Non-parametric Bayesian test for monotonicity

JPS 2014, Forges-Les-Eaux

J-B. Salomond

CREST & Université Paris Dauphine

April, 8$^{\text{th}}$ 2014

CREST
Centre de Recherche
en Économie et Statistique

DAUPHINE
UNIVERSITÉ PARIS

Contents

"We balance probability and choose the most likely. It is the scientific use of the imagination."

-Sherlock Holmes, *The hound of Baskervilles*

Introduction

Consider the usual Gaussian regression setting

$$Y_i = f(i/n) + \sigma\epsilon_i$$

where $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0,1)$. We want to test whether f satisfy some **shape constraints** (such as monotonicity).

Introduction

Consider the usual Gaussian regression setting

$$Y_i = f(i/n) + \sigma\epsilon_i$$

where $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0,1)$. We want to test whether f satisfy some **shape constraints** (such as monotonicity).

- Drug response models
- Temperatures anomalies problems
- Economic models when assessing a global trend

Introduction cnt'd

Let $\mathcal{F}$ be the set of **uniformly bounded monotone non increasing functions**. We want to test

$$H_0 : f \in \mathcal{F}, \text{ versus } H_1 : f \notin \mathcal{F}$$

This problem has already been address in the frequentist literature

- Baraud et al. (2005) use projection of the regression function on the set of piecewise constant functions,
- Hall and Heckman (2000) and Ghosal et al. (2000) test negativity of the derivative of the regression function,
- Akakpo et al. (2012) propose a procedure that detect departure from monotonicity via local least concave majorant.

Introduction cnt'd

Remarks

- These methods require in general heavy computations to be used in practice,
- There is only one other (to my knowledge) Bayesian procedure to test for monotonicity.

We thus want to construct a Bayesian testing procedure...

Introduction cnt'd

Remarks

- These methods require in general heavy computations to be used in practice,
- There is only one other (to my knowledge) Bayesian procedure to test for monotonicity.

We thus want to construct a Bayesian testing procedure...

- ... that is easy to implement/use in practice
- ... that is consistent
- ... that achieve the optimal separation rate

Prior construction

We consider the following prior

$$\Pi : \begin{cases} k \sim \pi_k \\ \sigma \sim \pi_\sigma \\ \omega_1 \dots \omega_k | k \sim \otimes_{i=1}^{n} \pi_\omega \end{cases}$$

and f is defined by

$$f_{\omega,k}(\cdot) = \sum_{i=1}^{k} \omega_i \mathbb{I}_{[(i-1)/k, i/k)}(\cdot)$$

We thus have a posterior distribution for f

$$\Pi(f, \sigma | \mathbf{Y}) \propto \Pi(f, \sigma) \mathcal{L}(\mathbf{Y}, f, \sigma)$$

Prior construction

We consider the following prior

$$\Pi : \begin{cases} k \sim \pi_k \\ \sigma \sim \pi_\sigma \\ \omega_1 \dots \omega_k | k \sim \otimes_{i=1}^n \pi_\omega \end{cases}$$

and f is defined by

$$f_{\omega,k}(\cdot) = \sum_{i=1}^{k} \omega_i \mathbb{I}_{[(i-1)/k, i/k)}(\cdot)$$

We thus have a posterior distribution for f

$$\Pi(f, \sigma | \mathbf{Y}) \propto \Pi(f, \sigma) \prod_{i=1}^{n} \sigma^{-1} \phi\left(\frac{Y_i - f(i/n)}{\sigma}\right)$$

Bayes Factor

The standard Bayesian approach will be to compute the Bayes Factor We thus consider the Bayes Factor

$$B_{0,1} = \frac{\Pi(f \in \mathcal{F}|\mathbf{Y}) \diagup \Pi(\mathcal{F})}{\Pi(f \notin \mathcal{F}|\mathbf{Y}) \diagup 1 - \Pi(\mathcal{F})}$$

Bayes Factor

The standard Bayesian approach will be to compute the Bayes Factor We thus consider the Bayes Factor

$$B_{0,1} = \frac{\Pi(f \in \mathcal{F}|\mathbf{Y}) \big/ \Pi(\mathcal{F})}{\Pi(f \notin \mathcal{F}|\mathbf{Y}) \big/ 1 - \Pi(\mathcal{F})}$$

Drawbacks

- The computation of $\Pi(\mathcal{F})$ is difficult and require some strong conditions on the prior
- This approach does not give satisfactory results
- We could not prove consistency for this method

Alternative approach

We thus consider an alternative test

$$H_0^a : \delta(f, \mathcal{F}) \leq \tau \quad \text{versus} \quad H_1^a : \delta(f, \mathcal{F}) > \tau$$

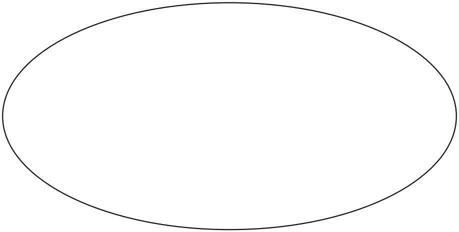
where $\delta(f, \mathcal{F})$ is a discrepancy measure between f and $\mathcal{F}$ and τ a threshold.

Alternative approach

We thus consider an alternative test

$$H_0^a : \delta(f, \mathcal{F}) \leq \tau \quad \text{versus} \quad H_1^a : \delta(f, \mathcal{F}) > \tau$$

where $\delta(f, \mathcal{F})$ is a discrepancy measure between f and $\mathcal{F}$ and τ a threshold.

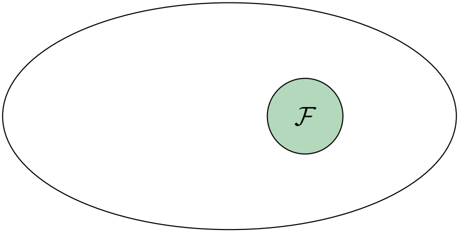


Alternative approach

We thus consider an alternative test

$$H_0^a : \delta(f, \mathcal{F}) \leq \tau \quad \text{versus} \quad H_1^a : \delta(f, \mathcal{F}) > \tau$$

where $\delta(f, \mathcal{F})$ is a discrepancy measure between f and $\mathcal{F}$ and τ a threshold.

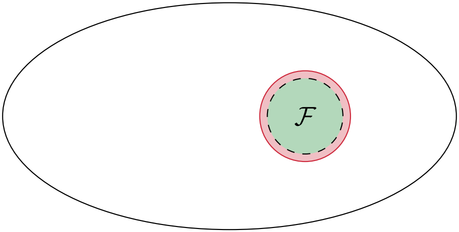


Alternative approach

We thus consider an alternative test

$$H_0^a : \delta(f, \mathcal{F}) \leq \tau \quad \text{versus} \quad H_1^a : \delta(f, \mathcal{F}) > \tau$$

where $\delta(f, \mathcal{F})$ is a discrepancy measure between f and $\mathcal{F}$ and τ a threshold.



Alternative approach

We thus consider an alternative test

$$H_0^a : \delta(f, \mathcal{F}) \leq \tau \quad \text{versus} \quad H_1^a : \delta(f, \mathcal{F}) > \tau$$

where $\delta(f, \mathcal{F})$ is a discrepancy measure between f and $\mathcal{F}$ and τ a threshold. We consider the standard $0-1$ loss and thus define our test as

$$\delta_n^\pi(\tau) := \begin{cases} 0 \text{ if } \Pi\left(\delta(f, \mathcal{F}) \leq \tau | \mathbf{Y}\right) \geq \frac{\gamma_0}{\gamma_0+\gamma_1} \\ 1 \text{ otherwise} \end{cases}$$

Alternative approach cnt'd

We want both test to be asymptotically equivalent

- We thus set $\tau = \tau_n \underset{n \to \infty}{\longrightarrow} 0$
- For some distance d and for all $\rho > 0$

$$\sup_{f \in \mathcal{F}} \mathrm{E}_f^n(\delta_n^\pi(\tau_n)) = o(1)$$

$$\sup_{f, d(f, \mathcal{F}) > \rho} \mathrm{E}_f^n(1 - \delta_n^\pi(\tau_n)) = o(1).$$

- We will ask that the minimum value $\rho = \rho_n$ such that the test is consistent is close to the minimax separation rate.

Construction of the test

Recall

$$f_{\omega,k}(\cdot) = \sum_{i=1}^{k} \omega_i \mathbb{I}_{[\frac{i-1}{k},\frac{i}{k}[}(\cdot)$$

We consider for $\delta(f_{\omega,k}, \mathcal{F})$ the supremum norm between $f_{\omega,k}$ and $\mathcal{F}$, more precisely, we take

$$\delta(f_{\omega,k}, \mathcal{F}) = H(\omega, k) = \max_{1 \leq i < j \leq k} (\omega_j - \omega_i)$$

We now need a good calibration of τ_n such that our procedure is consistent and achieve the **optimal separation rate**. We consider Hölderian alternatives with $\alpha \leq 1$

$$f \in \mathcal{H}(\alpha, L) = \left\{f, [0,1] \to \mathbb{R}, \forall x, y \in [0,1]^2 |f(y) - f(x)| \leq L|y - x|^{\alpha}\right\}$$

Conditions

Consider the model $Y_i = f(i/n) + \sigma\epsilon_i$ where $\epsilon_i \sim \mathcal{N}(0,1)$. We take Π of the form

$$d\Pi(f_{\omega,k},\sigma) = \pi_k(k)\pi_\sigma(\sigma)\prod_{i=1}^{k}\pi_\omega(\omega_i)$$

We get these two conditions

- C1 The densities π_ω and π_σ are continuous, $\pi_\omega(x) > 0$ for all $x \in \mathbb{R}$ and $\pi_\sigma(t) > 0$ for all $t \in \mathbb{R}_*^+$,
- C2 π_k is such that there exist positive constants C_d and C_u such that

$$e^{-C_d k L(k)} \leq \pi_k(k) \leq e^{-C_u k L(k)}$$

where $L(k)$ is either $\log(k)$ or 1.

Main Theorem

Theorem

*Under the assumptions **C1** and **C2**, setting* $\tau_n^k = M_0\sqrt{k\log(n)/n}$ *and* δ_n^π *the testing procedure*

$$\delta_n^\pi = \mathbb{I}\left\{\pi\left(H(\omega,k) > \tau_n^k | Y^n\right) > \frac{\gamma_0}{\gamma_1+\gamma_0}\right\}$$

thus

$$\begin{aligned} \sup_{f\in\mathcal{F}} \mathrm{E}_f^n(\delta_n^\pi) = o(1) \\ \sup_{f,d(f,\mathcal{F})>\rho, f\in\mathcal{H}(\alpha,L)} \mathrm{E}_f^n(1-\delta_n^\pi) = o(1) \end{aligned} \quad (1)$$

for all $\rho > \rho_n(\alpha) = M(n/\log(n))^{-\alpha/(2\alpha+1)}v_n$ *where* $v_n = 1$ *when* $L(k) = \log(k)$ *and* $v_n = \sqrt{\log(n)}$ *when* $L(k) = 1$.

Main Theorem

Remarks

- The conditions on the prior are mild and satisfied for a wide variety of distributions
- Neither the prior nor the hyperparameters depend on the regularity under the alternative,
- the separation rate $\rho_n(\alpha)$ is the minimax separation rate (up to a $\log(n)$ factor)

Contents

"Let us hear the suspicions. I will look after the proof."
-Sherlock Holmes, *The Adventure of Three Students*

Simple choice for the prior

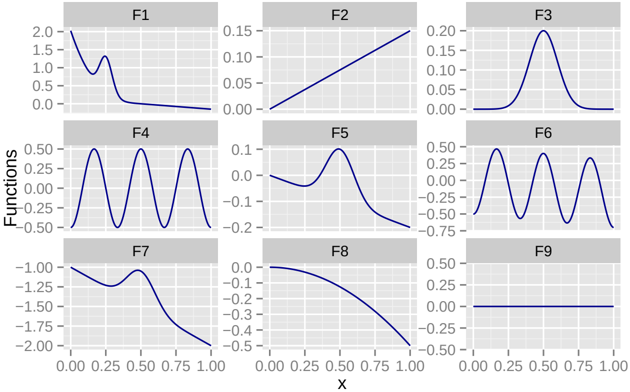
We choose

- $k \sim \mathrm{Geom}(\lambda)$
- $\sigma | k \sim \Gamma^{-1}(a, b)$
- $\omega_i | k, \sigma^2 \sim \mathcal{N}(m, \sigma^2/\mu)$

We also need to calibrate the hyperparameters λ, a, b, m and μ, and the constant M_0 in τ_n^k.

Simulated data

We run our test for nine functions adapted from the frequentist literature



Simulated data

Table : Percentage of rejection for the simulated examples

	f_0	σ^2	Barraud et al. $n = 100$	Akakpo et al. $n = 100$	Bayes Test, n : 100	250	500	1000
H_0	f_1	0.01	99	99	97	100	100	100
	f_2	0.01	99	100	95	100	100	100
	f_3	0.01	99	98	100	100	100	100
	f_4	0.01	100	99	100	100	100	100
	f_5	0.004	99	99	100	100	100	100
	f_6	0.006	98	99	100	100	100	100
	f_7	0.01	76	68	97	100	100	100
H_1	f_8	0.01	-	-	2	0	0	0
	f_9	0.01	-	-	2	3	2	2

Temperature annomalies

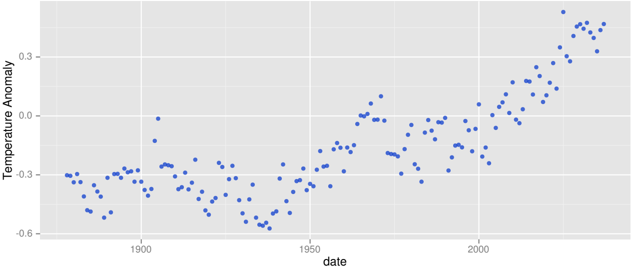


Figure : Temperature Anomalies

We compute $\hat{\Pi}(H(\omega, k) > \tau_n^k | \mathbf{Y}) = 0.98$, and thus **reject** monotonicity of the cuve.

Contents

Conclusion

"It is capital mistake to theorize before one has data."
-Sherlock Holmes, *A Scandal in Bohemia*

We proposed a test for monotonicity in a case where **the standard Bayesain approach fails** Our procedure is

- easy to implement,
- gives good results in practice,
- has good asymptotic properties.

Conclusion

"It is capital mistake to theorize before one has data."

-Sherlock Holmes, *A Scandal in Bohemia*

We proposed a test for monotonicity in a case where **the standard Bayesain approach fails** Our procedure is

- easy to implement,
- gives good results in practice,
- has good asymptotic properties.

extentions

- extend to other types of shape constrains
- study how our procedure behaves for non Gaussian errors

Thank You !

Go Bayes !

References

Akakpo, N., Balabdaoui, F., and Durot, C. (2012). Testing monotonicity via local least concave majorants.

Baraud, Y., Huet, S., and Laurent, B. (2005). Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function. *Ann. Statist.*, 33(1):214–257.

Ghosal, S., Sen, A., and Van der Vaart, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.*, 28(4):1054–1082.

Hall, P. and Heckman, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.*, 28(1):20–39.