Change-point detection in exponential families

Olivier Lopez[(1)], Vladimir Spokoiny [(2)]

(1) LSTA Université Paris VI, (2) WIAS, Berlin

Colloque Franco-Roumain, Poitiers 30 août 2010

Introduction 1/3

Change-point model

Observations : $(Y_i)_{1\leq i\leq n}$ independent random variables with law $\mathbb{P}_{v_i}$, where

$$v_i = v + a^* \mathbf{1}_{i\leq\tau^*},$$

with $a^* \in \mathcal{A}$, and $\tau^* \in \{0, ..., n\}$.

- We assume that $v \in \mathbb{R}$ is known.
- Aim : estimation of $\theta^* = (a^*, \tau^*)$.
- Identification problem : cases where $\tau^* = 0$, $a^* = 0$. The set of parameters is $\theta \in (\mathbb{R}^{*+} \times \{1, ..., n\}) \cup \{(0, 0)\}$.

Introduction 1/3

Change-point model

Observations : $(Y_i)_{1\leq i\leq n}$ independent random variables with law $\mathbb{P}_{v_i}$, where

$$v_i = v + a^* \mathbf{1}_{i\leq\tau^*},$$

with $a^* \in \mathcal{A}$, and $\tau^* \in \{0, ..., n\}$.

- We assume that $v \in \mathbb{R}$ is known.
- Aim : estimation of $\theta^* = (a^*, \tau^*)$.
- Identification problem : cases where $\tau^* = 0$, $a^* = 0$. The set of parameters is $\theta \in (\mathbb{R}^{*+} \times \{1, ..., n\}) \cup \{(0, 0)\}$.

Introduction 1/3

Change-point model

Observations : $(Y_i)_{1 \leq i \leq n}$ independent random variables with law $\mathbb{P}_{v_i}$, where

$$v_i = v + a^* \mathbf{1}_{i \leq \tau^*},$$

with $a^* \in \mathcal{A}$, and $\tau^* \in \{0, ..., n\}$.

- We assume that $v \in \mathbb{R}$ is known.
- Aim : estimation of $\theta^* = (a^*, \tau^*)$.
- Identification problem : cases where $\tau^* = 0$, $a^* = 0$. The set of parameters is $\theta \in (\mathbb{R}^{*+} \times \{1, ..., n\}) \cup \{(0, 0)\}$.

Introduction 2/3

Example 1

- Y_i = number of claims during time period i.
- Classical model : Poisson random variable with mean λ.
- τ^* = change of behavior of insured people, the mean number of claims becomes $\lambda + \mu$.

Exemple 2

- Y_i = amount of the $i-$th claim.
- Classical model : Pareto.
- τ^* = time after which this model is not adapted anymore.

Introduction 2/3

Example 1

- $Y_i =$ number of claims during time period i.
- Classical model : Poisson random variable with mean λ.
- $\tau^* =$ change of behavior of insured people, the mean number of claims becomes $\lambda + \mu$.

Exemple 2

- $Y_i =$ amount of the $i-$th claim.
- Classical model : Pareto.
- $\tau^* =$ time after which this model is not adapted anymore.

Introduction 3/3

- Some asymptotic results on change-point :
  - Csörgo, Horvath (1997)
  - Haccou, Meelis, Van de Geer (1987)
- For finite sample size in the Gaussian case :
  - Golubev, Spokoiny (2009)
- Aim : extend these results to the case of a canonical exponential family.

Introduction 3/3

- Some asymptotic results on change-point :
  - Csörgo, Horvath (1997)
  - Haccou, Meelis, Van de Geer (1987)
- For finite sample size in the Gaussian case :
  - Golubev, Spokoiny (2009)
- Aim : extend these results to the case of a canonical exponential family.

Introduction 3/3

- Some asymptotic results on change-point :
  - Csörgo, Horvath (1997)
  - Haccou, Meelis, Van de Geer (1987)
- For finite sample size in the Gaussian case :
  - Golubev, Spokoiny (2009)
- Aim : extend these results to the case of a canonical exponential family.

Outline

1. **Case where a^* is known**
   - Maximum likelihood estimation
   - Obtained results
2. **Case where a^* is unknown**
   - Comparison with the Gaussian case
   - Exponential bounds for a random field
   - Geometric problem
   - Obtained results

Outline

1. **Case where a^* is known**
   - Maximum likelihood estimation
   - Obtained results
2. **Case where a^* is unknown**
   - Comparison with the Gaussian case
   - Exponential bounds for a random field
   - Geometric problem
   - Obtained results

Likelihood expression

- Density of $\mathbb{P}_a$ with respect to a dominating measure μ :

$$p(y)\exp(ya - d(a)),$$

with d C^2 with $d'' > 0$.

- Without loss of generality, $v = 0$, $d(0) = d'(0) = 0$, change-point model becomes

$$v_i = a^* \mathbf{1}_{i \leq \tau^*}.$$

- Log-likelihood expression :

$$L(\tau) = a^* \sum_{i=1}^{\tau} Y_i - \tau d(a^*).$$

The Maximum Likelihood Estimator maximizes $L(\tau, \tau^*) = L(\tau) - L(\tau^*)$.

Likelihood expression

- Density of $\mathbb{P}_a$ with respect to a dominating measure μ :

$$p(y)\exp(ya - d(a)),$$

with d C^2 with $d'' > 0$.

- Without loss of generality, $v = 0$, $d(0) = d'(0) = 0$, change-point model becomes

$$v_i = a^* \mathbf{1}_{i \le \tau^*}.$$

- Log-likelihood expression :

$$L(\tau) = a^* \sum_{i=1}^{\tau} Y_i - \tau d(a^*).$$

The Maximum Likelihood Estimator maximizes $L(\tau, \tau^*) = L(\tau) - L(\tau^*)$.

Main result

- Let $\mathfrak{M}(\tau, \tau^*) = -\log E[\exp(1/2L(\tau, \tau^*))]$.

Theorem (Case where a^* is known)

We have

$$\mathfrak{D}(\alpha) = E\left[\sup_{\tau} \exp(1/2L(\tau, \tau^*) + \alpha\mathfrak{M}(\tau, \tau^*))\right] \leq C,$$

where C only depends of $\sup_{a \in [0,a^*]} d''(a)$, *and* $\alpha < 1$.

- Proof : Doob's inequality, or results of Golubev and Spokoiny (2009).

Main result

- Let $\mathfrak{M}(\tau, \tau^*) = -\log E[\exp(1/2L(\tau, \tau^*))]$.

Theorem (Case where a^* is known)

We have

$$\mathfrak{D}(\alpha) = E\left[\sup_{\tau} \exp(1/2L(\tau, \tau^*) + \alpha\mathfrak{M}(\tau, \tau^*))\right] \leq C,$$

where C only depends of $\sup_{a\in[0,a^*]} d''(a)$, *and* $\alpha < 1$.

- Proof : Doob's inequality, or results of Golubev and Spokoiny (2009).

Two corollaries

- Let $\hat{\tau}$ be the maximum likelihood estimator,

$$E\left[\exp(\alpha\mathfrak{M}(\hat{\tau},\tau^*))\right] \leq E[\exp(1/2L(\hat{\tau},\tau^*)+\alpha\mathfrak{M}(\hat{\tau},\tau^*))] \leq \mathfrak{D}(\alpha).$$

Corollary (Estimation quality)

We have

$$E\left[\mathfrak{M}(\hat{\tau},\tau^*)\right] \leq \tilde{C}.$$

Corollary (Confidence intervals)

Let $\mathcal{A}(z) = \{\tau : L(\hat{\tau},\tau) \leq z\}$. We have

$$\mathbb{P}(\tau^* \notin \mathcal{A}(z)) \leq \mathfrak{D}(0)\exp(-z/2).$$

Two corollaries

- Let $\hat{\tau}$ be the maximum likelihood estimator,

$$E\left[\exp(\alpha\mathfrak{M}(\hat{\tau},\tau^*))\right] \leq E[\exp(1/2L(\hat{\tau},\tau^*)+\alpha\mathfrak{M}(\hat{\tau},\tau^*))] \leq \mathfrak{D}(\alpha).$$

Corollary (Estimation quality)

We have

$$E\left[\mathfrak{M}(\hat{\tau},\tau^*)\right] \leq \tilde{C}.$$

Corollary (Confidence intervals)

Let $\mathcal{A}(z) = \{\tau : L(\hat{\tau},\tau) \leq z\}$. *We have*

$$\mathbb{P}(\tau^* \notin \mathcal{A}(z)) \leq \mathfrak{D}(0)\exp(-z/2).$$

Outline

1. Case where a^* is known
   - Maximum likelihood estimation
   - Obtained results
2. Case where a^* is unknown
   - Comparison with the Gaussian case
   - Exponential bounds for a random field
   - Geometric problem
   - Obtained results

Comparison with the Gaussian case

Estimation of a^*

- With τ fixed, the likelihood $L(a, \tau)$ is maximum for

$$\hat{a}(\tau) = d'^{-1}\left(\frac{1}{\tau}\sum_{i=1}^{\tau} Y_i\right).$$

- Gaussian case : $\hat{a}(\tau) = \tau^{-1}\sum_{i=1}^{\tau} Y_i$. The MLE $\hat{\tau}$ maximizes

$$\tilde{L}(\tau) = \left(\frac{1}{\tau^{1/2}}\sum_{i=1}^{\tau} Y_i\right)^2.$$

- Non-Gaussian case :

$$\tilde{L}(\tau) = d'^{-1}(\tau^{-1}\sum_{i=1}^{\tau} Y_i)\sum_{i=1}^{\tau} Y_i - \tau d(d'^{-1}(\tau^{-1}\sum_{i=1}^{\tau} Y_i)).$$

- Conclusion : Difficult to separate a and τ.

Comparison with the Gaussian case

Estimation of a^*

- With τ fixed, the likelihood $L(a, \tau)$ is maximum for

$$\hat{a}(\tau) = d'^{-1}\left(\frac{1}{\tau}\sum_{i=1}^{\tau} Y_i\right).$$

- Gaussian case : $\hat{a}(\tau) = \tau^{-1}\sum_{i=1}^{\tau} Y_i$. The MLE $\hat{\tau}$ maximizes

$$\tilde{L}(\tau) = \left(\frac{1}{\tau^{1/2}}\sum_{i=1}^{\tau} Y_i\right)^2.$$

- Non-Gaussian case :

$$\tilde{L}(\tau) = d'^{-1}(\tau^{-1}\sum_{i=1}^{\tau} Y_i)\sum_{i=1}^{\tau} Y_i - \tau d(d'^{-1}(\tau^{-1}\sum_{i=1}^{\tau} Y_i)).$$

- Conclusion : Difficult to separate a and τ.

Estimation of a^*

- With τ fixed, the likelihood $L(a,\tau)$ is maximum for

$$\hat{a}(\tau) = d'^{-1}\left(\frac{1}{\tau}\sum_{i=1}^{\tau} Y_i\right).$$

- Gaussian case : $\hat{a}(\tau) = \tau^{-1}\sum_{i=1}^{\tau} Y_i$. The MLE $\hat{\tau}$ maximizes

$$\tilde{L}(\tau) = \left(\frac{1}{\tau^{1/2}}\sum_{i=1}^{\tau} Y_i\right)^2.$$

- Non-Gaussian case :

$$\tilde{L}(\tau) = d'^{-1}(\tau^{-1}\sum_{i=1}^{\tau} Y_i)\sum_{i=1}^{\tau} Y_i - \tau d(d'^{-1}(\tau^{-1}\sum_{i=1}^{\tau} Y_i)).$$

- Conclusion : Difficult to separate a and τ.

Estimation of a^*

- With τ fixed, the likelihood $L(a, \tau)$ is maximum for
$$\hat{a}(\tau) = d'^{-1}\left(\frac{1}{\tau}\sum_{i=1}^{\tau} Y_i\right).$$
- Gaussian case : $\hat{a}(\tau) = \tau^{-1}\sum_{i=1}^{\tau} Y_i$. The MLE $\hat{\tau}$ maximizes
$$\tilde{L}(\tau) = \left(\frac{1}{\tau^{1/2}}\sum_{i=1}^{\tau} Y_i\right)^2.$$
- Non-Gaussian case :
$$\tilde{L}(\tau) = d'^{-1}(\tau^{-1}\sum_{i=1}^{\tau} Y_i)\sum_{i=1}^{\tau} Y_i - \tau d(d'^{-1}(\tau^{-1}\sum_{i=1}^{\tau} Y_i)).$$
- Conclusion : Difficult to separate a and τ.

Result of Golubev et Spokoiny (2009)

- Let $\mathfrak{d}(a, \tau, a', \tau')$ be a metric such as, if $B(\varepsilon, a, \tau) = \{(a', \tau') : \mathfrak{d}(a, \tau, a', \tau') \leq \varepsilon\}$,

$$E\left[\sup_{(a',\tau')\in B(\varepsilon,a,\tau)} \exp\left(2\lambda \frac{L(a, \tau, a', \tau') - EL(a, \tau, a', \tau')}{\mathfrak{d}(a, \tau, a', \tau')}\right)\right] \leq 2\nu_0^2\lambda^2$$

Theorem

Let π be a $\sigma-$finite measure on the parameter space. Under conditions on the metric $\mathfrak{d}$,

$$\mathfrak{D}(\alpha) = E\left[\sup_{a,\tau} \exp(1/2L(a, \tau, a^*, \tau^*) + \alpha\mathfrak{M}(a, \tau, a^*, \tau^*))\right] \leq C,$$

$$\text{où } \log C = C_0(\mathfrak{d}) + \log\left(\int \frac{\exp(-\alpha\mathfrak{M}_\varepsilon(a,\tau,a^*,\tau^*))d\pi(a,\tau)}{\pi(B(\varepsilon,a,\tau))}\right).$$

Natural metric

- Gaussian case : $\mathfrak{d}(a,\tau,a',\tau') = (a-a')^2\tau + a'^2(\tau'-\tau)$, for $\tau' > \tau$.
- a big ($a > \varepsilon/\tau^{1/2}$), we can approximate $B(\varepsilon, a, \tau)$ by a rectangle.
- a small ($a < \varepsilon/\tau^{1/2}$) : tends to a set delimited by an hyperbola.
- Consequence : difficulties to control the local entropy. Necessity to adapt the result of Golubev and Spokoiny to the case of a basis of neighborhoods instead of using a metric.

Natural metric

- Gaussian case : $\mathfrak{d}(a, \tau, a', \tau') = (a - a')^2\tau + a'^2(\tau' - \tau)$, for $\tau' > \tau$.
- a big ($a > \varepsilon/\tau^{1/2}$), we can approximate $B(\varepsilon, a, \tau)$ by a rectangle.
- a small ($a < \varepsilon/\tau^{1/2}$) : tends to a set delimited by an hyperbola.
- Consequence : difficulties to control the local entropy. Necessity to adapt the result of Golubev and Spokoiny to the case of a basis of neighborhoods instead of using a metric.

Natural metric

- Gaussian case : $\mathfrak{d}(a,\tau,a',\tau') = (a-a')^2\tau + a'^2(\tau'-\tau)$, for $\tau' > \tau$.
- a big ($a > \varepsilon/\tau^{1/2}$), we can approximate $B(\varepsilon, a, \tau)$ by a rectangle.
- a small ($a < \varepsilon/\tau^{1/2}$) : tends to a set delimited by an hyperbola.
- Consequence : difficulties to control the local entropy. Necessity to adapt the result of Golubev and Spokoiny to the case of a basis of neighborhoods instead of using a metric.

Obtained results

Result

The MLE satifsies

$$E\left[\mathfrak{M}(\hat{\tau}, \hat{a}, \tau^*, a^*)\right] \leq C \log\log n.$$

- Gaussian case :

$$E\left[a^{*2}\tau^*|\hat{\tau} - \tau^*|\right] \leq C\sigma^2 \log\log n.$$

- Poisson case :

$$E\left[\tau^* \log((\lambda+\mu)/\lambda)^2|\hat{\tau} - \tau^*|\right] \leq \frac{C \log\log n}{\lambda}.$$

Obtained results

Result

The MLE satifsies

$$E\left[\mathfrak{M}(\hat{\tau}, \hat{a}, \tau^*, a^*)\right] \leq C \log\log n.$$

- Gaussian case :

$$E\left[{a^*}^2 \tau^* |\hat{\tau} - \tau^*|\right] \leq C\sigma^2 \log\log n.$$

- Poisson case :

$$E\left[\tau^* \log((\lambda+\mu)/\lambda)^2 |\hat{\tau} - \tau^*|\right] \leq \frac{C \log\log n}{\lambda}.$$

Lower bound

Lower bound

Let $\mathfrak{R} = \inf_{\tilde{\theta}} \max_{\theta} E\left[\mathfrak{M}(\tilde{\theta}, \theta)\right]$. We have

$$\mathfrak{R} \geq c_0 \log \log n.$$

- Proof : modification of Fano's Lemma from Birgé (2001).

Conclusion

- Extension to the case of a misspecified model is possible.
- Extension to the case of multiple change-point.
- Extension to online detection.

Conclusion

- Extension to the case of a misspecified model is possible.
- Extension to the case of multiple change-point.
- Extension to online detection.

Conclusion

- Extension to the case of a misspecified model is possible.
- Extension to the case of multiple change-point.
- Extension to online detection.