Notes on Differential Entropy Calculation Using Particles

Umut Orguner
Division of Automatic Control
E-mail: umut@isy.liu.se

15th August 2008

Report no.: LiTH-ISY-R-2857

Address:
Department of Electrical Engineering
Linköpings universitet
SE-581 83 Linköping, Sweden

WWW: http://www.control.isy.liu.se
Abstract
This report outlines a method to calculate the differential entropy of a probability density represented by a number of particles and weights. When only the particles and the weights are given, entropy calculation is cumbersome and requires continuous approximation of the density by using some kernel functions. However, in a particle filtering framework, Bayes rule provides a direct sample based approximation to the problem.

Keywords: Entropy, particle filter
This document suggests a method to calculate the differential entropy of particle mixtures. The basic problem can be seen as that, when we are given only the approximation

\[
p(x_k|y_{0:k}) \approx \sum_{i=1}^{N} \pi_{k|i}^{(i)} \delta(x_k - x_{k|i}^{(i)})
\]

(1)

the differential entropy \(\mathcal{H}(p(x_k|y_{0:k}))\) goes to \(-\infty\) due to the fact that differential entropy of an impulse is \(-\infty\). However, in a recursive algorithm, if we use some other densities which are obtained in the algorithms’ operation, it can be shown that a relatively good approximation for \(\mathcal{H}(p(x_k|y_{0:k}))\) can be obtained.

We consider the following nonlinear state-space representation.

\[
x_{k+1} = f(x_k) + w_k
\]

(2)

\[
y_k = h(x_k) + v_k
\]

(3)

where \(w_k \sim p_{w_k}(.)\) and \(v_k \sim p_{v_k}(.)\) are white process and measurement noise sequences respectively.

1 Particle Filter

We now suppose that we are at an intermediate stage of the estimation process and we have the approximation of the density \(p(x_{k-1}|y_{0:k-1})\) as

\[
p(x_{k-1}|y_{0:k-1}) \approx \sum_{i=1}^{N} \pi_{k-1|i-1}^{(i)} \delta(x_{k-1} - x_{k-1|i-1}^{(i)})
\]

(4)

In other words, we have \(\{x_{k-1|i-1}^{(i)}, \pi_{k-1|i-1}^{(i)}\}_i=1^N\).

1.1 Prediction Update

By sampling from an importance density \(\mu(x_k|x_{k-1}, y_k)\), we can approximate the predicted density \(p(x_k|y_{0:k-1})\) as

\[
p(x_k|y_{0:k-1}) \approx \sum_{i=1}^{N} \pi_{k|i}^{(i)} \delta(x_k - x_{k|i}^{(i)})
\]

(5)

where

\[
x_{k|i}^{(i)} \sim \mu(., x_{k-1|i}^{(i)}, y_k)
\]

(6)

\[
\pi_{k|i}^{(i)} \propto \frac{p(x_{k|i}^{(i)}|x_{k-1|i}^{(i)}, y_k)}{\mu(x_{k|i}^{(i)}|x_{k-1|i}^{(i)}, y_k)} \pi_{k-1|i}^{(i)}
\]

(7)

The important point here is that, \(p(x_k|y_{0:k-1})\) in (5) is an approximation of the density

\[
p(x_k|y_{0:k-1}) = \sum_{i=1}^{N} \pi_{k|i}^{(i)} \mu(x_k|x_{k-1|i}^{(i)}, y_k)
\]

(8)

which will play a crucial role in the calculation of \(\mathcal{H}(p(x_k|y_{0:k}))\) below.
1.2 Measurement Update
Using the likelihoods $p(y_k|x_k)$, we can obtain the measurement updated density
$p(x_k|y_{0:k})$ as

$$p(x_k|y_{0:k}) \approx \sum_{i=1}^{N} \pi_{k|k}^{(i)} \delta(x_k - x_{k|k}^{(i)}) \tag{9}$$

where

$$x_{k|k}^{(i)} = x_{k|k-1}^{(i)} \tag{10}$$
$$\pi_{k|k}^{(i)} \propto p(y_k|x_{k|k}^{(i)}) \pi_{k|k-1}^{(i)} \tag{11}$$

Here, optionally, a resampling procedure can be run.

1.3 Calculation of $\mathcal{H}(p(x_k|y_{0:k}))$

The differential entropy, by definition, is given as

$$\mathcal{H}(p(x_k|y_{0:k})) = - \int p(x_k|y_{0:k}) \log(p(x_k|y_{0:k})) dx_k \tag{12}$$

The Bayes rule tells us that

$$p(x_k|y_{0:k}) = \frac{p(y_k|x_k)}{p(y_k|y_{0:k-1})} p(x_k|y_{0:k-1}) \tag{13}$$

If we substitute (13) into the log term in (12), we get

$$\mathcal{H}(p(x_k|y_{0:k})) = - \int p(x_k|y_{0:k}) \left[ \log(p(y_k|x_k)) + \log(p(x_k|y_{0:k-1})) \right] dx_k$$

$$- \log(p(y_k|y_{0:k-1})) dx_k \tag{14}$$

$$= - \int p(x_k|y_{0:k}) \left[ \log(p(y_k|x_k)) + \log(p(x_k|y_{0:k-1})) \right] dx_k$$

$$+ \log(p(y_k|y_{0:k-1})) \tag{15}$$

Now, substituting the analytical expression (8) for $p(x_k|y_{0:k-1})$ into (15), we obtain

$$\mathcal{H}(p(x_k|y_{0:k})) = - \int p(x_k|y_{0:k}) \left[ \log(p(y_k|x_k)) \right.$$}

$$+ \log \left( \sum_{i=1}^{N} \pi_{k|k-1}^{(i)} \delta(x_k - x_{k|k-1}^{(i)}), y_k \right) dx_k$$

$$+ \log(p(y_k|y_{0:k-1})) \tag{16}$$
We can, now, use the particle approximation of \( p(x_k|y_{0:k}) \) in (9) to evaluate the integral in (16) as follows,

\[
\mathcal{H}(p(x_k|y_{0:k})) = -\sum_{j=1}^{N} \pi_{k|k}^{(j)} \left[ \log(p(y_k|x_{k|k})) \\
+ \log \left( \sum_{i=1}^{N} \pi_{k|k-1}^{(i)} \mu(x_{k|k}^{(i)}|x_{k-1|k-1}^{(i)}, y_k) \right) \right] \\
+ \log(p(y_k|y_{0:k-1}))
\]  

(17)

Now the only thing remaining before the full calculation of \( \mathcal{H}(p(x_k|y_{0:k})) \) is the evaluation of the constant \( p(y_k|y_{0:k-1}) \). But, by the total probability theorem, we have

\[
p(y_k|y_{0:k-1}) = \int p(y_k|x_k)p(x_k|y_{0:k-1})dx_k
\]  

(18)

Now using the particle approximation for \( p(x_k|y_{0:k-1}) \) in (5), we can calculate \( p(y_k|y_{0:k-1}) \) as

\[
p(y_k|y_{0:k-1}) = \sum_{i=1}^{N} \pi_{k|k-1}^{(i)} p(y_k|x_{k|k-1}^{(i)})
\]  

(19)

When we substitute this result into (17), we obtain the final formula for \( \mathcal{H}(p(x_k|y_{0:k})) \) as

\[
\mathcal{H}(p(x_k|y_{0:k})) = -\sum_{j=1}^{N} \pi_{k|k}^{(j)} \left[ \log(p(y_k|x_{k|k})) \\
+ \log \left( \sum_{i=1}^{N} \pi_{k|k-1}^{(i)} \mu(x_{k|k}^{(i)}|x_{k-1|k-1}^{(i)}, y_k) \right) \right] \\
+ \log \left( \sum_{i=1}^{N} \pi_{k|k-1}^{(i)} p(y_k|x_{k|k-1}^{(i)}) \right)
\]  

(20)

2 Example

In this section, we just compare the exact entropy of a Gaussian mixture with the particle based approximation derived above. We consider the linear system

\[
x_{k+1} = 2x_k + w_k
\]

\[
y_k = x_k + v_k
\]

where the variance of \( w_k \) and \( v_k \) are 5000 and 30000 respectively. The initial density \( p(x_0) \) is given as in Figure 1. We get a measurement at \( k = 1 \) as \( y_1 = 300 \). We can calculate the exact prediction and measurement updated mixtures using Kalman filter equations on each of the components as in Figure 2. We can calculate the theoretical differential entropy of \( p(x_1|y_1) \) as 6.5203 using numerical integration.
Figure 1: Initial Gaussian Mixture.

Figure 2: Initial Gaussian mixture \( p(x_0) \), predicted Gaussian mixture \( p(x_1) \), and measurement updated Gaussian mixture \( p(x_1|y_1) \).
Figure 3: Histogram of the initial Gaussian mixture \( p(x_0) \), predicted Gaussian mixture \( p(x_1) \), and measurement updated Gaussian mixture \( p(x_1|y_1) \).

We now represent initial Gaussian mixture \( p(x_0) \) with a number of particles and using particle filtering equations with \( \mu(x_k|x_{k-1}, y_k) = p(x_k|x_{k-1}) \) we obtain an approximation of \( p(x_1|y_1) \). Histograms of particles for \( p(x_0) \), \( p(x_1) \) and \( p(x_1|y_1) \) are shown in Figure 3 for number of particles \( N_p = 10000 \). We now calculate the differential entropy of \( p(x_1|y_1) \) using (20) with different number of particles. We then compare the resulting values to the analytical result 6.5203. The results are shown in Figure 4. Note that since the analytical entropy is calculated using the optimal result (Kalman filter), the errors in the figure contain also the errors caused by the particle representation. With number of particles greater than 500 almost exact result can be obtained.
Figure 4: Comparison of analytical entropy with the particle based calculation with different number of particles.