

N. V. Kruglova, O. O. Dykhovychnyi, D.R. Lysenko, N. V. Bogdanova

SIMULATION OF DISTRIBUTION OF THE MAXIMUM OF THE CHENTSOV FIELD ON 'STEPS'

Abstract: The paper extends the results obtained by Paranjape, Park, Klesov and Kruglova. A method of finding the distribution of the maximum of a Chentsov field restriction on a polyline is suggested. To do this, it was proposed to simulate a random process that corresponds to the restriction. The suggested algorithm has high computational efficiency which is crucial due to a large number of points of division of the process trajectory. The selection of the distributions of extreme functionals from the field was performed by means of the statistical programming language R. The Weibull distribution was chosen as the most suitable distribution, which was confirmed by the Kolmogorov criterion, as well as by Q-Q and P-P diagrams.

Keywords: Chentsov field, percolation, Gaussian process, maximum distribution, modelling, language R.

Problem statement

The problem of finding the distribution of the maximum of a random field on certain subsets is a very common problem in random field statistics. We consider the problem of finding the distribution of the maximum on stepped polynomials occurring in solutions of percolation problems.

As a random field, we consider the two-parameter Chentsov field [1].

Definition 1. *A real separable Gaussian field $\{X(u, v) : (u, v) \in [0, 1] \times [0, 1]\}$ is called a Chentsov field, if it satisfies the following conditions:*

1. $X(0, v) = X(u, 0) = 0$ for all $u, v \in [0, 1]$;
2. $E[X(u, v)] = 0$ for all $(u, v) \in [0, 1] \times [0, 1]$;
3. $E[X(u_1, v_1)X(u_2, v_2)] = \min\{u_1, u_2\} \min\{v_1, v_2\}$ for all $(u_1, v_1) \in [0, 1] \times [0, 1]$ and $(u_2, v_2) \in [0, 1] \times [0, 1]$.

Analysis of recent research and publications

In general, the problem of finding the maximum of the Chentsov field on a unit square is still unsolved. Only partial cases have been studied. Thus, the exact distribution of $X(u, v)$ on a polyline with one point of refraction was found by Paranjape and Park [2].

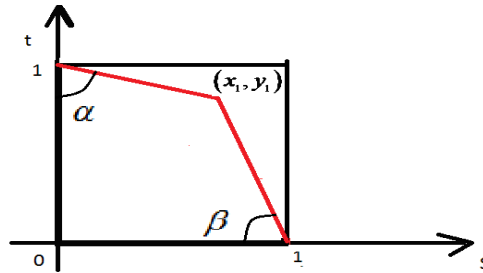


Figure 1. A polyline with one point of refraction

Let L be a polyline with one point of refraction shown in Fig. 1 and given by the following equation:

$$L = \left\{ (u, v) : v + ua^{-1} = 1, u \leq k; vb^{-1} + u = 1, u > k, (u, v) \in [0, 1] \times [0, 1] \right\}, \quad (1)$$

where $a = \frac{x_1}{1 - y_1}$, $b = \frac{y_1}{1 - x_1}$, $k = \frac{a(1 - b)}{1 - ab}$, $\pi/4 < \alpha, \beta < \pi/2$.

Theorem 1. (Paranjape, Park)[2] Let L be a polyline given by (1). Let $X(u, v)$ is Chentsov field. Then, for $x > 0$,

$$P \left\{ \sup_{(u,v) \in L} X(u, v) \leq x \right\} = \Phi \left(\frac{x(a + d)}{a\sqrt{d}} \right) - \Phi \left(\frac{x(d - a)}{a\sqrt{d}} \right) \exp \left\{ \frac{-2x^2}{a} \right\} -$$

$$+ \Phi \left\{ \frac{x(1 - bd)}{b\sqrt{d}} \right\} \exp \left\{ \frac{-2x^2}{b} \right\} + \exp \left\{ -2x^2(a^{-1} + b^{-1} - 2) \right\} \times$$

$$\times \Phi \left\{ x(b^{-1} - d - 2)d^{-1/2} \right\},$$

where $d = \frac{a(1 - b)}{b(1 - a)}$.

Park and Paranjape [2] have obtained the distribution of the maximum of the Chentsov field on the boundary of a square by considering a limit.

Theorem 2. [2]

$$P \left\{ \sup_{(u,v) \in \partial S} X(u; v) < x \right\} = -2\Phi(-x) + \Phi(-3x)e^{4x^2} + \Phi(x),$$

where ∂S — a boundary of the square.

Later I. I. Klesov, O. I. Klesov and N. V. Kruglova [4-6] obtained an exact distribution of the maximum of the Chentsov field over polylines with n points of refraction.

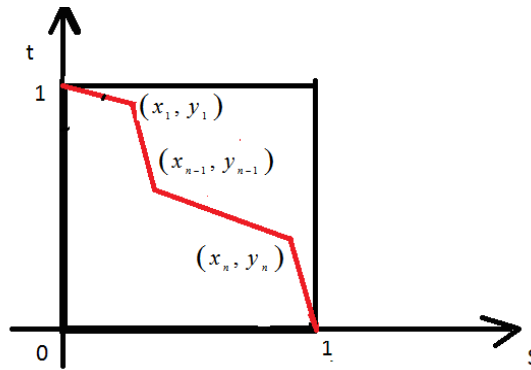


Figure 2. A polyline with n points of refraction

Let

$$0 = x_0 < x_1 < \dots < x_n < x_{n+1} = 1, \tag{2}$$

$$1 = y_0 \geq y_1 \geq \dots \geq y_n \geq y_{n+1} = 0. \tag{3}$$

Theorem 3. [6] *Let $X(u, v)$ be a Chentsov field on a unit square. Let $u_0 = 0$. Let L be a polyline with n points of refraction $(x_1, y_1), \dots, (x_n, y_n)$. Suppose the coordinates of these points satisfy conditions (2) - (3). Let's denote $\Delta_0 = 0, \Delta_i = x_i / y_i, i = \overline{1, n}$. Then, for $x > 0$,*

$$P \left\{ \sup_{(u;v) \in L} X(u;v) < x \right\} = \int_{-\infty}^{y_1} \dots \int_{-\infty}^{y_n} \left(1 - \exp \left\{ -2x \left(x / y_n - u_n \right) \right\} \right) \times \prod_{i=1}^n \left(1 - \exp \left\{ - \frac{2(x / y_{i-1} - u_{i-1})(x / y_i - u_i)}{(\Delta_i - \Delta_{i-1})} \right\} \right) \times \varphi_{0, \Delta_i - \Delta_{i-1}}(u_i - u_{i-1}) du_1 \dots du_n,$$

where $\varphi_{0, \Delta}(u) = \frac{1}{\sqrt{2\pi\Delta}} \cdot e^{-\frac{u^2}{2\Delta}}$ is a density of Gaussian distribution with parameters 0 and Δ .

Task definition

Note that the obtained theorems do not allow to find the maximum of the Chentsov field along stepped lines. The appropriate distributions could only be found by taking a corresponding limits, which may be problematic to justify. Instead, we will find an approximation of the Chentsov field maximum distribution by means of a numerical simulation.

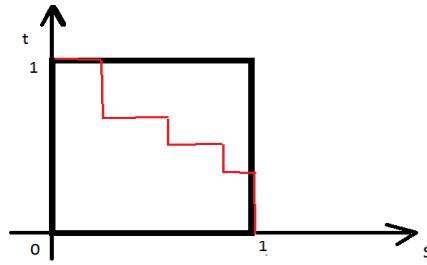


Figure 3. ‘Stepped’ polyline

An example of the proposed approach to finding the distribution of a polyline can be the polyline shown in Fig. 3. It is characterized by the fact that

$$y_{2i} = y_{2i+1}, \quad x_{2i+1} = x_{2i+2}, \quad i = \overline{0, \frac{n-1}{2}}. \quad (4)$$

This means that the first segment of polyline from Fig.3 is horizontal. The polyline has 99 refraction points, which have coordinates:

$Q_{2i-1} \left(\frac{i}{50}; 1 - \frac{i-1}{50} \right), i = \overline{1, 50}, \quad Q_{2i} \left(\frac{i}{50}; 1 - \frac{i}{50} \right), i = \overline{0, 50}$ and starts from point $(0, 1) = Q_0$ and ends in point $(1, 0) = Q_{100}$. Then polyline given by:

$$L = \left\{ (u, v) : u = \sum_{i=0}^{99} (e_i t + b_i) I_{[u_i, u_{i+1})}(t); v = \sum_{i=0}^{99} (c_i t + d_i) I_{[u_i, u_{i+1})}(t) \right\}, t \in [0, \sqrt{2}), \quad (5)$$

where $u_i = \frac{i}{50\sqrt{2}}, i = \overline{1, 100}, \quad e_{2(i-1)} = \sqrt{2}, e_{2i-1} = 0, i = \overline{1, 50},$

$$b_{2(i-1)} = -\frac{i-1}{50}, b_{2i-1} = \frac{i}{50}, i = \overline{1, 50},$$

$$c_{2(i-1)} = 0, c_{2i-1} = -\sqrt{2}, i = \overline{1, 50},$$

$$d_{2(i-1)} = 1 - \frac{i-1}{50}, d_{2i-1} = \frac{i}{50} + 1, i = \overline{1, 50}.$$

Let's denote $v_i(t) = c_i t + d_i, \quad a_i(t) = \frac{e_i t + b_i}{c_i t + d_i}, \quad i = \overline{0, 99}, \quad t \in [0, \sqrt{2}).$

The main research material

Let's model a process $X_L(t), t \in [0, \sqrt{2})$. It is a restriction of the Chentsov field to the above-mentioned polyline. Let's also find an empirical distribution

$P_{99}(x) = P \left\{ \max_{(u;v) \in L} X(u;v) < x \right\}$. We can take a modelling algorithm like in [7] because such

an algorithm is characterized by a high computational performance. By Dub's theorem [8], the following two processes are stochastically equivalent:

$$X_L(t) \sim \sum_{i=0}^{99} v_i(t) w(a_i(t)) I_{[u_i, u_{i+1})}(t), t \in [0, \sqrt{2}).$$

Let's denote

$$\Delta a(t_k) = a(t_k) - a(t_{k-1}), k = \overline{0, n};$$

$$t_k = \frac{k}{n}, k = \overline{0, n}.$$

Let's the discrete process $\hat{Y}(t)$ be modelled by:

$$\hat{Y}(0) = v(0) \xi_0 \sqrt{a(0)};$$

$$\hat{Y}(t_i) = v(t_i) \sum_{k=1}^i \sqrt{\Delta a(t_k)} \xi_k, i = \overline{0, n};$$

$\xi_0, \xi_1, \dots, \xi_n$ are standard independent Gaussian values.

Here is a fragment of code in the environment R. To find the empirical distribution of the maximum, we model 10^4 trajectories of the process $X_L(t)$, dividing a segment $[0; \sqrt{2}]$ into 1000 points. For each trajectory, we find the maximum that will form a sample of maxima and will determine the distribution.

To verify accuracy of the model, we find the mathematical expectation and the variance functions of the process $\hat{Y}(t)$. Let's graph the mathematical expectation of model $\hat{Y}(t)$ and process $X_L(t)$; variance functions $\hat{Y}(t)$ and $X_L(t)$.

Since $X_L(t)$ is the restriction of the Chentsov field to the polyline given by equation (5), by the definition of the Chentsov field:

$$E[X_L(t)] = 0, t \in [0, \sqrt{2}).$$

$$R_{X_L}(t, t) = E[(X_L(t))^2] = \sum_{i=0}^{99} (e_i t + b_i)(c_i t + d_i) I_{[u_i, u_{i+1})}(t) = \sum_{i=0}^{99} a_i(t) v_i^2 I_{[u_i, u_{i+1})}(t), t \in [0, \sqrt{2}).$$

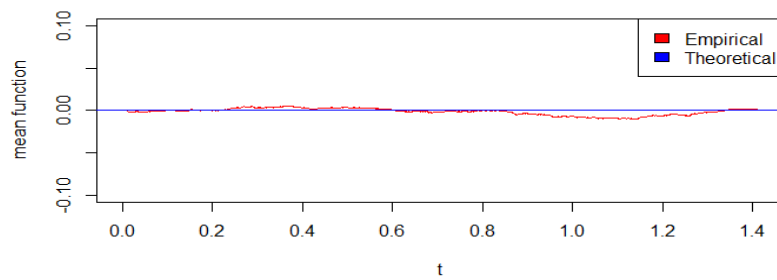


Figure 4. Plots of means

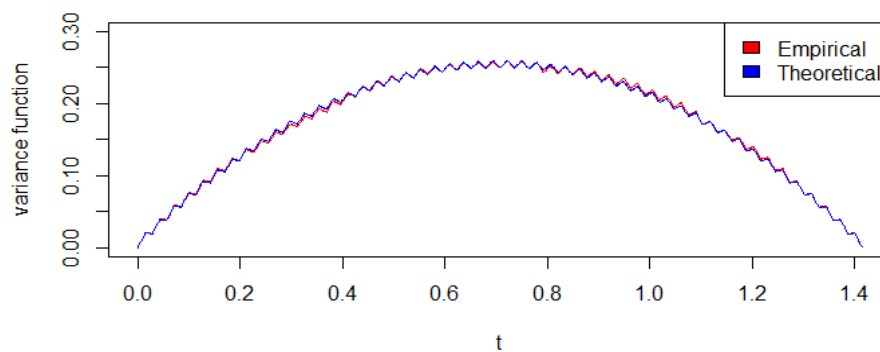


Figure 5. Plots of variance functions

```
# a sampling of the process trajectory
> e<-numeric(99)
> b<-numeric(99)
> c<-numeric(99)
> d<-numeric(99)
> for (i in 1:49)
+ {e[2*i]<-sqrt(2)
+ b[2*i]<--i/50
+ b[2*i-1]<-i/50
+ c[2*i-1]<--sqrt(2)
+ d[2*i]<-1-i/50
+ d[2*i-1]<-i/50+1
+ }
> e[99]<-0
> b[99]<-1
> c[99]<--sqrt(2)
> d[99]<-2
> u<-seq(1:100)/(50*sqrt(2))
> t<-seq(0,sqrt(2)-1/1000,length.out=1000)
```

```

> vt<-numeric(1000)
+ if (t[i]<u[1]){vt[i]<-1;at[i]<-sqrt(2)*t[i]} else if
> at<-numeric(1000)
> for (i in 1:1000)
+   for (k in 1:99)((t[i]>=u[k])&(t[i]<u[k+1])){vt[i]<-c[k]*t[i]+d[k];at[i]<-
(a[k]*t[i]+b[k])/vt[i]}
> D<-diff(at)
> m<-numeric(10^4)
> ex<-numeric(10^4)
> Dx<-numeric(10^4)
> x<-numeric(10^7)
> proc_l<-matrix(x,nrow=10^4)
> for(i in 1:10^4){
+
+   proc_l[i,]<-vt*c(0,cumsum(rnorm(999,0,sqrt(D))))
+ }
> ex<-colMeans(proc_l)
> plot(t,ex,col="red",type="l",xlab="t",ylab="mean function",ylim=c(-0.1,0.1))
> abline(a=0,b=0,col="blue")
> legend("topright", c("Empirical", "Theoretical"),fill=c("red","blue"))
> Dx<-colMeans((proc_l-ex)^2)
> plot(t,Dx,col="red",type="l",xlab="t",ylab="variance function",ylim=c(0,0.5))
> lines(t,at*vt^2,col="blue")
> legend("topright", c("Empirical", "Theoretical"),fill=c("red","blue"))

# Select the distribution
> library(fitdistrplus)
> dist1<-mledist(m[m>0],"weibull",lower = c(0.5, 0))
>   plotdist(m[m>0],      "weibull",      para=list(shape=dist1$estimate[1],
scale=dist1$estimate[2]))
>
ks.test(unique(m[m>0]),"pweibull",shape=dist1$estimate[1],scale=dist1$estimate[2])
> dgumbel<-function(x,a,b) 1/b*exp((a-x)/b)*exp(-exp((a-x)/b))
> pgumbel<-function(x,a,b) exp(-exp((a-x)/b))
> qgumbel<-function(x,a,b) (a-b*log(-log(x)))
> dist2 <- mledist(m[m>0],"gumbel",start=list(a=10,b=5))
> plotdist(m[m>0], "gumbel",para=list(a=dist2$estimate[1],b=dist2$estimate[2]))
> ks.test(unique(m[m>0]),"pgumbel",a=dist2$estimate[1],b=dist2$estimate[2])

```

We use the **fitdistrplus** package to find the "nearest" theoretical distribution for the empirical distribution law. We construct a histogram, an empirical distribution function and use P-P and Q-Q diagrams [9].

We compare the empirical distribution law with the Humbell and Weibull distributions.

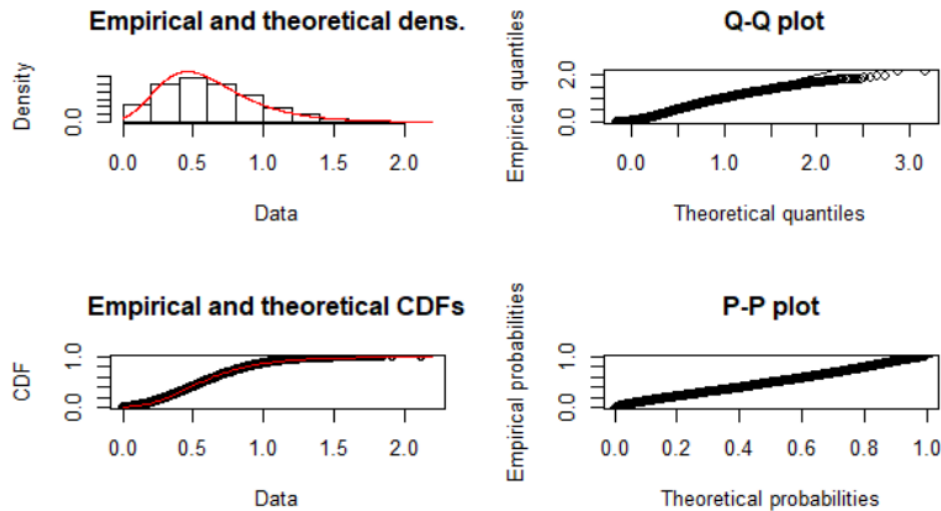


Figure 6. Humbell distribution

The Humbell distribution was considered with the parameters: $a = 0.460$, $b = 0.2748$, but the Kolmogorov's test showed that this distribution is not suitable ($p\text{-value} = 8.385e-07$).

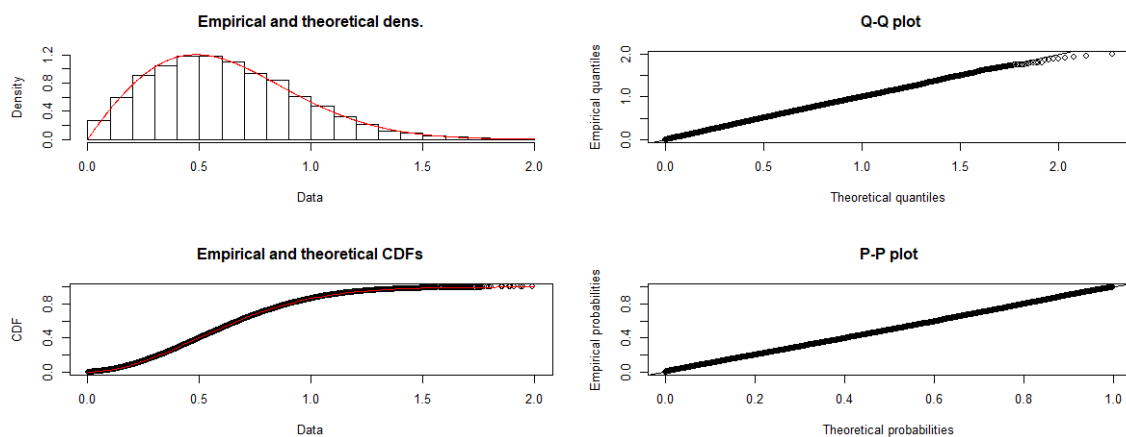


Figure 7. Weibull distribution

As can be seen from Figure 7 and the Kolmogorov's-Smirnov's test ($p\text{-value} = 0.3019$), the Weibull distribution with parameters shape = 1.94765, scale = 0.70068 can be taken as the theoretical distribution for $P_{99}(x) = P \left\{ \max_{(u;v) \in L} X(u;v) < x \right\}$. This does not contradict the results presented in [10].

Conclusions

A method for finding an exact approximation for the distribution of functionals from the Chentsov field was proposed. Numerical modelling by means of the language R of a restriction of the Chentsov field on a polyline is carried out. Graphs of mathematical expectations of the model and the given process are constructed. To maximize the restriction of the Chentsov field on a polyline with many refraction points, the "most appropriate" theoretical distribution was found, which turned out to be the Weibull distribution.

REFERENCES

1. Yeh J., Wiener Measure in a Space of Functions of Two Variables// Transactions of the American Mathematical Society, 95, –1960. –P. 433–450.
2. S. R. Paranjape and C. Park *Distribution of the supremum of the two-parameter Yeh-Wiener process on the boundary*, J. Appl. Probab., **10** (1973), no 4, –P. 875-880.
3. C. Park and F. J. Schuurmann *Evaluations of barrier-crossing probabilities of wiener paths*, J. Appl. Prob., **13** (1976), –P. 267-275.
4. N. V. Kruglova *Distribution of the maximum of the Chentsov random field*, Theor. Stoch. Process. (2008), no 1, –P. 76–81.
5. I.I. Klesov *On the probability of attainment of a curvilinear level by a Wiener field*, Probab and Math. Statist. (1995), no 51, –P. 63–67.
6. O.I. Klesov, N. V. Kruglova *Distribution Of the functionals of the Chentsov random field. R^3* , Naukovi visti NTUU-KPI, **6** (2007), –P. 145-150 (in ukrainian).
7. O.O. Dykhovychnyi, N.V. Kruglova, O.I. Virstiuk *Methods for identification of probability distribution of random variables from data samples with R statistical computing language*, Mathematics in Modern Technical University, **1** (2018), –P. 91-100 (in ukrainian).
8. Doob J. L. *Heuristic approach to Kolmogorov-Smirnov theorems*, Ann. Math. Statist., **20** (1949), –P. 393–403.
9. URL: <http://cran.us.r-project.org>.
10. Gnedenko B.V. *Sur la distribution limite du terme maximum d'une s'erie al'eatoire*. Ann. Math., 44 (1943), –P. 423-453.