

برآورد استوار بیضی‌گون کمینه حجم

اباذر خلجی، منوچهر خردمندنیا
گروه آمار، دانشکده علوم، دانشگاه اصفهان

چکیده

نقاط دورافتاده مشاهداتی هستند که از الگوی اکثر داده‌ها پیروی نمی‌کنند. به همین دلیل باعث تخریب برآوردگرهای بردار مکانی و ماتریس پراکندگی می‌شوند. در این مقاله روش بیضی‌گون کمینه حجم برای برآورد بردار مکانی و ماتریس پراکندگی معرفی می‌شود. در این روش از بین تمام زیرنمونه‌های h نقطه‌ای از n نقطه‌ای، زیر نمونه‌ای را که در عین پوشش h نقطه دارای حداقل حجم است، اساس محاسبه برآورد بردار مکان و ماتریس پراکندگی قرار می‌دهد. برآوردگرهای بدست آمده از روش بیضی‌گون کمینه حجم در مقابل داده‌های دورافتاده مقاوم هستند و علاوه بر این دارای خاصیت هم‌پایائی آفین و بیشینه فروریزش نیز می‌باشند. مشکل عمده این روش زمان‌بر بودن و حجم بالای محاسبات آن می‌باشد، زیرا تعداد زیرنمونه‌های مورد نیاز گاهی اوقات بسیار زیاد می‌شود. برای برطرف نمودن این مشکل الگوریتم‌های سریع معرفی می‌شود که با حداقل باز نمونه‌گیری به نتیجه مطلوب می‌رسند.

واژه‌های کلیدی: برآورد استوار، بیشینه فروریزش، هم‌پایائی آفین.

۱ مقدمه

همان نمونه تعدیل شده برآورد پراکندگی به دست می‌آید. در حالت تک متغیره امکان شناسایی نقاط دورافتاده از روی نمودار پراکنش داده‌ها نیز وجود دارد. در حالت دو و سه متغیره نیز اگر نمودارهای پراکنش را در فضاها دو یا سه بعدی ترسیم نماییم، احتمال شناسایی نقاط دورافتاده وجود دارد ولی لزوماً همیشه این کار شدنی نیست. در شناسایی نقاط دورافتاده زمانی که چند نقطه دورافتاده در مجموع داده‌ها وجود داشته باشد به علت

وجود نقاط دورافتاده در نمونه‌های یک یا چندمتغیره باعث تخریب برآوردگرها اعم از برآوردگر (بردار) مکان و (ماتریس) پراکندگی می‌شود. گاهی اوقات در نمونه‌های یک بعدی از میانه و گاهی نیز از میانگینی که با حذف درصدی از نقاط حدود بالایی و پایینی به دست آمده به عنوان برآورد مکان استفاده می‌شود و سپس با استفاده از

معکوسی بین فروریزش و کارایی بر قرار است به این شکل که افزایش فروریختگی با کاهش کارایی همراه است.

۴. در زمانی معقولی قابل محاسبه هستند، در این جا مقصود از زمانی معقول، کوتاه بودن زمان نیست بلکه بیش از اندازه زیاد و ناممکن بودن است.

۲ روش بیضی گون کمینه حجم

تعریف ۱.۲. اگر n مشاهده p متغیره داشته باشیم، در بین تمام زیرنمونه‌های h نقطه‌ای ($p < h < n$)، بردار میانگین و ماتریس کوواریانس زیرنمونه‌ای که در عین پوشش h نقطه، حجم بیضی‌گون متناظر با آن کمینه باشد، برآورد کمینه حجم به ترتیب برای μ و Σ می‌باشند. به عبارت دیگر فرض کنید:

الف) $x = (x_1, x_2, \dots, x_n)^T$ ماتریس $n \times p$ داده‌های نمونه است و μ و Σ به ترتیب بردار میانگین و ماتریس کوواریانس جامعه هستند.

ب) y زیر ماتریسی $h \times p$ از x است.

ج) $D^y(x_r, \tilde{\mu})$ فاصله نقطه r ام تا مرکز y است، یعنی:

$$D^y(x_r, \tilde{\mu}) = (x_r - \tilde{\mu})^T \tilde{\Sigma}^{-1} (x_r - \tilde{\mu}), r = 1, 2, \dots, n.$$

د) $D^y_{(h)}$ مقدار مرتب شده h ام عددهای $D^y(x_1, \tilde{\mu})$ تا $D^y(x_n, \tilde{\mu})$ است.

تحت شرایط فوق می‌گوییم که $\tilde{\mu}$ و $\tilde{\Sigma}$ برآورد کمینه حجم به ترتیب برای μ و Σ هستند هرگاه حجم بیضی‌گون زیر:

$$(x - \tilde{\mu})^T \tilde{\Sigma}^{-1} (x - \tilde{\mu}) = D^y_{(h)},$$

درون آوری^۱ شناسایی دیگر امکان‌پذیر نیست. پس مناسب‌تر است از برآوردگرهای استوار^۲ بردار مکانی و ماتریس پراکندگی برای شناسایی نقاط دورافتاده بهره جست. در این مقاله علاوه بر روش استوار بیضی‌گون کمینه حجم^۳ (MVE) برای برآورد μ و Σ الگوریتم‌های سریع برای محاسبه این روش نیز ارائه شده است. این روش به دنبال برآورد استواری است که تحت تاثیر نقاط دورافتاده نباشد. این روش نخستین بار توسط روسو [۷] معرفی شده است.

گوییم یک روش برای برآورد یا آزمون استوار است، اگر با حذف یا اصلاح درصد کوچکی از مجموعه داده‌ها نتایج برآورد یا آزمون تغییر قابل توجهی نکند. علاوه بر این‌ها برآوردگرهای MVE دارای خواص زیر نیز هستند:

۱. دارای نقطه فروریزش^۴ ۵۰ درصدی می‌باشند. نقطه فروریزش یک برآوردگر حداقل نسبتی از داده‌ها است که آلوده بودن آنها می‌تواند به برآوردی نامعقول منتهی گردد [۹].

۲. هم‌پایای آفین^۵ هستند، به این معنی که تغییر مقیاس اندازه‌گیری بر خصوصیت برآوردگرها اثر نمی‌گذارد [۱۰].

۳. کارایی بالایی دارند، یعنی واریانس تعمیم یافته یا واریانس کل این برآوردگرها نسبت به سایر برآوردگرها کمتر است. کارایی آماری برآوردگر مربوط می‌شود به این که چگونه از همه مشاهدات خوب در دسترس استفاده کنیم. معمولاً رابطه

^۱Masking

^۲Robust

^۳Minimum Volume Ellipsoid

^۴Breakdown Point

^۵Affine Equivariance

زیرنمونه (زیرماتریس) ممکن تعیین شود و سپس زیر نمونه‌ای که برای آن حجم بیضی‌گون $D^2(h)$ کمینه است، برای محاسبه $\bar{\mu}$ و $\bar{\Sigma}$ مورد استفاده قرار گیرد. واضح است که اگر n بزرگ باشد، تعداد زیرنمونه‌های h عضوی نیز افزایش یافته و به تبعیت از آن حجم محاسبات نیز بسیار زیاد می‌شود. اغلب تلاش‌های دو دهه اخیر مربوط به کم کردن حجم محاسبه‌ها از طریق روش‌های متنوع زیر نمونه گیری تصادفی بوده است. فرمول محاسبه حجم بیضی‌گون فوق به صورت زیر می‌باشد:

$$V = \frac{2\pi^{\frac{p}{2}}}{p\Gamma(\frac{p}{2})} [D^2(h) |\bar{\Sigma}|]^{1/2}$$

با توجه به این که p عدد ثابتی است، در عمل به جای کمینه کردن حجم کافی است که کمیت $v = [det(\bar{\Sigma})D^2(h)]^{1/2}$ مثال ساده زیر روشن‌گر بسیاری از ایده‌های مربوط به روش MVE می‌باشد.

مثال ۱۰.۲. ماتریس داده‌های زیر را در نظر بگیرید.

$$x = \begin{bmatrix} 4 & 15 & 6 & 12 & 5 \\ 13 & 25 & 12 & 15 & 17 \end{bmatrix}^T$$

در اینجا داریم، $n = 5$ و $p = 2$. بنابراین $h = 4 = \lfloor \frac{n+p+1}{2} \rfloor$. تعداد کل زیر نمونه‌های چهارتایی برابر $N = \binom{5}{4} = 5$ می‌باشد. این پنج زیر ماتریس 4×2 در زیر آورده شده‌اند.

$$y_1 = \begin{bmatrix} 15 & 25 \\ 6 & 12 \\ 12 & 15 \\ 5 & 17 \end{bmatrix}, \quad y_2 = \begin{bmatrix} 4 & 13 \\ 6 & 12 \\ 12 & 15 \\ 5 & 17 \end{bmatrix},$$

که به وسیله زیرماتریس y به دست آمده در مقایسه با حجم هر زیرماتریس $h \times p$ دیگری کمینه باشد.

کمیت h حجم زیرنمونه انتخابی برای محاسبه برآوردگرهای استوار $\bar{\mu}$ و $\bar{\Sigma}$ می‌باشد و نشان دهنده h مشاهده بالقوه خوب است که در محاسبه برآوردگرهای استوار شرکت دارند. اما $n-h$ مشاهده باقیمانده که فاصله بیشتری با مرکز داده‌ها دارند در محاسبه برآوردگرهای استوار شرکت داده نمی‌شوند و به این ترتیب برآوردگرهای استوار در برابر نقاط دورافتاده مقاوم می‌شوند. دیویس [۲] شرح داده است که با قرار دادن $h = \lfloor \frac{n+p+1}{2} \rfloor$ برآوردگرهای استوار بهبود می‌یابند. لذا ما نیز در محاسبات و مثال‌ها همین مقدار از h را در نظر گرفته‌ایم. در این جا مقصود از علامت [.] جزء صحیح است به طوری که h یک عدد صحیح مثبت است. در عمل h را می‌توان هر عدد صحیح در فاصله $\lfloor \frac{n+p+1}{2} \rfloor \leq h \leq n$ انتخاب نمود. یکی دیگر از انتخاب‌های مرسوم $h = 0.75n$ می‌باشد که منجر به افزایش کارایی برآوردگرها در مقابل کاهش نقطه فروریزش آن‌ها می‌شود. اگر $h = n$ آنگاه برآورد MVE مکانی همان بردار میانگین معمولی داده‌ها و برآورد MVE پراکندگی همان برآورد ماتریس کوواریانس نمونه است. در واقع با افزایش h برآوردگرهای استوار به برآوردگرهای معمولی نزدیک می‌شوند که منجر به کاهش نقطه فروریزش می‌شود. تعیین مقدار h بستگی به میزان آلودگی داده‌ها و هدف تحلیل‌گر دارد.

بنابراین برای محاسبه برآوردگرهای کمینه حجم برای μ و Σ لازم است که تمام،

$$N = \binom{n}{h} = \frac{n!}{h!(n-h)!}$$

چهارمین مقدار مرتب شده برابر $D_{(4)}^y = D_{(4)}^x = 1/6418$ می باشد. بنابراین بیضی زیر، ۴ مشاهده از پنج مشاهده را در بر دارد:

$$y_3 = \begin{bmatrix} 4 & 13 \\ 15 & 25 \\ 12 & 15 \\ 5 & 17 \end{bmatrix}, \quad y_4 = \begin{bmatrix} 4 & 13 \\ 15 & 25 \\ 6 & 12 \\ 5 & 17 \end{bmatrix},$$

$$(\underline{x} - \underline{\bar{y}}_1)^T s_1^{-1} (\underline{x} - \underline{\bar{y}}_1) = 1/6418.$$

به عبارت دیگر چهار نقطه از پنج نقطه در داخل و یا روی بیضی قرار دارند. در این جا نقطه x_5 دقیقاً روی بیضی قرار دارد. نقاط x_1 و x_3 و x_4 در داخل بیضی و نقطه x_2 در خارج از بیضی قرار دارند. مساحت این بیضی برابر است با:

$$y_5 = \begin{bmatrix} 4 & 13 \\ 15 & 25 \\ 6 & 12 \\ 12 & 15 \end{bmatrix}.$$

بردار میانگین و ماتریس زیر نمونه اول برابر است با:

$$\begin{aligned} V_1 &= cv_1 \\ &= c[\det(s_1)(1/6418)]^{1/2} \\ &= 24/1893c, \end{aligned}$$

$$\underline{\bar{y}}_1 = \begin{bmatrix} 9/5 \\ 17/25 \end{bmatrix},$$

$$s_1 = \begin{bmatrix} 23 & 18/8333 \\ 18/8333 & 30/9167 \end{bmatrix}.$$

که در آن $c = \frac{y\pi^{p/2}}{p\Gamma(p/2)} = 3/14159$ یک مقدار ثابت است.

مربع فاصله پنج مشاهده اولیه تا مرکز y_1 برابر است با:

بردار میانگین و ماتریس کوواریانس زیر نمونه دوم برابر است با:

$$D^y(x_1, \underline{\bar{y}}_1) = 1/3194, \quad D^y(x_2, \underline{\bar{y}}_1) = 1/9954,$$

$$D^y(x_3, \underline{\bar{y}}_1) = 0/8994, \quad D^y(x_4, \underline{\bar{y}}_1) = 1/4634,$$

$$D^y(x_5, \underline{\bar{y}}_1) = 1/6418.$$

$$\underline{\bar{y}}_2 = \begin{bmatrix} 6/75 \\ 14/25 \end{bmatrix},$$

$$s_2 = \begin{bmatrix} 12/9167 & 1/4167 \\ 1/4167 & 4/9167 \end{bmatrix}.$$

$D^y(x_r, \underline{\bar{y}}_1)$ در واقع فاصله نقطه x_r تا مرکز y_1 می باشند. فواصل مرتب شده عبارتند از:

مربع فاصله پنج مشاهده اولیه تا مرکز y_2 برابر است با:

$$D_{(1)}^y = 0/8994, \quad D_{(2)}^y = 1/3194,$$

$$D^y(x_1, \underline{\bar{y}}_2) = 0/7744, \quad D^y(x_2, \underline{\bar{y}}_2) = 25/6267,$$

$$D_{(3)}^y = 1/4634, \quad D_{(4)}^y = 1/6418,$$

$$D^y(x_3, \underline{\bar{y}}_2) = 1/0305, \quad D^y(x_4, \underline{\bar{y}}_2) = 2/1402,$$

$$D_{(5)}^y = 1/9954.$$

$$D^y(\underline{x}_5, \bar{y}_2) = 2/0549.$$

$$D^y(\underline{x}_1, \bar{y}_3) = 0/9349, \quad D^y(\underline{x}_2, \bar{y}_3) = 2/0451,$$

فواصل مرتب شده عبارتند از:

$$D^y(\underline{x}_3, \bar{y}_3) = 1/1862, \quad D^y(\underline{x}_4, \bar{y}_3) = 2/0168,$$

$$D^y(\underline{x}_5, \bar{y}_3) = 1/0032.$$

$$D^y_{(1)} = 0/7744, \quad D^y_{(2)} = 1/0305,$$

$$D^y_{(3)} = 2/0549, \quad D^y_{(4)} = 2/1402,$$

$$D^y_{(5)} = 25/6267.$$

فواصل مرتب شده عبارتند از:

$$D^y_{(1)} = 0/9349, \quad D^y_{(2)} = 1/0032,$$

$$D^y_{(3)} = 1/1862, \quad D^y_{(4)} = 2/0168,$$

$$D^y_{(5)} = 2/0451.$$

چهارمین مقدار مرتب شده برابر $D^y_{(4)} = 2/1402$ است. بنابراین بیضی زیر نیز ۴ مشاهده از پنج مشاهده را در بر دارد:

$$(\underline{x} - \bar{y}_2)^T s_2^{-1} (\underline{x} - \bar{y}_2) = 2/1402.$$

نقطه \underline{x}_2 در خارج این بیضی و نقطه \underline{x}_4 روی بیضی و بقیه در داخل بیضی قرار دارند. مساحت بیضی برابر است با:

بنابراین بیضی زیر نیز ۴ مشاهده از پنج مشاهده را در بر دارد:

$$(\underline{x} - \bar{y}_3)^T s_3^{-1} (\underline{x} - \bar{y}_3) = 2/0168.$$

نقطه \underline{x}_2 در خارج این بیضی و نقطه \underline{x}_4 روی بیضی و بقیه در داخل بیضی قرار دارند. مساحت بیضی برابر بردار میانگین و ماتریس کوواریانس زیر نمونه سوم است با:

$$\begin{aligned} V_2 &= cv_2 \\ &= c[\det(s_2)(2/1402)]^{1/2} \\ &= 11/4727c. \end{aligned}$$

برابر است با

$$\begin{aligned} V_3 &= cv_3 \\ &= c[\det(s_3)(2/0168)]^{1/2} \\ &= 27/1689c. \end{aligned}$$

$$\bar{y}_3 = \begin{bmatrix} 9 \\ 17/5 \end{bmatrix},$$

به طریق مشابه بر اساس زیر نمونه چهارم بیضی زیر حاصل می‌شود که این بیضی نیز $h = 4$ مشاهده از $n = 5$ مشاهده را پوشش می‌دهد.

$$s_3 = \begin{bmatrix} 28/6667 & 20/6667 \\ 20/6667 & 27/6667 \end{bmatrix}.$$

مربع فاصله پنج مشاهده اولیه تا مرکز y_3 برابر است با:

$$(\underline{x} - \bar{y}_4)^T s_4^{-1} (\underline{x} - \bar{y}_4) = 2/2073.$$

مساحت این بیضی برابر است با:

$$\begin{aligned} V_4 &= cv_4 \\ &= c[\det(s_4)(2/2073)]^{1/2} \\ &= 18/6848c. \end{aligned}$$

دوم، برآورد MVE به ترتیب برای μ و Σ می‌باشند. یعنی:

$$\tilde{\mu} = \bar{y}_2 = \begin{bmatrix} 6/75 \\ 14/25 \end{bmatrix},$$

$$\tilde{\Sigma} = s_2 = \begin{bmatrix} 12/9167 & 1/4167 \\ 1/4167 & 4/9167 \end{bmatrix}.$$

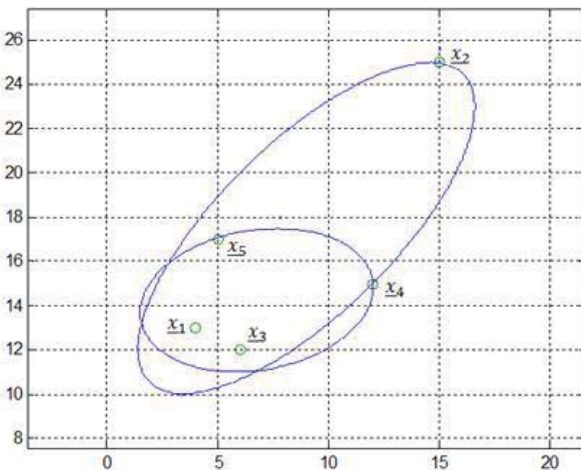
بر اساس زیر نمونه پنجم بیضی زیر حاصل می‌شود که این بیضی نیز $h = 4$ مشاهده از $n = 5$ مشاهده را پوشش می‌دهد.

$$(\underline{x} - \bar{y}_5)^T s_5^{-1} (\underline{x} - \bar{y}_5) = 2/2043.$$

در شکل ۱ پنج نقطه اولیه همراه با بیضی‌ای که کمترین مساحت را دارد و همراه با بیضی که بیشترین مساحت را دارد ملاحظه می‌شوند. همان طور که در شکل پیداست هر یک از بیضی‌ها چهار نقطه را در بر دارند که یکی از نقاط دقیقاً روی محیط بیضی قرار دارد.

مساحت این بیضی برابر است با:

$$\begin{aligned} V_5 &= cv_5 \\ &= c[\det(s_5)(2/2043)]^{1/2} \\ &= 24/0496c. \end{aligned}$$



به این ترتیب بر اساس هر یک از زیر نمونه‌های y_5, y_4, y_3, y_2, y_1 یک بیضی به دست آوردیم که هر یک $h = 4$ مشاهده از $n = 5$ مشاهده x را در بر دارد. بیضی‌ای که بر اساس زیر نمونه دوم (بر اساس y_2) به دست آمد دارای کمترین حجم است. به عبارتی دیگر:

$$\begin{aligned} \min(V_1, V_2, V_3, V_4, V_5) &= V_2 \\ &= 11/4727c. \end{aligned}$$

شکل ۱: بزرگترین و کوچکترین بیضی مبتنی بر چهار نقطه

در واقع هر بیضی‌گون مشخص کننده مجموعه نقاطی است که فاصله آن‌ها تا مرکز $\tilde{\mu}$ مقدار ثابتی است. اگر این مجموعه را به صورت بردار e در نظر بگیریم. طبق روسو [۷] بیضی‌گون e تحت تبدیل $Ae + b$ نیز یک بیضی‌گون است که در آن b هر بردار دلخواه و A ماتریسی ناکین می‌باشد. بنابراین خاصیت هم‌پایائی

بنابراین بردار میانگین و ماتریس کوواریانس زیر نمونه

آفین برآوردگر MVE به این معنی است که:

$$Volume(A\bar{e} + \underline{b}) = |det(A)|Volume(\bar{e}).$$

چون $det(A)$ ثابت است، حجم بیضی‌گون‌ها تحت تبدیل $A\bar{e} + \underline{b}$ تغییر نخواهد کرد. علاوه بر این‌ها روسو [۷] نشان داده است که برای هر اندازه نمونه n و هر بعد داده‌های p ، نقطه فروریزش برآوردگرهای MVE برابر مقدار زیر می‌باشد:

$$\frac{\left[\frac{n}{p}\right] - p + 1}{n},$$

که اگر n را به سمت بینهایت میل دهیم مقدار فروریزش به، ۵۰ درصد همگرا می‌شود.

مثال ۲.۲. (مطالعات شبیه‌سازی) در این مثال به کمک شبیه‌سازی و با استفاده از نرم افزار R عملکرد روش MVE را در آلودگی‌های مختلف مورد بررسی قرار می‌دهیم. ابتدا ماتریس 3×6 X را از توزیع $N_3(\underline{\mu}, \Sigma)$ تولید می‌کنیم. که در آن:

$$\underline{\mu} = \begin{bmatrix} 1 \\ 5 \\ 9 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{bmatrix}.$$

یکی از مشاهدات (مشاهده چهارم) را با مشاهده‌ای که هر بار به روشی آلوده شده، جایگزین می‌کنیم. با توجه به ابعاد ماتریس مشاهدات $h = 5$ و تعداد کل زیر نمونه‌های 3×5 برابر شش می‌باشد. شش زیر نمونه y_1, y_2, y_3, y_4, y_5 و y_6 به ترتیب از حذف $x_1^T, x_2^T, x_3^T, x_4^T, x_5^T, x_6^T$ از ماتریس مشاهدات x بدست می‌آیند که x_i^T نشان دهنده سطر i ام ماتریس x می‌باشد.

۱. آلوده کردن میانگین: در این شبیه‌سازی بردار میانگین متناظر با چهارمین مشاهده را به صورت

زیر آلوده می‌کنیم،

$$\underline{\mu}_{outlier} = \underline{\mu} + (0, 1, 0)^T$$

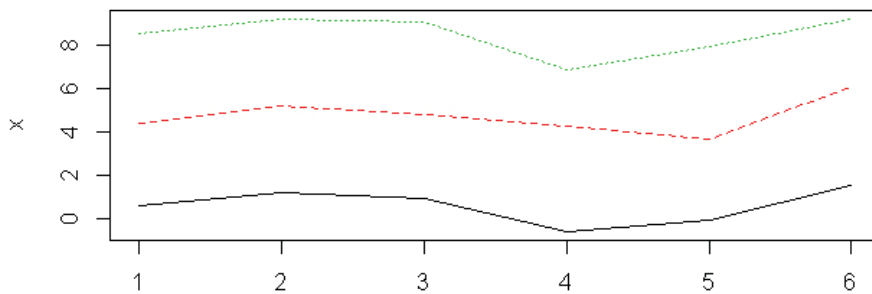
اما ماتریس کوواریانس را آلوده نمی‌کنیم. به عبارت دیگر برای مشاهده چهارم، میانگین دومین متغیر را به اندازه یک انحراف معیار افزایش می‌دهیم. به این ترتیب ماتریس مشاهدات زیر حاصل می‌گردد.

$$X_1 = \begin{bmatrix} 0.6 & 4.37 & 8.48 \\ 1.23 & 5.19 & 9.17 \\ 0.92 & 4.76 & 9.02 \\ -1.57 & 4.26 & 6.82 \\ -1.03 & 3.65 & 7.89 \\ 1.55 & 6.07 & 9.15 \end{bmatrix}.$$

نمودار دنباله‌ای مشاهدات در شکل ۲ ملاحظه می‌شود. نتایج کلیه محاسبات در ماتریس D و بردار v آورده شده است که سطر i ام ماتریس D مربع فاصله شش مشاهده اولیه تا مرکز زیر نمونه i ام و عضو i ام بردار v حجم بیضی‌گون متناظر با آن می‌باشد.

$$D = \begin{bmatrix} 2.87 & 0.55 & 2.63 & 3.13 & 3.06 & 2.63 \\ 1.21 & 0.96 & 3.18 & 3.13 & 1.59 & 2.90 \\ 1.41 & 3.12 & 1.91 & 3.11 & 1.59 & 2.90 \\ 2.00 & 1.43 & 2.67 & 1.31 & 2.74 & 3.16 \\ 3.02 & 0.44 & 2.54 & 3.17 & 5.05 & 2.84 \\ 1.57 & 2.04 & 2.81 & 3.19 & 2.39 & 1.71 \end{bmatrix},$$

$$\underline{v} = \begin{bmatrix} 0.06 \\ 0.07 \\ 0.03 \\ 0.01 \\ 0.05 \\ 0.04 \end{bmatrix}.$$



شکل ۲: نمودار پراکنش برای آلودگی میانگین

در واقع واریانس متغیر دوم را دو برابر و همبستگی بین متغیر اول و دوم را معکوس نموده‌ایم. اگر داده‌ها را در دو بعد تصور کنیم نمودار پراکنش داده‌ها به شکل بیضی خواهد بود که این نوع آلودگی مشاهده پرت را بر روی خطی موازی محور اصلی بیضی اما با فاصله بیشتری قرار می‌دهد که باعث می‌شود مشاهده آلوده دورتر از سایر مشاهدات قرار بگیرد (رنچر [۶]). نمودار دنباله‌ای ماتریس داده‌های فوق در شکل ۳ ملاحظه می‌شود. بنابراین داریم:

$$D = \begin{bmatrix} 8/23 & 1/59 & 3/04 & 3/05 & 1/14 & 3/18 \\ 1/99 & 5/02 & 3/15 & 3/18 & 3/17 & 0/51 \\ 2/36 & 2/85 & 6/15 & 3/12 & 5/52 & 1/50 \\ 2/03 & 3/03 & 3/07 & 98/8 & 3/12 & 0/75 \\ 1/97 & 3/17 & 3/12 & 3/19 & 2/97 & 0/54 \\ 3/19 & 1/57 & 3/02 & 3/06 & 1/16 & 1/15 \end{bmatrix},$$

$$\underline{v} = \begin{bmatrix} 0/1 \\ 0/12 \\ 0/05 \\ 0/04 \\ 0/13 \\ 1/015 \end{bmatrix}.$$

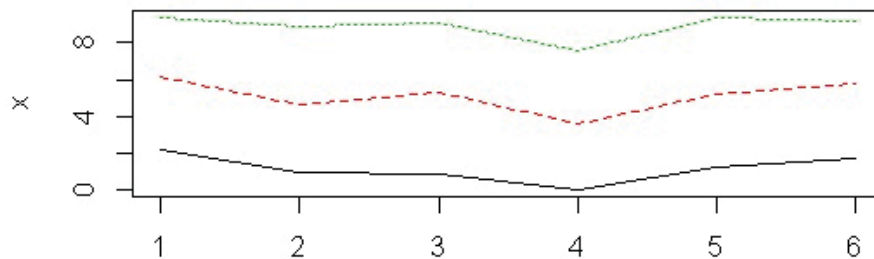
مشابه حالت قبل کمینه حجم مربوط به عضو چهارم

بنابراین کمینه حجم مربوط به عضو چهارم بردار \underline{v} می‌باشد، که زیر نمونه متناظر با آن به وسیله حذف مشاهده چهارم از ماتریس مشاهدات به وجود آمده است. می‌توان نتیجه گرفت که بردار میانگین و ماتریس کوواریانس زیر نمونه چهارم که مشاهده آلوده در آن وجود ندارد اساس برآورد $\underline{\mu}$ و Σ قرار می‌گیرد.

۲. آلوده کردن ماتریس کوواریانس: در شبیه‌سازی دیگری مشاهدات را از توزیع $N_3(\underline{\mu}, \Sigma)$ که در بالا شرح داده شد تولید کردیم، اما مشاهده آلوده (مشاهده چهارم) را بر اساس ماتریس کوواریانس زیر تولید کردیم ولی بردار میانگین را تغییر ندادیم. به این ترتیب ماتریس داده‌های زیر حاصل گردید.

$$\Sigma_{out} = \begin{bmatrix} 1 & -0/9 & 0/9 \\ -0/9 & 2 & 0/9 \\ 0/9 & 0/9 & 1 \end{bmatrix},$$

$$X_2 = \begin{bmatrix} 2/14 & 6/18 & 9/32 \\ 0/98 & 4/62 & 8/93 \\ 0/83 & 5/32 & 9/04 \\ 0/02 & 3/57 & 7/50 \\ 1/26 & 5/21 & 9/35 \\ 1/70 & 5/75 & 9/15 \end{bmatrix}.$$



شکل ۳: نمودار پراکنش برای آلودگی ماتریس کوواریانس

بردار \underline{v} می‌باشد. می‌توان نتیجه گرفت که بردار ۱.۳ الگوریتم استاندارد MVE

روسو و لروی [۸] یک روش باز نمونه گیری معرفی کرده‌اند که ما آن را به عنوان روش استاندارد نام گذاری کرده‌ایم. این نام گذاری از مقاله جووان و پریو [۳] اقتباس می‌شود.

اجرای این روش دارای مراحل زیر است:

۱. به طور تصادفی یک نمونه‌ای $k = p + 1$ تایی از n مشاهده p متغیره اولیه انتخاب می‌شود.

۲. بردار میانگین (\underline{t}) و ماتریس کوواریانس نمونه k تایی انتخاب شده (C) محاسبه می‌گردد.

$$\underline{t} = \frac{1}{p+1} \sum_{i=1}^{p+1} \underline{x}_i^*$$

$$C = \frac{1}{p} \sum_{i=1}^{p+1} (\underline{x}_i^* - \underline{t})(\underline{x}_i^* - \underline{t})^T.$$

۳. فاصله هر یک از n مشاهده اولیه تا مرکز نمونه انتخابی محاسبه می‌گردد.

$$D(\underline{x}_r, \underline{t}) = \left[(\underline{x}_r - \underline{t})^T C^{-1} (\underline{x}_r - \underline{t}) \right]^{\frac{1}{2}},$$

$$r = 1, 2, \dots, n.$$

مشاهده آلوده در آن وجود ندارد اساس برآورد $\underline{\mu}$ و Σ قرار می‌گیرد. لازم به ذکر است که با ۱۰۰ بار تکرار شبیه سازی، روش MVE در ۸۷ درصد مواقع قادر به ارائه برآوردهای استوار بود. تعداد دفعات باقیمانده ناشی از انحرافات شبیه سازی و خطاهای روش می‌باشد.

۳ الگوریتم‌های پرسرعت محاسبه

MVE

تمام روش‌های مشهور برآورد استوار بردار مکانی و ماتریس پراکنندگی از دیدگاه محاسبه‌ای بسیار سنگین هستند. لذا در عمل محاسبه‌ها می‌بایست به طور تقریبی بر اساس باز نمونه گیری صورت گیرد. جووان و پریو [۳] یک روش باز نمونه گیری خاص برای برآورد MVE معرفی کردند که دارای این خاصیت است که تعداد باز نمونه‌های مورد نیاز در مقایسه با روش استاندارد بسیار کمتر است. در بخش حاضر ابتدا روش استاندارد و سپس روش جووان و پریو معرفی می‌شوند.

۱. که دارای ۵ مشاهده دو متغیره است گام‌های الگوریتم را اجرا می‌کنیم. لازم به ذکر است که $h = 4$ می‌باشد.

۱. ابتدا به طور تصادفی زیر نمونه‌ای با حجم $p+1 = 3$

$$\begin{bmatrix} 12 & 6 & 5 \\ 15 & 12 & 17 \end{bmatrix}^T$$

اختیار می‌شود.

۲. بردار میانگین و ماتریس کوواریانس نمونه ۳ تایی محاسبه می‌شود.

$$\underline{t} = \begin{bmatrix} 7/67 \\ 14/67 \end{bmatrix}, \quad C = \begin{bmatrix} 14/34 & -1/17 \\ -1/17 & 6/34 \end{bmatrix}.$$

۳. فاصله هر یک از n مشاهده اولیه تا مرکز نمونه انتخابی محاسبه می‌گردد.

$$D(\underline{x}_1, \underline{t}) = 1/4, \quad D(\underline{x}_2, \underline{t}) = 26/6,$$

$$D(\underline{x}_3, \underline{t}) = 2/3, \quad D(\underline{x}_4, \underline{t}) = 1/4,$$

$$D(\underline{x}_5, \underline{t}) = 1/4.$$

۴. فاصله‌های بدست آمده از قسمت قبل را از کوچک به بزرگ مرتب می‌کنیم.

$$D(\underline{x}_1, \underline{t}) \leq D(\underline{x}_4, \underline{t})$$

$$\leq D(\underline{x}_5, \underline{t}) \leq D(\underline{x}_3, \underline{t}) \leq D(\underline{x}_2, \underline{t}).$$

۵. تعداد $h = 4$ نقطه که نزدیک‌ترین فاصله را تا نمونه ۳ تایی دارند را تعیین نموده بر اساس آن ۴ نقطه برآورد بردار میانگین $\tilde{\mu}$ و ماتریس کوواریانس $\tilde{\Sigma}$ محاسبه می‌شود.

۴. فاصله‌های $D(\underline{x}_1, \underline{t})$ تا $D(\underline{x}_n, \underline{t})$ به صورت صعودی مرتب می‌شوند.

$$D(\underline{x}_1, \underline{t}), D(\underline{x}_2, \underline{t}), \dots, D(\underline{x}_n, \underline{t}).$$

۵. تعداد h نقطه‌ای که نزدیک‌ترین فاصله را تا نمونه k تایی دارند تعیین نموده بر اساس آن h نقطه برآورد بردار میانگین $\tilde{\mu}$ و ماتریس کوواریانس $\tilde{\Sigma}$ محاسبه می‌شود.

$$\tilde{\mu} = \frac{1}{h} \sum_{j=1}^h \underline{x}_j,$$

$$\tilde{\Sigma} = \frac{1}{h-1} \sum_{j=1}^h (\underline{x}_j - \tilde{\mu})(\underline{x}_j - \tilde{\mu})^T.$$

۶. حجم با فرمول زیر محاسبه می‌گردد.

$$v = [\det(\tilde{\Sigma}) D_m^2]^{-\frac{1}{2}}.$$

۷. به مرحله ۱ برگشته و الگوریتم تا مرحله ۶ ام تکرار می‌شود. اگر حجمی که در مرحله ششم محاسبه شده از حجمی که در مرحله قبلی محاسبه شده کوچکتر باشد نمونه قبلی کنار گذاشته می‌شود و نمونه جدید جایگزین می‌شود، در غیر این صورت نمونه جدید کنار گذاشته می‌شود و نمونه قبلی حفظ می‌شود.

۸. مرحله های ۱ تا ۶ با توجه به هزینه و زمان در دسترس برای آنها ادامه می‌یابد.

۹. بردار میانگین و ماتریس کوواریانس باز نمونه h تایی که متناظر با کمترین حجم هستند به عنوان برآورد استوار MVE با روش استاندارد می‌باشند.

مثال ۱۰.۳. برای روشن شدن نحوه اجرای الگوریتم استاندارد MVE با استفاده از ماتریس مشاهدات مثال ۲

و با محاسبات کمتر به برآوردگر بهتری منتهی می‌شود. برآوردگرهایی که در این روش بدست می‌آیند قادر هستند که الگوهای آلودگی^۶ پیچیده را در نمونه مشخص کنند. اجرای این روش دارای مراحل زیر است.

$$\tilde{\mu} = \begin{bmatrix} ۶/۷۵ \\ ۱۴/۲۵ \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} ۱۲/۹ & ۱/۴ \\ ۱/۴ & ۴/۹ \end{bmatrix}.$$

۶. حجم با فرمول زیر محاسبه می‌گردد.

$$v = \left[\det(\tilde{\Sigma}) D_4^2 \right]^{\frac{1}{2}} = ۹/۲۷.$$

۱. به طور تصادفی یک نمونه $k = p + ۲$ تایی از n مشاهده p متغیره اولیه انتخاب می‌شود.

۷. به مرحله ۱ برگشته و الگوریتم تا مرحله ۶ ام تکرار می‌شود. اگر حجمی که در مرحله ششم محاسبه شده از حجمی که در مرحله قبلی محاسبه شده کوچکتر باشد باز نمونه قبلی کنار گذاشته می‌شود و باز نمونه جدید جایگزین می‌شود، در غیر این صورت باز نمونه جدید کنار گذاشته می‌شود و باز نمونه قبلی حفظ می‌شود. در جدول ۱ تعدادی از این تکرارها آمده است که مشاهده می‌شود کمترین حجم مربوط به زیر نمونه‌ای است که مشاهده دوم در آن حذف شده است. تفاوت حجم در تکرارهای سوم و چهارم با توجه به زیر نمونه‌های یکسان از این واقعیت ناشی می‌شود که D_h^2 در هر تکرار بسته به نمونه ۳ تایی مرحله اول متفاوت است.

۲. بردار میانگین و ماتریس کوواریانس نمونه k تایی انتخاب شده محاسبه می‌گردد.

۳. فاصله هر یک از k مشاهده اولیه را تا مرکز نمونه k تایی انتخابی محاسبه می‌گردد و مشاهده‌ای که دارای بیشترین فاصله است حذف می‌شود.

۴. بر اساس $۱ = p + ۱ = k - ۱$ نمونه باقیمانده بردار میانگین (\bar{t}) و ماتریس کوواریانس (C) محاسبه می‌گردد.

۵. $D(x_1, \bar{t})$ تا $D(x_n, \bar{t})$ به صورت صعودی مرتب می‌شوند.

$$D_{(1)}, D_{(2)}, \dots, D_{(n)}.$$

۶. تعداد h نقطه‌هایی که نزدیک‌ترین فاصله را تا نمونه k تایی دارند را تعیین نموده بر اساس آن h نقطه برآورد بردار میانگین $\tilde{\mu}$ و ماتریس کوواریانس $\tilde{\Sigma}$ محاسبه می‌شود.

$$\tilde{\mu} = \frac{1}{h} \sum_{j=1}^h x_j, \quad \tilde{\Sigma} = \frac{1}{h-1} \sum_{j=1}^h (x_j - \tilde{\mu})(x_j - \tilde{\mu})^T.$$

۷. حجم با فرمول زیر محاسبه می‌گردد.

$$v = \left[\det(\tilde{\Sigma}) D_m^2 \right]^{\frac{1}{2}}.$$

جدول ۱: تکرارهای الگوریتم

زیر نمونه	حجم	تکرار
x_1, x_3, x_4, x_5	۲/۲۷	اول
x_1, x_2, x_3, x_4	۱۸/۷	دوم
x_1, x_2, x_3, x_5	۱۴/۲۵	سوم
x_1, x_2, x_3, x_5	۳۲/۳۸	چهارم

۲.۳ الگوریتم سریع جوان و پریتمو

جوان و پریتمو [۳] الگوریتمی معرفی کردند که نسبت به روش استاندارد به تعداد باز نمونه گیری کمتری نیاز دارد

^۶Contamination

بازنمونه گیری به نتایج دقیقی که در آن همه باز نمونه‌های ممکن حضور دارند نزدیک خواهد بود. در جدول ۲ مقدار N_1 برای $\pi_1 = 0.95$ به تفکیک مقدارهای مختلف ε و p ملاحظه می‌شود. این جدول از جووان و پريتو [۳] می‌باشد. براساس این جدول، به عنوان مثال برای حالت $p = 6$ و $\varepsilon = 0.3$ اگر بخواهیم ۹۵ درصد مطمئن باشیم که حداقل یکی از باز نمونه‌ها شامل هیچ نقطه آلوده نیست باید تعداد باز نمونه‌ها برابر $N_1 = 35$ در نظر گرفته شود.

جدول ۲: مقدار N_1 برای $\pi_1 = 0.95$ به تفکیک مقدارهای مختلف ε و p

p/ε	0/1	0/2	0/3	0/4	0/5
۴	۴	۸	۱۷	۳۷	۹۵
۶	۵	۱۳	۳۵	۱۰۶	۳۸۲
۸	۷	۲۱	۷۳	۲۹۶	۱۵۳۳
۱۰	۸	۳۴	۱۵۰	۸۲۵	۶۱۳۴
۲۰	۲۶	۳۲۴	۵۳۶۲	۱۳۶۵۶۰	۲۶۸۲۵۰۶

• بر اساس جووان و پريتو، در الگوریتم سریع، اگر به تعداد N_2 باز نمونه تصادفی $p + 2$ تایی انتخاب کنیم احتمال این که حداقل یکی از باز نمونه‌ها پس از حذف بدترین مشاهده شامل هیچ نقطه آلوده‌ای نباشد برابر است با:

$$\pi_2 = 1 - (1 - (1 - \varepsilon)^{p+2} - (p + 2)(1 - \varepsilon)^{p+1} \varepsilon)^{N_2}$$

در جدول ۳ مقدار N_2 به ازای $\pi_2 = 0.95$ به تفکیک مقدارهای مختلف ε و p ملاحظه می‌شود. این جدول نیز از جووان و پريتو [۳] می‌باشد. بر اساس جدول ۳ با فرض $p = 6$ و $\varepsilon = 0.3$ اگر بخواهیم ۹۵ درصد مطمئن باشیم که حداقل یکی از باز نمونه‌ها شامل هیچ نقطه آلوده نیست باید تعداد باز نمونه‌ها برابر $N_2 = 11$ در نظر گرفته شود. مقایسه جدول ۲ و ۳ در واقع سرعت روش سریع را در مقایسه با روش استاندارد مشخص می‌کند. بر

۸. به مرحله ۱ برگشته و الگوریتم تا مرحله ۷ ام تکرار می‌شود. اگر حجمی که در مرحله هفتم محاسبه شده از حجمی که در مرحله قبلی محاسبه شده کوچکتر باشد نمونه قبلی کنار گذاشته می‌شود و نمونه جدید جایگزین می‌شود، در غیر این صورت نمونه جدید کنار گذاشته می‌شود و نمونه قبلی حفظ می‌شود.

۹. مرحله های ۱ تا ۷ با توجه به هزینه و زمان در دسترس برای آنها ادامه می‌یابد.

۱۰. بردار میانگین و ماتریس کوواریانس بازنمونه h تایی که متناظر با کمترین حجم هستند به عنوان برآورد استوار MVE با روش استاندارد می‌باشند.

در رابطه با الگوریتم های فوق نکات زیر قابل توجه می‌باشند:

• انتخاب اعداد کوچک $p + 1$ یا $p + 2$ به عنوان اندازه نمونه تصادفی با این انگیزه صورت می‌گیرد که احتمال قرار گرفتن یک نقطه دورافتاده در نمونه به حداقل خود برسد.

• در الگوریتم استاندارد روسو و لروی با این فرض که ε نسبت داده‌های آلوده است، احتمال این که باز نمونه‌های به اندازه $p + 1$ شامل هیچ نقطه آلوده‌ای نباشند برابر $(1 - \varepsilon)^{p+1}$ است. براساس مقاله جووان و پريتو [۳] اگر تعداد باز نمونه‌های انتخابی برابر N_1 باشد در الگوریتم استاندارد، احتمال این که حداقل یکی از بازنمونه‌ها شامل هیچ نقطه آلوده‌ای نباشد برابر است با:

$$\pi_1 = 1 - (1 - (1 - \varepsilon)^{p+1})^{N_1}$$

توجه کنید که اگر حداقل یکی از بازنمونه‌ها شامل هیچ نقطه آلوده‌ای نباشد نتیجه‌های حاصل از روش

اساس تجربه‌های شبیه‌سازی ما با فرض $n = 40$ و $11, 10, 7$ و 31 را مشکوک به دورافتاده بودن و سایر مشاهدات باقیمانده را نقاط دورافتاده شدید^۷ نامیدند. ما از این داده‌ها برای ارائه مثالی عملی از برآوردگر MVE بهره برده‌ایم. نکته بسیار جالب توجه این است که با تعداد تکرار نه چندان زیاد (مثلاً 50° تکرار)، رویت حتی یک مورد از 13 مورد نقطه دورافتاده در باز نمونه نهایی پیشامدی بسیار نادر است. عملاً با افزایش تعداد تکرارها می‌توان احتمال تحقق یافتن این پیشامد را به صفر نزدیک کرد. توجه کنید که از دیدگاه نظری برآوردهای MVE

جدول ۳: مقدار N_T برای $\pi_T = 0.95$ به تفکیک مقدارهای مختلف ε و p

p/ε	۰/۱	۰/۲	۰/۳	۰/۴	۰/۵
۴	۲	۳	۶	۱۲	۲۶
۶	۲	۵	۱۱	۲۷	۲۸
۸	۳	۷	۱۹	۶۴	۲۷۸
۱۰	۳	۱۰	۳۴	۱۵۲	۹۴۳
۲۰	۸	۶۱	۷۳۴	۱۴۵۲۷	۵۴۶۳۰۴

یکتا نیستند ولی احتمال یافتن دو زیر مجموعه h عضو که حجم آنها برابر حجم کمینه باشد بسیار ناچیز است. در شکل ۴ (به پیوست مراجعه شود) در 4 مورد با 100° تکرار بر اساس الگوریتم سریع جووان و پریتو دروند کاهش حجم بیضی‌گون‌ها و سرعت همگرایی الگوریتم در روش MVE ملاحظه می‌شود. در جدول ۴ (به پیوست مراجعه شود) برآوردهای ML با برآوردهای استوار مقایسه شده‌اند.

همان گونه ملاحظه می‌شود در تمام حالت‌های روش استوار برآورد واریانس هر یک از متغیرها از برآورد واریانس روش ML کوچکتر است. واریانس تعمیم یافته که بر اساس روش استوار بدست آمده نیز از واریانس تعمیم یافته بدست آمده از روش ML کمتر است. واریانس متغیرها در جدول ۴ در داخل پرانتز قرار گرفته‌اند.

• منطقی به نظر می‌رسد که تکرار اجرای الگوریتم‌های فوق تا رسیدن به زیرمجموعه‌ای که شامل اکثر مشاهدات بالقوه خوب است ادامه یابد، زیرا این زیرمجموعه دارای کمینه حجم است و اساس محاسبه برآورد استوار بردار میانگین و ماتریس کوواریانس قرار می‌گیرد. بنابراین برای افزایش دقت الگوریتم‌ها و تضمین رسیدن آن‌ها به یک مینیمم مطلق باید این الگوریتم‌ها در تعداد تکرار بالا انجام شوند، در تکرارهای کم مینیمم‌های موضعی به عنوان تقریبی مناسب از مینیمم مطلق استفاده می‌گردد.

۴ داده‌های کمپبل

کمپبل [۱] مجموعه داده‌هایی متشکل از $n = 38$ مشاهده در $p = 5$ بعد را ارائه کرده است. در متون آماری نشان داده شده که 13 مورد از این مشاهدات به شماره‌های 7 تا 11 و 31 تا 38 نقاط دورافتاده هستند. در این مورد مراجع مارونا و یوها [۵] و کوسینسکی [۴] را ملاحظه فرمایید. لازم به ذکر است که روك و ودروف [۱۰] نیز به همین نتایج رسیدند با این تفاوت که مشاهدات

مراجع

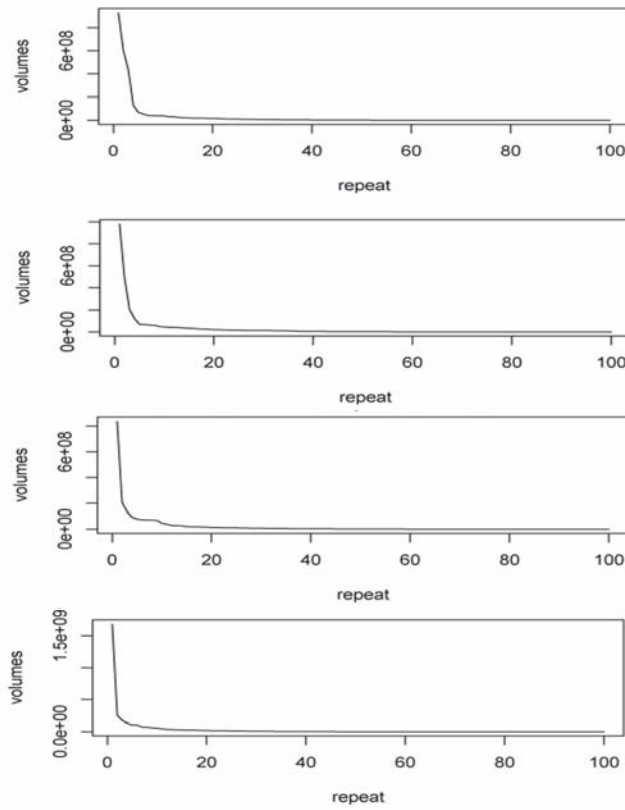
- [1] Campbell, N. A. (1989). Bushfire mapping using NOAA AVHRR data. Technical Report, CSIRO.

^۷Extrem Outlier

- Pflug, I. vineze and W. Werz, Dordrecht :Reidel.
- [8] Rousseeuw, P. J. and Leroy, A. M. (1987). Robust regression and outlier detection. John Wily and Snos, New York, NY.
- [9] Sakata, S. and With, H. (1998). Breakdown point. Encyclopedia of Statistical Sciences, 2, 84-89.
- [10] Woodruff, D. L. and Rocke, D. M. (1996). Identification of outlier in multivariate data. Journal of the American Statistical Association, 91(435), 1047-1061.
- [2] Davies, P. L. (1987). Asymptotic behavior of s-estimators of multivariate location parameteres and dispersion matrices. The Annals of Statistics, 15, 1269-1292.
- [3] Juan, J. and Prieto, F. J. (1995). A subsampling method of computation of multivariate estimators with high breakdown point. Journal of Computational and Graphical Statistics, 4(32), 319-334.
- [4] Kosinski, A. S. (1999). A procedure for the detection of multivariate outlier. Computational Statistic and Data Analysis, 29, 145-161.
- [5] Maronna, R. A. and Yohai, V. J. (1995). The behavior of the stahel-donoho robust multivariate location and shape. Journal of the American Statistical Association, 4, 51-67.
- [6] Rencher, A. C. (2002). Methods of multivariate analysis. New York: Wiley - interscience. Second Edition.
- [7] Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. In Mathematical Statistics and Application, VOL. B. eds. W. Grossmann, G.

۵ پیوست

در ادامه شکل ۴ و جدول ۴ ارائه شده اند.



شکل ۴: روند کاهش حجم بر اساس الگوریتم سریع جوان و پریتمو در 10^6 تکرار

جدول ۴: مقایسه برآوردهای ML و استوار

روش ML					برآورد
۱۰۳/۶	۱۲۹/۱	۲۸۸/۶	۲۲۷/۸	۲۸۶/۶	میانگین
(۴۰۶)	۵۶۵	-۲۰۹۱	-۶۳۹	-۵۱۶	ماتریس کوواریانس
۵۶۵	(۱۲۲۵)	-۳۲۵۸	-۱۱۸۴	-۹۴۳	
-۲۰۹۱	-۳۲۵۸	(۳۱۴۰۵)	۱۱۰۶۰	۹۰۲۱	
-۶۳۹	-۱۱۴۸	۱۱۰۶۰	(۴۱۰۳)	۳۳۴۰	
-۵۱۶	-۹۴۳	۹۰۲۱	۳۳۴۰	(۲۷۲۲)	
۱۱۹۵۲۰۰۰۰۰۰۰۰					واریانس تعمیم یافته
روش استوار					برآورد
۱۰۵/۵	۱۴۶/۹	۲۷۴/۴	۲۱۷/۶	۲۷۹/۶	میانگین
(۲۸۸)	۲۲۳	-۱۴۰۸	-۳۱۷	-۲۵۸	ماتریس کوواریانس
۲۲۳	(۱۹۷)	-۹۳۶	-۱۹۱	-۱۶۲	
-۱۴۰۸	-۹۳۶	(۸۳۱۴)	۲۰۴۳	۱۶۲۳	
-۳۱۷	-۱۹۱	۲۰۴۳	(۵۳۸)	۴۲۰	
-۲۵۸	-۱۶۲	۱۶۲۳	۴۲۰	(۳۳۱)	
۷۵۲۱۱۰۰۰					واریانس تعمیم یافته