A Bayesian approach to change point analysis of discrete time series

Ali Mohammad-Djafari and Olivier Féron

Laboratoire des Signaux et Systèmes

CNRS-ESE-UPS

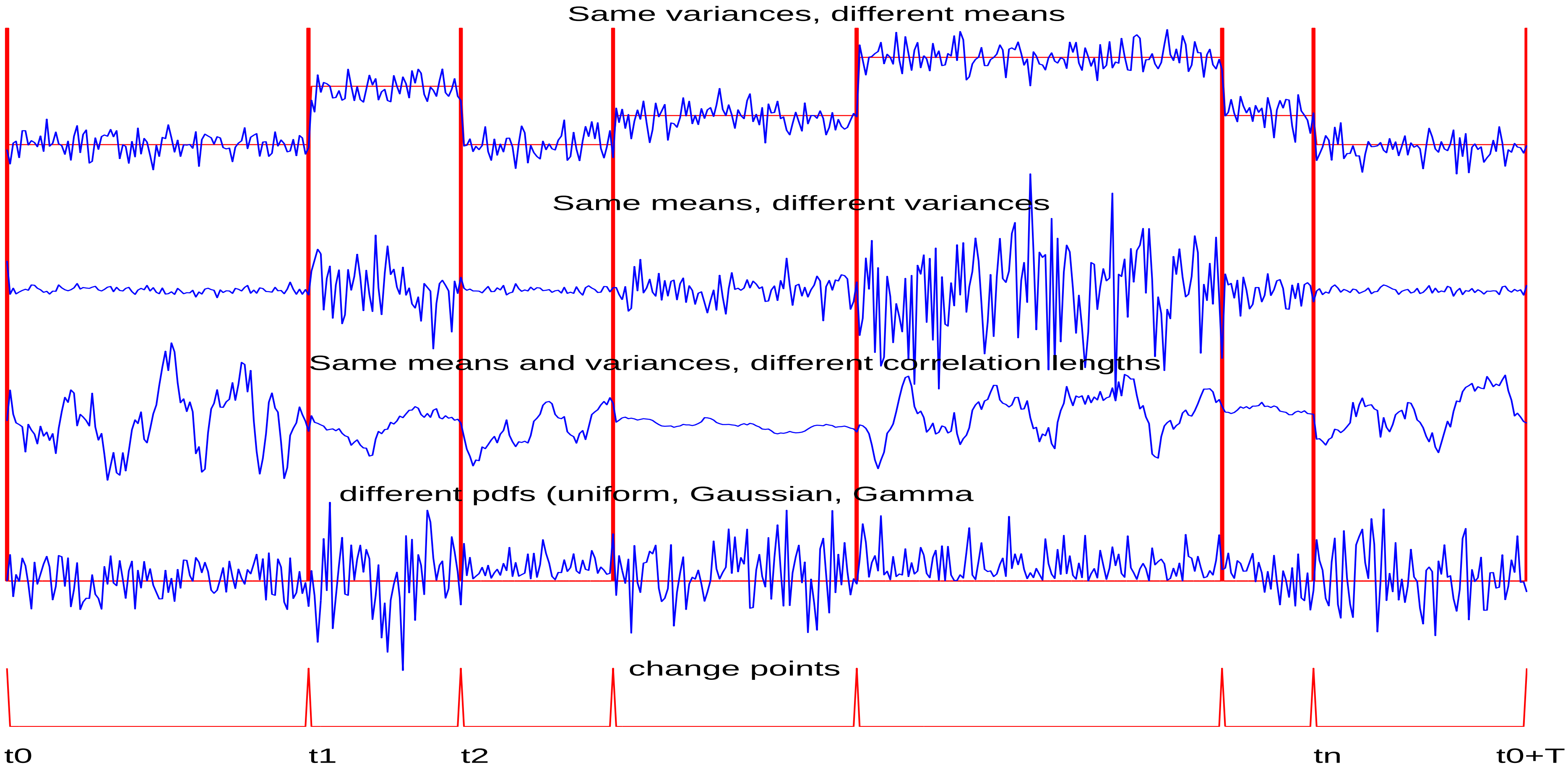
Supélec, Plateau de Moulon

91192 Gif-sur-Yvette, FRANCE.

{djafari,feron}@lss.supelec.fr

http://djafari.free.fr

1 Introduction



2 Notations and Hypothesis

$\boldsymbol{x} = [x(t_0), \cdots, x(t_0+T)]'$ observed samples

$\boldsymbol{t} = [t_1, \cdots, t_N]'$ change points instants

$\boldsymbol{x} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_N]'$ $\boldsymbol{x}_n = [x(t_n), x(t_n+1), \cdots, x(t_{n+1})]', \quad n = 0, \cdots, N$

$$
\begin{aligned}
p(x(t_n)) &= \mathcal{N}(\mu_n, \sigma_n^2) \\
p(x(t_n+l)|x(t_n+l-1)) &= \mathcal{N}(\rho_n\, x(t_n+l-1) + (1-\rho_n)\mu_n, \sigma_n^2(1-\rho_n^2)), \\
\text{with} & \quad l_n = t_{n+1} - t_n + 1 = \dim[\boldsymbol{x}_n], \quad l = 1, \cdots, l_n - 1
\end{aligned}
$$

$$
\begin{aligned}
p(\boldsymbol{x}_n) &= p(x(t_n)) \prod_{l=1}^{l_n} p(x(t_n+l)|x(t_n+l-1)) \\
&= \mathcal{N}(\mu_n \mathbf{1}, \Sigma_n) \ \text{ with } \ \boldsymbol{\Sigma}_n = \sigma_n^2 \,\text{Toeplitz}([1, \rho_n, \rho_n^2, \cdots, \rho_n^{l_n}])
\end{aligned}
$$

$$
p(\boldsymbol{x}|\boldsymbol{t}, \boldsymbol{\theta}, N) = \prod_{n=0}^{N} \mathcal{N}(\mu_n \mathbf{1}, \boldsymbol{\Sigma}_n) \propto \exp\left[-\frac{1}{2} \sum_{n=0}^{N} (\boldsymbol{x}_n - \mu_n \mathbf{1})' \boldsymbol{\Sigma}_n^{-1} (\boldsymbol{x}_n - \mu_n \mathbf{1})\right]
$$

where we noted $\boldsymbol{\theta} = \{\mu_n, \sigma_n, \rho_n, \ n = 0, \cdots, N\}$.

3 Bayesian Approach

- Likelihood:

$$p(\boldsymbol{x}|\boldsymbol{t},\boldsymbol{\theta},N) = \prod_{n=0}^{N} \mathcal{N}(\mu_n \mathbf{1}, \boldsymbol{\Sigma}_n) \propto \exp\left[-\frac{1}{2}\sum_{n=0}^{N}(\boldsymbol{x}_n - \mu_n\mathbf{1})'\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{x}_n - \mu_n\mathbf{1})\right]$$

- Priors:

$$t_n = t_{n-1} + \epsilon_n \quad \text{with} \quad \epsilon_n \sim \mathcal{P}(\lambda),$$

$$p(\boldsymbol{t}|\lambda, N) = \prod_{n=1}^{N+1} \mathcal{P}(t_n - t_{n-1}|\lambda) = \prod_{n=1}^{N+1} e^{-\lambda}\frac{\lambda^{(t_n - t_{n-1})}}{(t_n - t_{n-1})!}$$

$$p(\boldsymbol{\theta}|N) = \prod_{n} p(\theta_n) \quad \left\{\begin{array}{l} p(\mu_n) = \mathcal{N}(\mu_0, \sigma_0^2) \\ p(\sigma_n^2) = \mathcal{IG}(\alpha_0, \beta_0) \\ p(\rho_n) = \mathcal{U}([0,1]) \end{array}\right.$$

Conjugate priors

- Posterior:

$$p(\boldsymbol{t},\boldsymbol{\theta}|\boldsymbol{x},\lambda,N) \propto p(\boldsymbol{x}|\boldsymbol{t},\boldsymbol{\theta},N)\,p(\boldsymbol{t}|\lambda,N)\,p(\boldsymbol{\theta}|N)$$

- Inference: Estimate $\boldsymbol{t}$ and $\boldsymbol{\theta}$ using $p(\boldsymbol{t},\boldsymbol{\theta}|\boldsymbol{x},\lambda,N)$

  – MAP:

$$(\widehat{\boldsymbol{t}},\widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{t},\boldsymbol{\theta})}\{p(\boldsymbol{t},\boldsymbol{\theta}|\boldsymbol{x},\lambda,N)\}$$

  – MSE:

$$\{\widehat{\boldsymbol{t}},\widehat{\boldsymbol{\theta}}\} = \int \{\boldsymbol{t},\boldsymbol{\theta}\}\,p(\boldsymbol{t},\boldsymbol{\theta}|\boldsymbol{x},\lambda,N)\,\mathrm{d}\{\boldsymbol{t},\boldsymbol{\theta}\}$$

  MCMC Gibbs sampling:

  Generate samples $\{(\boldsymbol{t},\boldsymbol{\theta})^{(i)}\}$ from posterior and average to estimate the mean values

$$\begin{cases} \boldsymbol{t} & \sim & p(\boldsymbol{t}|\boldsymbol{x},\boldsymbol{\theta},\lambda,N) \\ \boldsymbol{\theta} & \sim & p(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{t},N) \end{cases}$$

MCMC Gibbs Sampling

- Initialize $\boldsymbol{\theta} = \{\mu_n = 0, \sigma_n = 1, \rho_n = 0,\ n = 0, \cdots, N\}$
- Iterate

. sample $\boldsymbol{t}$	using	$p(\boldsymbol{t}\|\boldsymbol{x}, \boldsymbol{\theta}, \lambda, N) \propto p(\boldsymbol{x}\|\boldsymbol{t}, \boldsymbol{\theta}, N)\, p(\boldsymbol{t}\|\lambda, N)$
. sample $\boldsymbol{\theta} = \{\mu_n, \sigma_n, \rho_n\}$	using	$p(\boldsymbol{\theta}\|\boldsymbol{x}, \boldsymbol{t}, N) \propto p(\boldsymbol{x}\|\boldsymbol{t}, \boldsymbol{\theta}, N)\, p(\boldsymbol{\theta}\|N)$
μ_n	using	$p(\mu_n\|\boldsymbol{x}, \boldsymbol{t}, N)$
σ_n^2	using	$p(\sigma_n^2\|\boldsymbol{x}, \boldsymbol{t}, N)$
ρ_n	using	$p(\rho_n\|\boldsymbol{x}, \boldsymbol{t}, N)$

$$p(\mu_n|\boldsymbol{x}, \boldsymbol{t}) = \mathcal{N}(\widehat{\mu}_n, \widehat{\sigma}_n^2) \text{ with } \begin{cases} \widehat{\mu}_n = \widehat{\sigma}_n^2 \left[\frac{\mu_0}{\sigma_0^2} + \mathbf{1}'\boldsymbol{\Sigma}_n^{-1}\boldsymbol{x}_n\right] \\ \widehat{\sigma}_n^2 = \left(\mathbf{1}'\boldsymbol{\Sigma}_n^{-1}\mathbf{1} + \frac{1}{\sigma_0^2}\right)^{-1} \end{cases}$$

$$p(\sigma_n^2|\boldsymbol{x}, \boldsymbol{t}) = \mathcal{IG}(\widehat{\alpha}_n, \widehat{\beta}_n) \text{ with } \begin{cases} \widehat{\alpha}_n = \alpha_0 + \frac{l_n}{2} \\ \widehat{\beta}_n = \beta_0 + \frac{1}{2}(\boldsymbol{x}_n - \mu_n\mathbf{1})\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{x}_n - \mu_n\mathbf{1}), \end{cases}$$

Sampling $p(\rho_n|\boldsymbol{x},\boldsymbol{t})$:

- $p(\rho_n|\boldsymbol{x},\boldsymbol{t})$ is not a classical law:

$$p(\rho_n|\boldsymbol{x},\boldsymbol{t},N) = \prod_{n=0}^{N} p(\rho_n|\boldsymbol{x}_n,\boldsymbol{t},N)$$
$$\propto \left(\frac{1}{\sigma_n^2(1-\rho_n^2)}\right)^{\frac{ln}{2}} \exp\left\{-\frac{1}{2\sigma_n^2(1-\rho_n^2)}(\boldsymbol{x}_n - \mu_n\mathbf{1})'\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{x}_n - \mu_n\mathbf{1})\right\}$$
$$\propto \left(\frac{1}{\sigma_n^2(1-\rho_n^2)}\right)^{\frac{ln}{2}} \exp\left\{-\frac{1}{2\sigma_n^2(1-\rho_n^2)}\sum_{l=1}^{ln}(x(t_n+l) - \rho_n x(t_n+l-1) - (1-\rho_n)\mu_n)^2\right\}$$

- Hasting-Metropolis MCMC

- Instrumental density: A Gaussian obtained by Laplace approximation

$$m_{\rho_n} \longrightarrow \int_0^1 \rho_n \quad p(\rho_n|\boldsymbol{x},\boldsymbol{t},N)$$
$$\sigma_{\rho_n}^2 \longrightarrow \int_0^1 \rho_n^2 \quad p(\rho_n|\boldsymbol{x},\boldsymbol{t},N) - m_{\rho_n}^2$$

Sampling of $p(\boldsymbol{t}|\boldsymbol{x},\boldsymbol{\theta})$:

- $x_{t:s} = [x(t), x(t+1), \ldots, x(s)]$

$$\begin{aligned} R(t,s) &= p(x_{t:s}|t, s \text{ in the same segment}) \\ Q(t) &= p(x_{t:s}| \text{ changepoint at } t-1), \quad Q(1) = p(\boldsymbol{x}) \end{aligned}$$

- Noting $g(t_j - t_{j-1})$ the a priori density of the interval between two changepoints, and $G(t)$ its associated cumulative distribution function, we have:

$$p(t_j|t_{j-1},\boldsymbol{x}) = \frac{R(t_{j-1},t_j)Q(t_j+1)g(t_j-t_{j-1})}{Q(t_{j-1})}$$

and

$$p(t_j = T|t_{j-1},\boldsymbol{x}) = \frac{P(t_{j-1},T)(1-G(T-t_{j-1}-1))}{Q(t_{j-1})}$$

- Gibbs sampler: $t_0 = 1$, Iterate $t_j \sim p(t_j|t_{j-1},\boldsymbol{x}), j = 1, \cdots, N$.

4 Change point problem with HMM

- Hidden variables:

$$\boldsymbol{z} = [z(t_0), \cdots, z(t_0+T)]' \text{ and } \boldsymbol{q} = [q(t_0), \cdots, q(t_0+T)]'$$

  where

$$q(t) = \begin{cases} 1 & \text{if } z(t) \neq z(t-1) \\ 0 & \text{elsewhere} \end{cases} = \begin{cases} 1 & \text{if } t = t_n, n = 0, \cdots, N \\ 0 & \text{elsewhere} \end{cases}.$$

- Bernouilli-Gaussian process:

$$P(\boldsymbol{Q} = \boldsymbol{q}) = \lambda^{\sum_j q_j} (1-\lambda)^{\sum_j (1-q_j)} = \lambda^{\sum_j q_j} (1-\lambda)^{N-\sum_j q_j}$$

- Markov Chain process: $\{z(t), t = 1, \cdots, T\}$ forms a Markov chain:

$$P(z(t) = k) = p_k, \quad k = 1, \cdots, K,$$
$$P(z(t) = k | z(t-1) = l) = p_{kl}, \quad \text{with } \sum_k p_{kl} = 1.$$

Bernouilli-Gaussian process:

$$P(\boldsymbol{Q}=\boldsymbol{q}) = \lambda^{\sum_j q_j}(1-\lambda)^{\sum_j (1-q_j)} = \lambda^{\sum_j q_j}(1-\lambda)^{N-\sum_j q_j}$$

$$\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{q},\boldsymbol{\theta}) &= (2\pi)^{-N/2}\left(\prod_{n=1}^{N} 1/\sigma_n\right)\exp\left[-\frac{1}{2\sigma_n^2}\sum_{n=1}^{N}\left(x(t_n)-\mu_n\right)^2\right] \\
&+ (2\pi)^{-(T-N)/2}\left(\prod_{n=1}^{N} 1/\sigma_n^{(l_n-1)}\right)\exp\left[-\frac{1}{2\sigma_n^2}\sum_{j=1}^{T}(1-q_j)\left(x_j-x_{j-1}\right)^2\right] \\
&= (2\pi)^{-T/2}\left(\prod_{n=1}^{N} 1/\sigma_n^{(l_n)}\right)\exp\left[-\frac{1}{2\sigma_n^2}\sum_{j=1}^{T}\left[(1-q_j)\left(x_j-x_{j-1}\right)^2+q_j\left(x_j-\right.\right.\right.
\end{aligned}$$

Contextual Hidden Markov process

- $\{z(t), t = 1, \cdots, T\}$ forms a Markov chain:

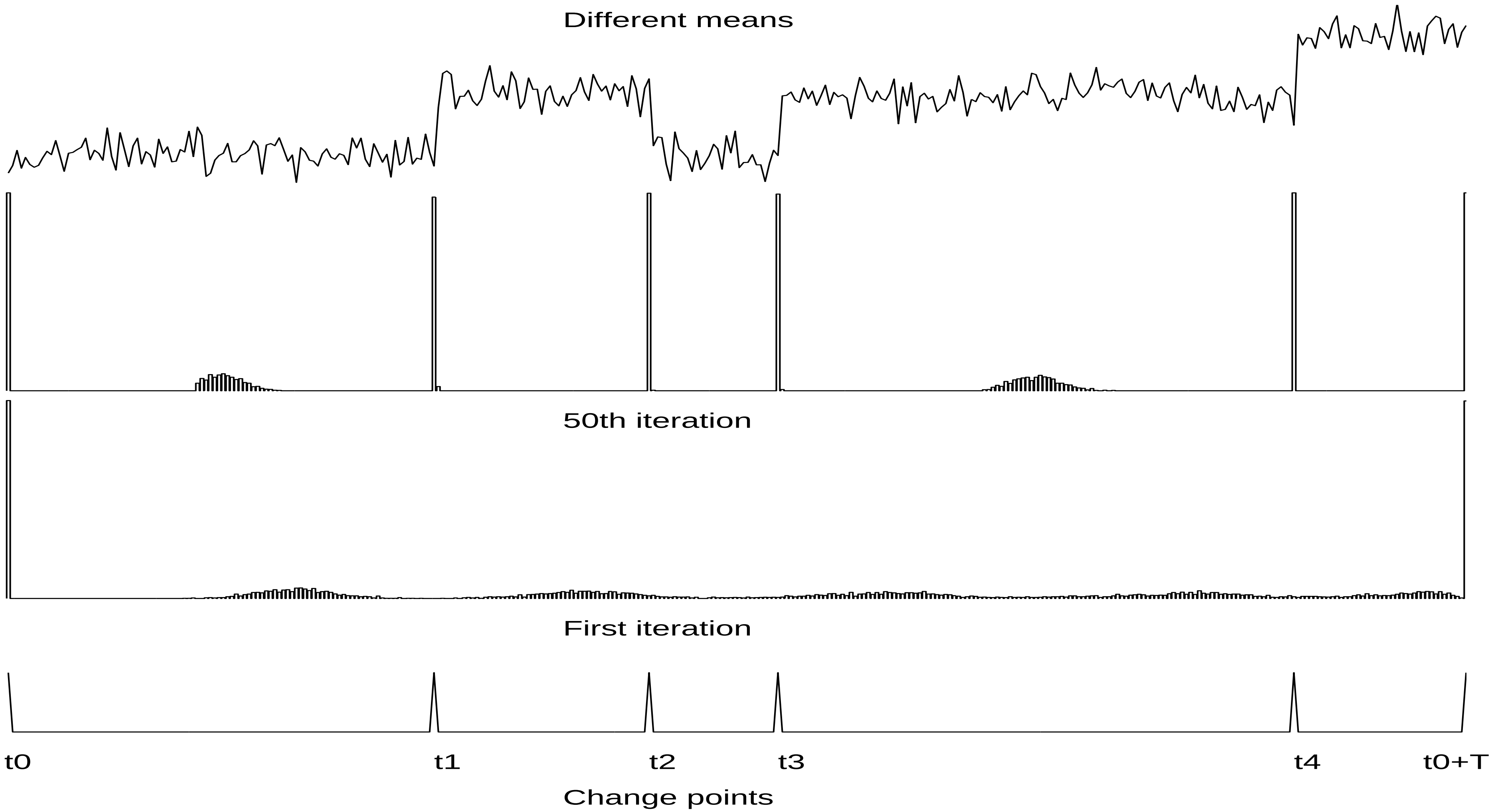
$$\begin{cases} P(z(t) = k) = p_k, \quad k = 1, \cdots, N, \\ P(z(t) = k | z(t-1) = l) = p_{kl}, \quad \text{with} \quad \sum_k p_{kl} = 1,, \quad k, l = 1, \cdots, N. \end{cases}$$

- $\boldsymbol{\theta} = \{N, \{\mu_n, \sigma_n, p_n,\ n = 1, \cdots, N\}, (p_{kl},\ k, l = 1, \cdots, N)\}$
- The model is a mixture of Gaussians.

$$\begin{aligned} p(\boldsymbol{x}_n | \boldsymbol{\theta}, z_n = n) &= \mathcal{N}(\mu_n \mathbf{1}, \boldsymbol{\Sigma}_n) \\ p(\boldsymbol{x}_n | \boldsymbol{\theta}) &= \sum_{n=1}^{N} p_n \mathcal{N}(\mu_n \mathbf{1}, \boldsymbol{\Sigma}_n) \end{aligned}$$

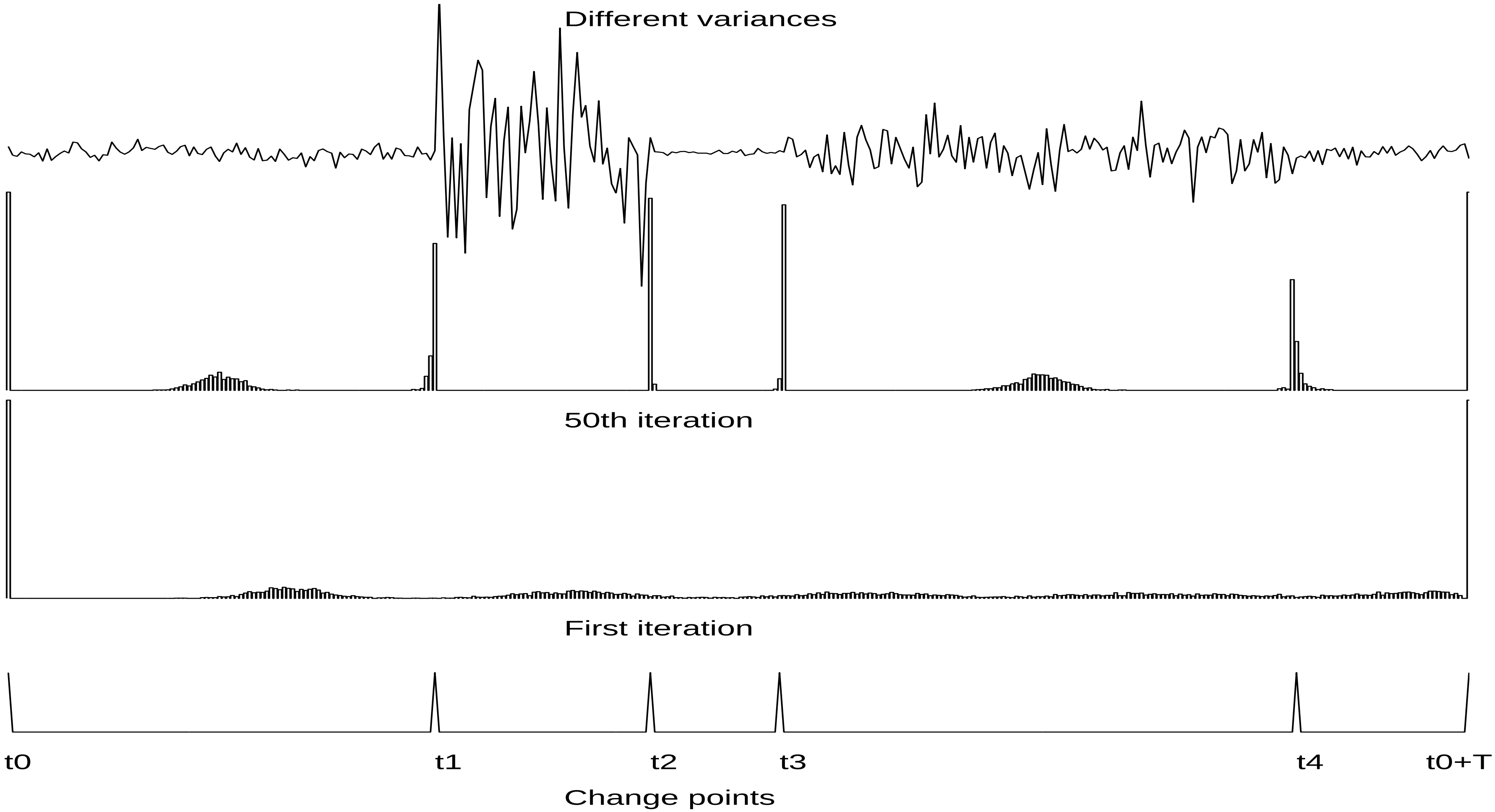
5 Simulation results

Change of the means



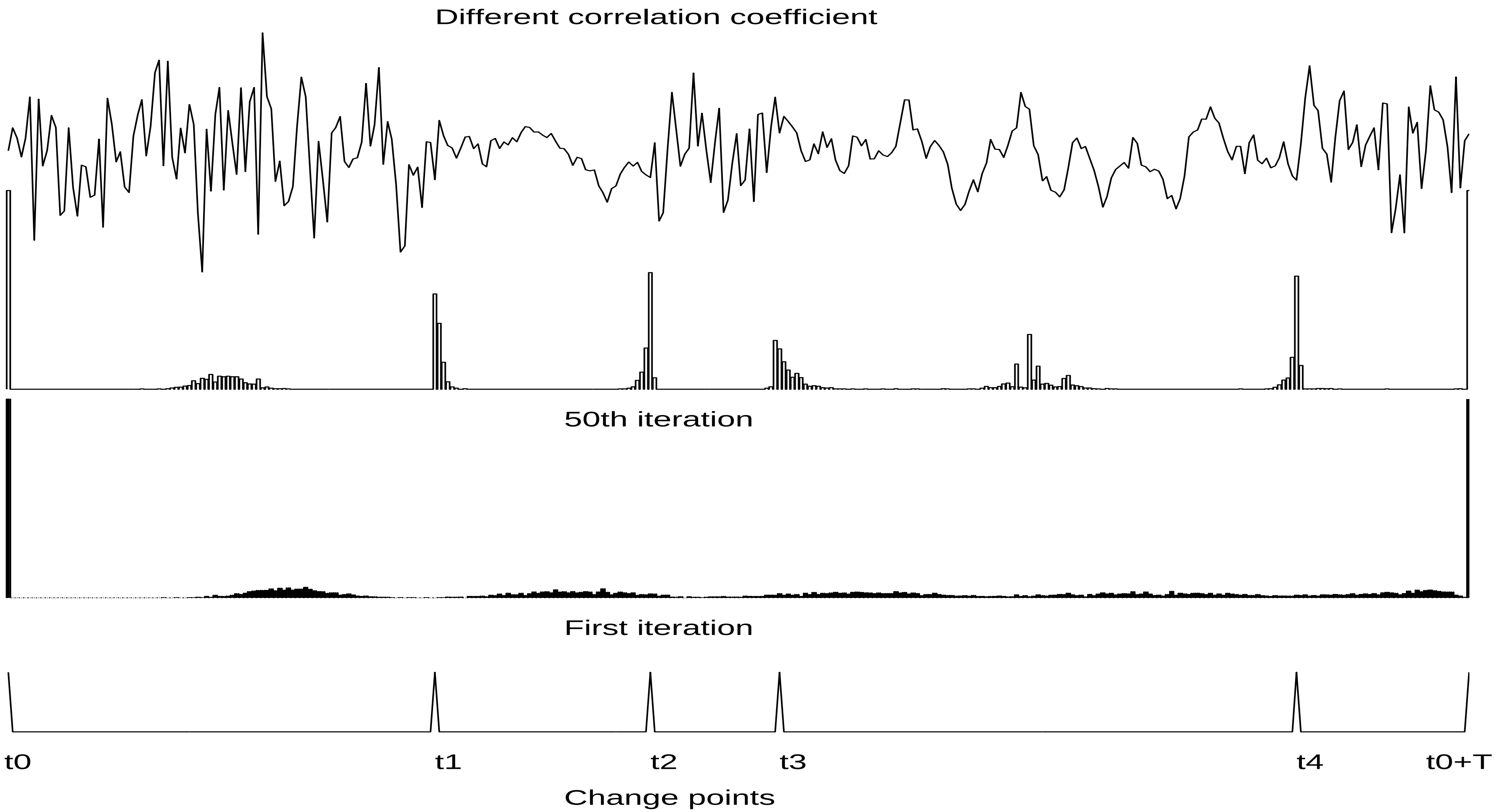
m	$\hat{m}\|\boldsymbol{x},\boldsymbol{t}$	$\hat{\sigma}^2\|\boldsymbol{x},\boldsymbol{t}$	$\hat{m}\|\boldsymbol{x}$	$\hat{\sigma}^2\|\boldsymbol{x}$
1.5	1.4966	0.0015	1.4969	0.0013
1.7	1.7084	0.0017	1.7013	0.0038
1.5	1.4912	0.0020	1.5015	0.0045
1.7	1.6940	0.0014	1.6929	0.0016
1.9	1.9012	0.0015	1.8915	0.0039

Change in the variances



σ^2	$\hat{\sigma}^2 \| \boldsymbol{x}, \boldsymbol{t}$	$\hat{\sigma}^2 \| \boldsymbol{x}$
0.01	0.0083	0.0081
1	0.9918	0.9598
0.001	0.0007	0.0026
0.1	0.0945	0.0940
0.01	0.0079	0.0107

Change in the correlation coefficient



a	$\hat{a}\|\boldsymbol{x}$
0	0.0988
0.9	0.7875
0.1	0.3737
0.8	0.8071
0.2	0.1710

6 Conclusions and Perspectives

- A Bayesian approach for estimating change points in time series is presented
- Detection of change points due to changes in the mean is easier than those due to changes in variances or changes in correlation coefficient.
- In this work, first we assumed to know the number N of change points.
- We also studied the role of the a priori parameter λ on the results.
- We are investigating the estimation of the number of change points in the same framework.
- Other modeling using other hidden variables than change point time instants are possible and are under investigation.
- We are also investigating the extension of this work to image processing (2D signals) where the change points are contours.