بسم الله الرحمن الرحیم

پروژه درس داده کاوی

تئوری

استاد محترم درس

جناب آقای دکتر پاینده

دانشجو

آیدا اعلابیکی

آمیزه‌ای متناهی از مدل‌های آماری

در آمار یک مدل آمیزه‌ای در واقع یک مدل احتمالاتی برای نمایش وجود زیرجامعه در داخل یک جامعه‌ی کلی است، در شرایطی که هیچ‌کدام از مشاهدات نشانی دقیق دال بر تعلق به هر یک از زیرجامعه‌ها ندارند.

به طور کلی هر آمیزه‌ای متناهی از مدل‌های آماری یک توزیع آمیزه‌ای است که به شکلی دقیق توزیع جامعه‌ی تحت پوشش خود را توضیح می‌دهد.

برخی از روش‌های اجرای مدل‌های آمیزه‌ای شامل مراحلی است که هویت‌های زیر جمعیتی را براساس مشاهدات فردی (یا وزنهای مربوط به چنین زیر جمعیتی) نسبت می دهد ، که در این صورت می توان این موارد را به عنوان انواع روش‌های یادگیری یا خوشه بندی بدون نظارت در نظر گرفت. با این وجود ، تمام مراحل استنتاج شامل چنین مراحلی نیست.

مدل های آمیزه‌ای را نباید با مدل‌هایی برای داده های ترکیبی اشتباه گرفت ، یعنی داده‌هایی که اجزای آنها محدود به یک مقدار ثابت هستند (1 ، 100٪ و غیره). با این حال ، مدلهای ترکیبی را می توان مدلهای آمیزه‌ای دانست که در آن اعضای جامعه بصورت تصادفی نمونه برداری می شوند. در مقابل ، مدلهای آمیزه‌ای را می توان مدلهای ترکیبی دانست که در آن کل احتمال پوشش داده شده توسط جامعه با توجه به ضریب نرمال سازی 1 شده است.

مدل‌های داده‌های ترکیبی بخصوص زمانی که اندازه گیری های مکرر در یک سری از واحدهای آماری (مطالعه طولی) انجام یا جایی که اندازه‌گیری‌ها بر روی خوشه‌ای از واحدهای آماری مرتبط انجام می‌شود، بسیار مفید هستند.

در ادامه تعریفی از مدل آمیزه‌ای متناهی ارائه می‌دهیم.

تعریف 9-1: فرض کنید $Y_1,\dots,Y_n$ یک نمونه‌ی تصادفی به اندازه‌ی n باشد، به‌طوری که Y_j یک بردار تصادفی p بعدی با تابع چگالی احتمال $f\left(y_j;\theta_i\right)$ برای $\left(i=1,\dots,g\right)$ روی فضای R^p است. اگر $f_i\left(y_j;\theta_i\right)$ تابع چگالی j امین متغیر در i امین جامعه باشد، در آن صورت آمیزه‌ای متناهی از توابع چگالی متغیر تصادفی Y بدین صورت نوشته می‌شود:

$$f\left(y_j,\psi\right)=\sum_{i=1}^{g}\pi_i f_i\left(y_j;\theta_i\right)$$

ψ برداری شامل تمام پارامترهای نامعلوم در مدل آمیزه‌ای است و به‌صورت $\psi=\left(\pi_1,\dots,\pi_{g-1},\xi^T\right)^T$ تعریف می‌شود. $\pi_1,\dots,\pi_g$ کمیت‌هایی نامنفی هستند که به‌عنوان وزن‌ها نامیده می شوند و مقادیری را بین صفر و یک اختیار می‌کنند به‌طوری که

$$\sum_{i=1}^{g}\pi_i=1$$

مثال‌هایی از کاربرد توزیع‌های آمیخته

- یک مدل مالی

بازده مالی اغلب در شرایط عادی و در زمان بحران متفاوت رفتار می کند. یک مدل آمیزه‌ای برای داده های بازده مالی منطقی به نظر می رسد. گاهی اوقات مدل مورد استفاده یک مدل پرش-انتشار یا به عنوان ترکیبی از دو توزیع نرمال است.

- قیمت خانه

فرض کنید قیمت N خانه را مشاهده کرده‌ایم. قطعا انواع مختلف خانه در محله‌های مختلف قیمت‌های متفاوتی با یکدیگر خواهند داشت، اما قیمت یک خانه مشخص در یک محله‌ی مشخص به میانگین خوشه‌ی مربوط به محله‌ی خود نزدیک است. یک مدل ممکن برای این قیمت‌ها می‌تواند آمیزه‌ای از K توزیع باشد که هر مولفه می‌تواند یک توزیع نرمال با میانگین و واریانس مجهول باشد. هر مولفه قیمت خانه را در یک محله‌ی مشخص مدل می‌کند.
برازش این مدل به داده‌ها بااستفاده از روش EM می‌تواند پراکندگی قیمت در هر محله را مشخص کند.
(توجه داشته باشید که برای شاخص‌هایی مانند قیمت یا درآمد که قطعا مثبت هستند و تمایل به رشد به صورت نمایی دارند، توزیع لگ-نرمال ممکن است انتخاب مناسب‌تری نسبت به توزیع نرمال باشد.)

- تشخیص دست خط

مثال زیر از کتاب کریستوفر م. بیشوپ، شناخت الگو و یادگیری ماشین انتخاب شده است.
فرض کنید به ما یک تصویر N*N پیکسل سیاه سفید داده شده است که می‌دانیم یک اسکن از دست نوشته‌ی فردی است که قطعا یکی از اعداد 0 تا 9 می‌باشد، اما ما نمی‌دانیم که کدام یک از اعداد نوشته شده است.
در این‌جا ما می‌توانیم یک مدل آمیزه‌ای با K=10 مولفه در نظر بگیریم که هر مولفه یک بردار به اندازه‌ی N^2 از توزیع‌های دوجمله‌ای می باشد (به ازای هر پیکسل). چنین مدلی می‌تواند توسط الگوریتم EM برروی تعدادی دست‌نوشته‌ی بدون برچسب آموزش داده شود، سپس مدل قادر است که دست‌نوشته‌ها را با توجه به عددی که برروی آن‌ها نوشته شده است خوشه‌بندی کند. همین مدل می‌تواند پس از آموزش برای دسته‌بندی دست‌نوشته‌های جدید استفاده شود.

- ✓ در بسیاری از مسئله‌های برآورد پارامترهای مدل آماری، به شکلی از تابع درستنمایی برخورد می‌کنیم که امکان بیشینه‌سازی آن به روش تحلیلی وجود ندارد. تعداد پارامترهای زیاد، محاسبات تحلیلی را افزایش می‌دهند. در چنین مواقعی با اضافه کردن متغیر پنهان (که البته مقدار آن نیز مشاهده نشده) ممکن است مدل تابع درستنمایی، ساده‌تر شده و امکان محاسبات عددی برای بیشینه‌سازی را فراهم آورد.این تکنیک در به‌دست آوردن برآورد ML پارامترهای یک مدل آمیزه‌ای استفاده می‌شود.

الگوریتم EM یکی از روش‌هایی است که براساس وجود متغیر پنهان امکان برآورد پارامترهای مدل آماری را میسر می‌سازد. این الگوریتم اولین بار توسط «آرتور دمپستر» دانشمند آمار و همکارانش در مقاله‌ای با عنوان «حداکثر درستنمایی برای داده‌های ناکامل برمبنای الگوریتم EM» که در سال 1977 منتشر شد معرفی گردید.

6-1- تقریب تغییرات

مقدمه

تحلیل دقیق بیزی داده‌های آمیزه‌ای از نقطه نظر محاسباتی بسیار پیچیده است، زیرا تابع درستنمایی دارای فرمی گسترده با تعداد زیادی مولفه است، در عمل استفاده از نوعی تقریب برای استنباط اجتناب‌ناپذیر است و در طول سالیان وقت و انرژی زیادی در جهت طراحی روش‌هایی برای تقریب توزیع پسین و پیشگو صرف شده است.

به طور معمول اینگونه استباط‌ها برمبنای روش‌های MCMC صورت می‌گیرد، و درواقع استنباط های مبتنی بر MCMC تا حد بسیارقابل‌قبولی صحیح هستند به عبارتی توزیع تجربی به‌دست آمده از این روش تا حد بسیار زیادی به توزیع واقعی نزدیک است؛ با این‌حال در عمل مشکلات فراوانی برای پیاده‌سازی این روش وجود دارد مخصوصا زمانی که تعداد مشاهدات بسیار زیاد بوده و پارامترهای زیادی وجود داشته باشد.

درنتیجه به‌خاطر فضای ذخیره‌سازی موردنیاز، زمان موردنیاز برای محاسبات و مسائلی از این دست این روش مورد قبول واقع نمی‌شود.

مسائل مطرح شده باعث طراحی روش‌های دیگری برای شبیه‌سازی توابع چگالی پیچیده گردید که از این روش‌ها می‌توان به دو روش

1. تقریب تغییرات
2. انتشار–امید

اشاره کرد.

این نکته را مدنظر داشته باشید که نتیجه‌ی روش تقریب تغییرات دقیق نیست مگر اینکه $n \to \infty$ اما به دلیل برخی از ویژگی‌های مجانبی بسیار محبوب است.

مقدمه‌ای بر تقریب تغییرات

ایده‌ی اولیه‌ی این روش بسیار منطقی و طبیعی است، به عبارتی ایده اولیه یافتن بهترین توزیع تقریبی (q) برای تابع چگالی هدف (p) است که در این‌جا معیار بهترین بودن براساس یک فاصله بین p و q است.

معیارهای زیادی برای سنجش این فاصله وجود دارد اما یکی از مهم‌ترین این معیارها، معیار کول‌بَک–لیبلِر یا آنتروپی نسبی می‌باشد که به صورت زیر تعریف می‌شود.

$$\mathrm{KL}(q,p) = \int q \log(q/p) \qquad (1)$$

نکته : می‌توان به راحتی نشان داد که

$$\mathrm{KL}\left(p \,\|\, q\right) = -\mathrm{H}(p) + \mathrm{H}(p,q)$$

که در آن H(.) آنتروپی و H(.,.) آنتروپی متقابل است.

واضح است که معیار (1) متقارن نیست درنتیجه نمی‌تواند یک متر باشد، اما این معیار نامنفی است و صفر است اگر و تنها اگر p=q پس هرچه مقدار معیار (1) بیشتر باشد فاصله‌ی بین توزیع تقریبی (q) و توزیع دقیق(p) بیشتر است. بدون اعمال هیچ‌گونه محدودیتی برروی q واضح است که بهترین انتخاب برای q خود p می‌باشد، اما در عمل تابع چگالی p بیش از حد پیچیده است در نتیجه محدودیت‌هایی برروی q اعمال می‌شود که عمل یافتن تقریب را تسهیل کند اما q را اندکی از p دور می‌کند.
در بسیاری از کاربردها $P_D = P(.|D)$ است که در واقع تابع چگالی شرطی مقادیر مشاهده نشده (u) به شرط داده‌های مشاهده شده (D) است و q به شکلی انتخاب می‌شود که مقدار زیر را با توجه به محدودیت‌های اعمال شده برروی q، حداقل کند.

$$\mathrm{KL}\left(q, p_D\right) = \int q(\mathbf{u}) \log[q(\mathbf{u}) / p(\mathbf{u}|\ D)] d\mathbf{u}$$

می‌توان برای تابع درستنمایی نوع دوم ((P(D)) با توجه به رابطه‌ی بالا یک کران پایین یافت، که به‌صورت زیر می‌باشد.

$$\log p(D) = \int q(\mathbf{u}) \log[p(D, \mathbf{u}) / q(\mathbf{u})] d\mathbf{u} + \mathrm{KL}\left(q, p_D\right)$$

از آنجایی که معیار KL نامنفی است در نتیجه می توان نوشت

$$\log p(D) \geq \int q(\mathbf{u}) \log[p(D, \mathbf{u}) / q(\mathbf{u})] du = F\ (q) \qquad (2)$$

می‌توان نشان داد qای که مقدار $\mathrm{KL}\left(q, p_D\right)$ را حداقل می‌کند بهترین کران پایین را برای $\log\left(P\left(D\right)\right)$ فراهم می‌کند.
عبارت سمت راست رابطه‌ی (2) در فیزیک آماری به عنوان انرژی آزاد توزیع q شناخته می‌شود.
بهترین q ، انرژی آزاد را حداکثر می‌کند و همین موضوع روش تقریب تغییرات را ممکن می‌کند.
در دیدگاه فراوانی‌گرا نتایج بالا اجازه می‌دهد که یک تابع درستنمایی را در حضور داده‌ی گمشده تقریب بزنیم. در اینجا u نشان‌دهنده‌ی داده گمشده و (D،u) مجموعه‌ی کامل داده‌ها می‌باشد. در نتیجه q توزیع شرطی داده‌های گمشده به شرط مقادیر مشاهده است، و (2) یک کران پایین برای میزان درستنمایی لگاریتم داده‌های مشاهده شده فراهم می‌کند.
در دیدگاه بیزی مقادیر مشاهده نشده شامل پارامترها و داده‌های گمشده است. در این‌صورت اگر θ بردار تمام پارامترهای مدل باشد در این‌صورت داریم $u = (\theta, z)$ که در آن zنشان‌دهنده‌ی مقادیر گمشده می‌باشد که در بحث توزیع‌های آمیزه‌ای z ، وزن‌های آمیزش می‌باشد. در نتیجه q در اینجا $P\left(\theta, z \mid D\right)$ می‌باشد، که در واقع توزیع پسین پارامترها و وزن‌های آمیزش است و(2)یک کران پایین برای لگاریتم تابع درستنمایی حاشیه‌ای مشاهدات ارائه می‌کند. این تابع به عنوان تابع درستنمایی نوع دوم شناخته می‌شود.
تابع درستنمایی نوع دوم نقش مهمی در بحث فاکتور بیزی برای مقایسه مدل‌ها دارد. همان‌طور که گفته شد باید یک سری از محدودیت‌ها برروی q اعمال شود، یکی از محدودیت‌های استاندارد به صورت زیر است.

$$q(\theta, \mathbf{z}) = q^{(\theta)}(\theta) q^{(\mathbf{z})}(\mathbf{z})$$

فرض کرده‌ایم که وزن‌ها و پارامترها دارای توزیع مستقل هستند. واضح است که این فرض، یک فرضیه‌ی پایه‌ای است. توجه کنید که در اینجا یک داد و ستد بین دقت و راحتی محاسبات برقرار است و باید به شکلی دقیق بررسی شود که افزایش سرعت محاسبات تا چه حد باعث کاهش دقت می‌شود.
در اینجا $q^{(\theta)}\left(\theta\right)$ نشان‌دهنده‌ی تقریب تغییراتی توزیع پسین $P\left(\theta \middle| D\right)$ است.

باتوجه به رابطه (2) داریم

$$\log p(D) \geq \int \sum_{\mathbf{z}} q^{(\theta)}(\theta) q^{(\mathbf{z})}(\mathbf{z}) \log\left[p(D,\mathbf{z},\theta) / q^{(\theta)}(\theta) q^{(\mathbf{z})}(\mathbf{z})\right] d\theta$$

$$= F\left(q^{(\theta)}, q^{(\mathbf{z})}\right) \qquad (3)$$

که می‌توان رابطه (3) را به شکل زیر بازنویسی کرد

$$\log p(D) \geq \int q^{(\theta)}(\theta) \left\{ \sum_{\mathbf{z}} q^{(\mathbf{z})}(\mathbf{z}) \log\left[p(D,\mathbf{z}|\,\theta) / q^{(\mathbf{z})}(\mathbf{z})\right] + \log\left[p(\theta) / q^{(\theta)}(\theta)\right] \right\} d\theta$$

می‌دانیم که بهترین انتخاب برای $q^{(\theta)}$ و $q^{(z)}$ سمت راست رابطه‌ی (3) را حداکثر می‌کند. پس باید دو عبارت زیر را حداکثر کنند.

$$\int q^{(\theta)}(\theta) \log\left(\frac{\exp\left\{\sum q^{(\mathbf{z})}(\mathbf{z}) \log[p(D,\mathbf{z},\theta)]\right\}}{q^{(\theta)}(\theta)} \right) d\theta \qquad (4)$$

$$\sum q^{(\mathbf{z})}(\mathbf{z}) \log\left(\frac{\exp\left\{\int q^{(\theta)}(\theta) \log[p(D,\mathbf{z},\theta)] d\theta\right\}}{q^{(\mathbf{z})}(\mathbf{z})} \right) \qquad (5)$$

می‌توان دید که $q^{(\theta)}$ و $q^{(z)}$ باید در روابط زیر صدق کنند.

$$q^{(\theta)}(\theta) \propto \exp\left\{\sum q^{(\mathbf{z})}(\mathbf{z}) \log[p(D,\mathbf{z},\theta)]\right\} \qquad (6)$$

$$q^{(\mathbf{z})}(\mathbf{z}) \propto \exp\left\{\int q^{(\theta)}(\theta) \log[p(D,\mathbf{z},\theta)] d\theta\right\} \qquad (7)$$

حل صریح این دو معادله ممکن نیست؛ اما با نوشتن معادلات در فرم خلاصه شده $q^{(\theta)} = T^{(\theta)}\left(q^{(z)}\right),\ q^{(z)} = T^{(z)}\left(q^{(\theta)}\right)$ می‌توان یک الگوریتم تکرارشونده را نوشت، ابتدا یک نقطه‌ی اولیه به‌صورت $q^{(z)} = q^{(z)(0)}$ را انتخاب می‌کنیم حال برای m=0,1,... داریم

$$q^{(\mathbf{z})(m+1)} = T^{(\mathbf{z})}\left(q^{(\theta)(m+1)}\right)$$

$$q^{(\theta)(m+1)} = T^{(\theta)}\left(q^{(\mathbf{z})(m)}\right)$$

ساختار این الگوریتم به شکلی است که به ازای $F\left(q^{(\theta)}, q^{(z)}\right)$ (انرژی آزاد) در رابطه‌ی (3) داریم.

$$F\left(q^{(\theta)(m)}, q^{(\mathbf{z})(m)}\right) \leq F\left(q^{(\theta)(m+1)}, q^{(\mathbf{z})(m)}\right) \leq F\left(q^{(\theta)(m+1)}, q^{(\mathbf{z})(m+1)}\right) \qquad (8)$$

پس انرژی آزاد به صورت یکنوا افزایش می‌یابد و به بهترین کران پایین $\log P(D)$ نزدیک می‌شود. در نتیجه دنباله تولید شده همگراست، اگرچه تابع هدف دارای تعدادی حداکثرمحلی است و مسئله‌ی وجود حداکثرهای محلی مشکلی شایع است.

در ادامه یک مثال از کاربرد این روش در بررسی مدل‌های آمیزه‌ای ارائه می‌دهیم.

فرض کنید $D = \{y_1, \ldots, y_n\}$ نمونه‌ای تصادفی از توزیع آمیزه‌ای زیر باشد.

$$p(y|\,\theta) = \sum_{j=1}^{k} \lambda_j f_j(y)$$

که در آن f_j کاملا شناخته شده است پس بردار پارامترهای θ تنها شامل λ_jها می‌باشد. در این مثال داریم

$$p(D,\mathbf{z},\theta)=\prod_{i=1}^{n}\prod_{j=1}^{k}\left(\lambda^{z_{ij}}f_{ij}^{z_{ij}}\right)p(\theta)$$

که در آن $z_{ij}=1$ اگر و تنها اگر مشاهده‌ی iام از مولفه‌ی jام باشد، $f_{ij}=f_j\left(y_i\right)$ و $P\left(\theta\right)$ توزیع پیشین برای θ است؛ اگر ما فرض کنیم $\mathrm{Dir}\left(a^{(0)}\right)$ توزیع پیشین θ باشد بنابراین

$$p(D,\mathbf{z},\theta)\propto\prod_{j=1}^{k}\lambda_j^{\sum_i z_{ij}+a_j^{(0)}-1}$$

بنابراین

$$\sum_{\mathbf{z}}q^{(\mathbf{z})}(\mathbf{z})\log[p(D,\mathbf{z},\theta)]=\sum_{j=1}^{k}\left(\sum_i q_{ij}^{(\mathbf{z})}+a_j^{(0)}-1\right)\log\lambda_j+\text{const}$$

که در آن const مستقل از θ و $q_{ij}^{(z)}$ احتمال حاشیه‌ای برای $z_{ij}=1$ طبق توزیع $q^{(z)}\left(z\right)$ می‌باشد.

از (8) داریم که بهینه‌ترین توزیع برای $q^{(\theta)}\left(\theta\right)$ توزیع $\mathrm{Dir}\left(a\right)$ با پارامتر $a_j=\sum_i q_{ij}^{(z)}+a_j^{(0)}$, $i=1,\ldots,k$ می‌باشد؛

در ادامه توزیع بهینه برای $q^{(z)}\left(z\right)$ را برحسب تابعی از $q^{(\theta)}\left(\theta\right)$ به‌دست می‌آوریم؛ داریم

$$\int q^{(\theta)}(\theta)\log[p(D,\mathbf{z},\theta)]d\theta=\sum_{i=1}^{n}\sum_{j=1}^{K}z_{ij}\left[\log\left(\phi_j^{(q)}\right)+\log f_{ij}\right]+\text{const}$$

که در آن داریم

$$\phi_j^{(q)}=\exp\left[E_{q^{(\theta)}}\log\left(\lambda_j\right)\right]$$

و در اینجا const مستقل از $\left\{z_{ij}\right\}$ می‌باشد.

اگر عبارت حاصل شده را در (7) جاگذاری کنیم مشاهده می‌کنیم که توزیع بهینه برای $q^{(z)}\left(z\right)$ به شکل یک فرم ضربی درمی‌آید که در آن هر مشاهده برای خود مضربی دارد و مضارب بهینه برای توزیع بهینه به‌شکل زیر هستند

$$q_{ij}^{(\mathbf{z})}\propto f_{ij}\phi_j^{(q)}$$

و داریم $\sum_j q_{ij}^{(z)}=1$, $\forall i,\ \ j=1,\ldots,k$ ، از خواص توزیع دیریکله داریم $E_{q^{(\theta)}}\left[\log\lambda_j\right]=\Psi\left(a_j\right)-\Psi\left(a_.\right)$ که در آن Ψ نشان‌دهنده‌ی تابع گاما است و $a_.=\sum_j a_j$ ؛ بنابراین داریم

$$q_{ij}^{(\mathbf{z})}=f_{ij}\exp\left[\Psi\left(a_j\right)-\Psi\left(a_.\right)\right]/\sum_{r=1}^{k}f_{ir}\exp\left[\Psi\left(a_r\right)-\Psi\left(a_.\right)\right]$$

همان‌طور که می‌شد حدس زد معادلات کاملا تحت‌تاثیر $\left\{a_j\right\}$ هستند و $\left\{q_{ij}^{(z)}\right\}$ به صورت صریح قابل محاسبه نیست اما الگوریتم تکرار شونده به صورت زیر مسئله را حل می‌کند.

فرض کنید به عنوان نقطه‌ی ابتدایی ، $q_{ij}^{(z)(0)}=f_{ij}a_j^{(0)}\Big/\left(\sum_{r=1}^{k}f_{ir}a_r^{(0)}\right)$ را انتخاب کنیم، در نتیجه گام‌های زیر را برای $m=0,1,\ldots$ داریم

گام 1 : برای $j=1,\ldots,k$ مقدار زیر را محاسبه کن

$$a_j^{(m+1)} = \sum_i q_{ij}^{(\mathbf{z})(m)} + a_j^{(0)}$$

گام 2 : برای $i = 1,\dots,n$ و $j = 1,2$ مقدار زیر را محاسبه می‌کنیم.

$$q_{ij}^{(\mathbf{z})(m+1)} = f_{ij}\exp\left[\Psi\left(a_j^{(m+1)}\right) - \Psi\left(a_{\cdot}^{(m+1)}\right)\right] / \sum_{r=1}^{k} f_{ir}\exp\left[\Psi\left(a_r^{(m+1)}\right) - \Psi\left(a_{\cdot}^{(m+1)}\right)\right]$$

به این نکته جالب توجه کنید که گام 1 الگوریتم مشابه گام M و گام 2 مشابه گام E الگوریتم EM است، تفاوت عمده‌ی دو الگوریتم جایگزینی $q_{ij}^{(z)(m)}$ به‌جای $z_{ij}^{(m)}$ است.

در برخی از کتب به این الگوریتم EM تغییراتی گفته می‌شود.

نتایج مجانبی

تقریب تغییرات بیزی برای بسیاری از مسائل طراحی شده است، اما مزایای واقع‌بینانه‌ی آن باید براساس مبانی تئوری مورد بررسی قراربگیرد.

آتیاس (1999 ، 2000) و پنی و رابرتز (2000) ادعا کردند که در مسائل خاص، برآورد بیزی تغییراتی، برمبنای میانگین پسین تغییراتی، در بزرگ نمونه به برآوردگر حداکثر درستنمایی میل می‌کند، اما هیچ اثبات دقیقی ارائه نشد. اخیرا تحقیقات بیشتری در این زمینه انجام شد، وانگ و تیترینگتون (2004) ثابت کردند در آمیزش k توزیع شناخته شده، برای حالتی که توزیع پسین وزن‌ها دیریکله باشد، برآورد بیزی تغییراتی (میانگین توزیع پسین تغییراتی) برای وزن‌های آمیزش به صورت مجانبی برابر با برآورد حداکثر درستنمایی می‌شود.

وانگ و تیترینگتون (2005) اثبات قبلی خود را بهبود بخشیده و ثابت کردند توزیع پسین مجانبی تغییراتی پارامترها، نرمال است.

درنتیجه می توان گفت در عموم مواقع تقریب تغییراتی به صورت مجانبی رفتار معقولی دارد، اگرچه مقالاتی که به بررسی رفتار مجانبی ماتریس واریانس-کوواریانس پرداخته است نشان داده است که برآورد حاصل از طریق روش تقریب تغییرات "بسیار کوچک‌تر" از برآورد حاصل از طریق روش حداکثر درستنمایی است؛ به عنوان یک نتیجه از این موضوع داریم که برآورد فاصله‌ای حاصل از روش تقریب تغییرات برای پارامترها به شکل غیرواقعی طول کمی خواهد داشت.

به عنوان مثالی از خواص مجانبی، برای $k=2$ (توزیع آمیخته دو توزیع) واریانس توزیع تقریبی تغییراتی برای پارامتر وزن λ برابر است با $\lambda(1-\lambda)/n$ که دقیقا برابر واریانس حاصل در روش حداکثر درستنمایی برمبنای داده‌های کامل است.

وانگ و تیترینگتون (2004) یک خانواده از توزیع‌های نمایی با مقادیر گمشده را در نظر گرفتند و همگرایی الگوریتم تکراری برای محاسبه‌ی تقریب تغییراتی را اثبات کردند؛ همچنین آن‌ها ثابت کردند که توزیع مجانبی برآوردگرها نرمال است.

مشاهدات نشان‌داده است در بحث زنجیرهای مارکوف پنهان استفاده از فرم کاملا ضربی $q^{(z)}$ نتایج مطلوبی برای تقریب تغییراتی حاصل نمی‌کند.