

# میانگین میدان تقریب تغییرات بیز

میلااد اسدزاده، آرش اردلان  
گروه ریاضی، دانشگاه یاسوج

## چکیده

تکنیک‌های برآوردیابی پارامترهای موجود در مدل‌های آماری امروزه بسیار به چالش کشیده شده است. در حالت کلی این تکنیک‌ها به دو دسته‌ی احتمالی و غیره احتمالی تقسیم می‌شود. روش‌های احتمالی به ویژه زنجیر مارکوف مونت کارلو (*MCMC*) با وجود دقت بالا، دارای روند بسیار کند در برآورد پارامترها می‌باشد. از این رو، برای رفع این کندی در روند برآورد از تکنیک غیره احتمالی با عنوان تقریب تغییرات با استفاده از محدودیت میانگین میدان تغییرات بیز (*MFVB*) می‌پردازیم. که این معیار بسیار سریعتر از *MCMC* و در عین حال دارای دقت بالایی نیز می‌باشد.

واژه‌های کلیدی: تقریب تغییرات، زنجیر مارکوف مونت کارلو، میانگین میدان تقریب تغییرات.

## ۱ مقدمه

استنباط تقریب تغییرات بیز، که در اواسط قرن هجدهم میلادی بر اساس کارهای اوایلر [۲] یک روش غیره فشرده محاسباتی در مقابل داده‌های حجیم و با سرعت رشد می‌باشد. نام "تقریب تغییرات"<sup>۲</sup> ریشه در عنوان شناخته شده ریاضی مانند تغییرات حساب دیفرانسل و جبر که مرتبط به مبحث بهینه‌سازی یک تابع بر یک کلاس از توابع وابسته، می‌باشد. با وجود مفاهیم آماری، تقریب تغییرات به طور گسترده در جامعه آماری شناخته شده نیست. بنابراین این روش به وسیله‌ی روش‌های مونت کارلو به ویژه *MCMC* برای استنباط تقریبی تحت شعاع قرار گرفته است. تقریب تغییرات یک بدنه از تکنیک‌های

در اواسط سال ۲۰۱۰ به بعد، ما شاهد حجم بزرگ داده‌ها و سرعت افزایش آن‌ها بودیم. رشد سریع اطلاعات (داده‌ها)، یکی از چالش برانگیزترین موضوع‌های جهان امروز است، که به عنوان انقلاب داده‌های بزرگ<sup>۱</sup> انگیزه اصلی برای این پژوهش می‌باشد. سرعت افزایش اطلاعات در سال‌های اخیر، انگیزه‌ای برای توسعه تکنیک‌های تحلیلی سریع‌تر به جای روش‌های مرسوم بوده است. روش‌های محاسباتی فشرده جوابگوی این حجم عظیم از داده‌ها نخواهد بود. از این رو، روش

<sup>۲</sup>Variational Approximation

<sup>۱</sup>Big Data

اگر  $h$  یک تابع چگالی پیوسته باشد این معیار به صورت

$$KL(h, f) = \int_{-\infty}^{\infty} h(x) \log \left\{ \frac{h(x)}{f(x)} \right\} dx,$$

تعریف می‌شود. اگر  $h$  یک تابع احتمال گسسته باشد آنگاه معیار کولبک-لیبلر به صورت

$$KL(h, f) = \sum_{i=1}^n h(x_i) \log \left\{ \frac{h(x_i)}{f(x_i)} \right\}.$$

تعریف می‌شود.

## ۱.۲ ویژگی های معیار اطلاع کولبک - لیبلر

معیار کولبک-لیبلر به عنوان معیار واگرایی بین دو توزیع احتمال معرفی شده است. این معیار یک متر بر روی فضای توزیع های احتمالی نیست زیرا:

(۱) اطلاع کولبک-لیبلر متقارن نیست یعنی

$$KL(h, f) \neq KL(f, h).$$

(۲) این معیار در نامساوی مثلثی<sup>۳</sup> صدق نمی‌کند. [۱].

$$KL(h, f) \not\leq KL(h, g) + KL(g, f).$$

معیار اطلاع کولبک-لیبلر دارای ویژگی های زیر است:

$$KL(h, f) \geq 0. \quad (۱)$$

$$KL(h, f) = 0 \text{ اگر و فقط اگر } h(x) = f(x). \quad (۲)$$

این ویژگی در روند معرفی روش تقریب تغییرات بیز مورد استفاده قرار خواهد گرفت.

قطعی و غیره احتمالی با استفاده از محدودیت میانگین میدان تقریب تغییرات، برای ساخت استنباط تقریبی برای پارامترها در مدل های پیچیده آماری می‌باشد.

در این پژوهش، ابتدا در بخش اول به بیان معیار کولبک - لیبلر و ویژگی های آن پراخته شده است. پس از آن در بخش های دوم و سوم، با معرفی محدودیت میانگین میدان تغییرات بیز به بیان روش تقریب تغییرات خواهیم پرداخت. دقت برآورد مدل با استفاده از این روش در بخش چهارم بررسی شده است. در آخر به بیان یک کاربرد از این روش در نمونه تصادفی نرمال پرداخته ایم.

## ۲ معیار اطلاع کولبک-لیبلر

معیار کولبک-لیبلر ( $KL$ )، یک معیار برای ارزیابی میزان نزدیکی مدل های آماری به توزیع درست جامعه است. معیار اطلاع  $KL$  در واقع میزان ریسک احتمالی انتخاب یک مدل رقابتی به جای توزیع درست است. هرچه این معیار کوچک تر باشد، مدل انتخابی، مدل مناسب تری خواهد بود.

فرض می‌شود که  $f(x)$  یک مدل آماری ساخته شده براساس مشاهدات باشد. هدف، ارزیابی میزان نزدیکی این مدل به تابع چگالی (تابع جرم احتمال) درست  $h(x)$  براساس معیار اطلاع کولبک-لیبلر می‌باشد. این معیار به صورت

$$KL(h, f) = E_h[\log \left\{ \frac{h(X)}{f(X)} \right\}],$$

است که در آن امید ریاضی تحت تابع چگالی یا تابع جرم احتمال  $h$  است [۴].

<sup>۳</sup>Triangle Inequality

### ۳ روش تبدیل چگالی

بود:

(۲)

$$\log p(y) = \log p(y) \int q(\theta) d\theta = \int q(\theta) \log p(y) d\theta,$$

با در نظر گرفتن رابطه‌ی (۱) و قرار دادن در رابطه‌ی (۲) خواهیم داشت:

$$\int q(\theta) \log \left\{ \frac{p(y, \theta)/q(\theta)}{p(\theta|y)/q(\theta)} \right\} d\theta, \quad (۳)$$

حال با استفاده از روابط لگاریتمی در معادله‌ی (۳) داریم:

(۴)

$$\underbrace{\int q(\theta) \log \left\{ \frac{p(y, \theta)}{q(\theta)} \right\} d\theta}_{\text{کران پایین}} + \underbrace{\int q(\theta) \log \left\{ \frac{q(\theta)}{p(\theta|y)} \right\} d\theta}_{KL[q, p] > 0}$$

$$\geq \int q(\theta) \log \left\{ \frac{p(y, \theta)}{q(\theta)} \right\} d\theta.$$

نامعادله‌ی (۴) رخ می‌دهد زیرا:

$$\int q(\theta) \log \left\{ \frac{q(\theta)}{p(\theta|y)} \right\} d\theta \geq 0, \quad (۵)$$

نامعادله‌ی (۵) برای تمام چگالی‌های  $q$  به مساوی تبدیل می‌شود زمانی که:

$$q(\theta) = p(\theta | y). \quad (۶)$$

انتگرال در رابطه‌ی (۵) را واگرایی کولبک-لیبلر یا اغلب فاصله کولبک-لیبلر بین  $q$  و  $p(\cdot | y)$  می‌نامند.

یکی از روش‌های انجام تقریب تغییرات، روش تبدیل چگالی<sup>۴</sup> می‌باشد. که شامل تقریب چگالی پسین به وسیله‌ی دیگر چگالی‌ها برای استنباطی راحت می‌باشد. این روش نیازمند استفاده از واگرایی کولبک-لیبلر<sup>۵</sup> نیز می‌باشد. که در ادامه به بیان آن خواهیم پرداخت.

### ۱.۳ واگرایی کولبک لیبلر در تقریب تغییرات بیز

یک نوع مدل بیزی با بردار پارامتر  $\theta$  عضو فضای کل پارامتر  $\Theta$  و بردار مشاهدات  $y$  را در نظر بگیرید. استنباط بیزی بر اساس تابع چگالی پسین در حالت پیوسته به صورت زیر است:

$$\overbrace{p(\theta | y)}^{\text{چگالی پسین}} = \frac{p(y, \theta)}{p(y)} = \frac{p(y | \theta)p(\theta)}{\int p(y, \theta)d\theta} \quad (۱)$$

بنابراین در رابطه (۱) هدف، برآورد چگالی پسین  $p(\theta | y)$  می‌باشد. ولی ارزیابی و برآورد چگالی پسین در عمل بسیار دشوار می‌باشد. زیرا ارزیابی  $p(y)$  به دو دلیل امکان پذیر نمی‌باشد:

(۱) راه حل‌های تحلیلی در دسترس نیستند.

(۲) انتگرال‌گیری عددی بسیار وقت‌گیر و پرهزینه می‌باشد. بنابراین با استفاده از تکنیک‌هایی که در ادامه بیان خواهد شد به برآورد چگالی پسین خواهیم پرداخت. فرض کنید،  $q$  یک تابع چگالی دلخواه از فضای پارامتر  $\Theta$  باشد. آنگاه لگاریتم تابع درستنمایی حاشیه‌ای به صورت زیر خواهد

<sup>۴</sup> Density Transform Approach

<sup>۵</sup> Kullback-Leibler Divergence

### ۲.۳ کران پایین درست‌نمایی حاشیه ای

در رابطه‌ی (۴) نتیجه می‌شود که

$$p(y) = p(y; q) + KL.$$

چون مقدار  $KL$  بیشتر از صفر می‌باشد خواهیم داشت:

$$p(y) \geq p(y; q), \quad (۷)$$

به طوری که  $q$  متعلق به کران پایین (باند پایین) <sup>۶</sup> درست‌نمایی حاشیه‌ای می‌باشد که با توجه به رابطه (۴) خواهیم داشت:

$$\begin{aligned} \underline{p}(y; q) &= \exp \int q(\theta) \log \left\{ \frac{p(y, \theta)}{q(\theta)} \right\} d\theta, \\ \log \underline{p}(y; q) &= \int q(\theta) \log p(y, \theta) d\theta - \int q(\theta) \log q(\theta) d\theta, \end{aligned}$$

$$\log \underline{p}(y; q) = E_q \{ \log p(y, \theta) \} - E_q \{ \log q(\theta) \}. \quad (۸)$$

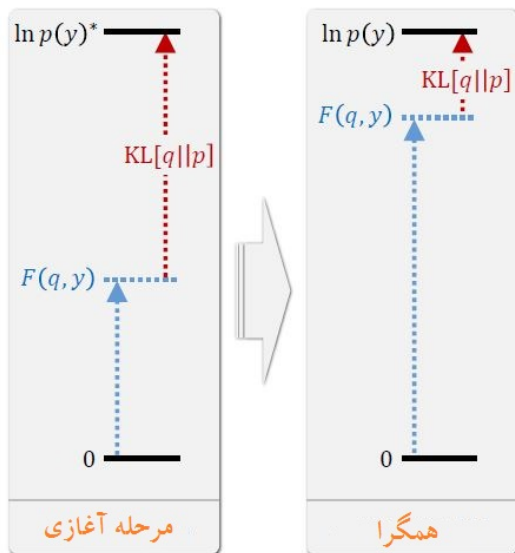
همانطور که در بخش‌های بعدی بیان خواهد شد لگاریتم کران پایین درست‌نمایی حاشیه به عنوان یک ابزار در روند روش تقریب تغییرات بیز استفاده خواهد شد.

### ۳.۳ اساس روش تبدیل چگالی

اساس روش تبدیل چگالی در تقریب تغییرات، دست یافتن به تقریبی از چگالی پسین  $p(\theta | y)$  به وسیله‌ی  $q(\theta)$  می‌باشد. زیرا استفاده از  $\underline{p}(y; q)$  راحت‌تر از  $p(y)$  می‌باشد. این راحتی ناشی از وجود محدودیت  $q$  در یک کلاس قابل کنترل بیشتر، از چگالی‌ها می‌باشد. که در این

<sup>۶</sup>Lower Bound

حالت تلاش محاسبگر در به حداکثر رساندن  $\underline{p}(y; q)$  در آن کلاس می‌باشد. مطابق (۴) به حداکثر رساندن  $\underline{p}(y; q)$



شکل ۱: همگرایی در کران پایین

معادل است با کمینه کردن واگرایی کولبک-لیبلر بین  $q$  و  $p(\theta | y)$  می‌باشد. به این منظور برای نزدیک شدن مقدار کران پایین درست‌نمایی حاشیه‌ای به مقدار اصلی درست‌نمایی حاشیه‌ای  $p(y)$ ، مطابق با شکل (۱) مقدار کران پایین درست‌نمایی حاشیه از مرحله‌ی آغازین طی انجام روندهای تکرار شوند افزایش می‌یابد و از طرفی مقدار معیار کولبک-لیبلر کاهش می‌یابد. در نهایت مقدار کران پایین درست‌نمایی حاشیه‌ای به مقدار  $p(y)$  همگرا می‌شود.

### ۴ میانگین میدان تغییرات بیز

(MFVB)

همانطور که بیان شد  $q$  یک محدودیت در کران پایین درست‌نمایی حاشیه برای انجام محاسبات ساده می‌باشد.

دو محدودیت عام برای  $q$  به صورت زیر معرفی می‌شود: از آنجایی که تابع چگالی توام  $\tilde{p}(y, \theta_1)$  به صورت زیر (۱) تجزیه کردن  $q(\theta)$  به  $\prod_{i=1}^M q_i(\theta_i)$  برای برخی از  $q$  می‌باشد: افزای‌های  $\{\theta_1, \theta_2, \dots, \theta_M\}$  از  $\theta$ .

$$(2) \text{ چگالی } q \text{ یک عضو از خانواده‌ی پارامتری توابع چگالی است.}$$

$$\tilde{p}(y, \theta_1) = \frac{\exp \int \log p(y, \theta) q_2(\theta_2) \dots q_M(\theta_M) d\theta_2 \dots d\theta_M}{\int \int \{\exp \int \log p(y, \theta) q_2(\theta_2) \dots q_M(\theta_M) d\theta_2 \dots d\theta_M\} d\theta_1 dy}$$

در مورد محدودیت اول توجه داشته باشید که به صورت یک فرم حاصلضرب فرض شده، از این رو محدودیت اولی یک نوع محدودیت ناپارامتری است. محدودیت اول اغلب به عنوان "میانگین میدان تقریب" شناخته شده و ریشه در علم فیزیک دارد. عبارت تغییرات بیز <sup>۸</sup> امری عادی برای استنباط تقریبی بیزی تحت محدودیت‌های حاصلضرب چگالی می‌باشد.

این روند برای هر یک از  $q_2, q_3, \dots, q_M$  به طور یکسان تکرار می‌شود و منجر به برآورد چگالی‌های بهینه رضایت بخش می‌شود. در نهایت چگالی‌های بهینه به صورت رابطه‌ی (۹) خواهد بود.

$$q_i^*(\theta_i) \propto \exp\{E_{-\theta_i} \log p(y, \theta)\}, \quad (9)$$

به طوری که  $E_{-\theta_i}$  به معنی مقدار امید ریاضی نسبت به چگالی  $\prod_{j \neq i} q_j(\theta_j)$  است.

اکنون با مراجعه به جزییات مدل بیزی (۲) فرض می‌کنیم که  $q$  از محدودیت اولی یا همان میانگین میدان تقریب باشد، آنگاه:

آپدیت‌ها براساس رابطه (۹) صورت می‌گیرد. این الگوریتم شامل یک حلقه می‌باشد که با اختیار کردن مقادیر اولیه‌ی

با توجه به مدل بیزی موجود، هر دو محدودیت می‌توانند اثرات با شدت کم یا شدت زیاد در نتیجه‌ی استنباط داشته باشد [۳].

مقدار ثابت +  $\int q_1(\theta_1) \log \frac{\tilde{p}(y, \theta_1)}{q(\theta_1)} d\theta_1$

$$\log p(y; q) = \int \prod_{i=1}^M q_i(\theta_i) \{ \log p(y, \theta) - \sum_{i=1}^M \log q_i(\theta_i) \} d\theta_1 d\theta_2 \dots d\theta_M,$$

حال با فاکتورگیری از  $q_1(\theta_1)$  خواهیم داشت:

حلقه‌ی (۴) شروع به چرخیدن می‌کند و برداری از چگالی‌های بروز شده از  $q$  بدست می‌آورد. در واقع مراحل الگوریتم (۱) به صورت زیر است:

$$\log p(y; q) = \int q_1(\theta_1) \log \frac{\tilde{p}(y, \theta_1)}{q(\theta_1)} d\theta_1 + \text{مقدار ثابت}$$

<sup>۷</sup>Mean Field Approximation

<sup>۸</sup>Variational Bayes

شرطی<sup>۹</sup> معرفی شده است. این فرم از چگالی‌های بهینه یک ارتباط آشکار با نمونه‌گیری گیبس است. بنابراین می‌توان از توزیع‌های تمام شرطی در روش تقریب تغییرات استفاده کرد [۵].

مقادیر اولیه :		
$q_1^*(\theta_1)$	$\leftarrow$	$\frac{\text{Exp}\{E_{-\theta_1} \log p(y, \theta)\}}{\int \text{Exp}\{E_{-\theta_1} \log p(y, \theta)\}}$
$q_2^*(\theta_2)$	$\leftarrow$	$\frac{\text{Exp}\{E_{-\theta_2} \log p(y, \theta)\}}{\int \text{Exp}\{E_{-\theta_2} \log p(y, \theta)\}}$
$\vdots$		
$q_M^*(\theta_M)$	$\leftarrow$	$\frac{\text{Exp}\{E_{-\theta_M} \log p(y, \theta)\}}{\int \text{Exp}\{E_{-\theta_M} \log p(y, \theta)\}}$

شرط توقف در الگوریتم افزایش ناچیز  $p(y; q)$  در حلقه

## ۵ دقت برآورد تقریب تغییرات بیز

همانطور که در بخش‌های قبل بیان شد، الگوریتم (۱) استنباط تقریبی سریع برای پارامترهای مدل ارائه می‌دهد. دقت برآورد با روش تقریب تغییرات بیز (MFVB) می‌تواند در محدوده بسیار عالی تا نسبتاً ناچیز قرار بگیرد. که این مقدار، وابستگی زیادی با ساختار و نوع حاصلضرب محدودیت اعمال شده در مدل خواهد داشت. اگر  $\theta$  یک نوع پارامتر تک متغیره در مدل فرض شود، آنگاه راه‌های متعددی برای اندازه‌گیری دقت یک چگالی تقریبی تغییرات بیز  $q^*(\theta)$  نسبت به چگالی پسین واقعی  $p(\theta | y)$  وجود دارد. فاصله‌ی کولبک-لیبلر یک انتخاب مناسب برای اندازه‌گیری این دقت نیز می‌باشد. بنابراین کار کردن با انتگرال قدر مطلق خطا (IAE)  $q^*(\theta)$  توصیه می‌شود. که به صورت رابطه زیر تعریف می‌شود:

$$IAE(q^*(\theta)) = \int_{-\infty}^{\infty} |q^*(\theta) - p(\theta | y)| d\theta, \quad (11)$$

مقدار عبارت (۱۱) متعلق به بازه‌ی بسته  $[0, 2]$  می‌باشد. بنابراین یک معیار برای اندازه‌گیری دقت،  $1^\circ$  به صورت

الگوریتم (۱)، با هر بار چرخش از مرحله ۱ تا مرحله  $M$ ، منجر به برآورد یک بردار به روز از چگالی‌های  $q$  می‌شود. بار دیگر با قرار گرفتن این بردار به عنوان مقادیر ثانویه منجر به چرخش دوباره‌ی الگوریتم می‌شود و این روند ادامه می‌یابد تا زمانی که افزایش در  $p(y, \theta)$  برآورد شده از هر چرخش ناچیز باشد. بنابراین بردار برآورد شده در حلقه آخر به عنوان بردار برآورد چگالی انتخاب می‌شود. لازم به ذکر است، در این روش اگر از پیشین‌های مزدوج مورد استفاده قرار گیرد آنگاه  $q_i^*$  متعلق به همان خانواده از توزیع می‌باشد و تعداد  $q_i^*$  های به روز رسانی کاهش می‌یابد. یک نمایش دیگر برای  $q_i^*(\theta_i)$  را می‌توان به صورت رابطه‌ی (۱۰) نشان داد.

$$q_i^*(\theta_i) \propto \exp\{E_{-\theta_i} \log p(\theta_i | rest)\}, \quad (10)$$

از آنجایی که

$$rest \equiv \{y, \theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_M\}.$$

$rest$  : مجموعه‌ای شامل بردارهای تصادفی در مدل به جز  $\theta_i$  می‌باشد.

در مقالات MCMC توزیع‌های  $\theta_i | rest$  به عنوان تمام

<sup>۹</sup>Full Conditionals

<sup>۱۰</sup>Accuracy

زیر تعریف می‌شود:

مهارت‌های استاندارد منجر به استفاده از تمام شرطی می‌شود. بنابراین چگالی‌های بهینه به صورت زیر بدست می‌آید:

$$\begin{aligned} Acc(q^*(\theta)) &= 1 - IAE(q^*(\theta))/supIAE(q^*(\theta)) \\ &= 1 - \frac{1}{\gamma} IAE\{q^*(\theta)\}, \end{aligned} \quad (12)$$

$$q_{\mu}^*(\mu) \propto Exp\{E_{\sigma^2} \log p(\mu | \sigma^2, x)\}.$$

$$q_{\sigma^2}^*(\sigma^2) \propto Exp\{E_{\mu} \log p(\sigma^2 | \mu, x)\}.$$

توزیع تمام شرطی  $\sigma^2 | \mu, x$  با وجود توزیع پیشین گامای معکوس، به فرم زیر می‌باشد:

از آنجایی که دقت برآورد متعلق به بازه  $[0, 1]$  می‌باشد، بنابراین دقت برآورد  $q^*(\theta)$  را می‌توان به صورت درصد بیان کرد که به صورت رابطه (۱۳) معرفی شده است [۵].

$$\begin{aligned} \pi(\sigma^2 | x) &= \pi(\sigma^2) L(\sigma^2) \propto (\sigma^2)^{-\frac{n}{\gamma}} (\sigma^2)^{-A-1} e^{-\frac{B}{\sigma^2}} \\ & e^{-\frac{1}{\sigma^2} \sum_{i=1}^n (x - \mu)_n^2} \\ & \propto (\sigma^2)^{-(A+\frac{n}{\gamma})-1} e^{-\frac{B+\frac{1}{\gamma} \|x - \mu\|_n^2}{\sigma^2}}. \end{aligned}$$

$$\begin{aligned} Acc(q^*(\theta)) &= 100 \left( 1 - \frac{1}{\gamma} \int_{-\infty}^{\infty} |q^*(\theta) - p(\theta | y)| d\theta \right) \end{aligned}$$

با توجه به توزیع گامای معکوس نتیجه می‌شود:

## ۶ کاربرد در نمونه تصادفی نرمال

$$\sigma^2 | \mu, x \sim IG(A + \frac{n}{\gamma}, B + \frac{1}{\gamma} \|x - \mu\|_n^2).$$

و همچنین با توجه به توزیع پیشین نرمال برای  $\mu$  توزیع تمام شرطی  $\mu | \sigma^2, x$  به فرم زیر می‌باشد:

اولین و دقیق‌ترین مثال از تقریب تغییرات شامل استنباط بیز است. برای آشنایی بیشتر از جزئیات آماری، یک نمونه تصادفی از توزیع نرمال را در نظر بگیرید.

$$\mu | \sigma^2, x \sim N\left(\frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu\mu}{\sigma_{\mu}^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_{\mu}^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_{\mu}^2}}\right).$$

$$x_i | \mu, \sigma^2 \sim N(\mu, \sigma^2)$$

با پیشین‌های  $\sigma^2 \sim IG(A, B)$  و  $\mu \sim N(\mu_{\mu}, \sigma_{\mu}^2)$  با تابع چگالی زیر

بنابراین با جایگذاری توزیع‌های تمام شرطی فوق در رابطه‌ی (۱۰) داریم:

$$f(\sigma^2) = \frac{B^A}{\Gamma(A)} (\sigma^2)^{-A-1} exp(-B/\sigma^2).$$

$$\begin{aligned} q_{\sigma^2}^*(\sigma^2) &\propto exp\left\{-\left(A + \frac{n}{\gamma} + 1\right) \log(\sigma^2) - \frac{B + \frac{1}{\gamma} \|x - \mu\|_n^2}{\sigma^2}\right\} \\ &\propto (\sigma^2)^{-(A+\frac{n}{\gamma}+1)} exp\left\{-\frac{B + \frac{1}{\gamma} \|x - \mu\|_n^2}{\sigma^2}\right\}, \end{aligned}$$

از طرفی تبدیل چگالی حاصلضرب تقریبی در  $(\mu, \sigma^2)$  به صورت رابطه‌ی زیر است:

$$q(\mu, \sigma^2) = q_{\mu}(\mu) \cdot q_{\sigma^2}(\sigma^2). \quad (13)$$

در نهایت توزیع چگالی‌های بهینه به صورت زیر نمایش داده می‌شود:

برنامه تکرارکننده برای بدست آوردن پارامترها در چگالی‌های بهینه  $q_{\sigma^2}^*$  و  $q_{\mu}^*$  در مثال نمونه تصادفی نرمال به صورت زیر می‌باشد.

$$q_{\sigma^2}^*(\sigma^2) \approx IG(A + \frac{n}{\nu}, B + \frac{1}{\nu} E_{\mu} \|x - \mu\|_n^2),$$

الگوریتم شماره ۲

مقدار اولیه :

$$B_{q(\sigma^2)} > 0$$

حلقه :

$$\begin{aligned} \sigma_{q(\mu)}^2 &\leftarrow \left\{ \frac{n(A + \frac{n}{\nu})}{B_{q(\sigma^2)}} - \frac{1}{\sigma_{\mu}^2} \right\}^2 \\ \mu_{q(\mu)} &\leftarrow \left\{ \frac{n\bar{x}(A + \frac{n}{\nu})}{B_{q(\sigma^2)}} - \frac{\mu_{\mu}}{\sigma_{\mu}^2} \right\} \sigma_{q(\mu)}^2 \\ B_{q(\sigma^2)} &\leftarrow \left\{ B + \frac{1}{\nu} (\|X - \mu_{q(\mu)}\|_n^2 + n\sigma_{q(\mu)}^2) \right\} \end{aligned}$$

شرط توقف در الگوریتم افزایش ناچیز  $p(y; q)$  در حلقه

که در اینجا

$$(14)$$

$$q_{\mu}^*(\mu) \approx N\left(\frac{n\bar{x}E_{\sigma^2}(\frac{1}{\sigma^2}) + \frac{\mu_{\mu}}{\sigma_{\mu}^2}}{nE_{\sigma^2}(\frac{1}{\sigma^2}) + \frac{1}{\sigma_{\mu}^2}}, \frac{1}{nE_{\sigma^2}(\frac{1}{\sigma^2}) + \frac{1}{\sigma_{\mu}^2}}\right),$$

$$E_{\mu} \|x - \mu\|_n^2 = \|x - E_{\mu}(\mu)\|_n^2 + n \cdot Var_{\mu}(u),$$

$$(15)$$

$$E_{\sigma^2}(\frac{1}{\sigma^2}) = \frac{A + \frac{n}{\nu}}{B + \frac{1}{\nu} \{ \|x - E_{\mu}(\mu)\|_n^2 + nvar_{\mu}(\mu) \}}.$$

فرم تابع چگالی‌های بهینه  $q_{\sigma^2}^*$  و  $q_{\mu}^*$  به ترتیب دارای توزیع نرمال و گاما معکوس هستند. اما لازم است پارامترهای موجود در چگالی‌ها برآورد شود.

فرض کنید:  $\sigma_{q(\mu)}^2 \equiv Var_{\mu}(\mu)$  و  $\mu_{q(\mu)} \equiv E_{\mu}(\mu)$  و  $B_{q(\sigma^2)} \equiv \frac{(A + \frac{n}{\nu})}{E_{\sigma^2}(\frac{1}{\sigma^2})}$  در (۱۴)

در این الگوریتم پارامترهای برآورد شده در هر تکرار به عنوان مقادیر اولیه در چرخه‌ی بعدی نیز قرار گرفته است. این روند ادامه یافته تا زمانی که افزایش در لگاریتم کران پایین درستی‌نمایی حاشیه‌ای ناچیز بوده است. پس از همگرایی مدل به  $\mu_{q(\mu)}^*$  و  $(\sigma_{q(\mu)}^2)^*$  و  $B_{q(\sigma^2)}^*$  تقریب‌ها برای چگالی‌های پسین هرکدام به صورت زیر خواهد بود:

$$p(\mu|x) \approx \{2\pi(\sigma_{q(\mu)}^2)^*\}^{-\frac{1}{2}} \cdot \exp\left\{-\frac{(\mu - \mu_{q(\mu)}^*)^2}{2(\sigma_{q(\mu)}^2)^*}\right\},$$

$$p(\sigma^2|x) \approx \frac{(B_{q(\sigma^2)}^*)^{A + \frac{n}{\nu}}}{\Gamma(A + \frac{n}{\nu})} (\sigma^2)^{-A - \frac{n}{\nu} - 1} \cdot \exp\left(-\frac{B_{q(\sigma^2)}^*}{\sigma^2}\right).$$

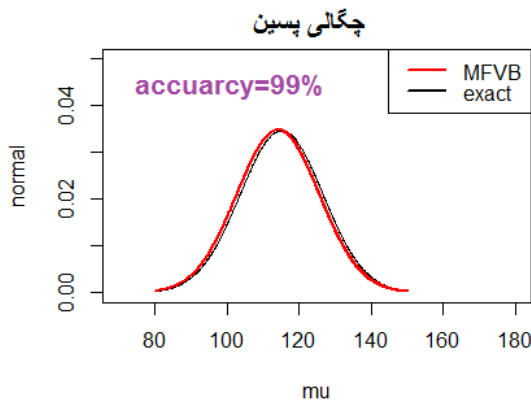
مثال ۱.۶. نمونه‌ای به اندازه  $n = 20$  از توزیع نرمال  $N(100, 225)$  شبیه‌سازی شده است. حال اگر پیشین‌های مورد استفاده به صورت زیر معرفی شود:  $\mu \sim N(0, 10^8)$  و  $\sigma^2 \sim IG(\frac{1}{100}, \frac{1}{100})$  و  $B_{q(\sigma^2)}^* = 1$  به عنوان مقدار اولیه برای انجام الگوریتم در نظر گرفته شده است. الگوریتم بعد از انجام ۴ تکرار منجر به برآورد پارامترهای

و (۱۵) به الگوریتم (۲) دست خواهیم یافت. که می‌توان برای بدست آوردن مقادیر بهینه  $\mu_{q(\mu)}$  و  $\sigma_{q(\mu)}^2$  و  $B_{q(\sigma^2)}$  مورد استفاده قرار داد. توجه داشته باشید که بیان ساده برای  $\log p(y; q)$  به صورت زیر می‌باشد:

$$\begin{aligned} \log p(x; q) &= \frac{1}{\nu} - \frac{n}{\nu} \log(2\pi) + \frac{1}{\nu} \log\left(\frac{\sigma_{q(\mu)}^2}{\sigma_{\mu}^2}\right) + A \log(B) \\ &\quad - \frac{(\mu_{q(\mu)} - \mu_{\mu})^2 + \sigma_{q(\mu)}^2}{2\sigma_{\mu}^2} - (A + \frac{n}{\nu}) \log(B_{q(\sigma^2)}) \\ &\quad + \log \Gamma(A + \frac{n}{\nu}) - \log \Gamma(A) \end{aligned}$$



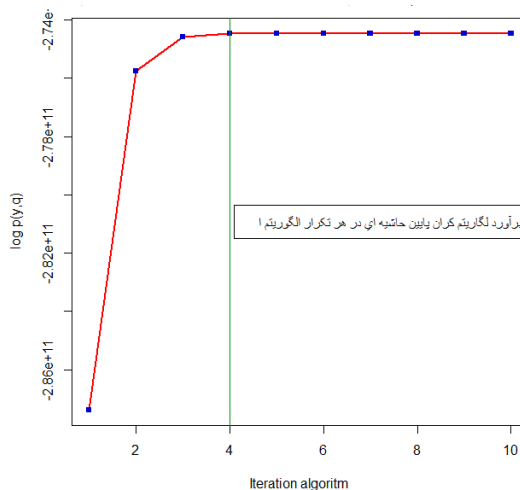
بهینه به صورت جدول زیر می شود.



شکل ۳: دقت برآورد چگالی پسین

تکرار	تکرار ۱	تکرار ۲	تکرار ۳	تکرار ۴
$\mu_q(\mu)$	۱۱۳.۲۶۳۲	۱۱۴.۲۶۲۹	۱۱۴.۲۶۲۸	۱۱۴.۲۶۲۸
$\sigma_q(\mu)$	۸۹۵.۴۰	۱۱۸۴.۵۲	۱۲۱۳.۴۱۰	۱۲۱۶.۲۹
$B_q(\sigma^2)$	۲۳۷۱۴۵.۱	۲۴۲۹۲۷.۷	۲۴۳۵۰۵.۴	۲۴۳۵۶۳.۲

شکل ۲ بیانگری سرعت برآورد پارامترهای بهینه می باشد. بطوریکه الگوریتم با ۴ تکرار به برآورد بهینه دست یافته است و افزایش  $\log p(y; q)$  از مرحله ی ۴ به بعد ناچیز بوده و الگوریتم همگرا شده است.



شکل ۲: نمودار کران پایین لگاریتم درستنمایی حاشیه ای برای ۱۰ تکرار در الگوریتم ۲

## ۷ بحث و نتیجه گیری

همانطور که بیان شد روش تقریب تغییرات یک روش بسیار سریع در برآورد پارامترها می باشد. این روش علاوه بر دارا بودن سرعت در برآورد دارای دقت بالایی نیز می باشد. موقعیت و حالاتی که تقریب تغییرات در برآورد مفید واقع شده است، نزدیک و متناظر با موقعیت و حالاتی است که  $MCMC$  مفید بوده است. روش تقریب تغییرات در انجام برآورد بسیار سریعتر از روش  $MCMC$  می باشد. به گونه ای که با صرف زمان کمتری به برآورد مورد نظر دست می توان یافت. به ویژه زمانی که مدل بسیار بزرگ و پیچیده باشد. دقت در روش تقریب تغییرات محدود است. اما در روش  $MCMC$  با افزایش سائز نمونه ای مونت کارلو دقت به طور دلخواه تغییر می کند. این تفاوت ها سبب برتری روش تقریب تغییرات نسبت به  $MCMC$  شده است [۳].

دقت برآورد در این مدل برای میانگین برابر با ۹۹ درصد می باشد. شکل ۳ نشان دهنده میزان دقت برآورد چگالی پسین  $\mu | x$  می باشد. که در این نمودار به وضوح دیده می شود که نمودار چگالی پسین واقعی و نمودار با استفاده از برآورد  $MFVB$  در الگوریتم (۲) کاملاً بر هم منطبق هستند و این منطبق بودن نمایانگر دقت بالا در این برآورد می باشد.

## مراجع

- [1] Challis, E. and Barber, D. (2013). Gaussian Kullback-Leibler approximate inference. *Journal Machine Learning Research*, 14, 2239-2286.
- [2] Euler, L. (1707). Variational Methods. *Mathematical Analysis*, 324-333.
- [3] Ormerod, J. T. and Wand, M. P. (2010). Explaining Variational Approximations. *The American Statistician*, 64, 140-153.
- [4] Sayyareh, A. (2010). Improved Kullback-Leibler Upper Bound Based on Convex Combination of  $k$  Rival Models. *Journal of Statistical Sciences, Autumn&Winter*, 2, 193-209.
- [5] Menictas, M. (2015). Variational inference for heteroscedastic and longitudinal regression models. *University of Technology: Sydney*.