سیزدهمین کنفرانس آمار ایران
۲-۴ شهریور ۱۳۹۵

الگوریتم EM مقید تحت قیدهای نابرابر روی پارامترها

مرضیه رضائی[۱]، جواد اطمینان[۲]

[۱] فارغ التحصیل کارشناسی ارشد دانشگاه بیرجند

[۲] استادیار گروه آمار دانشگاه بیرجند

چکیده: الگوریتم EM مقید برای پیدا کردن برآورد ماکسیمم درستنمایی تحت قیدهای خطی روی پارامترها اولین بار توسط کیم و تیلور بدست آمد. در عمل، اغلب در کاربردهای آماری با مجموعه‌ای از قیدهای نابرابر خطی نیز مواجه هستیم. واضح است در این موارد فضای پارامتر باز نیست و در نتیجه وقتی که مقادیر واقعی پارامتر روی مرز فضای پارامتر قرار می‌گیرند، پیدا کردن برآورد ماکسیمم درستنمایی مشکل است، زیرا گام M الگوریتم‌های نوع EM با پیچیدگی‌هایی همراه است. علاوه بر این در چنین مواردی ساختاربندی خصوصیات همگرایی الگوریتم‌های نوع EM کار ساده‌ای نیست، زیرا تکرارهای مکرر الگوریتم‌های نوع EM دنباله‌ای از برآوردها را نتیجه نمی‌دهد. در این مقاله الگوریتم EM مقید برای پیدا کردن برآورد ماکسیمم درستنمایی تحت قیدهای نابرابر روی پارامترها که توسط شی و همکارانش معرفی شده است بررسی و خصوصیات همگرایی آن ثابت می‌شود.

واژه‌های کلیدی: الگوریتم EM، داده ناکامل، برآورد ماکسیمم درستنمایی، مدل نرمال چندمتغیره

کد موضوع‌بندی ریاضی (۲۰۱۰): 99X99، 99X99، 99X99، 99X99.

---

۱ مقدمه

الگوریتم EMیکی از مفیدترین الگوریتم‌ها برای پیدا کردن برآورد ماکسیمم درستنمایی در مسائل مربوط به داده‌های غیرکامل است. ساختارهایی شامل نابرابری‌ها، اغلب در استنباط آماری مورد استفاده قرار می‌گیرد. در خیلی از مواقع، جریان مقید شدن پارامترهای مدل با نابرابری‌های خطی در کاربردهای عملی ظاهر می‌شود. برای مثال، ممکن است احتمال پاسخ یک تیمار خاص تابعی افزایشی از سطح تعیین دارو باشد یا ممکن است پاسخ تیمار به طور تصادفی، کنترل آزمایش را تحت تاثیر قرار دهد. به طور کلی استفاده از اطلاعات قیدهای نابرابر در استنباط آماری منجر به دستیابی به نتایج بهتر و کارآمدتر می‌شود.

---

[۱] مرضیه رضائی : rezaii.m2014@gmail.com

۲ فرمول‌بندی الگوریتم EM مقید

مدل خطی $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ را در نظر بگیرید که $\boldsymbol{X}$ ماتریس طرح معلوم، $\boldsymbol{\beta}$ بردار پارامتر نامعلوم و $\boldsymbol{e}$ بردار تصادفی نرمال با میانگین صفر و ماتریس کواریانس معلوم $\boldsymbol{\Sigma}$ است. فرض کنید $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{ik})^\top$، $i = 1, \cdots, n$ یک نمونه تصادفی به حجم n از $\boldsymbol{Y}$ باشد که Y_{ij}، jامین مؤلفه $\boldsymbol{y}_i$ می‌باشد. اگر $\boldsymbol{y}_i$ به طور کامل مشاهده نشود، آنگاه بردار داده‌های گمشده $\boldsymbol{y}_i$ با نماد $\boldsymbol{y}_{i(mis)}$ نشان داده می‌شود و $\boldsymbol{y}_{i(obs)}$ به بردار داده‌های مشاهده شده $\boldsymbol{y}_i$ اشاره دارد. چنانچه $\boldsymbol{y}_i$ به طور کامل مشاهده شود، $\boldsymbol{y}_{i(mis)}$ وجود ندارد و $\boldsymbol{y}_{i(obs)}$ همان $\boldsymbol{y}_i$ است. فرض کنید $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, \cdots, \boldsymbol{Y}_n^\top)^\top$ بردار داده‌های کامل باشد، سپس $\boldsymbol{y}_{mis} = (\boldsymbol{y}_{1(mis)}^\top, \cdots, \boldsymbol{y}_{n(mis)}^\top)^\top$، بردار داده‌های گمشده است و بردار داده‌های مشاهده شده به صورت $\boldsymbol{y}_{obs} = (\boldsymbol{y}_{1(obs)}^\top, \cdots, \boldsymbol{y}_{n(obs)}^\top)^\top$ می‌باشد. لگاریتم تابع درستنمایی داده‌های کامل با نماد $l(\boldsymbol{X}\boldsymbol{\beta}; \boldsymbol{y})$ نشان داده می‌شود و به صورت زیر تعریف می‌شود:

$$l(\boldsymbol{X}\boldsymbol{\beta}; \boldsymbol{y}) = -\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{X}\boldsymbol{\beta}) + c_0$$

به طوری که c_0 یک ثابت است که به بردار پارامتر نامعلوم $\boldsymbol{\beta}$ وابسته نیست. مشروط به اینکه بردار پارامتر $\boldsymbol{\beta}$ توسط تعدادی نابرابری به صورت $\boldsymbol{A}_0\boldsymbol{\beta} \geq \boldsymbol{0}$ مقید شده باشد، فرض کنید $\boldsymbol{A} = \boldsymbol{A}_0(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ و $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ باشند، در این صورت قید $\boldsymbol{A}_0\boldsymbol{\beta} \geq \boldsymbol{0}$ به فرم $\boldsymbol{A}\boldsymbol{\mu} \geq \boldsymbol{0}$ تبدیل می‌شود و در این حالت لگاریتم تابع درستنمایی داده‌های کامل با نماد $l(\boldsymbol{\mu}; \boldsymbol{y})$ نشان داده و به فرم زیر تبدیل می‌شود:

$$l(\boldsymbol{\mu}; \boldsymbol{y}) = -\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}) + c_0$$

واضح است که پیدا کردن برآورد ماکسیمم درستنمایی $\boldsymbol{\beta}$ در $l(\boldsymbol{X}\boldsymbol{\beta}; \boldsymbol{y})$ تحت قید $\boldsymbol{A}_0\boldsymbol{\beta} \geq \boldsymbol{0}$ معادل با پیدا کردن برآورد ماکسیمم درستنمایی $\boldsymbol{\mu}$ در $l(\boldsymbol{\mu}; \boldsymbol{y})$ تحت قید $\boldsymbol{A}\boldsymbol{\mu} \geq \boldsymbol{0}$ می‌باشد.

حال یک الگوریتم EM مقید برای پیدا کردن برآورد ماکسیمم درستنمایی $\boldsymbol{\mu}$ معرفی می‌کنیم. برای این منظور ابتدا امید ریاضی شرطی لگاریتم تابع درستنمایی داده‌های کامل به شرط داده‌های مشاهده شده را محاسبه می‌کنیم که،

$$\begin{aligned}
Q(\boldsymbol{\mu}; \boldsymbol{\mu}^{(m)}) &= -\frac{1}{2}\int \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}) f(\boldsymbol{y}_{mis}|\boldsymbol{y}_{obs}.\boldsymbol{\mu}^{(m)}) d\boldsymbol{y}_{mis} + c_0 \\
&= -\frac{1}{2}\sum_{i=1}^{n}\int \boldsymbol{y}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{y}_i f(\boldsymbol{y}_{mis}|\boldsymbol{y}_{obs}, \boldsymbol{\mu}^{(m)}) d\boldsymbol{y}_{mis} \\
&+ \frac{1}{2}\sum_{i=1}^{n}\int \boldsymbol{y}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} f(\boldsymbol{y}_{mis}|\boldsymbol{y}_{obs}, \boldsymbol{\mu}^{(m)}) d\boldsymbol{y}_{mis} \\
&+ \frac{1}{2}\sum_{i=1}^{n}\int \mu^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{y}_i f(\boldsymbol{y}_{mis}|\boldsymbol{y}_{obs}, \boldsymbol{\mu}^{(m)}) d\boldsymbol{y}_{mis} \\
&- \frac{1}{2}\sum_{i=1}^{n}\int \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} f(\boldsymbol{y}_{mis}|\boldsymbol{y}_{obs}, \boldsymbol{\mu}^{(m)}) d\boldsymbol{y}_{mis} + c_0 \\
&= -\frac{1}{2}\sum_{i=1}^{n} E(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)}) + \sum_{i=1}^{n} E(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)}) - \frac{n}{2}\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + c_0
\end{aligned}$$

با قرار دادن $\boldsymbol{V}_i^{(m)} = E(\boldsymbol{Y}_i|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})$ و $\bar{\boldsymbol{V}}^{(m)} = \frac{۱}{n}\sum_{i=۱}^n \boldsymbol{V}_i^{(m)}$ داریم:

$$\begin{aligned}
E(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}^{-۱}\boldsymbol{Y}_i|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)}) &= E[trac(\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}^{-۱}\boldsymbol{Y}_i)|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})] = E[trac(\boldsymbol{Y}_i\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}^{-۱})|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)}] \\
&= trac[E(\boldsymbol{Y}_i\boldsymbol{Y}_i^\top \boldsymbol{\Sigma}^{-۱})|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)}] = trac(Var(\boldsymbol{Y}_i|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})\boldsymbol{\Sigma}^{-۱} \\
&+ E(\boldsymbol{Y}_i|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})E(\boldsymbol{Y}_i^\top|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})\boldsymbol{\Sigma}^{-۱}) \\
&= trac[Var(\boldsymbol{Y}_i|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})\boldsymbol{\Sigma}^{-۱}] + trac[E(\boldsymbol{Y}_i|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})E(\boldsymbol{Y}_i^\top|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})\boldsymbol{\Sigma}^{-۱}] \\
&= trac[Var(\boldsymbol{Y}_i|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})\boldsymbol{\Sigma}^{-۱}] + trac[E(\boldsymbol{Y}_i^\top|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})\boldsymbol{\Sigma}^{-۱}E(\boldsymbol{Y}_i|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})] \\
&= trac[Var(\boldsymbol{Y}_i|\boldsymbol{Y}_{obs}, \boldsymbol{\mu}^{(m)})\boldsymbol{\Sigma}^{-۱}] + \boldsymbol{V}_i^{(m)\top}\boldsymbol{\Sigma}^{-۱}\boldsymbol{V}_i^{(m)}
\end{aligned}$$

همچنین با در نظرگرفتن $\boldsymbol{V}_i^{(m)} = (\boldsymbol{V}_i^{(m)} - \bar{\boldsymbol{V}}^{(m)}) + \bar{\boldsymbol{V}}^{(m)}$، تابع $Q(\boldsymbol{\mu};\boldsymbol{\mu}^{(m)})$ به صورت زیر نتیجه می‌شود:

$$\begin{aligned}
Q(\boldsymbol{\mu};\boldsymbol{\mu}^{(m)}) &= -\frac{۱}{۲}\sum_{i=۱}^n \Big[trac[Var(\boldsymbol{Y}_i|\boldsymbol{y}_{obs}, \boldsymbol{\mu}^{(m)})\boldsymbol{\Sigma}^{-۱}] + \boldsymbol{V}_i^{(m)\top}\boldsymbol{\Sigma}^{-۱}\boldsymbol{V}_i^{(m)}\Big] \\
&+ \sum_{i=۱}^n \boldsymbol{V}_i^{(m)\top}\boldsymbol{\Sigma}^{-۱}\boldsymbol{\mu} - \frac{n}{۲}\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-۱}\boldsymbol{\mu} + c_۰ \\
&= -\frac{۱}{۲}(\sum_{i=۱}^n trac[Var(\boldsymbol{Y}_i|\boldsymbol{y}_{obs}, \boldsymbol{\mu}^{(m)})\boldsymbol{\Sigma}^{-۱}] \\
&+ \sum_{i=۱}^n((\boldsymbol{V}_i^{(m)} - \bar{\boldsymbol{V}}^{(m)})^\top\boldsymbol{\Sigma}^{-۱}(\boldsymbol{V}_i^{(m)} - \bar{\boldsymbol{V}}^{(m)}) + \bar{\boldsymbol{V}}^{(m)\top}\boldsymbol{\Sigma}^{-۱}\bar{\boldsymbol{V}}^{(m)})) \\
&+ n\bar{\boldsymbol{V}}^{(m)\top}\boldsymbol{\Sigma}^{-۱}\boldsymbol{\mu} - \frac{n}{۲}\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-۱}\boldsymbol{\mu} + c_۰ \\
&= -\frac{n}{۲}(\boldsymbol{\mu} - \bar{\boldsymbol{V}}^{(m)})^\top\boldsymbol{\Sigma}^{-۱}(\boldsymbol{\mu} - \bar{\boldsymbol{V}}^{(m)}) + \boldsymbol{b}^{(m)}
\end{aligned}$$

که $\boldsymbol{b}^{(m)} = -\frac{۱}{۲}\sum_{i=۱}^n[trac(Var(\boldsymbol{Y}_i|\boldsymbol{y}_{obs}, \boldsymbol{\mu}^{(m)})\boldsymbol{\Sigma}^{-۱}) + (\boldsymbol{V}_i^{(m)} - \bar{\boldsymbol{V}}^{(m)})^\top\boldsymbol{\Sigma}^{-۱}(\boldsymbol{V}_i^{(m)} - \bar{\boldsymbol{V}}^{(m)})] + c_۰$ تابعی از $\boldsymbol{\mu}^{(m)}$ است. عبارت $\boldsymbol{V}_i^{(m)}$ به دو قسمت به صورت زیر افراز می‌شود:

$$\boldsymbol{V}_i^{(m)} = \boldsymbol{H}_i\boldsymbol{\mu}^{(m)} + \boldsymbol{Z}_i$$

که $\boldsymbol{H}_i$ به ماتریس $\boldsymbol{\Sigma}$ و $\boldsymbol{Z}_i$ به $\boldsymbol{y}_{obs}$ و $\boldsymbol{\Sigma}$ بستگی دارند. برای روشن شدن موضوع به مثال زیر توجه کنید.

مثال ۱.۲. فرض کنید $\boldsymbol{y}_i = (y_{i۱}, y_{i۲})^\top$ دارای توزیع نرمال با بردار میانگین $\boldsymbol{\mu}^{(m)} = (\mu_۱^{(m)}, \mu_۲^{(m)})^\top$ و ماتریس کواریانس $\boldsymbol{\Sigma}$ باشد به طوری که $y_{i۱}$ به مقدار مشاهده شده و $y_{i۲}$ به مقدار گمشده اشاره دارد و

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_۱^۲ & \sigma_۱\sigma_۲\rho \\ \sigma_۱\sigma_۲\rho & \sigma_۲^۲ \end{bmatrix}$$

سپس،

$$\boldsymbol{V}_i^{(m)} = E(\boldsymbol{Y}_i|y_{i۱}, \boldsymbol{\mu}^{(m)}) = E[(Y_{i۱}, Y_{i۲})^\top|y_{i۱}, \boldsymbol{\mu}^{(m)}] = \begin{bmatrix} y_{i۱} \\ \mu_۲^{(m)} + \frac{\sigma_۲\rho}{\sigma_۱}(y_{i۱} - \mu_۱^{(m)}) \end{bmatrix} = \boldsymbol{H}_i\begin{bmatrix} \mu_۱^{(m)} \\ \mu_۲^{(m)} \end{bmatrix} + \boldsymbol{Z}_i$$

به طوری که،

$$\boldsymbol{H}_i = \begin{bmatrix} \circ & \circ \\ -\frac{\sigma_{۲}\rho}{\sigma_{۱}} & ۱ \end{bmatrix} \quad , \boldsymbol{Z}_i = \begin{bmatrix} y_{i۱} \\ \frac{\sigma_{۲}\rho}{\sigma_{۱}} y_{i۱} \end{bmatrix}$$

فرض کنید $\boldsymbol{H} = \frac{۱}{n}\sum_{i=۱}^{n} \boldsymbol{H}_i$ و $\boldsymbol{Z}_\circ = \frac{۱}{n}\sum_{i=۱}^{n} \boldsymbol{Z}_i$ باشند، سپس $\bar{\boldsymbol{V}}^{(m)} = \boldsymbol{H}\boldsymbol{\mu}^{(m)} + \boldsymbol{Z}_\circ$ است. اکنون با قرار دادن،

$$\boldsymbol{\theta}^{(m)} = \boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{\mu}^{(m)} \;, \boldsymbol{C} = \boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{H}\boldsymbol{\Sigma}^{\frac{۱}{۲}} \;, \boldsymbol{\theta} = \boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{\mu} \;, \boldsymbol{Z} = \boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{Z}_\circ \;, \boldsymbol{B} = \boldsymbol{A}\boldsymbol{\Sigma}^{\frac{۱}{۲}}$$

داریم:

$$\bar{\boldsymbol{V}}^{(m)} = \boldsymbol{H}\boldsymbol{\mu}^{(m)} + \boldsymbol{Z}_\circ = \boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{H}\boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{\mu}^{(m)} + \boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{Z}_\circ$$

و

$$\begin{aligned}(\boldsymbol{\mu} - \bar{\boldsymbol{V}}^{(m)})^\top &= (\boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{\mu} - \boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{H}\boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{\mu}^{(m)} - \boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{Z}_\circ)^\top = (\boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\theta} - \boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{C}\boldsymbol{\theta}^{(m)} - \boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{Z})^\top \\ &= (\boldsymbol{\theta} - \boldsymbol{C}\boldsymbol{\theta}^{(m)} - \boldsymbol{Z})^\top \boldsymbol{\Sigma}^{\frac{۱}{۲}\top}\end{aligned}$$

بنابراین تابع $Q(\boldsymbol{\mu};\boldsymbol{\mu}^{(m)})$ به صورت زیر نتیجه می‌شود:

$$Q(\boldsymbol{\mu};\boldsymbol{\mu}^{(m)}) = -\frac{n}{۲}||\boldsymbol{C}\boldsymbol{\theta}^{(m)} + \boldsymbol{Z} - \boldsymbol{\theta}||^{۲} + \boldsymbol{b}^{(m)}$$

و قید $\boldsymbol{A}\boldsymbol{\mu} \geq \boldsymbol{0}$ به فرم $\boldsymbol{A}\boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{\mu} = \boldsymbol{B}\boldsymbol{\theta} \geq \boldsymbol{0}$ تبدیل می‌شود.

با قرار دادن $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(m)}) = -Q(\boldsymbol{\mu};\boldsymbol{\mu}^{(m)}) + \boldsymbol{b}^{(m)}$ و $\mathcal{D} = \{\boldsymbol{\theta} \in \mathcal{R}^p : \boldsymbol{B}\boldsymbol{\theta} \geq \boldsymbol{0}\}$، در مسائل برآوردیابی مقید قصد داریم، تابع $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(m)})$ را نسبت به $\boldsymbol{\theta} \in \mathcal{D}$ مینیمم کنیم که برآورد ماکسیمم درستنمایی مقید در این حالت با استفاده از الگوریتم کودو و دیکسترا بدست می‌آید. توجه کنید که $\boldsymbol{\theta}^{(m)}$ تابعی از $\boldsymbol{\mu}^{(m)}$ است.

الگوریتم EM مقید را می‌توان بر اساس $\boldsymbol{\mu}^{(m)}$ بدست آورد. در این الگوریتم نقطه آغازین به صورت تصادفی انتخاب می‌شود. به طور کلی الگوریتم EM مقید در این حالت به صورت زیر است:

۱- ابتدا $\boldsymbol{\mu}^{(\circ)}$ به عنوان مقدار آغازین برای $\boldsymbol{\mu}$ انتخاب می‌شود.

۲- درگام E مقدار تابع $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(m)}) \propto ||\boldsymbol{C}\boldsymbol{\theta}^{(m)} + \boldsymbol{Z} - \boldsymbol{\theta}||^{۲}$ برحسب $\boldsymbol{\theta}^{(m)}$ محاسبه می‌شود.

۳- درگام M با قرار دادن $\boldsymbol{\mu}^{(m+۱)} = \boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\theta}^{(m+۱)}$ با استفاده از الگوریتم دیکسترا یا الگوریتم کودو مقدار $\boldsymbol{\theta}^{(m+۱)}$که به صورت $\boldsymbol{\theta}^{(m+۱)} = \Pi(\boldsymbol{C}\boldsymbol{\theta}^{(m)} + \boldsymbol{Z}|\mathcal{D})$ یا همان تصویر $\boldsymbol{C}\boldsymbol{\theta}^{(m)} + \boldsymbol{Z}$ روی $\mathcal{D}$ است، محاسبه می‌شود.

این فرایند تا زمانی که برای یک عدد مثبت ثابت $t_\circ$، $||\boldsymbol{\mu}^{(m)} - \boldsymbol{\mu}^{(m+۱)}|| < ۱۰^{-t_\circ}$ باشد، تکرار می‌شود. مقدار $t_\circ$ در مسائل مختلف ۲ یا ۳ و غیره است. سپس $\boldsymbol{\mu}^{(m+۱)}$ به عنوان برآورد ماکسیمم درستنمایی مقید $\boldsymbol{\mu}$، تعیین می‌گردد و با نماد $\widehat{\boldsymbol{\mu}}$ نشان داده می‌شود. بنابراین برآورد ماکسیمم درستنمایی $\boldsymbol{\beta}$ یا همان $\widehat{\boldsymbol{\beta}}$ از حل معادله $\boldsymbol{X}\boldsymbol{\beta} = \widehat{\boldsymbol{\mu}}$ بدست می‌آید.

۳ همگرایی الگوریتم EM مقید

در این قسمت تعدادی از خصوصیات همگرایی الگوریتم معرفی شده را مورد بحث قرار می‌دهیم. توجه داشته باشید که $\{\boldsymbol{\mu}^{(m)}\}$ دنباله تولید شده توسط الگوریتم EM مقید است.

لم ۳.۱. اگر $\|\boldsymbol{C}\| < ۱$ باشد، آنگاه $\{\boldsymbol{\mu}^{(m)}\}$ به طور یکنواخت کراندار است که در آن $\boldsymbol{C} = \boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{H}\boldsymbol{\Sigma}^{\frac{۱}{۲}}$ است.

اثبات. با توجه به تعریف گام M الگوریتم EM مقید داریم، $\boldsymbol{\theta}^{(m+۱)} = \Pi(\boldsymbol{Z} + \boldsymbol{C}\boldsymbol{\theta}^{(m)}|\mathcal{D})$. بنابراین،

$$\begin{aligned}
\|\boldsymbol{\theta}^{(m+۱)}\| &\leq \|\boldsymbol{Z} + \boldsymbol{C}\boldsymbol{\theta}^{(m)}\| \leq \|\boldsymbol{Z}\| + \|\boldsymbol{C}\boldsymbol{\theta}^{(m)}\| = ||\boldsymbol{Z}|| + ||\boldsymbol{C}\boldsymbol{\theta}^{(m)}\frac{\|\boldsymbol{\theta}^{(m)}\|}{\|\boldsymbol{\theta}^{(m)}\|}|| \\
&\leq \|\boldsymbol{Z}\| + \|\boldsymbol{C}\|.\|\boldsymbol{\theta}^{(m)}\| \\
&\leq \|\boldsymbol{Z}\| + \|\boldsymbol{C}\|\|\boldsymbol{Z}\| + \|\boldsymbol{C}\|^{۲}\|\boldsymbol{Z}\| + \cdots + \|\boldsymbol{C}\|^{m-۱}\|\boldsymbol{Z}\| + \|\boldsymbol{C}\|^{m}\|\boldsymbol{Z}\| + \|\boldsymbol{C}\|^{m+۱}\|\boldsymbol{\theta}^{(۰)}\| \\
&\leq \|\boldsymbol{Z}\|(۱ + \|\boldsymbol{C}\| + \cdots + \|\boldsymbol{C}\|^{m}) + \|\boldsymbol{C}\|^{m+۱}\|\boldsymbol{\theta}^{(۰)}\| \\
&< \frac{\|\boldsymbol{Z}\|}{۱ - \|\boldsymbol{C}\|} + \|\boldsymbol{C}\|^{m+۱}\|\boldsymbol{\theta}^{(۰)}\| \leq \frac{\|\boldsymbol{Z}\|}{۱ - \|\boldsymbol{C}\|} + \|\boldsymbol{\theta}^{(۰)}\|
\end{aligned}$$

که $\|\boldsymbol{\mu}^{(m+۱)}\| = \|\boldsymbol{\Sigma}^{\frac{۱}{۲}}\boldsymbol{\theta}^{(m+۱)}\| \leq \|\boldsymbol{\Sigma}^{\frac{۱}{۲}}\|.\|\boldsymbol{\theta}^{(m+۱)}\|$. اکنون قرار می‌دهیم، $r_۰ = \frac{\|\boldsymbol{Z}\|}{۱ - \|\boldsymbol{C}\|} + \|\boldsymbol{\theta}_۰\|$ به طوری که $\boldsymbol{\theta}^{(۰)} = \boldsymbol{\Sigma}^{-\frac{۱}{۲}}\boldsymbol{\mu}^{(۰)}$ و $\boldsymbol{\mu}^{(۰)}$ نقطه آغازین برای $\boldsymbol{\mu}$ می‌باشند، بنابراین، $\|\boldsymbol{\Sigma}^{\frac{۱}{۲}}\|r_۰$ یک کران یکنواخت برای $\{\boldsymbol{\mu}^{(m)}\}$ است و برهان کامل می‌شود. □

با توجه به اینکه در مدل نرمال چند متغیره تابع $l(\boldsymbol{\mu}^{(m)};\boldsymbol{y}_{obs})$ کراندار و $l(\boldsymbol{\mu}^{(m+۱)};\boldsymbol{y}_{obs}) \geq l(\boldsymbol{\mu}^{(m)};\boldsymbol{y}_{obs})$ است، گزاره زیر نتیجه می‌شود.

گزاره ۳.۲. وقتی $m \longrightarrow \infty$، تابع $l(\boldsymbol{\mu}^{(m)};\boldsymbol{y}_{obs})$ همگرا می‌شود.

قضیه ۳.۳. اگر $\|\boldsymbol{C}\| < ۱$ باشد، آنگاه $\{\boldsymbol{\mu}^{(m)}\}$ همگراست.

اثبات. از آنجایی که $\boldsymbol{\theta}^{(m+p)} = \Pi(\boldsymbol{Z} + \boldsymbol{C}\boldsymbol{\theta}^{(m+p-۱)}|\mathcal{D})$ و $\boldsymbol{\theta}^{(m)} = \Pi(\boldsymbol{Z} + \boldsymbol{C}\boldsymbol{\theta}^{(m-۱)}|\mathcal{D})$، داریم:

$$\begin{aligned}
\|\boldsymbol{\theta}^{(m+p)} - \boldsymbol{\theta}^{(m)}\| &\leq \|\boldsymbol{C}(\boldsymbol{\theta}^{(m+p-۱)} - \boldsymbol{\theta}^{(m-۱)})\| \leq \|\boldsymbol{C}\|.\|\boldsymbol{\theta}^{(m+p-۱)} - \boldsymbol{\theta}^{(m-۱)}\| \\
&\leq \|\boldsymbol{C}\|^{m}\|\boldsymbol{\theta}^{(p)} - \boldsymbol{\theta}^{(۰)}\| \leq \|\boldsymbol{C}\|^{m}\|(\|\boldsymbol{\theta}^{(p)}\| + \|\boldsymbol{\theta}^{(۰)}\|) \leq ۲\|\boldsymbol{C}\|^{m}r_۰
\end{aligned}$$

که $r_۰ = \frac{\|\boldsymbol{Z}\|}{۱ - \|\boldsymbol{C}\|} + \|\boldsymbol{\theta}^{(۰)}\|$. بنابراین اگر $\|\boldsymbol{C}\| < ۱$ باشد، آنگاه برای هر p صحیح مثبت داده شده، داریم:

$$\lim_{m \longrightarrow \infty} \|\boldsymbol{\theta}^{(m+p)} - \boldsymbol{\theta}^{(m)}\| = ۰$$

همچنین،

$$\|\boldsymbol{\mu}^{(m+p)} - \boldsymbol{\mu}^{(m)}\| = \|\boldsymbol{\Sigma}^{\frac{۱}{۲}}(\boldsymbol{\theta}^{(m+p)} - \boldsymbol{\theta}^{(m)})\| \leq \|\boldsymbol{\Sigma}^{\frac{۱}{۲}}\|.\|\boldsymbol{\theta}^{(m+p)} - \boldsymbol{\theta}^{(m)}\|$$

بنابراین $\{\boldsymbol{\mu}^{(m)}\}$ یک دنباله کوشی است و لذا وقتی $m \longrightarrow \infty$، این دنباله همگرا می‌شود. □

۴ کاربرد الگوریتم EM مقید تحت قیدهای نابرابر

در مدل $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{e}$ می‌خواهیم آزمون فرضیه زیر را انجام دهیم:

$$\begin{cases} H_{0}:\boldsymbol{\mu}=\mathbf{0}\\ H_{1}:\boldsymbol{\mu}>\mathbf{0}\end{cases} \tag{۱.۴}$$

که $\boldsymbol{\mu}=\boldsymbol{X\beta}$ است.

حالت $k=2$ را در نظر بگیرید. با قرار دادن $\boldsymbol{\mu}_{0}=(\mu_{01},\mu_{02})^{\top}$ به عنوان مقدار واقعی $\boldsymbol{\mu}=(\mu_{1},\mu_{2})^{\top}$، فرض کنید برای $i=1,\cdots,m$، Y_{i1} مشاهده شده و Y_{i2} گمشده‌اند، برای $i=m+1,\cdots,k$، Y_{i1} گمشده و Y_{i2} مشاهده شده‌اند و برای $i=k+1,\cdots,n$، Y_{i1} و Y_{i2} هر دو مشاهده شده‌اند. بنابراین، بردار داده‌های مشاهده شده به صورت، $\boldsymbol{Y}_{obs}=(Y_{i1}\,;i=1,\cdots,m\,,Y_{i2}\,;i=m+1,\cdots,k\,\boldsymbol{Y}_{i}\,;i=k+1,\cdots,n)$ است. $\widetilde{\boldsymbol{\mu}}_{(n)}$ به برآورد ماکسیمم درستنمایی مقید (تحت فرض H_{1}) و $\widehat{\boldsymbol{\mu}}_{(n)}$ به برآورد ماکسیمم درستنمایی تحت فرض H_{0} اشاره دارند. همچنین $l_{(n)}(\boldsymbol{\mu};\boldsymbol{y}_{obs})$ به لگاریتم تابع درستنمایی داده‌های مشاهده شده اشاره دارد که برای سادگی با نماد $l_{(n)}(\boldsymbol{\mu})$ نشان داده می‌شود. بنابراین فرض (۱.۴) به فرم زیر تبدیل می‌شود:

$$\begin{cases} H_{0}:\mu_{1}=0,\mu_{2}=0\\ H_{1}:\mu_{1}>0,\mu_{2}>0\end{cases} \tag{۲.۴}$$

تابع $l_{(n)}(\boldsymbol{\mu})$ را به صورت زیر تعریف می‌کنیم:

$$\begin{aligned} l_{(n)}(\boldsymbol{\mu}) &= \sum_{i=1}^{m} log\ f_{1}(y_{i1};\boldsymbol{\mu})+\sum_{i=m+1}^{k} log\ f_{2}(y_{i2};\mu)+\sum_{i=k+1}^{n} log\ f_{3}(\boldsymbol{y}_{i};\boldsymbol{\mu})\\ &= \sum_{i=1}^{m}\frac{(y_{i1}-\mu_{1})^{2}}{2\sigma_{1}^{2}}+\sum_{i=m+1}^{k}\frac{(y_{i2}-\mu_{2})^{2}}{2\sigma_{2}^{2}}+\frac{1}{2}\sum_{i=k+1}^{n}(\boldsymbol{y}_{i}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_{i}-\boldsymbol{\mu})+c \end{aligned}$$

که c یک ثابت است.

قضیه ۱.۴. تحت فرض H_{0}، آماره نسبت درستنمایی $-2(l_{(n)}(\widehat{\boldsymbol{\mu}})-l_{(n)}(\widetilde{\boldsymbol{\mu}}))$ دارای توزیع زیر است:

$$\lim_{n\longrightarrow\infty}P(-2(l_{(n)}(\widehat{\boldsymbol{\mu}})-l_{(n)}(\widetilde{\boldsymbol{\mu}}))\geqslant t)=\sum_{i=1}^{2}w_{i}P(\chi_{i}^{2}\geqslant t)$$

که χ_{i}^{2} به توزیع خی‌دو با i درجه آزادی اشاره دارد. همچنین $w_{2}=\lim_{n\longrightarrow\infty}P_{H_{0}}(\widetilde{\mu}_{1_{(n)}}>0,\widetilde{\mu}_{2_{(n)}}>0)$ و $w_{1}=\lim_{n\longrightarrow\infty}P_{H_{0}}(\widetilde{\mu}_{1_{(n)}}>0,\widetilde{\mu}_{2_{(n)}}=0)+\lim_{n\longrightarrow\infty}P_{H_{0}}(\widetilde{\mu}_{1_{(n)}}=0,\widetilde{\mu}_{2_{(n)}}>0)$ می‌باشند.

اثبات. ابتدا حالتی را در نظر بگیرید که $\widetilde{\mu}_{1_{(n)}}>0$ و $\widetilde{\mu}_{2_{(n)}}>0$ باشند، یعنی، $\widetilde{\boldsymbol{\mu}}_{(n)}$ یک نقطه درونی است. بنابراین $\frac{\partial l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}_{(n)}}=\mathbf{0}$.

با بکار بردن قضیه مقدار میانگین برای $\frac{\partial l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}$ در بازه باز $(\mathbf{0},\widetilde{\boldsymbol{\mu}}_{(n)})$ داریم:

$$\frac{\partial l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}_{(n)}}=\frac{\partial l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}|_{\boldsymbol{\mu}=\mathbf{0}}+\frac{\partial^{2}l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^{\top}}|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}^{*}_{(n)}}(\widetilde{\boldsymbol{\mu}}_{(n)})$$

در نتیجه،

$$\frac{\partial l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}|_{\boldsymbol{\mu}=\mathbf{0}}=-\frac{\partial^{2}l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^{\top}}|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}^{*}_{(n)}}(\widetilde{\boldsymbol{\mu}}_{(n)})$$

که $\widetilde{\boldsymbol{\mu}}^{*}_{(n)}$ بین صفر و $\widetilde{\boldsymbol{\mu}}_{(n)}$ می‌باشد. با توجه به قانون اعداد بزرگ داریم:

$$\lim_{n\longrightarrow\infty}\frac{۱}{n}\frac{\partial^{۲}l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^{\top}}|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}^{*}_{(n)}}=-\boldsymbol{I}(\mathbf{0})$$

که،

$$\boldsymbol{I}(\mathbf{0})=\lim_{n\longrightarrow\infty}E[(\frac{۱}{n}\frac{\partial l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}})(\frac{۱}{n}\frac{\partial l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}})^{\top}|_{\boldsymbol{\mu}=\mathbf{0}}]$$

همچنین،

$$n^{\frac{۱}{۲}}\widetilde{\boldsymbol{\mu}}\stackrel{L}{\longrightarrow}\boldsymbol{I}^{-۱}(\mathbf{0})\boldsymbol{V}$$

به طوری که $\boldsymbol{V}$ یک متغیر تصادفی نرمال با میانگین صفر و ماتریس واریانس $\boldsymbol{I}(\mathbf{0})$ است و L به نماد همگرایی در توزیع اشاره دارد.
همچنین با توجه به قضیه حد مرکزی داریم:

$$\frac{۱}{n^{\frac{۱}{۲}}}\frac{\partial l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}|_{\boldsymbol{\mu}=\mathbf{0}}\stackrel{L}{\longrightarrow}\boldsymbol{V}$$

با بکار بردن بسط سری تیلور $l_{(n)}(\widehat{\boldsymbol{\mu}})$ حول نقطه $\widetilde{\boldsymbol{\mu}}_{(n)}$ داریم:

$$l_{(n)}(\widehat{\boldsymbol{\mu}})=l_{(n)}(\widetilde{\boldsymbol{\mu}}_{(n)})+\frac{۱}{۲}\widetilde{\boldsymbol{\mu}}^{\top}_{(n)}\frac{\partial^{۲}l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^{\top}}|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}^{*}_{(n)}}\widetilde{\boldsymbol{\mu}}(n)$$

که،

$$\frac{۱}{n}\frac{\partial^{۲}l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^{\top}}|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}^{*}_{(n)}}\stackrel{a.s}{\longrightarrow}-\boldsymbol{I}(\mathbf{0})$$

$a.s$ به نماد همگرایی تقریباً حتمی اشاره دارد. سپس،

$$-۲(l_{(n)}(\widehat{\boldsymbol{\mu}})-l_{(n)}(\widetilde{\boldsymbol{\mu}}_{(n)}))=-\widetilde{\boldsymbol{\mu}}^{\top}_{(n)}\frac{\partial^{۲}l_{(n)}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}^{\top}}|_{\boldsymbol{\mu}=\widetilde{\boldsymbol{\mu}}^{*}_{(n)}}\widetilde{\boldsymbol{\mu}}(n)$$

بنابراین،

$$-۲(l_{(n)}(\widehat{\boldsymbol{\mu}})-l_{(n)}(\widetilde{\boldsymbol{\mu}}_{(n)}))\stackrel{L}{\longrightarrow}\boldsymbol{V}^{\top}\boldsymbol{I}^{-۱}(\mathbf{0})\boldsymbol{V}$$

که $\boldsymbol{V}^{\top}\boldsymbol{I}^{-۱}(\mathbf{0})\boldsymbol{V}$ یک متغیر تصادفی است که از توزیع خی‌دو با ۲ درجه آزادی پیروی می‌کند. بنابراین زمانی که $\widetilde{\mu}_{۱_{(n)}}>۰$ و $\widetilde{\mu}_{۲_{(n)}}>۰$ برقرارند، داریم:

$$-۲(l_{(n)}(\widehat{\boldsymbol{\mu}})-l_{(n)}(\widetilde{\boldsymbol{\mu}}_{(n)}))\stackrel{L}{\longrightarrow}\chi^{۲}_{۲}$$

که $\chi^{۲}_{۲}$ به توزیع خی‌دو با ۲ درجه آزادی اشاره دارد. اکنون در حالتی که $\widetilde{\mu}_{۱_{(n)}}=۰$ و $\widetilde{\mu}_{۲_{(n)}}>۰$ باشند، $\mu_{۱}=۰$ است و $l_{(n)}(\boldsymbol{\mu})$ در نقطه $\widetilde{\mu}_{۲_{(n)}}$ ماکسیمم می‌شود. فرض کنید $l_{(n)}(\mu_{۲})$ مقدار لگاریتم تابع درستنمایی در نقطه $\mu_{۱}=۰$ باشد، در این صورت $\frac{\partial l_{(n)}(\mu_{۲})}{\partial\mu_{۲}}|_{\mu_{۲}=\widetilde{\mu}_{۲_{(n)}}}=۰$. با بکار بردن همان روش بکار رفته در بالا داریم:

$$-۲(l_{(n)}(\widehat{\boldsymbol{\mu}})-l_{(n)}(\widetilde{\boldsymbol{\mu}}_{(n)}))\stackrel{L}{\longrightarrow}\chi^{۲}_{۱}$$

که $\chi^{۲}_{۱}$ به توزیع خی‌دو با ۱ درجه آزادی اشاره دارد. با این فرض که $l_{(n)}(\mu_{۱})$ مقدار لگاریتم تابع درستنمایی در نقطه $\mu_{۲}=۰$ باشد، سپس با بکارگرفتن روش بالا داریم:

$$-۲(l_{(n)}(\widehat{\boldsymbol{\mu}})-l_{(n)}(\widetilde{\boldsymbol{\mu}}_{(n)}))\stackrel{L}{\longrightarrow}\chi^{۲}_{۱}$$

بنابراین با در نظر گرفتن همه حالت‌ها داریم:

$$\begin{aligned}\lim_{n\longrightarrow\infty} P(-۲(l_{(n)}(\widehat{\boldsymbol{\mu}}) - l_{(n)}(\widetilde{\boldsymbol{\mu}}_{(n)})) \geqslant t) &= \lim_{n\longrightarrow\infty} P(\widetilde{\mu}_{۱(n)} > ۰, \widetilde{\mu}_{۲(n)} > ۰)P(\chi^۲_۲ \geqslant t)\\ &+ \lim_{n\longrightarrow\infty} P(\widetilde{\mu}_{۱(n)} = ۰, \widetilde{\mu}_{۲(n)} > ۰)P(\chi^۲_۱ \geqslant t) + \lim_{n\longrightarrow\infty} P(\widetilde{\mu}_{۱(n)} > ۰, \widetilde{\mu}_{۲(n)} = ۰)P(\chi^۲_۱ \geqslant t)\end{aligned}$$

با تعریف $w_۱$ و $w_۲$ به صورت،

$$w_۱ = \lim_{n\longrightarrow\infty} P_{H_۰}(\widetilde{\mu}_{۱(n)} > ۰, \widetilde{\mu}_{۲(n)} = ۰) + \lim_{n\longrightarrow\infty} P_{H_۰}(\widetilde{\mu}_{۱(n)} = ۰, \widetilde{\mu}_{۲(n)} > ۰)$$

$$w_۲ = \lim_{n\longrightarrow\infty} P_{H_۰}(\widetilde{\mu}_{۱(n)} > ۰, \widetilde{\mu}_{۲(n)} > ۰)$$

نتیجه می‌گیریم که،

$$\lim_{n\longrightarrow\infty} P(-۲(l_{(n)}(\widehat{\boldsymbol{\mu}}) - l_{(n)}(\widetilde{\boldsymbol{\mu}})) \geqslant t) = \sum_{i=۱}^{۲} w_i P(\chi^۲_i \geqslant t)$$

□

بحث و نتیجه‌گیری

در این مقاله، الگوریتم EM مقید تحت نابرابری‌های خطی برای یک مدل با ماتریس کواریانس معلوم که توسط شی و همکاران معرفی شده است، مطرح شد. خصوصیات همگرایی الگوریتم پیشنهاد شده ثابت شد و مسائل مربوط به آزمون فرض مورد توجه قرار گرفت.

مراجع

[1] Kim, D. K. and Taylor, J. M. G. (1995). The restricted EM algorithm for maximum likelihood estimation under linear restrictions on the parameters. *Journal of the American Statistical Association . 90, 708–716.*

[2] Shi, N.Z., Zheng, S.R. and Guo, J.H. (2005). The restricted EM algorithm under inequality restrictions on the parameters. Journal of Multivariate Analysis. 92, 53–76.