Multiple ζ^* values in the one dimensional ERAP with stretched-exponentially distributed points

Matteo D'Achille

Short talk for the "ALÉA Days 2021"
Centre international de rencontres mathématiques
Marseilles Luminy (online)

Based on a joint work with Andrea Sportiello (CNRS, UParis-Nord)

Thursday, March 18, 2021, 14h30 CET

The Euclidean Random Assignment Problem (ERAP)

Let $\mathcal{B} = (B_1, \ldots, B_n)$ (blues) and $\mathcal{R} = (R_1, \ldots, R_n)$ (reds) be two n-samples of i.i.d. r.v.s. with common pdf $\rho : \Omega \to \mathbb{R}$, where $(\Omega, \mathcal{D})$ is a metric space (in the following we shall consider only the case of an **Euclidean** space equipped with $\mathcal{D}$ **Euclidean** distance). For $p \in \mathbb{R}$ and a permutation π consider the *hamiltonian*

$$\mathcal{H}(\pi) = \sum_{i=1}^{n} \mathcal{D}^p(B_i, R_{\pi(i)})$$

and the r.v. $\mathcal{H}^{(p)}_{\mathrm{opt},(n,d)} = \min_\pi \mathcal{H}(\pi)$ (and $\pi_{\mathrm{opt}} = \arg\min \mathcal{H}(\pi)$). Interest has developed around $E_{p,d}(n) := \mathbb{E}[\mathcal{H}^{(p)}_{\mathrm{opt},(n,d)}]$ as $n \to \infty$.

Some motivations :

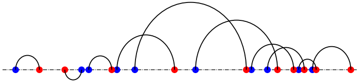
- Finding π_{opt} is **easy** (the assignment problem is P-complete);
- ERAP is the **Monge-Kantorovich problem** among the empirical measures of $\rho_{\mathcal{B}}$ and $\rho_{\mathcal{R}}$;
- It is a nice model of finite-dimensional **spin-glass**.

Some more motivations and connections can be found in the introduction of my PhD thesis (available here).

The one dimensional convex regime

Remark: in this talk $d = 1$, $\Omega = \mathbb{R}$ (or $\Omega = [0,1], \mathbb{R}^+$), and $p \geq 1$. Here convexity and strict monotonicity of $\mathcal{D}^p$ implies that, if realizations $(b_1, \dots, b_n)$, and $(r_1, \dots, r_n)$ are sorted in **natural order**, then

$$H^{(p)}_{\mathrm{opt},(n,1)}|_{p\geq 1} = \sum_{i=1}^{n} |b_i - r_i|^p,$$



This fact allowed to obtain detailed results about $E_{p,1}(n)|_{p\geq 1}$ (among other things) for **uniformly distributed points**, see Boniolo–Caracciolo–Sportiello 2014, Caracciolo–Sicuro 2014 and Caracciolo–**D'A**–Sicuro 2017, or Chapter 2 of my PhD thesis.

Bulk vs anomalous scaling

In the case of uniformly distributed points on $\Omega = [0,1]$ and $p \geq 1$ even an exact expression has been obtained using a connection with Selberg integrals (see Caracciolo–DiGioacchino–Malatesta–Molinari 2019) :

$$\begin{aligned} E^{\mathrm{U}}_{p,1}(n) &= n\frac{\Gamma\left(1+\frac{p}{2}\right)}{p+1}\frac{\Gamma(n+1)}{\Gamma\left(n+1+\frac{p}{2}\right)} \\ &= c_p n^{1-\frac{p}{2}}\left(1+o(1)\right) = c_p n^{\gamma_{p,1}}\left(1+o(1)\right) \quad (\textit{Bulk scaling}). \end{aligned}$$

What if ρ is non-uniform?

Depending on the choice of ρ, the large n behavior of $E^{(\rho)}_{p,1}$ may be considerably different from the bulk one due to contributions of few "elongated" edges in the region of low density of points which become more important than the "bulk" one. If $\gamma_{p,1} \neq 1 - \frac{p}{2}$ or in presence of a logarithmic correction we talk about an **anomalous scaling**.

Reduction to quadratures in the bulk case

If ρ is sufficiently regular the problem can be "reduced to quadratures". For a pdf ρ and R its cdf, call $\Psi^{(\rho)} := \rho \circ R^{-1}$, then

$$E_{p,1}^{(\rho)}(n) = \frac{2^p}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) \int_0^1 \left[\frac{\sqrt{s(1-s)}}{\Psi^{(\rho)}(s)}\right]^p \mathrm{d}\, s \; n^{1-p/2} + o\left(n^{1-p/2}\right).$$

Addressed in Caracciolo–**D'A**–Sicuro 2018 with non-rigorous regularization methods inspired by cutoff regularization in QFT. Analogous problem "in the continuum" considered by Bobkov-Ledoux 2019.

Example: for the exponential $\rho(x) = e^{-x}\theta(x)$, $\Psi^{\exp}(s) = 1 - s$.

- Cutoff method: $E_{p,1}^{\exp}(n) = \frac{2^p}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) \int_0^{1-c/n} \left(\frac{s}{1-s}\right)^{\frac{p}{2}} ds$
  at $p = 2$ we get $E_2(n) = 2 \ln n - 2 \log c - 2 + o(1)$;
- *Exact calculation* at $p = 2$ (via Beta integrals):
  $E_{2,1}^{\exp}(n) = 2 \sum_{k=1}^{n} \frac{1}{k} = 2 \ln n + 2\gamma_{\mathrm{E}} + o(1)$.

By comparison $c = e^{-\gamma_{\mathrm{E}} - 1} = 0.20655\ldots$

The general pattern

Emergence of logarithmic corrections to scaling suggested to study the contributions of edges at fixed n and then take the $n \to \infty$ limit.

General approach: 1) binomial expand $\mathcal{H}^{(p)}_{\mathrm{opt},(n,1)}|_{p\geq 1}$ at fixed (n,p) (for $p>1$ even[1]), 2) exploit linearity to compute expectation w.r.t. a certain Beta distrib., then 3) asymptotic analysis (case by case depending on ρ). More precisely

$$E^{(\rho)}_{p,1}(n) = \sum_{k=1}^{n} \mathbb{E}\left[(b_k - r_k)^p\right] = \sum_{k=1}^{n}\sum_{q=0}^{p} \binom{p}{q} (-1)^{p-q} M^{(\rho)}_{n,k,q} M^{(\rho)}_{n,k,p-q}$$

for

$$M^{(\rho)}_{n,k,\ell} = \langle R_\rho^{-1}(u)^\ell \rangle_{P_{n,k}}, \quad R_\rho^{-1}(u) \quad \text{the quantile function,}$$

where $\langle \dots \rangle_{P_{n,k}}$ is expectation w.r.t. $\mathrm{Beta}(k, n-k+1)$, i.e.

$$P_{n,k}(u)du := \frac{n!}{(k-1)!(n-k)!} u^{k-1}(1-u)^{n-k} du.$$

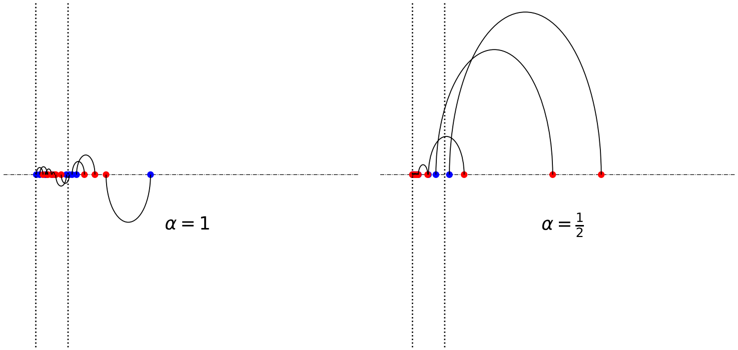
[1] The case $p=1$ is degenerate (there are many different solutions, see Caracciolo–Erba–Sportiello 2021) and the analysis is slightly different in this case.

The case of stretched exponential points

Let us consider the pdf (θ is Heaviside's function)

$$\rho_{\mathrm{ie},\alpha}(x) = \alpha x^{\alpha-1} \exp(-x^{\alpha})\theta(x).$$

Example solutions at $n = 10$:



Here $R_{\alpha}(x) = \exp(-x^{\alpha})\theta(x)$, hence $R_{\alpha}^{-1}(u) = (-\ln u)^{\frac{1}{\alpha}}$.

Sketch of the calculation

For $s = \frac{1}{\alpha}$ integer we wish to evaluate

$$M_{n,k;s} := \int_0^1 \mathrm{d}u \, P_{n,k}(u)(-\ln u)^s.$$

Lemma I (**D'A**-Sportiello 2020)

$$M_{n,k;s} = s! \, h_s(A_{k,n}),$$

for the alphabet $A_{k,n} := \left\{ \frac{1}{k}, \frac{1}{k+1}, \dots, \frac{1}{n} \right\}$, where h_s is the *complete homogeneous symmetric function* of degree s[2].

Two hints: 1) use the representation $-\ln u = \lim_{x\to 0} \frac{u^{-x}-1}{x}$ and 2) recall that for a polynomial $A(q) = a_p q^p + a_{p-1} q^{p-1} + \ldots + a_0$,

$$\sum_{q=0}^{p} \binom{p}{q} (-1)^{p-q} A(q) = p! \, a_p.$$

[2] I.e. for alphabet $(x_1, \dots, x_m)$, $h_s(x_1, \dots, x_m) := \sum_{1 \le j_1 \le \ldots \le j_s \le m} x_{j_1} x_{j_2} \cdots x_{j_s}$.

Sketch of the calculation

The contribution $E_{(s,p),n}(k)$ of the k-th edge in the solution to the total energy $E_{(s,p)}(n) = \sum_{k=1}^{n} E_{(s,p),n}(k)$ is thus

$$\begin{aligned} E_{(s,p),n}(k) &= \sum_{q=0}^{p} \binom{p}{q} (-1)^{p-q} M_{n,k;sq} M_{n,k;s(p-q)} \\ &= \sum_{q=0}^{p} (-1)^q \binom{p}{q} (sq)!(s(p-q))!\, h_{sq}(A_{k,n}) h_{s(p-q)}(A_{k,n}). \end{aligned}$$

In the next steps we make use of generating functions.

Exact result in the case $s = 1$ (exponential)

Here we have a remarkable simplification. Call

$$H(t, X) = \sum_{\ell \geq 0} h_\ell(X) t^\ell = \prod_{x \in X} (1 - tx)^{-1}.$$

Then

$$E_{(1,p),n}(k) = p! \sum_{q=0}^{p} (-1)^q h_q(A_{k,n}) h_{p-q}(A_{k,n}) = p! [t^p] H(t, A_{k,n}) H(-t, A_{k,n}).$$

Obviously for any alphabet X, $H(t, X)H(-t, X) = H(t^2, X^2)$, so that

$$E_{(1,p),n}(k) = p! [t^p] H(t^2, A_{k,n}^2) = p! h_{p/2} \left(A_{k,n}^2 \right).$$

Asymptotic result at $s > 1$ integer (stretched-exponential)

For the power-sum functions $p_k(X) = \sum_{x \in X} x^k$, define

$$P(t, X) = \sum_{\ell \geq 1} \ell^{-1} p_\ell(X) t^\ell = \log H(t, X)$$

Fact: $(p_\ell(A_{k,n})^{1/\ell})_\ell$ is monotone decreasing (due to the generalised-mean ineq.) In a regime in which $p_1 \gg \sqrt{p_2}$ (which we will justify later on) we consider the following perturbative expansion: put $P_+(t, X) = P(t, X) - tp_1(X)$, in terms of which

$$\begin{aligned} h_k(X) &= [t^k] H(t, X) = [t^k] \exp(P(t, X)) \\ &= \sum_{\ell=0}^{k} \frac{1}{(k-\ell)!} p_1(X)^{k-\ell} [t^\ell] \exp(P_+(t, X)). \end{aligned}$$

After some calculation we get

$$E_{(s,p),n}(k) \simeq p!\, s^p p_1(A_{k,n})^{(s-1)p} [t^p] \big(\exp(P_+(t, A_{k,n})) \exp(P_+(-t, A_{k,n})) \big).$$

Asymptotic result at $s > 1$ integer (stretched-exponential)

Since still obviously $\exp(P_+(t,X))\exp(P_+(-t,X)) = H(t^2,X^2)$, we get

$$E_{(s,p),n}(k) \simeq p!\, s^p p_1(A_{k,n})^{(s-1)p}\, h_{p/2}(A_{k,n}^2)$$

which recovers the exact result at $s = 1$. As $p_1(A_{k,n}) \sim \ln(n) - \ln(k) + \mathcal{O}(1)$ whenever $n \gg k$, while $p_h(A_{k,n}) \sim \frac{1}{h-1}(k^{1-h} - n^{1-h})$ if $n \gg k \gg 1$, in any case, whenever $n \gg k$, $p_1/\sqrt{p_2}$ is at least of order $\ln(n)$. We can thus replace $p_1(A_{k,n})$ with $\ln(n)$ to get

$$E_{(s,p),n}(k) = p!\, s^p \ln(n)^{(s-1)p}\, h_{p/2}(A_{k,n}^2)\big(1 + \mathcal{O}((1+\ln k)/\ln n)\big).$$

In order to estimate $E_{(s,p)}(n)$ we are left with the task of estimating ($q = \frac{p}{2}$)

$$F_{n,q} := \sum_{k=1}^{n} h_q(A_{k,n}^2).$$

Emergence of multiple ζ^* values in the $n \to \infty$ limit

Observe that each monomial contributing to $F_{n,q}$ is of the form $\frac{1}{i_1^2 i_2^2 \cdots i_q^2}$, and enters exactly i_1 times in the sum. Hence

$$F_{n,q} = \sum_{1 \le i_1 \le i_2 \le \cdots \le i_q \le n} \frac{1}{i_1\, i_2^2 \cdots i_q^2} = \sum_{k=1}^{n} \frac{1}{k} h_{q-1}(A_{k,n}^2). \tag{1}$$

In the series[3] associated to (1) we recognize a special kind of "**multiple ζ^* values**" (MZSVs), defined as

$$\zeta^*(a_1, \ldots, a_r) := \sum_{1 \le i_1 \le i_2 \le \cdots \le i_r} \frac{1}{i_1^{a_1} \cdots i_r^{a_r}}.$$

These numbers and the related multiple ζ values naturally arise in several contexts (e.g. in Vassiliev–Kontsevich theory, see Zagier 1992). Remarkably, they are not new in **physics**, as they naturally arise in the calculation of scattering amplitudes in perturbative quantum field theory (see Broadhurst–Kreimer 1997).

[3] In the limit $n \to \infty$ the sum is convergent since $h_{q-1}(A_{k,n}^2) \sim k^{-(q-1)}$ for $n \gg k$.

$E^{(s)}_{(p,1)}(n)$ for stretched-exponentially distributed points

For our MZSVs there is the remarkable simplification (Ohno–Wakabayashi 2006,Ohno–Zudilin 2008,Borwein–Zudilin 2011)

$$F_{+\infty,q} = \zeta^*(1, \underbrace{2,\dots,2}_{q-1}) = 2\zeta(2q-1)$$

where ζ is Riemann's function. In conclusion, we have the following

Lemma II (D'A-Sportiello 2020)

In the ERAP with stretched-exponentially distributed points with pdf $\rho_{\mathrm{ie},\alpha}(x) = \alpha x^{\frac{1}{s}-1}\exp(-x^{\frac{1}{s}})$, for $p > 1$ even and s integer

$$E^{(s)}_{p,1}(n) \simeq \begin{cases} 2s^2 \ln(n)^{2s-1} & p = 2 \\ 2\zeta(p-1)s^p p! \ln(n)^{(s-1)p} & p \geq 4 \end{cases}$$

Remark: the same asymptotics holds for $\rho^+_{\mathrm{ie},\alpha}(x) = \alpha x^{\alpha-1}\exp(1-x^\alpha)$ (the argument goes with minor modifications).

Some conclusions

Using basic combinatorial-analytics methods we have shown that for stretched–exponentially distributed points, $E(n)$ scales as a(n explicit) power of $\ln n$ and obtained explicit expressions for the asymptotic coefficients in terms of MZSVs. This result implies that for a generic distribution with both a uniform and a tail part the scaling is bulk if $p < 2$ and marginally anomalous if $p \geq 2$. We have also applied our approach to other classes of pdf, classified according to the local properties of ρ around the low density region (see Chapter 2 of my PhD thesis), exploiting at $p = 1$ the connection with recently introduced Dyck matchings at $p \leq 1$ (Caracciolo-**D'A**-Erba-Sportiello 2020).

Two open questions:

- Analytic continuation of results to $p \in \mathbb{R}, p \geq 1$;
- Extension to non-integer values of s (thus including gaussian tail at $\alpha = 2$, for which $E(n) \sim c \ln \ln n$, see Caracciolo–**D'A**–Sicuro 2018, Ledoux 2019,Berthet–Fort 2020).

Thank you for your attention!