**Manuscript Number:** PONE-D-22-05986  
**Article Type:** Research Article  
**Full Title:** Genetic algorithm with a new round-robin based tournament selection: statistical properties analysis  
**Short Title:** Genetic algorithm with a new selection operator  
**Corresponding Author:** Abid Hussain  
Government College  
Rawalpindi, PAKISTAN  
**Keywords:** Genetic algorithms; selection schemes; selection pressure; diversity; statistical analysis; traveling salesman problem.  
**Abstract:**  
A round-robin tournament is a contest where each and every player plays with all the other players. In this study, we propose a round-robin based tournament selection operator for the genetic algorithms (GAs). At first, we divide the whole population into two equal and disjoint groups, then each individual of a group competes with all the individuals of other group. Statistical experimental results reveal that the devised selection operator has a relatively better selection pressure along with a minimal loss of population diversity. For the consisting of assigned probability distribution with sampling algorithms, we employ the Pearson's chi-square and the empirical distribution function as goodness of fit tests for the analysis of statistical properties analysis. At the cost of a nominal increase of the complexity as compared to conventional selection approaches, it has improved the sampling accuracy. Finally, for the global performance, we considered the traveling salesman problem to measure the efficiency of the newly developed selection scheme with respect to other competing selection operators and observed an improved performance.  
**Order of Authors:**  
Abid Hussain  
Salma Riaz, PhD  
Muhammad Sohail Amjad, PhD  
Ehtasham ul Haq, PhD  
**Additional Information:**  
**Question**  
**Response**  
**Financial Disclosure**  
The author(s) received no specific funding for this work.
This statement is required for submission and will appear in the published article if the submission is accepted. Please make sure it is accurate.

**Unfunded studies**
Enter: The author(s) received no specific funding for this work.

**Funded studies**
Enter a statement with the following details:
- Initials of the authors who received each award
- Grant numbers awarded to each author
- The full name of each funder
- URL of each funder website
- Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript?
  - **NO** - Include this sentence at the end of your statement: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
  - **YES** - Specify the role(s) played.

* typeset

**Competing Interests**

Use the instructions below to enter a competing interest statement for this submission. On behalf of all authors, disclose any competing interests that could be perceived to bias this work—acknowledging all financial support and any other relevant financial or non-financial competing interests.

The authors have declared that no competing interests exist.
View published research articles from *PLOS ONE* for specific examples.

**NO authors have competing interests**

Enter: *The authors have declared that no competing interests exist.*

**Authors with competing interests**

Enter competing interest details beginning with this statement:

*I have read the journal’s policy and the authors of this manuscript have the following competing interests: [insert competing interests here]*

* typeset

**Ethics Statement**

Enter an ethics statement for this submission. This statement is required if the study involved:

- Human participants
- Human specimens or tissue
- Vertebrate animals or cephalopods
- Vertebrate embryos or tissues
- Field research

Write "N/A" if the submission does not require an ethics statement.

General guidance is provided below. Consult the [submission guidelines](https://www.plosone.org/submission-guidelines) for detailed instructions. **Make sure that all information entered here is included in the Methods section of the manuscript.**

N/A
| Format for specific study types |
|---------------------------------|
| **Human Subject Research (involving human participants and/or tissue)** |
| • Give the name of the institutional review board or ethics committee that approved the study |
| • Include the approval number and/or a statement indicating approval of this research |
| • Indicate the form of consent obtained (written/oral) or the reason that consent was not obtained (e.g. the data were analyzed anonymously) |

| Animal Research (involving vertebrate animals, embryos or tissues) |
|------------------------------------------------------------------|
| • Provide the name of the Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board that reviewed the study protocol, and indicate whether they approved this research or granted a formal waiver of ethical approval |
| • Include an approval number if one was obtained |
| • If the study involved non-human primates, add additional details about animal welfare and steps taken to ameliorate suffering |
| • If anesthesia, euthanasia, or any kind of animal sacrifice is part of the study, include briefly which substances and/or methods were applied |

| Field Research |
|----------------|
| Include the following details if this study involves the collection of plant, animal, or other materials from a natural setting: |
| • Field permit number |
| • Name of the institution or relevant body that granted permission |

| Data Availability |
|-------------------|
| Authors are required to make all data underlying the findings described fully available, without restriction, and from the time of publication. PLOS allows rare exceptions to address legal and ethical concerns. See the [PLOS Data Policy](https://www.plos.org/data-policy) and [FAQ](https://www.plos.org/data-faq) for detailed information. |

Yes - all data are fully available without restriction
A Data Availability Statement describing where the data can be found is required at submission. Your answers to this question constitute the Data Availability Statement and will be published in the article, if accepted.

**Important:** Stating ‘data available on request from the author’ is not sufficient. If your data are only available upon request, select ‘No’ for the first question and explain your exceptional situation in the text box.

Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?

Describe where the data may be found in full sentences. If you are copying our sample text, replace any instances of XXX with the appropriate details.

- If the data are **held or will be held in a public repository**, include URLs, accession numbers or DOIs. If this information will only be available after acceptance, indicate this by ticking the box below. For example: All XXX files are available from the XXX database (accession number(s) XXX, XXX).
- If the data are all contained **within the manuscript and/or Supporting Information files**, enter the following: All relevant data are within the manuscript and its Supporting Information files.
- If neither of these applies but you are able to provide **details of access elsewhere**, with or without limitations, please do so. For example:

  Data cannot be shared publicly because of [XXX]. Data are available from the XXX Institutional Data Access / Ethics Committee (contact via XXX) for researchers who meet the criteria for access to confidential data.

  The data underlying the results presented in the study are available from [include the name of the third party]
and contact information or URL).  

- This text is appropriate if the data are owned by a third party and authors do not have permission to share the data.

* typeset

Additional data availability information:
Genetic algorithm with a new round-robin based tournament selection: statistical properties analysis

Abid Hussain¹,*; Salma Riaz²; M.S. Amjad³; and Ehtasham ul Haq⁴

¹Department of Statistics, Government College Khayaban-e-sir Syed, Rawalpindi, Pakistan.
²Department of Applied Mathematics & Statistics, Institute of Space Technology, Islamabad, Pakistan.
³School of Applied Sciences & Humanities, National University of Technology, Islamabad, Pakistan.
⁴Department of Mathematics and Statistics, International Islamic University, Islamabad, Pakistan.

Abstract

A round-robin tournament is a contest where each and every player plays with all the other players. In this study, we propose a round-robin based tournament selection operator for the genetic algorithms (GAs). At first, we divide the whole population into two equal and disjoint groups, then each individual of a group competes with all the individuals of other group. Statistical experimental results reveal that the devised selection operator has a relatively better selection pressure along with a minimal loss of population diversity. For the consisting of assigned probability distribution with sampling algorithms, we employ the Pearson’s chi-square and the empirical distribution function as goodness of fit tests for the analysis of statistical properties analysis. At the cost of a nominal increase of the complexity as compared to conventional selection approaches, it has improved the sampling accuracy. Finally, for the global performance, we considered the traveling salesman problem to measure the efficiency of the newly developed selection scheme with respect to other competing selection operators and observed an improved performance.

Keywords: Genetic algorithms; selection schemes; selection pressure; diversity; statistical analysis; traveling salesman problem.

1 Introduction

Genetic algorithms (GAs) are stochastic approaches for optimization, based on natural mechanisms of genetics. These algorithms refer to the natural selection process where the most fitted individuals for reproduction. Generally, five stages are considered in GA: initial population, fitness function, selection, crossover and mutation operators. If parents have good fitness, their offspring will be better than them and have a better chance to survive. The process continues and eventually a generation is found with the most qualified individuals.

The development of GAs originates from the influential work of Holland (1975). Many scholars have acknowledged GA as it is considered a key member of optimization related research. The global environment, robustness actions and reliability are the main reasons for its popularity. For example, Ruxton and Beauchamp (2008) highlighted the applicability of GAs in the field of behavioral ecology to discover the vigilance behavior in animals. Further, Song et al. (2016) employed GAs to achieve optimal satellite selection for global positioning system (GPS) use. In recent times, Soui et al. (2019) explored GAs in modeling of membership behavior in the stock market and to increase the scalability in credit risk assessment. Other than these, many applications of GAs can be perceived in the multidisciplinary
research literature, such as, interlocking medicine (Forrest et al., 1993) and artificial intelligence (Rubio et al., 2016). The ability to solve complex multidimensional and multi-models optimization problems with the minimum information required about their objective functions, is also a reason of its popularity, see, for example, Beyer (1997) and Amsa et al. (2013). Someone can consult to Pandey et al. (2014) for a more detailed account of promising features and a comprehensive overview of GA.

In all domains of life, GAs are found to be effective but still there is the issue of premature convergence in pursuit for optimal solutions, see, for example, Fogel (1994) and Hussain and Muhammad (2020). The complications of premature convergence are entrenched in the philosophical orientation of GAs as shortened by Julstrom (1999). Maintaining good diversity in the population is required to GA success. Otherwise, this leads to stuck off on local optima, which is an undesirable situation in GA, called premature convergence. The optimization literature acknowledging the diversity of population is a vital factor in search of a global optimal solution. This is evident by the discussion on the relevance of premature convergence and population diversity, see, for example, Burke (2004); Koumousis and Katsaras (2006); Pandey et al. (2014) and Aibinu et al. (2016). So it is clear from these studies that the performance of GA is mostly affected by the choice of selection operator. The selection operator is the most crucial research area in the body of contributions associated with GAs.

Due to the importance of the selection phase in GAs, the current research contributes to the literature by introducing a novel selection operator, namely the round-robin based tournament selection (RRTS). The main focus of this research is on facilitating the convergence process by maintaining a desirable level of population diversity. This objective is achieved by achieving a tradeoff between exploration and exploitation. The fitness rank of participants in concordance with the normality of generations is used to aid the selection process and the encouraging results of this delicate selection scheme are documented in this article.

The remainder part of this paper is organized as follows. In Section 2, the selection operator as a two-stage procedure, with a detailed review, has been discussed. In Section 3, we propose a new selection operator with its theoretical and mathematical foundations. Further, several stochastic properties of the newly proposed operator are reviewed in Section 4. Inspired by the stochastic features, Section 5 delineates the applicability of the proposed methodology in solving one of the practical problems, i.e. the traveling salesman problem (TSP). Lastly, Section 6 summarizes the study along with a brief discussion of future perspective research.

2 Selection procedure

The selection process in GA can be split into two stages. In the first stage, a selection probability is assigned to each and every individual based on fitness values. These probabilities are denoted as: $P = (p_1, ..., p_K)$, where $p_i \in [0,1]$ and $\sum_{i=1}^{K} p_i = 1$ with $K$ is the population size. To investigation about selection probabilities, someone can consult to Julstrom (1999) and Hussain and Muhammad (2020). The second stage is the sampling process, which selects the most fitted parents (based on Darwin’s “survival-of-the-fittest” criterion) from the current population for mating process. A thoroughly discussion about sampling algorithms has been provided in Section 4. This study, additionally, has a significant effect on the GA’s selection methods. In this perspective, a new operator for the selection is projected that is expected to reinforce the typical character of the population and offers an improved tradeoff between exploitation and exploration.

2.1 Assignment of probability

The first and the most popular selection procedure known as fitness proportional selection (FPS), was proposed by Holland (1975). In this selection procedure, the selection probability of $i^{th}$ individual, say $p_i$, is directly proportional to its fitness. The theme of this method depends upon the understanding that fitter individual ought to have a high probability of selection, whereas each individual to become the
member of the parent population using the following formula:

\[ p_i = \frac{f_i}{\sum_{j=1}^{K} f_j}; \quad i \in \{1, 2, ..., K\}, \quad (1) \]

where, \( f_i \) denotes the fitness status of \( i^{th} \) individual.

The operational directives of FPS are similar to the probability proportional to size (PPS) sampling using with replacement approach. Throughout the entire selection process, there will be no alternation needed in size and possibilities. This method is easy to enforce and offers probabilities to all individuals according to their fitness values, but the scaling problem is its main drawback, see, for instance, Grefenstette (1986).

The linear rank selection (LRS) was introduced by Baker (1985), is catered as the remedy of premature convergence attributed with FPS. This method provides a relatively better opportunity to pick out weaker individuals and therefore offered a smoother selection function. In the LRS procedure, the \( i^{th} \) individual is assigned a selection probability using the following formula:

\[ p_i = \frac{1}{K} (\vartheta^- + (\vartheta^+ - \vartheta^-) \frac{i - 1}{K - 1}); \quad i \in \{1, 2, ..., K\}, \quad (2) \]

where, \( i \) is the rank of the individual based on fitness status and \( \vartheta^- \) and \( \vartheta^+ \) are the parameters for the selection probabilities of worst and best individuals based on their ranks, respectively. The two constraints which are associated with this scheme as: \( \vartheta^+ + \vartheta^- = 2 \) and \( \vartheta^- \geq 0 \). As a result, even if individuals differ notably in fitness status, the ranks remain work in uniform pattern, unable to reflect the difference with desirable intensity and so naturally compromise relevant data.

Along with numerous applications of LRS, as an example, see, Sharma and Mehta (2013), but on the other hand, it has a drawback of slower convergence of the algorithm. This is because of its internal methodology based on ranks instead of fitness values directly for selection, see, Pandey et al. (2014); Aibinu et al. (2016) and Hussain and Muhammad (2020). This issue becomes more serious in the case of a larger population where ranks are thought about as a realization from the uniform distribution. To resolve the difficulty of LRS, Michalewicz (1992) designed an alternative rank based selection operator, called exponential rank selection (ERS). To tell apart from LRS, Michalewicz (1992) prompt that the selection probabilities increase exponentially from worst individual to best one. The selection venue of individual medaled with \( i^{th} \) rank, is mathematically written as:

\[ p_i = \nu^{K-i} \frac{1 - \nu}{1 - \nu^K}; \quad i \in \{1, 2, ..., K\}, \quad (3) \]

where, \( 0 < \nu < 1, \nu \) is a fixed ratio defined as weights of individuals based on their fitness ranks and maximum gain values of \( \nu \) closer to unity (i.e. \( \nu \to 1 \)) is recommended by Michalewicz (1992). ERS as a popular selection method is evidenced by various applications, see, for example, Schell and Wegenkittl (2001) and Lee et al. (2008).

Another selection procedure, which is based on a real phenomenon of the tournament, known as binary tournament selection (BTS), was introduced by Back (1996). Using BTS, two competitors are randomly chosen, then a winner will be selected for mating process. In this case, the chance of choosing a good parent is very high, but if both of the selected parents are of low quality, the low quality parents will be selected. Spotting the significance of population diversity, Back (1996) insisted lower tournament size because pair wise comparisons remain the most common theme in tournament selection schemes. The selection probability of \( i^{th} \) ordered individual is given as:

\[ p_i = \frac{1}{K^r} ((i)^r - (i - 1)^r); \quad i \in \{1, 2, ..., K\}, \quad (4) \]

where, \( r \) represents the array of tournament size.
Julstrom (1999) employed a probability-based threshold level to select the winner of the tournament called probabilistic 2-tournament selection (PTS). In this scheme, the competition winner will be survive with a probability \(0.5 < q < 1\), where the loser will get another chance of competing, with probability \(1 - q\). For the PTS method, the \(i^{th}\) ordered individual is assigned the selection probability by the following rule:

\[
p_i = \frac{2(i - 1)}{K(K - 1)} q + \frac{2(K - i)}{K(K - 1)} (1 - q); \quad i \in \{1, 2, ..., K\}.
\]

This selection procedure has great applicability, see, for instance, Schell and Wegenkittl (2001); Lee et al. (2008) and Hussain and Muhammad (2020).

The most significant feature of the selection operator is the selection pressure because it is adjacent with a suitable balance between exploration and exploitation. Eiben et al. (2006) described a scenario, where a relatively lower selection pressure is required at the initial stage for diversity in the whole sampled population and enhance at the last stage to assist the convergence of algorithm. To tradeoff between the two extremes, an adjustable selection pressure should be required, see, for example, Pham and Castellani (2010).

This article proposes a new selection approach, which removes the weakness related to fitness based (i.e. FPS), rank based (i.e. LRS and ERS) and tournament based (i.e. BTS and PTS) approaches. It is predicted on the basis of a tournament scheme, whereas we split all the individuals into two equal and mutually exclusive groups and assign them probabilities for selection according to their ranks. The details that how an individual is competing with other group’s members to be survived as parent for mating process are provided in the adjacent section.

3 The proposed selection operator

3.1 Motivation

There are many selection mechanisms that have been proposed in the literature. As LRS emphasized maintaining higher levels of population diversity at the cost of selection pressure and results in slowest convergence of GAs. On the other hand, the FPS method has high selection pressure with sacrificing the diversity and as a result remains the prime candidate of suffering from premature convergence. In this section, a new operator capable of achieving more balance between exploration and exploitation is proposed, which provides sufficient selection pressure throughout the selection process.

3.2 Round-robin based tournament selection

An alternative selection scheme (round-robin based tournament selection (RRTS)) is proposed to maintain a precise balance between exploration and exploitation. In this approach, an adequate selection pressure with elimination of the fitness scaling problem is provided. Consider the following steps for the proposed selection procedure:

(i) In the RRTS method, all individuals are ranked according to their fitness measures and acquire a distinct rank even though they have equivalent fitness values.

(ii) The individuals are divided into two equal and disjoint groups, e.g. \(A\) and its complement \(A^c\). The population can be inserted in these groups in multiple ways, such as: randomly, first half is in one group and rest is in other (best-worst), the odd individuals are in one group and even in the other (even-odd), up to 25% and 51% to 75% in one group and rest in other etc.

(iii) Now, one individual, i.e. \(i\), is chosen at random from a group with a surviving process probability \(\theta\), also comparing the combined effect of all the other group members with probability \((1 - \theta)\). Thus,
Table 1. Selection probabilities using various criteria in proposed method.

| Rank(i) | Group | Even-odd | Best-worst |
|---------|-------|----------|------------|
|         |       | \( p_i \) | \( p_i \) | \( p_i \) |
| 1       | A     | 0.0489   | 0.0444     | 0.0222     |
| 2       | \( A^c \) | 0.0622   | 0.0667     | 0.0333     |
| 3       | A     | 0.0733   | 0.0667     | 0.0444     |
| 4       | A     | 0.0822   | 0.0889     | 0.0556     |
| 5       | \( A^c \) | 0.0956   | 0.0889     | 0.0667     |
| 6       | A     | 0.1044   | 0.1111     | 0.1333     |
| 7       | A     | 0.1156   | 0.1111     | 0.1444     |
| 8       | \( A^c \) | 0.1289   | 0.1333     | 0.1556     |
| 9       | A     | 0.1378   | 0.1333     | 0.1667     |
| 10      | \( A^c \) | 0.1511   | 0.1556     | 0.1778     |

The selection probability to select an individual as a parent is determined by the subsequent rule:

\[
p_i = \frac{2(i - 1)}{K(K - 1)} \theta + \frac{4(1 - \theta)}{K^2(K - 1)} \sum_{j \in A^c} (K - j),
\]

where if \( i \) belongs to one group then \( j \) belongs to other group and \( K \) is the population size.

Table 1 presents some rules to assign probabilities to all the individuals for \( K = 10 \) and \( \theta = 0.5 \). There are two tuning parameters to maintain a tradeoff between diversity and selection pressure in our proposed method, i.e. the value of \( \theta \) and group segmentation.

4 The sampling algorithms

The first stage in the selection phase is to assign probabilities to all competing individuals, whereas in the second stage, a sampling algorithm is requisite to fill the mating pool for parents, whereas this process reflects the selection probabilities, such that the expected and observed number of individuals are equal. In this study, two popular sampling methods, roulette wheel sampling (RWS) and stochastic universal samplings (SUS) are used for testing.

4.1 Roulette wheel sampling

The RWS was introduced by Holland and it is still one of the most popular sampling methods for GA. In the RWS procedure, each possible solution is appointed as a slice with respect to a portion, which is assigned by a desired selection probability method. By using a single marker at the border of the roulette wheel and the roulette wheel is spun \( K \) times to successively select individuals. This sampling method is very simple and easier to implement with a high probability for the better choice of chromosomes. Clearly, the vector \((o_1, o_2, ..., o_K)\) follows a multinomial distribution with parameters \( K \) and \( P \), where \( P = (p_1, ..., p_K) \). The mean and variance of this distribution are given below:

\[
E[(o_1, o_2, ..., o_K)] = (e_1, e_2, ..., e_K),
\]

\[
V[o_i] = Kp_i(1 - p_i).
\]

4.2 Stochastic universal sampling (SUS)

The mechanism of SUS was introduced by Baker and is quite similar to RWS, however instead of the number of markers. In this method, \( K \) markers spaced evenly are used at the border of the roulette wheel. The slices of wheel are consistent as in RWS. In the SUS method, the roulette wheel is spun one time only and select all individuals, which are pointed by the \( K \) markers and enclosed in the mating pool as parents.
Therefore, all the parents are chosen in just one cycle of the wheel and this method promotes the better individuals for selecting at least once.

The computational complexity of SUS (i.e. $O(N)$) is comparatively lower than the complexity of RWS (i.e. $O(N^2)$), to identify the selected candidates as parents, only one pass over the population is needed. However, their expectations close with each other but the variabilities of the most fitted individuals are significantly less than RWS.

A detailed comparison between two sampling methods, i.e. RWS and SUS can be found in the central moments of the distributions of the vectors $(o_1, o_2, ..., o_K)$. The absolute difference between an individual’s observed and its expected values is defined as bias, i.e. $|o_i - e_i|$. In sampling, each individual might be provided a certain number of copies that are placed into the mating pool. The possible range of the number of copies is called “spread”. The SUS provides the insurance of minimum spread and almost zero bias.

### 4.3 The chi-square test as a goodness-of-fit measure

For empirical analysis, the chi-square test is a measure to ascertain the accuracy of sampling algorithms, i.e. RWS and SUS, and will compare with the probability distribution of the selection operators. As a tool for measuring the expected accuracy, the $\chi^2$ test was first introduced by Schell and Wegenkittl (2001).

Let we consider, $\xi_j = \sum_{i \in C_j} e_i$ is an overall expectation, whereas $O_j = \sum_{i \in C_j} o_i$ be the observed (actual) copies of individuals in mating pool after the sampling procedure. The two disjoint classes are: \{C_1, C_2, ..., C_c\}, $C_j \subset \{1, 2, ..., K\}$ and $\cup_{j=1}^c C_j = \{1, 2, ..., K\}$. For expected behavior, the $\xi_j$ be of the enjoin $K/c$ members with $1 \leq j \leq c$, to regulate that each class maintains the same number of individuals (on average). To desired stochastic accuracy, at least 10 individuals should be in each class. The chi-square test is defined as:

$$\chi := \sum_{j=1}^c \frac{(\xi_j - O_j)^2}{\xi_j}.$$

In the RWS algorithm, the $\xi_j \geq 10$ as it minimizes the differences between expected and observed frequencies. On the other hand, for SUS, we expect that $\chi \approx 0$. In Table 2, the probability distributions of all competing selection operators with the corresponding overall expected individuals (i.e. close to 300/10).

---

**Table 2. The overall expected counts, $\xi_j$, with respect to their classes, $C_j (j = 1, 2, ..., 10)$.

| j | LRS | $\xi_j$ | ERS | $\xi_j$ | BTS | $\xi_j$ |
|---|-----|--------|-----|--------|-----|--------|
| 1 | 1–33 | 30.05 | 1–108 | 30.33 | 1–95 | 30.08 |
| 2 | 34–65 | 29.84 | 109–158 | 29.91 | 96–134 | 29.77 |
| 3 | 66–96 | 29.56 | 159–237 | 30.22 | 135–164 | 29.80 |
| 4 | 97–127 | 30.20 | 238–253 | 29.22 | 214–233 | 29.73 |
| 5 | 128–157 | 29.84 | 254–267 | 29.72 | 253–269 | 30.72 |
| 6 | 158–187 | 30.44 | 280–290 | 29.86 | 270–285 | 29.55 |
| 7 | 188–216 | 30.00 | 291–300 | 30.16 | 286–300 | 29.25 |
| 8 | 217–244 | 29.50 | 21–22 | 30.43 | 181–205 | 29.96 |
| 9 | 245–272 | 30.02 | 206–227 | 29.57 | 186–206 | 29.46 |
| 10 | 273–300 | 30.54 | 228–247 | 30.49 | 227–246 | 30.77 |
| j | PTS | $\xi_j$ | SRS | $\xi_j$ | RRTS | $\xi_j$ |
|---|-----|--------|-----|--------|-----|--------|
| 1 | 1–52 | 30.43 | 1–87 | 30.42 | 1–79 | 29.99 |
| 2 | 53–93 | 30.37 | 88–123 | 30.18 | 80–130 | 30.45 |
| 3 | 94–128 | 30.38 | 124–151 | 30.33 | 131–162 | 29.57 |
| 4 | 129–159 | 30.33 | 152–180 | 29.89 | 163–185 | 30.58 |
| 5 | 160–287 | 30.15 | 181–205 | 29.96 | 186–206 | 29.46 |
| 6 | 188–213 | 30.35 | 206–227 | 29.57 | 207–226 | 29.43 |
| 7 | 214–237 | 30.02 | 228–247 | 29.49 | 227–246 | 30.77 |
| 8 | 238–259 | 29.21 | 248–266 | 30.32 | 247–265 | 30.47 |
| 9 | 260–280 | 29.39 | 267–284 | 30.79 | 266–283 | 29.98 |
| 10 | 281–300 | 29.36 | 285–300 | 29.06 | 284–300 | 29.31 |
Table 3. Simulated means and variances of the Chi-squared test statistics.

|       | LRS   | ERS   | RTS   | PTS   | SRS   | RRTS  |
|-------|-------|-------|-------|-------|-------|-------|
|       | RWS   | SUS   | RWS   | SUS   | RWS   | SUS   | RWS   | SUS   | RWS   | SUS   | RWS   | SUS   | RWS   | SUS   |
| \(\hat{\mu}\) | 8.68  | 0.0634 | 8.92  | 0.0542 | 8.71  | 0.0637 | 9.39  | 0.0647 | 9.35  | 0.0561 | 8.98  | 0.0492 |
| \(\hat{\sigma}^2\) | 18.31 | 0.0003 | 17.94 | 0.0008 | 15.93 | 0.0002 | 20.48 | 0.0001 | 19.99 | 0.0003 | 17.92 | 0.0001 |

are presented. \(\chi^S,R\) is the measure of chi-square for operator \(S\) that assigns the probabilities to individuals and for sampling algorithm \(R\).

The basic purpose of this test is to describe the sample mean and sample variance. The initial sampled population is to be considered as randomly. The probability distribution \(S\) is used to assign selection probabilities to all the individuals and then one of two sampling schemes, i.e. \(R\) is utilized to obtain instances of \(o_i\), \(O_j\) and \(\chi^S,R\), respectively. By the succession of \((\chi^S,R)_k\) for \(1 \leq k \leq s\), the sample mean and variance can be computed as:

\[
\hat{\mu}^{(S,R)} = \frac{1}{s} \sum_{k=1}^{s} \chi^S,R_k,
\]

(8)

\[
\hat{\sigma}^2(S,R) = \frac{1}{s-1} \sum_{k=1}^{s} (\chi^S,R_k - \hat{e}(S,R))^2.
\]

(9)

For 99% confidence interval, it is compared with theoretical \(\chi^2_{c-1}\) distribution. The sample mean and variance of chi-square should be close to \(c-1 = 9\), \(2(c-1) = 18\), respectively, and their estimates of \(\hat{\mu}\) and \(\hat{\sigma}^2\) are provided in Table 3. The SUS results are also reported in this table, which are shown its sampling accuracy as well. The average accuracy of the sampling method with all competing selection schemes is observed from these empirical results.

4.4 Empirical distribution function analysis

In this section, the empirical distribution function (EDF) is compared with theoretical chi-square distribution \(\chi^2_{c-1}\) of roulette wheel sampling and it is given as:

\[
EDF^{S,R}_{(t)} := \frac{1}{s} \#\{1 \leq k \leq s : \chi^S,R_k \leq t\}, \quad t \in [0, \infty).
\]

(10)

In Figure 1, the behaviors of EDF (dashed line), for various selection operators for a population size \(K = 300\) with a similar number of tests are reported. The selection operators are being compared with the theoretical \(\chi^2_{c-1}\) distribution (dark thick line) using a 99% confidence band under the hypothesis of RWS (dashed thin double line). Here, the range on the x-axis values is \(t \in [0, 18]\), where we expect the value of \(\chi^S,R_k\), i.e. \(E[\chi^S,R_k] = c-1 = 9\). The RWS provides the empirical distribution function that is insignificant from the theoretical \(\chi^2_{c-1}\) distribution by \(\hat{e}\) and \(\hat{\sigma}^2\) statistics.

5 Global performance

5.1 The traveling salesman problem

The most illustrious benchmark, noteworthy and historic hard combinatorial optimization problem is the traveling salesman problem (TSP). In this problem, someone wants to find out the shortest Hamiltonian tour to starts his/her tour from a city and go to all other cities once and come back to the initial city. The first one, who documented this problem was Euler in 1759, see, for example, Larranaga et al. (1999). It is the most fundamental problem and has many applications in engineering, discrete mathematics, operations research, graph theory and computer science, etc.
Let $n$ cities with a distance (cost) matrix, $C = [c_{ij}]_{n \times n}$ is searched for a permutation $\lambda : \{0, ..., n-1\} \rightarrow \{0, ..., n-1\}$, where $c_{ij}$ is the distance between city $i$ and city $j$ and it minimizes the traveled distance $f(\lambda, C)$ as follows:

$$f(\lambda, C) = \sum_{i=0}^{n-1} d(c_{\lambda(i)}, c_{\lambda(i+1)}) + d(c_{\lambda(n)}, c_{\lambda(1)}), \quad (11)$$

where $\lambda(i)$ is the location of city $i$ in each tour, $d(c_i, c_j)$ is the distance from a city $i$ to another city $j$, whereas $(x_i, x_j)$ be a specified position of each city in a tour in the plane, and the Euclidean distances of
Table 4. The benchmark problems

| Problem name | No. of cities | Optimal tour length |
|--------------|---------------|---------------------|
| BERLIN52     | 52            | 7542                |
| PR144        | 144           | 58537               |
| KROB150      | 150           | 26130               |
| RAT195       | 195           | 2323                |
| KROA200      | 200           | 29368               |

Table 5. Parametric configuration for GA

| Parameter                | Setting                           |
|--------------------------|-----------------------------------|
| Representation           | Permutation                       |
| Population size          | 150                               |
| Crossover criteria       | OX                                |
| Crossover rate           | 80%                               |
| Mutation method          | EM                                |
| Mutation rate            | 5%                                |
| Maximum generation       | 5000                              |
| Number of trails         | 30                                |
| Replacement in GA        | Steady-state GA                   |

The distance matrix $C$ between the city $i$ and city $j$ is extracted in the following way:

$$ c_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. $$

(12)

TSP is easy to understand but very difficult to solve, e.g. with ‘100’ cities, there are $10^{155}$ possibilities to find the tour. This is the main reason to declare it as a non-deterministic polynomial (NP-hard) problem, see, for example, Helsgaun (2000) and Applegate et al. (2006). Hence, this type of problem is not possible to solve using traditional optimization algorithms, e.g. gradient-based methods. To attain the optimal or close to optimal solution within an adequate amount of time, the heuristic algorithms are better choices to manage the NP-hard problems (Huang et al., 2015; Ruiz et al., 2015; Hussain and Muhammad, 2020). The GA has also been applied for the solution of this problem in different ways, see, for example, Potvin (1996); Larrañaga et al. (1999); Moon et al. (2002); Nagata and Soler (2012) and Hussain et al. (2017). Some test problems are taken from the library of traveling salesman problem (TSPLIB) for the global performance of the newly devised selection operator with respect to existing ones and reported in Table 4.

5.2 The state-of-the-art settings

In the simulation study, all GA programs were coded in MATLAB software. Moreover, the two stopping criteria, i.e. not improvement found in 300 successive generations and the maximum number of generations (i.e. 5000) are used. The order crossover (OX) along with a well-known exchange mutation (EM) operator are used in this study. Table 5 is provided for further information about desired parameters.

5.3 Simulation results and discussion

In the above sections, the relative characteristics of the proposed operator with its competing selection methods with respect to sampling accuracy and population diversity have been determined. In this section, we test the performance of RRTS with other schemes by applying it to TSP. The results of six competing selection schemes with the most popular genetic operators, i.e. order crossover (OX) and exchange mutation (EM) are provided in Table 6, where all the tests are repeated thirty times. On the
Table 6. Results of various selection methods with respect to OX (crossover) and EM (mutation) operators.

| Selection Method | Statistics | BERLIN52 | PR144 | KROB150 | RAT195 | KROA200 |
|------------------|------------|----------|-------|---------|--------|---------|
| FPS              | Average    | 7978     | 61931 | 28378   | 2481   | 31134   |
|                  | S.D        | 297      | 1765  | 999     | 101    | 958     |
|                  | R.E        | 5.78     | 5.80  | 8.60    | 6.80   | 6.01    |
| LRS              | Average    | 8056     | 61643 | 27813   | 2412   | 30480   |
|                  | S.D        | 371      | 1531  | 916     | 109    | 917     |
|                  | R.E        | 6.82     | 5.31  | 6.44    | 3.83   | 3.79    |
| ERS              | Average    | 8437     | 63751 | 29558   | 2532   | 31878   |
|                  | S.D        | 452      | 1734  | 874     | 134    | 758     |
|                  | R.E        | 11.87    | 8.91  | 13.12   | 9.00   | 8.55    |
| BTS              | Average    | 7954     | 60666 | 27576   | 2431   | 29968   |
|                  | S.D        | 328      | 1254  | 941     | 101    | 858     |
|                  | R.E        | 5.46     | 3.64  | 5.53    | 4.65   | 2.04    |
| PTS              | Average    | 8077     | 60831 | 28165   | 2443   | 30825   |
|                  | S.D        | 432      | 989   | 918     | 117    | 869     |
|                  | R.E        | 7.09     | 3.92  | 7.79    | 5.17   | 4.96    |
| RRTS             | Average    | 7963     | 60632 | 27569   | 2404   | 29926   |
|                  | S.D        | 301      | 1024  | 875     | 103    | 831     |
|                  | R.E        | 5.58     | 3.58  | 5.71    | 3.49   | 1.90    |

basis of average, standard deviation (S.D) and relative efficiency (R.E), these computational results are compared. Since TSP is a minimization problem, we observed an improved performance, based on 5000 simulations, by the proposed operator from among all six competing selection operators. From these results, we can confirm that RRTS outperforms the others.

6 Conclusions

For every optimization algorithm, the main desire is to balance between two extremes, i.e. exploration and exploitation. This article presents a new round-robin based tournament selection operator for GAs, which is suggested a fine balance between exploitation and exploration. The individuals are sorted with respect to their fitness measures and then the whole population is divided into two equal and non-overlapping groups, i.e. $A$ and $A^c$. To determine the sampling accuracy, we employ $\chi^2$ test to confirm a close match between the expected and observed number of offspring (insignificant difference). A simulation study is performed to evaluate the performance of the newly devised selection operator along with some conventional operators. Based on this research, we suggest that the proposed operator might be used as a better alternative to get global optima or near to optimum results. Moreover, researchers might be apply it for any problems related to evolutionary algorithms.

Disclosure statement

No conflict of interest is declared by author(s).

References

Aibinu, A.M., Salau, H.B., Rahman, N.A., Nwohu, M.N., Akachukwu, C. (2016). A novel clustering based genetic algorithm for route optimization. *Engineering Science and Technology, an International Journal*, 19(4):2022-2034.
Amsa, M.G.B.A., Aibinu, A.M. and Salami, M.J.E. (2013). A novel hybrid artificial intelligence technique for colpitts oscillator design. *Journal of Control, Automation and Electrical Systems*, 25(1):10-21.

Applegate, D.L., Bixby, R.E., Chvatal, V., Cook, W.J. (2006). The Traveling Salesman Problem: A Computational Study. Princeton University Press.

Back, T. (1996). Evolutionary Algorithms in Theory and practice. Oxford Press.

Baker, J.E. (1985). Adaptive selection methods for genetic algorithms. In Proceedings of an International Conference on Genetic Algorithms and their applications, pages 101-111. Hillsdale, New Jersey.

Beyer, H.G. (1997). An alternative explanation for the manner in which genetic algorithms operate. *Biosystems*, 41(1):1-15.

Burke, K., Gustafson, S. and Kendall, G. (2004). Diversity in genetic programming: An analysis of measures and correlation with fitness. *IEEE Transactions on Evolutionary Computation*, 8(1):47-62.

Eiben A.E, Schut M.C, de-Wilde A.R (2006). Is self-adaptation of selection pressure and population size possible? A case study. In: Parallel problem solving from nature-PPSN IX, pp 900–909.

Fogel, D.B. (1994). An introduction to simulated evolutionary optimization. *IEEE Transactions on Neural Networks*, 5(1):3-14.

Forrest, S., Javornik, B., Smith, R.E. and Perelson, A.S. (1993). Using genetic algorithms to explore pattern recognition in the immune system. *Evolutionary Computation*, 1(3):191-211.

Grefenstette, J.J. (1986). Optimization of control parameters for genetic algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, 16(1):122-128.

Helsgaun, K. (2000). An effective implementation of the Lin-Kernighan traveling salesman heuristic. *European Journal of Operational Research*, 126(1):106-130.

Huang, H.X., Li, J.C. and Xiao, C.L. (2015). A proposed iteration optimization approach integrating backpropagation neural network with genetic algorithm. *Expert Systems with Applications*, 42(1):146-155.

Hussain, A., Muhammad, Y.S., Sajid, M.N., Hussain, I., Shoukry, M.A. and Gani, S. (2017). Genetic algorithm for traveling salesman problem with modified cycle crossover operator. *Computational Intelligence and Neuroscience*, 2017:1-7.

Hussain, A., and Muhammad, Y.S. (2020). Trade-off between exploration and exploitation with genetic algorithm using a novel selection operator. *Complex & Intelligent Systems*, 6(1):1-14.

Holland, J.H. (1975). Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. University of Michigan Press, Oxford, UK.

Julstrom, B.A. (1999). It’s all the same to me: Revisiting rank-based probabilities and tournaments. In Proceedings of the congress on evolutionary computation, pp 1501–1505, Volume 2, IEEE.

Koumousis, V.K. and Katsaras, C.P. (2006). A saw-tooth genetic algorithm combining the effects of variable population size and reinitialization to enhance performance. *IEEE Transactions on Evolutionary Computation*, 10(1):19-28.

Larranaga, P., Kuijpers, C.M., Murga, R.H., Inza, I., Dizdarevic, S. (1999). Genetic algorithms for the traveling salesman problem: A review of representations and operators. *Artificial Intelligence Review*, 13:129-170.
Lee, S., Soak, S., Kim, K., Park, H. and Jeon, M. (2007). Statistical properties analysis of real world tournament selection in genetic algorithms. *Applied Intelligence*, 28(2):195-205.

Michalewicz, Z. (1992). Genetic algorithms + Data structures = Evolution programs. Artificial Intelligence Book Series. Berlin: Springer.

Moon, C., Kim, J., Choi, G., Seo, Y. (2002). An efficient genetic algorithm for the traveling salesman problem with precedence constraints. *European Journal of Operational Research*, 140(3):606-617.

Nagata, Y., Soler, D. (2012). A new genetic algorithm for the asymmetric traveling salesman problem. *Expert Systems with Applications*, 39(10):8947-8953.

Pandey, H.M., Chaudhary, A. and Mehrotra, D. (2014). A comparative review of approaches to prevent premature convergence in GA. *Applied Soft Computing*, 24:1047-1077.

Pham, D.T., Castellani, M. (2010). Adaptive selection routine for evolutionary algorithms. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 224(6):623-633.

Potvin, J.Y. (1996). Genetic algorithms for the traveling salesman problem. *Annals of Operations Research*, 63(3):337-370.

Rubio, F., Llopis-Albert, C., Valero, F. and Suner, J.L. (2016). Industrial robot efficient trajectory generation without collision through the evolution of the optimal trajectory. *Robotics and Autonomous Systems*, 86:106-112.

Ruiz, E., Albareda-Sambola, M., Fernandez, E., Resende, M.G. (2015). A biased random-key genetic algorithm for the capacitated minimum spanning tree problem. *Computers & Operations Research*, 57:95-108.

Ruxton, G.D. and Beauchamp, G. (2008). The application of genetic algorithms in behavioural ecology, illustrated with a model of anti-predator vigilance. *Journal of Theoretical Biology*, 250(3):435-448.

Schell, T. and Wegenkittl, S. (2001). Looking beyond selection probabilities: adaptation of the $\chi^2$ measure for the performance analysis selection methods in GAs. *Evolutionary Computation*, 9(2):243–256.

Sharma, A. and Mehta, A. (2013). Review paper of various selection methods in genetic algorithm. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(7):1476-1479.

Song, J., Xue, G. and Kang, Y. (2016). A novel method for optimum global positioning system satellite selection based on a modified genetic algorithm. *Plos One*, 11(3):e0150005.

Soui, M., Gasmi, I., Smiti, S. and Ghedira, K. (2019). Rule-based credit risk assessment model using multi-objective evolutionary algorithms. *Expert Systems with Applications*, 126:144-157.