Journal of Decisions and Operational Research

**نشریه تصمیم‌گیری و تحقیق در عملیات**

**دوره ۳، شماره (۴)، زمستان ۱۳۹۷**

www.journal-dmor.ir

نشریه تصمیم‌گیری و تحقیق در عملیات، دوره ۳، شماره (۴)، زمستان ۱۳۹۷، صفحه: ۴۰۲-۴۱۲

# کاربرد الگوریتم نیوتن رافسون به‌عنوان یک‌روش عددی در برآورد بیش‌ترین درست‌نمایی

**عباس پرچمی*، مجید دوست محمدی، ماشاالله ماشین‌چی**

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

**چکیده**

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

دریافت: ۱۳۹۷/۳/۳۰ اصلاح: ۱۳۹۷/۷/۱۸ پذیرش: ۱۳۹۷/۹/۲۱

## 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 ایزاك نیوتن[1] در سال ۱۷۱۱ میلادی منصوب شده است، ولی احتمالاً روش نیوتن از روشی با دقت کمتر که قبلاً به‌وسیله‌ی فراکویز ویت[2] 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 هرحال ژوزف رافسون[3] 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، ازقبیل روش نیوتن رافسون، روش انتظار بیشینه[4]، روش دونیمه‌سازی[5] و روش فراز هم‌دسته‌سازی[6]، تقریب می‌زنند (پرچمی، ۱۳۸۳؛ بیکل و داکسام، ۲۰۱۵).

---

[1] Isaac Newton

[2] François Viète

[3] Joseph Raphson

[4] Expectation Maximization (EM)

[5] Bisection

[6] Coordinate Ascent

*عهده‌دار مکاتبات: parchami@uk.ac.ir

در بخش ۲ این مقاله، به‌تشریح الگوریتم نیوتن رافسون و ارائه‌ی چندین مثال می‌پردازیم. سپس در بخش ۳ ضمن ارائه‌ی چند مثال مختلف به بررسی پیرامون حالت‌های خاصی که این الگوریتم کارایی خود را ازدست می‌دهد، می‌پردازیم. برآورد پارامترهای مجهول یک جامعه‌ی آماری به‌روش بیش‌ترین درست‌نمایی و بر پایه‌ی الگوریتم نیوتن رافسون در بخش ۴ مطرح شده است. یک جمع بندی کلی که شامل چندین نکته برای استفاده‌ی صحیح‌تر از این‌الگوریتم در برآوردیابی می‌باشد در بخش ۵ ارائه شده است.

## ۲- تشریح الگوریتم نیوتن رافسون

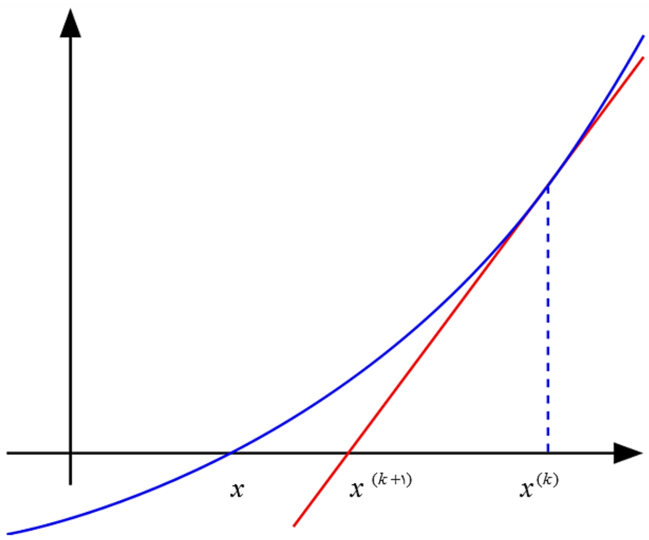
در سراسر این‌مقاله فرض می‌کنیم که $f:[a,b]\to R$ با ضابطه‌ی $f(x)$ یک تابع مشتق پذیر است که $a,b\in R$ بر حسب مورد، اختیار می‌شوند که در آن $R$ مجموعه‌ی اعداد حقیقی است. ایده‌ی اصلی در روش نیوتن رافسون به این‌صورت است که به‌منظور یافتن ریشه‌ی معادله $f(x)=0$، الگوریتم با یک جواب آغازین نه چندان دور از جواب واقعی، شروع می‌شود و سپس ریشه‌ی $f(x)=0$ به‌وسیله‌ی 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 به‌دست آوردن $x^{(k+1)}$ به‌عنوان تقریبی دقیق‌تر از ریشه‌ی واقعی معادله، از مفهوم اصلی مشتق به‌صورت زیر استفاده می‌شود:

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

و درنتیجه

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




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

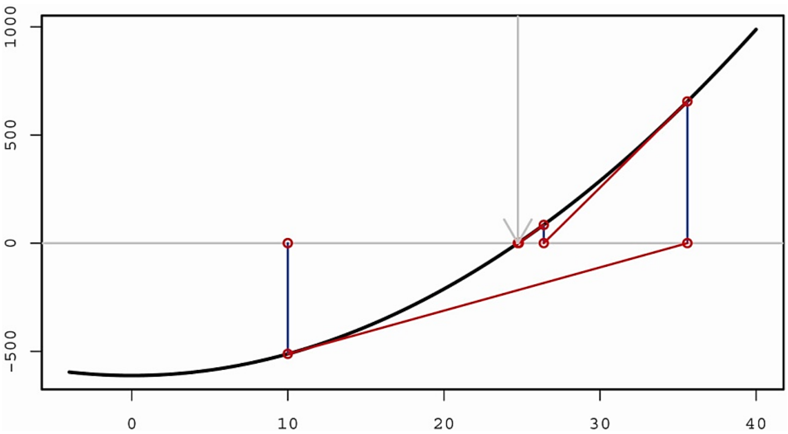
کار تقریب با جواب آغازین $x^{(0)}$ شروع و با توجه به رابطه‌ی بازگشتی (۱) در هر مرحله از الگوریتم به‌ازای $k=1,2,3,\ldots$ تقریب دقیق‌تری ارائه خواهد شد. این الگوریتم تا جایی ادامه می‌یابد که مقدار $\left|\frac{f(x^{(k)})}{f'(x^{(k)})}\right|$، یا به‌طور معادل مقدار $\left|f(x^{(k)})\right|$، به‌اندازه‌ی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

نشریه تصمیم‌گیری و تحقیق در عملیات، دوره ۳، شماره (۴)، زمستان ۱۳۹۷، صفحه: ۴۱۲-۴۰۲

**مثال ۱-** مسئله‌ی یافتن جذر یک عدد حقیقی را درنظر بگیرید. گرچه روش‌های متعددی، توانایی حل این مسئله را دارند، ولی در این‌جا می‌خواهیم از الگوریتم نیوتن رافسون جهت حل این‌مسئله یاری بگیریم. به‌طور مثال، برای یافتن جذر ۶۱۲، باید معادله‌ی $x^{۲} = ۶۱۲$ را حل نمود. بنابراین، تابعی که در روش نیوتن رافسون به‌کار برده می‌شود به‌صورت $f\,(x) = x^{۲} - ۶۱۲$ با مشتق $f\,'(x) = ۲x$ است. اگر به‌عنوان جواب آغازین از $x^{(۰)} = ۱۰$ استفاده شود، آن‌گاه با افزایش $k$ در مراحل زیر عمل تقریب زدن عدد ۶۱۲ دقیق‌تر صورت می‌گیرد.

$$
\begin{aligned}
x^{(۱)} &= x^{(۰)} - \frac{f\,(x^{(۰)})}{f\,'(x^{(۰)})} = ۱۰ - \frac{۱۰^{۲} - ۶۱۲}{۲\times ۱۰} = ۳۵/۰۶،\\
x^{(۲)} &= x^{(۱)} - \frac{f\,(x^{(۱)})}{f\,'(x^{(۱)})} = ۳۵/۰۶ - \frac{۳۵/۰۶^{۲} - ۶۱۲}{۲\times ۳۵/۰۶} = \underline{۲}۶/۳۹۵۵۰۵۶،\\
x^{(۳)} &= \quad \vdots \quad = \quad \vdots \quad = \underline{۲۴/۷}۹۰۶۳۵۵،\\
x^{(۴)} &= \quad \vdots \quad = \quad \vdots \quad = \underline{۲۴/۷۳۸۶}۸۸۳،\\
x^{(۵)} &= \quad \vdots \quad = \quad \vdots \quad = \underline{۲۴/۷۳۸۶۳۳۸}،
\end{aligned}
$$

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

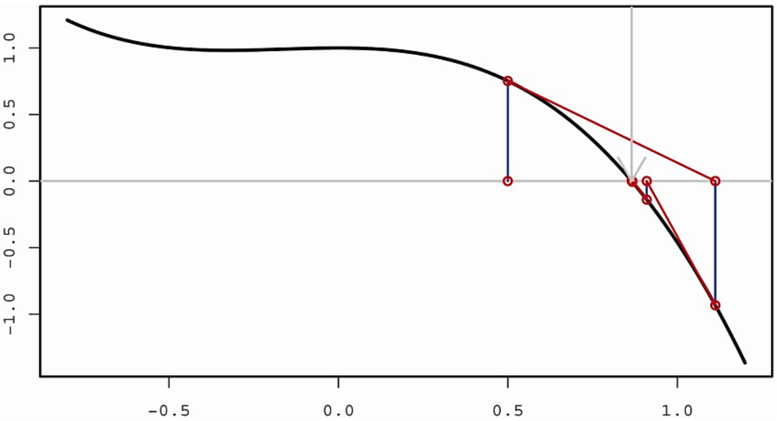



**شکل ۲- روند همگرایی الگوریتم برای یافتن ریشه‌ی تابع $f\,(x) = x^{۲} - ۶۱۲$ به‌ازای نقطه‌ی آغازین $x^{(۰)} = ۱۰$ در مثال ۱.**

**مثال ۲-** برای یافتن ریشه‌ی مثبت معادله $\cos(x) = x^{۳}$، در الگوریتم نیوتن رافسون $f\,(x) = \cos(x) - x^{۳}$ درنظر گرفته می‌شود. از آن‌جا که به‌ازای هر $x$ داریم $\cos(x) \leq ۱$ و به‌ازای هر $x > ۱$ داریم $x^{۳} > ۱$، می‌توان نتیجه گرفت که ریشه‌ی معادله در بازه‌ی $[۰,۱]$ قراردارد؛ بنابراین الگوریتم را با $x^{(۰)} = ۰/۵$ آغاز می‌کنیم.

$$
\begin{aligned}
x^{(۱)} &= x^{(۰)} - \frac{f\,(x^{(۰)})}{f\,'(x^{(۰)})} = ۰/۵ - \frac{\cos(۰/۵) - ۰/۵^{۳}}{-\sin(۰/۵) - ۳\times ۰/۵^{۲}} = ۱/۱۱۲۱۴۱۶۳۷۰۹۷،\\
x^{(۲)} &= x^{(۱)} - \frac{f\,(x^{(۱)})}{f\,'(x^{(۱)})} = \quad \vdots \quad = \underline{۰}/۹۰۹۶۷۷۲۶۹۳۶۳۶،\\
x^{(۳)} &= \quad \vdots \quad = \quad \vdots \quad = \underline{۰/۸۶}۷۲۶۳۸۱۸۲۰۹،\\
x^{(۴)} &= \quad \vdots \quad = \quad \vdots \quad = \underline{۰/۸۶۵۴۷}۷۱۳۵۲۹۸،\\
x^{(۵)} &= \quad \vdots \quad = \quad \vdots \quad = \underline{۰/۸۶۵۴۷۴۰۳۳۱۱}۱،\\
x^{(۶)} &= \quad \vdots \quad = \quad \vdots \quad = \underline{۰/۸۶۵۴۷۴۰۳۳۱۰۲}،
\end{aligned}
$$

که در آن‌ها، زیر ارقامی که به‌طور درست و صحیح در هر مرحله تقریب زده شده، خط کشیده شده است. همان‌گونه که مشاهده می‌شود تعداد ارقامی که بعد از اعشار درست تقریب زده شده است، از ۲ (برای $x^{(۳)}$) 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).

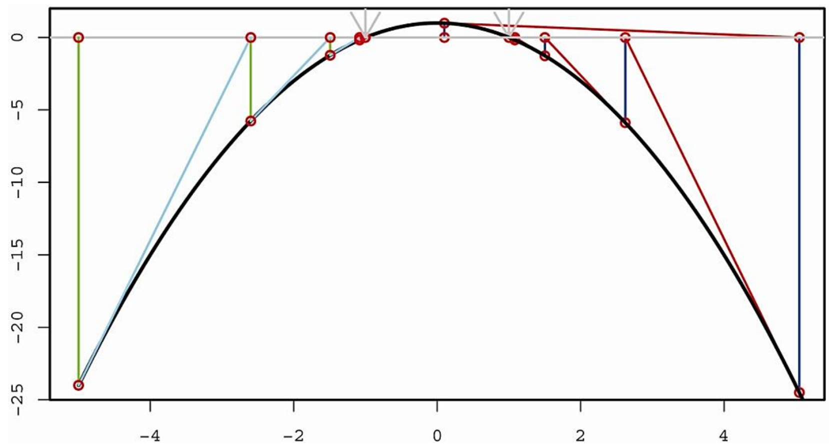



**شکل ۳- روند همگرایی الگوریتم برای یافتن ریشه‌ی تابع $f(x)=\cos(x)-x^{۳}$ به‌ازای نقطه‌ی آغازین $x^{(۰)}=۰/۵$ در مثال ۲.**

## 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 $f(x)=۱-x^{۲}$ در $x=۰$ دارای ماکزیمم بوده و $x=\pm۱$ ریشه‌های معادله‌ی $f(x)=۰$ می‌باشند. اگر نقطه‌ی آغازین $x^{(۰)}=۰$ که به‌ازای آن $f'(x)=۰$ برای شروع الگوریتم انتخاب گردد، آن‌گاه در اولین مرحله‌ی الگوریتم مقدار $x^{(۱)}=x^{(۰)}-\frac{f(x^{(۰)})}{f'(x^{(۰)})}=۰-\frac{۱}{۰}$ تعریف نشده است.



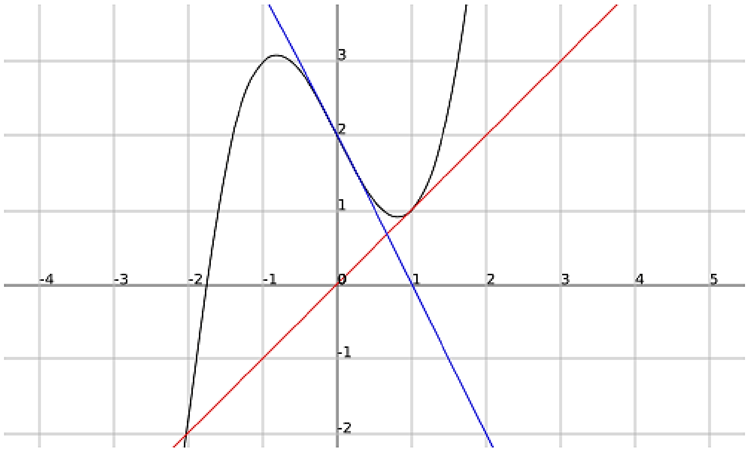

**شکل ۴- روند همگرایی الگوریتم برای یافتن ریشه‌های تابع $f(x)=۱-x^{۲}$ به‌ازای نقاط آغازین $x^{(۰)}=۰/۱$ (برای یافتن ریشه مثبت) و $x^{(۰)}=-۵$ (برای یافتن ریشه منفی) در مثال ۳.**

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

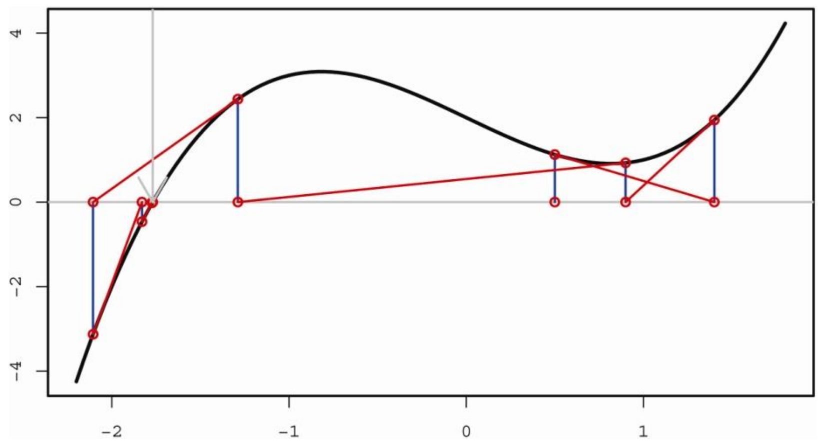
**مثال ۴-** نقطه‌ی آغازین $x^{(۰)}=۰$ را برای حل معادله غیرخطی $f(x)=x^{۳}-۲x+۲=۰$ درنظر بگیرید. در این‌حالت $x^{(k+1)}=x^{(k)}-\left.\frac{x^{۳}-۲x+۲}{۳x^{۲}-۲}\right|_{x=x^{(k)}}$ و لذا همانند شکل ۵، در نخستین مرحله‌ی الگوریتم $x^{(۱)}=۱$ و در جواب دومین‌مرحله به صفر یعنی مرحله‌ی آغازین برمی‌گردیم. بنابراین جواب‌های الگوریتم هم‌چنان دو عدد ۰ و ۱ باقی خواهند ماند که این امر از همگرایی

نشریه تصمیم‌گیری و تحقیق در عملیات، دوره ۳، شماره (۴)، زمستان ۱۳۹۷، صفحه: ۴۰۲-۴۱۲

الگوریتم به ریشه‌ی واقعی ممانعت می‌کند. بدیهی است که این مشکل با اختیار نقطه آغازینی به جز نقاط ۰ و ۱ (مثلا $x^{(0)} = 0/5$) رفع می‌گردد (شکل ۶ را ببینید).




**شکل ۵- خطوط مماس بر تابع $x^3 - 2x + 2$ به‌ترتیب در نقاط ۰ و ۱ محور $x$ ها را قطع می‌کند.**




**شکل ٦- روند همگرایی الگوریتم برای یافتن ریشه‌ی تابع $f(x) = x^3 - 2x + 2$ به ازای نقطه آغازین $x^{(0)} = 0/5$ در مثال ۴.**

ماهیت بعضی از توابع به‌گونه‌ای است که روش نیوتن رافسون به‌ازای 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، درنظر بگیرید. در هر مرحله از حل معادله $f(x) = \sqrt[3]{x} = 0$ به‌وسیله الگوریتم نیوتن رافسون و به‌ازای 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 یک‌طرف دیگر محور، با فاصله‌ای بیش‌تر از جواب واقعی نسبت به مرحله‌ی قبلی، پرتات می‌کند. درحقیقیت، جواب الگوریتم نیوتن رافسون برای تابع $f(x) = |x|^{\alpha}$ به‌ازای هر $0 < \alpha < 0/5$، همگرا نبوده و به‌سمت بی‌نهایت میل می‌کند.

## ۴- کاربرد الگوریتم نیوتن رافسون در برآورد بیش‌ترین درست‌نمایی[1]

در این بخش به یکی‌از 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 درنظر می‌گیریم و برای آن یک‌مثال ارائه می‌دهیم، سپس روش را به‌حالت چند پارامتری تعمیم می‌دهیم.



### ۴-۱ برآورد تک پارامتر جامعه

نمونه‌ی تصادفی $X_1,...,X_n$ را از توزیعی با تابع جرم احتمال یا تابع چگالی احتمال $f(x;\theta)$ درنظر می‌گیریم که در آن $\theta$ پارامتر مجهول جامعه می‌باشد. فرض کنید مقدار مشاهده‌شده $\underline{X}=(X_1,...,X_n)$ را با $\underline{x}$ و هم‌چنین تابع درست‌نمایی و لگاریتم آن را به‌ترتیب با $L(\theta;\underline{x})$ و $l(\theta)=\ln L(\theta;\underline{x})=\sum_{i=1}^{n}\ln f(x_i;\theta)$ نشان می‌دهیم. منظور از برآورد پارامتر مجهول $\theta$ به‌روش بیش‌ترین درست‌نمایی، یافتن $\theta$ ای در فضای پارامتر است، به‌طوری که به‌ازای آن $\theta$ تابع درست‌نمایی بیشینه شود. به بیان‌دیگر، اگر $\hat{\theta}$ برآورد بیش‌ترین درست‌نمایی برای $\theta$ باشد، آن‌گاه $l'(\hat{\theta})=0$. اگر برآورد بیش‌ترین درست‌نمایی $\theta$ که ازاین‌پس 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:

$$\hat{\theta}^{(k+1)}=\hat{\theta}^{(k)}-(l''(\hat{\theta}^{(k)}))^{-1}\,l'(\hat{\theta}^{(k)}). \qquad (۲)$$

بدیهی است که برای شروع الگوریتم به یک‌برآورد آغازین از $\theta$ نیاز داریم. یکی از روش‌های یافتن مقدار برآورد اولیه، استفاده از برآوردگرهای دیگر مانند برآورد به‌روش گشتاوری است. اگر برآورد آغازین به‌اندازه‌ی کافی برای $\theta$ خوب باشد، آن‌گاه برآوردگر حاصل در اولین‌گام از الگوریتم دارای خاصیت بزرگ نمونه‌ای برآوردگرهای بیش‌ترین درست‌نمایی است (بیکل و داکسام، ۲۰۱۵)؛ به‌عبارت دیگر

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

که در آن مبتنی‌بر نمونه‌ی تصادفی $\underline{X}=(X_1,...,X_n)$، $l(\theta)=\ln L(\theta;\underline{X})$ و $I_{\underline{X}}(\theta)=E\left(\left[\frac{\partial}{\partial\theta}\ln L(\theta;\underline{X})\right]^2\right)$ اطلاع فیشر $\underline{X}$ در مورد $\theta$ می‌باشد.

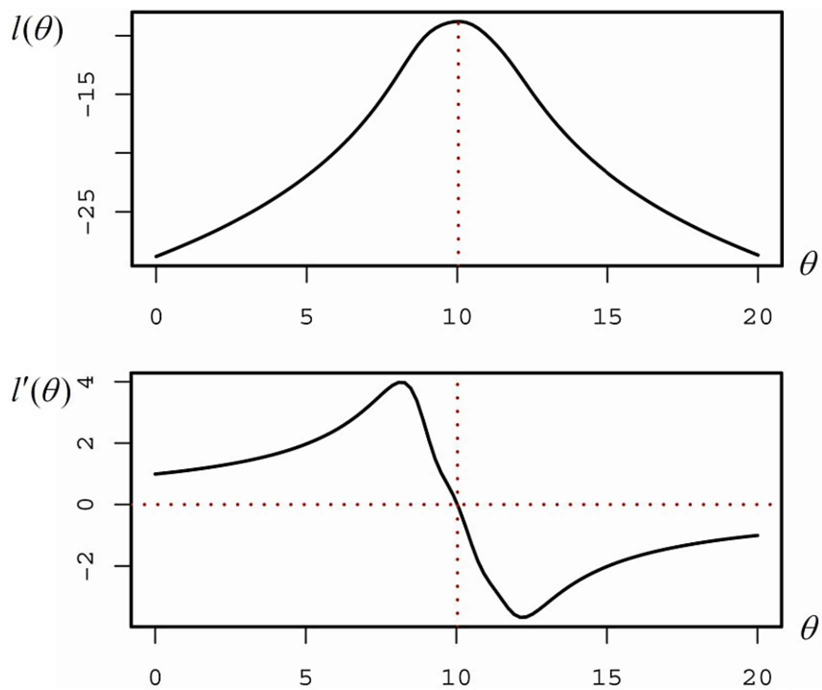
**مثال ۶- (برآورد پارامتر در توزیع کوشی تک پارامتری).** فرض کنید $\underline{X}=(X_1,...,X_n)$ یک نمونه‌ی تصادفی از توزیع کوشی، با تابع چگالی احتمال زیر باشد:

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

در این‌مثال فرض کنید نمونه‌ی تصادفی $x_1=۱۰/۵۷۵۱۴۰$، $x_2=۱۱/۶۲۴۹۸۸$، $x_3=۸/۹۸۵۹۱۳$، $x_4=۸/۹۳۵۰۰۰$، $x_5=۱۰/۱۶۸۳۰۲$ از جامعه‌ای با توزیع کوشی مشاهده شده است. براساس این نمونه‌ی مشاهده‌شده، توابع $l$ و $l'$ نسبت‌به $\theta$ در شکل ۷ رسم شده است.



[1] Maximum Likelihood Estimate (MLE)




**شکل ۷- تابع لگاریتم درست‌نمایی (تصویر بالا) و مشتق آن (تصویر پایین) در مثال ٦.**

با توجه به تقارن توزیع کوشی حول $\theta$، می‌توان هرکدام از میانگین و یا میانه‌ی نمونه را به‌عنوان نقطه‌ی آغازین $\hat{\theta}^{(0)}$ درنظر 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 $l'(\hat{\theta}^{(k)}) = 2\sum_{i=1}^{n}\frac{x_i-\hat{\theta}^{(k)}}{1+(x_i-\hat{\theta}^{(k)})^2}$ و $l''(\hat{\theta}^{(k)}) = 2\sum_{i=1}^{n}\frac{(x_i-\hat{\theta}^{(k)})^2-1}{(1+(x_i-\hat{\theta}^{(k)})^2)^2}$ در (۲)، رابطه‌ی بازگشتی الگوریتم مشخص 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 ۸/۷٤ و بیش‌تر از ۱۱/۸٤ انتخاب شود، دنباله‌ی $\{\hat{\theta}^{(k)}\}$ همگرا می گردد. 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.**

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

## ۴-۲ 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 $\theta = (\theta_1,\cdots,\theta_p)^T$ بردار پارامترهای مجهول جامعه باشد، آن‌گاه همانند حالت تک‌پارامتری، اساس کار، فرمول بازگشتی (۲) می‌باشد که در آن $l'(\hat{\theta}) = (\frac{\partial}{\partial\theta_1}l(\hat{\theta}),\ldots,\frac{\partial}{\partial\theta_p}l(\hat{\theta}))^T$ و هم‌چنین $l''(\hat{\theta}^{(k)})$ ماتریسی با درایه‌های $l''_{ij}(\hat{\theta}) = \frac{\partial^2}{\partial\theta_i\,\partial\theta_j}l(\hat{\theta})$ است به قسمی

که $i$ شماره سطر و $j$ شماره ستون ماتریس می‌باشد. گفتنی است که $H(\hat{\theta}) = l''(\hat{\theta}) = \left[\frac{\partial^{۲}}{\partial\theta_i\,\partial\theta_j} l(\hat{\theta})\right]_{p\times p}$ ماتریس هسین تابع $l(\theta)$ مبتنی‌بر مشاهدات نام دارد که رابطه‌ی $H(\hat{\theta}) = -I_{\underline{x}}(\hat{\theta})$ با ماتریس اطلاع فیشر دارد (نایت[1]، ۲۰۰۰).

**مثال ۷- (برآورد پارامترها در توزیع وایبل دو پارامتری).** فرض کنید $\underline{X} = (X_{۱},\ldots,X_n)$ یک نمونه‌ی تصادفی از توزیع وایبل با تابع چگالی احتمال

$$f(x;a,b) = \begin{cases} \frac{a}{b^a} x^{a-۱} \exp\left[-\left(\frac{x}{b}\right)^a\right], & if\ x \geq ۰,\ a,b > ۰ \\ ۰ & if\ \ x < ۰, \end{cases}$$

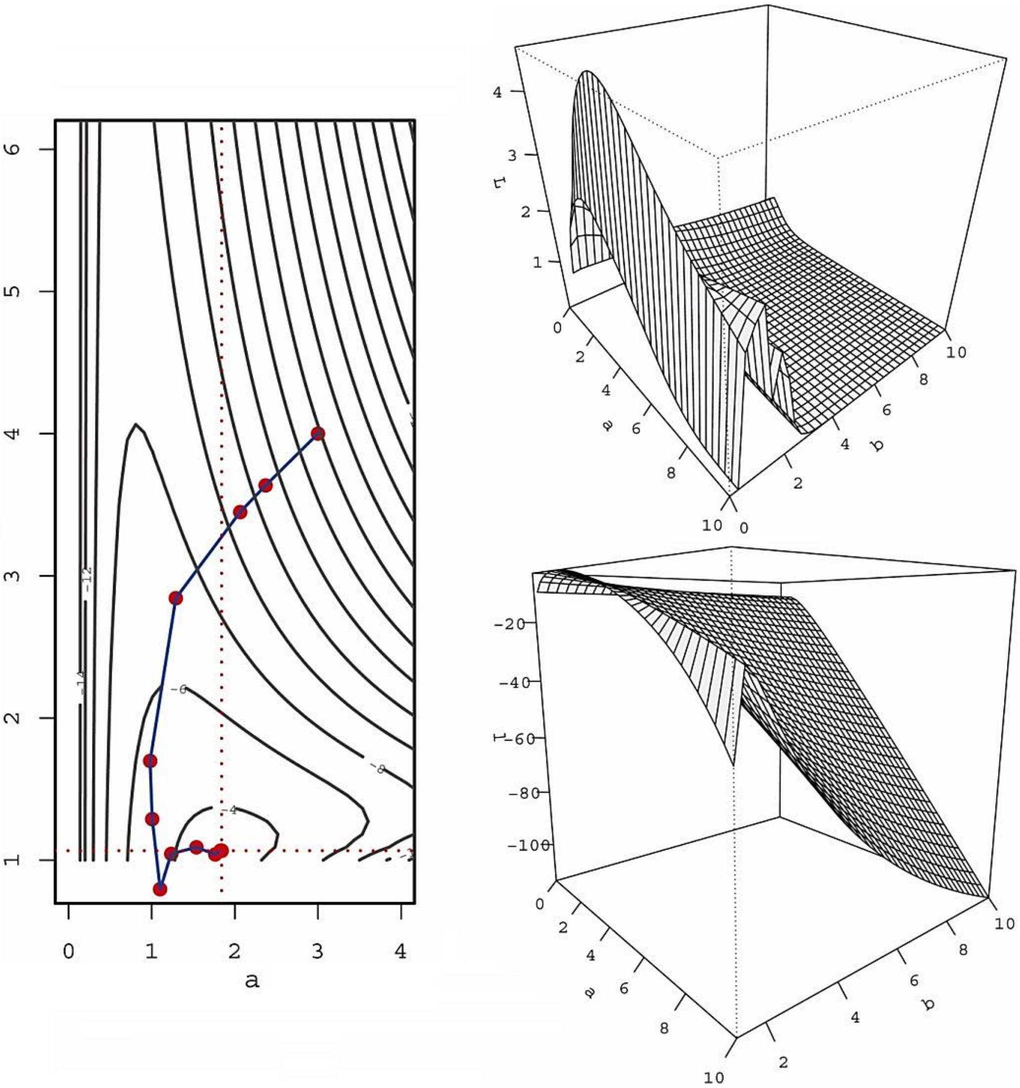
و با مقدار مشاهده‌شده $\underline{x} = (۰/۷۶۸,\ ۰/۳۷۵,\ ۱/۸۵۸,\ ۱/۲۳۵,\ ۰/۴۶۸)$ باشد. گفتنی است این‌توزیع، که اغلب در تحلیل داده‌های بقا مورد استفاده قرار می‌گیرد، در حالت خاص $a = ۱$ تبدیل‌به توزیع نمایی می‌شود. هدف، برآورد پارامتر $\theta = (a,b)^T$ به‌روش بیش‌ترین درست‌نمایی و بر پایه‌ی بردار مشاهدات $\underline{x}$ است. در این‌حالت، لگاریتم تابع درست‌نمایی برابر است با

$$l(\theta) = \sum_{i=۱}^{n}\left[\ln a - a\ln b + (a-۱)\ln x_i - \left(\frac{x_i}{b}\right)^a\right].$$

براساس تابع nlm، برآورد بیش‌ترین درست‌نمایی مبتنی‌بر نقطه‌ی آغازین $\hat{\theta}^{(۰)} = (\hat{a},\hat{b})^T = (۳,۴)^T$، پس از ۱۳ دور چرخش الگوریتم نیوتن رافسون، برابر $\hat{\theta}^{(۸)} = (۱/۸۴,\ ۱/۰۷)^T$ به‌دست 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 به‌روش نیوتن رافسون است. علاقه‌مندان جهت کسب اطلاعات و بررسی بیشتر در مورد توزیع وایبل سه‌پارامتره، می‌توانند به مرجع (کیو و توکاس[2]، ۱۹۹۵) مراجعه نمایند.

[1] knight

[2] Qiao & Tsokos




شکل۸- رویه تابع درست‌نمایی $L(\theta;\underline{x})$ (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 استفاده‌ی 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 متغییرهای 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 معادله‌ی $l'(\hat{\theta})=0$ برحسب $\theta$ (و نه $x$) بوده و ازاین‌رو، شرط مشتق‌پذیری و در نتیجه پیوستگی تابع $l'$ شرطی بدیهی است که مشمول متغییرهای تصادفی گسسته نیز می‌باشد.*
- *اگر $\sqrt{n}\left(\hat{\theta}^{(0)}-\theta\right)$ در توزیع هم‌گرا باشد، آن‌گاه $\sqrt{n}\left(\hat{\theta}^{(1)}-\hat{\theta}_n\right)\xrightarrow{P}0$ که در آن $\hat{\theta}_n$ برآورد $MLE(\theta)$ و $\hat{\theta}^{(1)}$ برآوردگر نخستین‌گام الگوریتم به‌ازای نقطه‌ی آغازین $\hat{\theta}^{(0)}$ می‌باشد (نایت، ۲۰۰۰).*
- *متاسفانه وقتی جواب آغازین از ریشه ی اصلی تابع دور باشد، روش نیوتن رافسون ممکن است تقریب نامناسبی را ارائه دهد. ازاین 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 به‌ازای تمام مقادیر $\theta$ در بازه‌ای که شامل ریشه‌ی واقعی تابع می‌باشد نامساوی*

$$\left|\frac{l'(\theta)l'''(\theta)}{(l'(\theta))^{2}}\right|<1 \qquad (۳)$$

***برقرار باشد، آن‌گاه به ازای هر مقدار آغازین $\hat{\theta}^{(0)}$ 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 درعین‌حال رویکرد نیوتن رافسون در برآورد بیش‌ترین درست‌نمایی هم‌گرا باشد (توماس و فینی، ۱۹۸۴).***

– ***وظیفه‌ی تابع nlm کمینه‌سازی تابعی است که به‌عنوان اولین آرگومان nlm درنظر گرفته شده است. لذا در برآورد بیش‌ترین درست‌نمایی به کمک تابع nlm، گرچه دنباله‌ی نقاطی 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 $-L$، $1/L$ و یا $-l$) بستگی دارد، اما نقطه‌ی برآوردشده‌ی نهایی (درصورت 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 درمورد استفاده‌ی صحیح از این الگوریتم، مطرح شده است.

## سپاس‌گزاری

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

## منابع

پرچمی، 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

Ypma, T. J. (1995). Historical development of the Newton–Raphson method. *SIAM review*, *37*(4), 531-551.

## ضمایم

در این بخش، برنامه‌های رایانه‌ای الگوریتم نیوتن رافسون براساس نرم افزار Maple 7 برای برخی‌از 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).
```



نشریه تصمیم‌گیری و تحقیق در عملیات، دوره ۳، شماره (۴)، زمستان ۱۳۹۷، صفحه: ۴۱۲-۴۰۲