

## کاربرد الگوریتم نیوتن رافسون به‌عنوان یک روش عددی در برآورد بیش‌ترین درست‌نمایی

عباس پرچمی\*، مجید دوست محمدی، ماشالله ماشین چی

گروه آمار، دانشکده ریاضی و کامپیوتر، دانشگاه شهید باهنر کرمان، کرمان، ایران.

### چکیده

روش نیوتن که مشهور به الگوریتم نیوتن رافسون نیز می‌باشد، به‌عنوان یکی از کارآمدترین روش‌های عددی در ریاضیات، برای تقریب زدن ریشه‌ی معادلات غیر خطی روی میدان اعداد حقیقی شناخته شده است. در این مقاله پس از بیان و تفسیر این روش، به یکی از پرکاربردترین موارد استفاده‌ی آن در علم آمار، یعنی برآورد پارامترهای مجهول جامعه به‌روش بیش‌ترین درست‌نمایی، می‌پردازیم. جهت تسهیل در انتقال مفاهیم، مطالب این مقاله با چندین مثال عددی مختلف و برنامه‌های رایانه‌ای آن‌ها همراه شده است.

**واژه‌های کلیدی:** روش‌های عددی، الگوریتم نیوتن رافسون، برآورد بیش‌ترین درست‌نمایی، توزیع کوشی، توزیع وایبل.

پذیرش: ۱۳۹۷/۹/۲۱

اصلاح: ۱۳۹۷/۷/۱۸

دریافت: ۱۳۹۷/۳/۳۰

### ۱ - مقدمه

به‌طور کلی، الگوریتم نیوتن رافسون برای به‌دست آوردن جواب یک معادله‌ی غیرخطی یا یک دستگاه معادلات غیرخطی استفاده می‌شود. گرچه امروزه ارائه‌ی اولیه و مقدماتی روش عددی نیوتن رافسون به ایزاک نیوتن<sup>۱</sup> در سال ۱۷۱۱ میلادی منسوب شده است، ولی احتمالاً روش نیوتن از روشی با دقت کمتر که قبلاً به‌وسیله‌ی فراکوینز ویت<sup>۲</sup> مطرح شده بود، الهام گرفته شده است، زیرا سال‌ها پیش‌تر شرف‌الدین طوسی، ریاضی‌دان برجسته قرن ششم هجری قمری در ایران، به اهمیت و ضرورت استفاده از روش ویت در حل معادله  $x^p - N = 0$  اشاره کرده بود (وای‌ما، ۱۹۹۵). به هر حال ژوزف رافسون<sup>۳</sup> در اواخر قرن شانزدهم میلادی، با نگاهی مجدد به روش نیوتن از دیدگاه جبری، توانست محاسبات این روش را بسیار ساده‌تر سازد (<https://en.wikipedia.org>). در بسیاری از مسائل آماری نیز، برآورد پارامترهای مجهول جامعه با استفاده از روش بیش‌ترین درست‌نمایی نیازمند حل یک یا چند معادله‌ی پیچیده‌ی غیرخطی می‌باشد. به‌دلیل مشکل بودن حل هم‌زمان چنین معادلاتی، معمولاً برآورد را به‌روش‌های عددی مختلفی، از قبیل روش نیوتن رافسون، روش انتظار بیشینه<sup>۴</sup>، روش دونیمه‌سازی<sup>۵</sup> و روش فراز هم‌دسته‌سازی<sup>۶</sup>، تقریب می‌زنند (پرچمی، ۱۳۸۳؛ بیگل و داکسام، ۲۰۱۵).

<sup>1</sup> Isaac Newton

<sup>2</sup> François Viète

<sup>3</sup> Joseph Raphson

<sup>4</sup> Expectation Maximization (EM)

<sup>5</sup> Bisection

<sup>6</sup> Coordinate Ascent



## ۲- تشریح الگوریتم نیوتن رافسون

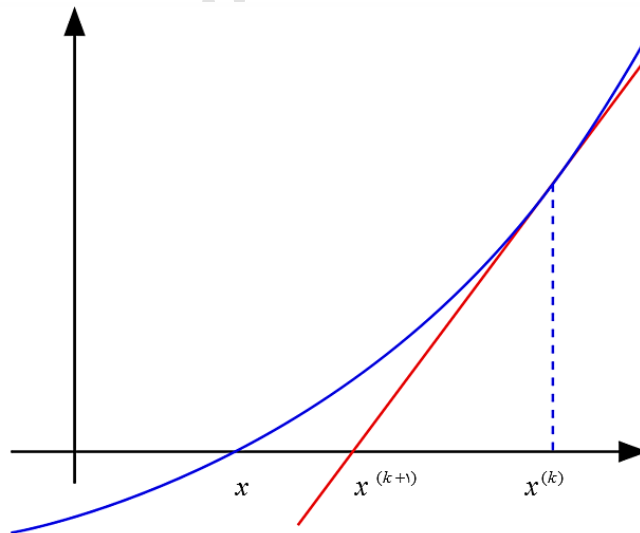
در سراسر این مقاله فرض می‌کنیم که  $f : [a, b] \rightarrow R$  با ضابطه‌ی  $f(x)$  یک تابع مشتق پذیر است که  $a, b \in R$  بر حسب مورد، اختیار می‌شوند که در آن  $R$  مجموعه‌ی اعداد حقیقی است. ایده‌ی اصلی در روش نیوتن رافسون به این صورت است که به منظور یافتن ریشه‌ی معادله  $f(x) = 0$ ، الگوریتم با یک جواب آغازین نه چندان دور از جواب واقعی، شروع می‌شود و سپس ریشه‌ی  $f(x) = 0$  به وسیله‌ی خط مماس همان نقطه بر منحنی  $f$ ، تخمین زده می‌شود. همان‌گونه که در شکل ۱ مشهود است، نقطه‌ی برخورد این خط مماس با محور  $x$  ها تقریبی بهتر برای ریشه واقعی  $f(x) = 0$ ، نسبت به نقطه‌ی آغازین می‌باشد.

برای بهبود جواب  $x^{(k)}$  در  $k$  امین مرحله از الگوریتم نیوتن رافسون و به دست آوردن  $x^{(k+1)}$  به عنوان تقریبی دقیق‌تر از ریشه‌ی واقعی معادله، از مفهوم اصلی مشتق به صورت زیر استفاده می‌شود:

$$f'(x^{(k)}) = \frac{\Delta Y}{\Delta X} = \frac{f(x^{(k)}) - 0}{x^{(k)} - x^{(k+1)}}$$

و در نتیجه

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})} \quad (1)$$



شکل ۱- عملکرد هندسی  $k$  امین مرحله از الگوریتم نیوتن رافسون جهت بهبود تقریب  $x^{(k+1)}$  نسبت به  $x^{(k)}$ .

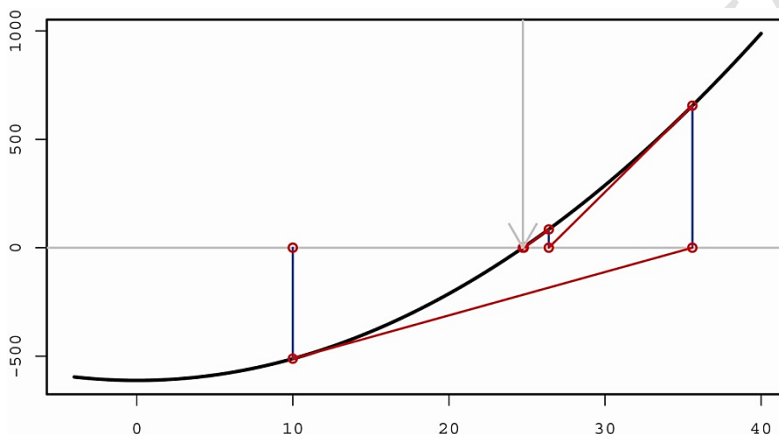
کار تقریب با جواب آغازین  $x^{(0)}$  شروع و با توجه به رابطه‌ی بازگشتی (۱) در هر مرحله از الگوریتم به ازای  $k = 1, 2, 3, \dots$  تقریب دقیق‌تری ارائه خواهد شد. این الگوریتم تا جایی ادامه می‌یابد که مقدار  $\left| \frac{f(x^{(k)})}{f'(x^{(k)})} \right|$ ، یا به طور معادل مقدار  $|f(x^{(k)})|$ ، به اندازه‌ی کافی کوچک شود و در نتیجه بتوان ادعا کرد که الگوریتم به همگرایی رسیده است (توماس و فینی، ۱۹۸۴).



مثال ۱- مسئله‌ی یافتن جذر یک عدد حقیقی را در نظر بگیرید. گرچه روش‌های متعددی، توانایی حل این مسئله را دارند، ولی در این جا می‌خواهیم از الگوریتم نیوتن رافسون جهت حل این مسئله یاری بگیریم. به طور مثال، برای یافتن جذر ۶۱۲، باید معادله‌ی  $x^2 = 612$  را حل نمود. بنابراین، تابعی که در روش نیوتن رافسون به کار برده می‌شود به صورت  $f(x) = x^2 - 612$  با مشتق  $f'(x) = 2x$  است. اگر به عنوان جواب آغازین از  $x^{(0)} = 10$  استفاده شود، آنگاه با افزایش  $k$  در مراحل زیر عمل تقریب زدن عدد ۶۱۲ دقیق‌تر صورت می‌گیرد.

$$\begin{aligned} x^{(1)} &= x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})} = 10 - \frac{10^2 - 612}{2 \times 10} = 35/06, \\ x^{(2)} &= x^{(1)} - \frac{f(x^{(1)})}{f'(x^{(1)})} = 35/06 - \frac{35/06^2 - 612}{2 \times 35/06} = 26/3955056, \\ x^{(3)} &= \vdots = \vdots = 24/7906355, \\ x^{(4)} &= \vdots = \vdots = 24/7386883, \\ x^{(5)} &= \vdots = \vdots = 24/7386338. \end{aligned}$$

که در آن‌ها، زیر ارقامی که به طور صحیح در هر مرحله تقریب زده شده، خط کشیده شده است (<https://en.wikipedia.org>).

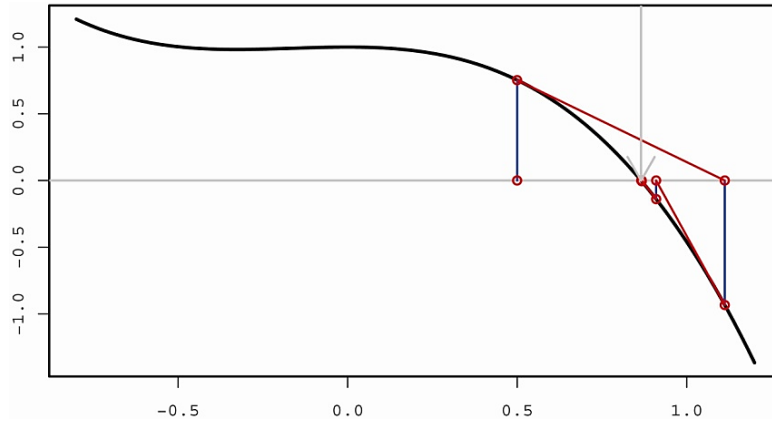


شکل ۲- روند همگرایی الگوریتم برای یافتن ریشه‌ی تابع  $f(x) = x^2 - 612$  به ازای نقطه‌ی آغازین  $x^{(0)} = 10$  در مثال ۱.

مثال ۲- برای یافتن ریشه‌ی مثبت معادله  $\cos(x) = x^2$ ، در الگوریتم نیوتن رافسون  $f(x) = \cos(x) - x^2$  در نظر گرفته می‌شود. از آن جا که به ازای هر  $x$  داریم  $\cos(x) \leq 1$  و به ازای هر  $x > 1$  داریم  $x^2 > 1$ ، می‌توان نتیجه گرفت که ریشه‌ی معادله در بازه‌ی  $[0, 1]$  قرار دارد؛ بنابراین الگوریتم را با  $x^{(0)} = 0/5$  آغاز می‌کنیم.

$$\begin{aligned} x^{(1)} &= x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})} = 0/5 - \frac{\cos(0/5) - 0/5^2}{-\sin(0/5) - 3 \times 0/5^2} = 1/112141637097, \\ x^{(2)} &= x^{(1)} - \frac{f(x^{(1)})}{f'(x^{(1)})} = \vdots = 0/9096772693636, \\ x^{(3)} &= \vdots = \vdots = 0/867263818209, \\ x^{(4)} &= \vdots = \vdots = 0/865477135298, \\ x^{(5)} &= \vdots = \vdots = 0/865474033111, \\ x^{(6)} &= \vdots = \vdots = 0/865474033102. \end{aligned}$$

که در آن‌ها، زیر ارقامی که به طور درست و صحیح در هر مرحله تقریب زده شده، خط کشیده شده است. همان گونه که مشاهده می‌شود تعداد ارقامی که بعد از اعشار درست تقریب زده شده است، از ۲ (برای  $x^{(2)}$ ) به ۵ و سپس به ۱۰ و ۱۲ افزایش یافته است که این موضوع ناشی از سرعت بالای همگرایی در این الگوریتم می‌باشد (<https://en.wikipedia.org>).



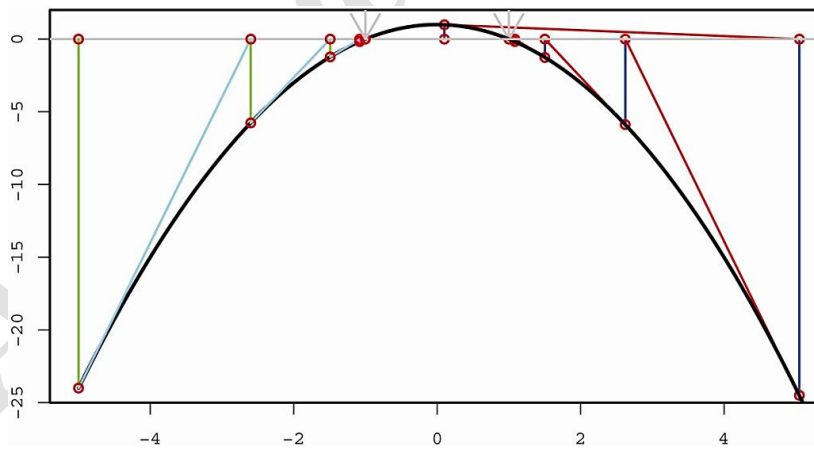
شکل ۳- روند همگرایی الگوریتم برای یافتن ریشه‌ی تابع  $f(x) = \cos(x) - x^2$  به‌ازای نقطه‌ی آغازین  $x^{(0)} = 0.5$  در مثال ۲.

### ۳- معادلات و نقاط آغازین نامناسب

روش نیوتن رافسون با توجه به شکل تابع  $f$  و نقطه‌ی آغازین، هیچ‌گونه تضمینی جهت همگرایی صحیح به ریشه‌ی تابع ندارد. در این بخش چند مثال خاص از رفتار نامناسب این الگوریتم مطرح شده است (<https://en.wikipedia.org>).

مثال ۳- می‌دانیم که تابع  $f$  با ضابطه‌ی  $f(x) = 1 - x^2$  در  $x = 0$  دارای ماکزیمم بوده و  $x = \pm 1$  ریشه‌های معادله‌ی  $f(x) = 0$  می‌باشند. اگر نقطه‌ی آغازین  $x^{(0)} = 0$  که به‌ازای آن  $f'(x) = 0$  برای شروع الگوریتم انتخاب گردد، آن‌گاه در اولین

مرحله‌ی الگوریتم مقدار  $x^{(1)} = x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})} = 0 - \frac{1}{0}$  تعریف نشده است.



شکل ۴- روند همگرایی الگوریتم برای یافتن ریشه‌های تابع  $f(x) = 1 - x^2$  به‌ازای نقاط آغازین  $x^{(0)} = 0/1$  (برای یافتن ریشه مثبت) و  $x^{(0)} = -5$  (برای یافتن ریشه منفی) در مثال ۳.

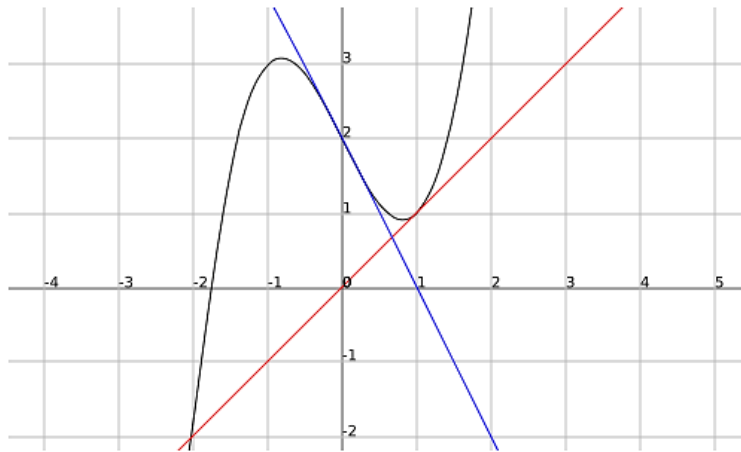
ممکن است برخی از نقاط آغازین به‌ازای بعضی از معادلات، داخل یک دور نامحدود بیافتند که از همگرایی به ریشه‌ی واقعی معادله جلوگیری کند.

مثال ۴- نقطه‌ی آغازین  $x^{(0)} = 0$  را برای حل معادله غیرخطی  $f(x) = x^3 - 2x + 2 = 0$  در نظر بگیرید. در این حالت

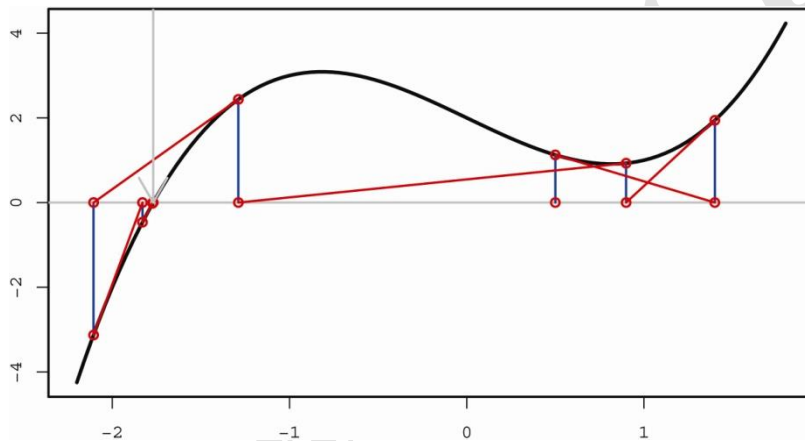
$$x^{(k+1)} = x^{(k)} - \frac{x^3 - 2x + 2}{3x^2 - 2} \Big|_{x=x^{(k)}}$$

به صفر یعنی مرحله‌ی آغازین برمی‌گردیم. بنابراین جواب‌های الگوریتم هم چنان دو عدد ۰ و ۱ باقی خواهند ماند که این امر از همگرایی

الگوریتم به ریشه‌ی واقعی ممانعت می‌کند. بدیهی است که این مشکل با اختیار نقطه آغازینی به جز نقاط ۰ و ۱ (مثلاً  $x^{(0)} = 0.5$ ) رفع می‌گردد (شکل ۶ را ببینید).



شکل ۵- خطوط مماس بر تابع  $x^2 - 2x + 2$  به ترتیب در نقاط ۰ و ۱ محور  $x$  ها را قطع می‌کند.



شکل ۶- روند همگرایی الگوریتم برای یافتن ریشه‌ی تابع  $f(x) = x^2 - 2x + 2$  به ازای نقطه آغازین  $x^{(0)} = 0.5$  در مثال ۴.

ماهیت بعضی از توابع به گونه‌ای است که روش نیوتن رافسون به ازای هر نقطه‌ی آغازینی واگراست مگر آن‌که در ابتدا، جواب آغازین دقیقاً همان ریشه‌ی معادله باشد که این خود امری نامحتمل است.

مثال ۵- معادله‌ی  $\sqrt[3]{x} = 0$  را که دارای جواب  $x = 0$  است، در نظر بگیرید. در هر مرحله از حل معادله  $f(x) = \sqrt[3]{x}$  به وسیله الگوریتم نیوتن رافسون و به ازای هر نقطه  $x^{(k)}$  داریم:

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})} = x^{(k)} - \frac{(x^{(k)})^{1/3}}{(x^{(k)})^{-2/3}/3} = -2x^{(k)}.$$

بدیهی است که در هر مرحله، الگوریتم جواب را به یک طرف دیگر محور، با فاصله‌ای بیش‌تر از جواب واقعی نسبت به مرحله‌ی قبلی، پرتات می‌کند. در حقیقت، جواب الگوریتم نیوتن رافسون برای تابع  $f(x) = |x|^\alpha$  به ازای هر  $0 < \alpha < 0.5$ ، همگرا نبوده و به سمت بی‌نهایت میل می‌کند.



## ۴- کاربرد الگوریتم نیوتن رافسون در برآورد بیشترین درست‌نمایی<sup>۱</sup>

در این بخش به یکی از مهم‌ترین کاربردهای روش نیوتن رافسون در آمار، یعنی محاسبه‌ی برآورد بیشترین درست‌نمایی برای پارامترهای مجهول جامعه می‌پردازیم. ابتدا ساده‌ترین حالت ممکن، یعنی جامعه‌ای با یک پارامتر مجهول را در نظر می‌گیریم و برای آن یک مثال ارائه می‌دهیم، سپس روش را به حالت چند پارامتری تعمیم می‌دهیم.

### ۴-۱ برآورد تک پارامتر جامعه

نمونه‌ی تصادفی  $X_1, \dots, X_n$  را از توزیعی با تابع جرم احتمال یا تابع چگالی احتمال  $f(x; \theta)$  در نظر می‌گیریم که در آن  $\theta$  پارامتر مجهول جامعه می‌باشد. فرض کنید مقدار مشاهده‌شده  $\underline{X} = (X_1, \dots, X_n)$  را با  $\underline{x}$  و هم‌چنین تابع درست‌نمایی و لگاریتم آن را به ترتیب با  $L(\theta; \underline{x})$  و  $l(\theta; \underline{x}) = \sum_{i=1}^n \ln f(x_i; \theta)$  نشان می‌دهیم. منظور از برآورد پارامتر مجهول  $\theta$  به روش بیشترین درست‌نمایی، یافتن  $\theta$  ای در فضای پارامتر است، به طوری که به ازای آن  $\theta$  تابع درست‌نمایی بیشینه شود. به بیان دیگر، اگر  $\hat{\theta}$  برآورد بیشترین درست‌نمایی برای  $\theta$  باشد، آن‌گاه  $l'(\hat{\theta}) = 0$ . اگر برآورد بیشترین درست‌نمایی  $\theta$  که از این پس آن را با  $MLE(\theta)$  نشان می‌دهیم موجود و یکتا باشد، آن‌گاه با جای‌گذاری  $f = l'$  و  $x = \hat{\theta}$  در فرمول بازگشتی (۱) از الگوریتم نیوتن رافسون داریم:

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - (l''(\hat{\theta}^{(k)}))^{-1} l'(\hat{\theta}^{(k)}). \quad (2)$$

بدیهی است که برای شروع الگوریتم به یک برآورد آغازین از  $\theta$  نیاز داریم. یکی از روش‌های یافتن مقدار برآورد اولیه، استفاده از برآوردگرهای دیگر مانند برآورد به روش گشتاوری است. اگر برآورد آغازین به اندازه‌ی کافی برای  $\theta$  خوب باشد، آن‌گاه برآوردگر حاصل در اولین گام از الگوریتم دارای خاصیت بزرگ نمونه‌ای برآوردگرهای بیشترین درست‌نمایی است (بیگل و داکسام، ۲۰۱۵)؛ به عبارت دیگر

$$\hat{\theta}^{(1)} = \hat{\theta}^{(0)} - (l''(\hat{\theta}^{(0)}))^{-1} l'(\hat{\theta}^{(0)}) \xrightarrow[n \rightarrow \infty]{D} N\left(\theta, \frac{1}{I_{\underline{X}}(\theta)}\right),$$

که در آن مبتنی بر نمونه‌ی تصادفی  $\underline{X} = (X_1, \dots, X_n)$  و  $l(\theta) = \ln L(\theta; \underline{X})$ ،  $I_{\underline{X}}(\theta) = E\left[\left[\frac{\partial}{\partial \theta} \ln L(\theta; \underline{X})\right]^T\right]$  اطلاع فیشر  $\underline{X}$  در مورد  $\theta$  می‌باشد.

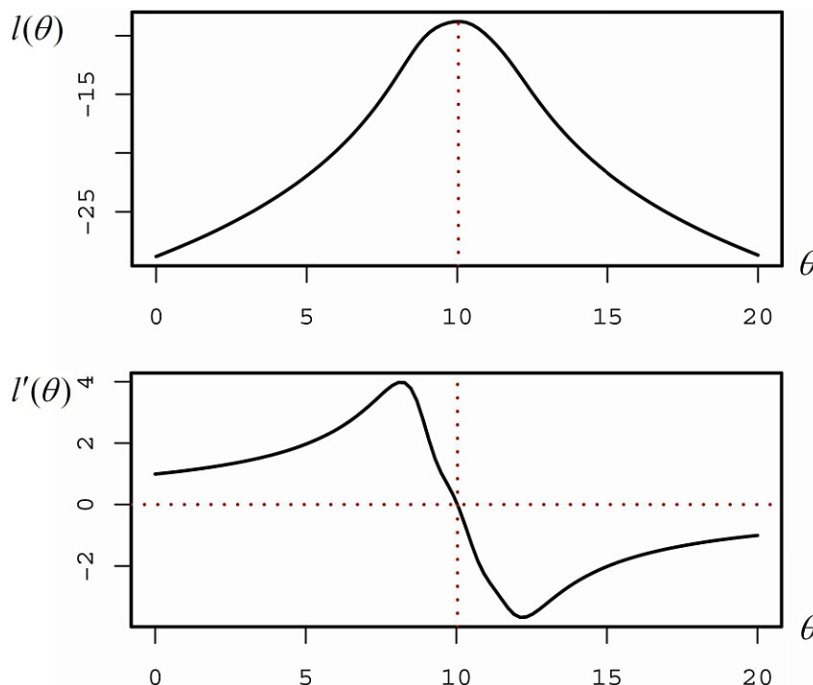
مثال ۶- (برآورد پارامتر در توزیع کوشی تک پارامتری). فرض کنید  $\underline{X} = (X_1, \dots, X_n)$  یک نمونه‌ی تصادفی از توزیع کوشی، با تابع چگالی احتمال زیر باشد:

$$f(x; \theta) = \frac{1}{\pi(1+(x-\theta)^2)}, \quad -\infty < \theta < +\infty.$$

در این مثال فرض کنید نمونه‌ی تصادفی  $x_1 = 10/575140$ ،  $x_2 = 11/624988$ ،  $x_3 = 1/985913$ ،  $x_4 = 1/935000$ ،  $x_5 = 10/168302$  از جامعه‌ای با توزیع کوشی مشاهده شده است. براساس این نمونه‌ی مشاهده‌شده، توابع  $l$  و  $l'$  نسبت به  $\theta$  در شکل ۷ رسم شده است.

<sup>1</sup> Maximum Likelihood Estimate (MLE)





شکل ۷- تابع لگاریتم درست‌نمایی (تصویر بالا) و مشتق آن (تصویر پایین) در مثال ۶.

با توجه به تقارن توزیع کوشی حول  $\theta$ ، می‌توان هرکدام از میانگین و یا میانه‌ی نمونه را به‌عنوان نقطه‌ی آغازین  $\hat{\theta}^{(0)}$  در نظر گرفت. ولی از آن‌جا که امید ریاضی  $X_i$ ها در توزیع کوشی وجود ندارد، از میانه‌ی نمونه به‌عنوان برآورد اولیه استفاده می‌کنیم. با جای‌گذاری شده و به‌کمک نقطه‌ی آغازین  $\hat{\theta}^{(0)}$  مراحل هم‌گرایی الگوریتم در جدول ۱ درج شده است. توجه داشته باشید که حل دقیق معادله  $l'(\hat{\theta}) = 0$  به منظور یافتن  $MLE(\theta)$  در این مثال ممکن نیست و به‌عنوان روشی جای‌گزین، برای یافتن  $MLE(\theta)$  می‌توان از روش‌های عددی استفاده کرد. با توجه به تصویر پایین شکل ۷، انتخاب نقطه‌ی آغازین، بحرانی می‌باشد، زیرا اگر  $\hat{\theta}^{(0)}$  کم‌تر از  $8/74$  و بیش‌تر از  $11/84$  انتخاب شود، دنباله‌ی  $\{\hat{\theta}^{(k)}\}$  همگرا می‌گردد. برای مشاهده‌ی برنامه‌ی رایانه‌ای الگوریتم نیوتن رافسون این‌مثال، قسمت الف از بخش ضمیمه را مشاهده نمایید.

جدول ۱- مراحل هم‌گرایی الگوریتم در مثال ۶ برای برآورد پارامتر توزیع کوشی.

$k$	$\hat{\theta}^{(k)}$	$l(\hat{\theta}^{(k)})$	$L(\hat{\theta}^{(k)}, \mathbf{x})$
۰	۱۰٫۱۶۸۳	۸٫۸۱۴۵۸۹-	۰٫۰۰۰۱۴۸۵۴
۱	۱۰٫۰۴۴۷	۸٫۷۹۳۲۷۸-	۰٫۰۰۰۱۵۱۷۴
۲	۱۰٫۰۳۷۲	۸٫۷۹۳۲۱۲-	۰٫۰۰۰۱۵۱۷۵

## ۲-۴ برآورد همزمان چند پارامتر جامعه

روش ارائه‌شده در بخش ۴-۱ را می‌توان برای حالتی که  $\theta$  برداری است، به صورت زیر تعمیم داد. اگر  $\theta = (\theta_1, \dots, \theta_p)^T$  بردار پارامترهای مجهول جامعه باشد، آن‌گاه همانند حالت تک‌پارامتری، اساس کار، فرمول بازگشتی (۲) می‌باشد که در آن

$$l'(\hat{\theta}) = \left( \frac{\partial}{\partial \theta_1} l(\hat{\theta}), \dots, \frac{\partial}{\partial \theta_p} l(\hat{\theta}) \right)^T \text{ و هم‌چنین } l''(\hat{\theta}^{(k)}) \text{ ماتریسی با درایه‌های } l''_{ij}(\hat{\theta}) = \frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\hat{\theta}) \text{ است به قسمی}$$



که  $i$  شماره سطر و  $j$  شماره ستون ماتریس می‌باشد. گفتنی است که  $H(\hat{\theta}) = l''(\hat{\theta}) = \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} l(\hat{\theta}) \right]_{p \times p}$  ماتریس هسین تابع  $l(\theta)$  مبتنی بر مشاهدات نام دارد که رابطه‌ی  $H(\hat{\theta}) = -I_{\underline{x}}(\hat{\theta})$  با ماتریس اطلاع فیشر دارد (نایت<sup>۱</sup>، ۲۰۰۰).

مثال ۷- (برآورد پارامترها در توزیع وایبل دو پارامتری). فرض کنید  $\underline{X} = (X_1, \dots, X_n)$  یک نمونه‌ی تصادفی از توزیع وایبل با تابع چگالی احتمال

$$f(x; a, b) = \begin{cases} \frac{a}{b^a} x^{a-1} \exp\left[-\left(\frac{x}{b}\right)^a\right], & \text{if } x \geq ab \\ 0, & \text{if } x < ab \end{cases}, a, b > 0.$$

و با مقدار مشاهده شده  $(0/468, 0/375, 1/858, 1/235, 0/468)$  باشد. گفتنی است این توزیع، که اغلب در تحلیل داده‌های بقا مورد استفاده قرار می‌گیرد، در حالت خاص  $a=1$  تبدیل به توزیع نمایی می‌شود. هدف، برآورد پارامتر  $\theta = (a, b)^T$  به روش بیشترین درست‌نمایی و بر پایه‌ی بردار مشاهدات  $\underline{x}$  است. در این حالت، لگاریتم تابع درست‌نمایی برابر است با

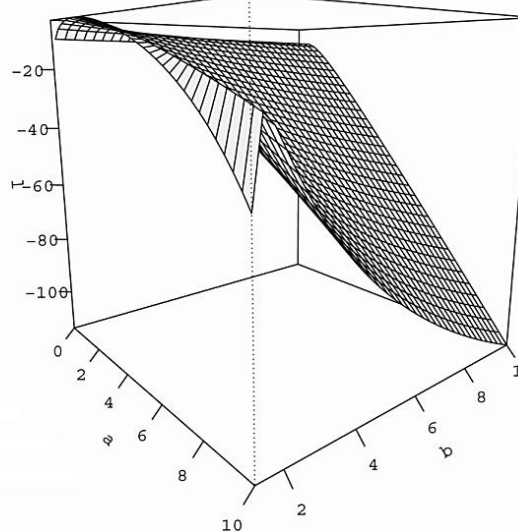
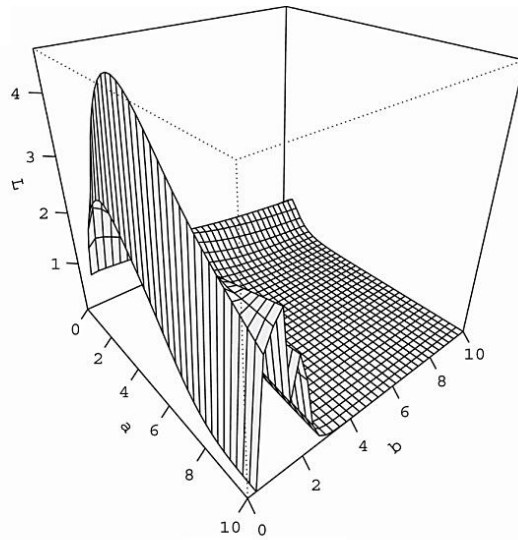
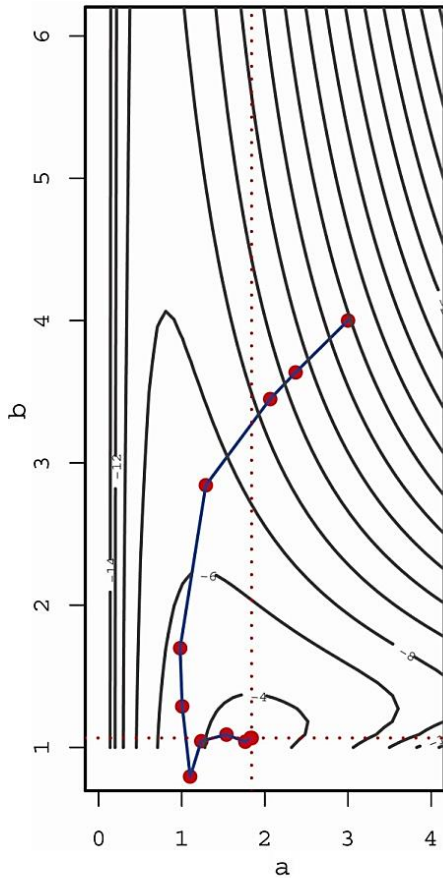
$$l(\theta) = \sum_{i=1}^n \left[ \ln a - a \ln b + (a-1) \ln x_i - \left(\frac{x_i}{b}\right)^a \right].$$

بر اساس تابع nlm، برآورد بیشترین درست‌نمایی مبتنی بر نقطه‌ی آغازین  $(\hat{a}, \hat{b})^T = (3, 4)^T$  پس از ۱۳ دور چرخش الگوریتم نیوتن رافسون، برابر  $\hat{\theta}^{(A)} = (1/84, 1/07)^T$  به دست آمده است (شکل ۸ و قسمت ب از بخش ضمیمه را ببینید). توضیح آنکه nlm یکی از توابع پکیج stats از نرم‌افزار R است که قادر به کمینه‌سازی توابع به روش نیوتن رافسون است. علاقه‌مندان جهت کسب اطلاعات و بررسی بیشتر در مورد توزیع وایبل سه پارامتری، می‌توانند به مرجع (کیو و توکاس<sup>۲</sup>، ۱۹۹۵) مراجعه نمایند.

<sup>1</sup> knight

<sup>2</sup> Qiao & Tsokos





شکل ۸- رویه تابع درست‌نمایی  $L(\theta; x)$  (تصویر بالا سمت راست)، رویه تابع لگاریتم درست‌نمایی  $l(\theta)$  (تصویر پایین سمت راست)، کانتورهای تابع لگاریتم درست‌نمایی و برآورد بیش‌ترین درست‌نمایی (تصویر سمت چپ) در مثال ۷.

## ۵- چند نکته برای استفاده صحیح از الگوریتم

- با توجه به شرط مشتق‌پذیری و در نتیجه پیوستگی تابع  $f$  در استفاده از الگوریتم نیوتن رافسون، شاید در ابتدا تصور شود که یافتن برآورد بیش‌ترین درست‌نمایی با استفاده از این الگوریتم تنها برای متغیرهای تصادفی پیوسته با تابع چگالی احتمال  $f(x; \theta)$  کاربرد دارد. ولی با توجه به بخش ۴-۱، باید توجه داشت که برای یافتن  $MLE(\theta)$  با استفاده از الگوریتم نیوتن رافسون، هدف حل معادله  $l'(\hat{\theta}) = 0$  برحسب  $\theta$  (و نه  $x$ ) بوده و از این رو، شرط مشتق‌پذیری و در نتیجه پیوستگی تابع  $l'$  شرطی بدیهی است که مشمول متغیرهای تصادفی گسسته نیز می‌باشد.
- اگر  $\sqrt{n}(\hat{\theta}^{(k)} - \theta) \xrightarrow{P} 0$  در توزیع هم‌گرا باشد، آن‌گاه  $\sqrt{n}(\hat{\theta}^{(k)} - \hat{\theta}_n)$  که در آن  $\hat{\theta}_n$  برآورد  $MLE(\theta)$  و  $\hat{\theta}^{(k)}$  برآوردگر نخستین‌گام الگوریتم به‌ازای نقطه‌ی آغازین  $\theta^{(0)}$  می‌باشد (نایت، ۲۰۰۰).
- متأسفانه وقتی جواب آغازین از ریشه‌ی اصلی تابع دور باشد، روش نیوتن رافسون ممکن است تقریب نامناسبی را ارائه دهد. از این رو، الگوریتم نیوتن رافسون بهتر است با جواب اولیه نه چندان دوری از جواب واقعی معادله شروع شود تا توانایی بهبود آن را از دست ندهد.
- با توجه به مثال ۳، اگر  $l''(\hat{\theta}^{(k)}) = 0$  باشد، آن‌گاه این روش کارایی ندارد.
- با توجه به مثال‌های ۴ و ۵، ممکن است الگوریتم نیوتن رافسون هم‌گرا نباشد.

- اگر به ازای تمام مقادیر  $\theta$  در بازه‌ای که شامل ریشه‌ی واقعی تابع می‌باشد نامساوی

$$\left| \frac{l'(\theta)l'''(\theta)}{(l'(\theta))^2} \right| < 1 \quad (3)$$

برقرار باشد، آن‌گاه به ازای هر مقدار آغازین  $\theta^{(0)}$  متعلق به این بازه، الگوریتم هم‌گراست. البته این شرط کافی است و نه لازم؛ یعنی ممکن است بازه‌ای شامل ریشه‌ی واقعی به‌گونه‌ای نتوان یافت که مقادیرش در رابطه‌ی (۳) صدق کند، اما درعین حال رویکرد نیوتن رافسون در برآورد بیش‌ترین درست‌نمایی هم‌گرا باشد (توماس و فینی، ۱۹۸۴).

- وظیفه‌ی تابع  $nlm$  کمینه‌سازی تابعی است که به‌عنوان اولین آرگومان  $nlm$  در نظر گرفته شده است. لذا در برآورد بیش‌ترین درست‌نمایی به کمک تابع  $nlm$ ، گرچه دنباله‌ی تقاطعی که طبق الگوریتم نیوتن رافسون به هم‌گرایی می‌رسند یکتا نیست و هم به نقطه‌ی آغازین و هم به تابع اکیدا نزولی برحسب تابع درست‌نمایی (مانند  $-L$ ،  $1/L$ ، یا  $-1$ ) بستگی دارد، اما نقطه‌ی برآورد شده‌ی نهایی (در صورت عدم وجود چندین ماکزیمم موضعی برای  $L$ ) یکتاست.

## ۶- نتیجه‌گیری

در این مقاله، کاربرد الگوریتم نیوتن رافسون در حل معادلات درست‌نمایی برای یافتن برآورد بیش‌ترین درست‌نمایی وقتی روش تحلیلی ریاضی برای حل آن‌ها وجود ندارد، توضیح داده شده است. به‌منظور تسهیل در انتقال مفاهیم، مطالب این مقاله به‌همراه مثال‌های متنوع، برنامه‌های رایانه‌ای و نکاتی در مورد استفاده‌ی صحیح از این الگوریتم، مطرح شده است.

## سپاس‌گزاری

بدین وسیله از زحمات آقایان دکتر محمود طاهری، مرتضی پورجعفردوین و علیرضا یآوری که نظراتشان موجب ارتقای نسخه‌ی اولیه‌ی این مقاله شده است، تشکر و قدردانی می‌گردد.

## منابع

- پرجمی، ع. (۱۳۸۳). مثال‌هایی از الگوریتم EM. اندیشه آماری، ۹ (۲)، ۵۴-۵۸.
- Bickel, P. J., & Doksum, K. A. (2015). *Mathematical statistics: basic ideas and selected topics, volume I*. CRC Press.
- Knight, K. (2000). *Mathematical statistics*. CHAPMAN & HALL/CRC: New York.
- Thomas, G. B., & Finney, R. L. (1984). *Calculus and analytic geometry*. Addison Wesley Publishing Company.
- Qiao, H., & Tsokos, C. P. (1995). Estimation of the three parameter Weibull probability distribution. *Mathematics and computers in simulation*, 39(1-2), 173-185.
- Newton's method. (n.d). Retrieved from [http://en.wikipedia.org/wiki/Newton-Raphson\\_method](http://en.wikipedia.org/wiki/Newton-Raphson_method)
- Ypma, T. J. (1995). Historical development of the Newton-Raphson method. *SIAM review*, 37(4), 531-551.



در این بخش، برنامه‌های رایانه‌ای الگوریتم نیوتن رافسون براساس نرم افزار Maple 7 برای برخی از مثال‌ها مطرح شده است.

الف) برنامه‌ی مثال ۶



```
> x(1):=10.575140:
> x(2):=11.624988:
> x(3):=8.985913:
> x(4):=8.935000:
> x(5):=10.168302:
> l:=Sum(-ln(Pi)-ln(1+((x(i)-a)^2)', i'=1..5);
> A:=diff(l,a);
> AA:=diff(A,a);
> A:=a-> sum(-(-2*x(i)+2*a)/(1+(x(i)-a)^2),i = 1 .. 5);
> AA:=a-> sum(-2*1/(1+(x(i)-a)^2)+(- 2*x(i)+2*a)^2/((1+(x(i)-a)^2)^2), i = 1.. 5);
> p:=a->a-(A(a)/AA(a));
> a_old(-1):=0:
> a_old(0):=9:
> k_a:=0:
> while abs(a_old(k_a-1)- a_old(k_a))>0.0001 do a_new:=p(a_old(k_a));
a_old(k_a+1):=a_new;
a_old(k_a):=a_old(k_a);
k_a:=k_a+1
od;
> plot(l,a=0..20);
> plot(A(a),a=0..20);
```

ب) برنامه‌ی مثال ۷

```
> x(1):=0.76818124:
> x(2):=0.37550860:
> x(3):=1.85810086:
> x(4):=1.23559542:
> x(5):=0.46788881:
> l:=Sum(ln(a)-ln(b)+(a-1)*(ln(x(i)))-(b^(-1))*((x(i))^a)', i'=1..5);
> A:=diff(l,a);
> AA:=diff(A,a);
> B:=diff(l,b);
> BB:=diff(B,b);
> AB:=diff(A,b);
> A:=(a,b)-> sum(1/a+ln(x(i))- x(i)^a*ln(x(i))/b,i = 1 .. 5);
> AA:=(a,b)-> sum(-1/(a^2)- x(i)^a*ln(x(i))^2/b,i = 1 .. 5);
> B:=(a,b)->sum(-1/b+x(i)^a/(b^2),i =1 .. 5);
> BB:=(a,b)->sum(1/(b^2)-2*x(i)^a/(b^3),i = 1 .. 5);
> AB:=(a,b)-> sum(x(i)^a*ln(x(i))/(b^2),i = 1 .. 5);
> with(linalg):
> IM:=(a,b)-> inverse(array([[AA(a,b),AB(a,b)],[AB (a,b),BB(a,b)]]));
> V:=(a,b)- array([[A(a,b)],[B(a,b)]]);
> P:=(a,b)->evalm(array([[a],[b]])-IM(a,b)&*V(a,b));
> E:=array([[10^(-3)],[10^(-3)]]);
> P(-1):=array([[0],[0]]);
> P(0):=array([[1.570193102], [.7312167594]]);
> k:=0:
> while abs(P(k-1)[1,1]-P(k)[1,1])>E[1,1] and abs(P(k-1)[2,1]-P(k)[2,1])>E[2,1]
do
P(k+1):=P(P(k)[1,1],P(k)[2,1]);
k:=k+1;
P(k-1):=P(k-1)
od;
> plot3d(((a/b)^n)*((product(x(i), i=1..5))^(a-1))*exp(-(b^(-1)) *(sum('((x(i))^a)', i'=1..5))),a=0..5,b=0..2).
```