Author: Rossum, A.C. van
Title: Nonparametric Bayesian methods in robotic vision
Issue date: 2021-06-03

CHAPTER 4

NONPARAMETRIC BAYESIAN SEGMENT ESTIMATION

Contents In this chapter, we introduce a Bayesian method to perform inference over line segments. In this model, infinite segment model (ISM), the prior for the location is given by a Normal distribution, the prior for the length of the segment is given by a Pareto distribution. Due to the fact that the prior and likelihood do not form a conjugate pair, a more general inference method is used (than the inference methods for the conjugate model in Chapter 3), namely Gibbs sampling with auxiliary variables.

Published in A.C. van Rossum, H.X. Lin, J. Dubbeldam, and H.J. van den Herik. Nonparametric Segment Detection. *Proceedings of the Eighth European Starting AI Researcher Symposium*, STAIRS 2016, the Hague, the Netherlands, August 26-27, 2016.

Outline Our proposed model is using both a Normal-Inverse-Gamma distribution and a Normal and Pareto distribution as priors for an individual line segment (Section 4.1). Inference over the infinite segment model is done using Gibbs sampling over auxiliary variables (Section 4.2). The results for inference over line segments are compared with those for lines (Section 4.3). Finally, weak aspects of the current MCMC method are established (Section 4.4). They will form the basis for new inference methods in the next chapters.

4.1 Infinite Segment Model

The application we would like to address in this chapter is that of the detection of multiple segments rather than lines. We will label the model is the infinite segment model. The term infinite relates to the use of a nonparametric Bayesian prior. The term does not reflect the size of the segments.

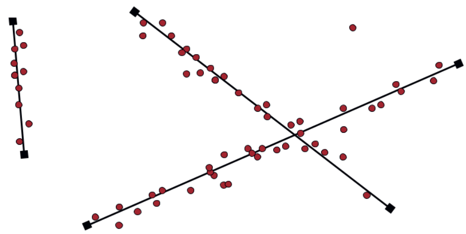


Figure 4.1: A mixture of segments. The segments have two more parameters compared to lines: the length of the segment and its center (or alternatively, the endpoints of the segment). Analogous to the line detection application, there are n points in 2D space, each point generated from a segment with parameters θ_k. The number of segments, k, is not known beforehand. Compare with Figure 3.1.

We will model the infinite segment model similar to the infinite line model, namely as a Dirichlet process mixture:

$$\begin{aligned} G &\sim DP(\alpha, H), \\ \theta_i \mid G &\overset{iid}{\sim} G, \\ w_i \mid \theta_i &\overset{iid}{\sim} F(w_i; \theta_i). \end{aligned} \tag{4.1}$$

The likelihood function F describes the mapping from parameters θ_i to observations w_i. In the previous chapter this has been a likelihood function that describes points on lines. In this chapter the likelihood function describes points on line segments.

Along the same lines as in Chapter 3 we have a base distribution H, a dispersion factor α, and hyperparameters for the base distribution λ_0.

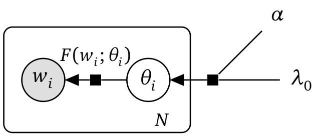


Figure 4.2: The Dirichlet process mixture with hyperparameter λ_0 for the base distribution H.

For each point w_i the segment parameters are given by θ_i. The parameters θ_i are not necessarily unique (for $i \neq j$). When we iterate over unique segments we will use the subscript k rather than i or we will mention this explicitly.

4.1.1 Posterior Predictive for a Segment given Other Segments

This follows exactly the same derivation as for the infinite line model in Section 3.2.1. The posterior predictive is given by (Neal, 2000):

$$\theta_n \mid \theta_1, \ldots, \theta_{n-1} \sim \frac{1}{\alpha + n - 1}\left(\alpha H + \sum_{j=1}^{n-1} \delta_{\theta_j}\right). \tag{4.2}$$

The prior distribution of parameters θ_i takes the form of conditional distributions:

$$\theta_i \mid \theta_{-i} \sim \frac{1}{\alpha + n - 1}\left(\alpha H + \sum_{j \neq i} \delta_{\theta_j}\right). \tag{4.3}$$

The notation θ_{-i} describes every other parameter than θ_i: the set of parameters, θ_j, with $j \neq i$.

4.1.2 Likelihood of Data given Segment Parameters

The likelihood $F(w_i, \theta_i)$ describes the mapping from parameters θ_i to observations w_i. We create a likelihood function by the combination of two probability density functions. The observation w_i has x-coordinate x_i and y-coordinate y_i. We sample x_i from a uniform distribution only giving it nonzero probability on a particular segment on the x-axis:

$$x_i \mid c, d \overset{iid}{\sim} U(c - d, c + d). \tag{4.4}$$

This defines a segment on the x-axis centered at c which extends in both directions with size d. We will use an intercept-slope representation (Chapter 3). Let us define $X_i = [1, x_i]$ with x_i distributed as in Eq. 4.4. The column vector $\beta = [\beta_0, \beta_1]$ contains two parameters: the y-intercept β_0 and the slope parameter β_1 (compare Section 3.2.2). And we assume a normally distributed random variable across $y - X\beta$, the same as in the line model (Eq. 3.5):

$$y_i \overset{iid}{\sim} N(X_i \beta_k, \sigma_k^2). \tag{4.5}$$

The combination of Eq. 4.4 and Eq. 4.5 generates points across a segment on a line.

4.1.3 Prior for a Segment

We postulate a prior that is a combination of Bayesian linear regression with restrictions on the size of the line:

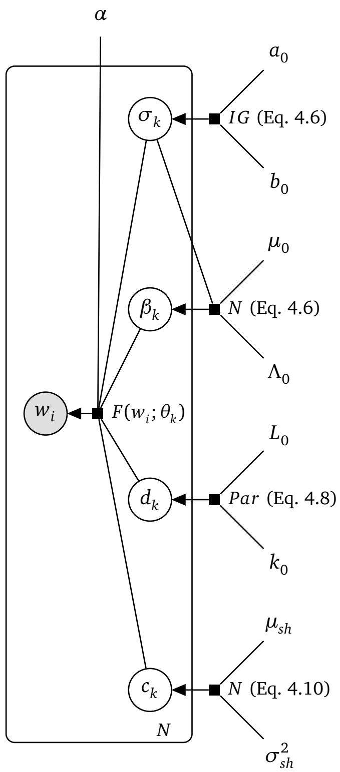


Figure 4.3: The segment parameters for segment k are $\theta_k = \{\sigma_k, \beta_k, d_k, c_k\}$. Here σ_k and β_k are sampled from the same distributions (an Inverse-Gamma, respectively, a Normal distribution) as in the infinite line model. The extend of the segment, d_k, is defined by a Pareto distribution and its center, c_k, by a Normal distribution.

The slope and intercept parameters of the segment are sampled according to a Normal-Inverse-Gamma distribution (compare Eq. 3.16 for line parameters):

$$\begin{aligned} \sigma_k^2 &\sim IG(a_0, b_0), \\ \beta_k &\sim N(\mu_0, \sigma_k^2 \Lambda_0^{-1}). \end{aligned} \tag{4.6}$$

Recall that the data on a line segment is distributed uniformly (Eq. 4.4). This is parametrized through two parameters, the center of the segment, c, and its extent, d:

$$x \mid c, d \sim U(c - d, c + d). \tag{4.7}$$

We propose as a prior for the extend of the line segment d, a Pareto distribution (Par):

$$d \mid L_0, k_0 \sim Par(L_0, k_0). \tag{4.8}$$

The Pareto distribution (Par) is given by:

$$p(d|L_0, k_0) = \begin{cases} k_0 L_0^{k_0} d^{-k_0-1} & if\, d \geq L_0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.9}$$

The parameter L_0 can be seen as a prior parameter that sets a minimal size to the line segment. The parameter k_0 is the shape parameter of the Pareto distribution.

The center of the segment is sampled from a Normal distribution:

$$c \mid \mu_{sh}, \sigma_{sh}^2 \sim N(\mu_{sh}, \sigma_{sh}^2). \tag{4.10}$$

The subscript in μ_{sh} and σ_{sh}^2 stands for shifted. The center of the segment is shifted along the line.

We will collect all priors and call it a Segment prior, abbreviated to Seg.

$$\theta_k \sim Seg(\lambda_0) \tag{4.11}$$

Writing out all parameters:

$$\beta_k, \sigma_k^2, d_k, c_k \sim Seg(a_0, b_0, \mu_0, \Lambda_0, L_0, k_0, \mu_{sh}, \sigma_{sh}^2). \tag{4.12}$$

This corresponds to:

$$\begin{aligned} \beta_k, \sigma_k^2 &\sim NIG(a_0, b_0, \mu_0, \Lambda_0), \\ d_k &\sim Par(L_0, k_0), \\ c_k &\sim N(\mu_{sh}, \sigma_{sh}^2) \end{aligned} \tag{4.13}$$

4.1.4 Sampling Segment Parameters given Data

In contrast to the infinite line model there is no conjugacy between prior and likelihood in the infinite segment model. We have no closed-form updates for hyperparameters given observed data. Hence, we have to resort to sampling parameters. The proposal distribution, Q, with which we sample new parameters can be using the current state, θ_k, or it can sample from the prior λ_0, or a combination thereof:

$$\theta_k \sim Q(\theta_k, \lambda_0) \tag{4.14}$$

Observations are sampled independently from line parameters (Section 4.1.2), hence the likelihood of a set of observations is described by the product.

$$L_k = \prod_i p(\theta_k | w_i) \tag{4.15}$$

We can sample θ_{new} from $Seg(\lambda_0$ and then accept with probability L_{new}/L_k. Alternatively we can sample using an MCMC proposal distribution around θ_k:

$$\theta_k \sim N(\theta_k, \sigma^2_{prop}). \tag{4.16}$$

Alternatively, we can sample in a way that reflects our priors. For example, taking turns and sample first β_k, σ_k^2 from a NIG distribution keeping d_k, c_k the same and the other way around, sample d_k, c_k from a Pareto-Normal distribution and keep β_k, σ_k^2 the same.

4.2 Inference for the Infinite Segment Model

Let us introduce *Gibbs sampling with auxiliary variables* (Neal, 2000), see Algorithm 10.

Algorithm 10 Gibbs sampling with auxiliary variables

procedure GIBBS ALGORITHM WITH AUXILIARY VARIABLES(w, λ_0, α) ▷ Accepts points w and hyperparameters λ_0 and α. Requires also the number of auxiliary variables V, a proposal distribution Q. Returns k line coordinates.
  for all $t = 1 : T$ **do**
    for all $i = 1 : N$ **do**
      for all $v = 1 : V$ **do**
        $\theta_v \sim Seg(\lambda_0)$ ▷ Sample from Eq. 4.11.
        $m_v = \alpha/V$
      end for
      $c = \text{cluster}(w_i)$ ▷ Get cluster c currently assigned to observation w_i.
      $m_c = m_c - 1$ ▷ Adjust cluster size m_c (and bookkeeping of K).
      for all $k = 1 : K + m$ **do**
        $L_k = m_k F(w_i; \theta_k)$ ▷ Calculate likelihood for all θ_k.
      end for
      $k \sim Mult(K + m, L_k)$ ▷ Sample k from all clusters (weighed by m_k cq m_v).
      $\theta_i = \theta_k$ ▷ Set θ_i to sampled cluster.
      $m_k = m_k + 1$ ▷ Increment m_k (set to 1 for m_v, and adjust K).
    end for
    for all $k = 1 : K$ **do**
      $\theta_{prop} \sim Q(\theta_k, \lambda_0)$ ▷ Sample from proposal distribution (Eq. 4.14).
      $L_{prop} = \prod_i F(w_i; \theta_{prop})$ ▷ Likelihood for all i at θ_{prop}.
      $L_k = \prod_i F(w_i; \theta_k)$ ▷ Likelihood for all i at θ_k.
      $u \sim U(0, 1)$
      if $(L_{prop}/L_k) > u$ **then** ▷ Accept/reject.
        $\theta_k = \theta_{prop}$
      end if
    end for
  end for
  return summary on θ_k for k line segments.
end procedure

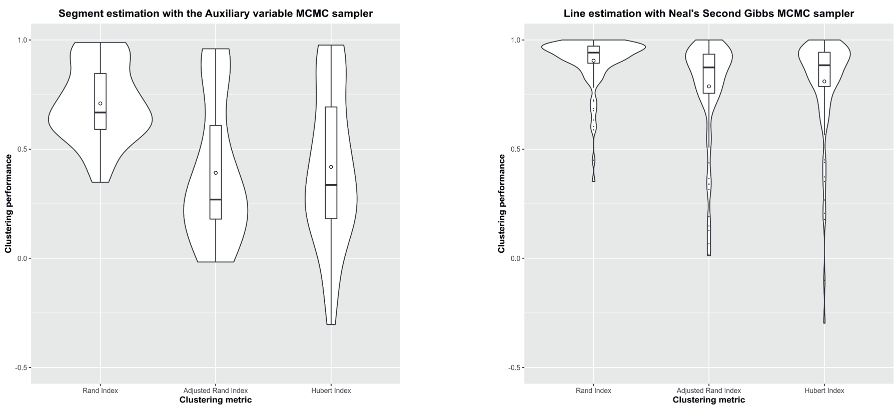
This Gibbs algorithm[1] has been described before in the context of a Dirichlet process mixture, without particular likelihoods or priors in mind (see algorithm 8 in Neal, 2000). The sampling process proposes V new values for the parameters from the hyperparameters. The V values are called auxiliary parameters. Now, to establish to which cluster a certain observation w_i needs to be assigned, the likelihood of each existing and new clusters alike are compared. The weight of an old cluster is defined through the number of data points assigned to it. The weight of a new cluster is defined through α/V. After every data item is assigned a cluster, the cluster parameters themselves are updated given the assigned data items.

4.3 Results

We show a drop in performance for segment detection compared to line detection in Section 4.3.1. Some examples of difficult to assign segments are given in Section 4.3.2. We visualize (the lack of) convergence in Section 4.3.3.

4.3.1 Clustering Performance

The results over a larger dataset can be measured with clustering metrics as visualized in Figure 4.4. The clustering performance of the segment detection algorithm, measured by the clustering index, such as the Rand Index, the Adjusted Rand Index, and the Hubert metric, show all reduced performance (see Figure 4.4) compared to line detection (without constraints on segment size).



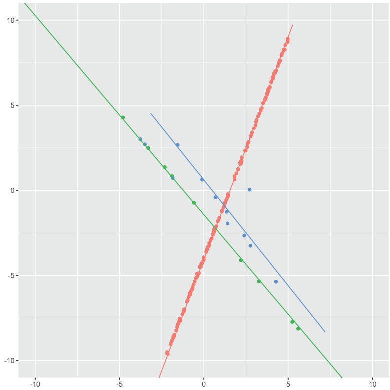
(a) Segment detection. **(b)** Line detection.

Figure 4.4: Segment detection performs worse than line detection across all three clustering performance indicators. Perfect clustering is indicated by 1.0 for Rand Index, Adjusted Rand Index, and Hubert.

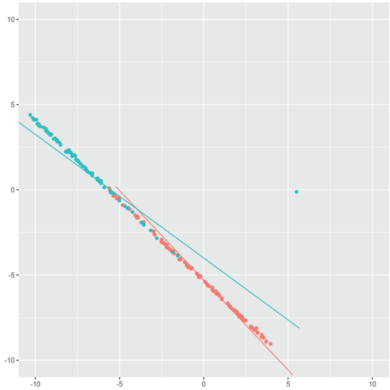
[1] The implementation can be found at `https://code.annevanrossum.nl/dpm` in the folder inference (gibbsDPM_algo8), written such that it is compatible with octave.

4.3.2 Examples

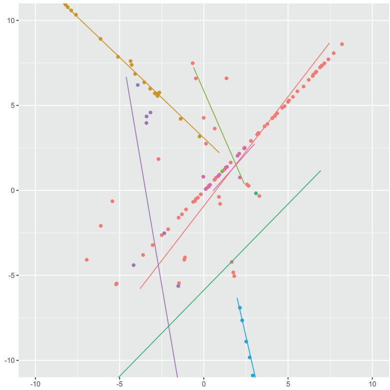
In Figure 4.5 we show four Bayesian point estimates of the sampling process. These are examples that demonstrate the type of errors that are made in the inference process. In example (a) the segments are correctly sampled. In (b) the type of error is that of recognizing multiple segments where there is only one segment to the human observer. In (c) the error is due to the fact that some segments contain very few points. In (d) the error stems from line segments being chosen orthogonal to the actual segment.



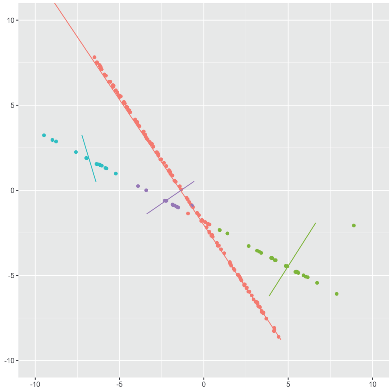
(a) There is an outlier right of the center. Also, the line segments that have fewer points, have end-points that are recognized less "tight" (to be expected given the Pareto prior).



(b) The single line segment is incorrectly recognized as multiple segments.



(c) The segments with fewer observations are recognized poorly.

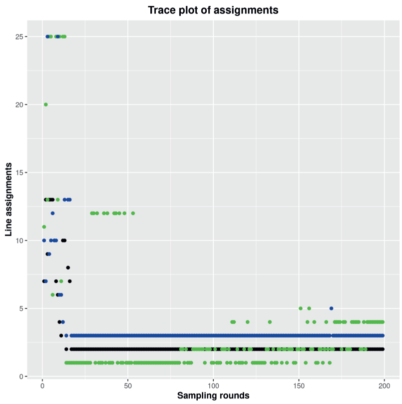


(d) Line segments are (incorrectly) chosen to be orthogonal to the lines.

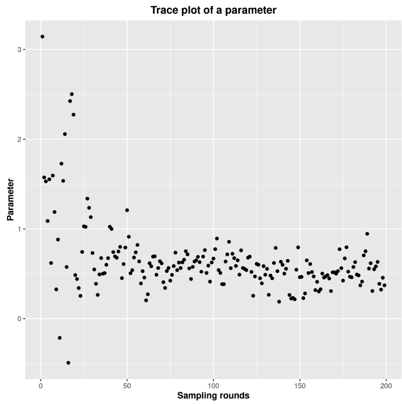
Figure 4.5: Bayesian point estimates of the sampling process with varying types of sampling errors. The descriptions indicate what type of sampling error is visualized per subfigure.

4.3.3 Trace Plots

To study the convergence of parameters in the infinite segment model, we use trace plots.



(a) This plot traces three points that are assigned to clusters (limited to around 30). Two of the points are assigned to one cluster. The other point to the cluster at the bottom. The plot only shows accepted assignments. The acceptance of a new assignment takes rarely hold.



(b) This plot traces a segment parameter belonging to point w_i. It exhibits exploratory behavior around a particular value (in this case 0.5). Compared to Figure 3.14 the variance is quite large.

Figure 4.6: Two examples of trace plots. Left: a trace plot of the assignment of points to cluster (it changes not so often). Right: a trace plot of one of the parameter values assigned to w_i.

4.4 Chapter Conclusions

From Chapter 3 we know that segment estimation is a much harder problem than line estimation. In this chapter we used an advanced method, namely Gibbs sampling with auxiliary variables to perform inference over an infinite set of line segments (Van Rossum et al., 2016c). The auxiliary variable Gibbs sampling method converges faster than the ordinary Metropolis-Hastings sampling algorithm by postulating multiple segments rather than only one.

This chapter contributes to answering our first research question.

RQ 1 How can we estimate the number of objects simultaneously with the fitting of these objects?

To estimate the number of objects simultaneously with the fitting of those objects, we have used a Bayesian method (as in the previous chapter). In this chapter, the prior and likelihood for the line segment model does not form a conjugate pair. Hence, different sampling methods had to be used to perform inference for the introduced Bayesian model.

However, the segment estimation problem remains a challenge for the inference method in this chapter. The target probability density has modes that each needs to be found and tend to be separated by very low probability regions. In Chapter 5 we will introduce new sampling methods that will cope with this challenge.