Bump Detection in Heterogeneous Gaussian Regression

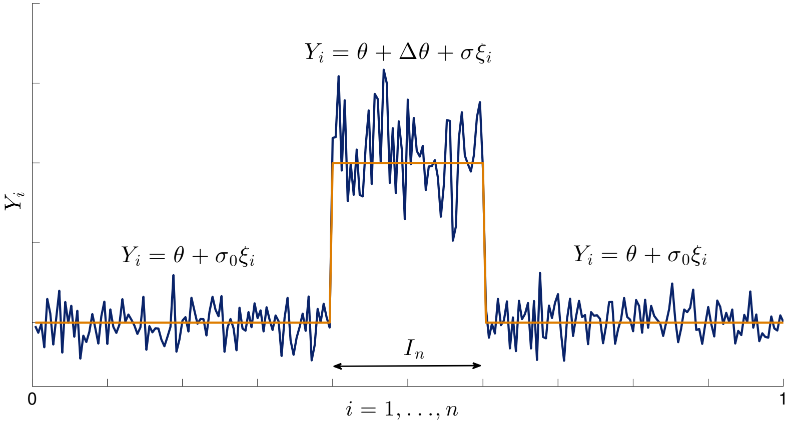
Farida Enikeeva[1], Axel Munk[2] and Frank Werner[2]

[1]LMA–University of Poitiers, [2]Max Planck Institute for Biophysical Chemistry

February 12, 2016

CIRM, Luminy

Problem



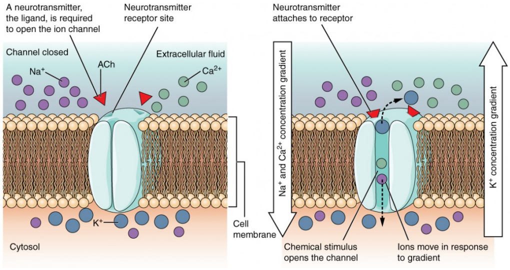
Heterogeneous Bump Detection :

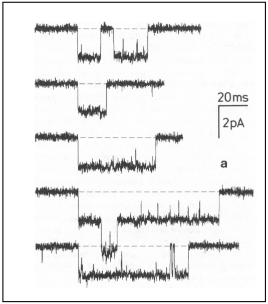
Detect a simultaneous change in mean and variance within an unknown interval I_n

Analysis of financial stock data

- Two features of financial asset prices :
  - the volatility changes over time
  - the trajectories can have large discontinuities
- The price jumps are often associated with simultaneous discontinuous changes in the level of volatility [Jacod & Todorov '10 ; Todorov & Tauchen '10]

Fluctuations of ionic currents





- Closed channel : no current (the signal mean is zero, some initial noise is present)
- Open channel : change in mean of the current and in increase in variance of the noise [Sigworth, '85 ; Schrimer '98]
- The noise is Gaussian

Image from https ://courses.candelalearning.com/ap2x1/chapter/the-action-potential/

Heterogeneous Bump Detection : Model

Heterogeneous Bump Model (HBR)

$$Y_i = \mu_n\left(\frac{i}{n}\right) + \lambda_n\left(\frac{i}{n}\right)\xi_i, \quad i = 1, \ldots, n$$

- ξ_i are $\mathcal{N}(0,1)$ i.i.d.
- Change in mean on the interval I_n :

$$\mu_n(x) = \begin{cases} \Delta_n & \text{if } x \in I_n, \\ 0 & \text{otherwise,} \end{cases}$$

- Simultaneous change in variance on I_n :

$$\lambda_n^2(x) = \sigma_0^2 + \sigma_n^2 1\{x \in I_n\}, \quad x \in [0,1]$$

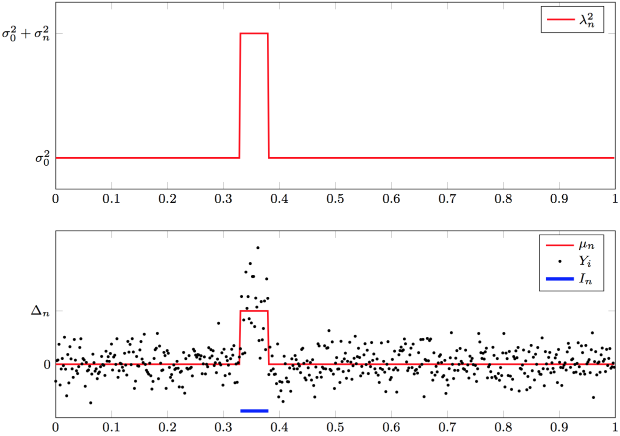


Figure: The HBR model : $\Delta_n = 4$, $\sigma_0^2 = 1$, $\sigma_n^2 = 4$ and $n = 512$.

Problem Statement

- Observations $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$:

$$Y_i = \mu_n\left(\frac{i}{n}\right) + \lambda_n\left(\frac{i}{n}\right)\xi_i, \quad i = 1, \ldots, n$$

- Let $|I_n|$ be the length of $I_n \subset [0,1]$.
- Define the set of contiguous intervals

$$\mathcal{A}_n := \Big\{ \big[(j-1)|I_n|, j|I_n|\big] \ ; \ 1 \leq j \leq 1/|I_n| \Big\}.$$

Testing problem

$$\mathrm{H}_0 : \ \mu_n \equiv 0, \lambda_n \equiv \sigma_0, \quad \sigma_0 > 0 \text{ fixed}$$

against

$$\mathrm{H}_1^n : \exists\ I_n \in \mathcal{A}_n \text{ s.t. } \mu_n = \Delta_n 1_{I_n},\ \lambda_n^2 = \sigma_0^2 + \sigma_n^2 1_{I_n}, \quad \sigma_n > 0.$$

Detection boundary conditions w.r.t. (Δ_n, σ_n)?

Problem Statement

Testing problem

$$\mathrm{H}_0 : \ \mu_n \equiv 0, \lambda_n \equiv \sigma_0, \quad \sigma_0 > 0 \text{ fixed}$$

against

$$\mathrm{H}_1^n : \exists \ I_n \in \mathcal{A}_n \text{ s.t. } \mu_n = \Delta_n 1_{I_n}, \ \lambda_n^2 = \sigma_0^2 + \sigma_n^2 1_{I_n}, \quad \sigma_n > 0.$$

For a test $\Phi_n : \mathbf{Y}_n \to \{0, 1\}$ define

- Test of level α : $\limsup_{n\to\infty} \mathbb{P}_{\mathrm{H}_0}\Big\{\Phi_n(\mathbf{Y}_n) = 1\Big\} \leq \alpha$;
- Type II error :

$$\mathbb{P}_{\mathrm{H}_1^n}\Big\{\Phi_n(\mathbf{Y}_n) = 0\Big\} := \sup_{I_n \in \mathcal{A}_n} \mathbb{P}_{\mu_n = \Delta_n \mathbf{1}_{I_n}, \ \lambda_n^2 = \sigma_0^2 + \sigma_n^2 \mathbf{1}_{I_n}}\Big\{\Phi_n(\mathbf{Y}_n) = 0\Big\}.$$

- The change of variance parameter : $\kappa_n^2 = \frac{\sigma_n^2}{\sigma_0^2} > 0$.

Remark : we consider only the situation of $\Delta_n > 0$ and increasing variance : σ_0^2 changes to $\sigma_0^2 + \sigma_n^2 := \sigma_0^2(1 + \kappa_n^2)$.

Detection boundary

The goal is to find a set of detection boundary conditions on the bump functions

$$S_n = \{\mu_n = \Delta_n \mathbf{1}_{I_n}, \lambda_n^2 = \sigma_0^2 + \sigma_n^2 \mathbf{1}_{I_n} \text{ s.t. } \Delta_n, |I_n|, \kappa_n \text{ satisfy conditions}\}$$

such that

- upper detection bound :
  there exists a test which can differentiate between the null hypothesis $\mu \equiv 0, \lambda_n \equiv \sigma_0$ and the signals vanishing slower than S_n at level α with a power greater or equal $1 - \alpha$.
- lower detection bound :
  no test at level α under the null-hypothesis $\mu \equiv 0, \lambda_n \equiv \sigma_0$ can differentiate between the signals vanishing faster than S_n with power strictly greater than α.

Homogeneous case : $\sigma_n^2 = 0$

Observations :

$$Y_i = \mu_n\Big(\frac{i}{n}\Big) + \sigma_0 \xi_i, \quad i = 1, \dots, n$$

Testing problem :

$$\mathrm{H}_0 : \mu_n \equiv 0 \quad \text{against} \quad \mathrm{H}_1^n : \exists\ I_n \in \mathcal{A}_n \text{ s.t. } \mu_n = \Delta_n \mathbf{1}_{I_n}$$

Detection boundary for Δ_n known [Dümbgen, Spokoiny '01 ; Frick, Munk, Siegel '14]

Let $|I_n| \to 0$ and $\varepsilon_n > 0$ be a sequence such that $\varepsilon_n \to 0$ and

$$\lim_{n\to\infty} \varepsilon_n \sqrt{-\log |I_n|} \to \infty.$$

Then the detection boundary set is given by

$$S_n = \Big\{\mu(t) = \Delta_n \mathbf{1}_{I_n}(t) : \ \sqrt{n\,|I_n|}\Delta_n = (\sqrt{2}\sigma_0 \pm \varepsilon_n)\sqrt{-\log(|I_n|)}\Big\}$$

Homogeneous case

Likelihood ratio

Let $I_n^j = \left[(j-1)|I_n|, j|I_n|\right] \in \mathcal{A}_n$. Then

$$\log L_{n,j}(\Delta_n; \mathbf{Y}_n) = \frac{\Delta_n}{\sigma_0^2} \sum_{i \in I_n^j} Y_i - \frac{\Delta_n^2 n |I_n|}{2\sigma_0^2}$$

Under H_0 we have

$$\log L_{n,j}((\Delta_n; \mathbf{Y}_n) = -\frac{\Delta_n^2 n |I_n|}{2\sigma_0^2} + \frac{\Delta_n}{\sigma_0} \sqrt{n|I_n|} Z_j, \quad Z_j \sim \mathcal{N}(0,1)$$

Homogeneous case : lower bound

For any test Φ_n such that $\Phi_n(\mathbf{Y}_n) \leq \alpha$ we have

$$\inf_{1 \leq j \leq 1/|I_n|} \mathbf{E}_{\mu_n,j} \Phi_n(\mathbf{Y}_n) - \alpha$$

$$\leq \mathbf{E}_{\mu_0} \left| |I_n| \sum_{j=1}^{1/|I_n|} L_{n,j}(\mathbf{Y}_n) - 1 \right|$$

$$\leq \mathbf{E} \left| |I_n| \sum_{j=1}^{1/|I_n|} \exp\left(\frac{\Delta_n}{\sigma_0} \sqrt{n|I_n|} Z_j - \frac{\Delta_n^2 n |I_n|}{2\sigma_0^2} \right) - 1 \right|.$$

Homogeneous case : lower bound

Lemma (Lepski, Tsybakov '00 ; Dümbgen, Spokoiny '01)

Let $Z_1, \ldots, Z_m$ *be i.i.d.* $\mathcal{N}(0,1)$. *If* $w_m = \sqrt{2\log m}(1-\varepsilon_m)$ *with* $\lim_{m\to\infty} \varepsilon_m = 0$ *and* $\lim_{m\to\infty} \varepsilon_m \sqrt{\log m} = \infty$, *then*

$$\mathbf{E}\left|\frac{1}{m}\sum_{j=1}^{m} \exp(w_m Z_j - w_m^2/2) - 1\right| \to 0, \quad m \to \infty.$$

Set

$$m = 1/|I_n|, \quad w_m = \frac{\Delta_n}{\sigma_0}\sqrt{n|I_n|}.$$

Thus $\lim_{n\to\infty} \inf_{1\leq j\leq 1/|I_n|} \mathbf{E}_{\mu_n,j}\Phi_n(\mathbf{Y}_n) \leq \alpha$ if

$$\sqrt{n\,|I_n|}\Delta_n = (\sqrt{2}\sigma_0 - \varepsilon_n)\sqrt{-\log(|I_n|)}$$

for $\varepsilon_n \to 0$ such that

$$\varepsilon_n\sqrt{\log(1/|I_n|)} \to 0.$$

Homogeneous case : lower bound

Lemma (Lepski, Tsybakov '00 ; Dümbgen, Spokoiny '01)

Let $Z_1, \dots, Z_m$ be i.i.d. $\mathcal{N}(0,1)$. If $w_m = \sqrt{2\log m}(1-\varepsilon_m)$ with $\lim_{m\to\infty} \varepsilon_m = 0$ and $\lim_{m\to\infty} \varepsilon_m \sqrt{\log m} = \infty$, then

$$\mathbf{E}\left|\frac{1}{m}\sum_{j=1}^{m} \exp(w_m Z_j - w_m^2/2) - 1\right| \to 0, \quad m \to \infty.$$

Set

$$m = 1/|I_n|, \quad w_m = \frac{\Delta_n}{\sigma_0}\sqrt{n|I_n|}.$$

Thus $\lim_{n\to\infty} \inf_{1\leq j\leq 1/|I_n|} \mathbf{E}_{\mu_n,j}\Phi_n(\mathbf{Y}_n) \leq \alpha$ if

$$\sqrt{n\,|I_n|}\Delta_n = (\sqrt{2}\sigma_0 - \varepsilon_n)\sqrt{-\log\left(|I_n|\right)}$$

for $\varepsilon_n \to 0$ such that

$$\varepsilon_n\sqrt{\log(1/|I_n|)} \to 0.$$

Heterogeneous case, Δ_n known

We observe

$$Y_i = \mu_n\left(\frac{i}{n}\right) + \lambda_n\left(\frac{i}{n}\right)\xi_i, \quad i = 1, \dots, n$$

- Under H_0 :

$$\mu_n \equiv 0, \quad \lambda_n^2 \equiv \sigma_0^2$$

- Under H_1 : for $\Delta_n > 0$, $\kappa_n > 0$

$$\mu_n\Big(\frac{i}{n}\Big) = \Delta_n \mathbf{1}\Big\{\frac{i}{n} \in I_n\Big\}, \quad \lambda_n^2\Big(\frac{i}{n}\Big) = \sigma_0^2 + \sigma_0^2\kappa_n^2\mathbf{1}\Big\{\frac{i}{n} \in I_n\Big\}$$

Likelihood

$$L_n(\Delta_n, I_n, \kappa_n; \mathbf{Y}_n) = \Big(\kappa_n^2 + 1\Big)^{-\frac{n|I_n|}{2}} \exp\left(\frac{\kappa_n^2}{2\sigma_0^2(1+\kappa_n^2)} \sum_{i:\frac{i}{n}\in I_n}\Big(Y_i + \frac{\Delta_n}{\kappa_n^2}\Big)^2 - \frac{n|I_n|\Delta_n^2}{2\sigma_0^2\kappa_n^2}\right)$$

Lower Bound

Theorem

Let $\Delta_n > 0$, $\kappa_n > 0$, and $|I_n|$ be known. If $|I_n| \searrow 0$ and there exists a sequence $\delta_n > 0$, satisfying $\delta_n < 1/\kappa_n^2$ such that for $n \to \infty$,

$$\frac{n\,|I_n|\,\Delta_n^2}{2\sigma_0^2}\frac{(1+\delta_n)\,\delta_n}{1-\delta_n\kappa_n^2} - \delta_n\frac{n\,|I_n|}{2}\log\left(1+\kappa_n^2\right) - \frac{n\,|I_n|}{2}\log\left(1-\delta_n\kappa_n^2\right) + \delta_n\log\left(|I_n|\right) \to -\infty,$$

then sequence $\mathcal{S}_n$ determining the asymptotic behavior of Δ_n, κ_n and $|I_n|$ is undetectable.

Proof is based on the inequality

$$\inf_{(\mu_n,\lambda_n)\in S_n} \mathbf{E}_{\mu_n,\lambda_n}\Phi_n(Y) - \alpha \le \mathbf{E}_{\mu\equiv 0,\lambda\equiv\sigma_0}\left|\frac{1}{I_n}\sum_{I_n\in\mathcal{A}_n} L_n(\Delta_n, I_n, \kappa_n, Y) - 1\right| + o(1)$$

and on the modification of the Lemma for the heterogeneous case.

Upper bound

- Likelihood

$$L_n(\Delta_n, I_n, \kappa_n; \mathbf{Y}_n) = \left(\kappa_n^2+1\right)^{-\frac{n|I_n|}{2}} \exp\left(\frac{\kappa_n^2}{2\sigma_0^2(1+\kappa_n^2)} \sum_{i:\frac{i}{n}\in I_n} \left(Y_i + \frac{\Delta_n}{\kappa_n^2} \right)^2 - \frac{n|I_n|\Delta_n^2}{2\sigma_0^2\kappa_n^2} \right)$$

- Likelihood ratio test statistic :

$$T_n^{\Delta_n,\kappa_n,|I_n|}(Y) := \sup_{I_n\in\mathcal{A}_n} \frac{1}{\sigma_0^2} \sum_{i:\frac{i}{n}\in I_n} \left(Y_i + \frac{\Delta_n}{\kappa_n^2} \right)^2$$

The LR test of level α is defined by

$$\Phi_n(Y) := \begin{cases} 1, & \text{if } T_n^{\Delta_n,\kappa_n,|I_n|}(Y) > c^*_{\alpha,n}, \\ 0, & \text{otherwise.} \end{cases}$$

Upper bound

Theorem

Let $|I_n| \searrow 0$, $\sigma_n > 0$ *and* $\alpha \in (0,1)$ *be a given significance level. Set*

$$c^*_{\alpha,n} = n\,|I_n| + \frac{n\,|I_n|\,\Delta_n^2}{\sigma_0^2\kappa_n^4} - 2\log\left(\alpha\,|I_n|\right) + 2\sqrt{n\,|I_n|\left(1 + 2\frac{\Delta_n^2}{\sigma_0^2\kappa_n^4}\right)\log\left(\frac{1}{\alpha\,|I_n|}\right)}.$$

Assume that $\Delta_n, |I_n|$ *and* κ_n *satisfy*

$$\begin{aligned} n\,|I_n|\left(\kappa_n^4 + 2\frac{\Delta_n^2}{\sigma_0^2}\right) + \frac{\kappa_n^2\Delta_n^2 n\,|I_n|}{\sigma_0^2} &\\ \geq \quad 2\kappa_n^2\log\left(\frac{1}{\alpha\,|I_n|}\right) + 2\sqrt{n\,|I_n|\left(\kappa_n^4 + 2\frac{\Delta_n^2}{\sigma_0^2}\right)\log\left(\frac{1}{\alpha\,|I_n|}\right)} &\\ + 2\left(1+\kappa_n^2\right)\sqrt{n\,|I_n|\left(\kappa_n^4 + 2\left(1+\kappa_n^2\right)\frac{\Delta_n^2}{\sigma_0^2}\right)\log\left(\frac{1}{\alpha}\right)} & \end{aligned}$$

Then the LR test $\Phi_n(\mathbf{Y}_n) = \mathbf{1}\left\{T_n^{\Delta_n,\kappa_n,|I_n|} > c^*_{\alpha,n}\right\}$ *satisfies*

$$\mathbf{E}_{\mathrm{H}_0}\Phi_n(Y) \leq \alpha \qquad \textit{and} \qquad \mathbf{E}_{\mathrm{H}_1^n}\Phi_n(Y) \geq 1-\alpha.$$

Upper bound

Recall the test statistic

$$T_n^{\Delta_n,\kappa_n,|I_n|}(Y) := \sup_{I_n\in\mathcal{A}_n} S(I_n) = \sup_{I_n\in\mathcal{A}_n} \frac{1}{\sigma_0^2} \sum_{i:\frac{i}{n}\in I_n} \left(Y_i + \frac{\Delta_n}{\kappa_n^2} \right)^2$$

Then we have

- Under H_0 : $S(I_n) \sim \chi^2_{n|I_n|}\left(\frac{n|I_n|\Delta_n^2}{\sigma_0^2\kappa_n^4}\right)$;
- Under H_1^n : $S(I_n) \sim \left(1+\kappa_n^2\right)\chi^2_{n|I_n|}\left(\left(1+\kappa_n^2\right)\frac{n|I_n|\Delta_n^2}{\sigma_0^2\kappa_n^4}\right)$.

Lemma (Laurent, Loubes, Marteau '12)

Let $Z = \sum_{i=1}^{k} b_i X_i$, where $b_i \geq 0$ and $X_i \sim \chi^2_{d_i}\left(a_i^2\right)$ are independent with $d_i \in \mathbb{N}$. Let $\|b\|_\infty = \max_{1\leq i\leq k} |b_i|$. Then for all $x > 0$ we have

$$\mathbb{P}\{Z \leq \mathbf{E}Z - \sqrt{2x\mathbf{Var}Z}\} \leq \exp\left(-x\right),$$

$$\mathbb{P}\{Z > \mathbf{E}Z + \sqrt{2x\mathbf{Var}Z} + 2\|b\|_\infty x\} \leq \exp\left(-x\right),$$

where $\mathbf{E}Z = \sum_{i=1}^k b_i(d_i + a_i^2)$ and $\mathbf{Var}Z = 2\sum_{i=1}^k b_i^2(d_i + 2a_i^2)$.

Three Regimes

We have the upper bound condition :

$$n\,|I_n|\left(\kappa_n^4 + 2\frac{\Delta_n^2}{\sigma_0^2}\right) + \frac{\kappa_n^2\Delta_n^2 n\,|I_n|}{\sigma_0^2}$$
$$\geq \quad 2\kappa_n^2\log\left(\frac{1}{\alpha\,|I_n|}\right) + 2\sqrt{n\,|I_n|\left(\kappa_n^4 + 2\frac{\Delta_n^2}{\sigma_0^2}\right)\log\left(\frac{1}{\alpha\,|I_n|}\right)}$$
$$+ 2\left(1+\kappa_n^2\right)\sqrt{n\,|I_n|\left(\kappa_n^4 + 2\left(1+\kappa_n^2\right)\frac{\Delta_n^2}{\sigma_0^2}\right)\log\left(\frac{1}{\alpha}\right)}$$

that suggests the separation in three regimes based on the asymptotics of κ_n, Δ_n and

$$\frac{\kappa_n^2}{\Delta_n\sigma_0} \asymp ?, \quad n \to \infty.$$

Three Regimes

- **Dominant Signal Regime (DSR) :**
  κ_n^2 vanishes faster than Δ_n :

$$\frac{\kappa_n^2}{\Delta_n} \to 0, \quad n \to \infty$$

- **Equilibrium Regime (ER) :**
  κ_n^2 and Δ_n are of the same order, $\kappa_n \to 0$, $\Delta_n \to 0$ and

$$c := \lim_{n\to\infty} \frac{\kappa_n^2}{\Delta_n/\sigma_0} = \lim_{n\to\infty} \frac{\sigma_n^2}{\Delta_n \sigma_0}, \quad c \in (0, \infty)$$

- **Dominant Variance Regime (DVR) :**
  Δ_n vanishes faster than κ_n^2,

$$\kappa_n \to 0 \quad \text{and} \quad \frac{\kappa_n^2}{\Delta_n} \to \infty, \quad n \to \infty.$$

Dominant Signal Regime

Theorem

Assume that $|I_n| \searrow 0$ *and let* (ε_n) *be any sequence such that* $\varepsilon_n \to 0$, $\varepsilon_n \sqrt{-\log(|I_n|)} \to \infty$. *Let* $\alpha > 0$ *be a significance level. Define the conditions*

(DSR1) $\sigma_n^2/\Delta_n \to 0$ *and* $\sigma_n^2 = o(\varepsilon_n)$, $n \to \infty$

(DSR2) $\sigma_n^2/\Delta_n \to 0$ *and* $\sigma_n^2 = \sigma^2(1 + o(\varepsilon_n))$, $n \to \infty$.

- **Upper bound :** *let the conditions (DSR1) and (DSR2) hold. If*

$$\sqrt{n|I_n|}\Delta_n \gtrsim \left(\sqrt{2}\sigma_0 + \varepsilon_n\right)\sqrt{-\log(|I_n|)},$$

then the LR test $T_n^{\Delta_n,\kappa_n,|I_n|}(\mathbf{Y}_n)$ *satisfies*

$$\lim_{n\to\infty} \mathbf{E}_{\mathrm{H}_0}\Phi_n(Y) \leq \alpha \qquad \text{and} \qquad \lim_{n\to\infty} \mathbf{E}_{\mathrm{H}_1^n}\Phi_n(Y) \geq 1 - \alpha.$$

- **Lower Bound :** *if (DSR1) is satisfied or (DSR2) with* $1/\Delta_n^2 = o(\varepsilon_n)$ *holds, then the sequence* $\mathcal{S}_n$ *satisfying*

$$\sqrt{n|I_n|}\Delta_n \lesssim \left(\sqrt{2}\sigma_0 - \varepsilon_n\right)\sqrt{-\log(|I_n|)},$$

is undetectable.

Dominant Variance Regime

Theorem

Assume that $|I_n| \searrow 0$ *and let* (ε_n) *be any sequence such that* $\varepsilon_n \to 0$, $\varepsilon_n \sqrt{-\log(|I_n|)} \to \infty$. *Define the conditions*

$$\text{(DVR1)} \quad \sigma_0 \Delta_n = \sigma_n^2 \theta_n,\ \Delta_n, \sigma_n, \theta_n \to 0,\ \sigma_n^2 = o(\varepsilon_n),\ \theta_n^2 = o(\varepsilon_n)$$

$$\text{(DVR2)} \quad \sigma_n = \sigma(1 + o(\varepsilon_n)),\ \Delta_n^2 = o(\varepsilon_n) \text{ as } n \to \infty.$$

- **Upper bound :** *if (DVR1) holds with*

$$\sqrt{n|I_n|}\kappa_n^2 \gtrsim (2 + \varepsilon_n)\sqrt{-\log(|I_n|)}$$

or (DVR2) holds with

$$\sqrt{n|I_n|} \gtrsim (C + \varepsilon_n)\sqrt{-\log(|I_n|)}, \quad C := \frac{\sqrt{2\kappa^2 + 1} + 1}{\kappa^2},$$

then the LR test $T_n^{\Delta_n, \kappa_n, |I_n|}(\mathbf{Y}_n)$ *satisfies*

$$\lim_{n\to\infty} \mathbf{E}_{\mathrm{H}_0} \Phi_n(Y) \leq \alpha \quad \text{and} \quad \lim_{n\to\infty} \mathbf{E}_{\mathrm{H}_1^n} \Phi_n(Y) \geq 1 - \alpha.$$

- **Lower Bound :** *if (DVR1) holds, then the sequence* $\mathcal{S}_n$ *is undetectable if*

$$\sqrt{n|I_n|}\kappa_n^2 \gtrsim (2 - \varepsilon_n)\sqrt{-\log(|I_n|)}.$$

If (DVR2) holds, then the sequence $\mathcal{S}_n$ *is undetectable if*

$$\sqrt{n|I_n|} \lesssim (C - \varepsilon_n)\sqrt{-\log(|I_n|)}, \quad C := \left(\frac{\kappa^2}{2} - \frac{1}{2}\log(1 + \kappa^2)\right)^{-1/2}.$$

Equilibrium Regime : vanishing jump in variance

Theorem

Assume that $|I_n| \searrow 0$ *and let* (ε_n) *be any sequence such that* $\varepsilon_n \to 0$, $\varepsilon_n\sqrt{-\log(|I_n|)} \to \infty$. *Let* $\alpha > 0$ *be a given significance level. Let for some* $c > 0$

$$\sigma_n^2 = c\sigma_0\Delta_n(1 + o(\varepsilon_n)) \quad \textit{and} \quad \sigma_n^2 = o(\varepsilon_n).$$

- **Upper bound** *: if* $\mathcal{S}_n$ *satisfies*

$$\sqrt{n\,|I_n|}\Delta_n \gtrsim (C + \varepsilon_n)\sqrt{-\log(|I_n|)}, \quad C := \sqrt{2}\sigma_0\sqrt{\frac{2}{2+c^2}},$$

then for the LR test $T_n^{\Delta_n,\kappa_n,|I_n|}(\mathbf{Y}_n)$ *we have*

$$\lim_{n\to\infty} \mathbf{E}_{\mathrm{H}_0}\Phi_n(Y) \le \alpha \qquad \textit{and} \qquad \lim_{n\to\infty} \mathbf{E}_{\mathrm{H}_1^n}\Phi_n(Y) \ge 1-\alpha.$$

- **Lower bound** *: the sequence* $\mathcal{S}_n$ *that satisfies*

$$\sqrt{n\,|I_n|}\Delta_n \lesssim (C - \varepsilon_n)\sqrt{-\log(|I_n|)}, \qquad C := \sqrt{2}\sigma_0\sqrt{\frac{2}{2+c^2}}.$$

is undetectable.

Equilibrium Regime : constant jump in variance

Theorem

Assume that $|I_n| \searrow 0$ *and let* (ε_n) *be any sequence such that* $\varepsilon_n \to 0$, $\varepsilon_n \sqrt{-\log(|I_n|)} \to \infty$. *Let* $\alpha > 0$ *be a given significance level. Let for some* $c > 0$

$$\Delta_n = \frac{\sigma^2}{c\sigma_0}(1 + o(\varepsilon_n)) \quad \text{and} \quad \sigma_n^2 = \sigma^2(1 + o(\varepsilon_n)).$$

- **Upper bound** *: if* $\mathcal{S}_n$ *satisfies* $\sqrt{n\,|I_n|}\Delta_n \gtrsim (C + \varepsilon_n)\sqrt{-\log(|I_n|)}$ *with*

$$C := \frac{\sqrt{2\kappa^2 + \frac{4\kappa^2}{c^2} + \frac{2\kappa^4}{c^2} + 1 + \frac{2}{c^2}} + \sqrt{1 + \frac{2}{c^2}}}{\kappa^2 + \frac{2\kappa^2}{c^2} + \frac{\kappa^4}{c^2}}.$$

  then for the LR test $T_n^{\Delta_n, \kappa_n, |I_n|}(\mathbf{Y}_n)$ *we have*

$$\lim_{n\to\infty} \mathbf{E}_{\mathrm{H}_0}\Phi_n(Y) \le \alpha \qquad \text{and} \qquad \lim_{n\to\infty} \mathbf{E}_{\mathrm{H}_1^n}\Phi_n(Y) \ge 1 - \alpha.$$

- **Lower bound** *:* $\mathcal{S}_n$ *that satisfies* $\sqrt{n\,|I_n|}\Delta_n \lesssim (C - \varepsilon_n)\sqrt{-\log(|I_n|)}$ *with*

$$C := \frac{1}{\sqrt{\frac{\kappa^2}{2c^2}(\kappa^2 + c^2) - \frac{1}{2}\log(1 + \kappa^2)}}$$

  is undetectable.

- Empirical mean over I_n : $\hat{\Delta}_n := \frac{1}{n|I_n|} \sum_{i:\frac{i}{n}\in I_n} Y_i$.
- Marginal likelihood ratio :

$$L_n\left(I_n, \kappa_n; Y\right) = \left(\kappa_n^2 + 1\right)^{-\frac{n|I_n|}{2}} \exp\left(\frac{\kappa_n^2}{2\sigma_0^2\left(\kappa_n^2+1\right)} \sum_{i:\frac{i}{n}\in I_n} \left(Y_i - \hat{\Delta}_n\right)^2 + \frac{n\left|I_n\right|}{2\sigma_0^2}\hat{\Delta}_n^2\right)$$

- The adaptive test statistic :

$$T_n^{\kappa_n,|I_n|}\left(Y\right) := \sup_{A_n\in\mathcal{A}_n}\left(\frac{\kappa_n^2}{\sigma_0^2\left(\kappa_n^2+1\right)} \sum_{i:\frac{i}{n}\in A_n} \left(Y_i - \hat{\Delta}_n\right)^2 + \frac{1}{\sigma_0^2 n\left|I_n\right|}\Big[\sum_{i:\frac{i}{n}\in A_n} Y_i\Big]^2\right).$$

- The adaptive LR test :

$$\Phi_n^A(\mathbf{Y}_n) = \mathbf{1}\Big\{ T_n^{\kappa_n,|I_n|}\left(Y\right) > c_{n,\alpha}^* \Big\}$$

Adaptation to an unknown change in mean

Theorem

Let $|I_n| \searrow 0$ *and* $\varepsilon_n > 0$ *be such that* $\varepsilon_n \sqrt{-\log(|I_n|)} \to \infty$ *as* $n \to \infty$. *Le* $\alpha \in [0,1]$ *be a significance level. Define the corresponding threshold by*

$$c^*_{n,\alpha} := \frac{\kappa_n^2(n\,|I_n| - 1)}{\kappa_n^2 + 1} + 1 + 2\sqrt{\left[\frac{\kappa_n^4\,(n\,|I_n| - 1)}{(\kappa_n^2+1)^2} + 1\right]\log\left(\frac{1}{\alpha\,|I_n|}\right)} + 2\log\left(\frac{1}{\alpha\,|I_n|}\right).$$

The adaptive test satisfies $\lim_{n\to\infty} \mathbf{E}_{\mathrm{H}_0}\Phi_n(Y)$ *and* $\lim_{n\to\infty} \mathbf{E}_{\mathrm{H}_1^{\mathrm{n}}}\Phi_n(Y) \geq 1-\alpha$ *in any of three following situations :*

- **DSR :** $\kappa_n^2/\Delta_n \to 0$ *and* $1/\Delta_n = o(\varepsilon_n)$ *as* $n \to \infty$ *and*

$$\sqrt{n\,|I_n|}\Delta_n \gtrsim \left(\sqrt{2}\sigma_0 + \varepsilon_n\right)\sqrt{-\log(|I_n|)}.$$

- **ER :** $\sigma_n^2 = c\sigma_0\Delta_n(1 + o(\varepsilon_n))$ *with* $c > 0$ *and* $\sigma_n^2 = o(\varepsilon_n)$ *as* $n \to \infty$ *and*

$$\sqrt{n\,|I_n|}\Delta_n \gtrsim (C + \varepsilon_n)\sqrt{-\log(|I_n|)}, \qquad C := \sigma_0\frac{c + \sqrt{2+3c^2}}{1+c^2}.$$

- **DVR :** $\sigma_0\Delta_n = \sigma_n^2\theta_n$ *with* $\Delta_n, \sigma_n, \theta_n \to 0$ *as* $n \to \infty$ *where* $\sigma_n^2 = o(\varepsilon_n)$, $\theta_n = o(\varepsilon_n)$ *and*

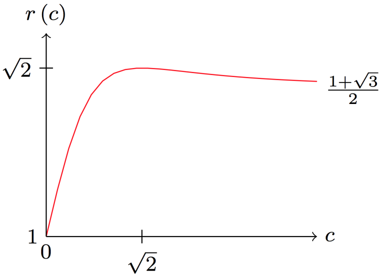
$$\sqrt{n\,|I_n|}\Delta_n \gtrsim \left(\left(1+\sqrt{3}\right)\sigma_0 + \varepsilon_n\right)\sqrt{-\log(|I_n|)}\theta_n.$$

Adaptation : the risk constant

The risk constant loss with respect to the lower bound :

$$r(c) = \begin{cases} 1 & \text{DSR}, \ c = 0, \\ \frac{\sqrt{2+c^2}\left(c+\sqrt{2+3c^2}\right)}{2(1+c^2)} & \text{ER}, \ 0 < c < \infty, \\ \frac{1+\sqrt{3}}{2} & \text{DVR}, \ c = \infty, \end{cases}$$

Figure: The constant loss $r(c)$ with respect to $c = \lim_{n\to\infty} \frac{\sigma_n^2}{\Delta_n \sigma_0}$.



Main results

Three regimes ; vanishing jump in variance $\sigma_0^2\kappa_n^2 := \sigma_n^2 = o(\varepsilon_n^2)$

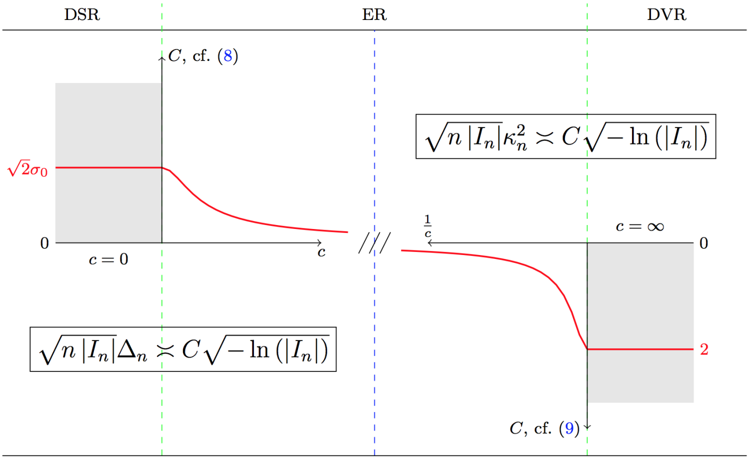
- DSR : $\sigma_n^2/\Delta_n \to 0$ as $n \to \infty$
- ER : $\sigma_0\kappa_n^2/\Delta \to c$ as $n \to \infty$
- DVR : $\sigma_n^2\theta_n = \sigma_0\Delta_n$, where $\Delta_n, \sigma_n, \theta_n \to 0$ as $n \to \infty$

Table: Rates for 3 regimes for $|I_n| \searrow 0$; $\varepsilon_n > 0$ s.t. $\varepsilon_n\sqrt{-\log(|I_n|)} \to \infty$

	Rate	Constant C_n		
		Lower bound	Upper bound	
			Δ_n known	Δ_n unknown
DSR	$\sqrt{n\lvert I_n\rvert}\Delta_n \sim C_n\sqrt{-\log(\lvert I_n\rvert)}$	$\sqrt{2}\sigma_0 - \varepsilon_n$	$\sqrt{2}\sigma_0 + \varepsilon_n$	$\sqrt{2}\sigma_0 + \varepsilon_n$
ER	$\sqrt{n\lvert I_n\rvert}\Delta_n \sim C_n\sqrt{-\log(\lvert I_n\rvert)}$	$\sqrt{2}\sigma_0\sqrt{\frac{2}{2+c^2}} - \varepsilon_n$	$\sqrt{2}\sigma_0\sqrt{\frac{2}{2+c^2}} + \varepsilon_n$	$\sigma_0\frac{c+\sqrt{2+3c^2}}{1+c^2} + \varepsilon_n$
	$\sqrt{n\lvert I_n\rvert}\kappa_n^2 \sim C_n\sqrt{-\log(\lvert I_n\rvert)}$	$2\sqrt{\frac{c^2}{2+c^2}} - \varepsilon_n$	$2\sqrt{\frac{c^2}{2+c^2}} + \varepsilon_n$	$c\frac{c+\sqrt{2+3c^2}}{1+c^2} + \varepsilon_n$
DVR	$\sqrt{n\lvert I_n\rvert}\kappa_n^2 \sim C_n\sqrt{-\log(\lvert I_n\rvert)}$	$2 - \varepsilon_n$	$2 + \varepsilon_n$	$1 + \sqrt{3} + \varepsilon_n$

Risk constants

- DSR : $\sigma_n^2/\Delta_n \to 0$ as $n \to \infty$
- ER : $\sigma_0 \kappa_n^2/\Delta \to c$ as $n \to \infty$
- DVR : $\sigma_n^2 \theta_n = \sigma_0 \Delta_n$, where $\Delta_n, \sigma_n, \theta_n \to 0$ as $n \to \infty$



- DSR : $\sigma_n^2/\Delta_n \to 0,\ n \to \infty$
- ER : $\sigma_0 \kappa_n^2/\Delta \to c,\ n \to \infty$
- DVR : $\sigma_n^2 \theta_n = \sigma_0 \Delta_n$, where $\Delta_n, \sigma_n, \theta_n \to 0,\ n \to \infty$

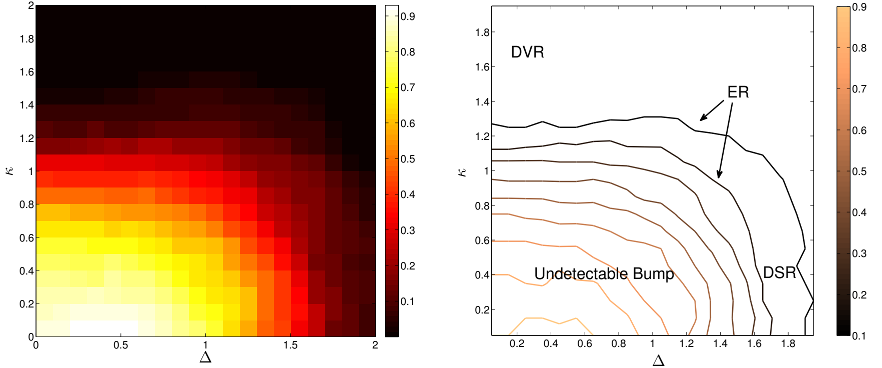


Figure: Type II error for the test of level $\alpha = 0.05$, $n = 100$, $l_n = 0.1$, $\sigma_0 = 1$; $\Delta\theta \in (0, 2]$, $\kappa \in (0, 2)$

Future work

- Adaptation to the unknown change in variance σ_n^2.
- Adaptation to I_n of unknown size.
- Multiple jumps
- Dependent noise
- High-dimensional case : the change in a Gaussian vector
  - Generalization to the case of independent channels with simultaneous jumps in all channels
  - Sparse situation : jumps of the same length in some channels

Bump Detection in Heterogeneous Gaussian Regression

Farida Enikeeva[1], Axel Munk[2] and Frank Werner[2]

[1]LMA–University of Poitiers, [2]Max Planck Institute for Biophysical Chemistry

February 12, 2016

CIRM, Luminy