Central Limit Theorem for the Euler Characteristic of excursion sets of random Gaussian fields

Anne Estrade.
Université Paris Descartes.
José R. León R.
Universidad Central de Venezuela.
Aline Bonami conference. Orleans 2014

Let us begin with two old formulas.

Area Formula

$$\int_{\mathbb{R}^d} F(x) N_T^X(x) dx = \int_T F(X(t)) |\det \nabla X(t)| dt,$$

where

$$N_T^X(x) = \#\{t \in T : X(t) = x\}, \quad X : D \subset \mathbb{R}^d \to \mathbb{R}^d.$$

and X smooth enough.

Coarea formula

$$\int_{\mathbb{R}^d} F(x)\sigma_{d-j}(\mathcal{C}_Q^X(x))dx = \int_Q F(X(t)) \det(\nabla X(t)\nabla X(t)^T)dt,$$

where

$$\mathcal{C}_Q^X(x) = \{t \in Q : X(t) = x\}, \quad X : D \subset \mathbb{R}^d \to \mathbb{R}^j.$$

whee $d > j$ and X smooth enough. We have denoted as σ_{d-j} the $(d-j)$-dimensional Hausdorff measure. Let us point out that if the function X has maximal rank for all x the $\mathcal{C}_Q^X(x)$ is a dimension $(d-j)$-dimensional manifold.

Considering X as a stationary centered Gaussian random field (smooth enough), we can take expectation in the Area formula. Then Fubini theorem and duality imply that for almost all x

$$\mathbb{E}[N_T^X(x)] = \sigma_d(T)\mathbb{E}[|\det \nabla X(0)|]\frac{e^{-\frac{1}{2}||x||^2}}{\pi^{d/2}}.$$

We have assumed that $X(0)$ is a $N(0, I_d)$. This formula is well known as the Rice Formula. It is true for all $x \in \mathbb{R}^d$ but its proof is a little subtle.

The second moment can be also computed in both cases. But in the case of roots, Rice formula must be written for the second factorial moment. Hence we get

$$\mathbb{E}[N_T^X(x)(N_T^X(x)-1)] =$$

$$\int_T \int_T \mathbb{E}[|\det \nabla X(t) \det \nabla X(s)|| X(t) = X(s) = x] \, p_{t,s}(x,x) \, dtds,$$

where $p_{t,s}(\cdot,\cdot)$ is the density of vector $(X(t), X(s))$. The above formula gives a tool for evaluating the variance of level functionals.

Now we can go to study our problem. **What is the problem?**

- $X : \mathbb{R}^d \to \mathbb{R}$ a Gaussian stationary random field
- $u \in \mathbb{R}$
- T bounded rectangle of $\mathbb{R}^d$
- $A(X, u, T) = \{t \in T\,;\, X(t) \geq u\}$ excursion set above u
- χ denotes the Euler Characteristic (Euler-Poincaré)

What is the asymptotic of $\chi(A(X, u, T))$ when $T \nearrow \mathbb{R}^d$?

Euler-Poincare characteristic : **EPC** denote by χ

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

- in dimension $d = 1$
  $$\chi(A) = \text{number of intervals disjoint of } A$$
- in dimension $d = 2$
  $$\chi(A) = \text{number of connected components- number of hollows of } A$$
- en dim $d = 3$
  $$\chi(A) = \text{ number of connected components}$$
  $$-\text{number of handles} + \text{number of hollows of } A$$

Axiomatic definition

ECP is the only functional χ

- $\chi(\emptyset) = 0$
- $\chi(A) = 1$ if A is basic ($\approx$ compact simply connected)
- $\chi(A \cup B) = \chi(A) + \chi(B) - \chi(A \cap B)$

whenever $A, B, A \cup B, A \cap B$ are the "basic complex" are $\approx$ unions of basic sets)

Constructive definition : definition iterative on the dimension

$$\chi(A) = \begin{cases} \#\{\text{disjoint of } A\} & \text{if } d = 1 \\ \sum_z \chi(A \cap E_z) - \chi(A \cap E_{z-}) & \text{if } d > 1 \end{cases}$$

or

- for $z \in \mathbb{R}$, $E_z = \{t \in \mathbb{R}^d \ : \ t_d = z\} = \mathbb{R}^{d-1} \times \{z\}$
- $\chi(A \cap E_{z-}) = \lim_{y \nearrow z} \chi(A \cap E_y)$
- $\sum_z$ =sum on z such that $\chi(A \cap E_z) \neq \chi(A \cap E_{z-})$

Excursions over a d-dimensional rectangle

Rectangle of $\mathbb{R}^d$: $T = [-N, +N]^d$

$\partial_\ell T$: sets of faces of T of dimension ℓ, $0 \leq \ell \leq d$.

Before studying the asymptotic of χ we will introduce an auxiliary functional that we define in the following.

$X : \mathbb{R}^d \to \mathbb{R}$ **a smooth stationary random field**

We denote by $\varphi(T)$ the term defined by :

$$\varphi(T) = \sum_{0 \leq k \leq d} (-1)^k \mu_k(T)$$

$$\mu_k(T) = \#\{t \in T : X(t) \geq u, \nabla X(t) = 0, \text{index}(\nabla^2 X(t)) = d - k\}$$

Formula “à la Rice” :

(*ref: Adler & Taylor, RFG, chap.11*)

$$\varphi(T) \overset{ps,L^2}{=} \lim_{\varepsilon \to 0} (-1)^d \int_T \det(\nabla^2 X(t)) 1_{[u,\infty)}(X(t)) \delta_\varepsilon(\nabla X(t)) dt$$

δ_ε is an approximate convolution unit, for example.
$\delta_\varepsilon = (2\varepsilon)^{-d} 1_{]-\varepsilon,+\varepsilon[^d}$.
In the cited book it was shown the a.s. convergence. The L^2 convergence requires a sharp bound of the second moment of the number of zeros of the ∇X, and also a systematical use of the convergence dominated theorem.

EPC expectation

X a smooth Gaussian random field **centered and isotropic**
$T = [-N, N]^d$ a rectangle, $|T| = (2N)^d$

$$\mathbb{E}\,\varphi([-N,N]^d) \quad = \quad e^{-u^2/2\sigma^2}\,(2N)^d\,\frac{(\lambda_2)^{d/2}}{\sigma^d(2\pi)^{(d+1)/2}}\,H_{d-1}(u/\sigma)$$

$(H_n)_{n\geq 0}$ are the Hermite's polynomials, $\sigma^2 = Var\,X(0)$, $\lambda_2 = Var\,X_j(0)$

Theorem: for $T = [-N, N]^d$, when $N \nearrow +\infty$,

$$\frac{\varphi(A(X,u,T)) - \mathbb{E}\varphi(A(X,u,T))}{(2N)^{d/2}} \xrightarrow{loi} \text{mean zero Gaussian r.v.}$$

In that follows we denote for ease of notation $\varphi(T) = \varphi(A(X,u,T))$.

If $\mathring{T}$ is the interior of T, it holds $\varphi(T) = \varphi(\mathring{T})$. To see that, we have

$$\varphi(\mathring{T}) = \sum_{k=0}^{d} (-1)^k \mu_k(\mathring{T})$$

with

$$\mu_k(\mathring{T}) = \#\{t \in \mathring{T} : X(t) \geq u, \nabla X(t) = 0, \text{index}(\nabla^2 X(t)) = d-k\}.$$

Let consider first the restriction of ∇X to the boundary set ∂T i.e. $\nabla X_{\partial T} : \partial T \to \mathbb{R}^d$ defined as $\nabla X_{\partial T}(t) = \nabla X(t)$ for $t \in \partial T$.
Bulinskaya Lemma entails

$$\mathbb{P}\{(\nabla X_{\partial T})^{-1}\{0\} = \emptyset\} = 1. \tag{1}$$

Secondly by the conditions of regularity of the field X we get

$$\mathbb{P}\{\omega : \nabla X(t) = \nabla^2 X(t) = 0, \text{ for some } t \in T\} = 0. \tag{2}$$

Since the above conditions entail that with probability one there are no points $t \in \partial T$ satisfying $\nabla X(t) = 0$, then $\mu_k(\overset{\circ}{T}) = \mu_k(T)$ and hence

$$\varphi(\overset{\circ}{T}) = \varphi(T).$$

We are going to show a Hermite expansion for $\varphi(T)$.
Let denote

$$\varphi(\varepsilon, T) = \int_T \det(\nabla^2 X(t)) 1_{[u,\infty)}(X(t)) \delta_\varepsilon(\nabla X(t)) dt.$$

We will look for first an expansion for this random variable.

We can write

$$\varphi(\varepsilon, T) = \int_T G_\varepsilon(\nabla X(t), X(t), \nabla^2 X(t))dt.$$

In the following we identify any symmetric matrix of size $d \times d$ with the $d(d+1)/2$-dimensional vector containing the coefficients above the diagonal and hence consider the map G_ε as

$$\begin{aligned} G_\varepsilon : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^{d(d+1)/2} &\rightarrow \mathbb{R} \\ (x, y, \mathbf{z}) &\mapsto \delta_\varepsilon(x) 1_{[u,\infty)}(y) \det(\mathbf{z}) \end{aligned}$$

Given that the field $(\nabla X(t), X(t), \nabla^2 X(t))$ is stationary it holds that $\nabla X(t)$ is independent from $(X(t), \nabla^2 X(t))$ for each fixed t.

Define $D = 1 + d + \frac{d(d+1)}{2}$ and let Λ be the $D \times D$ matrix s.t. $(\nabla X(t), X(t), \nabla^2 X(t)) = \Lambda Y(t)$ with $Y(t)$ a $N(0, I_D)$ Gaussian vector. It factorizes into

$$\begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix}.$$

We define

$$\widetilde{G}_\varepsilon(y) = G_\varepsilon(\Lambda y) = \delta_\varepsilon(\Lambda_1 \underline{y}) f(\Lambda_2 \bar{y}) = \delta_\varepsilon \circ \Lambda_1(\underline{y}) f \circ \Lambda_2(\bar{y}). \quad (3)$$

Since the map $\widetilde{G}_\varepsilon$ clearly belongs to $\mathbb{L}^2(\mathbb{R}^D, \phi_D(y)dy = \frac{1}{(2\pi)^{D/2}} e^{-\frac{1}{2}||y||^2} dy)$.

The following expansion converges in this space

$$\widetilde{G}_\varepsilon(y) = \sum_{q=0}^{\infty} \sum_{|\mathbf{n}|=q} c(\widetilde{G}_\varepsilon, \mathbf{n}) \widetilde{H_{\mathbf{n}}}(y)$$

where $\mathbf{n} = (n_1, n_2, \ldots, n_D)$, $|\mathbf{n}| = n_1 + n_2 + \cdots + n_D$,

$\widetilde{H_{\mathbf{n}}}(y) = \prod_{1\le j\le D} H_{n_j}(y_j)$ and H_n are the Hermite polynomials.

The **n**-th Hermite coefficient of $\widetilde{G}_\varepsilon$ is given by

$$c(\widetilde{G}_\varepsilon, \mathbf{n}) = \frac{1}{\mathbf{n}!} \int_{\mathbb{R}^D} \widetilde{G}_\varepsilon(y) \widetilde{H_\mathbf{n}}(y) \phi_D(y)\, dz,$$

with $\mathbf{n}! = n_1! n_2! \dots n_D!$. The factorization (3) induces a factorization of the Hermite coefficient into

$$c(\widetilde{G}_\varepsilon, \mathbf{n}) = c(\delta_\varepsilon \circ \Lambda_1, \underline{n})\, c(f \circ \Lambda_2, \bar{n}),$$

with self understanding notations concerning $\mathbf{n} = (\underline{n}, \bar{n})$ and the Hermite coefficients of the maps $\delta_\varepsilon \circ \Lambda_1$ and $f \circ \Lambda_2$.

Recall that $\Lambda Y(t) = (\nabla X(t), X(t), \nabla^2 X(t))$, we can define

$$G_\varepsilon(\nabla X(t), X(t), \nabla^2 X(t)) = G_\varepsilon(\Lambda Y(t)) = \widetilde{G}_\varepsilon(Y(t)).$$

So we can get the Hermite type expansion of $\varphi(\varepsilon, T)$

$$\varphi(\varepsilon, T) = \sum_{q=0}^{\infty} \sum_{\substack{\mathbf{n}=(\underline{n},\bar{n}) \\ |\mathbf{n}|=q}} c(\delta_\varepsilon \circ \Lambda_1, \underline{n})\, c(f \circ \Lambda_2, \bar{n}) \int_T \widetilde{H_{\mathbf{n}}}(Y(t))\, dt. \quad (4)$$

Taking formally the limit as ε goes to 0 yields the expansion for $\varphi(T)$

$$\varphi(T) = \sum_{q=0}^{\infty} \sum_{\substack{\mathbf{n}=(\underline{n},\bar{n}) \\ |\mathbf{n}|=q}} d(\underline{n})\, c(f \circ \Lambda_2, \bar{n}) \int_T \widetilde{H_{\mathbf{n}}}(Y(t))\, dt \qquad (5)$$

where $c(f \circ \Lambda_2, \bar{n})$ is the $\bar{n}$-th Hermite coefficient of $f \circ \Lambda_2$ and

$$d(\underline{n}) = |\det \Lambda_1|^{-1} \frac{1}{(2\pi)^{d/2} \underline{n}!} \prod_{1 \le k \le d} H_{n_k}(0).$$

As a consequence of the L^2 convergence one can show that this formal convergence takes place in $L^2(\Omega)$.

Our result is valid when the compact domain T has the following shape $T = [-N, N]^d$ with N a positive integer, and we let N go to infinity. We will prove that, as $N \to +\infty$, the random variable

$$\zeta([-N,N]^d)) = \frac{\varphi([-N,N]^d) - \mathbb{E}[\varphi([-N,N]^d)]}{(2N)^{d/2}}$$

converges in distribution to a centered Gaussian variable.

We need to prove three things.

- The variance of $\zeta([-N,N]^d))$ converges when $N \to \infty$.
- Each term in the expansion has a Gaussian limit.
- The variance is strictly positive.

Set $\Gamma^Y(v)$ the covariance matrix function of process Y and let introduce the coefficient $\psi(v) = \sup_{ij} |\Gamma^Y_{ij}(v)|$ for $i, j = 1, 2, \ldots D$.

Assume hypothesis **(H)** $\psi \in L^1(\mathbb{R}^d)$, and $\psi(v) \to 0$ when $|v| \to \infty$ and moreover X is three times continuously differentiable.

By using the orthogonality properties of the Chaos we get

$$Var\,\zeta([-N,N]^d)) = \sum_{q=1}^{\infty} \sum_{\substack{\mathbf{n},\mathbf{m}\in\mathbb{N}^D \\ |\mathbf{n}|=|\mathbf{m}|=q}} a(\mathbf{n})a(\mathbf{m})\mathbf{n}!\mathbf{m}!\, R^N(\mathbf{n},\mathbf{m}),$$

where $R^N(\mathbf{n},\mathbf{m})$ is equal to

$$= \int_{[-2N,2N]^d} Cov(\widetilde{H_{\mathbf{n}}}(Y(0)), \widetilde{H_{\mathbf{m}}}(Y(v)))\prod_{1\leq k\leq d}(1-\frac{|v_k|}{2N})\,du.$$

By using the hypothesis **(H)** about ψ and the convergence dominated theorem the above term converges towards

$$R(\mathbf{n},\mathbf{m}) = \int_{\mathbb{R}^d} Cov(\widetilde{H_{\mathbf{n}}}(Y(0)), \widetilde{H_{\mathbf{m}}}(Y(v)))\, du.$$

By using a covariance inequality due to Arcones we can prove that

$$\lim_{Q\to\infty}\lim_{N\to\infty}\sum_{q=Q+1}^{\infty}\sum_{\substack{\mathbf{n},\mathbf{m}\in\mathbb{N}^D\\|\mathbf{n}|=|\mathbf{m}|=q}} a(\mathbf{n})a(\mathbf{m})\mathbf{n}!\mathbf{m}!\,R^N(\mathbf{n},\mathbf{m})=0.$$

Thus the tail of the series of the variance limit tend to uniformly in N. Let define

$$\pi_Q(\zeta([-N,N]^d)):=\sum_{q=Q+1}^{\infty}\sum_{\substack{\mathbf{n}=(\underline{n},\bar{n})\\|\mathbf{n}|=q}} d(\underline{n})\,c(f\circ\Lambda_2,\bar{n})\int_T\widetilde{H_{\mathbf{n}}}(Y(t))\,dt,$$

and $\pi^Q=I-\pi_Q$.

The limit variance results

$$V = \sum_{q=0}^{\infty} \sum_{\substack{\mathbf{n},\mathbf{m}\in\mathbb{N}^D \\ |\mathbf{n}|=|\mathbf{m}|=q}} a(\mathbf{n})a(\mathbf{m})\mathbf{n}!\mathbf{m}!\, R(\mathbf{n},\mathbf{m}) < \infty.$$

Moreover

$$V \geq V_1(u) = f_X(0)\,\lambda_2^d\, H_d(u)^2\,\phi(u)^2,$$

where $f_X(\cdot)$ is the spectral density of X and $\lambda_2 > 0$ is the second spectral moment.

Now we can prove the CLT.

1. We have proved that $\pi_Q(\zeta([-N,N]^d)) = o_{\mathbb{P}}(1)$ for Q large enough.
2. Then we must show that $\pi^Q(\zeta([-N,N]^d))$ satisfies a CLT.
3. By the Nualart & Peccati CLT we only need to prove that each term for $|\mathbf{n}| = q$ in the expansion tends to a Gaussian.
4. But this term can be written

$$\frac{1}{N^{d/2}} \int_{[N,N]^d} \sum_{\substack{\mathbf{n}=(\underline{n},\bar{n}) \\ |\mathbf{n}|=q}} d(\underline{n})\, c(f \circ \Lambda_2, \bar{n}) \widetilde{H_{\mathbf{n}}}(Y(t))\, dt$$

$$:= \frac{1}{N^{d/2}} \int_{[N,N]^d} \mathfrak{G}_q(Y(t))\, dt$$

And this is a particular case of the Breuer & Major Theorem.

Excursions over a d-dimensional rectangle

Let us recall that we are working with a rectangle of $\mathbb{R}^d$:
$T = [-N, +N]^d$
$\partial_\ell T$: sets of faces of T of dimension ℓ, $0 \le \ell \le d$

In particular

- $\partial_d T = \{\overset{\circ}{T}\}$ with $\overset{\circ}{T} =]-N, +N[^d$
- $\partial_0 T$ is the set of vertex of T
- for $J \in \partial_\ell T$, there exists $\sigma(J) \subset \{1, \ldots, d\}$ of cardinal ℓ and a sequence $(\varepsilon_j)_{j \notin \sigma(J)}$ taking values in $\{-1, +1\}$ such that

$$J = \{v \in T \quad ; \quad -N < v_j < N \text{ for } j \in \sigma(J), \\ v_j = \varepsilon_j N \text{ for } j \notin \sigma(J)\}$$

Morse Theorem (*ref: Adler & Taylor, RFG, chap.9*)

Let $X : \mathbb{R}^d \to \mathbb{R}$ be a Morse's function

denoting : $\frac{\partial X}{\partial t_j} = X_j$ and $\frac{\partial^2 X}{\partial t_i \partial t_j} = X_{ij}$

$$\chi(A(X, u, T)) = \sum_{0 \leq \ell \leq d} \sum_{J \in \partial_\ell T} \varphi(J)$$

where for each $J \in \partial_\ell T$,

$$\begin{aligned} \varphi(J) &= \sum_{0 \leq k \leq \ell} (-1)^k \mu_k(J) \\ \mu_k(J) = &\# \quad \{t \in J : X(t) \geq u, \\ & \quad X_j(t) = 0 \text{ si } j \in \sigma(J), \varepsilon_j X_j(t) > 0 \text{ si } j \notin \sigma(J), \\ & \quad \text{index}((X_{ij}(t))_{i,j \in \sigma(J)}) = \ell - k\} \end{aligned}$$

Morse Theorem (continuation)

$$\chi(A(X,u,T)) = \sum_{v\in\partial_0 T}\varphi(\{v\}) + \sum_{0<\ell<d}\ \sum_{J\in\partial_\ell T}\varphi(J) + \varphi(\mathring{T})$$

with $\varphi(\{v\}) = 0$ or 1.

EPC expectation

X a smooth Gaussian random field **centered and isotropic**
$T = [-N, N]^d$ a rectangle, $|T| = (2N)^d$, Adler & Taylor have obtained

$$\begin{aligned} \mathbb{E}\,\chi(A(X,u,[-N,N]^d)) &= \psi(u/\sigma) \\ &+ e^{-u^2/2\sigma^2} \sum_{0<\ell<d} \frac{(2N)^\ell (\lambda_2)^{\ell/2}}{\sigma^\ell (2\pi)^{(\ell+1)/2}} H_{\ell-1}(u/\sigma) \end{aligned}$$

where $\psi(x) = 1 - \Phi(x)$, Φ is the standard Gaussian distribution.
$(H_n)_{n\geq 0}$ are the Hermite's polynomials, $\sigma^2 = Var\,X(0)$, $\lambda_2 = Var\,X_j(0)$

To finish our CLT for the complete Euler Characteristic we must show that for any $\ell = 1, \dots, d-1$ and any face J in $\partial_\ell[N,N]^d$, $|N|^{-d} Var(\varphi(J) - \mathbb{E}\varphi(J))$ vanishes as $[N,N]^d$ grows to $\mathbb{R}^d$.

The proof consists in representing the faces as a sum of level sets then construct a Hermite expansion for each, similar at the obtained for φ and then bounded the variance by $O(N^\ell)$.

CLT

Resuming we can establish our main result.

Theorem

Let $T = [-N, N]^d$ *and* X *as before* $+$ *hyp* **(H)**

$$\frac{\chi(A(X,T,u)) - \mathbb{E}\chi(A(X,T,u))}{(2N)^{d/2}} \xrightarrow{loi} N(0,V).$$

Perspectives

- speed of convergence in CLT (distances of Kolmogorov or Wasserstein)
- CLT functional for $u \mapsto \chi(A(X, T, u))$ whenever $T \nearrow \mathbb{R}^d$
- measurement of EPC for real data (see waves, mammographs)
- test of Gaussianity (versus ?) We can compute the EPC from the observations and compare it with the theoretical expected result. In the following slide we show the examples for $d = 2$ and $d = 3$.

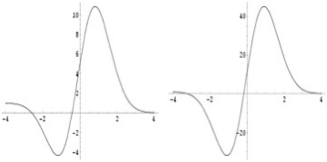

Fig. 2 Two EEC curves $\mathbb{E}\{\varphi(A_u)\}$ for Gaussian fields in two dimensions for different values of the second spectral moment λ_2. The horizontal axis gives values of u and the vertical axis the EEC.

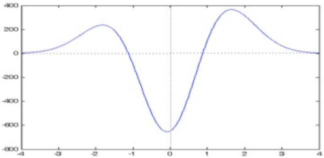

Fig. 3 EEC for a Gaussian field in three dimensions. Axes as for Figure 2.

14

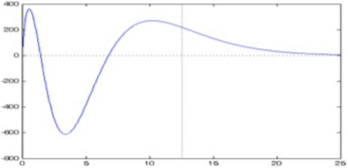

Fig. 4 EEC for a χ^2_5 field in three dimensions. Axes as for Figure 2.

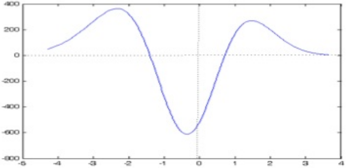

Fig. 5 EEC for a 'normalised' χ^2_5 field in three dimensions. Axes as for Figure 2.

- Adler R., Taylor J. *Random Fields and Geometry.* Springer Monographs in Mathematics. Springer (2007).
- Arcones M. A. *Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors.* Ann. of Prob. Vol 22. N° 4 2242-2274. (1994).
- Azaïs J-M., Wschebor M. *Level Sets and Extrema of Random Processes and Fields.* Wiley (2009).
- Breuer J., Major P. *Central limit theorems for non-linear functionals of Gaussian Fields.* J. Mult. An. 13, 425-444. (1983).