1 Introduction

Everywhere we encounter change in processes. To estimate this change and to reveal the underlying dynamics of some of these processes is the aim of this thesis. Knowledge about changes in processes is useful in a wide range of real world application. They arise from many different fields like engineering, meteorology, biology to name a few:

Engineering To ensure the quality of parts in running systems, the parts are monitored with sensors. They might monitor reading like temperature, current voltage or movement. If some of these readings change from the regular fluctuation, it might indicate a quality drop of certain parts. This might even lead to the break down of machines or production-lines.

Biology Genetic disorders and cancer may result in chromosomal DNA copy number changes. Some forms of cancer is probably caused by mutations in oncogenes and tumor suppressor genes. Identification of these genes is important for the development of medical diagnosis and treatment for these forms of cancer. The change in copy number can be compared to a baseline with aCGH experiments. The data resulting from aCGH experiments suffer from high noise. In several cases statistical change point analysis was used to identify DNA copy number changes[**?**][**?**][**?**].

Meteorology Weather events oft occur related to cyclic events like El Niño or the Pacific decadal oscillation. The later system is strongly connected to the cyclone activity in the North Pacific. Gaining knowledge about the temporal changes in climate data helps to predict cycles with increased bad weather events[**?**].

While for some of the applications a retrospective analysis might be sufficient, others require close to real time analysis. This concept is here called online change point detection. A change in the studied process should be identified with no knowledge about the future of the process. Also the sensitivity to changes should be managed in a way that outlier are not identified as change points.

1. example of a change point

2. what has been done. refer to the change point paper. state that there is this idea to improve on this paper by keeping information on the already seen states. The belief in a known state might increase faster than for a unknown state. so the belief in a change point can increase faster.

3. what is the conclusion

4. how is the text structured

TO CITE: Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance

2 Related Work/Background

This chapter will present the theoretical background for the following work. First the concept of time series and change points will be introduced briefly. Then an approach to detect change points in an online fashion with a Bayes framework is introduced.

2.1 Time Series and Change Points

Intuitively time series are data collections over time. For this thesis time series are defined as the sequence $x_1, x_2, ..., x_t$ of data ordered in equally spaced time intervals. In applications the requirement of equally spaced time intervals may be relaxed. Time series are a stochastic model and therefor a stochastic process yields the data.

One relevant attribute of time series is stationarity. A time series is stationary if the expected value or other moments is invariant to the time index. In other words a stationary time series does not change its properties over time. In time series analysis several methods are used to create a stationary time series from a non-stationary time series[?].

Change Points are the points in a time series where the generating stochastic process changes. They split a time series into non-overlapping sequences. For the frequentist approach the change point problem is to test the following null hypothesis

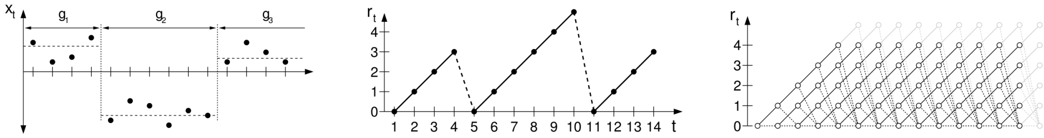
$$H_0 : z_1 = z_2 = ... = z_n \tag{2.1}$$

$$\begin{aligned} H_1 : z_1 = ...z_{k_1} \neq z_{k_1+1} = ... = z_{k_2} \neq z_{k_2+1} \\ = ... = z_{k_q} \neq z_{k_q+1} = ... = z_n \end{aligned} \tag{2.2}$$

with $1 < k_1 < k_2 < ... < k_q < n$ is a unknown number of change points and z_i is the probability distribution of the values x_i[?].

2.2 Bayesian Online detection of Change Points

While the previous approach is focused on retrospective analysis and segmentation, change points can also be detected in an online fashion with a Bayesian frame-



(a) A gull (b) A tiger (c) A mouse

Figure 2.1: Pictures of animals

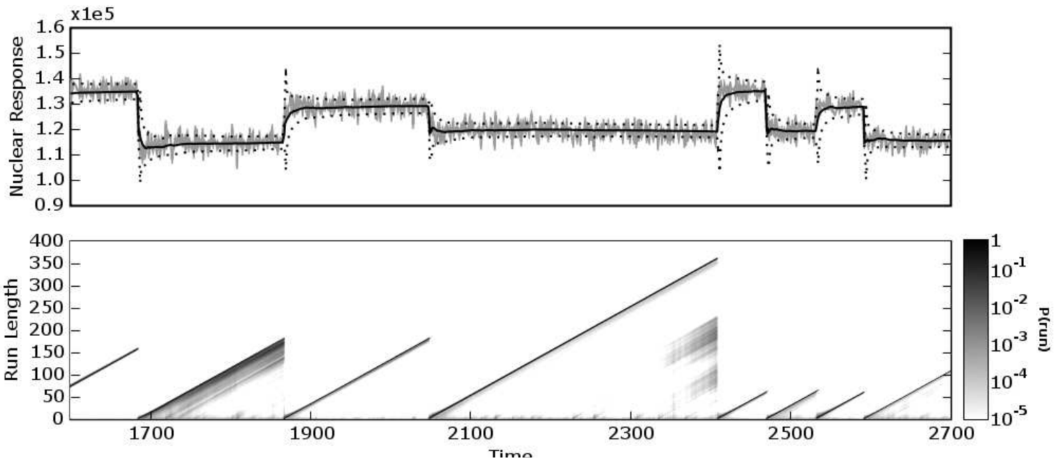


Figure 2.2: Caption

work. The work of Adams and MacKay focuses on causal predictive filtering. The posterior distribution over the run length since the last change point is estimated with the observed data.

The basic idea in their work is to model the change point detection as a decision problem. At each time step the run length since the last change point can either grow by one, or the current data point is a change point and the run length drops to zero. Each run length at time t has an probability assigned to it. The structure this approach builds is called trellis, through message passing the algorithm is computationally efficient.

While this method works well for some applications it provides some pitfalls. Some of them motivated this very thesis. Form the practitioner point of view there are at least two hyperparameter to choose, one being the expected time between two change points, the other being a distribution of the data. Both play a big part in the success of detecting change points correctly. Implementation-wise the trellis structure poses a problem. The structure is growing exponentially. At some point past data will either congest the memory or has to be discarded. It is unclear when data can be discarded without loosing vital data for change point prediction.

–ARE THERE OTHER PAPERS? HAVE NOT FOUND THEM–

3 Algorithm

In this chapter a new approach to detect change points in a stationary time series is presented. It starts by stating assumption on the time series. Then the general theoretical approach is presented. After that some variations of algorithms derived from an analytic analysis are shown. To relax some approximations and assumptions particle filters are introduced. They are the state of art class of algorithm to deploy Bayesian filter where the filter equation becomes intractable(CITE).

The data of a sequence between two change points is assumed to be Independent and identically distributed random variables. Further more the total number of possible generative processes is finite. The transition from one generative process to another is a Markov process, saying its conditional probability given the previous process depends only on the direct predecessor generative process. The generative process is called state z_i from now on.

3.1 Bayesian Filter Approach

This section presents the basic approach of this thesis. With several assumptions and approximation an analytic solution is found which yields to some basic filtering algorithms.

The basic model for this approach is to have a finite set of states $z_1, z_2, ..., z_n$ with their parameters for a generative processes described by $\theta_1, \theta_2, ..., \theta_n$.

– EXPLAIN THIS A BIT MORE –

This leads to the filter equation:

$$P_{t+1}(z, \theta_z | x_{t+1}) = \alpha \cdot P(x_{t+1}|\theta) \cdot \sum_{z'} P(z|z') \cdot P(z', \theta_z | x_{1:t}) \tag{3.1}$$

Where z is the current state, θ_z are the parameters of the generative process of state z. x is an observation, meaning that x_{t+1} is the observation at time $t+1$ and $x_{1:t}$ are all observations until time t. $P_{t+1}(z, \theta_z|x_{t+1})$ describes the probability of being in state z given the newly observed data x_{t+1}.

As a general assumption the number of states z_i and the transition model $P(z|z')$ is known.

To first derive an analytic solution, several approximation will be done. First the generative process is to be assumed as Gaussian with a variance of $\sigma^2 = 1$. This yields

$$\theta_i = \mu_i \tag{3.2}$$

where μ_i is the mean of a Gaussian.

The next approximation is $P(z,\theta) = P(z)P(\theta)$. With Integration over θ this leads to:

$$P(z|x_{t+1}) = \alpha \cdot \sum_{z'} P(z|z')P_t(z') \cdot \int P(x_{t+1}|\theta_z)P(\theta_z|x_{1:t})d\theta_z \tag{3.3}$$

$$= \alpha \cdot \sum_{z'} P(z|z')P_t(z') \cdot \int P(x_{t+1}|\theta_z)P(\theta_z|x_{1:t})d\theta_z \tag{3.4}$$

Which yields the probability of being in state z given the newly observed data x_{t+1}. Assuming the prior $P(x_i|\theta_i)$ and the likelihood $P(\theta_i|x_i)$ are Gaussian respectively makes the integral $\int P(\theta_2|x_{1:t})P(x_{t+1}|\theta_2)d\theta_2$ analytically traceable. For $\int P(x_{t+1}|\theta_z)P(\theta_z|x_{1:t})d\theta_z = \int \mathcal{N}(x_{t+1}|\mu_\theta, 1)\mathcal{N}(\mu_\theta|\mu_{1:t}, \sigma_{1:t})d\mu_\theta$:

$$\int \mathcal{N}(x_{t+1}|\mu_\theta, 1)\mathcal{N}(\mu_\theta|\mu_{1:t}, \sigma_{1:t})d\mu_\theta =$$
$$\frac{\sqrt{\frac{1}{1+\frac{1}{\sigma_{1:t}^2}}}}{\sqrt{2\pi}\sqrt{\sigma_{1:t}^2}} \exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma_{1:t}^2}}}\left(-\left(\frac{x+\frac{\mu_{1:t}}{\sigma_{1:t}^2}}{1+\frac{1}{\sigma_{1:t}^2}}\right)^2 + \frac{x^2}{1+\frac{1}{\sigma_{1:t}^2}} + \frac{\frac{\mu_{1:t}^2}{\sigma_{1:t}^2}}{1+\frac{1}{\sigma_{1:t}^2}}\right)\right) \tag{3.5}$$

The derivation of the integral is presented in the appendix at page 20. The probability of being in state z_i is accordingly:

$$P(z_i|x_{t+1}) = \alpha \cdot \sum_{z'} P(z_i|z')P_t(z') \cdot \frac{\sqrt{\frac{1}{1+\frac{1}{\sigma_{1:t}^2}}}}{\sqrt{2\pi}\sqrt{\sigma_{1:t}^2}}$$
$$\exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma_{1:t}^2}}}\left(-\left(\frac{x+\frac{\mu_{1:t}}{\sigma_{1:t}^2}}{1+\frac{1}{\sigma_{1:t}^2}}\right)^2 + \frac{x^2}{1+\frac{1}{\sigma_{1:t}^2}} + \frac{\frac{\mu_{1:t}^2}{\sigma_{1:t}^2}}{1+\frac{1}{\sigma_{1:t}^2}}\right)\right) \tag{3.6}$$

To get an approximation of θ it is assumed that the θ_i are independent, $P(\theta) = P(\theta_1)(\theta_2) \cdot ...(\theta_n)$. For the case of $n = 2$ the summation over z yields:

$$\begin{aligned}
P(\theta_1|x_{t+1}) &= \alpha \cdot P(\theta_1|x_{1:t}) \cdot \sum_{z} P(x_{t+1}|\theta_1) \sum_{z'} P(z|z')P_t(z') \qquad (3.7)\\
&= \alpha \cdot P(\theta_1|x_{1:t}) \cdot \Bigg(\sum_{z'} P(z_1|z')P_t(z')P(x_{t+1}|\theta_1) + \\
&\quad \sum_{z'} P(z_2|z')P_t(z') \int P(\theta_2|x_{1:t})P(x_{t+1}|\theta_2)\, d\theta_2 \Bigg) \qquad (3.8)\\
&= \alpha \cdot P(\theta_1|x_{1:t})P(x_{t+1}|\theta_1) \sum_{z'} P(z_1|z')P_t(z') + P(\theta_1|x_{1:t}) \\
&\quad \sum_{z'} P(z_2|z')P_t(z') \int P(\theta_2|x_{1:t})P(x_{t+1}|\theta_2)\, d\theta_2 \qquad (3.9)
\end{aligned}$$

Since every expression not containing θ_1 is merely a constant:

$$\begin{aligned}
P(\theta_1|x_{t+1}) &= \alpha \cdot P(\theta_1|x_{1:t})P(x_{t+1}|\theta_1) \sum_{z'} P(z_1|z')P_t(z') + P(\theta_1|x_{1:t}) \\
&\quad \sum_{z'} P(z_2|z')P_t(z') \int P(\theta_2|x_{1:t})P(x_{t+1}|\theta_2)\, d\theta_2 \qquad (3.10)\\
&= \alpha \cdot \beta P(\theta_1|x_{1:t})P(x_{t+1}|\theta_1) + \gamma P(\theta_1|x_{1:t}) \qquad (3.11)
\end{aligned}$$

Since $\beta >> \gamma$ if $P(\theta_1)$ is higher then $P(\theta_2)$

$$P(\theta_1|x_{t+1}) \approx \alpha \cdot \beta P(\theta_1|x_{1:t})P(x_{t+1}|\theta_1) \qquad (3.12)$$

Still assuming $P(\theta_1|x_{1:t}) = \mathcal{N}(x_{t+1}|\mu_\theta, 1)$, $P(x_{t+1}|\theta_1) = \mathcal{N}(\theta|\mu_{x_{1:t}}, \sigma^2_{x_{1:t}})$.

$$P(\theta_1|x_{t+1}) = \alpha \cdot \beta \mathcal{N}(x_{t+1}|\mu_\theta, 1)\mathcal{N}(\theta|\mu_{x_{1:t}}, \sigma^2_{x_{1:t}}) \qquad (3.13)$$

This can be used to derive an analytic update to θ_i:

$$P(\theta|x_{1:t+1}) = \alpha \cdot \mathcal{N}(x_{t+1}|\mu_\theta, 1)\mathcal{N}(\theta|\mu_{x_{1:t}}, \sigma^2_{x_{1:t}}) \tag{3.14}$$

$$= \alpha \cdot \exp\left(-\frac{1}{2}(x_{t+1} - \mu_\theta)^2 - \frac{1}{2\sigma^2_{x_{1:t}}}(\mu_\theta - \mu_{x_{1:t}})^2 \right) \tag{3.15}$$

$$= \alpha \cdot \exp\left(-\frac{1}{2}\left(x^2_{t+1} - 2x_{t+1}\mu_\theta + \mu^2_\theta + \frac{1}{\sigma^2_{x_{1:t}}}(\mu^2_\theta - 2\mu_\theta\mu_{1:t} + \mu^2_{1:t}) \right) \right) \tag{3.16}$$

$$= \alpha \cdot \exp\left(-\frac{1}{2}\left((1 + \frac{1}{\sigma^2_{1:t}})\mu^2_\theta - 2\mu_\theta(x_{t+1} + \frac{\mu_{1:t}}{\sigma^2_{1:t}}) + x^2_{t+1} + \frac{\mu^2_{1:t}}{\sigma^4_{1:t}} \right) \right) \tag{3.17}$$

Which leads to an update for the posterior Gaussian of:

$$\frac{1}{\sigma^2_{t+1}} = 1 + \frac{1}{\sigma^2_{1:t}} \tag{3.18}$$

$$\mu_{t+1} = \frac{1}{\frac{1}{\sigma^2_{t+1}} + 1}\left(x_{t+1} + \frac{\mu_{1:t}}{\sigma^2_{1:t}} \right) \tag{3.19}$$

3.2 Gaussian Likelihood Filter

Another ansatz is to formulate the joint probability of being in state $z_{t+1} = 1, 2$ with θ_1, θ_2 given all data $P(z_{t+1}, \theta_1, \theta_2|x_{1:t+1})$. Where θ_1, θ_2 are just the mean for a Gaussian describing the approximated generative process of state z_1, z_2 respectively.

$$P(z_{t+1}, \theta_1, \theta_2|x_{1:t+1}) = \frac{P(x_{t+1}, z_{t+1}, \theta_1, \theta_2|x_{1:t})}{P(x_{t+1}|x_{1:t})} \tag{3.20}$$

$$= \alpha \cdot P(x_{t+1}, z_{t+1}, \theta_1, \theta_2|x_{1:t}) \tag{3.21}$$

$$= \alpha \cdot P(x_{t+1}|z_{t+1}, \theta_1, \theta_2) \cdot P(z_{t+1}, \theta_1, \theta_2|x_{1:t}) \tag{3.22}$$

$$= \alpha \cdot P(x_{t+1}|z_{t+1}, \theta_1, \theta_2) \cdot \sum_{z_t} P(z_{t+1}|z_t) \cdot P(z_t, \theta_1, \theta_2|x_{1:t}) \tag{3.23}$$

$$= \alpha \cdot P(x_{t+1}|z_{t+1}, \theta_1, \theta_2) \cdot \sum_{z_t} P(z_{t+1}|z_t) \cdot P(z_t|x_{1:t}) \cdot P(\theta_1|x_{1:t}) \cdot P(\theta_2|x_{1:t}) \tag{3.24}$$

This general formula can now be deployed in an iterative algorithm. Again assuming the Gaussian case where the likelihood $P(x_{t+1}|\theta_i) = \mathcal{N}(x_{t+1}|\theta_i, 1)$ and the prior , $P(x_{t+1}|\theta_1) = \mathcal{N}(\theta|\mu_{x_{1:t}}, \sigma^2_{x_{1:t}})$. For each iteration the following terms would need to be evaluated:

$$z_{t+1} = i : P(\theta_i|x_{1:t+1}) = \alpha \cdot P(x_{t+1}|\theta_i) \cdot P(\theta_i|x_{1:t}) \tag{3.25}$$
$$= \alpha \cdot \mathcal{N}(x_{t+1}|\theta_i, 1) \cdot \mathcal{N}(\theta_i|\mu_{1:t}, \sigma^2_{1:t}) \tag{3.26}$$

The update of θ_i would also be:

$$\frac{1}{\sigma^2_{t+1}} = 1 + \frac{1}{\sigma^2_{1:t}} \tag{3.27}$$
$$\mu_{t+1} = \frac{1}{\frac{1}{\sigma^2_{t+1}} + 1} \left(x_{t+1} + \frac{\mu_{1:t}}{\sigma^2_{1:t}} \right) \tag{3.28}$$

3.3 Analytic Algorithms

The two approaches presented in the two previous sections differ in assigning a probability to the current state z_i. Otherwise the structure is similar. First they compute the probability of each state z_i, then the update of $theta_i$ is performed. There are two ways to apply the update. Either the algorithm follows a Maximum a Posteriori estimation or a Bayes estimation. In the following the four different algorithms are presented. The evaluation of the these algorithms can be found in the following chapter.

The first algorithm(Algorithm 1) is a Gaussian Filter with Maximum a Posteriori estimation. For each new data point it first computes the probability of each state z_i. The probability for a state z_i is computed as follows:

$$P(\tilde{z}_i|x_{t+1}) = \mathcal{N}(x_{t+1}|\theta_i, 1) \cdot \sum_j P(z_i|z_j)P(z_j|x_{1:t}) \tag{3.29}$$
$$P(z_i|x_{t+1}) = \frac{z_i}{\sum_z {}_j z_j} \tag{3.30}$$

Then it updates the the state z_{max}, i.e the most probable state $argmax_i\ P(z_i|x_{t+1})$. To update the estimated mean θ_{max} of the current state z_{max} equation 3.28 is used.

The following three algorithms differ only in small but relevant changes to the computation of z_i, as mentioned above, or the estimator.

The second algorithm(Algorithm 2) differs to Algorithm 1 by the estimator, i.e the update for θ_i. While the previous algorithm only updated the most probable state z_{max}, this algorithm updates all states weighted on the probability of each

Algorithm 1 Gaussian Filter with MAP Estimator

while new data **do**
  $x_{t+1} \leftarrow newdata$
  for all states **do**
    $likelihoodState_i \leftarrow normal(x_{t+1}|\theta_i, 1)$
    $\tilde{z}_i \leftarrow likelihoodState_i \cdot \sum_j P(z_i|z_j)P(z_j|x_{1:t})$
  for all states **do**
    $z_i \leftarrow z_i / \sum_j z_j$
  $i \leftarrow \max(z_1, z_2, ..., z_n)$
  $\sigma_i \leftarrow \frac{1}{1+\frac{1}{\sigma^2_{1:t}}}$
  $\theta_i \leftarrow \frac{1}{\frac{1}{\sigma^2_{t+1}}+1}\left(x_{t+1} + \frac{\mu_{1:t}}{\sigma^2_{1:t}}\right)$

state. The update of θ_i is the following

$$\theta_i = \frac{1}{\frac{1}{\sigma^2_{t+1}} + w_i}\left(w_i \cdot x_{t+1} + \frac{\mu_{1:t}}{\sigma^2_{1:t}}\right) \tag{3.31}$$

Where w_i is the weight of the new data point x_{t+1}. The weight is equivalent to the probability of being in state z_i given the previous observations and the new data point x_{t+1}, $w_i = z_i$.

The two presented algorithms differ in the update of θ_i. The following two algorithms will have the same difference in the update of θ_i, but compute z_i different then before:

$$\begin{aligned} P(z_i|x_{t+1}) = \alpha \cdot \sum_{z'} P(z_i|z')P_t(z') \cdot \frac{\sqrt{\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}}{\sqrt{2\pi}\sqrt{\sigma^2_{1:t}}} \\ \exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(-\left(\frac{x+\frac{\mu_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)^2 + \frac{x^2}{1+\frac{1}{\sigma^2_{1:t}}} + \frac{\frac{\mu^2_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)\right) \end{aligned} \tag{3.32}$$

Algorithm 2 Gaussian Filter with Bayes Estimator

while new data **do**
  $x_{t+1} \leftarrow newdata$
  for all states **do**
    $likelihoodState_i \leftarrow normal(x_{t+1}|\theta_i, 1)$
    $\tilde{z}_i \leftarrow likelihoodState_i \cdot \sum_j P(z_i|z_j)P(z_j|x_{1:t})$
  for all states **do**
    $z_i \leftarrow z_i / \sum_j z_j$
    $\sigma_i \leftarrow \frac{1}{1+\frac{1}{\sigma^2_{1:t}}}$
    $\theta_i \leftarrow \frac{1}{\frac{1}{\sigma^2_{t+1}}+z_i} \left(z_i \cdot x_{t+1} + \frac{\mu_{1:t}}{\sigma^2_{1:t}} \right)$

3.4 Particle Filter

In this section the basics of the vast field of particle filter or sequential Monte Carlo approaches are presented. Particle Filter are used to approximate the posterior probabilities of a Bayes Filter. They are known under many names and have been used in a wide variety of scientific fields and their applications.

Particle Filter estimate the posterior by Monte Carlo methods. Monte Carlo methods generally integrate any given function by drawing a large number of samples from the function of interest. Assuming it is possible to draw N independent samples $p^{(i)}_{0:t}; i = 1, 2, ..., N$ according to the posterior probability $P(z_{0:t}, \theta_{0:t}|x_{1:t})$. An estimate of the posterior distribution is

$$P_N(z_{0:t}, \theta_{0:t}|x_{1:t}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{p^{(i)}_{0:t}}(z_{0:t}, \theta_{0:t}) \tag{3.33}$$

where $\delta_{p^{(i)}_{0:t}}(z, \theta)$ is the delta-Dirac mass in $p^{(i)}_{0:t}$. This yields an estimation for the computation of functions of interest for z, θ

$$I(f_t) = E_{P(z_{0:t},\theta_{0:t}|x_{1:t})}[f_t(z_{0:t}, \theta_{0:t})] \tag{3.34}$$

$$= \int f_t(z_{0:t}, \theta_{0:t}) P_N(z_{0:t}, \theta_{0:t}|x_{1:t}) dz, \theta \tag{3.35}$$

$$= \frac{1}{N} \sum_{i=1}^{N} f(p^{(i)}_{0:t}) \tag{3.36}$$

An interesting function f_t might be the mean with $f_t(p^{(i)}_{0:t}) = p^{(i)}_{0:t}$.

Algorithm 3 Gaussian Filter with MAP Estimator Opper

while new data **do**
  $x_{t+1} \leftarrow newdata$
  for all states **do**
    $\tilde{z}_i \leftarrow \int \mathcal{N}(x_{t+1}|\mu_\theta, 1)\mathcal{N}(\mu_\theta|\mu_{1:t}, \sigma_{1:t})d\mu_\theta \cdot \sum_j P(z_i|z_j)P(z_j|x_{1:t})$
  for all states **do**
    $z_i \leftarrow z_i / \sum_j z_j$
  $i \leftarrow$max(z$_{1,z}$ $2, ..., z_n)$
  $\sigma_i \leftarrow \frac{1}{1+\frac{1}{\sigma^2_{1:t}}}$
  $\theta_i \leftarrow \frac{1}{\frac{1}{\sigma^2_{t+1}}+1}\left(x_{t+1} + \frac{\mu_{1:t}}{\sigma^2_{1:t}}\right)$

Often it is not possible to gain samples directly from $P(z_{0:t}, \theta_{0:t}|x_{1:t})$, bur from a distribution $\pi(z_{0:t}, \theta_{0:t}|x_{1:t})$ with larger support than $P(z_{0:t}, \theta_{0:t}|x_{1:t})$. This function is called importance distribution. Weighted samples drawn from the importance distribution can be used to estimate $I(f_t)$:

$$I(f_t) = \frac{\int f_t(z_{0:t}, \theta_{0:t})w(z_{0:t}, \theta_{0:t})\pi(z_{0:t}, \theta_{0:t}|x_{1:t})dz_{0:t}, \theta_{0:t}}{\int w(z_{0:t}, \theta_{0:t})\pi(z_{0:t}, \theta_{0:t}|x_{1:t})dz_{0:t}, \theta_{0:t}} \tag{3.37}$$

with $w(z_{0:t}, \theta_{0:t})$ as the importance weight

$$w(z_{0:t}, \theta_{0:t}) = \frac{P(z_{0:t}, \theta_{0:t}|x_{1:t})}{\pi(z_{0:t}, \theta_{0:t}|x_{1:t})} \tag{3.38}$$

Now drawing N i.i.d. samples $p^{(i)}_{0:t}; i = 1, 2, ..., N$ from $\pi(z, \theta|x)$ yields a Monte Carlo estimate of $I(f_t)$

$$I(f_t) = \frac{\frac{1}{N}\sum_{i=1}^N f(p^{(i)}_{0:t})w(p^{(i)}_{0:t})}{\frac{1}{N}\sum_{i=1}^N w(p^{(i)}_{0:t}} \tag{3.39}$$

$$= \sum_{i=1}^N f(p^{(i)}_{0:t})\tilde{w}^{(i)}_t \tag{3.40}$$

where $\tilde{w}^{(i)}_t$ is the normalized important weight

$$\tilde{w}^{(i)}_t = \frac{w(p^{(i)}_{0:t})}{\sum_{i=1}^N w(p^{(i)}_{0:t}} \tag{3.41}$$

Algorithm 4 Gaussian Filter with Bayes Estimator Opper

while new data **do**
  $x_{t+1} \leftarrow newdata$
  for all states **do**
    $\tilde{z}_i \leftarrow \int \mathcal{N}(x_{t+1}|\mu_\theta, 1)\mathcal{N}(\mu_\theta|\mu_{1:t}, \sigma_{1:t})d\mu_\theta \cdot \sum_j P(z_i|z_j)P(z_j|x_{1:t})$
  for all states **do**
    $z_i \leftarrow z_i / \sum_j z_j$
    $\sigma_i \leftarrow \frac{1}{1+\frac{1}{\sigma^2_{1:t}}}$
    $\theta_i \leftarrow \frac{1}{\frac{1}{\sigma^2_{t+1}}+z_i}\left(z_i \cdot x_{t+1} + \frac{\mu_{1:t}}{\sigma^2_{1:t}}\right)$

This algorithm is called *important sampling*. The algorithm can also be interpreted as estimating the posterior distribution $P(z, \theta|x)$. Using 3.40 in 3.35 and 3.33 yields

$$\hat{P}_N(z_{0:t}, \theta_{0:t}|x_{1:t}) = \sum_{i=1}^{N} \tilde{w}_t^{(i)} \delta_{p_{0:t}^{(i)}}(z_{0:t}, \theta_{0:t}) \tag{3.42}$$

which is an estimate of the posterior distribution using the weighted samples $p^{(i)}_{0:t}$. Importance sampling is a general Monte Carlo integration method, but in this form not applicable to an online filtering problem. To compute the estimate $\hat{P}_N(z_{0:t}, \theta_{0:t}|x_{1:t})$ the algorithm needs to get all data $x_{1:t}$. Each time a new datum x_{x+1} becomes available, the methods needs to recompute the important weights of the entire sequence $x_{1:t+1}$.

The algorithm can be altered to fit the demands of online estimation. *Sequential Importance Sampling* uses an importance function $\pi(z_{0:t}, \theta_{0:t}|x_{1:t}$ which at time t is based on the previous $\pi(z_{0:t-1}, \theta_{0:t-1}|x_{1:t-1})$

$$\pi(z_{0:t}, \theta_{0:t}|x_{1:t}) = \pi(z_{0:t-1}, \theta_{0:t-1}|x_{1:t-1})\pi(z_{0:t}, \theta_{0:t}|z_{0:t-1}, \theta_{0:t-1}, x_{1:t}) \tag{3.43}$$

Which leads to

$$\pi(z_{0:t}, \theta_{0:t}|x_{1:t}) = \pi(z_0, \theta_0) \prod_{k=1}^{t} \pi(z_{0:k}, \theta_{0:k}|z_{0:k-1}, \theta_{0:k-1}, x_{1:k}) \tag{3.44}$$

This important function allows to compute the importance weights recursively

$$\tilde{w}_t^{(i)} \propto \tilde{w}_{t-1}^{(i)} \frac{P(x_t|p_t^{(i)})P(p_t^{(i)}|p_{t-1}^{(i)})}{\pi(p_t^{(i)}|p_{0:t-1}^{(i)}, x_{1:t})} \tag{3.45}$$

An important special case of importance distribution is to chose the prior as importance distribution. This choice leads to

$$\pi(z_{0:t}, \theta_{0:t}|x_{1:t}) = P(z_{0:t}, \theta_{0:t}) \tag{3.46}$$

$$= P(z_0, \theta_0) \prod_{k=1}^{t} P(z_k|z_{k-1}) \tag{3.47}$$

where the importance weights are

$$\tilde{w}_t^{(i)} \propto \tilde{w}_{t-1}^{(i)} P(x_t|p_t^{(i)}) \tag{3.48}$$

sequence important sampling is not restricted to this choice of importance function, but in this work the choice of the prior is made.

While *sequence important sampling* works well for online filtering in theory. In practice with the increase of t the problem of degeneracy arises. The distribution of $\tilde{w}_t^{(i)}$ becomes more and more skewed until all but one sample has non-zero importance weight. Because of this the method does not approximate the posterior distribution well enough after some step t.

To elude degeneracy an extra step is introduced. When the samples start to degenerate, i.e. fail to approximate the posterior, new samples are drawn from the current approximation of the posterior. This idea is called resampling. This algorithm is know as *sequential importance resampling* or *particle filter*.There are many ideas on how to do the resampling step. Here one basic method called multinomial resampling is presented.

The new particle are drawn from the discrete distribution $\hat{P}_N(z_{0:t}, \theta_{0:t}|x_{1:t}) = \sum_{i=1}^{N} \tilde{w}_t^{(i)} \delta_{p_{0:t}^{(i)}}(z_{0:t}, \theta_{0:t})$. This is equivalent to sampling according to a multinomial distribution with $\tilde{w}_t^{(i)}$.

Algorithm 5 shows the pseudo code of a particle filter. Due to $P(p_t^{(i)}|p_{t-1}^{(i)})$ being a markov transition the path of the samples are not relevant and they are not necessary to keep in memory. Figure **??** illustrates one iteration of a *particle filter* with multinomial resampling. [**?**][**?**]

–WRITE AOUBT THE PARTICLE FILTER USDED IN THIS WORK–

– SOME NICE UEBERGANG –

3.5 Change Point Analysis with Filtering

Algorithm 5 Particle filter

Initialization: $t = 1$ For $i = 1, 2, ..., N$ sample $p_1^{(i)} \leftarrow P(x_0)$
while new data x_t **do**
  For $i = 1, 2, ..., N$ sample $\tilde{p_t}^{(i)} \leftarrow P(p_t|p_{t-1})$
  For $i = 1, 2, ..., N$ evaluate $\tilde{w}_t^{(i)} = P(\theta|\tilde{p_t}^{(i)})$
  Normalize the weights $\tilde{w}_t^{(i)}$
  $N_{eff} \leftarrow 1/\sum_{i=1}^{N}(w_t^{(i)})^2$
  if $N_{eff} < \frac{N}{2}$ **then**
    Resample $p_t^{(i)}$ from $\tilde{p_t}^{(i)}$ with the important weights $w_t^{(i)}$
    $t \leftarrow t + 1$
    a

    b

4 Evaluation

5 Conclusion

List of Acronyms

3GPP	3rd Generation Partnership Project
AJAX	Asynchronous JavaScript and XML
API	Application Programming Interface

Bibliography

Annex

.1 Analytic solution of Gaussian Integral

$$\int \mathcal{N}(x_{t+1}|\mu_\theta, 1)\mathcal{N}(\mu_\theta|\mu_{1:t}, \sigma_{1:t})d\mu_\theta \tag{.1}$$

$$= \int \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\mu_\theta)^2 \right) \cdot \frac{1}{\sqrt{2\pi\sigma_{1:t}^2}} \exp\left(-\frac{1}{2\sigma_{1:t}^2}(\mu_\theta - \mu_{1:t})^2 \right) d\mu_\theta \tag{.2}$$

$$= \int \frac{1}{\sqrt{2\pi}\sqrt{2\pi\sigma_{1:t}^2}} \exp\left(-\frac{1}{2}(x-\mu_\theta)^2 - \frac{1}{2\sigma_{1:t}^2}(\mu_\theta - \mu_{1:t})^2 \right) d\mu_\theta \tag{.3}$$

$\alpha = \frac{1}{\sqrt{2\pi}\sqrt{2\pi\sigma_{1:t}^2}}$

$$= \int \alpha \exp\left(-\frac{1}{2}\left((x-\mu_\theta)^2 - \frac{1}{\sigma_{1:t}^2}(\mu_\theta - \mu_{1:t})^2 \right) \right) d\mu_\theta \tag{.4}$$

$$= \int \alpha \exp\left(-\frac{1}{2}\left(x^2 - 2\mu_\theta x + \mu_\theta^2 + \frac{\mu_\theta^2}{\sigma_{1:t}^2} - 2\mu_\theta\frac{\mu_{1:t}}{\sigma_{1:t}^2} + \frac{\mu_{1:t}^2}{\sigma_{1:t}^2} \right) \right) d\mu_\theta \tag{.5}$$

$$= \int \alpha \exp\left(-\frac{1}{2}\left((1+\frac{1}{\sigma_{1:t}^2})\mu_\theta^2 - 2\mu_\theta(x + \frac{\mu_{1:t}}{\sigma_{1:t}^2}) + x^2 + \frac{\mu_{1:t}^2}{\sigma_{1:t}^2} \right) \right) d\mu_\theta \tag{.6}$$

$$= \int \alpha \exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma_{1:t}^2}}}\left(\mu_\theta^2 - 2\mu_\theta\frac{x + \frac{\mu_{1:t}}{\sigma_{1:t}^2}}{1+\frac{1}{\sigma_{1:t}^2}} + \frac{x^2}{1+\frac{1}{\sigma_{1:t}^2}} + \frac{\frac{\mu_{1:t}^2}{\sigma_{1:t}^2}}{1+\frac{1}{\sigma_{1:t}^2}} \right) \right) d\mu_\theta \tag{.7}$$

$\beta = \frac{x+\frac{\mu_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}$

$$= \int \alpha \exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(\mu_\theta^2 - 2\mu_\theta\beta + \beta^2 - \beta^2 + \frac{x^2}{1+\frac{1}{\sigma^2_{1:t}}} + \frac{\frac{\mu^2_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)\right)d\mu_\theta \tag{.8}$$

$$= \int \alpha \exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(\mu_\theta^2 - 2\mu_\theta\beta + \beta^2\right)\right)\cdot \tag{.9}$$

$$\exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(-\beta^2 + \frac{x^2}{1+\frac{1}{\sigma^2_{1:t}}} + \frac{\frac{\mu^2_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)\right)d\mu_\theta$$

$$= \int \alpha \exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(\mu_\theta - \beta\right)^2\right)\cdot \exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(-\beta^2 + \frac{x^2}{1+\frac{1}{\sigma^2_{1:t}}} + \frac{\frac{\mu^2_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)\right)d\mu_\theta \tag{.10}$$

$$= \alpha \exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(-\beta^2 + \frac{x^2}{1+\frac{1}{\sigma^2_{1:t}}} + \frac{\frac{\mu^2_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)\right)\int \exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(\mu_\theta - \beta\right)^2\right)d\mu_\theta \tag{.11}$$

$$= \alpha \exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(-\beta^2 + \frac{x^2}{1+\frac{1}{\sigma^2_{1:t}}} + \frac{\frac{\mu^2_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)\right)\sqrt{2\pi\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\cdot \tag{.12}$$

$$\int \frac{1}{\sqrt{2\pi\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}}\exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(\mu_\theta - \beta\right)^2\right)d\mu_\theta$$

$$= \frac{\sqrt{2\pi\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}}{\sqrt{2\pi}\sqrt{2\pi\sigma^2_{1:t}}}\exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(-\left(\frac{x+\frac{\mu_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)^2 + \frac{x^2}{1+\frac{1}{\sigma^2_{1:t}}} + \frac{\frac{\mu^2_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)\right) \tag{.13}$$

$$= \frac{\sqrt{\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}}{\sqrt{2\pi}\sqrt{\sigma^2_{1:t}}}\exp\left(-\frac{1}{2\frac{1}{1+\frac{1}{\sigma^2_{1:t}}}}\left(-\left(\frac{x+\frac{\mu_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)^2 + \frac{x^2}{1+\frac{1}{\sigma^2_{1:t}}} + \frac{\frac{\mu^2_{1:t}}{\sigma^2_{1:t}}}{1+\frac{1}{\sigma^2_{1:t}}}\right)\right) \tag{.14}$$