On the convergence of percolation probability functions to cumulative distribution functions on square lattices

Pavel V. Moskalev *
Voronezh State Agricultural University,
1 Michurin street, Voronezh, 394087, Russia
2018 May

Abstract. We consider a percolation model on square lattices whose sites are weighted by beta-distributed random variables $S \sim \text{Beta}(a, b)$ with a positive real parameters $a > 0$ and $b > 0$. Using the Monte Carlo method, we estimate the percolation probability $P_\infty$ as a relative frequency $P_\infty^*$ averaged over the target subset of sites on a square lattice. As a result of the comparative analysis, we formulate two empirical hypotheses: H1 on the correspondence of percolation thresholds $p_c$ to $p_0$-quantiles of random variables $S_i$ weighing sites of the square lattice, and H2 on the convergence of statistical estimates of percolation probability functions $P_\infty^*(p)$ to cumulative distribution functions $F_{S_i}(p)$ of these variables $S_i$ for the supercritical values of the occupation probability $p \geq p_c$.

PACS. 64.60.ah — Percolation in phase transitions; 02.50.Ng — Monte Carlo methods in probability theory and statistics.

1 Introduction

The percolation theory, born from the conjugation of graph theory and probability, attracts researchers from many fields of science [1, 2, 3]. Analyzing the structure of percolation models used in physics, we can divide them into lattice, continuous and potential models [4, 5]. The space of lattice percolation models form subsets of vertices (sites) and edges (bonds) of a graph with a given topology and random weights. If the random weights correspond only to the sites (or bonds), then we get a site (or bond) percolation model. On the other hand, if the random weights correspond to both sites and bonds, then we get a mixed percolation model [6, 7].

*e-mail: moskalefff@gmail.com
The object of research in this work is a site percolation model on a square lattice with a unit von Neumann neighborhood [8]. This neighborhood consists of four sites, only one coordinate of which differs from the current site’s coordinate by one $V(x, y) = \{(x \pm 1, y), (x, y \pm 1)\}$. One of the problems solved for this model is the search for a site cluster or an open site subset, connected to the starting site subset. The cluster sites satisfy condition $s_{xy} < p$, which weights the sample values $s_{xy}$ of the random variable $S$ with the occupation probability $p$. Lattice sites that satisfy the weighted inequality are called open, otherwise — closed sites. If both the start and target subsets of the lattice sites are included in the cluster simultaneously, we will call it the finite-dimensional approximation of the percolation cluster.

In the classical percolation model, sites on a square lattice are weighted by a standard uniformly distributed random variable $S \sim \text{Unif}(0, 1)$. However, a uniform distribution with linear cumulative functions is quite rare in natural phenomena. In such phenomena, the nonlinear form of cumulative functions is much more common [9,10]. In the proposed model, sites on a square lattice are weighted by beta-distributed random variables $S \sim \text{Beta}(a, b)$ with real shape parameters $a > 0$ and $b > 0$. Note that $S \sim \text{Unif}(0, 1) \equiv \text{Beta}(1, 1)$ and this model can be considered as an extension of the classical percolation model for the class of continuous distributions defined on the interval $[0, 1]$.

2 Algorithms and software

The main algorithm used to generate realizations of site clusters is based on iterative joining of open sites from the current neighborhood of the cluster. Iterations begin with a starting subset of sites and continue until the disappearance of open sites in the current neighborhood of the cluster.

To implement this algorithm, we used a free software environment for statistical computing and programming $R$ [11]. Listing 1 shows the source code of the "ssi20b()" function that generates the site clusters on a square lattice with a von Neumann neighborhood and impermeable boundary conditions.

```
ssi20b <- function(x=33, p=0.592746,
                   set=(x^2+1)/2, all=TRUE,
                   a=1, b=1) {
  e <- c(-1, 1, -x, x)
  acc <- array(rbeta(x^2,a,b), rep(x,2))
  if (all) acc[set] <- 2
  else acc[set <- set[acc[set] < p]] <- 2
  acc[c(1,x),] <- acc[,c(1,x)] <- 1
  while (length(set) > 0) {
    acc[set <- unique(c(
      set[acc[set+e[1]] < p] + e[1],
      set[acc[set+e[2]] < p] + e[2],
      set[acc[set+e[3]] < p] + e[3],
      set[acc[set+e[4]] < p] + e[4]))] <- 2
  }
}
```
This function uses the following variables to initialize: “x” — linear dimension of a square percolation lattice; “p” — occupational probability for one site; “set” — starting site subset; “all” — if “all=TRUE”, then the function uses all sites from the starting subset, or else only open sites; “a”, “b” — shape parameters of the beta-distributed random variable $S$, which weighs the percolation lattice sites. This function creates the following variables at runtime: “e” — shift vector defined by the neighborhood of the lattice site; “acc” — array with sample values of the random variable $S$, which weighs the percolation lattice sites.

At the beginning of this listing, we specify the vector “e”, which determines the shifts of site indexes from the unit von Neumann neighborhood of the internal lattice site. Next, we initialize the elements of the square matrix “acc” using a sample of the beta-distributed random variable $S$. Then, all or only the open elements of the starting subset of the sites “set” are marked with a numeric label exceeding the largest value of the random variable $S$ that weighs the lattice sites. To improve the performance of the function “ssi20b()”, all boundary lattice sites are marked as closed. Next, we define a loop with a precondition for continuing the iteration by the presence of sites in the “set” vector. Before the iterations begin, the “set” vector includes the indices of the starting site subset. On subsequent iterations, this vector includes indexes of open sites from the unit von Neumann perimeter for the “set” vector sites at the previous iteration.

To estimate statistically stable characteristics of site clusters on a square lattice, we need data on the relative frequencies with which the lattice sites will be included in the sample of percolation clusters. Listing 2 shows the source code of the “fssi20b()” function that calculates the site relative frequencies on a square percolation lattice with the unit von Neumann neighborhood and impermeable boundary conditions.

```r
fssi20b <- function(n=1000, x=33, p=0.592746,
    set=(x^2+1)/2, all=TRUE,
    a=1, b=1) {
  rfq <- array(0, rep(x,2))
  for (i in seq(n)) {
    rfq <- rfq + (ssi20b(x,p,set,all,a,b) > 1) } 
  return(rfq/n) }
```

In addition to the variables used to initialize the function “ssi20b()”, this function also requires the variable “n”, which determines the sample size. At runtime, the function “fssi20b()” creates “rfq” — array of relative frequencies with which the lattice sites are included in the cluster sample.

The function “ssi20b()” shown above are based on the similar function “ssi20()”, previously published in the SPSL package [12]. This package containing several functions for generating percolation clusters and their samples on
2D and 3D square lattices with different sizes, neighborhood types, percolation probabilities and starting site subsets, but with a standard uniform weighting distribution \( S \sim \text{Unif}(0, 1) \). The basic characteristics of percolation models realized by functions from the \textit{SPSL} package have been described by the author in previously published works \cite{13, 14} in Russian. The scope of these percolation models is limited by such stochastic phenomena that are satisfactorily described by uniformly distributed random variables with linear cumulative distribution functions. However, as we said above, the nonlinear form of cumulative distribution functions is much more common \cite{15, 16}. The function “\texttt{ssi20b()}” were developed for percolation modeling of just such phenomena.

3 Cluster generation

To generate individual clusters, we use a square lattice with a linear size of 65 sites and three weighting distributions whose functions \( F_S(p) \) we show in Fig. 1. The linear cumulative function for \( S_2 \sim \text{Beta}(1, 1) \) is shown as a green line, and the nonlinear cumulative functions for \( S_1 \sim \text{Beta}(1, 2) \) and \( S_3 \sim \text{Beta}(2, 1) \) is shown as red and blue lines.

![Figure 1: Cumulative functions \( F_S(p) \) of beta-distributed weights \( S_1 \sim \text{Beta}(1, 2) \) (red line), \( S_2 \sim \text{Beta}(1, 1) \) (green line), \( S_3 \sim \text{Beta}(2, 1) \) (blue line).](image)

In Fig. 2 we show individual realizations of site clusters for various parameters of the percolation lattice. As the starting subset, we selected open sites along the lower boundary of the square lattice at \( y = 1 \). The occupation probability in Fig. 2 changed from subcritical \( p < p_c \) (left column) to supercritical \( p > p_c \) (right column). The convexity of the functions \( F_S(p) \) for the weight distribution in Fig. 2 changed from the negative at \( S_1 \sim \text{Beta}(1, 2) \) (top row), through the zero at \( S_2 \sim \text{Beta}(1, 1) \) (center row), to the positive at \( S_3 \sim \text{Beta}(2, 1) \) (bottom row).
Figure 2: Site clusters for weight distributions $S_1 \sim \text{Beta}(1, 2)$ (top row), $S_2 \sim \text{Beta}(1, 1)$ (center row), $S_3 \sim \text{Beta}(2, 1)$ (bottom row), and occupation probabilities $p_1 < p_c$ (left column), $p_2 > p_c$ (right column).
The red color in Fig. 2 corresponds to lattice sites with close to zero weights, orange to sites with weights close to one, and white to sites marked on the cluster. In this case, the overall color of the lattice sites is redder, the frequency of its open sites is higher. A comparative analysis of the cluster realizations in Fig. 2 and the cumulative distribution functions \( F_S(p) \) in Fig. 1 shows that the large convexity of the cumulative function leads to a shift of the percolation threshold \( p_c \) to unity.

### 4 Relative frequencies of lattice sites

As is known, the percolation threshold \( p_c \), like any other parameters calculated from individual implementations of site clusters, is the values of some random variable [17]. To find a statistical estimate of the mathematical expectation of such a value, we use a sample of these values. We will need to perform a statistical estimation procedure for each implementation of the site cluster and the number of these calculations will increase linearly as the sample size grows. Then, if we assume the distribution of the estimated parameter is close to normal, then to reduce the statistical error in \( k \) times, we need to increase the sample size and the number of additional computations in \( k^2 \) times. We can overcome this problem by moving from statistical estimation of parameters for individual implementations of site clusters to estimating the parameters of their sample as a whole. Input data for such an estimate will be the frequencies with which each of the sites on the square lattice is used in the cluster sample.

In Fig. 3 we show the relative frequencies for sites at different parameters of the percolation lattice. The starting subset, occupational probabilities \( p \) and weighting variables distributions \( S \) for the percolation lattices in Fig. 3 are identical to those shown in Fig. 2.

The red color in Fig. 3 corresponds to the lattice sites with relative frequencies close to zero, and the yellow color — with frequencies close to one. The redder the average color of the lattice sites, the lower the frequency of percolation clusters. A comparative analysis of the distributions of the relative frequencies in Fig. 3 shows that with appropriate occupational probabilities, the spatial distribution of relative frequencies on bounded lattices will be determined by the structure of the starting and target subsets, and not by the convexity of the cumulative function for the weighting distribution.

### 5 Results and discussion

One of the main characteristics for the percolation process is the percolation cluster probability \( P_\infty \). This value corresponds to the probability of finding an infinite cluster on an unbounded lattice [17], and it is defined as a function of the occupational probability \( p \) at the given percolation threshold \( p_c \) for this
Figure 3: Relative frequencies of lattice sites for weight distributions $S_1 \sim \text{Beta}(1, 2)$ (top row), $S_2 \sim \text{Beta}(1, 1)$ (center row), $S_3 \sim \text{Beta}(2, 1)$ (bottom row), and occupation probabilities $p_1 < p_c$ (left column), $p_2 > p_c$ (right column).
lattice:

\[
P_\infty(p) = \begin{cases} 
0, & p < p_c; \\
> 0, & p \geq p_c.
\end{cases} \quad (1)
\]

Note that the definition (1) is rarely used by researchers for applications, since it requires a percolation threshold \( p_c \) and the form of the probability function is undefined for \( p > p_c \).

On bounded lattices, the value of \( P_\infty \) can be estimated from the frequency of clusters connecting the starting and target site subsets. Using the statistics shown in Fig. 3, we can estimate this probability from the averaged relative frequencies of the sites along the upper boundary of the square lattice at \( y = 63 \).

In Fig. 4 shows estimates of the probability function of percolation clusters on the probability of site occupation \( P_\infty(p) \) for three weighted beta-distributed variables \( S_1, S_2, S_3 \), whose cumulative functions \( F_S(p) \) were shown earlier in Fig. 1. Red circles, green squares and blue diamonds correspond to statistical estimates of the probability of percolation clusters \( P_\infty(p) \) when the convexity of the weighing distribution function \( F_S(p) \) changes from negative at \( S_1 \sim \text{Beta}(1,2) \), through zero at \( S_2 \sim \text{Beta}(1,1) \), to positive at \( S_3 \sim \text{Beta}(2,1) \). In Fig. 4 we see that the finite-dimensional estimates of the percolation probability function \( P_\infty(p) \) describe the transition between the subcritical lattice states for \( p < p_c \) and the supercritical for \( p \geq p_c \).

Figure 4: Convergence the percolation probability function \( P_\infty(p) \) (circle, square, diamond symbols) to cumulative functions \( F_S(p) \) for beta-distributed weights \( S_1 \sim \text{Beta}(1,2) \) (red line), \( S_2 \sim \text{Beta}(1,1) \) (green line), \( S_3 \sim \text{Beta}(2,1) \) (blue line).

The vertical dashed line of green color in Fig. 4 corresponds to the percolation threshold \( p_{c,2} = 0.592746 \ldots \) known from [18] for a uniformly weighted square lattice. Note that the cumulative function \( F_{S_2}(p) \) of a uniformly distributed random variable \( S_2 \sim \text{Beta}(1,1) \) coincides with its argument on \( 0 \leq p \leq 1 \). Given this property of the cumulative function \( F_{S_2}(p) \) for a uniformly distributed
random variable $S_2$, the percolation threshold on a classical square lattice can be considered a quantile with a level $p_0 = 0.592746\ldots$ Level $F_{S_2}(p_{c2}) = p_0$ in Fig. 4 is shown by a horizontal dashed line of gray color. A comparative analysis of the finite-dimensional estimates of the percolation probability functions $P_\infty^*(p)$ in Fig. 4 leads us to the first empirical hypothesis.

**H1:** The percolation threshold $p_c$ on square lattices with a unit von Neumann neighborhood, weighted by a continuous random variable $S$, is a priori determined by a $p_0$-quantile:

$$ p_c = F_{S}^{-1}(p_0), $$

where $p_0 = 0.592746\ldots$

The a priori estimates of the percolation thresholds $p_{c1}$, $p_{c2}$ and $p_{c3}$ for the weighting variables $S_1$, $S_2$ and $S_3$ found from (2) are shown in Fig. 4 in red, green and blue vertical dashed lines. In Fig. 4 we see that all three estimates of the percolation threshold $p_{c1} = 0.361835\ldots$, $p_{c2} = 0.592746\ldots$ and $p_{c3} = 0.769900\ldots$ are equally consistent with the results of statistical modeling. We also see that since in the subcritical state the percolation probability function $P_\infty(p)$ converges to zero, and in the supercritical state, to the corresponding cumulative distribution function $F_S(p)$, this allows us to formulate a second empirical hypothesis.

**H2:** The percolation probability function $P_\infty(p)$ on an unbounded square lattice with a unit von Neumann neighborhood weighted by a random variable $S$ has the form:

$$ P_\infty(p) = \begin{cases} 0, & p < p_c; \\ F_S(p), & p \geq p_c, \end{cases} $$

where $p_c$ — percolation threshold; $F_S(p)$ — cumulative distribution function.

6 Conclusion

The theoretical analysis of the empirical hypotheses H1 and H2 formulated in this paper is an actual and interesting problem. For example, one of the key tasks required to analyze H1 hypothesis is to theoretically derive the value $p_0 = 0.592746\ldots$ of $p_0$-quantile. In our opinion, this value should be related to the structure of the unit von Neumann neighborhood on a square lattice. The solution of this and other problems will help researchers understand more clearly the origins of the numerous interrelationships of modern percolation theory with other areas of mathematics, physics and computer science, and will lead to the development of more accurate percolation models of natural phenomena with nonlinear cumulative distribution functions.
References

[1] D. Stauffer, A. Aharony, *Introduction to percolation theory* (CRC Press, 1994) 194.

[2] G.R. Grimmett, *Percolation* (Springer, 1999) 444.

[3] B. Zeimetz, B.A. Glowacki, J.E. Evetts, *Application of percolation theory to current transfer in granular superconductors*. Eur. Phys. J. B 29, (2002) 359.

[4] J. Hoshen, R. Kopelman, E.M. Monberg, *Percolation and cluster distribution. II. layers, variable-range interactions, and exciton cluster model*. J. Stat. Phys. 19, (1978) 219.

[5] M.H. Cohen, J. Jortner, I. Webman, *Percolation conductivity in granular metal films*. Phys. Rev. B 17, (1978) 4555.

[6] S. Kirkpatrick, *Percolation and conduction*. Rev. Mod. Phys. 45, (1973) 574.

[7] J. Wierman, *Mixed percolation on the square lattice*. J. Appl. Probab. 21, (1984) 247.

[8] E. Domany, W. Kinzel, *Equivalence of cellular automata to Ising models and directed percolation*. Phys. Rev. Lett. 53, (1984) 311.

[9] K. Ricciardi, G.F. Pinderb, K. Belitz, *Comparison of the lognormal and beta distribution functions to describe the uncertainty in permeability*. J. Hydrol. 313, (2005) 248.

[10] W. Nowak, R.L. Schwede, O.A. Cirpka, I. Neuweiler, *Probability density functions of hydraulic head and velocity in three-dimensional heterogeneous porous media*. Water Resour. Res. 44, (2008) W08452.

[11] R Core Team, *R: A language and environment for statistical computing*. URL: https://www.r-project.org

[12] P.V. Moskalev, *SPSL: Site percolation on square lattices*. URL: http://cran.r-project.org/package=SPSL

[13] P.V. Moskalev, *The structure of site percolation models on three-dimensional square lattices*. Computer Research and Modeling 5, (2013) 607.

[14] P.V. Moskalev, *Estimates of threshold and strength of percolation clusters on square lattices with (1,d)-neighborhood*. Computer Research and Modeling 6, (2014) 405.

[15] A.K. Gupta, S. Nadarajah, *Handbook of beta distribution and its applications* (CRC Press, 2004) 600.

[16] H.M. Taylor, K. Samuel, *An introduction to stochastic modeling* (Academic press, 2014) 631.
[17] H. Kesten, *Percolation theory for mathematicians* (Birkhäuser, 1982) 423.

[18] M.E.J. Newman, R.M. Ziff, *Efficient Monte Carlo algorithm and high-precision results for percolation*. Phys. Rev. Lett. 85, (2000) 4104.