Chapter 4
Statistical Foundations of Nanoscale Photonic Imaging

Axel Munk, Thomas Staudt and Frank Werner

Essentially, all models are wrong, but some are useful.
— George Box

4.1 Introduction

4.1.1 Background and Examples

The term 'photonic imaging' describes an optical imaging setup where the available measurement data Y are counts of detected photons. The origin of these photons can be diverse in its nature. In coherent X-ray imaging (see e.g. Chap. 2), photons emitted by an X-ray source (like a free electron laser) are scattered (and/or absorbed) by a specimen. In fluorescence microscopy (see e.g. Chap. 1 or Chap. 7), marker molecules are excited by an excitation pulse and emit photons with a certain probability. These two examples are characteristic for the wide range of scenarios arising in photonic imaging: in coherent X-ray imaging we have on the one hand single-molecule diffraction data composed of only few photons [1], and on the other hand

A. Munk (✉) · T. Staudt · F. Werner
Institute for Mathematical Stochastics, Universität Göttingen, Goldschmidtstr. 7,
37077 Göttingen, Germany
e-mail: munk@math.uni-goettingen.de

T. Staudt
e-mail: thomas.staudt@uni-goettingen.de

F. Werner
e-mail: f.werner@math.uni-goettingen.de

A. Munk · F. Werner
Max Planck Institute for Biophysical Chemistry, Am Faßberg 11,
37077 Göttingen, Germany

holographic experiments where millions of photons can be collected from one sample [2]. In fluorescence microscopy, the number of photons is intrinsically limited to a few hundred or thousand per marker due to bleaching effects, and in case of temporally resolved measurements, only a handful of photons is available per time step [3]. Similar restrictions arise in related imaging modalities, including those based on Förster resonance energy transfer (FRET) or metal induced energy transfer (MIET), see e.g. Chap. 8 or [4, 5] for a discussion. Although not within the context of nanoscale imaging, statistically related is astrophysical imaging. Here, there is no a priori limit for the observation time and hence for the number of photons. However, the former is practically limited to several minutes to avoid severe motion blur, see e.g. [6, 7] for examples. We also mention positron emission tomography (PET), where the total number of emitted photons should be as small as possible to minimize the radiation dose for the patient [8]. In all of these applications, detected photons can also originate from undesired background contributions, whose nature strongly depends on the experimental setup, adding additional noise to the observations.

4.1.2 Purpose of the Chapter

The aim of this chapter is to give an overview over prototypical approaches to model the data emerging in photonic imaging from a statistical point of view, based on the physical modeling of photon observation. A sketch of the typical imaging setup we consider is presented in Fig. 4.1.

We assume that the imaging process is described by an underlying photon intensity $\lambda : \Omega \times [0, T] \to [0, \infty)$ at the detector interface, where Ω is the spa-

detection plane Ω

photon source   optical system

photons

bins

signal

photon intensity $\lambda(\cdot, t)$

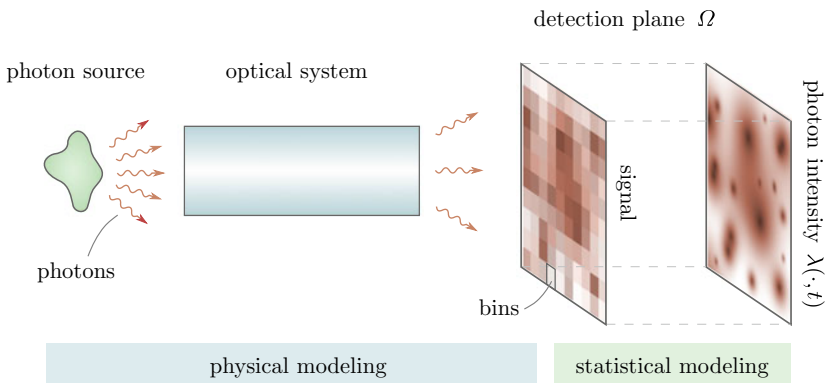physical modeling   statistical modeling

Fig. 4.1 Sketch of the imaging process. A source emits photons that are mapped on a (binned) detection interface Ω through the optical system. The underlying photon intensity $\lambda(\cdot, t)$ at time t, which is determined by the physics of the specific imaging setup, is used for statistical modeling of the detected signal

tial domain of observation (which can be two- or three-dimensional) and T is the total observation time. Let us enumerate the emitted photons by $1, \ldots, N$ and denote their specific detection position and time by $(\mathbf{x}_i, t_i) \in \Omega \times [0, T]$. For a given (measurable) subset $A \subset \Omega$ and time interval $I \subset [0, T]$ we write $Y(A \times I) :=$ $\#\left\{1 \leq i \leq N \mid \mathbf{x}_i \in A, t_i \in I\right\}$ to denote the number of photons observed in A during I. The expected number of photons detected in $A \times I$ is by definition of λ given by

$$\mathbb{E}\left[Y(A \times I)\right] = \int\limits_{I} \int\limits_{A} \lambda(\mathbf{y}, t) \, d\mathbf{y} \, dt .\tag{4.1}$$

Note that this includes all detected photons, including all background contributions. We will always assume $\lambda \geq 0$, which ensures that the integral in (4.1) is well-defined (however it might be ∞).

Throughout this manuscript, we will discuss statistical models for the distribution of the observations Y, depending on the physical measurement setup. We assume λ to be given, as deriving or estimating λ and/or other model parameters described (implicitly) by λ is the topic of other expositions (see e.g. Chap. 5 or Chap. 11).

4.1.3 Measurement Devices

Depending on the type of sensor used for photon detection, different models for photonic imaging settings have been proposed. One commonality of all measurement setups is that the spatial domain of observation Ω is discretized into detector regions, so-called bins. We will assume that the detectors on all bins have identical physical properties, and we denote the centers of such bins by $\mathbf{x} \in \Xi$ with Ξ being the set of all bin centers. If a charge-coupled device (CCD) camera is used for detection, all bins (the pixels of the sensor) can be observed simultaneously. This is e.g. the case in most coherent X-ray experiments or astrophysical imaging. PET requires a tomographic setup consisting of several photomultiplier tubes (PMT) surrounding the patient (see e.g. [9]). In confocal fluorescence microscopy the most widely applied detectors are based on avalanche photodiodes, which can measure photons in one bin at a time only. Hence, the domain of observation Ω is typically scanned by physically moving the specimen (or detector) at a fast pace. Temporal simultaneous photons can be measured as well, requiring a different experimental setup (see e.g. [10]).

Most photon detectors rely on the photoelectric effect. With a certain probability (the quantum efficiency), incident photons will release photo electrons on the detector surface. Since single electrons cannot be detected reliably, the signal is typically amplified by a cascade of electron multiplying systems. This introduces additional noise due to the stochastic nature of the multiplying steps. Another complication is the existence of dead times. The dead time of a detection device refers to the time interval (after activation) during which it is unable to record another event. Dead times can, for example, arise due to the necessity to recharge conductors in-between

measurements, or due to time delays caused by analog-to-digital conversion and data storage. Details on the statistics of different detectors can be found in [11, Cpt. 12].

4.1.4   Structure and Notation

For the remainder of this chapter we will develop and discuss models for the right part in Fig. 4.1 with different degrees of accuracy. The model choice mainly depends on the total number of detected photons and on the spatial and temporal dependency structure of the randomly generated photons. We will start with the Poisson model, which is well-known and most common for many applications. It can be derived immediately from (4.1) under the assumption of independence, which explains its wide use in photonic imaging (see e.g. the reviews [7, 12] and the references therein). However, if it is necessary to count photons on small time scales, or if independence is not given, a more refined modeling is on demand. In these situations, we turn towards Bernoulli and Binomial models subsequently, and discuss to what extend they are compatible with the aforementioned Poisson model. Finally we turn to the case of large counting rates, which lead to Gaussian models based on asymptotic normality. We discuss differences and commonalities arising from the different base models and indicate in which situation which model should be used. This will be linked to different examples from this book, where we argue if our assumptions are met or not.

Let us introduce the basic notation used in this chapter. We will always assume that any observation y is the realization of a random variable Y, and we will denote by $\mathbb{P}$ probabilities w.r.t. this random object. By $\mathbb{E}$ and $\mathbb{V}$ we will denote the expectation and variance w.r.t. $\mathbb{P}$, respectively. The letters $\mathcal{P}$, $\mathcal{B}$ and $\mathcal{N}$ will denote the Poisson, Binomial and normal distribution introduced below. Random variables will always be denoted by capital letters X, X_i, Z etc., and if we write i.i.d. for a sequence $X_1, X_2, \ldots$ of random variables, this stands for independent identically distributed.

4.2   Poisson Modeling

Suppose we have a perfect photon detector that registers the individual arrival times of all emitted photons reaching a bin without missing any. We will focus on describing a single bin for the moment to avoid notational difficulties. In this situation, the total number of collected photons often can be modeled as Poissonian. A random variable X follows a Poisson law with parameter (intensity) $\mu \geq 0$, if

$$\mathbb{P}[X = j] = \frac{\mu^j}{j!} \exp(-\mu), \quad j \in \mathbb{N}_0.$$

We write $X \sim \mathcal{P}(\mu)$. The following fundamental theorem about point processes explains why the Poisson distribution often comes into play when modeling photon counts:

Theorem 4.1 *Suppose we observe a random number N of photons at random arrival times $0 \leq t_1 < \cdots < t_N \leq T$ such that*

(a) *for each choice of disjoint intervals $I_1, \ldots, I_n \subset [0, T]$, the random variables $\#\{1 \leq k \leq N \mid t_k \in I_i\}$, $1 \leq i \leq n$, corresponding to the number of observed photons during I_i are independent, and*

(b) *there exists some integrable function μ on $[0, T]$ such that for any choice $0 \leq a < b \leq T$ it holds*

$$\mathbb{E}\left[\#\{1 \leq k \leq N \mid a \leq t_k \leq b\}\right] = \int_a^b \mu(t)\, \mathrm{d}t.$$

Then, for all $0 \leq a < b \leq T$, the number of photons observed between time a and time b is Poisson distributed with parameter $\int_a^b \mu(t)\, \mathrm{d}t$, i.e.

$$\#\{1 \leq k \leq N \mid a \leq t_k \leq b\} \sim \mathcal{P}\left(\int_a^b \mu(t)\, \mathrm{d}t\right).$$

For the proof we refer to [13, Theorem 1.11.8]. In terms of probability theory, this theorem implies that the point process $X := \sum_{i=1}^N \delta_{t_i}$, with δ_t denoting the Dirac measure at t, is a Poisson point process with intensity μ if the stated assumptions are satisfied.

Let us discuss these assumptions. Condition (b) underlies our whole modeling procedure as described in (4.1) and seems universally evident. Temporal independence of the arrival times in (a) is more critical but seems (at least approximately) reasonable in many imaging modalities where photons arise from a high-intensity source, including coherent X-ray imaging. However, if the photons arise from fluorescent markers, temporal independence can be violated due to hidden internal states of the fluorophores, energy transfer between different fluorophores on small time and spatial scales (e.g. FRET), or dead times of the detectors.

If temporal independence is given, then Theorem 4.1 states that the number $Y_{\mathbf{x},t}$ of collected photons within a bin $B_{\mathbf{x}}$ until time $t \in [0, T]$ can naturally be modeled by a Poissonian random variable with intensity $\int_0^t \int_{B_{\mathbf{x}}} \lambda(\mathbf{y}, \tau)\, \mathrm{d}\mathbf{y}\, \mathrm{d}\tau$. This gives rise to the following model:

Poisson model

Let the spatial domain of observation Ω be discretized into bins $B_{\mathbf{x}}$ with centers $\mathbf{x} \in \Xi$. We assume that our observations are given by a field $Y_t := \left(Y_{\mathbf{x},t}\right)_{\mathbf{x} \in \Xi}$ of random variables such that

$$Y_{\mathbf{x},t} \sim \mathcal{P}\left(\int_0^t \int_{B_{\mathbf{x}}} \lambda\left(\mathbf{y}, \tau\right) \, \mathrm{d}\mathbf{y} \, \mathrm{d}\tau\right), \qquad \mathbf{x} \in \Xi, t \in [0, T] \qquad (4.2)$$

for some intensity function $\lambda \geq 0$.

This is the basis of many popular models covering a variety of distinct applications. Examples include PET (see Vardi et al. [9]), astronomy and fluorescence microscopy (see Bertero et al. [7] or Hohage and Werner [12]), or a more subtle model for CCD cameras due to Snyder et al. [14, 15].

Note that so far we have assumed that all arriving photons are collected by the detector. This will however be never the case due to several physical limitations, see Fig. 4.2.

The specific efficiency depends strongly on the setup and can vary considerably. Additionally to different quantum efficiencies of different detectors, it might also happen that the detector does not cover all of Ω or has some dead subregions (like interfaces between individual elements). This causes a loss of measured photons and hence a statistical thinning of the random variable $Y_{\mathbf{x},t}$. In this case, the actually
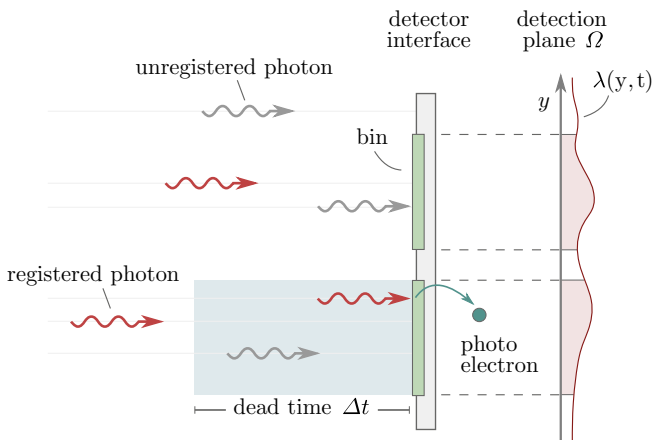

Fig. 4.2 Statistical photon thinning at the detector interface. Photons that reach the detection plane can stay undetected due to various reasons. For example, they can fail to free a photo electron, miss the sensitive regions of the detector bins, or arrive during the dead time caused by a previously recorded photon

observed random variable $\widetilde{Y}_{\mathbf{x},t}$ can be written as

$$\widetilde{Y}_{\mathbf{x},t} = \sum_{i=1}^{Y_{\mathbf{x},t}} X_i \tag{4.3}$$

with Bernoulli random variables X_i having success probability $\eta_i \in [0, 1]$ where each X_i indicates if the ith photon has been detected. If, in addition, the thinning happens identically and independently for each photon, i.e. $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{B}(1, \eta)$, only the parameter in the Poisson law (4.2) changes, but not its distributional structure. More precisely, in this case it follows (see the Appendix) that

$$\widetilde{Y}_{\mathbf{x},t} \sim \mathcal{P}\left(\eta \int_0^t \int_{B_{\mathbf{x}}} \lambda(\mathbf{y}, \tau) \, d\mathbf{y} \, d\tau \right).$$

Consequently, the imperfectness of a detector (as long as the induced thinning happens independent for each photon) can be seen as a scaling of the underlying photon intensity λ by an efficiency factor $\eta \in (0, 1]$. In agreement with Fig. 4.1 we can hence assume that all physical processes causing a thinning have already been treated when modeling λ in the following.

Besides this kind of independent thinning, a further important issue in many imaging modalities is the dead time Δt of the employed detector. Dead times can vary significantly depending on the type of detector, but usually are in the range of nanoseconds. If a photon arrives at time $t \in [0, T)$, the detector will only be able to record the next photon arriving after $t + \Delta t$. Note that whenever $\Delta t > 0$, at most $T/\Delta t$ photons can be detected during the whole measurement, which contradicts (4.2) in the sense that $\mathbb{P}\left[Y_{\mathbf{x},T} > T/\Delta t\right] = 0$ in this case. Such an upper limit on the total number of detected photons can crucially change the distribution, which can, e.g., be seen from the following fact proven in the appendix:

Theorem 4.2 *Fix* $\mathbf{x} \in \varXi$ *and let* $I_1, \ldots, I_m$ *be a decomposition of* $[0, T]$ *into disjoint intervals. Denote by* X_i *the number of photons observed during* I_i *in bin* $B_{\mathbf{x}}$. *Assume model* (4.2), *and suppose that* $X_1, \ldots, X_m$ *are independent. Then the conditional distribution given* $Y_{\mathbf{x},T} = N$ *of* $(X_1, \ldots, X_m)$ *is multinomial with parameter* N *and probability vector* $(p_1, \ldots, p_m)$ *where*

$$p_i = \frac{\int_{I_i} \int_{B_{\mathbf{x}}} \lambda(\mathbf{y}, \tau) \, d\mathbf{y} \, d\tau}{\int_0^T \int_{B_{\mathbf{x}}} \lambda(\mathbf{y}, \tau) \, d\mathbf{y} \, d\tau}.$$

In other words, Theorem 4.2 states that, conditioning on the total number of photons, the arrival times of individual photons behave like a Bernoulli process with intensity $\tau \mapsto \int_{B_{\mathbf{x}}} \lambda(\mathbf{y}, \tau) \, d\mathbf{y}$. This implies that conditioning on the total number

of photons introduces a dependency structure between the number of counts during different time intervals. Consequently, if Δt cannot be neglected, temporal independence is not given anymore, hence corrupting the Poisson law, and different modeling approaches are needed.

4.3   Bernoulli Modeling

To measure the temporal structure of the incoming photons, counting as described above is not sufficient. In such cases, photons are consecutively counted during (short) time frames. We suppose that the discretization of the temporal measurement process is refined such that temporal aggregation underlying the Poisson model is not appropriate anymore. This is described by (equidistant) time frames, which are consecutive intervals $I_1, I_2, \ldots, I_n \subset [0, T]$ of equal length $\delta > 0$, chosen such that the probability to observe more than one photon in each bin $B_{\mathbf{x}}$ during any interval is sufficiently close to 0, and separated by a waiting time $\epsilon > 0$, which allows to ignore the dead time. In this situation, the following model is a reasonable approximation:

Bernoulli model

For $\mathbf{x} \in \Xi$ and $1 \leq i \leq n$ the random variable $Y_{\mathbf{x},i}$ indicating if a photon arrives in bin $B_{\mathbf{x}}$ during the time interval I_i follows a Bernoulli distribution,

$$Y_{\mathbf{x},i} \sim \mathcal{B}\left(1, p_{\mathbf{x},i}\right), \tag{4.4}$$

with success probability

$$p_{\mathbf{x},i} \approx \int\limits_{I_i} \int\limits_{B_{\mathbf{x}}} \lambda\left(\mathbf{y}, \tau\right) \, \mathrm{d}\mathbf{y} \, \mathrm{d}\tau. \tag{4.5}$$

As mentioned before, the detector will hardly count all arriving photons, which causes a statistical thinning as in (4.3). If the thinning happens independently of the photon arrivals, we obtain $\widetilde{Y}_{\mathbf{x},i} \sim \mathcal{B}\left(1, \eta \cdot p_{\mathbf{x},i}\right)$ with the probability η that an incident photon is detected, which immediately follows from $X \cdot Z \sim \mathcal{B}\left(1, pp'\right)$ if $X \sim \mathcal{B}\left(1, p\right)$ is independent of $Z \sim \mathcal{B}\left(1, p'\right)$.

In many imaging setups, it would be difficult to store the whole time series $Y_{\mathbf{x},i}$, for instance due to memory limitations. Examples include fluorescence microscopy setups like confocal, STED or 4Pi microscopy, or coherent X-ray imaging, where millions of photons are observed in short times, which would require an unreasonably fine time discretization. For other examples like SMS microscopy, however, the temporal structure can be important (e.g. for adjusting temporal drifts, see e.g. [16, 17]) and hence most of the data of the above model has to be used. If temporal

dependencies are less important, it is sufficient to count photon arrivals in some interval $I \subset [0, T]$ larger than δ, i.e. to consider $Y_{\mathbf{x}, I} := \sum_{I_i \subset I} Y_{\mathbf{x}, i}$. The distribution of $Y_{\mathbf{x}, I}$ depends strongly on the temporal dependency structure of the $Y_{\mathbf{x}, i}$. In case that they are independent and $p_{\mathbf{x}, i} \equiv p_{\mathbf{x}}$ for all $1 \leq i \leq n$, we obtain a Binomial model:

> ### Binomial model
> For $\mathbf{x} \in \varXi$ and $I \subset [0, T]$, the number of photons observed in the bin centered at $\mathbf{x}$ during the time interval I is
>
> $$Y_{\mathbf{x}, I} \sim \mathcal{B} \left(\# \{ I_i \subset I \}, p_{\mathbf{x}} \right) \qquad (4.6)$$
>
> with $p_{\mathbf{x}, i} \equiv p_{\mathbf{x}}$ for all $1 \leq i \leq n$ and $p_{\mathbf{x}, i}$ as in (4.5).

Note that if we proceed similarly with the thinned observations $\widetilde{Y}_{\mathbf{x}, i}$, we obtain $\widetilde{Y}_{\mathbf{x}, I} \sim \mathcal{B} \left(\# \{ I_i \subset I \}, \eta p_{\mathbf{x}} \right)$, which is the canonical thinning of (4.6), see e.g. [18].

Independence of the $Y_{\mathbf{x}, i}$ is strongly connected to the photon source, as discussed above. If $\epsilon \geq \Delta t$, the dead times of the detectors have no influence on the temporal dependency structure anymore. The second assumption, $p_{\mathbf{x}, i} \equiv p_{\mathbf{x}}$ for all $1 \leq i \leq n$, is equivalent to stationarity of the underlying photon source, which again depends on the imaging modality. If, e.g., a freeze-dried sample is imaged sufficiently fast, then this assumption is reasonable.

Besides temporal dependencies, the field of random variables can also have a spatial dependency structure. In many modalities the random variables are independent for different pixels or voxels $\mathbf{x}$, but on sufficiently small scales some dependency can occur, e.g., due to energy transfer between molecules.

4.3.1 Law of Small Numbers

It is a fundamental and well-known fact that a Binomial distribution can in certain situations be approximated by a Poissonian distribution. In this section, we will discuss how this provides a link between the initial Poisson modeling (4.2) and the preceding Bernoulli modeling (4.6). To this end, we recall the so-called *law of small numbers*, which will be stated in terms of Le Cam's theorem [19]. For the moment we suppress dependencies on $\mathbf{x}$ and consider only a single Binomial random variable, corresponding to a fixed bin.

Theorem 4.3 (Law of small numbers) *Let $X_1, \ldots, X_m$ be independent and Bernoulli distributed with success probabilities $q_1, \ldots, q_m$. Then the distribution of $X := X_1 + \cdots + X_m$ can be approximated by $\mathcal{P}(\lambda_m)$ with $\lambda_m = -\sum_{i=1}^{m} \log (1 - q_i)$. More precisely it holds that*
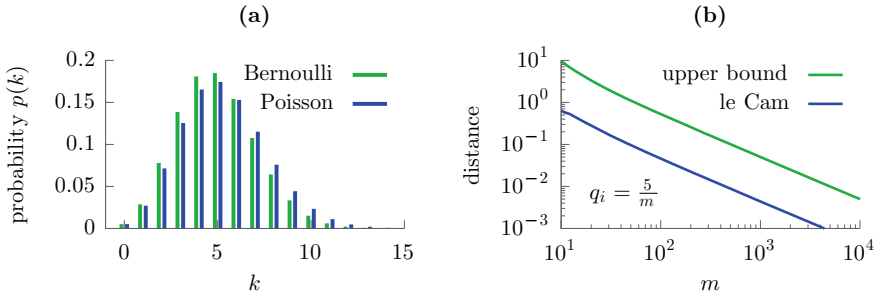
Fig. 4.3 Law of small numbers. Figure **(a)** shows the probabilities for a sum of 50 Bernoulli variables with $q_i = 0.1$ and the respective Poisson approximation with $\lambda = -50 \log(0.9) \approx 5$. Figure **(b)** depicts the left hand side of (4.7) and the corresponding upper bound (right hand side of (4.7)) of Theorem 4.3 for increasing m and $q_i := 5/m$

$$\sum_{k=0}^{\infty} \left| \mathbb{P}[X = k] - \frac{\lambda_m^k}{k!} \exp(-\lambda_m) \right| \leq 2 \sum_{i=1}^{m} (\log(1 - q_i))^2 . \qquad (4.7)$$

For a textbook proof we refer to [20, Theorem 5.1]. Figure 4.3 visualizes the law of small numbers. Note that the bound on the right-hand side of (4.7) can be simplified by using $\log(1 - x) \leq -x$, resulting in $\sum_{i=1}^{m} q_i^2$. We furthermore refer to [21, Proposition 4.3 and 4.4], where bounds of the supremum instead of the sum over k on the left-hand side are given. Note that Theorem 4.3 can be generalized to dependent Bernoulli random variables at the price of a worse upper bound, see e.g. [20, Theorem 5.5].

A classical example for this law is the situation when $q_i \equiv q_m$ for all $1 \leq i \leq m$ and $q_m \cdot m$ converges to some $\lambda > 0$, i.e., $q_m \sim 1/m$. In this case we may use $\log(1 - x) \approx -x$ for small x to obtain $\lambda_m \approx m q_m \to \lambda$ and $2 \sum_{i=1}^{m} (\log(1 - q_i))^2 \approx 2 \sum_{i=1}^{m} q_i^2 = 2m q_m^2 \sim 1/m \to 0$ as $m \to \infty$, i.e., the Binomial distribution of X converges rapidly to the Poisson distribution with parameter λ.

On the other hand, if the success probabilities $q_i \equiv q \in (0, 1)$ are fixed, the right-hand side of (4.7) diverges. This seems intuitive, as in this situation convergence towards a normal distribution has to be expected (cf. Sect. 4.4.1 below). This is in line with the observation that a Poisson distribution with growing parameter $\lambda_m = -m \log(1 - q)$ converges towards a normal distribution (cf. Sect. 4.4.2 below).

Let us now compare the two Poisson laws arising from Theorem 4.3 and (4.2). According to (4.4), our observations are Binomial random variables with success probability

$$p_{\mathbf{x},i} \approx \int_{I_i} \int_{B_{\mathbf{x}}} \lambda(\mathbf{y}, \tau) \, d\mathbf{y} \, d\tau,$$

where we used that the probability to observe more than one photon is close to 0. Hence, if we denote the largest time in I_m by t_m, and use again $\log(1 - x) \approx -x$,

then the number of photons observed until time t_m is approximately Poisson distributed with parameter $\lambda_{\mathbf{x},m} = p_{\mathbf{x},1} + \cdots + p_{\mathbf{x},m} \approx \int_0^{t_m} \int_{B_{\mathbf{x}}} \lambda\left(\mathbf{y}, \tau\right) \, \mathrm{d}\mathbf{y} \, \mathrm{d}\tau$ ignoring the waiting times. This is in good agreement with (4.2). According to (4.7), the error in this approximation is bounded by

$$2 \sum_{i=1}^{m} \left(\log\left(1 - p_{\mathbf{x},i}\right)\right)^2 \leq 2\lambda_{\mathbf{x},m} \max_{1 \leq i \leq m} p_{\mathbf{x},i} \leq C \max_{1 \leq i \leq m} |I_i|,$$

revealing it valid whenever the temporal discretization is sufficiently fine.

4.4   Gaussian Modeling

4.4.1   As Approximation of the Binomial Model

Besides the approximation by a Poisson distribution, it is well-known that a Binomial model can also be approximated by a Gaussian one under suitable circumstances. Let us start with the Bernoulli model (4.4) and suppose that all $Y_{\mathbf{x},i}$ are independent with $p_{\mathbf{x},i} \equiv p_{\mathbf{x}}$. If we are interested in the total number of counts $Y_{\mathbf{x}} := \sum_{i=1}^{n} Y_{\mathbf{x},i}$ in bin $\mathbf{x}$, the de Moivre-Laplace theorem states that

$$\frac{Y_{\mathbf{x}} - np_{\mathbf{x}}}{\sqrt{np_{\mathbf{x}}\left(1 - p_{\mathbf{x}}\right)}} \longrightarrow Z \quad \text{as} \quad n \to \infty \tag{4.8}$$

in distribution, where $Z \sim \mathcal{N}\left(0, 1\right)$ follows a standard normal distribution. Note that $\frac{Y_{\mathbf{x}} - np_{\mathbf{x}}}{\sqrt{np_{\mathbf{x}}(1-p_{\mathbf{x}})}}$ is just the centered and standardized version of the total number of counts $Y_{\mathbf{x}}$. This implies that the distribution of $Y_{\mathbf{x}}$ can be approximated by a Gaussian distribution with mean $np_{\mathbf{x}}$ and variance $np_{\mathbf{x}}\left(1 - p_{\mathbf{x}}\right)$ if n is sufficiently large. This gives rise to a first Gaussian model:

Gaussian model I

For each $\mathbf{x} \in \varXi$, the number of photons observed in the bin centered at $\mathbf{x}$ up to time T is

$$Y_{\mathbf{x}} \sim \mathcal{N}\left(np_{\mathbf{x}}, np_{\mathbf{x}}\left(1 - p_{\mathbf{x}}\right)\right) \tag{4.9}$$

where $n = n(T) \sim T/\delta$ with the length δ of the individual time frames.

The rate of convergence in (4.8) can be made more precise. For instance a special case of the Berry-Esseen theorem states

$$\sup_{y \in \mathbb{R}} \left| \mathbb{P}\left[\frac{Y_\mathbf{x} - np_\mathbf{x}}{\sqrt{np_\mathbf{x}(1 - p_\mathbf{x})}} \leq y \right] - \Phi(y) \right| < \frac{\sqrt{10} + 3}{6\sqrt{2\pi}} \frac{p_\mathbf{x}^2 + (1 - p_\mathbf{x})^2}{\sqrt{np_\mathbf{x}(1 - p_\mathbf{x})}} \qquad (4.10)$$

where Φ denotes the distribution function of $\mathcal{N}(0, 1)$, i.e.,

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \exp\left(-\frac{x^2}{2}\right) \, \mathrm{d}x.$$

In fact, the constant on the right-hand side of (4.10) cannot be improved [22]. An interpretation of this theorem is that the approximation leading to the model (4.9) is reasonable as soon as $np_\mathbf{x}(1 - p_\mathbf{x}) > 9$, which implies the right-hand side of (4.10) to be bounded by $\frac{\sqrt{10}+3}{18\sqrt{2\pi}} \approx 0.137$.

If the success probabilities $p_{\mathbf{x},i}$ do vary in i, the de Moivre-Laplace theorem (4.8) cannot be applied immediately. However, it is still possible, under certain conditions, to derive an approximate Gaussian model of the form (4.9) by applying the Lindeberg central limit theorem (see e.g. [23]). It states that the sum $Y_\mathbf{x}$, after centralization and standardization, still converges to $\mathcal{N}(0, 1)$ in distribution even for non identically distributed $Y_{\mathbf{x},i}$. This motivates a second Gaussian model:

Gaussian model II

For each $\mathbf{x} \in \varXi$, the number of photons observed in the bin centered at $\mathbf{x}$ up to time T is

$$Y_\mathbf{x} \sim \mathcal{N}\left(\sum_{i=1}^{n} p_{\mathbf{x},i}, \sum_{i=1}^{n} p_{\mathbf{x},i}(1 - p_{\mathbf{x},i}) \right). \qquad (4.11)$$

Note that, if the random variables $Y_{\mathbf{x},i}$ are dependent, the type of dependency very much determines whether a central limit theorem is still valid (with different limiting variance), see e.g. [24] or [25–27] for mixing sequences, and [28] for martingale difference sequences, to mention two large classes of examples.

4.4.2   As Approximation of the Poisson Model

The Poisson model in (4.2) can also be approximated by a Gaussian one. This relies on the fact that the Poisson distribution is infinitely divisible, which means that whenever $X \sim \mathrm{Poi}(\mu)$, then X can be represented as $X = X_1 + \cdots + X_n$ for any $n \in \mathbb{N}$ with i.i.d. random variables $X_1, \ldots, X_n \sim \mathrm{Poi}(\mu/n)$. Consequently, the central limit theorem states that

$$\frac{X - \mu}{\sqrt{\mu}} \longrightarrow Z, \quad \text{as} \quad \mu \to \infty$$

with $Z \sim \mathcal{N}(0, 1)$. The general Berry-Esseen theorem can also be used to bound the error of an approximation of $\frac{X-\mu}{\sqrt{\mu}}$ by Z, namely one obtains (see also [29])

$$\sup_{y \in \mathbb{R}} \left| \mathbb{P}\left[\frac{X-\mu}{\sqrt{\mu}} \leq y \right] - \Phi(y) \right| < \frac{5}{2} \frac{1}{\sqrt{\mu}}. \tag{4.12}$$

Hence, if μ is sufficiently large, the distribution of X can be approximated by a Gaussian distribution with mean and variance μ. If we suppose that $Y_{\mathbf{x},t}$ satisfies (4.2) and that $\int_0^t \int_{B_\mathbf{x}} \lambda(\mathbf{y}, \tau)\, \mathrm{d}\mathbf{y}\, \mathrm{d}\tau \to \infty$ as $t \to \infty$, then the above reasoning gives rise to another Gaussian model:

Gaussian model III

For each $\mathbf{x} \in \varXi$, the number of photons observed in the bin centered at $\mathbf{x}$ up to time t is

$$Y_{\mathbf{x},t} \sim \mathcal{N}\left(\int_0^t \int_{B_\mathbf{x}} \lambda(\mathbf{y}, \tau)\, \mathrm{d}\mathbf{y}\, \mathrm{d}\tau, \int_0^t \int_{B_\mathbf{x}} \lambda(\mathbf{y}, \tau)\, \mathrm{d}\mathbf{y}\, \mathrm{d}\tau \right). \tag{4.13}$$

4.4.3 *Comparison*

Let us briefly compare the Gaussian models I-III in (4.9), (4.11) and (4.13) respectively. It is clear that (4.11) is a generalization of (4.9) to the case of non-identical success probabilities $p_{\mathbf{x},i}$, and both coincide if $p_{\mathbf{x},i}$ is independent of i. To compare (4.11) with (4.13), we recall our previous computation that $p_{\mathbf{x},1} + \cdots + p_{\mathbf{x},n} = \int_0^{t_n} \int_{B_\mathbf{x}} \lambda(\mathbf{y}, \tau)\, \mathrm{d}\mathbf{y}\, \mathrm{d}\tau \to \infty$ where t_n is the largest time in the sub-interval I_n. Consequently, (4.11) and (4.13) differ only in the variance by $1 - p_{\mathbf{x},i}$, which is usually small. Hence, all three Gaussian models are in good agreement, and (4.13) can be considered the most simple one which should be used.

4.4.4 *Thinning*

Taking into account the detection efficiency $\eta \in [0, 1]$ as discussed before, we will arrive at models similar to (4.9), (4.11) and (4.13) with the only difference being that $p_\mathbf{x}$, $p_{\mathbf{x},i}$ or λ are multiplied by η. In this sense, the canonical thinning of the Poisson or Binomial models carries over to the Gaussian one.

4.4.5   Variance Stabilization

Note that the variance in the Gaussian models I-III is always inhomogeneous, which hinders data analysis with standard methods and causes further difficulties. This can be overcome by variance stabilization. The most popular choice is the celebrated Anscombe transform, which is applied to the Poisson model (4.2) to obtain asymptotically a normal distribution with variance 1. It is based on the following result (see e.g. [30, Lemma 1]):

Lemma 4.1 (Anscombe's transform) *Let $\mu > 0$ and $Y \sim \mathcal{P}(\mu)$ be a Poisson distributed random variable. Then it holds for all $c \geq 0$ that*

$$\mathbb{E}\left[2\sqrt{Y+c}\right] = 2\sqrt{\mu} + \frac{4c-1}{4\sqrt{\mu}} + \mathcal{O}\left(\frac{1}{\mu^{\frac{3}{2}}}\right),$$

$$\mathbb{V}\left[2\sqrt{Y+c}\right] = 1 + \frac{3-8c}{8\mu} + \mathcal{O}\left(\frac{1}{\mu^2}\right).$$

From this we can conclude that the choice $c = 3/8$ ensures that the variance of $2\sqrt{Y+c}$ does no longer depend on the parameter μ up to second order. To reduce the bias, $c = 1/4$ is the best choice. Furthermore, applying this result to the Poisson model in (4.2) gives rise to a fourth Gaussian model:

Gaussian model IV

For each $\mathbf{x} \in \varXi$, denote the number of photons observed in the bin centered at $\mathbf{x}$ up to time t by $Y_{\mathbf{x},t}$. Then we assume

$$2\sqrt{Y_{\mathbf{x},t} + \frac{3}{8}} \sim \mathcal{N}\left(2\left(\int\limits_0^t \int\limits_{B_\mathbf{x}} \lambda(\mathbf{y}, \tau)\, \mathrm{d}\mathbf{y}\, \mathrm{d}\tau\right)^{1/2}, 1\right) \qquad (4.14)$$

for each $\mathbf{x} \in \varXi$.

We emphasize the importance of the model (4.14) in statistics, as it turns out to be equivalent in a strict sense to the previously discussed Poisson model (4.2) as the total number of photons (and hence the parameter t) tends to ∞ (see e.g. [31–33]).

4.5   Conclusion

In this chapter we introduced models for photonic imaging setups with different degrees of accuracy. The most common and basic Poisson model (4.2) is accurate as

soon as the temporal dependency can be neglected and the detector has no significant dead time. If furthermore the number of observed photons is sufficiently large on each bin, then the Gaussian model (4.13) can be used. In case of significant temporal dependency, the Bernoulli model (4.4) with time resolved individual photon arrivals or the resulting Binomial model (4.6) should be considered instead.

An overview about appropriate model choices for the various imaging techniques discussed previously is provided in Fig. 4.4.

In fluorescence microscopy, STED based methods, which scan the sample pixelwise, record about 10–100 photons per fluorescent marker. Due to low temporal dependencies, we are thus in the scope of the binomial or Poisson models [3]. Even though a Gaussian approximation seems questionable as in regions of low intensities only a few photons per bin can be collected, it has been successfully applied employing variance stabilizing techniques [34]. In order to analyze STORM/PALM data, the full range of modeling approaches is applied. Individual frames contain spots with single or several photons and weak temporal dependency, calling for Bernoulli, binomial, or Poisson models, while Gaussian approximations are used successfully for drift and rotational corrections [17]. FRET/MIET based imaging heavily relies on the interactions of fluorescent markers, so that the assumption of temporal independence is violated. This makes the Bernoulli model the model of choice, or if more photons are counted, also the Binomial model can be applied [4, 5].

Another example in the scope of the Bernoulli model is the 3-photon correlation technique (see e.g. Chap. 16), where molecular structures are probed by femtosecond X-ray pulses. This leads to a high number of images consisting of a few photons
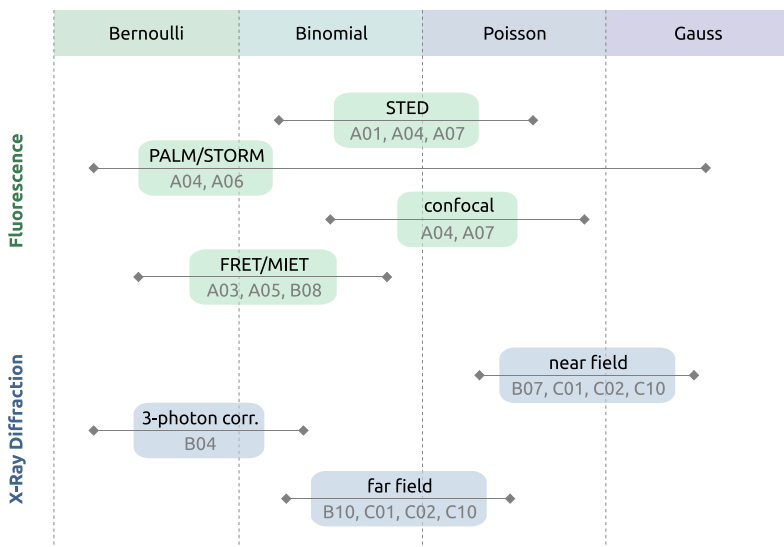

Fig. 4.4 Overview over viable model choices for different imaging methods. Projects of the SFB 755 associated to the respective methods are marked in gray

only, out of which only triples are used. Inference based on this sequence of images is additionally complicated by rotations of the single target molecules [1].

X-ray diffraction imaging also allows for a whole range of models. On first glance it seems that a Gaussian model is sufficient, as in total millions of photons are collected. However, depending on the specific setup, the photon intensity λ may vary strongly over the detection region. If imaging is performed in a near-field regime, as e.g. in many X-ray microscopy setups, the number of photons in the lower intensity regions is about one order of magnitude lower than in the high intensity regions, allowing for a Gaussian model. In contrast to this are far field methods where on high intensity bins 10^4 photons can be collected, but in low intensity regions only a handful of photons arrives, revealing a Binomial and/or Poisson model more suitable [12].

Appendix: Poisson Thinning

Let $\mu > 0$, $\eta \in (0, 1)$ and suppose $Y \sim \mathcal{P}(\mu)$, $X_1, X_2, \ldots \sim \mathcal{B}(1, \eta)$ independent. The η-thinning of Y is defined as

$$\tilde{Y} := \sum_{i=1}^{Y} X_i.$$

We will now show that the distribution of $\tilde{Y}$ is still Poissonian, but with parameter $\eta \cdot \mu$. To this end, observe that the probability of $\tilde{Y}$ being k is given by the sum over all probabilities of Y being l and exactly k out of the first l X_i's being 1, i.e.

$$\mathbb{P}\left[\tilde{Y} = k\right] = \sum_{l=k}^{\infty} \mathbb{P}\left[Y = l, \sum_{i=1}^{l} X_i = k\right] = \sum_{l=k}^{\infty} \mathbb{P}\left[Y = l\right] \mathbb{P}\left[\sum_{i=1}^{l} X_i = k\right]$$

by independence. Inserting the Poisson distribution of Y and the Binomial distribution of $\sum_{i=1}^{l} X_i$ gives

$$\mathbb{P}\left[\tilde{Y} = k\right] = \sum_{l=k}^{\infty} \frac{\mu^l}{l!} \exp(-\mu) \binom{l}{k} \eta^k (1-\eta)^{l-k}$$

$$= \frac{\exp(-\mu)}{k!} (\eta\mu)^k \sum_{l=k}^{\infty} \frac{(\mu(1-\eta))^{l-k}}{(l-k)!}$$

$$= \frac{\exp(-\mu)}{k!} (\eta\mu)^k \exp(\mu(1-\eta))$$

$$= \frac{(\eta\mu)^k}{k!} \exp(-\mu\eta),$$

which proves $\tilde{Y} \sim \mathcal{P}(\eta\mu)$.

Appendix: Conditioned Poisson Processes

Suppose we observe a random number N of photons at random arrival times $0 \le t_1 < \cdots < t_N \le T$ such that the number of photons between time a and time b is Poisson distributed with parameter $\int_a^b \mu(t)\,dt$ for a fixed function $\mu \ge 0$. Given a decomposition of $[0, T]$ into disjoint intervals $I_1, \ldots, I_m$, denote by

$$Y_i := \#\left\{1 \le j \le N \,\big|\, t_j \in I_i\right\}, \qquad 1 \le i \le m$$

the number of photons observed during I_i. Assume furthermore that $Y_1, \ldots, Y_m$ are independent. We will now show that the conditional distribution given $N = n$ of $(Y_1, \ldots, Y_m)$ is multinomial with parameter n and probability vector $(p_1, \ldots, p_m)$ where

$$p_i = \frac{\int_{I_i} \mu(t)\,dt}{\int_0^T \mu(t)\,dt}.$$

Therefore let $n_1, \ldots, n_m \in \mathbb{N}_0$ such that $\sum_{i=1}^m n_i = n$. Then we have

$$\mathbb{P}\left[Y_1 = n_1, \ldots, Y_m = n_m \mid N = n\right] = \frac{\mathbb{P}\left[Y_1 = n_1, \ldots, Y_m = n_m, N = n\right]}{\mathbb{P}\left[N = n\right]}$$

$$= \frac{\prod_{i=1}^m \exp\left(-\int_{I_i} \mu(t)\,dt\right) \frac{1}{n_i!} \left(\int_{I_i} \mu(t)\,dt\right)^{n_i}}{\exp\left(-\int_0^T \mu(t)\,dt\right) \frac{1}{n!} \left(\int_0^T \mu(t)\,dt\right)^n}$$

$$= n! \prod_{i=1}^m \frac{p_i^{n_i}}{n_i!}$$

which proves the claim.

References

1. von Ardenne, B., Mechelke, M., Grubmüller, H.: Structure determination from single molecule x-ray scattering with three photons per image. Nat. Commun. **9**, 2375 (2018)
2. Bartels, M., Krenkel, M., Haber, J., Wilke, R.N., Salditt, T.: X-ray holographic imaging of hydrated biological cells in solution. Phys. Rev. Lett. **114**, 048,103 (2015). https://doi.org/10.1103/PhysRevLett.114.048103
3. Aspelmeier, T., Egner, A., Munk, A.: Modern statistical challenges in high-resolution fluorescence microscopy. Annu. Rev. Stat. Appl. **2**, 163–202 (2015)
4. Graen, T., Hoefling, M., Grubmüller, H.: Amber-dyes: Characterization of charge fluctuations and force field parameterization of fluorescent dyes for molecular dynamics simulations. J. Chem. Theory Comput. **10**(12), 5505–5512 (2014). https://doi.org/10.1021/ct500869p. PMID: 26583233
5. Michalet, X., Weiss, S., Jäger, M.: Single-molecule fluorescence studies of protein folding and conformational dynamics. Chem. Rev. **106**(5), 1785–1813 (2006). https://doi.org/10.1021/cr0404343
6. Adorf, H.M.: Hubble space telescope image restoration in its fourth year. Inverse Probl. **11**(4), 639 (1995). http://stacks.iop.org/0266-5611/11/i=4/a=003
7. Bertero, M., Boccacci, P., Desiderà, G., Vicidomini, G.: Image deblurring with Poisson data: from cells to galaxies. Inverse Probl. **25**(12), 025,004, 18 (2009). https://doi.org/10.1088/0266-5611/25/12/123006
8. Sawatzky, A., Brune, C., Wubbeling, F., Kosters, T., Schafers, K., Burger, M.: Accurate em-tv algorithm in pet with low snr. In: 2008 IEEE Nuclear Science Symposium Conference Record, pp. 5133–5137 (2008). https://doi.org/10.1109/NSSMIC.2008.4774392
9. Vardi, Y., Shepp, L.A., Kaufman, L.: A statistical model for positron emission tomography. J. Am. Stat. Assoc. **80**(389), 8–37 (1985). With discussion
10. Ta, H., Keller, J., Haltmeier, M., Saka, S.K., Schmied, J., Opazo, F., Tinnefeld, P., Munk, A., Hell, S.W.: Mapping molecules in scanning far-field fluorescence nanoscopy. Nat. Commun. **6**, 7977 (2015)
11. Pawley, J. (ed.): Handbook of Biological Confocal Microscopy. Springer (2006)
12. Hohage, T., Werner, F.: Inverse problems with poisson data: statistical regularization theory, applications and algorithms. Inverse Probl. **32**, 093,001, 56 (2016)
13. Kerstan, J., Matthes, K., Mecke, J.: Infinitely divisible point processes. Wiley Series in Probability and Mathematical Statistics. Wiley (1978)
14. Snyder, D.L., Helstrom, C.W., Lanterman, A.D., White, R.L., Faisal, M.: Compensation for readout noise in CCD images. J. Opt. Soc. Am. **12**(2), 272–283 (1995)
15. Snyder, D.L., White, R.L., Hammoud, A.M.: Image recovery from data acquired with a charge-coupled-device camera. J. Opt. Soc. Am. **10**(5), 1014–1023 (1993)
16. Geisler, C., Hotz, T., Schönle, A., Hell, S.W., Munk, A., Egner, A.: Drift estimation for single marker switching based imaging schemes. Opt. Express **20**(7), 7274–7289 (2012). https://doi.org/10.1364/OE.20.007274. http://www.opticsexpress.org/abstract.cfm?URI=oe-20-7-7274
17. Hartmann, A., Huckemann, S., Dannemann, J., Laitenberger, O., Geisler, C., Egner, A., Munk, A.: Drift estimation in sparse sequential dynamic imaging, with application to nanoscale fluoresence microscopy. J. Roy. Stat. Soc. Ser. B **78**(3), 563–587 (2016). https://doi.org/10.1111/rssb.12128
18. Harremoës, P., Johnson, O., Kontoyiannis, I.: Thinning and the law of small numbers. In: IEEE International Symposium on Information Theory, 2007. ISIT 2007, pp. 1491–1495. IEEE (2007)
19. Le Cam, L.: An approximation theorem for the poisson binomial distribution. Pac. J. Math. **10**(4), 1181–1197 (1960)
20. den Hollander, F.: Probability Theory: The Coupling Method (2012)
21. Novak, S.Y.: Extreme value methods with applications to finance. Monographs on Statistics and Applied Probability, vol. 122. CRC Press, Boca Raton, FL (2012)

22. Schulz, J.: The optimal berry-esseen constant in the binomial case. Ph.D. thesis, Univeristy of Trier (2016)
23. Billingsley, P.: Probability and Measure. Wiley (2008)
24. Peligrad, M.: On the central limit theorem for triangular arrays of ϕ-mixing sequences. In: Asymptotic Methods in Probability and Statistics (Ottawa, ON, 1997), pp. 49–55. North-Holland, Amsterdam (1998). https://doi.org/10.1016/B978-044450083-0/50005-8
25. Bradley, R.C.: Introduction to Strong Mixing Conditions, vol. 1. Kendrick Press, Heber City, UT (2007)
26. Bradley, R.C.: Introduction to Strong Mixing Conditions, vol. 2. Kendrick Press, Heber City, UT (2007)
27. Bradley, R.C.: Introduction to Strong Mixing Conditions, vol. 3. Kendrick Press, Heber City, UT (2007)
28. Shorack, G.R.: Probability for Statisticians. Springer Texts in Statistics. Springer (2000)
29. Lane, J.A.: The Berry-Esseen bound for the Poisson shot-noise. Adv. Appl. Probab. **19**(2), 512–514 (1987). https://doi.org/10.2307/1427432
30. Brown, L., Cai, T.T., Zhang, R., Zhao, L., Zhou, H.: The root-unroot algorithm for density estimation as implemented via wavelet block thresholding. Probab. Theory Relat. Fields **146**(3–4), 401–433 (2010)
31. Grama, I.: Gaussian approximation for nonparametric models. Tatra Mt. Math. Publ. **17**, 219–226 (1999)
32. Grama, I., Nussbaum, M.: Asymptotic equivalence for nonparametric regression. Math. Methods Stat. **11**(1), 1–36 (2002)
33. Ray, K., Schmidt-Hieber, J.: The le cam distance between density estimation, poisson processes and gaussian white noise. Math. Stat. Learn. **1**, 101–170 (2018)
34. Frick, K., Marnitz, P., Munk, A.: Statistical multiresolution estimation for variational imaging: with an application in Poisson-biophotonics. J. Math. Imaging Vis. **46**(3), 370–387 (2013). https://doi.org/10.1007/s10851-012-0368-5