تحلیل مدل بیز سلسله مراتبی چند جمله‌ای سه مرحله‌ای[1]

بالگوبین ناندرام[2]

مترجم: امید حمیدی[3]

چکیده

مسئله‌ی ادغام داده‌ها از چند جامعه‌ی چند جمله‌ای با استفاده از مدل بیز سلسله مراتبی سه مرحله‌ای را در نظر بگیرید. روش‌های شناخته‌شده‌ای برای ادغام داده‌های چند جامعه‌ی دو جمله‌ای وجود دارد و در صورت نبود روش مبتنی بر نمونه‌گیری، معمولاً پارامترها معلوم فرض می‌شوند یا حتی بدتر از آن تقریب‌هایی بدون در نظرگرفتن عدم حتمیت در مورد درستی آن‌ها لحاظ می‌شود. در این جا با استفاده از یک پیشین دیریکله برای مدل‌بندی احتمال‌های سلول‌ها، یک تحلیل بیز سلسله مراتبی کامل برای جوامع چند جمله‌ای بدون هیچ تقریب جبری و بدون ثابت و معلوم فرض کردن ابر پارامترها ارائه می‌شود. این روش نیاز به

---

[1] *این مقاله در نشریه‌ی آمار، رایانه، مانستگر شماره‌ی ۶۱- ۱۹۹۸ منتشر شده است.*

[2] *Balgobin Nandram*

[3] *دانشجوی کارشناسی ارشد آمار اقتصادی و اجتماعی دانشگاه شهید بهشتی*

محاسبه‌های پیچیده‌ای دارد که با استفاده از الگوریتم متروپلیس هستینگس انجام می‌شود. برای نشان دادن مزیت‌های محاسبات، سه کاربرد شرح داده می‌شود. کاربرد نخست یک مطالعه‌ی شبیه‌سازی بر اساس استنباط بیز پیشگو است. همچنین برای مدل بیز کامل در مطالعه‌ی شبیه‌سازی نشان می‌دهیم که روش ما بهتر از یک روش جایگزین با استفاده از چگالی پیشنهادی متفاوت است. با انتخاب بهترین جامعه‌ی چند جمله‌ای از کاربرد دوم برای مقایسه‌ی مدل‌های لوژستیک خطی و لوژستیک غیرخطی استفاده می‌کنیم. در کاربردهای اول و دوم نشان می‌دهیم که زمانی که ابر پارامترها معلوم فرض می‌شوند در مورد تغییرپذیری، کم براوردی ناخالص وجود دارد. در کاربرد سوم تقریب لاپلاس و نمونه‌گیر گیبس با الگوریتم متروپلیس هستینگس مقایسه می‌شوند. روش ما حتی برای جداول با پراکندگی بسیار زیاد به خوبی عمل می‌کند.

واژه‌های کلیدی: استنباط بیز پیشگو؛ پیشین دیریکله؛ ابر پارامتر؛ مدل لوژستیک؛ الگوریتم متروپلیس ـ هستینگس؛ جداول پراکنده.

مقدمه

مسئله‌ای که به وفور در اپیدمیولوژی، تعلیم و تربیت و علوم اجتماعی رخ می‌دهد براورد نسبت‌های دو جمله‌ای در گروه‌های مختلف است که حالت ویژه‌ای از مسئله‌ی عمومی براورد نسبت‌های چند جمله‌ای است. چندین روش برای تحلیل نسبت‌های دو جمله‌ای پیشنهاد شده است، اما در مورد مسئله‌ی چند جمله‌ای چندان بحث نشده است. مشکل اصلی در این باره این است که کاربرد ابرپارامترهای مدل چند جمله‌ای ـ دیریکله برای یک تحلیل بیز کامل، دشوار است. بنابراین بیزین‌ها به طور معمول ابر پارامترها را معلوم فرض می‌کنند. مشکل دیگر پراکندگی است که در مورد داده‌های چند جمله‌ای با تعداد سلول‌های زیاد به وجود می‌آید. بنابراین هدف توسعه‌ی تحلیل بیز کامل مدل چند جمله‌ای ـ دیریکله بدون در نظرگرفتن معلوم بودن پارامترها است.

سابقه‌ی گسترده‌ای در مورد روش بیز برای تحلیل داده‌های رسته‌ای وجود دارد. یک بررسی منتخب در مورد این سابقه برای تحلیل بیز داده‌های رسته‌ای توسط لئونارد

و هسو[1] (۱۹۹۴) توضیح داده شده است. به ویژه براورد بیز برای چند توزیع دو جمله‌ای (لئونارد، ۱۹۷۲) که به چند جمله‌ای (لئونارد، ۱۹۷۷) تعمیم داده شده، بیان شده است.

بخش بیش‌تر روش بیزی در مورد تحلیل جدول پیش‌آیندی باید به مدل‌های خطی سلسله مراتبی بیزی (لیندلی و اسمیت[2]، ۱۹۷۲) با مراحل و پیشین‌های نرمال گوناگون منتسب شود. برای نمونه لئونارد (۱۹۷۲) از تبدیل لوجیت روی احتمال‌های دو جمله‌ای استفاده کرد و نویک[3] و همکاران (۱۹۷۳) تبدیل آرک سینوس را روی داده‌ها و احتمال‌های دو جمله‌ای برای در نظرگرفتن یک فرمول‌بندی بیز سلسله مراتبی مشابه مورد استفاده قرار دادند. متأسفانه درستی این تقریب‌ها نامطمئن است و به طور معمول در چارچوب روش مبتنی بر نمونه‌گیری، حتی مدل‌های پیچیده‌تر قابل بررسی است.

آلبرت و گوپتا[4] (۱۹۸۳) بدون هیچ تقریبی، تشابه در یک مجموعه داده‌ی دو جمله‌ای را با استفاده از پیشین بتای مزدوج که در آن مجموع پارامترهای بتا با همبستگی مرتبط است وارد کردند. هر چند، آن‌ها فرض کردند که مجموع پارامترهای بتا معلوم است. به آلبرت و گوپتا (۱۹۸۵) که فرض مشابهی را بیان کردند مراجعه کنید. ناندرام و سدرانسک[5] (۱۹۹۳) با استفاده از توزیع‌های بتا به عنوان پیشین کار آلبرت و گوپتا را به داده‌های رسته‌ای تعمیم دادند. دوباره آن‌ها فرض کردند که مجموع پارامترهای بتا معلوم است. حال فرض معلوم بودن مجموع پارامترهای بتا را می‌توان حذف کرد.

---

1. Leonard and Hsu
2. Lindley and Smith
3. Novick
4. Albert and Gupta
5. Sedransk

در همه‌ی کارهای بحث شده در مورد سابقه‌های مربوط به چند توزیع دو جمله‌ای (یا کلی‌تر چندین توزیع چند جمله‌ای) بجز در مورد وجود متغیرهای کمکی (که در آن رگرسیون لجستیک استفاده می‌شود) از فرض تعویض‌پذیری پارامترهای مرحله‌ی اول به شکلی (لجستیک در برابر بتا) استفاده می‌شود. برای نمونه ونگ و ماسون[1] (۱۹۸۵) را ببینید. ما فرض تعویض‌پذیری را ادامه می‌دهیم زیرا در کنترل تغییرات برون‌زا سودمند است. به طور کلی مسئله‌ی اصلی با راهبرد سلسله مراتبی در روش‌های بیزی این است که ابر پارامترها را چگونه مشخص کنیم. برای نمونه، لئونارد (۱۹۷۷) در مدل چند جمله‌ای ـ دیریکله، دو تقریب مهم را به منظور ترکیب کردن عدم حتمیت ابر پارامترها انجام داد. یکی از این تقریب‌ها روی تابع درستنمایی و دیگری روی توزیع پسین اجرا شد. اگر از این تقریب‌ها استفاده نشود بدون شک تحلیل‌ها و بررسی‌ها پیچیده‌تر خواهد شد. شایان توجه است که کاس و استفی[2] (۱۹۸۹) تقریب لاپلاس و جرج[3] و همکاران (۱۹۹۴) نمونه‌گیری گیبس را برای مدل دو جمله‌ای ـ بتا به کار بردند. زمانی که عدم حتمیت در مورد ابر پارامترها به مدل اضافه می‌شود تحلیل مدل چند جمله‌ای ـ دیریکله مشکل می‌شود.

برای حل این مسئله نشان می‌دهیم که محاسبه‌های دقیق چگونه انجام می‌شود و مدل را در سه مورد مهم به چه شیوه‌ای به کار می‌بریم. با به کارگیری یک روش مبتنی بر نمونه‌گیری از تقریب‌های لئونارد (۱۹۷۷) اجتناب می‌شود. با استفاده از الگوریتم متروپلیس‌ـ هستینگس (چیب و گرینبرگ[4]، ۱۹۹۵) یک براورد از توزیع پسین مرتبط به دست می‌آید. متأسفانه حتی در داخل چارچوب مبتنی بر نمونه‌گیری نیز موانعی برای به دست آوردن توزیع پسین وجود دارد. البته این محاسبات مسئله‌ی مدل دو جمله‌ای ـ بتای سلسله مراتبی را حل می‌کند. با استفاده از الگوریتم متروپلیس ـــ

---

[1] Wong and Mason

[2] Kass and Steffey

[3] George

[4] Chib and Greenberg

هستینگس محاسبات لئونارد را برای به دست آوردن براوردهای توزیع پسین بهبود می‌بخشیم. افزون بر این، به این ترتیب یک جواب بیز کامل برای مدل چند جمله‌ای سلسله مراتبی سه مرحله‌ای به دست می‌آید.

ادامه‌ی مقاله به شرح زیر تنظیم شده است. در بخش (۲) روش کار را شرح می‌دهیم. به ویژه مدل چند جمله‌ای بیز سلسله مراتبی سه مرحله‌ای شرح داده می‌شود و الگوریتم متروپلیس ـ هستینگس برای تحلیل مدل سلسله مراتبی سه مرحله‌ای ساخته می‌شود. مسائل آماری اصلی، به دست آوردن دو چگالی است که به عنوان چگالی‌های مولد به کار می‌روند که این امر با تقریب دو توزیع پسین شرطی با استفاده از چگالی‌هایی که به سادگی از آن‌ها نمونه می‌گیریم انجام می‌شود.

روش کار

در این بخش مدل دو مرحله‌ای استاندارد که برای تحلیل داده‌های رسته‌ای به کار می‌رود در مورد سه مرحله که در براورد ابر پارامترهای آن قطعیت وجود ندارد، تعمیم داده می‌شود.

ـ مدل‌سازی

فرض کنید n_{ij} شمارش سلول‌ها، p_{ij} احتمال‌های مربوط به سلول‌ها که $i = 1,\dots,I$ و $j = 1,\dots,J$، $n_{i\circ} = \sum_{j=1}^{J} n_{ij}$، $p_i = (p_{i1},\dots,p_{iJ})'$، $\sum_{j=1}^{J} p_{ij}$ و $n_i = (n_{i1},\dots,n_{iJ})'$ باشند. همچنین فرض کنید:

$$n_i \big| p_i \sim_{ind} Multinomial(n_{i\circ}, p_i),\ \ i = 1,2,\dots,I \tag{1}$$

که در آن:

$$f(n_i \mid p_i) = n_{i\circ}! \prod_{j=1}^{J} \left\{ \frac{p_{ij}^{n_{ij}}}{n_{ij}!} \right\},\ \ 0 < n_{ij} < n_{i\circ},\ \ j = 1,2,\dots,J$$

در مرحله‌ی دوم حالت زیر را در نظر می‌گیریم:

$$p_i | \mu, \tau \sim_{iid} Dirichlet(\mu\tau), \quad i = 1, 2, \ldots, I \qquad (2)$$

که در آن:

$$\pi(p_i \mid \mu, \tau) = \frac{\prod_{j=1}^{J} p_{ij}^{\mu_j \tau - 1}}{D(\mu\tau)} \quad 0 < p_{ij} < 1 \ , \ \sum_{j=1}^{J} p_{ij} = 1 \ , \ \mu = (\mu_1, \mu_2, \ldots, \mu_J)'$$

در حالی که $D(\mu\tau) = \left\{\prod_{j=1}^{J} \Gamma(\mu_j \tau)\right\} / \Gamma(\tau)$، $0 < \mu_j < 1$ و $\sum_{j=1}^{J} \mu_j = 1$ است. دوباره پارامتری کردن توزیع دیریکله در (۲) سودمند است زیرا همبستگی بین μ و τ را هنگامی که براورد می‌شوند، از بین می‌برد. همچنین مدل مشخص شده با (۱) و (۲) را می‌توان در مورد بسیاری از مسائل داده‌های رسته‌ای به کار برد. برای اطمینان از یک تحلیل بیز کامل در مرحله‌ی سوم فرض می‌کنیم که:

$$\mu \sim Dirichlet(\mu^{(0)} \tau^{(0)}), \ \ 0 < \mu_j < 1 \ , \ \sum_{j=1}^{J} \mu_j = 1 \ , \qquad (3)$$
$$\mu^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)}, \ldots, \mu_J^{(0)})'$$

و

$$\tau \sim \Gamma(\eta^{(0)}, \nu^{(0)}), \ \ \tau > 0 \qquad (4)$$

که در آن $\mu^{(0)}$، $\tau^{(0)}$، $\eta^{(0)}$ و $\nu^{(0)}$ باید مشخص شوند. در رابطه‌ی (۴) فرض می‌شود:

$$\pi(\tau)=(\nu^{(\circ)})^{\eta^{(\circ)}}\tau^{\eta^{(\circ)}-1}e^{-\nu^{(\circ)}\tau}/\Gamma(\eta^{(\circ)}),\quad \tau>\circ$$

بنابر این (۱)ـ (۴) یک مدل بیز سلسله مراتبی سه مرحله‌ای را برای داده‌های چند جمله‌ای فراهم می‌کند. ابر پارامترهای $j=1,2,\dots,J$ ، $\mu_j^{(\circ)}$ ، $\tau^{(\circ)}$ ، $\eta^{(\circ)}$ و $\nu^{(\circ)}$ را می‌توان به وسیله‌ی استخراج به دست آورد. ولی، بدون هیچ اطلاعات پیشین درباره‌ی μ و τ می‌توان $\mu_j^{(\circ)}\tau^{(\circ)}=1$ که $j=1,2,\dots,J$ ، $\eta^{(\circ)}=\circ$ و $\nu^{(\circ)}=\circ$ را درباره‌ی یک توزیع پیشین یکنواخت روی μ و پیشین ناآگاهی بخش روی τ در نظر گرفت. مشکل اصلی در مدل به دست آوردن توزیع پسین توأم μ و τ است.

هر چند، با استفاده از قضیه‌ی بیز نتیجه می‌شود که توزیع پسین توأم μ و τ به شرط $n=(n_1',n_2',\dots,n_I')'$ به صورت زیر است:

$$\pi(\mu,\tau\mid n)\propto\prod_{i=1}^{I}\left\{\left(\prod_{j=1}^{J}\frac{\Gamma(n_{ij}+\mu_j\tau)}{\Gamma(\mu_j\tau)}\right)/\left(\frac{\Gamma(n_{i\circ}+\tau)}{\Gamma(\tau)}\right)\right\} \tag{۵}$$

$$\left\{\prod_{j=1}^{J}\mu_j^{\mu_j^{(\circ)}\tau^{(\circ)}-1}\right\}\times\{\tau^{\eta^{(\circ)}-1}e^{-\nu^{(\circ)}\tau}\},\ \circ<\mu_j<1,\ \sum_{j=1}^{J}\mu_j=1,\ \tau>\circ$$

توزیع پسین (۵) به فرم بسته وجود ندارد، زیرا ثابت تناسب به یک مقدار ثابت همگرا نمی‌شود. به طور معمول هدف اصلی در بسیاری از مسائل به دست آوردن چگالی توأم $p=(p_1',p_2',\dots,p_I')'$ به شرط n به جای چگالی توأم μ و τ به شرط n است.

بنابر این استنباط کامل درباره‌ی p با استفاده از

$$\pi(p\mid n)=\int_S\int_\circ^\infty\left\{\prod_{i=1}^{I}\pi(p_i\mid\mu,\tau,n)\right\}\pi(\mu,\tau\mid n)\,d\tau\, d\mu \tag{۶}$$

به دست می‌آید که $S=\{\mu: ۰<\mu_j<۱, j=۱,۲,\ldots,J, \sum_{j=۱}^{J}\mu_j=۱\}$ و

$$p_i \mid \mu,\tau,n \sim_{ind} Dirichlet(\mu^*\tau_i^*),\ i=۱,۲,\ldots,I \qquad (۷)$$

در حالی که $\mu_{ij}^*=(\mu_j\tau+n_{ij})/\tau_i^*$ و $\tau_i^*=\tau+n_{i.}$ ، $i=۱,۲,\ldots,I$ و $j=۱,۲,\ldots,J$.

وقتی که متغیر کمکی وجود ندارد مدل بیز سلسله مراتبی سه مرحله‌ای داده شده در (۱)ـ (۴) ترجیح داده می‌شود. روش‌های دیگری نیز وجود دارد که در آن‌ها از لوجیت‌های چند جمله‌ای استفاده می‌شود (برای مثال لئونارد[۱]، ۱۹۷۵). اما چنین روش‌هایی برای مسائلی بدون متغیر کمکی لازم نیست. وقتی که نمونه‌هایی از $\pi(\mu,\tau \mid n)$ به دست می‌آید، نمونه‌هایی از $\pi(p \mid n)$ را نیز می‌توان به روش مستقیم به دست آورد.

ـ الگوریتم متروپلیس ـ هستینگس

فرض کنید $\mu_{(j)}$ نشان‌دهنده‌ی بردار همه‌ی اجزای μ بجز عضو jام باشد که $j=۱,۲,\ldots,J$. آنگاه توزیع پسین شرطی $\mu_j \mid \mu_{(j)},\tau,n, j=۱,۲,\ldots,J-۱$ و توزیع پسین شرطی $\tau \mid \mu,n$ به دست می‌آید. توجه کنید که توزیع پسین شرطی μ_J با استفاده از $\mu_J=۱-\sum_{j=۱}^{J-۱}\mu_j$ به دست می‌آید.

از رابطه‌ی (۵) یک نمونه به روش زیر به دست می‌آید که μ_j با استفاده از $\sum_{i=۱}^{I} n_{ij} / \sum_{i=۱}^{I}\sum_{j=۱}^{J} n_{ij}$ براورد می‌شود، یک برامد از توزیع پسین شرطی $\mu_j \mid \mu_{(j)},\tau,n, j=۱,۲,\ldots,J$ به دست می‌آید و سپس یک برامد از توزیع پسین شرطی $\tau \mid \mu,n$ حاصل می‌شود. بعد از دوره‌ی سوختن اولیه، فرایند را M بار برای به

---

[۱] Leonard

دست آوردن $(\mu^{(h)}, \tau^{(h)}), h=1,2,\ldots,M$ تکرار می‌کنیم. متأسفانه توزیع‌های پسین شرطی $\mu_j \mid \mu_{(j)}, \tau, n$ و $\tau \mid \mu, n$ به فرم‌های بسته وجود ندارند و بنابراین استفاده از نمونه‌گیر گیبس (گلفند و اسمیت، ۱۹۹۰) مشکل است. جرج و همکاران (۱۹۹۴) با استفاده از نمونه‌گیری گیبس، یک تحلیل بیز کامل سلسله مراتبی سه مرحله‌ای را برای خانواده‌ی توزیع‌های نمایی در نظر گرفتند. آن‌ها در مورد مدل دو جمله‌ای ـ بتای سه مرحله‌ای به عنوان یک حالت ویژه بحث کردند. در این کاربردها به دست آوردن ابر پارامترها مشکل است و توزیع‌های پیشین به گونه‌ای ساخته می‌شوند که توزیع‌های پسین شرطی آن‌ها لگ مقعر باشند تا کاربرد الگوریتم نمونه‌گیری ردی انطباقی گیلکز و وایلد[1] (۱۹۹۲) امکان‌پذیر باشد. در مدل چند جمله‌ای با چندین سلول ممکن است کسی بخواهد هر ابر پارامتر را با استفاده از یک الگوریتم ARS به دست آورد. بجز برای به دست آوردن پسین شرطی برای τ، نیاز به لگ مقعر بودن نداریم زیرا می‌توانیم توابع مهم بسیار کارایی برای توزیع پیشین μ_j بسازیم. حتی برای دریافتن مد به لگ مقعر بودن نیازی نداریم.

افزون بر این، الگوریتم متروپلیس ـ هستینگس مفیدتر از الگوریتم گیبس است، زیرا الگوریتم متروپلیس ـ هستینگس نیازمند، به دست آوردن توزیع‌های پسین‌شرطی p_i نیست. این گام در تحلیل خروجی رخ خواهد داد. بنابراین، اگر جوامع چند جمله‌ای زیادی برای مثال در براورد ناحیه‌ی کوچک داشته باشیم زمان زیادی صرفه‌جویی می‌شود.به این ترتیب الگوریتم متروپلیس ـ هستینگس که در این جا شرح داده شده ساده‌تر است و در واقع گسترش کار جرج و همکاران (۱۹۹۴) است؛ تیرنی (۱۹۹۴) را ببینید. ضرورتاً از اصل کرنل استفاده می‌کنیم که به دست آوردن موفقیت‌آمیز برامد از هر توزیع پسین را به جای اجرای هر یک از توزیع‌های پسین شرطی برای همگرایی برای هر مقدار از متغیرهای شرطی امکان‌پذیر می‌سازد. (چیب و گرین برگ، ۱۹۹۵ را برای دیدن یک مثال ساده ببینید).

---

[1] Gilks and Wild

با در نظر گرفتن $a_j = 1 - \sum_{s=1, s\neq j}^{J-1} \mu_s$ و $n_{\circ j} = \sum_{i=1}^{I} n_{ij}$ نانـدرام (۱۹۹۸) نشـان داد که

(۸)

$$\pi(\mu_i \mid \mu_{(i)}, \tau, n) \propto \prod_{i=1}^{I} \left\{ \frac{\Gamma(n_{ij} + \mu_j \tau)}{\Gamma(\mu_j \tau)} \frac{\Gamma(n_{iJ} + \mu_J \tau)}{\Gamma(\mu_J \tau)} \right\}$$
$$\times \left(\frac{\mu_j}{a_j} \right)^{\mu_j^{(\circ)} \tau^{(\circ)} - 1} \left(1 - \frac{\mu_j}{a_j} \right)^{\mu_J^{(\circ)} \tau^{(\circ)} - 1}, \quad \circ < \mu_j < a_j \ , \ j = 1, 2, \ldots, J-1$$

(توجه کنید که μ_J بخشی از مجموع که a_j تعریف شده، نیست).

قابل توجه است که توزیع شرطی μ_j برای هر $j = 1, 2, \ldots, J-1$ بـه سـلول J بسـتگی دارد. اگر هر دوی j امین سلول که $j = 1, 2, \ldots, J-1$ و J امین سلول شمارش‌هـای صـفر داشته باشند، مشکلاتی به وجود می‌آید. بنابراین نکته‌ی اساسی در محاسبه‌هـا بـه ویـژه برای جداول پراکنده مرتب کردن سلول‌ها است، به طوری که J امین سلول یک شـمارش غیر صفر و در صورت امکان بزرگ‌تر از سایر سلول‌ها در هر توزیع چنـد جملـه‌ای داشـته باشد. البته این مرتب کردن باید سازگار با جوامع چند جمله‌ای صورت گیرد. بعد از ایـن که متروپلیس ـ هستینگس کامل شد، سلول‌ها در تحلیل خروجی براساس ترتیب اصـلی سلول‌ها مرتب می‌شوند. این سیاست زمانی که آخرین سلول پراکنده اسـت در مثـال‌هـا مورد استفاده قرار می‌گیرد.

با فرض این که نمونه‌گیر متـروپلیس ـــ هسـتینگس در $\mu_j^{(r)}$ قـرار دارد، آنگـاه احتمـال پذیرش $\mu_j^{(r+1)}$ برابر است با

$$U_{r,r+1} = \min \left\{ \frac{\psi(\mu_j^{(r+1)}, \mu_j^{(r)}, \tau^{(r)}, n)}{\psi(\mu_j^{(r)}, \mu_j^{(r)}, \tau^{(r)}, n)}, 1 \right\}$$

که در آن

$$\psi(\mu_j,\mu_j,\tau,n)=\frac{\prod_{i=1}^{I}\left\{\frac{\Gamma(n_{ij}+\mu_j\tau)}{\Gamma(\mu_j\tau)}\frac{\Gamma(n_{iJ}+\mu_J\tau)}{\Gamma(\mu_J\tau)}\right\}}{\left(\frac{\mu_j}{a_j}\right)^{\mu_j^{(a)}\tau^{(a)}-1}\left(1-\frac{\mu_j}{a_j}\right)^{\mu_j^{(a)}\tau^{(a)}-1}}$$

و $\mu_j^{(a)}$ که $j=1,2,\ldots,J$ و $\tau^{(a)}$ در پیوست A نشان داده شده است. توجه کنید که با استفاده از نتایج پیوست A، توزیع شرطی $\mu|\tau,n$ را برحسب یک توزیع $Dirichlet(\mu'\tau')$ همراه با $\mu'_j=(\tau^{(a)}\mu_j^{(a)}+\tau^{(0)}\mu_j^{(0)}-)/(\tau^{(a)}+\tau^{(0)})$ و $\tau'=\tau^{(a)}+\tau^{(0)}$ که از آن چگالی شرطی حاشیه‌ای به روش مشابه با (۸) به دست آمده است، تقریب زده می‌شود. در مطالعه‌ی شبیه‌سازی بخش۳، $Dirichlet(\mu'\tau')$ چند متغیره به عنوان چگالی پیشنهادی به کار رفته است.

گام متروپلیس به صورت زیر به دست می‌آید: با فرض این که زنجیر مارکف در $\mu_j^{(r)}$ قرار دارد، یک برامد تصادفی $\mu_j^{(r+1)}$ از این چگالی پیشنهادی به دست می‌آید و $U_{r,r+1}$ محاسبه می‌شود. یک برامد تصادفی یکنواخت U در $[0,1]$ حاصل می‌شود و اگر $U<U_{r,r+1}$ بود انحراف تصادفی $\mu_j^{(r+1)}$ پذیرفته می‌شود و در غیر این صورت زنجیر در وضعیت $\mu_j^{(r)}$ باقی می‌ماند.

توزیع پسین شرطی τ به شرط μ و n برابر است با

(۹)

$$\pi(\tau|\mu,n)\propto\prod_{i=1}^{I}\left\{\left\{\prod_{j=1}^{J}\frac{\Gamma(n_{ij}+\mu_j\tau)}{\Gamma(\mu_j\tau)}\right\}/\frac{\Gamma(n_{i.}+\tau)}{\Gamma(\tau)}\right\}\tau^{\eta^{(0)}-1}e^{-\upsilon^{(0)}\tau},\quad \tau>0$$

در پیوست B توزیع پسین شرطی τ به شرط μ و n داده شده در (۹) برحسب یک توزیع گاما که به عنوان چگالی پیشنهادی استفاده شده است، تقریب زده می‌شود. در واقع چگالی پیشنهادی $\Gamma(\eta^{(a)}+\eta^{(0)}-1, \upsilon^{(a)}+\upsilon^{(0)})$ است که $\eta^{(a)}$ و $\upsilon^{(a)}$ در پیوست B شرح داده شده‌اند.

با فرض این که نمونه‌گیر متروپلیس ـ هستینگس در $\tau^{(r)}$ قرار دارد، احتمال پذیرش $\tau^{(r+1)}$ برابر است با

$$T_{r,r+1} = \min\left\{\frac{\psi(\tau_j^{(r+1)}, \mu_j^{(r)}, \tau^{(r)}, n)}{\psi(\tau_j^{(r)}, \mu_j^{(r)}, \tau^{(r)}, n)}, 1\right\}$$

و

$$\psi(\tau,\mu,\tau,n) = \frac{\prod_{i=1}^{I}\left\{\prod_{j=1}^{J}\Gamma(n_{ij}+\mu_j\tau)/\Gamma(\mu_j\tau)\right\}/(\Gamma(n_{i_0}+\tau)/\Gamma(\tau))}{\tau^{\eta^{(a)}-1}e^{-\upsilon^{(a)}\tau}}$$

گام متروپلیس برای توزیع پسین شرطی $\tau \mid \mu, n$ در (۹) به روشی مشابه با روشی که برای $\mu_j \mid \tau, \mu_j, n$ به کار رفت، انجام می‌شود.

مطالعه‌ی عددی و مقایسه‌ها

برای ارزیابی محاسبه‌ها از سه کاربرد استفاده خواهد شد. نخست یک مطالعه‌ی شبیه‌سازی است که در آن از استنباط بیز پیشگو با هدف براورد نسبت‌های جامعه‌ی متنابهی، برای یک جامعه‌ی خوشه‌ای استفاده می‌شود. دومین کاربرد در رتبه‌بندی و انتخاب است، که در آن بهترین جامعه‌ی چند جمله‌ای براساس داده‌های حسی با استفاده از مدل‌های لوژستیک خطی و غیرخطی انتخاب می‌شود. کاربرد دوم دارای دو ویژگی، تعداد زیاد سلول‌ها و پراکندگی داده‌ها می‌باشد. کاربرد سوم مقایسه‌ی الگوریتم متروپلیس ـ هستینگس با روش‌های نمونه‌گیری لاپلاس و گیبس برای تحلیل مدل

سلسله مراتبی دو جمله‌ای ـ بتا است.

در دو مثال اول برای تحلیل و مقایسه، سه حالت در نظر گرفته می‌شود: (۱) هر دوی μ و τ مجهول هستند، (۲) μ مجهول اما τ معلوم است و (۳) هردوی μ و τ معلوم هستند در چارچوب بیز تجربی می‌توان از (۲) یا (۳) استفاده کرد. وقتی که حالت (۱) یا (۲) به کار می‌رود براوردهای مدل به دست آمده از حالت (۱) جایگزین می‌شود. متوجه می‌شویم که موضوع به حالت (۲) ارتباط پیدا می‌کند که در آن فعالیت زیادی انجام شده است.

در همه‌ی مثال‌ها $\mu_j^{(\circ)}\tau^{(\circ)} = 1$ که $j = 1, 2, \dots, J$ ، $\eta^{(\circ)} = \circ$ و $\upsilon^{(\circ)} = \circ$ در نظر گرفته شده است. برای الگوریتم متروپلیس ـ هستینگس از یک دوره‌ی سوختن با ۵۰۰ تکرار استفاده می‌شود و هر پنج تکرار برای به دست آوردن ۲۰۰۰ تکرار در نظر گرفته می‌شود. این یک روش محافظه کارانه است و همه‌ی خود همبستگی‌ها را حذف می‌کند. ما از ۵۰ تا از ۲۰۰۰ تکرار ناهمبسته برای براورد خطای استاندارد عددی به روش میانگین بچ استفاده می‌کنیم. تکرار با ۱۰۰۰۰ تکرار نشان می‌دهد که در براوردهای مرتبط تغییری ایجاد نمی‌شود.

ـ مطالعه‌ی شبیه‌سازی

جامعه‌ی متناهی شامل I خوشه با N_i واحد در خوشه‌ی i ام است. پاسخ هر واحد می‌تواند در یکی از J رسته قرار گیرد. در مرحله‌ی اول همه‌ی خوشه‌ها انتخاب می‌شوند و در مرحله‌ی دوم n_i واحد از هر خوشه‌ی نمونه‌گیری شده، نمونه‌گیری می‌شود. فرض کنید N_{ij} تعداد کل واحدهای (مجهول) پاسخ داده در رسته‌ی j در خوشه‌ی i ام در همه‌ی جامعه باشد. هدف براورد نسبت جامعه‌ی متناهی برای رسته‌ی j ام به صورت زیر است:

$$P_j = N^{-1} \sum_{i=1}^{I} N_{ij}, \quad j = 1, 2, \dots, J \qquad (10)$$

که در آن $N = \sum_{i=1}^{I} \sum_{j=1}^{I} N_{ij}$ تعداد کل افراد در همه‌ی جامعه است. برای سادگی، N معلوم فرض می‌شود. در شبیه‌سازی $N_{i_1} = ۱۰۰$ و $N_{i_2} = ۲۰$ ، $i = ۱,\dots,I$ در نظر گرفته می‌شود که در آن $I = ۲۰$ است. در این جا سه مدل داده‌های چند جمله‌ای در نظر گرفته می‌شود که اولی دارای دو سلول، دومی دارای سه سلول و سومی دارای پنج سلول می‌باشد. برای همه‌ی مدل‌ها با $\tau = ۵,۱۰,۲۵$ شروع می‌شود. برای مدل دو سلولی $\mu_1 = ۰/۴ = ۱ - \mu_2$ ، برای مدل سه سلولی $\mu_1 = ۰/۱$ ، $\mu_2 = ۰/۳$ و $\mu_3 = ۰/۶$ و برای مدل پنج سلولی $\mu_1 = \mu_2 = ۰/۱$ ، $\mu_3 = \mu_4 = ۰/۲$ و $\mu_5 = ۰/۴$ در نظر گرفته می‌شود. با استفاده از این مقادیر داده‌ها را از جامعه‌ی چند جمله‌ای به دست می‌آوریم. نخست p_{ij} از توزیع دیریکله و سپس n_{ij} به شرط p_{ij} از توزیع چند جمله‌ای حاصل می‌شود. این داده‌های شبیه‌سازی شده در تحلیل به عنوان داده‌های واقعی مورد استفاده قرار می‌گیرند.

جدول۱ـ مقایسه‌ی چگالی‌های پسین نسبت جامعه‌ی متناهی (P_1) برای مدل چند جمله‌ای دو سلولی برحسب حالتی که $\mu_1 = 0/4$ است

τ	طرح	میانگین	P_1 انحراف استاندارد	خطای استاندارد عددی
۵	۱	۰/۴۱۰۲	۰/۰۲۰۴	۰/۰۰۲۸
	۲	۰/۴۱۰۳	۰/۰۲۰۲	۰/۰۰۳۵
	۳	۰/۴۱۲۷	۰/۰۱۸۲	۰/۰۰۲۶
۱۰	۱	۰/۴۰۳۳	۰/۰۲۰۹	۰/۰۰۲۵
	۲	۰/۴۰۳۵	۰/۰۲۰۹	۰/۰۰۳۴
	۳	۰/۴۰۵۲	۰/۰۱۷۶	۰/۰۰۲۴
۲۵	۱	۰/۴۱۱۲	۰/۰۲۱۱	۰/۰۰۲۶
	۲	۰/۴۱۱۵	۰/۰۲۱۲	۰/۰۰۳۶
	۳	۰/۴۱۲۷	۰/۰۱۶۹	۰/۰۰۲۲

یک براورد از نسبت‌های جامعه‌ی متناهی در (۱۰) به روش زیر به دست می‌آید. نخست توجه کنید که

(۱۱)

$$(N_{i1}-n_{i1}, N_{i2}-n_{i2}, \dots, N_{iJ}-n_{iJ}) \mid p_i, N_{i_0}, n_{i_0} \sim_{ind} Multinomial(N_{i_0}-n_{i_0}, p_i).$$

بنابر این در هر تکرار الگوریتم متروپلیس ـ هستینگس $(N_{i1}-n_{i1}, N_{i2}-n_{i2}, \dots, N_{iJ}-n_{iJ})$ از (۱۱) برای $i=1,\dots,I$ و $j=1,\dots,J$ به دست می‌آید. آنگاه

$$P_j = \left\{\left(\sum_{i=1}^{I} n_{ij}\right) + \left(\sum_{i=1}^{I} (N_{ij} - n_{ij})\right)\right\} N^{-1},\ j = 1, 2, \ldots, J \qquad (12)$$

محاسبه می‌شود. توجه کنید که به شرط n در (۱۲) مقدار $\sum_{i=1}^{I} n_{ij}$ از داده‌های مشاهده شده مشخص است. بنابراین میانگین‌های پسین و انحراف استانداردهای p_j با استفاده از توزیع‌های تجربی براورد شده‌ی p_j به دست می‌آیند.

در جدول (۱) نتایج شبیه‌سازی برای حالت دو جمله‌ای با $\mu_1 = 0/4$ نشان داده شده است. براوردهای نسبت‌های جامعه‌ی متناهی برای همه‌ی حالت‌ها بسیار شبیه است، شاید برای حالت (۳) کمی بزرگ‌تر باشد. برای سه مقدار τ بجز برای $\tau = 10$ تغییرات اندکی وجود دارد به طوری که براورد نسبت‌های جامعه‌ی متناهی کمی کوچک‌تر هستند. انحراف استانداردهای براوردها با حالت‌های طرح‌های (۱) و (۲) بسیار شبیه هستند اما انحراف استانداردهای حالت (۳) بسیار کمتر است. از حالت (۳) به حالت (۲) افزایش انحراف استاندارد برای $\tau = 5$، $\tau = 10$ و $\tau = 25$ به ترتیب به میزان ۱۱، ۱۸ و ۲۵ درصد وجود دارد. افزایش‌های مربوط در واریانس‌ها به ترتیب برابر ۲۳، ۴۱ و۵۷ درصد است. این شبیه‌سازی نشان می‌دهد که با افزایش τ، افزایش‌های بزرگ‌تری وجود دارد. مشاهده می‌شود که خطاهای استاندارد عددی مقایسه پذیرند.

جدول ۲ـ مقایسه‌ی چگالی‌های پسین نسبت جامعه‌ی متناهی P_1، P_2 و P_3 مربوط به مدل چند جمله‌ای سه سلولی برای حالت ۱ $\mu_1 =$ ۰/۱، $\mu_2 =$ ۰/۳ و $\tau =$ ۱۰

طرح	کمیت	P_1	P_2	P_3
۱	میانگین	۰/۰۹۱۶	۰/۲۶۸۷	۰/۶۳۹۶
	انحراف استاندارد	۰/۰۱۱۶	۰/۰۱۸۶	۰/۰۲۰۲
	خطای استاندارد عددی	۰/۰۰۱۶	۰/۰۰۲۵	۰/۰۰۲۳
۱	میانگین	۰/۰۹۱۴	۰/۲۶۸۸	۰/۶۳۹۸
	انحراف استاندارد	۰/۰۱۱۶	۰/۰۱۸۶	۰/۰۲۰۲
	خطای استاندارد عددی	۰/۰۰۱۴	۰/۰۰۲۲	۰/۰۰۲۴
۱	میانگین	۰/۰۹۳۷	۰/۲۷۳۶	۰/۶۲۲۷
	انحراف استاندارد	۰/۰۱۰۷	۰/۰۱۶۰	۰/۰۱۷۴
	خطای استاندارد عددی	۰/۰۰۱۵	۰/۰۰۲۱	۰/۰۰۲۴

جدول (۲) نتایج شبیه‌سازی را برای حالت سه جمله‌ای با $\mu_1 =$ ۰/۲، $\mu_2 =$ ۰/۳، $\mu_3 =$ ۰/۵ و $\tau =$ ۱۰ نشان می‌دهد. بین براوردهای نسبت‌های جامعه‌ی متناهی، تفاوت‌های مهمی وجود ندارد. در ضمن تغییرات برای حالت (۳) نسبت به حالت (۱) یا (۲) کم براورد می‌شود. برای P_1، P_2 و P_3 افزایش در انحراف استانداردها به ترتیب برابر ۸، ۱۶ و ۱۵ درصد و افزایش در واریانس‌های مربوط به ترتیب برابر ۱۸، ۳۵ و ۳۲ درصد است. خطاهای استاندارد عددی برای همه‌ی حالت‌ها مشابه هستند اما کمترین آن‌ها درباره‌ی P_1 است. در آخر، مثال چند جمله‌ای با پنج سلول در جدول (۳) نشان داده شده است که در آن $\tau =$ ۱۰، $\mu_1 = \mu_2 =$ ۰/۱ و $\mu_3 = \mu_4 =$ ۰/۲ است. با توجه به این که $n_i =$ ۲۰ است، انتظار می‌رود دو سلول اول پراکنده باشند. همچنین براوردهای نقطه‌ای نسبت‌های متناهی بین سه حالت یاد شده بسیار مشابه هستند. هر چند، افزایش در

انحراف استانداردها بسیار چشم‌گیرتر از دو مثال پیش است. تغییرها در انحراف استانداردها از حالت (۱) به حالت (۲) بجز برای P_5 با ۹ درصد افزایش واریانس ناچیز است که شگفت‌آور نیست. هر چند، افزایش در تغییرات از حالت (۳) به حالت (۲) فراوان است. انحراف استانداردها به اندازه‌ی ۵۰، ۴۱، ۵۶، ۵۰ و ۵۵ درصد با افزایش در واریانس برابر ۱۲۷، ۹۸، ۱۴۳، ۱۲۴ و ۱۴۰ درصد به ترتیب برای P_1، P_2، P_3، P_4 و P_5 افزایش می‌یابند. در این مثال خطای استاندارد عددی برای حالت (۳) تا حدودی کوچک‌تر است، اما این مطلب، تفاوت‌های عمده در تغییرات را از بین نمی‌برد.

جدول ۳ـ مقایسه‌ی چگالی‌های پسین نسبت جامعه‌ی متناهی P_1، P_2، P_3، P_4 و P_5 مربوط به مدل چند جمله‌ای پنج سلولی برای حالت‌ها

$\mu_1 = \mu_2 = ۰/۱$، $\mu_3 = \mu_4 = ۰/۲$ و $\tau = ۲۵$

طرح	کمیت	P_1	P_2	P_3	P_4	P_5
۱	میانگین	۰/۰۸۳۱	۰/۰۶۰۹	۰/۲۱۴۱	۰/۲۲۲۱	۰/۴۱۹۸
	انحراف استاندارد	۰/۰۱۳۲	۰/۰۱۱۲	۰/۰۲۰۸	۰/۰۲۰۶	۰/۰۲۴۳
	خطای استاندارد عددی	۰/۰۰۱۴	۰/۰۰۱۷	۰/۰۰۲۹	۰/۰۰۳۲	۰/۰۰۳۱
۱	میانگین	۰/۰۸۲۳	۰/۰۶۱۳	۰/۲۱۲۶	۰/۲۲۱۹	۰/۴۲۱۹
	انحراف استاندارد	۰/۰۱۳۴	۰/۰۱۱۴	۰/۰۲۰۹	۰/۰۲۱۱	۰/۰۲۵۴
	خطای استاندارد عددی	۰/۰۰۱۸	۰/۰۰۲۰	۰/۰۰۲۸	۰/۰۰۲۷	۰/۰۰۳۴
۱	میانگین	۰/۰۸۱۷	۰/۰۶۲۸	۰/۲۱۴۹	۰/۲۲۰۸	۰/۴۱۹۹
	انحراف استاندارد	۰/۰۰۸۹	۰/۰۰۸۱	۰/۰۱۳۴	۰/۰۱۴۱	۰/۰۱۶۴
	خطای استاندارد عددی	۰/۰۰۱۰	۰/۰۰۱۴	۰/۰۰۱۶	۰/۰۰۱۳	۰/۰۰۱۸

به طور خلاصه به نظر می‌رسد که اگر حالت (۲) (یعنی وقتی که μ نامعلوم اما τ معلوم باشد) رخ دهد، مقدار کمی از روش بیز کامل را از دست می‌دهیم. ولی برای حالت (۳) (یعنی روش بیز تجربی کامل) کم براوردی زیادی در تغییرات وجود دارد.

افزون بر این برای مطالعه‌ی شبیه‌سازی، یک روش جایگزین برای انجام تحلیل بیز کامل پیش‌بینی شده است. برای تولید توزیع پسین شرطی همه‌ی بردار μ به شرط τ و n از چگالی پیشنهادی به دست آمده در پیوست A (یعنی چگالی چند متغیره‌ی $Dirichlet(\mu\tau)$) استفاده می‌شود. یعنی روش اصلی که در آن توزیع‌های پسین شرطی μ_j گاهی به دست می‌آیند با روش جایگزین مقایسه می‌شود. برای همه‌ی ۹ طرح (تعداد سلول‌های چند جمله‌ای ۲، ۳ و ۵ نسبت به $\tau=$ ۵,۱۰,۲۵) با استفاده از روش جایگزین محاسبه‌ها تکرار شده است. برای مقایسه‌ی آسان از ۲۰۰۰ تکرار به دست آمده بعد از یک دوره‌ی سوختن ۵۰۰ تکراری و بعد از انتخاب هرم l امین تکرار یعنی $l=$ ۵,۱۰,۲۵ استفاده شده است با این هدف که نمونه‌ای تصادفی به دست می‌آید (به خاطر داشته باشید که برای روش اصلی از $l=$ ۵ استفاده شد).

مسئله‌ی اصلی در روش جایگزین این است که برای ثابت‌های میزان‌سازی نیاز به جستجوی وقت‌گیر است (مثال، چیب و گرین برگ[۱] ۱۹۹۵) تا این که خود همبستگی‌های ناچیز بین تکرارها و احتمال‌های پرش معقول (مثال،گلمن[۲] و همکاران ۱۹۹۶) به دست آید. دو ثابت میزان‌ساز انتخاب می‌شود، یکی برای تولید از توزیع پسین شرطی μ و دیگری برای تولید از توزیع پسین τ. دو ثابت میزان‌ساز ψ_1 و ψ_2 هستند که در آن τ' به صورت $\psi_1\tau' = \psi_1(\tau^{(a)} + \tau^{(\circ)})$ (پیوست A را ببینید) و σ_*^2 به صورت $\psi_2\sigma_*^2$ تعدیل می‌شوند (پیوست B را ببینید). توجه کنید که در روش اصلی ثابت‌های میزان‌ساز با مقادیر واحد خوب کار می‌کنند و در حقیقت انتخاب ثابت‌های

---

۱ Chib and Greenberg

۲ Gelman

میزان‌ساز مشکل نبود.

جدول ۴ـ مقایسه‌ی چگالی‌های پسین نسبت جامعه‌ی متناهی P_1، P_2، P_3، P_4 و P_5 برای مدل چند جمله‌ای پنج سلولی برای چگالی پیشنهادی جایگزین برحسب l

	l	P_1	P_2	P_3	P_4	P_5
میانگین	۵	۰/۰۸۱۲	۰/۰۶۲۱	۰/۲۱۵۲	۰/۲۲۰۴	۰/۴۲۱۱
	۱۰	۰/۰۸۰۶	۰/۰۶۲۶	۰/۲۱۵۷	۰/۲۱۹۹	۰/۴۲۱۲
	۲۰	۰/۰۸۰۶	۰/۰۶۲۴	۰/۲۱۵۸	۰/۲۲۰۴	۰/۴۲۰۹
انحراف استاندارد	۵	۰/۰۱۰۹	۰/۰۰۹۴	۰/۰۱۷۱	۰/۰۱۷۱	۰/۰۱۹۹
	۱۰	۰/۰۱۰۶	۰/۰۰۹۹	۰/۰۱۷۱	۰/۰۱۶۹	۰/۰۲۰۲
	۲۰	۰/۰۱۰۹	۰/۰۰۹۶	۰/۰۱۶۹	۰/۰۱۷۱	۰/۰۱۹۸
خطای استاندارد عددی	۵	۰/۰۰۴۱	۰/۰۰۳۱	۰/۰۰۷۴	۰/۰۰۵۹	۰/۰۰۷۶
	۱۰	۰/۰۰۲۸	۰/۰۰۳۶	۰/۰۰۶۰	۰/۰۰۵۱	۰/۰۰۶۸
	۲۰	۰/۰۰۲۵	۰/۰۰۲۹	۰/۰۰۴۰	۰/۰۰۴۱	۰/۰۰۴۹

در جدول (۴) نتایج برای چند جمله‌ای پنج سلولی و حالت (۱) نشان داده شده است (جدول (۳) را ببینید). ثابت‌های میزان‌ساز برابر ۱۰ = ψ_1 و ۱/۰۵ = ψ_1 انتخاب شده‌اند. (توجه کنید که ۱ = ψ_2 = ψ_1 انتخاب بسیار ضعیفی است). برای هر سه حالت احتمال‌های پرش ۰/۱۹ برای μ، ۰/۷۹ برای τ و ۰/۱۵ برای هر دوی μ و τ

است این ثابت‌های میزان‌ساز و احتمال‌های پرش و مثال‌های شبیه‌سازی شده تغییر می‌کنند.

میانگین‌های پسین و انحراف استانداردها به ترتیب P_iها با تغییر l به طور چشم‌گیر تغییر نمی‌کنند. در مقایسه با مقادیر مربوط در جدول (۳) (حالت ۱) تفاوت‌های بسیار کوچکی بین میانگین‌های پسین وجود دارد اما انحراف استانداردهای پسین برای روش جایگزین کوچک‌تر از انحراف استانداردهای روش اصلی هستند. جدی‌ترین مسئله در مورد روش جایگزین دقت آن است. خطای استاندارد عددی بیش از دو برابر بزرگ‌تر از روش اصلی است (مثال، ۰/۰۰۴۱ را با ۰/۰۰۱۷ مقایسه کنید). هر چند، خطاهای یادشده به سرعت با تغییر l در $l=۲۰$ به مقادیر مربوط به روش اصلی (گر چه هنوز بزرگ‌تر است) نزدیک شده و کاهش می‌یابند. حتی در $l=۲۰$ خود همبستگی بین تکرارها برای هر یک از پارامترهای μ_i و τ که همگی در این مثال مثبت هستند، از حدود ۰/۵۰ تا حدود ۰/۲۰ تغییر می‌کند. این مطلب تفاوت بین خطاهای استاندارد برای روش‌های اصلی و جایگزین را بیان می‌کند.

سرانجام، زمان اجرای هر نه طرح برای مطالعه‌ی شبیه‌سازی را با استفاده از دو روش بیز کامل با کامپیوتر آلفای ۳۳۰ مگا هرتز در نظر گرفتیم. برای ۱۰۵۰۰ تکرار با دوره‌ی سوختن برابر ۵۰۰ تکرار و پس از انتخاب هر پنجمین تکرار، روش اصلی ۸۳/۹ دقیقه و روش جایگزین ۳۰/۵ دقیقه زمان گرفته است. آزمایش‌های دیگر روی روش اصلی مشخص می‌کند که ۴۲۰۰ تکرار با دوره‌ی سوختن برابر با دوره‌ی سوختن برابر ۲۰۰ تکرار و انتخاب هر دومین تکرار ۳۵/۲ دقیقه زمان گرفته است. با ترکیب اخیر، میانگین‌های پسین، انحراف‌های استاندارد و NSEها بسیار شبیه به میانگین‌های پسین، انحراف‌های استاندارد NSE های روش اصلی در ۱۰۵۰۰ تکرار با دوره‌ی سوختن برابر با ۵۰۰ تکرار و انتخاب هر پنجمین تکرار است.

اجرای روش جایگزین مشکل است، زیرا ثابت‌های میزان‌ساز به راحتی به دست نمی‌آیند. با هر مجموعه‌ی داده‌ی جدید باید ثابت‌های میزان‌ساز جدیدی را پیدا کنیم. اگر ثابت‌های میزان‌ساز معقولی پیدا نکردیم، خود همبستگی‌های زیادی، بین تکرارها

حتی در تأخیرهای بزرگ وجود دارد و احتمال‌های پرش به طور غیرقابل قبولی بزرگ یا کوچک خواهند بود. آنگاه خطای استاندارد عددی بسیار بزرگ خواهد بود که روش جایگزین را نادقیق خواهد کرد. براساس مثال‌های زیاد تجربه نشان می‌دهد که مشکل و نادقیق بودن، با افزایش تعداد سلول‌های چند جمله‌ای افزایش می‌یابد.

ـ انتخاب بهترین جامعه‌ی چند جمله‌ای

انتخاب بهترین قلم غذایی از بین دوازده نوع غذا در جیره‌ی نظامی را در نظر بگیرید. اقلام (شامل قلم غذایی موردنظر) موجود در هر غذا در یک آزمون چشایی به ترتیب تصادفی توسط ۳۶ مشاور با مشاورهای مختلف برای غذاهای گوناگون در یک مقیاس ۹ گزینه‌ای درباره‌ی خوشمزگی ارزیابی می‌شود. بنابراین، دوازده جامعه‌ی چند جمله‌ای مستقل برای انتخاب بهترین غذا وجود دارد. در جدول (۵) پاسخ‌های درباره‌ی دوازده قلم غذایی نشان داده شده است. به پراکندگی داده‌ها در نقاط انتهایی پایین و بالای مقیاس خوشمزگی توجه کنید.ناندرام (۱۹۹۷) از چندین معیار (برای مثال، میانگین و انحراف استاندارد) و یک روش مبتنی بر نمونه‌گیری نقاط مهم به منظور انتخاب بهترین قلم غذایی برای داده‌های جدول (۵) استفاده کرده است. نیول[1] (۱۹۸۲) روش غیر بیزمک کولاف[2] (۱۹۸۰) را برای تحلیل داده‌های حسی به کاربرد و با استفاده از این روش برخی معایب استفاده از نمره‌ها (مقادیر مقیاس خوشمزگی) را رفع کرد. در این روش برای داده‌های ترتیبی، رسته‌های پاسخ به عنوان فاصله‌های مجاور در یک مقیاس پیوسته با نقاط قطع مجهول $\theta_1 < \theta_2 < \ldots < \theta_{J-1}$ در نظر گرفته می‌شود که برای نقطه‌ی J مقیاس $J = ۹$ است. نظر به این که P_{ij} نشان‌دهنده‌ی این احتمال است که یک مشاور در i امین جامعه در j امین رسته پاسخ دهد. فرض کنید $\gamma_{ij} = \sum_{s=1}^{j} p_{is}$ احتمال تجمعی جامعه‌ی i ام باشد. با این فرض‌ها نیول (۱۹۸۲) مدل را به صورت

---

[1] Newell

[2] Mccullagh

$$\ln\{\gamma_{ij}/(۱-\gamma_{ij})\} = (\theta_j - \beta_i)/\delta_i\,, \quad i = ۱,۲,\ldots,\mathrm{I}\,,\ j = ۱,۲,\ldots,\mathrm{J}-۱ \qquad (۱۳)$$

جدول ۵ ـ پاسخ‌های مشاوران برای دوازده قلم غذایی از ارزیابی حسی نظامی

غذای اصلی		رسته‌های پاسخ								
		۱	۲	۳	۴	۵	۶	۷	۸	۹
۱	سوسیس خوک	۰	۱	۱	۲	۲	۸	۱۱	۸	۳
۲	گوشت خوک و مرغ	۱	۰	۳	۴	۵	۷	۱۲	۴	۰
۳	گوشت گوساله	۰	۱	۲	۴	۱	۷	۱۴	۷	۰
۴	باربکیو گوساله	۰	۱	۱	۲	۳	۴	۱۸	۴	۳
۵	کباب گوساله	۰	۰	۱	۲	۳	۹	۱۳	۷	۱
۶	فرانک فورتر	۲	۸	۶	۵	۲	۹	۴	۰	۰
۷	بوقلمون	۰	۱	۳	۳	۳	۵	۱۲	۹	۰
۸	گوشت آب‌پز گوساله	۱	۳	۲	۶	۵	۷	۸	۴	۰
۹	مرغ	۰	۱	۲	۲	۰	۷	۱۱	۱۰	۳
۱۰	کله گنجشکی	۱	۰	۲	۶	۱	۹	۹	۸	۰
۱۱	ژامبون	۰	۰	۰	۰	۱	۱۱	۱۴	۹	۱
۱۲	آب‌پز گوساله با سس	۱	۳	۴	۲	۳	۶	۱۰	۷	۰

در نظر گرفت که در آن β_i و δ_i به ترتیب ملاک‌های نسبی مکان و پراکنش جامعه‌ی i ام هستند. توجه کنید که در (۱۳) محدودیت‌های $\sum_{i=1}^{I}\beta_i = 0$ و $\sum_{i=1}^{I}\ln(\delta_i) = 0$ استفاده می‌شود، مک کولاف (۱۹۸۰) را ببینید. این مدل مکان ارزیابی‌ها و سازگاری پاسخ‌های مشاورها را به طور مستقیم ترکیب می‌کند. β_i، δ_i و θ_j در (۱۳) توزیع‌های پسین دارند که ویژگی‌های آن‌ها از توزیع‌های پسین p_{ij} گرفته شده است. برای این کاربرد، مدل بیز سلسله مراتبی سه مرحله‌ای به داده‌ها برازش می‌شود که به این وسیله به داده‌ها تا حدودی امکان هموارسازی داده می‌شود. هر تکرار الگوریتم متروپلیس ـ هستینگس مقادیری برای p_{ij} از لحاظ جایگزینی می‌توان از مدل بسیار ساده‌تر لوژستیک خطی استفاده کرد که در آن δ_i در (۱۳) مقدار یک می‌گیرد. در این حالت در h امین تکرار الگوریتم متروپلیس- هستینگس براوردگرهای کمترین توان‌های دوم β_i و θ_j به فرم بسته وجود دارند. برای در نظرگرفتن $\bar{z}_{\circ j}^{(h)} = \sum_{i=1}^{I} z_{ij}^{(h)} / I$، $\bar{z}_{i\circ}^{(h)} = \sum_{j=1}^{J-1} z_{ij}^{(h)} /(J-1)$ و $\bar{z}_{\circ\circ}^{(h)} = \sum_{i=1}^{I} z_{i\circ}^{(h)} / I$ براوردگرهای کمترین توان‌های دوم برابر $\theta_j^{(h)} = \bar{z}_{\circ j}^{(h)}$ و $\beta_i^{(h)} = \bar{z}_{\circ\circ}^{(h)} - \bar{z}_{i\circ}^{(h)}$ هستند.

جدول ۶ ـ مقایسه‌ی توزیع‌های پسین β_i، $i = ۱, \ldots, ۱۲$ برحسب حالت‌ها و مدل‌ها

کمیت	حالت ۱ میانگین	انحراف استاندارد	حالت ۲ میانگین	انحراف استاندارد	حالت ۳ میانگین	انحراف استاندارد
(الف) مدل لوژستیک خطی						
$\beta_{۱}$	۰/۴۰۸	۰/۳۲۵	۰/۴۰۴	۰/۳۱۵	۰/۲۳۵	۰/۲۹۷
$\beta_{۲}$	−۰/۲۵۴	۰/۳۰۴	−۰/۲۵۲	۰/۲۹۸	−۰/۱۲۸	۰/۲۵۰
$\beta_{۳}$	−۰/۰۷۱	۰/۳۴۱	−۰/۰۸۴	۰/۳۴۶	۰/۱۹۶	۰/۲۸۹
$\beta_{۴}$	۰/۳۷۹	۰/۳۲۱	۰/۳۷۳	۰/۳۱۹	۰/۳۲۸	۰/۲۹۳
$\beta_{۵}$	۰/۳۵۴	۰/۳۴۰	۰/۳۶۹	۰/۳۳۷	۰/۳۲۷	۰/۳۱۳
$\beta_{۶}$	−۰/۹۵۰	۰/۲۹۶	−۰/۹۳۷	۰/۲۷۴	−۰/۹۹۶	۰/۲۱۱
$\beta_{۷}$	−۰/۰۵۷	۰/۳۳۶	۰/۰۵۳	۰/۳۴۳	۰/۱۴۶	۰/۲۹۷
$\beta_{۸}$	−۰/۴۶۱	۰/۳۰۶	−۰/۴۶۱	۰/۲۸۴	−۰/۴۰۹	۰/۲۳۸
$\beta_{۹}$	۰/۴۲۳	۰/۳۱۷	۰/۴۱۷	۰/۳۱۴	۰/۲۲۹	۰/۳۰۴
$\beta_{۱۰}$	−۰/۱۷۶	۰/۳۰۷	−۰/۱۸۳	۰/۳۱۱	−۰/۱۲۰	۰/۲۴۶
$\beta_{۱۱}$	۰/۵۱۴	۰/۳۵۶	۰/۵۱۵	۰/۳۵۱	۰/۴۸۶	۰/۳۲۲
$\beta_{۱۲}$	−۰/۳۶۴	۰/۲۸۷	−۰/۳۸۱	۰/۲۸۹	−۰/۲۹۲	۰/۲۲۸
(ب) مدل لوژستیک غیرخطی						
$\beta_{۱}$	۰/۴۶۷	۰/۲۸۱	۰/۴۷۲	۰/۲۷۶	۰/۱۴۱	۰/۱۸۹
$\beta_{۲}$	−۰/۲۴۰	۰/۳۰۰	−۰/۲۴۰	۰/۲۹۳	−۰/۰۱۴	۰/۲۱۲
$\beta_{۳}$	−۰/۰۲۷	۰/۲۸۵	−۰/۰۰۷	۰/۲۹۳	۰/۱۵۹	۰/۱۹۶
$\beta_{۴}$	۰/۴۳۱	۰/۲۸۲	۰/۴۱۷	۰/۲۷۲	۰/۳۳۲	۰/۲۰۸
$\beta_{۵}$	۰/۲۹۱	۰/۲۷۴	۰/۲۹۶	۰/۲۶۲	۰/۲۲۱	۰/۱۸۹
$\beta_{۶}$	−۰/۹۲۶	۰/۳۰۶	−۰/۹۱۳	۰/۲۸۱	−۰/۹۵۰	۰/۲۱۶
$\beta_{۷}$	−۰/۰۲۷	۰/۲۸۹	۰/۰۳۰	۰/۲۹۲	۰/۰۷۹	۰/۱۹۱
$\beta_{۸}$	−۰/۴۲۸	۰/۳۰۵	−۰/۴۲۶	۰/۲۸۷	−۰/۳۴۲	۰/۲۰۵
$\beta_{۹}$	۰/۴۹۸	۰/۲۷۷	۰/۴۹۰	۰/۲۷۴	۰/۱۵۹	۰/۱۹۲
$\beta_{۱۰}$	−۰/۱۵۹	۰/۳۰۹	−۰/۱۶۱	۰/۳۰۶	−۰/۰۳۸	۰/۲۰۶
$\beta_{۱۱}$	۰/۴۳۹	۰/۲۸۶	۰/۴۳۶	۰/۲۸۰	۰/۴۰۲	۰/۲۰۷
$\beta_{۱۲}$	−۰/۳۱۹	۰/۲۹۸	−۰/۳۳۳	۰/۳۰۱	−۰/۱۵۰	۰/۲۱۵

از لحاظ جایگزینی می‌توان از مدل بسیار ساده‌تر لوژستیک خطی استفاده کرد که در آن δ_i در (۱۳) مقدار یک می‌گیرد. در این حالت در h امین تکرار الگوریتم متروپلیس- هستینگس براوردگرهای کمترین توان‌های دوم β_i و θ_j به فرم بسته وجود دارند. برای در نظرگرفتن $\bar{z}_{\cdot j}^{(h)} = \sum_{i=1}^{I} z_{ij}^{(h)} / I$، $\bar{z}_{i\cdot}^{(h)} = \sum_{i=1}^{J-1} z_{ij}^{(h)} / (J-1)$ و $\bar{z}_{\cdot\cdot}^{(h)} = \sum_{i=1}^{I} z_{i\cdot}^{(h)} / I$ براوردگرهای کمترین توان‌های دوم برابر $\theta_j^{(h)} = \bar{z}_{\cdot j}^{(h)}$ و $\beta_i^{(h)} = \bar{z}_{\cdot\cdot}^{(h)} - \bar{z}_{i\cdot}^{(h)}$ هستند.

روش کمترین توان‌های دوم به وسیله‌ی چند نفر در متون بیزی استفاده شده است. برای مثال ناندرام و همکاران (۱۹۹۷) و چن و همکاران (۱۹۹۶) را ببینید. البته رگرسیون می‌تواند داخل مدل بیزی به طور مستقیم به فرم مدل خطی تعمیم یافته، قرار گیرد. اما این کار ما را از تأکید اصلی در این تحلیل دور می‌کند.

جدول ۷ ـ مقایسه‌ی توزیع‌های پسین θ_j، $j = ۱, \dots, ۸$ برحسب حالت‌ها و مدل‌ها

کمیت	حالت ۱ میانگین	انحراف استاندارد	حالت ۲ میانگین	انحراف استاندارد	حالت ۳ میانگین	انحراف استاندارد
(الف) مدل لوژستیک خطی						
θ_1	-۵/۰۸۰	۰/۶۶۴	-۵/۰۴۵	۰/۶۸۶	-۴/۹۴۸	۰/۳۷۴
θ_2	-۳/۰۵۴	۰/۳۳۶	-۳/۰۴۹	۰/۳۳۳	-۳/۹۰۵	۰/۱۵۵
θ_3	-۲/۱۱۲	۰/۲۱۸	-۲/۱۱۱	۰/۲۱۹	-۲/۰۹۱	۰/۱۰۲
θ_4	-۱/۴۰۴	۰/۱۶۹	-۱/۳۴۹	۰/۱۷۳	-۱/۳۹۰	۰/۰۸۱
θ_5	-۱/۰۰۹	۰/۱۵۱	-۱/۰۰۳	۰/۱۵۲	-۱/۰۰۰	۰/۰۷۲
θ_6	-۰/۰۷۴	۰/۱۳۵	-۰/۰۶۹	۰/۱۳۶	-۰/۰۷۱	۰/۰۶۳
θ_7	۱/۴۱۱	۰/۱۶۹	۱/۴۱۳	۰/۱۷۴	۰/۷۳۴	۰/۰۶۷
θ_8	۴/۲۵۲	۰/۷۲۵	۴/۲۳۳	۰/۷۲۷	۱/۷۹۸	۰/۰۹۵
(ب) مدل لوژستیک غیرخطی						
θ_1	-۵/۰۴۳	۰/۶۶۶	-۵/۰۰۷	۰/۶۸۸	-۴/۹۲۸	۰/۳۶۹
θ_2	-۲/۹۸۲	۰/۳۳۱	۲/۹۸۵	۰/۳۲۵	-۲/۹۰۶	۰/۱۶۵
θ_3	-۲/۰۶۱	۰/۲۱۴	-۲/۰۶۱	۰/۲۱۵	-۲/۰۰۷	۰/۱۱۲
θ_4	-۱/۳۷۲	۰/۱۶۶	-۱/۳۶۳	۰/۱۷۰	-۱/۳۲۸	۰/۰۸۷
θ_5	-۰/۹۹۰	۰/۱۵۰	-۰/۹۸۴	۰/۱۵۱	-۰/۹۵۱	۰/۰۷۷
θ_6	-۰/۰۸۸	۰/۱۳۸	-۰/۰۸۲	۰/۱۳۹	-۰/۰۵۷	۰/۰۶۷
θ_7	۱/۳۴۲	۰/۱۸۰	۱/۳۴۸	۰/۱۸۲	۰/۷۱۴	۰/۰۷۲
θ_8	۴/۱۸۶	۰/۷۰۱	۴/۱۶۸	۰/۷۰۶	۱/۷۳۴	۰/۱۰۳

در جدول (۶ .الف) براوردهای پسین β_i برای مدل لوژستیک خطی (LLM) و در جدول (۶. ب) براوردهایی برای مدل لوژستیک غیرخطی (NLM) برای طرح‌های مختلف نشان داده شده است. به همین ترتیب در جدول ۷. الف و جدول ۷. ب براوردهای پسین θ_j نشان داده شده است.

برای β_i تفاوت‌هایی بین سه حالت وجود دارد. در جدول (۶ .الف) تفاوت‌های بین براوردهای نقطه‌ای برای حالت‌های (۱) و (۲) ناچیز است، اما زمانی که با حالت (۳) مقایسه می‌شوند تفاوت‌های بزرگی وجود دارد. انحراف استانداردهای حالت (۳) بسیار کمتر هستند. در جدول (۶ .ب) الگوهای مشابهی وجود دارد. هر چند مشاهده می‌شود که در پایان، انحراف استانداردها برای مدل لوژستیک غیرخطی کوچک‌تر از انحراف استانداردهای مدل لوژستیک خطی هستند، که وقتی مقادیر بزرگ‌ترند این تفاوت‌ها کمتر است. برای حالت (۳) تفاوت‌ها بسیار بیش‌ترند و برای مدل لوژستیک غیرخطی تفاوت‌های بین حالت‌ها بیش‌تر است.

برای براورد θ_j تفاوت آشکاری بین مدل لوژستیک غیرخطی و مدل لوژستیک خطی وجود ندارد. تا زمانی که حالت‌های (۱) و (۲) بسیار شبیه به هم هستند، تفاوت‌های بزرگی بین حالت‌های (۱) و (۲) و حالت (۳) برای هر دو مدل لوژستیک خطی و مدل لوژستیک غیرخطی وجود دارد. انحراف استانداردها برای حالت (۲) دست کم دو برابر بزرگ‌تر از انحراف استانداردهای حالت (۳) هستند. برای مثال وقتی برای θ_8 مدل لوژستیک غیرخطی استفاده می‌شود افزایش بیش از دو برابر میانگین پسین و تقریباً افزایش هشت برابر در انحراف استاندارد پسین از حالت (۳) به حالت (۲) وجود دارد.

جدول۸ ـ مقایسه‌ی احتمال‌ای پسین که قلم غذایی i که$i = 1, \ldots, 12$ برحسب حالت و مدل بهترین قلم غذایی است

غذای اصلی		حالت ۱ (خطی)			حالت ۲ (غیرخطی)		
		۱	۲	۳	۱	۲	۳
۱	سوسیس خوک	۰/۱۷۶	۰/۱۶۱	۰/۰۹۷	۰/۲۲۷	۰/۲۱۴	۰/۰۵۳
۲	گوشت خوک و مرغ	۰/۰۰۲	۰/۰۰۲	۰/۰۰۴	۰/۰۰۴	۰/۰۰۴	۰/۰۲۱
۳	گوشت گوساله	۰/۰۴۱	۰/۰۴۱	۰/۰۷۳	۰/۰۰۸	۰/۰۱۳	۰/۰۷۶
۴	باربکیو گوساله	۰/۱۴۴	۰/۱۴۳	۰/۱۶۱	۰/۱۷۸	۰/۱۸۹	۰/۲۵۱
۵	کباب گوساله	۰/۱۴۱	۰/۱۴۹	۰/۱۶۸	۰/۰۸۰	۰/۰۷۷	۰/۱۱۱
۶	فرانک فورتر	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰
۷	بوقلمون	۰/۰۳۹	۰/۰۳۷	۰/۰۶۷	۰/۰۱۱	۰/۰۱۱	۰/۰۲۹
۸	گوشت آب‌پز گوساله	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰
۹	مرغ	۰/۱۷۷	۰/۱۸۱	۰/۱۰۳	۰/۲۷۳	۰/۲۷۰	۰/۰۷۲
۱۰	کله گنجشکی	۰/۰۰۶	۰/۰۰۶	۰/۰۰۸	۰/۰۰۷	۰/۰۰۷	۰/۰۱۲
۱۱	ژامبون	۰/۲۷۶	۰/۲۸۴	۰/۳۲۰	۰/۲۱۱	۰/۲۱۶	۰/۳۷۲
۱۲	آب‌پز گوساله با سس	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۳	۰/۰۰۱	۰/۰۰۷

در مورد بهترین قلم غذایی محاسبه‌ی احتمال‌های پسین

$$\psi_i = P\{\beta_i > \max(\beta_s, s = 1,2,\ldots,I, s \neq i) \mid n\}, \quad i = 1,2,\ldots,I$$

قضاوت می‌شود که به سادگی از تحلیل خروجی الگوریتم متروپلیس ـ هستینگس به دست می‌آید. بهترین قلم غذایی که انتخاب می‌شود غذایی است که بزرگ‌ترین ψ_iرا دارد. در جدول (۸) این احتمال‌ها برحسب مدل‌ها و حالت‌ها نشان داده شده است.

هنگامی که مدل لوژستیک خطی استفاده می‌شود ژامبون برحسب هر سه حالت با $16 \approx 100 \times (\frac{0/32}{0/0276} - 1)$ درصد افزایش در احتمال از حال (۱) به حالت (۳) برای انتخاب به عنوان بهترین پذیرفته می‌شود. اما وقتی که مدل لوژستیک غیرخطی مورد استفاده

قرار می‌گیرد برحسب حالت‌های (۱) و (۲) مرغ و برحسب حالت (۳) ژامبون به عنوان بهترین پذیرفته می‌شوند و احتمال به اندازه‌ی $۳۸ \approx ۱۰۰ \times (۱ - \frac{۰/۳۷۲}{۰/۲۷۰})$ درصد افزایش می‌یابد. افزون بر این تا زمانی که مدل لوژستیک خطی نشان می‌دهد که برای حالت‌های (۱) و (۲) مرغ و برای حالت (۳) کباب گوساله دومین غذای بهتر است مدل لوژستیک غیرخطی نشان می‌دهد که سوسیس خوک، ژامبون و گوشت گوساله به ترتیب برای حالت‌های (۱) و (۲) و (۳) دومین غذای بهتر هستند. به طور آشکار، باید از مدل غیرخطی با روش بیز کامل، به جای مدل خطی که ناسازگاری رتبه‌ها را لحاظ نمی‌کند، استفاده شود. زمانی که τ معلوم فرض می‌شود برای هر مدلی کم براورد بزرگی در تغییرپذیری وجود دارد.

ـ مقایسه‌ی الگوریتم متروپلیس ـ هستینگس با گیبس و لاپلاس

در این بخش با به کاربردن مدل دو جمله‌ای ـ بتا، نمونه‌گیر گیبس (جرج و همکاران، ۱۹۹۴) و تقریب لاپلاس (کاس و استفی[1]، ۱۹۸۹) را با الگوریتم متروپلیس ـ هستینگس مقایسه می‌کنیم. به ویژه از داده‌های توکسوپلاسم که حاوی شمارش‌های دو جمله‌ای تعداد افرادی است که برای توکسوپلاسم مورد آزمون قرار گرفته‌اند و تعداد افرادی که نتیجه‌ی آزمون آن‌ها مثبت بوده است، استفاده می‌شود. یک مجموعه از کمیت‌های موردنظر، نسبت افرادی است که در یک جامعه‌ی نامتناهی بزرگ نتیجه‌ی آزمون آن‌ها مثبت بوده است. نانـدرام و سدرانسک[2] (۱۹۹۳) روشی برای جامعه‌ی متناهی تعدادی بیمار که در طول سال گذشته دست کم یک بار توسط پزشک معاینه شده‌اند، بیان کرده‌اند. داده‌ها در سه ستون اول جدول (۹) نشان داده شده‌اند. ستون چهارم براوردهای ماکسیمم درستنمایی p_i را نشان می‌دهد (نسبت مثبت‌های نمونه برای شهر i ام).

---

[1] Kass and Steffey

[2] Sedransk

در واقع در این جا از مدل کاس و استفی (۱۹۸۹) استفاده شده است. ولی در مدل استفاده شده توسط جرج و همکاران (۱۹۹۴) فرض می‌شود که احتمال‌ها از یک توزیع پیشین $Beta(\alpha, \beta)$ با پیشین توأم روی ابرپارامترها به صورت $\pi(\alpha, \beta) \propto (\alpha+\beta)^{-2}$ که به جای $\alpha, \beta > 0$ در این جا $\alpha, \beta \geq 1$ آمده است. با استفاده از تبدیل توأم $\alpha = \mu\tau$ و $\beta = (1-\mu)\tau$ پیشین توأم به صورت $\pi(\tau, \mu) \propto \tau^{-1}$ یعنی $\tau \geq 2$ و $0 \leq \mu \leq 1$ مشخص می‌شود. توجه کنید که این مدل $\tau > 0$ را امکان‌پذیر می‌سازد. جرج و همکاران (۱۹۹۴) برای به دست آوردن لگ مقعر بودن توزیع‌های پسین شرطی α و β به منظور کاربرد در نمونه‌گیری رد و پذیرش گیلکز و وایلد (۱۹۹۲) و امکان‌پذیر ساختن روش نمونه‌گیری گیبس نیاز به فرض $\alpha, \beta \geq 1$ دارند.

در جدول (۹) میانگین‌های پسین p_i با هم مقایسه شده‌اند. برای سادگی کار این میانگین‌ها برای نمونه‌گیر گیبس از جدول (۱۳.۳) جرج و همکاران (۱۹۹۴) و روش لاپلاس از جدول (۲) کاس و استفی (۱۹۸۹) نشان داده می‌شود. میانگین‌های پسین به دست آمده از الگوریتم متروپلیس - هستینگس در ستون آخر جدول نشان داده شده است.

جدول۹ـ مقایسه‌ی تقریب لاپلاس، نمونه‌گیر گیبس و الگوریتم و متروپلیس ـ هستینگس برای داده‌های توکسوپلاسم در براورد نسبت مثبت‌ها برحسب شهر

شهر	مثبت	منفی	MLE	گیبس	لاپلاس	متروپلیس
۱	۲۴	۲۷	۰/۴۷۱	۰/۵۳۳	(۰/۰۶۷) ۰/۵۳۱	(۰/۰۶۰) ۰/۴۸۹
۲	۷	۹	۰/۴۳۸	۰/۵۳۹	(۰/۰۷۸) ۰/۵۳۵	(۰/۰۸۷) ۰/۴۹۵
۳	۴۶	۳۶	۰/۵۶۱	۰/۵۶۲	(۰/۰۴۵) ۰/۵۶۲	(۰/۰۵۰) ۰/۵۶۰
۴	۹	۴	۰/۶۹۲	۰/۵۸۱	(۰/۰۷۳) ۰/۵۸۴	(۰/۰۹۱) ۰/۶۱۰
۵	۲۳	۲۰	۰/۵۳۵	۰/۵۵۴	(۰/۰۵۶) ۰/۵۵۳	(۰/۰۶۶) ۰/۵۳۶
۶	۵۳	۲۲	۰/۷۰۷	۰/۶۱۷	(۰/۰۷۶) ۰/۶۲۵	(۰/۰۵۰) ۰/۶۷۷
۷	۸	۵	۰/۶۱۵	۰/۵۶۸	(۰/۰۶۷) ۰/۵۷۰	(۰/۰۹۲) ۰/۵۷۵
۸	۳	۷	۰/۳۰۰	۰/۵۲۶	(۰/۰۹۵) ۰/۵۲۰	(۰/۱۰۲) ۰/۴۵۰
۹	۱	۵	۰/۱۶۷	۰/۵۲۴	(۰/۰۹۸) ۰/۵۱۸	(۰/۱۰۹) ۰/۴۳۹
۱۰	۲۳	۱۴	۰/۶۲۲	۰/۵۷۹	(۰/۰۶۰) ۰/۵۸۰	(۰/۰۶۷) ۰/۵۹۸

در مرحله‌ی نخست براوردهای ابرپارامترها مقایسه می‌شوند. جرج و همکاران (۱۹۹۴) براوردهای ۲۰۱۶/۱ و ۱۵۶۶/۳ را به ترتیب برای α و β که براورد ۳۵۸۲/۴ را برای τ نتیجه می‌دهد، گزارش کرده‌اند. همچنین کاس و استفی (۱۹۸۹) براوردهای ۳۱/۳ و ۲۵/۰ را به ترتیب برای α و β به دست می‌دهد به طوری که براورد ۱۷/۰ برای τ با انحراف استاندارد ۳/۸ حاصل می‌شود. براورد μ برای سه روش مشابه است و برابر ۰/۵۵۶ برای روش لاپلاس، ۰/۵۶۳ برای گیبس و ۰/۵۴۱ برای متروپلیس ـ هستینگس است؛ براوردهای ماکسیمم درستنمایی نسبت‌ها برای هر ده شهر با فرض یک احتمال دو جمله‌ای مشترک، برابر ۰/۵۶۹ است. براورد τ مشکلاتی را در اجرای نمونه‌گیر گیبس نشان می‌دهد. ممکن است انتظار داشته باشیم که مقدار τ نزدیک برخی جمع‌بندی‌های

ده حجم نمونه باشد (برای مثال، ۳۵ برای میانگین یا ۲۷ برای میانه)، اما نمونه‌گیر گیبس براوردی از ۱۰۰ بار از چنین جمع‌بندی‌هایی را می‌دهد؛ نانـدرام و سدرانسک (۱۹۹۳) را ببینید. مشکل اصلی در مدل دو جمله‌ای ـ بتا داشتن یک براورد حساس از توزیع پسین τ است.

در مرحله‌ی دوم براوردهای نقطه‌ای p_i در نظر گرفته می‌شود. مطلوب است ببینیم که براوردهای الگوریتم متروپلیس ـ هستینگس در کل نزدیک به براورد ماکسیمم درستنمایی باشد. برای شهر ۵ براورد به طور دقیق برابر براورد ماکسیمم درستنمایی است. عامل انقباض مربوط به براوردگر نقطه‌ای p_i برای شهر i به شرط τ برابر $\tau/(n_i + \tau)$ است، که این مقدار تفاوت را شرح می‌دهد؛ τ برای الگوریتم متروپلیس ـ هستینگس کوچک‌تر است. اما برای شهر سوم براوردهای نقطه‌ای برای همه‌ی روش‌ها مشابه‌اند. پیش‌بینی می‌شود، احتمال می‌رود برای شهرهایی با شمارش‌های بزرگ مثلاً شهرهای ۳ و ۶ و شاید شهرهای ۱ و ۵ و ۱۰ کمی نیرو گرفته شود. ممکن است استفاده از براوردهای ماکسیمم درستنمایی به عنوان نقطه‌ی مرجع پذیرفته شود. تجربه نشان داده است که استنباط درباره‌ی p_i تاحدودی زیرنظر مشخص کردن τ به خصوص برای τ ی بزرگ قرار می‌گیرد. به نظر نمی‌رسد مقدار τ ی بزرگ برای نمونه‌گیر گیبس در براورد مقادیر p_i وقتی که با تقریب لاپلاس مقایسه می‌شود، تفاوتی ایجاد کند. در جدول (۹) براوردهای خطاهای استاندارد p_i (گزارش شده توسط کاس و استفی، ۱۹۸۹) و این براورد خطاهای استاندارد حاصل از الگوریتم متروپلیس ـ هستینگس (این براوردها توسط جرج، ۱۹۹۴ گزارش نشده است) نشان داده شده است.

جرج و همکاران (۱۹۹۴) از یک پیشین جایگزین برای ابرپارامترها استفاده کردند. آن‌ها با استفاده از نمونه‌گیر گیبس براوردهایی هستند که ما به دست آوردیم. این بار براورد τ ی آن‌ها ۴/۹ است. به جز براورد بزرگ τ برای نمونه‌گیر گیبس، تفاوت‌های موجود بین سه روش برای استنباط درباره‌ی p_i چشم‌گیر نیست.

نتیجه

نشان داده شد که محاسبه‌های مدل چند جمله‌ای بیز سلسله مراتبی چگونه انجام می‌شود. روشن است که اگر حالت (۲) (μ مجهول اما τ معلوم) رخ دهد مقدار خیلی کمی از بیز کامل از دست می‌رود. برای حالت (۳) (روش بیز تجربی کامل) در مقایسه با حالت (۱) (روش بیز کامل) کم براوردی چشم‌گیری از تغییرپذیری وجود دارد.

روش سلسله مراتبی چندین ویژگی دارد. نخست این که این روش برای هر تعداد ($I \geq ۲$) جامعه‌ی چند جمله‌ای و هر تعداد سلول ($J \geq ۲$) مفید است. دوم این که این روش برای جداول چند جمله‌ای با بدون رسته‌های ترتیبی به همان میزان مفید است. سرانجام جداول پراکنده را بدون مشکل می‌توان مورد استفاده قرار داد. هر چند برای محاسبه‌ها لازم است که برای هر i که $i = ۱, ۲, \ldots, I$ و $n_{ij} \geq ۰$ برای برخی jها که $j = ۱, ۲, \ldots, J$ و یکی از این شمارش‌های غیر صفر (ترجیحاً بزرگ‌ترین شمارش‌ها) برای آخرین سلول به کار رود.

همچنین از سه کاربرد برای نشان دادن توانایی این روش استفاده شد. برای کاربرد اول روی یک استنباط پیش‌بینی‌کننده‌ی بیز برای نسبت‌های جامعه‌ی متناهی با استفاده از شبیه‌سازی، کم براورد شدن تغییرپذیری نشان داده شد. تجربه نشان می‌دهد که برای جداول پراکنده کم براوردی بیش‌تری وجود دارد. همچنین با استفاده از مطالعه‌ی شبیه‌سازی نشان داده شد که روش یاد شده با استفاده از چگالی پیشنهادی برای مدل بیز کامل بهتر از روش جایگزین عمل می‌کند. برای کاربرد دوم روی انتخاب بهترین قلم غذایی براساس داده‌های حسی، دوباره کم براوردی تغییرپذیری پارامترهای به کار رفته در قضاوت در مورد بهترین قلم غذایی نشان داده شد. زمانی که θ_j براورد می‌شود کم براوردی بزرگی در تغییرپذیری وجود دارد، اما زمانی که β_j براورد می‌شود کم براوردی کمتری وجود دارد. همچنین مدل لوژستیک خطی و مدل لوژستیک غیرخطی مقایسه شد و دقت بیش‌تری حاصل شد وقتی که مدل لوژستیک غیرخطی برای براورد β_j به کار رفت، اما نه برای براورد θ_j. سرانجام در مثال سوم (داده‌های توکسوپلاسم) براوردهای p_i به دست آمده از الگوریتم متروپلیس ـ هستینگس بیش‌تر

از براوردهای به دست آمده از تقریب لاپلاس (کاس و استفی، ۱۹۸۹) و نمونه‌گیر گیبس (جرج، ۱۹۹۴) به براوردهای ماکسیمم درستنمایی نزدیک است.

ممکن است بتوان این روش را برای کاربرد مدل‌های سلسله مراتبی چند مرحله‌ای (نمونه‌گیری خوشه‌ای چند مرحله‌ای) با بیش از سه مرحله تعمیم داد. افزون بر این مسئله‌ای که در آن محدودیت‌های ترتیب بین p_{ij}ها وجود دارد بسیار مشکل است.

البته برای کاربردهای دارای متغیرهای کمکی، باید مدل‌های خطی تعمیم یافته‌ای به کار برده شوند که در آن‌ها توابع ربط (برای مثال، لوژستیک، لگ لگ مکمل و ربط t) مورد استفاده قرار می‌گیرند. اما چنین روش‌هایی اکنون برای کاربردهای بدون متغیر کمکی لازم نیست.

فهرست منابع:

Albert, J. H. and Gupta, A. K (1933). Estimation in contingency tables using prior information. J. Roy. Statist. Soc. B 45, 60-69.
Albert, J. H. and Gupta, A. K (1985). Beyesian methods for binomial data with applications to nonresponse problem. J. Am. Statist. Assoc. 80, 167-174.
J. Roy. Statist. Soc. B 45, 60-69.
Chen, M. H., Nandram, B. amd Rpss. E/ W/ (1996). Bayesian prediction of the shelf-life of a military raiton with sensory data. J. Agri. Bio. Envi. Stat. 1, 167-174.
Chib, S. and Greenberg, E. (1995). Understanding the metropois-hastings algorithm. Am. Statist. 49, 327-335.
Feller, W. (1968). An introduction to probability theory and its applications. New York. Wiley.
Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. J. Am. Statist. Assoc. 85, 398-409.
Gelman, A. E., Roberts, G. O. and Gilks, W. R. (1996). Efficient metropolis jumping rules. In Bayesian Statistics, 5, Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (Eds), Oxford, U.K, Oxford University Press, 559-607.
George, E. I., Markov, U. E. and Smith, A. F. M. (1994). Fully bayesian hierarchical analysis for exponential families via monte carlo computation. In aspects of uncertainty: A tribute to D. V. Lindley: Smith, A. F. M. and Freeman, P. (Eds). New York, Wiley, 181-199.
Giks, W. R. and Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. J. Roy. Statist. Soc. C 41, 337-348.

Kass, R. E. and Steffy, D. (1989). Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). J. Am. Statist. Assoc. 84, 717-726.

Lindley, D. and Smith, A. F. M. (1972). Bayes estimaties for the linear model (with discussion). J. Roy. Statist. Soc. B 34, 1-41.

Leonard, T. (1972). Bayesian methods for binomial data. Biometrika 59, 581-589.

Leonard, T. (1975). Bayes methods for two-way contingency tables. J. Roy. Statist. Soc. B 37, 23-37.

Leonard, T. (1977). Bayes simultaneous estimation for several multinomial distributions. Communications in statistics, Part A-Theory and methods, 6, 619-630.

Leonard, T. and Hsu, J. S. J (1994). The bayesian andalysis of categorical data-A selective reviw, In Aspects of Uncertainty: Atribute to D. V. Lindley: Smith, A. F. M. and Freeman, P. (Eds.) NewYork, Wiley, 283.-310.

Mccullagh, P. (1980). Regression models for ordinal data. (With discussin) J. Roy. Statist. Soc. B 42, 109-142.

Nandram, B. (1997) Beyesian inference for the best ordinal multinomial population in a taste test, In case studies in Bayesian statistics, Gatsonis, C., Hogges, J. S., Hass, R. E., McCulloch, R. and Singpurwalla, N.D. (Eds), Volume III, Springer Verlag, 399-418.

Nandram, B. and Sedransk, J. (1993). Bayesian predictiva inference for a finite population proportion: Two stage cluster sampling. J. Roy. Statist. Soc. B 55, 399-408.

Nandram, B. and Sedransk, J. and Smith, S. J. (1997). Order restricted Bayesian estimation fo the age composition of a population atlantic cod. J. Am. Statist. Assoc. 92, 33-40. Newell, G. J. (1982). Use of linear logistic model for the analysis of sensory evaluation data. Journal of Food Science. 47, 818-820.

Novick, M. R., Lewis, C. and Jackson, P. H. (1973). The estimation of proportions in m groups. Psychometrika. 38, 19-46.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). Annals of statistics. 22, 1701-1762.

Wong, G. Y. and Mason, W. M. (1985). The hierarchical logestic regression model for multi-level analysis J. Am. Statist. Assoc. 80, 513-524.

پیوست

A چگالی پیشنهادی برای چگالی پسین شرطی μ به شرط τ

با شروع با پیشین $\pi(\mu,\tau)=1$، توزیع پسین شرطی $\mu \mid \tau, n$ برابر است با

$$\pi(\mu \mid \tau, n) \propto \prod_{i=1}^{I}\prod_{j=1}^{J}\left\{\frac{\Gamma(n_{ij}+\mu_j\tau)}{\Gamma(\mu_j\tau)}\right\},\ \ 0<\mu_j<1,\ \ \sum_{s=1}^{J}\mu_s=1$$

با استفاده از فرمول استرلینگ، تقریبی برای $\pi(\mu \mid \tau, n)$ به دست می‌آید. در فرمول استرلینگ برای a بزرگ رابطه‌ی

$$\Gamma(a)\approx\sqrt{\frac{2\pi}{a}}\left(\frac{a}{e}\right)^{a}$$

برقرار است و همچنین فرمول استرلینگ بیان می‌کند که برای هر h رابطه‌ی زیر برقرار است:

$$\frac{\Gamma(a+h)}{\Gamma(a)}\approx a^{h}$$

$$\frac{\Gamma(a+h)}{\Gamma(a)}\approx h^{h}e^{-h}\left(1+\frac{h}{a}\right)^{a}\left(1+\frac{a}{h}\right)^{h}$$

به فیلتر (۱۹۶۸)، فصل ۲، بخش ۹ و تمرین ۲۲ مراجعه کنید.

با فرض بزرگ بودن $\mu_j\tau$ از کاربرد اول فرمول استرلینگ رابطه‌ی زیر حاصل می‌شود:

$$\frac{\Gamma(n_{ij}+\mu_j\tau)}{\Gamma(\mu_j\tau)}\approx(\mu_j\tau)^{n_{ij}} \qquad (1.A)$$

در حالی که از کاربرد دوم رابطه زیر به دست می‌آید:

$$\frac{\Gamma(n_{ij}+\mu_j\tau)}{\Gamma(\mu_j\tau)} \approx n_{ij}^{n_{ij}} e^{-n_{ij}} (1+\frac{n_{ij}}{\mu_j\tau})^{\mu_j\tau} (1+\frac{\mu_j\tau}{n_{ij}})^{n_{ij}} \qquad (2.A)$$

کـــه در آن $i=1,2,\dots,I$ و $j=1,2,\dots,J$ اســـت. بـــا اســـتفاده از رابطـــه‌ی $(1.A)$ می‌توان $\pi(\mu\mid\tau,n)$ را از شیوه‌ی رابطه‌ی زیر

$$p_1(\mu\mid n,\tau)=\frac{\prod_{j=1}^{J}\mu_j^{n_{\circ j}}}{D(a)},\quad 0<\mu_j<1,\quad \sum_{j=1}^{J}\mu_{ij}=1$$

تقریـــب زد کـــه در آن $D(a)=\prod_{j=1}^{J}\Gamma(a_j)/\Gamma(\sum_{j=1}^{J}a_j)$ و $a_j=n_{\circ j}=\sum_{i=1}^{I}n_{ij}$ و $j=1,2,\dots,J$ است. توجه کنید که $p_1(\mu\mid n,\tau)$ دارای توزیع مستقل از τ است و

$$E(\mu_j\mid\tau,n)=\frac{n_{\circ j}+1}{\sum_{s=1}^{J}(n_{\circ s}+1)}=\hat{\theta}_j,\qquad j=1,2,\dots,J \qquad (3.A)$$

با استفاده از رابطه‌ی $(2.A)$ می‌توان $\pi_1(\mu\mid n,\tau)$ را به صورت

$$p_2(\mu\mid n,\tau)\propto\prod_{i=1}^{I}\prod_{j=1}^{J}(1+\frac{n_{ij}}{\mu_j\tau})^{\mu_j\tau}(1+\frac{\mu_j\tau}{n_{ij}})^{n_{ij}},\quad 0<\mu_j<1,\quad \sum_{j=1}^{J}\mu_j=1$$

تقریب زد. حال با فرض $\mu_{(J)}=(\mu_1,\mu_2,\dots,\mu_{J-1})'$ می‌توان توزیع پســین $\mu_{(J)}\mid\tau,n$ را به وسیله‌ی

$$p_2(\mu_{(J)}\mid\tau,n)\propto\exp(D_1+D_2)$$

تقریب زد که در آن

$$D_1=\sum_{j=1}^{J-1}\sum_{i=1}^{I}\left\{\tau\mu_j\ln(1+n_{ij}/\tau\mu_j)+n_{ij}\ln(1+\tau\mu_j/n_{ij})\right\}$$

و

$$D_2=\sum_{i=1}^{I}[\tau(1-\sum_{j=1}^{J-1}\mu_j)\ln\{1+n_{iJ}/\tau(1-\sum_{j=1}^{J-1}\mu_j)\}+n_{iJ}\ln\{1+\tau(1-\sum_{j=1}^{J-1}\mu_j)/n_{iJ}\}]$$

و $0<\mu_j<1$ و $j=1,2,\dots,J-1$ است.

سپس $p_2(\mu_{(J)}\mid\tau,n)$ با استفاده از بسط مرتبه‌ی دوم سری تیلور چند متغیره حول $(\hat{\theta}_1,\hat{\theta}_2,\dots,\hat{\theta}_{J_1})$ در (A .۳) تقریب زده می‌شود. سپس با فرض این که k_1 و k_2 ثابت است

$$D_1\approx k_1+\sum_{j=1}^{J-1}(\mu_j-\hat{\theta}_j)A_j-\frac{1}{2}\sum_{j=1}^{J-1}(\mu_j-\hat{\theta}_j)^2B_j$$

و

$$D_2\approx k_2-A_J\sum_{j=1}^{J-1}(\mu_j-\hat{\theta}_j)-\frac{1}{2}B_J(\sum_{j=1}^{J-1}(\mu_j-\hat{\theta}_j))^2$$

که در آن

$$A_j=\tau\sum_{i=1}^{I}\ln(1+n_{ij}/\tau\hat{\theta}_j)$$

و

$$B_j=\tau\sum_{i=1}^{I}\left\{\hat{\theta}_j^{-1}-(\hat{\theta}_j+n_{ij}/\tau)^{-1}\right\}$$

و $j=1,2,\dots,J$ است. بنابراین $p_2(\mu_{(J)}\mid\tau,n)$ به وسیله‌ی

$$p_{۲}(\mu_{(J)} \mid \tau, n) \propto \exp[-\frac{۱}{۲}\{B_J\{\sum_{j=۱}^{J-۱}(\mu_j - \hat{\theta}_j)\}^{۲}$$

$$+\sum_{j=۱}^{J-۱}(\mu_j - \hat{\theta}_j)^{۲} B_j - ۲\sum_{j=۱}^{J-۱}(A_j - A_J)(\mu_j - \hat{\theta}_j)\}],\ \ ۰ < \mu_j < ۱ \qquad (۴.\ A)$$

تقریب زده می‌شود.

از رابطه‌ی $(۴.\ A)$ نتیجه می‌شود که $\mu \mid \tau, n$ به طور تقریبی دارای توزیع نرمال با میانگین

$$\hat{\mu}_j = \hat{\theta}_j + B_j^{-۱}(A_j - \overline{A})$$

است که در آن $\overline{A} = \sum_{j=۱}^{J} B_j^{-۱} A_j / \sum_{j=۱}^{J} B_j^{-۱}$ است و با فرض $\nu_j = B_j^{-۱} / (\sum_{s=۱}^{J} B_s^{-۱})^{۱/۲}$ کوواریانس آن به صورت

$$\operatorname{cov}(\mu_j, \mu'_j \mid n, \tau) = \begin{cases} B_j^{-۱} - \nu_j^{۲} & j = j' \quad \text{اگر} \\ -\nu_j \nu_{j'} & j \neq j' \quad \text{اگر} \end{cases} \qquad (۵.\ A)$$

است که $j, j' = ۱, ۲, \ldots, J$. چون $\hat{\mu}_j$ می‌تواند خارج از بازه‌ی $[۰,۱]$ قرار گیرد بنابراین می‌توان رابطه‌ی زیر را برقرار کرد:

$$\tilde{\mu}_j = \begin{cases} \hat{\mu}_j & ۰ < \hat{\mu}_j < ۱ \quad \text{اگر} \\ \hat{\theta}_j & \hat{\mu}_j \leq ۰ \quad \hat{\mu}_j \geq ۱ \quad \text{اگر} \end{cases}$$

سپس $E(\mu_j \mid n, \tau)$ به وسیله‌ی $\hat{\mu}_j$ تقریب زده می‌شود که در آن

$$\hat{\mu}_j = \tilde{\mu}_j / \sum_{s=۱}^{J} \tilde{\mu}_s, \quad j = ۱, ۲, \ldots, J \quad (۶.\ A)$$

چنان چه $\tau > 1$ باشد $\mathrm{cov}(\mu_j, \mu'_j \mid n, \tau)$ کاملاً تعریف شده است.
بالاخره یک تقریب دیریکله برای توزیع پسین شرطی $\mu \mid n, \tau$ مطلوب است. ماتریس کوواریانس توزیع دیریکله با پارامترهای $(\mu_1^{(a)}\tau^{(a)}, \mu_2^{(a)}\tau^{(a)}, \ldots, \mu_J^{(a)}\tau^{(a)})$ را با پارامترهای(A .۴) مساوی قرار می‌دهیم که در آن $\mu_j^{(a)} = \hat{\mu}_j$ و $j = 1, \ldots, J$ است. حال

$$d_{jj'} = \begin{cases} \mu_j^{(a)^2} / (B_j^{-1} - \nu_j^2) & j = j' \quad \text{اگر} \\ \mu_j^{(a)} \mu_{j'}^{(a)} / \nu_j \nu_{j'} - 1 & j \neq j' \quad \text{اگر} \end{cases}$$

را در نظر بگیرید که $j, j' = 1, 2, \ldots, J$ و قرار دهید

$$\tau^{(a)} = \{\sum_{j=1}^{J} \sum_{j'=1}^{J} d_{jj'} I_{jj'}\} / \sum_{j=1}^{J} \sum_{j'=1}^{J} I_{jj'} \quad (\text{A .۷})$$

که در آن $I_{jj'} = 1$ اگر $d_{jj'} > 0$ و $I_{jj'} = 0$ اگر $d_{jj'} \leq 0$ است.
پس $\mu \mid n, \tau$ به طور تقریبی دارای توزیع دیریکله با پارامترهای $(\mu_1^{(a)}\tau^{(a)}, \mu_2^{(a)}\tau^{(a)}, \ldots, \mu_J^{(a)}\tau^{(a)})$ است. توجه کنید که $\mu_j^{(a)}$ و $\tau^{(a)}$ توابعی از τ هستند.

B چگالی پیشنهادی برای چگالی پسین شرطی τ به شرط μ

با شروع توزیع پیشین $\pi(\mu, \tau) = 1$، توزیع پسین شرطی $\tau \mid \mu, n$ برابر است با

$$\pi(\tau \mid \mu, n) \propto \prod_{i=1}^{I} \left\{ \prod_{j=1}^{J} \frac{\Gamma(n_{ij} + \mu_j \tau)}{\Gamma(\mu_j \tau)} \right\} / \{\Gamma(n_{i\cdot} + \tau) / \Gamma(\tau)\}, \ \tau > 0$$

با استفاده از فرمول استرلینگ، $\pi(\tau \mid \mu, n)$ به وسیله‌ی

$$p_1(\tau \mid \mu, n) \propto \exp(E_1 + E_2) \quad (1.B)$$

تقریب زده می‌شود که در آن

$$E_1 = \sum_{i=1}^{I}\sum_{j=1}^{J}\{\mu_j \tau \ln(1 + \frac{n_{ij}}{\mu_j \tau}) + n_{ij}\ln(1 + \frac{\mu_j \tau}{n_{ij}})\}$$

$$E_2 = \sum_{i=1}^{I}\{\tau \ln(1 + \frac{n_{i\circ}}{\tau}) + n_{i\circ}\ln(1 + \frac{\tau}{n_{i\circ}})\}$$

و $n_{i\circ} = \sum_{j=1}^{J} n_{ij}$ برای $i = 1, 2, \ldots, I$ است.

سپس E_1 و E_2 به وسیله‌ی بسط مرتبه‌ی دوم سری تیلور یک متغیره حول τ_* تقریب زده می‌شوند که τ_* مد پسین $\pi(\tau \mid \mu, n)$ برای هر μ است. داریم

$$E_1 \approx k_1 + (\tau - \tau_*)\sum_{i=1}^{I}\sum_{j=1}^{J}\mu_j \ln(1 + \frac{n_{ij}}{\mu_j \tau_*})$$
$$-\frac{1}{2}(\tau - \tau_*)^2\sum_{i=1}^{I}\sum_{j=1}^{J}\mu_j^2\{\frac{1}{\mu_j \tau_*} - \frac{1}{n_{ij} + \mu_j \tau_*}\} \quad (2.B)$$

که در آن k_1 یک مقدار ثابت است. اولین گام مهم تقریب E_2 به وسیله‌ی

$$E_2^* = \sum_{i=1}^{I}\{\tau^* \ln(1 + \frac{n_{i\circ}}{\tau}) + n_{i\circ}\ln(1 + \frac{\tau}{n_{i\circ}})\}$$

است و با بسط E_2^* حول τ_* تا مرتبه‌ی اول رابطه زیر حاصل می‌شود

$$E_2 \approx k_2 + \frac{1}{2}(\tau - \tau_*)^2\sum_{i=1}^{I}\{\frac{1}{\tau_*} - \frac{1}{\tau_* + n_{i\circ}}\} \quad (3.B)$$

که در آن k_2 یک مقدار ثابت است. مد پسین یعنی τ_* به وسیله‌ی بهینه‌سازی به دست می‌آید (مثلاً الگوریتم نلدر ـ میلد). از روابط (B .۱)، (B .۲) و (B .۳) نتیجه می‌شود که میانگین و واریانس $\tau \mid \mu, n$ را می‌توان با استفاده از μ_* و σ_*^2 به صورت زیر تقریب زد که در آن

$$\mu_* = \tau_* + \sigma_*^2 \sum_{i=1}^{I}\sum_{j=1}^{J} \mu_j \ln(1 + \frac{n_{ij}}{\mu_j \tau_*})$$

و

$$\sigma_*^2 = \left[\sum_{i=1}^{I}\{(\frac{1}{\tau_*} - \frac{1}{\tau_* + n_{i_\circ}}) + \sum_{j=1}^{J}\mu_j^2(\frac{1}{\mu_j\tau_*} - \frac{1}{n_{ij} + \mu_j\tau_*})\}\right]^{-1} \qquad (B .۴)$$

چون $\tau > 0$ است، توزیع پسین $\tau \mid \mu, n$ به وسیله‌ی یک توزیع گاما با شاخص $\eta^{(a)}$ و مقیاس $\nu^{(a)}$ تقریب زده می‌شود. بنابراین با مساوی قرار دادن میانه‌ی کاندید با مقدار واقعی میانه یعنی τ_* نتیجه می‌شود که $(\eta^{(a)} - 1)/\nu^{(a)} = \tau_*$ است و با مساوی قرار دادن واریانس‌ها نتیجه می‌شود که $(\eta^{(a)} / \nu^{(a)})^2 = \sigma_*^2$ است. به این ترتیب به طور تقریبی

$$\tau \mid \mu, n \sim \Gamma(\eta^{(a)}, \nu^{(a)}) \qquad (B .۵)$$

به دست می‌آید که در آن

$$\eta^{(a)} = \left\{\frac{\tau_*}{2\sigma_*} + \sqrt{\left(\frac{\tau_*}{2\sigma_*}\right)^2 + 1}\right\}^2, \quad \nu^{(a)} = \sqrt{\eta^{(a)}} / \sigma_{*_\circ}$$

توجه کنید که $\eta^{(a)}$ و $\nu^{(a)}$ توابعی از μ هستند.