A TEST OF GAUSSIANITY BASED ON THE EXCURSION SETS OF A RANDOM FIELD

Elena Di Bernardino & Anne Estrade & José R. León
CNAM-Paris & Université Paris Descartes & Universidad Central de Venezuela

What is the question?

- $X : \mathbb{R}^d \to \mathbb{R}$ is a stationary isotropic random field, smooth enough, with covariance that verifies a decreasing assumption,
- it is observed through some of its excursion sets

$$\{t \in T_i\,;\, X(t) \geq u_i\}\ ,\ i = 1, \ldots, m$$

with $T_1, \ldots, T_m$ cubes in $\mathbb{R}^d$ s.t. $|T_i|$ and $dist(T_i, T_j)$ large, and $u_1, \ldots, u_m$ various levels in $\mathbb{R}$

Question: Is X Gaussian or not?
Main tool: Euler characteristic of the $\{t \in T_i\,;\, X(t) \geq u_i\}$'s

Euler characteristic(χ)?

Important fact: χ is an **additive** functional $\mathcal{C} \subset \mathcal{P}(\mathbb{R}^d) \to \mathbb{Z}$

Heuristic definition for compact $A \subset \mathbb{R}^d$

- $d = 1$: $\chi(A) =$ nber of disjoint intervals in A
- $d = 2$: $\chi(A) =$ nber of connected components $-$ nber of holes in A

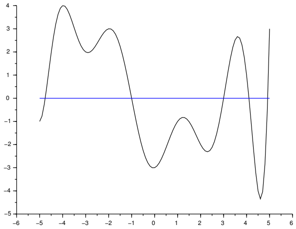
Morse's theory when A=excursion set $\{t \in T \,;\, X(t) \geq u\} \subset \mathbb{R}^d$

$$\chi(A) = \sum_{k=0}^{d} \sum_{\text{face } J \in \partial_k T} \sum_{\ell=0}^{k} (-1)^\ell \mu_\ell(J,u)$$

with $\mu_\ell(J,u) = \#\{t \in J \;;\; X(t) \geq u, X'_{/J}(t) = 0, X'_{/J^c}(t) > 0,$
$\mathrm{ind}(X''_{/J}(t)) = k - \ell\}$

Case $d = 1,\ T = [a, b]$

$$\chi(\{t \in T\,;\, X(t) \geq u\}) = ?$$



Morse's theory

$$= \#\{\text{max. of } X \text{ above } u \text{ in } \overset{\circ}{T}\} - \#\{\text{min. of } X \text{ above } u \text{ in } \overset{\circ}{T}\} + \mathbf{1}_{\{X(a)\geq u, X'(a)<0\}} + \mathbf{1}_{\{X(b)\geq u, X'(b)>0\}}$$

Crossings theory (up-crossings)

$$= \#\{t \in \overset{\circ}{T}\,;\, X(t) = u,\ X'(t) \geq 0\} + \mathbf{1}_{\{X(a)\geq u\}}$$

Case $d \geq 1$, T cube in $\mathbb{R}^d$

$$\chi(\{t \in T\,;\, X(t) \geq u\}) = \sum_{k=0}^{d-1} \sum_{\text{face } J \in \partial_k T} \cdots + \varphi(X, \overset{\circ}{T}, u)$$

where

$$\varphi(X, \overset{\circ}{T}, u) = \sum_{\ell=0}^{d} (-1)^\ell \mu_\ell(\overset{\circ}{T}, u)$$

$$\mu_\ell(\overset{\circ}{T}, u) = \#\{t \in \overset{\circ}{T} : X(t) \geq u, X'(t) = 0, \text{ind } (X''(t)) = d - \ell\}$$

Actually, $\displaystyle\sum_{k=0}^{d-1} \sum_{\text{face } J \in \partial_k T} \cdots = o(|T|)$ as $|T| \to \infty$

$\Rightarrow$ we focus on the "modified" Euler characteristic $\varphi(X, T, u)$

Outline of the talk

signal: X stationary isotropic random field $\mathbb{R}^d \to \mathbb{R}$

observations: $Y_i = \dfrac{\varphi(\{t \in T_i\,;\, X(t) \geq u_i\})}{|T_i|}\ ,\ i = 1, \ldots, m$

Outline

1. Under H0: "X is Gaussian"
2. Alternative hypothesis H1: "X is chi-square" or "X is Poisson"
3. Test

$\oplus$ One-dimensional Monte-Carlo illustrations

$$N = 300,\ |T| = 200,\ X = \text{ stat. process with } \mathbb{E}X(0) = 0,\ \mathrm{Var}X(0) = 1$$

Under H0 (X is Gaussian)

Let $X : \mathbb{R}^d \to \mathbb{R}$ be a Gaussian stationary isotropic random field with $\mathbb{E}X(0) = 0$ and $\mathrm{Var}X(0) = 1$, and let

$$\varphi(X, T, u) = \varphi\left(\{t \in T\,;\, X(t) \geq u\}\right)$$

Theorem [Rice'45, Adler'76]

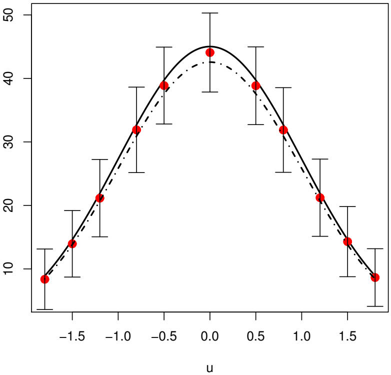
$$\mathbb{E}\left(\frac{\varphi(X, T, u)}{|T|}\right) = (2\pi)^{-(d+1)/2}\,\lambda^{d/2}\,H_{d-1}(u)\,e^{-u^2/2} := C(u, \lambda)$$

with H_k: Hermite polynomial of order k and λ: 2nd spectral moment of X

Note that $H_0(u) = 1$, $H_1(u) = u$ and $\mathrm{Cov}\left(X'(0)\right) = \lambda\, I_d$

Empirical / theoretical mean Euler characteristic

$N = 300$, $|T| = 200$, X Gaussian with $r(t) = e^{-t^2}$ ($\Rightarrow \lambda = 2$)



red dots and boxplots:	empirical mean $\widehat{\mathbb{E}}\varphi(X, T, u_i) = \frac{1}{N}\sum_{n=1}^{N} \varphi(X_n, T, u_i)$
full line:	$u \mapsto \mathbb{E}\varphi(X, T, u) = |T|\, C(u, \lambda)$ with $\lambda = 2$
dashed line:	$u \mapsto |T|\, C(u, \widehat{\lambda})$ with $\widehat{\lambda}$ ... (see next slide)

Estimation of λ

X is still observed through $Y_i = \frac{\varphi(X, T_i, u_i)}{|T_i|}$, $i = 1, \ldots, m$
with levels $u_1, \ldots, u_m$ that are assumed to be **different**

Theorem [Lindgren'74]

Assume $d = 1$. A good estimator of $\lambda^{1/2}$ is given by

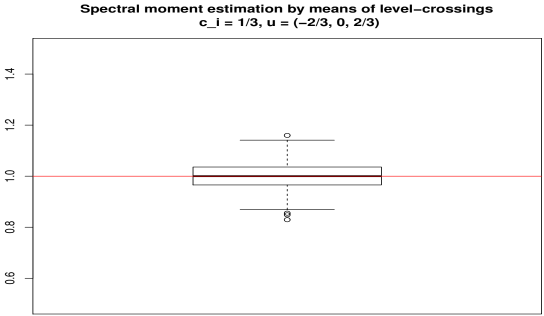
$$\widehat{\gamma} = \sum_{i=1}^{m} c_i \widehat{\gamma}_{u_i} \quad \text{where} \quad \widehat{\gamma}_{u_i} = 2\pi e^{u_i^2/2} \frac{\varphi(X, T_i, u_i)}{|T_i|} \quad \text{and} \quad \sum_{i=1}^{m} c_i = 1$$

Rule of thumb:

$$m = 3 \ , \ c_1 = c_2 = c_3 = \frac{1}{3} \ , \ (u_i) = (-u, 0, u) \text{ with } u = \frac{2}{3}\sigma(X(0))$$

Estimation of λ

Monte-carlo simulations: $N = 300$, X Gaussian with $r(t) = e^{-t^2}$, $|T| = 200$



Boxplot of the ratio between $\lambda^{1/2}$ and its estimation $\widehat{\gamma}$

Second moment of the Euler characteristic

Theorem [DEL]

Under **H0**,

$$\mathbb{E}\left(\frac{\varphi(X,T,u)^2}{|T|}\right) = g(u)\, p_{X'(0)}(0) + \int_{\mathbb{R}^d} \frac{|T \cap (T-t)|}{|T|} G(u,t) p_{X'(0),X'(t)}(0,0)\, dt$$

$$g(u) = \mathbb{E}[\, \mathbf{1}_{[u,\infty)}(X(0))\, |\det(X''(0))|\,]$$

$$G(u,t) = \mathbb{E}[\mathbf{1}_{[u,\infty)^2}(X(0),X(t))\ \det(X''(0)\, X''(t))\, /\, X'(0) = X'(t) = 0]$$

Alternative hypothesis H1: chi-square field

Choose a positive integer s as **degrees of freedom**

Chi-square field

Let $\{X_i(.)\}_{i=1}^s$ be an iid sample of stationary Gaussian fields on $\mathbb{R}^d$ with $\mathbb{E}X_i(0) = 0$ and $\mathrm{Var}X_i(0) = 1$ and let

$$Z^s(.) = \frac{1}{\sqrt{2s}}(\sum_{i=1}^{s} X_i(\cdot)^2 - s)$$

Then

- $\forall t \in \mathbb{R}^d$, $\sum_{i=1}^s X_i(t)^2$ is a χ_s^2 random variable
- Z^s is a stationary random field with $\mathbb{E}Z^s(0) = 0$ and $\mathrm{Var}Z^s(0) = 1$

Chi-square field (2)

Again $\{X_i(.)\}_{i=1}^s$ iid stationary Gaussian fields and

$$Z^s(.) = \frac{1}{\sqrt{2s}}(\sum_{i=1}^{s} X_i(\cdot)^2 - s)$$

- $\forall t \in \mathbb{R}^d$, $\mathbb{E}[Z^s(0)Z^s(t)] = \mathbb{E}[X_i(0)X_i(t)]^2$
- second spectral moment of Z^s = 2 $\times$ second spectral moment of X_i
- as $s \to \infty$, $Z^s \overset{distrib}{\to}$ stationary Gaussian field Z^∞ with $\mathbb{E}Z^\infty(0) = 0$ and $\mathrm{Var}Z^\infty(0) = 1$

Mean Euler characteristic of a chi-square excursion set

Since $Z^s(.) = \frac{1}{\sqrt{2s}}(\chi^2_s(\cdot) - s)$,

$$\mathbb{E}\varphi(Z^s, T, u) = \mathbb{E}\varphi(\chi^2_s, T, s + u\sqrt{2s})$$

Theorem [Worsley'94]

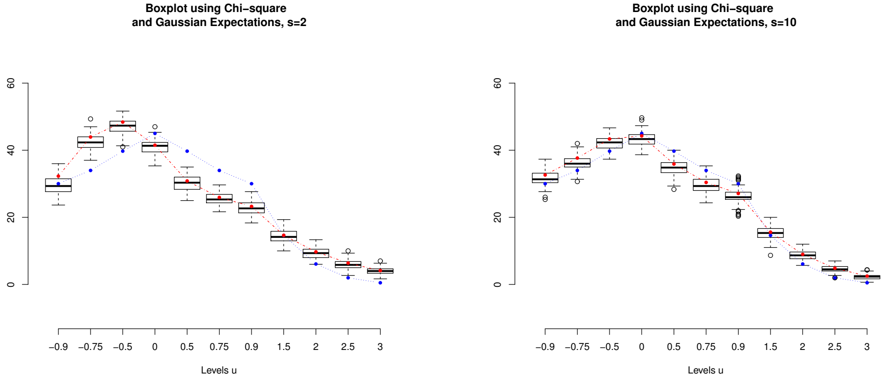
$$\mathbb{E}\left(\frac{\varphi(\chi^2_s, T, u)}{|T|}\right) = \frac{\lambda^{d/2}\, e^{-u/2}}{(2\pi)^{d/2} 2^{(s-2)/2}\Gamma(s/2)}\, u^{(s-d)/2}\, P_{d,s}(u)\, \mathbf{1}_{[0,\infty)}(u)$$

with $P_{d,s}$ a polynomial of degree $d-1$ with integer coefficients and λ second spectral moment of Z^s

Note that $P_{1,s}(u) = 1$ and $P_{2,s}(u) = u - s + 1$

Empirical / theoretical mean Euler characteristic

$N = 300$, $|T| = 200$, X "normalised" i.e. $\mathbb{E}X(0) = 0, \mathrm{Var}X(0) = 1$ and $\lambda_X = 2$



red dots: theoretical $\mathbb{E}\varphi(X, T, u)$ for $X = Z^s$ chi-square field
box plots: empirical mean $\widehat{\mathbb{E}}\varphi(X, T, u)$ for $X = Z^s$ chi-square field
blue dots: theoretical $\mathbb{E}\varphi(X, T, u)$ for X Gaussian field

Another alternative: shot-noise process

Let Φ be a Poisson point process on $\mathbb{R}$ with intensity $\lambda > 0$

Theorem [Biermé-Desolneux'12]

Let S be the shot-noise process based on Φ defined by

$$S(t) = \sum_{\xi \in \Phi} \mathbf{1}_{[0,1]}(t - \xi) \ , \ \ t \in \mathbb{R}$$

Then, for any $u \in \mathbb{R}^{+} \setminus \mathbb{N}$,

$$\mathbb{E}\left(\frac{\varphi(S, T, u)}{|T|}\right) = 2\, e^{-\lambda} \sum_{k \geq 0} \frac{\lambda^{k+1}}{k!} \mathbf{1}_{k < u < k+1}$$

Test of Gaussianity

Null hypothesis
H0: *"X is Gaussian"*
Observations: X is observed through

$$Y_i = \frac{\varphi(X, T_i, u_i)}{|T_i|}\ ,\ i = 1, \ldots, m$$

with $|T_i|$ *and* $dist(T_i, T_j) > 0$ *"large" and various levels* $u_1, \ldots, u_m$

Then, under **H0**,

$$Y_i = C(u_i, \lambda)\ +\ \epsilon_i\ ,\ i = 1, \ldots, m$$

where $C(u,\lambda) = (2\pi)^{-(d+1)/2}\,\lambda^{d/2}\,H_{d-1}(u)\,e^{-u^2/2}$

and the $(\epsilon_i)_{i=1,\ldots,m}$ *are* {
- *centered*
- *variance?* • *distribution?* • *independent?*

Asymptotic variance

Observations on a **large** domain:

$$T^{(N)} = \{Nt \,:\, t \in T\} \text{ with } T \text{ a cube in } \mathbb{R}^d$$

Theorem [DEL]

Under **H0**,

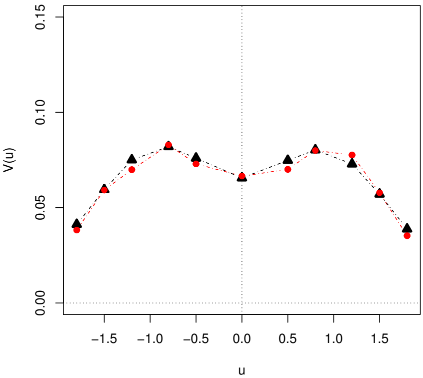
$$\lim_{N\to+\infty} \mathrm{Var}\left(\frac{\varphi(X, T^{(N)}, u)}{|T^{(N)}|^{1/2}}\right) = V(u) \in (0, +\infty)$$

with $V(u) = \int_{\mathbb{R}^d} (G(u,t)\, D(t)^{-1/2} - C(u,\lambda)^2)\, dt + (2\pi\lambda)^{-d/2}\, g(u)$

Moreover, in dimension $d = 1$, explicit formula for $V(u)$ in term of Gaussian integrals

Theoretical / empirical asymptotic variance

Monte-Carlo simulations: $N = 300$, X Gaussian with $r(t) = e^{-t^2}$, $|T| = 200$



black triangles: theoretical $V(u)$

red dots: empirical variance $\widehat{\mathrm{Var}}\left(\frac{\varphi(X,T,u)}{|T|^{1/2}}\right)$

Asymptotic normality

Observations on a **large** domain:

$$T^{(N)} = \{Nt \; : \; t \in T\} \;\; \text{with } T \text{ a cube in } \; \mathbb{R}^d$$

Theorem [EL'15]

Under **H0**,

$$\frac{\varphi(X, T^{(N)}, u) - \mathbb{E}\varphi(X, T^{(N)}, u)}{|T^{(N)}|^{1/2}} \xrightarrow[N\to\infty]{distrib} \mathcal{N}(0, V(u))$$

Corollary: $\dfrac{\varphi(X, T^{(N)}, u)}{|T^{(N)}|} - C(u, \lambda) \; \sim \; \mathcal{N}(0, \dfrac{V(u)}{|T^{(N)}|})$

Disjoint large domains and various levels

Let T_1 and T_2 be two cubes in $\mathbb{R}^d$ s.t. $|T_1| = |T_2|$ and $dist(T_1, T_2) > 0$ and let u_1 and u_2 belong to $\mathbb{R}$ ($u_1 \neq u_2$ or $u_1 = u_2$).

Proposition [DEL]

Let

$$Z_i^{(N)} = \frac{\varphi(X, T_i^{(N)}, u_i) - \mathbb{E}\varphi(X, T_i^{(N)}, u_i)}{|T_i^{(N)}|^{1/2}}, \; i = 1, 2.$$

Then, under **H0**,

$$\left(Z_1^{(N)}, Z_2^{(N)}\right) \underset{N\to\infty}{\overset{distrib}{\longrightarrow}} \mathcal{N}\left(0, \begin{pmatrix} V(u_1) & 0 \\ 0 & V(u_2) \end{pmatrix}\right)$$

Note that $dist(T_1^{(N)}, T_2^{(N)}) \underset{N\to\infty}{\to} \infty$

Statistical model

Observations: $Y_i = \dfrac{\varphi(X, T_i, u_i)}{|T_i|}$, $i = 1, \ldots, m$

with $|T_i|$ and $dist(T_i, T_j) > 0$ "large" and various levels $u_1, \ldots, u_m$

Under **H0** ("X is Gaussian"),

$$Y_i = C(u_i, \lambda) \ + \ \epsilon_i \, , \, i = 1, \ldots, m$$

where $C(u, \lambda) = (2\pi)^{-(d+1)/2} \, \lambda^{d/2} \, H_{d-1}(u) \, e^{-u^2/2}$

and the $(\epsilon_i)_{i=1,\ldots,m}$ are independent $\mathcal{N}(0, V(u_i)/|T_i|)$

Note that $V(u_i)/|T_i|$ is "small"

Chi-square statistics

X is observed through $Y_i = \frac{\varphi(X,T_i,u_i)}{|T_i|}$, $i = 1, \dots, m$
with $|T_i|$ and $dist(T_i, T_j) > 0$ "large" and various levels $u_1, \dots, u_m$

Then, under **H0**,

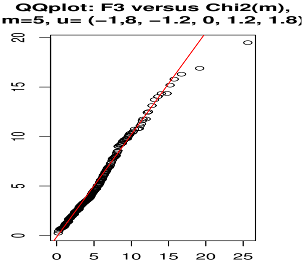
$$F_3 = \sum_{i=1}^{m} \left(\frac{Y_i - C(u_i, \lambda)}{(V(u_i)/|T_i|)^{1/2}} \right)^2 \approx \chi^2_m \text{ distributed}$$

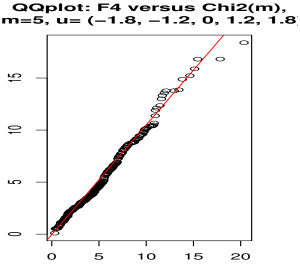
and

$$F_4 = \sum_{i=1}^{m} \left(\frac{\varphi(X, T_i, u_i) - \widehat{\mathbb{E}}\varphi(X, T_i, u_i)}{(\widehat{\mathrm{Var}}\, \varphi(X, T_i, u_i))^{1/2}} \right)^2 \text{ also}$$

Monte-Carlo simulations

$N = 300$, X Gaussian with $r(t) = e^{-t^2}$, $|T| = 200$





$$F_3 = \sum_{i=1}^{m} \frac{(Y_i - C(u_i, \lambda))^2}{V(u_i)/|T_i|} \qquad F_4 = \sum_{i=1}^{m} \frac{(\varphi(X, T_i, u_i) - \widehat{\mathbb{E}}\varphi(X, T_i, u_i))^2}{\widehat{Var}\,\varphi(X, T_i, u_i)}$$

Still in progress...

- Monte-Carlo illustration with 2D simulations
- more examples of Poisson alternative hypothesis (continuous shot-noise, 2D spot-noise)
- real data (sea waves, signals in neurobiology,...)

And to go further

- explicit formula for the asymptotic variance in dimension > 1
- use of other Minkovski functionals (length, area of excursion sets)
- anisotropic fields and test of anisotropy

Conclusion

Take home message

1. shape of $u \mapsto \mathbb{E}\left(\frac{\varphi(X,T,u)}{|T|}\right)$ characterizes X-distribution
2. it can be estimated through a single realization of X on a very large domain by only observing

$$Y_i = \frac{\varphi(X, T_i, u_i)}{|T_i|}, \quad i = 1, \ldots, m$$

with $|T_i|$ and $dist(T_i, T_j) > 0$ "large" and distinct levels $u_1, \ldots, u_m$

Main references

- Adler, Taylor, *Random Fields and Geometry.* Springer (2007)
- Biermé, Desolneux, *A Fourier approach for the level crossings of shot noise processes with jumps.* J. Appl. Probab. (2012)
- Estrade, León, *A CLT for the Euler characteristic of a Gaussian excursion set.* Ann. Probab. (2016)
- Lindgren, *Spectral Moment Estimation by Means of Level Crossings.* Biometrika (1974)
- Worsley, *Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields.* Adv. Appl. Probab. (1994)