PARAMETER ESTIMATION BY A MARKOV CHAIN MONTE CARLO TECHNIQUE FOR THE CANDY MODEL

X. Descombes [1], *M.N.M. van Lieshout* [2],

R. Stoica [2], *J. Zerubia* [1]

[1] Ariana, CNRS/UNSA/INRIA joint research group
INRIA, PO Box 93,
06902 Sophia Antipolis Cedex,
France

[2] CWI, PO Box 94079,
1090 GB Amsterdam
The Netherlands

ABSTRACT

This paper presents a parameter estimation method for the Candy model based on Monte Carlo approximation of the likelihood function. In order to produce such an approximation a Metropolis-Hastings style algorithm [3] for simulating the Candy model [10, 11] is introduced.

1. SET-UP AND NOTATION

In the last decade in image processing, a few researchers moved away from pixel-based methods to more high-level image analysis based on point process models. In this spirit, Stoica, Descombes and Zerubia [11] introduced a marked point process model for line segments, dubbed *Candy*, as prior distribution for the image analysis problem of extracting linear networks such as roads or rivers from images obtained by aerial and high resolution satellite photography.

More formally, represent a line segment as a point in some compact subset $K \subset \mathbb{R}^2$ of strictly positive volume $0 < \nu(K) < \infty$ with an attached mark taking values in $[l_{\min}, l_{\max}] \times [0, \pi)$ for some $0 < l_{\min} < l_{\max} < \infty$. Each marked point (k, l, θ) can be interpreted as a line segment with midpoint k, length l, and orientation θ. When applying the model to road extraction, it is natural to include marks for characteristics such as width and color as well. A configuration of line segments is a finite set of marked points. The probabilistic model is defined by its density p with respect to a unit rate Poisson process on K with independently and uniformly distributed marks as follows. At $s = \{s_1, \ldots, s_n\}$ with $s_i = (k_i, l_i, \theta_i) \in K \times [l_{\min}, l_{\max}] \times [0, \pi)$, $i = 1, \ldots, n$,

$$p(s) = \alpha \prod_{i=1}^{n} \exp\left[\frac{l_i - l_{\max}}{l_{\max}}\right]$$
$$\times \ \gamma_f^{n_f(s)} \gamma_s^{n_s(s)} \gamma_d^{n_d(s)} \gamma_o^{n_o(s)} \gamma_r^{n_r(s)} \quad (1)$$

where $\gamma_f, \gamma_s, \gamma_d > 0$ and $\gamma_o, \gamma_r \in (0, 1)$, are the model parameters. Stoica et al. recommend $\gamma_f < \gamma_s < \gamma_d$, in order to favor configurations containing more connected segments than free ones. The sufficient statistics $n_f(s)$, $n_s(s)$, $n_d(s)$, $n_o(s)$, $n_r(s)$ respectively represent the number of 'free' segments, the number of segments with a single one of its endpoints near another segment endpoint, the number of segments with both extremities connected, the number of pairs of segments crossing at too sharp angles, and the number of pairs that are disoriented. Thus, there are penalties attached to each free and singly connected segment, as well as to each sharp crossing and to every disagreement in orientation. For more details on the model and its applications to network extraction see [11], and [9] where the authors prove existence and Ruelle stability of p and establish various Markov properties.

2. METROPOLIS–HASTINGS ALGORITHMS

The Candy model (1) is too complicated to sample from directly. Hence, we apply Markov chain Monte Carlo techniques [6] to construct a Markov chain which has the Candy model π as its equilibrium distribution. Here we use the Metropolis–Hastings sampler, a flexible proposal-acceptance technique that is well adapted to point processes [3, 7]. In its generic form, the transition proposals are uniformly distributed births and deaths. The acceptance probabilities are based on the likelihood ratio of the new state compared to the old one. Due to the results in [2], the algorithm converges in total variation to π for π-almost all initial configurations provided. The theorem applies equally to any other pair of strictly positive birth and death kernels.

In order to improve mixing, we incorporate transitions that are tailor-made for the Candy model. Thus, we include a birth kernel that tends to add a segment in order to prolongate the current network. The idea is that when adding a segment, preference should be given to positions that 'fit' the current configuration. More specifically, a new seg-

ment might be positioned in such a way that it is connected to an endpoint of a segment in the configuration, see [9]. For computational convenience, we only connect to segment endpoints that are sufficiently far from the boundary of K.

Another option is to include transition types other than births and deaths. For instance in [2] change transitions that do not alter the number of segments are described.There are many valid choices for the proposal kernel. For instance, we may shift a segment center a bit, modify the orientation and/or the length, or even discard a segment altogether and generate a new one randomly. For more details see [9].
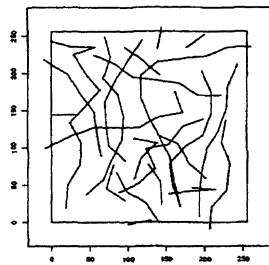
Model parameters
$\gamma_f = 0.0002$
$\gamma_s = 0.05$
$\gamma_d = 12.2$
$\gamma_o = 0.08$
$\gamma_r = 0.08$

Sufficient statistics
$n_f = 4$
$n_s = 34$
$n_d = 63$
$n_o = 12$
$n_r = 9$

Fig. 1. Realization (top) of the Candy model given by the parameters in the middle table. The observed values of the sufficient statistics are listed below.

In Figure 1 we present a sample of the Candy model, its parameters and the observed values of the sufficient statistics. We carried out 2×10^7 iterations. The sufficient statistics were taken every 10^3 iterations. The point space is $K = [0, 256] \times [0, 256]$ while marks take values in $[30, 40] \times [0, \pi)$.The weights of the different transition kernels were fixed empirically. The Candy model is very complex, hence

it is difficult to assess convergence. However, we may analyze the evolution of the cumulative means of the sufficient statistics during the simulation. These are plotted in Figure 2.

$\bar{n}_f = 3.64$

$\bar{n}_s = 28.63$

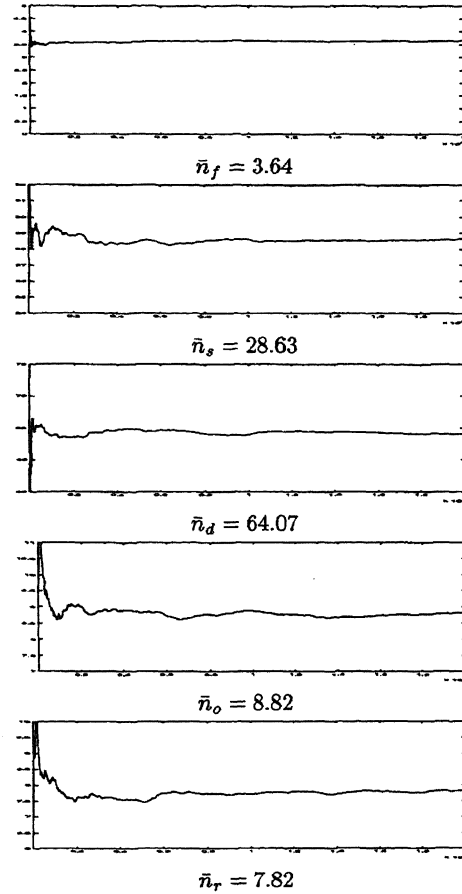$\bar{n}_d = 64.07$

$\bar{n}_o = 8.82$

$\bar{n}_r = 7.82$

Fig. 2. Evolution of the empirical moments of the sufficient statistics during the simulation of the Candy model. The cumulative means $\bar{n}_f, \bar{n}_s, \bar{n}_d, \bar{n}_o, \bar{n}_r$ (from top to bottom) are plotted as a function of the number of iterations.

3. MAXIMUM LIKELIHOOD ESTIMATION

The Candy model (1) is a five-parameter exponential family

$$p_\omega(s) = \alpha(\omega) \exp \left[t(s)^T \omega \right] h(s)$$

with canonical sufficient statistic

$$t(s) = \left(n_f(s), n_s(s), n_d(s), n_o(s), n_r(s) \right)^T,$$

parameter vector

$$\omega = (\log \gamma_f, \log \gamma_s, \log \gamma_d, \log \gamma_o, \log \gamma_r)^T,$$

and $h(s) = \prod_{i=1}^{n} \exp\left[\frac{l_i - l_{max}}{l_{max}}\right]$. Using the importance sampling ideas outlined in [4, 5] the ratio of normalizing constants can be expressed as

$$\alpha(\omega_0)/\alpha(\omega) = E_{\omega_0} \exp\left[t(S)^T(\omega - \omega_0)\right]$$

and the log likelihood ratio with respect to some reference value ω_0 can be written as

$$l(\omega) = \log \frac{p_\omega(s)}{p_{\omega_0}(s)}$$
$$= t(s)^T(\omega - \omega_0) - \log E_{\omega_0} \exp\left[t(S)^T(\omega - \omega_0)\right].$$
$$(2)$$

The score equations $\nabla l(\omega) = t(s) - E_\omega T(S)$ and Fisher information matrix $I(\omega) = -\nabla^2 l(\omega) = \mathrm{Var}_\omega t(S)$ are easily derived, so that under the maximum likelihood estimator $\hat{\omega}$, the expected values of the sufficient statistics must be equal to the observed values. Now, since the covariance matrix of $t(S)$ is positive definite, (2) is concave in ω. Therefore, provided the score equations have a solution $\hat{\omega}$ in $\mathbb{R} \times \mathbb{R}^4_-$, a unique maximum likelihood estimator exists and equals $\hat{\omega}$. Otherwise, a maximum may be found on the boundary of the parameter space.

Numerically, the expectation in (2) can be approximated [4, 5] by its Monte Carlo counterpart

$$\frac{1}{n} \sum_{i=1}^{n} \exp\left[t(S_i)^T(\omega - \omega_0)\right]$$

based on a single sample $S_1, \ldots, S_n$ from p_{ω_0}.

Considering the true unknown MLE $\hat{\omega}$, due to [2, Theorem 7] the Monte Carlo maximum likelihood estimator is consistent and satisfies the central limit theorem :

$$\sqrt{(n)} (\hat{\omega}^n - \hat{\omega}) \to \mathcal{N}(0, I(\hat{\omega})^{-1}\Sigma I(\hat{\omega})^{-1})$$

where Σ is the asymptotic covariance matrix of the normalized Monte Carlo score $\sqrt{(n)}\nabla l_n(\hat{\omega})$ and $I(\hat{\omega})$ denotes the Fisher information matrix at the maximum likelihood estimator.

However, the method described above relies on a reference value ω_0 that is not too far from the maximum likelihood estimator. Here we used the iterative gradient method [1].

$$\begin{cases} l_n(\omega_k + \rho(\omega_k)\nabla l_n(\omega_k)) = \\ \quad = \max_{\rho \in \mathbb{R}} l_n(\omega_k + \rho \nabla l_n(\omega_k)) \\ \omega_{k+1} = \omega_k + \rho(\omega_k)\nabla l_n(\omega_k) \end{cases} \quad (3)$$

to find a reasonable value. Here $\rho(\omega_k)$ is the optimal step, which is computed using a one-dimensional minimization of the likelihood function.

We implemented the procedure for the data of Figure 1. Starting with some arbitrary initial values (see Figure 3, first column) we ran (3) for 1000 steps to obtain the vector ω_0 listed in the second column of Figure 3. Based on a sample of size $n = 2 \times 10^7$ from p_{ω_0}, we calculated the Monte Carlo approximation $l_n(\omega)$, cross sections of which are shown in Figure 5. The maximum of $l_n(\omega)$ is located at $\hat{\omega}^n$ as listed in Figure 3 (third column).

In Figure 4 we show the asymptotic standard deviation of the true MLE, and the Monte Carlo Standard Error (MCSE) which approximates the difference between the unknown MLE and its Monte Carlo approximation. We notice that by increasing n, we can make the MCSE negligible.

Initial parameters	Iterative method	Monte Carlo MLE
$\omega_f^i = -9.5$	$\hat{\omega}_f^0 = -8.37$	$\hat{\omega}_f^n = -8.32$
$\omega_s^i = -4.0$	$\hat{\omega}_s^0 = -2.74$	$\hat{\omega}_s^n = -2.73$
$\omega_d^i = 1.5$	$\hat{\omega}_d^0 = 2.46$	$\hat{\omega}_d^n = 2.47$
$\omega_o^i = -3.5$	$\hat{\omega}_o^0 = -2.13$	$\hat{\omega}_o^n = -2.17$
$\omega_r^i = -3.5$	$\hat{\omega}_r^0 = -2.42$	$\hat{\omega}_r^n = -2.42$

Fig. 3. Estimation of the parameters for the data of Figure 1.

Asymptotic standard deviation of MLE	MCSE
0.51	0.002
0.23	0.003
0.17	0.001
0.30	0.002
0.31	0.005

Fig. 4. Estimation errors.

4. CONCLUSION AND FUTURE WORK

In practice, the main challenges in working with point processes are the following: to build appropriate moves, to find the optimal way of combining them into a simulation algorithm, and to carry out statistical inference. Here, we have built a Metropolis–Hastings sampler, that combines uniform birth and death proposals that guarantee the convergence of the Markov chain to the target equilibrium distribution (1) with transitions designed to exploit specific characteristics of the model, in our case connectivity properties.

The main application of the Candy model is that of thin network extraction. This was the topic of [10, 11], where results were obtained using fixed pareameters as well as approximations to the Metropolis–Hastings proposal kernels and acceptance probabilities. The results here, and in [9],

remove the need for approximate sampling, and may be a starting point for unsupervised network extraction.

5. REFERENCES

[1] X. Descombes, R. D. Morris, J. Zerubia and M. Berthod. Estimation of Markov random field prior parameters using Markov Chain Monte Carlo maximum likelihood. *IEEE Transactions on Image Processing* **8**, 954-963, 1999.

[2] C.J. Geyer. On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society, Series B* **56**, 261-274, 1994.

[3] C.J. Geyer and J. Møller. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics* **21**, 359-373, 1994.

[4] C.J. Geyer. Likelihood inference for spatial point processes. In O. Barndorff-Nielsen, W.S. Kendall, and M.N.M. van Lieshout, editors, *Stochastic geometry, likelihood, and computation*, CRC Press/Chapman and Hall, Boca Raton, 1999.

[5] C.J. Geyer and E.A. Thompson. Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, Series B* **54**, 657-699, 1992.

[6] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall, London, 1996.

[7] P.J. Green. Reversible jump MCMC computation and Bayesian model determination. *Biometrika* **82**, 711-732, 1995.

[8] M.N.M. van Lieshout. *Markov point processes and their applications*. Imperial College Press/World Scientific Publishing, London/Singapore, 2000.

[9] M.N.M. van Lieshout and R.S. Stoica. The Candy model revisited : Markov properties and inference. CWI Research Report, 2001.

[10] R. Stoica. *Processus ponctuels pour l'extraction des réseaux linéiques dans les images satellitaires et aériennes*. PhD Thesis (in French), University of Nice-Sophia Antipolis, 2001.

[11] R. Stoica, X. Descombes and J. Zerubia. A Gibbs point process for road extraction in remotely sensed images. Research Report 3923, INRIA Sophia Antipolis, 2000.
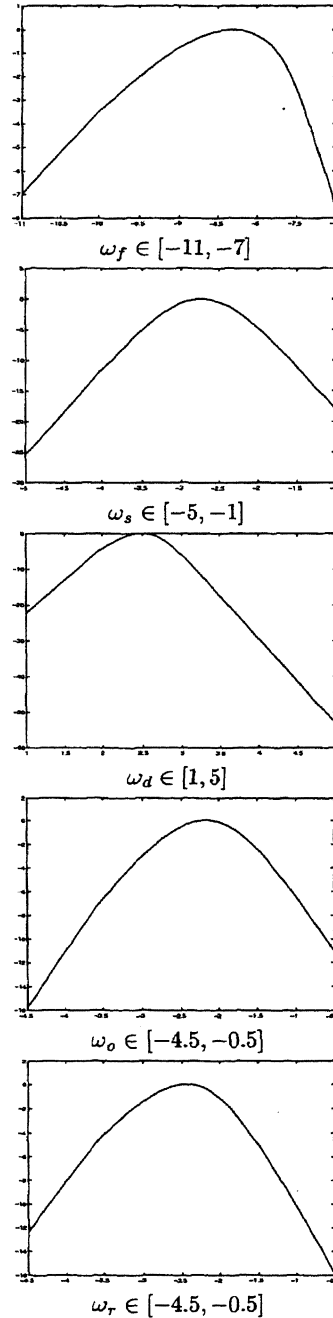
Fig. 5. Monte Carlo approximation of the log likelihood function. The X axis represents the variation of a single component. The Y axis represents the values of the Monte Carlo log likelihood with all other components of $\hat{\omega}^0$ fixed.