Part IV

Stochastic Image Analysis

Contents

Chapter 7

Introduction to Stochastic Processes

Every statistical analysis must be built upon a *mathematical model* linking observable reality with the mechanism generating the observations. This model should be a parsimonious description of nature: its functional form should be simple and the number of its parameters and components should be a minimum. The model should be parameterized in such a way that each parameter can be interpreted easily and identified with some aspect of reality. The functional form should be sufficiently tractable to permit the sort of mathematical manipulations required for the estimation of its parameters and other inference about nature.

Mathematical models may be divided into three general classes:

1) Purely deterministic.

2) Deterministic with simple random components.

3) Stochastic.

Examples:

- Observation using a TV camera of a falling object. Newtonian physics states that the distance traveled by a falling object is directly related to the squared time of fall. If atmospheric turbulence, observing errors and other transient effects can be ignored, the displacement can be calculated and predicted exactly.

- In the second kind of model each observation is a function of a strictly deterministic component and a random term, e.g. camera electrical noise.The random components are assumed not to depend one of another for different observations.

- Stochastic models are constructed from fundamental random events as components, e.g. the observation of a butterfly flight using a noisy camera.

Further, to define more precisely these approaches, we present an overview of the following concepts: probability, random variable, stochastic process, random signal and information.

7.1 Probability

"Probability theory is nothing but common sense reduced to calculations." (Laplace, 1819).

7.1.1 Random events and subjective experience

Definition 1. "The probability of one event is the ratio of the number of cases favorable to it, to the number of all cases possible when nothing leads us to expect that every one of these cases should occur more than any other, which renders them, for us, equally possible" (Laplace 1812).

Definition 2. "The probability is a representation of degrees of plausibility by real numbers" (Jeffrey, today's understanding) .

The first definition is the classical "$f$requentist" approach. The second is the modern "Bayesian" interpretation. In these lectures the first definition will be used, the second will be mentioned only few times to turn the attention to fields of advanced statistics.

7.1.2 Axioms of probability

The concept of probability Pr associated to an event E is further axiomatically introduced.

Set operation	Probability statement
$E_1 \cup E_2$	At least one of E_1 or E_2 occurs
$E_1 \cap E_2$	Both E_1 and E_2 occur
$\bar{E}$	E does not occur
$\emptyset$	The impossible event
$E_1 \cap E_2 = \emptyset$	E_1 and E_2 are mutually exclusive
$E_1 \subset E_2$	E_1 implies E_2

Table 7.1: Some set operations and corresponding probability statements

Axioms:

1) $\Pr(E) \geq 0$

2) $\Pr(I) = 1$

3) $\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2)$ if $E_1 \cup E_2 = \emptyset$ (sum axiom)

4) $\Pr(E_1 \cap E_2) = \Pr(E_2|E_1)\Pr(E_1) = \Pr(E_1|E_2)\Pr(E_2)$ (product axiom)

Observation The product axiom does not assume any exclusivity of the events. As a remainder the notation $\Pr(A|B)$ refers the probability of event A conditioned by B.

As consequence:

$$\Pr(\emptyset) = 0$$

The impossible event has zero probability. However it does not follow that an event of zero probability is impossible (e.g. the real numbers in the interval $[0, 1]$).

$$0 \leq \Pr(E) \leq 1$$

Remark These two conclusions are in the spirit of the second definition of the probability.

7.1.3 Statistical independence

Two events are statistical independent if

$$\Pr(E_1 \cap E_2) = \Pr(E_1)\Pr(E_2) \tag{7.1}$$

thus deducing that

$$\Pr(E_2|E_1) = \Pr(E_2) \tag{7.2}$$

and

$$\Pr(E_1|E_2) = \Pr(E_1). \tag{7.3}$$

The probability of an event is not influenced by the fact that another event takes place.

7.1.4 The Bayes formula

Consider a mutually exclusive and complete set of events $\{H_1, H_2, \ldots, H_n\}$ that is not independent of an event E in a certain experiment. We call the events H_i hypotheses and interpret them as hypothetical causes of the event E. The following decomposition formula can be written:

$$\Pr(E) = \sum_{i=1}^{n} \Pr(E|H_i)\Pr(H_i) \tag{7.4}$$

and also the formula of Bayes:

$$\Pr(H_i|E) = \frac{\Pr(E|H_i)\Pr(H_i)}{\Pr(E)}. \tag{7.5}$$

The probability $\Pr(H_i|E)$ is the probability satisfying the hypothesis H_i knowing that the event E was effectively produced. This is called "*a posteriori*" probability of H_i (one knows that E take place) and $\Pr(H_i)$ is called "*a priori*" probability.

Remark The Bayes formula can be understood as a formula for inverting conditional probabilities, i.e. compute $\Pr(H_i|E)$ given $\Pr(E|H_i)$ and $\Pr(H_i)$.

7.2 Random variables

Consider an experiment characterized by its elementary events supposed mutually independent and exclusive. A particular event consists of the union of several elementary events. Its probability is the sum of the probabilities of the elementary events. A random variables (r.v.) is defined by the biunivoque correspondence with an ensemble of elementary events and is characterized by the probability distribution of these events.

Example In an image the fact that a pixel value has a certain grey value is an elementary event characterized by its probability. Thus the ensemble of pixel grey levels, its intensity, is a random variable.

7.2.1 Distribution function and probability density function

Given a r.v. X defined on $(-\infty, \infty)$ the *distribution function*is defined as

$$F(x) = \Pr(X \leq x) \tag{7.6}$$

and has the following properties:

$$\begin{aligned} F(-\infty) &= 0; \\ F(\infty) &= 1 \\ F(b) - F(a) &= \Pr(a < X < b) \end{aligned} \tag{7.7}$$

and $F(x)$ is a monotonic non decreasing function. The probability density function p.d.f. is defined as

$$p(x) = \frac{dF(x)}{dx} \tag{7.8}$$

and is interpreted as

$$\Pr(x < X \leq x + dx) = p(x)dx. \tag{7.9}$$

A *distribution function* and a p.d.f. can be defined for a multidimensional[1] r.v. $X = (X_1, \ldots, X_n)^T$ as:

$$F(x_1, \ldots, x_n) = \Pr(X_1 \leq x_1, \ldots, X_n \leq x_n \tag{7.10}$$

and

$$p(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n} \tag{7.11}$$

knowing the interpretation

$$\Pr(x_1 < X_1 \leq x_1 + dx_1, \ldots, x_n < X_n \leq x_n + dx_n) = p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n. \tag{7.12}$$

Example In an image "interesting" patterns are e.g. lines. One considers the neighborhood of pixel ω (Fig. 7.1, and defines the p.d.f. of a line on the neighborhood variables. The joint distribution $p(x_\alpha, x_\beta, \ldots, x_\epsilon)$ captures the orientation of the line.

[1]Also called joint p.d.f. of $X_1, \ldots, X_n$.

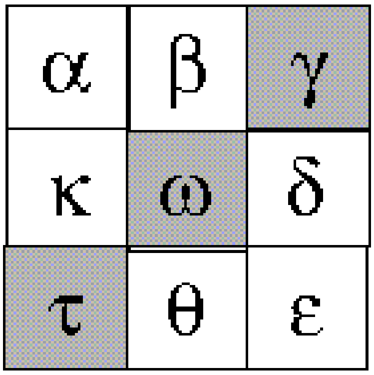

Figure 7.1: Example of 3x3 pixels neighborhood and diagonal image line.

Marginal p.d.f. Given a n-dimensional r.v. from which only $k < n$ components are of interest, a marginal p.d.f. is defined as:

$$p(x_1, \ldots x_k) = \int_{-\infty}^{+\infty} dx_{k+1} \cdots \int_{-\infty}^{+\infty} p(x_1, \ldots x_k, x_{k+1}, \ldots, x_n). \tag{7.13}$$

Example Assuming the knowledge of a joint p.d.f. characterizing edges on a 9 pixels neighborhood, the p.d.f. of the grey level in the center of the neighborhood is obtained as a marginal p.d.f. by integration over the 8 r.v. attached to the surrounding neighbors.

Conditional p.d.f. An n-dimensional joint p.d.f. is called conditional relative to $n-k$ variables if these $n-k$ variables have predefined values.

$$\begin{aligned} & p(x_1, \ldots, x_k \,|\, x_{k+1}, \ldots, x_n) = \\ & \Pr(x_1 < X_1 \le x_1 + dx_1, \ldots, x_k < X_k \le x_k + dx_k \,|\, X_{k+1} = x_{k+1}, \ldots, X_n = x_n). \end{aligned} \tag{7.14}$$

Using the Bayes formula one obtains:

$$p(x_1, \ldots, x_k | x_{k+1}, \ldots, x_n) = \frac{p(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n)}{p(x_{k+1}, \ldots, x_n)}. \tag{7.15}$$

In the more general case and using a vector notation $x = (x_1, \ldots, x_n)^T$, $y = (y_1, \ldots, y_n)^T$

$$\begin{aligned} p(x, y) &= p(x|y)\, p(y) = p(y|x)\, p(x) \\ p(x) &= \int_y p(x, y)\, dy = \int_y p(x|y) p(y)\, dy \\ p(y) &= \int_x p(x, y)\, dx = \int_x p(y|x) p(x)\, dx \end{aligned} \tag{7.16}$$

Statistical independence The generalization of the statistical independence gives:

$$p(x_1, \ldots, x_k \,|\, x_{k+1}, \ldots, x_n) = p(x_1, \ldots, x_k) \tag{7.17}$$

and

$$p(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) = p(x_1, \ldots, x_k)\, p(x_{k+1}, \ldots, x_n). \tag{7.18}$$

7.2.2 Expectation and moments

Given a function $f(X_1, \ldots, X_n)$ of the r.v. $X_1, \ldots, X_n$, the expectation operator $\mathcal{E}[\cdot]$ is introduced:

$$\mathcal{E}[f(X_1, \ldots, X_n)] = \int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_n f(x_1, \ldots, x_n) p(x_1, \ldots, x_n). \tag{7.19}$$

The expectation operator allows the definition of the moments of a r.v. For the 2-dimensional case we have:

$$\begin{aligned} m_1 &= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2\, x_1\, p(x_1, x_2) = \int_{-\infty}^{+\infty} dx_1\, x_1\, p(x_1) && (7.20) \\ m_2 &= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2\, x_2\, p(x_1, x_2) = \int_{-\infty}^{+\infty} dx_2\, x_2\, p(x_2) \\ R_1 &= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2\, x_1^2\, p(x_1, x_2) = \int_{-\infty}^{+\infty} dx_1\, x_1^2\, p(x_1) && (7.21) \\ R_{12} &= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2\, x_1 x_2\, p(x_1, x_2). \end{aligned}$$

The variance

$$\begin{aligned} \sigma_1^2 &= \mathcal{E}[(X_1 - m_1)^2] = R_1 - m_1^2 && (7.22) \\ \sigma_2^2 &= \mathcal{E}[(X_2 - m_2)^2] = R_2 - m_2^2, \end{aligned}$$

and the covariance

$$C_{12} = \mathcal{E}[(X_1 - m_1)(X_2 - m_2)] = R_{12} - m_1 m_2. \tag{7.23}$$

Remark If the r.v. are independent their covariance is zero.

$$\begin{aligned} C_{12} &= \mathcal{E}[(X_1 - m_1)(X_2 - m_2)] && (7.24) \\ &= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, (x_1 - m_1)(x_2 - m_2)\, p(x_1, x_2) \\ &= \int_{-\infty}^{\infty} dx_1\, (x_1 - m_1)\, p(x_1) \int_{-\infty}^{\infty} dx_2\, (x_2 - m_2)\, p(x_2) && (7.25) \\ &= 0 \end{aligned}$$

However if the covariance is zero one can not conclude that the r.v. are independent. A simple example follows.

Example We consider a r.v. Θ uniformly distributed on the interval $[0, 2\pi]$:

$$p(\theta) = \begin{cases} \frac{1}{2\pi} & \theta \in [0, 2\pi] \\ 0 & else \end{cases}. \tag{7.26}$$

We define two others r.v. as

$$\begin{aligned} X_1 &= \sin\theta && (7.27) \\ X_2 &= \cos\theta. \end{aligned}$$

We note that

$$\begin{aligned} m_1 &= \frac{1}{2\pi}\int_0^{2\pi} \sin\theta = 0 \\ m_2 &= \frac{1}{2\pi}\int_0^{2\pi} \cos\theta = 0. \end{aligned} \tag{7.28}$$

The covariance can be computed as:

$$\begin{aligned} c_{12} &= \int_0^{2\pi} \sin\theta\cos\theta\, p(\theta) d\theta \\ &= \frac{1}{4\pi}\int_0^{2\pi} \sin 2\theta\, d\theta = 0. \end{aligned} \tag{7.29}$$

We found that the r.v. X_1 and X_2 are not correlated. However we observe that X_1 and X_2 are strongly dependent

$$X_1^2 + X_2^2 = 1, \tag{7.30}$$

($\cos^2\theta + \sin^2\theta = 1$) but in a non linear relationship. The covariance can describe only linear interrelations.

7.2.3 The covariance matrix

We define a vector X of random variables

$$X = [X_1, \ldots, X_n]^T \tag{7.31}$$

and its mean

$$m_X = [m_1, \ldots, m_n]^T. \tag{7.32}$$

The covariance matrix C_X of X is defined as

$$C_X = \mathcal{E}\left[(X - m_X)(X - m_X)^T\right] \tag{7.33}$$

and has the form

$$C_X = \begin{bmatrix} \sigma_1^2 & c_{12} & \ldots & c_{1n} \\ c_{21} & \sigma_2^2 & \ldots & c_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ c_{n1} & c_{n2} & \ldots & \sigma_X^2 \end{bmatrix} \tag{7.34}$$

of a symmetric, squared, non negative matrix.

If the components of a random vector are independent, the covariance matrix is diagonal. However if the covariance matrix is diagonal there is non implication of independence of the two random vectors, only their components are mutually uncorrelated.

7.2.4 Complex random variables

A complex r.v. Z has two components, real X and imaginary Y, each at its turn being a r.v.

$$Z = X + jY. \tag{7.35}$$

The characterization of complex r.v. is done by a 2-dimensional p.d.f.: the joint p.d.f. of the real and imaginary parts. The mean of a complex r.v. Z is also complex and given by

$$m_Z = \mathcal{E}[Z] = \mathcal{E}[X + jY] = \mathcal{E}[X] + j\mathcal{E}[Y]. \tag{7.36}$$

The variance of a complex r.v. is

$$\sigma_Z^2 = \mathcal{E}[(Z - m_Z)(Z - m_Z)^*] = \sigma_X^2 + \sigma_Y^2, \tag{7.37}$$

which is real. The covariance of Z is defined as:

$$C_Z = C_X + C_Y + j(C_{YX} - C_{XY}). \tag{7.38}$$

7.2.5 Functions of random variables

We suppose a r.v. X of known p.d.f. $p(X)$ and define another r.v. Y

$$Y = f(X), \tag{7.39}$$

where $f(\cdot)$ is a bijection. We would like to find the p.d.f. $p(y)$ of Y. Knowing that $f(\cdot)$ is a bijective function

$$\Pr(x_0 < X \leq x_0 + dx) = \Pr(y_0 < Y \leq y_0 + dy) \tag{7.40}$$

where

$$y_0 = f(x_0) \tag{7.41}$$

and

$$dy = f'(x_0)dx. \tag{7.42}$$

Thus we obtain

$$|p(x)dx| = |p(y)dy| \tag{7.43}$$

and

$$p(y) = p(x)\left|\frac{dx}{dy}\right| = p(f^{-1}(y))\left|\frac{dx}{dy}\right| \tag{7.44}$$

which gives the p.d.f. of the r.v. Y.

In the multidimensional case with

$$Y_1 = f_1(X_1, \ldots, X_n) \tag{7.45}$$

$$\cdots = \cdots$$

$$Y_n = f_n(X_1, \ldots, X_n) \tag{7.46}$$

where f_i are continuous and differentiable functions and the Jacobian of the transformation is non negative. We have:

$$|p(x_1, \ldots, x_n)\, dx_1 \ldots dx_n| = |p(y_1, \ldots, y_n)\, dy_1 \ldots dy_n|. \tag{7.47}$$

The p.d.f. of the r.v. after the transformation will be:

$$p(y_1, \ldots, y_n) = p(x_1, \ldots, x_n)\left|\frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)}\right| \tag{7.48}$$

where the Jacobian is defined by:

$$\frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \cdots & \cdots & \cdots \\ \frac{\partial x_1}{\partial y_n} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}. \tag{7.49}$$

Example To exemplification we use the transformation from cartesian to polar coordinates in a 2-dimensional space.

$$\begin{aligned} Y_1 &= \sqrt{X_1^2 + X_2^2} \\ Y_2 &= \arctan \frac{X_2}{X_1} \end{aligned} \tag{7.50}$$

The inverse transformation is:

$$\begin{aligned} X_1 &= Y_1 \cos Y_2 \\ X_2 &= Y_1 \sin Y_2 \end{aligned} \tag{7.51}$$

and has the Jacobian

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{vmatrix} = \rho \tag{7.52}$$

when

$$\begin{aligned} Y_1 &= \rho \\ Y_2 &= \theta. \end{aligned} \tag{7.53}$$

Given the p.d.f. of the r.v. X_1, X_2 in cartesian coordinates $p(X_1, X_2)$ one finds the p.d.f. in polar coordinates

$$\begin{aligned} p(\rho, \theta) &= \rho\, p(x_1, x_2) \\ p(\rho, \theta) &= \rho\, p(\rho\cos\theta, \rho\sin\theta) \end{aligned} \tag{7.54}$$

with

$$\rho > 0 \text{ and } 0 \leq \theta \leq 2\pi. \tag{7.55}$$

Further we consider an important case, X_1 and X_2 are independent r.v. with normal p.d.f. of zero mean and the same variance σ^2:

$$\begin{aligned} p(x_1, x_2) &= p(x_1)\, p(x_2) \\ &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x_1^2 + x_2^2}{2\sigma^2}\right). \end{aligned} \tag{7.56}$$

For exemplification in Fig. 7.2.5 are presented the real X_1 and imaginary X_2 parts of a Synthetic Aperture Radar (SAR) image[2] and their p.d.f.ś.

In polar coordinates the p.d.f. is obtained in the form

$$p(\rho, \theta) = \rho \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \tag{7.57}$$

where

$$p(\rho) = \frac{\rho}{\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \tag{7.58}$$

[2]Synthetic Aperture Radar images are obtained by coherently illuminating a scene with microwave radiation. SAR images are in the class of important images as laser, sonar, ecographical, computer tomography.

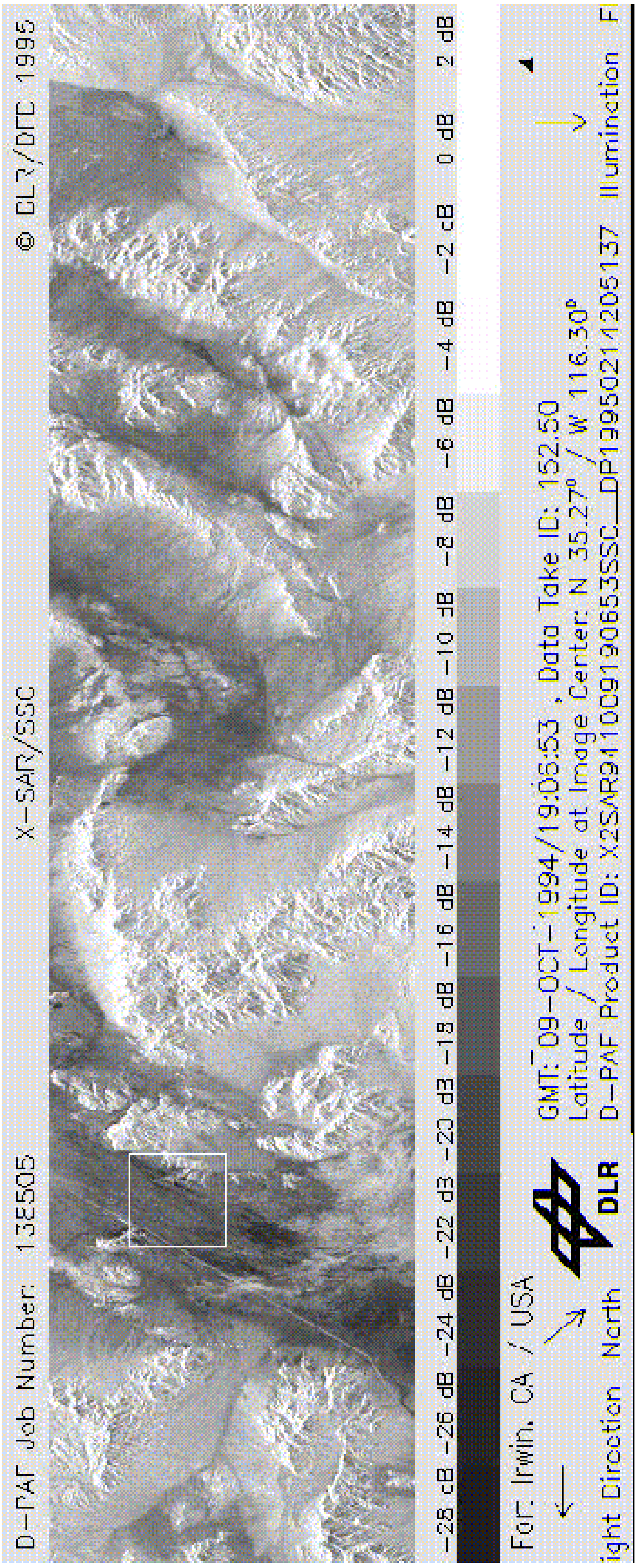

Figure 7.2: Synthetic Aperture Radar image, part of which (marked) was used for the example of section 7.2.5

is a Rayleigh distribution and

$$p(\theta) = \frac{1}{2\pi} \tag{7.59}$$

is a uniform distribution. The images in polar coordinates and their p.d.f.'s are presented in Fig. 7.2.5. The "ρ-image" is the amplitude image and is Rayleigh distributed. The θ-image is the phase image and is uniformly distributed.

Remark

- In this example we can observe that the "information" is contained in the amplitude image. The phase image has the maximum incertitude, there is no preferred phase value.
- The Rayleigh distribution of the intensity suggests that such an image can be modeled as a noise-less amplitude *multiplied* with a noise. That is why is interesting to consider the exercise to compute the p.d.f. of the logarithmic transformed Rayleigh distributed r.v.The logarithm transforms the multiplicative nature of the noise in an additive one.

Linear transformation of r.v. Further we consider the vectors of r.v. $X = [X_1, \ldots, X_n]^T$ and $Y = [Y_1, \ldots, Y_n]$. Without loss of generality we consider the r.v. to have zero mean.

A linear transformation is given by

$$Y = AX \tag{7.60}$$

where $A(n \times n)$ is the matrix characterizing the transformation. With simple computations we obtain:

$$\begin{aligned} \mathcal{E}[Y] &= A\,\mathcal{E}[X] && (7.61)\\ C_Y &= \mathcal{E}[YY^T] = A\,\mathcal{E}[XX^T]\,A^T = AC_XA^T \\ C_{XY} &= \mathcal{E}[XY^T] = \mathcal{E}[XX^T]A^T = C_XA^T \\ C_{XY} &= C_{YX}^T && (7.62) \end{aligned}$$

An important linear transformation is the Karhunen-Loève transform. We can write

$$C_X = M\Lambda M^T \tag{7.63}$$

where Λ is the diagonal matrix of eigenvalues of C_X, and M is the orthogonal matrix having the eigenvectors of C_X as columns. It follows:

$$\Lambda = M^TC_XM. \tag{7.64}$$

The Karhunen-Loève transform has the matrix

$$A = M^T \tag{7.65}$$

and transforms X in Y having a diagonal matrix C_Y. The Karhunen-Loève transform uncorrelates the components of the r.v. vector X.

7.3 Stochastic processes

7.3.1 Definition

In the previous section we introduced two concepts: random variables and vector of random variables. A r.v. was defined by a biunivoque correspondence to the elementary events obtained as a result of an experiment. The vector of r.v.'s was introduced as a formal notation, but now we give a more precise meaning: each component is in correspondence with a time moment[3]. In this way a stochastic process is defined by the biunivoque correspondence of the results of an experience with time (space) functions. We can interpret a stochastic process as a process to generate random functions. Each random function is a particular realization of the stochastic process called random signal. Thus a stochastic process is the ensemble of random signals which evolve in time (Fig. 7.3.1).

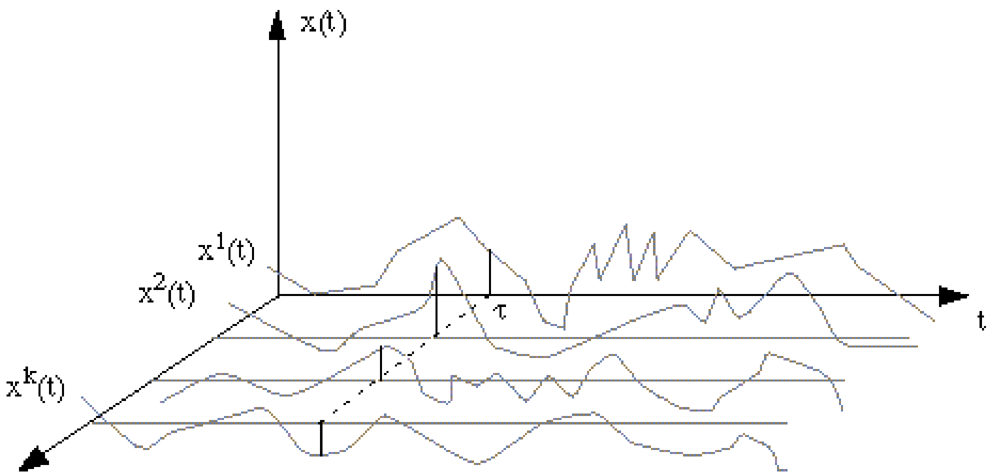

Figure 7.3: Exemplification of a stochastic process $x(t)$. Each $x^k(t)$, for a fixed value k, is a particular realization, a random signal. Each ensemble $\{x^1(t), x^2(t), \ldots\}$ for a given time moment t, is a random variable. A stochastic process is characterized by its across-realizations behavior for all time moments.

The problematic of stochastic processes is to make evidence and to characterize not only the individual r.v. at fixed moments of time, but also the relationship and statistical dependencies in between r.v. obtained at different time measurements.

Stochastic image analysis considers images as particular realizations of a spatial stochastic process and aims at the local and neighborhood analysis of the statistical dependencies of the pixel intensities.

[3]Historically time is the variable used in signal theory. In these lectures time will be substitute to space $t \to (i,j)$, where (i,j) are the coordinates of a pixel in an image.

7.3.2 Characterization of a stochastic process

A stochastic process is fully determined by the knowledge of all p.d.f.'s

$$\lim_{n\to\infty} p(x(t_1),\ldots,x(t_n)) \tag{7.66}$$

where $x(t_k)$ is a r.v. obtained sampling the process at the time t_k.

7.3.3 Statistical vs. time characterization of a random signal

The statistical characterization is done at a fixed time moment t_k across all the realizations of the stochastic process. It is the characterization of a r.v. $x(t_k)$.

The time characterization refers one particular representation of the stochastic process $x^{(k)}(t)$ as a time dependent random signal. The following table presents comparatively the stochastic and temporal characterization of random signals in the first and second order statistics.

	Stochastic	Temporal
Mean	$m_X(t) = \mathcal{E}[X(t)] = \int_{-\infty}^{+\infty} xp(x(t))dx$	$\eta^{(k)}(t_0) = \lim_{T\to\infty} \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} x^{(k)}(t_0+t)dt$
Variance**	$\sigma_X^2 = \mathcal{E}[X(t)X^T(t)]$	$\mathrm{var}^{(k)}(t_0) = \lim_{T\to\infty} \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} [x^{(k)}(t_0+t)]^2 dt$
Covariance**	$C_{XX}(t_1,t_2) = \mathcal{E}[X(t_1)X^T(t_2)]$	$\mathrm{R}_{XX}^{(k)}(t_1,t_2) = \lim_{T\to\infty} \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} x^{(k)}(t_1+t)x^{(k)}(t_2+t)dt$ (autocorrelation)
Mutual covariance	$C_{XY}(t_1,t_2) = \mathcal{E}[X(t_1)Y^T(t_2)]$	$\mathrm{R}_{XY}^{(k)}(t_1,t_2) = \lim_{T\to\infty} \frac{1}{T}\int_{-\frac{T}{2}}^{+\frac{T}{2}} x^{(k)}(t_1+t)y^{(k)}(t_2+t)dt$ (cross correlation)

(**) For simplicity the mean is assumed to be zero.

7.3.4 Stationary processes

A stochastic process is called *strict-sense stationary* if all its statistical properties are invariant to a shift of the time origin.

A stochastic process is called *wide-sense stationary* if its mean is constant and its autocorrelation depends only on the distance $t_1 - t_2$

$$\begin{aligned} \mathcal{E}[x(t)] &= \text{constant} \\ C_{XX} &= C_{XX}(t_1-t_2). \end{aligned} \tag{7.67}$$

Example We consider a random variable Φ uniformly distributed on the interval $[0, 2\pi]$ and define the random vector $X(t)$

$$X(t) = \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega t + \varphi) \\ \sin(\omega t + \varphi) \end{bmatrix}. \tag{7.68}$$

The expectations are

$$\begin{aligned}\mathcal{E}[X_1(t)] &= 0 \\ \mathcal{E}[X_2(t)] &= 0.\end{aligned} \tag{7.69}$$

Thus the covariance of $X(t)$ can be evaluated as

$$\begin{aligned}C_X(t_1,t_2) &= \mathcal{E}\left[\begin{bmatrix} X_1(t_1) \\ X_2(t_2)\end{bmatrix}\begin{bmatrix} X_1(t_1) & X_2(t_2)\end{bmatrix}\right] \\ &= \begin{bmatrix} C_{X_1}(t_1,t_2) & C_{X_1,X_2}(t_1,t_2) \\ C_{X_2,X_1}(t_1,t_2) & C_{X_2}(t_1,t_2)\end{bmatrix}.\end{aligned} \tag{7.70}$$

And further:

$$\begin{aligned}C_{X_1}(t_1,t_2) &= \mathcal{E}[X_1(t_1)X_2(t_2)] \\ &= \mathcal{E}[\cos(\omega t_1+\varphi)\cos(\omega t_2+\varphi)] \\ &= \frac{1}{2}\cos[\omega(t_2-t_1)].\end{aligned} \tag{7.71}$$

In similar way

$$C_{X_2}(t_1,t_2) = \frac{1}{2}\cos[\omega(t_1-t_2)] \tag{7.72}$$

and

$$C_{X_1,X_2}(t_1,t_2) = -\frac{1}{2}\sin[\omega(t_1-t_2)]. \tag{7.73}$$

Because the mean is zero and the covariance matrix depends only on the time difference $t_2 - t_1$ we conclude that the process $X(t)$ is wide-sense stationary.

7.3.5 Ergodicity

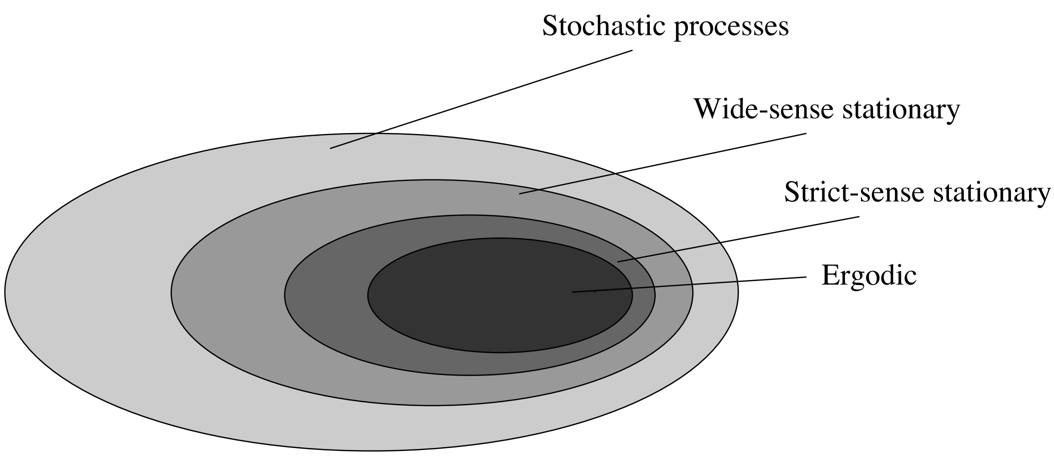

Figure 7.4: Classification of stochastic processes.

We already have seen from the definition of a stochastic process that two modalities of analysis are possible:

- stochastic
- temporal.

Ergodicity is the theory dealing with the study of stochastic vs. temporal characterization of stochastic processes, e.g. is a temporal average equal with a stochastic expectation?

A stochastic process is called ergodic if all its statistical properties can be determined from a single realization.

Further we comment only a limited case: the ergodicity of the mean. Supposing that $X(t)$ is a stochastic process with constant mean

$$\mathcal{E}[x(t)] = \eta, \tag{7.74}$$

we study under what condition the time average

$$\eta_T = \frac{1}{2T}\int_{-T}^{T} x(t)dt \tag{7.75}$$

is close to η. The necessary and sufficient condition for the process $X(t)$ to be mean-ergodic is

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T\left(1-\frac{|\tau|}{T}C_X(\tau)d\tau\right) = 0. \tag{7.76}$$

Remarks

- Ergodicity necessarily requires strict-sense stationary. The classification of stochastic processes presented in Fig. 7.3.5
- In practice is very difficult to have access to enough realizations of a stochastic process, thus the ensemble statistical analysis is not possible. It remain only the possibility for temporal analysis of one realization.

7.3.6 Spectral analysis of random signals

The harmonic analysis of deterministic functions can not be directly applied to random signals because in general

$$\int_{-\infty}^{+\infty} |x^{(k)}(t)dt| \to \infty. \tag{7.77}$$

The power spectrum density concept is introduced:

$$\phi(\omega) = \lim_{T\to\infty}\frac{\mathcal{E}[|X_T^{(k)}(\omega)|^2]}{T} \tag{7.78}$$

where $X_T^{(k)}(\omega)$ is the Fourier transform of the realization $X^{(k)}$ of the stochastic process $X(t)$, restricted to the time interval T. One of the basic results of the spectral analysis of random signals is the *Wiener-Khintchine theorem*: the power spectrum density is the Fourier transform of the autocorrelation function

$$\phi(\omega) = \int_{-\infty}^{+\infty} R_{XX}(\tau)\exp(-j\omega\tau)d\tau. \tag{7.79}$$

The Wiener-Khintchine theorem was reminded to show the connection of stochastic and spectral analyses.

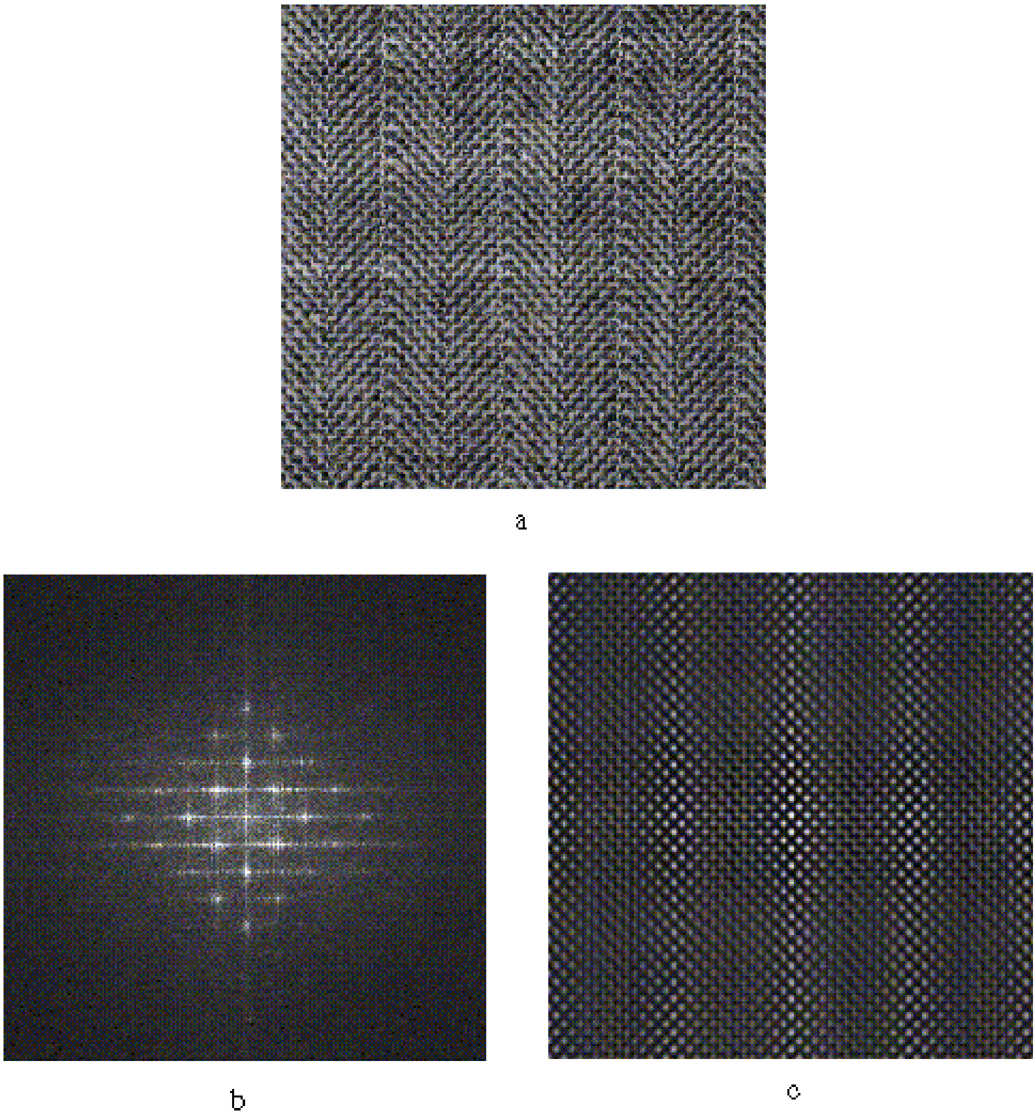

Figure 7.5: Exemplification of the Wiener-Khintchine theorem. a) texture image from Brodatz album, b) the power spectrum density of the image and, c) the autocorrelation function of the image. One could note the strong quasi-periodicity of the texture reflected in the stochastic characterization of it, power spectrum density and autocorrelation function. The signals in images b) and c) are Fourier pairs.

7.4 Outlook

This introduction was a short recapitulation of some of the basics concepts in statistics which will be used in the next chapters.

Three nested notions have been referred: probabilistic event, random variable and stochastic process.

The random signal (image) was defined as being a particular realization of a stochastic process.

The statistical independence and correlation have been commented, intuitively showing the importance of nonlinear interrelationships.

Stationarity and ergodicity have been introduced from an "engineering" point of view.

Bibliography

[1] Al. Spataru, *Fondéments de la théorie de la transmission de l'information,* Presses Polytechniques Romandes , 1987

[2] A. K. Jain, *Fundamentals of Digital Image Processing,* Prentice Hall, 1989.

[3] J.-R. Ohm, *Digitale Bildcodierung,* Springer, 1995.

[4] D. Cox, H. Miller, *The theory of stochastic processes,* Methuen, 1965

[5] A. Papoulis, *Probability, random variables and stochastic processes,* McGraw-Hill, 1991.