Poisson Approximation of Image Processes in Computer Tomography

D. Pfeifer
Fachbereich Mathematik, Universität Oldenburg
Ammerländer Heerstr. 114–118, D–2900 Oldenburg

Abstract: We present estimations and asymptotic expansions for the total variation distance between the superposition of independent Bernoulli point processes and the Poisson point process with the same intensity measure. Special emphasis is given to the lattice case which arises in connection with the image reconstruction in computer tomography.

1. Introduction.

In the field of X–ray tomographic image reconstruction it is usually supposed that the emission of radioactive particles implanted in a patient's body follows a spatial Poisson point process with some absolutely continuous intensity measure μ. Emission of particles is recorded from outside the body by e.g. PET † scanners consisting of detector elements surrounding the patient's body; collected data are then used to estimate the intensity measure μ for instance by maximum likelihood methods (see Geman and McClure (1987) or Vardi, Shepp and Kaufman (1985)). Actually, the image domain is partitioned into a large number of *pixels* which means that instead of the hypothetical Poisson point process governed by a continuous intensity measure μ the Poisson proces with the corresponding discrete intensity measure ν which groups the mass of the pixel areas is considered. Hence the estimation problem reduces to the estimation of a vector of parameters for a Poisson process with discrete intensity measure ν, or, equivalently, Poisson distributed random vectors with independent components.

The assumption of a Poisson point process for the radioactive decay process carries, however, a systematic error since it would require an infinite source of particles with substitution of particles after each decay. Actually, implanting nuclear material in a patient's body creates a superposition of a finite (but large) number of independent Bernoulli processes, each of which describes the spatial position and the life length of the individual isotopes. Grouping by pixels then results into a Bernoulli point process with discrete intensity measure ν^* again, which will in general be close in distribution to a Poisson point process with the same intensity $\nu = \nu^*$. In general, the discrete Bernoulli process can be described by a string †† $\boldsymbol{S} = (S_1, \ldots, S_m)$ of independent Bernoulli $\mathfrak{B}(n_j, p)$–distributed random variables S_j, $1 \leq j \leq m$, with $n = \sum_{j=1}^{m} n_j$, and $p \in (0,1)$ where m is the number of pixels, n is the total number of particles implanted, n_j is the number (concentration) of isotopes in pixel no. j and p is the probability of decay during the observation period. Typically, p is very small (in the range $< 10^{-k}$ for some $k \in \mathbb{N}$), whereas the n_j are rather large (note that one mol of isotopes always contains $\approx 6 \cdot 10^{23}$ atoms, according to Avogadro's law). The discrete intensity measure ν^* hence has the form

$$\nu^*(B) = p \sum_{j \in B} n_j, \quad B \subseteq \mathbb{R}. \tag{1}$$

† **p**ositron **e**mission **t**omography

†† Note that the actual 2– or 3–dimensional structure of pixels can always be one–to–one transformed to such a string by a convenient numbering, preserving the independence structure of individual pixels.

H.-H. Bock et al. (eds.), *Classification, Data Analysis, and Knowledge Organization*

In what follows we shall present some estimations and asymptotic expansions for the total variation distance of such processes which have recently been obtained by Deheuvels and Pfeifer (1988).

2. The Model.

Let $(\Omega, \mathcal{A})$ be a measurable space and P, Q be probability distributions over $\mathcal{A}$. The *total variation distance* $d(P,Q)$ between P and Q is defined as

$$d(P,Q) = \sup_{A \in \mathcal{A}} \big|P(A) - Q(A)\big|. \tag{2}$$

We shall throughout use the same symbol "d" for the variational distance, even though spaces of different dimensions are considered in the sequel which, however, will not lead to confusions.

With the notations to (1), let $\boldsymbol{P}$ denote the distribution of the random vector $\boldsymbol{S}$, and $\boldsymbol{Q}$ denote the Poisson product distribution $\boldsymbol{Q} = \bigotimes_{j=1}^{m} \mathfrak{P}(n_j p)$. We have the following simple estimations for the variational distance between $\boldsymbol{P}$ and $\boldsymbol{Q}$:

$$d\big(\mathfrak{B}(n,p), \mathfrak{P}(np)\big) \le d(\boldsymbol{P}, \boldsymbol{Q}) \le \min\Big\{1, \sum_{j=1}^{m} d\big(\mathfrak{B}(n_j,p), \mathfrak{P}(n_j p)\big)\Big\}. \tag{3}$$

The left inequality can be obtained by choosing in (2) the particular sets $A = \bigcup_{k \in I} A_k$, $I \subseteq \{0,\dots,n\}$, with $A_k = \big\{(k_1,\dots,k_m) \in \mathbb{Z}^{+m} \mid \sum_{j=1}^{m} k_j = k\big\}$, $0 \le k \le n$, whereas the second inequality follows directly from $d(\boldsymbol{P}, \boldsymbol{Q})$ using the independence assumption and the well–known inequality

$$\Big|\prod_{j=1}^{m} a_j - \prod_{j=1}^{m} b_j\Big| \le \sum_{j=1}^{m} |a_j - b_j|$$

for all $a_1,\dots,a_m,\ b_1,\dots,b_m \in [0,1],\ m \in \mathbb{N}$.

With the estimations in Barbour and Hall (1984), we eventually obtain

$$\frac{1}{32} \min\{p, np^2\} \le d(\boldsymbol{P}, \boldsymbol{Q}) \le \min\Big\{1, \sum_{j=1}^{m} \min\{p, n_j p^2\}\Big\} \le \min\{1, mp, np^2\}. \tag{4}$$

When np gets large relation (4) (or (3)) gives a bound of the order $\min\{1, mp\}$ only where m is the number of pixels which in practical applications might be large also. Note that np is the average number of overall decays registrated by the scanner which may not be taken too small in order to enable sharp image reconstructions, and which will therefore in general be much larger than the number m of pixels. On the other hand, it has been shown by Prohorov (1953) that whenever np is large, the left hand side of (3) behaves like

$$d\big(\mathfrak{B}(n,p), \mathfrak{P}(np)\big) \sim \frac{p}{\sqrt{2\pi e}} \tag{5}$$

which indicates that the true variational distance should be proportional to p whenever np is large. By functional analytic methods, using convolution semigroups of operators, and normal approximations to quadratic forms of Poisson distributed random variables Prohorov's expansion (5) has been extended to the higher dimensional case in Deheuvels and Pfeifer (1988); an application of Theorem 2.3 there to the situation under consideration gives

$$d(\boldsymbol{P}, \boldsymbol{Q}) \sim \frac{1}{2} V_m \Big(\frac{m}{2\pi e}\Big)^{m/2} mp \tag{6}$$

whenever $\min\limits_{1\le j\le m}\{n_j p\}$ is large and p is small where V_m is the volume of the m–dimensional unit sphere, i.e.

$$V_m = \begin{cases} \dfrac{\pi^r}{r!} & \text{if } m = 2r \\ \dfrac{2^{2r+1} r! \pi^r}{(2r+1)!} & \text{if } m = 2r+1, \end{cases} \qquad r \in \mathbb{N}. \tag{7}$$

For m being large, $V_m\Big(\dfrac{m}{2\pi e}\Big)^{m/2} \sim \dfrac{1}{\sqrt{\pi m}}$, such that (6) becomes

$$d(\boldsymbol{P}, \boldsymbol{Q}) \sim \frac{1}{2}\sqrt{\frac{m}{\pi}}\, p = 0.28\sqrt{m}p \tag{8}$$

whenever $\min_{1\le j\le m}\{n_j p\}$ is large and p is small. This means that in the situation under consideration, the right magnitude of the approximation error is of the order $\sqrt{m}p$.
Similar considerations hold when the total variation distance between *point processes* ξ and ζ (as random elements in a space of measures) is considered since in $\mathbb{R}^2$ or $\mathbb{R}^3$ (or more generally, a polish space $\mathcal{X}$ which carries a metric that renders the space separable and complete) this distance can be reduced to the calculation of

$$d\big(P^{\xi}, P^{\zeta}\big) = \sup_{k\in\mathbb{N}} \sup_{B_1,\dots,B_k\in\mathcal{B}} d\Big(P^{(\xi(B_1),\dots,\xi(B_k))}, P^{(\zeta(B_1),\dots,\zeta(B_k))}\Big) \tag{9}$$

where $\mathcal{B}$ denotes the Borel σ–field over $\mathcal{X}$ and P is the underlying probability measure. In our case, if ξ denotes the Bernoulli point process and ζ the Poisson point process with intensity measure ν^* given by (1), (9) reduces once more to

$$d\big(P^{\xi}, P^{\zeta}\big) = \sup_{I\in\mathfrak{I}} d(\boldsymbol{P}_I, \boldsymbol{Q}_I) \tag{10}$$

where $\mathfrak{I}$ consists of all disjoint partitions $I = (I_1, \dots, I_k)$ $(1 \le k \le m)$ of subsets of $\{1, \dots, m\}$, and

$$\boldsymbol{P}_I = \bigotimes_{i=1}^{k} \mathfrak{B}\Big(\sum_{j\in I_i} n_j, p\Big), \quad \boldsymbol{Q}_I = \bigotimes_{i=1}^{k} \mathfrak{P}\Big(\sum_{j\in I_i} n_j p\Big), \quad I \in \mathfrak{I}. \tag{11}$$

Since the right hand side of (4) and (6), resp. is independent of the n_j (with $\sum_{i=1}^{k}\sum_{j\in I_i} n_j$ being bounded by n) and is monotonically increasing with m a comparison with (10) and (11) shows that the bound in (4) and the asymptotic evaluation in (8) remain also valid for $d\big(P^{\xi}, P^{\zeta}\big)$.

(Different approaches to this problem have been studied by Barbour (1988) using Stein's method, and Karr (1986, Proposition 1.46) who uses a particular coupling approach with a stochastically larger Poisson process; see also Witte (1988).)

3. Numerical evaluations.

According to Geman and McClure (1987) or Vardi, Shepp and Kaufman (1985) the number of pixels is typically in the range of $m \approx 100 \times 100 = 10^4$, whereas the overall number of particle counts is in the range of $\approx 10^6$. Further, the decay probability p w.r.t. one second of observation time can be expressed as

$$p = 0.0116 \cdot h^{-1} \tag{12}$$

where h is the half–life measured in minutes. If one assumes that the half–life of the material used is around 20 minutes (for instance, ^{11}C carbon), then the decay probability per second would be roughly $p \approx 6 \cdot 10^{-4}$. This shows that with one second of observation time the right hand side of (3) or (4) would not give a reasonable error bound for $n > 3 \cdot 10^6$. From (8), however, we obtain an error of approximately $28p \approx 0.015$ or 1.5% which is acceptable. Seemingly, the approximation error increases when the half–life of the isotope in use decreases. For oxygen ^{15}O which is for instance used in quantitative measurements of regional cerebral blood flow (see Frackowiak et al.) the half–life is only 2 minutes. In that paper, it is stated that the total number of particle counts is typically in the range of 7.5 to $10 \cdot 10^5$ within 5 minutes of scanning time (i.e. $n \approx 9$ to 12×10^5), i.e. the overall approximation error increases to around 15% for one second of observation time † which is not neglegible, especially since there is an additional image error of 3% due to prompt scatter coincidences, not to speak of the estimation error by the maximum–likelihood method itself. Here again the upper bounds in (3) and (4) are larger than one (with $n \approx 5200$ for one second of scanning time). Indeed, medical experiments seem to show that it is not very useful to choose too many pixels in the procedure of image reconstruction since this does not necessarily lead to a sharper image, which might in part also be due to the theoretical deviations in the model outlined above.

Acknowledgements.

We would very much like to thank Dr. Udo Kamps and Dr. Ingrid Kamps, Aachen, for some inspiring discussions and hints concerning the medical aspects of PET, and especially for bringing the paper by Frackowiak et al. to our attention.

References.

A.D. Barbour (1988): Stein's method and Poisson process convergence. J. Appl. Prob. 25A (special volume): *A Celebration of Applied Probability*, 175 – 184.

A.D. Barbour and P. Hall (1984): On the rate of Poisson convergence. Math. Proc. Camb. Phil. Soc. **95**, 473 – 480.

P. Deheuvels and D. Pfeifer (1988): Poisson approximations of multinomial distributions and point processes. J. Multivar. Analysis **25**, 65 – 89.

R.S.J. Frackowiak, G.-L. Lenzi, T. Jones and J.D. Heather (1980): Quantitative measurement of regional cerebral blood flow and oxygen metabolism in man using ^{15}O and positron emission tomography: theory, procedure, and normal values. J. Computer Ass. Tomography **6**, 727 – 736.

S. Geman and D.E. McClure (1987): Statistical methods for tomographic image reconstruction. Invited paper No. 21.1, 46^{th} Session of the ISI. Bulletin of the International Statistical Institute, Proceedings of the 46^{th} Session, Tokyo 1987, Vol. **4**, 5 – 21.

A.F. Karr (1986): *Point Processes and their Statistical Inference.* Marcel Dekker.

Y.V. Prohorov (1953): Asymptotic behavior of the binomial distribution. Usp. Mat. Nauk **8**, no. 3 (55), 135 – 142.

Y. Vardi, L.A. Shepp and L. Kaufman (1985): A statistical model for positron emission tomography. JASA **80**, 8 – 20.

H.J. Witte (1988): *Multivariate Poissonapproximation.* Ph.D. Thesis, RWTH Aachen.

† Note that for essentially larger time periods the asymptotic expansions in (6) will not necessarily hold.