## Chapter 8

## Emission Imaging

ch, emis

## Contents

8.1 Introduction (s,emis,intro) ..... 8.1
8.2 Emission process (s,emis, proc) ..... 8.2
8.2.1 Radiotracer spatial distribution ..... 8.2
8.2.2 The density estimation problem (s,emis,dens) ..... 8.3
8.2.3 Direct density estimation methods ..... 8.3
8.2.4 Poisson $N$ (s,emis,poisN) ..... 8.4
8.2.5 Poisson spatial point processes ..... 8.5
8.3 Radioactive decay statistics (s,emis,decay) ..... 8.6
8.3.1 Statistics of an ideal decay counting process ..... 8.7
8.3.2 Properties of decay counting processes ..... 8.8
8.3.3 Precision of source activity estimates ..... 8.9
8.4 Ideal counting detectors (s,emis,detect) ..... 8.10
8.4.1 Statistics of ideal detector units (binned mode) ..... 8.11
8.4.2 Background events (s,emis,ri) ..... 8.13
8.4.3 Compton scattered photons (s,emis,scatter) ..... 8.13
8.4.4 Poisson log-likelihood for binned data (s,emis,like) ..... 8.14
8.5 List-mode likelihood (s,emis,list) ..... 8.14
8.5.1 Static object ..... 8.16
8.5.2 Object discretization ..... 8.17
8.5.3 Binned-mode data revisited ..... 8.17
8.5.4 Sensitivity calculations (s,emis,sens) ..... 8.18
8.5.5 Disk detector: 2D sensitivity and position distribution (s,emis,disk) ..... 8.19
8.6 PET-specific topics ..... 8.22
8.6.1 Randoms-precorrected PET scans (s,emis,randoms) ..... 8.22
8.6.2 Time-of-flight PET (s,emis,tof) ..... 8.23
8.7 Summary (s,emis,summ) ..... 8.23
8.8 Problems (s,emis,prob) ..... 8.23
8.9 Appendix A: 2D gamma camera (s,emis,spect2) ..... 8.24
8.10 Bibliography ..... 8.24

### 8.1 Introduction (s,emis, intro)

This chapter describes statistical models for emission imaging, also known in the medical field as nuclear medicine or nuclear imaging or radiotracer imaging. The presentation focuses on positron emission tomography (PET) and single photon emission computed tomography (SPECT), but the methods are also useful in non-medical applications such as imaging radioactive waste drums, [1] and detecting nuclear materials [2]. The primary purpose of this chapter
is to provide a rigorous development of the Poisson log-likelihood model (8.4.12) that is used nearly ubiquitously in statistical methods for PET and SPECT image reconstruction, articulating what quantities are imaged in emission imaging.

### 8.2 Emission process (s,emis,proc)

The first goal in a statistical treatment of emission imaging is to define precisely the quantity of interest. This section argues that, in emission imaging, we would like to form images of the radiotracer distribution. Ultimately we will parameterize this quantity so that we can apply parametric estimation methods (although nonparametric estimation methods are also of interest). Because estimation methods are built upon statistical models, we define the radiotracer distribution in a statistical framework.

### 8.2.1 Radiotracer spatial distribution

At time $t=0$, the patient is injected with (or inhales) a radiotracer, containing a very large number ${ }^{1}, N$, of metastable atoms of some radionuclide family. We refer to these as tracer atoms ${ }^{2}$. Each tracer atom is a part of some compound or molecule of physiological interest ${ }^{3}$ such as ${ }^{11} \mathrm{C}$ glucose. The flow of blood, in conjunction with other physiological processes, distributes the finite number of injected tracer atoms (and associated molecules) throughout the body. At some time instant, say $t_{0}$, after the onset of injection, these $N$ tracer atoms are located in various distinct positions within the body. Denote these positions $\overrightarrow{\mathrm{X}}_{1}\left(t_{0}\right), \overrightarrow{\mathrm{X}}_{2}\left(t_{0}\right), \ldots, \overrightarrow{\mathrm{X}}_{N}\left(t_{0}\right)$, where $\overrightarrow{\mathrm{X}}_{k}(t)$ is the three-tuple describing the spatial location (with respect to some fixed coordinate system) of the $k$ th tracer atom at time $t$.

An ideal imaging instrument would have infinitesimal spatial resolution, $100 \%$ sensitivity, and would require an infinitesimal scan duration. Such an ideal instrument would in principle provide a list of the $N$ spatial locations of the tracer atoms at any time $t_{0}$ of interest, or at many such times if dynamic processes are of interest.

Even if a hypothetical system could provide the list $\overrightarrow{\mathrm{X}}_{1}\left(t_{0}\right), \overrightarrow{\mathrm{X}}_{2}\left(t_{0}\right), \ldots, \overrightarrow{\mathrm{X}}_{N}\left(t_{0}\right)$, this list would not constitute the ultimate description of the behavior of the radiotracer (as influenced by the patient's physiology). Imagine that the patient were scanned again under "essentially identical" conditions, meaning that the same quantity of radiotracer is injected, and that the patient's physiological state is the same as in the first scan. In this second scan, the individual tracer atoms would not occupy exactly the same spatial locations as in the first scan at a time $t_{0}$ after onset of injection, because there are many unpredictable phenomena that influence the trajectory of a given tracer atom, such as turbulent blood flow and random diffusion (Brownian motion). If the imaging study were performed repeatedly under essentially identical circumstances, there would be certain regions in the body which would tend to have more tracer atoms at time $t_{0}$, and other regions that have fewer, but the particular locations of the $N$ tracer atoms in each replication would be random variables. The distribution of these random variables is the quantity of most interest, rather than the particular random values that they take on a given realization (scan), even if they could be determined [4]. So even if a hypothetical scanner could provide a list of the exact locations of all injected tracer atoms, there would still be a data-processing problem: estimating the general radiotracer distribution from these spatial locations.

Consider a hypothetical scenario where a single tracer atom (and molecule) is injected at time $t=0$. At some time $t_{0}>0$ after the injection, that tracer atom could, in principle, be located anywhere in the body. But certain locations in the body are more likely than other locations, depending on the patient's physiology, anatomy, and the radiotracer properties.

Define $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$ to be the probability density function (pdf) for the location of a single tracer atom at time $t$ after injection. When multiple tracer atoms are injected, we make the following assumption.

Assumption 8.2.1 The spatial locations of individual tracer atoms at any time $t \geq 0$ are statistically independent random variables that are all identically distributed according to a pdf $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$.

This independence assumption should be very reasonable when trace quantities of radiotracer are injected. If a very large quantity of a radiotracer such as a neuroreceptor agent were injected, then the first wave of tracer atoms (and

[^0]molecules) to reach the brain could occupy all or most available receptor sites, denying access to later arriving tracer atoms. This would lead to a statistical dependence between the tracer atom locations. Rarely are such large quantities injected, so this i.i.d. assumption is a reasonable starting point.

Thus, the quantity of interest in radiotracer imaging is the pdf $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$, which has units inverse volume, i.e., tracer atoms per $\mathrm{cm}^{3}$. This pdf reflects the mean local "concentration" of radiotracer. We refer to this quantity as the radiotracer distribution function or just radiotracer distribution, even though "density function" might be more appropriate. Regions where the radiotracer distribution has relatively larger values are the "hot" regions of the object where tracer atoms are more likely to be located.

An ideal imaging instrument and data processing method would provide $p_{t}(\overrightarrow{\mathrm{x}})$ for all spatial locations $\overrightarrow{\mathrm{x}}$ and for all times $t$ spanning a large time interval so that one can investigate dynamic properties of the physiology. In practice the instrument usually has a finite field of view (FOV) that is much smaller than the patient size, limiting the range of $\vec{x}$ examined. In typical emission imaging instruments the scan time interval is split into several time segments, and at best one can measure information related to the time integral of the radiotracer distribution over each time segment i.e.,

$$
\int_{t_{1}}^{t_{2}} \mathrm{p}_{t}(\overrightarrow{\mathrm{x}}) \mathrm{d} t
$$

Alternatively, some emission tomography systems collect "list mode" events [5-7] rather than "binned mode" acquisitions. In principle, list-mode scans can provide something closer to continuous time information. We describe statistical models for both types of scans in subsequent sections.

### 8.2.2 The density estimation problem (s,emis,dens)

As an example, Fig. 8.2.1 illustrates the distinction between the underlying radiotracer distribution $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$ and the particular spatial locations of tracer atoms that occur in one experiment that an "ideal" scanner would acquire. It is difficult to see the ramp trend in the underlying density within the upper region. We really want a picture like on the left of Fig. 8.2.1 rather than on its right.


Figure 8.2.1: Left: an example of a radiotracer distribution function $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$. Right: collection of $N=2000$ tracer atom positions $\overrightarrow{\mathrm{X}}_{1}(t), \ldots, \overrightarrow{\mathrm{X}}_{2000}(t)$ (shown with dots) at some time $t$. Note the great dissimilarity between this representation and the distribution.

### 8.2.3 Direct density estimation methods

Hypothetically, if we could observe directly the spatial locations $\left\{\overrightarrow{\mathrm{X}}_{k}(t)\right\}_{k=1}^{N}$ at some time $t$, then a simple approach to estimating the radiotracer distribution $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$ would be to discretize the spatial domain into small cubic voxels, and simply count how many of the $\overrightarrow{\mathrm{X}}_{k}(t)$ values fall within each voxel. Let $\mathcal{B}_{j}$ denote the spatial region of the $j$ th
voxel. The number of tracer atoms in the $j$ th voxel at time $t$ is

$$
N_{j}(t) \triangleq \sum_{k=1}^{N} \mathbb{I}_{\left\{\overrightarrow{\mathrm{x}}_{k}(t) \in \mathcal{B}_{j}\right\}}
$$

where $\mathbb{I}_{\{\cdot\}}$ denotes the indicator function. The histogram density estimate is then

$$
\hat{\mathrm{p}}_{t}(\overrightarrow{\mathrm{x}}) \triangleq \sum_{j} \frac{N_{j}(t)}{N} \frac{\mathbb{I}_{\left\{\overrightarrow{\mathrm{x}} \in \mathcal{B}_{j}\right\}}}{\operatorname{vol}\left(\mathcal{B}_{j}\right)},
$$

where $\operatorname{vol}\left(\mathcal{B}_{j}\right)$ denotes the volume of the $j$ th voxel (included so that density estimate integrates to unity).
Histogram density estimators are simple, but suboptimal [8]. A popular method for density estimation is the kernel density estimator defined by

$$
\begin{equation*}
\hat{\mathrm{p}}_{t}(\overrightarrow{\mathrm{x}}) \triangleq \frac{1}{N} \sum_{k=1}^{N} g\left(\overrightarrow{\mathrm{x}}-\overrightarrow{\mathrm{X}}_{k}(t)\right) \tag{8.2.1}
\end{equation*}
$$

where $g(\overrightarrow{\mathrm{x}})$ is some (typically smooth) function called the kernel, e.g., a gaussian function. Basically a "blob" is centered at each "point" $\overrightarrow{\mathrm{X}}_{k}(t)$ and these blobs are summed to form an estimate.

Example 8.2.2 Fig. 8.2.2 illustrates the gaussian kernel density estimator (8.2.1), where $g(\vec{x})$ is a $2 D$ gaussian kernel with $F H W M=0.5$, computed from the data shown in Fig. 8.2.1. Fig. 8.2.3 shows a horizontal profile through the true density $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$ and the two preceding density estimates. The ramp structure is much easier to identify in the kernel density estimate.

The mean of a kernel density estimator (8.2.1) is a blurred version of the radiotracer distribution $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$ :

$$
\mathrm{E}\left[\hat{\mathrm{p}}_{t}(\overrightarrow{\mathrm{x}})\right]=\int g\left(\overrightarrow{\mathrm{x}}-\overrightarrow{\mathrm{x}}^{\prime}\right) \mathrm{p}_{t}\left(\overrightarrow{\mathrm{x}}^{\prime}\right) \mathrm{d} \overrightarrow{\mathrm{x}}^{\prime}=\mathrm{p}_{t}(\overrightarrow{\mathrm{x}}) * * g(\overrightarrow{\mathrm{x}})
$$

where * denotes 2D (or 3D) convolution. Choosing a narrower kernel $g(\overrightarrow{\mathrm{x}})$ induces less blur (improving spatial resolution), but at the expense of increased variance (noise).

There are also penalized likelihood methods for density estimation, e.g., [8-11].


Figure 8.2.2: Estimate of the radiotracer distribution $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$ formed by binning (left) and formed by a gaussian kernel density estimator (right).

### 8.2.4 Poisson $N$ (s,emis,poisN)

In practice, one cannot control the exact number of tracer atoms administered, so it is appropriate to consider the actual number $N$ to be a random variable. (If the radionuclide production and radiotracer synthesis processes were repeated under essentially identical conditions, the number of tracer atoms administered would vary between replications.)


Figure 8.2.3: Horizontal profiles at $x_{2}=0.5$ through density estimates in Fig. 8.2.2 and through the true density shown in Fig. 8.2.1

Assumption 8.2.3 The number of administered tracer atoms $N$ is a Poisson distributed random variable with some mean

$$
\mu_{N} \triangleq \mathrm{E}[N]=\sum_{n=0}^{\infty} n \mathrm{P}\{N=n\}
$$

Furthermore, $N$ is statistically independent of the tracer atom locations $\left\{\overrightarrow{\mathrm{X}}_{k}(t)\right\}$.
Assumption 8.2.3 is reasonable because the tracer atoms are activated by various nuclear processes involving enormous numbers of atoms and small probabilities (e.g., in a cyclotron target), so the Poisson approximation to the Binomial distribution is well justified. There are other methods for formulating the problem that also lead to the same Poisson spatial point-process model given below.

### 8.2.5 Poisson spatial point processes

Under the assumptions that a Poisson number of tracer atoms are administered and that these tracer atoms occupy independent and identically distributed spatial locations at some time $t$ after injection, the collection of such spatial locations is called a Poisson spatial point process [12]. The following property of such processes is relevant to our development. Let $N_{t}(\mathcal{B})$ denote the number of tracer atoms that have spatial locations in set $\mathcal{B} \subset \mathbb{R}^{3}$ at some time $t$ after injection. Specifically:

$$
N_{t}(\mathcal{B})=\sum_{k=1}^{N} \mathbb{I}_{\left\{\overrightarrow{\mathrm{X}}_{k}(t) \in \mathcal{B}\right\}}
$$

Lemma 8.2.4 For a Poisson spatial point process, $N_{t}(\mathcal{B})$ is a Poisson random variable with mean

$$
\mathrm{E}\left[N_{t}(\mathcal{B})\right]=\mathrm{E}[N] \int_{\mathcal{B}} \mathrm{p}_{t}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}
$$

Proof:
Using iterated expectation: and the assumption of independence in Assumption 8.2.3:

$$
\begin{aligned}
\mathrm{E}\left[N_{t}(\mathcal{B})\right] & =\mathrm{E}_{N}\left[\mathrm{E}\left[N_{t}(\mathcal{B}) \mid N\right]\right]=\mathrm{E}_{N}\left[\mathrm{E}\left[\sum_{k=1}^{N} \mathbb{I}_{\left\{\overrightarrow{\mathrm{x}}_{k}(t) \in \mathcal{B}\right\}} \mid N\right]\right] \\
& =\mathrm{E}_{N}\left[N \mathrm{P}\left\{\overrightarrow{\mathrm{X}}_{k}(t) \in \mathcal{B}\right\}\right]=\mathrm{E}[N] \mathrm{P}\left\{\overrightarrow{\mathrm{X}}_{k}(t) \in \mathcal{B}\right\}
\end{aligned}
$$

The remainder of the argument is a case of Bernoulli thinning described in §31.3.2.
In general we are interested in many regions (such as a collection of voxels), so we explore their statistics next.
Lemma 8.2.5 If $\mathcal{B}_{1}, \mathcal{B}_{2}, \ldots, \mathcal{B}_{M}$ are disjoint subsets of $\mathbb{R}^{3}$, then conditioned on $N=n$, the random variables $N_{t}\left(\mathcal{B}_{1}\right), N_{t}\left(\mathcal{B}_{2}\right), \ldots, N_{t}\left(\mathcal{B}_{M}\right)$ have a multinomial distribution with associated probabilities $p_{m}=\int_{\mathcal{B}_{m}} \mathrm{p}_{t}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}$, for $m=1, \ldots, M$, i.e.,

$$
\mathrm{P}\left\{N_{t}\left(\mathcal{B}_{1}\right)=n_{1}, \ldots, N_{t}\left(\mathcal{B}_{M}\right)=n_{M} \mid N=n\right\}=\binom{n}{n_{1} \ldots n_{M} n_{0}} p_{1}^{n_{1}} \cdots p_{M}^{n_{M}} p_{0}^{n_{0}}
$$

where $p_{0} \triangleq 1-p_{1}-\cdots-p_{M}$ and $n_{0} \triangleq n-n_{1}-\cdots-n_{M}$.
Proof:
This follows from the independence of the $\overrightarrow{\mathrm{X}}_{k}(t)$ variables and the definition of a multinomial distribution.

Lemma 8.2.6 If $\mathcal{B}_{1}, \mathcal{B}_{2}, \ldots, \mathcal{B}_{M}$ are disjoint subsets of $\mathbb{R}^{3}$, then $N_{t}\left(\mathcal{B}_{1}\right), N_{t}\left(\mathcal{B}_{2}\right), \ldots, N_{t}\left(\mathcal{B}_{M}\right)$ are independent random variables.
Proof:
We consider the case $M=2$, leaving the more general case as an exercise (Problem 8.2).
Let $N_{1}=N_{t}\left(\mathcal{B}_{1}\right)$ and $N_{2}=N_{t}\left(\mathcal{B}_{2}\right)$. For $n_{1} \geq 0$ and $n_{2} \geq 0$,

$$
\begin{aligned}
\mathrm{P}\left\{N_{1}=n_{1}, N_{2}=n_{2}\right\} & =\sum_{n=n_{1}+n_{2}}^{\infty} \mathrm{P}\left\{N_{1}=n_{1}, N_{2}=n_{2} \mid N=n\right\} \mathrm{P}\{N=n\} \\
& =\sum_{n=n_{1}+n_{2}}^{\infty} \frac{n!}{n_{1}!n_{2}!\left(n-n_{1}-n_{2}\right)!} p_{1}^{n_{1}} p_{2}^{n_{2}}\left(1-p_{1}-p_{2}\right)^{n-n_{1}-n_{2}}\left[\mathrm{e}^{-\mu_{N}} \mu_{N}^{n} / n!\right] \\
& =\frac{\left(\mu_{N} p_{1}\right)^{n_{1}}}{n_{1}!} \frac{\left(\mu_{N} p_{2}\right)^{n_{2}}}{n_{2}!} e^{-\mu_{N}} \sum_{k=0}^{\infty} \frac{1}{k!}\left[\left(1-p_{1}-p_{2}\right) \mu_{N}\right]^{k} \\
& =\left[e^{-\mu_{N} p_{1}} \frac{\left(\mu_{N} p_{1}\right)^{n_{1}}}{n_{1}!}\right]\left[e^{-\mu_{N} p_{2}} \frac{\left(\mu_{N} p_{2}\right)^{n_{2}}}{n_{2}!}\right] \\
& =\mathrm{P}\left\{N_{1}=n_{1}\right\} \mathrm{P}\left\{N_{2}=n_{2}\right\},
\end{aligned}
$$

using total probability, Lemma 8.2.5, $k=n-n_{1}-n_{2}$, the Taylor series for $\mathrm{e}^{x}$, and finally Lemma 8.2.4.
In other words, the number of tracer atoms in disjoint regions are independent Poisson random variables. Fig. 8.2.4 illustrates this property.


Figure 8.2.4: Illustration of four realizations of a Poisson spatial point process. The number of points within a fixed ROI is a Poisson random variable, and is independent of the number of points in other regions.

### 8.3 Radioactive decay statistics (s,emis,decay)

The preceding description covers two of the random phenomena in nuclear imaging: the number of injected tracer atoms, and the spatial locations of those tracer atoms. Unfortunately, we do not get to observe the spatial locations of the tracer atoms at all times. We can only observe a tracer atom indirectly through its emitted photon(s) when it decays (and even then only for some of the tracer atoms). By "decay" of a tracer atom, we mean the time at which
the tracer atom emits a photon (or positron in PET) from its nucleus. The time at which any given tracer atom decays is unpredictable, hence it is a random variable. (See [13, p. 267] for a nice introduction to decay.) For emission tomography or nuclear imaging, we need to examine the statistics of the decay of radioactive materials.

We simplify the analysis by assuming that all decays consist of the emission of a single photon (or positron). One can generalize the analysis to radionuclides having multiple modes of decay [14, p. 27] [15].

Suppose the patient contains $N=n$ tracer atoms at time $t=0$. For $k=1, \ldots, n$, let $T_{k} \geq 0$ denote the random variable that denotes the time at which the $k$ th tracer atom decays, e.g., emits a $\gamma$ photon.
Assumption 8.3.1 The $T_{k}$ random variables are independent, $k=1,2, \ldots, n$, and are independent of the $\overrightarrow{\mathrm{X}}_{k}(\cdot)$ values.

The assumption of independence is reasonable physically except in cases of stimulated emissions, or chain disintegrations [16] (sometimes called the parent-daughter problem [17]). In the usual circumstances of interest, the decay of a given nucleus is not "affected to any significant extent by events occurring outside the nucleus" [14, p. 22].

Assumption 8.3.2 Each $T_{k}$ has an exponential distribution with mean $\mu_{T}$.
The cumulative distribution function of the exponential distribution is given by

$$
\begin{equation*}
\mathrm{P}\left\{T_{k} \leq t\right\}=1-\mathrm{e}^{-t / \mu_{T}}, \quad t \geq 0 \tag{8.3.1}
\end{equation*}
$$

and is shown in Fig. 8.3.1. The half life $t_{1 / 2}$ of a radionuclide is the time at which $\mathrm{P}\left\{T_{k} \leq t_{1 / 2}\right\}=1 / 2$. Solving yields

$$
\begin{equation*}
t_{1 / 2}=\mu_{T} \ln 2 \tag{8.3.2}
\end{equation*}
$$



Figure 8.3.1: Cumulative distribution function of the exponential distribution for radioactive decay.
The exponential distribution (8.3.1) is consistent with empirical observations [18]. Also, the exponential distribution is the unique distribution that is consistent with the assumption that "the probability of decay of an atom is independent of the age of that atom" [18, p. 470]. In statistical terms, this characteristic is called the memoryless property, and can be expressed mathematically as follows:

$$
\mathrm{P}\left\{T_{k} \leq t \mid T_{k} \geq t_{0}\right\}=\mathrm{P}\left\{T_{k} \leq t-t_{0}\right\} \text { for } t \geq t_{0}
$$

See [19, p. 349] for further discussion of the "at random" property.
Assumption 8.3 .2 is appropriate in typical cases that use only a single radioisotope. If multiple isotopes are present with different half lives, then the conditional distribution of each $T_{k}$ given the particular isotope is again exponential, but the overall distribution of the $T_{k}$ values is a mixture of exponential distributions with mixture proportions that depend on the relative abundance of each isotope in the object. (See Problem 8.4.)

### 8.3.1 Statistics of an ideal decay counting process

We now combine the statistical model for the radiotracer distribution (Assumption 8.2.1) with the statistical model for decay (8.3.1). Let $\mathcal{B} \subseteq \mathbb{R}^{3}$ denote an arbitrary spatial set and let $K_{t}(\mathcal{B})$ denote the random process that counts the number of tracer atoms that have decayed by time $t$, relative to some arbitrary starting time $t=0$, and that were located in the set $\mathcal{B}$ at the time they decayed. By definition: $K_{0}(\mathcal{B})=0$ and $K_{0}(\mathcal{B}) \leq N$. We can express this counting process as

$$
K_{t}(\mathcal{B})=\sum_{k=1}^{N} Z_{k}(t ; \mathcal{B}), \text { where } Z_{k}(t ; \mathcal{B})= \begin{cases}1, & T_{k} \leq t \text { and } \overrightarrow{\mathrm{X}}_{k}\left(T_{k}\right) \in \mathcal{B}  \tag{8.3.3}\\ 0, & \text { otherwise }\end{cases}
$$

The following theorem describes the statistics of this counting process in terms of the following emission rate density function:

$$
\begin{equation*}
\lambda_{t}(\overrightarrow{\mathrm{x}}) \triangleq \mu_{N} \frac{\mathrm{e}^{-t / \mu_{T}}}{\mu_{T}} \mathrm{p}_{t}(\overrightarrow{\mathrm{x}}) \tag{8.3.4}
\end{equation*}
$$

The units of $\lambda_{t}(\overrightarrow{\mathrm{x}})$ are "counts" (i.e., decays) per unit time per unit volume. Because $\int \lambda_{t}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}=\mu_{N} \mathrm{e}^{-t / \mu_{T}} / \mu_{T}$, we can also express the pdf $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$ in terms of $\lambda_{t}(\overrightarrow{\mathrm{x}})$ as follows

$$
\begin{equation*}
\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})=\frac{\lambda_{t}(\overrightarrow{\mathrm{x}})}{\int \lambda_{t}\left(\overrightarrow{\mathrm{x}}^{\prime}\right) \mathrm{d} \overrightarrow{\mathrm{x}}^{\prime}} \tag{8.3.5}
\end{equation*}
$$

This expression is useful for formulating log-likelihoods in terms of $\lambda_{t}(\vec{x})$.
Theorem 8.3.3 Under Assumptions 8.2.1-8.3.2, $K_{t}(\mathcal{B})$ is a Poisson counting process with mean

$$
\begin{equation*}
\mathrm{E}\left[K_{t}(\mathcal{B})\right]=\int_{0}^{t} \int_{\mathcal{B}} \lambda_{s}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}} \mathrm{~d} s \tag{8.3.6}
\end{equation*}
$$

where $\lambda_{t}(\overrightarrow{\mathrm{x}})$ was defined in (8.3.4). Note that $K_{t}(\mathcal{B})$ is an inhomogeneous Poisson process because its rate varies with time (due to decay).
Proof:
It is obvious from (8.3.3) that $K_{t}(\mathcal{B})$ is a counting process. Because the $T_{k}$ values and $\overrightarrow{\mathrm{X}}_{k}(\cdot)$ values are independent by Assumption 8.3.1, $K_{t}(\mathcal{B})$ has independent increments. Thus it suffices to show that the marginal distribution of $K_{t}(\mathcal{B})$ is Poisson with mean (8.3.6). For $t \geq 0$ define

$$
\begin{align*}
p_{t} & =\mathrm{P}\left\{Z_{k}(t ; \mathcal{B})=1\right\}=\mathrm{P}\left\{T_{k} \leq t, \overrightarrow{\mathrm{X}}_{k}\left(T_{k}\right) \in \mathcal{B}\right\}=\int_{0}^{\infty} \mathrm{P}\left\{T_{k} \leq t, \overrightarrow{\mathrm{X}}_{k}\left(T_{k}\right) \in \mathcal{B} \mid T_{k}=s\right\} \mathrm{p}_{T_{k}}(s) \mathrm{d} s  \tag{8.3.7}\\
& =\int_{0}^{t} \mathrm{P}\left\{\overrightarrow{\mathrm{X}}_{k}(s) \in \mathcal{B} \mid T_{k}=s\right\} \mathrm{p}_{T_{k}}(s) \mathrm{d} s=\int_{0}^{t}\left[\int_{\mathcal{B}} \mathrm{p}_{s}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}\right] \frac{1}{\mu_{T}} \mathrm{e}^{-s / \mu_{T}} \mathrm{~d} s \tag{8.3.8}
\end{align*}
$$

Because $N$ is Poisson by Assumption 8.2.3, it follows from (8.3.6) and by the Bernoulli thinning discussed in §31.3.2.2 that $K_{t}(\mathcal{B})$ is Poisson with mean $\mathrm{E}[N] \mathrm{P}\left\{Z_{k}(t ; \mathcal{B})=1\right\}$. In particular, its mean is

$$
\mathrm{E}\left[K_{t}(\mathcal{B})\right]=\mu_{N} p_{t}=\frac{\mu_{N}}{\mu_{T}} \int_{0}^{t}\left[\int_{\mathcal{B}} \mathrm{p}_{s}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}\right] \mathrm{e}^{-s / \mu_{T}} \mathrm{~d} s
$$

and its rate is

$$
\frac{\mathrm{d}}{\mathrm{~d} t} \mathrm{E}\left[K_{t}(\mathcal{B})\right]=\int_{\mathcal{B}}\left[\mu_{N} \frac{\mathrm{e}^{-t / \mu_{T}}}{\mu_{T}} \mathrm{p}_{t}(\overrightarrow{\mathrm{x}})\right] \mathrm{d} \overrightarrow{\mathrm{x}}
$$

We define the bracketed quantity to be the emission rate density function $\lambda_{t}(\overrightarrow{\mathrm{x}})$ in (8.3.4).
Importantly, the conclusions of this theorem hold even for dynamic objects, i.e., even for time-varying radiotracer distributions. Now we proceed from analyzing the statistics of one region to considering a collection of regions.

Theorem 8.3.4 Under Assumptions 8.2.1-8.3.2, if $\mathcal{B}_{1}, \mathcal{B}_{2}, \ldots, \mathcal{B}_{M}$ are disjoint subsets of $\mathbb{R}^{3}$, then the random processes $K_{t}\left(\mathcal{B}_{1}\right), K_{t}\left(\mathcal{B}_{2}\right), \ldots, K_{t}\left(\mathcal{B}_{M}\right)$ are independent.
(See Problem 8.3.)
In summary, the preceding two theorems provide a rigorous mathematical justification for the generally held notion that "radioactive decay is governed by Poisson statistics." (See [20, Sec. 11.3] for more Poisson process details.) However, these theorems alone do not ensure that the measurements made by emission imaging systems have Poisson distributions. That conclusion requires additional assumptions discussed in §8.4.

### 8.3.2 Properties of decay counting processes

Because all tracer atoms must decay eventually (with probability one), one expects that

$$
\begin{equation*}
\lim _{t \rightarrow \infty} \mathrm{P}\left\{K_{t}\left(\mathbb{R}^{3}\right)=n \mid N=n\right\}=1 \tag{8.3.9}
\end{equation*}
$$

Note that for $\mathcal{B}=\mathbb{R}^{3}$, from (8.3.8):

$$
p_{t}=\int_{0}^{t}\left[\int_{\mathbb{R}^{3}} \mathrm{p}_{s}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}\right] \frac{1}{\mu_{T}} \mathrm{e}^{-s / \mu_{T}} \mathrm{~d} s=\int_{0}^{t} \frac{1}{\mu_{T}} \mathrm{e}^{-s / \mu_{T}} \mathrm{~d} s=1-\mathrm{e}^{-t / \mu_{T}}
$$

so from (31.3.3) we can confirm (8.3.9) as follows:

$$
\mathrm{P}\left\{K_{t}\left(\mathbb{R}^{3}\right)=n \mid N=n\right\}=\binom{n}{n} p_{t}^{n}\left(1-p_{t}\right)^{n-n}=p_{t}^{n}=\left(1-\mathrm{e}^{-t / \mu_{T}}\right)^{n} \rightarrow 1 \text { as } t \rightarrow \infty
$$

To count all decays, we examine $K_{t}\left(\mathbb{R}^{3}\right)$, which is Poisson distributed with mean

$$
\mathrm{E}\left[K_{t}\left(\mathbb{R}^{3}\right)\right]=\mu_{N} p_{t}=\mu_{N}\left(1-\mathrm{e}^{-t / \mu_{T}}\right)
$$

The rate of this process, defined as $\frac{\mathrm{d}}{\mathrm{d} t} \mathrm{E}\left[K_{t}\left(\mathbb{R}^{3}\right)\right]$, varies with time. However, by the Taylor series, $1-\mathrm{e}^{-x} \approx x$ for small $x$, for small $t$ we have

$$
\mathrm{E}\left[K_{t}\left(\mathbb{R}^{3}\right)\right] \approx \frac{\mu_{N}}{\mu_{T}} t
$$

so $K_{t}\left(\mathbb{R}^{3}\right)$ behaves very much like a homogeneous Poisson process for $t \ll \mu_{T}$, with rate $\lambda=\mu_{N} / \mu_{T}$, known as the activity of the source. Furthermore, for small $t$ :

$$
\mathrm{E}\left[K_{t}\left(\mathbb{R}^{3}\right) \mid N=n\right] \approx \frac{n}{\mu_{T}} t .
$$

An alternative to the criterion $\mathrm{P}\left\{T_{k} \leq t_{1 / 2}\right\}=1 / 2$ used in deriving the half life (8.3.2) is to define $t_{1 / 2}$ to be the time at which $\mathrm{E}\left[K_{t}\left(\mathbb{R}^{3}\right) \mid N=n\right]=\frac{1}{2} n$. Solving for $t$ again yields (8.3.2). Alternatively, $t_{1 / 2}$ is the time at which the rate of decays $r_{t}$ is half of initial rate $r_{0}$.

$$
r_{t}=\frac{\mathrm{d}}{\mathrm{~d} t} \mathrm{E}\left[K_{t}\left(\mathbb{R}^{3}\right) \mid N=n\right]=\frac{\mathrm{d}}{\mathrm{~d} t} n\left(1-\mathrm{e}^{-t / \mu_{T}}\right)=\frac{n}{\mu_{T}} \mathrm{e}^{-t / \mu_{T}} .
$$

Solving $r_{t_{1 / 2}}=\frac{1}{2} r_{0}$ again yields the same expression for $t_{1 / 2}$.

### 8.3.3 Precision of source activity estimates

The fact that the variance of a Poisson random variable equals its mean sometimes creates confusion about whether a "larger variance" is better or not. Given a source (such as tissue sample extracted from an animal previously injected with a radiotracer) where $\lambda$ is unknown, we can measure its radioactivity in a well counter for time $t_{0}$ and observe $K\left(t_{0}\right)$ (the total number of counts in $t_{0}$ seconds), where $K\left(t_{0}\right) \sim \operatorname{Poisson}\left\{\lambda t_{0}\right\}$. Then the (maximum-likelihood) estimator ( $c f$. Problem 1.11) for rate $\lambda$ is

$$
\hat{\lambda}=\frac{K\left(t_{0}\right)}{t_{0}}
$$

This ML estimator is unbiased:

$$
\mathrm{E}[\hat{\lambda}]=\mathrm{E}\left[\frac{K\left(t_{0}\right)}{t_{0}}\right]=\frac{\mathrm{E}\left[K\left(t_{0}\right)\right]}{t_{0}}=\frac{\lambda t_{0}}{t_{0}}=\lambda
$$

Furthermore, the estimator variance is

$$
\operatorname{Var}\{\hat{\lambda}\}=\operatorname{Var}\left\{\frac{K\left(t_{0}\right)}{t_{0}}\right\}=\frac{\operatorname{Var}\left\{K\left(t_{0}\right)\right\}}{t_{0}^{2}}=\frac{\lambda t_{0}}{t_{0}^{2}}=\frac{\lambda}{t_{0}}
$$

so as the measurement time $t_{0}$ increases, the variance of $\hat{\lambda}$ decreases, i.e., the precision of the estimate improves. This decrease occurs despite the fact that the variance of the counting process $K\left(t_{0}\right)$ increases with $t_{0}$.

### 8.4 Ideal counting detectors (s,emis,detect)

Nuclear imaging systems record measurements using at least one of two possible acquisition methods: list mode or binned mode. This section focuses on the statistics of binned-mode scans, whereas $\S 8.5$ considers list-mode scans.

A nuclear imaging system operating in binned mode can be modeled as consisting of $n_{\mathrm{d}}$ conceptual detector units, each of which can record certain emitted photons. In our terminology, detector units need not correspond to physical detectors. For example, in a 2D PET system consisting of a ring of $n$ physical detectors, there could be as many as $n_{\mathrm{d}}=n(n-1)$ detector units, each of which corresponds to a pair of physical detectors in electronic coincidence ${ }^{4}$. (Or even more if the system is wobbled [22,23].) In a SPECT system based on a single rotating gamma camera that collects projection views of size $n_{\mathrm{u}} \times n_{\mathrm{v}}$ at each of $n_{\varphi}$ projection angles, there would be $n_{\mathrm{d}}=n_{\mathrm{u}} n_{\mathrm{v}} n_{\varphi}$ detector units. Systems that bin measurements into multiple energy windows have even more detector units [24-26], as do systems that bin data temporally using some type of gating mechanism such as phase of respiration.

Usually imaging systems are configured deliberately so that different detector units are relatively more likely to record photons originating from different spatial locations.

Typically each emitted photon (or pair in PET) that is recorded by the system is assigned (by the electronics and binning circuitry) to a single detector unit, e.g., to a single sinogram bin. We make this an explicit assumption.

Assumption 8.4.1 Each decay results in a recorded count in at most one detector unit.
This assumption is a key component of the argument below that leads to the conclusion that the measurements have Poisson distributions. Systems that distribute fractions of each detected event over multiple detector units, or that increment more than one detector unit for each event would almost certainly have non-Poisson measurements and would need to be analyzed differently.

Let $S_{k} \in\left\{0,1,2, \ldots, n_{\mathrm{d}}\right\}$ be the random variable that denotes the index of the detector unit that records the $k$ th decay. If the $k$ th decay goes undetected, assign $S_{k}$ to be zero. Let $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})$ denote the probability that a single isolated decay at spatial location $\vec{x}$ will be detected by the system and recorded (counted) by the $i$ th detector unit, i.e.,

$$
\begin{equation*}
\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})=\mathrm{P}\left\{S_{k}=i \mid \overrightarrow{\mathrm{X}}_{k}\left(T_{k}\right)=\overrightarrow{\mathrm{x}}\right\}, \quad i=1, \ldots, n_{\mathrm{d}} \tag{8.4.1}
\end{equation*}
$$

We refer to $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})$ as the detector unit sensitivity pattern, or simply as the system model. Typically $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}}) \ll 1$. Usually we assume that $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})$ is "known," i.e., it is determined by considering the geometry of the imaging system and the design of the system components such as the collimator. Ideally the detector unit sensitivity patterns $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})$ will include all the physical effects that affect the detection of a single decay, such as scatter, attenuation, detector response, and detector efficiency. In PET, factors that affect spatial resolution (and thus should be modeled in $s_{i}(\overrightarrow{\mathrm{x}})$ for best results) include positron range [27-29], positron non-colinearity, crystal penetration, inter-crystal scatter, crystal identification error. The notation $s_{i}(\overrightarrow{\mathrm{x}})$ is appropriate for a stationary imaging system. See Problem 8.6 for consideration of moving imaging systems, e.g., a rotating SPECT system, where the sensitivity patterns $s_{i}(\overrightarrow{\mathrm{x}}, t)$ vary with time.

Under Assumption 8.4.1, the system sensitivity pattern, defined as the overall probability that a decay at location $\overrightarrow{\mathrm{x}}$ is recorded by some detector unit, is given by:

$$
\begin{equation*}
\mathrm{s}(\overrightarrow{\mathrm{x}}) \triangleq \sum_{i=1}^{n_{\mathrm{d}}} \mathrm{~s}_{i}(\overrightarrow{\mathrm{x}})=1-\mathrm{s}_{0}(\overrightarrow{\mathrm{x}}) \leq 1, \quad \forall \overrightarrow{\mathrm{x}} \tag{8.4.2}
\end{equation*}
$$

When $n_{\mathrm{d}}$ is large, computing $\mathrm{s}(\overrightarrow{\mathrm{x}})$ exactly can be impractical; see §8.5.4.
Example 8.4.2 Appendix $\S 8.9$ derives example $s_{i}(\overrightarrow{\mathrm{x}})$ functions for SPECT imaging using an Anger camera with parallel hole collimation in front of a flat pane of scintillator.

Fig. 8.4.1 shows two examples of such $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})$ patterns. These patterns are somewhat "tube shaped" and this characteristic distinguishes tomography from image restoration problems where the impulse response is usally much more localized [30]. One can also examine $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})$ as a function of $i$ for a fixed spatial location $\overrightarrow{\mathrm{x}}_{0}$, as illustrated in Fig. 8.4.2. Fig. 8.4.1 shows the corresponding overall system sensitivity pattern $\mathrm{s}(\overrightarrow{\mathrm{x}})$ for a $180^{\circ}$ SPECT scan with uniform attenuation. The sensitivity is higher in the portion of the FOV that is closer to the camera.

Next we analyze the statistics of the recorded counts, assuming that the recorded bin for the $k$ th tracer atom's decay depends only on its position when it decays, and is independent of all other tracer atoms.

[^1]

Figure 8.4.1: Illustration of two $s_{i}(\vec{x})$ functions (left, middle) and of a system sensitivity pattern $s(\vec{x})$ (right) for a collimated Anger camera model including uniform attenuation for a $180^{\circ}$ SPECT scan.


Figure 8.4.2: Illustration of $\mathrm{s}_{i}\left(\overrightarrow{\mathrm{x}}_{0}\right)$ vs detector unit index $i$ (right) for a particular spatial location $\overrightarrow{\mathrm{x}}_{0}$ (left). Each $\varphi, r$ pair corresponds to a unique value of $i$.

Assumption 8.4.3 The $S_{k}$ values are statistically independent, and $S_{k}$ is statistically independent of $T_{j}$, for $j \neq k$. Specifically, we assume

$$
\mathrm{P}\left\{S_{1}, \ldots, S_{n} \mid N=n, T_{1}, \ldots, T_{n}, \overrightarrow{\mathrm{X}}_{1}(\cdot), \ldots, \overrightarrow{\mathrm{X}}_{n}(\cdot)\right\}=\prod_{k=1}^{n} \mathrm{P}\left\{S_{k} \mid \overrightarrow{\mathrm{X}}_{k}\left(T_{k}\right)\right\}
$$

This assumption is reasonable when deadtime losses and detector pileup conditions are minimal. For high count rates, the detection probabilities are reduced due to deadtime losses, meaning that $S_{k}$ in fact becomes dependent on the other $T_{k}$ values. Counting statistics in the presence of deadtime were analyzed in [31, 32]. The overall conclusion in those analyses is that the Poisson model is reasonable even in the presence of deadtime, provided that the "first-order" effect of deadtime losses is included in $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})$. Therefore we assume deadtime losses are included in $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})$ hereafter if needed.

Under Assumption 8.4.3, each $S_{k}$ has the following conditional distribution:

$$
\mathrm{P}\left\{S_{k}=i \mid \overrightarrow{\mathrm{X}}_{k}\left(T_{k}\right)=\overrightarrow{\mathrm{x}}\right\}= \begin{cases}\mathrm{s}_{i}(\overrightarrow{\mathrm{x}}), & i=1, \ldots, n_{\mathrm{d}} \\ \mathrm{~s}_{0}(\overrightarrow{\mathrm{x}})=1-\mathrm{s}(\overrightarrow{\mathrm{x}}), & i=0 \\ 0, & \text { otherwise }\end{cases}
$$

### 8.4.1 Statistics of ideal detector units (binned mode)

So far we have analyzed the statistics of unobservable random variables. Now we turn to the statistics of the recorded measurements. Suppose we scan (i.e., record emissions) over the time interval $\left[t_{1}, t_{2}\right]$ for $0<t_{1}<t_{2}$. Let $Y_{i}$ denote the number of events recorded by the $i$ th detector unit during this scan interval. Mathematically:

$$
\begin{equation*}
Y_{i}=\sum_{k=1}^{N} \mathbb{I}_{\left\{S_{k}=i, T_{k} \in\left[t_{1}, t_{2}\right]\right\}} \tag{8.4.3}
\end{equation*}
$$

This is called a binned mode acquisition, because both the spatial position on the detector and the time of decay are quantized. To characterize the distribution of $Y_{i}$, it is helpful to first use (8.3.4) to define the average emission rate density function over the scan:

$$
\begin{equation*}
\lambda(\overrightarrow{\mathrm{x}}) \triangleq \frac{1}{\tau} \int_{t_{1}}^{t_{2}} \lambda_{t}(\overrightarrow{\mathrm{x}}) \mathrm{d} t=\frac{\mu_{N}}{\tau} \int_{t_{1}}^{t_{2}} \mathrm{p}_{t}(\overrightarrow{\mathrm{x}}) \frac{1}{\mu_{T}} \mathrm{e}^{-t / \mu_{T}} \mathrm{~d} t \tag{8.4.4}
\end{equation*}
$$

where the scan duration is $\tau=t_{2}-t_{1}$. Hereafter we refer to $\lambda(\overrightarrow{\mathrm{x}})$ as simply the emission density. It has units counts per unit time per unit volume and reflects the average rate of decays per unit volume during the scan interval. It is a function of the injected dose, $\mu_{N}$, of the radiotracer distribution, $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$, and (in the bracketed exponential terms) of the radionuclide decay during the scan.

If the radiotracer distribution $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$ is static (time invariant) over the time interval $\left[t_{1}, t_{2}\right]$, then the expression (8.4.4) for the emission density $\lambda(\vec{x})$ simplifies to the following:

$$
\begin{equation*}
\lambda(\overrightarrow{\mathrm{x}})=\frac{\mu_{N}}{\tau} \mathrm{p}_{t_{1}}(\overrightarrow{\mathrm{x}})\left[\mathrm{e}^{-t_{1} / \mu_{T}}-\mathrm{e}^{-t_{2} / \mu_{T}}\right] \tag{8.4.5}
\end{equation*}
$$

The following theorem characterizes the distribution of the $Y_{i}$ values in the general case where the emission density need not be static.

Theorem 8.4.4 Under Assumptions 1-6, each $Y_{i}$ has a Poisson distribution with mean

$$
\begin{align*}
\mathrm{E}\left[Y_{i}\right] & =\mathrm{E}[N] \mathrm{P}\left\{S_{k}=i, T_{k} \in\left[t_{1}, t_{2}\right]\right\}  \tag{8.4.6}\\
& =\tau \int \mathrm{s}_{i}(\overrightarrow{\mathrm{x}}) \lambda(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}} \tag{8.4.7}
\end{align*}
$$

Furthermore, the $Y_{i}$ values are statistically independent random variables.
Proof:
Because $Y_{i}$ is a sum of a Poisson number of independent Bernoulli random variables, the fact that $Y_{i}$ is Poisson with mean (8.4.6) follows from §31.3.2.2.

To show that (8.4.7) is the mean, we must evaluate the probability in (8.4.6). By total probability:

$$
\begin{aligned}
\mathrm{P}\left\{S_{k}\right. & \left.=i, T_{k} \in\left[t_{1}, t_{2}\right]\right\} \\
& =\int_{0}^{\infty} \mathrm{P}\left\{S_{k}=i, T_{k} \in\left[t_{1}, t_{2}\right] \mid T_{k}=t\right\} \mathrm{p}_{T_{k}}(t) \mathrm{d} t \\
& =\int_{t_{1}}^{t_{2}} \mathrm{P}\left\{S_{k}=i \mid T_{k}=t\right\} \frac{\mathrm{e}^{-t / \mu_{T}}}{\mu_{T}} \mathrm{~d} t \\
& =\int_{t_{1}}^{t_{2}}\left[\int \mathrm{P}\left\{S_{k}=i \mid T_{k}=t, \overrightarrow{\mathrm{X}}_{k}\left(T_{k}\right)=\overrightarrow{\mathrm{x}}\right\} \mathrm{p}_{t}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}\right] \frac{\mathrm{e}^{-t / \mu_{T}}}{\mu_{T}} \mathrm{~d} t \\
& =\int_{t_{1}}^{t_{2}}\left[\int \mathrm{~s}_{i}(\overrightarrow{\mathrm{x}}) \mathrm{p}_{t}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}\right] \frac{\mathrm{e}^{-t / \mu_{T}}}{\mu_{T}} \mathrm{~d} t \\
& =\frac{\tau}{\mu_{N}} \int \mathrm{~s}_{i}(\overrightarrow{\mathrm{x}})\left[\frac{\mu_{N}}{\tau} \int_{t_{1}}^{t_{2}} \mathrm{p}_{t}(\overrightarrow{\mathrm{x}}) \frac{\mathrm{e}^{-t / \mu_{T}}}{\mu_{T}} \mathrm{~d} t\right] \mathrm{d} \overrightarrow{\mathrm{x}} .
\end{aligned}
$$

The bracketed term is simply $\lambda(\overrightarrow{\mathrm{x}})$ defined in (8.4.4), thus establishing (8.4.7).
Showing that the $Y_{i}$ values are statistically independent is left as an exercise (Problem 8.7).

Corollary 8.4.5 The total number of recorded events

$$
M \triangleq \sum_{i=1}^{n_{\mathrm{d}}} Y_{i}
$$

has mean, using (8.4.2):

$$
\begin{equation*}
\mathrm{E}[M]=\tau \int \mathrm{s}(\overrightarrow{\mathrm{x}}) \lambda(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}=\int_{\tau_{1}}^{\tau_{2}} \int \mathrm{~s}(\overrightarrow{\mathrm{x}}) \lambda_{t}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}} \tag{8.4.8}
\end{equation*}
$$

### 8.4.2 Background events (s,emis,ri)

In practice, nuclear imaging systems record not only photons that are emitted from the object, but also photons from normal background radiation.

Assumption 8.4.6 The recorded background events are all statistically independent Poisson random variables that are independent of the emission process.

Let $\bar{r}_{i}$ denote the mean number of background events recorded by the $i$ th detector unit during the scan. Then by Assumption 8.4.6, the Poisson model (8.4.7) is replaced by the following model:

$$
\begin{equation*}
Y_{i} \sim \operatorname{Poisson}\left\{\bar{y}_{i}(\lambda)\right\}, \quad \bar{y}_{i}(\lambda)=\mathrm{E}\left[Y_{i}\right]=\tau \int \mathrm{s}_{i}(\overrightarrow{\mathrm{x}}) \lambda(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}+\bar{r}_{i} \tag{8.4.9}
\end{equation*}
$$

In operator-vector notation, we write

$$
\begin{equation*}
\mathrm{E}[\boldsymbol{Y}]=\overline{\boldsymbol{y}}(\lambda)=\mathcal{A} \lambda+\overline{\boldsymbol{r}} \tag{8.4.10}
\end{equation*}
$$

where $\boldsymbol{Y}=\left(Y_{1}, \ldots, Y_{n_{\mathrm{d}}}\right)$ and

$$
\begin{equation*}
[\mathcal{A} \lambda]_{i}=\tau \int \mathrm{s}_{i}(\overrightarrow{\mathrm{x}}) \lambda(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}} \tag{8.4.11}
\end{equation*}
$$

When estimating the emission density $\lambda(\overrightarrow{\mathrm{x}})$, usually one assumes that the $\bar{r}_{i}$ values are "known," meaning that they are determined separately. For example, in PET scans, the $\bar{r}_{i}$ values denote the mean number of recorded random coincidence events. These $\bar{r}_{i}$ values can be estimated using delayed coincidence counting, as well as other methods [33-37].

It is important to note that $\bar{r}_{i}$ is not a random variable, so the expression " $+\overline{\boldsymbol{r}}$ " in (8.4.10) above must not be interpreted as "additive noise." The noise arises from the Poisson variability, not from $\overline{\boldsymbol{r}}$. To elaborate on this point, we could rewrite (8.4.9) as

$$
\begin{array}{ll}
Y_{i}=Y_{i}^{\text {true }}+Y_{i}^{\text {back }}, & Y_{i}^{\text {true }} \sim \operatorname{Poisson}\left\{\tau \int \mathrm{s}_{i}(\overrightarrow{\mathrm{x}}) \lambda(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}\right\} \\
Y_{i}^{\text {back }} \sim \operatorname{Poisson}\left\{\bar{r}_{i}\right\}
\end{array}
$$

Even in this form, $Y_{i}^{\text {back }}$ is not additive noise in the usual sense because $Y_{i}^{\text {true }}$ is also random.

### 8.4.3 Compton scattered photons (s,emis,scatter)

In both PET and SPECT, a large fraction of the emitted photons will undergo Compton scatter before leaving the object [38]. (Some will also scatter in the detector.) Fortunately, using energy-sensitive detectors will exclude many of these photons. Nevertheless, the energy resolution of practical detector materials is imperfect so a non-negligible fraction of scattered photons will also be recorded, i.e., included in the measurement $\boldsymbol{Y}$. There are several possible approaches to dealing with scatter.

- Ignore scatter. This expedient option leads to biased estimates of $\lambda(\vec{x})$, resulting in inaccurate quantification and reduced contrast.
- Estimate the scatter contribution, e.g., by using convolution / deconvolution methods [39, 40] [41, 42], model-based methods [43-45], multiple energy windows [25, 46-50], and a wide variety of other methods, e.g., [24, 51, 52]. Then one could subtract the scatter estimate from the measurements $\boldsymbol{Y}$. However, such subtraction corrupts the Poisson statistical model for the measurements.
- Estimate the scatter contribution (prior to iterating, using one of the above methods for example) and then include that scatter estimate in the additive background $(\overline{\boldsymbol{r}})$ term in (8.4.10). This increases the computation per iteration only very slightly relative to the two previous approaches, but preserves the Poisson statistical model [53]. If the scatter estimate is accurate, the resulting estimate of $\lambda(\vec{x})$ should be free of scatter-induced bias. This approach treats the scattered photons as being non-informative, so an increased scatter fraction would increase the noise in the estimate of $\lambda(\vec{x})$.
- At each iteration, first forward project the current object estimate to compute "ideal" projections, and then estimate the scatter contribution from those projections. This approach was investigated in PET with cylindrical phantoms using a convolution method in [53]. This approach may be insufficiently accurate when imaging objects with more complicated attenuation properties.
- Include the effects of scatter in the system model $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})$. This is an active area of research, e.g., [54-60]. These approaches attempt to treat scattered photons as bearing information [61] about the underlying object $\lambda(\vec{x})$, and in essence try to "put the scattered photons back where they originated." If one can model $s_{i}(\vec{x})$ accurately, then this approach has the potential to reduce the variance in estimates of $\lambda(\vec{x})$ because all recorded events are used. In fact, an increased scatter fraction could potentially decrease the noise in the estimate of $\lambda(\vec{x})$ if the scattering model is accurate, because every recorded photon (scattered or unscattered) bears information about the activity distribution. On the other hand, if the system model is inaccurate, the previous two approaches may be more robust. In particular, the detectors can record counts from scattered photons that originate from outside of the field of view. The contribution of such photons cannot be calculated from the portion of $\lambda(\vec{x})$ seen within the field of view.
Another complication of including scatter in the system model $\mathcal{A}$ in (8.4.10) is that in principle one should then also include the effects of scatter in the corresponding back projector (the adjoint of $\mathcal{A}$ or the transpose of its discretized version $\boldsymbol{A}$ ). This could increase computation substantially and can slow algorithm convergence, leading to proposals to use a mismatched back projector that ignores scatter [62]. This simplification accelerates computation but invalidates convergence conditions in general.


### 8.4.4 Poisson log-likelihood for binned data (s,emis,like)

Under the above assumptions, the joint distribution of the recorded events $\boldsymbol{Y}$ is given by

$$
\mathrm{P}\{\boldsymbol{Y}=\boldsymbol{y}\}=\prod_{i=1}^{n_{\mathrm{d}}} \mathrm{P}\left\{Y_{i}=y_{i}\right\}=\prod_{i=1}^{n_{\mathrm{d}}} \frac{1}{y_{i}!} \mathrm{e}^{-\bar{y}_{i}(\lambda)}\left(\bar{y}_{i}(\lambda)\right)^{y_{i}}
$$

where $\lambda(\vec{x})$ was defined in (8.4.4) and $\bar{y}_{i}$ was defined in (8.4.9). It follows that the log-likelihood for $\lambda$ has the following form:

$$
\begin{equation*}
\mathrm{L}(\lambda ; \boldsymbol{y})=\sum_{i=1}^{n_{\mathrm{d}}} y_{i} \log \bar{y}_{i}(\lambda)-\bar{y}_{i}(\lambda) \tag{8.4.12}
\end{equation*}
$$

The goal in emission tomography is to reconstruct $\lambda(\overrightarrow{\mathrm{x}})$ from $\boldsymbol{y}$, usually by using this log-likelihood [63-67].

### 8.5 List-mode likelihood (s,emis, list)

The statistical analysis in the preceding section was for the case of binned-mode scans. Some nuclear imaging systems collect measurements in a list-mode format [5-7, 68, 69]. List-mode formats have the advantage of retaining complete temporal information, at the price of increased data storage for scanners with a small number of bins. However, for scanners with numerous bins, such as 3D PET systems and Compton imaging systems, the number of recorded events can be fewer than the number of bins, so the list-mode format can in fact require less storage.

There are two types of list-mode scans: preset-time scans, where the scan time is predetermined but the total number of recorded events is a random variable, and preset-counts scans, where scan continues until a predetermined number of events are recorded so the total duration of the scan is a random variable [70]. We focus here on the presettime case. The present-count mode is less common, but has been used in practice for some transmission scans to avoid image quality degradation as the radioisotope transmission source decays [71].

Suppose the system records a total of $M$ events over the time interval $\left[\tau_{1}, \tau_{2}\right]$. For each recorded event, the system records a time $T_{m}$ and an attribute vector $\boldsymbol{V}_{m}$ that typically includes position information and possibly other characteristics such as recorded energy ${ }^{5}$. The log-likelihood associated with these observations is

$$
\begin{equation*}
\mathrm{L}(\lambda)=\log \left(\mathrm{p}\left(\left(\boldsymbol{v}_{1}, t_{1}\right), \ldots,\left(\boldsymbol{v}_{n}, t_{n}\right) \mid M=n ; \lambda\right) \mathrm{P}\{M=n ; \lambda\}\right) \tag{8.5.1}
\end{equation*}
$$

To develop model-based reconstruction methods we need to simplify this log-likelihood expression.
By Corollary 8.4.5, in the absence of background events, the total number of recorded events $M$ would have a Poisson distribution with mean given by (8.4.8). In practice, the system will record not only events that originated from the object but also events originating from the natural background (or from random coincidences in PET). Let $\lambda_{\mathrm{b}}(t)$ denote the rate of the Poisson counting process associated with recorded background events, which has units counts per second. Then the mean of $M$ is

$$
\begin{equation*}
\mathrm{E}_{\lambda}[M]=\int_{\tau_{1}}^{\tau_{2}} \lambda_{\mathrm{s}}(t)+\lambda_{\mathrm{b}}(t) \mathrm{d} t \tag{8.5.2}
\end{equation*}
$$

[^2]where, using (8.4.8), the counting rate associated with recorded events that originated in the object of interest is
\[

$$
\begin{equation*}
\lambda_{\mathrm{s}}(t) \triangleq \int \mathrm{s}(\overrightarrow{\mathrm{x}}) \lambda_{t}(\overrightarrow{\mathrm{x}}) \mathrm{d} t \tag{8.5.3}
\end{equation*}
$$

\]

Thus the distribution of $M$ is

$$
\mathrm{P}\{M=n ; \lambda\}=\frac{1}{n!} \mathrm{e}^{-\mathrm{E}_{\lambda}[M]}\left(\mathrm{E}_{\lambda}[M]\right)^{n}
$$

Rewrite the joint distribution of the event times and attributes as

$$
\mathrm{p}\left(\left(\boldsymbol{v}_{1}, t_{1}\right), \ldots,\left(\boldsymbol{v}_{n}, t_{n}\right) \mid M=n ; \lambda\right)=\mathrm{p}\left(\boldsymbol{v}_{1}, \ldots, \boldsymbol{v}_{n} \mid M=n, t_{1}, \ldots, t_{n} ; \lambda\right) \mathrm{p}\left(t_{1}, \ldots, t_{n} \mid M=n ; \lambda\right)
$$

We assume that the recorded attributes $\left\{\boldsymbol{V}_{m}\right\}$ are conditionally independent random variables given the event times, (i.e., we ignore deadtime):

$$
\begin{equation*}
\mathrm{p}\left(\boldsymbol{v}_{1}, \ldots, \boldsymbol{v}_{n} \mid M=n, t_{1}, \ldots, t_{n} ; \lambda\right)=\prod_{m=1}^{n} \mathrm{p}\left(\boldsymbol{v}_{m} \mid t_{m}, D ; \lambda\right), \tag{8.5.4}
\end{equation*}
$$

where $\mathrm{p}(\boldsymbol{v} \mid t, D ; \lambda)$ denotes the distribution of a recorded attribute vector $\boldsymbol{v}$ given that the event is detected at time $t$. A concrete example of this distribution is derived in $\S 8.5 .5$ for a simple 2D disk detector.

For a Poisson process, the conditional distribution of the (ordered) arrival times $T_{1}<T_{2}<\cdots<T_{n}$ is [73, p. 37, 53]:

$$
\mathrm{p}\left(t_{1}, \ldots, t_{n} \mid M=n ; \lambda\right)= \begin{cases}n!\prod_{m=1}^{n} q\left(t_{m} ; \lambda\right), & \tau_{1}<t_{1}<t_{2}<\cdots<t_{n}<\tau_{2} \\ 0, & \text { otherwise }\end{cases}
$$

where

$$
q(t ; \lambda) \triangleq\left\{\begin{array}{ll}
\frac{\lambda_{\mathrm{s}}(t)+\lambda_{\mathrm{b}}(t)}{\int_{\tau_{1}}^{\tau_{2}} \lambda_{\mathrm{s}}\left(t^{\prime}\right)+\lambda_{\mathrm{b}}\left(t^{\prime}\right) \mathrm{d} t^{\prime}} & \tau_{1} \leq t \leq \tau_{2} \\
0, & \text { otherwise }
\end{array}= \begin{cases}\frac{\lambda_{\mathrm{s}}(t)+\lambda_{\mathrm{b}}(t)}{\mathrm{E}_{\lambda}[M]} & \tau_{1} \leq t \leq \tau_{2} \\
0, & \text { otherwise }\end{cases}\right.
$$

In words, the unordered arrival times are independently and identically distributed on the interval $\left[\tau_{1}, \tau_{2}\right]$ according to $q(t ; \lambda)$ whereas the ordered arrival times follow the usual distribution of order statistics. In practice there might be some error (e.g., due to quantization) in the recorded value of the time $T_{m}$, but for simplicity we ignore such errors. As long as those errors are small relative to the time scale of the temporal changes of $\lambda_{t}(\overrightarrow{\mathrm{x}})$ and $\lambda_{\mathrm{b}}(t)$, this approximation should be adequate. Note that the time measurement $T_{m}$ cannot be absorbed into the attribute vector $\boldsymbol{V}_{m}$ because $T_{1}<T_{2}<\ldots<T_{M}$ which precludes independence. The influence of the recorded times $T_{m}$ was not considered in [6, 7], but they are important for time-varying objects or systems [69].

Substituting these distributions into the list-mode log-likelihood (8.5.1) leads to the following considerably simplified form:

$$
\begin{equation*}
\mathrm{L}(\lambda)=\sum_{m=1}^{M} \log \left(\mathrm{p}\left(\boldsymbol{v}_{m} \mid t_{m}, D ; \lambda\right)\left(\lambda_{\mathrm{s}}\left(t_{m}\right)+\lambda_{\mathrm{b}}\left(t_{m}\right)\right)\right)-\mathrm{E}_{\lambda}[M] \tag{8.5.5}
\end{equation*}
$$

To further simplify, we must analyze the distribution of recorded attributes $\mathrm{p}(\boldsymbol{v} \mid t, D ; \lambda)$.
In general, the recorded attributes for background events may differ from those associated with emissions from the object, so $\mathrm{p}(\boldsymbol{v} \mid t, D ; \lambda)$ is a mixture distribution:

$$
\mathrm{p}(\boldsymbol{v} \mid t, D ; \lambda)=\mathrm{p}\left(\boldsymbol{v} \mid t, D, R_{\mathrm{S}} ; \lambda\right) \mathrm{P}\left\{R_{\mathrm{S}} \mid t, D ; \lambda\right\}+\mathrm{p}\left(\boldsymbol{v} \mid t, D, R_{\mathrm{B}} ; \lambda\right) \mathrm{P}\left\{R_{\mathrm{B}} \mid t, D ; \lambda\right\}
$$

where $R_{\mathrm{S}}$ and $R_{\mathrm{B}}$ denote the events that a recording originated from the source object and background respectively. Because $\lambda_{\mathrm{s}}$ and $\lambda_{\mathrm{b}}$ denote the source and background rates for recorded events respectively:

$$
\begin{aligned}
\mathrm{P}\left\{R_{\mathrm{S}} \mid t, D ; \lambda\right\} & =\frac{\lambda_{\mathrm{s}}(t)}{\lambda_{\mathrm{s}}(t)+\lambda_{\mathrm{b}}(t)} \\
\mathrm{P}\left\{R_{\mathrm{B}} \mid t, D ; \lambda\right\} & =\frac{\lambda_{\mathrm{b}}(t)}{\lambda_{\mathrm{s}}(t)+\lambda_{\mathrm{b}}(t)}=1-\mathrm{P}\left\{R_{\mathrm{S}} \mid t, D ; \lambda\right\}
\end{aligned}
$$

Typically $\mathrm{p}\left(\boldsymbol{v} \mid t, D, R_{\mathrm{B}} ; \lambda\right)$ is independent of $\lambda$ and must be determined separately by some type of calibration process. An exception is in PET where most background events are accidental coincidences whose distribution does depend on $\lambda$. We leave implicit the dependence on $D$ and write $\mathrm{p}\left(\boldsymbol{v} \mid t, R_{\mathrm{B}}\right)$ hereafter. If the background does not vary with time, then we could write simply $\mathrm{p}\left(\boldsymbol{v} \mid R_{\mathrm{B}}\right)$, the distribution of recorded attributes for background events.

Next we consider recorded events that that originate from the source object. By total probability:

$$
\begin{equation*}
\mathrm{p}\left(\boldsymbol{v} \mid t, D, R_{\mathrm{S}} ; \lambda\right)=\int \mathrm{p}\left(\boldsymbol{v} \mid T=t, \overrightarrow{\mathrm{X}}(T)=\overrightarrow{\mathrm{x}}, D, R_{\mathrm{S}} ; \lambda\right) \mathrm{p}_{\overrightarrow{\mathrm{x}}(T)}\left(\overrightarrow{\mathrm{x}} \mid T=t, D, R_{\mathrm{S}} ; \lambda\right) \mathrm{d} \overrightarrow{\mathrm{x}} \tag{8.5.6}
\end{equation*}
$$

The first term in the integral is simply

$$
\mathrm{p}\left(\boldsymbol{v} \mid T=t, \overrightarrow{\mathrm{X}}(T)=\overrightarrow{\mathrm{x}}, D, R_{\mathrm{S}} ; \lambda\right)=\mathrm{p}(\boldsymbol{v} \mid \overrightarrow{\mathrm{x}})
$$

which is the distribution of attributes of recorded events emitted from source location $\vec{x}$. It is independent of the emission density $\lambda$. For a motionless measurement system (with ideal components that do not drift over time), this function is independent of time. See Problem 8.6 and Problem 8.8 for generalizations to systems that change with time, e.g., SPECT systems with rotating Anger cameras or rotating slat collimators, or hand-held detectors in security applications. The function $p(\boldsymbol{v} \mid \overrightarrow{\mathrm{x}})$ is the key part of the system model in a list-mode framework.

The second term in the integral in (8.5.6) is more subtle, because of the conditioning on $D$ (recorded events). Recall that (8.3.5) relates the distribution of tracer atom locations $\mathrm{p}_{t}(\overrightarrow{\mathrm{x}})$ to the emission rate density function $\lambda_{t}(\overrightarrow{\mathrm{x}})$. However, an attribute is recorded for a decay event only if the photon is detected. For an event detected at time $t$, the conditional distribution of the decay spatial location is

$$
\begin{align*}
\mathrm{p}_{\overrightarrow{\mathrm{x}}(T)}\left(\overrightarrow{\mathrm{x}} \mid T=t, D, R_{\mathrm{S}} ; \lambda\right) & =\frac{\mathrm{P}\left\{D \mid \overrightarrow{\mathrm{X}}(T)=\overrightarrow{\mathrm{x}}, T=t, R_{\mathrm{S}} ; \lambda\right\} \mathrm{p}_{\overrightarrow{\mathrm{x}}(T)}\left(\overrightarrow{\mathrm{x}} \mid T=t, R_{\mathrm{S}} ; \lambda\right)}{\mathrm{P}\left\{D \mid T=t, R_{\mathrm{S}} ; \lambda\right\}} \\
& =\frac{\mathrm{s}(\overrightarrow{\mathrm{x}}) \mathrm{p}_{t}(\overrightarrow{\mathrm{x}})}{\mathrm{P}\left\{D \mid T=t, R_{\mathrm{S}} ; \lambda\right\}}=\frac{\mathrm{s}(\overrightarrow{\mathrm{x}}) \mathrm{p}_{t}(\overrightarrow{\mathrm{x}})}{\int \mathrm{s}\left(\overrightarrow{\mathrm{x}}^{\prime}\right) \mathrm{p}_{t}\left(\overrightarrow{\mathrm{x}}^{\prime}\right) \mathrm{d} \overrightarrow{\mathrm{x}}^{\prime}}=\frac{\mathrm{s}(\overrightarrow{\mathrm{x}}) \lambda_{t}(\overrightarrow{\mathrm{x}})}{\int \mathrm{s}\left(\overrightarrow{\mathrm{x}}^{\prime}\right) \lambda_{t}\left(\overrightarrow{\mathrm{x}}^{\prime}\right) \mathrm{d} \overrightarrow{\mathrm{x}}^{\prime}}=\frac{\mathrm{s}(\overrightarrow{\mathrm{x}}) \lambda_{t}(\overrightarrow{\mathrm{x}})}{\lambda_{\mathrm{s}}(t)} \tag{8.5.7}
\end{align*}
$$

by Bayes rule and (8.3.5), because $s(\vec{x})$ denotes the overall detection probability for a decay at position $\vec{x}$. The function $s(\vec{x})$ is the other key part of the system model in a list-mode framework. Substituting into (8.5.6) yields

$$
\mathrm{p}\left(\boldsymbol{v} \mid t, D, R_{\mathrm{S}} ; \lambda\right)=\int \mathrm{p}(\boldsymbol{v} \mid \overrightarrow{\mathrm{x}}) \frac{\mathrm{s}(\overrightarrow{\mathrm{x}}) \lambda_{t}(\overrightarrow{\mathrm{x}})}{\lambda_{\mathrm{s}}(t)} \mathrm{d} \overrightarrow{\mathrm{x}}=\frac{\int \mathrm{p}(\boldsymbol{v} \mid \overrightarrow{\mathrm{x}}) \mathrm{s}(\overrightarrow{\mathrm{x}}) \lambda_{t}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}}{\lambda_{\mathrm{s}}(t)}
$$

Finally, substituting into (8.5.5) and simplifying yields the following general expression for the list-mode loglikelihood for $M$ and $\left\{\left(\boldsymbol{v}_{m}, t_{m}\right): m=1, \ldots, M\right\}$ :

$$
\begin{equation*}
\mathrm{L}(\lambda)=\sum_{m=1}^{M} \log \left(\int \mathrm{p}\left(\boldsymbol{v}_{m} \mid \overrightarrow{\mathrm{x}}\right) \mathrm{s}(\overrightarrow{\mathrm{x}}) \lambda_{t_{m}}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}+\mathrm{p}\left(\boldsymbol{v}_{m} \mid t_{m}, R_{\mathrm{B}}\right) \lambda_{\mathrm{b}}\left(t_{m}\right)\right)-\mathrm{E}_{\lambda}[M] \tag{8.5.8}
\end{equation*}
$$

where $\mathrm{E}_{\lambda}[M]$ is defined above in (8.5.2).

### 8.5.1 Static object

The list-mode log-likelihood expression simplifies when the scan duration is sufficiently small relative to the radionuclide's half life and any other temporal variations of the emission rate density function $\lambda_{t}(\vec{x})$ and the background rate $\lambda_{\mathrm{b}}(t)$. If $\lambda_{t}(\overrightarrow{\mathrm{x}})$ can be considered static, then $\lambda_{t}(\overrightarrow{\mathrm{x}})=\lambda(\overrightarrow{\mathrm{x}})$, where $\lambda(\overrightarrow{\mathrm{x}})$ is the emission density defined in (8.4.4) In this case $\mathrm{E}_{\lambda}[M]=\tau \int \mathrm{s}(\overrightarrow{\mathrm{x}}) \lambda(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}+\tau \lambda_{\mathrm{b}}$, where $\tau$ is the (preset) scan duration. The list-mode log-likelihood (8.5.8) simplifies to:

$$
\begin{equation*}
\mathrm{L}(\lambda) \equiv \sum_{m=1}^{M} \log \left(\tau \int \mathrm{p}\left(\boldsymbol{v}_{m} \mid \overrightarrow{\mathrm{x}}\right) \mathrm{s}(\overrightarrow{\mathrm{x}}) \lambda(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}+\mathrm{p}\left(\boldsymbol{v}_{m} \mid R_{\mathrm{B}}\right) \tau \lambda_{\mathrm{b}}\right)-\left(\tau \int \mathrm{s}(\overrightarrow{\mathrm{x}}) \lambda(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}+\tau \lambda_{\mathrm{b}}\right) \tag{8.5.9}
\end{equation*}
$$

One can parameterize $\lambda(\vec{x})$ using a finite series and perform statistical reconstruction of the series coefficients using this log-likelihood.

Often the attributes $\boldsymbol{V}_{m}$ are stored as integers and hence are discrete random variables. If the dimension of $\boldsymbol{V}_{m}$ is not too large, then a natural data format is the timogram approach of Nichols et al. [68].

### 8.5.2 Object discretization

For numerical implementation we parameterize the emission density of the object using a finite set of basis functions (see Chapter 10) as follows:

$$
\lambda(\overrightarrow{\mathrm{x}})=\sum_{j=1}^{n_{\mathrm{p}}} x_{j} b_{j}(\overrightarrow{\mathrm{x}})
$$

So now the unknown parameter vector is $\boldsymbol{x}=\left(x_{1}, \ldots, x_{n_{\mathrm{p}}}\right)$. Recall that $\lambda(\overrightarrow{\mathrm{x}})$ has units counts per unit volume per unit time, which must also be the units of the product of the units of $x_{j}$ and $b_{j}(\overrightarrow{\mathrm{x}})$. One option is for $x_{j}$ to have the same units as $\lambda(\overrightarrow{\mathrm{x}})$ and $b_{j}(\overrightarrow{\mathrm{x}})$ to be unitless. However, for later simplicity, we choose to define $x_{j}$ to have units of counts per unit time and $b_{j}(\overrightarrow{\mathrm{x}})$ to have units of inverse unit volume. Substituting into (8.5.9) and simplifying yields the discrete-object list-mode log likelihood:

$$
\begin{equation*}
\mathrm{L}(\boldsymbol{x})=\sum_{m=1}^{M} \log \left(\sum_{j=1}^{n_{\mathrm{p}}} \bar{a}_{j}\left(\boldsymbol{v}_{m}\right) x_{j}+\mathrm{p}\left(\boldsymbol{v}_{m} \mid R_{\mathrm{B}}\right) \tau \lambda_{\mathrm{b}}\right)-\left(\sum_{j=1}^{n_{\mathrm{p}}} a_{j} x_{j}+\tau \lambda_{\mathrm{b}}\right) \tag{8.5.10}
\end{equation*}
$$

where $\left\{\bar{a}_{j}\left(\boldsymbol{v}_{m}\right)\right\}$ is a kind of $n \times n_{\mathrm{p}}$ system matrix and

$$
\begin{align*}
\bar{a}_{j}(\boldsymbol{v}) & \triangleq \tau \int \mathrm{p}(\boldsymbol{v} \mid \overrightarrow{\mathrm{x}}) \mathrm{s}(\overrightarrow{\mathrm{x}}) b_{j}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}  \tag{8.5.11}\\
a_{j} & \triangleq \tau \int \mathrm{~s}(\overrightarrow{\mathrm{x}}) b_{j}(\overrightarrow{\mathrm{x}}) \mathrm{d} \overrightarrow{\mathrm{x}}=\int \bar{a}_{j}(\boldsymbol{v}) \mathrm{d} \boldsymbol{v} \tag{8.5.12}
\end{align*}
$$

When $b_{j}(\overrightarrow{\mathrm{x}})$ has units of inverse unit volume as mentioned above, then the following sensitivity map is unitless:

$$
\begin{equation*}
\mathrm{s}_{j} \triangleq a_{j} / \tau \tag{8.5.13}
\end{equation*}
$$

It can be interpreted as a probability in the usual cases where $b_{j}(\vec{x}) \geq 0$. In such cases, $\mathbf{s}=\left(s_{1}, \ldots, s_{n_{\mathrm{p}}}\right)$ is a discrete sensitivity map that describes the probability that an emission from the $j$ th "voxel" is recorded.
See demo_list_mode_em.m.

### 8.5.3 Binned-mode data revisited

We can partition the attribute space $\mathcal{V}$ into $n_{\mathrm{d}}$ disjoint sets or "bins" as follows: $\mathcal{V}=\mathcal{V}_{1} \cup \cdots \cup \mathcal{V}_{n_{\mathrm{d}}}$ and define binned-mode measurements to be the number of recorded attributes that fall within each bin as follows:

$$
Y_{i}=\sum_{m=1}^{M} \mathbb{I}_{\left\{\boldsymbol{v}_{m} \in \mathcal{V}_{i}\right\}}
$$

Note that $\sum_{i=1}^{n_{\mathrm{d}}} Y_{i}=M$. It follows from Theorem 8.4.4 that the $Y_{i}$ random variables are independent, and by (8.5.11):

$$
Y_{i} \sim \operatorname{Poisson}\left\{\bar{y}_{i}\right\}, \quad \bar{y}_{i}=\mathrm{E}\left[Y_{i}\right]=\mathrm{E}_{\lambda}[M] \mathrm{P}\left\{\boldsymbol{v} \in \mathcal{V}_{i} \mid D\right\}
$$

where

$$
\mathrm{E}_{\lambda}[M]=\sum_{j=1}^{n_{\mathrm{p}}} a_{j} x_{j}+\tau \lambda_{\mathrm{b}}
$$

By total probability:

$$
\begin{aligned}
\mathrm{P}\left\{\boldsymbol{v} \in \mathcal{V}_{i} \mid D\right\} & =\mathrm{P}\left\{\boldsymbol{v} \in \mathcal{V}_{i} \mid D, R_{\mathrm{S}}\right\} \mathrm{P}\left\{R_{\mathrm{S}} \mid D\right\}+\mathrm{P}\left\{\boldsymbol{v} \in \mathcal{V}_{i} \mid D, R_{\mathrm{B}}\right\} \mathrm{P}\left\{R_{\mathrm{B}} \mid D\right\} \\
& =\left(\int_{\mathcal{V}_{i}} \frac{\sum_{j=1}^{n_{\mathrm{p}}} \bar{a}_{j}(\boldsymbol{v}) x_{j}}{\sum_{j=1}^{n_{\mathrm{p}}} a_{j} x_{j}} \mathrm{~d} \boldsymbol{v}\right)\left(\frac{\sum_{j=1}^{n_{\mathrm{p}}} a_{j} x_{j}}{\mathrm{E}_{\lambda}[M]}\right)+\left(\int_{\mathcal{V}_{i}} \mathrm{p}\left(\boldsymbol{v} \mid R_{\mathrm{B}}\right) \mathrm{d} \boldsymbol{v}\right)\left(\frac{\tau \lambda_{\mathrm{b}}}{\mathrm{E}_{\lambda}[M]}\right) \\
& =\frac{1}{\mathrm{E}_{\lambda}[M]}\left(\sum_{j=1}^{n_{\mathrm{p}}} a_{i j} x_{j}+\bar{r}_{i}\right),
\end{aligned}
$$

where

$$
\begin{align*}
a_{i j} & \triangleq \int_{\mathcal{V}_{i}} \bar{a}_{j}(\boldsymbol{v}) \mathrm{d} \boldsymbol{v}  \tag{8.5.14}\\
\bar{r}_{i} & \triangleq \tau \lambda_{\mathrm{b}} \int_{\mathcal{V}_{i}} \mathrm{p}\left(\boldsymbol{v} \mid R_{\mathrm{B}}\right) \mathrm{d} \boldsymbol{v} \tag{8.5.15}
\end{align*}
$$

Combining, the mean is

$$
\bar{y}_{i}=\mathrm{E}\left[Y_{i}\right]=\sum_{j=1}^{n_{\mathrm{p}}} a_{i j} x_{j}+\bar{r}_{i}
$$

In the literature, the system matrix elements $a_{i j}$ are often described as the probability $\mathrm{p}_{i j}$ that a decay in the $j$ th voxel is recorded in the $i$ th bin. This interpretation is correct only if the units of $b_{j}(\vec{x})$ are chosen to be inverse unit volume per unit time so that $a_{i j}$ becomes a unitless probability (and $x_{j}$ becomes unitless "counts"). More typically, the elements $a_{i j}$ are proportional to $\mathrm{p}_{i j}$.

For binned-mode data, the sensitivity map in (8.5.13) "simplifies" to

$$
\begin{equation*}
\mathrm{s}_{j}=\frac{1}{\tau} a_{j}=\frac{1}{\tau} \sum_{i=1}^{n_{\mathrm{d}}} a_{i j}, \quad \text { i.e., } \quad \mathbf{s}=\frac{1}{\tau} \boldsymbol{A}^{\prime} \mathbf{1} \tag{8.5.16}
\end{equation*}
$$

using (8.5.12).
In the binned-mode case, the log-likelihood (8.5.9) simplifies to the usual binned-mode log likelihood:

$$
\begin{equation*}
\mathrm{L}(\boldsymbol{x})=\sum_{i=1}^{n_{\mathrm{d}}} Y_{i} \log \left([\boldsymbol{A} \boldsymbol{x}]_{i}+\bar{r}_{i}\right)-\left([\boldsymbol{A} \boldsymbol{x}]_{i}+\bar{r}_{i}\right)=\boldsymbol{Y}^{\prime} \log (\boldsymbol{A} \boldsymbol{x}+\overline{\boldsymbol{r}})-\mathbf{1}^{\prime}(\boldsymbol{A} \boldsymbol{x}+\overline{\boldsymbol{r}}) \tag{8.5.17}
\end{equation*}
$$

For systems where list-mode acquisitions are used, typically $n_{d} \gg \mathrm{E}_{\lambda}[M]$ and (8.5.10) is more practical than (8.5.17). For such systems, most of the terms in the first sum in (8.5.17) are zero, and for efficient implementation we write

$$
\sum_{i=1}^{n_{\mathrm{d}}} Y_{i} \log \left([\boldsymbol{A} \boldsymbol{x}]_{i}+\bar{r}_{i}\right)=\sum_{i: Y_{i}>0} Y_{i} \log \left([\boldsymbol{A} \boldsymbol{x}]_{i}+\bar{r}_{i}\right)
$$

Many of the "list mode" methods in the literature are simply this "binned" version of list mode [74]. Note however that by (8.5.16) the second term in (8.5.17) depends on all rows of the system matrix:

$$
\mathbf{1}^{\prime} \boldsymbol{A} \boldsymbol{x}=\tau \mathbf{s}^{\prime} \boldsymbol{x}=\tau \sum_{j=1}^{n_{\mathrm{p}}} \mathbf{s}_{j} x_{j} .
$$

Unfortunately, no simplification is possible here without making approximations.

### 8.5.4 Sensitivity calculations (s,emis,sens)

When $n_{\mathrm{d}}$ is large, computing the sensitivity map $\mathrm{s}(\overrightarrow{\mathrm{x}})$ or $\mathrm{s}_{j}$ exactly per (8.4.2) or (8.5.13) or (8.5.16) is particularly challenging if not completely impractical. Typically the only viable option is to estimate or approximate the sensitivity map using some form of Monte Carlo simulation.

The simplest approach conceptually is to simulate several emissions from each voxel and propagate the emissions through the physics of the imaging system and count how many of the emissions are recorded. Then an unbiased estimate of the system sensitivity $s_{j}$ for the $j$ th voxel is simply the ratio of recorded counts over simulated emissions. If $N$ emissions are simulated, then the variance of this simple estimate is $\operatorname{Var}\left\{\hat{\mathbf{s}}_{j}\right\}=\mathrm{s}_{j}\left(1-\mathrm{s}_{j}\right) / N$. This imagedomain approach would require a very large number of simulated emissions to generate reasonably precise estimates. (Simulating 100 emissions per voxel would provide only $10 \%$ precision, which is probably insufficient.) Thus this image domain is impractical and usually alternative methods are used.

The sensitivity map is the back projection of a uniform measurement vector: $\mathbf{s}=\mathbf{P}^{\prime} \mathbf{1}_{n_{\mathrm{d}}}$, where we define $\mathbf{P} \triangleq \frac{1}{\tau} \boldsymbol{A}$. Practical Monte Carlo methods are based on approximations to this expression. Instead of back projecting every possible measurement, suppose we select a random subset of the indices $\left\{1, \ldots, n_{d}\right\}$. Let $q_{i} \in(0,1)$ denote the probability that we select the $i$ th index, where $\sum_{i=1}^{n_{\mathrm{d}}} \mathrm{q}_{i}=M \ll n_{\mathrm{d}}$ and $M$ is the number of events to be simulated. Let $B_{i}$ denote a Bernoulli random variable with success probability $\mathrm{q}_{i}$. Then an unbiased estimate of the sensitivity
image is $\hat{\mathrm{s}}_{j}=\sum_{i=1}^{n_{\mathrm{d}}} \frac{B_{i}}{\mathrm{q}_{i}} \mathrm{p}_{i j}$ because $\mathrm{E}\left[B_{i}\right]=\mathrm{q}_{i}$. Of course the practical implementation of this expression uses only the nonzero terms:

$$
\hat{\mathrm{s}}_{j}=\sum_{i \in \mathcal{I}} \frac{\mathrm{p}_{i j}}{\mathrm{q}_{i}}, \quad \mathcal{I}=\left\{i: B_{i} \neq 0\right\},
$$

where typically $|\mathcal{I}| \approx M \ll n_{\mathrm{d}}$. The simplest option would be a uniform distribution [75] $q_{i}=M / n_{\mathrm{d}}$, but this choice is suboptimal [76]. Other options include $\mathrm{q}_{i}=M a_{i} . / \sum_{k=1}^{n_{\mathrm{d}}} a_{k}$. where $a_{i} . \triangleq \sum_{j=1}^{n_{\mathrm{p}}} a_{i j}$ is the forward projection of a uniform image and $\mathrm{q}_{i}=M \sqrt{c_{i}} / \sum_{k=1}^{n_{\mathrm{d}}} \sqrt{c_{k}}$ where $c_{i}=\sum_{j=1}^{n_{\mathrm{p}}} a_{i j}^{2}$.

Yet another alternative is to simulate (via Monte Carlo) a set of $M$ recorded events of list-mode data from a uniform image, i.e., $\boldsymbol{x}=\frac{M}{n_{\mathrm{p}}} 1$. Let $i_{m} \in\left\{i=1, \ldots, n_{\mathrm{d}}\right\}$ denote the index of recorded events and set $i_{m}=0$ if the $m$ th event is not recorded. Then estimate s as follows: [77, p. 130]:

$$
\begin{equation*}
\hat{\mathbf{s}}_{j} \triangleq \frac{n_{\mathrm{p}}}{M} \sum_{\substack{m=1 \\ i_{m}=0}}^{M} \frac{a_{i_{m}, j}}{a_{i_{m}}} \tag{8.5.18}
\end{equation*}
$$

This estimator was motivated by the multiplicative form of the ML-EM algorithm (18.4.3). It is invariant to $i$ dependent scale factors [77] and can be shown to be unbiased [77, p. 132]. (See Problem 8.10.)

Examining how system model mismatch (errors in the $a_{i j}$ values) will propagate into errors in the sensitivity map estimates is an interesting open problem. The effects of sensitivity map errors on the reconstructed image has been analyzed [78]. The image domain approach yields sensitivity map estimates that are independent, whereas the other methods yield correlated errors.

### 8.5.5 Disk detector: 2D sensitivity and position distribution (s,emis,disk)

This section derives an concrete example of the recorded attribute distribution $\mathrm{p}(\boldsymbol{v} \mid D)$ introduced in (8.5.4), for a hypothetical 2D problem. This presentation illustrates that such derivations can be subtle due to the conditioning on the event $D$ that something was recorded, i.e., an emission was detected.

Consider a 2D disk of radius $R$ with attenuation coefficient $\mu$ centered at the origin of the 2D plane, as illustrated in Fig. 8.5.1. Due to the circular symmetry, it is natural to use polar coordinates for the location of a point source in the 2D plane: $\overrightarrow{\mathrm{x}}=(r, \phi)$. Using total probability, the sensitivity pattern of such a disk detector is

$$
s(\overrightarrow{\mathrm{x}})=s(r, \phi)=s(r)=\mathrm{p}(D ; r)=\frac{1}{2 \pi} \int_{-\theta_{r}}^{\theta_{r}} \mathrm{p}(D \mid \theta ; r) \mathrm{d} \theta
$$

where $\theta_{r} \triangleq \arcsin (R / r)$ for $r>R$ and $D$ denotes the event that an emission is recorded. The intersection length of a ray at angle $\theta$ from the source through the detector is $2 \sqrt{R^{2}-r^{2} \sin ^{2} \theta}$ so

$$
\begin{equation*}
\mathrm{p}(D \mid \theta ; r)=\left(1-\mathrm{e}^{-\mu 2 \sqrt{R^{2}-r^{2} \sin ^{2} \theta}}\right) \mathbb{I}_{\left\{|\theta| \leq \theta_{r}\right\}} \tag{8.5.19}
\end{equation*}
$$

and

$$
s(r)=\frac{1}{\pi} \int_{0}^{\theta_{r}} \mathrm{p}(D \mid \theta ; r) \mathrm{d} \theta=\frac{\theta_{r}}{\pi}-\frac{1}{\pi} \int_{0}^{\theta_{r}} \mathrm{e}^{-2 \mu \sqrt{R^{2}-r^{2} \sin ^{2} \theta}} \mathrm{~d} \theta
$$

Making the change of variables $t=\frac{r}{R} \sin \theta$ yields:

$$
\begin{equation*}
s(r)=\frac{\theta_{r}}{\pi}-\frac{1}{\pi} \int_{0}^{1} \mathrm{e}^{-2 \alpha \sqrt{1-t^{2}}} \frac{R}{r \sqrt{1-(R t / r)^{2}}} \mathrm{~d} t \tag{8.5.20}
\end{equation*}
$$

where $\alpha \triangleq \mu R$. Thus, as $r \rightarrow \infty$ :

$$
r s(r) \rightarrow \frac{R}{\pi}\left(1-\int_{0}^{1} \mathrm{e}^{-2 \alpha \sqrt{1-t^{2}}} \mathrm{~d} t\right)
$$

so the far-field sensitivity decreases as $1 / r$.
Fig. 8.5.2 compares the exact sensitivity and the far-field approximation for $\mu=2$. There is very good agreement between the two for $r>3 R$.

Suppose that each recorded attribute vector $\boldsymbol{v}$ is simply the interaction location within the disk: $\boldsymbol{v}=(x, y)$. To derive the distribution of interaction locations $(x, y) \in \mathcal{S}_{R} \triangleq\left\{(x, y): x^{2}+y^{2} \leq R\right\}$ for recorded events emitted
from $\overrightarrow{\mathrm{x}}=(r, 0)$, we first consider the alternative coordinate system $(l, \phi)$ over the disk $\mathcal{S}_{R}$ for which $(x, y)=$ $(r-l \sin \phi, l \sin \phi)$, where $l=\sqrt{(r-x)^{2}+y^{2}}$ is the distance from the point $(r, 0)$ to $(x, y)$ and $\phi=\arctan \left(\frac{y}{r-x}\right)$ is the corresponding angle.

Using Beer's law:

$$
\begin{equation*}
\mathrm{p}(l \mid \phi, D ; r)=\frac{\mu \mathrm{e}^{-\mu\left(l-r \cos \phi+\sqrt{R^{2}-r^{2} \sin ^{2} \phi}\right)}}{1-\mathrm{e}^{-\mu 2 \sqrt{R^{2}-r^{2} \sin ^{2} \phi}}} \mathbb{I}_{\left\{r \cos \phi-\sqrt{R^{2}-r^{2} \sin ^{2} \phi} \leq l \leq r \cos \phi+\sqrt{R^{2}-r^{2} \sin ^{2} \phi}\right\}}, \tag{8.5.21}
\end{equation*}
$$

and using Bayes rule and (8.5.19):

$$
\begin{equation*}
\mathrm{p}(\phi \mid D ; r)=\frac{\mathrm{p}(D \mid \phi ; r) \mathrm{p}(\phi ; r)}{\mathrm{p}(D ; r)}=\frac{\mathrm{p}(D \mid \phi ; r) \frac{1}{2 \pi}}{s(r)}=\frac{1}{2 \pi s(r)}\left(1-\mathrm{e}^{-2 \mu \sqrt{R^{2}-r^{2} \sin ^{2} \phi}}\right) \mathbb{I}_{\left\{|\phi| \leq \theta_{r}\right\}} \tag{8.5.22}
\end{equation*}
$$

Thus by the definition of conditional probability:

$$
\begin{align*}
\mathrm{p}(l, \phi \mid D ; r) & =\mathrm{p}(l \mid \phi, D ; r) \mathrm{p}(\phi \mid D ; r) \\
& =\frac{\left.\mu \mathrm{e}^{-\mu\left(l-r \cos \phi+\sqrt{R^{2}-r^{2} \sin ^{2} \phi}\right.}\right)}{2 \pi s(r)} \mathbb{I}_{\left\{r \cos \phi-\sqrt{R^{2}-r^{2} \sin ^{2} \phi} \leq l \leq r \cos \phi+\sqrt{R^{2}-r^{2} \sin ^{2} \phi}\right\}} \mathbb{I}_{\left\{|\phi| \leq \theta_{r}\right\}} \tag{8.5.23}
\end{align*}
$$

Now using the formula for transformation of random variables (31.4.4) yields our final expression for the attribute distribution $\mathrm{p}(\boldsymbol{v} \mid D ; r)$ :

$$
\begin{equation*}
\mathrm{p}(x, y \mid D ; r)=\left.\frac{1}{l} \mathrm{p}(l, \phi \mid D ; r)\right|_{l=\sqrt{(r-x)^{2}+y^{2}}, \phi=\arctan \left(\frac{y}{r-x}\right)} . \tag{8.5.24}
\end{equation*}
$$

Fig. 8.5.3 illustrates this pdf for $\mu=2$ and $r / R=3$.


Figure 8.5.1: Geometry of simple 2D disk detector.


Figure 8.5.2: Sensitivity of a radial disk detector of radius $R$ and absorption coefficient $\mu$.
To derive the pdf (8.5.24) properly, it is essential to use the probability of $\phi$ conditional on the event $D$ that the emission was recorded, as in (8.5.22).

Fig. 8.5.4 compares the conditional distribution of $\phi$ given $D$ in (8.5.22). to the distribution where we are given only that the emitted photon was incident on the detector:

$$
\begin{equation*}
\mathrm{p}(\varphi \mid \text { incident } ; r)=\frac{1}{2 \theta_{r}} \mathbb{I}_{\left\{|\phi| \leq \theta_{r}\right\}} . \tag{8.5.25}
\end{equation*}
$$



Figure 8.5.3: Distribution of recorded positions for emissions from point $(3 R, 0)$ when $\mu=2$.


Figure 8.5.4: Conditional distribution $\mathrm{p}(\phi \mid D ; r)$ of emission angle $\phi$ compared to the simpler distribution of $\phi$ given only that the emitted photon was incident on the detector. Conditioning on $D$ leads to a highly nonuniform distribution.

Conditioning on $D$ leads to a highly nonuniform distribution for $\phi$ where smaller angles are more probable.
If we ignore the " $D$ " part of (8.5.22) and use (8.5.25) instead, then we get the (inaccurate) model

$$
\tilde{\mathrm{p}}(l, \phi ; r)=\frac{1}{2 \theta_{r}} \frac{\mu \mathrm{e}^{-\mu\left(l-r \cos \varphi+\sqrt{\left.R^{2}-r^{2} \sin ^{2} \varphi\right)}\right.}}{1-\mathrm{e}^{-2 \mu \sqrt{R^{2}-r^{2} \sin ^{2} \varphi}}} \mathbb{I}_{\left\{r \cos \phi-\sqrt{R^{2}-r^{2} \sin ^{2} \phi} \leq l \leq r \cos \phi+\sqrt{R^{2}-r^{2} \sin ^{2} \phi}\right\}^{\mathbb{I}_{\left\{|\phi| \leq \theta_{r}\right\}}} . . .2{ }^{2} .}
$$

This model has singularities for $\phi \approx \pm \theta_{r}$; therefore it is crucial to condition on $D$ when deriving the probability models for list-mode data.

Note that the distribution expression (8.5.23) depends on the sensitivity $s(r)$. In this case the sensitivity has the relatively simple expression (8.5.20). Typically the sensitivity is a much more complicated function and the dependence of the distribution (8.5.23) on it is an unfortunate but apparently unavoidable complication.

### 8.6 PET-specific topics

### 8.6.1 Randoms-precorrected PET scans (s,emis,randoms)

In PET scans, true coincidence events are those that originate from a single positron-electron annihilation, whereas random coincidence events are those that originate from two or more positron-electron annihilations [33]. PET scanners usually detect two types of events: prompt coincidences are coincidences that occur within a small time interval that is just large enough to include all true coincidence events; delayed coincidences are those that occur with a time delay that is large enough to exclude all true coincidence events. Random coincidence events will contaminate both such measurements, and because of the time scales involved, the mean contribution of randoms to the two types of coincidences will be essentially equal. Thus, the standard statistical model for prompt and delayed coincidence measurements is:

$$
\begin{aligned}
Y_{i}^{\text {prompt }} & \sim \operatorname{Poisson}\left\{[\mathcal{A} \lambda]_{i}+\bar{r}_{i}\right\} \\
Y_{i}^{\text {delay }} & \sim \operatorname{Poisson}\left\{\bar{r}_{i}\right\},
\end{aligned}
$$

using (8.4.11). One can form an unbiased estimate of the true coincidences by subtracting the delayed events from the prompt events:

$$
Y_{i}^{\text {diff }} \triangleq Y_{i}^{\text {prompt }}-Y_{i}^{\text {delay }}
$$

Many PET scanners perform this subtraction in real time during the scan, recording only the difference $Y_{i}^{\text {diff }}$. Unfortunately, this subtraction destroys the Poisson statistics. In particular, the variance and the mean differ:

$$
\begin{align*}
\mathrm{E}\left[Y_{i}^{\text {diff }}\right] & =[\mathcal{A} \lambda]_{i} \\
\operatorname{Var}\left\{Y_{i}^{\text {diff }}\right\} & =[\mathcal{A} \lambda]_{i}+2 \bar{r}_{i} \tag{8.6.1}
\end{align*}
$$

The exact log-likelihood for $Y_{i}^{\text {diff }}$ is complicated, so approximations have been investigated for both transmission scans [79-82] and emission scans [83-85]. The conclusion of this work is that the "ordinary Poisson" approach of assuming that $Y_{i}^{\text {diff }}$ has a Poisson distribution with mean $\mathrm{E}\left[Y_{i}^{\text {diff }}\right]$ leads to suboptimal estimates. The most useful approximation is to first form estimates $\hat{r}_{i}$ of the $\bar{r}_{i}$ values, for example using the block singles events that usually are recorded in PET scanners [86], and then use the following shifted Poisson model:

$$
\begin{equation*}
Y_{i}^{\mathrm{diff}}+2 \hat{r}_{i} \sim \operatorname{Poisson}\left\{[\mathcal{A} \lambda]_{i}+2 \hat{r}_{i}\right\} \tag{8.6.2}
\end{equation*}
$$

There can still be negative values even after "shifting" by adding $2 \hat{r}_{i}$, but this can be accommodated with appropriate algorithms [84, 85].

An alternative approach would be to use model-weighted least squares (MWLS) [87-90] where the weights depend on the mean and variance in (8.6.1), as follows:

$$
\mathfrak{Ł}(\lambda)=\sum_{i=1}^{n_{\mathrm{d}}} \mathrm{~h}_{i}\left([\mathcal{A} \lambda]_{i}\right), \quad \mathrm{h}_{i}(l)=\frac{1}{2} \frac{\left(y_{i}-l\right)^{2}}{l+2 \bar{r}_{i}} .
$$

Interestingly, this function is convex and

$$
\ddot{\mathrm{h}}_{i}(l)=\frac{y_{i}^{2}+4 \bar{r}_{i} y_{i}+4 y_{i}^{2}}{\left(l+2 \bar{r}_{i}\right)^{3}}
$$

Nevertheless, it seems no more convenient for optimization than the shifted Poisson log-likelihood (8.6.2).
Yet another approach is related to iteratively reweighted least-squares (IRLS) methods [91]. First use some method to form an initial reconstructed image $\hat{\lambda}^{(0)}$. Then estimate the variances of $Y_{i}^{\text {diff }}$ as follows:

$$
\hat{\sigma}_{i}^{2} \triangleq\left[\mathcal{A} \hat{\lambda}^{(0)}\right]_{i}+2 \hat{r}_{i}
$$

Finally, estimate a refined reconstructed image using a weighted least-squares (WLS) cost function of this form:

$$
\begin{equation*}
\mathrm{Ł}(\lambda)=\sum_{i=1}^{n_{\mathrm{d}}} \mathrm{~h}_{i}\left([\mathcal{A} \lambda]_{i}\right), \quad \mathrm{h}_{i}(l)=\frac{1}{2} \frac{\left(y_{i}-l\right)^{2}}{\hat{\sigma}_{i}^{2}} . \tag{8.6.3}
\end{equation*}
$$

Because the weights $w_{i}=1 / \hat{\sigma}_{i}^{2}$ are held fixed while (re)estimating $\lambda$, the cost function $Ł$ is quadratic (and convex). One could iterate this process by estimating $\sigma_{i}^{2}$ from the new $\hat{\lambda}$, but often this is unnecessary [91].

### 8.6.2 Time-of-flight PET (s,emis,tof)

In the 1980's, some PET systems were investigated that could measure time of flight (TOF) information, e.g., [92-94]. Although TOF information has the potential to improve spatial resolution and reduce noise, the scintillators used in these systems had other significant disadvantages. More recently, TOF systems are being reconsidered due to advances in scintillator technology and electronics [95]. The principles of statistical reconstruction for TOF PET are the same as for non-TOF PET [5, 35, 96, 97]. The primary difference is the system model.

### 8.7 Summary (s,emis,summ)

Applying the various assumptions made throughout this chapter, the main results presented are the Poisson loglikelihood for binned-mode data (8.4.12) (8.5.17) and for list-mode data for dynamic (8.5.8) and for static (8.5.9) objects. These statistical models form the foundation of the reconstruction algorithms developed in Chapter 18 and Chapter 19.

### 8.8 Problems (s,emis,prob)

Problem 8.1 Find an expression for the Fisher information for the discrete-object list-mode log-likelihood (8.5.10).
Problem 8.2 Generalize the proof of Lemma 8.2.6 to the case $M>2$.
Problem 8.3 Prove Theorem 8.3.4.
Problem 8.4 Generalize the analysis in $\S 8.3$ to the case of radiotracers containing multiple radioisotopes.
Problem 8.5 Generalize the analysis in $\$ 8.4$ to the case of radioisotopes that produce multiple photons (with various probabilities) that are emitted and recorded [15].
(Solve?)
p,emis,sixt
Problem 8.6 Generalize Theorem 8.4.4 and Corollary 8.4.5 for the case of a moving imaging system, where the detector sensitivity patterns vary with time, i.e., $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}}, t)$.

Problem 8.7 Complete Theorem 8.4.4 by showing that the random variables $Y_{i}$ are statistically independent.
Problem 8.8 Generalize $\S 8.5$ to the case of a dynamic or time-varying object model $\lambda_{t}(\vec{x})$ with a possibly moving or time-varying system described by time-varying sensitivity function $\mathrm{s}_{t}(\overrightarrow{\mathrm{x}})$ and time-varying recording distribution $\mathrm{p}_{t}(\boldsymbol{v} \mid \overrightarrow{\mathrm{x}})$.

Problem 8.9 After studying Chapter 25, use Fisher information matrices to justify the claims made in §8.4.3 about the noise effects of Compton scatter.

Problem 8.10 The sensitivity map estimate (8.5.18) uses a uniform distribution in the image domain ( $\boldsymbol{x}=M / n_{\mathrm{p}} \mathbf{1}$ ) which may be suboptimal. Generalize (8.5.18) to the case of a nonuniform distribution $\mathrm{p}_{j}, j=1, \ldots, n_{\mathrm{p}}$ for sampling from the voxels, and prove that your proposed estimator is unbiased.

### 8.9 Appendix A: 2D gamma camera (s,emis,spect2)

To provide a concrete example of a detector unit sensitivity pattern $s_{i}(\vec{x})$, consider the 2D SPECT geometry illustrated in Fig. 8.9.1. The scintillator of a collimated Anger camera is placed at a distance $D_{0 \mathrm{~d}}$ from the origin of the coordinate system (typically the center of rotation). Assume that the scintillator is very thin but has a very high attenuation coefficient so that each gamma ray that reaches it is recorded in one of the pixels behind it. The $i$ th detector element covers the interval $\left[y_{i}^{-}, y_{i}^{+}\right]$. In front of the scintillator is a parallel-hole collimator having holes of length $l$ and width $w$.

Consider a given position $\overrightarrow{\mathrm{x}}=(x, y)$ and let $D_{i}$ denote the event that an emission from position $\overrightarrow{\mathrm{x}}$ is recorded by the $i$ th detector element. Then by total probability:

$$
\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})=\mathrm{P}\left\{D_{i}\right\}=\int_{-\pi}^{\pi} \mathrm{P}\left\{D_{i} \mid \varphi\right\} \frac{1}{2 \pi} \mathrm{~d} \varphi=\int_{\varphi_{i}^{-}}^{\varphi_{i}^{+}} \mathrm{P}\left\{D_{i} \mid \varphi\right\} \frac{1}{2 \pi} \mathrm{~d} \varphi
$$

where $\varphi$ denotes the (random) angle of $\gamma$-ray emission within the 2D plane, and $\varphi_{i}^{ \pm}=\arctan \left(\frac{y_{i}^{ \pm}-y}{d}\right)$, where $d=$ $d(\overrightarrow{\mathrm{x}}) \triangleq D_{0 \mathrm{~d}}-x$ is the distance from the point $\overrightarrow{\mathrm{x}}$ to the detector.

The exact response of a collimator is quite complicated so we simplify by assuming infinitesimal collimator septa and averaging over all possible translations of the collimator along the vertical axis [98]. Then one can show that the probability that a photon at angle $\varphi$ reaches the scintillator is

$$
\begin{equation*}
\mathrm{P}\left\{D_{i} \mid \varphi\right\}=\left[1-\frac{l}{w} \tan |\varphi|\right]_{+} \tag{8.9.1}
\end{equation*}
$$

so

$$
\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})=\int_{\varphi_{i}^{-}}^{\varphi_{i}^{+}}\left[1-\frac{l}{w} \tan (|\varphi|)\right]_{+} \frac{1}{2 \pi} \mathrm{~d} \varphi=\int_{y_{i}^{-}-y}^{y_{i}^{+}-y} \frac{1}{2 \pi}\left[1-\frac{l}{w} \frac{|r|}{d}\right]_{+} \frac{1}{d} \frac{1}{1+(r / d)^{2}} \mathrm{~d} r
$$

making the change of variables $r=d \tan \varphi$. The FWHM of the integrand is $2 d\left(\sqrt{1+(l / w)^{2}}-l / w\right)$. This width increases with distance $d$ from the point to the detector, as seen in Fig. 8.4.1. Fig. 8.9.2 illustrates how the detector unit sensitivity pattern changes as a function of the ratio $l / w$. Profiles of these functions along the $y$ direction appear nearly triangular.

The overall system sensitivity pattern is

$$
\mathrm{s}(\overrightarrow{\mathrm{x}})=\sum_{i=1}^{n_{\mathrm{d}}} \mathrm{~s}_{i}(\overrightarrow{\mathrm{x}})=\int \frac{1}{2 \pi}\left[1-\frac{l}{w} \frac{|r|}{d}\right]_{+} \frac{1}{d} \frac{1}{1+(r / d)^{2}} \mathrm{~d} r \approx \frac{1}{2 \pi} \frac{w}{l}
$$

Increasing the collimator aspect ratio $l / w$ improves spatial resolution but decreases sensitivity.


Figure 8.9.1: Illustration of geometry of 2D collimated gamma camera.

### 8.10 Bibliography

[1] R. J. Estep, T. H. Prettyman, and G. A. Sheppard. "Tomographic gamma scanning to assay heterogeneous radioactive waste." In: Nucl. Sci. Eng. 118.3 (Nov. 1994), 145-52 (cit. on p. 8.1).


Figure 8.9.2: Illustration of detector unit sensitivity patterns $\mathrm{s}_{i}(\overrightarrow{\mathrm{x}})$ for $l / w=3$ (left) and $l / w=8$ (right) for idealized 2D gamma camera with $D_{0 \mathrm{~d}}=150$.
[2] D. Xu and Z. He. "Gamma-ray energy-imaging integrated spectral deconvolution." In: Nucl. Instr. Meth. Phys. Res. A. 574.1 (Apr. 2007), 98-109. DOI: $10.1016 /$ j.nima. 2007.01 .171 (cit. on pp. 8.1, 8.2).
[3] H. de Jong et al. "Simultaneous $99 \mathrm{mTc} / 201 \mathrm{Tl}$ dual-isotope SPET with Monte Carlo-based down-scatter correction." In: Eur. J. Nuc. Med. 29.8 (Aug. 2002), 1063-71. DOI: 10.1007 / s $00259-002$ - 0834 -1 (cit. on p. 8.2).
[4] C-M. Kao et al. "A continuous-coordinate image reconstruction method for list-mode TOF PET." In: Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf. 2013 (cit. on p. 8.2).
[5] D. L. Snyder and D. G. Politte. "Image reconstruction from list-mode data in emission tomography system having time-of-flight measurements." In: IEEE Trans. Nuc. Sci. 20.3 (June 1983), 1843-9. DOI: $10.1109 /$ TNS .1983 .4332660 (cit. on pp. 8.3, 8.14, 8.23).
[6] H. H. Barrett, T. White, and L. C. Parra. "List-mode likelihood." In: J. Opt. Soc. Am. A 14.1 (Nov. 1997), 2914-23. DOI: 10.1364 /JOSAA. 14.002914 (cit. on pp. 8.3, 8.14, 8.15).
[7] L. C. Parra and H. H. Barrett. "List-mode likelihood - EM algorithm and noise estimation demonstrated on 2D-PET." In: IEEE Trans. Med. Imag. 17.2 (Apr. 1998), 228-35. DOI: 10.1109/42. 700734 (cit. on pp. 8.3, 8.14, 8.15).
[8] B. W. Silverman. Density estimation for statistics and data analysis. New York: Chapman and Hall, 1986 (cit. on p. 8.4).
[9] I. J. Good and R. A. Gaskins. "Nonparametric roughness penalties for probability densities." In: Biometrika 58.2 (1971), 255-77. DOI: $10.1093 /$ biomet/58.2. 255 (cit. on p. 8.4).
[10] I. J. Good and R. A. Gaskins. "Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering meteorites data." In: J. Am. Stat. Assoc. 75.369 (Mar. 1980), 42-56. URL: http://www. jstor.org/stable/2287379 (cit. on p. 8.4).
[11] B. W. Silverman. "On the estimation of a probability density function by the maximum penalized likelihood method." In: Ann. Stat. 10.3 (Sept. 1982), 795-810. DOI: 10.1214 / aos /1176345872 (cit. on p. 8.4).
[12] D. L. Snyder and M. I. Miller. Random point processes in time and space. New York: Springer-Verlag, 1991 (cit. on p. 8.5).
[13] S. Ross. A first course in probability. 4th ed. Englewood Cliffs, NJ: Macmillan, 1994 (cit. on p. 8.7).
[14] J. A. Sorenson and M. E. Phelps. Physics in nuclear medicine. 2nd ed. Philadelphia: Saunders, 1987 (cit. on p. 8.7).
liang: 90:com
[15] Z. Liang and R. Jaszczak. "Comparisons of multiple photon coincidence imaging techniques." In: IEEE Trans. Nuc. Sci. 37.3 (June 1990), 1282-92. DOI: $10.1109 / 23.57378$ (cit. on pp. 8.7, 8.23).
schiff: 1936:sao
[16] L. I. Schiff. "Statistical analysis of counter data." In: Phys. Rev. 50.1 (July 1936), 88-96. DOI: $10.1103 /$ PhysRev. 50.88 (cit. on p. 8.7).
[17] J. W. Müller. "Dead-time problems." In: Nucl. Instr. Meth. 112.1 (Sept. 1973), 47-57. DOI: 10.1016/0029-554X(73) 90773-8 (cit. on p. 8.7).
[18] R. D. Evans. The atomic nucleus. New York: McGraw-Hill, 1955 (cit. on p. 8.7).
[19] A. Leon-Garcia. Probability and random processes for electrical engineering. 2nd ed. New York: Addison-Wesley, 1994 (cit. on p. 8.7).
[20] H. H. Barrett and K. J. Myers. Foundations of image science. New York: Wiley, 2003 (cit. on p. 8.8).
[21] K. Kacperski, N. M. Spyrou, and F. A. Smith. "Three-gamma annihilation imaging in positron emission tomography." In: IEEE Trans. Med. Imag. 23.4 (Apr. 2004), 525-9. DOI: $10.1109 /$ TMI 2004.824150 (cit. on p. 8.10).
[22] G. T. Herman. "The mathematics of wobbling a ring of positron annihilation detectors." In: IEEE Trans. Nuc. Sci. 26.2 (Apr. 1979), 2756-59. DOI: $10.1109 /$ TNS .1979 .4330530 (cit. on p. 8.10).
[23] M. R. Palmer et al. "Effects of detector wobble motion on image noise in positron emission tomography." In: IEEE Trans. Med. Imag. 4.1 (Mar. 1985), 58-62. DOI: $10.1109 /$ TMI. 1985.4307694 (cit. on p. 8.10).
[24] R. J. Jaszczak, C. E. Floyd, and R. E. Coleman. "Scatter compensation techniques for SPECT." In: IEEE Trans. Nuc. Sci. 32.1 (Feb. 1985), 786-93. DOI: $10.1109 /$ TNS .1985 .4336941 (cit. on pp. 8.10, 8.13).
[25] K. F. Koral et al. "SPECT Compton-scattering correction by analysis of energy spectra." In: J. Nuc. Med. 29.2 (Feb. 1988), 195-202. URL:
http://jnm.snmjournals.org/cgi/content/abstract/29/2/195 (cit. on pp. 8.10, 8.13).
[26] L. Shao, R. Freifelder, and J. S. Karp. "Triple energy window scatter correction technique in PET." In: IEEE Trans. Med. Imag. 13.4 (Dec. 1994), 641-8. DOI: $10.1109 / 42.363104$ (cit. on p. 8.10).
[27] M. R. Palmer and G. L. Brownell. "Annihilation density distribution calculations for medically important positron emitters." In: IEEE Trans. Med. Imag. 11.3 (Sept. 1992), 373-8. DOI: $10.1109 / 42.158941$ (cit. on p. 8.10).
[28] C. S. Levin and E. J. Hoffman. "Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution." In: Phys. Med. Biol. 44.3 (Mar. 1999), 781-800. DOI: 10.1088/0031-9155/44/3/019 (cit. on p. 8.10).
[29] A. Sánchez-Crespo, P. Andreo, and S. A. Larsson. "Positron flight in human tissues and its influence on PET image spatial resolution." In: Eur. J. Nuc. Med. 31.1 (Jan. 2004), 44-51. DOI: 10.1007/s00259-003-1330-y (cit. on p. 8.10).
[30] R. M. Lewitt and S. Matej. "Overview of methods for image reconstruction from projections in emission computed tomography." In: Proc. IEEE 91.10 (Oct. 2003), 1588-611. DOI: $10.1109 /$ JPROC. 2003.817882 (cit. on p. 8.10).
[31] D. F. Yu and J. A. Fessler. "Mean and variance of singles photon counting with deadtime." In: Phys. Med. Biol. 45.7 (July 2000), 2043-56. DOI: $10.1088 / 0031$-9155/45/7/324 (cit. on p. 8.11).
[32] D. F. Yu and J. A. Fessler. "Mean and variance of coincidence photon counting with deadtime." In: Nucl. Instr. Meth. Phys. Res. A. 488.1-2 (Aug. 2002), 362-74. DOI: 10.1016/S0168-9002 (02) 00460-6 (cit. on p. 8.11).
[33] E. J. Hoffman et al. "Quantitation in positron emission computed tomography: 4 Effect of accidental coincidences." In: J. Comp. Assisted Tomo. 5.3 (June 1981), 391-400. URL:
http://gateway.ovid.com/ovidweb.cgi?T=JS\&MODE=ovid\&NEWS=n\&PAGE=toc\&D= ovft \&AN=00004728-198106000-00015 (cit. on pp. 8.13, 8.22).
[34] T. Holmes, D. Ficke, and D. Snyder. "Modeling of accidental coincidences in both conventional and TOF PET." In: IEEE Trans. Nuc. Sci. 31.1 (Feb. 1984), 627-31. DOI: $10.1109 /$ TNS .1984 .4333332 (cit. on p. 8.13).
[35] T. J. Holmes, D. L. Snyder, and D. C. Ficke. "The effect of accidental coincidences in time-of-flight positron emission tomography." In: IEEE Trans. Med. Imag. 3.2 (June 1984), 68-79. DOI:
$10.1109 / \mathrm{TMI} .1984 .4307659$ (cit. on pp. 8.13, 8.23).
casey: 86 :qip
[36] M. E. Casey and E. J. Hoffman. "Quantitation in positron emission computed tomography: 7 A technique to reduce noise in accidental coincidence measurements and coincidence efficiency calibration." In: J. Comp. Assisted Tomo. 10.5 (Sept. 1986), 845-50. URL: http: / / gateway.ovid.com/ovidweb.cgi?T=JS\& MODE=Ovid\&NEWS=n\&PAGE=toc\&D=Ovft\&AN=00004728-198609000-00027 (cit. on p. 8.13).
badawi:99: rvr
[37] R. D. Badawi et al. "Randoms variance reduction in 3D PET." In: Phys. Med. Biol. 44.4 (Apr. 1999), 941-54. DOI: 10.1088/0031-9155/44/4/010 (cit. on p. 8.13).
[38] H. H. Barrett et al. "Scattered radiation in nuclear medicine: A case study on the Boltzmann transport equation." In: Computational radiology and imaging. Ed. by F Natterer C Borgers. New York: Springer, 1999, pp. 71-100 (cit. on p. 8.13).
[39] M. Bergström et al. "Correction for scattered radiation in a ring detector positron camera by integral transformation of the projections." In: J. Comp. Assisted Tomo. 7.1 (Feb. 1983), 42-50. URL:
http://gateway.ovid.com/ovidweb.cgi?T=JS\&MODE=ovid\&NEWS=n\&PAGE=toc\&D= ovft\&AN=00004728-198302000-00008 (cit. on p. 8.13).
[40] B. T. A. McKee et al. "A deconvolution scatter correction for a 3-D PET system." In: IEEE Trans. Med. Imag. 11.4 (Dec. 1992), 560-9. DOI: $10.1109 / 42.192692$ (cit. on p. 8.13).
[41] C. E. Floyd et al. "Deconvolution of Compton scatter in SPECT." In: J. Nuc. Med. 26.4 (Apr. 1985), 403-8. URL: http://jnm.snmjournals.org/cgi/content/abstract/26/4/403 (cit. on p. 8.13).
[42] M. A. King et al. "Digital restoration of Indium-111 and Iodine-123 SPECT images with optimized Metz filters." In: J. Nuc. Med. 27.8 (Aug. 1986), 1327-36. URL:
http://jnm.snmjournals.org/cgi/content/abstract/27/8/1327 (cit. on p. 8.13).
[43] J. M. Ollinger. "Model-based scatter correction for fully 3D PET." In: Phys. Med. Biol. 41.1 (Jan. 1996), 153-76. DOI: $10.1088 / 0031-9155 / 41 / 1 / 012$ (cit. on p. 8.13).
[44] C. C. Watson et al. "Evaluation of simulation-based scatter correction for 3-D PET cardiac imaging." In: IEEE Trans. Nuc. Sci. 44.1 (Feb. 1997), 90-7. DOI: $10.1109 / 23.554831$ (cit. on p. 8.13).
[45] S. D. Wollenweber. "Parameterization of a model-based 3-D PET scatter correction." In: IEEE Trans. Nuc. Sci. 49.3 (June 2002), 722-7. DOI: $10.1109 /$ TNS .2002 .1039554 (cit. on p. 8.13).
[46] D. Gagnon et al. "Introduction to holospectral imaging in nuclear medicine for scatter subtraction." In: IEEE Trans. Med. Imag. 8.3 (Sept. 1989), 245-50. DOI: $10.1109 / 42.34713$ (cit. on p. 8.13).
[47] D. Gagnon, N. Pouliot, and L. Laperriere. "Statistical and physical content of low-energy photons in holospectral imaging." In: IEEE Trans. Med. Imag. 10.3 (Sept. 1991), 284-9. DOI: $10.1109 / 42.97577$ (cit. on p. 8.13).
[48] X. Wang and K. F. Koral. "A regularized deconvolution-fitting method for Compton-scatter correction in SPECT." In: IEEE Trans. Med. Imag. 11.3 (Sept. 1992). Corrections Sep. 1993, 351-60. DOI: 10.1109/42.158939 (cit. on p. 8.13).
[49] B. Guérin and G. El Fakhri. "Novel scatter compensation of list-mode PET data using spatial and energy dependent corrections." In: IEEE Trans. Med. Imag. 30.3 (Mar. 2011), 759-73. DOI: 10.1109/TMI. 2010.2095025 (cit. on p. 8.13).
[50] M. Conti, I. Hong, and C. Michel. "Reconstruction of scattered and unscattered PET coincidences using TOF and energy information." In: Phys. Med. Biol. 57.15 (Aug. 2012), N307-. DOI: 10.1088/0031-9155/57/15/N307 (cit. on p. 8.13).
[51] K. F. Koral. "Correction for patient Compton scattering-Current status." In: J. Nuc. Med. 32.6 (June 1991), 1291-3. URL: http://jnm.snmjournals.org/cgi/content/abstract/32/6/1291 (cit. on p. 8.13).
[52] H. Zaidi and K. F. Koral. "Scatter modelling and compensation in emission tomography." In: Eur. J. Nuc. Med. 31.5 (May 2004), 761-82. DOI: $10.1007 / \mathrm{s} 00259-004-1495-\mathrm{z}$ (cit. on p. 8.13).
[53] M. E. Daube-Witherspoon et al. "Scatter correction in maximum likelihood reconstruction of PET data." In: Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf. Vol. 2. 1992, 945-7. DOI: $10.1109 /$ NSSMIC.1992. 301098 (cit. on p. 8.13).
[54] J. E. Bowsher and C. E. F. Jr. "Treatment of Compton scattering in maximum-likelihood, expectation-maximization reconstructions of SPECT images." In: J. Nuc. Med. 32.6 (June 1991), 1285-91. URL: http://jnm.snmjournals.org/cgi/content/abstract/32/6/1285 (cit. on p. 8.14).
[55] E. C. Frey et al. "A fast projector-backprojector pair modeling the asymmetric, spatially varying scatter response function for scatter compensation in SPECT imaging." In: IEEE Trans. Nuc. Sci. 40.4-1 (Aug. 1993), 1192-7. DOI: $10.1109 / 23.256735$ (cit. on p. 8.14).
[56] F. J. Beekman et al. "SPECT scatter modelling in non-uniform attenuating objects." In: Phys. Med. Biol. 42.6 (June 1997), 1133-42. DOI: $10.1088 / 0031-9155 / 42 / 6 / 010$ (cit. on p. 8.14).
[57] F. J. Beekman, C. Kamphuis, and E. C. Frey. "Scatter compensation methods in 3D iterative SPECT reconstruction: A simulation study." In: Phys. Med. Biol. 42.8 (Aug. 1997), 1619-32. DOI: 10.1088/0031-9155/42/8/011 (cit. on p. 8.14).
[58] A. Welch and G. T. Gullberg. "Implementation of a model-based nonuniform scatter correction scheme for SPECT." In: IEEE Trans. Med. Imag. 16.6 (Dec. 1997), 717-26. DOI: $10.1109 / 42.650869$ (cit. on p. 8.14).
[59] G. L. Zeng, C. Bai, and G. T. Gullberg. "A projector/backprojector with slice-to-slice blurring for efficient three-dimensional scatter modeling." In: IEEE Trans. Med. Imag. 18.8 (Aug. 1999), 722-32. DOI: 10.1109/42. 796285 (cit. on p. 8.14).
[60] F. J. Beekman, H. W. A. M. de Jong, and S. van Geloven. "Efficient fully 3D iterative SPECT reconstruction with Monte Carlo based scatter compensation." In: IEEE Trans. Med. Imag. 21.8 (Aug. 2002), 867-77. DOI: $10.1109 /$ TMI. 2002.803130 (cit. on p. 8.14).
[61] R. D. Speller and J. A. Horrocks. "Photon scattering-a 'new' source of information in medicine and biology?" In: Phys. Med. Biol. 36.1 (Jan. 1991), 1-6. DOI: $10.1088 / 0031-9155 / 36 / 1 / 001$ (cit. on p. 8.14).
[62] C. Kamphuis et al. "Dual matrix ordered subsets reconstruction for accelerated 3D scatter compensation in single-photon emission tomography." In: Eur. J. Nuc. Med. 25.1 (Jan. 1998), 8-18. DOI:
10.1007 /s002590050188 (cit. on p. 8.14).
[63] C. Byrne. "Likelihood maximization for list-mode emission tomographic image reconstruction." In: IEEE Trans. Med. Imag. 20.10 (Oct. 2001), 1084-92. DOI: $10.1109 / 42.959305$ (cit. on p. 8.14).
[64] R. Levkovitz et al. "The design and implementation of COSEM, an iterative algorithm for fully 3-D listmode data." In: IEEE Trans. Med. Imag. 20.7 (July 2001), 633-42. DOI: $10.1109 / 42.932747$ (cit. on p. 8.14).
[65] A. J. Reader et al. "One-pass list-mode EM algorithm for high-resolution 3-D PET image reconstruction into large arrays." In: IEEE Trans. Nuc. Sci. 49.3 (June 2002), 693-9. DOI: $10.1109 /$ TNS . 2002.1039550 (cit. on p. 8.14).
[66] A. Rahmim et al. "Motion compensation in histogram-mode and list-mode EM reconstructions: beyond the event-driven approach." In: IEEE Trans. Nuc. Sci. 51.5 (Oct. 2004), 2588-96. DOI: 10.1109/TNS. 2004.835763 (cit. on p. 8.14).
[67] P. Khurd et al. "A globally convergent regularized ordered-subset EM algorithm for list-mode reconstruction." In: IEEE Trans. Nuc. Sci. 51.3 (June 2004), 719-25. DOI: $10.1109 /$ TNS . 2004.829780 (cit. on p. 8.14).
[68] T. E. Nichols et al. "Spatiotemporal reconstruction of list-mode PET data." In: IEEE Trans. Med. Imag. 21.4 (Apr. 2002), 396-404. DOI: $10.1109 /$ TMI .2002 .1000263 (cit. on pp. 8.14, 8.16).
[69] J. M. Jaworski et al. "Model-based reconstruction of spectral and spatial source distribution for objects with known motion." In: IEEE Trans. Nuc. Sci. 60.5 (Oct. 2013), 3981-9. DOI: $10.1109 /$ TNS . 2013.2278283 (cit. on pp. 8.14, 8.15).
[70] D. R. Barr, A. G. Glen, and H. F. Graf. "The "straightforward" nature of arrival rate estimation?" In: American Statistician 52.4 (Nov. 1998), 346-50. URL: http: / /www. jstor.org/stable/2685439 (cit. on p. 8.14).
[71] B-L. Hsu et al. "Reconstruction of rapidly acquired Germanium-68 transmission scans for cardiac PET attenuation correction." In: J. Nucl. Card. 14.5 (Sept. 2007), 706-14. DOI: $10.1016 /$ j.nuclcard. 2007.05 .009 (cit. on p. 8.14).
[72] R. A. Thompson and J. W. LeBlanc. "A novel PET detector block with three dimensional hit position encoding." In: Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf. Vol. 3. 2003, 1718-22. DOI: $10.1109 / \mathrm{NSSMIC} .2003 .1352210$ (cit. on p. 8.14).
[73] S. Ross. Stochastic processes. New York: Wiley, 1983 (cit. on p. 8.15).
hesman: 00:1mm
[74] R. H. Huesman et al. "List mode maximum likelihood reconstruction applied to positron emission mammography with irregular sampling." In: IEEE Trans. Med. Imag. 19.5 (May 2000), 532-7. DOI: 10.1109/42.870263 (cit. on p. 8.18).
[75] R. E. Carson et al. "Design of a motion-compensation OSEM list-mode algorithm for resolution-recovery reconstruction for the HRRT." In: Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf. Vol. 5. 2003, 3281-5. DOI: 10.1109/NSSMIC. 2003.1352597 (cit. on p. 8.19).
[76] J. Qi. "Calculation of the sensitivity image in list-mode reconstruction for PET." In: IEEE Trans. Nuc. Sci. 53.5 (Oct. 2006), 2746-51. DOI: $10.1109 /$ TNS .2006 .882298 (cit. on p. 8.19).
[77] D. Xu. "Gamma-ray imaging and polarization measure using 3-d position-sensitive CdZnTe detectors." PhD thesis. NERS: UM, 2006 (cit. on p. 8.19).
[78] J. Qi and R. H. Huesman. "Propagation of errors from the sensitivity image in list-mode reconstruction." In: IEEE Trans. Med. Imag. 23.9 (Sept. 2004), 1094-9. DOI: $10.1109 /$ TMI. 2004.829333 (cit. on p. 8.19).
[79] M. Yavuz and J. A. Fessler. "New statistical models for randoms-precorrected PET scans." In: Information Processing in Medical Im. Ed. by J Duncan and G Gindi. Vol. 1230. Lecture Notes in Computer Science. Berlin: Springer-Verlag, 1997, pp. 190-203. DOI: 10.1007/3-540-63046-5_15 (cit. on p. 8.22).
[80] M. Yavuz and J. A. Fessler. "Statistical image reconstruction methods for randoms-precorrected PET scans." In: Med. Im. Anal. 2.4 (Dec. 1998), 369-78. DoI: $10.1016 /$ S1361-8415 (98) $80017-0$ (cit. on p. 8.22).
[81] M. Yavuz and J. A. Fessler. "Penalized-likelihood estimators and noise analysis for randoms-precorrected PET transmission scans." In: IEEE Trans. Med. Imag. 18.8 (Aug. 1999), 665-74. DOI: 10.1109/42.796280 (cit. on p. 8.22).
[82] M. Yavuz. "Statistical tomographic image reconstruction methods for randoms-precorrected PET measurements." PhD thesis. Ann Arbor, MI: Univ. of Michigan, Ann Arbor, MI, 48109-2122, 2000. URL: https://hdl.handle. net/2027.42/132500 (cit. on p. 8.22).
[83] M. Yavuz and J. A. Fessler. "Maximum likelihood emission image reconstruction for randoms-precorrected PET scans." In: Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf. Vol. 2. 2000, 15, 229-15, 233. DoI: 10.1109/NSSMIC. 2000.950109 (cit. on p. 8.22).
[84] S. Ahn and J. A. Fessler. "Statistical emission image reconstruction for randoms-precorrected PET scans using negative sinogram values." In: Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf. Vol. 5. 2003, 3062-6. DOI: $10.1109 /$ NSSMIC. 2003.1352544 (cit. on p. 8.22).
[85] S. Ahn and J. A. Fessler. "Emission image reconstruction for randoms-precorrected PET allowing negative sinogram values." In: IEEE Trans. Med. Imag. 23.5 (May 2004), 591-601. DoI: $10.1109 /$ TMI. 2004.826046 (cit. on p. 8.22).
[86] R. J. Smith and J. S. Karp. "A practical method for randoms subtraction in volume imaging PET from detector singles countrate measurements." In: IEEE Trans. Nuc. Sci. 43.3-2 (June 1996), 1981-7. DOI: 10.1109/23.507257 (cit. on p. 8.22).
[87] B. A. Mair, M. Rao, and J. M. M. Anderson. "A minimum chi-squared method for indirect parameter estimation from Poisson data." In: Statistica Neerlandica 56.2 (May 2002), 165-78. DOI: 10.1111/1467-9574.00191 (cit. on p. 8.22).
[88] H. Q. Zhu et al. "A novel weighted least squares PET image reconstruction method using adaptive variable index sets." In: Digital Signal Process 16.2 (Mar. 2006), 106-19. DOI: $10.1016 /$ j.dsp. 2005.04 .011 (cit. on p. 8.22).
[89] A. Staglianò, P. Boccacci, and M. Bertero. "Analysis of an approximate model for Poisson data reconstruction and a related discrepancy principle." In: Inverse Prob. 27.12 (Dec. 2011), p. 125003. DOI: 10.1088/0266-5611/27/12/125003 (cit. on p. 8.22).
[90] M. Kubínová and J. G. Nagy. Robust regression for mixed Poisson-Gaussian model. 2016. URL: http://arxiv.org/abs/1611. 07774 (cit. on p. 8.22).
[91] P. J. Green. "Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives." In: J. Royal Stat. Soc. Ser. B 46.2 (1984), 149-92. URL:
http://www.jstor.org/stable/2345503 (cit. on p. 8.23).
[92] M. M. Ter-Pogossian et al. "Photon time-of-flight-assisted positron emission tomography." In: J. Comp. Assisted Tomo. 5.2 (Apr. 1981), 227-39. URL: http: / / gateway. ovid.com/ovidweb.cgi?T=JS \& MODE=Ovid\&NEWS=n\&PAGE=tOC\&D=Ovft\&AN=00004728-198104000-00014 (cit. on p. 8.23).
[93] M. M. Ter-Pogossian et al. "Super PETT I: A positron emission tomograph utilizing photon time-of-flight information." In: IEEE Trans. Med. Imag. 1.3 (Nov. 1982), 179-87. DOI: $10.1109 /$ TMI. 1982.4307570 (cit. on p. 8.23).
[94] M. Yamamoto, D. C. Ficke, and M. Ter-Pogossian. "Experimental assessment of the gain achieved by the utilization of time-of-flight information in a positron emission tomograph (Super PETT I)." In: IEEE Trans. Med. Imag. 1.3 (Nov. 1982), 187-92. DOI: $10.1109 /$ TMI. 1982.4307571 (cit. on p. 8.23).
[95] W. W. Moses. "Time of flight in PET revisited." In: IEEE Trans. Nuc. Sci. 50.5-2 (Oct. 2003), 1325-30. DOI: 10.1109/TNS. 2003.817319 (cit. on p. 8.23).
[96] D. L. Snyder, L. J. Thomas, and M. M. Ter-Pogossian. "A mathematical model for positron-emission tomography systems having time-of-flight measurements." In: IEEE Trans. Nuc. Sci. 28.3 (June 1981), 3575-83. DOI: 10.1109/TNS. 1981.4332168 (cit. on p. 8.23).
[97] D. G. Politte and D. L. Snyder. "Results of a comparative study of a reconstruction procedure for producing improved estimates of radioactivity distributions in time-of-flight emission tomography." In: IEEE Trans. Nuc. Sci. 31.1 (Feb. 1984), 614-9. DOI: $10.1109 /$ TNS .1984 .4333330 (cit. on p. 8.23).
[98] C. E. Metz, F. B. Atkins, and R. N. Beck. "The geometric transfer function component for scintillation camera collimators with straight parallel holes." In: Phys. Med. Biol. 25.6 (Nov. 1980), 242-50. DOI:
10.1088/0031-9155/25/6/003 (cit. on p. 8.24).


[^0]:    ${ }^{1}$ I mean "large" in the statistical sense that for large $N$ and small success probability, the probability mass function (PMF) of the Binomial distribution approximates that of the Poisson distribution. However, in the biological sense $N$ is tiny relative to the number of atoms in the body, hence a "tracer."
    ${ }^{2}$ We define tracer atoms to be those which were in the metastable state at time $t=0$. Any previously metastable atoms of the radionuclide that have already undergone radioactive decay by time $t=0$ are irrelevant to our considerations.
    ${ }^{3}$ Generalizations to the case of multiple radionuclides, each bound to different compounds, are of interest as well, e.g., [3]. For simplicity we focus the description on the single radionuclide case. However, all the mathematical formulas generalize directly to the case of a spectrum of emitted photon energies [2] simply by letting $\overrightarrow{\mathrm{X}}_{k}(t)$ and $\overrightarrow{\mathrm{x}}$ denote 4-tuples of 3D spatial position and energy, i.e., $\overrightarrow{\mathrm{x}}=(x, y, z, \mathcal{E})$.

[^1]:    ${ }^{4}$ Or triplets of detectors for 3-gamma annihilation imaging [21]

[^2]:    ${ }^{5}$ For example, for a system composed of detectors with multiple recording elements, e.g., [72], the attribute vector could include all of the individual values recorded by each element

