

Is Pool Testing Method of COVID-19 Employed in Germany and India Effective?

Jia Liu (✉ whutrobin@163.com)

School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China

Yi Chen

School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China

Kefan Xie

School of Management, Wuhan University of Technology, Wuhan 430070, China

Xiaohong Chen

School of Business, Central South University, Changsha 410083, China

Research Article

Keywords: COVID-19, Novel Coronavirus, Nucleic Acid Testing, Pool Testing Method

Posted Date: May 15th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-29389/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Is Pool Testing Method of COVID-19 Employed in Germany and India Effective?

LIU Jia^{1*}, CHEN Yi¹, XIE Kefan², CHEN Xiaohong³

¹School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073, China

²School of Management, Wuhan University of Technology, Wuhan 430070, China

³School of Business, Central South University, Changsha 410083, China

*Corresponding author, email: whutrobin@163.com

Abstract

At present, several countries, such as Germany and India, have employed a pool testing method on the nucleic acid testing of COVID-19 for the shortage of detection kits. In this method, the testing is performed on several samples of the cases together as a bunch. If the test result of the bunch is negative, then it is shown that none of the cases in the bunch has been infected with the novel coronavirus. On the contrary, if the test result of the bunch is positive, then the samples are tested one by one to confirm which cases are infected. We verified that the pool testing method of COVID-19 is effective in the situation of the shortage of nucleic acid detection kits based on probabilistic modeling. Moreover, the following interesting results are also obtained. (1) If the infection rate is extremely low, while the same number of detection kits are used, the expected number of cases that can be tested by the pool testing method is far more than that by the one-by-one testing method. (2) The pool testing method is effective only when the infection rate is less than 0.3078. While the infection rate decreases from 0.3078 to 0.0018, the optimal sample sizes in one bunch increases from 3 to 25. In general, the higher the infection rate, the smaller the optimal sample size in one bunch. (3) If N samples are tested by the pool testing method, while the sample size in one bunch is G , the number of detection kits required is in the interval $(N/G, N)$. Additionally, the lower the infection rate, the fewer detection kits are needed. Therefore, the pool testing method is not only suitable for the situation of the shortage of detection kits, but also the situation of the overall or sampling detection for a large population.

Keywords

COVID-19, Novel Coronavirus, Nucleic Acid Testing, Pool Testing Method

1 Introduction

At present, COVID-19 epidemic outbreaks in most countries all over the world. The epidemic should last for a long time due to its strong infectivity (Yang and Wang, 2020). To minimize the negative impact of the epidemic, each country has adopted its characteristic response methods.

The recent research on the response of COVID-19 focuses on the techniques and methods to physically interrupt the transmission chain of the novel coronavirus. The analysis results of Haider et al. (2020) and Wells et al. (2020) implies that the virus spreads rapidly around the world as a large number of infected people travel internationally, and border controls between countries can help stem the spread of the virus. Hellewell et al. (2020) simulated the effects of tracing close contacts and isolating suspected cases in different basic reproduction numbers. The results show that the two measures could interrupt the transmission of the virus in three months. Boulos and Geraghty (2020) proposed the technique and method of tracking the infected cases and their close contacts based on geographic information systems. Additionally, the media plays a pivotal role in slowing the spread of the epidemic. In this respect, Zhou et al. (2020) discovered the quantitative relationship between the media reports and the progress of the epidemic. Yin et al. (2020) proposed the mode of public opinion transmission and its guidance method.

The experience of China suggests that rapid detection and isolation of the infected cases are of vital importance in controlling the development of the epidemic (Cyranski, 2020). Nucleic acid testing (NAT) is a normal way to detect the infection of the novel coronavirus (Hu et al., 2020). In the early stage of the epidemic in many countries and regions, NAT capacity should be insufficient to enable the one-by-one test of all the suspected cases. To improve the pertinence of the NAT, Ghinai et al. (2020) made a statistical analysis of the NAT results of the close contacts of confirmed cases and found the ones who are most likely to be infected. Additionally, several countries, such as Germany and India, have employed a pool testing method. Specifically, NAT is performed on several samples of the cases together as a bunch. If the test result of the bunch is negative (hereinafter referred to as the “uninfected bunch”), then none of the cases in the bunch has been infected with the novel coronavirus. On the contrary, if the test result of the bunch is positive (hereinafter referred to as the “infected bunch”), then at least one of the cases in the bunch has been infected with the novel coronavirus. In this case, the samples are tested one by one to determine which cases are infected.

Therefore, this research aims to analyze the effectiveness of the pool testing method and propose an approach to maximize the detection efficiency with limited nucleic acid testing kits. The remainder of this paper is organized as follows. Section 2 calculates the expected number of testing times of overall infection cases in a country or region. Section 3 compares the efficiency of the pool testing and the one-by-one testing method with limited detection kits. Section 4 proposes the optimal sample sizes in one bunch at different infection rates.

2 Expected number of testing times of overall infection cases

While NAT is carried out by the pool testing method, the expected number of testing times is affected by the total population and infection rate of the country or region, as well as the sample size in one bunch. Let T denote the total testing number, N denote the total population, R denote infection rate, G denote the sample size in one bunch, and U denote the number of uninfected bunches. Each bunch is tested for nucleic acids in sequence.

The probability that the first bunch is an uninfected bunch is

$$P(1) = \frac{C_{(1-R)N}^G}{C_N^G}. \quad (1)$$

Let x_1 ($x_1 = 0, 1, \dots, G$) denote the number of uninfected cases in the first bunch. The probability that there are x_1 uninfected cases detected in the first bunch is $\frac{C_{(1-R)N}^{x_1} \cdot C_{RN}^{G-x_1}}{C_N^G}$. Based on the total probability formula, the probability that the second bunch is an uninfected bunch can be obtained as

$$P(2) = \frac{\sum_{x_1=0}^G C_{(1-R)N}^{x_1} \cdot C_{RN}^{G-x_1} \cdot C_{(1-R)N-x_1}^G}{C_N^G \cdot C_{N-G}^G}. \quad (2)$$

In equation (2), $\sum_{x_1=0}^G C_{(1-R)N}^{x_1} \cdot C_{RN}^{G-x_1} \cdot C_{(1-R)N-x_1}^G$ means the number of basic events that x_1 uninfected cases are randomly selected from $(1-R)N$ uninfected ones, and then $G-x_1$ infected cases are randomly selected from RN infected ones (here a bunch of G cases is formed), and finally G uninfected cases are randomly selected from the rest $(1-R)N-x_1$ uninfected ones. This is equivalent to the number of basic events that G uninfected cases are randomly selected from $(1-R)N$ uninfected ones, and then G cases are randomly selected from the rest $N-G$ cases. Therefore, the number of basic events above is $C_{(1-R)N}^G \cdot C_{N-G}^G$. Consequently, we have

$$\sum_{x_1=0}^G C_{(1-R)N}^{x_1} \cdot C_{RN}^{G-x_1} \cdot C_{(1-R)N-x_1}^G = C_{(1-R)N}^G \cdot C_{N-G}^G.$$

Therefore, we obtain

$$P(2) = \frac{C_{(1-R)N}^G}{C_N^G} = P(1).$$

Let x_2 ($x_2 = 0, 1, \dots, G$) denote the number of uninfected cases in the second bunch. Based on the total probability formula, the probability that the third bunch is an uninfected bunch can be obtained as

$$P(3) = \frac{\sum_{x_1=0}^G C_{(1-R)N}^{x_1} \cdot C_{RN}^{G-x_1}}{C_N^G} \cdot \frac{\sum_{x_2=0}^G C_{(1-R)N-x_1}^{x_2} C_{RN-G+x_1}^{G-x_2} C_{(1-R)N-x_2-x_1}^G}{C_{N-G}^G \cdot C_{N-2G}^G}. \quad (3)$$

In equation (3), $\sum_{x_2=0}^G C_{(1-R)N-x_1}^{x_2} C_{RN-G+x_1}^{G-x_2} C_{(1-R)N-x_2-x_1}^G$ means the number of basic events that after the first bunch is designated, x_2 uninfected cases are randomly selected from $(1-R)N-x_1$ uninfected ones, and then $G-x_2$ infected cases are randomly selected from $RN-G+x_1$ infected ones, and finally G uninfected cases are randomly selected from the rest $(1-R)N-x_1-x_2$ uninfected ones. This is equivalent to the number of basic events that G uninfected cases are randomly selected from $(1-R)N-x_1$ uninfected ones, and then G cases are randomly selected from the rest $N-2G$ cases. Therefore, the number of basic events is $C_{(1-R)N-x_1}^G \cdot C_{N-2G}^G$.

Consequently, we have

$$\sum_{x_2=0}^G C_{(1-R)N-x_1}^{x_2} C_{RN-G+x_1}^{G-x_2} C_{(1-R)N-x_2-x_1}^G = C_{(1-R)N-x_1}^G \cdot C_{N-2G}^G.$$

Therefore, we obtain

$$P(3) = \frac{\sum_{x_1=0}^G C_{(1-R)N}^{x_1} \cdot C_{RN}^{G-x_1}}{C_N^G} \cdot \frac{C_{(1-R)N-x_1}^G}{C_{N-G}^G} = P(2).$$

Similarly, the following equation can be also obtained successively:

$$P(k+1) = P(k), (k = 1, 2, \dots).$$

Therefore, the probability of whichever bunch being an uninfected bunch is equal to $P(1)$. While $N \gg G$, this probability can be considered as

$$P = (1-R)^G.$$

Additionally, the bunch number can be considered as $\frac{N}{G}$. Consequently, the expected number of uninfected bunches is

$$E(U) = \frac{N}{G} \cdot (1-R)^G. \quad (4)$$

It can be seen from equation (4), the value of $E(U)$ is equal to that of sampling with replacement. Therefore, the expected number of testing times of overall infection cases is

$$E(T) = N \left(\frac{1}{G} + 1 - (1 - R)^G \right). \quad (5)$$

In practice, equation (5) can only be applied in the condition of $G \geq 2$. Whereas, in one-by-one testing (i.e. $G = 1$), the expected number of testing times of overall infection cases is $E(T) = N$.

3 Testing Efficiency with Limited Detection Kits

Let m denote the upper limit of the number of testing times per unit time in a country or region. While employing the one-by-one testing method, the number of cases that can be tested is m . Then, for a country or region with a total population of N , the probability of each person being tested is $\frac{m}{N}$, and the expected number of infected cases that can be detected is mR .

While employing the pool testing method in which the sample size in one bunch is G , the expected number of cases who can be tested is $\frac{m}{\frac{1}{G} + 1 - (1 - R)^G}$. Then, for a country or region with a total population of N , the probability of each person being tested is $\frac{m}{N(\frac{1}{G} + 1 - (1 - R)^G)}$, and the expected number of infected cases which can be detected is $\frac{mR}{\frac{1}{G} + 1 - (1 - R)^G}$.

It can be seen that the expected number of cases who can be tested by the pool testing method is approximately G times higher than by the one-by-one testing method while the infection rate is extremely low.

4 Optimal Sample Size in One Bunch

The optimal sample size in one bunch, denoted by $Opt(G)$, is defined as the sample size that minimizes the expected number of testing times (L) per case in the bunch. According to the previous analysis, L is determined both by the sample size in one bunch (G) and infection rate (R). While the detection kits are sufficient, the expected number of testing times per case in the one-by-one testing is 1; the expected number of testing times per case in the pool testing is

$$L(G, R) = \frac{1}{G} + 1 - (1 - R)^G. \quad (6)$$

While $L(G, R)$ is larger than 1, the expected number of testing times per case in the pool testing is more than that in one-by-one testing. That is,

$$G \ln(1 - R) + \ln G < 0.$$

Then, we set

$$F(G) = G \ln(1 - R) + \ln G.$$

Obviously, $F(G)$ increases monotonically in the interval $(2, \frac{1}{-\ln(1-R)})$ and decreases monotonically in the interval $(\frac{1}{-\ln(1-R)}, N)$. Therefore, the maximum of

$F(G)$ is $F(\frac{1}{-\ln(1-R)})$. For any $R \in (0, 1)$,

$$F\left(\frac{1}{-\ln(1-R)}\right) = -1 + \ln\left(\frac{1}{-\ln(1-R)}\right) \leq 0$$

gives

$$F(G) < 0.$$

In this case, we have

$$R > 1 - e^{-\frac{1}{e}}, \quad (7)$$

That is, in the condition of $R \geq 0.3078$, the one-by-one testing method is strictly superior to the pool testing method.

On the contrary, in the condition of $R < 0.3078$, the expected number of testing times per case in the pool testing is smaller than that in the one-by-one testing. Moreover, we obtain

$$L(2, R) - L(3, R) = \frac{1}{6} - R(1 - R)^2$$

which is greater than zero in the condition of $R \in (0, 1)$. Therefore, the expected number of testing times per case in the condition of $G = 2$ is strictly greater than that in the condition of $G = 3$. In other words, the optimal sample size in one bunch should be no less than 3. In the condition of $G \geq 3$ and $0 < R < 0.3078$, the function graph of $L(G, R)$ is depicted in Figure 1.

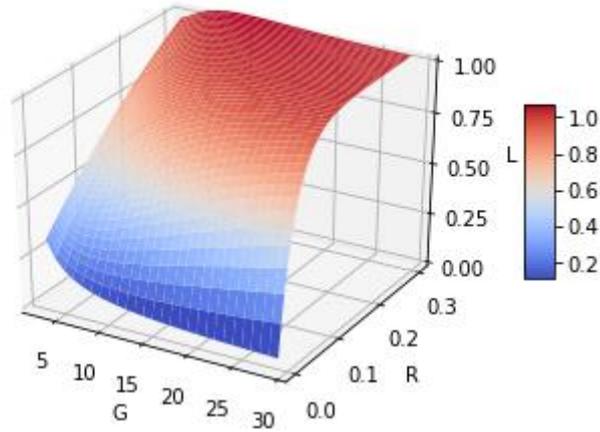


Figure 1 function graph of $L(G, R)$

In the condition of $R \in (0, 0.3078)$, as G increases, L increases significantly from its minimum, then decreases slightly, and finally approaches 1. It also can be seen that the larger the value of R , the smaller the optimal sample size in one bunch. Obviously, while G is equal to 1 (the minimum value), the number of testing times per case is 1; while G is equal to N (the maximum value), the expected number of testing times per case is also 1. On this basis, we set

$$K(G, R) = L(G + 1, R) - L(G, R) = R(1 - R)^G - \frac{1}{G(G + 1)}. \quad (8)$$

The function graph of $K(G, R)$ is depicted in Figure 2.

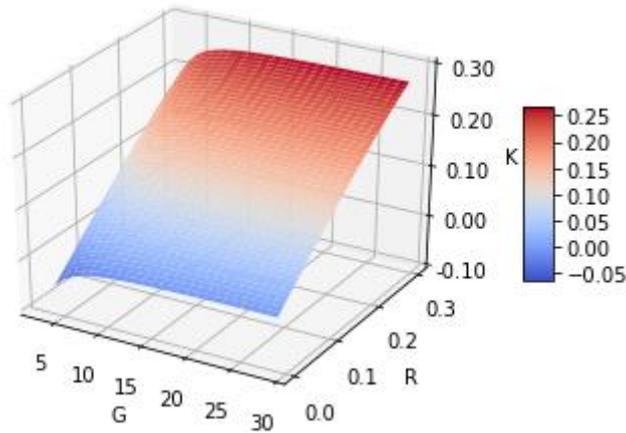


Figure 2 function graph of $K(G, R)$

In the condition of $R \in (0, 0.3078)$, as G increases from 3, K increases monotonically from negative to positive. We set $G = g$ if $K = 0$. $[g]$ is the integer part of g . Then, $G \leq [g]$ gives $K < 0$ which is $L(G + 1, R) < L(G, R)$, and $G \geq [g] + 1$ gives $K > 0$ which is $L(G + 1, R) > L(G, R)$. At this point, the optimal sample size in one bunch $Opt(G)$ is $[g] + 1$. In other words, in the condition of $G \in$

$([g], [g] + 1]$, we obtain $Opt(G) = [g] + 1$.

In the condition of $G \geq 3$ and $R \in (0, 0.3078)$, as R increases, K increases monotonically from negative to positive. We set $R = r_1$ if $G = [g]$ and $K = 0$; $R = r_2$ if $G = [g] + 1$ and $K = 0$. Therefore, we obtain $Opt(G) = [g] + 1$ in the condition of $R \in (r_1, r_2]$.

Based on the analysis above, we figure out the value of r_1, r_2 by traversing R in the interval $(0, 0.3078)$ by a step size of 0.0001 in the condition of $Opt(G) = 3, 4, 5, \dots$. Consequently, we obtain the optimal sample size in one bunch $Opt(G)$ corresponding to different infection rate intervals $(r_1, r_2]$, which is demonstrated in Table 1.

Table 1 optimal sample size in one bunch corresponding to different infection rates

Infection rate intervals	Optimal sample size in one bunch
No less than 0.3078	1 (one-by-one testing)
0.1240-0.3078	3
0.0656-0.1240	4
0.0412-0.0656	5
0.0283-0.0412	6
0.0207-0.0283	7
0.0158-0.0207	8
0.0125-0.0158	9
0.0101-0.0125	10
0.0084-0.0101	11
0.0070-0.0084	12
0.0060-0.0070	13
0.0052-0.0060	14
0.0045-0.0052	15
0.0040-0.0045	16
0.0035-0.0040	17
0.0031-0.0035	18
0.0028-0.0031	19
0.0026-0.0028	20
0.0023-0.0026	21
0.0021-0.0023	22

0.0019-0.0021	23
0.0018-0.0019	24
Less than 0.0018	25 or larger

5 Conclusions

This research verifies the effectiveness of the pool testing method of COVID-19 employed in Germany and India. Additionally, the expected number of testing times of overall infection cases in a country or region by the pool testing method is proposed. The research compares the efficiency of the pool testing method and the one-by-one testing method in the condition of limited nucleic acid testing kits. The results show that the cases that can be tested by the pool testing method are several times more than that by the one-by-one testing method while the infection rate is extremely low. However, the pool testing method is effective only when the infection rate is less than 0.3078. For the pool testing, the higher the infection rate, the smaller the optimal sample sizes in one bunch. This research also proposes the optimal sample sizes in one bunch corresponding to different infection rates. In practice, the infection rate in a country or region can be estimated by random sampling and testing, and then the sample size in one bunch can be determined.

Acknowledgement

This research was funded by the National Natural Science Foundation of China (Nos. 71603284 and 71790615). The authors declare no conflict of interest.

References

- [1] Hellewell J. et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health*, 2020, 8(4): E488-E496.
- [2] Wells C.R. et al. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(13): 7504-7509.
- [3] Boulos M.K.N., Geraghty E.M. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *International Journal*

of Health Geographics, 2020, 19(1): 8.

[4] Yin F.L., Lv J.H., Zhang X.J., Xia X.Y., Wu J.H. COVID-19 information propagation dynamics in the Chinese Sina-microblog. *Mathematical Biosciences and Engineering*, 2020, 17(3): 2676-2692.

[5] Zhou W.K., Wang A.L., Xia F., Xiao Y.N., Tang S.Y. Effects of media reporting on mitigating spread of COVID-19 in the early phase of the outbreak. *Mathematical Biosciences and Engineering*, 2020, 17(3): 2693-2707.

[6] Yang C.Y., Wang J. A mathematical model for the novel coronavirus epidemic in Wuhan, China. *Mathematical Biosciences and Engineering*, 2020, 17(3): 2708-2724.

[7] Haider et al. Passengers' destinations from China: low risk of Novel Coronavirus (2019-nCoV) transmission into Africa and South America. *Epidemiology and Infection*, 2020, 148: e41.

[8] Hu et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Science China-Life Sciences*, 2020, DOI: 10.1007/s11427-020-1661-4 (Online)

[9] Ghinai et al. First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. *LANCET*, 2020, 395(10230): 1137-1144.

[10] Cyranoski D. What China's coronavirus response can teach the rest of the world. *Nature*, 2020, 579(7800): 479-480.

Figures

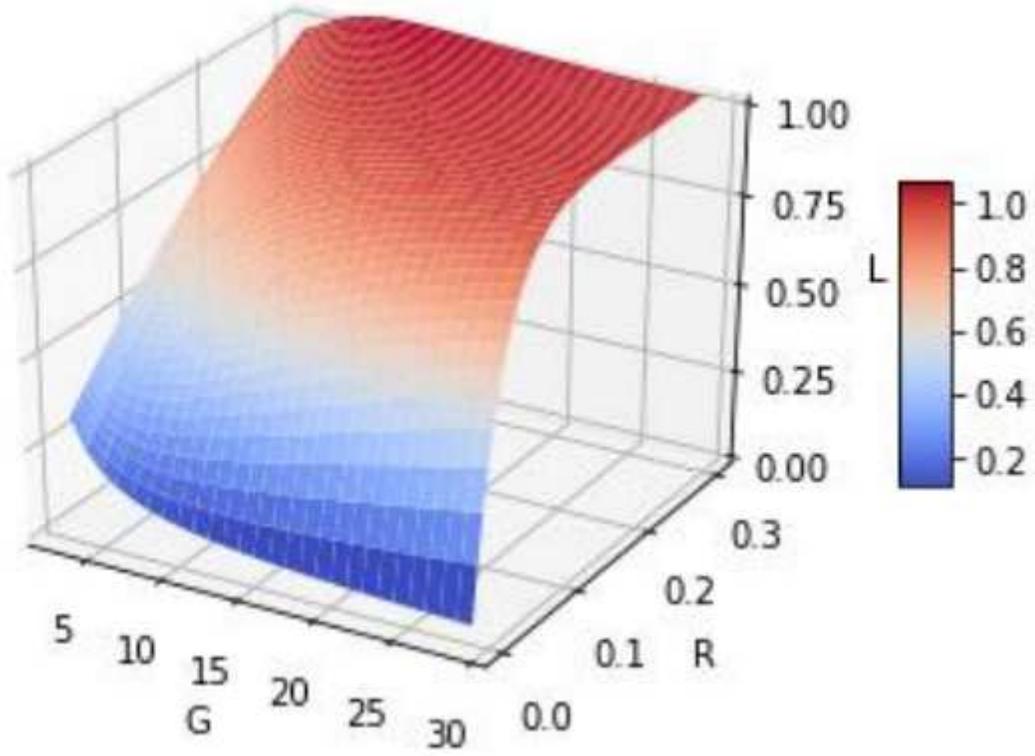


Figure 1

Function graph of $L(G, R)$

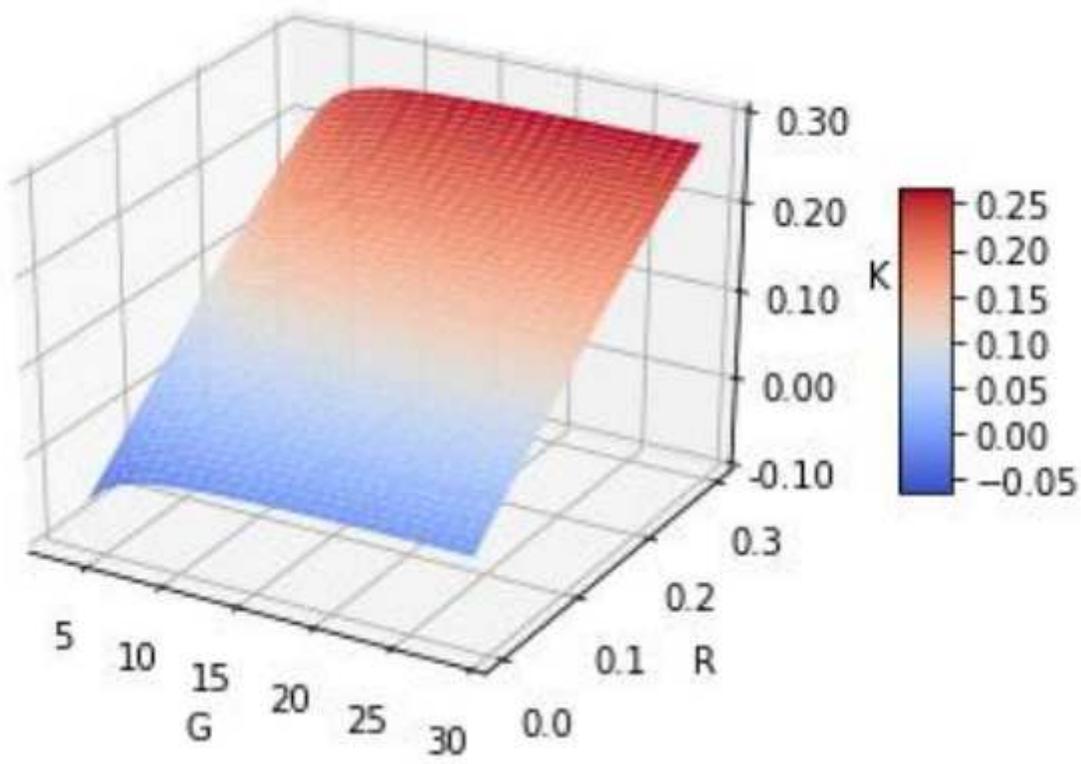


Figure 2

Function graph of $K(G, R)$