Porównanie klasycznej ewolucji różnicowej DE/rand/1 z odmianą wykorzystująca punkt środkowy DE/mid/1

Adam Stelmaszczyk

Streszczenie

W pracy porównano klasyczny algorytm ewolucji różnicowej z jego odmianą DE/mid/1. W części teoretycznej wyznaczono współczynnik skalujący dla DE/mid/1. W części praktycznej przeprowadzono szereg eksperymentów na 7 wybranych funkcjach z BBOB 2013[1]. DE/mid/1 okazał się lepszy na 6 z nich.

1 Część teoretyczna

Klasyczna ewolucja różnicowa DE/rand/1 oraz jej odmiana DE/mid/1 różnią się jedynie operatorem mutacji. W DE/rand/1, mutant i-tego osobnika w populacji P powstaje w następujący sposób:

$$u_i = P_j + F(P_k - P_l), i \neq j \neq k \neq l \tag{1}$$

W DE/mid/1:

$$u'_i = m + a(P_k - P_l), i \neq k \neq l \tag{2}$$

m to punkt środkowy populacji:

$$m = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{3}$$

$a \in \mathbb{R}$ jest parametrem skalującym analogicznym do F. Żeby macierz kowariancji populacji w DE/mid/1 była taka sama jak w DE/rand/1, macierz

[1] http://coco.gforge.inria.fr/doku.php?id=bbob-2013-downloads

kowariancji mutanta w DE/mid/1 musi być taka sama jak w DE/rand/1. Można to osiągnąć tak dobierając a, żeby było spełnione równanie:

$$\operatorname{cov}(u_i) = \operatorname{cov}(u_i') \tag{4}$$

Zakładając, że osobniki są liniowo niezależne od siebie:

$$\begin{aligned}\operatorname{cov}(u_i) &\overset{(1)}{=} \operatorname{cov}(P_j + F(P_k - P_j)) = \operatorname{cov}(P_j) + \operatorname{cov}(F(P_k - P_j)) \\ &= \operatorname{cov}(P_j) + F^2\operatorname{cov}(P_k - P_j) = \operatorname{cov}(P_j) + F^2(\operatorname{cov}(P_k) + cov(P_j))\end{aligned}$$

$\forall i \operatorname{cov}(P_i) = V$, ponieważ każdy osobnik ma taki sam rozkład prawdopodobieństwa:

$$\operatorname{cov}(u_i) = \operatorname{cov}(P_j) + F^2(\operatorname{cov}(P_k) + cov(P_j)) = V + F^2(V + V) = V(1 + 2F^2)$$

Rozwijając prawą stronę równania (4):

$$\begin{aligned}\operatorname{cov}(u_i') &\overset{(2)}{=} \operatorname{cov}(m + a(P_k - P_l)) = \operatorname{cov}(m) + \operatorname{cov}(a(P_k - P_l)) \\ &\overset{(2)}{=} \operatorname{cov}(\frac{1}{n}\sum_{i=1}^{n} P_i) + a^2(\operatorname{cov}(P_k) + \operatorname{cov}(P_j)) = \frac{1}{n^2}\operatorname{cov}(\sum_{i=1}^{n} P_i) + a^2(V + V) \\ &= \frac{1}{n^2}(nV) + 2a^2V = (\frac{1}{n} + 2a^2)V\end{aligned}$$

Zatem $\operatorname{cov}(u_i) = (1 + 2F^2)V$, $\operatorname{cov}(u_i') = (\frac{1}{n} + 2a^2)V$. Podstawiając do (4):

$$\begin{aligned}(1 + 2F^2)V &= (\frac{1}{n} + 2a^2)V \\ 1 + 2F^2 &= \frac{1}{n} + 2a^2 \\ 2a^2 &= 1 + 2F^2 - \frac{1}{n} \\ a^2 &= \frac{1}{2} + F^2 - \frac{1}{2n}\end{aligned}$$

Obie strony są nieujemne, więc:

$$a = \sqrt{\frac{1}{2} + F^2 - \frac{1}{2n}} \tag{5}$$

$$\lim_{n\to\infty} \sqrt{\frac{1}{2} + F^2 - \frac{1}{2n}} = \sqrt{\frac{1}{2} + F^2}$$

$a \approx 1.142$ dla $F = 0.9$ i $n = 100$. Opuszczając człon $-\frac{1}{2n}$, $a \approx 1.145$.

a jest większe od F dla $n > 1$. Dowód:

$$\begin{aligned} a &> F \\ \sqrt{\frac{1}{2} + F^2 - \frac{1}{2n}} &> F \\ \frac{1}{2} + F^2 - \frac{1}{2n} &> F^2 \\ \frac{1}{2} &> \frac{1}{2n} \\ 1 &> \frac{1}{n} \\ n &> 1 \end{aligned}$$

W DE/mid/1, zamiast przesuwać losowo wybranego osobnika P_j, przesuwamy punkt środkowy. Punkt środkowy jest mniej zmienny, tzn. norma macierzy kowariancji punktu środkowego jest mniejsza niż norma macierzy kowariancji dowolnego osobnika. $cov(m) = \frac{1}{n}V$, natomiast $cov(P_i) = V$. Dlatego na ogół DE/mid/1 potrzebuje większego współczynnika skalującego niż DE/rand/1.

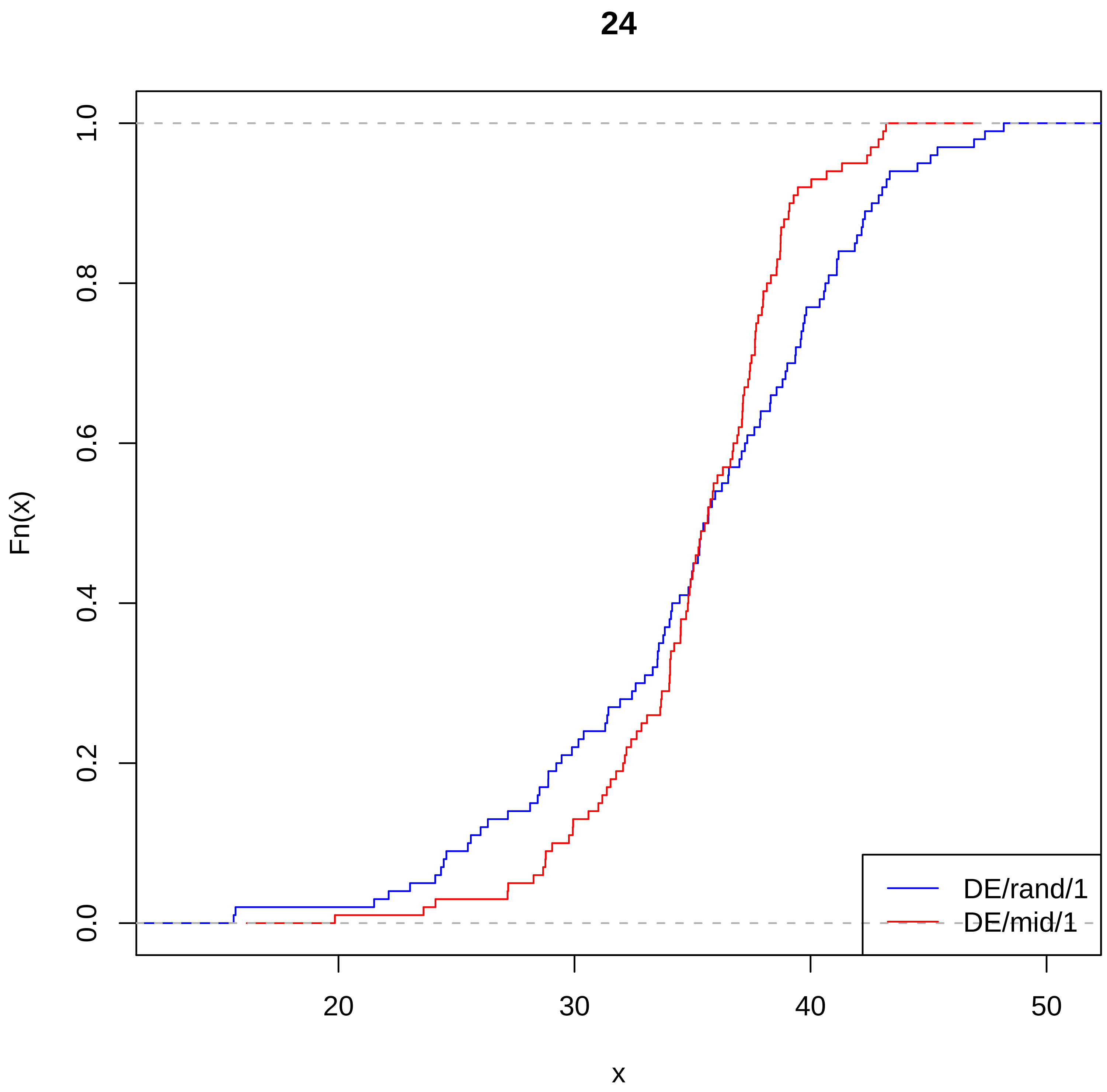
2 Część praktyczna

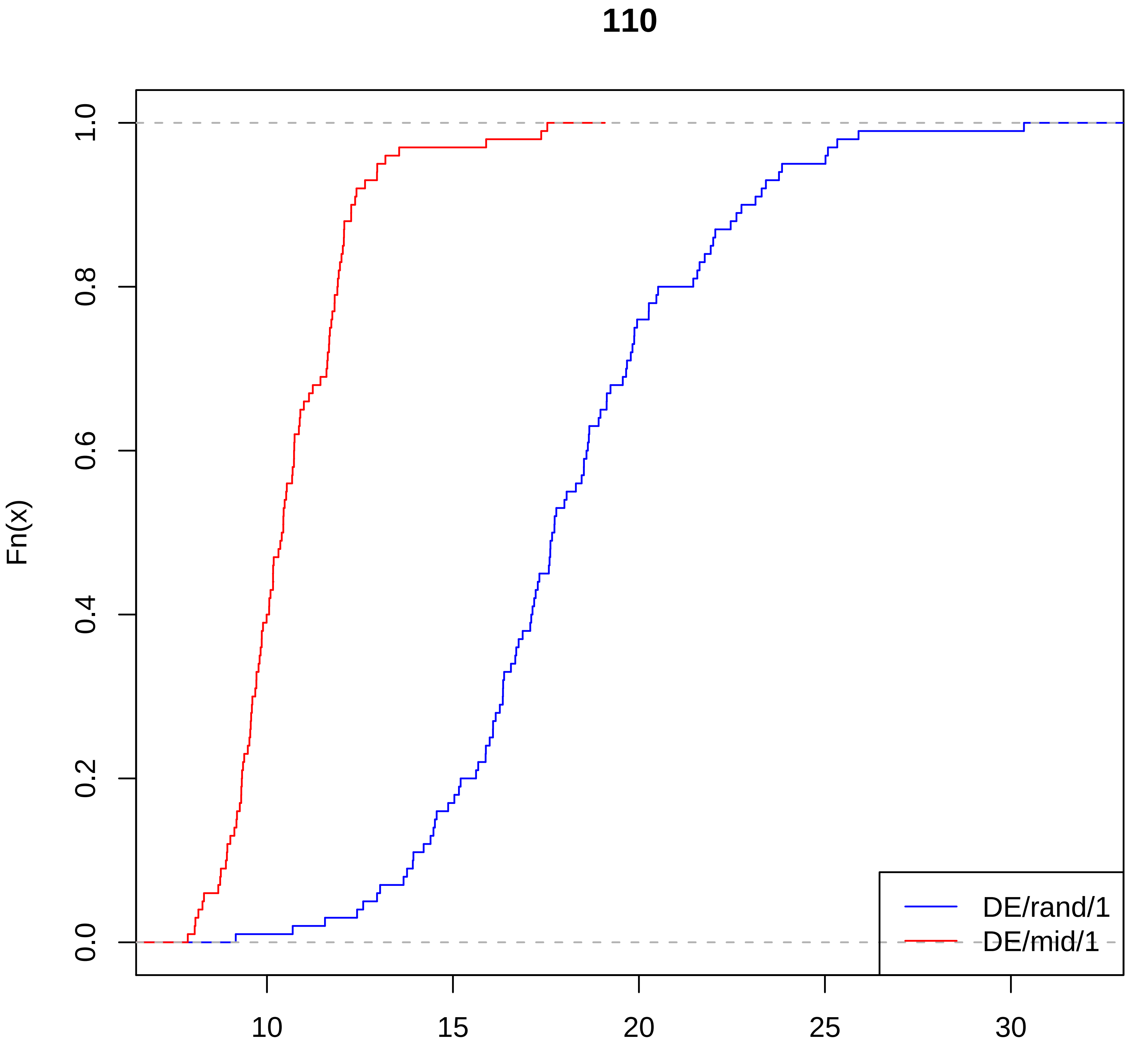
Ekperymenty przeprowadzono na 7 funkcjach testowych o numerach 24, 110, 113, 116, 119, 122, 126 z BBOB 2013, zaimplementowanych w C. Funkcje testowe są wywoływane z Javy, w której napisano algorytmy oraz procedurę testującą. Liczba wymiarów $D = 10$. Maksymalna liczba wywołań funkcji oceny $FEs = 10^6$. Rozmiar populacji $n = 100$. Jeśli algorytm nie znajdował minimum, wówczas w jednym podejściu, na jednej funkcji, generował 10^4 pokoleń. Najlepszy wynik z jednego podejścia był zapisywany do pliku. Do każdej funkcji algorytm miał 100 podejść, z każdego podejścia zapisywany był najlepszy wynik. W celu porównania, wykreślano dystrybuanty empiryczne najlepszych wyników z każdego podejścia dla obu algorytmów na jednej funkcji. Wykresy przedstawiono poniżej. Najlepszym wynikiem była najmniejsza odległość funkcji oceny osobnika od minimum. Współczynnik skalujący $F = 0.9$ dla DE/rand/1. Dla DE/mid/1 współczynnik $a \approx 1.142$, zgodnie ze wzorem (5). Prawdopodobieństwo krzyżowania $Cr = 0.9$ dla obu algorytmów.

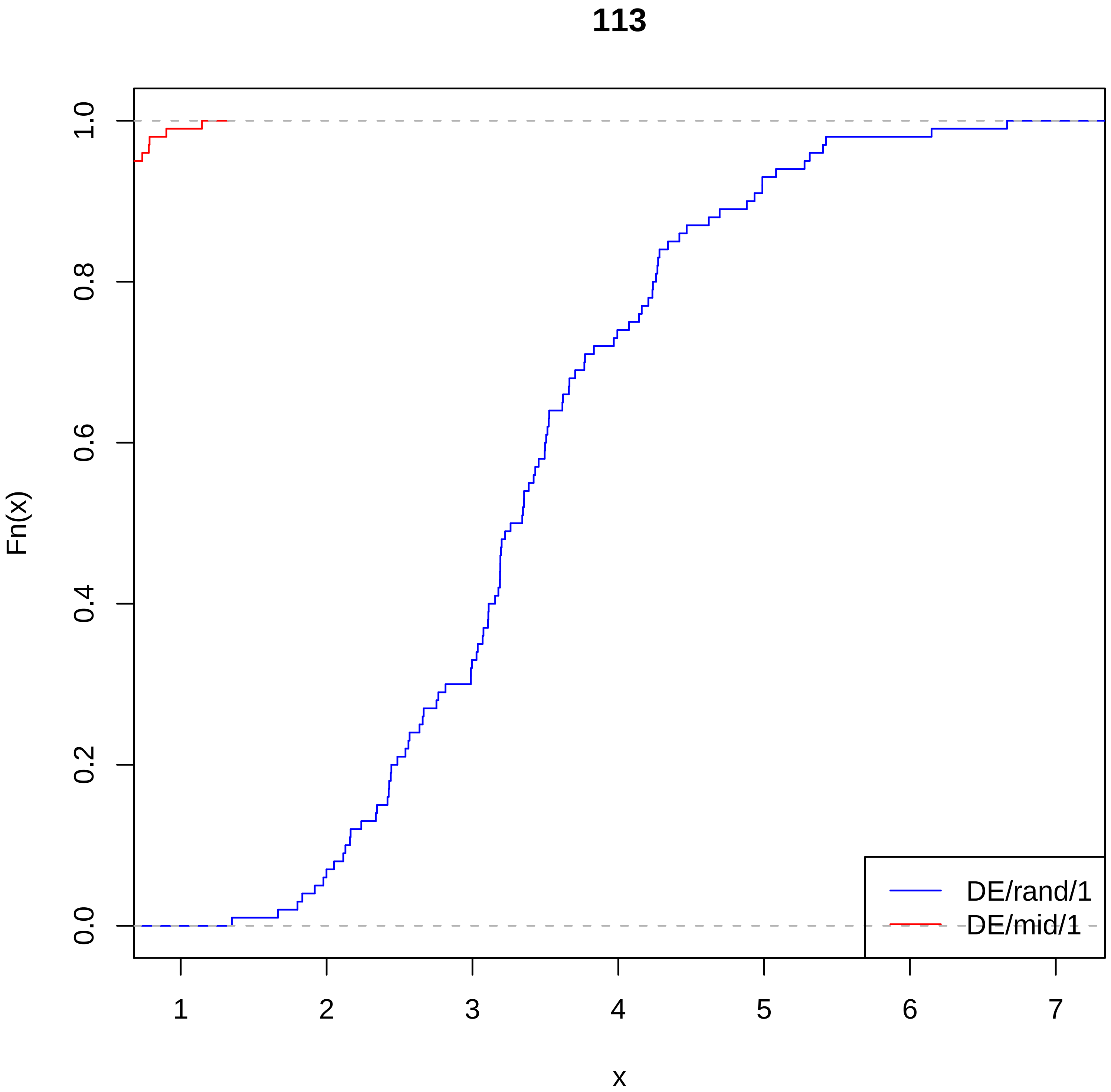
Numer funkcji	24	110	113	116	119	122	126
Wynik porównania	r	+	+	+	+	+	+

Tabela 1: Porównanie DE/mid/1 z DE/rand/1, r = remis, + = wygrana

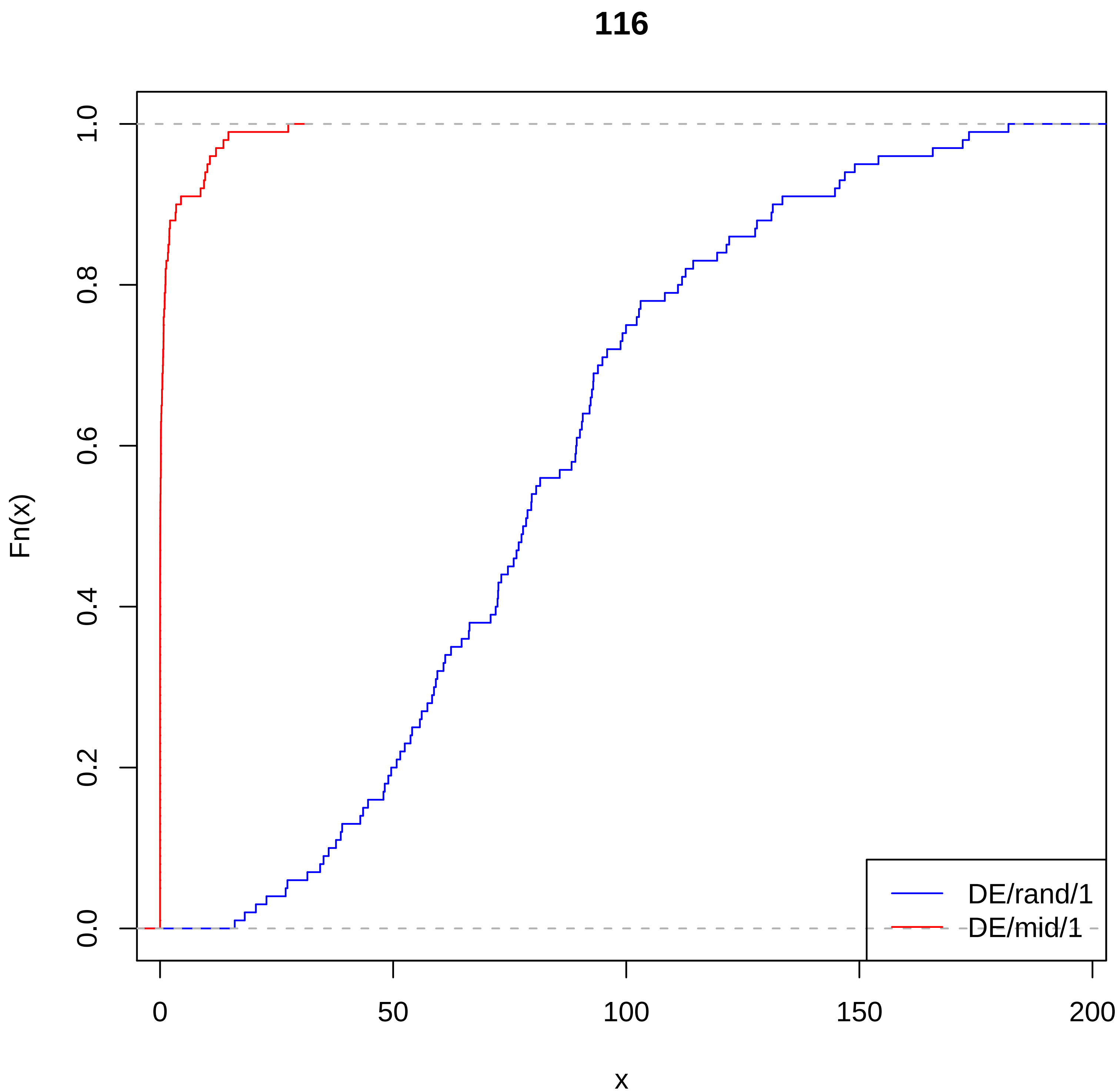
Porównanie algorytmów przedstawia tabela 1. + oznacza znaczną przewagę DE/mid/1, r oznacza remis. DE/mid/1 był lepszy na 6 spośród 7 funkcji. Na 113, 116, 119, 122 i 126 DE/mid/1 znalazł minimum, podczas gdy DE/rand/1 nigdy się to nie udało.

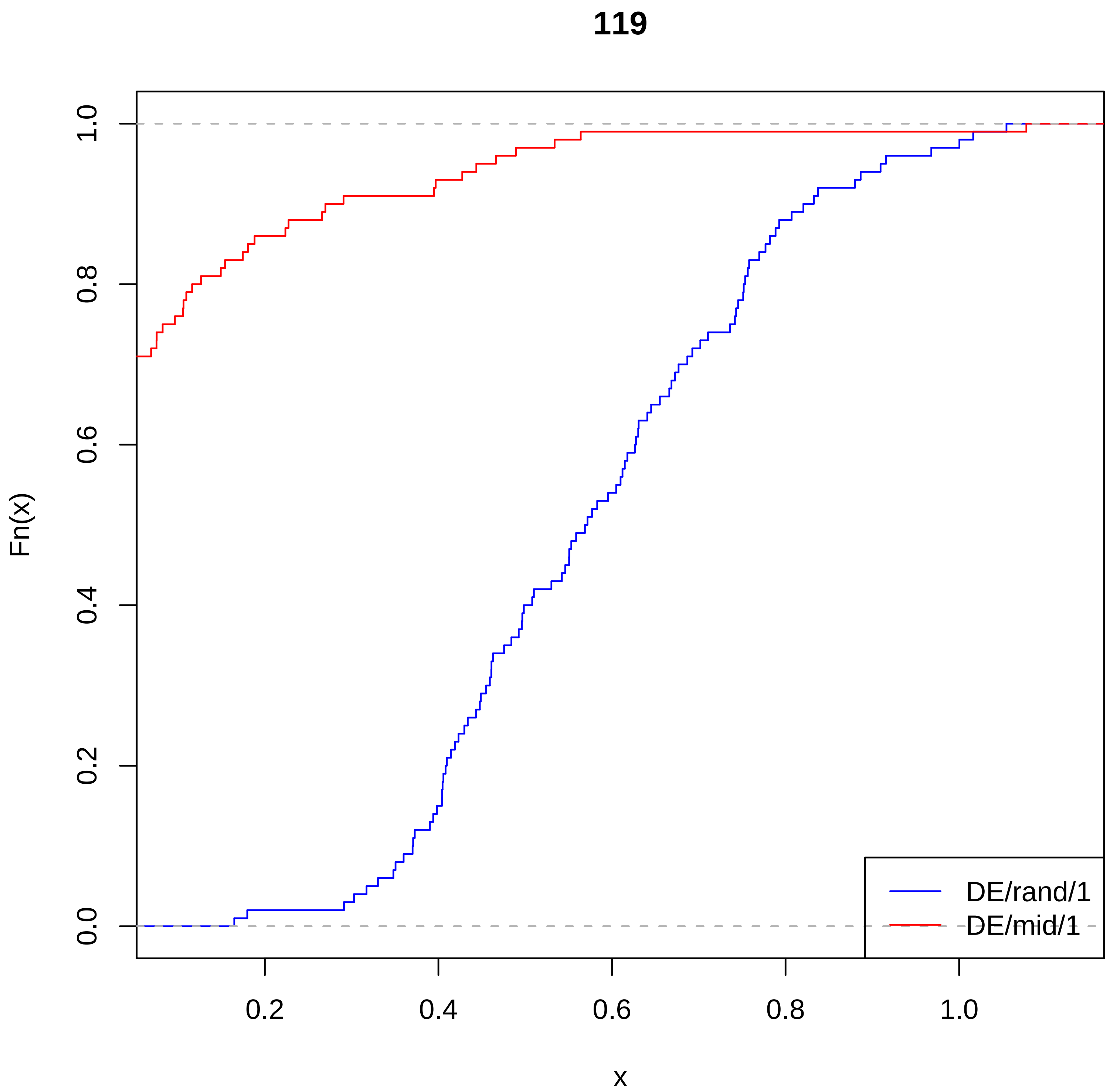

24
Fn(x)
1.0
0.8
0.6
0.4
0.2
0.0
20
30
40
50
x
DE/rand/1
DE/mid/1

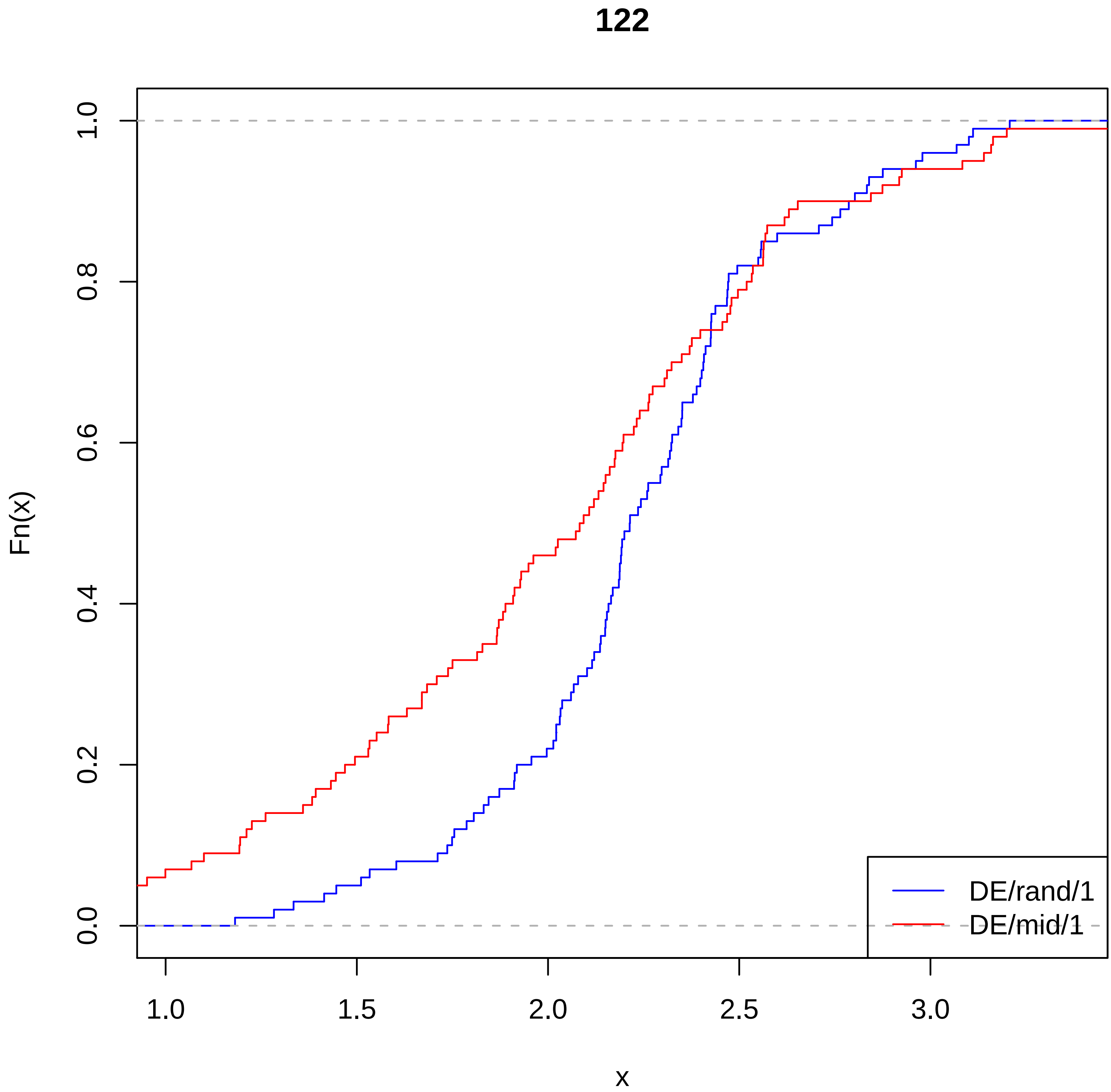

110
Fn(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
10
15
20
25
30
DE/rand/1
DE/mid/1



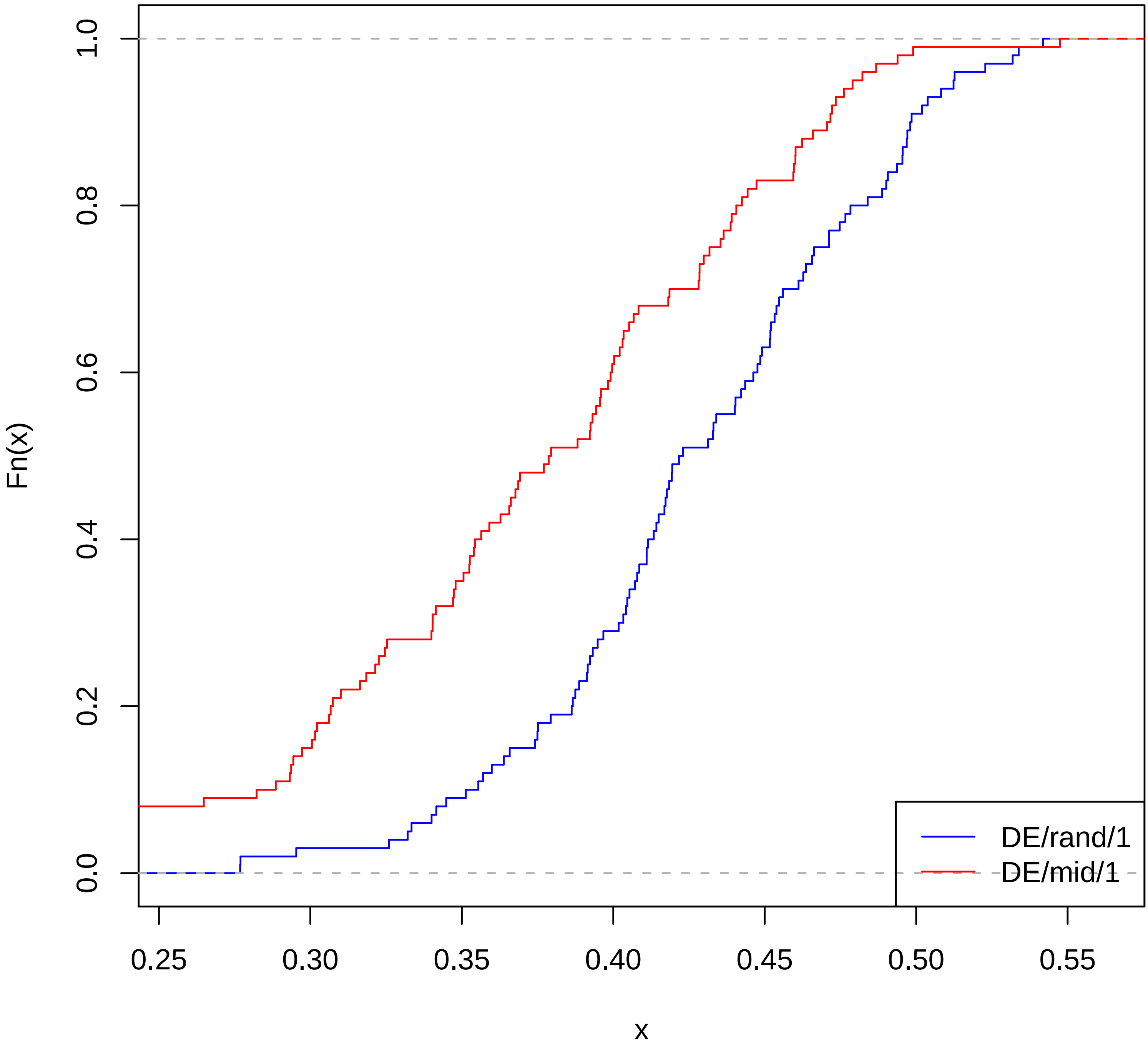
113
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
x
DE/rand/1
DE/mid/1


116
Fn(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
DE/rand/1
DE/mid/1


119
1.0
0.8
0.6
0.4
0.2
0.0
Fn(x)
0.2
0.4
0.6
0.8
1.0
x
DE/rand/1
DE/mid/1


122
Fn(x)
x
DE/rand/1
DE/mid/1
0.0
0.2
0.4
0.6
0.8
1.0
1.0
1.5
2.0
2.5
3.0

126



1.0
0.8
0.6
0.4
0.2
0.0
Fn(x)
0.25
0.30
0.35
0.40
0.45
0.50
0.55
x
DE/rand/1
DE/mid/1