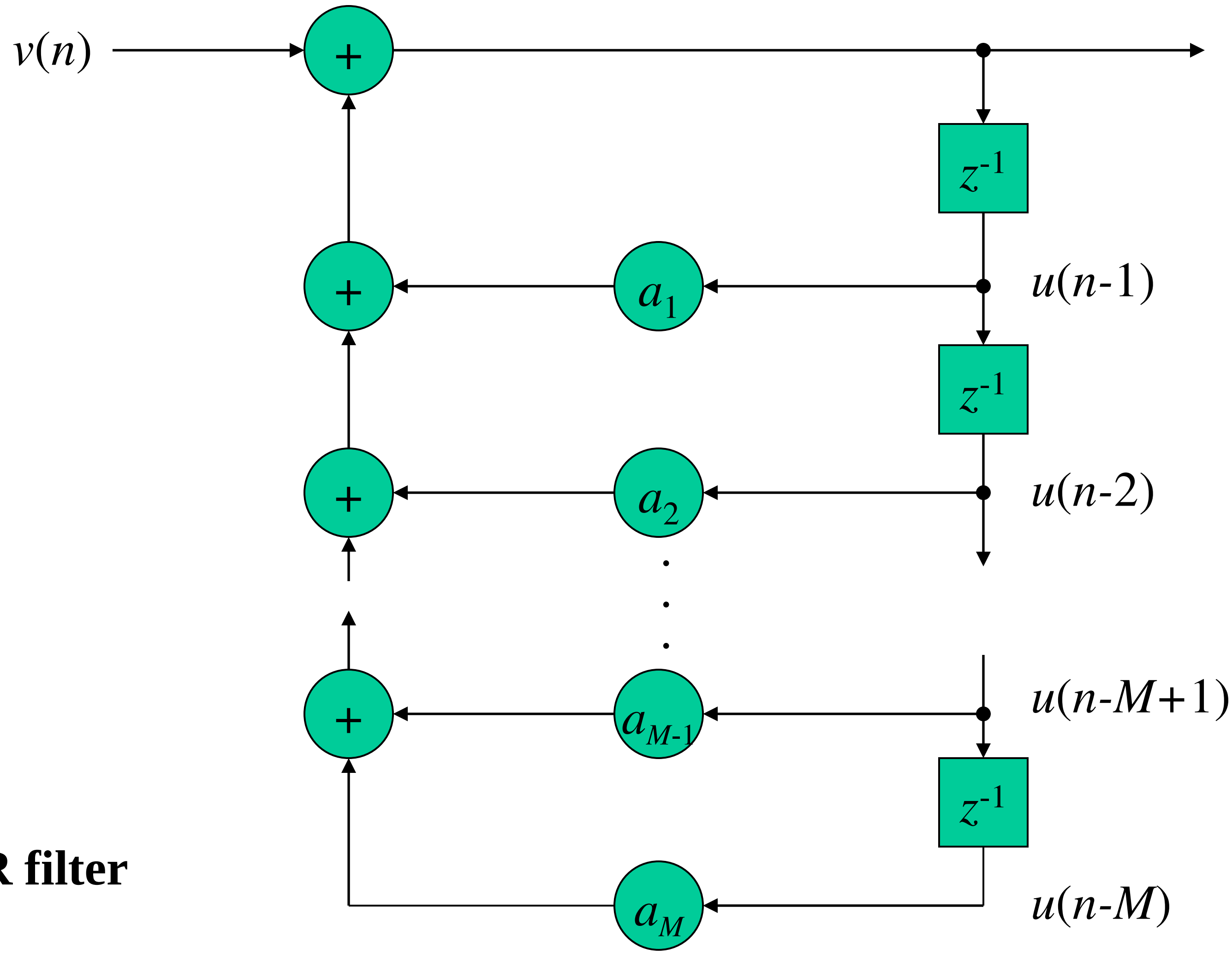
- From (1.9), two distinct types of LTI filters:
  - *Finite-duration impulse response (FIR) filters:* d_k=0 for all k.
    $\rightarrow$ *all-zero filter*, $h(n)$ has finite duration.
  - *Infinite-duration impulse response (IIR) filters:* $H(z)$ has at least one non-zero pole, $h(n)$ has infinite duration. When c_k=0 for all k $\rightarrow$ *all-pole filter*.
  - See examples of FIR and IIR filters in next two slides.

1. LTI Filters (4)



FIR filter

1. LTI Filters (5)



IIR filter

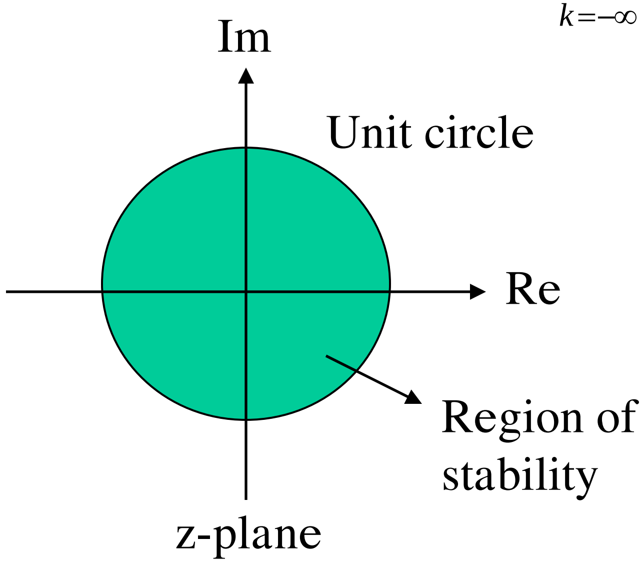
1. LTI Filters (6)

- **Causality and stability:**
  - LTI filter is *causal* if:

$$h(n) = 0 \quad \text{for} \quad n < 0 \qquad (1.10)$$

  - LTI filter is *stable* if output sequence is bounded for all bounded input sequences. From (1.5), necessary and sufficient condition:

$$\sum_{k=-\infty}^{\infty} |h(k)| < \infty \qquad (1.11)$$



A causal LTI filter is stable if and only if all of the poles of the filter's transfer function lie inside the unit circle in the z-plane

(See more in [3])

1. Discrete Fourier Transform (DFT) (1)

- **Fourier transform** of a sequence $u(n)$ is obtained from z-transform by setting z=exp($j2\pi f$), f: real frequency variable.

  When $u(n)$ has a finite duration, its Fourier representation → **discrete Fourier transform (DFT).**

  For numerical computation of DFT → efficient **fast Fourier transform (FFT).**

- $u(n)$: finite-duration sequence of length N, DFT of $u(n)$:

$$U(k)=\sum_{n=0}^{N-1}u(n)\exp\left(-\frac{j2\pi kn}{N}\right),\quad k=0,\ldots,N-1 \qquad (1.12)$$

  Inverse DFT (IDFT) of $U(k)$:

$$u(n)=\frac{1}{N}\sum_{k=0}^{N-1}U(k)\exp\left(\frac{j2\pi kn}{N}\right),\quad n=0,\ldots,N-1 \qquad (1.13)$$

  $u(n)$, $U(k)$: same length N → "N-point DFT"

Chapter 2:
Stochastic Processes and Models

- ❑ Review of Probability and Random Variables.
- ❑ Review of Stochastic Processes and Stochastic Models.

2. Review of Probability and Random Variables

- Axioms of Probability.
- Repeated Trials.
- Concepts of Random Variables.
- Functions of Random Variables.
- Moments and Conditional Statistics.
- Sequences of Random Variables.

2. Review of Probability Theory: Basics (1)

- Probability theory deals with the study of random phenomena, which under repeated experiments yield different outcomes that have certain underlying patterns about them. The notion of an experiment assumes a set of repeatable conditions that allow any number of identical repetitions. When an experiment is performed under these conditions, certain elementary events ξ_i occur in different but *completely uncertain* ways. We can assign nonnegative number $P(\xi_i)$, as the probability of the event ξ_i in various ways:

  Laplace's Classical Definition: The Probability of an event A is defined a-priori without actual experimentation as

$$P(A) = \frac{\text{Number of outcomes favorable to } A}{\text{Total number of possible outcomes}} \quad ,$$

  provided all these outcomes are *equally likely*.

2. Review of Probability Theory: Basics (2)

Relative Frequency Definition: The probability of an event A is defined as

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

where n_A is the number of occurrences of A and n is the total number of trials.

The axiomatic approach to probability, due to Kolmogorov, developed through a set of axioms (below) is generally recognized as superior to the above definitions, as it provides a solid foundation for complicated applications.

2. Review of Probability Theory: Basics (3)

The totality of all ξ_i, *known a priori*, constitutes a set Ω, the set of all experimental outcomes.

$$\Omega = \{ \xi_1, \xi_2, \cdots, \xi_k, \cdots \}$$

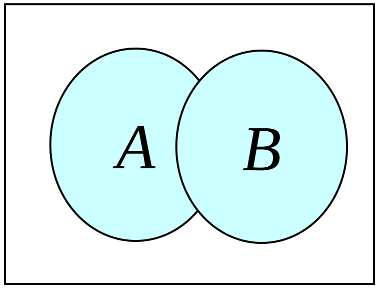
Ω has subsets $A, B, C, \cdots$. Recall that if A is a subset of Ω, then $\xi \in A$ implies $\xi \in \Omega$. From A and B, we can generate other related subsets $A \cup B, A \cap B, \overline{A}, \overline{B}$,

$$A \cup B = \{ \xi \mid \xi \in A \text{ or } \xi \in B \}$$
$$A \cap B = \{ \xi \mid \xi \in A \text{ and } \xi \in B \}$$

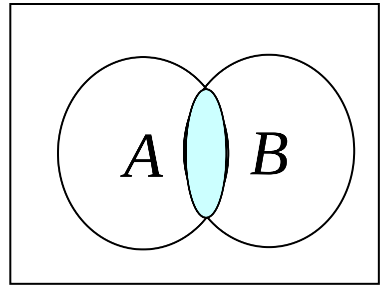
and

$$\overline{A} = \{ \xi \mid \xi \notin A \}$$

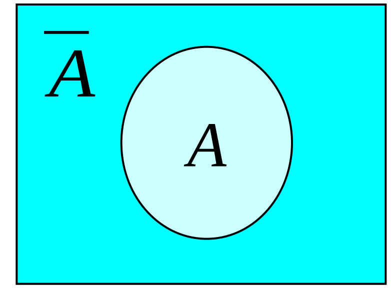
2. Review of Probability Theory: Basics (4)



$A \cup B$



$A \cap B$

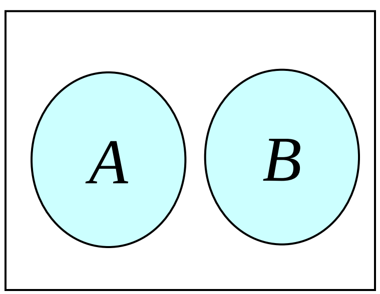


$\overline{A}$

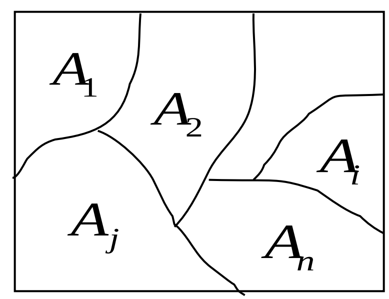
If $A \cap B = \phi$, the empty set, then A and B are said to be mutually exclusive (M.E).

A partition of Ω is a collection of mutually exclusive subsets of Ω such that their union is Ω.

$$A_i \cap A_j = \phi, \quad \text{and} \quad \bigcup_{i=1} A_i = \Omega.$$



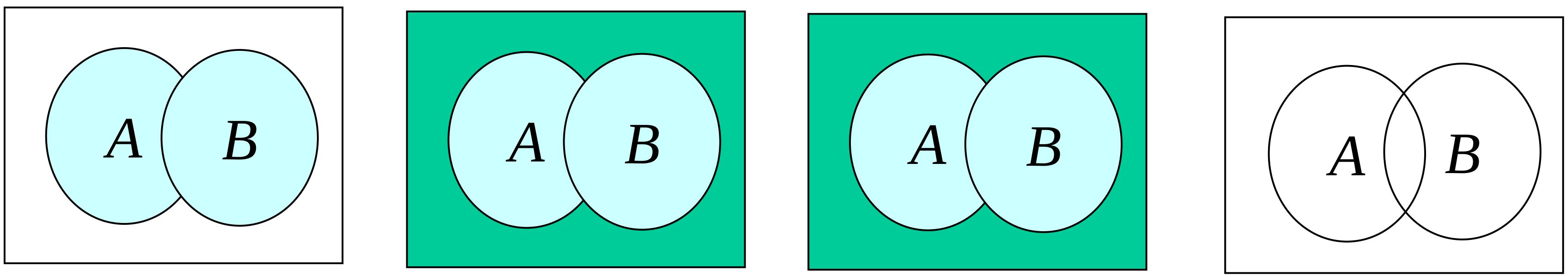
$A \cap B = \phi$



2. Review of Probability Theory: Basics (5)

De-Morgan's Laws:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}; \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}$$



$A \cup B$ $\qquad$ $\overline{A \cup B}$ $\qquad$ $\overline{A} \cap \overline{B}$

Often it is meaningful to talk about at least some of the subsets of Ω as events, for which we must have mechanism to compute their probabilities.

<u>Example</u>: Consider the experiment where two coins are simultaneously tossed. The various elementary events are

2. Review of Probability Theory: Basics (6)

$$\xi_1 = (H, H), \ \ \xi_2 = (H, T), \ \ \xi_3 = (T, H), \ \ \xi_4 = (T, T)$$

and

$$\Omega = \{ \xi_1, \xi_2, \xi_3, \xi_4 \}.$$

The subset $A = \{ \xi_1, \xi_2, \xi_3 \}$ is the same as "*Head has occurred at least once*" and qualifies as an event.

Suppose two subsets A and B are both events, then consider

"*Does an outcome belong to A or B* $= A \cup B$ "

"*Does an outcome belong to A and B* $= A \cap B$"

"*Does an outcome fall outside A*"?

2. Review of Probability Theory: Basics (7)

Thus the sets $A \cup B$, $A \cap B$, $\overline{A}$, $\overline{B}$, etc., also qualify as events. We shall formalize this using the notion of a Field.

Field: A collection of subsets of a nonempty set Ω forms a field F if

(i) $\Omega \in F$

(ii) If $A \in F$, then $\overline{A} \in F$

(iii) If $A \in F$ and $B \in F$, then $A \cup B \in F$.

Using (i) - (iii), it is easy to show that $A \cap B$, $\overline{A} \cap B$, etc., also belong to F.

For example, from (ii) we have $\overline{A} \in F$, $\overline{B} \in F$, and using (iii) this gives $\overline{A} \cup \overline{B} \in F$; applying (ii) again we get $\overline{\overline{A} \cup \overline{B}} = A \cap B \in F$, where we have used De Morgan's theorem in slide 22.

2. Review of Probability Theory: Basics (8)

- **Axioms of Probability**

For any event A, we assign a number $P(A)$, called the probability of the event A. This number satisfies the following three conditions that act the *axioms of probability*.

$$\text{(i)} \quad P(A) \geq 0 \quad \text{(Probability is a nonnegative number)}$$

$$\text{(ii)} \quad P(\Omega) = 1 \quad \text{(Probability of the whole set is unity)}$$

$$\text{(iii)} \quad \text{If } A \cap B = \phi, \quad \text{then} \quad P(A \cup B) = P(A) + P(B).$$

(Note that (iii) states that if A and B are mutually exclusive (M.E.) events)

2. Review of Probability Theory: Basics (9)

The following conclusions follow from these axioms:

a. $P(A \cup \overline{A}) = P(A) + P(\overline{A}) = 1$ or $P(\overline{A}) = 1 - P(A).$

b. $P\{\phi\} = 0.$

c. Suppose A and B are *not* mutually exclusive (M.E.)?

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

❑ Conditional Probability and Independence

In N independent trials, suppose $N_A,\ N_B,\ N_{AB}$ denote the number of times events A, B and AB occur respectively, for large N

$$P(A) \approx \frac{N_A}{N}, \quad P(B) \approx \frac{N_B}{N}, \quad P(AB) \approx \frac{N_{AB}}{N}.$$

Among the N_A occurrences of A, only N_{AB} of them are also found among the N_B occurrences of B. Thus the ratio

$$\frac{N_{AB}}{N_B} = \frac{N_{AB}/N}{N_B/N} = \frac{P(AB)}{P(B)}$$

2. Review of Probability Theory: Basics (10)

is a measure of "*the event A given that B has already occurred*". We denote this conditional probability by

$P(A|B) =$ Probability of "*the event A given that B has occurred*".

We define

$$P(A \mid B) = \frac{P(AB)}{P(B)},$$

provided $P(B) \neq 0.$ As we show below, the above definition satisfies all probability axioms discussed earlier.

Independence: *A* and *B* are said to be independent events, if

$$P(AB) = P(A) \cdot P(B).$$

Then

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$

Thus if *A* and *B* are independent, the event that *B* has occurred does not shed any more light into the event *A*. It makes no difference to *A* whether *B* has occurred or not.

2. Review of Probability Theory: Basics (11)

Let

$$A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n , \qquad (*)$$

a union of n independent events. Then by De-Morgan's law

$$\overline{A} = \overline{A}_1 \overline{A}_2 \cdots \overline{A}_n$$

and using their independence

$$P(\overline{A}) = P(\overline{A}_1 \overline{A}_2 \cdots \overline{A}_n) = \prod_{i=1}^{n} P(\overline{A}_i) = \prod_{i=1}^{n} (1 - P(A_i)).$$

Thus for any A as in (*)

$$P(A) = 1 - P(\overline{A}) = 1 - \prod_{i=1}^{n} (1 - P(A_i)),$$

a useful result.

2. Review of Probability Theory: Basics (12)

Bayes' theorem:

$$P(A|B) = \frac{P(B|A)}{P(B)} \cdot P(A)$$

Although simple enough, Bayes' theorem has an interesting interpretation: $P(A)$ represents the a-priori probability of the event A. Suppose B has occurred, and assume that A and B are not independent. How can this new information be used to update our knowledge about A? Bayes' rule above take into account the new information ("B has occurred") and gives out the a-posteriori probability of A given B.

We can also view the event B as new knowledge obtained from a fresh experiment. We know something about A as $P(A)$. The new information is available in terms of B. The new information should be used to improve our knowledge/understanding of A. Bayes' theorem gives the exact mechanism for incorporating such new information.

2. Review of Probability Theory: Basics (13)

Let $A_1, A_2, ..., A_n$ are pair wise disjoint $A_i A_j = \phi$, we have

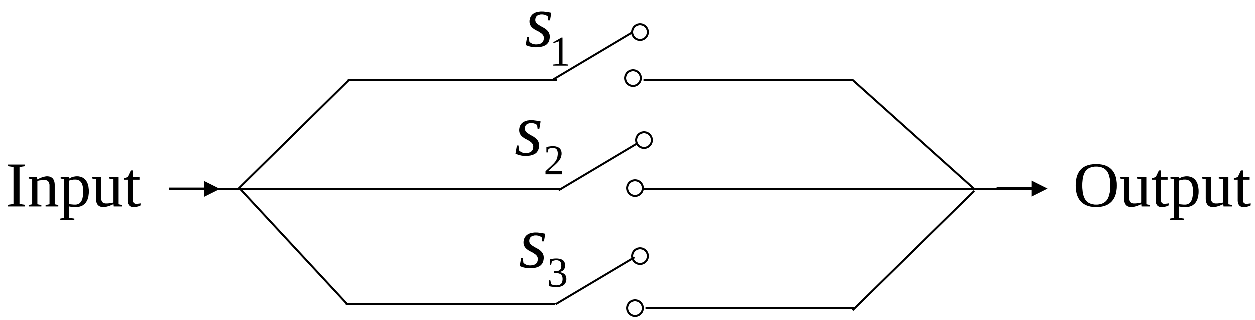
$$P(B) = \sum_{i=1}^{n} P(BA_i) = \sum_{i=1}^{n} P(B \mid A_i) P(A_i).$$

A more general version of Bayes' theorem

$$P(A_i \mid B) = \frac{P(B \mid A_i) P(A_i)}{P(B)} = \frac{P(B \mid A_i) P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i) P(A_i)},$$

2. Review of Probability Theory: Basics (14)

Example: Three switches connected in parallel operate independently. Each switch remains closed with probability p. (a) Find the probability of receiving an input signal at the output. (b) Find the probability that switch S_1 is open given that an input signal is received at the output.



Solution:

a. Let A_i = "*Switch* S_i *is closed*". Then $P(A_i) = p, \; i = 1 \to 3.$
Since switches operate independently, we have

$$P(A_iA_j) = P(A_i)P(A_j); \; P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3).$$

2. Review of Probability Theory: Basics (15)

Let R = "input signal is received at the output". For the event R to occur either switch 1 or switch 2 or switch 3 must remain closed, i.e.,

$$R = A_1 \cup A_2 \cup A_3.$$

Using results in slide 38,

$$P(R) = P(A_1 \cup A_2 \cup A_3) = 1-(1-p)^3 = 3p - 3p^2 + p^3.$$

b. We need $P(\overline{A}_1 \,|\, R).$ From Bayes' theorem

$$P(\overline{A}_1 \,|\, R) = \frac{P(R \,|\, \overline{A}_1)P(\overline{A}_1)}{P(R)} = \frac{(2p-p^2)(1-p)}{3p-3p^2+p^3} = \frac{2-2p+p^2}{3p-3p^2+p^3}.$$

Because of the symmetry of the switches, we also have

$$P(\overline{A}_1 \,|\, R) = P(\overline{A}_2 \,|\, R) = P(\overline{A}_3 \,|\, R).$$

2. Review of Probability Theory: Repeated Trials (1)

❑ Consider two independent experiments with associated probability models (Ω_1, F_1, P_1) and (Ω_2, F_2, P_2). Let $\xi \in \Omega_1$, $\eta \in \Omega_2$ represent elementary events. A joint performance of the two experiments produces an elementary events $\omega = (\xi, \eta)$. **How to characterize an appropriate probability to this "combined event" ?**

Consider the Cartesian product space $\Omega = \Omega_1 \times \Omega_2$ generated from Ω_1 and Ω_2 such that if $\xi \in \Omega_1$ and $\eta \in \Omega_2$, then every ω in Ω is an ordered pair of the form $\omega = (\xi, \eta)$. To arrive at a probability model we need to define the combined trio (Ω, F, P).

Suppose $A \in F_1$ and $B \in F_2$. Then $A \times B$ is the set of all pairs (ξ, η), where $\xi \in A$ and $\eta \in B$. Any such subset of Ω appears to be a legitimate event for the combined experiment. Let F denote the field composed of all such subsets $A \times B$ together with their unions and compliments. In this combined experiment, the probabilities of the events $A \times \Omega_2$ and $\Omega_1 \times B$ are such that

$$P(A \times \Omega_2) = P_1(A), \quad P(\Omega_1 \times B) = P_2(B).$$

2. Review of Probability Theory: Repeated Trials (2)

Moreover, the events $A \times \Omega_2$ and $\Omega_1 \times B$ are independent for any $A \in F_1$ and $B \in F_2$. Since

$$(A \times \Omega_2) \cap (\Omega_1 \times B) = A \times B,$$

then

$$P(A \times B) = P(A \times \Omega_2) \cdot P(\Omega_1 \times B) = P_1(A) P_2(B)$$

for all $A \in F_1$ and $B \in F_2$. This equation extends to a unique probability measure $P(\equiv P_1 \times P_2)$ on the sets in F and defines the combined trio (Ω, F, P).

❑ **Generalization**: Given n experiments $\Omega_1, \Omega_2, \cdots, \Omega_n,$ and their associated F_i and P_i, $i = 1 \to n,$ let

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$$

represent their Cartesian product whose elementary events are the ordered n-tuples $\xi_1, \xi_2, \cdots, \xi_n,$ where $\xi_i \in \Omega_i$. Events in this combined space are of the form

$$A_1 \times A_2 \times \cdots \times A_n$$

where $A_i \in F_i,$

2. Review of Probability Theory: Repeated Trials (3)

If all these n experiments are independent, and $P(A_i)$ is the probability of the event A_i in F_i then as before

$$P(A_1 \times A_2 \times \cdots \times A_n) = P_1(A_1)P_2(A_2)\cdots P_n(A_n).$$

Example: An event A has probability p of occurring in a single trial. Find the probability that A occurs exactly k times, $k \le n$ in n trials.

Solution: Let (Ω, F, P) be the probability model for a single trial. The outcome of n experiments is an n-tuple

$$\omega = \left\{ \xi_1, \xi_2, \cdots, \xi_n \right\} \in \Omega_0,$$

where every $\xi_i \in \Omega$ and $\Omega_0 = \Omega \times \Omega \times \cdots \times \Omega.$ The event A occurs at trial # i , if $\xi_i \in A.$ Suppose A occurs exactly k times in ω. Then k of the ξ_i belong to A, say $\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}$, and the remaining n-k are contained in its compliment in $\overline{A}.$

2. Review of Probability Theory: Repeated Trials (4)

Using independence, the probability of occurrence of such an ω is given by

$$P_0(\omega) = P(\{\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}, \cdots, \xi_{i_n}\}) = P(\{\xi_{i_1}\})P(\{\xi_{i_2}\})\cdots P(\{\xi_{i_k}\})\cdots P(\{\xi_{i_n}\})$$

$$= \underbrace{P(A)P(A)\cdots P(A)}_{k} \; \underbrace{P(\overline{A})P(\overline{A})\cdots P(\overline{A})}_{n-k} = p^k q^{n-k}.$$

However the k occurrences of A can occur in any particular location inside ω. Let $\omega_1, \omega_2, \cdots, \omega_N$ represent all such events in which A occurs exactly k times. Then

$$\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"} = \omega_1 \cup \omega_2 \cup \cdots \cup \omega_N.$$

But, all these ω_i s are mutually exclusive, and equiprobable. Thus

$$P(\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"})$$

$$= \sum_{i=1}^{N} P_0(\omega_i) = NP_0(\omega) = Np^k q^{n-k},$$

2. Review of Probability Theory: Repeated Trials (5)

Recall that, starting with n possible choices, the first object can be chosen n different ways, and for every such choice the second one in (n-1) ways, … and the kth one (n-k+1) ways, and this gives the total choices for k objects out of n to be $n(n\text{-}k)\ldots(n\text{-}k+1)$. But, this includes the $k!$ choices among the k objects that are indistinguishable for identical objects. As a result

$$N = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

represents the number of combinations, or choices of n identical objects taken k at a time. Thus, we obtain **Bernoulli formula**

$$P_n(k) = P(\text{"}A \text{ occurs exactly } k \text{ times in } n \text{ trials"})$$
$$= \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,2,\cdots,n,$$

2. Review of Probability Theory: Repeated Trials (6)

Independent repeated experiments of this nature, where the outcome is either a "success" $(= A)$ or a "failure" $(= \overline{A})$ are characterized as **Bernoulli trials**, and the probability of k successes in n trials is given by Bernoulli formula, where p represents the probability of "success" in any one trial.

❑ **Bernoulli trial**: consists of repeated independent and identical experiments each of which has only two outcomes A or $\overline{A}$ with $P(A) = p,$ and $P(\overline{A}) = q.$ The probability of exactly k occurrences of A in n such trials is given by Bernoulli formula. Let

$$X_k = \text{"exactly } k \text{ occurrences in } n \text{ trials"}.$$

Since the number of occurrences of A in n trials must be an integer $k = 0, 1, 2, \ldots, n,$ either X_0 or X_1 or X_2 or ... or X_n must occur in such an experiment. Thus

$$P(X_0 \cup X_1 \cup \cdots \cup X_n) = 1.$$

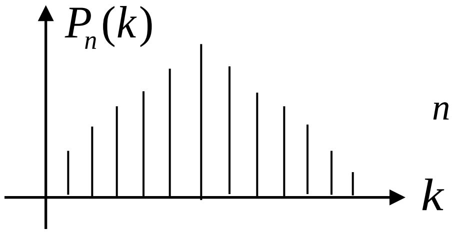
2. Review of Probability Theory: Repeated Trials (7)

But X_i, X_j are mutually exclusive. Thus

$$P(X_0 \cup X_1 \cup \cdots \cup X_n) = \sum_{k=0}^{n} P(X_k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.$$

$$P(X_i \cup \cdots \cup X_j) = \sum_{k=i}^{j} P(X_k) = \sum_{k=i}^{j} \binom{n}{k} p^k q^{n-k}.$$

For a given n and p what is the most likely value of k ?



From Figure, the most probable value of k is that number which maximizes $P_n(k)$ in Bernoulli formula. To obtain this value, consider the ratio

$$\frac{P_n(k-1)}{P_n(k)} = \frac{n!\, p^{k-1} q^{n-k+1}}{(n-k+1)!(k-1)!} \frac{(n-k)!k!}{n!\, p^k q^{n-k}} = \frac{k}{n-k+1} \frac{q}{p}.$$

2. Review of Probability Theory: Repeated Trials (8)

$$\frac{P_n(k-1)}{P_n(k)} = \frac{n!\,p^{k-1}q^{n-k+1}}{(n-k+1)!(k-1)!}\,\frac{(n-k)!k!}{n!\,p^k q^{n-k}} = \frac{k}{n-k+1}\,\frac{q}{p}.$$

Thus, $P_n(k) \geq P_n(k\text{-}1)$, if $k(1\text{-}p) \leq (n\text{-}k\text{+}1)p$ or $k \leq (n\text{+}1)p$. Thus $P_n(k)$ as a function of k increases until

$$k = (n+1)\,p$$

if it is an integer, or the largest integer $k_{\max}$ less than $(n\text{+}1)p$, it represents the most likely number of successes (or heads) in n trials.

Example: Parity check coding.

2. Review of Probability Theory: Repeated Trials (9)

- **Approximate evaluation of** $P(X_i \cup \cdots \cup X_j)$, [4], with $npq >> 1$:

$$P(X_i \cup \cdots \cup X_j) = \sum_{k=i}^{j} \binom{n}{k} p^k q^{n-k} = G\left(\frac{j-np}{\sqrt{npq}}\right) - G\left(\frac{i-np}{\sqrt{npq}}\right)$$

where $G(x)$ is expressed in terms of the **error function**:

$$erf(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy = G(x) - \frac{1}{2}$$

(see the table next slide)

If not only $n >> 1$ but also $np >> 1$ then

$$\sum_{k=0}^{j} \binom{n}{k} p^k q^{n-k} = G\left(\frac{j-np}{\sqrt{npq}}\right)$$

<u>Example</u>: see Example 3-16, p. 52, [4].

2. Review of Probability Theory: Repeated Trials (10)

x	erf(x)	x	erf(x)	x	erf(x)	x	erf(x)
0.05	0.01994	0.80	0.28814	1.55	0.43943	2.30	0.48928
0.10	0.03983	0.85	0.30234	1.60	0.44520	2.35	0.49061
0.15	0.05962	0.90	0.31594	1.65	0.45053	2.40	0.49180
0.20	0.07926	0.95	0.32894	1.70	0.45543	2.45	0.49286
0.25	0.09871	1.00	0.34134	1.75	0.45994	2.50	0.49379
0.30	0.11791	1.05	0.35314	1.80	0.46407	2.55	0.49461
0.35	0.13683	1.10	0.36433	1.85	0.46784	2.60	0.49534
0.40	0.15542	1.15	0.37493	1.90	0.47128	2.65	0.49597
0.45	0.17364	1.20	0.38493	1.95	0.47441	2.70	0.49653
0.50	0.19146	1.25	0.39435	2.00	0.47726	2.75	0.49702
0.55	0.20884	1.30	0.40320	2.05	0.47982	2.80	0.49744
0.60	0.22575	1.35	0.41149	2.10	0.48214	2.85	0.49781
0.65	0.24215	1.40	0.41924	2.15	0.48422	2.90	0.49813
0.70	0.25804	1.45	0.42647	2.20	0.48610	2.95	0.49841
0.75	0.27337	1.50	0.43319	2.25	0.48778	3.00	0.49865

$$\mathrm{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-y^2/2} dy = G(x) - \frac{1}{2}$$

2. Review of Probability Theory: Repeated Trials (11)

<u>Example</u> : A fair coin is tossed 5,000 times. Find the probability that the number of heads is between 2,475 to 2,525.

Solution: We need $P(2{,}475 \leq X \leq 2{,}525)$. Here n is large so that we can use the normal approximation. In this case $p = 0.5$ so that $np = 2{,}500$ and $\sqrt{npq} \approx 35$. Since $np - \sqrt{npq} = 2{,}465$, and $np + \sqrt{npq} = 2{,}535$, the approximation is valid for $k_1 = 2{,}475$ and $k_2 = 2{,}525$. Thus

$$P(k_1 \leq X \leq k_2) = \int_{x_1}^{x_2} \frac{1}{2\pi} e^{-y^2/2} dy.$$

Here

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = -\frac{5}{7}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{5}{7}.$$

Since $x_1 < 0$, from Fig. 5.1(b), the above probability is given by

$$P(2{,}475 \leq X \leq 2{,}525) = \text{erf}(x_2) - \text{erf}(x_1) = \text{erf}(x_2) + \text{erf}(|x_1|)$$

$$= 2\text{erf}\left(\frac{5}{7}\right) = 0.516,$$

2. Review of Probability Theory: Repeated Trials (12)

- **Poisson theorem**: if np is of order of one, $np \to a$, $p \to 0$, $n \to \infty$, then

$$\binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \xrightarrow[n\to\infty]{} e^{-a} \frac{a^k}{k!}$$

Therefore,

$$P(X_i \cup \cdots \cup X_j) = \sum_{k=i}^{j} \binom{n}{k} p^k q^{n-k} \approx e^{-np} \sum_{k=i}^{j} \frac{(np)^k}{k!}$$

<u>Example</u>: see Example 3-21, p. 56, [4].

2. Review of Probability Theory: Repeated Trials (13)

The Poisson Approximation

As we have mentioned earlier, for large n, the Gaussian approximation of a binomial r.v is valid only if p is fixed, i.e., only if $np >> 1$ and $npq >> 1$. What if np is small, or if it does not increase with n?

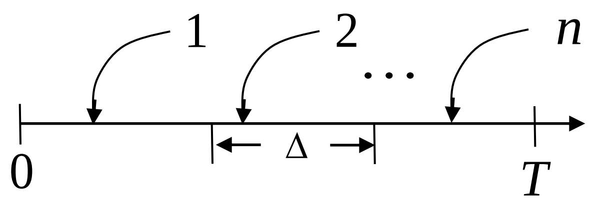
Obviously that is the case if, for example, $p \to 0$ as $n \to \infty$, such that $np = \lambda$ is a fixed number.

2. Review of Probability Theory: Repeated Trials (14)

Many random phenomena in nature in fact follow this pattern: Total number of calls on a telephone line, claims in an insurance company etc. tend to follow this type of behavior.

Consider random arrivals such as telephone calls over a line. Let n represent the total number of calls in the interval $(0, T)$. From our experience, as $T \to \infty$ we have $n \to \infty$ so that we may assume $n = \mu T$.

Consider a small interval of duration Δ as in Figure below. If there is only a single call coming in, the probability p of that single call occurring in that interval must depend on its relative size with respect to T.



2. Review of Probability Theory: Repeated Trials (15)

Hence we may assume $p = \frac{\Delta}{T}$. Note that $p \to 0$ as $T \to \infty$. However in this case

$$np = \mu T \cdot \frac{\Delta}{T} = \mu\Delta = \lambda$$

is a constant, and the normal approximation is invalid here.

Suppose the interval Δ is of interest to us. A call inside that interval is a "success" (H), whereas one outside is a "failure" (T). This is equivalent to the coin tossing situation, and hence the probability of obtaining k calls (in any order) in an interval of duration Δ is given by the binomial p.m.f. Thus

$$P_n(k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}, \qquad (*)$$

and here as $n \to \infty$, $p \to 0$ such that $np = \lambda$. It is easy to obtain an excellent approximation to (*) in that situation.

2. Review of Probability Theory: Repeated Trials (16)

To see this, rewrite (*) as

$$P_n(k) = \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{(np)^k}{k!}(1-np/n)^{n-k}$$

$$= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\frac{\lambda^k}{k!}\frac{(1-\lambda/n)^n}{(1-\lambda/n)^k}.$$

Thus

$$\lim_{n\to\infty,\ p\to 0,\ np=\lambda} P_n(k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad (**)$$

since the finite products $\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)$ as well as $\left(1-\frac{\lambda}{n}\right)^k$ tend to unity as $n\to\infty$, and

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

2. Review of Probability Theory: Repeated Trials (17)

The right side of (**) represents the Poisson p.m.f and the Poisson approximation to the binomial r.v is valid in situations where the binomial r.v parameters *n* and *p* diverge to two extremes $(n \to \infty, p \to 0)$ such that their product *np* is a constant.

<u>Example:</u> Winning a Lottery:

Suppose two million lottery tickets are issued with 100 winning tickets among them. (a) If a person purchases 100 tickets, what is the probability of winning? (b) How many tickets should one buy to be 95% confident of having a winning ticket?

Solution: The probability of buying a winning ticket

$$p = \frac{\text{No. of winning tickets}}{\text{Total no. of tickets}} = \frac{100}{2 \times 10^6} = 5 \times 10^{-5}.$$

Here n = 100 and the number of winning tickets X in the n purchased tickets has an approximate Poisson distribution with parameter

$$\lambda = np = 100 \times 5 \times 10^{-5} = 0.005.$$

2. Review of Probability Theory: Repeated Trials (18)

Thus

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!},$$

and (a) Probability of winning $= P(X \geq 1) = 1 - P(X=0) = 1 - e^{-\lambda} \approx 0.005.$

(b) In this case we need $P(X \geq 1) \geq 0.95.$

$$P(X \geq 1) = 1 - e^{-\lambda} \geq 0.95 \quad \text{implies} \quad \lambda \geq \ln 20 = 3.$$

But $\lambda = np = n \times 5 \times 10^{-5} \geq 3$ or $n \geq 60{,}000.$ Thus one needs to buy about 60,000 tickets to be 95% confident of having a winning ticket!

2. Review of Probability Theory: Repeated Trials (19)

Example: A space craft has 100,000 components $(n \to \infty)$. The probability of any one component being defective is $2 \times 10^{-5} (p \to 0)$. The mission will be in danger if five or more components become defective. Find the probability of such an event.

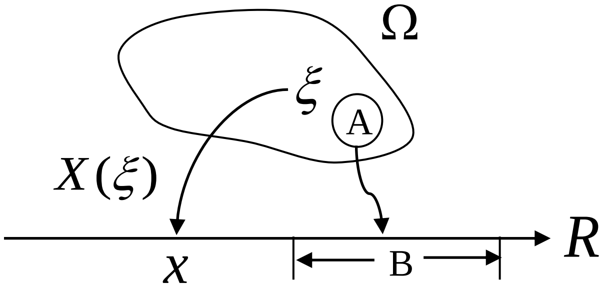
Solution: Here n is large and p is small, and hence Poisson approximation is valid. Thus $np = \lambda = 100{,}000 \times 2 \times 10^{-5} = 2$, and the desired probability is given by

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{k=0}^{4} e^{-\lambda} \frac{\lambda^k}{k!} = 1 - e^{-2} \sum_{k=0}^{4} \frac{\lambda^k}{k!}$$

$$= 1 - e^{-2} \left(1 + 2 + 2 + \frac{4}{3} + \frac{2}{3} \right) = 0.052.$$

2. Review of Probability Theory: Random Variables (1)

- Let (Ω, F, P) be a probability model for an experiment, and X a function that maps every $\zeta \in \Omega$ to a unique point $x \in R$, the set of real numbers. Since the outcome ζ is not certain, so is the value $X(\zeta) = x$. Thus if B is some subset of R, we may want to determine the probability of "$X(\zeta) \in B$". To determine this probability, we can look at the set $A = X^{-1}(B) \in \Omega$ that contains all $\zeta \in \Omega$ that maps into B under the function X.



Obviously, if the set $A = X^{-1}(B)$ also belongs to the associated field F, then it is an event and the probability of A is well defined; in that case we can say

$$\text{Probability of the event } "X(\xi) \in B" = P(X^{-1}(B)).$$

2. Review of Probability Theory: R. V (2)

However, $X^{-1}(B)$ may not always belong to F for all B, thus creating difficulties. The notion of **random variable** (r.v) makes sure that the inverse mapping always results in an event so that we are able to determine the probability for any $B \in R$.

❑ **Random Variable (r.v)**: A finite single valued function $X(.)$ that maps the set of all experimental outcomes Ω into the set of real numbers R is said to be a r.v, if the set $\{ \zeta \mid X(\zeta) \le x \}$ is an event $\in F$ for every x in R.

Alternatively, X is said to be a r.v, if $X^{-1}(B) \in F$ where B represents semi-definite intervals of the form $\{-\infty < x \le a\}$ and all other sets that can be constructed from these sets by performing the set operations of union, intersection and negation any number of times. Thus if X is a r.v, then

$$\{ \xi \mid X(\xi) \le x \} = \{ X \le x \}$$

is an event for every x. (see Example 4-2, p. 65, [4])

2. Review of Probability Theory: R. V (3)

What about $\{ a < X \le b \}$, $\{ X = a \}$? Are they also events ?

In fact with $b > a$, since $\{ X \le a \}$ and $\{ X \le b \}$ are events, $\{ X \le a \}^c = \{ X > a \}$ is an event and hence

$$\{ X > a \} \cap \{ X \le b \} = \{ a < X \le b \},$$

is also an event. Thus, $\{ a - 1/n < X \le a \}$, is an event for every n.

Consequently,

$$\bigcap_{n=1}^{\infty} \left\{ a - \frac{1}{n} < X \le a \right\} = \{X = a\}$$

is also an event. All events have well defined probability. Thus the probability of the event $\{ \zeta \mid X(\zeta) \le x \}$ must depend on x. Denote

$$P\left\{ \xi \mid X(\xi) \le x \right\} = F_X(x) \ge 0$$

which is referred to as the **Probability Distribution Function** (**PDF**) associated with the r.v X. (see Example 4-4, p. 67, [4])

2. Review of Probability Theory: R. V (4)

- **Distribution Function**: If $g(x)$ is a distribution function, then

  (i) $g(+\infty) = 1$; $g(-\infty) = 0$
  (ii) if $x_1 < x_2$, then $g(x_1) \le g(x_2)$
  (iii) $g(x^+) = g(x)$ for all x.

It is shown that the PDF $F_X(x)$ satisfies these properties for any r.v X.

Additional Properties of a PDF

(iv) if $F_X(x_0) = 0$ for some x_0, then $F_X(x) = 0$ for $x \le x_0$

(v) $P\{\, X(\xi) > x \,\} = 1 - F_X(x).$

(vi) $P\{\, x_1 < X(\xi) \le x_2 \,\} = F_X(x_2) - F_X(x_1), \quad x_2 > x_1.$

(vii) $P\big(X(\xi) = x\big) = F_X(x) - F_X(x^-).$

2. Review of Probability Theory: R. V (5)

X is said to be a continuous-type r.v if its distribution function $F_X(x)$ is continuous. In that case $F_X(x^-) = F_X(x)$ for all x, and from property (vii) we get $P\{X = x\} = 0$.

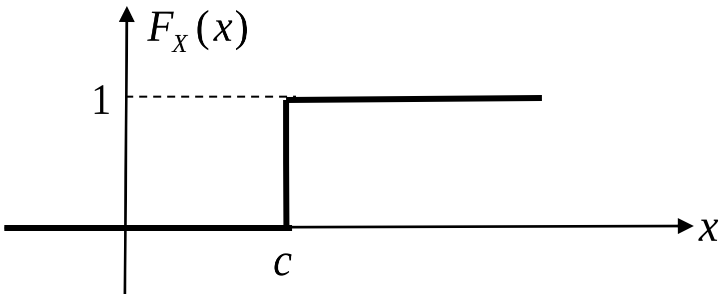
If $F_X(x)$ is constant except for a finite number of jump discontinuities (piece-wise constant; step-type), then X is said to be a discrete-type r.v. If x_i is such a discontinuity point, then from (vii)

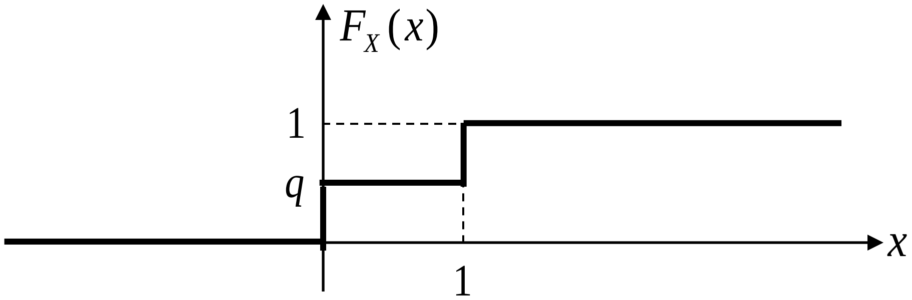
$$p_i = P\{X = x_i\} = F_X(x_i) - F_X(x_i^-).$$

2. Review of Probability Theory: R. V (6)

Example: X is a r.v such that $X(\zeta) = c$, $\zeta \in \Omega$. Find $F_X(x)$.

Solution: For $x < c$, $\{X(\zeta) \le x\} = \{\phi\}$, so that $F_X(x) = 0$, and for $x > c$, $\{X(\zeta) \le x\} = \Omega$, so that $F_X(x) = 1$.





Example: Toss a coin. $\Omega = \{H, T\}$. Suppose the r.v X is such that $X(T) = 0$, $X(H) = 1$. Find $F_X(x)$.

Solution: For $x < 0$, $\{X(\zeta) \le x\} = \{\phi\}$, so that $F_X(x) = 0$.

$$0 \le x < 1, \ \{X(\xi) \le x\} = \{T\}, \ \text{so that } F_X(x) = P\{T\} = 1 - p,$$
$$x \ge 1, \ \{X(\xi) \le x\} = \{H, T\} = \Omega, \ \text{so that } F_X(x) = 1.$$

2. Review of Probability Theory: R. V (7)

Example: A fair coin is tossed twice, and let the r.v X represent the number of heads. Find $F_X(x)$.

Solution: In this case $\Omega = \{HH, HT, TH, TT\}$, and $X(HH) = 2$, $X(HT) = X(TH) = 1$, $X(TT) = 0$.

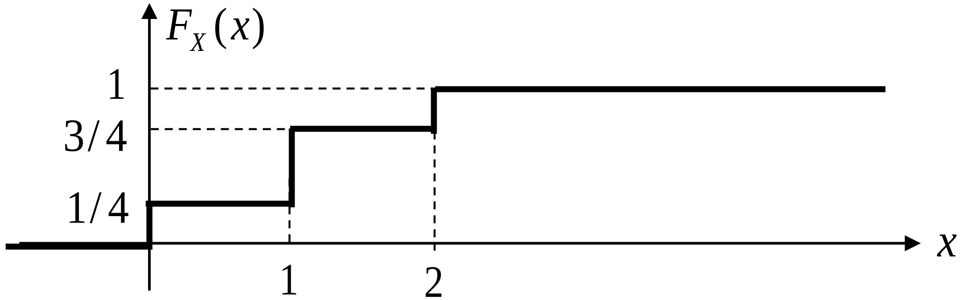
$$x < 0, \quad \{X(\xi) \le x\} = \phi \Rightarrow F_X(x) = 0,$$

$$0 \le x < 1, \{X(\xi) \le x\} = \{TT\} \Rightarrow F_X(x) = P\{TT\} = P(T)P(T) = \frac{1}{4},$$

$$1 \le x < 2, \{X(\xi) \le x\} = \{TT, HT, TH\} \Rightarrow F_X(x) = P\{TT, HT, TH\} = \frac{3}{4},$$

$$x \ge 2, \quad \{X(\xi) \le x\} = \Omega \Rightarrow F_X(x) = 1.$$

$$P\{X = 1\} = F_X(1) - F_X(1^-) = 3/4 - 1/4 = 1/2.$$



2. Review of Probability Theory: R. V (8)

❑ **Probability Density Function (p.d.f)**: The derivative of the distribution function $F_X(x)$ is called the **probability density function** $f_X(x)$ of the r.v X. Thus

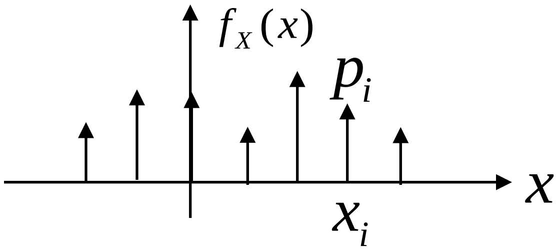
$$f_X(x) = \frac{dF_X(x)}{dx}.$$

Since

$$\frac{dF_X(x)}{dx} = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} \geq 0,$$

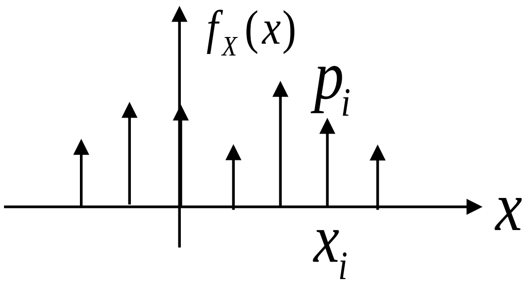
from the monotone-nondecreasing nature of $F_X(x)$, it follows that $f_X(x) \geq 0$ for all x. $f_X(x)$ will be a continuous function, if X is a continuous type r.v. However, if X is a discrete type r.v, then its p.d.f has the general form

$$f_X(x) = \sum_i p_i \delta(x - x_i),$$



2. Review of Probability Theory: R. V (9)

where x_i represent the jump-discontinuity points in $F_X(x)$. In the figure, $f_X(x)$ represents a collection of positive discrete masses, and it is known as the probability mass function (p.m.f) in the discrete case.



We also obtain

$$F_X(x) = \int_{-\infty}^{x} f_x(u)du.$$

Since $F_X(+\infty) = 1$, it yields

$$\int_{-\infty}^{+\infty} f_x(x)dx = 1,$$

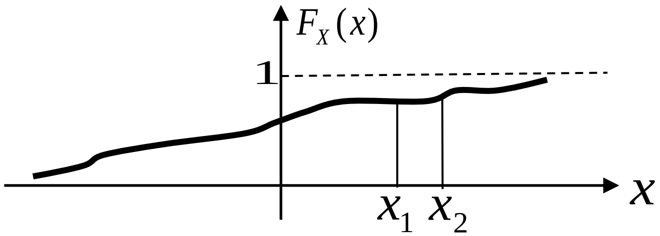
which justifies its name as the density function. Further, we also get

$$P\{ x_1 < X(\xi) \le x_2 \} = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)dx.$$

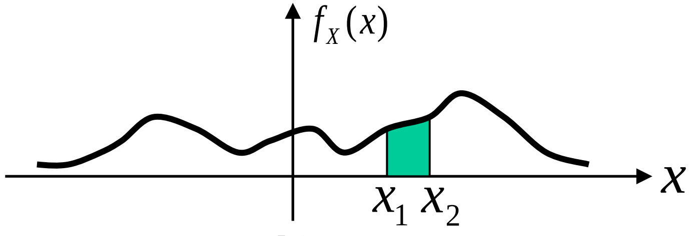
2. Review of Probability Theory: R. V (10)

Thus the area under $f_X(x)$ in the interval (x_1, x_2) represents the probability in the equation

$$P\left\{ x_1 < X(\xi) \le x_2 \right\} = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)dx.$$



(a)



(b)

Often, r.vs are referred by their specific density functions - both in the continuous and discrete cases - and in what follows we shall list a number of them in each category.

<u>Example</u>: Transmitting a 3-digit message over a noisy channel

2. Review of Probability Theory: R. V (11)

❑ Continuous-type Random Variables

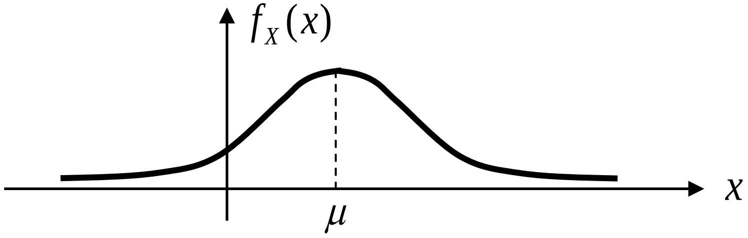
1. *Normal (Gaussian):* X is said to be normal or Gaussian r.v, if

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

This is a bell shaped curve, symmetric around the parameter μ and its distribution function is given by

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} dy = G\left(\frac{x-\mu}{\sigma}\right),$$

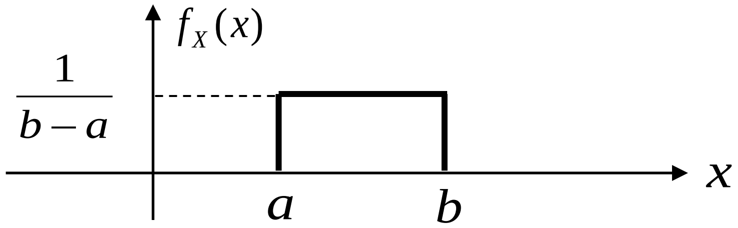
where $G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ is often tabulated. Since $f_X(x)$ depends on two parameters μ and σ^2, the notation $X \sim N(\mu, \sigma^2)$ will be used to represent $f_X(x)$.



2. Review of Probability Theory: R. V (12)

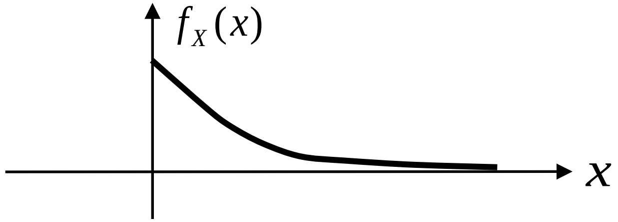
2. *Uniform*: $X \sim U(a,b)$, $a < b$, if

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise.} \end{cases}$$



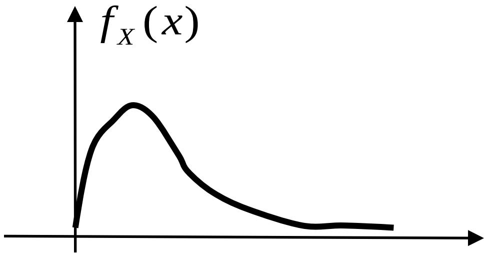
3. *Exponential*: $X \sim \varepsilon(\lambda)$, if

$$f_X(x) = \begin{cases} \dfrac{1}{\lambda} e^{-x/\lambda}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$



4. *Gamma*: $X \sim G(\alpha, \beta)$, if $\alpha > 0$, $\beta > 0$

$$f_X(x) = \begin{cases} \dfrac{x^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} e^{-x/\beta}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$

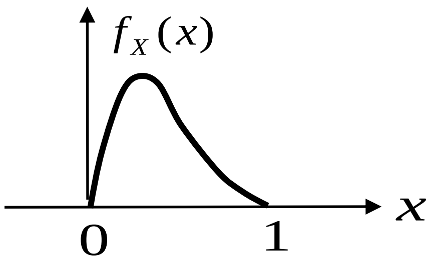


If $\alpha = n$ an integer, $\Gamma(n) = (n\text{-}1)!$

2. Review of Probability Theory: R. V (13)

5. <u>*Beta*</u>: $X \sim \beta(a,b)$ if $(a>0,\ b>0)$

$$f_X(x) = \begin{cases} \dfrac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}, & 0<x<1, \\ 0, & \text{otherwise.} \end{cases}$$

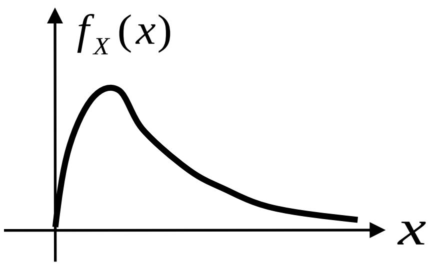


where the Beta function $\beta(a,b)$ is defined as

$$\beta(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1}du.$$

6. <u>*Chi-Square*</u>: $X \sim \chi^2(n)$ if

$$f_X(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1}e^{-x/2}, & x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$

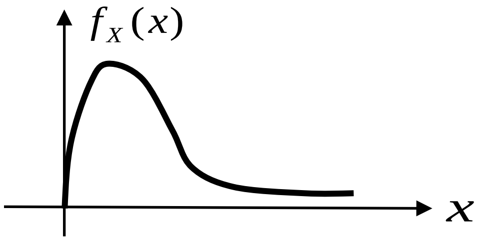


Note that $\chi^2(n)$ is the same as Gamma $(n/2, 2)$

2. Review of Probability Theory: R. V (14)

7. <u>*Rayleigh*</u>: $X \sim R(\sigma^2)$, if

$$f_X(x) = \begin{cases} \dfrac{x}{\sigma^2} e^{-x^2/2\sigma^2}, & x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

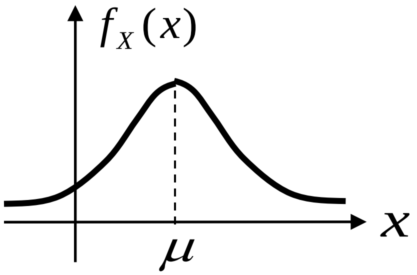


8. <u>*Nakagami – m distribution*</u>:

$$f_X(x) = \begin{cases} \dfrac{2}{\Gamma(m)} \left(\dfrac{m}{\Omega}\right)^m x^{2m-1} e^{-mx^2/\Omega}, & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

9. <u>*Cauchy*</u>: $X \sim C(\alpha, \mu)$, if

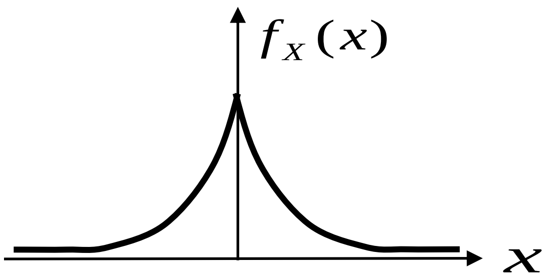
$$f_X(x) = \frac{\alpha/\pi}{\alpha^2 + (x-\mu)^2}, \quad -\infty < x < +\infty.$$



2. Review of Probability Theory: R. V (15)

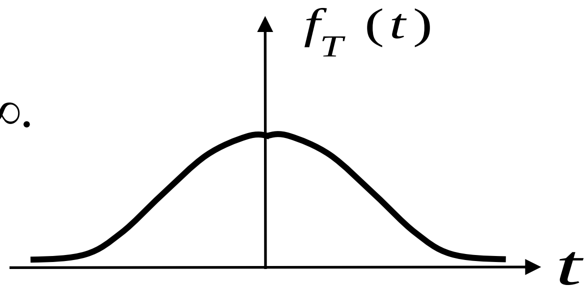
10. <u>*Laplace*</u>:

$$f_X(x) = \frac{1}{2\lambda} e^{-|x|/\lambda}, \quad -\infty < x < +\infty.$$



11. <u>*Student's t-distribution with n degrees of freedom*</u>:

$$f_T(t) = \frac{\Gamma\left((n+1)/2\right)}{\sqrt{\pi n}\,\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad -\infty < t < +\infty.$$



12. <u>*Fisher's F-distribution*</u>:

$$f_z(z) = \begin{cases} \dfrac{\Gamma\{(m+n)/2\}\, m^{m/2} n^{n/2}}{\Gamma(m/2)\,\Gamma(n/2)} \dfrac{z^{m/2-1}}{(n+mz)^{(m+n)/2}}, & z \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

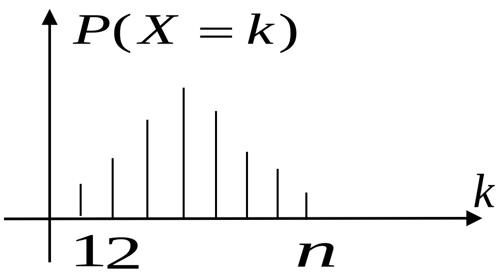
❑ Discrete-type Random Variables

1. *Bernoulli*: X takes the values (0,1), and

$$P(X=0)=q, \quad P(X=1)=p.$$

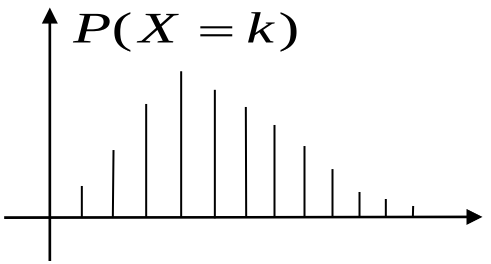
2. *Binomial*: $X \sim B(n,p)$ if

$$P(X=k)=\binom{n}{k}p^k q^{n-k}, \quad k=0,1,2,\cdots,n.$$



3. *Poisson*: $X \sim P(\lambda)$ if

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}, \quad k=0,1,2,\cdots,\infty.$$



4. *Discrete-Uniform*

$$P(X=k)=\frac{1}{N}, \quad k=1,2,\cdots,N.$$

2. Review of Probability Theory: Function of R. V (1)

❑ Let X be a r.v defined on the model (Ω, F, P) and suppose $g(x)$ is a function of the variable x. Define

$$Y = g(X)$$

Is Y necessarily a r.v ? If so what is its PDF $F_Y(y)$ and pdf $f_Y(y)$?

Consider some of the following functions to illustrate the technical details.

<u>Example 1</u>: $Y = aX + b$

- Suppose $a > 0$,

$$F_Y(y) = P(Y(\xi) \le y) = P(aX(\xi) + b \le y) = P\left(X(\xi) \le \frac{y-b}{a} \right) = F_X\left(\frac{y-b}{a} \right).$$

and

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a} \right).$$

2. Review of Probability Theory: Function of R. V (2)

- If $a < 0$, then

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P\big(aX(\xi) + b \le y\big) = P\left(X(\xi) > \frac{y-b}{a} \right)$$

$$= 1 - F_X\left(\frac{y-b}{a} \right),$$

and hence

$$f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a} \right).$$

- For all a

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a} \right).$$

2. Review of Probability Theory: Function of R. V (3)

Example 2: $Y = X^2$

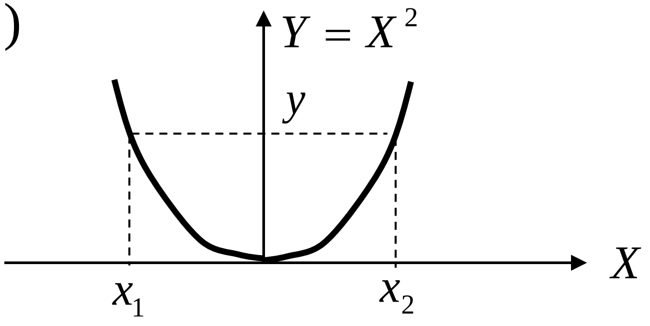
$$F_Y(y) = P(Y(\xi) \le y) = P(X^2(\xi) \le y).$$

If $y < 0$ then the event $\{X^2(\zeta) \le y\} = \phi$, and hence

$$F_Y(y) = 0, \quad y < 0.$$

For $y > 0$, from figure, the event $\{Y(\zeta) \le y\} = \{X^2(\zeta) \le y\}$ is equivalent to $\{x_1 \le X(\zeta) \le x_2\}$. Hence

$$F_Y(y) = P(x_1 < X(\xi) \le x_2) = F_X(x_2) - F_X(x_1)$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \quad y > 0.$$



By direct differentiation, we get

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right), & y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

2. Review of Probability Theory: Function of R. V (4)

If $f_X(x)$ represents an even function, then $f_Y(y)$ reduces to

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X\left(\sqrt{y}\right) U(y).$$

In particular, if $X \sim N(0,1)$, so that

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

then, we obtain the p.d.f of $Y = X^2$ to be

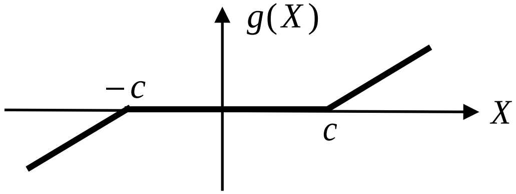
$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} U(y).$$

we notice that this equation represents a Chi-square r.v with $n = 1$, since $\Gamma(1/2) = \sqrt{\pi}$.

Thus, if X is a Gaussian r.v with $\mu = 0$, *then* $Y = X^2$ *represents a Chi-square r.v with one degree of freedom (n = 1).*

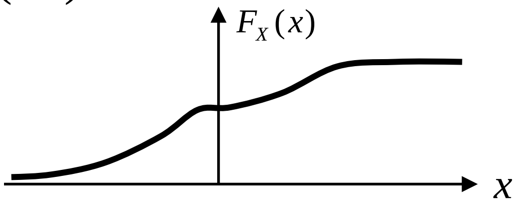
Example 3: Let

$$Y = g(X) = \begin{cases} X - c, & X > c, \\ 0, & -c < X \le c, \\ X + c, & X \le -c. \end{cases}$$



In this case

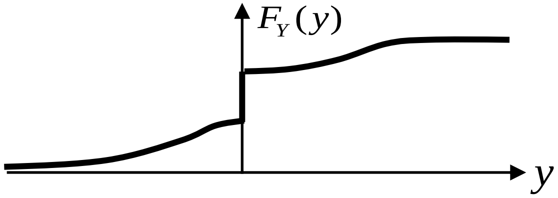
$$P(Y = 0) = P(-c < X(\xi) \le c) = F_X(c) - F_X(-c).$$



For $y > 0$, we have $x > c$, and $Y(\zeta) = X(\zeta) - c$, so that

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P(X(\xi) - c \le y)$$
$$= P\big(X(\xi) \le y + c\big) = F_X(y + c), \quad y > 0.$$

Similarly $y < 0$, if $x < -c$, and $Y(\zeta) = X(\zeta) + c$, so that



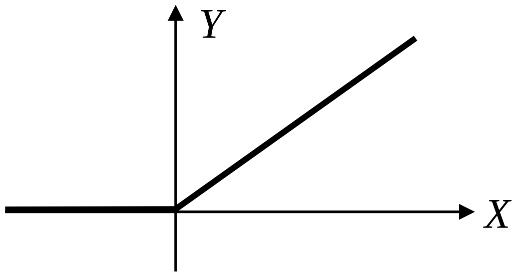
$$F_Y(y) = P\big(Y(\xi) \le y\big) = P(X(\xi) + c \le y)$$
$$= P\big(X(\xi) \le y - c\big) = F_X(y - c), \quad y < 0.$$

Thus

$$f_Y(y) = \begin{cases} f_X(y + c), & y > 0, \\ [F_X(c) - F_X(-c)]\delta(y), \\ f_X(y - c), & y < 0. \end{cases}$$

2. Review of Probability Theory: Function of R. V (6)

Example 4: Half-wave rectifier



$$Y = g(X); \quad g(x) = \begin{cases} x, & x > 0, \\ 0, & x \le 0. \end{cases}$$

In this case

$$P(Y = 0) = P(X(\xi) \le 0) = F_X(0).$$

and for $y > 0$, since $Y = X$

$$F_Y(y) = P\big(Y(\xi) \le y\big) = P\big(X(\xi) \le y\big) = F_X(y).$$

Thus

$$f_Y(y) = \begin{cases} f_X(y), & y > 0, \\ F_X(0)\delta(y) & y = 0, \\ 0, & y < 0, \end{cases} \quad = \quad f_X(y)U(y) + F_X(0)\delta(y).$$

2. Review of Probability Theory: Function of R. V (7)

- Note: As a general approach, given $Y = g(X)$, first sketch the graph $y = g(x)$, and determine the range space of y. Suppose $a < y < b$ is the range space of $y = g(x)$. Then clearly for $y < a$, $F_Y(y) = 0$, and for $y > b$, $F_Y(y) = 1$, so that $F_Y(y)$ can be nonzero only in $a < y < b$. Next, determine whether there are discontinuities in the range space of y. If so, evaluate $P(Y(\zeta) = y_i)$ at these discontinuities. In the continuous region of y, use the basic approach

$$F_Y(y) = P\big(g(X(\xi)) \leq y\big)$$

and determine appropriate events in terms of the r.v X for every y. Finally, we must have $F_Y(y)$ for $-\infty < y < +\infty$ and obtain

$$f_Y(y) = \frac{dF_Y(y)}{dy} \quad \text{in} \quad a < y < b.$$

2. Review of Probability Theory: Function of R. V (8)

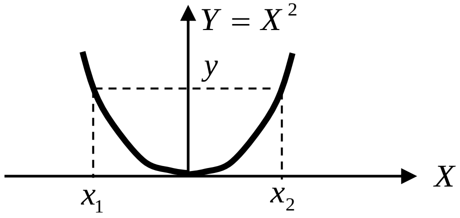
❑ However, if $Y = g(X)$ is a *continuous function*, it is easy to obtain $f_Y(y)$ directly as

$$f_Y(y) = \sum_i \frac{1}{\left| dy/dx \right|_{x_i}} f_X(x_i) = \sum_i \frac{1}{\left| g'(x_i) \right|} f_X(x_i). \qquad (\blacklozenge)$$

The summation index i in this equation depends on y, and for every y the equation $y = g(x_i)$ must be solved to obtain the total number of solutions at every y, and the actual solutions $x_1, x_2, \ldots$ all in terms of y.

For example, if $Y = X^2$, then for all $y > 0$, $x_1 = -\sqrt{y}$ and $x_2 = +\sqrt{y}$ represent the two solutions for each y. Notice that the solutions x_i are all in terms of y so that the right side of $(\blacklozenge)$ is only a function of y. Moreover

$$\frac{dy}{dx} = 2x \quad \text{so that} \quad \left| \frac{dy}{dx} \right|_{x=x_i} = 2\sqrt{y}$$



2. Review of Probability Theory: Function of R. V (9)

Using (♦), we obtain

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right), & y > 0, \\ 0, & \text{otherwise}, \end{cases}$$

which agrees with the result in Example 2.

<u>Example 5</u>: $Y = 1/X$. Find $f_Y(y)$

Here for every y, $x_1 = 1/y$ is the only solution, and

$$\frac{dy}{dx} = -\frac{1}{x^2} \quad \text{so that} \quad \left|\frac{dy}{dx}\right|_{x=x_1} = \frac{1}{1/y^2} = y^2,$$

Then, from (♦), we obtain

$$f_Y(y) = \frac{1}{y^2} f_X\left(\frac{1}{y}\right).$$

2. Review of Probability Theory: Function of R. V (10)

In particular, suppose X is a Cauchy r.v with parameter α so that

$$f_X(x) = \frac{\alpha/\pi}{\alpha^2 + x^2}, \quad -\infty < x < +\infty.$$

In this case, $Y = 1/X$ has the p.d.f

$$f_Y(y) = \frac{1}{y^2} \frac{\alpha/\pi}{\alpha^2 + (1/y)^2} = \frac{(1/\alpha)/\pi}{(1/\alpha)^2 + y^2}, \quad -\infty < y < +\infty.$$

But this represents the p.d.f of a Cauchy r.v with parameter $(1/\alpha)$. Thus if

$X \sim C(\alpha)$, then $1/X \sim C(1/\alpha)$.

<u>Example 6</u>: Suppose $f_X(x) = 2x/\pi^2, \; 0 < x < \pi$ and $Y = \sin X$. Determine $f_Y(y)$.

Since X has zero probability of falling outside the interval $(0, \pi)$, $y = \sin x$ has zero probability of falling outside the interval $(0, 1)$. Clearly $f_Y(y) = 0$ outside this interval.

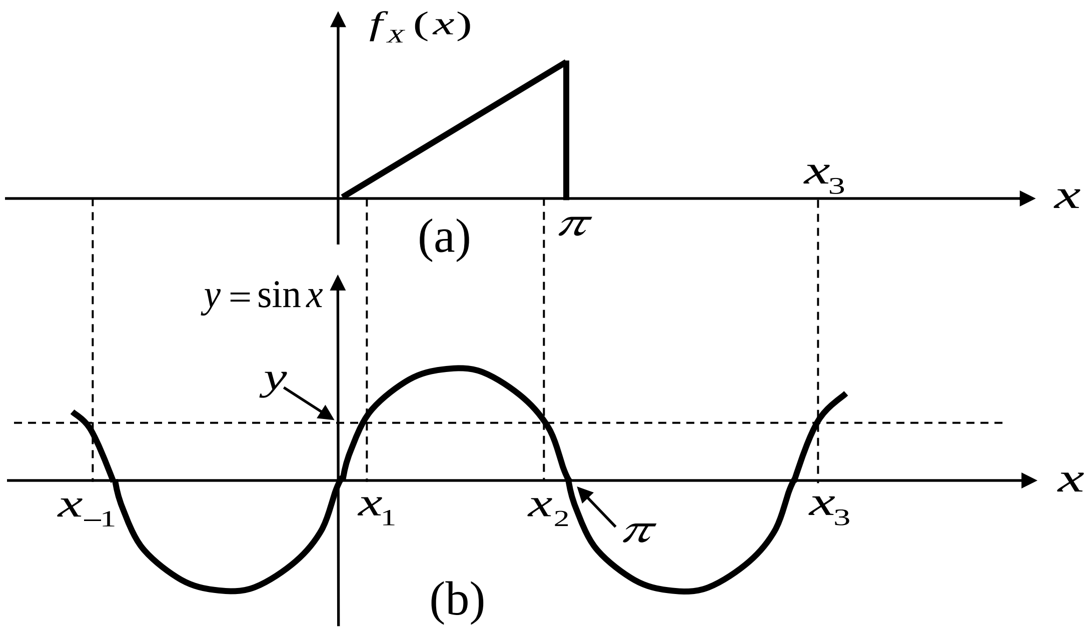
2. Review of Probability Theory: Function of R. V (11)

For any $0 < y < 1$, the equation $y = \sin x$ has an infinite number of solutions $\ldots x_1, x_2, x_3, \ldots$ (see the Figure below), where $x_1 = \sin^{-1} y$ is the principal solution. Moreover, using the symmetry we also get $x_2 = \pi - x_1$ etc. Further,

$$\frac{dy}{dx} = \cos x = \sqrt{1 - \sin^2 x} = \sqrt{1 - y^2}$$

so that

$$\left| \frac{dy}{dx} \right|_{x = x_i} = \sqrt{1 - y^2}.$$



2. Review of Probability Theory: Function of R. V (12)

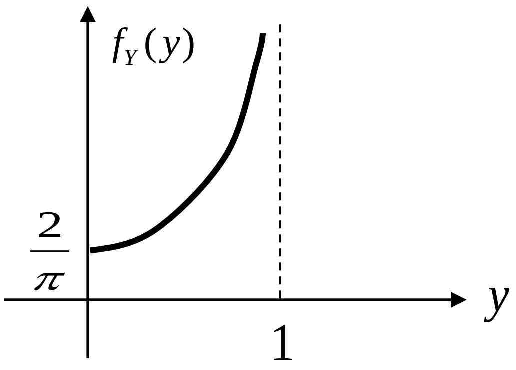
From (♦), we obtain for $0 < y < 1$

$$f_Y(y) = \sum_{\substack{i=-\infty \\ i \neq 0}}^{+\infty} \frac{1}{\sqrt{1-y^2}} f_X(x_i).$$

But from the figure, in this case $f_X(-x_1) = f_X(x_3) = f_X(x_4) = \ldots = 0$ (Except for $f_X(x_1)$ and $f_X(x_2)$ the rest are all zeros). Thus

$$f_Y(y) = \frac{1}{\sqrt{1-y^2}}\left(f_X(x_1) + f_X(x_2)\right) = \frac{1}{\sqrt{1-y^2}}\left(\frac{2x_1}{\pi^2} + \frac{2x_2}{\pi^2}\right)$$

$$= \frac{2(x_1 + \pi - x_1)}{\pi^2\sqrt{1-y^2}} = \begin{cases} \dfrac{2}{\pi\sqrt{1-y^2}}, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$



2. Review of Probability Theory: Function of R. V (13)

❑ Functions of a discrete-type r.v

Suppose X is a discrete-type r.v with $P(X = x_i) = p_i$, $x = x_1, x_2, \ldots, x_i, \ldots$
and $Y = g(X)$. Clearly Y is also of discrete-type, and when $x = x_i$, $y_i = g(x_i)$ and for those y_i , $P(Y = y_i) = P(X = x_i) = p_i$, $y = y_1, y_2, \ldots, y_i, \ldots$

Example 7: Suppose $X \sim P(\lambda)$, so that

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0,1,2,\cdots$$

Define $Y = X^2 + 1$. Find the p.m.f of Y.

X takes the values 0, 1, 2,..., k, ... so that Y only takes the value 1, 2, 5,..., k^2+1,... and $P(Y = k^2+1) = P(X = k)$ so that for $j = k^2+1$

$$P(Y = j) = P\left(X = \sqrt{j-1}\right) = e^{-\lambda}\frac{\lambda^{\sqrt{j-1}}}{(\sqrt{j-1})!}, \quad j = 1, 2, 5, \cdots, k^2+1, \cdots.$$

2. Review of Probability Theory: Moments… (1)

❑ **Mean** or the **Expected Value** of a r.v X is defined as

$$\eta_X = \overline{X} = E(X) = \int_{-\infty}^{+\infty} x\, f_X(x) dx.$$

If X is a discrete-type r.v, then we get

$$\eta_X = \overline{X} = E(X) = \int x \sum_i p_i \delta(x - x_i) dx = \sum_i x_i p_i \underbrace{\int \delta(x - x_i) dx}_{1}$$

$$= \sum_i x_i p_i = \sum_i x_i \; P(X = x_i)\,.$$

Mean represents the average (mean) value of the r.v in a very large number of trials. For example if $X \sim U(a,b)$, then

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left.\frac{x^2}{2}\right|_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

is the midpoint of the interval (a,b).

2. Review of Probability Theory: Moments… (2)

On the other hand if X is exponential with parameter λ, then

$$E(X) = \int_0^\infty \frac{x}{\lambda} e^{-x/\lambda} dx = \lambda \int_0^\infty y e^{-y} dy = \lambda,$$

implying that the parameter λ represents the mean value of the exponential r.v.

Similarly if X is Poisson with parameter λ , we get

$$E(X) = \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Thus the parameter λ also represents the mean of the Poisson r.v.

2. Review of Probability Theory: Moments… (3)

In a similar manner, if X is binomial, then its mean is given by

$$E(X)=\sum_{k=0}^{n} kP(X=k)=\sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k}=\sum_{k=1}^{n} k\frac{n!}{(n-k)!k!}p^k q^{n-k}$$

$$=\sum_{k=1}^{n}\frac{n!}{(n-k)!(k-1)!}p^k q^{n-k}=np\sum_{i=0}^{n-1}\frac{(n-1)!}{(n-i-1)!i!}p^i q^{n-i-1}=np(p+q)^{n-1}=np.$$

Thus np represents the mean of the binomial r.v.

For the normal r.v,

$$E(X)=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} x e^{-(x-\mu)^2/2\sigma^2}dx=\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty}(y+\mu)e^{-y^2/2\sigma^2}dy$$

$$=\frac{1}{\sqrt{2\pi\sigma^2}}\underbrace{\int_{-\infty}^{+\infty} y e^{-y^2/2\sigma^2}dy}_{0}+\mu\cdot\underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} e^{-y^2/2\sigma^2}dy}_{1}=\mu.$$

2. Review of Probability Theory: Moments… (4)

Given $X \sim f_X(x)$, suppose $Y = g(X)$ defines a new r.v with p.d.f $f_Y(y)$, the new r.v Y has a mean μ_Y given by

$$\mu_Y = E(Y) = E\big(g(X)\big) = \int_{-\infty}^{+\infty} y\, f_Y(y)dy = \int_{-\infty}^{+\infty} g(x) f_X(x)dx.$$

In the discrete case,

$$E(Y) = \sum_i g(x_i)P(X = x_i).$$

From the equations, $f_Y(y)$ is not required to evaluate $E(Y)$ for $Y=g(X)$.

2. Review of Probability Theory: Moments… (5)

Example: Determine the mean of $Y = X^2$ where X is a Poisson r.v.

$$E\left(X^2\right) = \sum_{k=0}^{\infty} k^2 P(X = k) = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{i=0}^{\infty} (i+1) \frac{\lambda^{i+1}}{i!}$$

$$= \lambda e^{-\lambda} \left(\sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} + \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \right) = \lambda e^{-\lambda} \left(\sum_{i=1}^{\infty} i \frac{\lambda^i}{i!} + e^{\lambda} \right)$$

$$= \lambda e^{-\lambda} \left(\sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} + e^{\lambda} \right) = \lambda e^{-\lambda} \left(\sum_{m=0}^{\infty} \frac{\lambda^{m+1}}{m!} + e^{\lambda} \right)$$

$$= \lambda e^{-\lambda} \left(\lambda e^{\lambda} + e^{\lambda} \right) = \lambda^2 + \lambda.$$

In general, $E(X^k)$ is known as the kth **moment** of r.v X. Thus if $X \sim P(\lambda)$, its second moment is given by the above equation.

2. Review of Probability Theory: Moments… (6)

- For a r.v X with mean μ , $X - \mu$ represents the deviation of the r.v from its mean. Since this deviation can be either positive or negative, consider the quantity $(X - \mu)^2$ and its average value $E[(X - \mu)^2]$ represents the average mean square deviation of X around its mean. Define

$$\sigma_X^2 = E[(X - \mu)^2] > 0.$$

With $g(X) = (X - \mu)^2$, we get

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x)dx > 0.$$

where σ_X^2 is known as the **variance** of the r.v X, and its square root $\sigma_X = \sqrt{E(X - \mu)^2}$ is known as the **standard deviation** of X. Note that the standard deviation represents the root mean square spread of the r.v X around its mean μ .

2. Review of Probability Theory: Moments… (7)

Alternatively, the variance can be calculated by

$$Var(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} \left(x^2 - 2x\mu + \mu^2\right) f_X(x)dx$$

$$= \int_{-\infty}^{+\infty} x^2 f_X(x)dx - 2\mu \int_{-\infty}^{+\infty} x \; f_X(x)dx + \mu^2$$

$$= E\left(X^2\right) - \mu^2 = E\left(X^2\right) - \left[E(X)\right]^2 = \overline{X^2} - \overline{X}^2.$$

Example: Determine the variance of Poisson r.v

$$\sigma_X^2 = \overline{X^2} - \overline{X}^2 = \left(\lambda^2 + \lambda\right) - \lambda^2 = \lambda.$$

Thus for a Poisson r.v, mean and variance are both equal to its parameter λ

2. Review of Probability Theory: Moments… (8)

Example: Determine the variance of the normal r.v $N(\mu\ ,\ \sigma^2\ )$

We have

$$Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx.$$

Use of the identity

$$\int_{-\infty}^{+\infty} f_X(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = 1$$

for a normal p.d.f. This gives

$$\int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} dx = \sqrt{2\pi}\sigma.$$

Differentiating both sides with respect to σ, we get

$$\int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sigma^3} e^{-(x-\mu)^2/2\sigma^2} dx = \sqrt{2\pi}$$

or

$$\int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx = \sigma^2,$$

2. Review of Probability Theory: Moments… (9)

❑ **Moments**: In general

$$m_n = \overline{X^n} = E(X^n), \quad n \geq 1$$

are known as the **moments** of the r.v X, and

$$\mu_n = E[(X-\mu)^n]$$

are known as the **central moments** of X. Clearly, the mean $\mu = m_1$, and the variance $\sigma 2 = \mu_2$. It is easy to relate m_n and μ_n. Infact

$$\mu_n = E[(X-\mu)^n] = E\left(\sum_{k=0}^{n}\binom{n}{k}X^k(-\mu)^{n-k}\right)$$

$$= \sum_{k=0}^{n}\binom{n}{k}E\left(X^k\right)(-\mu)^{n-k} = \sum_{k=0}^{n}\binom{n}{k} m_k (-\mu)^{n-k}.$$

In general, the quantities $E[(X - a)^n]$ are known as the generalized moments of X about a, and $E[|X|^n]$ are known as the absolute moments of X.

2. Review of Probability Theory: Moments… (10)

❑ **Characteristic Function** of a r.v X is defined as

$$\Phi_X(\omega) = E\left(e^{jX\omega}\right) = \int_{-\infty}^{+\infty} e^{jx\omega} f_X(x)dx.$$

Thus $\Phi_X(0) = 1$, and $\left|\Phi_X(\omega)\right| \le 1$ for all ω.

For discrete r.vs the characteristic function reduces to

$$\Phi_X(\omega) = \sum_k e^{jk\omega} P(X = k).$$

<u>Example</u>: If $X \sim P(\lambda)$, then its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} e^{jk\omega} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{j\omega})^k}{k!} = e^{-\lambda} e^{\lambda e^{j\omega}} = e^{\lambda(e^{j\omega}-1)}.$$

<u>Example</u>: If X is a binomial r.v, its characteristic function is given by

$$\Phi_X(\omega) = \sum_{k=0}^{n} e^{jk\omega} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^{j\omega})^k q^{n-k} = (pe^{j\omega} + q)^n.$$

2. Review of Probability Theory: Moments… (11)

The characteristic function can be used to compute the mean, variance and other higher order moments of any r.v X.

$$\Phi_X(\omega) = E\left(e^{jX\omega}\right) = E\left[\sum_{k=0}^{\infty} \frac{(j\omega X)^k}{k!}\right] = \sum_{k=0}^{\infty} j^k \frac{E(X^k)}{k!}\omega^k$$

$$= 1 + jE(X)\omega + j^2 \frac{E(X^2)}{2!}\omega^2 + \cdots + j^k \frac{E(X^k)}{k!}\omega^k + \cdots.$$

Taking the first derivative with respect to ω, and letting it to be equal to zero, we get

$$\left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0} = jE(X) \quad \text{or} \quad E(X) = \frac{1}{j}\left.\frac{\partial \Phi_X(\omega)}{\partial \omega}\right|_{\omega=0}.$$

Similarly, the second derivative gives

$$E(X^2) = \frac{1}{j^2}\left.\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2}\right|_{\omega=0},$$

and repeating this procedure k times, we obtain the kth moment of X to be

$$E(X^k) = \frac{1}{j^k}\left.\frac{\partial^k \Phi_X(\omega)}{\partial \omega^k}\right|_{\omega=0}, \quad k \geq 1.$$

2. Review of Probability Theory: Moments… (12)

Example: If $X \sim P(\lambda)$, then

$$\frac{\partial \Phi_X(\omega)}{\partial \omega} = e^{-\lambda} e^{\lambda e^{j\omega}} \lambda j e^{j\omega},$$

so that $E(X) = \lambda.$

The second derivative gives

$$\frac{\partial^2 \Phi_X(\omega)}{\partial \omega^2} = e^{-\lambda}\left(e^{\lambda e^{j\omega}} (\lambda j e^{j\omega})^2 + e^{\lambda e^{j\omega}} \lambda j^2 e^{j\omega}\right),$$

so that

$$E(X^2) = \lambda^2 + \lambda,$$

2. Review of Probability Theory: Two r.vs (1)

❑ Let X and Y denote two random variables (r.v) based on a probability model (Ω, F, P). Then

$$P\left(x_1 < X(\xi) \le x_2\right) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)dx,$$

and

$$P\left(y_1 < Y(\xi) \le y_2\right) = F_Y(y_2) - F_Y(y_1) = \int_{y_1}^{y_2} f_Y(y)dy.$$

What about the probability that the pair of r.vs (X,Y) belongs to an arbitrary region D? In other words, how does one estimate, for example,

$$P\left[(x_1 < X(\xi) \le x_2) \cap (y_1 < Y(\xi) \le y_2)\right] = ?$$

Towards this, we define the **joint probability distribution function** of X and Y to be

$$F_{XY}(x, y) = P\left[(X(\xi) \le x) \cap (Y(\xi) \le y)\right]$$
$$= P(X \le x, Y \le y) \ge 0,$$

where x and y are arbitrary real numbers.

2. Review of Probability Theory: Two r.vs (2)

❑ Properties

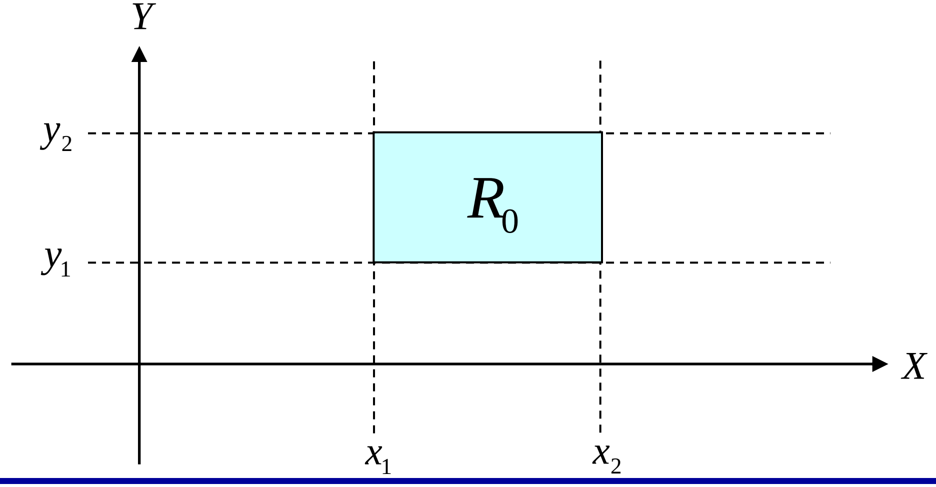
(i) $$F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0, \quad F_{XY}(+\infty, +\infty) = 1.$$

(ii) $$P\big(x_1 < X(\xi) \le x_2, Y(\xi) \le y\big) = F_{XY}(x_2, y) - F_{XY}(x_1, y).$$

$$P\big(X(\xi) \le x, y_1 < Y(\xi) \le y_2\big) = F_{XY}(x, y_2) - F_{XY}(x, y_1).$$

(iii) $$P\big(x_1 < X(\xi) \le x_2, y_1 < Y(\xi) \le y_2\big) = F_{XY}(x_2, y_2) - F_{XY}(x_2, y_1) - F_{XY}(x_1, y_2) + F_{XY}(x_1, y_1).$$

This is the probability that (X,Y) belongs to the rectangle R_0.



2. Review of Probability Theory: Two r.vs (3)

❑ **Joint probability density function (Joint p.d.f):** By definition, the joint p.d.f of X and Y is given by

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x\, \partial y}.$$

and hence we obtain the useful formula

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

Using property (i), we also get

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)\, dx dy = 1.$$

The probability that (X,Y) belongs to an arbitrary region D is given by

$$P\big((X,Y)\in D\big) = \int\int_{(x,y)\in D} f_{XY}(x,y) dx dy.$$

2. Review of Probability Theory: Two r.vs (4)

❑ **Marginal Statistics**: In the context of several r.vs, the statistics of each individual ones are called marginal statistics. Thus $F_X(x)$ is the **marginal probability distribution function** of X, and $f_X(x)$ is the **marginal p.d.f** of X. It is interesting to note that all marginals can be obtained from the joint p.d.f. In fact

$$F_X(x) = F_{XY}(x, +\infty), \quad F_Y(y) = F_{XY}(+\infty, y).$$

Also

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dx.$$

If X and Y are discrete r.vs, then $p_{ij} = P(X = x_i, Y = y_i)$ represents their joint p.d.f, and their respective marginal p.d.fs are given by

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij}$$

$$P(Y = y_j) = \sum_i P(X = x_i, Y = y_j) = \sum_i p_{ij}$$

The joint P.D.F and/or the joint p.d.f represent complete information about the r.vs, and their marginal p.d.fs can be evaluated from the joint p.d.f. However, given marginals, (most often) it will not be possible to compute the joint p.d.f.

2. Review of Probability Theory: Two r.vs (5)

❑ Independence of r.vs

Definition: The random variables X and Y are said to be statistically independent if the events $\{X(\zeta) \in A\}$ and $\{Y(\zeta) \in B\}$ are independent events for any two sets A and B in x and y axes respectively. Applying the above definition to the events $\{X(\zeta) \leq x\}$ and $\{Y(\zeta) \leq y\}$, we conclude that, if the r.vs X and Y are independent, then

$$P\big((X(\xi) \leq x) \cap (Y(\xi) \leq y)\big) = P(X(\xi) \leq x)P(Y(\xi) \leq y)$$

i.e.,

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

or equivalently, if X and Y are independent, then we must have

$$f_{XY}(x, y) = f_X(x)\, f_Y(y).$$

If X and Y are discrete-type r.vs then their independence implies

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \text{for all} \quad i, j.$$

2. Review of Probability Theory: Two r.vs (6)

The equations in previous slide give us the procedure to test for independence. Given $f_{XY}(x,y)$, obtain the marginal p.d.fs $f_X(x)$ and $f_Y(x)$ and examine whether these equations (last two equations) are valid. If so, the r.vs are independent, otherwise they are dependent.

Example: Given

$$f_{XY}(x,y) = \begin{cases} xy^2 e^{-y}, & 0 < y < \infty, \quad 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Determine whether X and Y are independent.

We have $$f_X(x) = \int_0^{+\infty} f_{XY}(x,y)dy = x\int_0^{\infty} y^2 e^{-y} dy$$

$$= x\left(-2ye^{-y}\Big|_0^{\infty} + 2\int_0^{\infty} ye^{-y}dy \right) = 2x, \quad 0 < x < 1.$$

Similarly,

$$f_Y(y) = \int_0^1 f_{XY}(x,y)dx = \frac{y^2}{2}e^{-y}, \quad 0 < y < \infty.$$

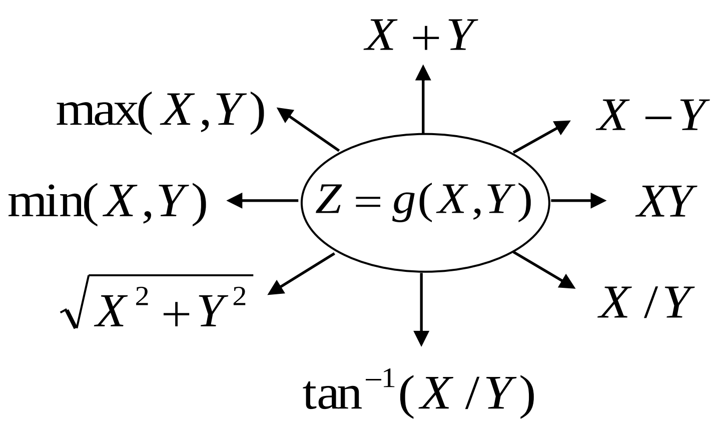
In this case, $f_{XY}(x,y) = f_X(x) f_Y(y)$, and hence X and Y are independent r.vs.

2. Review of Probability Theory: Func. of 2 r.vs (1)

- Given two random variables X and Y and a function $g(x,y)$, we form a new random variable $Z = g(X, Y)$.

Given the joint p.d.f $f_{XY}(x, y)$, how does one obtain $f_Z(z)$ the p.d.f of Z ? Problems of this type are of interest from a practical standpoint. For example, a receiver output signal usually consists of the desired signal buried in noise, and the above formulation in that case reduces to $Z = X + Y$.

It is important to know the statistics of the incoming signal for proper receiver design. In this context, we shall analyze problems of the following type:



2. Review of Probability Theory: Func. of 2 r.vs (2)

- Start with: $F_Z(z) = P\big(Z(\xi) \le z\big) = P\big(g(X,Y) \le z\big) = P\big[(X,Y) \in D_z\big]$

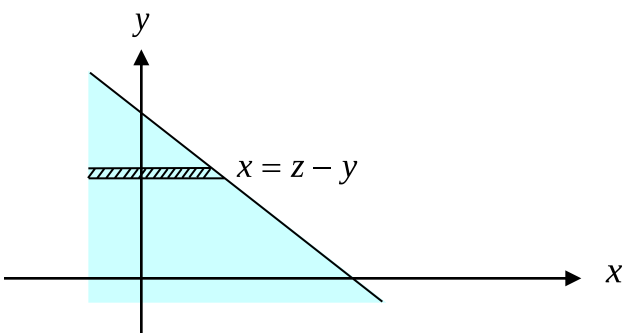
$$= \int\int_{x,y \in D_z} f_{XY}(x,y)dxdy,$$

where D_z in the XY plane represents the region such that $g(x,y) \le z$ is satisfied. To determine $F_Z(z)$, it is enough to find the region D_z for every z, and then evaluate the integral there.

Example 1: $Z = X + Y$. Find $f_Z(z)$.

$$F_Z(z) = P\big(X + Y \le z\big) = \int_{y=-\infty}^{+\infty}\int_{x=-\infty}^{z-y} f_{XY}(x,y)dxdy,$$

since the region D_z of the xy plane where $x+y \le z$ is the shaded area in the figure to the left of the line $x+y = z$. Integrating over the horizontal strip along the x-axis first (inner integral) followed by sliding that strip along the y-axis from $-\infty$ to $+\infty$ (outer integral) we cover the entire shaded area.



2. Review of Probability Theory: Func. of 2 r.vs (3)

We can find $f_Z(z)$ by differentiating $F_Z(z)$ directly. In this context, it is useful to recall the differentiation rule due to Leibnitz. Suppose

$$H(z) = \int_{a(z)}^{b(z)} h(x, z)dx.$$

Then

$$\frac{dH(z)}{dz} = \frac{db(z)}{dz} h\big(b(z), z\big) - \frac{da(z)}{dz} h\big(a(z), z\big) + \int_{a(z)}^{b(z)} \frac{\partial h(x, z)}{\partial z} dx.$$

Using above equations, we get

$$f_Z(z) = \int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial z} \int_{-\infty}^{z-y} f_{XY}(x, y)dx \right) dy = \int_{-\infty}^{+\infty} \left(f_{XY}(z-y, y) - 0 + \int_{-\infty}^{z-y} \frac{\partial f_{XY}(x, y)}{\partial z} \right) dy$$

$$= \int_{-\infty}^{+\infty} f_{XY}(z-y, y)dy.$$

If X and Y are independent, $f_{XY}(x,y) = f_X(x)f_Y(y)$, then we get

$$f_Z(z) = \int_{y=-\infty}^{+\infty} f_X(z-y) f_Y(y)dy = \int_{x=-\infty}^{+\infty} f_X(x) f_Y(z-x)dx. \quad \text{(Convolution!)}$$

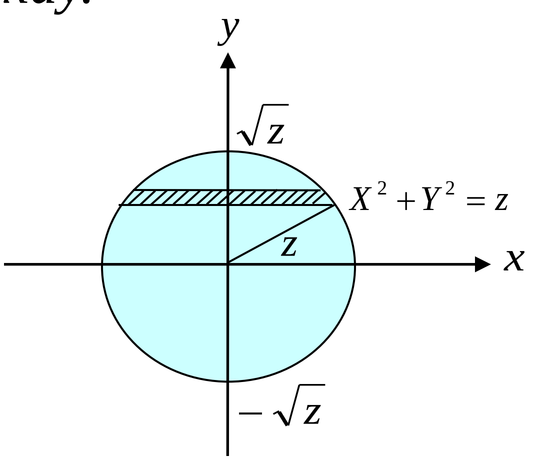
Example 2: X and Y are independent normal r.vs with zero mean and common variance σ^2 . Determine $f_Z(z)$ for $Z = X^2 + Y^2$.

We have

$$F_Z(z) = P\left(X^2 + Y^2 \le z\right) = \int\int_{X^2+Y^2\le z} f_{XY}(x,y)dxdy.$$

But, $X^2 + Y^2 \le z$ represents the area of a circle with radius $\sqrt{z}$ and hence

$$F_Z(z) = \int_{y=-\sqrt{z}}^{\sqrt{z}} \int_{x=-\sqrt{z-y^2}}^{\sqrt{z-y^2}} f_{XY}(x,y)dxdy.$$



This gives after repeated differentiation

$$f_Z(z) =$$

$$\int_{y=-\sqrt{z}}^{\sqrt{z}} \frac{1}{2\sqrt{z-y^2}}\left(f_{XY}(\sqrt{z-y^2},y) + f_{XY}(-\sqrt{z-y^2},y)\right)dy. \qquad (*)$$

2. Review of Probability Theory: Func. of 2 r.vs (5)

Moreover, X and Y are said to be jointly normal (Gaussian) distributed, if their joint p.d.f has the following form:

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-r^2}}e^{\frac{-1}{2(1-r^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2}-\frac{2r(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)},$$

$$-\infty<x<+\infty,\quad -\infty<y<+\infty,\quad |r|<1.$$

with zero mean

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_1^2}-\frac{2rxy}{\sigma_1\sigma_2}+\frac{y^2}{\sigma_2^2}\right)}.$$

with $r = 0$ and $\sigma_1 = \sigma_2 = \sigma$, then direct substitution into (*), we get

$$f_Z(z)=\int_{y=-\sqrt{z}}^{\sqrt{z}}\frac{1}{2\sqrt{z-y^2}}\left(2\cdot\frac{1}{2\pi\sigma^2}e^{-(z-y^2+y^2)/2\sigma^2}\right)dy=\frac{e^{-z/2\sigma^2}}{\pi\sigma^2}\int_0^{\sqrt{z}}\frac{1}{\sqrt{z-y^2}}\,dy$$

$$=\frac{e^{-z/2\sigma^2}}{\pi\sigma^2}\int_0^{\pi/2}\frac{\sqrt{z}\cos\theta}{\sqrt{z}\cos\theta}\,d\theta=\frac{1}{2\sigma^2}e^{-z/2\sigma^2}U(z),$$

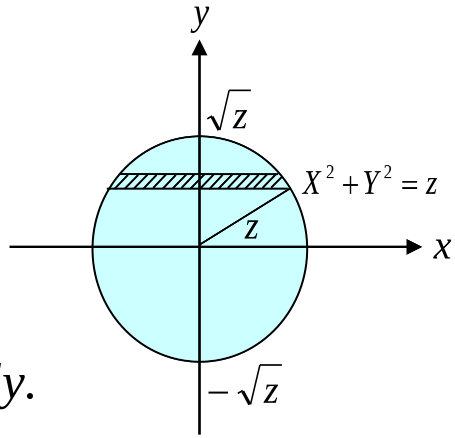
with $y=\sqrt{z}\sin\theta.$

2. Review of Probability Theory: Func. of 2 r.vs (6)

Thus, *if X and Y are independent zero mean Gaussian r.vs with common variance σ^2 then X^2+Y^2 is an exponential r.vs with parameter $2\sigma^2$.*

<u>Example 3</u>: Let $Z=\sqrt{X^2+Y^2}$. Find $f_Z(z)$

From the figure, the present case corresponds to a circle with radius z^2. Thus



$$F_Z(z)=\int_{y=-z}^{z}\int_{x=-\sqrt{z^2-y^2}}^{\sqrt{z^2-y^2}} f_{XY}(x,y)dxdy.$$

$$f_Z(z)=\int_{-z}^{z}\frac{z}{\sqrt{z^2-y^2}}\left(f_{XY}(\sqrt{z^2-y^2},y)+f_{XY}(-\sqrt{z^2-y^2},y)\right)dy.$$

Now suppose X and Y are independent Gaussian as in Example 2, we obtain

$$f_Z(z)=2\int_0^z\frac{z}{\sqrt{z^2-y^2}}\frac{1}{2\pi\sigma^2}e^{(z^2-y^2+y^2)/2\sigma^2}dy=\frac{2z}{\pi\sigma^2}e^{-z^2/2\sigma^2}\int_0^z\frac{1}{\sqrt{z^2-y^2}}dy$$

$$=\frac{2z}{\pi\sigma^2}e^{-z^2/2\sigma^2}\int_0^{\pi/2}\frac{z\cos\theta}{z\cos\theta}d\theta=\frac{z}{\sigma^2}e^{-z^2/2\sigma^2}U(z),\qquad \textit{Rayleigh distribution!}$$

2. Review of Probability Theory: Func. of 2 r.vs (7)

Thus, *if* $W = X + iY$, *where X and Y are real, independent normal r.vs with zero mean and equal variance, then the r.v* $|W| = \sqrt{X^2 + Y^2}$ *has a Rayleigh density. W is said to be a complex Gaussian r.v with zero mean, whose real and imaginary parts are independent r.vs and its magnitude has Rayleigh distribution.*

What about its phase $\theta = \tan^{-1}\left(\dfrac{X}{Y}\right)$?

Clearly, the principal value of θ lies in the interval $(-\pi/2, +\pi/2)$. If we let $U = \tan\,\theta = X/Y$, then it is shown that U has a Cauchy distribution with

$$f_U(u) = \frac{1/\pi}{u^2+1}, \quad -\infty < u < \infty.$$

As a result

$$f_\theta(\theta) = \frac{1}{|\,d\theta/du\,|} f_U(\tan\theta) = \frac{1}{(1/\sec^2\theta)} \frac{1/\pi}{\tan^2\theta+1} = \begin{cases} 1/\pi, & -\pi/2 < \theta < \pi/2, \\ 0, & \text{otherwise}. \end{cases}$$

2. Review of Probability Theory: Func. of 2 r.vs (8)

To summarize, *the magnitude and phase of a zero mean complex Gaussian r.v has Rayleigh and uniform distributions respectively. Interestingly, as we will see later, these two derived r.vs are also independent of each other!*

<u>Example 4</u>: Redo example 3, where X and Y have nonzero means μ_X and μ_Y respectively.

Since

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2} e^{-[(x-\mu_X)^2 + (y-\mu_Y)^2]/2\sigma^2},$$

Similar to example 3, we obtain the *Rician probability density function* to be

$$f_Z(z) = \frac{z e^{-(z^2+\mu^2)/2\sigma^2}}{2\pi\sigma^2} \int_{-\pi/2}^{\pi/2} \left(e^{z\mu\cos(\theta-\phi)/\sigma^2} + e^{-z\mu\cos(\theta+\phi)/\sigma^2} \right) d\theta$$

$$= \frac{z e^{-(z^2+\mu^2)/2\sigma^2}}{2\pi\sigma^2} \left(\int_{-\pi/2}^{\pi/2} e^{z\mu\cos(\theta-\phi)/\sigma^2}\, d\theta + \int_{\pi/2}^{3\pi/2} e^{z\mu\cos(\theta-\phi)/\sigma^2}\, d\theta \right)$$

$$= \frac{z e^{-(z^2+\mu^2)/2\sigma^2}}{2\pi\sigma^2} I_0\left(\frac{z\mu}{\sigma^2}\right),$$

2. Review of Probability Theory: Func. of 2 r.vs (9)

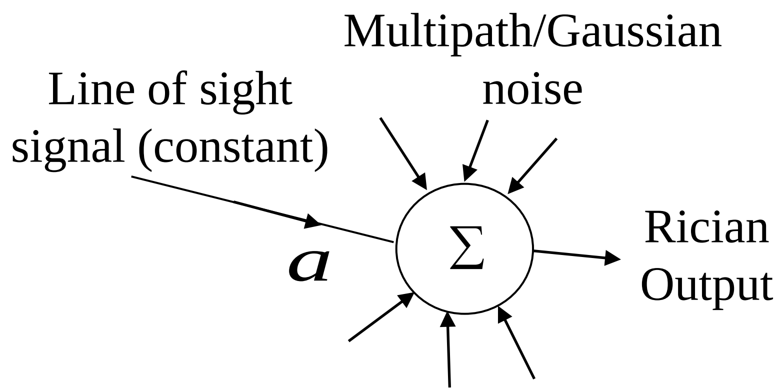
where $x = z\cos\theta,\ y = z\sin\theta,\ \ \mu = \sqrt{\mu_X^2 + \mu_Y^2},\ \ \mu_X = \mu\cos\phi,\ \ \mu_Y = \mu\sin\phi,$

and

$$I_0(\eta) = \frac{1}{2\pi}\int_0^{2\pi} e^{\eta\cos(\theta-\phi)}d\theta = \frac{1}{\pi}\int_0^{\pi} e^{\eta\cos\theta}d\theta$$

is the modified Bessel function of the first kind and zeroth order.

Thus, *if X and Y have nonzero means μ_X and μ_Y, respectively. Then* $Z = \sqrt{X^2 + Y^2}$ *is said to be a Rician r.v*. Such a scene arises in fading multipath situation where there is a dominant constant component (mean) in addition to a zero mean Gaussian r.v. The constant component may be the line of sight signal and the zero mean Gaussian r.v part could be due to random multipath components adding up incoherently (see diagram). The envelope of such a signal is said to have a Rician p.d.f.



2. Review of Probability Theory: Joint Moments… (1)

❑ Given two r.vs X and Y and a function $g(x,y)$, define the r.v $Z = g(X,Y)$, the **mean** of Z can be defined as

$$\mu_Z = E(Z) = \int_{-\infty}^{+\infty} z\, f_Z(z)dz.$$

or more useful formula

$$E(Z) = \int_{-\infty}^{+\infty} z\, f_Z(z)dz = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y) f_{XY}(x,y)dxdy.$$

If X and Y are discrete-type r.vs, then

$$E[g(X,Y)] = \sum_i \sum_j g(x_i, y_j) P(X = x_i, Y = y_j).$$

Since expectation is a linear operator, we also get

$$E\left(\sum_k a_k g_k(X,Y)\right) = \sum_k a_k E[g_k(X,Y)].$$

2. Review of Probability Theory: Joint Moments… (2)

If X and Y are independent r.vs, it is easy to see that $V=g(X)$ and $W=h(Y)$ are always independent of each other. We get the interesting result

$$E[g(X)h(Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x)h(y)\, f_X(x)\, f_Y(y)\,dxdy$$

$$= \int_{-\infty}^{+\infty} g(x)\, f_X(x)dx \int_{-\infty}^{+\infty} h(y)\, f_Y(y)dy = E[g(X)]E[h(Y)].$$

In the case of one random variable, we defined the parameters mean and variance to represent its average behavior. How does one parametrically represent similar cross-behavior between two random variables? Towards this, we can generalize the variance definition given as

- **Covariance**: Given any two r.vs X and Y, define

$$Cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right].$$

or

$$Cov(X,Y) = E(XY) - \mu_X\mu_Y = E(XY) - E(X)E(Y)$$

$$= \overline{XY} - \overline{X}\,\overline{Y}.$$

2. Review of Probability Theory: Joint Moments… (3)

It is easy to see that

$$|Cov(X,Y)| \le \sqrt{Var(X)Var(Y)}\,.$$

We define the normalized parameter

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \le \rho_{XY} \le 1,$$

and it represents the **correlation coefficient** between X and Y.

Uncorrelated r.vs: If $\rho_{XY} = 0$, then X and Y are said to be uncorrelated r.vs. If X and Y are uncorrelated, then $E(XY) = E(X)E(Y)$

Orthogonality: X and Y are said to be orthogonal if $E(XY) = 0$

If either X or Y has zero mean, then orthogonality implies uncorrelatedness also and vice-versa. Suppose X and Y are independent r.vs, it also implies they are uncorrelated.

2. Review of Probability Theory: Joint Moments… (4)

Naturally, if two random variables are statistically independent, then there cannot be any correlation between them $\rho_{XY} = 0$. However, the converse is in general not true. As the next example shows, random variables can be uncorrelated without being independent.

Example 5: Let $Z = aX + bY$. Determine the variance of Z in terms of σ_X, σ_Y and ρ_{XY} .

We have

$$\mu_Z = E(Z) = E(aX + bY) = a\mu_X + b\mu_Y$$

$$\begin{aligned}\sigma_Z^2 &= Var(Z) = E\left[(Z-\mu_Z)^2\right] = E\left[\left(a(X-\mu_X)+b(Y-\mu_Y)\right)^2\right] \\ &= a^2E(X-\mu_X)^2 + 2abE\left((X-\mu_X)(Y-\mu_Y)\right) + b^2E(Y-\mu_Y)^2 \\ &= a^2\sigma_X^2 + 2ab\rho_{XY}\sigma_X\sigma_Y + b^2\sigma_Y^2.\end{aligned}$$

In particular, if X and Y are independent, then $\rho_{XY} = 0$ and

$$\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2.$$

2. Review of Probability Theory: Joint Moments… (5)

❑ **Moments**: represents the joint moment of order (k,m) for X and Y

$$E[X^k Y^m] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^k y^m f_{XY}(x,y)dx\,dy,$$

Following the one random variable case, we can define the joint characteristic function between two random variables which will turn out to be useful for moment calculations.

❑ **Joint characteristic functions**: between X and Y is defined as

$$\Phi_{XY}(u,v) = E\left(e^{j(Xu+Yv)}\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{j(Xu+Yv)} f_{XY}(x,y)dxdy.$$

Note that $\left|\Phi_{XY}(u,v)\right| \le \Phi_{XY}(0,0) = 1.$

It is easy to show that

$$E(XY) = \frac{1}{j^2}\frac{\partial^2 \Phi_{XY}(u,v)}{\partial u \partial v}\bigg|_{u=0,v=0}.$$

2. Review of Probability Theory: Joint Moments… (6)

If X and Y are independent r.vs, then we obtain

$$\Phi_{XY}(u,v) = E(e^{juX})E(e^{jvY}) = \Phi_X(u)\Phi_Y(v).$$

Also

$$\Phi_X(u) = \Phi_{XY}(u,0), \quad \Phi_Y(v) = \Phi_{XY}(0,v).$$

❑ **More on Gaussian r.vs**: the joint characteristic function of two jointly Gaussian r.vs to be

$$\Phi_{XY}(u,v) = E(e^{j(Xu+Yv)}) = e^{j(\mu_X u + \mu_Y v) - \frac{1}{2}(\sigma_X^2 u^2 + 2r\sigma_X\sigma_Y uv + \sigma_Y^2 v^2)}.$$

<u>Example 6</u>: Let X and Y be jointly Gaussian r.vs with parameters $N(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, r)$. Define $Z = aX + bY$. Determine $f_Z(z)$.

In this case we can use characteristic function to solve.

$$\Phi_Z(u) = E(e^{jZu}) = E(e^{j(aX+bY)u}) = E(e^{jauX+jbuY})$$
$$= \Phi_{XY}(au, bu).$$

2. Review of Probability Theory: Joint Moments… (7)

or

$$\Phi_Z(u) = e^{j(a\mu_X + b\mu_Y)u - \frac{1}{2}(a^2\sigma_X^2 + 2rab\sigma_X\sigma_Y + b^2\sigma_Y^2)u^2} = e^{j\mu_Z u - \frac{1}{2}\sigma_Z^2 u^2},$$

where

$$\mu_Z \triangleq a\mu_X + b\mu_Y,$$

$$\sigma_Z^2 \triangleq a^2\sigma_X^2 + 2rab\sigma_X\sigma_Y + b^2\sigma_Y^2.$$

Thus, $Z = aX + bY$ is also Gaussian with mean and variance as above. We conclude that *any linear combination of jointly Gaussian r.vs generate a Gaussian r.v*.

Example 7: Suppose X and Y are jointly Gaussian r.vs as in the example 6. Define two linear combinations: $Z = aX + bY$ and $W = cX + dY$. Determine their joint distribution?

The characteristic function of Z and W is given by

$$\Phi_{ZW}(u,v) = E(e^{j(Zu+Wv)}) = E(e^{j(aX+bY)u + j(cX+dY)v})$$

$$= E(e^{jX(au+cv) + jY(bu+dv)}) = \Phi_{XY}(au+cv, bu+dv).$$

2. Review of Probability Theory: Joint Moments… (8)

Similar to example 6, we get

$$\Phi_{ZW}(u,v) = e^{j(\mu_Z u + \mu_W v) - \frac{1}{2}(\sigma_Z^2 u^2 + 2r_{ZW}\sigma_X \sigma_Y uv + \sigma_W^2 v^2)},$$

where

$$\mu_Z = a\mu_X + b\mu_Y,$$
$$\mu_W = c\mu_X + d\mu_Y,$$
$$\sigma_Z^2 = a^2\sigma_X^2 + 2abr\sigma_X\sigma_Y + b^2\sigma_Y^2,$$
$$\sigma_W^2 = c^2\sigma_X^2 + 2cdr\sigma_X\sigma_Y + d^2\sigma_Y^2,$$

and

$$\rho_{ZW} = \frac{ac\sigma_X^2 + (ad+bc)r\sigma_X\sigma_Y + bd\sigma_Y^2}{\sigma_Z\sigma_W}.$$

Thus, Z and W are also jointly distributed Gaussian r.vs with means, variances and correlation coefficient as above.

2. Review of Probability Theory: Joint Moments… (9)

To summarize, *any two linear combinations of jointly Gaussian random variables (independent or dependent) are also jointly Gaussian r.vs.*

Gaussian input $\longrightarrow$ Linear operator $\longrightarrow$ Gaussian output

Gaussian r. vs are also interesting because of the following result:

- **Central Limit Theorem**: Suppose $X_1, X_2, \ldots, X_n$ are a set of *zero mean independent*, *identically distributed (i.i.d) random variables* with *some common distribution*. Consider their scaled sum

$$Y = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}.$$

Then asymptotically (as $n \to \infty$)

$$Y \to N(0, \sigma^2).$$

2. Review of Probability Theory: Joint Moments… (10)

The central limit theorem states that *a large sum of independent random variables each with finite variance tends to behave like a normal random variable. Thus the individual p.d.fs become unimportant to analyze the collective sum behavior. If we model the noise phenomenon as the sum of a large number of independent random variables (eg: electron motion in resistor components), then this theorem allows us to conclude that noise behaves like a Gaussian r.v.*

<u>General:</u> Given n independent r.vs X_i, we form their sum: $X = X_1 + \ldots + X_n$

X is a r.v with mean $\eta = \eta_1 + \ldots + \eta_n$ and the variance $\sigma^2 = \sigma_1^2 + \ldots + \sigma_n^2$
The central limit theorem states that under certain general conditions, the distribution of X approaches a **normal distribution** with mean η and variance σ^2 as n increases

$$F(x) \approx G\left(\frac{x-\eta}{\sigma}\right)$$

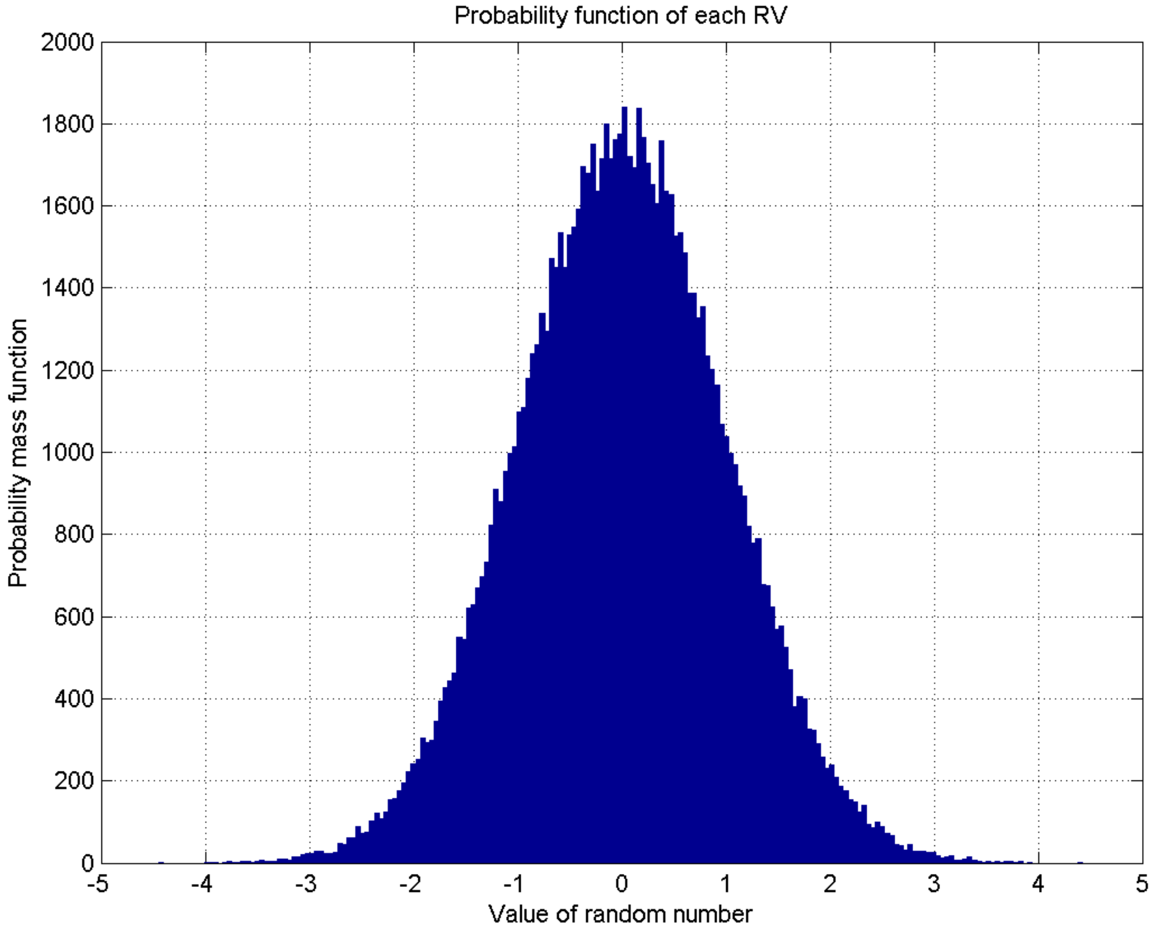
2. Review of Probability Theory: Joint Moments… (11)

Furthermore, if the r.vs X_i are continuous type, the p.d.f $f(x)$ of X approaches a **nomal density**:

$$f(x) \approx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\eta)^2/2\sigma^2}$$

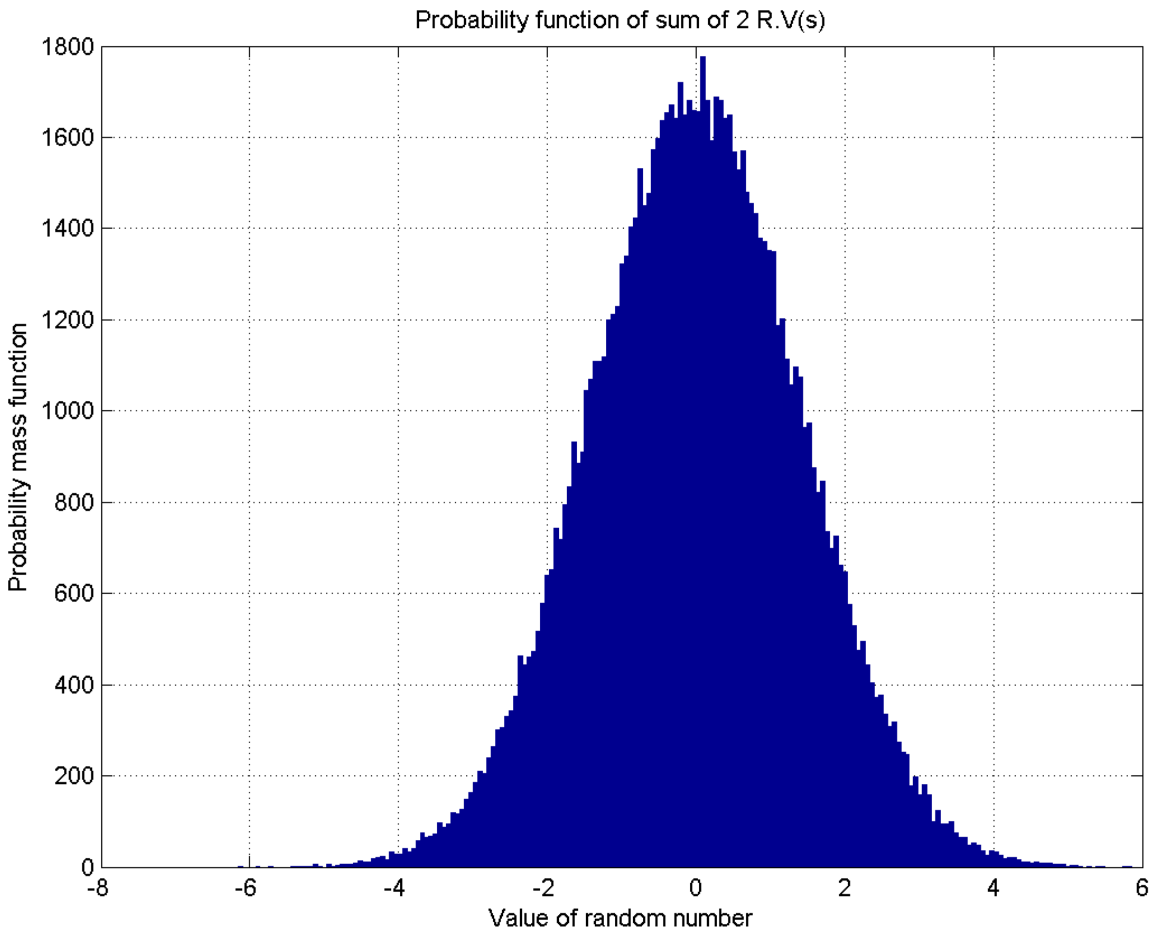
Example: See example 8-15, p.215, [4]
and
http://en.wikipedia.org/wiki/Illustration_of_the_central_limit_theorem

2. Review of Probability Theory: Joint Moments… (12)



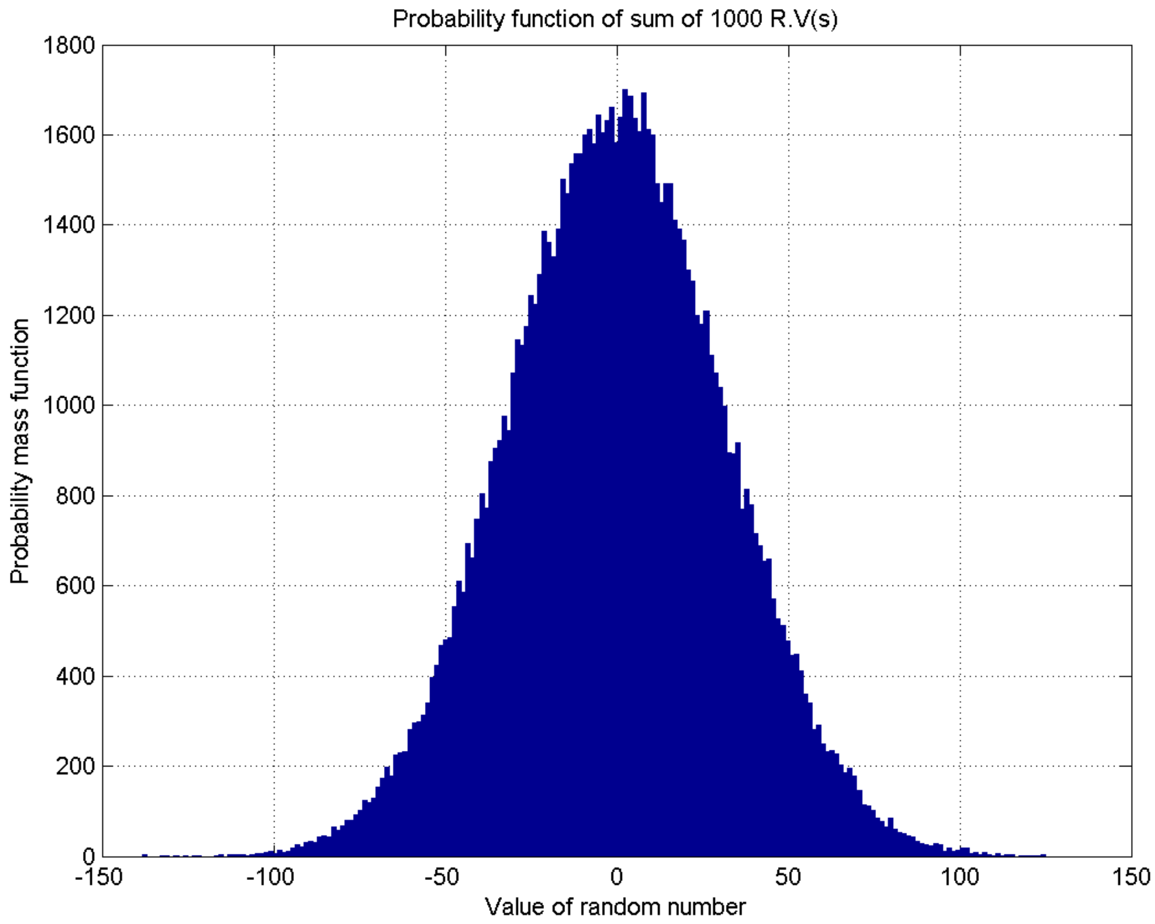
Normal distribution μ=0, σ=1

2. Review of Probability Theory: Joint Moments… (13)



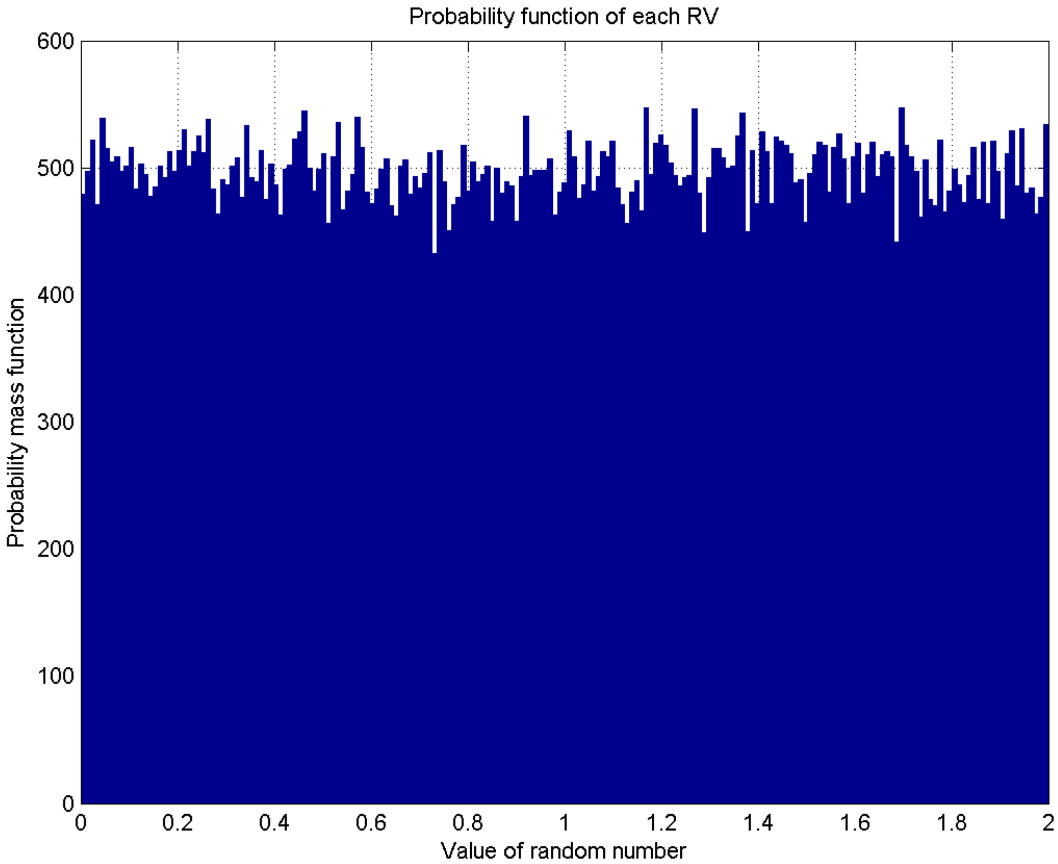
Central limit theorem with sum of 2 identically distributed r.vs (normal dis. μ=0,σ=1)

2. Review of Probability Theory: Joint Moments… (14)



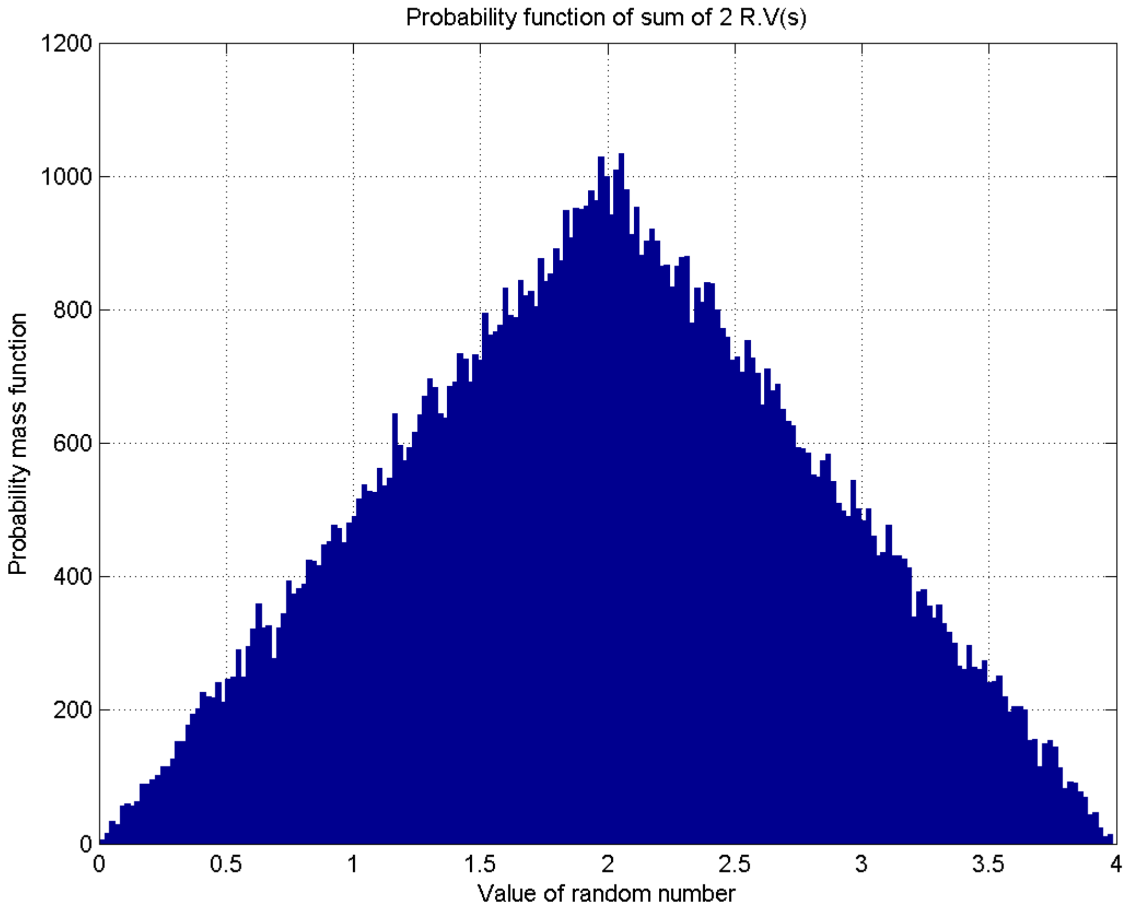
Central limit theorem with sum of 1000 identically distributed r.vs (normal dis. $\mu=0,\sigma=1$)

2. Review of Probability Theory: Joint Moments… (15)



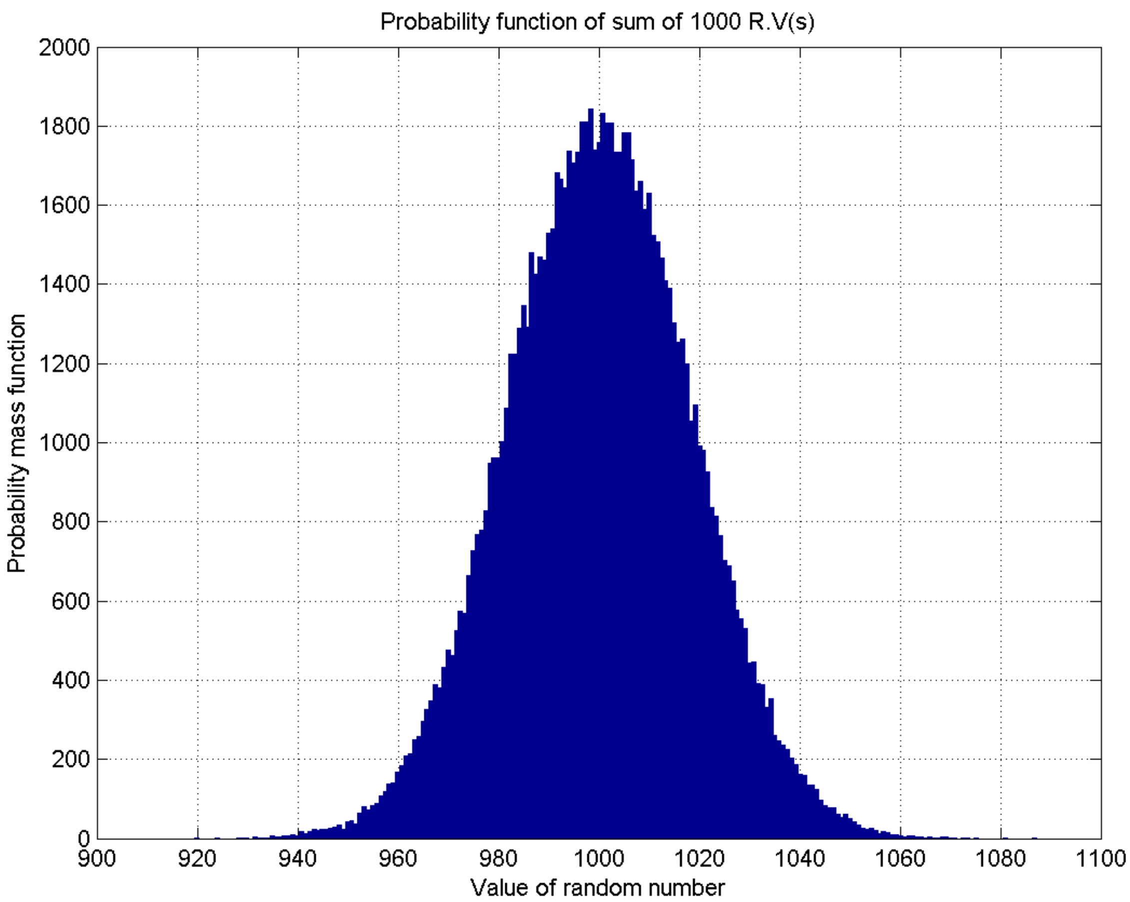
Uniform distribution [0, 2]

2. Review of Probability Theory: Joint Moments… (16)



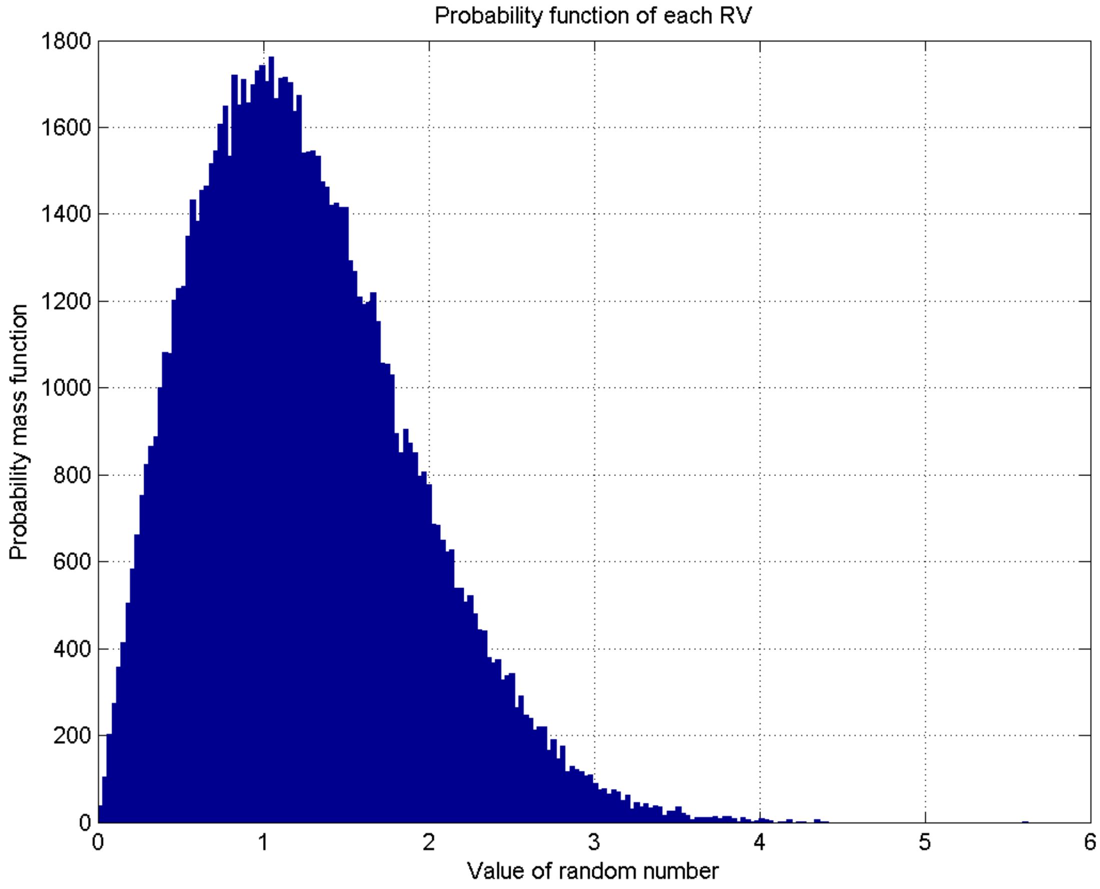
Central limit theorem with sum of 2 identically distributed r.vs (uniform dis. [0,2])

2. Review of Probability Theory: Joint Moments… (17)



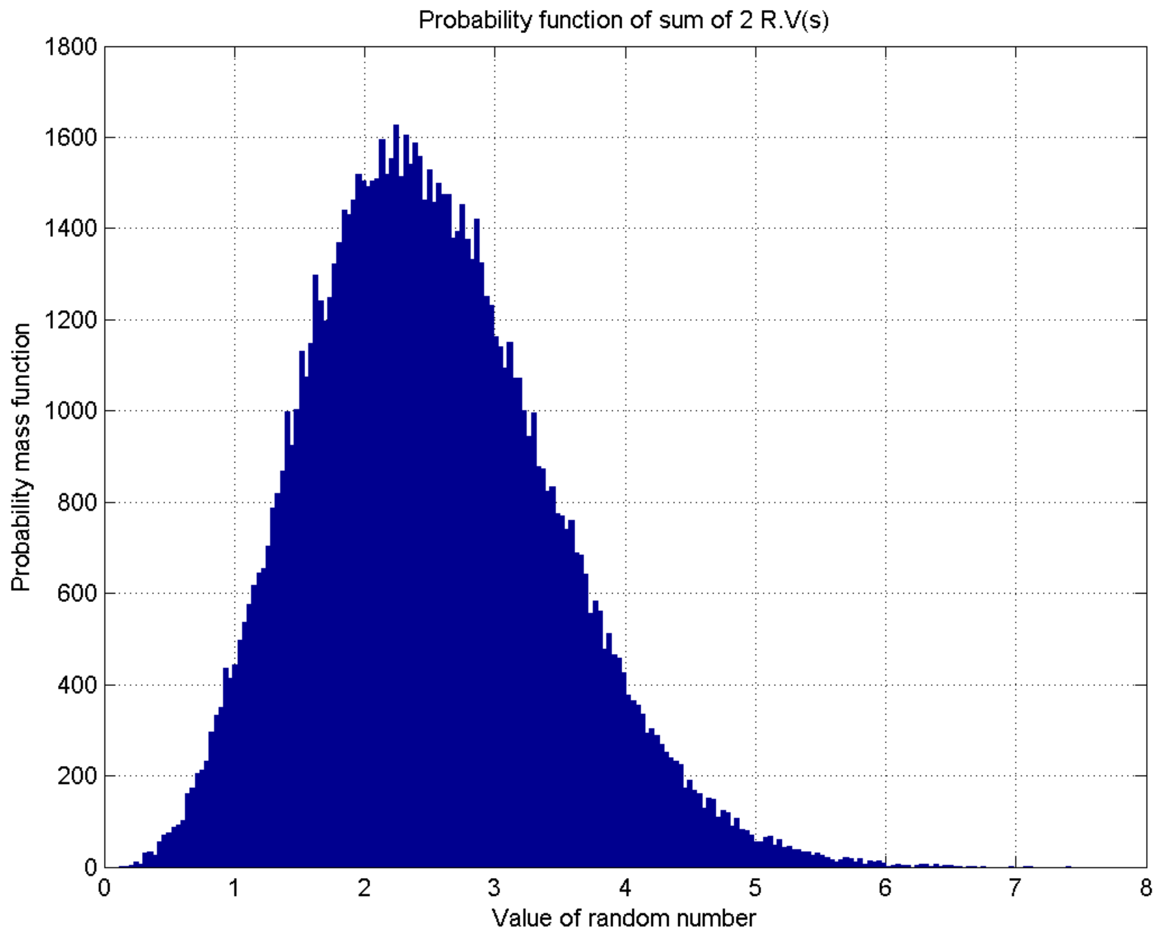
Central limit theorem with sum of 1000 identically distributed r.vs (uniform dis. [0,2])

2. Review of Probability Theory: Joint Moments… (18)



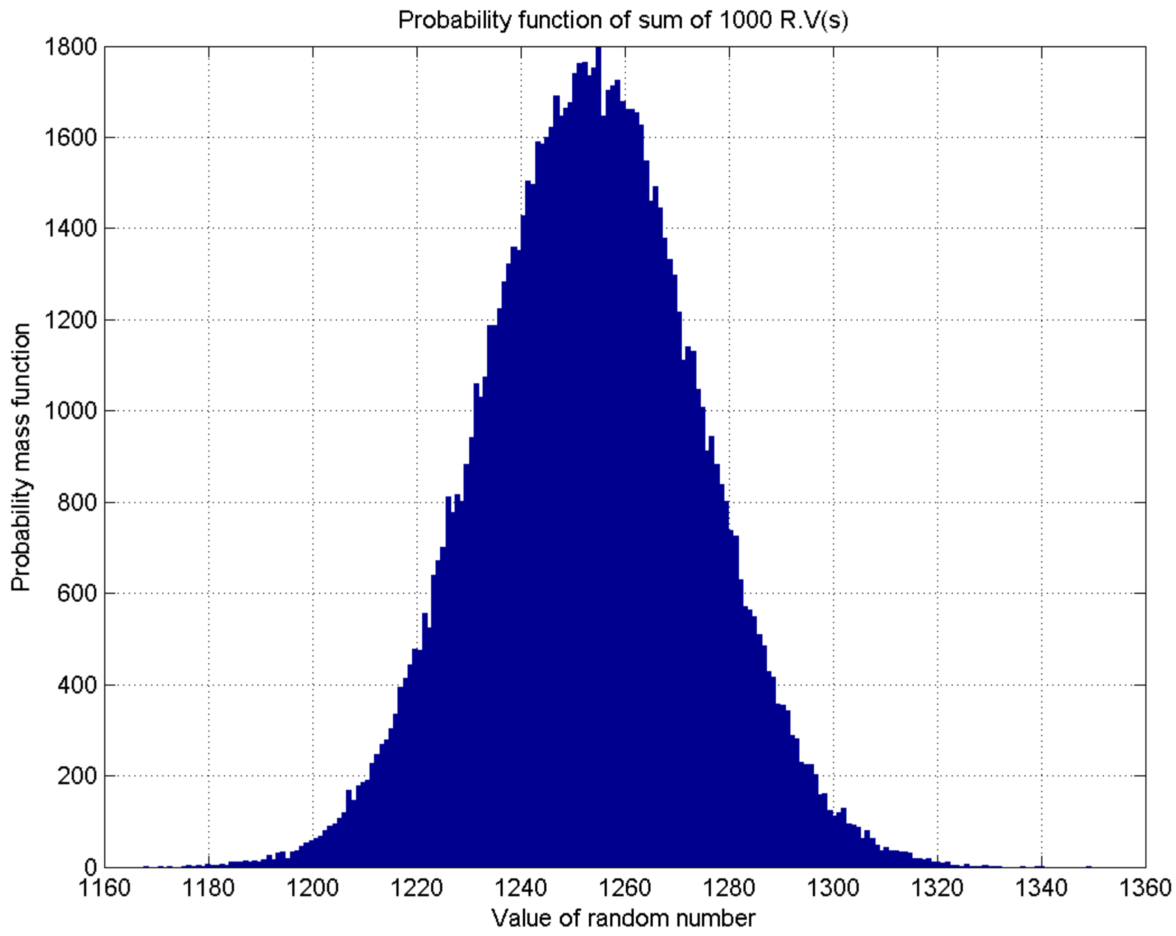
Rayleigh distribution, σ=1

2. Review of Probability Theory: Joint Moments… (19)



Central limit theorem with sum of 2 identically distributed r.vs (Rayleigh dis. ,σ=1)

2. Review of Probability Theory: Joint Moments… (20)



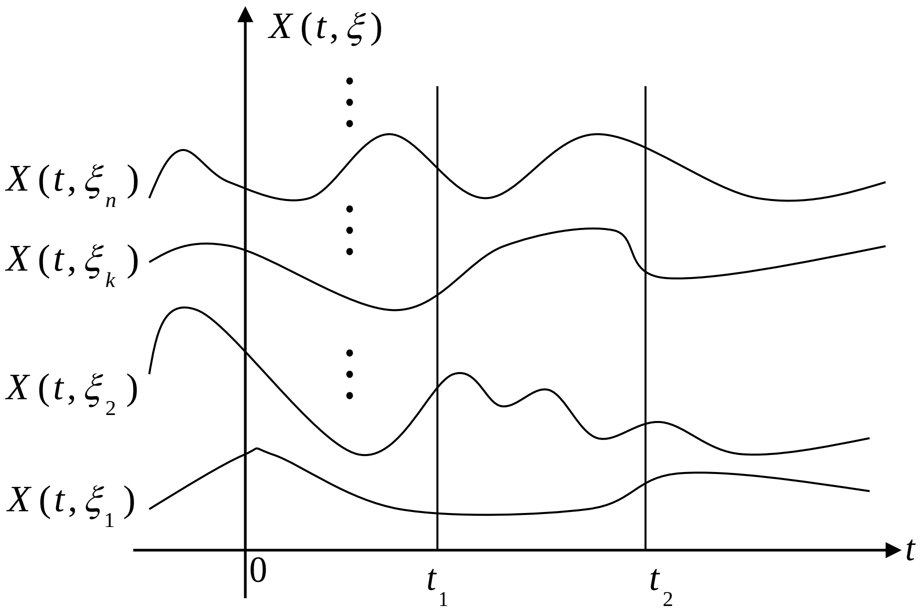
Central limit theorem with sum of 1000 identically distributed r.vs (Rayleigh dis. ,σ=1)

2. Review of Stochastic Processes and Models

- Mean, Autocorrelation.
- Stationarity.
- Deterministic Systems.
- Discrete Time Stochastic Process.
- Stochastic Models.

2. Review of Stochastic Processes: Introduction (1)

- Let ζ denote the random outcome of an experiment. To every such outcome suppose a waveform $X(t, \zeta)$ is assigned. The collection of such waveforms form a stochastic process. The set of $\{\zeta_k\}$ and the time index t can be continuous or discrete (countably infinite or finite) as well. For fixed $\zeta_i \in S$ (the set of all experimental outcomes), $X(t, \zeta)$ is a specific time function. For fixed t, $X_1 = X(t_1, \zeta_i)$ is a random variable. The ensemble of all such realizations $X(t, \zeta)$ over time represents the **stochastic process** (or **random process**) $X(t)$ (see the figure).



For example:

$$X(t) = a\cos(\omega_0 t + \phi)$$

where ϕ is a uniformly distributed random variable in $(0, 2\pi)$, represents a stochastic process.

2. Review of Stochastic Processes: Introduction (2)

If $X(t)$ is a stochastic process, then for fixed t, $X(t)$ represents a random variable. Its distribution function is given by

$$F_X(x,t) = P\{X(t) \le x\}$$

Notice that $F_X(x, t)$ depends on t, since for a different t, we obtain a different random variable. Further

$$f_X(x,t) \triangleq \frac{dF_X(x,t)}{dx}$$

represents the first-order probability density function of the process $X(t)$.

For $t = t_1$ and $t = t_2$, $X(t)$ represents two different random variables $X_1 = X(t_1)$ and $X_2 = X(t_2)$, respectively. Their joint distribution is given by

$$F_X(x_1,x_2,t_1,t_2) = P\{X(t_1) \le x_1, X(t_2) \le x_2\}$$

and

$$f_X(x_1,x_2,t_1,t_2) \triangleq \frac{\partial^2 F_X(x_1,x_2,t_1,t_2)}{\partial x_1 \, \partial x_2}$$

represents the second-order density function of the process $X(t)$.

2. Review of Stochastic Processes: Introduction (3)

Similarly, $f_X(x_1, x_2,\dots x_n, t_1, t_2,\dots t_n)$ represents the nth order density function of the process $X(t)$. Complete specification of the stochastic process $X(t)$ requires the knowledge of $f_X(x_1, x_2,\dots x_n, t_1, t_2,\dots t_n)$ for all t_i , $i = 1, 2\dots, n$ and for all n. (an almost impossible task in reality!).

❑ **Mean of a stochastic process:**

$$\mu(t) \triangleq E\{X(t)\} = \int_{-\infty}^{+\infty} x\, f_X(x,t)dx$$

represents the mean value of a process $X(t)$. In general, the mean of a process can depend on the time index t.

❑ **Autocorrelation** function of a process $X(t)$ is defined as

$$R_{XX}(t_1,t_2) \triangleq E\{X(t_1)X^*(t_2)\} = \int\int x_1 x_2^* f_X(x_1,x_2,t_1,t_2)dx_1 dx_2$$

and it represents the interrelationship between the random variables $X_1 = X(t_1)$ and $X_2 = X(t_2)$ generated from the process $X(t)$.

2. Review of Stochastic Processes: Introduction (4)

Properties:

(i) $$R_{xx}(t_1,t_2) = R^*_{xx}(t_2,t_1) = [E\{X(t_2)X^*(t_1)\}]^*$$

(ii) $$R_{xx}(t,t) = E\{|X(t)|^2\} > 0.$$ (Average instantaneous power)

(iii) $R_{XX}(t_1, t_2)$ represents a nonnegative definite function, i.e., for *any* set of constants $\{a_i\}^n_{i=1}$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j^* R_{xx}(t_i,t_j) \geq 0.$$

this follows by noticing that $E\{|Y|^2\} \geq 0$ for $Y = \sum_{i=1}^{n} a_i X(t_i).$

The function

$$C_{xx}(t_1,t_2) = R_{xx}(t_1,t_2) - \mu_x(t_1)\mu_x^*(t_2)$$

represents the **autocovariance** function of the process $X(t)$.

2. Review of Stochastic Processes: Introduction (5)

<u>Example:</u>

$$X(t) = a\cos(\omega_0 t + \varphi), \qquad \varphi \sim U(0,2\pi).$$

This gives

$$\mu_X(t) = E\{X(t)\} = aE\{\cos(\omega_0 t + \varphi)\}$$
$$= a\cos\omega_0 t\, E\{\cos\varphi\} - a\sin\omega_0 t\, E\{\sin\varphi\} = 0,$$

since $\quad E\{\cos\varphi\} = \frac{1}{2\pi}\int_0^{2\pi}\cos\varphi\, d\varphi = 0 = E\{\sin\varphi\}.$

Similarly,

$$R_{XX}(t_1,t_2) = a^2 E\{\cos(\omega_0 t_1 + \varphi)\cos(\omega_0 t_2 + \varphi)\}$$
$$= \frac{a^2}{2} E\{\cos\omega_0(t_1 - t_2) + \cos(\omega_0(t_1 + t_2) + 2\varphi)\}$$
$$= \frac{a^2}{2}\cos\omega_0(t_1 - t_2).$$

2. Review of Stochastic Processes: Stationary (1)

❑ Stationary processes exhibit statistical properties that are invariant to shift in the time index. Thus, for example, second-order stationarity implies that the statistical properties of the pairs $\{X(t_1), X(t_2)\}$ and $\{X(t_1+c), X(t_2+c)\}$ are the same for *any* c. Similarly, first-order stationarity implies that the statistical properties of $X(t_i)$ and $X(t_i+c)$ are the same for any c.

In strict terms, the statistical properties are governed by the joint probability density function. Hence a process is nth-order **Strict-Sense Stationary (S.S.S)** if

$$f_x(x_1, x_2, \cdots x_n,\ t_1, t_2 \cdots, t_n) \equiv f_x(x_1, x_2, \cdots x_n,\ t_1 + c, t_2 + c \cdots, t_n + c) \quad (*)$$

for *any* c, where the left side represents the joint density function of the random variables $X_1 = X(t_1)$, $X_2 = X(t_2)$, ..., $X_n = X(t_n)$, and the right side corresponds to the joint density function of the random variables $X'_1 = X(t_1+c)$, $X'_2 = X(t_2+c)$, ..., $X'_n = X(t_n+c)$. A process $X(t)$ is said to be **strict-sense stationary** if (*) is true for all t_i, $i = 1, 2, ..., n$; $n = 1, 2, ...$ and *any* c.

2. Review of Stochastic Processes: Stationary (2)

For a **first-order strict sense stationary process**, from (*) we have

$$f_X(x,t) \equiv f_X(x,t+c)$$

for any c. In particular $c = -t$ gives

$$f_X(x,t) = f_X(x)$$

i.e., the first-order density of $X(t)$ is independent of t. In that case

$$E[X(t)] = \int_{-\infty}^{+\infty} x\, f(x) dx = \mu, \;\; a \; constant.$$

Similarly, for a **second-order strict-sense stationary process** we have from (*)

$$f_X(x_1, x_2, \; t_1, t_2) \equiv f_X(x_1, x_2, \; t_1 + c, t_2 + c)$$

for any c. For $c = -t_2$ we get

$$f_X(x_1, x_2, \; t_1, t_2) \equiv f_X(x_1, x_2, \; t_1 - t_2)$$

i.e., the second order density function of a strict sense stationary process depends only on the difference of the time indices $t_1 - t_2 = \tau$.

2. Review of Stochastic Processes: Stationary (3)

In that case the autocorrelation function is given by

$$\begin{aligned} R_{XX}(t_1,t_2) &\triangleq E\{X(t_1)X^*(t_2)\} \\ &= \int\int x_1 x_2^* f_X(x_1,x_2,\tau = t_1 - t_2)dx_1 dx_2 \\ &= R_{XX}(t_1 - t_2) \triangleq R_{XX}(\tau) = R_{XX}^*(-\tau), \end{aligned}$$

i.e., the autocorrelation function of a second order strict-sense stationary process depends only on the difference of the time indices τ.

However, the basic conditions for the first and second order stationarity are usually difficult to verify. In that case, we often resort to a looser definition of stationarity, known as **Wide-Sense Stationarity (W.S.S)**. Thus, a process $X(t)$ is said to be **Wide-Sense Stationary** if

(i) $E\{X(t)\} = \mu$

and

(ii) $E\{X(t_1)X^*(t_2)\} = R_{XX}(t_1 - t_2),$

2. Review of Stochastic Processes: Stationary (4)

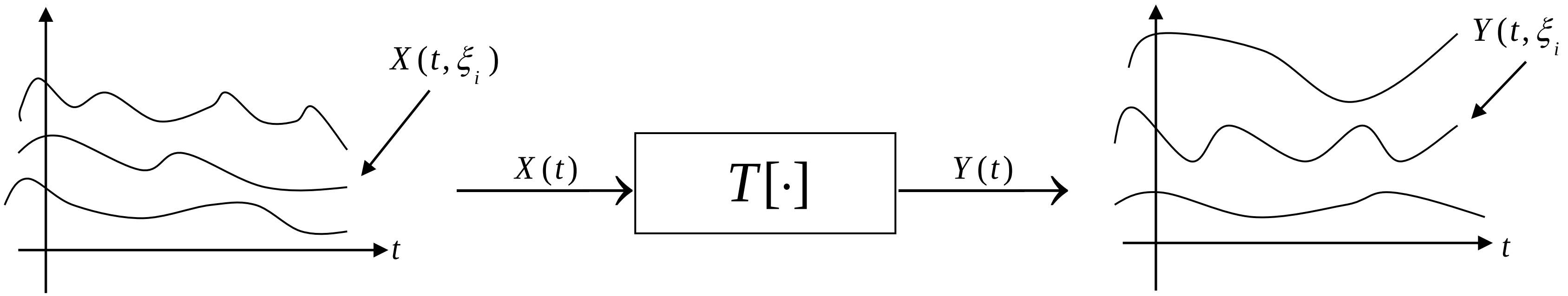
i.e., for wide-sense stationary processes, the mean is a constant and the autocorrelation function depends only on the difference between the time indices. Strict-sense stationarity always implies wide-sense stationarity. However, the converse is *not true* in general, the only exception being the Gaussian process. If $X(t)$ is a Gaussian process, then

$$\text{wide-sense stationarity (w.s.s)} \quad \Rightarrow \quad \text{strict-sense stationarity (s.s.s).}$$

2. Review of Stochastic Processes: Systems (1)

❑ A **deterministic system** transforms each input waveform $X(t, \zeta_i)$ into an output waveform $Y(t, \zeta_i) = T[X(t, \zeta_i)]$ by operating only on the time variable t. A **stochastic system** operates on both the variables t and ζ.

Thus, in deterministic system, a set of realizations at the input corresponding to a process $X(t)$ generates a new set of realizations $Y(t, \zeta)$ at the output associated with a new process $Y(t)$.



Our goal is to study the output process statistics in terms of the input process statistics and the system function.

2. Review of Stochastic Processes: Systems (2)

Deterministic Systems

Memoryless Systems

$$Y(t) = g[X(t)]$$

Systems with Memory

Time-varying systems

Time-invariant systems

Linear systems

$$Y(t) = L[X(t)]$$

Linear-Time Invariant (LTI) systems

$$X(t) \longrightarrow \boxed{h(t)} \longrightarrow Y(t) = \int_{-\infty}^{+\infty} h(t-\tau) X(\tau) d\tau$$

LTI system

$$= \int_{-\infty}^{+\infty} h(\tau) X(t-\tau) d\tau.$$

2. Review of Stochastic Processes: Systems (3)

❑ Memoryless Systems:

The output $Y(t)$ in this case depends only on the present value of the input $X(t)$.
i.e., $Y(t) = g\{X(t)\}$

Strict-sense stationary input → Memoryless system → Strict-sense stationary output.

Wide-sense stationary input → Memoryless system → Need *not* be stationary in any sense.

$X(t)$ stationary Gaussian with $R_{XX}(\tau)$ → Memoryless system → $Y(t)$ stationary,but *not* Gaussian with $R_{XY}(\tau) = \eta R_{XX}(\tau)$.

2. Review of Stochastic Processes: Systems (4)

Theorem: If $X(t)$ is a zero mean stationary Gaussian process, and $Y(t) = g[X(t)]$, where g(.) represents a nonlinear memoryless device, then

$$R_{XY}(\tau) = \eta R_{XX}(\tau), \qquad \eta = E\{g'(X)\}.$$

where $g'(x)$ is the derivative with respect to x.

❑ **Linear Systems**: $L[.]$ represents a linear system if

$$L\{a_1 X(t_1) + a_2 X(t_2)\} = a_1 L\{X(t_1)\} + a_2 L\{X(t_2)\}.$$

Let $Y(t) = L\{X(t)\}$ represent the output of a linear system.

❑ **Time-Invariant System**: $L[.]$ represents a time-invariant system if

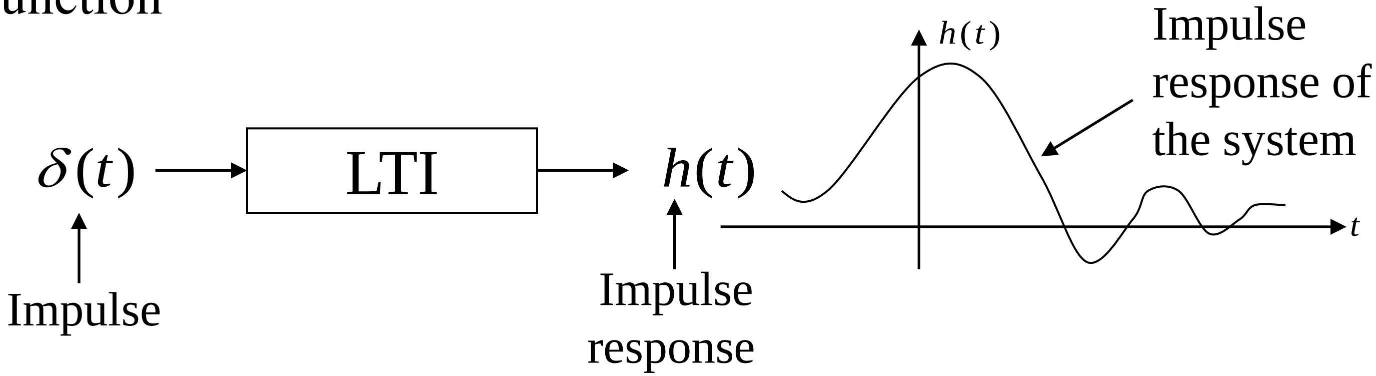
$$Y(t) = L\{X(t)\} \Rightarrow L\{X(t-t_0)\} = Y(t-t_0)$$

i.e., shift in the input results in the same shift in the output also.

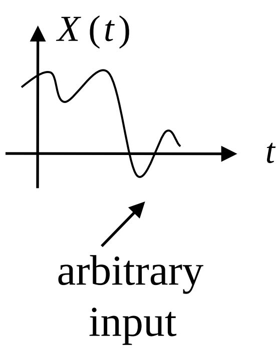
❑ If $L[.]$ satisfies both conditions for linear and time-invariant, then it corresponds to a **linear time-invariant (LTI) system**.

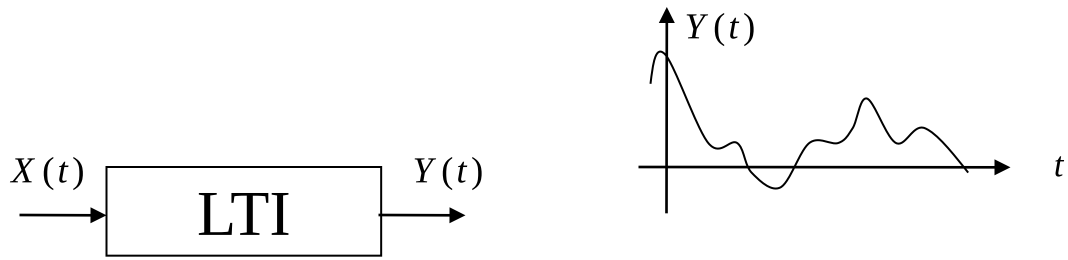
2. Review of Stochastic Processes: Systems (5)

LTI systems can be uniquely represented in terms of their output to a delta function



then





$$Y(t) = \int_{-\infty}^{+\infty} h(t-\tau)X(\tau)d\tau$$
$$= \int_{-\infty}^{+\infty} h(\tau)X(t-\tau)d\tau$$

where

$$X(t) = \int_{-\infty}^{+\infty} X(\tau)\delta(t-\tau)d\tau$$

2. Review of Stochastic Processes: Systems (6)

Thus

$$Y(t) = L\{X(t)\} = L\{\int_{-\infty}^{+\infty} X(\tau)\delta(t-\tau)d\tau\}$$

$$= \int_{-\infty}^{+\infty} L\{X(\tau)\delta(t-\tau)d\tau\}$$

By Linearity

$$= \int_{-\infty}^{+\infty} X(\tau)L\{\delta(t-\tau)\}d\tau$$

By Time-invariance

$$= \int_{-\infty}^{+\infty} X(\tau)h(t-\tau)d\tau = \int_{-\infty}^{+\infty} h(\tau)X(t-\tau)d\tau.$$

Then, the mean of the output process is given by

$$\mu_Y(t) = E\{Y(t)\} = \int_{-\infty}^{+\infty} E\{X(\tau)h(t-\tau)d\tau\}$$

$$= \int_{-\infty}^{+\infty} \mu_X(\tau)h(t-\tau)d\tau = \mu_X(t) * h(t).$$

2. Review of Stochastic Processes: Systems (7)

Similarly, the cross-correlation function between the input and output processes is given by

$$
\begin{aligned}
R_{XY}(t_1,t_2) &= E\{X(t_1)Y^*(t_2)\} \\
&= E\{X(t_1)\int_{-\infty}^{+\infty} X^*(t_2-\alpha)h^*(\alpha)d\alpha\} \\
&= \int_{-\infty}^{+\infty} E\{X(t_1)X^*(t_2-\alpha)\}h^*(\alpha)d\alpha \\
&= \int_{-\infty}^{+\infty} R_{XX}(t_1,t_2-\alpha)h^*(\alpha)d\alpha \\
&= R_{XX}(t_1,t_2) * h^*(t_2).
\end{aligned}
$$

2. Review of Stochastic Processes: Systems (8)

Finally the output autocorrelation function is given by

$$\begin{aligned} R_{YY}(t_1,t_2) &= E\{Y(t_1)Y^*(t_2)\} \\ &= E\{\int_{-\infty}^{+\infty} X(t_1-\beta)h(\beta)d\beta\, Y^*(t_2)\} \\ &= \int_{-\infty}^{+\infty} E\{X(t_1-\beta)Y^*(t_2)\}h(\beta)d\beta \\ &= \int_{-\infty}^{+\infty} R_{XY}(t_1-\beta,t_2)h(\beta)d\beta \\ &= R_{XY}(t_1,t_2)*h(t_1), \end{aligned}$$

or

$$R_{YY}(t_1,t_2) = R_{XX}(t_1,t_2)*h^*(t_2)*h(t_1).$$

In particular, if $X(t)$ is wide-sense stationary, then we have $\mu_X(t) = \mu_X$

Also

$$\begin{aligned} R_{XY}(t_1,t_2) &= \int_{-\infty}^{+\infty} R_{XX}(t_1-t_2+\alpha)h^*(\alpha)d\alpha \\ &= R_{XX}(\tau)*h^*(-\tau) \triangleq R_{XY}(\tau), \quad \tau = t_1 - t_2. \end{aligned}$$

2. Review of Stochastic Processes: Systems (9)

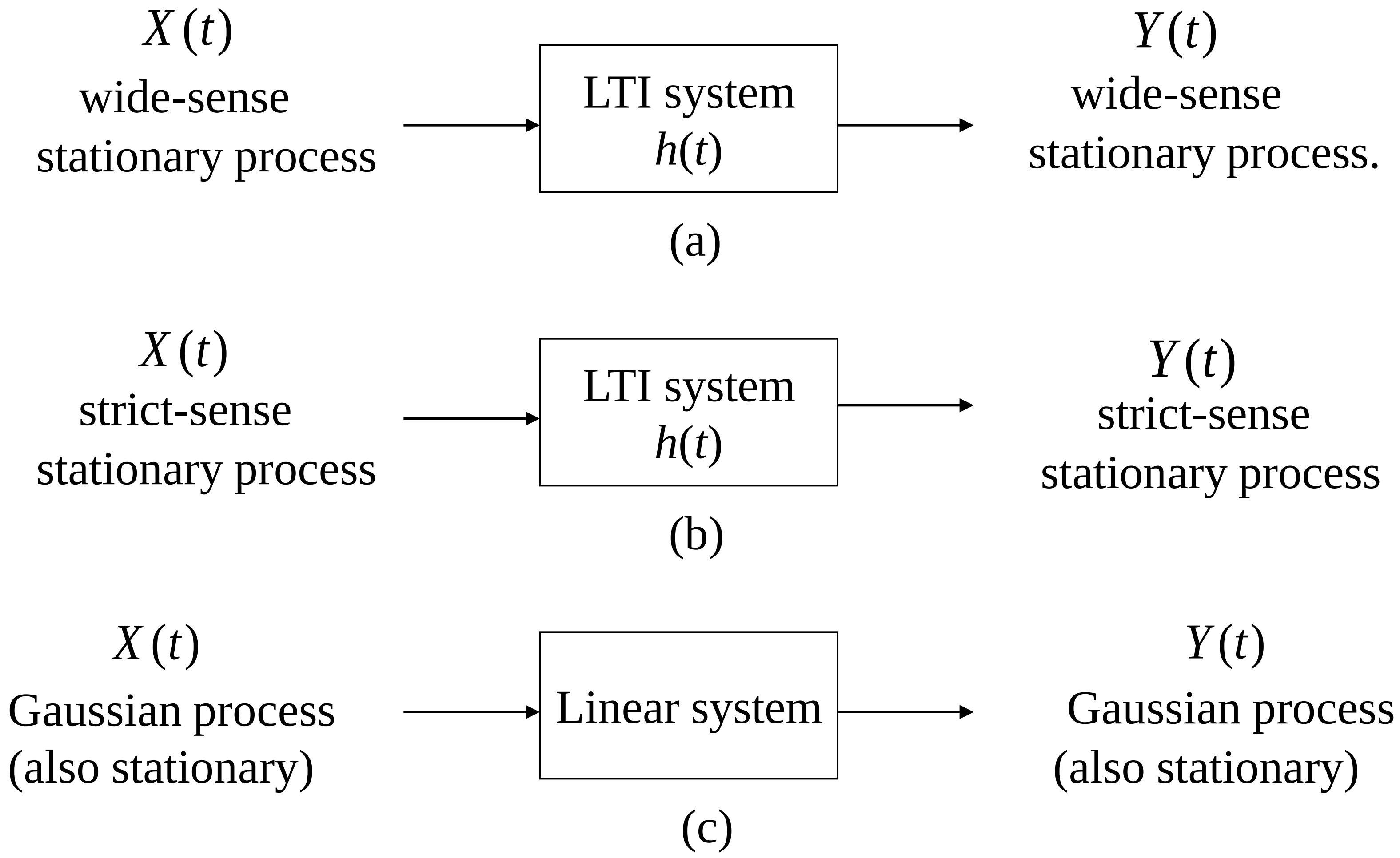
Thus $X(t)$ and $Y(t)$ are jointly w.s.s. Further, the output autocorrelation simplifies to

$$R_{YY}(t_1, t_2) = \int_{-\infty}^{+\infty} R_{XY}(t_1 - \beta - t_2) h(\beta) d\beta, \qquad \tau = t_1 - t_2$$

$$= R_{XY}(\tau) * h(\tau) = R_{YY}(\tau).$$

or

$$R_{YY}(\tau) = R_{XX}(\tau) * h^*(-\tau) * h(\tau).$$



$X(t)$
wide-sense
stationary process
LTI system
$h(t)$
$Y(t)$
wide-sense
stationary process.
(a)
$X(t)$
strict-sense
stationary process
LTI system
$h(t)$
$Y(t)$
strict-sense
stationary process
(b)
$X(t)$
Gaussian process
(also stationary)
Linear system
$Y(t)$
Gaussian process
(also stationary)
(c)

2. Review of Stochastic Processes: Systems (11)

❑ **White Noise Process:** $W(t)$ is said to be a white noise process if

$$R_{WW}(t_1,t_2)=q(t_1)\delta(t_1-t_2),$$

i.e., $E[W(t_1)\, W^*(t_2)] = 0$ unless $t_1 = t_2$.

$W(t)$ is said to be wide-sense stationary (w.s.s) white noise if
$E[W(t)]$ = constant, and

$$R_{WW}(t_1,t_2)=q\delta(t_1-t_2)=q\delta(\tau).$$

If $W(t)$ is also a Gaussian process (white Gaussian process), then all of its samples are independent random variables (why?).

For w.s.s. white noise input $W(t)$, we have

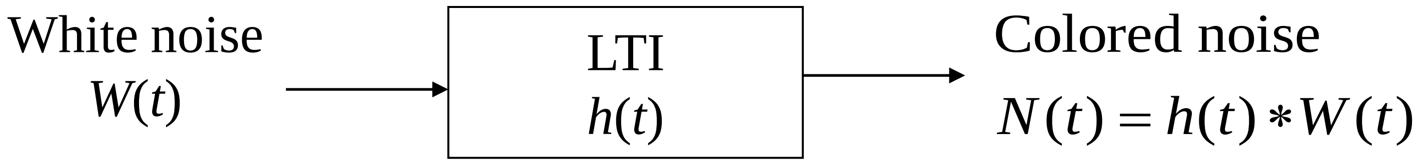
$$E[N(t)]=\mu_W\int_{-\infty}^{+\infty}h(\tau)d\tau,\quad a\ constant \quad \text{and} \quad R_{nn}(\tau)=q\delta(\tau)*h^*(-\tau)*h(\tau)$$
$$=qh^*(-\tau)*h(\tau)=q\rho(\tau)$$

where $\rho(\tau)=h(\tau)*h^*(-\tau)=\int_{-\infty}^{+\infty}h(\alpha)h^*(\alpha+\tau)d\alpha.$

2. Review of Stochastic Processes: Systems (12)

Thus the output of a white noise process through an LTI system represents a **(colored) noise process**.

Note: White noise need not be Gaussian.
"White" and "Gaussian" are two different concepts!



2. Review of Stochastic Processes: Discrete Time (1)

❑ A **discrete time stochastic process (DTStP)** $X_n = X(nT)$ is a sequence of random variables. The mean, autocorrelation and auto-covariance functions of a discrete-time process are gives by

$$\mu_n = E\{X(nT)\}$$

$$R(n_1, n_2) = E\{X(n_1T)X^*(n_2T)\}$$

$$C(n_1, n_2) = R(n_1, n_2) - \mu_{n_1}\mu_{n_2}^*$$

respectively. As before strict sense stationarity and wide-sense stationarity definitions apply here also. For example, $X(nT)$ is wide sense stationary if

$$E\{X(nT)\} = \mu, \quad a\ constant$$

and

$$E[X\{(k+n)T\}X^*\{(k)T\}] = R(n) = r_n \stackrel{\Delta}{=} r_{-n}^*$$

i.e., $R(n_1, n_2) = R(n_1 - n_2) = R^*(n_2 - n_1)$.

2. Review of Stochastic Processes: Discrete Time (2)

- If $X(nT)$ represents a wide-sense stationary input to a discrete-time system $\{h(nT)\}$, and $Y(nT)$ the system output, then as before the cross correlation function satisfies

$$R_{XY}(n) = R_{XX}(n) * h^*(-n)$$

and the output autocorrelation function is given by

$$R_{YY}(n) = R_{XY}(n) * h(n)$$

or

$$R_{YY}(n) = R_{XX}(n) * h^*(-n) * h(n).$$

Thus wide-sense stationarity from input to output is preserved for discrete-time systems also.

2. Review of Stochastic Processes: Discrete Time (3)

❑ **Mean** (or **ensemble average** μ) of a stochastic process is obtained by averaging **across** the process, while **time average** is obtained by averaging **along** the process as

$$\hat{\mu} = \frac{1}{M}\sum_{n=0}^{M-1} X_n$$

where M is total number of time samples used in estimation.

Consider wide-sense DTStP X_n, time average *converge* to ensemble average if:

$$\lim_{M\to\infty}\left\lfloor(\mu-\hat{\mu})^2\right\rfloor = 0$$

the process X_n is said **mean ergodic** (in the mean-square error sense).

2. Review of Stochastic Processes: Discrete Time (4)

❑ Let define an $(M\times1)$-**observation vector** $\mathbf{x}_n$ represents elements of time series $X_n, X_{n-1}, \ldots X_{n-M+1}$

$$\mathbf{x}_n = [X_n, X_{n-1}, \ldots X_{n-M+1}]^T$$

An $(M\times M)$-**correlation matrix R** (using condition of wide-sense stationary) can be defined as

$$\mathbf{R} = E\left[\mathbf{x}_n \mathbf{x}_n^H\right] = \begin{bmatrix} R(0) & R(1) & \cdots & R(M-1) \\ R(-1) & R(0) & \cdots & R(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ R(-M+1) & R(-M+2) & \cdots & R(0) \end{bmatrix}$$

Superscript H denotes Hermitian transposition.

2. Review of Stochastic Processes: Discrete Time (5)

Properties:

(i) Correlation matrix $\mathbf{R}$ of stationary DTStP is *Hermitian*: $\mathbf{R}^H = \mathbf{R}$ or $R(k)=R^*(k)$. Therefore:

$$\mathbf{R} = \begin{bmatrix} R(0) & R(1) & \cdots & R(M-1) \\ R^*(1) & R(0) & \cdots & R(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ R^*(M-1) & R^*(M-2) & \cdots & R(0) \end{bmatrix}$$

(ii) Matrix $\mathbf{R}$ of stationary DTStP is *Toeplitz*: all elements on main diagonal are equal, elements on any subdiagonal are also equal.

(iii) Let $\mathbf{x}$ be arbitrary (nonzero) $(M\times1)$-complex-valued vector, then $\mathbf{x}^H\mathbf{R}\mathbf{x} \geq 0$ (*nonnegative definition*).

(iv) If $\mathbf{x}^B{}_n$ is backward arrangement of $\mathbf{x}_n$: $\mathbf{x}_n^B = [x_{n-M+1}, x_{n-M+2}, \ldots, x_n]^T$

Then, $E[\mathbf{x}^B{}_n \mathbf{x}^{BH}{}_n] = \mathbf{R}^T$

2. Review of Stochastic Processes: Discrete Time (6)

(v) Consider correlation matrices $\mathbf{R}_M$ and $\mathbf{R}_{M+1}$, corresponding to M and $M+1$ observations of process, these matrices are related by

$$\mathbf{R}_{M+1} = \begin{bmatrix} R(0) & \mathbf{r}^H \\ \mathbf{r} & \mathbf{R}_M \end{bmatrix}$$

or

$$\mathbf{R}_{M+1} = \begin{bmatrix} \mathbf{R}_M & \mathbf{r}^{B*} \\ \mathbf{r}^{BT} & R(0) \end{bmatrix}$$

where $\mathbf{r}^H = [R(1), R(2),..., R(M)]$ and $\mathbf{r}^{BT} = [r(-M), r(-M+1),..., r(-1)]$

2. Review of Stochastic Processes: Discrete Time (7)

❑ Consider a time series consisting of complex sine wave plus noise:

$$u_n = u(n) = \alpha \exp(j\omega n) + v(n), \qquad n = 0, \dots, N-1$$

Sources of sine wave and noise are independent. Assumed that $v(n)$ has zero mean and autocorrelation function given by

$$E\left[v(n)v^*(n-k)\right] = \begin{cases} \sigma_v^2 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

For a lag k, autocorrelation function of process $u(n)$:

$$r(k) = E\left[u(n)u^*(n-k)\right] = \begin{cases} |\alpha|^2 + \sigma_v^2, & k = 0 \\ |\alpha|^2 e^{j\omega k}, & k \neq 0 \end{cases}$$

2. Review of Stochastic Processes: Discrete Time (8)

Therefore, correlation matrix of $u(n)$:

$$\mathbf{R} = |\alpha|^2 \begin{bmatrix} 1+\dfrac{1}{\rho} & \exp(j\omega) & \cdots & \exp(j\omega(M-1)) \\ \exp(-j\omega) & 1+\dfrac{1}{\rho} & \cdots & \exp(j\omega(M-2)) \\ \vdots & \vdots & \ddots & \vdots \\ \exp(-j\omega(M-1)) & \exp(-j\omega(M-2)) & \cdots & 1+\dfrac{1}{\rho} \end{bmatrix}$$

where $\rho = \dfrac{|\alpha|^2}{\sigma_v^2}$: *signal to noise ratio (SNR).*

2. Review of Stochastic Models (1)

❑ Consider an input – output representation

$$X(n) = -\sum_{k=1}^{p} a_k X(n-k) + \sum_{k=0}^{q} b_k W(n-k),$$

where $X(n)$ may be considered as the output of a system $\{h(n)\}$ driven by the input $W(n)$. Using z – transform, it gives

$$X(z)\sum_{k=0}^{p} a_k z^{-k} = W(z)\sum_{k=0}^{q} b_k z^{-k}, \quad a_0 \equiv 1$$

or

$$H(z) = \sum_{k=0}^{\infty} h(k) z^{-k} = \frac{X(z)}{W(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_q z^{-q}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_p z^{-p}} \triangleq \frac{B(z)}{A(z)}$$

represents the *transfer function* of the associated system response $\{h(n)\}$ so that

$$X(n) = \sum_{k=0}^{\infty} h(n-k) W(k).$$

$$W(n) \longrightarrow \boxed{h(n)} \longrightarrow X(n)$$

2. Review of Stochastic Models (2)

Notice that the transfer function $H(z)$ is rational with p poles and q zeros that determine the *model order* of the underlying system. The output $X(n)$ undergoes *regression* over p of *its previous values* and at the same time a *moving average* based on $W(n)$, $W(n$-1), ..., $W(n$-$q)$ of the input over (q + 1) values is added to it, thus generating an **Auto Regressive Moving Average** (ARMA (p, q)) process $X(n)$.

Generally the input $\{W(n)\}$ represents a sequence of uncorrelated random variables of zero mean and constant variance so that

$$R_{ww}(n) = \sigma_w^2 \delta(n).$$

If in addition, $\{W(n)\}$ is *normally distributed* then the output $\{X(n)\}$ also represents a *strict-sense stationary normal process*.

If $q = 0$, then $X(n)$ represents an **Auto Regressive** AR(p) process (all-pole process), and if $p = 0$, then $X(n)$ represents an **Moving Average** MA(q) process (all-zero process).

2. Review of Stochastic Models (3)

AR(1) process: An AR(1) process has the form

$$X(n) = aX(n-1) + W(n)$$

and the corresponding system transfer function

$$H(z) = \frac{1}{1-az^{-1}} = \sum_{n=0}^{\infty} a^n z^{-n}$$

provided $|a| < 1$. Thus

$$h(n) = a^n, \qquad |a| < 1$$

represents the impulse response of an AR(1) stable system. We get the output autocorrelation sequence of an AR(1) process to be

$$R_{XX}(n) = \sigma_W^2 \delta(n) * \{a^{-n}\} * \{a^n\} = \sigma_W^2 \sum_{k=0}^{\infty} a^{|n|+k} a^k = \sigma_W^2 \frac{a^{|n|}}{1-a^2}$$

The normalized (in terms of $R_{XX}(0)$) output autocorrelation sequence is given by

$$\rho_X(n) = \frac{R_{XX}(n)}{R_{XX}(0)} = a^{|n|}, \quad |n| \geq 0.$$

2. Review of Stochastic Models (4)

It is instructive to compare an AR(1) model discussed above by superimposing a random component to it, which may be an error term associated with observing a first order AR process $X(n)$. Thus

$$Y(n) = X(n) + V(n)$$

where $X(n) \sim$ AR(1), and $V(n)$ is an uncorrelated random sequence with zero mean and variance σ^2_V that is also uncorrelated with $\{W(n)\}$. Then, we obtain the output autocorrelation of the observed process $Y(n)$ to be

$$R_{YY}(n) = R_{XX}(n) + R_{VV}(n) = R_{XX}(n) + \sigma_V^2 \delta(n)$$

$$= \sigma_W^2 \frac{a^{|n|}}{1-a^2} + \sigma_V^2 \delta(n)$$

so that its normalized version is given by

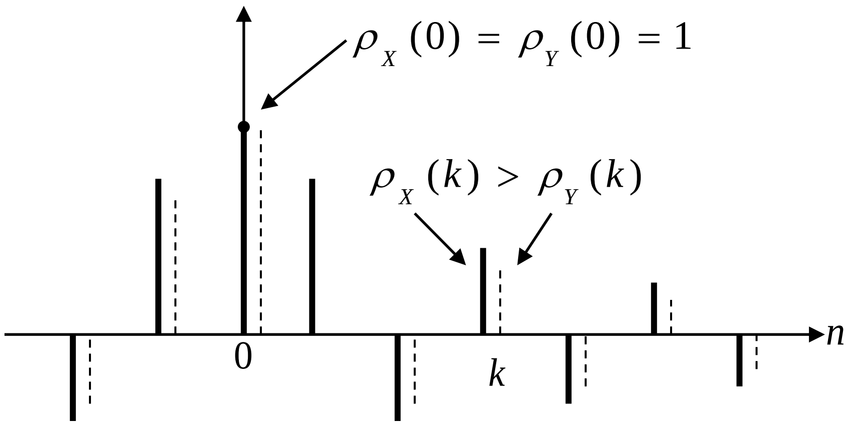
$$\rho_Y(n) \triangleq \frac{R_{YY}(n)}{R_{YY}(0)} = \begin{cases} 1 & n = 0 \\ c\, a^{|n|} & n = \pm 1, \ \pm 2, \cdots \end{cases}$$

where

$$c = \frac{\sigma_W^2}{\sigma_W^2 + \sigma_V^2 (1-a^2)} < 1.$$

2. Review of Stochastic Models (5)

The results demonstrate the effect of superimposing an error sequence on an AR(1) model. For non-zero lags, the autocorrelation of the observed sequence $\{Y(n)\}$ is reduced by a constant factor compared to the original process $\{X(n)\}$. The superimposed error sequence $V(n)$ only affects the corresponding term in $Y(n)$ (term by term). However, a particular term in the "input sequence" $W(n)$ affects $X(n)$ and $Y(n)$ as well as *all* subsequent observations.



2. Review of Stochastic Models (6)

AR(2) Process: An AR(2) process has the form

$$X(n) = a_1 X(n-1) + a_2 X(n-2) + W(n)$$

and the corresponding transfer function is given by

$$H(z) = \sum_{n=0}^{\infty} h(n) z^{-n} = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}} = \frac{b_1}{1 - \lambda_1 z^{-1}} + \frac{b_2}{1 - \lambda_2 z^{-1}}$$

so that

$$h(0) = 1,\ \ h(1) = a_1,\ \ h(n) = a_1 h(n-1) + a_2 h(n-2), \quad n \ge 2$$

and in term of the poles of the transfer function, we have

$$h(n) = b_1 \lambda_1^n + b_2 \lambda_2^n, \qquad n \ge 0$$

that represents the impulse response of the system. We also have

$$\lambda_1 + \lambda_2 = a_1, \qquad \lambda_1 \lambda_2 = -a_2,$$

$$b_1 + b_2 = 1, \qquad b_1 \lambda_1 + b_2 \lambda_2 = a_1.$$

and $H(z)$ stable implies $|\lambda_1| < 1,\ \ |\lambda_2| < 1.$

2. Review of Stochastic Models (7)

Further, the output autocorrelations satisfy the recursion

$$R_{xx}(n) = E\{X(n+m)X^*(m)\}$$

$$= E\{[a_1 X(n+m-1) + a_2 X(n+m-2)]X^*(m)\}$$

$$+ \overbrace{E\{W(n+m)X^*(m)\}}^{0}$$

$$= a_1 R_{xx}(n-1) + a_2 R_{xx}(n-2)$$

and hence their normalized version is given by

$$\rho_x(n) \triangleq \frac{R_{xx}(n)}{R_{xx}(0)} = a_1 \rho_x(n-1) + a_2 \rho_x(n-2).$$

By direct calculation using, the output autocorrelations are given by

$$R_{xx}(n) = R_{ww}(n) * h^*(-n) * h(n) = \sigma_w^2 h^*(-n) * h(n)$$

$$= \sigma_w^2 \sum_{k=0}^{\infty} h^*(n+k) * h(k)$$

$$= \sigma_w^2 \left(\frac{|b_1|^2 (\lambda_1^*)^n}{1-|\lambda_1|^2} + \frac{b_1^* b_2 (\lambda_1^*)^n}{1-\lambda_1^* \lambda_2} + \frac{b_1 b_2^* (\lambda_2^*)^n}{1-\lambda_1 \lambda_2^*} + \frac{|b_2|^2 (\lambda_2^*)^n}{1-|\lambda_2|^2} \right)$$

2. Review of Stochastic Models (8)

Then, the normalized output autocorrelations may be expressed as

$$\rho_X(n) = \frac{R_{XX}(n)}{R_{XX}(0)} = c_1 \lambda_1^{*n} + c_2 \lambda_2^{*n}$$

2. Review of Stochastic Models (9)

❑ An ARMA (p, q) system has only $p + q + 1$ independent coefficients, (a_k, $k = 1... p$; b_i, $i = 0... q$) and hence its impulse response sequence $\{h_k\}$ also must exhibit a similar dependence among them. In fact, according to P. Dienes (1931), and Kronecker (1881) states that the necessary and sufficient condition for $H(z) = \sum_{k=0}^{\infty} h_k z^{-k}$ to represent a rational system (ARMA) is that

$$\det H_n = 0, \qquad n \geq N \quad \text{(for all sufficiently large } n\text{)},$$

where

$$H_n \triangleq \begin{pmatrix} h_0 & h_1 & h_2 & \cdots & h_n \\ h_1 & h_2 & h_3 & \cdots & h_{n+1} \\ \vdots & & & & \\ h_n & h_{n+1} & h_{n+2} & \cdots & h_{2n} \end{pmatrix}.$$

i.e., in the case of rational systems for all sufficiently large n, the Hankel matrices H_n all have the same rank.