The Annals of Statistics
1994, Vol. 22, No. 4, 1871–1883

ASYMPTOTIC CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATE IN POSITRON EMISSION TOMOGRAPHY AND APPLICATIONS[1]

BY I-SHOU CHANG AND CHAO A. HSIUNG

National Central University and Academia Sinica

This paper indicates that a minor modification of the maximum likelihood estimate of Vardi, Shepp and Kaufman can be regarded as a step in the standard nonparametric MLE by the method of sieves and establishes the asymptotic consistency for it. This method of sieves suggests that the number of pixels needs to be in line with the number of detectors in order to avoid poor image reconstructions. A simulation study is also presented to support this suggestion.

1. Introduction. Positron emission tomography (PET) deals with the estimation of the amount and location of a radioactively labeled metabolite in the organ under study on the basis of particle decays indirectly observed outside the body. When a positron is emitted from the metabolite introduced into the organ, it annihilates with a nearby electron. The annihilation creates two gamma ray photons that fly off the point of annihilation, at the speed of light, in (nearly) opposite directions with a completely random orientation. The two photons are then detected in coincidence by two of the many detectors mounted on a ring, positioned around the body. Based on data collected this way, Vardi, Shepp and Kaufman (1985) (henceforth VSK) proposed a mathematical model and the statistical methodology needed for it.

The mathematical model of VSK assumes that the emission location follows a spatial Poisson process. Assuming further that the spatial Poisson process has a discrete intensity function, called emission density, of a certain form, VSK studied the maximum likelihood estimation (MLE) of the emission density, including the relevant EM algorithm.

In this paper we will indicate that a minor modification of the MLE approach of VSK can be regarded as a step in the standard nonparametric MLE by the method of sieves and establish the asymptotic consistency for it. In fact, the asymptotic consistency is established when the number of detectors, pixels and data counts go to ∞ satisfying certain relations among them. These relations have practical implications. They suggest we design the experiment according to these relations so as to avoid possibly poor image reconstruction.

This paper is organized as follows. Section 2 fixes the notation and describes the statistical problem for positron emission tomography.

Received June 1992; revised October 1993.
[1]Research partially supported by NSC Grant NSC81-0208-M008-17 of ROC.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62G20, 62M99.
Key words and phrases. Asymptotics, density estimation, method of sieves, nonparametric maximum likelihood estimation, positron emission tomography.

Section 3 introduces the sieves so as to establish the asymptotic consistency of the MLE of VSK. In particular, we give explicit relations among the number of pixels, detectors and the emission intensity so that the asymptotic consistency is valid. In Section 4 we will take these relations as a guide for better image reconstruction. Our simulation study indicates that with a given number of detectors and data counts, increasing the number of pixels produces better images initially and poorer images afterwards.

There is a substantial literature on PET image reconstruction and related statistical theory. We refer the reader to Jones and Silverman (1989), Johnstone and Silverman (1990) and Silverman, Jones, Wilson and Nychka (1990) and the references therein for some of the recent developments.

2. Preliminaries. This section fixes the notation, presents the basic facts of the Radon transform needed in this paper and describes the relevant statistical problem for PET.

2.1. *Notation and the Radon transform.* This subsection adapts some of the notation and framework set in Johnstone and Silverman (1990) for the purpose of a concise presentation. The reader is encouraged to consult Deans (1983), Natterer (1986), Jones and Silverman (1989) and Johnstone and Silverman (1990) for more details.

Let both B and D denote the closed unit disk in the plane (B is usually referred to as the brain space and D the detector space). Define a mapping from $B' = B \times [0, \pi)$ to R^3 by sending $(x^{(1)}, x^{(2)}, \psi)$ to (s, φ, t) as follows:

$$
\begin{aligned}
s &= \left|x^{(1)} \cos\psi + x^{(2)} \sin\psi\right|,\\
\varphi &= \begin{cases} \psi, & \text{if } x^{(1)} \cos\psi + x^{(2)} \sin\psi > 0,\\ \psi + \pi, & \text{otherwise,} \end{cases} \qquad (2.1)\\
t &= -x^{(1)} \sin\psi + x^{(2)} \cos\psi.
\end{aligned}
$$

Figure 1 depicts this mapping.

Let D' be the range of this mapping in R^3. Consider a further mapping from D' to D by sending (s, φ, t) to $(s\cos\varphi, s\sin\varphi)$. Denote by u the composition of these two mappings, which is from B' to D.

Let $L^2(B, m_B)$ be the space of square-integrable functions on B with respect to the measure $m_B = \pi^{-1}\times$ Lebesgue measure. Let $L^2(D, m_D)$ be that with respect to the measure $m_D = 2\pi^{-2}(1 - s^2)^{1/2}\, ds\, d\varphi$.

Let P be the linear operator from $L^2(B, m_B)$ to $L^2(D, m_D)$ defined by

$$Pf(s, \varphi) = E\big(f(X) \mid X^{(1)} \cos\varphi + X^{(2)} \sin\varphi = s\big),$$

with $X = (X^{(1)}, X^{(2)})$ distributed uniformly on B. P is called the Radon transform. In fact, if f is a density function relative to m_B, then $Pf \equiv g$ is a density function relative to m_D and

$$g(s, \varphi) = \frac{1}{2}\left(1 - s^2\right)^{-1/2} \int_{-\sqrt{1-s^2}}^{\sqrt{1-s^2}} f(s\cos\varphi - t\sin\varphi, s\sin\varphi + t\cos\varphi)\, dt. \qquad (2.2)$$

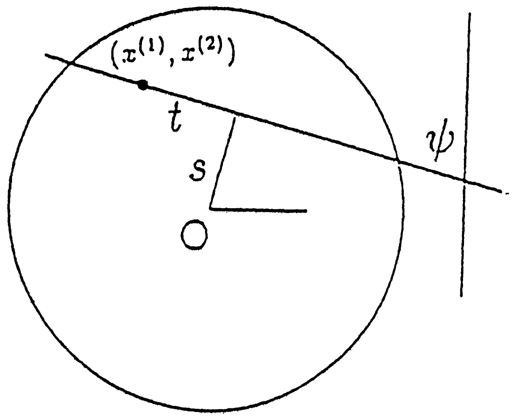

FIG. 1.

2.2. *The statistical problem.* In a typical positron emission tomography study, a detector ring surrounds the patient's head. To simplify the discussion, we will assume that the whole problem takes place in a plane and the detector ring is positioned on the boundary of the unit disk B.

Let $\widetilde{N}(G)$ denote the number of positrons emitted in a region $G \subset B$ during the study. Then $\widetilde{N}(G)$ is a Poisson random variable with mean

$$E\widetilde{N}(G) = \Lambda\left(\int_G f_0\left(x^{(1)}, x^{(2)}\right) dm_B\right), \tag{2.3}$$

where f_0 is a density function relative to m_B and Λ is a constant. We note that f_0 and Λ represent respectively the relative concentration and the total amount of the radioactively labeled metabolite in B. The density f_0 will be called the emission density and Λ the emission rate in the study.

It is clear that every pair of detectors defines a specific subset C' of $B' = B \times [0, \pi)$. A point $(x^{(1)}, x^{(2)}, \psi) \in B'$ is in C' if the line passing through $(x^{(1)}, x^{(2)})$ in direction ψ intersects this pair of detectors. This set C' specifies in turn a subset C of D, by the mapping u from B' to D. We assume that the detector space D is the disjoint union of sets $C_1, \dots, C_K$ with each C_k specified by a pair of detectors.

Let C_k be a subset of D defined by a certain pair of detectors. We will denote by $N(C_k)$ the number of photon pairs detected by this pair of detectors during the study. It follows from VSK and (2.3) that $N(C_1), \dots, N(C_K)$ is a finite sequence of independent Poisson random variables with mean

$$EN(C_k) = \Lambda \int_{C_k} (Pf_0)(s, \varphi)\, dm_D. \tag{2.4}$$

The statistical problem is to estimate the emission density f_0 based on the data

$$\{N(C_1), \dots, N(C_K)\}. \tag{2.5}$$

We will denote by $N(\cdot)$ the spatial Poisson point process on D satisfying (2.4).

3. Asymptotic consistency of the maximum likelihood estimate. In this section we will establish the asymptotic consistency of the nonparametric maximum likelihood estimation (MLE) by the method of sieves for the statistical problem stated in Section 2.2; namely, to estimate the emission density based on the data (2.5).

In order to simplify the mathematics and avoid the situation of estimating infinitely many parameters in this statistical problem, VSK assumed that the emission density is constant on each of the pixels artificially created in the brain space B and discussed the existence and uniqueness of the MLE under this assumption. We will see that the MLE approach of VSK can be regarded as a nonparametric MLE by the method of sieves and as such its asymptotic consistency could be expected. We refer the reader to Grenander (1981), Geman and Hwang (1982) and McKeague (1986) and the references therein for general discussions of the asymptotic consistency of the method of sieves.

The main result of this section says roughly that, with a relatively large number of detectors, moderate number of pixels and large data counts, the MLE of VSK will be close to the true emission density in the sense of (3.8).

3.1. *Nonparametric MLE by the method of sieves.* In this subsection we will define the sieve and the entropy concept and present the likelihood to be used in establishing the asymptotic consistency of a nonparametric MLE. The sieve we use corresponds to the pixelization of VSK. Although there are other possibilities [cf. Silverman, Jones, Wilson and Nychka (1990)], we will focus only on this sieve for the sake of a concise presentation.

Let A be the set of all density functions, bounded by $\widetilde{C}$, relative to the measure m_B on the brain space B. For every positive m, let $A_m = \{f \in A \mid f$ is constant on every pixel specified by the lattice points $(k/m, l/m)$, and $f \geq \pi/m\}$.

For $f, h \in A$, we define

$$H(f,h) = \int_D (Pf)(\log Ph)\, dm_D, \tag{3.1}$$

where P is the Radon transform in (2.2). In fact, H is the ordinary formal entropy for the Radon transform of densities on B.

Let $f_m \in A_m$ be chosen so that

$$\lim_{m\to\infty} H(f_0, f_m) = H(f_0, f_0), \tag{3.2}$$

where f_0 denotes the true emission density. Let

$$W_m = \{f \in A_m \mid H(f_0, f) \leq H(f_0, f_m) - 2\delta\}, \tag{3.3}$$

where $\delta > 0$ is a constant.

In the rest of this subsection we will give the relevant likelihood and some related concepts. We will consider a sequence of PET studies indexed by $n = 1, 2, \ldots$.

Let $\pi(n) = \{C_1^n, \dots, C_{K(n)}^n\}$ be a partition of the detector space D defined by a ring of identical equally spaced detectors. Let $N_n(\cdot)$ be a spatial Poisson point process on D satisfying

$$EN_n\left(C_k^n\right) = \Lambda_n \int_{C_k^n} (Pf_0)\, dm_D. \tag{3.4}$$

To simplify the notation, we set

$$\mathcal{I}(C_k^n, f) \equiv \left(\int_{C_k^n} (Pf)\, dm_D \right).$$

Hence the likelihood for $N_{\pi(n)} = (N_n(C_1^n), \dots, N_n(C_{K(n)}^n))$ specified by f is

$$L_n(f) = \prod_{k=1}^{K(n)} L_{n,k}(f), \tag{3.5}$$

where $L_{n,k}(f)$ satisfies

$$\begin{aligned} \log L_{n,k}(f) &= \log \frac{\mathcal{I}(C_k^n, f)}{\mathcal{I}(C_k^n, 1)} N_n\left(C_k^n\right) \\ &\quad + \left(\mathcal{I}(C_k^n, 1) - \mathcal{I}(C_k^n, f)\right)\Lambda_n \end{aligned} \tag{3.6}$$

and

$$\log L_n(f) = \sum_{k=1}^{K(n)} \log \frac{\mathcal{I}(C_k^n, f)}{\mathcal{I}(C_k^n, 1)} N_n\left(C_k^n\right).$$

Here we used the fact that $\Sigma_{k=1}^{K(n)}(\mathcal{I}(C_k^n, 1) - \mathcal{I}(C_k^n, f))\Lambda_n = (\int_D (P1)\, dm_D - \int_D (Pf)\, dm_D)\Lambda_n = 0$.

Let $m(n)$ be a sequence of increasing integers with $\lim_{n \to \infty} m(n) = \infty$. Let

$$\Theta_{m(n)}^n = \left\{ f \in A_{m(n)} \mid L_n(f) = L_n\left(A_{m(n)}\right) \right\} \tag{3.7}$$

be the set of all maximum likelihood estimators in $A_{m(n)}$, given the data $N_{\pi(n)}$. In (3.7) we adapted the convention that for a real-valued function h and a subset W of its domain, $h(W)$ denotes the sup of h over W. This convention is used throughout this section.

According to the arguments in VSK, $\Theta_{m(n)}^n$ is not empty. We note that if we did not require $f \geq \pi/m$ in the definition of A_m, $\Theta_{m(n)}^n$ would be precisely the MLE of VSK.

3.2. *Asymptotic consistency.* Suppose $K(n)$, the number of detector tubes, goes to ∞ fast enough, $m(n)$ goes to ∞ slowly enough and Λ_n goes to ∞ as described in (3.11), (3.12), (3.14) and (3.15). Then we have the following result.

THEOREM 3.1.

$$\lim_{n\to\infty} \sup_{f\in\Theta^n_{m(n)}} \int_D |Pf - Pf_0|\, dm_D = 0 \quad a.e. \tag{3.8}$$

For the proof of Theorem 3.1, we need the following lemma.

LEMMA 3.1. *For every* $\delta > 0$,

$$\mathcal{P}\big(\Theta^n_{m(n)} \cap W_{m(n)} \neq \emptyset \ i.o.\big) = 0. \tag{3.9}$$

PROOF. Observe that

$$\begin{aligned}
&\mathcal{P}\big(\Theta^n_{m(n)} \cap W_{m(n)} \neq \emptyset\big)\\
&\leq \mathcal{P}\Big(\sup_{f\in W_{m(n)}} L_n(f) \geq L_n\big(f_{m(n)}\big)\Big)\\
&\leq \mathcal{P}\Bigg(\sup_{f\in W_{m(n)}} \bigg(\sum_k \log \frac{\mathcal{I}(C^n_k, f)}{\mathcal{I}(C^n_k, 1)} N_n(C^n_k)\bigg)\\
&\qquad\qquad \geq \sum_k \log \frac{\mathcal{I}(C^n_k, f_{m(n)})}{\mathcal{I}(C^n_k, 1)} N_n(C^n_k)\Bigg)\\
&\leq \mathcal{P}\bigg(\sup_{W_{m(n)}} \eta_n \sum_k \log \frac{\mathcal{I}(C^n_k, f)}{\mathcal{I}(C^n_k, f_{m(n)})} N_n(C^n_k) \geq 0\bigg)\\
&\leq \mathcal{P}\bigg(\eta_n \sum_k \log \frac{\mathcal{I}(C^n_k, W_{m(n)})}{\mathcal{I}(C^n_k, f_{m(n)})} N_n(C^n_k) \geq 0\bigg)\\
&\leq E\exp\bigg\{\eta_n \sum_k \log \frac{\mathcal{I}(C^n_k, W_{m(n)})}{\mathcal{I}(C^n_k, f_{m(n)})} N_n(C^n_k)\bigg\}\\
&= \prod_k e^{\mathcal{I}(C^n_k, f_0)\Lambda_n Q_{nk}},
\end{aligned} \tag{3.10}$$

where $Q_{nk} = \exp\{\eta_n \log[\mathcal{I}(C^n_k, W_{m(n)})/\mathcal{I}(C^n_k, f_{m(n)})]\} - 1$ and $\eta_n > 0$ is any constant.

Assume now that $K(n)$ and $m(n)$ go to ∞ at rates such that

$$\sup_{f\in W_{m(n)}} \left| \sum_{k=1}^{K(n)} \Big(\log |C^n_k|^{-1} \mathcal{I}(C^n_k, f)\Big) \mathcal{I}(C^n_k, f_0) - H(f_0, f)\right| = o(1) \tag{3.11}$$

and

$$\left| \sum_{k=1}^{K(n)} \Big(\log |C^n_k|^{-1} \mathcal{I}(C^n_k, f_{m(n)})\Big) \mathcal{I}(C^n_k, f_0) - H(f_0, f_{m(n)})\right| = o(1), \tag{3.12}$$

where $|C_k^n| = m_D(C_k^n)$.

Let $f \in W_{m(n)}$. Then, using (3.11) and (3.12), for n large,

$$\begin{aligned}\sum_k \left(\log \frac{\mathcal{I}(C_k^n, f)}{\mathcal{I}(C_k^n, f_{m(n)})}\right) \mathcal{I}(C_k^n, f_0) \\ \le \left(H(f_0, f) - H(f_0, f_{m(n)}) + \frac{\delta}{2}\right) \\ \le -\frac{3\delta}{2}.\end{aligned} \tag{3.13}$$

Assume further that

$$\eta_n \sup_k \left(\log \frac{\mathcal{I}(C_k^n, W_{m(n)})}{\mathcal{I}(C_k^n, f_{m(n)})}\right)^2 = o(1), \tag{3.14}$$

$$\sum_n e^{-\delta \Lambda_n \eta_n} < \infty. \tag{3.15}$$

Then, using (3.10), (3.13), (3.14) and (3.15), we have, for suitably large n_0,

$$\begin{aligned}&\sum_{n=n_0}^{\infty} \mathcal{P}(\Theta_{m(n)}^n \cap W_{m(n)} \neq \emptyset) \\ &\le \sum_n \prod_k \exp\left\{\mathcal{I}(C_k^n, f_0)\Lambda_n \left(\eta_n \log \frac{\mathcal{I}(C_k^n, W_{m(n)})}{\mathcal{I}(C_k^n, f_{m(n)})}\right.\right. \\ &\qquad\qquad \left.\left. + \eta_n^2 \left(\log \frac{\mathcal{I}(C_k^n, W_{m(n)})}{\mathcal{I}(C_k^n, f_{m(n)})}\right)^2\right)\right\} \\ &\le \sum_n \exp\left\{\Lambda_n \eta_n \sum_k \mathcal{I}(C_k^n, f_0) \left(\log \frac{\mathcal{I}(C_k^n, W_{m(n)})}{\mathcal{I}(C_k^n, f_{m(n)})}\right.\right. \\ &\qquad\qquad \left.\left. + \eta_n \left(\log \frac{\mathcal{I}(C_k^n, W_{m(n)})}{\mathcal{I}(C_k^n, f_{m(n)})}\right)^2\right)\right\} \\ &\le \sum_n \exp\left\{\Lambda_n \eta_n \left(-\frac{3\delta}{2} + \frac{\delta}{2}\right)\right\} \\ &= \sum_n \exp\{-\Lambda_n \eta_n \delta\} \\ &< \infty.\end{aligned} \tag{3.16}$$

Thus Lemma 3.1 follows from (3.16) and the Borel–Cantelli lemma. □

PROOF OF THEOREM 3.1. Using Lemma 3.1 and a standard argument [cf. Geman and Hwang (1982), pages 405 and 406], we can show

$$\lim_{n\to\infty} \sup_{f\in\Theta^n_{m(n)}} |H(f_0, f) - H(f_0,f_0)| = 0 \quad \text{a.e.} \tag{3.17}$$

Therefore, Theorem 3.1 follows from (3.17) and the Kullback–Csiszar–Kemperman inequality [cf. Devroye (1987), page 10], which says, in particular,

$$\left(\int_D |Pf - Pf_0|\, dm_D\right)^2 \le 2\big(H(f_0,f_0) - H(f_0,f)\big).$$

This completes the proof. □

3.3. *An explicit sufficient condition.* The asymptotic consistency of the MLE was established under the condition (3.11), (3.12), (3.14) and (3.15). This is a general condition. There are many simpler sufficient conditions for which this general condition holds. We will give one such useful sufficient condition in this subsection and use it as a guide in image reconstruction in Section 4.

We note that item (i) in Theorem 3.2 is a relation between the number of pixels and the number of detectors and item (ii) is about the emission rates. If we choose $m(n) = n$, then (ii) says that the emission rate needs to be as fast as $n(\log n)^3$. We would like to remark that condition (3.14) and (3.15) depends on the sieve we use; in particular, the condition that $\pi/m \le f \le \widetilde{C}$ for $f \in A_m$ is crucial.

THEOREM 3.2. (i) *If*

$$\lim_{n\to\infty} \frac{m^2(n)\log^2 m(n)}{d(n)} = 0,$$

then (3.11) *and* (3.12) *are satisfied, where $d(n)$ is the number of detectors.* (ii) *If* $\Lambda_n \ge n(\log m(n))^3$, *then* (3.14) *and* (3.15) *are satisfied with the choice of* $\eta_n = (\log m(n))^{-3}$ *in* (3.14).

PROOF. The proof for (ii) is straightforward, recognizing that $\pi m(n)^{-1} \le f \le \widetilde{C}$ for every $f \in A_{m(n)}$.

We will now start to prove (i). Let

$$\mathcal{R}_1 = \{(s,\theta) \in D \mid s \ge s_{m(n)}\},$$

$$\mathcal{R}_2 = \left\{(s,\theta) \in D \mid \max\left\{|\theta|, \left|\theta - \frac{\pi}{2}\right|, |\theta - \pi|, \left|\theta - \frac{3}{2}\pi\right|, |\theta - 2\pi|\right\} \le \theta_{m(n)}\right\},$$

where both $s_{m(n)}$ and $\theta_{m(n)}$ are constants to be specified later. Let I be the collection of the C^n_k intersecting $\mathcal{R}_1$, II be that intersecting $\mathcal{R}_2$ but not in I and III be the rest. Observe that

$$\sum_{k=1}^{K(n)}\left(\log\frac{1}{|C_k^n|}\mathfrak{I}(C_k^n,f)\right)\mathfrak{I}(C_k^n,f_0)-H(f_0,f)$$

$$=\sum_k\int_{C_k^n}(Pf_0)\left[\log\frac{1}{|C_k^n|}\int_{C_k^n}(Pf)\,dm_D-\log Pf\right]dm_D \tag{3.18}$$

$$=S_{\mathrm{I}}+S_{\mathrm{II}}+S_{\mathrm{III}},$$

where $S_{\mathrm{I}}, S_{\mathrm{II}}$ and S_{III} denote, respectively, the sum of the integrals in (3.18) with C_k^n in I, II and III. Note that we shall consider only f in $A_{m(n)}$ for (3.18).

A careful examination of the geometry shows that each C_k^n is contained in a set of the form

$$\left\{(s,\theta)\in D\,\|s-s_0|\le\frac{2\pi}{d(n)},\ |\theta-\theta_0|\le\frac{2\pi}{d(n)}\right\}$$

for some (s_0,θ_0). Thus we can bound S_{I} as follows.

Let

$$\mathcal{R}_1'=\left\{(s,\theta)\in D\,|\,s\ge s_{m(n)}-\frac{2\pi}{d(n)}\right\}.$$

Then, with a little calculation,

$$|S_{\mathrm{I}}|\le 2\widetilde{C}\cdot\log m(n)\cdot m_D(\mathcal{R}_1')$$

$$\le\frac{16}{\pi}\cdot\widetilde{C}\cdot\left(\log m(n)\right)\left(\left(1-s_{m(n)}\right)^{3/2}+\left(\frac{2\pi}{d(n)}\right)^{3/2}\right). \tag{3.19}$$

Similarly, let

$$\mathcal{R}_2'=\left\{(s,\theta)\in D\,|\,\max\left\{|\theta|,\ \left|\theta-\frac{\pi}{2}\right|,|\theta-\pi|,\left|\theta-\frac{3}{2}\pi\right|,\ |\theta-2\pi|\right\}\right.$$

$$\left.\le\theta_{m(n)}+\frac{2\pi}{d(n)}\right\}.$$

Then

$$|S_{\mathrm{II}}|\le 2\widetilde{C}\cdot\left(\log m(n)\right)\cdot m_D(\mathcal{R}_2')$$

$$=10\widetilde{C}\cdot\left(\log m(n)\right)\cdot\left(\theta_{m(n)}+\frac{2\pi}{d(n)}\right). \tag{3.20}$$

We recall that $\widetilde{C}$ is a constant depending only on the sup-norm of f_0.

Next we are going to find a bound for S_{III}.

Suppose $(s_1, \theta), (s_2, \theta) \in (\mathcal{R}_1 \cup \mathcal{R}_2)^c, |s_1 - s_2| \leq s$. Then the difference in length of the two line segments (s_1, θ) and (s_2, θ) in each pixel specified by the lattice points $(k/m, l/m)$ is bounded by

$$\left(s \wedge \left(\frac{1}{m(n)}\right)\right)\left(\tan\theta_{m(n)} + \cot\theta_{m(n)}\right). \tag{3.21}$$

Suppose $(s, \theta_1), (s, \theta_2) \in (\mathcal{R}_1 \cup \mathcal{R}_2)^c, |\theta_1 - \theta_2| \leq \theta \leq 2\sin^{-1}\sqrt{1 - s^2_{m(n)}}$. Then the line segments (s, θ_1) and (s, θ_2) meet in the unit disk and their difference in length in each pixel is bounded by

$$\left((2\sin\theta) \wedge \left(\frac{1}{m(n)}\right)\right)\left(\tan\theta_{m(n)} + \cot\theta_{m(n)}\right). \tag{3.22}$$

Both (3.21) and (3.22) can be shown by examining the relative positions of the line segments in the pixels. We omit the details.

With (3.21), we get, for large n,

$$\begin{aligned} &|\log(Pf)(s_1, \theta) - \log(Pf)(s_2, \theta)| \\ &= \left|\log\frac{(Pf)(s_1, \theta)}{(Pf)(s_2, \theta)}\right| \\ &\leq \left|\frac{(Pf)(s_1, \theta) - (Pf)(s_2, \theta)}{(Pf)(s_2, \theta)}\right| \\ &\leq C \cdot m^2(n) \cdot \left(|s_1 - s_2| \wedge \frac{1}{m(n)}\right) \cdot \left(\tan\theta_{m(n)} + \cot\theta_{m(n)}\right), \end{aligned} \tag{3.23}$$

for some constant C.

Similarly, we have

$$\begin{aligned} &|\log(Pf)(s, \theta_1) - \log(Pf)(s, \theta_2)| \\ &\leq C \cdot m^2(n) \cdot \left((2\sin\theta) \wedge \frac{1}{m(n)}\right) \cdot \left(\tan\theta_{m(n)} + \cot\theta_{m(n)}\right). \end{aligned} \tag{3.24}$$

Let $s_{m(n)} = 1 - (\log m(n))^{-1}, \theta_{m(n)} = 1/(\log m(n))^2$. Then it follows from (3.19), (3.20), (3.23) and (3.24) that (3.18) goes to 0 as n goes to ∞ uniformly for $f \in A_{m(n)}$. This completes the proof. □

4. Simulation results. Our results in Section 3.3 show that the MLE image reconstruction gets close to the true emission density when the number of data counts, pixels and detectors becomes large in a certain way. In particular,

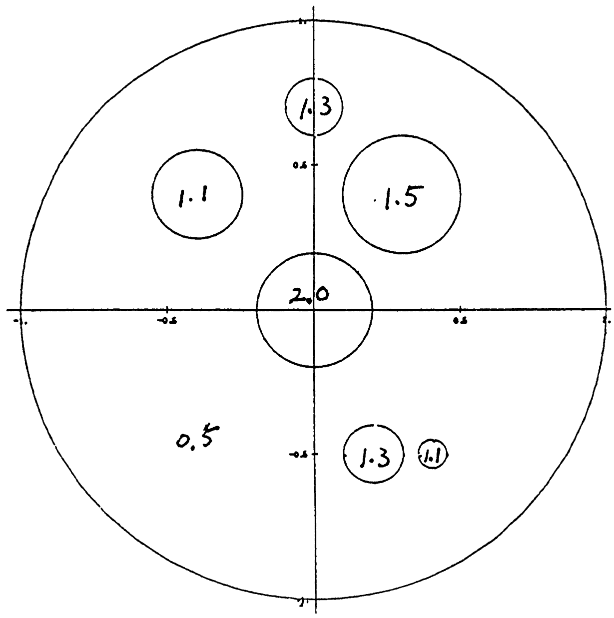

FIG. 2. *ET phantom used in the experiments in Section* 4.

it is desired that with a given number of detectors and number of data counts, we sacrifice little in resolution when reconstructing the image with bigger pixels. It seems that this theoretical conclusion is also supported by the simulation study presented in the following discussion.

We follow Shepp and Vardi (1982) very closely to do the simulations. In fact, we essentially use their computer program. We recall that, because the exact MLE is difficult to calculate, Shepp and Vardi (1982) provided an approximation by the EM algorithm, which forms part of their computer program.

In our simulation data counts are generated from a spatial Poisson process with emission density f_0 indicated in Figure 2. The emission density f_0 in each region of Figure 2 is constant and is proportional to the number written in that region. The precise definition of f_0 is as follows. Let $B(a,r)$ denote the disk centered at a with radius r. Then $f_0 = c \cdot (2.0I_{B((0,0),.2)} + 1.5I_{B((.3,.4),.2)} + 1.3I_{B((0,.7),.1)} + 1.3I_{B((.2,-.5),.1)} + 1.1I_{B((-.4,.4),.15)} + 1.1I_{B((.4,-.5),.05)} + 0.5I_{\{\text{rest area}\}})$, where c is a constant such that $\int\int_D f_0 = 1$.

Let $\widehat{f}_{n,d,b,e}(\cdot)$ denote the approximate maximum likelihood estimate of f_0 obtained by the EM algorithm, where n is the number of data counts, d is the number of detectors, b is the number of pixels on each row or column and e is the number of iterations in using the EM algorithm.

The integrated square error of the reconstructed image $\int\int_D |f_0(x,y) - \widehat{f}_{n,d,b,e}(x,y)|^2\, dx\, dy$ for this simulated data is presented in Table 1, which is computed by the subroutine DT2ODQ(DTWODQ) in IMSL Math/Library with

TABLE 1
Integrated square errors over the unit disk D for different values of (n, d, b, e) with $e = 32$

$d = 128$					
	$b = 16$	$b = 32$	$b = 64$	$b = 128$	$b = 256$
$n = 10^5$	0.03892	0.04206	0.04529	0.04804	0.04954
$n = 2 \times 10^5$	0.02880	0.02813	0.02892	0.03006	0.03072
$n = 3 \times 10^5$	0.02492	0.02264	0.02228	0.02294	0.02332
$n = 4 \times 10^5$	0.02360	0.02106	0.02043	0.02081	0.02106
$n = 5 \times 10^5$	0.02230	0.01912	0.01833	0.01855	0.01872
$n = 6 \times 10^5$	0.02189	0.01844	0.01743	0.01752	0.01763
$n = 7 \times 10^5$	0.02109	0.01724	0.01611	0.01611	0.01617
$n = 8 \times 10^5$	0.02067	0.01629	0.01510	0.01507	0.01511
$n = 9 \times 10^5$	0.02040	0.01615	0.01496	0.01487	0.01489
$n = 10^6$	0.02037	0.01590	0.01458	0.01449	0.01447

$d = 256$					
	$b = 16$	$b = 32$	$b = 64$	$b = 128$	$b = 256$
$n = 10^5$	0.04291	0.06798	0.09203	0.10655	0.11497
$n = 2 \times 10^5$	0.02991	0.03884	0.04837	0.05461	0.05815
$n = 3 \times 10^5$	0.02509	0.03039	0.03560	0.03944	0.04173
$n = 4 \times 10^5$	0.02403	0.02708	0.03023	0.03294	0.03459
$n = 5 \times 10^5$	0.02207	0.02184	0.02410	0.02608	0.02729
$n = 6 \times 10^5$	0.02167	0.02151	0.02263	0.02414	0.02512
$n = 7 \times 10^5$	0.02078	0.01954	0.02029	0.02149	0.02231
$n = 8 \times 10^5$	0.02024	0.01758	0.01812	0.01904	0.01970
$n = 9 \times 10^5$	0.01986	0.01700	0.01695	0.01762	0.01816
$n = 10^6$	0.02007	0.01657	0.01592	0.01645	0.01688

IRULE = 2.

It is clear from Table 1 that with given detectors and data counts, increasing the number of pixels produces better reconstructed images initially and poorer images eventually. For example, with $d = 128$, $n = 5 \times 10^5$, the reconstructed image giving the least integrated square error is obtained with $b = 64$.

Acknowledgments. We are grateful to Y. Vardi and L. Shepp for helpful conversations, to H. H. Wu for explaining the relevant physics to us and to Y. C. Chuang for his help in the simulation studies. We would also like to thank the referees whose suggestions simplified both the arguments and the presentation of the paper.

REFERENCES

DEANS, S. R. (1983). *The Radon Transform and Some of Its Applications.* Wiley, New York.

DEVROYE, L. (1987). *A Course in Density Estimation.* Birkhäuser, Boston.

GEMAN, S. and HWANG, C. R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 401–414.

GRENANDER, U. (1981). *Abstract Inference.* Wiley, New York.

JOHNSTONE, I. M. and SILVERMAN, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.* **18** 251–280.

JONES, M. C. and SILVERMAN, B. W. (1989). An orthogonal series density estimation approach to reconstructing positron emission tomography images. *Journal of Applied Statistics* **16** 177–191.

MCKEAGUE, I. W. (1986). Estimation for a semimartingale regression model using the method of sieves. *Ann. Statist.* **14** 579–589.

NATTERER, F. (1986). *The Mathematics of Computerized Tomography.* Teubner, Stuttgart.

SHEPP, L. A. and VARDI, Y. (1982). Maximum likelihood reconstruction in positron emission tomography. *IEEE Transactions on Medical Imaging* **1** 113–122.

SILVERMAN, B. W., JONES, M. C., WILSON, J. D. and NYCHKA, D. W. (1990). A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography (with discussion). *J. Roy. Statist. Soc. Ser. B* **52** 271–324.

VARDI, Y., SHEPP, L. A. and KAUFMAN, L. (1985). A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.* **80** 8–37.

DEPARTMENT OF MATHEMATICS
NATIONAL CENTRAL UNIVERSITY
CHUNGLI
TAIWAN
REPUBLIC OF CHINA

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI 11529
TAIWAN
REPUBLIC OF CHINA