How to prove tightness for the size of strange random sets? Based on [GPS 2008].

Gábor Pete
http://www.math.toronto.edu/∼gabor

Recall the Fourier spectral sample

The space $L^2(\Omega, \mu)$, where $\Omega = \{\pm 1\}^V$, μ uniform probability measure, inner product $\mathbf{E}[\, fg \,]$, has a nice orthonormal basis:

For $S \subset V$, let $\chi_S(\omega) := \prod_{v \in S} \omega(v)$, the parity inside S.

Any function $f \in L^2(\Omega, \mu)$ decomposes in this basis (Fourier-Walsh series):

$$\hat{f}(S) := \mathbf{E}[\, f \chi_S \,]\,; \qquad f(\omega) = \sum_{S \subset V} \hat{f}(S)\, \chi_S(\omega)\,.$$

By Parseval, $\sum_S \hat{f}(S)^2 = \mathbf{E}[\, f^2 \,]$. So can define probability measure $\mathbf{P}\big[\, \mathscr{S}_f = S \,\big] := \hat{f}(S)^2 / \mathbf{E}[\, f^2 \,]$, the spectral sample $\mathscr{S}_f \subset V$.

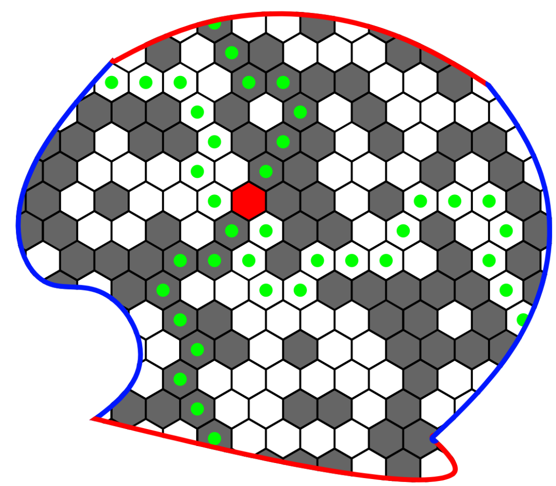
$$\frac{\mathbf{E}[\, f(\omega^\epsilon) f(\omega) \,] - \mathbf{E}[\, f \,]^2}{\mathbf{E}[\, f^2 \,]} = \sum_{S \neq \emptyset} \frac{\hat{f}(S)^2}{\mathbf{E}[\, f^2 \,]} (1-\epsilon)^{|S|} = \mathbf{E}\big[\, (1-\epsilon)^{|\mathscr{S}_f|}\,;\, |\mathscr{S}_f| > 0 \,\big],$$

hence small $\mathbf{P}\big[\, 0 < |\mathscr{S}_f| < K/\epsilon \,\big]$ means small covariance after ϵ-noise.

Although goal is to understand size, Gil Kalai suggested trying to understand entire distribution. A strange random set of bits.

Effective sampling? If f is an effectively computable Boolean function, then there is an effective quantum algorithm for $\mathscr{S}_f$ [Bernstein-Vazirani 1993].

For critical planar percolation, [Smirnov '01] + [Tsirelson '04] + [Schramm-Smirnov] implies that $\mathscr{S}_{\mathcal{Q},n}$ (left-right crossing in a conformal rectangle $\mathcal{Q}$, mesh $1/n$) has a conformally invariant scaling limit.



For ± 1-valued f, can consider pivotal bits.
$\mathbf{P}\big[\, x, y \in \mathrm{Piv}_f \,\big] = \mathbf{P}\big[\, x, y \in \mathscr{S}_f \,\big]$, but not for more points.
Both random subsets measure the "influence" or "relevance" of bits.
$\mathbf{P}\big[\, \mathscr{S}_{\mathcal{Q},n} \cap B \neq \emptyset \,\big] \asymp \mathbf{P}\big[\, B \text{ is pivotal for crossing } \mathcal{Q} \,\big]$
$= \alpha_4(B, \mathcal{Q})$, the 4-arm event.
$\mathbf{P}\big[\, \emptyset \neq \mathscr{S}_{\mathcal{Q},n} \subseteq B \,\big] \asymp \alpha_4(B, \mathcal{Q})^2$. But $\mathbf{P}\big[\, \emptyset \neq \mathrm{Piv}_{\mathcal{Q},n} \subseteq B \,\big] \asymp \alpha_6(B, \mathcal{Q})$.

Three very simple examples

$\text{Dictator}_n(x_1, \ldots, x_n) := x_1$.
Here $\text{Cov}\big[\,\text{Dic}_n(x), \text{Dic}_n(x^\epsilon)\,\big] = 1 - \epsilon$, so noise-stable.
And $\mathbf{P}\big[\,\mathscr{S}_n = \{x_1\}\,\big] = 1$.

$\text{Majority}_n(x_1, \ldots, x_n) := \operatorname{sgn}(x_1 + \cdots + x_n) \approx \frac{1}{\sqrt{n}}(x_1 + \cdots + x_n)$.
Here $\text{Cov}\big[\,\text{Maj}_n(x), \text{Maj}_n(x^\epsilon)\,\big] = 1 - O(\epsilon)$, so noise-stable.
And $\mathbf{P}\big[\,\mathscr{S}_n = \{x_i\}\,\big] \asymp 1/n$, most of the weight is on singletons.
On the other hand, $\mathbf{E}|\mathscr{S}_n| = \mathbf{E}|\text{Piv}_n| \asymp \frac{1}{\sqrt{n}}\,n \asymp \sqrt{n}$.

$\text{Parity}_n(x_1, \ldots, x_n) := x_1 \cdots x_n$
Here $\text{Cov}\big[\,\text{Par}_n(x), \text{Par}_n(x^\epsilon)\,\big] = (1 - \epsilon)^n$, the most sensitive to noise.
And $\mathbf{P}\big[\,\mathscr{S}_n = \{x_1, \ldots, x_n\}\,\big] = 1$.

Self-similarity for left-right crossing of $n \times n$ square

$$\mathbf{E}|\mathscr{S}_n| = \mathbf{E}|\mathrm{Piv}_n| \asymp n^2\, \alpha_4(1,n) \overset{\Delta}{\asymp} n^{3/4+o(1)}\,,$$

$$\mathbf{E}|\mathscr{S}_n(r)| := \mathbf{E}\Big[\,\#\big\{r\text{-boxes } \mathscr{S}_n \cap B_r \neq \emptyset\big\}\,\Big] \asymp \frac{n^2}{r^2}\,\alpha_4(r,n) \asymp \mathbf{E}|\mathscr{S}_{n/r}|\,,$$

$$\mathbf{E}\Big[\,|\mathscr{S}_n \cap B_r| \;\Big|\; \mathscr{S}_n \cap B_r \neq \emptyset\Big] \asymp r^2\,\alpha_4(1,r) \asymp \mathbf{E}|\mathscr{S}_r|\,.$$

Of course, $r^2\,\alpha_4(1,r) \cdot \frac{n^2}{r^2}\,\alpha_4(r,n) \asymp n^2\,\alpha_4(1,n)$, by quasi-multiplicativity.

Self-similarity for left-right crossing of $n \times n$ square

$$\mathbf{E}|\mathscr{S}_n| = \mathbf{E}|\mathrm{Piv}_n| \asymp n^2\,\alpha_4(1,n) \overset{\Delta}{\asymp} n^{3/4+o(1)}\,,$$

$$\mathbf{E}|\mathscr{S}_n(r)| := \mathbf{E}\Big[\#\big\{r\text{-boxes } \mathscr{S}_n \cap B_r \neq \emptyset\big\}\Big] \asymp \frac{n^2}{r^2}\,\alpha_4(r,n) \asymp \mathbf{E}|\mathscr{S}_{n/r}|\,,$$

$$\mathbf{E}\Big[\,|\mathscr{S}_n \cap B_r|\;\Big|\;\mathscr{S}_n \cap B_r \neq \emptyset\Big] \asymp r^2\,\alpha_4(1,r) \asymp \mathbf{E}|\mathscr{S}_r|\,.$$

Of course, $r^2\,\alpha_4(1,r)\cdot\frac{n^2}{r^2}\,\alpha_4(r,n) \asymp n^2\,\alpha_4(1,n)$, by quasi-multiplicativity.

Similar to the zero-set of simple random walk: $\mathbf{E}|\mathscr{Z}_n| \asymp n\,n^{-1/2} = n^{1/2}$,

$$\mathbf{E}|\mathscr{Z}_n(r)| := \mathbf{E}\Big[\#\big\{r\text{-intervals } \mathscr{Z}_n \cap I_r \neq \emptyset\big\}\Big] \asymp \frac{n}{r}\,(n/r)^{-1/2} \asymp \mathbf{E}|\mathscr{Z}_{n/r}|\,,$$

$$\mathbf{E}\Big[\,|\mathscr{Z}_n \cap I_r|\;\Big|\;\mathscr{Z}_n \cap I_r \neq \emptyset\Big] \asymp r\,r^{-1/2} \asymp \mathbf{E}|\mathscr{Z}_r|\,.$$

The $\mathscr{S}_n(r)$ and $\mathscr{Z}_n(r)$ results are related to the existence of scaling limits.

What concentration can we expect?

$\mathscr{S}_n$ is very different from uniform set of similar density:
i.i.d. $\mathbf{P}[\, x \in \mathscr{U}_n \,] = n^{-5/4}$. Hence $\mathbf{E}|\mathscr{U}_n| = n^{3/4}$.

For large r ($\gg n^{5/8}$), this $\mathscr{U}_n$ intersects every r-box;
for small r, if it intersects one, there is just one point there.

Concentration of size: roughly within $\sqrt{\mathbf{E}|\mathscr{U}_n|} = n^{3/8}$.

A bit more similar: for $i = 1, \ldots, (n/r)^2$, i.i.d. $\mathbf{P}\big[\, X_i = r^{3/4} \,\big] = (n/r)^{-5/4}$,
$X_i = 0$ otherwise. Then $S_{n,r} := \sum_i X_i$. Hence $\mathbf{E}|S_{n,r}| = n^{3/4}$.

For $r = n^\gamma$, size $|S_{n,r}|$ is concentrated within $n^{3/8(1+\gamma)}$, still $o(\mathbf{E}|S_{n,r}|)$.

For self-similar sets, we expect only tightness around the mean:
$\mathbf{P}\big[\, 0 < |\mathscr{S}_n| < \lambda\, \mathbf{E}|\mathscr{S}_n| \,\big] \to 0$ as $\lambda \to 0$, uniformly in n.

Proving tightness with a lot of independence

Assume we have the following ingredients, true for the zeroes:

(1) $\quad \mathbf{P}\Big[\, |\mathcal{Z}_n \cap I_r| > c\, \mathbf{E}|\mathcal{Z}_r| \,\Big|\, \mathcal{Z}_n \cap I_r \neq \emptyset,\ \mathscr{F}_{[n]\setminus I_r} \Big] \geqslant c > 0.$

(2) $\quad \mathbf{P}\big[\, |\mathcal{Z}_n(r)| = k \,\big] \leqslant g(k)\, \mathbf{P}\big[\, |\mathcal{Z}_n(r)| = 1 \,\big]$, with sub-exponential $g(k)$:

when the r-intervals intersected are scattered, have to pay k times to get to and leave them, and this cost is not balanced by combinatorial entropy.

Proving tightness with a lot of independence

Assume we have the following ingredients, true for the zeroes:

(1) $\mathbf{P}\Big[\,|\mathscr{Z}_n \cap I_r| > c\,\mathbf{E}|\mathscr{Z}_r| \,\Big|\, \mathscr{Z}_n \cap I_r \neq \emptyset,\ \mathscr{F}_{[n]\setminus I_r}\Big] \geqslant c > 0.$

(2) $\mathbf{P}\big[\,|\mathscr{Z}_n(r)| = k\,\big] \leqslant g(k)\,\mathbf{P}\big[\,|\mathscr{Z}_n(r)| = 1\,\big]$, with sub-exponential $g(k)$:

when the r-intervals intersected are scattered, have to pay k times to get to and leave them, and this cost is not balanced by combinatorial entropy.

$$\mathbf{P}\big[\,0 < |\mathscr{Z}_n| < c\,\mathbf{E}|\mathscr{Z}_r|\,\big] = \sum_{k\geqslant 1} \mathbf{P}\Big[\,0 < |\mathscr{Z}_n| < c\,\mathbf{E}|\mathscr{Z}_r|\,,\ |\mathscr{Z}_n(r)| = k\,\Big]$$

$$\text{by (1):} \quad \leqslant \sum_{k\geqslant 1} (1-c)^k\,\mathbf{P}\big[\,|\mathscr{Z}_n(r)| = k\,\big]$$

$$\text{by (2):} \quad \leqslant O(1)\,\mathbf{P}\big[\,|\mathscr{Z}_n(r)| = 1\,\big] \asymp (n/r)^{1-3/2},$$

which, using $\lambda = \frac{c\,\mathbf{E}|\mathscr{Z}_r|}{\mathbf{E}|\mathscr{Z}_n|} \asymp (r/n)^{1/2}$, reads as $\mathbf{P}\big[\,0 < |\mathscr{Z}_n| < \lambda\,\mathbf{E}|\mathscr{Z}_n|\,\big] \asymp \lambda$.

But we know much less independence for $\mathscr{S}_n$

$$(1') \quad \mathbf{P}\Big[\, |\mathscr{S}_n \cap B_r/3| > c\,\mathbf{E}|\mathscr{S}_r| \,\Big|\, \mathscr{S}_n \cap B_r \neq \emptyset = \mathscr{S}_n \cap W \,\Big] \geqslant c > 0,$$

for any W that is not too close to B_r.

Why only this negative conditioning? Inclusion formula:

$$\mathbf{P}[\,\mathscr{S}_f \subset U\,] = \sum_{S\subset U} \hat{f}(S)^2 = \mathbf{E}\Big[\Big(\sum_{S\subset U} \hat{f}(S)\,\chi_S\Big)^2\Big] = \mathbf{E}\Big[\mathbf{E}[\,f \mid \mathscr{F}_U\,]^2\Big].$$

From this, for disjoint subsets A and B,

$$\begin{aligned}
\mathbf{P}[\,\mathscr{S}_f \cap B \neq \emptyset = \mathscr{S}_f \cap A\,] &= \mathbf{P}[\,\mathscr{S}_f \subseteq A^c\,] - \mathbf{P}[\,\mathscr{S}_f \subseteq (A\cup B)^c\,] \\
&= \mathbf{E}\Big[\mathbf{E}[\,f \mid \mathscr{F}_{A^c}\,]^2 - \mathbf{E}[\,f \mid \mathscr{F}_{(A\cup B)^c}\,]^2\Big] \\
&= \mathbf{E}\Big[\big(\mathbf{E}[\,f \mid \mathscr{F}_{A^c}\,] - \mathbf{E}[\,f \mid \mathscr{F}_{(A\cup B)^c}\,]\big)^2\Big].
\end{aligned}$$

So, what are we going to do?

With quite a lot of work for both items,

(1') $\quad \mathbf{P}\Big[\, |\mathscr{S}_n \cap B_r/3| > c\,\mathbf{E}|\mathscr{S}_r| \;\Big|\; \mathscr{S}_n \cap B_r \neq \emptyset = \mathscr{S}_n \cap W \,\Big] \geqslant c > 0.$

(2) $\quad \mathbf{P}\big[\, |\mathscr{S}_n(r)| = k \,\big] \leqslant g(k)\, \mathbf{P}\big[\, |\mathscr{S}_n(r)| = 1 \,\big]$, with sub-exponential $g(k)$.

We could repeat (1') for many r-boxes only if "not enough points in one box" meant "we found nothing in that box".

So, take an independent random dilute sample: $\mathbf{P}[\, x \in \mathcal{R} \,] = 1/\mathbf{E}|\mathscr{S}_r|$ i.i.d.
Then, $|\mathscr{S}_n \cap B_r/3|$ is small $\implies \mathcal{R} \cap \mathscr{S}_n \cap B_r/3 = \emptyset$ is likely,
and $|\mathscr{S}_n \cap B_r/3|$ is large $\implies \mathcal{R} \cap \mathscr{S}_n \cap B_r/3 \neq \emptyset$ is likely.

So, what are we going to do?

With quite a lot of work for both items,

(1') $\quad \mathbf{P}\Big[\, |\mathscr{S}_n \cap B_r/3| > c\,\mathbf{E}|\mathscr{S}_r| \;\Big|\; \mathscr{S}_n \cap B_r \neq \emptyset = \mathscr{S}_n \cap W \,\Big] \geqslant c > 0.$

(2) $\quad \mathbf{P}\big[\, |\mathscr{S}_n(r)| = k \,\big] \leqslant g(k)\, \mathbf{P}\big[\, |\mathscr{S}_n(r)| = 1 \,\big]$, with sub-exponential $g(k)$.

We could repeat (1') for many r-boxes only if "not enough points in one box" meant "we found nothing in that box".

So, take an independent random dilute sample: $\mathbf{P}[\, x \in \mathcal{R}\,] = 1/\mathbf{E}|\mathscr{S}_r|$ i.i.d.
Then, $|\mathscr{S}_n \cap B_r/3|$ is small $\implies \mathcal{R} \cap \mathscr{S}_n \cap B_r/3 = \emptyset$ is likely,
and $|\mathscr{S}_n \cap B_r/3|$ is large $\implies \mathcal{R} \cap \mathscr{S}_n \cap B_r/3 \neq \emptyset$ is likely.

But $\mathbf{P}\Big[\, \mathscr{S}_n \neq \emptyset = \mathcal{R} \cap \mathscr{S}_n \;\Big|\; |\mathscr{S}_n(r)| = k \,\Big]$ is still problematic conditioning.

Or, if we scan sequentially the r-boxes until $\mathcal{R} \cap \mathscr{S}_n \cap B_r/3 \neq \emptyset$, how would (2) imply that we had a good chance of success several times? We don't know how $\mathbf{P}\big[\, \mathscr{S}_n \cap B_r(t) \neq \emptyset \;\big|\; \mathscr{S}_n \cap W(t) = \emptyset \,\big]$ changes with the steps t.

Oded's first solution: a filtered Markov inequality

If $\mathscr{F}_k$ is a monotone increasing filtration, X_k are non-negative variables, and $Y_k := \mathbf{E}\big[\, X_k \,\big|\, \mathscr{F}_k \,\big]$, then, for any $s, t \geqslant 0$,

$$\mathbf{P}\Big[\sum_k Y_k \leqslant s,\ \sum_k X_k \geqslant t\Big] \leqslant s/t\,.$$

In the application, $\mathscr{F}_k$ is the σ-algebra generated by the random sets $\{\mathcal{R} \cap \mathscr{S}_n \cap B_j : j \leqslant k-1\}$, and $X_k = 1_{\{\mathscr{S}_n \cap B_k \neq \emptyset\}}$.

Since (2) says that $\sum_k X_k$ is probably large, we get the same for $\sum_k Y_k$. Hence, with large probability, there are several boxes where the scanning has a positive chance to succeed, so it is unlikely that it remains unsuccessful.

However, the Markov-type bound is too weak, we don't get the sharp result.

Oded's second solution: a large deviation lemma

Suppose $X_i, Y_i \in \{0,1\}$, $i = 1, \dots, n$, and that $\forall J \subset [n]$ and $\forall i \in [n] \setminus J$

$$\mathbf{P}\big[\, Y_i = 1 \,\big|\, \forall_{j \in J} Y_j = 0 \,\big] \geqslant c\, \mathbf{P}\big[\, X_i = 1 \,\big|\, \forall_{j \in J} Y_j = 0 \,\big].$$

Then

$$\mathbf{P}\Big[\forall_i Y_i = 0\Big] \leqslant c^{-1}\, \mathbf{E}\Big[\exp\Big(-(c/e) \sum_i X_i\Big)\Big].$$

We use this with $X_j := 1_{\{\mathscr{S} \cap B_j \neq \emptyset\}}$ and $Y_j := 1_{\{\mathscr{S} \cap B_j \cap \mathcal{R} \neq \emptyset\}}$.

Proof: Instead of sequential scan, average everything together.
Choose $J \subset [n]$ randomly, Bernoulli$(1-p)$. Get $\mathbf{E}\big[\, Y\, p^Y \,\big] \geqslant c\, \mathbf{E}\big[\, X\, p^{Y+1} \,\big]$.

So, $\mathbf{E}[\, Z\,] \geqslant 0$, where $Z := (Y - c\, p\, X)\, p^Y$. Choose $p := e^{-1}$. Maximize Z over Y, and get the bound $Z \leqslant \exp(-1 - c\, X/e)$. Altogether, $c\, e^{-1}\, \mathbf{P}\big[\, Y = 0 < X \,\big] \leqslant \mathbf{E}\big[\, 1_{X>0} \exp(-1 - c\, X/e) \,\big]$, and done.

Final result for the spectral sample

If $r \in [1, n]$, then $\{|\mathscr{S}_n| < \mathbf{E}|\mathscr{S}_r|\}$ is basically equivalent to being contained inside some $r \times r$ sub-square:

$$\mathbf{P}\big[\, 0 < |\mathscr{S}_n| < \mathbf{E}|\mathscr{S}_r| \,\big] \asymp \alpha_4(r,n)^2 \left(\frac{n}{r}\right)^2 .$$

In particular, on the triangular lattice Δ,

$$\mathbf{P}\big[\, 0 < |\mathscr{S}_n| < \lambda\, \mathbf{E}|\mathscr{S}_n| \,\big] \asymp \lambda^{2/3} .$$

The *scaling limit* of $\mathscr{S}_n$ is a conformally invariant Cantor-set with Hausdorff-dimension $3/4$.

Remark. The same strategy gives $\mathbf{P}\big[\, 0 < |\mathsf{Piv}_n| < \lambda\, \mathbf{E}|\mathsf{Piv}_n| \,\big] \asymp \lambda^{11/9}$, but it's an overkill, given all the independence in Piv_n.

Some related questions

Question 1: Can one build similar proofs for other Boolean functions?

Question 2: Self-similarity of Piv_n and $\mathscr{S}_n$ is a lot of restriction.

Conjecture [Gil Kalai]: The entropy of such random sets X_n is at most $\mathbf{E}|X_n|$, i.e., there is no $\log$ factor as in uniform.

In particular, Influence-Entropy conjecture [Friedgut-Kalai 1996]: For some universal constant C, for any Boolean function f,

$$\mathrm{SpecEnt}(f) := \sum_{S \subset [n]} \widehat{f}(S)^2 \log \frac{1}{\widehat{f}(S)^2} \leqslant C \times$$

$$\times\, \mathrm{Influence}(f) := \mathbf{E}|\mathscr{S}_f| = \mathbf{E}|\mathrm{Piv}_f| = \sum_{S \subset [n]} \widehat{f}(S)^2 |S|\,.$$

I think I can do it for Piv_n, but have no idea about $\mathscr{S}_n$.