CEFET-MG

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

ANÁLISE E SÍNTESE DE REGRAS DE ADAPTAÇÃO EM ESTRATÉGIAS EVOLUTIVAS USANDO FUNÇÕES DE LYAPUNOV ESTOCÁSTICAS

CLÁUDIA RAQUEL MARTINS CORRÊA

Orientadora: Elizabeth Fialho Wanner
Departamento de Engenharia de Computação, CEFET-MG

Coorientador: Carlos Manuel Mira da Fonseca
CISUC - Departamento de Engenharia Informática, Universidade de Coimbra - Portugal

BELO HORIZONTE
FEVEREIRO DE 2019

Cláudia Raquel Martins Corrêa

Análise e Síntese de Regras de Adaptação em Estratégias Evolutivas usando Funções de Lyapunov Estocásticas

Tese apresentada ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Doutor em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional

Linha de pesquisa: Sistemas Inteligentes

Orientadora: Elizabeth Fialho Wanner
Departamento de Engenharia de Computação, CEFET-MG

Coorientador: Carlos Manuel Mira da Fonseca
CISUC - Departamento de Engenharia Informática, Universidade de Coimbra - Portugal

Belo Horizonte

Fevereiro de 2019

Corrêa, Cláudia Raquel Martins
C824a Análise e síntese de regras de adaptação em estratégias evolutivas usando funções de Lyapunov estocásticas / Cláudia Raquel Martins Corrêa. – 2019.
193 f.

Tese de doutorado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional.
Orientadora: Elizabeth Fialho Wanner.
Coorientador: Carlos Manuel Mira da Fonseca.
Tese (doutorado) – Centro Federal de Educação Tecnológica de Minas Gerais.

1. Algorítmo genético – Teses. 2. Computação evolutiva – Teses. 3. Funções Lyapunov – Teses. 4. Convergência – Teses. I. Wanner, Elizabeth Fialho. II. Fonseca, Carlos Manuel Mira da. III. Centro Federal de Educação Tecnológica de Minas Gerais. IV. Título.

CDD 519.6

Elaboração da ficha catalográfica pela Biblioteca-Campus II / CEFET-MG

SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
COORDENAÇÃO DO PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

ANÁLISE E SÍNTESE DE REGRAS DE ADAPTAÇÃO EM ESTRATÉGIAS EVOLUTIVAS USANDO FUNÇÕES DE LYAPUNOV ESTOCÁSTICAS.

Tese de Doutorado apresentada por **Cláudia Raquel Martins Corrêa**, em 28 de fevereiro de 2019, ao Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG, e aprovada pela banca examinadora constituída pelos professores:

Profª. Drª. Elizabeth Fialho Wanner
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Carlos Manoel Mira da Fonseca
Universidade de Coimbra

Prof. Dr. Pedro Luis Dias Peres
Universidade Estadual de Campinas

Prof. Dr. Ricardo Hiroshi Caldeira Takahashi
Universidade Federal de Minas Gerais

Prof. Dr. Sérgio Ricardo de Souza
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Rodrigo Tomás Nogueira Cardoso
Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida à impressão,

Prof. Dr. Thiago de Souza Rodrigues
Coordenador do Programa de Pós-Graduação em
Modelagem Matemática e Computacional

À minha amada sobrinha Flora (in memorian),
um anjo de luz, presente de Deus!

Agradecimentos

Agradeço a Deus acima de tudo. Agradeço aos meus pais, pelo amor, pelo apoio e por me proporcionarem condições para conclusão das etapas básicas e primordiais dos meus estudos. Agradeço ao meu marido pelo amor, pela renúncia e pela paciência nesses últimos anos, período em que não consegui lhe dedicar atenção e que a distância física se fez necessária. Agradeço também aos meus irmãos pelo incentivo.

Agradeço à minha orientadora, Professora Elizabeth, por ter aceito esta tarefa e por tê-la executado com competência, dedicação e humanidade, também por sua paciência e amizade. Agradeço ao meu coorientador, Professor Carlos Fonseca, por também ter aceito me guiar nesse caminho com paciência e dedicação. Sua contribuição foi mais que fundamental! Eles me conduziram nos meus primeiros passos rumo ao campo da computação.

Agradeço aos colegas, professores do Departamento de Educação e Tecnologias (DEETE) da Universidade Federal de Ouro Preto, no qual sou lotada, pelo apoio e colaboração que me permitiram, por um período, dedicar-me exclusivamente a esta capacitação.

Agradeço ao Professor Rémy de Paiva Sanchis, meu orientador no mestrado, por ser um dos maiores incentivadores das minhas realizações profissionais e acadêmicas. E também por sua contribuição nesta tese.

Aos colegas do PPGMMC agradeço pela convivência e pelas valiosíssimas contribuições.

Agradeço ao CEFET- MG pelo apoio financeiro.

Agradeço ao DEI/CISUC Departamento de Engenharia e Informática da Universidade de Coimbra - Portugal, por me receber (através do Professor Carlos Fonseca), mesmo que por pouco tempo.

Resumo

As Estratégias Evolutivas (EEs) constituem uma classe particular de Algoritmos Evolutivos (AEs). As pesquisas que tratam da análise de estratégias evolutivas têm sido focadas nas aplicações destes algoritmos, usados para resolver problemas das mais diversas áreas, principalmente em espaços de busca contínuos, mas também em espaços discretos. As investigações das EEs também demonstram que estes algoritmos, ainda suscetíveis a diversas pesquisas, são eficientes e populares. Uma análise contendo provas rigorosas de convergência de EEs é uma tarefa difícil devido à estocasticidade destes algoritmos, apesar desta aleatoriedade permitir sua análise sob uma perspectiva matemática.

Neste trabalho são propostas Estratégias Evolutivas com um único progenitor e λ descendentes $(1 \overset{+}{,} \lambda)$-EE. Na primeira delas, o tamanho do passo do algoritmo é modificado de acordo com uma regra de adaptação simples baseada em sucesso, denominada **Regra 1**. Uma extensão desta regra de adaptação do tamanho do passo, denominada **Regra 2**, também é proposta. Nesta estratégia o tamanho do passo é adaptado de acordo com o número de descendentes bem sucedidos. Além destas, a generalização a qualquer número de descendentes e qualquer limiar, de uma EE com controle de mutação baseado no tamanho do passo, também é explorada e nomeada **Regra 3**. Finalmente é formulada uma EE com regra de adaptação do tamanho do passo baseada na combinação da **Regra 1** com a **Regra 3**. A análise teórica destes algoritmos evolutivos concentra-se na investigação sobre sua convergência, quando aplicados em problemas de otimização em espaços de busca contínuos em uma classe de funções estritamente unimodais de uma variável. O estudo sobre a convergência das EEs segue uma abordagem usando funções de Lyapunov estocásticas, no contexto da teoria de martingais. Expressões gerais para as esperanças condicionais dos próximos valores do tamanho do passo e para a distância para o ótimo são analiticamente derivadas para todas as Estratégias, e uma função de Lyapunov apropriada é construída. Os limites superiores da taxa de convergência, bem como os valores dos parâmetros de adaptação, são obtidos através da otimização numérica para valores crescentes de λ, permitindo também uma seleção informada desse parâmetro. Os resultados experimentais contribuem para uma análise dos limites de convergência teóricos obtidos e fornecem uma visão adicional sobre os pontos fortes e fracos da metodologia adotada.

Palavras-chave: Estratégias Evolutivas. Adaptação do tamanho do passo. Teoria de Função de Lyapunov. Taxa de Convergência.

Abstract

The Evolution Strategies (ESs) are a particular class of Evolutionary Algorithms (EAs). The researches dealing with the analysis of evolutionary strategies have been focused on applications of these algorithms used to solve problems in the most diverse fields, mainly in continuous search spaces, but also in discrete spaces. The EEs investigations also show that these algorithms, yet susceptible to various researches, are efficient and popular. An analysis containing rigorous proofs of the convergence of ESs is a difficult task due to the stochastic nature of those algorithms. However this randomness allows this analysis under a mathematical perspective.

Three adaptation rules are proposed in this work as a single parent λ offspring $(1 \overset{+}{,} \lambda)$-ES. The first rule, the algorithm step size is modified according to a simple adaptation rule based on success, named Rule 1. An extension of this rule for adjusting the size of the step, called Rule 2, is also proposed. In this strategy the step size is adapted according to the number of successful offspring. In addition to these, the generalization to any number of descendants and any threshold, EE with mutation control based on step size, is also explored and called Rule 3. An ES with the step size adaptation based on the combination of Rule 1 and Rule 3 is also formulated. The theoretical analysis of these evolutionary algorithms focuses on research on its convergence, when applied to problems in continuous search spaces in a strictly unimodal unidimensional function class. The study on the convergence of EEs follows an approach using Lyapunov stochastic functions, in the context of the theory of martingais. General expressions for the conditional expectations of the next values of step size and distance to the optimum under $(1 \overset{+}{,} \lambda)$-selection are analytically derived for all rules, and an appropriate Lyapunov function is constructed. Convergence rate upper bounds, as well as adaptation parameter values, are obtained through numerical optimization for increasing values of λ, allowing for an informed selection of this parameter, as well. Experimental results are well within the theoretical convergence bounds, and provide additional insight into the strengths and weaknesses of the design methodology adopted.

Keywords: Step-size Adaptation. Evolution Strategy. Lyapunov Function Theory. Convergence Rate.

Lista de Figuras

Lista de Tabelas

Lista de Símbolos

$(1+1)$-EE Estratégia Evolutiva elitista com um único progenitor e um único descendente.

$(\mu+\lambda)$-EE Estratégia Evolutiva elitista com μ progenitores e λ descendentes.

$(\mu{,}\lambda)$-EE Estratégia Evolutiva não elitista com μ progenitores e λ descendentes.

μ Número de indivíduos da população ou número de progenitores.

λ Número de descentes.

ρ Número de indivíduos que serão recombinados.

$(\mu/\rho+\lambda)$-EE Estratégia Evolutiva elitista com μ progenitores, ρ progenitores escolhidos para recombinação e λ descendentes.

$(\mu/\rho,\lambda)$-EE Estratégia Evolutiva não elitista com μ progenitores, ρ progenitores escolhidos para recombinação e λ descendentes.

$(\mu/\rho \overset{+}{,} \lambda)$ Estratégias Evolutivas elitista e não elitista com μ progenitores, ρ progenitores escolhidos para recombinação e λ descendentes.

$(\mu \overset{+}{,} \lambda)$ Estratégias Evolutivas elitista e não elitista com μ progenitores e λ descendentes.

$(1 \overset{+}{,} \lambda)$-EE Estratégias Evolutivas elitista e não elitista com um único progenitor e λ descendentes.

$\{0{,}1\}^n$ Conjunto da $n-$úplas cujas entradas assumem o valor 0 ou o valor 1.

$(1+(\lambda,\lambda))$-AG Algoritmo Genético com um único progenitor em $\{0{,}1\}^n$, λ descendentes gerados a partir de mutações adaptadas segundo a regra de sucesso 1/5 e λ descendentes gerados pelo cruzamento.

Sumário

Apêndices **120**

1 Introdução

As Estratégias Evolutivas (EEs), propostas originalmente por Rechenberg na década de 60, e posteriormente, desenvolvida por Schwefel, constituem uma classe particular de Algoritmos Evolutivos (AEs) (RECHENBERG, 1965; RECHENBERG, 1971; SCHWEFEL, 1965; SCHWEFEL, 1975) conforme descrito em (BACK; HOFFMEISTER; SCHWEFEL, 1991). Esses algoritmos implementam um processo aleatório e dinâmico, inspirado em princípios da evolução natural, que manipula uma população de soluções. Atualmente, são explorados como técnicas de otimização, devido à flexibilidade e adaptação.

Os AEs incluem várias abordagens que foram desenvolvidas de forma independente: *Programação Evolutiva* (PE) (FOGEL; OWENS; WALSH, 1966), *Estratégias Evolutivas* (EEs), *Programação Genética* (PG) (KOZA, 1992) , *Algoritmos Genéticos* (AGs) (HOLLAND, 1992; GOLDBERG, 1989) e combinações desses. Os métodos de Estratégias Evolutivas foram desenvolvidos para resolver problemas de otimização de parâmetros (BACK; HOFFMEISTER; SCHWEFEL, 1991) e têm recebido atenção considerável nas últimas décadas, principalmente devido à sua simplicidade de implementação. As EEs iniciam o processo de busca a partir de uma população constituída de determinado número de indivíduos que são potenciais soluções geradas aleatoriamente. A cada iteração, a partir dos progenitores, novos indivíduos são gerados (descendentes). As primeiras EEs usavam apenas dois operadores genéticos: seleção e mutação. O operador de seleção atua na população corrente e os melhores indivíduos são escolhidos para comporem a população seguinte. O operador de mutação consiste na geração de novos filhos (soluções candidatas), a cada iteração, a partir de uma solução candidata anteriormente avaliada, adicionando a essa solução um vetor aleatório.

1.1 Breve História

As EEs surgiram a partir de experimentos realizados por Ingo Rechenberg e Hans-Paul Schwefel na Universidade Técnica de Berlim para o desenvolvimento de corpos ótimos (com menor resistência ao ar possível) em túneis de vento (BEYER; SCHWEFEL, 2002). Tendo obtido resultados insatisfatórios, Rechenberg realizou mudanças aleatórias sutis nos parâmetros que definiam a forma do corpo, gerando a ideia de mutação.

Os primeiros algoritmos de estratégia evolutiva operavam com um único indivíduo na população, que sofria a ação do operador genético mutação. Rechenberg e Schwefel desenvolveram a estratégia $(1+1)$-EE na qual um pai gera um único descendente e ambos competem pela sobrevivência. Essa forma mais simples de uma EE, mais tarde, foi chamada *estratégia dois membros*. Em 1965, Schwefel concluiu seu mestrado (SCHWEFEL, 1965 apud BEYER; SCHWEFEL, 2002) investigando o algoritmo $(1+1)$-EE com mutações com distribuição binomial em um "cume" parabólico discretizado. Com seu estudo, foi descoberto que o processo poderia ficar preso em certas posições. Dessa forma, surgiu a ideia de usar variáveis contínuas e mutações com distribuições gaussianas (inspirada no fato de que, na natureza, pequenas mudanças ocorrem mais frequentemente do que grandes mudanças).

Em 1971, Rechenberg incorporou a *auto-adaptação* nas EEs (RECHENBERG, 1971 apud MEYER-NIEBERG; BEYER, 2007). A característica essencial da auto-adaptação consiste na introdução dos parâmetros de controle da estratégia na representação genética dos indivíduos, de forma que o processo evolutivo modifica o parâmetro localmente para cada indivíduo.

A auto-adaptação é um método estado da arte para controle de parâmetros e tem, como um de seus objetivos, manter a diversidade adequada entre os indivíduos, a fim de permitir maior capacidade de evolução.

Com o objetivo de otimizar a taxa de convergência de um agoritmo $(1+1)$-EE, Rechenberg propôs a *regra de sucesso* $1/5$. Nesta regra, o tamanho do passo é adaptado com base na taxa de mutações bem sucedidas, de modo que esta taxa se mantenha próxima de $1/5$. O valor $1/5$ foi obtido a partir da análise do algoritmo em duas funções objetivo, conhecidas como modelo esférico e modelo de corredor.

Em 1981, Schwefel (1981) estendeu a abordagem mutativa para a adaptação de parâmetros de mutação com distribuições normais: *Estratégia Mutativa de Controle de Parâmetros* **(MSC)**. O objetivo era encontrar os parâmetros da estratégia com a maior probabilidade de seleção, pressupondo que os parâmetros que produzem indivíduos que vêm a ser selecionados são os mais adequados para o futuro (próximo).

Depois dessa primeira formulação, que poderia conter algumas deficiências, como convergência lenta ou convergência prematura, as EEs foram desenvolvidas em versões denominadas *multi-membros*, introduzindo o conceito de população nesses algoritmos (BEYER; SCHWEFEL, 2002). Os dois principais tipos são $(\mu+\lambda)$-EE (propostas primeiro) e (μ,λ)-EE, sendo μ a quantidade de indivíduos da população inicial que produz λ filhos:

- $(\mu+\lambda)$-EE: De $(\mu+\lambda)$ indivíduos, μ são selecionados.
- (μ,λ)-EE: De λ indivíduos $(\lambda > \mu)$, μ são selecionados.

Nos algoritmos $(\mu+\lambda)$-EE, os pais participam da seleção e sobrevivem enquanto não forem superados por algum dos filhos. Este tipo de seleção é elitista. Na seleção não elitista (μ,λ)-EE, boas soluções podem ser perdidas.

1.2 Algoritmo Padrão de uma EE

Para definir um AE específico, é necessário identificar alguns de seus componentes como (EIBEN; SMITH, 2003):

- Representação (definição de indivíduos): Nas estratégias evolutivas, os indivíduos são representados por um par de vetores reais $(\vec{x},\vec{\sigma})$, sendo $\vec{x}$ o vetor das variáveis da função objetivo, e $\vec{\sigma}$ o vetor de desvio padrão associado, que denota o tamanho do passo de mutação. As entradas do vetor $\vec{\sigma}$ compõem os parâmetros do algoritmo, que se alteram ao longo do tempo. Esses parâmetros também são chamados de parâmetros de controle;
- Função de avaliação (ou função de *fitness*): Definida pela função objetivo;
- População: Geralmente inicializada aleatoriamente, ou a partir de uma aproximação ao ótimo;
- Mecanismo de seleção: A aptidão dos progenitores e dos descendentes (ou apenas dos descendentes) é avaliada e os μ indivíduos com os melhores valores de aptidão são selecionados para comporem a próxima geração. O esquema de seleção é determinístico;
- Operadores de variação, recombinação e mutação: O operador de recombinação pode ser introduzido em determinadas EEs com o objetivo de acelerar a busca. Ele é executado antes da mutação e restringe-se em recombinar um conjunto de progenitores, gerando novas soluções. Considerando ρ $(1 \leq \rho \leq \mu)$ como o número de indíviduos que serão recombinados, as EEs são denotadas por $(\mu/\rho + \lambda)$-EE e $(\mu/\rho, \lambda)$-EE. A recombinação pode ser discreta ou intermediária. O operador de mutação introduz mudanças pequenas e aleatórias em um indivíduo.

O Algoritmo 1 apresenta a estrutura padrão de EEs elitistas e não elitistas com μ progenitores, dos quais ρ são escolhidos para serem recombinados. Os descendentes gerados a partir desta recombinação são afetados pelo operador de mutação e produzem λ novos filhos. Assim como nas versões mais comuns de EEs formuladas, os algoritmos abordados nesta tese são mais simples e os únicos operadores genéticos utilizados são mutação e seleção.

Algoritmo 1: ESTRATÉGIA EVOLUTIVA $(\mu/\rho \overset{+}{,} \lambda)$

Entrada: $\mu,\rho, \lambda, n \in \mathbb{N}^+$

início

$t \leftarrow 0$

$P^{(t)} \leftarrow \{(x_k^{(t)}, \sigma_k^{(t)}, f(x_k^{(t)}))\ ;\ 1 \leq k \leq \mu\}$, sendo $x_k^{(t)} = (x_1,x_2,\ldots,x_n)^{(t)}$

Enquanto critério de parada não for satisfeito

para *cada* $k \in \{1,2,\ldots,\lambda\}$ **faça**

$g_k = (x_k,\sigma_k) \leftarrow$ selecionar indivíduos para recombinação $(P^{(t)},\rho)$

$\sigma_k^{'} \leftarrow mutar_s(\sigma_k)$

$x_k^{'} \leftarrow mutar_x(x_k,\sigma_k)$

fim

fim

$D^{(t)} \leftarrow \{(x_k^{'(t)}, \sigma_k^{'(t)}, f(x_k^{'(t)}))\ ;\ 1 \leq k \leq \mu\}$

$P^{(t+1)} \leftarrow$ seleção $D^{(t)}$, se $(\mu/\rho,\lambda)$

$P^{(t+1)} \leftarrow$ seleção $D^{(t)} \cup P^{(t)}$, se $(\mu/\rho + \lambda)$

$t \leftarrow t + 1$

$x_{final} \leftarrow$ melhor $P^{(t)}$

retorna x_{final}

Neste algoritmo, P representa a população com μ indivíduos $(x_k; \sigma_k; f(x_k))$, $k = 1,\ldots,\mu$, sendo $x_k \in \mathbb{R}^n$ um vetor solução e σ_k contém os parâmetros de controle ou parâmetros endógenos da estratégia, por exemplo, um contador de sucesso ou o tamanho de passo, que serve principalmente para controlar a mutação de x_k. Os valores de σ_k podem ser idênticos para todos os k. Em cada geração, são gerados λ descendentes, por recombinação de $\rho \leq \mu$ indivíduos de P, seguida de mutação de σ e de x_k. As funções $mutar_s$ e $mutar_x$ representam as mutações respectivamente, dos desvios-padrão e das variáveis de decisão. A nova população é adicionada a P e, finalmente, os melhores μ indivíduos são escolhidos para comporem a nova geração. O algoritmo retorna o melhor indivíduo encontrado durante toda a execução (o melhor indivíduo pode não estar na população final caso não haja elitismo).

1.3 Controle de Parâmetros

Algoritmos evolutivos são especificados por vários parâmetros que podem ser configurados de duas maneiras: (i) *ajustados* antes da execução do algoritmo ou (ii)*controlados* durante a execução do algoritmo de otimização (KRAMER, 2010). Assim, a abordagem de ajuste de parâmetros é de natureza estática, enquanto a abordagem de controle é de natureza dinâmica. Eiben, Hinterding e Michalewicz (1999) apresentam uma taxonomia global da configuração de parâmetros em AAs (ver Figura 1).

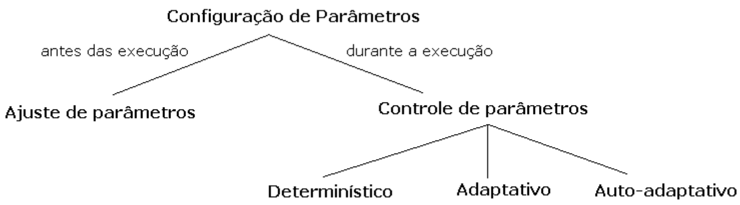

Figura 1 – Taxonomia global da configuração de parâmetros em AEs - retirada de (EIBEN; HINTERDING; MICHALEWICZ, 1999).

Segundo Eiben e Smith (2003), controlar os parâmetros durante a execução do algoritmo potencializa a modelagem do algoritmo na resolução do problema. Por exemplo, grandes passos de mutação podem ser úteis nas primeiras gerações, ajudando a explorar o espaço de busca, e pequenos passos de mutação podem ser necessários nas gerações futuras, para aprimoramento de soluções candidatas. Para Eiben e Smith (2003), as técnicas de controle de parâmetros de um algoritmo evolutivo devem ser classificadas, levando em conta principalmente o que deve ser alterado (por exemplo, representação, função de avaliação, operadores, processo de seleção, taxa de mutação, tamanho da população, etc).

Na técnica de controle determinista, os parâmetros são modificados por uma regra predeterminada, de acordo com um programa de tempo fixo, dependendo explicitamente do número de gerações (KRAMER, 2010). Nos métodos adaptativos de controle de parâmetros a alteração dos parâmetros é feita com base em algum *feedback* na busca, que determina a magnitude e direção da mudança de parâmetro. A famosa regra de sucesso 1/5 é um exemplo de mecanismo de controle adaptativo.

"As estratégias $(\mu \overset{+}{,} \lambda)$-EE deram origem a possibilidades de formas mais sofisticadas de controle de tamanho do passo e conduziram ao desenvolvimento de uma característica muito útil na computação evolutiva: autoadaptação de parâmetros da estratégia"(EIBEN; SMITH, 2003).

No controle auto-adaptativo, os parâmetros a serem adaptados são codificados no indivíduo e passam por mutação e recombinação. Os melhores valores destes levam a melhores indivíduos que, por sua vez, são mais propensos a sobreviver, produzem descendentes e, portanto, propagam esses melhores valores de parâmetros. Esta é uma distinção importante entre esquemas adaptativos e auto-adaptativos: nos últimos, os mecanismos de atribuição de crédito e atualização de diferentes parâmetros de estratégia são inteiramente implícitos, ou seja, representam operadores de seleção e variação do ciclo evolutivo.

Os parâmetros de auto-adaptação mais bem-sucedidos são os parâmetros de mutação. Esta é uma consequência da influência direta do operador de mutação no comportamento de exploração do algoritmo de otimização.

Nas EEs, o controle do tamanho do passo tem papel crucial no processo de busca, afetando diretamente o desempenho destes algoritmos. Se o tamanho do passo de mutação for muito pequeno, o número de iterações necessárias para obter a estimativa do ótimo pode ser muito elevado. Por outro lado, se o tamanho do passo de mutação for muito grande, a aproximação ao ótimo pode ser ruim.

Após essas primeiras configurações, as EEs evoluíram em diversos outros formatos. As principais pesquisas que tratam da análise de estratégias evolutivas têm sido focadas na aplicação destes algoritmos, usados para resolver problemas das mais diversas áreas, principalmente em espaços de busca contínuos, mas também discretos. As investigações das EEs também demonstram que esses algoritmos têm propriedades interessantes como:

- Facilidade de implementação;
- Adaptação para resolução de problemas em cenário *caixa-preta*[1];
- Aplicabilidade em otimização de funções cujas propriedades matemáticas, como continuidade, diferenciabilidade e convexidade não necessitam ser satisfeitas;
- Irrestritos quanto à inicialização dentro do espaço de busca;
- Trabalham com populações de indivíduos, o que permite a realização de buscas simultâneas.

Essas propriedades fazem das EEs métodos populares, como demonstra a literatura, e suas aplicações em problemas de otimização ocupam mais espaço nos diversos trabalhos do que uma análise teórica. Tal análise, contendo provas rigorosas de convergência de EEs, é uma tarefa difícil devido à estocasticidade destes algoritmos, apesar de sua aleatoriedade ainda assim permitir a análise sob uma perspectiva matemática.

Além da propriedade de convergência, outras questões fundamentais na análise de AEs são a rapidez com a qual estes algoritmos convergem para o ótimo, definida como *taxa de convergência*, e o tempo de execução esperado. As pesquisas que tratam da estimativa do tempo de execução esperado, numa perspectiva de complexidade computacional, têm recebido maior atenção nos últimos tempos em relação às pesquisas que tratam da análise de taxa de convergência.

1.4 Objetivos

Este trabalho tem, como objetivo geral, realizar uma análise rigorosa sobre a convergência de estratégias evolutivas adaptativas, aplicadas em problemas de otimização em espaços de busca contínuos, obtendo provas analíticas formais. A análise apresenta uma abordagem usando funções de Lyapunov estocásticas e teoria de martingais.

A partir dos estudos realizados em (SEMENOV; TERKEL, 2003; WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016), trabalhos que investigam a convergência de algorti-

[1] Problemas nos quais a expressão matemática da função objetivo e, consequentemente, o gradiente analítico não são conhecidos.

mos $(1,\lambda)$-EE, é possível analisar EEs adaptativas alternativas, apresentando importantes informações sobre a dinâmica do algoritmo. O algoritmo $(1,2)$-EE adaptativo é talvez a mais simples das EEs não-elitistas para o qual se pode esperar a convergência para o ótimo de uma função. As condições sob as quais isso pode ocorrer não são evidentes e foram verificadas em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016). Contudo, algumas propostas de desdobramento dos resultados obtidos nesse trabalho surgem, principalmente com o objetivo de gerar o uso prático das EEs.

Os objetivos específicos deste trabalho são:

(i) Realizar um estudo análogo ao apresentado em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016) para obter provas formais de convergência de estratégias $(1 \overset{+}{,} \lambda)$-EE, com regras de adaptação baseadas em sucesso na classe de funções unimodais de uma variável, simétricas em torno do eixo que contém o ótimo, obtendo limites para a taxa de convergência e parâmetros que os otimizam com respeito ao número de avaliações da função objetivo;

(ii) Generalizar a análise de EEs com controle de mutação baseado no tamanho do passo (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016) para os casos $(1,\lambda)$-EE com $\lambda > 2$ e $(1+\lambda)$-EE com $\lambda \geq 1$, considerando ainda a otimização de parâmetros de adaptação que foram mantidos fixos no estudo original;

(iii) Combinar as diversas regras de adaptação estudadas num só algoritmo, obtendo um tratamento analítico unificado dessas abordagens;

(iv) Avaliar experimentalmente a qualidade dos parâmetros obtidos para as diversas EEs analisadas por comparação com limites teóricos, e possível refinamento da análise.

1.5 Organização do trabalho

O restante deste trabalho está organizado da seguinte forma: no Capítulo 2 apresentamos uma revisão bibliográfica, contendo alguns dos principais trabalhos relativos à análise teórica de estratégias evolutivas, detalhando, em seção específica, os referenciais teóricos fundamentais nos quais este trabalho é baseado.

As formulações adaptativas dos algoritmos $(1 \overset{+}{,} \lambda)$-EE, com regras de adaptação do tamanho do passo, são formalmente introduzidos no Capítulo 3. Nesse capítulo também é realizada a análise da taxa de convergência destas Estratégias.

O Capítulo 4 contém o estudo da generalização para qualquer número de descendentes e qualquer limiar da estratégia $(1,2)$-EE tratada em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016) e descrita na Subseção 2.3.2 do capítulo 2 .

No Capítulo 5 apresentamos novos cenários derivados das investigações tratadas nos capítulos anteriores, associando os dois tipos de EEs abordados separadamente nos Capítulos 3 e 4, buscando melhores resultados a partir do refinamento da análise.

As conclusões resultantes desta tese, as possibilidades de futuros avanços e as contribuições derivadas deste trabalho são apresentadas no Capítulo 6.

2 Metodologia para Análise de Convergência de Algoritmos Evolutivos

As aplicações dos métodos de otimização têm avançado muito com a utilização de AEs na resoluções de problemas de otimização. A elaboração de algoritmos mais sofisticados se torna necessária quando se deseja resolver problemas complexos, difíceis de serem tratados. A investigação de tais algoritmos se concentra essencialmente na análise de taxa de convergência. No entanto, pesquisas direcionadas para uma análise do comportamento assintótico de algoritmos são realizadas com o auxílio de versões simplificadas dos AEs. Este vácuo existente entre o estudo teórico e a aplicação de AEs representa uma área de pesquisa ampla e promissora.

Diversos trabalhos teóricos sobre AEs apontam resultados sobre tempo de execução desses algoritmos limitados a espaços de busca discretos. No entanto, os resultados para espaços de busca contínuos se concentram em investigar a convergência de AEs e, muitas vezes, dependem de validação através de experimentos, o que é insatisfatório do ponto de vista teórico. Ainda que a aplicabilidade e a eficiência dos AEs tenham sido apresentadas em vários resultados, esses métodos ainda precisam ser analisados sob o ponto de vista teórico, a partir de fundamentos matemáticos rigorosos para um melhor entendimento do funcionamento desses algoritmos. Tal análise teórica também permite o alcance de estratégias mais eficazes e contribui para sua adoção pela comunidade científica. Questões como a convergência e a velocidade de convergência são fundamentais em otimização e devem ser analisadas de forma sistemática. Esses importantes tópicos de pesquisa em algoritmos evolutivos têm sido aplicados de formas distintas em problemas de otimização em espaços de busca contínuos e discretos.

A análise teórica de AEs em problemas discretos constitui-se basicamente na determinação de limites inferiores e superiores para o tempo computacional, o que permite a classificação dos problemas nos quais estes algoritmos são aplicados e a formulação de AEs mais eficientes. O tempo de execução de AEs, *Runtime*, é medido pela quantidade (média) de avaliações da função objetivo necessárias para que um AE encontre a solução ótima pela primeira vez. O *Runtime* de um algoritmo também pode ser considerado como o *First Hitting Time*[1] (traduzem a complexidade do tempo computacional, medida em função do tamanho do problema). Uma ferramenta poderosa, conhecida como *Drift Analysis*[2], é uma das técnicas estado da arte na análise do *Runtime*. Em diversos trabalhos esta técnica é aplicada juntamente com a teoria de Cadeias de Markov.

[1] Número de gerações para um AE encontrar uma solução ótima pela primeira vez.
[2] A expressão não será traduzida, pois é comumente apresentada dessa forma na literatura.

Pesquisas que abordam propriedades de convergência de AEs em problemas de otimização contínuos são encontradas em vários trabalhos. Um grande número de resultados foi obtido através de uma abordagem que descreve o algoritmo como um processo estocástico, por exemplo, Processos Markovianos ou Martingais. Geralmente, deseja-se encontrar regiões de estabilização para os parâmetros das estratégias, para os quais é possível garantir a convergência dos algoritmos.

Existe uma lacuna a ser preenchida em termos da análise teórica entre algoritmos de otimização contínuos e discretos. Com o objetivo de preencher esse vazio, torna-se necessário estabelecer uma ponte entre duas das mais importantes questões teóricas sobre algoritmos evolucionários: o tempo de execução esperado e a taxa de convergência. Para tanto, é interessante realizar um paralelo entre a análise de tempo de execução para problemas discretos e a análise de taxa de convergência para problemas contínuos.

Um método recente, chamado *Método de Otimização de Informação Geométrica* (IGO - Information-Geometric Optimization Algorithms), serve de suporte no preenchimento da lacuna existente na análise teórica entre problemas de otimização contínuos e discretos, pois pode ser aplicado a ambos os problemas (OLLIVIER; ARNOLD; AUGER; HANSEN, 2017). Estes algoritmos se baseiam em conceitos relacionados com *Geometria da Informação*[3]. A nova abordagem IGO é caracterizada por um arcabouço uniforme de algoritmos estocásticos para problemas de otimização, fornecendo informações teóricas para algoritmos existentes e, principalmente, possibilitando a geração de novos algoritmos para problemas de otimização mais complexos. Contudo, evidenciamos que esta tese está direcionada à investigação de convergência de AEs aplicados em problemas de otimização em espaço de busca contínuos.

Nas próximas seções abordamos separadamente resultados que envolvem problemas de otimização em espaços de busca contínuos e discretos. Na seção seguinte apresentamos uma revisão da literatura, envolvendo trabalhos que tratam da análise de tempo de execução de algoritmos evolucionários, mais especificamente, trabalhos que abordam as estimativas de *Hitting Time* e *Runtime* em espaços discretos que usam a metodologia de *Drift Analysis*. Posteriormente, destacamos uma fundamentação teórica, abordando pesquisas que tratam da análise de convergência, incluindo estimativas de taxas de convergência para problemas em espaço de busca contínuos.

[3] Área da matemática que utiliza ferramentas de geometria diferencial no estudo de modelos estatísticos.

2.1 Runtime e Hitting Time para Algoritmos Evolutivos em Espaços de Busca Discretos

A análise do tempo de execução de AEs tem se desenvolvido desde meados de 1990 para espaços de busca discretos, mais geralmente para $\{0,1\}^n$. Uma das ferramentas mais importantes e poderosas utilizadas para estimar o tempo computacional de algoritmos evolutivos e de outras heurísticas de busca randomizadas é chamada *Drift Analysis*. Esta técnica, introduzida por He e Yao (HE; YAO, 2001; HE; YAO, 2002), é considerada estado da arte para o estudo de *First Hitting Time* esperado de AEs aplicados a problemas discretos. Tal análise é baseada em propriedades de um processo estocástico a partir de seu desvio médio, permitindo obter limites inferior e superior para o runtime esperado a partir da variação da distância entre um estado do algoritmo e o conjunto de estados ótimos em uma iteração.

O uso de *Drift Analysis*, combinado com teoria de Cadeias de Markov, com o objetivo de obter limites inferior e superior para o *First Hitting Time* esperado de algoritmos evolutivos em espaços de busca discretos está presente em (HE; YAO, 2001; HE; YAO, 2002; HE; YAO, 2004; JÄGERSKÜPPER, 2008; DOERR; JOHANNSEN; WINZEN, 2012; DOERR; FOUZ; WITT, 2011; DOERR; GOLDBERG, 2013; LEHRE; WITT, 2013; HE; HE; YAO, 2013). Nesses trabalhos, os autores analisam algoritmos $(1+1)$-AE para otimizar funções lineares, modelando a evolução do AE como um processo estocástico, mais especificamente uma Cadeia de Markov. Em seguida o *drift* do processo é calculado e condições sobre o mesmo são analisadas para estimar o tempo computacional.

Em (HE; YAO, 2001; HE; YAO, 2002; HE; YAO, 2004) são apresentadas condições envolvendo o *drift*, sob as quais os AEs convergem em tempo computacional de ordem polinomial ou em tempo de ordem exponencial. He e Yao estudaram em (HE; YAO, 2004) um $(1+1)$-AE sobre funções lineares pseudo-booleanas e um $(n+n)$-AE sobre o problema OneMax de tamanho n e mostraram que a escolha da função de distância influenciou crucialmente a qualidade dos limites médios de complexidade obtidos.

Algoritmos $(1+1)$-AE aplicados a funções lineares pseudo-booleanas também foram submetidos a análise de tempo computacional em (JÄGERSKÜPPER, 2008; DOERR; JOHANNSEN; WINZEN, 2012; WITT, 2013). Em (JÄGERSKÜPPER, 2008) os autores combinam Drift Analysis e cadeias de Markov para obter limites no tempo de otimização esperado, provando uma propriedade de invariância da distribuição do indivíduo sobre o espaço de busca ao longo da execução. O limite superior sobre o tempo de execução esperado do algoritmo foi melhorado em (DOERR; JOHANNSEN; WINZEN, 2012), usando Drift Analysis multiplicativo (versão multiplicativa do teorema clássico de Drift Analysis). Em (DOERR; FOUZ; WITT, 2011) os autores obtiveram limites mais apertados usando Drift Analysis variável (*variable drift* - mais uma versão do teorema clássico de drift analysis).

Mais tarde, em (WITT, 2013), o autor melhora ainda mais o limite superior, também usando Drift Analysis multiplicativo. Em (DOERR; GOLDBERG, 2013) é encontrada um prova alternativa do teorema de Drift Analysis multiplicativo, permitindo que limites de runtime real sejam declarados com alta probabilidade. Em (LEHRE; WITT, 2013) é desenvolvido um teorema universal que generaliza todos os outros existentes, obtendo limites em ambas as caudas da distribuição do *Hitting Time*.

A percepção desenvolvida de AEs simples, promovida pela análise de Runtime, incluindo a determinação de configurações de parâmetros ótimos, alterou o foco da análise de *Runtime* para projeto de algoritmo. A título de exemplo, o *Runtime* também foi analisado para uma classe de algoritmos genéticos $(1 + (\lambda, \lambda))$-AG com tamanho populacional adaptado pela regra de sucesso 1/5. Esta regra foi originalmente proposta por Rechenberg (RECHENBERG, 1973) para adaptação de tamanho de passo em $(1 + 1)$-EE. Enquanto em domínios contínuos esta regra é bem compreendida e amplamente aplicada na análise de convergência de estratégias elitistas (ver (JÄGERSKÜPPER, 2003), (JÄGERSKÜPPER, 2005)), nenhuma teoria comparável está disponível para problemas em domínios discretos. Em (DOERR; DOERR, 2015a) os autores mostram que esta regra também pode ser eficaz para problemas discretos, investigando um algoritmo genético $(1+(\lambda, \lambda))$-AG. Doerr e Doerr (2015a) provam que em funções simples lineares o tempo de otimização do AG $(1 + (\lambda, \lambda))$ é linear, ao considerar que a população é escolhida de acordo com a regra 1/5. Além disso, garantem que o desempenho do algoritmo $(1 + (\lambda, \lambda))$-AG supera qualquer escolha estática de parâmetros e é também o melhor dentre todos as escolhas dinâmicas de parâmetros. Em (DOERR; DOERR, 2015b) os autores usam o operador crossover na formulação do algoritmo $(1 + (\lambda, \lambda))$-AG de forma diferente de outras obras que também tratam da análise deste algoritmo. Doerr e Doerr (2015b) apresentam um melhor entendimento de como utilizar a velocidade de cruzamento sobre o tempo de execução do algoritmo.

Na maior parte da literatura que apresenta estudos sobre análise de convergência e tempo de execução de algoritmos evolutivos, os autores apresentam apenas a análise de convergência para problemas contínuos ou estimam o tempo de execução para problemas em espaço de busca discretos. Um dos problemas em determinar o tempo de execução de um AE para problemas contínuos reside na dificuldade de calcular esse tempo em relação ao erro de aproximação absoluto (distância real ao ótimo) devendo, portanto ser considerada para essa finalidade, a melhoria relativa da aproximação (JÄGERSKÜPPER, 2006).

Em (HE; HE; YAO, 2013) os autores apresentam uma abordagem unificada das análises de convergência, taxas de convergência e tempo de execução de AEs. He, He e Yao (2013) definem condições necessárias e suficientes para a *convergência em distribuição*[4] da

[4]A convergência em distribuição é a mais fraca dentre os tipos de convergência de sequências estocásticas: Uma sequência de variáveis aleatórias X_n converge em distribuição para a variável aleatória X se $F_n(x)$ converge para $F(x)$, para todo ponto de continuidade de F, sendo F_n e F as funções de distribuição de X_n e

Cadeia de Markov Absorvente que modela o algoritmo. Para tanto, a iteração do algoritmo é relacionada com a matriz de probabilidades de transição entre populações intermediárias. Além disso, encontram limites inferior e superior para a taxa de convergência média e usam *Drift Analysis* para encontrar limites para o *First Hitting Time* esperado.

O foco deste trabalho é apresentar uma metodologia de análise de convergência de determinados algoritmos evolutivos (Estratégias Evolutivas), quando aplicados a uma classe específica de funções. Tal metodologia, baseada na Teoria de Estabilidade de Lyapunov, busca encontrar parâmetros ótimos para os quais a taxa de convergência do algoritmo é maximizada.

Cabe aqui ressaltar uma correspondência entre a abordagem usada em (DOERR; JOHANNSEN; WINZEN, 2012) e a metodologia adotada em (SEMENOV; TERKEL, 2003; WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016), trabalhos que constituem a principal fundamentação teórica para esta tese. Em ambos os métodos, tanto para a estimação do tempo de execução em configurações discretas quanto para a análise de convergência para problemas contínuos, o comportamento de um algoritmo evolutivo é analisado através de uma função auxiliar, que deve ser escolhida de maneira que a convergência do algoritmo sobre a verdadeira função objetivo possa ser comprovada verificando as condições na função auxiliar. Em (DOERR; JOHANNSEN; WINZEN, 2012), os autores mostram a existência de *funções de potencial* que atendem uma certa condição sobre seu *drift* (relacionado com o tamanho do problema) e obtêm, a partir destas, um limite para o *First Hitting Time*. Encontrar uma tal função potencial adequada que se comporte de forma suficientemente similar à função objetivo nem sempre é uma tarefa fácil. De forma semelhante, no presente trabalho, assim como em (SEMENOV; TERKEL, 2003; WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016), uma função de Lyapunov estocástica, satisfazendo a propriedade de supermartingal (similar à propriedade de *Drift*) é usada para obter limites para a taxa de convergência.

Na seção que segue, destacamos uma revisão bibliográfica contendo pesquisas que tratam da análise de convergência, incluindo estimativas de taxas de convergência para problemas em espaço de busca contínuos.

2.2 Convergência de Algoritmos Evolutivos em Espaços de Busca Contínuos

Os AEs, apesar de possuírem diversas aplicações, não foram submetidos a provas de convergência para um problema geral. Realizar uma análise teórica desses métodos com demontrações rigorosas de convergência é uma tarefa difícil. Por isso, é natural buscar

X respectivamente (JAMES, 2004).

tais provas a partir de versões mais simplificadas de AEs, que otimizam funções também mais elementares, como exemplo, no caso das EEs, os algoritmos evolutivos mais simples utilizam uma população que consiste de um único indivíduo e usam mutações aleatórias como o único operador de busca: $(1+1)$-EE.

Os métodos de Estratégias Evolutivas passaram por um longo processo de melhoria desde a sua concepção em meados da década de 1960. Eles foram analisados inicialmente por Beyer (1995), que utilizaram uma abordagem de taxa de progresso, calculando a expectativa de progresso de uma etapa de algoritmo para a próxima. Esse tipo de procedimento de análise de convergência não leva em conta a natureza estocástica do tamanho do passo e, consequentemente, a dinâmica do algoritmo não pode ser usada como uma ferramenta matemática.

Dentre os trabalhos nos quais são apresentadas análises teóricas de EEs, alguns dos primeiros em que se avaliam condições em que a convergência deste tipo de algoritmo pode ser provada são (RUDOLPH, 1997a; RUDOLPH, 1997b). Em (RUDOLPH, 1997b) são obtidos limites para a taxa de convergência de AEs que minimizam funções fortemente convexas, pertencentes a uma classe de funções quadráticas com matriz Hessiana definida positiva. Mais tarde, em (RUDOLPH, 2001), a possibilidade de convergência prematura para um ótimo não global do algoritmo $(1+1)$-EE é provada. Nessa publicação, comprovou-se que os algoritmos evolutivos elitistas com um mecanismo de auto-adaptação que se assemelham à regra de sucesso de Rechenberg $1/5$ são atraídos para ótimo não global com probabilidade positiva. Em (JÄGERSKÜPPER, 2006) os resultados da análise de $(1+1)$-EE na função esfera são estendidos a uma classe mais ampla de funções, as formas quadráticas definidas positivas. O autor usa mutações Gaussianas adaptadas pela regra $1/5$. Utilizando este mesmo tipo de mutação, uma primeira análise algorítmica do tempo de execução esperado do algoritmo para espaço de busca contínuo ($\mathbb{R}^n$) é encontrada em (JÄGERSKÜPPER, 2007). Em (AGAPIE, 2001), os autores utilizam sistemas aleatórios com conexões completas, que são extensões Markovianas para modelar métodos guiados pelo algoritmo $(1+1)$-EE.

As EEs são consideradas métodos potentes por serem capazes de realizar outro processo evolutivo embutido no processo global. Esse mecanismo adapta internamente os parâmetros da estratégia originando o conceito de *auto-adaptação*. Métodos adaptativos que redimensionam dinamicamente o tamanho do passo de mutação representam um desenvolvimento chave na computação evolutiva, tendo-se revelado muito eficientes (BEYER, 1995; BÄCK, 1998; EIBEN; HINTERDING; MICHALEWICZ, 1999).

A primeira técnica introduzida para adaptar o tamanho do passo é a conhecida *regra de sucesso* $1/5$, na qual o tamanho do passo é adaptado com base na taxa de mutações bem sucedidas (RECHENBERG, 1971 apud MEYER-NIEBERG; BEYER, 2007). Essa regra foi utilizada em (RUDOLPH, 2001; AGAPIE, 2001) e em (JÄGERSKÜPPER, 2006; JÄGERSKÜPPER, 2007), mutações gaussianas adaptadas pela regra $1/5$ são utilizadas em um algoritmo $(1+1)$-EE. Em seguida, foi proposta a auto-adaptação, sendo o tamanho do passo é alterado e selecionado de acordo com a aptidão. Em (BÄCK, 1998) a auto-adaptação da distribuição da mutação foi implementada em três níveis: adaptações do tamanho do passo "global", "individual" e "adaptação das mutações correlacionadas". O conceito de auto-adaptação foi originalmente desenvolvido para $(1,\lambda)$-EE e posteriormente para $(\mu/\rho,\lambda)$-EE. No algoritmo $(\mu/\rho,\lambda)$-EE, a cada geração a EE mantém uma população de μ soluções candidatas e, a partir dessa população, λ filhos são gerados através de recombinações de $\rho \leq \mu$ pais e mutações. Com o objetivo de superar algumas deficiências da auto-adaptação, foi implementado o *Controle do Comprimento do Passo Cumulativo* (CMA-ES). A *Estratégia Evolutiva de Adaptação da Matriz de Covariância* (CMA-ES) proposta em (HANSEN; OSTERMEIER; GAWELCZYK, 1995) e desenvolvida em (HANSEN; OSTERMEIER, 2001), representa o estado da arte em algoritmos de otimização evolutivos com codificação real. Esse é um método estocástico para otimização de funções não-convexas e não-lineares. Esses algoritmos apresentam importantes propriedades de invariância como, por exemplo, invariância em relação às transformações estritamente monótonas e invariância escalar.

Extensões do CMA-ES e abordagens alternativas para a adaptação da matriz covariância foram propostos na literatura. Em (AUGER; SCHOENAUER; VANHAECKE, 2004) os autores propuseram um método alternativo para calcular a matriz de covariância, estimando localmente a matriz hessiana.

Beyer e Sendhoff (2008) revisitam o processo de auto-adaptação no contexto da adaptação da matriz de covariância. Nesse trabalho, os autores descrevem um novo algoritmo CMSA-ES:$(\mu/\mu_I,\lambda)$-CMA -σ -SA-ES que usa auto-adaptação mutativa em vez de adaptação do comprimento do passo cumulativa para ajustar o tamanho do passo global σ durante a busca. Em (BEYER; SENDHOFF, 2008) também é realizada uma comparação entre o CMSA-ES e o estado da arte CMA-ES: enquanto o CMA-ES obtém performance ligeiramente melhor para tamanhos pequenos de população, o CMSA-ES proposto alcança resultados muito melhores para grandes tamanhos populacionais.

Em (ARNOLD; HANSEN, 2010) é proposta uma nova variante do algoritmo $(1+1)$-CMA-ES, cuja adaptação do passo de mutação é baseada em sucesso: **Adaptação da Matriz Covariância Ativa**. Este algoritmo mantém todas as propriedades de invariância de CMA-ES. No algoritmo CMA-ES, a matriz de covariância da distribuição dos vetores de mutação é adaptada com base nos passos bem-sucedidos, com o intuito de aumentar a variância da distribuição em direções que têm sido bem sucedidas no passado recente. Na adaptação da matriz de covariância ativa, tal variância é aumentada nas direções bem sucedidas e reduzida nas direções mal sucedidas.

Como os AEs representam processos estocásticos, eles podem ser descritos como processos Markovianos ou ainda como Martingais ou Supermartingais. A ligação entre a estabilidade da cadeia normalizada $||x_t||/\sigma_t$ e a convergência linear ou divergência da EE foi apontada pela primeira vez em (BIENVENÜE; FRANÇOIS, 2003) e explorada em (AUGER, 2005). Em (BIENVENÜE; FRANÇOIS, 2003), é apresentada a análise da convergência global de uma estratégia $(1,\lambda)$-EE auto adaptativa na função esfera. Os autores investigam o processo $z_t = ||x_t||/\sigma_t$, sendo σ_t o passo de mutação e x_t o vetor de estados, mostrando que esse é uma Cadeia de Markov homogênea. Seguindo a mesma metodologia, em (AUGER, 2005) a função esfera $f(x) = \|x\|^2$ também foi utilizada. Bienvenüe e FRANÇOIS (2003), Auger (2005) provam a convergência de $(1,\lambda)$-EE, considerando válidas as propriedades de recorrência e Harris-recorrência para o processo $z_t = ||x_t||/\sigma_t$ associado ao algoritmo. Os resultados de (BIENVENÜE; FRANÇOIS, 2003; AUGER, 2005) são baseados na teoria de Cadeias de Markov φ-irredutíveis em espaços de busca contínuos, mais especificamente nas chamadas *condições de drift Foster-Lyapunov*.

Alguns autores (SEMENOV, 2002; SEMENOV; TERKEL, 2003; HART, 2005) utilizam o conceito de martingais e supermartingais para mostrar a convergência de EEs ou para estimar a velocidade de convergência. Em (SEMENOV, 2002; SEMENOV; TERKEL, 2003) os autores tratam da análise de um algoritmo $(1, \lambda)$-EE auto-adaptativo mutativo, com mutação uniforme. Eles utilizaram funções de Lyapunov e relacionaram a teoria de convergência de Martingais com a convergência do algoritmo. Os autores fazem uma analogia entre a convergência de processos estocásticos que descrevem algoritmos evolucionários e o método da Função de Lyapunov estocástica utilizado na teoria de estabilidade de processos estocásticos. A função de Lyapunov descrita em função do processo estocástico é um supermartingal. Seguindo a mesma analogia, em (SEMENOV; TERKEL, 2003), a velocidade de convergência do algoritmo também foi estimada. Contudo, a prova analítica da convergência do algoritmo analisado não foi alcançada. Simulações de Monte Carlo serviram de base para garantir as condições de convergência. Em (HART, 2005), o autor analisa a convergência de AEs com codificação real que realizam a mutação usando passos gerados por variáveis aleatórias discretas. Um dos algoritmos considerados é um $(1,\lambda)$-EE auto-adaptativo para o qual a atualização do comprimento do passo é discretizada para mostrar que este algoritmo converge em problemas unidimensionais simétricos e unimodais. O uso de variáveis aleatórias discretas simplifica a análise e permite a descrição matemática das principais propriedades desses métodos, possibilitando o uso de teoremas de convergência de supermartingais que caracterizam a dinâmica assintótica do algoritmo.

Uma análise rigorosa com demonstrações analíticas para a convergência de uma estratégia evolutiva pode ser vista em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016), referencial teórico para o presente trabalho. Nesse estudo, os autores investigaram a convergência de um algoritmo $(1, 2)$-EE, auto-adaptativo derandomizado com mutações uniformemente distribuídas sobre a classe de funções de uma variável, unimodais e simétricas. Eles seguiram a metodologia desenvolvida em (SEMENOV; TERKEL, 2003), reformulando a estratégia evolutiva e propuseram um método de síntese de Lyapunov para os parâmetros de adaptação, encontrando uma região de estabilização para os mesmos, na qual o algoritmo converge. Também foi estimada a taxa de convergência do algoritmo, tendo sido obtidos limites para o decaimento exponencial da distância ao ótimo e do tamanho do passo. Para isso, os autores utilizaram os resultados apresentados em (SEMENOV; TERKEL, 2003) sobre o comportamento assintótico de supermartingais que satisfazem determinadas condições. Eles utilizam uma função de Lyapunov mais elaborada para provar que o supermartingal dado por tal função satisfaz as propriedades que determinam seu comportamento assintótico. A fim de determinar todos os parâmetros envolvidos no problema, foi utilizada programação quadrática sequencial para resolver um problema de programação não linear resultante da análise. Em (SEMENOV; TERKEL, 2003), esses resultados não foram rigorosamente demonstrados e simulações de Monte Carlo serviram novamente de base para tais provas.

2.3 Fundamentação Teórica

Nesta seção apresentamos os resultados obtidos em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016; SEMENOV; TERKEL, 2003), fundamentos para o desenvolvimento deste trabalho.

A convergência de algoritmos $(1,\lambda)$-EE adaptativos com mutação uniforme é investigada em (SEMENOV; TERKEL, 2003). Neste trabalho, os autores se amparam em duas teorias: Teoria de Funções de Lyapunov Estocásticas e Teoria de Martingais e formulam um teorema de convergência do Supermartingal associado à Função de Lyapunov Estocástica. Tal formulação é usada para provar a convergência do processo estocástico definido pelo algoritmo. Seguindo o quadro teórico desenvolvido em (SEMENOV; TERKEL, 2003), em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016) um procedimento de síntese de Lyapunov para os parâmetros de adaptação de um algoritmo adaptativo simples $(1,2)$-EE é proposto.

A abordagem utilizada nestes trabalhos é análoga a uma ferramenta primária denominada *Síntese de Lyapunov de controle de feedback*, elaborada pela área da Teoria de Controle, para projeto de controle de sistemas não lineares e para análise de estabilidade desses sistemas. A concepção do *design* de Lyapunov é delinear uma lei de controle de feedback que retorna a derivada de uma candidata a função de Lyapunov específica, definida negativa ou semi-definida negativa (GE, 2009).

As funções de Lyapunov, introduzidas pelo russo Aleksandr Mikhailovich Lyapunov em sua tese de doutorado, intitulada *The general problem of the stability of motion*, em 1892, foram utilizadas por muito tempo com o intuito de analisar estabilidade de sistemas dinâmicos. Na década de 1960, esforços consideráveis foram feitos para elaborar métodos sistemáticos de construção das funções de Lyapunov (PARKS, 1992). Informalmente, pode-se definir uma Função de Lyapunov como uma função V que localmente: (i) tenha derivadas parciais contínuas, (ii) seja estritamente positiva, exceto em zero e (iii) tenha derivadas negativas sobre a trajetória para todas as soluções $\frac{dV(y(t))}{dt} < 0$.

É intuitivo pensar em um algoritmo de otimização, juntamente com uma função objetivo, como um sistema dinâmico discreto, pois as variáveis de estado do algoritmo são iteradas de acordo com uma lei de recursão. Dessa forma, os autores de (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016) realizam uma analogia entre o conjunto de pontos para os quais uma Estratégia Evolutiva, que busca encontrar o ótimo de uma determinada função objetivo, pode convergir e o conjunto de pontos de equilíbrio estáveis do sistema dinâmico correspondente. Mais especificamente, eles determinaram a região de estabilização para os parâmetros de adaptação, de modo que o ponto de mínimo da função se torne o único ponto fixo estável do sistema dinâmico associado.

Uma rápida revisão sobre Teoria de Martingais e Teoria de Estabilidade segundo Lyapunov podem ser encontradas nos Apêndices A e B.

2.3.1 Análise de Convergência de Algoritmos Evolutivos usando Função de Lyapunov Estocástica

Nesta seção apresentamos de forma resumida os resultados de Semenov e Terkel (2003), que analisaram a convergência e a velocidade de convergência de algoritmos evolutivos. Eles realizaram uma correspondência entre a análise de convergência de processos estocásticos, que descrevem o comportamento de algoritmos evolutivos e o método de função de Lyapunov estocástica para análise de estabilidade de processos estocásticos.

Segundo Kushner (1967), o método de função de Lyapunov busca obter informações sobre uma família de soluções de um sistema dinâmico $x_{t+1} = F(x_t)$ para $t \in \mathbb{N}$, sendo $x_t \in \mathbb{R}^n$ o vetor de estados do sistema, sem computar os valores numéricos de cada solução. A abordagem de função de Lyapunov para estudo de estabilidade de sistemas determinísticos assegura a propriedade básica de uma função $V(x_t)$ não negativa, tal que $V(0) = 0;\ V(x_t) > 0$ se $x_t \neq 0$, que possui derivada parcial primeira contínua em um conjunto limitado, decrescer monotonicamente para uma constante não negativa υ. A função $V(x)$ é denominada *função de Lyapunov*. Na abordagem de função de Lyapunov estocástica, o objetivo é alcancar informações sobre o comportamento assintótico de $V(x_t)$ a partir de propriedades locais, sendo intuitiva a aplicação de teoremas de Martingais.

Seja $(\Omega, \mathit{A}, \mathit{P})$ um espaço de probabilidade, sendo A uma σ-álgebra de conjuntos mensuráveis e $(\mathit{A}_t, t \in \mathbb{N})$ uma família crescente de sub-σ-álgebras $\mathit{A}_t \subset \mathit{A}$.

Se $V(x_t) \geq 0$ e, dado qualquer $\Delta > 0$, $E^{\mathit{A}_t}[V(x_{t+\Delta})]$[5] $- V(x_t) \leq 0$, ou seja,

[5] $E^{\mathit{A}_t}[V(x_{t+\Delta})] = E[V(x_{t+\Delta})|\mathit{A}_t]$ denota a esperança condicional de V condicionada à sub-σ-álgebra A_t.

se $V(x_t)$ é um supermartingal com probabilidade 1, então teoremas de convergência de supermartingais podem ser aplicados e é possível inferir que $V(x_t) \to v \geq 0$ com probabilidade 1.

Dado x_t um processo estocástico, então a função de Lyapunov associada a x_t é a função numérica:

$$V : \Omega \to \mathbb{R}$$

tal que $V(x_t)$ é um supermartingal. Garantindo propriedades sobre esse supermartingal, pode-se concluir a convergência do processo estocástico.

A estrutura da função de Lyapunov estocástica que descreve o comportamento do algoritmo evolutivo analisado em (SEMENOV; TERKEL, 2003) consiste em um par de processos estocásticos V_t e V_t^* (que caracteriza a variação de V_t), adaptados à família de sub-σ-álgebras $A_t \subset A$ tais que $V_t : \Omega \to U \subset \mathbb{R}$ sendo U fechado.

Os elementos de Ω são denotados por ω e, por simplicidade de notação, são usadas as expressões V_t e V_t^* no lugar de $V_t(\omega)$ e $V_t^*(\omega)$ nas definições e resultados que seguem.

A caracterização de (V_t^*) como um processo que controla a variação de (V_t) é dada pela seguinte condição:

Definição 1 *(**A**- condição)*

*O par de processos estocásticos (V_t, V_t^*) satisfaz a **A**- condição em $G \subset U$ se $\forall\, V \in G, V^* \in \mathbb{R}, \exists\, \delta > 0, \epsilon > 0$ tais que*

$$P[\cup_{k=1}^{\infty}\{|V_{t+k} - V_t| > \delta\}|A_t] > \epsilon$$

quase certamente em $D_1 = \{V_t^* > V^*\} \cap \{|V_t - V| < \delta\}$.

Como consequência desta definição, tem-se o seguinte resultado:

Proposição 1 *Se o par (V_t, V_t^*) de processos estocásticos satisfaz a **A**- condição em $G \subset U$, então*

$$\lim_{t\to\infty} V_t^* = -\infty$$

quase certamente em $\{\omega \in \Omega : \; \exists \lim_{t\to\infty} V_t \in G\}$.

A segunda propriedade a ser garantida sobre o par de processos estocásticos (V_t, V_t^*), com o objetivo de mostrar a convergência de x_t afirma que (V_t^*) é um submartingal se V_t está próximo de um estado V e se V_t^* é suficientemente pequeno. A propriedade é assim formulada:

Definição 2 *(**B**- condição)*

*O par de processos estocásticos (V_t, V_t^*) satisfaz a **B**- condição em $G \subset U$ se $\forall\, V \in G, \exists\, \delta > 0, V^* \in \mathbb{R}$ e $c \in \mathbb{R}$ tais que as seguintes desigualdades*

a) $E^{A_t}[V_{t+1}^*] > V_t^*$

b) $E^{A_t}[V_{t+1}^* I_{\{V_{t+1}^* \geq V^*\}}] \leq c \cdot P(V_{t+1}^* > V^* | A_t)$

seguem quase certamente em $D_2 = \{V_t^* < V^*\} \cap \{|V_t - V| < \delta\}$.

O seguinte teorema trata da convergência do supermartingal V_t:

Teorema 2 *Sejam $V_t : \Omega \to U \subset \mathbb{R}$ um supermartingal limitado, sendo U fechado e V_t^* um processo estocástico, adaptados à família de σ-álgebras (A_t). Se o par (V_t, V_t^*) satisfaz a **A**-condição e a **B**-condição em em um conjunto $U - \Gamma$, com $\Gamma \subset U$, então existe $\lim_{t\to\infty} V_t \in \Gamma$ quase certamente. Além disso, se a **A**-condição é satisfeita em U, então $\lim_{t\to\infty} V_t^* = -\infty$ quase certamente.*

As demonstrações da Proposição 1 e do Teorema 2 são apresentadas em (SEMENOV; TERKEL, 2003). Demonstrações mais detalhadas podem ser encontradas nas seções C.1 e C.2 do Apêndice C.

Como consequência deste resultado pode-se formular o seguinte corolário:

Corolário 3 *Seja (V_t, V_t^*) um par de processos estocásticos que satisfaz todas as hipóteses do Teorema 2. Considerando $U = \mathbb{R}_+$ e $\Gamma = \{0\}$ tem-se:*

$$\lim_{t\to\infty} V_t = 0 \text{ e } \lim_{t\to\infty} V_t^* = -\infty \text{ quase certamente}$$

O Teorema 2 foi aplicado para analisar a convergência de um algoritmo evolutivo auto-adaptativo, cuja população é formada por um único indivíduo representado por um par (x_t, x_t^*), sendo $x_t \in \mathbb{R}$ a variável de decisão da função objetivo e x_t^* responsável por controlar a variação de x_t. Em cada etapa da evolução, o indivíduo (x_t, x_t^*) gera λ filhos, com mutações aleatórias e independentes para ambos os parâmetros, e um novo indivíduo é selecionado, de acordo com a função fitness $f(x)$.

A seguinte formulação é considerada:

$$\begin{aligned} x_{t+1} &= \underset{x \in \left\{x_{i,t+1} = x_t + \xi_{i,t+1} \cdot x_{i,t}^*,\ i=1,\ldots,\lambda\right\}}{\arg\max} f(x) \\ x_{i,t+1}^* &= x_t^* \exp(\vartheta_{i,t+1}) \text{ para } i = 1, \ldots, \lambda \end{aligned} \tag{1}$$

em que $\xi_{i,t}$ para $i = 1, \ldots, \lambda$, são variáveis aleatórias independentes uniformemente distribuídas em $[-1{,}1]$; $\vartheta_{i,t}$ para $i = 1, \ldots, \lambda$, são variáveis aleatórias independentes uniformemente distribuídas em $[-2{,}2]$ e $f(x) = -|x|$.

Assumindo a seguinte função de Lyapunov:

$$V_t = V(x_t, x_t^*) = \left[- \int f(x_{t+1}) dP_{x_t, x_t^*} \right]^{\alpha}$$

sendo $0 < \alpha < 1$ e dP_{x_t, x_t^*} a probabilidade de transição do processo de Markov (x_t, x_t^*) e, considerando $V_t^* = \ln(x_t^*)$, Semenov e Terkel (2003) provam que (V_t, V_t^*) satisfaz a **A**-condição e a **B**-condição em $\mathbb{R}^*$. Utilizando simulações numéricas de Monte Carlo, eles verificam que V_t é um supermartingal e concluem que $x_t \to 0$ e $x_t^* \to 0$ para o algoritmo descrito pela Equação (1).

Os resultados teóricos apresentados em (SEMENOV; TERKEL, 2003) também tratam da análise da velocidade de convergência de supermartingais. O comportamento assintótico de um supermartingal V_t é analizado sobre as seguintes condições: a cada passo, V_t decresce em média por pelo menos uma constante $a > 0$, e a variância condicional de V_t é no máximo $b > 0$. O seguinte resultado é provado:

Proposição 4 *Seja V_t um supermartingal tal que $V_0 = 0$. Se são válidas as seguintes condições*

$$E^{A_t}(V_{t+1}) \leq V_t - a \tag{2}$$

$$E^{A_t}[(V_{t+1} - E^{A_t}(V_{t+1}))^2] \leq b \tag{3}$$

sendo $a > 0$, $b > 0$, então $\forall\, \epsilon > 0$ a seguinte desigualdade segue quase certamente

$$V_t \leq -at + o(t^{0.5+\epsilon}). \tag{4}$$

A demonstração da Proposição 4 pode ser encontrada na Seção C.3 do Apêndice C.

Com o objetivo de mostrar a velocidade de convergência exponencial do algoritmo evolutivo com auto-adaptação descrito pela Equação (1) a Proposição 4 é empregada no seguinte resultado:

Proposição 5 *Considere o processo estocástico (x_t, x_t^*) definido na Equação (1). Se $f(x_t) = -|x_t|$, então este processo converge para $(0,0)$ quase certamente, e as seguintes inequações*

$$|x_t| \leq \exp(-at) \quad e \quad d_t \leq \exp(-at) \tag{5}$$

seguem assintoticamente quase certamente [6].

[6] Um evento ocorre assintoticamente quase certamente se ocorre com probabilidade $1 - o(1)$, ou seja, a probabilidade de sucesso tende a 1 no limite quando $n \to \infty$ (JANSON; LUCZAK; RUCINSKI, 2000).

A seguinte função de Lyapunov estocástica mais complexa é selecionada para obter tal resultado:

$$V_t = V(x_t, x_t^*) = \ln\left[E(|x_{t+1}|)\right] - k\ln(x_t^*) \tag{6}$$

Semenov e Terkel mostram que, para esta função V_t, as Inequações (2) e (3) são satisfeitas. Porém, a Inequação (2) é verificada numericamente através de simulações de Monte Carlo. Além disso, demonstram a Proposição 5 a partir da Inequação (4).

A abordagem adotada por Semenov e Terkel para avaliar a velocidade de convergência de um algoritmo evolutivo pode ser considerada como uma prova de convergência independente pois não foi necessário utilizar o Teorema 2. Apesar de (SEMENOV; TERKEL, 2003) não contemplar uma análise de convergência formal, com demonstrações analíticas, a estrutura desenvolvida nesse trabalho pode ser empregada como um guia genérico para demonstrações de convergência de algoritmos evolutivos. Na subseção seguinte é exposto como os autores em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016) utilizam este guia para investigar a convergência de um algoritmo $(1{,}2)$-EE.

2.3.2 Análise de Estabilidade de Lyapunov e Síntese de Lei de Adaptação de um $(1{,}2)$-EE Auto-Adaptativo Derandomizado

Em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016) os autores propuseram um método de síntese de Lyapunov para os parâmetros de adaptação de um algoritmo $(1{,}2)$-EE adaptativo derandomizado, encontrando uma região de estabilização para estes parâmetros na qual o algoritmo converge. As candidatas a funções de Lyapunov se tornam funções de Lyapunov estocásticas através de escolhas adequadas dos parâmetros de adaptação do algoritmo. A metodologia foi baseada na abordagem contemplada em (SEMENOV; TERKEL, 2003) e apresentada na Subseção 2.3.1.

O algoritmo $(1{,}2)$-EE adaptativo derandomizado considerado tem como objetivo minimizar uma função real $f : \mathbb{R} \to \mathbb{R}$, estritamente unimodal e simétrica em relação ao eixo vertical que contém o ótimo. Neste algoritmo, a cada unidade de tempo discreto t, a população é constituída de um único indivíduo, descrito pelo vetor (x_t, d_t) com $x_t \in \mathbb{R}$ e $d_t \in \mathbb{R}_+^*$, sendo x_t a variável de decisão que determina a função de aptidão do indivíduo e d_t a variável que representa o tamanho do passo do algoritmo que controla a variação de x_t. Cada etapa da evolução é constituída da replicação do indivíduo em dois indivíduos, utilizando mutações aleatórias e independentes, e da seleção de uma nova população (um novo indivíduo) de acordo com a função fitness $f(x)$, determinando simultaneamente o tamanho do passo associado d_{t+1} para a nova geração:

$$\begin{aligned} x_{t+1} &= F(x_t,d_t,\mu_{1,t+1},\mu_{2,t+1}) \\ &= \arg\min_{x\in\{x_{i,t+1}=x_t+\mu_{i,t+1}d_t,\ i=1,2\}} f(x) \\ \\ d_{t+1} &= G(x_t,d_t,\mu_{1,t+1},\mu_{2,t+1}) \\ &= d_t\cdot\beta\cdot\alpha^{\mathsf{sign}\left(\frac{|x_{t+1}-x_t|}{d_t}-\frac{1}{2}\right)} \end{aligned} \tag{7}$$

em que $\mu_{i,t+1},\ i = 1,2$ são variáveis aleatórias uniformemente distribuídas em $[-1,1]$, $\alpha \in (1,+\infty)$ e $1/\alpha < \beta < \alpha$.

Pode-se observar que qualquer função f real de uma variável unimodal, simétrica em torno do eixo que contém o ótimo, pode ser empregada na análise desenvolvida. Isso se deve ao fato de que a evolução do algoritmo se dá de forma independente da função. Mesmo não conhecendo o valor mínimo de f calculada nos filhos $x_{i,t+1}, i = 1,2$, a escolha de x_{t+1} é feita entre a menor distância entre os dois filhos e o valor mínimo de f. A seleção revela-se invariante à translação e por isso pode-se considerar o ótimo como zero, sem perda de generalidade.

Para determinar a região de estabilização para os parâmetros de adaptação, o algoritmo foi interpretado como um sistema dinâmico discreto não-linear e foi verificada a equivalência entre o conjunto dos pontos para os quais o algoritmo em questão converge e o conjunto dos pontos de equilíbrio estáveis do sistema dinâmico correspondente. A ideia principal é escolher adequadamente os parâmentros do algoritmo, de forma a obter, como mínimo da função objetivo, o único ponto de equilíbrio estável do sistema dinâmico associado.

Como o algoritmo $(1,2)$-EE considerado representa um processo estocástico, a condição (c) do Teorema 16 do Apêndice B não pode ser aplicada diretamente para obter a condição de estabilidade uniformemente assintótica e, portanto, foi descrita em termos da esperança condicional:

$$E^{\mathcal{A}_t}[V(x_{t+1},d_{t+1})] \triangleq E[V(x_{t+1},d_{t+1})|x_1,\ldots,x_t,d_1,\ldots,d_t]$$

de uma função de Lyapunov $V(x_t,d_t)$ do processo estocástico (x_t,d_t) adequada (que também compartilha das propriedades da função objetivo).

Assim, para garantir a validade da terceira condição no Teorema 16, deve-se ter:

$$E^{A_t}[V(x_{t+1},d_{t+1})] \le V(x_t,d_t)$$

ou seja, $V(x_t,d_t)$ deve ser um supermartingal. A convergência de tal processo a algum escalar não negativo vem diretamente do Corolário 12 do Teorema 11 do Apêndice A.2 .

Considerando o algoritmo descrito pela Equação (7), e utilizando os resultados da seção anterior, os autores determinaram a região de estabilização para os parâmetros α e β na qual

$$\lim_{t\to\infty} x_t = 0 = \lim_{t\to\infty} d_t$$

quase certamente.

A função de Lyapunov candidata considerada foi

$$V_t = V(x_t, d_t) = |x_t| + wd_t \text{ sendo } w \in \mathbb{R}_+. \tag{8}$$

Além disso foram considerados

$$V_t^* = V^*(x_t, d_t) = \ln(d_t) \tag{9}$$

e $\Gamma = \{0\}$ e o seguinte resultado foi demonstrado.

Proposição 6 *Considere o algoritmo adaptativo $(1{,}2)$-EE descrito pela Equação (7). Para uma escolha adequada dos parâmetros α e β, o par de processos estocásticos (V_t, V_t^*), definido pelas Equações (8) e (9) satisfaz a **A**-condição (descrita na Definição 1 da seção anterior) em $\mathbb{R}$ e a **B**-condição (descrita na Definição 2 da seção anterior) em $\mathbb{R}_+^*$, e, além disso, V_t é um supermartingal.*

Como consequência da Proposição 6 e do Teorema 2 obtém-se a seguinte proposição:

Proposição 7 *Considere o seguinte algoritmo adaptativo $(1{,}2)$-EE descrito pela Equação (7). Existem α e β tais que*

$$\lim_{t\to\infty} x_t = 0 \text{ e } \lim_{t\to\infty} d_t = 0 \text{ quase certamente.}$$

Os resultados teóricos apresentados em (SEMENOV; TERKEL, 2003) também permitiram a investigação da taxa de convergência em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016).

Considere que a desigualdade

$$|x_t| \leq \exp(-at), \tag{10}$$

segue assintoticamente quase certamente para algum escalar $a > 0$.

Então a taxa de convergência de x_t ao ponto de equilíbrio $x^* = 0$ é $e^{-\bar{a}}$, em que $\bar{a}$ é o supremo do conjunto de valores a para os quais (10) segue assintoticamente quase certamente. Se somente um limite inferior a na constante de decaimento exponencial $\bar{a}$ pode ser determinado, com $0 < a \leq \bar{a}$, então $|x_t| \leq \exp(-\bar{a}t) \leq \exp(-at)$.

Considerando a seguinte função de Lyapunov[7] mais elaborada:

$$V_t = V(x_t, d_t) = \ln(|x_t| + wd_t) - k\ln(d_t) \tag{11}$$

sendo $w,k \in \mathbb{R}$, $w > 0$, e $0 \leq k < 1$, as Inequações (2) e (3) foram verificadas e um resultado análogo à Proposição 5 é provado.

Proposição 8 *Considere o processo estocástico (x_t,d_t) definido na Equação (7). Se f é uma função de uma variável, estritamente unimodal, que é simétrica em torno do mínimo, então este processo converge para $(0,0)$ quase certamente, e as seguintes inequações*

$$|x_t| \leq \exp(-at) \quad \textit{e} \quad d_t \leq \exp(-at) \tag{12}$$

seguem assintoticamente quase certamente para algum α_f, α_s, e $a > 0$.

Para verificar a validade da condição dada pela Inequação (2) da Proposição 4 e para provar que V_t defindo pela Equação (8) é um supermartingal, foram obtidas expressões de $E^{A_t}(|x_{t+1}|)$ e $E^{A_t}(d_{t+1})$. Dessa forma foram descritas expressões analíticas para um limite superior (derivado da desigualdade de Jensen[8]) para $E^{A_t}(V_{t+1}) - V_t$, denotado por $\Psi\left(\frac{|x_t|}{d_t}\right)$ e a validade da Inequação (2) foi verificada mostrando que $\Psi\left(\frac{|x_t|}{d_t}\right) \leq -a$.

O limite superior para a taxa de convergência, assim como os parâmetros de adaptação ótimos foram obtidos através da otimização numérica.

2.4 Metodologia Proposta

Seguindo a metodologia exibida nos textos (SEMENOV; TERKEL, 2003; WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016), nos capítulos subsequentes, a análise de conververgência das Estratégias Evolutivas propostas neste trabalho é realizada cumprindo as seguintes etapas:

1. Definição de uma estratégia de adaptação que seja adequada para uma determinada classe de funções a serem otimizadas, bem como uma estrutura de função de Lyapunov estocástica com as mesmas propriedades dessas funções.
2. Obtenção de expressões gerais para $E^{A_t}(|x_{t+1}|)$ e $E^{A_t}(d_{t+1})$, considerando ambas seleções $(1 \overset{+}{,} \lambda)$-EE.

[7]Neste trabalho o termo **Função de Lyapunov**, é usado, realizando um abuso de linguagem, para referir a função que atende condições equivalentes àquelas que definem uma função de Lyapunov.

[8]Definição 3 do Apêndice A.1.

3. Usando a função de Lyapunov definida, estimar a taxa de convergência da estratégia, considerando o decrescimento esperado do supermartingal associado. Para tanto deve-se determinar valores específicos dos parâmetros da estratégia que maximizam o limite superior sobre a constante de decaimento exponencial, provando um resultado análogo à Proposição 8 da seguinte forma:
   3.1. Determinar os parâmentros que satisfazem a Inequação (2).
   3.2. Provar que a função de Lyapunov escolhida possui variância condicional limitada, ou seja, provar que V_t satisfaz a Inequação (3).
   3.3. Provar que a partir da Proposição 4 (Inequação (4)), as Inequações (12) seguem assintoticamente quase certamente.

3 Adaptação do Tamanho do Passo Baseada em Sucesso

Neste capítulo desenvolvemos formulações adaptativas de Estratégias Evolutivas $(1,\lambda)$-EEs e $(1+\lambda)$-EEs, com regras distintas de adaptação do tamanho do passo baseadas em sucesso. Analisamos a convergência destes algoritmos segundo a abordagem para Estratégias Evolutivas Adaptativas Derandomizadas apresentada em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016) e descrita na Subseção 2.3.2 do capítulo anterior.

As EEs $(1 \overset{+}{,} \lambda)$-EE analisadas aqui têm como objetivo minimizar uma função $f : \mathbb{R} \to \mathbb{R}$ estritamente unimodal e simétrica em torno do eixo vertical que contém o mínimo. O vetor de estados do algoritmo é descrito pelo par (x_t, d_t), sendo $x_t \in \mathbb{R}$ a variável de decisão que determina a função de aptidão do indivíduo na iteração t, e $d_t \in \mathbb{R}_+^*$ a variável que representa o tamanho de passo da mutação correspondente.

O processo de evolução é constituído de duas etapas:

- a replicação do indivíduo x_t em λ indivíduos $x_{i,t+1}$, $i = 1, \ldots, \lambda$, utilizando mutações aleatórias e independentes;
- a seleção de uma nova população (um novo indivíduo) x_{t+1} a partir dos filhos (e possivelmente do indivíduo atual, x_t, dependendo do tipo de seleção), de acordo com a função fitness $f(x)$, determinando simultaneamente o tamanho do passo associado d_{t+1} para a nova geração.

Duas regras de adaptação do tamanho do passo são consideradas:

Regra 1 A primeira regra de adaptação do tamanho do passo considerada é a regra de adaptação baseada em sucesso, originalmente formulada por (KERN et al., 2004), que pode ser descrita como segue:

- Se qualquer um dos filhos é pelo menos tão bom quanto o pai (sucesso), então o tamanho do passo aumenta;
- Se todos os filhos são piores que o pai (fracasso), então o tamanho do passo diminui.

Regra 2 Na segunda regra de adaptação do tamanho do passo considerada, o tamanho do passo aumenta ou diminui em função do número de descendentes bem-sucedidos.

As equações de estados de um algoritmo $(1,\lambda)$-EE com a **Regra 1** de adaptação podem ser escritas da seguinte maneira:

$$
\begin{aligned}
x_{t+1} &= F(x_t, d_t, \mu_{1,t+1}, \ldots, \mu_{\lambda,t+1}) \\
&= \underset{x \in \{x_{i,t+1} = x_t + \mu_{i,t+1} \cdot d_t,\ i=1,\ldots,\lambda\}}{\arg\min} f(x)
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
d_{t+1} &= G(x_t, d_t, \mu_{1,t+1}, \ldots, \mu_{\lambda,t+1}) \\
&= \begin{cases} \alpha_s \cdot d_t & \text{se } \exists\, i \in \{1, \ldots, \lambda\} : f(x_{i,t+1}) \leq f(x_t) \\ \alpha_f \cdot d_t & \text{caso contrário} \end{cases}
\end{aligned} \tag{14}
$$

em que $\mu_{i,t+1}$, $i = 1, \ldots, \lambda$, são variáveis aleatórias uniformemente distribuídas em $[-1,1]$, $0 < \alpha_f \leq 1$, e $\alpha_s \geq 1$ são os parâmetros do algoritmo.

A única diferença entre $(1,\lambda)$-EE e $(1+\lambda)$-EE é o método de seleção: neste último caso, o melhor entre os λ filhos e o pai x_t é selecionado para se tornar x_{t+1}:

$$
x_{t+1} = \underset{x \in \{x_{i,t+1} = x_t + \mu_{i,t+1} d_t,\ i=1,\ldots,\lambda\} \cup \{x_t\}}{\arg\min} f(x) \tag{15}
$$

Como consequência, o pai sobrevive até que filhos com valores de função objetivo inferiores ao dele sejam gerados.

Observamos que qualquer função real f de uma variável, estritamente unimodal e simétrica em relação ao eixo vertical que contém o mínimo, pode ser empregada na análise a ser desenvolvida. Isso se deve ao fato de que a evolução do algoritmo se dá de forma independente da função. Mesmo não conhecendo o valor mínimo de f calculada nos filhos $x_{i,t+1}, i = 1, 2, \ldots, \lambda$, a escolha de x_{t+1} é feita entre a menor distância entre os filhos e o valor mínimo de f. Além disso, a seleção é invariante por translação e por isso podemos considerar o ótimo como zero, sem perda de generalidade. Também observamos que d_{t+1} depende indiretamente de $(\mu_{1,t+1}, \ldots, \mu_{\lambda,t+1})^T$ através de $x_{i,t+1}$.

Podemos argumentar que a **Regra 1** é baseada em eventos de sucesso e fracasso, no sentido de realizar o ajuste do tamanho do passo com base na informação sobre o quão próximo do ótimo o atual pai está. De fato, a probabilidade de sucesso é zero quando o pai está localizado no ótimo, enquanto essa probabilidade deve ser grande quando o pai está longe do ótimo. Para $\lambda > 1$, espera-se que essa estimativa de sucesso seja melhorada considerando o número de filhos bem sucedidos, em vez de simples eventos de sucesso e fracasso.

Considerando $S = \{i \in \{1, \ldots, \lambda\}; f(x_{i,t+1}) \leq f(x_t)\}$ e $|S| \in \{0, 1, \ldots\}$ a cardinalidade de S (número de descendentes melhores que o pai), independentemente da estratégia de seleção a ser empregada, a **Regra 2** de adaptação é formalizada pela equação:

$$
d_{t+1} = \alpha_j \, d_t, \quad \text{se } |S| = j \tag{16}
$$

sendo $\alpha_j > 0$, para cada $j = 0, \ldots, \lambda$.

A **Regra 2** é uma extensão da **Regra 1** porque esta última pode ser recuperada a partir da primeira simplesmente definindo $\alpha_0 = \alpha_f$ e $\alpha_1 = \cdots = \alpha_\lambda = \alpha_s$. Vale ressaltar que as duas regras são idênticas quando $\lambda = 1$.

3.1 Análise da Taxa de Convergência

Definidas as estratégias de adaptação baseadas em sucesso e a classe de funções a serem otimizadas, é necessário estabelecer a estrutura de função de Lyapunov estocástica com as mesmas propriedades dessas funções.

Assim como em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016), a função de Lyapunov adotada é definida pela Equação (11) e é reproduzida novamente para conveniência do leitor:

$$V_t = V(x_t, d_t) = \ln(|x_t| + wd_t) - k\ln(d_t)$$

sendo $w, k \in \mathbb{R}$, $w > 0$, e $0 \leq k < 1$,

Para a classe de funções reais de uma variável, estritamente unimodais e simétricas em torno do eixo que contém o mínimo, esta função satisfaz propriedades equivalentes às que definem uma função de Lyapunov. Com efeito, como V_t descreve a energia de um sistema dinâmico que deve diminuir com o tempo, podemos verificar que um aumento de d_t não permite um aumento da energia. De forma equivalente à Condição a) do Teorema 16, observamos que $V_t \to -\infty$ somente quando $x_t \to 0$ e $d_t \to 0$. De fato, $V_t = \ln(|x_t|d_t^{-k} + wd_t^{1-k})$ e portanto $V_t \to -\infty \Leftrightarrow |x_t|d_t^{-k} + wd_t^{1-k} \to 0$. Como $w > 0, |x_t| \geq 0$ e $d_t \geq 0$, então $|x_t|d_t^{-k} + wd_t^{1-k} \to 0 \Leftrightarrow |x_t|d_t^{-k} \to 0$ e $wd_t^{1-k} \to 0$. Mas $|x_t|d_t^{-k} \to 0 \Leftrightarrow |x_t| \to 0$ ou $d_t \to \infty$. Além disso, assumindo $1 - k > 0,\ wd_t^{1-k} \to 0 \Leftrightarrow d_t \to 0$.

Com o objetivo de estabelecer a convergência exponencial do algoritmo proposto, apresentamos aqui a demonstração do seguinte resultado (análogo à Proposição 8 do Capítulo 2):

Proposição 9 *Considere o processo estocástico (x_t, d_t) definido por um dos pares de Equações (13) e (14) (**Regra 1** não elitista) ou (15) e (14) (**Regra 1** elitista) ou (13) e (16) (**Regra 2** não elitista) ou (15) e (16) (**Regra 2** elitista). Se f é uma função de uma variável real, estritamente unimodal que é simétrica em torno do mínimo, então este processo converge para $(0,0)$ quase certamente, e as seguintes inequações*

$$|x_t| \leq \exp(-at) \quad e \quad d_t \leq \exp(-at) \tag{17}$$

*seguem assintoticamente quase certamente para algum α_f, α_s (**Regra 1**), ou α_j com $j = 0, \ldots, \lambda$ (**Regra 2**) e $a > 0$.*

A prova da Proposição 9 consiste de três etapas:

1. Determinar valores para α_f, α_s, (ou α_i, com $i = 0,1,\ldots,\lambda$ para a **Regra 2**), w, k e a tais que a Desigualdade (2) da Proposição 4 é verificada.
2. Provar que a variância condicional de V_t é limitada (Desigualdade (3)).
3. Provar que, a partir da Proposição 4, as desigualdades em (17) seguem assintoticamente quase certamente.

Notamos que as Etapas 2 e 3 dependem da estrutura da função de Lyapunov, do tipo de distribuição da mutação e do tipo de adaptação do tamanho do passo considerado. No entanto, as Etapas 2 e 3 não dependem do tipo de seleção da estratégia ($(1,\lambda)$ ou $(1+\lambda)$). As demonstrações destas etapas são idênticas às apresentadas em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016), com as seguintes substituições de variáveis no caso da etapa 2:

Regra 1 $\beta/\alpha = \alpha_f$ e $\beta\alpha = \alpha_s$,

Regra 2 $\beta/\alpha = \min_{j\in\{0,\ldots,\lambda\}} \alpha_j$ e $\beta\alpha = \max_{j\in\{0,\ldots,\lambda\}} \alpha_j$.

Para a conveniência do leitor, reproduzimos integralmente tais demonstrações para a **Regra 1**.

A Etapa 2 pode ser provada, mostrando que V_{t+1} possui suporte de comprimento limitado para todo x_t e d_t. Como

$$V_{t+1}-E^{A_t}(V_{t+1}) = \underbrace{\ln(|x_{t+1}|+wd_{t+1}) - E^{A_t}[\ln(|x_{t+1}|+wd_{t+1})]}_{A} + k\{\underbrace{E^{A_t}[\ln(d_{t+1})] - \ln(d_{t+1})}_{B}\},$$

é suficiente mostrar que existem $0 < b_1,\, b_2 < \infty$ tais que $|A| \leq b_1$ e $|B| \leq b_2$.

Como o próximo valor do tamanho do passo satisfaz $d_t\alpha_f \leq d_{t+1} \leq d_t\alpha_s$, então

$$\ln(d_{t+1}) - E^{A_t}[\ln(d_{t+1})] \leq \ln(d_t\alpha_s) - \ln(d_t\alpha_f) = \ln\left(\frac{\alpha_s}{\alpha_f}\right).$$

Portanto, existe $0 < b_2 = \ln\left(\frac{\alpha_s}{\alpha_f}\right) < \infty$ tal que $|B| \leq b_2$.

Por construção, $|x_t| - d_t \leq |x_{t+1}| \leq |x_t| + d_t$. Considerando $|x_t| < d_t$, podemos escrever $\ln(wd_t\alpha_f) \leq \ln(|x_{t+1}| + wd_{t+1}) \leq \ln(2d_t + wd_t\alpha_s)$, ou equivalentemente

$$\ln(d_t) + \ln(w\alpha_f) \leq \ln(|x_{t+1}| + wd_{t+1}) \leq \ln(d_t) + \ln(2 + w\alpha_s).$$

Neste caso, existe $b_1' = \ln(2 + w\alpha_s) - \ln(w\alpha_f)$ tal que $|A| \leq b_1'$.

Se $|x_t| \geq d_t$, então

$$\ln(|x_t| - d_t + wd_t\alpha_f) \leq \ln(|x_{t+1}| + wd_{t+1}) \leq \ln(|x_t| + d_t + wd_t\alpha_s)$$

e consequentemente

$$|A| \leq \ln\left(\frac{|x_t| + d_t + wd_t\alpha_f}{|x_t| - d_t + wd_t\alpha_s}\right) = \ln\left(\frac{\frac{|x_t|}{d_t} + 1 + w\alpha_f}{\frac{|x_t|}{d_t} - 1 + w\alpha_s}\right) = \xi\left(\frac{|x_t|}{d_t}\right).$$

Como ξ é contínua e os seguintes limites são finitos

$$\lim_{\frac{|x_t|}{d_t}\to 1} \xi\left(\frac{|x_t|}{d_t}\right) = b_1'$$

$$\lim_{\frac{|x_t|}{d_t}\to \infty} \xi\left(\frac{|x_t|}{d_t}\right) = \ln(1) = 0$$

então ξ é limitada. Neste caso, existe $0 < b_1'' < \infty$ tal que $|A| \leq b_1''$. Assim, basta considerar $b_1 = \max\{b_1', b_1''\}$.

Dessa forma

$$\begin{aligned} E^{A_t}\{[V_{t+1} - E^{A_t}(V_{t+1})]^2\} &= E^{A_t}(A^2 + 2kAB + k^2B^2) \\ &= E^{A_t}(|A|^2 + 2kAB + k^2|B|^2) \ . \\ &\leq b_1^2 + 2kb_1b_2 + k^2b_2 = b \end{aligned}$$

Para a demonstração da Etapa 3, notamos que:

$$\begin{aligned} V_t \leq -at + o(t^{0,5+\epsilon}) \ &\Leftrightarrow \ \ln(|x_t| + wd_t) - k\ln(d_t) \leq -at + o(t^{0,5+\epsilon}) \\ &\Leftrightarrow \ \frac{|x_t| + wd_t}{d_t^k} \leq \exp(-at + o(t^{0,5+\epsilon})) \\ &\Leftrightarrow \ |x_t|d_t^{-k} + wd_t^{1-k} \leq \exp(-at + o(t^{0,5+\epsilon})). \end{aligned}$$

Como $|x_t|d_t^{-k} \geq 0$ e $wd_t^{1-k} \geq 0$, então

$$|x_t|d_t^{-k} \leq \exp(-at + o(t^{0,5+\epsilon})) \tag{18}$$

$$wd_t^{1-k} \leq \exp(-at + o(t^{0,5+\epsilon})) \tag{19}$$

Da Desigualdade (19) temos que:

$$\begin{aligned} d_t &\leq \frac{\exp(-at + o(t^{0,5+\epsilon}))^{1/(1-k)}}{w^{1/(1-k)}} \\ &= \frac{\exp(-at)^{1/(1-k)}\exp(o(t^{0,5+\epsilon}))^{1/(1-k)}}{w^{1/(1-k)}} \\ &= \exp(-at)\left[w^{1/(1-k)} + \exp\left(\frac{-akt}{1-k} + \frac{1}{1-k}o(t^{0,5+\epsilon})\right)\right]. \end{aligned}$$

Como $L = o(t^{0,5+\epsilon}) \Leftrightarrow \dfrac{L}{t^{0,5+\epsilon}} \to 0$, podemos tomar $\epsilon < 1/2$ e t suficientemente grande de forma que $w^{1/(1-k)} + \exp\left(\dfrac{-akt}{1-k} + \dfrac{1}{1-k}o(t^{0,5+\epsilon})\right) \leq 1$.

Logo $d_t \leq \exp(-at)$.

Da Desigualdade (18) temos que

$$\begin{aligned} |x_t| &\leq \exp(-at + o(t^{0,5+\epsilon}))d_t^k \\ &\leq \exp(-at + o(t^{0,5+\epsilon}))\exp(-atk) \\ &= \exp(-at)[\exp(-atk) + o(t^{0,5+\epsilon})]. \end{aligned}$$

Novamente, podemos tomar $\epsilon < 1/2$ e t suficientemente grande de forma que $\exp(-atk) + o(t^{0,5+\epsilon}) \leq 1$.

Logo $|x_t| \leq \exp(-at)$.

Para demonstrar a Etapa 1 para os algoritmos $(1 \overset{+}{,} \lambda)$-EEs, é necessário determinar valores para α_f, α_s, (ou α_i, com $i = 0{,}1,\ldots,\lambda$ para a **Regra 2**), w, k e a tais que a $E^{A_t}(V_{t+1}) \leq V_t - a$.

Utilizando a função V_t escolhida, a desigualdade anterior é descrita como:

$$E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] - kE^{A_t}[\ln(d_{t+1})] - \ln(|x_t| + wd_t) + k\ln(d_t) \leq -a. \tag{20}$$

Como $\ln(\cdot)$ é uma função côncava, usando a desigualdade de Jensen (ver Definição 3 do Apêndice A.1), obtemos

$$E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \leq \ln[E^{A_t}(|x_{t+1}| + wd_{t+1})] = \ln[E^{A_t}(|x_{t+1}|) + wE^{A_t}(d_{t+1})] \tag{21}$$

e, assim, é suficiente provar que existem α_f, α_s (**Regra 1**), α_j, $j = 0,\ldots,\lambda$ (**Regra 2**), w, k e a tais que a seguinte desigualdade é válida:

$$\ln[E^{A_t}(|x_{t+1}|) + wE^{A_t}(d_{t+1})] - kE^{A_t}[\ln(d_{t+1})] - \ln(|x_t| + wd_t) + k\ln(d_t) \leq -a. \tag{22}$$

Devido à forma particular do seu lado esquerdo, a Desigualdade (22) pode ser reescrita como:

$$\ln\left[\frac{E^{A_t}(|x_{t+1}|) + wE^{A_t}(d_{t+1})}{|x_t| + wd_t}\right] - k(E^{A_t}[\ln(d_{t+1})] - \ln(d_t)) \leq -a. \tag{23}$$

A fim de obter um limite superior para $E^{A_t}(V_{t+1}) - V_t$ para um número arbitrário de filhos λ, as esperanças $E^{A_t}(|x_{t+1}|)$, $E^{A_t}(d_{t+1})$ e $E^{A_t}[\ln(d_{t+1})]$ são calculadas primeiramente para cada tipo de seleção e cada tipo de regra de adaptação.

Em todas as variantes de EEs consideradas na Proposição 9, o próximo pai x_{t+1} é selecionado independentemente da regra de adaptação do tamanho do passo. Da mesma forma, o tamanho do passo correspondente d_{t+1} é adaptado independentemente da estratégia de seleção. Portanto, $E^{A_t}(|x_{t+1}|)$ depende exclusivamente da estratégia de seleção, e $E^{A_t}(d_{t+1})$ e $E^{A_t}[\ln(d_{t+1})]$ dependem apenas da regra de adaptação. Além disso, como a função objetivo f é assumida simétrica em torno do eixo que contém o mínimo, o cálculo dessas esperanças pode ser restrito a valores não negativos de x_t.

Como o esquema de adaptação do tamanho do passo é exatamente o mesmo para $(1,\lambda)$-EE e $(1+\lambda)$-EE, $E^{A_t}(d_{t+1})$ é o mesmo em ambos os algoritmos. No entanto, dependendo do tipo de seleção, $E^{A_t}(|x_{t+1}|)$ resulta em diferentes expressões.

Para calcular $E^{A_t}(|x_{t+1}|)$, verificamos que, para cada $i = 1,\ldots,\lambda$, $x_{i,t+1} = x_t + \mu_{i,t+1}d_t$ é uma variável aleatória definida no espaço[1] de probabilidade $(\mathbb{R}, B, P)$ com a seguinte função de distribuição:

$$F_{x_{i,t+1}}(x) = F_{\mu_{i,t+1}}\left(\frac{x-x_t}{d_t}\right) = \begin{cases} 0, \text{ se } x < x_t - d_t \\ \frac{x-x_t+d_t}{2d_t}, \text{ se } x_t - d_t \leq x \leq x_t + d_t \\ 1, \text{ se } x > x_t + d_t \end{cases}$$

Assim sua função de densidade é $f_{x_{i,t+1}}(x) = f_i(x) = \begin{cases} \frac{1}{2d_t}, \text{ se } x_t - d_t \leq x \leq x_t + d_t \\ 0, \text{ caso contrário} \end{cases}$

Sejam $\tilde{X} = (x_{1,t+1}, x_{2,t+1}, \ldots, x_{\lambda,t+1})$ um vetor aleatório e $\varphi : \mathbb{R}^\lambda \to \mathbb{R}$ definida por $\varphi(x_1,x_2,\ldots,x_\lambda) = \min\{|x_1|,|x_2|,\ldots,|x_\lambda|\}$.

Então $\varphi(\tilde{X}) = |x_{t+1}|$ e

$$\begin{aligned} E^{A_t}(|x_{t+1}|) &= E^{A_t}[\varphi(\tilde{X})] \\ &= \underbrace{\int_{x_t-d_t}^{x_t+d_t}\int_{x_t-d_t}^{x_t+d_t}\cdots\int_{x_t-d_t}^{x_t+d_t}}_{\lambda \text{ vezes}} \varphi(x_1,x_2,\ldots,x_\lambda)f_1(x_1)\cdots f_\lambda(x_\lambda)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_\lambda \\ &= \frac{1}{(2d_t)^\lambda}\underbrace{\int_{x_t-d_t}^{x_t+d_t}\int_{x_t-d_t}^{x_t+d_t}\cdots\int_{x_t-d_t}^{x_t+d_t}}_{\lambda \text{ vezes}} \min\{|x_1|,\ldots|x_\lambda|\}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_\lambda \qquad (24) \end{aligned}$$

A Equação (24) permite obter $E^{A_t}(|x_{t+1}|)$ para λ fixo. Com o objetivo de obter uma expressão para $E^{A_t}(|x_{t+1}|)$ em função de λ, denotamos $x_{t+1} = z$ por simplicidade de notação.

[1] *B* representa a $\sigma-$álgebra de *Borel* na reta real, que é a menor $\sigma-$álgebra contendo todos os intervalos.

Considerando $\mathbb{R}^n$ munido com a $\sigma-$álgebra[2] de *Borel*, B^n e $H \subset \mathbb{R}^n$ uma região de $\mathbb{R}^n$, o volume n-dimensional de H é definido por

$$\text{Vol}H = \underbrace{\int \dots \int}_{n \text{ vezes}} 1\mathrm{d}x_1 \dots \mathrm{d}x_n.$$

Quando $n = 2$, por exemplo, Vol$H =$ área H.

Dados o intervalo $I = [x_t - d_t, x_t + d_t]$, seu subconjunto $B_i = \{x_{i,t+1} \in I\ ; |x_{i,t+1}| \geq |z|\} \subset I$ e os subconjuntos de $\mathbb{R}^\lambda$: $G = I^\lambda$ e $B = \prod_{i=1}^{\lambda} B_i$ (B é um boreliano[3] da σ-álgebra $\mathbb{R}^\lambda$), como o vetor aleatório $\tilde{X}$ é uniformemente distribuído em G com distribuição:

$$P_{\tilde{X}}(B) = P(\tilde{X} \in B) = \frac{\text{Vol}(B \cap G)}{\text{Vol}G}$$

e $B \in B^\lambda$ (σ-álgebra de Borel em $\mathbb{R}^\lambda$), então a Equação (24) pode ser escrita da forma:

$$E^{A_t}(|x_{t+1}|) = \sum_{i=1}^{\lambda} \int_{x_t-d_t}^{x_t+d_t} \varphi(\tilde{X}) P_{\tilde{X}}(B)(\mathrm{d}x) = \lambda \int_{x_t-d_t}^{x_t+d_t} |z| \frac{\text{Vol}(B \cap G)}{\text{Vol}G} \mathrm{d}z$$

Além disso,

$\text{Vol}(B \cap G) = \prod_{i=1}^{\lambda} \text{Vol}B_i = (\prod_{i=1}^{\lambda-1} \text{Vol}B_i)\text{Vol}B_z = \prod_{i=1}^{\lambda-1} \text{Vol}B_i = (\text{Vol}B_i)^{\lambda-1}$, sendo $B_z = \{z\}$ e $\text{Vol}G = (2d_t)^\lambda$.

Logo

$$E^{A_t}(|x_{t+1}|) = \frac{\lambda}{(2d_t)^\lambda} \int_{x_t-d_t}^{x_t+d_t} |z|(\text{Vol}B_i)^{\lambda-1} \mathrm{d}z. \tag{25}$$

Considerando $\lambda = 2$, é possível representar graficamente as regiões de G em que $\min\{|x_{1,t+1}|,|x_{2,t+1}|\} = |x_{1,t+1}|$ e $\min\{|x_{1,t+1}|,|x_{2,t+1}|\} = |x_{2,t+1}|$ e calcular $E^{A_t}(|x_{t+1}|)$ a partir da Equação (24) . No Apêndice D apresentamos o esboço dessas regiões e o cálculo de $E^{A_t}(|x_{t+1}|)$ para a estratégia $(1{,}2)$-EE.

Para determinar $E^{A_t}(|x_{t+1}|)$, para $(1{,}\lambda)$-EE primeiramente, dois casos devem ser considerados: x_t próximo à origem e x_t longe da origem.

Caso 1 - $(1{,}\lambda)$**-EE:** $0 \leq x_t < d_t$ que significa que x_t está "próximo" à origem.

Neste caso, três intervalos de integração precisam ser considerados de acordo com as possibilidades de sinal de z.

Se $z < 0$, então $|z| = -z$ e o primeiro intervalo de integração no cálculo de $E^{A_t}(|x_{t+1}|)$ é $x_t - d_t \leq z < 0$.

[2] Menor $\sigma-$álgebra contendo todos os retângulos n dimensionais.

[3] Um conjunto que pode ser obtido a partir de um número enumerável de retângulos λ-dimensionais e aplicando as operações $\cup, \cap$, e ^{c}(complementar) um número enumerável de vezes.

Assim $|x_{i,t+1}| > |z| \Leftrightarrow |x_{i,t+1}| > -z \Leftrightarrow x_{i,t+1} > -z$ ou $x_{i,t+1} < z$ e consequentemente

$$\begin{aligned} \text{Vol}B_i &= \text{Vol}\{x_{i,t+1} \in I \,; x_{i,t+1} > -z\} + \text{Vol}\{x_{i,t+1} \in I \,; x_{i,t+1} < z\} \\ &= [x_t + d_t - (-z)] + [z - (x_t - d_t)] = 2d_t + 2z \end{aligned}$$

Se $z \geq 0$, $|z| = z$. Neste caso podemos ter $-z > x_t - d_t$ ou $-z < x_t - d_t$.

Se $-z > x_t - d_t$, então $z < -x_t + d_t$ e o segundo intervalo de integração no cálculo de $E^{A_t}(|x_{t+1}|)$ é $0 < z < -x_t + d_t$.

Assim $|x_{i,t+1}| > |z| \Leftrightarrow |x_{i,t+1}| > z \Leftrightarrow x_{i,t+1} > z$ ou $x_{i,t+1} < -z$ e consequentemente

$$\begin{aligned} \text{Vol}B_i &= \text{Vol}\{x_{i,t+1} \in I \,; x_{i,t+1} > z\} + \text{Vol}\{x_{i,t+1} \in I \,; x_{i,t+1} < -z\} \\ &= (x_t + d_t - z) + [(-z) - (x_t - d_t)] = 2d_t - 2z. \end{aligned}$$

Se $-z < x_t - d_t$, ou seja, $z > -x_t + d_t$, então o terceiro intervalo de integração no cálculo de $E^{A_t}(|x_{t+1}|)$ é $-x_t + d_t < z < x_t - d_t$.

Neste caso Vol$\{x_{i,t+1} \in I \,; x_{i,t+1} < -z\} = 0$ e consequentemente

VolB_i = Vol$\{x_{i,t+1} \in I \,; x_{i,t+1} > z\} = x_t + d_t - z$.

Logo

$$\begin{aligned} E^{A_t}(|x_{t+1}|) &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{0} -z(2z+2d_t)^{\lambda-1}\mathrm{d}z + \int_0^{-x_t+d_t} z(2d_t-2z)^{\lambda-1}\mathrm{d}z \right. \\ &\quad \left. + \int_{-x_t+d_t}^{x_t+d_t} z(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= \left(\frac{1}{\lambda+1}\right)\left[\left(\frac{x_t}{d_t}\right)^{\lambda+1} + 1\right] d_t. \end{aligned} \tag{26}$$

Caso 2 - $(1,\lambda)$-EE: $x_t \geq d_t$ que significa que x_t está "longe" da origem.

Neste caso $z > 0$, $x_{i,t+1} > 0$ e consequentemente

VolB_i = Vol$\{x_{i,t+1} \in I \,; x_{i,t+1} > z\} = x_t + d_t - z$.

Logo

$$\begin{aligned} E^{A_t}(|x_{t+1}|) &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{x_t+d_t} z(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= \left[\frac{x_t}{d_t} - \left(\frac{\lambda-1}{\lambda+1}\right)\right] d_t. \end{aligned} \tag{27}$$

Fazendo $\lambda = 2$, temos

$E^{A_t}(|x_{t+1}|) = \left(\frac{1}{3}\right)\left[\left(\frac{x_t}{d_t}\right)^3 + 1\right] d_t$ para $0 \leq x_t < d_t$ e

$$E^{A_t}(|x_{t+1}|) = \left(\frac{x_t}{d_t} - \frac{1}{3}\right) d_t \text{ para } x_t \geq d_t.$$

Este resultado coincide com o obtido em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016), que foi calculado usando a Equação (24) para $\lambda = 2$, e reproduzido no Apêndice D. Dessa forma, conseguimos confirmar que a expressão de $E^{A_t}(|x_{t+1}|)$, definida nas Equações (26) e (27) foi generalizada em relação à λ corretamente.

Para a seleção $(1+\lambda)$-EE, é necessário analisar quando $x_{t+1} = x_t$, ou seja, quando $|x_t| \leq |x_{i,t+1}|\ \forall i = 1, \ldots, \lambda$.

Como $|x_t| = x_t$, então $x_t \leq |x_{i,t+1}| \Leftrightarrow x_{i,t+1} > x_t$ ou $x_{i,t+1} < -x_t$.

Dessa forma, para o caso em que x_t está próximo à origem é preciso verificar se $-x_t > x_t - d_t$ ou $-x_t < x_t - d_t$, isto é, se $x_t < d_t/2$ ou $x_t > d_t/2$.

Considerando$(1+\lambda)$-EE, temos então três casos:

Caso 1 - $(1+\lambda)$**-EE:** $0 \leq x_t < \dfrac{d_t}{2}$

$$\begin{aligned} E^{A_t}(|x_{t+1}|) &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{-x_t} x_t(2z+2d_t)^{\lambda-1}\mathrm{d}z + \int_{-x_t}^{0} -z(2d_t+2z)^{\lambda-1}\mathrm{d}z \right. \\ &\quad + \int_{0}^{x_t} z(2d_t-2z)^{\lambda-1}\mathrm{d}z + \int_{x_t}^{-x_t+d_t} x_t(2d_t-2z)^{\lambda-1}\mathrm{d}z \\ &\quad \left. + \int_{-x_t+d_t}^{x_t+d_t} x_t(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= \left(\frac{1}{\lambda+1}\right)\left[1 - \left(1 - \frac{x_t}{d_t}\right)^{\lambda+1}\right] d_t. \end{aligned} \tag{28}$$

Caso 2 - $(1+\lambda)$**-EE:** $\dfrac{d_t}{2} \leq x_t < d_t$

$$\begin{aligned} E^{A_t}(|x_{t+1}|) &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{0} -z(2z+2d_t)^{\lambda-1}\mathrm{d}z + \int_{0}^{-x_t+d_t} z(2d_t-2z)^{\lambda-1}\mathrm{d}z \right. \\ &\quad \left. + \int_{-x_t+d_t}^{x_t} z(x_t+d_t-z)^{\lambda-1}\mathrm{d}z + \int_{x_t}^{x_t+d_t} x_t(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= \left(\frac{1}{\lambda+1}\right)\left[1 - 2^{-\lambda} + \left(\frac{x_t}{d_t}\right)^{\lambda+1}\right] d_t. \end{aligned} \tag{29}$$

Caso 3 - $(1+\lambda)$**-EE:** $x_t \geq d_t$

$$\begin{aligned} E^{A_t}(|x_{t+1}|) &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{x_t} z(x_t+d_t-z)^{\lambda-1}\mathrm{d}z + \int_{x_t}^{x_t+d_t} x_t(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= \left[\frac{x_t}{d_t} + \left(\frac{1}{1+\lambda}\right)(1-\lambda-2^{-\lambda})\right] d_t. \end{aligned} \tag{30}$$

Como mencionado, a escolha da regra de adaptação do tamanho do passo afeta somente o cálculo de $E^{A_t}(d_{t+1})$ e $E^{A_t}[\ln(d_{t+1})]$.

Para a adaptação dada pela **Regra 2**, temos

$$E^{A_t}(d_{t+1}) = d_t\left(\alpha_f \cdot P_F + \sum_{j=1}^{\lambda} \alpha_j \cdot P_{S_j}\right) = d_t \sum_{j=0}^{\lambda} \alpha_j \cdot P_{S_j} \text{ e}$$

$$E^{A_t}[\ln(d_{t+1})] = \ln(d_t) + \sum_{j=0}^{\lambda} \ln(\alpha_j) \cdot P_{S_j}$$

sendo $P_F = P_{S_0} =$ Probabilidade de Falha (nenhum filho é melhor que o pai), $\alpha_f = \alpha_0$ e $P_{S_j} =$ Probabilidade de Sucesso (probabilidade de exatamente $0 \leq j \leq \lambda$ filhos serem melhores que o pai).

Logo

$$P_{S_j} = \binom{\lambda}{j} P(|x_{i,t+1}| \leq |x_t|)^j P(|x_{i,t+1}| > |x_t|)^{\lambda - j}$$

para $1 \leq i \leq \lambda$ arbitrário, em que $P(\cdot)$ denota a *probabilidade*.

Se $|x_t| = x_t$, então $|x_t| < |x_{i,t+1}| \Leftrightarrow x_t < |x_{i,t+1}| \Leftrightarrow x_{i,t+1} > x_t$ ou $x_{i,t+1} < -x_t$.

Dessa forma é preciso considerar se $-x_t > x_t - d_t$ ou $-x_t < x_t - d_t$, isto é, se $x_t < d_t/2$ ou $x_t > d_t/2$ e, portanto temos os seguintes casos:

Caso 1 - Regra 2 : $0 \leq x_t < \dfrac{d_t}{2}$

$$\begin{aligned} P(|x_{i,t+1}| > |x_t|) &= P(|x_{i,t+1}| > x_t) \\ &= P(x_{i,t+1} > x_t) + P(x_{i,t+1} < -x_t) \\ &= \frac{(x_t + d_t - x_t) + [(-x_t) - (x_t - d_t)]}{2d_t} = \frac{d_t - x_t}{d_t} = 1 - \frac{x_t}{d_t} \end{aligned}$$

$$\Rightarrow P(|x_{i,t+1}| \leq |x_t|) = P(|x_{i,t+1}| \leq x_t) = \frac{x_t}{d_t}.$$

Assim

$$P_{S_j} = \binom{\lambda}{j}\left(\frac{x_t}{d_t}\right)^j \left(1 - \frac{x_t}{d_t}\right)^{\lambda - j}$$

e consequentemente

$$E^{A_t}(d_{t+1}) = d_t \sum_{j=0}^{\lambda} \alpha_j P_{S_j} = d_t \sum_{j=0}^{\lambda} \alpha_j \binom{\lambda}{j} \left(\frac{x_t}{d_t}\right)^j \left(1 - \frac{x_t}{d_t}\right)^{\lambda - j} \tag{31}$$

$$E^{A_t}[\ln(d_{t+1})] = \ln(d_t) + \sum_{j=0}^{\lambda} \ln(\alpha_j) \binom{\lambda}{j} \left(\frac{x_t}{d_t}\right)^j \left(1 - \frac{x_t}{d_t}\right)^{\lambda - j} \tag{32}$$

Caso 2 - Regra 2 : $x_t \geq \dfrac{d_t}{2}$

Neste caso, $-x_t \leq x_t - d_t$, o que implica $P(x_{i,t+1} < -x_t) = 0$.

Assim

$$P(|x_{i,t+1}| > |x_t|) = P(x_{i,t+1} > x_t) = \frac{x_t + d_t - x_t}{2d_t} = \frac{1}{2} = P(|x_{i,t+1}| \leq |x_t|)$$

Logo

$$P_{S_j} = \binom{\lambda}{j} \left(\frac{1}{2}\right)^j \left(\frac{1}{2}\right)^{\lambda - j} = \binom{\lambda}{j} \left(\frac{1}{2}\right)^{\lambda}$$

e consequentemente

$$E^{A_t}(d_{t+1}) = \frac{d_t}{2^{\lambda}} \sum_{j=0}^{\lambda} \alpha_j \binom{\lambda}{j} \tag{33}$$

$$E^{A_t}[\ln(d_{t+1})] = \ln(d_t) + 2^{-\lambda} \sum_{j=0}^{\lambda} \ln(\alpha_j) \binom{\lambda}{j} \tag{34}$$

A adaptação dada pela **Regra 1** é um caso especial da **Regra 2**, com $\alpha_0 = \alpha_f$ e $\alpha_1 = \ldots = \alpha_\lambda = \alpha_s$. Portanto, basta realizar essas substituições nas expressões acima para obter $E^{A_t}(d_{t+1})$ e $E^{A_t}[\ln(d_{t+1})]$ para a **Regra 1**.

Efetuando as substituições, obtemos:

Caso 1 - Regra 1 : $0 \leq x_t < \dfrac{d_t}{2}$

$$E^{A_t}(d_{t+1}) = d_t \left[\left(1 - \frac{x_t}{d_t}\right)^{\lambda} (\alpha_f - \alpha_s) + \alpha_s \right] \tag{35}$$

$$E^{A_t}[\ln(d_{t+1})] = \ln(d_t) + \left[\left(1 - \frac{x_t}{d_t}\right)^{\lambda} \ln\left(\frac{\alpha_f}{\alpha_s}\right) + \ln(\alpha_s) \right] \tag{36}$$

Caso 2 - Regra 1 : $x_t \geq \dfrac{d_t}{2}$

$$E^{A_t}(d_{t+1}) = d_t \left[\frac{1}{2^\lambda}(\alpha_f - \alpha_s) + \alpha_s\right] \tag{37}$$

$$E^{A_t}[\ln(d_{t+1})] = \ln(d_t) + \left[\frac{1}{2^\lambda}\ln\left(\frac{\alpha_f}{\alpha_s}\right) + \ln(\alpha_s)\right] \tag{38}$$

De posse das expressões das esperanças condicionais necessárias e considerando

$$\Psi\left(\frac{|x_t|}{d_t}\right) = \ln\left[\frac{E^{A_t}(|x_{t+1}|) + wE^{A_t}(d_{t+1})}{|x_t| + wd_t}\right] - k(E^{A_t}[\ln(d_{t+1})] - \ln(d_t)),$$

a fim de verificar a existência de parâmetros α_f, α_s (**Regra 1**), α_j, $j = 0, \ldots, \lambda$ (**Regra 2**), w, k e a que satisfaçam à Inequação (23), devemos mostrar que $\Psi(r) \leq -a$, com $r = \frac{|x_t|}{d_t}$.

É possível obter uma expressão analítica para a função $\Psi(r)$, combinando os intervalos obtidos para o cálculo de $E^{A_t}(|x_{t+1}|)$, $E^{A_t}(d_{t+1})$ e $E^{A_t}[\ln(d_{t+1})]$ e as expressões (26) a (38) conforme cada intervalo e cada tipo de seleção.

Sendo a **Regra 1** um caso particular da **Regra 2**, apresentamos aqui a expressão de Ψ apenas para a **Regra 2**.

Articulando os casos obtidos, obtemos a descrição de Ψ para os seguintes intervalos:

(A) $0 \leq x_t < \dfrac{d_t}{2}$;

(B) $\dfrac{d_t}{2} \leq x_t < d_t$;

(C) $x_t \geq d_t$,

Para a seleção $(1,\lambda)$-EE as expressões de Ψ são:

$$\begin{aligned}\Psi_A(r) &= \ln\left[\frac{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1) + w\sum_{j=0}^{\lambda}\alpha_j\binom{\lambda}{j}(1-r)^{\lambda-j}r^j}{w+r}\right] \\ &- k\left[\sum_{j=0}^{\lambda}\ln(\alpha_j)\binom{\lambda}{j}(1-r)^{\lambda-j}r^j\right]\end{aligned} \tag{39}$$

$$\Psi_B(r) = \ln\left[\frac{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1) + \frac{w}{2^\lambda}\sum_{j=0}^{\lambda}\alpha_j\binom{\lambda}{j}}{w+r}\right] - k\frac{1}{2^\lambda}\sum_{j=0}^{\lambda}\ln(\alpha_j)\binom{\lambda}{j} \tag{40}$$

$$\Psi_C(r) = \ln\left[\frac{r - \left(\frac{\lambda-1}{\lambda+1}\right) + \frac{w}{2^\lambda}\sum_{j=0}^{\lambda}\alpha_j\binom{\lambda}{j}}{w+r}\right] - k\frac{1}{2^\lambda}\sum_{j=0}^{\lambda}\ln(\alpha_j)\binom{\lambda}{j} \tag{41}$$

Ou seja, $\Psi : \mathbb{R}^+ \to \mathbb{R}$ é tal que

$$\Psi(r) = \begin{cases} \Psi_A(r) \text{ se } r \in [0, 1/2) \\ \Psi_B(r) \text{ se } r \in [1/2, 1) \\ \Psi_C(r) \text{ se } r \in [1, +\infty) \end{cases}$$

As expressões de Ψ para a seleção $(1+\lambda)$-EE com a **Regra 2** de adaptação são apresentadas no Apêndice E.1. Também são listadas no Apêndice E.2 as expressões de Ψ para ambas seleções com a adaptação dada pela **Regra 1**.

Para mostrar que $\Psi(r) \leq -a$ para todo $r \geq 0$ e algum $a > 0$, é suficiente verificar esta condição nos extremos de cada intervalo A, B e C ($r = 0$, $1/2$, 1 e $r \to +\infty$) e em quaisquer pontos críticos dentro desses intervalos ($r = r^*$ tais que $\Psi'(r^*) = 0$).

É possível observar que, independentemente do esquema de seleção, $\Psi(r)$ é representada em cada intervalo como uma soma de logaritmos de funções racionais, e que $\Psi'(r)$ é representada por uma função racional, na variável r (todos os parâmetros da estratégia, incluindo λ, são fixos). De fato, observamos que para a seleção $(1,2)$-EE, com a **Regra 2**, as derivadas em cada intervalo são dadas por:

$$\begin{aligned} \Psi'_A(r) &= \frac{[2(3\alpha_2 - 6\alpha_1 + 3\alpha_0)r + 6\alpha_1 - 6\alpha_0]w + 3r^2}{r^3 + (3\alpha_2 - 6\alpha_1 + 3\alpha_0)wr^2 + (6\alpha_1 - 6\alpha_0)wr + 3\alpha_0 w + 1} - \frac{1}{w+r} \\ &+ 2kr[-\ln(\alpha_0) + 2\ln(\alpha_1) - \ln(\alpha_2)] + k[2\ln(\alpha_0) - 2\ln(\alpha_1)] \end{aligned} \tag{42}$$

$$\begin{aligned} \Psi'_B(r) &= \frac{12r^2}{4r^3 + 3w(\alpha_0 + 2\alpha_1 + \alpha_2) + 4} - \frac{1}{w+r} \\ &= \frac{8r^3 - 4 - 3w(-4r^2 + \alpha_0 + 2\alpha_1 + \alpha_2)}{(w+r)[4r^3 + 3w(\alpha_0 + 2\alpha_1 + \alpha_2) + 4]} \end{aligned} \tag{43}$$

$$\begin{aligned} \Psi'_C(r) &= \frac{12}{12r - 4 + 3w(\alpha_0 + 2\alpha_1 + \alpha_2)} - \frac{1}{w+r} \\ &= \frac{12w + 4 - 3w(\alpha_0 + 2\alpha_1 + \alpha_2)}{[12r - 4 + 3w(\alpha_0 + 2\alpha_1 + \alpha_2)](w+r)} \end{aligned} \tag{44}$$

Portanto, os pontos críticos dentro de cada intervalo, e os valores correspondentes de Ψ podem ser determinados numericamente para quaisquer valores dados dos parâmetros da estratégia, e o seguinte problema de otimização não-linear restrito pode ser formulado:

$$(\vec{\alpha}^*, w^*, k^*, a^*) = \underset{\vec{\alpha},w,k,a}{\arg\max} \quad a$$

$$\text{sujeito a:} \begin{cases} \vec{\alpha} > \vec{0} \\ \alpha_f < 1 \\ w > 0 \\ 0 \leq k < 1 \\ a > 0 \\ \Psi(r) + a \leq 0, \quad r = 0,\, 1/2,\, 1,\, +\infty \\ \Psi(r^*) + a \leq 0, \quad r^* : \Psi'(r^*) = 0 \end{cases} \tag{45}$$

sendo que $\vec{\alpha} = (\alpha_f, \alpha_s)$ para a adaptação dada pela **Regra 1** e $\vec{\alpha} = (\alpha_0, \ldots, \alpha_\lambda)$ para a **Regra 2**, e a relação $>$ se aplica a cada componente do vetor $\vec{\alpha}$.

Ao obter soluções numéricas para o Problema (45), para cada estratégia de seleção, regra de adaptação e número de filhos (ver Tabelas 1 a 4), verificamos a validade da Desigualdade (2) em cada caso, e consequentemente a demonstração da Etapa 1 da Proposição 9 é concluída para esses casos. Na próxima seção exibimos essas soluções numéricas.

3.2 Configurações de Parâmetros e Limite Superior para Taxa de Convergência

Resolvendo numericamente o Problema (45) para valores de λ dados, obtemos valores dos parâmetros de adaptação do tamanho do passo $\vec{\alpha}$ que minimizam o limite superior da taxa de convergência e^{-a}. Simultaneamente, são determinados valores dos parâmetros da função de Lyapunov e o próprio limite.

Tabela 1 – Valores dos parâmetros ótimos para $(1,\lambda)$-EE com **Regra 1**.

$(1,\lambda)$-EE	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$
α_f	0.72873	0.52779	0.42236	0.36531	0.33489	0.31895
α_s	1.15180	1.19847	1.19591	1.18729	1.18036	1.17567
w	2.51320	1.44610	1.09380	0.90415	0.77973	0.68860
k	0.31395	0.30306	0.29601	0.29158	0.28901	0.28761
a	0.00843	0.02379	0.03370	0.03932	0.04223	0.04362
a/λ	0.00422	0.00793	**0.00842**	0.00786	0.00704	0.00623

Tabela 2 – Valores dos parâmetros ótimos para $(1+\lambda)$-EE com **Regra 1**.

$(1+\lambda)$-EE	$\lambda=1$	$\lambda=2$	$\lambda=3$	$\lambda=4$	$\lambda=5$	$\lambda=6$	$\lambda=7$
α_f	0.75675	0.60518	0.51496	0.46382	0.43603	0.42133	0.41370
α_s	1.89274	1.73948	1.64704	1.59411	1.56531	1.55027	1.54260
w	0.11219	0.10567	0.09626	0.08560	0.07521	0.06598	0.05819
k	0.23988	0.24304	0.24488	0.24591	0.24648	0.24681	0.24701
a	0.04309	0.07039	0.08660	0.09569	0.10060	0.10318	0.10453
a/λ	**0.04309**	0.03518	0.02887	0.02392	0.02012	0.01720	0.01493

Tabela 3 – Valores dos parâmetros ótimos para $(1,\lambda)$-EE com **Regra 2**.

$(1,\lambda)$-EE	$\lambda=2$	$\lambda=3$	$\lambda=4$	$\lambda=5$	$\lambda=6$	$\lambda=7$	$\lambda=8$
α_0	0.71039	0.49155	0.35742	0.27651	0.22974	0.19367	0.17511
α_1	1.0826	0.91243	0.76964	0.65213	0.56452	0.50027	0.45250
α_2	1.3779	1.5546	1.3609	1.1888	1.0424	0.92767	0.83696
α_3		2.1146	2.1679	1.8892	1.6418	1.4786	1.3183
α_4			2.7640	2.6921	2.3571	2.0835	1.8543
α_5				3.2872	3.1068	2.7544	2.4600
α_6					3.5728	3.4852	3.0790
α_7						3.9436	3.6243
α_8							4.8683
w	1.9729	0.93414	0.66790	0.55511	0.49894	0.45210	0.42762
k	0.32744	0.32563	0.33051	0.33903	0.35075	0.35407	0.36404
a	0.011239	0.044260	0.080238	0.11452	0.14564	0.17345	0.19825
a/λ	0.0056195	0.014753	0.020059	0.02290	0.02427	**0.02478**	**0.02478**

Tabela 4 – Valores dos parâmetros ótimos para $(1+\lambda)$-EE com **Regra 2**.

$(1+\lambda)$-EE	$\lambda=1$	$\lambda=2$	$\lambda=3$	$\lambda=4$	$\lambda=5$	$\lambda=6$	$\lambda=7$
α_0	0.75675	0.62312	0.54357	0.48162	0.42880	0.38776	0.35206
α_1	1.89274	1.2143	0.85249	0.65294	0.52762	0.44394	0.38441
α_2		4.6226	2.6358	1.6620	1.1488	0.82451	0.62111
α_3			12.929	6.9717	4.4011	2.9785	2.0902
α_4				24.140	12.705	8.7671	6.4816
α_5					18.461	12.663	9.9223
α_6						13.251	11.697
α_7							13.807
w	0.11219	0.10656	0.10211	0.099812	0.099032	0.098947	0.097633
k	0.23988	0.24322	0.24485	0.24927	0.25613	0.26293	0.26753
a	0.04309	0.087939	0.13401	0.18016	0.22241	0.26086	0.29578
a/λ	0.04309	0.043969	0.04467	**0.04504**	0.044482	0.04348	0.04225

As Tabelas 1 a 4 mostram esses valores obtidos para as quatro estratégias evolutivas variantes, apresentadas pelas combinações das Equações (13) e (14), (15) e (14), (13) e (16) e (15) e (16), de acordo com os valores crescentes de λ. Os resultados foram alcançados usando a programação quadrática sequencial (SQP) numérica do pacote GNU Octave. A partir dessas tabelas, podemos conjectur sobre o número de filhos λ que minimiza o limite superior da taxa de convergência para cada algoritmo $(1 \overset{+}{,} \lambda)$ - EE em relação ao número de avaliações da função objetivo. É intuitivo que um aumento do número de descendentes provoca um aumento nos valores de a. Dessa forma buscamos encontrar λ para o qual o aumento relativo a/λ seja o maior possível, obtendo assim o número de descedentes (ótimo) que maximiza a/λ . Os valores de a/λ se encontram na última linha de cada tabela.

Para a **Regra 1**, as configurações ótimas são $(1{,}4)$-EE e $(1+1)$-EE, enquanto para a **Regra 2** são $(1{,}8)$-EE (marginalmente, em comparação com $(1{,}7)$-EE, e dada a falta de dados para $\lambda > 8$) e $(1+4)$-EE. Comparando ambas as regras, a **Regra 2** melhora o parâmetro de taxa de convergência normalizada a/λ, que poderia ser alcançado com $(1{,}\lambda)$-EE, por quase um fator de 3, enquanto que para $(1+\lambda)$-EE, a melhoria é muito pequena (cerca de 4,5%).

A Figura 2 mostra os gráficos das funções correspondentes $\Psi(r)/\lambda$ para as quatro combinações de estratégia de seleção e regra de adaptação consideradas. Essas funções representam um limite superior (devido à desigualdade de Jensen) no decaimento esperado do processo estocástico V_t por avaliação da função objetivo, em função da distância ao ótimo normalizada, r. Quando necessário, usaremos no decorrer deste trabalho a notação Ψ_λ sempre que desejarmos explicitar a dependência em λ da função Ψ.

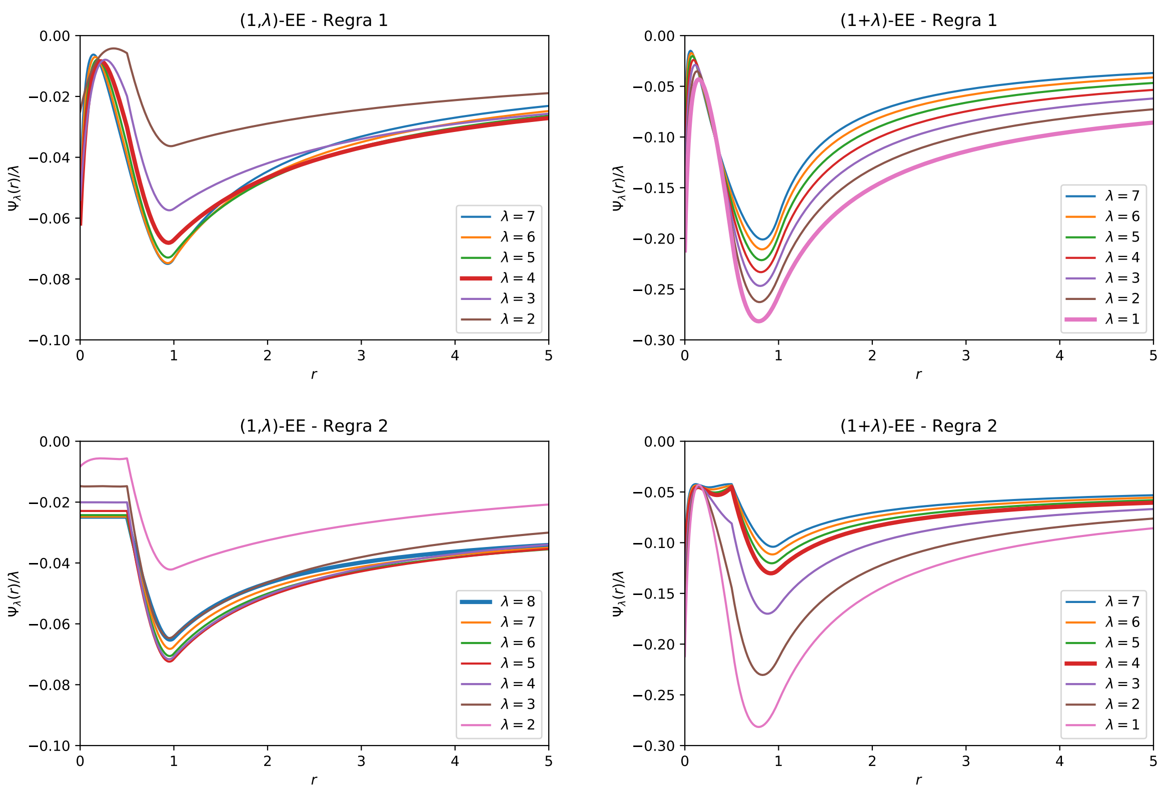

Figura 2 – Funções $\Psi_\lambda(r)/\lambda$ para adaptação baseada em sucesso.

Analisando a Figura 2, duas características dessas funções Ψ_λ merecem destaque. A primeira é que todas as funções correspondentes à **Regra 1**, independentemente do mecanismo de seleção, exibem um pico próximo de $r = 0$, enquanto que as correspondentes à **Regra 2** tendem a exibir uma região quase plana no intervalo $0 \leq r \leq 1/2$ à medida que λ aumenta. Além disso, enquanto o melhor limite[4] para a taxa de convergência para a seleção $(1,\lambda)$-EE com a **Regra 1** corresponde a $\lambda = 4$ para os valores de λ calculados, com a **Regra 2** este limite é muito menor[5] e continua a diminuir, embora muito lentamente, até $\lambda = 8$, pelo menos. Isso sugere que a **Regra 2** em combinação com seleção não elitista permite que a disponibilidade de filhos adicionais seja explorada muito mais efetivamente do que a **Regra 1**, de tal forma que o custo computacional associado seja totalmente compensado por uma convergência mais rápida por geração.

Outra característica importante verificada na forma da função Ψ/λ é que, à medida que λ aumenta, com a seleção $(1,\lambda)$, as curvas tendem a formar um vale em torno de $r = 1$, comparativamente mais profundo, pelo menos até certo ponto. Com a seleção $(1 + \lambda)$, o oposto acontece. Isso sugere que, para $(1 + \lambda)$-EE, as taxas de convergência reais podem aumentar com λ na prática, apesar do fato de que o limite superior normalizado na taxa de convergência realmente diminui (ligeiramente) entre $\lambda = 1$ e $\lambda = 4$. A relação entre os limites da taxa de convergência obtida e as taxas reais de convergência é investigada experimentalmente na próxima seção.

3.3 Resultados Experimentais e Discussões

Considerando a função objetivo $f(x) = |x|$ e usando as configurações que levaram aos melhores limites de convergência para cada combinação de estratégia de seleção e regra de adaptação do tamanho do passo, conforme identificado na última seção, executamos cada EE correspondente 10001 vezes com os valores ótimos de todos os parâmetros. De acordo com a Proposição 4, e para facilitar a comparação, escolhemos o estado inicial (x_0,d_0) dos algoritmos de modo que $V_0 = 0$ em todos os casos. Realizamos dois conjuntos de experimentos, um no qual os algoritmos se iniciam no ponto ótimo $x_0 = 0$ (porém, excluindo esta solução inicial da seleção, mesmo para $(1 + \lambda)$-EE) e outro com $|x_0|/d_0 = 10^{15}$, ou seja, com um tamanho de passo inicial muito pequeno.

Todos os quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t ao longo das gerações para $(1 \overset{+}{,} \lambda)$-EE com **Regra 1** e **Regra 2** podem ser vistos no Apêndice F.1.

A Figura 3 mostra quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t ao longo das avaliações para $(1,4)$-EE com **Regra 1** e $(1,8)$-EE com **Regra 2**, tendo iniciado em $x_0 = 0$,

[4]Maior valor de a/λ.

[5]A taxa de convergência corresponde a $e^{-\bar{a}}$, sendo que $\bar{a}$ é o supremo do conjunto dos valores de a para os quais $|x_t| \leq \exp(-at)$. Dessa forma, um aumento na constante de decaimento exponencial a, provoca uma diminuição no limite da taxa de convergência.

juntamente com os limites de convergência correspondentes.

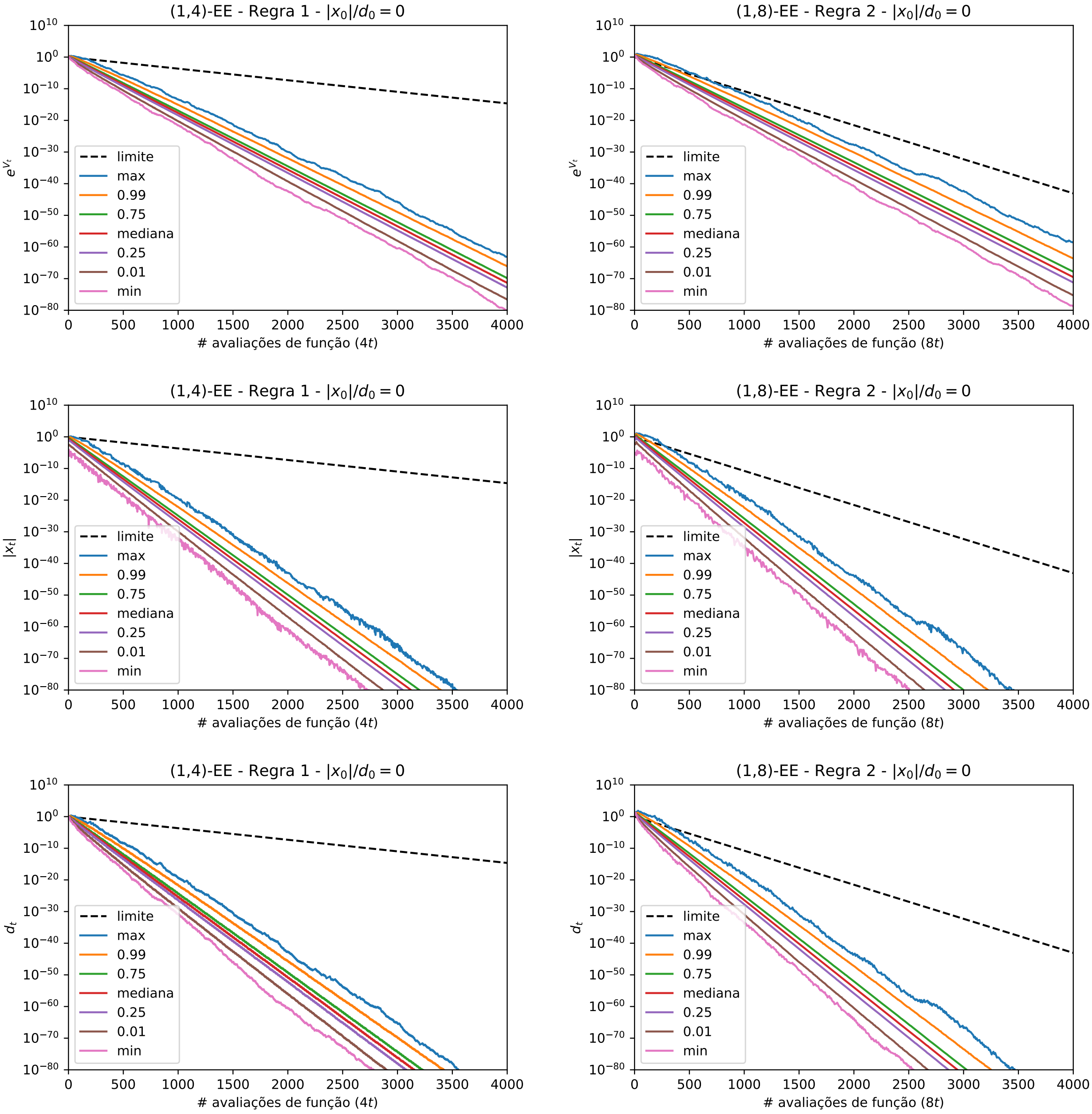

Figura 3 – Quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 1** e **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

A Figura 4 mostra quantis empíricos das distribuições de $|x_t|$ ao longo das gerações para $(1,\lambda)$-EE ($\lambda \in \{2, 3, 4, 5, 6, 7, 8\}$), com a **Regra 2**, tendo iniciado em $x_0 = 0$, juntamente com o limites de convergência teóricos correspondentes.

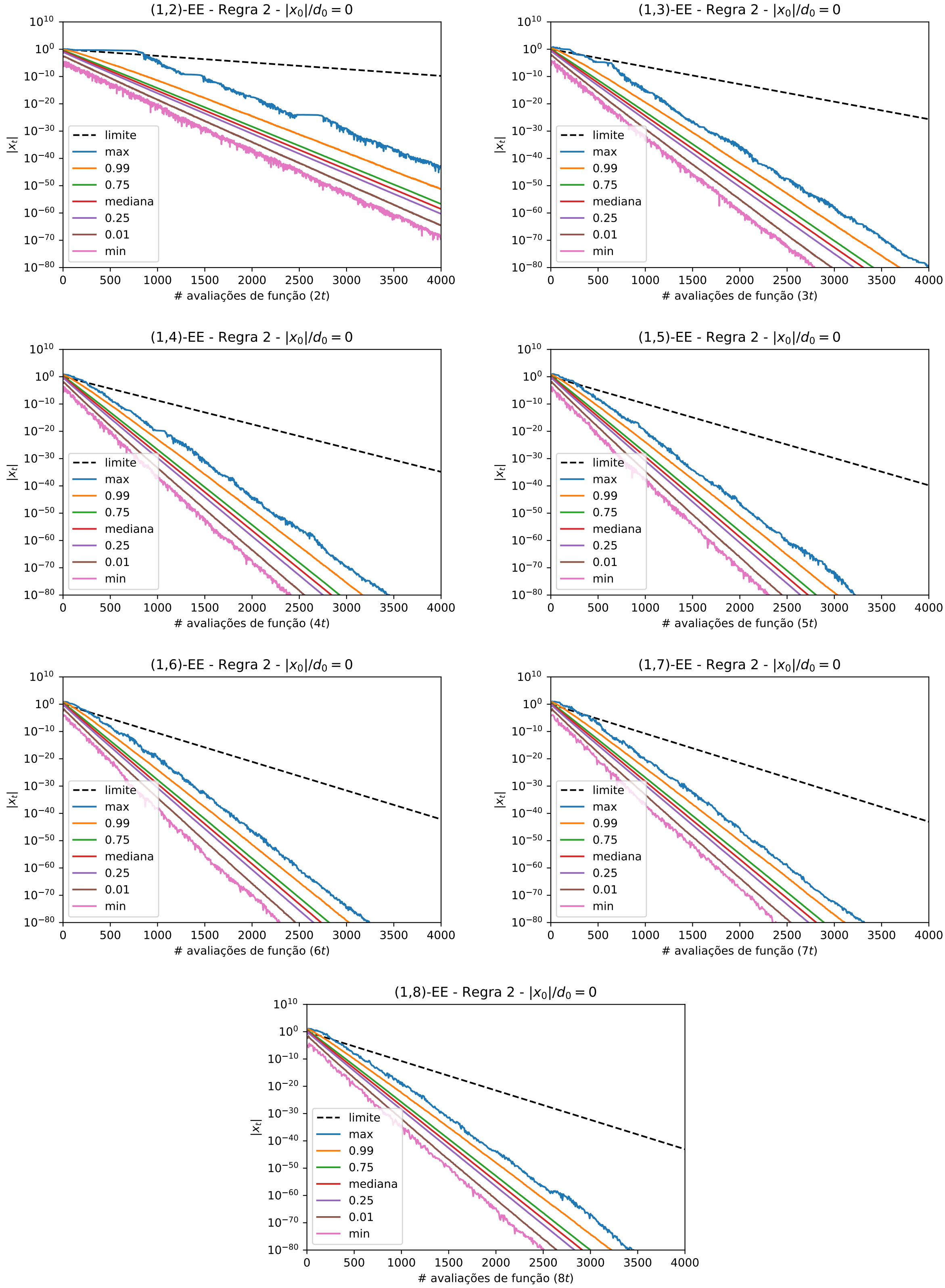

Figura 4 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1{,}\lambda)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Analisando a Figura 3, pode parecer um pouco surpreendente que a **Regra 2** leve a uma convergência ligeiramente mais rápida para o ótimo que a **Regra 1**, apesar de possuir um limite de convergência muito melhor, o que poderia levantar dúvidas sobre os méritos da **Regra 2**, sobre a qualidade dos limites obtidos ou ambos.

Por um lado, observa-se também que o algoritmo $(1{,}6)$-EE com a **Regra 2** converge um pouco mais rápido do que o algoritmo $(1{,}8)$-EE, apesar de possuir um limite ligeiramente pior (ver Figura 4), o que pode ser explicado pelo fato da função $\Psi_6(r)/6$ apresentar um vale mais profundo perto de $r = 1$ do que para $\Psi_8(r)/8$ (Figura 2, esquerda). Por outro lado, isso simplesmente mostra que ambas as estratégias convergem de forma confiável, quando iniciadas perto do ótimo ou, em outras palavras, quando o tamanho do passo inicial é grande. De fato, isso é verdade para todas as estratégias consideradas neste estudo, como mostra a Figura 4 para $(1{,}\lambda)$-EE com a **Regra 2**. Os resultados para a **Regra 1** são semelhantes, mas a convergência é mais lenta para valores idênticos de λ (ver Figura 5).

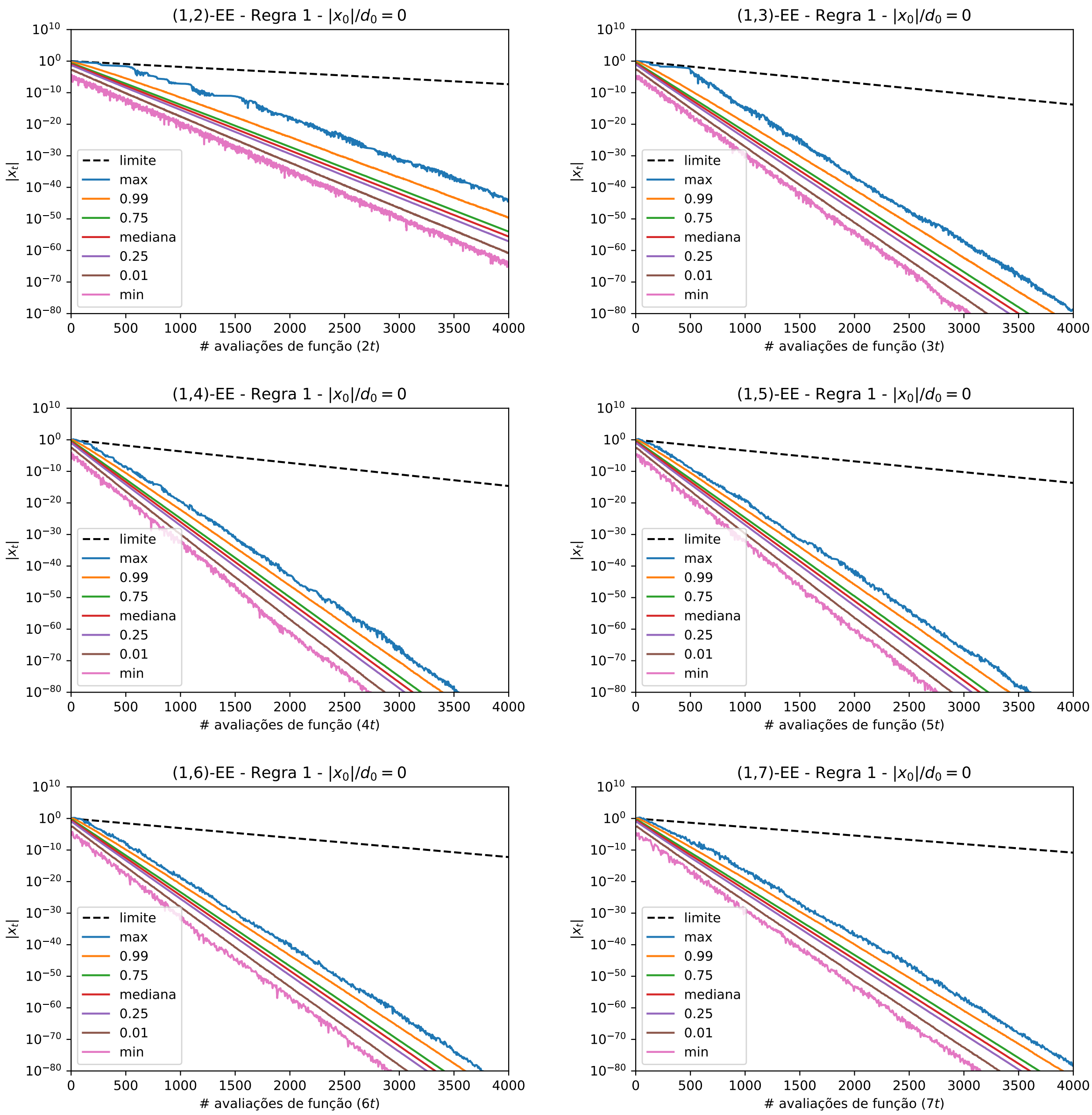

Figura 5 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1{,}\lambda)$-EE com **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

A Figura 6 mostra quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t ao longo das gerações para $(1{,}4)$-EE com **Regra 1** e $(1{,}8)$-EE, com **Regra 2**, tendo iniciado longe do ótimo, juntamente com os limites de convergência correspondentes.

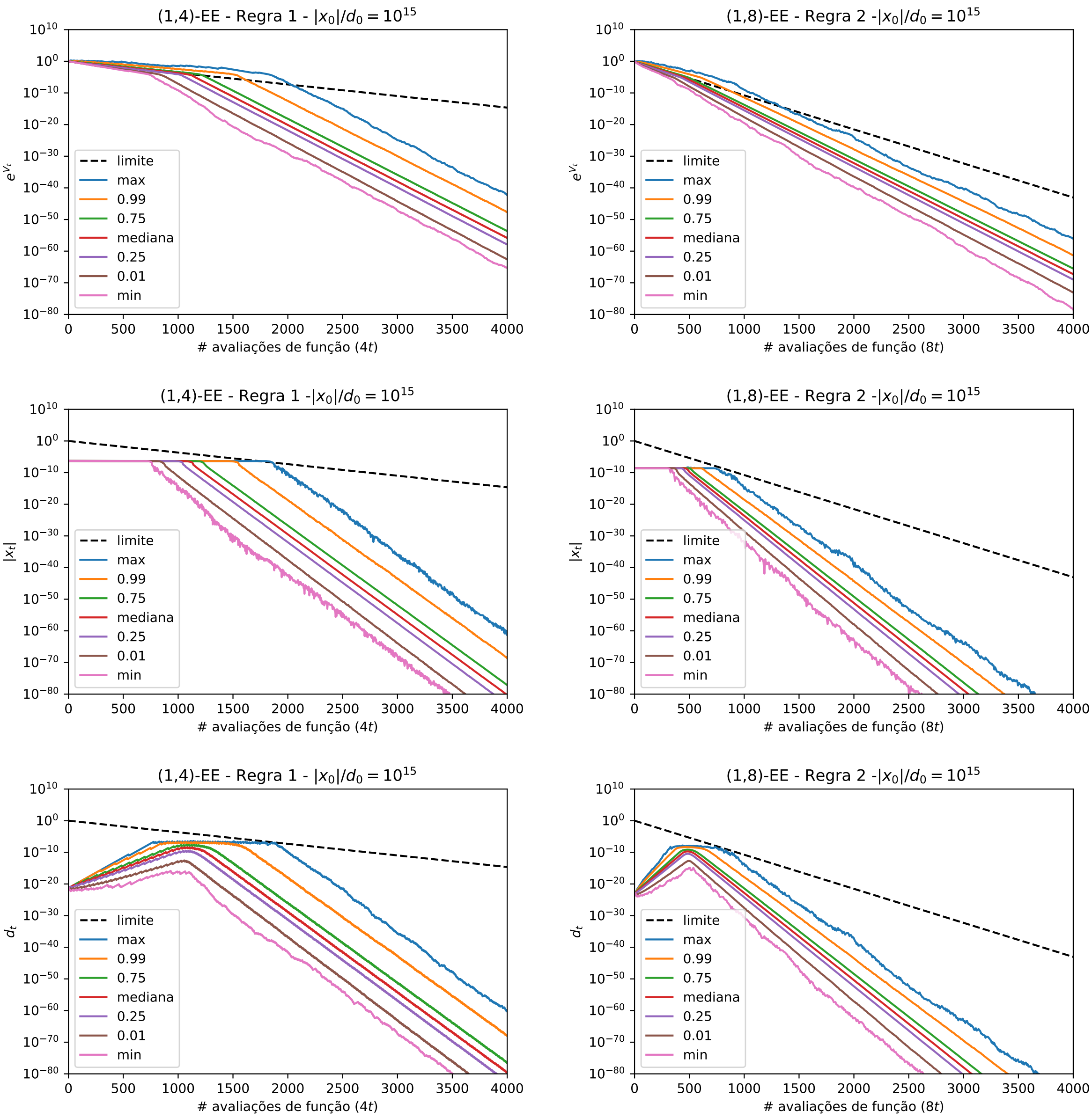

Figura 6 – Quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t estimadas de 10001 execuções de $(1{,}\lambda)$-EE com **Regra 1** e **Regra 2** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

As diferenças de desempenho entre $(1{,}4)$-EE com a **Regra 1** e $(1{,}8)$-EE com a **Regra 2** são muito mais evidentes quando o tamanho do passo inicial é muito pequeno, como mostra a Figura 6. Neste caso, podemos verificar que o tamanho do passo d_t cresce muito mais rápido nos estágios iniciais das execuções para a **Regra 2** do que para a **Regra 1**, levando a uma convergência global mais rápida. As distribuições de e^{V_t} também seguem os limites teóricos mais de perto no início das corridas, até que o tamanho do passo seja suficientemente grande para que o ótimo seja atingido mais rápido.

Os resultados para a seleção elitista são apresentados nas Figuras 7 e 8. Mais uma vez, o limite de convergência ligeiramente melhorado pela **Regra 2** com $\lambda = 4$, em comparação com $(1+1)$-EE para a **Regra 1**, não significa uma convergência mais rápida na prática, possivelmente porque $\Psi_1(r)/1$ possui um vale muito mais profundo perto de $r = 1$ do que $\Psi_4(r)/4$. Além disso, os limites superiores correspondentes $-a/\lambda$ em $\Psi_\lambda(r)/\lambda$ são tão semelhantes que nenhuma melhoria relevante pode ser observada nos estágios iniciais da busca quando o tamanho do passo inicial é muito pequeno (ver Figura 8).

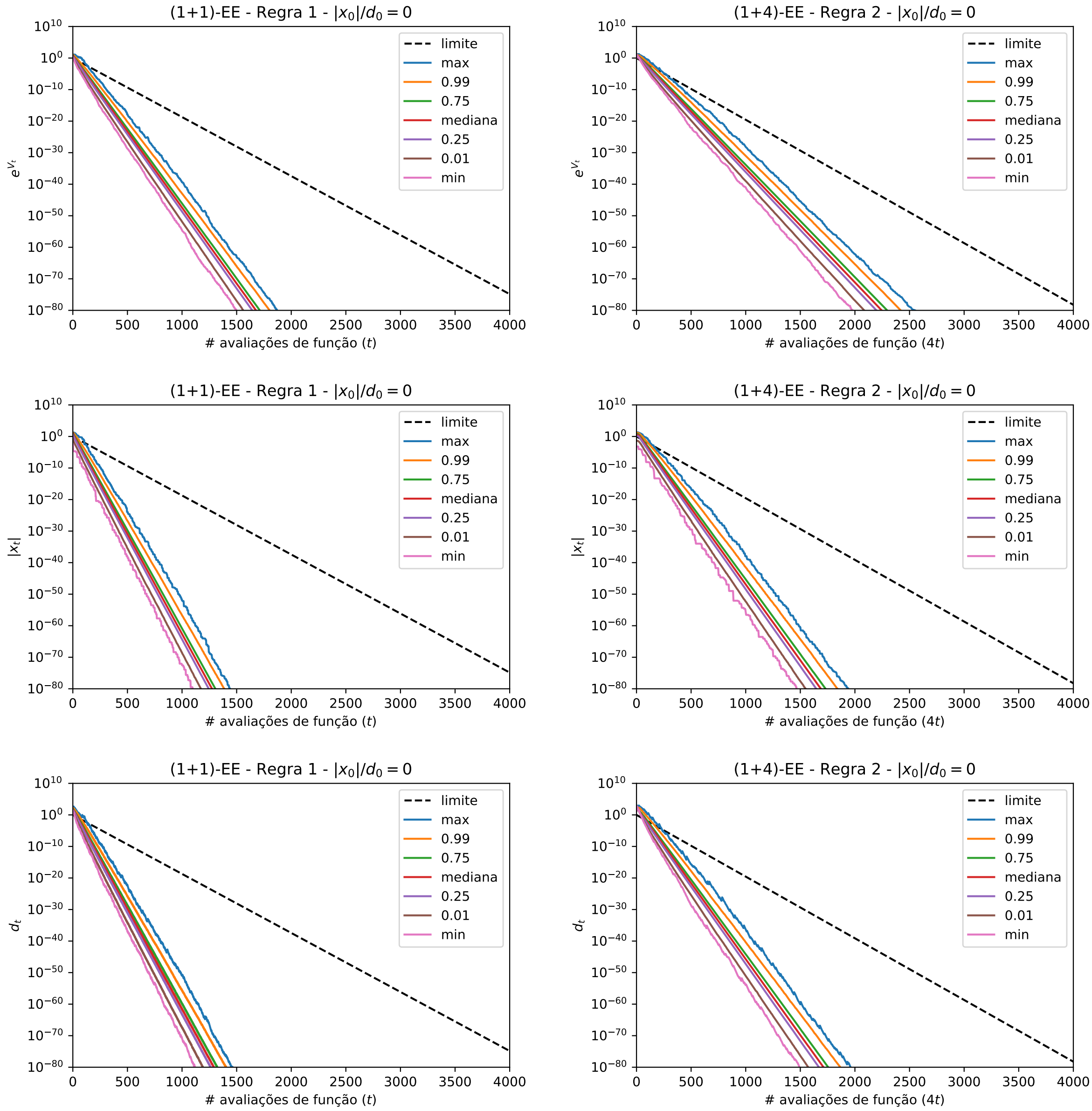

Figura 7 – Quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t estimadas de 10001 execuções de $(1+\lambda)$-ES com **Regra 1** e **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$)

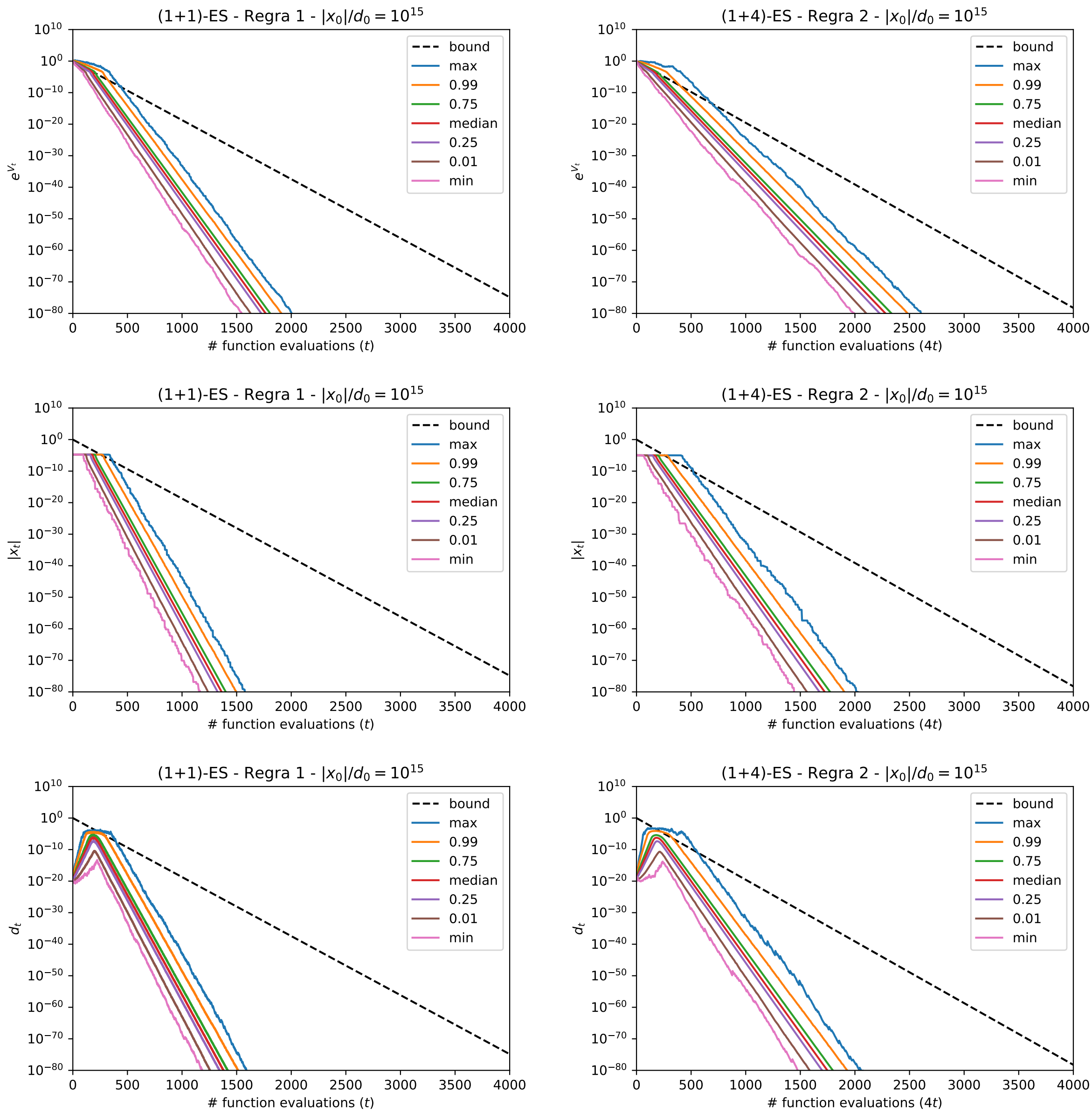

Figura 8 – Quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t estimadas de 10001 execuções de $(1+\lambda)$-ES om **Regra 1** e **Regra 2** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$)

Em todos os experimentos, todas as configurações de algoritmos levam a uma convergência muito mais rápida da distância ao ótimo $|x_t|$ e do tamanho do passo d_t, para zero do que garantido pelos limites teóricos. Tal fato pode ser justificado por quatro razões principais:

- A convergência de V_t é uma condição suficiente para a convergência de $|x_t|$ e d_t, que convergem estritamente mais rápido do que V_t por construção, pois $k > 0$.
- A taxa de convergência assintótica real de V_t está relacionada com $\Psi(r)$ como um todo, e não apenas ao seu valor máximo.
- $\Psi(r)$ foi obtida usando a desigualdade de Jensen para simplificar a análise.

- A otimização numérica do Problema (45) com um método de otimização local, como o SQP, não garante a otimização global das soluções obtidas.

Pelas razões acima, os limites superiores resultantes são considerados conservadores, e selecionar λ com base exclusivamente na maximização de a/λ pode não fornecer os melhores resultados na prática. No entanto, estes limites proporcionam um ponto de partida útil. O principal desafio colocado pela implementação da **Regra 2** de adaptação na prática, antes que a presente análise seja estendida a cenários mais realistas, provavelmente é a escolha apropriada dos $\lambda + 1$ parâmetros de adaptação do tamanho do passo correspondentes.

Embora este estudo seja limitado pelo tipo de mutação e classe de funções consideradas, ainda pode ser estendido, considerando outras funções de uma variável e até funções de múltiplas variáveis, desde que as esperanças condicionais necessárias possam ser computadas. Ainda considerando a mesma classe de funções, no próximo capítulo é realizada a generalização a qualquer número de filhos, da estratégia $(1{,}2)$-EE considerada em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016). Também é proposta a seleção elitista com a mesma adaptação do tamanho do passo. Assim é possivel uma comparação entre as regras de adapatação baseadas em sucesso tratadas neste capítulo com a regra generalizada sugerida em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016).

4 Generalização do Controle de Mutação Baseado no Tamanho do Passo

Segundo Ostermeier, Gawelczyk e Hansen (1994) existem duas maneiras de adaptar o tamanho do passo em EEs. A primeira delas é a regra de sucesso 1/5 de Rechenberg (1973). Nesta regra, o tamanho do passo deve ser aumentado se a proporção de mutações bem sucedidas for maior que 1/5 e deve diminuir, caso contrário. A segunda forma é através do *Controle de Tamanho de Passo Mutativo* proposto por Rechenberg e Schwefel (RECHENBERG, 1973; SCHWEFEL, 1977; SCHWEFEL, 1981), na qual os parâmetros da estratégia são afetados por mutação e seleção.

Em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016), os autores analisam um algoritmo $(1{,}2)$-EE com adaptação do tamanho do passo baseada nas modificações do controle de tamanho de passo mutativo realizadas em (OSTERMEIER; GAWELCZYK; HANSEN, 1994). Nesse algoritmo, os valores absolutos das mutações selecionadas representam fatores de adaptação do tamanho do passo ao serem comparados com um *parâmetro limiar*, que neste caso foi fixado igual a $1/2$. Neste capítulo apresentamos uma generalização deste algoritmo para qualquer número de descendentes $(1{,}\lambda)$-EE e qualquer *limiar*, bem como a estratégia elitista $(1+\lambda)$-EE.

A etapa de seleção dos algoritmos $(1 \overset{+}{,} \lambda)$-EE é realizada da mesma forma como nas estratégias com adaptação do tamanho do passo baseadas em sucesso descritas no capítulo anterior. As equações de estados com esta regra de adaptação generalizada são formalizadas pelas seguintes equações.

$$\begin{aligned} x_{t+1} &= F(x_t, d_t, \mu_{1,t+1}, \ldots, \mu_{\lambda,t+1}) \\ &= \begin{cases} \arg\min_{x \in \{x_{i,t+1} = x_t + \mu_{i,t+1} \cdot d_t,\ i=1,\ldots,\lambda\}} f(x) \text{ para seleção não elitista } (1{,}\lambda) \\ \arg\min_{x \in \{x_{i,t+1} = x_t + \mu_{i,t+1} \cdot d_t,\ i=1,\ldots,\lambda\} \cup \{x_t\}} f(x) \text{ para seleção elitista } (1+\lambda) \end{cases} \end{aligned} \tag{46}$$

$$\begin{aligned} d_{t+1} &= G(x_t, d_t, \mu_{1,t+1}, \ldots, \mu_{\lambda,t+1}) \\ &= \begin{cases} \alpha^{+} \cdot d_t \ \text{ se } \frac{|x_{t+1} - x_t|}{d_t} - \theta \geq 0 \\ \alpha^{-} \cdot d_t \ \text{ caso contrário} \end{cases} \end{aligned} \tag{47}$$

sendo $\mu_{i,t+1}$, $i = 1, \ldots, \lambda$, variáveis aleatórias uniformemente distribuídas em $[-1{,}1]$, $\alpha^{+} > 1$, $0 < \alpha^{-} < 1$ e $0 < \theta < 1$.

A regra de adaptação do tamanho do passo pode ser descrita como segue:

- Se o filho selecionado está longe do pai, sendo a distância entre o pai e o filho pelo menos θd_t, então o tamanho do passo aumenta;
- Se o filho selecionado está perto do pai, sendo a distância entre o pai e o filho menor que θd_t, então o tamanho do passo diminui.

Observamos que

$$\frac{|x_{t+1} - x_t|}{d_t} - \theta = \begin{cases} |\mu_{i,t+1}| - \theta \ \text{ se } x_{i,t+1} \neq x_t \\ -\theta \ \text{ se } x_{i,t+1} = x_t \end{cases}$$

Assim,

$$d_{t+1} = \begin{cases} \alpha^+ \cdot d_t, & \text{se } |\mu_{i,t+1}| \geq \theta \\ \alpha^- \cdot d_t, & \text{se } |\mu_{i,t+1}| < \theta \ \text{ ou se } \ x_{i,t+1} = x_t \end{cases}$$

sendo que $x_{i,t+1}$ é o filho selecionado.

Dessa forma, mutações selecionadas com magnitude maior ou igual a θ resultam em um aumento do tamanho do passo e mutações com magnitude menor que θ geram uma diminuição do tamanho do passo, originando a nomenclatura de tal regra de adaptação.

Além disso, para a seleção elitista $(1+\lambda)$-EE, o controle da mutação baseado no tamanho do passo com $\theta = 0$ se comporta da mesma forma que a adaptação baseada em sucesso dada pela **Regra 1**. De fato, considerando a adaptação baseada em sucesso, temos

$$x_{t+1} = \begin{cases} x_t, & \text{se } |x_t| < |x_{i,t+1}| \ \ \forall\, i \\ x_{i,t+1} = x_t + \mu_{i,t+1} d_t, & \text{caso contrário} \end{cases}$$

Portanto a ocorrência de fracasso implica que o pai é selecionado para compor a nova geração e, consequentemente, não existe a possibilidade do filho selecionado estar longe do pai.

$$\frac{|x_{t+1} - x_t|}{d_t} = \begin{cases} 0, & \text{se } |x_t| < |x_{i,t+1}| \ \ \forall \ i \ \ \text{(fracasso)} \ \ \Rightarrow d_{t+1} = \alpha^- \cdot d_t \\ |\mu_{i,t+1}|, & \text{caso contrário} \ \ \text{(sucesso)} \ \ \Rightarrow d_{t+1} = \alpha^+ \cdot d_t \end{cases}$$

Com a finalidade de padronizar e facilitar a nomenclatura das EEs estudadas, intitulamos a generalização do controle de mutação baseado no tamanho do passo por **Regra 3**.

4.1 Análise da Taxa de Convergência

Utilizando a metodologia para análise de convergência de EEs descrita no final do Capítulo 2, queremos determinar valores dos parâmetros da estratégia que maximizam o limite superior sobre a constante de decaimento exponencial da taxa de convergência para a **Regra 3**. Para tanto nosso propósito é demonstrar um resultado análogo à Proposição 8:

Proposição 10 *Considere o processo estocástico* (x_t,d_t) *definido por qualquer uma das Equações em (46) e pela Equação (47). Se* f *é uma função de uma variável real, estritamente unimodal que é simétrica em relação ao eixo vertical que contém o mínimo, então este processo converge para* $(0,0)$ *quase certamente, e as seguintes inequações*

$$|x_t| \leq \exp(-at) \quad \textit{e} \quad d_t \leq \exp(-at) \tag{48}$$

seguem assintoticamente quase certamente para α^-*,* α^+ *e* $a > 0$*.*

Para a demonstração da Proposição 10, considerando a mesma estrutura de função de Lyapunov estocástica adotada no capítulo anterior, é necessário determinar valores para $0 < \alpha^- < 1,\, \alpha^+ > 1,\, w > 0,\, 0 < k < 1,\; a > 0$ e $b > 0$ tais que:

a) $E^{A_t}(V_{t+1}) \leq V_t - a$.

b) $E^{A_t}[(V_{t+1} - E^{A_t}(V_{t+1}))^2] \leq b$.

c) $V_t \leq -at + o(t^{0,5+\epsilon}) \Rightarrow |x_t| \leq \exp(-at)$ e $d_t \leq \exp(-at)$.

sendo $V_t = \ln(|x_t| + wd_t) - k\ln(d_t)$ e x_t e d_t descritos respectivamente por uma das Equações em (46) e pela Equação (47).

Assim como na prova da Proposição 9, não é necessário demonstrar os itens b) e c). Tais demonstrações são idênticas às apresentadas em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016), considerando $\beta/\alpha = \alpha^-$ e $\beta\alpha = \alpha^+$.

Como a seleção é a mesma dos algoritmos do capítulo anterior, é necessário calcular apenas $E^{A_t}(d_{t+1})$ e $E^{A_t}[\ln(d_{t+1})]$ para obter um limite superior (derivado da desigualdade de Jensen) para $E^{A_t}(V_{t+1}) - V_t$.

Segue da definição do tamanho do passo que

$$E^{A_t}(d_{t+1}) = d_t(\alpha^+ \cdot P^+ + \alpha^- \cdot P^-)$$

considerando:

$P^+ = Prob(|x_{t+1} - x_t| \geq \theta d_t)$ a probabilidade do novo indivíduo selecionado estar longe do pai;

$P^- = Prob(|x_{t+1} - x_t| < \theta d_t)$ a probabilidade do novo indivíduo selecionado estar perto do pai.

Para o cálculo de P^+, algumas observações devem ser levadas em conta:

(i) Como a função a ser otimizada é simétrica em torno do eixo que contém o ótimo, será considerado $x_t > 0$;

(ii) $x_{t+1} \in [x_t - d_t,\, x_t + d_t]$;

(iii) $|x_{t+1} - x_t| \geq \theta d_t \Leftrightarrow x_{t+1} \geq x_t + \theta d_t$ ou $x_{t+1} \leq x_t - \theta d_t$;

(iv) $[x_t - \theta d_t,\, x_t + \theta d_t] \subset [x_t - d_t,\, x_t + d_t]$, pois $0 < \theta < 1$;

(v) $-(x_t - d_t) < x_t + d_t$ e $-(x_t + d_t) < x_t - d_t$.

Dessa forma, P^+ é calculado para vários intervalos, de acordo com as possibilidades de ordenação dos seguintes valores:

$$0,\, x_t - d_t,\, x_t - \theta d_t,\, -x_t + d_t,\, x_t + \theta d_t \text{ e } x_t + d_t$$

Como $x_t > 0$ e $d_t > 0$, então $0 < x_t + \theta d_t < x_t + d_t$, porém, não sabemos o sinal de $x_t - d_t$ e $x_t - \theta d_t$. Considerando $x_t - d_t < 0$, temos $-x_t + d_t > 0$. Neste caso, podemos ter $x_t - \theta d_t < 0$ ou $x_t - \theta d_t > 0$. Se $x_t - \theta d_t < 0$, então resta analisar as possibilidades para $-x_t + d_t$: entre 0 e $x_t + \theta d_t$ ou entre $x_t + \theta d_t$ e $x_t + d_t$.

Se $x_t + \theta d_t < -x_t + d_t < x_t + d_t$, então a seguinte possibilidade de ordenação é estabelecida:

$$x_t - d_t < x_t - \theta d_t < 0 < x_t + \theta d_t < -x_t + d_t < x_t + d_t$$

Neste caso: $\frac{x_t}{d_t} < \theta$ e $\frac{x_t}{d_t} < \frac{1-\theta}{2}$, ou seja, $0 \leq \frac{x_t}{d_t} \leq \min\left\{\frac{1-\theta}{2}, \theta\right\}$, com:

$$\begin{aligned} P^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{x_t-\theta d_t}(2d_t+2z)^{\lambda-1}\mathrm{d}z + \int_{x_t+\theta d_t}^{-x_t+d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z \right. \\ &\quad \left. + \int_{-x_t+d_t}^{x_t+d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= \frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^\lambda + \left(1-\theta-\frac{x_t}{d_t}\right)^\lambda\right] \end{aligned}$$

$$E^{A_t}(d_{t+1}) = d_t\left\{\frac{1}{2}(\alpha^+ - \alpha^-)\left[\left(1-\theta+\frac{x_t}{d_t}\right)^\lambda + \left(1-\theta-\frac{x_t}{d_t}\right)^\lambda\right] + \alpha^-\right\} \quad (49)$$

$$E^{A_t}[\ln(d_{t+1})] = \ln(d_t) + \frac{1}{2}\ln\left(\frac{\alpha^+}{\alpha^-}\right)\left[\left(1-\theta+\frac{x_t}{d_t}\right)^\lambda + \left(1-\theta-\frac{x_t}{d_t}\right)^\lambda\right] + \ln(\alpha^-) \quad (50)$$

Se $0 < -x_t + d_t < x_t + \theta d_t$, temos a seguinte possibilidade de ordenação:

$$x_t - d_t < x_t - \theta d_t < 0 < -x_t + d_t < x_t + \theta d_t < x_t + d_t$$

Neste caso $\frac{1-\theta}{2} \leq \frac{x_t}{d_t} \leq \theta$, com

$$\begin{aligned} P^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{x_t-\theta d_t}(2d_t+2z)^{\lambda-1}\mathrm{d}z + \int_{x_t+\theta d_t}^{x_t+d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= \frac{1}{2}\left[\left(\frac{x_t}{d_t}+1-\theta\right)^\lambda - \left(\frac{x_t}{d_t}\right)^\lambda\right] + \left(\frac{1-\theta}{2}\right)^\lambda \end{aligned}$$

$$E^{A_t}(d_{t+1}) = d_t\left\{(\alpha^+ - \alpha^-)\left[\frac{1}{2}\left(\frac{x_t}{d_t}+1-\theta\right)^\lambda - \frac{1}{2}\left(\frac{x_t}{d_t}\right)^\lambda + \left(\frac{1-\theta}{2}\right)^\lambda\right] + \alpha^-\right\} \quad (51)$$

$$E^{A_t}[\ln(d_{t+1})] = \ln(d_t) + \ln\left(\frac{\alpha^+}{\alpha^-}\right)\left[\frac{1}{2}\left(\frac{x_t}{d_t}+1-\theta\right)^\lambda - \frac{1}{2}\left(\frac{x_t}{d_t}\right)^\lambda + \left(\frac{1-\theta}{2}\right)^\lambda\right] + \ln(\alpha^-) \quad (52)$$

Se $x_t - \theta d_t > 0$, novamente podemos ter $-x_t + d_t$ entre $x_t + \theta d_t$ e $x_t + d_t$ ou entre $x_t - \theta d_t$ e $x_t + \theta d_t$, ou ainda, entre 0 e $x_t - \theta d_t$.

Se $x_t + \theta d_t < -x_t + d_t < x_t + d_t$, temos a seguinte ordenação:

$$x_t - d_t < 0 < x_t - \theta d_t < x_t + \theta d_t < -x_t + d_t < x_t + d_t$$

Neste caso $\theta \leq \frac{x_t}{d_t} \leq \frac{1-\theta}{2}$, com:

$$\begin{aligned} P^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{0}(2d_t+2z)^{\lambda-1}\mathrm{d}z + \int_{0}^{x_t-\theta d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z\right. \\ &\quad \left. + \int_{x_t+\theta d_t}^{-x_t+d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z + \int_{-x_t+d_t}^{x_t+d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= 1 - \frac{1}{2}\left[\left(1+\theta-\frac{x_t}{d_t}\right)^\lambda - \left(1-\theta-\frac{x_t}{d_t}\right)^\lambda\right] \end{aligned}$$

$$E^{A_t}(d_{t+1}) = d_t\left\{\frac{1}{2}\left[\left(1+\theta-\frac{x_t}{d_t}\right)^\lambda(\alpha^- - \alpha^+) + \left(1-\theta-\frac{x_t}{d_t}\right)^\lambda(\alpha^+ - \alpha^-)\right] + \alpha^+\right\} \quad (53)$$

$$\begin{aligned} E^{A_t}[\ln(d_{t+1})] &= \ln(d_t) + \frac{1}{2}\left[\ln\left(\frac{\alpha^-}{\alpha^+}\right)\left(1+\theta-\frac{x_t}{d_t}\right)^\lambda\right. \\ &\quad \left. + \ln\left(\frac{\alpha^+}{\alpha^-}\right)\left(1-\theta-\frac{x_t}{d_t}\right)^\lambda\right] + \ln(\alpha^+) \end{aligned} \quad (54)$$

Se $x_t - \theta d_t < -x_t + d_t < x_t + \theta d_t$, então observamos a seguinte ordenação:

$$x_t - d_t < 0 < x_t - \theta d_t < -x_t + d_t < x_t + \theta d_t < x_t + d_t$$

Neste caso $\max\left\{\frac{1-\theta}{2}, \theta\right\} \leq \frac{x_t}{d_t} \leq \frac{\theta+1}{2}$ com:

$$\begin{aligned} P^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{0}(2d_t+2z)^{\lambda-1}\mathrm{d}z + \int_{0}^{x_t-\theta d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z \right. \\ &\quad \left. + \int_{x_t+\theta d_t}^{x_t+d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= 1 + \left(\frac{1-\theta}{2}\right)^\lambda - \frac{1}{2}\left[\left(\theta+1-\frac{x_t}{d_t}\right)^\lambda + \left(\frac{x_t}{d_t}\right)^\lambda\right] \end{aligned}$$

$$E^{A_t}(d_{t+1}) = d_t\left\{\frac{1}{2}\left[\left(\theta+1-\frac{x_t}{d_t}\right)^\lambda + \left(\frac{x_t}{d_t}\right)^\lambda\right](\alpha^- - \alpha^+) + \left(\frac{1-\theta}{2}\right)^\lambda(\alpha^+ - \alpha^-) + \alpha^+\right\} \quad (55)$$

$$\begin{aligned} E^{A_t}[\ln(d_{t+1})] &= \ln(d_t) + \frac{1}{2}\ln\left(\frac{\alpha^-}{\alpha^+}\right)\left[\left(\theta+1-\frac{x_t}{d_t}\right)^\lambda + \left(\frac{x_t}{d_t}\right)^\lambda\right] \\ &\quad + \ln\left(\frac{\alpha^+}{\alpha^-}\right)\left(\frac{1-\theta}{2}\right)^\lambda + \ln(\alpha^+) \end{aligned} \quad (56)$$

Se $0 < -x_t + d_t < x_t - \theta d_t$, então verificamos a seguinte ordenação:

$$x_t - d_t < 0 < -x_t + d_t < x_t - \theta d_t < x_t + \theta d_t < x_t + d_t$$

Neste caso $\frac{1+\theta}{2} \leq \frac{x_t}{d_t} \leq 1$, com:

$$\begin{aligned} P^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{0}(2d_t+2z)^{\lambda-1}\mathrm{d}z + \int_{0}^{-x_t+d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z \right. \\ &\quad \left. + \int_{-x_t+d_t}^{x_t-\theta d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z + \int_{x_t+\theta d_t}^{x_t+d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= 1 - \left(\frac{1+\theta}{2}\right)^\lambda + \left(\frac{1-\theta}{2}\right)^\lambda \end{aligned}$$

$$E^{A_t}(d_{t+1}) = d_t\left[\left(\frac{1+\theta}{2}\right)^\lambda(\alpha^- - \alpha^+) + \left(\frac{1-\theta}{2}\right)^\lambda(\alpha^+ - \alpha^-) + \alpha^+\right] \quad (57)$$

$$E^{A_t}[\ln(d_{t+1})] = \ln(d_t) + \ln\left(\frac{\alpha^-}{\alpha^+}\right)\left(\frac{1+\theta}{2}\right)^\lambda + \ln\left(\frac{\alpha^+}{\alpha^-}\right)\left(\frac{1-\theta}{2}\right)^\lambda + \ln(\alpha^+) \quad (58)$$

Se $x_t - d_t > 0$, a única possibilidade de ordenação é:

$$-x_t + d_t < 0 < x_t - d_t < x_t - \theta d_t < x_t + \theta d_t < x_t + d_t$$

Neste caso $\frac{x_t}{d_t} \geq 1$, e:

$$\begin{aligned} P^{+} &= \frac{\lambda}{(2d_t)^{\lambda}} \left[\int_{x_t - d_t}^{x_t - \theta d_t} (x_t + d_t - z)^{\lambda - 1} \mathrm{d}z + \int_{x_t + \theta d_t}^{x_t + d_t} (x_t + d_t - z)^{\lambda - 1} \mathrm{d}z \right] \\ &= 1 - \left(\frac{1+\theta}{2} \right)^{\lambda} + \left(\frac{1-\theta}{2} \right)^{\lambda} \end{aligned}$$

que coincide com o resultado do intervalo anterior.

Ressaltamos que as expressões de $E^{A_t}(d_{t+1})$ dependem não só dos parâmetros α^- e α^+, mas também do limiar θ.

Como $\min\left\{\frac{1-\theta}{2},\theta\right\} = \frac{1-\theta}{2} \Leftrightarrow \theta > \frac{1}{3}$ e $\min\left\{\frac{1-\theta}{2},\theta\right\} = \theta \Leftrightarrow \theta < \frac{1}{3}$, a representação analítica de

$$\Psi(r) = \ln\left[\frac{E^{A_t}(|x_{t+1}|) + wE^{A_t}(d_{t+1})}{|x_t| + wd_t}\right] - k(E^{A_t}[\ln(d_{t+1})] - \ln(d_t)), \text{ sendo } r = |x_t|/d_t$$

pode ser obtida combinando as expressões de $E^{A_t}(|x_{t+1}|)$ dadas pelas Equações (26) a (30) (para as estratégias $(1 \overset{+}{,} \lambda)$-EE) com as expressões de $E^{A_t}(d_{t+1})$ e $E^{A_t}[\ln(d_{t+1})]$ dadas pelas Equações (49) a (58).

Para a seleção $(1,\lambda)$-EE, Ψ é representada por cinco expressões conforme os intervalos indicados na Tabela 5.

Tabela 5 – Intervalos para representação de $\Psi(r)$ de acordo com os valores de θ para $(1,\lambda)$-EE com **Regra 3**.

Intervalos	**Caso 1:** $\theta < 1/3$	**Caso 2:** $\theta > 1/3$
(A)	$0 \leq r < \theta$	$0 \leq r < \frac{1-\theta}{2}$
(B)	$\theta \leq r < \frac{1-\theta}{2}$	$\frac{1-\theta}{2} \leq r < \theta$
(C)	$\frac{1-\theta}{2} \leq r < \frac{1+\theta}{2}$	$\theta \leq r < \frac{1+\theta}{2}$
(D)	$\frac{1+\theta}{2} \leq r < 1$	$\frac{1+\theta}{2} \leq r < 1$
(E)	$r \geq 1$	$r \geq 1$

Apenas para o intervalo (B), Ψ apresenta expressões distintas para cada um dos casos, denotados por B_1 e B_2.

$$\begin{aligned}\Psi_A(r) &= \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}) + \frac{w}{2}(\alpha^+ - \alpha^-)\left[(1-\theta+r)^\lambda + (1-\theta-r)^\lambda\right] + w\alpha^-}{w+r}\right\} \\ &- k\left\{\frac{1}{2}[(1-\theta+r)^\lambda + (1-\theta-r)^\lambda][\ln(\alpha^+) - \ln(\alpha^-)] + \ln(\alpha^-)\right\}\end{aligned} \tag{59}$$

$$\begin{aligned}\Psi_{B_1}(r) &= \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}) + \frac{w}{2}(\alpha^- - \alpha^+)\left[(1+\theta-r)^\lambda - (1-\theta-r)^\lambda\right] + w\alpha^+}{w+r}\right\} \\ &- k\left\{\frac{1}{2}[(1+\theta-r)^\lambda - (1-\theta-r)^\lambda][\ln(\alpha^-) - \ln(\alpha^+)] + \ln(\alpha^+)\right\}\end{aligned} \tag{60}$$

$$\begin{aligned}\Psi_{B_2}(r) &= \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}) + \frac{w}{2}(\alpha^+ - \alpha^-)\left[(1-\theta+r)^\lambda - r^\lambda + 2\left(\frac{1-\theta}{2}\right)^\lambda\right] + w\alpha^-}{w+r}\right\} \\ &- k\left\{\frac{1}{2}\left[(1-\theta+r)^\lambda - r^\lambda + 2\left(\frac{1-\theta}{2}\right)^\lambda\right][\ln(\alpha^+) - \ln(\alpha^-)] + \ln(\alpha^-)\right\}\end{aligned} \tag{61}$$

$$\begin{aligned}\Psi_C(r) &= \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}) + \frac{w}{2}(\alpha^- - \alpha^+)\left[(1+\theta-r)^\lambda + r^\lambda - 2\left(\frac{1-\theta}{2}\right)^\lambda\right] + w\alpha^+}{w+r}\right\} \\ &- k\left\{\frac{1}{2}\left[(1+\theta-r)^\lambda + r^\lambda - 2\left(\frac{1-\theta}{2}\right)^\lambda\right][\ln(\alpha^-) - \ln(\alpha^+)] + \ln(\alpha^+)\right\}\end{aligned} \tag{62}$$

$$\begin{aligned}\Psi_D(r) &= \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}) + w(\alpha^- - \alpha^+)\left[\left(\frac{1+\theta}{2}\right)^\lambda - \left(\frac{1-\theta}{2}\right)^\lambda\right] + w\alpha^+}{w+r}\right\} \\ &- k\left\{\left[\left(\frac{1+\theta}{2}\right)^\lambda - \left(\frac{1-\theta}{2}\right)^\lambda\right][\ln(\alpha^-) - \ln(\alpha^+)] + \ln(\alpha^+)\right\}\end{aligned} \tag{63}$$

$$\Psi_E(r) \;=\; \ln\left\{\frac{r-\left(\frac{\lambda-1}{\lambda+1}\right)+w(\alpha^- - \alpha^+)\left[\left(\frac{1+\theta}{2}\right)^\lambda - \left(\frac{1-\theta}{2}\right)^\lambda\right]+w\alpha^+}{w+r}\right\}$$
$$- \; k\left\{\left[\left(\frac{1+\theta}{2}\right)^\lambda - \left(\frac{1-\theta}{2}\right)^\lambda\right][\ln(\alpha^-)-\ln(\alpha^+)]+\ln(\alpha^+)\right\} \qquad (64)$$

Para a seleção $(1+\lambda)$-EE, é necessário analisar três casos distintos, para os quais Ψ é representada para seis intervalos conforme a Tabela 6.

Tabela 6 – Intervalos para representação de $\Psi(r)$ de acordo com os valores de θ para $(1+\lambda)$-EE com **Regra 3**.

Intervalos	**Caso 1**: $\theta < \frac{1}{3}$	**Caso 2**: $\frac{1}{3} < \theta < \frac{1}{2}$	**Caso 3**: $\theta > \frac{1}{2}$
(A)	$0 \le r < \theta$	$0 \le r < \frac{1-\theta}{2}$	$0 \le r < \frac{1-\theta}{2}$
(B)	$\theta \le r < \frac{1-\theta}{2}$	$\frac{1-\theta}{2} \le r < \theta$	$\frac{1-\theta}{2} \le r < \frac{1}{2}$
(C)	$\frac{1-\theta}{2} \le r < \frac{1}{2}$	$\theta \le r < \frac{1}{2}$	$\frac{1}{2} \le r < \theta$
(D)	$\frac{1}{2} \le r < \frac{1+\theta}{2}$	$\frac{1}{2} \le r < \frac{1+\theta}{2}$	$\theta \le r < \frac{1+\theta}{2}$
(E)	$\frac{1+\theta}{2} \le r < 1$	$\frac{1+\theta}{2} \le r < 1$	$\frac{1+\theta}{2} \le r < 1$
(F)	$r \ge 1$	$r \ge 1$	$r \ge 1$

Apenas para os intervalos (B) e (C), Ψ apresenta expressões distintas, sendo Ψ_B igual para os casos 2 e 3 e Ψ_C igual para os casos 1 e 2. Todas as expressões de Ψ para a seleção $(1+\lambda)$-EE com a **Regra 3** são apresentadas no Apêndice E.3.

Para mostrar que $\Psi(r) \le -a$ para todo $r \ge 0$ e algum $a > 0$, é suficiente verificar esta condição nos extremos de cada intervalo A, B, C, D,E e F ($r = 0$, $\frac{1}{2}$, θ, $\frac{1-\theta}{2}$, $\frac{1+\theta}{2}$, 1 e $r \to +\infty$) e em quaisquer pontos críticos dentro desses intervalos ($r = r^*$ tais que $\Psi'(r^*) = 0$).

Assim como na adaptação baseada em sucesso, $\Psi(r)$ é representada em cada intervalo como uma soma de logaritmos de funções racionais para ambas seleções $(1 \overset{+}{,} \lambda)$-EE. Além disso $\Psi'(r)$ é representada por uma função racional, na variável r. Portanto, os pontos críticos dentro de cada intervalo, e os valores correspondentes de Ψ podem ser determinados numericamente para quaisquer valores dados dos parâmetros da estratégia, e o seguinte problema de otimização não-linear restrito pode ser formulado:

$$((\alpha^-)^*,(\alpha^+)^*,\theta^*,w^*,k^*,a^*) = \arg\max_{\alpha^-,\alpha^+,\theta,w,k,a} \quad a$$

$$\text{sujeito a:} \begin{cases} 0 < \alpha^- < 1 \\ \alpha^+ > 1 \\ 0 < \theta < 1 \\ w > 0 \\ 0 \leq k < 1 \\ a > 0 \\ \Psi(r) + a \leq 0, \quad r = 0,\, \frac{1}{2},\, \theta,\, \frac{1-\theta}{2},\, \frac{1+\theta}{2},\, 1,\, +\infty \\ \Psi(r^*) + a \leq 0, \quad r^* : \Psi'(r^*) = 0 \end{cases} \tag{65}$$

As Tabelas 7 e 8 apresentam os valores ótimos dos parâmetros de adaptação do controle de mutação e dos parâmetros da função de Lyapunov que minimizam o limite superior para a taxa de convergência e^{-a}, obtidos pela resolução numérica do Problema (65) (utilizando a programação quadrática sequencial (SQP) numérica do pacote GNU Octave). A partir dessas tabelas, podemos admitir que o número de filhos λ que maximiza a/λ para a seleção $(1,\lambda)$-EE é $\lambda = 6$ (marginalmente, em comparação com $(1,5)$-EE, e dada a falta de dados para $\lambda > 6$) e para a seleção $(1+\lambda)$-EE é $\lambda = 1$.

Tabela 7 – Valores dos parâmetros ótimos para $(1,\lambda)$-EE com **Regra 3**.

$(1,\lambda)$-EE	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$
α^-	0.90330	0.71288	0.54538	0.43002	0.35521
α^+	1.30280	1.6649	1.9202	2.0883	2.2082
θ	0.68779	0.60419	0.56534	0.54857	0.54368
w	10.7030	2.3367	1.1752	0.80462	0.63608
k	0.34305	0.35426	0.36068	0.36268	0.36274
a	0.0043339	0.027859	0.065973	0.10789	0.14731
a/λ	0.0021669	0.009286	0.016493	0.021578	0.024552

Tabela 8 – Valores dos parâmetros ótimos para $(1+\lambda)$-EE com **Regra 3**.

$(1+\lambda)$-EE	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$
α^-	0.80939	0.66699	0.55945	0.47598	0.40740	0.35506
α^+	5.57060	5.6834	5.8587	5.9980	5.8753	5.7372
θ	0.46320	0.48515	0.50723	0.52535	0.53078	0.53903
w	0.085053	0.094717	0.10455	0.11387	0.12126	0.12686
k	0.23398	0.24658	0.25834	0.26930	0.28003	0.29046
a	0.071659	0.13713	0.19702	0.25156	0.29957	0.33962
a/λ	0.071659	0.068565	0.065673	0.06289	0.059914	0.056603

Executando o algoritmo com θ fixo igual a $1/2$ verificamos que os parâmetros ótimos obtidos para $(1{,}2)$-EE são os mesmos encontrados em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016).

A Figura 9 mostra os gráficos das funções correspondentes $\Psi_\lambda(r)/\lambda$ para as duas estratégias de seleção e regra de adaptação do controle de mutação baseado no tamanho do passo. Essas funções representam um limite superior (devido à desigualdade de Jensen) no decaimento esperado do processo estocástico V_t por avaliação da função objetivo, em função da distância ao ótimo normalizada, r.

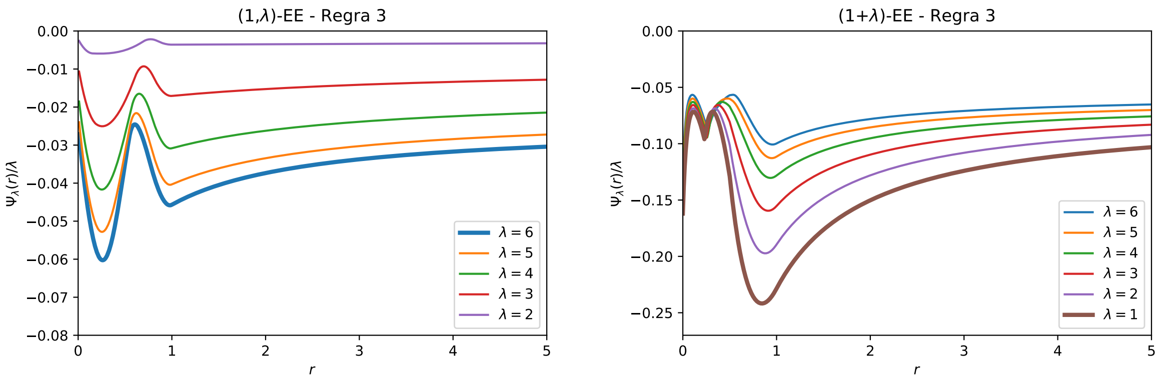

Figura 9 – Funções $\Psi_\lambda(r)/\lambda$ para controle de mutação baseado no tamanho do passo.

Analisando a Figura 9, observamos que para a seleção $(1{,}\lambda)$-EE, $\Psi_\lambda(r)/\lambda$ decresce até um certo valor $r \in (0{,}1)$. À medida que λ aumenta, as curvas de $\Psi_\lambda(r)/\lambda$ tendem a formar um vale próximo a esse mesmo valor, diferentemente do que ocorre com a mesma seleção com a **Regra 1**, na qual $\Psi_\lambda(r)/\lambda$ apresenta um pico mais ou menos próximo de $r = 0$.

Uma característica importante verificada na forma da função $\Psi_\lambda(r)/\lambda$ é que, à medida que λ aumenta, com a seleção $(1{,}\lambda)$-EE, as curvas tendem a formar um vale raso em torno de $r = 1$. Com a seleção $(1 + \lambda)$-EE, esses vales são mais profundos, indicando que, para $(1 + \lambda)$-EE, as taxas de convergência reais aumentam com λ, apesar do limite superior da taxa normalizada diminuir.

4.2 Resultados Experimentais e Discussões

Nesta seção apresentamos alguns experimentos resultantes de 10001 execuções das estatégias evolutivas $(1 \overset{+}{,} \lambda)$-EE com controle de mutação baseado no tamanho do passo. Estes resultados são comparados com os resultados equivalentes para as regras de adaptação do tamanho do passo baseadas em sucesso exibidas no capítulo anterior. Assim como no Capítulo 3, os algoritmos foram executados com os parâmetros ótimos exibidos nas Tabelas 7 e 8 e considerando $f(x) = |x|$ como função objetivo.

Todos os quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t ao longo das gerações para $(1 \overset{+}{,} \lambda)$-EE com **Regra 3** podem ser vistos no Apêndice F.2.

Analisando as Figuras 10 e 11, observamos que $(1{,}2)$-EE, tanto com a **Regra 1**, quanto com a **Regra 2**, converge mais rápido que $(1{,}2)$-EE com **Regra 3**. Estes resultados validam os limites teóricos obtidos.

Para $\lambda > 2$ a adaptação baseada no tamanho do passo converge mais rápido que a **Regra 1** (comparar Figuras 35 e 59 do Apêndice F). Porém, para $\lambda = 3, 4, 5$, apesar do limite teórico apresentar uma taxa de convergência maior para a **Regra 2**, os experimentos indicam uma convergência mais rápida para a **Regra 3** (ver Figura 12). Nos experimentos em que o tamanho do passo inicial é muito pequeno, o mesmo comportamento é observado e à medida que λ aumenta fica mais evidente a convergência mais rápida da **Regra 3** (ver Figura 13).

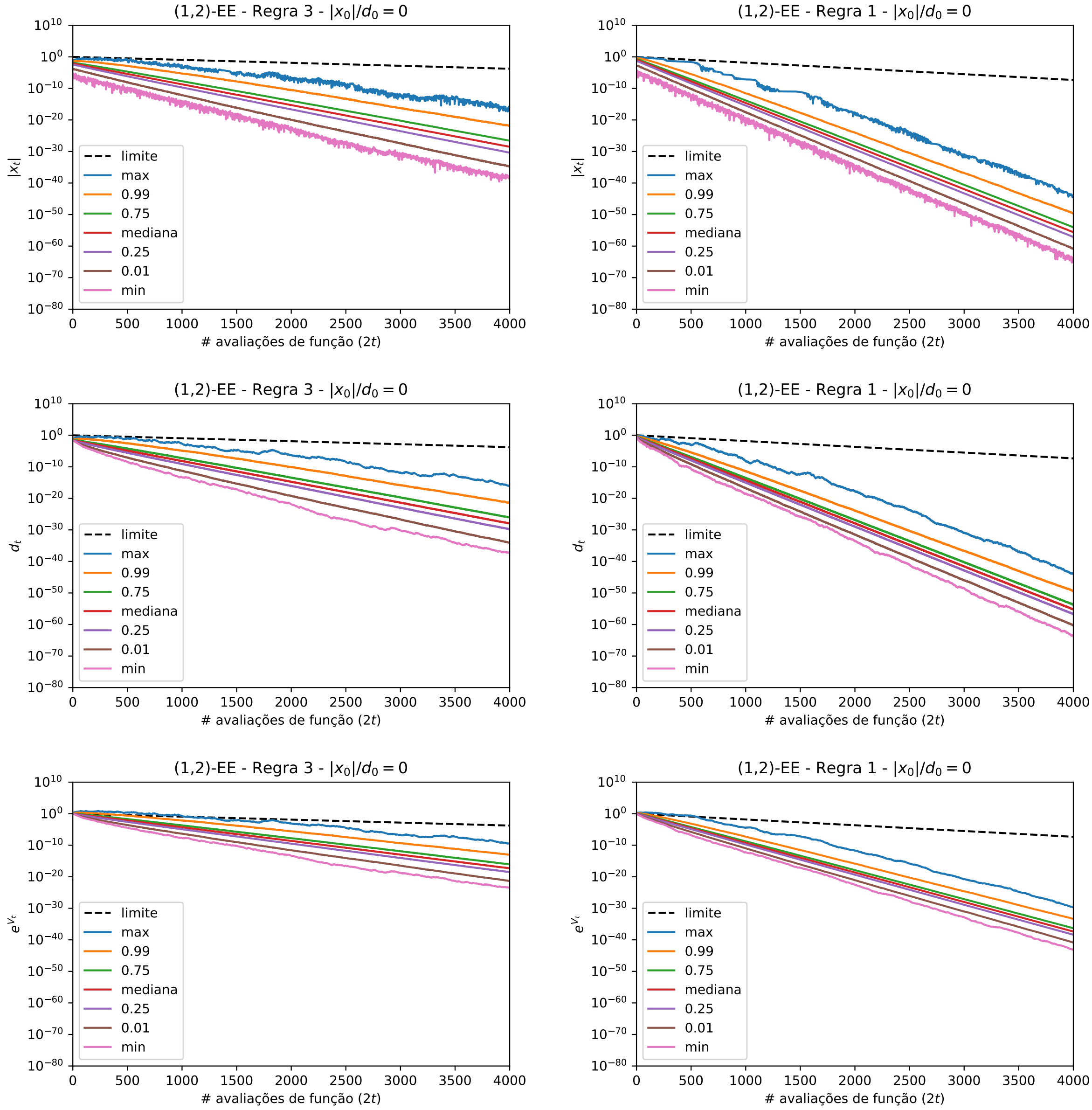

Figura 10 – Quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 3** e **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

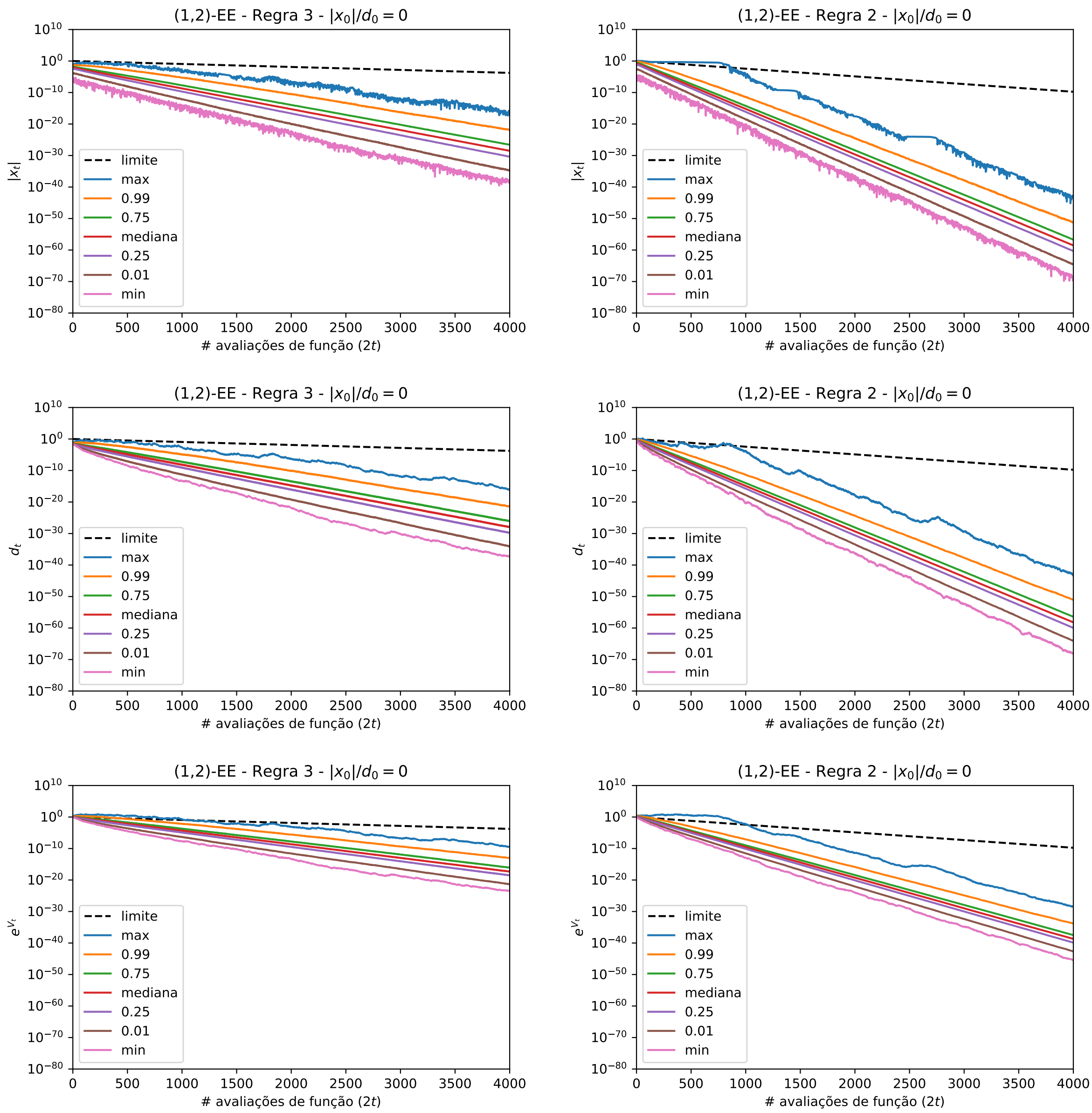

Figura 11 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 3** e **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

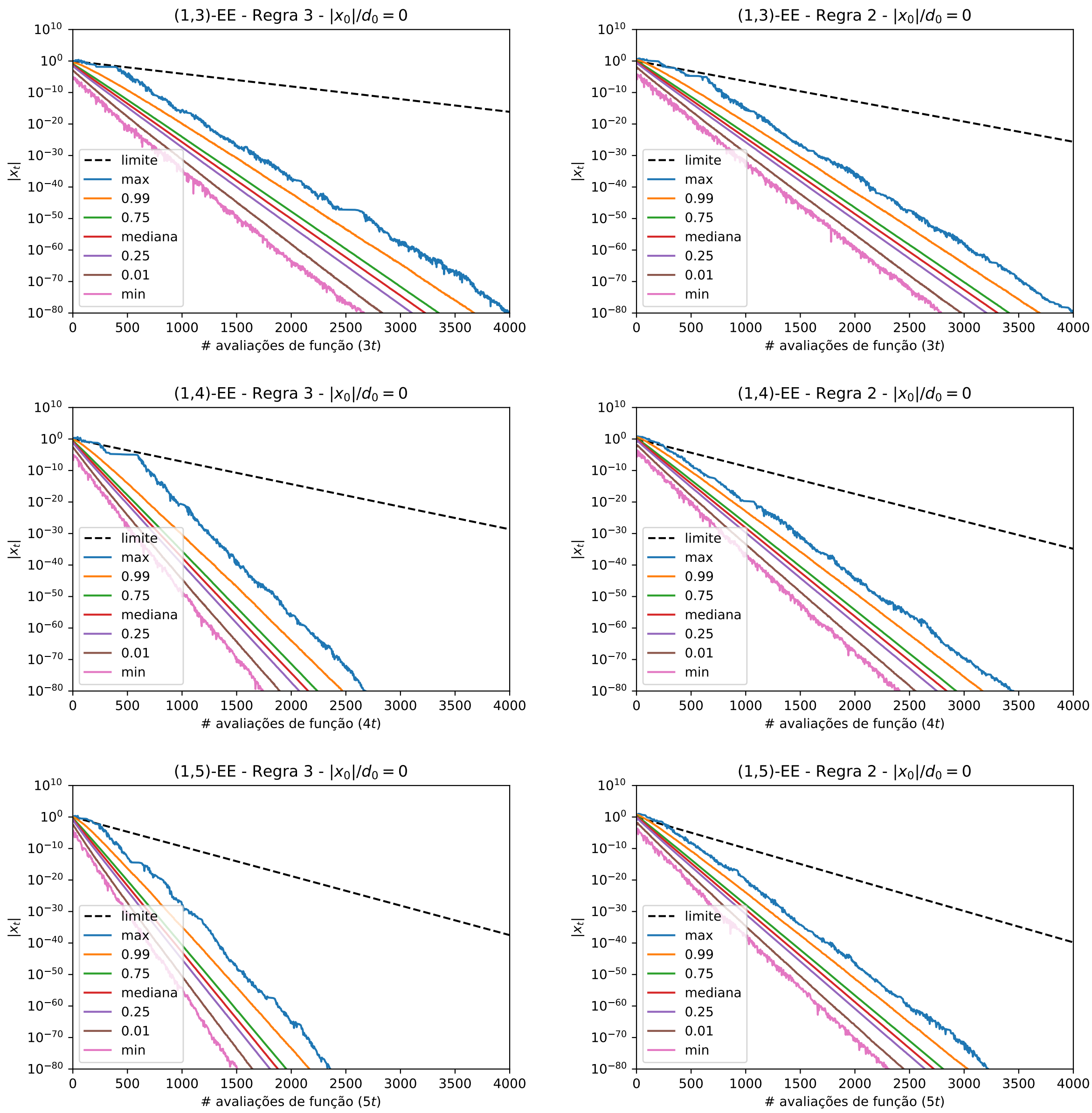

Figura 12 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 3** e **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Diferentemente da adaptação baseada em sucesso, considerando os resultados experimentais apresentados referentes à execução do algoritmo iniciado fora do ótimo ($|x_0|/d_0 = 10^{15}$), para a adaptação definida no presente capítulo, a probabilidade do tamanho do passo diminuir pode ser bastante diferente do valor teórico se este mesmo ponto inicial for considerado. Tal fato ocorre devido a um problema de precisão numérica no arredondamento dos valores de x_t que afeta diretamente a probabilidade do tamanho do passo diminuir. Por isso, nos experimentos em que os algoritmos são iniciados com tamanho de passo muito pequeno consideramos (x_0, d_0) tal que $|x_0|/d_0 = 10^{10}$.

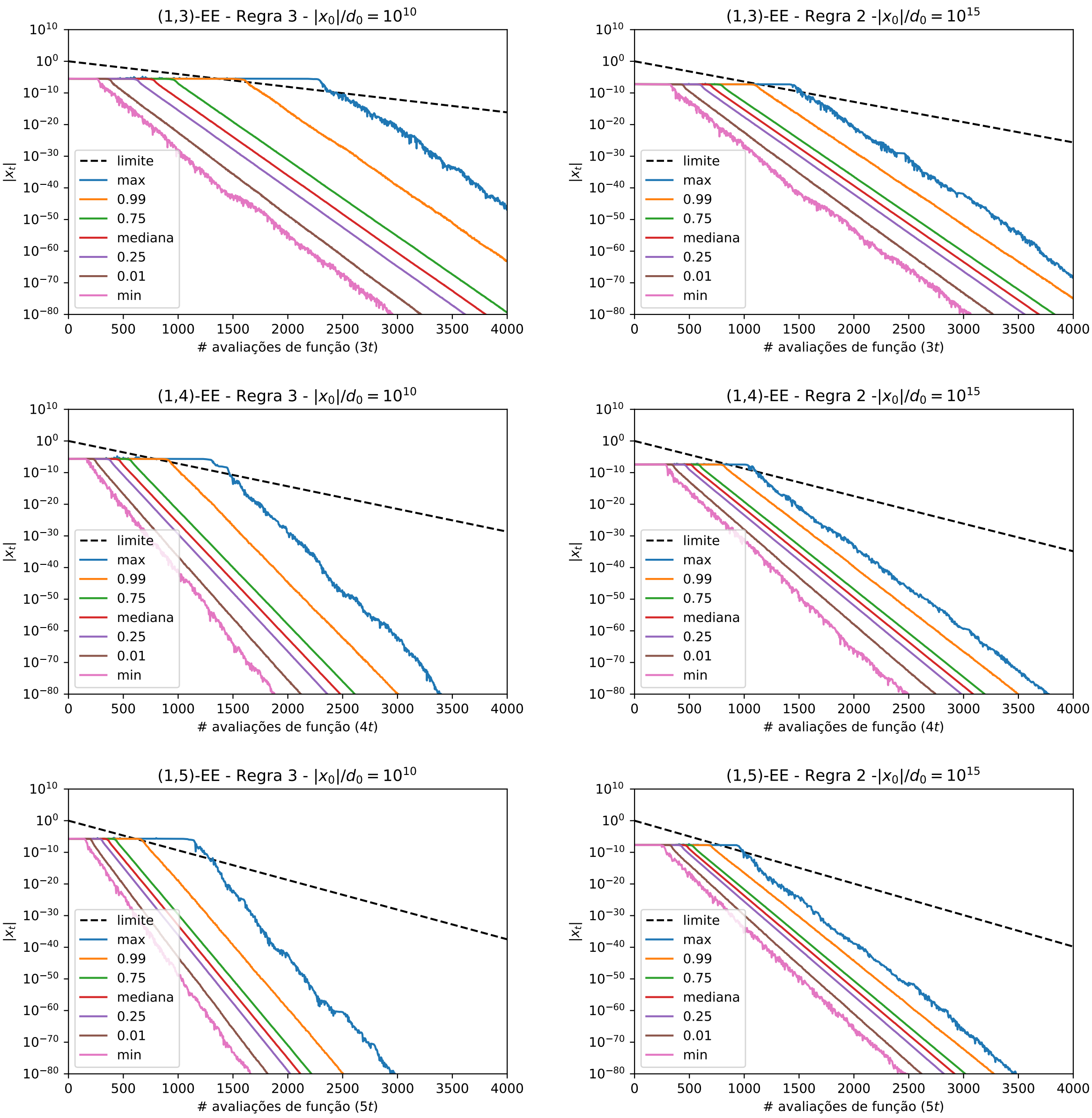

Figura 13 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1{,}\lambda)$-EE com **Regra 3** e **Regra 2** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{10}$).

A Figura 14 mostra quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t ao longo das gerações para $(1{,}6)$-EE com **Regra 3** e $(1{,}4)$-EE com **Regra 1**, tendo iniciado em $x_0 = 0$, juntamente com os limites de convergência correspondentes.

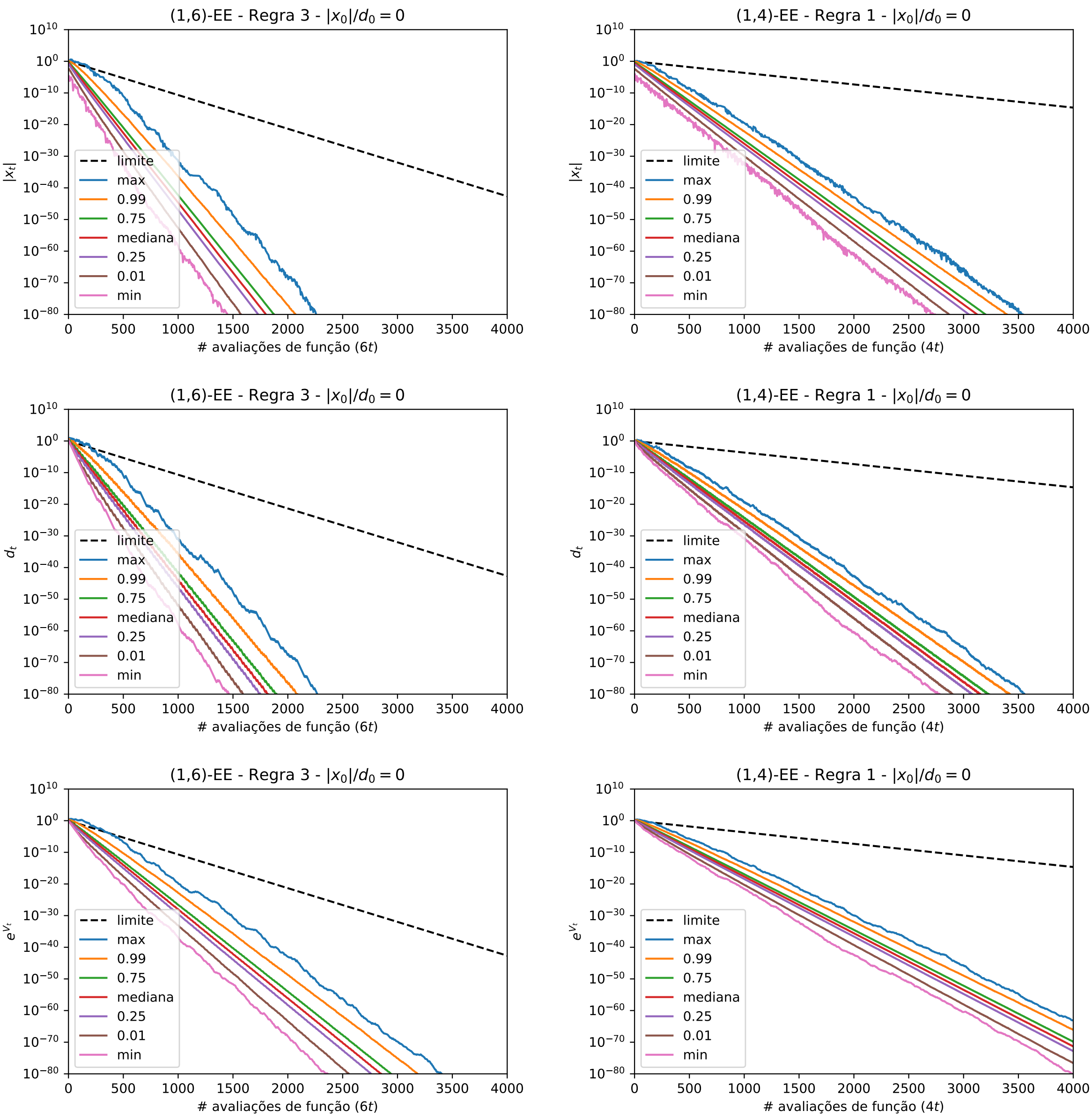

Figura 14 – Quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} estimadas de 10001 execuções de $(1{,}6)$-EE com **Regra 3** e $(1{,}4)$-EE com **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Analisando a Figura 14 notamos que o declive do máximo das distribuições de $|x_t|, d_t$ e e^{V_t} para $(1{,}6)$-EE com **Regra 3** é mais acentuado que o declive do mínimo das distribuições de $|x_t|, d_t$ e e^{V_t} para $(1{,}4)$-EE com **Regra 1**, indicando de fato uma convergência ao ótimo muito mais rápida da **Regra 3**. Apesar do limite teórico da taxa de convergência para a **Regra 2** ser maior que o da **Regra 3**, a Figura 15 mostra uma convergência mais rápida da **Regra 3**.

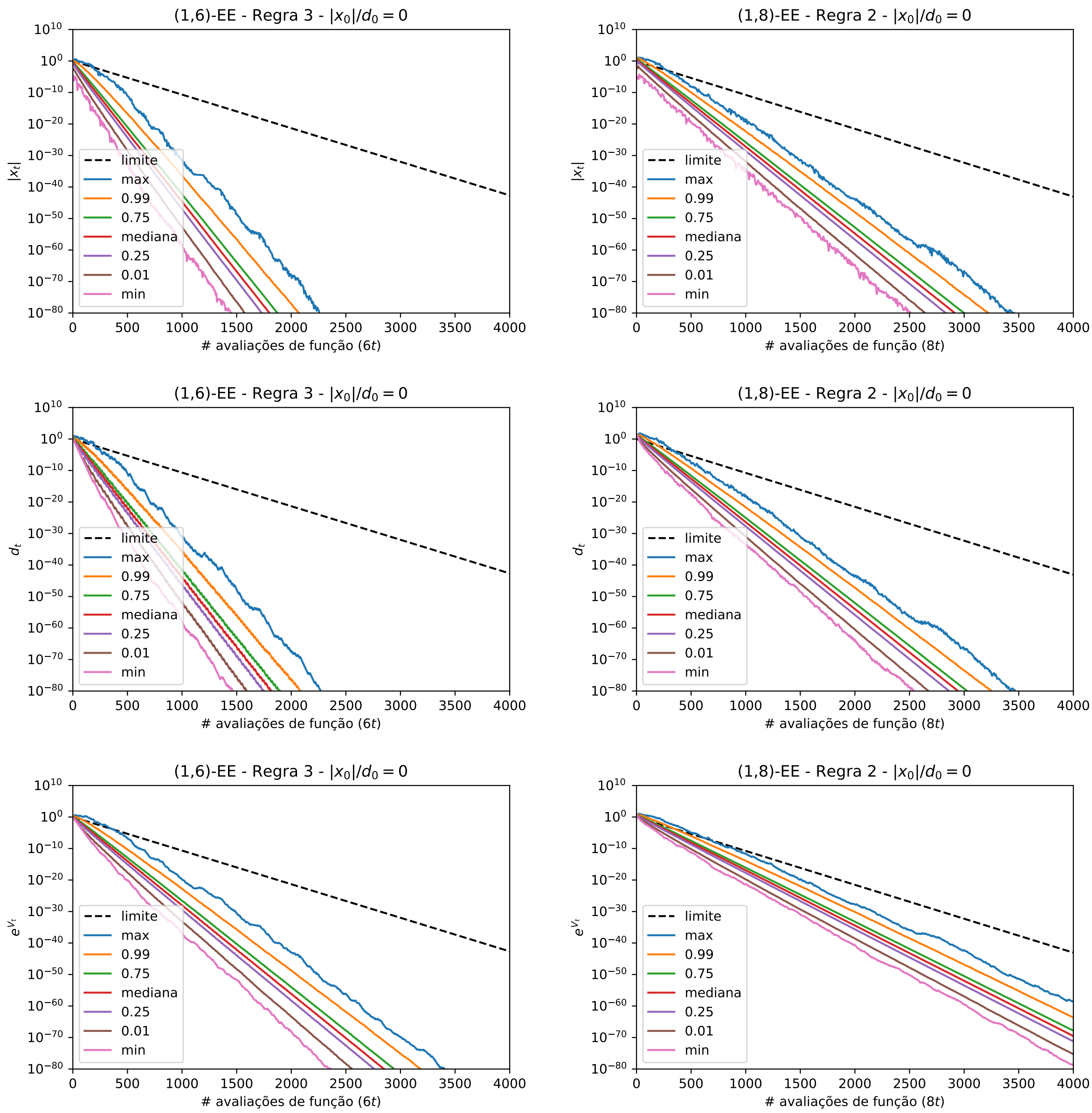

Figura 15 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1{,}6)$-EE com **Regra 3** e $(1{,}8)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

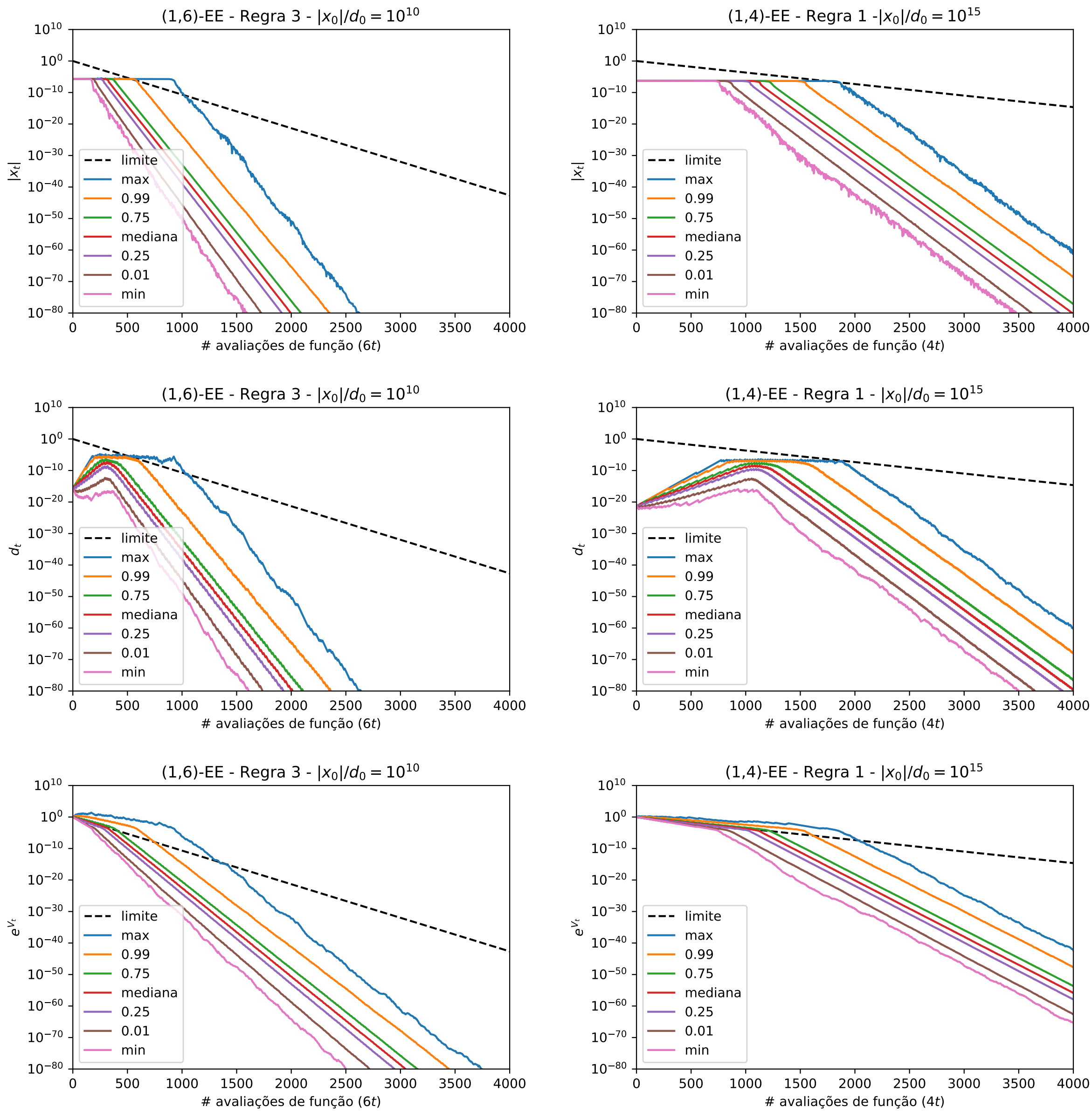

Figura 16 – Quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} estimadas de 10001 execuções de $(1{,}6)$-EE com **Regra 3** e $(1{,}4)$-EE com **Regra 1** iniciadas longe do ótimo.

A diferenças de desempenho entre $(1{,}6)$-EE com **Regra 3** e $(1{,}4)$-EE com a **Regra 1** são muito mais evidentes quando o tamanho do passo inicial é muito pequeno, como mostra a Figura 16. Neste caso, verificamos que o tamanho do passo d_t cresce muito mais rápido nos estágios iniciais das corridas para a **Regra 3** do que para a **Regra 1**, levando a uma convergência global mais rápida.

No que diz respeito à seleção elitista, os limites teóricos obtidos corroboram os resultados experimentais e mostram que a **Regra 3** converge mais rápido que ambas as regras de adaptação do tamanho do passo baseadas em sucesso. Nesses experimentos todos os quantis das distribuições de $|x_t|$, d_t e e^{V_t} estão mais próximos ao limite teórico (ver

Figuras 65, 66 e 67 no Apêndice F).

A Figura 17 nos permite verificar que quase não existe diferença entre a **Regra 1** e a **Regra 3** para as configurações ótimas. A convergência mais rápida da estratégia com **Regra 3** é mais evidente quando comparada com a **Regra 2** (ver Figura 18) .

Também podemos verificar que quando os algoritmos são iniciados fora do ótimo, d_t cresce muito mais rápido nos estágios iniciais dos experimentos para a **Regra 3** do que para a **Regra 1**. Essa diferença não é tão evidente quando a **Regra 3** é comparada à **Regra 2** (ver Figuras 19 e 20).

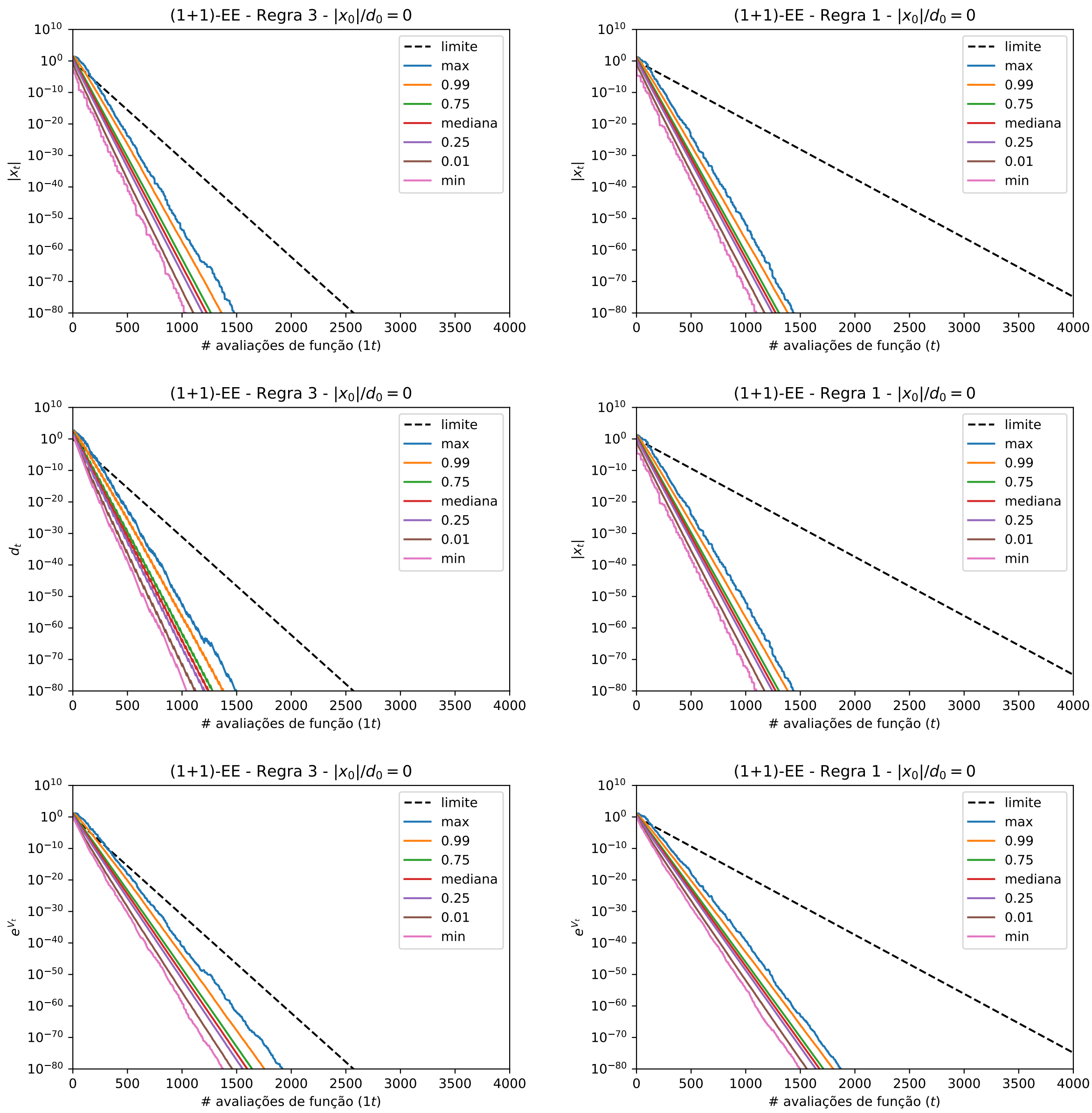

Figura 17 – Quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} estimadas de 10001 execuções de $(1+1)$-EE com **Regra 3** e **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

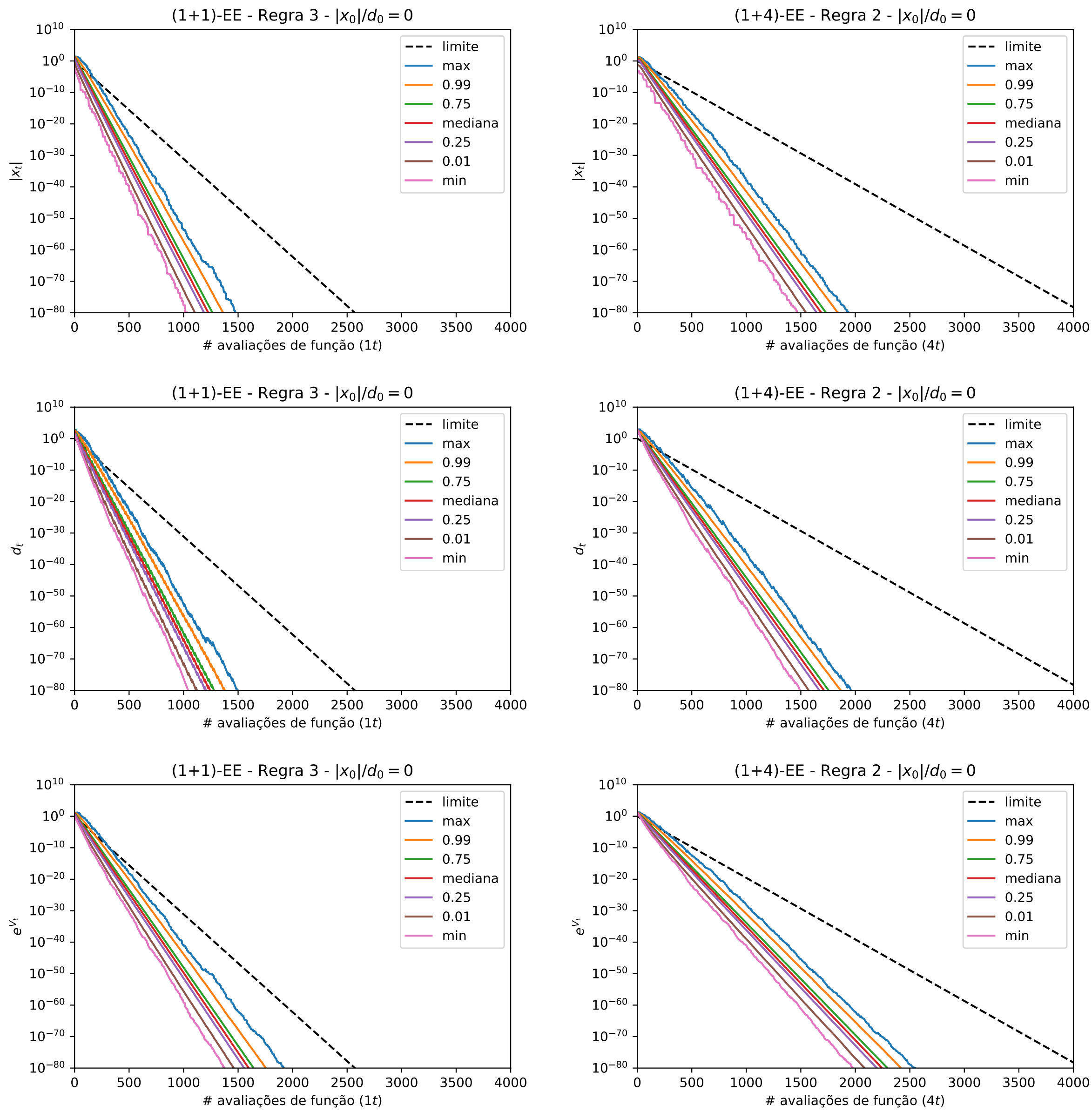

Figura 18 – Quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} estimadas de 10001 execuções de $(1+4)$-EE com **Regra 2** e $(1+1)$-EE com **Regra 3** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

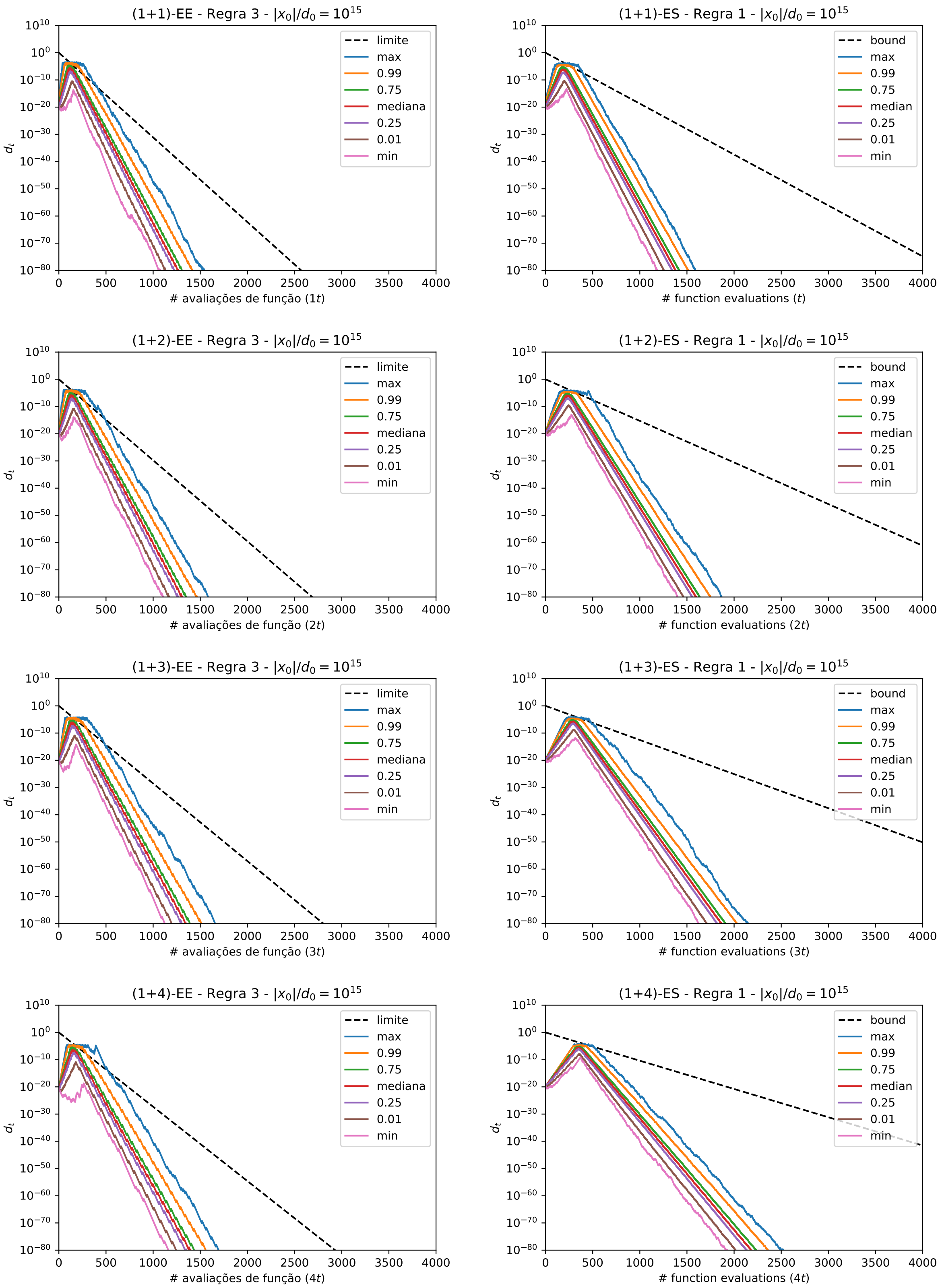

Figura 19 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 3** e **Regra 1** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

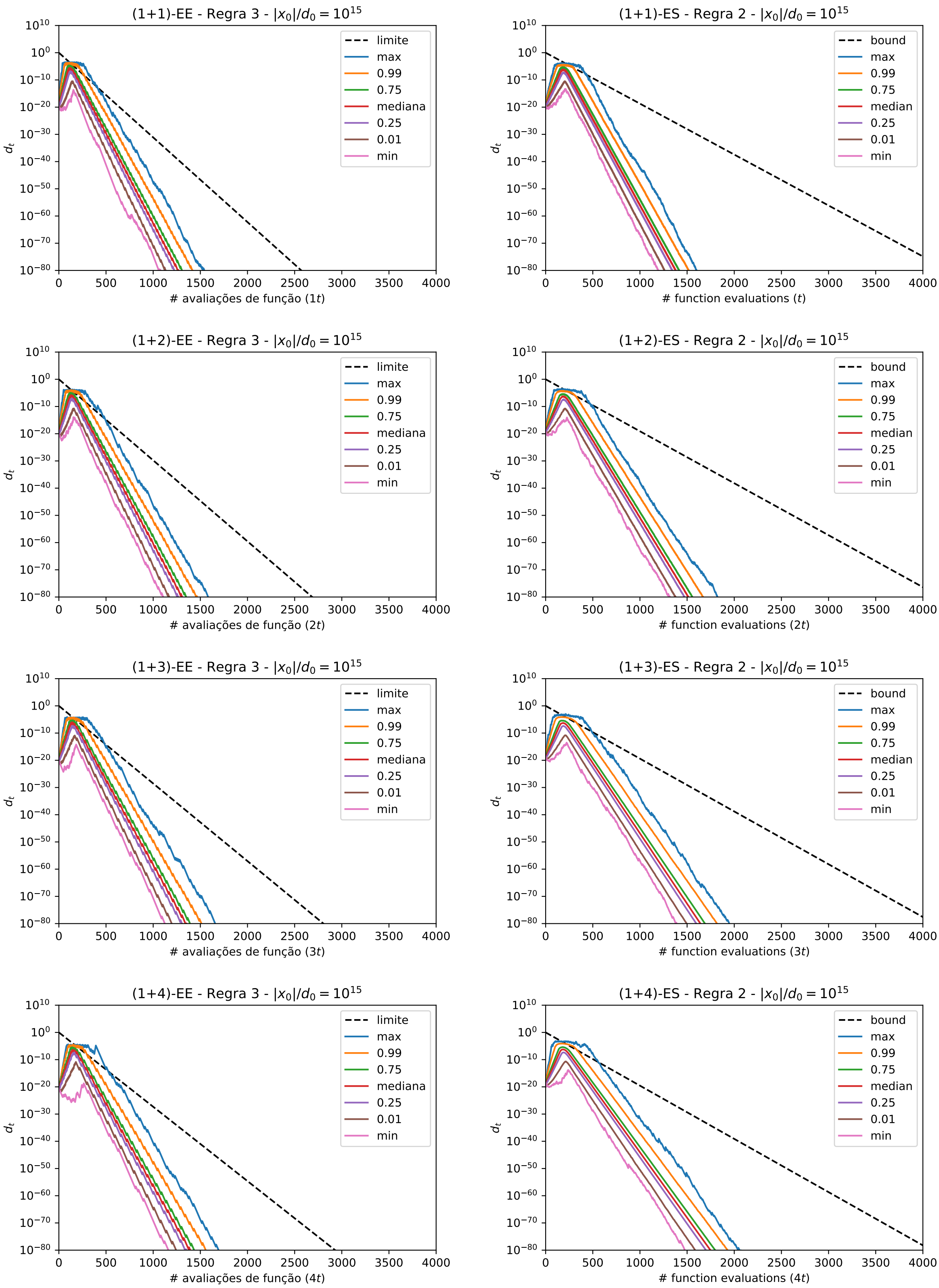

Figura 20 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 3** e com **Regra 2** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

Na Tabela 9 listamos os valores de a/λ com λ ótimo para cada uma das EEs analisadas até o momento.

Regra 1	**Regra 2**	**Regra 3**
$(1,4)$-EE = 0.00842	$(1,8)$-EE = **0.02478**	$(1,6)$-EE = 0.024552
$(1+1)$-EE = 0.04309	$(1+4)$-EE = 0.04504	$(1+1)$-EE = **0.071659**

Tabela 9 – Valores de a/λ com λ ótimo para cada regra de adaptação.

Notamos que para a seleção $(1,\lambda)$, a **Regra 2** apresenta o maior limite ótimo da taxa de convergência normalizada. Na seleção elitista, o maior dentre estes limites é obtido com a **Regra 3**. Apesar da **Regra 2** apresentar um limite de convergência ligeiramente melhor que a **Regra 3** para a seleção não elitista, na prática isso não ocorre conforme os experimentos apontam.

De forma geral, novamente todas as configurações dos algoritmos expostos neste capítulo levam a uma convergência muito mais rápida da distância ao ótimo, $|x_t|$ e do tamanho do passo, d_t, para zero do que garantido pelos limites teóricos e também quando comparadas às estratégias analisadas no Capítulo 3.

No próximo capítulo utilizaremos o mesmo procedimento de síntese de regras de controle do passo de mutação para analisar e estimar a velocidade de convergência de versões de EEs formuladas a partir das regras adaptativas estudadas até o momento. Admitindo a mesma classe de funções, formularemos $(1 \overset{+}{,} \lambda)$-EE com o passo de mutação controlado por meio da combinação entre a **Regra 1** e a **Regra 3** . Também realizaremos um refinamento da análise, obtendo limites para a taxa de convergência sem o auxílio da desigualdade de Jensen.

5 Avanços Iniciais para Análise e Síntese de Estratégias Evolutivas

Como mencionado ao final do Capítulo 3, a metodologia utilizada para obter os resultados apresentados até o momento pode ser desenvolvida também para algumas variações das Estratégias Evolutivas estudadas. A proposta deste capítulo é apresentar novos cenários provenientes das formulações vistas nos capítulos anteriores. Primeiramente, buscamos desenvolver Estratégias Evolutivas $(1 \overset{+}{,} \lambda)$-EE com controle de mutação combinado, no qual o tamanho do passo é adaptado com base na **Regra 1**, articulada com a **Regra 3**.

Os limites superiores da taxa de convergência, bem como os parâmetros que os otimizam, resultantes da abordagem empregada, são considerados conservadores. Conforme alegado, acredita-se que um dos motivos para tal fato seja o uso da desigualdade de Jensen para o alcance de uma expressão simplificada para $\Psi(r)$ (limite superior para $E^{A_t}(V_{t+1}) - V_t$). Dessa forma revisamos neste capítulo os resultados obtidos nos capítulos anteriores a partir da mesma metodologia, porém determinando expressões explícitas para $E^{A_t}(V_{t+1}) - V_t$ e estimando a taxa de convergência das estratégias, considerando o decrescimento esperado do supermartingal associado V_t.

5.1 Adaptação Combinada

Nesta seção propomos a implementação de uma formulação híbrida para Estratégias Evolutivas $(1 \overset{+}{,} \lambda)$-EE, nas quais o tamanho do passo é controlado segundo uma combinação da adaptação baseada em sucesso com o controle de mutação baseado no tamanho do passo. Ao combinar a **Regra 2** com a **Regra 3**, $2(\lambda + 1)$ casos precisam ser considerados para descrever as atualizações do tamanho do passo. Como a **Regra 1** é um caso particular da **Regra 2**, apenas a primeira é combinada com a **Regra 3**.

É intuitivo que a adaptação combinada produza taxas de convergência melhores que as taxas obtidas em cada adaptação separadamente. É possível que se chegue ao ótimo mais rapidamente, se os tamanhos dos passos são controlados com base na distância do indivíduo selecionado ao pai e também na ocorrência de sucesso ou fracasso.

Considerando a seleção da mesma forma como nos algoritmos descritos até o momento, as atualizações do tamanho do passo para a adaptação combinada com seleção

$(1,\lambda)$-EE são descritas da seguinte forma:

$$
\begin{aligned}
d_{t+1} &= G(x_t,d_t,\mu_{1,t+1},\ldots,\mu_{\lambda,t+1}) \\
&= \begin{cases}
\alpha_s^+ \cdot d_t & \text{se } \exists\, i \in \{1,\ldots,\lambda\} : f(x_{i,t+1}) \leq f(x_t) \text{ e } |x_{t+1}-x_t| \geq \theta d_t \\
\alpha_s^- \cdot d_t & \text{se } \exists\, i \in \{1,\ldots,\lambda\} : f(x_{i,t+1}) \leq f(x_t) \text{ e } |x_{t+1}-x_t| < \theta d_t \\
\alpha_f^- \cdot d_t & \text{se } f(x_{i,t+1}) > f(x_t)\ \forall\, i \in \{1,\ldots,\lambda\} \text{ e } |x_{t+1}-x_t| < \theta d_t \\
\alpha_f^+ \cdot d_t & \text{se } f(x_{i,t+1}) > f(x_t)\ \forall\, i \in \{1,\ldots,\lambda\} \text{ e } |x_{t+1}-x_t| \geq \theta d_t
\end{cases}
\end{aligned}
\tag{66}
$$

com $\alpha_s^+, \alpha_s^-, \alpha_f^-, \alpha_f^+ > 0$ e $0 < \theta < 1$.

Na seleção elitista $(1+\lambda)$-EE, se $f(x_{i,t+1}) > f(x_t)\ \forall\, i \in \{1,\ldots,\lambda\}$, então $x_{t+1} = x_t$ e consequentemente $|x_{t+1} - x_t| \geq \theta d_t$ se, e somente se, $\theta = 0$ ou $d_t = 0$. Logo a ocorrência de fracasso não permite que o indivíduo selecionado para compor a nova geração esteja longe do pai. Neste caso, as equações que definem as atualizações do tamanho do passo para a adaptação combinada são:

$$
\begin{aligned}
d_{t+1} &= G(x_t,d_t,\mu_{1,t+1},\ldots,\mu_{\lambda,t+1}) \\
&= \begin{cases}
\alpha_s^+ \cdot d_t & \text{se } \exists\, i \in \{1,\ldots,\lambda\} : f(x_{i,t+1}) \leq f(x_t) \text{ e } |x_{t+1}-x_t| \geq \theta d_t \\
\alpha_s^- \cdot d_t & \text{se } \exists\, i \in \{1,\ldots,\lambda\} : f(x_{i,t+1}) \leq f(x_t) \text{ e } |x_{t+1}-x_t| < \theta d_t \\
\alpha_f \cdot d_t & \text{se } f(x_{i,t+1}) > f(x_t)\ \forall\, i \in \{1,\ldots,\lambda\}
\end{cases}
\end{aligned}
\tag{67}
$$

Para obter o valor esperado do próximo tamanho do passo d_{t+1} nesta abordagem combinada, é necessário o cálculo de quatro probabilidades, diferentemente do que ocorre com as formulações individuais, em que era necessário apenas o cálculo de duas probabilidades.

Para $(1,\lambda)$-EE com controle de mutação combinado temos $E^{A_t}(d_{t+1}) = d_t(\alpha_s^+ \cdot P_s^+ + \alpha_s^- \cdot P_s^- + \alpha_f^- \cdot P_f^- + \alpha_f^+ \cdot P_f^+)$, com:

$P_s^+ =$ probabilidade de existir um filho que seja pelo menos tão bom quanto o pai e do filho selecionado estar longe do pai.

$P_s^- =$ probabilidade de existir um filho que seja pelo menos tão bom quanto o pai e do filho selecionado estar perto do pai.

$P_f^- =$ probabilidade de todos os filhos serem piores que o pai e do filho selecionado estar perto do pai.

$P_f^+ =$ probabilidade de todos os filhos serem piores que o pai e do filho selecionado estar longe do pai.

Para a seleção $(1+\lambda)$-EE , $\alpha_f^+ = \alpha_f^- = \alpha_f$ e $E^{A_t}(d_{t+1}) = d_t(\alpha_s^+ \cdot P_s^+ + \alpha_s^- \cdot P_s^- + \alpha_f \cdot P_f)$, sendo $P_f = P_f^- + P_f^+$ a probabilidade de fracasso.

Observamos que $P_s = P_s^+ + P_s^- =$ Probabilidade de sucesso (para a **Regra 1**), $P^+ = P_f^+ + P_s^+$ e $P^- = P_f^- + P_s^-$.

Para o cálculo de P_s^+, P_s^-, P_f^- e P_f^+, além das observações (i) a (v), listadas na Seção 4.1, é preciso notar que no caso de sucesso temos $|x_{t+1}| \leq x_t$, ou seja, $-x_t \leq x_{t+1} \leq x_t$ e no caso de fracasso, $|x_{t+1}| > x_t$, isto é, $x_{t+1} > x_t$ ou $x_{t+1} < -x_t$. Portanto é necessário analisar as possibilidades de ordenação dos valores:

$$x_t - d_t; \;\; x_t - \theta d_t; \;\; -x_t; \;\; 0; \;\; x_t; \;\; -x_t + d_t; \;\; x_t + \theta d_t; \;\; x_t + d_t$$

Como $x_t > 0$ e $d_t > 0$, então $-x_t < 0 < x_t + \theta d_t < x_t + d_t$. Porém não sabemos o sinal de $x_t - d_t$ e $x_t - \theta d_t$.

Considerando $x_t - d_t < 0$, é possível que ocorra $x_t - \theta d_t < 0$ ou $x_t - \theta d_t > 0$.

Se $\boldsymbol{x_t - \theta d_t < 0}$, as possibilidades de ordenação são:

1) $x_t - d_t < x_t - \theta d_t < -x_t < 0 < x_t < x_t + \theta d_t < -x_t + d_t < x_t + d_t$

Neste caso $0 \leq \dfrac{x_t}{d_t} < \min\left\{\dfrac{\theta}{2}, \dfrac{1-\theta}{2}\right\}$, $P_s^+ = 0$,

$$P_s^- = \frac{\lambda}{(2d_t)^\lambda}\left[\int_{-x_t}^{0} (2d_t + 2z)^{\lambda-1}\mathrm{d}z + \int_{0}^{x_t} (2d_t - 2z)^{\lambda-1}\mathrm{d}z\right] = 1 - \left(1 - \frac{x_t}{d_t}\right)^\lambda \tag{68}$$

$$\begin{aligned} P_f^- &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-\theta d_t}^{-x_t} (2d_t + 2z)^{\lambda-1}\mathrm{d}z + \int_{x_t}^{x_t+\theta d_t} (2d_t - 2z)^{\lambda-1}\mathrm{d}z\right] \\ &= -\frac{1}{2}\left[\left(\frac{x_t}{d_t} + 1 - \theta\right)^\lambda + \left(-\frac{x_t}{d_t} + 1 - \theta\right)^\lambda\right] + \left(1 - \frac{x_t}{d_t}\right)^\lambda \end{aligned} \tag{69}$$

$$\begin{aligned} P_f^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{x_t-\theta d_t} (2d_t + 2z)^{\lambda-1}\mathrm{d}z + \int_{x_t+\theta d_t}^{-x_t+d_t} (2d_t - 2z)^{\lambda-1}\mathrm{d}z\right. \\ &\quad \left. + \int_{-x_t+d_t}^{x_t+d_t} (x_t + d_t - z)^{\lambda-1}\mathrm{d}z\right] \\ &= \frac{1}{2}\left[\left(\frac{x_t}{d_t} + 1 - \theta\right)^\lambda + \left(-\frac{x_t}{d_t} + 1 - \theta\right)^\lambda\right] \end{aligned} \tag{70}$$

$$E^{A_t}(d_{t+1}) = d_t\left\{\frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^{\lambda}+\left(1-\theta-\frac{x_t}{d_t}\right)^{\lambda}\right](\alpha_f^+-\alpha_f^-) + \left(1-\frac{x_t}{d_t}\right)^{\lambda}(\alpha_f^- - \alpha_s^-)+\alpha_s^-\right\} \quad (71)$$

$$E^{A_t}[\ln(d_{t+1})] = \ln(d_t)+\frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^{\lambda}+\left(1-\theta-\frac{x_t}{d_t}\right)^{\lambda}\right]\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right) + \left(1-\frac{x_t}{d_t}\right)^{\lambda}\ln\left(\frac{\alpha_f^-}{\alpha_s^-}\right)+\ln(\alpha_s^-) \quad (72)$$

2) $x_t - d_t < x_t - \theta d_t < -x_t < 0 < x_t < -x_t + d_t < x_t + \theta d_t < x_t + d_t$

Neste caso $\frac{1-\theta}{2} \leq \frac{x_t}{d_t} < \frac{\theta}{2}$, se $\theta > \frac{1}{2}$, $P_s^+ = 0$, P_s^- é descrita pela Equação (68),

$$\begin{aligned} P_f^- &= \frac{\lambda}{(2d_t)^{\lambda}}\left[\int_{x_t-\theta d_t}^{-x_t}(2d_t+2z)^{\lambda-1}\mathrm{d}z + \int_{x_t}^{-x_t+d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z\right] \\ &+ \frac{\lambda}{(2d_t)^{\lambda}}\int_{-x_t+d_t}^{x_t+\theta d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z \\ &= -\frac{1}{2}\left[\left(\frac{x_t}{d_t}+1-\theta\right)^{\lambda}-\left(\frac{x_t}{d_t}\right)^{\lambda}\right]+\left(1-\frac{x_t}{d_t}\right)^{\lambda}-\left(\frac{1-\theta}{2}\right)^{\lambda} \end{aligned} \quad (73)$$

$$\begin{aligned} P_f^+ &= \frac{\lambda}{(2d_t)^{\lambda}}\left[\int_{x_t-d_t}^{x_t-\theta d_t}(2d_t+2z)^{\lambda-1}\mathrm{d}z + \int_{x_t+\theta d_t}^{x_t+d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= \frac{1}{2}\left[\left(\frac{x_t}{d_t}+1-\theta\right)^{\lambda}-\left(\frac{x_t}{d_t}\right)^{\lambda}\right]+\left(\frac{1-\theta}{2}\right)^{\lambda} \end{aligned} \quad (74)$$

$$\begin{aligned} E^{A_t}(d_{t+1}) &= d_t\left\{\frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^{\lambda}-\left(\frac{x_t}{d_t}\right)^{\lambda}\right]+\left(\frac{1-\theta}{2}\right)^{\lambda}\right\}(\alpha_f^+-\alpha_f^-) \\ &+ d_t\left[\left(1-\frac{x_t}{d_t}\right)^{\lambda}(\alpha_f^- - \alpha_s^-)+\alpha_s^-\right] \end{aligned} \quad (75)$$

$$\begin{aligned} E^{A_t}[\ln(d_{t+1})] &= \ln(d_t)+\left\{\frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^{\lambda}-\left(\frac{x_t}{d_t}\right)^{\lambda}\right]+\left(\frac{1-\theta}{2}\right)^{\lambda}\right\}\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right) \\ &+ \left(1-\frac{x_t}{d_t}\right)^{\lambda}\ln\left(\frac{\alpha_f^-}{\alpha_s^-}\right)+\ln(\alpha_s^-) \end{aligned} \quad (76)$$

3) $x_t - d_t < -x_t < x_t - \theta d_t < 0 < x_t < x_t + \theta d_t < -x_t + d_t < x_t + d_t$

Neste caso $\dfrac{\theta}{2} \leq \dfrac{x_t}{d_t} < \min\left\{\dfrac{1-\theta}{2}, \theta\right\}$, se $\theta < \dfrac{1}{2}$, com

$$P_s^+ = \frac{\lambda}{(2d_t)^\lambda} \int_{-x_t}^{x_t - \theta d_t} (2d_t + 2z)^{\lambda-1} \mathrm{d}z = \frac{1}{2}\left[\left(\frac{x_t}{d_t} + 1 - \theta\right)^\lambda - \left(1 - \frac{x_t}{d_t}\right)^\lambda\right] \tag{77}$$

$$\begin{aligned} P_s^- &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-\theta d_t}^{0} (2d_t + 2z)^{\lambda-1}\mathrm{d}z + \int_0^{x_t} (2d_t - 2z)^{\lambda-1}\mathrm{d}z\right] \\ &= 1 - \frac{1}{2}\left[\left(\frac{x_t}{d_t} + 1 - \theta\right)^\lambda + \left(1 - \frac{x_t}{d_t}\right)^\lambda\right] \end{aligned} \tag{78}$$

$$P_f^- = \frac{\lambda}{(2d_t)^\lambda} \int_{x_t}^{x_t + \theta d_t} (2d_t - 2z)^{\lambda-1} \mathrm{d}z = -\frac{1}{2}\left[\left(-\frac{x_t}{d_t} + 1 - \theta\right)^\lambda - \left(1 - \frac{x_t}{d_t}\right)^\lambda\right] \tag{79}$$

$$\begin{aligned} P_f^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{-x_t} (2d_t + 2z)^{\lambda-1}\mathrm{d}z + \int_{x_t+\theta d_t}^{-x_t+d_t} (2d_t - 2z)^{\lambda-1}\mathrm{d}z \right. \\ &\quad \left. + \int_{-x_t+d_t}^{x_t+d_t} (x_t + d_t - z)^{\lambda-1}\mathrm{d}z\right] \\ &= \frac{1}{2}\left[\left(-\frac{x_t}{d_t} + 1 - \theta\right)^\lambda + \left(1 - \frac{x_t}{d_t}\right)^\lambda\right] \end{aligned} \tag{80}$$

$$\begin{aligned} E^{A_t}(d_{t+1}) &= d_t \left\{\frac{1}{2}\left[\left(1 - \theta + \frac{x_t}{d_t}\right)^\lambda (\alpha_s^+ - \alpha_s^-) + \left(1 - \theta - \frac{x_t}{d_t}\right)^\lambda (\alpha_f^+ - \alpha_f^-)\right.\right. \\ &\quad \left.\left. + \left(1 - \frac{x_t}{d_t}\right)^\lambda (\alpha_f^+ + \alpha_f^- - \alpha_s^- - \alpha_s^+)\right] + \alpha_s^-\right\} \end{aligned} \tag{81}$$

$$\begin{aligned} E^{A_t}[\ln(d_{t+1})] &= \ln(d_t) + \frac{1}{2}\left[\left(1 - \theta + \frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_s^+}{\alpha_s^-}\right) + \left(1 - \theta - \frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)\right. \\ &\quad \left. + \left(1 - \frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_f^+ \alpha_f^-}{\alpha_s^- \alpha_s^+}\right)\right] + \ln(\alpha_s^-) \end{aligned} \tag{82}$$

4) $x_t - d_t < -x_t < x_t - \theta d_t < 0 < x_t < -x_t + d_t < x_t + \theta d_t < x_t + d_t$

Neste caso $\frac{1-\theta}{2} \leq \frac{x_t}{d_t} < \theta$, $\left(\text{se } \frac{1}{3} < \theta < \frac{1}{2}\right)$ ou $\frac{\theta}{2} \leq \frac{x_t}{d_t} < \frac{1}{2}$, $\left(\text{se } \theta > \frac{1}{2}\right)$ e P_s^+ e P_s^- são respectivamente descritas pelas Equações (77) e (78),

$$
\begin{aligned}
P_f^- &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t}^{-x_t+d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z + \int_{-x_t+d_t}^{x_t+\theta d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\
&= \frac{1}{2}\left[\left(\frac{x_t}{d_t}\right)^\lambda + \left(1-\frac{x_t}{d_t}\right)^\lambda\right] - \left(\frac{1-\theta}{2}\right)^\lambda
\end{aligned} \tag{83}
$$

$$
\begin{aligned}
P_f^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{-x_t}(2d_t+2z)^{\lambda-1}\mathrm{d}z + \int_{x_t+\theta d_t}^{x_t+d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\
&= -\frac{1}{2}\left[\left(\frac{x_t}{d_t}\right)^\lambda - \left(1-\frac{x_t}{d_t}\right)^\lambda\right] + \left(\frac{1-\theta}{2}\right)^\lambda
\end{aligned} \tag{84}
$$

$$
\begin{aligned}
E^{A_t}(d_{t+1}) &= d_t\left\{\frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^\lambda(\alpha_s^+-\alpha_s^-) + \left(\frac{x_t}{d_t}\right)^\lambda(\alpha_f^- - \alpha_f^+)\right.\right. \\
&\quad \left. + \left(1-\frac{x_t}{d_t}\right)^\lambda(\alpha_f^+ + \alpha_f^- - \alpha_s^+ - \alpha_s^-)\right] \\
&\quad \left. + \left(\frac{1-\theta}{2}\right)^\lambda(\alpha_f^+ - \alpha_f^-) + \alpha_s^-\right\}
\end{aligned} \tag{85}
$$

$$
\begin{aligned}
E^{A_t}[\ln(d_{t+1})] &= \ln(d_t) + \frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_s^+}{\alpha_s^-}\right) + \left(\frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_f^-}{\alpha_f^+}\right)\right. \\
&\quad \left. + \left(1-\frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_f^+\alpha_f^-}{\alpha_s^-\alpha_s^+}\right)\right] + \left(\frac{1-\theta}{2}\right)^\lambda \ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right) + \ln(\alpha_s^-)
\end{aligned} \tag{86}
$$

5) $-x_t < x_t - d_t < x_t - \theta d_t < 0 < -x_t + d_t < x_t < x_t + \theta d_t < x_t + d_t$

Neste caso $\frac{1}{2} \leq \frac{x_t}{d_t} < \theta$, se $\theta > \frac{1}{2}$, com

$$
P_s^+ = \frac{\lambda}{(2d_t)^\lambda}\int_{x_t-d_t}^{x_t-\theta d_t}(2d_t+2z)^{\lambda-1}\mathrm{d}z = \frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^\lambda - \left(\frac{x_t}{d_t}\right)^\lambda\right] \tag{87}
$$

$$
\begin{aligned}
P_s^- &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-\theta d_t}^{0}(2d_t+2z)^{\lambda-1}\mathrm{d}z+\int_{0}^{-x_t+d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z\right]\\
&+ \frac{\lambda}{(2d_t)^\lambda}\int_{-x_t+d_t}^{x_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\\
&= -\frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^\lambda-\left(\frac{x_t}{d_t}\right)^\lambda\right]+1-\frac{1}{2^\lambda} \qquad (88)
\end{aligned}
$$

$$
P_f^- = \frac{\lambda}{(2d_t)^\lambda}\int_{x_t}^{x_t+\theta d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z = -\left(\frac{1-\theta}{2}\right)^\lambda+\frac{1}{2^\lambda} \qquad (89)
$$

$$
P_f^+ = \frac{\lambda}{(2d_t)^\lambda}\int_{x_t+\theta d_t}^{x_t+d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z = \left(\frac{1-\theta}{2}\right)^\lambda \qquad (90)
$$

$$
\begin{aligned}
E^{A_t}(d_{t+1}) &= d_t\left\{\frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^\lambda-\left(\frac{x_t}{d_t}\right)^\lambda\right](\alpha_s^+-\alpha_s^-)+\frac{1}{2^\lambda}(\alpha_f^--\alpha_s^-)\right.\\
&\left.+\left(\frac{1-\theta}{2}\right)^\lambda(\alpha_f^+-\alpha_f^-)+\alpha_s^-\right\} \qquad (91)
\end{aligned}
$$

$$
\begin{aligned}
E^{A_t}[\ln(d_{t+1})] &= \ln(d_t)+\frac{1}{2}\left[\left(1-\theta+\frac{x_t}{d_t}\right)^\lambda-\left(\frac{x_t}{d_t}\right)^\lambda\right]\ln\left(\frac{\alpha_s^+}{\alpha_s^-}\right)+\frac{1}{2^\lambda}\ln\left(\frac{\alpha_f^-}{\alpha_s^-}\right)\\
&+ \left(\frac{1-\theta}{2}\right)^\lambda\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)+\ln(\alpha_s^-) \qquad (92)
\end{aligned}
$$

Se $\boldsymbol{x_t-\theta d_t>0}$, as possibilidades de ordenação são:

6) $x_t-d_t<-x_t<0<x_t-\theta d_t<x_t<x_t+\theta d_t<-x_t+d_t<x_t+d_t$

Neste caso $\theta\le\dfrac{x_t}{d_t}<\dfrac{1-\theta}{2}$, se $\theta<\dfrac{1}{3}$, com

$$
\begin{aligned}
P_s^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{-x_t}^{0}(2d_t+2z)^{\lambda-1}\mathrm{d}z+\int_{0}^{x_t-\theta d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z\right]\\
&= 1-\frac{1}{2}\left[\left(1+\theta-\frac{x_t}{d_t}\right)^\lambda+\left(1-\frac{x_t}{d_t}\right)^\lambda\right] \qquad (93)
\end{aligned}
$$

$$P_s^- = \frac{\lambda}{(2d_t)^\lambda}\int_{x_t-\theta d_t}^{x_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z = \frac{1}{2}\left[\left(-\frac{x_t}{d_t}+1+\theta\right)^\lambda - \left(1-\frac{x_t}{d_t}\right)^\lambda\right] \tag{94}$$

P_f^- e P_f^+ descritas respectivamente pelas Equações (79) e (80),

$$\begin{aligned} E^{A_t}(d_{t+1}) &= d_t\left\{\frac{1}{2}\left[\left(1+\theta-\frac{x_t}{d_t}\right)^\lambda(\alpha_s^- - \alpha_s^+) + \left(1-\theta-\frac{x_t}{d_t}\right)^\lambda(\alpha_f^+ - \alpha_f^-)\right.\right. \\ &\quad \left.\left. + \left(1-\frac{x_t}{d_t}\right)^\lambda(\alpha_f^+ + \alpha_f^- - \alpha_s^+ - \alpha_s^-)\right] + \alpha_s^+\right\} \end{aligned} \tag{95}$$

$$\begin{aligned} E^{A_t}[\ln(d_{t+1})] &= \ln(d_t) + \frac{1}{2}\left[\left(1+\theta-\frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_s^-}{\alpha_s^+}\right) + \left(1-\theta-\frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)\right. \\ &\quad \left. + \left(1-\frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_f^+ + \alpha_f^-}{\alpha_s^+ - \alpha_s^-}\right)\right] + \ln(\alpha_s^+) \end{aligned} \tag{96}$$

7) $x_t - d_t < -x_t < 0 < x_t - \theta d_t < x_t < -x_t + d_t < x_t + \theta d_t < x_t + d_t$

Neste caso $\max\left\{\frac{1-\theta}{2}, \theta\right\} \leq \frac{x_t}{d_t} < \frac{1}{2}$, $\left(\text{se } \theta < \frac{1}{2}\right)$, e P_s^+, P_s^-, P_f^- e P_f^+ são descritas respectivamente pelas Equações (93), (94), (83) e (84).

$$\begin{aligned} E^{A_t}(d_{t+1}) &= d_t\left\{\frac{1}{2}\left[\left(1+\theta-\frac{x_t}{d_t}\right)^\lambda(\alpha_s^- - \alpha_s^+) + \left(\frac{x_t}{d_t}\right)^\lambda(\alpha_f^- - \alpha_f^+)\right.\right. \\ &\quad \left. + \left(1-\frac{x_t}{d_t}\right)^\lambda(\alpha_f^+ + \alpha_f^- - \alpha_s^+ - \alpha_s^-)\right] \\ &\quad \left. + \left(\frac{1-\theta}{2}\right)^\lambda(\alpha_f^+ - \alpha_f^-) + \alpha_s^+\right\} \end{aligned} \tag{97}$$

$$\begin{aligned} E^{A_t}[\ln(d_{t+1})] &= \ln(d_t) + \frac{1}{2}\left[\left(1+\theta-\frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_s^-}{\alpha_s^+}\right) + \left(\frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_f^-}{\alpha_f^+}\right)\right. \\ &\quad \left. + \left(1-\frac{x_t}{d_t}\right)^\lambda \ln\left(\frac{\alpha_f^+\alpha_f^-}{\alpha_s^+ - \alpha_s^-}\right)\right] \\ &+ \left(\frac{1-\theta}{2}\right)^\lambda \ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right) + \ln(\alpha_s^+) \end{aligned} \tag{98}$$

8) $-x_t < x_t - d_t < 0 < x_t - \theta d_t < -x_t + d_t < x_t < x_t + \theta d_t < x_t + d_t$

Neste caso $\max\left\{\theta, \frac{1}{2}\right\} \leq \frac{x_t}{d_t} < \frac{1+\theta}{2}$, com

$$\begin{aligned} P_s^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{0}(2d_t+2z)^{\lambda-1}\mathrm{d}z + \int_{0}^{x_t-\theta d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z\right] \\ &= -\frac{1}{2}\left[\left(\frac{x_t}{d_t}\right)^\lambda + \left(1+\theta-\frac{x_t}{d_t}\right)^\lambda\right] + 1 \end{aligned} \tag{99}$$

$$\begin{aligned} P_s^- &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-\theta d_t}^{-x_t+d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z + \int_{-x_t+d_t}^{x_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \\ &= \frac{1}{2}\left[\left(\frac{x_t}{d_t}\right)^\lambda + \left(1+\theta-\frac{x_t}{d_t}\right)^\lambda\right] - \frac{1}{2^\lambda} \end{aligned} \tag{100}$$

$$P_f^- = \frac{\lambda}{(2d_t)^\lambda}\int_{x_t}^{x_t+\theta d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z = -\left(\frac{1-\theta}{2}\right)^\lambda + \frac{1}{2^\lambda} \tag{101}$$

$$P_f^+ = \frac{\lambda}{(2d_t)^\lambda}\int_{x_t+\theta d_t}^{x_t+d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z = \left(\frac{1-\theta}{2}\right)^\lambda \tag{102}$$

$$\begin{aligned} E^{A_t}(d_{t+1}) &= d_t\left\{\frac{1}{2}\left[\left(1+\theta-\frac{x_t}{d_t}\right)^\lambda + \left(\frac{x_t}{d_t}\right)^\lambda\right](\alpha_s^- - \alpha_s^+)\right. \\ &\quad \left. + \left(\frac{1-\theta}{2}\right)^\lambda(\alpha_f^+ - \alpha_f^-) + \frac{1}{2^\lambda}(\alpha_f^- - \alpha_s^-) + \alpha_s^+\right\} \end{aligned} \tag{103}$$

$$\begin{aligned} E^{A_t}[\ln(d_{t+1})] &= \ln(d_t) + \frac{1}{2}\left[\left(1+\theta-\frac{x_t}{d_t}\right)^\lambda + \left(\frac{x_t}{d_t}\right)^\lambda\right]\ln\left(\frac{\alpha_s^-}{\alpha_s^+}\right) \\ &+ \left(\frac{1-\theta}{2}\right)^\lambda \ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right) + \frac{1}{2^\lambda}\ln\left(\frac{\alpha_f^-}{\alpha_s^-}\right) + \ln(\alpha_s^+) \end{aligned} \tag{104}$$

9) $-x_t < x_t - d_t < 0 < -x_t + d_t < x_t - \theta d_t < x_t < x_t + \theta d_t < x_t + d_t$

Neste caso $\frac{1+\theta}{2} \leq \frac{x_t}{d_t} < 1$, com

$$
\begin{aligned}
P_s^+ &= \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{0}(2d_t+2z)^{\lambda-1}\mathrm{d}z+\int_{0}^{-x_t+d_t}(2d_t-2z)^{\lambda-1}\mathrm{d}z\right]\\
&+ \frac{\lambda}{(2d_t)^\lambda}\int_{-x_t+d_t}^{x_t-\theta d_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z=-\left(\frac{1+\theta}{2}\right)^\lambda+1
\end{aligned}
\tag{105}
$$

$$
P_s^- = \frac{\lambda}{(2d_t)^\lambda}\int_{x_t-\theta d_t}^{x_t}(x_t+d_t-z)^{\lambda-1}\mathrm{d}z=\left(\frac{1+\theta}{2}\right)^\lambda-\frac{1}{2^\lambda}
\tag{106}
$$

e P_f^- e P_f^+ são descritas respectivamente pelas Equações (101) e (102),

$$
E^{A_t}(d_{t+1})=\frac{d_t}{2^\lambda}\left[(1+\theta)^\lambda(\alpha_s^- - \alpha_s^+)+(1-\theta)^\lambda(\alpha_f^+ - \alpha_f^-)+\alpha_f^- - \alpha_s^-\right]+d_t\alpha_s^+
\tag{107}
$$

$$
\begin{aligned}
E^{A_t}[\ln(d_{t+1})] &= \ln(d_t)+\frac{1}{2^\lambda}\left[(1+\theta)^\lambda\ln\left(\frac{\alpha_s^-}{\alpha_s^+}\right)+(1-\theta)^\lambda\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)\right.\\
&\left.+\ln\left(\frac{\alpha_f^-}{\alpha_s^-}\right)\right]+\ln(\alpha_s^+)
\end{aligned}
\tag{108}
$$

Se $x_t - d_t > 0$, a única possibilidade de ordenação é:

10) $-x_t < -x_t + d_t < 0 < x_t - d_t < x_t - \theta d_t < x_t < x_t + \theta d_t < x_t + d_t$

Neste caso $\frac{x_t}{d_t} \geq 1$, e P_s^+, P_s^-, P_f^- e P_f^+ são descritas respectivamente pelas Equações (105), (106), (101) e (102).

$E^{A_t}(d_{t+1})$ e $E^{A_t}[\ln(d_{t+1})]$ também são descritas respectivamente pelas Equações (107) e (108).

$\Psi(r)$ é, portanto, descrita em sete intervalos para as duas seleções conforme a Tabela 10.

Tabela 10 – Intervalos para expressões de $\Psi(r)$ de acordo com os valores de θ para $(1 \overset{+}{,} \lambda)$-EE com controle combinado.

Intervalos	**Caso 1**: $\theta \leq \frac{1}{3}$	**Caso 2**: $\frac{1}{3} < \theta \leq \frac{1}{2}$	**Caso 3**: $\theta > \frac{1}{2}$
(A)	$0 \leq r < \frac{\theta}{2}$	$0 \leq r < \frac{\theta}{2}$	$0 \leq r < \frac{1-\theta}{2}$
(B)	$\frac{\theta}{2} \leq r < \theta$	$\frac{\theta}{2} \leq r < \frac{1-\theta}{2}$	$\frac{1-\theta}{2} \leq r < \frac{\theta}{2}$
(C)	$\theta \leq r < \frac{1-\theta}{2}$	$\frac{1-\theta}{2} \leq r < \theta$	$\frac{\theta}{2} \leq r < \frac{1}{2}$
(D)	$\frac{1-\theta}{2} \leq r < \frac{1}{2}$	$\theta \leq r < \frac{1}{2}$	$\frac{1}{2} \leq r < \theta$
(E)	$\frac{1}{2} \leq r < \frac{1+\theta}{2}$	$\frac{1}{2} \leq r < \frac{1+\theta}{2}$	$\theta \leq r < \frac{1+\theta}{2}$
(F)	$\frac{1+\theta}{2} \leq r < 1$	$\frac{1+\theta}{2} \leq r < 1$	$\frac{1+\theta}{2} \leq r < 1$
(G)	$r \geq 1$	$r \geq 1$	$r \geq 1$

Apresentamos as expressões que definem $\Psi(r)$ apenas para a seleção $(1,\lambda)$-EE, que são derivadas das Equações (26), (27) com (71) à (108). Para a seleção $(1+\lambda)$-EE, Ψ é representada para os mesmos intervalos combinando as Equações (28), (29) e (30) com (71) a (108). As expressões não são exibidas, pois diferem da seleção não-elitista em apenas uma pequena parcela do primeiro logaritmando, nos valores referentes à esperança $E^{A_t}(|x_{t+1}|)$.

Para o intervalo A, $\Psi(r)$ possui mesma expressão para os três casos:

$$\begin{aligned}
\Psi_A(r) \quad = \quad & \ln\left\{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1) + \frac{w}{2}[(1-\theta+r)^{\lambda} + (1-\theta-r)^{\lambda}](\alpha_f^{+} - \alpha_f^{-})\right. \\
& \left. + w[(1-r)^{\lambda}(\alpha_f^{-} - \alpha_s^{-}) + \alpha_s^{-}]\right\} - \ln(w+r) \\
& -k\left\{\frac{1}{2}[(1-\theta+r)^{\lambda} + (1-\theta-r)^{\lambda}]\ln\left(\frac{\alpha_f^{+}}{\alpha_f^{-}}\right)\right. \\
& \left. + (1-r)^{\lambda}\ln\left(\frac{\alpha_f^{-}}{\alpha_s^{-}}\right) + \ln(\alpha_s^{-})\right\} \qquad (109)
\end{aligned}$$

Para o intervalo B, Ψ possui mesma expressão apenas nos casos 1 e 2, sendo denotada por $\Psi_{B_1}(r)$. Para o caso 3, denotamos Ψ por $\Psi_{B_2}(r)$.

$$
\begin{aligned}
\Psi_{B_1}(r) &= \ln\left\{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1)+\frac{w}{2}[(1-\theta+r)^{\lambda}(\alpha_s^+-\alpha_s^-)+(1-\theta-r)^{\lambda}(\alpha_f^+-\alpha_f^-)]\right. \\
&\quad \left.+\,w[(1-r)^{\lambda}(\alpha_f^++\alpha_f^--\alpha_s^--\alpha_s^+)+\alpha_s^-]\right\}-\ln(w+r) \\
&\quad -k\left\{\frac{1}{2}\left[(1-\theta+r)^{\lambda}\ln\left(\frac{\alpha_s^+}{\alpha_s^-}\right)+(1-\theta-r)^{\lambda}\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)\right.\right. \\
&\quad \left.\left.+\,(1-r)^{\lambda}\ln\left(\frac{\alpha_f^+\alpha_f^-}{\alpha_s^-\alpha_s^+}\right)\right]+\ln(\alpha_s^-)\right\}
\end{aligned} \tag{110}
$$

$$
\begin{aligned}
\Psi_{B_2}(r) &= \ln\left\{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1)+w\left[\frac{(1-\theta+r)^{\lambda}-r^{\lambda}}{2}+\left(\frac{1-\theta}{2}\right)^{\lambda}\right](\alpha_f^+-\alpha_f^-)\right. \\
&\quad \left.+\,w[(1-r)^{\lambda}(\alpha_f^--\alpha_s^-)+\alpha_s^-]\right\}-\ln(w+r) \\
&\quad -k\left\{\left[\frac{(1-\theta+r)^{\lambda}-r^{\lambda}}{2}+\left(\frac{1-\theta}{2}\right)^{\lambda}\right]\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)\right. \\
&\quad \left.+\,(1-r)^{\lambda}\ln\left(\frac{\alpha_f^-}{\alpha_s^-}\right)+\ln(\alpha_s^-)\right\}
\end{aligned} \tag{111}
$$

Para o intervalo C, Ψ possui mesma expressão apenas nos casos 2 e 3, sendo denotada por $\Psi_{C_2}(r)$. Para o caso 1, denotamos Ψ por $\Psi_{C_1}(r)$.

$$
\begin{aligned}
\Psi_{C_1}(r) &= \ln\left\{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1)+\frac{w}{2}[(1+\theta-r)^{\lambda}(\alpha_s^--\alpha_s^+)+(1-\theta-r)^{\lambda}(\alpha_f^+-\alpha_f^-)\right. \\
&\quad \left.+\,(1-r)^{\lambda}(\alpha_f^-+\alpha_f^+-\alpha_s^+-\alpha_s^-)]+w\alpha_s^+\right\}-\ln(w+r) \\
&\quad -k\left\{\frac{1}{2}\left[(1+\theta-r)^{\lambda}\ln\left(\frac{\alpha_s^-}{\alpha_s^+}\right)+(1-\theta-r)^{\lambda}\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)\right.\right. \\
&\quad \left.\left.+\,(1-r)^{\lambda}\ln\left(\frac{\alpha_f^+\alpha_f^-}{\alpha_s^-\alpha_s^+}\right)\right]+\ln(\alpha_s^+)\right\}
\end{aligned} \tag{112}
$$

$$
\begin{aligned}
\Psi_{C_2}(r) &= \ln\left\{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1)+\frac{w}{2}[(1-\theta+r)^{\lambda}(\alpha_s^+-\alpha_s^-)+r^{\lambda}(\alpha_f^--\alpha_f^+)\right. \\
&\quad \left.+\,(1-r)^{\lambda}(\alpha_f^-+\alpha_f^+-\alpha_s^+-\alpha_s^-)]+w\left[\left(\frac{1-\theta}{2}\right)^{\lambda}(\alpha_f^+-\alpha_f^-)+\alpha_s^-\right]\right\} \\
&\quad -\ln(w+r)-\frac{k}{2}\left[(1-\theta+r)^{\lambda}\ln\left(\frac{\alpha_s^+}{\alpha_s^-}\right)+r^{\lambda}\ln\left(\frac{\alpha_f^-}{\alpha_f^+}\right)\right. \\
&\quad \left.+\,(1-r)^{\lambda}\ln\left(\frac{\alpha_f^+\alpha_f^-}{\alpha_s^-\alpha_s^+}\right)\right]-k\left[\left(\frac{1-\theta}{2}\right)^{\lambda}\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)+\ln(\alpha_s^-)\right]
\end{aligned} \tag{113}
$$

Para o intervalo D, Ψ possui mesma expressão apenas nos casos 1 e 2, sendo denotada por $\Psi_{D_1}(r)$. Para o caso 3, denotamos Ψ por $\Psi_{D_2}(r)$.

$$\begin{aligned}\Psi_{D_1}(r) &= \ln\left\{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1)+\frac{w}{2}[(1+\theta-r)^{\lambda}(\alpha_s^- - \alpha_s^+)+r^{\lambda}(\alpha_f^- - \alpha_f^+)\right. \\ &\quad \left. +(1-r)^{\lambda}(\alpha_f^- + \alpha_f^+ - \alpha_s^+ - \alpha_s^-)]+w\left[\left(\frac{1-\theta}{2}\right)^{\lambda}(\alpha_f^+ - \alpha_f^-)+\alpha_s^+\right]\right\} \\ &\quad -\ln(w+r)-\frac{k}{2}\left[(1+\theta-r)^{\lambda}\ln\left(\frac{\alpha_s^-}{\alpha_s^+}\right)+r^{\lambda}\ln\left(\frac{\alpha_f^-}{\alpha_f^+}\right)\right. \\ &\quad \left. +(1-r)^{\lambda}\ln\left(\frac{\alpha_f^+\alpha_f^-}{\alpha_s^-\alpha_s^+}\right)\right]-k\left[\left(\frac{1-\theta}{2}\right)^{\lambda}\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)+\ln(\alpha_s^+)\right] \end{aligned} \quad (114)$$

$$\begin{aligned}\Psi_{D_2}(r) &= \ln\left\{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1)+\frac{w}{2}[(1-\theta+r)^{\lambda}-r^{\lambda}](\alpha_s^+ - \alpha_s^-)\right. \\ &\quad \left. +w\left[\frac{1}{2^{\lambda}}(\alpha_f^- - \alpha_s^-)+\left(\frac{1-\theta}{2}\right)^{\lambda}(\alpha_f^+ - \alpha_f^-)+\alpha_s^-\right]\right\} \\ &\quad -\ln(w+r)-k\left\{\frac{1}{2}[(1-\theta+r)^{\lambda}-r^{\lambda}]\ln\left(\frac{\alpha_s^+}{\alpha_s^-}\right)+\frac{1}{2^{\lambda}}\ln\left(\frac{\alpha_f^-}{\alpha_s^-}\right)\right. \\ &\quad \left. +\left(\frac{1-\theta}{2}\right)^{\lambda}\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)+\ln(\alpha_s^-)\right\} \end{aligned} \quad (115)$$

Para os intervalos E, F e G, $\Psi(r)$ possui mesma expressão para os três casos:

$$\begin{aligned}\Psi_{E}(r) &= \ln\left\{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1)+\frac{w}{2}[(1+\theta-r)^{\lambda}+r^{\lambda}](\alpha_s^- - \alpha_s^+)\right. \\ &\quad \left. +w\left[\frac{1}{2^{\lambda}}(\alpha_f^- - \alpha_s^-)+\left(\frac{1-\theta}{2}\right)^{\lambda}(\alpha_f^+ - \alpha_f^-)+\alpha_s^+\right]\right\}-\ln(w+r) \\ &\quad -k\left\{\frac{1}{2}[(1+\theta-r)^{\lambda}+r^{\lambda}]\ln\left(\frac{\alpha_s^-}{\alpha_s^+}\right)+\frac{1}{2^{\lambda}}\ln\left(\frac{\alpha_f^-}{\alpha_s^-}\right)\right. \\ &\quad \left. +\left(\frac{1-\theta}{2}\right)^{\lambda}\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)+\ln(\alpha_s^+)\right\} \end{aligned} \quad (116)$$

$$\begin{aligned}\Psi_{F}(r) &= \ln\left\{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1)+\frac{w}{2^{\lambda}}[(1+\theta)^{\lambda}(\alpha_s^- - \alpha_s^+)\right. \\ &\quad \left. +(1-\theta)^{\lambda}(\alpha_f^+ - \alpha_f^-)+\alpha_f^- - \alpha_s^-]+w\alpha_s^+\right\}-\ln(w+r) \\ &\quad -\frac{k}{2^{\lambda}}\left[(1+\theta)^{\lambda}\ln\left(\frac{\alpha_s^-}{\alpha_s^+}\right)+(1-\theta)^{\lambda}\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)+\ln\left(\frac{\alpha_f^-}{\alpha_s^-}\right)\right] \\ &\quad -k\ln(\alpha_s^+) \end{aligned} \quad (117)$$

$$
\begin{aligned}
\Psi_G(r) \;=\; & \ln\left\{r-\left(\frac{\lambda-1}{\lambda+1}\right)+\frac{w}{2^\lambda}[(1+\theta)^\lambda(\alpha_s^- - \alpha_s^+)\right. \\
& \left.+(1-\theta)^\lambda(\alpha_f^+ - \alpha_f^-)+\alpha_f^- - \alpha_s^-]+w\alpha_s^+\right\}-\ln(w+r) \\
& -\frac{k}{2^\lambda}\left[(1+\theta)^\lambda\ln\left(\frac{\alpha_s^-}{\alpha_s^+}\right)+(1-\theta)^\lambda\ln\left(\frac{\alpha_f^+}{\alpha_f^-}\right)+\ln\left(\frac{\alpha_f^-}{\alpha_s^-}\right)\right] \\
& -k\ln(\alpha_s^+)
\end{aligned} \tag{118}
$$

Para mostrar que $\Psi(r) \leq -a$ para todo $r \geq 0$ e algum $a > 0$, é suficiente verificar esta condição nos extremos de cada intervalo A, B, C, D, E, F e G ($r = 0,\ \frac{1}{2},\ \frac{\theta}{2},\ \theta,\ \frac{1-\theta}{2},\ \frac{1+\theta}{2},\ 1$ e $r \to +\infty$) e em quaisquer pontos críticos dentro desses intervalos ($r = r^*$ tais que $\Psi'(r^*) = 0$).

Observamos que, independentemente do esquema de seleção, assim como para as duas adaptações, $\Psi(r)$ é representado em cada intervalo como uma soma de logaritmos de frações racionais, e que $\Psi'(r)$ é representado por uma fração racional, na variável r. Portanto, os pontos críticos dentro de cada intervalo, e os valores correspondentes de Ψ podem ser determinados numericamente para quaisquer valores dados dos parâmetros da estratégia, e o seguinte problema de otimização não-linear restrito pode ser formulado:

$$
((\alpha_s^+)^*,(\alpha_s^-)^*,(\alpha_f^-)^*,(\alpha_f^+)^*,\theta^*,w^*,k^*,a^*) = \text{arg max}_{\alpha_s^+,\alpha_s^-,\alpha_f^-,\alpha_f^+,\theta,w,k,a}\, a
$$

$$
\text{sujeito a: }\begin{cases}
(\alpha_s^+,\alpha_s^-,\alpha_f^-,\alpha_f^+) > 0 \\
0 < \theta < 1 \\
w > 0 \\
0 \leq k < 1 \\
a > 0 \\
\Psi(r) + a \leq 0, & r = 0,\ \frac{1}{2},\ \frac{\theta}{2},\ \theta,\ \frac{1-\theta}{2},\ \frac{1+\theta}{2},\ 1,\ +\infty \\
\Psi(r^*) + a \leq 0, & r^* : \Psi'(r^*) = 0
\end{cases} \tag{119}
$$

Notamos que qualquer solução viável para a **Regra 1** com ambas as seleções é também uma solução viável para a combinação, considerando $\alpha_s^+ = \alpha_s^-$ e $\alpha_f^+ = \alpha_f^-$. Além disso qualquer solução exequível para **Regra 3** é também uma solução exequível para a combinação, tomando $\alpha_s^+ = \alpha_f^+$ e $\alpha_s^- = \alpha_f^-$ para a seleção $(1,\lambda)$-EE e $\alpha_s^- = \alpha_f^- = \alpha_f^+$ para a seleção $(1+\lambda)$-EE.

Nas Tabelas 11 e 12 apresentamos os valores ótimos dos parâmetros de adaptação do controle combinado e dos parâmetros da função de Lyapunov que minimizam o limite superior para a taxa de convergência e^{-a}, obtidos variando λ.

Tabela 11 – Valores dos parâmetros ótimos para $(1,\lambda)$-EE com adaptação combinada.

$(1,\lambda)$-ES	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$
α_s^+	1.5729	2.3606	2.2971	2.3468	2.4113
α_s^-	1.0644	0.94630	0.81903	0.69802	0.61011
α_f^-	0.67427	0.42549	0.30426	0.24425	0.21057
α_f^+	0.65111	0.37757	0.28140	0.23542	0.20704
θ	0.72498	0.69814	0.61173	0.58659	0.58153
w	1.8015	0.95106	0.75701	0.64312	0.57311
k	0.34409	0.34217	0.35117	0.35713	0.36204
a	0.016392	0.068111	0.11754	0.15687	0.19099
a/λ	0.008196	0.022704	0.029385	0.031374	0.031832

Tabela 12 – Valores dos parâmetros ótimos para $(1+\lambda)$-EE com adaptação combinada.

$(1+\lambda)$-ES	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$
α_s^+	5.8704	5.9619	6.0796	6.0961	6.3250
α_s^-	1.1357	0.90148	0.73171	0.60207	0.51172
α_f	0.75134	0.57881	0.45740	0.37239	0.31569
θ	0.51413	0.52547	0.53798	0.54568	0.57016
w	0.093490	0.10348	0.11391	0.12348	0.13243
k	0.23328	0.24632	0.25869	0.27045	0.28215
a	0.074589	0.14179	0.20265	0.25740	0.30457
a/λ	0.074589	0.070895	0.06755	0.06435	0.060914

A Figura 21 mostra as funções correspondentes $\Psi(r)/\lambda$ para as duas estratégias de seleção e regra de adaptação do tamanho do passo combinada. Essas funções representam um limite superior (devido à desigualdade de Jensen) no decaimento esperado do processo estocástico V_t por avaliação da função objetivo, em função da distância ao ótimo normalizada, r.

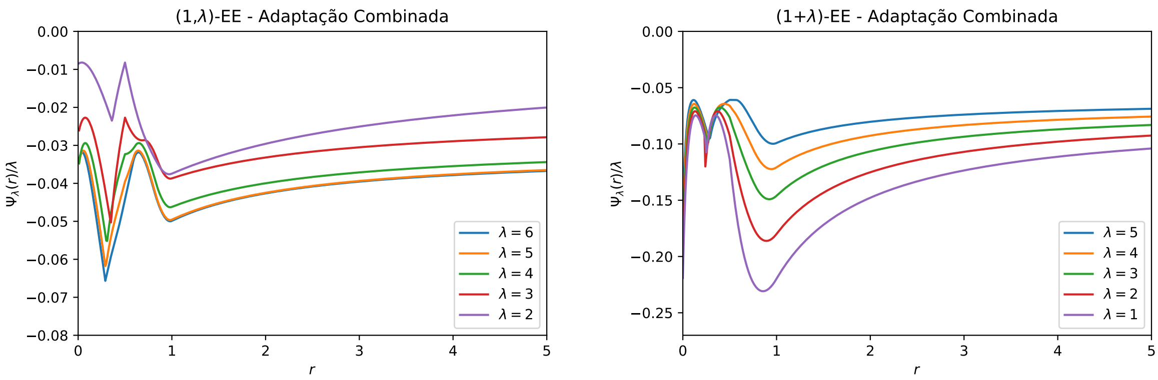

Figura 21 – Funções $\Psi_\lambda(r)/\lambda$ para adaptação combinada.

Observamos que os gráficos de Ψ/λ para a adaptação combinada são semelhantes aos mesmos gráficos para a **Regra 3** , principalmente no caso da seleção elitista, o que sugere que o controle combinado sofra maior influência da **Regra 3**. Os limites para as taxas de convergência apresentados nas tabelas com os valores ótimos dos parâmetros também apontam nesse sentido. De fato, comparando as Tabelas 8 e 12, notamos que a diferença entre os limites da taxa de convergência é da ordem de $0{,}01$. Na próxima subseção veremos que limites reais para a **Regra 3** superam os limites reais para a adaptação combinada.

5.1.1 Resultados Experimentais e Discussões

Apresentamos nesta seção uma análise experimental dos benefícios do controle combinado em relação às Regras 1 e 3. Todos os quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t ao longo das gerações para $(1 \overset{+}{,} \lambda)$-EE com controle combinado podem ser vistos no Apêndice F.3.

Analisando a Figura 22 verificamos que a adaptação combinada apresenta convergência mais rápida que a **Regra 1**. Na comparação entre o controle combinado e a **Regra 3**, a primeira estratégia converge mais rápido que a segunda para $\lambda = 2$, pouco mais rápido para $\lambda = 3$ e mais devagar para $\lambda = 4, 5, 6$, apesar de apresentar melhor limite teórico (ver Figura 23). O mesmo comportamento é observado nos experimentos em que o tamanho do passo inicial é muito pequeno.

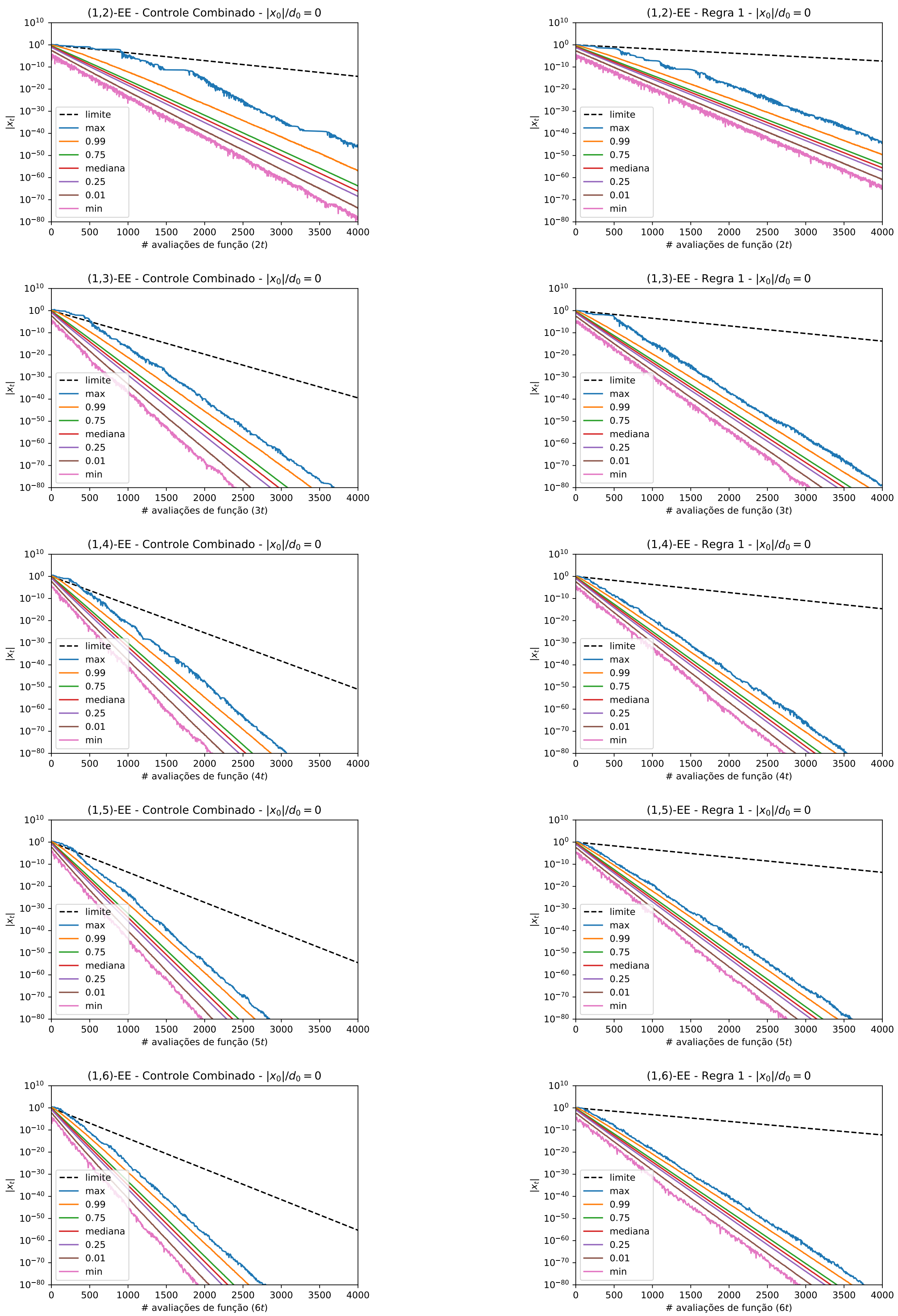

Figura 22 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1{,}\lambda)$-EE com adaptação combinada e **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

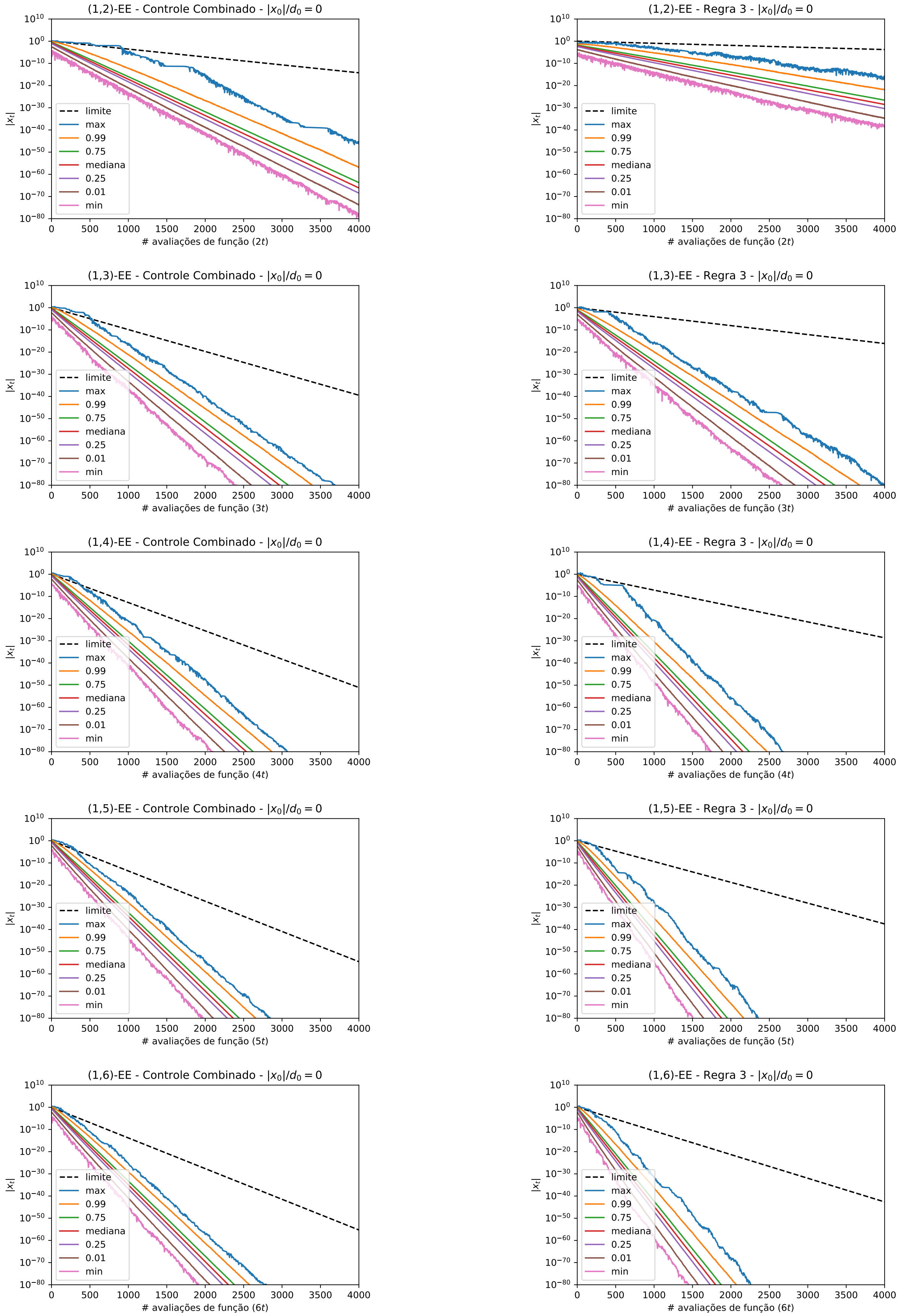

Figura 23 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1,\lambda)$-EE com adaptação combinada e **Regra 3** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Para a seleção elitista, a velocidade de convergência não aumenta com o crescimento de λ. Pelo contrário, à medida que λ aumenta, $(1+\lambda)$-EE converge ligeiramente mais devagar, mesmo nos experimentos em que o tamanho do passo inicial é muito pequeno. Comparando as configurações com os valores ótimos de λ, observamos que a adaptação combinada apresenta convergência mais rápida que a **Regra 1** e que a **Regra 3**, tendo comportamento quase indistinguível ao desta última (ver Figura 24).

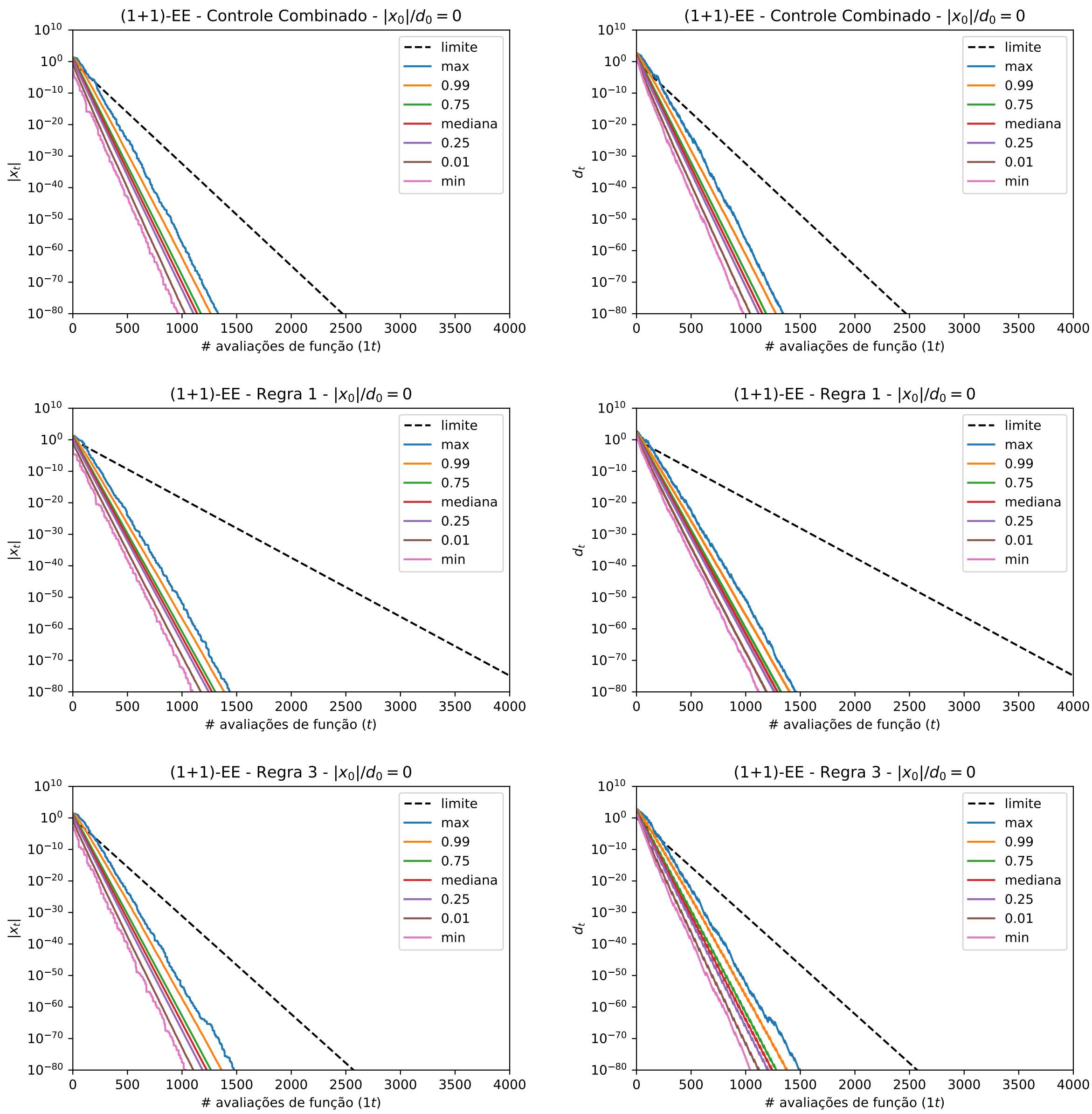

Figura 24 – Quantis empíricos das distribuições de $|x_t|$ e d_t estimadas de 10001 execuções de $(1+1)$-EE com adaptação combinada, **Regra 1** e **Regra 3** , iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Tendo em vista os aspectos observados, o método de demonstração de convergência de EEs empregado em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016; SEMENOV;

TERKEL, 2003) também pode ser utilizado na análise de uma abordagem unificada da regra de adaptação do tamanho do passo baseada em sucesso, associada com o controle de mutação baseado no tamanho do passo. Esta regra combinada gerou melhores limites teóricos para a taxa de convergência, apesar dos experimentos não indicarem vantagem na comparação com a estratégia $(1,\lambda)$-EE, formulada no Capítulo 4 para determinados valores de λ. Também podemos perceber influência preponderante da **Regra 3** sobre a combinação, quando analisamos as estratégias $(1+\lambda)$-EE, comparando os limites teóricos e os resultados experimentais.

5.2 Em Busca de Resultados Menos Conservadores

Para todas a formulações de EEs apresentadas até o momento, o limite superior para a taxa de convergência, assim como os parâmetros que maximizam tal limite, foram estimados com o auxílio da desigualdade de Jensen para obter uma cota superior para $E^{A_t}(V_{t+1}) - V_t$. Acreditamos que a utilização desta desigualdade é um dos motivos para o alcance de limites teóricos que prevêem para os algoritmos uma convergência muito mais lenta do que a convergência apresentada pelos resultados experimentais. Esta seção é dedicada à busca de limites superiores menos conservadores para a taxa de convergência das Estratégias Evolutivas formuladas nos capítulos anteriores e na seção anterior.

Utilizando a mesma metodologia descrita no final do Capítulo 2, a única tarefa distinta a ser realizada é o cálculo explícito de

$$E^{A_t}(V_{t+1}) - V_t = E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] - \ln(|x_t| + wd_t) - k\{E^{A_t}[\ln(d_{t+1})] - \ln(d_t)\}$$

que demanda a avaliação de $E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})]$ e não mais de $E^{A_t}(|x_{t+1}|)$ e $E^{A_t}(d_{t+1})$ como anteriormente. Para evitar ambiguidades, denotamos $E^{A_t}(V_{t+1}) - V_t$ por $\Phi(r)$, com $r = \frac{|x_t|}{d_t}$.

O cálculo da esperança $E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})]$ é realizado apenas para $(1,\lambda)$-EE com controle combinado, da mesma forma como foram obtidas as expressões de $E^{A_t}(d_{t+1})$. Claramente é possível descrever estas expressões para cada uma das adaptações separadamente a partir da esperança da combinação.

Dessa forma, $\Phi(r)$ pode ser descrita por sete expressões para os intervalos indicados na Tabela 10, apresentada na Seção 5.1.

As expressões de $E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})]$ para cada um dos intervalos da Tabela 10 são dadas por:

(A) $0 \leq r < \min\left\{\frac{\theta}{2}, \frac{1-\theta}{2}\right\}$

$$\begin{aligned}
E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \;=\; & \frac{\lambda}{(2d_t)^{\lambda}}\left[\int_{x_t-d_t}^{x_t-\theta d_t} \ln(-z + wd_t\alpha_f^+)(2d_t+2z)^{\lambda-1}\mathrm{d}z\right. \\
& + \int_{x_t-\theta d_t}^{-x_t} \ln(-z + wd_t\alpha_f^-)(2d_t+2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{-x_t}^{0} \ln(-z + wd_t\alpha_s^-)(2d_t+2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{0}^{x_t} \ln(z + wd_t\alpha_s^-)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t}^{x_t+\theta d_t} \ln(z + wd_t\alpha_f^-)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t+\theta d_t}^{-x_t+d_t} \ln(z + wd_t\alpha_f^+)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& \left.+ \int_{-x_t+d_t}^{x_t+d_t} \ln(z + wd_t\alpha_f^+)(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \qquad (120)
\end{aligned}$$

(B) para $\theta > 1/2$**:** $\frac{\theta}{2} \leq r < \min\left\{\theta, \frac{1-\theta}{2}\right\}$

$$\begin{aligned}
E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \;=\; & \frac{\lambda}{(2d_t)^{\lambda}}\left[\int_{x_t-d_t}^{-x_t} \ln(-z + wd_t\alpha_f^+)(2d_t+2z)^{\lambda-1}\mathrm{d}z\right. \\
& + \int_{-x_t}^{x_t-\theta d_t} \ln(-z + wd_t\alpha_s^+)(2d_t+2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t-\theta d_t}^{0} \ln(-z + wd_t\alpha_s^-)(2d_t+2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{0}^{x_t} \ln(z + wd_t\alpha_s^-)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t}^{x_t+\theta d_t} \ln(z + wd_t\alpha_f^-)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t+\theta d_t}^{-x_t+d_t} \ln(z + wd_t\alpha_f^+)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& \left.+ \int_{-x_t+d_t}^{x_t+d_t} \ln(z + wd_t\alpha_f^+)(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\right] \qquad (121)
\end{aligned}$$

(B) para $\theta > 1/2$**:** $\dfrac{1-\theta}{2} \leq r < \dfrac{\theta}{2}$

$$
\begin{aligned}
E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \quad = \quad & \frac{\lambda}{(2d_t)^\lambda}\Bigg[\int_{x_t-d_t}^{x_t-\theta d_t} \ln(-z + wd_t\alpha_f^+)(2d_t+2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t-\theta d_t}^{-x_t} \ln(-z + wd_t\alpha_f^-)(2d_t+2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{-x_t}^{0} \ln(-z + wd_t\alpha_s^-)(2d_t+2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{0}^{x_t} \ln(z + wd_t\alpha_s^-)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t}^{-x_t+d_t} \ln(z + wd_t\alpha_f^-)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{-x_t+d_t}^{x_t+\theta d_t} \ln(z + wd_t\alpha_f^-)(x_t+d_t-z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t+\theta d_t}^{x_t+d_t} \ln(z + wd_t\alpha_f^+)(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\Bigg]
\end{aligned} \tag{122}
$$

(C) para $\theta < 1/3$**:** $\theta \leq r < \dfrac{1-\theta}{2}$

$$
\begin{aligned}
E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \quad = \quad & \frac{\lambda}{(2d_t)^\lambda}\Bigg[\int_{x_t-d_t}^{-x_t} \ln(-z + wd_t\alpha_f^+)(2d_t+2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{-x_t}^{0} \ln(-z + wd_t\alpha_s^+)(2d_t+2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{0}^{x_t-\theta d_t} \ln(z + wd_t\alpha_s^+)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t-\theta d_t}^{x_t} \ln(z + wd_t\alpha_s^-)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t}^{x_t+\theta d_t} \ln(z + wd_t\alpha_f^-)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t+\theta d_t}^{-x_t+d_t} \ln(z + wd_t\alpha_f^+)(2d_t-2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{-x_t+d_t}^{x_t+d_t} \ln(z + wd_t\alpha_f^+)(x_t+d_t-z)^{\lambda-1}\mathrm{d}z\Bigg]
\end{aligned} \tag{123}
$$

(C) para $\frac{1}{3} < \theta < \frac{1}{2}$**:** $\dfrac{1-\theta}{2} \leq r < \theta$ **ou** $\dfrac{\theta}{2} \leq r < \dfrac{1}{2}$ **para** $\theta > \frac{1}{2}$

$$
\begin{aligned}
E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \quad = \quad & \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{-x_t} \ln(-z + wd_t\alpha_f^+)(2d_t + 2z)^{\lambda-1}\mathrm{d}z\right. \\
& + \int_{-x_t}^{x_t-\theta d_t} \ln(-z + wd_t\alpha_s^+)(2d_t + 2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t-\theta d_t}^{0} \ln(-z + wd_t\alpha_s^-)(2d_t + 2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{0}^{x_t} \ln(z + wd_t\alpha_s^-)(2d_t - 2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t}^{-x_t+d_t} \ln(z + wd_t\alpha_f^-)(2d_t - 2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{-x_t+d_t}^{x_t+\theta d_t} \ln(z + wd_t\alpha_f^-)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\
& \left. + \int_{x_t+\theta d_t}^{x_t+d_t} \ln(z + wd_t\alpha_f^+)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z\right] \quad (124)
\end{aligned}
$$

(D) para $\theta < 1/2$**:** $\max\left\{\dfrac{1-\theta}{2}, \theta\right\} \leq r < \dfrac{1}{2}$

$$
\begin{aligned}
E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \quad = \quad & \frac{\lambda}{(2d_t)^\lambda}\left[\int_{x_t-d_t}^{-x_t} \ln(-z + wd_t\alpha_f^+)(2d_t + 2z)^{\lambda-1}\mathrm{d}z\right. \\
& + \int_{-x_t}^{0} \ln(-z + wd_t\alpha_s^+)(2d_t + 2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{0}^{x_t-\theta d_t} \ln(z + wd_t\alpha_s^+)(2d_t - 2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t-\theta d_t}^{x_t} \ln(z + wd_t\alpha_s^-)(2d_t - 2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t}^{-x_t+d_t} \ln(z + wd_t\alpha_f^-)(2d_t - 2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{-x_t+d_t}^{x_t+\theta d_t} \ln(z + wd_t\alpha_f^-)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\
& \left. + \int_{x_t+\theta d_t}^{x_t+d_t} \ln(z + wd_t\alpha_f^+)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z\right] \quad (125)
\end{aligned}
$$

(D) para $\theta > 1/2$: $\dfrac{1}{2} \leq r < \theta$

$$\begin{aligned} E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \;=\; & \frac{\lambda}{(2d_t)^\lambda} \left[\int_{x_t-d_t}^{x_t-\theta d_t} \ln(-z + wd_t\alpha_s^+)(2d_t + 2z)^{\lambda-1}\mathrm{d}z \right. \\ & + \int_{x_t-\theta d_t}^{0} \ln(-z + wd_t\alpha_s^-)(2d_t + 2z)^{\lambda-1}\mathrm{d}z \\ & + \int_{0}^{-x_t+d_t} \ln(z + wd_t\alpha_s^-)(2d_t - 2z)^{\lambda-1}\mathrm{d}z \\ & + \int_{-x_t+d_t}^{x_t} \ln(z + wd_t\alpha_s^-)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\ & + \int_{x_t}^{x_t+\theta d_t} \ln(z + wd_t\alpha_f^-)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\ & \left. + \int_{x_t+\theta d_t}^{x_t+d_t} \ln(z + wd_t\alpha_f^+)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \right] \qquad (126) \end{aligned}$$

(E) $\max\left\{\dfrac{1}{2}, \theta\right\} \leq r < \dfrac{1+\theta}{2}$

$$\begin{aligned} E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \;=\; & \frac{\lambda}{(2d_t)^\lambda} \left[\int_{x_t-d_t}^{0} \ln(-z + wd_t\alpha_s^+)(2d_t + 2z)^{\lambda-1}\mathrm{d}z \right. \\ & + \int_{0}^{x_t-\theta d_t} \ln(z + wd_t\alpha_s^+)(2d_t - 2z)^{\lambda-1}\mathrm{d}z \\ & + \int_{x_t-\theta d_t}^{-x_t+d_t} \ln(z + wd_t\alpha_s^-)(2d_t - 2z)^{\lambda-1}\mathrm{d}z \\ & + \int_{-x_t+d_t}^{x_t} \ln(z + wd_t\alpha_s^-)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\ & + \int_{x_t}^{x_t+\theta d_t} \ln(z + wd_t\alpha_f^-)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\ & \left. + \int_{x_t+\theta d_t}^{x_t+d_t} \ln(z + wd_t\alpha_f^+)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \right] \qquad (127) \end{aligned}$$

(F) $\dfrac{1+\theta}{2} \leq r < 1$

$$\begin{aligned}
E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \;=\; & \frac{\lambda}{(2d_t)^\lambda}\Bigg[\int_{x_t-d_t}^{0} \ln(-z + wd_t\alpha_s^+)(2d_t + 2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{0}^{-x_t+d_t} \ln(z + wd_t\alpha_s^+)(2d_t - 2z)^{\lambda-1}\mathrm{d}z \\
& + \int_{-x_t+d_t}^{x_t-\theta d_t} \ln(z + wd_t\alpha_s^+)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t-\theta d_t}^{x_t} \ln(z + wd_t\alpha_s^-)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t}^{x_t+\theta d_t} \ln(z + wd_t\alpha_f^-)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t+\theta d_t}^{x_t+d_t} \ln(z + wd_t\alpha_f^+)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z\Bigg] \qquad (128)
\end{aligned}$$

(G) $r \geq 1$

$$\begin{aligned}
E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})] \;=\; & \frac{\lambda}{(2d_t)^\lambda}\Bigg[\int_{x_t-d_t}^{x_t-\theta d_t} \ln(-z + wd_t\alpha_s^+)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t-\theta d_t}^{x_t} \ln(z + wd_t\alpha_s^-)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t}^{x_t+\theta d_t} \ln(z + wd_t\alpha_f^-)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z \\
& + \int_{x_t+\theta d_t}^{x_t+d_t} \ln(z + wd_t\alpha_f^+)(x_t + d_t - z)^{\lambda-1}\mathrm{d}z\Bigg] \qquad (129)
\end{aligned}$$

As expressões de $E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})]$ foram obtidas com o GPL Maxima. As integrais derivadas do cálculo de $E^{A_t}[\ln(|x_{t+1}| + wd_{t+1})]$ só podem ser calculadas se atribuirmos valores a λ.

Para cada valor de λ fixado, podemos descrever expressões analíticas para $\Phi(r)$, no caso da adaptação combinada, a partir das Equações (71) à (108) e (120) à (129). Apresentamos as expressões de $\Phi(r)$ para o caso mais simples $(1+1)$-EE com **Regra 1** de adaptação do tamanho do passo para os seguintes intervalos:

(A) $0 \leq r < \frac{1}{2}$

(B) $\frac{1}{2} \leq r < 1$

(C) $r \geq 1$

$$
\begin{aligned}
\Phi_A(r) &= (\alpha_s w + r)\ln(\alpha_s w + r) + (1-r)\ln(\alpha_f w + r) - \ln(w+r) \\
&- \alpha_s w[\ln(w) + \ln(\alpha_s)] + k[(r-1)\ln(\alpha_f) - r\ln(\alpha_s)] - r \\
\Phi_B(r) &= \frac{(\alpha_s w + r)\ln(\alpha_s w + r) + (\alpha_s w - r + 1)\ln(\alpha_s w - r + 1) + \ln(\alpha_f w + r)}{2} \\
&+ \frac{-2\ln(w+r) - 2\alpha_s w[\ln(w) + \ln(\alpha_s)] - k[\ln(\alpha_s) + \ln(\alpha_f)] - 1}{2} \\
\Phi_C(r) &= \frac{(\alpha_s w + r)\ln(\alpha_s w + r) + (-\alpha_s w - r + 1)\ln(\alpha_s w + r - 1) + \ln(\alpha_f w + r)}{2} \\
&+ \frac{-2\ln(w+r) - k[\ln(\alpha_s) + \ln(\alpha_f)] - 1}{2}
\end{aligned}
$$

Notamos que $\Phi(r)$ não pode ser representada em cada intervalo como uma soma de logaritmos de funções racionais. Consequentemente não é possível descrever as derivadas de $\Phi(r)$ em cada um dos intervalos como uma função racional na variável r e, portanto, não podemos empregar a metodologia adotada anteriormente para verificar a existência de parâmetros α_s, α_f (**Regra 1**), $\alpha_j, j = 0, \ldots, \lambda$ (**Regra 2**), α^+, α^- (**Regra 3**), $\alpha_s^+, \alpha_s^-, \alpha_f^-, \alpha_f^+$ (Combinação), w, k e a tais que $\Phi(r) = E^{A_t}(V_{t+1}) - V_t \leq -a$. Segundo tal abordagem, é suficiente mostrar a validade desta última desigualdade para os extremos de cada intervalo em que Φ é definida e em quaisquer pontos críticos dentro desses intervalos. Entretanto, para mostrar que $\Phi(r) \leq -a$, realizamos, primeiramente uma partição no domínio de Φ em intervalos crescentes, suficientemente pequenos. Através de uma amostragem tão fina quanto necessario, o valor máximo de Φ é calculado numericamente utilizando o método de otimização local da seção áurea.

Assim o seguinte problema de otimização não-linear restrito pode ser formulado:

$$
((\alpha_s^+)^*, (\alpha_s^-)^*, (\alpha_f^-)^*, (\alpha_f^+)^*, \theta^*, w^*, k^*, a^*) = \arg\max_{\alpha_s^+, \alpha_s^-, \alpha_f^-, \alpha_f^+, \theta, w, k, a} \quad a
$$

$$
\text{sujeito a:} \begin{cases}
(\alpha_s^+, \alpha_s^-, \alpha_f^-, \alpha_f^+) > 0 \\
0 < \theta < 1 \\
w > 0 \\
0 \leq k < 1 \\
a > 0 \\
\Phi(r) + a \leq 0, \qquad r = 0,\ \frac{1}{2},\ \frac{\theta}{2},\ \theta,\ \frac{1-\theta}{2},\ \frac{1+\theta}{2},\ 1,\ +\infty
\end{cases} \tag{130}
$$

Resolvendo numericamente o Problema 130, encontramos os valores ótimos alcançados pelas estratégias $(1,2)$-EE e $(1+1)$-EE para a **Regra 1**, **Regra 3** e para o controle combinado. Esses valores são apresentados nas Tabelas 13 e 14.

Tabela 13 – Valores dos parâmetros ótimos para $(1{,}2)$-EE (sem usar Jensen).

Parâmetros	**Regra 1**	**Regra 3**	Controle Combinado
α_s^+	1.2011	1.4098	2.2695
α_s^-	1.2011	0.65032	0.96029
α_f^-	0.65815	0.65032	0.52130
α_f^+	0.65815	1.4098	1.0119
θ	——-	0.39576	0.66002
w	2.1244	3.9933	1.3144
k	0.37754	0.38748	0.35774
a	0.012405	0.011419	0.032617

Tabela 14 – Valores dos parâmetros ótimos para $(1+1)$-EE (sem usar Jensen).

Parâmetros	**Regra 1**	**Regra 3**	Controle Combinado
α_s^+	1.9178	11.092	11.096
α_s^-	1.9178	0.73370	0.74885
α_f^-	0.73921	0.73370	0.73087
α_f^+	0.73921	0.73370	0.73087
θ	——	0.51057	0.51195
w	0.15534	0.11116	0.11142
k	0.26526	0.27562	0.27448
a	0.046286	0.097794	0.097802

Pelos dados apresentados nessas tabelas, constatamos um aumento significativo dos limites teóricos da taxa de convergência para a seleção $(1{,}2)$-EE. Para a **Regra 1** o aumento foi de mais 47%. Para a combinação o aumento foi de mais 100%. Para **Regra 3**, o aumento foi mais expressivo (maior que 185%).

A análise da Figura 25 sugere que os algoritmos $(1{,}2)$-EE, principalmente com a **Regra 3**, apresentam melhores limites para as taxas de convergência, quando estes são obtidos sem a desigualdade de Jensem. Neste caso, observamos o gráfico de $\Phi/2$ completamente abaixo do gráfico de $\Psi/2$ para $(1{,}2)$-EE.

Para a seleção $(1+1)$-EE, o aumento não foi tão surpreendente. Para a **Regra 1**, pouco mais de 7%, para a **Regra 3** mais de 36% e para a combinação mais de 31%.

Na Figura 25 exibimos os gráficos de $\Psi(r)/2$ e $\Phi(r)/2$ para $(1{,}2)$-EE e $\Psi(r)$ e $\Phi(r)$ para $(1+1)$-EE.

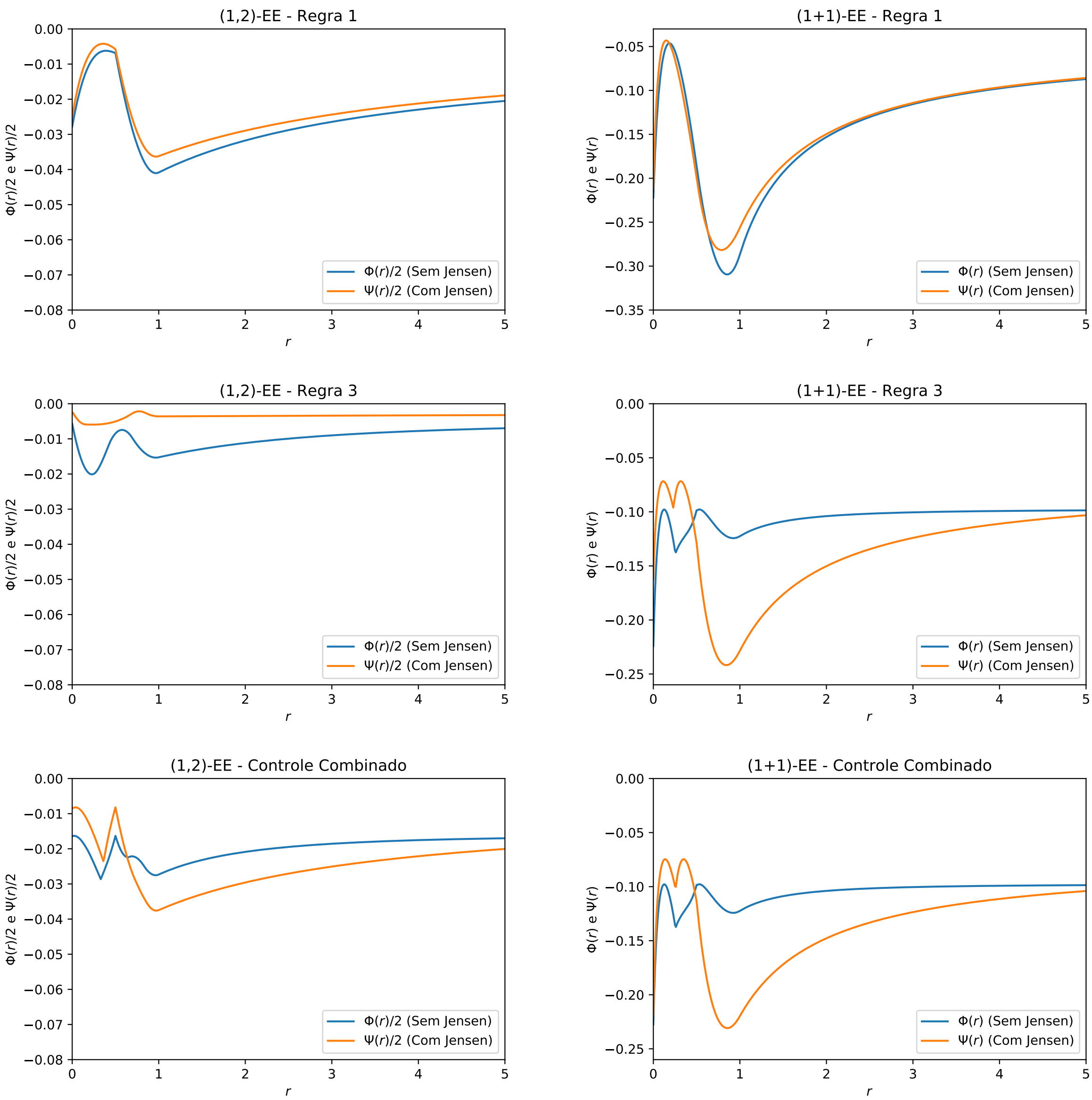

Figura 25 – Gráficos de $\Psi_2(r)/2$, $\Phi_2(r)/2$, $\Psi(r)$ e $\Phi(r)$.

5.2.1 Resultados Experimentais e Discussões

Podemos executar todas as EEs formuladas neste trabalho com os valores ótimos dos parâmetros obtidos sem o auxílio da desigualdade de Jensen na estimativa do decrescimento esperado do supermartingal associado V_t. Nesta subseção realizamos a comparação dos resultados de tais experimentos com os apresentados anteriormente. Iniciamos contrastando o comportamento da estratégia $(1{,}2)$-EE para cada uma das regras estudadas.

Analisando a Figura 26 observamos que, quando os algoritmos são iniciados no ótimo, o máximo das distribuições de $|x_t|$, d_t e e^{V_t} em $(1{,}2)$-EE com a **Regra 1** está muito mais próximo ao limite teórico, principalmente no início dos experimentos, em comparação com o máximo das mesmas distribuições com os parâmetros obtidos com o uso da desigualdade de Jensen para a mesma adaptação. Nos experimentos nos quais os algoritmos foram iniciados com tamanho de passo inicial muito pequeno, as diferenças não são expressivas (figura encontrada no Apêndice F.4).

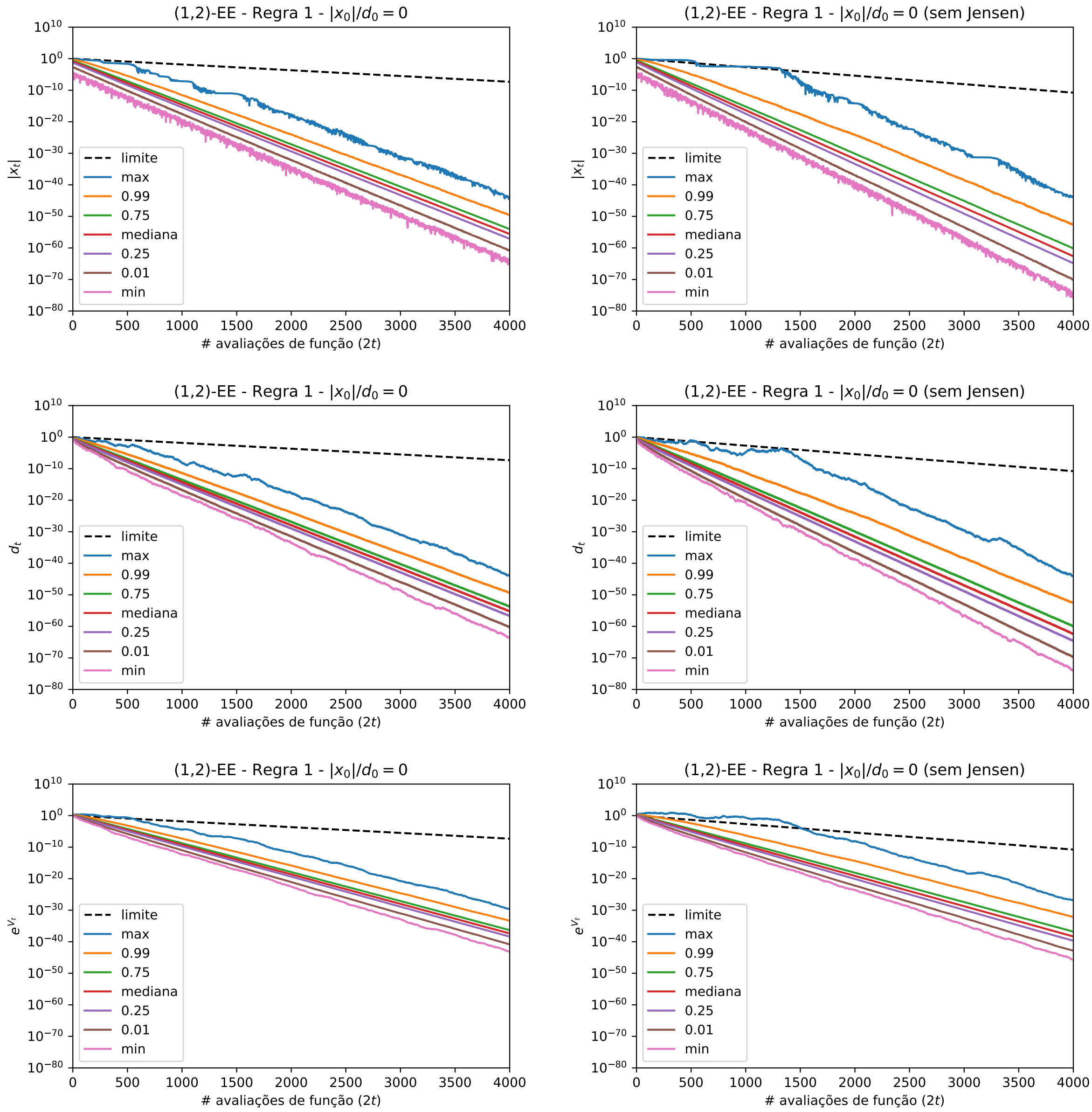

Figura 26 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1{,}2)$-EE com **Regra 1** (com e sem Jensen), iniciadas no ótimo ($|x_0|/d_0 = 0$).

Para a **Regra 3**, a Figura 27 evidencia a convergência mais rápida do algoritmo $(1{,}2)$-EE quando a desigualdade de Jensen não é aplicada para estimar um limite superior para a taxa de convergência. Observamos que não só o maximo está mais próximo ao limite teórico, o 99º percentil tembém está mais próximo ao limite teórico no início dos experimentos.

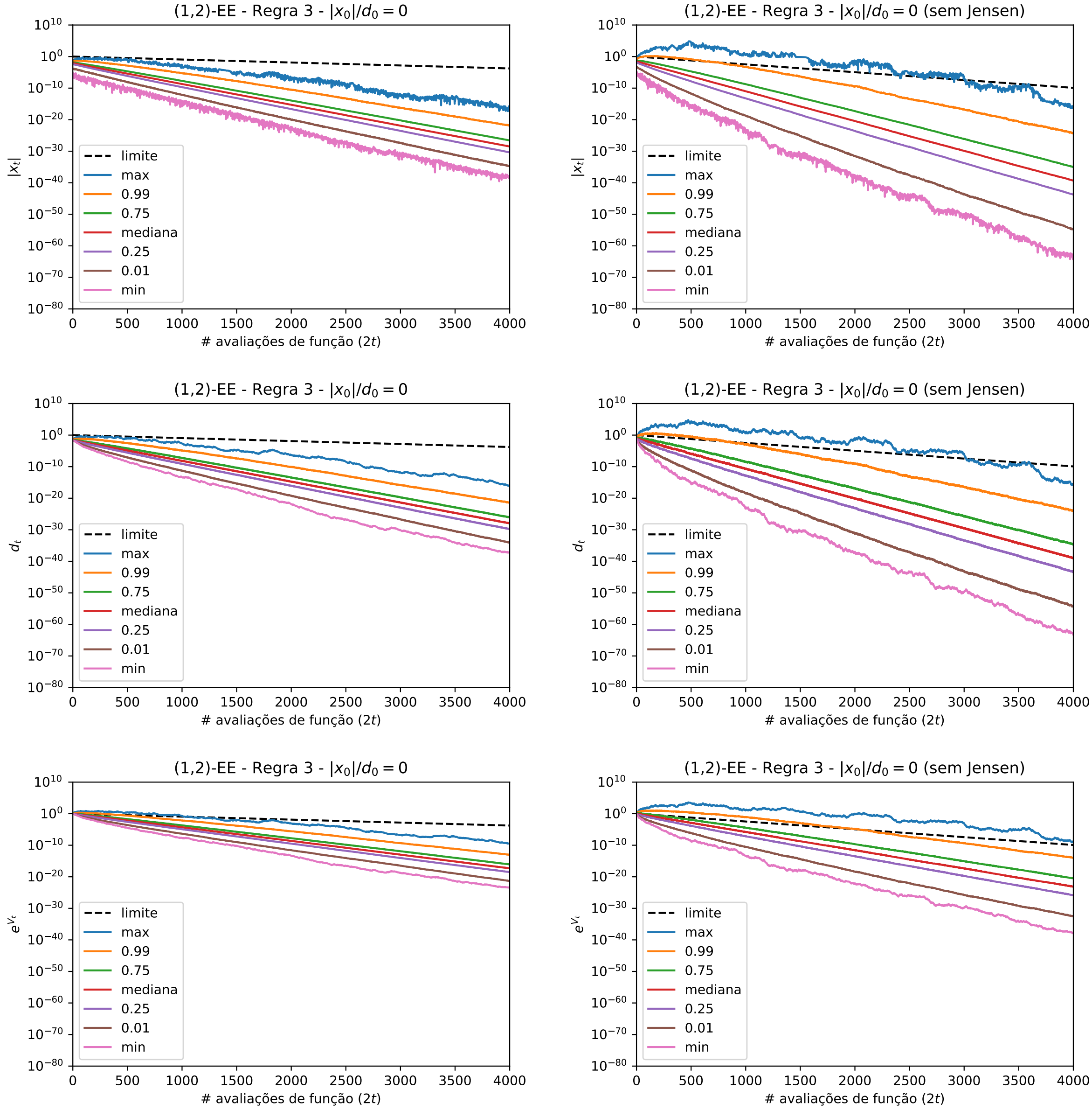

Figura 27 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1{,}2)$-EE com **Regra 3** (com e sem Jensen), iniciadas no ótimo ($|x_0|/d_0 = 0$).

O limite teórico da taxa de convergência do algoritmo $(1{,}2)$-EE com controle combinado, obtido com o uso da desigualdade de Jensen é superior, quando comparado com o limite da mesma EE, alcançado sem o auxílio da desigualdade de Jensen (ver última linha e última coluna das Tabelas 13 e 14). A Figura 29 apresenta uma comparação para os experimentos iniciados longe do ótimo e nos permite validar a observação feita anteriormente em relação aos limites teóricos. Entretanto, os experimentos iniciados no ótimo apresentam um cenário diferente conforme nos indica a Figura 28. Nesses experimentos, notamos a convergência ligeiramente mais rápida dos algoritmos $(1{,}2)$-EE com controle combinado com o uso da desigualdade de Jensen. O mais importante a ser observado é o fato de que, neste último caso, todos os quantis empíricos estão muito mais próximos ao limite teórico.

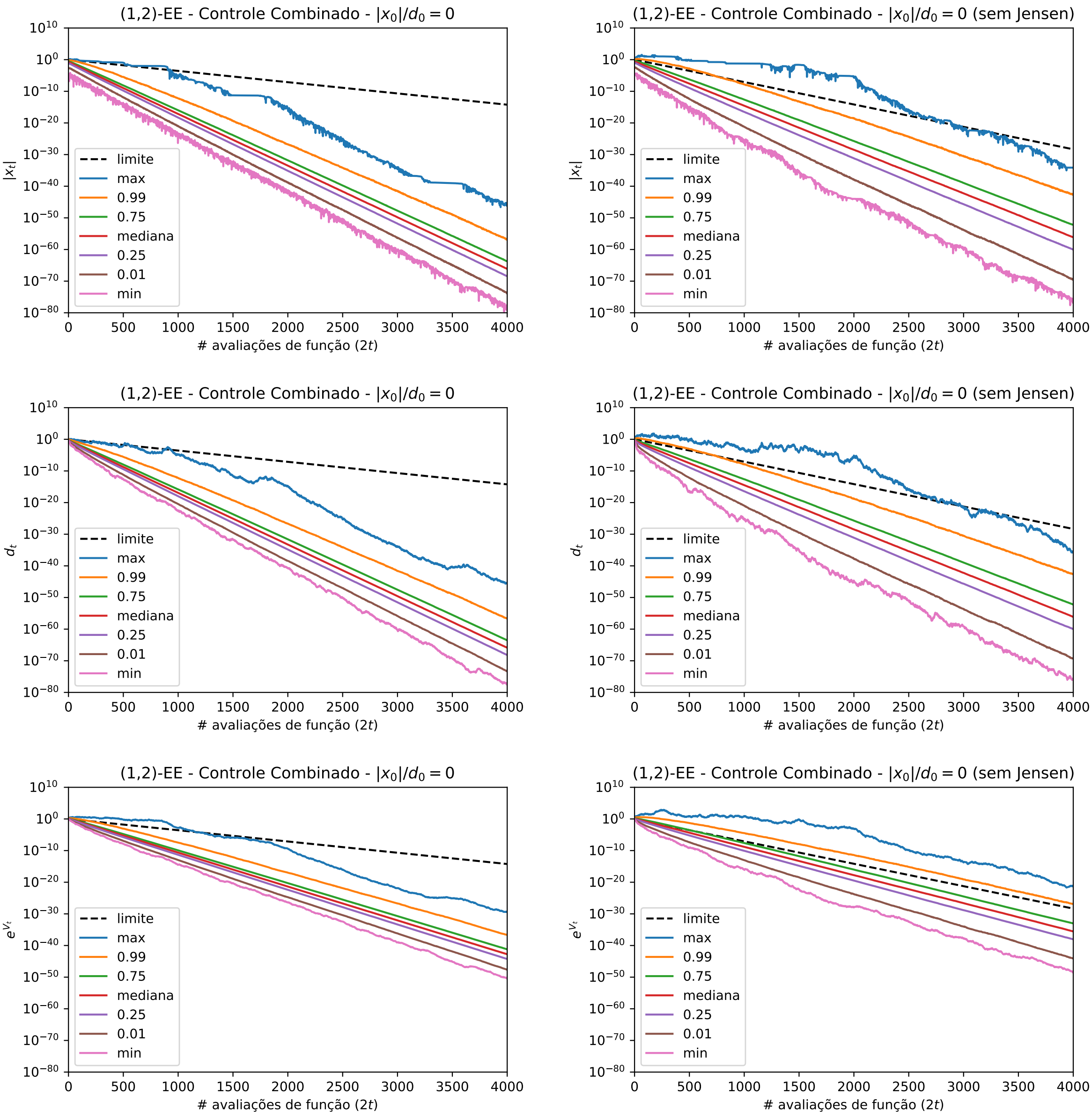

Figura 28 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1{,}2)$-EE com controle combinado (com e sem Jensen), iniciadas no ótimo ($|x_0|/d_0 = 0$).

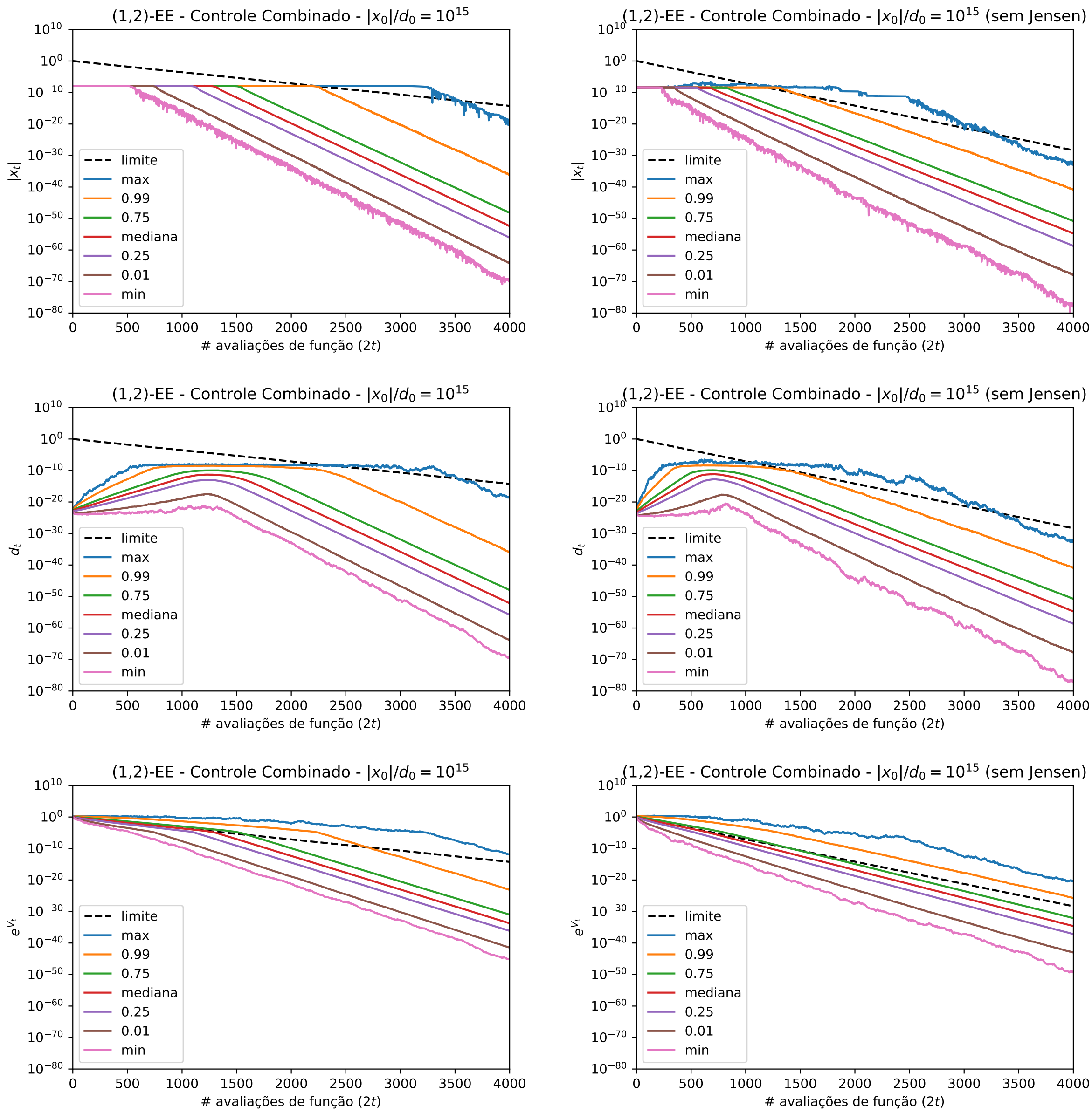

Figura 29 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1,2)$-EE com controle combinado (com e sem Jensen), iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

Os experimentos para $(1+1)$-EE com **Regra 1**, **Regra 3** e controle combinado, com os parâmetros obtidos sem o uso da desigualdade de Jensen, apresentam diferenças bem sutis em relação aos experimentos da mesma regra de adaptação com os parâmetros obtidos com a desigualdade de Jensen (ver Figuras 30, 31 e 32). Para a **Regra 3** e o controle combinado, todos os quantis estão mais próximos ao limite teórico. Os mesmos comportamentos podem ser observados nos experimentos em que os algoritmos foram iniciados longe do ótimo.

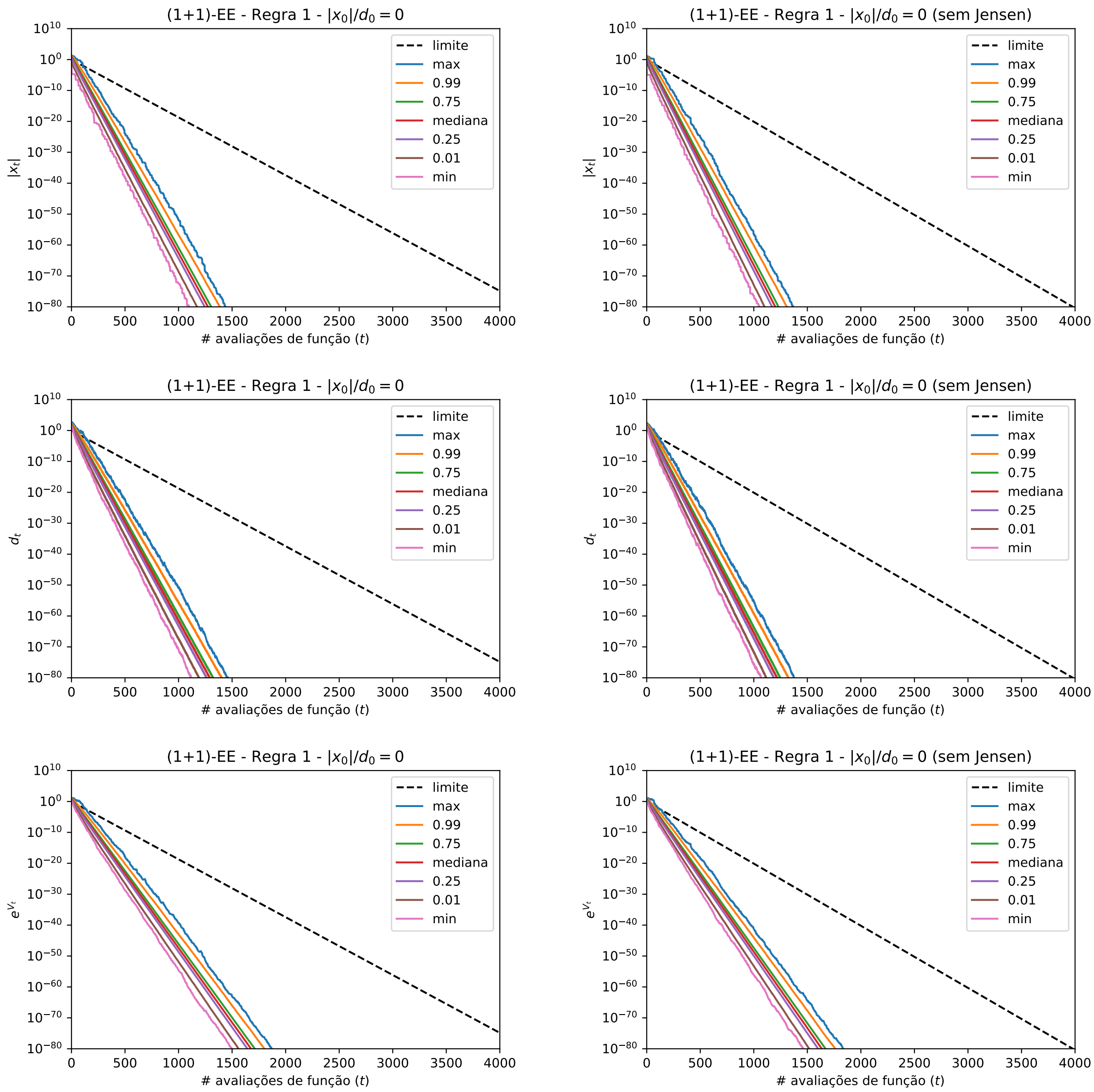

Figura 30 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1+1)$-EE com **Regra 1** (com e sem Jensen), iniciadas no ótimo ($|x_0|/d_0 = 0$).

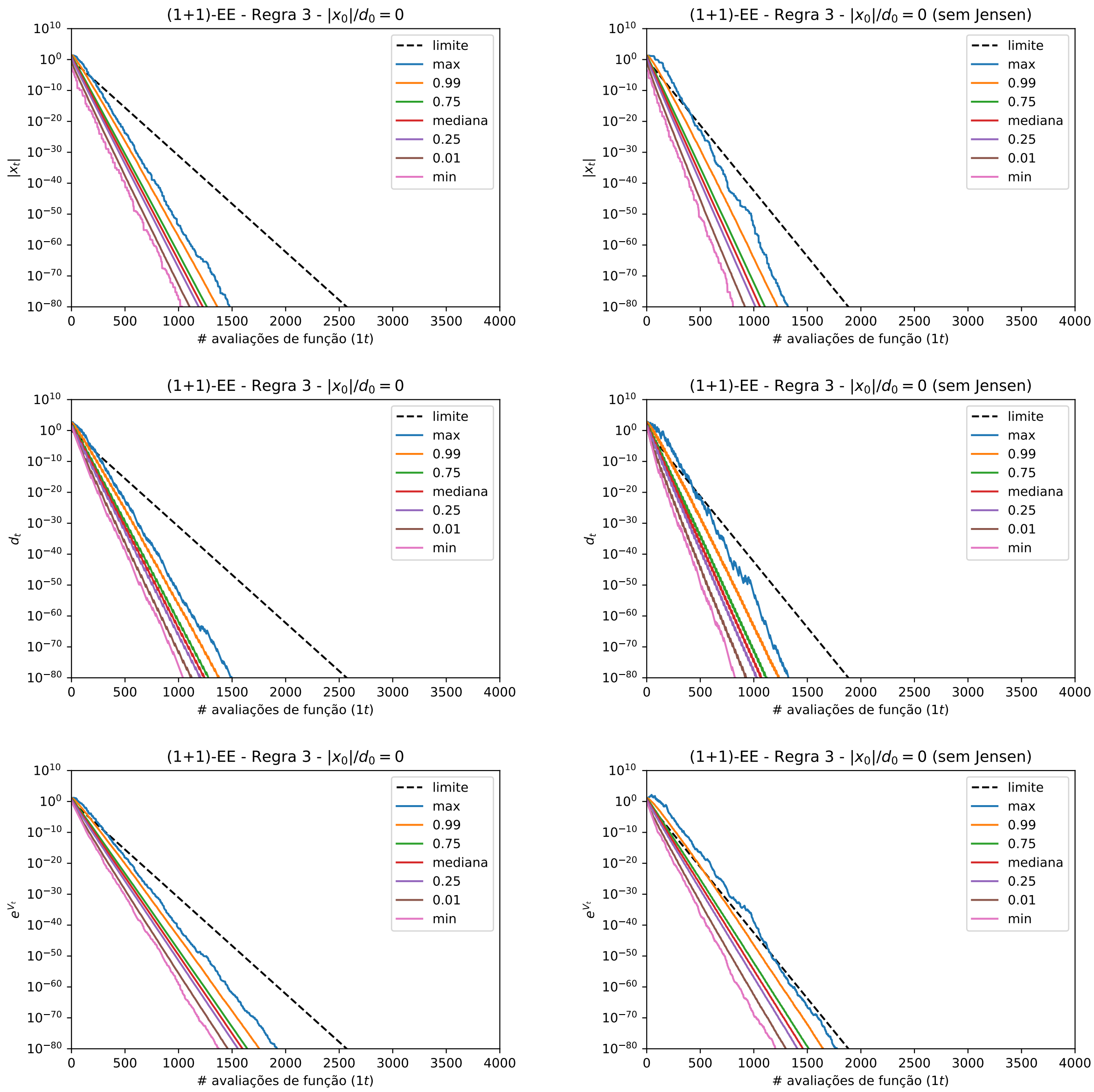

Figura 31 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1+1)$-EE com **Regra 3** (com e sem Jensen), iniciadas no ótimo ($|x_0|/d_0 = 0$).

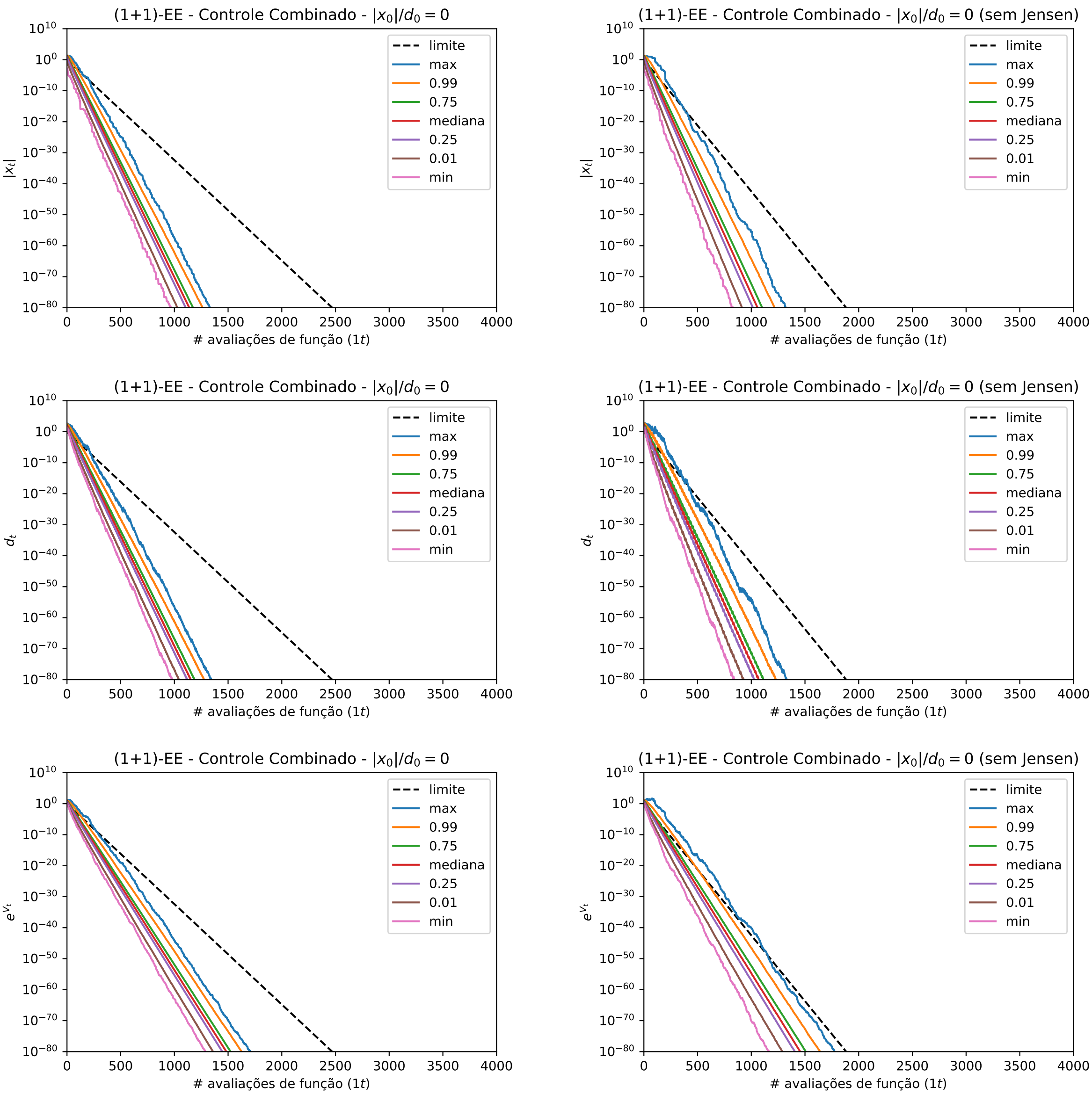

Figura 32 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1+1)$-EE com controle combinado (com e sem Jensen), iniciadas no ótimo ($|x_0|/d_0 = 0$).

6 Conclusão

Os Algoritmos Evolutivos (AEs) são ferramentas úteis para se encontrar extremos locais e globais e parâmetros ótimos em modelos de simulação computacional. Diversos modelos de AEs têm sido desenvolvidos e aplicados e propostas híbridas têm sido criadas com o objetivo de potencializar estes algoritmos. Os experimentos com simulações têm fundamentado as principais pesquisas que tratam da análise de AEs com foco nas suas aplicações. A análise rigorosa de AEs, por sua vez, não é amplamente vista e as investigações teóricas do comportamento dessas técnicas de otimização em espaços de busca contínuos vêm sendo realizadas, principalmente na exploração das estratégias evolutivas. Demonstrações analíticas de convergência de AEs não são o foco da maioria das pesquisas dessa área. Contudo, a análise e síntese das Estratégias Evolutivas (EEs) descritas neste trabalho permitiu o alcance de provas formais da convergência destes algoritmos.

Seguindo o paradigma no qual o comportamento de uma EE é tratado como um sistema dinâmico, a metologia utilizada neste trabalho realizou uma analogia entre a análise de convergência de processos estocásticos, que descrevem o comportamento de algoritmos evolutivos, e o método de função de Lyapunov estocástica para análise de estabilidade de processos estocásticos. A partir dessa abordagem foi possível demonstrar a convergência de estratégias $(1 \overset{+}{,} \lambda)$-EE que otimizam uma classe de funções unimodais de uma variável que são simétricas em torno do eixo vertical que contém o ótimo. Na primeira EE formulada, o tamanho do passo do algoritmo foi modificado, de acordo com uma regra de adaptação simples baseada em sucesso, denominada **Regra 1**. Em seguida, uma extensão desta regra, denominada **Regra 2**, foi proposta. Nesta estratégia, o tamanho do passo foi adaptado de acordo com o número de descendentes bem sucedidos. Também foi desenvolvida uma EE com controle de mutação baseado no tamanho do passo, intitulada **Regra 3**. Esta regra é uma generalização a qualquer número de descendentes e qualquer limiar do algoritmo $(1{,}2)$-EE proposto em (WANNER; FONSECA; CARDOSO; TAKAHASHI, 2016). Além disso, a **Regra 1** e a **Regra 3** foram combinadas em um único algoritmo.

Como o objetivo de estimar o decrescimento esperado do supermartingal associado à função de Lyapunov definida, expressões gerais para as esperanças condicionais dos próximos valores do tamanho do passo e para a distância ao ótimo foram analiticamente derivadas para todas as Estratégias. Isto permitiu a formulação de problemas de otimização numérica, através dos quais foram obtidos limites superiores para a taxa de convergência das EEs, bem como valores dos parâmetros ótimos que derivam estes limites.

Tendo em vista os valores ótimos dos limites teóricos, normalizados por λ, para as taxas de convergência, obtidos para as regras de adaptação do tamanho do passo baseadas em sucesso, a **Regra 2** supera a **Regra 1**. Essa diferença não é tão expressiva quando examinamos a seleção elitista.

Como esperado, dentre todas as regras de adaptação formuladas, a combinação da **Regra 1** com a **Regra 3** apresentou melhores limites teóricos superiores para a taxa de convergência dos algoritmos $(1 \overset{+}{,} \lambda)$-EE. Apesar disso, os resultados experimentais mostraram uma convergência mais acelerada da **Regra 3** para $(1,\lambda)$, com $\lambda = 4, 5, 6$. Em relação à seleção elitista, os experimentos apresentaram diferença ínfima entre a regra combinada e a **Regra 3**. De modo geral os resultados experimentais nos permitiram verificar que todos os limites teóricos obtidos podem ser considerados conservadores. Também confirmamos que o uso da desigualdade de Jensen favoreceu essa característica. De fato, observamos que os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} estão mais próximos aos limites teóricos, quando as EEs são executadas com os parâmetros obtidos sem o auxílio da desigualdade de Jensen.

Os resultados apresentados nesta tese revelam uma importante contribuição para o campo teórico das Estratégias Evolutivas, principalmente por apresentar uma metodologia para provar a convergência destes algoritmos, que pode ser estendida considerando outras funções de uma variável e até funções de múltiplas variáveis, desde que as esperanças condicionais necessárias possam ser computadas. Ainda assim este trabalho pode ser consideravelmente ampliado, gerando maior impacto na utilização prática das Estratégias Evolutivas.

Destacamos alguns avanços que podem ser derivados desta tese:

- Ainda considerando a mesma classe de funções objetivo e utilizando a mesma metodologia, esperamos obter melhores taxas de convergência adotando outra estrutura de função de Lyapunov $V_t = V(x_t, d_t)$. Para tanto é fundamental que a estrutura de função de Lyapunov estocástica adotada atenda às mesmas propriedades que caracterizam a classe da função objetivo considerada e às condições necessárias impostas pela Proposição 4. Lembramos que V_t representa a energia de um sistema dinâmico. Neste trabalho definimos $V_t = \ln(|x_t| + wd_t) - k\ln(d_t)$ de forma que seu primeiro termo garante que a função tende a $-\infty$ somente quanto $|x_t|$ e d_t tendem a zero. Seu segundo termo garante que quando x_t está muito longe do ótimo, valores elevados de d_t não permitem um aumento desta energia. Quando x_t está perto do ótimo, é possível controlar o comportamento de V_t, mantendo suas características fundamentais, acrescentanto mais um termo à sua expressão.
- Uma questão fundamental que se apresenta nesta pesquisa é se os mesmos procedimentos para estimar limites para taxa de convergência podem ser aplicados na análise de EEs que otimizam outras classes de funções objetivo. Considerando funções de uma variável, unimodais e **assimétricas** em torno do ótimo e a mesma função de Lyapunov, alcançamos resultados preliminares para a estratégia não elitista $(1,2)$-EE com

a **Regra 3**. Para a mesma estratégia de seleção com a **Regra 1**, a metodologia não se mostrou eficaz. Acreditamos que ao modificar a estrutura de função de Lyapunov, seja possivel obter limites para a taxa de convergência de $(1 \overset{+}{,} \lambda)$-EE com a **Regra 1**.

- Se a abordagem da análise teórica de EE's executada nesta tese também for exequível na análise de EEs que otimizam funções assimétricas, poderemos encontrar parâmetros ótimos que maximizem simultaneamente os limites das taxas de convergência para ambas as classes de funções (simétrica e assimétrica).
- Este trabalho apresentou formas distintas de atualizar o tamanho do passo em EEs, considerando o mesmo modo de selecionar as gerações futuras. Uma maneira de modificar a seleção do novo pai em $(1 \overset{+}{,} \lambda)$-EE pode ser executada escolhendo, por exemplo x_{t+1} como uma combinação linear dos melhores filhos.
- Resultados mais realistas, porém mais difíceis de serem alcançados, podem ser investigados ao considerar funções objetivo de várias variáveis de decisão. Com este objetivo, talvez seja necessário utilizar outra distribuição para a mutação (distribuição normal).

6.1 Contribuições da Tese

Esta tese originou os seguintes trabalhos:

- CORRÊA, C. R.; WANNER, E. F.; FONSECA, C. M. Lyapunov design of a simple step-size adaptation strategy based on success. In: **Proceedings of the 14th International Conference on Parallel Problem Solving from Nature**. Edinburgh - Scotland: Springer, 2016. (Lecture Notes in Computer Science, v. 9921), p. 101–110.
  Neste trabalho formulamos uma regra de adaptação de tamanho de passo simples baseada em sucesso para Estratégias Evolutivas $(1 \overset{+}{,} \lambda)$-EE (**Regra 1**) e consideramos a configuração dos parâmetros correspondentes. A convergência teórica sobre a classe de funções estritamente unimodais de uma variável que é simétrica em torno do ótimo foi investigada usando um método de função estocástica de Lyapunov desenvolvido por Semenov e Terkel (SEMENOV; TERKEL, 2003) no contexto da teoria do martingais.
- CORRÊA, C. R.; WANNER, E. F.; FONSECA, C. M. Lyapunov design of simple and extended success-based step-size adaptation rules. **Evolutionary Computation - Submitted.**, 2018.
  Este trabalho é uma extensão do anterior, no qual a **Regra 2**, baseada no número de descendentes bem sucedidos, foi formulada e investigada para as mesmas EEs aplicadas à mesma classe de funções objetivo. A configuração dos parâmetros correspondentes também foi obtida através de procedimento idêntico.

Referências

AGAPIE, A. Theoretical analysis of mutation-adaptive evolutionary algorithms. **Evolutionary Computation**, MITP, v. 9, n. 2, p. 127–146, 2001.

ARNOLD, D. V.; HANSEN, N. Active covariance matrix adaptation for the (1+1)-CMA-ES. In: ACM. **Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation**. [S.l.], 2010. p. 385–392.

AUGER, A. Convergence results for the $(1, \lambda)$-SA-ES using the theory of ϕ-irreducible Markov chains. **Theoretical Computer Science**, Elsevier, v. 334, n. 1-3, p. 35–69, 2005.

AUGER, A.; SCHOENAUER, M.; VANHAECKE, N. LS-CMA-ES: A second-order algorithm for covariance matrix adaptation. In: SPRINGER. **International Conference on Parallel Problem Solving from Nature**. [S.l.], 2004. v. 8, n. 4, p. 182–191.

BÄCK, T. An overview of parameter control methods by self-adaptation in evolutionary algorithms. **Fundamenta Informaticae**, IOS Press, v. 35, n. 1-4, p. 51–66, 1998.

BACK, T.; HOFFMEISTER, F.; SCHWEFEL, H.-P. A Survey of Evolution Strategies. In: MORGAN KAUFMANN PUBLISHERS SAN MATEO, CA. **Proceedings of the Fourth International Conference on Genetic Algorithms**. [S.l.], 1991. v. 2, n. 9.

BEYER, H.-G. Toward a theory of evolution strategies: Self-adaptation. **Evolutionary Computation**, MIT Press, v. 3, n. 3, p. 311–347, 1995.

BEYER, H.-G.; SCHWEFEL, H.-P. Evolution strategies–a comprehensive introduction. **Natural Computing**, Springer, v. 1, n. 1, p. 3–52, 2002.

BEYER, H.-G.; SENDHOFF, B. Covariance matrix adaptation revisited–the CMSA- Evolution Strategy. In: SPRINGER. **International Conference on Parallel Problem Solving from Nature**. [S.l.], 2008. p. 123–132.

BIENVENÜE, A.; FRANÇOIS, O. Global convergence for evolution strategies in spherical problems: some simple proofs and difficulties. **Theoretical Computer Science**, Elsevier, v. 306, n. 1-3, p. 269–289, 2003.

CORRÊA, C. R.; WANNER, E. F.; FONSECA, C. M. Lyapunov design of a simple step-size adaptation strategy based on success. In: **Proceedings of the 14th International Conference on Parallel Problem Solving from Nature**. Edinburgh - Scotland: Springer, 2016. (Lecture Notes in Computer Science, v. 9921), p. 101–110.

CORRÊA, C. R.; WANNER, E. F.; FONSECA, C. M. Lyapunov design of simple and extended success-based step-size adaptation rules. **Evolutionary Computation - Submitted.**, 2018.

DOERR, B.; DOERR, C. Optimal parameter choices through self-adjustment: Applying the $1/5$-th rule in discrete settings. In: ACM. **Proceedings of the 24th Genetic and Evolutionary Computation Conference**. [S.l.], 2015. p. 1335–1342.

DOERR, B.; DOERR, C. A tight runtime analysis of the (1+(λ, λ)) genetic algorithm on OneMax. In: ACM. **Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation**. [S.l.], 2015. p. 1423–1430.

DOERR, B.; FOUZ, M.; WITT, C. Sharp bounds by probability-generating functions and variable drift. In: ACM. **Proceedings of the 13th annual conference on Genetic and evolutionary computation**. [S.l.], 2011. p. 2083–2090.

DOERR, B.; GOLDBERG, L. A. Adaptive drift analysis. **Algorithmica**, Springer, v. 65, n. 1, p. 224–250, 2013.

DOERR, B.; JOHANNSEN, D.; WINZEN, C. Multiplicative drift analysis. **Algorithmica**, v. 64, n. 4, p. 673–697, 2012.

EIBEN, A. E.; HINTERDING, R.; MICHALEWICZ, Z. Parameter control in evolutionary algorithms. **IEEE Transactions on Evolutionary Computation**, IEEE, v. 3, n. 2, p. 124–141, 1999.

EIBEN, A. E.; SMITH, J. E. **Introduction to Evolutionary Computing**. [S.l.]: Springer, 2003. v. 53.

FOGEL, L. J.; OWENS, A. J.; WALSH, M. J. **Artificial intelligence through simulated evolution**. [S.l.]: John Wiley, 1966.

GE, S. Lyapunov design. **CONTROL SYSTEMS, ROBOTICS AND AUTOMATION–Volume XIII: Nonlinear, Distributed, and Time Delay Systems-II**, EOLSS Publications, p. 90, 2009.

GOLDBERG, D. **Genetic Algorithms in Search, Optimization and Machine Learning, 1989, 1st**. [S.l.]: Addison-Wesley Longman Publishing Co., Inc, 1989.

HAHN, W. **Stability of Motion**. [S.l.]: Springer, 1967.

HANSEN, N.; OSTERMEIER, A. Completely derandomized self-adaptation in evolution strategies. **Evolutionary Computation**, MIT Press, Cambridge, MA, USA, v. 9, n. 2, p. 159 –195, 2001.

HANSEN, N.; OSTERMEIER, A.; GAWELCZYK, A. On the adaptation of arbitrary normal mutation distributions in evolution strategies: The generating set adaptation. In: MORGAN KAUFMANN. **Proceedings of the 6th International Conference on Genetic Algorithms**. San Francisco - California, 1995. p. 57–64.

HART, W. E. Rethinking the design of real-coded evolutionary algorithms: Making discrete choices in continuous search domains. **Soft Computing Journal**, Springer, v. 9, n. 4, p. 225–235, 2005.

HE, J.; HE, F.; YAO, X. A unified markov chain approach to analysing randomised search heuristics. **arXiv preprint arXiv:1312.2368**, 2013. Disponível em: <https://arxiv.org/abs/1312.2368#>.

HE, J.; YAO, X. Drift analysis and average time complexity of evolutionary algorithms. **Artificial Intelligence**, Elsevier, v. 127, n. 1, p. 57–85, 2001.

HE, J.; YAO, X. Erratum to: Drift analysis and average time complexity of evolutionary algorithms:[artificial intelligence 127 (2001) 57–85]. **Artificial Intelligence**, Elsevier, v. 140, n. 1-2, p. 245–248, 2002.

HE, J.; YAO, X. A study of drift analysis for estimating computational time of evolutionary algorithms. **Natural Computing**, Springer, v. 3, p. 21–35, 2004.

HOLLAND, J. H. **Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence**. [S.l.]: MIT press, 1992.

JÄGERSKÜPPER, J. Analysis of a simple evolutionary algorithm for minimization in euclidean spaces. **Proceedings of the 30th International Colloquium on Automata, Languages and Programming - Lecture Notes in Computer Science**, Springer Verlag, v. 2719, p. 1068–1079, 2003.

JÄGERSKÜPPER, J. Rigorous runtime analysis of the $(1+1)$ ES: $1/5$-rule and ellipsoidal fitness landscapes. In: SPRINGER. **Proceedings of International Workshop on Foundations of Genetic Algorithms**. [S.l.], 2005. p. 260–281.

JÄGERSKÜPPER, J. How the $(1+1)$-ES using isotropic mutations minimizes positive definite quadratic forms. **Theoretical Computer Science.**, v. 361, n. 1, p. 38–56, 2006.

JÄGERSKÜPPER, J. Algorithmic analysis of a basic evolutionary algorithm for continuous optimization. **Theoretical Computer Science**, Elsevier, v. 379, n. 3, p. 329–347, 2007.

JÄGERSKÜPPER, J. A blend of Markov-chain and drift analysis. In: SPRINGER. **Proceedings of 10th International Conference on Parallel Problem Solving from Nature**. [S.l.], 2008. p. 41–51.

JAMES, B. **Probabilidade: um curso em nıvel intermediário. ed**. Rio de Janeiro: IMPA, 2004.

JANSON, S.; LUCZAK, T.; RUCINSKI, A. **Random Graphs**. [S.l.]: Wiley, 2000.

KERN, S.; Müller, S. D.; HANSEN, N.; BÜCHE, D.; OCENASEK, J.; KOUMOUTSAKOS, P. Learning probability distributions in continuous evolutionary algorithms – a comparative review. **Natural Computing**, Kluwer Academic Publishers, v. 3, p. 77–112, 2004.

KOZA, J. R. **Genetic Programming: On the Programming of Computers by Means of Natural Selection**. [S.l.]: MIT press, 1992. v. 1.

KRAMER, O. Evolutionary self-adaptation: a survey of operators and strategy parameters. **Evolutionary Intelligence**, Springer, v. 3, n. 2, p. 51–65, 2010.

KUSHNER, H. **Stochastic Stability and Control**. [S.l.]: Academic Press, 1967.

LEHRE, P. K.; WITT, C. General drift analysis with tail bounds. **arXiv:1307.2559v1**, 2013. Disponível em: <https://arxiv.org/abs/1307.2559>.

LYAPUNOV, A. M. The General Problem of Stability of Motion (reprint of the original paper of 1892). **International Journal of Control**, v. 55, n. 3, p. 531–773, 1992.

MEYER-NIEBERG, S.; BEYER, H.-G. Self-adaptation in evolutionary algorithms. In: **Parameter setting in evolutionary algorithms**. [S.l.]: Springer, 2007. p. 47–75.

NEVEU, J. **Discrete-Parameter Martingales**. [S.l.]: Elsevier, 1975. v. 10.

OLLIVIER, Y.; ARNOLD, L.; AUGER, A.; HANSEN, N. Information-geometric optimization algorithms: A unifying picture via invariance principles. **Journal of Machine Learning Research**, v. 18, n. 18, p. 1–65, 2017.

OSTERMEIER, A.; GAWELCZYK, A.; HANSEN, N. A derandomized approach to self-adaptation of evolution strategies. **Evolutionary Computation**, MIT Press, v. 2, n. 4, p. 369–380, 1994.

PARKS, P. C. Am Lyapunov's Stability Theory—100 years on. **IMA Journal of Mathematical Control and Information**, Oxford University Press, v. 9, n. 4, p. 275–303, 1992.

RECHENBERG, I. Cybernetic solution path of an experimental problem. **Royal Aircraft Establishment Library Translation 1122**, 1965.

RECHENBERG, I. **Evolutionsstrategie: Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution. Dr.-Ing**. Tese (Doutorado) — Thesis, Technical University of Berlin, Department of Process Engineering, 1971.

RECHENBERG, I. Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. **Stuttgart-Bad Cannstatt: Friedrich Frommann Verlag**, 1973.

RUDOLPH, G. **Convergence Properties of Evolutionary Algorithms**. Hamburg: Kovac, 1997.

RUDOLPH, G. Convergence rates of evolutionary algorithms for a class of convex objective functions. **Control and Cybernetics**, POLISH ACADEMY OF SCIENCES, v. 26, p. 375–390, 1997.

RUDOLPH, G. Self-adaptive mutations may lead to premature convergence. **IEEE Transactions on Evolutionary Computation**, v. 5, n. 4, p. 410–414, 2001.

SCHWEFEL, H.-P. **Kybernetische Evolution als Strategie der exprimentellen Forschung in der Strömungstechnik**. Tese (Doutorado) — Thesis, Technical University of Berlin, Hermann Föttinger-Institute for Hydrodynamics, 01 1965.

SCHWEFEL, H.-P. Evolutionsstrategie und numerische optimierung. dr.-ing. **Diss., Technical University of Berlin**, 1975.

SCHWEFEL, H.-P. **Numerische Optimierung Von Computer-Modellen Mittels der Evolutionsstrategie**. [S.l.]: Birkhäuser, Basel Switzerland, 1977. v. 1.

SCHWEFEL, H.-P. **Numerical Optimization of Computer Models**. Chichester: Wiley, 1981.

SEMENOV, M. A. Convergence velocity of an evolutionary algorithm with self-adaptation. In: MORGAN KAUFMANN PUBLISHERS INC. **Proceedings of the 4th Annual Conference on Genetic and Evolutionary Computation**. [S.l.], 2002. p. 210–213.

SEMENOV, M. A.; TERKEL, D. A. Analysis of convergence of an evolutionary algorithm with self-adaptation using a stochastic Lyapunov function. **Evolutionary Computation**, v. 11, n. 4, p. 363–379, 2003.

SHIRYAYEV, A. N. **Probability**. [S.l.]: Springer, 1984.

WANNER, E. F.; FONSECA, C. M.; CARDOSO, R. T. N.; TAKAHASHI, R. H. C. Lyapunov stability analysis and adaptation law synthesis of a derandomized self-adaptive $(1,2)$-ES. **Evolutionary Computation - ACCEPTED.**, 2016.

WITT, C. Tight bounds on the optimization time of a randomized search heuristic on linear function. **Combinatorics Probability and Computing**, Cambridge University Press, Cambridge, MA, USA, v. 22, n. 02, p. 294–318, 2013.

Apêndices

APÊNDICE A – Breve Teoria de Martingais

Neste apêndice descrevemos definições e resultados relevantes sobre Probabilidade e Teoria de Martingais, baseados em (JAMES, 2004; SHIRYAYEV, 1984), para o desenvolvimento da metodologia utilizada na análise de convergência das EEs trabalhadas.

A.1 Primeiras Definições

Seja $(\Omega, \mathit{A}, \mathit{P})$ um espaço de probabilidade, em que Ω é o espaço amostral, *A* é uma σ-álgebra de conjuntos mensuráveis, $(\mathit{A}_t, t \in \mathbb{N})$ é uma família de sub-σ-álgebras $\mathit{A}_t \subset A$ e *P* é uma medida de probabilidade.

Definição 3 (Desigualdade de Jensen) *Sejam $X : \Omega \to \mathbb{R}_+$ uma variável aleatória e φ uma função convexa definida na reta. Se a variável aleatória X é integrável, então*

$$E[\varphi(X)] \geq \varphi(E(X))$$

Definição 4 *Seja G $\subset$ A uma sub-σ-álgebra e $X : \Omega \to \mathbb{R}_+$ uma variável aleatória positiva integrável. Qualquer variável aleatória $Y : \Omega \to \mathbb{R}$, G-mensurável que satisfaz*

$$\nu(B) = \int_B Y \, d\mathsf{P} = \int_B X \, d\mathsf{P}, \forall B \in \mathit{G} \tag{131}$$

*é chamada uma versão da **esperança condicional**[1] de X com respeito a G.*

ν é uma medida definida em (Ω, A) absolutamente contínua ($\nu << \mathit{P}$).

Notação: $Y = E[X|\mathit{G}]$ ou $Y = E^{\mathit{G}}[X]$.

No caso geral em que $X : \Omega \to \mathbb{R}$, $X = X^+ - X^-$, tem-se $E[X|\mathit{G}] = E[X^+|\mathit{G}] - E[X^-|\mathit{G}]$.

Definição 5 *Seja $(\Omega, \mathsf{A}, \mathsf{P})$ um espaço de probabilidade:*

a) Uma filtração é uma sequência crescente $(\mathsf{A}_t)_{t \geq 1}$ de sub-σ-álgebras $\mathsf{A}_1 \subset \mathsf{A}_2 \subset \cdots \subset$ A.

[1] O Teorema de Radon-Nikodym afirma que existe uma única função Y quase certamente *G*-mensurável que satisfaz (131), chamada derivada de Radon-Nikodym: $Y = \frac{d\nu}{dP}$.

b) Uma sequência de variáveis aleatórias $(X_t)_{t\geq 1}$ em (Ω, A) é adaptada à filtração $(\mathsf{A}_t)_{t\geq 1}$ se X_t é A_t-mensurável para todo $t \geq 1$.

c) Uma sequência de variáveis aleatórias $(X_t)_{t\geq 1}$ em (Ω, A) é previsível com respeito à filtração $(\mathsf{A}_t)_{t\geq 1}$ se X_t é A_{t-1}-mensurável para todo $t \geq 1$.

d) Uma sequência dupla $(X_t, \mathsf{A}_t)_{t\geq 1}$, em que $(\mathsf{A}_t)_{t\geq 1}$ é uma filtração e $(X_t)_{t\geq 1}$ é adaptada a $(\mathsf{A}_t)_{t\geq 1}$, é chamada uma ***sequência estocástica***[2].

Definição 6 *Uma sequência estocástica $(X_t, \mathsf{A}_t)_{t\geq 1}$, tal que X_t é integrável ($E(|X_t|) < +\infty, \forall t \geq 1$) é chamada:*

a) Um ***Martingal****, se para todo $t \geq 1$: $E[X_{t+1}|\mathsf{A}_t] = X_t$*

b) Um ***Submartingal****, se para todo $t \geq 1$: $E[X_{t+1}|\mathsf{A}_t] \geq X_t$*

c) Um ***Supermartingal****, se para todo $t \geq 1$: $E[X_{t+1}|\mathsf{A}_t] \leq X_t$*

A.2 Teoremas de Convergência de Submartingais e Martingais

Nesta subseção são apresentados alguns teoremas que tratam da convergência de Martingais e Submartingais.

Teorema 11 *(Doob)*

Seja (X_t, A_t) um submartingal com $\sup_t E(|X_t|) < +\infty$. Então existe quase certamente $\lim_{t\to\infty} X_t = X_\infty$. Além disso $E(|X_\infty|) < +\infty$.

Corolário 12 (Doob) *Se (X_t, A_t) é um submartingal não positivo, então existe quase certamente $\lim_{t\to\infty} X_t = X_\infty < +\infty$.*

Corolário 13 (Doob) *Se (X_t, A_t) é um martingal não negativo, então existe quase certamente $\lim_{t\to\infty} X_t = X_\infty < +\infty$.*

A.3 Martingais Quadrado Integráveis e Tempo de Parada

Teorema 14 (Teorema da decomposição de Doob) *Seja (X_t, A_t) um submartingal. Então existem quase certamente, um martingal (M_t, A_t) e uma sequência previsível crescente (H_t, A_{t-1}) tais que, para todo $t \geq 0$,*

$$X_t = M_t + H_t. \tag{132}$$

Além disso, tal decomposição é única.

[2] Terminologia encontrada em (SHIRYAYEV, 1984).

Definição 7 *Uma sequência previsível crescente* $(H_t, \mathcal{A}_{t-1})$ *definida como na decomposição descrita pela Equação (132) é nomeada um **compensador** (do submartingal* X_t*).*

A decomposição de Doob desempenha um papel fundamental no estudo de martingais quadrados integráveis[3] $(X_t, \mathcal{A}_t)$. Como $(X_t^2, \mathcal{A}_t)$ é um submartingal, então existem quase certamente, um martingal $(M_t, \mathcal{A}_t)$ e uma sequência previsível crescente $(H_t, \mathcal{A}_{t-1})$ tais que, $X_t^2 = M_t + H_t$ para todo $t \geq 0$.

$H_t = \langle X_t \rangle$ é chamado **variação quadrática** de X_t.

Definição 8 *Uma variável aleatória* $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ *é um **tempo de parada** com respeito à filtração* $(\mathcal{A}_t)$ *se,* $P(\tau < \infty) = 1$ *e para cada* $t \geq 0$*,*

$$\{\tau = t\} \in \mathcal{A}_t \text{ ou } \{\tau \leq t\} \in \mathcal{A}_t$$

Proposição 15 *Se* T *é um tempo de parada e se* $k \geq 1$ *é um número inteiro, então* $T \wedge k$[4] *é um tempo de parada. Além disso, se* T_1 *e* T_2 *são tempos de parada, então* $T_1 + T_2, T_1 \wedge T_2, T_1 \vee T_2$ *também são tempos de parada*

Sejam $(X_t, \mathcal{A}_t)$ uma sequência estocástica e τ um tempo de parada com respeito à $(\mathcal{A}_t)$. Definindo $X_\tau : \Omega \to \mathbb{R}$ tal que $X_\tau = \begin{cases} X_t(\omega), \text{ se } T(\omega) = t \\ 0, \text{ se } T(\omega) = \infty \end{cases}$, podemos escrever $X_\tau = \sum_{t \geq 0} X_t I_{\{\tau = t\}}$. Observamos que X_τ é é uma variável aleatória.

[3]Martingais $(X_t, \mathcal{A}_t)$ para os quais $E[X_t^2] < \infty, n \geq 0$.
[4]$T \wedge k = \min\{T, k\}$

APÊNDICE B – Teoria de Estabilidade de Lyapunov

Neste apêndice são apresentadas definições e resultados referentes à Teoria de Estabilidade de Lyapunov.

De modo geral, a estabilidade de um sistema dinâmico significa insensibilidade do estado do sistema para pequenas mudanças no estado inicial ou nos parâmetros do sistema. Ela pode ser descrita em termos do conceito de energia: diz-se que um sistema é assintoticamente estável se sua energia tende a zero à medida que o tempo tende ao infinito. Os métodos de Lyapunov realizam a verificação de estabilidade de um sistema dinâmico via uma função escalar que procura descrever a energia do sistema.

Em (SEMENOV; TERKEL, 2003) um método de funções de Lyapunov estocásticas da teoria de estabilidade de processos estocásticos é utilizado como uma abordagem para investigar a convergência de processos estocásticos que descrevem o comportamento de algoritmos evolutivos.

Como um sistema dinâmico discreto determinístico é um sistema para o qual se deseja saber seu estado no decorrer de uma sequência discreta de tempos $\{t_1,t_2,t_3,\ldots\}$, sua representação (em dimensão n) pode ser dada na forma:

$$x_{t+1} = F(t,x_t) \tag{133}$$

sendo $x_t \in \mathbb{R}^n$ o vetor estado do sistema, $t \in \mathbb{N}$ denota o tempo e $F : \mathbb{N} \times \mathbb{R}^n \to \mathbb{R}^n$ uma função C^p com $p \geq 1$.

Definição 9 *Seja $x^* \in \mathbb{R}^n$ um vetor estado do sistema dinâmico (133).*

a) x^* é um ponto de equilíbrio do sistema dinâmico (133), se $x_{t_0} = x^* \Rightarrow x_t = x^*, \forall t \geq t_0$.

b) Um ponto de equilíbrio x^* do sistema (133) é estável, segundo Lyapunov, se dado $\epsilon > 0$ existe $\delta(\epsilon, t_0)$ tal que $\|x_{t_0} - x^*\| < \delta \Rightarrow \|x_t - x\| < \epsilon, \forall t \geq t_0$.

c) Um ponto de equilíbrio x^* do sistema (133) é dito assintoticamente estável, se é estável e se $x^* = \lim_{t\to\infty} x_t$.

Na definição anterior, quando δ não depende de t_0, o ponto de equilíbrio é chamado *uniformemente assintoticamente estável*. Esse conceito de estabilidade uniforme é importante na análise de estabilidade de sistemas cujo campo vetorial depende explicitamente do tempo.

Não existe um método geral para construir ou encontrar uma função candidata de Lyapunov que demonstre a estabilidade de um ponto de equilíbrio dado. Porém a incapacidade de encontrar uma função de Lyapunov não implica na instabilidade do ponto de equilíbrio automaticamente.

O método direto de Lyapunov (segundo método de Lyapunov - versão discreta do método de Lyapunov (LYAPUNOV, 1992): reimpressão do artigo original de 1892) apresenta condições sob as quais x^* é um ponto estável. Tais condições tratam da existência de uma função (contínua no segundo argumento) de Lyapunov candidata $V(t, x_t) : \mathbb{N} \times \mathbb{R}^N \to \mathbb{R}$ que seja *definida positiva*, *radialmente ilimitada* e que satisfaça: $\frac{dV}{dt} < 0$.

O método direto de Lyapunov é uma interpretação matemática da propriedade física de que, se a energia total de um sistema estiver se dissipando, os estados do sistema chegarão finalmente a um ponto de equilíbrio.

Em forma de teorema (versão do Teorema de Estabilidade de Lyapunov encontrada em (HAHN, 1967)), tem-se o seguinte:

Teorema 16 *Seja x^* um ponto de equilíbrio do sistema definido pela Equação (133). Se existe uma função de Lyapunov $V(t, x_t) : \mathbb{N} \times \mathbb{R}^N \to \mathbb{R}$ contínua na segunda coordenada que satisfaz:*

a) $V(t, x_t) = 0 \Leftrightarrow x_t = x^*$ *e* $V(t, x_t) > 0, \forall x_t \neq x^*$ *(definida positiva);*

b) $\lim_{\|x_t\| \to \infty} V(t, x_t) = \infty$ *(radialmente ilimitada);*

c) $V(t+1, x_{t+1}) < V(t, x_t), \forall\, V(t, x_t) > 0$ *(equivalente a $\frac{dV}{dt} < 0$).*

então x^* é uniformemente assintoticamente estável.

OBS: Uma função $V(t, x_t)$ que seja continuamente diferenciável e definida positiva é chamada de Função de Lyapunov Candidata. Segundo o método Direto de Lyapunov, se a função de Lyapunov candidata não for uma função de Lyapunov, ou seja, se falhar a condição (c) do Teorema 16, então nada se pode concluir sobre a estabilidade do ponto de equilíbrio.

De forma geral, segundo Lyapunov, um sistema dinâmico discreto iniciado perto suficientemente de um ponto de equilíbrio estável, mantém os estados do sistema próximos a ele indefinidamente.

APÊNDICE C – Demonstrações dos Resultados de Semenov e Terkel

Neste apêndice realizamos uma releitura das demonstraçoes dos resultados utilizados nesta tese (Proposição 1, Teorema 2 e Proposição 4 da Subseção 2.3.1 do Capítulo 2), obtidos em (SEMENOV; TERKEL, 2003). As demonstrações seguem exatamente a mesma ideia dos autores de (SEMENOV; TERKEL, 2003), porém, utilizamos uma notação matemática mais compreensível, explicitando pontos não exibidos no trabalho original.

C.1 Demonstração da Proposição 1

Se o par (V_t, V_t^*) de processos estocásticos satisfaz a A- condição em $G \subset U$, então

$$\lim_{t\to\infty} V_t^* = -\infty$$

quase certamente em $\{\omega \in \Omega : \; \exists \lim_{t\to\infty} V_t \in G\}$.

A-condição: O par de processos estocásticos (V_t, V_t^*) satisfaz a **A**- condição em $G \subset U$ se $\forall \; V \in G, V^* \in \mathbb{R}, \exists \; \delta > 0, \epsilon > 0$ tais que

$$P[\cup_{k=1}^{\infty}\{|V_{t+k} - V_t| > \delta\}|A_t] > \epsilon$$

quase certamente em $D = \{V_t^* > V^*\} \cap \{|V_t - V| < \delta\}$.

Prova:

Seja $E = \{\omega \in \Omega : \; \exists \lim_{t\to\infty} V_t \in G\}$.

Devemos mostrar que

$$P\left[\left\{\lim_{t\to\infty} V_t \in G\right\} \cap \left\{\lim_{t\to\infty} V_t^* \neq -\infty\right\}\right] = P\left[E \cap \left\{\lim_{t\to\infty} V_t^* \neq -\infty\right\}\right] = 0 \qquad (134)$$

Se $\lim_{t\to\infty} V_t^* \neq -\infty$, então existe pelo menos um número racional L e um inteiro t_0, tais que $V_t^* \geq L \; \forall \; t > t_0$. Assim é possível escrever o evento $\{\lim_{t\to\infty} L_t \neq -\infty\}$ como uma união enumerável de eventos da seguinte forma:

$$\left\{\lim_{t\to\infty} V_t^* \neq -\infty\right\} = \bigcup_{L\in\mathbb{Q}} \bigcap_{s\in\mathbb{N}} \bigcup_{t\geq s} \{V_t^* \geq L\}.$$

Como

$$P\left[E\cap\left\{\lim_{t\to\infty}V_t^*\neq-\infty\right\}\right]=P\left[E\cap\bigcup_{L\in\mathbb{Q}}\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^*\geq L\}\right]=$$

$$=P\left[\bigcup_{L\in\mathbb{Q}}\left(E\cap\left\{\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^*\geq L\}\right\}\right)\right]\leq\sum_{L\in\mathbb{Q}}P\left[E\cap\left\{\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^*\geq L\}\right\}\right],$$

para provar (134) é suficiente mostrar que $\forall\, L\in\mathbb{Q}$, temos:

$$P\left[E\cap\left\{\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^*\geq L\}\right\}\right]=0. \tag{135}$$

Para qualquer $V\in G$ e $L\in\mathbb{Q}$ escolhendo $\delta=\delta(V,L)$ de acordo com a A-condição e um conjunto enumerável de intervalos $I_{V_i,\delta_i/2}=(V_i-\delta_i/2,V_i+\delta_i/2)$, tal que $G\subset\bigcup_{i\in\mathbb{N}}I_{V_i,\delta_i/2}$, temos

$$E\subset\bigcup_{i\in\mathbb{N}}\left\{\lim_{t\to\infty}V_t\in G\cap I_{V_i,\delta_i/2}\right\}.$$

$$\Rightarrow P\left[E\bigcap\left\{\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^*\geq L\}\right\}\right]\leq\sum_{i\in\mathbb{N}}P\left[\left\{\lim_{t\to\infty}V_t\in G\cap I_{V_i,\delta_i/2}\right\}\bigcap\left\{\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^*\geq L\}\right\}\right].$$

Consequentemente para provar (135) é suficiente provar que $\forall\, i\in\mathbb{N}$:

$$P_i\equiv P\left[\left\{\lim_{t\to\infty}V_t\in G\cap I_{V_i,\delta_i/2}\right\}\bigcap\left\{\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^*\geq L\}\right\}\right]=0. \tag{136}$$

É possível observar que se o evento $\{\lim_{t\to\infty}V_t\in G\cap I_{V_i,\delta_i/2}\}$ ocorre, então para todos os valores de t, exceto para um número finito, tem-se $V_t\in I_{V_i,\delta_i/2}$, ou seja:

$$\{\lim_{t\to\infty}V_t\in G\cap I_{V_i,\delta_i/2}\}\subset\bigcup_{k\in\mathbb{N}}\bigcap_{t>k}\{V_t\in I_{V_i,\delta_i/2}\}.$$

Como a união dos intervalos acima é crescente, é possível estimar P_i como:

$$P_i\leq\lim_{k\to\infty}P\left[\bigcap_{t>k}\{V_t\in I_{V_i,\delta_i/2}\}\bigcap\left\{\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^*\geq L\}\right\}\right].$$

O objetivo então é mostrar que, sempre que a A-condição for satisfeita em G, tem-se para qualquer $k\in\mathbb{N}$:

$$P\left[\bigcap_{t>k}\{V_t\in I_{V_i,\delta_i/2}\}\bigcap\left\{\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^*\geq L\}\right\}\right]=0. \tag{137}$$

Consideremos $\delta = \delta(V, L)$ e $\epsilon = \epsilon(V, L)$ dados na A-condição, tais que:

$$P\left[\bigcup_{k\geq 0}\{|V_{t+k} - V_t| > \delta\}\middle| A_t\right] \geq \epsilon I_{D_{t,V,L,\delta}},$$

sendo I a função indicadora.

Para qualquer $t \in \mathbb{N}$ e $\lambda > 0$ definamos $K(t,\lambda)$ da seguinte forma:

$$K(t,\lambda) = \min_{k:k>t}\left\{\exists\, \Omega_t,\ \text{com } P(\Omega_t) \leq \lambda/2^t \text{ tal que } P\left(\bigcup_{t\leq j\leq k}\{|V_j - V_t| > \delta\}\middle| A_t\right) \geq \frac{\epsilon}{2} I_{D/\Omega_t}\right\}.$$

Pela A-condição, $K(t,\lambda) < \infty$ quase certamente.

Dado T um tempo de parada no evento $\{T < \infty\}$, definamos o tempo de parada$T^{+,\lambda} = K(T,\lambda)$

Pela definição acima, temos

$$P\left[\bigcup_{j=T}^{T^{+,\lambda}}\{|V_j - V_T| > \delta\}\middle| A_T\right] \geq \frac{\epsilon}{2} I_{\{D/\Omega_T\}} I_{\{T<\infty\}}.$$

Declarando $T_0^{+,\lambda} \equiv 0$, para qualquer sequência de λ_l dada, definamos indutivamente para $l \geq 1$:

$$T_l = \min\{t > T_{l-1}^{+,\lambda_l} :\ V_t^* > L\}.$$

Consequentemente $T_l < T_l^{+,\lambda_l} < T_{l+1}$.

Notamos que $\cap_{l\in\mathbb{N}}\{T_l < \infty\} =\subset \left\{\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^* \geq L\}\right\}$ e portanto

$P\left[\bigcap_{t>k}\{V_t \in I_{V_i,\delta_i/2}\}\bigcap\left\{\bigcap_{s\in\mathbb{N}}\bigcup_{t\geq s}\{V_t^* \geq L\}\right\}\right] \leq$

$\leq \lim_l P\left[\left\{\bigcap_{T\leq t\leq T_l}\{V_t \in I_{V_i,\delta_i/2}\}\right\}\bigcap\left\{\bigcap_{l\in\mathbb{N}}\{T_l < \infty\}\right\}\right].$

Seja $C_l = \left\{\bigcap_{k<t\leq T_l}\{|V_t - V_i| \leq \delta_i/2\}\right\}\bigcap\left\{\bigcap_{j=1}^{l}\{T_j < \infty\}\right\}.$

Usando as definições dos tempos de parada T_l e T_l^{+,λ_l} é possível inferir que:

$$P\left[\bigcap_{k<t\leq T_l^{+,\lambda_l}}\{V_t \in I_{V_i,\delta_i/2}\}\bigcap\left\{\bigcap_{j=1}^{\infty}\{T_j < \infty\}\right\}\right] \leq$$

$$P\left[\bigcap_{k<t\leq T_l^{+,\lambda_l}}\{V_t \in I_{V_i,\delta_i/2}\}\bigcap\left\{\bigcap_{j=1}^{l-1}\{T_j < \infty\}\right\}\right] =$$

$$E\left[E\left[\underbrace{I_{\bigcap_{t=k}^{T_{l-1}}\{V_t\in I_{V_i,\delta_i/2}\}}}_{\text{é } A_{T_l} \text{ mensurável}} I_{\bigcap_{t=T_l}^{T_l^{+,\lambda_l}}\{V_t\in I_{V_i,\delta_i/2}\}} \underbrace{I_{\bigcap_{j=1}^{l-1}\{T_j<\infty\}}}_{\text{é } A_{T_l} \text{ mensurável}} \middle| A_{T_l}\right]\right]=$$

$$E\left[I_{\bigcap_{t=k}^{T_{l-1}}\{V_t\in I_{V_i,\delta_i/2}\}} I_{\bigcap_{j=1}^{l-1}\{T_j<\infty\}} P\left(\bigcap_{t=T_l}^{T_l^{+,\lambda_l}}\{V_t\in I_{V_i,\delta_i/2}\}\middle| A_{T_l}\right)\right]=$$

$$\int_{C_{l-1}} P\left(\bigcap_{t=T_l}^{T_l^{+,\lambda_l}}\{V_t\in I_{V_i,\delta_i/2}\}\middle| A_{T_l}\right)dP\leq$$

$$\int_{C_{l-1}}\left[1-P\left(\bigcup_{t=T_l}^{T_l^{+,\lambda_l}}\{|V_t-V_{T_l}|>\delta\}\middle| A_{T_l}\right)\right]dP=$$

$$\int_{C_{l-1}-\Omega_{\lambda_l}}\left[1-P\left(\bigcup_{t=T_l}^{T_l^{+,\lambda_l}}\{|V_t-V_{T_l}|>\delta\}\middle| A_{T_l}\right)\right]dP+\int_{\Omega_{\lambda_l}}\left[1-P\left(\bigcup_{t=T_l}^{T_l^{+,\lambda_l}}\{|V_t-V_{T_l}|>\delta\}\middle| A_{T_l}\right)\right]dP\leq$$

$$\int_{D_{T_l}-\Omega_{\lambda_l}}\left[1-P\left(\bigcup_{t=T_l}^{T_l^{+,\lambda_l}}\{|V_t-V_{T_l}|>\delta\}\middle| A_{T_l}\right)\right]dP+\int_{\Omega_{\lambda_l}}\left[1-P\left(\bigcup_{t=T_l}^{T_l^{+,\lambda_l}}\{|V_t-V_{T_l}|>\delta\}\middle| A_{T_l}\right)\right]dP\leq$$

$$\left(1-\frac{\epsilon}{2}\right)\underbrace{P\left[\left\{\bigcap_{k<t\leq T_{l-1}}\{|V_t-V|\leq\delta/2\}\right\}\bigcap\left\{\bigcap_{j=1}^{l-1}\{T_j<\infty\}\right\}\right]}_{P(C_{l-1})}+\lambda_l=$$

$$\sum_{r=1}^{l}\left(1-\frac{\epsilon}{2}\right)^r\lambda_{l-r}\leq\sum_{r=1}^{l}\lambda_{l-r}.$$

É possível escolher $\sum_{r=1}^{l}\lambda_{l-r}$ arbitrariamente pequeno, e consequentemente (137) é verdadeira.

C.2 Demonstração do Teorema 2

Sejam $V_t : \Omega \to U \subset \mathbb{R}$ um supermartingal limitado, sendo U fechado e V_t^* um processo estocástico, adaptados a família de σ-álgebras (A_t). Se o par (V_t, V_t^*) satisfaz a **A**-condição e a **B**-condição em em um conjunto $U - \Gamma$, com $\Gamma \subset U$, então existe $\lim_{t\to\infty} V_t \in \Gamma$ quase certamente. Além disso, se a **A**-condição é satisfeita em U, então $\lim_{t\to\infty} V_t^* = -\infty$ quase certamente.

B-condição: O par de processos estocásticos (V_t, V_t^*) satisfaz a **B**- condição em $G \subset U$ se $\forall\, V \in G, \exists\, \delta > 0, V^* \in \mathbb{R}$ e $c \in \mathbb{R}$ tais que as seguintes desigualdades

a) $E^{A_t}[V_{t+1}^*] > V_t^*$

b) $E^{A_t}[V_{t+1}^* I_{\{V_{t+1}^* \geq V^*\}}] \leq c \cdot P(V_{t+1}^* > V^* | A_t)$

seguem quase certamente em $D = \{V_t^* < V^*\} \cap \{|V_t - V| < \delta\}$.

Prova:

Para simplificar a notação serão considerados $L_t = V_t^*$ e $L = V^*$.

Dados $V \in G$, selecionemos $\delta > 0$ e $L \in \mathbb{R}$ de acordo com as definições da A-condição e B-condição.

Estabelecidos V, L, δ e $n \in \mathbb{N}$, definamos o seguinte tempo de parada T:

$$T = T_{V,L,\delta,n} = \min_{i \geq n}\{i \in \mathbb{N} : \{|V_i - V| > \delta\} \cup \{L_i > L\}\}. \tag{138}$$

Se for possível mostrar que o processo $L_{t\wedge T}$ é um submartingal para $t > n$, e que $\sup_t E(L_{t\wedge T}^+) < \infty$ (isto é, $L_{t\wedge T}$ é limitado) , então pelo teorema de convergência de martingais, existe $\lim L_{t\wedge T}$ quase certamente.

Temos $E[L_{t+1\wedge T} | A_t] = E[L_{t+1\wedge T} I_{\{T>t\}} | A_t] + E[L_{t+1\wedge T} I_{\{T\leq t\}} | A_t]$.

Como a variável aleatória $L_{t+1\wedge T} I_{\{T\leq t\}} = L_{t\wedge T} I_{\{T\leq t\}}$ é A_t- mensurável, então

$$E[L_{t+1\wedge T} I_{\{T\leq t\}} | A_t] = L_{t\wedge T} I_{\{T\leq t\}}.$$

Pela desigualdade a) da B-condição é possível escrever:

$$E[L_{t+1\wedge T} I_{\{T>t\}} | A_t] = E[L_{t+1} I_{\{T>t\}} | A_t] \geq L_t I_{\{T>t\}} = L_{t\wedge T} I_{\{T>t\}}.$$

Consequentemente

$$E[L_{t+1\wedge T} | A_t] \geq L_{t\wedge T} I_{\{T\leq t\}} + L_{t\wedge T} I_{\{T>t\}} = L_{t\wedge T}$$

ou seja, $L_{t\wedge T}$ é um submartingal.

Assim

$$E[L_{t\wedge T}|A_t] \leq E\left[E[L_{t+1\wedge T}|A_t]\middle|A_t\right] = E[L_{t+1\wedge T}|A_t].$$

Seja $D_i = \{T = i\}$. Assim, podemos considerar $\mathcal{D} = \{D_1, D_2, ...\}$ uma decomposição enumerável com respeito a probabilidade P. Considerando $A_T = \sigma(\mathcal{D})$, temos $E[XI_{D_i}|A_T] = E[X|A_T]P(D_i)$ se X for uma variável aleatória para a qual sua esperança está definida.

Como $L_{t+1\wedge T} = \sum_{k\geq t+1} L_{t+1}I_{\{T=k\}} + \sum_{k\leq t} L_k I_{\{T=k\}}$ então

$$E[L_{t+1\wedge T}] = \sum_{k\geq t+1} E[L_{t+1}I_{\{T=k\}}] + \sum_{k\leq t} E[L_k I_{\{T=k\}}]$$

Para $k \geq t+1: \; E[L_{t+1}I_{\{T=k\}}] = E[E[L_{t+1}I_{\{T=k\}}|A_T]]$

$$E[L_{t+1}I_{\{T=k\}}|A_T] = E[L_{t+1}I_{\{L_{t+1}\geq L\}}I_{\{T=k\}}|A_T] + E[L_{t+1}I_{\{L_{t+1}<L\}}I_{\{T=k\}}|A_T]$$

$$= \underbrace{E[L_{t+1}I_{\{L_{t+1}\geq L\}}|A_T]}_{\leq cP(L_{t+1}\geq L|A_T)} P(T=k) + \underbrace{E[L_{t+1}I_{\{L_{t+1}\geq L\}}|A_T]}_{\leq L} P(T=k)$$

$$\leq \max\{c, L\}P(T=k)$$

Para $k \leq t < t+1: \; E[L_k I_{\{t=k\}}] = E[E[L_k I_{\{t=k\}}|A_{k-1}]]$

$$E[L_k I_{\{t=k\}}|A_{k-1}] = E[L_k I_{\{L_k\geq L\}}I_{\{T=k\}}|A_{k-1}] + E[L_k I_{\{L_k<L\}}I_{\{T=k\}}|A_{k-1}]$$

$$= \underbrace{E[L_k I_{\{L_k\geq L\}}|A_{k-1}]}_{\leq cP(L_k\geq L|A_{k-1})} P(T=k) + \underbrace{E[L_k I_{\{L_k<L\}}|A_{k-1}]}_{\leq L} P(T=k)$$

$$\leq \max\{c, L\}P(T=k)$$

$$\Rightarrow E[L_{t+1\wedge T}] = \max\{c, L\}(\sum_{k\geq t+1} P(T=k) + \sum_{k\leq t} P(T=k)) = \max\{c, L\}.$$

Consequentemente

$$E[L^+_{t\wedge T}|A_t] \leq \max\{c, L\}. \tag{139}$$

Como V_t é um supermatingal limitado, então $\exists\ \lim V_t = V_\infty$ e $P(V_\infty \in U) = 1$. Para provar a primeira afirmação do Teorema 2 , isto é, $\exists\ \lim V_t \in \Gamma$ quase certamente, é suficiente provar que $P(V_\infty \in U - \Gamma) = 0$.

A segunda afirmação do Teorema 2: $\lim L_t = -\infty$, é uma conclusão direta da Proposição 1.

Suponhamos que

$$P(V_\infty \in U - \Gamma) > 0 \tag{140}$$

por absurdo. A ideia é concluir a partir dessa desigualdade que $\sup E(L^+_{t\wedge T}) = \infty$, o que contradiz a desigualdade 139.

Como a A-condição é válida em $U - \Gamma$, se a desigualdade 140 é verdadeira, então pela Proposição 1 temos

$$P(\{V_\infty \in U - \Gamma\} \bigcap \{\lim L_t = -\infty\}) > 0. \tag{141}$$

Além disso, a B-condição é verdadeira para (V_t, L_t) em $U - \Gamma$. Consequentemente $\forall\ V \in U - \Gamma$ é possível selecionar $\delta > 0$ de acordo com as definições da A-condição e B-condição. De um conjunto de invervalos $I_{V,\delta/2}$ centrados em $V \in U - \Gamma$ de raio $\delta/2$, selecionando um subconjunto enumerável deles tal que $U - \Gamma \subset \bigcup_{i\in\mathbb{N}}\{I_{V_i,\delta_i/2}\}$, temos

$$0 < P(V_\infty \in U - \Gamma) \leq \sum_{i\in\mathbb{N}} P\left(V_\infty \in (U-\Gamma)\bigcap I_{V_i,\delta_i/2}\right).$$

Dessa forma, existe $m \in \mathbb{N}$ tal que $P\left(V_\infty \in (U-\Gamma)\bigcap I_{V_m,\delta_m/2}\right) > 0$ e por (141)

$$P\left(\{V_\infty \in (U-\Gamma)\bigcap I_{V_m,\delta_m/2}\}\bigcap\{\lim L_t = -\infty\}\right) > 0 \tag{142}$$

Se $V_\infty = \lim V_t$, então $\forall\ \delta > 0$, $\exists\, k(w) \in \mathbb{N}$ tal que $|V_t - V_\infty| < \delta\ \forall\, t > k(w)$. Em particular $|V_t - V_\infty| < \delta_m/2\ \forall\, t > k(w)$.

Se $V_\infty \in (U-\Gamma)\bigcap I_{V_m,\delta_m/2}$, então $|V_\infty - V_m| < \delta_m/2$

$\Rightarrow |V_t - V_m| = |V_t - V_\infty + V_\infty - V_m| \leq |V_t - V_\infty| + |V_\infty - V_m| < \delta_m/2 + \delta_m/2 = \delta_m$.

Assim

$$\left\{V_\infty \in (U-\Gamma)\bigcap I_{V_m,\delta_m/2}\right\} \subset \left\{\bigcap_{t>k(w)} \{V_t \in I_{V_m,\delta_m/2}\}\right\} \tag{143}$$

Usando (142) e (143) é possível selecionar $k \in \mathbb{N}$ tal que

$$P\left[\bigcap_{t>k}\{V_t \in I_{V_m,\delta_m/2}\}\bigcap\{\lim L_t = -\infty\}\right] > 0.$$

Usando a B-condição, é possível escolher L selecionando anteriormente valores de V_m and δ_m.

Como $\{\lim L_t = -\infty\} = \bigcap_{L\in\mathbb{Q}} \bigcup_n \bigcap_{t>n} \{L_t < L\}$ é possível selecionar $L_m < L$ com $L_m \in \mathbb{Q}$ e $n_m \in \mathbb{N}$ tais que

$$P\left[\bigcap_{t>k}\{V_t \in I_{V_m,\delta_m/2}\} \bigcap_{t>n_m} \{L_t < L_m\} \bigcap \{\lim L_t = -\infty\}\right] > 0 \tag{144}$$

Sejam

$$T = T_{V_m,L_m,\delta_m,n_m}$$

o tempo de parada definido no início da demonstração em (138) e $\Omega = \{T = \infty\} \bigcap \{\lim L_t = -\infty\}$.

Por (144) e pela definição de T, temos que $P(\Omega) > 0$.

Como $E[L^-_{t\wedge T}] = E[L^-_{t\wedge T};\Omega] + E[L^-_{t\wedge T};\Omega^C] \geq \int_\Omega L^-_{t\wedge T} dP$ e $L^-_{t\wedge T} = L^-_t$ em Ω e $L^-_t \to \infty$, então $E[L^-_{t\wedge T}] = \infty$.

Mas como $E[L_{t\wedge T}] = E[L^+_{t\wedge T}] - E[L^-_{t\wedge T}]$, então $E[L^+_{t\wedge T}] = E[L_{t\wedge T}] + E[L^-_{t\wedge T}]$.

Logo

$$\sup_t E[L^+_{t\wedge T}] = \infty$$

já que $\lim E[L_{t\wedge T}] \neq -\infty$.

C.3 Demonstração da Proposição 4

O supermartingal (V_t) pode ser escrito como uma soma de um martingal Y_t com um processo crescente H_t

$$V_t = Y_t - H_t \tag{145}$$

sendo

$$Y_0 = V_0,\ Y_{t+1} - Y_t = V_{t+1} - E[V_{t+1}|A_t]$$

$$H_0 = 0,\ H_{t+1} - H_t = V_t - E[V_{t+1}|A_t]$$

Por hipótese temos $E[V_{t+1}|A_t] \leq V_t - a$ e consequentemente H_t pode ser estimado por:

$$H_t = \sum_{i=1}^{t}(H_i - H_{i-1}) + H_0 \geq at \tag{146}$$

É possível mostrar que o martingal $Y_t \in L^2$.

De fato $Y_t = \sum_{i=0}^{t-1}(V_{i+1} - E[V_{i+1}|A_t]) + V_0$ e consequentemente

$$E[(Y_t)^2] = E[(\sum_{i=0}^{t-1}(V_{i+1} - E[V_{i+1}|A_t]))^2] < E[(\sum_{i=0}^{t-1}(V_{t+1} - E[V_{t+1}|A_t]))^2] =$$

$$= t^2 \sum E[(V_{t+1} - E[V_{t+1}|A_t])^2] = t^2 E\left[E[(V_{t+1} - E[V_{t+1}|A_t])^2]\right] \leq t^2 b.$$

Assim $Y_t \in L^2$, e portanto $(Y_t^2)_{t\geq 1}$ é um submartingal.

Pela decomposição de Doob é possível escrever $Y_t^2 = M_t + K_t$ em que M_t é um martingal e K_t é um processo crescente.

Então $K_{t+1} - K_t = Y_{t+1}^2 - M_{t+1} - Y_t^2 + M_t = Y_{t+1}^2 - Y_t^2 + M_t - M_{t+1}$.

$\Rightarrow E[(K_{t+1} - K_t)|A_t] = E[Y_{t+1}^2|A_t] - Y_t^2 + M_t - E[M_{t+1}|A_t] = E[Y_{t+1}^2|A_t] - Y_t^2$.

Como $E[(Y_{t+1} - Y_t)^2|A_t] = E[Y_{t+1}^2 - 2Y_{t+1}Y_t + Y_t^2|A_t] = E[Y_{t+1}^2|A_t] - Y_t^2$, então

$$E[(K_{t+1} - K_t)|A_t] = E[(Y_{t+1} - Y_t)^2|A_t] = E[(V_{t+1} - E[V_{t+1}|A_t])^2|A_t] \leq b$$

Consequentemente $E[(K_{t+1} - K_t)|A_t] = E[K_{t+1}|A_t] - k_t = K_{t+1} - k_t$.

$\Rightarrow K_{t+1} - k_t \leq b$.

É possível escrever

$$K_t = \sum_{i=1}^{t}(k_i - k_{i-1}) + k_0 \leq bt \tag{147}$$

Usando a Proposição VII (2-4) de (NEVEU, 1975) e a desigualdade (147), concluimos que

$$Y_t = o(K_t^{0.5+\epsilon}) = o(t^{0.5+\epsilon}),\ \forall \epsilon > 0 \tag{148}$$

Trocando Y_t e H_t em (145), (146) e (148) obtemos

$$V_t \leq -at + o(t^{0.5+\epsilon}).$$

APÊNDICE D – Cálculo de $E^{\boldsymbol{A}_t}(|x_{t+1}|)$ para $(1{,}2)$-EE, usando interpretação geométrica

Na Seção 3.1 do Capítulo 3, $E^{A_t}(|x_{t+1}|)$ foi obtida em função de λ, a partir da equação $E^{A_t}(|x_{t+1}|) = \frac{\lambda}{(2d_t)^\lambda}\int_{x_t-d_t}^{x_t+d_t}|z|(\text{Vol}B_i)^{\lambda-1}\mathrm{d}z$, com o cálculo do volume de $B_i = \{x_{i,t+1} \in I\ ; |x_{i,t+1}| \geq |z|\} \subset I$, sendo $I = [x_t - d_t, x_t + d_t]$. Com o objetivo de facilitar o entedimento do leitor, neste apêndice calculamos esta esperança para $(1{,}2)$-EE através de uma interpretação gráfica.

Considerando $\lambda = 2$, pela Equação (24), temos

$$E^{A_t}(|x_{t+1}|) = \frac{1}{4d_t^2}\int_{x_t-d_t}^{x_t+d_t}\int_{x_t-d_t}^{x_t+d_t}\min\{|x_{1,t+1}|, |x_{2,t+1}|\}\mathrm{d}x_{2,t+1}\mathrm{d}x_{1,t+1}.$$

Assim, definindo $\varphi : \mathbb{R}^2 \to \mathbb{R}$ por $\varphi(x_{1,t+1}, x_{2,t+1}) = \min\{|x_{1,t+1}|,|x_{2,t+1}|\}$, $E^{A_t}(|x_{t+1}|)$ é dada pela soma (dividida por $(2d_t)^2$) dos volumes dos sólidos que estão acima de $I \times I$ e abaixo da superfície $\gamma = \varphi(x_{1,t+1}, x_{2,t+1})$.

Para o caso em que x_t está próximo à origem ($0 \leq x_t < d_t$), a Figura 33 contém o esboço das regiões de $I \times I$, em que $\min\{|x_{1,t+1}|, |x_{2,t+1}|\} = |x_{1,t+1}|$ (destacadas em azul) e $\min\{|x_{1,t+1}|, |x_{2,t+1}|\} = |x_{2,t+1}|$ (destacadas em rosa).

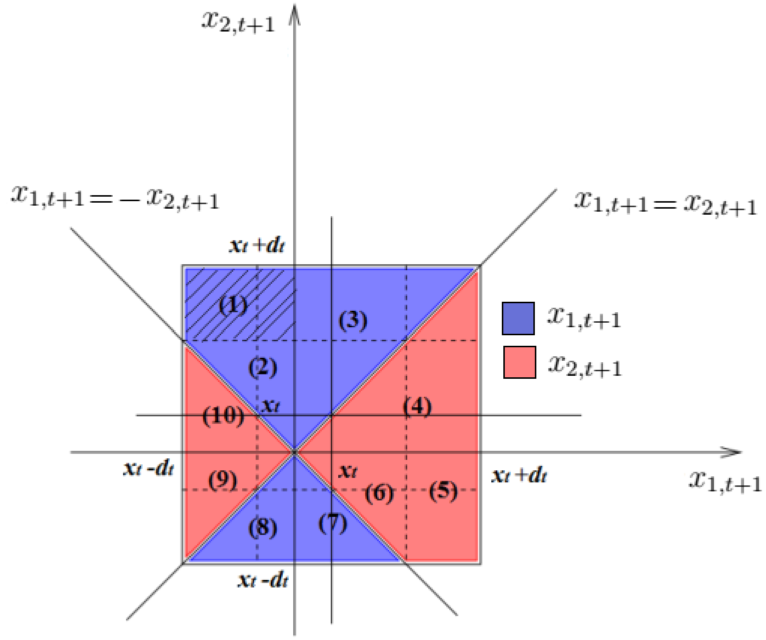

Figura 33 – $0 \leq x_t < d_t$.

Notamos que as áreas das regiões (1) (retângulo hachurado) e (5) são iguais. Também são iguais as áreas de (3) (triângulo azul acima do eixo $x_{1,t+1}$ e à direita do eixo $x_{2,t+1}$) e (4). Além disso, as regiões (2) (triângulo abaixo do retângulo hachurado) , $(6), (7), (8), (9)$ e (10) têm mesma área. Podemos então escrever:

$$E^{A_t}(|x_{t+1}|) = \frac{1}{4d_t^2} \sum_{i=1}^{10} V(i) = 2V(1) + 2V(3) + 6V(2),$$

sendo $V(i)$ o volume do sólido que está acima da região (i) e abaixo de $\gamma = \varphi(x_{1,t+1}, x_{2,t+1})$

Assim

$$\begin{aligned}
E^{A_t}(|x_{t+1}|) &= \frac{1}{4d_t^2} \left\{ 2 \int_{x_t - d_t}^{0} \int_{d_t - x_t}^{x_t + d_t} (-x_{1,t+1}) \, \mathrm{d}x_{2,t+1} \, \mathrm{d}x_{1,t+1} \right. \\
&+ 2 \int_{0}^{x_t + d_t} \int_{x_{1,t+1}}^{x_t + d_t} (x_{1,t+1}) \, \mathrm{d}x_{2,t+1} \, \mathrm{d}x_{1,t+1} \\
&+ \left. 6 \int_{x_t - d_t}^{0} \int_{-x_{1,t+1}}^{d_t - x_t} (-x_{1,t+1}) \, \mathrm{d}x_{2,t+1} \, \mathrm{d}x_{1,t+1} \right\}
\end{aligned}$$

Logo

$$E^{A_t}(|x_{t+1}|) = \frac{1}{4d_t^2}\left\{2[(x_t-d_t)^2x_t] + 2\left[\frac{(x_t+d_t)^3}{6}\right] + 6\left[-\frac{(x_t-d_t)^3}{6}\right]\right\}$$

$$= \frac{1}{4d_t^2}\cdot\frac{4}{3}(x_t^3+d_t^3) = \frac{1}{3}(x_t^3+d_t^3) = \frac{1}{3}\left[\left(\frac{x_t}{d_t}\right)^3+1\right]d_t.$$

Considerando o caso em que x_t está longe da origem ($x_t \geq d_t$), a Figura 34 contém o esboço das regiões de $I \times I$ nas quais o filho escolhido é $x_{1,t+1}$ (região sombreada de azul) e $x_{2,t+1}$ (região sombreada de rosa). Notamos que as áreas das regiões (1) (triângulo azul) e (2) (triângulo rosa) são iguais.

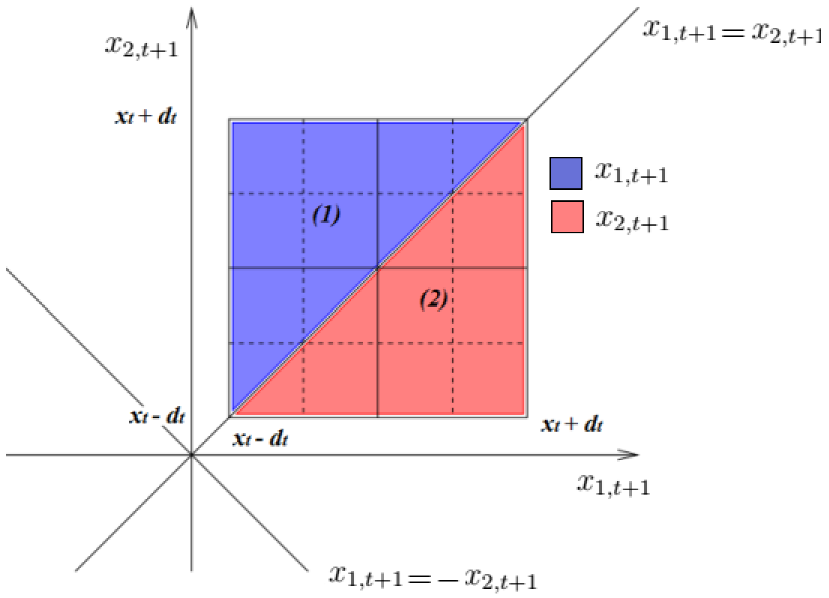

Figura 34 – $x_t \geq d_t$.

Então $E^{A_t}(|x_{t+1}|) = \dfrac{1}{4d_t^2}2V(1)$.

Assim

$$E^{A_t}(|x_{t+1}|) = \frac{2}{4d_t^2}\int_{x_t-d_t}^{x_t+d_t}\int_{x_{1,t+1}}^{x_t+d_t}(x_{1,t+1})\,\mathrm{d}x_{2,t+1}\,\mathrm{d}x_{1,t+1}$$

$$= \frac{2}{4d_t^2}\left(\frac{12x_td_t^2-4d_t^3}{6}\right) = x_t - \frac{1}{3}d_t = \left(\frac{x_t}{d_t}-\frac{1}{3}\right)d_t.$$

APÊNDICE E – Expressões de Ψ

A função $\Psi(r) = \ln\left[\frac{E^{A_t}(|x_{t+1}|)+wE^{A_t}(d_{t+1})}{|x_t|+wd_t}\right] - k(E^{A_t}[\ln(d_{t+1})] - \ln(d_t))$, com $r = \frac{|x_t|}{d_t}$, $w,k \in \mathbb{R}$, $w > 0$, e $0 \leq k < 1$, representa uma cota superior para o decrescimento esprado do supermatingal V_t. Sua representação em função do parâmetro λ das EEs $(1 \overset{+}{,} \lambda)$ estudadas nesta tese foi fundamental para encontrarmos limites superiores para as taxas de convergência. Por isso, apresentamos neste apêndice todas as expressões de $\Psi(r)$ omitidas no texto.

E.1 $(1+\lambda)$-EE com Adaptação do Tamanho do Passo baseada em Número de Descendentes bem sucedidos - **Regra 2**

Seguem as expressões de $\Psi(r)$ para $(1+\lambda)$-EE com a **Regra 2** de adaptação do tamanho do passo.

$$\begin{aligned}\Psi_A(r) &= \ln\left[\frac{\left(\frac{1}{\lambda+1}\right)[1-(1-r)^{\lambda+1}]+w\sum_{j=0}^{\lambda}\alpha_j\binom{\lambda}{j}(1-r)^{\lambda-j}r^j}{w+r}\right] \\ &-k\left[\sum_{j=0}^{\lambda}\ln(\alpha_j)\binom{\lambda}{j}(1-r)^{\lambda-j}r^j\right]\end{aligned} \tag{149}$$

$$\Psi_B(r) = \ln\left[\frac{\left(\frac{1}{\lambda+1}\right)\left(r^{\lambda+1}+\frac{2^\lambda-1}{2^\lambda}\right)+\frac{w}{2^\lambda}\sum_{j=0}^{\lambda}\alpha_j\binom{\lambda}{j}}{w+r}\right] - k\frac{1}{2^\lambda}\sum_{j=0}^{\lambda}\ln(\alpha_j)\binom{\lambda}{j} \tag{150}$$

$$\Psi_C(r) = \ln\left[\frac{r+\frac{2^\lambda(1-\lambda)-1}{2^\lambda(\lambda+1)}+\frac{w}{2^\lambda}\sum_{j=0}^{\lambda}\alpha_j\binom{\lambda}{j}}{w+r}\right] - k\frac{1}{2^\lambda}\sum_{j=0}^{\lambda}\ln(\alpha_j)\binom{\lambda}{j} \tag{151}$$

E.2 $(1 \overset{+}{,} \lambda)$-EE com Adaptação do Tamanho do Passo Baseada em Sucesso - **Regra 1**

Seguem as expressões de $\Psi(r)$ para $(1,\lambda)$-EE com a **Regra 1** de adaptação do tamanho do passo.

$$\begin{aligned}\Psi_A(r) &= \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1)+w\left[(1-r)^\lambda(\alpha_f-\alpha_s)+\alpha_s\right]}{r+w}\right\} \\ &-k\{(1-r)^\lambda[\ln(\alpha_f)-\ln(\alpha_s)]+\ln(\alpha_s)\}\end{aligned} \tag{152}$$

$$\Psi_B(r) = \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)(r^{\lambda+1}+1)+w\left[\frac{1}{2^\lambda}(\alpha_f-\alpha_s)+\alpha_s\right]}{r+w}\right\}$$
$$-k\left\{\tfrac{1}{2^\lambda}[\ln(\alpha_f)-\ln(\alpha_s)]+\ln(\alpha_s)\right\} \tag{153}$$

$$\Psi_C(r) = \ln\left\{\frac{r-\left(\frac{\lambda-1}{\lambda+1}\right)+w\left[\frac{1}{2^\lambda}(\alpha_f-\alpha_s)+\alpha_s\right]}{r+w}\right\}$$
$$-k\left\{\tfrac{1}{2^\lambda}[\ln(\alpha_f)-\ln(\alpha_s)]+\ln(\alpha_s)\right\} \tag{154}$$

Seguem as expressões de $\Psi(r)$ para $(1+\lambda)$-EE com a **Regra 1** de adaptação do tamanho do passo.

$$\Psi_A(r) = \ln\left\{\frac{\frac{1}{\lambda+1}-(1-r)^\lambda\left(\frac{1-r}{\lambda+1}\right)+w[(1-r)^\lambda(\alpha_f-\alpha_s)+\alpha_s]}{r+w}\right\}$$
$$-k\{(1-r)^\lambda[\ln(\alpha_f)-\ln(\alpha_s)]+\ln(\alpha_s)\} \tag{155}$$

$$\Psi_B(r) = \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)r^{\lambda+1}+\left(\frac{1}{\lambda+1}\right)\left(1-\frac{1}{2^\lambda}\right)+w\left[\frac{1}{2^\lambda}(\alpha_f-\alpha_s)+\alpha_s\right]}{r+w}\right\}$$
$$-k\left\{\tfrac{1}{2^\lambda}[\ln(\alpha_f)-\ln(\alpha_s)]+\ln(\alpha_s)\right\} \tag{156}$$

$$\Psi_C(r) = \ln\left\{\frac{r+\frac{1-\lambda}{\lambda+1}-\frac{1}{2^\lambda(\lambda+1)}+w\left[\frac{1}{2^\lambda}(\alpha_f-\alpha_s)+\alpha_s\right]}{r+w}\right\}$$
$$-k\left\{\tfrac{1}{2^\lambda}[\ln(\alpha_f)-\ln(\alpha_s)]+\ln(\alpha_s)\right\} \tag{157}$$

E.3 $(1+\lambda)$-EE com Controle de Mutação Baseado no Tamanho do Passo

Seguem as expressões de $\Psi(r)$ para $(1+\lambda)$-EE com controle de mutação baseado no tamanho do passo.

$$\begin{aligned}\Psi_A(r) &= \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)\left[1-(1-r)^{\lambda+1}\right]+\frac{w}{2}(\alpha^+-\alpha^-)\left[(1-\theta+r)^\lambda+(1-\theta-r)^\lambda\right]+w\alpha^-}{w+r}\right\}\\ &- k\left\{\frac{1}{2}[(1-\theta+r)^\lambda+(1-\theta-r)^\lambda][\ln(\alpha^+)-\ln(\alpha^-)]+\ln(\alpha^-)\right\}\end{aligned} \tag{158}$$

$$\begin{aligned}\Psi_{B_1}(r) &= \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)\left[1-(1-r)^{\lambda+1}\right]+\frac{w}{2}(\alpha^--\alpha^+)\left[(1+\theta-r)^\lambda-(1-\theta-r)^\lambda\right]+w\alpha^+}{w+r}\right\}\\ &- k\left\{\frac{1}{2}[(1+\theta-r)^\lambda-(1-\theta-r)^\lambda][\ln(\alpha^-)-\ln(\alpha^+)]+\ln(\alpha^+)\right\}\end{aligned} \tag{159}$$

$$\Psi_{B_2}(r) = \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)\left[1-(1-r)^{\lambda+1}\right]+\frac{w}{2}(\alpha^+-\alpha^-)\left[(1-\theta+r)^\lambda - r^\lambda + 2\left(\frac{1-\theta}{2}\right)^\lambda\right]+w\alpha^-}{w+r}\right\}$$

$$- k\left\{\frac{1}{2}\left[(1-\theta+r)^\lambda - r^\lambda + 2\left(\frac{1-\theta}{2}\right)^\lambda\right]\left[\ln(\alpha^+)-\ln(\alpha^-)\right]+\ln(\alpha^-)\right\} \tag{160}$$

$$\Psi_{C_1}(r) = \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)\left[1-(1-r)^{\lambda+1}\right]+\frac{w}{2}(\alpha^--\alpha^+)\left[(1+\theta-r)^\lambda + r^\lambda - 2\left(\frac{1-\theta}{2}\right)^\lambda\right]+w\alpha^+}{w+r}\right\}$$

$$- k\left\{\frac{1}{2}\left[(1+\theta-r)^\lambda + r^\lambda - 2\left(\frac{1-\theta}{2}\right)^\lambda\right]\left[\ln(\alpha^-)-\ln(\alpha^+)\right]+\ln(\alpha^+)\right\} \tag{161}$$

$$\Psi_{C_2}(r) = \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)\left(r^{\lambda+1}+\frac{2^\lambda-1}{2^\lambda}\right)+\frac{w}{2}(\alpha^+-\alpha^-)\left[(1-\theta+r)^\lambda - r^\lambda + 2\left(\frac{1-\theta}{2}\right)^\lambda\right]+w\alpha^-}{w+r}\right\}$$

$$- k\left\{\frac{1}{2}\left[(1-\theta+r)^\lambda - r^\lambda + 2\left(\frac{1-\theta}{2}\right)^\lambda\right]\left[\ln(\alpha^+)-\ln(\alpha^-)\right]+\ln(\alpha^-)\right\} \tag{162}$$

$$\Psi_{D}(r) = \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)\left(r^{\lambda+1}+\frac{2^\lambda-1}{2^\lambda}\right)+\frac{w}{2}(\alpha^--\alpha^+)\left[(1+\theta-r)^\lambda + r^\lambda - 2\left(\frac{1-\theta}{2}\right)^\lambda\right]+w\alpha^+}{w+r}\right\}$$

$$- k\left\{\frac{1}{2}\left[(1+\theta-r)^\lambda + r^\lambda - 2\left(\frac{1-\theta}{2}\right)^\lambda\right]\left[\ln(\alpha^-)-\ln(\alpha^+)\right]+\ln(\alpha^+)\right\} \tag{163}$$

$$\Psi_{E}(r) = \ln\left\{\frac{\left(\frac{1}{\lambda+1}\right)\left(r^{\lambda+1}+\frac{2^\lambda-1}{2^\lambda}\right)+w(\alpha^--\alpha^+)\left[\left(\frac{1+\theta}{2}\right)^\lambda - \left(\frac{1-\theta}{2}\right)^\lambda\right]+w\alpha^+}{w+r}\right\}$$

$$- k\left\{\left[\left(\frac{1+\theta}{2}\right)^\lambda - \left(\frac{1-\theta}{2}\right)^\lambda\right]\left[\ln(\alpha^-)-\ln(\alpha^+)\right]+\ln(\alpha^+)\right\} \tag{164}$$

$$
\begin{aligned}
\Psi_F(r) &= \ln\left\{\frac{r + \frac{2^\lambda(1-\lambda)-1}{2^\lambda(\lambda+1)} + w(\alpha^- - \alpha^+)\left[\left(\frac{1+\theta}{2}\right)^\lambda - \left(\frac{1-\theta}{2}\right)^\lambda\right] + w\alpha^+}{w+r}\right\} \\
&- k\left\{\left[\left(\frac{1+\theta}{2}\right)^\lambda - \left(\frac{1-\theta}{2}\right)^\lambda\right]\left[\ln(\alpha^-) - \ln(\alpha^+)\right] + \ln(\alpha^+)\right\}
\end{aligned} \tag{165}
$$

APÊNDICE F – Resultados Experimentais

Neste apêndice são exibidos todos os experimentos que geram os quantis empíricos das distribuições de e^{V_t}, $|x_t|$ e d_t ao longo das gerações para $(1 \overset{+}{,} \lambda)$-EE. Estes resultados foram obtidos considerando a função objetivo $f(x) = |x|$ e as configurações que levaram aos melhores limites de convergência. As EE foram executadas 10001 vezes com os valores ótimos de todos os parâmetros. O estado inicial (x_0,d_0) dos algoritmos foi escolhido de modo que $V_0 = 0$ em todos os casos.

F.1 Regras de Adaptação do Tamanho do Passo Baseadas em Sucesso

Nas Figuras 35, 36 e 37 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1,\lambda)$-EE (para $\lambda \in \{2, 3, 4, 5, 6, 7\}$) com **Regra 1** , tendo iniciado em $x_0 = 0$, juntamente com os limites de convergência correspondentes.

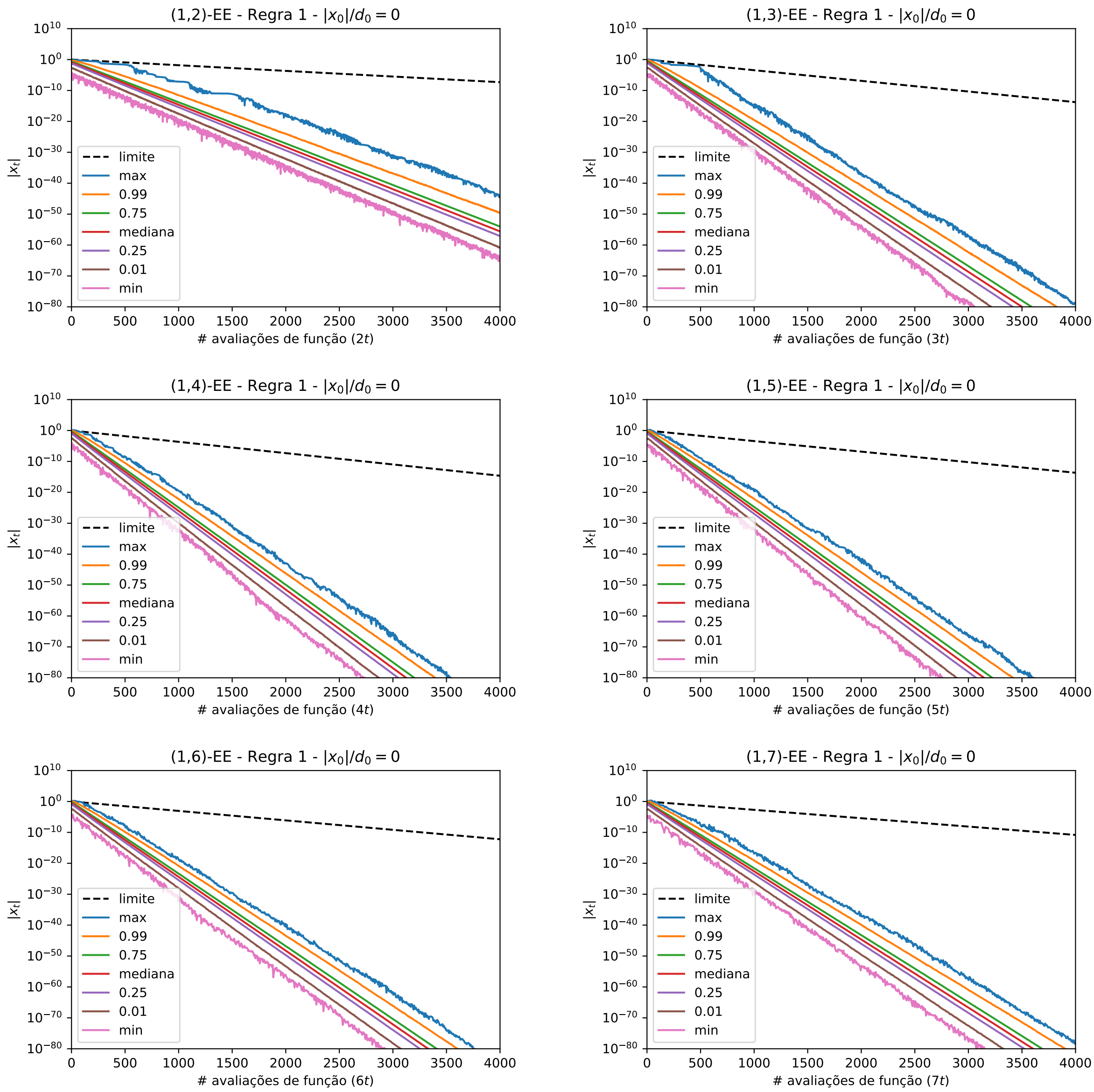

Figura 35 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

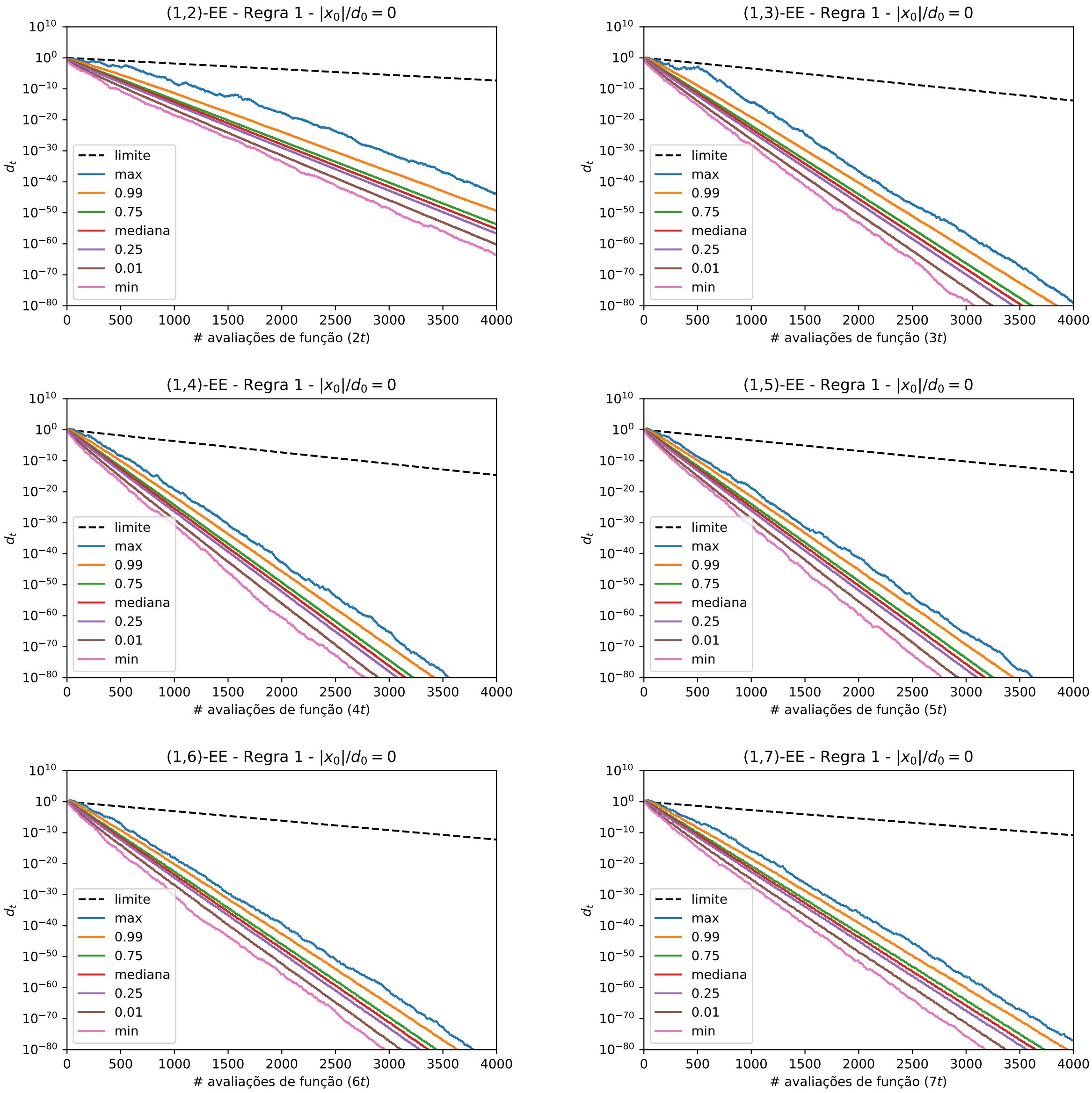

Figura 36 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1{,}\lambda)$-EE com **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

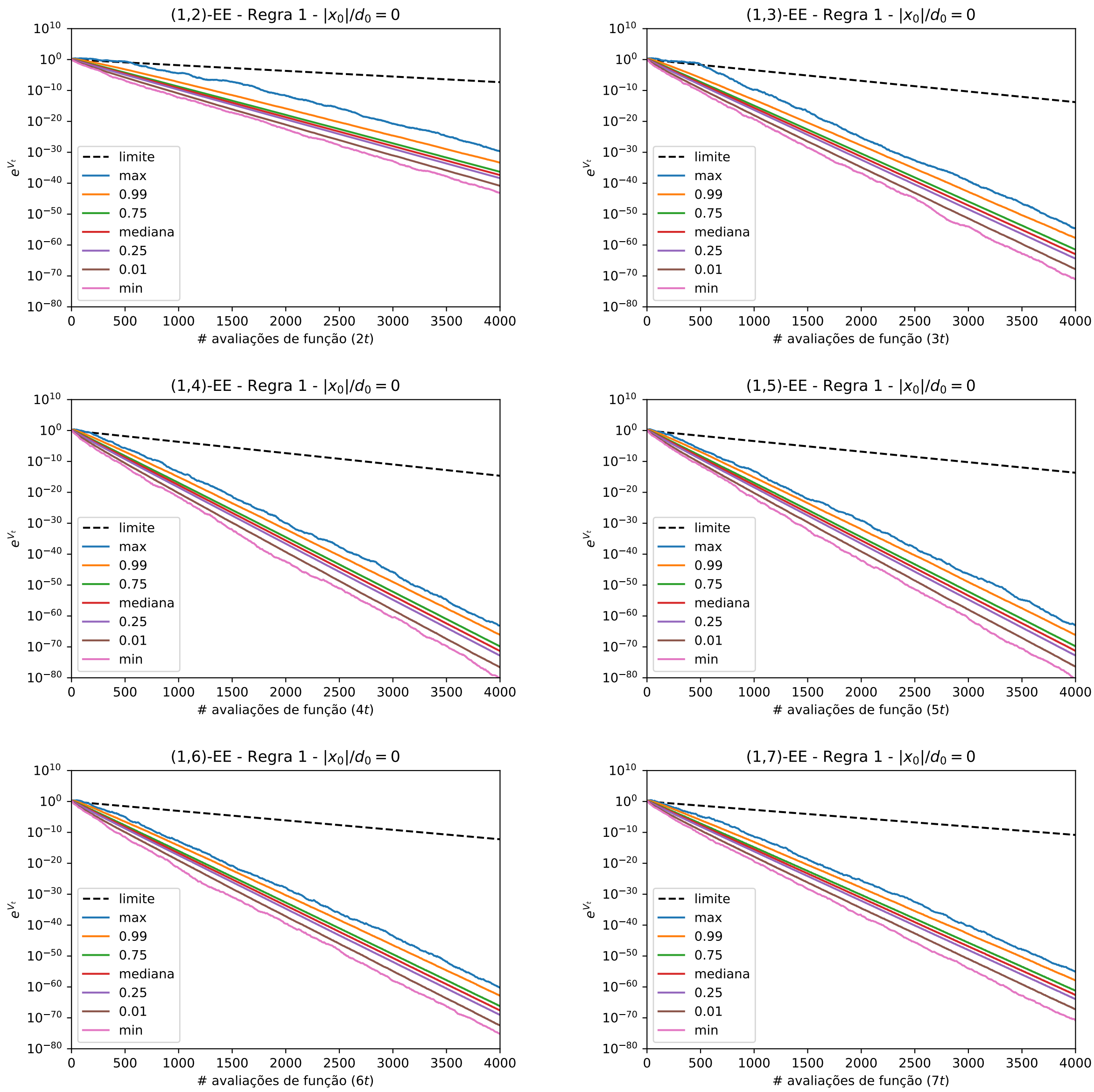

Figura 37 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Nas Figuras 38, 39 e 40 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1,\lambda)$-EE (para $\lambda \in \{2, 3, 4, 5, 6, 7, 8\}$) com **Regra 2** , tendo iniciado em $x_0 = 0$, juntamente com os limites de convergência correspondentes.

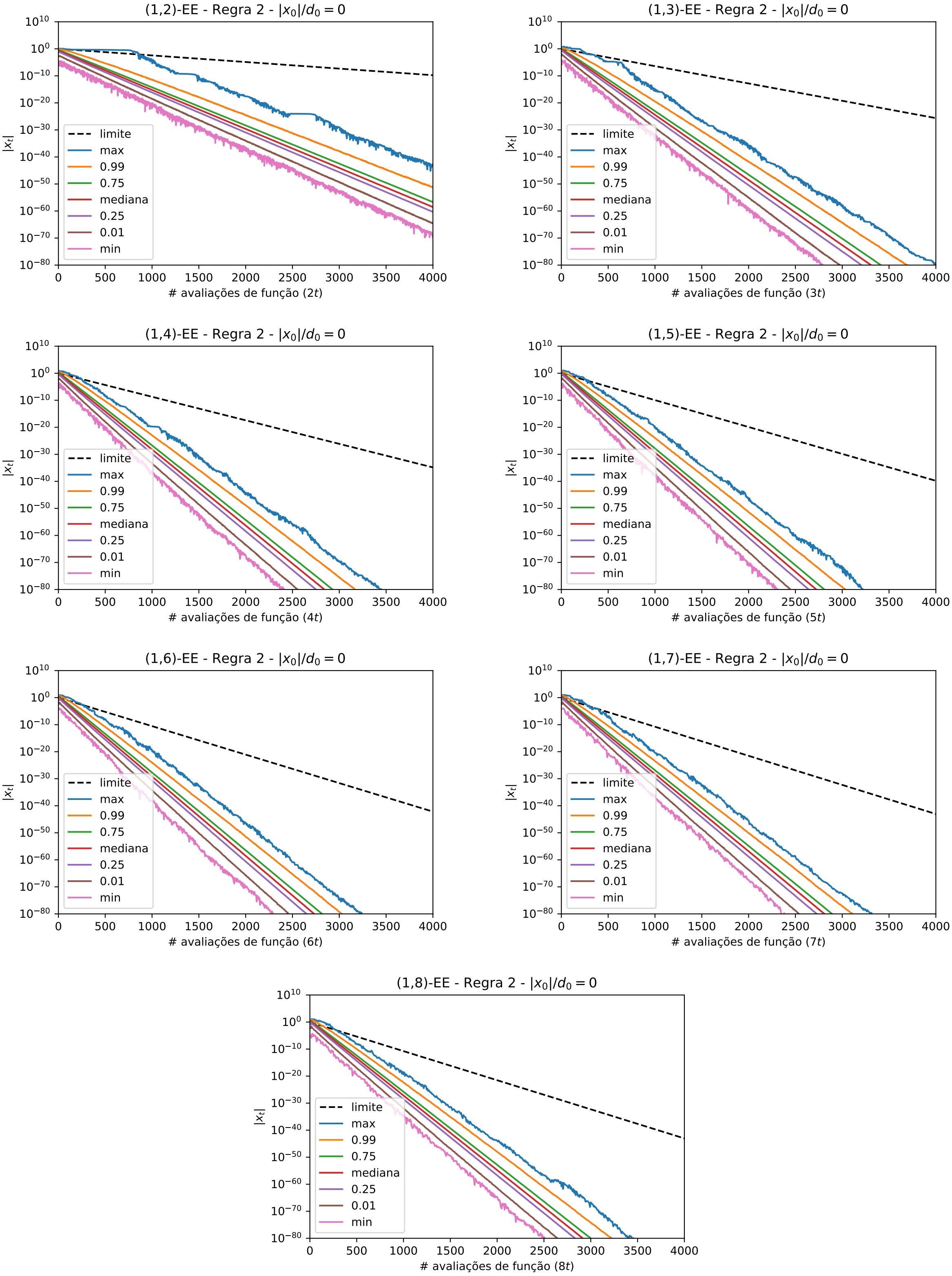

Figura 38 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

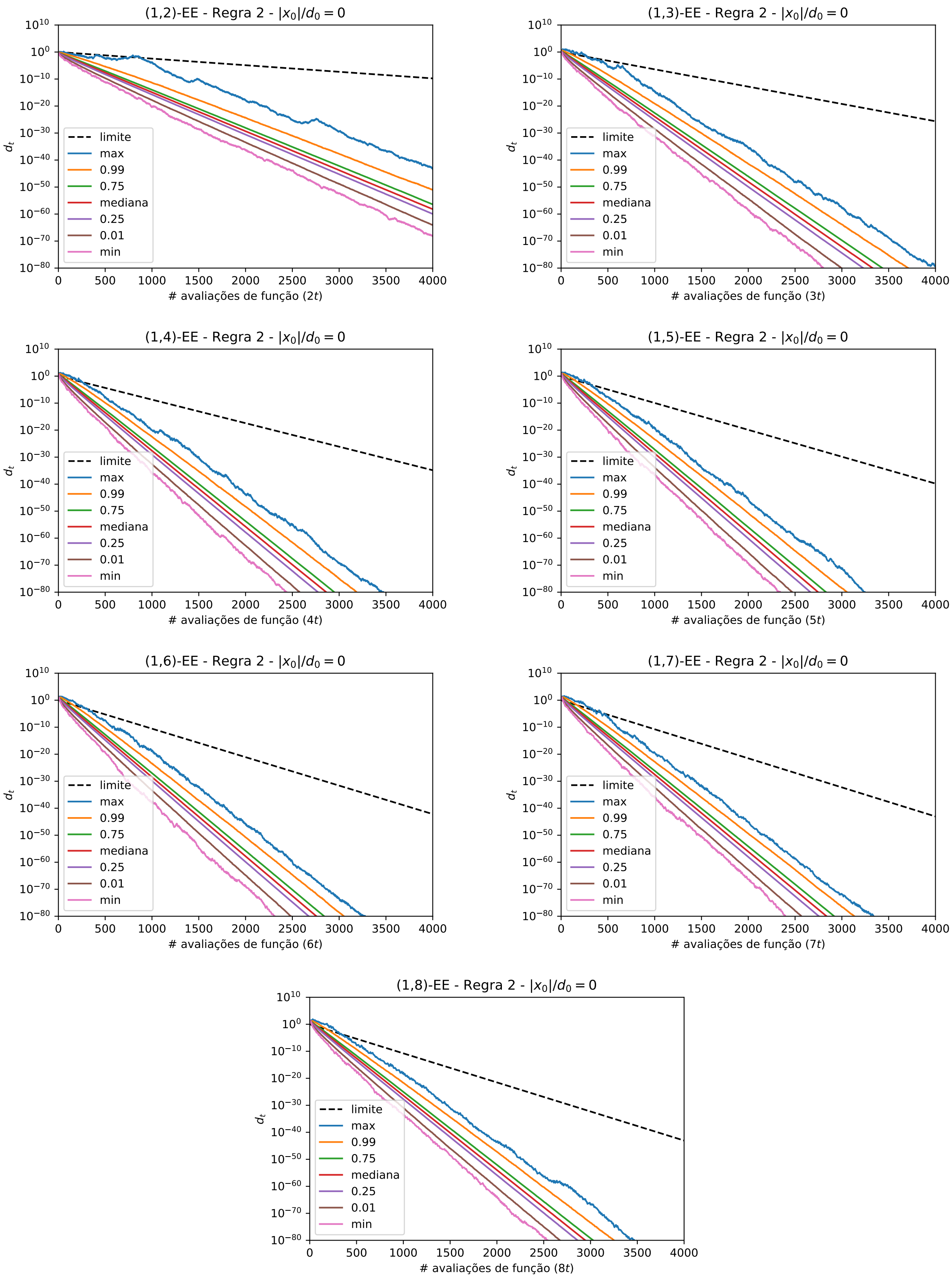

Figura 39 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

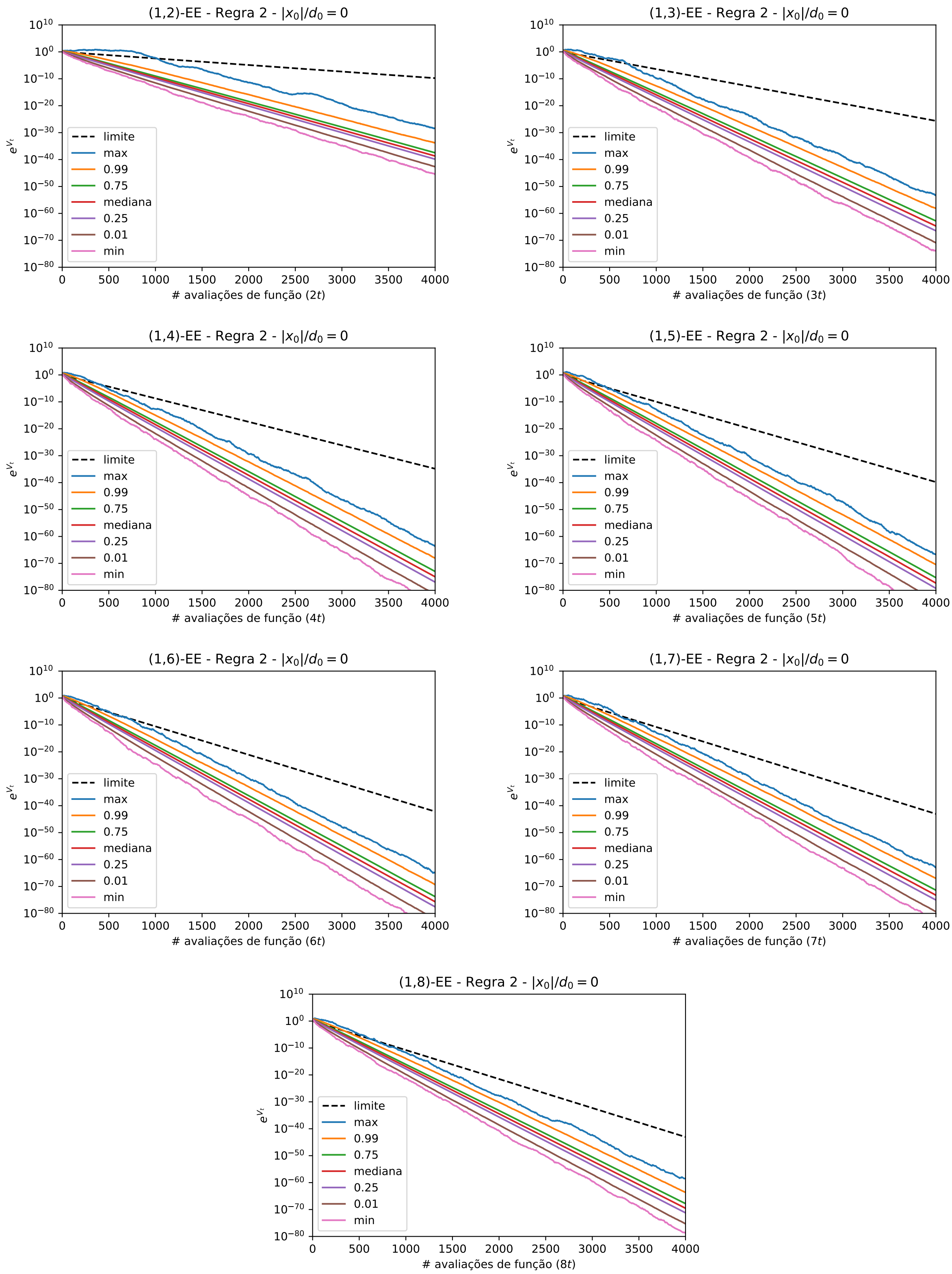

Figura 40 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Nas Figuras 41, 42 e 43 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1,\lambda)$-EE (para $\lambda \in \{2,3,4,5,6,7\}$) com **Regra 1** , tendo iniciado em x_0 tal que $|x_0|/d_0 = 10^{15}$, juntamente com os limites de convergência correspondentes.

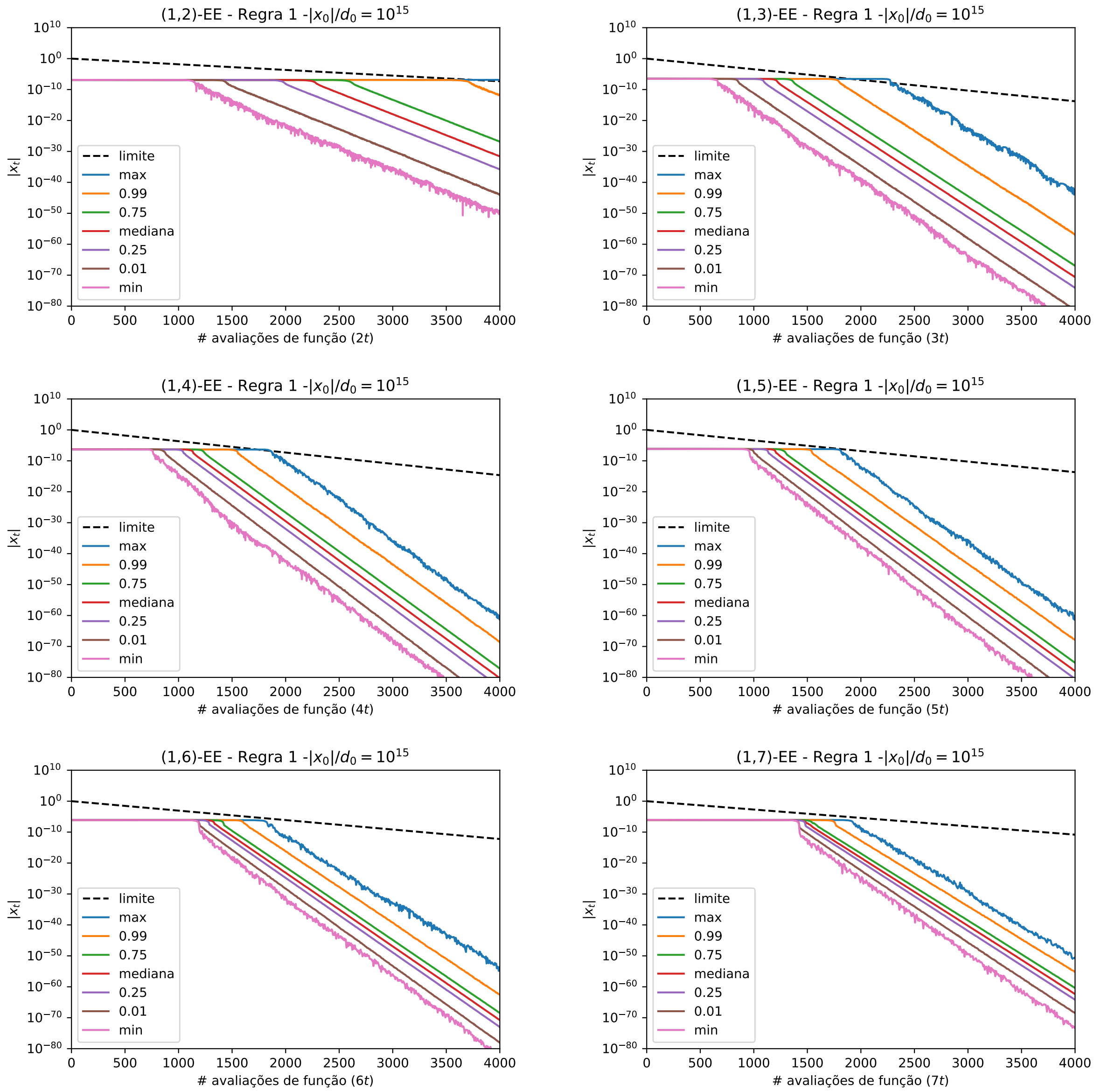

Figura 41 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 1** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

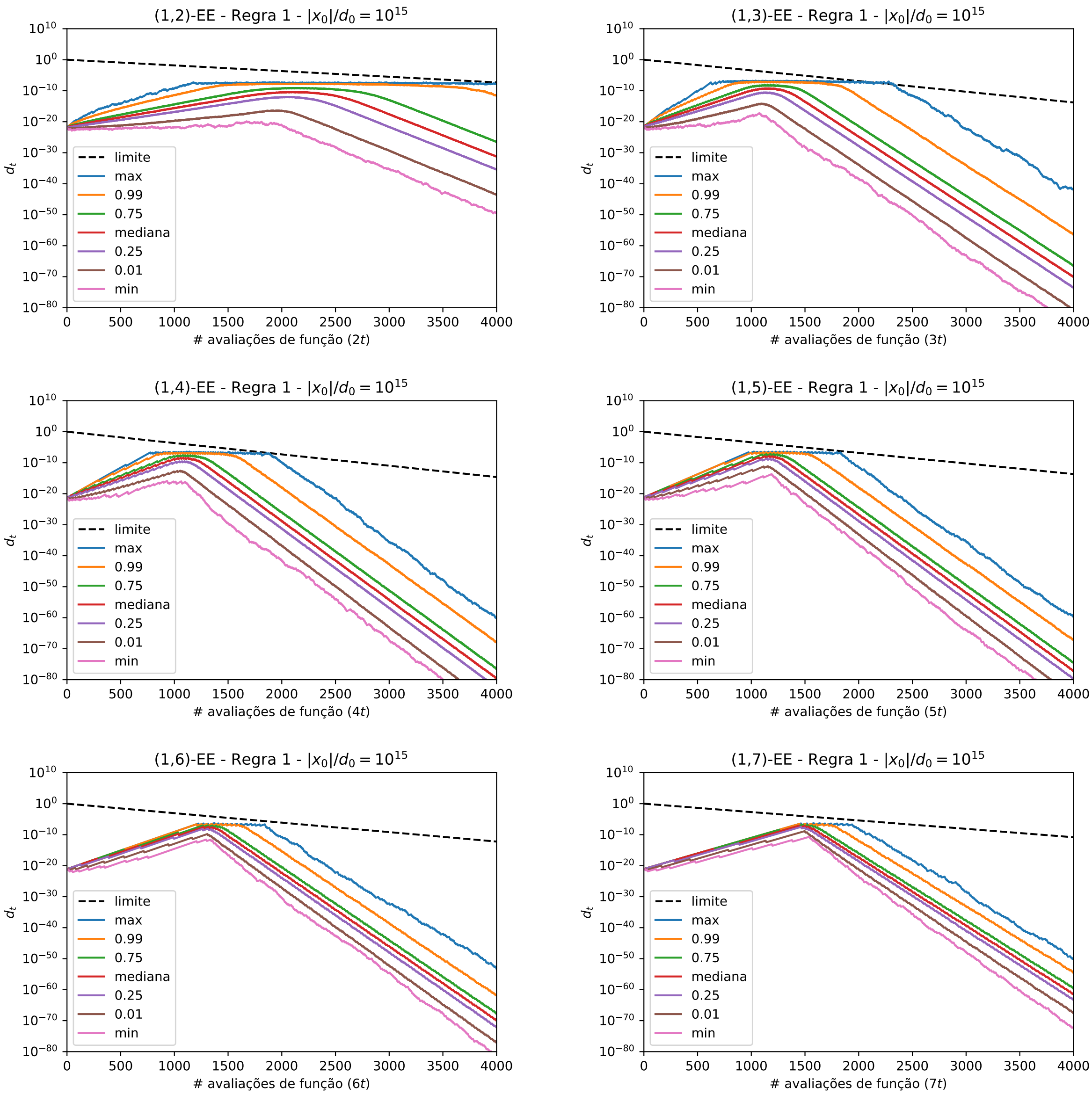

Figura 42 – Quantis empíricos das distribuições de $|d_t|$ estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 1** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

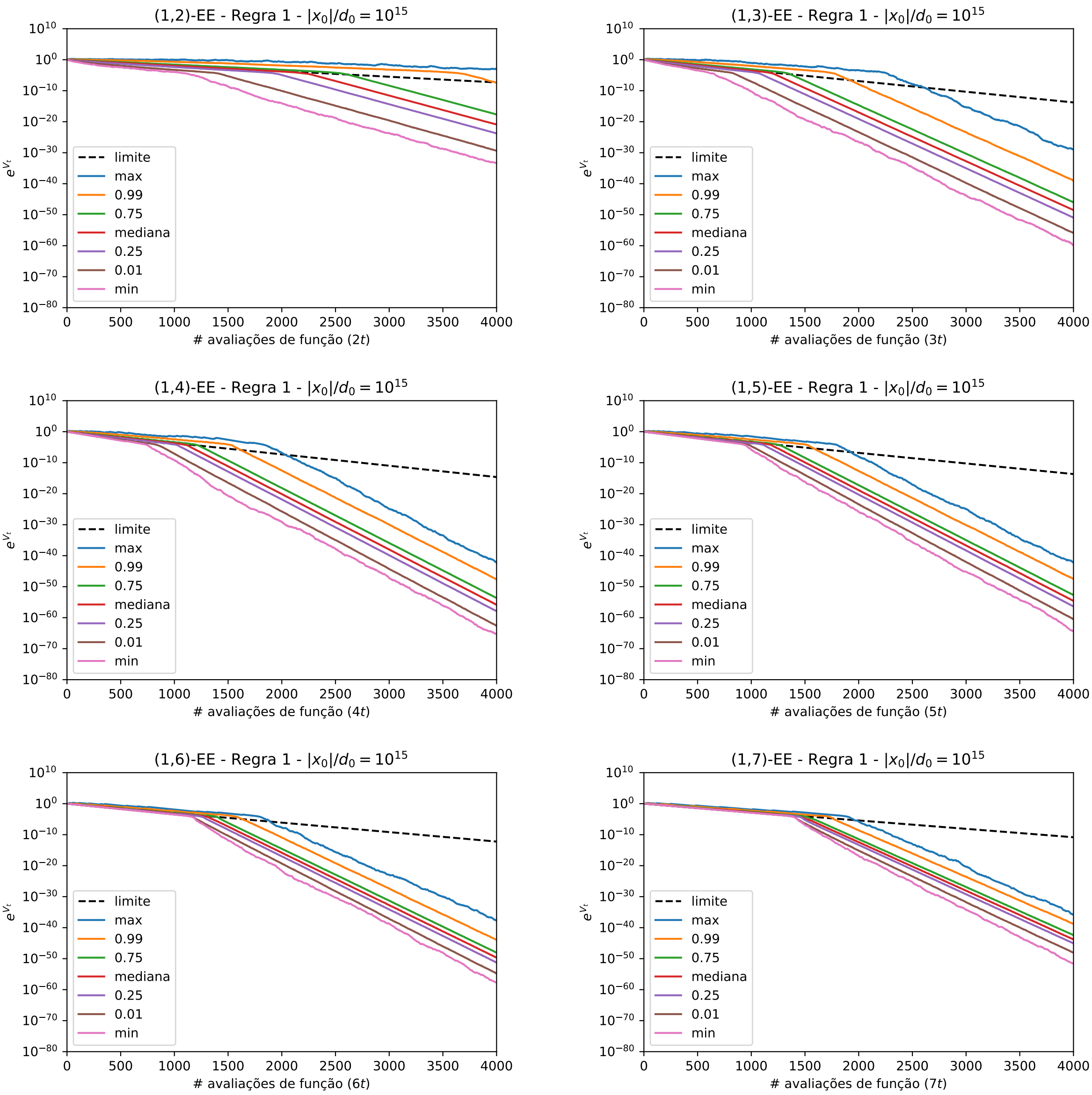

Figura 43 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 1** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

Nas Figuras 44, 45 e 46 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1,\lambda)$-EE (para $\lambda \in \{2, 3, 4, 5, 6, 7, 8\}$) com **Regra 2** , tendo iniciado em x_0 tal que $|x_0|/d_0 = 10^{15}$, juntamente com os limites de convergência correspondentes.

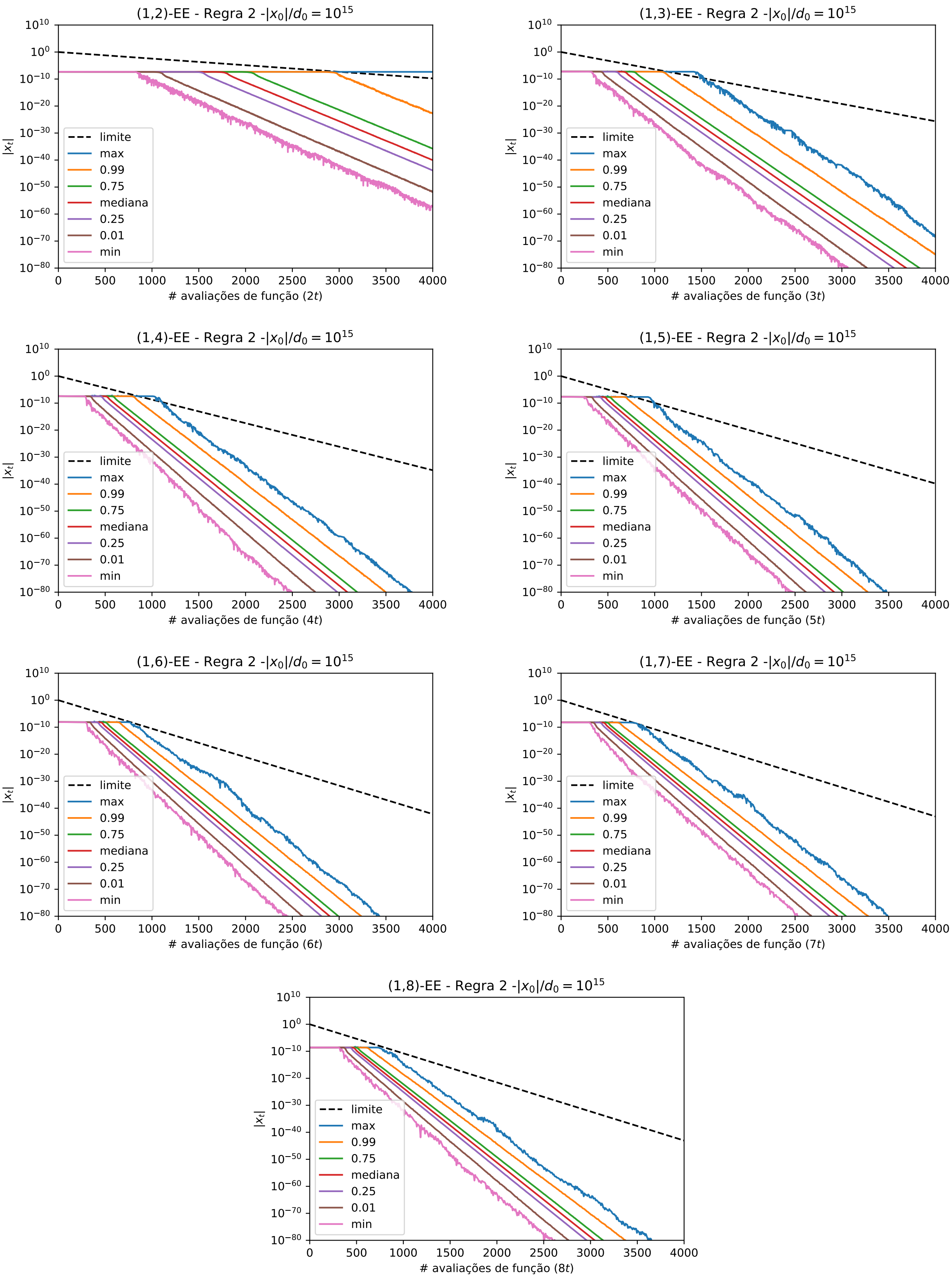

Figura 44 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 10^{15}$).

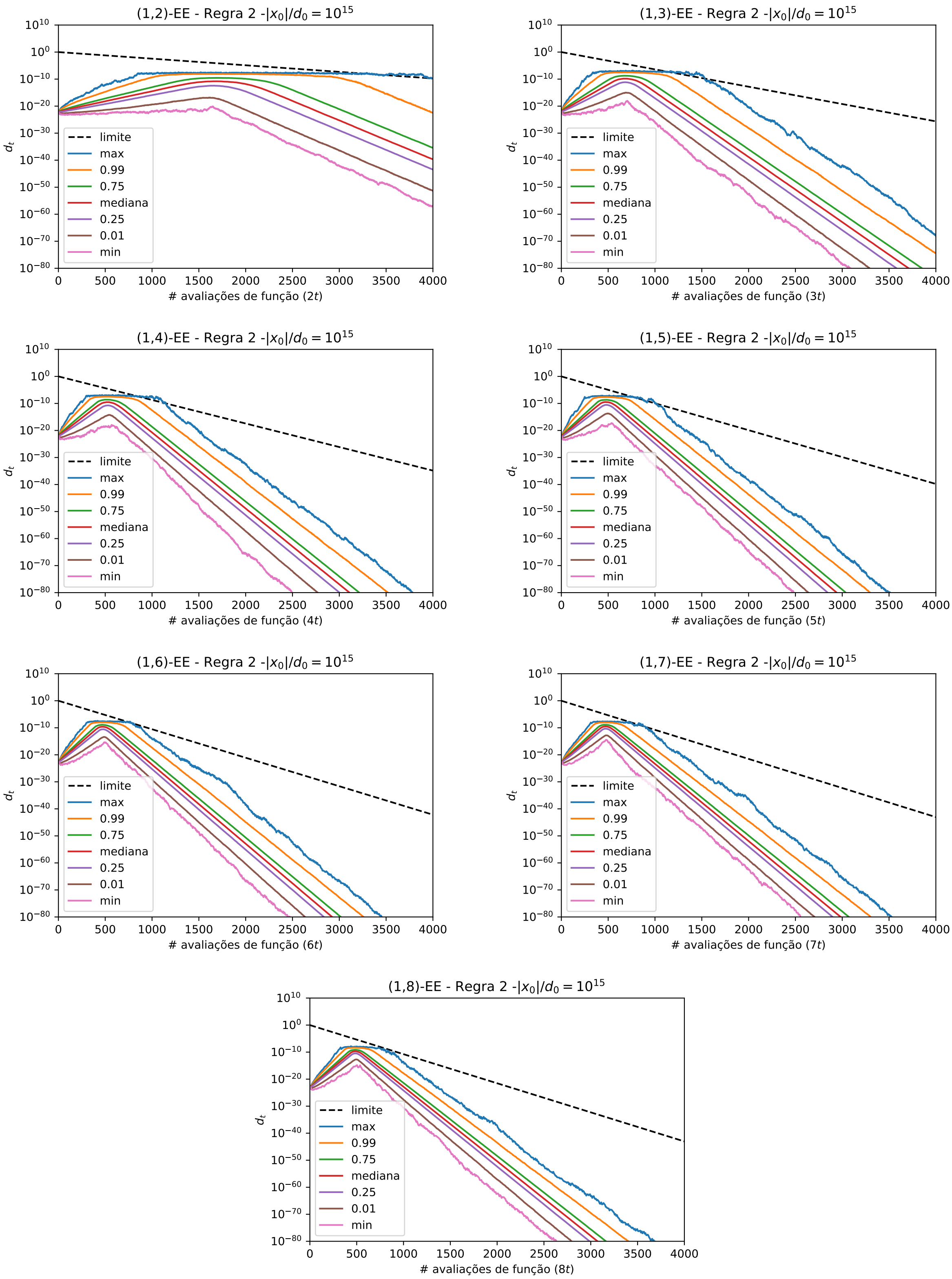

Figura 45 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 10^{15}$).

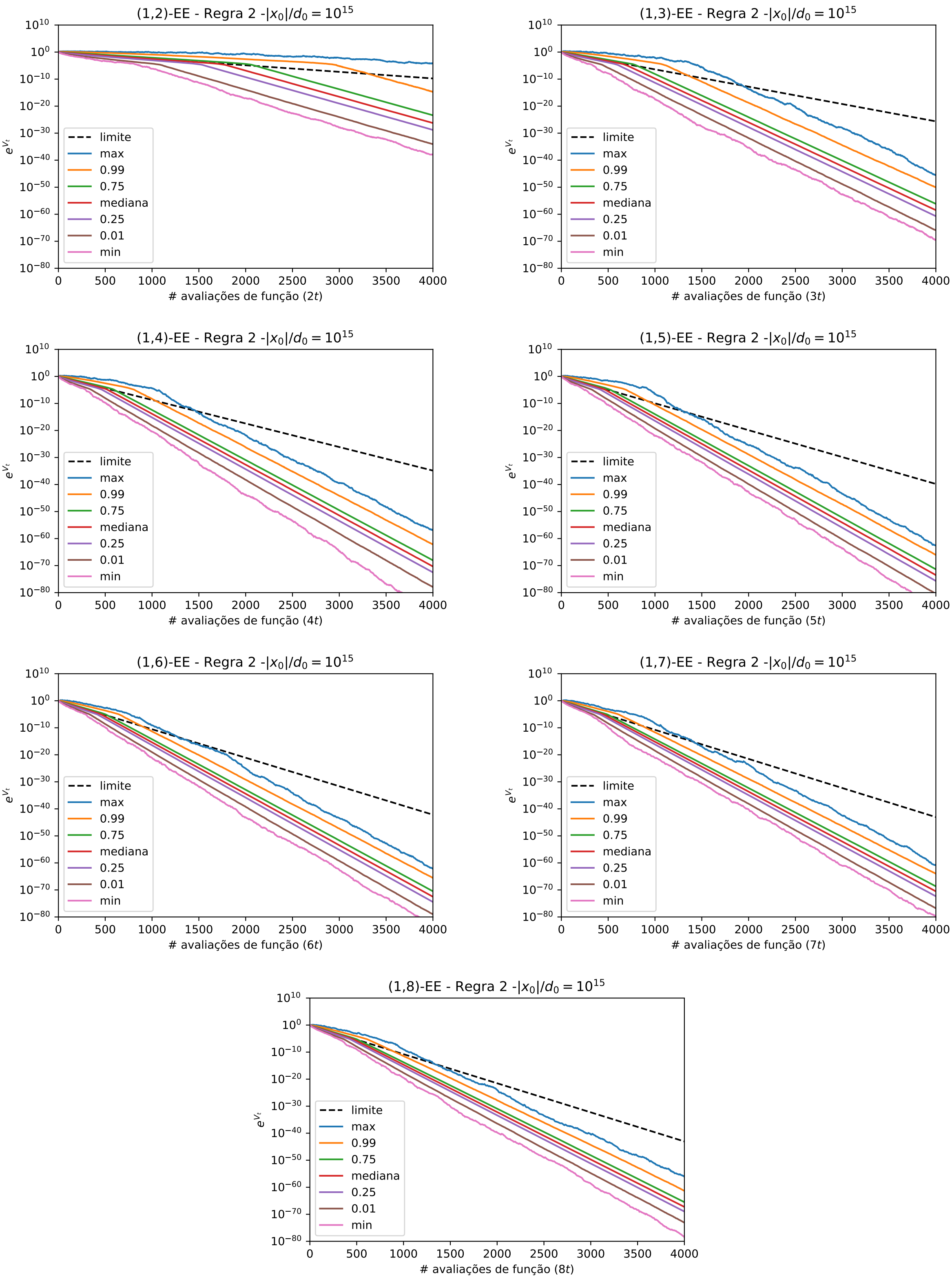

Figura 46 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1{,}\lambda)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 10^{15}$).

Nas Figuras 47, 48 e 49 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1+\lambda)$-EE (para $\lambda \in \{1,2,3,4,5,6,7\}$) com **Regra 1**

, tendo iniciado em $x_0 = 0$, juntamente com os limites de convergência correspondentes.

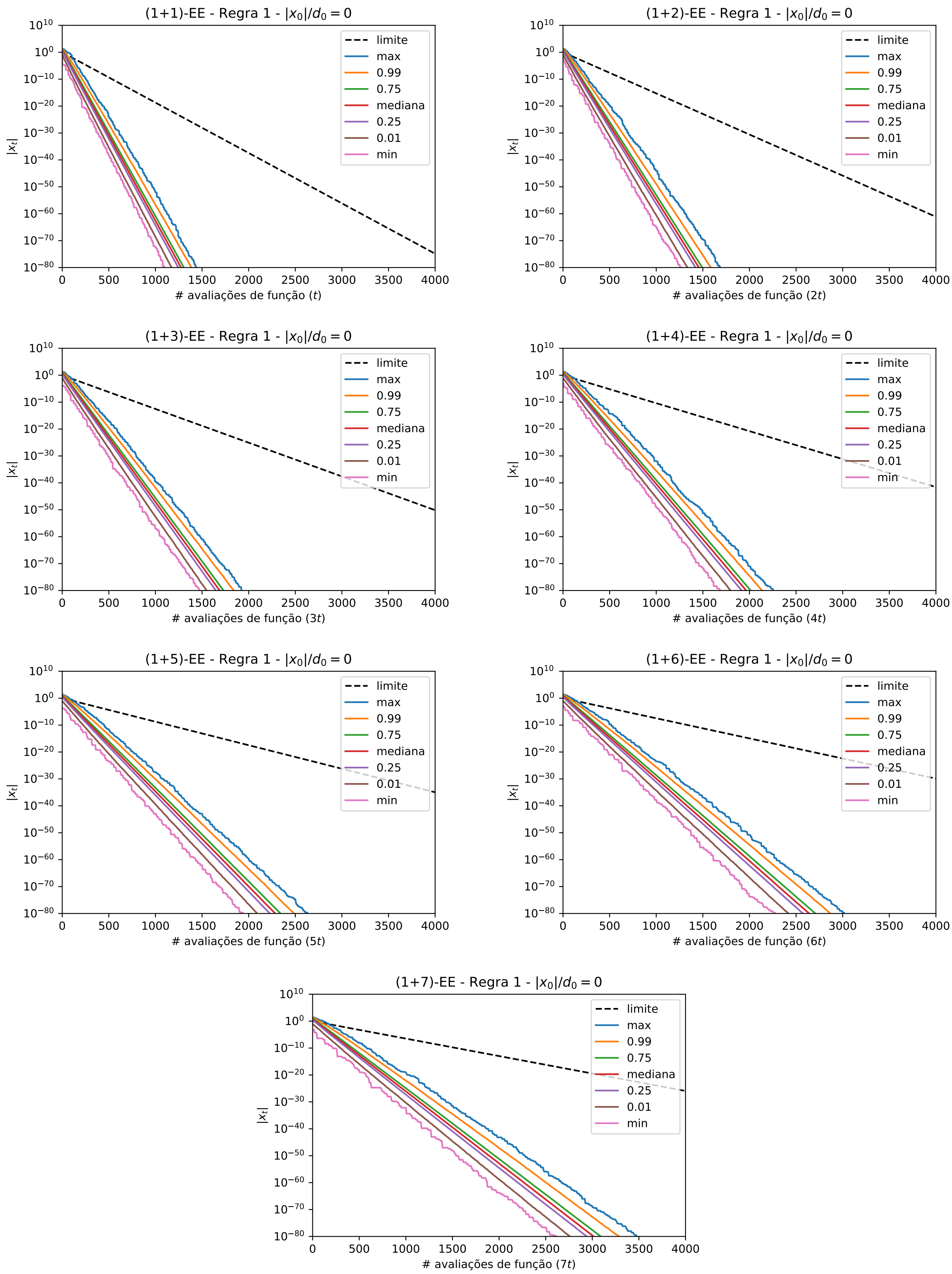

Figura 47 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

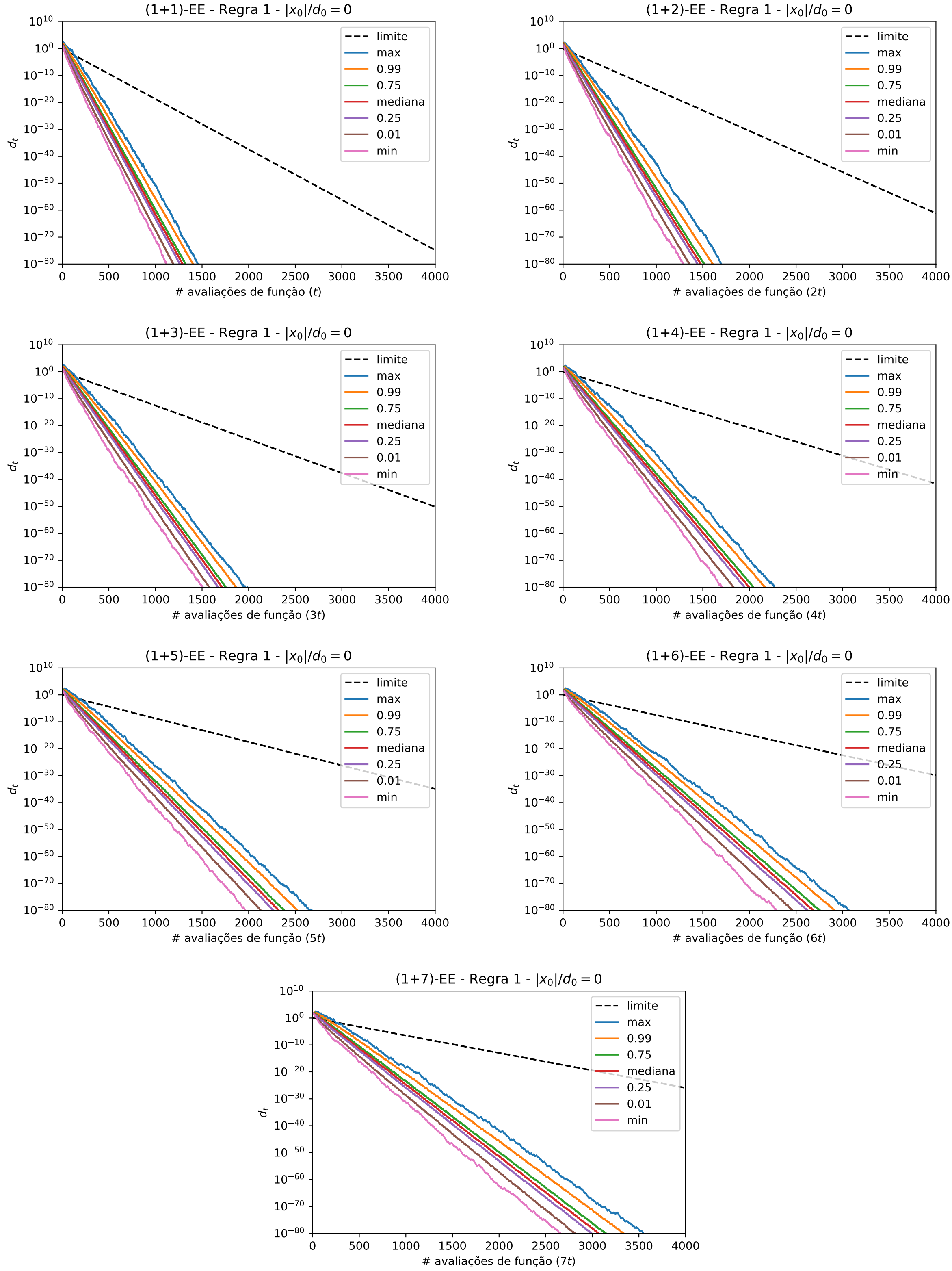

Figura 48 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

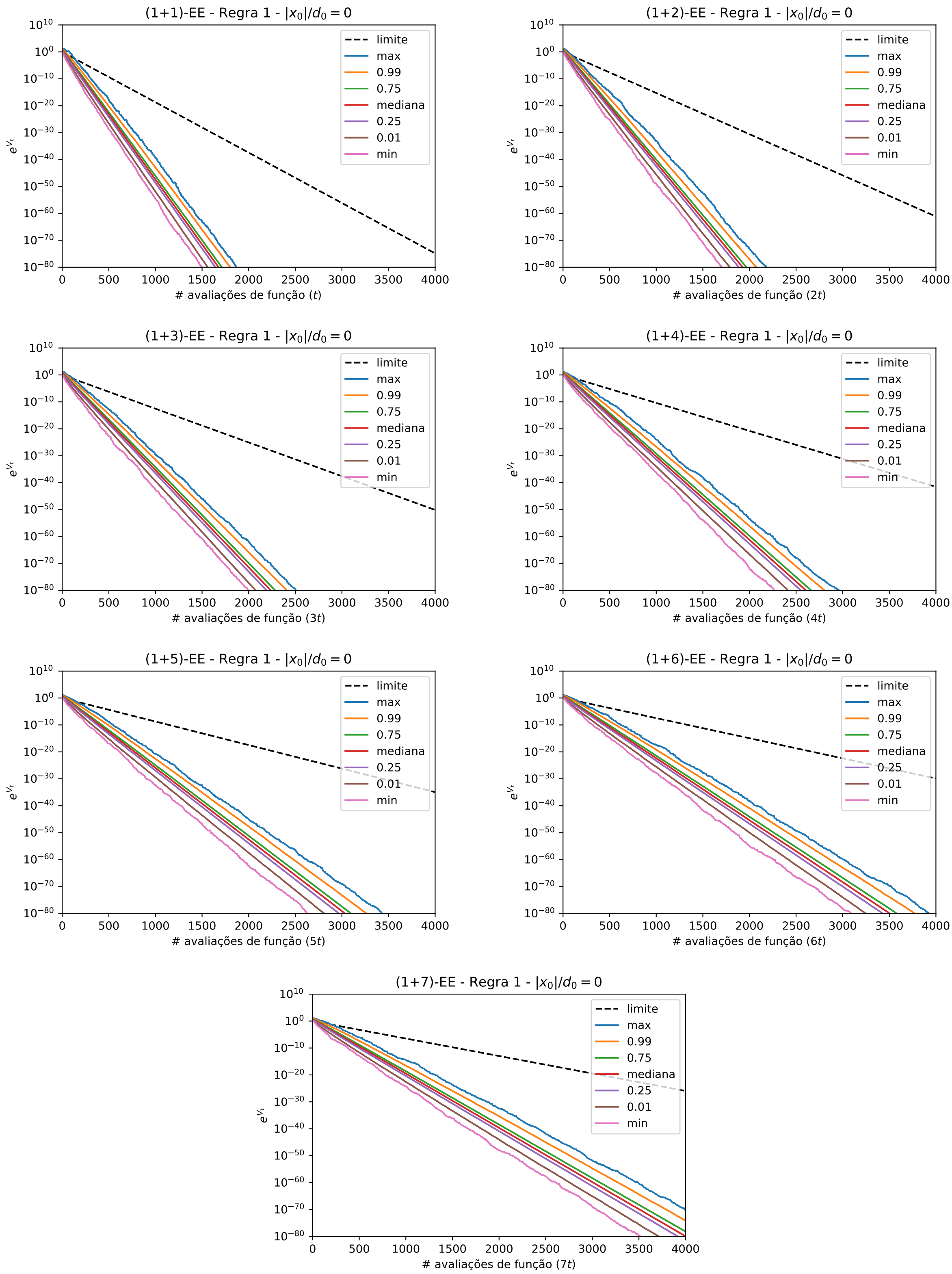

Figura 49 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 1** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Nas Figuras 50, 51 e 52 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1+\lambda)$-EE (para $\lambda \in \{1,2,3,4,5,6,7\}$) com **Regra 2**

, tendo iniciado em $x_0 = 0$, juntamente com os limites de convergência correspondentes.

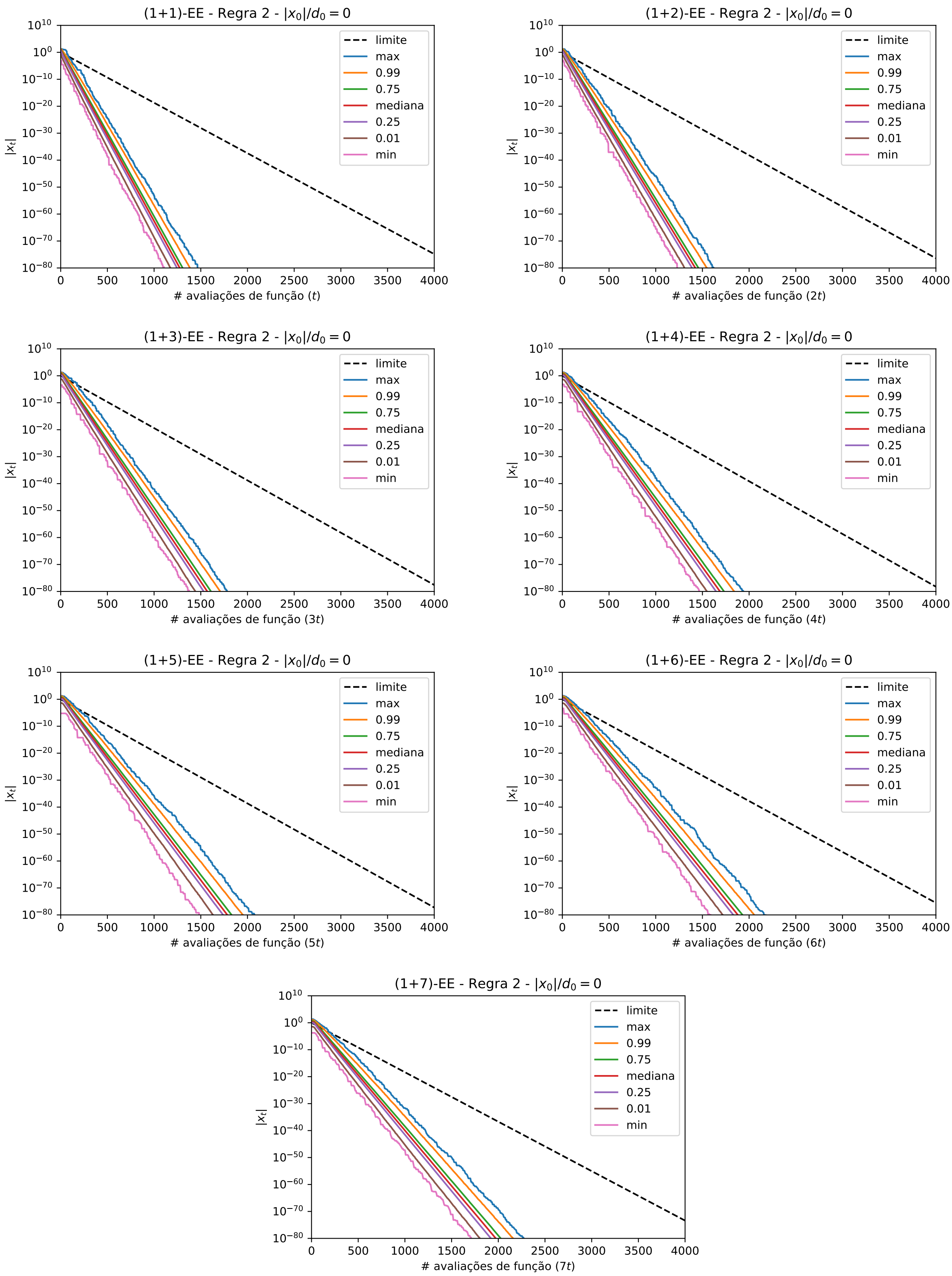

Figura 50 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

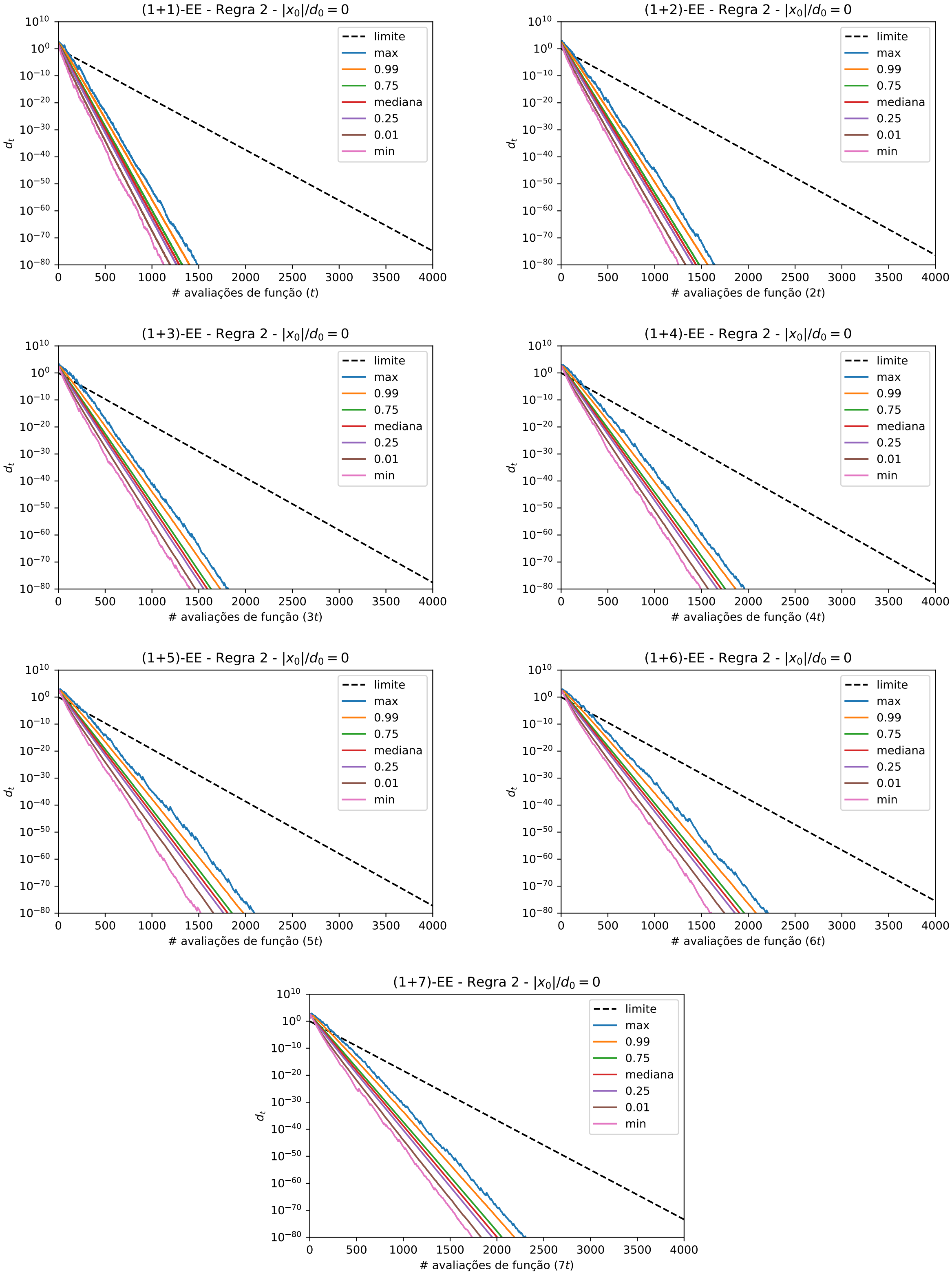

Figura 51 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

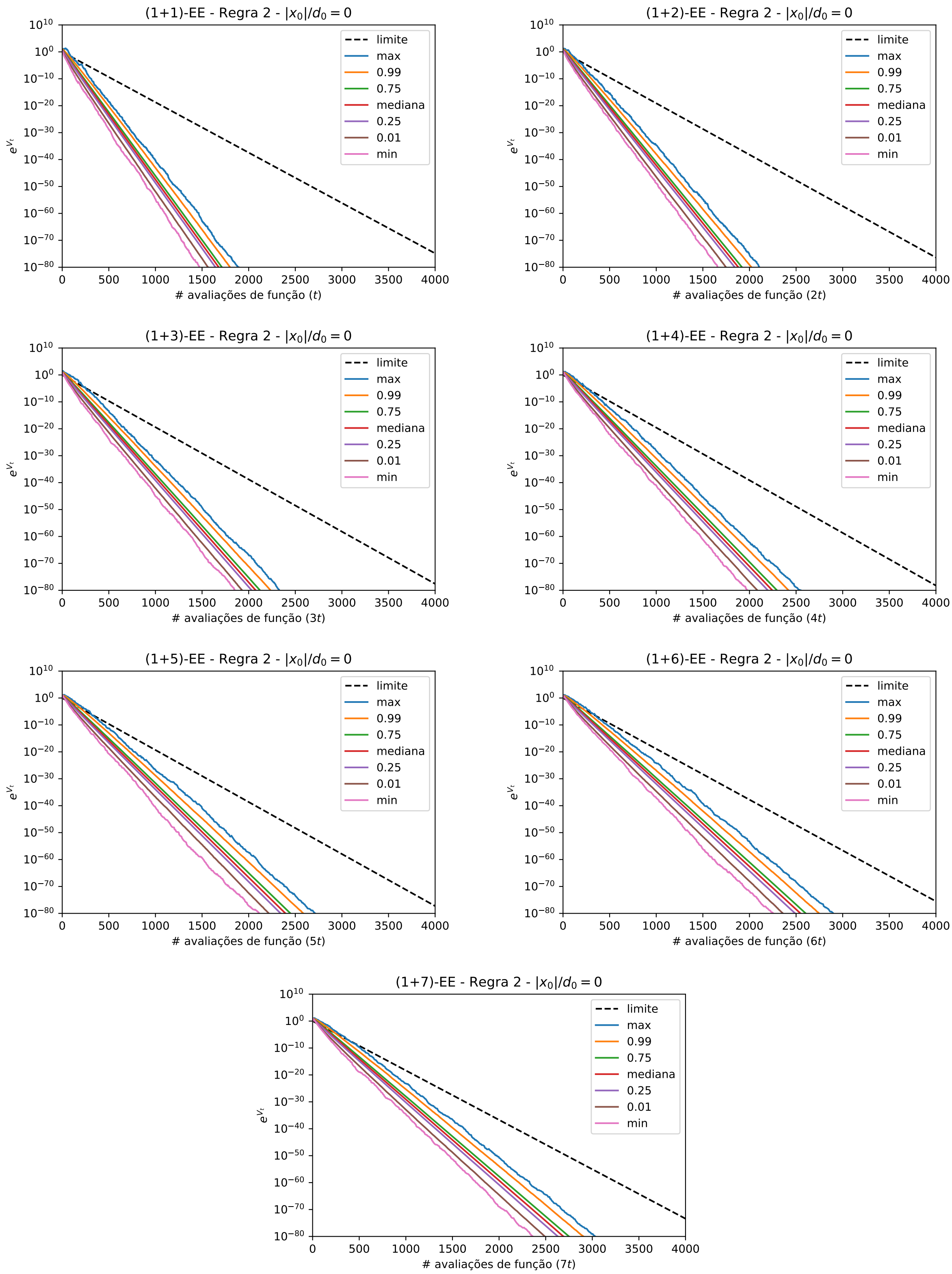

Figura 52 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 2** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Nas Figuras 53, 54 e 55 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1+\lambda)$-EE (para $\lambda \in \{1,2,3,4,5,6,7\}$) com **Regra 1**

, tendo iniciado em x_0 tal que $|x_0|/d_0 = 10^{15}$, juntamente com os limites de convergência correspondentes.

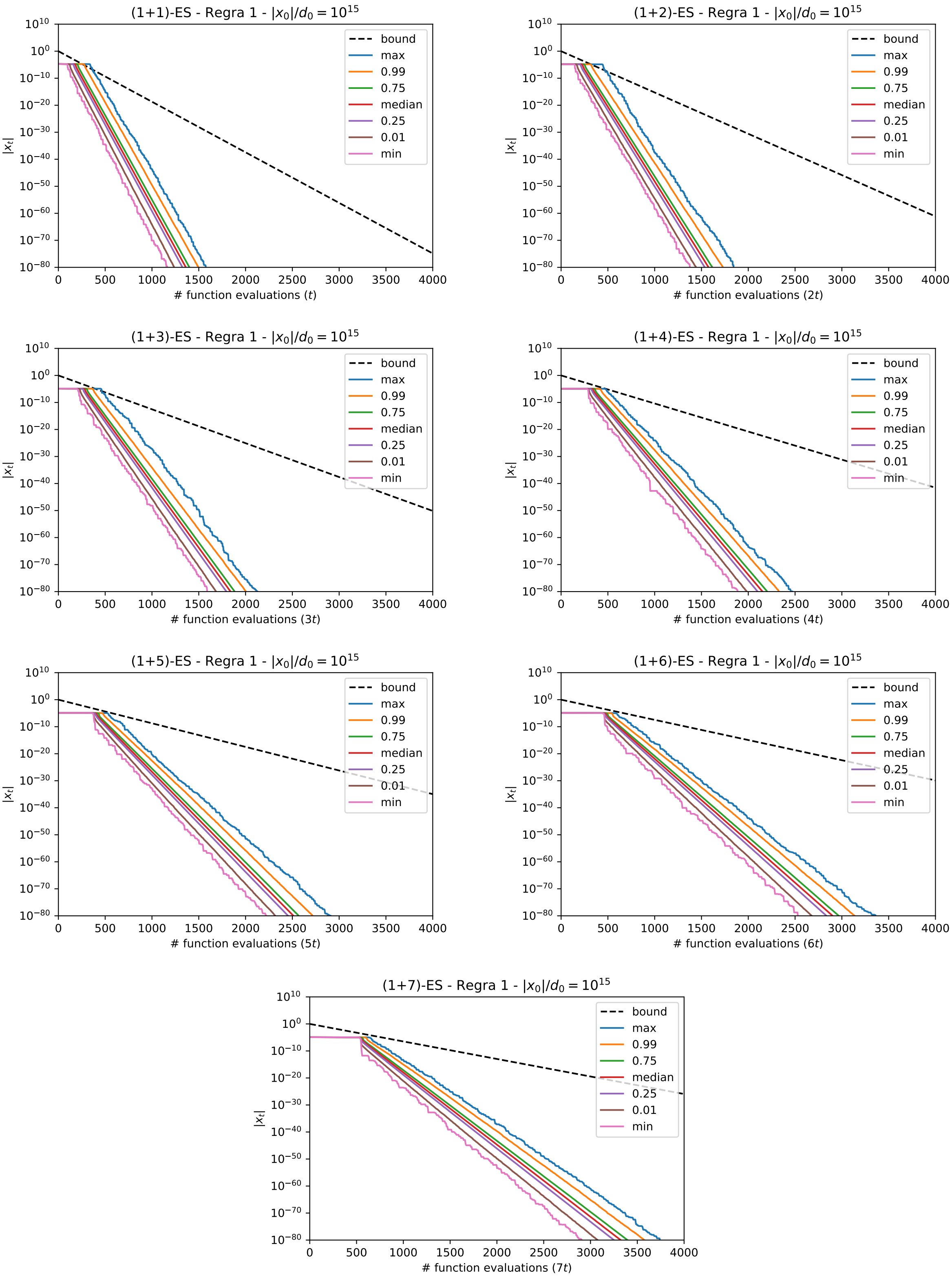

Figura 53 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1 + \lambda)$-EE com **Regra 1** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

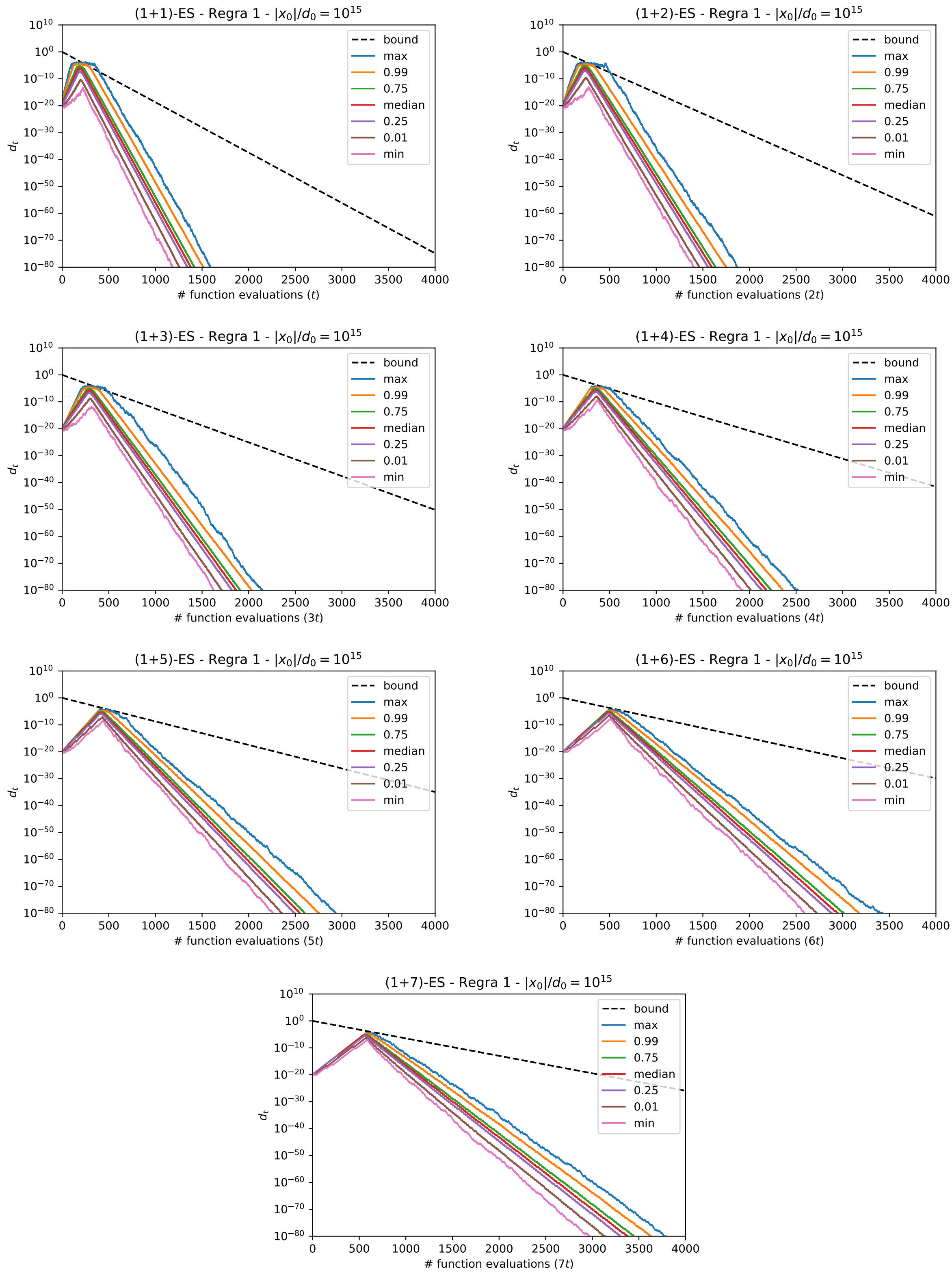

Figura 54 – Quantis empíricos das distribuições de d_t com estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 1** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

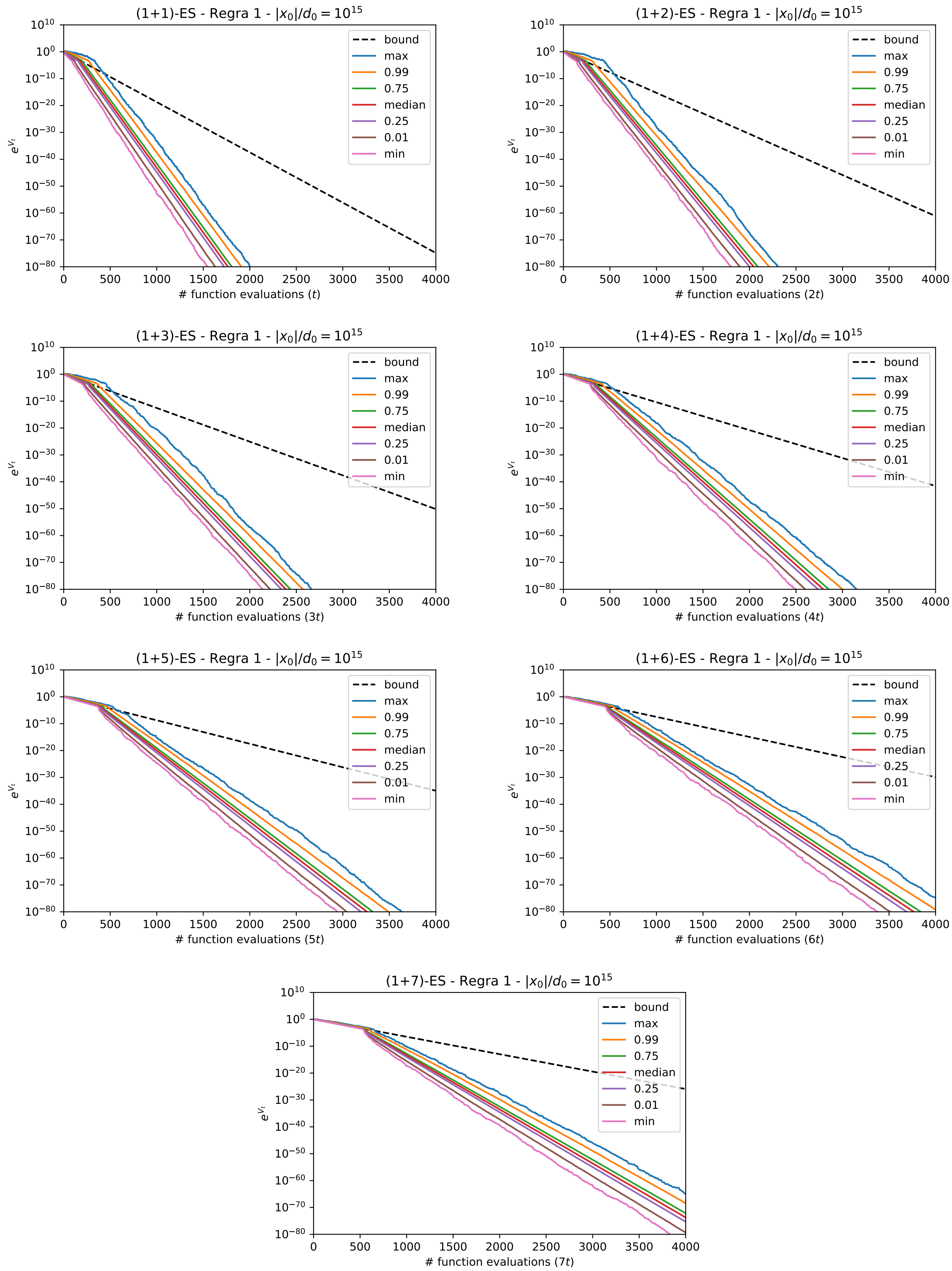

Figura 55 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 1** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

Nas Figuras 56, 57 e 58 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1+\lambda)$-EE (para $\lambda \in \{1,2,3,4,5,6,7\}$) com **Regra 2**

, tendo iniciado em x_0 tal que $|x_0|/d_0 = 10^{15}$, juntamente com os limites de convergência correspondentes.

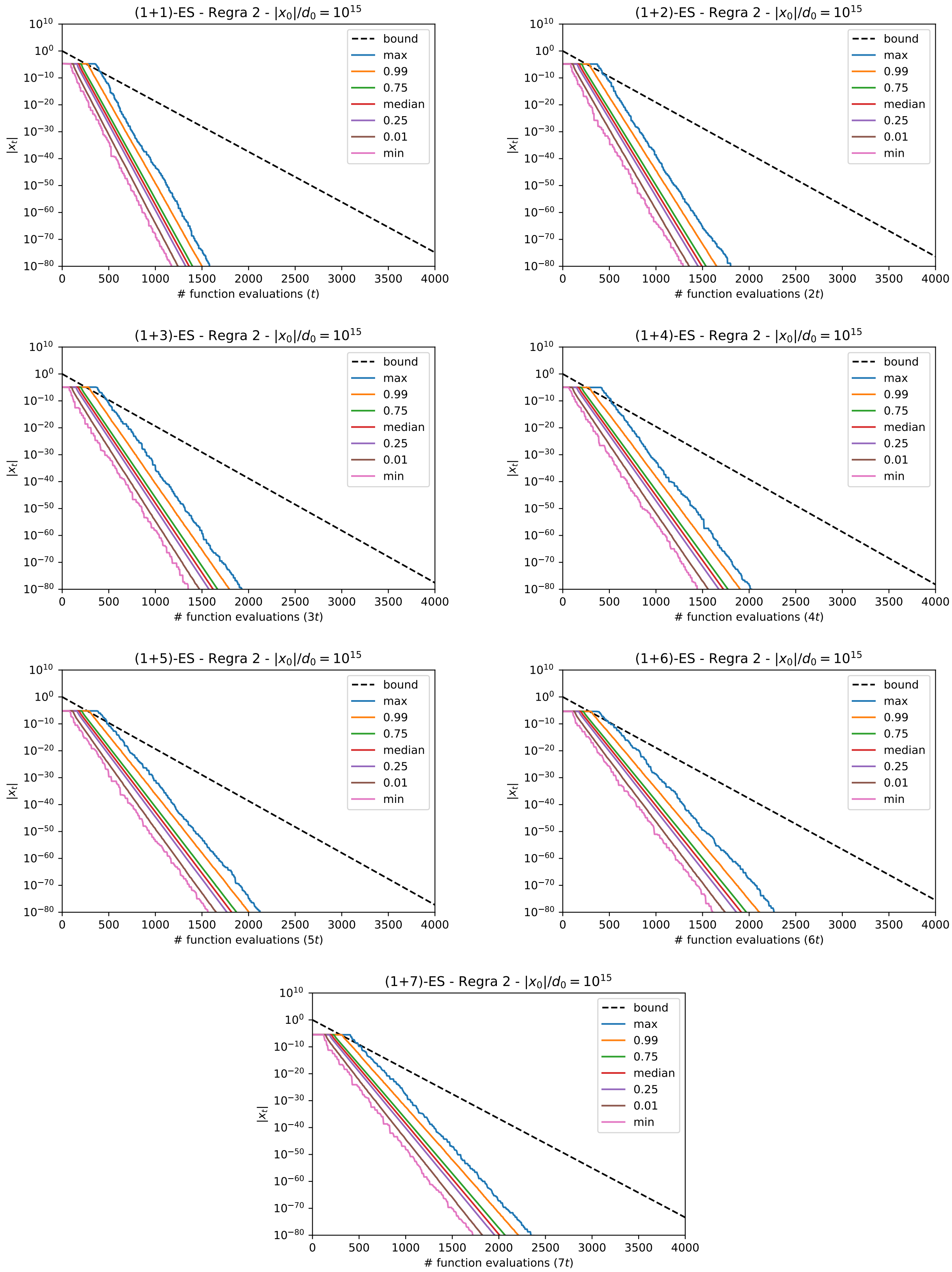

Figura 56 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 2** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

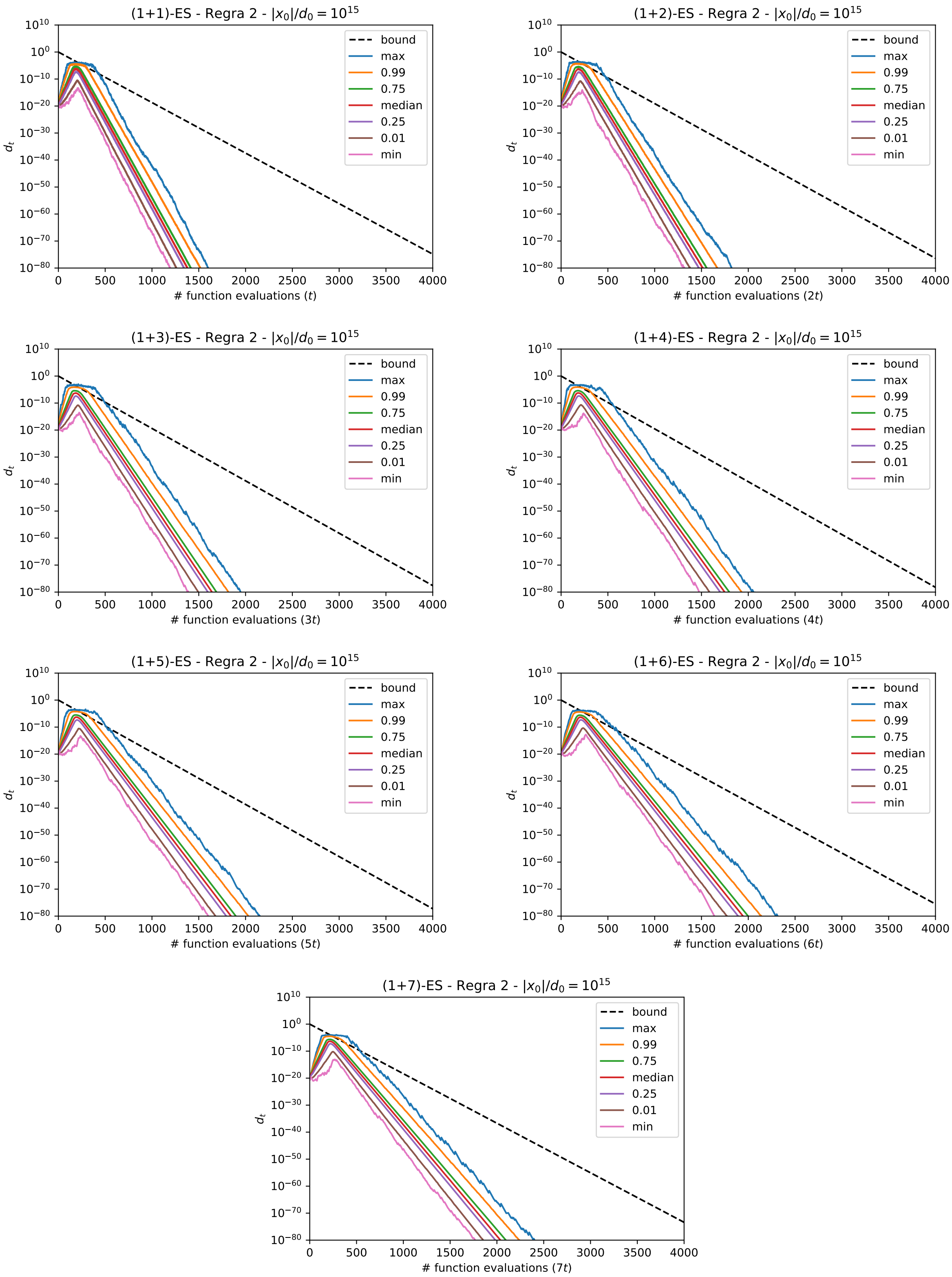

Figura 57 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 2** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

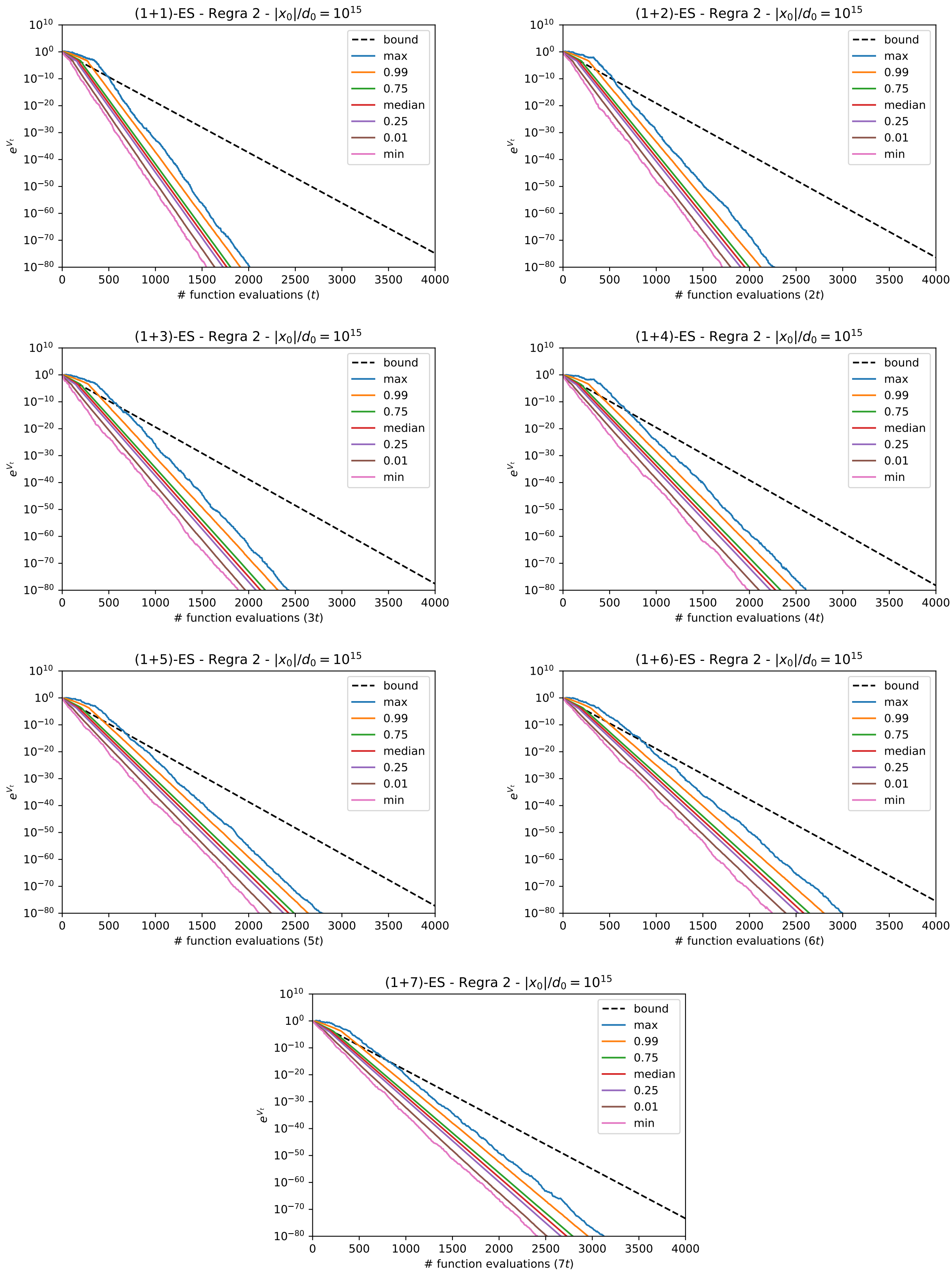

Figura 58 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 2** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

F.2 Controle de Mutação Baseado no Tamanho do Passo

Nas Figuras 59, 60 e 61 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1,\lambda)$-EE (para $\lambda \in \{2, 3, 4, 5, 6\}$) com controle de mutação baseado no tamanho do passo , tendo iniciado em $x_0 = 0$, juntamente com os limites de convergência correspondentes.

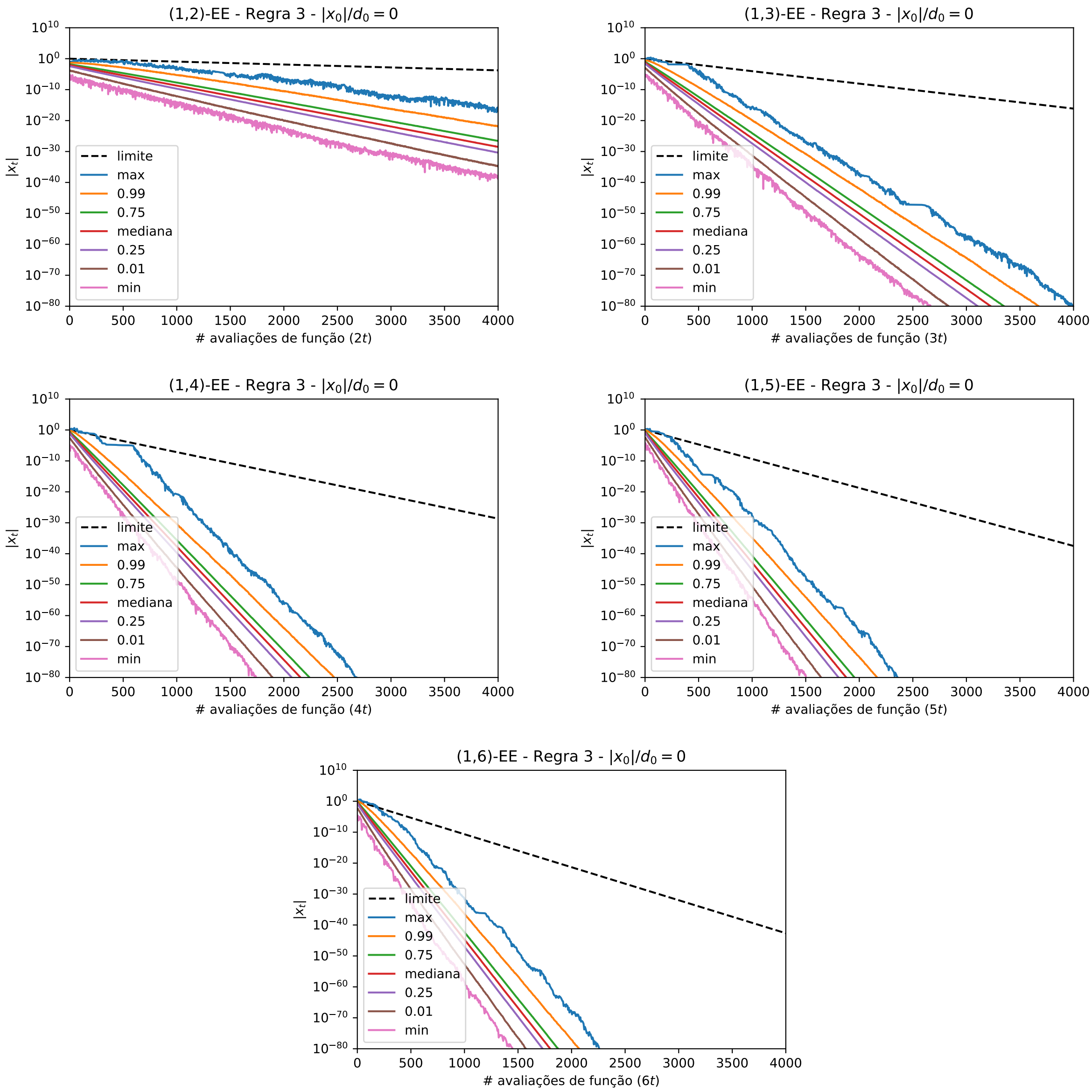

Figura 59 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 3** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

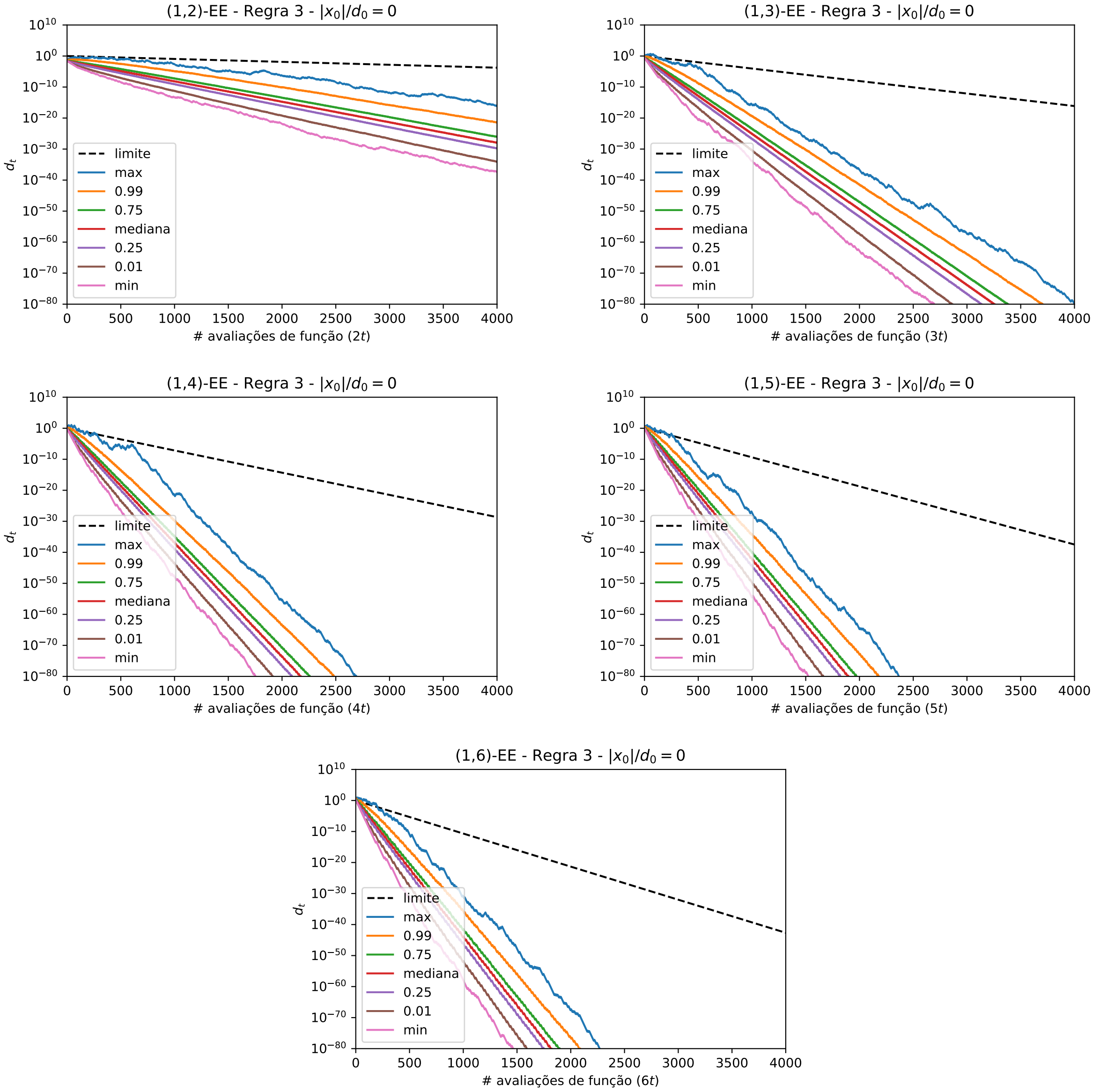

Figura 60 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 3** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

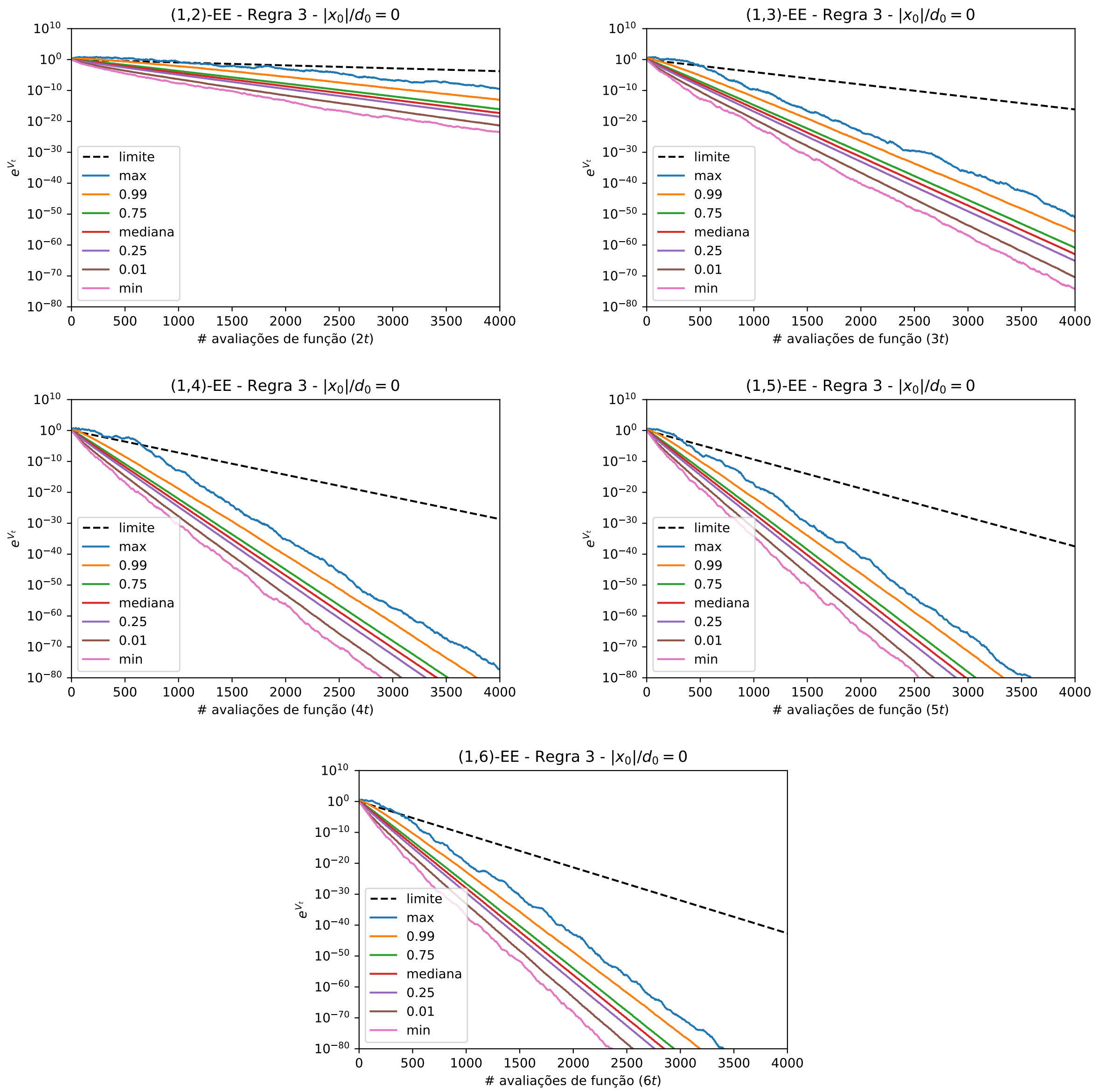

Figura 61 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 3** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Nas Figuras 62, 63 e 64 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1,\lambda)$-EE (para $\lambda \in \{2, 3, 4, 5, 6\}$) com controle de mutação baseado no tamanho do passo , tendo iniciado em x_0, tal que $|x_0|/d_0 = 10^{10}$, juntamente com os limites de convergência correspondentes.

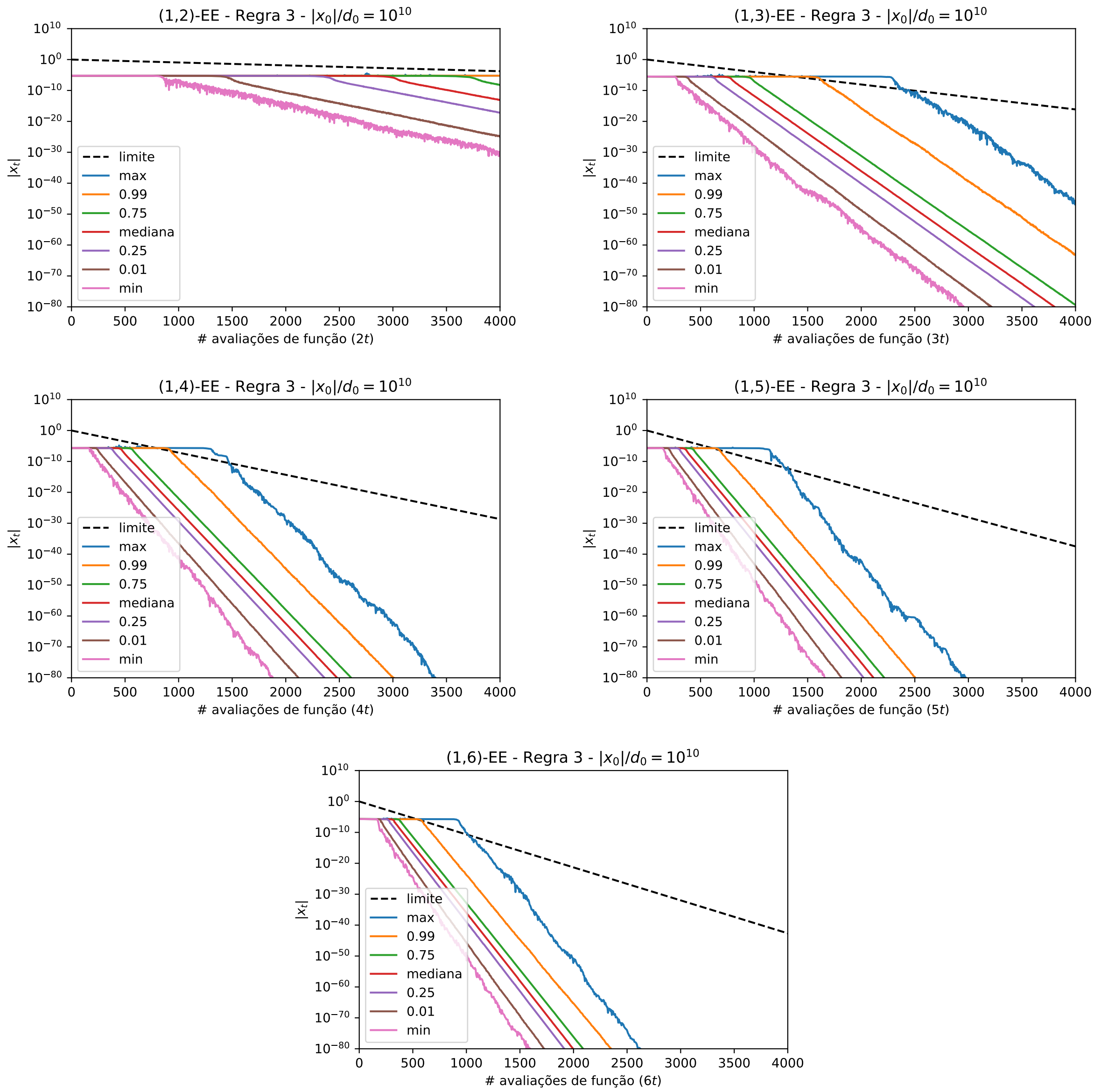

Figura 62 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 3** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{10}$).

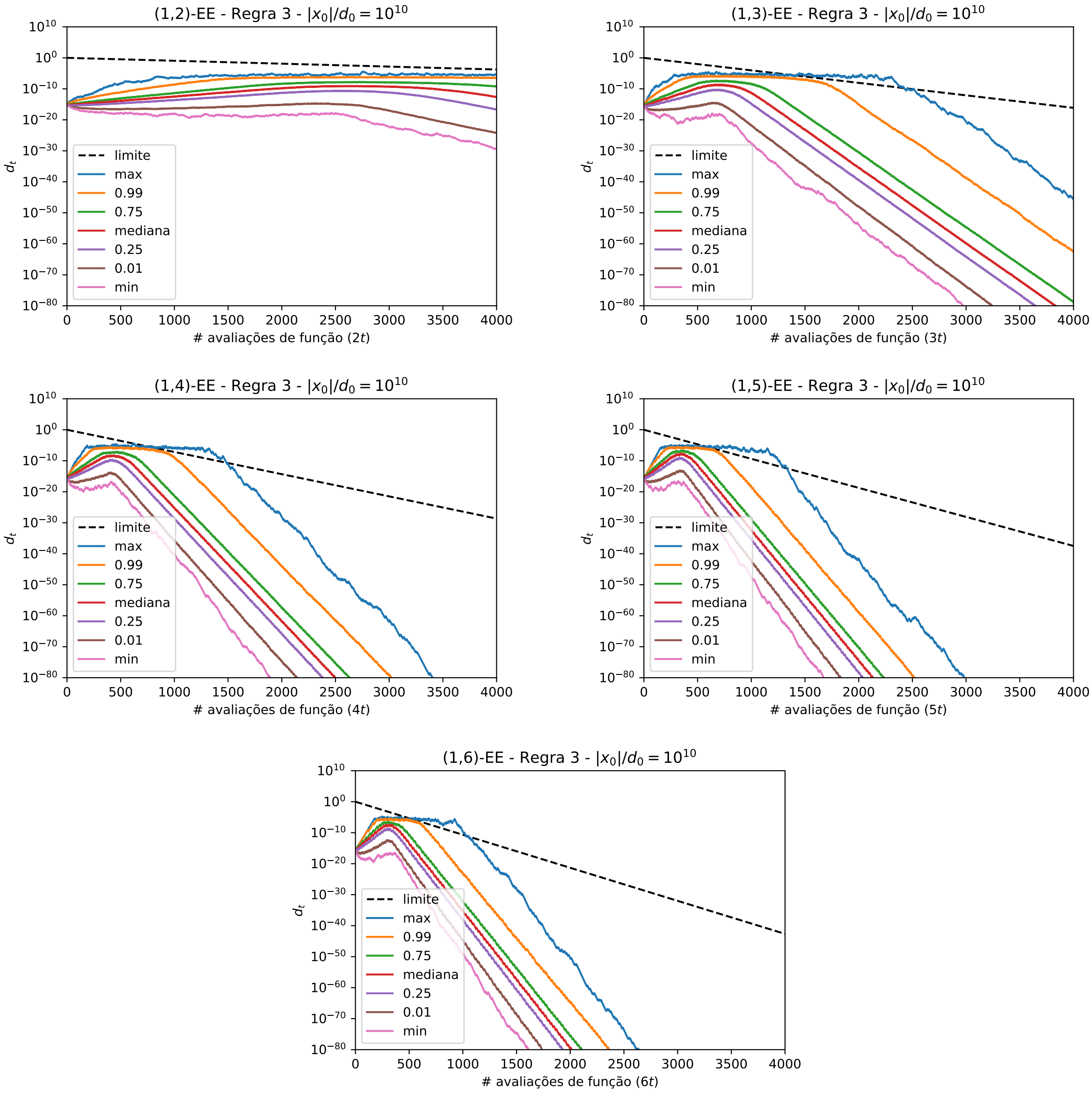

Figura 63 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 3** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{10}$).

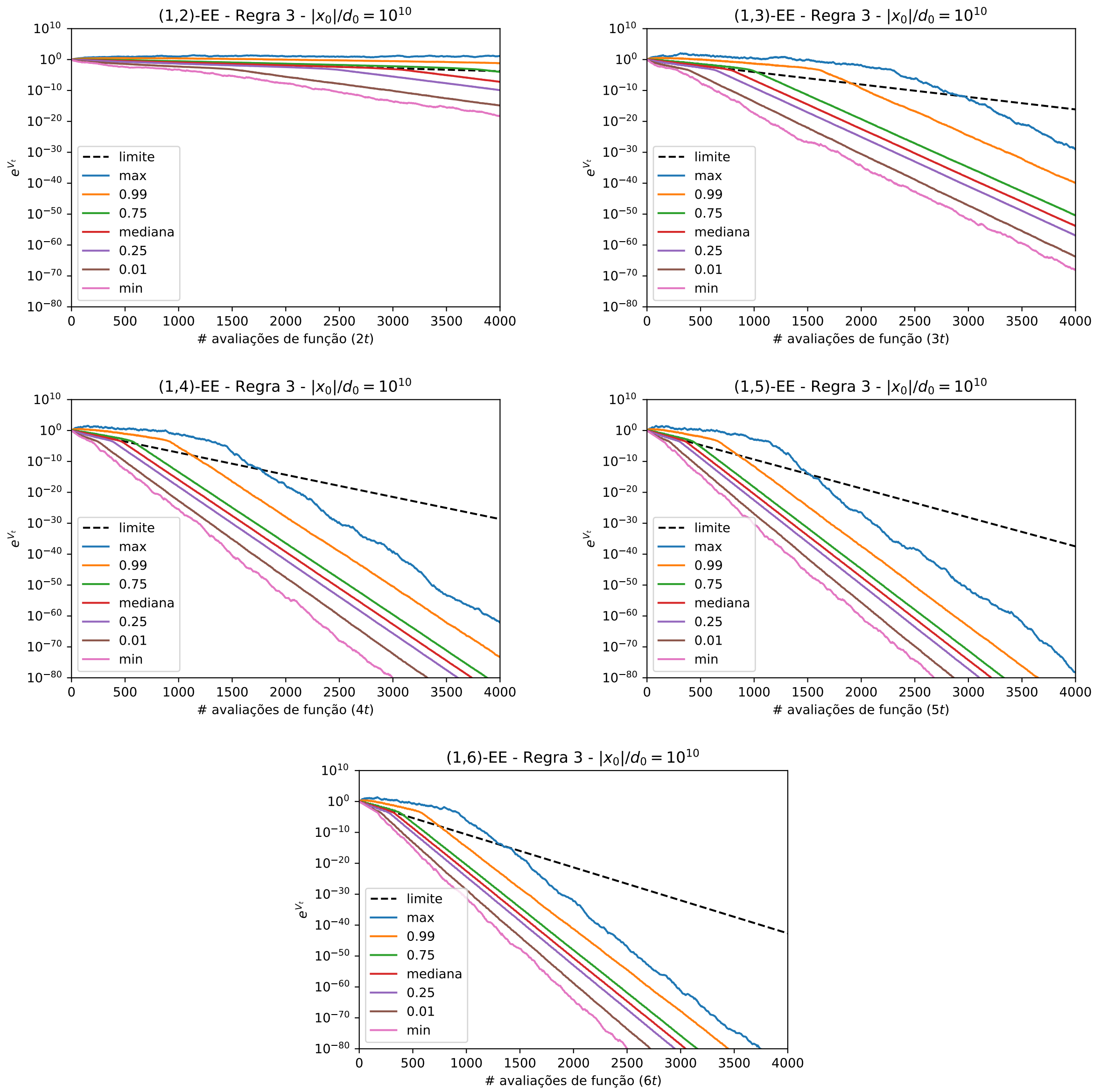

Figura 64 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1,\lambda)$-EE com **Regra 3** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{10}$).

Nas Figuras 65, 66 e 67 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1+\lambda)$-EE (para $\lambda \in \{1,2,3,4,5,6\}$) com controle de mutação baseado no tamanho do passo , tendo iniciado em $x_0 = 0$, juntamente com os limites de convergência correspondentes.

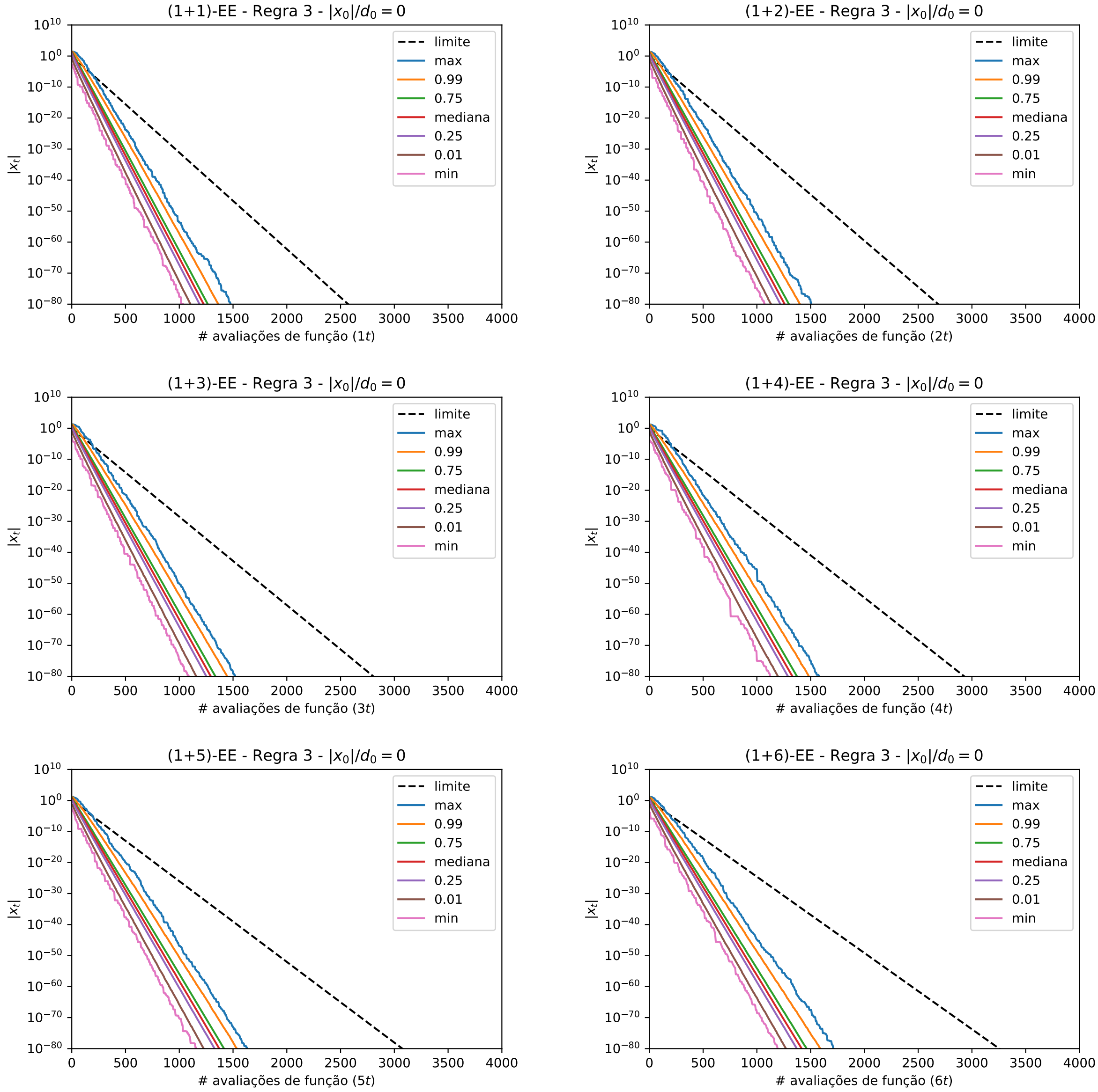

Figura 65 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 3** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

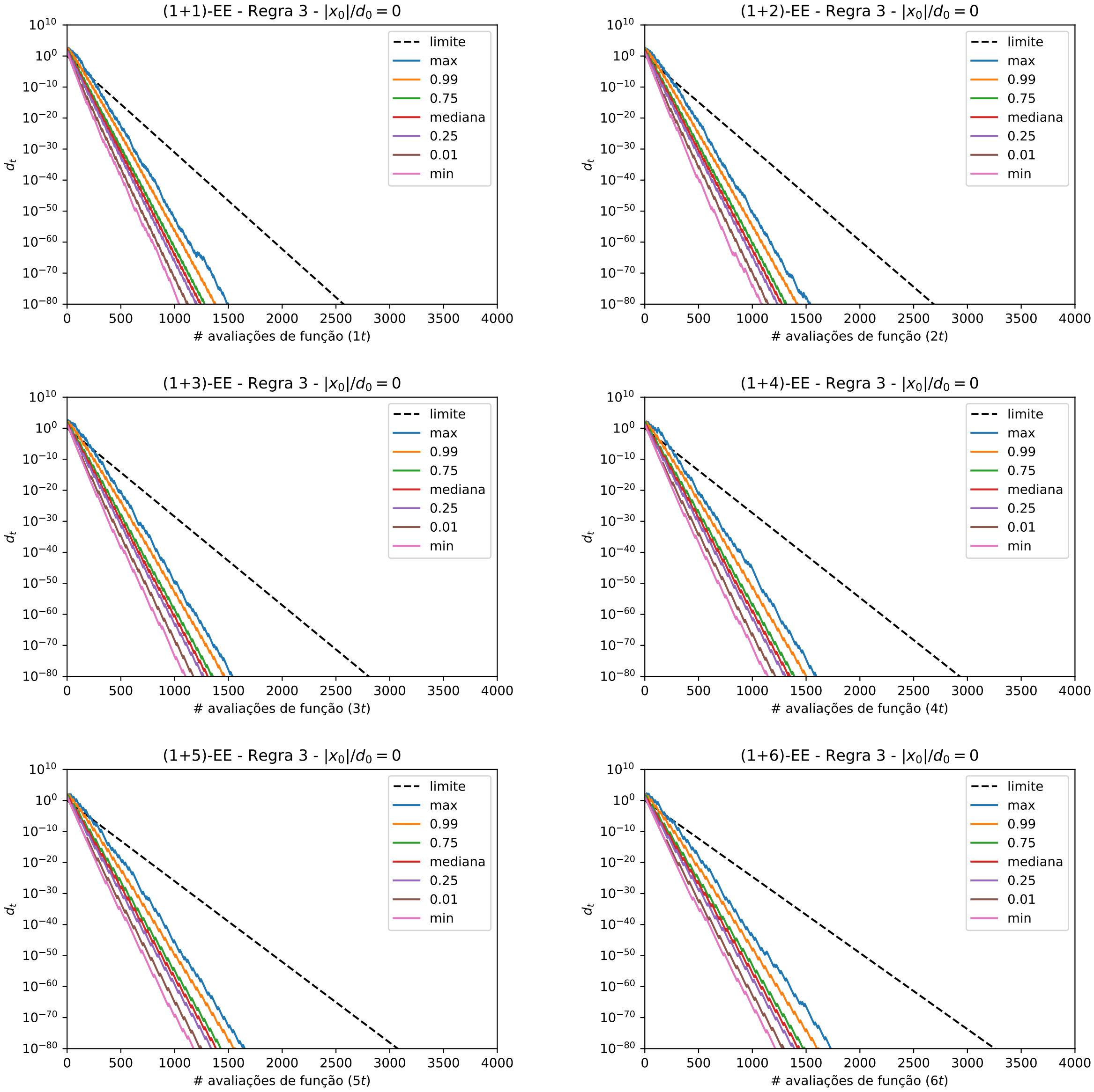

Figura 66 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 3** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

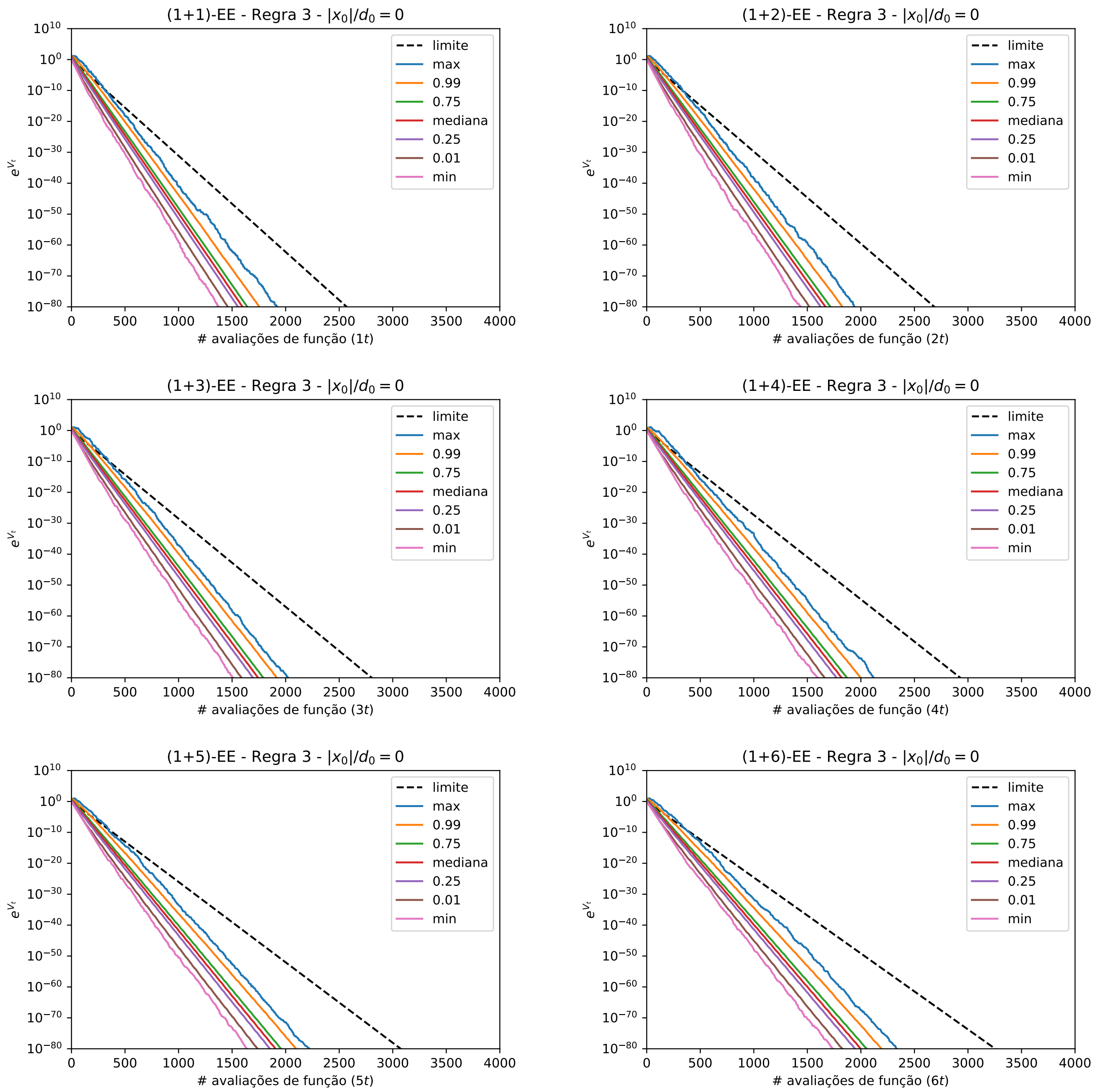

Figura 67 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 3** iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Nas Figuras 68, 69 e 70 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1+\lambda)$-EE (para $\lambda \in \{1,2,3,4,5,6\}$) com controle de mutação baseado no tamanho do passo , tendo iniciado em x_0 tal que $|x_0|/d_0 = 10^{15}$, juntamente com os limites de convergência correspondentes.

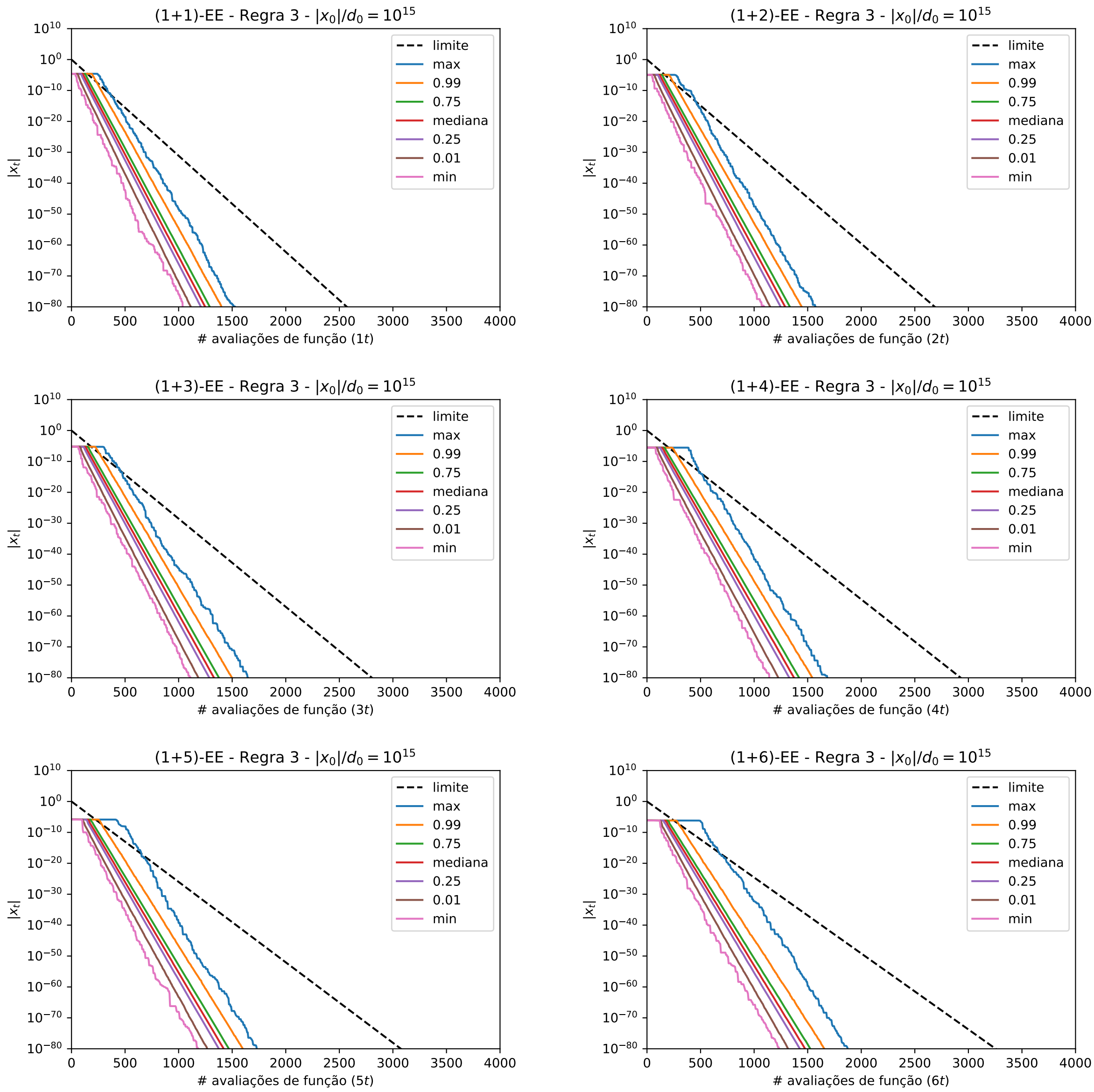

Figura 68 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 3** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

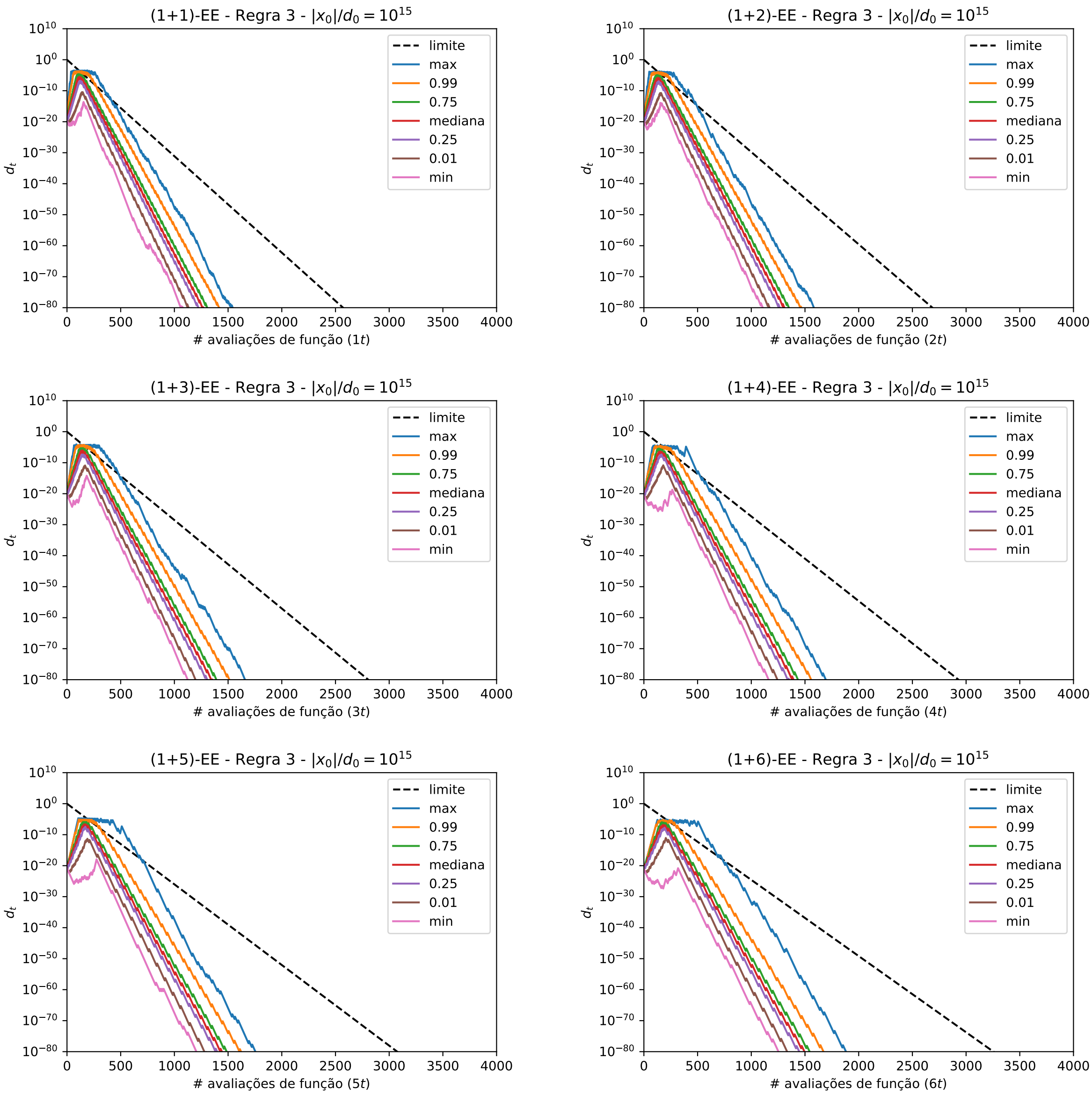

Figura 69 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1+\lambda)$-EE com **Regra 3** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

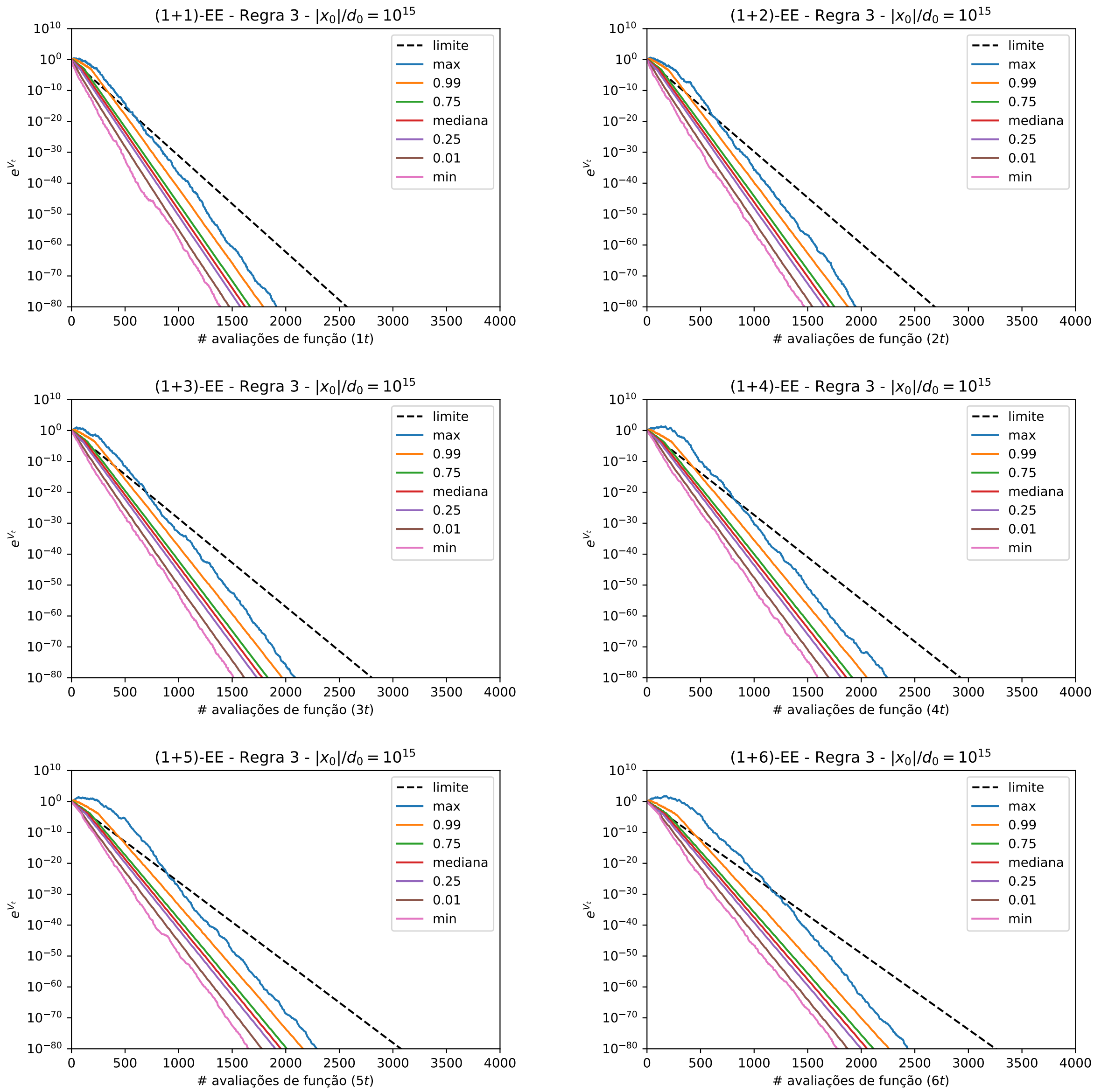

Figura 70 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1 + \lambda)$-EE com **Regra 3** iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

F.3 Controle de Mutação Combinado

Nas Figuras 71, 72 e 73 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1,\lambda)$-EE (para $\lambda \in \{2, 3, 4, 5, 6\}$) com controle de mutação combinado , tendo iniciado em $x_0 = 0$, juntamente com os limites de convergência correspondentes.

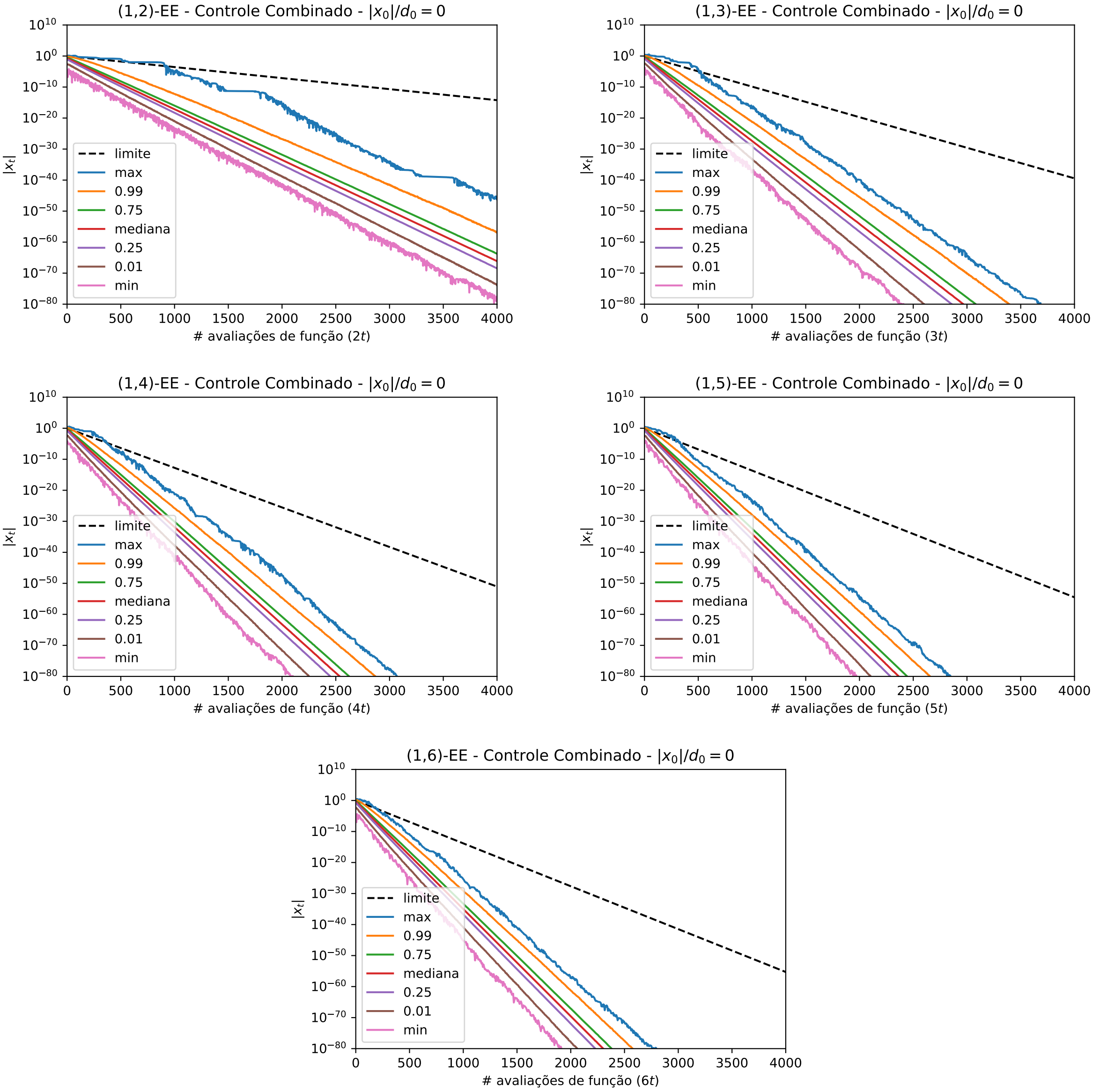

Figura 71 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1{,}\lambda)$-EE com adaptação combinada iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

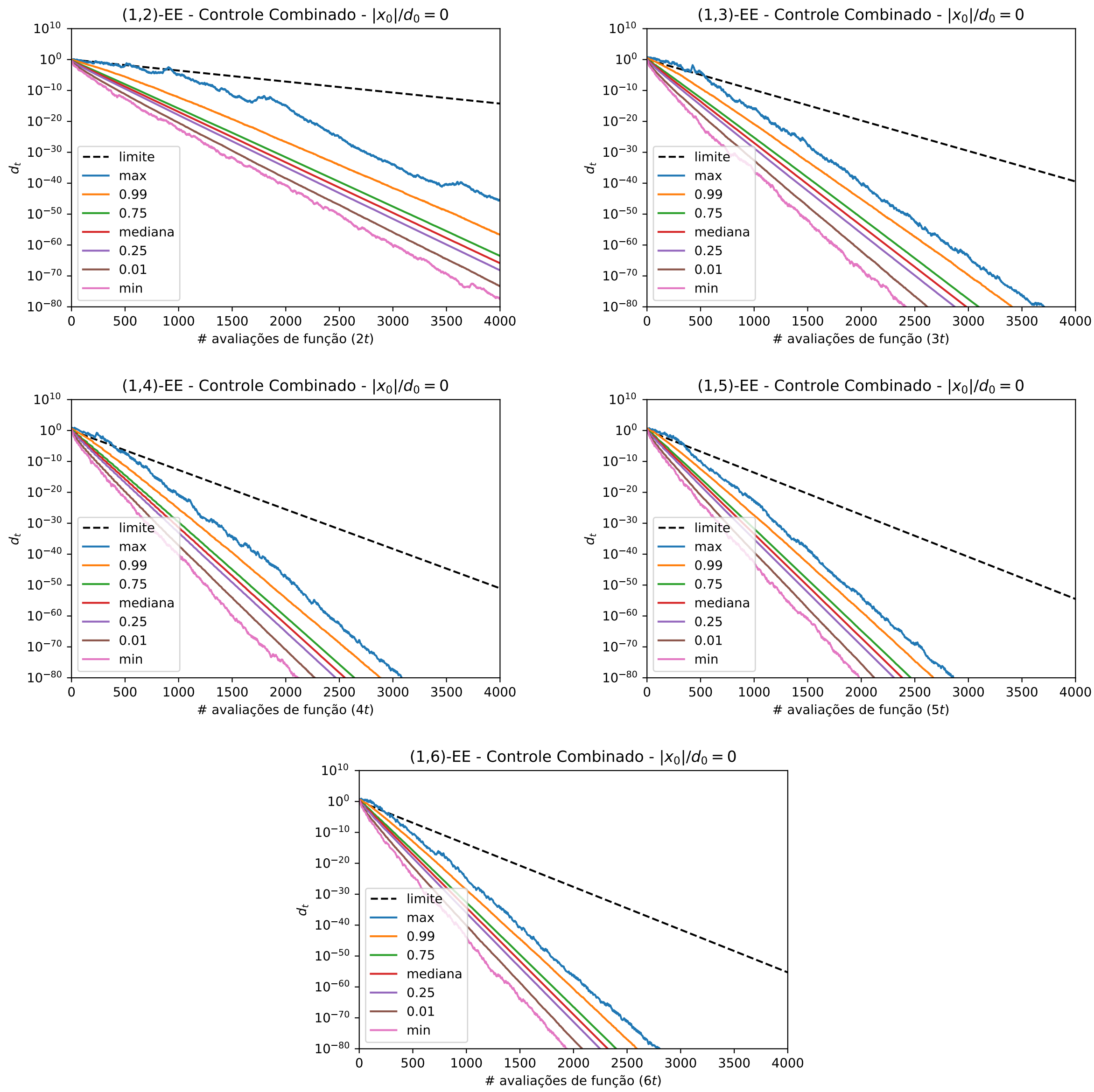

Figura 72 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1{,}\lambda)$-EE com adaptação combinada iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

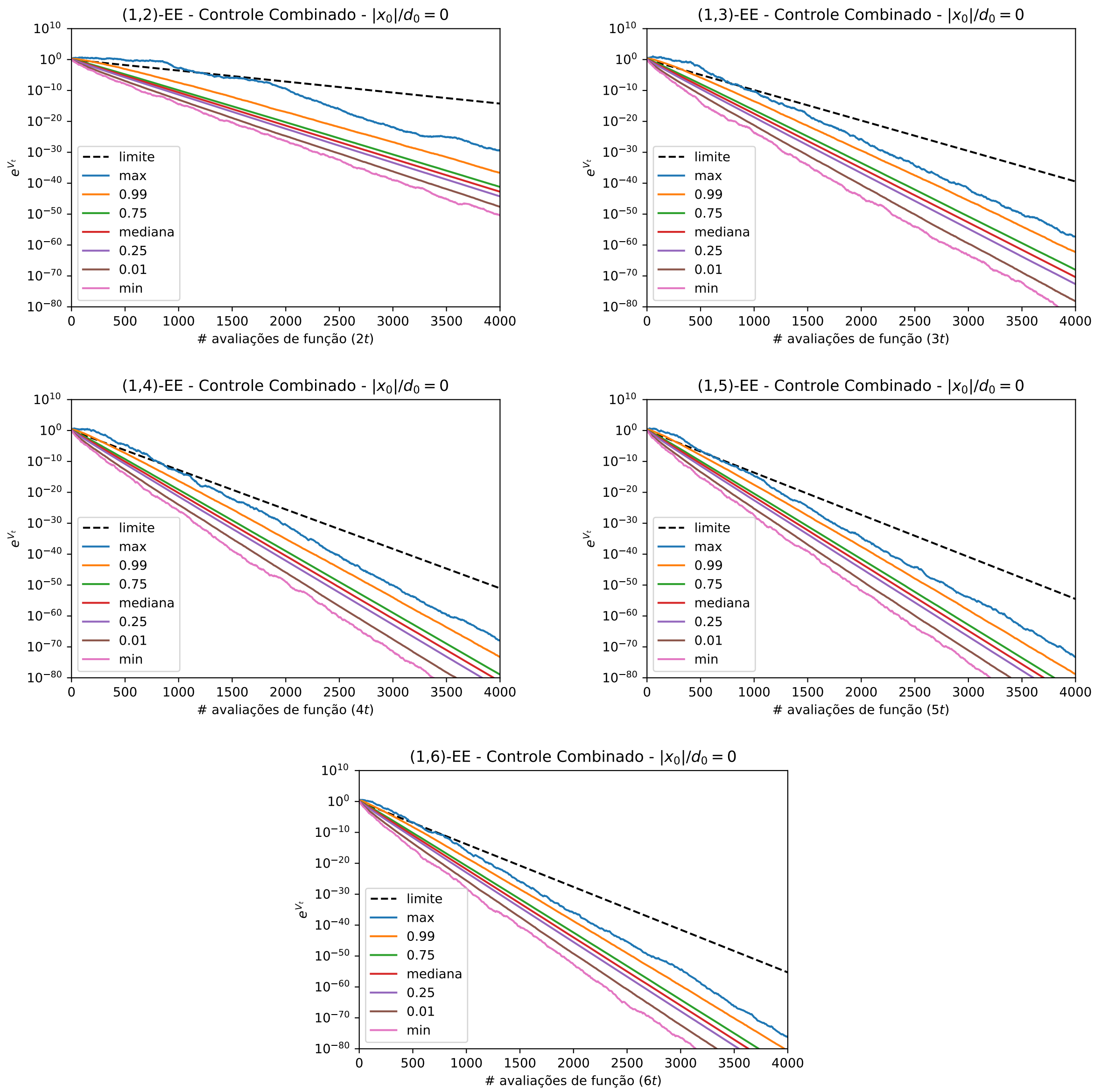

Figura 73 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1{,}\lambda)$-EE com adaptação combinada iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Nas Figuras 74, 75 e 76 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1{,}\lambda)$-EE (para $\lambda \in \{2, 3, 4, 5, 6\}$) com controle de mutação combinado , tendo iniciado em x_0 tal que $|x_0|/d_0 = 10^{15}$, juntamente com os limites de convergência correspondentes.

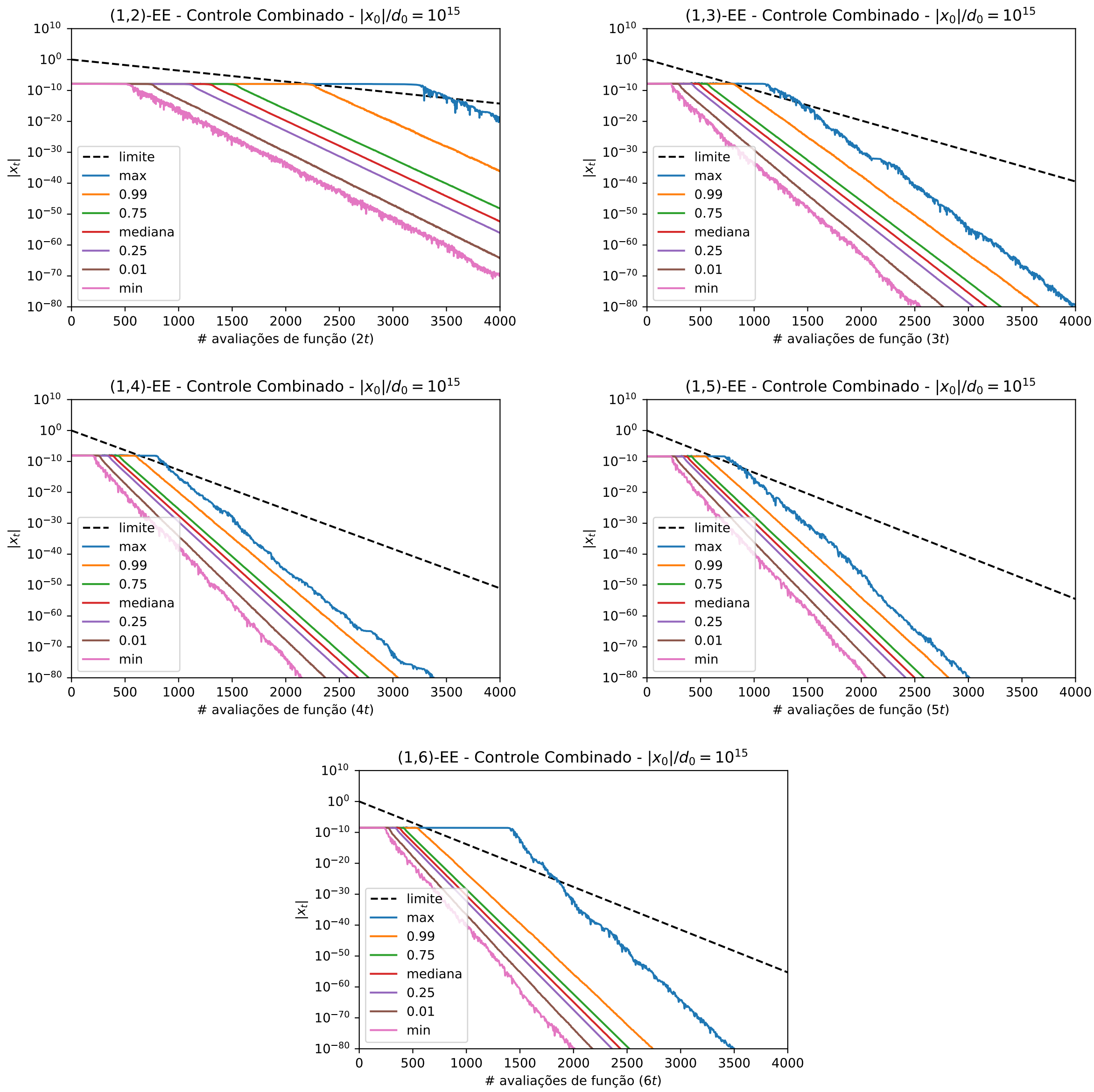

Figura 74 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1,\lambda)$-EE com adaptação combinada iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

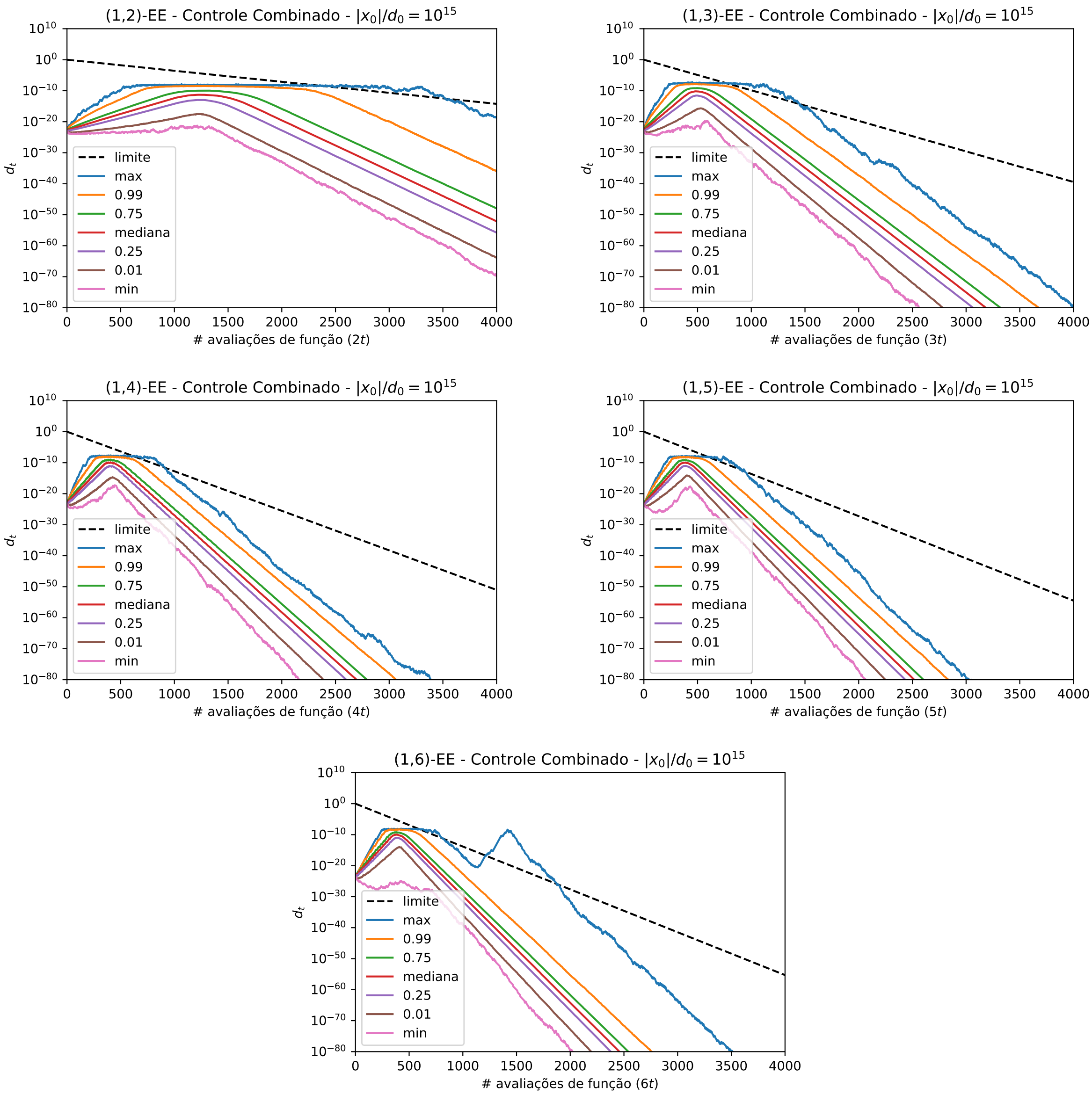

Figura 75 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1,\lambda)$-EE com adaptação combinada iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

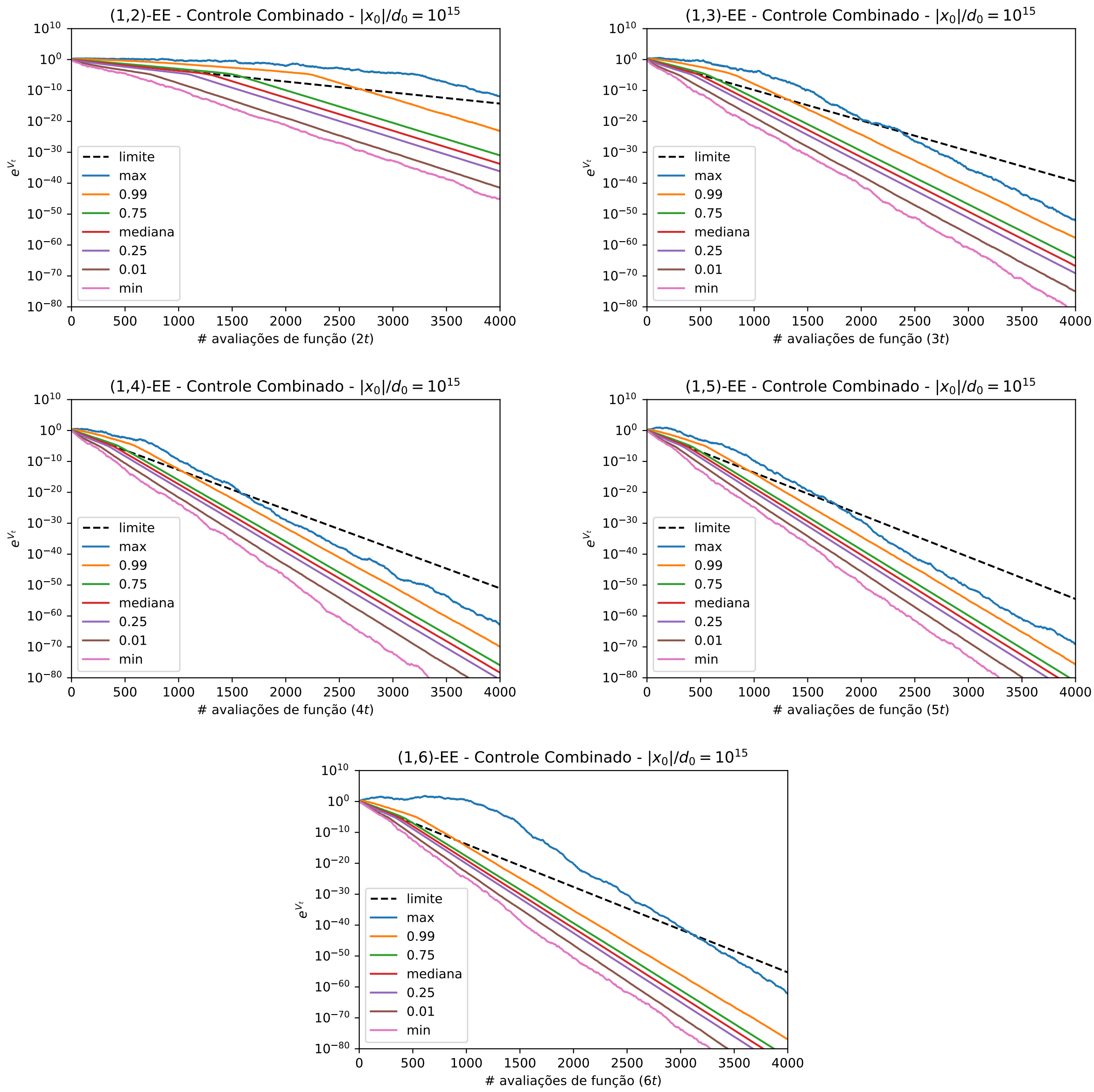

Figura 76 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1{,}\lambda)$-EE com adaptação combinada iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

Nas Figuras 77, 78 e 79 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1+\lambda)$-EE (para $\lambda \in \{1, 2, 3, 4, 5\}$) com controle de mutação combinado , tendo iniciado em $x_0 = 0$, juntamente com os limites de convergência correspondentes.

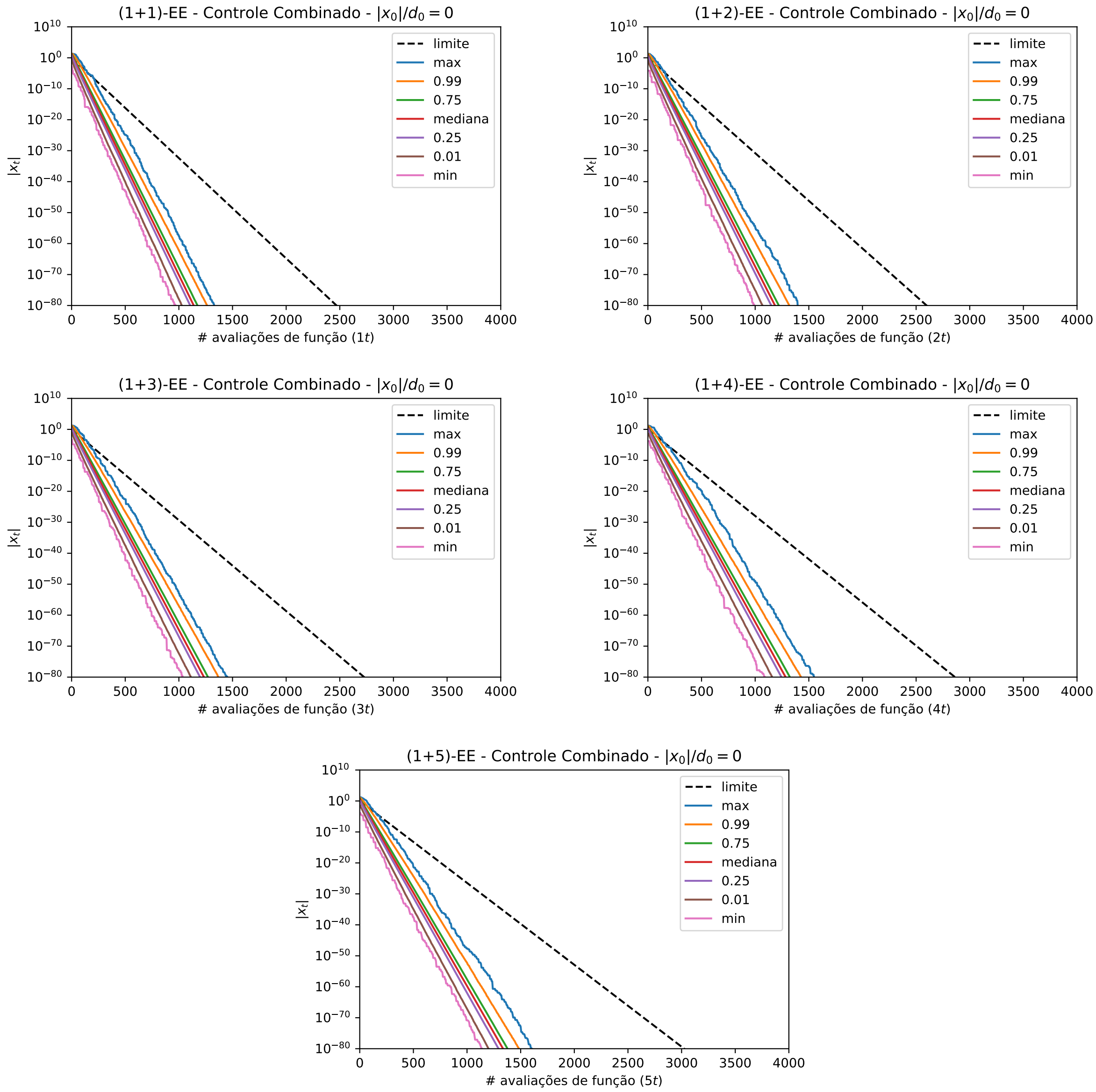

Figura 77 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1+\lambda)$-EE com adaptação combinada iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

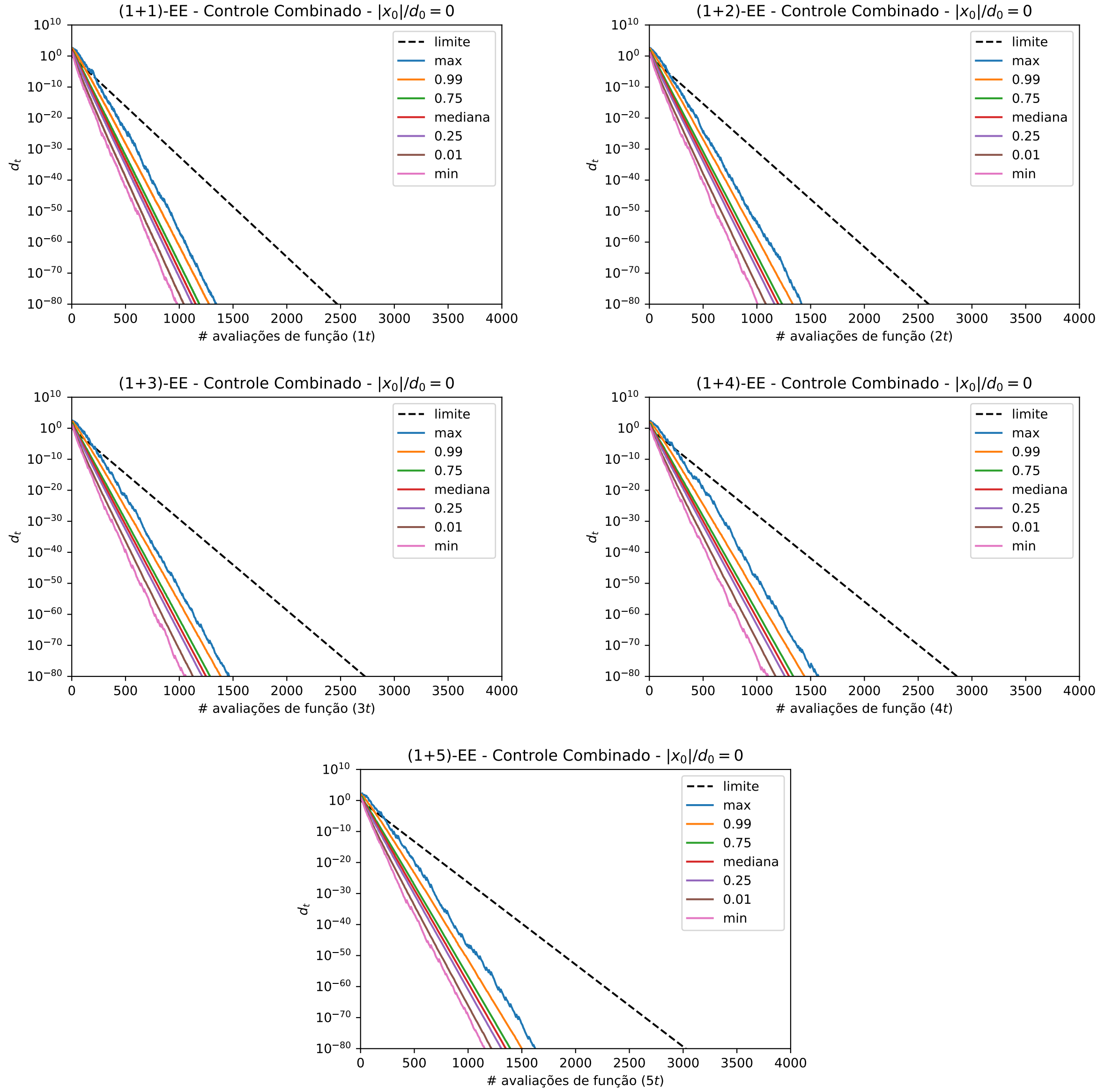

Figura 78 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1+\lambda)$-EE com adaptação combinada iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

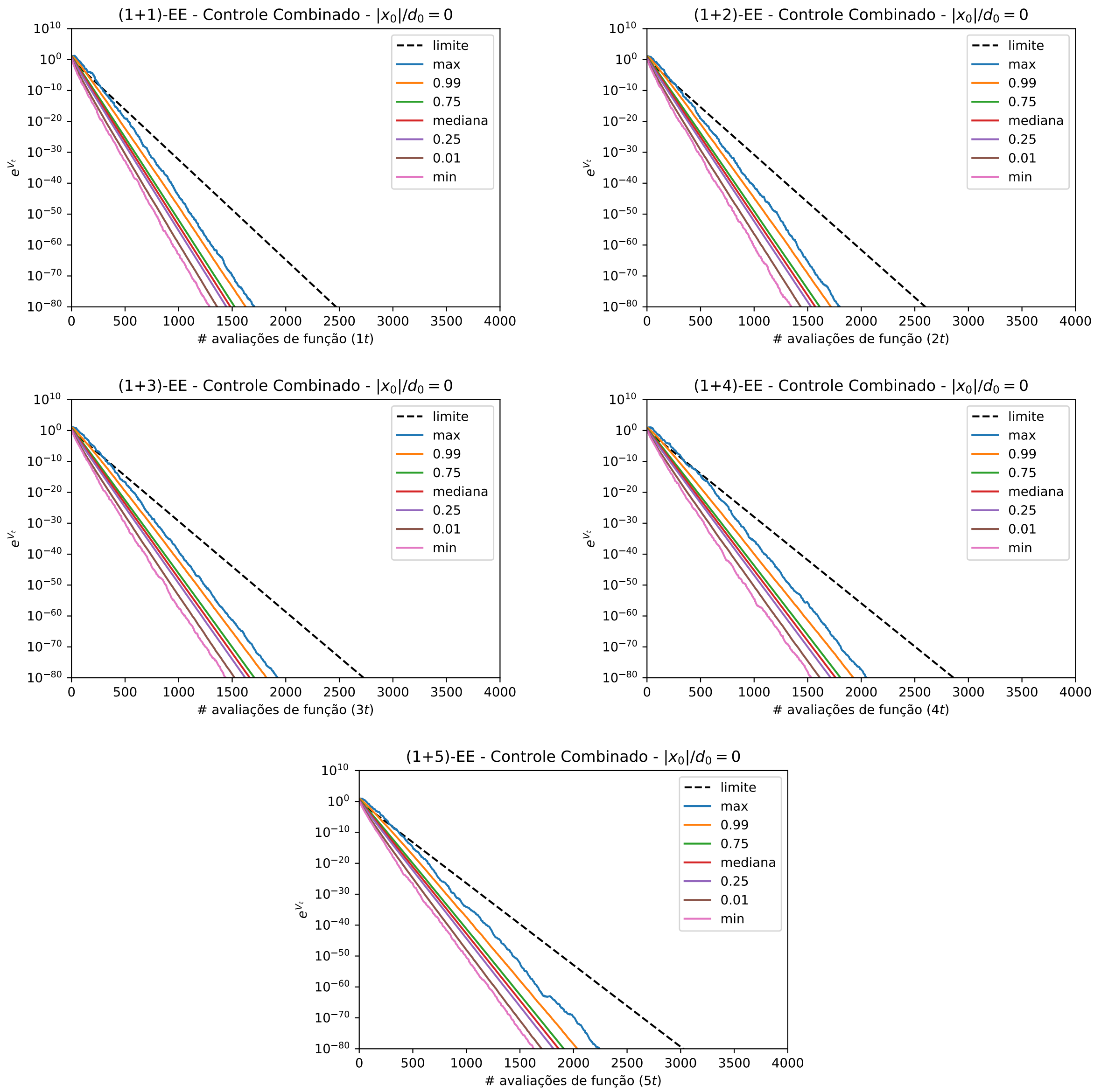

Figura 79 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1+\lambda)$-EE com adaptação combinada iniciadas perto do ótimo ($|x_0|/d_0 = 0$).

Nas Figuras 80, 81 e 82 são exibidos os quantis empíricos das distribuições de $|x_t|, d_t$ e e^{V_t} ao longo das gerações de $(1+\lambda)$-EE (para $\lambda \in \{1,2,3,4,5\}$) com controle de mutação combinado , tendo iniciado em x_0 tal que $|x_0|/d_0 = 10^{15}$, juntamente com os limites de convergência correspondentes.

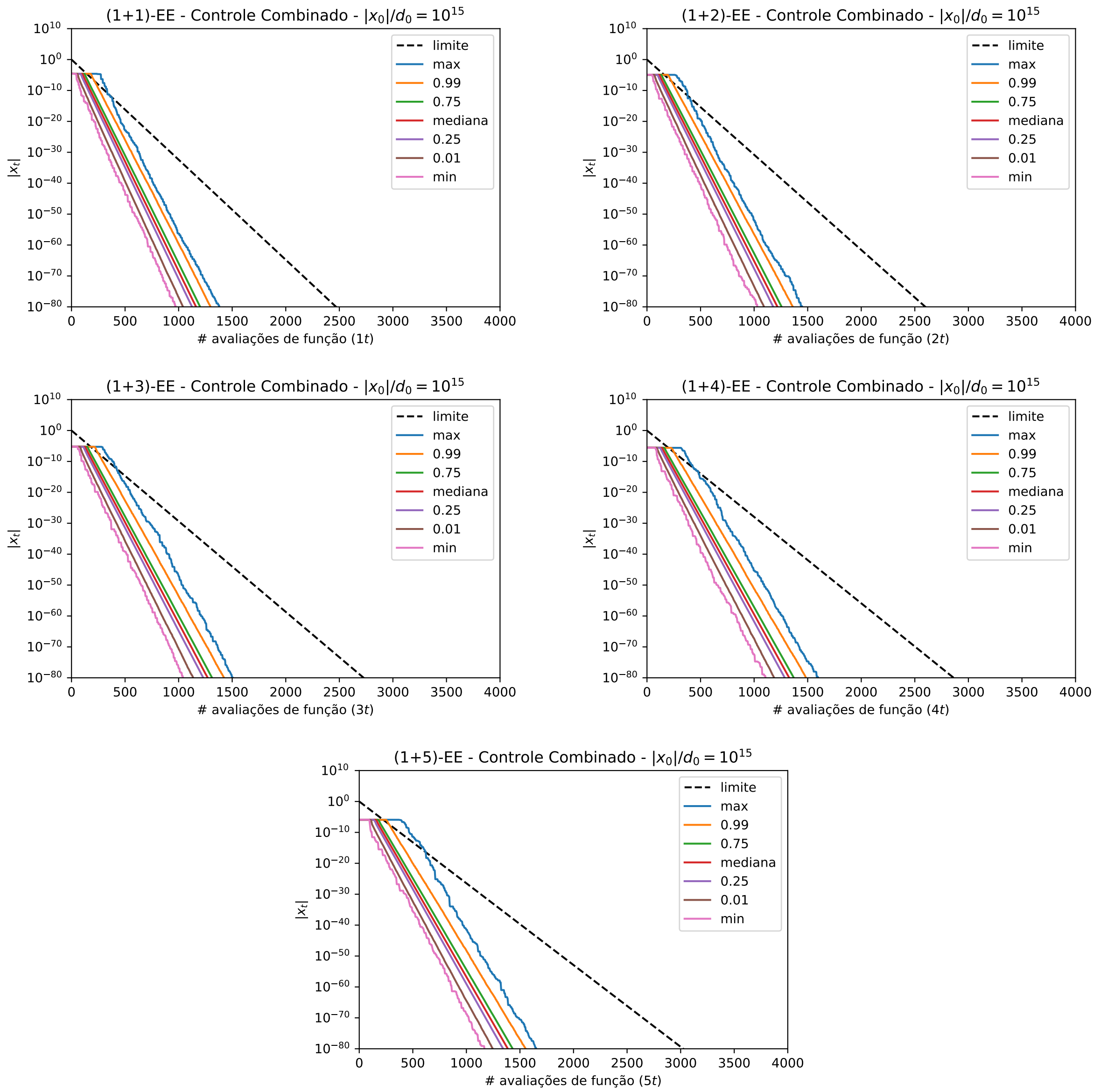

Figura 80 – Quantis empíricos das distribuições de $|x_t|$ estimadas de 10001 execuções de $(1+\lambda)$-EE com adaptação combinada iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

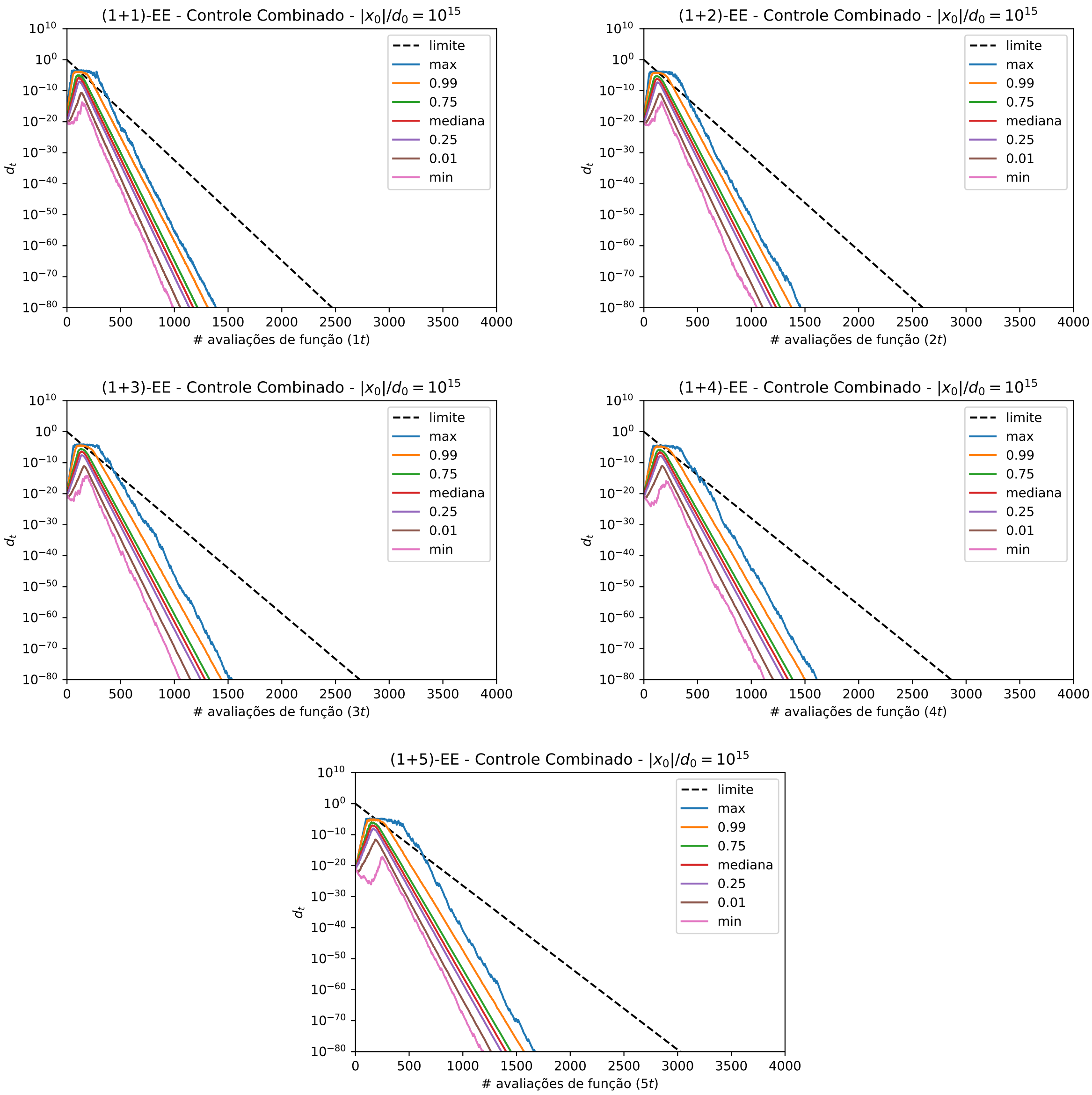

Figura 81 – Quantis empíricos das distribuições de d_t estimadas de 10001 execuções de $(1+\lambda)$-EE com adaptação combinada iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

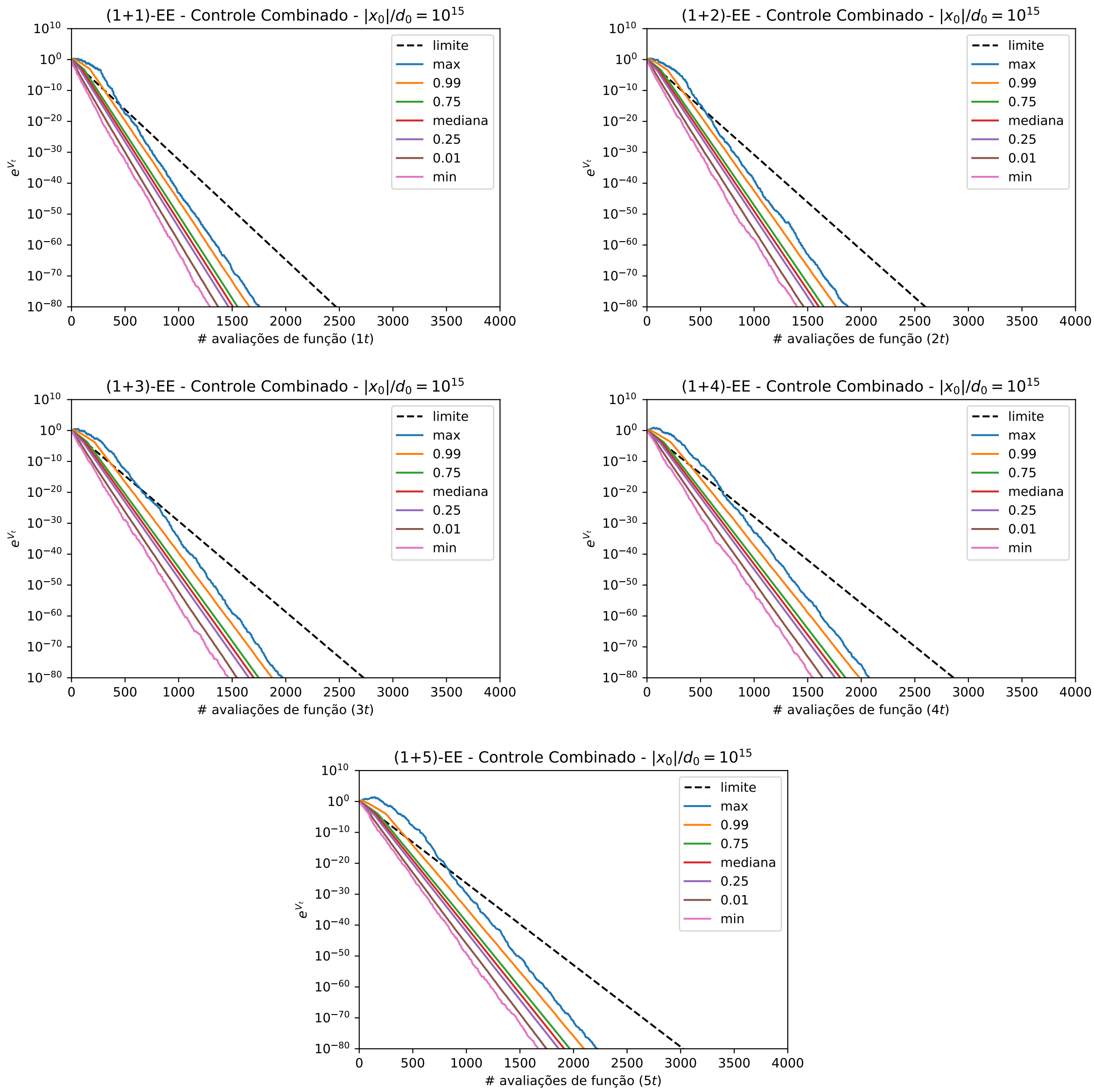

Figura 82 – Quantis empíricos das distribuições de e^{V_t} estimadas de 10001 execuções de $(1+\lambda)$-EE com adaptação combinada iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

F.4 Resultados com parâmetros obtidos sem o uso da desigualdade de Jensen

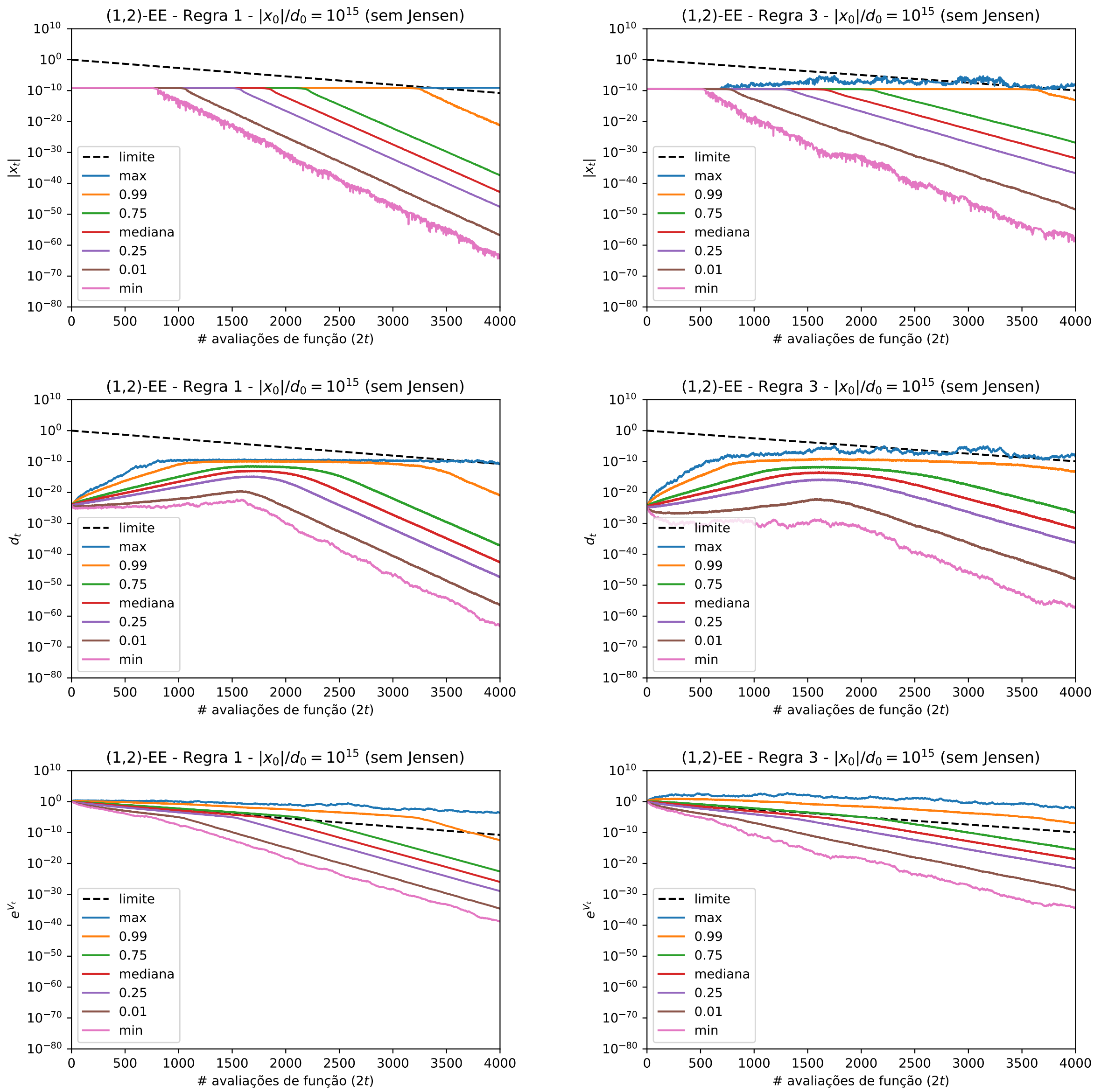

Figura 83 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1{,}2)$-EE com **Regra 1** e **Regra 3** (sem Jensen), iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

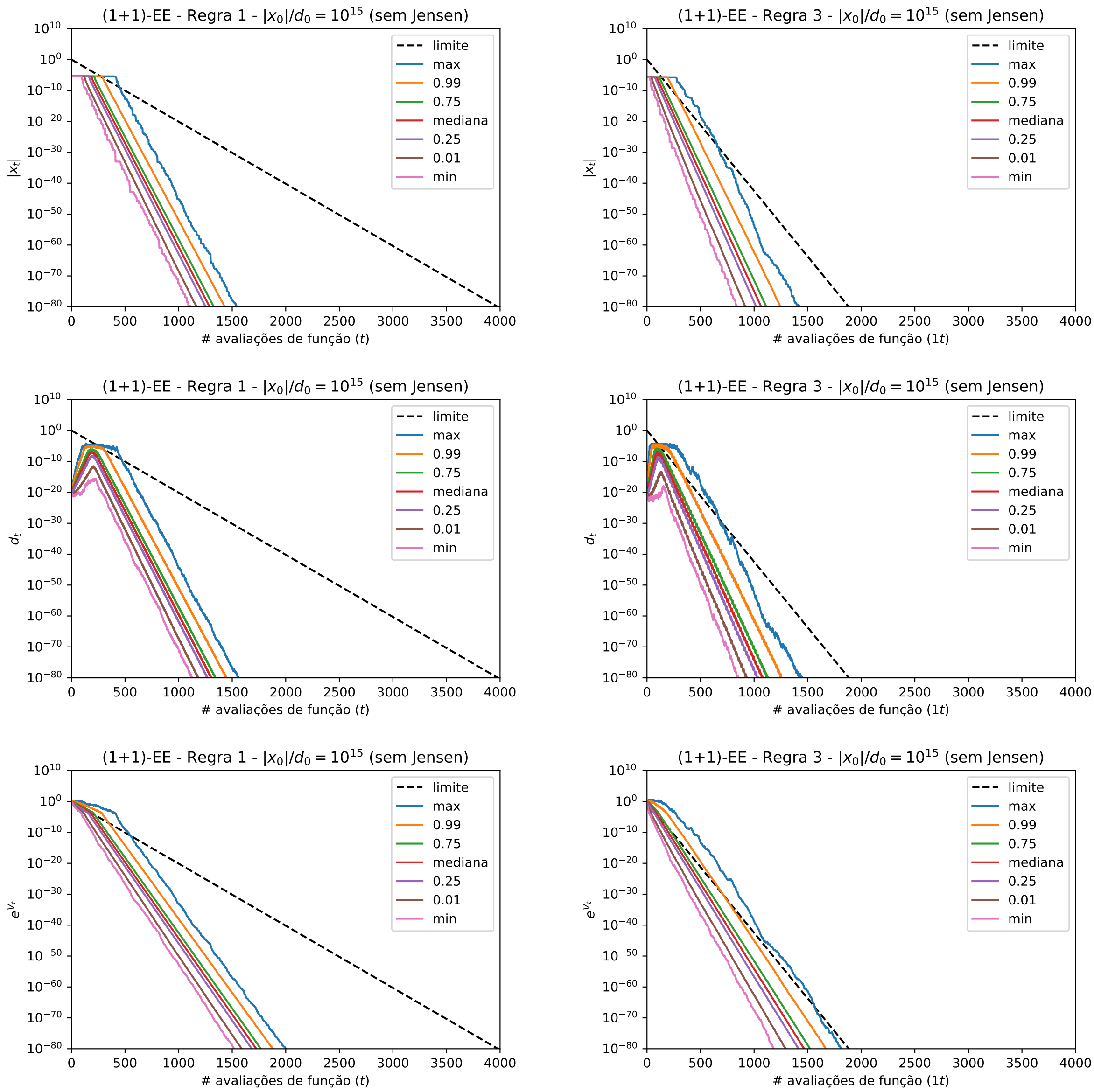

Figura 84 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1+1)$-EE com **Regra 1** e controle combinado (sem Jensen), iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).

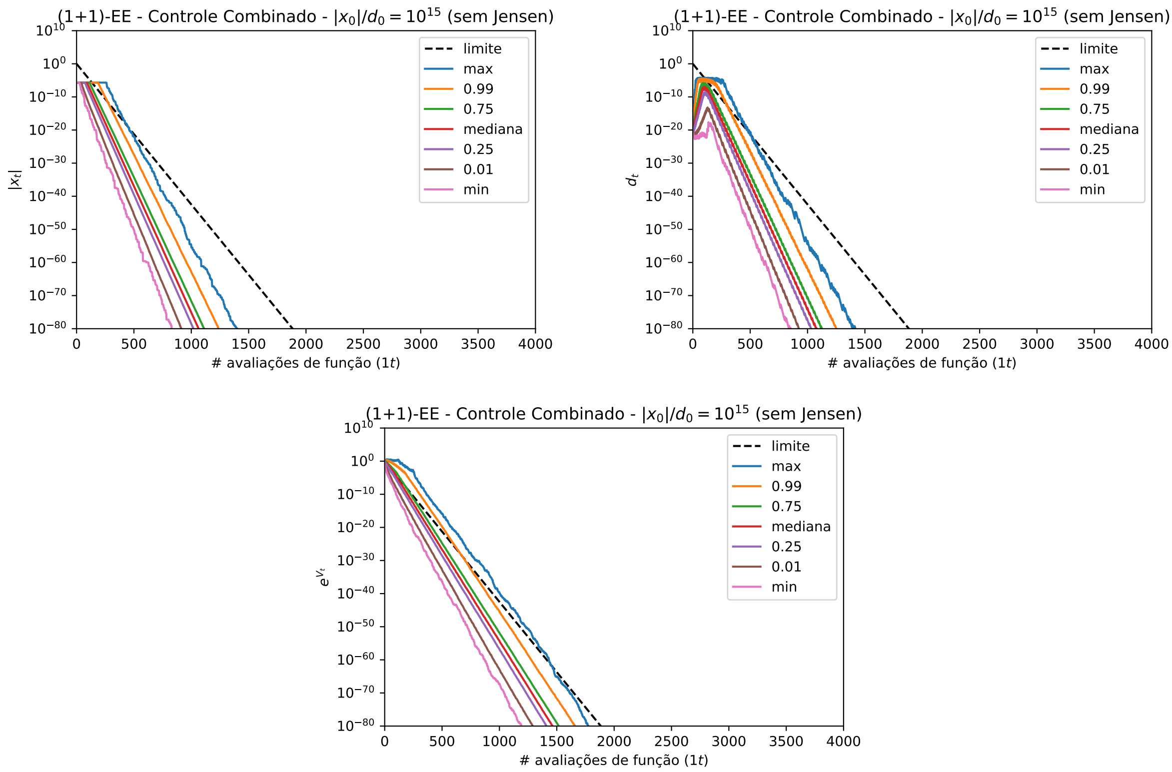

Figura 85 – Quantis empíricos das distribuições de $|x_t|$, d_t e e^{V_t} estimadas de 10001 execuções de $(1+1)$-EE com controle combinado (sem Jensen), iniciadas longe do ótimo ($|x_0|/d_0 = 10^{15}$).