Fluctuations of random 2-SAT

Arnab Chatterjee, Amin Coja-Oghlan, Noëla Müller,
Connor Riddlesden, Maurice Rolvien, Pavel Zakharov, Haodong Zhu

technische universität dortmund

Introduction

- 2-CNF is a boolean formula, written as $(l_1 \vee l_2) \wedge \ldots \wedge (l_{2m-1} \vee l_{2m})$, where l_i is either a variable or a negation of a variable.
- Satisfiable assignment maps each variable to a value, such that in each clause there is at least one true literal.
- Consequently, random 2-CNF consists of n variables and $m = \lfloor \frac{dn}{2} \rfloor$ uniformly chosen clauses. Here d can be viewed as an average number of clauses which contain any chosen variable.

Known results

1 The well-known result [1] states that $d = 2$ is a threshold for satisfiability of a random 2-CNF.

In other words, for any $\varepsilon > 0$ probability that random 2-CNF is satisfiable tends to one if $d < 2 - \varepsilon$ and to zero if $d < 2 + \varepsilon$, as $n \to +\infty$.

2 Another question is how much solution does the random 2-CNF have below the satisfiability threshold (e.g. $d < 2$). The first-order approximation was given by Coja-Oghlan et al. [2]. By $\mathbf{\Phi_n}$ define a random formula n variables and $m = \lfloor \frac{dn}{2} \rfloor$ and by Z the number of the solutions of such formula. Then the aforementioned work proves that

$$\frac{\log Z(\mathbf{\Phi_n})}{n} \xrightarrow{\mathsf{P}} \mu_d, \tag{1}$$

where μ_d is a constant which does not depend on n.

Our contribution

Actually, we can say even more about the number of solutions of a random 2-CNF.

Theorem 1 ([3])

$$\frac{\log Z(\mathbf{\Phi_n}) - \mathsf{E} \log Z(\mathbf{\Phi_n})}{\eta_d \sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1), \tag{2}$$

where $\eta_d \geq 0$ does not depend on n.

References

[1] Andreas Goerdt. A threshold for unsatisfiability. *Journal of Computer and System Sciences*, 53(3):469–486, 1996.

[2] D. Achlioptas, A. Coja-Oghlan, M. Hahn-Klimroth, J. Lee, N. Müller, M. Penschuck, and G. Zhou. The number of satisfying assignments of random 2-sat formulas. *Random Structures & Algorithms*, 58(4):609–647, 2021.

[3] A. Chatterjee, A. Coja-Oghlan, N. Müller, C. Riddlesden, M. Rolvien, P. Zakharov, and H. Zhu. The number of random 2-sat solutions is asymptotically log-normal, 2024.

Proof idea

Two main questions are:

1. How to show assymptotic normality?
2. How to calculate the variance?

First part is just an application of a martingale central limit theorem, however the variance calculation is a bit more tricky.

For $M \in \{0, \ldots, m\}$ let us generate a pair of random 2-CNFs $(\mathbf{\Phi}_1^{(M)}, \mathbf{\Phi}_2^{(M)})$. First we generate M random clauses and add them to both formulas. Second we generate two independent sets of $m - M$ clauses and add them to corresponding formulas.

It is easy to see that $(\mathbf{\Phi}_1^{(0)}, \mathbf{\Phi}_2^{(0)})$ are two independent random formulas, and $(\mathbf{\Phi}_1^{(M)}, \mathbf{\Phi}_2^{(M)})$ is a pair of copies of a random 2-CNF.

Therefore,

$$\mathrm{Var} \log Z(\mathbf{\Phi_n}) = \mathsf{E}\left[\log Z(\mathbf{\Phi}_1^{(m)}) \log Z(\mathbf{\Phi}_2^{(m)})\right] - \mathsf{E}\left[\log Z(\mathbf{\Phi}_1^{(0)}) \log Z(\mathbf{\Phi}_2^{(0)})\right] \tag{3}$$

$$= \sum_{M=0}^{m-1} \mathsf{E}\left[\log Z(\mathbf{\Phi}_1^{(M+1)}) \log Z(\mathbf{\Phi}_2^{(M+1)})\right] - \mathsf{E}\left[\log Z(\mathbf{\Phi}_1^{(M)}) \log Z(\mathbf{\Phi}_2^{(M)})\right]. \tag{4}$$

Some intuition

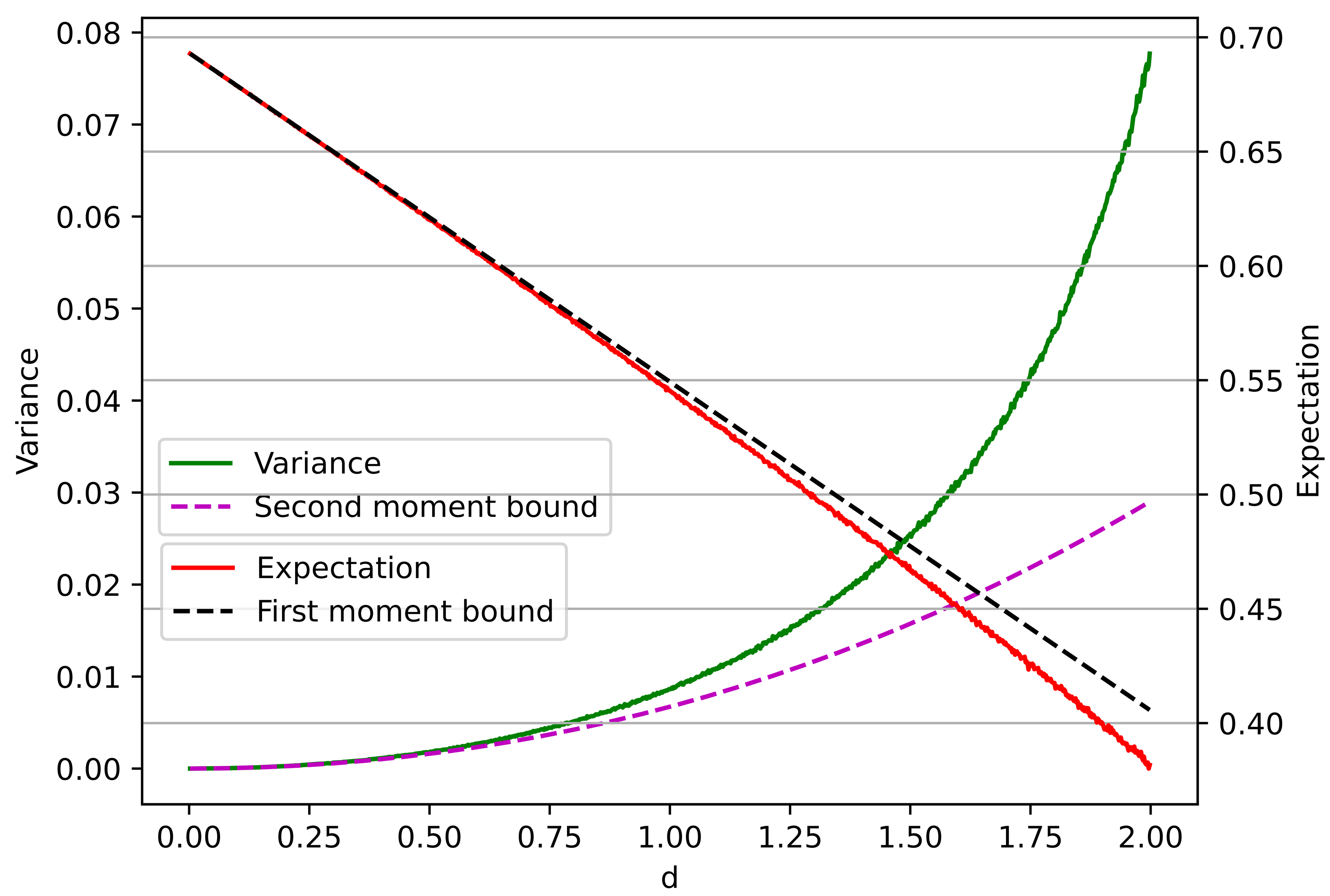

Figure 1: Plots of μ_d and η_d as well as their first/second moment estimates;

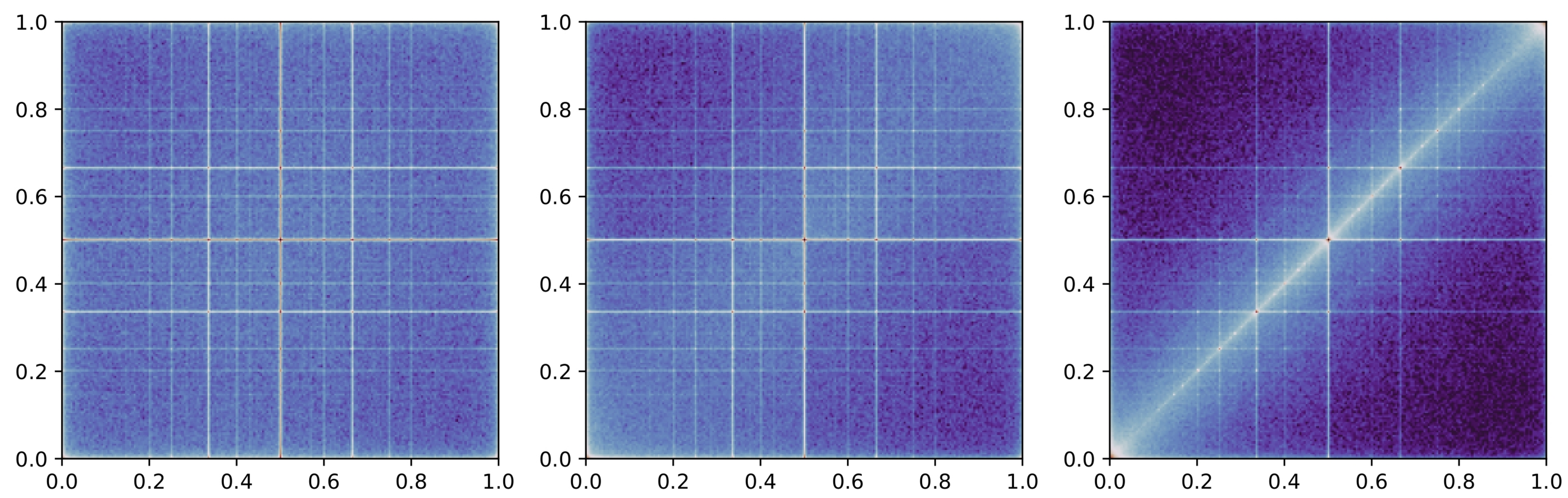

Figure 2: Marginal distribution on two correlated formulas for $d = 1.9$ and $M = 0.1m$, $0.5m$, $0.9m$

To understand the second figure we should generate a pair of random formulas and take a uniformly random pair of satisfying assignments. The joint distribution of any of n coordinates can be viewed on the heatmaps: almost independet forlumas on the left and highly correlated formulas on the right.