Solutions for Week 10

Bruno Loureiro and Luca Saglietti

20.05.2021

1 Exercise 11.1

The random-subcube model is defined by its solution space $S \subset \{0,1\}^N$ (not by a graphical model). We define S as the union of $\lfloor 2^{(1-\alpha)N} \rfloor$ random clusters (where $\lfloor x \rfloor$ denotes the integer value of x). A random cluster A being defined as:

$$A = \{\boldsymbol{\sigma} | \sigma_i \in \pi_i^A, \quad \forall\, i \in \{1, \dots, N\}\} \tag{1.1}$$

where π^A is a random mapping:

$$\begin{aligned} \pi^A \colon \{1, \dots, N\} &\to \{\{0\}, \{1\}, \{0,1\}\} \\ i &\mapsto \pi_i^A \end{aligned} \tag{1.2}$$

such that for each variable i, $\pi_i^A = \{0\}$ with probability $p/2$, $\{1\}$ with probability $p/2$, and $\{0,1\}$ with probability $1-p$. A cluster is thus a random subcube of $\{0,1\}^N$. If $\pi_i^A = \{0\}$ or $\{1\}$, variable i is said "frozen" in A; otherwise it is said "free" in A. One given configuration $\boldsymbol{\sigma}$ might belong to zero, one or several clusters. A "solution" belongs to at least one cluster.

We will analyze the properties of this model in the limit $N \to \infty$, the two parameters α and p being fixed and independent of N. The internal entropy s of a cluster A is defined as $\frac{1}{N}\log_2\{|A|\}$, i.e. the fraction of free variables in A. We also define complexity $\Sigma(s)$ as the (base 2) logarithm of the number of clusters of internal entropy s per variable (i.e. divide by N).

Comments: Note that in a usual constraint satisfaction problem (e.g. graph colouring), we are given the space in which the variables live (a.k.a. the configuration space: $\boldsymbol{\sigma} \in \{1, \cdots, q\}^N$ for graph colouring, where q is the number of colours) and a set of constraints $f(\boldsymbol{\sigma})$ that the variables need to satisfy (e.g. two neighbouring variables cannot share the same colour). In the Statistical Physics approach, we define a Gibbs-Boltzmann measure $\mu_\beta(\boldsymbol{\sigma})$ over the space of configurations that enforce the constraints (and sometimes we soften them, by introducing a "temperature" parameter $\beta > 0$). Sampling from μ_β therefore is equivalent to sampling from the solution space $\mathcal{S}$, which is the set of all configurations satisfying the constraint.

In the Random Subcube Model (RSM), the configuration space is an N dimensional hypercube $\boldsymbol{\sigma} \in \{0,1\}^N$. However, instead of giving a set of constraints over the configurations, we define the solution space by directly specifying the solution space $\mathcal{S}$. This is defined as a union $\mathcal{S} = \bigcup_{\mu=1}^M A_\mu$ of $M = \lfloor 2^{(1-\alpha)N} \rfloor$ subcubes $A_\mu \subset \{0,1\}^N$ which are sampled identically and independently. Therefore, the Gibbs-Boltzmann measure is simply the uniform distribution over $\mathcal{S}$. More specifically, each subcube A_μ composing the solution space is generated as follows:

1. Draw a vector $\boldsymbol{\pi}^\mu \in \{0, 1, \{0,1\}\}^N$ with components π_i i.i.d. on:

$$\pi_i^\mu = \begin{cases} 0 & \text{with probability } p/2 \\ 1 & \text{with probability } p/2 \\ \{0,1\} & \text{with probability } 1-p \end{cases} \tag{1.3}$$

2. Given $\boldsymbol{\pi}^\mu$, the subcube $A_\mu \subset \{0,1\}^N$ is constructed by fixing $\sigma_i = 0,1$ if $\pi_i^\mu = 0,1$. These variables are called "frozen" since they are fixed by the random map. If $\pi_i^\mu = \{0,1\}$, then σ_i can be 0 or 1. These variables are called free, since they are free to take any value.

Note that a given configuration $\boldsymbol{\sigma}$ can belong to more than one subcube A_μ. If it belongs to at least one, it is in the solution set $\mathcal{S}$. In summary, the parameter $\alpha \in [0,1)$ fixes the number of subcubes in the model, and the parameter p determine the size of the subcubes.

(a) What is the analog of the satisfiability threshold α_s in this model?

Solution: By definition, the satisfiability threshold is the value α_s such that for $\alpha > \alpha_s$ there are no solutions, i.e. no configuration satisfying all the constraints. In the RSM, a solution is a configuration $\boldsymbol{\sigma}$ that belongs to at least one cluster A, and the size of each cluster is $|A| = 2^{\#\text{ free variables}} \geq 1$. Therefore, the SCM has no solution if and only if there are no clusters, i.e. $\alpha > 1$ such that $(1-\alpha) < 0$ and $S = \emptyset$. Hence, $\alpha_s = 1$.

(b) Compute the α_d threshold below which most configurations belong to at least one cluster.

Solution: Let's start by computing the probability that a configuration $\boldsymbol{\sigma} \in \{0,1\}^N$ belongs to a random cluster A:

$$\mathbb{P}(\boldsymbol{\sigma} \in A) \overset{(a)}{=} \prod_{i=1}^N \mathbb{P}\left(\sigma_i \in \pi_i^A\right) \overset{(b)}{=} \prod_{i=1}^N \mathbb{P}\left(\pi_i^A \in \{\{\sigma_i\},\{0,1\}\}\right) = \prod_{i=1}^N \left(1 - p + \frac{p}{2}\right) = \left(1 - \frac{p}{2}\right)^N$$

where in (a) we used that each component of the configuration is distributed independently and in (b) we used that the probability for σ_i to be the image of a random map π_i^A is equivalent to the probability that the random map π_i^A takes value σ_i or $\{0,1\}$. Therefore a given configuration $\boldsymbol{\sigma}$ belongs on average to:

$$\begin{aligned} \mathbb{E}|\{A|\boldsymbol{\sigma} \in A\}| &= (\#\text{ of clusters with at least one } \boldsymbol{\sigma}) \times (\text{prob. } \boldsymbol{\sigma} \text{ belongs to } A) \\ &= 2^{(1-\alpha)N}\left(1 - \frac{p}{2}\right)^N = 2^{N\left(1-\alpha+\log_2\left(1-\frac{p}{2}\right)\right)} = 2^{N(\log_2(2-p)-\alpha)} \end{aligned} \tag{1.4}$$

Therefore, as $N \to \infty$ we can identify two regimes: a) $\alpha > \log_2(2-p)$ the average above goes to zero exponentially in N, meaning that a given configuration on average doesn't belong to any cluster; b) $\alpha < \log_2(2-p)$ then the average grows exponentially with N, meaning that on average a given configuration belong to many clusters. In particular, this means that all configurations belong at least to one cluster, and therefore with high probability all configurations are solutions. Thus, the clustering threshold $\alpha_d = \log_2(2-p)$, and for $\alpha < \alpha_d$ we have $s_\mathrm{tot} = \frac{1}{N}\log_2(|S|) = 1$.

(c) For $\alpha > \alpha_\mathrm{d}$ write the expression for the complexity $\Sigma(s)$ as a function of the parameters p and α. Compute the total entropy defined as $s_\mathrm{tot} = \max_s\{\Sigma(s) + s|\Sigma(s) \geq 0\}$. Observe that there are two regimes in the interval $\alpha \in (\alpha_\mathrm{d}, 1)$, discuss their properties and write the value of the "condensation" threshold α_c.

Solution: To compute $\Sigma(s)$, we need to count the number of clusters with a given internal entropy s. Let $s(A) = \frac{1}{N}\log_2(|A|)$ denote the internal entropy of cluster A, and recall that this is the fraction of free variables in A. The probability that a cluster A has entropy s given by:

$$\begin{aligned} \mathbb{P}(s(A) = s) &= (\#\text{ possible free variables}) \times (\text{prob. frozen})^{\#\text{ frozen}} \times (\text{prob. free})^{\#\text{ free}} \\ &= \binom{N}{Ns} p^{(1-s)N}(1-p)^{Ns} \equiv \rho(s) \end{aligned} \tag{1.5}$$

We are now in a position to compute $\Sigma(s)$. Let $\mathcal{N}(s)$ be the number of clusters with internal entropy s, and recall that $\Sigma(s) = \frac{1}{N}\log_2 \mathcal{N}(s)$. Note that $\mathcal{N}(s)$ and $\Sigma(s)$ are not deterministic quantities: they are random variables. But if we are lucky, as $N \to \infty$ we can show that even though $\mathcal{N}(s)$ fluctuates, $\Sigma(s)$ concentrates around its mean - just as we typically do for the free entropy. Indeed, this is the case: it is not hard to see that $\mathcal{N}(s)$ is a binomial distribution $\text{Binom}\left(2^{(1-\alpha)N}, \rho(s)\right)$, and therefore $\Sigma(s)$ is simply the entropy of this binomial distribution:

$$\Sigma(s) = \begin{cases} 1-\alpha - H(s, 1-p) & \text{if } H(s,1-p) \leq 1-\alpha \\ -\infty & \text{otherwise} \end{cases} \tag{1.6}$$

where $H(x, y)$ is given by:

$$H(x,y) = x\log_2\left(\frac{x}{y}\right) - (1-x)\log_2\left(\frac{1-x}{1-y}\right) \tag{1.7}$$

Now consider the regime $\alpha > \alpha_d$. The total entropy can be computed using Laplace's method:

$$\sum_A 2^{Ns(A)} \approx N \int \mathrm{d}s \; 2^{N(\Sigma(s)+s)}\mathbb{I}\left(\Sigma(s) \geq 0\right) \asymp 2^{N(\Sigma(s^\star)+s^\star)} \tag{1.8}$$

where $s^\star = \text{argmin}_s \left\{\Sigma(s) + s|\Sigma(s) \geq 0\right\}$. Note that even though this seems to over count the solutions, since a configuration can belong to one or more clusters and this sum just add the cluster sizes, in the regime $\alpha > \alpha_d$ we have shown that the fraction of solutions belonging to more than one cluster is exponentially small. To solve the minimisation problem, search for zero derivative points:

$$\frac{\partial \Sigma}{\partial s} = -1 \qquad \Leftrightarrow \qquad s^0 = 2\frac{1-p}{2-p}, \qquad \Sigma(s^0) = \frac{p}{2-p} - \alpha + \log_2(2-p) \tag{1.9}$$

We can distinguish two cases: a) when $\Sigma(s^0) \geq 0$, i.e. $\alpha < \alpha_c \equiv \frac{p}{2-p} + \log_2(2-p)$, then the minimiser is given by $s^\star = s^0$; b) when $\Sigma(s^0) \leq 0$, i.e. $\alpha > \alpha_c$, the correct minimiser is $s^\star = s_M = \max_s\{s|\Sigma(s) \geq 0\}$. Therefore:

$$s_{\text{tot}} = \begin{cases} 1-\alpha + \log_2(1-p) & \text{for } \alpha \leq \alpha_c \\ s_M & \text{for } \alpha > \alpha_c \end{cases} \tag{1.10}$$

Summarising, we have four phases:

Liquid phase: $\alpha < \alpha_d$, almost all configurations are solutions, and $s_{\text{tot}} = 1$.

Clustered phase: $\alpha_d < \alpha < \alpha_c$, the solutions set S is partitioned into exponentially many non-overlapping clusters. Most of solutions are in the $e^{N\Sigma(\tilde{s})}$ with internal entropy $\tilde{s}$.

Condensed clustered phase: $\alpha_c < \alpha < \alpha_s$, the solution set S is partitioned into exponentially many non-overlapping clusters. However, most solutions are in clusters with entropy s_M, but their number is not exponentially large since $\Sigma(s_M) = 0$.

Unsatisfiable phase: $\alpha > \alpha_s$, there are no clusters, and therefore no solutions.

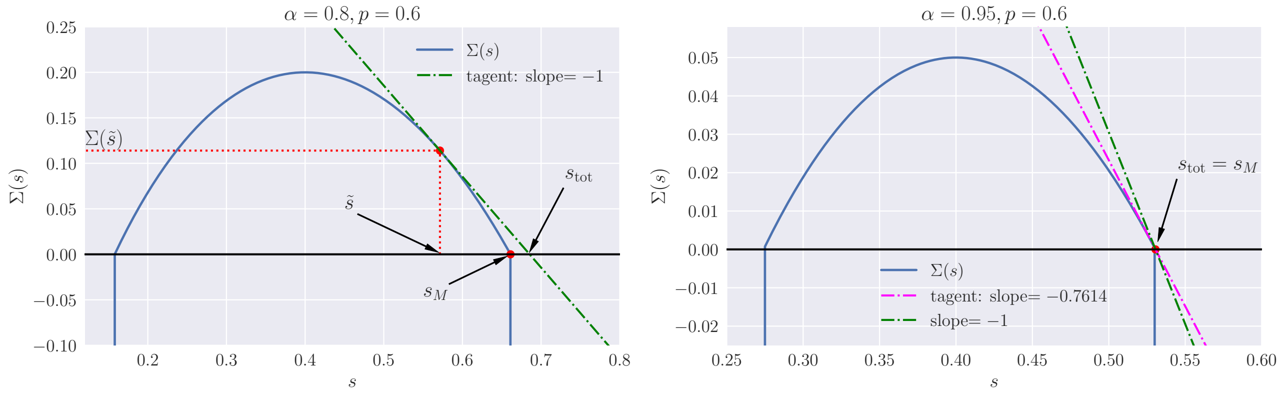

Figure 1: Complexity function $\Sigma(s)$ under clustered phase (left, $\alpha = 0.8, p = 0.6$) and condensed clustered phase (right, $\alpha = 0.95, p = 0.6$)

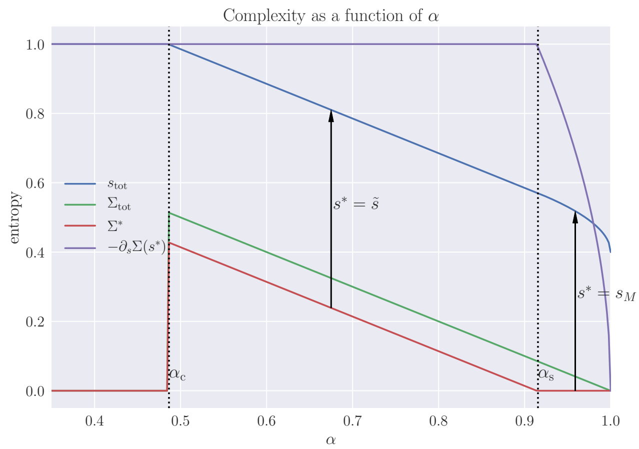

Figure 2: Complexity as a function of α for $p = 0.6$.