Non-parametric Bayesian test for monotonicity
JPS 2014, Forges-Les-Eaux

J-B. Salomond
CREST & Université Paris Dauphine

April, 8th 2014
"We balance probability and choose the most likely. It is the scientific use of the imagination."

-Sherlock Holmes, *The hound of Baskervilles*

1 **Introduction**
   - Prior construction
   - Alternative approach
   - Construction of the test
   - Conditions on the prior and main Theorem

2 **Practical implementation**

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

$$H_0 : f \in \mathcal{F}, \text{ versus } H_1 : f \not\in \mathcal{F}$$

This problem has already been addressed 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 detects departure from monotonicity via local least concave majorant.
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

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

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 \( f \)

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

We consider the following prior

\[ \Pi : \begin{cases} 
    k \sim \pi_k \\ 
    \sigma \sim \pi_\sigma \\ 
    \omega_1 \ldots \omega_k | k \sim \bigotimes_{i=1}^n \pi_\omega 
\end{cases} \]

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 \( f \)

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

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

\[ B_{0,1} = \frac{\prod(f \in \mathcal{F}|Y) / \prod(\mathcal{F})}{\prod(f \notin \mathcal{F}|Y) / 1 - \prod(\mathcal{F})} \]
Bayes Factor

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

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

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
We thus consider an alternative test

\[ H_0^a : \delta(f, F) \leq \tau \ \text{versus} \ \ H_1^a : \delta(f, F) > \tau \]

where \( \delta(f, F) \) is a discrepancy measure between \( f \) and \( F \) and \( \tau \) a threshold.
We thus consider an alternative test

\[ H_0^a : \delta(f, F) \leq \tau \text{ versus } H_1^a : \delta(f, F) > \tau \]

where \( \delta(f, F) \) is a discrepancy measure between \( f \) and \( F \) and \( \tau \) a threshold.
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 \( \tau \) a threshold.
We thus consider an alternative test

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

where \( \delta(f, F) \) is a discrepancy measure between \( f \) and \( F \) and \( \tau \) a threshold.
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 \( \tau \) a threshold. We consider the standard 0–1 loss and thus define our test as

\[
\delta_n^\pi(\tau) := \begin{cases} 
0 & \text{if } \Pi(\delta(f, \mathcal{F}) \leq \tau | Y) \geq \frac{\gamma_0}{\gamma_0 + \gamma_1} \\
1 & \text{otherwise}
\end{cases}
\]
We want both test to be asymptotically equivalent

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

\[
\sup_{f \in \mathcal{F}} E^n_f (\delta_n^\pi (\tau_n)) = o(1)
\]

\[
\sup_{f, d(f, \mathcal{F}) > \rho} E^n_f (1 - \delta_n^\pi (\tau_n)) = o(1).
\]

- 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 \( \tau_n \) such that our procedure is consistent and achieve the optimal separation rate. We consider Hölderian alternatives with \( \alpha \leq 1 \)

\[ f \in \mathcal{H}(\alpha, L) = \{ f, [0, 1] \to \mathbb{R}, \forall x, y \in [0, 1]^2 |f(y) - f(x)| \leq L|y - x|^\alpha \} \]
Consider the model $Y_i = f(i/n) + \sigma \epsilon_i$ where $\epsilon_i \sim N(0, 1)$. We take $\Pi$ 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 $\pi_\omega$ and $\pi_\sigma$ are continuous, $\pi_\omega(x) > 0$ for all $x \in \mathbb{R}$ and $\pi_\sigma(t) > 0$ for all $t \in \mathbb{R}_+^*$,

**C2** $\pi_k$ is such that there exist positive constants $C_d$ and $C_u$ such that

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

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 \( \delta_n^\pi \) the testing procedure*

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

*thus*

\[
\sup_{f \in \mathcal{F}} E_f^n(\delta_n^\pi) = o(1)
\]

\[
\sup_{f, d(f, \mathcal{F}) > \rho, f \in \mathcal{H}(\alpha, L)} E_f^n(1 - \delta_n^\pi) = o(1)
\] (1)

*for all \( \rho > \rho_n(\alpha) = M(n/ \log(n))^{-\alpha/(2\alpha+1)}v_n \) where \( v_n = 1 \) when \( L(k) = 1 \log(k) \) 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)
"Let us hear the suspicions. I will look after the proof."
-Sherlock Holmes, *The Adventure of Three Students*

1 Introduction
   Prior construction
   Alternative approach
   Construction of the test
   Conditions on the prior and main Theorem

2 Practical implementation

3 Conclusion
Simple choice for the prior

We choose

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

We also need to calibrate the hyperparameters $\lambda, a, b, m$ and $\mu$, and the constant $M_0$ in $\tau_n^k$. 
Simulated data

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

Table: Percentage of rejection for the simulated examples

<table>
<thead>
<tr>
<th></th>
<th>$f_0$</th>
<th>$\sigma^2$</th>
<th>Barraud et al. $n=100$</th>
<th>Akakpo et al. $n=100$</th>
<th>Bayes Test, $n$ :</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>$H_0$</td>
<td>$f_1$</td>
<td>0.01</td>
<td>99</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>0.01</td>
<td>99</td>
<td>100</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>$f_3$</td>
<td>0.01</td>
<td>99</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$f_4$</td>
<td>0.01</td>
<td>100</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$f_5$</td>
<td>0.004</td>
<td>99</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$f_6$</td>
<td>0.006</td>
<td>98</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>$f_7$</td>
<td>0.01</td>
<td>76</td>
<td>68</td>
<td>97</td>
</tr>
<tr>
<td>$H_1$</td>
<td>$f_8$</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$f_9$</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
</tbody>
</table>
Temperature anomalies

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

**Extensions**

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

Go Bayes!
References


