Testing One Hypothesis Multiple Times

Samuel Davenport

BDI

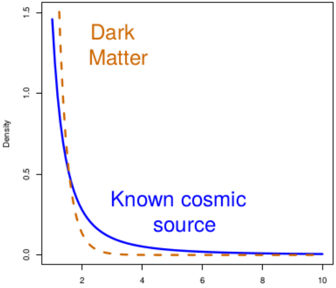
November 12, 2018

Dark Matter - Problem Set up

Non-nested models comparison in physics

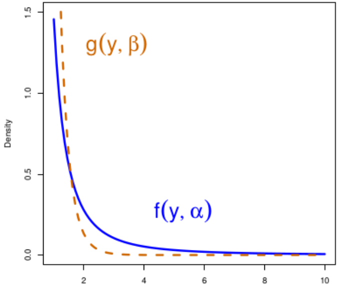
Goal

We would like to distinguish known astrophysics from new signals.



- **E.g.,** Dark Matter.
- We wish to distinguish a dark matter signal from a "fake" signal that mimics it.

The statistical problem



- The model for the know cosmic source is $f(y, \alpha)$;
- The model for the new source is $g(y, \beta)$;
- $f \not\equiv g$ for any α and β.

Is f sufficient to explain the data, or does g provide a better fit?

Problem

f and g are non-nested.

Mathematical Formulation

Given an observation Y if there is a signal present then Y has probability η of being distributed according to the density function $g(y,\theta)$ where $\theta \in \Theta$ for some parameter space Θ. So Y has density:

$$(1-\eta)f(y,\gamma) + \eta g(y,\theta)$$

Mathematical Formulation

Given an observation Y if there is a signal present then Y has probability η of being distributed according to the density function $g(y, \theta)$ where $\theta \in \Theta$ for some parameter space Θ. So Y has density:

$$(1-\eta)f(y,\gamma) + \eta g(y,\theta)$$

and we wish to test the hypothesis

$$H_0 : \eta = 0 \quad \text{versus} \quad H_1 : \eta > 0.$$

Mathematical Formulation

Given an observation Y if there is a signal present then Y has probability η of being distributed according to the density function $g(y, \theta)$ where $\theta \in \Theta$ for some parameter space Θ. So Y has density:

$$(1 - \eta) f(y, \gamma) + \eta g(y, \theta)$$

and we wish to test the hypothesis

$$H_0 : \eta = 0 \quad \text{versus} \quad H_1 : \eta > 0.$$

However θ is unknown and we wish to search over some uncountable region Θ. Eg $\Theta \subset \mathbb{R}^D$ for some D.

The Test

Suppose that we observe $Y_1, \dots, Y_n$ iid with the same distribution as Y. Suppose that there exist random fields $W_n = W_n(Y_1, \dots, Y_n)$ s.t

$$W_n : \Theta \longrightarrow \mathbb{R}$$

such that W_n has a known (potentially asymptotic) distribution under H_0 which we denote by W:

$$W_n \xrightarrow{d} W.$$

The Test

Suppose that we observe $Y_1, \ldots, Y_n$ iid with the same distribution as Y. Suppose that there exist random fields $W_n = W_n(Y_1, \ldots, Y_n)$ s.t

$$W_n : \Theta \longrightarrow \mathbb{R}$$

such that W_n has a known (potentially asymptotic) distribution under H_0 which we denote by W:

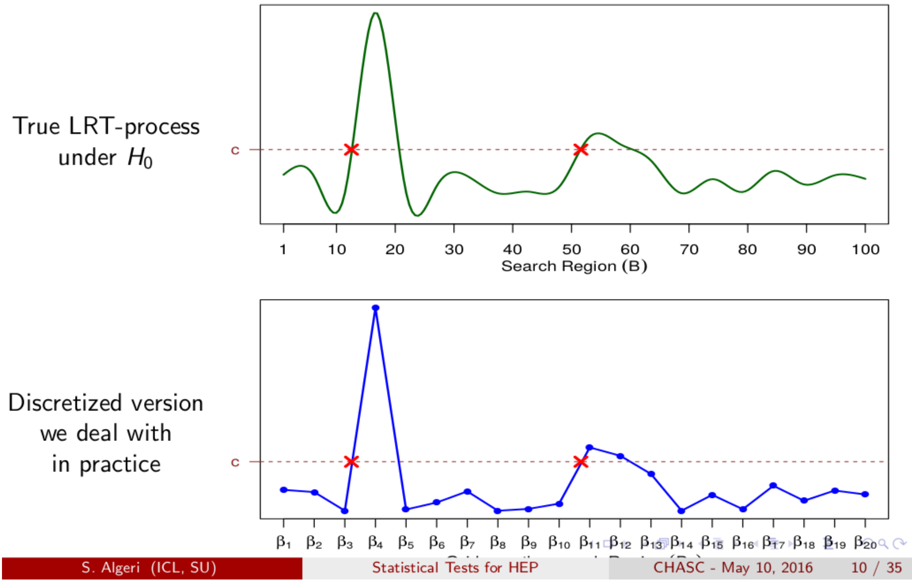
$$W_n \xrightarrow{d} W.$$

Then in order to test the null hypothesis we can consider the global p-value:

$$\mathbb{P}\left(\sup_{\theta \in \Theta}\{W_n(\theta) > c\}\right) \approx \mathbb{P}\left(\sup_{\theta \in \Theta}\{W(\theta) > c\}\right)$$

for some threshold u and large enough n.

Suppose that $\Theta = [L, U]$ is a closed interval and that W is a stationary process. Let M_u be the number of upcrossings of u by W.



Claim

If W is a 1D process, and $\Theta = [L, U]$ and M_u is the number of upcrossings of u by W, then

$$\mathbb{P}\left(\sup_{\theta \in \Theta} W(\theta) > u\right) \leq \mathbb{P}(W(L)) > u) + \mathbb{E}M_u.$$

Claim

If W is a 1D process, and $\Theta = [L, U]$ and M_u is the number of upcrossings of u by W, then

$$\mathbb{P}\left(\sup_{\theta \in \Theta} W(\theta) > u\right) \leq \mathbb{P}(W(L)) > u) + \mathbb{E}M_u.$$

Proof.

$$\text{We have: } \mathbb{P}\left(\sup_{\theta \in \Theta} W(\theta) > u\right) = \mathbb{P}(W(t)) > u, \text{ some } t \in [L, U])$$
$$= \mathbb{P}(W(L) > u \text{ or } M_u \geq 1)$$
$$\leq \mathbb{P}(W(L) > u) + \mathbb{P}(M_u \geq 1)$$
$$\leq \mathbb{P}(W(L) > u) + \mathbb{E}M_u$$

We have used the fact that $\mathbb{E}M_u = \sum_{n \geq 1} \mathbb{P}(M_u) \geq n)$ to justify the 2nd inequality. □

Multi-Dimensional Parameter Spaces

Instead of upcrossings let M_u be the number of local maxima of W (note that N_c is a random variable) including maxima that lie on the boundary. Assume that W is a smooth process (need specific condition here), then

$$\mathbb{P}\left(\sup_{\theta\in\Theta} W(\theta) > u\right) = \mathbb{P}(M_u \geq 1) \leq \mathbb{E}[M_u].$$

because W exceeds u if and only if there is at least one local maxima.

Multi-Dimensional Parameter Spaces

Instead of upcrossings let M_u be the number of local maxima of W (note that N_c is a random variable) including maxima that lie on the boundary. Assume that W is a smooth process (need specific condition here), then

$$\mathbb{P}\left(\sup_{\theta \in \Theta} W(\theta) > u\right) = \mathbb{P}(M_u \geq 1) \leq \mathbb{E}[M_u].$$

because W exceeds u if and only if there is at least one local maxima. This is easy to see.

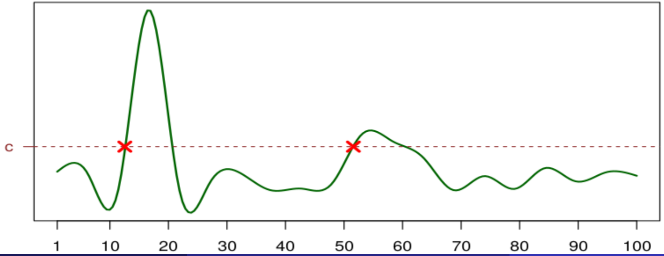
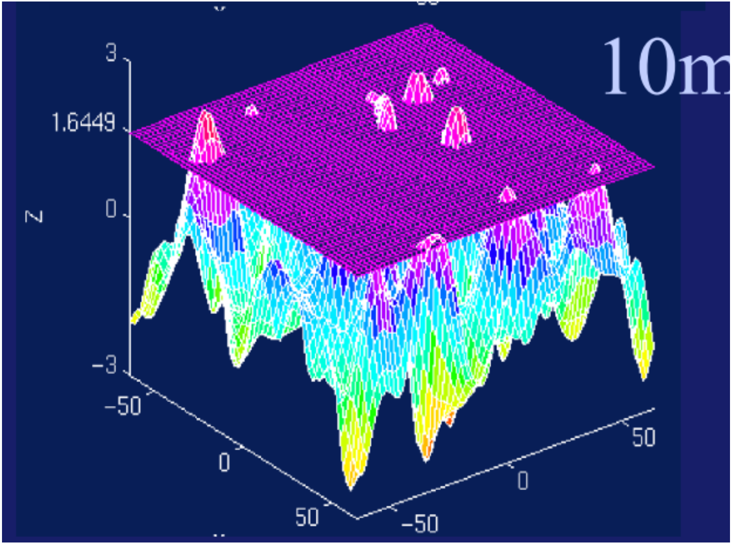


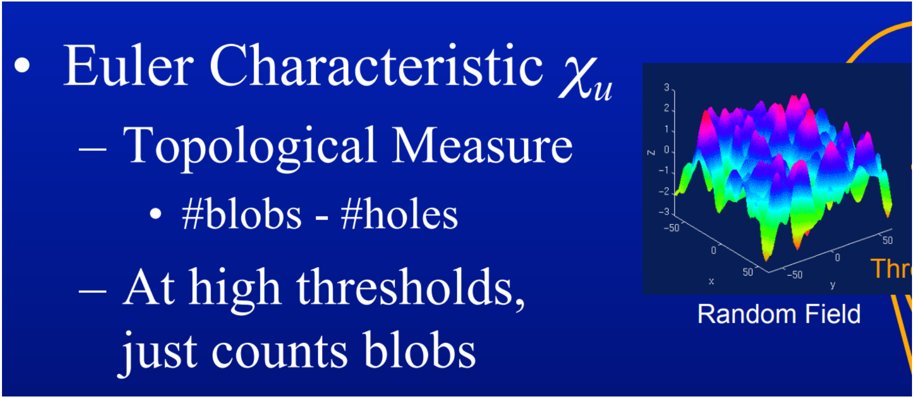
Illustration in 2D



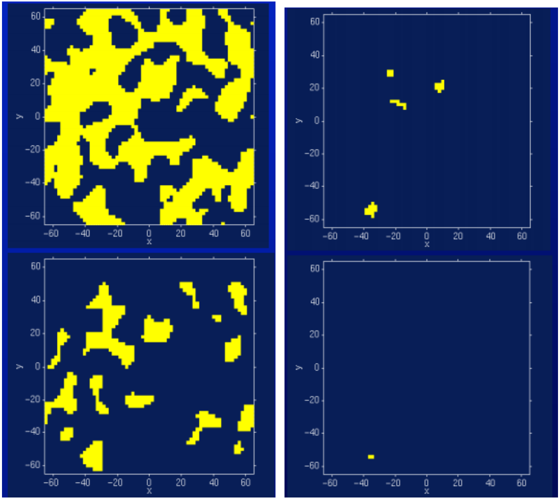
Random Field Theory

The Euler Characteristic

$\mathbb{E}[M_u]$ is difficult to estimate and requires us to be clever.
To do so we introduce a topological quantity called the Euler Characteristic $\chi_u$which looks at the excursion set and calculates the number of blobs minus the number of holes.



At High Thresholds, Euler Char is the number of Maxima



Using the Euler Characteristic

Let $\mathcal{A}_u$ be the excursion set when the threshold is u ie

$$\mathcal{A}_u = \{\theta \in \Theta : W(\theta) \geq u\}.$$

Using the Euler Characteristic

Let $\mathcal{A}_u$ be the excursion set when the threshold is u ie

$$\mathcal{A}_u = \{\theta \in \Theta : W(\theta) \geq u\}.$$

$\chi_u = \chi(\mathcal{A}_u)$ is the number of blobs of $\mathcal{A}_u$ minus the number of its holes. Then for high thresholds, $M_u = \chi_u$. So

$$\mathbb{E}[M_u] = \mathbb{E}[\chi(\mathcal{A}_u)].$$

Using the Euler Characteristic

Let $\mathcal{A}_u$ be the excursion set when the threshold is u ie

$$\mathcal{A}_u = \{\theta \in \Theta : W(\theta) \geq u\}.$$

$\chi_u = \chi(\mathcal{A}_u)$ is the number of blobs of $\mathcal{A}_u$ minus the number of its holes. Then for high thresholds, $M_u = \chi_u$. So

$$\mathbb{E}[M_u] = \mathbb{E}[\chi(\mathcal{A}_u)].$$

So

$$\mathbb{P}\left(\sup_{\theta\in\Theta} W(\theta) > u\right) = \mathbb{P}(M_u \geq 1) \leq \mathbb{E}[M_u] \approx \mathbb{E}[\chi(\mathcal{A}_u)].$$

LKCs

Disclaimer: Complicated Equation Alert!

LKCs

Disclaimer: Complicated Equation Alert!

Theorem

Under certain regularity conditions,

$$\mathbb{E}[\chi(\mathcal{A}_u)] = \sum_{d=0}^{D} \mathcal{L}_d \rho_d(u)$$

$\rho_d : \mathbb{R} \longrightarrow \mathbb{R}$ are the euler-characteristic densities

$\mathcal{L}_d$ are the Lipshitz Killing Curvatures (LKCs) which depend on Θ and on the covariance structure and partial derivatives of W.

Explaining the LKC formula

The Euler Characteristic densities are dependent only on the marginal distribution of the field. So these are typically easy to compute.

For instance if W is a gaussian random field ie $W(\theta) \sim N(0,1)$ for each θ, then:

$$\rho_0(u) = 1 - \Phi(u), \rho_1(u) = \exp(-2u^2)/2\pi$$

In general $\rho_0(u) = \mathbb{P}(W(\theta) > u)$.

The LKCs are more complicated and we need to estimate them.

Solving with regression, see (Adler, Bartz, Kou, & Monod, 2017)

Theorem

Let $u \in \mathbb{R}$ *and define* $u_1 \neq u_2 \neq \cdots \neq c_D$ *all in* $\mathbb{R}$*, then*

$$\mathbb{E}[\chi(\mathcal{A}_u)] = \mathcal{L}_0\rho_0(u) + \sum_{d=1}^{D} \mathcal{L}_d^*\rho_d(u)$$

where the $\mathcal{L}_d^*$ *are the solutions of the system of* D *equations:*

$$\mathbb{E}[\chi(\mathcal{A}_{u_1})] - \mathcal{L}_0\rho_0(u_1) = \sum_{d=1}^{D} \mathcal{L}_d\rho_d(u_1)$$

$$\vdots$$

$$\mathbb{E}[\chi(\mathcal{A}_{u_D})] - \mathcal{L}_0\rho_0(u_D) = \sum_{d=1}^{D} \mathcal{L}_d\rho_d(u_D)$$

Estimating the Euler Characteristic

Lattice Approximations

To do this we need to calculate $\mathbb{E}[\chi(\mathcal{A}_{u_k})]$ for $k = 1, \dots, D$.
Let us now consider lattice approximations. Suppose that

$$\Theta = [0,1]^D = [0,1] \times \cdots \times [0,1] \subset \mathbb{R}^D$$

We can only evaluate $W(\theta)$ at a finite number of values of θ so given an integer n suitably large, consider the finite subset:

$$\left\{0, 2^{-n}, 2 * 2^{-n}, 3 * 2^{-n}, \dots, (2^n - 1) * 2^{-n}, 1\right\} \subset [0,1]$$

that divides $[0,1]$ into $2^n + 1$ points.

Lattice Approximations

To do this we need to calculate $\mathbb{E}[\chi(\mathcal{A}_{u_k})]$ for $k = 1, \dots, D$. Let us now consider lattice approximations. Suppose that

$$\Theta = [0,1]^D = [0,1] \times \cdots \times [0,1] \subset \mathbb{R}^D$$

We can only evaluate $W(\theta)$ at a finite number of values of θ so given an integer n suitably large, consider the finite subset:

$$\left\{0, 2^{-n}, 2 * 2^{-n}, 3 * 2^{-n}, \dots, (2^n - 1) * 2^{-n}, 1\right\} \subset [0,1]$$

that divides $[0,1]$ into $2^n + 1$ points. (Illustrate with line on board.) And define the lattice:

$$L_n = \left\{0, 2^{-n}, 2 * 2^{-n}, 3 * 2^{-n}, \dots, (2^n - 1) * 2^{-n}, 1\right\}^D \subset [0,1]^D = \Theta$$

Euler Characteristic Lattice Approximation

We can't calculate W at every $\theta \in \Theta$ as this is infinitely many points, so we will instead calculate $W(\theta)$ for

$$\theta \in L_n = \left\{0, 2^{-n}, 2 * 2^{-n}, 3 * 2^{-n}, \dots, (2^n - 1) * 2^{-n}, 1\right\}^D$$

Definition

Define the lattice excursion set $\tilde{\mathcal{A}}_u = \{\theta \in L_n : W(\theta) \geq u\}$

(Note that we will drop the dependence on n when talking about excursion sets.) Recall that

$$\mathcal{A}_u = \{\theta \in \Theta : W(\theta) \geq u\}.$$

So the idea is that we can use $\tilde{\mathcal{A}}_u$ to approximate $\mathcal{A}_u$

Slide to Skip: (General Notation)

The previous slide was a simplified version for understanding, this can be written more generally in the form: $\mathbb{E}[\chi(\mathcal{A}_{u_k})]$ for $k = 1, \ldots, D$.
Let us now consider lattice approximations. Suppose that

$$\Theta = \Theta_1 \times \cdots \times \Theta_D \subset \mathbb{R}^D$$

We can only evaluate $W(\theta)$ at a finite number of values of θ so let $P_d \subset \Theta_d$ be a finite subset. Then we will calculate $W(\theta)$ for

$$\theta \in P = P_1 \times \cdots \times P_D$$

Let $\tilde{\mathcal{A}}_u = \{\theta \in P : W(\theta) \geq u\}$
Then

$$\chi(\tilde{\mathcal{A}}_u) \approx \chi(\mathcal{A}_u)$$

so long as our lattice P is fine enough.
Can say something like wlog assume an equally spaced lattice (with spacing k) as this doesn't change the euler characteristic. In fact wlog $P_i \subset \mathbb{Z}$. But more for inclusion in a paper rather than a talk.

Lattice Graphs

Definition

Given a set of vertices V and a set of edges E connecting some of the vertices define the graph (V, E) to be the collection of these vertices and edges.

Lattice Graphs

Definition

Given a set of vertices V and a set of edges E connecting some of the vertices define the graph (V, E) to be the collection of these vertices and edges.

Definition

Define the **lattice graph** to be the graph with $V = L_n$ and such that given two vertices $v = (v_1, \dots, v_D)$ and $w = (w_1, \dots, w_D) \in L_n$ the edge vw is in E if $v = w$ except for some $i \in \{1, \dots, D\}$, we have $w_i = v_i + 2^{-n}$ or $v_i = w_i + 2^{-n}$.

Lattice Graphs

Definition

Given a set of vertices V and a set of edges E connecting some of the vertices define the graph (V, E) to be the collection of these vertices and edges.

Definition

Define the **lattice graph** to be the graph with $V = L_n$ and such that given two vertices $v = (v_1, \ldots, v_D)$ and $w = (w_1, \ldots, w_D) \in L_n$ the edge vw is in E if $v = w$ except for some $i \in \{1, \ldots, D\}$, we have $w_i = v_i + 2^{-n}$ or $v_i = w_i + 2^{-n}$.

Definition

Let the **excursion graph** be the subgraph of the lattice graph corresponding to the vertex set $V = \tilde{\mathcal{A}}_u = \{\theta \in L_n : W(\theta) \geq u\}$. Such that for vertices v and w, the edge vw is in edge set E iff vw lies in $\mathcal{A}_u$. We denote this graph by $\mathcal{G}$.

Example to understand the notation

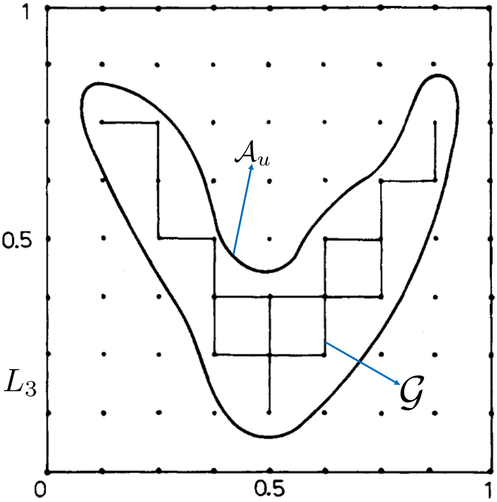


Image generated by (Adler, 1981).

$$\mathcal{A}_u = \left\{\theta \in [0,1]^2 : W(\theta) \geq u\right\}$$

$$L_3 = \left\{0, \frac{1}{8}, \frac{2}{8}, \frac{3}{8}, \ldots, \frac{7}{8}, 1\right\}$$

$$\tilde{\mathcal{A}}_u = \{\theta \in L_3 : W(\theta) \geq u\}$$

$\mathcal{G}$ is the graph of the vertices in $\tilde{\mathcal{A}}_u$ and the edges that connect them and which lie in $\mathcal{A}_u$.

Euler Characteristic on a Lattice

Theorem

For a fine enough lattice (ie large enough n)

$$\chi(\mathcal{A}_u) \approx \chi(\tilde{\mathcal{A}}_u)$$

see (Adler, 1981), chapter 5.5.

Note that we write $\chi(\tilde{\mathcal{A}}_u)$ to denote $\chi(\mathcal{G})$.

Euler Characteristic on a Lattice

Theorem

For a fine enough lattice (ie large enough n)

$$\chi(\mathcal{A}_u) \approx \chi(\tilde{\mathcal{A}}_u)$$

see (Adler, 1981), chapter 5.5.

Note that we write $\chi(\tilde{\mathcal{A}}_u)$ to denote $\chi(\mathcal{G})$.

Theorem

Given a subgraph A of the lattice graph,

$$\chi(A) = \sum_{d=0}^{D} (-1)^d R_d(A)$$

where $R_d(A)$ is the number of cubes of dimension d in the subgraph A.

2D example

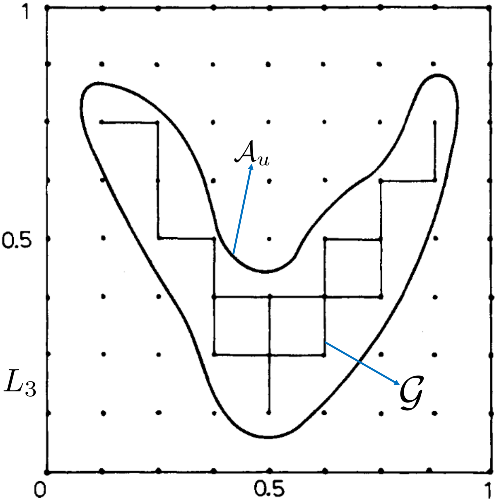


Image generated by (Adler, 1981).

The number of vertices here is 18, the number of edges is 20 and the number of squares is 3!

2D example

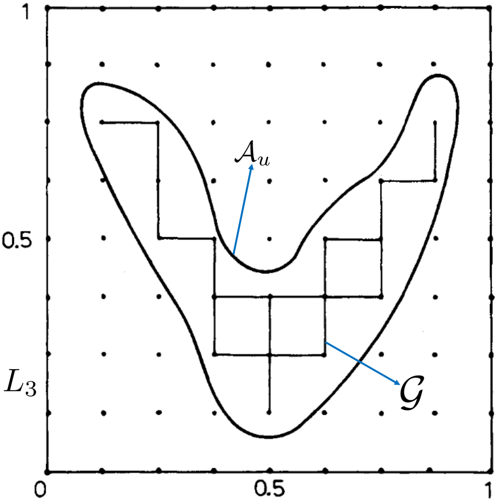


Image generated by (Adler, 1981).

The number of vertices here is 18, the number of edges is 20 and the number of squares is 3!And so

$$\chi(A) = \sum_{d=0}^{2} (-1)^d R_d(A)$$
$$= 18 - 20 + 3 = 1.$$

which is the number of connected components on this graph.

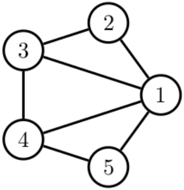
Graph Theory

Definition

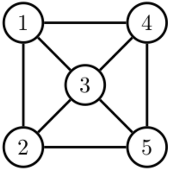
Given a graph: (V, E) with vertices V and edges E, we define a **clique** of size m to be a set of vertices $\{v_1, \ldots, v_m\} \subset V$ such that $v_i v_j \in E$ for all i and j.

Definition

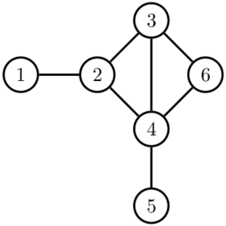
Given a graph: (V, E) with vertices V and edges E, we define a **clique** of size m to be a set of vertices $\{v_1, \ldots, v_m\} \subset V$ such that $v_i v_j \in E$ for all i and j.



(a)

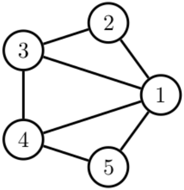


(b)

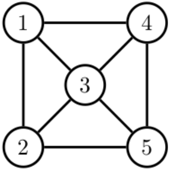


Definition

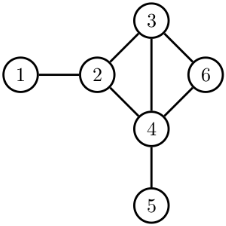
Given a graph: (V, E) with vertices V and edges E, we define a **clique** of size m to be a set of vertices $\{v_1, \ldots, v_m\} \subset V$ such that $v_i v_j \in E$ for all i and j.



(a)



(b)



2D example

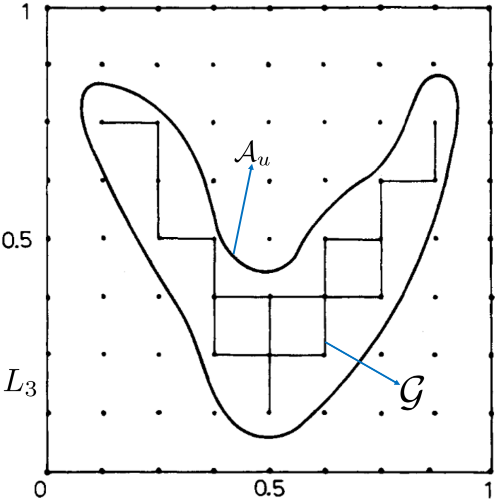


Image generated by (Adler, 1981).

In this graph the number of cliques of size $2^0 = 1$ is the number of vertices.

2D example

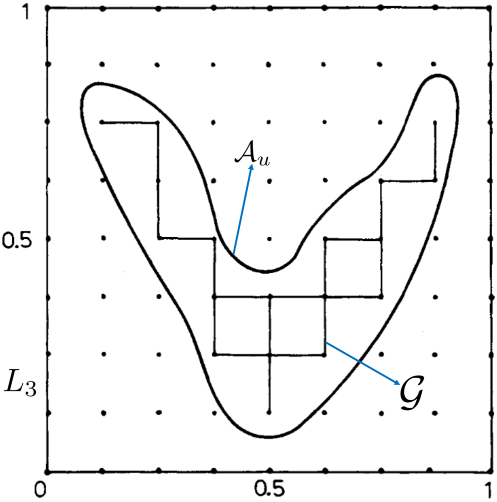
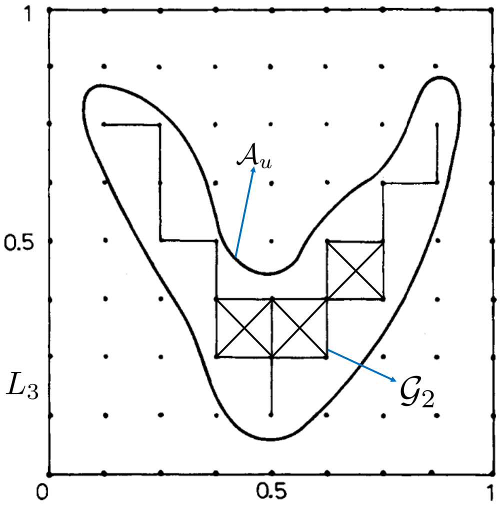


Image generated by (Adler, 1981).

In this graph the number of cliques of size $2^0 = 1$ is the number of vertices.

The number of cliques of size $2^1 = 2$ is the number of edges.

2D example



Suppose we connect all vertices v, w in the lattice such that there is a path from v to w of length 2 which has one edge in the x direction and one edge in the y direction.
Then the number cliques of size $2^2 = 4$ in the new graph is the number of squares eg 3!

Rectanguloids as cliques

Definition

Given $v, w \in L_n$, define the distance $\rho(v, w)$ to be 0 if there is some i such that $|v_i - w_i| > 2^{-n}$. Else set $\rho(v, w)$ equal to the number of $i \in \{1, \dots, D\}$ such that $|v_i - w_i| = 1$.

Go back to the picture above and explain this.

Rectanguloids as cliques

Definition

Given $v, w \in L_n$, define the distance $\rho(v,w)$ to be 0 if there is some i such that $|v_i - w_i| > 2^{-n}$. Else set $\rho(v,w)$ equal to the number of $i \in \{1, \dots, D\}$ such that $|v_i - w_i| = 1$.

Go back to the picture above and explain this. Recall $\mathcal{G}$ is the subset of the lattice graph corresponding to the vertices $\tilde{\mathcal{A}}_u = \{\theta \in L_n : W(\theta) \geq u\}$.

Definition

For $d = 1, \dots, D$, let $\mathcal{G}_d$ be a new graph on the vertex set $\tilde{\mathcal{A}}_u$ such that vertices v and w are connected in $\mathcal{G}_d$ if $\rho(v,w) = d$ and there there is a path from v to w in $\mathcal{G}$.

($\mathcal{G}_1$ and $\mathcal{G}_2$ can be seen on the previous slides.)

Rectanguloids as cliques

Recall:

Theorem

Given a subgraph A of the lattice graph,

$$\chi(\mathcal{G}) = \sum_{d=0}^{D} (-1)^d R_d(\mathcal{G})$$

where $R_d(\mathcal{G})$ is the number of cubes of dimension d in the excursion graph $\mathcal{G}$. Then we can compute $R_d(\mathcal{G})$

Claim

$R_d(\mathcal{G})$ *is equal to the number of cliques of size* 2^d *in the graph* $\mathcal{G}_d$.

So we can just count the number of cliques! There are efficient methods to do this, eg:
(Eppstein, Loffler, & Strash, 2010) and (Csardi & Nepusz, 2006).

Examples

Go to ARXIV version of the paper to look at the examples!

Bibliography

Adler, R. J. (1981). *The Geometry of Random Fields*.

Adler, R. J., Bartz, K., Kou, S. C., & Monod, A. (2017). Estimating thresholding levels for random fields via Euler characteristics.

Csardi, G., & Nepusz, T. (2006). iGraph. *InterJournal*, *Complex Sy*, 1695. Retrieved from `http://igraph.org`

Eppstein, D., Loffler, M., & Strash, D. (2010). *Algorithms and Computation: Listing All Maximal Cliques in Sparse Graphs in Near-Optimal Time*.