# STATISTICAL ESTIMATION OF THE AVERAGE NUMBER OF CLUSTER SITES ON UNIFORMLY WEIGHTED SQUARE LATTICES

**P. V. Moskalev[1,2], L. S. Onishchenko[2]**

[1] Moscow State Technological University "STANKIN", Moscow, Russia,
127055, Vadkovsky lane, 1
[2] Voronezh State Technical University, Voronezh, Russia,
394006, 20th Anniversary of October st., 84
E-mail: moskaleff@mail.ru

A significant part of statistical estimates of the characteristics of clusters on finite-size percolation lattices are related to the probability of formation of these clusters, which makes such estimates very sensitive not only to sample sizes, but also to lattice sizes. Approximations of such empirical relationships are often based on various sigmoid functions. For example, to approximate the relative frequency of formation of contracting (percolation) clusters from a sample $\{w_i(p_i)\}$ of $n$ points on finite-size lattices, logistic models can be used [1, 2]:

$$w_i = 1/(1 + \exp(-(p_i - p_c)/s)) + e_i, \quad \text{for} \quad i = 1, 2, \ldots, n, \tag{1}$$

where $p_c$, $s$ are shift and scale parameters of the model (1), the first of which is usually subject to estimation, and the latter is determined by the percolation lattice size. The logistic function $w(p)$ is bounded and strictly monotonic, describing the asymptotic transition from the lower limit value to the upper one as the argument increases:

$$w(p \to -\infty) \to 0+ \quad \text{and} \quad w(p \to +\infty) \to 1-. \tag{2}$$

Note that when statistically estimating the parameters of functions of the form (1), only those points whose ordinates differ significantly from their limiting values (2) have information significance. In the absence of a priori estimates of the percolation threshold, this can increase the computational complexity of the problem and leads to a significant dependence of the errors in statistical estimates on the available computing resources.

One of the possible ways to reduce the computational complexity of such a problem is to estimate the required characteristics from samples of quantities based on statistics not of the clusters as a whole, but of the percolation lattice sites included in the clusters. We will form estimation data based on the number of cluster sites with a sample size of $N = 300$ implementations on uniformly weighted square lattices with linear sizes $x = 33, 65, 129$ sites for a given grid of open site probabilities $p = 0, 0.05, ..., 1$. The correlation fields of these samples are shown in Figure 1. The "triangle", "circle" and "square" symbols correspond to samples built on percolation lattices of sizes $x = 33, 65$, and 129 sites.

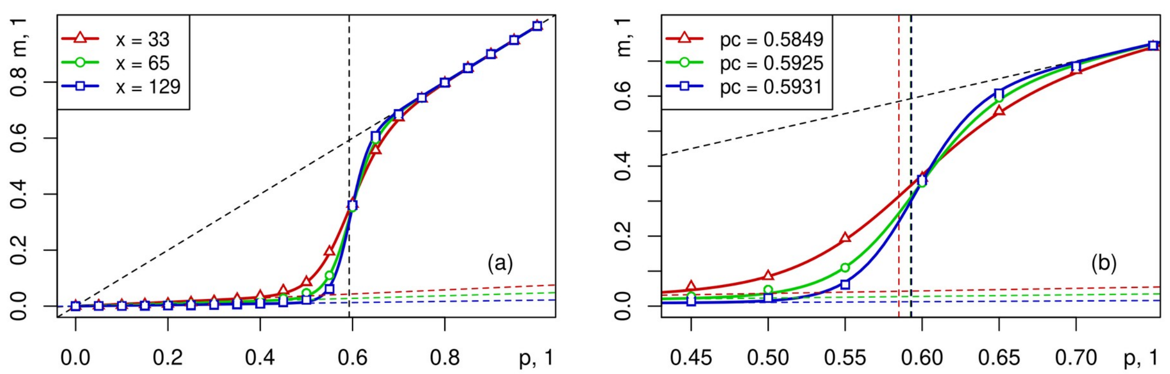


**Fig. 1.** Approximation (3) for the number of cluster sites $m$ on uniformly weighted square lattices at $x = 33, 65, 129$ sites vs. open site probabilities $p = 0, 0.05, ..., 1$

It can be shown that all presented samples $\{m_i(p_i)\}$ can be approximated by monotonic sigmoid functions describing the transition between the lower and upper slant asymptotes passing through the origin. This allows us to propose a generalized logistic model:

$$m_i = kp_i + (1 - k)p_i/(1 + \exp(-(p_i - p_c)/s)) + e_i, \quad \text{for} \quad i = 1, 2, \ldots, n, \tag{3}$$

where $k$ is a parameter that determines the angular coefficient of the lower slant asymptote; $p_c$, $s$ are the shift and scale parameters of the model (3) that determine the position of the inflection point and the radius of the neighborhood within which the second derivative of the approximating function (3) differs significantly from zero.

The slant dashed lines in Figure 1 correspond to: a) lower slant asymptotes of the form $m = kp$, approximating the behavior of model (3) at subcritical values of the probability of open sites $p < p_c$; b) an upper slant asymptote of the form $m = p$, approximating the behavior of model (3) at supercritical values of the probability of open sites $p > p_c$ and coincides on the interval (0, 1) with the cumulative distribution function of a uniformly distributed random variable $s \sim \text{unif}(0, 1)$ weighing sites on the percolation lattice.

The vertical dashed line in Figure 1(a) corresponds to the value of the site percolation threshold on a square lattice with a unit von Neumann neighborhood $p_c = 0.592746...$, known from the literature [2]. The vertical dashed lines in Figure 1(b) correspond to statistical estimates of the shift parameter of models (3), found for samples of site clusters on lattices with different linear sizes: a) $\mathbf{I}_{0.95}(p_{c1}) = (0.5849 \pm 0.0026)$ for $x_1 = 33$ sites; b) $\mathbf{I}_{0.95}(p_{c2}) = (0.5925 \pm 0.0017)$ for $x_2 = 65$ sites; c) $\mathbf{I}_{0.95}(p_{c3}) = (0.5931 \pm 0.0014)$ for $x_3 = 129$ sites.

All results were obtained using the "*ssi20()*" function from the "*SPSL*" package released by the author for the R system under the GNU GPL-3 license [3]. To estimate the parameters of the regression model, the "*gsl_nls()*" function from the "*gslnls*" package, released under the GNU GPL-3 license for the R system, was used [4]. A summary of the results obtained when constructing approximation (3) for the number of cluster sites on the probability of open sites for square lattices with size $x = 129$ sites is given in this listing:

```
> print(summary(f3))

Formula: w3 ~ k * pp + (1 - k) * pp/(1 + exp(-(pp - pc)/s))

Parameters:
    Estimate Std. Error t value Pr(>|t|)
k  0.0213085  0.0047065   4.527 0.000261 ***
pc 0.5930757  0.0006710 883.896  < 2e-16 ***
s  0.0201649  0.0006513  30.963  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.004525 on 18 degrees of freedom

Number of iterations to convergence: 7
Achieved convergence tolerance: 5.236e-12

> print(confint(f3, parm="pc"))
       2.5 %    97.5 %
pc 0.591666 0.5944853
```

When using only statistical indicators, the quality of the constructed approximation can be characterized as good. Note, however, that the radius of the 0.95-confidence intervals decreases asymptotically with increasing the lattice size $x$. For example, the absolute error in estimating the shift parameter $p_c$ of model (3) will not exceed $\varepsilon \leq 0.001$, with a lattice size of at least $x \geq 250$ sites, and the level $\varepsilon \leq 0.0005$ will be achieved with a lattice size of at least $x \geq 1000$ sites. This behavior suggests that the quality of model (3) can be improved by choosing functions that better approximate the latent features of the data obtained through computational experiments.

The research was funded by the Russian Science Foundation (project No. 23-21-00376).

**References:**

[1] Moskalev P. V. Percolation Modeling of Porous Structures. – Moscow: URSS, 2018 [in Russian].
[2] Newman M. E. J., Ziff R. M. // Phys. Rev. Lett., V. 85, 2000, P. 4104, DOI: 10.1103/PhysRevLett.85.4104.
[3] Moskalev P. V., SPSL: Site Percolation on Square Lattices. – CRAN, 2019, URL: https://cran.r-project.org/package=SPSL.
[4] Chau J. gslnls: GSL Nonlinear Least-Squares Fitting. – CRAN, 2023. – URL: https://cran.r-project.org/package=gslnls.