List-mode likelihood

Harrison H. Barrett,
Department of Radiology and Optical Sciences Center, University of Arizona, Tucson, Arizona

Timothy White, and
Idaho National Engineering Laboratory, Pocatello, Idaho

Lucas C. Parra
Siemens Corporate Research Center, Princeton, New Jersey

Abstract

As photon-counting imaging systems become more complex, there is a trend toward measuring more attributes of each individual event. In various imaging systems the attributes can include several position variables, time variables, and energies. If more than about four attributes are measured for each event, it is not practical to record the data in an image matrix. Instead it is more efficient to use a simple list where every attribute is stored for every event. It is the purpose of this paper to discuss the concept of likelihood for such list-mode data. We present expressions for list-mode likelihood with an arbitrary number of attributes per photon and for both preset counts and preset time. Maximization of this likelihood can lead to a practical reconstruction algorithm with list-mode data, but that aspect is covered in a separate paper [IEEE Trans. Med. Imaging (to be published)]. An expression for lesion detectability for list-mode data is also derived and compared with the corresponding expression for conventional binned data.

1. INTRODUCTION

As photon-counting imaging systems become more complex, there is a trend toward measuring more attributes of each individual event. As an example, consider a planar nuclear-medicine imaging system in which the detector is a scintillation camera. This detector measures (or estimates) two coordinates for each gamma ray. For some kinds of scatter correction, the energy of the photon is also estimated and recorded.[1–3] With high photon energies and thick scintillation crystals, it can also be useful to estimate the depth of interaction of the photon in the crystal.[4–6] All of these attributes are estimated from the basic raw data, the photomultiplier signals, and in fact these signals can themselves be regarded as measured attributes of the scintillation event.[7,8]

Additional attributes arise in dynamic and tomographic imaging. One way of conducting a dynamic study in nuclear medicine is to record the time of occurrence for each event. Similarly, in single-photon emission computed tomography (SPECT) systems, it is necessary, at a minimum, to record the projection angle along with the event coordinates.

The number of attributes increases further in a fully three-dimensional positron emission tomography (PET) system with two scintillation cameras. There the minimal set of attributes consists of four coordinates (two for each of the coincident photons) plus a rotation angle. In addition, the attribute set might include estimates of the energies of each photon, depth of

interaction, or time-of-flight difference. Similarly, in a Compton camera[9] each primary gamma-ray photon produces a Compton-scattered photon, and the coordinates and energy of both primary and secondary photon are measured. Thus the attribute set consists of at least four measured coordinates and two energies, and two more coordinates can be measured with thick detectors.

The concept of measuring multiple attributes for each event is not restricted to gamma rays; optical photon-counting detectors with multiple outputs exist as well. A position-sensitive photomultiplier, for example, can have multiple anodes. If optical photons are incident on the photocathode and well resolved temporally, so that distinct anode signals are obtained for each photon, then the signal on each anode can be regarded as an attribute of a single optical photon.

It is clear from these examples that a substantial number of attributes can be measured for each detected event. One way of recording these data is to bin them into one large data matrix, with one index for each attribute. This method of data recording encounters difficulties as the number of attributes increases, since the number of elements in the data matrix can be huge. If N attributes are measured to a precision of B bits each, there must be 2^{NB} elements. If we assign one byte to each element, we can acquire a maximum of 255 events in one bin. With four attributes, we thus require approximately 2^{24} bytes (16 megabytes) of storage if each attribute is measured with 6-bit precision or 2^{32} bytes (4 gigabytes) for 8-bit precision. With more than four attributes it is usually out of the question to assign one bin to each possible combination of measured attributes. Even if we could afford the storage, it would be very inefficient to have many bins with no recorded events.

If we attempted to reduce the storage requirements by reducing the number of bits per attribute, there would be a danger of information loss. As an extreme example, photon energy in a scintillation camera is often reduced to a single bit, set to one if the estimated energy lies in a preset window. There is evidence that this results in a loss of image quality as measured by lesion detectability.[8,10]

Another form of data reduction is to use the initial set of measured attributes to estimate values for some smaller set of attributes. From the photomultiplier outputs in a scintillation camera, for example, we can estimate the coordinates and energy of each photon, reducing the number of attributes from the number of photomultipliers to three. Though this practice is virtually universal, it is difficult to establish in general that it entails no information loss.

In an important alternative mode of data storage, called list mode, the measured attributes of each event are simply stored in a list. If J events are observed and N attributes are measured for each, then NJ memory locations are required. Each location can be one byte if 8-bit precision is adequate for each attribute. Whenever 2^{NB} exceeds NJ, as it must for large N, list mode is more efficient than binning. Moreover, since all measured attributes are recorded for later processing, there is no loss of information in the data-acquisition stage.

The goal of this paper is to present a comprehensive treatment of the important concept of likelihood for list-mode data. A familiar use of likelihood is in maximum-likelihood parameter estimation or object reconstruction, and the theory presented here provides the mathematical basis for these applications. Another use of likelihood is in signal detection and discrimination problems, where it is known that the likelihood ratio is the optimum test statistic. Performance on these detection and discrimination tasks can then be used for the objective assessment of image quality.

In Section 2 we consider three modes of representing data from photon-counting imaging systems: conventional binning, list mode, and an impulse-valued random process. We present the relevant multivariate probability distributions needed to describe the data in each mode.

In Section 3 we use the statistical distributions developed in Section 2 to derive expressions for list-mode likelihood of the data given a particular object. We consider separately the situations where data are acquired for a preset time or a preset number of counts. We have used one of these expressions to develop a maximum-likelihood reconstruction algorithm, but the algorithm itself is derived and discussed in a separate paper.[11] Previous work on maximum-likelihood reconstruction from list-mode data has been published by Snyder[12] and Snyder and Politte,[13] and it is a subject of current interest in high-energy physics.

In Section 4 we derive expressions for the likelihood ratio and detectability index for list-mode data when the task is detection of a nonrandom object or discrimination between two such objects. The results are compared with previously published expressions for binned data.

2. STATISTICAL MODELS

A. Kinds of Data Sets

In any event-counting system, data can be collected either for a given time or until a given number of events is reached. In the nuclear-medicine literature, these two methods are referred to as preset time and preset counts, respectively, and we adopt that terminology here also. The key distinction is that the total number of events J is a random variable for preset time but a fixed number for preset counts. Another possible data set could be obtained by collecting a preset number of counts but also recording the (random) time required to reach this number. This option is rarely used in practice and is not treated here.

Let $\mathbf{r}_j, j = 1 \ldots J$, denote the N-dimensional attribute vector for the jth event. The nth component ($n = 1 \ldots N$) of the vector $\mathbf{r}_j$ will be denoted x_{jn}. In the simple case of planar imaging, $N = 2$ and x_{j1} and x_{j2} are the Cartesian coordinates of a point in the image plane representing event j. For the more interesting situation where $N > 2$, the components x_{jn} do not necessarily signify a physical position, but they can be regarded as Cartesian coordinates in an N-dimensional hyperspace which we call attribute space.

There are three different ways of representing these data. The simplest is the attribute list $\{\mathbf{r}_j, j = 1 \ldots J\}$ plus J itself if data are collected for a preset time. The other two modes, conventional binned data and an impulse-valued random process, are easily constructed from the list.

In a binned image, it is convenient to use a single index m to denote the bin rather than using one index for each component of the attribute vector. The jth event is assigned to bin m if $X_{mn} - \frac{1}{2}\varepsilon_n \leq x_{jn} < X_{mn} + \frac{1}{2}\varepsilon_n$ for all n, where X_{mn} is the nth Cartesian coordinate of the center of the mth bin in attribute space and ε_n is the bin width associated with the nth attribute. Thus the center of the mth bin is specified by an N-dimensional vector $\mathbf{X}_m$. For $N = 2$ with the attributes being positions, $\mathbf{X}_m$ is a two-dimensional vector in the image plane centered on the mth pixel. In that case, we usually take $\varepsilon_1 = \varepsilon_2 = \varepsilon$, which is the pixel width. For an energy attribute, ε_n is the width of the energy window.

The number of bins associated with the nth attribute is M_n, given by the range of allowed values of x_{jn} divided by the bin width ε_n. In practice, M_n may be as small as 1 for an energy attribute or as large as 256 or 512 for spatial coordinates.

After all J events have been binned, a total of g_m events will have accumulated in bin m. We can denote the binned data set by an $M \times 1$ data vector $\mathbf{g}$ with components $g_m, m = 1 \ldots M$. The total number of bins is

$$M = \prod_{n=1}^{N} M_n.$$

(1)

As noted in the introduction, M can be huge even for modest N, so the binned data representation may not be feasible for N larger than 4 or so.

Another representation of the data, very useful for theoretical analysis, is the impulse-valued random process.[14] In this representation, we assign an N-dimensional Dirac delta function to each event in the list. The resulting random process is a generalized function in attribute space defined by

$$g(\mathbf{r}) = \sum_{j=1}^{J} \delta(\mathbf{r} - \mathbf{r}_j).$$

(2)

This function is parameterized by the random attributes $\{\mathbf{r}_j\}$ and by J itself.

Given the random process $g(\mathbf{r})$, we can obtain the binned data vector $\mathbf{g}$ simply by integrating. The number of events in bin m is given by

$$g_m = \int_{\text{bin } m} g(\mathbf{r}) \mathrm{d}^N r,$$

(3)

where $\mathrm{d}^N r$ is a volume element in attribute space and the integral is over the region of attribute space associated with the mth bin.

Equation (3) demonstrates a key difference between $g(\mathbf{r})$ and g_m: the latter is a pure number, while the former must have dimensions associated with it. Just what these dimensions are depends on the specific attributes that make up $\mathbf{r}$. In the simplest case of $N = 2$, with both attributes being spatial coordinates, $\mathrm{d}^N r$ has dimensions of area, so $g(\mathbf{r})$ must have dimensions of reciprocal area.

B. Statistical Independence

The key assumption in the analysis that follows is that individual events are statistically independent. Although we often take this condition for granted in photon-counting problems, there are several important situations that could invalidate it.

The first is detector saturation, manifested as dead time or loss of resolution at high counting rates. If one photon temporarily paralyzes the detector and there is a significant probability of another photon arriving before it recovers, the probability of detection of the second photon is dependent on the presence of the first. Even if the second photon is detected, the transient response of the detector or the electronics may cause errors in the measured position, energy, or other attributes of the second photon. We shall neglect all of these effects in this paper, which amounts to restricting the analysis to relatively low count rates.

Statistical independence also fails in random multiplication processes where one primary event gives rise to a random number of secondary events.[14,15] In a scintillation detector, for example, a single gamma-ray photon produces a large number of optical photons, and these secondary events are not statistically independent since they arise from the same gamma-ray photon.[16]

One can conceive of systems that measure attributes of individual secondary events, though the authors know of no current systems that do so. A scintillator could be viewed by an optical photon-counting imaging system, for example, but currently such systems do not have sufficient temporal resolution to report coordinates or other attributes for the individual optical photons. Since this paper is concerned only with counting systems where multiple attributes of individual events are measured, we rule out the possibility that the events under consideration are secondaries associated with a single primary event.

On the other hand, random multiplication processes may be present in the systems considered here, scintillation cameras being a prime example. The distinction is that the attributes being measured are properties of the primary event, and the randomness of the secondaries simply leads to error in the estimates of the attributes. As long as the primary events (the gamma rays in the case of a scintillation camera) are statistically independent, the analysis given here is valid.

Another problem that can invalidate the independence assumption is randomness in the object being imaged. There are two different statistical ensembles that we might consider. The first is all realizations of the random data set for one particular object; the second allows the objects themselves to be drawn at random from some distribution. In the latter case, the independence assumption may hold conditionally for a fixed object but not when the ensemble of objects is taken into account. Object randomness is not considered in this paper.

C. Statistical Properties of List-Mode Data

We adopt a discrete object model and represent the object by a $K \times 1$ vector $\mathbf{f}$. The kth component of $\mathbf{f}$, denoted f_k, is the mean number of photons per second emitted from the kth voxel (volume element) of the object. Though we regard $\mathbf{f}$ as nonrandom, it is useful to write the data probabilities as conditional on $\mathbf{f}$ as a way of emphasizing the dependence of the data on the object.

The random vector $\mathbf{r}_j$ is the result of a measurement of the set of attributes associated with an individual event. As with any measurement, there can be both systematic and random errors. In the language of estimation theory, there is both a bias and a variance associated with the estimate or measurement. If we denote the true attribute vector for the jth event by $\mathbf{R}_j$, we can represent the measurement in full generality as

$$\mathbf{r}_j = \mathbf{R}_j + \mathbf{b}_j + \eta_j, \tag{4}$$

where $\mathbf{b}_j$ is the bias or systematic error and η_j is the random error.

Thus the measured attribute vector $\mathbf{r}_j$ has two random components, $\mathbf{R}_j$ and η_j. The statistics of the true attribute vector $\mathbf{R}_j$ are determined by the object $\mathbf{f}$ and the image-forming elements (collimator or lens, for example). The bias and the random error, on the other hand, are associated with the measurement process, including the detector, the electronics, and any subsequent data processing. It is reasonable to assume that the bias and the random error depend on $\mathbf{R}_j$ but not directly on $\mathbf{f}$.

With these considerations in mind, we write the probability density function for $\mathbf{r}_j$ as

$$\mathrm{pr}(\mathbf{r}_j|\mathbf{f}) = \int_{\mathrm{att}} \mathrm{d}^N R_j \, \mathrm{pr}_{\mathrm{det}}(\mathbf{r}_j|\mathbf{R}_j) \, \mathrm{pr}_{\mathrm{im}}(\mathbf{R}_j|\mathbf{f}), \tag{5}$$

where the integral is over the range of each component in attribute space, $\text{pr}_{\text{det}}(\mathbf{r}_j|\mathbf{R}_j)$ is the conditional probability density function for $\mathbf{r}_j$ given that event j has true attribute vector $\mathbf{R}_j$, and $\text{pr}_{\text{im}}(\mathbf{R}_j|\mathbf{f})$ is the conditional probability density function for $\mathbf{R}_j$ given an object $\mathbf{f}$. The subscripts on the densities imply that $\text{pr}_{\text{det}}(\mathbf{r}_j|\mathbf{R}_j)$ represents characteristics of the detector system and $\text{pr}_{\text{im}}(\mathbf{R}_j|\mathbf{f})$ represents the image-forming system.

In practice, the two factors in the integrand in Eq. (5) can be calculated from an analytical or numerical model of the detector and imaging system. The second factor, $\text{pr}_{\text{im}}(\mathbf{R}_j|\mathbf{f})$, is computed from knowledge of the deterministic laws of propagation of photons from source to detector, and the first factor, $\text{pr}_{\text{det}}(\mathbf{r}_j|\mathbf{R}_j)$, requires a model of the random estimation errors, which are associated with $\mathbf{b}_j$ and $\boldsymbol{\eta}_j$ in Eq. (4). A specific example of how such a computation is performed is given by Parra and Barrett.[11]

Since nothing distinguishes one photon from another, $\text{pr}(\mathbf{r}_j|\mathbf{f})$ has the same functional form for all j. The multivariate probability density function for the list-mode data with preset counts is then given by

$$\text{pr}(\{\mathbf{r}_j\}|\mathbf{f}, J)=\prod_{j=1}^{J}\text{pr}(\mathbf{r}_j|\mathbf{f}).$$

(6)

For an acquisition with preset time, the list-mode data set consists of $J + 1$ random variables, namely, all of the $\mathbf{r}_j$ and J itself. The probability law for this data set is

$$\begin{aligned}\text{pr}(\{\mathbf{r}_j\}, J|\mathbf{f})&=\text{pr}(\{\mathbf{r}_j\}|\mathbf{f}, J)\text{Pr}(J|\mathbf{f})\\&=\text{Pr}(J|\mathbf{f})\prod_{j=1}^{J}\text{pr}(\mathbf{r}_j|\mathbf{f}).\end{aligned}$$

(7)

The notation is a bit tricky here since $\text{pr}(\{\mathbf{r}_j\}, J|\mathbf{f})$ is a probability *density* function for each of the $\mathbf{r}_j$ but a probability for the discrete random variable J. Usually we denote probabilities Pr $(\cdot)$ and probability density functions pr $(\cdot)$, but in a mixed case like $\text{pr}(\{\mathbf{r}_j\}, J|\mathbf{f})$ we still use the lower case.

For later convenience we define two unnormalized densities $h(\mathbf{r}_j)$ and $\bar{h}(\mathbf{r}_j)$ by

$$h(\mathbf{r}_j)=J\,\text{pr}(\mathbf{r}_j|\mathbf{f}),$$

(8)

$$\bar{h}(\mathbf{r}_j)=\bar{J}\,\text{pr}(\mathbf{r}_j|\mathbf{f}),$$

(9)

where $\bar{J}$ is the mean number of events, averaged over many acquisitions with the same object and the same preset time:

$$\bar{J}=\sum_{J=0}^{\infty}J\,\text{Pr}(J|\mathbf{f}).$$

(10)

For preset counts, $h(\mathbf{r})\mathrm{d}^N r$ is the mean number of events with attributes in the differential volume $\mathrm{d}^N r$ centered at point $\mathbf{r}$ in attribute space. A similar interpretation applies to $\bar{h}(\mathbf{r})$ with preset time.

D. Statistical Properties of Binned Data

In this section we review some well-known results concerning statistics of binned data. The relation of these results to list-mode data will be discussed in Subsection 2.F.

The probability that event j is recorded in bin m equals the probability that $\mathbf{r}_j$ falls in the region of attribute space associated with that bin; this probability is given by the integral

$$\mathrm{Pr}(\mathbf{r}_j \text{ in bin } m) = \int_{\text{bin } m} \mathrm{pr}(\mathbf{r}_j|\mathbf{f})\mathrm{d}^N r_j \equiv \alpha_m. \tag{11}$$

Since we have assumed that $\mathrm{pr}(\mathbf{r}_j|\mathbf{f})$ is the same for all j, we can denote $\mathrm{Pr}(\mathbf{r}_j$ in bin $m)$ simply as α_m, which is the probability of any photon being recorded in bin m.

For a preset-count data acquisition, the mean number of counts in bin m is simply

$$\bar{g}_m = \alpha_m J = \int_{\text{bin } m} h(\mathbf{r})\mathrm{d}^N r \quad \text{(preset counts)}. \tag{12}$$

Since the events are independent and the total number is fixed, the univariate probability law for g_m in this case is a binomial,

$$\mathrm{Pr}(g_m|\mathbf{f}, J) = \frac{J!}{(g_m!)(J - g_m)!}(\alpha_m)^{g_m}(1 - \alpha_m)^{J-g_m}, \tag{13}$$

and the corresponding multivariate law for the entire data vector $\mathbf{g}$ is a multinomial,

$$\mathrm{Pr}(\mathbf{g}|\mathbf{f}, J) = J! \prod_{m=1}^{M} \frac{(\alpha_m)^{g_m}}{g_m!}. \tag{14}$$

Note that the g_m are not statistically independent, because their sum must be J.

For preset time, we can write

$$\mathrm{Pr}(\mathbf{g}|\mathbf{f}) = \sum_{J=0}^{\infty} \mathrm{Pr}(\mathbf{g}|\mathbf{f}, J)\mathrm{Pr}(J|\mathbf{f}). \tag{15}$$

If J is a Poisson random variable, as it almost invariably is in practical photon-counting problems, this sum can readily be performed.[14] The result is that the g_m are also Poisson and statistically independent, with a probability law given by

$$\Pr(\mathbf{g}|\mathbf{f})=\prod_{m=1}^{M}\exp(-\bar{g}_m)\frac{(\bar{g}_m)^{g_m}}{g_m!}$$

(16)

where now

$$\bar{g}_m=\alpha_m\bar{J}=\int_{\text{bin } m}\bar{h}(r)\text{d}^N r \quad \text{(preset time)}.$$

(17)

Thus, with preset time and the assumption that J is a Poisson random variable, the full probability law $\Pr(\mathbf{g}|\mathbf{f})$ is determined by knowledge of $h(r)$.

E. Statistical Properties of the Impulse-Valued Random Process

The statistics of a spatial or temporal, impulse-valued, Poisson random process are well understood.[14,17] In this section we extend these properties to a more general attribute space.

Since $g(\mathbf{r})$ is a generalized function with no finite values other than zero, a probability density function does not have much meaning. Instead we shall discuss the first- and second-order statistics of $g(\mathbf{r})$, or its mean and autocorrelation function.

The conditional expectation of $g(\mathbf{r})$, given $\mathbf{f}$ and J, is

$$E\{g(\mathbf{r})|\mathbf{f}, J\}=\int_{\text{att}}\text{d}^N r_1 \text{ pr}(\mathbf{r}_1|\mathbf{f})\int_{\text{att}}\text{d}^N r_2 \text{ pr}(\mathbf{r}_2|\mathbf{f})\cdots\times\int_{\text{att}}\text{d}^N r_J \text{ pr}(\mathbf{r}_J|\mathbf{f})g(\mathbf{r}).$$

(18)

The procedure for performing this kind of average is detailed by Barrett and Swindell[14]; the result is

$$E\{g(\mathbf{r})|\mathbf{f}, J\}=h(\mathbf{r}).$$

(19)

If we further average over J in a preset-time mode and assume that J is Poisson, we find[14]

$$E\{g(\mathbf{r})|\mathbf{f}\}=\bar{h}(\mathbf{r}).$$

(20)

The nonstationary autocorrelation function of $g(\mathbf{r})$ is defined by

$$\Gamma(\mathbf{r}, \mathbf{r}')=\left\langle g(\mathbf{r})g(\mathbf{r}')\right\rangle,$$

(21)

where the angle brackets denote averaging over the set $\{\mathbf{r}_j\}$ and, for preset time, J itself. For preset time and Poisson J, a calculation analogous to the one in Ref. [14] shows that

$$\Gamma(\mathbf{r}, \mathbf{r}')=\bar{h}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')+\bar{h}(\mathbf{r})\bar{h}(\mathbf{r}').$$

(22)

Counterparts of Eqs. (18)–(22) for spatial or temporal random processes are derived by Barrett and Swindell,[14] among other sources. The main point of this section, however, is that they apply in attribute space as well.

F. Some Interrelations

We have considered two kinds of data acquisition (preset counts and preset time) and three kinds of data representation (list, bins, and random process). In this section we point out some connections among the results.

First, with binned data the distinction between preset time and preset counts is not great if the total number of bins M is large and the number of events J is also large. If M is large, chances are that no single bin will have a large probability α_m of getting a particular event. By Eq. (12), the mean number in bin m with present counts is $\bar{g}_m = \alpha_m J$, but if $\alpha_m \to 0$ and $J \to \infty$ in such a way that $\alpha_m J$ remains constant, it is well known[18] that the multinomial law of Eq. (14) approaches the Poisson of Eq. (16), originally derived for preset time. Conversely, if J is Poisson and $\bar{J}$ is large, the standard deviation of J is small compared with its mean, so it does not matter much if J is fixed.

The distinction between binned data and list mode disappears if the size of each bin is made small enough, since in that case the average number of counts in any bin becomes much less than one. Then the actual random number of events recorded in a bin is either 0 or 1 with high probability, and the list of attribute vectors is simply a list of addresses of bins with one count. This limit is not a practical one, since it is wasteful of memory, but it shows the theoretical relation between binning and list mode: binned data approaches a list as bin size tends to zero.

Also, the density functions $h(\mathbf{r})$ and $\bar{h}(\mathbf{r})$ play a key role in the statistics of all three kinds of data. These functions were originally introduced as unnormalized versions of $\mathrm{pr}(\mathbf{r}_j|\mathbf{f}, J)$ and $\mathrm{pr}(\mathbf{r}_j|\mathbf{f})$, respectively, for list-mode data [see Eqs. (8) and (9)]. In the random-process description of the data, these same functions reappeared with a different interpretation. In that case they are directly the mean values of the impulse-valued random process [see Eqs. (19) and (20)], and they determine the autocorrelation function, at least for preset time and Poisson counts [see Eq. (22)].

In the case of binned data, $h(\mathbf{r})$ and $\bar{h}(\mathbf{r})$ have yet another meaning: their integrals determine the mean number of counts in each bin [see Eqs. (12) and (17)]. Moreover, if the bins are small enough that these density functions do not vary appreciably over a bin width, we can approximate the integrals as

$$\bar{g}_m \simeq h(\mathbf{X}_m)\prod_{n=1}^{N}\varepsilon_n,$$

(23)

where $\mathbf{X}_m$ and ε_n are defined in Subsection 2.A. Thus, except for a constant of proportionality, the sampled value $h(\mathbf{X}_m)$ is approximately the mean number of counts in bin m for preset counts, and $\bar{h}(\mathbf{X}_m)$ has the same meaning for preset time.

Comparison of Eqs. (7) and (15) reveals an interesting distinction between list-mode and binned data. The list-mode data set consists of $J + 1$ random variables, the attribute vectors plus J itself, so the probability law in Eq. (7) includes J as a random variable. With binned data, however, J is not a separate random variable since it is the sum of the g_m. Thus a sum over J appears in Eq. (15) but not in Eq. (7). When Eq. (7) is used to compute expectations, however, the sum over J reappears since J is then a random variable to be averaged over.

3. MAXIMUM-LIKELIHOOD IMAGE RECONSTRUCTION

Suppose we are given a data vector **g** described by a probability density function pr(**g**|$\boldsymbol{\theta}$), where the vector $\boldsymbol{\theta}$ is a set of unknown parameters. For a given realization of the data, the likelihood $L(\boldsymbol{\theta})$ of the parameter vector $\boldsymbol{\theta}$ is simply pr(**g**|$\boldsymbol{\theta}$), regarded as a function of $\boldsymbol{\theta}$ with **g** fixed at the observed value. Parameter estimation is frequently performed by choosing the parameter vector $\boldsymbol{\theta}$ that maximizes the likelihood. The resulting $\boldsymbol{\theta}$, denoted $\hat{\boldsymbol{\theta}}$, is called the maximum-likelihood estimator of $\boldsymbol{\theta}$. It has many desirable properties, especially in an asymptotic sense.[19−21] We can apply this concept to image reconstruction by regarding the components of **f** as the parameters to be estimated.[22]

A. Binned Data

For binned data, the likelihood is given by Eq. (14) or (16), where the dependence on **f** is contained in α_m or $\bar{g}_m$. To make this dependence explicit, we assume that the imaging system is linear, so that we can write

$$\bar{g}_m = \sum_{k=1}^{K} H_{mk} f_k = (\mathbf{Hf})_m,$$

(24)

where H_{mk} is an element of an $M \times K$ matrix **H** describing the imaging system, and $(\mathbf{Hf})_m$ is the mth component of the vector **Hf**. Since f_k is the mean number of photons emitted per second, H_{mk} is proportional to the acquisition time in a preset-time mode. For preset counts, H_{mk} is proportional to J.

With this system model, the likelihood for binned data and preset counts is given, from Eqs. (12), (14), and (24), by

$$L_{\text{bin}}(\mathbf{f}|J) = J! \prod_{m=1}^{M} \frac{[(\mathbf{Hf})_m / J]^{g_m}}{g_m!}.$$

(25)

For preset time, and with the assumption that J is a Poisson random variable, the likelihood for binned data is [see Eq. (16)]

$$L_{\text{bin}}(\mathbf{f}) = \sum_{J=0}^{\infty} \text{Pr}(J|\mathbf{f}) L_{\text{bin}}(\mathbf{f}|J)$$
$$= \prod_{m=1}^{M} \exp[-(\mathbf{Hf})_m] \frac{[(\mathbf{Hf})_m]^{g_m}}{g_m!}.$$

(26)

With this form of the likelihood, a popular method for finding the maximum-likelihood estimator of **f** is the expectation-maximization algorithm.[23,24] This algorithm finds the vector **f** that maximizes $L_{\text{bin}}(\mathbf{f})$ (or, equivalently, its logarithm), subject to the constraint that all components f_k be nonnegative.

B. List-Mode Data

For list-mode data, the likelihood is given by Eq. (6) or (7). The dependence on **f** in these equations is contained in pr($\mathbf{r}_j$|**f**), which is related to the unnormalized densities $h(\mathbf{r}_j)$ and $\bar{h}$

($\mathbf{r}_j$) in Eqs. (8) and (9). It follows from Eqs. (6) and (8) that the likelihood for list-mode data and preset counts is

$$L_{\text{list}}(\mathbf{f}|J) = J^{-J} \prod_{j=1}^{J} h(\mathbf{r}_j).$$

(27)

A constant factor such as J^{-J} is irrelevant in most applications of the likelihood, so Eq. (27) shows that the list-mode likelihood for preset counts is simply the product of the densities h ($\mathbf{r}_j$) for each event.

For preset time, Eqs. (7) and (27) show that

$$L_{\text{list}}(\mathbf{f}) = \Pr(J|\mathbf{f}) L_{\text{list}}(\mathbf{f}|J) = \Pr(J|\mathbf{f}) J^{-J} \prod_{j=1}^{J} h(\mathbf{r}_j).$$

(28)

In contrast to Eq. (26), no sum over J appears in this equation, since J is a separate random variable in list mode.

To express the list-mode likelihoods as explicit functions of $\mathbf{f}$, we note that

$$\text{pr}(\mathbf{r}_j|\mathbf{f}) = \sum_{k=1}^{K} \text{pr}(\mathbf{r}_j|j \text{ from } k) \Pr(j \text{ from } k|\mathbf{f}),$$

(29)

where "j from k" is shorthand indicating that the jth event originated with the emission of a photon from the kth voxel. Since all photons are equivalent, the probability that any one of them originated from the kth voxel is proportional to the object strength associated with that voxel, so $\Pr(j \text{ from } k|\mathbf{f}) \propto f_k$.

To determine the constant of proportionality in general, we must consider the possibility that the overall system sensitivity can vary with voxel location. The sensitivity S_k is the probability that a photon emitted from voxel k is detected, independent of what attribute vector is assigned to it by the detector. Note that information about S_k is *not* contained in $\text{pr}(\mathbf{r}_j|j \text{ from } k)$, since the latter is the probability density on $\mathbf{r}_j$ given that the photon was emitted from voxel k *and detected*. Thus S_k is a separate system specification. In many practical photon-counting systems, it is reasonable to take S_k as a constant, but we shall leave it general.

The probability that the jth event originated in the kth voxel is the mean number of photons emitted from the kth voxel and detected divided by the total number emitted and detected. The mean number emitted from the voxel per unit time is f_k, and the probability that an emitted photon is detected is S_k. Thus, for all j,

$$\Pr(j \text{ from } k|\mathbf{f}) = \frac{f_k S_k}{\sum\limits_{k=1}^{K} f_k S_k}.$$

(30)

Equation (29) now becomes

$$\text{pr}(\mathbf{r}_j|\mathbf{f}) = \frac{\sum_{k=1}^{K} f_k S_k \, \text{pr}(\mathbf{r}_j|j \text{ from } k)}{\sum_{k=1}^{K} f_k S_k}.$$

(31)

From Eqs. (27), (29) and (31), the list-mode likelihood for preset counts is

$$L_{\text{list}}(\mathbf{f}|J) = \frac{\prod_{j=1}^{J} \left[\sum_{k=1}^{K} f_k S_k \, \text{pr}(\mathbf{r}_j|j \text{ from } k) \right]}{\left(\sum_{k=1}^{K} f_k S_k \right)^{J}}.$$

(32)

The corresponding log likelihood is

$$\log[L_{\text{list}}(\mathbf{f}|J)] = -J\log\left(\sum_{k=1}^{K} f_k S_k \right) + \sum_{j=1}^{J} \log\left[\sum_{k=1}^{K} f_k S_k \, \text{pr}(\mathbf{r}_j|j \text{ from } k) \right].$$

(33)

For preset time, the list-mode likelihood is

$$L_{\text{list}}(\mathbf{f}) = \text{Pr}(J|\mathbf{f}) L_{\text{list}}(\mathbf{f}|J)$$
$$= \frac{\text{Pr}(J|\mathbf{f})}{\left(\sum_{k=1}^{K} f_k S_k \right)^{J}} \prod_{j=1}^{J} \left[\sum_{k=1}^{K} f_k S_k \, \text{pr}(\mathbf{r}_j|j \text{ from } k) \right].$$

(34)

But note that the mean number of detected counts in a preset-time acquisition is given by

$$\bar{J} = \tau \sum_{k=1}^{K} S_k f_k,$$

(35)

where τ is the acquisition time. Thus

$$L_{\text{list}}(\mathbf{f}) = \frac{\text{pr}(J|\mathbf{f})}{(\bar{J})^{J}} \prod_{j=1}^{J} \left[\sum_{k=1}^{K} \tau f_k S_k \, \text{pr}(\mathbf{r}_j|j \text{ from } k) \right]$$
$$= \frac{1}{J!} \exp(-\bar{J}) \prod_{j=1}^{J} \left[\sum_{k=1}^{K} \tau f_k S_k \, \text{pr}(\mathbf{r}_j|j \text{ from } k) \right],$$

(36)

where the last step follows by use of a Poisson law for $\Pr(J|\mathbf{f})$. The log likelihood is now

$$\log[L_{\mathrm{list}}(\mathbf{f})] = -\log(J!) - \bar{J} + \sum_{k=1}^{J} \log\left(\sum_{k=1}^{K} \tau f_k S_k \, \mathrm{pr}(\mathbf{r}_j | j \text{ from } k)\right).$$

(37)

The first term in this expression, $-\log(J!)$, is an irrelevant constant, but $\mathbf{f}$ appears in the other two terms. (One must resist the temptation to approximate $\bar{J}$ with the observed J in the second term, since that would throw out an essential dependence on $\mathbf{f}$, the variable of interest in the reconstruction). Thus maximum-likelihood reconstruction from list-mode, preset-time data consists of maximizing the sum of the second and third terms in Eq. (36), subject to the constraint that all of the f_k be nonnegative. An algorithm for doing this is given in a separate paper.[11]

4. HYPOTHESIS TESTING

A. Discrimination of Nonrandom Signals

Objective assessment of image quality can be based on the ability of an ideal observer to perform a specified task, using the image data.[21,25] One particular task that has received considerable attention for this purpose is detection of an exactly known, nonrandom signal. [26] For the ideal observer, this task is essentially equivalent to discrimination between two specified nonrandom signals. It amounts to testing the binary hypothesis that either signal 1 or signal 2 is present; in the detection problem, signal 1 is zero. In an imaging context, the signals in question are the objects being imaged, denoted $\mathbf{f}$ in this paper. Thus the task is to determine whether $\mathbf{f}_1$ or $\mathbf{f}_2$ is present. Performance on this task can be measured by the area under a receiver operating characteristic (ROC) curve, or, equivalently, by the detectability index d to be defined below.

It is well known[19,20,25] that the optimum strategy for performing a binary discrimination task is to first compute the likelihood ratio λ, defined by

$$\lambda = \frac{L(\mathbf{f}_2)}{L(\mathbf{f}_1)} = \frac{\mathrm{pr}(\mathbf{g}|\mathbf{f}_2)}{\mathrm{pr}(\mathbf{g}|\mathbf{f}_1)}.$$

(38)

The discrimination is then performed by comparing λ with a threshold λ_{th} and choosing the hypothesis that $\mathbf{f}_2$ is present if $\lambda > \lambda_{\mathrm{th}}$. Equivalently, we can compute the log of the likelihood ratio,

$$\ell = \log(\lambda)$$

(39)

and compare it with $\log(\lambda_{\mathrm{th}})$. The result is the same, and the log is often more convenient mathematically.

The ROC curve is generated by varying λ_{th} and, for each value, plotting the true-positive rate (probability of choosing $\mathbf{f}_2$ when it is actually present) versus the false-positive rate (probability of choosing $\mathbf{f}_2$ when $\mathbf{f}_1$ is actually present). In a detection problem where $\mathbf{f}_1$ is zero, the true-positive rate is the probability of detection and the false-positive rate is the false-alarm rate.

The detectability index d is defined by

$$d^2 = \frac{[E(\ell|\mathbf{H}_2) - E(\ell|\mathbf{H}_1)]^2}{\frac{1}{2}\text{var}(\ell|\mathbf{H}_1) + \frac{1}{2}\text{var}(\ell|\mathbf{H}_2)}, \tag{40}$$

where $E(\ell|\mathbf{H}_i)$ is the expected value of ℓ, given that hypothesis i ($i = 1, 2$) is true, and var($\ell|$ $\mathbf{H}_i$) is the corresponding variance. If ℓ is a normal random variable, which we can often argue on the basis of the central-limit theorem, then the area under the ROC curve is uniquely related to d.

B. Binned Data

The log of the likelihood ratio can easily be constructed from any of the likelihood expressions given in Section 3. For example, with binned data and preset time, it is given by

$$\ell_{bin} = \log[\lambda_{\text{bin}}(\mathbf{f}_2)] - \log[\lambda_{\text{bin}}(\mathbf{f}_2)]$$
$$= \sum_{m=1}^{M} \{(\mathbf{Hf}_1)_m - (\mathbf{Hf}_2)_m + g_m\log[(\mathbf{Hf}_2)_m] - g_m\log[(\mathbf{Hf}_1)_m]\}. \tag{41}$$

Terms independent of the data $\mathbf{g}$ can be lumped into the threshold without affecting the ROC curve, so ℓ_{bin} becomes

$$\ell_{\text{bin}} = \sum_{m=1}^{M} g_m\log\left[\frac{(\mathbf{Hf}_2)_m}{(\mathbf{Hf}_1)_m}\right] + \text{dull terms}. \tag{42}$$

This form shows that ℓ_{bin} can be realized as a linear filter where each datum g_m is multiplied by a weight given by the logarithmic expression in Eq. (42). When the weight is simply the difference of the means under the two hypotheses, the process is called matched filtering; the present process is therefore referred to as logarithmic matched filtering.

Since ℓ_{bin} is a linear function of the data and the g_m are independent Poisson random variables, it is straightforward to compute d^2; the result is

$$d^2 = \frac{\left\{\sum_{m=1}^{M}[(\mathbf{Hf}_2)_m - (\mathbf{Hf}_1)_m]\log\left[\frac{(\mathbf{Hf}_2)_m}{(\mathbf{Hf}_1)_m}\right]\right\}^2}{\frac{1}{2}\sum_{m=1}^{M}[(\mathbf{Hf}_2)_m - (\mathbf{Hf}_1)_m]\log^2\left[\frac{(\mathbf{Hf}_2)_m}{(\mathbf{Hf}_1)_m}\right]}. \tag{43}$$

This expression was derived by Cunningham $et\ al.$[27] and used as a figure of merit for image quality by Wagner $et\ al.$[26]

C. List Mode

In this section we present expressions analogous to Eqs. (42) and (43) for the log likelihood and detectability but for list-mode data. Only the case of preset time will be considered, but the case of preset counts is similar.

We define a density $\bar{h}_i(\mathbf{r}_j)$ as in Eq. (9) but with a subscript to distinguish which of the two objects is present. Thus, from Eqs. (9), (31), and (35), we write

$$\bar{h}_i(\mathbf{r}_j) = \bar{J}_i \, \mathrm{pr}(\mathbf{r}_j|\mathbf{f}_i) = \tau \sum_{k=1}^{K} f_{ik} S_k \, \mathrm{pr}(\mathbf{r}_j|j \text{ from } k), \quad i=1,2. \tag{44}$$

Using Eq. (37), we find

$$\ell_{\mathrm{list}} = \bar{J}_1 - \bar{J}_2 + \sum_{j=1}^{J} \log\left[\frac{\bar{h}_2(\mathbf{r}_j)}{\bar{h}_1(\mathbf{r}_j)} \right]. \tag{45}$$

The constant terms $\bar{J}_1 - \bar{J}_2$ are irrelevant for purposes of hypothesis testing since the objects to be discriminated are specified. The term $-\bar{J}$ could not be dropped in Eq. (37) since it was a function of the unknown $\mathbf{f}$ in a reconstruction problem, but we can drop the corresponding terms in Eq. (45).

Comparison of Eq. (45) with Eq. (42) shows that there is a similar logarithmic structure, but the dependence on the data is quite different. In ℓ_{bin}, the data g_m appear linearly, but in ℓ_{list} the attribute vectors $\mathbf{r}_j$ are buried in the arguments of the densities. In contrast to ℓ_{bin}, ℓ_{list} cannot be realized by a linear filter.

Next we turn to the detectability index, which requires computation of the mean and variance of ℓ_{list}. It is shown in Appendix A that

$$d^2 = \frac{\left\{ \int_{\mathrm{att}} \mathrm{d}^N r [\bar{h}_2(\mathbf{r}) - \bar{h}_1(\mathbf{r})] \log\left[\frac{\bar{h}_2(\mathbf{r})}{\bar{h}_1(\mathbf{r})} \right] \right\}^2}{\frac{1}{2} \int_{\mathrm{att}} \mathrm{d}^N r [\bar{h}_1(\mathbf{r}) + \bar{h}_2(\mathbf{r})] \log^2\left[\frac{\bar{h}_2(\mathbf{r})}{\bar{h}_1(\mathbf{r})} \right]}. \tag{46}$$

The structure of this expression is very similar to that of the Cunningham formula, Eq. (43). Evaluation of Eq. (43) requires computation of the mean data vector for each object, while evaluation of Eq. (46) requires computation of the mean data density in attribute space for each object.

5. SUMMARY AND CONCLUSIONS

The main contribution of this paper has been the development of a likelihood formalism for photon-counting imaging systems in which multiple attributes are measured for each photon. Since binning of the count data may not be practical when the number of attributes per photon is larger than about four, attention was concentrated on list-mode data storage. Separate likelihood formulas were derived for acquisitions with preset time and ones with present number of events. In Section 3 we derived expressions for list-mode likelihood that can serve as the starting point for a specific reconstruction algorithm. The algorithm itself is derived and discussed in a separate paper.[11] In Section 4 we derived expressions for the likelihood ratio and the detectability index for list-mode data when the task was detection of a nonrandom object or discrimination between two such objects. The detectability expression is the list-mode

counterpart of a formula due to Cunningham *et al.*[27] that has been used for image-quality assessment.

A limitation of the theory developed here is that it considers only nonrandom objects. It has been noted in the literature[28–30] that classification tasks based on nonrandom objects can lead to counterintuitive results on image quality. It is argued in these papers that it is better to allow some degree of randomness in the objects to be discriminated.

Unfortunately, it has not been possible to develop a useful theory for the likelihood ratio or the detectability when object randomness is included. With binned data, this difficulty has led to consideration of suboptimal linear discriminants in place of the likelihood ratio.[21,25,29,31,32] These discriminants have proved to be an effective tool for optimizing imaging systems and predicting performance of human observers.

In the case of list-mode data, the natural interpretation of a linear discriminant is that it is linear not in the attributes themselves but in the random process $g(\mathbf{r})$. Future work will investigate the properties of such linear discriminants.

Acknowledgments

References

1. Gagnon D, Todd-Pokropek A, Arsenault A, Dupras G. Introduction to holospectral imaging in nuclear medicine for scatter subtraction. IEEE Trans Nucl Sci 1989;8:245–250.

2. Jaszczak RJ, Greer KL, Floyd CE Jr, Harris CC, Coleman RE. Improved SPECT quantification using compensation for scattered photons. J Nucl Med 1984;25:893–900. [PubMed: 6611390]

3. King MA, Hademenos GJ, Glick SJ. A dual-photopeak window method for scatter correction. J Nucl Med 1992;33:605–612. [PubMed: 1552349]

4. Cook WR, Finger M, Prince TA. A thick Anger camera for gamma-ray astronomy. IEEE Trans Nucl Sci 1985;NS-32:129–133.

5. Rogers JG, Saylor DP, Harrop R, Yao XG, Leitao CVM, Pate BD. Design of an efficient position sensitive gamma ray detector for nuclear medicine. Phys Med Biol 1986;31:1061–1090. [PubMed: 3786398]

6. Gagnon D. Maximum likelihood positioning in the scintillation camera using depth of interaction. IEEE Trans Med Imaging MI-12 1993:101–107.

7. White, TA.; Barrett, HH.; Rowe, RK. Direct three-dimensional SPECT reconstruction from photomultiplier signals. presented at the International Meeting on Fully 3D Image Reconstruction in Nuclear Medicine and Radiology; Corsendonk, Belgium. June 1991;

8. White, TA. PhD dissertation. University of Arizona; Tucson, Ariz: 1994. SPECT reconstruction directly from photo-multiplier tube signals.

9. Singh M. An electronically collimated gamma camera for single photon emission computed tomography. Part 1: theoretical considerations and design criteria. Med Phys 1983;10:421–427. [PubMed: 6604216]

10. Kijewski MF, Moore SC, Mueller SP. Cramer–Rao bounds for estimation tasks using multiple energy window data. J Nucl Med 1994;35:4P.

11. Parra L, Barrett HH. List-mode likelihood—EM algorithm and noise estimation demonstrated on 2D-PET. IEEE Trans Med Imaging. (to be published).

12. Snyder DL. Parameter estimation for dynamic studies in emission-tomography systems having list-mode data. IEEE Trans Nucl Sci NS-3 1984:925–931.

13. Snyder DL, Politte DG. Image reconstruction from list-mode data in an emission tomography system having time-of-flight measurements. IEEE Trans Nucl Sci NS-20 1983:1843–1849.

14. Barrett, HH.; Swindell, W. Radiological Imaging: The Theory of Image Formation, Detection and Processing. paperback. Vol. Chap. 3. Academic; San Diego, Calif: 1996.

15. Rabbani M, Shaw R, van Metter R. Detective quantum efficiency of imaging systems with amplifying and scattering mechanisms. J Opt Soc Am A 1987;4:895–901. [PubMed: 3598742]

16. Barrett, HH.; Swindell, W. Radiological Imaging: The Theory of Image Formation, Detection and Processing. paperback. Vol. Chap. 5. Academic; San Diego, Calif: 1996.

17. Goodman, JW. Statistical Optics. Wiley; New York: 1985.

18. Frieden, BR. Probability, Statistical Optics, and Data Testing: A Problem Solving Approach. Springer-Verlag; New York: 1983.

19. Van Trees, HL. Detection, Estimation, and Modulation Theory. Vol. I. Wiley; New York: 1968.

20. Melsa, JL.; Cohn, DL. Decision and Estimation Theory. McGraw-Hill; New York: 1978.

21. Barrett HH, Denny JL, Wagner RF, Myers KJ. Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance. J Opt Soc Am A 1995;12:834–852.

22. Rockmore AJ, Macovski A. A maximum likelihood approach to emission image reconstruction from projections. IEEE Trans Nucl Sci NS-23 1976:1428–1432.

23. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Statistical Soc Ser B 1977;39:1–38.

24. Shepp L, Vardi Y. Maximum likelihood reconstruction for emission tomography. IEEE Trans Med Imaging, MI-1 1982:113–121.

25. Barrett HH. Objective assessment of image quality: effects of quantum noise and object variability. J Opt Soc Am A 1990;7:1266–1278. [PubMed: 2370589]

26. Wagner, RF.; Brown, DG.; Metz, CE. On the multiplex advantage of coded source/aperture photon imaging. In: Brody, WR., editor. Digital Radiography; Proc SPIE; 1981. p. 72-76.

27. Cunningham DR, Laramore RD, Barrett E. Detection in image dependent noise. IEEE Trans Inf Theory IT-10 1976:603–610.

28. Myers KJ, Rolland JP, Barrett HH, Wagner RF. Aperture optimization for emission imaging: effect of a spatially varying background. J Opt Soc Am A 1990;7:1279–1293. [PubMed: 2370590]

29. Barrett HH, Gooley TA, Girodias KA, Rolland JP, White TA, Yao J. Linear discriminants and image quality. Image Vis Comput 1992;10:451–460.

30. Rolland JP, Barrett HH. Effect of random background inhomogeneity on observer detection performance. J Opt Soc Am A 1992;9:649–658. [PubMed: 1588452]

31. Smith WE, Barrett HH. Hotelling trace criterion as a figure of merit for the optimization of imaging systems. J Opt Soc Am A 1986;3:717–725.

32. Fiete RD, Barrett HH, Smith WE, Myers KJ. Hotelling trace criterion and its correlation with human observer performance. J Opt Soc Am A 1987;4:945–953. [PubMed: 3598746]

APPENDIX A: COMPUTATION OF d2 FOR LIST-MODE DATA

With preset time, ℓ_{list} is a function of the $J+1$ random variables $\{\mathbf{r}_j\}$ and J itself. Its expectation when hypothesis H_i is true (or object $\mathbf{f}_i$ is present) is given by[14]

$$E(\ell_{\text{list}}|H_i) = \sum_{J=0}^{\infty} \Pr(J|H_i) \int_{\text{att}} \mathrm{d}^N r_1 \, \mathrm{pr}(\mathbf{r}_1|H_i) \times \int_{\text{att}} \mathrm{d}^N r_2 \, \mathrm{pr}(\mathbf{r}_2|H_i) \cdots \times \int_{\text{att}} \mathrm{d}^N r_N \, \mathrm{pr}(\mathbf{r}_N|H_i) \ell_{\text{list}}.$$

(A1)

We now use Eq. (45) for ℓ_{list} and drop the constant terms $J_1 - J_2$, yielding

$$E(\ell_{\text{list}}|H_i) = \sum_{J=0}^{\infty} \Pr(J|H_i) \int_{\text{att}} d^N r_1 \, \text{pr}(\mathbf{r}_1|H_i) \times \int_{\text{att}} d^N r_2 \, \text{pr}(\mathbf{r}_2|H_i) \cdots \times \int_{\text{att}} d^N r_J \, \text{pr}(\mathbf{r}_J|H_i) \sum_{j=1}^{J} \log \left[\frac{\overline{h}_2(\mathbf{r}_j)}{\overline{h}_1(\mathbf{r}_j)} \right].$$

(A2)

Consider one particular value of j, say, $j = 17$. Then all of the integrals except the one over $\mathbf{r}_{17}$ yield unity, and that one gives

$$\int_{\text{att}} d^{17} r_{17} \, \text{pr}(\mathbf{r}_{17}|H_i) \log \left[\frac{\overline{h}_2(\mathbf{r}_{17})}{\overline{h}_1(\mathbf{r}_{17})} \right] = \frac{1}{\overline{J}_i} \int_{\text{att}} d^N r_{17} \, \overline{h}_i(\mathbf{r}_{17}) \log \left[\frac{\overline{h}_2(\mathbf{r}_{17})}{\overline{h}_1(\mathbf{r}_{17})} \right].$$

(A3)

Since r_{17} is just a dummy variable of integration, it does not matter which value of j was chosen. There are J identical terms in the sum over j, and we have

$$\begin{aligned} E(\ell_{\text{list}}|H_i) &= \sum_{J=0}^{\infty} \Pr(J|H_i) \frac{1}{\overline{J}_i} \int_{\text{att}} d^N r \, \overline{h}_i(\mathbf{r}) \log \left[\frac{\overline{h}_2(\mathbf{r})}{\overline{h}_1(\mathbf{r})} \right] \\ &= \int_{\text{att}} d^N r \, \overline{h}_i(\mathbf{r}) \log \left[\frac{\overline{h}_2(\mathbf{r})}{\overline{h}_1(\mathbf{r})} \right]. \end{aligned}$$

(A4)

As a step toward computing the variance, we consider the average of the square of ℓ_{list}, given by

$$E\{(\ell_{\text{list}})^2|H_i\} = \sum_{J=0}^{\infty} \Pr(J|H_i) \int_{\text{att}} d^N r_1 \, \text{pr}(\mathbf{r}_1|H_i) \times \int_{\text{att}} d^N r_2 \, \text{pr}(\mathbf{r}_2|H_i) \cdots \times \int_{\text{att}} d^N r_J \, \text{pr}(\mathbf{r}_J|H_i) \cdot \sum_{j=1}^{J} \log \left[\frac{\overline{h}_2(\mathbf{r}_j)}{\overline{h}_1(\mathbf{r}_j)} \right] \sum_{j'=1}^{J} \log \left[\frac{\overline{h}_2(\mathbf{r}_{j'})}{\overline{h}_1(\mathbf{r}_{j'})} \right].$$

(A5)

In the double sum over j and j', there are J terms for which $j = j'$, and there are $J^2 - J$ terms for which $j \neq j'$. Performing each of these averages separately and making use of the statistical independence of $\mathbf{r}_j$ and $\mathbf{r}_{j'}$ for $j \neq j'$, we find

$$E\{(\ell_{\text{list}})^2|H_i\} = \sum_{J=0}^{\infty} \Pr(J|H_i) \left\{ \frac{J}{\overline{J}_i} \int_{\text{att}} d^N r \, \overline{h}_i(\mathbf{r}) \log^2 \left[\frac{\overline{h}_2(\mathbf{r})}{\overline{h}_1(\mathbf{r})} \right] \right\} + \sum_{J=0}^{\infty} \Pr(J|H_i) \times \left(\frac{J^2 - J}{\overline{J}_i^2} \left\{ \int_{\text{att}} d^N r \, \overline{h}_i(\mathbf{r}) \log \left[\frac{\overline{h}_2(\mathbf{r})}{\overline{h}_1(\mathbf{r})} \right] \right\}^2 \right).$$

(A6)

For Poisson J,

$$\sum_{J=0}^{\infty} \Pr(J|H_i)(J^2 - J) = \overline{J}_i^{\,2},$$

(A7)

so

$$E\{(\ell_{\text{list}})^2|H_i\}=\int_{\text{att}}\text{d}^N r\,\overline{h}_i(\mathbf{r})\log^2\left[\frac{\overline{h}_2(\mathbf{r})}{\overline{h}_1(\mathbf{r})}\right]+[E(\ell_{list}|H_i)]^2.$$

(A8)

Thus the variance of the log of the likelihood ratio is

$$\text{var}(\ell_{\text{list}}|H_i)=\int_{\text{att}}\text{d}^N r\,\overline{h}_i(\mathbf{r})\log^2\left[\frac{\overline{h}_2(\mathbf{r})}{\overline{h}_1(\mathbf{r})}\right].$$

(A9)

Equations (A4) and (A9) are inserted into Eq. (40) to obtain the expression for d^2 in Eq. (46).