The number of random 2-SAT solutions is asymptotically log-normal

Noela Müller
Eindhoven University of Technology

Based on joint work with Arnab Chatterjee, Amin Coja-Oghlan, Connor Riddlesden, Maurice Rolvien, Pavel Zakharov and Haodong Zhu

Workshop on Partition Functions, CWI
28^{th} March 2025

Boolean satisfiability problem

The Boolean satisfiability problem from logic and computer science asks the following:

Given a propositional formula Φ,
determine whether its variables can be consistently replaced by
TRUE or **FALSE**
such that the overall formula evaluates to **TRUE**.

Restrict to k-**CNF formulas** $\Phi = \Phi_{n,m}$:

Given $\boldsymbol{n}$ **variables** $\{x_1, \ldots, x_n\}$, assume that

$$\Phi_{n,m} = (\ell_{1,1} \vee \ldots \vee \ell_{1,k}) \wedge (\ell_{2,1} \vee \ldots \vee \ell_{2,k}) \wedge \ldots \wedge (\ell_{m,1} \vee \ldots \vee \ell_{m,k}),$$

is a conjunction of $\boldsymbol{m}$ **clauses** of the form $a_i = \ell_{i,1} \vee \ldots \vee \ell_{i,k}$, where $\ell_{i,j} \in \{x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n\}$.

Restrict to k-**CNF formulas** $\Phi = \Phi_{n,m}$:

Given $\boldsymbol{n}$ **variables** $\{x_1, \ldots, x_n\}$, assume that

$$\Phi_{n,m} = (\ell_{1,1} \vee \ldots \vee \ell_{1,k}) \wedge (\ell_{2,1} \vee \ldots \vee \ell_{2,k}) \wedge \ldots \wedge (\ell_{m,1} \vee \ldots \vee \ell_{m,k}),$$

is a conjunction of $\boldsymbol{m}$ **clauses** of the form $a_i = \ell_{i,1} \vee \ldots \vee \ell_{i,k}$, where $\ell_{i,j} \in \{x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n\}$.

$\{x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n\}$ is the set of **literals**.

Restrict to k-**CNF formulas** $\Phi = \Phi_{n,m}$:

Given $\boldsymbol{n}$ **variables** $\{x_1, \ldots, x_n\}$, assume that

$$\Phi_{n,m} = (\ell_{1,1} \vee \ldots \vee \ell_{1,k}) \wedge (\ell_{2,1} \vee \ldots \vee \ell_{2,k}) \wedge \ldots \wedge (\ell_{m,1} \vee \ldots \vee \ell_{m,k}),$$

is a conjunction of $\boldsymbol{m}$ **clauses** of the form $a_i = \ell_{i,1} \vee \ldots \vee \ell_{i,k}$, where $\ell_{i,j} \in \{x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n\}$.

$\{x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n\}$ is the set of **literals**.

1. Does there exist a satisfying variable assignment?

2. If a satisfying assignment exists, how many are there?

Restrict to k-**CNF formulas** $\Phi = \Phi_{n,m}$:

Given $\boldsymbol{n}$ **variables** $\{x_1, \ldots, x_n\}$, assume that

$$\Phi_{n,m} = (\ell_{1,1} \vee \ldots \vee \ell_{1,k}) \wedge (\ell_{2,1} \vee \ldots \vee \ell_{2,k}) \wedge \ldots \wedge (\ell_{m,1} \vee \ldots \vee \ell_{m,k}),$$

is a conjunction of $\boldsymbol{m}$ **clauses** of the form $a_i = \ell_{i,1} \vee \ldots \vee \ell_{i,k}$, where $\ell_{i,j} \in \{x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n\}$.

$\{x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n\}$ is the set of **literals**.

1. Does there exist a satisfying variable assignment?
   NP-complete for $k \geq 3$ and in P for $k = 2$.
2. If a satisfying assignment exists, how many are there?
   #P-complete for $k \geq 2$.

Random satisfiability

Early observation:

Many ‘industrial’ instances of Boolean formulas can be efficiently tackled by existing SAT-solvers, despite (conjectured) theoretical hardness.

Random satisfiability

Early observation:

Many ‘industrial’ instances of Boolean formulas can be efficiently tackled by existing SAT-solvers, despite (conjectured) theoretical hardness.

Selman, Mitchell and Levesque (1996):
Random instances of 3-SAT with certain clause-to-variable ratios appear to be **very difficult to solve**.

Random satisfiability

Early observation:

Many ‘industrial’ instances of Boolean formulas can be efficiently tackled by existing SAT-solvers, despite (conjectured) theoretical hardness.

Selman, Mitchell and Levesque (1996):
Random instances of 3-SAT with certain clause-to-variable ratios appear to be **very difficult to solve**.

- What are characteristic features of random formulas?
- How are these related to the performance of algorithms?

Random 2-SAT

Denote by $\mathbf{\Phi} = \mathbf{\Phi}_{n,m}$ a random **2-CNF** on n Boolean variables $x_1, \ldots, x_n$ with m clauses,
drawn independently and uniformly from all $4\binom{n}{2}$ possible 2-clauses:

$$\mathbf{\Phi} = (\boldsymbol{\ell}_{1,1} \vee \boldsymbol{\ell}_{1,2}) \wedge \ldots \wedge (\boldsymbol{\ell}_{m,1} \vee \boldsymbol{\ell}_{m,2}),$$

where $\boldsymbol{\ell}_{1,1}, \boldsymbol{\ell}_{1,2}, \ldots, \boldsymbol{\ell}_{m,2} \in \{x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n\}$.

Random 2-SAT

Denote by $\mathbf{\Phi} = \mathbf{\Phi}_{n,m}$ a random **2-CNF** on n Boolean variables $x_1, \ldots, x_n$
with m clauses,
drawn independently and uniformly from all $4\binom{n}{2}$ possible 2-clauses:

$$\mathbf{\Phi} = (\boldsymbol{\ell}_{1,1} \vee \boldsymbol{\ell}_{1,2}) \wedge \ldots \wedge (\boldsymbol{\ell}_{m,1} \vee \boldsymbol{\ell}_{m,2}),$$

where $\boldsymbol{\ell}_{1,1}, \boldsymbol{\ell}_{1,2}, \ldots, \boldsymbol{\ell}_{m,2} \in \{x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n\}$.

Suppose that

$$m \sim dn/2$$

for a fixed real $d > 0$.

Random 2-SAT

Denote by $\mathbf{\Phi} = \mathbf{\Phi}_{n,m}$ a random **2-CNF** on n Boolean variables $x_1, \ldots, x_n$ with m clauses,
drawn independently and uniformly from all $4\binom{n}{2}$ possible 2-clauses:

$$\mathbf{\Phi} = (\boldsymbol{\ell}_{1,1} \vee \boldsymbol{\ell}_{1,2}) \wedge \ldots \wedge (\boldsymbol{\ell}_{m,1} \vee \boldsymbol{\ell}_{m,2}),$$

where $\boldsymbol{\ell}_{1,1}, \boldsymbol{\ell}_{1,2}, \ldots, \boldsymbol{\ell}_{m,2} \in \{x_1, \neg x_1, x_2, \neg x_2, \ldots, x_n, \neg x_n\}$.

Suppose that

$$m \sim dn/2$$

for a fixed real $d > 0$.

$\Longrightarrow$ The parameter d represents the average number of clauses in which any variable x_i appears.

Satisfiability threshold

$m = \Theta(n)$ is the 'right' clause-to-variable ratio to observe a (sharp) transition from satisfiability to unsatisfiability:

Satisfiability threshold

$m = \Theta(n)$ is the 'right' clause-to-variable ratio to observe a (sharp) transition from satisfiability to unsatisfiability:

Theorem (Chvátal & Reed (1992), Goerdt (1992), Fernandez de la Vega (1992))

Let $\Phi = \Phi_{n,m}$ be a random 2-CNF on n Boolean variables with $m \sim dn/2$ for a fixed real $d > 0$. Then for any $\varepsilon > 0$:

- *If $d \leq 2 - \varepsilon$, w.h.p. Φ is satisfiable.*
- *If $d \geq 2 + \varepsilon$, w.h.p. Φ is not satisfiable.*

Proof spirit

Approach:

Translate satisfiability question into graph-theoretical question and apply techniques from the theory of random (di)graphs.

Proof spirit

Approach:

Translate satisfiability question into graph-theoretical question and apply techniques from the theory of random (di)graphs.

More specifically:

Both satisfiability and unsatisfiability of a formula are related to the (non-)existence of cycles with a certain structure.

Then apply first and second moment method on cycle counts.

General $k \geq 3$

Let $\mathbf{\Phi} = \mathbf{\Phi}_{n,m}$ be a **random k-CNF**
on n Boolean variables $x_1, \ldots, x_n$ with m clauses,
drawn independently and uniformly from all $2^k \binom{n}{k}$ possible k-clauses.

General $k \geq 3$

Let $\mathbf{\Phi} = \mathbf{\Phi}_{n,m}$ be a **random k-CNF**
on n Boolean variables $x_1, \ldots, x_n$ with m clauses,
drawn independently and uniformly from all $2^k\binom{n}{k}$ possible k-clauses.

Suppose that $m \sim dn/k$ for a fixed real $d > 0$.

General $k \geq 3$

Let $\mathbf{\Phi} = \mathbf{\Phi}_{n,m}$ be a **random k-CNF**
on n Boolean variables $x_1, \ldots, x_n$ with m clauses,
drawn independently and uniformly from all $2^k \binom{n}{k}$ possible k-clauses.

Suppose that $m \sim dn/k$ for a fixed real $d > 0$.

Theorem (Friedgut (1999))

For each $k \geq 3$ there is a function $d_k(n)$ bounded above and below by constants so that for every $\varepsilon > 0$ the following hold:

- *If $d \leq (1-\varepsilon)d_k(n)$, w.h.p. $\mathbf{\Phi}$ is satisfiable.*
- *If $d \geq (1+\varepsilon)d_k(n)$, w.h.p. $\mathbf{\Phi}$ is not satisfiable.*

General $k \geq 3$

Let $\mathbf{\Phi} = \mathbf{\Phi}_{n,m}$ be a **random k-CNF**
on n Boolean variables $x_1, \ldots, x_n$ with m clauses,
drawn independently and uniformly from all $2^k \binom{n}{k}$ possible k-clauses.

Suppose that $m \sim dn/k$ for a fixed real $d > 0$.

Theorem (Friedgut (1999))

For each $k \geq 3$ there is a function $d_k(n)$ bounded above and below by constants so that for every $\varepsilon > 0$ the following hold:

- *If $d \leq (1-\varepsilon)d_k(n)$, w.h.p. $\mathbf{\Phi}$ is satisfiable.*
- *If $d \geq (1+\varepsilon)d_k(n)$, w.h.p. $\mathbf{\Phi}$ is not satisfiable.*

Does $(d_k(n))_n$ converge?

Random k-SAT as a spin glass model

In the early 2000's, physicists observed that random k-SAT (and, more generally, random constraint satisfaction problems) can be studied as examples of **diluted spin glasses**.

Random k-SAT as a spin glass model

In the early 2000's, physicists observed that random k-SAT (and, more generally, random constraint satisfaction problems) can be studied as examples of **diluted spin glasses**.

Based on the **cavity method**, Mézard, Parisi & Zecchina and Mertens, Mézard & Zecchina (early 2000's) put forward an explicit characterisation of the conjectured limit d_k.

Proof of the satisfiability threshold conjecture for large k

Theorem (Ding, Sly, Sun (2015))

Let $\mathbf{\Phi} = \mathbf{\Phi}_{n,m}$ be a random k-CNF on n Boolean variables with $m \sim dn/k$ for a fixed real $d > 0$. Moreover, assume that $k \geq k_0$ for an absolute constant k_0. Then there exists d_k that matches the physics predictions such that for all $\varepsilon > 0$:

- *If $d \leq d_k - \varepsilon$, w.h.p. $\mathbf{\Phi}$ is satisfiable.*
- *If $d \geq d_k + \varepsilon$, w.h.p. $\mathbf{\Phi}$ is not satisfiable.*

Counting solutions

Let's return to the second initial question:

If a satisfying assignments exist,
how many of them are there typically?

Counting solutions

Let's return to the second initial question:

If a satisfying assignments exist,
how many of them are there typically?

The number of solutions is related to structural properties of the solution space geometry.

$\longrightarrow$ Connection to the computational nature of finding or sampling solutions.

Counting solutions

Let's return to the second initial question:

If a satisfying assignments exist,
how many of them are there typically?

The number of solutions is related to structural properties of the solution space geometry.
$\longrightarrow$ Connection to the computational nature of finding or sampling solutions.

For random 2-SAT:
Monasson & Zecchina (1996) put forward a statistical physics based prediction about the leading exponential order of the number of solutions.

Denote by $Z(\mathbf{\Phi})$ the number of satisfying assignments of $\mathbf{\Phi}$.

Denote by $Z(\mathbf{\Phi})$ the number of satisfying assignments of $\mathbf{\Phi}$.

Theorem (Achlioptas, Coja-Oghlan, Hahn-Klimroth, Lee, M., Penschuk, Zhou (2021))

Fix $0 < d < 2$. There exists a probability distribution π_d on $(0,1)$ such that for i.i.d. samples $(\boldsymbol{\mu}_{\pi_d,i})_{i\geq 1}$ from π_d and $\boldsymbol{d}^-, \boldsymbol{d}^+ \sim \mathrm{Po}(d/2)$, all independent, we have

$$\frac{1}{n}\log Z(\mathbf{\Phi}) \xrightarrow{\mathbb{P}} \mathbb{E}\left[\log\left(\prod_{i=1}^{\boldsymbol{d}^-}\boldsymbol{\mu}_{\pi_d,i} + \prod_{i=1}^{\boldsymbol{d}^+}\boldsymbol{\mu}_{\pi_d,\boldsymbol{d}^-+i}\right) - \frac{d}{2}\log\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right]$$
$$=: \phi(d).$$






Related work

- **Boufkhad and Dubois** (1999) obtain best prior lower bound on $\frac{1}{n}\log Z(\mathbf{\Phi})$.
- **Franz & Leone** (2003), **Panchenko & Talagrand** (2004) obtain an asymptotically tight upper bound on $\frac{1}{n}\log Z(\mathbf{\Phi})$ via the interpolation method.
- The analysis of a general approximation algorithm by **Montanari and Shah** (2007) implies analogous results (correlation decay, performance of BP, limit of the log-partition function) for a 'soft' version of random 2-SAT for $d < 1.16$.
- **Abbe and Montanari** (2015): $\frac{1}{n}\log Z(\mathbf{\Phi})$ converges in probability to a deterministic limit $\phi(d)$ for Lebesgue-almost all $d \in (0,2)$. Their approach does not give information on the value of $\phi(d)$.

High-level proof idea

The expected value

Consider the 'simpler' task of determining the asymptotics of

$$\frac{1}{n}\mathbb{E}[\log(Z(\mathbf{\Phi}_n) \vee 1)].$$

The expected value

Consider the 'simpler' task of determining the asymptotics of

$$\frac{1}{n}\mathbb{E}[\log(Z(\mathbf{\Phi}_n) \vee 1)].$$

One approach to this problem:
Aizenman-Sims-Starr scheme from the mathematics of spin glasses:

Compute the asymptotic mean of a random variable on a formula of size n by estimating the change of that mean upon going to a formula of size $n+1$.

$$\frac{1}{n}\mathbb{E}[\log(Z(\mathbf{\Phi}_n) \vee 1)] = \frac{1}{n}\sum_{N=2}^{n-1}\big(\mathbb{E}[\log(Z(\mathbf{\Phi}_{N+1}) \vee 1)] - \mathbb{E}[\log(Z(\mathbf{\Phi}_N) \vee 1)]\big) + \frac{1}{n}\mathbb{E}[\log(Z(\mathbf{\Phi}_2) \vee 1)].$$

The expected value

Proposition

We have

$$\lim_{n\to\infty} \mathbb{E}[\log(Z(\boldsymbol{\Phi}_{n+1}) \vee 1)] - \mathbb{E}[\log(Z(\boldsymbol{\Phi}_n) \vee 1)]$$

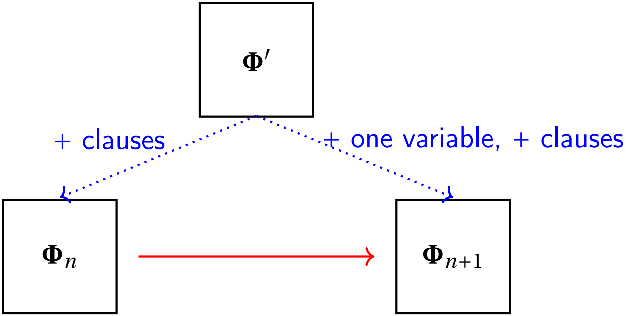
$$= \mathbb{E}\left[\log\left(\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi_d,i} + \prod_{i=1}^{\boldsymbol{d}^+} \boldsymbol{\mu}_{\pi_d,i+\boldsymbol{d}^-}\right) - \frac{d}{2}\log\left(1 - \boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right].$$

$$= \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\log(Z(\boldsymbol{\Phi}_n) \vee 1)]$$

by Stolz-Cesàro Theorem.

Coupling

The difference is calculated by coupling the formulas of size n and $n+1$ such that the latter is obtained from the former by a small expected number of local changes.



Goal:
Get a handle on the expected change of the effects of

1. adding a (small) number of clauses and
2. adding a variable and a (small) number of clauses.

Goal:
Get a handle on the expected change of the effects of

1. adding a (small) number of clauses and
2. adding a variable and a (small) number of clauses.

Both can be expressed in terms of the joint marginals of a bounded number of variables with respect to the uniform distribution over satisfying assignments.

From digraphs to marginals

Denote by $S(\mathbf{\Phi})$ the set of all satisfying assignments of $\mathbf{\Phi}$.

From digraphs to marginals

Denote by $S(\mathbf{\Phi})$ the set of all satisfying assignments of $\mathbf{\Phi}$.
Assuming that $S(\mathbf{\Phi}) \neq \emptyset$, let

$$\mu_{\mathbf{\Phi}}(\sigma) = \frac{\mathbb{1}\{\sigma \in S(\mathbf{\Phi})\}}{Z(\mathbf{\Phi})}, \qquad \sigma \in \{\pm 1\}^{\{x_1,\ldots,x_n\}},$$

denote the uniform distribution on $S(\mathbf{\Phi})$, where $Z(\mathbf{\Phi}) = |S(\mathbf{\Phi})|$.

(Encoding *'true'* by $+1$ and *'false'* by -1.)

From digraphs to marginals

Denote by $S(\mathbf{\Phi})$ the set of all satisfying assignments of $\mathbf{\Phi}$.
Assuming that $S(\mathbf{\Phi}) \neq \emptyset$, let

$$\mu_{\mathbf{\Phi}}(\sigma) = \frac{\mathbb{1}\{\sigma \in S(\mathbf{\Phi})\}}{Z(\mathbf{\Phi})}, \qquad \sigma \in \{\pm 1\}^{\{x_1,\ldots,x_n\}},$$

denote the uniform distribution on $S(\mathbf{\Phi})$, where $Z(\mathbf{\Phi}) = |S(\mathbf{\Phi})|$.

(Encoding *'true'* by $+1$ and *'false'* by -1.)

Samples from $\mu_{\mathbf{\Phi}}$ are denoted by the boldface notation $\boldsymbol{\sigma}$.

Back to the expectation

For simplicity, let $\Phi + \boldsymbol{a}$ denote the formula that is obtained from Φ by adding a uniformly random clause

$$\boldsymbol{a} = \boldsymbol{s}_1 x_{\boldsymbol{i}} \vee \boldsymbol{s}_2 x_{\boldsymbol{j}}.$$

Assume that Φ is satisfiable.

Back to the expectation

For simplicity, let $\Phi + \boldsymbol{a}$ denote the formula that is obtained from Φ by adding a uniformly random clause

$$\boldsymbol{a} = \boldsymbol{s}_1 x_{\boldsymbol{i}} \vee \boldsymbol{s}_2 x_{\boldsymbol{j}}.$$

Assume that Φ is satisfiable.

Then

$$\log(Z(\Phi + \boldsymbol{a})) - \log(Z(\Phi)) = \log\left(\frac{Z(\Phi + \boldsymbol{a})}{Z(\Phi)}\right)$$

and

$$\frac{Z(\Phi + \boldsymbol{a})}{Z(\Phi)} = \sum_{\sigma:\sigma \models \Phi} \frac{\mathbb{1}\{\sigma \models \boldsymbol{a}\}}{Z(\Phi)} = \mu_\Phi(\boldsymbol{\sigma} \models \boldsymbol{a})$$
$$= 1 - \mu_\Phi(\boldsymbol{\sigma}_{\boldsymbol{i}} \neq \boldsymbol{s}_1, \boldsymbol{\sigma}_{\boldsymbol{j}} \neq \boldsymbol{s}_2).$$

Local changes to a given formula

Having expressed $\mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)]$ as a sum of local changes, to analyse $\mu_{\mathbf{\Phi}}$, we next perform the following steps:

1. Analyse (joint) marginals on the local limit of $\mathbf{\Phi}$:
   $\longrightarrow$ Multitype Galton-Watson tree for formulas.

Local changes to a given formula

Having expressed $\mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)]$ as a sum of local changes, to analyse $\mu_{\mathbf{\Phi}}$, we next perform the following steps:

1. Analyse (joint) marginals on the local limit of $\mathbf{\Phi}$:
   $\longrightarrow$ Multitype Galton-Watson tree for formulas.
   - Establish decorrelation properties for random 2-SAT on the local limit tree:
     $\longrightarrow$ Gibbs uniqueness.

Local changes to a given formula

Having expressed $\mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)]$ as a sum of local changes, to analyse $\mu_{\mathbf{\Phi}}$, we next perform the following steps:

1. Analyse (joint) marginals on the local limit of $\mathbf{\Phi}$:
   $\longrightarrow$ Multitype Galton-Watson tree for formulas.
   - Establish decorrelation properties for random 2-SAT on the local limit tree:
     $\longrightarrow$ Gibbs uniqueness.
   - Characterize the root marginals for random 2-SAT on the local limit tree via stochastic fixed point equation:
     $\longrightarrow$ analysis of belief propagation algorithm for marginals.

Local changes to a given formula

Having expressed $\mathbb{E}[\log(Z(\mathbf{\Phi}) \vee 1)]$ as a sum of local changes, to analyse $\mu_{\mathbf{\Phi}}$, we next perform the following steps:

1. Analyse (joint) marginals on the local limit of $\mathbf{\Phi}$:
   $\longrightarrow$ Multitype Galton-Watson tree for formulas.
   - Establish decorrelation properties for random 2-SAT on the local limit tree:
     $\longrightarrow$ Gibbs uniqueness.
   - Characterize the root marginals for random 2-SAT on the local limit tree via stochastic fixed point equation:
     $\longrightarrow$ analysis of belief propagation algorithm for marginals.
2. Show that $\log(Z(\mathbf{\Phi}) \vee 1)/n$ concentrates about its mean.

$\mathscr{P}(0,1)$: set of Borel probability measures on $(0,1)$.
Define $\mathrm{BP}_d : \mathscr{P}(0,1) \to \mathscr{P}(0,1)$ as follows: Let $\boldsymbol{d}^+, \boldsymbol{d}^- \sim \mathrm{Po}(d/2)$ and $(\boldsymbol{\mu}_{\pi,i})_{i\geq 1}$ be a sequence of i.i.d. samples from $\pi \in \mathscr{P}(0,1)$ (all independent). Then

$$\mathrm{BP}_d(\pi) = \mathscr{L}\left(\frac{\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi,i}}{\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi,i} + \prod_{i=1}^{\boldsymbol{d}^+} \boldsymbol{\mu}_{\pi,i+\boldsymbol{d}^-}}\right).$$

$\mathscr{P}(0,1)$: set of Borel probability measures on $(0,1)$.
Define $\mathrm{BP}_d : \mathscr{P}(0,1) \to \mathscr{P}(0,1)$ as follows: Let $\boldsymbol{d}^+, \boldsymbol{d}^- \sim \mathrm{Po}(d/2)$ and $(\boldsymbol{\mu}_{\pi,i})_{i \geq 1}$ be a sequence of i.i.d. samples from $\pi \in \mathscr{P}(0,1)$ (all independent). Then

$$\mathrm{BP}_d(\pi) = \mathscr{L}\left(\frac{\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi,i}}{\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi,i} + \prod_{i=1}^{\boldsymbol{d}^+} \boldsymbol{\mu}_{\pi,i+\boldsymbol{d}^-}} \right).$$

Theorem (Achlioptas, Coja-Oghlan, Hahn-Klimroth, Lee, M., Penschuk, Zhou (2021))

For any $0 < d < 2$ *the limit* $\pi_d = \lim_{\ell \to \infty} \mathrm{BP}^{\ell}(\delta_{\frac{1}{2}})$ *exists and*

$$\frac{1}{n} \log Z(\boldsymbol{\Phi}) \xrightarrow{\mathbb{P}} \mathbb{E}\left[\log\left(\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi_d,i} + \prod_{i=1}^{\boldsymbol{d}^+} \boldsymbol{\mu}_{\pi_d,\boldsymbol{d}^-+i} \right) - \frac{d}{2} \log\left(1 - \boldsymbol{\mu}_{\pi_d,1} \boldsymbol{\mu}_{\pi_d,2}\right) \right].$$

Proposition

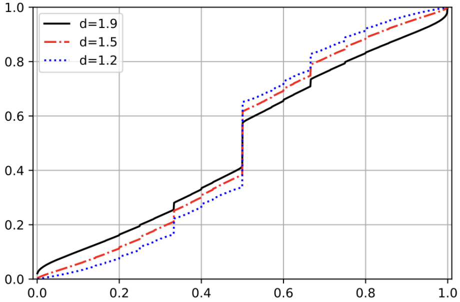
For any $0 < d < 2$, the random probability measure

$$\pi_{\mathbf{\Phi}} = \frac{1}{n} \sum_{i=1}^{n} \delta_{\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_i}=1)}$$

converges to π_d weakly in probability.

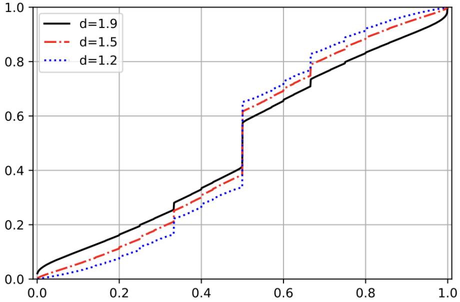
π_d corresponds to the asymptotic probability that a uniformly chosen variable within a uniformly random solution is set to 'true'.

How bad can the marginal structure get?



An approximation to the c.d.f. corresponding to π_d*, for* $d \in \{1.2, 1.5, 1.9\}$*.*

How bad can the marginal structure get?



An approximation to the c.d.f. corresponding to π_d*, for* $d \in \{1.2, 1.5, 1.9\}$*.*

'Complex' marginal structure arises from *inhomogeneity* among variable marginals: Variables are highly sensitive to imbalances in their local neighbourhood.

Let

$$\mathbb{A}_{\text{p.p.}} := \{x \in [0,1] : \pi_d(\{x\}) > 0\}$$

denote the pure point support of π_d.

Let

$$\mathbb{A}_{\text{p.p.}} := \{x \in [0,1] : \pi_d(\{x\}) > 0\}$$

denote the pure point support of π_d.

Theorem (M., Neininger, Zhu (2025+))

For any $d \in (0,2)$, the pure point support of π_d is

$$\mathbb{A}_{\text{p.p.}} = \mathbb{Q} \cap (0,1).$$

Let

$$\mathbb{A}_{\text{p.p.}} := \{x \in [0,1] : \pi_d(\{x\}) > 0\}$$

denote the pure point support of π_d.

Theorem (M., Neininger, Zhu (2025+))

For any $d \in (0,2)$, the pure point support of π_d is

$$\mathbb{A}_{\text{p.p.}} = \mathbb{Q} \cap (0,1).$$

Moreover:

- *For $d \in (0,1]$, π_d is a discrete measure;*
- *For $d \in (1,2)$, π_d has a non-trivial continuous part $\pi_{d,\text{c}}$ with supp$(\pi_{d,\text{c}}) = [0,1]$.*

Figure: Ralph Neininger

Figure: Haodong Zhu

- $\mathbb{Q} \cap (0,1)$ is a not too surprising subset of the pure point support: A uniformly chosen variable has asymptotically non-negligible probability to come from a *small component* (e.g. isolated vertex), such that its marginal still corresponds to a *proportion*.

- $\mathbb{Q} \cap (0,1)$ is a not too surprising subset of the pure point support: A uniformly chosen variable has asymptotically non-negligible probability to come from a *small component* (e.g. isolated vertex), such that its marginal still corresponds to a *proportion*.

- Less immediate: Irrespective of d, the pure point support of π_d contains *all* rational numbers in $(0,1)$, and a *non-trivial continuous part* $\pi_{d,\mathrm{c}}$ exists for $d \in (1,2)$.

Fluctuations

Having a 'law-of-large-numbers-type' result, can we derive the precise limiting distribution of a rescaled version of $\log Z(\mathbf{\Phi})$?

Fluctuations

Having a ‘law-of-large-numbers-type’ result, can we derive the precise limiting distribution of a rescaled version of $\log Z(\mathbf{\Phi})$?

In many previously studied random constraint satisfaction problems, the logarithm of the number of solutions **superconcentrates** for constraint densities up to the so-called condensation threshold (a phase transition that shortly precedes the satisfiability threshold):

It has only **bounded fluctuations**.

Example: Random k-SAT with regular literal degrees

Let $\tilde{\mathbf{\Phi}} = \tilde{\mathbf{\Phi}}_{n,m}$ be a **random k-CNF**
on n Boolean variables $x_1,\ldots,x_n$ with $m = 2dn/k$ clauses of length k,
where $k \mid 2dn$, defined as follows:

For each variable x_i, choose exactly d "positive" and d "negative" literal slots out of the km available literal slots (without replacement).

Example: Random k-SAT with regular literal degrees

Theorem (Coja-Oghlan, Wormald 2016)

There exists a constant k_0 such that for all $k \geq k_0$ and $d > 0$ such that $2d/k \leq 2k\ln 2 - k\ln 2/2 - 4$ the following is true. Let $q = q(k) \in (0,1)$ be the unique solution to the equation

$$2q = 1 - (1-q)^k.$$

Then there exists a random variable W with finite second moment such that as $n \to \infty$,

$$Z \cdot \frac{\left(4q(1-q)\right)^{dn}\sqrt{2+2(k-1)q-k}}{2^n\left(2q\right)^m} \xrightarrow{d} W.$$

More superconcentration

Superconcentration also occurs in

- **random k-XORSAT** up to the satisfiability threshold [Ayre, Coja-Oghlan, Gao, M. (2020)].
- **random graph q-coloring** up to the condensation threshold [Coja-Oghlan, Jaafari, Efthymiou, Kang, Kapetanopoulos (2018)].
- **random k-NAESAT** up to the condensation threshold [Coja-Oghlan, Kapetanopoulos, M. (2020)].
- **symmetric perceptron**, but with slightly different limiting distribution (log-normal with bounded variance) [Abbe, Li, Sly (2021)].

Fluctuations in random 2-SAT

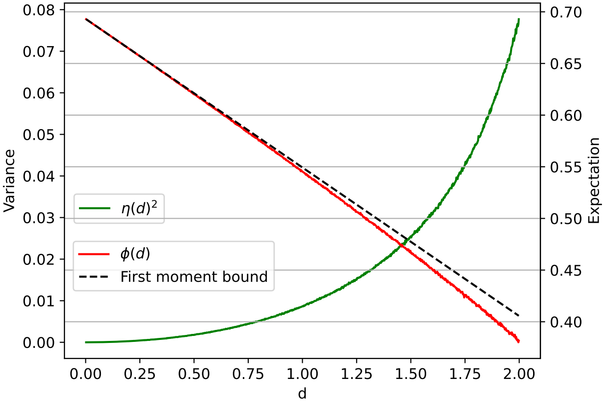
Theorem (Chatterjee, Coja-Oghlan, M., Riddlesden, Rolvien, Zakharov, Zhu (2025+))

For any $0 < d < 2$*, there exists* $\eta(d)^2 \in (0, \infty)$ *such that*

$$\frac{\log Z(\mathbf{\Phi}) - \mathbb{E}[\log Z(\mathbf{\Phi}) \mid Z(\mathbf{\Phi}) > 0]}{\sqrt{m}} \xrightarrow{d} \mathcal{N}\left(0, \eta(d)^2\right).$$





Red: The function $d \mapsto \phi(d)$.

Black: First moment bound $d \mapsto (1-d)\log 2 + \frac{d}{2}\log 3$.

Green: Approximation of the variance $\eta(d)^2$.

High-level proof idea

The variance

Consider the 'simpler' task of determining the asymptotics of the 'variance' of $\mathbf{\Phi}$.

For now, assume that $\hat{\mathbf{\Phi}}$ is some satisfiable modification of $\mathbf{\Phi}$:

$$\mathrm{Var}(\log Z(\hat{\mathbf{\Phi}})) = \mathbb{E}\left[\log Z(\hat{\mathbf{\Phi}})^2\right] - \mathbb{E}\left[\log Z(\hat{\mathbf{\Phi}})\right]^2.$$

The variance

Consider the 'simpler' task of determining the asymptotics of the 'variance' of $\mathbf{\Phi}$.

For now, assume that $\hat{\mathbf{\Phi}}$ is some satisfiable modification of $\mathbf{\Phi}$:

$$\mathrm{Var}(\log Z(\hat{\mathbf{\Phi}})) = \mathbb{E}\left[\log Z(\hat{\mathbf{\Phi}})^2\right] - \mathbb{E}\left[\log Z(\hat{\mathbf{\Phi}})\right]^2.$$

In particular, for two independent copies $\hat{\mathbf{\Phi}}_1, \hat{\mathbf{\Phi}}_2$ of $\hat{\mathbf{\Phi}}$,

$$\mathrm{Var}(\log Z(\hat{\mathbf{\Phi}})) = \mathbb{E}\left[\log Z(\hat{\mathbf{\Phi}}_1)^2\right] - \mathbb{E}\left[\log Z(\hat{\mathbf{\Phi}}_1) \log Z(\hat{\mathbf{\Phi}}_2)\right].$$

The variance

Consider the 'simpler' task of determining the asymptotics of the 'variance' of $\mathbf{\Phi}$.

For now, assume that $\hat{\mathbf{\Phi}}$ is some satisfiable modification of $\mathbf{\Phi}$:

$$\mathrm{Var}(\log Z(\hat{\mathbf{\Phi}})) = \mathbb{E}\left[\log Z(\hat{\mathbf{\Phi}})^2\right] - \mathbb{E}\left[\log Z(\hat{\mathbf{\Phi}})\right]^2.$$

In particular, for two independent copies $\hat{\mathbf{\Phi}}_1, \hat{\mathbf{\Phi}}_2$ of $\hat{\mathbf{\Phi}}$,

$$\mathrm{Var}(\log Z(\hat{\mathbf{\Phi}})) = \mathbb{E}\left[\log Z(\hat{\mathbf{\Phi}}_1)^2\right] - \mathbb{E}\left[\log Z(\hat{\mathbf{\Phi}}_1) \log Z(\hat{\mathbf{\Phi}}_2)\right].$$

$\longrightarrow$ Key idea (morally also employed in spin glass theory; see e.g. **Chen, Dey, Panchenko** (2017)):

Set up a family of correlated random formulas.

Setting up correlated formulas

For integers $M, M' \geq 0$ we construct a correlated pair $(\mathbf{\Phi}_1(M, M'), \mathbf{\Phi}_2(M, M'))$ of formulas on the same variable set $V_n = \{x_1, \ldots, x_n\}$ as follows:

Setting up correlated formulas

For integers $M, M' \geq 0$ we construct a correlated pair $(\mathbf{\Phi}_1(M, M'), \mathbf{\Phi}_2(M, M'))$ of formulas on the same variable set $V_n = \{x_1, \ldots, x_n\}$ as follows:

Let $(\boldsymbol{a}_i)_{i\geq 1}$, $(\boldsymbol{a}'_i)_{i\geq 1}$, $(\boldsymbol{a}''_i)_{i\geq 1}$ be sequences of mutually independent uniformly random clauses on V_n, and set

$$\mathbf{\Phi}_1(M, M') = \boldsymbol{a}_1 \wedge \cdots \wedge \boldsymbol{a}_M \wedge \boldsymbol{a}'_1 \wedge \cdots \wedge \boldsymbol{a}'_{M'},$$
$$\mathbf{\Phi}_2(M, M') = \boldsymbol{a}_1 \wedge \cdots \wedge \boldsymbol{a}_M \wedge \boldsymbol{a}''_1 \wedge \cdots \wedge \boldsymbol{a}''_{M'}.$$

$\mathbf{\Phi}_1(M, M')$ and $\mathbf{\Phi}_2(M, M')$ share clauses $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M$. Additionally, each contains another M' independent clauses.

Setting up correlated formulas

For integers $M, M' \geq 0$ we construct a correlated pair $(\mathbf{\Phi}_1(M, M'), \mathbf{\Phi}_2(M, M'))$ of formulas on the same variable set $V_n = \{x_1, \ldots, x_n\}$ as follows:

Let $(\boldsymbol{a}_i)_{i \geq 1}$, $(\boldsymbol{a}'_i)_{i \geq 1}$, $(\boldsymbol{a}''_i)_{i \geq 1}$ be sequences of mutually independent uniformly random clauses on V_n, and set

$$\mathbf{\Phi}_1(M, M') = \boldsymbol{a}_1 \wedge \cdots \wedge \boldsymbol{a}_M \wedge \boldsymbol{a}'_1 \wedge \cdots \wedge \boldsymbol{a}'_{M'},$$
$$\mathbf{\Phi}_2(M, M') = \boldsymbol{a}_1 \wedge \cdots \wedge \boldsymbol{a}_M \wedge \boldsymbol{a}''_1 \wedge \cdots \wedge \boldsymbol{a}''_{M'}.$$

$\mathbf{\Phi}_1(M, M')$ and $\mathbf{\Phi}_2(M, M')$ share clauses $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M$. Additionally, each contains another M' independent clauses.

In particular, $\mathbf{\Phi}_1(m, 0) = \mathbf{\Phi}_2(m, 0)$, while $\mathbf{\Phi}_1(0, m)$, $\mathbf{\Phi}_2(0, m)$ are independent.

Telescoping sum

Interpolating between the extreme cases, we can write a telescoping sum for the variance of $\hat{\mathbf{\Phi}}$:

$$\log Z(\hat{\mathbf{\Phi}}_1(m,0)) \cdot \log Z(\hat{\mathbf{\Phi}}_2(m,0)) - \log Z(\hat{\mathbf{\Phi}}_1(0,m)) \cdot \log Z(\hat{\mathbf{\Phi}}_2(0,m))$$

$$= \sum_{M=1}^{m} \log Z(\hat{\mathbf{\Phi}}_1(M,m-M)) \cdot \log Z(\hat{\mathbf{\Phi}}_2(M,m-M))$$

$$- \log Z(\hat{\mathbf{\Phi}}_1(M-1,m-M+1)) \cdot \log Z(\hat{\mathbf{\Phi}}_2(M-1,m-M+1)).$$

Telescoping sum

Interpolating between the extreme cases, we can write a telescoping sum for the variance of $\hat{\mathbf{\Phi}}$:

$$\log Z(\hat{\mathbf{\Phi}}_1(m,0)) \cdot \log Z(\hat{\mathbf{\Phi}}_2(m,0)) - \log Z(\hat{\mathbf{\Phi}}_1(0,m)) \cdot \log Z(\hat{\mathbf{\Phi}}_2(0,m))$$
$$= \sum_{M=1}^{m} \log Z(\hat{\mathbf{\Phi}}_1(M,m-M)) \cdot \log Z(\hat{\mathbf{\Phi}}_2(M,m-M))$$
$$- \log Z(\hat{\mathbf{\Phi}}_1(M-1,m-M+1)) \cdot \log Z(\hat{\mathbf{\Phi}}_2(M-1,m-M+1)).$$

Each summand on the r.h.s. corresponds to a local change of swapping a shared clause for a pair of independent clauses.

Taking expectations

Problem:
We are actually interested in

$$\log Z(\mathbf{\Phi}_1(m,0)) \cdot \log Z(\mathbf{\Phi}_2(m,0)) - \log Z(\mathbf{\Phi}_1(0,m)) \cdot \log Z(\mathbf{\Phi}_2(0,m))$$
$$= \sum_{M=1}^{m} \log Z(\mathbf{\Phi}_1(M,m-M)) \cdot \log Z(\mathbf{\Phi}_2(M,m-M))$$
$$- \log Z(\mathbf{\Phi}_1(M-1,m-M+1)) \cdot \log Z(\mathbf{\Phi}_2(M-1,m-M+1)),$$

but each $\mathbf{\Phi}_h(M,m-M)$ has a non-zero probability of being unsatisfiable.

Taking expectations

Problem:
We are actually interested in

$$\log Z(\mathbf{\Phi}_1(m,0)) \cdot \log Z(\mathbf{\Phi}_2(m,0)) - \log Z(\mathbf{\Phi}_1(0,m)) \cdot \log Z(\mathbf{\Phi}_2(0,m))$$
$$= \sum_{M=1}^{m} \log Z(\mathbf{\Phi}_1(M,m-M)) \cdot \log Z(\mathbf{\Phi}_2(M,m-M))$$
$$- \log Z(\mathbf{\Phi}_1(M-1,m-M+1)) \cdot \log Z(\mathbf{\Phi}_2(M-1,m-M+1)),$$

but each $\mathbf{\Phi}_h(M,m-M)$ has a non-zero probability of being unsatisfiable.

Solution:
Turn each $\mathbf{\Phi}_h(M,m-M)$ by a satisfiable version $\hat{\mathbf{\Phi}}_h(M,m-M)$ s.t. typically, $\log Z(\mathbf{\Phi}_h(M,m-M)), \log Z(\hat{\mathbf{\Phi}}_h(M,m-M))$ are close.
The construction of $\hat{\mathbf{\Phi}}$ is based on the Unit Clause Propagation algorithm.

Local changes in correlated formula pairs

Having expressed the variance of $\log Z(\hat{\mathbf{\Phi}})$ as a sum of local changes, to analyse these, we next perform the following steps:

1. Derive the local limit of pairs of correlated formulas:
   $\longrightarrow$ Multitype Galton-Watson tree for formula pairs.

Local changes in correlated formula pairs

Having expressed the variance of $\log Z(\hat{\mathbf{\Phi}})$ as a sum of local changes, to analyse these, we next perform the following steps:

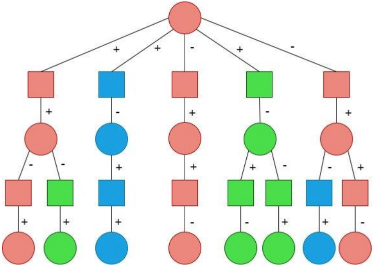
1. Derive the local limit of pairs of correlated formulas:
   $\longrightarrow$ Multitype Galton-Watson tree for formula pairs.
2. Establish decorrelation properties for random 2-SAT on the local limit tree:
   $\longrightarrow$ Gibbs uniqueness for formula pairs.

Local changes in correlated formula pairs

Having expressed the variance of $\log Z(\hat{\mathbf{\Phi}})$ as a sum of local changes, to analyse these, we next perform the following steps:

1. Derive the local limit of pairs of correlated formulas:
   $\longrightarrow$ Multitype Galton-Watson tree for formula pairs.
2. Establish decorrelation properties for random 2-SAT on the local limit tree:
   $\longrightarrow$ Gibbs uniqueness for formula pairs.
3. Characterize the root marginals for random 2-SAT on the local limit tree via stochastic fixed point equation:
   $\longrightarrow$ analysis of belief propagation algorithm for marginals in formula pairs.

Multitype Galton-Watson tree



Visualization of the local limit of a pair of correlated random 2-SAT formulas.

The variance formula

Let $\mathscr{P}(\mathbb{R}^2)$ be the set of all (Borel) probability measures on $\mathbb{R}^2$.
For $0 < d < 2$ and $0 \le t \le 1$ we define an operator

$$\mathtt{logBP}^{\otimes}_{d,t} : \mathscr{P}\left(\mathbb{R}^2\right) \to \mathscr{P}\left(\mathbb{R}^2\right), \qquad \rho \mapsto \hat{\rho} = \mathtt{logBP}^{\otimes}_{d,t}(\rho),$$

as follows.

The variance formula

Let

$$(\boldsymbol{\xi}_{\rho,i})_{i\geq 1}, (\boldsymbol{\xi}'_{\rho,i})_{i\geq 1}, (\boldsymbol{\xi}''_{\rho,i})_{i\geq 1}, \quad \boldsymbol{\xi}_{\rho,i} = \begin{pmatrix} \boldsymbol{\xi}_{\rho,i,1} \\ \boldsymbol{\xi}_{\rho,i,2} \end{pmatrix}, \boldsymbol{\xi}'_{\rho,i} = \begin{pmatrix} \boldsymbol{\xi}'_{\rho,i,1} \\ \boldsymbol{\xi}'_{\rho,i,2} \end{pmatrix}, \boldsymbol{\xi}''_{\rho,i} = \begin{pmatrix} \boldsymbol{\xi}''_{\rho,i,1} \\ \boldsymbol{\xi}''_{\rho,i,2} \end{pmatrix}$$

be random vectors with distribution ρ, let $\boldsymbol{d} \stackrel{\text{dist}}{=} \text{Po}(td)$, $\boldsymbol{d}', \boldsymbol{d}'' \stackrel{\text{dist}}{=} \text{Po}((1-t)d)$ and let $\boldsymbol{s}_i, \boldsymbol{s}'_i, \boldsymbol{s}''_i, \boldsymbol{r}_i, \boldsymbol{r}'_i, \boldsymbol{r}''_i$ for $i \geq 1$ be uniformly random on $\{\pm 1\}$, all mutually independent.

The variance formula

Let

$$(\boldsymbol{\xi}_{\rho,i})_{i\geq 1},\ (\boldsymbol{\xi}'_{\rho,i})_{i\geq 1},\ (\boldsymbol{\xi}''_{\rho,i})_{i\geq 1},\quad \boldsymbol{\xi}_{\rho,i}=\begin{pmatrix}\boldsymbol{\xi}_{\rho,i,1}\\ \boldsymbol{\xi}_{\rho,i,2}\end{pmatrix},\ \boldsymbol{\xi}'_{\rho,i}=\begin{pmatrix}\boldsymbol{\xi}'_{\rho,i,1}\\ \boldsymbol{\xi}'_{\rho,i,2}\end{pmatrix},\ \boldsymbol{\xi}''_{\rho,i}=\begin{pmatrix}\boldsymbol{\xi}''_{\rho,i,1}\\ \boldsymbol{\xi}''_{\rho,i,2}\end{pmatrix}$$

be random vectors with distribution ρ, let $\boldsymbol{d}\overset{\mathrm{dist}}{=}\mathrm{Po}(td)$, $\boldsymbol{d}',\boldsymbol{d}''\overset{\mathrm{dist}}{=}\mathrm{Po}((1-t)d)$ and let $\boldsymbol{s}_i,\boldsymbol{s}'_i,\boldsymbol{s}''_i,\boldsymbol{r}_i,\boldsymbol{r}'_i,\boldsymbol{r}''_i$ for $i\geq 1$ be uniformly random on $\{\pm 1\}$, all mutually independent.

Then $\hat{\rho}$ is the distribution of the vector

$$\begin{pmatrix}\sum_{i=1}^{\boldsymbol{d}}\boldsymbol{s}_i\log\left(\frac{1}{2}\left(1+\boldsymbol{r}_i\tanh(\boldsymbol{\xi}_{\rho,i,1}/2)\right)\right)+\sum_{i=1}^{\boldsymbol{d}'}\boldsymbol{s}'_i\log\left(\frac{1}{2}\left(1+\boldsymbol{r}'_i\tanh(\boldsymbol{\xi}'_{\rho,i,1}/2)\right)\right)\\ \sum_{i=1}^{\boldsymbol{d}}\boldsymbol{s}_i\log\left(\frac{1}{2}\left(1+\boldsymbol{r}_i\tanh(\boldsymbol{\xi}_{\rho,i,2}/2)\right)\right)+\sum_{i=1}^{\boldsymbol{d}''}\boldsymbol{s}''_i\log\left(\frac{1}{2}\left(1+\boldsymbol{r}''_i\tanh(\boldsymbol{\xi}''_{\rho,i,2}/2)\right)\right)\end{pmatrix}.$$

The variance formula

For any $0 < d < 2$, $t \in [0,1]$ there exists a unique probability measure $\rho_{d,t} \in \mathscr{P}(\mathbb{R}^2)$ such that

$$\rho_{d,t} = \texttt{logBP}^{\otimes}_{d,t}(\rho_{d,t}) \qquad \text{and} \qquad \int_{\mathbb{R}^2} \|\xi\|_2^2 \mathrm{d}\rho_{d,t}(\xi) < \infty.$$

The variance formula

For any $0 < d < 2$, $t \in [0,1]$ there exists a unique probability measure $\rho_{d,t} \in \mathscr{P}(\mathbb{R}^2)$ such that

$$\rho_{d,t} = \texttt{logBP}^{\otimes}_{d,t}(\rho_{d,t}) \qquad \text{and} \qquad \int_{\mathbb{R}^2} \|\xi\|_2^2 \mathrm{d}\rho_{d,t}(\xi) < \infty.$$

In addition, define a function $\mathscr{B}^{\otimes}_{d,t} : \mathscr{P}(\mathbb{R}^2) \to (0,\infty]$ by letting

$$\mathscr{B}^{\otimes}_{d,t}(\rho) = \mathbb{E}\left[\prod_{h=1}^{2} \log\left(1 - \frac{1}{4}(1 + \boldsymbol{r}_1 \tanh(\boldsymbol{\xi}_{\rho,1,h}/2))(1 + \boldsymbol{r}_2 \tanh(\boldsymbol{\xi}_{\rho,2,h}/2))\right)\right].$$

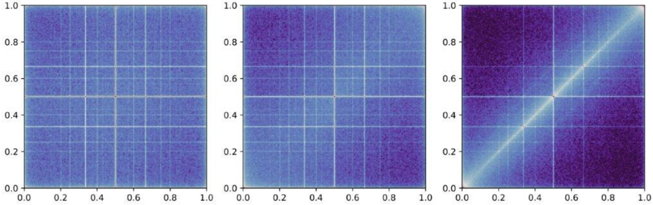
The variance formula

Theorem

We have $\eta(d) > 0$ *and* $\mathrm{Var}\log Z(\hat{\mathbf{\Phi}}) \sim m \cdot \eta_d^2$*, where*

$$\eta(d)^2 = \int_0^1 \mathscr{B}^{\otimes}_{d,t}(\rho_{d,t})\mathrm{d}t - \mathscr{B}^{\otimes}_{d,0}(\rho_{d,0}) \in (0, \infty).$$

Visualization of (a function of) $\rho_{d,t}$



Visualization of (a function of) $\rho_{d,t}$ for $d = 1.9$ and different values of t: $t = 0.1, 0.5, 0.9$ (left to right).

As t increases, the correlations between the two coordinates of the random vector increase (brighter diagonal).

From increments to CLT

Overall proof approach:

Combine techniques from variance computation with a generic martingale CLT.

From increments to CLT

Overall proof approach:

Combine techniques from variance computation with a generic martingale CLT.

For $0 \le M \le m_n$, set

$$\boldsymbol{Z}_{n,M} = \frac{\mathbb{E}\left[\log Z(\hat{\boldsymbol{\Phi}}) \mid \boldsymbol{a}_1, \ldots, \boldsymbol{a}_M\right]}{\sqrt{m}}.$$

Then for any fixed n, $(\boldsymbol{Z}_{n,M})_{0 \le M \le m_n}$ is a martingale (**clause-exposure Doob martingale**).

From increments to CLT

Overall proof approach:

Combine techniques from variance computation with a generic martingale CLT.

For $0 \le M \le m_n$, set

$$\boldsymbol{Z}_{n,M} = \frac{\mathbb{E}\left[\log Z(\hat{\boldsymbol{\Phi}}) \mid \boldsymbol{a}_1, \ldots, \boldsymbol{a}_M\right]}{\sqrt{m}}.$$

Then for any fixed n, $(\boldsymbol{Z}_{n,M})_{0 \le M \le m_n}$ is a martingale (**clause-exposure Doob martingale**).

Let $\boldsymbol{X}_{n,M} = \boldsymbol{Z}_{n,M} - \boldsymbol{Z}_{n,M-1}$ be its martingale differences.

The martingale differences

Also the squared martingale differences $\boldsymbol{X}_{n,M}^2$ can be related to the operation of exchanging common for independent clauses in pairs of correlated formulas:

The martingale differences

Also the squared martingale differences $\boldsymbol{X}_{n,M}^2$ can be related to the operation of exchanging common for independent clauses in pairs of correlated formulas:

$$\boldsymbol{\Delta}(M) = \log\left(\frac{Z(\hat{\boldsymbol{\Phi}}_1(M, m-M))}{Z(\hat{\boldsymbol{\Phi}}_1(M-1, m-M))}\right) \cdot \log\left(\frac{Z(\hat{\boldsymbol{\Phi}}_2(M, m-M))}{Z(\hat{\boldsymbol{\Phi}}_2(M-1, m-M))}\right),$$

$$\boldsymbol{\Delta}'(M) = \log\left(\frac{Z(\hat{\boldsymbol{\Phi}}_1(M-1, m-M+1))}{Z(\hat{\boldsymbol{\Phi}}_1(M-1, m-M))}\right) \cdot \log\left(\frac{Z(\hat{\boldsymbol{\Phi}}_2(M-1, m-M+1))}{Z(\hat{\boldsymbol{\Phi}}_2(M-1, m-M))}\right),$$

$$\boldsymbol{\Delta}''(M) = \log\left(\frac{Z(\hat{\boldsymbol{\Phi}}_1(M, m-M))}{Z(\hat{\boldsymbol{\Phi}}_1(M-1, m-M))}\right) \cdot \log\left(\frac{Z(\hat{\boldsymbol{\Phi}}_2(M-1, m-M+1))}{Z(\hat{\boldsymbol{\Phi}}_2(M-1, m-M))}\right).$$

Lemma

We have $m_n \boldsymbol{X}_M^2 = \mathbb{E}\left[\boldsymbol{\Delta}(M) + \boldsymbol{\Delta}(M)' - 2\boldsymbol{\Delta}''(M) \mid \boldsymbol{a}_1, \ldots, \boldsymbol{a}_M\right]$.

The martingale differences

⟶ Using ideas and techniques from the variance computation, we show the following:

Proposition

For all $0 < d < 2$ *the martingale array* $(\boldsymbol{Z}_{n,M})_{n \geq 1, 0 \leq M \leq m_n}$ *satisfies*

$$\lim_{n\to\infty} \mathbb{E}\left[\max_{1\leq M\leq m} |\boldsymbol{X}_{n,M}|\right] = 0 \qquad \textit{and} \qquad \lim_{n\to\infty} \mathbb{E}\left|\eta(d)^2 - \sum_{M=1}^{m} \boldsymbol{X}_{n,M}^2\right| = 0.$$

General martingale CLT

Theorem (Hall & Heyde, Theorem 3.2)

Let $(Z_{n,i}, \mathfrak{F}_{n,i})_{0 \le i \le m_n, n \ge 1}$ *be a zero-mean, square-integrable martingale array with differences* $X_{n,i} = Z_{n,i} - Z_{n,i-1}$ *for* $1 \le i \le m_n$*. Assume that there exists a constant* η^2 *such that*

$$\lim_{n \to \infty} \max_{1 \le i \le m_n} |X_{n,i}| = 0 \qquad \text{in probability,}$$

$$\lim_{n \to \infty} \sum_{i=1}^{m_n} X_{n,i}^2 = \eta^2 \qquad \text{in probability,}$$

$$\mathbb{E}\left[\max_{1 \le i \le m_n} X_{n,i}^2\right] \qquad \text{is bounded in } n.$$

Then Z_{n,m_n} *converges in distribution to a Gaussian random variable with mean zero and variance* η^2*.*

Take away

- The **satisfiability threshold** for random 2-SAT can be determined by a first and second moment analysis in the associated random digraph.
- The **logarithm of the number of solutions** in random 2-SAT, normalized by n, converges to a constant that matches the predictions from statistical physics.
- The logarithm of the number of solutions in random 2-SAT **does not superconcentrate**, which is different from previously known behaviour of other random CSPs.
- The proof of the last result does not proceed via moment analysis, but via the study of **pairs of correlated random formulas**.