3 Random Variables and Fields

3.1 Introduction

Digital image processing can be regarded as a subarea of *digital signal processing*. As such, all the methods for taking and analyzing measurements and their errors can also be applied to image processing. In particular, any measurement we take from images — e. g., the size or the position of an object or its mean gray value — can only be determined with a certain precision and is only useful if we can also estimate its uncertainty. This basic fact, which is well known to any scientist and engineer, was often neglected in the initial days of image processing. Using empirical and ill-founded techniques made reliable error estimates impossible. Fortunately, knowledge in image processing has advanced considerably. Nowadays, many sound image processing techniques are available that include reliable error estimates.

In this respect, it is necessary to distinguish two important classes of errors. The *statistical error* describes the scatter of the measured value if one and the same measurement is repeated over and over again as illustrated in Fig. 3.1. A suitable measure for the width of the distribution gives the statistical error and its centroid, the mean measured value.

This mean value may, however, be much further off the true value than given by the statistical error margins. Such a deviation is called a *systematic error*. Closely related to the difference between systematic and statistical errors are the terms *precise* and *accurate*. A precise but inaccurate measurement is encountered when the statistical error is low but the systematic error is high (Fig. 3.1a). If the reverse is true, i. e., the statistical error is large and the systematic error is low, the individual measurements scatter widely but their mean value is close to the true value (Fig. 3.1b).

It is easy — at least in principle — to get an estimate of the statistical error by repeating the same measurement many times. But it is much harder to control systematic errors. They are often related to a lack in understanding of the measuring setup and procedure. Unknown or uncontrolled parameters influencing the measuring procedure may easily lead to systematic errors. Typical sources of systematic errors are *calibration errors* or temperature-dependent changes of a parameter in an experimental setup without temperature control.

B. Jähne, Digital Image Processing
ISBN 3-540-24035-7

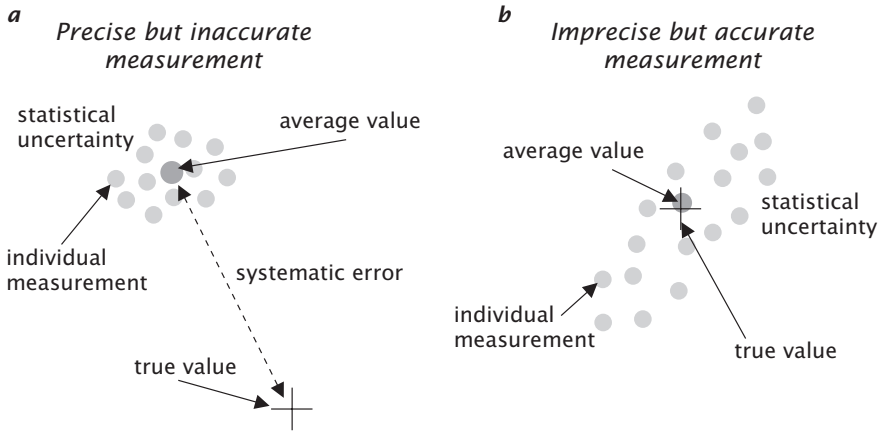

Figure 3.1: *Illustration of* ***a*** *systematic and* ***b*** *statistical error distinguishing precision and accuracy for the measurement of position in 2-D images. The statistical error is given by the distribution of the individual measurements, while the systematic error is the difference between the true value and the average of the measured values.*

In this chapter, we learn how to handle image data as statistical quantities or random variables. We start with the statistical properties of the measured gray value at an individual *sensor element* or *pixel* in Section 3.2. Then we can apply the classical concepts of statistics used to handle point measurements. These techniques are commonly used in most scientific disciplines. The type of statistics used is also known as *first-order statistics* because it considers only the statistics of a single measuring point.

Image processing operations take the measured gray values to compute new quantities. In the simplest case, only the gray value at a single point is taken as an input by so-called *point operations*. In more complex cases, the gray values from many pixels are taken to compute a new point. In any case, we need to know how the statistical properties, especially the precision of the computed quantity depends on the precision of the gray values taken to compute this quantity. In other words, we need to establish how errors are propagating through image processing operations. Therefore, the topic of Section 3.3 is multiple random variables and error propagation.

As a last step, we turn to time series of random variables (*stochastic processes*) and spatial arrays of random variables (*random fields*) in Section 3.5. This allows us to discuss random processes in the Fourier domain.

3.2 Random Variables

3.2.1 Probability Density Functions and Histograms

Imagine an experimental setup in which we are imaging a certain object. The measured quantity at a certain point in the image plane (a pixel) is the irradiance. Because of the statistical nature of the observed process, each measurement will give a different value.

This means that the observed signal is not characterized by a single value but rather a *probability density function* (*PDF*) $f(g)$. This function indicates the probability of observing the value g. A measurable quantity which is governed by a random process is denoted as a *random variable*, or short *RV*.

In the following, we discuss continuous and discrete random variables and probability functions together. We need discrete probabilities as only discrete numbers can be handled by a digital computer. Discrete values are obtained after a process called *quantization* which was introduced in Section 2.2.4. Many equations in this section contain a continuous formulation on the left side and their discrete counterparts on the right side. In the continuous case, $f(g)\mathrm{d}g$ gives the non-negative probability to measure a value within the interval g and $g + \mathrm{d}g$. In the discrete case, we can only measure a finite number, Q, of values g_q ($q = 1, 2, \ldots, Q$) with probability f_q. Normally, the value of a pixel is stored in one byte so that we can measure $Q = 256$ different gray values. As the total probability to observe any value at all is 1 by definition, the PDF must meet the requirement

$$\int_{-\infty}^{\infty} f(g)\mathrm{d}g = 1, \qquad \sum_{q=1}^{Q} f_q = 1. \tag{3.1}$$

The integral of the PDF

$$F(g) = \int_{-\infty}^{g} f(g')\mathrm{d}g', \qquad F_q = \sum_{q'=1}^{q} f_{q'} \tag{3.2}$$

is known as the *distribution function*. Because the PDF is a non-negative function, the distribution function increases monotonically from 0 to 1.

Generally, the probability distribution is not known a priori. Rather it is estimated from measurements. If the observed process is *homogeneous*, that is, if it does not depend on the position of the pixel in the image, there is a simple way to estimate the PDF using a *histogram*.

A histogram of an image is a list (vector) that contains one element for each quantization level. Each element contains the number of pixels whose gray value corresponds to the index of the element. Histograms can be calculated easily for data of any dimension. First we set the whole

histogram vector to zero. Then we scan each pixel of the image, match its gray value to an index in the list, and increment the corresponding element of the list by one. The actual scanning algorithm depends on how the image is stored.

An estimate of the probability density function can also be obtained for image data with higher resolution, for instance 16-bit images or floating-point images. Then the range of possible values is partitioned into Q equally wide bins. The value associated with the bin is the center of the bin, whereas we take the values in between the bins to decide which bin is to be incremented. If we do not make this distinction, values computed from the histogram, such as mean values, are biased.

3.2.2 Mean, Variance, and Moments

The two basic parameters that describe a RV g are its *mean* (also known as the *expectation value*) and its *variance*. The mean $\mu = Eg$ is defined as

$$\boxed{\mu = \int_{-\infty}^{\infty} g f(g) \mathrm{d}g, \qquad \mu = \sum_{q=1}^{Q} g_q f_q.} \tag{3.3}$$

The mean can also be determined without knowing the probability density function explictly by averaging an infinite number of measurements:

$$\mu = \lim_{P \to \infty} \frac{1}{P} \sum_{p=1}^{P} g_p. \tag{3.4}$$

As we cannot take an infinite number of measurements, the determination of the mean by Eq. (3.4) remains an estimate with a residual uncertainty that depends on the form of the PDF, i. e., the type of the random process and the number of measurements taken.

The *variance* $\sigma^2 = \operatorname{var} g = E\left((g - \mu)^2\right)$ is a measure of the extent to which the measured values deviate from the mean value:

$$\boxed{\sigma^2 = \int_{-\infty}^{\infty} (g - \mu)^2 f(g) \mathrm{d}g, \qquad \sigma^2 = \sum_{q=1}^{Q} (g_q - \mu)^2 f_q.} \tag{3.5}$$

The PDF can be characterized in more detail by quantities similar to the variance, the *central moments* of nth order $\mu_n = E\left((g - \mu)^n\right)$:

$$\mu_n = \int_{-\infty}^{\infty} (g - \mu)^n f(g) \mathrm{d}g, \qquad \mu_n = \sum_{q=1}^{Q} (g_q - \mu)^n f_q. \tag{3.6}$$

The first central moment is by definition zero. The second moment μ_2 is the variance σ^2. The third moment μ_3, the *skewness*, is a measure

for the asymmetry of the PDF around the mean value. If the PDF is a function of even symmetry, $f(-(g-\mu)) = f(g-\mu)$, the third and all higher-order odd moments vanish.

3.2.3 Functions of Random Variables

Any image processing operation changes the signal g at the individual pixels. In the simplest case, g at each pixel is transformed into h by a function p: $h = p(g)$. Such a function is known in image processing as a *point operator*. Because g is a RV, h will also be a RV and we need to know its PDF in order to know the statistical properties of the image after processing it.

It is obvious that the PDF f_h of h has same form as the PDF f_g of g if p is a linear function: $h = a_0 + a_1 g$:

$$f_h(h) = \frac{f_g(g)}{|a_1|} = \frac{f_g((h-a_0)/a_1)}{|a_1|}, \tag{3.7}$$

where the inverse linear relation $g = p^{-1}(h) : g = (h-a_0)/a_1$ is used to express g as a function of h.

From Eq. (3.7) it is intuitive that in the general case of a nonlinear function $p(g)$, the slope a_1 will be replaced by the first derivative $p'(g)$ of $p(g)$. Further complications arise if the inverse function has more than one branch. A simple and important example is the function $h = g^2$ with the two inverse functions $g_{1,2} = \pm\sqrt{h}$. In such a case, the PDF of h needs to be added from all branches of the inverse function.

Theorem 3.1 (PDF of the function of a random variable) *If f_g is the PDF of the random variable g and p a differentiable function $h = p(g)$, then the PDF of the random variable h is given by*

$$\boxed{f_h(h) = \sum_{s=1}^{S} \frac{f_g(g_s)}{|p'(g_s)|}}, \tag{3.8}$$

where g_s are the S real roots of $h = p(g)$.

A monotonic function p has a unique inverse function $p^{-1}(h)$. Then Eq. (3.8) reduces to

$$f_h(h) = \frac{f_g(p^{-1}(h))}{|p'(p^{-1}(h))|}. \tag{3.9}$$

In image processing, the following problem is encountered with respect to probability distributions. We have a signal g with a certain PDF and want to transform g by a suitable transform into h in such a way

that h has a specific probability distribution. This is the inverse problem to what we have discussed so far and it has a surprisingly simple solution. The transform

$$h = F_h^{-1}(F_g(g)) \tag{3.10}$$

converts the $f_g(g)$-distributed random variable g into the $f_h(h)$-distributed random variable h. The solution is especially simple for a transformation to a *uniform distribution* because then F^{-1} is a constant function and $h = F_g(g))$.

Now we consider the mean and variance of functions of random variables. By definition according to Eq. (3.3), the mean of h is

$$Eh = \mu_h = \int_{-\infty}^{\infty} h f_h(h) \mathrm{d}h. \tag{3.11}$$

We can, however, also express the mean directly in terms of the function $p(g)$ and the PDF $f_g(g)$:

$$Eh = E\left(p(g)\right) = \int_{-\infty}^{\infty} p(g) f_g(g) \mathrm{d}g. \tag{3.12}$$

Intuitively, you may assume that the mean of h can be computed from the mean of g: $Eh = p(Eg)$. This is, however, only possible if p is a linear function. If $p(g)$ is approximated by a polynomial

$$p(g) = p(\mu_g) + p'(\mu_g)(g - \mu_g) + p''(\mu_g)(g - \mu_g)^2/2 + \ldots \tag{3.13}$$

then

$$\mu_h \approx p(\mu_g) + p''(\mu_g)\sigma_g^2/2. \tag{3.14}$$

From this equation we see that $\mu_h \approx p(\mu_g)$ is only a good approximation if both the curvature of $p(g)$ and the variance of g are small, i. e., p(g) can be well approximated by a linear function in an interval $[\mu - 3\sigma, \mu + 3\sigma]$.

The first-order estimate of the variance of h is given by

$$\boxed{\sigma_h^2 \approx \left|p'(\mu_g)\right|^2 \sigma_g^2.} \tag{3.15}$$

This expression is only exact for linear functions p.

The following simple relations for means and variances follow directly from the discussion above (a is a constant):

$$E(ag) = aEg, \quad \mathrm{var}(ag) = a^2 \mathrm{var}\, g, \quad \mathrm{var}\, g = E(g^2) - (Eg)^2. \tag{3.16}$$

3.3 Multiple Random Variables

In image processing, we have many random variables and not just one. Image processing operations compute new values from values at many pixels. Thus, it is important to study the statistics of multiple RVs. In this section, we make the first step and discuss how the statistical properties of multiple RVs and functions of multiple RVs can be handled.

3.3.1 Joint Probability Density Functions

First, we need to consider how the random properties of multiple RVs can be described. Generally, the random properties of two RVs, g_1 and g_2, cannot be described by their individual PDFs, $f(g_1)$ and $f(g_2)$. It is rather necessary to define a *joint probability density function* $f(g_1, g_2)$. Only if the two random variables are *independent*, i. e., if the probability that g_1 takes a certain value does not depend on the value of g_2, we can compute the joint PDF from the individual PDFs, known as *marginal PDFs*:

$$f(g_1, g_2) = f_{g_1}(g_1) f_{g_2}(g_2) \quad \Leftrightarrow \quad g_1, g_2 \text{ independent.} \tag{3.17}$$

For P random variables g_p, the random vector $\boldsymbol{g}$, the joint probability density function is $f(g_1, g_2, \ldots, g_P) = f(\boldsymbol{g})$. The P RVs are called independent if the joint PDF can be written as a product of the marginal PDFs

$$f(\boldsymbol{g}) = \prod_{p=1}^{P} f_{g_p}(g_p) \quad \Leftrightarrow \quad g_p \text{ independent, } p = 1, \ldots, P. \tag{3.18}$$

3.3.2 Covariance and Correlation

The covariance measures to which extent the fluctuations of two RVs, g_p and g_q, are related to each other. In extension of the definition of the *variance* in Eq. (3.5), the *covariance* is defined as

$$\sigma_{pq} = E\left((g_p - \mu_p)(g_q - \mu_q)\right) = E(g_p g_q) - E(g_p)E(g_q). \tag{3.19}$$

For P random variables, the covariances form a $P \times P$ symmetric matrix, the *covariance matrix* $\boldsymbol{\Sigma} = \operatorname{cov} \boldsymbol{g}$. The diagonal of this matrix contains the variances of the P RVs.

The *correlation coefficient* relates the covariance to the corresponding variances:

$$c_{pq} = \frac{\sigma_{pq}}{\sigma_p \sigma_q} \quad \text{where} \quad |c| \leq 1. \tag{3.20}$$

Two RVs g_p and g_q are called *uncorrelated* if the covariance σ_{pq} is zero. Then according to Eqs. (3.19) and (3.20) the following relations are true for uncorrelated RVs:

$$\sigma_{pq} = 0 \Leftrightarrow c_{pq} = 0 \Leftrightarrow E(g_p g_q) = E(g_p)E(g_q) \Leftrightarrow g_p, g_q \text{ uncorrelated.} \tag{3.21}$$

From the last of these conditions and Eq. (3.17), it is evident that independent RVs are uncorrelated.

At first glance it appears that only the statistical properties of independent RVs are easy to handle. Then we only need to consider the marginal PDFs of the individual variables together with their mean and variance. Generally, the interrelation of random variations of the variables as expressed by the covariance matrix $\mathbf{\Sigma}$ must be considered. Because the covariance matrix is symmetric, however, we can always find a coordinate system, i. e., a linear combination of the RVs, in which the covariance matrix is diagonal and thus the RVs are uncorrelated.

3.3.3 Linear Functions of Multiple Random Variables

In extension to the discussion of functions of a single RV in Section 3.2.3, we can express the mean of a function of multiple random variables $h = p(g_1, g_2, \ldots, g_P)$ directly from the joint PDF:

$$Eh = \int_{-\infty}^{\infty} p(g_1, g_2, \ldots, g_P) f(g_1, g_2, \ldots, g_P) \mathrm{d}g_1 \mathrm{d}g_2 \ldots \mathrm{d}g_P. \tag{3.22}$$

From this general relation it follows that the mean of any linear function

$$h = \sum_{p=1}^{P} a_p g_p \tag{3.23}$$

is given as the linear combination of the means of the RVs g_p:

$$E\left(\sum_{p=1}^{P} a_p g_p\right) = \sum_{p=1}^{P} a_p E\left(g_p\right). \tag{3.24}$$

Note that this is a very general result. We did not assume that the RVs are independent, and this is not dependent on the type of the PDF. As a special case Eq. (3.24) includes the simple relations

$$E(g_1 + g_2) = Eg_1 + Eg_2, \quad E(g_1 + a) = Eg_1 + a. \tag{3.25}$$

The variance of functions of multiple RVs cannot be computed that easy even in the linear case. Let $\boldsymbol{g}$ be a vector of P RVs, $\boldsymbol{h}$ a vector of

Q RVs that is a linear combination of the P RVs $\boldsymbol{g}$, $\boldsymbol{M}$ a $Q \times P$ matrix of coefficients, and $\boldsymbol{a}$ a column vector with Q coefficients. Then

$$\boldsymbol{h} = \boldsymbol{M}\boldsymbol{g} + \boldsymbol{a} \quad \text{with} \quad E(\boldsymbol{h}) = \boldsymbol{M}E(\boldsymbol{g}) + \boldsymbol{a} \tag{3.26}$$

in extension to Eq. (3.24). If $P = Q$, Eq. (3.26) can be interpreted as a coordinate transformation in a P-dimensional vector space. Therefore it is not surprising that the symmetric covariance matrix transforms as a second-order tensor [149]:

$$\boxed{\operatorname{cov}(\boldsymbol{h}) = \boldsymbol{M}\operatorname{cov}(\boldsymbol{g})\boldsymbol{M}^T.} \tag{3.27}$$

To illustrate the application of Eq. (3.27), we discuss three examples.

Variance of the mean of RVs. First, we discuss the computation of the variance of the mean $\overline{g}$ of P RVs with the same mean and variance σ^2. We assume that the RVs are uncorrelated. Then the matrix $\boldsymbol{M}$ and the covariance matrix $\operatorname{cov}\boldsymbol{g}$ are

$$\boldsymbol{M} = \frac{1}{P}[1,1,1,\ldots,1] \quad \text{and} \quad \operatorname{cov}(\boldsymbol{g}) = \begin{bmatrix} \sigma^2 & 0 & \ldots & 0 \\ 0 & \sigma^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2 \end{bmatrix} = \sigma^2 \boldsymbol{I}.$$

Using these expressions in Eq. (3.27) yields

$$\sigma_{\overline{g}}^2 = \frac{1}{P}\sigma^2. \tag{3.28}$$

Thus the variance $\sigma_{\overline{g}}^2$ is proportional to P^{-1} and the *standard deviation* $\sigma_{\overline{g}}$ decreases only with $P^{-1/2}$. This means that we must take four times as many measurements in order to double the precision of the measurement of the mean. This is not the case for correlated RVs. If the RVs are fully correlated ($c_{pq} = 1$, $\sigma_{pq} = \sigma^2$), according to Eq. (3.27), the variance of the mean is equal to the variance of the individual RVs. In this case it is not possible to reduce the variance by averaging.

Variance of the sum of uncorrelated ZRVs with unequal variances. In a slight variation, we take P uncorrelated RVs with unequal variances σ_p^2 and compute the variance of the sum of the RVs. From Eq. (3.25), we know already that the mean of the sum is equal to the sum of the means (even for correlated RVs). Similar as for the previous example, it can be shown that for uncorrelated RVs the variance of the sum is also the sum of the individual variances:

$$\operatorname{var} \sum_{p=1}^{P} g_p = \sum_{p=1}^{P} \operatorname{var} g_p. \tag{3.29}$$

Linear Combinations of multiple RVs. As a second example we take Q RVs h_q that are a linear combination of the P uncorrelated RVs g_p with equal variance σ^2:

$$h_q = \boldsymbol{a}_q^T \boldsymbol{g}. \tag{3.30}$$

Then the vectors $\boldsymbol{a}_q^T$ form the rows of the $Q \times P$ matrix $\boldsymbol{M}$ in Eq. (3.26) and the covariance matrix of $\boldsymbol{h}$ results according to Eq. (3.27) in

$$\operatorname{cov}(\boldsymbol{h}) = \sigma^2 \boldsymbol{M}\boldsymbol{M}^T = \sigma^2 \begin{bmatrix} \boldsymbol{a}_1^T\boldsymbol{a}_1 & \boldsymbol{a}_1^T\boldsymbol{a}_2 & \dots & \boldsymbol{a}_1^T\boldsymbol{a}_Q \\ \boldsymbol{a}_1^T\boldsymbol{a}_2 & \boldsymbol{a}_2^T\boldsymbol{a}_2 & \dots & \boldsymbol{a}_2^T\boldsymbol{a}_Q \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_1^T\boldsymbol{a}_Q & \boldsymbol{a}_2^T\boldsymbol{a}_Q & \dots & \boldsymbol{a}_Q^T\boldsymbol{a}_Q \end{bmatrix}. \tag{3.31}$$

From this equation, we can learn two things. First, the variance of the RV h_q is given by $\boldsymbol{a}_q^T\boldsymbol{a}_q$, i. e., the sum of the squares of the coefficients

$$\sigma^2(h_q) = \sigma^2 \boldsymbol{a}_q^T \boldsymbol{a}_q. \tag{3.32}$$

Second, although the RVs g_p are uncorrelated, two RVs h_p and h_q are only uncorrelated if the scalar product of the coefficient vectors, $\boldsymbol{a}_p^T\boldsymbol{a}_q$, is zero, i. e., the coefficient vectors are orthogonal. Thus, only orthogonal transform matrixes $\boldsymbol{M}$ in Eq. (3.26) leave uncorrelated RVs uncorrelated.

For correlated RVs, we can conclude that it is always possible to apply a suitable transform $\boldsymbol{M}$ to get a set of linear combinations of RVs that are uncorrelated. This follows from the elementary theorem in linear algebra: Each symmetric square matrix can be diagonalized by a transform, which is called the *principal component transform* [15, 215]. The uncorrelated set of linear combinations constitutes the axis of the principal component system and is known as the set of *eigenvectors* of the matrix. The eigenvectors meet the condition

$$\operatorname{cov}(\boldsymbol{h})\boldsymbol{e}_p = \sigma_p^2 \boldsymbol{e}_p. \tag{3.33}$$

This means that the multiplication of the covariance matrix with the eigenvector reduces to a multiplication by a scalar. This factor is called the *eigenvalue* to the eigenvector $\boldsymbol{e}_p$. For the covariance matrix the pth eigenvalue is the variance σ_p^2 in the direction of the eigenvector $\boldsymbol{e}_p$.

3.3.4 Nonlinear Functions of Multiple Random Variables

The above analysis of the variance for functions of multiple RVs can be extended to nonlinear functions provided that the function is sufficiently linear around the mean value. As in Section 3.2.3, we expand the nonlinear function $p_q(\boldsymbol{g})$ into a Taylor series around the mean value:

$$h_p = p_q(\boldsymbol{g}) \approx p_q(\boldsymbol{\mu}) + \sum_{p=1}^{P} \frac{\partial p_q}{\partial g_p}(g_p - \mu_p). \tag{3.34}$$

We compare this equation with Eq. (3.26) and find that the $Q \times P$ matrix $\boldsymbol{M}$ has to be replaced by the matrix

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial p_1}{\partial g_1} & \frac{\partial p_1}{\partial g_2} & \cdots & \frac{\partial p_1}{\partial g_P} \\ \frac{\partial p_2}{\partial g_1} & \frac{\partial p_2}{\partial g_2} & \cdots & \frac{\partial p_2}{\partial g_P} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial p_Q}{\partial g_1} & \frac{\partial p_Q}{\partial g_2} & \cdots & \frac{\partial p_Q}{\partial g_P} \end{bmatrix}, \tag{3.35}$$

known as the *Jacobian matrix* of the transform $\boldsymbol{h} = \boldsymbol{p}(\boldsymbol{g})$. Thus the covariance of $\boldsymbol{h}$ is given by

$$\boxed{\operatorname{cov}(\boldsymbol{h}) \approx \boldsymbol{J} \operatorname{cov}(\boldsymbol{g}) \boldsymbol{J}^T.} \tag{3.36}$$

3.4 Probability Density Functions

In the previous sections, we derived a number of general properties of random variables without any knowledge about the probability distributions. In this section, we discuss a number of specific probability density functions that are of importance for image processing.

As an introduction to handling of PDFs, we discuss the PDFs of function of multiple RVs. We restrict the discussion to two simple cases. First, we consider the addition of two RVs. If two RVs g_1 and g_2 are independent, the resulting probability density function of an additive superposition $g = g_1 + g_2$ is given by the *convolution* integral

$$p_g(g) = \int_{-\infty}^{\infty} p_{g_1}(h) p_{g_2}(g - h) \mathrm{d}h. \tag{3.37}$$

This general property results from the multiplicative nature of the superposition of probabilities. The probability $p_g(g)$ to measure the value g is the product of the probabilities to measure $g_1 = h$ and $g_2 = g - h$. The integral in Eq. (3.37) itself is required because we have to consider all combinations of values that lead to a sum g.

Second, the same procedure can be applied to the multiplication of two RVs if the multiplication of two variables is transformed into an addition by applying the logarithm: $\ln g = \ln g_1 + \ln g_2$. The PDFs of the logarithm of an RV can be computed using Eq. (3.9).

3.4.1 Poisson Distribution

First, we consider image acquisition. An imaging sensor element that is illuminated with a certain irradiance receives within a time interval

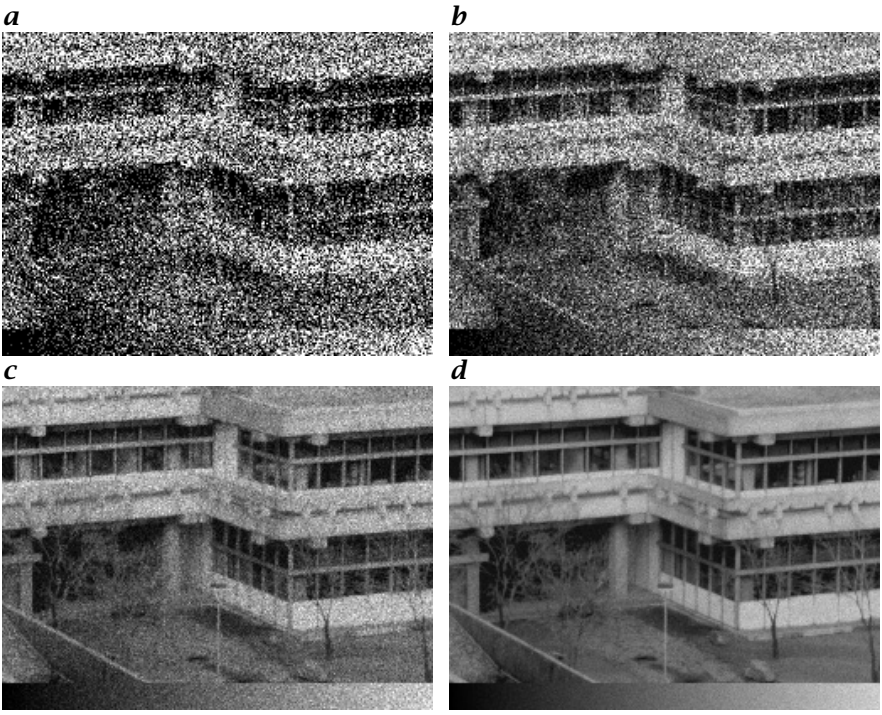

Figure 3.2: *Simulation of low-light images with Poisson noise that have collected maximal* ***a*** *3,* ***b*** *10,* ***c*** *100, and* ***d*** *1000 electrons. Note the linear intensity wedge at the bottom of images* ***c*** *and* ***d*** *.*

Δt, the *exposure time*, on average N electrons by absorption of photons. Thus the mean rate of photons per unit time λ is given by

$$\lambda = \frac{N}{\Delta t}. \tag{3.38}$$

Because of the random nature of the stream of photons a different number of photons arrive during each exposure. A random process in which we count on average $\lambda\Delta t$ events is known as a *Poisson process* $P(\lambda\Delta t)$. It has the discrete probability density distribution

$$\boxed{P(\lambda\Delta t): \quad f_n = \exp(-\lambda\Delta t)\frac{(\lambda\Delta t)^n}{n!}, \quad n \geq 0} \tag{3.39}$$

with the mean and variance

$$\boxed{\mu = \lambda\Delta t \quad \text{and} \quad \sigma^2 = \lambda\Delta t.} \tag{3.40}$$

Simulated low-light images with Poisson noise are shown in Fig. 3.2. For low mean values, the Poisson PDF is skewed with a longer tail towards

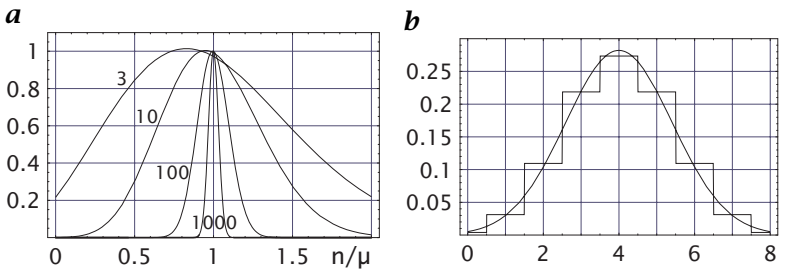

Figure 3.3: ***a*** *Poisson PDFs $P(\mu)$ for mean values μ of 3, 10, 100, and 1000. The x axis is normalized by the mean: the mean value is one; $P(\lambda\Delta t)$ is multiplied by $\sigma\sqrt{2\pi}$;* ***b*** *Discrete binomial PDF $B(8, 1/2)$ with a mean of 4 and variance of 2 and the corresponding normal PDF $N(4, 2)$.*

higher values (Fig. 3.3a). But even for a moderate mean (100), the density function is already surprisingly symmetric.

A typical CCD image sensor element (Section 1.7.1, ≻ R2) collects in the order of 10000 or more electrons that are generated by absorbed photons. Thus the standard deviation of the number of collected electrons is 100 or 1%. From this figure, we can conclude that even a perfect image sensor element that introduces no additional electronic noise will show a considerable noise level just by the underlying Poisson process.

The Poisson process has the following important properties:

1. The standard deviation σ is not constant but is equal to the square root of the number of events. Therefore the noise level is signal-dependent.
2. It can be shown that nonoverlapping exposures are statistically independent events [149, Section. 3.4]. This means that we can take images captured with the same sensor at different times as independent RVs.
3. The Poisson process is additive: the sum of two independent Poisson-distributed RVs with the means μ_1 and μ_2 is also Poisson distributed with the mean and variance $\mu_1 + \mu_2$.

3.4.2 Normal and Binomial Distributions

Many processes with continuous RVs can adequately be described by the *normal* or *Gaussian probability density* $N(\mu, \sigma)$ with the mean μ and the variance σ^2:

$$N(\mu, \sigma): \quad f(g) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(g-\mu)^2}{2\sigma^2}\right). \tag{3.41}$$

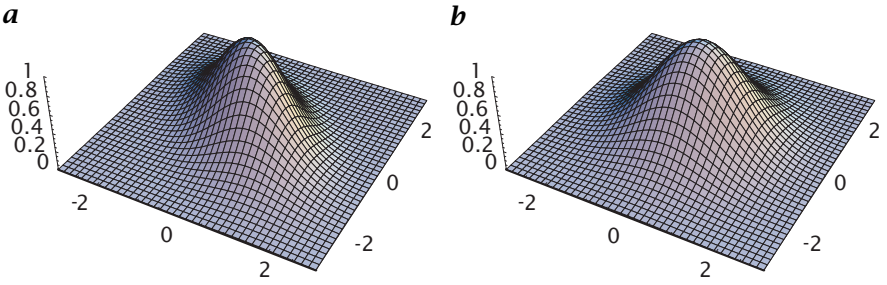

Figure 3.4: *Bivariate normal densities:* ***a*** *correlated RVs with* $\sigma_1^2 = \sigma_2^2 = 1$*, and* $r_{12} = -0.5$*;* ***b*** *isotropic uncorrelated RVs with variances* $\sigma_1^2 = \sigma_1^2 = 1$*.*

From Eq. (3.41) we can see that the normal distribution is completely described by the mean and the variance.

For discrete analogue to the normal distribution is the *binomial distribution* $B(Q, p)$

$$\boxed{B(Q,p): \quad f_q = \frac{Q!}{q!\,(Q-q)!} p^q (1-p)^{Q-q}, \quad 0 \le q < Q.} \tag{3.42}$$

The natural number Q denotes the number of possible outcomes and the parameter $p \in]0,1[$ determines together with Q the mean and the variance:

$$\boxed{\mu = Qp \quad \text{and} \quad \sigma^2 = Qp(1-p).} \tag{3.43}$$

Even for moderate Q, the binomial distribution comes very close to the Gaussian distribution as illustrated in Fig. 3.3b.

In extension to Eq. (3.41), the joint normal PDF $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for multiple RVs, i.e., the random vector $\boldsymbol{g}$ with the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ is given by

$$N(\boldsymbol{\mu}, \boldsymbol{\Sigma}): \; f(\boldsymbol{g}) = \frac{1}{(2\pi)^{P/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left(-\frac{(\boldsymbol{g}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{g}-\boldsymbol{\mu})}{2}\right). \tag{3.44}$$

At first glance this expression looks horribly complex. It is not. We must just consider that the symmetric covariance matrix becomes a diagonal matrix by rotation into its principle-axis system. Then the joint normal density function becomes a separable function

$$f(\boldsymbol{g}') = \prod_{p=1}^{P} \frac{1}{(2\pi\sigma_p^2)^{1/2}} \exp\left(-\frac{(g_p' - \mu_p)^2}{2\sigma_p^2}\right) \tag{3.45}$$

with the variances σ_p^2 along the principle axes (Fig. 3.4a) and the components g_p' are independent RVs.

For uncorrelated RVs with equal variance σ^2, the $N(\boldsymbol{\mu}, \boldsymbol{C})$ distribution reduces to the isotropic normal PDF $N(\boldsymbol{\mu}, \sigma)$ (Fig. 3.4b):

$$N(\boldsymbol{\mu}, \sigma): \quad f(\boldsymbol{g}) = \frac{1}{(2\pi\sigma^2)^{P/2}} \exp\left(-\frac{|(\boldsymbol{g}-\boldsymbol{\mu})|^2}{2\sigma^2}\right). \tag{3.46}$$

3.4.3 Central Limit Theorem

The central importance of the normal distribution stems from the *central limit theorem* (Theorem 2.6, p. 56), which we discussed with respect to cascaded convolution in Section 2.3.4. Here we emphasize its significance for RVs in image processing. The central limit theorem states that under conditions that are almost ever met for image processing applications the PDF of a sum of RVs tends to a normal distribution. As we discussed in Section 3.3, in image processing weighted sums from many values are often computed. Consequently, these combined variables have a normal PDF.

3.4.4 Other Distributions

Despite the significance of the *normal distribution*, other probability density functions also play a certain role for image processing. They occur when RVs are combined by nonlinear functions.

As a first example, we discuss the conversion from *Cartesian* to *polar coordinates*. We take the random vector $\boldsymbol{g} = [g_1, g_2]^T$ with independent $N(0, \sigma)$-distributed components. Then it can be shown [149, Section 6.3] that the magnitude of this vector $r = (g_1^2, g_2^2)^{1/2}$ and the polar angle $\phi = \arctan(g_2/g_1)$ are independent random variables. The magnitude has a *Rayleigh density*

$$R(\sigma): \quad f(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad \text{for} \quad r > 0 \tag{3.47}$$

with the mean and variance

$$\mu_R = \sigma\sqrt{\pi/2} \quad \text{and} \quad \sigma_R^2 = \sigma^2 \frac{4-\pi}{2}, \tag{3.48}$$

and the angle ϕ has a *uniform density*

$$f(\phi) = \frac{1}{2\pi}. \tag{3.49}$$

In generalization of the Rayleigh density, we consider the magnitude of a P dimensional vector. It has a *chi density* with P degrees of freedom

$$\chi(P, \sigma): \quad f(r) = \frac{2r^{P-1}}{2^{P/2}\Gamma(P/2)\sigma^P} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad \text{for} \quad r > 0 \tag{3.50}$$

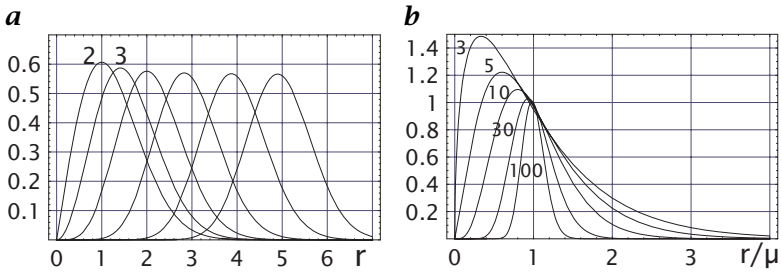

Figure 3.5: ***a*** *Chi density for 2 (Rayleigh density), 3 (Maxwell density), and higher degrees of freedom as indicated;* ***b*** *chi-square density in a normalized plot (mean at one) with degrees of freedom as indicated.*

with the mean

$$\mu_\chi = \sigma \frac{\sqrt{2}\,\Gamma(P/2 + 1/2)}{\Gamma(P/2)} \approx \sigma\sqrt{P - 1/2} \quad \text{for} \quad P \gg 1 \tag{3.51}$$

and variance

$$\sigma_\chi^2 = \sigma^2 P - \mu_\chi^2 \approx \sigma^2/2 \quad \text{for} \quad P \gg 1. \tag{3.52}$$

The mean of the chi density increases with the square root of P while the variance is almost constant. For large degrees of freedom, the density quickly approaches the normal density $N(\sigma\sqrt{P/2 - 1/2}, \sigma/\sqrt{2})$ (Fig. 3.5a).

The PDF of the square of the magnitude of the vector has a different PDF because squaring is a nonlinear function (Section 3.2.3). Using Theorem 3.1 the PDF, known as the *chi-square density* with P degrees of freedom, can be computed as

$$\chi^2(P, \sigma): \quad f(r) = \frac{r^{P/2-1}}{2^{P/2}\Gamma(P/2)\sigma^P} \exp\left(-\frac{r}{2\sigma^2}\right) \quad \text{for} \quad r > 0 \tag{3.53}$$

with the mean and variance

$$\mu_{\chi^2} = \sigma^2 P \quad \text{and} \quad \sigma_{\chi^2}^2 = 2\sigma^4 P. \tag{3.54}$$

The sum of squares of RVs is of special importance to obtain the error in the estimation of the *sample variance*

$$s^2 = \frac{1}{P-1}\sum_1^P (g_p - \overline{g})^2 \quad \text{with} \quad \overline{g} = \frac{1}{P}\sum_1^P g_p. \tag{3.55}$$

Papoulis [149, Section 8.2] shows that the normalized sample variance

$$\frac{(P-1)s^2}{\sigma^2} = \sum_1^P \left(\frac{g_p - \overline{g}}{\sigma}\right)^2 \tag{3.56}$$

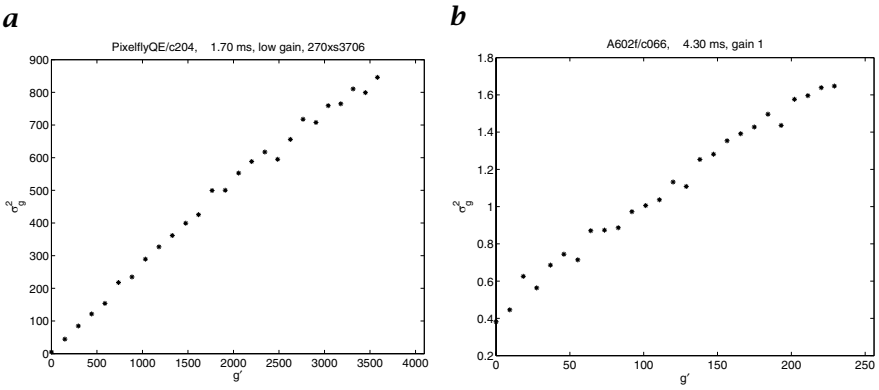

Figure 3.6: *Noise variance as a function of the digital gray value for* ***a*** *Pixelfly QE from PCO with Sony interline CCD ICX285AL, 12 Bit,* $\sigma_0 = 2.2$ *(8* e^-*) and* ***b*** *Basler A602f with Micron MT9V403 CMOS, 8 Bit,* $\sigma_0 = 0.61$ *(91* e^-*) [90].*

has a *chi-square density* with $P - 1$ degrees of freedom. Thus the mean of the sample variance is σ^2 (unbiased estimate) and the variance is $2\sigma^4/(P-1)$. For low degrees of freedom, the chi-square density shows significant deviations from the normal density (Fig. 3.5b). For more than 30 degrees of freedom the density is in good approximation normally distributed. A reliable estimate of the variance requires many measurements. For $P = 100$, the relative standard deviation of the variance is still about 20% (for the standard deviation of the standard deviation it is half, 10%).

3.4.5 Noise Model for Image Sensors

After the detailed discussion on random variables, we can now conclude with a simple noise model for an image sensor. In Section 3.4.1 we saw that the photo signal for a single pixel is Poisson distributed. Except for very low-level imaging conditions, where only a few electrons are collected per sensor element, the Poisson distribution is well approximated by a normal distribution $N(Q_e, \sqrt{Q_e})$, where Q_e is the number of electrons absorbed during an exposure. Not every incoming photon causes the excitation of an electron. The fraction of electrons excited by the photons irradiating onto the sensor element (Q_p) is known as *quantum efficiency* η:

$$\eta = \frac{Q_e}{Q_p}. \tag{3.57}$$

The electronic circuits add a number of other noise sources. For practical purposes, it is only important to know that these noise sources are normal distributed and independent of the photon noise. Therefore

the total number of generated charge units and their variances are

$$Q = Q_0 + Q_e \quad \text{und} \quad \sigma_Q^2 = \sigma_{Q_0}^2 + \sigma_{Q_e}^2. \tag{3.58}$$

We assume that the electronic circuits are linear. Therefore the resulting digital signal g is given by

$$g = KQ. \tag{3.59}$$

The conversion factor K is dimensionless und expresses the entire amplification of the signal in bits/charge unit. The variance of the digital signal is easy to compute by using the rules of error propagation (Eqs. (3.15) and (3.29)), the fact that $\sigma_{Q_e}^2 = Q_e$ (3.40), and (3.59):

$$\boxed{\sigma_g^2 = K^2\sigma_{Q_0}^2 + K^2\sigma_{Q_e}^2 = \sigma_0^2 + Kg.} \tag{3.60}$$

Equation (3.60) predicts a linear increase of the variance with the digital signal g. Measurements generally show a good agreement with this simple model (Fig. 3.6). Interestingly, noise has a benefit here. The conversion factor K can be determined from the $\sigma_g^2(g)$ relation without knowing any detail about the electronic circuits.

3.5 *Stochastic Processes and Random Fields*

The statistics developed so far do not consider the spatial and temporal relations between the points of a multidimensional signal. If we want to analyze the content of images statistically, we must consider the whole image as a statistical quantity, known as a *random field* for spatial data and as a *stochastic process* for time series.

In case of an $M \times N$ image, a random field consists of an $M \times N$ matrix whose elements are random variables. This means that a joint probability density function has MN variables. The mean of a random field is then given as a sum over all possible states q:

$$\overline{G_{m,n}} = \sum_{q=1}^{Q^{MN}} f_q(\boldsymbol{G})\boldsymbol{G}_q. \tag{3.61}$$

If we have Q quantization levels, each pixel can take Q different states. In combination of all $M \times N$ pixels we end up with Q^{MN} states $\boldsymbol{G}_q$. This is a horrifying concept, rendering itself useless because of the combinatory explosion of possible states. Thus we have to find simpler concepts to treat multidimensional signals as random fields. In this section, we will approach this problem in a practical way.

We start by estimating the mean and variance of a random field. We can do that in the same way as for a single value (Eq. (3.55)), by taking the mean $\boldsymbol{G}_p$ of P measurements under the same conditions and computing the average as

$$\overline{\boldsymbol{G}} = \frac{1}{P}\sum_{p=1}^{P} \boldsymbol{G}_p. \tag{3.62}$$

This type of averaging is known as an *ensemble average*. The estimate of the *variance*, the *sample variance*, is given by

$$S_{\overline{G}}^2 = \frac{1}{P-1}\sum_{p=1}^{P}\left(\boldsymbol{G}_p - \overline{\boldsymbol{G}}\right)^2. \tag{3.63}$$

At this stage, we know already the mean and variance at each pixel in the image. From these values we can make a number of interesting conclusions. We can study the uniformity of both quantities under given conditions such as a constant illumination level.

3.5.1 *Correlation and Covariance Functions*

In a second step, we now relate the gray values at different positions in the images with each other. One measure for the correlation of the gray values is the mean for the product of the gray values at two positions, the *autocorrelation function*

$$R_{gg}(m,n;m',n') = \overline{G_{mn}G_{m'n'}}. \tag{3.64}$$

As in Eqs. (3.62) and (3.63), an ensemble mean is taken.

The autocorrelation function is not of much use if an image contains a deterministic part with additive *zero-mean noise*

$$\boldsymbol{G}' = \boldsymbol{G} + \boldsymbol{N}, \quad \text{with} \quad \overline{\boldsymbol{G}'} = \boldsymbol{G} \quad \text{and} \quad \overline{\boldsymbol{N}'} = \boldsymbol{0}. \tag{3.65}$$

Then it is more useful to subtract the mean so that the properties of the random part in the signal are adequately characterized:

$$C_{gg}(m,n;m',n') = \overline{(G_{mn} - \overline{G_{mn}})(G_{m'n'} - \overline{G_{m'n'}})}. \tag{3.66}$$

This function is called the *autocovariance function*. For zero shift ($m = m'$ and $n = n'$) it gives the variance at the pixel $[m,n]^T$, at all other shifts the *covariance*, which was introduced in Section 3.3.2, Eq. (3.19). New here is that the autocovariance function includes the spatial relations between the different points in the image. If the autocovariance is zero, the random properties of the corresponding points are uncorrelated.

The autocovariance function as defined in Eq. (3.66) is still awkward because it is four-dimensional. Therefore even this statistic is only of use for a restricted number of shifts, e. g., short distances, because we suspect that the random properties of distant points are uncorrelated.

Things become easier if the statistics do not explicitly depend on the position of the points. This is called a *homogeneous random field*. Then the autocovariance function becomes *shift invariant*:

$$\begin{aligned} C_{gg}(m+k,n+l;m'+k,n'+l) &= C_{gg}(m,n;m',n') \\ &= C_{gg}(m-m',n-n';0,0) \\ &= C_{gg}(0,0;m'-m,n'-n). \end{aligned} \tag{3.67}$$

The last two identities are obtained when we set $(k,l) = (-m',-n')$ and $(k,l) = (-m,-n)$. This also means that the variance of the noise $C_{gg}(m,n;m,n)$ no longer depends on the position in the image but is equal at all points.

Because the autocorrelation function depends only on the distance between points, it reduces from a four- to a two-dimensional function. Fortunately, many stochastic processes are homogeneous. Because of the shift invariance, the autocovariance function for a homogeneous random field can be estimated by spatial averaging:

$$C_{gg}(m,n) = \frac{1}{MN}\sum_{m'=0}^{M-1}\sum_{n'=0}^{N-1}(G_{m'n'} - \overline{G_{m'n'}})(G_{m'+m,n'+n} - \overline{G_{m'+m,n'+n}}). \quad (3.68)$$

Generally, it is not certain that spatial averaging leads to the same mean as the ensemble mean. A random field that meets this criterion is called *ergodic*.

Another difficulty concerns indexing. As soon as $(m,n) \neq (0,0)$, the indices run over the range of the matrix. We then have to consider the periodic extension of the matrix, as discussed in Section 2.3.4. This is known as *cyclic correlation*.

Now we illustrate the meaning of the autocovariance function. We consider an image that contains a deterministic part plus zero-mean homogeneous noise, see Eq. (3.65). Let us further assume that all points are statistically independent. Then the mean is the deterministic part and the autocovariance vanishes except for zero shift, i. e., for a zero pixel distance:

$$\boldsymbol{C}_{gg} = \sigma^2 {}^{oo}\boldsymbol{P} \quad \text{or} \quad C_{gg}(m,n) = \sigma^2 \delta_m \delta_n. \quad (3.69)$$

For zero shift, the autocovariance is equal to the variance of the noise. In this way, we can examine whether the individual image points are statistically uncorrelated. This is of importance because the degree of correlation between the image points determines the statistical properties of image processing operations as discussed in Section 3.3.3.

In a similar manner to correlating one image with itself, we can correlate two different images $\boldsymbol{G}$ and $\boldsymbol{H}$ with each other. These could be either images from different scenes or images of a dynamic scene taken at different times. By analogy to Eq. (3.68), the *cross-correlation function* and *cross-covariance function* are defined as

$$R_{gh}(m,n) = \frac{1}{MN}\sum_{m'=0}^{M-1}\sum_{n'=0}^{N-1} G_{m'n'} H_{m'+m,n'+n} \quad (3.70)$$

$$C_{gh}(m,n) = \frac{1}{MN}\sum_{m'=0}^{M-1}\sum_{n'=0}^{N-1}(G_{m'n'} - \overline{G_{m'n'}})(H_{m+m',n+n'} - \overline{H_{m+m',n+n'}}). \quad (3.71)$$

The cross-correlation operation is very similar to *convolution* (Section 2.3.4, ≻R7). The only difference is the sign of the indices (m',n') in the second term.

3.5.2 *Random Fields in Fourier Space*

In the previous sections we studied random fields in the spatial domain. Given the significance of the *Fourier transform* for image processing (Section 2.3), we now turn to random fields in the Fourier domain. For the sake of simplicity, we restrict the discussion here to the 1-D case. All arguments put forward in this section can, however, be applied analogously in any dimension.

The Fourier transform requires complex numbers. This constitutes no additional complications, because the random properties of the real and imaginary part can be treated separately. The definitions for the mean remains the same, the definition of the covariance, however, requires a slight change as compared to Eq. (3.19):

$$C_{pq} = E\left((g_p - \mu_p)^*(g_q - \mu_q)\right), \tag{3.72}$$

where $*$ denotes the conjugate complex. This definition ensures that the variance

$$\sigma_p^2 = E\left((g_p - \mu_p)^*(g_p - \mu_p)\right) \tag{3.73}$$

remains a real number.

The 1-D DFT maps a vector $\boldsymbol{g} \in \mathbb{C}^N$ onto a vector $\hat{\boldsymbol{g}} \in \mathbb{C}^N$. The components of $\hat{\boldsymbol{g}}$ are given as scalar products with orthonormal base vectors for the vector space $\mathbb{C}^N$ (compare Eqs. (2.29) and (2.30)):

$$\hat{g}_v = \boldsymbol{b}_v{}^T \boldsymbol{g} \quad \text{with} \quad \boldsymbol{b}_v{}^T \boldsymbol{b}_{v'} = \delta_{v-v'}. \tag{3.74}$$

Thus the complex RVs in Fourier space are nothing else but linear combinations of the RVs in the spatial domain. If we assume that the RVs in the spatial domain are uncorrelated with equal variance (homogeneous random field), we arrive at a far-reaching conclusion. According to Eq. (3.74) the coefficient vectors $\boldsymbol{b}_v$ are orthogonal to each other with a unit square magnitude. Therefore we can conclude from the discussion about functions of multiple RVs in Section 3.3.3, especially Eq. (3.32), that the RVs in the Fourier domain remain uncorrelated and have the same variance as in the spatial domain.

3.5.3 *Power Spectrum, Cross-correlation Spectrum, and Coherence*

In Section 3.5.1 we learnt that random fields in the space domain are characterized by the auto- and the cross-correlation functions. Now we consider random fields in the Fourier space.

Correlation in the space domain corresponds to multiplication in the Fourier space with the complex conjugate functions (≻ R4):

$$\boldsymbol{G} \star \boldsymbol{G} \circ\!\!-\!\!\bullet\, P_{gg}(\boldsymbol{k}) = \hat{g}(\boldsymbol{k})^* \hat{g}(\boldsymbol{k}) \tag{3.75}$$

and

$$\boldsymbol{G} \star \boldsymbol{H} \circ\!\!-\!\!\bullet\, P_{gh}(\boldsymbol{k}) = \hat{g}(\boldsymbol{k})^* \hat{h}(\boldsymbol{k}). \tag{3.76}$$

In these equations, correlation is abbreviated with the $\star$ symbol, similar to convolution for which we use the $*$ symbol. For a simpler notation, the spectra are written as continuous functions. This corresponds to the transition to an infinitely extended random field (Section 2.3.2, Table 2.1).

The Fourier transform of the autocorrelation function is the *power spectrum* P_{gg}. The power spectrum is a real-valued quantity. Its name is related to the fact that it represents the distribution of power of a physical signal in the Fourier domain, i. e., over frequencies and wave numbers, if the signal amplitude squared is related to the energy of a signal. If the power spectrum is averaged over several images, it constitutes a sum of squares of independent random variables. If the RVs have a normal density, the power spectrum has, according to the discussion in Section 3.4.4, a chi-square density.

The autocorrelation function of a field of uncorrelated RVS is zero except at the origin, i. e., a δ-function (Eq. (3.69)). Therefore, its power spectrum is a constant ($\succ$ R7). This type of noise is called *white noise*.

The Fourier transform of the cross-correlation function is called the *cross-correlation spectrum* P_{gh}. In contrast to the power spectrum, it is a complex quantity. The real and imaginary parts are termed the *co-* and *quad-spectrum*, respectively.

To understand the meaning of the cross-correlation spectrum, it is useful to define another quantity, the *coherence function* Φ:

$$\Phi^2(\boldsymbol{k}) = \frac{|P_{gh}(\boldsymbol{k})|^2}{P_{gg}(\boldsymbol{k})P_{hh}(\boldsymbol{k})}. \tag{3.77}$$

Basically, the coherence function contains information on the similarity of two images. We illustrate this by assuming that the image $\boldsymbol{H}$ is a shifted copy of the image $\boldsymbol{G}$: $\hat{h}(\boldsymbol{k}) = \hat{g}(\boldsymbol{k}) \exp(-\mathrm{i}\boldsymbol{k}\boldsymbol{x}_s)$. In this case, the coherence function is one and the cross-correlation spectrum P_{gh} reduces to

$$P_{gh}(\boldsymbol{k}) = P_{gg}(\boldsymbol{k}) \exp(-\mathrm{i}\boldsymbol{k}\boldsymbol{x}_s). \tag{3.78}$$

Because P_{gg} is a real quantity, we can compute the shift $\boldsymbol{x}_s$ between the two images from the phase factor $\exp(-\mathrm{i}\boldsymbol{k}\boldsymbol{x}_s)$.

If there is no fixed phase relationship of a periodic component between the two images, then the coherency decreases. If the phase shift is randomly distributed from image to image in a sequence, the cross-correlation vectors in the complex plane point in random directions and add up to zero. According to Eq. (3.77), then also the coherency is zero.

3.6 *Exercises*

3.1: Noise in images and image sequences

Interactive simulation of Poisson-distributed noise, additive normal-distributed noise und multiplicative normal-distributed noise; computation of mean and variance (dip6ex03.01).

3.2: **Poisson distribution and normal distribution

An image sensor receives a spatially and temporally constant irradiation. During the exposure time 9 and 100 charge units are generated in the mean. We further assume that the sensor is ideal, i. e., the electronic circuits produce no additional noise.

1. Compute the absolute standard deviation and the relative standard deviation (σ/μ) for both cases
2. How much does the Poisson distribution deviate from the normal distribution with the same variance?
   Answer this question by computing the probability density functions for the values $\mu - n\sigma$ with $n \in \{-3, -2, -1, 0, 1, 2, 3\}$.

3.3: *__Binomial and normal distribution__

The Binomial distribution $B(Q, 1/2)$ converges for increasing Q quickly to the normal distribution. Check the statement by comparing all values of the binomial distributions $B(4, 1/2)$ and $B(8, 1/2)$ to the normal distribution with equal mean and variance.

3.4: *__Uniform distribution__

A random variable (RV) has a uniform probability density function (PDF) in the interval between g and $g + \Delta g$. The PDF is zero outside of this interval. Compute the mean and variance of this RV.

3.5: **__PDFs, mean and variance__

Let g_1 and g_2 be two uncorrelated RVs with zero mean ($\mu = 0$) and variance $\sigma^2 = 1$. Compute the PDF, mean and variance of the following RVs:

1. $h = g_1 + g_2$
2. $h = ag_1 + b$ (a and b are deterministic constants)
3. $h = g_1 + g_1$
4. $h = g_1^2$
5. $h = \sqrt{g_1^2 + g_2^2}$ (Magnitude of vector $[g_1\ g_2]^T$)
6. $h = \arctan(g_2/g_1)$ (Angle of vector $[g_1\ g_2]^T$)

3.6: *__Error propagation__

Let g be a RV with mean $\overline{g}$ and variance σ_g^2. The PDF is unknown. Compute, if possible, the variance and the relative error $\sigma_h/\overline{h}$ of the following RVs h assuming that the variance is small enough so that the nonlinearity of the following functions is negligible:

1. $h = g^2$
2. $h = \sqrt{g}$
3. $h = 1/g$
4. $h = \ln(g)$

3.7: Central limit theorem

Interactive simulation to illustrate the central limit theorem (dip6ex03.02).

3.8: **__Selection of an image sensor__

In Section 3.4.5 we discussed a simple linear noise model for imaging sensors, which proved worthwhile. You have two cameras at hand with the following noise characteristics:

$$\begin{array}{ll} \text{Camera A} & \sigma^2 = 1.0 + 0.1g \\ \text{Camera B} & \sigma^2 = 2.5 + 0.025g \end{array}$$

Both cameras deliver digital signals with 12-bit resolution. Thus gray values g between 0 and 4095 can be measured. Both cameras have a quantum efficiency of 0.5. Which of the two cameras is better suited for the following tasks:

1. Measurement of high gray values with the best possible relative resolution
2. Measurement of the smallest possible irradiation.

In order to decision correctly, compute the standard deviation at the highest digital gray value ($g = 4095$) and at the lowest value (dark image, $g = 0$). Further compute the number of photons that are equal to the standard deviation of the dark image.

3.9: **Covariance propagation

A line sensor has five sensor elements. In a first post processing step, the signals of two neighboring elements are averaged (so called running mean) According to Section 3.3.3 this corresponds to the linear transform

$$\boldsymbol{h} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \boldsymbol{g}.$$

Compute the covariance matrix of $\boldsymbol{h}$ assuming that $\boldsymbol{g}$ is a vector with 5 uncorrelated RVs with equal variance σ^2. Also compute the variance of the mean of $\boldsymbol{h}$ $((h_1 + h_2 + h_3 + h_4)/4)$ and compare it with the variance of the mean of $\boldsymbol{g}$ $((g_1 + g_2 + g_3 + g_4 + g_5)/5)$. Analyze the results!

3.7 *Further Readings*

An introduction to random signals is given by Rice [164]. A detailed account of the theory of probability and random variables can be found in Papoulis [149]. The textbook of Rosenfeld and Kak [172] gives a good introduction to stochastic processes with respect to image processing. Spectral analysis is discussed in Marple Jr. [131].