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

X. Descombes 1, M.N.M. van Lieshout 2,

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

Note: This research was supported by NWO grant 'Inference for random sets' (613-03-045) and INRIA grant via ERCIM.

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 \([l_{\text{min}}, l_{\text{max}}] \times [0, \pi)\) for some \( 0 < l_{\text{min}} < l_{\text{max}} < \infty \). Each marked point \((k, l, \theta)\) can be interpreted as a line segment with midpoint \(k\), length \(l\), and orientation \(\theta\). 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 \( s = \{s_i, \ldots, s_n\} \) with \( s_i = (k_i, l_i, \theta_i) \in K \times [l_{\text{min}}, l_{\text{max}}] \times [0, \pi)\), \( i = 1, \ldots, n \),

\[
p(s) = \alpha \prod_{i=1}^{n} \exp \left[ \frac{l_i - l_{\text{min}}}{l_{\text{max}} - l_{\text{min}}} \right] \\
\times \gamma_f \gamma_r \gamma_t \gamma_d \gamma_0 \gamma_c \rho_f(s) \rho_r(s) \rho_d(s) \rho_0(s) \rho_c(s) (1)
\]

where \( \gamma_f, \gamma_r, \gamma_t, \gamma_d > 0 \) and \( \gamma_0, \gamma_c \in (0, 1) \), are the model parameters. Stoica et al. recommend \( \gamma_f < \gamma_r < \gamma_d \), in order to favor configurations containing more connected segments than free ones. The sufficient statistics \( n_f(s), n_r(s), n_t(s), n_d(s), n_0(s), n_c(s) \) 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 \( \pi \) 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 \( \pi \) 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 prolong 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 seg­
ment 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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_f$</td>
<td>0.0002</td>
</tr>
<tr>
<td>$\gamma_a$</td>
<td>0.05</td>
</tr>
<tr>
<td>$\gamma_d$</td>
<td>12.2</td>
</tr>
<tr>
<td>$\gamma_o$</td>
<td>0.08</td>
</tr>
<tr>
<td>$\gamma_r$</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Sufficient statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_f$</td>
<td>4</td>
</tr>
<tr>
<td>$n_a$</td>
<td>34</td>
</tr>
<tr>
<td>$n_d$</td>
<td>63</td>
</tr>
<tr>
<td>$n_o$</td>
<td>12</td>
</tr>
<tr>
<td>$n_r$</td>
<td>9</td>
</tr>
</tbody>
</table>

Fig. 1. Realization (top) of the Candy model given by the
parameters in the middle. 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 statis­
tics. We carried out $2 \times 10^7$ 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 ana­
lyze the evolution of the cumulative means of the sufficient
statistics during the simulation. These are plotted in Fig­
ure 2.

![Graphs showing evolution of empirical moments of sufficient statistics](image)

**Fig. 2.** Evolution of the empirical moments of the sufficient
statistics during the simulation of the Candy model. The
cumulative means $n_f, n_a, n_d, n_o, n_r$ (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

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

with canonical sufficient statistic

$$t(s) = (n_f(s), n_a(s), n_d(s), n_o(s), n_r(s))^T,$$
parameter vector
\[ \omega = (\log \gamma_1, \log \gamma_2, \log \gamma_3, \log \gamma_4, \log \gamma_5)^T, \]
and
\[ h(s) = \prod_{i=1}^{n} \exp \left[ \frac{\log_{\text{max}}}{\log_{\text{min}}} \right]. \]

Using the importance sampling ideas outlined in [4, 5], the ratio of normalizing constants can be expressed as
\[ \alpha_0(\omega_0) / \alpha(\omega) = E_{\omega_0} \exp \left[ t(S)^T(\omega - \omega_0) \right] \]
and the log likelihood ratio with respect to some reference value \( \omega_0 \) 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 \( \nabla l(\omega) = t(s) - E_{\omega_0} T(S) \) and Fisher information matrix \( I(\omega) = -\nabla^2 l(\omega) = \text{Var}_{\omega_0}(t(S)) \) 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 \( \omega \). Therefore, provided the score equations have a solution \( \hat{\omega} \in \mathbb{R} \times \mathbb{R}^n \), 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
\[ \frac{1}{n} \sum_{i=1}^{n} \exp \left[ t(S_i)^T(\omega - \omega_0) \right] \]
and
\[ \frac{1}{n} \sum_{i=1}^{n} \exp \left[ t(S_i)^T(\omega - \omega_0) \right]. \]
(3)

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 \( \omega_0 \) listed in the second column of Figure 3. Based on a sample of size \( n = 2 \times 10^7 \) from \( p_{\omega_0} \), 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 \( \hat{\omega} \) 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.

![Fig. 3. Estimation of the parameters for the data of Figure 1.](image)

<table>
<thead>
<tr>
<th>Initial parameters</th>
<th>Iterative method</th>
<th>Monte Carlo MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega_1 ) = -9.5</td>
<td>( \omega_1 ) = -8.37</td>
<td>( \omega_1 ) = -8.32</td>
</tr>
<tr>
<td>( \omega_2 ) = -4.0</td>
<td>( \omega_2 ) = -2.74</td>
<td>( \omega_2 ) = -2.73</td>
</tr>
<tr>
<td>( \omega_3 ) = 1.5</td>
<td>( \omega_3 ) = 2.46</td>
<td>( \omega_3 ) = 2.47</td>
</tr>
<tr>
<td>( \omega_4 ) = -3.5</td>
<td>( \omega_4 ) = -2.13</td>
<td>( \omega_4 ) = -2.17</td>
</tr>
<tr>
<td>( \omega_5 ) = -3.5</td>
<td>( \omega_5 ) = -2.42</td>
<td>( \omega_5 ) = -2.42</td>
</tr>
</tbody>
</table>

![Fig. 4. Estimation errors.](image)

### 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 parameters 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


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 $\omega^0$ fixed.