



Paper Type: Original Article

# Estimation of Weibull Distribution Parameters Using a Genetic Algorithm

Hashem Talamkhani<sup>1</sup>, Akram Kohansal<sup>1,\*</sup>, Kimia Samavati<sup>1</sup>, Zahra Barikbin<sup>2</sup>

<sup>1</sup> Department of Statistics, Imam Khomeini International University; hashemtalam1@gmail.com; kohansal@sci.ikiu.ac.ir; kimia.samavati@gmail.com.

<sup>2</sup> Department of Applied Mathematics, Imam Khomeini International University; barikbin@sci.ikiu.ac.ir.

## Citation:

Received: 05 February 2025

Revised: 19 May 2025

Accepted: 10 June 2025

Talamkhani, H., Kohansal, A., Samavati, K., & Barikbin, Z. (2025).

Estimation of Weibull distribution parameters using a genetic algorithm. *Journal of Quality Engineering and Management*, 15(4), 415-432.

## Abstract

**Purpose:** This paper aims to estimate the parameters of the Weibull distribution using a genetic algorithm and compare its performance with traditional estimation methods.

**Methodology:** A simulation study was conducted under different sample sizes and censoring levels. The genetic algorithm was applied to maximize the likelihood function.

**Findings:** The results show that the genetic algorithm provides more accurate and stable parameter estimates compared to the maximum likelihood method, especially in the presence of censored data.

**Originality/Value:** This study presents a novel application of genetic algorithms in reliability analysis, demonstrating their effectiveness in parameter estimation for censored datasets.

**Keywords:** Weibull distribution, Censored data, Genetic algorithm.



## برآورد پارامترهای توزیع وایبول آمیخته با استفاده از الگوریتم ژنتیک

هاشم تلمخانی<sup>۱</sup>، اکرم کهن سال<sup>۱\*</sup>، کیمیا سماواتی<sup>۱</sup>، زهرا باریک‌بین<sup>۲</sup>

<sup>۱</sup>گروه آمار، دانشگاه بین‌المللی امام خمینی (ره)، قزوین، ایران.

<sup>۲</sup>گروه ریاضی کاربردی، دانشگاه بین‌المللی امام خمینی (ره)، قزوین، ایران.

### چکیده

هدف: این مقاله با هدف برآورد پارامترهای توزیع وایبول آمیخته با استفاده از الگوریتم ژنتیک و مقایسه عملکرد آن با روش‌های سنتی برآورد رایج شده است.

روش‌شناسی پژوهش: یک مطالعه شبیه‌سازی تحت شرایط اندازه نمونه‌ها و سطوح سانسور متفاوت انجام شد. الگوریتم ژنتیک برای بیشینه‌سازی تابع درستنمایی به کار گرفته شد.

یافته‌ها: نتایج نشان می‌دهد که الگوریتم ژنتیک نسبت به روش ماکسیمم درستنمایی، به‌ویژه در حضور داده‌های سانسور شده، برآوردهای پارامتری دقیق‌تر و پایدارتری ارائه می‌دهد.

اصالت/ارزش‌افزوده علمی: این مطالعه کاربردی نوین از الگوریتم‌های ژنتیک در تحلیل قابلیت اعتماد ارائه می‌دهد و اثربخشی آن‌ها در برآورد پارامترها برای داده‌های سانسور شده را نشان می‌دهد.

کلیدواژه‌ها: توزیع وایبول، الگوریتم ژنتیک، داده‌های سانسور شده.

### ۱- مقدمه

تحلیل داده‌های طول عمر و زمان شکست یکی از حوزه‌های مهم در آمار کاربردی و مدل‌سازی پدیده‌های تصادفی است که به بررسی زمان وقوع رویدادهایی با ماهیت زمانی می‌پردازد. این نوع داده‌ها در بسیاری از مطالعات مهندسی، پزشکی و علوم داده مشاهده می‌شوند و هدف اصلی تحلیل آن‌ها، توصیف رفتار زمانی سیستم‌ها و پیش‌بینی زمان وقوع رویدادهای آینده است. با این حال، در بسیاری از کاربردهای عملی، داده‌های مربوط به زمان وقوع رویداد به‌طور کامل مشاهده نمی‌شوند و به‌صورت داده‌های سانسور شده در دسترس قرار می‌گیرند؛ وضعیتی که معمولاً به دلیل محدودیت زمانی مطالعه، پایان زود هنگام فرآیند مشاهده یا خروج برخی نمونه‌ها از مطالعه رخ می‌دهد [1]. وجود داده‌های سانسور شده باعث می‌شود که استفاده از روش‌های آماری کلاسیک بدون در نظر گرفتن این ویژگی، منجر به برآوردهای اریب و نتایج نادرست شود.

در ادبیات آماری، مدل‌سازی داده‌های طول عمر با استفاده از توزیع‌های پارامتریک متداول است و در این میان، توزیع وایبول به‌عنوان یکی از پرکاربردترین توزیع‌ها شناخته می‌شود. انعطاف‌پذیری این توزیع در مدل‌سازی نرخ‌های شکست افزایشی، کاهشی و ثابت سبب شده است که در طیف وسیعی از مسایل مربوط به تحلیل زمان شکست مورد استفاده قرار گیرد [2]. با وجود این مزیت، مدل وایبول همگن مبتنی بر فرض یکنواختی رفتار داده‌هاست؛ فرضی که در بسیاری از مجموعه داده‌های واقعی برقرار نیست. داده‌ها اغلب از ترکیب چند زیرجمعیت با ویژگی‌های متفاوت تشکیل شده‌اند و نادیده گرفتن این ناهمگنی می‌تواند منجر به برازش ضعیف مدل و تفسیر نادرست نتایج شود.

برای غلبه بر این محدودیت، مدل‌های آمیخته به‌عنوان ابزارهای انعطاف‌پذیر برای مدل‌سازی ناهمگنی داده‌ها معرفی شده‌اند. در این مدل‌ها، توزیع کلی داده‌ها به‌صورت ترکیبی از چند توزیع پایه در نظر گرفته می‌شود. توزیع وایبول آمیخته متناهی یکی از پرکاربردترین این مدل‌ها در تحلیل داده‌های طول عمر است و مطالعات مختلف نشان داده‌اند که این مدل می‌تواند ساختار واقعی داده‌ها را بهتر از مدل‌های همگن توصیف کند [3]. در سال‌های اخیر، پژوهش‌های متعددی به بررسی روش‌های مختلف برآورد پارامترهای توزیع‌های وایبول آمیخته پرداخته‌اند و کارایی آن‌ها را در شرایط گوناگون مورد ارزیابی قرار داده‌اند [4]. برآورد پارامترهای مدل‌های آمیخته، به‌ویژه در حضور داده‌های سانسور شده، یک مساله غیرخطی و محاسباتی پیچیده است. روش‌های کلاسیک مانند برآورد ماکسیمم درستمایی و الگوریتم *EM* به‌طور گسترده در این زمینه به‌کار رفته‌اند، اما این روش‌ها ممکن است به مقادیر اولیه حساس باشند یا در بهینه‌های محلی متوقف شوند، به‌خصوص زمانی که تعداد پارامترها افزایش می‌یابد [5]. این چالش‌ها باعث شده است که پژوهشگران در سال‌های اخیر به استفاده از روش‌های بهینه‌سازی ابتکاری و فراابتکاری برای برآورد پارامترهای مدل‌های آمیخته روی آورند.

در میان روش‌های فراابتکاری، الگوریتم ژنتیک به‌عنوان یکی از الگوریتم‌های تکاملی شناخته شده، توانایی بالایی در جست‌وجوی فضای پارامتری پیچیده دارد. این الگوریتم با الهام از فرآیند تکامل طبیعی، قادر است بدون نیاز به مشتق‌گیری و با کاهش وابستگی به مقادیر اولیه، راه‌حل‌های مناسبی برای مسایل بهینه‌سازی غیرخطی ارائه دهد. مطالعات اخیر نشان داده‌اند که الگوریتم ژنتیک می‌تواند در برآورد پارامترهای توزیع‌های آمیخته، به‌ویژه در شرایط وجود سانسور و ناهمگنی داده‌ها، عملکرد رقابتی و پایدارتری نسبت به روش‌های کلاسیک داشته باشد [6]. بر این اساس، در این مقاله هدف برآورد پارامترهای توزیع وایبول آمیخته متناهی در حضور داده‌های سانسور شده با استفاده از الگوریتم ژنتیک است. برای ارزیابی کارایی روش پیشنهادی، یک مطالعه شبیه‌سازی انجام شده و نتایج حاصل با روش ماکسیمم درستمایی مقایسه می‌شود تا عملکرد این رویکرد در چارچوب برآورد پارامترهای مدل‌های آمیخته به‌صورت کمی بررسی گردد. همچنین، برای نشان دادن قابلیت کاربرد روش پیشنهادی در مسایل واقعی، یک مثال کاربردی مبتنی بر داده‌های واقعی قابلیت اعتماد با سانسور راست مورد بررسی قرار گرفته است.

## ۲- معرفی خانواده توزیع وایبول آمیخته

در حال حاضر علاقه بسیاری در میان پژوهشگران برای ارائه طبقه‌بندی‌های جدید به‌منظور توصیف بهتر داده‌ها و افزایش انعطاف‌پذیری مدل‌های آماری به‌وجود آمده است. یکی از رویکردهای موثر در این زمینه، استفاده از خانواده توزیع‌های آمیخته است. در این چارچوب، توزیع وایبول آمیخته به‌عنوان یکی از پرکاربردترین خانواده‌های توزیعی شناخته می‌شود. این خانواده با آمیختن چندین جز از توزیع وایبول با پارامترهای متفاوت، امکان مدل‌سازی ساختارهای پیچیده‌تر در داده‌ها را فراهم می‌آورد. در این راستا، هر جز از توزیع آمیخته می‌تواند معرف یک زیرگروه از جامعه آماری با ویژگی‌های خاص خود باشد و مدل نهایی ناهمگنی موجود در داده‌ها را به‌خوبی منعکس کند [7].

به‌عنوان مثال، در تحلیل داده‌های صنعتی ممکن است بخشی از محصولات دارای خرابی زود هنگام ناشی از نقص تولید باشند، درحالی‌که بخش دیگر طول عمر بیشتری دارند. در چنین شرایطی، یک مدل تک‌توزیعی قادر به نمایش این رفتار ترکیبی نیست، اما مدل آمیخته وایبول می‌تواند این ناهمگنی را به‌طور مناسب مدل‌سازی کند. روش‌های متعددی برای برآورد پارامترهای این خانواده توزیعی ارائه شده است که از جمله می‌توان به روش ماکسیمم درستمایی و الگوریتم *EM* اشاره کرد. یان و همکاران [7] با معرفی یک رویکرد نوآورانه برای برآورد پارامترهای توزیع وایبول آمیخته، به چالش همگرایی به ماکسیمم‌های محلی پرداخته و نشان دادند که ترکیب برآوردگرهای چندک با الگوریتم‌های عددی می‌تواند دقت برآورد را به‌طور قابل‌توجهی افزایش دهد.

در ادامه، جوکیل-روکیتا و پیانک [8] چند برآوردگر جدید برای پارامترها و چندک‌های توزیع وایبول پیشنهاد دادند که نتایج شبیه‌سازی نشان داد این برآوردگرها، به‌ویژه در نمونه‌های با حجم کوچک، عملکرد بهتری در برآورد چندک‌های انتهایی دارند. همچنین کاراکوکا و همکاران [9] با انجام یک مطالعه مقایسه‌ای جامع، روش‌های مختلف برآورد پارامتر برای توزیع وایبول آمیخته را از طریق شبیه‌سازی و داده‌های واقعی بررسی کرده و معیارهای مناسبی برای انتخاب روش برآورد متناسب با ویژگی‌های داده‌ها ارائه دادند.

از سوی دیگر، هلند [10] با تمرکز بر مدل‌سازی داده‌های طول عمر سیستم‌هایی با حالت‌های شکست متعدد نشان داد که استفاده از مدل آمیخته متناهی توزیع‌های وایبول می‌تواند دقت برازش توابع چگالی احتمال و نرخ خطر را نسبت به روش‌های مرسوم افزایش دهد. براینده این مطالعات بیانگر آن است که خانواده توزیع‌های وایبول آمیخته، به دلیل توانایی بالا در مدل‌سازی ناهمگنی و الگوهای پیچیده، گزینه‌ای مناسب برای تحلیل آماری در حوزه‌هایی نظیر تحلیل بقا و مهندسی پزشکی محسوب می‌شود.

لم ۱- برآورد ماکسیمم درست‌نمایی پارامترهای توزیع وایبول آمیخته

اگر  $X$  دارای توزیع خانواده وایبول باشد، تابع چگالی آن به صورت

$$f(x) = \alpha h(x) e^{-\alpha H(x)}, \quad (1)$$

تعریف می‌شود که در آن بسته به نوع توزیع پایه‌ای که انتخاب می‌کنیم، تابع چگالی  $h(x)$  و تابع توزیع تجمعی  $H(x)$  تغییر می‌کنند. تابع توزیع تجمعی  $F(x)$  نیز به صورت

$$F(x) = 1 - e^{-\alpha H(x)}, \quad (2)$$

است.

آنگاه برآوردگرهای ماکسیمم درست‌نمایی پارامترهای مدل با بیشینه‌سازی تابع لگ درست‌نمایی حاصل از داده‌های کامل و سانسور شده به دست می‌آیند. معادلات لازم برای این برآورد از صفر قرار دادن مشتق‌های جزئی لگ درست‌نمایی نسبت به پارامترهای مدل حاصل می‌شوند. به دلیل غیرخطی بودن ساختار این معادلات در مدل‌های وایبول آمیخته، حل تحلیلی بسته برای آن‌ها عموماً وجود ندارد و برآورد پارامترها با استفاده از روش‌های عددی یا الگوریتم‌های بهینه‌سازی انجام می‌شود.

توزیع نمایی یکی از توزیع‌های پرکاربرد در مدل‌سازی داده‌های طول عمر است. توابع پایه این توزیع به صورت

$$h(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad x \geq 0, \quad H(x) = 1 - e^{-\frac{x}{\theta}}, \quad x \geq 0,$$

تعریف می‌شوند که با جایگذاری در روابط اصلی، توابع خانواده وایبول آمیخته برای توزیع نمایی

$$f(x) = \alpha \frac{1}{\theta} e^{-\frac{x}{\theta}} e^{-\alpha \left(1 - e^{-\frac{x}{\theta}}\right)}, \quad x \geq 0,$$

$$F(x) = 1 - e^{-\alpha \left(1 - e^{-\frac{x}{\theta}}\right)}, \quad x \geq 0,$$

$$S(x) = e^{-\alpha \left(1 - e^{-\frac{x}{\theta}}\right)}, \quad x \geq 0,$$

$$h(x) = \alpha \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad x \geq 0,$$

$$H(x) = 1 - e^{-\frac{x}{\theta}}, \quad x \geq 0,$$

به دست می‌آیند. آمیخته دو توزیع نمایی با پارامترهای مقیاس متفاوت  $\theta_1$  و  $\theta_2$  و وزن‌های  $w_1$  و  $w_2$  (با شرط  $w_1 + w_2 = 1$ ) مورد بررسی قرار می‌گیرد. توابع پایه به صورت

$$h(x) = w_1 \frac{1}{\theta_1} e^{-\frac{x}{\theta_1}} + (1 - w_1) \frac{1}{\theta_2} e^{-\frac{x}{\theta_2}}, \quad x \geq 0,$$

$$H(x) = w_1(1 - e^{-\frac{x}{\theta_1}}) + (1 - w_1)(1 - e^{-\frac{x}{\theta_2}}), \quad x \geq 0,$$

تعریف می‌شوند که با اعمال تبدیل وایبول آمیخته، توابع نهایی به صورت

$$f(x) = w_1 \alpha \frac{1}{\theta_1} e^{-\frac{x}{\theta_1}} e^{-\alpha(1 - e^{-\frac{x}{\theta_1}})} + (1 - w_1) \alpha \frac{1}{\theta_2} e^{-\frac{x}{\theta_2}} e^{-\alpha}, \quad x \geq 0,$$

$$F(x) = w_1(1 - e^{-\alpha(1 - e^{-\frac{x}{\theta_1}})}) + (1 - w_1)(1 - e^{-\alpha(1 - e^{-\frac{x}{\theta_2}})}), \quad x \geq 0,$$

$$S(x) = w_1 e^{-\alpha(1 - e^{-\frac{x}{\theta_1}})} + (1 - w_1) e^{-\alpha(1 - e^{-\frac{x}{\theta_2}})}, \quad x \geq 0,$$

$$h(x) = \frac{w_1 \alpha \frac{1}{\theta_1} e^{-\frac{x}{\theta_1}} e^{-\alpha(1 - e^{-\frac{x}{\theta_1}})} + (1 - w_1) \alpha \frac{1}{\theta_2} e^{-\frac{x}{\theta_2}} e^{-\alpha(1 - e^{-\frac{x}{\theta_2}})}}{w_1 e^{-\alpha(1 - e^{-\frac{x}{\theta_1}})} + (1 - w_1) e^{-\alpha(1 - e^{-\frac{x}{\theta_2}})}}, \quad x \geq 0,$$

به دست می‌آیند. برای آمیخته دو توزیع نمایی، تابع لگ درستمایی به صورت

$$\begin{aligned} \ell(w_1, \theta_1, \theta_2, \alpha) = & \sum_{i \in D} \ln(w_1 \alpha \frac{1}{\theta_1} e^{-\frac{t_i}{\theta_1}} e^{-\alpha(1 - e^{-\frac{t_i}{\theta_1}})} + (1 - w_1) \alpha \frac{1}{\theta_2} e^{-\frac{t_i}{\theta_2}} e^{-\alpha(1 - e^{-\frac{t_i}{\theta_2}})}) \\ & + \sum_{i \in C} \ln(1 - [w_1(1 - e^{-\alpha(1 - e^{-\frac{v_i}{\theta_1}})}) + (1 - w_1)(1 - e^{-\alpha(1 - e^{-\frac{v_i}{\theta_2}})})]), \end{aligned} \quad (3)$$

محاسبه می‌شود.

توزیع وایبول با داشتن پارامتر شکل  $\beta$ ، انعطاف پذیری بیشتری در مدل سازی داده‌ها ارائه می‌دهد و قادر است رفتارهای مختلف توزیع احتمال و نرخ خطر را شبیه سازی کند. توابع پایه این توزیع به صورت

$$h(x) = \frac{\beta}{\theta} x^{\beta-1} e^{-\frac{x^\beta}{\theta}}, \quad x \geq 0,$$

$$H(x) = 1 - e^{-\frac{x^\beta}{\theta}}, \quad x \geq 0,$$

هستند. توابع خانواده وایبول آمیخته برای این توزیع به صورت

$$f(x) = \alpha \frac{\beta}{\theta} x^{\beta-1} e^{-\frac{x^\beta}{\theta}} e^{-\alpha(1 - e^{-\frac{x^\beta}{\theta}})}, \quad x \geq 0,$$

$$F(x) = 1 - e^{-\alpha(1 - e^{-\frac{x^\beta}{\theta}})}, \quad x \geq 0,$$

$$S(x) = e^{-\alpha(1 - e^{-\frac{x^\beta}{\theta}})}, \quad x \geq 0,$$

$$h(x) = \alpha \frac{\beta}{\theta} x^{\beta-1} e^{-\frac{x^\beta}{\theta}}, \quad x \geq 0,$$

محاسبه می‌شوند. برای آمیخته دو توزیع وایبول با پارامترهای  $\theta_1$  و  $\theta_2$  و وزن‌های  $w_1$  و  $w_2$ ، توابع پایه به صورت

$$h(x) = w_1 \frac{\beta_1}{\theta_1} x^{\beta_1-1} e^{-\frac{x^{\beta_1}}{\theta_1}} + (1 - w_1) \frac{\beta_2}{\theta_2} x^{\beta_2-1} e^{-\frac{x^{\beta_2}}{\theta_2}}, \quad x \geq 0,$$

$$H(x) = w_1(1 - e^{-\frac{x^{\beta_1}}{\theta_1}}) + (1 - w_1)(1 - e^{-\frac{x^{\beta_2}}{\theta_2}}), \quad x \geq 0,$$

هستند که با جایگذاری به صورت

$$\begin{aligned}
 f(x) &= w_1 \alpha \frac{\beta_1}{\theta_1} x^{\beta_1-1} e^{-\frac{x^{\beta_1}}{\theta_1}} e^{-\alpha(1-e^{-\frac{x^{\beta_1}}{\theta_1}})} + (1-w_1) \alpha \frac{\beta_2}{\theta_2} x^{\beta_2-1} e^{-\frac{x^{\beta_2}}{\theta_2}} e^{-\alpha(1-e^{-\frac{x^{\beta_2}}{\theta_2}})}, & x \geq 0, \\
 F(x) &= w_1 (1 - e^{-\alpha(1-e^{-\frac{x^{\beta_1}}{\theta_1}})}) + (1-w_1) (1 - e^{-\alpha(1-e^{-\frac{x^{\beta_2}}{\theta_2}})}), & x \geq 0, \\
 S(x) &= w_1 e^{-\alpha(1-e^{-\frac{x^{\beta_1}}{\theta_1}})} + (1-w_1) e^{-\alpha(1-e^{-\frac{x^{\beta_2}}{\theta_2}})}, & x \geq 0, \\
 h(x) &= \frac{w_1 \alpha \frac{\beta_1}{\theta_1} x^{\beta_1-1} e^{-\frac{x^{\beta_1}}{\theta_1}} e^{-\alpha(1-e^{-\frac{x^{\beta_1}}{\theta_1}})} + (1-w_1) \alpha \frac{\beta_2}{\theta_2} x^{\beta_2-1} e^{-\frac{x^{\beta_2}}{\theta_2}} e^{-\alpha(1-e^{-\frac{x^{\beta_2}}{\theta_2}})}}{w_1 e^{-\alpha(1-e^{-\frac{x^{\beta_1}}{\theta_1}})} + (1-w_1) e^{-\alpha(1-e^{-\frac{x^{\beta_2}}{\theta_2}})}}, & x \geq 0,
 \end{aligned}$$

به دست می آید. برای آمیخته دو توزیع وایبول، تابع لگ درستمایی به صورت

$$\begin{aligned}
 & \ell(w_1, \theta_1, \beta_1, \theta_2, \beta_2, \alpha) \\
 &= \sum_{i \in D} \ln(w_1 \alpha \frac{\beta_1}{\theta_1} t_i^{\beta_1-1} e^{-\frac{t_i^{\beta_1}}{\theta_1}} e^{-\alpha(1-e^{-\frac{t_i^{\beta_1}}{\theta_1}})} + (1-w_1) \alpha \frac{\beta_2}{\theta_2} t_i^{\beta_2-1} e^{-\frac{t_i^{\beta_2}}{\theta_2}} e^{-\alpha(1-e^{-\frac{t_i^{\beta_2}}{\theta_2}})}) \\
 &+ \sum_{i \in C} \ln(1 - [w_1 (1 - e^{-\alpha(1-e^{-\frac{v_i^{\beta_1}}{\theta_1}})}) + (1-w_1) (1 - e^{-\alpha(1-e^{-\frac{v_i^{\beta_2}}{\theta_2}})})])
 \end{aligned} \tag{4}$$

است.

توزیع رایلی حالتی خاص از توزیع وایبول با  $\beta = 2$  است که در مدل سازی داده هایی با نرخ خطر افزایشی-کاهشی کاربرد دارد. این توزیع به دلیل سادگی ساختار و ویژگی های ریاضیاتی قابل تحلیل، برای مدل سازی زمان های وقوع پدیده هایی که ابتدا احتمال وقوع آن ها افزایش و سپس کاهش می یابد، بسیار مناسب است. توابع پایه آن به صورت

$$\begin{aligned}
 h(x) &= \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}, & x \geq 0, \\
 H(x) &= 1 - e^{-\frac{x^2}{2\theta}}, & x \geq 0,
 \end{aligned}$$

می باشند. توابع خانواده وایبول آمیخته برای توزیع رایلی به صورت

$$\begin{aligned}
 f(x) &= \alpha \frac{x}{\theta} e^{-\frac{x^2}{2\theta}} e^{-\alpha(1-e^{-\frac{x^2}{2\theta}})}, & x \geq 0, \\
 F(x) &= 1 - e^{-\alpha(1-e^{-\frac{x^2}{2\theta}})}, & x \geq 0, \\
 S(x) &= e^{-\alpha(1-e^{-\frac{x^2}{2\theta}})}, & x \geq 0, \\
 h(x) &= \alpha \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}, & x \geq 0,
 \end{aligned}$$

هستند. آمیخته دو توزیع رایلی با پارامترهای  $\theta_1$  و  $\theta_2$  وزن های  $w_1$  و  $w_2$  دارای توابع پایه

$$\begin{aligned}
 h(x) &= w_1 \frac{x}{\theta_1} e^{-\frac{x^2}{2\theta_1}} + (1-w_1) \frac{x}{\theta_2} e^{-\frac{x^2}{2\theta_2}}, & x \geq 0, \\
 H(x) &= w_1 (1 - e^{-\frac{x^2}{2\theta_1}}) + (1-w_1) (1 - e^{-\frac{x^2}{2\theta_2}}), & x \geq 0,
 \end{aligned}$$

است.

توابع تبدیل شد رایلی پس از اعمال تبدیل وایبول آمیخته به صورت

$$\begin{aligned}
 f(x) &= w_1 \alpha \frac{x}{\theta_1} e^{-\frac{x^2}{2\theta_1}} e^{-\alpha(1-e^{-\frac{x^2}{2\theta_1}})} + (1-w_1) \alpha \frac{x}{\theta_2} e^{-\frac{x^2}{2\theta_2}} e^{-\alpha(1-e^{-\frac{x^2}{2\theta_2}})}, & x \geq 0, \\
 F(x) &= w_1 (1 - e^{-\alpha(1-e^{-\frac{x^2}{2\theta_1}})}) + (1-w_1) (1 - e^{-\alpha(1-e^{-\frac{x^2}{2\theta_2}})}), & x \geq 0, \\
 S(x) &= w_1 e^{-\alpha(1-e^{-\frac{x^2}{2\theta_1}})} + (1-w_1) e^{-\alpha(1-e^{-\frac{x^2}{2\theta_2}})}, & x \geq 0, \\
 h(x) &= \frac{w_1 \alpha \frac{x}{\theta_1} e^{-\frac{x^2}{2\theta_1}} e^{-\alpha(1-e^{-\frac{x^2}{2\theta_1}})} + (1-w_1) \alpha \frac{x}{\theta_2} e^{-\frac{x^2}{2\theta_2}} e^{-\alpha(1-e^{-\frac{x^2}{2\theta_2}})}}{w_1 e^{-\alpha(1-e^{-\frac{x^2}{2\theta_1}})} + (1-w_1) e^{-\alpha(1-e^{-\frac{x^2}{2\theta_2}})}}, & x \geq 0,
 \end{aligned}$$

هستند. برای آمیخته دو توزیع رایلی، تابع لگدرستمایی به صورت

$$\begin{aligned}
 \ell(w_1, \theta_1, \theta_2, \alpha) &= \sum_{i \in D} \ln \left( w_1 \alpha \frac{t_i}{\theta_1} e^{-\frac{t_i^2}{2\theta_1}} e^{-\alpha(1-e^{-\frac{t_i^2}{2\theta_1}})} + (1-w_1) \alpha \frac{t_i}{\theta_2} e^{-\frac{t_i^2}{2\theta_2}} e^{-\alpha(1-e^{-\frac{t_i^2}{2\theta_2}})} \right) \\
 &+ \sum_{i \in C} \ln \left( 1 - \left[ w_1 \left( 1 - e^{-\alpha(1-e^{-\frac{v_i^2}{2\theta_1}})} \right) + (1-w_1) \left( 1 - e^{-\alpha(1-e^{-\frac{v_i^2}{2\theta_2}})} \right) \right] \right),
 \end{aligned} \tag{5}$$

است.

### ۳- الگوریتم ژنتیک

الگوریتم ژنتیک یکی از روش‌های بهینه‌سازی فراابتکاری مبتنی بر اصول انتخاب طبیعی و تکامل زیستی است که به‌عنوان ابزاری کارآمد برای حل مسایل پیچیده و غیرخطی معرفی شده است. این الگوریتم با استفاده از یک جمعیت اولیه از جواب‌های ممکن و به‌کارگیری عملگرهایی نظیر انتخاب، ترکیب و جهش، فرآیند جست‌وجویی تکراری را در فضای پارامترها انجام می‌دهد و به تدریج جمعیت را به سمت نواحی با برازندگی بالاتر هدایت می‌کند [11]. مرورهای پژوهشی جدید نشان می‌دهند که این الگوریتم در سال‌های اخیر به‌طور گسترده در مسایل آماری و مهندسی، به‌ویژه در مدل‌های دارای ساختار پیچیده، مورد استفاده قرار گرفته است [12]. الگوریتم ژنتیک نخستین بار در دهه ۱۹۷۰ میلادی توسط هالند و در چارچوب نظریه سازگاری در سامانه‌های طبیعی و مصنوعی به‌صورت نظام‌مند معرفی شد و پس‌از آن به تدریج توسعه یافته و در حوزه‌های مختلف بهینه‌سازی و تحلیل داده به‌کار گرفته شد [13].

در مسایل برآورد پارامتر مدل‌های آمیخته، از جمله توزیع وایبول آمیخته، تابع لگدرستمایی معمولاً دارای ساختاری غیرخطی و چندقله‌ای است و در بسیاری از موارد، روش‌های کلاسیک بهینه‌سازی به مقاداردهی اولیه حساس بوده یا به بیشینه‌های محلی همگرا می‌شوند. این مساله زمانی برجسته‌تر می‌شود که ناهمگنی در داده‌ها وجود داشته باشد یا تعداد پارامترهای مدل افزایش یابد. در چنین شرایطی، استفاده از الگوریتم‌های فراابتکاری به‌عنوان جایگزین یا مکمل روش‌های کلاسیک مورد توجه قرار گرفته است [12]. با وجود کارایی مناسب الگوریتم ژنتیک در بسیاری از کاربردها، نتایج نظری ارائه‌شده در ادبیات بهینه‌سازی نشان می‌دهد که هیچ الگوریتمی را نمی‌توان برای تمامی مسایل بهینه‌سازی به‌طور مطلق برتر دانست. این اصل که به وابستگی عملکرد الگوریتم‌ها به ساختار مساله اشاره دارد، بیان می‌کند که انتخاب روش بهینه‌سازی باید متناسب با ویژگی‌های مساله مورد بررسی انجام شود [14]. بر همین اساس، به‌کارگیری الگوریتم ژنتیک در این پژوهش به دلیل سازگاری آن با ساختار خاص مساله برآورد پارامترهای توزیع وایبول آمیخته و پیچیدگی تابع هدف صورت گرفته است.

در این مطالعه، الگوریتم ژنتیک در سه قالب متفاوت به‌کار گرفته شده است که تفاوت آن‌ها در نحوه تعریف معیار برازندگی و سنجش کیفیت جواب‌هاست. در قالب نخست، معیار برازندگی بر اساس خطای مطلق بین مقادیر مشاهده‌شده و مقادیر برآوردشده تعریف شده است؛ این رویکرد نسبت به داده‌های پرت حساسیت کمتری دارد و می‌تواند در نمونه‌هایی با نوسانات شدید، پایداری بیشتری از خود نشان دهد. در قالب دوم، معیار برازندگی بر پایه مجموع مربعات خطا در نظر گرفته شده است که وزن بیشتری به خطاهای بزرگ اختصاص می‌دهد و الگوریتم را به سمت برازش دقیق‌تر در بخش‌های مرکزی داده‌ها هدایت می‌کند. در قالب سوم، خطای درصدی مطلق به‌عنوان معیار برازندگی استفاده شده است که امکان مقایسه نسبی دقت برآوردها را فراهم می‌سازد و در شرایطی که مقیاس پارامترها متفاوت است، کاربرد ویژه‌ای دارد [15]. در هر سه قالب، هر کروموزوم

نمایانگر یک بردار پارامتر شامل وزن‌های آمیختگی و پارامترهای توزیع وایبول آمیخته است و مقدار برازندگی آن بر اساس معیار متناظر محاسبه می‌شود. عملگرهای الگوریتم به‌گونه‌ای طراحی شده‌اند که قیود مدل، از جمله مثبت بودن پارامترها و برقرار بودن شرط جمع‌پذیری وزن‌ها، در تمامی مراحل حفظ شود. پژوهش‌های اخیر نشان داده‌اند که تنظیم مناسب این عملگرها و پارامترهای کنترلی نقش مهمی در بهبود همگرایی و کیفیت جواب‌ها ایفا می‌کند [16].

تابع هدف به‌صورت کلی مطابق با رابطه

$$F_i = F(x_i), \quad (6)$$

تعریف می‌شود که در آن  $x_i$  کروموزوم شماره  $i$  و  $F(x_i)$  مقدار تابع برازندگی محاسبه شده برای این کروموزوم است. نتیجه این ارزیابی به‌صورت  $F_i$  ثبت می‌شود که می‌تواند در مرحله انتخاب کروموزوم‌ها برای نسل بعدی استفاده شود. تابع برازندگی بسته به نوع مساله می‌تواند شکل‌های مختلفی داشته باشد. در مسایل بیشینه‌سازی، خود تابع هدف می‌تواند مستقیماً به‌عنوان برازندگی در نظر گرفته شود، اما در مسایل کمینه‌سازی باید ابتدا تابع هدف به‌گونه‌ای تبدیل شود که مقادیر کوچک‌تر آن برازندگی بزرگ‌تر ایجاد کنند. یکی از متداول‌ترین روش‌ها برای تبدیل مسایل کمینه‌سازی به بیشینه‌سازی در رابطه

$$F_i = \frac{1}{1 + g(x_i)}, \quad (7)$$

نشان داده شده است که در این رابطه  $g(x_i)$  تابع هدفی است که باید کمینه شود. با این تبدیل، هر چه مقدار  $g(x_i)$  کوچک‌تر باشد، برازندگی بزرگ‌تر می‌شود و احتمال انتخاب کروموزوم  $x_i$  افزایش می‌یابد. پس از محاسبه مقادیر برازندگی برای تمام کروموزوم‌های موجود در جمعیت، معمولاً لازم است این مقادیر نرمال شوند تا بتوان آن‌ها را به‌عنوان احتمال انتخاب به‌کار برد. نرمال‌سازی مطابق رابطه

$$p_i = \frac{F_i}{\sum_{j=1}^N F_j}, \quad (8)$$

انجام می‌شود که در این رابطه  $N$  تعداد کل کروموزوم‌ها در جمعیت است و  $p_i$  احتمال انتخاب کروموزوم شماره  $i$  را نشان می‌دهد. نرمال‌سازی باعث می‌شود مجموع تمام  $p_i$  برابر یک شود و انتخاب کروموزوم‌ها در مرحله انتخاب مانند یک توزیع احتمال انجام گیرد. این فرآیند در عمل به روش چرخ رولت شباهت دارد که در ادامه شرح داده می‌شود. در این روش هر کروموزوم به تناسب برازندگی خود سهمی از چرخ را در اختیار دارد و هر چه این سهم بزرگ‌تر باشد، احتمال انتخاب آن نیز بیشتر خواهد بود.

روش چرخ رولت یکی از متداول‌ترین مکانیزم‌های انتخاب در الگوریتم ژنتیک است که بر پایه احتمال‌های برازندگی عمل می‌کند. در این روش، هر کروموزوم متناسب با مقدار برازندگی خود، سهمی از یک چرخ فرضی دریافت می‌کند؛ به‌گونه‌ای که مجموع همه سهم‌ها معادل یک دایره کامل یا مجموع احتمال یک است. در نتیجه، کروموزوم‌هایی که برازندگی بالاتری دارند، سهم بزرگ‌تری از چرخ خواهند داشت. سپس با تولید یک عدد تصادفی بین صفر و یک یا انتخاب نقطه‌ای تصادفی روی محیط چرخ، مشخص می‌شود که چرخ در کدام بخش متوقف شود و به‌این ترتیب، کروموزوم متناظر با آن بخش انتخاب می‌شود. این فرآیند باعث می‌شود احتمال انتخاب هر کروموزوم متناسب با برازندگی آن باشد و به‌صورت طبیعی، به کروموزوم‌های قوی‌تر شانس بیشتری برای حضور در نسل بعدی بدهد، درحالی‌که همچنان به کروموزوم‌های ضعیف‌تر نیز شانس غیرصفر برای انتخاب می‌دهد تا تنوع حفظ شود. در بسیاری از کاربردها، به‌ویژه در مدل‌سازی‌های آماری و قابلیت اعتماد، می‌توان تابع برازندگی را بر پایه‌ی تابع درست‌نمایی تعریف کرد. در این حالت، هر کروموزوم  $x_i$  بردار پارامترهای مدل را نمایش می‌دهد و برازندگی آن از طریق تابع درست‌نمایی یا منفی لگاریتم درست‌نمایی محاسبه می‌شود. این حالت مطابق رابطه

$$F_i = L(\theta|x_i) \quad \text{یا} \quad F_i = -\ell(\theta|x_i), \quad (9)$$

است که در آن،  $\theta$  بردار پارامترهای مدل است و  $x_i$  مقادیر پیشنهادی این پارامترها را در قالب کروموزوم  $i$  نشان می‌دهد. استفاده از تابع درست‌نمایی به‌عنوان برازندگی باعث می‌شود کروموزوم‌هایی که مقادیر پارامترهای آن‌ها بهتر داده‌ها را برازش می‌کنند، برازندگی بیشتری داشته باشند و احتمال انتخاب آن‌ها برای نسل بعدی افزایش یابد. در مجموع، تعریف و انتخاب مناسب تابع برازندگی در الگوریتم ژنتیک تعیین‌کننده کیفیت جست‌وجو و

سرعت همگرایی به سمت جواب بهینه است. ابتدا مقدار  $Fi$  بر اساس تابع هدف یا درست‌نمایی محاسبه می‌شود، سپس مقادیر نرمال شده  $p_i$  به‌عنوان احتمال انتخاب هر کروموزوم به کار می‌رود و در نهایت، کروموزوم‌های با برزندگی بیشتر شانس بیشتری برای بازتولید و انتقال ویژگی‌های خود به نسل‌های بعدی خواهند داشت. این مراحل موجب می‌شوند جمعیت به تدریج به سمت نواحی بهینه فضای جست‌وجو حرکت کرده و جواب‌های پایدارتر و دقیق‌تری حاصل شوند.

#### ۴- الگوریتم EM

الگوریتم EM یکی از روش‌های کلاسیک و پرکاربرد برای برآورد پارامترها در مدل‌هایی است که شامل داده‌های ناقص یا متغیرهای پنهان هستند. ایده اصلی این الگوریتم بر پایه ساده‌سازی مساله بیشینه‌سازی درست‌نمایی از طریق معرفی متغیرهای پنهان و تجزیه مساله پیچیده به دو گام تکرارشونده بنا شده است. این الگوریتم نخستین بار در دهه ۱۹۷۰ میلادی برای مسایل برآورد درست‌نمایی در حضور داده‌های ناقص ارایه شد و از آن زمان تاکنون در حوزه‌های مختلفی همچون مدل‌های آمیخته، تحلیل بقا و قابلیت اعتماد به‌طور گسترده مورد استفاده قرار گرفته است [17].

در مدل‌های آمیخته، از جمله توزیع وایبول آمیخته، تعلق هر مشاهده به یکی از اجزای آمیخته به‌طور مستقیم قابل مشاهده نیست و همین موضوع باعث پیچیدگی تابع درست‌نمایی می‌شود. الگوریتم EM با در نظر گرفتن این تعلق به‌عنوان یک متغیر پنهان، امکان بازنویسی تابع درست‌نمایی کامل را فراهم می‌کند. در گام امید، مقدار امید ریاضی تابع لگ درست‌نمایی کامل با شرط داده‌های مشاهده‌شده و برآوردهای فعلی پارامترها محاسبه می‌شود. سپس در گام بیشینه‌سازی، این تابع کمکی نسبت به پارامترهای مدل بیشینه می‌گردد و مقادیر جدید پارامترها به دست می‌آیند [5]. این فرآیند به‌صورت تکراری ادامه می‌یابد تا همگرایی حاصل شود. یکی از مزایای اصلی الگوریتم EM، سادگی مفهومی و قابلیت پیاده‌سازی آن در مدل‌های آمیخته است. با این حال، این الگوریتم نیز همانند سایر روش‌های مبتنی بر بیشینه‌سازی درست‌نمایی، می‌تواند به مقداردهی اولیه حساس باشد و در برخی موارد به بیشینه‌های محلی همگرا شود. این مساله به‌ویژه در مدل‌های آمیخته با تعداد پارامترهای زیاد یا ساختار پیچیده‌تر نمود بیشتری دارد [6]. در زمینه توزیع‌های وایبول آمیخته، الگوریتم EM به‌عنوان یکی از روش‌های مرجع برای برآورد پارامترها شناخته می‌شود و در مطالعات متعددی به‌کار گرفته شده است. پژوهش‌های اخیر نشان داده‌اند که اگرچه این الگوریتم در بسیاری از موارد نتایج قابل قبولی ارایه می‌دهد، اما عملکرد آن می‌تواند تحت تأثیر کیفیت مقداردهی اولیه و ناهمگنی داده‌ها قرار گیرد.

از این رو، در مسایل پیچیده، استفاده از روش‌های بهینه‌سازی جایگزین یا ترکیبی در کنار الگوریتم EM مورد توجه قرار گرفته است [7]. در این مقاله، الگوریتم EM به‌عنوان یک روش مرجع برای برآورد پارامترهای مدل وایبول آمیخته در نظر گرفته شده و نتایج حاصل از آن با نتایج به دست آمده از الگوریتم ژنتیک مقایسه می‌شود. هدف از این مقایسه، بررسی نقاط قوت و محدودیت‌های هر یک از این دو رویکرد در مواجهه با تابع درست‌نمایی پیچیده و داده‌های ناهمگن است.

گام امیدگیری (E-step): در این مرحله، احتمال تعلق هر مشاهده به هر مولفه محاسبه می‌شود. این احتمال، امید ریاضی شرطی لگ درست‌نمایی کامل نسبت به متغیرهای پنهان است و داده‌های ناقص را به شکل قابل استفاده برای مرحله بعد تبدیل می‌کند.

گام بیشینه‌سازی (M-step): در این مرحله، پارامترهای مدل با بیشینه کردن تابع لگ درست‌نمایی کامل وزنی به‌روز می‌شوند. این به‌روزرسانی تضمین می‌کند که لگ درست‌نمایی کامل در هر تکرار افزایش یابد یا حداقل تغییر نکند. این دو گام تا رسیدن به معیار همگرایی تکرار می‌شوند. معیار همگرایی معمولاً به صورت

$$|\ell(\theta^{(k+1)}) - \ell(\theta^{(k)})| < \epsilon, \quad (10)$$

تغییرات کوچک در لگ درست‌نمایی تعریف می‌شود. در گام E، احتمال تعلق هر مشاهده  $x_i$  به مولفه  $j$  از مدل به صورت

$$\gamma_{ij}^{(k)} = P(Z_i = j | x_i, \theta^{(k)}) = \frac{\pi_j^{(k)} f_j(x_i | \theta_j^{(k)})}{\sum_{l=1}^m \pi_l^{(k)} f_l(x_i | \theta_l^{(k)})}, \quad i = 1, \dots, n, \quad j = 1, \dots, m,$$

محاسبه می‌شود. در گام M، پارامترهای وزنی و مولفه‌ها به صورت

$$\pi_j^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \gamma_{ij}^{(k)},$$

$$\theta_j^{(k+1)} = \operatorname{argmax}_{\theta_j} \sum_{i=1}^n \gamma_{ij}^{(k)} \log f_j(x_i | \theta_j),$$

به روزرسانی می‌شوند. فرآیند تکرار گام‌های  $E$  و  $M$  تا زمانی ادامه می‌یابد که تغییر لگ درستنمایی از آستانه مشخصی کوچک‌تر شود و در نهایت، پارامترهای نهایی مدل با نماد  $\hat{\theta}$  به صورت

$$\hat{\theta}, \hat{\pi} = \operatorname{argmax}_{\theta, \pi} \ell(\theta, \pi),$$

گزارش می‌شوند. در مدل آمیخته دو مولفه، وزن مولفه اول با  $w_1$  و وزن مولفه دوم با  $1 - w_1$  مشخص می‌شود. احتمال تعلق هر مشاهده به مولفه اول، برای داده‌های غیرسانسور شده و سانسور شده به ترتیب به صورت

$$\gamma_{i1}^{(D)} = \frac{w_1 f_1(t_i)}{w_1 f_1(t_i) + (1 - w_1) f_2(t_i)}, \quad (11)$$

$$\gamma_{i1}^{(C)} = \frac{w_1 S_1(v_i)}{w_1 S_1(v_i) + (1 - w_1) S_2(v_i)}, \quad (12)$$

محاسبه می‌گردد. وزن جدید مولفه اول با میانگین وزنی احتمال تعلق هر مشاهده به آن مولفه به شکل

$$\hat{w}_1 = \frac{\sum_{i \in D} \gamma_{i1}^{(D)} + \sum_{i \in C} \gamma_{i1}^{(C)}}{|D| + |C|}, \quad (13)$$

به دست می‌آید و وزن مولفه دوم برابر با  $1 - \hat{w}_1$  خواهد بود. برای هر مولفه توزیع نمایی، تابع چگالی و بقا به صورت

$$f(t|\theta) = \frac{1}{\theta} e^{-\frac{t}{\theta}}, S(t|\theta) = e^{-\frac{t}{\theta}}.$$

تعریف می‌شود که پارامترهای مقیاس برای هر مولفه با میانگین وزنی به صورت

$$\hat{\theta}_1 = \frac{\sum_{i \in D} \gamma_{i1}^{(D)} t_i + \sum_{i \in C} \gamma_{i1}^{(C)} v_i}{\sum_{i \in D} \gamma_{i1}^{(D)}},$$

$$\hat{\theta}_2 = \frac{\sum_{i \in D} (1 - \gamma_{i1}^{(D)}) t_i + \sum_{i \in C} (1 - \gamma_{i1}^{(C)}) v_i}{\sum_{i \in D} (1 - \gamma_{i1}^{(D)})},$$

است. با پارامترهای مقیاس  $\theta > 0$  و شکل  $\beta > 0$ ، توابع چگالی و بقا توزیع وایبول به صورت

$$f(t|\theta, \beta) = \frac{\beta}{\theta} \left(\frac{t}{\theta}\right)^{\beta-1} e^{-(t/\theta)^\beta}, S(t|\theta, \beta) = e^{-(t/\theta)^\beta}.$$

تعریف می‌شوند که برای مولفه اول با  $\beta_1$  ثابت

$$\hat{\theta}_1 = \frac{\sum_{i \in D} \gamma_{i1}^{(D)} t_i^{\beta_1} + \sum_{i \in C} \gamma_{i1}^{(C)} v_i^{\beta_1}}{\sum_{i \in D} \gamma_{i1}^{(D)}},$$

است و معادله نمره برای برآورد  $\beta_1$  به صورت

$$\sum_{i \in D} \gamma_{i1}^{(D)} \left( \frac{1}{\beta_1} + \ln t_i \right) - \frac{1}{\hat{\theta}_1} \left[ \sum_{i \in D} \gamma_{i1}^{(D)} t_i^{\beta_1} \ln t_i + \sum_{i \in C} \gamma_{i1}^{(C)} v_i^{\beta_1} \ln v_i \right] = 0,$$

است. با پارامتر مقیاس  $\theta > 0$ ، توابع چگالی و بقا توزیع راپلی به صورت

$$f(t|\theta) = \frac{t}{\theta} e^{-\frac{t^2}{2\theta}}, S(t|\theta) = e^{-\frac{t^2}{2\theta}},$$

هستند. پارامترهای مقیاس هر مولفه به شکل

$$\hat{\theta}_1 = \frac{\sum_{i \in D} \gamma_{i1}^{(D)} t_i^2 + \sum_{i \in C} \gamma_{i1}^{(C)} v_i^2}{2 \sum_{i \in D} \gamma_{i1}^{(D)}},$$

$$\hat{\theta}_2 = \frac{\sum_{i \in D} (1 - \gamma_{i1}^{(D)}) t_i^2 + \sum_{i \in C} (1 - \gamma_{i1}^{(C)}) v_i^2}{2 \sum_{i \in D} (1 - \gamma_{i1}^{(D)})},$$

برآورد می‌شوند. این فرآیند در هر تکرار ادامه می‌یابد و معیار توقف می‌تواند بر اساس کاهش تغییرات لگاریتم درست‌نمایی ناقص یا تغییرات جزئی پارامترها در دو تکرار متوالی تعریف شود. الگوریتم EM به این ترتیب امکان برآورد کارآمد و پایدار پارامترهای آمیخته‌های نمای، وایبول و رایلی را حتی در حضور داده‌های سانسور شده فراهم می‌کند.

## ۵- مطالعه شبیه‌سازی

در این فصل، به منظور ارزیابی عملکرد روش پیشنهادی مبتنی بر الگوریتم ژنتیک، مجموعه‌ای از مطالعات شبیه‌سازی عددی طراحی و اجرا شده است. در این مطالعات، داده‌های سانسور شده با نسبت‌های مختلف آمیختگی، اندازه‌های نمونه و سطوح گوناگون سانسور تولید شده و پارامترهای مدل‌های آمیخته نمای، وایبول و رایلی با استفاده از الگوریتم ژنتیک برآورد گردیده‌اند. سپس نتایج به دست آمده با خروجی‌های الگوریتم EM مقایسه شده تا دقت، پایداری و کارایی روش پیشنهادی در شرایط مختلف بررسی شود. هدف از این فصل، سنجش میزان کارآمدی و توانایی الگوریتم ژنتیک در تخمین پارامترهای مدل‌های آمیخته در حضور داده‌های سانسور شده و مقایسه آن با روش‌های کلاسیک است.

برای ارزیابی کارایی برآوردگرها در مدل‌های آمیخته نمای، وایبول و رایلی تعمیم یافته دومولفه‌ای، یک مطالعه شبیه‌سازی طراحی و اجرا شد. داده‌ها مطابق تابع چگالی آمیخته آن‌ها تولید شدند که ساختار آن از خانواده‌های تعمیم یافته پیروی می‌کند که در آن‌ها  $\theta_1, \theta_2 > 0$  و  $\alpha > 0$  و  $w_1 \in (0,1)$  هستند. نمونه‌گیری از این مدل‌ها به صورت استاندارد آمیخته انجام شد. ابتدا مولفه با  $Z \sim Bernoulli(w_1)$  انتخاب شد به طوری که اگر  $Z = 1$  باشد  $\theta = \theta_1$  و اگر  $Z = 0$  باشد  $\theta = \theta_2$  قرار داده شد. برای اعمال سانسور راست با نرخ هدف  $c \in \{0.1, 0.2, 0.3, 0.4\}$ ، متغیرهای سانسور مستقل  $G_i$  تولید شدند و در تمامی سناریوها پارامترها ثابت  $\theta_1 = 3$ ،  $\theta_2 = 4$ ،  $\alpha = 2$  در نظر گرفته شد.

سطوح نسبت آمیختگی  $\{0.9, 0.7, 0.5, 0.3, 0.1\}$ ، اندازه نمونه  $n \in \{25, 100, 250, 500\}$  و نرخ سانسور  $c \in \{0.1, 0.2, 0.3, 0.4\}$  آزموده شدند. برای هر ترکیب  $(w_1, n, c)$  تعداد  $m = 100$  تکرار مستقل اجرا و چهار برآوردگر مقایسه شدند. الگوریتم‌های مورد استفاده شامل الگوریتم EM و سه نسخه الگوریتم ژنتیک  $GAMAE$ ،  $GAMSE$  و  $GAMAPE$  هستند. معیارهای ارزیابی برای هر پارامتر  $\theta$  به صورت

$$\text{میانگین توان دوم خطای انتظاری}(\theta) = \frac{1}{m} \sum_{r=1}^m (\hat{\theta}^{(r)} - \theta)^2, \quad \text{میانگین اریبی}(\theta) = \frac{1}{m} \sum_{r=1}^m (\hat{\theta}^{(r)} - \theta),$$

محاسبه شد و برای گزارش، میانگین توان دوم خطای انتظاری و قدر مطلق میانگین اریبی به تفکیک  $w_1$ ،  $n$  و  $c$  تجمیع شدند.

از آنجاکه مدل‌های آمیخته مورد بررسی شامل بیش از یک پارامتر مجهول هستند، استفاده از معیارهای تک پارامتری نظیر میانگین توان دوم خطای انتظاری و میانگین اریبی به تنهایی قادر به توصیف کامل رفتار همزمان برآوردگرها نیست. به ویژه، این معیارها اطلاعاتی درباره وابستگی و همبستگی بین مولفه‌های مختلف بردار پارامترها ارائه نمی‌کنند. از این رو، به منظور ارزیابی جامع‌تر عملکرد روش‌ها در چارچوب چند پارامتری، شاخص واریانس تعمیم یافته نیز مورد توجه قرار گرفت. این شاخص به صورت دترمینان ماتریس کوواریانس تجربی بردار برآوردگرها تعریف می‌شود و معیاری کلی از میزان پراکندگی مشترک برآوردها در فضای پارامترها فراهم می‌آورد. مقادیر کوچک‌تر این شاخص بیانگر تمرکز بیشتر برآوردگرها و پایداری بالاتر آن‌ها در تخمین همزمان پارامترها است. محاسبه واریانس تعمیم یافته بر اساس خروجی تکرارهای شبیه‌سازی و با استفاده از ماتریس

کوواریانس تجربی بردار پارامترها انجام شد. نتایج حاصل از بررسی سناریوهای نماینده نشان داد که الگوی کلی واریانس تعمیم یافته با نتایج به دست آمده از معیارهای  $EMSE$  و  $EBIAS$  همسو بوده و روش های مبتنی بر الگوریتم ژنتیک، به ویژه  $GAMSE$  و  $GAMAPE$ ، از نظر پایداری چند پارامتری عملکرد مطلوب تری نسبت به الگوریتم  $EM$  ارائه می کنند.

جدول ۱- میانگین خطای دوم انتظاری به ازای نسبت آمیختگی.

Table 1- EMSE by  $w_1$ .

توزیع	0.1	0.3	0.5	0.7	0.9	روش/ $w_1$
نمایی	0.0291	0.0272	0.0266	0.0272	0.0291	EM
	0.0260	0.0276	0.0292	0.0308	0.0324	GA-MAE
	0.0250	0.0266	0.0282	0.0298	0.0314	GA-MAPE
	0.0254	0.0270	0.0286	0.0302	0.0318	GA-MSE
وایبول	0.0343	0.0321	0.0314	0.0321	0.0343	EM
	0.0292	0.0306	0.0320	0.0334	0.0348	GA-MAE
	0.0282	0.0296	0.0310	0.0324	0.0338	GA-MAPE
	0.0286	0.0300	0.0314	0.0328	0.0342	GA-MSE
رایلی	0.3198	0.2994	0.2926	0.2994	0.3198	EM
	0.2731	0.2881	0.3031	0.3181	0.3331	GA-MAE
	0.2641	0.2791	0.2941	0.3091	0.3241	GA-MAPE
	0.2671	0.2821	0.2971	0.3121	0.3271	GA-MSE

مقدار میانگین توان دوم خطای انتظاری در حوالی  $w_1 \approx 0.5$  برای روش  $EM$  به حداقل می رسد و با دور شدن از این مقدار، چه به سمت صفر و چه به سمت یک، به تدریج افزایش می یابد. این رفتار بیانگر آن است که الگوریتم  $EM$  در شرایطی که میزان اختلاط متعادل تر است، عملکرد دقیق تری دارد، اما در ترکیب های نامتقارن تر دچار ناپایداری نسبی می شود و حساسیت بیشتری نسبت به تغییرات نسبت آمیختگی نشان می دهد. در مقابل، روش های ژنتیک در دو سر بازه آمیختگی مقادیر کوچک تری از میانگین توان دوم خطای انتظاری را ثبت می کنند و روند تغییراتشان ملایم تر است؛ به عبارتی، این روش ها در مواجهه با درجات مختلف اختلاط، پایداری و سازگاری بیشتری از خود نشان می دهند و قادرند حتی در شرایط نامتقارن نیز برآوردهای قابل اعتماد ارائه کنند. همچنین، در اغلب سطوح، روش  $GAMSE$  کمینه نسبی را ارائه کرده است که حاکی از توانایی بالاتر آن در برآورد پارامترها حتی در شرایط ترکیب های نابرابر است. روش  $GAMAPE$  نیز به طور مشابه عملکرد پایداری دارد و تفاوت آن با  $GAMSE$  در اغلب نقاط اندک است. در میانه بازه، همگرایی عملکرد روش ها بیشتر است و مقادیر میانگین مربعات خطای انتظاری به هم نزدیک می شوند، اما در نسبت های آمیختگی نامتوازن، شکاف عملکردی به سود  $GAMSE$  و  $GAMAPE$  افزایش می یابد.

جدول ۲- میانگین خطای دوم انتظاری به ازای اندازه نمونه.

Table 2- EMSE by n.

توزیع	500	250	100	25	روش/n
نمایی	0.0238	0.0259	0.0287	0.0329	EM
	0.0252	0.0273	0.0301	0.0343	GA-MAE
	0.0242	0.0263	0.0291	0.0333	GA-MAPE
	0.0246	0.0267	0.0295	0.0337	GA-MSE

جدول ۲- ادامه

Table 2- Continued.

توزیع	500	250	100	25	روش/ n
وایبول	0.0286	0.0308	0.0338	0.0383	EM
	0.0277	0.0300	0.0330	0.0375	GA-MAE
	0.0267	0.0290	0.0320	0.0365	GA-MAPE
	0.0271	0.0294	0.0324	0.0369	GA-MSE
رایلی	0.2648	0.2846	0.3151	0.3584	EM
	0.2617	0.2833	0.3120	0.3553	GA-MAE
	0.2527	0.2743	0.3030	0.3463	GA-MAPE
	0.2557	0.2773	0.3060	0.3493	GA-MSE

کاهش یکنواخت میانگین توان دوم خطای انتظاری را با افزایش  $n$  در همه روش‌ها مشاهده می‌شود. شیب کاهش برای روش‌های ژنتیک بزرگ‌تر است و در بیشتر سطوح،  $GAMSE$  کمینه نسبی را ارائه می‌کند. افزایش اندازه نمونه موجب همگرایی عملکرد روش‌ها می‌شود، هرچند در بازه‌های کوچک تا متوسط، مزیت نسبی روش‌های ژنتیک چشمگیر است. این نکته در طراحی اندازه نمونه کارآمد قابل استفاده است.

جدول ۳- میانگین خطای دوم انتظاری به ازای نرخ سانسور.

Table 3- EMSE by censor.

توزیع	0.4	0.3	0.2	0.1	روش/ c
نمایی	0.0312	0.0290	0.0268	0.0246	EM
	0.0325	0.0303	0.0281	0.0259	GA-MAE
	0.0315	0.0293	0.0271	0.0249	GA-MAPE
	0.0319	0.0297	0.0275	0.0253	GA-MSE
وایبول	0.3651	0.3411	0.3171	0.2931	EM
	0.3567	0.3327	0.3087	0.2847	GA-MAE
	0.3467	0.3227	0.2987	0.2747	GA-MAPE
	0.3507	0.3267	0.3027	0.2787	GA-MSE
رایلی	0.0734	0.1773	0.9472	0.7172	EM
	0.7633	0.1463	0.9162	0.6862	GA-MAE
	0.8632	0.5630	0.8262	0.5962	GA-MAPE
	0.1633	0.8630	0.8562	0.6262	GA-MSE

افزایش یکنواخت میانگین توان دوم خطای انتظاری را با افزایش نرخ سانسور مشاهده می‌کنیم که توازی نسبی خطوط مربوط به روش‌های ژنتیک نسبت به منحنی  $EM$  را آشکار می‌سازد. این توازی بیانگر پایداری نسبی الگوریتم‌های ژنتیک در مواجهه با کاهش اطلاعات ناشی از سانسور است و حاکی از توانایی آن‌ها در حفظ عملکرد قابل قبول حتی در شرایط از دست رفت نمونه می‌باشد. از منظر کاربردی، این ویژگی برای طراحی طرح‌های پیگیری و مطالعات بقا که با افت نمونه مواجه هستند، اهمیت ویژه‌ای دارد، زیرا امکان برآورد دقیق‌تر پارامترها را در شرایط نامطلوب فراهم می‌کند. این نتایج نشان می‌دهند که الگوریتم‌های ژنتیک می‌توانند گزینه مطمئن‌تری نسبت به  $EM$  در سناریوهای با سانسور بالا باشند.

جدول ۴- قدر مطلق میانگین اربیی به ازای نسبت آمیختگی.

Table 4- |EBIAS| by  $w_1$ .

توزیع	0.1	0.3	0.5	0.7	0.9	روش/ $w_1$
نمایی	0.0782	0.0720	0.0662	0.0662	0.0720	EM
	0.0615	0.0651	0.0687	0.0723	0.0759	GA-MAE

جدول ۴- ادامه.

Table 4- Continued.

توزیع	0.1	0.3	0.5	0.7	0.9	روش / $w_1$
	0.0559	0.0603	0.0647	0.0691	0.0735	GA-MAPE
	0.0577	0.0617	0.0657	0.0697	0.0737	GA-MSE
وایبول	0.0767	0.0690	0.0657	0.0668	0.0723	EM
	0.0596	0.0630	0.0664	0.0698	0.0732	GA-MAE
	0.0538	0.0580	0.0622	0.0664	0.0706	GA-MAPE
	0.0553	0.0591	0.0629	0.0667	0.0705	GA-MSE
رایلی	0.0746	0.0673	0.0642	0.0652	0.0704	EM
	0.0613	0.0645	0.0677	0.0709	0.0741	GA-MAE
	0.0554	0.0594	0.0634	0.0674	0.0714	GA-MAPE
	0.0572	0.0608	0.0644	0.0680	0.0716	GA-MSE

میانگین اریبی برای روش های ژنتیک در بخش عمده بازه  $w_1$  کمتر از EM است و این نشان دهنده دقت بالاتر و پایداری بیشتر الگوریتم های ژنتیک در مواجهه با تغییرات نسبت آمیختگی است. تنها در حوالی میانه بازه، الگوریتم EM گاه رقابتی ظاهر می شود و مقدار اریبی آن به روش های ژنتیک نزدیک است، که بیانگر توانایی نسبی EM در شرایط اختلاط متعادل است. با دور شدن از میانه، به ویژه در نقاطی که یکی از مولفه ها سهم غالب دارد، مزیت الگوریتم های ژنتیک برجسته تر می شود. این رفتار با دشواری تفکیک مولفه کوچک تر در انتهای بازه همخوان است و نشان می دهد که جست و جوی سراسری در GAMSE و GAMAPE توانایی آن ها را در یافتن برآوردهای دقیق تر و کاهش میانگین اریبی افزایش می دهد. همچنین همواری خطوط مربوط به روش های ژنتیک نشان می دهد که این الگوریتم ها نسبت به تغییرات شدید در نسبت آمیختگی حساسیت کمتری دارند و برآوردهایی پایدارتر ارائه می کنند.

جدول ۵- قدر مطلق میانگین اریبی به ازای اندازه نمونه.

Table 5- | EBIAS | by n.

توزیع	500	250	100	25	روش / n
نمایی	0.0505	0.0606	0.0747	0.0569	EM
	0.0484	0.0595	0.0737	0.0949	GA-MAE
	0.0444	0.0555	0.0696	0.0909	GA-MAPE
	0.0454	0.0565	0.0707	0.0919	GA-MSE
وایبول	0.0482	0.0597	0.0748	0.0977	EM
	0.0451	0.0563	0.0710	0.0933	GA-MAE
	0.0409	0.0520	0.0668	0.0890	GA-MAPE
	0.0416	0.0528	0.0675	0.0898	GA-MSE
رایلی	0.0476	0.0585	0.0728	0.0945	EM
	0.0475	0.0581	0.0720	0.0931	GA-MAE
	0.0433	0.0538	0.0677	0.0888	GA-MAPE
	0.0443	0.0548	0.0687	0.0898	GA-MSE

کاهش یکنواخت قدر مطلق میانگین اریبی را با افزایش اندازه نمونه  $n$  به وضوح مشاهده می شود. هر چند با بزرگ تر شدن نمونه، اختلاف میان روش ها کاهش یافته و تفاوت ها در مقادیر بزرگ  $n$  تقریباً ناچیز می شود، اما در بازه کاربردی نمونه های متوسط، مزیت نسبی الگوریتم GAMSE همچنان محسوس است و نشان می دهد که این روش در شرایطی که جمع آوری نمونه های بزرگ امکان پذیر نیست، عملکرد پایدارتری ارائه می دهد. این روند

همچنین تاکید می‌کند که روش‌های ژنتیک قادرند با بهره‌گیری موثر از داده‌های محدود، اریب را به‌طور قابل‌توجهی کاهش دهند و دقت برآوردها را در سناریوهای عملی افزایش دهند.

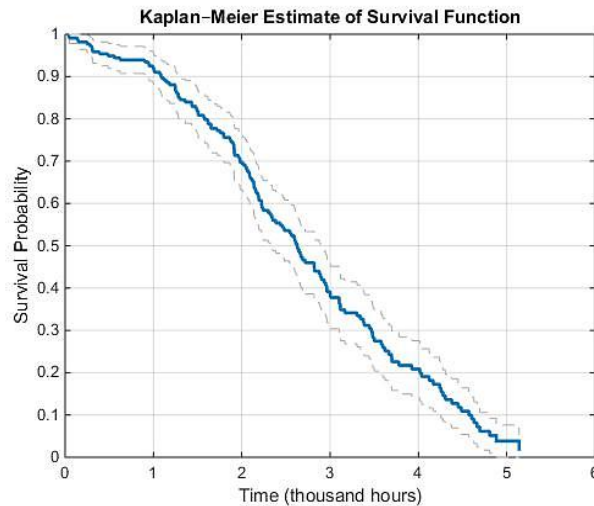
جدول ۶- قدر مطلق میانگین اریبی به ازای نرخ سانسور.  
Table 6- | EBIAS | by censor.

توزیع	0.4	0.3	0.2	0.1	روش/c
نمایی	0.0822	0.0742	0.0662	0.0582	EM
	0.0807	0.0727	0.0647	0.0567	GA-MAE
	0.0767	0.0687	0.0607	0.0527	GA-MAPE
	0.0777	0.0697	0.0617	0.0527	GA-MSE
وایبول	0.0828	0.0743	0.0658	0.0573	EM
	0.0787	0.0705	0.0623	0.0541	GA-MAE
	0.0743	0.0662	0.0581	0.0500	GA-MAPE
	0.0749	0.0669	0.0589	0.0509	GA-MSE
رایلی	0.0806	0.0724	0.0642	0.0560	EM
	0.0797	0.0717	0.0637	0.0557	GA-MAE
	0.0753	0.0674	0.0595	0.0516	GA-MAPE
	0.0763	0.0684	0.0605	0.0526	GA-MSE

با افزایش میزان سانسور، میانگین اریبی در همه روش‌ها افزایش می‌یابد و شدت این اثر برای الگوریتم *EM* بیشتر است. در بسیاری از سطوح نرخ سانسور، *GAMSE* و *GAMAPE* همچنان کمینه نسبی را ارائه می‌کنند و نشان می‌دهند که توانایی بیشتری در مقابله با کاهش اطلاعات دارند. این الگو حاکی از کارایی بهتر روش‌های ژنتیک در شرایطی است که داده‌ها ناقص یا از دست رفته‌اند. کاهش اریبی در سطوح بالای سانسور اهمیت ویژه‌ای برای تحلیل‌های بقا دارد، زیرا دقت برآورد پارامترها در حضور داده‌های ناقص تعیین‌کننده صحت نتیجه‌گیری‌ها و پیش‌بینی‌ها است. علاوه بر این، همواری رفتار روش‌های ژنتیک نسبت به تغییرات نرخ سانسور، پایداری و اعتمادپذیری بالاتر این الگوریتم‌ها را در شرایط واقعی داده‌های ناقص نشان می‌دهد. به‌منظور ارزیابی عملی کارایی روش پیشنهادی در شرایط واقعی، مجموعه داده‌ای شامل زمان‌های خرابی شیشه جلوی یک مدل خاص از هواپیما مورد بررسی قرار گرفت. این داده‌ها پیش‌تر در مطالعات قابلیت اعتماد گزارش شده‌اند و شامل مشاهداتی هستند که بخشی از آن‌ها زمان خرابی واقعی را نشان می‌دهند، درحالی‌که برای بخشی دیگر، خرابی تا پایان دوره مشاهده رخ نداده است؛ بنابراین، داده‌های مربوط به این دسته دوم به‌عنوان داده‌های سانسور شده از راست در تحلیل‌ها لحاظ شدند.

مجموعه داده مورد استفاده شامل ۱۵۳ مشاهده است که از این تعداد، ۸۸ مشاهده به زمان‌های خرابی ثبت شده و ۶۵ مشاهده به زمان‌های سرویس بدون خرابی اختصاص دارند. تمامی زمان‌ها برحسب هزار ساعت اندازه‌گیری شده‌اند. برای مطالعات بیشتر و مشاهده داده‌ها به مقاله [18] مراجعه شود. در مرحله نخست تحلیل، به‌منظور دستیابی به یک تصویر تجربی از رفتار قابلیت اعتماد سیستم، تابع بقا با استفاده از روش غیرپارامتری کاپلان-مایر برآورد شد. این برآورد غیرپارامتری که شامل داده‌های سانسور شده نیز می‌شود، به‌عنوان مرجع اولیه و معیار مقایسه‌ای برای ارزیابی دقت و کفایت مدل‌های پارامتری پیشنهادی به‌کار گرفته شد.

استفاده از کاپلان-مایر امکان بررسی توزیع تجربی زمان‌های خرابی و شناخت الگوهای بقا بدون اعمال فرضیات پارامتری را فراهم می‌کند و بنابراین پایه‌ای برای مقایسه مدل‌های آمیخته و دومولفه‌ای ارائه می‌دهد. این برآورد به پژوهشگر اجازه می‌دهد تا اختلاف بین مدل پارامتری و داده‌های واقعی را به‌صورت بصری و کمی ارزیابی کند. نمودار کاپلان-مایر به همراه تابع قابلیت اعتماد مدل آمیخته وایبول در شکل زیر نشان داده شده است. در ادامه، یک مدل آمیخته از خانواده توزیع وایبول دومولفه‌ای برای توصیف الگوی خرابی داده‌ها در نظر گرفته شد. پارامترهای این مدل با بهره‌گیری از الگوریتم ژنتیک و از طریق کمینه‌سازی فاصله بین تابع قابلیت اعتماد مدل و برآورد کاپلان-مایر تخمین زده شدند. سه معیار خطا شامل میانگین



شکل ۱- برآورد تابع بقا به روش کاپلان-مایر.

Figure 1- Kaplan-Meier estimate of survival function.

مربعات خطا<sup>۱</sup>، میانگین قدر مطلق خطا<sup>۲</sup> و میانگین قدر مطلق درصد خطا<sup>۳</sup> به عنوان توابع برازندگی در الگوریتم ژنتیک مورد استفاده قرار گرفتند که به ترتیب منجر به تعریف سه نسخه از روش پیشنهادی با عناوین *GAMSE*، *GAMAE* و *GAMAPE* شدند.

جدول ۷- مقایسه روش های پیشنهادی در توزیع های نمایی، رایلی و وایبول.

Table 7- Comparison of the proposed methods for exponential, Rayleigh, and Weibull distributions.

w	$\theta_1$	$\theta_2$	$\alpha$	$\beta_1$	$\beta_2$	MSE	MAE	MAPE	روش	توزیع
0.5354	3.3201	1.3019	3.3198			0.2734	0.5036	69.648	EM	
0.1881	2.703	2.349	1.2628			0.2055	0.4238	59.627	$GA_{MSE}$	نمایی
0.1925	2.6273	2.349	1.2352	-	-	0.2056	0.4225	59.472	$GA_{MAE}$	
0.2101	2.7039	2.349	1.3913			0.2057	0.4227	59.482	$GA_{MAPE}$	
0.4879	1.7518	1.7518	0.9266			0.1568	0.3430	0.45.339	EM	
0.3721	0.9597	9.8788	1.0000			0.0860	0.17301	20.298	$GA_{MSE}$	رایلی
0.5581	1.3649	6.3729	0.9999	-	-	0.0988	0.1478	15.7500	$GA_{MAE}$	
0.3083	1.1637	3.9557	0.49819			0.1047	0.1511	16.014	$GA_{MAPE}$	
0.4686	1.66	1.66	0.8234	1	1	0.3824	0.6047	83.091	EM	
1.0000	2.651	0.4489	1	1	0.9805	0.2861	0.5068	70.756	$GA_{MSE}$	وایبول
0.9479	7.1942	4.1727	1	0.9963	0.7893	0.2891	0.5041	70.647	$GA_{MAE}$	
0.9738	1.8588	5.7333	1	0.9836	0.9543	0.2891	0.5045	70.636	$GA_{MAPE}$	

عملکرد روش پیشنهادی مبتنی بر الگوریتم ژنتیک با برخی روش های متداول مقایسه شد. به منظور افزایش اعتبار نتایج و کاهش خطر بیش برآزش، از روش اعتبارسنجی متقاطع ده بخشی در هر دو چارچوب *EM* و *GA* استفاده شد. نتایج حاصل نشان داد که الگوریتم *EM* با افت قابل توجهی مواجه می شود، در حالی که روش های مبتنی بر الگوریتم ژنتیک از پایداری بیشتر و توان تعمیم بالاتری در مواجهه با داده های جدید برخوردارند؛ بنابراین، می توان گفت اگرچه روش *EM* مبتنی بر بیشینه سازی مستقیم تابع درست نمایی است، اما در حضور سانسور و هم خطی پارامترها، این روش مستعد ناپایداری عددی است. در مقابل، روش های فراابتکاری نظیر الگوریتم ژنتیک به دلیل استفاده از معیارهای برآزش غیرمستقیم و کران گذاری پارامترها، برآوردهای پایدارتر اما بالقوه بایاس دار ارائه می دهند. به طور کلی، نتایج این مطالعه کاربردی حاکی از آن است که روش

<sup>1</sup> Mean Square Error (MSE)

<sup>2</sup> Mean Absolute Error (MAE)

<sup>3</sup> Mean Absolute Percentage Error (MAPE)

پیشنهادی مبتنی بر الگوریتم ژنتیک می‌تواند در تحلیل داده‌های واقعی سانسور شده، برآوردهایی پایدارتر و قابل اعتمادتر نسبت به روش‌های کلاسیک ارائه دهد و از این منظر، گزینه‌ای مناسب برای کاربردهای عملی در تحلیل قابلیت اعتماد محسوب می‌شود.

## ۶- نتیجه‌گیری

یافته‌های شبیه‌سازی در سه خانواده آمیخته نمایی، وایبول و رایلی نشان داد که کارایی برآوردها به صورت نظام‌مند تابع نسبت آمیختگی، اندازه نمونه و نرخ سانسور است. در هر سه خانواده، برآوردها در حوالی نسبت‌های میانی آمیختگی، به ویژه  $w_1 \approx 0.5$ ، عملکرد رقابتی تری ارائه کرد و مقادیر میانگین توان دوم خطای انتظاری و قدر مطلق میانگین اریبی آن نسبتاً کوچک بود. با دور شدن از ناحیه میانی و حرکت به سوی مقادیر بسیار کوچک یا بسیار بزرگ  $w_1$ ، هر دو معیار خطا و اریب در  $EM$  افزایش یافت؛ در مقابل، روش‌های ژنتیک، به ویژه  $GAMSE$  و  $GAMAPE$  رفتاری پایدارتر نشان دادند و در دو سر بازه آمیختگی برتری نسبی خود را حفظ کردند. این الگو در همه خانواده‌ها مشاهده شد و اهمیت هندسه تابع هدف و دشواری تفکیک مولفه‌ها را در شکل‌دهی به سطح خطا برجسته کرد. افزایش اندازه نمونه در تمامی سناریوها موجب کاهش یکنواخت میانگین توان دوم خطای انتظاری و قدر مطلق میانگین اریبی شد و همگرایی عملکرد روش‌ها را در سه خانواده رقم زد. با این حال، در بازه عملی اندازه‌های نمونه کوچک تا متوسط که در مسایل بقا و قابلیت اعتماد رایج است، مزیت نسبی روش‌های ژنتیک محسوس باقی ماند. در حضور سانسور راست مستقل، همان‌گونه که انتظار می‌رفت، خطا و اریب افزایش یافت و حساسیت  $EM$  در نرخ‌های بالای سانسور بیشتر بود؛ در حالی که روش‌های ژنتیک مقاومت بیشتری در برابر از دست رفت اطلاعات نشان دادند. در مجموع، برای شرایط نامتوازن آمیختگی یا سانسورهای متوسط تا زیاد، به کارگیری  $GAMSE$  و در بسیاری از مواقع  $GAMAPE$  از منظر دقت و بی‌طرفی توصیه‌پذیر است. با وجود انسجام نتایج، این مقاله با محدودیت‌هایی همراه است. داده‌ها به صورت مصنوعی و تحت مدل درست مشخص شده تولید شدند؛ بنابراین، اثر ناهمخوانی مدل، وجود داده‌های پرت یا ساختارهای پیچیده سانسور مانند سانسور وابسته یا دوطرفه بررسی نشد. همچنین، تنظیمات الگوریتم ژنتیک شامل اندازه جمعیت، نرخ جهش، راهبرد گزینش و معیار خاتمه به صورت یکنواخت و ابتکاری تعیین گردید و امکان بهینه‌سازی سناریو ویژه این تنظیمات ارزیابی نشد. در نهایت، تحلیل مقایسه‌ای زمان اجرای دقیق و هزینه محاسباتی در مقیاس‌های بزرگ در دستور کار این مطالعه نبود و می‌تواند بر تصمیم‌گیری عملی اثرگذار باشد.

به منظور تعمیم‌پذیری نتایج در محیط‌های واقعی، پیشنهاد می‌شود ارزیابی جامع روش‌ها بر مجموعه داده‌های میدانی بقا و قابلیت اعتماد انجام شود؛ به گونه‌ای که ناهمخوانی‌های محتمل مدل، وجود برون‌دادها و الگوهای سانسور پیچیده به صورت ساختاری در طراحی مطالعه منظور گردد. در این مسیر، می‌توان چارچوب‌های مقاوم در برابر پرت و انحراف از فرض‌های توزیعی را به کار گرفت و علاوه بر گزارش برآورد نقطه‌ای، بازه‌های اطمینان و سنجه‌های پیش‌بینی را برای سنجش عملی کارایی لحاظ کرد. برای ارتقای کارایی محاسباتی و دقت برآورد، توسعه نسخه‌های تطبیقی الگوریتم ژنتیک با تنظیم خودکار ابر پارامترها و استفاده از اپراتورهای مساله‌محور پیشنهاد می‌شود. یک مسیر سودمند، رویکردهای هیبریدی است که در آن آغازش سراسری با  $GA$  انجام و پالایش نهایی با  $EM$  یا روش‌های گرادینتی اجرا شود تا هم از جست‌وجوی سراسری و هم از همگرایی محلی سریع بهره برده شود. افزون بر این، مطالعه آمیخته‌های چندمولفه‌ای و گسترش مدل به چارچوب‌های رگرسیونی وابسته به هم‌متغیرها، همراه با معیارهای انتخاب مدل و تنظیمی نظیر اعتبارسنجی متقاطع، مسیر طبیعی توسعه این پژوهش است. برای برآورد عدم قطعیت و تصمیم‌گیری مبتنی بر مخاطره، بررسی چارچوب‌های بیزی و تقریب‌های محاسباتی کارا مانند روش‌های تقریب پسین می‌تواند مکملی موثر باشد. در کنار آن، تحلیل مبادله دقت-هزینه محاسباتی و ارائه دستورالعمل‌های عملی برای انتخاب روش بر حسب بودجه محاسباتی، اندازه نمونه و نرخ سانسور، کاربردپذیری نتایج را افزایش می‌دهد و به پژوهشگران کمک می‌کند در محیط‌های محدودیت‌دار، انتخابی آگاهانه و کارآمد داشته باشند.

## تشکر و قدردانی

نویسندگان بدین وسیله از سردبیر محترم و داوران گرامی نشریه به دلیل ارائه نظرات علمی دقیق و سازنده که نقش موثری در بهبود کیفیت و انسجام این مقاله داشته است، صمیمانه قدردانی می‌کنند.

## منابع مالی

این پژوهش هیچ‌گونه حمایت مالی مشخصی از سوی سازمان‌ها، نهادهای دولتی، موسسات خصوصی یا غیرانتفاعی دریافت نکرده است.

## تعارض با منافع

ما نویسندگان این مقاله، بدین‌وسیله اعلام می‌داریم که هیچ‌گونه تعارض منافع شخصی یا سازمانی در ارتباط با انجام و انتشار این پژوهش وجود ندارد. همچنین تایید می‌کنیم که نسخه نهایی مقاله را مطالعه و تایید کرده‌ایم، این مقاله حاصل کار اصیل ماست، پیش‌تر منتشر نشده و هم‌اکنون نیز تحت بررسی یا چاپ در نشریه دیگری نیست.

## منابع

- [1] Klein, J. P., & Moeschberger, M. L. (2006). Multivariate survival analysis. In *Survival analysis: techniques for censored and truncated data* (pp. 425–441). Springer. [https://doi.org/10.1007/0-387-21645-6\\_13](https://doi.org/10.1007/0-387-21645-6_13)
- [2] Dodson, B. (2006). *The Weibull analysis handbook*. ASQ Quality Press. <https://cir.nii.ac.jp/crid/1971993809695576965>
- [3] Kececioglu, D. B., & Wang, W. (1998). Parameter estimation for mixed-weibull distribution. *Annual reliability and maintainability symposium. 1998 proceedings. international symposium on product quality and integrity* (pp. 247–252). IEEE. <https://doi.org/10.1109/RAMS.1998.653782>
- [4] Lakshmi, R. V., & Vaidyanathan, V. S. (2016). Parameter estimation in gamma mixture model using normal-based approximation. *Journal of statistical theory and applications*, 15(1), 25–35. <https://doi.org/10.2991/jsta.2014.13.3.1>
- [5] Elmahdy, E. E., & Aboutahoun, A. W. (2013). A new approach for parameter estimation of finite Weibull mixture distributions for reliability modeling. *Applied mathematical modelling*, 37(4), 1800–1810. <https://doi.org/10.1080/02664763.2014.1000275>
- [6] Kaled, H. R., José, L. R., Vega, E., Becerra-Rozas, M., & Andrés, R. (2025). Robust metaheuristic optimization for algorithmic trading: A comparative study of optimization techniques. *Mathematics*, 14(1), 69. <https://doi.org/10.3390/math14010069>
- [7] Yan, T., Fang, K. T., & Yin, H. (2024). A novel approach for parameter estimation of mixture of two Weibull distributions in failure data modeling. *Statistics and computing*, 34(6), 221. <https://doi.org/10.1007/s11222-024-10534-1>
- [8] Jokieli-Rokita, A., & Piątek, S. (2024). Estimation of parameters and quantiles of the Weibull distribution. *Statistical papers*, 65(1), 1–18. <https://doi.org/10.1007/s00362-022-01379-9>
- [9] Karakoca, A., Erisoglu, U., & Erisoglu, M. (2015). A comparison of the parameter estimation methods for bimodal mixture Weibull distribution with complete data. *Journal of applied statistics*, 42(7), 1472–1489. <https://doi.org/10.1080/02664763.2014.1000275>
- [10] Holland, J. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press, ann arbor. [https://books.google.com/books/about/Adaptation\\_in\\_Natural\\_and\\_Artificial\\_Sys.html?id=JE5RAAAAMA\\_AJ](https://books.google.com/books/about/Adaptation_in_Natural_and_Artificial_Sys.html?id=JE5RAAAAMA_AJ)
- [11] Alhijawi, B., & Awajan, A. (2024). Genetic algorithms: Theory, genetic operators, solutions, and applications. *Evolutionary intelligence*, 17(3), 1245–1256. <https://doi.org/10.1007/s12065-023-00822-6>
- [12] Wolpert, D. H., & Macready, W. G. (2002). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1), 67–82. <https://doi.org/10.1109/4235.585893>
- [13] Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: Past, present, and future. *Multimedia tools and applications*, 80, 8091–8126. <https://link.springer.com/article/10.1007/s11042-020-10139-6>
- [14] Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms* John Wiley & Sons. John Wiley & Sons. <https://www.wiley.com/en-us/Multi-Objective+Optimization+using+Evolutionary+Algorithms-p-9780471873396>

- 
- [15] Alexakis, K., Benekis, V., Kokkinakos, P., & Askounis, D. (2025). Genetic algorithm-based multi-objective optimisation for energy-efficient building retrofitting: A systematic review. *Energy and buildings*, 328, 115216. <https://doi.org/10.1016/j.enbuild.2024.115216>
- [16] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series b (methodological)*, 39(1), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- [17] McLachlan, G. J., & Krishnan, T. (2008). *The EM algorithm and extensions*. John Wiley & Sons. <https://www.wiley.com/en-br/The+EM+Algorithm+and+Extensions%2C+2nd+Edition-p-9780471201700>
- [18] Tekeli, E., & Yüksel, G. (2022). Estimating the parameters of twofold Weibull mixture model in right-censored reliability data by using genetic algorithm. *Communications in statistics-simulation and computation*, 51(11), 6621–6634. <https://doi.org/10.1080/03610918.2020.1808681>