Raimund J. Ober, E. Sally Ward, Jerry Chao

Appendix
Mathematical Background

Quantitative Bioimaging
An Introduction to Biology, Instrumentation, Experiments, and Data Analysis for Scientists and Engineers

Contents

A

Probability and Statistics

A.1 Tutorial: Poisson and Gaussian random variables

In this tutorial we provide a brief review of the basic definitions and properties of Poisson and Gaussian random variables. This material is of relevance primarily in the last part of this book on data analysis, although we also briefly discuss related random phenomena in Sections 4.3 and 12.7.1. We begin with a concise treatment of the basics of random variables.

The mathematical language of probability theory and statistics relies on the concept of a *random variable*. A random variable is used to provide a mathematical description of the random events that are being analyzed. If we would like to analyze the results of a coin tossing experiment, we can do it with the help of a random variable, denoted here with X. The possible values of this random variable are $\{-1, 1\}$, where -1 stands for heads and 1 stands for tails. We also make use of a very abstract set Ω of possible random events. The random variable is formally a function that attributes to each random event $\omega \in \Omega$ an outcome of the coin tossing experiment, meaning an element in $\{-1, 1\}$. For example, $X(\omega) = 1$ if the result of the coin toss is tails.

The random variable that we just discussed is an example of a *discrete random variable* as it has discrete values. Continuous random variables, meaning random variables with values in the real numbers, play an equally important part in probability theory and statistics. For example, the distance that an athlete will throw a javelin in a certain attempt might be modeled as a continuous random variable with values in the real numbers.

The probability with which a certain random event occurs is described by the probability mass function p for a discrete random variable X, i.e.,

$$p(k) := P(X = k), \quad k = \ldots, -2, -1, 0, 1, 2, \ldots$$

denotes the probability that the outcome of the random event is k. As all probabilities have to sum to 1, we have that

$$\sum_{k=-\infty}^{\infty} p(k) = \sum_{k=-\infty}^{\infty} P(X = k) = 1.$$

For a probabilistic description of continuous random variables, we consider the probability that the outcome of the continuous random variable is in a subset A of the real numbers, i.e., $P(X \in A)$. We will often encounter the situation where, for each (reasonably well-behaved) subset A of the real numbers, this expression can be written as

$$P(X \in A) = \int_A p(x)dx$$

for a positive function p, which is called the *probability density function* for the probability distribution P. The fact that P defines a probability distribution on the real numbers implies that

$$\int_{-\infty}^{\infty} p(x)dx = 1.$$

Also note that for probability distributions P for which a probability density exists, we immediately have that the probability that the outcome of the random event is a specific real number must be 0, as for any real number a, we have

$$P(X \in \{a\}) = \int_{\{a\}} p(x)dx = \int_a^a p(x)dx = 0.$$

Continuous random variables can also be defined such that they have vector values. For example, in Part IV of this book, we will have occasion to consider the random impact points of photons in an idealized infinite detector plane.

An important notion in probability theory is that of the *expected value* of a random variable. The expected value indicates an "average" or "expected" outcome of a random experiment. In fact, the definition of the expected value for a discrete random variable illustrates this well. For a discrete random variable, any number k is a potential outcome. We "average" all the possible outcomes, each weighted by its probability. The expected value $\mathrm{E}[X]$ of a discrete random variable X is therefore defined by

$$\mathrm{E}[X] := \sum_{k=-\infty}^{\infty} k \cdot P(X = k) = \sum_{k=-\infty}^{\infty} k \cdot p(k). \tag{A.1}$$

For the coin tossing example, if the probabilities of heads and tails are $\frac{1}{2}$ each, the expected value of the associated random variable X is immediately calculated as

$$\begin{aligned}\mathrm{E}[X] &= \sum_{k=-\infty}^{\infty} k \cdot p(k) \\ &= \cdots + (-2) \cdot p(-2) + (-1) \cdot p(-1) + 0 \cdot p(0) + 1 \cdot p(1) + 2 \cdot p(2) + \cdots \\ &= \cdots + (-2) \cdot 0 + (-1) \cdot \frac{1}{2} + 0 \cdot 0 + 1 \cdot \frac{1}{2} + 2 \cdot 0 + \cdots \\ &= (-1) \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0.\end{aligned}$$

This shows that the expected value of the random coin tossing experiment is 0. Intuitively, this is the correct result because with equal probability, the outcome of the experiment is either positive or negative. Therefore the average outcome is 0. For a continuous random variable, the analogous definition is

$$\mathrm{E}[X] := \int_{-\infty}^{\infty} x \cdot p(x)dx.$$

These definitions immediately imply the following so-called linearity property of the expectation. If we have two random variables X and Y and real numbers α and β, then the expectation of the linear combination of X and Y is the linear combination of the expectations of X and Y, or formally,

$$\mathrm{E}[\alpha X + \beta Y] = \alpha \mathrm{E}[X] + \beta \mathrm{E}[Y].$$

While the expected value of a random variable gives an expression for the "average" outcome of a random experiment, the variance is defined so that it provides an expression for the variation of the outcome around the expected value. The variation is given in terms of the square of the difference between an outcome and the expected value. The variance $\mathrm{Var}(X)$ of a random variable X is defined in the sense of an expected value of the variation, i.e.,

$$\mathrm{Var}(X) := \mathrm{E}\left[(X - \mathrm{E}[X])^2\right]. \tag{A.2}$$

For the coin tossing example, we can immediately compute the variance to be

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}\left[(X-0)^2\right] = \mathrm{E}\left[X^2\right] = \sum_{k=-\infty}^{\infty} k^2 \cdot p(k) \\ &= \cdots + (-2)^2 \cdot p(-2) + (-1)^2 \cdot p(-1) + 0^2 \cdot p(0) + 1^2 \cdot p(1) + 2^2 \cdot p(2) + \cdots \\ &= \cdots + (-2)^2 \cdot 0 + (-1)^2 \cdot \frac{1}{2} + 0^2 \cdot 0 + 1^2 \cdot \frac{1}{2} + 2^2 \cdot 0 + \cdots \\ &= 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 1.\end{aligned}$$

This shows that the variance of the coin tossing experiment is 1.

More generally, we define the expectation of a function g of a random variable X by

$$\mathrm{E}[g(X)] := \sum_{k=-\infty}^{\infty} g(k) \cdot P(X = k)$$

in the case of a discrete random variable, and by

$$\mathrm{E}[g(X)] := \int_{-\infty}^{\infty} g(x) \cdot p(x) dx$$

for a continuous random variable with probability density p.

A.1.1 Poisson random variables

The number of photons emitted by a light source, such as by a fluorescence-emitting object, can be modeled as a Poisson random variable. Consequently, the light signal impacting a camera pixel can be modeled with a Poisson random variable.

A random variable X with values in $\{0, 1, 2, \dots\}$ is called a *Poisson random variable* if its probability distribution function (i.e., probability mass function) is given by

$$p_\lambda(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots, \tag{A.3}$$

where λ, $0 < \lambda < \infty$, is the parameter of the distribution. For verification that p_λ is indeed a probability distribution, we show that it sums to 1 over all possible values of k:

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

Applying the definition of the expectation of a random variable (Eq. (A.1)), the mean of a Poisson random variable X is given by

$$\begin{aligned}\mathrm{E}[X] &= \sum_{k=0}^{\infty} k \cdot p_\lambda(k) = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^k}{k!} \\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.\end{aligned} \tag{A.4}$$

The mean of a Poisson random variable is therefore the parameter λ of its probability distribution. To obtain the variance of a Poisson random variable X, we start with the general expression for the variance of a random variable (Eq. (A.2)), and obtain

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}\left[(X-\mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2 - 2X\cdot\mathrm{E}[X] + (\mathrm{E}[X])^2\right] \\ &= \mathrm{E}\left[X^2\right] - \mathrm{E}[2X\cdot\mathrm{E}[X]] + \mathrm{E}\left[(\mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2\right] - 2\mathrm{E}[X]\,\mathrm{E}[X] + (\mathrm{E}[X])^2 \\ &= \mathrm{E}\left[X^2\right] - 2(\mathrm{E}[X])^2 + (\mathrm{E}[X])^2 = \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2 = \mathrm{E}\left[X^2\right] - \lambda^2.\end{aligned}$$

We then calculate $\mathrm{E}\left[X^2\right]$, obtaining

$$\begin{aligned}\mathrm{E}\left[X^2\right] &= \sum_{k=0}^{\infty} k^2\cdot P(X=k) = \sum_{k=0}^{\infty} k^2\cdot e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k^2\cdot e^{-\lambda}\frac{\lambda^k}{k!} \\ &= \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{k\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\left[\sum_{k=1}^{\infty}\frac{(k-1)\lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\right] \\ &= \lambda\left[e^{-\lambda}\sum_{k=1}^{\infty}\frac{(k-1)\lambda^{k-1}}{(k-1)!} + e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}\right] \\ &= \lambda\left[E[X] + e^{-\lambda}e^{\lambda}\right] = \lambda[\lambda+1] = \lambda^2+\lambda.\end{aligned}$$

Hence the variance of X is given by

$$\mathrm{Var}(X) = \mathrm{E}\left[X^2\right] - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \tag{A.5}$$

This shows that the variance of a Poisson random variable equals its mean.

A.1.1.1 Additivity of Poisson random variables

Let X_1 be a Poisson random variable with parameter λ_1, and let X_2 be a Poisson random variable with parameter λ_2. Let X_1 and X_2 be stochastically independent. Then the sum

$$X := X_1 + X_2 \tag{A.6}$$

is a Poisson random variable with parameter $\lambda_1 + \lambda_2$. The difference $X_1 - X_2$, however, is in general not a Poisson random variable (Exercise A.2).

A.1.2 Gaussian random variables

In a pixelated detector, the signal detected from a light source in each pixel is Poisson distributed. Therefore, the detected signal X_s is modeled by a Poisson random variable. The detector's readout process, however, produces a noise signal X_a in the readout electronics that is typically modeled as a Gaussian random variable. Therefore, the final signal X_m that is measured in a pixel is given by

$$X_m = X_s + X_a. \tag{A.7}$$

A random variable X with real values is called a *Gaussian random variable* if its probability density function is given by

$$p_{\eta,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\eta)^2}, \quad x\in\mathbb{R}, \tag{A.8}$$

where $\eta \in \mathbb{R}$ and $\sigma > 0$. This random variable has mean η and variance σ^2 (Exercise A.10).

A.2 Expectation of a Poisson random variable given a sum of Poisson random variables

The following result is central to our derivation of the Shepp-Vardi algorithm in Section 21.5. Let $X_1, \ldots, X_L$ be independent Poisson-distributed random variables with means $\lambda_1, \ldots, \lambda_L$. Let $Y = \sum_{j=1}^{L} X_j$. Then for $k = 1, \ldots, L$ and m a nonnegative integer, the conditional expectation $\mathrm{E}[X_k \mid Y = m]$ is given by

$$\mathrm{E}[X_k \mid Y = m] = \frac{m\lambda_k}{\lambda_1 + \cdots + \lambda_L}.$$

In order to show this, we first compute, using $\sum_{n_1+\cdots+n_L=m}$ to denote the summation over all combinations of nonnegative integers $n_1, \ldots, n_L$ that add up to m,

$$\begin{aligned} P(Y = m) &= P(X_1 + \cdots + X_L = m) \\ &= \sum_{n_1+\cdots+n_L=m} P(X_1 = n_1, \ldots, X_L = n_L) = \sum_{n_1+\cdots+n_L=m} \prod_{j=1}^{L} e^{-\lambda_j} \frac{\lambda_j^{n_j}}{n_j!} \\ &= e^{-(\lambda_1+\cdots+\lambda_L)} \frac{1}{m!} \sum_{n_1+\cdots+n_L=m} \frac{m!}{\prod_{j=1}^{L} n_j!} \prod_{j=1}^{L} \lambda_j^{n_j} \\ &= \frac{e^{-(\lambda_1+\cdots+\lambda_L)}}{m!} \left(\sum_{j=1}^{L} \lambda_j \right)^m, \end{aligned}$$

where we have used the multinomial theorem

$$(x_1 + x_2 + \cdots + x_D)^q = \sum_{k_1+\cdots+k_D=q} \frac{q!}{k_1!k_2!\cdots k_D!} \prod_{j=1}^{D} x_j^{k_j}$$

for D a positive integer, q a nonnegative integer, and $k_1, \ldots, k_D$ all nonnegative integers.

For $Y = \sum_{j=1}^{L} X_j$, we next compute, for $k = 1, \ldots, L$,

$$P(X_k = n_k \mid Y = m) = \frac{P(X_k = n_k, Y = m)}{P(Y = m)}.$$

Using the above result, we evaluate the numerator as

$$\begin{aligned} P(X_k = n_k, Y = m) &= P(X_k = n_k, X_1 + \cdots + X_L = m) \\ &= P(X_k = n_k, X_1 + \cdots + X_{k-1} + X_{k+1} + \cdots + X_L = m - n_k) \\ &= P(X_k = n_k)\, P(X_1 + \cdots + X_{k-1} + X_{k+1} + \cdots + X_L = m - n_k) \\ &= e^{-\lambda_k} \frac{\lambda_k^{n_k}}{n_k!} \frac{e^{-(\lambda_1+\cdots+\lambda_{k-1}+\lambda_{k+1}+\cdots+\lambda_L)}}{(m-n_k)!} \left(\sum_{j=1, j\neq k}^{L} \lambda_j \right)^{m-n_k}. \end{aligned}$$

Therefore, we have

$$
\begin{aligned}
P\left(X_k = n_k \mid Y = m\right) &= \frac{P\left(X_k = n_k, Y = m\right)}{P\left(Y = m\right)} \\
&= \frac{e^{-\lambda_k} \frac{\lambda_k^{n_k}}{n_k!} \frac{e^{-(\lambda_1 + \cdots + \lambda_{k-1} + \lambda_{k+1} \cdots \lambda_L)}}{(m-n_k)!} \left(\sum_{j=1, j\neq k}^{L} \lambda_j\right)^{m-n_k}}{e^{-(\lambda_1 + \cdots + \lambda_L)} \frac{(\lambda_1 + \cdots + \lambda_L)^m}{m!}} \\
&= \frac{\frac{\lambda_k^{n_k}}{n_k!} \frac{m!}{(m-n_k)!} \left(\sum_{j=1, j\neq k}^{L} \lambda_j\right)^{m-n_k}}{(\lambda_1 + \cdots + \lambda_L)^m} \\
&= \binom{m}{n_k} \lambda_k^{n_k} \frac{\left(\sum_{j=1, j\neq k}^{L} \lambda_j\right)^{m-n_k}}{\left(\lambda_1 + \cdots + \lambda_L\right)^m}.
\end{aligned}
$$

The conditional expectation is then given by

$$
\begin{aligned}
\mathrm{E}\left[X_k \mid Y = m\right] &= \sum_{n_k=0}^{\infty} n_k P\left(X_k = n_k \mid Y = m\right) = \sum_{n_k=0}^{m} n_k P\left(X_k = n_k \mid Y = m\right) \\
&= \sum_{n_k=1}^{m} n_k \binom{m}{n_k} \lambda_k^{n_k} \frac{\left(\sum_{j=1, j\neq k}^{L} \lambda_j\right)^{m-n_k}}{\left(\lambda_1 + \cdots + \lambda_L\right)^m} \\
&= \frac{1}{\left(\lambda_1 + \cdots + \lambda_L\right)^m} \sum_{n_k=1}^{m} n_k \binom{m}{n_k} \lambda_k^{n_k} \left(\sum_{j=1, j\neq k}^{L} \lambda_j\right)^{m-n_k} \\
&= \frac{1}{\left(\lambda_1 + \cdots + \lambda_L\right)^m} \sum_{n_k=1}^{m} n_k \frac{m!}{n_k!(m-n_k)!} \lambda_k^{n_k} \left(\sum_{j=1, j\neq k}^{L} \lambda_j\right)^{m-n_k} \\
&= \frac{m\lambda_k}{\left(\lambda_1 + \cdots + \lambda_L\right)^m} \\
&\quad \times \sum_{n_k=1}^{m} \frac{(m-1)!}{(n_k-1)!(m-1-(n_k-1))!} \lambda_k^{n_k-1} \left(\sum_{j=1, j\neq k}^{L} \lambda_j\right)^{m-1-(n_k-1)} \\
&= \frac{m\lambda_k}{\left(\lambda_1 + \cdots + \lambda_L\right)^m} \sum_{n_k=0}^{m-1} \frac{(m-1)!}{n_k!(m-1-n_k)!} \lambda_k^{n_k} \left(\sum_{j=1, j\neq k}^{L} \lambda_j\right)^{m-1-n_k} \\
&= \frac{m\lambda_k}{\left(\lambda_1 + \cdots + \lambda_L\right)^m} \left(\lambda_1 + \cdots + \lambda_L\right)^{m-1} \\
&= \frac{m\lambda_k}{\lambda_1 + \cdots + \lambda_L},
\end{aligned}
$$

where we have used the binomial theorem $(x+y)^{q-1} = \sum_{k=0}^{q-1} \binom{q-1}{k} x^{q-1-k} y^k$ for q a positive integer.

A.3 Additivity of Fisher information matrices

An important and useful property of the Fisher information matrix is that when the underlying data is composed of stochastically independent subsets of data, the overall Fisher information matrix for the entire data set can be obtained by summing the Fisher information matrices corresponding to the component data subsets. For example, if $\mathbf{I}_1(\theta)$ is the Fisher information matrix corresponding to one image, and $\mathbf{I}_2(\theta)$ is the Fisher information matrix corresponding to another image that is stochastically independent from the first image, then the overall Fisher information matrix $\mathbf{I}(\theta)$ corresponding to the pair of images, taken as one data set, is given by $\mathbf{I}(\theta) = \mathbf{I}_1(\theta) + \mathbf{I}_2(\theta)$.

This property of additivity can be demonstrated generally by considering data consisting of K independent measurements $z_1, z_2, \ldots, z_K$, from which we wish to estimate the parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_N)$. By the independence of the measurements, the log-likelihood function $\mathcal{L}\left(\theta \mid z_1, \ldots, z_K\right)$ needed to calculate the Fisher information matrix $\mathbf{I}(\theta)$ is given, as shown in Eq. (16.3), by the sum of the logarithms of the K probability distributions $p_{\theta,1}(z_1), \ldots, p_{\theta,K}(z_K)$ corresponding to the K measurements. Using the Fisher information matrix definition of Eq. (17.8), we obtain

$$
\begin{aligned}
\mathbf{I}(\theta) &= \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\mathcal{L}\left(\theta \mid z_1, \ldots, z_K\right)\right)^T \left(\frac{\partial}{\partial\theta}\mathcal{L}\left(\theta \mid z_1, \ldots, z_K\right)\right)\right] \\
&= \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\sum_{i=1}^{K}\ln p_{\theta,i}(z_i)\right)^T \left(\frac{\partial}{\partial\theta}\sum_{i=1}^{K}\ln p_{\theta,i}(z_i)\right)\right] \\
&= \mathrm{E}\left[\left(\sum_{i=1}^{K}\frac{\partial}{\partial\theta}\ln p_{\theta,i}(z_i)\right)^T \left(\sum_{i=1}^{K}\frac{\partial}{\partial\theta}\ln p_{\theta,i}(z_i)\right)\right] \\
&= \mathrm{E}\left[\left(\sum_{i=1}^{K}\frac{\partial}{\partial\theta}\mathcal{L}\left(\theta \mid z_i\right)\right)^T \left(\sum_{i=1}^{K}\frac{\partial}{\partial\theta}\mathcal{L}\left(\theta \mid z_i\right)\right)\right] \\
&= \mathrm{E}\begin{bmatrix}
\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1} & \sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2} & \cdots & \sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_K} \\
\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1} & \sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2} & \cdots & \sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_K} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_K}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1} & \sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_K}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2} & \cdots & \sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_K}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_K}
\end{bmatrix}
\end{aligned}
$$

$$
= \begin{bmatrix}
\mathrm{E}\left[\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\right] & \mathrm{E}\left[\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\right] & \cdots & \mathrm{E}\left[\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\right] \\
\mathrm{E}\left[\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\right] & \mathrm{E}\left[\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\right] & \cdots & \mathrm{E}\left[\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\right] \\
\vdots & \vdots & \ddots & \vdots \\
\mathrm{E}\left[\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\right] & \mathrm{E}\left[\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\right] & \cdots & \mathrm{E}\left[\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\sum_{i=1}^{K}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\right]
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\sum_{i=1}^{K}\mathrm{E}\left[\left(\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\right)^2\right] & \sum_{i=1}^{K}\mathrm{E}\left[\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\right] & \cdots & \sum_{i=1}^{K}\mathrm{E}\left[\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\right] \\
\sum_{i=1}^{K}\mathrm{E}\left[\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\right] & \sum_{i=1}^{K}\mathrm{E}\left[\left(\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\right)^2\right] & \cdots & \sum_{i=1}^{K}\mathrm{E}\left[\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\right] \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{K}\mathrm{E}\left[\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\right] & \sum_{i=1}^{K}\mathrm{E}\left[\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\right] & \cdots & \sum_{i=1}^{K}\mathrm{E}\left[\left(\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\right)^2\right]
\end{bmatrix}
$$

$$
= \sum_{i=1}^{K}\left(\mathrm{E}\begin{bmatrix}
\left(\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\right)^2 & \frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2} & \cdots & \frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N} \\
\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1} & \left(\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\right)^2 & \cdots & \frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_1} & \frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_2} & \cdots & \left(\frac{\partial\mathcal{L}(\theta|z_i)}{\partial\theta_N}\right)^2
\end{bmatrix}\right)
$$

$$
= \sum_{i=1}^{K}\mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\mathcal{L}\left(\theta \mid z_i\right)\right)^T\left(\frac{\partial}{\partial\theta}\mathcal{L}\left(\theta \mid z_i\right)\right)\right] = \sum_{i=1}^{K}\mathbf{I}_i(\theta), \tag{A.9}
$$

showing that the overall Fisher information matrix $\mathbf{I}(\theta)$ is the sum of the Fisher information matrices $\mathbf{I}_i(\theta)$, $i = 1, \ldots, K$, corresponding to the K independent measurements $z_1, \ldots, z_K$. In going from the expectation of the product of sums to the sum of the expectations of products, only products of partial derivatives involving the same measurement z_i, $i = 1, \ldots, K$, are kept. This is because by the independence of the measurements, we have, for $i, j = 1, \ldots, K$, $i \neq j$, and $n, m = 1, \ldots, N$,

$$
\mathrm{E}\left[\frac{\partial\mathcal{L}\left(\theta \mid z_i\right)}{\partial\theta_m}\frac{\partial\mathcal{L}\left(\theta \mid z_j\right)}{\partial\theta_n}\right] = \mathrm{E}\left[\frac{\partial\mathcal{L}\left(\theta \mid z_i\right)}{\partial\theta_m}\right]\cdot\mathrm{E}\left[\frac{\partial\mathcal{L}\left(\theta \mid z_j\right)}{\partial\theta_n}\right] = 0\cdot 0 = 0.
$$

The additivity property of the Fisher information matrix is easily seen in the results of the examples of Sections 17.1.3 and 17.1.4. Since in both examples the independent measurements consist of repeat measurements of a random variable, the overall Fisher information matrix $\mathbf{I}(\theta)$ is the sum of N_{im} identical Fisher information matrices corresponding to the N_{im} independent measurements.

B

Analysis

B.1 Delta function

A *delta function* δ_{x_0} is defined for a point $x_0 \in \mathbb{R}$ by the property that it has in regards to integration. Specifically, for a piecewise continuous function f, the integral of the product of f with the delta function produces the value of the function f at the point x_0, i.e.,

$$\int_{-\infty}^{\infty} f(x)\delta_{x_0}(x)dx = f(x_0).$$

The delta function is not a function in a strict mathematical sense and a rigorous definition is more complex (Note B.1).

B.2 Taylor series approximation

If a function $f(x)$, $x \in \mathbb{R}$, is infinitely often differentiable at the point $x_0 \in \mathbb{R}$, then the *Taylor series* of f around the point x_0 is given by

$$f(x_0)+\frac{f'(x_0)}{1!}(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\frac{f'''(x_0)}{3!}(x-x_0)^3+\cdots.$$

Given that certain conditions are satisfied, the Taylor series converges to the function f in a neighborhood around the expansion point x_0.

An example that is important for us is the Taylor series expansion of the function $f(x) = \sqrt{1+x}$, $x \geq -1$, around the point $x_0 = 0$, which is given by

$$1+\frac{1}{2}x-\frac{1}{8}x^2+\frac{1}{16}x^3-\frac{5}{128}x^4+\cdots.$$

As a result, for x small, $\sqrt{1+x}$ can be approximated well by the first two terms, i.e.,

$$\sqrt{1+x} \approx 1+\frac{x}{2}.$$

B.3 Change of variables theorem

Let $g = (g_1, g_2, \ldots, g_n)$: $B \subseteq \mathbb{R}^n \mapsto \mathbb{R}^n$ be an injective and continuously differentiable function. Let f: $\mathbb{R}^n \mapsto \mathbb{R}$ be an integrable function and $A \subseteq \mathbb{R}^n$, then the *change of* *variables theorem* is given by

$$\int_{g(A)} f(y_1, y_2, \ldots, y_n) dy_1 dy_2 \cdots dy_n$$
$$= \int_A f(g(x_1, x_2, \ldots, x_n)) \left|\det\left(J(g)\left(x_1, x_2, \ldots, x_n\right)\right)\right| dx_1 dx_2 \cdots dx_n,$$

where $J(g)$ is the *Jacobian matrix*

$$J(g) := \begin{pmatrix} \frac{\partial g_1(x_1,x_2,\ldots,x_n)}{\partial x_1} & \frac{\partial g_1(x_1,x_2,\ldots,x_n)}{\partial x_2} & \cdots & \frac{\partial g_1(x_1,x_2,\ldots,x_n)}{\partial x_n} \\ \frac{\partial g_2(x_1,x_2,\ldots,x_n)}{\partial x_1} & \frac{\partial g_2(x_1,x_2,\ldots,x_n)}{\partial x_2} & \cdots & \frac{\partial g_2(x_1,x_2,\ldots,x_n)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n(x_1,x_2,\ldots,x_n)}{\partial x_1} & \frac{\partial g_n(x_1,x_2,\ldots,x_n)}{\partial x_2} & \cdots & \frac{\partial g_n(x_1,x_2,\ldots,x_n)}{\partial x_n} \end{pmatrix}.$$

B.3.1 Change of Cartesian coordinates to polar coordinates

Of particular interest for us is the application of the change of variables theorem to the translation of an integral when the coordinate system is changed from the Cartesian coordinate system to the polar coordinate system. We would like to integrate an integrable function $f\colon \mathbb{R}^2 \to \mathbb{R}$.

If $g\colon [0, \infty[\, \times\, [0, 2\pi[\, \mapsto \mathbb{R}^2$ is the coordinate transformation from the polar coordinates (r, ϕ) to the Cartesian coordinates (x, y), i.e.,

$$g(r, \phi) = (g_1(r, \phi), g_2(r, \phi)) = (r\cos(\phi), r\sin(\phi)) =: (x, y),$$

then we have for the Jacobian

$$J(g) = \begin{pmatrix} \frac{\partial g_1(r,\phi)}{\partial r} & \frac{\partial g_1(r,\phi)}{\partial \phi} \\ \frac{\partial g_2(r,\phi)}{\partial r} & \frac{\partial g_2(r,\phi)}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \frac{\partial r\cos(\phi)}{\partial r} & \frac{\partial r\cos(\phi)}{\partial \phi} \\ \frac{\partial r\sin(\phi)}{\partial r} & \frac{\partial r\sin(\phi)}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \cos(\phi) & -r\sin(\phi) \\ \sin(\phi) & r\cos(\phi) \end{pmatrix},$$

and the modulus of its determinant is given by

$$\left|\det\begin{pmatrix} \cos(\phi) & -r\sin(\phi) \\ \sin(\phi) & r\cos(\phi) \end{pmatrix}\right| = |r| = r.$$

Therefore, by the change of variables theorem,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y) dx dy = \int_0^{\infty}\int_0^{2\pi} r f\left(g(r, \phi)\right) d\phi dr. \tag{B.1}$$

C

Fourier Transform

C.1 Fourier transform

The Fourier transformation is an important mathematical technique that plays a major role in optics, and in particular in the analysis of wave propagation and diffraction theory. Let $f : \mathbb{R} \to \mathbb{C}$ be a function such that $\int_{-\infty}^{\infty} |f(x)|dx < \infty$. Then the *Fourier transform* $\mathcal{F}(f)$ of f is defined as

$$(\mathcal{F}(f))(\xi) := \int_{-\infty}^{\infty} f(x)e^{-2\pi i\xi x}dx, \quad \xi \in \mathbb{R}.$$

Similarly, for a function $g : \mathbb{R} \to \mathbb{C}$ such that $\int_{-\infty}^{\infty} |g(\xi)|d\xi < \infty$, we can define the *inverse Fourier transform*

$$\left(\mathcal{F}^{-1}(g)\right)(x) := \int_{-\infty}^{\infty} g(\xi)e^{2\pi i\xi x}d\xi, \quad x \in \mathbb{R}.$$

The *convolution theorem* states one of the most important properties of the Fourier transform, i.e., that the Fourier transform of the convolution of two functions is the pointwise product of the Fourier transforms of the functions. We have, according to this theorem,

$$(\mathcal{F}(f \star g))(\xi) = (\mathcal{F}(f))(\xi) \cdot (\mathcal{F}(g))(\xi), \quad \xi \in \mathbb{R},$$

where the convolution $f \star g$ is defined by

$$(f \star g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)dy, \quad x \in \mathbb{R}.$$

We will also frequently use multidimensional generalizations of the Fourier transform. For an n-dimensional function $f : \mathbb{R}^n \to \mathbb{C}$ such that $\int_{\mathbb{R}^n} |f(x)|dx_1dx_2\cdots dx_n < \infty$, with $x = (x_1, \ldots, x_n)$, the *n-dimensional Fourier transform* $\mathcal{F}(f)$ of f is defined by

$$(\mathcal{F}(f))(\xi) = \int_{\mathbb{R}^n} f(x)e^{-2\pi i\langle \xi, x\rangle}dx_1dx_2\cdots dx_n, \quad \xi \in \mathbb{R}^n.$$

For an n-dimensional function $g : \mathbb{R}^n \to \mathbb{C}$ such that $\int_{\mathbb{R}^n} |g(\xi)|d\xi_1d\xi_2\cdots d\xi_n < \infty$, with $\xi = (\xi_1, \ldots, \xi_n)$, the *$n$-dimensional inverse Fourier transform* is defined by

$$\left(\mathcal{F}^{-1}(g)\right)(x) = \int_{\mathbb{R}^n} g(\xi)e^{2\pi i\langle \xi, x\rangle}d\xi_1d\xi_2\cdots d\xi_n, \quad x \in \mathbb{R}^n.$$

The term inverse Fourier transform is justified since the inverse Fourier transform inverts the Fourier transform operation, i.e., for integrable functions f and g, we have

$$\mathcal{F}^{-1}(\mathcal{F}(f)) = f \quad \text{and} \quad \mathcal{F}(\mathcal{F}^{-1}(g)) = g.$$

C.1.1 Fourier transform of circularly symmetric functions

We now specialize the prior definitions to investigate the case of the Fourier transform applied to functions of two variables, i.e., we assume that

$$f : \mathbb{R}^2 \to \mathbb{C}.$$

In Cartesian coordinates the Fourier transform is then given by

$$\left(\mathcal{F}(f)\right)(y_1, y_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2) e^{-2\pi i (x_1 y_1 + x_2 y_2)} dx_1 dx_2.$$

We now also assume that f is *circularly symmetric*, i.e., for some single variable function f^s, we have that for all $(x_1, x_2) \in \mathbb{R}^2$,

$$f(x_1, x_2) = f^s\left(\sqrt{x_1^2 + x_2^2}\right) =: f^s(r), \quad r = \sqrt{x_1^2 + x_2^2}.$$

We will now calculate the 2D Fourier transform under this assumption. Using polar coordinates for both the pairs (x_1, x_2) and (y_1, y_2), i.e., setting

$$\begin{aligned} (x_1, x_2) &=: (r\cos(\phi), r\sin(\phi)), \quad r \geq 0, \quad \phi \in [0, 2\pi), \\ (y_1, y_2) &=: (\rho\cos(\psi), \rho\sin(\psi)), \quad \rho \geq 0, \quad \psi \in [0, 2\pi), \end{aligned}$$

we have, using the change of variables theorem (Eq. (B.1)), that

$$\begin{aligned} \left(\mathcal{F}(f)\right)(y_1, y_2) &= \int_0^{\infty}\int_0^{2\pi} r f(r\cos(\phi), r\sin(\phi)) e^{-2\pi i (r\cos(\phi) y_1 + r\sin(\phi) y_2)} d\phi dr \\ &= \int_0^{\infty}\int_0^{2\pi} r f^s(r) e^{-2\pi i (r\rho\cos(\phi)\cos(\psi) + r\rho\sin(\phi)\sin(\psi))} d\phi dr \\ &= \int_0^{\infty}\int_0^{2\pi} r f^s(r) e^{-2\pi i r\rho\cos(\phi - \psi)} d\phi dr = \int_0^{\infty} r f^s(r) \int_0^{2\pi} e^{-2\pi i r\rho\cos(\phi)} d\phi dr \\ &= \int_0^{\infty} r f^s(r) \int_0^{2\pi} e^{2\pi i r\rho\cos(\phi)} d\phi dr = 2\pi \int_0^{\infty} r f^s(r) J_0(2\pi r\rho) dr, \end{aligned}$$

where $\rho = \sqrt{y_1^2 + y_2^2}$. In the last step, we made use of the zeroth order Bessel function identity

$$J_0(x) := \frac{1}{2\pi}\int_0^{2\pi} e^{ix\cos(\phi)} d\phi, \quad x \in \mathbb{R}.$$

This result implies that the 2D Fourier transform of a circularly symmetric function is itself circularly symmetric.

C.2 Discrete Fourier transform

Having reviewed some of the basic results for the Fourier transform, we now give a similar review of the basic results for the discrete Fourier transform. We will use the discrete Fourier transform primarily for the analysis of data from pixelated detectors.

Given a sequence $x = (x_n)_{0 \le n \le N-1}$, $x_n \in \mathbb{C}$, $n = 0, 1, \ldots, N-1$, we define the *discrete Fourier transform* (DFT) of the sequence by

$$(DFT\,(x))_k := \sum_{n=0}^{N-1} x_n e^{-2\pi i k n/N}, \quad k = 0, 1, \ldots, N-1.$$

The *inverse discrete Fourier transform* ($IDFT$) for a sequence $X = (X_k)_{0 \le k \le N-1}$, $X_k \in \mathbb{C}$, $k = 0, 1, \ldots, N-1$, is defined by

$$(IDFT\,(X))_n := \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{2\pi i k n/N}, \quad n = 0, 1, \ldots, N-1.$$

The inverse discrete Fourier transform is indeed the inverse of the discrete Fourier transform, and we have that

$$IDFT((DFT)(x)) = x \quad \text{and} \quad DFT(IDFT(X)) = X.$$

Two important properties are *Plancherel's theorem* and the convolution property, with Plancherel's theorem stating for the vectors above that

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X_k|^2,$$

or equivalently that the square of the l^2 norm of the vector x equals the square of the l^2 norm of the vector associated with the discrete Fourier transform up to the constant $\frac{1}{N}$.

The definition of the convolution of two finite sequences creates technical problems due to the finiteness of the vectors. To deal with these issues, we define the *periodic extension* w^e of a finite vector w of length N by

$$w_n^e := w_{(n \bmod N)}, \quad n \in \mathbb{Z},$$

where (n mod N) is the integer a, $0 \le a \le N-1$, such that for some integer b, we have that $n = a + bN$. With this notation we can now define the convolution $x \star y$ of two finite sequences $x = (x_n)_{0 \le n \le N-1}$ and $y = (y_n)_{0 \le n \le N-1}$, $x_n, y_n \in \mathbb{C}$, $n = 0, 1, \ldots, N-1$, by

$$(x \star y)_n := \sum_{l=0}^{N-1} x_l y_{n-l}^e = \sum_{l=0}^{N-1} x_{n-l}^e y_l, \quad n = 0, 1, \ldots, N-1.$$

Analogous to the continuous case, we have that the discrete Fourier transform of the convolution of two sequences can be obtained as the point-wise multiplication of the discrete Fourier transforms of the individual sequences, i.e.,

$$DFT(x \star y) = DFT(x) \cdot DFT(y).$$

To write the identity more explicitly, if $(X_k)_{0 \le k \le N-1} := DFT(x)$, $(Y_k)_{0 \le k \le N-1} := DFT(y)$, and $(C_k)_{0 \le k \le N-1} := DFT(x * y)$, then

$$C_k = X_k \cdot Y_k, \quad k = 0, 1, \ldots, N-1.$$

C.3 Multidimensional discrete Fourier transform

We now generalize the above definitions and identities for the discrete Fourier transform to multidimensional arrays. If $x = (x_{n_1,n_2,\dots,n_K})_{0\le n_1\le N_1-1, 0\le n_2\le N_2-1,\dots,0\le n_K\le N_K-1}$ is a K-dimensional finite array of real or complex numbers, we can define the K-dimensional discrete Fourier transform analogously to the 1D case as $X = (X_{k_1,k_2,\dots,k_K})_{0\le k_1\le N_1-1,0\le k_2\le N_2-1,\dots,0\le k_K\le N_K-1} := DFT(x)$, where for $k_l = 0, 1, \dots, N_l - 1$, $l = 1, 2, \dots, K$,

$$X_{k_1,k_2,\dots,k_K} := \sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}\cdots\sum_{n_K=0}^{N_K-1} x_{n_1,n_2,\dots,n_K} e^{-2\pi i k_1 n_1/N_1} e^{-2\pi i k_2 n_2/N_2}\cdots e^{-2\pi i k_K n_K/N_K}.$$

Similarly, the inverse multidimensional discrete Fourier transform is defined as $x := IDFT(X)$, where for $n_l = 0, 1, \dots, N_l - 1$, $l = 1, 2, \dots, K$,

$$\begin{aligned} x_{n_1,n_2,\dots,n_K} := &\frac{1}{N_1 N_2\cdots N_K} \\ &\times \sum_{k_1=0}^{N_1-1}\sum_{k_2=0}^{N_2-1}\cdots\sum_{k_K=0}^{N_K-1} X_{k_1,k_2,\dots,k_K} e^{2\pi i n_1 k_1/N_1} e^{2\pi i n_2 k_2/N_2}\cdots e^{2\pi i n_K k_K/N_K}. \end{aligned}$$

Again we have that the inverse multidimensional discrete Fourier transform inverts the multidimensional discrete Fourier transform, and therefore

$$IDFT((DFT)(x)) = x \quad \text{and} \quad DFT(IDFT(X)) = X.$$

In order to define the convolution between two multidimensional arrays

$$\begin{aligned} x &= (x_{n_1,n_2,\dots,n_K})_{0\le n_1\le N_1-1,0\le n_2\le N_2-1,\dots,0\le n_K\le N_K-1} \quad \text{and} \\ y &= (y_{n_1,n_2,\dots,n_K})_{0\le n_1\le N_1-1,0\le n_2\le N_2-1,\dots,0\le n_K\le N_K-1} \end{aligned}$$

of equal dimensions, we need to define the periodic extension of a multidimensional array. We do this analogously to the 1D case by defining the extension for each of the dimensions of a multidimensional array w, obtaining

$$w^e_{n_1,n_2,\dots,n_K} := w_{(n_1 \bmod N_1),(n_2 \bmod N_2),\dots,(n_K \bmod N_K)}, \quad n_1, n_2, \dots, n_K \in \mathbb{Z}.$$

We can now define the convolution $x \star y$ as

$$\begin{aligned} (x\star y)_{n_1,n_2,\dots,n_K} &:= \sum_{l_1=0}^{N_1-1}\sum_{l_2=0}^{N_2-1}\cdots\sum_{l_K=0}^{N_K-1} x_{l_1,l_2,\dots,l_K} y^e_{n_1-l_1,n_2-l_2,\dots,n_K-l_K} \\ &= \sum_{l_1=0}^{N_1-1}\sum_{l_2=0}^{N_2-1}\cdots\sum_{l_K=0}^{N_K-1} x^e_{n_1-l_1,n_2-l_2,\dots,n_K-l_K} y_{l_1,l_2,\dots,l_K}, \end{aligned}$$

$n_l = 0, 1, \dots, N_l - 1$, $l = 1, 2, \dots, K$. With this definition, we again obtain that the discrete Fourier transform of the convolution of two multidimensional arrays is the point-wise product of the discrete Fourier transforms of the arrays, i.e.,

$$DFT(x \star y) = DFT(x) \cdot DFT(y).$$

The relationship between a multidimensional array and its multidimensional discrete

Fourier transform is even deeper, as expressed by *Parseval's theorem* and Plancherel's theorem, which state that the "geometry" of the space of arrays is identical to that of the space of Fourier transforms. As a result, we can perform least squares optimization tasks and related computations interchangeably in either domain.

The multidimensional Parseval's theorem is given by

$$\sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \cdots \sum_{n_K=0}^{N_K-1} x_{n_1,n_2,\ldots,n_K} \overline{y_{n_1,n_2,\ldots,n_K}} = \frac{1}{N_1 N_2 \cdots N_K} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} \cdots \sum_{k_K=0}^{N_K-1} X_{k_1,k_2,\ldots,k_K} \overline{Y_{k_1,k_2,\ldots,k_K}}.$$

This theorem tells us that the inner product of the arrays x and y is, up to a constant that is only dependent on the sizes of the arrays, identical to the inner product of their transforms. If we set $y = x$, we immediately obtain the multidimensional Plancherel's theorem, which is given by

$$\sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \cdots \sum_{n_K=0}^{N_K-1} |x_{n_1,n_2,\ldots,n_K}|^2 = \frac{1}{N_1 N_2 \cdots N_K} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} \cdots \sum_{k_K=0}^{N_K-1} |X_{k_1,k_2,\ldots,k_K}|^2.$$

D

Least Squares Minimization

D.1 Least squares minimization problem

We discuss here the problem of least squares minimization. This problem is very easily posed and solved using the language of linear algebra. The solution is derived here for the general scenario where the entries of a linear transformation can be complex. We therefore use the conjugate transpose, denoted by the superscript *. In Section 21.3, the solution derived here is written instead with the transpose operator T as it is assumed there that the entries of a linear transformation are all real.

Let $A : U \to V$ be a linear transformation between the inner product spaces U and V with inner products $\langle \cdot, \cdot \rangle_U$ and $\langle \cdot, \cdot \rangle_V$, respectively. Assume that A^*A is invertible. As examples, we can think of $U = \mathbb{R}^n$ and $V = \mathbb{R}^m$ with A given by a matrix that we also denote by A for simplicity. The inner products for U and V are then simply given by the standard scalar products between vectors in $\mathbb{R}^n$ and between vectors in $\mathbb{R}^m$, respectively.

We want to consider, for a vector $b \in V$ and a linear transformation $C : U \to V$, the minimization problem

$$\inf_{x \in U} \left(\|Ax - b\|^2 + \|Cx\|^2 \right). \tag{D.1}$$

Before addressing this general regularized problem, we will derive a solution for the non-regularized problem

$$\inf_{x \in U} \|Ax - b\|^2,$$

where we have set $C = 0$. In fact, we are primarily interested in finding x_0 such that

$$\|Ax_0 - b\|^2 = \inf_{x \in U} \|Ax - b\|^2. \tag{D.2}$$

First, consider the orthogonal projection $P_{range(A)}$ of V onto the range $range(A) = \{Ax \mid x \in U\}$ of A. It is important to note that in fact $P_{range(A)} = A\left(A^*A\right)^{-1} A^*$. This can be seen as follows. For $y \in range(A)$ we have that $y = A\tilde{x}$ for some $\tilde{x} \in U$. Therefore, $A\left(A^*A\right)^{-1} A^*y = A\left(A^*A\right)^{-1} A^*A\tilde{x} = A\tilde{x} = y \in range(A)$. For $y_\perp \in (range(A))^\perp$, the orthogonal complement of the range of A, we have $\left\langle y, A\left(A^*A\right)^{-1} A^*y_\perp \right\rangle = \left\langle A\left(A^*A\right)^{-1} A^*y, y_\perp \right\rangle = 0$ for all $y \in V$ as $A\left(A^*A\right)^{-1} A^*y \in range(A)$ and $y_\perp \in (range(A))^\perp$. This implies that $A\left(A^*A\right)^{-1} A^*y_\perp = 0$ and hence that $P_{range(A)} = A\left(A^*A\right)^{-1} A^*$.

Now we use this projection to decompose the expression into terms that are in the range of A and its orthogonal complement. For $x \in U$ and denoting by $P_{(range(A))^\perp}$ the orthogonal projection of V onto the orthogonal complement of the range of A, we have, by Pythagoras' theorem,

$$\begin{aligned}\|Ax-b\|^2 &= \left\|P_{range(A)}(Ax-b)\right\|^2 + \left\|P_{(range(A))^\perp}(Ax-b)\right\|^2\\ &= \left\|Ax - P_{range(A)}b\right\|^2 + \left\|P_{(range(A))^\perp}b\right\|^2\\ &= \|Ax-b_A\|^2 + \|b_{A^\perp}\|^2,\end{aligned}$$

where $b_A = P_{range(A)}b$ and $b_{A^\perp} = P_{(range(A))^\perp}b$.

As $\|b_{A^\perp}\|^2$ does not depend on x, we have changed the optimization into the analogous optimization problem

$$\inf_{x\in U}\|Ax-b_A\|^2,$$

with the important property that b_A is in the range of A.

To solve this problem, we introduce a new inner product space $\langle x,y\rangle_Q := \langle x, A^*Ay\rangle$, $x,y\in U$, with norm $\|x\|_Q := \sqrt{\langle x,x\rangle_Q} = \sqrt{\langle x, A^*Ax\rangle}$. Using this inner product space, we obtain

$$\begin{aligned}\|Ax-b_A\|^2 &= \langle Ax,Ax\rangle - 2\langle x, A^*b_A\rangle + \langle b_A, b_A\rangle\\ &= \langle x, A^*Ax\rangle - 2\left\langle x, A^*A\left(A^*A\right)^{-1}A^*b\right\rangle + \left\langle A\left(A^*A\right)^{-1}A^*b, A\left(A^*A\right)^{-1}A^*b\right\rangle\\ &= \langle x,x\rangle_Q - 2\left\langle x, \left(A^*A\right)^{-1}A^*b\right\rangle_Q + \left\langle \left(A^*A\right)^{-1}A^*b, A^*A\left(A^*A\right)^{-1}A^*b\right\rangle\\ &= \langle x,x\rangle_Q - 2\left\langle x, \left(A^*A\right)^{-1}A^*b\right\rangle_Q + \left\langle \left(A^*A\right)^{-1}A^*b, \left(A^*A\right)^{-1}A^*b\right\rangle_Q\\ &= \left\|x - \left(A^*A\right)^{-1}A^*b\right\|_Q^2.\end{aligned}$$

Therefore, with $x_0 = \left(A^*A\right)^{-1}A^*b$, we have that $\left\|x_0 - \left(A^*A\right)^{-1}A^*b\right\|_Q^2 = \|Ax_0-b_A\|^2 = 0$. Hence x_0 is the optimizing solution and

$$\inf_{x\in U}\|Ax-b\|^2 = \inf_{x\in U}\|Ax-b_A\|^2 + \|b_{A^\perp}\|^2 = \|Ax_0-b_A\|^2 + \|b_{A^\perp}\|^2 = \|b_{A^\perp}\|^2.$$

Having obtained a solution to the non-regularized problem, we can now address the regularized problem by reducing it to a non-regularized one. We can rewrite the original regularized minimization problem as

$$\inf_{x\in U}\left(\|Ax-b\|^2 + \|Cx\|^2\right) = \inf_{x\in U}\left\|\begin{pmatrix}A\\C\end{pmatrix}x - \begin{pmatrix}b\\0\end{pmatrix}\right\|^2 = \inf_{x\in U}\|\tilde{A}x-\tilde{b}\|^2,$$

where

$$\tilde{A} := \begin{pmatrix}A\\C\end{pmatrix},\quad \tilde{b} := \begin{pmatrix}b\\0\end{pmatrix}.$$

The problem is thus written in the form for which we have derived a solution. Assuming that $\left(\tilde{A}\right)^*\tilde{A} = (A^*A + C^*C)$ is invertible, we have for the solution to the original regularized minimization problem (Eq. (D.1))

$$x_0 = \left(\left(\tilde{A}\right)^*\tilde{A}\right)^{-1}\left(\tilde{A}\right)^*\tilde{b} = (A^*A+C^*C)^{-1}A^*b. \tag{D.3}$$

D.2 Linear least squares in the Fourier domain

We now consider the deconvolution problem where we want to minimize the regularized least squares criterion

$$\hat{X}_{RLR} = \arg\min_{X} \left(\|I - psf \star X\|^2 + \|W \star X\|^2 \right), \tag{D.4}$$

where I, psf, W, and X are functions of compatible dimensions and $\star$ is the convolution operator. Applying Plancherel's theorem, we can translate this problem into the frequency domain as

$$\hat{X}_{RLR} = \mathcal{F}^{-1}\left[\arg\min_{\mathcal{F}(X)} \left(\|\mathcal{F}(I) - \mathcal{F}(psf) \cdot \mathcal{F}(X)\|^2 + \|\mathcal{F}(W) \cdot \mathcal{F}(X)\|^2 \right)\right].$$

We want to apply the solution of Eq. (D.3), and therefore have to write our minimization problem in terms of linear algebra. We first sketch how we can introduce notions of linear algebra into the discussion of Fourier transforms. The Fourier transform of an n-dimensional function is an n-dimensional function on $\mathbb{R}^n$ mapping into $\mathbb{C}$. Let F be the vector space of all such functions that are both integrable and square integrable, i.e., $f \in F$ if $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |f(\xi_1, \ldots, \xi_n)|^p \, d\xi_1 \cdots d\xi_n < \infty$ for $p = 1, 2$. It is straightforward to show that F is in fact a vector space, as the sum of two such functions and the product of such a function with a complex scalar are again in F. In fact, we can consider F to be an inner product space by setting, for $f_1, f_2 \in F$, $\langle f_1, f_2 \rangle := \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_1(\xi_1, \ldots, \xi_n) \overline{f_2(\xi_1, \ldots, \xi_n)} d\xi_1 \cdots d\xi_n$. Now assume we have another function g on $\mathbb{R}^n$ such that $\sup \{|g(\xi_1, \ldots, \xi_n)| : (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n\} < \infty$. With this function, we define the multiplication map

$$\begin{aligned} M_g : F &\to F \\ f &\mapsto gf, \end{aligned}$$

where gf is nothing but the point-wise multiplication

$$(gf)(\xi_1, \ldots, \xi_n) = g(\xi_1, \ldots, \xi_n) \, f(\xi_1, \ldots, \xi_n).$$

The adjoint map M_g^* of this map is easily determined as $\langle M_g f_1, f_2 \rangle = \langle g f_1, f_2 \rangle = \langle f_1, \overline{g} f_2 \rangle = \langle f_1, M_{\overline{g}} f_2 \rangle$, $f_1, f_2 \in F$. Hence $M_g^* = M_{\overline{g}}$.

Using the above, let us denote by $M_{\mathcal{F}(psf)}$ ($M_{\mathcal{F}(W)}$) the multiplication operator that multiplies point-wise any function of compatible dimension with $\mathcal{F}(psf)$ ($\mathcal{F}(W)$). With this we can write

$$\hat{X}_{RLR} = \mathcal{F}^{-1}\left[\arg\min_{\mathcal{F}(X)} \left(\left\|\mathcal{F}(I) - M_{\mathcal{F}(psf)} \cdot \mathcal{F}(X)\right\|^2 + \left\|M_{\mathcal{F}(W)} \cdot \mathcal{F}(X)\right\|^2 \right)\right].$$

Now we can apply Eq. (D.3) to obtain

$$\begin{aligned} \hat{X}_{RLR} &= \mathcal{F}^{-1}\left[\left(M^*_{\mathcal{F}(psf)} \cdot M_{\mathcal{F}(psf)} + M^*_{\mathcal{F}(W)} \cdot M_{\mathcal{F}(W)}\right)^{-1} M^*_{\mathcal{F}(psf)} \cdot \mathcal{F}(I)\right] \\ &= \mathcal{F}^{-1}\left[\left(\overline{\mathcal{F}(psf)} \cdot \mathcal{F}(psf) + \overline{\mathcal{F}(W)} \cdot \mathcal{F}(W)\right)^{-1} \overline{\mathcal{F}(psf)} \cdot \mathcal{F}(I)\right] \\ &= \mathcal{F}^{-1}\left[\left(|\mathcal{F}(psf)|^2 + |\mathcal{F}(W)|^2\right)^{-1} \overline{\mathcal{F}(psf)} \cdot \mathcal{F}(I)\right]. \end{aligned}$$

Here we have used that the adjoint map $M^*_{\mathcal{F}(psf)}$ of the multiplication operator $M_{\mathcal{F}(psf)}$ is given by the multiplication operator that multiplies with the complex conjugate $\overline{\mathcal{F}(psf)}$ of the function $\mathcal{F}(psf)$, i.e., by $M^*_{\mathcal{F}(psf)} = M_{\overline{\mathcal{F}(psf)}}$. Similarly, $M^*_{\mathcal{F}(W)} = M_{\overline{\mathcal{F}(W)}}$. The above expression for $\hat{X}_{RLR}$ shows how the solution to the regularized deconvolution problem can be written in a compact fashion through Fourier transforms.

E

Fisher Information for the Fundamental Data Model

We derive here the Fisher information matrix corresponding to the fundamental data model, i.e., Eq. (17.11). We follow closely a previous derivation (Note E.1), but use the zero-truncated Poisson distribution in place of the Poisson distribution to model the number of photons detected in an image.

We begin by repeating here Eq. (16.10), the log-likelihood function corresponding to an image of N_0 photons detected on the detector C during the acquisition time interval $[t_0, t]$. Letting $w_k := (r_k, \tau_k)$, $k = 1, \ldots, N_0$, where $r_k = (x_k, y_k) \in C$ denotes the spatial coordinates and $t_0 \leq \tau_k \leq t$ denotes the time point at which the kth photon is detected, we have the log-likelihood function

$$\mathcal{L}\left(\theta \mid w_1, \ldots, w_{N_0}\right) = \sum_{k=1}^{N_0} \ln f_{\theta,\tau_k}(r_k) + \sum_{k=1}^{N_0} \ln \Lambda_\theta(\tau_k) - \int_{t_0}^{t} \Lambda_\theta(\tau) d\tau - \ln\left(1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}\right),$$

where $\theta \in \Theta$ is the vector of parameters to be estimated from the image and Θ is the parameter space. In this expression, $f_{\theta,\tau_k}(r_k)$ denotes the density function of the spatial coordinates r_k of the kth detected photon, and $\Lambda_\theta(\tau)$, $\tau \geq t_0$, denotes the rate of photon detection. The latter is the rate function of the Poisson process $\{\mathcal{N}(\tau); \tau \geq t_0\}$ that models the arrival of photons to the detector. The number of photons detected during the acquisition time interval $[t_0, t]$ is therefore represented by the Poisson random variable $\mathcal{N}(t)$, and the mean number of photons detected during this interval is given by $\mathrm{E}\left[\mathcal{N}(t)\right] = \int_{t_0}^{t} \Lambda_\theta(\tau) d\tau$. However, N_0, the strictly positive number of photons that comprise the image, is a realization of $\mathcal{N}(t)$ given that $\mathcal{N}(t) > 0$. In what follows, the expected value of a function of the spatial coordinates and time points of the N_0 photons that comprise the image will therefore be evaluated under the condition $\mathcal{N}(t) > 0$.

For $\theta \in \Theta$ and $\tau \geq t_0$, define

$$\mathcal{J}_{\theta,\tau}(r) := \frac{1}{f_{\theta,\tau}(r)} \frac{\partial f_{\theta,\tau}(r)}{\partial \theta}, \quad r \in C, \qquad \mathcal{M}_\theta(\sigma) := \frac{1}{\Lambda_\theta(\sigma)} \frac{\partial \Lambda_\theta(\sigma)}{\partial \theta}, \quad \sigma \geq t_0.$$

Then the derivative of the log-likelihood function with respect to $\theta \in \Theta$ is given by (Note E.2)

$$\begin{aligned}\frac{\partial \mathcal{L}\left(\theta \mid w_1,\ldots,w_{N_0}\right)}{\partial \theta} &= \sum_{k=1}^{N_0} \frac{1}{f_{\theta,\tau_k}(r_k)} \frac{\partial f_{\theta,\tau_k}(r_k)}{\partial \theta} + \sum_{k=1}^{N_0} \frac{1}{\Lambda_\theta(\tau_k)} \frac{\partial \Lambda_\theta(\tau_k)}{\partial \theta} - \frac{\partial}{\partial \theta} \int_{t_0}^{t} \Lambda_\theta(\tau) d\tau \\ &\quad + \frac{e^{-\int_{t_0}^{t} \Lambda_\theta(\tau)d\tau}}{1-e^{-\int_{t_0}^{t} \Lambda_\theta(\tau)d\tau}} \left(-\frac{\partial}{\partial \theta} \int_{t_0}^{t} \Lambda_\theta(\tau) d\tau \right) \\ &= \sum_{k=1}^{N_0} \frac{1}{f_{\theta,\tau_k}(r_k)} \frac{\partial f_{\theta,\tau_k}(r_k)}{\partial \theta} + \sum_{k=1}^{N_0} \frac{1}{\Lambda_\theta(\tau_k)} \frac{\partial \Lambda_\theta(\tau_k)}{\partial \theta} \\ &\quad - \frac{1}{1-e^{-\int_{t_0}^{t} \Lambda_\theta(\tau)d\tau}} \cdot \frac{\partial}{\partial \theta} \int_{t_0}^{t} \Lambda_\theta(\tau) d\tau \\ &= \sum_{k=1}^{N_0} \mathcal{J}_{\theta,\tau_k}(r_k) + \sum_{k=1}^{N_0} \mathcal{M}_\theta(\tau_k) - \frac{1}{1-e^{-\int_{t_0}^{t} \Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau.\end{aligned}$$

Using this derivative in Eq. (17.8), we obtain, for $\theta \in \Theta$, the Fisher information matrix

$$\begin{aligned}\mathbf{I}(\theta) &= \mathrm{E}\left[\left(\frac{\partial \mathcal{L}\left(\theta \mid w_1,\ldots,w_{N_0}\right)}{\partial \theta} \right)^T \left(\frac{\partial \mathcal{L}\left(\theta \mid w_1,\ldots,w_{N_0}\right)}{\partial \theta} \right) \right] \\ &= \mathrm{E}\left[\left(\sum_{k=1}^{N_0} \mathcal{J}_{\theta,\tau_k}^T(r_k) + \sum_{k=1}^{N_0} \mathcal{M}_\theta^T(\tau_k) - \frac{1}{1-e^{-\int_{t_0}^{t} \Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T d\tau \right) \times \right. \\ &\qquad \left. \left(\sum_{l=1}^{N_0} \mathcal{J}_{\theta,\tau_l}(r_l) + \sum_{l=1}^{N_0} \mathcal{M}_\theta(\tau_l) - \frac{1}{1-e^{-\int_{t_0}^{t} \Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \right) \right] \\ &= \mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{J}_{\theta,\tau_k}^T(r_k) \sum_{l=1}^{N_0} \mathcal{J}_{\theta,\tau_l}(r_l) \right] + \mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{J}_{\theta,\tau_k}^T(r_k) \sum_{l=1}^{N_0} \mathcal{M}_\theta(\tau_l) \right] \\ &\quad - \mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{J}_{\theta,\tau_k}^T(r_k) \right] \cdot \frac{1}{1-e^{-\int_{t_0}^{t} \Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \\ &\quad + \mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{M}_\theta^T(\tau_k) \sum_{l=1}^{N_0} \mathcal{J}_{\theta,\tau_l}(r_l) \right] + \mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{M}_\theta^T(\tau_k) \sum_{l=1}^{N_0} \mathcal{M}_\theta(\tau_l) \right] \\ &\quad - \mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{M}_\theta^T(\tau_k) \right] \cdot \frac{1}{1-e^{-\int_{t_0}^{t} \Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \\ &\quad - \frac{1}{1-e^{-\int_{t_0}^{t} \Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T d\tau \left(\mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{J}_{\theta,\tau_k}(r_k) \right] \right. \\ &\quad \left. + \mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{M}_\theta(\tau_k) \right] - \frac{1}{1-e^{-\int_{t_0}^{t} \Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \right). \qquad \text{(E.1)}\end{aligned}$$

To evaluate further the terms of $\mathbf{I}(\theta)$, we first consider a general function $\mathcal{H}$ that depends on $\{w_1,\ldots,w_{N_0}\}$. By the assumptions that the temporal and spatial components of photon detection are independent of each other, and that the locations at which the N_0 photons are detected are mutually independent, we can write, for the expectation of $\mathcal{H}$,

$$\mathrm{E}\left[\mathcal{H}\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0}\right)\right]=\sum_{N_0=1}^{\infty}\int_{t_0}^{t}\cdots\int_{t_0}^{t}\int_C\cdots\int_C\mathcal{H}\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0}\right)\times$$
$$p_\theta\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0},\mathcal{N}(t)=N_0\mid\mathcal{N}(t)>0\right)dr_1\cdots dr_{N_0}d\tau_1\cdots d\tau_{N_0}$$
$$=\sum_{N_0=1}^{\infty}\int_{t_0}^{t}\cdots\int_{t_0}^{t}\int_C\cdots\int_C\mathcal{H}\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0}\right)\times$$
$$p_\theta\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0},\mathcal{N}(t)=N_0,\mathcal{N}(t)>0\right)/P\left(\mathcal{N}(t)>0\right)$$
$$dr_1\cdots dr_{N_0}d\tau_1\cdots d\tau_{N_0}$$
$$=\frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\int_{t_0}^{t}\cdots\int_{t_0}^{t}\int_C\cdots\int_C\mathcal{H}\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0}\right)\times$$
$$p_\theta\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0},\mathcal{N}(t)=N_0\right)dr_1\cdots dr_{N_0}d\tau_1\cdots d\tau_{N_0}$$
$$=\frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\left(\int_{t_0}^{t}\cdots\int_{t_0}^{t}\left[\int_C\cdots\int_C\mathcal{H}\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0}\right)\times\right.\right.$$
$$\left.p_\theta\left(r_1,\ldots,r_{N_0}\mid\tau_1,\ldots,\tau_{N_0},\mathcal{N}(t)=N_0\right)dr_1\cdots dr_{N_0}\right]\times$$
$$\left.p_\theta\left(\tau_1,\ldots,\tau_{N_0}\mid\mathcal{N}(t)=N_0\right)d\tau_1\cdots d\tau_{N_0}\right)P\left(\mathcal{N}(t)=N_0\right)$$
$$=\frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\left(\int_{t_0}^{t}\cdots\int_{t_0}^{t}\left[\int_C\cdots\int_C\mathcal{H}\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0}\right)\times\right.\right.$$
$$\left.f_{\theta,\tau_1}\left(r_1\right)\cdots f_{\theta,\tau_{N_0}}\left(r_{N_0}\right)dr_1\cdots dr_{N_0}\right]\times$$
$$\left.p_\theta\left(\tau_1,\ldots,\tau_{N_0}\mid\mathcal{N}(t)=N_0\right)d\tau_1\cdots d\tau_{N_0}\right)P\left(\mathcal{N}(t)=N_0\right),\qquad\text{(E.2)}$$

where we have used p_θ to denote a probability density function.

Now, consider the case where $\mathcal{H}$ is the sum $\mathcal{H}\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0}\right)=\sum_{k=1}^{N_0}\mathcal{U}\left(r_k,\tau_k\right)$, where $\mathcal{U}$ is a real-valued vector function defined on $C\times[t_0,\infty)$. For a given value of N_0, a random reordering of the acquired data $\left(r_1,\tau_1\right),\ldots,\left(r_{N_0},\tau_{N_0}\right)$ does not change the value of the sum $\mathcal{H}$. However, with the reordering the time points of photon detection can be treated as independent random variables identically distributed with the probability density s_θ given by (Note E.3)

$$s_\theta(\tau)=\frac{\Lambda_\theta(\tau)}{\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau},\quad \tau\in[t_0,t],\quad \theta\in\Theta.$$

This implies that in this case we can set

$$p_\theta\left(\tau_1,\ldots,\tau_{N_0}\mid\mathcal{N}(t)=N_0\right)=s_\theta\left(\tau_1\right)\cdots s_\theta\left(\tau_{N_0}\right)$$

in Eq. (E.2), and the expectation of the sum $\mathcal{H}$ is then given by

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{k=1}^{N_0}\mathcal{U}\left(r_k,\tau_k\right)\right]\\
&=\frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\left(\int_{t_0}^{t}\cdots\int_{t_0}^{t}\left[\int_C\cdots\int_C\left(\mathcal{U}\left(r_1,\tau_1\right)+\cdots+\mathcal{U}\left(r_{N_0},\tau_{N_0}\right)\right)\right.\right.\\
&\quad\left.\left.\times f_{\theta,\tau_1}\left(r_1\right)\cdots f_{\theta,\tau_{N_0}}\left(r_{N_0}\right)dr_1\cdots dr_{N_0}\right]s_\theta\left(\tau_1\right)\cdots s_\theta\left(\tau_{N_0}\right)d\tau_1\cdots d\tau_{N_0}\right)P\left(\mathcal{N}(t)=N_0\right)\\
&=\frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\left(\int_{t_0}^{t}\cdots\int_{t_0}^{t}\left[\int_C\mathcal{U}\left(r,\tau_1\right)f_{\theta,\tau_1}(r)dr\int_C f_{\theta,\tau_2}(r)dr\right.\right.\\
&\quad\left.\cdots\int_C f_{\theta,\tau_{N_0}}(r)dr+\cdots+\int_C f_{\theta,\tau_1}(r)dr\int_C f_{\theta,\tau_2}(r)dr\cdots\int_C\mathcal{U}\left(r,\tau_{N_0}\right)f_{\theta,\tau_{N_0}}(r)dr\right]\\
&\quad\left.\times s_\theta\left(\tau_1\right)\cdots s_\theta\left(\tau_{N_0}\right)d\tau_1\cdots d\tau_{N_0}\right)P\left(\mathcal{N}(t)=N_0\right)\\
&=\frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\left(\int_{t_0}^{t}\cdots\int_{t_0}^{t}\left[\sum_{k=1}^{N_0}\int_C\mathcal{U}\left(r,\tau_k\right)f_{\theta,\tau_k}(r)dr\right]\right.\\
&\quad\left.\times s_\theta\left(\tau_1\right)\cdots s_\theta\left(\tau_{N_0}\right)d\tau_1\cdots d\tau_{N_0}\right)P\left(\mathcal{N}(t)=N_0\right)\\
&=\frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\left(\int_{t_0}^{t}\left[\int_C\mathcal{U}\left(r,\tau_1\right)f_{\theta,\tau_1}(r)dr\right]s_\theta\left(\tau_1\right)d\tau_1\int_{t_0}^{t}s_\theta\left(\tau_2\right)d\tau_2\right.\\
&\quad\cdots\int_{t_0}^{t}s_\theta\left(\tau_{N_0}\right)d\tau_{N_0}+\cdots+\int_{t_0}^{t}s_\theta\left(\tau_1\right)d\tau_1\int_{t_0}^{t}s_\theta\left(\tau_2\right)d\tau_2\\
&\quad\left.\cdots\int_{t_0}^{t}\left[\int_C\mathcal{U}\left(r,\tau_{N_0}\right)f_{\theta,\tau_{N_0}}(r)dr\right]s_\theta\left(\tau_{N_0}\right)d\tau_{N_0}\right)P\left(\mathcal{N}(t)=N_0\right)\\
&=\frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\left(N_0\int_{t_0}^{t}\left[\int_C\mathcal{U}\left(r,\tau\right)f_{\theta,\tau}(r)dr\right]s_\theta\left(\tau\right)d\tau\right)P\left(\mathcal{N}(t)=N_0\right)\\
&=\frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}N_0P\left(\mathcal{N}(t)=N_0\right)\int_{t_0}^{t}\left[\int_C\mathcal{U}\left(r,\tau\right)f_{\theta,\tau}(r)dr\right]s_\theta\left(\tau\right)d\tau\\
&=\frac{\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\int_{t_0}^{t}\left[\int_C\mathcal{U}\left(r,\tau\right)f_{\theta,\tau}(r)dr\right]s_\theta\left(\tau\right)d\tau\\
&=\frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\int_{t_0}^{t}\Lambda_\theta(\tau)\left[\int_C\mathcal{U}\left(r,\tau\right)f_{\theta,\tau}(r)dr\right]d\tau,
\end{aligned}
\tag{E.3}
$$

where we have used that $\sum_{N_0=1}^{\infty}N_0P\left(\mathcal{N}(t)=N_0\right)$ is the mean $\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau$ of the Poisson random variable $\mathcal{N}(t)$ representing the number of photons detected during the interval $[t_0,t]$. Note also that by the definitions of the density functions f_{θ,τ_k}, $k=1,\ldots,N_0$, and the density function s_θ, their integrals over C and $[t_0,t]$, respectively, evaluate to 1. If $\mathcal{U}$ is only a function of τ, then Eq. (E.3) becomes

$$\mathrm{E}\left[\sum_{k=1}^{N_0}\mathcal{U}\left(\tau_k\right)\right] = \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\int_{t_0}^{t}\Lambda_\theta(\tau)\left[\int_C \mathcal{U}\left(\tau\right)f_{\theta,\tau}(r)dr\right]d\tau$$
$$= \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\int_{t_0}^{t}\Lambda_\theta(\tau)\mathcal{U}\left(\tau\right)\left[\int_C f_{\theta,\tau}(r)dr\right]d\tau$$
$$= \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\int_{t_0}^{t}\Lambda_\theta(\tau)\mathcal{U}\left(\tau\right)d\tau. \quad \text{(E.4)}$$

Next, consider the case where $\mathcal{H}$ is given by

$$\mathcal{H}\left(r_1,\ldots,r_{N_0},\tau_1,\ldots,\tau_{N_0}\right) = \sum_{k=1}^{N_0}\mathcal{U}\left(r_k,\tau_k\right)\sum_{l=1}^{N_0}\mathcal{V}^T\left(r_l,\tau_l\right),$$

where $\mathcal{V}$ is likewise a real-valued vector function defined on $C\times[t_0,\infty)$. In this scenario, a random reordering of the acquired data for a given value of N_0 also does not change the value of $\mathcal{H}$. Therefore, using $p_\theta\left(\tau_1,\ldots,\tau_{N_0}\mid\mathcal{N}(t)=N_0\right)=s_\theta\left(\tau_1\right)\cdots s_\theta\left(\tau_{N_0}\right)$ in Eq. (E.2), and also using Eq. (E.3), we obtain, for $\theta\in\Theta$, the expectation of $\mathcal{H}$ as

$$\mathrm{E}\left[\sum_{k=1}^{N_0}\mathcal{U}\left(r_k,\tau_k\right)\sum_{l=1}^{N_0}\mathcal{V}^T\left(r_l,\tau_l\right)\right]$$
$$= \mathrm{E}\left[\sum_{k=1}^{N_0}\mathcal{U}\left(r_k,\tau_k\right)\mathcal{V}^T\left(r_k,\tau_k\right)\right] + \mathrm{E}\left[\sum_{k,l=1,k\neq l}^{N_0}\mathcal{U}\left(r_k,\tau_k\right)\mathcal{V}^T\left(r_l,\tau_l\right)\right]$$
$$= \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\int_{t_0}^{t}\Lambda_\theta(\tau)\left[\int_C \mathcal{U}\left(r,\tau\right)\mathcal{V}^T\left(r,\tau\right)f_{\theta,\tau}(r)dr\right]d\tau$$
$$+ \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\left(\int_{t_0}^{t}\cdots\int_{t_0}^{t}\left[\int_C\cdots\int_C\left(\sum_{k,l=1,k\neq l}^{N_0}\mathcal{U}\left(r_k,\tau_k\right)\mathcal{V}^T\left(r_l,\tau_l\right)\right)\right.\right.$$
$$\left.\left. f_{\theta,\tau_1}\left(r_1\right)\cdots f_{\theta,\tau_{N_0}}\left(r_{N_0}\right)dr_1\cdots dr_{N_0}\right]s_\theta\left(\tau_1\right)\cdots s_\theta\left(\tau_{N_0}\right)d\tau_1\cdots d\tau_{N_0}\right)P\left(\mathcal{N}(t)=N_0\right)$$
$$= \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\int_{t_0}^{t}\Lambda_\theta(\tau)\left[\int_C \mathcal{U}\left(r,\tau\right)\mathcal{V}^T\left(r,\tau\right)f_{\theta,\tau}(r)dr\right]d\tau$$
$$+ \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\left(\int_{t_0}^{t}\cdots\int_{t_0}^{t}\left[\sum_{k,l=1,k\neq l}^{N_0}\int_C \mathcal{U}\left(r,\tau_k\right)f_{\theta,\tau_k}(r)dr\right.\right.$$
$$\left.\left.\int_C \mathcal{V}^T\left(r,\tau_l\right)f_{\theta,\tau_l}(r)dr\right]s_\theta\left(\tau_1\right)\cdots s_\theta\left(\tau_{N_0}\right)d\tau_1\cdots d\tau_{N_0}\right)P\left(\mathcal{N}(t)=N_0\right)$$
$$= \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\int_{t_0}^{t}\Lambda_\theta(\tau)\left[\int_C \mathcal{U}\left(r,\tau\right)\mathcal{V}^T\left(r,\tau\right)f_{\theta,\tau}(r)dr\right]d\tau$$
$$+ \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}}\sum_{N_0=1}^{\infty}\left(N_0^2-N_0\right)P\left(\mathcal{N}(t)=N_0\right)$$
$$\times\int_{t_0}^{t}\left[\int_C \mathcal{U}\left(r,\tau\right)f_{\theta,\tau}(r)dr\right]s_\theta\left(\tau\right)d\tau\int_{t_0}^{t}\left[\int_C \mathcal{V}^T\left(r,\tau\right)f_{\theta,\tau}(r)dr\right]s_\theta\left(\tau\right)d\tau$$

$$
\begin{aligned}
&= \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \Lambda_\theta(\tau) \left[\int_C \mathcal{U}(r,\tau)\,\mathcal{V}^T(r,\tau)\, f_{\theta,\tau}(r)dr\right] d\tau \\
&\quad + \frac{\left(\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau\right)^2}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}} \\
&\quad \times \int_{t_0}^{t} \left[\int_C \mathcal{U}(r,\tau)\, f_{\theta,\tau}(r)dr\right] s_\theta(\tau)\, d\tau \int_{t_0}^{t} \left[\int_C \mathcal{V}^T(r,\tau)\, f_{\theta,\tau}(r)dr\right] s_\theta(\tau)\, d\tau \\
&= \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \Lambda_\theta(\tau) \left[\int_C \mathcal{U}(r,\tau)\,\mathcal{V}^T(r,\tau)\, f_{\theta,\tau}(r)dr\right] d\tau \\
&\quad + \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}} \\
&\quad \times \int_{t_0}^{t} \Lambda_\theta(\tau) \left[\int_C \mathcal{U}(r,\tau)\, f_{\theta,\tau}(r)dr\right] d\tau \int_{t_0}^{t} \Lambda_\theta(\tau) \left[\int_C \mathcal{V}^T(r,\tau)\, f_{\theta,\tau}(r)dr\right] d\tau. \qquad \text{(E.5)}
\end{aligned}
$$

Here, we have used that $\sum_{N_0=1}^{\infty}\left(N_0^2 - N_0\right) P\left(\mathcal{N}(t) = N_0\right) = \mathrm{E}\left[(\mathcal{N}(t))^2\right] - \mathrm{E}\left[\mathcal{N}(t)\right] = \mathrm{Var}\left(\mathcal{N}(t)\right) + \left(\mathrm{E}\left[\mathcal{N}(t)\right]\right)^2 - \mathrm{E}\left[\mathcal{N}(t)\right] = \left(\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau\right)^2$.

With the above results, we now have the tools we need to evaluate the terms of the Fisher information matrix of Eq. (E.1). Applying Eq. (E.5), and using that $\int_C \mathcal{J}_{\theta,\tau}(r) f_{\theta,\tau}(r)dr = \int_C \frac{\partial f_{\theta,\tau}(r)}{\partial\theta}dr = 0$ for $\theta \in \Theta$ and $\tau \geq t_0$ (Note E.4), we obtain, for $\theta \in \Theta$,

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{J}_{\theta,\tau_k}^T(r_k) \sum_{l=1}^{N_0} \mathcal{J}_{\theta,\tau_l}(r_l)\right] \\
&= \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \Lambda_\theta(\tau) \left[\int_C \mathcal{J}_{\theta,\tau}^T(r)\mathcal{J}_{\theta,\tau}(r) f_{\theta,\tau}(r)dr\right] d\tau \\
&\quad + \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \Lambda_\theta(\tau) \left[\int_C \mathcal{J}_{\theta,\tau}^T(r) f_{\theta,\tau}(r)dr\right] d\tau \int_{t_0}^{t} \Lambda_\theta(\tau) \left[\int_C \mathcal{J}_{\theta,\tau}(r) f_{\theta,\tau}(r)dr\right] d\tau \\
&= \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \Lambda_\theta(\tau) \left[\int_C \frac{1}{f_{\theta,\tau}(r)} \left(\frac{\partial f_{\theta,\tau}(r)}{\partial\theta}\right)^T \frac{1}{f_{\theta,\tau}(r)} \frac{\partial f_{\theta,\tau}(r)}{\partial\theta} f_{\theta,\tau}(r)dr\right] d\tau \\
&= \frac{1}{1-e^{-\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau}} \int_{t_0}^{t} \int_C \frac{\Lambda_\theta(\tau)}{f_{\theta,\tau}(r)} \left(\frac{\partial f_{\theta,\tau}(r)}{\partial\theta}\right)^T \frac{\partial f_{\theta,\tau}(r)}{\partial\theta} dr d\tau.
\end{aligned}
$$

Similarly, application of Eq. (E.5) yields, for $\theta \in \Theta$,

$$\begin{aligned}
&\mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{M}_\theta^T(\tau_k) \sum_{l=1}^{N_0} \mathcal{J}_{\theta,\tau_l}(r_l)\right] \\
&= \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \Lambda_\theta(\tau)\left[\int_C \mathcal{M}_\theta^T(\tau)\mathcal{J}_{\theta,\tau}(r) f_{\theta,\tau}(r)dr\right] d\tau \\
&\quad + \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \Lambda_\theta(\tau)\left[\int_C \mathcal{M}_\theta^T(\tau) f_{\theta,\tau}(r)dr\right] d\tau \int_{t_0}^t \Lambda_\theta(\tau)\left[\int_C \mathcal{J}_{\theta,\tau}(r) f_{\theta,\tau}(r)dr\right] d\tau \\
&= \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \Lambda_\theta(\tau)\mathcal{M}_\theta^T(\tau)\left[\int_C \mathcal{J}_{\theta,\tau}(r) f_{\theta,\tau}(r)dr\right] d\tau \\
&\quad + \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \Lambda_\theta(\tau)\mathcal{M}_\theta^T(\tau)\left[\int_C f_{\theta,\tau}(r)dr\right] d\tau \int_{t_0}^t \Lambda_\theta(\tau)\left[\int_C \mathcal{J}_{\theta,\tau}(r) f_{\theta,\tau}(r)dr\right] d\tau \\
&= 0.
\end{aligned}$$

Using Eq. (E.5) once again, we have, for $\theta \in \Theta$,

$$\begin{aligned}
&\mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{M}_\theta^T(\tau_k) \sum_{l=1}^{N_0} \mathcal{M}_\theta(\tau_l)\right] \\
&\quad = \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \Lambda_\theta(\tau)\mathcal{M}_\theta^T(\tau)\,\mathcal{M}_\theta(\tau)\,d\tau \\
&\qquad + \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \Lambda_\theta(\tau)\mathcal{M}_\theta^T(\tau)\,d\tau \int_{t_0}^t \Lambda_\theta(\tau)\mathcal{M}_\theta(\tau)\,d\tau \\
&\quad = \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \frac{\Lambda_\theta(\tau)}{\Lambda_\theta(\tau)}\left(\frac{\partial\Lambda_\theta(\tau)}{\partial\theta}\right)^T \frac{1}{\Lambda_\theta(\tau)} \frac{\partial\Lambda_\theta(\tau)}{\partial\theta} d\tau \\
&\qquad + \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \left(\frac{\partial\Lambda_\theta(\tau)}{\partial\theta}\right)^T d\tau \int_{t_0}^t \frac{\partial\Lambda_\theta(\tau)}{\partial\theta} d\tau \\
&\quad = \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \frac{1}{\Lambda_\theta(\tau)}\left(\frac{\partial\Lambda_\theta(\tau)}{\partial\theta}\right)^T \frac{\partial\Lambda_\theta(\tau)}{\partial\theta} d\tau \\
&\qquad + \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \left(\frac{\partial\Lambda_\theta(\tau)}{\partial\theta}\right)^T d\tau \int_{t_0}^t \frac{\partial\Lambda_\theta(\tau)}{\partial\theta} d\tau.
\end{aligned}$$

Using that $\int_{t_0}^t \Lambda_\theta(\tau)\mathcal{M}_\theta(\tau)d\tau = \int_{t_0}^t \frac{\partial\Lambda_\theta(\tau)}{\partial\theta} d\tau$, $\theta \in \Theta$, and applying Eq. (E.3), we obtain, for $\theta \in \Theta$,

$$\mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{J}_{\theta,\tau_k}(r_k)\right] = \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \Lambda_\theta(\tau)\left[\int_C \mathcal{J}_{\theta,\tau}(r) f_{\theta,\tau}(r)dr\right] d\tau = 0.$$

Finally, application of Eq. (E.4) yields, for $\theta \in \Theta$,

$$\begin{aligned}
\mathrm{E}\left[\sum_{k=1}^{N_0} \mathcal{M}_\theta(\tau_k)\right] &= \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \Lambda_\theta(\tau)\mathcal{M}_\theta(\tau)d\tau \\
&= \frac{1}{1-e^{-\int_{t_0}^t \Lambda_\theta(\tau)d\tau}} \int_{t_0}^t \frac{\partial\Lambda_\theta(\tau)}{\partial\theta} d\tau.
\end{aligned}$$

Substituting these results into Eq. (E.1), we arrive at the Fisher information matrix

$$
\begin{aligned}
\mathbf{I}(\theta) = {} & \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \int_C \frac{\Lambda_\theta(\tau)}{f_{\theta,\tau}(r)} \left(\frac{\partial f_{\theta,\tau}(r)}{\partial \theta} \right)^T \frac{\partial f_{\theta,\tau}(r)}{\partial \theta} dr d\tau \\
& + \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \frac{1}{\Lambda_\theta(\tau)} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \\
& + \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T d\tau \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \\
& - \left(\frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \right)^2 \int_{t_0}^{t} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T d\tau \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \\
& - \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T d\tau \\
& \qquad \times \left(\frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau - \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \right) \\
= {} & \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \int_C \frac{\Lambda_\theta(\tau)}{f_{\theta,\tau}(r)} \left(\frac{\partial f_{\theta,\tau}(r)}{\partial \theta} \right)^T \frac{\partial f_{\theta,\tau}(r)}{\partial \theta} dr d\tau \\
& + \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \frac{1}{\Lambda_\theta(\tau)} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \\
& + \left(\frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} - \left(\frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \right)^2 \right) \int_{t_0}^{t} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T d\tau \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \\
= {} & \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \int_C \frac{\Lambda_\theta(\tau)}{f_{\theta,\tau}(r)} \left(\frac{\partial f_{\theta,\tau}(r)}{\partial \theta} \right)^T \frac{\partial f_{\theta,\tau}(r)}{\partial \theta} dr d\tau \\
& + \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \int_C \frac{f_{\theta,\tau}(r)}{\Lambda_\theta(\tau)} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} dr d\tau \\
& + \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \int_C \left(\frac{\partial f_{\theta,\tau}(r)}{\partial \theta} \right)^T \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} dr d\tau \\
& + \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \int_C \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T \frac{\partial f_{\theta,\tau}(r)}{\partial \theta} dr d\tau \\
& - \frac{e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}}{\left(1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}\right)^2} \int_{t_0}^{t} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T d\tau \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau \\
= {} & \frac{1}{1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}} \int_{t_0}^{t} \int_C \frac{1}{\Lambda_\theta(\tau) f_{\theta,\tau}(r)} \left(\frac{\partial \left[\Lambda_\theta(\tau) f_{\theta,\tau}(r) \right]}{\partial \theta} \right)^T \frac{\partial \left[\Lambda_\theta(\tau) f_{\theta,\tau}(r) \right]}{\partial \theta} dr d\tau \\
& - \frac{e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}}{\left(1 - e^{-\int_{t_0}^{t} \Lambda_\theta(\tau) d\tau}\right)^2} \int_{t_0}^{t} \left(\frac{\partial \Lambda_\theta(\tau)}{\partial \theta} \right)^T d\tau \int_{t_0}^{t} \frac{\partial \Lambda_\theta(\tau)}{\partial \theta} d\tau.
\end{aligned}
$$

Note that here we have used the fact that $\int_C \frac{\partial f_{\theta,\tau}(r)}{\partial \theta} dr = 0$, $\theta \in \Theta$, $\tau \geq t_0$, to add zero-valued terms that allow us to express $\mathbf{I}(\theta)$ in its final compact form.

F

Fisher Information for the Deterministic Data Models

F.1 Simple form of the deterministic data model

For the simple form of the deterministic data model, in which the Gaussian noise component of the data does not depend on the deterministic signal, the signal component of the data in the kth pixel is denoted as D_k (Eq. (15.22)). Making the assumption that information about the parameter θ is contained only in this signal component, we can derive the Fisher information matrix for the data in the kth pixel as we did in Section 17.3.1. Specifically, we denote D_k as $D_{\theta,k}$, and start by taking exactly the same steps as in Eq. (17.18), but with $\nu_{\theta,k}$ replaced by $D_{\theta,k}$, to obtain the Fisher information matrix $\mathbf{I}_k(\theta)$ as

$$\mathbf{I}_k(\theta) = \left(\frac{\partial D_{\theta,k}}{\partial \theta}\right)^T \frac{\partial D_{\theta,k}}{\partial \theta} \cdot \mathrm{E}\left[\left(\frac{\partial \ln p_{\theta,k}(z_k)}{\partial D_{\theta,k}}\right)^2\right]. \tag{F.1}$$

To evaluate the expectation term, we replace $\ln p_{\theta,k}(z_k)$ with the log-likelihood function for the deterministic data model as shown in Table 16.1, and we obtain

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\partial \ln p_{\theta,k}(z_k)}{\partial D_{\theta,k}}\right)^2\right] &= \mathrm{E}\left[\left(\frac{\partial}{\partial D_{\theta,k}}\left(-\ln\left(\sqrt{2\pi\xi_k^2}\right) - \frac{(z_k - D_{\theta,k} - \zeta_k)^2}{2\xi_k^2}\right)\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{z_k - D_{\theta,k} - \zeta_k}{\xi_k^2}\right)^2\right] = \int_{-\infty}^{\infty}\left(\frac{z - D_{\theta,k} - \zeta_k}{\xi_k^2}\right)^2 p_{\theta,k}(z)dz \\
&= \frac{1}{\xi_k^4} \cdot \frac{1}{\sqrt{2\pi\xi_k^2}} \int_{-\infty}^{\infty} (z - D_{\theta,k} - \zeta_k)^2 e^{-\frac{(z - D_{\theta,k} - \zeta_k)^2}{2\xi_k^2}} dz \\
&= \frac{1}{\xi_k^4} \cdot \frac{1}{\sqrt{2\pi\xi_k^2}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2\xi_k^2}} dz = \frac{1}{\xi_k^4} \cdot \xi_k^2 = \frac{1}{\xi_k^2},
\end{aligned}$$

where we have used the change of variables $x = z - D_{\theta,k} - \zeta_k$ and the fact that the second central moment of a Gaussian distribution is the variance of the distribution.

Since the expectation term is the Fisher information with respect to $D_{\theta,k}$, we could also arrive at the same expression by applying, for a scalar parameter θ, the following result for the Fisher information $\mathbf{I}_{Gaussian}(\theta)$ of a Gaussian random variable with mean χ_θ and variance ϵ_θ^2:

$$\mathbf{I}_{Gaussian}(\theta) = \frac{1}{\epsilon_\theta^2}\left(\frac{\partial \chi_\theta}{\partial \theta}\right)^2 + \frac{1}{2\epsilon_\theta^4}\left(\frac{\partial \epsilon_\theta^2}{\partial \theta}\right)^2. \tag{F.2}$$

One can easily verify that by letting $\theta = D_{\theta,k}$, $\chi_\theta = D_{\theta,k} + \zeta_k$, and $\epsilon_\theta^2 = \xi_k^2$, we obtain the same result.

Now that we have $\mathbf{I}_k(\theta)$, the Fisher information matrix corresponding to an N_p-pixel image is just

$$\mathbf{I}(\theta) = \sum_{k=1}^{N_p} \mathbf{I}_k(\theta) = \sum_{k=1}^{N_p} \left(\frac{\partial D_{\theta,k}}{\partial \theta}\right)^T \left(\frac{\partial D_{\theta,k}}{\partial \theta}\right) \cdot \frac{1}{\xi_k^2}. \tag{F.3}$$

F.2 Gaussian approximation for the CCD/CMOS data model

As discussed in Section 15.2.3.1, when signal levels are high, the data acquired in the pixels of a CCD/CMOS image may be described using the deterministic data model. Specifically, the mean of the Poisson signal detected in a pixel may be viewed as the value of a constant signal, while the variance of the Poisson signal may be accounted for by the variance of a Gaussian random variable. For an N_p-pixel image that is modeled in this way, the log-likelihood function for the data z_k in the kth pixel, $k = 1, \ldots, N_p$, is as given in Table 16.1 for the Gaussian approximation for the CCD/CMOS data model. The noise coefficient can be obtained by substituting $\ln p_{\theta,k}(z_k)$ in the expectation term of Eq. (17.19) with this log-likelihood function, and simplifying. A faster approach, however, is to calculate the expectation term by applying Eq. (F.2) with $\theta = \nu_{\theta,k}$, $\chi_\theta = \nu_{\theta,k} + \eta_k$, and $\epsilon_\theta^2 = \nu_{\theta,k} + \sigma_k^2$. Either approach will yield the noise coefficient

$$\gamma_k = \nu_{\theta,k} \cdot \mathrm{E}\left[\left(\frac{\partial \ln p_{\theta,k}(z_k)}{\partial \nu_{\theta,k}}\right)^2\right] = \nu_{\theta,k} \cdot \left(\frac{1}{\nu_{\theta,k} + \sigma_k^2} + \frac{1}{2\left(\nu_{\theta,k} + \sigma_k^2\right)^2}\right).$$

F.3 Gaussian approximation for the EMCCD data model

As seen in Section 15.2.4.2, an analogous approach to the Gaussian approximation for the CCD/CMOS data model may be taken to model the data in the pixels of an EMCCD image. Provided that the initial Poisson signal detected in a pixel is large, the mean of the amplified signal in that pixel may be taken to be the value of a deterministic signal, while the variance of the amplified signal may be accounted for by the variance of a Gaussian random variable. For an N_p-pixel image modeled in this manner, the log-likelihood function for the data z_k in the kth pixel, $k = 1, \ldots, N_p$, is as shown in Table 16.1 for the Gaussian approximation for the EMCCD data model. As noted in Section 15.2.4.2, this Gaussian approximation assumes that the EMCCD detector's signal amplification is described by the geometric model of electron multiplication (Section G.1).

Analogous to the case of the Gaussian approximation for the CCD/CMOS data model (Section F.2), the expectation term in the general noise coefficient expression of Eq. (17.19) can be directly evaluated using the log-likelihood function from Table 16.1, or it can be calculated by utilizing Eq. (F.2), in this case with $\theta = \nu_{\theta,k}$, $\chi_\theta = g\nu_{\theta,k} + \eta_k$, and $\epsilon_\theta^2 = (2g^2 - g)\nu_{\theta,k} + \sigma_k^2$. Either method will produce the noise coefficient

$$\begin{aligned}\gamma_k &= \nu_{\theta,k} \cdot \mathrm{E}\left[\left(\frac{\partial \ln p_{\theta,k}(z_k)}{\partial \nu_{\theta,k}}\right)^2\right] \\ &= \nu_{\theta,k} \cdot \left(\frac{g^2}{(2g^2 - g)\nu_{\theta,k} + \sigma_k^2} + \frac{(2g^2 - g)^2}{2\left((2g^2 - g)\nu_{\theta,k} + \sigma_k^2\right)^2}\right).\end{aligned}$$

G

Models of EMCCD Electron Multiplication

This part of the appendix is devoted to the presentation of further derivations and results relating to the modeling of EMCCD image data.

G.1 Geometric multiplication-based EMCCD probability mass function

To obtain the probability mass function of Eq. (15.26), which results from the modeling of the multiplication of electrons through the typically hundreds of stages of an EMCCD detector's multiplication register, we start with an introduction to branching processes.

G.1.1 Branching processes

A branching process is a stochastic process whereby particles are passed through a sequence of stages, and in each stage, a particle that enters the stage (i.e., an input particle) generates a number of secondary, or offspring, particles with a certain probability. All particles, input or offspring, that exit a given stage enter the next stage as input particles. In this way, after many stages of multiplication, a potentially large number of particles are generated from a comparatively small number of initial particles.

The described cascading effect of a branching process is captured in a sequence of random variables characterized by two probability distributions: one for the initial number of particles to be multiplied, and one for the number of offspring particles that result from an input particle at a given stage of multiplication. More formally, a branching process with an initial particle count probability distribution $(q_i)_{i=0,1,\ldots}$ and an individual offspring count probability distribution $(p_i)_{i=0,1,\ldots}$ is defined as a sequence of nonnegative and integer-valued random variables $(X_n)_{n=0,1,\ldots}$, given by

$$X_n = \begin{cases} U, & n = 0, \\ \sum_{j=1}^{X_{n-1}} Y_j, & n = 1, 2, \ldots. \end{cases} \tag{G.1}$$

The random variable X_n, $n = 0, , 1, \ldots$, represents the number of particles at the output of the nth stage of multiplication. The output of the zeroth stage, $X_0 = U$, is understood to be the initial particle count prior to multiplication, and is distributed according to the probability distribution $(q_i)_{i=0,1,\ldots}$. For $n = 1, 2, \ldots$, the output of the nth stage, X_n, is the sum of the offspring particle counts Y_j, $j = 1, \ldots, X_{n-1}$, that arise from the X_{n-1} input particles to the nth stage. The random variables Y_j, $j = 1, \ldots, X_{n-1}$, are nonnegative and integer-valued, and are mutually independent and identically distributed according to the probability distribution $(p_i)_{i=0,1,\ldots}$. Note that if $X_{n-1} = 0$, then $\sum_{j=1}^{X_{n-1}} Y_j = 0$. Note also that the offspring particle count Y_j includes the jth input particle that produced the offspring particles.

G.1.2 Electron multiplication as a branching process

For our application, the particles are electrons, and the stages of multiplication are the stages of an EMCCD detector's multiplication register. Furthermore, the initial electron count is the number of electrons resulting from the object and background signals detected at a given pixel k, and is therefore Poisson-distributed with mean ν_k. Hence the initial electron count probability distribution is given by

$$q_i = \frac{e^{-\nu_k}\nu_k^i}{i!}, \quad i = 0, 1, \ldots.$$

Suppose the EMCCD detector has an N-stage multiplication register. Then by the law of total probability, the probability distribution of the number of electrons at the output of the register, i.e., the amplified signal, is given by

$$\begin{aligned} p_N(x) = P(X_N = x) &= \sum_{i=0}^{\infty} P(X_N = x \mid X_0 = i) \cdot P(X_0 = i) \\ &= \sum_{i=0}^{\infty} P(X_N = x \mid X_0 = i) \cdot q_i \\ &= \sum_{i=0}^{\infty} P(X_N = x \mid X_0 = i) \cdot \frac{e^{-\nu_k}\nu_k^i}{i!}, \quad x = 0, 1, \ldots. \end{aligned} \tag{G.2}$$

G.1.3 Geometric distribution of offspring electrons

To arrive at an explicit expression for $p_N(x)$, we need to determine $P(X_N = x \mid X_0 = i)$, the probability of obtaining x electrons at the output of the multiplication register, given that there are i electrons to begin with. In order to do so, we need to define $(p_i)_{i=0,1,\ldots}$, the probability distribution of the number of offspring electrons that result from a single input electron at any given multiplication stage. We consider the geometric probability distribution

$$p_i = (1-b)b^{i-1}, \quad i = 1, 2, \ldots, \quad 0 \le b < 1. \tag{G.3}$$

Note that this distribution does not allow for the possibility of losing an input electron, as it is not defined for $i = 0$ (Note G.1).

The geometric probability distribution specifies the individual offspring count to take on increasing positive integer values with decreasing probabilities. It is therefore suitable for modeling electron multiplication in an EMCCD detector, where the generation of more than one offspring electron by a single input electron at any given stage in the multiplication register can be assumed to be an extremely rare event.

Importantly, the geometric probability distribution has a corresponding probability generating function that is of linear fractional form. As we demonstrate in the following subsections, this special property allows us to obtain, by way of probability generating functions, an explicit expression (i.e., a non-recursive expression) for $P(X_N = x \mid X_0 = i)$ (Note G.2).

G.1.4 Probability generating function of output electron count given one initial electron

We begin with the probability generating function of the geometric probability distribution $(p_i)_{i=1,2,\ldots}$ which, by definition, is given by

$$f(s) = \sum_{i=1}^{\infty} p_i s^i = \sum_{i=1}^{\infty}(1-b)b^{i-1}s^i = (1-b)s\sum_{i=1}^{\infty}(bs)^{i-1} = \frac{(1-b)s}{1-bs}, \quad s \in \mathbb{C}, \quad |s| \le 1.$$

Given that there is exactly one initial electron, the probability generating function corresponding to the probability distribution of the electron count at the output of the N-stage multiplication register is the Nth iterate of $f(s)$ (Note G.2). Denoted by $f_N(s)$, $N = 0, 1, \ldots,$ the Nth iterate of $f(s)$ is given by

$$f_N(s) = \frac{(1-b)^N s}{1-(1-(1-b)^N)s}, \quad s \in \mathbb{C}, \quad |s| \leq 1, \quad b \neq 0,$$

which we prove by induction. For our base case, we have that the expression satisfies the definition that the zeroth iterate of a probability generating function $f(s)$ is equal to s:

$$f_0(s) = s = \frac{(1-b)^0 s}{1-(1-(1-b)^0)s}.$$

For the inductive step, we assume that the expression is true for any nonnegative integer $N = n$, and show that it is true for $N = n + 1$ using the fact that $f_{n+1}(s) = f_n(f(s))$:

$$\begin{aligned} f_{n+1}(s) = f_n(f(s)) &= \frac{(1-b)^n \left(\frac{(1-b)s}{1-bs}\right)}{1-(1-(1-b)^n)\left(\frac{(1-b)s}{1-bs}\right)} = \frac{\frac{(1-b)^n(1-b)s}{1-bs}}{\frac{1-bs-(1-(1-b)^n)(1-b)s}{1-bs}} \\ &= \frac{(1-b)^{n+1}s}{1-bs-(1-b-(1-b)^{n+1})s} = \frac{(1-b)^{n+1}s}{1-(1-(1-b)^{n+1})s}. \end{aligned}$$

G.1.5 Probability generating function of output electron count given i initial electrons

It can be inferred from the branching process definition of Eq. (G.1) that while all initial particles will multiply according to a common stochastic model, each of them will multiply independently of the others. Provided that the initial particle count is a positive integer i, the probability distribution of the particle count at the output of an N-stage process will then be the i-fold convolution of the probability distribution of the output particle count, given one initial particle, with itself. In our current case, the output electron count probability distribution, given i initial electrons, is therefore the i-fold convolution of the probability distribution corresponding to $f_N(s)$ with itself. By a well-known result, the probability generating function corresponding to the probability distribution that results from the i-fold convolution is given simply by $[f_N(s)]^i$.

G.1.6 Probability mass function of output electron count given i initial electrons

Given $[f_N(s)]^i$, we can now obtain an expression for the conditional probability $P(X_N = x \mid X_0 = i)$ using well-known properties of the probability generating function. For an output electron count of $x = 0$, the conditional probability is given by $[f_N(s)]^i$ evaluated at $s = 0$, i.e.,

$$P(X_N = 0 \mid X_0 = i) = [f_N(0)]^i = \left[\frac{(1-b)^N \cdot 0}{1-(1-(1-b)^N)\cdot 0}\right]^i = 0, \quad i = 1, 2, \ldots. \qquad \text{(G.4)}$$

The probability of 0 for $X_N = 0$ is as expected, since given the geometric model of multiplication, which does not allow for the possibility of electron loss during the multiplication process, it should not be possible to have no electrons at the output of the multiplication register when there is a positive number of electrons to begin with.

For an output electron count of $x = 1, 2, \ldots$, the conditional probability is given by the xth derivative of $[f_N(s)]^i$, evaluated at $s = 0$ and divided by $x!$. This gives the expression

$$P(X_N = x \mid X_0 = i) = \begin{cases} \binom{x-1}{i-1}\left[(1-b)^N\right]^i\left[1-(1-b)^N\right]^{x-i}, & i = 1, 2, \ldots, x, \\ 0, & i \geq x+1, \end{cases} \quad \text{(G.5)}$$

where the probability of 0 for $i \geq x+1$ is again due to the impossibility of electron loss given the geometric model of multiplication. To derive the conditional probability of Eq. (G.5), we first show, by induction, that the xth derivative of $[f_N(s)]^i$ is given by

$$\frac{\partial^x [f_N(s)]^i}{\partial s^x} = \sum_{j=1}^{x} \binom{x}{j} \frac{(x-1)!}{(j-1)!} \frac{\left[(1-b)^N\right]^j \left[1-(1-b)^N\right]^{x-j}}{(1-(1-(1-b)^N)s)^{x+j}} \frac{i!}{(i-j)!} \left[\frac{(1-b)^N s}{1-(1-(1-b)^N)s}\right]^{i-j}. \quad \text{(G.6)}$$

For our base case $x = 1$, we show that the first derivative of $[f_N(s)]^i$ is given by Eq. (G.6) with $x = 1$, i.e.,

$$\begin{aligned}
\frac{\partial [f_N(s)]^i}{\partial s} &= \frac{(1-b)^N}{(1-(1-(1-b)^N)s)^2}\, i \left[\frac{(1-b)^N s}{1-(1-(1-b)^N)s}\right]^{i-1} \\
&= \binom{1}{1}\frac{0!}{0!}\frac{\left[(1-b)^N\right]^1\left[1-(1-b)^N\right]^0}{(1-(1-(1-b)^N)s)^2}\frac{i!}{(i-1)!}\left[\frac{(1-b)^N s}{1-(1-(1-b)^N)s}\right]^{i-1} \\
&= \sum_{j=1}^{1}\binom{1}{j}\frac{(1-1)!}{(j-1)!}\frac{\left[(1-b)^N\right]^j\left[1-(1-b)^N\right]^{1-j}}{(1-(1-(1-b)^N)s)^{1+j}}\frac{i!}{(i-j)!}\left[\frac{(1-b)^N s}{1-(1-(1-b)^N)s}\right]^{i-j}.
\end{aligned}$$

For the inductive step, we assume that Eq. (G.6) is correct for any positive integer x, and show that it is also correct for $x+1$ by simply taking its derivative and demonstrating that the $(x+1)$th derivative is of the exact same form, i.e.,

$$\begin{aligned}
\frac{\partial^{x+1} [f_N(s)]^i}{\partial s^{x+1}} &= \frac{\partial}{\partial s}\left[\frac{\partial^x [f_N(s)]^i}{\partial s^x}\right] \\
&= \sum_{j=1}^{x}\binom{x}{j}\frac{(x-1)!}{(j-1)!}\frac{\left[(1-b)^N\right]^{j+1}\left[1-(1-b)^N\right]^{x-j}}{(1-(1-(1-b)^N)s)^{x+2+j}}\frac{i!}{(i-j-1)!} \\
&\qquad \times \left[\frac{(1-b)^N s}{1-(1-(1-b)^N)s}\right]^{i-j-1} \\
&\quad + \sum_{j=1}^{x}\binom{x}{j}\frac{(x-1)!}{(j-1)!}(x+j)\frac{\left[(1-b)^N\right]^j\left[1-(1-b)^N\right]^{x+1-j}}{(1-(1-(1-b)^N)s)^{x+1+j}}\frac{i!}{(i-j)!} \\
&\qquad \times \left[\frac{(1-b)^N s}{1-(1-(1-b)^N)s}\right]^{i-j} \\
&= \frac{\left[(1-b)^N\right]^{x+1}}{(1-(1-(1-b)^N)s)^{2(x+1)}}\frac{i!}{(i-(x+1))!}\left[\frac{(1-b)^N s}{1-(1-(1-b)^N)s}\right]^{i-(x+1)}
\end{aligned}$$

$$+\sum_{j=1}^{x-1}\binom{x}{j}\frac{(x-1)!}{(j-1)!}\frac{\left[(1-b)^N\right]^{j+1}\left[1-(1-b)^N\right]^{x-j}}{\left(1-\left(1-(1-b)^N\right)s\right)^{x+2+j}}\frac{i!}{(i-j-1)!}$$

$$\times\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-j-1}$$

$$+\sum_{j=2}^{x}\binom{x}{j}\frac{(x-1)!}{(j-1)!}(x+j)\frac{\left[(1-b)^N\right]^{j}\left[1-(1-b)^N\right]^{x+1-j}}{\left(1-\left(1-(1-b)^N\right)s\right)^{x+1+j}}\frac{i!}{(i-j)!}$$

$$\times\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-j}$$

$$+(x+1)!\frac{(1-b)^N\left[1-(1-b)^N\right]^{x}}{\left(1-\left(1-(1-b)^N\right)s\right)^{x+2}}\ i\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-1}$$

$$=\frac{\left[(1-b)^N\right]^{x+1}}{\left(1-\left(1-(1-b)^N\right)s\right)^{2(x+1)}}\frac{i!}{(i-(x+1))!}\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-(x+1)}$$

$$+\sum_{j=2}^{x}\binom{x}{j-1}\frac{(x-1)!}{(j-2)!}\frac{\left[(1-b)^N\right]^{j}\left[1-(1-b)^N\right]^{x+1-j}}{\left(1-\left(1-(1-b)^N\right)s\right)^{x+1+j}}\frac{i!}{(i-j)!}$$

$$\times\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-j}$$

$$+\sum_{j=2}^{x}\binom{x}{j}\frac{(x-1)!}{(j-1)!}(x+j)\frac{\left[(1-b)^N\right]^{j}\left[1-(1-b)^N\right]^{x+1-j}}{\left(1-\left(1-(1-b)^N\right)s\right)^{x+1+j}}\frac{i!}{(i-j)!}$$

$$\times\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-j}$$

$$+(x+1)!\frac{(1-b)^N\left[1-(1-b)^N\right]^{x}}{\left(1-\left(1-(1-b)^N\right)s\right)^{x+2}}\ i\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-1}$$

$$=\binom{x+1}{x+1}\frac{x!}{(x+1-1)!}\frac{\left[(1-b)^N\right]^{x+1}\left[1-(1-b)^N\right]^{x+1-(x+1)}}{\left(1-\left(1-(1-b)^N\right)s\right)^{x+1+x+1}}\frac{i!}{(i-(x+1))!}$$

$$\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-(x+1)}$$

$$+\sum_{j=2}^{x}\binom{x+1}{j}\frac{x!}{(j-1)!}\frac{\left[(1-b)^N\right]^{j}\left[1-(1-b)^N\right]^{x+1-j}}{\left(1-\left(1-(1-b)^N\right)s\right)^{x+1+j}}\frac{i!}{(i-j)!}$$

$$\times\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-j}$$

$$+\binom{x+1}{1}\frac{x!}{(1-1)!}\frac{\left[(1-b)^N\right]^{1}\left[1-(1-b)^N\right]^{x+1-1}}{\left(1-\left(1-(1-b)^N\right)s\right)^{x+1+1}}\frac{i!}{(i-1)!}$$

$$\times\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-1}$$

$$=\sum_{j=1}^{x+1}\binom{x+1}{j}\frac{x!}{(j-1)!}\frac{\left[(1-b)^N\right]^{j}\left[1-(1-b)^N\right]^{x+1-j}}{\left(1-\left(1-(1-b)^N\right)s\right)^{x+1+j}}\frac{i!}{(i-j)!}\left[\frac{(1-b)^N s}{1-\left(1-(1-b)^N\right)s}\right]^{i-j}.$$

To obtain Eq. (G.5), we now evaluate Eq. (G.6) at $s = 0$ and divide the result by $x!$, yielding

$$\begin{aligned}
P\left(X_N = x \mid X_0 = i\right) &= \frac{1}{x!} \cdot \left.\frac{\partial^x \left[f_N(s)\right]^i}{\partial s^x}\right|_{s=0} \\
&= \frac{1}{x!} \sum_{j=1}^{x} \binom{x}{j} \frac{(x-1)!}{(j-1)!} \left[(1-b)^N\right]^j \left[1-(1-b)^N\right]^{x-j} \frac{i!}{(i-j)!} \cdot 0^{i-j} \\
&= \frac{1}{x!} \binom{x}{i} \frac{(x-1)!}{(i-1)!} \left[(1-b)^N\right]^i \left[1-(1-b)^N\right]^{x-i} \frac{i!}{(i-i)!} \\
&= \binom{x-1}{i-1} \left[(1-b)^N\right]^i \left[1-(1-b)^N\right]^{x-i}, \quad i = 1, 2, \ldots, x.
\end{aligned}$$

Note that in going from the second to the third equality, the summation is removed by observing that 0^{i-j} is nonzero only when $j = i$, when we have $0^0 = 1$.

G.1.7 Probability mass function of output electron count

We now have the necessary tools to use the general expression of Eq. (G.2) to obtain the the probability distribution of X_N, the number of electrons at the output of an N-stage EMCCD multiplication register, assuming the electrons are multiplied according to the geometric model. We know from Eq. (G.4) that the probability of having $X_N = 0$ electrons at the output of the register is 0 when the initial electron count X_0 is greater than or equal to 1. The event of having no output electrons does occur, however, with probability 1, when the initial electron count is 0 (i.e., when there is nothing to multiply in the first place). For $X_N = 0$, we therefore have, from Eq. (G.2),

$$\begin{aligned}
p_N(0) = P\left(X_N = 0\right) &= \sum_{i=0}^{\infty} P\left(X_N = 0 \mid X_0 = i\right) \cdot \frac{e^{-\nu_k} \nu_k^i}{i!} \\
&= P\left(X_N = 0 \mid X_0 = 0\right) \cdot \frac{e^{-\nu_k} \nu_k^0}{0!} = e^{-\nu_k}. \qquad \text{(G.7)}
\end{aligned}$$

For output electron counts $X_N = x$, where $x = 1, 2, \ldots$, we use Eq. (G.5) in Eq. (G.2) to obtain

$$\begin{aligned}
p_N(x) = P\left(X_N = x\right) &= \sum_{i=0}^{\infty} P\left(X_N = x \mid X_0 = i\right) \cdot \frac{e^{-\nu_k} \nu_k^i}{i!} \\
&= \sum_{i=1}^{x} \binom{x-1}{i-1} \left[(1-b)^N\right]^i \left[1-(1-b)^N\right]^{x-i} \frac{e^{-\nu_k} \nu_k^i}{i!} \\
&= e^{-\nu_k} \sum_{i=1}^{x} \frac{\binom{x-1}{i-1}}{i!} \left[(1-b)^N \nu_k\right]^i \left[1-(1-b)^N\right]^{x-i} \\
&= e^{-\nu_k} \sum_{i=0}^{x-1} \frac{\binom{x-1}{i}}{(i+1)!} \left[(1-b)^N \nu_k\right]^{i+1} \left[1-(1-b)^N\right]^{x-1-i} \\
&= e^{-\nu_k} \sum_{i=0}^{x-1} \frac{\binom{x-1}{i}}{(i+1)!} \left(\frac{\nu_k}{g}\right)^{i+1} \left(1-\frac{1}{g}\right)^{x-1-i}, \qquad \text{(G.8)}
\end{aligned}$$

where $g = \frac{1}{(1-b)^N}$ is the *mean gain* of the geometrically multiplied branching process, and in our current application is referred to as the electron multiplication gain. The mean gain is the mean output particle count given a single initial particle (i.e, given that $X_0 = 1$), and is well known to be the mean of the individual offspring count probability distribution $(p_i)_{i=0,1,...}$ raised to the power of the number of multiplication stages N. In our current case, the mean of the geometric distribution of Eq. (G.3) is $\frac{1}{1-b}$, and hence $g = \frac{1}{(1-b)^N}$. Together, Eqs. (G.7) and (G.8) give us Eq. (15.26), with μ_k replaced by ν_k.

G.2 Exponential multiplication-based EMCCD probability density function

The exponential multiplication-based probability density function of Eq. (15.28) approximates well the geometric multiplication-based probability mass function of Eq. (15.26) when the electron multiplication gain g is large. To demonstrate how it is obtained, we start with the fact that if $(X_n)_{n=0,1,...}$ is a branching process with a single initial particle (i.e., $X_0 = 1$) and an individual offspring count probability distribution $(p_i)_{i=0,1,...}$ given by the geometric distribution of Eq. (G.3) with mean $\frac{1}{1-b} > 1$ (i.e., with $0 < b < 1$), then as the number of multiplication stages n converges to infinity, the sequence of random variables $(Y_n = X_n/g)_{n=0,1,...}$, where $g = \frac{1}{(1-b)^N}$ is the mean gain of the branching process $(X_n)_{n=0,1,...}$, converges, with probability 1 and in mean square, to a random variable Y with an exponential distribution given by

$$p(y) = e^{-y}, \quad y > 0. \tag{G.9}$$

This result allows us to approximate, for a large N, the output particle count X_N of an N-stage, geometric multiplication-based branching process with $X_0 = 1$ and $\frac{1}{1-b} > 1$ by the random variable gY. The probability density function of $X_N \approx gY$ is given by the following scaled version of Eq. (G.9):

$$p(x) = \frac{1}{g}e^{-\frac{x}{g}}, \quad x > 0. \tag{G.10}$$

If the branching process has instead an initial particle count of $X_0 = i$, $i = 1, 2, \ldots$, then, by the fact that the i particles are multiplied independently, but each according to the same geometric distribution for the individual offspring count, the output particle count X_N may be approximated as the sum of i independent exponential random variables, each with parameter $\frac{1}{g}$. The output particle count X_N, conditioned on the initial particle count i, is then distributed according to the Erlang probability density function

$$p(x \mid i) = \frac{e^{-\frac{x}{g}}}{g(i-1)!}\left(\frac{x}{g}\right)^{i-1}, \quad x > 0, \tag{G.11}$$

which results from the i-fold convolution of $p(x)$ of Eq. (G.10) with itself.

We are now in a position to determine $p_N(x)$, the approximate probability distribution for the output particle count X_N. For $X_N = 0$, we have

$$p_N(0) = P(X_N = 0) = e^{-\nu_k}, \tag{G.12}$$

just like Eq. (G.7) since the event of having no output electrons only occurs, with probability 1, when the initial electron count is 0. For $X_N > 0$, we use the conditional probability density function of Eq. (G.11) in

$$p_N(x) = \sum_{i=1}^{\infty} p(x \mid i) \cdot P(X_0 = i) = \sum_{i=1}^{\infty} p(x \mid i) \cdot q_i = e^{-\nu_k} \sum_{i=1}^{\infty} \frac{p(x \mid i)\, \nu_k^i}{i!}, \quad x > 0, \tag{G.13}$$

a continuous analogue of Eq. (G.2), to obtain

$$
\begin{aligned}
p_N(x) &= e^{-\nu_k} \sum_{i=1}^{\infty} \frac{e^{-\frac{x}{g}}}{g(i-1)!} \left(\frac{x}{g}\right)^{i-1} \frac{\nu_k^i}{i!} = e^{-\left(\nu_k + \frac{x}{g}\right)} \sum_{i=1}^{\infty} \frac{\left(\frac{x}{g}\right)^{i-1} \nu_k^i}{g(i-1)!i!} \\
&= e^{-\left(\nu_k + \frac{x}{g}\right)} \frac{\nu_k}{g} \sum_{i=0}^{\infty} \frac{\left(\frac{\nu_k x}{g}\right)^i}{i!(i+1)!} \\
&= e^{-\left(\nu_k + \frac{x}{g}\right)} \frac{\sqrt{\nu_k x/g}}{x} \left[\sqrt{\nu_k x/g} \sum_{i=0}^{\infty} \frac{\left(\frac{1}{4}\left(2\sqrt{\nu_k x/g}\right)^2\right)^i}{i!(i+1)!} \right] \\
&= \frac{e^{-\left(\nu_k + \frac{x}{g}\right)} \sqrt{\nu_k x/g}\; I_1\left(2\sqrt{\nu_k x/g}\right)}{x}, \quad x > 0,
\end{aligned}
\tag{G.14}
$$

where we have used the modified Bessel function identity (Note G.3)

$$
I_w(u) = \left(\frac{1}{2}u\right)^w \sum_{j=0}^{\infty} \frac{\left(\frac{1}{4}u^2\right)^j}{j!(w+j)!}, \quad w \in \{0, 1, \dots\}, \quad u \in \mathbb{R},
\tag{G.15}
$$

with $w = 1$. Together, Eqs. (G.12) and (G.14) give us Eq. (15.28), with μ_k replaced by ν_k (Note G.4).

G.3 Geometric multiplication-based EMCCD noise coefficient

Assume that the log-likelihood function $\ln p_{\theta,k}(z_k)$ for the data in the kth pixel of an EMCCD image is as given in Table 16.1 for the EMCCD data model, which is based on the geometric model of electron multiplication detailed in Section G.1. Then using Eq. (17.19), the noise coefficient γ_k for the kth pixel is given by

$$
\begin{aligned}
\gamma_k &= \nu_{\theta,k} \cdot \mathrm{E}\left[\left(\frac{\partial \ln p_{\theta,k}(z_k)}{\partial \nu_{\theta,k}}\right)^2\right] \\
&= \nu_{\theta,k} \cdot \mathrm{E}\left[\left(\frac{\partial}{\partial \nu_{\theta,k}} \left(-\nu_{\theta,k} - \ln\left(\sqrt{2\pi}\sigma_k\right) \right.\right.\right. \\
&\qquad \left.\left.\left. + \ln\left[e^{-\left(\frac{z_k - \eta_k}{\sqrt{2}\sigma_k}\right)^2} + \sum_{l=1}^{\infty} e^{-\left(\frac{z_k - l - \eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1 - \frac{1}{g}\right)^{l-j-1}}{(j+1)!\left(\frac{g}{\nu_{\theta,k}}\right)^{j+1}}\right]\right)\right)^2\right] \\
&= \nu_{\theta,k} \cdot \mathrm{E}\left[\left(-1 + \frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k \cdot p_{\theta,k}(z_k)} \sum_{l=1}^{\infty} e^{-\left(\frac{z_k - l - \eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1 - \frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}}\right)^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \nu_{\theta,k} \cdot \left(\mathrm{E}\left[\left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k \cdot p_{\theta,k}(z_k)} \sum_{l=1}^{\infty} e^{-\left(\frac{z_k - l - \eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} \right)^2 \right] \right. \\
&\qquad \left. -2 \cdot \mathrm{E}\left[\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k \cdot p_{\theta,k}(z_k)} \sum_{l=1}^{\infty} e^{-\left(\frac{z_k - l - \eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} \right] + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \sum_{l=1}^{\infty} e^{-\left(\frac{z - l - \eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} \right)^2 dz \right. \\
&\qquad \left. -2 \cdot \sum_{l=1}^{\infty} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\left(\frac{z - l - \eta_k}{\sqrt{2}\sigma_k}\right)^2} dz \right) e^{-\nu_{\theta,k}} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \sum_{l=1}^{\infty} e^{-\left(\frac{z - l - \eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} \right)^2 dz \right. \\
&\qquad \left. -2 \cdot \sum_{l=1}^{\infty} e^{-\nu_{\theta,k}} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \sum_{l=1}^{\infty} e^{-\left(\frac{z - l - \eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} \right)^2 dz \right. \\
&\qquad \left. -2 \cdot \sum_{l=1}^{\infty} \left(\frac{\partial}{\partial \nu_{\theta,k}} p_{\nu_{\theta,k},g}(l) + p_{\nu_{\theta,k},g}(l) \right) + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \sum_{l=1}^{\infty} e^{-\left(\frac{z - l - \eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} \right)^2 dz \right. \\
&\qquad \left. -2 \cdot \left[\frac{\partial}{\partial \nu_{\theta,k}} \left(\sum_{l=1}^{\infty} p_{\nu_{\theta,k},g}(l) \right) + \sum_{l=1}^{\infty} p_{\nu_{\theta,k},g}(l) \right] + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \sum_{l=1}^{\infty} e^{-\left(\frac{z - l - \eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} \right)^2 dz \right. \\
&\qquad \left. -2 \cdot \left[\frac{\partial}{\partial \nu_{\theta,k}} \left(1 - e^{-\nu_{\theta,k}}\right) + 1 - e^{-\nu_{\theta,k}} \right] + 1 \right)
\end{aligned}
$$

$$= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \sum_{l=1}^{\infty} e^{-\left(\frac{z-l-\eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} \right)^2 dz \right.$$
$$\left. -2 \cdot 1 + 1 \right)$$
$$= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \sum_{l=1}^{\infty} e^{-\left(\frac{z-l-\eta_k}{\sqrt{2}\sigma_k}\right)^2} \sum_{j=0}^{l-1} \frac{\binom{l-1}{j}\left(1-\frac{1}{g}\right)^{l-j-1} \nu_{\theta,k}^j}{j! g^{j+1}} \right)^2 dz - 1 \right),$$

where $p_{\nu_{\theta,k},g}$ in the intermediate steps is the probability mass function of Eq. (15.26) with μ_k replaced by $\nu_{\theta,k}$, and we have used the fact that it sums to $1 - e^{-\nu_{\theta,k}}$ from $l = 1$ to infinity.

G.4 Exponential multiplication-based EMCCD noise coefficient

Assume that the log-likelihood function $\ln p_{\theta,k}(z_k)$ for the data in the kth pixel of an EMCCD image is as given in Table 16.1 for the high gain approximation for the EMCCD data model. The high gain approximation is obtained, as shown in Section G.2, by using an exponential random variable to describe the output of a geometrically multiplied branching process with a single initial particle. Using Eq. (17.19), the noise coefficient γ_k for the kth pixel is then given by

$$\gamma_k = \nu_{\theta,k} \cdot \mathrm{E}\left[\left(\frac{\partial \ln p_{\theta,k}(z_k)}{\partial \nu_{\theta,k}} \right)^2 \right]$$
$$= \nu_{\theta,k} \cdot \mathrm{E}\left[\left(\frac{\partial}{\partial \nu_{\theta,k}} \left(-\nu_{\theta,k} - \ln\left(\sqrt{2\pi}\sigma_k\right) \right.\right.\right.$$
$$\left.\left.\left. + \ln\left[e^{-\left(\frac{z_k-\eta_k}{\sqrt{2}\sigma_k}\right)^2} + \int_0^{\infty} e^{-\left(\frac{z_k-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}} \frac{\sqrt{\frac{\nu_{\theta,k}u}{g}} I_1\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right)}{u} du \right] \right) \right)^2 \right]$$
$$= \nu_{\theta,k} \cdot \mathrm{E}\left[\left(-1 + \frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k \cdot p_{\theta,k}(z_k)} \int_0^{\infty} \frac{e^{-\left(\frac{z_k-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du \right)^2 \right]$$
$$= \nu_{\theta,k} \cdot \left(\mathrm{E}\left[\left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k \cdot p_{\theta,k}(z_k)} \int_0^{\infty} \frac{e^{-\left(\frac{z_k-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du \right)^2 \right] \right.$$
$$\left. -2 \cdot \mathrm{E}\left[\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k \cdot p_{\theta,k}(z_k)} \int_0^{\infty} \frac{e^{-\left(\frac{z_k-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du \right] + 1 \right)$$

$$
\begin{aligned}
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \int_0^{\infty} \frac{e^{-\left(\frac{z-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du \right)^2 dz \right. \\
&\quad \left. -2 \cdot \int_0^{\infty} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\left(\frac{z-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2} dz \right) \frac{e^{-\frac{u}{g}-\nu_{\theta,k}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \int_0^{\infty} \frac{e^{-\left(\frac{z-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du \right)^2 dz \right. \\
&\quad \left. -2 \cdot \int_0^{\infty} \frac{e^{-\frac{u}{g}-\nu_{\theta,k}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \int_0^{\infty} \frac{e^{-\left(\frac{z-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du \right)^2 dz \right. \\
&\quad \left. -2 \cdot \int_0^{\infty} \left(\frac{\partial}{\partial \nu_{\theta,k}} p_{\nu_{\theta,k},g}(u) + p_{\nu_{\theta,k},g}(u) \right) du + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \int_0^{\infty} \frac{e^{-\left(\frac{z-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du \right)^2 dz \right. \\
&\quad \left. -2 \cdot \left[\frac{\partial}{\partial \nu_{\theta,k}} \left(\int_0^{\infty} p_{\nu_{\theta,k},g}(u) du \right) + \int_0^{\infty} p_{\nu_{\theta,k},g}(u) du \right] + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \int_0^{\infty} \frac{e^{-\left(\frac{z-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du \right)^2 dz \right. \\
&\quad \left. -2 \cdot \left[\frac{\partial}{\partial \nu_{\theta,k}} \left(1 - e^{-\nu_{\theta,k}}\right) + 1 - e^{-\nu_{\theta,k}} \right] + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \int_0^{\infty} \frac{e^{-\left(\frac{z-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du \right)^2 dz \right. \\
&\quad \left. -2 \cdot 1 + 1 \right) \\
&= \nu_{\theta,k} \cdot \left(\int_{-\infty}^{\infty} \frac{1}{p_{\theta,k}(z)} \left(\frac{e^{-\nu_{\theta,k}}}{\sqrt{2\pi}\sigma_k} \int_0^{\infty} \frac{e^{-\left(\frac{z-u-\eta_k}{\sqrt{2}\sigma_k}\right)^2 - \frac{u}{g}}}{g} I_0\left(2\sqrt{\frac{\nu_{\theta,k}u}{g}}\right) du \right)^2 dz - 1 \right),
\end{aligned}
$$

where $p_{\nu_{\theta,k},g}$ in the intermediate steps is the probability density function of Eq. (15.28) with μ_k replaced by $\nu_{\theta,k}$, which integrates to $1-e^{-\nu_{\theta,k}}$ from $u=0$ to infinity. In the calculations, we have made use of the identity $\frac{\partial I_1(x)}{\partial x} = I_0(x) - \frac{1}{x}I_1(x)$, $x \in \mathbb{R}$, between I_0 and I_1, the zeroth and first order modified Bessel functions of the first kind, respectively.

Notes

Chapter B

1. For a mathematically rigorous definition of the delta function, we refer the reader to the literature (e.g., [24]).

Chapter E

1. A derivation of the Fisher information matrix corresponding to an image acquired under the fundamental data model is provided in [85], where it is presented as a proof of a theorem. See also [107].
2. In writing out the expression for $\frac{\partial \mathcal{L}\left(\theta|w_1,\ldots,w_{N_0}\right)}{\partial \theta}$, we equate the quantity $\frac{\partial}{\partial \theta}\int_{t_0}^{t}\Lambda_\theta(\tau)d\tau$ with the quantity $\int_{t_0}^{t}\frac{\partial \Lambda_\theta(\tau)}{\partial \theta}d\tau$. The validity of this equality is explained in [85].
3. For more details on the probability density of an arrival time of a Poisson process given a positive number of arrivals, see [107].
4. For a formal statement and proof of the result $\int_C \frac{\partial f_{\theta,\tau}(r)}{\partial \theta}dr = \frac{\partial}{\partial \theta}\int_C f_{\theta,\tau}(r)dr = \frac{\partial}{\partial \theta}1 = 0$, see [85].

Chapter G

1. A more general geometric distribution, which does define a probability for the loss of an input electron, has been used to derive a model for electron multiplication in [30].
2. To learn more about the use of probability generating functions, including those of linear fractional form, in the analysis of branching processes, see, e.g., [45, 7, 42].
3. All Bessel function identities used in this book can be found in, e.g., [3].
4. The derivation of the exponential multiplication-based probability density function in Section G.2 is founded on theoretical results presented in [45]. An analogous derivation that uses a more general exponential distribution, which allows for the possibility of particle loss during multiplication, can be found in [30].

Exercises

Chapter A

1. Let X_1 and X_2 be two independent Poisson random variables with parameters λ_1 and λ_2. Show that $X := X_1 + X_2$ is a Poisson random variable with parameter $\lambda := \lambda_1 + \lambda_2$.
2. Let X_1 and X_2 be two independent Poisson random variables with parameters λ_1 and λ_2. Show that $X := X_1 - X_2$ is not a Poisson random variable.
3. If X is a Poisson random variable and c is a constant, show that cX is a Poisson random variable only if $c = 1$.
4. Use a computational software program to plot the Poisson probability distribution p_λ with λ values of 0.01, 1, 10, 100, and 1000.
5. If X is a Poisson random variable with parameter $\lambda = 100$, calculate the probability that $X > 100$, $X > 120$, $X > 150$, $X > 200$, and $X > 1000$. Also calculate the probability that $X < 100$, $X < 80$, $X < 50$, and $X < 10$.
6. Use a computational software program to simulate 1000 random samples of a Poisson random variable with parameter $\lambda = 100$. How many of these random samples are between 80 and 120?
7. Verify that the Gaussian probability distribution function is indeed a probability density function, i.e., verify that
   (a) $p_{\eta,\sigma^2}(x) \geq 0$ for all $x \in \mathbb{R}$, and
   (b) $\int_{-\infty}^{\infty} p_{\eta,\sigma^2}(x)dx = 1$.
8. Let X_1 and X_2 be two independent Gaussian random variables with means η_1 and η_2 and variances σ_1^2 and σ_2^2. Let $a, b, a_1, a_2 \in \mathbb{R}$.
   (a) Show that $X := aX_1 + b$ is a Gaussian random variable with mean $a\eta_1 + b$ and variance $a^2\sigma_1^2$.
   (b) Show that $X := a_1X_1 + a_2X_2$ is a Gaussian random variable with mean $a_1\eta_1 + a_2\eta_2$ and variance $a_1^2\sigma_1^2 + a_2^2\sigma_2^2$.
9. Use a computational software program to plot the Gaussian probability density with parameters
   (a) $\eta = 0$ and $\sigma^2 = 0.1, 0.2, 1, 10, 100$,
   (b) $\eta = 5$ and $\sigma^2 = 0.1, 0.2, 1, 10, 100$,
   (c) $\eta = -15$ and $\sigma^2 = 0.1, 0.2, 1, 10, 100$.
10. Verify analytically that the mean and variance of a Gaussian random variable are indeed η and σ^2, respectively, if the probability density function is given by

$$p_{\eta,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2\sigma^2}(x-\eta)^2}, \quad x \in \mathbb{R}.$$

11. Let X be a Gaussian random variable with mean η and variance σ^2. Show that $\frac{X-\eta}{\sigma}$ is a standard Gaussian random variable, i.e., it has mean 0 and variance 1.
12. Consider three Poisson random variables X_1, X_2, and X_3 with parameters $\lambda_1 = 10$, $\lambda_2 = 100$, and $\lambda_3 = 1000$. Using a computational software program, for each case investigate whether the corresponding Poisson probability distribution function can be well approximated by the probability distribution function of a Gaussian random variable. If a good approximation is possible, what parameter values for the Gaussian probability distribution function provide the good approximation?