Computational Experiments on Random Chromatic Persistent Homology

Sophie Rosenmeier SOPHIE.ROSENMEIER@UNIVIE.AC.AT
Max Planck Institute of Biochemistry, Munich, Germany; University of Vienna, Vienna, Austria

Ondřej Draganov
INRIA, Sophia Antipolis, France

Morteza Saghafian
Technical University Vienna, Vienna, Austria

Sebastiano Cultrera di Montesano
Broad Institute of MIT and Harvard, Boston, USA

Herbert Edelsbrunner
Institute of Science and Technology Austria, Klosterneuburg, Austria

Editors: Michael Bleher, Freya Jensen, Levin Maier, Diaaeldin Taha, and Anna Wienhard

Abstract

Chromatic alpha complexes serve as a generalization of alpha complexes for chromatic point sets and were developed beyond two colors by Cultrera di Montesano et al. (2025). Instead of only one as in the case in standard persistent homology, six different persistence diagrams result from this construction. Here we present the findings of Rosenmeier (2025), in which we study the expected number and total length of persistence pairs for each diagram, assuming uniformly distributed 2-colored points in the unit square. Additionally, we highlight deeper connections to the research area of Euclidean minimum spanning trees.

Background

With persistent homology we can extract topological features such as cavities or clusters in point clouds in Euclidean space. The idea is to grow balls around the points with the radius as a parameter. As the radius increases, the balls may overlap to merge components, form loops or higher-dimensional holes and their union results in a new topological space. Eventually, these holes are filled in again by the growing balls and it is exactly these lifespans of present topological features that are recorded by persistent homology (Edelsbrunner and Harer, 2010). Motivated by spatial biology (see e.g. (Cámara, 2017), (Rosenberger et al., 2025)), Cultrera di Montesano et al. (2025) developed a theory of persistent homology for chromatic point sets in Euclidean space, where each point can be thought of as a cell together with a color indicating its type. Based on Reani and Bobrowski (2021), the authors define a chromatic version of the Delaunay complex with a chromatic alpha filtration. By further choosing some colored subcomplex, which they want to relate to the rest, and using results from Cohen-Steiner et al. (2009) on the computation of image, kernel and cokernel persistence, a total of six persistence diagrams, the *six-pack*, arises out of this process.

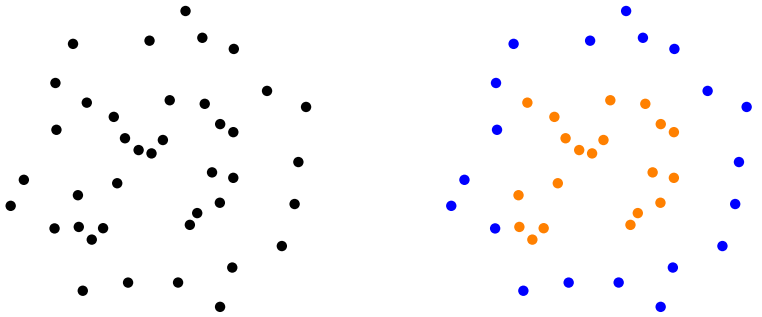

Figure 1: An example of a point set C in $\mathbb{R}^2$ (*left*) that turns into a chromatic point set (C, χ) by assigning a 2-coloring $\chi : C \to \{0, 1\}$ (*right*). The colors 0 and 1 are represented by blue and orange, respectively.

Let $C \subseteq \mathbb{R}^d$ be finite. Together with a coloring map $\chi : C \to \omega \subseteq \mathbb{N}$ we call the tuple (C, χ) a *chromatic point set* (see Figure 1). For a fixed *color* $j \in \omega$, the preimage $C_j = \chi^{-1}(\{j\})$ contains all j-colored points. We define the *chromatic Delaunay complex of* χ to be

$$\mathrm{Del}(\chi) = \{\alpha \subseteq C \mid \bigcap_{c \in \alpha} V(c, C_{\chi(c)}) \neq \emptyset\},$$

where $V(c, C_{\chi(c)})$ is the *Voronoi domain of* c *with respect to* $C_{\chi(c)}$ and consists of all points in $\mathbb{R}^d$ which are closest to c compared to the remaining points of the same color. Given a radius parameter $r \geq 0$ we can induce the *chromatic alpha filtration* $\{\mathrm{Alpha}(r, \chi)\}_{r \geq 0}$ of the simplicial complex $\mathrm{Del}(\chi)$ by setting the *chromatic Alpha complex of* χ *at* r

$$\mathrm{Alpha}(r, \chi) = \{\alpha \subseteq C \mid \bigcap_{c \in \alpha} \big(B_c(r) \cap V(c, C_{\chi(c)})\big) \neq \emptyset\},$$

where $B_c(r)$ denotes the closed ball of radius r around c. Note that for one color this gives us the usual Delaunay complex with alpha filtration of the growing clipped balls and in this case we write $\mathrm{Del}(C) = \mathrm{Del}(\chi)$ as well as $\mathrm{Alpha}(r, C) = \mathrm{Alpha}(r, \chi)$. Moreover, it is easily checked that $\mathrm{Alpha}(r, C_j)$ is a subcomplex of $\mathrm{Alpha}(r, \chi)$ for every $j \in \omega$ and every $r \geq 0$.

Let us now describe the process for obtaining the six-pack with the concrete example shown on the right in Figure 1. Taking the chromatic Delaunay complex $\mathrm{Del}(\chi)$ as well as the 0-colored subcomplex $\mathrm{Del}(C_0)$, we build two chromatic alpha filtrations and consider the radii $0 \leq r_1 \leq \cdots \leq r_n$ where either of the complexes changes topologically. These filtrations then yield two persistence modules by applying the p^{th} homology functor $\mathcal{H}_p$ ($p \in \mathbb{Z}$):

$$\begin{array}{ccccccc} \mathcal{H}_p(\mathrm{Alpha}(r_1, \chi)) & \longrightarrow & \ldots & \longrightarrow & \mathcal{H}_p(\mathrm{Alpha}(r_n, \chi)) & = & \mathcal{H}_p(\mathrm{Del}(\chi)) \\ {\scriptstyle (\iota_1)_\star}\uparrow & & & & {\scriptstyle (\iota_n)_\star}\uparrow & & \\ \mathcal{H}_p(\mathrm{Alpha}(r_1, C_0)) & \longrightarrow & \ldots & \longrightarrow & \mathcal{H}_p(\mathrm{Alpha}(r_n, C_0)) & = & \mathcal{H}_p(\mathrm{Del}(C_0)) \end{array} \tag{1}$$

$(\iota_k)_\star$ is the induced homomorphism of the inclusion map $\iota_k : \mathrm{Alpha}(r_k, C_0) \hookrightarrow \mathrm{Alpha}(r_k, \chi)$ for $k = 1, \ldots, n$. At each radius $r \geq 0$ one can show that $\mathrm{Alpha}(r, \chi)$ is homotopy equivalent

to $\mathrm{Alpha}(r, C)$. This means that the upper persistence module in (1) results in the exact same persistence diagram we would expect in the non-chromatic case. In the following we refer to it as the *codomain diagram*. The lower persistence module in (1) gives us the same persistence diagram as when viewing only the blue points C_0 to begin with and we call it the *domain diagram*. Both diagrams are part of the six-pack as Figure 2 displays for the chromatic point set in Figure 1.

Considering the simplicial pairs $\big(\mathrm{Alpha}(r_k, \chi), \mathrm{Alpha}(r_k, C_0)\big)$, we construct a persistence module out of their relative homology groups, which yields the *relative diagram*. It captures topological features of $\mathrm{Del}(\chi)$ with the subcomplex $\mathrm{Del}(C_0)$ identified as a single point. Moreover, all induced homomorphisms $(\iota_k)_\star$ are linear maps, so we further obtain persistence modules of kernels, images and cokernels and, therefore, three more persistence diagrams: the *kernel*, *image* and *cokernel diagram*.

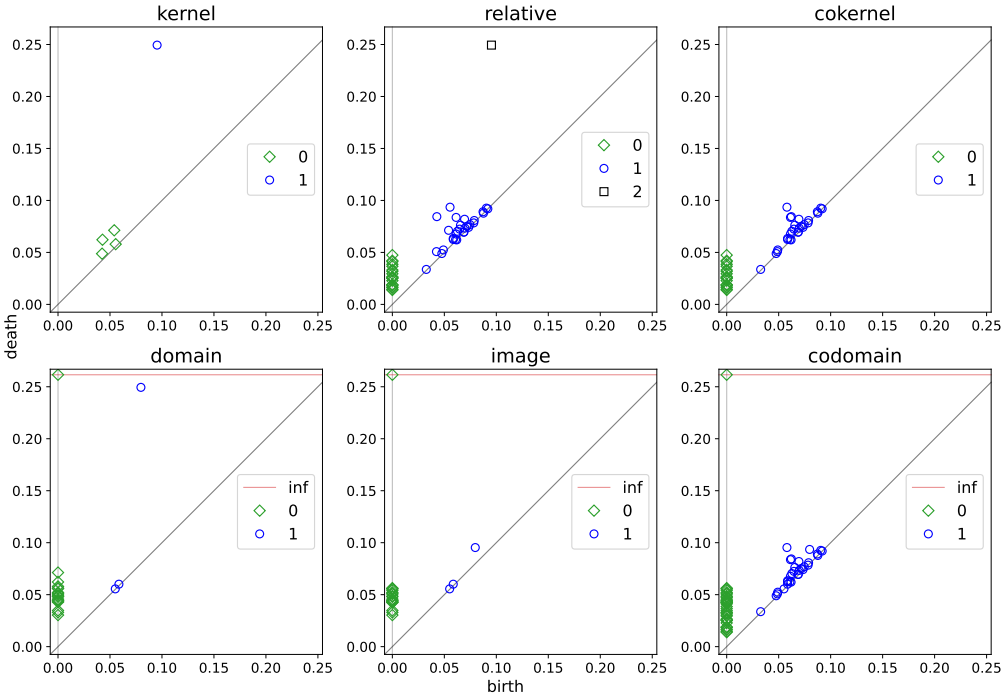

Figure 2: The six-pack of persistence diagrams of the chromatic point cloud with two colors from Figure 1 and with the inclusion $\mathrm{Del}(C_0) \subseteq \mathrm{Del}(\chi)$.

The image diagram shows all persistence pairs obtained from the chosen subcomplex embedded into the complex. In our case this translates to all occurrences of blue holes within the whole point cloud. In case a hole has died in the image but not yet in the domain, it appears in the kernel diagram where it persists until it finally dies in the domain as well. In Figure 2 this is best seen for the circular structure that is formed by the blue points in Figure 1 and that lives the longest out of all 1-dimensional holes in the domain diagram. Finally, the cokernel diagram gathers all persistence pairs from the codomain that are essentially not present in the image. Notice that the six-pack depends on the choice of subcomplex. In our example, taking for instance $\mathrm{Del}(C_1)$ instead of $\mathrm{Del}(C_0)$ leads to different diagrams.

Computational Setup

We investigated the six-pack for random point sets, the initial motivation being that real data often comes with noise and we are interested in whether or how the six-pack machinery is affected by it and how the "non-signal" looks like. More precisely, in our setup we consider uniformly distributed 2-colored point sets $\chi^{(n,t)} : C^{(n,t)} \to \{0, 1\}$ in the unit square $[0, 1]^2$ with varying number of points n and varying probability of the two colors $(t, 1-t)$ with $t \in [0, 1]$ (see Appendix A for more details). In particular, we focus on the number (0-norm) and total length (1-norm) of persistence pairs in each of the six diagrams in expectation. For these computations we do 100 random colorings for each particular point cloud $(C^{(n,t)}, \chi^{(n,t)})$ as well as for each of the two interesting choices of subcomplexes – $\mathrm{Del}(C_0^{(n,t)})$ and $\mathrm{Del}(C_0^{(n,t)}) \sqcup \mathrm{Del}(C_1^{(n,t)})$.[1]

Results

For each specific configuration with varying n we calculate the means of the number and total length of persistence pairs. These outcomes are then plotted against n, respectively, and we fit curves through them. Repeating this with different values of t yields results like the ones shown as an example in Figure 3 in Appendix A for the domain 0^{th} persistence diagram. An analysis of these curves reveals that their slopes depend on the given probability. We, therefore, collect all the slopes of specific configurations and do another fit against t. For the domain 0^{th} persistence diagram and with the inclusion $\mathrm{Del}(C_0^{(n,t)}) \sqcup \mathrm{Del}(C_1^{(n,t)}) \subseteq \mathrm{Del}(\chi^{(n,t)})$ this is depicted in Figure 4 in Appendix A. For the other choice of subcomplex and the remaining diagrams similar steps are done. We arrive at two tables of results, one for 0^{th} and one for 1^{st} degree persistent homology. In both tables, each row represents one of the six diagrams and is further divided into two subrows — one for the expected 0-norm and one for the expected 1-norm. The two columns represent the two kinds of filtrations we obtain depending on the choice of subcomplex.

Currently known results on expected 0- and 1-norms are given in blue and detailed proofs thereof can be found in (Rosenmeier, 2025). In some instances we do not have a proof for the corresponding expectation, but we state conjectures of possible explicit formulae in yellow based on the experimental observations. White boxes with a question mark indicate the cases where we neither have a theoretical idea nor a concrete computational hint to what the explicit form could look like.

Observe that nearly all the different cases are worked out for $\mathcal{H}_0$ in Table 1. In the first column no formulae are given for the relative and cokernel diagram, because both 0- and 1-norms are zero for any given point cloud $(C^{(n,t)}, \chi^{(n,t)})$. Also, due to the symmetry in t for $\mathrm{Del}(C_0^{(n,t)}) \sqcup \mathrm{Del}(C_1^{(n,t)})$, we obtain rather simple functions in this setting. This changes for $\mathrm{Del}(C_0^{(n,t)})$, where the conjectured explicit formulae seem to be more delicate in the 1-norm setting. Fortunately, however, once we are able to show either one of these conjectures in the left table, we can deduce the rest of them with the help of three powerful relations between the lengths of six-pack diagrams (see Theorem 5.3, Cultrera di Montesano et al.

1. Because all points are uniformly distributed with probability t of being blue and $1-t$ of being orange, the filtrations with subcomplexes $\mathrm{Del}(C_0^{(n,t)})$ and $\mathrm{Del}(C_1^{(n,t)})$ are symmetric to each other.

$\mathcal{H}_0$		$\mathrm{Del}(C_0^{(n,t)}) \sqcup \mathrm{Del}(C_1^{(n,t)})$	$\mathrm{Del}(C_0^{(n,t)})$
Ker	0	$2t(1-t)n$	$t(1-t)n$
	1	$\frac{c}{2}\big(\sqrt{t}+\sqrt{1-t}-1\big)\sqrt{n}$	$\big((\frac{c}{2}-\frac{1}{4})t^2-\frac{c}{2}t+\frac{1}{4}\sqrt{t}\big)\sqrt{n}$
Rel	0	—	$(1-t)n$
	1		$\big((\frac{c}{2}-\frac{1}{4})t^2-\frac{c}{2}t-(\frac{c}{2}-\frac{1}{4})\sqrt{t}+\frac{c}{2}\big)\sqrt{n}$
Cok	0		$(1-t)n$
	1		$\big((\frac{c}{2}-\frac{1}{4})t^2-\frac{c}{2}t-(\frac{c}{2}-\frac{1}{4})\sqrt{t}+\frac{c}{2}\big)\sqrt{n}$
Dom	0	n	tn
	1	$\frac{c}{2}\big(\sqrt{t}+\sqrt{1-t}\big)\sqrt{n}$	$\frac{c}{2}\sqrt{t}\sqrt{n}$
Img	0	n	tn
	1	$\frac{c}{2}\sqrt{n}$	$\big(-(\frac{c}{2}-\frac{1}{4})t^2+\frac{c}{2}t+(\frac{c}{2}-\frac{1}{4})\sqrt{t}\big)\sqrt{n}$
Cod	0	n	
	1	$\frac{c}{2}\sqrt{n}$	

$\mathcal{H}_1$		$\mathrm{Del}(C_0^{(n,t)}) \sqcup \mathrm{Del}(C_1^{(n,t)})$	$\mathrm{Del}(C_0^{(n,t)})$
Ker	0	$3t(1-t)n$	$t(1-t)(1+t)n$
	1	?	?
Rel	0	$4t(1-t)n$	$(2t+1)(1-t)n$
	1	?	?
Cok	0	$\mu t(1-t)n$, $\mu \approx 2.8$	?
	1	?	?
Dom	0	n	tn
	1	$(\frac{c}{2}-\frac{1}{4})\big(\sqrt{t}+\sqrt{1-t}\big)\sqrt{n}$	$(\frac{c}{2}-\frac{1}{4})\sqrt{t}\sqrt{n}$
Img	0	$1-\kappa t(1-t)n$, $\kappa \approx 2.4$	?
	1	?	?
Cod	0	n	
	1	$(\frac{c}{2}-\frac{1}{4})\sqrt{n}$	

Table 1: Explicit proven formulae (*blue*), conjectures (*yellow*) or completely unknown cases (*white with question mark*) for expectations in the 0^{th} (*left*) and 1^{st} (*right*) degree persistent homology and for the number of points n and probability parameter t. Each row for the diagrams is subdivided into expected 0-norm (*upper part*) and 1-norm (*lower part*). The asymptotic constant c arises due to a connection between the 1-norm of the 0^{th} codomain diagram and expected lengths of Euclidean minimum spanning trees.

(2025)). In the right table roughly half of the boxes are still given with a question mark, most of which are regarding the 1-norm.

Notice the appearance of a constant $c > 0$ in all known entries about the expected length of persistence pairs. This is the constant such that the expected length of the Euclidean minimum spanning tree of n points sampled uniformly at random in the unit square is $c\sqrt{n}$ as n goes to infinity (Steele, 1988). It is no coincidence that this asymptotic constant shows up here, as the 1-norm of the 0^{th} codomain diagram is half the length of the Euclidean minimum spanning tree. No explicit value of c is known, although in the past there was some partial success in developing a formula for computing this constant by Avram and Bertsimas (1992). Unfortunately, this formula contains a series expansion whose summands are increasingly harder to calculate analytically. Nevertheless, Avram and Bertsimas achieved a lower bound of 0.6008 by approximating the series up to the fifth summand. Very recently, Draganov et al. (2025) were able to improve this number by means of birth- and death-giving critical simplices and we have

$$0.6289 \leq c \leq \frac{\sqrt{2}}{2} \approx 0.7072,$$

where the best upper bound is due to a theoretical argument of Gilbert (1965). Our experiments are consistent with these bounds and they suggest c to lie between 0.6364 and 0.6524.

References

Florin Avram and Dimitris Bertsimas. The minimum spanning tree constant in geometrical probability and under the independent model: A unified approach. *The Annals of Applied Probability*, 2(4):113–130, 1992.

David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov. Persistent homology for kernels, images, and cokernels. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1011–1020. Society for Industrial and Applied Mathematics, 2009.

Sebastiano Cultrera di Montesano, Ondřej Draganov, Herbert Edelsbrunner, and Morteza Saghafian. Chromatic alpha complexes. *Foundations of Data Science*, 2025.

Pablo G. Cámara. Topological methods for genomics: Present and future directions. *Current Opinion in Systems Biology*, 1:95–101, 2017.

Ondřej Draganov, Herbert Edelsbrunner, Sophie Rosenmeier, and Morteza Saghafian. Expected length of the euclidean minimum spanning tree and 1-norms of chromatic persistence diagrams in the plane. *preprint arXiv:2510.23373v1*, 2025.

Herbert Edelsbrunner and John Harer. *Computational Topology - An Introduction*. American Mathematical Society, 2010. ISBN 978-0-8218-4925-5.

E. N. Gilbert. Random minimal trees. *Journal of the Society for Industrial and Applied Mathematics*, 13(2):376–387, 1965.

Yohai Reani and Omer Bobrowski. A coupled alpha complex. *Journal of Computional Geometry*, 14:221–256, 2021.

Florian A. Rosenberger, Sophia C. Mädler, and Katrine et al Holtz Thorhauge. Deep visual proteomics maps proteotoxicity in a genetic liver disease. *Nature*, 642:484–491, 2025.

Sophie Rosenmeier. Computational experiments on random chromatic persistent homology. Master's thesis, University of Vienna, 2025.

J. Michael Steele. Growth rates of euclidean minimal spanning trees with power weighted edges. *The Annals of Probability*, 16(4):1767–1787, 1988.

Appendix A. Computational Details

Number of experiments: 100
Number of points: parameter n
100, 300, 500, 700, 1000, 1200, 1500, 1700, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 12000, 15000, 17000 20000
Colors: $\{0, 1\}$
Probability of colors: parameter $(t, 1-t)$
$(0.01, 0.99)$, $(0.05, 0.95)$, $(0.07, 0.93)$, $(0.1, 0.9)$, $(0.15, 0.85)$, $(0.2, 0.8)$, $(0.25, 0.75)$, $(0.3, 0.7)$, $(0.35, 0.65)$, $(0.4, 0.6)$, $(0.45, 0.55)$, $(0.5, 0.5)$, $(0.55, 0.45)$, $(0.6, 0.4)$, $(0.65, 0.35)$, $(0.7, 0.3)$, $(0.75, 0.25)$, $(0.8, 0.2)$, $(0.85, 0.15)$, $(0.9, 0.1)$, $(0.93, 0.07)$, $(0.95, 0.05)$, $(0.99, 0.01)$
Filtrations:
Complex $\mathrm{Del}(\chi^{(n,t)})$ with subcomplex $\mathrm{Del}(C_0^{(n,t)})$ or $\mathrm{Del}(C_0^{(n,t)}) \sqcup \mathrm{Del}(C_1^{(n,t)})$
Spatial dimension: 2
Diagrams: kernel, relative, cokernel, domain, image, codomain
Distribution of points: uniform
Location of points: unit square
Degree of persistent homology: 0^{th}, 1^{st}

We observe a linear behavior with respect to n when taking the 0-norm and a square-root behavior with respect to n when taking the 1-norm (see Figure 3). For a fixed choice of subcomplex, n points and probability t that a point is 0-colored we, thus, fit curves through the means of the ...

... 0-norm of the form

$$n \mapsto c_1 \cdot n + c_0,$$

... 1-norm of the form

$$n \mapsto c_1 \cdot \sqrt{n} + c_0,$$

where c_1 and c_0 might depend on t. For large enough n the offset c_0 is, however, negligible in both cases. Figure 4 shows the dependence of the fitting coefficient c_1 on t for the expected number of connected components in the domain diagram and with chosen subcomplex $\mathrm{Del}(C_0^{(n,t)}) \sqcup \mathrm{Del}(C_1^{(n,t)})$. All other configurations with different diagrams, norms, degree of persistent homology or filtration were analyzed in a similar fashion.

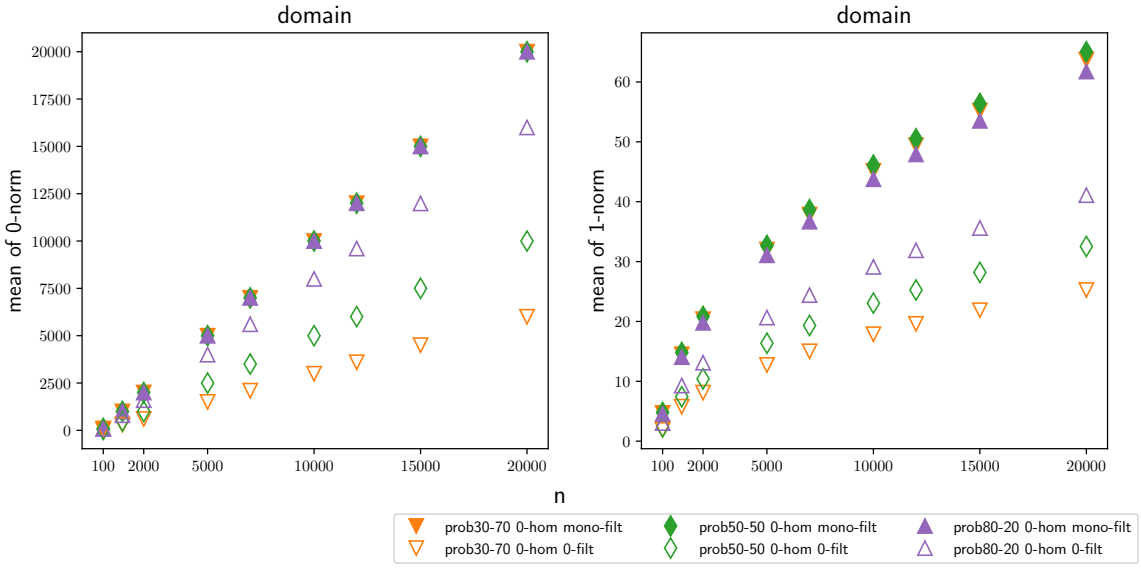

Figure 3: Different behavior for means of the 0-norm (*left*) and 1-norm (*right*) in 0^{th} persistent homology for the domain diagram and both subcomplexes $\mathrm{Del}(C_0^{(n,t)})$ ("0-filt", *empty markers*) and $\mathrm{Del}(C_0^{(n,t)}) \sqcup \mathrm{Del}(C_1^{(n,t)})$ ("mono-filt", *solid markers*). For better visualization, only six values of n and three values of t are shown.

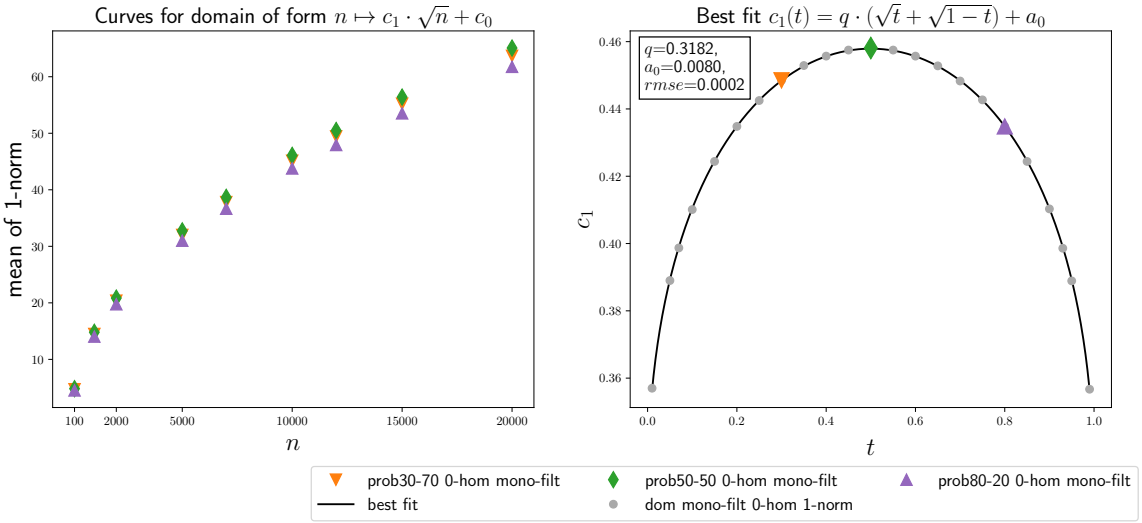

Figure 4: Different behavior for means of the 0-norm in 0^{th} persistent homology for the domain diagram with $\mathrm{Del}(C_0^{(n,t)}) \sqcup \mathrm{Del}(C_1^{(n,t)})$ (*left*) and corresponding best fit for the coefficient c_1 of different probabilities t (*right*). Like in Figure 3, only six values of n are shown for clarity and three values of t are highlighted in color.